CENTRO UNIVERSITÁRIO DA FEI

GUILHERME ALBERTO WACHS LOPES

RECONHECIMENTO DE OBJETOS UTILIZANDO PERCEPÇÃO MULTISSENSORIAL COMPETITIVA BASEADA EM REDES COMPLEXAS

São Bernardo do Campo

2016

GUILHERME ALBERTO WACHS LOPES

RECONHECIMENTO DE OBJETOS UTILIZANDO PERCEPÇÃO MULTISSENSORIAL COMPETITIVA BASEADA EM REDES COMPLEXAS

Tese de Doutorado apresentada ao Centro Universitário da FEI para obtenção do título de Doutor em Engenharia Elétrica, orientado pelo Prof. Dr. Paulo Sérgio Silva Rodrigues.

São Bernardo do Campo
2016

Wachs Lopes, Guilherme Alberto.
Reconhecimento de Objetos Utilizando Percepção Multissensorial Competitiva Baseada em Redes Complexas / Guilherme Alberto Wachs Lopes. São Bernardo do Campo, 2016.
170 p. : il.

Tese - Centro Universitário FEI.
Orientador: Prof. Dr. Paulo Sérgio Silva Rodrigues.

1. Reconhecimento de Objetos. 2. Segmentação de Imagens. 3. Redes Complexas. 4. Visão Computacional. I. Silva Rodrigues, Paulo Sérgio, orient. II. Título.

Elaborada pelo sistema de geração automática de ficha catalográfica da FEI com os dados fornecidos pelo(a) autor(a).

CENTRO UNIVERSITÁRIO DA FEI

APRESENTAÇÃO DE TESE
ATA DA BANCA EXAMINADORA

Programa de Pós-Graduação Stricto Sensu em Engenharia Elétrica

Doutorado

PGE-10

Aluno: Guilherme Alberto Wachs Lopes **Matrícula:** 512303-9

Título do Trabalho: Reconhecimento de objetos utilizando percepção multissensorial competitiva baseada em redes complexas.

Área de Concentração: Processamento de Sinais

Orientador: Prof. Dr. Paulo Sergio Silva Rodrigues

Data da realização da defesa: 04/12/2015

ORIGINAL ASSINADA

Avaliação da Banca Examinadora

São Bernardo do Campo, / / .

MEMBROS DA BANCA EXAMINADORA	
Prof. Dr. Paulo Sergio Silva Rodrigues	Ass.: _____
Prof. Dr. Vagner Bernal Barbeta	Ass.: _____
Prof. Dr. Roberto Baginski Batista Santos	Ass.: _____
Profª. Drª. Denise Guliato	Ass.: _____
Prof. Dr. Nelson Delfino D'Ávila Mascarenhas	Ass.: _____

A Banca Examinadora acima-assinada atribuiu ao aluno o seguinte:

APROVADO ☐ REPROVADO ☐

VERSÃO FINAL DA TESE

ENDOSSO DO ORIENTADOR APÓS A INCLUSÃO DAS RECOMENDAÇÕES DA BANCA EXAMINADORA

_____

Aprovação do Coordenador do Programa de Pós-graduação

_____

Prof. Dr. Carlos Eduardo Thomaz

PREFÁCIO

Estudar o voo dos pássaros examinando as penas. Era esse o sentimento que pairava quando revisávamos a literatura atual de Visão Computacional para reconhecimento de objetos em cenas digitais. O grande tempo e trabalho dedicado pela literatura técnico-científica a esse tema sob essa abordagem corrobora com esse sentimento. Talvez o grande David Marr, quando propôs o "Mundo de Bloquinhos" para tentar entender melhor a enorme plasticidade de que é capaz o mecanismo de visão humana enquanto dedicado a nossa inconsciente tarefa de reconhecer um objeto familiar de quase infinitas perspectivas, tenha induzido a comunidade científica interessada em engenharia aplicada à processamento de imagens e inteligência artificial a focar inexoravelmente nesse túnel, ignorando quase que por completo as suas diversas aberturas ramificadas. David Marr nos deixou muito jovem, chocando e desanimando a comunidade científica interessada no tema, mas possivelmente também sem a chance de concluir e explicar melhor suas teorias. O que será que o próprio David Marr teria inventado mais se estivesse entre nós até hoje? Como será que seriam os modelos e aplicações atuais na área de Visão Computacional? Com certeza esse grande pesquisador precoce teria avançado muito pelo túnel que abriu, e possivelmente encontrado outras aberturas para mais túneis ramificados.

Paralelamente, é notável hoje em dia o que a comunidade de neurocientistas, psicologistas, neurofisiologistas e muitos outros pesquisadores da área médica, interessada no cérebro humano, tem descoberto a respeito desse enorme mecanismo de controle de quase tudo que há nesse mundo. Ao escrever isso, lembrei o que um jovem aluno de terceiro ciclo de Ciência da Computação me falou com propriedade e conhecimento altamente intuitivo, mas não menos racionalista, e em ton quase apocalíptico: "quem entender o cérebro humano controlará o mundo". Independente desse pensamento arrepiante, basta só dar uma passada geral pela literatura das ciências biológicas do cérebro para perceber o volume enorme de informações descobertas e postas à disposição da humanidade a respeito dessa ainda enigmática máquina de controle terráqueo. E para aqueles não muito familiarizados com esse campo fascinante da ciência, uma conversa rápida com seu neurologista perceberá o quanto ele sabe como seu cérebro funciona. Conhecimento descoberto e acumulado pela ciência mais nas últimas três décadas do que nos últimos séculos de pesquisa na área da medicina. Justamente porque talvez esse conhecimento tenha se acumulado rápido demais, ainda não houve tempo hábil para ser absorvido formalmente pela comunidade científica de inteligência artificial, engenharias e ciência da computação intencionada em aplicações práticas para o nosso dia-dia. Talvez não

seja a toa que foi nessas últimas três décadas que foram cunhados os nomes de novas áreas modernas de pesquisa: neurociência computacional, inteligência computacional, bioinformática, computação cognitiva, e muitas outras surgidas do cruzamento genético de ramos das ciências biológicas e exatas.

Ainda há muito trabalho a ser feito para acomodar harmoniosamente esses conhecimentos. Esforços dedicados de exuberantes performances como o Deep Learning, Redes Conexionistas e Redes Convolucionárias, com certeza estão no caminho certo, mas ainda precisando de ajuda para caminhar mais precisamente e mais rápido. Ideias ainda como mecanismos de atenção top-down, bottom-up, plasticidade, competição, colaboração e consciência, já são quase um consenso na área das ciências cerebrais, mas ainda temas de acalorados debates nas áreas das ciências exatas. Vale destacar que, recentemente, um artigo na influente revista Current Opinion in Neurobiology, opina claramente que desconsiderar processos de re-informação oriundas de áreas relacionadas a níveis de abstração mais altas que exercem um papel tão importante quanto os processos de informação forward, é um erro fatal nos modelos pensados para as peremptórias tarefas de reconhecimento de objetos. Eu me atrevo a ir mais longe, os modelos auto-intitulados bio-inspirados devem se preparar e abrir caminhos para anexação de mecanismos de re-informação oriundos, não somente das áreas mais altas nas vias visuais, mas também das muitas áreas funcionais e sensoriais, com o objetivo de integração homogênea e não heterogênea de forças convergentes para diversas tarefas paralelas, endógenas e exógenas. O rastro deixado pelos sinais visuais nas regiões retinianas, núcleo geniculado lateral, V1, V2 e ínfero-temporais são apenas as penas e plumas das aves participantes do fascinante voo que queremos entender.

A ideia de propor um meta-modelo que agregasse mecanismos funcionais recentemente descobertos na neurociência para reconhecimento de objetos em imagens digitais como tese de doutorado partiu do próprio Guilherme, possivelmente induzido sob o clima, nem sempre amistoso, de nossos debates e discussões quase diários sobre o assunto. Desde o início, sabíamos que enveredar por esse tema era o mesmo que pisar em terreno desconhecido; e o que é pior, andar em campo minado; afinal tratava-se de uma tese de doutorado, e como tal deveria ser aprovada pela comunidade científica e por uma banca. Mas, justamente por se tratar de uma tese de doutorado, ela deve, parafraseando Steve Jobs: “abrir fronteiras, e se possível construir pontes e acender fogueiras”. Tem que ousar. O que vem depois não pode ser mais ignorado, deve ser criticado, debatido e explorado, criando novos horizontes. Assim, nem sei o momento que foi decidido aceitar esse desafio. Quando percebemos, estávamos caminhando no túnel, ousando; ou, “dando a cara à tapa”, como disse o excepcional professor Roberto Baginsky, ao usar

seu tempo de arguição na banca. Embora cientes dos riscos, a única certeza de que tínhamos era de que não queríamos ousar menos. Então, caro leitor, após ler a tese com atenção, pode bater à vontade. A gente gosta. E foi nesse clima fascinante e de adrenalina, cientes também de nossas limitações e contexto local, que esta tese foi construída. E garanto que foi usado todo o conhecimento que acumulamos pessoalmente em mais de seis anos de parceria. O maior ganho? acho que foi e é mais pessoal. Para mim, particularmente como orientador, pude extravasar meu conhecimento e convicções científicas através de um aluno, e agora colega de profissão, que é ao mesmo tempo um profissional e colega de trabalho excepcional, pesquisador perspicaz e ser humano de valor inestimável, que é o Guilherme. Para o Guilherme, acho melhor perguntarem a ele ..

Paulo Sérgio Rodrigues
orientador, 15:05, 30/01/2016

A Deus e a minha amada família.

AGRADECIMENTOS

Pode-se dizer que o Doutorado é um momento de provação. Durante esse período, diversos desafios são enfrentados semanalmente: artigos devem ser publicados, erros de programação devem ser sanados, a literatura da área deve ser constantemente revisada, o trabalho deve ser documentado.

Esses são só alguns desafios que um aluno de Doutorado deve enfrentar. Claro que eu não consegui fazer esse trabalho sozinho. Muitas pessoas foram essenciais para a execução dessa Tese. Assim, esta seção é dedicada exclusivamente aos meus agradecimentos.

Primeiramente, dedico este trabalho a Deus. Há momentos de desânimo em que precisamos encontrar uma força extra para ir adiante. Nesse ponto, acho que Deus é o grande responsável por eu conseguir encontrar essa força.

Agradeço também à minha família, base de toda minha educação. Apesar de todo o estresse que passava, eles me entendiam e ajudavam sempre que podiam. Assim, agradeço ao meu pai, Pedro Lopes Crivelare, pela sua compreensão infinita que tem comigo; minha mãe, Helenice Ingrid Wachs, pelos conselhos (mesmo que duros, mas corretos) e pela transparência de que tudo dará certo; minha irmã, Gisele Ingrid Wachs Lopes, que deposita uma confiança enorme em minha capacidade que às vezes eu até acredito! Para todos os outros membros da família que colaboraram de maneira indireta: saibam que reconheço todos e que foram essenciais.

Há também duas pessoas que foram essenciais em minha conquista. A primeira delas é Tamires Ino Alves, mais que uma companheira durante esses seis últimos anos e responsável por me mostrar um amor incondicional, capaz de abrir mão de suas próprias vontades para me ajudar durante todo esse período. Tamires, espero ter você comigo para sempre! A segunda pessoa é a Sra. Mitsue Ino, avó de Tamires, no auge de sua sabedoria de vida e aos 91 anos, que constantemente me dá valiosas lições de vida.

Gostaria de agradecer também aos meus professores da Pós-Graduação da FEI: Dr. Plinio Thomaz Aquino Júnior, Dr. Flavio Tonidandel, Dr. Carlos Eduardo Thomaz, Dr. Paulo Eduardo Santos, Dr. Reinaldo A. da Costa Bianchi, Dr. Ivandro Sanches, Dra. Maria Claudia F. de Castro, Dr. João Chang Júnior e Dr. Fabrizio Leonardi, que me ensinaram muitos conceitos utilizados nesta Tese.

Claro que não poderia deixar de comentar toda a ajuda do pessoal da Secretaria de Pós-Graduação. Assim, gostaria de deixar registrado aqui o meu agradecimento. Obrigado Jorge

Ricardo Aguiar Mendes, Márcia Ferrareto de Jesus, Adriana Miguel Ramos e Daniele Couto de Abreu Neto.

Finalmente, mas não menos importante, vem meu orientador Prof. Dr. Paulo Sergio Rodrigues. Não há palavras para descrever a parceria que foi construída nesses quase dez anos de trabalho em conjunto. Com ele aprendi mais do que pode-se aprender em um Doutorado. Ele me passou muito conhecimento sobre a vida acadêmica e minha personalidade na profissão com certeza foi "esculpida" por ele. Contudo, não acho que essa sua primeira orientação de Doutorado foi como serão as outras. Acredito que ele me depositou uma confiança acima do normal e me escolheu como uma pessoa que irá continuar todo seu trabalho desenvolvido até aqui. Na finalização desta Tese, eu já desconfiava que o trabalho não acabaria aqui e já sentia a responsabilidade em ser a próxima geração para retransmitir seus ensinamentos. Prof. Paulo Sergio, sinta-se orgulhoso por sua conquista e saiba que eu reconheço seus esforços. Obrigado! A parceria continua ...

"When you see that trading is done, not by consent, but by compulsion – when you see that in order to produce, you need to obtain permission from men who produce nothing – when you see money flowing to those who deal, not in goods, but in favors – when you see that men get richer by graft and pull than by work, and your laws don't protect you against them, but protect them against you – when you see corruption being rewarded and honesty becoming a self-sacrifice – you may know that your society is doomed."
Ayn Rand

RESUMO

Na área de Visão Computacional, especificamente nas tarefas de Reconhecimento de Objetos, pode-se modelar os problemas principalmente de duas maneiras: orientada à engenharia ou orientada ao comportamento visual humano. A primeira maneira busca soluções mais eficientes do que um ser humano faria, e a segunda busca resolver problemas mais complexos. Desde o surgimento tanto da Ciência da Computação quanto da Neurociência, muitos aspectos neurais foram incorporados em modelos computacionais. Exemplos clássicos são as Redes Neurais Artificiais e modelos mais modernos como $HMAX$ e Redes Convolucionárias. No entanto, ainda não há na literatura de Ciência da Computação um Meta-Modelo funcional que inclua ao mesmo tempo importantes aspectos cognitivos. A presente Tese de Doutorado é a proposição de um Meta-Modelo Computacional, inspirado na biologia funcional do sistema visual neurológico, com flexibilidade incluindo módulos separados para o armazenamento do conhecimento visual aprendido, proposição de processos de baixo e médio nível para inferências de segmentacão e inclusão de novas características competitivas-colaborativas, além de diversas novas possibilidades, incluindo aprendizado infinito. Por outro lado, acompanhando resultados recentes da Neurociência, o Meta-Modelo proposto é estudado topologicamente do ponto de vista das Redes Complexas. Foram encontrados resultados similares aos da literatura para outros sistemas já bem fundamentados, tal como a linguística. Assim, a rede estudada sobre uma base de dados anotada apresenta comportamento de redes livres de escala, indicando a formação de "linguagem visual" baseada em contextos. Os experimentos foram conduzidos em um cluster acadêmico com poder de processamento de 20.6 TFlops com capacidade para até 2600 núcleos em paralelo, o que permitiu a redução do tempo computacional em até 90% do tempo que levaria em um caso de implementação serial.

Palavras-chave: Reconhecimento de Objetos, Segmentação de Imagens, Redes Complexas, Visão Computacional

ABSTRACT

In the Computer Vision area, specifically in Object Recognition tasks, one can model problems mainly in two approaches: oriented to engineering or oriented to human behavior. The first approach seeks more efficient solutions than a human would, and the second approach solves more complex problems. Since the emergence Computer Science and Neuroscience, many neural aspects have been incorporated into computational models. Classic examples are the Artificial Neural Networks and more modern models such as $HMAX$ and Convolutional Neural Networks. However, there is not yet in computer science literature an unique functional Meta-Model that includes important cognitive aspects simultaneously. This Doctoral Thesis is the proposition of a Computational Meta-Model inspired by the functional biology of neurological visual system including separate modules for storage of visual knowledge, proposition of low and mid-level processes for segmentation inferences and inclusion of new competitive-collaborative features, among many others new possibilities, including infinite learning. On the other hand, following recent results of Neuroscience, the proposed meta-model is topologically studied from the viewpoint of complex networks. Similar results were found to those reported for other systems already well founded, such as Human language. Thus, the network studied on an annotated database behaves as scale-free networks, indicating the generation of "visual language" based on contexts. The experiments were conducted in an academic cluster of 20.6 TFlops processing power for up to 2600 cores in parallel, allowing the reduction of the computation time by up to 90% of the time it would take in a case of serial implementation.

Keywords: Object Recognition, Image Segmentation, Complex Networks, Computer Vision

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A área de Visão Computacional tem sido uma das mais estratégicas dentro da Ciência da Computação, com aplicações em diversas outras áreas importantes e bem conhecidas, tais como: engenharia, física, biologia, medicina, astronomia, entre muitas outras. Por outro lado, o avanço da multimídia aliado à popularização da internet tem contribuído para ampliar o potencial dessa área, haja vista a quantidade de vídeos e imagens que todos os dias são despejados às "toneladas" por usuários comuns em todo o mundo, tornado nossas vidas mais fáceis, contudo dependentes do gerenciamento adequado desse volume de informação gerado. Um exemplo disso são os dados médicos que, ao mesmo tempo que facilitam a interpretação de diagnósticos e gerenciamento de exames, também desafiam o tráfego, armazenamento e gerenciamento dessas informações.

Da mesma forma, mesmo para aplicações voltadas ao entretenimento ou dados de informações governamentais, os avanços tecnológicos relacionados a vídeos e imagens podem ter benefícios imediatos muito vantajosos.

Sendo assim, o volume de informação multimídia demanda por sistemas de gerenciamento cada vez mais elaborados. Tradicionalmente, a área de Visão Computacional lida com esses problemas de duas maneiras. A primeira delas trata de problemas específicos sem a preocupação de que a solução ou o método utilizado seja inspirado em modelos biológicos. Geralmente, chamados de sistemas especialistas, esses modelos tendem a focar na resolução do problema sem a preocupação com o mecanismo de solução. Exemplos clássicos podem ser a análise do controle de qualidade em linhas de produção, a detecção de placas de veículos, o reconhecimento de digitais, íris ou faces.

A segunda maneira, mais utilizada em interpretação de objetos em cenas, procura solucionar problemas através de modelos inspirados em mecanismos biológicos, particularmente inspirados em modelos neurais de visão cognitiva, a medida que esses mecanismos também vão sendo compreendidos e propostos.

Um exemplo clássico bem conhecido foi a proposição das Redes Neurais Artificiais (R.N.A.) para reconhecimento de padrões. Quando surgiram, desde a década de 50, criou-se a expectativa de que tratava-se de um modelo representativo da própria estrutura cerebral humana. Sabemos hoje, no entanto, que as R.N.A. são um ferramental limitado considerando este ponto de vista.

Ao longo das últimas décadas, muitos modelos inspirados na biologia surgiram dentro da área da neurociência e psicologia para explicar os mecanismos da visão humana (veja o artigo de Khaligh-Razavi (2014) para um survey da área). Muitos desses mecanismos foram trazidos, totalmente ou em parte, para a área de Visão Computacional não somente com o objetivo de interpretar cenas ou reconhecer objetos, mas também interessados em validar os modelos propostos. Nesse sentido, a área de Visão Computacional obteve grande sucesso ao analisar características individuais, como cor, forma, textura ou movimento. No entanto, muito trabalho ainda deve ser feito na integração dessas características individuais. Por exemplo, sabe-se que cor, contraste e relacionamento espacial são informações fundamentais para interpretação de objetos em cena, mas não há um consenso geral sobre como realmente essas características cooperam entre si em um único mecanismo de interpretação de objetos ou cenas.

Existem muitos aspectos cognitivos relacionados ao processo de interpretação de objetos ou cenas que ainda são ensejo de profundo debate na área da neurociência e psicologia. Um exemplo são informações de contexto, que consideram objetos distribuídos na cena para inferência de outros objetos. Apesar da informação contextual ser um fator determinante na interpretação da cena, não há na área de Visão Computacional muitos trabalhos que explorem esses aspectos, nem como característica, nem como modelo matemático. Outros mecanismos ainda mal compreendidos são os mecanismos de visão Top-Down, Bottom-Up, movimento e foco.

Embora, hoje em dia, na área da Neurociência já haja uma melhor compreensão desses mecanismos ao que se sabia até a metade do século passado, poucos deles foram incorporados ou, pelo menos, considerados parcialmente na área de Visão Computacional para criação de modelos matemáticos visando a interpretação de cena e soluções de problemas de I.A. e Engenharia. O que se encontra na literatura científica da área de Visão Computacional geralmente é o estudo e a introdução individual desses mecanismos. Por exemplo, sistemas de *tracking* costumam se basear apenas no movimento ou características locais dos objetos para rastrear o alvo. Propostas mais recentes, como o SIFT ou Bag-Of-Words, apostam na mistura de características, mas sem considerar nenhum mecanismo neural.

Na última década, houveram grandes progressos na área da Neurociência Cognitiva que ainda não foram completamente absorvidos na área de Visão Computacional, visando não somente contribuir com aplicações na Ciência da Computação como também buscar confirmações nas Neurociências.

Duas importantes descobertas recentes graças a experimentos psicofísicos apoiados por avanços de ferramentas tecnológicas como IRM, IRMF, TC e TEP, foram o processo de atenção

precoce Broadbent (1970) e a disjunção de áreas corticais destinadas à interpretação de objetos e localização espacial Ungerleider e Mishkin (1982).

A primeira é área occipitotemporal (via ventral), destinada à percepção de objetos, e a segunda é a área occipitoparietal (ou via dorsal), destinada à percepção da localização de objetos. Leslie Ungerleider e Mortimer Mishkin mostraram que a separação colaborativa dessas duas vias e consequentemente suas funcionalidades são de fundamental importância para processos de alto nível de consciência cognitivo. Sem esse mecanismo, qualquer sistema de visão, seja ele biológico ou artificial, fica seriamente comprometido. Portanto, nas próximas décadas, a sua compreensão do ponto de vista biológico e sua modelagem do ponto de vista matemático-computacional, seria de fundamental importância para as duas áreas.

Por outro lado, a confirmação do processo de atenção precoce Broadbent (1970) jogou luz ao debate da visão *bottom-up* e *top-down*, iniciado na Psicologia e carregado para área de Visão Computacional cognitiva por pesquisadores interessados em confirmar fenômenos humanos ou criar aplicações na área de Inteligência Artificial que reproduzissem ao máximo comportamentos até então privilégios exclusivos de sistemas biológicos.

De uma maneira geral, — e para não entrar em detalhes que não são o foco deste texto — o processo de atenção precoce diz que informações ignoradas, no entanto destinadas ao córtex visual primário, são filtradas precocemente, em contrapartida com o processo de atenção tardia, que diz que um mecanismo cognitivo processado em altas áreas do córtex visual é o responsável pela filtragem de informações ignoradas Broadbent (1970). Essas ideias parecem estar de acordo com as definições de interpretação *bottom-up* e *top-down* de objetos em cenas. Embora seja um debate ainda longe de acabar, na área de Visão Computacional foi muito pouco explorado. Talvez porque haja, no entanto, a necessidade de um modelo matemático-computacional para explorá-lo.

Da mesma forma, a confirmação de que áreas diferentes do córtex cerebral lidam com problemas de reconhecimento (região do *o que?*) no córtex temporal e percepção espacial (região do *onde?*), no córtex parietal superior, parece afirmar que informações de naturezas diversas (exemplo: características espaciais e de identidade) competem e colaboram entre si para executar tarefas ditas primárias. Tais confirmações são sugeridas em relatos médicos bem conhecidos. Por exemplo, pacientes que tiveram um AVC no córtex parietal, são capazes de descrever objetos em cenas de seu campo visual, mas ao mesmo tempo, incapazes de definir suas localizações espaciais com precisão. De maneira inversa, quando o AVC atinge somente a área temporal, os pacientes podem ser capazes de localizar com precisão vários objetos que aparecem em seu campo visual sem, no entanto, conseguirem descrever o que são exatamente

esses objetos. Interessantemente, em alguns casos, quando o paciente é solicitado a tocar esses objetos, muitos são finalmente capazes de descrevê-los. Uma indicação clara de que o sensoriamento baseado em tato está cooperando para a percepção visual. Um outro caso intrigante descrito na literatura médica é de um paciente que, após um AVC, era incapaz de reconhecer o rosto de seus familiares muito próximos, até ouvir suas vozes (veja Gazzaniga, Ivry e Mangun (2002), Capítulo 5, para leitura de relatos).

Exemplos visuais e não-visuais de cooperação multi-sensorial inundam a literatura científica da Neurociência, no entanto levam décadas para serem incorporados como modelos matemáticos-computacionais na área de Visão Computacional. Mesmo que muitos desses mecanismos tenham sido tratados isoladamente, ainda há a necessidade de estudá-los de forma integrada, ainda que não os compreendamos por completo.

Por outro lado, a concepção e implementação desses mecanismos de interesse da área médica como modelos na área computacional demanda por ferramentas adequadas que permitam não somente aplicações comportamentalmente semelhantes aos mecanismos biológicos, mas também permitam fazer inferências científicas de maneira a ajudar neurocientistas a compreender ainda mais esses mecanismos.

Ao longo da última década, surgiram inúmeras ferramentas com esse propósito, destacando-se as Redes Complexas, pensadas na junção entre a Teoria dos Grafos e da Estatística, com aplicações em uma vasta gama de problemas físicos. Entre tais problemas, ressaltam-se as redes de inter-relacionamento social e biológicos, comunicação, transportes e interações químicas, além de fenômenos físicos, atômicos e espaciais, tais como turbulência e interação intra e intergalácticas. Em cada um desses contextos, as Redes Complexas foram capazes de melhorar a compreensão e até explicar fenômenos até então inalcançáveis para pesquisadores interessados.

Na área da linguística, outro ramo bastante explorado, recentemente surgiram diversas propostas que utilizam Redes Complexas para modelar o inter-relacionamento entre palavras de um texto em vários idiomas, nos diversos níveis léxico, sintático e semântico. Este último ainda demandando muito trabalho de pesquisa.

Há cerca de três anos, a linguística, modelada como uma Rede Complexa, tem fomentado intenso debate da comunidade científica. No artigo de Wachs-Lopes e Rodrigues (2015), um resumo desse debate é apresentado como introdução de um modelo de Redes Complexas para a análise dos idiomas Português e Inglês, do ponto de vista do dinamismo das redes em bases de dados específicas, contribuindo com importantes resultados na área.

O entusiasmo nas áreas de Neurociência e Redes Complexas, gerado por suas descobertas recentes, também tem entusiasmado pesquisadores interessados em modelar o cérebro

humano como uma grande rede neural, onde cada nó não é exatamente um neurônio, e a conexão entre eles não representa exatamente suas sinapses — ainda é extremamente difícil modelar a esse nível — mas sim cada nó é uma área particular do córtex que se inter-relaciona com muitas outras, representando assim arestas de um grande grafo. Exemplos desse tipo de modelagem podem ser encontrados em Goni et al. (2014), Heuvel e Sporns (2013), Sporns (2002). Recentemente, pesquisadores da Universidade de Yale publicaram um artigo Finn et al. (2015) sugerindo que as conexões nesse nível (baseado em regiões e não em neurônios) podem identificar unicamente um habitante do mundo, tal como as descobertas em impressões digitais e íris.

A modelagem em termos de regiões é uma abordagem importante e interessante. Mas outra alternativa, pouco explorada em áreas tecnológicas, é a modelagem em termos de blocos de atividades, onde blocos computacionais responsáveis por executar algoritmos ou ações se inter-relacionam com o objetivo de cooperar ou competir por estímulos ou sensores visando a execução de uma ou mais tarefas.

Em contrapartida, esse tipo de modelagem também produz Redes Complexas extremamente grandes, difíceis de compreender e principalmente computar. Assim, demandam estratégias específicas do ponto de vista computacional com o objetivo de ultrapassar a barreira nos limites de custos computacionais.

Assim, no presente trabalho de Tese de Doutorado, a tarefa de reconhecimento de objetos em uma cena de uma imagem é concebido como um Meta-Modelo composto de diversos blocos de atividades. Um desses blocos é uma Rede Complexas de inter-relacionamento de objetos, construída sob uma importante base de dados anotada: a SUN database Xiao et al. (2010).

Nessa rede, o modelo proposto é estudado sob o ponto de vista de diversas características cognitivas tradicionais e já bem conhecidas na Neurociência mas também em descobertas recentes (ou que são melhores compreendidas hoje). Essa rede representa informações contextuais, onde uma proposta para manipulá-la é apresentada. Também é apresentada uma proposta para modelagem do sensoriamento competitivo-colaborativo baseado em apenas três, mas importantes informações visuais — entre tantas que sabemos existir —: cor, orientação e área relativa dos objetos nas cenas, mas que é facilmente estendido para muitas outras.

O modelo proposto também é inspirado no comportamento de visão top-down, bottom-up e de atenção precoce e tardia, todos ainda em debate, cuja modelagem matemático-computacional simultânea pode ajudar a jogar luz sobre essa discussão.

1.1 OBJETIVO

Apresentar e validar um Meta-Modelo para reconhecimento de objetos em cenas naturais, que incorpore aspectos cognitivos atualmente estudados na Neurociência, mas ainda não plenamente incorporados como um único modelo matemático em Visão Computacional. O Meta-Modelo proposto é estudado experimentalmente de duas maneiras: através de uma base de dados anotada de imagens naturais, e através da modelagem de Redes Complexas.

1.2 PRINCIPAIS CONTRIBUIÇÕES DA TESE

Esta Tese apresenta um Meta-Modelo para reconhecimento de objetos utilizando o modelo de Redes Complexas para gerenciamento de informações contextuais para reconhecimento de objetos em cenas. Dentre as contribuições desta Tese, destacam-se:

a) Incorporação de alguns aspectos da Neurociência que influenciam no reconhecimento de objetos em cena, e que foram pouco abordados na área de Visão Computacional, de maneira integrada.

b) Modelagem contextual de objetos utilizando a co-ocorrência entre diferentes características extraídas a partir de imagens.

c) Um modelo único que incorpora os modelos de visão top-down/bottom-up, competição/colaboração entre características extraídas e atenção precoce e tardia a partir de imagens.

d) Estudo de novas características extraídas de imagens e suas influências sobre o reconhecimento de objetos.

e) Utilização dos conceitos de Redes Complexas para estudar e analisar a organização topológica da estrutura aprendida pelo modelo.

1.3 ORGANIZAÇÃO DA TESE

Os items abaixo descrevem a ordem e um resumo dos principais assuntos abordados na Tese:

Conceitos Fundamentais: Neste capítulo são abordados os principais conceitos utilizados no trabalho, bem como os trabalhos relacionados. Inicialmente, são apresentados a

área de Visão Computacional e os modelos de representação de imagens. Alguns tópicos de aprendizagem de máquina são vistos na Seção 2.2. Posteriormente, na Seção 2.3 são apresentados alguns aspectos da Neurociência que serviram de inspiração para esta Tese. Alguns modelos tradicionais de reconhecimento de objetos inspirados biologicamente são vistos na Seção 2.4. Finalmente, na Seção 2.5 alguns modelos e propriedades físicas são apresentados.

Proposta: Neste capítulo, é apresentada o Meta-Modelo bem como a base de dados utilizada para os experimentos. O capítulo é dividido nos 3 módulos integrantes do modelo.

Experimentos Discussão: Este capítulo apresenta a metodologia de experimentação, os resultados bem como as discussões dos resultados obtidos.

Conclusões: Finalmente, este capítulo resume os achados nos experimentos, destaca alguns pontos não considerados no modelo e direciona trabalhos futuros.

2 CONCEITOS FUNDAMENTAIS

2.1 VISÃO COMPUTACIONAL E MODELOS DE RECONHECIMENTO DE OBJETOS

Visão Computacional é um ramo da Inteligência Artificial que surgiu em meados dos anos 60 com o problema de reconhecimento de textos a partir de folhas impressas. Sua crescente utilização em ambientes industriais tem gerado novos desafios e chamou a atenção de muitos cientistas. Com o passar do tempo, novas aplicações foram desenvolvidas utilizando estudos nessa área. Exemplos dessas aplicações são: reconhecimento de chips eletrônicos, sistemas de controle de qualidade, sistemas para análise de imagens médicas, sistemas de segurança automotivos, sistemas de rastreamento de objetos, entre muitos outros.

Um sistema de visão computacional deve extrair de uma imagem, ou sequência de imagens, informações capazes de descrever seu conteúdo, propriedades, forma, iluminação, distribuição de cores entre outras características Szeliski (2010) relevantes para o problema a ser tratado. Essa definição mostra o quanto genérico, complexo e abrangente é esse campo de pesquisa.

Há décadas, cientistas, tanto da área da psicologia quanto da área da computação, tentam entender como o sistema visual humano trabalha com o objetivo de criar um modelo fidedigno. Porém, sua complexidade vai além da *percepção* do mundo como o vemos, envolvendo também a *cognição* humana. Por isso, muitos pesquisadores acabam subestimando os problemas de visão computacional e adotam heurísticas ou até mesmo diminuem o escopo do projeto a fim de ter um resultado aceitável sem tanta complexidade no modelo computacional.

Dos diversos tipos de problemas tratados em Visão Computacional, podemos generalizá-los em duas categorias: problemas em ambientes controlados e problemas genéricos Andreopoulos e Tsotsos (2013). O primeiro tipo é mais abordado em casos onde é definida uma tarefa restrita, tais como: controle de qualidade de produtos industriais, identificação de chips eletrônicos, leitura de digitais, reconhecimento óptico de caracteres, inspeção industrial, fotometria para geração de objetos 3D, análise de imagens médicas, segurança automotiva e vigilância. Os sistemas de visão para tais tarefas são superiores a humanos, uma vez que são mais rápidos e construídos especificamente para atingir um objetivo específico. Por outro lado, os sistemas de visão computacional genéricos envolvem meta-modelos e vão ao encontro das questões de análise contextual e semântica. Neste caso, ainda não se conhece um modelo matemático que consiga armazenar e processar informações tal como os humanos.

Os sistemas de visão computacional genéricos são mais difíceis de serem modeladas porque lidam com o *problema inverso* Szeliski (2010); isto é, devemos procurar algo desconhecido considerando informações insuficientes para especificar totalmente a solução.

Apesar dos avanços da área nos últimos 50 anos, ainda não há um modelo definitivo capaz de interpretar e extrair informações básicas de uma imagem, tal como faz uma criança Szeliski (2010), Palmer (1999), Marr (1982). Uma tarefa simples como contar o número de pessoas em um ambiente não controlado ainda é uma realidade complexa, mesmo para novas tecnologias.

Dentre os problemas encontrados na área, mais especificamente em ambientes não controlados, destaca-se o de reconhecimento de objetos. Segundo Prasad (2012), Dickinson (1999), reconhecimento de objetos lida com a identificação da presença de diversos objetos em uma imagem, sendo também uma das principais tarefas do sistema visual humano. Porém, para ambientes não controlados o problema ainda é um desafio, com muitos pontos em aberto.

O reconhecimento automático de objetos é uma tarefa que envolve frequentemente execução em tempo real. Nessa tarefa, erros podem ser toleráveis até uma especificação, mas o mais importante é que seja genérico o bastante para ajudar a resolver problemas muitas vezes intratáveis[1]. Geralmente, esses problemas estão relacionados à busca de uma solução em um espaço n-dimensional. No domínio de imagens e vídeos, esses espaços de buscas são grandes o bastante para tornar algoritmos de força bruta inviáveis. Uma possível solução pode ser a escolha de uma meta-heurística, necessária para guiar os algoritmos na busca por uma solução ideal.

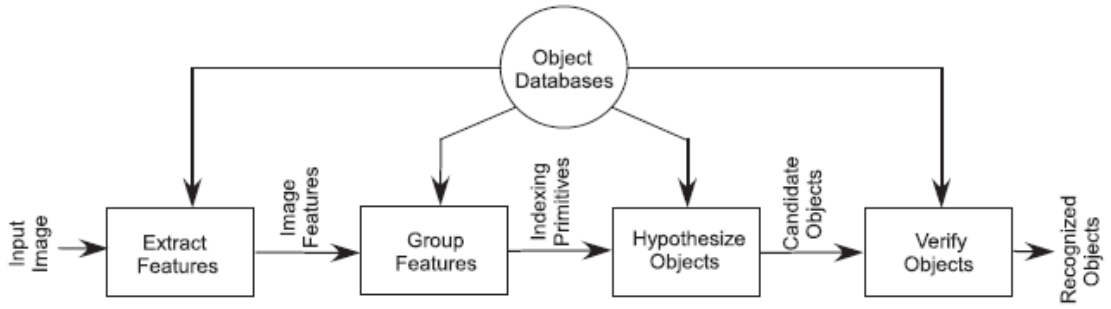
O trabalho de Prasad (2012) sugere que a acuracidade dos modelos atuais não são satisfatórios, alcançando somente cerca de 16% de acerto sobre 22.000 categorias de objetos no caso do *Google Glass*.

O trabalho de Andreopoulos e Tsotsos (2013) faz um estudo sobre a área de reconhecimento de objetos nos últimos 50 anos e as técnicas e modelos tradicionais. Além disso, o trabalho identifica as principais tarefas envolvidas no processo.

Como mostrado na Figura 1, um sistema de reconhecimento de objetos tem como entrada uma imagem digital. Uma vez adquirida a imagem, o próximo passo é extrair uma ou mais características ou *features*. Esse processo transforma a imagem de entrada em estruturas que capturam informações de cor, forma e textura, tais como: arestas (descontinuidades de brilho e/ou cor), cantos, regiões (áreas com cores homogêneas). Além disso, o objetivo é extrair

[1]Entenda como "intratável" um problema que não pode ser solucionado em tempo polinomial.

Figura 1 – Componentes básicos de um típico sistema de reconhecimento de objetos Andreopoulos e Tsotsos (2013)



Fonte: Autor “adaptado de” Andreopoulos e Tsotsos (2013)

características que serão suficientes para possibilitar a localização e identificação dos objetos de interesse.

As características extraídas de uma imagem são então processadas e agrupadas como primitivas de indexação na etapa “Group Features”. Essas primitivas têm a mesma função de uma linguagem de consulta em um banco de dados. Tome como exemplo um sistema de reconhecimento de aeroportos a partir de imagens de satélite. Neste caso, as primitivas de indexação podem ser as linhas que delimitam as pistas.

De posse das primitivas, o sistema de reconhecimento deverá buscar em uma base de conhecimento por algum objeto que contém a mesma primitiva. Esse processo é nomeado como *matching algorithm* ou algoritmo de casamento. O resultado desse processo é um conjunto de objetos candidatos.

Contudo, se o conjunto de objetos candidatos for maior que um, o sistema deverá escolher o objeto com maior semelhança. A esse processo damos o nome de *evaluation* ou avaliação. Nesse processo, deve-se escolher um candidato por métodos de ordenação baseado em medidas heurísticas. Essas medidas dão uma nota para cada objeto candidato. O candidato que tiver a maior nota será o objeto escolhido.

Do ponto de vista da Neurociência, o processo de reconhecimento de uma cena, ou objetos em uma cena, pelo sistema visual humano, pode ser visto de duas maneiras: *top-down* ou *bottom-up*. Ainda não há um consenso geral sobre qual a estratégia deve ser utilizada, ou mesmo se ambas devem ser utilizadas. No processo *top-down*, considera-se o reconhecimento de uma cena de forma abstrata antes de serem identificados os componentes principais. Por outro lado, o processo *bottom-up* considera-se o inverso. Inicialmente, características isoladas são identificadas, juntando-as para formar regiões e, finalmente, cria-se uma cena abstrata.

Na literatura de Visão Computacional, é comum dividir as estratégias em três principais níveis: baixo, médio e alto-nível, descritos a seguir:

a) Baixo nível: onde se trabalha com os *pixels*, elementos individuais de textura ou pequenas regiões. Algoritmos típicos ou métodos para lidar com esse nível são: filtros espaciais ou filtros de frequência, conversores de formatos, restauradores de artefatos como ruídos, efeitos indesejados, compressão de informações e esterografia. De forma geral, algoritmos de baixo nível normalmente recebem como entrada uma imagem e retornam como saída uma outra imagem.

b) Médio nível: onde trabalha-se com regiões ou agrupamento de regiões. Algoritmos típicos para lidar com problemas nesse nível pode ser: segmentadores de imagens, métodos morfológicos, clusterizadores de regiões, restauradores de imagens baseados em regiões, algoritmos de inter-relacionamento de regiões para descrição da estrutura da cena, entre outros. O processamento de médio nível normalmente recebe uma imagem como entrada e devolve atributos, tais como: bordas, contornos e regiões com objetos de interesse.

c) Alto nível: neste nível, trabalha-se com a cena como um todo, considerando as regiões e objetos como sendo seus elementos cognitivos. Algoritmos típicos para lidar com problemas de médio nível são: classificadores, localizadores de objetos em cenas, rastreamento de objetos ou classificadores de cena.

Uma referência didática, que descreve a área em três níveis, pode ser encontrada em GONZALEZ e WOODS (2002).

Dentro dessa ideia de divisão do processo de visão computacional, vários autores têm proposto diferentes trabalhos e modelos de reconhecimento, tanto em sentido *top-down*, quanto em sentido *bottom-up*.

Um exemplo de arquitetura *bottom-up* é encontrado em Lowe (1987), Lowe (1999), Huttenlocher e Ullman (1990). Em Lowe (1987), o autor cria um modelo probabilístico para encontrar características locais, como linhas, que pertencem a um mesmo objeto. Após essa localização há um processo de união desses elementos em uma estrutura de dados. Essa transformação relaxa o objeto que se quer encontrar na imagem e torna o processo de busca dos objetos mais eficiente. O autor ainda lembra que a identificação de grupos estáveis (sem ruído, ou elementos acidentais) na imagem é uma tarefa típica do sistema visual humano. Contudo, o modelo proposto pelo autor é restrito a objetos primitivos e mais simples de detectar. Assim, para identificar uma face humana, por exemplo, seria necessário acrescentar novos módulos ao reconhecedor.

Em Khan, Weijer e Vanrell (2009), os autores exploram a modelagem *top-down* a partir de informações que relacionam cor com objetos; isto é, a coloração das regiões de uma imagem ajuda o sistema a encontrar regiões onde há objetos. Após esse processo, novas características são extraídas das regiões delimitadas para serem classificadas. Contudo, o modelo sugerido não contém parâmetros de ajuste, fazendo com que, para algumas classes de objetos, o reconhecimento não tenha sido satisfatório. A falta de parâmetros pode impedir o uso de classificadores supervisionados (veja Seção 2.2).

Quando se fala em reconhecimento de objetos, estamos nos referindo a modelos matemáticos capazes de *classificar* como região de interesse um conjunto de dados obtidos a partir de uma imagem digital. Contudo, pesquisadores sempre buscam modelos mais representativos e capazes de reconhecer diversas classes de objetos. A compreensão desses modelos é estudada na área de aprendizagem de máquina. Portanto, o estudo de aprendizagem de máquina em conjunto com classificadores é necessário quando se deseja criar um modelo de reconhecimento de objetos.

2.1.1 Modelos de Representação de Imagens e Características

Do ponto de vista do sistema visual humano, existem muitas características que podem ser utilizadas para a interpretação de um objeto ou cena. No entanto, somente um sub-conjunto delas é utilizado para aplicações computacionais. Entre elas, as mais populares envolvem a representação de luminância, cor, forma, textura e movimento. O principal motivo é que são as mais simples de implementar e compreender.

2.1.1.1 Intensidades

Para imagens digitais binárias, duas formas de representação de luminância são essenciais. A primeira delas é através de uma função $f : (x,y) \rightarrow \{0,1\}$, onde x,y são as coordenadas para um ponto da imagem. A imagem dessa função é formada pelos valores de luminância 0 para preto e 1 para branco. Para imagens em tons de cinza, essa função é dada por $f : (x,y) \rightarrow \mathbb{R}$, sendo capaz de representar valores entre preto e branco, sendo esta uma maneira tradicional de representação de imagens sempre que é necessária sua exibição ou leitura de valores individuais de luminância.

Por outro lado, uma segunda maneira de representar imagens digitais em tons de cinza é através da representação como histograma. A função correspondente a essa representação é

dada por $f : k \rightarrow \mathbb{R}^+$, onde k é um valor de luminância e a imagem da função é a quantidade de *pixels* que têm k como valor de luminância. Uma conclusão imediata desse modelo de representação é que as informações de localização espaciais dos *pixels* são perdidas. Contudo, essa técnica é utilizada em aplicações cujas informações espaciais não são relevantes, tais como segmentação, ajuste de contraste, ajuste de brilho, entre outras.

2.1.1.2 Cores

Para representação de imagens coloridas, é necessário primeiramente definir um sistema de cores, que é um espaço multidimensional que as descreve. Existem diversos sistemas com este objetivo, contudo dois deles se destacam na área de Visão Computacional: RGB e HSV. O sistema RGB (Red, Green, Blue) é composto por três componentes $(r,g,b) \in \mathbb{R}^3$, no qual cada uma descreve a intensidade de cada canal de cor primária. Por outro lado, o sistema de cores HSV é uma transformação não-linear do sistema RGB, normalmente utilizado em aplicativos de edição de imagens por serem mais intuitivos e perceptivos ao sistema visual humano. O sistema HSV é formado também por três componentes (h,s,v), matiz, saturação e valor, respectivamente.

Uma função de representação de imagens digitais coloridas f é feita utilizando um sistema de cores de tal forma que $f : (x,y) \rightarrow \mathbb{R}^d$, onde d é o número de componentes do sistema. De maneira semelhante às imagens em tons de cinza, também é possível representar uma imagem digital colorida através de seu histograma de cores. Em Wachs-Lopes, Fukuma e Rodrigues (2012), os autores criam um sistema de detecção de tomadas em vídeos de futebol. O objetivo do trabalho é classificar a cena como principal ou secundária. Sendo assim, foi utilizado histogramas gerados a partir do sistema HSV e a teoria da informação como uma maneira de comparar as semelhanças entre esses histogramas. Para isso, os autores discretizaram o sistema HSV possibilitando 18 valores para matiz, 3 valores para saturação e 3 valores para valor, gerando um histograma de 162 entradas. Para cada imagem, os autores as descreveram através dos histogramas. Em seguida, foi gerado um histograma médio sobre o conjunto de histogramas gerados a partir de tomadas principais supervisionadas. De posse do histograma médio das cenas de tomada principal, os autores utilizaram a divergência de Kullback-Leibler com uma cena de tomada desconhecida. Um limiar foi então definido para determinar se a tomada desconhecida era principal ou secundária. Os resultados mostraram que o histograma HSV foi capaz de classificar com as tomadas como principal ou secundária com 97% de acerto. Esse trabalho mostra que a informação de frequência foi suficiente para classificar as tomadas.

2.1.1.3 Texturas

A característica de textura em imagens é outro domínio estudado na área de visão computacional para representação de imagens. Ao contrário da característica de cor, a textura de uma imagem é dada a partir de uma região e não somente sobre um ponto. A textura é uma característica que geralmente descreve a disposição espacial de cores e padrões de intensidades em uma imagem inteira ou em uma região específica. Três modelos de representação de imagem a partir da característica de textura serão apresentados nesta tese: baseado em estatística, baseado em fatores psicológicos e utilizando processamento de sinais.

A representação de imagens utilizando características estatísticas de textura pode ser feita através de uma matriz de co-ocorrência P gerada a partir do relacionamento entre as vizinhanças dos *pixels* Howarth e Rüger (2004a). Cada elemento $p_{i,j}$ dessa matriz contém a frequência relativa entre dois *pixels* vizinhos, um com luminância i e outro com luminância j, de acordo com uma regra de vizinhança estabelecida previamente. Contudo, a determinação da vizinhança de um pixel envolve duas constantes: a distância d e o ângulo θ. Assim, uma matriz de co-ocorrência gerada com $d = 1$ e $\theta = 0^o$ será capaz de representar as relações de luminância entre *pixels* esquerdos/direitos. O trabalho de Haralick (1979), propõe a utilização de algumas funções aplicadas sobre a matriz P com o objetivo de extrair informações de textura. Algumas dessas equações são mostradas abaixo:

Energia	$\sum_{i,j} p_{i,j}^2$
Entropia	$\sum_{i,j} p_{i,j} \log p_{i,j}$
Probabilidade Máxima	$\max_{i,j} p_{i,j}$
Homogeneidade	$\sum_{i} \sum_{j} \frac{p_{i,j}}{1+\lvert i-j \rvert}$

Uma segunda abordagem para representação de imagens utilizando características de texturas foi proposta em Tamura, Mori e Yamawaki (1978). Na área de Visão Computacional, essa abordagem é conhecida como método de *Tamura*. Para essa representação, os autores se basearam no sistema visual humano para definir algumas características extraídas das imagens, na qual três delas obtiveram resultados semelhantes quando comparadas a humanos. A primeira característica é denominada *coarseness* (rispidez), que é responsável pela quantificação do tamanho do menor elemento da textura. Essa característica é obtida primeiramente extraindo-se

as cores média de cada pixel sobre uma vizinhança de tamanho variável, segundo a Equação (1)

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} \frac{f(i,j)}{2^{2k}} \tag{1}$$

onde k é o tamanho da vizinhança e f é a função da imagem. Os valores de A_k para cada pixel da imagem são então comparados com outros *pixels* que estão na fronteira vertical e horizontal. A Equação (2) mostra a comparação das médias na projeção horizontal.

$$E_{k,hor}(x,y) = |A_k(x + 2^{k-1},y) - A_k(x - 2^{k-1},y)| \tag{2}$$

A escolha de k que maximiza $E_{k,hor}$ e $E_{h,ver}$ para um determinado pixel determinará o valor de sua rispidez. Para obter a rispidez de uma região da imagem, extrai-se a média da rispidez sobre todos os *pixels* dessa região.

A segunda característica extraída pelo método de *Tamura* é o contraste. Essa característica é calculada em termos do desvio padrão e da curtose do histograma de cores da imagem, segundo a Equação (3),

$$contraste = \frac{\sigma}{(\alpha_4)^n} \tag{3}$$

onde σ^2 é a variância, n é uma constante empiricamente definida como $1/4$ e α_4 é a curtose, calculada através da Equação (4).

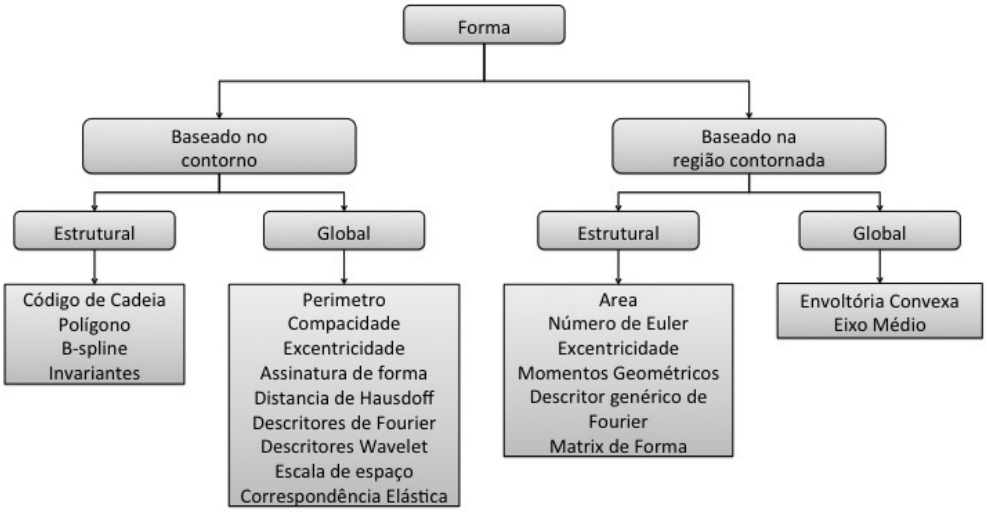
$$\alpha_4 = \frac{\frac{1}{XY} \sum_x \sum_y (f(x,y) - \mu)^4}{\sigma^4} \tag{4}$$

onde σ^2 é a variância das luminâncias dos *pixels*, μ é a média das luminâncias e X,Y são respectivamente a largura e altura da imagem.

A terceira característica extraída pelo método de *Tamura* é a direção. Essa característica utiliza o operador *Prewitt* para detectar a direção das bordas da imagem. Um histograma contendo as frequências de cada ângulo é gerado e utilizado como descritor.

Outro método empregado na representação de imagens utilizando características de textura é calculado através do filtro de *Gabor*, que permite a filtragem de imagens no domínio espacial e da frequência através da transformada *Wavelet* Howarth e Rüger (2004b). A transformada *Wavelet* é semelhante à transformada de *Fourier*, porém com um *kernel* que está localizado tanto no espaço de *Fourier* quanto no espaço Real. O filtro de *Gabor* permite compor um vetor de características que armazena informações de orientação e frequência através da utili-

Figura 2 – Taxonomia das características de forma apresentadas em Zhang e Lu (2004)



Fonte: Autor "adaptado de" Zhang e Lu (2004)

zação de diversas configurações de escala e angulação. Os valores obtidos nesse vetor podem ser utilizados como um modelo de representação de imagens.

2.1.1.4 Topologias e Formas

Outra característica utilizada para a representação de imagens é a topologia da forma. Para a utilização dessa característica é necessário primeiramente definir uma área de interesse sobre a imagem original utilizando, por exemplo, um segmentador. As características de forma podem ser classificadas em dois grupos: baseada no contorno e baseada na região delimitada. A diferença entre essas características está em determinar se as informações serão extraídas a partir da linha de contorno ou sobre toda a área delimitada pelo contorno Zhang e Lu (2004). As características de forma também podem ser classificadas em globais ou estruturais. Características estruturais descrevem a forma a partir do próprio domínio espacial e as características globais descrevem a forma através de uma transformação. A Figura 2 mostra a taxonomia das características de forma que são descritas em Zhang e Lu (2004). Dentre as características apresentadas nessa figura, três se destacam na área de visão computacional: descritores de Fourier, assinatura de forma e código de cadeia.

Os descritores de Fourier utilizam a representação de pontos de um contorno discretizado através de um plano complexo. Mais formalmente, seja $C = \{(x_0,y_0),(x_1,y_1),\cdots,(x_{N-1},y_{N-1})\}$ o conjunto de pontos de um contorno discretizado. Essas coordenadas podem ser representadas através de duas funções: $x(k) = x_k$ e $y(k) = y_k$. Para fazer a transformação do domínio

cartesiano para o domínio imaginário utiliza-se a Equação (5).

$$s(k) = x(k) + jy(k) \tag{5}$$

Nessa equação, o eixo x é representado pela parte real do espaço e o eixo y pela parte imaginária. Através da transformada discreta de *Fourier* é possível encontrar os $N-1$ descritores de C segundo a Equação (6).

$$a(u) = \frac{1}{N} \sum_{i=0}^{N-1} s(i) e^{-j2\pi ui/N} \tag{6}$$

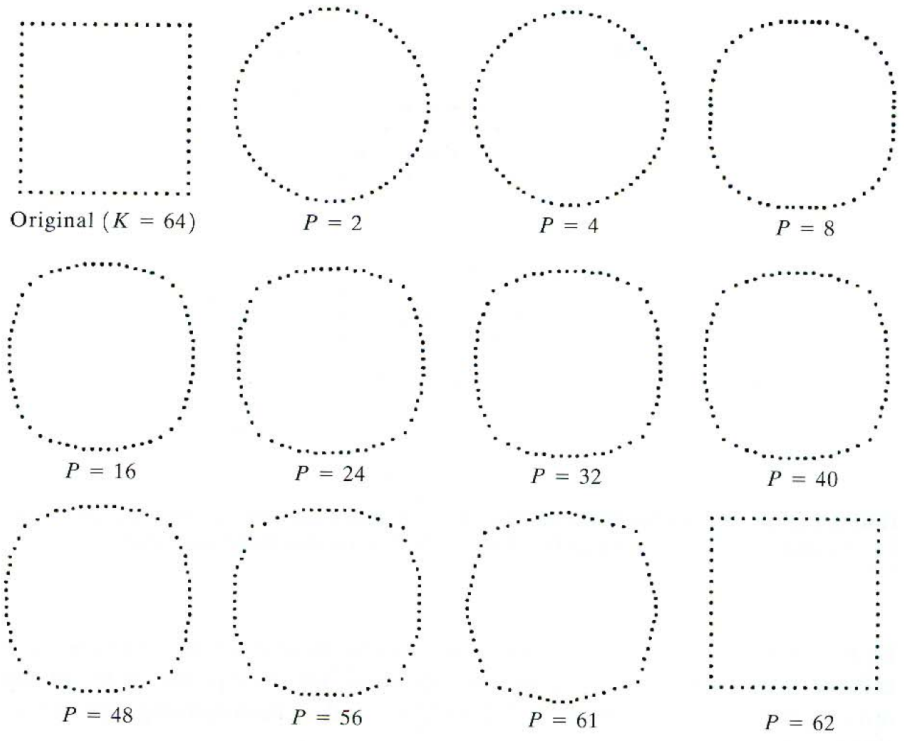
Uma vez que essa equação é apenas uma transformação espacial, é possível fazer a transformada inversa de *Fourier* e obter as coordenadas originais de C. Embora a Equação (6) revele $N-1$ descritores, é possível desprezar alguns desses descritores de tal forma que somente as tendências principais da forma sejam capturadas. Para isso, é necessário desconsiderar os componentes de $a(u)$ que representam as altas frequências; isto é, para todo $u > P$, onde P é o número de componentes que se deseja considerar, faz-se $s(u) = 0$. A Figura 3 ilustra a transformada inversa dos descritores de Fourier sobre um quadrado para diferentes valores de P. Uma propriedade interessante dos descritores de Fourier é que as baixas frequências da série estão associadas às características globais da forma, enquanto as altas frequências estão associadas aos detalhes da forma. Isso explica o comportamento da Figura 3, uma vez que quanto maior o valor de P, maior é a quantidade de detalhes da forma.

Outro método para descrição de imagens utilizando informações de forma é através do método de assinatura de forma. Esse método baseia-se na criação de um vetor gerado a partir do contorno da forma. O método tradicional de assinatura de forma utiliza as distâncias dos pontos do contorno igualmente espaçados até seu centróide, conforme demonstrado na Equação (7).

$$s(u) = \sqrt{(x(u) - c_x)^2 + (y(u) - c_y)^2} \tag{7}$$

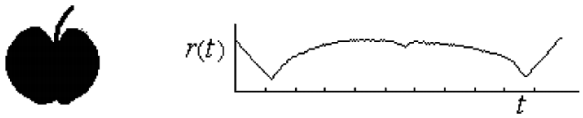
Nesta equação, $x(u)$ e $y(u)$ são as coordenadas do ponto u do contorno, e c_x e c_y são as coordenadas do centróide da forma. A Figura 4 mostra um exemplo de assinatura de forma. Na área de Recuperação de Informação, esta representação é frequentemente normalizada, tornando-a invariante à escala. Contudo, a invariância de rotação é feita através de *shift matching*; isto é, ao comparar duas assinaturas, mantém-se uma assinatura fixa e desloca-se outra assinatura sobre seu eixo das ordenadas até que se encontre a menor distância entre elas. Essa técnica deve ser analisada com cautela, uma vez que a complexidade para encontrar a forma mais semelhante a

Figura 3 – Descritores de Fourier aplicados a um quadrado com diferentes valores de P GONZALEZ e WOODS (2002)



Fonte: Autor “adaptado de” GONZALEZ e WOODS (2002)

Figura 4 – Uma forma descrita através das distâncias entre o centróide e os pontos discretizados de seu contorno Zhang e Lu (2004)



Fonte: Autor “adaptado de” Zhang e Lu (2004)

partir de outra forma dada em um banco de tamanho F toma tempo computacional $O(F.n^2)$, visto que são feitos $n - 1$ deslocamentos sobre um vetor e a medida de distância é $\theta(n)$ para cada forma do banco.

As representações vistas até aqui são chamadas de Modelos de Representação de Imagens, uma vez que o objetivo é descrever uma imagem utilizando algumas características extraídas a partir das regiões. O histograma, template ou domínio de frequência e domínio espacial são os domínios que são popularmente utilizados.

Recentemente, foram propostos modelos mistos; isto é, modelos que agregam mais do que uma característica para a representação de imagens. O *SIFT* é um dos mais conhecidos dentre eles Chang et al. (2010). A palavra *SIFT* é um acrônimo para *Scale-Invariant Feature*

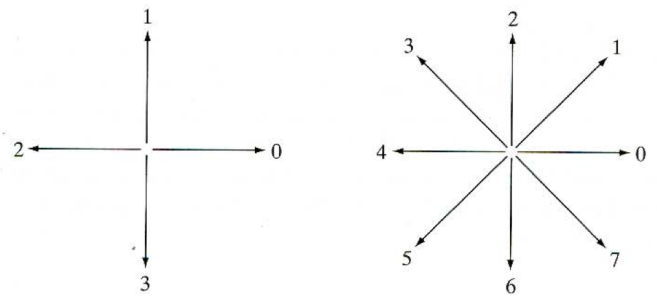
Transform. Trata-se de um método para representação de imagens através da identificação de pontos discriminantes que são menos afetados por fatores de rotação e escala, os chamados pontos fiduciais. Este método é utilizado nas áreas de Reconhecimento de Objetos, Reconhecimento de Gestos, Monitoramento de Vídeo, Identificação de Indivíduos, entre outras áreas correlatas à Visão Computacional.

O método *SIFT* utiliza a distribuição de gradiente em pontos específicos e em diferentes escalas para determinar pontos-chaves. De forma mais específica, o *SIFT* é composto por quatro etapas. A primeira etapa tem por objetivo localizar um conjunto P de pontos que não mudam de características quando a imagem sofre uma operação de transformação espacial. Para isso, é utilizado o método de diferença de gaussianas, onde cada gaussiana é aplicada a imagens de diferentes escalas. A segunda etapa filtra os P pontos utilizando a matriz Lapaciana da imagem. Com essa transformação, é possível localizar os pontos que apresentam maiores contrastes que fazem parte do novo conjunto P filtrado. A terceira etapa, insere a informação de orientação para cada ponto $p \in P$. Essa informação poderá ser então utilizada mais tarde quando uma imagem semelhante, porém com uma orientação diferente, for empregada em uma comparação. A quarta etapa é responsável por atribuir características locais aos pontos fiduciais obtidas a partir do gradiente ao seu redor. Essas características são representadas de forma a não permitir que valores de luminância e forma interfiram de maneira acentuada na representação do gradiente.

O último método, abordado nesta tese, capaz de representar imagens através das características de forma é o código de cadeia GONZALEZ e WOODS (2002). O principal objetivo da representação através do código de cadeia é descrever o contorno de uma forma utilizando segmentos de reta imaginários que conectam pontos vizinhos do contorno. Para cada segmento de reta, estabelece-se um número, ou código, que representa a direção desse segmento.

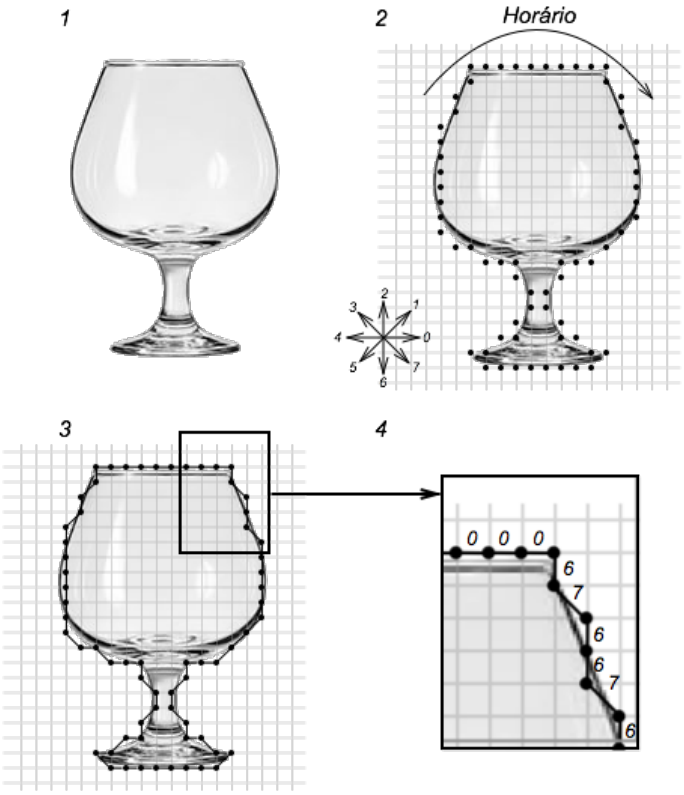
De maneira mais prática, para se obter o código de cadeira, deve-se inicialmente estabelecer três propriedades essenciais do algoritmo: a discretização dos pontos do contorno, o sentido de varredura e a discretização das direções. O primeiro parâmetro estabelece a discretização dos pontos do contorno representado para geração dos segmentos de reta imaginários. Esta propriedade estabelece o quanto a imagem será sub-amostrada para que o algoritmo inicie a detecção dos pontos do contorno. O sentido de varredura é uma propriedade que estabelece a ordem com que os segmentos de retas serão obtidos: horário ou anti-horário. O terceiro parâmetro estabelece o grau de discretização para a codificação da direção de cada segmento de reta. A Figura 5 ilustra duas possíveis discretizações. Na Figura 6, todo o processo de representação de forma através do código de cadeia é ilustrado.

Figura 5 – Discretizações para representação de forma através de código de cadeia. **Esquerda**: representação de direção utilizando 4 valores distintos. **Direita**: representação de direção utilizando 8 valores distintos. GONZALEZ e WOODS (2002)



Fonte: Autor “adaptado de” GONZALEZ e WOODS (2002)

Figura 6 – Exemplo de aplicação de código de cadeia sobre uma forma. **1:** Imagem original. **2:** Definição das propriedades do algoritmo e identificação dos pontos do contorno. **3:** Identificação dos segmentos de reta e rotulação. **4:** Detalhe da rotulação



Fonte: Autor

Os métodos abordados nessa seção são estudados a partir de duas linhas de pesquisas diferentes. A primeira explora as características e a segunda explora o modelo de representação. Os modelos de representação podem ser ajustados para atividades específicas. Na proposta dessa tese, apresentamos um modelo de representação misto para reconhecimento de objetos em cenas naturais, utilizando algumas das características discutidas aqui. O modelo proposto apresenta flexibilidade, uma vez que mais características podem ser incorporadas para o aumento de precisão.

2.1.1.5 Orientação

A orientação com que objetos são visualizados dentro no nosso campo visual sem dúvida é uma característica importante que influencia como interpretamos tais objetos e até mesmo a cena. Essa ideia é fácil de constatar com um experimento também simples. Se observarmos fotografias familiares de ponta-cabeça, para a maioria de nós é que interpretemos contextos totalmente diferentes. O arguto pintor italiano do Século XVI Giuseppe Arcimbaldo brincou com essa nossa peculiaridade (Figura 7).

Da mesma forma, é provável que deixemos de reconhecer rostos bem comuns (e até de familiares muito próximos) se invertermos suas fotos de ponta-cabeça (Figura 8).

Com certeza, o processo de atenção humana está preparado para visualizar objetos em orientação comuns, e ao nos depararmos com algo em orientação inusitada, como um carro na vertical, por exemplo (Figura 9), é provável que outros mecanismos de percepção visual entrem em ação para permitir a interpretação da cena. Provavelmente, mecanismos que cooperam entre si.

Assim, parece provável que a orientação dos objetos em cena exerça forte influência no tempo que levamos para interpretar o que vemos. De maneira intuitiva, é pouco provável que a orientação sozinha seria capaz de nos dar a interpretação completa do que vemos, sendo mais plausível afirmar que a orientação, combinada com outras características ao redor da cena, são fatores importante que colaboram.

Vários pesquisadores da área da Neurociência comprovaram os efeitos de nossa dependência da orientação que também não se limita a seres humanos Ashbridge et al. (2000). Os chimpanzés também são melhores no reconhecimento de faces de outros chimpanzés, quando são apresentados em sua correta posição vertical LA, T e WD (1998). Curiosamente, esses animais não apresentam o efeito de inversão quando lhes são apresentadas faces de outra espécie de primatas, os macacos cebos (*Cebus sp.*), ou um conjunto de objetos inanimados, os automóveis.

Figura 7 – Pinturas de Giuseppe Arcimbaldo (Século XVI). O artista até hoje nos deixa intrigados ao contemplarmos suas obras de orientações inusitadas.

Fonte: Giuseppe Arcimbaldo

Figura 8 – A percepção de faces é provavelmente uma das habilidade humanas extremamente dependente da orientação. O reconhecimento pode ser bastante difícil quando as faces são vistas de cabeça para baixo (esquerda). Ainda é mais surpreendente que não conseguimos notar uma grave distorção criada pela inversão dos olhos e da boca (direita) — Algo que seria logo evidente quando a foto fosse virada em sua posição correta vertical.

Fonte: Autor

Figura 9 – Um carro na orientação vertical é um evento muito raro de acontecer. Provavelmente a maioria de nós estranhará a cena até perceber o que realmente significa.

Fonte: <http://www.ndtv.com/world-news/snow-storm-hits-half-of-us-446801>

A orientação de objetos e cenas é uma característica também estudada em Visão Computacional. Ullman (1998) propôs um modelo computacional que utiliza a combinação de um pequeno número de vistas e orientações de objetos para lidar com o problema de reconhecimento. Ullmann mostra que, de acordo com evidências psicofísicas, a abordagem de combinação de vistas proposta utiliza vistas de diferentes de objetos diferentes, no lugar de múltiplas vistas de um mesmo objeto, para obter generalização de classes. A proposta é interessante porque tenta jogar luz ao entendimento de como conseguimos generalizar objetos e cenas da mesma classe sob uma enorme quantidade de perspectivas e orientações diferentes. Uma possível limitação do trabalho de Ullmann é o fato de que trata a percepção de objetos em uma cena como possível combinação linear de várias vistas 2D. Se $M_1, M_2, \ldots, M_k$ é um conjunto de vistas de um mesmo objeto, e P é a imagem 2D de um objeto a ser reconhecido, então P é considerado uma instância de M se:

$$P = \sum_{i=1}^{k} \alpha_i M_i, \text{para alguma constante } \alpha_i$$

Na abordagem proposta por Ullman, a percepção do contexto da cena para a interpretação generalizada de objetos é com certeza perdida. Provavelmente, em condições severas de orientação e vistas, o modelo proposto é limitado.

Plebe e Domenella (2007) propuseram um modelo baseado em Redes Neurais como uma série de blocos que representam as principais áreas das vias visuais encontradas em primatas e humanos para reconhecimento de objetos. Cada bloco corresponde a uma área cerebral, tais como as vias V1, V2 e V4. Como no sistema visual, cada via é responsável por um aspecto do

modelo, processando uma característica diferente das imagens de entrada. Cada rede é auto-organizável, como as redes de Kohonen. O sistema foi testado apenas para objetos individuais com característica locais e sem background, por esse motivo não considera reconhecimento baseado em contexto.

Recentemente, Lu et al. (2015) propôs uma alteração, batizada de S-HMAX, no modelo HMAX (Hierarquical Model and X)Riesenhuber e Poggio (1999) com o objetivo de obter melhor performance na detecção de objetos em cenas com uma certa quantidade de variação em termos de orientação espacial. A proposta dos autores baseia-se em características locais de pontos invariantes e foi testada em diversas bases de dados naturais, alcançando performance superior ao HMAX.

A grande maioria dos trabalhos para reconhecimento de objetos baseiam-se fundamentalmente em características locais. Com a orientação não é diferente, onde essa característica é extraída em ralação a uma região de interesse local ou a pontos fiduciais Akagündüz (2010), Sun et al. (2009), Ulrich, Wiedemann e Steger (2012), Xu e Xu (2005), Han, Yun e Lee (2004), Ekvall, Kragic e Hoffmann (2005), Zhu e Malsburg (2004), Kim, Yoon e Kweon (2008).

A característica de orientação regional definida nesta Tese objetiva estimar esta informação em um contexto global e não local, combinada com outras características contextuais em um modelo único de reconhecimento de objetos. A estratégia utilizada será descrita em detalhes no Capítulo 3.

2.1.1.6 Área

A área relativa de uma região de interesse é uma característica com pouco poder discriminativo. Assim como a orientação, parece mais uma característica auxiliar do que completa para diferenciar objetos em uma cena. Intuitivamente, parece estar mais associada à discriminação de ambientes outdoor e indoor: é mais provável que uma grande região em uma imagem, associada a uma textura específica e amorfa possa ser relacionada ao céu ou a uma praia ou montanha do que a objetos familiares de ambientes internos. No entanto, alguns trabalhos em processamento de imagens se beneficiaram da área das regiões de interesse para identificar objetos na cena. Um exemplo são os trabalhos de Rodrigues, Chang e Suri (2006), Rodrigues e Giraldi (2011) que utilizam a área de uma lesão dentro de um modelo bayesiano como fator de detecção para definir a região de tumores de câncer de mama. E Rathi e Palani (2012) utilizam a área de tumores cerebrais como discriminadores da lesão.

Em termos do sistema visual humano, a percepção do tamanho da área de um objeto não parece estar atrelada a nenhuma região cortical específica. Por outro lado, Tyler, Hardage e Miller (1995) demonstraram com experimentos psicofísicos como o sistema visual é sensível à simetria de objetos no seu campo visual e Dio C.D. e Rizzolatti (2007) estudaram via fMRI as áreas corticais relacionadas à percepção estética de proporções áureas.

Na presente Tese de Doutorado, estamos interessados em utilizar a área de um objeto no modelo proposto, observando o padrão de discriminação no modelo.

2.1.1.7 Contexto

Um objeto raramente ocorre isolado em uma cena. Geralmente são partes de uma cena mais completa e genérica. Por exemplo, muitos objetos aparecem associados a uma determinada região, como um carro sobre um estrada, talheres sobre uma mesa, ou mesmo animais em seu ambiente natural, que geralmente restringem seu aparecimento em cenários específicos. Além disso, um típico leiaute de cena 3D de um objeto e sua subsequente projeção em uma imagem 2D gera normalmente um cenário típico que pode ser explorado por algoritmos de reconhecimento. Contexto, nesse caso, é então um elemento crucial em um processo automático de reconhecimento de cena. Fato que foi estabelecido tanto na psicofísica Biederman (1981), Graef, Christiaens e d'Ydewalle (1990), Torralba A. e Henderson (2006) quanto na computação Hoiem D. e Hebert (2006), Torralba A., Freeman e Rubin (2003).

Diferente tipos de informações contextuais podem contribuir para este efeito. Por exemplo, cadeiras são mais prováveis de serem encontradas em ambientes indoor, enquanto árvores são um bom sinal de que trata-se de um ambiente outdoor. Este tipo de informação global de uma cena tem sido utilizado por Torralba A., Freeman e Rubin (2003) para construir modelos de reconhecimento de objetos. Uma importante ferramenta para esse tipo de aplicação são as representações integradas de objetos das imagens que capturam propriedades essenciais da cena, tal como proposta por Oliva e Torralba (2001).

Por outro lado, existe o contexto geométrico, que é resultado de como a cena foi capturada do mundo real, onde tamanho está mais relacionado à distância e oclusão indica ordem de profundidade. Além disso, muitos objetos do mundo real são mais prováveis de aparecer sobre uma superfície típica, geralmente um plano, que restringem sua possível localização dentro da cena. Abordagem que envolvem a detecção de automóveis em estradas geralmente se prevalecem dessas restrições. Como exemplo, temos Labayrade e Aubert (2003), Bombini et al. (2006), Gavrila e Munder (2007), Gerónimo et al. (2010). Com o objetivo de reconhecimento

de objetos em cenários mais genéricos, Hoiem D. e Hebert (2006) propôs uma abordagem para detectar simultaneamente aspectos geométricos de uma cena e objetos típicos relacionados.

Por último, mas não menos importante, existe o chamado contexto espacial, que diz que um tipo de região ou textura é mais provável de aparecer adjacente ou muito próximo a uma outra região ou textura específica do que outras. Por exemplo, a localização dos pés de um pedestre é mais provável de aparecer ao redor de superfícies do chão, e as partes do corpo ou da face de uma pessoa seguem geralmente uma ordem bem conhecida. Esse tipo de informação contextual foi explorada por He, Zemel e Carreira-Perpindn (2004).

Finalmente, os benefícios de informações semânticas de alto-nível foram demonstradas em uma série de artigos Rabinovich et al. (2007), Galleguillos, Rabinovich e Belongie (2008), Desai, Ramanan e Fowlkes (2011), Felzenszwalb et al. (2010) e um artigo relativamente recente demonstrou que, sem esse tipo de informação, sob baixa resolução, o reconhecimento de objetos é severamente afetado Parikh, Zitnick e Chen (2008).

Particularmente, o trabalho de Felzenszwalb et al. (2010) apresenta um modelo de reconhecimento de objetos baseado no contexto de co-ocorrência que é capaz de rotular de forma totalmente automática, região de imagens previamente segmentadas. A base de dados utilizada foi a popular PASCAl, utilizada para competição. O trabalho utiliza Bag-of-Visual-Words e um algoritmo de clusterização para criar um dicionário de objetos que é posteriormente usado para rotular regiões. Embora este objetivo se assemelhe bastante com uma das principais contribuições desta Tese, a metodologia e os bons resultados alcançados por Felzenszwalb et al. (2010), ao utilizar informações de características locais em uma assinatura única digital (bag-of-visual-words) e um clusterizador para criar o dicionário, se afasta da inspiração biológica e se assemelha muito mais a uma solução voltada à engenharia, não obstante, o trabalho de Felzenszwalb et al. (2010) nasceu da meta de ganhar uma competição. O objetivo da presente Tese é propor um modelo com o mesmo fim, mas separando as características de modo a introduzi-las em um modelo competitivo único, tal como se espera em um modelo neural.

Na área da Neurociência, como já foi dito na introdução desta Tese, há fortes indícios de que os processos de alto-nível interferem na percepção de baixo-nível (processo de atenção tardio interferindo no processo de atenção precoce). Pode-se mesmo levantar a hipótese de que, sem um processo de atenção de alto-nível, oriundo de um mecanismo de aprendizagem contextual, não é possível filtrar toda a quantidade de informação que chega nas vias visuais primárias.

A presente Tese de Doutorado apresenta como principal aspecto do modelo proposto a informação contextual, que foi construída a partir de uma base de dados anotada (SUN da-

tabase Xiao et al. (2010)) oferecida pela laboratório de inteligência artificial do Instituto de Tecnologia de Massachusetts (MIT). A partir de uma rede de relacionamento de objetos anotados, a presente Tese demonstra a forte ocorrência de contextos (processos de alto-nível), que são utilizados para interferir em processos de baixo nível, como segmentação de imagens.

2.1.1.8 Movimento

Como muitas outras características que não foram citadas, a detecção de movimento, não menos importante, não será tratada nessa tese. Há, na literatura, entretanto, uma vasta literatura, tanto na área de Neurociências quanto em Visão Computacional, demonstrando a eficiência da captura do movimento para reconhecimento de objetos. Como foi dito, esse tópico está fora do escopo da Tese.

Apesar disso, o modelo proposto, como será visto mais adiante, é capaz de agregar facilmente um conjunto ilimitado de características. Por esse motivo, as características em si não serão o foco deste trabalho, mas sim o modelo. Nesse sentido, a eficiência do modelo, por questões de tempo computacional e foco, será estudada e demonstrada apenas do ponto de vista de três características: Área relativa de uma região de interesse, Orientação e Cor.

2.1.2 Características Estudadas Neste Trabalho

Conforme será visto mais adiante, este trabalho utiliza características extraídas a partir de diferentes regiões de uma imagem. Assim, o modelo de representação utilizado baseia-se em regiões. Neste trabalho, considera-se uma região como um subconjunto de pixels que formam um componente conexo, de acordo com alguma regra de associação.

Utilizando a definição acima, pode-se considerar que uma região R_i de uma imagem I é um conjunto de pixels tal que $R \subseteq I$. Além disso, as regiões de uma imagem I são disjuntas; isto é, $\bigcup_i R_i = I$ e $\bigcap_i R_i = \emptyset$.

O motivo de ilustrarmos o modelo proposto com apenas um subconjunto de características descritas na Seção 2.1.1, é que o objetivo principal da tese é o modelo e não as características em si. Estas podem ser facilmente introduzidas no modelo a medida do interesse em investigação. Assim, os experimentos que serão conduzidos no Capítulo 4, serão feitos observando apenas área, orientação e cor.

Como dito anteriormente, neste trabalho serão utilizadas 3 características extraídas de regiões: Cor, Área e Orientação. A característica cor foi escolhida por ser utilizada em muitos

trabalhos da área de visão computacional e por ser uma característica considerada fundamental para o sistema visual humano S.Gazzaniga (1995).

A área e a orientação são outras características consideradas nesse trabalho. Há fortes evidências na Neurociência indicando que essas características já são observadas nas primeiras regiões do sistema visual humano (V1) e contribuem para as próximas áreas do sistema Kobatake e Tanaka (1994), Tanaka (1996).

Apesar dos estudos serem focados nessas três características, pode-se estender o trabalho para outras, tais como: textura, forma, posição espacial, posição relativa, movimento, entre outras.

A característica de cor é modelada através do "Histograma HSV" e segue o mesmo princípio explicado na Seção 2.1.1.2. Contudo, o histograma será criado tendo como amostra apenas os pixels pertencentes ao conjunto R_i.

A característica de "Área" extrai a quantidade de pixels dentro de uma região R_i e pode ser obtida através da Equação (8), onde x_i é a posição do ponto i no eixo x, y_i é a posição do ponto i no eixo y e $P_i = \{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$ é o polígono que delimita a região R_i descrito através do pontos ordenados de forma horária no plano xy.

$$S = \frac{1}{2}\left|\sum_{i=1}^{n-1} x_i y_{i+1} + x_n y_1 - \sum_{i=1}^{n-1} x_{i+1} y_i - x_1 y_n\right| \tag{8}$$

Finalmente, a terceira característica estudada neste trabalho é a "Orientação". Essa característica extrai o ângulo formato entre o eixo horizontal e a linha que conecta os dois pontos mais distantes de uma região. Essa característica é descrita através da Equação (9), onde p e q são os pontos mais distantes em P_i.

$$O = atan\left(\frac{|p_y - q_y|}{|p_x - q_x|}\right) \tag{9}$$

Todas as três características aqui apresentadas são mapeadas no domínio discreto; isto é, há um processo de discretização da característica para um intervalo determinado. A variação desse parâmetro será um dos pontos de estudo desse trabalho.

2.2 APRENDIZAGEM DE MÁQUINA

A área de aprendizagem de máquina está cada vez mais presente em sistemas inteligentes. Com o aumento do volume de dados, existe cerca de 1 trilhão de páginas na Web, fazendo com que a demanda por métodos de análise de dados também cresça da mesma forma.

Dentre os métodos de análise de dados, a aprendizagem de máquina é um importante arcabouço que permite aos pesquisadores detectar padrões de comportamento e até mesmo desenvolver modelos estatísticos previsíveis. Outro importante uso dessa área são os modelos de tomada de decisão baseados na incerteza.

A área de aprendizagem de máquina é normalmente dividida em 3 sub-áreas, distinguindo-se pelo seus modos de aprendizagem: supervisionados, não-supervisionados e aprendizagem por reforço.

Uma extensa fonte de informação sobre esta área pode ser encontrada em Mitchell (1997), Wernick et al. (2010), DUDA, HART e STORK (2000).

2.2.1 Aprendizagem Supervisionada

A aprendizagem supervisionada é um conjunto de técnicas de aprendizagem de máquina que utiliza dados de treinamento para inferir uma função classificadora. Mais formalmente, um classificador supervisionado é uma função

$$\hat{g}(x) : X \rightarrow Y$$

onde X é o espaço de entrada e Y é o espaço de saída. Normalmente, o espaço de saída Y é o conjunto de classes $Y = \{1, \ldots, c\}$, onde c é o número total de classes. A função $\hat{g}(x)$ é responsável por encontrar a classe esperada e, muitas vezes, toma a forma de uma maximização, como mostrado a seguir:

$$\hat{g}(x) = \arg\max_{y} f(x,y)$$

A função $f : X \times Y \rightarrow \mathbb{R}$ confronta uma entrada x com uma classe y e retorna um valor que representa a nota da classificação de x como pertencente à classe y. Assim, a classe esperada $\hat{y}$ para a entrada x será dada pelo valor de y que maximiza a função $f(x,y)$.

Os algoritmos de aprendizagem supervisionada possuem um papel fundamental quando se deseja trabalhar com bases de dados de exemplos de treinamento. De fato, nesse tipo de metodologia, a base de exemplos atua como um "professor", fornecendo custos ou classes esperadas, ajudando-o na busca de uma função $f(x,y)$.

Dentro do contexto de aprendizagem supervisionada, encontramos algoritmos que podem trabalhar em duas categorias de problemas: regressão e classificação. A regressão é uma técnica que relaciona variáveis dependentes e independentes através de uma função. De maneira

mais específica, a regressão estuda como o valor de uma ou mais variáveis (variável dependente) variam em função de outra, ou outras variáveis.

Por outro lado, a classificação é o processo de aprender uma função que mapeia dados de entrada, x, em dados de saída y. O que se espera de um classificador é que ele consiga identificar um padrão em x que esteja associado a um valor em y.

Exemplos práticos de aplicação das duas categorias podem ser dados a partir de dois problemas: interpolação e reconhecimento de padrões. No primeiro caso, a interpolação estima uma valor desconhecido no conjunto de dados a partir de valores conhecidos. Diferentemente da interpolação, o reconhecimento de padrões tenta associar um dado a uma determinada categoria.

Uma das primeiras técnicas em aprendizagem supervisionada para detecção de objetos utilizava as *Redes Neurais* Yang, Shu e Shah (2013), Rowley, Baluja e Kanade (1998), Fukushima (1980), Behnke (2003a). A vantagem de se utilizar essa ferramenta está em sua habilidade de classificar dados não lineares, tais como aqueles sem correlação entre as variáveis, erro aleatório com média zero e homoscedasticidade. Contudo, as redes neurais são modelos auto-organizáveis; isto é, a partir de um conjunto de exemplos ou amostras, seu modelo fará ajustes para minimizar o erro. Isso gera dois pontos negativos. O primeiro ponto é que a complexidade do modelo torna-o difícil de se depurar, uma vez que o mesmo se torna uma "caixa-preta" (conjunto de pesos sinápticos). Outro problema é que não se pode incluir fatos conhecidos na rede sem que ela seja treinada por novas amostras (ou exemplos). O efeito disso é que deve-se estar muito atento à criação do conjunto de treinamento, uma vez que o mesmo deve representar de maneira global a todas as classes.

Um exemplo clássico da aplicação de redes neurais pode ser encontrado em Fukushima (1980). Neste trabalho, o autor define um modelo, chamado de *Neocognitron*, para reconhecimento de padrões invariantes à posição em imagens digitais. O autor elabora um modelo multi-camadas e argumenta que esse modelo segue a organização cerebral em humanos. No entanto, modelos de reconhecimento de objetos ou mesmo interpretação automática de cenas, inspirados na biologia, ainda são raros e apresentam muitos pontos em aberto, sendo portanto motivos de controvérsia na literatura.

Outra limitação importante sobre redes neurais é que sua implementação baseia-se em 2 etapas: treinamento e predição. A etapa de treinamento é feita para configurar os pesos sinápticos para que a rede "aprenda" a função de classificação. A segunda etapa, a predição, é utilizada para classificar os dados novos de fato. A separação destas duas etapas torna as redes neurais limitadas, pois não se pode aprender novos casos sem antes treiná-las novamente. Assim, em sistemas dinâmicos, sempre será necessário voltar à etapa de treinamento para reconfigurar a

rede. A partir dessa limitação, criou-se o conceito de aprendizagem infinita (*Never-Ending Learning* Carlson et al. (2010)).

Em Carlson et al. (2010), os autores propõem um modelo computacional de aprendizagem infinita para extrair informações da *WEB*. Esse nome é dado pois espera-se que esse sistema permaneça aprendendo sempre que estiver operando. O modelo é constituído por um agente computacional inteligente capaz de ler informações da internet e popular uma crescente base de conhecimento. Basicamente, um modelo como esse busca por ontologias; isto é, uma coleção de predicados que definem relações entre entidades. Cada predicado encontrado na web é então analisado por sub-módulos que dão uma probabilidade do mesmo estar correto. Por exemplo, se este predicado é encontrado com muita frequência, ele deve ser promovido à base de conhecimento.

Atualmente, está sendo utilizado com frequencia em reconhecimento de objetos máquinas de vetores de suporte (SVM). O SVM é um classificador supervisionado baseado em busca de um hiperplano de separação que trabalha em duas classes, portanto esse é um classificador binário. O que diferencia o SVM de outros classificadores binários é que seu hiperplano de separação está localizado na região mais crítica de separação: as fronteiras. A função de maximização deste modelo cria um hiperplano que está localizado na maior distância entre os pontos mais próximos; isto é, há uma grande margem entre a fronteira das classes, tornando o erro da generalização, em geral, menor.

Em Lan et al. (2013), é proposto um modelo para reconhecimento de objetos baseado em classificadores SVM. O objetivo do artigo é propor um arcabouço para reconhecer e rotular objetos em cenas. O artigo utiliza diversos classificadores SVM, cada um representando uma classe, para classificar os objetos em cena. Além disso, o processo de reconhecimento utiliza objetos já reconhecidos como informação extra para inferir outros objetos na cena. O problema deste modelo é que deve-se ter um classificador SVM para cada classe de objetos. Para sistemas onde existem muitas classes, ter múltiplas instâncias de um classificador pode consumir excessiva memória computacional e tempo para o treinamento.

Outra teoria conhecida para reconhecimento de padrões em aprendizagem de máquina supervisionada é a teoria de Redes Bayesianas Neapolitan (2003). Esse método baseia-se fundamentalmente em uma análise estatística do problema e tem sido amplamente utilizado por trabalhos recentes para reconhecimento de objetos.

Um exemplo de classificador utilizando redes bayesianas é encontrado em Vasconcelos e Lippman (1998), onde os autores propõem um sistema de recuperação de imagens baseado em inferência estatística. O trabalho mantém foco em três principais aspectos: deve ser um sis-

tema intuitivo de recuperação; deve integrar diferentes modalidades de conteúdo; deve suportar aprendizagem estatística e, portanto, pode ser treinado automaticamente.

Um exemplo de trabalho que utiliza Redes Bayesianas para reconhecimento de objetos é o trabalho de Meng et al. (2011). Neste trabalho, os autores utilizam a contagem de eventos atômicos para relacionar as probabilidades *a priori* e *a posteriori*, com a finalidade de identificar padrões em imagens. Além disso, a rede proposta é implementada em um dispositivo FPGA (Dispositivo programável de portas lógicas), comparando seus resultados com uma Rede Neural SOM. Tais resultados mostraram que a rede bayesiana utilizou menos recursos computacionais que a SOM, alcançando resultados similares.

Recentemente, uma ferramenta que tem sido utilizada para modelar sistemas de reconhecimento de objetos são as chamadas Redes Complexas. Esses modelos têm sido inspirados devido a problemas atuais que só recentemente foram observados, tais como: sistemas biológicos (cadeias de DNA), sistemas sociais, redes de utilidade pública (elétrica, hidrovias, rodovias, aeroportos, etc...) e comportamentos de sistemas biológicos como o espalhamento de vírus e patologias, entre muitas outras. Uma revisão recente da área pode ser encontrada em Newman, Barabasi e Watts (2006), Newman (2010) e Caldarelli e Catanzaro (2012).

As redes complexas, conforme será abordado em mais detalhes na Seção 2.5, são basicamente grafos. Contudo, o seu poder de relacionamento de entidades, em conjunto com a área de sistemas complexos e estatística, permite o estudo das dinâmicas dos sistemas Wachs-Lopes e Rodrigues (2015), Newman, Barabasi e Watts (2006), ALBERT e BARABÁSI (2002), Barabási e Albert (1999). Há tempos alguns sistemas eram por vezes estudados a partir de hipóteses e experimentos empíricos, ou por serem difíceis de se modelar ou por envolverem consumo excessivo de processamento computacional, o que os tornavam inviáveis. Um exemplo disso ocorre com Milgram (1967). Nesse trabalho, o autor fala sobre a organização da rede social e o problema do *mundo pequeno*. De maneira geral, o trabalho trata sobre a seguinte questão: dada uma rede de amizades, quantas pessoas, no máximo, são necessárias para conectar quaisquer duas outras?

Esse estudo foi feito de maneira empírica, porém não computacionalmente, e sim de fato, pois foram enviadas cartas a pessoas geograficamente distantes com mensagens informando o nome de destinatários finais que deveriam receber essas mensagens. Cada pessoa que recebia uma mensagem informava seu nome e passava para outra pessoa que acreditava conhecer o destinatário final. O experimento indicou que existem, em média, 6 passos intermediários. Porém, o estudo foi conduzido por pessoas que vivem somente nos Estados Unidos e havia diversas limitações que poderiam invalidar o resultado final do experimento. Recen-

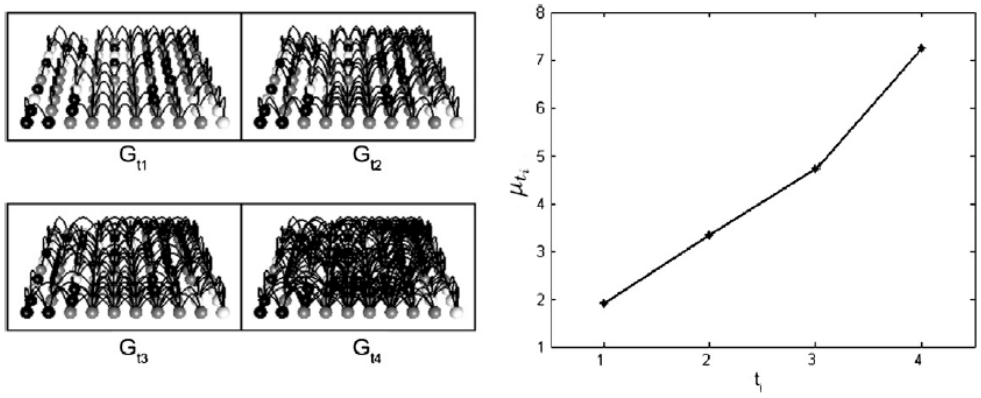
temente, trabalhos tratando do mesmo problema foram publicados graças ao surgimento das Redes Complexas Newman, Barabasi e Watts (2006), WATTS e STROGATZ (1998), Watts (2004).

Atraídos pelos resultados encontrados na área social, pesquisadores da área cognitiva Sporns (2002), linguística Wachs-Lopes e Rodrigues (2015), CANCHO e SOLÉ (2001) e visão computacional incorporaram as redes complexas em seus estudos Tang et al. (2012), Costa (2009), Zhang et al. (2012), Backes, Casanova e Bruno (2013b), Cuadros et al. (2012), Backes, Casanova e Bruno (2013a), Casanova, Backes e Bruno (2013). Em Backes, Casanova e Bruno (2013a), os autores propõem uma nova representação computacional de texturas de imagens para servir como descritores de texturas. De uma forma genérica, texturas são definidas como uma estrutura bidimensional de *pixel*. Nesse artigo, os autores representam cada *pixel* da textura através de um nó da rede complexa. A conexão entre dois nós é feita sempre que a distância euclidiana entre dois *pixels* for menor que um valor r, sendo representada por uma aresta não direcionada. Além disso, essas arestas possuem um peso que está relacionado à diferença da intensidade de cores entre os dois *pixels* que ela conecta. De posse dessa estrutura, o autor pode extrair diversas características como conectividade média, grau médio, coeficiente de clusterização médio, entre outras. Contudo, o trabalho vai além da extração de características em uma rede estática. Para dar dinamismo à rede, o autor elimina gradativamente arestas que possuem pesos menores que um limiar t. Ao variar t, obtém-se uma sub-rede com menos arestas, porém representando ainda a mesma textura. Ao medir novamente as características da nova rede, é possível analisar o seu dinamismo. Cada textura irá produzir diferentes comportamentos, sendo essa diferenciação que pode ser utilizada como descritor de texturas.

A Figura 10 ilustra a evolução e a variação do grau médio de uma rede complexa para valores crescentes de t. Observa-se na figura da direita que o gráfico do grau médio da rede aumenta a medida que t aumenta. Essa curva é vista como um descritor da textura. O que se obtém nesse tipo de modelagem é uma rede que relaciona a diferença de intensidades de cores para cada par de *pixels* próximos. Os autores desenvolvem diversas associações de redes complexas com texturas e propõem uma assinatura própria para cada textura. Os resultados mostraram que os descritores são robustos o suficiente para discriminar as texturas em suas respectivas classes, superando métodos tradicionais.

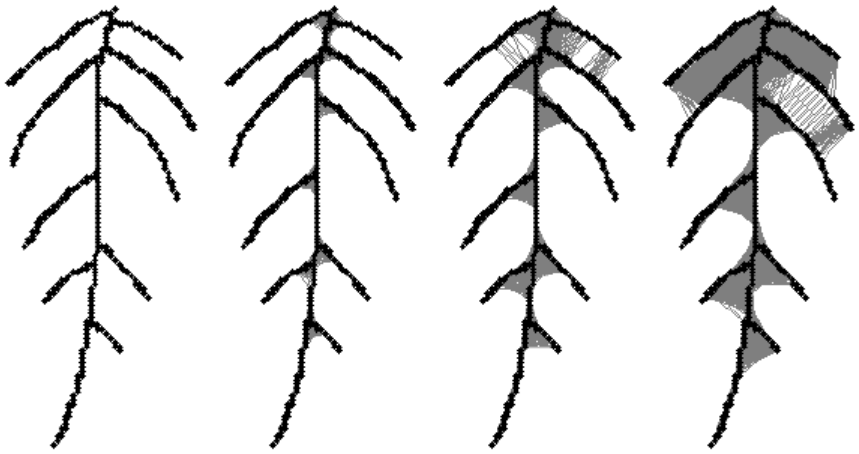
Resultados similares foram encontrados em Casanova, Backes e Bruno (2013). Nesse artigo, os autores estão interessados em desenvolver um modelo genérico o bastante para caracterização de sinais, curvas e conjuntos de pontos. O artigo é apresentado e validado utilizando como apoio o problema de reconhecimento de tipos de plantas; mais especificamente, identi-

Figura 10 – **Esquerda**: Visualização de uma rede complexa com diferentes valores de t.
Direita: Grau médio da rede para cada valor t Backes, Casanova e Bruno (2013a).



Fonte: Autor “adaptado de” Backes, Casanova e Bruno (2013a)

Figura 11 – Representação das nervuras de uma folha utilizando a evolução de redes complexasCasanova, Backes e Bruno (2013).



Fonte: Autor “adaptado de” Casanova, Backes e Bruno (2013)

ficando plantas a partir da análise de suas folhas. Nesse caso, a rede complexa é construída a partir dos pontos discretizados das nervuras das folhas; isto é, cada nó da rede complexa representa um ponto da nervura de uma folha. Inicialmente, todos os pontos são conectados entre si através das arestas, onde o peso de cada aresta representa a distância euclidiana entre os dois pontos. Esses pesos são então discretizados entre 0 e 1 e passam por um processo de limiarização. Nesse processo, diversos limiares são aplicados à rede com o objetivo de extrair informações de sua dinâmica. Aqui é utilizado o mesmo processo proposto em Backes, Casanova e Bruno (2013a), o dinamismo da rede complexa é gerado ao remover arestas que possuem pesos abaixo de um limiar variável t . A Figura 11 apresenta a evolução de uma rede complexa de acordo com a variação desse limiar. De posse da evolução dessa rede, o trabalho extrai, para cada valor de limiar, o grau máximo e o grau médio da rede, compondo um vetor de caracterís-

ticas. Para cada planta, o processo é repetido e são gerados diversos vetores de características. Finalmente, o trabalho aplica o LDA para separar linearmente as classes. A taxa de classificação obtida pelos experimentos foi de 93%. Os autores sugerem que esse método pode ser utilizado para outros problemas de classificação.

Outros artigos, tais como Chalumeau L. da F. Costa (2012), Backes e Bruno (2010), Backes, Casanova e Bruno (2009) mostram mais exemplos na literatura onde as redes complexas são utilizadas para lidar com problemas de classificação supervisionada. De forma geral, a maioria destes artigos modela um pixel ou uma região da imagem através dos nós e estabelecem o relacionamento entre eles através de arestas. Esses relacionamentos podem conter informações de cores, formas, texturas, entre outras características.

O uso de redes complexas para modelos de classificação supervisionada é promissor e traz novas possibilidades para a área de reconhecimento de objetos.

Esta seção apresentou três técnicas baseadas em aprendizagem supervisionada: redes neurais, redes bayesianas e redes complexas. Existem outras técnicas que se enquadram nessa categoria de aprendizagem, porém não serão abordados nesta tese, uma vez que o objetivo principal aqui é entender sua evolução na literatura de uma maneira mais ampla.

2.2.2 Aprendizagem Não Supervisionada

Dentre as técnicas de aprendizagem de máquina, a aprendizagem não supervisionada utiliza a própria estrutura estatística do conjunto de entrada de dados como uma base para discriminar padrões. Desse modo, diferentemente da aprendizagem supervisionada, não se tem um valor de saída esperado para cada entrada Dayan, Sahani e Deback (1999).

Um dos primeiros trabalhos que apresentou a aprendizagem não supervisionada é o de David Marr (1970), também encontrado em Vaina (1991). Neste trabalho, Marr propõe teorias computacionais baseadas no *neocortex* cerebral. Uma de suas teorias diz que as células do *neocortex* cerebral são classificadores flexíveis; ou seja, elas aprendem a estrutura estatística dos padrões de entradas a medida que combinações repetidas de padrões são apresentadas. A partir dessa constatação, os cientistas puderam concentrar seus esforços para criar novas técnicas de aprendizagem que fossem capazes de aprender sem qualquer tipo de supervisionamento. Marr publicou outros trabalhos apresentando teorias que continuam a influenciar modelos matemáticos modernos.

De forma mais explicativa, podemos dizer que a aprendizagem não supervisionada deve estimar um modelo probabilístico dos dados. Por exemplo, considere um sistema de previsão

de tempo. Nesse sistema, deve-se inferir um modelo probabilístico que represente a distribuição de probabilidade de um novo valor x_t a partir de dados passados $x_1, \cdots, x_{t-1}$. O que temos no final da aprendizagem é o modelo de $P(x_t|x_1, \cdots, x_{t-1})$.

Uma técnica derivada da aprendizagem não supervisionada é a clusterização. Esta técnica é responsável por encontrar, de maneira automática, padrões nos dados de entrada para a formação de grupos (ou *clusters*), de tal forma que objetos em um mesmo grupo sejam mais semelhantes que objetos entre grupos distintos.

De uma maneira mais formal, muitos algoritmos de clusterização são iterativos, onde a cada iteração, a clusterização é submetida a uma equação que avalia sua qualidade. A Equação (10) mostra um exemplo de como avaliar uma clusterização, baseada no conhecido algoritmo k-means JAIN (2010).

$$J = \sum_{j=1}^{k} \sum_{i \in C_j} ||x_i - \mu_j||^2 \tag{10}$$

Na Equação (10), k é o número de *clusters*, C_j é o conjunto de elementos pertencente da mesma classe e $||x_i - \mu_j||$ é uma medida de distância que relaciona o elemento x_i do cluster C_j ao centro μ_j do cluster C_j.

O que se espera de um algoritmo de clusterização é que a distância $||x_i - \mu_j||$ seja a menor possível para os elementos dentro de um mesmo cluster. Assim, os algoritmos de clusterização muitas vezes utilizam técnicas de minimização e aplicam a Equação (10) como uma função objetivo. A cada arranjo dos elementos nos grupos, o algoritmo testa a qualidade da clusterização utilizando esta equação.

No trabalho de Rodrigues e Giraldi (2011) foi feito um estudo detalhado do comportamento do algoritmo k-means aplicado à clusterização de imagens. Nele, os autores mostraram que essa técnica é altamente dependente dos parâmetros estatísticos da distribuição dos dados, como média e desvio-padrão. Como consequência, não é uma tarefa fácil generalizar aplicações baseadas nesse algoritmo.

Outro tipo de aplicação onde a aprendizagem não supervisionada é utilizada é a redução de dimensionalidade. Considere um sistema capaz de classificar se uma pessoa está doente ou não. O sistema é alimentado através de um prontuário contendo as seguintes informações: tipo sanguíneo, número de plaquetas no sangue, idade da pessoa, tom de pele e hora da internação. Dessas 5 características pode-se retirar algumas que não estão relacionadas à presença de doença como, por exemplo, o tipo sanguíneo, tom de pele e hora da internação. As duas variáveis restantes (número de plaquetas e idade da pessoa) estão mais relacionadas às doenças e são as que mais contribuem para a classificação. Ao utilizar um algoritmo de clusterização, somente as

duas variáveis relacionadas ao diagnóstico devem ser utilizadas, do contrário as outras variáveis poderiam prejudicar a taxa de classificação do sistema. É nesse tipo de problema que muitas vezes a redução de dimensionalidade é utilizada.

Existem muitos métodos capazes de fazer redução de dimensionalidade, cada um com abordagens diferentes. Um método bem conhecido baseia-se na decomposição SVD (*Singular Value Decomposition*), que basicamente fatora uma matriz M em três novas matrizes: U, Σ e V. O SVD é então uma transformação linear da matriz original M, de tal forma que:

$$M = U\Sigma V$$

Pode-se interpretar a equação acima através de 3 operações lineares. A partir de uma matriz com vetores unitários V, um escalonamento Σ e uma rotação final U pode-se obter M. Uma observação importante deve ser feita sobre a matriz Σ. Uma vez que esta matriz faz uma operação de escalonamento sobre V, os maiores auto-valores de Σ estão associados às dimensões que apresentam maiores variâncias na matriz M. Isso significa que a matriz Σ pode dar informações sobre as dimensões mais importantes nos dados de entrada. É nesse ponto que podemos descobrir quais dimensões são as mais importantes para separar os dados em grupos. Muitos trabalhos utilizam as dimensões com maiores variâncias como entrada em classificadores.

Contudo, qualquer método de redução de dimensionalidade deve ser utilizado com cautela, uma vez que os dados de entrada devem ser tratados para que haja uma proporção correta entre as dimensões. Tome como exemplo o sistema de classificação anterior de doentes. A escala de medida para contagem de plaquetas no sangue em uma pessoa saudável varia entre 150.000 a 450.000 por mililitro cúbico, enquanto a idade da pessoa varia entre 0 a 100 anos aproximadamente. Evidentemente, não se pode tentar fazer uma redução de dimensionalidade sem antes normalizar estes dados para uma faixa comum, caso contrário isso levaria a matriz de escalonamento Σ a ter valor maior sobre a dimensão de número de plaquetas do sangue, por conter números de maior grandeza.

Além da normalização, um ponto importante deve ser discutido. Atualmente não se conhece um método geral que informe quantas dimensões são suficientes para representar o conjunto de entrada de forma fidedigna e sem interferência no classificador. Assim, muitos trabalhos recorrem a métodos empíricos a fim de observar o efeito da dimensionalidade sobre a classificação dos dados.

Há diversas vantagens em se utilizar técnicas de aprendizagem não supervisionada. A primeira delas está relacionada à base de dados. Quando trabalhamos com técnicas não su-

pervisionadas, podemos dispensar o uso de bases supervisionadas e simplesmente treinar o classificador com dados reais. Essa técnica é interessante principalmente para problemas que demandam por grandes bases de dados. Muitas vezes não temos uma base de dados supervisionada grande o suficiente para treinar um classificador, pois poderia ser muito trabalhoso criar uma base que conseguisse representar genericamente um problema. Assim, se um classificador supervisionado fosse treinado com uma base de tamanho menor que o necessário (ou não genérica), o mesmo poderia se tornar muito especializado aos dados de treinamento e não ser genérico o suficiente para classificar novos dados de entrada. Isso mostra como os classificadores não supervisionados podem ser úteis como um ferramental para se trabalhar com grandes bases de dados não supervisionadas.

Uma outra vantagem da aprendizagem não supervisionada é que podemos treinar um classificador não supervisionado, encontrando as principais classes de dados e, posteriormente, utilizar técnicas supervisionadas para identificar as classes encontradas DUDA, HART e STORK (2000). Tome como exemplo um reconhecedor de caracteres (OCR). Podemos apresentar diversos padrões das letras do alfabeto sem qualquer supervisão e separá-los em 52 grupos (26 pares de letras maiúsculas e minúsculas). Após essa separação, podemos usar técnicas de supervisão para rotular os grupos encontrados apenas comparando um elemento de cada grupo com um caractere supervisionado.

Além das duas vantagens apresentadas anteriormente, podemos acrescentar outros benefícios de se utilizar técnicas não supervisionadas. Muitas vezes, lidamos com sistemas dinâmicos em que o meio externo está em constante mudança. Nesses casos, utilizar uma técnica de aprendizagem que possa se adaptar a essas variações é essencial para o modelo. Para esses casos, a aprendizagem não supervisionada também pode ser utilizada para fazer ajustes frequentes, à medida que o ambiente se altera.

Finalmente, a aprendizagem não supervisionada pode ser utilizada como uma ferramenta para encontrar características discriminantes nos dados de entrada. Frequentemente, nos deparamos com bases de dados contendo muitas dimensões (ou características). Para esses tipos de bases, escolher quais características considerar no modelo do sistema é uma tarefa inerentemente difícil. Ao utilizar técnicas de classificação não supervisionadas, podemos obter as características que mais separam, ou discriminam, os dados de entrada.

Por outro lado, um ponto negativo da aprendizagem não supervisionada é a excessiva generalidade dos algoritmos, que pode limitar a performance em uma atividade específica. Uma técnica não supervisionada utiliza noções gerais para identificar padrões ou encontrar estruturas

interessantes nos dados de entrada. Isso significa que qualquer especialização necessária para algum tipo de entrada não será possível de se treinar.

Outro ponto negativo é que, como seu método de aprendizagem não é guiado (ou supervisionado), a presença de erros na base de dados pode induzir a aprendizagem a um caminho indesejado. Suponha um sistema que faça o controle de qualidade em uma linha de produção. Se, durante o treinamento, houve erro de leitura dos sensores, esse erro poderá ser "aprendido", o que pode trazer resultados indesejáveis durante a classificação de novos dados.

Na literatura, diversos trabalhos recentes utilizam a aprendizagem não supervisionada para área de visão computacional e reconhecimento de objetos em cenas. Nessa linha de pesquisa, Bo, Ren e Fox (2012), Vincent et al. (2008), Collins e Singer (1999), Lee et al. (2009), Coates e Ng (2011), Bo, Ren e Fox (2011) são exemplos de trabalhos que utilizam esta técnica.

Um dos problemas envolvidos na área de reconhecimento de objetos é descobrir quais características (ou dimensões) nos dados de entrada são fundamentais para separá-los em grupos. A essas características, damos o nome de *características discriminantes*.

Em Bo, Ren e Fox (2011), Bo, Ren e Fox (2012), os autores propõem um método para encontrar características discriminantes utilizando o *K-SVD* Aharon, Elad e Bruckstein (2006), que é um decompositor de sinais de tal forma que um sinal y pode ser reconstruído através de uma matriz D de tamanho reduzido e de um vetor de coeficientes x, conforme a seguinte equação:

$$y \simeq D.x$$

onde D é chamada de matriz dicionário e x contém os coeficientes de representação do sinal y. Em Bo, Ren e Fox (2011), os autores utilizam essa ferramenta para construir um dicionário a partir de diversas grades de *pixels* obtidas de imagens coloridas com informações de profundidade. Assim, um dicionário para uma grade de 5×5 *pixels* irá formar vetores de tamanho $5 \times 5 \times 8$, onde o último componente é responsável por guardar informações de tons de cinza, RGB, profundidade e normais das superfícies. Contudo, o trabalho treina cada característica separadamente, havendo um dicionário para cada tipo de característica.

Um ponto em aberto deixado pelo trabalho é o valor de K escolhido para os experimentos. É mostrado, visualmente, que um dicionário aprendido com $K = 5$ pode ser bem próximo da imagem original, veja Figura 12. Contudo, esta afirmação não pôde ser esclarecida matematicamente e nem uma justificativa para esse valor foi discutida além da similaridade na aparência.

Figura 12 – Reconstrução de uma imagem a partir de uma matriz de dicionário. **Esquerda:** Imagem original, **Centro:** Imagem reconstruída com $k = 2$. **Direita:** Imagem reconstruída com $k = 5$

Fonte: Autor “adaptado de” Bo, Ren e Fox (2011)

A escolha por um valor de K (ou quantidade de grupos) ideal ainda é um problema em aberto. Em Erdmann et al. (2013), os autores utilizam uma meta-heurística chamada *Firefly* para encontrar valores de limiares para segmentação de imagens. Uma base de dados de imagens segmentadas manualmente foi utilizada para comparar cada supervisão humana com a segmentação automática. Os resultados mostraram que existe um valor de K ideal associado aos valores de limiares que aproxima a segmentação das imagens às imagens supervisionadas. Isso significa que o valor de K (ou quantidade de limiares) é tão importante quanto os próprios valores dos limiares.

Alguns trabalhos como Rose, Gurewitz e Fox (1990), Temel (2014), Newell et al. (2013), Yu, Liu e Wang (2014) estudam métodos capazes de computar o valor de K automaticamente. Em Yu, Liu e Wang (2014) os autores propõem um método bayesiano associado à na teoria dos conjuntos aproximativos (ou *rough set model*) para clusterização de dados. De uma maneira geral, o trabalho cria um modelo capaz de representar estatisticamente a tomada de uma ação α sobre um objeto O e uma classe C com respeito a três possíveis ações. A primeira delas é classificar $O \in C$. A segunda é classificar $O \notin C$. A última ação é classificar O como um objeto na fronteira da classe C. O algoritmo para determinação de K é desenvolvido utilizando o conceito bottom-up; isto é, o algoritmo assume inicialmente que cada objeto está contido em sua própria classe. A cada passo iterativo, o algoritmo compara a similaridade entre dois objetos em relação à similaridade média do conjunto de dados. Caso haja algum par de objeto entre *clusters* diferentes que tenham similaridade menor do que a similaridade média entre todos os pares de objetos, haverá uma junção entre estes *clusters*. O algoritmo termina assim que não haja mais qualquer par de objetos em *clusters* diferentes que tenham similaridade menor do que a similaridade média entre todos os pares.

Apesar da literatura propor novas maneiras de computar o valor de K, ainda não há um método suficientemente preciso e aceito pelos cientistas. Além disso, não se sabe exatamente como medir a qualidade de um cluster. Isso deve-se ao fato de que a clusterização está relacionada à intuição humana. Por isso, muitos trabalhos assumem o valor de K empiricamente, deixando em aberto discussões mais profundas relacionadas à cognição humana.

Em Mesnil et al. (2013) são utilizadas outras técnicas de redução de dimensionalidade como PCA e CAEs (Contractive Auto-Encoders) Rifai et al. (2011) para tratar grandes vetores de características de objetos para classificação de cenas. Os autores utilizam dados obtidos a partir de diversos extratores de características e reduzem a dimensionalidade dos dados aplicando técnicas não supervisionadas, tal como o PCA. Os resultados mostram que a técnica pode ser relevante para a área, uma vez que conseguiu melhorar a performance da classificação em até 10% em relação aos trabalhos recentes. Os autores destacam que a redução de dimensionalidade sobre os vetores de características estudados conseguem extrair dados de alto nível (mais representativo) eliminando-se a dependência linear entre os atributos.

Além dos problemas específicos discutidos até aqui, destacamos mais dois problemas comuns na área de aprendizagem não supervisionada. O primeiro deles está relacionado ao vetor de características modelado. Muitos trabalhos, como é o caso de Mesnil et al. (2013), constroem um vetor de características a partir de diferentes características heterogêneas. Dependendo da técnica utilizada, essa modelagem pode trazer efeitos indesejáveis à modelagem, uma vez que o algoritmo de otimização pode apresentar restrições quanto às variáveis. Uma dessas restrições, a normalização das dimensões, foi discutida anteriormente. Nesse caso específico, o modelo poderá se tornar inválido devido a mistura dos tipos de variáveis, tais como: discretas e contínuas, qualitativa e quantitativa, nominal e ordinal, entre outras. Portanto, a mistura entre tipos de variáveis em um mesmo vetor de características deve ser analisada cuidadosamente para que não haja violação das restrições.

Outro tipo de problema ocorre quando o método utilizado não está adaptado para sistemas dinâmicos. Normalmente, os métodos de aprendizagem supervisionada e não-supervisionada trabalham em três estágios. O primeiro é a aprendizagem, o segundo é a validação e o terceiro é a classificação. Contudo, para sistemas dinâmicos, não se pode treinar o sistema em um determinado momento e classificar em outro momento. Nesse caso, a aprendizagem deve ser "on-line"; ou seja, o treinamento ocorre conjuntamente com a classificação.

Uma aplicação que demanda por um sistema de aprendizagem e classificação de forma conjunta é a navegação de robôs em ambientes dinâmicos. Alguns trabalhos, como em Hadsell et al. (2007), consideram que o ambiente onde o robô, ou agente, navega contém obstáculos que

mudam de posição durante a fase de classificação. Assim, o ambiente em que eles interagem é dinâmico e necessita de um aprendizado adaptativo e constante em termos temporais.

Esta seção destacou alguns pontos positivos e negativos sobre aprendizagem não supervisionada. Na literatura científica, na área de reconhecimento de objetos, frequentemente são encontrados trabalhos utilizando aprendizagem não supervisionada como uma técnica para redução de dimensionalidade do vetor de características dos objetos. Contudo, essa prática pode ter efeitos indesejáveis como eliminação excessiva de dados, valor de corte ideal desconhecido e limitações devido à junção de características distintas em um mesmo vetor de características.

2.2.3 Aprendizagem por Reforço

Uma terceira técnica para aprendizagem de máquina é conhecida como *reinforcement learning*, ou aprendizagem por reforço. Esta técnica baseia-se no relacionamento entre recompensas e ações feitas por agentes em um determinado ambiente. Portanto, esta técnica de aprendizagem está mais relacionada com as interações de um agente com o seu ambiente de operação.

Como descrito em Sutton e Barto (1998), a aprendizagem por reforço é uma técnica que está intrinsecamente relacionada à definição do problema a ser resolvido, ao invés da elaboração de algoritmos para resolvê-lo. Assim, a modelagem do problema que será resolvido com aprendizagem por reforço é feita de acordo com um ou mais objetivos. A partir da definição desses objetivos, alguns algoritmos utilizam conceitos oriundos da psicologia, como é o caso de "reforço", para alcançar o objetivo.

Na área da psicologia, Ivan Pavlov define que o reforço é um estímulo dado a um indivíduo com o objetivo de associar e consolidar a resposta gerada a partir de uma ação Saunders (2006). Para os humanos, o reforço pode ser gerado a partir dos estímulos senso-motores ou até mesmo de outros indivíduos.

Um exemplo prático de reforço pode ser observado em crianças desenvolvendo a habilidade de andar. Nota-se que a aprendizagem acontece a medida que a criança testa diversas combinações de movimentos, ganhando mais equilíbrio com o passar do tempo. Nesse caso, as quedas agem como reforços negativos e são os estímulos que a criança precisa para não repetir o que havia testado anteriormente.

Na área da computação, este princípio se mantém. O agente aprende através de tentativas e erros e a informação de reforço é processada a cada nova tentativa. As técnicas de

aprendizagem por reforço têm por objetivo mapear situações (ou estados) para cada possível ação que um agente pode realizar.

Dentro do contexto de aprendizagem por reforço alguns conceitos são importantes para compreender diferentes métodos. O *objetivo* é um estado do agente que deve ser alcançado. Uma *política* define uma sequência de ações do agente na tentativa de se alcançar um *objetivo*. A *função de recompensa* mapeia estados (ou pares de estados-ações) em números que representem a *recompensa* por um agente estar em um estado (ou tomar uma determinada ação em um determinado estado). A *política ótima* é uma política que maximiza a *função de recompensa*.

Há duas formas de encontrar uma política ótima. A primeira delas é a chamada *on-line* e tem características exploratórias. A técnica *on-line* parte de uma política inicial (aleatória ou não) e a maximiza iterativamente até chegar à política ótima. Por outro lado, a técnica *off-line* descobre uma política ótima sem se basear em outra política. Assim, essa última é útil quando deseja-se treinar um sistema sem caráter exploratório e utilizá-lo em um momento posterior.

Muitos trabalhos que implementam a aprendizagem por reforço utilizam o modelo *MDP (Markov Decision Process)* como um ferramental matemático para descrever o sistema estudado. O *MDP* é uma tupla $\langle S,A,\delta,\tau\rangle$, onde S é o conjunto de estados, A é o conjunto de ações, δ é a função de transição e τ é o conjunto de recompensas. De posse desse modelo, os trabalhos utilizam técnicas de otimização para encontrar o conjunto de ações (política ótima) para que se maximize a recompensa.

Alguns trabalhos, como em Bandera et al. (1996), Draper, Bins e Baek (1999), Minut e Mahadevan (2001), Piñol et al. (2012), utilizam a aprendizagem por reforço na área de reconhecimento de objetos em imagens. Em Draper, Bins e Baek (1999), os autores propõem um modelo adaptativo para detecção de tipos de moradias. Essa detecção é feita a partir de imagens aéreas e são estudados 4 tipos de casas. O trabalho utiliza um modelo *MDP* e o *Q-learning* como técnica de otimização para definir um conjunto de módulos de visão computacional para reconhecimento de objetos. A metodologia proposta tem como objetivo encontrar a melhor combinação entre os módulos definidos que possa levar o sistema a obtenção da melhor performance possível. Neste caso, a função de recompensa está associada à qualidade da detecção. Sendo assim, foi necessária um base de dados supervisionada. A partir desse meta-modelo, os autores desenvolvem 4 modelos definitidos para detecção de 4 tipos de casas. Os resultados mostraram que o trabalho encontrou modelos que se aproximam dos modelos ótimos para detecção de cada classe de casa. Contudo, para a criação desse meta-modelo é necessário que haja uma biblioteca de procedimentos de visão computacional, tais como: correlação, detecção de picos, segmentação, modelos deformáveis, entre outros. Porém, criar uma biblioteca suficiente-

mente grande para detectores genéricos é uma tarefa difícil. Assim, no caso deste trabalho, que tem por objetivo detectar somente casas, a modelagem se torna viável pois é possível trabalhar em um subconjunto seletivo de procedimentos e técnicas de visão computacional.

Outros exemplos de aplicações de aprendizagem por reforço na área de visão computacional são encontrados em Paletta e Pinz (2000), Paletta, Fritz e Seifert (2005). Em Paletta e Pinz (2000), um sistema ativo de reconhecimento de objetos é proposto utilizando esse tipo de método. A ideia do trabalho é encontrar o melhor ponto de visão em uma cena para que se tenha maior discriminação dos objetos. Inicialmente, o trabalho apresenta um modelo utilizando funções de base radial capaz de aprender a aparência visual dos objetos em diferentes pontos de visão. Para isso, algumas características são extraídas e projetadas em um espaço 3-dimensional. Após o modelo ser aprendido, ele é então utilizado como uma função de utilidade capaz de quantificar a hipótese de presença de objetos na cena. De posse desse modelo, os autores criam um *MDP* para representar os processos de decisão para que se tenha o melhor ângulo de visão possível. Os estados desse modelo representam a probabilidade de se discriminar os objetos naquele ponto de visão, as ações são os movimentos possíveis do ponto de visão e a recompensa é encontrada a partir da função de utilidade. O trabalho é validado através de um experimento composto por um sistema senso-motor com uma câmera. Os resultados sugerem que o *Q-Learning* é um algoritmo promissor para detecção de objetos em cenas 3D e sugere que é um modelo promissor para a área de reconhecimento de objetos. Como trabalhos futuros, deve-se estudar formas alternativas para a função de utilidade, uma vez que há muita dependência entre os resultados e a função de utilidade. Outro ponto em aberto é que a modelagem dos objetos em um subespaço pode encontrar os mesmos problemas discutidos na Seção 2.2.2.

Em Piñol et al. (2012), os autores propõem um método de aprendizagem por reforço com o objetivo de selecionar as melhores características em imagens para que um classificador tenha a maior performance possível. O trabalho utiliza um *MDP* como modelo de representação dos processos de decisão. A estratégia é fazer com que os estados sejam os conjuntos de características que descrevem a imagem de entrada, tais como: cor mediana, cor média, número de cores, entre outras. O conjunto de ações do modelo *MDP* representa as escolhas que o sistema poderá fazer para melhorar o processo de reconhecimento de objetos da cena. Assim, o *MDP* foi definido como 4 possíveis extratores de características: *SIFT (Scale-Invariant Feature Transform)* Lowe (2004), *SPIN* Lazebnik, Schmid e Ponce (2005), *SURF (Speeded Up Robust Feature)* Bay et al. (2008) e *PHOW (Pyramid Histogram Of visual Words)* Bosch, Zisserman e noz (2007). A função de transição δ foi ligeiramente alterada no trabalho. Ao invés desta função retornar um novo estado, ela retorna a nova representação da imagem após aplicar uma ação α.

Finalmente, a função de recompensas τ retorna recompensas altas sempre que o classificador optar pela classe correta do objeto, portanto é necessário o uso de um *ground truth*.

Dada esta modelagem, o autor encontra a política ótima utilizado o *Q-Learning*Sutton e Barto (1998). Após o treinamento, o sistema estará apto a escolher corretamente os extratores de características a partir de uma imagem de entrada e enviar para um classificador. Apesar desse trabalho apresentar uma abordagem interessante, não se pode garantir que para qualquer imagem o sistema irá extrair as melhores características, uma vez que o conjunto S é descrito a partir de características genéricas. Outro ponto negativo é que a base de treinamento deve ser suficientemente grande para generalizar todo o universo de imagens.

Uma vantagem em se utilizar a aprendizagem por reforço é que o problema pode ser descrito a partir de um modelo *MDP*. Isso significa que devemos apenas definir os estados e ações do sistema, as transições e criar uma função de recompensa que esteja relacionada ao objetivo. Assim, o foco do modelo está muito mais ligado à função recompensa que propriamente uma saída esperada.

No entanto, o uso de métodos iterativos para encontrar a política ótima tem algumas desvantagens. A primeira delas está relacionada com a função de recompensa, que deve ser avaliada a cada passo da iteração. Muitas vezes, como é o caso dos trabalhos abordados nesta seção, a função de recompensa é computacionalmente cara. Isso significa que o uso de outras técnicas e modelos, como estrutura de dados avançadas, devem ser utilizados para diminuir a complexidade computacional.

Outra desvantagem em utilizar aprendizagem por reforço é a dificuldade em encontrar parâmetros ótimos para o modelo. Por exemplo, muitos algoritmos utilizando esta técnica possuem um parâmetro associado à taxa de aprendizagem. Se este parâmetro for muito pequeno, o sistema pode demorar para convergir e, em alguns casos, um sistema aleatório pode convergir mais rapidamente que a aprendizagem por reforço. Se a taxa de aprendizagem for muito alta, poderá ocorrer uma especialização nos últimos dados de entrada do sistema. Isso poderá levar a um sistema muito sensível e instável. Assim, um valor ótimo para a taxa de aprendizagem normalmente é encontrado empiricamente. Alguns trabalhos, como é o caso de Draper, Bins e Baek (1999), comparam o sistema de aprendizagem com um sistema aleatório na busca de convencer o leitor de que o modelo converge de maneira mais rápida.

As técnicas estudadas até aqui mostraram alguns recursos para a modelagem de sistemas capazes de identificar padrões a partir de uma massa de dados. Essas ténicas foram apresentadas como ferramentas de aprendizagem de máquina. Na Seção 2.3, apresentaremos alguns aspectos

que devem ser levados em consideração quando deseja-se incorporar ao modelo características relacionadas à consciência.

2.3 INFLUÊNCIA DA NEUROCIÊNCIA NOS MODELOS DE VISÃO COMPUTACIONAL

A inteligência artificial sempre buscou criar modelos inspirados nos seres humanos, adotando a cognição, ou aquisição de conhecimento, como um dos seus principais pilares Russell e Norvig (2003). Contudo, pesquisadores da área de Neurociência parecem concordar que existem diversos aspectos, interligados ou não, relacionados à tarefa de interpretação de uma cena ou objeto em uma cena.

Alguns desses aspectos, possíveis de mencionar hoje, são: mecanismo de consciência, mecanismo de atencão, visão *top-down* e *bottom-up* e competição/colaboração dos mecanismos que constituem as diversas características de percepção visual, tais como: cor, forma, textura, relacionamento espacial, co-ocorrência entre objetos, entre muitos outros. Esta seção apresenta a influência da área da Neurociência nos modelos de Visão Computacional propostos na literatura. Dentre esses quatro aspectos citados aqui, dois deles, *top-down*/*bottom-up* e competição/colaboração, estão entre os mais abordados na literatura de visão computacional. Por esse motivo, esses dois serão abordados com maior abrangência nesta seção.

O primeiro aspecto discutido aqui que a Neurociência relaciona ao processo de interpretação de uma cena é a consciência. Além da área de Neurociência, a consciência é um conceito estudado em diversas outras áreas, tais como: Filosofia, Neurociência, Psicologia, Física, Inteligência Artificial, entre outras. Para cada área citada, o mecanismo de consciência é definido a partir de diferentes pontos de vista, levando a várias definições para cada uma delas.

Embora não haja na literatura, até onde sabemos, uma definição única e formal sobre a consciência, pode-se dizer que ela está relacionada com alguns conceitos secundários, tais como: mecanismo de atenção, processo de decisão, percepção e cognição Calvin e Ojemann (1994), STARZYK e PRASAD (2011), Cohen, Alvarez e Nakayama (2011).

Alguns autores fundamentam o conceito de consciência a partir do conceito de experiência e objetivo Baars e Gage (2010). Uma experiência pode ser vista como uma memória entre causa e efeito, ou a habilidade em fazer alguma coisa, entre outras generalidades relacionadas com o conhecimento e cognição Baars e Gage (2010). Por outro lado, um objetivo pode ser entendido como um alvo ou lugar que deve ser alcançado por um indivíduo. Assim, objetivo é uma ferramenta para fornecer a um indivíduo uma orientação futura. Da mesma forma, em

Chella e Manzotti (2007) argumenta-se que a consciência é um mecanismo capaz de relacionar experiências passadas a fim de criar novas experiências e definir novos objetivos.

Podemos então utilizar a definição acima para argumentar que um robô em um ambiente dinâmico, como uma sala, não é consciente, uma vez que ele implementa, no máximo, um algoritmo estruturado capaz de modificar seu estado, mantendo seu objetivo inicial de explorar o ambiente. Uma vez que o objetivo do robô se mantém o mesmo durante toda sua atividade, não se pode considerar que esse robô é consciente, segundo as definições de Chella e Manzotti (2007), Baars e Gage (2010).

Na área da Inteligência Artificial, o opção pelo estudo da consciência é motivo de controvérsias entre os pesquisadores, uma vez que há pelo menos duas opiniões conhecidas a respeito Arkin (1998). A primeira opinião, muito difundida entre os roboticistas, é que esse assunto deve ser tratado apenas como um conhecimento filosófico, uma vez que não há consequências práticas diretas. Assim, eles preferem discutir questões mais práticas como visão computacional de baixo nível, representação de conhecimento, resolução de problemas, meta-heurísticas, entre outros problemas.

Contudo, outro grupo de cientistas é adepto de que a consciência é mais do que uma questão filosófica, podendo trazer informações sobre um modelo computacional hierárquico que pode ser utilizado pela área de visão computacional.

Recentemente, foi cunhado um novo termo chamado "Consciência Artificial" Chella e Manzotti (2007). Em STARZYK e PRASAD (2011), os autores propõem um modelo computacional teórico com características relacionadas à consciência. Neste mesmo trabalho, os autores sugerem que um modelo de Consciência Artificial deve ter uma percepção sensorial e ser capaz de prever tomadas de decisões em ambientes através de experiências obtidas anteriormente. Esta ideia está de acordo com as definições de Baars e Gage (2010), Chella e Manzotti (2007).

A conclusão que pode-se tomar com base na literatura da Neurociência até o momento, com relação ao mecanismo de consciência, é que trata-se de um processo que envolve o binômio experiência/objetivo. No entanto, esse processo ainda não é completamente esclarecido, uma vez que muitos aspectos relacionados ainda estão no campo especulativo.

Por exemplo, é razoável pensar que consciência envolve tanto experiências quanto objetivos mutáveis. Além disso, pode abranger múltiplos objetivos que se inter-relacionam em diferentes contextos, dependendo da experiência acumulada.

Por outro lado, na área de Visão Computacional, são muito populares os mecanismos de aprendizagem de máquina (experiência) para classificação de objetos (específicos). Alguém poderia dizer que trata-se de um tipo de consciência, mesmo sendo em um menor grau. Embora

isso pareça uma verdade, contudo considerando os amplos aspectos relacionados ao processo de consciência abordados ainda especulativamente na área de Neurociência, pode-se dizer que os modelos até agora implementados na área de Visão Computacional tratam-se de modelos em estágios ainda embrionários em se tratando do mecanismo de consciência.

Isso significa que, para ampliar a capacidade dos modelos de Visão Computacional no sentido de implementar maior grau de consciência, é necessário incluir no modelo muitos aspectos que na área da Neurociência ainda estão em debate. No entanto, ainda é possível incluir alguns desses aspectos, tais como: inclusão de múltiplos objetivos através da inclusão de detecção de múltiplos objetos na cena; realimentação de experiência, através da inclusão de reforços ou enfraquecimentos de padrões e talvez a inclusão de aprendizagem infinita.

Esses mecanismos são possíveis de serem implementados. No entanto, não há atualmente na literatura de Visão Computacional um modelo que inclua todos esse aspectos simultaneamente de modo a ampliar a implementação de um mecanismo de experiência em relação ao mecanismo tradicional de aprendizagem de máquima para o simples reconhecimento de padrões específicos.

O segundo aspecto da Neurociência, citado nesta tese, que está relacionado com o processo de interpretação de uma cena é o mecanismo de atenção. De acordo com Baars e Gage (2010), o mecanismo de atenção é responsável por definir o objetivo da consciência. A atenção é a habilidade mental em selecionar estímulos, memórias ou pensamentos que são relevantes entre muitos outros que são irrelevantes Corbetta et al. (1990).

De uma maneira mais prática, considere um ser humano em um local totalmente escuro sem qualquer noção de onde se encontra. Suponha, inicialmente, que o objetivo principal desse indivíduo seja responder a pergunta: “onde estou?”. Diversos componentes sensoriais enviarão sinais que o ajudarão a entender melhor o cenário. Primeiramente, a audição poderá informar se está perto de uma estrada, cachoeira ou cidade. O sistema nervoso periférico enviará informações sobre a temperatura do local. O seu conhecimento prévio sobre o mundo o ajudará a utilizar estas informações para compor uma resposta final. Contudo, muitos outros sensores, que podem não ser tão relevantes como os citados anteriormente (paladar, dores superficiais, pensamentos sobre outros assuntos), podem desviar o indivíduo de sua resposta correta. Além disso, a quantidade de informação, caso todos esses sinais fossem considerados, seria tão grande que a pergunta poderia não ser respondida em tempo hábil.

Outro ponto importante é que, mesmo em um único sensor, várias informações competem por atenção entre si. Por exemplo, quando um pessoa conversa com outra em uma festa, muitas outras conversas acabam sendo escutadas por ambas. Contudo, os sinais capturados

pelo sistema auditivo parecem ser “filtrados”, de tal modo que somente as informações relevantes para o diálogo são focadas pelo indivíduo.

Dessa maneira, sem um mecanismo de atenção para focar em cada um dos sensores, poderia não ser possível de obter respostas e ações em tempo hábil e de maneira correta. Assim, como também argumentado em STARZYK e PRASAD (2011), não pode haver consciência sem o mecanismo de atenção.

Especificamente para o sistema visual humano, estudos na área da psicofísica indicam que há dois estágios para o mecanismo de atenção: pré-atentivo e atentivo Frintrop, Rome e Christensen (2010). O estágio pré-atentivo é responsável por extrair informações de baixo-nível sobre a cena em observação. Algumas dessas informações são: cor, forma, movimento, textura, linhas, simetria, excentricidade, posicionamento, entre outras.

Na área de visão computacional, alguns pesquisadores implementam o mecanismo pré-atentivo como parte integrante de seus modelos. Uma das maneiras tradicionais de representar esse mecanismo é através do uso dos mapas de saliência Mesquita e Mello (2013), Cheng Guo-Xin Zhang (2011), Achanta e Susstrunk (2010), Achanta et al. (2009), Oliva et al. (2003), Itti, Koch e Niebur (1998). De uma maneira geral, os mapas de saliência são gerados a partir da utilização de uma ou mais medidas que relacionam uma região da imagem original com suas regiões vizinhas.

Tradicionalmente, os mapas de saliência são representados por imagens em tons de cinza de tal forma que, quanto maior a intensidade de um pixel ou conjunto de *pixels* em um determinada região, mais atenção deve ser dada a essa região. Exemplos de mapas de saliências são mostrados na Figura 13. O valor da intensidade de cada pixel no mapa de saliência será proporcional ao seu destaque em relação a sua vizinhança; isto é, quanto mais uma região se diferencia de sua vizinhança, maior destaque ela terá.

O outro estágio envolvido no mecanismo de atenção é o atentivo. Este estágio é responsável efetivamente pelo processo de decisão para a escolha da região mais importante para receber o foco do sistema. Esse estágio relaciona informações advindas dos próprios conhecimentos e experiências aprendidas pelo indivíduo como também os sinais recebidos pelo estágio pré-atentivo. Essas relações de alto-nível são referenciadas na literatura como modelo *top-down*, pois partem de um processamento de alto nível, como conhecimento, e descem para níveis mais baixo, como as características processadas no estágio pré-atentivo.

Na área de Visão Computacional, o trabalho de Torralba Oliva et al. (2003) propõe um modelo de atenção visual para reconhecimento de objetos. Esse modelo é baseado nas probabilidades de ocorrência de características locais (orientação, cores, texturas e contraste) em um

Figura 13 – Exemplos de mapas de saliência. As imagens localizadas na parte superior são originais. As imagens localizadas na parte inferior são os mapas de saliência. Cheng et al. (2011)

Fonte: Autor “adaptado de” Cheng et al. (2011)

determinado ponto (x,y) das imagens da base. Inicialmente, são computadas as probabilidades de cada característica ocorrer em cada ponto sobre todas as imagens da base. Em seguida, o sistema estima qual característica ocorre com maior frequência em cada região. O mapa de saliência é criado em função dessas probabilidades. Os autores argumentam que as regiões relacionadas com as altas probabilidades de ocorrência não são interessantes para o mapa de saliência, uma vez que elas aparecem com muita frequência nas imagens da base e, portanto, são comuns e não oferecem nenhum grau de discriminação. Assim, os pontos com menores probabilidades são selecionados para fazerem parte do mapa de saliência, que pode vir a ser utilizado diretamente em um modelo de Visão Computacional capaz de implementar o aspecto relacionado ao mecanismo de atenção.

Neste mesmo trabalho, Oliva et al. (2003), também foi proposta a adição de um modelo *top-down* probabilístico baseado em informações contextuais. Para essa implementação, foi necessária uma base supervisionada contendo a localização dos objetos de interesse. De posse dessas localizações, foi aplicado o PCA sobre o vetor de características locais nas regiões onde haviam objetos de interesse. A redução da dimensionalidade sobre o vetor de características foi necessária uma vez que esses dados fariam parte de um modelo probabilístico. Essa transformação nos dados torna a aprendizagem mais eficiente e mantém apenas informações relevantes no modelo. Assim sendo, foi criado um modelo probabilístico associando uma classe de objetos as suas possíveis características locais. Finalmente, um mapa de saliência é gerado a partir dos dois modelos probabilísticos utilizando uma regra de associação. Os mapas de saliência obtidos a partir dos experimentos gerados nesse trabalho foram comparados com os padrões de escaneamento dos olhos de observadores humanos. Os resultados mostraram que as informa-

ções contextuais consideradas no trabalho foram capazes de aproximar os mapas de saliência aos padrões de escaneamento dos humanos.

Considerando os conceitos vistos até aqui, podemos definir a atenção como a habilidade de fazer a escolha de um estímulo e/ou processo mental que está mais associado às intenções (ou objetivos) de um sistema.

O terceiro aspecto abordado nessa tese que está relacionado com reconhecimento visual de objetos são modelos *bottom-up* e *top-down*. Alguns cientistas da área da Neurociência Baars e Gage (2010), e também da computação Laar, Heskes e Gielen (1997), acreditam que esses modelos fazem parte do mecanismo de atenção.

Em Laar, Heskes e Gielen (1997), os autores discutem sobre diversas sub-tarefas que o mecanismo de atenção pode executar. Dentre elas, duas estão relacionadas com os modelos hierárquicos *bottom-up* e *top-down*: atenção exógena e atenção endógena.

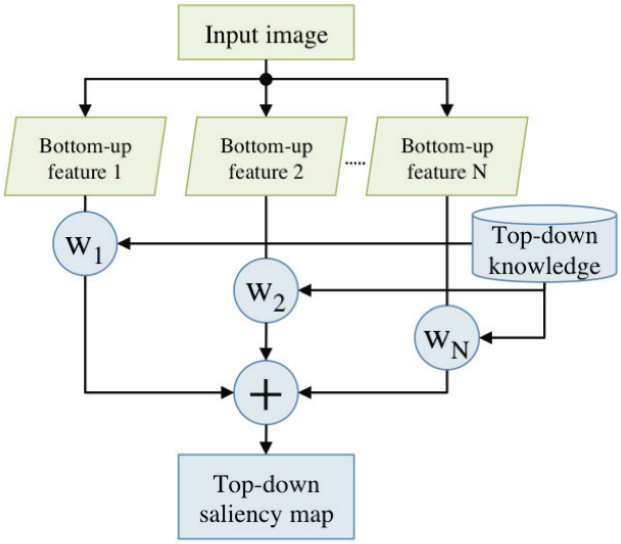
A atenção exógena, também chamada de atenção *bottom-up*, é exercida sempre que uma interferência externa chama atenção de um indivíduo Laar, Heskes e Gielen (1997). Neste caso, imagine uma foto de um cenário como um céu claro. Nele, é comum identificarmos a cor azul de nossa atmosfera e algumas poucas nuvens brancas pairando. Contudo, sempre que um objeto se diferencia muito desse padrão de cores, chamará nossa atenção. Assim, se uma foto semelhante for apresentada, mas com um pássaro cruzando o céu, será rapidamente focada pelo nosso sistema visual. Nesse exemplo, apenas a característica de cor foi ressaltada. Contudo, outras características como forma, textura, tamanho, simetria, movimento, cantos, entre outras, podem ser capturadas pelo sistema de atenção *bottom-up*.

Outra sub-tarefa do mecanismo de atenção relacionada ao modelo hierárquico é a atenção endógena. Também chamada de atenção *top-down*, a atenção endógena é exercida sempre que se utiliza as próprias experiências e conhecimentos de um indivíduo. Assim, a atenção endógena é guiada apenas pelas intenções, estímulos e níveis cognitivos mais altos do observador, sem qualquer interferência direta externa Laar, Heskes e Gielen (1997).

Para exemplificar o uso da atenção *top-down*, pode-se utilizar o exemplo de uma pessoa em uma festa conversando com um grupo de pessoas em particular. Assim que o assunto em discussão não faz mais parte do interesse dessa pessoa, sua atenção pode se voltar para algo que ache mais interessante na festa. Note que a mudança de atenção foi exercida a partir da própria intenção do indivíduo.

Na área de reconhecimento de objetos, alguns modelos computacionais baseiam-se em estruturas *top-down*. O trabalho de Wolfe, Cave e Franzel (1989) sugere um modelo *bottom-up* e *top-down* para representação da atenção visual. O processo *bottom-up* é feito a partir da repre-

Figura 14 – Modelo de atenção proposto por Wolfe, Cave e Franzel (1989). Note a presença do modelo *bottom-up*, através da extração de características; e a presença do modelo *top-down*, através das informações *a priori* da base de conhecimento.



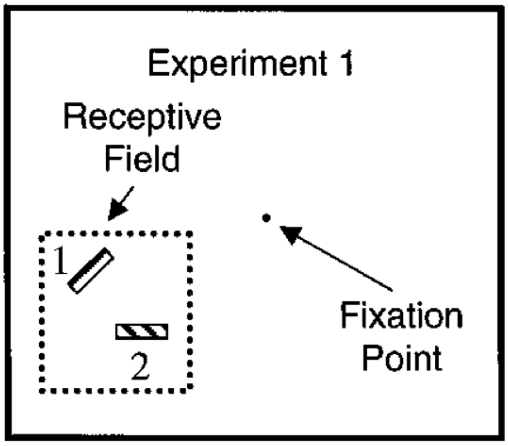
Fonte: Autor “adaptado de” Wolfe, Cave e Franzel (1989)

sentação da imagem por características de baixo nível. A representação do processo *top-down* é feita a partir de um conhecimento prévio sobre a importância de cada característica. Assim, há uma ponderação de cada sinal (característica) de acordo com as próprias experiências (base de conhecimento) do sistema (veja Figura 14). Finalmente, o mapa de saliência é feito a partir da somatória de todas as características observadas ponderadas pela base de conhecimento.

O último aspecto citado nesta tese relacionado ao processo de detecção de objetos em uma cena é a competição e colaboração entre mecanismos sensoriais. Sabe-se que o cérebro humano trabalha com muitos sinais ao mesmo tempo. Essa característica abre uma questão importante sobre seu funcionamento: como todas essas informações em paralelo são gerenciadas e utilizadas pelo cérebro?

Com o objetivo de explicar melhor essa questão, considere o exemplo de um animal em busca da presa. Em um primeiro momento, sua atenção poderá estar focada em seu sistema auditivo, capaz de informar sobre a presença de uma presa ao seu redor. Tão logo sua visão volte-se ao local indicado pelo sistema auditivo, seu sistema visual será capaz de informá-lo mais precisamente o local possível da presa. Nesse momento, há duas possíveis interpretações para o predador: o sistema visual informa que o alvo é de fato uma presa; ou o sistema visual informa que o alvo não é uma presa e não é de seu interesse. Neste caso, o sinal enviado pelo sistema auditivo entrará em competição com o sinal advindo do sistema visual. A interpretação dos dois sinais fará com que o predador julgue a presença ou não dessa presa.

Figura 15 – Ilustração dos estímulos presentes nos experimentos de Reynolds, Chelazzi e Desimone (1999).



Fonte: Autor “adaptado de” Reynolds, Chelazzi e Desimone (1999)

A questão da competição é apenas uma característica neural para tratar as informações vindas não somente de diferentes sistemas sensoriais, mas também de um mesmo sistema sensorial. Além dessa característica, o cérebro parece ter outro mecanismo capaz de associar informações através de um aprendizado. Considerando o mesmo exemplo dado anteriormente, se o predador encontrar de fato a presa através do sistema auditivo e visual, o sistema olfativo poderá se beneficiar, uma vez que seu cérebro irá associar o cheiro sentido no momento da caça com os sinais auditivos e visuais. Em um segundo momento, o animal poderá utilizar o sistema olfativo para detectar outra presa semelhante, melhorando sua capacidade de identificação de presas.

Em Moran e Desimone (1985), Desimone (1998), Reynolds, Chelazzi e Desimone (1999), Boynton (2005), os autores estudam a influência de dois estímulos visuais competitivos em cérebros de primatas. A Figura 15 ilustra a presença desses dois estímulos, indicados pelos números 1 e 2. Nessa figura, o ponto central representa a região onde um primata foi treinado para manter seu foco visual durante todos os experimentos.

Inicialmente, foi elaborado um conjunto de 16 estímulos, todos formados com retângulos de diferentes orientações e cores. No total, 4 orientações (0^o,45^o,90^o,135^o) e 4 cores (vermelho, azul, verde e amarelo) foram escolhidas.

O experimento consistiu em mostrar um dos 16 tipos de retângulos na região 1 ou 2 isoladamente para o primata e medir a quantidade de disparos neurais na região V4 (responsável principalmente pela percepção de orientações) de seu cérebro. Quando o primata foi submetido a um teste com dois estímulos, apresentados simultaneamente nas regiões 1 e 2, a quantidade

de atividade cerebral na região V4 ficou entre as medições obtidas quando cada estímulo foi apresentado isoladamente.

Mais tarde, em Reynolds, Chelazzi e Desimone (1999), os autores constataram que a quantidade de atividade cerebral presente quando os dois estímulos são apresentados ao mesmo tempo é uma média ponderada sobre quantidade de atividade medida quando os dois estímulos foram apresentados isoladamente. Esse resultado sugere que os dois estímulos são considerados pelo cérebro ao mesmo tempo e, além disso, que pode haver uma "competição" por atenção sobre cada estímulo.

Na área de Visão Computacional, em Rodrigues, Giraldi e Araujo (2005) é proposto um modelo bayesiano para recuperação de imagens baseado no conteúdo (CBIR). Para este sistema, duas características na forma de vetores são analisadas: cor (KC) e forma (KF). Os autores propõem um modelo matemático baseado em probabilidade condicional. O modelo estima a probabilidade condicional de uma imagem I_j ser retornada dada uma imagem de busca Q. Mais formalmente, o sistema pode ser representado através da equação bayesiana $P(I_j|Q) = P(I_j|KC, KF)$.

Contudo, $P(I_j|KC, KF)$ não é obtido da maneira convencional. Neste caso, os autores propõem a Equação (11) como uma maneira alternativa baseada na lógica de *ou-exclusivo* na forma $1 - (1 - t_1) \times (1 - t_2) \times \ldots \times (1 - t_n)$, onde t é um termo probabilístico.

$$P(I_j|KC, KF) = 1 - (1 - P(I_j|KC))^u \times (1 - P(I_j|KF))^v \tag{11}$$

A Equação (11) apresenta diversos pontos interessantes. O primeiro deles é que ela pode ser facilmente estendida para mais características, simplesmente adicionando novos termos no padrão $(1 - P(I_j|Kx)$, onde Kx é um novo vetor de características. O segundo ponto interessante é que há parâmetros potencializadores para cada termo; isto é, as variáveis u e v podem modelar o grau de importância de cada característica. No trabalho de Rodrigues, Giraldi e Araujo (2005), os valores dos parâmetros u e v, nomeados como "parâmetros de replicação de evidências semânticas", foram calculados experimentalmente.

O terceiro ponto, mais relevante para esta Tese, é que podemos interpretar esta equação como um modelo competitivo e colaborativo. No caso da característica de competição, podemos interpretar que todas as imagens I_j do banco de dados estão competindo entre si para serem escolhidas pelo sistema.

Para explicar o comportamento colaborativo, pode-se dividir o modelo em 3 casos principais: quando a probabilidade de um dos termos é igual a 0; quando a probabilidade de um dos termos é igual a 1, e quando a probabilidade dos dois (ou mais) termos são diferentes de 0 e 1.

Quando o valor da probabilidade de um termo é igual a 0, o seu termo $(1 - P(x))$ será 1. Isso significa que outro termo será responsável por alterar o valor final da equação. Assim, se um termo "não tem certeza" sobre sua característica, ele deixará a cargo de outro(s) termo(s) decidir(em), havendo a colaboração desse termo.

Quando o valor da probabilidade de um termo é igual a 1, o seu termo $(1 - P(x))$ será igual a 0. Isso significa que, o valor da equação será 1. Note que, basta um termo ser igual a 0 que a equação ficará igual a 1. Isso significa que, se um termo "tem certeza" sobre uma característica, sua escolha é "cedida" pelos outros termos, havendo a colaboração dos outros termos.

Finalmente, quando os valores das probabilidades de todos os termos são diferentes de 0 e 1, todos contribuirão para o resultado final. Isso significa que "o conhecimento" de cada termo será considerado na resposta final, havendo colaboração entre todos os termos.

Outros trabalhos na área da Inteligência Artificial também exploram esses e outros conceitos da Neurociência para a criação de modelos computacionais.

A Tabela 1 apresenta alguns modelos computacionais citados nessa Tese que foram utilizados em trabalhos com o objetivo de refletir alguns comportamentos baseados na Neurociência. Nesta tabela, são apresentadas algumas vantages e desvantagens da utilização de cada modelo.

Modelo Computacional	Utilização do modelo em conceitos da Neurociência	Vantagens (V) e Desvantagens (D)	Referências
Modelos Supervisionados			
Redes Neurais	• Cognição Humana • Reconhecimento de Objetos	**V:** Habilidade de classificar dados não lineares. **D:** O modelo aprendido se torna uma "caixa-preta" (não é possível de depurar. **D:** Só é possível incluir conhecimento na rede a partir de exemplos de treinamento.	Yang, Shu e Shah (2013) Rowley, Baluja e Kanade (1998) Fukushima (1980) Behnke (2003a)
Never Ending Learning (Aprendizagem Infinita)	• Cognição Humana • Adaptação a ambientes dinâmicos	**V:** Capacidade de aprender em tempo de classificação. **V:** Capaz de aprender novos casos em ambientes dinâmicos.	Carlson et al. (2010)
Máquinas de Vetores de Suporte (SVM)	• Cognição Humana • Reconhecimento de padrão	**V:** Hiperplano de separação estabelecido entre a maior distância entre os pontos da fronteira entre as classes. **D:** Trabalha somente com 2 classes. Para mais classes, deve-se adicionar mais classificadores SVN, o que pode tornar o modelo ineficiente.	Lan et al. (2013)
Redes Bayesianas	• Cognição Humana • Reconhecimento de Objetos • Mecanismo pré-atentivo	**V:** Capacidade de modelar incerteza de maneira estatística. **D:** A falta de dados estatísticos pode demandar por técnicas alternativas para estimar as probabilidades *a posteriori*.	Neapolitan (2003) Vasconcelos e Lippman (1998) Meng et al. (2011)

Redes Complexas	• Cognição Humana • Relacionamento entre entidades • Memória • Adaptação a ambientes dinâmicos	**V:** As propriedades das redes complexas podem prever comportamentos de sistemas dinâmicos. **V:** Flexibilidade para descrever diversos sistemas naturais. **D:** Algoritmos estatísticos são de alta complexidade computacional e técnicas heurísticas devem ser propostas para trabalhar com grandes massas de dados.	Backes, Casanova e Bruno (2013a) Newman, Barabasi e Watts (2006) Watts (2004) WATTS e STROGATZ (1998)
Modelos Não-Supervisionados			
Decomposição de Valores Singulares (SVD)	• Aprendizagem autônoma • Capacidade de generalização • Identificação das características mais importantes de um sinal	**V:** Redução de dimensionalidade considerando as principais tendências de um sinal. **D:** Para uma matriz $m \times n$, sua complexidade é $O(\min(n^2m,nm^2))$. **D:** As matrizes U e V são densas.	Bo, Ren e Fox (2011) Bo, Ren e Fox (2012)

PCA	• Aprendizagem autônoma • Capacidade de generalização • Identificação das características mais importantes de um sinal	**V:** Redução de dimensionalidade considerando as principais tendências de um sinal. **D:** As direções que maximizam a variância dos dados podem não ser as direções que maximizam as informações. Para maximizar informações utiliza-se LDA (Linear Discriminant Analysis), porém os dados devem ser supervisionados.	Mesnil et al. (2013)
Modelos de Aprendizagem por Reforço			
Q-Learning	• Mecanismo de recompensa • Mecanismo de Decisão	**V:** Modelagem através de MDP (*Markov Decision Process*). **D:** Para modelos com muitos estados e ações, há excessivo uso de memória principal.	Draper, Bins e Baek (1999) Paletta e Pinz (2000) Piñol et al. (2012)

Tabela 1 – Modelos computacionais e a influência da Neurociência na computação

Esta seção apresentou algumas características sobre o processo de reconhecimento de objetos advindas da Neurociência que podem ser utilizadas em modelos computacionais. Além disso, foram apresentadas algumas influências da Neurociência em modelos computacionais utilizados em trabalhos científicos. No Capítulo 3, alguns aspectos da Neurociência serão considerados para criação de um modelo computacional para reconhecimento de objetos.

2.4 MODELOS COMPUTACIONAIS DE RECONHECIMENTO DE OBJETOS INSPIRADOS BIOLOGICAMENTE

2.4.1 Redes Convolucionárias

Redes convolucionárias (CNN) foram inspiradas na estrutura hierárquica do cérebro de primatas por Fukushima (1980). O modelo foi chamado de Neocognitron. Posteriormente inspiraram outros modelos melhorados como Lecun et al. (1998). Desde essa época, até os dias de hoje, estão sendo melhoradas por vários grupos Behnke (2003b), Simard, Steinkraus e Platt (2003) e tem demonstrado performances impressionantes em vários problemas relacionados ao aprendizado de máquina Schmidhuber (2012).

As redes CNN são uma família de modelos hierárquicos em vários estágios, cada um com níveis diferentes de características. Cada nível recebe como entrada uma imagem que é varrida em diferentes lugares. A saída de cada nível é um mapa de característica LeCun e Bengio (1998).

Apesar do grande entusiasmo, principalmente com o desenvolvimento de hardware que permite níveis de abstração cada vez maiores e consequentemente resultados cada vez mais impressionantes, a inspiração biológica-neural ainda é limitada. Estruturas cerebrais importantes não são contempladas, a não ser que sejam adaptadas via a necessidade de uma aplicação específica, tais como: reconhecimento baseado em contexto, atenção precoce, tardia e competição-colaboração de características. A sua principal alegação de inspiração biológica é a sua estrutura hierárquica que pode representar processos de reconhecimento cerebrais top-down.

2.4.2 Transformações Biológicas (BT)

Transformações biológicas (BT) também é um modelo bio-inspirado Sountsov, Santucci e Lisman (2011), baseado na hierarquia do sistema visual de primatas. A principal alegação é de que é capaz de mimetizar a área V1 das vias visuais. Para isso, implementa uma série de

filtros sensíveis à orientação em intervalos de tempo. Embora a inspiração tenha sido realmente biológica, muitos aspectos do sistema visual de primatas não foram levados em consideração. Consequentemente, características como visão baseada em contexto, competição e colaboração de vários tipos de características não são contemplados.

2.4.3 VisNET

Trata-se de um modelo hierárquico que simula as vias visuais ventrais para reconhecimento de objetos. Possui quatro sucessivas camadas de redes neurais SOM (mapas auto-organizáveis de Kohonen). Neurônios, que são dispostos em hierarquias cada vez mais altas, possuem grandes campos receptivos (áreas nas imagens que avaliam). Cada nível no modelo é uma área específica nas vias visuais dos primatas, em termos de tamanho e campo receptivo Tromans, Harris e Stringer (2011), Riesenhuber e Poggio (1999). O modelo pode ser treinado para simular o aprendizado infinito Stringer e Rolls (2008). Os dois principais aspectos que simulam o sistema visual são: (a) hierarquia e (b) competitividade - colaboração de redes e neurônios. O modelo proposto lida bem com generalizações de múltiplas visões de objetos. Entretanto, os tipos de estímulos produzidos ainda são simples como aqueles utilizados em testes psicofísicos controlados.

2.4.4 Modelo V1

A área V1 é o primeiro estágio do processamento de informação no sistema visual de primatas, sendo uma porta de entrada para subsequentes processamentos de baixo, médio e alto-níveis. O modelo V1 é uma representação básica inspirada na área V1. Neste modelo, uma população de células são modeladas e alimentadas somente pela característica de luminância das imagens. Para isso, são utilizados filtros de Gabor em 4 direções diferentes (0^o, 90^o, 45^o e -45^o). Assim, os campos receptivos de células complexas são modeladas realizando a operação MAX, que consiste em escolher a maior entrada para determinar a saída. As saídas de todos os neurônios são então conectados em um vetor, que simula o padrão de saída da V1.

2.4.5 Modelo Piramidal Baseado em Wavelets de Funções de Gabor (GWP)

O modelo GWP foi usado por Kay et al. (2008) para simular a resposta de células no sistema visual primário de humanos. Esse modelo representa cada estágio do sistema visual

humano por um conjunto de funções Gabor de vários tamanhos e direções, frequências espaciais e fases em um grid regular. Assim como o BT, VisNet e V1 não implementa muitos aspectos de alto-nível do sistema visual, e é especializado em um conjunto fixo de imagens de treinamento.

2.4.6 HMax

O primeiro modelo qualitativo de células simples do sistema visual humano foi proposto por Hubel e Wiesel (1968), Hubel e Wiesel (1959). O sistema deles era um modelo hierárquico que respondia a spots de luz. Em seguida, simples células respondiam a barras de orientação dentro de um campo receptivo (imagem). Os próximos estágios simulavam células que respondiam a estímulos mais complexos, invariantes a posição e orientação. A partir desse trabalho pioneiro, muitos outros foram propostos para reconhecimento de objetos, entre os quais o trabalho de Riesenhuber e Poggio (1999) é provavelmente o mais popular. Mais tarde, esse modelo foi chamado de HMAX ("Hierarquical Model and X") por Tarr (1999).

O modelo genérico HMAX consiste de uma hierarquia de quatro estágios computacionais (S1, C1, S2 e C2), intercalados. As camadas S realizam operações lineares e as camadas C operações não-lineares, chamada MAX, que usa valores de entrada máximo para processar as saídas. Essa estratégia favoreceu a especificidade do sistema e invariância com relação à translação e rotação.

Embora argumente-se que o modelo HMAX simula o processamento nas principais vias visuais, ainda não apresenta, em sua forma original, aspectos importantes do sistema visual humano, tais como visão contextual e os processos de atenção precoce e tardia. Além disso, os modelos sofrem de especificidade de aplicações, não sendo uma tarefa trivial o treinamento para generalização de cenas.

2.5 REDES COMPLEXAS

As Redes Complexas são utilizadas para descrever os mais diversos tipos de sistemas Newman, Barabasi e Watts (2006). Exemplos de redes que só recentemente foram possíveis de modelar e estudar são: redes sociais (tais como *orkut*, *facebook*, comunidades sociais em geral), redes biológicas (cadeias de *DNA*, Genoma), redes epidêmicas Bose et al. (2013), redes de predador e presa Dunne, Williams e Martinez (2004), redes econômicas Glattfelder (2010), entre outras. Essas redes, até pouco tempo, eram desconhecidas ou difíceis de representar

matematicamente, ou por falta de um modelo matemático adequado, ou devido à ausência de hardware com capacidade compatível para sua implementação.

A principal característica das Redes Complexas é que as ligações entre seus elementos baseiam-se em regras[2] previamente estabelecidas (Figura 16) Newman, Barabasi e Watts (2006). Contudo, seu crescimento acontece de uma forma natural e, algumas vezes, de forma imprevisível. Ao definir-se a regra das ligações, tem-se ao final uma rede composta de muitas interações, tornando possível a extração de características físicas importantes para o entendimento do comportamento do sistema, tais como: tamanho, diâmetro, grau de conexão médio, densidade, entre outras. Assim, Redes Complexas por ser definida como a união da teoria dos grafos e estatística.

Figura 16 – Rede complexa representando todos os produtos comercializáveis Hidalgo et al. (2007). Com esse tipo de modelagem é possível tomar decisões estratégicas locais da economia emergente. O tamanho de cada nó representa a grandeza (em dólares) dos produtos, e as ligações representam a utilização de um produto para produzir outro.

A modelagem de uma rede complexa é semelhante a de um grafo Newman, Barabasi e Watts (2006). Há algumas formas de se modelar um grafo CORMEN et al. (2001). Uma delas é através da utilização de lista de adjacência de arestas, onde normalmente é implementada utilizando uma tabela de três colunas, onde cada linha define uma aresta do grafo. A primeira coluna contém o índice do nó de onde a aresta parte, a segunda informa qual é o nó de destino dessa aresta e a última informa o peso da aresta. Uma vez que a tabela da lista de adjacência

[2]Essas regras são definidas pelo próprio problema proposto. Por exemplo, no caso de rede social, cada nó representa um indivíduo e cada aresta representa amizade entre dois indivíduos.

apresenta um número fixo de colunas e a quantidade de linhas é proporcional a quantidade de arestas, esta modelagem consome espaço da ordem de $O(E)$, onde E é o número de arestas.

Outro tipo de modelagem, cria uma matriz de adjacência de tamanho $N \times N$, onde N é o número de nós. Portanto, o espaço para armazenamento da rede utilizando essa modelagem é da ordem de $\theta(N^2)$. Cada célula dessa matriz representa uma aresta, podendo assumir os valores 0 ou 1, para redes não ponderadas ou outros valores para redes ponderadas. Além disso, se a rede for não-dirigida essa matriz contém somente $N^2/2$ células, dado que somente a parte superior da diagonal principal precisa ser utilizada.

A escolha da modelagem está diretamente relacionada à densidade do grafo. Para redes esparsas, a modelagem por lista de adjacência parece ser a melhor opção, uma vez que o consumo de memória é $O(E)$. Para redes densas, a matriz de adjacência parece ser a melhor escolha, pois seu consumo é $O(N^2)$.

Na Seção 2.5.1 serão apresentados os três modelos de redes complexas mais estudados na literatura, e na Seção 2.5.2 será visto algumas de suas aplicações.

2.5.1 Modelos de Redes

Nesta seção serão abordados os três tipos de modelos teóricos de Redes Complexas. A primeira delas, Redes Aleatórias, é o modelo mais antigo que possui suas conexões baseadas em uma propriedade conhecida por probabilidade de conexão. O segundo modelo, Redes de Mundo Pequeno, é usado em muitos trabalhos para modelagem de redes sociais e, finalmente, Redes Livres de Escala que baseia-se na dinâmica do crescimento das redes.

2.5.1.1 Redes Aleatórias

A teoria das Redes Aleatórias foi desenvolvida inicialmente por Solomonoff e Rapoport (1951) e posteriormente estudada por ERDŐS e RÉNYI (1959). Esse tipo de rede é construída conectando aleatoriamente seus nós em uma proporção conhecida. Há duas representações matemáticas para este tipo de rede. A primeira delas, $G_{n,p}$, onde n é o número de nós e p é a probabilidade de conexão entre quaisquer dois nós. A segunda representação, $G_{n,m}$, onde n é o número de nós e m é o número de arestas. Essas duas representações matemáticas têm interpretações diferentes, uma vez que em $G_{n,p}$ o número de arestas é alterado proporcionalmente a n e, em $G_{n,m}$, o número de arestas é fixo.

Alguns estudos de Redes Complexas Aleatórias encontram algumas características físicas fundamentais para o estudo de seu crescimento, tais como: existência de *giant component*, fase de transição, *small components*, entre outros STAUFFER e AHARONY (1992).

As Redes Aleatórias conseguem modelar algumas redes do mundo real. Porém, alguns fatores podem interferir nos resultados uma vez que, frequentemente, a distribuição de "graus" dos nós em redes reais tende a ser exponencial, *power-law* ou, principalmente, distribuição com picos concentrados em graus pequenos. Uma vez que as redes aleatórias apresentam distribuição de graus binomial, isso impossibilita as redes aleatórias de representar alguns sistemas reais e pode trazer resultados imprecisos caso sejam utilizadas para descrevê-los Albert, Jeong e Barabási (2000), Cohen et al. (2000).

Como será visto mais adiante, as redes aleatórias servem como parâmetro de medição para comparação entre os modelos. A Equação (12) fornece uma medida de quanto uma rede tem características de rede aleatória para as medidas de coeficiente de clusterização e diâmetro. Essa equação resulta em 1 para redes aleatórias e um número muito maior que 1 para redes de mundo pequeno WALSH (1999).

$$\mu = \frac{C/C_{rg}}{\ell/\ell_{rg}} \tag{12}$$

Na Equação (12), C e ℓ são respectivamente o coeficiente de clusterização médio e a distância média entre todos os nós da rede estudada, C_{rg} e ℓ_{rg} são respectivamente o coeficiente de clusterização médio e distância média entre todos os nós estimados para uma rede randômica com o mesmo número de nós e arestas (ver também Seção 2.5.3.3 para coeficiente de clusterização).

2.5.1.2 Redes de Mundo Pequeno

Com o objetivo de modelar redes sociais e estudar o comportamento da proliferação de doenças, internet, redes metabólicas, entre outras, foi criado um modelo de rede complexa que utiliza uma variedade de técnicas da física estatística WATTS e STROGATZ (1998). Duas principais características observadas em redes do mundo real levaram à criação das redes de mundo pequeno. A primeira delas é que a média das distâncias entre os nós da rede cresce logaritmicamente de acordo com o número total de nós. Isso significa que a medida em que a rede cresce, suas distâncias crescem mais lentamente. Para fazer a medição dessa característica nas redes, deve-se calcular a média aritmética das distâncias entre todos os nós da rede. A segunda característica é que redes de mundo pequeno possuem alto Coeficiente de Clusterização

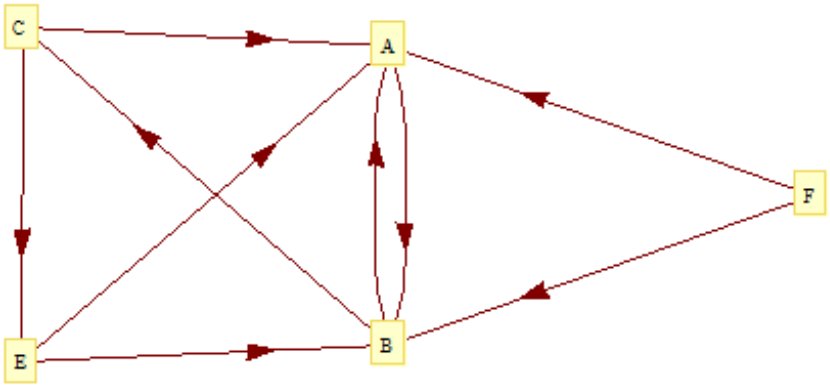
médio (visto adiante na Seção 2.5.3.3). Para isso acontecer, a vizinhança de um nó deve ser altamente conectada entre si.

O cálculo do coeficiente de clusterização de um nó i é feito a partir da relação entre a quantidade de conexões existentes entre os vizinhos de i e a quantidade máxima possível Newman, Barabasi e Watts (2006). Por exemplo, considere a rede apresentada pela Figura 17. Pode-se calcular o coeficiente de clusterização do nó *"A"* da seguinte forma:

$$CC_A = \frac{\textit{(número de vizinhos de } A \textit{ conectados entre si)}}{|\textit{Vizinhança}(A)| \times (|\textit{Vizinhança}(A)| - 1)} = \frac{4}{4 \times 3} \tag{13}$$

Onde *Vizinhança*(A) é o conjunto de nós que se conectam a A (em qualquer direção).

Figura 17 – Rede com 5 nós



As redes de mundo pequeno assumiram uma importância fundamental para o estudo teórico e prático das redes complexas, uma vez que permitiram a análise do comportamento de alguns sistemas naturais e a possibilidade de observar o comportamento da distribuição de graus, fundamental para o surgimento das Redes Livres de Escala, abordadas na Seção 2.5.1.3 Newman, Barabasi e Watts (2006).

2.5.1.3 Redes Livres de Escala

As redes livres de escala surgiram a partir da observação da distribuição de graus de alguns modelos estudados nos trabalhos de Price (1965), ALBERT, JEONG e BARABASI (1999), FALOUTSOS, FALOUTSOS e FALOUTSOS (1999), BRODER et al. (2000), ALBERT e BARABÁSI (2002). Algumas contribuições científicas mostraram que a lei de potência é uma característica comum observada sobre a distribuição de graus extraídas a partir de

redes do mundo real, tais como: Web ALBERT, JEONG e BARABASI (1999), atores de filmes WATTS e STROGATZ (1998) e redes de citações científicas REDNER (1998).

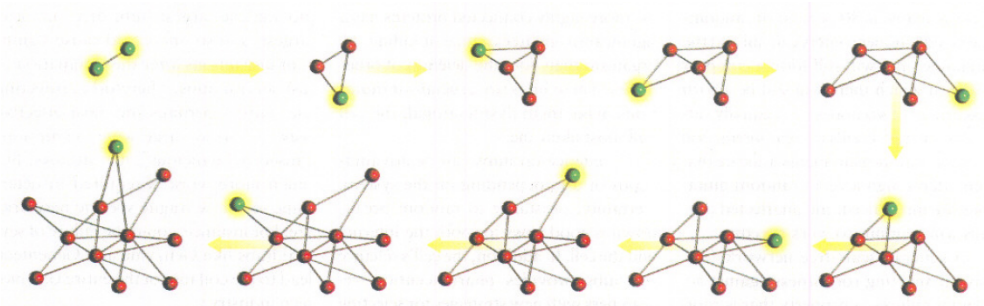
A lei de potência, descrita pela Equação (14), contém um parâmetro de ajuste γ e fornece uma estimativa sobre a probabilidade de um nó ter grau k.

$$P(k) \sim k^{-\gamma} \tag{14}$$

As redes livres de escala também apresentam o efeito de mundo pequeno. Sendo assim, as distâncias entre os nós da rede tendem a crescer logaritmicamente em função da quantidade de nós.

As redes livres de escala são fundamentadas na dinâmica do crescimento de redes naturais; isto é, diferentemente das redes aleatórias, onde o número de nós é fixo e as arestas são inseridas aleatoriamente, as redes livres de escala crescem de acordo com o conceito *ligação preferencial* a cada inserção de um novo nó. A *ligação preferencial* é uma característica que rege a forma com que novas arestas são inseridas na rede. Mais especificamente, quando um novo nó é adicionado, a probabilidade desse nó ligar-se com outro nó de grau elevado é proporcionalmente maior do que ligar-se com um nó de baixo grau ALBERT e BARABÁSI (2002). A Figura 18 ilustra a evolução de uma rede livre de escala quando novos nós (verdes) são inseridos na rede.

Figura 18 – Evolução de uma rede livre de escala BARABASI e BONABEAU (2003). Nesta figura, um novo nó é representado pela cor verde e nós antigos são representados pela cor vermelha



2.5.2 Exemplos e Aplicações de Redes Complexas

Todo o ferramental desenvolvido para as Redes Complexas permitiu sua aplicação em diversas áreas. Esta seção abordará algumas delas.

2.5.2.1 Redes Sociais

Redes sociais são construídas para estudos de relacionamentos interpessoais. Nestas redes, os nós representam pessoas e as arestas suas relações. Essas relações podem modelar amizades, doenças, casamentos entre famílias, comunidades de negócio, colaboração no trabalho, contatos telefônicos, comunicação por e-mail e até mesmo relações sexuais LILJEROS et al. (2001).

2.5.2.2 Redes de Informação

As redes de informação são construídas a partir de bases de conhecimento formal. Nestas redes, os nós representam informações e as arestas a relação entre essas informações.

Em REDNER (1998), foi estudado um modelo de redes complexas para representar as citações entre artigos acadêmicos. As características dessa rede permitiram entender a dependência entre a distribuição de graus dos artigos, que é descrita por uma lei de potência, e o *rank* de classificação de acordo com o número de citações Price (1965).

2.5.2.3 Redes Tecnológicas

Redes tecnológicas são redes complexas utilizadas para modelar a distribuição de facilidade ou recursos, tais como: água, malha elétrica, transporte, linhas aéreas, telefonia (fixa), internet (cabeamento e roteadores), entre outros.

Em GOVINDAN e TANGMUNARUNKIT (2000), foi proposto um modelo de redes complexas para elaboração de heurísticas capazes de aumentar a fidelidade dos roteadores. O mesmo estudo faz uma análise sobre o mapa físico da internet.

2.5.2.4 Redes Biológicas

Alguns sistemas biológicos tais como vascular, nervoso, circulatório, entre outros, são naturalmente identificados como redes complexas. Um exemplo da aplicação dessas redes é o trabalho de Sporns (2002), no qual foi analisado o cortex cerebral de primatas através do estudo das características físicas da rede.

Outro exemplo da aplicação de redes complexas nessa área é o estudo da dependência entre proteínas. Nesse caso, uma rede complexa pode ser modelada utilizando seus vértices

para representar as proteínas e as arestas suas dependências para sua síntese FARKAS et al. (2002), Guelzim et al. (2002), SHEN-ORR et al. (2002). O estudo de cadeias alimentares também é outro exemplo onde pode-se aplicar redes complexas para o entendimento de diversos ecossistemas Allesina e Pascual (2008).

2.5.3 Características Físicas de Redes Complexas

Frequentemente, ao modelar um sistema através de redes complexas não é possível classificá-la entre aleatória, mundo pequeno ou livres de escala até que algumas propriedades sejam extraídas Newman, Barabasi e Watts (2006). Algumas dessas propriedades, necessárias para este trabalho, são citadas a seguir:

*2.5.3.1 Grau de Entrada e Saída (*In-Out Degree*)*

Dentre muitas características importantes de um nó, pode-se encontrar a quantidade de arestas que chegam ou saem dele: grau de entrada ou grau de saída, respectivamente. Por definição, o somatório do grau de entrada com o de saída resulta no grau de conexão k de um nó. O grau de conexão médio de uma rede complexa é denotado por $< k >$, que é computado através da Equação (15), onde E é o número total de arestas e N é o número total de nós da rede.

$$< k >= \frac{E}{N} \tag{15}$$

Uma vez que os nós de uma rede complexa podem ter diferentes graus de conexão, utilizando a função de densidade de probabilidade desses graus pode ajudar a entender o tipo específico da rede.

Teorias sólidas desenvolvidas indicam que redes livres de escala seguem a distribuição da lei de potência (Equação (14)), onde $P(k)$ é a probabilidade do grau k ocorrer na rede Newman, Barabasi e Watts (2006). Esse comportamento sustenta a idéia que existem muitos nós com baixo grau de conectividade e poucos nós com alto grau de conectividade (chamados *hubs*).

Outra utilidade importante encontrada para o grau de entrada e saída está na checagem de erros na implementação rede. Seja $In(i)$ o grau de entrada de um nó I e $Out(i)$ o grau de

saída. A Equação (16) mostra o relacionamento entre as duas propriedades físicas.

$$\sum_i In(i) = \sum_i Out(i) \tag{16}$$

2.5.3.2 Distribuição de Pesos

Muitas das características apresentadas nesta tese podem ajudar a extrair propriedades estruturais topológicas das redes, uma vez que elas podem não só fornecer informações de grupos e relações entre as entidades, mas também semelhanças entre elas.

No caso da distribuição de pesos, esta é uma propriedade que apresenta a frequência com que cada peso aparece na rede. No trabalho de Barrat et al. (2004) os autores propõem um método que utiliza informações topológicas em conjunto com a distribuição de grau para extrair informações de heterogeneidade de duas redes complexas: uma gerada a partir de dados de transporte aéreo e outra de colaboração científica.

Uma maneira interessante de se observar como o comportamento da distribuição dos pesos está associado à topologia da rede é calcular a entropia da distribuição. Uma entropia baixa significa concentração de informação, levando possivelmente a uma rede sem topologia clara, uma vez que nesse caso, isso só é possível com muitas arestas de pesos semelhantes. No caso contrário, quando a tendência da entropia é alta, isso significa uma distribuição de pesos mais heterogênea, significando que existem vários conjuntos, de mesma quantidades de arestas, com pesos iguais dentro dos conjuntos e diferentes entre esses conjuntos. Nesse caso, não é possível inferir a topologia da rede sem a análise de outras propriedades.

Para a extração da distribuição de pesos deve-se antes determinar o número de discretizações que será utilizada. Após esta etapa, inicia-se a varredura de e contagem de todas as arestas sobre cada intervalo discretizado. Ao final, tem-se a quantidade com que cada faixa de valores de arestas está presente na rede, gerando um histograma de valores absolutos.

O algoritmo para extração da distribuição de pesos é da ordem de $O(E)$ (proporcional ao número de arestas que, no pior caso, pode chegar a $O(N^2)$), uma vez que deve-se percorrer todas as arestas da rede.

2.5.3.3 Coeficiente de Clusterização

A média geral do Coeficiente de Clusterização (ACC) é uma importante característica física de redes complexas com implicações em diversas aplicações. Por exemplo, WATTS e

STROGATZ (1998) definiu uma rede complexa como mundo pequeno se ela apresenta duas propriedades em conjunto. A primeira delas é que a média das distâncias entre todos os vértices da rede (ℓ) deve comparável àquela de redes aleatórias, $\ell/\ell_{rg} \sim 1$. A segunda é que o coeficiente de clusterização médio deve ser muito maior do que o de uma rede aleatória, $ACC/ACC_{rg} \gg 1$. Ambas as propriedades são definidas para uma rede de mesma quantidade de nós e arestas Newman, Barabasi e Watts (2006).

Algumas aplicações importantes como a Internet, World Wide Web, Colaboração Biológica, Co-ocorrência de Palavras, entre outras, apresentam tais características Newman, Barabasi e Watts (2006), enfatizando a necessidade de computar o coeficiente de clusterização da rede.

O cálculo do coeficiente de clusterização expressa o quanto os nós de uma vizinhança de um nó i estão conectados entre si mesmas. Este valor varia entre 0 e 1, onde 0 é uma vizinhança totalmente desconectada e 1 é uma vizinhança totalmente conectada.

Formalmente, considere i como qualquer nó de uma rede complexa, L_i o conjunto de nós que têm uma conexão com i, e $W = \{w_{u,v}|u,v \in L_i\}$ um conjunto de pesos de arestas que conectam cada nó de L_i a outro nó de L_i. A Equação (17) mostra o cálculo do coeficiente de clusterização para uma rede dirigida.

$$CC_i = \frac{|W|}{|L_i| \times (|L_i| - 1)} \tag{17}$$

Esta equação relaciona o número $|W|$ de arestas existentes na vizinhança de i com o máximo número de arestas possíveis para a quantidade de L_i nós. A Equação (17) é capaz de calcular o coeficiente de clusterização, porém ela não considera os pesos das arestas, mas apenas a quantidade de elementos. Em casos de redes ponderadas, essa equação é obviamente limitada e, em alguns casos, não se aplica.

Tendo isso em mente, propomos uma nova equação para o cálculo do CC e adicionamos o peso às arestas, trocando a quantidade $|W|$ na Equação (17) das arestas pela soma dos pesos dessas arestas $\sum W_e$:

$$CC_i = \frac{\sum W_e}{|L_i| \times (|L_i| - 1)} \tag{18}$$

Então, para uma rede com N nós, o coeficiente de clusterização médio ACC é dado por

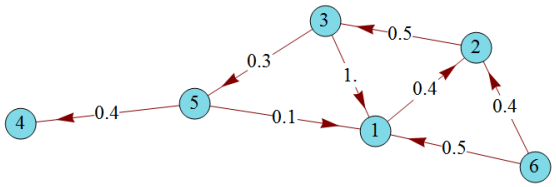
$$ACC = \frac{1}{N} \sum_i CC_i \tag{19}$$

A Equação (18) é idêntica à Equação (17) com as seguintes modificações: o numerador $|W|$ na Equação (17) é o tamanho do conjunto de pesos $w_i \in \{0,1\}$, e na Equação (18) $0 \leq$

$w_i \leq 1$. Isso permite que consideremos valores ponderados entre 0 e 1. Dessa forma, na Equação (17), o coeficiente de clusterização é computado apenas para cada nó j vizinho à i conectado por uma aresta $w_{i,j} = 1.0$. Porém, na Equação (18), a conexão entre i e um nó vizinho j é ponderada. A consequência dessa modelagem é que o coeficiente de clusterização de i não é somente computado levando em consideração a conexão entre a vizinhança, mas também o quanto i está conectado a ela.

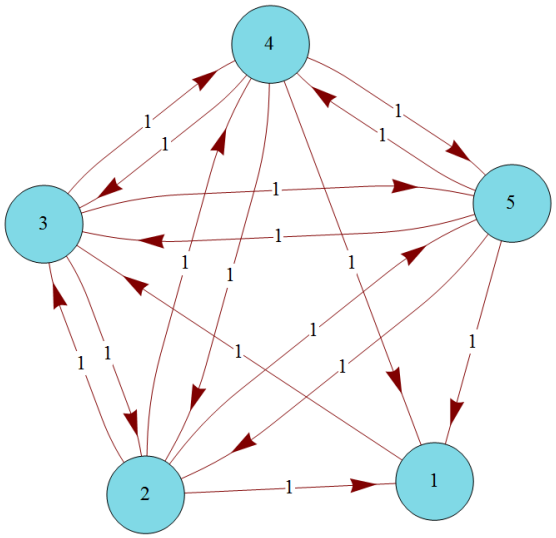
A Figura 19 mostra um exemplo. Para calcular o coeficiente de clusterização do nó $i = 1$, deve-se primeiramente identificar seus vizinhos, nesse caso $L_1 = \{2,3,5,6\}$. O método usado aqui considera todas as arestas conectando quaisquer dois nós do conjunto L_1, gerando o conjunto de arestas $E = \{2 \rightarrow 3, 3 \rightarrow 5, 6 \rightarrow 2\}$ e o conjunto $W_e = \{0.5, 0.3, 0.4\}$. Então, aplicamos a Equação (18), encontrando $CC_1 = 0.1$. Para a Equação (17), $CC_1 = 0.25$.

Figura 19 – Coeficiente de Clusterização



Note que, para o coeficiente de um nó ser igual a 1, é necessário que todas a arestas possíveis da vizinhança tenham $w_{i,j} = 1$. Nesse caso, a Equação (18) se reduz à Equação (17), portanto, ela é uma generalização (veja Figura 20).

Figura 20 – Coeficiente de Clusterização Máximo



Note que a Equação (18) também pode ser utilizada para redes não dirigidas. Para isso, teremos a metade do total de arestas possíveis entre os vizinhos. Sendo assim, o denominador da equação será $|L_i| \times (|L_i| - 1)/2$ e a Equação (18) pode ser re-escrita como:

$$CC_i = \frac{2 \times \sum W_e}{|L_i| \times (|L_i| - 1)} \tag{20}$$

2.5.3.4 Densidade de Conexão Média (K_{den})

A Densidade de Conexão Média (K_{den}) é uma medida física que relaciona a quantidade de arestas existentes em uma rede complexa com a quantidade de arestas existentes em uma rede totalmente conectada (completa), para o mesmo número de nós. Portanto, essa medida varia entre 0 e 1, onde 0 significa uma rede esparsa e 1 significa uma rede densa. Dessa forma essa característica está relacionada com a completude de um grafo.

Em Sporns (2002), o autor comenta estudos que inferem a topologia de redes complexas a partir de dados neurológicos através de medições de densidades locais (em uma determinada parte da rede) e global (toda a rede). Assim, essa medida também pode fornecer informações estruturais das redes.

A Equação (22) apresenta o cálculo do K_{den} para redes não dirigidas:

$$K_{den} = 2 \cdot \frac{|E|}{n^2 - n} \tag{21}$$

Para redes dirigidas, o K_{den} é calculado segundo a equação:

$$K_{den} = \frac{|E|}{n^2 - n} \tag{22}$$

onde $|E|$ é o conjunto de células da matriz de pesos com pesos diferentes de zero.

Conforme as Equações (21) e (22), quando seu valor está próximo a zero, há poucas arestas na rede e ela é considerada esparsa. Caso contrário, próximo a um, a rede contém muitas arestas e é considerada densa.

No caso da rede proposta por este trabalho, ao extrairmos esta característica, estaremos analisando a quantidade de relações em uma mesma característica da imagem. Se o valor obtido for alto, isso sugere que a característica representada na rede apresenta alta conectividade entre diversas instâncias de valores diferentes. Caso contrário, há poucos relacionamentos entre valores distintos em uma determinada característica.

2.5.3.5 Índice de Semelhança de Conexão (ϱ)

O Índice de Semelhança de Conexão $\varrho_{i,j}$ (para $i \neq j$) de uma Rede Complexa mede o quanto o conjunto de conexões de um nó i é semelhante ao conjunto de conexões de um nó j HILGETAG et al. (2000), Sporns (2002). O cálculo de $\varrho_{i,j}$ é feito somando-se a quantidade de nós semelhantes de i e j (que se conectam aos mesmos nós) e dividindo-os pela quantidade total de nós conectados a i e j. Por exemplo, suponha A como conjunto de nós conectados a i, e B o conjunto de nós conectados a j. O índice de semelhança de conexão $\varrho_{i,j}$ é dado pela equação

$$\varrho_{i,j} = \frac{|A \bigcap B|}{|A \bigcup B|} \tag{23}$$

Em SCHAEFFER (2007), a autora comenta sobre clusterização baseada em similaridade de vértices. Um dos métodos abordados por ela consiste em comparar a vizinhança dos vértices e agrupá-los de acordo com a semelhança dessas vizinhanças. Contudo, essa abordagem gera um algoritmo de ordem $O(N^3)$.

Trazendo esta mesma idéia para a área de Redes Complexas, podemos comparar dois nós através de suas conexões e propor uma clusterização baseada nesta medida. Porém, estima-se que a Rede Complexa gerada neste trabalho apresenta dimensão muito elevada, o que inviabilizaria o tempo computacional envolvido na clusterização.

Uma possível solução para isso seria um método para redução da dimensionalidade da matriz de pesos utilizando descritores estatísticos. Diversos descritores podem ser utilizados, tais como: média, mediana, desvio-padrão, moda, variância, percentis, entre outros. A fim de se obter uma maior precisão sobre a comparação das distribuições, pode-se calcular os descritores em algumas faixas da distribuição original. Por exemplo, considere v_i um vetor com os pesos de conexão de um nó i. Para representar essa distribuição, desmembra-se o vetor em p partes. Para cada uma dessas partes calcula-se todos os descritores. Esse processo é repetido para todos os nós da rede, gerando-se uma matriz de descritores, onde cada linha representa as características da distribuição de vizinhança de um nó da rede complexa e cada coluna um descritor.

Uma vez que o foco desse trabalho foca na clusterização de redes complexas, optamos por deixar esta idéia em aberto para futuros estudos.

2.5.3.6 Grau de Reciprocidade (ρ)

Conexões Recíprocas são pares de arestas que conectam dois nós em ambos os sentidos. Formalmente, uma conexão recíproca existe se $M_{i,j} > 0$ e $M_{j,i} > 0$ para $i \neq j$ Sporns (2002). A divisão da quantidade de arestas recíprocas pela quantidade de arestas da rede é uma medida conhecida como "Grau de Reciprocidade" (ou "Fração de Conexões Recíprocas") e é simbolizada por ρ.

2.5.3.7 Probabilidade de Ciclos

Ciclos são caminhos que conectam um nó j a ele mesmo com vértices e arestas distintas. Basicamente, esta medida informa a probabilidade de um caminho ser cíclico Sporns (2002). Por ser uma medida que demanda alto custo computacional, não será abordada nesse trabalho, uma vez que trataremos com bases de dados da ordem de Gigabytes.

2.5.3.8 Matriz de Distâncias, Excentricidade, Raio, Diâmetro

A Matriz de Distâncias armazena o tamanho do menor caminho, d_{ij}, entre um nó i e um nó j HARARY (1969). Se nenhum caminho existe entre i e j então $d_{i,j} = \infty$.

A Excentricidade de um nó i é a distância finita máxima para todos os outros nós da rede. Dessa forma, pode-se obter a excentricidade de um nó i a partir da linha da matriz de distâncias: $ecc(i) = \max_{j=1}^{N}\{d_{i,j}\}$ para $d_{i,j} \neq \infty$.

O Raio de uma Rede Complexa é a excentricidade mínima da rede: $\min_{i=1}^{N}\{ecc(i)\}$. O Diâmetro de uma Rede Complexa é a excentricidade máxima da rede: $\max_{i=1}^{N}\{ecc(i)\}$.

2.5.3.9 Matriz de Alcance

A Matriz de Alcance informa se existe pelo menos um caminho que conecta um nó i a um nó j. Se o caminho existe, denotado por $r_{i,j}$, ele recebe o valor 1 (caso contrário, recebe 0) Sporns (2002).

Assim como a probabilidade de ciclos, as outras medidas físicas conhecidas mencionadas aqui (matriz de distâncias, excentricidade, raio, diâmetro e matriz de alcance, bem como, coeficiente espectral, são computacionalmente inviáveis para grandes bases de dados, como

as estudadas nessa tese. Sendo assim, como já foi explicado, decidimos não implementá-las, optando por um aprofundamento nas discussões das demais medidas.

2.5.3.10 Modularidade

A modularidade baseia-se na comparação de Redes Aleatórias com o modelo observado. Isso significa que, quanto mais distante uma rede for em relação à uma rede aleatória de mesmas características, mais organizada ela será. Essa organização está relacionada ao conceito de *cluster*. Um conjunto de nós é um *cluster* se a quantidade de arestas entre eles é maior do que aquela esperada se a rede fosse totalmente aleatória. A qualidade de um *cluster* C_k para um conjunto de nós $V \in C_k$ pode ser obtida através da Equação (24).

$$Q(C_k) = \sum_{i,j \in V} \left[\frac{A_{ij}}{2m} - P_{ij} \right] \delta(c_i,c_j) \tag{24}$$

Onde

$$P_{ij} = \frac{k_i^{out} k_j^{in}}{m^2} \tag{25}$$

é a probabilidade esperada de conexão entre os nós i e j para uma rede aleatória dirigida de mesmas características (quantidade de nós (N) e densidade de arestas (K_{den})), m é a soma dos pesos da matriz de pesos, A_{ij} é o peso de uma aresta que sai do nó i em direção ao nó j e

$$\delta(c_i,c_j) = \begin{cases} 1 & : \text{ se } i \text{ e } j \text{ estiverem na mesma comunidade.} \\ 0 & : \text{ caso contrário.} \end{cases}$$

Em NEWMAN e GIRVAN (2004) a Equação (24) foi proposta como medida de qualidade da clusterização e também foi apresentado e discutido várias aplicações onde essa medida foi usada com sucesso.

3 PROPOSTA

Nesta seção será abordada a metodologia proposta por esta tese. Na Seção 3.1 será apresentada a base de dados utilizada no trabalho. Na Seção 3.2 os detalhes do meta-modelo serão descritos.

3.1 BASE DE DADOS

Conforme será explicado com maiores detalhes na Seção 3.2.2, este trabalho demanda, inicialmente, por uma base de dados com imagens supervisionadas. Uma vez que este trabalho utiliza informações específicas das imagens para inferir um objeto, é necessária uma base que contenha tanto as regiões que delimitam os objetos em uma cena como também os rótulos dos objetos.

Em Xiao et al. (2010), os autores propõem a criação de uma base de dados, nomeada "SUN Database", que contém 899 categorias de objetos em mais de 130.000 imagens supervisionadas por humanos. Além da supervisão das classes dos objetos, a base também contém informações da classificação das cenas de uma forma geral. A Tabela 2 mostra os 20 rótulos de objetos e cenas mais referenciados pelas supervisões da base de dados. Trabalhos recentes da área de reconhecimento de objetos e classificação de cenas têm utilizado esta base como suporte para treinamento e validação Mottaghi et al. (2014), Agrawal, Girshick e Malik (2014). Nessa Tese, esta base de dados será utilizada como fonte inicial de aprendizado (veja Seção 3.2.2).

A "SUN Database" está organizada através de uma estrutura hierárquica de pastas. No primeiro nível, existem duas pastas: "Images" e "Annotations". A pasta "Images" contém subpastas cujos nomes se referem aos rótulos das imagens. Dentro de cada pasta, um conjunto de imagens que se enquadram no rótulo são armazenadas. Todas as imagens estão em formato JPEG.

A pasta "Annotations" contém as mesmas subpastas da pasta "Images", porém dentro dessas subpastas existem diversos arquivos XML de supervisão das regiões e objetos. Para cada imagem da base um arquivo XML de supervisão é disponibilizado.

Os arquivos XML são padronizados para descreverem um conjunto de regiões (polígonos) com seus respectivos rótulos. A Figura 21 mostra um exemplo de um arquivo XML de supervisão e a imagem correspondente com sobreposição das posições informadas no XML. Na Figura 21 a tag "name", dentro da tag "object", é o rótulo da região descrita pelos pontos

Tabela 2 – As 20 categorias mais rotuladas para cenas e objetos

Categorias de Cena	Núm. de Imagens	Categorias de Objeto	Núm. de Imagens
Living room	2385	Wall	20213
Bedroom	2117	Window	16080
Kitchen	1755	Chair	7971
Beach	1223	Floor	7227
Dining room	1187	Sky	6328
Airport terminal	1152	Ceiling lamp	6268
Castle	1126	Person	6202
Church outdoor	1058	Building	6043
House	972	Trees	5785
Bathroom	956	Ceiling	5284
Playground	909	Tree	4956
Conference room	872	Car	4240
Bridge	870	Door	4135
Highway	861	Cabinet	3102
Market outdoor	853	Plant	3095
Golf course	841	Table	2999
Gazebo exterior	818	Painting	2784
Skyscraper	807	Person sitting	2696
Restaurant	800	Curtain	2525
Warehouse indoor	793	Grass	2427

dentro da tag “polygon”. Neste trabalho, usaremos estas informações para compor o conjunto de entrada de treinamento.

3.2 METODOLOGIA

Neste trabalho, é proposto um modelo de reconhecimento de objetos em cenas naturais que considera várias características armazenadas em uma topologia baseada em redes complexas. Esse modelo é inspirado em modelos biológicos recentemente discutidos na literatura de Neurociências.

Também é proposta uma metodologia experimental, que visa investigar tanto a eficiência do modelo, quanto o seu propósito. De maneira geral, esse modelo baseia-se no registro de observações de características de imagens e seus relacionamentos, aqui chamados de co-ocorrências.

Quando se fala em observação de características, o ponto de interesse está no valor obtido após a extração de algum valor feito sobre uma imagem inteira ou região de interesse (*region of interest*, ROI). Contudo, dependendo da característica observada, este valor pode ser

Figura 21 – Exemplo de arquivo XML descrevendo as regiões da imagem.

```
<annotation>
 <filename>sun_bhlazecjfmiyxbtg.jpg</filename>
 <source>
  <sourceImage>The MIT-CSAIL database of objects and scenes</sourceImage>
  <sourceAnnotation>LabelMe Webtool</sourceAnnotation>
 </source>
 <object>
  <name>trees</name>
  <deleted>0</deleted>
  <verified>0</verified>
  <date>09-Feb-2010 16:41:17</date>
  <id>0</id>
  <polygon>
   <username>anonymous</username>
   <pt>
    <x>0</x>
    <y>51</y>
   </pt>
      ...
   <pt>
    <x>0</x>
    <y>0</y>
   </pt>
  </polygon>
 </object>
 <object>
     ...
 </object>
 <imagesize>
  <nrows>240</nrows>
  <ncols>320</ncols>
 </imagesize>
</annotation>
```

Fonte: Autor “adaptado de” Xiao et al. (2010)

unidimensional ou multidimensional. Nesta Tese, cada valor (uni ou multi-dimensional) será referenciado como uma “instância de característica”.

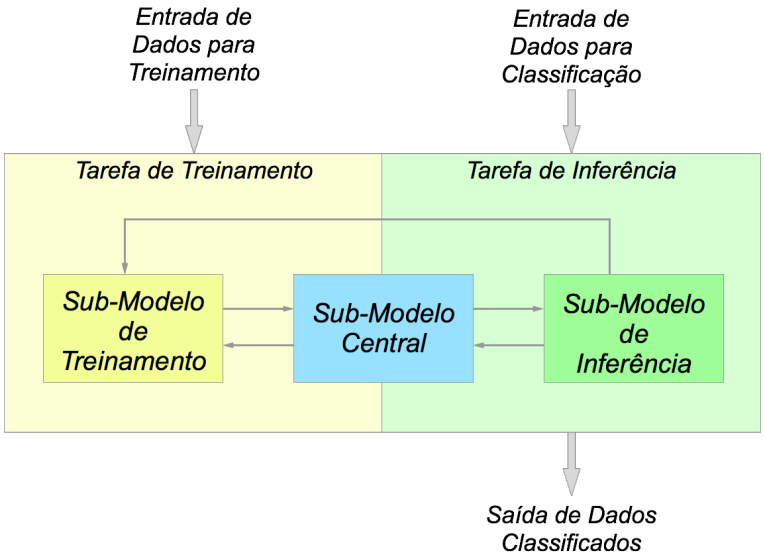
A hipótese central desta tese é que um modelo de co-ocorrência entre as instâncias de características é o principal fator que pode melhorar a qualidade da classificação de um sistema de reconhecimento de objetos. Por exemplo, considere as imagens da Figura 22. Nesta figura, a imagem à direita contém regiões segmentadas por humanos Xiao et al. (2010). Nela, é possível identificar 3 grandes objetos segmentados: céu, oceano e praia. De acordo com a hipótese desta tese, a informação de frequência com que cada par de objetos ocorre em cenas da mesma classe pode contribuir significativamente com a qualidade do classificador. No caso da Figura 22, as co-ocorrências do céu com o oceano, oceano com praia e céu com praia podem ser úteis quando uma nova imagem que contenha oceano for observada no sistema. A criação da base de co-ocorrência, bem como o uso destas informações no reconhecedor, serão detalhados mais adiante neste capítulo.

Esta tese sugere um modelo geral composto por sub-modelos e estruturas de dados. Contudo, esses sub-modelos e estruturas de dados não estão rigorosamente atrelados ao Modelo principal, pois podem ser alterados ou adaptados. Desta forma, o Modelo principal proposto pode ser considerado como um Meta-Modelo, que pode ser alterado para outras aplicações.

Figura 22 – Imagens ilustrando a segmentação feita por humanos. Esquerda: Imagem original. Direita: Imagem supervisionada

Fonte: Autor “adaptado de” Xiao et al. (2010)

Figura 23 – Estrutura do Modelo principal proposto



Fonte: Autor

O Modelo principal proposto opera sobre duas tarefas principais: treinamento e inferência. A tarefa de treinamento é responsável por ajustar os componentes do Modelo principal a partir da apresentação de dados de entrada e saída esperada. Esta tarefa é semelhante aos processos de aprendizagem supervisionada, vistos na Seção 2.2.1. Por outro lado, a tarefa de inferência é responsável por utilizar informações aprendidas para compor uma saída de dados classificados. A Figura 23 ilustra a estrutura geral do modelo proposto. Como pode ser observado nesta figura, para cada uma das duas tarefas do Modelo principal, é proposto um sub-modelo. Embora cada tarefa tenha seu próprio sub-modelo, um terceiro sub-modelo, chamado de Sub-Modelo central, é sugerido para ser responsável por centralizar as informações que são manipuladas em ambas as tarefas.

Este capítulo está organizado da seguinte forma. A Seção 3.2.1 descreve o Modelo principal de forma abrangente. A Seção 3.2.2 aborda o Sub-Modelo de Treinamento; na Seção 3.2.3, o Sub-Modelo Central será detalhado. Finalmente, a Seção 3.2.4 apresenta o Sub-Modelo de inferência. A Figura 24 apresenta o Modelo principal proposto de forma esquemática. Nas próximas seções, essa figura poderá ser utilizada como apoio.

3.2.1 Descrição Geral do Modelo

No Modelo principal proposto, a Rede Complexa contém informações de frequência com que cada par de instâncias de características aparece em uma mesma imagem.

Uma vez que a rede deve estar previamente construída, antes que a segmentação e classificação ocorram de fato, é necessário um módulo capaz de enviar informações dessas frequências. Assim, o Sub-Modelo de Treinamento é responsável por fazer a leitura de bases de dados supervisionadas, extrair características dessas imagens e enviá-las para o construtor da rede complexa.

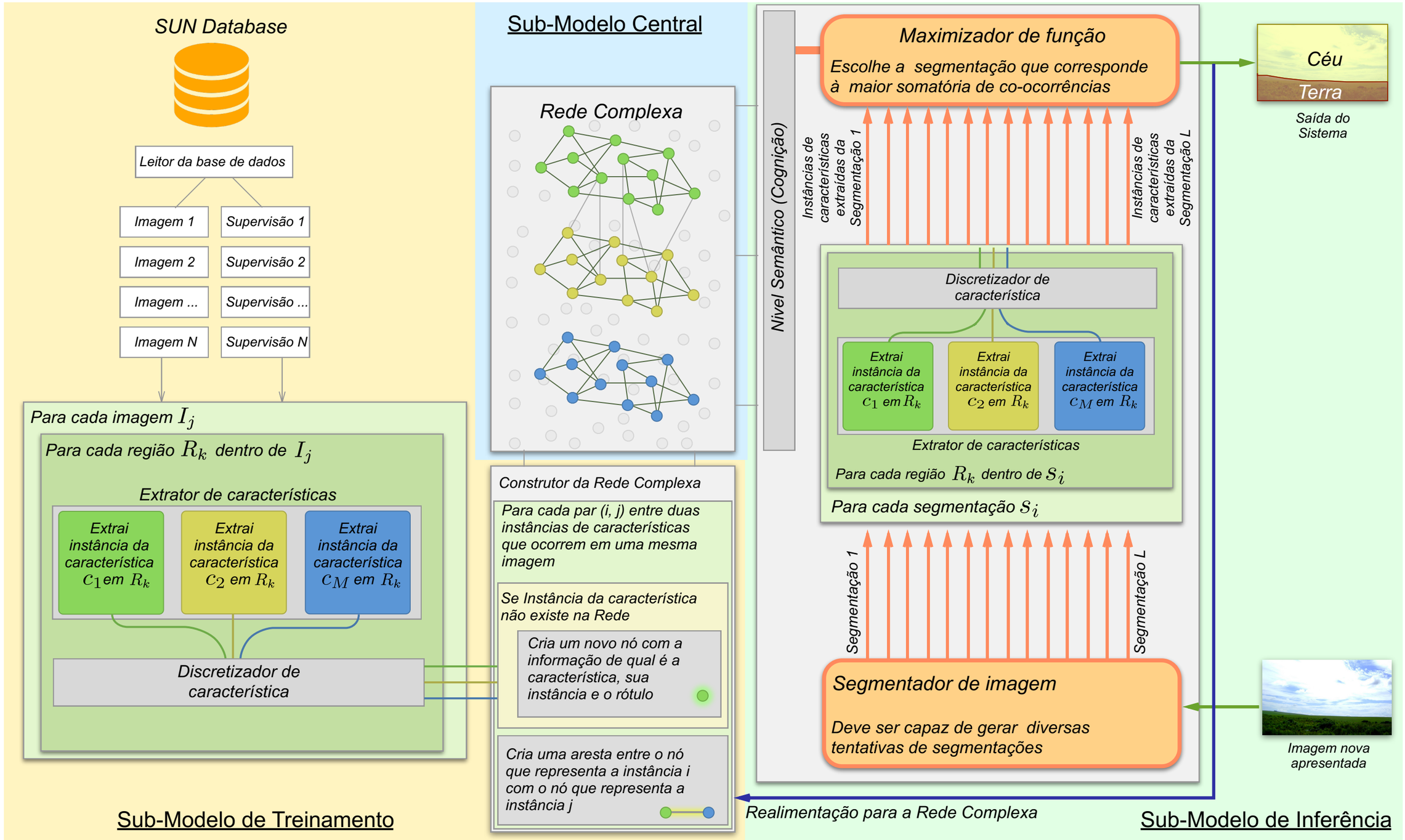
Uma vez treinada, a rede poderá então ser utilizada pelo segmentador e classificador para compor a saída. O módulo responsável por esta tarefa é o Sub-Modelo de Inferência. Neste módulo, uma imagem de entrada é recebida e encaminhada diretamente para um segmentador. Esse segmentador deve ser capaz de gerar um conjunto de diversas segmentações e enviá-lo para o extrator de características. Essas características são enviadas ao maximizador de função, que analisa qual segmentação deve ser escolhida para compor o resultado final. Essa escolha é feita através da maximização do somatório das co-ocorrências entre todos os pares de instâncias de características geradas por cada segmentação. A segmentação que maximizar o somatório é escolhida como vencedora.

3.2.2 Sub-Modelo de Treinamento

O Sub-Modelo de Treinamento é responsável por fazer a leitura de bases de treinamento, extrair as características sobre todas as regiões de todas as imagens, discretizar estas informações e enviá-las para o Sub-Modelo Central.

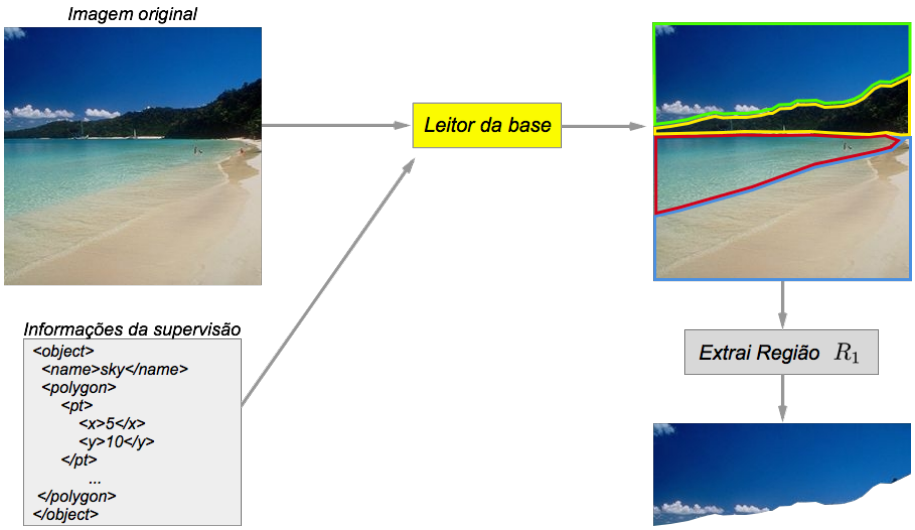
Nesta seção, usaremos a notação I_j para representar a j-ésima imagem do banco de dados e R_k para representar a k-ésima região de uma determinada imagem. Além disso, definimos o conjunto de características $\chi = \{c_1,c_2,c_3,\ldots,c_m\}$ e uma instância da característica c_i extraída a partir de uma região R_k como (c_i,i_k).

Figura 24 – Modelo proposto detalhado



Fonte: Autor

Figura 25 – Leitor da base efetuando a extração da Região R_1



Fonte: Autor

3.2.2.1 Leitor da Base de Dados

O Leitor da Base de Dados supervisionada recebe como entrada uma base da dados com imagens e arquivos de supervisão, normalmente em XML ou JSON, contendo a descrição de cada região R_k de cada imagem I_j. O conjunto de regiões obtido é então enviado ao módulo Extrator de Características. A Figura 25 ilustra essa operação.

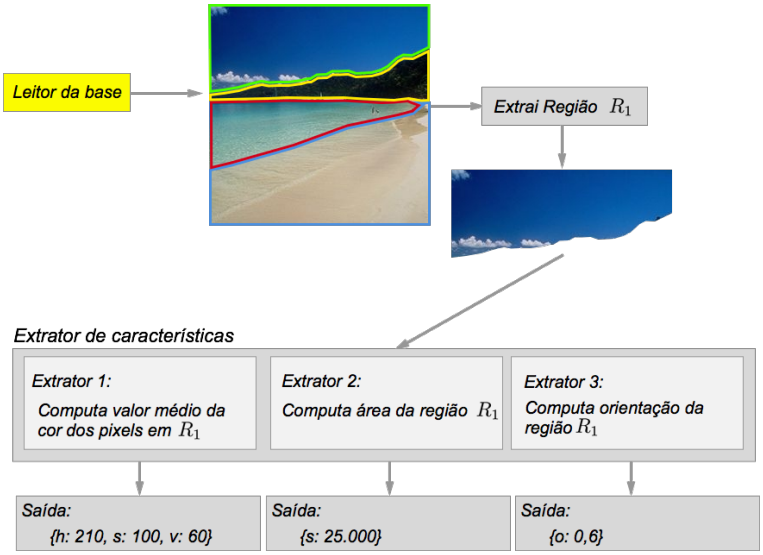
Um ponto adicional a ser informado é que o Leitor da Base pode ser especializado para outras bases de dados. Assim, este módulo não está restrito somente à base de dados SUN.

3.2.2.2 Extrator de Características

O Extrator de Características é um módulo da tarefa de treinamento responsável por extrair todas as instâncias de características em todas as regiões de todas as imagens da base e enviá-las ao Sub-Modelo Central.

O Extrator de Características contém um conjunto de sub-extratores, específicos para cada característica $c_i \in \chi$. Na Seção 2.1.1 (pág. 32), algumas dessas características são citadas como, por exemplo: histograma de cores, orientação, área, Tamura, entre outros. Na Figura (24) esses sub-extratores são representados pelos quadros verde, amarelo e azul. O conjunto de sub-extratores pode ser alterado para que se tenha maior precisão no Sub-Modelo Central. Assim, será definido que o tamanho do conjunto de sub-extratores é M.

Figura 26 – Extração de 3 características a partir de uma região



Fonte: Autor

O extrator de característica inicia sua operação tendo como entrada o conjunto de N imagens da base de dados já processadas pelo Leitor da Base. Para cada imagem I_j desse conjunto, são extraídas M instâncias de características sobre cada uma das R_k regiões de I_j. Isso significa que, se $M = 3$, $N = 1000$ e cada imagem contém 5 regiões, o processamento da base inteira irá gerar 15.000 instâncias de características. Essas 15.000 instâncias farão parte do conjunto de entrada do Discretizador de Característica.

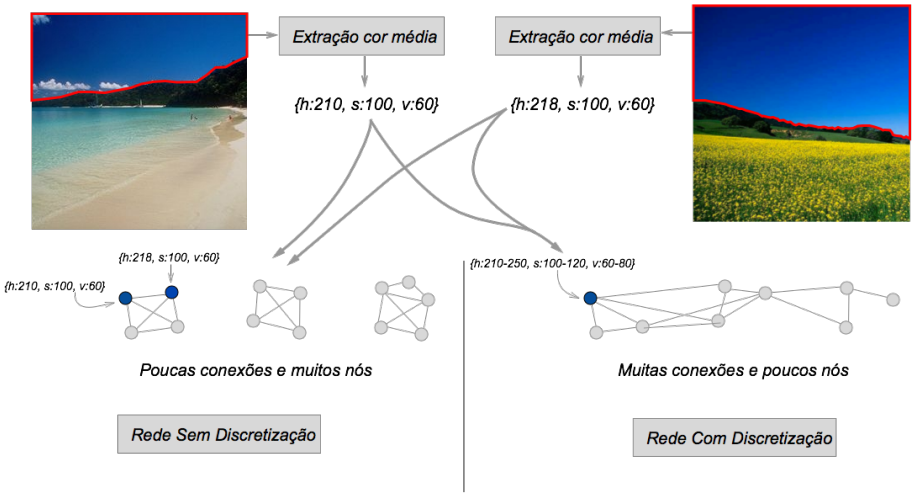
A Figura 26 exemplifica a extração de 3 características sobre uma região de uma imagem supervisionada. Como pode ser observado nessa figura, as características retornam um vetor com valores escalares.

3.2.2.3 Discretizador de Característica

O discretizador de característica exerce um papel fundamental no modelo proposto. Conforme explicado na seção anterior, um banco de dados poderá gerar milhares de instâncias de características. Contudo, se todas as instâncias geradas tiverem valores distintos entre si, não será possível capturar informações suficientes de co-ocorrência para compor a Rede Complexa, uma vez que cada instância será representada por um nó na rede e a co-ocorrência é proporcional ao peso da aresta entre os nós.

Para exemplificar, considere as duas imagens da Figura 27. Ao extrair a característica de cor média sobre as duas figuras, obtemos um conjunto de instâncias. Considere apenas as 2 instâncias extraídas a partir das regiões destacadas em vermelho. Essas 2 instâncias farão parte

Figura 27 – Exemplo de duas redes geradas: uma com discretização de instâncias e outra sem discretização



Fonte: Autor

da entrada de dados do Sub-Modelo Central. Conforme será abordado na Seção 3.2.3, os nós correspondentes a essas 2 instâncias serão buscados e, caso não existam, serão criados. Observe que essas instâncias são muito semelhantes entre si. Entretanto, essa semelhança não justifica a criação de dois nós distintos. Por essa razão, o Discretizador de Característica estabelece o número máximo de instâncias para cada característica discriminando seus valores. Dessa forma, um nó já pertencente à rede que contém uma instância de característica próxima o suficiente será re-aproveitado, gerando assim, mais co-ocorrências. Uma pergunta experimental, que será investigada nessa tese, é a dimensão da discretização que maximiza a separação das regiões ou objetos da cena.

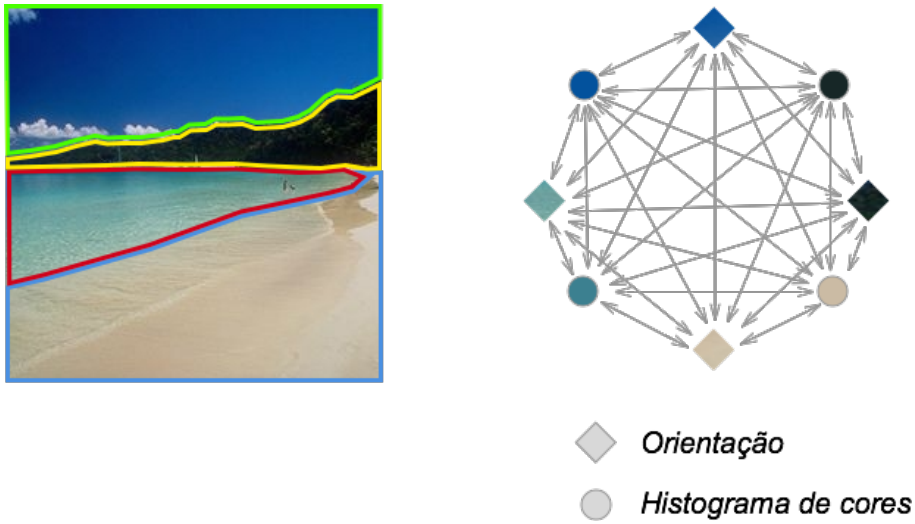
3.2.3 Sub-Modelo Central

Conforme explicado anteriormente, o Sub-Modelo Central é responsável por agregar e gerenciar todas as informações que serão utilizadas em seus sub-modelos vizinhos.

Uma vez que a hipótese central desta tese refere-se às informações de co-ocorrência, é necessário ter um modelo capaz de representá-las. Dessa forma, foi escolhido o modelo de Redes Complexas, uma vez que ele não só é capaz de relacionar entidades, assim como um grafo, como também contém ferramentas capazes de fornecer informações topológicas da rede.

Conforme citado na Seção 3.2, cada instância de característica é relacionada a outra instância de característica sempre que ambas forem encontradas em uma mesma imagem. Essa informação é armazenada através dos nós e arestas da rede; isto é, cada nó da rede é uma

Figura 28 – Representação das regiões da imagem a partir de duas características diferentes: histograma de cores e orientação. Todos os nós estão conectados entre si, uma vez que todas as características co-ocorrem em uma mesma imagem.



Fonte: Autor

instância de característica e cada aresta é a co-ocorrência entre duas instâncias. Mais especificamente, cada nó da rede guarda o valor obtido a partir da extração de uma característica em uma determinada região da imagem.

A Seção 2.1.1 apresentou alguns modelos de representação de imagens através de características. Algumas dessas características serão utilizadas para fazer parte da Rede Complexa[1]. Dessa forma, tanto a informação da instância da característica quanto a própria informação de qual é a característica extraída devem ser modeladas na Rede Complexa. Essa informação será adicionada como um novo atributo nos nós da rede. A Figura 28 ilustra a representação de duas características extraídas a partir de uma imagem (cor e orientação).

De acordo com a Seção 2.5, as Redes Complexas são baseadas em grafos e utilizam a estatística para extrair informações sobre o sistema modelado. A primeira vantagem em se utilizar um modelo baseado em grafos está no fato de que as arestas podem armazenar pesos. Podemos associar os pesos com maior valor às co-ocorrências mais frequentes entre as instâncias. Esta informação pode ser utilizada posteriormente pelo Sub-Modelo de Inferência para aperfeiçoar a classificação dos objetos em cena.

Uma segunda vantagem, também presente na teoria dos grafos, é que as arestas também podem carregar informações de relacionamento espacial entre as regiões. Por exemplo, no caso da Figura 22, o céu co-ocorreu acima da praia. Pode-se então utilizar mais um atributo na aresta

[1]Conforme explicado na Seção 2.1.2, o objetivo principal da Tese não é estudar quais características devem ser consideradas no modelo, uma vez que novas características podem ser incorporadas facilmente.

para guardar a informação de localização espacial relativa. Esta estratégia, no entanto, não será investigada nesta Tese.

A terceira vantagem em se utilizar este modelo está no fato de que os nós também são entidades que podem ter seus atributos customizados. Isto é útil quando se deseja guardar informações extras, assim como qual é a característica que está sendo representada por ele. Na Figura 24 essa informação está sendo representada através das cores dos nós.

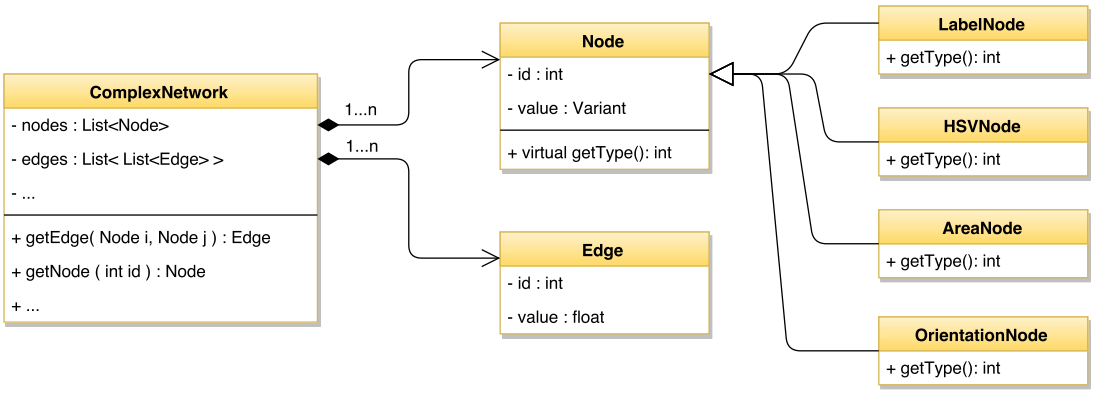
Finalmente, a quarta - e a mais importante - vantagem do uso exclusivo de Redes Complexas para um modelo de reconhecimento de objetos, está no fato de se poder estudar do ponto de vista estatístico e topológico as relações e processos físicos na rede, comparando com modelos já estudados na área. Utilizando como base importantes trabalhos na área de Redes Complexas dedicados a estudos de aplicações bem conhecidas, pode-se conduzir estudos semelhantes para a área de reconhecimento de objetos (que é o caso desta Tese). Um exemplo bem conhecido foram os trabalhos apresentados por linguistas (veja Wachs-Lopes e Rodrigues (2015) para uma revisão geral) que utilizaram redes complexas para a modelagem de textos em diversos idiomas. Como descrito na Seção 2.2.1, a partir dessas modelagens, foram extraídas informações topológicas e parâmetros físicos da rede que permitem caracterizar os idiomas. Um objetivo específico aqui nesta Tese é repetir estudos semelhantes para a área de reconhecimento de objetos. Até onde sabemos, esse tipo de estudo não há na literatura científica.

3.2.3.1 Estrutura de Dados Sugerida

Conforme abordado na Seção 2.5, há basicamente dois tipos de modelagem quando se fala em grafos. A primeira delas utiliza uma matriz de pesos A, tal que cada elemento $A_{i,j}$ representa o peso da aresta que sai do nó de índice i e vai para o nó de índice j. Por outro lado, a segundo tipo de modelagem é utilizando uma lista de adjacência. Nesse caso, cada nó i possui uma lista L_i composta por elementos na forma (j, p), onde j é o nó de destino e p é o peso da aresta que sai de i e vai para j.

No caso específico deste trabalho, foi optado pela utilização da segunda modelagem. O principal motivo por essa escolha foi devido à baixa densidade da rede obtida nos experimentos iniciais na base da SUN. A Figura **??** apresenta um diagrama UML com a modelagem da rede proposta nesse trabalho.

Nessa figura, a classe Node apresenta quatro heranças: LabelNode, HSVNode, AreaNode e OrientationNode. Cada herança representa uma característica que é extraída de uma região de uma imagem. Assim, pode-se agregar novas características ao modelo implemen-



Fonte: Autor

tando novas heranças. Considerando este ponto, pode-se dizer também que trata-se de uma Rede Complexa mista, uma vez que apresenta diversos tipos de nós em uma única rede.

É importante deixar claro que o modelo de implementação proposto não é rigoroso e pode ser implementado de outras formas. Assim, o diagrama de Figura **??** deve servir apenas como uma sugestão para a implementação da rede.

3.2.3.2 Construção da Rede Complexa

O Sub-Modelo Central é responsável por criar e gerenciar a Rede Complexa propriamente dita. Para construir uma rede de modo a minimizar inconsistências (ex. dois nós com a mesma instância de característica, duplicidade de arestas ou nós isolados), um conjunto de regras é definido.

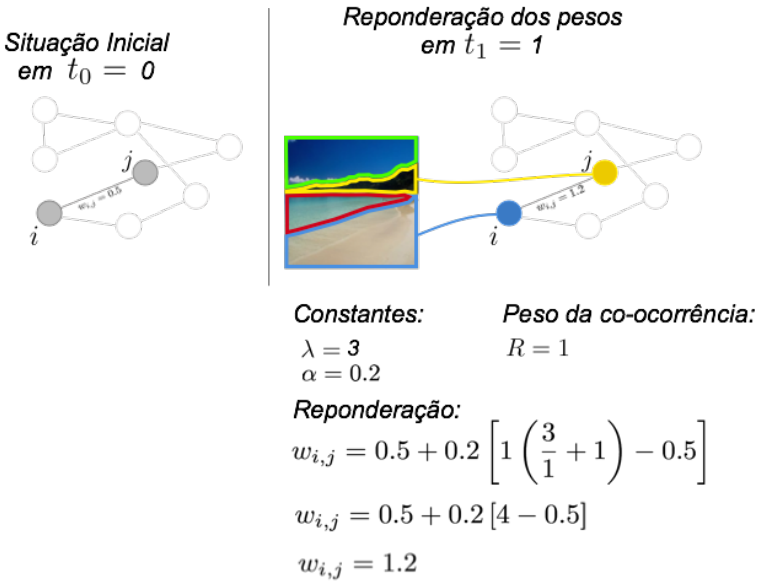
A manutenção da rede é feita sempre que uma nova imagem é analisada. Este módulo recebe um conjunto de informações obtidas a partir do Discretizador de Características sobre todas as regiões da imagem. Matematicamente, o conjunto dessas informações pode ser representado por:

$$F = \{i_{m,k} : \forall m,k\} \tag{26}$$

Nessa expressão, m é o índice da característica que fôra extraída da imagem e i_k é a instância discretizada dessa característica em uma determinada região R_k da imagem.

A modelagem proposta sugere que cada nó da rede seja uma representação de uma instância de característica. É importante destacar que uma mesma instância pode ser obtida a partir de muitas imagens diferentes, bastando para isso elas possuírem o mesmo valor para que seja considerada a mesma instância. Isso significa que um nó poderá representar regiões de diversas imagens.

Figura 29 – Reponderação da aresta $\overline{ij}$ na rede



Fonte: Autor

A relação entre duas instâncias de características é feita através das arestas da Rede Complexa. Essa relação pode ser criada utilizando diversos critérios. Um deles é associar a quantidade de co-ocorrência incrementando o peso da aresta com valores unitários. Uma desvantagem dessa abordagem é que ela utiliza valores absolutos durante toda sua operação. Isso significa que não haverá um ponto de estabilidade no aprendizado e sim, somente incrementos.

Uma abordagem mais precisa é considerar uma unidade de tempo para que seja possível re-aprender as co-ocorrências. Apesar da abordagem anterior re-ponderar os pesos das arestas, ela não "desaprende" as co-ocorrências que não foram reforçadas com o tempo. Essa característica é importante quando se fala em aprendizado infinito, ou quando se deja simular a plasticidade biológica. Matematicamente, essa abordagem pode ser modelada através da Equação (27):

$$w_{i,j} = w_{i,j} + \alpha \left[R \left(\frac{\lambda}{\Delta t} + 1 \right) - w_{i,j} \right] \tag{27}$$

Na equação acima, $w_{i,j}$ é o peso da aresta que liga o nó i ao nó j, $0 < \alpha \leq 1$ é uma constante denominada taxa de aprendizagem, R é um kernel de reforço, λ é uma constante potencializadora do tempo e Δt é uma variação do tempo. Para ilustrar o comportamento dessa equação na construção da Rede Complexa, considere a Figura 29. Nessa figura, pode-se observar que houve uma reponderação na aresta $\overline{ij}$ para um valor maior. Dois fatores contribuíram para esse incremento. O primeiro deles é o valor baixo de Δt e o segundo é o alto valor de R. Isso significa que, quanto menor for a diferença do tempo entre as reponderações, maior será $w_{i,j}$. O kernel de

reforço R é uma medida que representa o grau de relacionamento entre duas regiões que geram a co-ocorrência. Nesse exemplo, consideramos $R = 1$ sempre que duas regiões co-ocorrem. Contudo, pode-se atribuir ao R o inverso da distância entre os centróides das regiões envolvidas na co-ocorrência. Isso significa que, quanto mais próximas duas regiões co-ocorrerem em uma mesma imagem, maior será o valor de R.

Apesar desta Tese prever um sistema de realimentação inspirado na plasticidade biológica, ela não será implementada nos experimentos, uma vez que não faz parte do escopo deste trabalho.

3.2.4 Sub-Modelo de Inferência

O Sub-Modelo de Inferência é o componente responsável por fazer a análise de novas imagens. Esse modelo tem duas ligações com o Sub-Modelo Central. A primeira ligação é com a Rede Complexa e a segunda é com a re-alimentação para o aprendizado infinito.

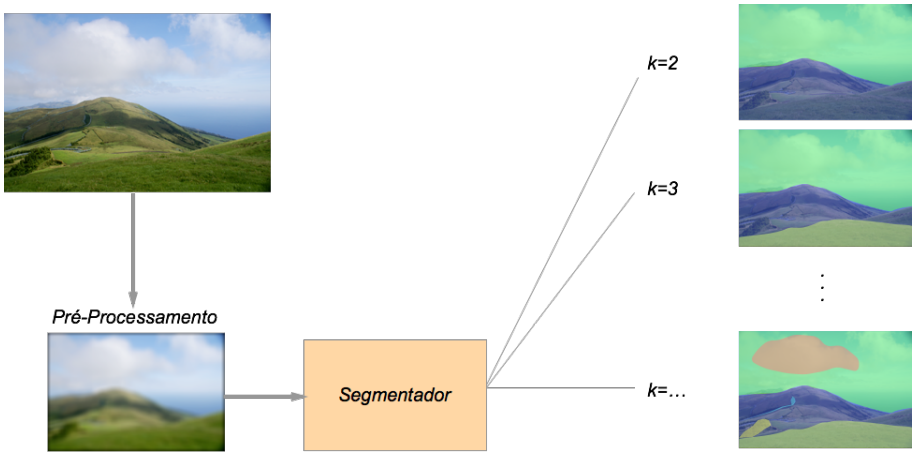
Inicialmente, esse sub-modelo pode receber uma imagem não-supervisionada diretamente em um segmentador. Esse segmentador gera um conjunto de segmentações e o envia para o extrator de características (módulo semelhante ao descrito na Seção 3.2.2.2). Após esta etapa, os conjuntos de instâncias de características de cada segmentação serão analisados pelo maximizador de função. O maximizador de função contém um meta-modelo para escolher a melhor segmentação utilizando como base a Rede Complexa. A segmentação que for escolhida terá suas regiões rotuladas pelos nós correspondentes a cada instância de característica. Finalmente, após a classificação e rotulagem das regiões, esse resultado é apresentado ao construtor da rede para efetuar ajustes no modelo. As próximas seções detalham o processo descrito acima.

3.2.4.1 Segmentador

O segmentador é o módulo responsável por gerar um conjunto de regiões para enviá-lo ao maximizador de função. Para essa tarefa, esse módulo é composto por uma etapa de pré-processamento e outra etapa de segmentação. A etapa de pré-processamento deve ser capaz de fazer um tratamento na imagem para eliminar, por exemplo, grande parte dos ruídos e artefatos que possam interferir negativamente na segmentação.

O modelo proposto por este trabalho prevê o uso de um segmentador capaz de gerar um conjunto de segmentações, dada uma imagem de entrada. O ideal que poderia se esperar deste módulo é que pelo menos uma segmentação se aproxime de uma feita por humanos. Essa

Figura 30 – Exemplo de segmentações a partir de diferentes valores de k



Fonte: Autor

abordagem é válida, uma vez que diversos segmentadores apresentam parâmetros de ajuste para melhorar a segmentação.

Por exemplo, considere o k-means como um algoritmo de segmentação. A quantidade de regiões segmentadas será o valor de k fornecido ao algoritmo. O módulo proposto aqui poderá utilizar este parâmetro para gerar diversas segmentações, variando k, e enviá-las para o próximo módulo. A Figura 30 ilustra algumas segmentações feitas utilizando o k-means.

Outros segmentadores podem ser utilizados nesse módulo. Podemos dividi-los em dois subgrupos: inspirados em modelos biológicos e baseado em análise estatística. Exemplos do primeiro tipo estratégia são: Firefly, Algoritmos Genéricos, Redes Neurais, Deep Learning, entre outros. Como exemplo do segundo tipo de estratégia, pode-se citar: Watershed, Modelos Deformáveis, Level-Set, k-means.

3.2.4.2 Extrator e Discretizador de Características do Sub-modelo de Inferência

O módulo Extrator e Discretizador de Características do Sub-Modelo de Inferência são semelhantes ao Extrator e Discretizador de Características do Sub-Modelo de Treinamento descritos nas Seções 3.2.2.2 e 3.2.2.3. Contudo, neste módulo, um conjunto de segmentações S é apresentado na entrada, ao invés de uma única segmentação supervisionada. Para cada segmentação s_l, um conjunto de instâncias de características é extraído e discretizado utilizando as informações das regiões segmentadas. O resultado desse processo é enviado ao Maximizador de Função. Conforme explicado na Seção 3.2.3.2, o conjunto de instâncias de características é dado por $F = \{i_{m,k}\}$. Nesse módulo, uma segmentação s_l terá um conjunto de instâncias de

características F_l definida por:

$$F_l = \{i_{m,k} : \forall m,k\} \tag{28}$$

onde m é o índice da característica, $i_{m,k}$ é a instância de característica da característica m na região k da segmentação s_l. O resultado final desse módulo será o conjunto composto por todas as instâncias de características de todas as segmentações, dada por:

$$F = \{F_l : \forall l\} \tag{29}$$

3.2.4.3 Maximizador de Função

Das diversas segmentações geradas na etapa de segmentação, uma delas estará mais próxima da segmentação padrão-ouro. O módulo responsável por escolher a melhor segmentação é o Maximizador de Função. Como o próprio nome diz, o maximizador de função encontra uma entrada s_{opt} para uma função f de tal forma que $f(s_{opt}) \geq f(s_i)\ \forall i$. Assim, $s_{opt} = \arg\max_{s_k} f(s_k)$.

Nesse trabalho, a função f utiliza a Rede Complexa do Sub-Modelo Central para dar uma "nota" a partir das instâncias de características F_l. Diversas funções podem ser estudadas, porém, nesse trabalho, focaremos na função ou-exclusivo, proposta inicialmente por RIBEIRO e MUNTZ (1996) para recuperação de informação textual e depois adaptada por Rodrigues, Giraldi e Araujo (2005), para recuperação de imagem com base no conteúdo.

$$f(s) = 1 - \prod_{(a,b)\in\chi\times\chi} \left(1 - \frac{1}{t}\sum_{(k,l)} w(i_{a,k},i_{b,l})\right) \tag{30}$$

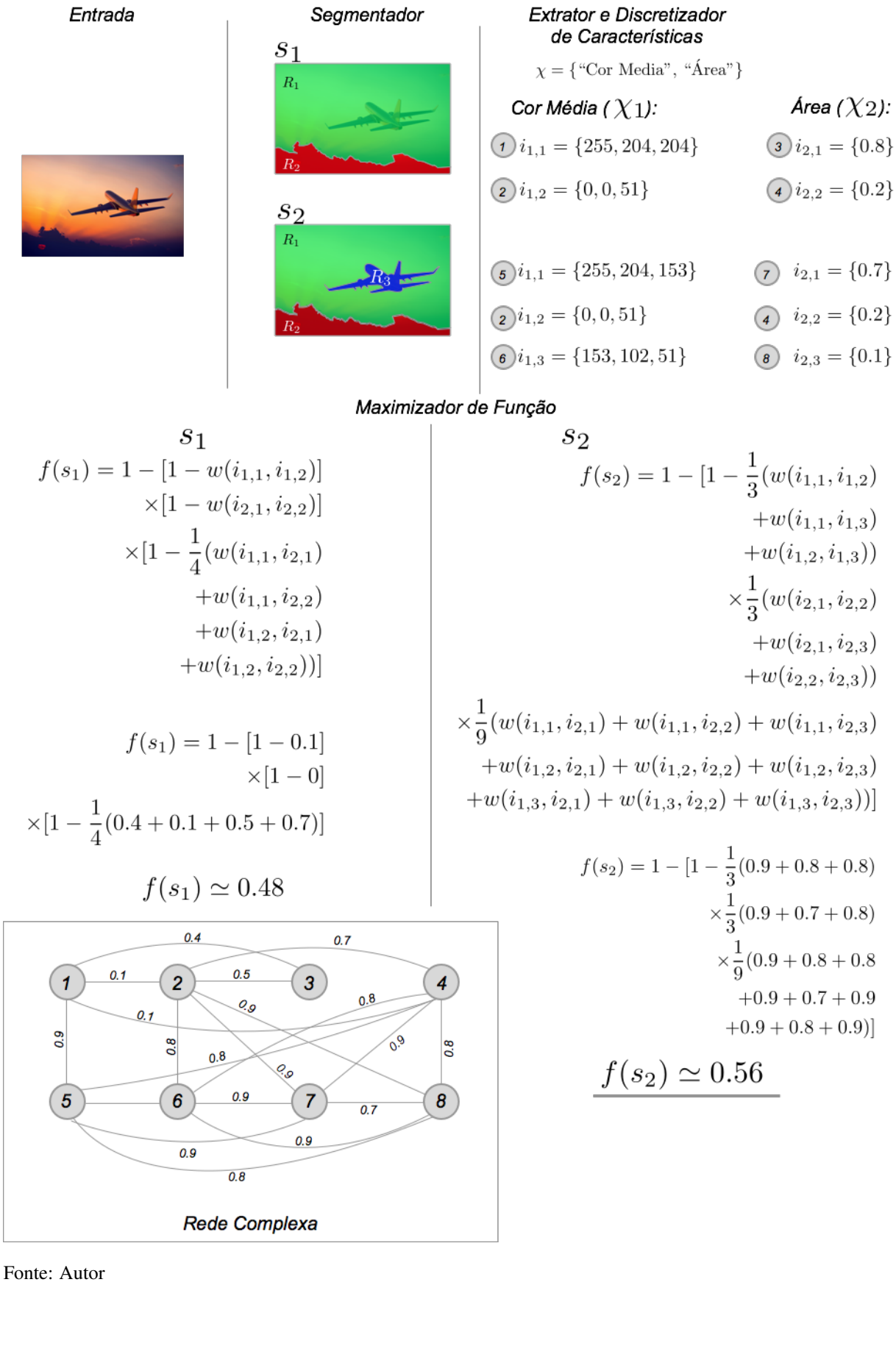
Na Equação (30), χ é o conjunto de características (pág. 100), t é um fator para normalizar o somatório entre 0 e 1, $i_{a,k}$ é a instância da característica a na região R_k da segmentação s e $i_{b,l}$ é a instância da característica b na região R_l da segmentação s, $w(i_{a,k}, i_{b,l})$ é o peso da aresta que conecta o nó que representa a instância $i_{a,k}$ ao nó que representa a instância $i_{b,l}$ na Rede Complexa.

O comportamento dessa equação é o mesmo observado na Seção 2.3 (pág. 74); isto é, se um termo do produtório resultar em 1, a função f será 1. A Figura 31 ilustra o processo de escolha da melhor segmentação segundo a equação proposta. Nessa figura, duas segmentações, s_1 e s_2, são ilustradas como exemplo para avaliação. Uma rede treinada, parte integrante do Sub-Modelo central, é apresentada no canto inferior esquerdo. A primeira etapa do processo é a detecção dos nós na rede a partir das características extraídas de cada segmentação. Assim,

os nós 1,2,3,4 foram identificados a partir das regiões da segmentação s_1 e os nós 2,4,5,6,7 e 8 a partir de s_2. Finalmente, cada segmentação é avaliada utilizando a Equação (30).

Após identificar a melhor segmentação, os rótulos das regiões são preenchidos a partir dos rótulos que estão armazenados nos nós que foram utilizados para fazer a maximização. O resultado desse processo é uma imagem com as regiões delimitadas e com rótulos.

Figura 31 – Ilustração do processo de escolha da melhor segmentação. Segundo a Equação (30), a melhor segmentação é s_2.



Fonte: Autor

4 EXPERIMENTOS E DISCUSSÕES

O meta-modelo de co-ocorrências proposto por este trabalho é uma tentativa de representar em um único modelo outras informações que são pouco exploradas na literatura para o reconhecimento de objetos em cena. A estratégia aqui é que esse modelo seja capaz de armazenar informações relacionadas ao contexto de uma cena; isto é, objetos de uma mesmo contexto possuem maior probabilidade de aparecer em uma mesma cena.

Assim, a hipótese central desta tese, conforme apresentada na Seção 3.2, é que um modelo de co-ocorrência entre as instâncias de características está entre os fatores que mais podem melhorar a qualidade da classificação de um sistema de reconhecimento de objetos. Como já definido, a co-ocorrência são todos os pares de instâncias de características ou rótulos de objetos diferentes que aparecem em uma mesma cena.

Neste capítulo, serão investigadas algumas características que podem indicar a presença de contextos no meta-modelo proposto implementado na base SUN que dão suporte à hipótese central. Contudo, é importante destacar que esses experimentos são empíricos e conduzidos sobre uma base de imagens específica, no entanto, muito conhecida no meio acadêmico-científico.

Dessa forma, os experimentos foram organizados tendo em vista duas linhas de investigação. A primeira linha está relacionada com os aspectos da Neurociência que serviram de inspiração para o modelo proposto, tais como: visão bottom-up/top-down e competição/colaboração entre os mecanismos sensoriais.

A outra linha de investigação é baseada nas características físicas que serão extraídas do modelo, tais como: distribuição de graus, coeficiente de clusterização, raio, diâmetro, entre outras. Essas características podem ajudar a revelar comportamentos do modelo que sustentam a confirmação da hipótese central.

Tabela 3 – Tabela de experimentos e principais linhas de investigação

		Linha de Investigação no Modelo	
Experimento	Seção (Pág.)	Aspectos da Neurociência	Características Físicas
Grau de discretização das instâncias de características	4.1 (116)	-	Interferência das configurações de características para discriminalidade de objetos.
Grau de discretização das instâncias de características utilizando um sistema de reconhecimento de objetos	4.2 (121)	-	Interferência das configurações de características para discriminalidade de objetos.
Estudo da rede de co-ocorrência dos rótulos supervisionados	4.3 (124)	Análise sobre a formação de contextos entre objetos na base de dados estudada. Análise de processos top-down	Análise da organização topológica de uma rede complexa.
Estudo da rede de co-ocorrências dos rótulos e características extraídas das regiões	4.6 (138)	Análise de processos top-down e bottom-Up	Extração de características físicas de uma rede complexa para verificar características do sistema.
Identificação de objetos a partir do conjunto de co-ocorrências	4.4 (129)	Verificação sobre indícios de formação de contextos. Análise de processos top-down	-
Avaliação da segmentação utilizando a rede de co-ocorrências	4.5 (133)	Verificação dos mecanismos de colaboração/cooperação entre mecanismos sensoriais. Análise de processos bottom-up e atenção precoce	-

Tendo em vista as duas linhas de investigação, é apresentado na Tabela 3 um plano de experimentos que serão conduzidos nas próximas seções.

4.1 GRAU DE DISCRETIZAÇÃO DAS INSTÂNCIAS DE CARACTERÍSTICAS

Observando o Modelo Geral na Figura 23, o Sub-Modelo de Treinamento contém dois principais aspectos que podem interferir nos Sub-Modelos seguintes. Um deles é a escolha das características que devem ser consideradas para a construção da rede. O segundo aspecto está relacionado com o grau de discretização dessas características, que afeta diretamente a quantidade de nós da rede, bem como a densidade de arestas (veja Seção 3.2.2.3 para uma descrição mais detalhada).

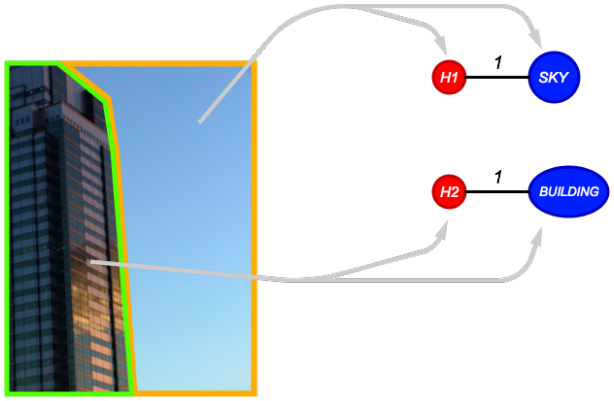
Para avaliar o quanto o grau de discretização de uma característica afeta um sistema de reconhecimento de objetos, propomos um experimento baseado na criação de um grafo bipartido. A ideia aqui é testar o quanto a discretização das características envolvidas afeta a eficiência do modelo geral proposto. Podemos questionar, por exemplo, o grau de modularização da rede se o histograma HSV for muito ou pouco discretizado; o mesmo vale para área das regiões dos objetos ou a sua orientação. Assim, estudamos a discretização de três características: histograma HSV, Área e Orientação (Veja Seção 2.1.2).

Para estudar o impacto dessas discretizações, variamos, para cada característica, a discretização de cada componente dos vetores de características e construímos o grafo bipartido proposto. Em seguida, para cada grafo gerado, medimos a eficiência do modelo com relação à topologia do grafo.

De maneira mais específica, a construção do grafo bipartido é feita considerando as regiões supervisionadas de cada imagem. Sempre que dois rótulos ocorrem em uma mesma imagem, é criada uma aresta entre esses nós que representam os rótulos dessas regiões. A Figura 32 ilustra um exemplo de construção. Nessa figura, nota-se que existem dois tipos de nós: nós-característica (vermelhos); e nós-rótulo (azuis). Sempre que um nó-rótulo r_i ocorre com um nó-característica h_j em uma mesma região, é criada uma aresta $r_i h_j$. Se a aresta $r_i h_j$ já existir no grafo que está sendo construído, e o par instância-rótulo (r_i,h_j) for observado novamente, adiciona-se o valor 1 ao peso da aresta $r_i h_j$.

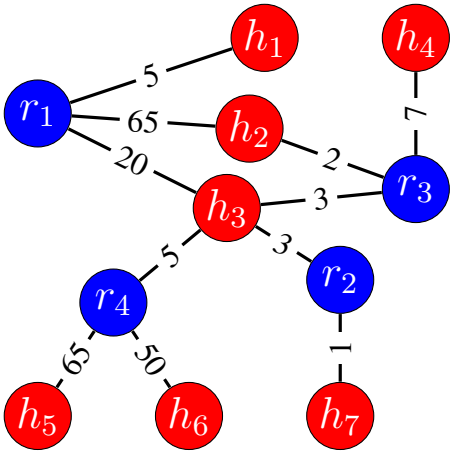
O processo de criação do grafo é continuado para todas as imagens I da base SUN e todas as regiões $r_i \in I$. Dessa forma, o grafo gerado representa as ocorrências entre nós-características e nós-rótulos de todas as regiões das imagens.

Figura 32 – Criação do grafo bipartido a partir de uma imagem supervisionada



Fonte: Autor

Figura 33 – Exemplo de grafo bipartido gerado a partir da análise de algumas imagens.

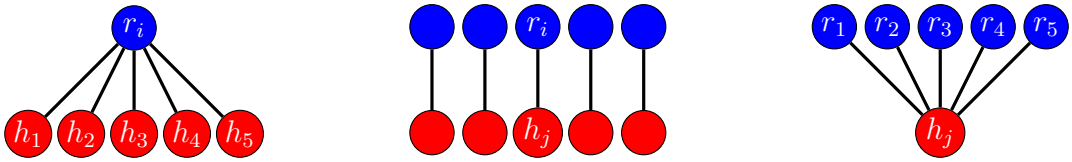


Fonte: Autor

A Figura 33 ilustra um exemplo de grafo construído após a análise de poucas imagens. Nesse grafo, nota-se a presença de quatro rótulos e sete instâncias de características. No domínio de reconhecimento de objetos, as instâncias de características são utilizadas como "pistas" para identificação de um objeto. Fazendo a analogia dessa tarefa utilizando o grafo gerado, caso uma instância de característica h_j seja encontrada, para saber quais rótulos co-ocorrem mais, consulta-se quais os vizinhos (nós-rótulos) r_i que se conectam com h_j com maiores pesos.

Por exemplo, se a característica h_4 for encontrada em uma região de uma imagem, o rótulo sugerido deveria ser o r_3, uma vez que é o único rótulo conectado com h_4. Contudo, poderão existir alguns casos onde um mesmo nó-característica ocorre com diferentes nós-rótulos como, por exemplo, o nó h_3. Assim, pode-se especular que, para um nó-característica específico h_j e um nó-rótulo específico r_i, há três formas extremas de conexão. Essas conexões são mostradas na Figura 34.

Figura 34 – Formas de conexão entre um nó-rótulo h_j e um nó-característica r_i.



Fonte: Autor

A Figura 34(a) ilustra um nó-rótulo conectado com muitos nós-características. Em um sistema de reconhecimento de objetos, esse comportamento de conexão sugere que, se qualquer instância $h_{1...5}$ for observada em uma região, o rótulo sugerido deverá ser r_i. Neste caso, as instâncias $h_{1...5}$ são discriminantes para r_i.

Na Figura 34(b), há uma relação de 1 para 1 entre r_i e h_j. Nesse caso, h_j é a única instância que descobre r_i. Portanto, h_j é discriminante para r_i.

Finalmente, a terceira forma de conexão é mostrada na Figura 34(c), onde uma instância h_j se conecta a vários nós-rotulo $r_{1...5}$. Isso significa que, se h_j for observado em uma imagem, qualquer rótulo em $r_{1...5}$ pode ser escolhido. Assim, h_j não é discriminante para esses rótulos.

Considerando esses três tipos de conexões, propomos uma métrica de vértice para estimar o quanto uma característica é discriminante para um rótulo r_i. A métrica proposta é apresentada na Equação (31).

$$g(r_i) = \frac{\sum\limits_{h_j \in N(r_i)} deg(h_j)}{deg(r_i)} \tag{31}$$

Nesta equação, $g(r_i) \geq 1$, r_i é um nó-rótulo, $N(i)$ é o conjunto de nós ligados ao nó r_i, $deg(h_j)$ é o grau de h_j, dado pela Equação $deg(h_j) = \sum_{r_i} w(r_i,h_j)$, onde $w(r_i, h_j)$ é o peso da aresta $r_i h_j$.

A equação proposta relaciona o somatório dos graus de todos os vizinhos de um rótulo r_i pelo próprio grau de r_i. Assim, considerando as Figuras 34(a) e 34(b), $g(r_i)$ será igual a um. Nesse caso o nó-rótulo r_i apresenta o maior grau de discretização.

Por outro lado, se $g(r_i) \gg 1$, isso significa que um mesmo nó-característica h_j, vizinho ao nó r_i, se conecta com outros nós-rótulo, assim como ilustrado na Figura 34(c). Nesse caso, h_j é pouco discriminante para r_i.

Considerando as medidas locais dos nós-rótulos, estendemos esta métrica para uma métrica global do grafo. Esta métrica considera a área sg sob a curva gerada pela função g, onde

Tabela 4 – Configurações das discretizações de HSV, Área e Orientação

Nº.	H	S	V	F
1	5	3	3	4
2	10	3	3	4
3	15	3	3	4
4	18	3	3	4
5	20	3	3	4
6	25	3	3	4
7	30	3	3	4

(a) HSV

Nº.	Área
8	4
9	8
10	16
11	32
12	64
13	128
14	256
15	512
16	1024
17	2048
18	4096
19	8192

(b) Área

Nº.	Orientação
20	4
21	8
22	16
23	32
24	64
25	128
26	256
27	512
28	1024
29	2048
30	4096
31	8912

(c) Orientação

r_i é um nó-rótulo. Portanto, sg é expressa através da Equação (32).

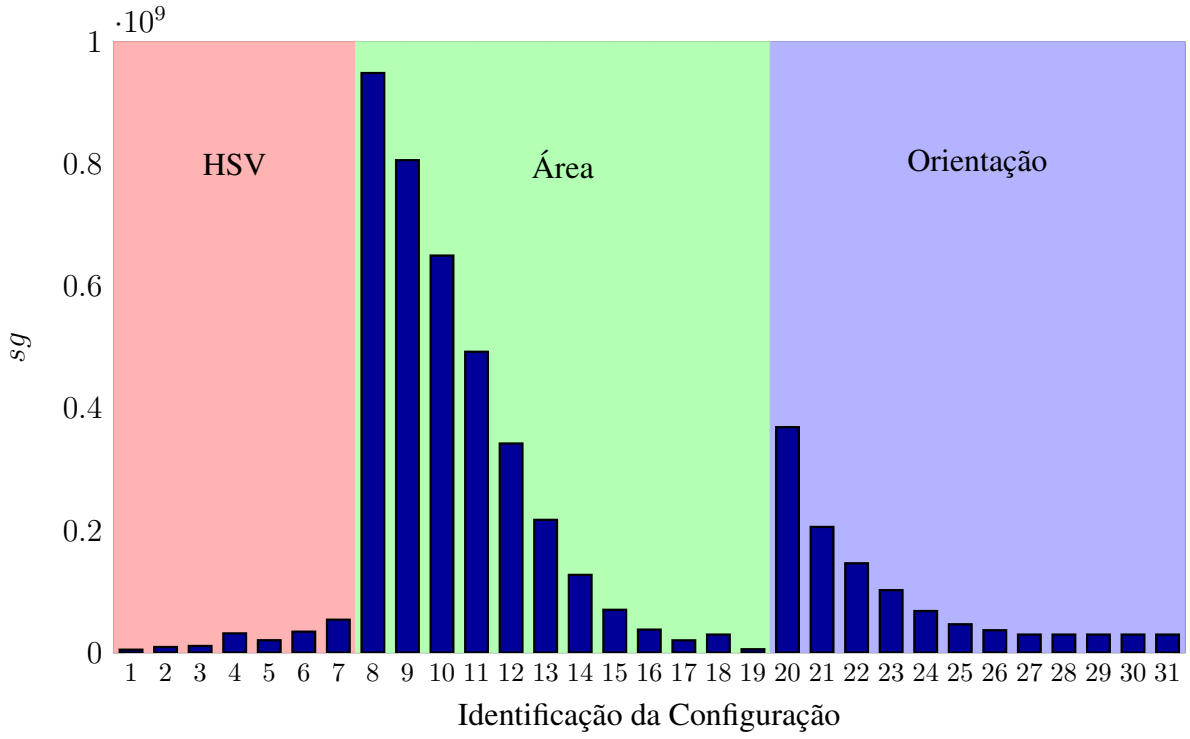
$$sg = \sum_{r_i} g(r_i) \tag{32}$$

Nessa equação, quanto menor for o valor de sg, menor será o somatório de $g(r_i)$ para todo i. Isso significa que a característica em estudo apresenta alto grau de discriminação. Contudo, caso $g(r_i)$ for um valor alto para todo i, sg terá um valor também alto, indicando que a característica em estudo não é discriminante.

Através desta metodologia, foram construídos um total de 31 grafos bipartidos, utilizando diferentes discretizações para características distintas. Cada uma dessas configurações são mostradas na Tabela 4. Esta tabela é composta por 3 sub-tabelas. A tabela mais à esquerda apresenta os parâmetros das 7 discretizações utilizadas para o Histograma HSV (H: Hue, S: Saturation, V: Value, Y: eixo y do histograma). A segunda sub-tabela apresenta o parâmetro das 12 discretizações utilizadas para a característica Área. Finalmente, a última sub-tabela apresenta o parâmetro das 12 discretizações utilizadas para a característica Orientação.

Para cada um dos 31 grafos construídos, foi aplicada a Equação (31). Assim, é esperado que se $g(r_i)$ for próximo a 1, a característica seja discriminante para o nó rótulo r_i. Para se obter uma métrica global do grafo (considerando todos os nós), utilizamos a área abaixo da curva, segundo a Equação (32). A Figura 35 mostra os resultados obtidos para cada uma das 31 entradas da Tabela 4. Nesta figura, o eixo x representa cada uma das configurações da Tabela 5 e o eixo y representa a área abaixo da curva g (Equação (32)).

Figura 35 – Resultados da métrica de qualidade de discriminação para cada configuração (Tabela 4) de construção do grafo.



Fonte: Autor

Fazendo uma análise inicial deste resultado, nota-se que a configuração 1 foi aquela que gerou a menor área sob $g(r_i)$. Isso significa que a configuração $H = 5$, $S = 3$, $V = 3$ e $Y = 4$ foi a melhor discretização encontrada para a base estudada, uma vez que, quanto menor sg, melhor a discretização em termos de discriminância.

De acordo com BRAMÃO et al. (2011), um dos principais aspectos estudados com relação à cor é a sua diagnosticidade. Diagnosticidade é a capacidade de um objeto poder ser discriminado por uma cor específica. Por exemplo, é consenso geral que a cor vermelha possui alta diagnosticidade para o objeto morango. Por outro lado, uma cadeira pode ter qualquer cor, que continuará a ser uma cadeira. Portanto, a cor possui baixa diagnosticidade para o objeto cadeira.

No caso específico da base de dados SUN, muitas das cenas são de ambientes externos e com grande variedade de objetos distintos. Contudo, pode-se especular que, se um mesmo objeto aparece com cores diferentes entre as cenas da base, há uma tendência de uma mesma cor representar mais do que um único objeto. De acordo com esses resultados, o aumento da discretização do componente H aumentou a quantidade de valores possíveis desse componente. Esse aumento significa, na prática, uma maior especificidade da matiz e maior número de nós-

instâncias no grafo. Contudo, a piora no comportamento do experimento pode indicar que os nós-instâncias estão conectados a, pelo menos, mais que um nó-rótulo.

Contudo, considerando os picos de intesidade nos gráficos da Figura 35, nota-se que o HSV obteve melhores resultados, seguido pela orientação e, por último, a área. Assim, para a base SUN, o uso de características baseadas em cores é a que possui maior discriminância para os objetos dentro das outras características estudadas. Um dos pontos comentado em BRAMÃO et al. (2011) é que as pesquisas relacionadas à importância da cor para discriminar objetos utilizam dois tipos de classes de objetos: diagnosticável por cor e não-diagnosticável por cor.

O problema da classificação de objetos entre diagnosticável e não-diagnosticável ainda é uma questão em aberto nessa área. Contudo, os resultados dos experimentos para cor apresentados aqui parecem revelar que os objetos da base estudada tratam-se de diagnosticáveis por cor em sua grande maioria, uma vez que o histograma HSV foi a característica que apresentou a menor média entre os experimentos. A elaboração de novas métricas no grafo bipartido gerado podem servir de base para a criação de métodos automáticos para classificação desses tipos de objetos.

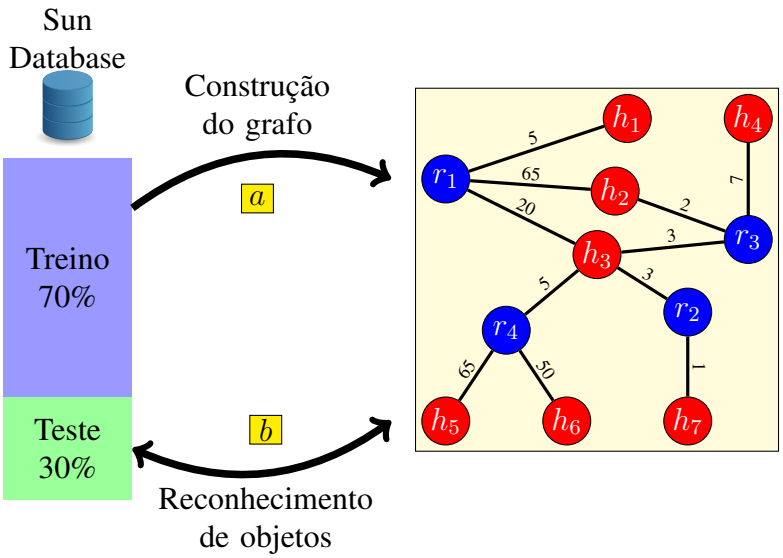
Apesar desse experimento mostrar que a cor é a característica que apresentou melhores resultados, ainda não é suficiente para validar a questão do quanto essa característica é discriminante, pois a métrica baseia-se exclusivamente nas hipóteses de conexões mostradas na Figura 34, ou suas combinações. Assim, propomos um segundo experimento, agora totalmente empírico, utilizando um sistema de reconhecimento de objetos tradicional. Esse sistema baseia-se no reconhecimento de um objeto previamente selecionado em uma imagem através da extração de sua característica. Este experimento será visto com maiores detalhes na Seção 4.2.

4.2 GRAU DE DISCRETIZAÇÃO DAS INSTÂNCIAS DE CARACTERÍSTICAS UTILIZANDO UM SISTEMA DE RECONHECIMENTO DE OBJETOS

Conforme explicado na Seção 4.1, a hipótese de discretização das características, baseada na topologia do grafo bipartido, não é suficiente para confirmar o quanto uma configuração de característica é discriminante. Assim, é necessário um experimento empírico capaz de validar os resultados do experimento anterior.

Neste experimento, a base de dados SUN é dividida em duas partes de maneira aleatória. A primeira parte, chamada de Treino, contém 70% das imagens. A segunda parte, chamada de Teste, contém os 30% restante das imagens. Nesta seção, será adotada a notação I para o

Figura 36 – Esquema geral dos experimentos relacionados ao grau de discretização das características.



Fonte: Autor

conjunto de imagens Teste. A Figura 36 apresenta um esquema geral sobre o experimento anterior e o descrito por esta seção.

A primeira etapa deste experimento é a criação do grafo bipartido da mesma forma que foi realizado na Seção 4.1, contudo, utilizando somente os 70% de imagens escolhidas aleatoriamente (marcação a na Figura 36). A segunda parte do experimento utiliza o grafo gerado como base para o reconhecedor de objetos e as imagens da base Teste para testar esse sistema (marcação b na Figura 36).

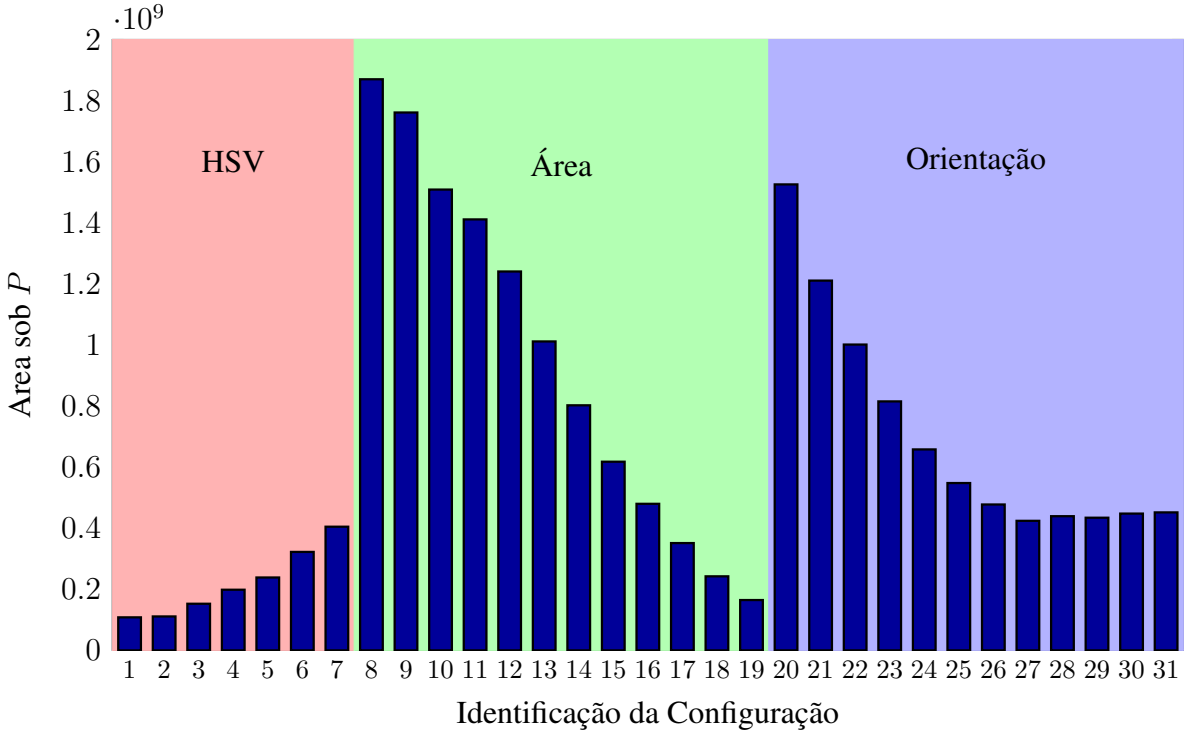
A segunda etapa do experimento é feita da seguinte forma. Considere primeiramente o conjunto de imagens I. Para cada imagem $I_{im} \in I$, uma região k dessa imagem é escolhida aleatoriamente. Então, é extraída a instância de característica de k e buscado o nó-instância h_k que contém os mesmos valores da instância de característica k no grafo estudado (gerado na etapa anterior com uma das configurações da Tabela 4).

A partir de h_k, todos os rótulos $r_i \in N(h_k)$[1] são ordenados de forma decrescente por $w(r_i,h_k)$ formando uma lista L de nós-rótulos ordenados. Finalmente, o rótulo r_k é buscado em L e sua posição é armazenada em uma segunda lista P. Assim, P_{im} irá armazenar a posição onde o rótulo r_k foi encontrado na imagem I_{im}.

Nesse experimento, espera-se que se $P_{im} = 1$, para todo im, então a característica estudada é altamente discriminante para os objetos da base, uma vez que o rótulo correto foi

[1] Primeira vizinhança de h_k

Figura 37 – Resultado do experimento com um sistema de reconhecimento de objetos



Fonte: Autor

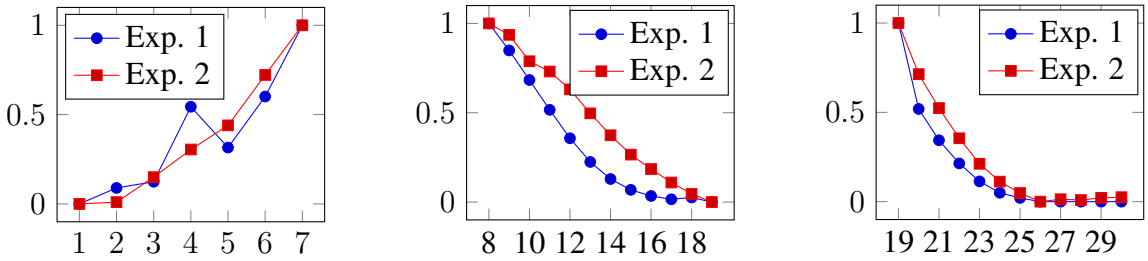
encontrado com maior peso dentre os vizinhos de h_j. Contudo, quanto mais $P_{im} \gg 1$, mais a característica estudada não será discriminante para os objetos da base.

Esse experimento é repetido para todas as configurações de grafos da Tabela 4. A Figura 37 mostra os resultados obtidos para esse sistema de reconhecimento de objetos. No eixo x está o número da configuração do grafo (de acordo com a Tabela 4) e no eixo y está o somatório do vetor de posições P do respectivo grafo.

Com o objetivo de verificar o quanto a métrica proposta na Seção 4.1 consegue descrever o quanto uma característica é discriminante, foi extraída a correlação entre os dois experimentos, considerando cada característica de forma independente. Os gráficos da Figura 38 ilustram essa correlação. Nesta figura, Exp. 1 é a curva gerada pelo experimento da Seção 4.1 e Exp. 2 é a curva gerada pelo experimento desta seção. Note que, em todas as características, a correlação foi no mínimo 94%, indicando que a métrica proposta (Equação (31)) segue o mesmo comportamento de uma máquina de reconhecimento de objetos.

A partir desses resultados, pode-se discutir duas implicações. A primeira delas é que a Equação (31) é uma métrica válida para descrever o quanto uma característica é discriminante. Isso significa que, para uma característica ser discriminante, espera-se que um nó-característica só esteja relacionado com um nó-rótulo (caso ilustrado pela Figura 34(a) e Figura 34(b)). Con-

Figura 38 – Comparação normalizada entre o experimento da Seção 4.1 (em azul) e da Seção 4.2 (em vermelho).



(a) HSV (correlação = 94%)

(b) Área (correlação = 95%)

(c) Orientação (correlação = 97%)

Fonte: Autor

tudo, uma característica não-discriminante utiliza diversos nós-características relacionados com um mesmo nó-rótulo. Esse caso é ilustrado na Figura 34(c).

A segunda implicação é que o resultado do sistema de reconhecimento de objetos proposto é dependente das configurações de características utilizadas. Dessa forma, não é suficiente estudar quais características são discriminantes em uma base de imagens. Deve-se também estudar como as configurações afetam o sistema.

Finalmente, a conclusão geral dos experimentos da Seção 4.1 e 4.2 é que a Cor apresentou os melhores resultados para discriminância de objetos, seguida da orientação e área. Os resultados apresentados aqui nortearam os próximos experimentos desse trabalho. Dessa forma, alguns deles serão estudados variando não só as características (ou combinação entre elas), mas também a configuração de cada uma. Sem perda de generalização, neste trabalho optou-se por estudar somente essas 3 características (Histograma HSV, Área e Orientação).

4.3 ESTUDO DA REDE DE CO-OCORRÊNCIA DOS RÓTULOS SUPERVISIONADOS

Com o objetivo de estudar o comportamento das co-ocorrências entre rótulos da base SUN (Seção 3.1), propomos um experimento para extrair as informações de co-ocorrências entre os rótulos de cada região das imagens dessa base. Nesse experimento, somente os dados supervisionados foram utilizados e nenhum processamento de imagem foi realizado.

A Seção 3.1 apresentou o formato dos arquivos de supervisão da “SUN Database”. Nesse experimento, todos os 16.873 arquivos de supervisão (um para cada imagem) foram utilizados para computar o número de co-ocorrências entre todos os pares de rótulos. O Algoritmo 1 descreve esse processo.

Algoritmo 1 – Algoritmo para contagem de co-ocorrências na base “SUN Database”

Entrada: S - Conjunto de Imagens Supervisionadas
Saída: D - Dicionário Contendo o número de co-ocorrências para cada par de rótulos (r_i, r_j) que ocorrem em uma mesma imagem
$D \leftarrow$ Inicializa Dicionário Vazio
para cada *Imagem* $I \in S$ **faça**
  para cada *Para cada região supervisionada* r_i *na imagem* I **faça**
    para cada *Para cada região supervisionada* r_j *na imagem* I **faça**
      se $r_i \neq r_j$ **então**
        $o_i \leftarrow$ extrai rótulo da região r_i
        $o_j \leftarrow$ extrai rótulo da região r_j
        $c \leftarrow (o_i,o_j)$
        se $c \in D$ **então**
          $D[c] = D[c] + 1$
        senão
          $D[c] = 1$
        fim
      fim
    fim
  fim
fim
retorna D

O Algoritmo 1 recebe como entrada um conjunto S de imagens supervisionadas e um dicionário, D, que associa um par de rótulos ao número de co-ocorrências é inicializado. Para cada imagem I, o algoritmo percorre todos os pares de co-ocorrências de diferentes regiões, extraindo todos os pares de rótulos $(o_i, o_j) \in I$. Para cada par de rótulos, o algoritmo verifica se o par (o_i,o_j) já existe no dicionário de co-ocorrências. Caso não exista, é criada uma entrada em D e atribuído o valor 1. Caso a entrada já exista, seu valor é atualizado através de um incremento.

A saída desse algoritmo será o dicionário D contendo a quantidade de co-ocorrências entre duas regiões previamente rotuladas. Esse dicionário foi ordenado de forma decrescente por número de co-ocorrências. As 30 co-ocorrências mais altas são mostradas na Tabela 5. Nessa tabela, nota-se que muitos rótulos co-ocorrem com os mesmos rótulos. É importante destacar que, apesar dos rótulos serem os mesmos, os objetos são diferentes; isto é, existe mais do que um objeto de mesmo rótulo na cena. Esse caso pode ser observado na Figura 39, onde há diversas regiões objetos diferentes com o mesmo rótulo “window”.

Nessa tabela, pode-se notar de maneira intuitiva que os rótulos que co-ocorrem têm uma relação semântica (de mesmo contexto) entre si, tais como: “floor” e “wall”, “ceiling” e “wall”,

Tabela 5 – As 30 co-ocorrências mais altas na base “SUN Database”.

Rótulo 1	Rótulo 2	Número de co-ocorrências
window	window	168680
chair_occluded	chair_occluded	38914
books	books	31040
ceiling_lamp	wall	22784
floor	wall	22416
window	building	19503
person_sitting_occluded	person_sitting_occluded	19422
ceiling	wall	18887
window	wall	18248
wall	chair_occluded	15106
box	box	14466
plants	plants	13830
tree	window	13277
ball	ball	11046
wall	cabinet	10332
sky	window	10160
trees	trees	9938
door	window	9786
chair_occluded	ceiling_lamp	9495
cabinet	cabinet	9296
ceiling_lamp	ceiling	8520
magazines	magazines	7912
wall	painting	7621
floor	ceiling_lamp	7320
book	book	7242
tree	trees	7136
wall	door	7128
wall	curtain	7122
person_occluded	wall	6635
picture	wall	6471

“window” e “sky”, entre outros. Este pode ser um indício de que pode haver uma formação natural de clusters (contextos) através das co-ocorrências entre os rótulos.

Este resultado induz a realização de um segundo experimento: analisar a distribuição de graus de uma rede complexa gerada a partir da co-ocorrência dos rótulos. A justificativa por detrás desse experimento é que muitos sistemas que apresentam uma distribuição de graus que seguem a lei de potência (Eq. (14)) apresentam cliques. Um clique é um subconjunto de vértices de tal forma que o subgrafo induzido por eles é completo; isto é, todos os vértices do subgrafo são conectados entre si Newman (2003). Portanto, a constatação desse tipo de distribuição será um indício de que há componentes fortemente conexos na rede complexa, uma característica

Figura 39 – Exemplo de imagem da base contendo diversos objetos diferentes de mesmo rótulo “window”.

Fonte: Xiao et al. (2010)

importante que pode revelar a existência de uma topologia na rede, ao contrário de um sistema aleatório.

A constatação de uma topologia na rede é um indício de há formação de contextos, uma vez que haverão grupos de nós que estarão fortemente conectados entre si. Além disso, essa constatação mostrará que trata-se de um sistema que se afasta de um aleatório. O que se espera aqui é que as informações de co-ocorrência aprendidas na fase de treinamento tenham formado grupos de nós naturalmente.

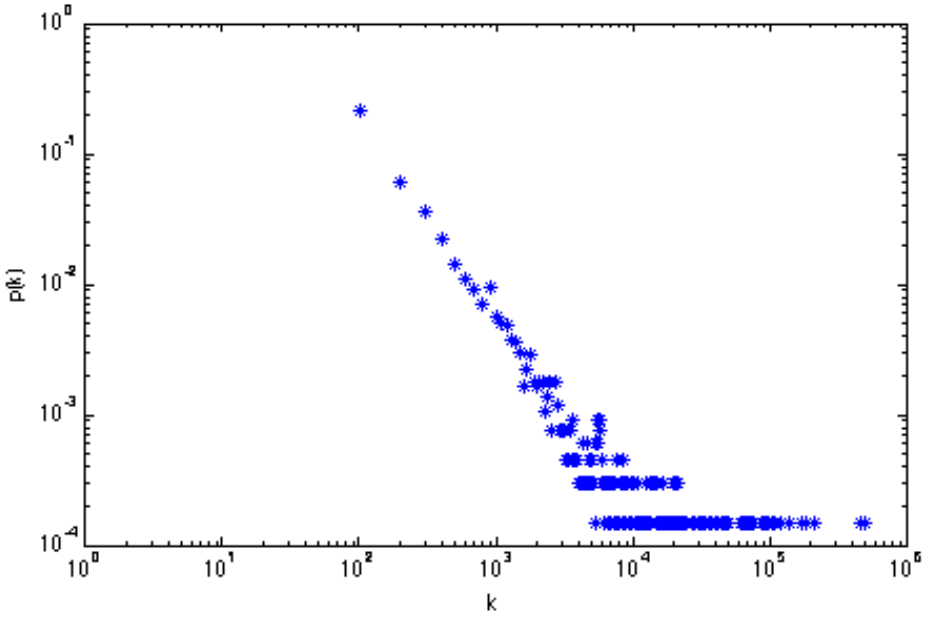
Assim, foi construída uma rede considerando os nós como os rótulos e as arestas como as co-ocorrências entre os rótulos que aparecem em uma mesma imagem. Para este experimento, uma varredura sobre toda a rede foi realizada para gerar uma lista contendo o grau de cada nó (veja Seção 2.5.3.1, pág. 88). Nesse experimento, consideramos que o grau $d(u)$ de cada nó u é o somatório de todos os pesos das arestas que incidem sobre ele, de acordo com a Equação (33):

$$d(u) = \sum_{v} w(u,v) \tag{33}$$

O histograma de graus é apresentado na Figura 40. Nesta figura, o eixo das abscissas representa o grau k dos nós e o eixo das ordenadas representa a probabilidade do grau k ocorrer na rede estudada.

Nesta rede, nota-se que há alta probabilidade para a ocorrência de nós com graus baixos e baixa probabilidade para nós de graus altos. Esse tipo de comportamento é notado em

Figura 40 – Histograma de graus segundo a Equação (33)



Fonte: Autor

redes complexas de mundo-pequeno e livres-de-escala, que apresentam características de agrupamento e cliques.

A partir da distribuição de graus, foi extraído o valor de λ da Equação (14), resultando em $\lambda = 1.59$. Sabe-se que, para $\lambda \gg 0$, o efeito da lei de potência é alto, gerando uma distribuição concentrada à esquerda (maior número de nós com menor grau). Esses tipos de redes normalmente são estudadas do ponto de vista da formação e identificação de comunidades (grupos de vértices altamente conectados entre si Newman, Barabasi e Watts (2006)).

No caso específico desta Tese, as comunidades formadas naturalmente é o que chamamos por contextos; isto é, um grupo de rótulos que aparecem frequentemente em diferentes imagens tendem a se conectar fortemente. Os resultado apresentados aqui são um indício de que os relacionamentos (co-ocorrências) entre os nós (rótulos) são comparáveis com os encontrados na literatura de análise de textos, como em Wachs-Lopes e Rodrigues (2015), reforçando a ideia de que trata-se de um sistema não aleatório.

Os próximos experimentos deste capítulo serão conduzidos para verificar outros aspectos da rede que podem sugerir sua organização topológica.

4.4 IDENTIFICAÇÃO DE OBJETOS A PARTIR DO CONJUNTO DE CO-OCORRÊNCIAS

Um dos principais componentes do modelo geral proposto por este trabalho é o Maximizador de Função (veja Seção 3.2.4.3). Este componente é responsável por integrar as informações extraídas a partir de uma imagem com as informações aprendidas no Sub-Modelo Central (veja Seção 3.2.3). Assim, pode-se utilizar o Maximizador de Função como uma forma de avaliar a rotulação de objetos desconhecidos em uma imagem utilizando as informações aprendidas na etapa de treinamento. Nesse ponto, é importante destacar que essa comunicação entre a rede complexa e as informações da imagem implementa o mecanismo Top-Down da Neurociência; isto é, as informações de alto-nível (contextos) ajudam o sistema a inferir as informações vindas das camadas inferiores.

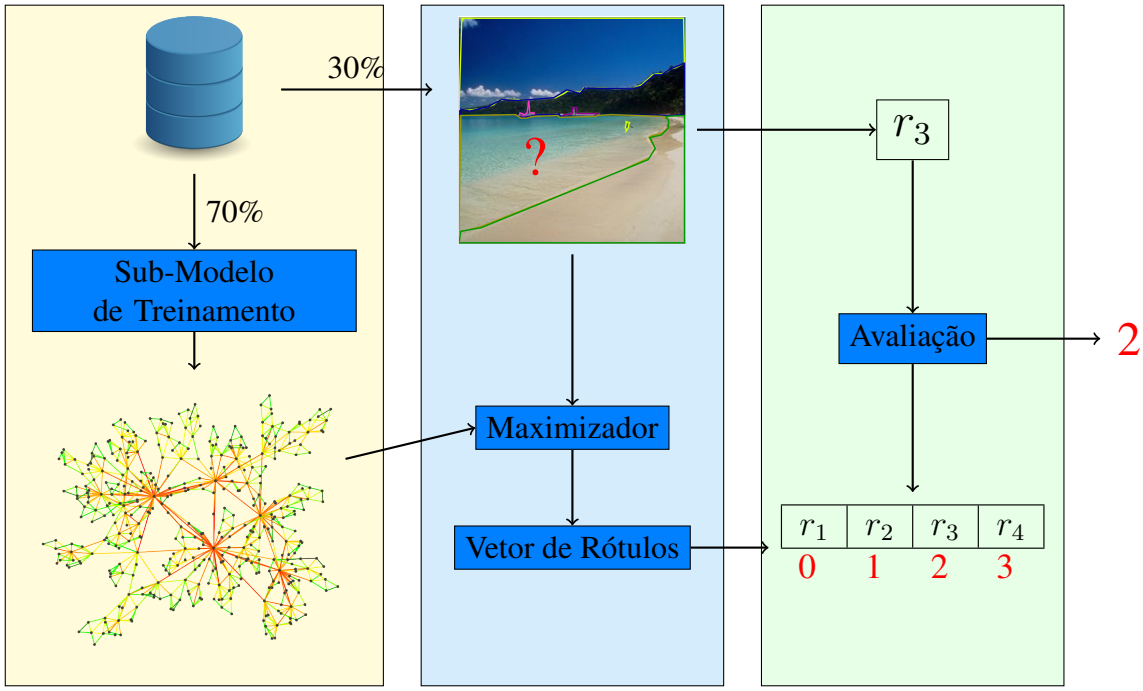
Neste experimento, o principal interesse de estudo é verificar a taxa de reconhecimento de objetos dada, uma imagem segmentada. Para esse estudo, são utilizadas 3.400 redes treinadas considerando diferentes combinações entre características e discretizações. Nesta seção, o conjunto de redes complexas será denotado por C.

Por motivos de clareza, será utilizada uma notação própria para descrever as configurações de diferentes características utilizadas aqui. Para a característica HSV, será utilizada na notação $H_hS_sV_vY_y$, onde h,s,v,y são números inteiros que representam a discretização em cada componente. Por exemplo, a configuração $H_{18}S_5V_3Y_{10}$ será feita para indicar que a discretização do histograma HSV foi feita utilizando 18 valores para o componente H, 5 valores para o componente S, 3 valores para o componente V e 10 valores para o eixo Y do histograma, que representa. De forma semelhante, a característica Área será representada pela notação A_a para indicar que foi discretizada em a valores distintos. A Orientação também seguirá o mesmo padrão: O_o indicará que a orientação foi discretizada em o valores distintos. Além dessas três características, será também considerada como característica os rótulos das regiões das imagens, tendo a notação RT_0 [2].

Além das notações apresentadas anteriormente, a notação $A_{\{1,...,5\}}$ indicará o conjunto de configurações $\{A_1, A_2, A_3, A_4, A_5\}$. Essa notação poderá ser estendida para qualquer outra característica. Por exemplo, $H_{\{5,6,7\}}S_3V_3Y_5$ representa o conjunto $\{H_5S_3V_3Y_5, H_6S_3V_3Y_5, H_7S_3V_3Y_5\}$. Combinações entre as características também poderão ser feitas. Por exemplo, para representar a configuração de uma rede complexa gerada a partir da extração dos rótulos e áreas discretizadas em 5 valores distintos, utiliza-se a notação RT_0A_5.

[2] O número subscrito 0 está presente na notação apenas para enfatizar que a característica Rótulo não apresenta um parâmetro de discretização.

Figura 41 – Etapas do experimento para rotulação de regiões. O quadro à esquerda representa o Sub-Modelo de treinamento;o quadro central representa a etapa de inferência; e o quadro à direita representa a avaliação dos resultados.



Fonte: Autor

Assim, considerando as notações, o conjunto de configurações C utilizado para os experimentos dessa seção, é definido como:

$$
\begin{aligned}
C \;=\; & \{RT_0 H_{\{5,10,15,20,25\}} S_{\{3,6,9\}} V_{\{3,6,9\}} F_{\{3,6,9\}}\} \\
\cup\; & \{RT_0 A_{\{5,10,25,45\}}\} \\
\cup\; & \{RT_0 O_{\{5,10,20,30\}}\} \\
\cup\; & \{RT_0 H_{\{5,10,15,20,25\}} S_{\{3,6,9\}} V_{\{3,6,9\}} F_{\{3,6,9\}} A_{\{5,10,25,45\}}\} \\
\cup\; & \{RT_0 H_{\{5,10,15,20,25\}} S_{\{3,6,9\}} V_{\{3,6,9\}} F_{\{3,6,9\}} O_{\{5,10,20,30\}}\} \\
\cup\; & \{RT_0 A_{\{5,10,25,45\}} O_{\{5,10,20,30\}}\} \\
\cup\; & \{RT_0 H_{\{5,10,15,20,25\}} S_{\{3,6,9\}} V_{\{3,6,9\}} F_{\{3,6,9\}} A_{\{5,10,25,45\}} O_{\{5,10,20,30\}}\} \\
\cup\; & \{RT_0\}
\end{aligned}
$$

A Figura 41 ilustra as etapas do experimento. Nesta figura, o quadro à esquerda representa o sub-modelo de treinamento. Nesta etapa, 70% das imagens supervisionadas da base SUN são utilizadas para a construção de uma Rede Complexa. Além disso, é nesta etapa que são configuradas as características e seus parâmetros de discretização.

O quadro central da figura representa a etapa de inferência. Aqui, são apresentadas 30% das imagens da base incluindo a segmentação humana. Contudo, uma região aleatória (representada por "?" na figura) é escolhida para ter seu rótulo removido. A ideia aqui é que a partir das informações das instâncias de características e rótulos das regiões vizinhas possa-se inferir qual o rótulo da região escondida.

Na etapa de inferência, todos os nós rótulos da Rede Complexa treinada são avaliados como candidatos à rotulação. Essa avaliação é baseada na mesma equação do Maximizador apresentada na Seção 3.2.4.3 (Equação (30)). Para cada imagem I_i analisada, um vetor V_i de rótulos é criado considerando todos os rótulos da rede ordenados de maneira decrescente pelo maximizador.

Por último, a etapa de avaliação irá armazenar a posição onde o rótulo correto da região escolhida de cada imagem I_i foi encontrado no vetor V_i. Essa informação é armazenada em um outro vetor A.

Considerando a proposta experimental apresentada, o vetor A terá então todas as posições onde o rótulo correto foi encontrado no ranque gerado pelo maximizador. Se $A_i = 0$ para todo i, isso significa que todos os rótulos desconhecidos tiveram uma inferência correta pelo maximizador. Contudo, caso $A_i = 6668$ (a quantidade de rótulos distintos da base - 1) terá o pior resultado.

Considerando essas observações, a avaliação final α do desempenho para rotulação de objetos desconhecidos se dará através da seguinte equação:

$$\alpha = 1 - \frac{\sum\limits_{i} A_i}{6668 \times |A|} \tag{34}$$

onde A_i é a posição do rótulo correto no vetor V_i e $|A|$ é o tamanho do vetor A (quantidade de imagens analisadas). Nesta equação, note que α terá seus valores entre 0 e 1. Para que α seja igual a 0, é necessário que o numerador da equação seja o maior possível; isto é, os rótulos corretos de cada região escondida devem estar na última posição do de cada vetor V_i, o pior resultado possível. Contudo, caso o numerador resulte em 0, α será igual a 1. Isso significa que todos os rótulos foram encontrados na primeira posição (índice 0) de cada vetor V_i, o melhor resultado possível.

A Figura 42 ilustra os valores de α para todas as 3.400 redes analisadas. Por motivos de visualização, o eixo x foi ordenado de maneira decrescente pelos valores de y. As configurações das 10 melhores e piores redes, bem como o valor de α e a posição média do rótulo correto, são mostrados na Tabela 6.

Figura 42 – Valor de α para cada Rede Complexa estudada no experimento.

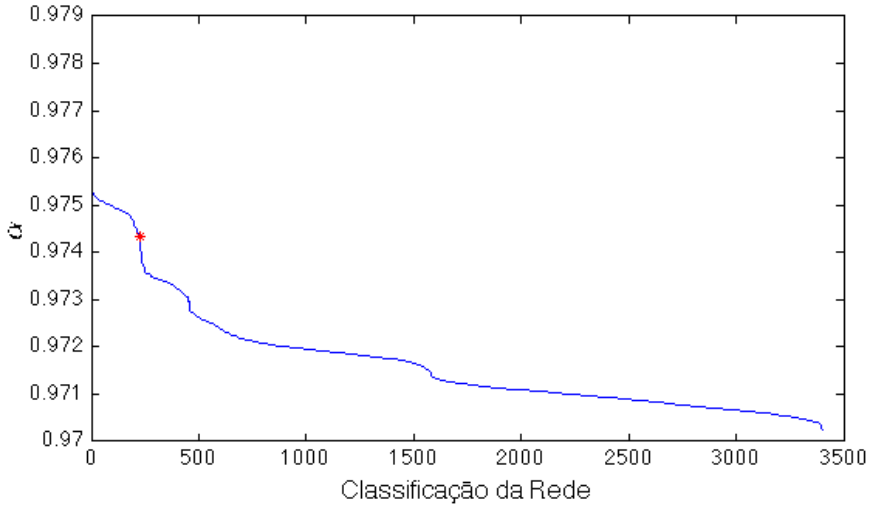


Tabela 6 – Melhores e Piores 10 configurações de Redes Complexas com relação ao valor α

$\alpha.10^2$	Posição Média	Configuração
97.530	164.69	$RT_0H_{10}S_3V_6Y_{10}O_{20}$
97.528	164.79	$RT_0H_5S_6V_6Y_{10}$
97.528	164.79	$RT_0H_{10}S_3V_6Y_{10}$
97.528	164.80	$RT_0H_5S_6V_6Y_{10}O_{20}$
97.528	164.80	$RT_0H_{10}S_3V_6Y_{10}O_{30}$
97.527	164.83	$RT_0H_{10}S_3V_6Y_{10}O_5$
97.527	164.85	$RT_0H_5S_6V_6Y_{10}O_{10}$
97.527	164.88	$RT_0H_5S_6V_6Y_{10}O_{30}$
97.526	164.90	$RT_0H_5S_6V_6Y_{10}O_5$
97.526	164.94	$RT_0H_{10}S_3V_6Y_{10}O_{10}$
⋮	⋮	⋮
97.038	197.50	$RT_0H_{20}S_9V_9Y_3A_{45}O_{10}$
97.036	197.60	$RT_0H_{20}S_9V_9Y_3A_{45}O_{20}$
97.036	197.60	$RT_0H_{20}S_9V_9Y_3A_{45}O_{30}$
97.036	197.63	$RT_0H_{25}S_9V_9Y_3A_{45}$
97.035	197.65	$RT_0H_{15}S_9V_9Y_3A_{45}$
97.033	197.82	$RT_0H_{15}S_9V_9Y_3A_5$
97.033	197.83	$RT_0H_{25}S_9V_9Y_3A_5$
97.031	197.96	$RT_0H_{20}S_9V_9Y_3A_{45}$
97.029	198.09	$RT_0H_{20}S_9V_9Y_3A_5$
97.025	198.31	RT_0A_5

Analisando os resultados, nota-se que $\alpha > 97\%$. Esse primeiro resultado é uma constatação de que a Rede Complexa gerada não é aleatória, do contrário $\alpha \rightarrow 50\%$. Então, isso significa que o rótulo e as características das regiões visíveis pelo algoritmo ajudaram a inferir a região escondida.

Esse comportamento é semelhante ao mecanismo Top-Down abordados pela Neurociência. Nesse caso, as informações cognitivas são representadas pelo modelo de Rede Complexa e o mecanismo Top-Down é representado inferência feita pelo Maximizador.

Outra constatação importante é que, dentre as 3.400 redes analisadas se encontra a rede RT_0 (formada somente por rótulos, uma característica importante para a formação formação de contexto, assim como visto na Seção 4.3). Essa rede serve como uma referência para avaliar o quanto a adição de novos sensores (ou características) colaboram com a inferência. A avaliação $alpha$ desta rede esta marcada com o símbolo vermelho no gráfico da Figura 42, na posição $x = 226$.

Considerando este fato, isso significa que 225 redes construídas (utilizando rótulos e mais alguma(s) configuração(ões) de característica(s)) obtiveram melhores resultados. Pode-se então considerar que o modelo proposto contém uma característica parecida com o processo de competição e colaboração entre mecanismos sensoriais, aspecto também abordado pela Neurociência que serviu de inspiração para o modelo proposto por esta Tese.

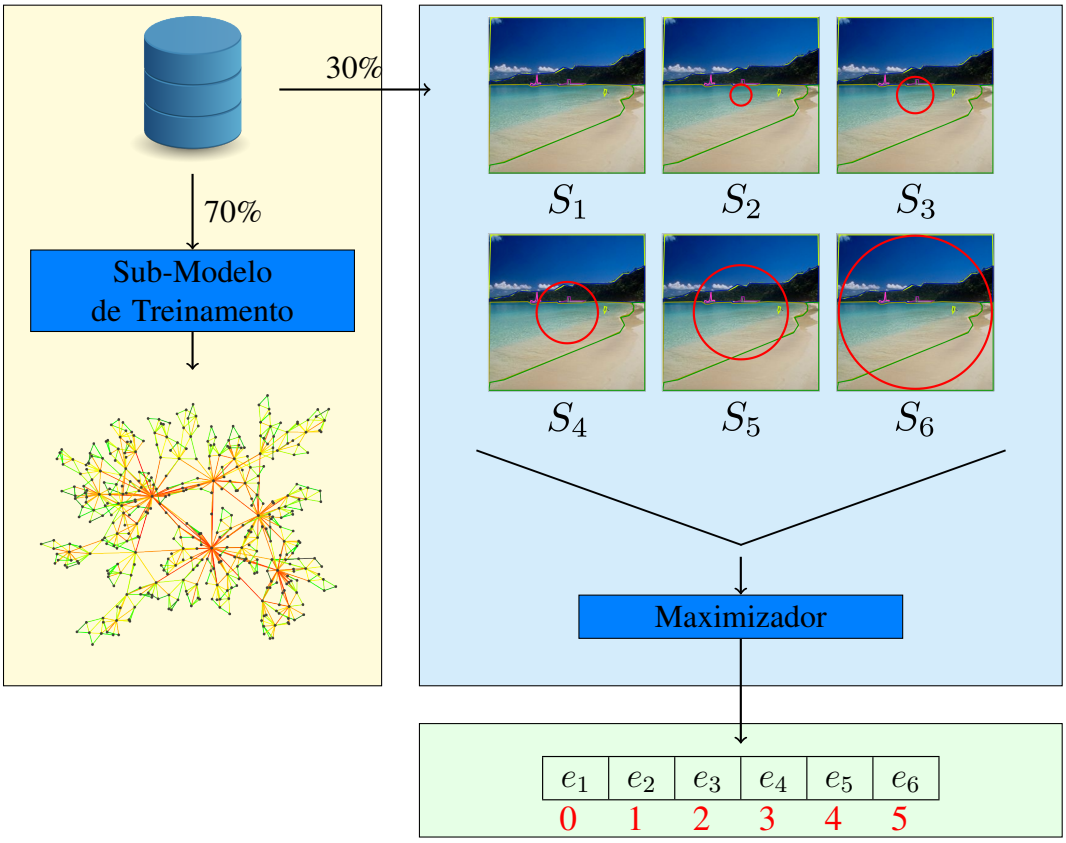
Apesar do desvio padrão entre os valores de α ser baixo (cerca de 0.0012), os resultados das melhores redes estão de acordo com os resultados encontrados no experimento da Seção 4.2, uma vez que a Cor e a Orientação apresentaram os melhores resultados. Por outro lado, a Área obteve o pior desempenho, conforme previsto anteriormente.

4.5 AVALIAÇÃO DA SEGMENTAÇÃO UTILIZANDO A REDE DE CO-OCORRÊNCIAS

Conforme descrito na Seção 3.2.4.3, o maximizador de função deve escolher a melhor segmentação segundo a Equação (30). Contudo, o conjunto de segmentações enviado ao maximizador deve conter, pelo menos, uma segmentação mais próxima ao padrão-ouro. Isso porque o objetivo principal do maximizador é avaliar segmentações e, portanto, não há nenhuma responsabilidade em se fazer de fato a segmentação. Assim, nesse experimento, estamos interessados em investigar a qualidade processo de avaliação da segmentação enviada ao maximizador.

Uma vez que o objetivo deste trabalho não é desenvolver um segmentador, propomos o uso da própria segmentação humana e outras 5 segmentações consideradas inadequadas geradas automaticamente a partir de deformações da original para fazer parte do conjunto de avaliação.

Figura 43 – Etapas do experimento para avaliação de segmentações. O quadro à esquerda representa o Sub-Modelo de treinamento; o quadro à direita representa a etapa avaliação; e o quadro abaixo representa a saída da avaliação.



Fonte: Autor

As deformações adicionadas artificialmente às segmentações são baseadas na criação de uma nova região artificialmente introduzida, delimitada por uma circunferência, centrada na imagem, de raio incremental. A ideia aqui é que, quanto maior o raio, menor deve ser a nota da segmentação, uma vez que será maior a deformação introduzida.

A Figura 43 ilustra as etapas desse experimento. A primeira etapa consiste na construção de diversas Redes Complexas a partir de 70% das imagens da base SUN (assim como foi conduzido no experimento da Seção 4.4). O conjunto C de configurações de Redes Complexas

utilizado neste experimento é dado por:

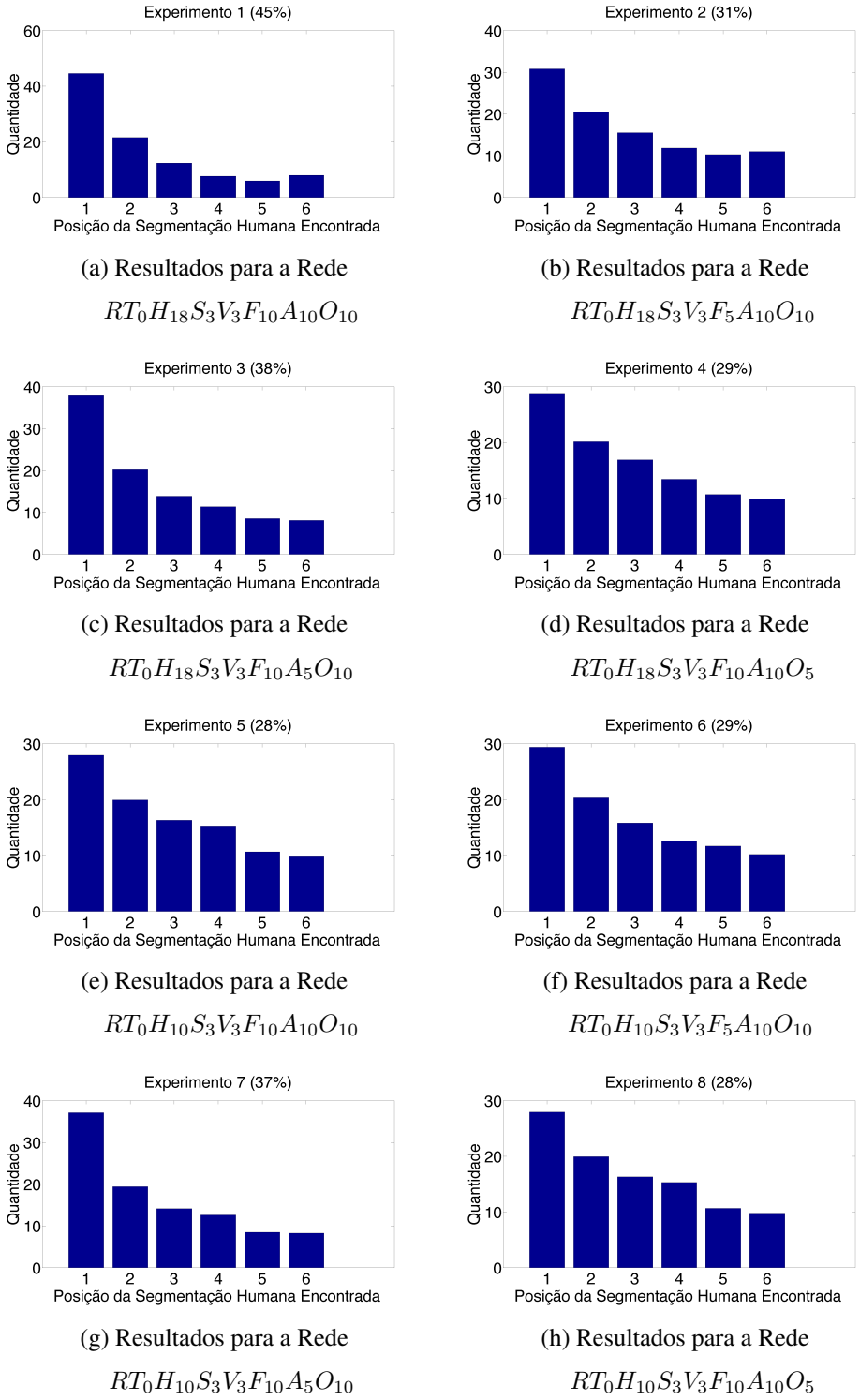
$$
\begin{array}{rcccc}
C & = & \{RT_0H_{18}S_3V_3F_{10}A_{10}O_{10}\} & \cup & \{RT_0H_{18}S_3V_3F_5A_{10}O_{10}\} \\
 & \cup & \{RT_0H_{18}S_3V_3F_{10}A_5O_{10}\} & \cup & \{RT_0H_{18}S_3V_3F_{10}A_{10}O_5\} \\
 & \cup & \{RT_0H_{10}S_3V_3F_{10}A_{10}O_{10}\} & \cup & \{RT_0H_{10}S_3V_3F_5A_{10}O_{10}\} \\
 & \cup & \{RT_0H_{10}S_3V_3F_{10}A_5O_{10}\} & \cup & \{RT_0H_{10}S_3V_3F_{10}A_{10}O_5\} \\
 & \cup & \{RT_0H_{18}S_3V_3F_5\} & \cup & \{RT_0O_{10}\} \\
 & \cup & \{RT_0A_{10}\} & \cup & \{RT_0H_{18}S_3V_3F_{10}\} \\
 & \cup & \{RT_0H_{10}S_3V_3F_{10}\} & \cup & \{RT_0H_{10}S_3V_3F_5\} \\
 & \cup & \{RT_0H_5S_2V_2F_3\} & \cup & \{RT_0H_{18}S_3V_3F_{10}O_{10}\}
\end{array}
$$

A segunda etapa, representada pelo quadro à direita, compreende a etapa de avaliação de segmentações. Para isso, um conjunto de 6 segmentações de uma imagem é apresentado por vez para o maximizador. Nesta figura, S_1 representa a segmentação humana (padrão-ouro) e $S_{2,\ldots,6}$ representam as segmentações geradas automaticamente adicionando uma nova região delimitada por uma circunferência centrada na imagem e de raio cada vez maior. Da mesma forma que no experimento da Seção 4.4, este experimento utiliza a Equação (30) como estratégia para ranquear as segmentações. Para cada imagem I_i do conjunto de imagens de teste, a saída do Maximizador é dada pelo vetor E_i de segmentações, ordenado pela nota da segmentação de maneira decrescente.

O que se espera deste experimento é que a segmentação padrão-ouro esteja localizada na posição 1 de cada vetor E_i. Os resultados iniciais desse experimento são mostrados na Figura 44. Para cada figura é apresentado um histograma com 6 entradas. Cada entrada representa a quantidade de vezes que a segmentação padrão-outro foi encontrada no vetor E_i para todas as imagens de teste.

Inicialmente, os histogramas mostram que as segmentações padrão-ouro foram escolhidas com maior frequência como a melhor. Trata-se de um indício de que o Maximizador apresenta melhor performance quando utiliza informações contextuais. Contudo, pode-se especular que, conforme demonstrados nos experimentos anteriores, a característica Área é a de menor performance para o sistema de avaliação.

Figura 44 – Resultados das primeiras 8 configurações de características do conjunto C



(a) Resultados para a Rede $RT_0H_{18}S_3V_3F_{10}A_{10}O_{10}$

(b) Resultados para a Rede $RT_0H_{18}S_3V_3F_5A_{10}O_{10}$

(c) Resultados para a Rede $RT_0H_{18}S_3V_3F_{10}A_5O_{10}$

(d) Resultados para a Rede $RT_0H_{18}S_3V_3F_{10}A_{10}O_5$

(e) Resultados para a Rede $RT_0H_{10}S_3V_3F_{10}A_{10}O_{10}$

(f) Resultados para a Rede $RT_0H_{10}S_3V_3F_5A_{10}O_{10}$

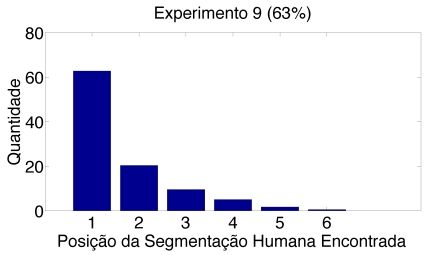
(g) Resultados para a Rede $RT_0H_{10}S_3V_3F_{10}A_5O_{10}$

(h) Resultados para a Rede $RT_0H_{10}S_3V_3F_{10}A_{10}O_5$

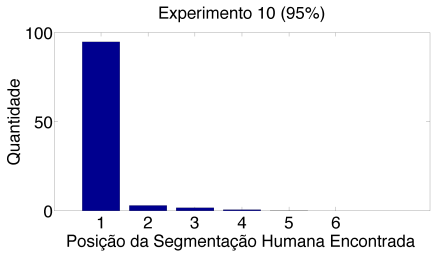
Fonte: Autor

Essa observação é confirmada ao avaliar o histograma da Rede da Figura 43a com o histograma da Rede da Figura 44h. Note que em 77% das classificações da Rede 16, o sistema classificou a segmentação padrão-ouro como a melhor. Contudo, no caso da Rede 1, o único

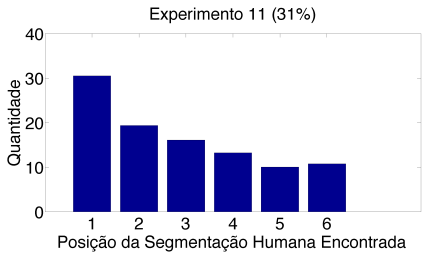
Figura 45 – Resultados das últimas 8 configurações do conjunto C



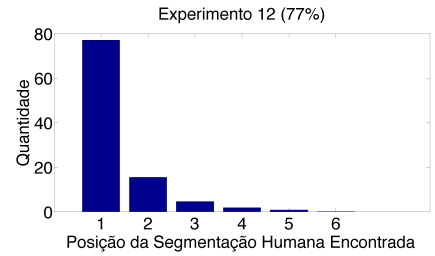
(a) Resultados para a Rede $RT_0H_{18}S_3V_3F_5$



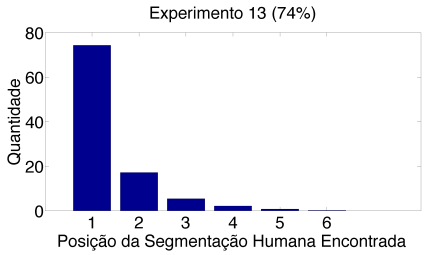
(b) Resultados para a Rede RT_0O_{10}



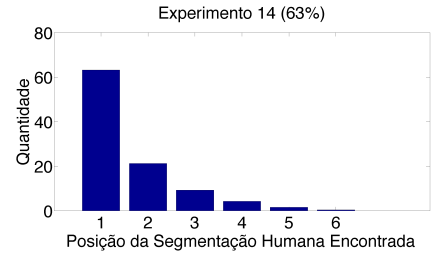
(c) Resultados para a Rede RT_0A_{10}



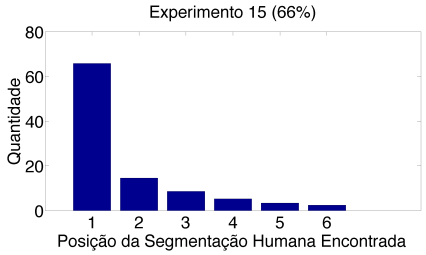
(d) Resultados para a Rede $RT_0H_{18}S_3V_3F_{10}$



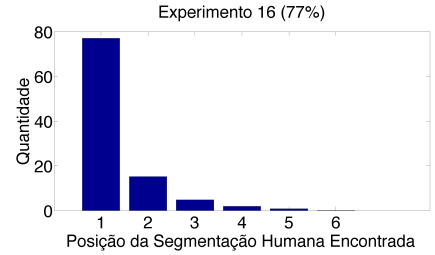
(e) Resultados para a Rede $RT_0H_{10}S_3V_3F_{10}$



(f) Resultados para a Rede $RT_0H_{10}S_3V_3F_5$



(g) Resultados para a Rede $RT_0H_5S_2V_2F_3$



(h) Resultados para a Rede $RT_0H_{18}S_3V_3F_{10}O_{10}$

Fonte: Autor

parâmetro diferente (A_{10}) fez esse resultado cair para 45%. Assim, esse resultado reforça a ideia de qua a característica Área não contribuiu positivamente com a discriminação contextual.

Considerando todos os experimentos de maneira geral, o melhor resultado foi obtido pela Rede da Figura 44b, cuja sua configuração é RT_0O_10. Esse resultado mostra que a informação de Orientação (considerada de forma isolada) foi a que mais avaliou de maneira correta a segmentação padrão-ouro como a melhor. A justificativa para este resultado pode ser feita considerando dois pontos. O primeiro deles está relacionado com a maneira de extração da ca-

racterística Orientação. Note que, como o ruído introduzido na segmentação foi feito na forma de circunferência, quaisquer dois pontos extremos do contorno terá sua distância máxima. A implementação feita nessa Tese escolhe o menor ângulo formado entre o eixo horizontal da imagem e a linha que conecta os pontos extremos do contorno. Assim, a orientação para o ruído sempre será a mesma, o valor 0. O segundo ponto da justificativa está relacionado com as co-ocorrências aprendidas pela Rede Complexa. Pode-se especular que, se a instância de valor 0 da característica Orientação estiver fracamente conectada ou simplesmente não existir na Rede, a nota do Maximizador será baixa.

Aqui, pode-se dizer que a Orientação serviu como uma característica que colaborou com a avaliação de uma segmentação. Este comportamento reforça os achados no experimento da Seção 4.4.

Um ponto importante a se dizer desses resultados é que o classificador proposto pode ser incorporado em algoritmos Meta-Heurísticos (Firefly, Algoritmos Genéricos, Redes Neurais Retro-Alimentadas, Q-Learning) utilizando o mesmo como uma função avaliadora. Assim, é interessante destacar que o modelo proposto por esta Tese prevê o mecanismo Bottom-Up, já discutido na Seção 2.3; isto é, um mecanismo de baixo nível (nesse caso o algoritmo meta-heurístico), gera diversas soluções possíveis que são enviadas para camadas mais altas do modelo (nesse caso a Rede Complexa), produzindo um processo de visão tardia, mais associada, como já foi discutido, a altos processos mentais de análise de objetos em cenas.

Apesar da Rede Complexa conter informações capazes de avaliar o quanto uma segmentação "faz sentido", segundo as co-ocorrências, a inferência nessa rede é algo custoso computacionalmente. Contudo, até mesmo essa característica ocorre com o sistema visual. Por exemplo, uma pessoa que acorda em um lugar não previsto por ela (fora do último contexto) levará um tempo maior para entender a cena ao abrir os olhos. A necessidade por consultar camadas superiores cognitivas é a principal responsável por esse atraso Broadbent (1970).

4.6 ESTUDO DA REDE DE CO-OCORRÊNCIAS DOS RÓTULOS E CARACTERÍSTICAS EXTRAÍDAS DAS REGIÕES

Os experimentos conduzidos nas seções anteriores foram planejados especificamente do ponto de vista do Meta-Modelo proposto por esta Tese. Contudo, uma investigação mais profunda através da teoria de Redes Complexas pode trazer novos indícios sobre o comportamento do sistema. Nesta seção, o principal interesse é a extração das principais características físicas

de redes complexas e a discussão sobre as implicações dos resultados, traçando um paralelo com os aspectos da Neurociência que inspiraram a elaboração desta Tese.

Entre as 3400 redes construídas, aquela escolhida foi a de configuração $RT_0H_5S_3V_3F_6O_5$. Há duas principais razões por esta escolha. A primeira delas está relacionada com seu tamanho (em termos de nós e densidade de arestas), uma vez que a extração da excentricidade é um dos algoritmos mais caros computacionalmente. O segundo motivo é que esta rede foi classificada na posição 365 do gráfico mostrado na Figura 42, relativamente próximo às redes que obtiveram melhores desempenhos em termos do valor de α.

Nestes experimentos, serão estudadas 5 características físicas de redes complexas: coeficiente de clusterização médio, distribuição de graus, distribuição de pesos, raio e diâmetro. Cada característica será abordada individualmente nas próximas seções. As justificativas para a escolha dessas características serão explicadas ao longo de cada experimento.

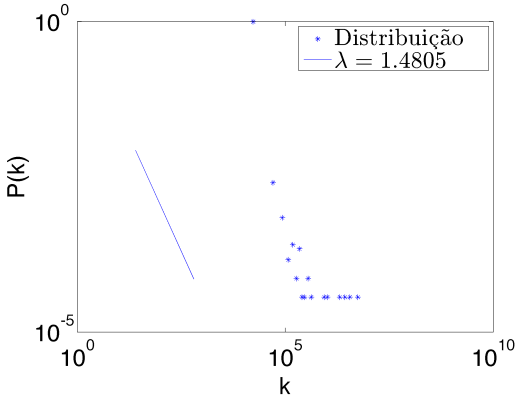
4.6.1 Distribuição de Graus

De acordo com Newman, Barabasi e Watts (2006), a distribuição de graus apresenta alguns padrões quando extraída a partir de determinados tipos de redes. Por exemplo, sabe-se que uma rede aleatória segue uma distribuição binomial e que redes livres de escala seguem uma distribuição com cauda exponencial (lei de potência).

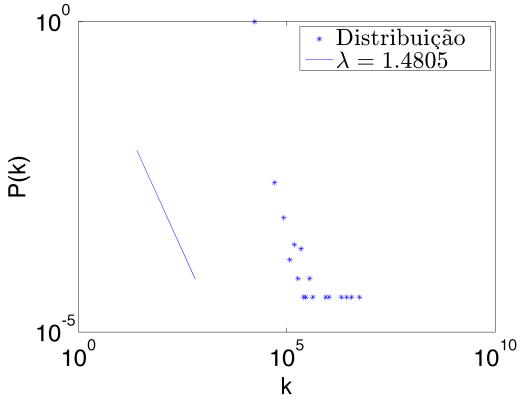
Neste experimento, o objetivo principal é analisar o comportamento da distribuição de graus para os dois tipos de redes estudados: aleatória e a construída utilizando o modelo proposto por esta Tese.

A extração da característica foi feita de duas formas. A primeira delas considera que o grau (de entrada ou saída) é o somatório das arestas (de entrada ou saída). A segunda forma considera o número de arestas (de entrada ou saída) incidentes aos nós. A Figura 46 mostra os resultados obtidos.

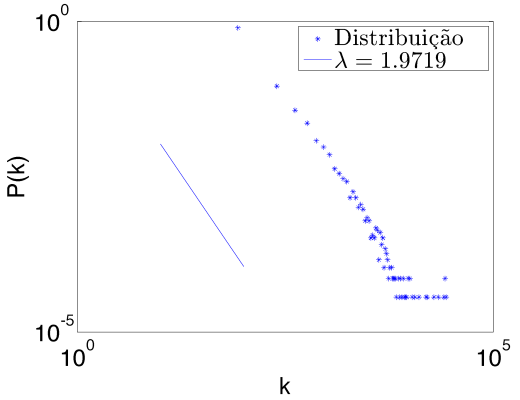
Figura 46 – Distribuição de graus para a rede estudada. Os gráficos da primeira linha foram extraídos considerando a soma dos pesos das arestas. A segunda linha mostra os resultados da contagem de arestas incidentes aos nós.



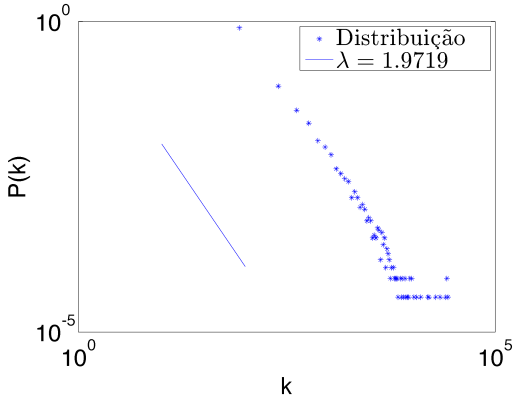
(a) Distribuição de graus de saída ponderado da rede estudada.



(b) Distribuição de graus de entrada ponderado da rede estudada.



(c) Distribuição de graus de saída não ponderado da rede estudada.



(d) Distribuição de graus de entrada não ponderado da rede estudada.

Fonte: Autor

Observando os resultados, nota-se que a distribuição de graus apresenta uma calda seguindo a lei de potência (representada pela linha contínua). Este é um indício de que a rede apresenta comportamento das redes livres de escala (Veja Seção 2.5.1.3). Contudo, este resultado ainda não é suficiente para esta classificação. Outras características devem ser analisadas em conjunto como, por exemplo: raio, diâmetro e coeficiente de clusterização médio.

Outra observação importante é que todas as distribuições possuem $\lambda > 1$. Esse comportamento significa que a quantidade de nós com alto grau é muito menor que os nós com baixo grau decaindo exponencialmente. Pode-se especular que os nós com alto grau são aqueles que aparecem com alta frequência entre imagens distintas; isto é, são vistos em diversos contex-

tos. Por outro lado a existência de nós com baixos graus pode estar relacionada com os nós específicos de um determinado contexto.

A presença deste tipo de comportamento é semelhante à notada recentemente na área da linguística. Em Wachs-Lopes e Rodrigues (2015), foi observado que os nós com alto grau (*hubs*) são os conectivos de uma língua. Esse conectivos estão relacionados com a ordem em que as palavras aparecem no texto, enquanto em uma base de imagens genéricas são objetos ou regiões que aparecem frequentemente nas cenas com muitos outros objetos, não necessariamente em uma ordem específica.

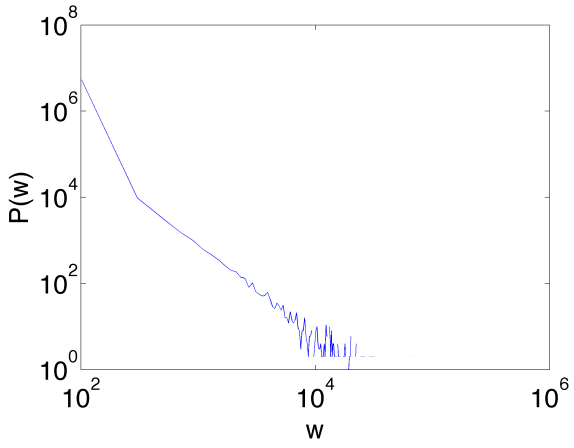
4.6.2 Distribuição de Pesos

A distribuição de pesos, apesar de não ser uma distribuição capaz de classificar as redes em termos de aleatórias, mundo-pequeno ou livres de escala, é importante para que se analise os valores de co-ocorrências entre as instâncias de características/rótulos.

Um valor alto de uma co-ocorrência significa que há uma forte ligação entre os nós envolvidos (foram vistos frequentemente nas mesmas imagens). A Figura 47 mostra os resultados obtidos.

Nessa figura, o eixo x representa o valor do peso $w(i,j)$ de uma aresta e o eixo y representa a probabilidade de $w(i,j)$ ocorrer na rede. Note que, de forma semelhante como o ocorrido com a distribuição de graus, a distribuição de pesos também parece seguir um comportamento de lei de potência. Isso significa que há muitas conexões com baixo peso e poucas conexões com alto peso de co-ocorrência.

Figura 47 – Distribuição de pesos para a rede estudada.



Fonte: Autor

O que se pode concluir desse resultado é que os pesos da rede também apresentam um comportamento que pode ser previsto através da lei de potência e modelos sintéticos podem ser propostos para replicar este tipo de rede. Dessa forma, ao invés de utilizar a etapa de treinamento para construção da rede, um modelo sintético com parâmetros de ajuste pode servir de base como um modelo de rede.

4.6.3 Coeficiente de Clusterização Médio

O coeficiente de clusterização médio (CCM), conforme explicado na Seção 2.5.3.3, é uma das características fundamentais para a classificação de redes complexas entre rede aleatória ou de mundo-pequeno. Sabe-se que um CCM de uma rede de mundo-pequeno deve ser muito maior que um CCM de uma rede aleatória de mesmas características (mesmo número de nós e arestas).

Para o experimento proposto, foi utilizado o mesmo algoritmo aplicado em Wachs-Lopes e Rodrigues (2015), proposto para extração dessa característica a partir de uma amostra dos nós da rede. Uma vez que o CCM é uma média dos coeficientes de clusterização de cada nó da rede, este algoritmo se beneficia do Teorema do Limite Central; isto é, a média de uma amostra tende à média de uma população. Em Wachs-Lopes e Rodrigues (2015) demonstramos empiricamente que uma amostra de 25% dos nós da rede o erro máximo do CCM foi de 4%.

Assim, foram escolhidos aleatoriamente 50% dos nós da rede e extraído o CCM a partir dessa amostra. O resultado final obtido foi $CCM = 0.022599$. Apesar desse valor parecer baixo para uma rede onde espera-se a formação de clusters, ainda não se pode concluir nada sem uma referência. Portanto, uma rede aleatória de mesmas características (mesmo número de nós e arestas) foi construída e o mesmo algoritmo foi executado. O resultado final obtido foi $CCM < 1.0 \times 10^{-5}$. Sendo assim, a rede aleatória obteve um CCM cerca de 2000 vezes menor que a rede em estudo.

Outros trabalhos encontrados na literatura fazem comparações com o modelo aleatório para medir o quanto a rede estudada se afasta de um comportamento aleatório. Um resumo geral sobre alguns desses trabalhos podem ser encontrados em ALBERT e BARABÁSI (2002). A Figura 48 ilustra alguns desses trabalhos. Note que os valores são compatíveis com os obtidos na rede estudada nesta Tese.

Traçando um paralelo do coeficiente de clusterização médio com a tarefa de reconhecimento de objetos, pode-se dizer que um CCM alto comparado com o CCM de uma rede aleatória apresenta características de formação de clusters. Sendo assim, a presença de objetos

Figura 48 – Um comparativo dos estudos feitos em ALBERT e BARABÁSI (2002). Note que o CCM das redes (coluna C) são sempre muito maiores que os CCM das redes aleatórias (coluna C_{rand})

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153, 127	35.21	3.1	3.35	0.1078	0.00023	Adamic 1999	1
Internet, domain level	3015 - 6209	3.52 - 4.11	3.7 - 3.76	6.36 - 6.18	0.18 - 0.3	0.001	Yook *et al.* 2001a, Pastor-Satorras *et al.* 2001	2
Movie actors	225, 226	61	3.65	2.99	0.79	0.00027	Watts, Strogatz 1998	3
LANL coauthorship	52, 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman 2001a,b	4
MEDLINE coauthorship	1, 520, 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman 2001a,b	5
SPIRES coauthorship	56, 627	173	4.0	2.12	0.726	0.003	Newman 2001a,b,c	6
NCSTRL coauthorship	11, 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman 2001a,b	7
Math coauthorship	70, 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.* 2001	8
Neurosci. coauthorship	209, 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.* 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner, Fell 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner, Fell 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya, Solé 2000	12
Silwood park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya, Solé 2000	13
Words, cooccurence	460.902	70.13	2.67	3.03	0.437	0.0001	Cancho, Solé 2001	14
Words, synonyms	22, 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.* 2001	15
Power grid	4, 941	2.67	18.7	12.4	0.08	0.005	Watts, Strogatz 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts, Strogatz 1998	17

Fonte: Autor

de uma mesmo contexto (grupo ou cluster) pode ajudar a identificar outro objeto também presente no mesmo grupo. Por outro lado, um CCM similar ao de uma rede aleatória pode indicar uma desordem na topologia da rede. Sendo assim, a ocorrência de objetos em uma cena pode não ser suficiente para identificar outro objeto.

Os resultados apresentados nessa seção parecem estar de acordo com a literatura. Contudo, um estudo mais profundo é necessário para que se possa classificar a rede aqui estudada. Nas próximas seções, as excentricidades dos nós serão extraídas para obter os valores de raio e diâmetro da rede.

4.6.4 Raio e Diâmetro

A seção anterior mostrou que o coeficiente de clusterização médio da rede foi superior ao encontrado em uma rede aleatória de mesmas características. Aqui, o principal interesse é avaliar o tamanho (tanto ponderados como não ponderados) dos menores caminhos encontrados na rede.

Conforme explicado na Seção 2.5.3.8, o raio e diâmetro são características que são extraídas a partir das excentricidades dos nós. A excentricidade de um nó i é feita computando a menor distância de i para todos os nós da rede. O valor da menor distância de i até o nó mais dis-

tante j, alcançavel por i, será a excentricidade de i. O raio será o valor da menor excentricidade da rede. Por outro lado, o diâmetro é o valor da maior excentricidade.

Considerando esses conceitos, o interesse pelo estudo desta característica está no fato de investigar qual é a menor distância que conecta os nós mais distantes da rede; isto é, pode-se investigar o quão longe dois contextos diferentes podem estar distantes entre si.

O algoritmo executado para a extração das excentricidade foi o Algoritmo de Dijkstra. Esse algoritmo é utilizado para obter a menor distância entre um nó de origem i e todos os outros nós de um grafo. A ordem de complexidade desse algoritmo é $O(E + N \log N)$, onde N é o número de nós e E é o número de arestas. No pior caso, considerando que $E = O(N^2)$, descobrir as menores distâncias entre todos os pares de nós resulta em uma complexidade $O(N^3)$. Assim, para redes densas e com grande quantidade de nós, a obtenção das menores distâncias ainda é um desafio para a computação. Assim, para a extração dessa característica, o algoritmo original Dijkstra foi adaptado para ser executado em paralelo para ser obtido em tempo viável.

Uma vez que o objetivo principal é extrair distâncias, uma transformação nos pesos das arestas foi necessária. No caso da modelagem proposta por esta Tese, quanto maior o peso de uma aresta mais estreita é a ligação entre os nós envolvidos. Contudo, quando se fala em distâncias, quanto maior o valor de uma aresta, mais distante é um nó em relação a outro.

Por esse motivo, propomos o uso da Equação (35) para a transformação necessária.

$$d(i,j) = \frac{1}{w(i,j)} \tag{35}$$

Nesta equação $d(i,j)$ é a distância entre os nós i e j, e $w(i,j)$ é o peso da aresta que conecta os respectivos nós na rede original. Note que, a medida que o peso da aresta aumenta, menor será o valor da distância, sendo $0 < d(i,j) \leq 1 \; \forall i,j$.

Uma vez feita a transformação, o algoritmo é executado. Para cada nó i da rede, o valor da distância do nó mais distante é armazenado em um vetor E na posição i (E_i). Portanto, o vetor E contém as excentricidades de cada nó da rede. Esses resultados serão chamados de excentricidades ponderadas.

A partir desse vetor, foram extraídos o raio e o diâmetro da rede, conforme apontados na Seção 2.5.3.8. Um segundo vetor de excentricidades foi criado porém com uma segunda equação de distância:

$$d(i,j) = \begin{cases} 1 & \text{caso } w(i,j) > 0 \\ \infty & \text{c.c.} \end{cases} \tag{36}$$

Tabela 7 – Resultados de raio e diâmetro para a rede estudada e uma rede aleatória

	Ponderado		Não ponderado	
	Raio	Diâmetro	Raio	Diâmetro
Rede Estudada	1.00	2.00	2	3
Rede Aleatória	0.37	0.37	3	3

Considerando esta equação, o vetor de excentricidades terá como significado o número de arestas de cada nó i até o nó mais distante j alcançavel por i. Portanto, aqui não são considerados os pesos das arestas.

Os mesmos experimentos foram conduzidos em uma rede aleatória de mesmas características. Os resultados são mostrados na Tabela 7.

Os resultados obtidos aqui mostram que a rede aleatória possui um raio e diâmetro ponderado menores que o da rede estudada. A explicação para este fato é que o peso de todas as arestas da rede aleatória considerado foi 8.089, a fim de manter a densidade da rede semelhante à rede estudada. Isso significa que a transformação deste peso em distância será dada por $d(i,j) = 1/8.09 = 0.124$. Portanto, são necessários 3 passos de arestas para se alcançar o valor do raio e o valor do diâmetro. Por esse motivo, os valores de raio e diâmetro não ponderados da rede aleatória resultaram em 3.

Por outro lado, considerando os valores não ponderados de raio e diâmetro, a rede estudada obteve um raio menor que o da rede aleatória. Isso é um indicativo de que há um nó i de tal forma que um outro nó j, mais distante que i, e alcançavel por i, se conecta por meio de 2 saltos (arestas intermediárias).

Nesse ponto, e de acordo com a Seção 2.5.1.3, pode-se aplicar a Equação (37), proposta por WALSH (1999), que relaciona as propriedades de uma rede qualquer com a de uma rede aleatória para fins de classificação de redes de mundo-pequeno.

$$\mu = \frac{C/C_{rg}}{\ell/\ell_{rg}} \gg 1 \tag{37}$$

onde C e l são respectivamente o coeficiente de clusterização médio e diâmetro da rede estudada e C_{rg} e L_{rg} são respectivamente o coeficiente de clusterização médio e diâmetro de uma rede aleatória de mesmas características.

Aplicando esta equação com os resultados dessa seção e da anterior, pode-se notar que o valor de $\mu \gg 1$:

$$\mu = \frac{0.02/1.0 \times 10^{-5}}{2/0.37}$$

Tabela 8 – Tempos absolutos estimados para um Servidor Intel Xeon (A) e o Cluster Titânio da UFABC. Tempo indicador por - não foram estimados. O número de processos e Threads variam para cada experimento. No máximo, 8 threads foram utilizadas em A e 64 em B.

Experimento	Intel Xeon E5-2400 (A)	Cluster Titânio (B)	Escolha
Exp. Seção 4.1	3 dias	-	A
Exp. Seção 4.2	3 dias	-	A
Exp. Seção 4.3	5 min.	-	A
Exp. Seção 4.4	31 dias	3 dias	B
Exp. Seção 4.5	$\geq$ 31 dias	3 dias	B
Exp. Seção 4.6.1	5 min.	-	A
Exp. Seção 4.6.2	5 min.	-	A
Exp. Seção 4.6.3	3 hr.	1 hr.	B
Exp. Seção 4.6.4	12 hr.	1 hr.	B

Esse resultado é um forte indício de que a rede em questão apresenta propriedades de rede de mundo-pequeno.

Contudo, combinando os resultados desta seção com os da Seção 4.6.1, nota-se que o comportamento de lei de potência na distribuição de graus, pode-se ainda sugerir de que esta rede é uma rede livre de escala, uma vez que também apresenta as características de redes de mundo-pequeno.

Os resultados encontrados nesta seção justificam os achados nos experimentos da Seção 4.4 e 4.5, uma vez que um modelo aleatório não faria avaliações com uma taxa de acerto elevada. Assim, esses experimentos se completam e apoiam a ideia de que há uma topologia organizada, formando contextos visuais.

4.7 TEMPOS COMPUTACIONAIS E ESTRATÉGIA DE EXECUÇÃO DOS EXPERIMENTOS

Alguns dos experimentos abordados nas seções anteriores apresentam elevado tempo computacional assintótico. O primeiro deles, apresentado na Seção 4.4, gerou 3400 redes e executou tarefas de análise probabilística do Maximizador para candidato da região escondida. Assim, foram feitas projeções de tempo para a execução dos experimentos em um servidor Intel Xeon E5-2400 com 2.2 GHz de clock e 48GB de memória RAM. Essas projeções são mostradas na segunda coluna (A) da Tabela 8.

Considerando os tempos estimados para os experimentos da Seção 4.4 e 4.5, optou-se pela utilização do cluster Titânio da UFABC. Este cluster possui 40 nós computacionais com

64 núcleos cada um. Além disso, cada nó tem capacidade para 256GB de memória principal. Assim, os algoritmos propostos foram paralelizados e configurados para a execução no cluster. Os tempos estimados absolutos são mostrados na terceira coluna (B) da Tabela 8.

A última coluna dessa tabela informa qual foi a escolha para execução dos experimento. Note que, no caso dos experimentos da Seção 4.4 e 4.5, o tempo foi reduzido em mais de 90%. Sem a utilização do cluster Titânio, a finalização desta Tese no tempo ocorrido seria pouco provável.

5 COMENTÁRIOS E CONCLUSÕES

Neste capítulo serão apresentadas as implicações dos resultados encontrados neste trabalho considerando os aspectos inspirados na Neurociência implementados no modelo proposto por esta Tese. Primeiramente, uma conclusão sobre os experimentos do Capítulo 4 será feita na Seção 5.1, tendo em paralelo as implicações nos sistemas de reconhecimento de objetos e efeitos da Neurociência. Na Seção 5.2, as contribuições deste trabalho para a área de reconhecimento de objetos, bem como a viabilidade de incorporação de novos aspectos do sistema visual humano serão comentados. Finalmente, na Seção 5.4, as aplicações e trabalhos futuros serão descritos.

5.1 RESULTADOS E IMPLICAÇÕES DA NEUROCIÊNCIA

O Capítulo 4 apresentou diversos experimentos relacionados com o Meta-Modelo proposto. Foram realizados experimentos frente a três pontos de vista: análise de características, análise de reconhecimento de objetos e análise de características de Redes Complexas. Com relação à análise de características, os resultados mostraram que há influência tanto do conjunto de características escolhido quanto das próprias configurações na tarefa de reconhecimento de objetos. Esse aspecto do modelo proposto é o que permite simular o comportamento multisensorial competitivo do modelo, comportamento também observado na área da Neurociência visual. Por exemplo (como já foi apresentado), pacientes com agnosia em algum sentido, como para uma determinada cor, ou outro tipo de sensoriamento, pode compensar a percepção visual com outros tipos de sensores ou sensibilidades a outras características.

A segunda frente de estudos, análise de reconhecimento de objetos, mostrou que a informação contextual é uma importante característica que deve ser levada em conta nos sistemas de reconhecimento de objetos. De forma análoga, a informação contextual também é utilizada por humanos.

Finalmente, a terceira linha de estudo, análise de Redes Complexas, mostrou que estudar o modelo do ponto de vista das Redes Complexas apresentou características de alta modularidade (clusterização). Além disso, foi possível encontrar indícios de que tratam-se de redes livres de escala e nenhum comportamento de redes aleatórias foi encontrado. A formação natural de clusters indica que houve o surgimento de contextos, que parecem ter sido fundamentais para os resultados de performance encontrados na Seção 4.4 e Seção 4.5. A modelagem das

co-ocorrências entre instâncias de características (incluindo rótulos) em uma Rede Complexa e a Equação (30) proposta foram resultados das inspirações de aspectos competitivos e colaborativos encontrados na literatura de Neurociência Cognitiva Visual.

5.2 CONTRIBUIÇÕES RELACIONADAS AO MODELO PROPOSTO

Nesta seção, são apresentadas as principais contribuições do modelo proposto considerando a Neurociência, bem como alguns aspectos da Neurociência que não foram implementados no modelo.

5.2.1 Aspectos Implementados

Esta seção discute os aspectos relacionados à Neurociência Cognitiva do sistema visual humano que serviram de inspiração no modelo proposto nesta Tese.

O primeiro aspecto levado em consideração foi o mecanismo de competição e colaboração entre vários tipos de características. Esse aspecto foi incorporado através da rede de co-ocorrências (Seção 3.2.3) e da equação probabilística descrita na Seção (3.2.4.3). Um ponto importante a ser mencionado aqui é que outras características podem ser facilmente incorporadas no modelo, uma vez que cada característica é tratada separadamente. Em um processo de reconhecimento de objetos, as características competem entre si no modelo, ao mesmo tempo que podem ser ponderadas para uma avaliação final.

Outro aspecto levado em consideração no modelo proposto é o mecanismo de reconhecimento Top-Down. As próprias informações de co-ocorrência na Rede Complexa construída na fase de treinamento contempla informações de alto-nível (simulando o envolvimento da área temporal, função *o que?*). Essas informações são responsáveis por ajudar no processo de reconhecimento de objetos sempre que uma região desconhecida é encontrada (vide Seção 4.4).

O reconhecimento Bottom-Up é outro aspecto que foi utilizado de inspiração para o modelo proposto. O módulo de múltiplas segmentações é um processo de baixo e médio nível, que pode produzir informações não interpretáveis cognitivamente. Contudo, o uso de informações de Alto-Nível (co-ocorrências da rede) podem ser utilizadas em fases posteriores de segmentação de forma a colaborar no mecanismo de interpretação de uma cena (veja Seção 4.5).

Além dos aspectos citados anteriormente, o mecanismos de atenção endógena, também implementado através da rede de conhecimento supervisionado pode ser utilizada como forma

de guiar tanto a segmentação quanto o reconhecimento de objetos, dependendo de qual contexto é percebido pelo sistema.

Finalmente, a atenção tardia (ou exógena) é um processo Top-Down em que um objeto pode ser revelado através de altos valores de conexões entre os nós da rede de conhecimento previamente através de conhecimento aprendido. Nesse ponto, pode-se dizer que é o mesmo mecanismo de reconhecimento Top-Down.

5.2.2 Aspectos Não Implementados, Mas Plausíveis de Implementar no Modelo Proposto

Nesta seção, outros aspectos inspirados na Neurociência Cognitiva Visual, que não foram implementados no modelo proposto, mas foram considerados para implementação futura sem grandes mudanças, são discutidos.

O Aprendizado Infinito, um processo de re-adaptação da rede, pode ser constantemente induzido através da saída do maximizador do segmentador e induzido diretamente à rede do Sub-Modelo Central. Essa re-adaptação dos pesos da rede podem remodelar sua topologia e tornar o processo de aprendizagem constante. Se um evento ocorrer infinita e conjuntamente com outros eventos, suas conexões serão reforçadas ao máximo.

Um mecanismo de atenção precoce pode ser implementado no modelo proposto para induzir processos de segmentação em níveis cada vez mais baixos do modelo (mais próximo do segmentador e processamento de baixo-nível) através da alimentação de informações de alto-nível. Por exemplo, quando determinados contextos forem detectados pelo sistema, essa informação poderá ser encaminhada como parâmetros adicionais dos módulos de baixo-nível como, por exemplo, controlar os parâmetros de uma segmentação inicial (processo precoce). Um exemplo simples é a segmentação com o k-means, onde o valores de k pode ser enviado pelo níveis mais altos para que a segmentação ocorra de forma mais rápida sem a necessidade da avaliação constante de segmentações (como apresentado na Seção 4.5). Outros algoritmos de segmentação podem também serem conduzidos de modo que uma segmentação induzida pela rede possa ser construída cada vez mais precocemente.

Conforme abordado na Seção 2.3, sabe-se que o sistema visual humano apresenta pelo menos duas linhas de processamento de alto-nível: *o que?* e *onde?*. Nesse ponto, no modelo proposto por esta Tese não foi previsto o processamento destas duas linhas de processamento de forma independente. Isso significa que um modelo mais preciso poderia ser proposto para identificar a presença de objetos (*onde?*) em determinadas regiões (através de segmentação) sem necessariamente reconhecê-los. Por outro lado, informações contextuais podem ajudar a

inferir a presença de outros objetos (*o que?*) sem necessariamente que o sistema saiba onde se encontra o objeto desconhecido.

Outro comportamento encontrado no sistema visual humano é o mecanismo de atenção exógena reflexa. Neste tipo de mecanismo de atenção, uma instância de característica observada em níveis baixos do modelo poderia disparar eventos às camadas mais altas para que o foco do reconhecimento mude de contexto e o objeto seja o novo foco do sistema.

Além dos pontos apontados aqui, outras características que podem ser extraídas de imagens (tais como as apresentadas na Seção 2.1.1) podem ser adicionadas ao modelo para que haja outros sinais que possam colaborar no processo de reconhecimento de objetos.

5.2.3 Não Implementados e Sem Previsão No Modelo

Alguns aspectos importantes da Neurociência Cognitiva Visual não foram previstos no modelo proposto por esta Tese. O primeiro deles é o processo de reconhecimento de faces. Sabe-se que esta tarefa é feita em regiões específicas do cérebro (lóbulo temporal), assim como a área da escrita (área de broca). Assim, cabe a discussão de novos módulos específicos para estes tipos de tarefas.

A consciência, um mecanismo muito pouco conhecido, também é algo não previsto ainda por nenhum modelo conhecido. Conforme abordado na Seção 2.3, sabe-se que a consciência está ligada à definição e escolha de objetivos. Neste ponto, poderia-se vislumbrar a implementação de um aspecto que pudesse dar autonomia ao processo de reconhecimento de objetos para decidir qual é o objetivo de interesse.

5.3 CONTRIBUIÇÕES RELACIONADAS AO ESTUDO DE REDES COMPLEXAS

Nesse trabalho, as informações ditas “cognitivas” do modelo proposto foram armazenadas na forma de um grafo. Contudo, a análise desse grafo do ponto de vista das Redes Complexas permitiu a melhor compreensão de sua topologia e as implicações desses resultados para o reconhecimento de objetos.

De acordo com a Seção 4.6, foram encontrados principalmente três indícios de que a rede construída segue as características de redes livres de escala. A primeira delas é a distribuição de graus segue a lei de potência. Isso significa que há nós (conhecidos como *hubs*) que exercem o papel de se conectarem a muitos nós. No área da linguística, esses *hubs* são conhecidos como os conectivos de uma língua Wachs-Lopes e Rodrigues (2015).

O segundo indício encontrado nos experimentos é que o coeficiente de clusterização médio da rede construída é relativamente muito maior que o de uma rede aleatória de mesmas características. Isso significa que há formação de grupos de nós fortemente conectados entre si; isto é, o modelo de co-ocorrências formou naturalmente grupos de nós que chamamos aqui de contextos.

Finalmente, o terceiro indício é que, tanto o diâmetro quanto o raio da rede construída, é baixo (no máximo 3). Isso sugere que os maiores caminhos "mais curtos" conectam os nós extremos da rede em poucos passos. Pode-se concluir que a mudança de contexto entre características visuais apresenta características de mundo pequeno, facilitando a navegação na rede e implementação do modelo.

Os resultados encontrados nesse trabalho também se assemelham com os encontrados em ALBERT e BARABÁSI (2002). Assim, pode-se especular que esses sistemas apresentam dinâmicas semelhantes e uma discussão considerando uma analogia pode ser feita.

Até onde sabemos, nenhum trabalho de reconhecimento de objetos fez um mapeamento sobre as co-ocorrências das características na base SUN ou outra semelhante. Assim, os resultados encontrados nesse trabalho podem servir como uma referência para fins de comparações.

Finalmente, a extração de características utilizando algoritmos paralelizáveis em um cluster permitiu a computação de algoritmos antes considerados impraticáveis para esta Tese, como é o caso das excentricidades.

5.4 TRABALHOS FUTUROS

Considerando o Meta-Modelo proposto por esta Tese, há diversos pontos que devem ser foco de estudos mais profundos. Há principalmente duas linhas que podem ser seguidas a partir desse ponto: aplicação e extensão do modelo.

A primeira delas, aplicação do modelo, tem por objetivo implementar este meta-modelo em problemas relacionados com o reconhecimento de objetos. A primeira aplicação prática pode ser feita em um robô autônomo, com uma câmera instalada, que pode ser treinado para detecção de objetos e cenas em um ambiente específico. Esse robô pode ser capaz de andar por esse ambiente, desviar de obstáculos estáticos e em movimento, reconhecer elementos familiares (ao ambiente) e aprender novos elementos (aprendizagem infinita). O ambiente pode ser dinâmico, com variação de elementos diferentes (embora familiares), iluminação e textura. Ao andar pelo ambiente e se deparar com um objeto que não esteja em sua base de conhecimento, pode isolá-lo (atenção reflexa), reavaliar (atenção tardia) e considerar o objeto como um novo

elemento integrante do ambiente. A todo momento, o robô utiliza diversos tipos de sensores para avaliar a cena com características diferentes. Cada característica é avaliada separadamente e combinada de maneira competitiva e colaborativa. Quando um objeto está fora do lugar ou em posição anormal, não percebe e trata como se fosse um objeto diferente. No entanto, pode reaprender essa nova posição se ela passar a ser frequente.

Outra aplicação desse modelo pode ser feita em um sistema de vigilância. Esse sistema pode aprender sobre um determinado ambiente e ser capaz de inferir se houve a intrusão de elementos estranhos na cena. Além disso, pode ser capaz de detectar atividades humanas específicas e ser capaz de inferir, por exemplo, se uma pessoa está andando, descansando, trabalhando, etc.

Um terceiro exemplo de aplicação pode ser feito considerando sistemas de controle de acesso à conteúdo. Determinados contextos podem ser inferidos de imagens e vídeos para que se restrinja o acesso dependendo de qual assunto é tratado na mídia.

A segunda linha que pode ser seguida neste trabalho é a extensão do modelo considerando a união da linguística com imagens e, até mesmo, vídeos. Por exemplo, uma imagem pode ser mostrada a um sistema e o mesmo pode elaborar textos que a descrevam. Com esse tipo de estudo, sistemas de busca de imagens poderiam utilizar esta descrição como indexador e ser capaz de juntar em uma única base de conhecimento os contextos aprendidos tanto através de textos quanto através das imagens lidos na WEB.

Por fim, mas não menos importante, a extensão deste trabalho para análise de vídeos é uma das principais linhas de pesquisa futuras do projeto. Espera-se que algumas características possam ser extraídas utilizando também a informação temporal, uma vez que, na maioria dos casos, não espera-se mudança abrupta de contexto entre os quadros dos vídeos.

5.5 COMENTÁRIOS FINAIS

Parece plausível afirmar que o desempenho do modelo proposto se apoia principalmente na característica de mundo pequeno e livre de escala da topologia da rede. Essa característica, que tem como principal aspecto um organização modular, sem dúvida é o que permite inferências em tempo aceitável, uma vez que é também plausível afirmar que uma topologia de rede aleatória inviabilizaria qualquer aplicação.

Embora a rede tenha sido inspirada no comportamento funcional e não anatômico do sistema visual humano, há cada vez mais indícios na literatura da Neurociência Cognitiva em geral que a estrutura anatômica, do ponto de vista dos agrupamentos e funções neuronais, tam-

bém segue um padrão de mundo-pequeno ou livre de escala Goni et al. (2014), Heuvel e Sporns (2013), Rubinov e Sporns (2010). Isso pode ser devido a um forte indício de uma indução funcional de mesmo comportamento, o que suportaria as implementações realizadas nesta Tese. Se considerarmos que essa afirmação é uma verdade, podemos especular a respeito do parâmetro λ da lei de potência e seu efeito no comportamento do sistema, tanto biológico quanto computacional.

Um fator λ muito grande inviesa o decaimento do gráfico da lei de potência para a esquerda com uma calda muito maior, produzindo uma quantidade menor de *hubs* na rede e uma quantidade maior de objetos visuais específicos. Isso pode significar uma "linguagem visual" mais rica, onde muitos objetos diferentes fazem parte da base de conhecimento e seus relacionamentos são inferidos de maneira mais consciente, difícil de modelar com o conhecimento psicológico de hoje.

Por outro lado, um fator λ muito pequeno inviesa o decaimento da rede para o lado direito, com uma calda muito menor. Isso tende a produzir uma rede com uma maior quantidade de *hubs* e menor quantidade de objetos específicos, produzindo uma "linguagem visual" mais pobre e com muitos objetos pertencentes ao mesmo contexto. É razoável pensar que isso pode ser previsto na área da Neurociência.

Uma dúvida que fica entre muitas é que, seria também razoável especular que tipo de rede seria produzida se os grupos de voluntários utilizados para rotular a base da SUN fossem de diversos tipos, de modo a poder-se realizar experimentos psicofísicos coletivos.

Obviamente, muitos aspectos estudados na Neurociência não foram considerados e não serviram de inspiração para o modelo proposto. Portanto, este modelo ainda é uma pequena contribuição que deve ser gradualmente repensada para agregar outros mecanismos ainda em fase de estudo pela Neurociência. Cabe dizer aqui que a comunicação estreita da área da Ciência da Computação com a Neurociência deve ser fomentada para que haja uma colaboração mútua, onde a Computação e os modelos matemáticos possam simular e prever alguns comportamentos cognitivos e, por outro lado, para que a Neurociência, através de experimentos psicofísicos, abra portas para outros aspectos relacionados à interpretação de cena.

REFERÊNCIAS

ACHANTA, R. et al. Frequency-tuned Salient Region Detection. In: **IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)**. [S.l.: s.n.], 2009. p. 1597 – 1604. For code and supplementary material, click on the url below.

ACHANTA, R.; SUSSTRUNK, S. Saliency Detection using Maximum Symmetric Surround. In: **Proceedings of IEEE International Conference on Image Processing**. [S.l.]: Ieee Service Center, 445 Hoes Lane, Po Box 1331, Piscataway, Nj 08855-1331 Usa, 2010. (IEEE International Conference on Image Processing ICIP).

AGRAWAL, P.; GIRSHICK, R.; MALIK, J. Analyzing the performance of multilayer neural networks for object recognition. In: **Proceedings of the European Conference on Computer Vision (ECCV)**. [S.l.: s.n.], 2014.

AHARON, M.; ELAD, M.; BRUCKSTEIN, A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. **Signal Processing, IEEE Transactions on**, IEEE, v. 54, n. 11, p. 4311–4322, nov. 2006. ISSN 1053-587X.

AKAGüNDüZ, E. 3d object recognition from range images using transform invariant object representation. **Electronics Letters**, Institution of Engineering and Technology, v. 46, p. 1499–1500(1), October 2010. ISSN 0013-5194.

ALBERT, R.; BARABÁSI, A. L. Statistical mechanics of complex networks. **Reviews of Modern Physics**, American Physical Society, v. 74, n. 1, p. 47–97, Jan 2002.

ALBERT, R.; JEONG, H.; BARABASI, A. L. The diameter of the world wide web. **Nature**, v. 401, p. 130–131, 1999.

ALBERT, R.; JEONG, H.; BARABáSI, A.-L. Attack and error tolerance of complex networks. **Nature**, n. 406, p. 376–382, 2000.

ALLESINA, S.; PASCUAL, M. Network structure, predator prey modules, and stability in large food webs. **Theoretical Ecology**, v. 1, n. 1, p. 55–64, mar. 2008.

ANDREOPOULOS, A.; TSOTSOS, J. K. 50 years of object recognition: Directions forward. **Computer Vision and Image Understanding**, v. 117, n. 8, p. 827–891, 2013.

ARKIN, R. C. **An Behavior-based Robotics**. 1st. ed. Cambridge, MA, USA: MIT Press, 1998. ISBN 0262011654.

ASHBRIDGE, E. et al. Effect of image orientation and size on object recognition: Responses of single units in the macaque monkey temporal cortex. **Cognitive Neuropsychology**, v. 17, n. 1-3, p. 13–34, 2000. Cited By 39.

BAARS, B.; GAGE, N. **Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience**. Elsevier Science, 2010. (Introduction to Cognitive Neuroscience Series). ISBN 9780123814401. Disponível em: <http://books.google.com.br/books?id=lEDyN5-80E8C>.

BACKES, A. R.; BRUNO, O. M. Shape classification using complex network and multi-scale fractal dimension. **Pattern Recognition Letters**, v. 31, n. 1, p. 44–51, 2010.

BACKES, A. R.; CASANOVA, D.; BRUNO, O. M. A complex network-based approach for boundary shape analysis. **Pattern Recognition**, v. 42, n. 1, p. 54–67, 2009.

____. Texture analysis and classification: A complex network-based approach. **Inf. Sci.**, Elsevier Science Inc., New York, NY, USA, v. 219, p. 168–180, jan. 2013. ISSN 0020-0255.

____. Texture analysis and classification: A complex network-based approach. **Inf. Sci.**, v. 219, p. 168–180, 2013.

BANDERA, C. et al. Residual q-learning applied to visual attention. In: SAITTA, L. (Ed.). **ICML**. [S.l.]: Morgan Kaufmann, 1996. p. 20–27. ISBN 1-55860-419-7.

BARABáSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **Science**, American Association for the Advancement of Science, Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA., v. 286, n. 5439, p. 509–512, out. 1999. ISSN 1095-9203.

BARABASI, A.-L.; BONABEAU, E. Scale-free networks. **Scientific American**, p. 50–59, Mai 2003.

BARRAT, A. et al. The architecture of complex weighted networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 11, p. 3747–3752, mar. 2004.

BAY, H. et al. Speeded-up robust features (surf). **Comput. Vis. Image Underst.**, Elsevier Science Inc., New York, NY, USA, v. 110, n. 3, p. 346–359, jun. 2008. ISSN 1077-3142.

BEHNKE, S. **Hierarchical Neural Networks for Image Interpretation**. [S.l.]: Springer-Verlag, 2003. v. 2766. (Lecture Notes in Computer Science, v. 2766).

____. **Hierarchical Neural Networks for Image Interpretation (Lecture Notes in Computer Science)**. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003. ISBN 3540407227.

BIEDERMAN, I. On the semantics of a glance at a scene. Lawrence Erlbaum, New Jersey, p. 213–263, 1981.

BO, L.; REN, X.; FOX, D. Hierarchical matching pursuit for image classification: Architecture

and fast algorithms. In: **NIPS**. [S.l.: s.n.], 2011. p. 2115–2123.

____. Unsupervised Feature Learning for RGB-D Based Object Recognition. In: **ISER**. [S.l.: s.n.], 2012.

BOMBINI, L. et al. An evaluation of monocular image stabilization algorithms for automotive applications. In: **Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE**. [S.l.: s.n.], 2006. p. 1562–1567.

BOSCH, A.; ZISSERMAN, A.; NOZ, X. M. Image classification using random forests and ferns. In: **ICCV**. [S.l.]: IEEE, 2007. p. 1–8.

BOSE, S. et al. The cost of an epidemic over a complex network: A random matrix approach. **CoRR**, abs/1309.2236, 2013.

BOYNTON, G. M. Attention and visual perception. **Current Opinion in Neurobiology**, v. 15, n. 4, p. 465 – 469, 2005. ISSN 0959-4388. Sensory systems.

BRAMÃO, I. et al. The role of color information on object recognition: A review and meta-analysis. **Acta Psychologica**, v. 138, n. 1, p. 244 – 253, 2011. ISSN 0001-6918. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0001691811001338>.

BROADBENT, D. E. Stimulus set and response set: Two kinds of selective attention. In: MOSTOFSKY, D. (Ed.). **Attention: Contemporary Theory and Analysis**. [S.l.]: Appleton-Century-Crofts, 1970. p. 51–60.

BRODER, A. et al. Graph structure in the web. **Comput. Netw.**, Elsevier North-Holland, Inc., New York, NY, USA, v. 33, n. 1-6, p. 309–320, 2000. ISSN 1389-1286.

CALDARELLI, G.; CATANZARO, M. **Networks: A Very Short Introduction**. United Kingdom: Oxford University Press, Inc., 2012.

CALVIN, W.; OJEMANN, G. **Conversations with Neil's brain: the neural nature of thought and language**. Addison-Wesley Pub. Co., 1994. (A William Patrick Book). ISBN 9780201632170. Disponível em: <http://books.google.com.br/books?id=iKLuAAAAMAAJ>.

CANCHO, R. F. .; SOLÉ, R. V. The small world of human language. **Proceedings of The Royal Society of London. Series B, Biological Sciences**, v. 268, p. 2261–2266, 2001.

CARLSON, A. et al. Toward an architecture for never-ending language learning. In: **Proceedings of the Twenty-Fourth Conference on Artificial Intelligence (AAAI 2010)**. [S.l.: s.n.], 2010.

CASANOVA, D.; BACKES, A. R.; BRUNO, O. M. Pattern recognition tool based on complex network-based approach. **Journal of Physics: Conference Series**, v. 410, n. 1, p. 012048,

2013. Disponível em: <http://stacks.iop.org/1742-6596/410/i=1/a=012048>.

CHALUMEAU L. DA F. COSTA, O. L. F. Texture discrimination using hierarchical complex networks. In: **Proceedings of the Second International Conference on Signal-Image Technology and Internet-Based Systems**. [S.l.: s.n.], 2012. p. 543–550.

CHANG, L. et al. Object class recognition using sift and bayesian networks. In: **MICAI (2)**. [S.l.: s.n.], 2010. p. 56–66.

CHELLA, A.; MANZOTTI, R. **Artificial consciousness**. [S.l.]: Imprint Academic, 2007. ISBN 9781845400705.

CHENG GUO-XIN ZHANG, N. J. M. X. H. S.-M. H. M.-M. Global contrast based salient region detection. In: . [S.l.: s.n.], 2011. p. 409–416.

CHENG, M.-M. et al. **Salient Object Detection and Segmentation**. [S.l.], 2011. Submission NO. TPAMI-2011-10-0753. Disponível em: <http://mmcheng.net/salobj/>.

COATES, A.; NG, A. The importance of encoding versus training with sparse coding and vector quantization. In: GETOOR, L.; SCHEFFER, T. (Ed.). **Proceedings of the 28th International Conference on Machine Learning (ICML-11)**. New York, NY, USA: ACM, 2011. (ICML '11), p. 921–928. ISBN 978-1-4503-0619-5.

COHEN, M. C.; ALVAREZ, G.; NAKAYAMA, K. Natural scene perception requires attention. **Psychological Science**, 2011.

COHEN, R. et al. Resilience of the internet to random breakdowns. **Phys. Rev. Lett.**, n. 85, p. 3682–3685, 2000.

COLLINS, M.; SINGER, Y. Unsupervised models for named entity classification. In: **In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora**. [S.l.: s.n.], 1999. p. 100–110.

CORBETTA, M. et al. Attentional modulation of neural processing of shape, color, and velocity in humans. **Science**, v. 248, p. 1556–1559, 1990.

CORMEN, T. H. et al. **Introduction to Algorithms**. 2nd revised edition. ed. [S.l.]: The MIT Press, 2001. Taschenbuch. ISBN 0262531968.

COSTA, L. da F. Systems biology through complex networks, signal processing, image analysis, and artificial intelligence. In: **Proceedings of the 16th International Conference on Digital Signal Processing**. Piscataway, NJ, USA: IEEE Press, 2009. (DSP'09), p. 1306–1313. ISBN 978-1-4244-3297-4.

CUADROS, O. et al. Segmentation of large images with complex networks. **2012 25th**

SIBGRAPI Conference on Graphics, Patterns and Images, IEEE Computer Society, Los Alamitos, CA, USA, v. 0, p. 24–31, 2012. ISSN 1530-1834.

DAYAN, P.; SAHANI, M.; DEBACK, G. Unsupervised learning. In: **In The MIT Encyclopedia of the Cognitive Sciences**. [S.l.]: The MIT Press, 1999.

DESAI, C.; RAMANAN, D.; FOWLKES, C. Discriminative models for multi-class object layout. **International Journal of Computer Vision**, 2011.

DESIMONE, R. Visual attention mediated by biased competition in extrastriate visual cortex. **Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences**, v. 353, n. 1373, p. 1245–1255, ago. 1998. ISSN 0962-8436.

DICKINSON, S. J. Rutgers university lectures on cognitive science. In: ____. [S.l.]: Basil Blackwell, 1999. cap. Object Representatino and Recognition, p. 172–207.

DIO C.D., M. M.; RIZZOLATTI, G. The golden beauty: Brain response to classical and renascence sculptures. **PLOS ONE**, 2007.

DRAPER, B. A.; BINS, J.; BAEK, K. Adore: Adaptive object recognition. In: CHRISTENSEN, H. I. (Ed.). **ICVS**. [S.l.]: Springer, 1999. (Lecture Notes in Computer Science, v. 1542), p. 522–537. ISBN 3-540-65459-3.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification (2nd Edition)**. 2. ed. [S.l.]: Wiley-Interscience, 2000. Hardcover. ISBN 0471056693.

DUNNE, J. A.; WILLIAMS, R. J.; MARTINEZ, N. D. Network structure and robustness of marine food webs. **Marine Ecology-Progress Series**, v. 273, p. 291–302, 2004.

EKVALL, S.; KRAGIC, D.; HOFFMANN, F. Object recognition and pose estimation using color cooccurrence histograms and geometric modeling. **Image Vision Comput.**, Butterworth-Heinemann, Newton, MA, USA, v. 23, n. 11, p. 943–955, out. 2005. ISSN 0262-8856.

ERDMANN, H. et al. A study of a firefly meta-heuristics for multithreshold image segmentation. **Computational Vision and Medical Image Processing IV: VIPIMAGE 2013**, CRC Press, p. 211, 2013.

ERDŐS, P.; RÉNYI, A. On random graphs. I. **Publ. Math. Debrecen**, v. 6, p. 290–297, 1959.

FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the internet topology. In: **SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication**. New York, NY, USA: ACM, 1999. p. 251–262. ISBN 1-58113-135-6.

FARKAS, I. J. et al. The topology of the transcription regulatory network in the yeast, s. cerevisiae. **Physica A**, v. 318, n. cond-mat/0205181, p. 3–4. 18 p, May 2002.

FELZENSZWALB, P. F. et al. Object detection with discriminatively trained part-based models. **IEEE Trans. Pattern Anal. Mach. Intell.**, IEEE Computer Society, Washington, DC, USA, v. 32, n. 9, p. 1627–1645, set. 2010. ISSN 0162-8828.

FINN, E. S. et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. **Nat Neurosci**, Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., v. 18, n. 11, p. 1664–1671, 11 2015.

FRINTROP, S.; ROME, E.; CHRISTENSEN, H. I. Computational visual attention systems and their cognitive foundations: A survey. **ACM Trans. Appl. Percept.**, ACM, New York, NY, USA, v. 7, n. 1, p. 6:1–6:39, jan. 2010. ISSN 1544-3558.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological Cybernetics**, v. 36, p. 193–202, 1980.

GALLEGUILLOS, C.; RABINOVICH, A.; BELONGIE, S. Object categorization using co-occurrence, location and appearance. In: **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Anchorage, AK: [s.n.], 2008.

GAVRILA, D. M.; MUNDER, S. Multi-cue pedestrian detection and tracking from a moving vehicle. **Int. J. Comput. Vision**, Kluwer Academic Publishers, Hingham, MA, USA, v. 73, n. 1, p. 41–59, jun. 2007. ISSN 0920-5691.

GAZZANIGA, M. S.; IVRY, R. B.; MANGUN, G. R. **Cognitive neuroscience: The biology of the mind**. 2nd. ed. New York: Norton, 2002.

GERóNIMO, D. et al. 2d-3d-based on-board pedestrian detection system. **Comput. Vis. Image Underst.**, Elsevier Science Inc., New York, NY, USA, v. 114, n. 5, p. 583–595, maio 2010. ISSN 1077-3142.

GLATTFELDER, J. **Ownership Networks and Corporate Control: Mapping Economic Power in a Globalized World**. ETH, 2010. Disponível em: <http://books.google.com.br/books?id=wwzUXwAACAAJ>.

GONI, J. et al. Resting-brain functional connectivity predicted by analytic measures of network communication. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 111, n. 2, p. 833–838, jan. 2014. ISSN 1091-6490.

GONZALEZ, R.; WOODS, R. **Digital Image Processing**. 2nd. ed. New Jersey: Prentice-Hall, 2002.

GOVINDAN, R.; TANGMUNARUNKIT, H. Heuristics for internet map discovery.

INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, v. 3, p. 1371–1380 vol.3, 2000.

GRAEF, P. D.; CHRISTIAENS, D.; D'YDEWALLE, G. Perceptual effects of scene context on object identification. **Psychological Research**, Springer-Verlag, v. 52, n. 4, p. 317–329, 1990. ISSN 0340-0727.

GUELZIM, N. et al. Topological and causal structure of the yeast transcriptional regulatory network. **Nat Genet**, v. 31, n. 1, p. 60–3, 2002. ISSN 1061-4036.

HADSELL, R. et al. Online learning for offroad robots: Spatial label propagation to learn long-range traversability. In: **Proceedings of Robotics: Science and Systems**. Atlanta, GA, USA: [s.n.], 2007.

HAN, I.; YUN, I. D.; LEE, S. U. Modified hausdorff distance for model-based 3-d object recognition from a single view. **Journal of Visual Communication and Image Representation**, v. 15, n. 1, p. 27–43, 2004.

HARALICK, R. Statistical and structural approaches to texture. **Proceedings of the IEEE**, v. 67, n. 5, p. 786–804, May 1979. ISSN 0018-9219.

HARARY, F. **Graph Theory**. [S.l.]: Addison-Wesley, 1969.

HE, X.; ZEMEL, R.; CARREIRA-PERPINDN, M. Multiscale conditional random fields for image labeling. In: **Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on**. [S.l.: s.n.], 2004. v. 2, p. II–695–II–702 Vol.2. ISSN 1063-6919.

HEUVEL, M. P. van den; SPORNS, O. Network hubs in the human brain. **Trends in Cognitive Sciences**, v. 17, n. 12, p. 683 – 696, 2013. ISSN 1364-6613. Special Issue: The Connectome. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1364661313002167>.

HIDALGO, C. A. et al. The product space conditions the development of nations. **Science**, v. 317, p. 482, July 2007.

HILGETAG, C. C. et al. Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, Department of Psychology, University of Newcastle upon Tyne, UK. claush@bu.edu, v. 355, n. 1393, p. 91–110, January 2000. ISSN 0962-8436.

HOIEM D., E. A.; HEBERT, M. Putting objects into perspective. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2006.

HOWARTH, P.; RüGER, S. Evaluation of texture features for content-based image retrieval. In: **In: Proceedings of the International Conference on Image and Video Retrieval, Springer-Verlag**. [S.l.: s.n.], 2004.

HOWARTH, P.; RüGER, S. M. Evaluation of texture features for content-based image retrieval. In: **CIVR**. [S.l.]: Springer, 2004. (Lecture Notes in Computer Science, v. 3115), p. 326–334. ISBN 3-540-22539-0.

HUBEL, D. H.; WIESEL, T. N. Receptive fields of single neurons in the cat's striate cortex. **Journal of Physiology**, v. 148, p. 574–591, 1959.

____. Receptive fields and functional architecture of monkey striate cortex. **Journal of Physiology (London)**, v. 195, p. 215–243, 1968.

HUTTENLOCHER, D. P.; ULLMAN, S. Recognizing solid objects by alignment with an image. **International Journal of Computer Vision**, v. 5, n. 2, p. 195–212, 1990.

ITTI, L.; KOCH, C.; NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. **IEEE Trans. Pattern Anal. Mach. Intell.**, IEEE Computer Society, Washington, DC, USA, v. 20, n. 11, p. 1254–1259, nov. 1998. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/34.730558>.

JAIN, A. K. Data clustering: 50 years beyond k-means. **Pattern Recognition Letters**, v. 31, n. 8, p. 651–666, June 2010. ISSN 01678655.

KAY, K. N. et al. Identifying natural images from human brain activity. **Nature**, Nature Publishing Group, v. 452, n. 7185, p. 352–355, 03 2008.

KHALIGH-RAZAVI, S. What you need to know about the state-of-the-art computational models of object-vision: A tour through the models. **CoRR**, abs/1407.2776, 2014. Disponível em: <http://arxiv.org/abs/1407.2776>.

KHAN, F. S.; WEIJER, J. van de; VANRELL, M. Top-down color attention for object recognition. In: **ICCV**. [S.l.]: IEEE, 2009. p. 979–986.

KIM, S.; YOON, K.-J.; KWEON, I.-S. Object recognition using a generalized robust invariant feature and gestalt's law of proximity and similarity. **Pattern Recognition**, v. 41, n. 2, p. 726–741, 2008.

KOBATAKE, E.; TANAKA, K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. **J. Neurophysiol.**, v. 71, p. 856–867, 1994.

LA, P.; T, D.; WD, H. **Journal of Cognitive Neuroscience**, v. 10, n. 5, 1998.

LAAR, P. van de; HESKES, T.; GIELEN, S. Task-dependent learning of attention. **Neural Networks**, v. 10, n. 6, p. 981 – 992, 1997. ISSN 0893-6080. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0893608097000312>.

LABAYRADE, R.; AUBERT, D. A single framework for vehicle roll, pitch, yaw estimation

and obstacles detection by stereovision. In: IEEE. **Intelligent Vehicles Symposium, 2003. Proceedings. IEEE**. [S.l.], 2003. p. 31–36.

LAN, T. et al. From subcategories to visual composites: A multi-level framework for object detection. In: . [S.l.: s.n.], 2013.

LAZEBNIK, S.; SCHMID, C.; PONCE, J. A sparse texture representation using local affine regions. **IEEE Trans. Pattern Anal. Mach. Intell.**, IEEE Computer Society, Washington, DC, USA, v. 27, n. 8, p. 1265–1278, ago. 2005. ISSN 0162-8828.

LECUN, Y.; BENGIO, Y. The handbook of brain theory and neural networks. In: ARBIB, M. A. (Ed.). Cambridge, MA, USA: MIT Press, 1998. cap. Convolutional Networks for Images, Speech, and Time Series, p. 255–258. ISBN 0-262-51102-9. Disponível em: <http://dl.acm.org/citation.cfm?id=303568.303704>.

LECUN, Y. et al. Gradient-based learning applied to document recognition. In: **Proceedings of the IEEE**. [S.l.: s.n.], 1998. p. 2278–2324.

LEE, H. et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: **Proceedings of the 26th Annual International Conference on Machine Learning**. New York, NY, USA: ACM, 2009. (ICML '09), p. 609–616. ISBN 978-1-60558-516-1.

LILJEROS, F. et al. The web of human sexual contacts. **Nature**, Nature Publishing Group, v. 411, n. 6840, p. 907–908, June 2001. ISSN 0028-0836.

LOWE, D. G. Three-dimensional object recognition from single two-dimensional images. **Artif. Intell.**, Elsevier Science Publishers Ltd., Essex, UK, v. 31, n. 3, p. 355–395, mar. 1987. ISSN 0004-3702.

____. Object recognition from local scale-invariant features. In: **Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2**. Washington, DC, USA: IEEE Computer Society, 1999. (ICCV '99), p. 1150–. ISBN 0-7695-0164-8.

____. Distinctive image features from scale-invariant keypoints. **Int. J. Comput. Vision**, Kluwer Academic Publishers, Hingham, MA, USA, v. 60, n. 2, p. 91–110, nov. 2004. ISSN 0920-5691.

LU, Y.-F. et al. Enhanced hierarchical model of object recognition based on a novel patch selection method in salient regions. **Computer Vision, IET**, v. 9, n. 5, p. 663–672, 2015. ISSN 1751-9632.

MARR, D. **Vision: A Computational Investigation into the Human Representation and Processing of Visual Information**. New York, NY, USA: Henry Holt and Co., Inc., 1982. ISBN 0716715678.

MENG, H. et al. Fpga implementation of naive bayes classifier for visual object recognition. In: . [S.l.: s.n.], 2011. p. 123–128.

MESNIL, G. et al. Unsupervised and transfer learning under uncertainty - from object detections to scene categorization. In: MARSICO, M. D.; FRED, A. L. N. (Ed.). **ICPRAM**. [S.l.]: SciTePress, 2013. p. 345–354. ISBN 978-989-8565-41-9.

MESQUITA, R. G.; MELLO, C. A. B. Segmentation of natural scenes based on visual attention and gestalt grouping laws. In: **IEEE International Conference on Systems, Man, and Cybernetics, Manchester, SMC 2013, United Kingdom, October 13-16, 2013**. [S.l.]: IEEE, 2013. p. 4237–4242.

MILGRAM, S. The small world problem. **Psychology Today**, v. 2, p. 60–67, 1967.

MINUT, S.; MAHADEVAN, S. A reinforcement learning model of selective visual attention. In: **Proceedings of the Fifth International Conference on Autonomous Agents**. New York, NY, USA: ACM, 2001. (AGENTS '01), p. 457–464. ISBN 1-58113-326-X.

MITCHELL, T. M. **Machine Learning**. 1. ed. New York, NY, USA: McGraw-Hill, Inc., 1997. ISBN 0070428077, 9780070428072.

MORAN, J.; DESIMONE, R. Selective attention gates visual processing in the extrastriate cortex. **Science**, v. 229, p. 782–784, 1985.

MOTTAGHI, R. et al. The role of context for object detection and semantic segmentation in the wild. In: **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2014.

NEAPOLITAN, R. E. **Learning Bayesian Networks**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2003. ISBN 0130125342.

NEWELL, M. A. et al. An algorithm for deciding the number of clusters and validation using simulated data with application to exploring crop population structure. **The Annals of Applied Statistics**, The Institute of Mathematical Statistics, v. 7, n. 4, p. 1898–1916, 12 2013.

NEWMAN, M. **Networks: An Introduction**. New York, NY, USA: Oxford University Press, Inc., 2010. ISBN 0199206651, 9780199206650.

NEWMAN, M.; BARABASI, A.-L.; WATTS, D. J. **The structure and dynamics of networks**. [S.l.]: Princeton University Press, 2006. ISBN 978-0-691-11356-2.

NEWMAN, M. E. J. The structure and function of complex networks. **SIAM REVIEW**, v. 45, p. 167–256, 2003.

NEWMAN, M. E. J.; GIRVAN, M. Finding and evaluating community structure in networks.

Physical Review E, American Physical Society, v. 69, n. 2, p. 026113+, Feb 2004.

OLIVA, A.; TORRALBA, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. **International Journal of Computer Vision**, Kluwer Academic Publishers, v. 42, n. 3, p. 145–175, 2001. ISSN 0920-5691.

OLIVA, A. et al. Top-down control of visual attention in object detection. In: **Proc. of the IEEE Int'l Conference on Image Processing (ICIP '03)**. [S.l.: s.n.], 2003.

PALETTA, L.; FRITZ, G.; SEIFERT, C. Q-learning of sequential attention for visual object recognition from informative local descriptors. In: **Proceedings of the 22Nd International Conference on Machine Learning**. New York, NY, USA: ACM, 2005. (ICML '05), p. 649–656. ISBN 1-59593-180-5.

PALETTA, L.; PINZ, A. Active object recognition by view integration and reinforcement learning. **Robotics and Autonomous Systems**, v. 31, p. 71 – 86, 2000. ISSN 0921-8890.

PALMER, S. E. **Vision science : photons to phenomenology**. [S.l.]: MIT Press, 1999. 810 p. Stephen E. Palmer.; "A Bradford book."; Bibliography: Includes bibliographical references (p. [737]-769) and indexes. ISBN 0262161834 Thanks for using Barton, the MIT Libraries' catalog http.

PARIKH, D.; ZITNICK, C.; CHEN, T. From appearance to context-based recognition: Dense labeling in small images. **Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on**, p. 1–8, June 2008. ISSN 1063-6919.

PIÑOL, M. et al. Feature selection based on reinforcement learning for object recognition. In: **Adaptive Learning Agents Workshop**. [S.l.: s.n.], 2012. p. 33–39.

PLEBE, A.; DOMENELLA, R. G. Object recognition by artificial cortical maps. **Neural Networks**, v. 20, n. 7, p. 763 – 780, 2007. ISSN 0893-6080. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0893608007000895>.

PRASAD, D. K. Survey of the problem of object detection in real images. **International Journal of Image Processing (IJIP)**, v. 6, p. 441–466, 2012.

PRICE, D. d. S. Network of scientific papers. **Science**, n. 149, p. 510–515, 1965.

RABINOVICH, A. et al. Objects in context. In: **Proceedings of the International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2007.

RATHI, V. P. G. P.; PALANI, S. Brain tumor mri image classification with feature selection and extraction using linear discriminant analysis. **CoRR**, abs/1208.2128, 2012.

REDNER, S. How popular is your paper? an empirical study of the citation distribution. **The**

European Physical Journal B - Condensed Matter and Complex Systems, v. 4, n. 2, p. 131–134, August 1998. ISSN 1434-6028.

REYNOLDS, J.; CHELAZZI, L.; DESIMONE, R. Competitive mechanisms subserve attention in macaque areas v2 and v4. **Journal of Neuroscience**, v. 19, p. 1736–1753, 1999.

RIBEIRO, B. A. N.; MUNTZ, R. A belief network model for ir. In: **SIGIR '96: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval**. New York, NY, USA: ACM, 1996. p. 253–260. ISBN 0-89791-792-8.

RIESENHUBER, M.; POGGIO, T. Hierarchical models of object recognition in cortex. **Nature Neuroscience**, v. 2, p. 1019–1025, 1999.

RIFAI, S. et al. Higher order contractive auto-encoder. In: **Proceedings of the 2011 European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part II**. Berlin, Heidelberg: Springer-Verlag, 2011. (ECML PKDD'11), p. 645–660. ISBN 978-3-642-23782-9.

RODRIGUES, P.; GIRALDI, G. Improving the non-extensive medical image segmentation based on tsallis entropy. **Pattern Analysis and Applications**, Springer-Verlag, v. 14, n. 4, p. 369–379, 2011. ISSN 1433-7541.

RODRIGUES, P. S.; CHANG, R.; SURI, J. S. Non-extensive entropy for CAD systems of breast cancer images. In: **19th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2006), 8-11 October 2006, Manaus, Amazonas, Brazil**. [S.l.]: IEEE Computer Society, 2006. p. 121–128. ISBN 0-7695-2686-1.

RODRIGUES, P. S.; GIRALDI, G. A.; ARAUJO, A. A. Using tsallis entropy into a bayesian network for cbir. In: IEEE (Ed.). **Proceedings of International International Conference on Image Processing (ICIP'05)**. Genova, Italy: [s.n.], 2005. v. 3, p. 1028–1031.

ROSE; GUREWITZ; FOX. Statistical mechanics and phase transitions in clustering. **Physical Review Letters**, v. 65, n. 8, p. 945–948, maio 1990.

ROWLEY, H. A.; BALUJA, S.; KANADE, T. Rotation invariant neural network-based face detection. In: **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**. Washington, DC, USA: IEEE Computer Society, 1998. (CVPR '98), p. 38–. ISBN 0-8186-8497-6.

RUBINOV, M.; SPORNS, O. Complex network measures of brain connectivity: Uses and interpretations. **NeuroImage**, v. 52, n. 3, p. 1059–1069, 2010.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 2. ed. [S.l.]: Pearson Education, 2003. ISBN 0137903952.

SAUNDERS, B. **Ivan Pavlov: Exploring the Mysteries of Behavior**. [S.l.]: Enslow

Publishers, 2006. (Great minds of science). ISBN 9780766025066.

SCHAEFFER, S. E. Graph clustering. **Computer Science Review**, v. 1, n. 1, p. 27 – 64, 2007. ISSN 1574-0137.

SCHMIDHUBER, J. Multi-column deep neural networks for image classification. In: **Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Washington, DC, USA: IEEE Computer Society, 2012. (CVPR '12), p. 3642–3649. ISBN 978-1-4673-1226-4. Disponível em: <http://dl.acm.org/citation.cfm?id=2354409.2354694>.

S.GAZZANIGA, M. (Ed.). **The Cognitive Neurosciences**. [S.l.]: MIT Press, 1995. (A Bradford book).

SHEN-ORR, S. S. et al. Network motifs in the transcriptional regulation network of escherichia coli. **Nature genetics**, Nature Publishing Group, v. 31, n. 1, p. 64–68, May 2002. ISSN 1061-4036.

SIMARD, P. Y.; STEINKRAUS, D.; PLATT, J. C. Best practices for convolutional neural networks applied to visual document analysis. In: . Institute of Electrical and Electronics Engineers, Inc., 2003. Disponível em: <http://research.microsoft.com/apps/pubs/default.aspx?id=68920>.

SOLOMONOFF, R.; RAPOPORT, A. Connectivity of random nets. **Bulletin of Mathematical Biophysics**, n. 13, p. 107–117, 1951.

SOUNTSOV, P.; SANTUCCI, D. M.; LISMAN, J. E. A Biologically Plausible Transform for Visual Recognition that is Invariant to Translation, Scale, and Rotation. **Frontiers in computational neuroscience**, v. 5, 2011. ISSN 1662-5188.

SPORNS, O. Graph theory methods for the analysis of neural connectivity patterns. **Complex.**, v. 8, n. 1, p. 56–60, 2002. ISSN 1076-2787.

STARZYK, J. A.; PRASAD, D. K. A computational model of machine consciousness. **International Journal of Machine Consciousness**, v. 03, n. 02, p. 255–281, 2011.

STAUFFER, D.; AHARONY, A. Book. **Introduction to percolation theory / Dietrich Stauffer and Amnon Aharony**. 2nd ed.. ed. [S.l.]: Taylor and Francis, London, 1992. x, 181 p. : p. ISBN 0748400273.

STRINGER, S. M.; ROLLS, E. T. Learning transform invariant object recognition in the visual system with multiple stimuli present during training. **Neural Netw.**, Elsevier Science Ltd., Oxford, UK, UK, v. 21, n. 7, p. 888–903, set. 2008. ISSN 0893-6080.

SUN, T.-H. et al. Invariant 2d object recognition using kra and gra. **Expert Systems with Applications**, v. 36, n. 9, p. 11517–11527, 2009.

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning: An Introduction**. [S.l.]: The MIT Press, 1998. Hardcover. ISBN 0262193981.

SZELISKI, R. **Computer Vision: Algorithms and Applications**. 1st. ed. New York, NY, USA: Springer-Verlag New York, Inc., 2010. ISBN 1848829345, 9781848829343.

TAMURA, H.; MORI, S.; YAMAWAKI, T. Texture features corresponding to visual perception. **IEEE Transactions on Systems, Man and Cybernetics**, v. 8, n. 6, 1978.

TANAKA, K. Inferotemporal cortex and object vision. **Annual Review of Neuroscience**, v. 19, p. 109–139, 1996.

TANG, J. et al. Graph structure analysis based on complex network. **Digital Signal Processing**, v. 22, n. 5, p. 713–725, 2012.

TARR, M. J. News on views: Pandemonium revisited. **Nat Neurosci**, v. 2, n. 11, p. 932–935, 11 1999. Disponível em: <http://dx.doi.org/10.1038/14714>.

TEMEL, T. Finding number of clusters in single-step with similarity-based information-theoretic algorithm. **Electronics Letters**, Institution of Engineering and Technology, v. 50, p. 29–30(1), January 2014. ISSN 0013-5194.

TORRALBA A., M. K.; FREEMAN, W.; RUBIN, M. Context-based vision system for place and object recognition. **Procedures of IEEE International Conference on Computer Vision**, 2003.

TORRALBA A., O. A. C. M.; HENDERSON, J. Contextual guindance of attention in natural scenes: The role of global features on object search. **Psychological Review**, 2006.

TROMANS, J. M.; HARRIS, M.; STRINGER, S. M. A computational model of the development of separate representations of facial identity and expression in the primate visual system. **PloS one**, v. 6, n. 10, p. e25616, 2011. ISSN 1932-6203.

TYLER, C.; HARDAGE, L.; MILLER, R. Multiple mechanisms for the detection of mirror symmetry. **Spatial Vision**, v. 9, n. 1, p. 79–100, 1995. Cited By 32.

ULLMAN, S. Three-dimensional object recognition based on the combination of views. **Cognition**, v. 67, n. 1a2, p. 21 – 44, 1998. ISSN 0010-0277.

ULRICH, M.; WIEDEMANN, C.; STEGER, C. Combining scale-space and similarity-based aspect graphs for fast 3d object recognition. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 34, n. 10, p. 1902–1914, Oct 2012.

UNGERLEIDER, L. G.; MISHKIN, M. Two Cortical Visual Systems. In: ____. **Analysis of Visual Behaviour**. [S.l.: s.n.], 1982. cap. 18, p. 549–586.

VAINA, L. (Ed.). **From Retina to Neocortex: Selected Papers of David Marr**. Boston, MA: Birkhauser, 1991.

VASCONCELOS, N.; LIPPMAN, A. A bayesian framework for content-based indexing and retrieval. In: **Data Compression Conference**. [S.l.]: IEEE Computer Society, 1998. p. 580. ISBN 0-8186-8406-2.

VINCENT, P. et al. Extracting and composing robust features with denoising autoencoders. In: **Proceedings of the 25th International Conference on Machine Learning**. New York, NY, USA: ACM, 2008. (ICML '08), p. 1096–1103. ISBN 978-1-60558-205-4.

Wachs-Lopes, G. A.; FUKUMA, W.; RODRIGUES, P. S. Detecção de tipos de tomadas de vídeos de futebol utilizando a divergência de Kullback-Liebler. In: NEVES, L. A. P.; Vieira Neto, H.; GONZAGA, A. (Ed.). **Avanços em Visão Computacional**. 1. ed. Curitiba, PR: Omnipax, 2012. cap. 11, p. 201–218. ISBN 978-85-64619-09-8.

WACHS-LOPES, G. A.; RODRIGUES, P. S. Analyzing natural human language from the point of view of dynamic of a complex network. **Expert Systems with Applications**, v. 45, p. 8 – 22, 2015. ISSN 0957-4174. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0957417415006429>.

WALSH, T. Search in a small world. In: **IJCAI '99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999. p. 1172–1177. ISBN 1-5860-613-0.

WATTS, D. J. **Small Worlds: The Dynamics of Networks between Order and Randomness (Princeton Studies in Complexity)**. [S.l.]: Princeton University Press, 2004.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of /'small-world/' networks. **Nature**, v. 393, n. 6684, p. 440–442, 1998. ISSN 00280836.

WERNICK, M. et al. Machine learning in medical imaging. **Signal Processing Magazine, IEEE**, IEEE, v. 27, n. 4, p. 25–38, jul. 2010. ISSN 1053-5888.

WOLFE, J. M.; CAVE, K. R.; FRANZEL, S. L. Guided search: An alternative to the feature integration model for visual search. **Journal of Experimental Psychology: Human Perception & Performance**, v. 15(3), p. 419–433, 1989.

XIAO, J. et al. Sun database: Large-scale scene recognition from abbey to zoo. In: **CVPR**. [S.l.]: IEEE, 2010. p. 3485–3492.

XU, D.; XU, W. Description and recognition of object contours using arc length and tangent orientation. **Pattern Recogn. Lett.**, Elsevier Science Inc., New York, NY, USA, v. 26, n. 7, p. 855–864, maio 2005. ISSN 0167-8655. Disponível em: <http://dx.doi.org/10.1016/j.patrec.2004.09.030>.

YANG, Y.; SHU, G.; SHAH, M. Semi-supervised learning of feature hierarchies for object detection in a video. **2013 IEEE Conference on Computer Vision and Pattern Recognition**, IEEE Computer Society, Los Alamitos, CA, USA, v. 0, p. 1650–1657, 2013. ISSN 1063-6919.

YU, H.; LIU, Z.; WANG, G. An automatic method to determine the number of clusters using decision-theoretic rough set. **International Journal of Approximate Reasoning**, v. 55, n. 1, Part 2, p. 101 – 115, 2014. ISSN 0888-613X. Special issue on Decision-Theoretic Rough Sets.

ZHANG, D.; LU, G. Review of shape representation and description techniques. **Pattern Recognition**, v. 37, p. 1–19, 1 2004.

ZHANG, Y. et al. A complex network-based approach for interest point detection in images. In: **2012 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)**. [S.l.]: IEEE Press, 2012. p. 1–4. ISBN 978-1-4673-0293-7.

ZHU, J.; MALSBURG, C. von der. Maplets for correspondence-based object recognition. **Neural Networks**, v. 17, n. 8a9, p. 1311 – 1326, 2004. ISSN 0893-6080. New Developments in Self-Organizing Systems.