UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO- UFES

CENTRO DE CIÊNCIAS AGRÁRIAS

DEPARTAMENTO DE COMPUTAÇÃO

ANA PAULA SIMIQUELI BORGES

RECONHECIMENTO DE IMAGENS COM REDES NEURAIS ARTIFICIAS

ALEGRE - ES

2015

ANA PAULA SIMIQUELI BORGES

RECONHECIMENTO DE IMAGENS COM REDES NEURAIS ARTIFICIAIS

Trabalho de Conclusão de Curso apresentado ao Departamento de Computação do Centro de Ciências Agrárias da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.
Orientador: Prof. M.Sc. Jacson Rodrigues Correia da Silva.

Alegre - ES

2015

ANA PAULA SIMIQUELI BORGES

RECONHECIMENTO DE IMAGENS COM REDES NEURAIS ARTIFICIAS

Trabalho de conclusão de curso apresentado ao Departamento de Computação do Centro de Ciências Agrárias da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Aprovado em 25 de novembro de 2015.

COMISSÃO EXAMIDADORA

Prof. M.Sc. Jacson Rodrigues Correia da Silva.
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. Geraldo Regis Mauri
Universidade Federal do Espírito Santo

Prof. M.Sc. Edmar Hell Kampke
Universidade Federal do Espírito Santo

“São as nossas escolhas, mais do que as nossas capacidades, que mostram quem realmente somos”.

JK Rowling

AGRADECIMENTOS

Primeiramente a Deus.

A minha família, principalmente a minha mãe Rosineide, meu pai Gilsimar , e meu padrasto Roberto por serem como pilares durante esses anos de estudo. Ao meu irmão Abilio por ser uma distração nos momentos de angustia e desespero e a minha avó Penha, que sempre me incentivou a estudar, e onde quer que esteja, sei que estará orgulhosa.

Ao Professor Jacson, por ser um mestre durante a graduação, pela paciência, por acreditar em mim mesmo nos momentos em que eu não acreditava, por ser um excelente profissional e um bom amigo.

Aos amigos que foram uma variável essencial ao longo desses anos, que suportaram meus dramas e desesperos. Em especial ao Leandro e o André pela força, apoio, zoações e verdades quando era preciso. Aos amigos do Social das Cachaças, que tornaram tudo mais agradável. E a Thaislane, que mesmo de longe, sempre esteve presente.

Aos professores do Departamento de Computação pelo conhecimento compartilhado e por serem bons colegas, pois isso faz a diferença.

Ao professor Geraldo por disponibilizar um computador para a execução dos testes.

A esta universidade, sua direção, administração e o Restaurante universitário.

A empresa Júnior Click Júnior por me proporcionar uma pequena experiência do mercado de trabalho aumentando assim o conhecimento adquirido nas aulas.

A todos, muito obrigada.

RESUMO

Inconscientemente, nós reconhecemos padrões. Desenvolver sistemas com tal capacidade possibilita a automatização de tarefas que necessitam do raciocínio humano. O Reconhecimento de Padrões realizado em imagens possui diversas aplicações, como no campo de diagnósticos, na área de segurança, e na robótica. Este trabalho abordou o reconhecimento de padrões em imagens por meio de métodos de Inteligência Artificial. Para isso foram testados métodos de segmentação de imagem e classificação. No que diz respeito aos métodos de segmentação de imagens, foram utilizados dois métodos: o filtro de Sobel e o JSEG. Os resultados de cada método são utilizados para comparação. Em relação aos métodos de classificação foram selecionadas duas Redes Neurais Artificiais, a *Multi Layer Perceptron* (MLP), que utiliza a técnica de aprendizado S*hallow Learning*, e a *Convolucional Neural Network* (CNN)*,* que utiliza como técnica de aprendizado o *Deep Learning.* Os testes utilizaram a MLP nas imagens, com a aplicação de métodos de segmentação, o Filtro de Sobel e o algoritmo JSEG. Já o teste da CNN foi feito com as imagens sem a aplicação de nenhum método de segmentação, seguindo as referências da área. Como resultados, a CNN apresentou uma taxa de acerto 27,7% maior em relação a classificação com a MLP.

Palavras-Chave: Reconhecimento de Padrões em Imagens. Inteligência Artificial. Pré-processamento de Imagens. *Deep Learning*. Redes Neurais Artificias.

ABSTRACT

We recognized patterns unconsciously. Developing systems able to recognize patterns automatizes tasks which need for human reasoning. The Pattern Recognition of images can be applied in several fields such as diagnosis, security and robotics. This work treated pattern recognition for images using Artificial Intelligence methods. Image segmentation and classification methods were employed. The segmentation methods Sobel filter and JSEG were used. Two Artificial Neural Networks were used for classification: a Multi Layer Perceptron (MLP) using the Shallow Learning and a Convolucional Neural Network (CNN) using a Depp Learning. Computational experiments were performed using the MLP applied on images obtained from the segmentation methods, and the CNN was applied on images without using segmentation methods. The CNN results improved the hit rate over the MLP in 27.7%.

Keywords: Pattern Recognition in Images. Artificial intelligence. Preprocessing of Images. Deep Learning. Neural Networks,

LISTA DE FIGURAS

Figura 1 - Principais etapas do Reconhecimento de Padrões segundo Erpen (2004). 18
Figura 2 - Exemplo da aplicação da técnica de Motion Deblurring. 20
Figura 3 - Exemplo da aplicação do Filtro de Wiener. 20
Figura 4 - Exemplo da aplicação de filtros para detecção de borda em imagens. 22
Figura 5 - Máscaras de Sobel (NEVES; PELAES, 2008). 23
Figura 6 - Antes e depois da aplicação do JSEG. 24
Figura 7 - Modelo do Neurônio Artificial de McCulloch e Pitts (1943). 29
Figura 8 - Exemplo de arquitetura básica de uma Rede Neural Artificial. 30
Figura 9 - Tipos de Arquiteturas de uma Rede Neural Artificial. 31
Figura 10 - Processo de funcionamento do Backpropagation (SILVA, 2011). 32
Figura 11 - Cótex Visual dos mamíferos (LECUN; RANZATO, 2013). 35
Figura 12 - Características de Convolução de uma Rede Neural utilizando o Deep Learning (LECUN; RANZATO, 2013). 36
Figura 13 - Exemplo de Convolução de uma Imagem (STANFORD, 2015). 38
Figura 14 - Arquitetura de uma Rede Neural de Convolução (LAWRENCE, 1997). 39
Figura 15 - Etapas do Desenvolvimento do Trabalho. 41
Figura 16 - Testes realizados com os filtros monocromáticos para detecção de borda. 42
Figura 17 - Arquitetura da MLP. 45
Figura 18 - Arquitetura da CNN. 45
Figura 19 - Comparação Geral dos Resultados. 49

LISTA DE TABELAS

Tabela 1 - Top-3 parâmetros da MLP. 46
Tabela 2 - Melhores Resultados da Classificação com MLP. 47
Tabela 3 - Classificação com a MLP e a CNN. 47

LISTA DE ABREVIATURAS E SIGLAS

BN - Rede Bayesiana

CNN - *Convolutional Neural Networks*

DBN - Rede Bayesiana Dinâmica

DL - *Deep Learning*

IA - Inteligência Artificial

MLP - *Multi Layer Perceptron*

RNA - Redes Neurais Artificiais

RP - Reconhecimento de Padrões

SVM - Máquinas de Suporte Vetorial

SUMÁRIO

1. INTRODUÇÃO 12
1.1. JUSTIFICATIVA 12
1.2. ESTRUTURA DO TRABALHO 14
2. OBJETIVO 15
2.1. OBJETIVO GERAL 15
2.2. OBJETIVOS ESPECIFICOS 15
3. REFERENCIAL TEÓRICO 16
3.1. RECONHECIMENTO DE PADRÕES 16
3.2 RECONHECIMENTO DE PADRÕES EM IMAGENS DIGITAIS 18
3.2.1 Aquisição da imagem 19
3.2.2 Pré-processamento 19
3.2.3 Segmentação 21
3.2.4 Extração de Feições 24
3.2.5 Classificação 25
3.3 TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA RECONHECIMENTOS DE PADRÕES 25
3.3.1 Redes Neurais Artificiais 28
3.4 DEEP LEARNING 34
3.4.1 Convolutional Neural Networks 37
3.4.2 Deep Learning Frameworks 39
4. METODOLOGIA 41
5. RESULTADOS 46
6. CONCLUSÕES 50
REFERÊNCIAS BIBLIOGRÁFICAS 51
APENDICE I - Código do Filtro de Sobel 57
APENDICE II - Código do MLP 58

1. INTRODUÇÃO

O reconhecimento de padrões é uma tarefa constante em nosso dia a dia. O ser humano inconscientemente reconhece padrões. Ao identificar um rosto amigo em meio à multidão, ao assimilar um cheiro a determinado alimento, ao reconhecer alguém apenas pelo som da voz, até mesmo a leitura desse texto é possível porque estes traços foram reconhecidos como letras que em conjunto formam uma palavra e então é possível identificar o significado de cada palavra.

Desenvolver máquinas capazes de realizar tal procedimento possibilita a automatização de tarefas que carecem do raciocínio humano, como diagnósticos médicos e inspeção automática de componentes industriais (MOREIRA, 2002; GODOI, 2005; PINHEIRO, 2006; SILVA, 2009). O Reconhecimento de Padrões (RP) aplicado em imagens é uma área que está em constante evolução, e possui diversas aplicações, como na área de robótica, possibilitando a locomoção do robô por meio de análise do seu ambiente. Nos radares, em que é possível detectar anomalias nas imagens, entre outros. Cada uma dessas áreas requer soluções diferentes, sendo assim, é importante conhecer sobre imagens digitais, e também quais as etapas que um sistema de reconhecimento de padrões em imagens digitais deve possuir.

Este trabalho aborda o reconhecimento de imagens utilizando métodos de Inteligência Artificial (IA), mais especificamente, Redes Neurais Artificiais, após o pré-processamento adequado da imagem. Para isso, são discutidos os métodos mais utilizados de reconhecimento de padrões em IA, e as técnicas mais comuns em Pré-processamento de imagens.

1.1. JUSTIFICATIVA

Projetar uma Inteligência Artificial tão hábil quanto a capacidade de raciocínio humano encanta os cientistas há anos, não só pelas aplicações que ela viria a oferecer, mas principalmente pela possibilidade de entender o funcionamento da nossa própria inteligência.

Ribeiro (2003) considera que o ato de Reconhecer Padrões é apenas uma das habilidades que a inteligência natural possui, mas, sem dúvida nenhuma, é uma das primordiais devido aos acontecimentos na vida humana tomarem formas de padrões. O modo de falar, a capacidade de ler, reconhecer amigos, o entendimento

das imagens, tudo envolve padrões. O RP é uma tarefa complicada, na qual o homem busca sempre avaliar os acontecimentos em relação aos padrões que os constituem, descobrir a conexão entre os fatos e assim melhor entendê-los e, com isso, adaptar-se à realidade.

A assimilação dos padrões ocorre de forma tão natural que o ser humano não consegue imaginar a complexidade cognitiva que está por trás de reconhecer um padrão. Tais complexidades são perceptíveis quando busca-se reproduzi-las artificialmente em um computador. Mesmo com o desenvolvimento das técnicas inteligentes, como Redes Neurais Artificiais e Lógica Fuzzy, o reconhecimento de padrões continua sendo um dos grandes desafios a ser superado na ciência (NOGUEIRA et al., 2006; SOUZA, 1999).

O Reconhecimento de Padrões em imagens é um dos grandes desafios para os pesquisadores, pois as imagens possuem padrões complexos, com um grande número de características, o que consequentemente torna difícil o seu processamento.

As tarefas de Reconhecimento de Padrões em Imagens Digitais podem ser utilizadas em várias áreas, como em aplicações médicas (MOREIRA, 2002; PINHEIRO, 2006; SILVA, 2009; SIQUEIRA, 2010), aplicações na agricultura (KHATCHATOURIAN; PADILHA, 2008; ANDRADE, 2011), aplicações militares (BURIOUL et al, 2008) e aplicações científicas (RAMOS; ALMEIDA, 2004; QUIÑONEZ et al, 2015).

As aplicações de análise e reconhecimento de imagens são fundamentais na busca pela visão de máquina e tem inúmeras aplicações, como o reconhecimento de faces humanas e a simulação da memória visual de um robô (PAWAN et al, 2005).

Descobrir quais técnicas são mais eficazes no pré-processamento de imagens e quais métodos são capazes de realizar seu reconhecimento com uma alta taxa de acerto representa um avanço na criação de sistemas independentes, que têm a capacidade de imitar o raciocínio humano, alavancando as pesquisas não só na área de Inteligência Artificial, mas também em toda a ciência.

1.2. ESTRUTURA DO TRABALHO

O presente trabalho está organizado da seguinte forma: No Capítulo 2 são apresentados o objetivo geral e os objetivos específicos. O Capítulo 3 expõe uma introdução ao tema Reconhecimento de Padrões, Reconhecimento de Padrões em Imagens, as Técnicas de Inteligência Artificial utilizadas para Reconhecer Padrões, as Redes Neurais Artificiais e o *Deep Learning*. O Capítulo 4 traz a metodologia que foi abordada no desenvolvimento do trabalho. Por fim, os resultados obtidos estão descritos no Capítulo 5, e no Capítulo 6 é feita uma conclusão do trabalho.

2. OBJETIVO

A seguir são descritos o objetivo geral e os objetivos específicos deste trabalho.

2.1. OBJETIVO GERAL

O objetivo geral consistiu em realizar o reconhecimento de padrões em imagens por meio de métodos de Inteligência Artificial, mais especificamente, com Redes Neurais Artificiais.

2.2. OBJETIVOS ESPECIFICOS

Os seguintes passos foram realizados para a obtenção do objetivo geral:

1. Obter um conjunto de imagens para treinamento e classificação.
2. Levantar quais os métodos de processamento de imagens que podem ser utilizados para a quantificação da imagem.
3. Pesquisar os métodos de Inteligência Artificial utilizados para reconhecer padrões em imagens.
4. Com o conjunto inicial de treino, testar os métodos para reconhecimento dos padrões nas imagens.
5. Testar e validar os resultados, juntamente ao conjunto de teste.

3. REFERENCIAL TEÓRICO

A Inteligência Artificial (IA) é uma área da computação que procura representar de forma computacional capacidades como: raciocinar, planejar, resolver problemas e aprender. É possível dizer então que a IA tenta simular o comportamento inteligente do ser humano utilizando como artefato a máquina, mais especificamente, o computador. Norvig e Russell (2014) fazem a separação dos sistemas inteligentes em 4 categorias: Os que pensam como seres humanos, os que pensam racionalmente, os que atuam como seres humanos e os que atuam racionalmente. O escopo deste trabalho são os sistemas que pensam racionalmente, ou seja, que fazem a ação mais correta se baseando no conhecimento que possuem e sempre buscando maximizar seu objetivo.

Bellman (1978) considera a IA como a “automatização de atividades que nós associamos como pensamento humano, tais como tomada de decisão, resolução de problemas e aprendizado”. No momento da tomada de decisão ou na resolução de problemas existem dois caminhos: tentativas por impulso e de forma aleatória ou tentativas baseadas em experiências anteriores que tenham características em comum com a situação atual. A segunda opção define bem uma tarefa conhecida como Reconhecimento de Padrões (RP), discutido na próxima seção.

3.1. RECONHECIMENTO DE PADRÕES

O Reconhecimento de Padrões consiste em extrair as características de uma entrada e identificar padrões nesses dados, permitindo que tal objeto seja classificado em relação às estruturas já conhecidas pelo sistema (RIBEIRO, 2003; SOUZA,2011). Para Tagliaferro (2007), o tema central do conceito de Reconhecimento de Padrões é a ideia de “aprender” através de exemplos, de uma forma totalmente automática ou reduzindo ao máximo o envolvimento do ser humano.

Como Souza (1999) afirma o RP é por natureza uma ciência inexata, logo, pode admitir diversas abordagens, de forma a aproximar, indicar ou encontrar as soluções de um dado problema. A primeira etapa para se entender um acontecimento ou explicar um fato está ligada a perceber quais os padrões do mesmo.

O Reconhecimento de Padrões realizado pelo computador é uma das mais importantes ferramentas usadas no campo da Inteligência Artificial (SOUZA, 1999). Atualmente está presente em inúmeras áreas de conhecimento e possui aplicações de forma direta em visão computacional, comandos de voz, previsão de comportamentos em mercados financeiros, entre outros.

Já existe uma boa quantidade de trabalhos desenvolvidos relacionados com o reconhecimento de padrões e suas aplicações, tais como o de Andrade (2011) que utiliza o Reconhecimento de Padrões para a identificação automática de áreas cafeeiras em imagens de satélite, o de Khatchatourian e Padilha (2008) que realizam o reconhecimento de variedades de soja por meio de processamento de imagens digitais usando Redes Neurais Artificiais, o de Siqueira (2010) que apresenta o Reconhecimento Automático de Padrões em Imagens Ecocardiográficas e os de Nogueira et al. (2006) que apresentam os principais tipos de reconhecimento de padrão utilizados em aplicações com o processo de autenticação de usuários, com intuito de oferecer maior segurança e confiabilidade aos sistemas.

De modo geral, todo reconhecimento de padrões trabalha da seguinte forma: Extrai as características de determinado objeto, as identifica como um padrão conhecido e realiza a classificação. Ribeiro (2003) define “característica” como uma grandeza que é extraída do objeto observado e tem por finalidade ajudar a identificá-lo e o “padrão” como um conjunto organizado das características que foram escolhidas para identificar o objeto, devendo possuir características semelhantes de forma a agrupá-las como pertencentes a uma “classe”, que por sua vez é uma categoria dos padrões presentes no sistema.

Dentre os problemas de Reconhecimento de Padrões, o reconhecimento de imagens é um dos grandes desafios para os pesquisadores. Isso por que as imagens possuem padrões complexos, com um grande número de características, o que consequentemente torna difícil o seu processamento.

A próxima seção descreve o Reconhecimento de Padrões com foco em imagens, apresentando suas características e também métodos atuais utilizados nesse escopo.

3.2 RECONHECIMENTO DE PADRÕES EM IMAGENS DIGITAIS

O Reconhecimento de Padrões aplicado em imagens digitais é uma área que está em constante evolução e possui muitas áreas de aplicação, como no sensoriamento remoto, no processamento de imagens médicas, na robótica, dentre outros. Para cada aplicação existe uma solução diferente, muitas vezes mais de uma solução para uma mesma aplicação. Dessa forma, é importante ter conhecimento sobre imagens digitais, e também saber as etapas do reconhecimento de padrões nas imagens digitais (ERPEN, 2004).

Silva (2005) define uma imagem digital como:

> "A representação numérica e discreta de um objeto, ou especificamente, é uma função quantificada e amostrada, de duas dimensões, geradas por meios ópticos, disposta em uma grade padrão, retangular igualmente espaçada, quantificada em iguais intervalos de amplitude. Assim, uma imagem digital é um vetor retangular bidimensional de amostras de valores quantificados. "

Em sistemas de análise de imagens, o Reconhecimento de Padrão é a parte principal. No entanto, alguns outros processamentos podem ser feitos, como aplicar alguns métodos e técnicas para reduzir a dimensão e a redundância dos dados, além de realizar um pré-processamento nas imagens com intuito de melhorar sua qualidade, que está fortemente relacionada com o sucesso da classificação (ERPEN, 2004; TAGLIAFERRO, 2007). De acordo com Erpen (2004), para um sistema de reconhecimento ser flexível ele deve ser capaz de reconhecer um objeto sem se importar com a sua orientação, tamanho e localização no campo de visão. As etapas do Reconhecimento de Padrões nas imagens digitais são apresentadas

4).

na Figura 1 e descritas nas subseções seguintes.

3.2.1 Aquisição da imagem

A etapa de Aquisição da Imagem compreende a etapa de obter a imagem, seja por um escâner, um sensor, uma câmera ou outro meio de representá-la digitalmente. O meio de aquisição da imagem varia de acordo com o domínio do problema abordado pelo sistema. Os dados também podem ser obtidos por meio de coleções de imagens que contemplem o problema.

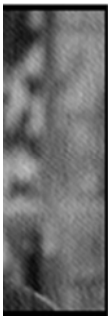
Atualmente existem bancos de dados na internet que disponibilizam arquivos para o treinamento e validação das técnicas utilizadas na classificação de imagens, como é o caso do MNIST (LECUN, 1998), que é um banco de dados com o alfabeto alfanumérico escrito por diversas caligrafias diferentes e do ImageNet (DENG, 2009), que é um repositório que atualmente possui mais de 10 milhões de imagens separadas em mais de 20 mil grupos.

O ImageNet, que é um banco de dados, que utiliza a estrutura hierárquica do WordNet (MILLER, 1995), onde cada conceito relevante no WordNet é descrito por várias palavras chamadas de “Conjunto de Sinônimos”, ou "*synsets*". O ImageNet possui grupos de imagens, como mamíferos, aves, veículos, flor, fruta e ferramentas. Os grupos possuem em média um total de 5000 subgrupos e em média de 500 a 1000 imagens para representar cada subconjunto (DENG, 2009). Após a aquisição da imagem, realiza-se a etapa de pré-processamento, que busca a melhoria da imagem para a aplicação da segmentação.

3.2.2 Pré-processamento

A etapa de pré-processamento tem o objetivo de melhorar a imagem, corrigindo imperfeições e defeitos agregados à imagem no momento da aquisição, a fim de aumentar as chances de sucesso da classificação. Os métodos operam *pixel* a *pixel*, onde o *pixel* é a menor unidade de uma imagem digital. Nessa etapa, uma série de procedimentos são aplicados, como técnicas para aumentar o contraste, filtragem de ruídos, realce, normalização, operações de rotação, realce de bordas, entre outros (ERPEN, 2004; SIQUEIRA, 2010).

Segundo Marques Filho e Neto (1999), as técnicas que são utilizadas para realizar o realce da imagem dependem dos problemas a serem resolvidos, como o *Motion Deblurring*. Na Figura 2 é apresentado um exemplo de sua utilização, que é adequado quando a imagem está borrada devido a um movimento da câmera. E o Filtro de Wiener, apresentado na Figura 3, que é utilizado quando o ruído é muito grande.



após a

Esquef, Alburquerque e Alburquerque (2003) separam as técnicas de pré-processamento em imagens em duas categorias básicas: Os métodos que atuam no domínio espacial, baseados em filtros que tendem a manipular o plano da imagem; e os que atuam no domínio de frequência, baseados em filtros que exercem influência no espectro da imagem. Ao finalizar as correções da imagem, tem-se início a fase de segmentação, onde as principais características da imagem são destacadas.

3.2.3 Segmentação

A segmentação é um passo importante no processo de Reconhecimento de Padrões, pois é responsável por fragmentar uma imagem em regiões ou objetos que apresentam propriedades similares. Porém, o nível da fragmentação varia de acordo com o problema abordado (ERPEN, 2004; TAGLIAFERRO, 2007). De acordo com Siqueira (2010), os algoritmos de segmentação geralmente são baseados em duas propriedades da intensidade dos valores:

- Descontinuidade: a imagem é dividida com base no subgrupo de pontos de um objeto que o separa dos outros na imagem, como por exemplo a borda.

- Similaridade: a imagem é agrupada de acordo com as características semelhantes, utilizando por exemplo, técnicas de limiarização para realizar a separação.

Algumas técnicas baseadas em similaridades são: limiarização, crescimento de regiões, junção e separação, e aglomeração. Já baseadas na descontinuidade encontram-se a detecção de pontos, retas e bordas, por meio de máscaras (NEVES; PELAES, 2008). Em trabalhos com reconhecimento de imagens, a segmentação por

detecção de bordas é a mais indicada, e vem ocasionando bons resultados, como no trabalho de Wang (2014).

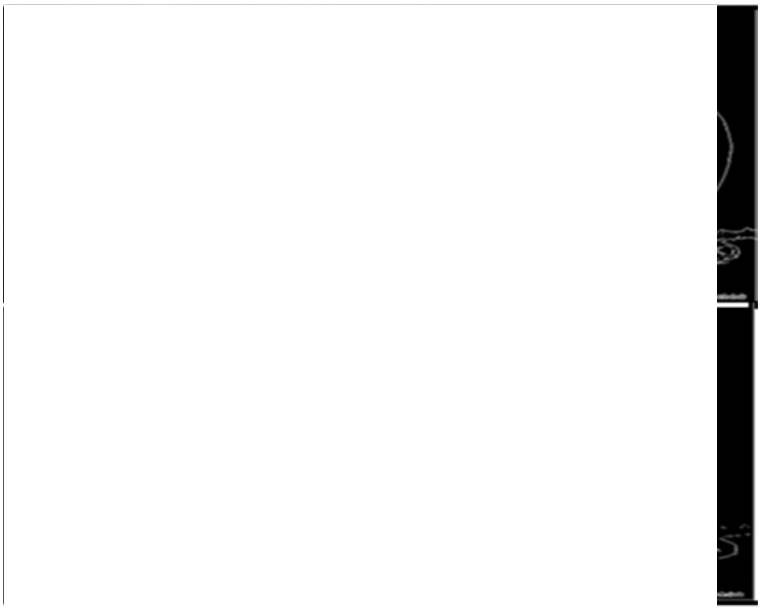
Os algoritmos utilizados para encontrar a descontinuidade realizam uma busca pela imagem utilizando uma máscara a fim de encontrar pontos em que a tonalidade da cor varie, para realçar os *pixels* e assim aumentar o contraste entre o fundo e a borda. Alguns exemplos de filtros para se trabalhar com a detecção de borda em imagens são Sobel, Prewitt, e Roberts, e um exemplo de aplicação é apresentado na Figura 4 (GONZALES, WOODS e EDDINS, 2004).

O filtro de Roberts faz cálculos utilizando matrizes de dimensão 2x2, para encontrar onde as mudanças ocorrem em x e y, com isso ele consegue determinar se o *pixel* é ou não uma borda. O filtro de Prewitt utiliza uma aproximação por derivada e retorna as bordas no local onde o gradiente é máximo. Já o filtro de Sobel faz uso de máscaras para encontrar o gradiente vertical e o horizontal das bordas, para isso utiliza matrizes de dimensão 3x3 como apresentado na Figura 5. A utilização destas máscaras torna o filtro de Sobel menos sensível aos ruídos das imagens, o que torna seus resultados mais precisos, porém sua computação mais complexa do que os demais filtros citados (MATURANA, 2010; NEVES, PELAES, 2008).

de

$$\begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$$

$$|Gx|$$



Em Reconhecimento de Objetos é interessante que além de reconhecer a variação das cores da imagem, o algoritmo de segmentação seja capaz de identificar a textura, uma vez que um objeto pode ser formado por regiões de cor não homogênea, portanto para a segmentação de tais imagens pode ser utilizado algoritmos que formem *clusters* no espaço de cores (CAVANI et al, 2006). Na segmentação de imagens com texturas é importante levar em consideração as diferentes escalas das imagens. Ao longo dos ultimos anos algoritmos para realizar tal tarefa veem sendo desenvolvidos, e um deles é o JSEG proposto por Deng e Manjunath (2002).

O JSEG é um algoritmo não supervisionado para segmentação de imagens utilizando sua cor e textura. Seu funcionamento é constituído de 2 etapas: a quantização do espaço de cores e a segmentação espacial. A primeira etapa consiste em definir as regiões da imagem e utilizar o mínimo de cores possível.

Cada cor é relacionada com uma classe, e a imagem original tem os *pixels* substituídos pelas classes. Na segunda etapa é feita uma análise utilizando as informações das classes geradas na quantização, onde é possível detectar onde é a fronteira das classes e onde ocorre a mudança de cores, feito isso é possível identificar os vales e as regiões com menos valores, o que possibilita identificar onde o crescimento de valores da classe inicial, e com isso o algoritmo integra as regiões que são mais semelhantes aos vales (CAVANI et all, 2006). É possível ver na Figura 6 o exemplo de aplicação do JSEG. Após a aplicação do método escolhido, é preciso encontrar uma forma de extrair as características da imagem para assim poder passar para o classificador.

Figura 6 - Antes e depois da aplicação do JSEG.

3.2.4 Extração de Feições

A saída da etapa de segmentação são os dados brutos de *pixel*. A partir disso, é necessário converter esses dados para uma forma mais apropriada para o processamento pelo computador (SOUZA, 2011). A etapa de extração de feições tem como objetivo encontrar alguma informação quantitativa de interesse ou que seja básica para a separação das classes dos objetos. Para a divisão da imagem é preciso escolher qual será a forma de representação dos dados, que de acordo com Erpen (2004) pode ser por fronteira ou por regiões completas.

A representação por fronteira é apropriada quando o foco do reconhecimento é baseado nas características externas do objeto, como os cantos e os pontos de inflexão. Já a representação por regiões é conveniente quando o interesse é em reconhecer as características internas da imagem, como cor e textura. As duas representações podem ser utilizadas em conjunto, isso depende do problema a ser abordado (ERPEN, 2004). A forma com que a extração de feições é feita influencia diretamente o reconhecimento de imagens e qual abordagem aplicar varia dependendo do escopo do trabalho, como, por exemplo, na classificação de frutas, a representação por fronteiras representa bem o objeto, pois a borda de um abacaxi é diferente da borda de uma maçã.

Para que seja possível reconhecer uma imagem independentemente de variações como tamanho, rotação e translação, Erpen (2004) afirma que os traços

selecionados para serem os descritores devem ser afetados o mínimo possível por tais variações. Selecionado a forma de representação das características é chegada a hora de escolher qual técnica de reconhecimento de padrões será utilizada na classificação.

3.2.5 Classificação

Feita a separação das classes dos objetos que possuem características semelhantes, o passo seguinte é identificar o que elas representam. Para Erpen (2004), a classificação parte do princípio que a similaridade entre objetos implica a existência de características similares, formando assim classes. É nessa etapa que as técnicas de reconhecimento de padrões são aplicadas.

Atualmente existem muitas técnicas de reconhecimento de padrões já desenvolvidas, que podem ser divididas em 3 abordagens (Silva, 2005):

- Estatística: Um conjunto de características são extraídas da imagem e métodos estatísticos são empregados para separar as classes. Alguns métodos são: Classificadores *bayesianos* e regras de decisão.

- Estrutural: Os padrões são apresentados em forma simbólica e os métodos utilizados para o reconhecimento são baseados em casamento de símbolos ou em modelos que tratam os padrões como sentenças.

- Neural: O reconhecimento é feito utilizando as Redes Neurais Artificiais (RNA).

De uma forma geral, a fase da classificação equivale ao reconhecimento de um objeto ou de uma forma, com base no conjunto de classes e padrões que tal imagem apresenta (ERPEN, 2004).

Dentre as abordagens atuais, a de grande destaque é a neural. A próxima seção apresenta o necessário para compreender melhor essa abordagem.

3.3 TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA RECONHECIMENTOS DE PADRÕES

O Reconhecimento de Padrões (RP) é um dos campos tratados em Aprendizagem de Máquina, a qual, por sua vez, é uma das divisões da Inteligência Artificial. Alguns

métodos utilizados no RP são as Redes Neurais Artificiais, a Logica *Fuzzy* e as Máquinas de Suporte Vetorial (SVM).

Um dos campos com maior desenvolvimento nos últimos anos na Inteligência Artificial foi o das Redes Neurais Artificiais. Para Silva (2005) tal crescimento se dá ao fato da capacidade da Rede Neural Artificial trabalhar de uma forma mais delicada com a distribuição dos dados de entrada e a sua capacidade de auxiliar na resolução de problemas não-lineares. Andrade (2011) define as Redes Neurais Artificiais como técnicas computacionais que possuem modelos matemáticos que objetivam a solução de problemas.

As Redes Neurais Artificiais podem utilizar estruturas diferentes, com base em conceitos de probabilidade e lógica por exemplo, gerando assim novos modelos de Redes Neurais (SAHEKI, 2005). Como é o caso da Rede Bayesiana, uma combinação das Redes Neurais e da estatística Bayesiana, que se apresenta como um modelo robusto por conta das capacidades de uma Rede Neural e a solidez fornecida pelos métodos estatísticos (LIMA *et al.,* 2010). Da Silva *et al.* (2008) define as Redes Bayesianas como grafos acíclicos direcionados em que os nós atuam como as variáveis que possuem a medida de incerteza associada. Os arcos mostram se existe a dependência entre duas características, onde a força da dependência é obtida por meio de probabilidades condicionais.

A Lógica *Fuzzy* é uma técnica que trabalha com informações subjetivas de natureza vaga e incerta. O motivo inicial para se utilizar a Lógica *Fuzzy* se dá ao fato das classes encontradas no mundo real não possuírem critérios precisos para a sua separação. Os conjuntos *Fuzzy* são baseados no grau de pertinência dos termos, onde esse grau de pertinência é a probabilidade de um elemento pertencer a um determinado conjunto (MORAES, 2008). Para Marques *et al*. (2005) uma das vantagens que a Lógica *Fuzzy* tem sobre as outras técnicas é que as características apresentam um formato mais fácil de examinar.

A Máquina de Suporte Vetorial (SVM) é uma técnica nova utilizada em Reconhecimento de Padrões, sua abordagem é estatística. Para Lima (2002), enquanto os métodos tradicionais utilizados no Reconhecimento de Padrões trabalham tentando minimizar o erro sobre os dados de treinamento, a SVM busca

minimizar a chance de classificar de forma errônea padrões que ainda não foram apresentados na amostra de treinamento. Ou seja, ela busca maximizar a capacidade de generalização. De acordo com Soares (2008), para realizar a classificação, uma SVM gera um hiperplano entre as classes, onde esse hiperplano tende estar a uma distância máxima da fronteira das classes.

Du et al. (2006) utilizam uma Rede Bayesiana Dinâmica (DBN) para realizar o reconhecimento da interação entre pessoas. Em sistemas inteligentes onde o foco é a percepção de movimentos de um objeto em relação ao ambiente, é usual separar as características em duas classes: As globais que representam a alteração do ambiente como um todo e a relação entre o objeto e o ambiente; e as locais que representam os detalhes do objeto que se tem interesse, nesse caso, as pessoas. A DBN se adequa ao problema pois combina as características globais e locais de uma forma coerente, tornando o reconhecimento mais eficaz, e não precisa sofrer alteração quando o reconhecimento envolve múltiplos objetos.

Park e Aggarwal (2003) apresentam um método para o reconhecimento de interações entre duas pessoas, utilizando uma Rede Bayesiana (BN). Para isso, as partes do corpo que são movimentadas são consideradas como nível baixo e o corpo como um todo é considerado como alto nível. O algoritmo de reconhecimento é precedido por um algoritmo de extração de características que extrai as características que são consideradas nível alto e nível baixo. Então, usando elipses e cascos convexos para representar as partes dos corpos, foi possível construir uma Rede Bayesiana para estimar o corpo do indivíduo em cada quadro. Com a dinâmica dos corpos na sequência de quadros, foi possível reconhecer a interação realizada na cena.

Medjahed et al. (2009) retratam em seu trabalho o desenvolvimento de um sistema para o reconhecimento de atividades da vida diária do ser humano. Para obter os dados foram utilizados sensores de som e infravermelhos, a fim de coletar informações sobre o ambiente e os movimentos de uma pessoa. Na escolha do método utilizado no reconhecimento dos movimentos, várias técnicas foram estudadas, como a Rede Neural Bayesiana, porém a sua performance é limitada quando os dados são heterogêneos. As Redes Neurais possuem capacidade de reconhecimento de dados não lineares, no entanto, possui uma arquitetura

complexa, o que causa a falta de inteligibilidade. Com base nas considerações anteriores, na complexidade dos dados e na falta de conjuntos que possam reproduzir as atividades humanas, a Lógica *Fuzzy* foi escolhida como método de reconhecimento, pois possui uma boa performance, é mais legível, e consegue lidar com imprecisão e incerteza.

Liang *et al.* (2007), propõem em seu trabalho a criação de um sistema para a detecção de distrações que ocorrem aos motoristas nas estradas a fim de melhorar a segurança. A distração pode ser percebida através da movimentação dos olhos dos motoristas e os dados da performance da condução. Para realizar o reconhecimento do padrão de distração foram usados dois métodos para teste: a Máquina de Suporte Vetorial e a abordagem tradicional de regressão logística. Nos resultados, a SVM obteve um melhor desempenho em relação ao método de regressão logística, obtendo a taxa média de precisão de 81,1%.

Para Souza (2011) as RNA's representam um dos mais bem-sucedidos ramos da Inteligência Artificial quando o assunto é reconhecimento de padrões. Vários são os artigos e pesquisas atuais, como o trabalho de Lopes (2015) e Oquab (2014), e vem sendo o foco das pesquisas de grandes empresas, como o Google (GOOGLE, 2015), IBM (IBM, 2015) e Microsoft (MICROSOFT, 2015). Com o intuito de investigar essas técnicas e aplicá-las, esse foi o foco desse trabalho, sendo melhor descrito nas subseções seguintes.

3.3.1 Redes Neurais Artificiais

As Redes Neurais Artificiais são técnicas computacionais que utilizam de modelos matemáticos para obter a solução para determinado problema, elas são inspiradas no sistema nervoso dos seres vivos, de forma a tentar imitar o funcionamento do cérebro humano (SILVA & MARTINS, 2015). Assim como na natureza, um neurônio artificial é uma estrutura relativamente simples, que reage aos estímulos que os neurônios conectados a ele enviam. A Figura 7 mostra o neurônio artificial de McCulloch e Pitts (1943), o primeiro neurônio artificial proposto na literatura, onde cada sinal de entrada do neurônio é multiplicado por seu peso equivalente, e após isso é feito o somatório dos valores resultando em um valor, $X = \sum_{i=1}^{N} x_i * w_i$, que

representa o nível de atividade do neurônio, o resultado passa por uma função de ativação, $y_i = f(x)$, onde o resultado é a saída do neurônio.

Figura 7 - Modelo do Neurônio Artificial de McCulloch e Pitts (1943).

As Redes Neurais Artificiais possuem a capacidade de adquirir conhecimento baseado em informações fornecidas, e através de tais informações fazer a manutenção do conhecimento previamente obtido (SILVA & MARTINS, 2015). O

da

i

fundamento central da teoria das Redes Neurais Artificiais está no fato de que disponibilizar exemplos do relacionamento entre variáveis de uma entrada *x* e um alvo *t*, a rede neural absorverá a relação entre as variáveis, e com isso, ser capaz de generalizar essas informações para novos casos (MOREIRA, 2014).

Uma RNA é determinada por uma arquitetura, um método de aprendizado ou treinamento e uma função de ativação. Como é possível ver na Figura 8, uma RNA

possui um conjunto de neurônios em sua camada de entrada, que são responsáveis por receber os dados (x_1, x_2) a serem classificados. Os neurônios da camada de saída, são os responsáveis por apresentar a saída da rede (y), e usar esse resultado de alguma forma a fim de alterar o contexto em que se encontra. Os neurônios que não são nem de entrada, nem de saída são conhecidos como neurônios internos, e estão inseridos nas camadas ocultas.

A camada oculta tem a função de extrair características dos dados que são enviados das camadas anteriores, seja da camada de entrada ou de outra camada oculta. Todas as camadas são conectadas por meio de pesos, que são responsáveis pela sinapse da Rede Neural. Além disso, uma Rede Neural pode ser totalmente ou parcialmente conectada. Na totalmente conectada cada um dos neurônios de uma camada estão conectados a todos os neurônios da camada adjacente, e na

ıl.

parcialmente conectada algumas dessas conexões não existem.

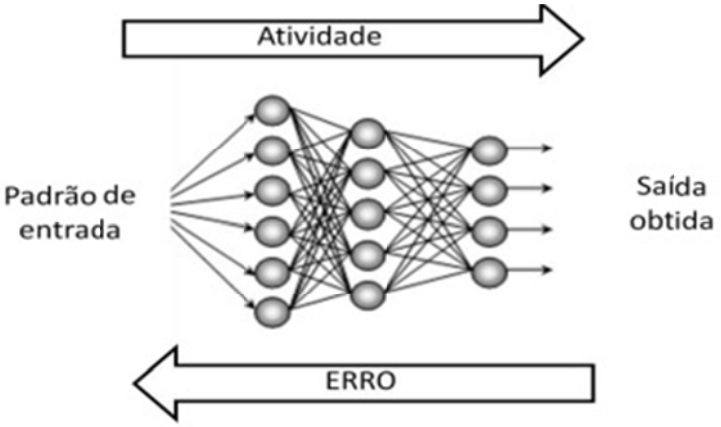
A arquitetura de uma RNA é definida pela maneira como seus neurônios estão estruturados. No geral é possível identificar três classes de arquiteturas de redes, que são apresentadas na Figura 9: as redes alimentadas com camada única onde os neurônios são organizados em uma única camada e a saída desses neurônios são consideradas as saídas da rede (Figura 9 (a)); as redes alimentadas diretamente com múltiplas camadas, onde os neurônios são organizados em mais de uma camada, e os neurônios que não estão na camada de saída, são chamados de neurônios internos (Figura 9 (b)); e as redes recorrentes, na qual a diferença para as outras arquiteturas é que nesta existe um laço de realimentação, onde a saída de um neurônio pode estar ligado a sua própria entrada (Figura 9 (c)) (DORN; FRISKE, 2012).

O processo de aprendizagem de uma Rede Neural significa ajustar a sua matriz de pesos, de forma que o vetor de saída coincida com certo valor desejado para cada vetor de entrada, esse processo pode ser feito de forma supervisionada ou não

(PEREIRA, 2013). De acordo com Haykin (2001) o tipo de aprendizagem é determinado pela maneira com a qual a modificação dos parâmetros ocorre.

Segundo Haykin (2001), no treinamento supervisionado são apresentados a rede pares de entrada e saída, esse grupo consiste em uma entrada (*x)* e a saída desejada para essa entrada (*y_est),* se a saída produzida pela RNA for diferente da desejada, é feito um ajuste nos pesos sinápticos para que a rede consiga armazenar o conhecimento desejado. Tal forma de aprendizado consiste em um processo iterativo de atualização dos pesos sinápticos e, por tal motivo uma RNA é capaz de aprender e generalizar. Já no treinamento não-supervisionado, nenhum conjunto de pares de entrada e saída é apresentado à rede. A própria rede neural "pondera" sobre os conjuntos de dados que são apresentados e determina algumas propriedades dos conjuntos de dados e com isso "aprende" a reproduzir estas propriedades em sua saída. (SOUZA, 2011; PEREIRA, 2013)

Um algoritmo muito utilizado no treinamento supervisionado de Redes Neurais Artificiais é o *Backpropagation,* ele possui duas fases de treinamento, a propagação e a retropropagação. Na propagação a RNA recebe os dados de entrada e realiza sua propagação pela rede, gerando uma saída (y). Essa saída é comparada com a saída estimada (y_est) e gera um valor de erro, após obter esse erro é dado o início ao processo de retropropagação, onde o valor de erro é propagado de volta a RNA de forma a ajustar os pesos de forma a tentar reduzir o erro. O esquema de funcionamento do *Backpropagation* é apresentado na Figura 10 (SILVA, 2011). Esse processo é repetido até que o valor de erro não sofra variações, ou variações



mínimas.

Dado os tipos de treinamentos é possível separar as Redes Neurais em dois grupos: as RNAs Supervisionadas, que utilizam do aprendizado supervisionado, como o *Perceptron*, a *Adaline* e as *Multi-Layer Perceptron*; as RNAs Não-Supervisionadas que utilizam o aprendizado não-supervisionado, como a Rede de Hopfield, o Mapa de Kohonen e os Mapas Auto-organizáveis (SOMs); e as RNAs que podem utilizar as duas abordagens, como é o caso da Convolutional Neural Network, que utilizam o aprendizado não-supervisionado em suas primeiras camadas, para que a rede possa aprender quais as características representam melhor os padrões de entrada, e utiliza o aprendizado supervisionado nas camadas finais para fazer a adaptação dos pesos (LECUN; RANZATO, 2013).

As RNA podem ser aplicadas em diversos tipos de atividade, como: O reconhecimento de padrões, a classificação, a transformação de dados, a predição, o controle de processos e aproximação de funções. Tais atividades podem ser reunidas em dois grupos principais com base no tipo de saída fornecido pela rede neural e o comportamento buscado. As redes para a aproximação de funções devem possuir uma saída com valores contínuos, que são utilizadas para realizar a aproximação de funções. Elas são aplicadas a problemas de regressão. Já as redes para a classificação de padrões: devem atribuir para cada exemplo de entrada, uma classe ao qual ele pertence. Sua saída então é classe que tal grupo provavelmente pertence, logo, deve ser um valor discreto (OSÓRIO; BITTENCOURT; OSÓRIO, 2010).

Quando utilizada para o reconhecimento de padrões o processo de treinamento consiste em corrigir os pesos nas conexões afim de estabelecer relações entre as características e as classes que apontam a melhor discriminação entre os padrões presentes nas diferentes classes. Dessa forma, quando um novo padrão é apresentado a rede, está terá a capacidade de indicar qual a classe que melhor o representa na camada de saída (SILVA, 2005).

Uma RNA que é muito utilizada no RP é a *Multi Layer Perceptron* (MLP), que é uma RNA formada por no mínimo uma camada oculta, onde essas camadas são

responsáveis por realizar o processamento dos dados, ou seja, extrair as características, e utiliza o backpropagation como algoritmo de treinamento (ANDRADE, 2011). Segundo Souza (2011) a MLP foi concebida para solucionar problemas complexos que os neurônios simples não são capazes de solucionar, como a classificação de padrões que possuem a classes que não são linearmente separáveis.

O *Perceptron*, a ADALINE e o *Multi Layer Perceptron* são exemplos de Redes Neurais Artificiais que utilizam o que chamamos de *shallow learning* (Aprendizagem rasa), tal estrutura é chamada assim pois utiliza geralmente uma ou duas camadas de transformação não linear. Esse tipo de aprendizagem se mostra muito eficiente quando se trata de solucionar problemas simples, no entanto, quando se trata de trabalhar com problemas complexos que envolvem o mundo real, como o reconhecimento de sons e o reconhecimento de imagens naturais que possuem um grande número de informações, os resultados são um tanto quanto insatisfatórios. Por isso uma nova área de aprendizagem de máquina vem se destacando nas pesquisas nos últimos anos, o *Deep Learning*.

3.4 DEEP LEARNING

O *Deep Learning* (DL) ou Aprendizagem Profunda é uma subárea do aprendizado de máquina que trabalha com muitas camadas de processamento de informações não-lineares para realizar a extração e transformação de características para fazer a análise e classificação de padrões. De acordo com Deng e Yo (2013) os dois aspectos principais sobre o *Deep Learning* são: que seus modelos trabalham com várias camadas e possuem fases de processamento de informação não-lineares; e que os métodos utilizados para o aprendizado tanto supervisionado quando não supervisionado trabalham com diferentes tipos de abstração dos dados nas camadas.

Para Bengio (2009) existem 2 fatores principais para se estudar o *Deep Learning*, o primeiro é o fato de que o cérebro humano possui uma arquitetura profunda, como o exemplo do córtex visual que mostra uma sequência de áreas, nas quais cada uma possui uma representação da entrada e sinais de fluxo de comunicação de uma área para outra e cada nível dessa hierarquia recebe a entrada com um diferente nível de abstração, como pode ser observado na Figura 11, onde por exemplo o V1, é uma

das áreas que processa a informação visual, e é responsável por reconhecer as formas simples, como bordas (KRUGER, 2013). O segundo motivo se dá fato do processo cognitivo humano apresentar características de aprendizado profundo, e pelo ser humano trabalhar com suas ideias de forma hierárquica, isso é, aprender primeiro conceitos mais simples, e em seguida, segue compondo o conhecimento a fim de torná-lo mais abstrato.

Figura 11 - Cótex Visual dos mamíferos (LECUN; RANZATO, 2013).

A aprendizagem profunda faz interseção entre os campos de redes neurais, Inteligência Artificial, modelagem gráfica, otimização, reconhecimento de padrões, processamento de sinal e de investigação. Deng e Yo (2013) acreditam que três razões que impulsionaram as pesquisas do DL foram: o aumento da capacidade de processamento dos chips, como os GPGPUs, o aumento significativo do tamanho dos dados utilizados para o treinamento e os avanços na aprendizagem de máquina.

Tais fatores permitem que os métodos de aprendizagem profunda sejam capazes de explorar de uma forma eficaz funções não-lineares de composição e com isso aprender de forma distribuída a trabalhar com as características de forma hierárquica.

Diferente da forma convencional de Reconhecimento de Padrões, onde a extração de características é feita manualmente, o *Deep Learning* faz a extração de características de forma treinável, ou seja, o algoritmo é responsável por decidir qual característica representa de forma melhor determinado objeto. A hierarquia de representação das características é feita com um aumento de níveis de abstração, onde cada etapa representa uma transformação de característica treinável, indo do nível baixo, que é considerado o das características que são compartilhadas entre as categorias, para o nível alto, que são as características mais globais e invariantes. No caso do reconhecimento de imagens, começa com o *pixel*, depois a borda, a textura, o motivo, a parte e por fim, o objeto em si, como é possível ver na Figura 11, que é uma visualização das características de convolução de uma rede treinada com o ImageNet de Zeiler e Fergus (2014).

Learning

*D*e acordo com Lecun e Ranzato (2013) o *Deep Learning* possui 3 tipos de protocolo para treinamento. O supervisionado, onde os pesos são inicializados de forma

randômica e o algoritmo é treinado de forma supervisionada, usando por exemplo o backpropagation para fazer o cálculo do gradiente, sendo essa abordagem é comumente usada no reconhecimento de fala e de imagens. O não supervisionado com um classificador supervisionado no topo, onde cada camada é treinada de forma não supervisionada, e o classificador é treinado no topo, mantendo as outras fixas, sendo que essa abordagem é recomendada quando se tem poucos exemplos de saída. E o não supervisionado com um supervisor global de ajuste fino, no qual, como no anterior, cada camada é treinada de forma não supervisionada, só que após isso, adiciona-se a camada de classificação e é feito um retreinamento de todas as camadas.

É possível ver a aplicação do *Deep Learning* em métodos como *Deep Boltzmann Machines, Stacked Auto-Encoders,* as *Convolutional Neural Networks* e as *Deep Belief Networks*(LECUN; RANZATO, 2013). Neste trabalho, daremos enfoque nas Convolutional Neural Networks – Redes Neurais de Convolução (CNNs), pois de acordo com Nielsen (2015) essas redes possuem uma arquitetura especial que é particularmente bem ajustada para a classificação de imagens.

3.4.1 Convolutional Neural Networks

Segundo Arel et al (2010) as CNNs são da família das redes neurais com multicamadas, e foram propostas como um *framework* do DL que é motivado pela necessidade mínima de se realizar um pré-processamento dos dados. Uma rede de convolução é projetada para reconhecer formas bidimensionais que possuam um alto grau de invariância como por exemplo, a translação. Elas foram a primeira abordagem de *Deep Learning* que obtiveram realmente sucesso, onde muitas camadas foram treinadas com sucesso em uma forma robusta.

As CNN's tem conjuntos como entrada e saída de suas camadas, que são chamados de mapas de características. Na saída, cada mapa representa uma característica particular que foi extraída dos dados de entrada. E todas as camadas são treinadas de forma simultânea a fim de minimizar a função global (HUANG; LECUN, 2006). Os pesos das camadas são aprendidos através do treinamento, mas vale ressaltar que a rede aprende a extrair suas próprias características automaticamente. Uma CNN consiste em um conjunto de camadas de convolução e subamostragem (agrupamento) que podem ser ou não completamente conectadas.

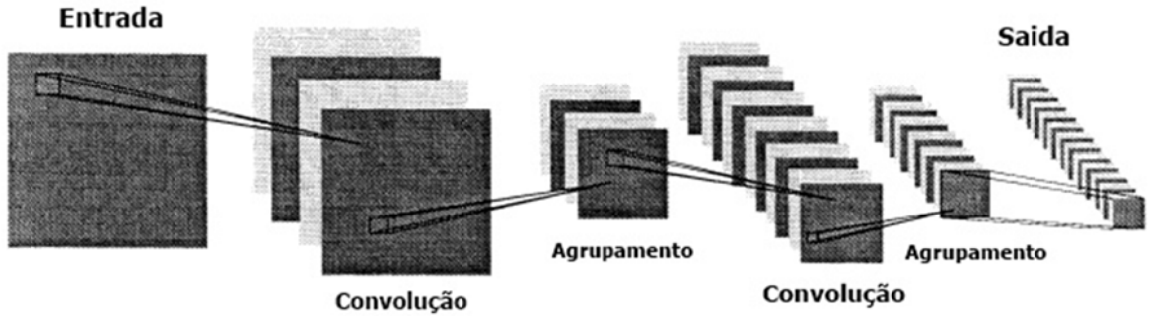
Em uma Rede Neural Artificial, ao se trabalhar com imagens é comum utilizar o *pixel* como dado de entrada (x_i, x=padrão de entrada, i=índice do *pixel*), então considerando uma imagem com tamanho de 256x256 *pixeis*, como as utilizadas neste trabalho, só para a camada de entrada são mais de 65 mil neurônios, o que se estende por toda a rede, e se torna um custo computacional alto e uma tarefa demorada.

Levando em consideração que as imagens naturais possuem uma propriedade chamada de estacionária, onde as estatísticas relacionadas a uma região são as mesmas do que de outra região, compreende-se que as características que se aprendem de uma parte da imagem também podem ser aplicadas a outras partes. Dessa forma as mesmas características podem ser usadas em outros locais (STANFORD, 2015).

A ideia da convolução é escolher um local e extrair suas características, e aprender com elas para obter um valor de ativação daquele local, como é apresentado na Figura 13. Após a convolução o vetor de características ainda é consideravelmente grande, e isso é ruim para a aprendizagem da rede, portanto, levando ainda em consideração a propriedade de ser estacionária, pode-se representar uma imagem com grande escala agrupando a estatística de suas características em vários locais

(STANFORD, 2015).

De acordo com Nielsen (2015), as CNN's possuem três ideias básicas: Campos receptivos locais, pesos compartilhados e agrupamento. O campo receptivo local é uma janela sobre os neurônios da entrada, onde uma seleção dos dados de entrada é feita para forçar a cada neurônio a extrair características locais da camada anterior, no caso, a convolução. Além disso, como citado acima, cada camada



possui seus mapas de características, nesses mapas os neurônios individuais são restringidos a compartilhar o mesmo conjunto de pesos sinápticos. Vale destacar que cada camada de convolução é seguida por uma camada que realiza cálculos como da média local e agrupamento, para reduzir a resolução do mapa de características. A Figura 14 apresenta uma arquitetura típica de uma Rede Neural Aritificial de Convolução.

3.4.2 Deep Learning Frameworks

Atualmente existem vários *frameworks* para se trabalhar com o Deep Learning, como o Torch (COLLOBERT, 2002) que é um *framework* de computação cientifica em Lua e tem suporte para a aprendizagem de máquina. O Theano (BASTIEN, 2012) que também é um *framework* de computação cientifica, só que implementado e aplicado em Python e o Caffe (JIA, 2014) que é um *framework* especifico para se trabalhar com o *Deep Learning*. Tais frameworks são bastante parecidos, eles conseguem representar os modelos de aprendizado profundo, são da iniciativa *open-source* e todos incluem scripts para se trabalhar, portanto a escolha de qual *framework* utilizar ao se trabalhar com o *Deep Learning* fica a critério do usuário. Para este trabalho o framework escolhido foi o Caffe, pois trabalha com o DL de uma

forma mais simples, como, por exemplo, seus modelos são em arquivos de texto, possui modelos prontos para se trabalhar com o *dataset* ImageNet, e atualmente vem sendo utilizado em várias pesquisas envolvendo o *Deep Learning*.

4. METODOLOGIA

A metodologia aplicada para o desenvolvimento do presente trabalho foi dividida em etapas, como é possível ver na Figura 15. A primeira etapa foi selecionar o conjunto de dados, no caso as imagens, que foi utilizado no treinamento e teste dos métodos de IA. O *dataset* escolhido foi ImageNet, e para os experimentos do trabalho foi escolhido o conjunto das frutas, mais especificamente 6 subconjuntos: maçã, morango, laranja, abacaxi, limão e banana. O ImageNet foi escolhido por ser o banco de dados mais utilizado quando se trata de Reconhecimento de Padrões, e está presente em diversos trabalhos, como em Russakovsky et al. (2014), Girshick et al (2013) e Deng et al. (2010).

Após a obtenção das imagens, iniciou-se a segunda etapa, na qual foi feito o pré-processamento e segmentação das imagens escolhidas. Como citado na subseção 3.2.2 o pré-processamento é utilizado para corrigir na imagem os defeitos agregados obtidos no momento de sua aquisição. Para uma imagem compor o *dataset* do ImageNet é importante que ela esteja em boa resolução (DENG et al, 2009). Portanto o único pré-processamento aplicado foi o redimensionamento das imagens, transformando-as para o tamanho de 256X256 *pixeis*, uniformizando o tamanho das imagens, que possuíam tamanhos distintos, e para uma Rede Neural Artificial é interessante ter um tamanho padrão de entrada, pois os *pixeis* os dados passados para a camada de entrada da RNA.

Para a segmentação das imagens, foram testados algoritmos monocromáticos e de segmentação por textura e cor. É possível ver na Figura 16 o exemplo de aplicação de alguns filtros monocromáticos sobre uma amostra dos dados. Para os testes do trabalho foi escolhido o filtro de Sobel pois mesmo tendo uma maior complexidade em sua computação apresenta uma representação visual melhor das bordas. O código para a aplicação do filtro de Sobel nas imagens foi desenvolvido na linguagem Python (PYTHON, 2015), e está disponível no APENDICE I. O algoritmo utilizado para segmentação de cores foi o JSEG, o qual teve o código fonte disponibilizado pelos autores (Deng e Manjunath, 2002). Para a utilização do algoritmo é preciso informar o valor escolhido para ser o limiar da quantização das cores, para este trabalho foi utilizado o limiar 16, tal valor foi obtido através de testes com valores empíricos. O tempo computacional para se aplicar a segmentação com o JSEG foi de em média 5 minutos para cada imagem, levando 2 dias para cada

conjunto escolhido.

Figura 16 - Testes realizados com os filtros monocromáticos para detecção de borda.

berts, (c)
obel

Após as imagens segmentadas deu-se início a terceira etapa do desenvolvimento, a pesquisa dos métodos de IA. De acordo com Nielsen (2015) as Redes Neurais Artificiais (RNA) e o *Deep Learning* atualmente fornecem as melhores soluções para muitos problemas no reconhecimento de imagem, reconhecimento de voz e processamento de linguagem natural, sendo assim, esses serão os métodos utilizados no trabalho. A implementação da Rede Neural foi feita em Python, utilizando a biblioteca *scikit-neuralnetwork (*SKNN, 2015*)*, e o reconhecimento com *Deep Learning* foi implementado utilizando o *framework* Caffe (CAFFE, 2015; JIA, 2014).

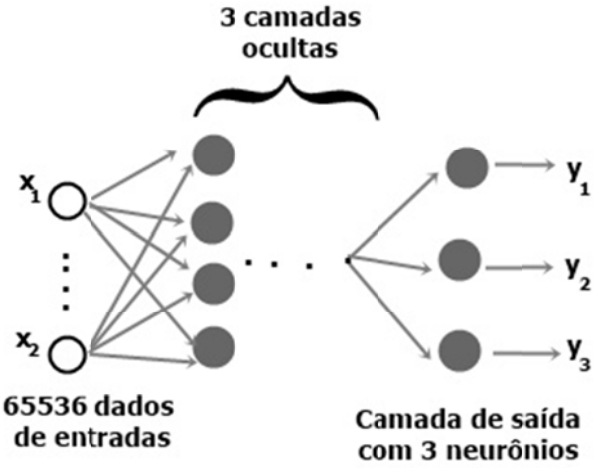
Foi feita a implementação de uma *Multi Layer Perceptron* (MLP) para classificar as imagens sem a aplicação de técnicas de segmentação, com o Filtro de Sobel, e com a Segmentação por JSEG (RNA-1), sendo o código fonte apresentado do APENDICE II, e uma *Convolutional Neural Network* (RNA-2). Vale ressaltar que os dados a serem classificados pela CNN são as imagens sem nenhum tipo de segmentação. Os parâmetros da RNA-1 foram sendo calibrados durante os testes. Como por exemplo o número de camadas ocultas e o número de neurônios. A RNA-2 foi construída utilizando o modelo “bvlc_caffenet” que foi desenvolvido por Krizhevsky (2012), tal modelo quando aplicado por seus criados no conjunto total de imagens do ImageNet obteve 67,4% como taxa de acerto.

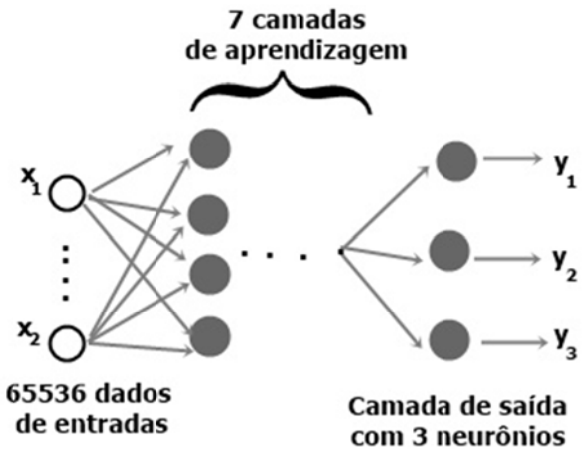
A quarta etapa do desenvolvimento do presente trabalho foi referente ao teste dos códigos. Uma dificuldade encontrada foi em relação a memória utilizada pelo treinamento das Redes Neurais. Executando o código para treinar a RNA-1 em uma máquina com processador i5 e com 4 GB de memória RAM, obteve-se o erro por falta de memória. Por esse motivo foi necessário adicionar memória a máquina para realizar os testes. Em uma máquina com 8 GB de memória RAM e processador i5 também não foi possível realizar os testes por conta do erro de memória. Vale salientar que o erro de memória se dá ao fato de que na RNA-1 todas as informações precisam ser carregadas para a RAM para realizar o treinamento, já na RNA-2 pode-se realizar o carregamento das imagens por lotes. Portanto, para o treinamento foi usada uma máquina com 12 GB de memória RAM e processador i5, porém os testes utilizando as 6 classes descritas acima, um total de 3664 imagens,

ainda provocava o erro de falta de memória para a RNA-1. Por não dispor de uma máquina com a memória maior, o número de classes foi reduzido para 3, no caso maçã, morango e banana, mantendo um total de 1446 imagens. Durante os treinos no momento da calibragem dos parâmetros da Rede Neural, como o número de camadas, a quantidade de neurônios em cada camada, ocorreram muitos erros devido a memória, o que limitou as variações para teste da RNA-1.

Os testes com a RNA-2 foram realizados na mesma máquina que na RNA-1, com 12 GB de memória RAM e processador i5, e a placa de vídeo GeForce GTS 240 da NVIDIA (NVIDIA, 2015). Ao se trabalhar com métodos que utilizam o Deep Learning, é recomendado utilizar a placa de vídeo para acelerar o processo de aprendizagem da rede. A RNA-2 tinha possibilidade de realizar o treinamento com as 6 classes de imagens, no entanto, para manter a comparação dos métodos e do conjunto usado, a RNA-2 foi treinada e testada com as mesmas amostras da RNA-1.

A Figura 17 apresenta a arquitetura da RNA-1, que possui 65536 dados de entrada, pois cada *pixel* de uma imagem é tratado como um dado de entrada, 4 camadas, das quais 3 são camadas ocultas mais a camada de saída, que por sua vez possui 3 neurônios, sendo que cada neurônio de saída representa uma classe, ou seja, maçã, morango ou banana. A Figura 18 apresenta a arquitetura da RNA-2, que possui oito camadas de aprendizagem, sendo cinco camadas de convolução e 3 são totalmente conectadas, de forma similar a RNA-1, possui 65536 dados de entrada e a camada de saída apresenta três neurônios que representam as classes.


3 camadas
ocultas
x1
x2
y1
y2
y3
65536 dados
de entradas
Camada de saída
com 3 neurônios


7 camadas
de aprendizagem
x1
x2
y1
y2
y3
65536 dados
de entradas
Camada de saída
com 3 neurônios

5. RESULTADOS

Ao trabalhar com Redes Neurais Artificiais é necessário realizar a calibragem dos parâmetros, como o número de camadas e qual a taxa de aprendizagem a ser utilizada. Os melhores resultados obtidos na calibração são apresentados na Tabela 1, vale ressaltar que os valores apresentados na Taxa de Acerto são referentes ao acerto de classificação da Rede Neural na fase de teste.

Tabela 1 - Top-3 parâmetros da MLP.

N° de Camadas	Taxa de Aprendizagem	Tempo Médio (s)	Taxa de Acerto
4	0.02	1045.9	40 %
3	0.02	795.0s	35 %
4	0.2	1165.2	32 %

Não foi possível realizar testes com um maior número de camadas, os testes com, por exemplo, 5 camadas causaram erros de memória, como citado na seção 4. É importante salientar que no treinamento de uma Rede Neural o tempo não é um parâmetro externo, comumente uma RNA treina até o momento em que sua taxa de erro chega a um ponto estável.

Neste trabalho foi utilizado como método de parada a estabilidade da taxa de erro por 10 iterações, com variação de 0.001. Portanto os parâmetros utilizados para o treinamento da RNA-1 neste trabalho foi o número de camadas igual a 4 e a taxa de aprendizagem de 0.02. Como citado acima, o modelo que é utilizado no treinamento da RNA-2 foi o utilizado por Krizhevsky (2012), não necessitando de uma calibração de parâmetros.

Os testes realizados no presente trabalho foram: Classificação com a MLP com as imagens sem nenhum tipo de segmentação; Classificação com a MLP utilizando o filtro de Sobel; Classificação com MLP utilizando o algoritmo de segmentação JSEG; e a Classificação com a CNN com as imagens sem nenhum tipo de segmentação. A classificação com a CNN não foi feita utilizando filtros pois iria fugir do princípio do método de Deep Learning, que é tentar simular o processo de aprendizagem humano, como descreve Bengio (2009).

A primeira comparação realizada foi com a MLP utilizando os métodos de segmentação. Foram utilizadas 1204 amostras de imagens para o treinamento e 302 amostras para o teste. É possível ver na Tabela 2 uma comparação dos melhores resultados obtidos.

Tabela 2 - Melhores Resultados da Classificação com MLP.

Método utilizado para a Segmentação	**Taxa de Acerto**	**Tempo de Segmentação**	**Tempo de Treinamento**	**Tempo Total**
Sem Segmentação	35,67%	0d 0h 0m 0s	0d 0h 16m 32s	0d 0h 16m 32s
Filtro de Sobel	41.72 %	0d 0h 4m 0s	0d 0h 17m 31s	0d 0h 21m 31s
JSEG	46.98 %	5 d 0h 50m 0s	0d 0h 18m 50s	5d 1h 8m 50s

O JSEG mostrou-se o melhor na segmentação das imagens, conseguindo uma taxa de 46,98% de acerto da classificação dos dados. É importante destacar o tempo levado para se segmentar as imagens, enquanto o Filtro de Sobel levou em média 4 minutos para realizar a segmentação no conjunto de imagens, o JSEG para realizar o mesmo procedimento levou aproximadamente 5 dias, pois o método utilizado para encontrar a fronteira entre as texturas é de custo computacional alto. No entanto, vale ressaltar que enquanto o filtro de Sobel destaca a borda, o JSEG separa a imagem em regiões levando em consideração sua textura e cores, o que fornece uma representação melhor da imagem.

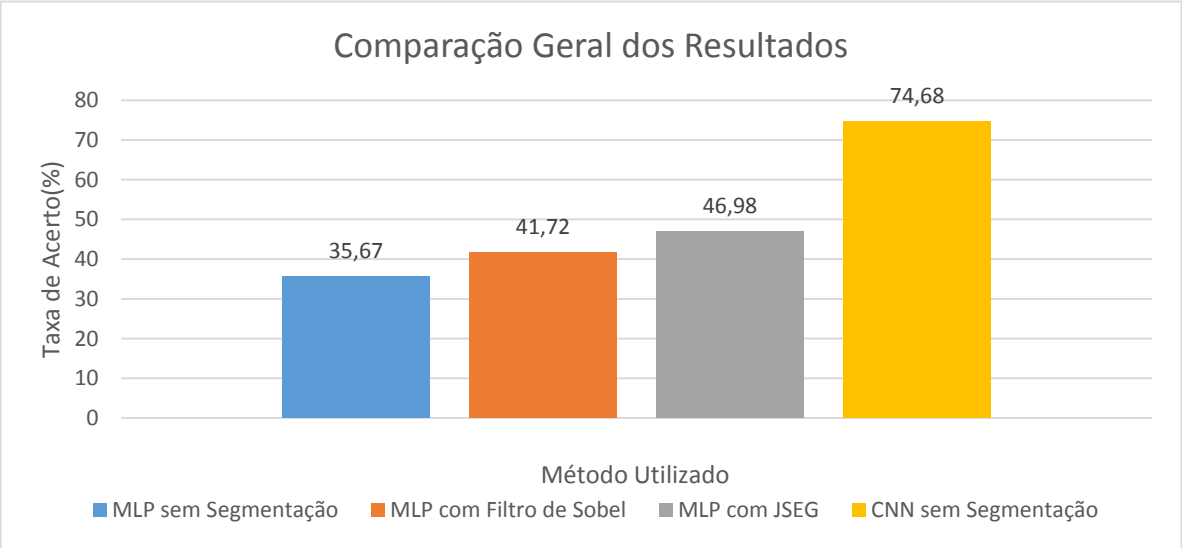
Tabela 3 - Classificação com a MLP e a CNN.

	Taxa de Acerto	**Tempo Médio de Treinamento**	**Tempo Total**
MLP utilizando JSEG	46.98 %	0d 0h 18m 50s	5d 1h 8m 50s
CNN	74,68%	2 d 0h 0m 0s	2 d 0h 0m 0s

A Tabela 3 apresenta uma comparação da classificação utilizando a *Multi Layer Perceptron* (MLP), com os dados segmentados com o JSEG, que foi o melhor resultado obtido, e o resultado do teste da *Convolucional Neural Network* (CNN). É possível notar que a CNN realizou melhor a classificação dos dados de teste. Vale ressaltar que foram utilizadas as mesmas imagens no treinamento e teste das RNA's. O tempo de treinamento da CNN é consideravelmente maior, isso ocorre pois a Rede Neural Artificial tem que aprender quais características representam melhor

os dados, enquanto a MLP recebe os dados com as características de representação destacadas. O tempo de processamento da CNN também foi influenciado pela configuração do computador. Ao se trabalhar com o Deep Learing é recomendado utilizar GPU's de alta performance para agilizar o processamento dos dados. Em comparação ao consumo de memória a CNN consumiu em média 50% da memória disponível, enquanto a MLP, 70%. Isso ocorreu pois a CNN trabalha com lotes, ou seja, carrega os dados de treinamento por partes, e a MLP trabalha com o conjunto completo.

Ainda analisando os dados da Tabela 3, a CNN apresentou 27,7% mais acertos do que a MLP, no entanto, seu tempo de treinamento foi 145 vezes maior. Vale considerar que a Rede Neural Artificial após treinada pode ser utilizada sem um novo treinamento para a classificação de qualquer outro padrão, mesmo desconhecido. Após o treinamento, a rede é armazenada em um registro, no qual os pesos são salvos, portanto, para se utilizar, basta carregar as informações na memória do computador, sendo assim, tal conhecimento pode ser utilizado para várias finalidades, como na memória visual de um robô. Ou utilizando imagens de locais, com pontos de referências, no treinamento, é possível utilizar a RNA treinada em um sistema de navegação de um carro, por exemplo, para que ele possa ser capaz de identificar o local em que se encontra. Portanto, devido a taxa de acerto



encontrada, é válido investir um maior tempo no treinamento da RNA.

Figura 19 - Comparação Geral dos Resultados.

A Figura 19 traz uma comparação geral dos resultados deste trabalho. É possível observar que a utilização de filtros trouxe uma melhora significativa ao percentual de acerto da classificação da Rede Neural Artificial MLP, e que o método de segmentação que apresentou um melhor resultado foi o JSEG, com uma taxa de acerto 11,31% maior em comparação a classificação das mesmas imagens sem a utilização de nenhum método de segmentação. O método que apresentou o melhor resultado na classificação de imagens no presente trabalho foi o CNN, que em comparação com a MLP sem a utilização de métodos de segmentação teve uma taxa de acerto superior a 40,01%.

Os resultados do presente trabalho não são comparados com outros resultados na literatura pelo motivo da diferença dos dados utilizados para a classificação. Mesmo os trabalhos que utilizam o ImageNet não seria uma comparação justa, tendo em vista que este trabalho classificou apenas 3 subconjuntos, enquanto os demais trabalham com o ImageNet por completo.

6. CONCLUSÕES

Este trabalho apresentou uma comparação de métodos de Segmentação e classificação de Imagens. Para Segmentação foi utilizado para comparação o Filtro de Sobel e o algoritmo JSEG, e para classificação foi utilizado um *Multi Layer Perceptron* e uma *Convolucional Neural Network*.

Como esperado, o trabalho atingiu seu objetivo geral, e os resultados obtidos demonstraram que o JSEG é melhor para a segmentação das imagens, no entanto exige um maior esforço computacional. E a *Convolucional Neural Network* apresentou melhores resultados se tratando de classificação de imagens.

Foi possível concluir que ao se trabalhar com os métodos de classificação de imagens que utilizam técnicas de aprendizado *shallow learning* é necessário a aplicação de técnicas para destacar as principais características dos objetos e facilitar o método na tarefa de encontrar padrões. Já em métodos de classificação de imagens que utilizam técnicas de aprendizado *Deep Learning* não é necessário a aplicação de técnicas de segmentação, pois a Rede Neural Artificial aprende durante o seu treinamento quais são as características que melhor descrevem o objeto.

Dessa forma, conclui-se que Redes Neurais Artificiais que trabalham com o *Deep Learning*, como as CNN, apresentam um melhor desempenho ao trabalhar com o aprendizado de máquina. Uma sugestão de trabalhos futuros é a aplicação da CNN para o reconhecimento de imagens utilizando o dataset completo do ImageNet, a utilização do Mapa de Kohonen ao invés da MLP e a realização do treinamento e testes em uma máquina com maior capacidade de processamento e com a disponibilidade de GPU's.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, Lívia Naiara de; VIEIRA, Tatiana Grossi Chquiloff; LACERDA, Wilian Soares; ALVES, Helena Maria Ramos; VOLPATO; Margarete Marin Lordelo; BRAGA, Ricardo César Arsillo; SOUZA, Vanessa Cristina Oliveira de. Redes neurais artificiais na classificação de áreas cafeeiras da região de Guaxupé. 2011.

AREL, Itamar; ROSE, Derek C.; KARNOWSKI, Thomas P. Deep machine learning-a new frontier in artificial intelligence research [research frontier].**Computational Intelligence Magazine, IEEE**, v. 5, n. 4, p. 13-18, 2010.

BASTOS, Igor Leonardo Oliveira; ANGELO, Michele Fúlvia. Reconhecimento de pele em imagens digitais utilizando redes neurais artificiais. 2014

BELLMAN, Richard. **An introduction to artificial intelligence: Can computers think?**. Boyd & Fraser Publishing Company, 1978.

BENGIO, Yoshua. Learning deep architectures for AI. **Foundations and trends in Machine Learning**, v. 2, n. 1, p. 1-127, 2009.

BURIOL, Tiago Martinuzzi et al. Uso de Redes Neurais Artificiais e Processamento de Imagens para Controle da Qualidade dos Dados do Radar Meteorológico do SIMEPAR. 2008.

CAVANI, F. A. Segmentaçao e classificaçao de imagens de laranjeiras utilizando JSEG e perceptron multicamadas. **Revista Minerva**, v. 3, n. 2, p. 189-197, 2006.

DA SILVA, Rosangela Westphal; SIMÕES, Priscyla Waleska Targino de Azevedo; MATTOS, Merisandra Côrtes de; CECHINEL, Cristian; RAMOS, Rozenir; LAZZARI, Rafael . Utilização da Inferência Bayesiana e Lógica Fuzzy na Modelagem de um Sistema Inteligente. 2008.

CAFFE. Disponível em: <http://caffe.berkeleyvision.org/>. Acessado em 11 de novembro de 2015.

COLLOBERT, Ronan; BENGIO, Samy; MARIÉTHOZ, Johnny. **Torch: a modular machine learning software library**. IDIAP, 2002.

DA SILVA, Marcelo Ponciano. **Processamento de consultas por similaridade em imagens médicas visando à recuperação perceptual guiada pelo usuário**. 2009. Tese de Doutorado. Instituto de Ciências Matemáticas e de Computação.

DENG, Jia . Imagenet: A large-scale hierarchical image database. In:**Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on**. IEEE, 2009. p. 248-255.

DENG, Jia. What does classifying more than 10,000 image categories tell us?. In: **Computer Vision–ECCV 2010**. Springer Berlin Heidelberg, 2010. p. 71-84.

DENG, Li; YU, Dong. Deep learning: methods and applications. **Foundations and Trends in Signal Processing**, v. 7, n. 3–4, p. 197-387, 2014.

DENG, Yining; MANJUNATH, B. S. Unsupervised segmentation of color-texture regions in images and video. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 23, n. 8, p. 800-810, 2001.

DORN, Márcio; FRISKE, Letícia Maria. Utilização de Raciocínio Subsimbólico Baseado em Redes Neurais Artificiais em Tarefas de Predição de Dados. **Anais SULCOMP**, v. 1, n. 1, 2012.

DU, Youtian et al. Recognizing interaction activities using dynamic bayesian network. In: **Pattern Recognition, 2006. ICPR 2006. 18th International Conference on**. IEEE, 2006. p. 618-621.

ERPEN, Luis Renato Cruz. Reconhecimento de padrões em imagens por descritores de forma. 2004.

ESQUEF, Israel Andrade; ALBUQUERQUE, Márcio Portes de; ALBUQUERQUE, Marcelo Portes de. Processamento Digital de Imagens.**CENTRO BRASILEIRO DE PESQUISAS FÍSICAS–CBPF**, 2003.

BASTIEN, Fredéric et al . “Theano: new features and speed improvements”. NIP deep learning workshop. 2012

GIRSHICK, Ross et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: **Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on**. IEEE, 2014. p. 580-587.

GODOI, Walmor Cardoso. DETECC AO DE DEFEITOS EM ISOLADORES POLIM ERICOS POR RADIOGRAFIA DIGITAL. 2005.

GOOGLE, Research at. Disponivel em <http://research.google.com/>. Acessado em 16 de novembro de 2015.

HUANG, F. J.; LECUN, Y. Large-scale learning with svm and convolutional netw for generic object recognition. In: **2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition**. 2006.

HAYKIN, Simon S. Redes neurais artificiais: princípio e prática. **2ª Edição, Bookman, São Paulo, Brasil**, 2000.

IBM, Research. Disponível em <http://researcher.watson.ibm.com/researcher/view.php?person=us-bedk>. Acessado em 16 de novembro de 2015.

JIA, Yangqing et al. Caffe: Convolutional architecture for fast feature embedding. In: **Proceedings of the ACM International Conference on Multimedia**. ACM, 2014. p. 675-678.

KHATCHATOURIAN, Oleg; PADILHA, Fábio RR. Reconhecimento de variedades de soja por meio do processamento de imagens digitais usando redes neurais artificiais. **Eng Agric Jaboticabal**, v. 28, n. 4, p. 759-69, 2008.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. 2012. p. 1097-1105.

KRUGER, Norbert et al. Deep hierarchies in the primate visual cortex: What can we learn for computer vision?. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 35, n. 8, p. 1847-1871, 2013.

LAWRENCE, Steve et al. Face recognition: A convolutional neural-network approach. **Neural Networks, IEEE Transactions on**, v. 8, n. 1, p. 98-113, 1997.

LECUN, Yann et al. Gradient-based learning applied to document recognition.**Proceedings of the IEEE**, v. 86, n. 11, p. 2278-2324, 1998.

LECUN, Yann; RANZATO, M. Deep learning tutorial. In: **Tutorials in International Conference on Machine Learning (ICML'13)**. 2013.

LIANG, Yulan; REYES, Michelle L.; LEE, John D. Real-time detection of driver cognitive distraction using support vector machines. **Intelligent Transportation Systems, IEEE Transactions on**, v. 8, n. 2, p. 340-350, 2007.

LIMA, Allan Reffson Granja. Máquinas de vetores suporte na classificaçao de impressoes digitais. **Universidade Federal do Ceará, Departamento de Computação, Fortaleza-Ceará**, 2002.

LIMA, S. L. et al. RNB Aplicada Ao Reconhecimento De Padrões De Falhas Em Imagens Térmicas De Equipamentos De Subestações. **III Simpósio Brasileiro de Sistemas Elétricos (SBSE/2010)**, 2010.

LOPES, Andre Teixeira; DE AGUIAR, Edilson; OLIVEIRA-SANTOS, Thiago. A Facial Expression Recognition System Using Convolutional Networks. In:**Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on**. IEEE, 2015. p. 273-280.

MARQUES FILHO, Ogê; NETO, Hugo Vieira. **Processamento digital de imagens**. Brasport, 1999.

MARQUES, Isaac R. et al. Guia de apoio à decisão em enfermagem obstétrica: aplicação da técnica da lógica fuzzy. **Rev Bras Enferm**, v. 58, n. 3, p. 349-54, 2005.

MATURANA, Patrícia Salles. Algoritmos de detecção de bordas implementados em FPGA. 2010.

MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, n. 4, p. 115-133, 1943.

MEDJAHED, Hamid et al. Human activities of daily living recognition using fuzzy logic for elderly home monitoring. In: **Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on**. IEEE, 2009. p. 2001-2006.

MICROSOFT, Research. Disponivel em <http://research.microsoft.com/en-us/people/deng/>. Acessado em 16 de novembro de 2015.

MILLER, George A. WordNet: a lexical database for English.**Communications of the ACM**, v. 38, n. 11, p. 39-41, 1995.

MORAES, Odair Barbosa de. **Método de análise de dados para avaliação de áreas urbanas recuperadas-uma abordagem utilizando a lógica fuzzy**. 2008. Tese de Doutorado. Universidade de São Paulo.

MOREIRA, édimo Fernando Alves. Redes neurais artificiais e análise discriminante linear como alternativas para seleção entre famílias de cana-de-açúcar. 2014.

MOREIRA, Fabiano Cordeiro et al. Reconhecimento e classificação de padrões de imagens de núcleos de linfócitos do sangue periférico humano com a utilização de redes neurais artificiais. 2002.

Nielsen, Michael , "Neural Networks and Deep Learning", Determination Press, 2015

NORVIG, Peter; RUSSELL, Stuart. **Inteligência Artificial, 3ª Edição**. Elsevier Brasil, 2014.

NEVES, Samuel Clayton Maciel; PELAES, Evaldo Gonçalves; DE SINAIS, Laboratório de Processamento. Estudo e implementação de técnicas de segmentação de imagens. **Revista Virtual de Iniciação Acadêmica da UFPA-Universidade Federal do Pará–Departamento de Engenharia Elétrica e de Computação**, v. 1, n. 2, 2008.

NEVES, Samuel Clayton Maciel; PELAES, Evaldo Gonçalves; D. Estudo e implementação de técnicas de segmentação de imagens. **Revista Virtual de Iniciação Acadêmica da UFPA-Universidade Federal do Pará–Departamento de Engenharia Elétrica e de Computação**, v. 1, n. 2, 2008.

NG, Andrew et al. Unsupervised feature learning and deep learning. 2013.

NOGUEIRA, A. et al. Um Overview Sobre Reconhecimento de Padrões. **II Simpósio de Excelência em Gestão e Tecnologia, Rio de Janeiro**, 2006.

NVIDIA. Disponível em <http://www.nvidia.com.br/ >. Acessado em 17 de novembro de 2015.

OQUAB, Maxime et al. Learning and transferring mid-level image representations using convolutional neural networks. In: **Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on**. IEEE, 2014. p. 1717-1724.

OSÓRIO, Fernando S.; BITTENCOURT, João Ricardo; OSÓRIO, Fernando Santos. Sistemas Inteligentes baseados em redes neurais artificiais aplicados ao processamento de imagens. In: **Workshop de Inteligência Artificial**. 2000.

PARK, Sangho; AGGARWAL, J. K. Recognition of two-person interactions using a hierarchical Bayesian network. In: **First ACM SIGMM international workshop on Video surveillance**. ACM, 2003. p. 65-76.

PEREIRA, Liliany Aparecida dos Anjos. Classificação automática de áreas cafeeiras em imagens de satélite, utilizando Redes Neurais Artificiais. 2015.

PINHEIRO, Erick Freire Pinheiro; CARVALHO, Edson Costa de Barros. Investigação da abordagem de sistemas imunológicos artificiais para reconhecimento de padrões. 2006.

PYTHON. Disponivel em: <https://www.python.org/>. Acessado em 11 de novembro de 2015.

QUIÑONEZ, Yadira et al. Autonomous Robot Navigation Based on Pattern Recognition Techniques and Artificial Neural Networks. In: **Bioinspired Computation in Artificial Systems**. Springer International Publishing, 2015. p. 320-329.

RAMOS, Vitorino; ALMEIDA, Filipe. Artificial ant colonies in digital image habitats-a mass behaviour effect study on pattern recognition. **arXiv preprint cs/0412086**, 2004.

RIBEIRO, João Henrique Burckas et al. Desenvolvimento de uma técnica de reconhecimento de padrões baseada em distância. 2003.

SAHEKI, André Hideaki. **Construção de uma rede Bayesiana aplicada ao diagnóstico de doenças cardíacas**. Tese de Doutorado. Universidade de São Paulo. 2005

SILVA, Renato Ramos da. Reconhecimento de imagens digitais utilizando redes neurais artificiais. 51 f. **Monografia (Graduação em Ciência da Computação). Universidade Federal de Lavras, Lavras–MG**. 2005.

SILVA, M; MARTINS, R. Comparação entre Multilayer Perceptron e Redes Neurais Lógicas no problema Presa-Predador usando Algoritmos Genéticos. **III SFCT**, p. 13, 2015.

SIQUEIRA, M. L. Reconhecimento Automático de Padrões em Imagens Ecocardiogáficas. **Porto Alegre: PPGC-UFRGS**, 2010.

SKNN. Disponivel em: <http://scikit-neuralnetwork.readthedocs.org/en/latest/>. Acessado em 3 de novembro de 2015.

SOARES, Heliana Bezerra. Análise e classificação de imagens de lesões da pele por atributos de cor, forma e textura utilizando máquina de vetor de suporte. 2008.

SOUZA, João Artur de et al. Reconhecimento de padrões usando indexação recursiva. 1999.

SOUZA, Lenivaldo Ribeiro. Algoritmo para Reconhecimento e Acompanhamento de Trajetória de Padrões em Vídeos. Bahia. Universidade Federal do Vale do São Francisco, 2011.

STANFORD, Deep Learning Computer Science Department. “Unsupervised Feature Learning And Deep Learning Tutorial”. Disponivel em <http://deeplearning.stanford.edu/tutorial/> acessado em 16 de novembro de 2015.

TAGLIAFERRO, Fatima Aparecida. Técnicas de reconhecimento automatizado de padrões baseados em imagens digitais. Jaguariúna. Faculdade de Jaguariúna, 2007.

ZEILER, Matthew D.; FERGUS, Rob. Visualizing and understanding convolutional networks. In: **Computer Vision–ECCV 2014**. Springer International Publishing, 2014. p. 818-833.

WANG, Chao et al. A multi-scale segmentation method of oil spills in sar images based on jseg and spectral clustering. **International Journal of Signal Processing, Image Processing and Pattern Recognition**, v. 7, n. 1, p. 425-432, 2014.

APENDICE I - Código do Filtro de Sobel

Código desenvolvido na linguagem Python para a aplicação do Filtro de Sobel nas imagens.

```
import numpy as np
import scipy as sc
from scipy import ndimage
from sys import argv
import os
# argv[1] é o caminho onde está a pasta das imagens
# argv[2] é o caminho onde irá se salvar a imagem
dirPath = argv[1]
savePath = argv[2]
filesDir = os.listdir(dirPath)
for files in filesDir:
    if files[-4:] == "JPEG":
        pathFile = dirPath + "/" + files
        im = sc.misc.imread(pathFile)
        im = im.astype('int32')
        dx = ndimage.sobel(im, 0)  # faz a derivada horizontal
        dy = ndimage.sobel(im, 1)  # faz a derivada vertical
        mag = np.hypot(dx, dy)  # magnitude
        mag *= 255.0 / np.max(mag)  # nornaliza
        sc.misc.imsave(savePath + '/' + files, mag)
```

APENDICE II - Código do MLP

Código desenvolvido na linguagem Python para o teste da RNA MLP.

```
from sys import argv, exit
import os, sys
import numpy as np
import cv2
from sknn.mlp import Classifier, Convolution, Layer
import pickle
import io

#Definindo os caminhos das variaveis para se treinar.
#dirPaths =
["/home/simiqueli/Documentos/tcc/etapa2/segmentacao_sobel/n07739125_Apple",
"/home/simiqueli/Documentos/tcc/etapa2/segmentacao_ sobel
/n07745940_strawberry", "/home/simiqueli/Documentos/tcc/etapa2/segmentacao_
sobel /n07753592_banana"]
dirPaths =
["/home/simiqueli/Documentos/tcc/etapa2/segmentacao_jseg/n07739125_Apple","
/home/simiqueli/Documentos/tcc/etapa2/segmentacao_jseg/n07745940_strawberry
","/home/simiqueli/Documentos/tcc/etapa2/segmentacao_jseg/n07753592_banana"
]


##Variaveis
x_treining = []
y_treining = []

#Leitura das imagens
print "Realizando a Leitura das Imagens"
for i in range(0, len(dirPaths)):
        path = dirPaths[i]
        print path
        dirs = os.listdir(path)
        for file in dirs:

                img = cv2.imread(path+'/'+file)
                imagem = np.array(img, dtype=np.float64) / 255.0
                w, h, d = original_shape = tuple(imagem.shape)
                if w == h:
                        image_array = np.reshape(imagem, (w * h, d))
                        x_treining.append(image_array)
                        y_treining.append(i)



print 'Converter para array'
x_treining = np.array(x_treining)
y_treining = np.array(y_treining)



print 'Embaralhando as entradas'
rand = np.random.permutation(y_treining.size)
```

```
x_treining = x_treining[rand]
y_treining = y_treining[rand]

auxTrei = int (y_treining.size * 0.8)
auxTest = y_treining.size - auxTrei

print 'Separando as amostras para teste e treino'
x_treino = x_treining[:auxTrei]
y_treino = y_treining[:auxTrei]

x_teste  = x_treining[auxTrei:]
y_teste  = y_treining[auxTrei:]

print '-', auxTrei, 'amostras para treinamento'
print '-', auxTest, 'amostras para teste'


print 'criando a rede neural'
nn = Classifier(
    layers=[
        Convolution('Rectifier', channels=6, kernel_shape=(3, 3), bor-
der_mode='full'),
        Convolution('Rectifier', channels=10, kernel_shape=(3, 3), bor-
der_mode='valid'),
        Layer('Rectifier', units=64),
        Layer('Softmax')],
    learning_rate=0.02,
    valid_size=0.2,
    n_stable=1,
     n_iter=5,
    verbose=True)

print 'Treinando a Rede Neural'
nn.fit(x_treino, y_treino)
print 'Teste:'
print('\nTEST SCORE', nn.score(x_teste, y_teste))
```