

**CENTRO UNIVERSITÁRIO CAMPO LIMPO PAULISTA  
(UNIFACCAMP)**

**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**Reconhecimento Facial com Redes Convolucionais em Imagens de  
Regiões Oculares**

**Eliana Pereira da Silva**

**Orientador: Prof. Dr. José Hiroki Saito**

**Campo Limpo Paulista – SP**

**Dezembro/2019**

**CENTRO UNIVERSITÁRIO CAMPO LIMPO PAULISTA  
(UNIFACCAMP)**

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**Reconhecimento Facial com Redes Convolucionais em Imagens de  
Regiões Oculares**

**Eliana Pereira da Silva**

**Orientador: Prof. Dr. José Hiroki Saito**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro Universitário Campo Limpo Paulista como parte dos requisitos para a obtenção do título de Mestra em Ciência da Computação. Orientador: Dr. José Hiroki Saito

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

**Campo Limpo Paulista – SP**

**Dezembro/2019**

## FICHA CATALOGRÁFICA

Ficha catalográfica elaborada pela  
Biblioteca Central da Unifaccam

S579r

Silva, Eliana Pereira da

Reconhecimento facial com redes convolucionais em  
imagens de regiões oculares / Eliana Pereira da Silva. Campo  
Limpo Paulista, SP: Unifaccamp, 2019.

Orientador: Prof<sup>o</sup>. Dr. José Hiroki Saito

Dissertação (Programa de Mestrado em Ciência da  
Computação) – Centro Universitário Campo Limpo Paulista –  
Unifaccamp.

1. Reconhecimento facial. 2. Região ocular. 3. Rede neural. 4.  
*Neocognitron*. 5. *Deep Learning*. 6. CNN. 7. LeNet. 8. AlexNet.  
I. Saito, José Hiroki. II. Campo Limpo Paulista. III. Título.

CDD – 006.32

## ***Resumo***

A proposta deste trabalho é realizar o reconhecimento facial de pessoas utilizando redes que recebem como entrada imagens contendo somente a região ocular. Entre as várias arquiteturas de Deep Learning, redes neurais convolucionais são adequadas para o reconhecimento de padrões em imagens, pois levam em consideração a localização espacial dos atributos. São considerados três exemplos de redes: Neocognitron, LeNet e AlexNet. Na fase preliminar do trabalho foi utilizada a base de dados pré-processada da Universidade Essex com 395 imagens entre homens e mulheres, sendo 20 imagens por indivíduo. Visando atingir resultados mais expressivos, novos experimentos foram realizados usando o banco de dados de rosto CMU-PIE da Universidade de Carnegie Mellon que contém 41.368 imagens de 68 pessoas. Foram realizados comparativos entre os resultados obtidos no processo de reconhecimento facial utilizando como padrão de entrada imagens completas da face e imagens da região ocular da face. Foi realizado um estudo comparativo de desempenho entre as três redes neurais consideradas.

***Palavras-chave:*** reconhecimento facial, região ocular, rede neural, Neocognitron, Deep Learning, CNN, LeNet, AlexNet.

### *Abstract*

The purpose of this paper is to perform facial recognition of people using networks that receive as input images containing only the ocular region. Among the various Deep Learning architectures, convolutional neural networks are suitable for pattern recognition in images because they take into account the spatial location of attributes. Three examples of networks are considered: Neocognitron, LeNet and AlexNet. In the preliminary phase of the work, the pre-processed Essex University database was used, with 395 images between men and women, 20 images per individual. In order to achieve more expressive results, new experiments were performed using the Carnegie Mellon University CMU-PIE face database containing 41,368 images of 68 people. Comparisons were made between the results obtained in the facial recognition process using as input standard complete images of the face and images of the ocular region of the face. A comparative performance study was performed between the three neural networks considered.

**Keywords:** facial recognition, ocular region, neural network, Neocognitron, Deep Learning, CNN, LeNet, AlexNet.

# Sumário

<b>CAPÍTULO 1</b> .....	<b>5</b>
1.1 Introdução .....	5
1.2 Objetivos gerais.....	7
1.3 Principais contribuições.....	7
1.4 Organização do trabalho .....	7
<b>CAPÍTULO 2 - Redes Neurais</b> .....	<b>9</b>
2.1 O neurônio biológico .....	9
2.2 Neurônios e Redes neurais artificiais .....	10
2.2.1 Estrutura das redes neurais artificiais .....	12
2.2.2 Redes Progressivas .....	12
2.2.3 <i>Deep Learning</i> .....	13
2.2.4 Redes Neurais Convolucionais .....	14
2.3 Trabalhos correlatos de reconhecimentos faciais usando redes convolucionais .....	17
<b>CAPÍTULO 3 – Redes Convolucionais Neocognitron, LeNet e AlexNet</b> .....	<b>20</b>
3.1 Rede neural Neocognitron.....	20
3.1.1 Estrutura da rede neural Neocognitron.....	20
3.1.6 Treinamento da rede neural Neocognitron .....	29
3.2 LeNet .....	37
3.2.1 Treinamento do LeNet.....	40
3.3 AlexNet .....	42
3.3.2 Treinamento do AlexNet.....	45
<b>CAPÍTULO 4 - Descrição do Trabalho e Metodologia</b> .....	<b>46</b>
4.1 Tipos de Experimentos .....	46
4.2 Base de Imagens.....	47
4.2.1 Base de imagens 1 .....	47
4.2.2 Base de imagens 2 .....	48
4.2.3 Pré-processamento para a Obtenção de Imagens da Região Ocular .....	48
4.3 Simulador do Neocognitron.....	49
4.4 Simulador da Rede LeNet .....	52
4.5 Simulador computacional AlexNet.....	52
4.6 Métodos para a Validação dos Resultados .....	53

4.6.1 Coeficiente Kappa.....	53
4.6.2 Outros Métodos de Validação.....	54
4.7 Considerações Finais .....	55
<b>CAPÍTULO 5 – Experimentos Realizados e Resultados.....</b>	<b>56</b>
5.1 Experimentos com o Neocognitron Usando a Base de Imagens 1.....	57
5.2 Experimentos com LeNet-5 usando Base de Imagens 1.....	62
5.3 Experimentos com AlexNet e Base de Imagens 1 .....	66
5.4 Experimentos com Neocognitron e Base de imagens 2 .....	69
5.5 Experimentos com LeNet para Base de Imagens 2 .....	72
5.6 Experimentos com AlexNet usando Base de Imagens 2.....	74
5.7 Análise dos Resultados .....	76
5.8 Considerações Finais .....	82
<b>CAPÍTULO 6 - Conclusão .....</b>	<b>84</b>
6.1 Propostas de Trabalhos Futuros .....	85
6.2: Artigos Publicados ou Submetidos.....	85
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>87</b>

## **Glossário**

CNN – redes neurais convolucionais

LGN - núcleo geniculado lateral

MLP – Multi Layer Perceptron

RNDPC - Rede Neural para Detecção de Pontos de Controle

SSP - Seed Selection Plain

UFSCar- Universidade Federal de São Carlos

RNRF – Rede Neural para Reconhecimento Facial

SGD - Stochastic Gradient Descent

## **Lista de Tabelas**

Tabela 1 - Estrutura da rede AlexNet

Tabela 2 – Significado dos valores do coeficiente Kappa

Tabela 3. Síntese dos experimentos com o Neocognitron com amostras de treinamento coincidentes com as amostras de teste usando Base de Imagens 1

Tabela 4. Síntese dos experimentos com o Neocognitron com amostras de treinamento distintas das amostras de teste usando Base de Imagens 1

Tabela 5. Experimentos com LeNet-5 com mesmas amostras de treinamento e teste

Tabela 6. Experimentos com o LeNet-5 com amostras de treinamento e teste distintas

Tabela 7. Síntese dos experimentos com o AlexNet

Tabela 8. Síntese dos experimentos com o AlexNet

Tabela 9- Base de imagens 2 - Síntese dos experimentos com o Neocognitron

Tabela 10 - Base de imagens 2 - Síntese dos experimentos com o LeNet-5

Tabela 11 – Base de imagens 2 - Síntese dos experimentos com AlexNet

Tabela 12. Neocognitron e Base de Imagens 1

Tabela 13. LeNet e Base de Imagens 1

Tabela 14. AlexNet e Base de Imagens 1

Tabela 15- Banco de imagens 2 - Síntese dos experimentos com o Neocognitron

Tabela 16 - Base de imagens 2 - Síntese dos experimentos com o LeNet-5

Tabela 17 – Base de imagens 2 - Síntese dos experimentos com AlexNet

Tabela 18 – Base de imagens 1- Melhores resultados das redes

Tabela 19 – Base de imagens 2- Melhores resultados das redes

## Lista de Figuras

Figura 1 – Interconexões entre neurônios biológicos

Figura 2 - (a) Neurônio artificial: calcula soma ponderada de  $n$  entradas e passa o resultado através de uma função de transferência / ativação (b) quatro principais funções de ativação.

Figura 3 – Estrutura de uma rede neural progressiva.

Figura 4 - Ilustração da arquitetura CNN e suas camadas: convolucionais, de pooling e totalmente conectadas.

Figura 5- Exemplo de aplicação da convolução na imagem: (a) à esquerda, uma imagem de entrada de dimensão  $8 \times 8$  que recebe a aplicação de um filtro  $3 \times 3$  e à direita, ilustra-se o somatório com o resultado da convolução. (b) à esquerda, a imagem de entrada e à direita, o resultado da convolução, destacando-se o resultado do cálculo feito em (a).

Figura 6- Estrutura da rede Neocognitron: (a) adaptado de Fukushima, 1988, (b) em formato genérico de redes convolucionais.

Figura 7- Fatores reconhecidos pelos neurônios nos 3 estágios da rede.

Figura 8. Ilustração das conexões de entrada de uma célula-S.

Figura 9- Ilustração das conexões de entrada e saída da célula vc.

Figura 10- Estrutura da rede neocognitron com camadas de células-S, células-C e células-V.

Figura 11. Ilustração das conexões de entrada das células-C.

Figura 12. Exemplos de detecção de bordas em linhas horizontal e inclinada.

Figura 13- Padrões de bordas divididos em dois lados (A e B) de uma linha.

Figura 14- Parâmetros usados para a detecção de linhas com espessuras diferentes.

Figura 15- Espessura e direção das linhas variando os parâmetros  $i_x$ ,  $i_{xm}$ ,  $\Delta x$  e  $\Delta y$

Figura 16- Detecção de pontos extremos

Figura 17 - Estrutura da rede neural convolucional LeNet. Fonte: <http://yann.lecun.com>

Figura 18- Combinação das matrizes de características da camada S2 com as matrizes da camada C3.

Figura 19 – Estrutura da rede CNN AlexNet

Figura 20- Exemplos de imagens de regiões oculares: (a) base de imagens 1, região ocular incluindo ambos os olhos e (b) base de imagens 2, região ocular exclusivo do lado esquerdo.

Figura 21 – Ilustração da interface do simulador durante a detecção. Adaptado pela autora

Figura 22 – Interface do reconhecimento de padrão usando RNN. Adaptado pela autora.

## CAPÍTULO 1

No presente capítulo é feita uma introdução do Trabalho de Dissertação, iniciando-se com uma contextualização do tema, seguida da descrição dos objetivos e descrição dos capítulos.

### 1.1 Introdução

É próprio do ser humano reconhecer objetos, lugares e pessoas. Podemos realizar esse processo de reconhecimento sob as mais variadas condições de observação tais como em imagens distorcidas ou deformadas, com pouca luminosidade ou observando apenas partes da face. Nossa habilidade natural de reconhecimento é capaz de tratar estas variações presentes na face com facilidade. No entanto, não é trivial implementar esse processo em uma máquina (Agarwal *et al.*, 2010).

Diversas pesquisas utilizando técnicas de reconhecimento facial têm sido desenvolvidas há muitos anos, e dentre as muitas encontradas na literatura destacamos: (Agarwal *et al.*, 2010), que utilizou *Principal Component Analysis* (PCA), (Gorokhovatskyi, 2016), que utilizou redes neurais e (Bianchini, 2001), (Saito *et al.*, 2005), (Poli *et al.*, 2008), que utilizaram a rede neural Neocognitron na tarefa de reconhecer faces. Contudo, apesar das décadas de pesquisas sobre reconhecimento facial, vários desafios persistem, tais como: problemas de deficiência na iluminação, variações de posição do rosto, oclusão (Elaiwat *et al.*, 2015) e alteração da imagem devido às expressões faciais (Dalapicola *et al.*, 2019).

Várias são as abordagens que podem ser utilizadas na tarefa de reconhecimento facial. A abordagem usando redes neurais artificiais, que será discutida nesse trabalho, tem sido aplicada com sucesso, uma vez que, as redes neurais têm capacidade de extrair

as características que maximizam o seu poder de classificação e ainda permitem usar a sua capacidade de generalização, dentre outras habilidades (Bengio e LeCun, 2007). Segundo Haykin (2001), redes neurais são estruturas computacionais que procuram simular ou reproduzir em pequena escala o funcionamento do cérebro humano. A técnica de *Deep Learning*, denominada rede neural convolucional, têm se apresentado como uma boa alternativa nos mais diferentes problemas de visão computacional, como por exemplo, reconhecimento de caracteres (Fukushima, 1980), reconhecimento facial (Saito *et al.*, 2005), e esta popularidade está associada ao fato de possuírem a capacidade de ser tolerante a distorções dos padrões de entrada.

Neocognitron é uma rede neural convolucional desenvolvida inicialmente para o reconhecimento de dígitos manuscritos, no entanto pode ser aplicada também para o reconhecimento facial (Saito *et al.*, 2005). O presente trabalho pretende investigar de forma comparativa o uso das redes neurais convolucionais Neocognitron, LeNet-5 e AlexNet, no reconhecimento facial utilizando como padrão de entrada imagens da região ocular da face. A escolha dessas três arquiteturas de redes convolucionais está baseada no fato das mesmas possuírem histórico de alta acurácia para tarefas de classificação, porém em contextos diferentes. Além disso, o orientador desta dissertação tem desenvolvido trabalhos com a rede Neocognitron, incluindo um estágio de pós-doutorado no Japão, realizado na Universidade de Osaka, junto ao Laboratório do Prof. Kunihiko Fukushima, criador da rede Neocognitron. O orientador, tem realizado trabalhos sobre o reconhecimento de imagens de regiões oculares de face (Barcellos *et al.*, 2019) (Dalapicola *et al.*, 2019) e em contextos diferentes dos propostos para esta dissertação.

## 1.2 Objetivos gerais

O objetivo desta dissertação de mestrado é o estudo da arquitetura de *Deep Learning*, rede neural convolucional, especificamente Neocognitron, LeNet-5 e AlexNet no processo de reconhecimento de imagens parciais da face, somente da região ocular.

Como objetivos específicos temos:

- 1) estudo comparativo de redes neurais convolucionais Neocognitron, LeNet-5 e AlexNet;
- 2) estudo da possibilidade de reconhecimento facial levando-se em conta imagens da região ocular da face; e
- 3) estudo comparativo entre o reconhecimento facial levando-se em conta apenas as imagens da região ocular e o reconhecimento facial levando-se em conta as imagens faciais completas.

## 1.3 Principais contribuições

Como principais contribuições da presente dissertação podem ser citadas as descrições estruturais das redes convolucionais de uma forma genérica; a verificação de como o número de amostras, número de classes e tipos de imagens influenciam no desempenho das redes convolucionais; e a proposta de verificação do número de épocas e seleção de amostras de treinamento, que melhoram o desempenho dessas redes.

## 1.4 Organização do trabalho

A presente dissertação de mestrado, após este capítulo introdutório, prossegue com os seguintes capítulos. O Capítulo 2 inicia-se com a contextualização sobre redes neurais estendendo-se até aos conceitos de redes convolucionais e trabalhos correlatos.

No Capítulo 3 serão descritas as três redes neurais convolucionais, Neocognitron, LeNet e AlexNet, utilizadas na pesquisa.

No Capítulo 4 é descrita a metodologia usada no desenvolvimento deste trabalho. No Capítulo 5 são apresentados os experimentos realizados com os simuladores computacionais baseados na rede neural Neocognitron, LeNet e AlexNet. Por fim, conclusão e trabalhos futuros são descritos no Capítulo 6.

O trabalho encerra com as referências bibliográficas, e os Apêndices que mostram as imagens utilizadas nos experimentos.

## **CAPÍTULO 2 - Redes Neurais**

Este capítulo apresenta alguns conceitos básicos referentes ao funcionamento do cérebro humano, diga-se dos neurônios biológicos, seguidos da descrição de alguns aspectos de redes neurais artificiais. Em seguida são apresentados os conceitos envolvidos nas redes neurais convolucionais que serão abordados na pesquisa, concluindo-se com algumas descrições principais de trabalhos relacionados ao reconhecimento de faces usando essas redes.

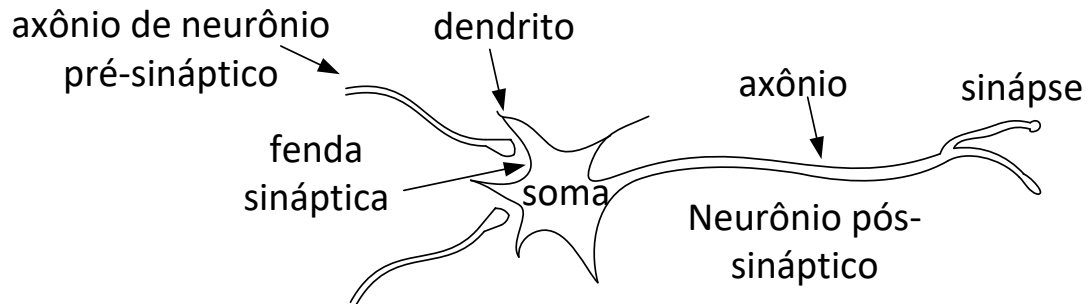
### **2.1 O neurônio biológico**

Os neurônios biológicos são unidades essenciais do cérebro de animais, com a função elementar de receber, processar e enviar informações.

O neurônio é formado por 3 partes essenciais: a soma ou corpo celular, do qual derivam algumas ramificações denominadas de dendritos, e por uma outra ramificação descendente da soma, porém mais longa, chamada de axônio (Arbib, 1995). Nas pontas dos axônios estão os terminais sinápticos onde se situam as vesículas, as quais contêm neurotransmissores que são substâncias que agem sobre os canais iônicos situados nos dendritos a fim de permitir a passagem de íons para o interior da célula pós-sináptica, seletivamente (Haykin, 2001).

O processo de comunicação entre os neurônios se dá através da sinapse. É por meio dela que ocorre a passagem do sinal neural de um neurônio para outro, denominada de transmissão sináptica. Em uma sinapse os neurônios não se tocam, há um espaço entre eles denominado de fenda sináptica, onde o neurônio pré-sináptico se comunica com o outro denominado neurônio pós-sináptico. É pelas sinapses que os neurônios se unem

funcionalmente e formam as redes neurais. As sinapses são muito variadas em suas formas e características: algumas são inibidoras e algumas excitadoras (Arbib, 1995). A Figura 1 ilustra interconexões entre neurônios pré-sináptico e pós-sináptico.



**Figura 1 – Interconexões entre neurônios biológicos**

## 2.2 Neurônios e Redes neurais artificiais

Neurônios artificiais são modelos matemáticos que simulam o funcionamento do neurônio biológico e se organizam em camadas com conexões programáveis, quando formam uma estrutura conhecida como rede neural artificial. Em 1943 McCulloch e Pitts elaboraram o primeiro modelo de neurônio artificial e o mesmo obteve aceitação favorável na comunidade científica (McCulloch e Pitts, 1943).

A estrutura de um neurônio artificial é formada por sinais de entrada, pesos sinápticos, função de ativação e as saídas. Ainda pode existir na camada de entrada e nas camadas ocultas, que são camadas intermediárias entre as camadas de entrada e de saída, um parâmetro adicional chamado de viés ou polarização (bias).

Por meio da camada de entrada os dados são disponibilizados para a rede. Os pesos sinápticos são definidos nas camadas ocultas. Por fim os resultados são exibidos na camada de saída. Uma característica importante sobre os pesos é que os mesmos podem assumir valores positivos ou negativos.

E por fim, cabe à função de transferência ou ativação, por exemplo, logística ou sigmoide (1), *Rectified Linear Unit* ou ReLU (2), tangente hiperbólica (3), degrau (4), definir se um neurônio será ativado ou não (Steck, 1992; Tissot, 2012; Fan *et al.*, 2015; Mikaeil *et al.*, 2018). A seguir são mostradas as equações 1, 2, 3 e 4, que representam as respectivas funções de ativação, onde  $f(u)$  é a saída do neurônio e  $u$  é a soma ponderada das entradas e o viés, com os respectivos pesos:

Sigmoide: 
$$f(u) = \frac{1}{1+e^{-u}} \quad (1)$$

ReLU: 
$$f(u) = \begin{cases} u, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0 \end{cases} \quad (2)$$

Tangente Hiperbólica: 
$$f(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} \quad (3)$$

Degrau: 
$$f(u) = \begin{cases} 1 & \text{se } u \geq 0 \\ 0, & \text{se } u < 0. \end{cases} \quad (4)$$

A Figura 2 ilustra a estrutura de um neurônio artificial.

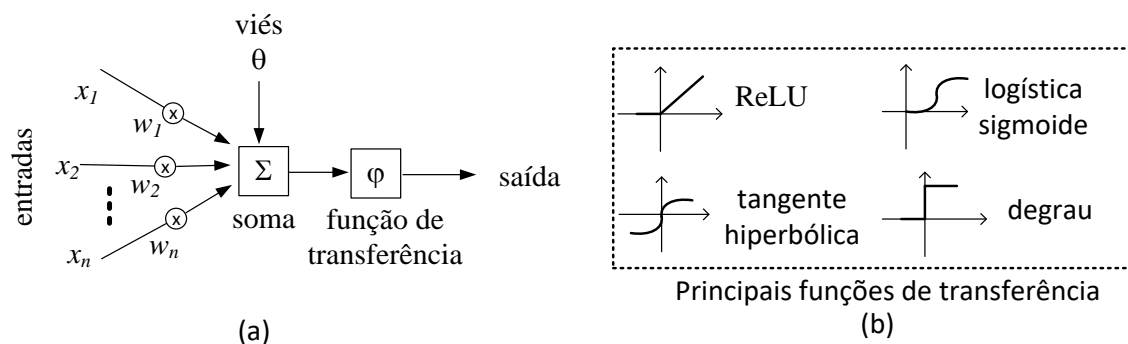


Figura 2 - (a) Neurônio artificial: calcula a soma ponderada de  $n$  entradas e passa o resultado por uma função de transferência / ativação (b) quatro principais funções de ativação.

A seguir, procura-se descrever a parte conceitual sobre redes neurais, de modo particular as redes convolucionais, afim de fornecer os requisitos necessários para a compreensão das redes neurais utilizadas no trabalho proposto.

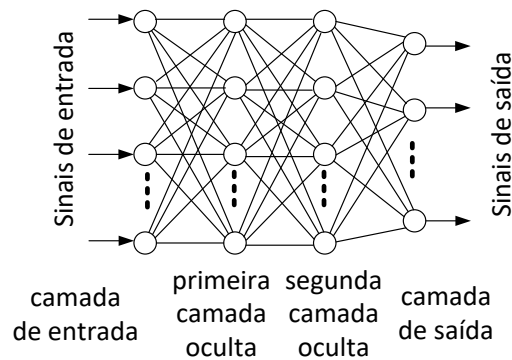
### 2.2.1 Estrutura das redes neurais artificiais

A estrutura de uma rede neural está diretamente relacionada à forma como seus elementos estão organizados, como por exemplo, o número de camadas ocultas ou intermediárias entre a camada de entrada e a camada de saída; o número de unidades por camada; a função de ativação usada para a obtenção da saída dos neurônios; os pesos e as conexões. Enfim todos esses elementos dizem respeito à estrutura da rede (Miyazaki e Yamada, 1993; Fan *et al.*, 2015; Komyakov *et al.*, 2016; Ertugrul *et al.*, 2017; Mikaeil *et al.*, 2018).

Uma característica relevante em uma modalidade de redes neurais artificiais é o fluxo de informação acíclico, conforme descrito na seção a seguir.

### 2.2.2 Redes Progressivas

Na estrutura das redes neurais progressivas multicamadas o fluxo de informação é acíclico, os neurônios da camada anterior funcionam como padrão de entrada para a camada seguinte, ou seja, não há laços. Essa arquitetura possibilita uma melhor qualidade do treinamento. Além das camadas de entrada e saída, as redes progressivas possuem camadas ocultas ou intermediárias destinadas a cálculos intermediários que auxiliam a rede a encontrar os valores finais. Estudos comprovam que as redes neurais progressivas com várias camadas ocultas podem melhorar significativamente a capacidade de classificação das redes neurais, com o aumento do número dessas camadas (Guo e Liu, 2017). Também foi observado que a retirada de uma única camada reduz a performance da rede, mostrando que a profundidade da rede é determinante para o alcance dos resultados. As redes neurais progressivas com muitas camadas ocultas são chamadas de redes de aprendizado profundo ou *Deep Learning*. A Figura 3 ilustra a estrutura de uma rede neural progressiva.



**Figura 3 – Estrutura de uma rede neural progressiva.**

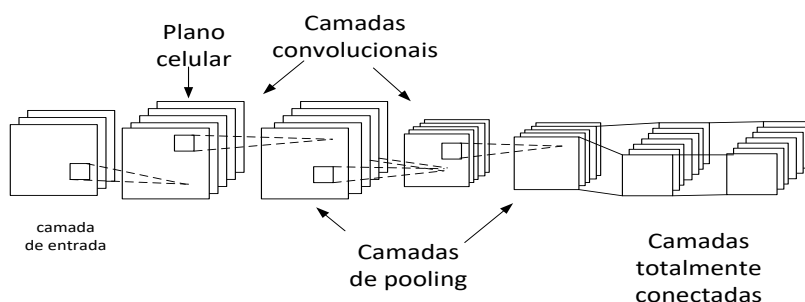
### 2.2.3 Deep Learning

*Deep Learning* pode ser definido como uma classe de técnicas de aprendizado de máquina que exploram muitas camadas de processamento de informação linear e não linear, esse processo pode ser supervisionado ou não supervisionado, para extração e análise de padrões e classificação. As técnicas de *Deep Learning* estão sendo aplicadas com sucesso em diversas áreas, tais como: Visão computacional, reconhecimento facial, reconhecimento automático de fala, processamento de linguagem natural além disso, tem superado outras técnicas de aprendizado de máquina em prever a atividade de potenciais drogas moleculares, análise de dados de acelerador de partículas e reconstrução de circuitos cerebrais (LeCun *et. al*, 2015)

Entre os diversos métodos de *Deep Learning* destaca-se as redes neurais convolucionais especialmente projetada para lidar com a variabilidade em dados bidimensionais (2D), como por exemplo, imagens em sua forma matricial (LeCun et al., 1998). Essas redes podem possuir uma arquitetura chamada de convolucional em que os neurônios se conectam a uma certa quantidade de neurônios da camada anterior, mas não a todos os neurônios. As redes neurais convolucionais, representam atualmente o estado da arte no reconhecimento de imagens e objetos (LeCun et al., 2015).

## 2.2.4 Redes Neurais Convolucionais

Redes Neurais Convolucionais (em inglês, *Convolutional Neural Network* - CNN ou ConvNet) são formadas por camada de entrada, camadas de convolução, camadas de *pooling*, camadas totalmente conectadas e camada de saída, e cada uma destas camadas possui uma função específica na propagação do sinal de entrada. A Figura 4 mostra a estrutura de uma rede convolucional. Cada uma dessas camadas é constituída de vários planos celulares, e em cada plano celular, todos os neurônios têm as mesmas características, ou seja, realizam uma mesma função, porém em locais diferentes da imagem.



**Figura 4 - Arquitetura CNN e suas camadas: convolucionais, de *pooling* e totalmente conectadas.**

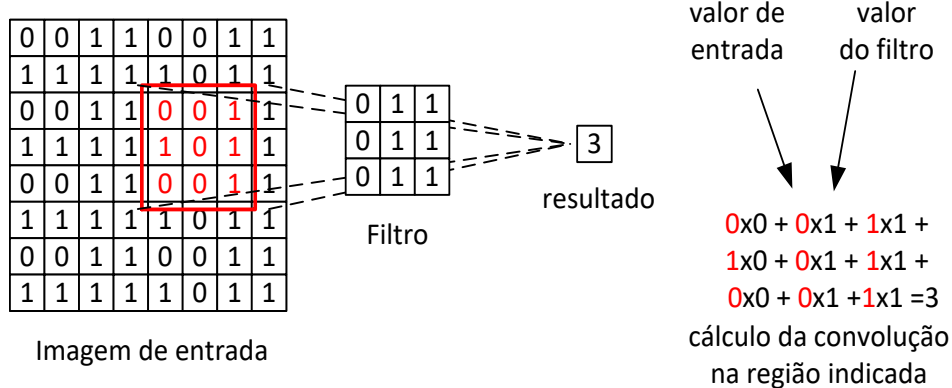
As CNN's são usadas atualmente para solucionar diversos tipos de problemas, com grande ênfase para aplicações que lidam com reconhecimento de padrões em imagens, como por exemplo, reconhecimento de caracteres manuscritos, classificação e detecção de vários tipos de imagens (LeCun *et al.*, 1998; Hayat *et al.*, 2018).

As CNN's fornecem vantagens como por exemplo: capacidade de extrair características relevantes necessitando de uma quantidade menor de parâmetros e ou pesos de ajustes em comparação com as redes totalmente conectadas. Como uma unidade de uma camada não é conectada com todas as unidades da camada anterior, em CNN há menos pesos para serem atualizados, facilitando assim o treinamento (Huang e LeCun,

2006). A rede Neocognitron estudada nesse trabalho é um exemplo de CNN. As redes LeNet e AlexNet são evoluções da rede neural Neocognitron. No Capítulo 3 veremos em detalhes os funcionamentos dessas redes.

### 2.2.3.1 Convolução

Convolução é uma operação matemática entre duas funções  $f$  e  $g$ , produzindo uma terceira função, que pode ser interpretada como uma função alterada de  $f$ . No processamento de imagens, a convolução é útil para detecção de bordas, suavização de imagem, extração de atributos, entre outras aplicações (Huang e LeCun, 2006; Abdel-Hamid *et al.*, 2014). A Figura 5 é uma ilustração de um processo de convolução em que numa imagem de entrada  $8 \times 8$  é aplicada a convolução usando um filtro  $3 \times 3$ . Na Figura 5(a) ilustra-se o procedimento de cálculo, onde cada elemento de uma região da imagem de entrada, é ponderado por um filtro  $3 \times 3$  para a obtenção do resultado, e na Figura 5(b) é mostrado o resultado total da convolução. Nota-se que devido ao fato de se aplicar o filtro sobre uma área preenchida totalmente com os elementos da imagem de entrada, sem considerar as bordas, o resultado da convolução tem tamanho  $6 \times 6$ .



(a)



(b)

**Figura 5 - Exemplo de aplicação da convolução na imagem: (a) à esquerda, uma imagem de entrada de dimensão 8x8 que recebe a aplicação de um filtro 3x3 e à direita, ilustra-se o somatório com o resultado da convolução. (b) à esquerda, a imagem de entrada e à direita, o resultado da convolução, destacando-se o resultado do cálculo feito em (a).**

### 2.2.3.2. Camada convolucional

Numa camada convolucional, a imagem de entrada é percorrida por filtros, que enxergam pequenos quadros na imagem e ao percorrer as mesmas capturam os traços mais relevantes. Os filtros também chamados *kernels*, são matrizes de pesos, que em geral são inicializados aleatoriamente. A pequena região onde o filtro é aplicado é chamada de campo receptivo e a aplicação do filtro resulta num valor que deve compor o mapa de características, que é a matriz resultante da convolução (Johnson e Zhang, 2015; Shen e Wang, 2018).

### 2.2.3.3 Mapa de características

Cada representação ou imagem gerada por um filtro da camada convolucional é conhecida como mapa de características, e os mesmos são empilhados formando um tensor. O tamanho do mapa de características é controlado por três parâmetros: profundidade (*depth*), passo (*stride*) e preenchimento (*padding*) (Arel *et al.*, 2010; Cardoso e Wichert, 2010; Abdel-Hamid *et al.*, 2014; Le *et al.*, 2015). Numa camada

convolucional atuam vários filtros convolucionais. Cada filtro é responsável por uma determinada característica e, portanto, resulta num mapa de características diferente.

#### **2.2.3.4 Camadas de pooling**

De acordo com Abdel-Hamid *et al.*, (2014) as ConvNets são organizadas em pares formados por uma camada de convolução e uma camada de *pooling*. A camada de *pooling* pode ser definida como uma forma de reduzir o tamanho da dimensão da imagem e em geral é aplicada sobre o resultado da camada de convolução. Uma das possibilidades de realização dessa redução é pelo uso do método denotado como *max-pooling*, que consiste na escolha do valor máximo que consta na região considerada (Arel *et al.*, 2010; Abdel-Hamid *et al.*, 2014).

#### **2.2.3.5 Camadas totalmente conectadas**

A camada totalmente conectada junta as entradas de todas as posições para realizar a classificação global das entradas. A informação de cada posição é obtida pela saída de um filtro convolucional. As saídas de todas as posições convergem para a camada totalmente conectada, em direção à saída da rede, objetivando encerrar a classificação (Li *et al.*, 2017; Zhang *et al.*, 2018).

### **2.3 Trabalhos correlatos de reconhecimentos faciais usando redes convolucionais**

O uso de redes convolucionais no reconhecimento facial demonstrou que essa é uma técnica robusta, principalmente quando existe uma variedade em pose, expressão e iluminação, em banco de dados com uma quantidade grande de indivíduos, como a LFW (*Labeled Face in the Wild*) (Huang *et al.*, 2007).

Um dos trabalhos envolvendo *deep learning* foi descrito por Yi *et al.*, (2014), que propõe uma forma semiautomática para coletar imagens faciais da Internet, coletando

imagens de 10.000 indivíduos em 500.000 imagens, chamado *CASIA-WebFace*. Tendo em vista esse banco de dados, foi desenvolvida uma rede convolucional de onze camadas para o reconhecimento de faces, obtendo uma precisão de 97% a 99%.

O trabalho de Taigman *et al.* (2014) descreve o reconhecimento em quatro etapas: detecção, alinhamento, representação e classificação. Para a classificação ou reconhecimento, é usada uma rede *deep learning* com nove camadas, envolvendo mais de 120 milhões de parâmetros, usando várias camadas localmente conectadas, sem o compartilhamento de pesos como em redes convolucionais padrões. A rede foi treinada usando um banco de dados contendo quatro milhões de imagens faciais pertencendo a mais de quatro mil indivíduos, e atingiu uma precisão de 97,35% nas faces rotuladas do LFW.

Sun *et al.* (2014) descreve o uso do *deep learning* para a identificação e verificação. A identificação consiste no treinamento variando as diversas partes ou formas de posicionamento dos indivíduos, e a verificação é a etapa do uso do resultado desse treinamento para a identificação dos indivíduos. Segundo os autores, o resultado teve uma precisão de 99.15% na verificação para o banco LFW.

O trabalho de Zhou *et al.* (2015) relata um sistema de reconhecimento facial usando *deep learning* com uma estrutura simples, que teve uma precisão de 99.50% para o banco denotado *Megvii Face Classification* (MFC), contendo cinco milhões de faces rotuladas de aproximadamente vinte milhões de indivíduos.

Esses e outros trabalhos recentes descrevem a alta precisão encontrada nos reconhecimentos de faces usando grandes bancos de dados, com uma quantidade grande de indivíduos. O trabalho proposto visa explorar o uso de redes convolucionais aplicando

o reconhecimento de faces a partir das regiões oculares, porém usando bancos de dados menores, diferentes do tipo LFW.

## **2.5 Considerações Finais**

Neste capítulo foram descritos os fundamentos de redes neurais biológicas, artificiais e redes neurais convolucionais. Foram também descritos alguns trabalhos correlatos na área de reconhecimento de faces usando *deep learning*, mostrando a alta precisão obtida.

## **CAPÍTULO 3 – Redes Convolucionais Neocognitron, LeNet e AlexNet**

Nesse capítulo serão descritas as redes convolucionais Neocognitron, LeNet e AlexNet, objetos de estudo deste trabalho, bem como as principais características destas redes, diga-se, estrutura e treinamento.

### **3.1 Rede neural Neocognitron**

Neocognitron foi proposta por Fukushima para fazer reconhecimento de dígitos manuscritos (Fukushima, 1980). Essa rede consiste de camadas de células simples e células complexas, conceito descrito por Hubel e Wiesel (1977).

As células são modificáveis e crescem gradualmente de acordo com a quantidade de amostras de treinamento que são apresentadas repetidamente à rede. Durante o treinamento, as células da rede tornam-se seletivas aos atributos morfológicos contidos nessas amostras (Fukushima, 2016). As seções subsequentes descrevem em detalhes o funcionamento desta rede.

#### **3.1.1 Estrutura da rede neural Neocognitron**

Neocognitron é formado por vários estágios hierarquicamente posicionados cada qual com duas camadas de células:  $U_S$  de células-S (onde a letra S é originada de células simples) responsáveis pela extração de atributos, e  $U_C$  de células-C (onde C é originada de células complexas) que aplica o filtro da média correspondente à generalização (White e Elmasry, 1992; Shi e Tan, 2000).

Cada camada é composta de um conjunto de planos celulares, sendo que em cada plano, todas as suas células têm conexões de entrada com uma pequena região da camada

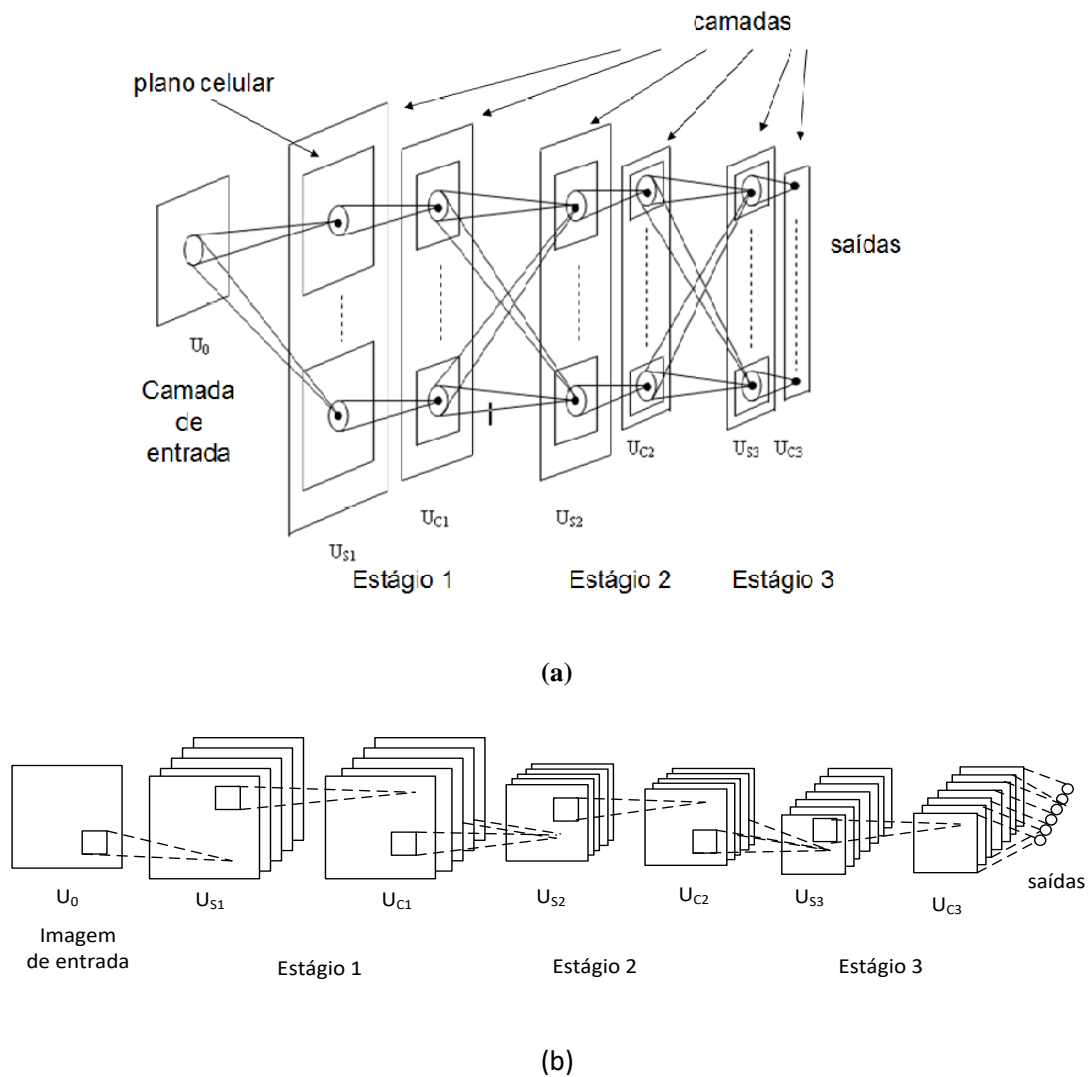
anterior e os respectivos pesos são iguais, ou seja, todas as células fazem o papel de um mesmo tipo de filtro, porém, cada uma atuando em posição ou campo receptivo diferente.

A entrada da rede é denominada de camada de entrada  $U_0$ , que corresponde às células fotorreceptoras da retina, na visão biológica; e a saída é denominada de camada de saída, onde cada saída corresponde a uma única célula, que representa uma classe.

A Figura 6(a) ilustra a estrutura da rede Neocognitron em diagrama original e a Figura 6(b), mostra a rede em forma de redes convolucionais. Na sequência veremos em detalhes essa estrutura.

### 3.1.2 Estágios

Os estágios da rede Neocognitron são responsáveis pelo processo de extração de atributos e em cada estágio existe um certo nível de complexidade da característica extraída da imagem de entrada. No estágio zero, que corresponde à camada de entrada,  $U_0$ , é apresentada a imagem a ser processada pela rede. As características mais simples como bordas e linhas em direções variadas são extraídas pelos estágios iniciais. Nos estágios seguintes são reconhecidos fatores mais complexos como ângulos, extremidades e polígonos. Em princípio o número de estágios pode variar e depende do tamanho da imagem de entrada. Quanto maior o tamanho da imagem de entrada, maior o número de estágios necessários para a rede (Teo e Sim, 1995; Poli *et al.*, 2008).



**Figura 6- Estrutura da rede Neocognitron: (a) adaptado de Fukushima, 1988, (b) em formato genérico de redes convolucionais.**

As características extraídas por um estágio têm como entrada apenas as informações enviadas pelo estágio anterior, o que caracteriza uma rede progressiva.

A Figura 7 mostra o processo de reconhecimento da letra 'A' em cada estágio. A figura ilustra, para cada matriz de células dos estágios da rede, o fator correspondente à letra 'A', manipulado. Nota-se que à medida que os estágios avançam, os atributos manipulados pelas células aumentam de complexidade, até o último estágio, em que uma célula reconhece a letra 'A' como um todo.

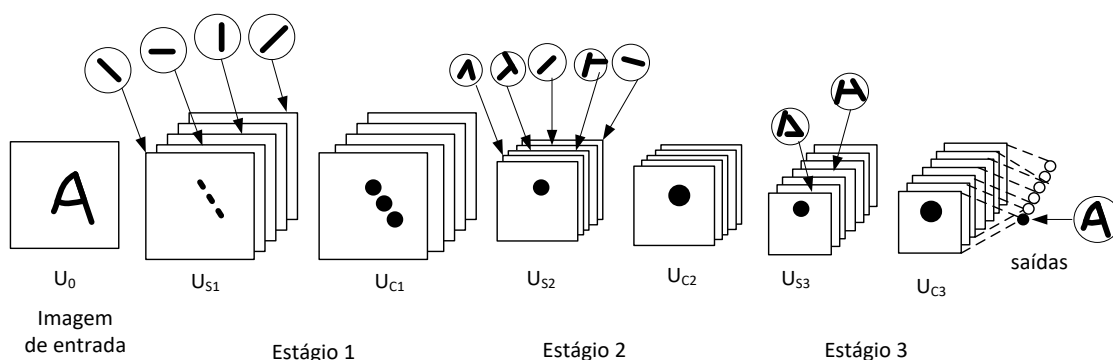
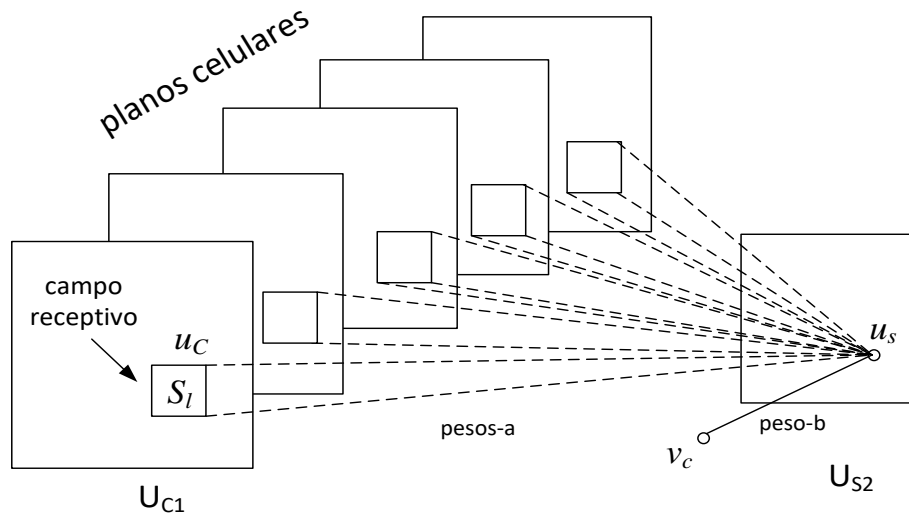


Figura 7- Características reconhecidas pelos neurônios nos 3 estágios da rede.

### 3.1.3 Camada de Células-S

As células-S têm várias entradas excitatórias ou inibitórias. As células-S nos estágios iniciais extraem atributos locais tais como linhas ou bordas. A camada  $U_{S1}$ , que é a primeira camada de células-S, consiste de células de extração de bordas (Myler, 1997; Kangin *et al.*, 2012). As células-S da camada  $U_{S2}$  extraem linhas usando informações das bordas extraídas em  $U_{S1}$ . Essa camada é seguida por uma camada denominada  $U_{S2}$ , que consiste de células de detecção de pontos extremos. A denotação não-convencional da camada  $U_{S2}$ , é devida à especificidade da mesma quanto às conexões. Assim, a camada  $U_{S3}$ , posterior, recebe conexões de entrada não somente de  $U_{S2}$  mas também de  $U_{S2}$  (Fukushima e Shouno, 1998).

Cada célula-S conecta-se a uma pequena área de células, campo receptivo, da camada anterior, que deve ser uma camada de células-C ou camada de entrada. Como existem vários planos celulares na camada-C anterior, uma célula-S, denotada  $u_s$ , conecta-se a todos esses planos celulares. Para cada conexão existe um peso, que é utilizado para influenciar na quantidade de informação que é transferida (Shi e Tan, 2000). A Figura 8 ilustra uma célula  $u_s$  da camada  $U_{S2}$  conectada à camada-C anterior,  $U_{C1}$ , ponderada pelo *peso-a*. Nota-se também a conexão da célula  $u_s$  com a célula  $v_c$ , ponderada pelo *peso-b*.



**Figura 8. Ilustração das conexões de entrada de uma célula-S.**

O valor de saída da célula-S é determinado resumidamente pela equação 6,

$$u_s = \varphi[x] = \varphi \left[ \frac{1 + \sum_{\text{planos}} \sum_{\text{recep}} a \times u_c}{1 + \Theta \times b \times v_c} - 1 \right] \quad (6)$$

onde, no numerador, o primeiro somatório representa todos os planos celulares conectados, conforme Figura 8; o segundo somatório representa todas as conexões dentro do campo receptivo  $S_l$  num plano de células-C; e cada entrada  $u_c$  é multiplicada pelo respectivo *peso-a*, denotado simplesmente  $a$ . No denominador,  $\Theta$  é o limiar responsável pela habilidade de extrair características,  $b$  representa o *peso-b*, e  $v_c$  é um valor correspondente a uma célula auxiliar denotada célula-V que será descrito na seção seguinte. No resultado do cálculo do argumento  $x$  da equação 6 é aplicada a função de ativação  $\varphi$ , definida pela Equação 7, que é uma função ReLU.

$$\varphi[x] = \left. \begin{array}{l} x \quad x \geq 0 \\ 0 \quad x < 0 \end{array} \right\} \quad (7)$$

Nota-se que o argumento  $x$  deve ser maior ou igual a zero quando o numerador  $1 + \sum_{\text{planos}} \sum_{\text{recep}} a \times u_c$  for maior ou igual ao denominador  $1 + \theta \times b \times v_c$ .

Devido a um valor maior do denominador em relação ao numerador evitar o disparo da célula  $u_s$ , fazendo com que a saída seja zero, o papel do denominador é de inibir o disparo, ou seja, a função da célula  $v_c$  é de inibição, ou seja, a conexão é inibitória.

Os pesos- $a$  e o peso- $b$  são variáveis e modificados durante o treinamento. Em geral esses pesos são maiores ou iguais a zero (Fukushima e Tanigawa, 1996; Mousa e Yusof, 2015).

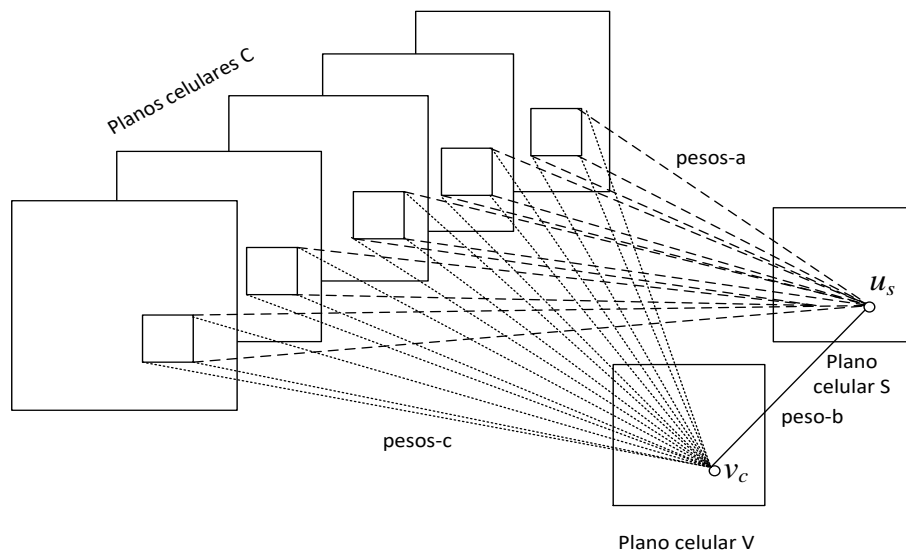
A descrição completa de todos os parâmetros para o cálculo da célula-S é dada pela Equação 8:

$$u_{S_l}(k_l, \mathbf{n}) = \frac{\theta}{1-\theta} \cdot \varphi \left[ \frac{1 + \sum_{k_{l-1}=1}^{K_{l-1}} \sum_{i \in S_l} a_l(k_{l-1}, i, k_l) \cdot u_{C_{l-1}}(k_{l-1}, \mathbf{n} + i)}{1 + \theta \cdot b_l(k_l) \cdot v_{C_l}(\mathbf{n})} - 1 \right] \quad (8)$$

onde  $k_l$  representa o plano celular dentro da camada-S,  $\mathbf{n}$  é a posição do campo receptivo da célula  $u_s$  dentro do plano celular,  $K_{l-1}$  é a quantidade de planos celulares na camada-C anterior,  $K_{l-1}$  representa o índice para todos os planos celulares dessa camada-C, e  $i$  representa o índice para todas as células dentro do campo receptivo  $S_l$ .

### 3.1.4 Camada de Células-V

As células-V são responsáveis por prover informações para o cálculo das células-S sobre a quantidade de atividade nas entradas (Lovell *et al.*, 1997; Gorokhovatskyi, 2016). Na Figura 9 são mostradas as conexões de entrada para uma célula  $v_c$ , e também a sua conexão de saída com a célula  $u_s$  correspondente.



**Figura 9- Ilustração das conexões de entrada e saída da célula  $v_c$ .**

A Equação 9 descreve o cálculo de uma célula-V, onde  $v_c$  é calculada como raiz quadrada da soma ponderada pelo *peso-c*, denotado simplesmente  $c$ , do quadrado de todas as entradas conectadas,  $u_c$ .

$$v_c = \sqrt{\sum_{\text{planos}} \sum_{\text{recep}} c \times u_c^2} \quad (9)$$

A Equação 10 descreve de forma completa o cálculo de  $v_c$  com todos os parâmetros envolvidos, onde  $n$  indica a posição da célula  $v_c$  no campo receptivo do plano celular,  $K_{l-1}$  é a quantidade de planos celulares na camada-C anterior,  $k_{l-1}$  representa o índice para todos os planos celulares dessa camada-C, e  $i$  representa o índice para todas as células dentro do campo receptivo  $S_l$ .

$$v_c(n) = \sqrt{\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{i \in S_l} c_l(i) \times u_{cl-1}^2(n+i, k_{l-1})} \quad (10)$$

Os *pesos-c* são fixos e são usados nas conexões de entrada das células-V. Seus valores são determinados no momento da construção da rede. Estes pesos são maiores na posição do centro do campo receptivo e diminuem à medida que se distancia desse centro

de forma monotônica, de forma gaussiana positiva e o seu somatório é normalizado para 1 (Fukushima e Wake, 1991; Ting e Chuang, 1993).

Como os pesos- $c$  são fixos, existe um único plano celular  $V$  para o cálculo de todos os planos celulares de uma dada camada  $S$ .

A estrutura da rede Neocognitron completa com camadas de células- $C$ , células- $S$  e células- $V$ , em três estágios hierárquicos, é mostrada na Figura 10. Nota-se que em cada estágio existem vários planos celulares  $S$  e  $C$ , e apenas um plano celular  $V$ .

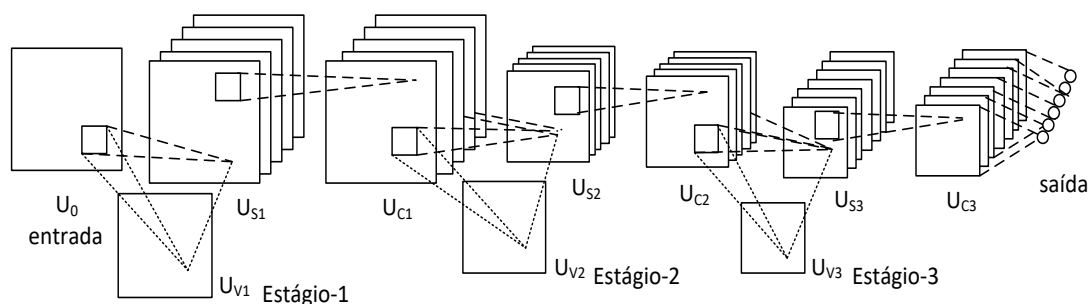
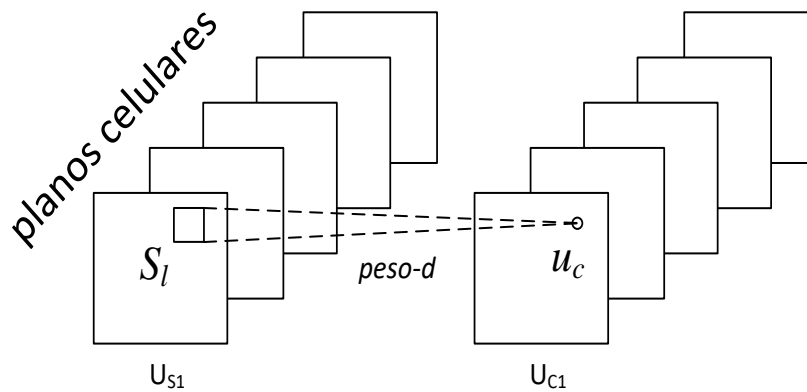


Figura 10- Estrutura da rede Neocognitron com camadas de células- $S$ , células- $C$  e células- $V$ .

### 3.1.5 Camadas de Células- $C$

Ao contrário das células- $S$ , as células- $C$  têm conexões excitatórias fixas e invariáveis, que não necessitam de treinamento. Através destas conexões cada célula- $C$  calcula a média das respostas da camada anterior das células- $S$ . A operação de média das respostas das células- $C$  é importante, não somente por dotar a rede neural com habilidade para reconhecer robustamente padrões deformados, mas também por adicionar suavização nos ruídos contidos nas respostas das células- $S$  (Myler, 1997; Fukushima, 2011).



**Figura 11. Ilustração das conexões de entrada das células-C.**

A Figura 11 é uma ilustração de uma célula-C,  $u_c$ , conectada às entradas da camada-S anterior, no caso, no estágio 1. Cada célula-C atua como um filtro convolucional e recebe sinais do campo receptivo de células-S. Nota-se que as células-C se conectam a apenas um plano celular da camada-S anterior, diferente das células-S. Isso porque cada plano celular da camada-S é responsável pelo filtro de um determinado atributo extraído da imagem de entrada, e as células-C apenas fazem a média por campo receptivo dessas características. Caso uma célula-C estivesse conectada a todos os planos celulares da camada-S, as características contidas seriam perdidas, pois cada plano celular da camada-S representa uma matriz de características diferente dos demais. A Equação 11 descreve o cálculo realizado por uma célula-C, denotado  $u_c$ , onde o argumento é calculado pelo somatório ao longo do campo receptivo das células  $u_s$  ponderadas pelo *peso-d*, denotado simplesmente  $d$ .

$$u_c = \Psi[x] = \Psi[\sum_{recep} d \times u_s] \quad (11)$$

A função  $\Psi$  é descrita pela Equação 12, notando-se que é uma forma de normalização do valor obtido como argumento  $x$ .

$$\Psi[x] = \begin{cases} \frac{x}{1+x} & \text{se } x \geq 0 \\ 0, & \text{caso contrário} \end{cases} \quad (12)$$

A equação 13 apresenta todos os parâmetros necessário para o cálculo de  $u_c$ :

$$u_c(k_l, n) = \Psi \left[ \sum_{i \in S_l} d_l(i) \times u_{sl}(k_l, n + i) \right] \quad (13)$$

onde,  $k_l$  é o plano celular correspondente à célula-C do estágio  $l$ , que está sendo calculada,  $n$  é a posição central do campo receptivo,  $d_l$  é o *peso-d* correspondente ao estágio  $l$ , e  $i$  representa o índice para todas as células dentro do campo receptivo  $S_l$ .

Os *pesos-d* são fixos e determinados no momento da construção da rede e também têm os valores reduzidos monotonicamente a partir do centro da área de conexão, no entanto o somatório dos pesos não é normalizado como ocorre com os *pesos-c* (Fukushima, 1988).

### 3.1.6 Treinamento da rede neural Neocognitron

Das várias camadas existentes na rede, os neurônios das camadas-S precisam de treinamento, usando o conceito de treinamento supervisionado ou não-supervisionado. A aprendizagem não-supervisionada pode fazer uso de regras de aprendizado competitivo, em que o sistema de aprendizagem deve competitivamente aprender a reconhecer novas características da camada de entrada. Tendo em vista que a rede neural Neocognitron utiliza as regras de aprendizado competitivo decidiu-se inserir uma seção no presente capítulo para abordar essa técnica de forma minuciosa.

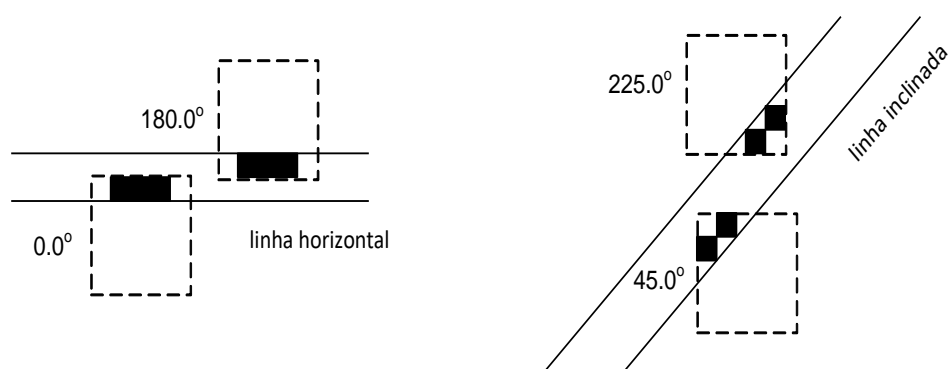
#### 3.1.6.1 Aprendizado supervisionado

Algoritmos de aprendizado supervisionado são comumente utilizados no treinamento de redes neurais convolucionais. Seu funcionamento consiste na inserção de um “professor” no ciclo de aprendizado, responsável por dizer ao modelo se suas previsões estão corretas ou não.

O aprendizado supervisionado caracteriza-se pelo uso de padrões de treinamento em que as respectivas classes são conhecidas. O algoritmo de treinamento consiste em conduzir várias iterações com a apresentação desses padrões à entrada da rede, com a verificação da saída. A cada iteração, é calculado o erro, que consiste na diferença entre a saída verificada e a saída conhecida, e os pesos das conexões devem ser ajustados para que minimize esse erro. O treinamento se encerra no momento em que o erro verificado for menor que um valor pré-definido, ou quando um número de iterações previamente definido for atingido.

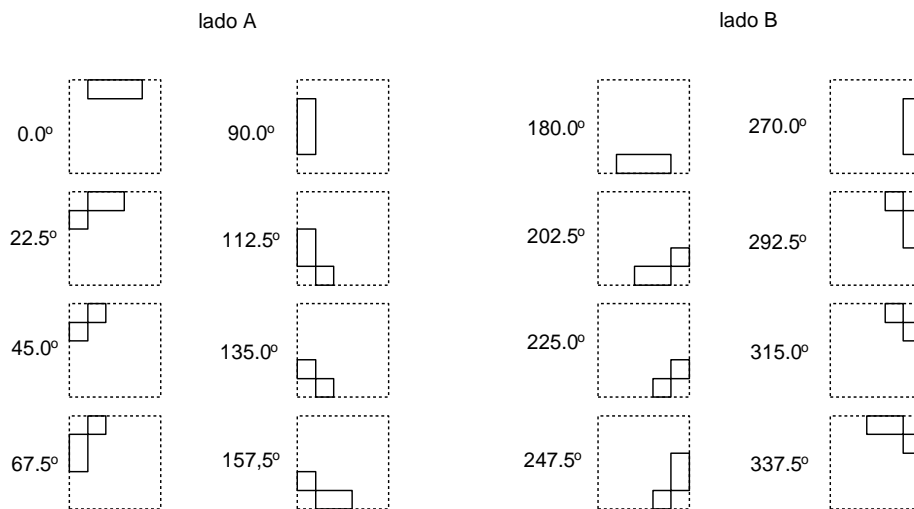
No Neocognitron, o treinamento supervisionado ocorre quando se quer fazer o treinamento dos extratores de características como bordas, linhas e pontos extremos. A seguir serão descritos esses extratores de características.

- **Deteção de Bordas:** Durante a extração de bordas, são verificadas nas imagens de entrada as ocorrências de bordas em diferentes direções. A Figura 12 (a) ilustra uma extração de bordas na direção horizontal e a Figura 12 (b), na direção inclinada. Nota-se que os quadrados tracejados representam os padrões de detecção de bordas, e que são verificadas as bordas nos dois lados de uma linha, que são defasados de  $180^\circ$ . Os quadrados tracejados são filtros de  $4 \times 4$  conexões, sendo que as conexões com peso 1 são destacadas em cor preta e as demais conexões têm peso 0.



**Figura 12. Exemplos de detecção de bordas em linhas horizontal e inclinada.**

Os 16 padrões de borda de tamanho 5x5, divididos em lado A e lado B são mostrados na Figura 13. Os padrões correspondentes a uma defasagem de 180° referem-se aos lados opostos de uma linha.

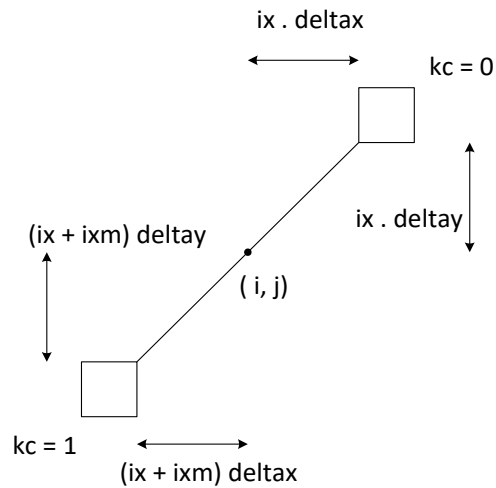


**Figura 13- Padrões de bordas divididos em dois lados (A e B) de uma linha.**

Esses padrões de borda são usados como entrada durante o treinamento supervisionado, para a obtenção dos planos S de detecção de bordas. Como o número de bordas é fixo em 16, o número de planos celulares S na camada de detecção de bordas é fixo em 16, sendo um plano para cada padrão de borda.

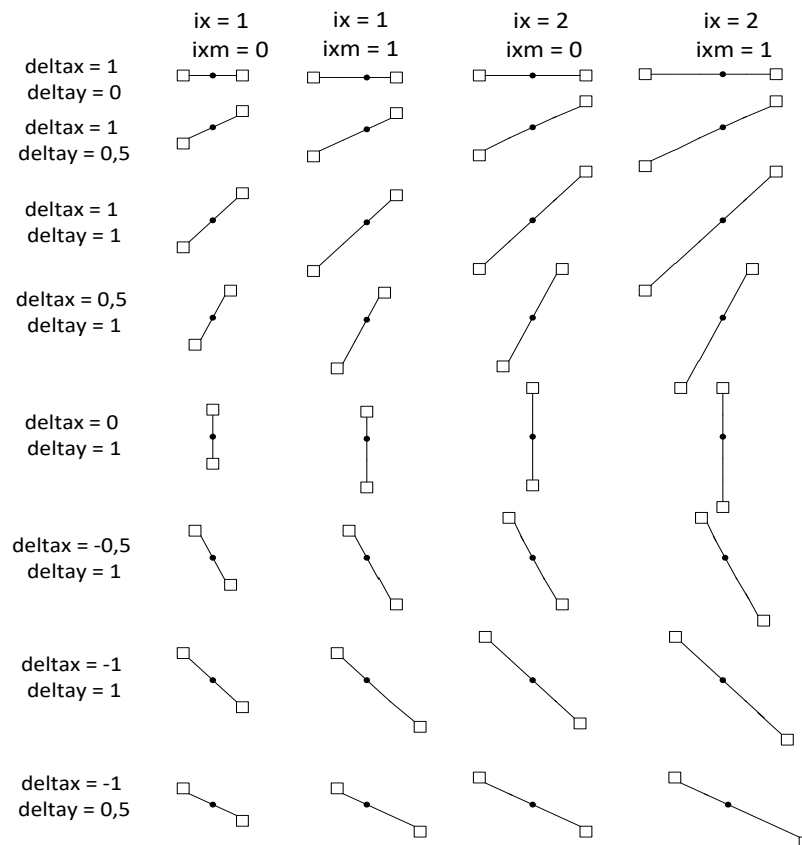
**Detecção de Linhas:** Durante o procedimento de detecção de linhas, 3 espessuras de linhas são consideradas, além das 8 direções de bordas, obtidas na camada anterior. Para a obtenção de uma dada espessura, varia-se os parâmetros conforme ilustrados na Figura 14. As duas bordas detectadas na camada anterior são representadas pelas variáveis  $kc$ , onde  $kc = 0$  para uma borda e  $kc = 1$  para a borda do lado oposto. As coordenadas da borda  $kc = 0$  são representadas pelas variáveis  $ix$ ,  $deltax$  e  $deltay$ , e as coordenadas da borda  $kc = 1$  são,  $ix + ixm$ ,  $deltax$  e  $deltay$ . As variáveis  $deltax$  e  $deltay$  determinam a

direção da linha. Com a variação de  $ix$  e  $ixm$ , são verificadas as três possibilidades de espessura consideradas. A coordenada  $(i, j)$  representa a posição onde a linha é detectada.



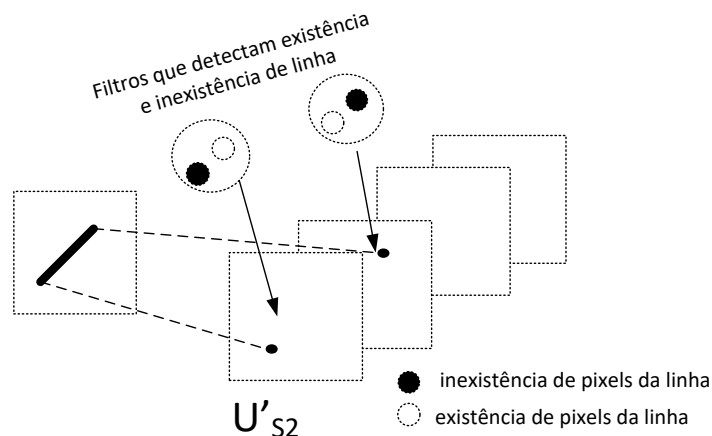
**Figura 14- Parâmetros usados para a detecção de linhas com espessuras diferentes.**

A Figura 15 ilustra as espessuras obtidas variando esses valores, onde o comprimento do segmento mostrado representa a espessura da linha.



**Figura 15- Espessura e direção das linhas variando os parâmetros  $ix$ ,  $ixm$ ,  $deltax$  e  $deltay$**

**Detecção de Pontos Extremos:** Uma outra operação de pré-processamento é a detecção de pontos extremos, nas linhas detectadas pela camada  $U_{S2}$ , que passam para a camada  $U_{C2}$  para a operação convolucional de suavização. Essa detecção de pontos extremos é realizada usando filtros especiais que detectam a existência de objetos de um lado e a inexistência de outro em várias direções em que as linhas são detectadas. A Figura 16 ilustra a obtenção desses pontos extremos numa linha inclinada, resultando em duas ocorrências, cada qual num plano celular diferente, pois correspondem a direções opostas.



**Figura 16- Detecção de pontos extremos**

Os três pré-processamentos descritos resultam em dados de entrada para os estágios posteriores da rede Neocognitron, que normalmente é constituída de mais três estágios, além da camada de saída. O aprendizado utilizado nesses estágios posteriores é o aprendizado competitivo que será descrito a seguir.

### 3.1.6.2 Aprendizado competitivo

O aprendizado competitivo é um aprendizado não-supervisionado, portanto, para um determinado padrão de entrada não é conhecida a sua saída. No Neocognitron, podemos definir aprendizado competitivo como a ação entre neurônios de um plano celular que competem entre si para tornar-se o neurônio vencedor. O treinamento ocorre

para as camadas de células-S, em que os seus *pesos-a* e *peso-b* são variáveis. Lembra-se que os *pesos-c* das células-V e *pesos-d* das células-C são fixos. Apenas um neurônio se torna vencedor e somente o neurônio vencedor tem suas conexões de entrada reforçadas. Esse tipo de aprendizado é denominado de vencedor leva tudo (em inglês, *winner-take-all*).

Para o processo de competição entre as células, no caso do Neocognitron, é usado um plano celular especial, denominado Plano de Seleção de Semente – PSS (*Seed-Selection Plane*, em inglês). Todas as células desse plano são inicializadas com os mesmos pesos mínimos nas conexões de entrada para começar a competição, ao ser apresentada uma dada amostra de entrada. Aquela célula que tiver valor maior do que qualquer outra célula no PSS é selecionada como célula vencedora, que se torna uma célula-semente (*seed cell*) e os seus pesos variáveis são reforçados. Como o aprendizado tem a função de que a rede aprenda a reconhecer uma nova característica em todas as posições, um novo plano celular é criado “plantando” em todas as posições do novo plano a ser criado, essa célula-semente (Fukushima, 1982; Cardoso e Wichert, 2010).

No Neocognitron, o procedimento de treinamento se inicia a partir dos estágios inferiores até terminar o último estágio. O treinamento de uma determinada camada-S, se inicia sem nenhum plano celular, apenas com o PSS. Quando é encontrada uma célula vencedora, os pesos da mesma são reforçadas, e cria-se um novo plano celular. Dessa forma, a cada iteração, um novo plano celular é acrescentado na camada-S. Esse procedimento termina quando não for possível obter mais nenhum novo vencedor, com as amostras de treinamento consideradas. Quando isso acontece, o treinamento prossegue, avançando para o estágio seguinte, quando inicia-se o procedimento de determinação de novos vencedores na nova camada-S, usando como entradas os mesmos padrões de

treinamento usados para o estágio anterior, e usando já os estágios anteriores já treinados juntamente com as camadas-C, cujos pesos são previamente definidos, sem treinamento.

### 3.1.6.3 Evoluções do Aprendizado Competitivo no Neocognitron

Ao longo dos anos algumas melhorias tem sido incorporadas ao Neocognitron e novas regras de aprendizado competitivo tem sido usadas, como por exemplo *winner-kill-loser* e *add-if-silent*.

A regra *winner-kill-loser*, (Fukushima 2010), deve ser usada para treinamento das camadas  $U_{S2}$  e  $U_{S3}$  das camadas de células-S do segundo e terceiro estágio. Durante a fase de treinamento, a saída da célula-C funciona como um vetor de treinamento para células-S. Quando o vetor de treinamento é apresentado, as células-S competem com outras células-S e a célula que emitir maior resposta para o vetor de treinamento torna-se o vencedor. O vencedor aprende o vetor de treinamento e tem suas conexões de entradas reforçadas. Ao mesmo tempo, células perdedoras são removidas da rede. Quanto às células que estão silenciosas para o vetor de treinamento, as mesmas não participam da competição e não são removidas da rede. Elas são esperadas para trabalhar para outros estímulos.

Quanto à regra *add-if-silent*, durante o aprendizado, padrões de treinamento de um conjunto de treinamento são apresentados um a um na camada de entrada  $U_0$ . Se todas as células-S estão silenciosas, uma nova célula-S é gerada e adicionada à rede. A célula adicionada aprende esse estímulo de treinamento. A geração de uma nova célula-S significa a geração de um novo plano de célula, para manter a condição de conexões compartilhadas. Diferente de outras regras de aprendizado uma vez que a célula-S é gerada e adicionada à rede, as conexões de entrada para a célula-S não mudam mais (Fukushima, 2013).

### 3.1.6.4 Reforço dos Pesos no Aprendizado

Tanto no aprendizado supervisionado como no não-supervisionado, os pesos  $a$  e  $b$  são reforçados, com os incrementos definidos de acordo com a Equação 14.

$$\begin{aligned}\Delta a &= q \times c \times u_c \\ \Delta b &= q \times v_c\end{aligned}\quad (14)$$

ou seja, o novo valor de um peso- $a$  passa a ser  $a + \Delta a$  e o novo valor de um peso- $b$  passa a ser  $a + \Delta b$ .

A Equação 15 mostra esses incrementos com todos os parâmetros envolvidos, referente a uma das conexões de entrada,  $i$ , de uma célula-S:

$$\begin{aligned}\Delta a_l(k_{l-1}, i, \hat{k}_l) &= q_l \times c_{l-1}(i) \times u_{c_{l-1}}(k_{l-1}, \hat{n} + i) \\ \Delta b(\hat{k}_l) &= q_l \times v_{c_{l-1}}(\hat{n})\end{aligned}\quad (15)$$

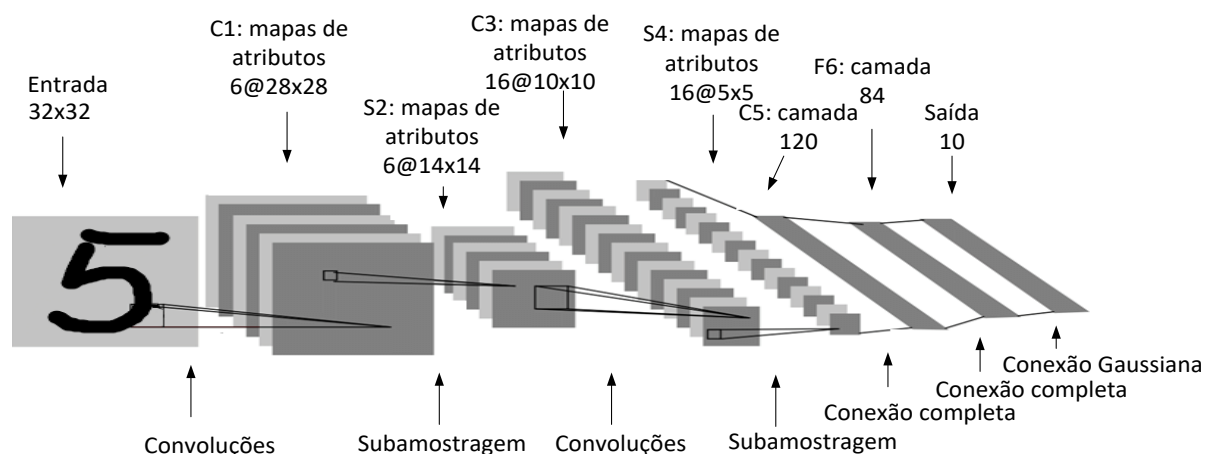
onde  $k_{l-1}$  é o número do plano celular da camada anterior,  $\hat{k}_l$  é o número do plano celular que está sendo acrescentado,  $\hat{n}$  é a posição do neurônio vencedor no PSS e  $q_l$  é o fator de reforço, escolhido empiricamente. A variável  $c_{l-1}(i)$  corresponde ao valor do *peso-c*,  $u_{c_{l-1}}$  é o valor de entrada, e  $v_{c_{l-1}}$  é o valor de entrada proveniente da célula-V.

O treinamento é realizado a partir do estágio inferior até o estágio superior. Após o treinamento do estágio inferior ter sido completamente finalizado, o treinamento do estágio seguinte é iniciado. Os mesmos padrões de treinamento são usados para o treinamento de todos os estágios, porém, as entradas para a camada S a ser treinada são as saídas das camadas C do estágio anterior. Portanto, com os planos celulares já treinados nos estágios anteriores, as saídas das camadas C do estágio anterior devem ser geradas com os mesmos padrões de treinamento, para que o treinamento ocorra nesse novo estágio.

Nesta seção foi descrita a estrutura da rede Neocognitron, bem como o processo de treinamento das camadas desta rede.

### 3.2 LeNet

LeNet foi proposta em 1998 por LeCun e seus colaboradores com o objetivo de fazer reconhecimento de dígitos manuscritos e impressos (LeCun *et al.*, 1998). A Figura 17 mostra a estrutura de rede LeNet. Nota-se que LeNet é uma arquitetura inspirada na rede Neocognitron e apresenta uma estrutura convolucional semelhante.



**Figura 17 - Estrutura da rede neural convolucional LeNet. Fonte: (LeCun *et al.*, 1998).**

O LeNet é uma rede neural convolucional composta por 7 camadas, possuindo parâmetros ou pesos treináveis. Na estrutura original a camada de entrada recebe imagens com resolução 32x32.

A camada C1 é uma camada convolucional de seis planos celulares de tamanho 28x28, contendo 156 parâmetros treináveis, sendo 26 em cada plano, 25 dos pesos nas entradas com campo receptivo 5x5 e um do viés, com um total de 122.304 conexões. Cada plano celular é responsável pela detecção de características específicas da imagem de entrada, sendo que todas as células de um plano celular são idênticas como no Neocognitron. Cada célula obtém uma característica na posição correspondente da imagem de entrada, na região conhecida como campo receptivo e o resultado constitui o plano denominado mapa de características ou mapa de fatores. Um fator obtido numa

posição usa entradas parcialmente sobrepostas em relação a um fator adjacente, ou seja, os campos receptivos se sobrepõem para posições adjacentes do mapa de características.

Uma propriedade interessante é que como todas as células num plano são responsáveis pela obtenção de um mesmo atributo, caso o objeto contido na imagem de entrada for deslocado espacialmente, os atributos no mapa de características se deslocam, porém, não são alterados quanto à forma. Essa propriedade é a base da robustez de redes convolucionais para deslocamentos e distorções na entrada.

A camada S2 é a camada de subamostragem com 6 mapas de características de tamanho 14x14. Essa camada contém 12 parâmetros treináveis, sendo 2 parâmetros em cada mapa: um do coeficiente multiplicativo e outro do *bias* ou viés. Cada célula é conectada a um campo receptivo 2x2, cujos valores são somados e o resultado multiplicado pelo coeficiente treinável, e adicionado a um viés treinável. Os campos receptivos 2x2 não são sobrepostos, portanto, o mapa de características resultante tem a metade da dimensão do mapa da camada anterior C1. O total de conexões da camada S2 é 5.880, considerando 5 conexões por célula, 4 entradas e um viés, dando um total de 980 conexões por mapa de características.

Nota-se que a subamostragem refere-se a uma redução do tamanho do mapa de características, fazendo uma média dos valores no campo receptivo da camada anterior, e reduzindo a precisão espacial do mapa de características. Isso reduz a sensibilidade da saída quanto a distorções de deslocamentos de objetos na imagem.

A camada C3 é uma camada convolucional com 16 mapas de características de tamanho 10x10. Cada unidade em C3 é conectada a um conjunto de campos receptivos de tamanho 5x5 localizados em mapas diferentes da camada S2, conforme a Figura 18,

que mostra com quais mapas da camada S2, as unidades dos mapas da camada C3 combinam.

**Matrizes de características da camada C3**

Matrizes de características da camada S2	0	x			x	x	x		x	x	x	x		x	x		
	1	x	x			x	x	x		x	x	x	x		x		
	2	x	x	x			x	x	x		x			x	x	x	
	3		x	x	x		x	x	x	x				x		x	x
	4			x	x	x		x	x	x	x			x	x		x
	5				x	x	x		x	x	x	x			x	x	x

**Figura 18- Combinação das matrizes de características da camada S2 com as matrizes da camada C3.**

A camada S4 é uma outra camada de subamostragem com 16 mapas de características com tamanho 5x5. Cada unidade em S4 é conectado ao campo receptivo correspondente de tamanho 2x2. São 32 parâmetros treinados, sendo dois em cada mapa, e 2000 conexões, considerando-se 5 conexões, 4 entradas e um viés por unidade, 25 unidades por mapa e 16 mapas.

A camada C5 convolucional tem 120 mapas de características. Cada unidade em C5 é conectada a todos os 16 campos receptivos em S4. Como o campo receptivo de uma célula da camada C5 é 5x5, mesmo tamanho dos mapas da camada S4, o tamanho de um mapa da camada C5 é 1x1. Nesta camada existem 48.120 parâmetros treináveis, sendo 400 conexões de entrada e um viés, por mapa, distribuídos em 120 mapas (LeCun *et al.*, 1998).

A camada F6 contém 84 unidades e é totalmente conectada à camada C5, tendo 10.164 parâmetros treináveis, sendo 120 parâmetros de entrada e um de viés, distribuídos em 84 unidades.

Os neurônios no LeNet, até a camada F6, são computados calculando a soma ponderada das entradas, com o adição de um viés. Essa soma ponderada é aplicada a uma função de ativação tangente hiperbólica.

A camada de saída da rede é composta de dez unidades de Funções de Base Radial (FBR) Euclidiana, uma por classe, com 84 entradas cada. Cada unidade representa um dígito decimal, uma vez que a LeNet foi proposta para reconhecimento de dígitos numéricos decimais. Cada uma das 84 entradas representa um pixel na imagem de entrada de um dígito distribuídos em 7x12 pixels. As saídas de cada unidade  $y_i$  é computada pela equação 16:

$$y_i = \sum_j (x_j - w_{ij})^2 \quad (16)$$

onde  $x_j$  representa uma entrada e  $w_{ij}$  o parâmetro respectivo. Em outras palavras, cada unidade FBR calcula a distância Euclidiana entre o vetor de entrada e o seu vetor de parâmetro. Quanto maior for a distância entre o vetor de entrada do vetor de parâmetro, maior é a saída da função FBR. A saída de uma particular FBR pode ser interpretada, então, como um termo de penalidade medido entre o padrão de entrada e um modelo de classe associada com a FBR. Dado um padrão de entrada, a função de perda poderia ser projetada tal que a configuração de F6 seja tão perto quanto possível ao vetor de parâmetro da FBR que corresponda ao padrão da classe desejada.

### 3.2.1 Treinamento do LeNet

A forma mais comum de aprendizado de máquina, profundo ou não, é o aprendizado supervisionado (LeCun *et al.*, 2015).

Algumas abordagens técnicas são usadas no treinamento da rede, entre elas destaca-se:

### 1) Descida Estocástica de Gradiente (do inglês *Stochastic Gradient Descent*, SGD)

O algoritmo SGD acelera o processo de treinamento pois analisa apenas uma amostra entre as disponíveis, ao invés de analisar toda as instâncias fornecidas (Wijnhoven e With, 2010).

Algoritmos de otimização como o SGD, são usados para ajustar os parâmetros (pesos e *bias*) iniciais aleatórios visando minimizar o custo da função de maneira a convergir para um modelo que forneça boas previsões.

### 2) Algoritmo de aprendizado *backpropagation* com SGD

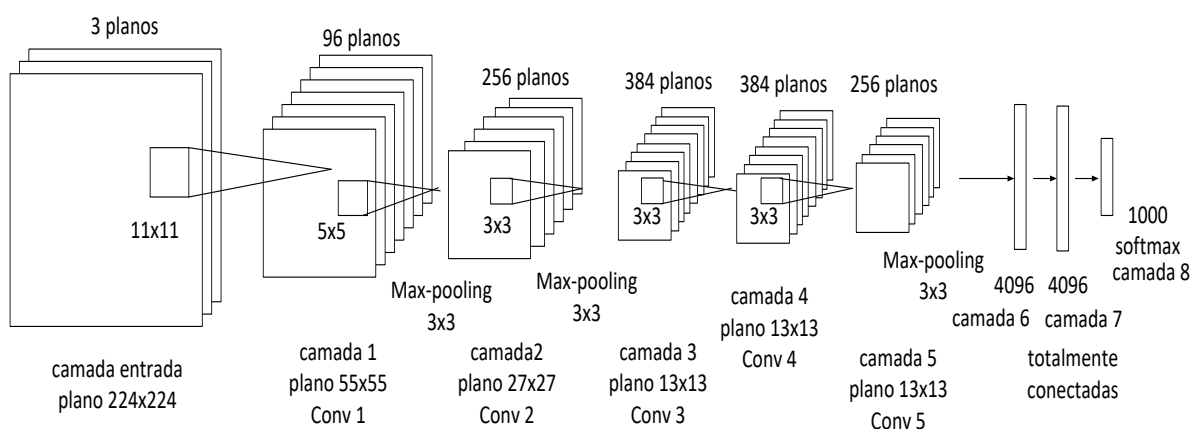
O algoritmo de treinamento *backpropagation* (retro-propagação) consiste em ajustar os pesos em cada uma das camadas da rede e é idealizado para reduzir o erro médio quadrático. Inicialmente quando um padrão é apresentado à rede, ela produz uma saída e, após medir a distância entre essa saída e a desejada, são realizados os ajustes adequados nos pesos de modo a reduzir esta distância. A equação de retropropagação pode ser aplicada repetidamente até que o erro de classificação computado dentro do conjunto de treinamento seja suficientemente baixo ou estabilize (Musakulova *et al.*, 2018).

Em 1998, LeCun e seus colaboradores, aplicaram o algoritmo de treinamento *backpropagation* para a rede neural convolucional LeNet alcançando sucesso na classificação de dígitos manuscritos (LeCun *et al.*, 1998).

Nesta seção foram descritas as peculiaridades referentes ao processo de treinamento das camadas da rede LeNet. A próxima seção trata da rede AlexNet.

### 3.3 AlexNet

AlexNet é uma rede neural convolucional, originalmente criada por Alex Krizhevsky, Geoffrey Hinton, e Ilya Sutskever, inspirada na rede LeNet, porém ela é mais profunda, com mais camadas convolucionais possuindo muito mais mapas de característica. Em 2012 Krizhevsky e seus colaboradores ganharam a competição ImageNet LSVRC-2012 em que classificaram 1,2 milhões de imagens de alta resolução, em mil categorias, por uma grande margem de diferença em relação ao segundo colocado (Krizhevsky *et al.*, 2012; Almisreb *et al.*, 2018; Zhang *et al.*, 2018). Essa competição popularizou a rede AlexNet, abrindo caminho para o estudo de sua arquitetura. A Figura 19 mostra a estrutura da AlexNet.



**Figura 19 – Estrutura da rede CNN AlexNet**

AlexNet é formada por 8 camadas sendo, cinco camadas convolucionais (Conv 1, Conv 2, Conv 3, Conv 4 e Conv 5) e em cada uma dessas camadas é utilizada a função *ReLU* para ativação dos neurônios. Após uma camada convolucional, a rede é composta por uma operação de redução por seleção, *Maxpooling*, como em outras redes convolucionais. Essa operação de *Maxpooling* é realizada em janelas dentro da matriz de características anterior, sendo que o deslocamento de uma janela para outra (*stride* em

inglês), pode variar de uma camada para outra. Além disso, quando a região corresponde a uma borda da matriz, existe um preenchimento (*padding*) das entradas vazias. Após essas camadas, a rede é composta por três camadas totalmente conectadas, sendo na camada de saída a implementação da função *softmax*. A última camada totalmente conectada é constituída de 1000 neurônios e o restante da rede é considerada como um extrator de características. A rede AlexNet, adota a função de ativação *ReLU*, para realizar o treinamento várias vezes mais rápido que outras funções de ativação como por exemplo tangente hiperbólica, utilizada em LeNet. A rede AlexNet foi a primeira rede a utilizar dropout para auxiliar no treinamento da camada totalmente conectada (Srivastava *et al.*, 2014).

A camada de entrada contém 3 imagens com resolução 224x224, contendo 50.176 pixels, totalizando 150.528 atributos. Na camada de entrada é aplicada uma convolução através da camada 1, denotada Conv 1, contendo 96 kernels 11x11, um para cada plano, com deslocamento de 4 pixels, e preenchimento de bordas com zeros. Sobre a matriz de características obtida pela Conv 1, de dimensão 55x55, é aplicada a operação de *Maxpooling* 3x3 com deslocamento de 2 pixels e preenchimento de bordas com zeros.

Sobre a saída do Maxpooling é aplicada a convolução Conv 2, de dimensão de kernel 5x5, obtendo 256 planos usando deslocamento de 1 pixel, e preenchimento de bordas com zeros, resultando em planos de 27x27, seguida de *Maxpooling* 3x3. Aplicando a camada Conv 3, de dimensão de kernel 3x3, sobre esse resultado anterior, obtém-se 512 planos de dimensão 13x13. Sobre esse resultado aplica-se a camada Conv 4, de dimensão de kernel 3x3, resultando em 1024 planos de dimensão 13x13. A última camada de convolução é a Conv 5, também de dimensão 3x3, resultando também em

1024 planos de dimensão 13x13. Nesse último resultado aplica-se *Maxpooling* 3x3. A Tabela 1 resume a estrutura da rede (Krizhevsky *et al.*, 2012; Almisreb *et al.*, 2018).

**Tabela 1 - Estrutura da rede AlexNet**

Camada	Descrição
Entrada	Imagens 224x224x3 totalizando 150.528 atributos na entrada
Camada 1	96 filtros de convolução 11x11, stride [4 4], <i>padding</i> [0 0 0 0]
RELU	Função de ativação
Pool1	<i>Max-Pooling</i> 3x3, stride[2 2], <i>padding</i> [0 0 0 0]
Camada 2	256 <i>kernels</i> ou filtros de convolução com tamanho 5x5
RELU	Função de ativação
Pool2	<i>Max-Pooling</i> 3x3
Camada 3	512 filtros de convolução 3x3, stride [1 1], <i>padding</i> [0 0 0 0]
RELU	Função de ativação
Camada 4	1024 filtros de convolução 3x3, stride [1 1], <i>padding</i> [0 0 0 0]
RELU	Função de ativação
Camada 5	1024 filtros com tamanho 3x3, stride [1 1], <i>padding</i> [0 0 0 0]
RELU	Função de ativação
Pool5	<i>Max-Pooling</i> 3x3, stride [2 2], <i>padding</i> [0 0 0 0]
Totalmente Conectada	Cada camada totalmente conectada tem 4096 neurônios, exceto a última que contém 1000 neurônios.

### 3.3.1 Função Softmax

A saída da rede AlexNet é obtida pela função *softmax*. A função *softmax* corresponde à função exponencial normalizada e é uma função que toma como entrada um vetor de k números reais, e a normaliza numa distribuição de probabilidade consistindo de k probabilidades. Isso significa que números de valores que podem ser negativos ou maiores do que um, serão convertidos em números num intervalo de zero a

um, pela função *softmax*, tal que esses números possam ser interpretados como probabilidades.

Uma unidade padrão de função *softmax*  $\sigma: \mathbb{R}^K \rightarrow \mathbb{R}^K$  é definida pela Equação 17,

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \text{ para } i = 1, \dots, K \text{ e } \mathbf{z} = (z_1, \dots, z_K) \in \mathbb{R}^K \quad (17)$$

### 3.3.2 Treinamento do AlexNet

Segundo Krizhevsky e seus colaboradores o modelo AlexNet deve ser treinado usando *Stochastic Gradient Descent* (SGD), conforme descrito na seção 3.2.1, com 128 exemplos de treinamento ou *batch\_size* a cada iteração. Além disso, é usada uma pequena quantidade de decaimento de peso no valor de 0,0005 (quanto maior, mais lento o decaimento). Essa pequena quantidade de decaimento de peso é importante para o modelo aprender, ou seja, o decaimento de peso não é apenas um mero regularizador, ele reduz a taxa de erro de aprendizagem do modelo.

Outro requisito necessário no treinamento do modelo AlexNet é o *momentum*, que no modelo original foi configurado com valor 0,9. O *momentum* age como um novo hiper-parâmetro que permite controlar a velocidade das mudanças nos pesos usados na rede. Isso é feito criando um fator de *momentum*, que fornece peso para a atual direção do gradiente, e previne que uma nova atualização dos parâmetros se desvie muito da atual direção de busca no espaço dos parâmetros (Krizhevsky *et al.*, 2012).

### 3.4 Considerações Finais

Neste capítulo foram descritas as redes neurais convolucionais Neocognitron, LeNet e AlexNet, suas particularidades, estruturas e também a forma como as mesmas são treinadas.

## **CAPÍTULO 4 - Descrição do Trabalho e Metodologia**

Neste capítulo será descrita a metodologia aplicada nesta dissertação, além dos aspectos de configuração dos simuladores das redes convolucionais a serem utilizadas.

### **4.1 Tipos de Experimentos**

Para esta dissertação foram realizados experimentos de reconhecimento facial usando as redes convolucionais Neocognitron, LeNet-5 e AlexNet, com a intenção de investigar as diferenças nesses três tipos de arquiteturas de redes convolucionais. Para tanto foram utilizadas imagens de faces completas e imagens da região ocular. Nota-se que as regiões oculares são partes que melhor caracterizam as diferenças entre os indivíduos, devido aos detalhes que envolvem essas regiões.

Para que esses experimentos pudessem ser realizados alguns procedimentos foram necessários:

- 1) Tratar as imagens utilizadas, ou seja, redimensionar o tamanho, retirar as partes da face que não contemplam a região ocular, considerando-se que a maioria dos bancos de imagens faciais contém imagens de faces inteiras;
- 2) Implementar as alterações necessárias nos simuladores computacionais, afim de que, os mesmos possam efetuar o reconhecimento. Essas alterações foram feitas interativamente, até atingir as condições de melhor reconhecimento, com os padrões de teste usados;
- 3) Por fim foram realizados testes utilizando imagens de faces completas com o intuito de comparar a diferença entre realizar testes com imagens faciais da região ocular e imagens de faces completas.

## **Motivos para o Reconhecimento de Faces por Regiões Oculares**

A região ocular é altamente discriminatória devido à diferença da expressão ocular entre os indivíduos. A biometria ocular ganhou popularidade devido ao processo ocorrido com o reconhecimento da íris dos indivíduos. Contudo, o reconhecimento de íris é desfavorável quando não ocorre em uma vista frontal. Em tais casos, outras partes dos olhos como a esclera podem ser significantes. Em imagens faciais, a esclera em conjunto com os componentes externos da visão, como os cílios, sobrancelhas, bem como a distância inter-ocular são importantes na diferenciação entre indivíduos (Crihalmeanu *et al.*, 2012, Barcellos *et al.*, 2019).

## **4.2 Base de Imagens**

Neste trabalho pretendeu-se fazer uso de bases de imagens que facilitem a obtenção de imagens de regiões oculares. A seguir será feito um breve descritivo sobre essas bases de imagens.

### **4.2.1 Base de imagens 1**

Para os experimentos iniciais, foi utilizada a base de imagens da Universidade de Essex (Essex, 2019). Os dados dessa base de imagens são mantidos nos diretórios faces94, faces95, faces96, grimace, em ordem crescente de dificuldade. As duas últimas são mais difíceis devido à variação no fundo, escala e expressões. As imagens são armazenadas em formato RGB de 24 bits e formato JPEG. O número total de indivíduos, entre masculinos e femininos, de várias origens raciais, é de 395 elementos, sendo 20 imagens por indivíduo, e total de imagens de 7.900. A idade varia entre 18 a 20 anos, mas imagens de indivíduos mais idosos também estão presentes. Tem indivíduos com óculos e também barbas. A iluminação é artificial, misturando tungstênio e fluorescente.

#### **4.2.2 Base de imagens 2**

A base 1, possui poucas imagens por classe, (20 amostras por indivíduo), neste sentido fez-se necessário utilizar uma base de imagens com um número maior de amostras por indivíduo. Optou-se pelo banco da Universidade de Carnegie Mellon, com variações de pose, iluminação e expressão (Carnegie Mellon University – Pose, Illumination, e Expression Database - CMU-PIE) (Sim et al., 2002).

A base de dados de faces CMU-PIE contém 41.368 imagens de 68 pessoas. Os indivíduos foram fotografados em 13 poses diferentes, 43 condições de iluminação diferentes, com e sem luz de fundo, 4 expressões diferentes (sorrindo, falando e piscando), com 15 pontos de vista e 19 condições de iluminação. Além de imagens frontais de alta resolução.

#### **4.2.3 Pré-processamento para a Obtenção de Imagens da Região Ocular**

Levando-se em consideração os problemas detectados nos experimentos preliminares usando a base de imagens 1, decidiu-se por destacar subimagens exclusivas do olho esquerdo da base de imagens CMU-PIE, redimensionando o tamanho da imagem. Nota-se que nos experimentos preliminares as subimagens da região ocular incluíram ambos os olhos.

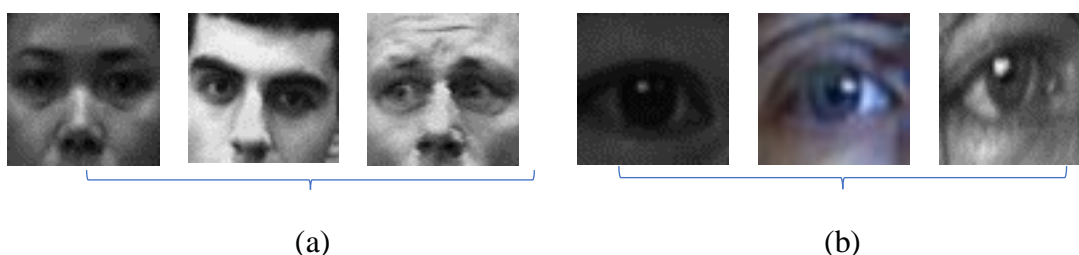
Nesse sentido, foi utilizado o editor de imagem Gimp versão 2.10, para as seguintes operações:

- 1) Conversão do formato JPEG para o formato BMP, requisito necessário para o funcionamento da rede Neocognitron usada;
- 2) Conversão da resolução 180 x 200, base de imagens 1, para 57 x 57 em função da estrutura da rede neural Neocognitron;

3) Conversão da resolução 640 x 486, base de imagens 2, para 57 x 57 em função da estrutura da rede neural Neocognitron; e

3) Representação da imagem do olho esquerdo em escala de cinza.

A Figura 20 exhibe algumas amostras de imagens de regiões oculares extraídas da base de imagens 1 (a) e base de imagens 2 (b).



**Figura 20- Exemplos de imagens de regiões oculares: (a) base de imagens 1, região ocular incluindo ambos os olhos e (b) base de imagens 2, região ocular exclusivo do lado esquerdo.**

#### 4.3 Simulador do Neocognitron

O simulador da rede neural Neocognitron utilizado neste trabalho foi implementado por Saito e seus colaboradores no Departamento de Computação da UFSCar usando o ambiente Delphi™ da Borland® (Bianchini, 2001).

A estrutura do simulador computacional está dividida em duas partes, sendo a primeira parte responsável pelo pré-processamento da rede, ou seja, a extração de bordas, linhas e pontos de controle, denotada de RNDPC (Rede Neural para Detecção de Pontos de Controle) e constitui os estágios inferiores da rede, cujos neurônios são fixos e pré-treinados com aprendizado supervisionado, conforme descrito na Seção 3.1.6.1. Na segunda parte do simulador constam os estágios de nível superior, cujos neurônios são treinados pelo esquema de aprendizado competitivo, seguidos da última camada que é a camada de saída, sendo um neurônio de saída por classe.

O simulador computacional permite controlar os limiares de disparo na detecção de segmentos de bordas e segmentos de retas correspondentes ao primeiro e segundo estágios respectivamente, e nos estágios superiores. O limiar de disparo relaciona-se à excitabilidade dos neurônios envolvidos em função da intensidade das entradas. A Figura 21 mostra uma cópia da tela durante a obtenção dos pontos extremos, que servem de pontos de controle para o algoritmo de treinamento. Da esquerda para a direita temos: imagens de entrada divididas em R, G e B e monocromática, as bordas detectadas nas 16 direções, as oito direções de linhas detectadas em função das bordas, e os 16 pontos extremos em 8 direções de linhas, ou seja, um ponto extremo para cada extremidade de uma linha. No canto direito é mostrado o ajustamento dos parâmetros, dentre eles, o limiar de disparo dos detectores de bordas ( $\theta_1$ ) e linhas ( $\theta_2$ ).

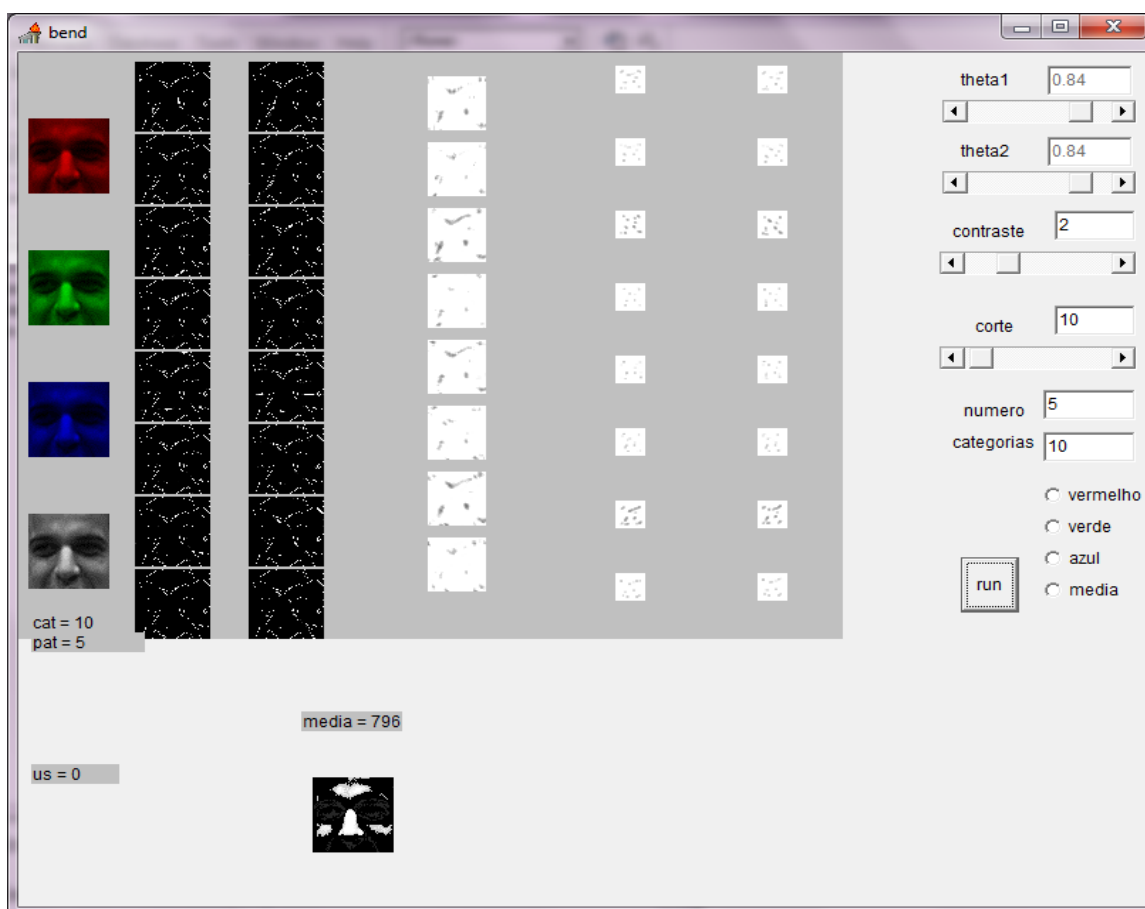


Figura 21 – Ilustração da interface do simulador durante a detecção. Adaptado pela autora

A segunda parte do simulador computacional consiste da Rede Neural para Reconhecimento Facial (RNRF), e seu treinamento é realizado por meio de aprendizado não supervisionado, que usa nos estágios iniciais além do padrão de entrada, os pontos de controle obtidos pela RNDPC. A RNRF obtida é composta de três estágios,  $Us_3-Uc_3$ ,  $Us_4-Uc_4$ ,  $Us_5-Uc_5$ . A Figura 22 mostra um exemplo de tela após o processamento do reconhecimento de padrão, pelas respectivas camadas, na RNRF. No exemplo, uma amostra correspondente à classe 4 padrão 1. A partir da esquerda, tem-se as imagens de entrada separadas em vermelho (R), verde (G) e azul (B), e monocromático, os vários disparos nos planos celulares dos diversos estágios e a saída com o disparo da classe correspondente à imagem de entrada reconhecida como classe 4. O painel de parâmetros mostra os valores de limiar de disparo  $\theta$ , com os valores  $\theta_3 = 0,97$ ,  $\theta_4 = 0,93$  e  $\theta_5 = 0,33$ , nos estágios 3, 4 e 5, respectivamente.

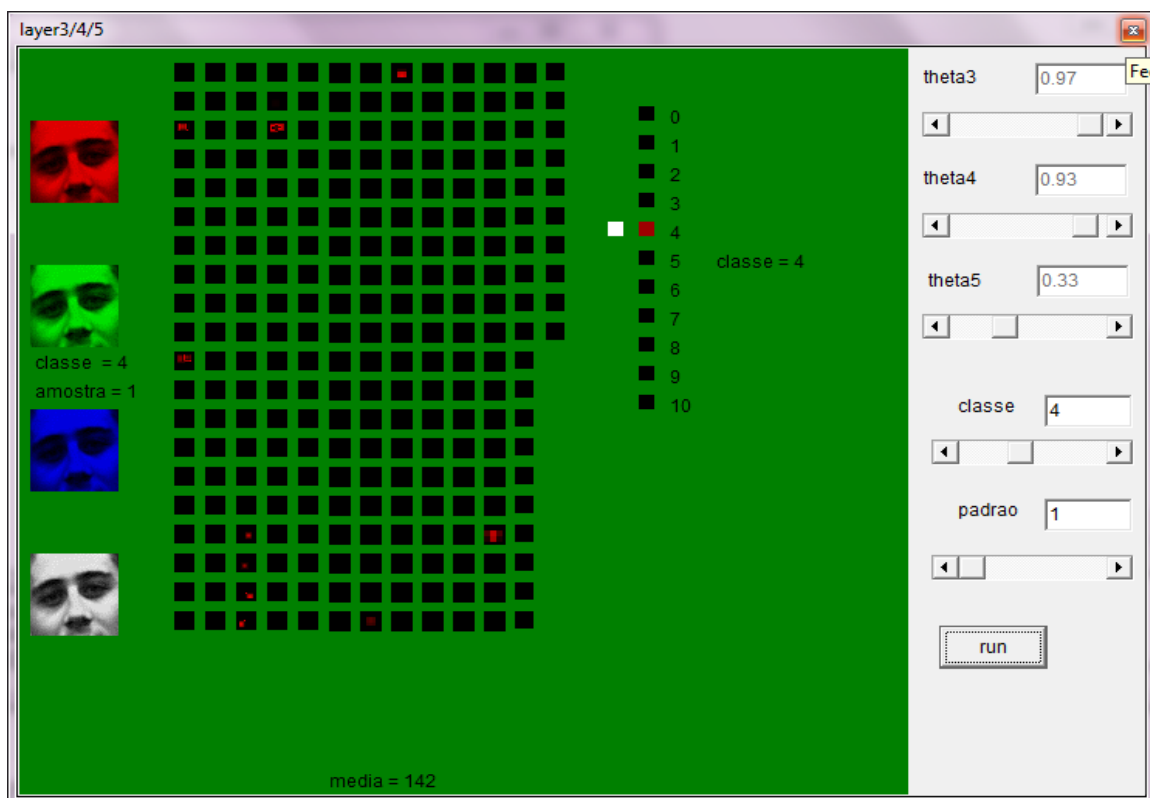


Figura 22 – Interface do reconhecimento de padrão usando RNRF. Adaptado pela autora.

#### 4.4 Simulador da Rede LeNet

O simulador computacional da rede LeNet utilizado neste trabalho é a sua versão 5, denotado LeNet-5, disponível no site (LeNet, 2019). Este simulador foi configurado usando a linguagem de programação Python 3.7 em conjunto com as bibliotecas *Keras* e *TensorFlow*. O simulador permite as seguintes configurações:

- 1) definição das camadas de convolução, camada de *pooling*, camadas totalmente conectadas, a função de ativação a ser usada, sendo que originalmente a rede está configurada para a função *ReLU*, e a função de ativação na saída da rede;
- 2) definição das características das amostras, tais como, resolução (altura e largura), quantidade de classes, proporção de amostras usadas para o treinamento e amostras usadas para os testes; e
- 3) definição do número de épocas de treinamento.

#### 4.5 Simulador computacional AlexNet

O simulador computacional da rede AlexNet, disponível no site (AlexNet, 2019) foi configurado usando a linguagem de programação Python 3.7 em conjunto com as bibliotecas *Keras* e *TensorFlow*. Os seguintes itens são configurados no simulador:

- 1) Definição das camadas de convolução, camada de *pooling*, camada totalmente conectada e função de ativação a ser usada (a rede está configurada originalmente com a função *ReLU*). Por fim configuramos a função de ativação na saída da rede;
- 2) Definição das características das amostras, tais como, resolução (altura e largura), quantidade de classes e definição da quantidade de amostras usadas para o treinamento e amostras usadas para os testes; e

3) Definição do número de épocas de treinamento.

O otimizador utilizado foi o gradiente descendente (*Stochastic Gradient Descent*, SGD) com taxa de aprendizado de 0.01.

## 4.6 Métodos para a Validação dos Resultados

Para a avaliação dos resultados obtidos nos experimentos realizados foi feito o uso do coeficiente Kappa, além do cálculo da porcentagem de acertos, ou taxa de acertos, que corresponde à quantidade de amostras classificadas corretamente dividida pelo número total de amostras classificadas, multiplicada por 100.

### 4.6.1 Coeficiente Kappa

O coeficiente Kappa foi proposto por Cohen (Cohen, 1960) e é definido para determinar a concordância entre resultados obtidos entre procedimentos. As equações (18) e (19) permitem obter o coeficiente Kappa,

$$K = \frac{p_o - p_e}{1 - p_e} \quad (18)$$

onde

$$p_o = \sum_{i=1}^r \frac{n_{ii}}{n} \quad \text{e} \quad p_e = \sum_{i=1}^r \frac{n_{i*} n_{*i}}{n^2} \quad (19)$$

em que  $n_{ii}$  é a quantidade de casos em que existe concordância de resultados entre os dois procedimentos para a classe  $i$ , existindo  $r$  classes de resultados, e  $n$  elementos amostrais de avaliação. A variável  $n_{i*}$  representa a quantidade de elementos amostrais resultante na classe  $i$  pelo primeiro procedimento e  $n_{*i}$ , representa a quantidade de elementos amostrais resultantes na classe  $i$  pelo segundo procedimento.

O significado do coeficiente Kappa pode ser assumido conforme Tabela 2:

**Tabela 2- Significado dos valores do coeficiente Kappa**

Intervalo de valores	significância
Menor que zero	Insignificante

Entre 0 e 0,2	Fraco
Entre 0,21 e 0,4	Razoável
Entre 0,41 e 0,6	Moderado
Entre 0,61 e 0,8	Alto
Entre 0,81 e 1,0	Excelente

#### 4.6.2 Outros Métodos de Validação

Os seguintes casos podem acontecer numa classificação: VP, FP, FN, VN, verdadeiro positivo, falso positivo, falso negativo, verdadeiro negativo sendo que: VP significa que a classificação foi positiva e correta, FP significa o caso em que a classificação resulta em positiva, mas erroneamente, FN é o caso em que o resultado da classificação foi negativo, porém deveria ser positiva e VN é o caso em que o resultado da classificação foi de fato negativo.

Considerando esses valores, podem ser obtidos os valores denotados como erro relativo (ER), precisão e revocação ou sensibilidade (Albuquerque, 2016).

O valor de erro relativo é calculado pela equação 20, e retorna os valores que forem classificados com erro considerando uma dada classificação:

$$ErroRelativo = \frac{FP+FN}{VP+FN} \quad (20).$$

A medida de precisão retorna a razão de predições positivas corretas, dentre todas as predições feitas, VP + FP, para uma determinada classe, pela Equação 21:

$$Precisão = \frac{VP}{VP+FP} \quad (21).$$

A revocação, também conhecida como sensibilidade, é a razão de todas as predições positivas corretas, dentre todas as ocorrências, VP + FN, numa determinada classe, Equação 22:

$$Revocação = \frac{VP}{VP+FN} \quad (22).$$

Tanto o coeficiente Kappa, quanto o erro relativo, precisão e revocação foram utilizados para a validação dos resultados dos experimentos.

#### **4.7 Considerações Finais**

No presente capítulo foram apresentadas as bases de dados, e a metodologia da dissertação de mestrado e os métodos de avaliação utilizados. A seguir serão apresentados os resultados dos experimentos.

## CAPÍTULO 5 – Experimentos Realizados e Resultados

Os testes foram iniciados de modo a avaliar a operacionalidade para diferentes tipos de amostras. Assim sendo, foram obtidos os resultados considerando amostras da região ocular com e sem acessórios (óculos) e amostras da face completa com e sem acessórios (óculos). Além disso, foi verificado o comportamento da rede quando a mesma é influenciada com a mudança no limiar (*threshold*) que determina o nível de intensidade das entradas para o disparo.

Inicialmente os experimentos foram realizados com a Base de Imagens 1 e posteriormente com a Base de Imagens 2.

Na Figura A-1, disponível no Apêndice 1, são exemplificadas 5 amostras de imagens da Base de Imagens 1 de 5 indivíduos usadas nos experimentos sobre região ocular incluindo ambos os olhos. Nota-se que as amostras incluem, além dos olhos, a região nasal. Na Figura A-2, são mostradas 5 amostras de imagens da Base de Imagens 1 de 5 indivíduos ou classes de imagens de faces completas, as mesmas constam no Apêndice 2. Exemplos de 10 amostras de imagens da Base de Imagens 1 de 10 indivíduos, da região ocular de faces que podem conter acessórios como óculos, bigode, ou barba, as mesmas estão na Figura A-3, no Apêndice 3.

Na Figura B-1 são exemplificadas amostras de imagens da região ocular exclusiva do olho direito, da Base de Imagens 2, conforme pode ser visto no Apêndice 4, e na Figura B-2 são exemplificadas amostras de imagens de faces completas, da mesma base, como pode ser visto no Apêndice 5.

## 5.1 Experimentos com o Neocognitron Usando a Base de Imagens 1

Os experimentos seguintes denotados Experimentos 1 a 11 são referentes ao uso do Neocognitron e a Base de Imagens 1. Os Experimentos 1 a 5 referem-se a imagens de região ocular, incluindo ambos os olhos, variando o número de amostras e classes. Os Experimentos 6 a 8 referem-se a imagens da face completa, e os Experimentos 9 a 11 referem-se a imagens de regiões oculares com acessórios.

### Experimentos 1 a 5: Imagens da região ocular da face

No Experimento 1 utilizou-se de cinco classes, utilizou-se as mesmas amostras para a fase de treinamento e para a fase de teste, sendo cinco amostras por classe. Após vários ajustes nos limiares, visando obter resultados mais expressivos, os limiares utilizados nos vários estágios foram: pontos de controle, borda igual a 0,84; linha igual a 0,84; treinamento  $\theta_3$ ,  $\theta_4$  e  $\theta_5$ , iguais a 0,97; teste  $\theta_3$  igual a 0,97;  $\theta_4$  igual a 0,93 e  $\theta_5$  igual a 0,33. A taxa de acerto obtido foi de 96%.

Os valores de limiares (*thresholds*) para os pontos de controle (Borda e Linha) referem-se aos itens denotados como  $\theta_1$  e  $\theta_2$ , respectivamente, visíveis no canto superior à direita da Figura 21; e os valores de limiares de treinamento referem-se aos valores  $\theta_3$ ,  $\theta_4$  e  $\theta_5$ , visíveis no canto direito da Figura 22. Os valores de limiares de reconhecimento/teste são visíveis da mesma forma que os limiares de treinamento, porém durante a fase de teste. Os valores de limiares para os pontos de controle e treinamento utilizados nos experimentos seguintes foram iguais aos utilizados no Experimento 1.

No Experimento 2, o número de classes consideradas é de 10, e as amostras utilizadas para treinamento e teste são distintas, sendo 5 amostras por classe. Os valores

de limiares utilizados para teste foram:  $\theta_3$  igual a 0,94;  $\theta_4$  igual a 0,62 e  $\theta_5$  igual a 0,01. O resultado do teste teve como taxa de acerto, 42%.

No Experimento 3, o número de classes consideradas é também de 10, e as amostras utilizadas para treinamento e teste são distintas, sendo 10 amostras por classe. Os valores de limiares utilizados para teste foram:  $\theta_3$  igual a 0,94;  $\theta_4$  igual a 0,62 e  $\theta_5$  igual a 0,01. O resultado do teste teve como taxa de acerto, 36,66%.

No Experimento 4, o número de classes consideradas é também de 10, e as amostras utilizadas para treinamento e teste são distintas, sendo 15 amostras por classe para treinamento e 5 amostras por classe para teste. Os valores de limiares utilizados para teste foram:  $\theta_3$  igual a 0,90;  $\theta_4$  igual a 0,45 e  $\theta_5$  igual a 0,15. O resultado do teste teve como taxa de acerto, 30%.

O Experimento 5 teve 13 classes, e as amostras utilizadas para treinamento e teste são distintas, sendo 15 amostras por classe para treinamento e 5 amostras por classe para teste. Os valores de limiares utilizados para teste foram:  $\theta_3$  igual a 0,99;  $\theta_4$  igual a 0,30 e  $\theta_5$  igual a 0,01. O resultado do teste teve como taxa de acerto, 18,4%.

### **Experimentos 6 a 8: Imagens da face completa**

O Experimento 6, refere-se ao uso de 5 classes, sendo as mesmas amostras para treinamento e teste. O mesmo número de 5 amostras por classe foi utilizado tanto para treinamento como para teste. Os valores de limiares utilizados para teste foram:  $\theta_3$  igual a 0,97;  $\theta_4$  igual a 0,97 e  $\theta_5$  igual a 0,97. O resultado do teste teve como taxa de acerto, 100%.

No Experimento 7 foram consideradas 5 classes, sendo utilizadas amostras distintas para treinamento e teste. O mesmo número de 5 amostras por classe foi utilizado

tanto para treinamento como para teste. Os valores de limiares utilizados para teste foram: theta3 igual a 0,97; theta4 igual a 0,97 e theta5 igual a 0,97. O resultado do teste teve como taxa de acerto, 80%.

No Experimento 8 foram consideradas 7 classes, sendo utilizadas amostras distintas para treinamento e teste. Foram utilizadas 7 amostras por classe para treinamento e 3 amostras por classe para teste. Os valores de limiares utilizados para teste foram: theta3 igual a 0,97; theta4 igual a 0,97 e theta5 igual a 0,97. O resultado do teste teve como taxa de acerto, 20,4%.

### **Experimentos 9 a 11: Imagens da região ocular com acessórios**

O Experimento 9, refere-se ao uso de 5 classes, sendo utilizadas as mesmas amostras para as fases de treinamento e teste. Os valores de limiares utilizados para teste foram: theta3 igual a 0,97; theta4 igual a 0,97 e theta5 igual a 0,97. O resultado do teste teve como taxa de acerto 100%.

No Experimento 10 foram consideradas 5 classes, com amostras distintas para treinamento e teste. O mesmo número de 5 amostras por classe foi utilizado tanto para treinamento como para teste. Os valores de limiares utilizados para teste foram: theta3 igual a 0,97; theta4 igual a 0,97 e theta5 igual a 0,97. O resultado do teste teve como taxa de acerto, 76%.

No Experimento 11 foram consideradas 7 classes, sendo utilizadas amostras distintas para treinamento e teste. Foram utilizadas 7 amostras por classe para treinamento e 3 amostras por classe para teste. Os valores de limiares utilizados para teste foram:

theta3 igual a 0,97; theta4 igual a 0,97 e theta5 igual a 0,97. O resultado do teste teve como taxa de acerto, 24,4%.

### **5.1.1 Síntese dos Experimentos usando Neocognitron e Base de Imagens 1**

Para sintetizar os experimentos realizados consideramos a Tabela 3 em que constam os experimentos nos quais as amostras de treinamento coincidiram com as de teste: Experimento 1, da região ocular, Experimento 6, da face completa, e Experimento 9, da região ocular com acessórios; e a tabela 4 em que constam os experimentos cujas amostras de treinamento são distintas das amostras de teste, que são os Experimentos 2, 3, 4, 5, 6, 7, 10 e 11.

Nas tabelas 3 e 4, a primeira linha (experim.) indica os experimentos realizados. A segunda linha (imagens) descreve o tipo de imagens utilizadas dentre região ocular (ocular), face completa (comp) e região ocular com acessórios (access). A terceira linha descreve o número de classes relativo ao experimento. A quarta linha refere-se ao número de amostras utilizadas para o treinamento. A quinta linha refere-se ao número de amostras utilizadas para os testes e finalmente na sexta linha são mostrados os resultados em taxa de acerto nos testes.

Nota-se que quando foram usadas as mesmas amostras para treinamento e teste (Tabela 3), a taxa de acerto variou de 96,0% a 100%. Quando as amostras de treinamento foram distintas das amostras de teste (Tabela 4), a taxa de acerto reduziu significativamente. Para melhorar essa taxa de acerto, é necessário aumentar o número de amostras no treinamento, aumentando o número de iterações. Por outro lado, o fato de realizar os testes com um número reduzido de amostras deve resultar em uma taxa de

reconhecimento com um desvio padrão grande, razão pela qual o número de amostras para teste necessitar ser ampliado.

**Tabela 3. Síntese dos experimentos com o Neocognitron com amostras de treinamento coincidentes com as amostras de teste usando Base de Imagens 1**

experim	Exp1	Exp6	Exp9
imagens	ocular	Comp	access
classes	5	5	5
treina	25	25	25
teste	25	25	25
Taxa acertos	96,0%	100%	100%

**Tabela 4. Síntese dos experimentos com o Neocognitron com amostras de treinamento distintas das amostras de teste usando Base de Imagens 1**

experim	Exp2	Exp3	Exp4	Exp5	Exp7	Exp8	Exp10	Exp11
imagens	ocular	ocular	ocular	ocular	comp	comp	Access	access
classes	10	10	10	13	5	7	5	7
treina	50	100	150	195	25	49	25	49
teste	50	30	50	65	25	21	25	21
Taxa acertos	42,0%	36,6%	30%	18,4%	80,0%	30,6%	76,0%	24,4%

Nota-se na tabela 3 que a taxa de acerto varia de 96,0% a 100%. Como as amostras testadas foram as mesmas do treinamento, é esperado uma taxa de acerto dessa ordem. Quanto a tabela 4, a taxa de acerto variou muito, desde 18,4% a 42%, em imagens de região ocular (Experimentos 2 a 5), sendo a menor taxa de 18,4% quando o número de classes foi maior, de 13 classes. Para imagens de face completa a taxa de acerto variou de 80,0% a 30,6%, quando o número de classes variou de 5 para 7 classes. Essa variação grande é justificada considerando-se o número relativamente pequeno de amostras de teste, 25 e 21, respectivamente. Para as imagens de região ocular com acessórios, a taxa de acerto também foi muito variado quando o número de classes variou de 5 para 7, sendo de 76,0% e 24,4%, respectivamente. Essa variação também seria justificada pelo número pequeno de amostras de teste, 25 e 21, respectivamente, como no caso de imagens faciais completas.

A conclusão em relação a esses experimentos é que outros experimentos com um número maior de amostras de treinamento e teste sejam necessários.

## **5.2 Experimentos com LeNet-5 usando Base de Imagens 1**

Os experimentos desta seção se referem ao uso da rede LeNet-5, usando Base de Imagens 1. Os Experimentos 12 a 16, referem-se a imagens de região ocular da face, incluindo ambos os olhos. Os Experimentos 17 a 19 referem-se a imagens de face completa, e os Experimentos 20 a 22 referem-se a imagens de região ocular com acessórios.

### **Experimentos 12 a 16: imagens da região ocular**

O Experimento 12 refere-se ao treinamento e teste em imagens da região ocular, considerando as mesmas amostras para as fases de treinamento e teste. O número de classes consideradas é de 5, e o número de amostras de treinamento é de 25 por classe. O número de épocas para treinamento é de 20. O resultado do teste com 25 amostras por classe foi de 80% para taxa de acerto.

O Experimento 13 caracteriza-se pela existência de 10 classes, e não coincidência das amostras de treinamento e teste. O número de amostras foi de 50 por classe para treinamento, e o número de épocas de 50. O resultado do teste com 50 amostras por classe foi uma taxa de acerto de 60%.

O Experimento 14 caracteriza-se pela existência de 10 classes, e não coincidência das amostras de treinamento e teste. O número de amostras de treinamento foi de 100 por classe, e o número de épocas de 50. O resultado do teste com 30 amostras por classe foi uma taxa de acerto de 80%.

O Experimento 15 também refere-se a 10 classes, e não coincidência das amostras de treinamento e teste. O número de amostras de treinamento foi de 150 por classe, e o número de épocas de 50. O resultado do teste com 50 amostras por classe foi uma taxa de acerto de 66%.

O Experimento 16 refere-se a 13 classes, e não coincidência das amostras de treinamento e teste. O número de amostras de treinamento foi de 195 por classe, e o número de épocas de 50. O resultado do teste com 65 amostras por classe foi uma taxa de acerto de 76,9%.

#### **Experimentos 17 a 19: imagens da face completa**

O Experimento 17 refere-se ao treinamento e teste em imagens da face completa, considerando as mesmas amostras para as fases de treinamento e teste. O número de classes considerado é de 5, e o número de amostras de treinamento é de 25 por classe. O número de épocas para treinamento é de 50. O resultado do teste com 25 amostras por classe foi de 80% para taxa de acerto.

O Experimento 18 refere-se a 5 classes, e não coincidência das amostras de treinamento e teste. O número de amostras foi de 25 por classe para treinamento, e o número de épocas de 50. O resultado do teste com 25 amostras por classe foi uma taxa de acerto de 80%.

O Experimento 19 caracteriza-se pela existência de 7 classes, e não coincidência das amostras de treinamento e teste. O número de amostras de treinamento foi de 49 por classe, e o número de épocas de 50. O resultado do teste com 21 amostras por classe foi uma taxa de acerto de 10%.

### **Experimentos 20 a 22: imagens da região ocular com acessórios**

O Experimento 20 refere-se ao treinamento e teste em imagens da região ocular com acessórios, considerando as mesmas amostras para as fases de treinamento e teste. O número de classes considerado é de 5, e o número de amostras de treinamento é de 25 por classe. O número de épocas para treinamento é de 50. O resultado do teste com 25 amostras por classe foi de 80% para taxa de acerto.

O Experimento 21 refere-se a 5 classes, e não coincidência das amostras de treinamento e teste. O número de amostras foi de 25 por classe para treinamento, e o número de épocas de 50. O resultado do teste com 25 amostras por classe foi uma taxa de acerto de 76%.

O Experimento 22 caracteriza-se pela existência de 7 classes, e amostras de treinamento e teste distintas. O número de amostras de treinamento foi de 49 por classe, e o número de épocas de 50. O resultado do teste com 21 amostras por classe foi uma taxa de acerto de 76%.

#### **5.2.1 Síntese dos Experimentos usando LeNet-5 e Base de Imagens 1**

Nas tabelas 5 e 6 são mostradas as sínteses dos experimentos realizados com a rede LeNet-5, onde na primeira linha (experim.) são identificados os experimentos realizados, numerados de 12 a 22. A segunda linha (imagens) corresponde ao tipo de imagens utilizadas dentre região ocular, face completa (comp) e região ocular com acessórios (access). A terceira linha refere-se ao número de classes relativo ao experimento. A quarta linha refere-se ao número de amostras utilizadas para o treinamento. A quinta linha refere-se ao número de amostras utilizadas para os testes. A sexta linha refere-se ao número de épocas utilizadas durante o treinamento e finalmente na sétima linha são mostrados os resultados em taxa de acerto nos testes. A tabela 5 refere-

se aos Experimentos 12, 17 e 20, em que as amostras utilizadas para o treinamento são as mesmas das amostras de teste. A tabela 6 refere-se aos Experimentos 13, 14, 15, 16, 18, 19, 21 e 22, em que as amostras são distintas para as fases de treinamento e teste.

**Tabela 5. Experimentos com LeNet-5 com mesmas amostras de treinamento e teste**

<b>Experim</b>	<b>12</b>	<b>17</b>	<b>20</b>
<b>Imagens</b>	<b>Ocular</b>	<b>Comp</b>	<b>Acess</b>
<b>Classes</b>	<b>5</b>	<b>5</b>	<b>5</b>
<b>Treina</b>	<b>25</b>	<b>25</b>	<b>25</b>
<b>Teste</b>	<b>25</b>	<b>25</b>	<b>25</b>
<b>Épocas</b>	<b>20</b>	<b>50</b>	<b>50</b>
<b>Taxa de acertos</b>	<b>80%</b>	<b>80%</b>	<b>80%</b>

**Tabela 6. Experimentos com o LeNet-5 com amostras de treinamento e teste distintas**

<b>Experim</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>18</b>	<b>19</b>	<b>21</b>	<b>22</b>
<b>imagens</b>	<b>ocular</b>	<b>ocular</b>	<b>ocular</b>	<b>ocular</b>	<b>comp</b>	<b>comp</b>	<b>Acess</b>	<b>access</b>
<b>classes</b>	<b>10</b>	<b>10</b>	<b>10</b>	<b>13</b>	<b>5</b>	<b>7</b>	<b>5</b>	<b>7</b>
<b>treina</b>	<b>50</b>	<b>100</b>	<b>150</b>	<b>195</b>	<b>25</b>	<b>49</b>	<b>25</b>	<b>49</b>
<b>teste</b>	<b>50</b>	<b>30</b>	<b>50</b>	<b>65</b>	<b>25</b>	<b>21</b>	<b>25</b>	<b>21</b>
<b>épocas</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>
<b>Taxa de acertos</b>	<b>60%</b>	<b>80%</b>	<b>66%</b>	<b>76,9%</b>	<b>80%</b>	<b>10%</b>	<b>76%</b>	<b>76%</b>

Nota-se nesses experimentos que a taxa de acerto teve uma pequena variação ao longo dos experimentos, considerando-se todos os tipos de imagens: região ocular, face completa e região ocular com acessórios. Mesmo usando as mesmas amostras para treinamento e reconhecimento (tabela 5), a taxa de acerto não sofreu muitas alterações.

Taxa de acerto da tabela 5, em que as amostras de treinamento e teste foram as mesmas, ficou fixa em 80%. Quanto à tabela 6, a taxa de acerto variou de 60% a 80%, com exceção do Experimento 19, de faces completas, em que a taxa de acerto foi de 10%. Esse resultado pode ser explicado se considerarmos que o número de amostras de treinamento foi pequeno, bem como o número de amostras de teste.

### 5.3 Experimentos com AlexNet e Base de Imagens 1

Com a rede AlexNet, foram realizados Experimentos 23 a 27 para imagens de região ocular; Experimentos 28 a 30 para imagens de face completa; e Experimentos 31 a 33 para imagens de região ocular com acessórios.

#### **Experimentos 23 a 27: imagens da região ocular**

O Experimento 23, com a rede AlexNet, foram iniciados considerando 5 classes com 25 amostras cada, para o treinamento. Neste primeiro experimento, as amostras de treinamento e teste são coincidentes. O número de 10 épocas de treinamento foi aplicado. Os testes obtidos com 25 amostras por classe, resultaram numa taxa de acerto de 20%.

No Experimento 24, considerando-se 5 classes, e 50 amostras cada, para o treinamento, as amostras de treinamento e teste foram distintas. O número de 20 épocas de treinamento foi aplicado. Os testes obtidos com 50 amostras por classe, resultaram numa taxa de acerto de 50%.

No Experimento 25, considerando-se 10 classes, e 100 amostras cada, para o treinamento, as amostras de treinamento e teste foram distintas. O número de 20 épocas de treinamento foi aplicado. Os testes obtidos com 30 amostras por classe, resultaram numa taxa de acerto de 23,3%.

No Experimento 26, considerou-se 10 classes e 150 amostras cada, para o treinamento, sendo que as amostras de treinamento e teste foram distintas. O número de 20 épocas de treinamento foi aplicado. Os testes obtidos com 50 amostras por classe, resultaram numa taxa de acerto de 74,0%.

O Experimento 27 refere-se a 13 classes e 195 amostras cada, para o treinamento, sendo as amostras de treinamento e teste distintas. O número de 20 épocas de treinamento

foi aplicado. Os testes obtidos com 65 amostras por classe, resultaram numa taxa de acerto de 92,3%.

### **Experimentos 28 a 30: imagens da face completa**

O Experimento 28 refere-se a 5 classes e 25 amostras cada, para o treinamento, sendo iguais as amostras de treinamento e teste. O número de 20 épocas de treinamento foi aplicado. Os testes obtidos com 25 amostras por classe, resultaram numa taxa de acerto de 76,0%.

No Experimento 29, foram consideradas 5 classes e 25 amostras cada, para o treinamento, sendo as amostras de treinamento e teste distintas. O número de 20 épocas de treinamento foi aplicado. Os testes obtidos com 25 amostras por classe, resultaram numa taxa de acerto de 80,0%.

No Experimento 30, foram consideradas 7 classes e 49 amostras cada, para o treinamento, sendo as amostras de treinamento e teste distintas. O número de 20 épocas de treinamento foi aplicado. Os testes obtidos com 21 amostras por classe, resultaram numa taxa de acerto de 88,8%.

### **Experimentos 31 a 33: imagens da região ocular com acessórios**

O Experimento 31 refere-se a 5 classes e 25 amostras de treinamento por classe, sendo iguais as amostras de treinamento e teste. O número de 20 épocas de treinamento foi aplicado. Os testes obtidos com 25 amostras por classe, resultaram numa taxa de acerto de 100%.

No Experimento 32, foram consideradas 5 classes e 25 amostras cada, para o treinamento, sendo as amostras de treinamento e teste distintas. O número de 20 épocas

de treinamento foi aplicado. Os testes obtidos com 25 amostras por classe, resultaram numa taxa de acerto de 96,0%.

No Experimento 33, foram consideradas 7 classes e 49 amostras cada, para o treinamento, sendo as amostras de treinamento e teste distintas. O número de 20 épocas de treinamento foi aplicado. Os testes obtidos com 21 amostras por classe, resultaram numa taxa de acerto de 88,8%.

### 5.3.1 Síntese dos Experimentos usando AlexNet com a Base de Imagens 1

Nas tabelas 7 e 8 são sintetizados os experimentos realizados com a rede AlexNet, da mesma forma que os experimentos com o LeNet-5 e Neocognitron.

Na tabela 7 são resumidos os experimentos onde as amostras de treinamento coincidiram com as amostras de teste, sendo o Experimento 23 referente a imagens da região ocular, o Experimento 28, de faces completas, e o Experimento 31, de região ocular com acessórios. Na tabela 8 são resumidos os experimentos em que as amostras de treinamento foram distintas das amostras de teste, quais sejam de números: 24, 25, 26, 27, 29, 30, 32 e 33.

**Tabela 7. Síntese dos experimentos com o AlexNet**

<b>Experim.</b>	<b>23</b>	<b>28</b>	<b>31</b>
<b>imagens</b>	<b>ocular</b>	<b>comp</b>	<b>acess</b>
<b>Classes</b>	<b>5</b>	<b>5</b>	<b>5</b>
<b>Treina</b>	<b>25</b>	<b>25</b>	<b>25</b>
<b>Teste</b>	<b>25</b>	<b>25</b>	<b>25</b>
<b>Épocas</b>	<b>10</b>	<b>20</b>	<b>20</b>
<b>Taxa de acertos</b>	<b>20%</b>	<b>76%</b>	<b>100%</b>

Tabela 8. Síntese dos experimentos com o AlexNet

Experim.	24	25	26	27	29	30	32	33
imagens	ocular	ocular	ocular	ocular	comp	comp	Acess	access
classes	10	10	10	13	5	7	5	7
treina	50	100	150	195	25	49	25	49
teste	50	30	50	65	25	21	25	21
épocas	20	20	20	20	20	20	20	20
Taxa de acertos	50%	23,3%	74%	92,3%	80%	88,8%	96%	88,8%

Nota-se nesses experimentos que a taxa de acerto teve variações maiores em comparação com a rede LeNet-5 ao longo dos experimentos. Na tabela 7, a taxa de acerto variou de 20%, 76% e 100%, para amostras da região ocular, face completa e região ocular com acessórios. A justificativa para essa variação seria a quantidade pequena de amostras de treinamento, teste, e número de épocas de treinamento. Na tabela 8 a taxa de acerto variou de 23,3% a 92,3%, para amostras de região ocular; de 80% a 88,8%, para amostras de face completa; e de 88,8% a 96%, para amostras de região ocular com acessórios. O maior valor de taxa de acerto de 92,3% ocorreu quando o número de classes foi maior, de 13, porém, o número de amostras de treinamento também foi maior, de 195, o que é razoável. O caso em que a taxa de acerto foi de 23,3%, Experimento 25, pode ser justificado se considerarmos o número reduzido de amostras de teste, de 30.

#### 5.4 Experimentos com Neocognitron e Base de imagens 2

Os experimentos descritos a partir desta seção correspondem às amostras contidas na base de imagens 2. Os experimentos desta seção referem-se ao uso da rede Neocognitron, cujos os limiares utilizados foram sempre: Borda, 0,84; Linha, 0,84; treinamento  $U_{S3}$ , 0,8;  $U_{S4}$ , 0,75;  $U_{S5}$ , 0,60; teste  $U_{S3}$ , 0,75;  $U_{S4}$ , 0,70 e  $U_{S5}$ , 0,60.

Em todos os experimentos foram utilizadas amostras distintas para as fases de treinamento e teste. Os Experimentos 34 a 39 referem-se ao uso de imagens de região ocular exclusiva do olho do lado esquerdo da imagem facial, variando o número de classes e número de amostras para treinamento e teste; e os Experimentos 40 a 42, referem-se ao uso de imagens completas da face, variando o número de classes e número de amostras para treinamento e teste.

### **Experimentos 34 a 39: Amostras da região ocular exclusiva de um olho**

No Experimento 34, foram consideradas 5 classes, com 100 amostras de treinamento para cada classe. O teste foi realizado com 40 amostras e a taxa de acerto foi de 46,50%.

No Experimento 35, foram consideradas 10 classes, com 100 amostras de treinamento para cada classe. O teste foi realizado com 40 amostras por classe e a taxa de acerto foi de 46,75%.

No Experimento 36, foram consideradas 20 classes, com 100 amostras de treinamento para cada classe. O teste foi realizado com 40 amostras por classe e a taxa de acerto foi de 65,87%.

No Experimento 37, foram consideradas 10 classes, com 200 amostras de treinamento para cada classe. O teste foi realizado com 100 amostras por classe e a taxa de acerto foi de 82,00%.

No Experimento 38, foram consideradas 20 classes, com 200 amostras de treinamento para cada classe. O teste foi realizado com 100 amostras por classe e a taxa de acerto foi de 49,45%.

No Experimento 39, foram consideradas 20 classes, com 500 amostras de treinamento para cada classe. O teste foi realizado com 100 amostras por classe e a taxa de acerto foi de 69,40%.

#### **Experimentos 40 a 42: Amostras da face completa**

No Experimento 40, foram consideradas 10 classes, com 100 amostras de treinamento para cada classe. O teste foi realizado com 40 amostras por classe e a taxa de acerto foi de 64,20%.

No Experimento 41, foram consideradas 20 classes, com 200 amostras de treinamento para cada classe. O teste foi realizado com 100 amostras por classe e a taxa de acerto foi de 47,87%.

No Experimento 42, foram consideradas 20 classes, com 500 amostras de treinamento para cada classe. O teste foi realizado com 100 amostras por classe e a taxa de acerto foi de 41,80%.

#### **5.4.1 Síntese dos Experimentos usando Neocognitron para Base de Imagens 2**

Na tabela 9 é mostrada uma síntese dos experimentos realizados com a rede Neocognitron, onde na primeira linha (experim.) são indicados os nove experimentos realizados, numerados de 34 a 42. A segunda linha (imagens) descreve o tipo de imagens utilizada dentre região ocular e face completa (comp). A terceira linha descreve o número de classes relativo ao experimento, que variou entre 5 a 20 classes. A quarta linha refere-se ao número de amostras utilizadas para o treinamento, que variou entre 100 e 500. A quinta linha refere-se ao número de amostras utilizadas para os testes, que variou entre 40 a 200 amostras. A sexta, sétima, oitava e nona linhas referem-se aos métodos de avaliação usados nos experimentos, índice Kappa, erro relativo (ER), revocação e

precisão, respectivamente. E finalmente na última linha são mostrados os resultados em taxa de acerto nos testes.

Nota-se que a taxa de acerto variou de 41,80% a 82,00% considerando-se todos os experimentos realizados com o Neocognitron. Verifica-se também que a melhor taxa de acerto ocorreu para 10 classes, com treinamento de 200 amostras por classe, no Experimento 37.

**Tabela 9- Base de imagens 2 - Síntese dos experimentos com o Neocognitron**

<b>Experim.</b>	<b>34</b>	<b>35</b>	<b>36</b>	<b>37</b>	<b>38</b>	<b>39</b>	<b>40</b>	<b>41</b>	<b>42</b>
<b>Imagens</b>	<b>ocular</b>	<b>ocular</b>	<b>ocular</b>	<b>ocular</b>	<b>ocular</b>	<b>ocular</b>	<b>comp</b>	<b>comp</b>	<b>Comp</b>
<b>Classes</b>	<b>5</b>	<b>10</b>	<b>20</b>	<b>10</b>	<b>20</b>	<b>10</b>	<b>10</b>	<b>16</b>	<b>20</b>
<b>Treina</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>200</b>	<b>200</b>	<b>500</b>	<b>100</b>	<b>200</b>	<b>500</b>
<b>Teste</b>	<b>40</b>	<b>40</b>	<b>40</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>40</b>	<b>100</b>	<b>100</b>
<b>Kappa</b>	<b>0,44</b>	<b>0,45</b>	<b>0,64</b>	<b>0,82</b>	<b>0,49</b>	<b>0,70</b>	<b>0,63</b>	<b>0,44</b>	<b>0,37</b>
<b>ER</b>	<b>0,53</b>	<b>1,05</b>	<b>0,43</b>	<b>0,22</b>	<b>0,64</b>	<b>0,37</b>	<b>0,57</b>	<b>0,11</b>	<b>0,10</b>
<b>Revocação</b>	<b>0,5</b>	<b>0,93</b>	<b>0,82</b>	<b>1</b>	<b>0,92</b>	<b>0,98</b>	<b>0,90</b>	<b>0,55</b>	<b>0,47</b>
<b>Precisão</b>	<b>0,87</b>	<b>0,49</b>	<b>0,77</b>	<b>0,82</b>	<b>0,78</b>	<b>0,71</b>	<b>0,65</b>	<b>0,78</b>	<b>0,74</b>
<b>Taxa de acertos</b>	<b>46,50%</b>	<b>46,75%</b>	<b>65,87%</b>	<b>82%</b>	<b>49,45%</b>	<b>69,40%</b>	<b>64,20%</b>	<b>47,87%</b>	<b>41,80%</b>

### **5.5 Experimentos com LeNet para Base de Imagens 2**

Os experimentos descritos nesta seção correspondem à rede LeNet com Base de Imagens 2. Foram realizados os Experimentos 43 a 45, usando imagens de região ocular exclusiva; e os Experimentos 46 e 47, usando imagens da face completa, variando o número de classes e número de amostras de treinamento e teste. Em todos os experimentos, as amostras da fase de treinamento foram distintas das amostras da fase de teste, e o número de épocas de treinamento adotado foi 30.

#### **Experimentos 43 a 45: Imagens da região ocular da face**

No experimento 43 foram consideradas 10 classes, com 200 amostras por classe, para treinamento. O número de épocas de treinamento foi 30. O teste com 100 amostras por classe teve uma taxa de acerto de 40,50%.

No experimento 44 foram consideradas 20 classes, com 500 amostras por classes, para treinamento. O número de épocas de treinamento foi 30. O teste com 100 amostras por classe teve uma taxa de acerto de 42,35%.

No experimento 45 foram consideradas 20 classes, com 700 amostras por classes, para treinamento. O número de épocas de treinamento foi 30. O teste com 100 amostras por classe teve uma taxa de acerto de 48,55%.

### **Experimentos 46 e 47: Imagens da face completa**

O Experimento 46 refere-se ao uso de 20 classes e treinamento com 200 amostras por classe. O número de épocas considerada para treinamento foi de 30. O resultado de teste com 100 amostras por classe foi de uma taxa de acerto de 33,25%.

No Experimento 47 refere-se ao uso de 20 classes e treinamento com 500 amostras por classe. O número de épocas considerada para treinamento foi de 30. O resultado de teste com 100 amostras por classe foi de uma taxa de acerto de 42,15%.

#### **5.5.1 Síntese dos Experimentos usando LeNet-5 com Base de Imagens 2**

Na tabela 10 é mostrada uma síntese dos experimentos realizados com a rede LeNet-5, onde na primeira linha são identificados os 5 experimentos realizados, numerados de 43 a 47. A segunda linha (imagens) descreve o tipo de imagens utilizada dentre região ocular, face completa (completa). A terceira linha descreve o número de classes relativo ao experimento, que variou entre 10 a 20 classes. A quarta linha refere-se ao número de amostras utilizadas para o treinamento, que variou entre 200 e 700. A quinta linha refere-se ao número de amostras utilizadas para os testes, que variou entre 100 e 300 amostras. A sexta linha refere-se ao número de épocas utilizadas durante o treinamento. As sétima, oitava, nona e décima linhas referem-se aos índices de avaliação

usados nos experimentos, índice Kappa, erro relativo (ER), revocação e precisão respectivamente e finalmente na última linha são mostrados os resultados em taxa de acerto nos testes.

Nota-se que a taxa de acertos variou entre 33,25% a 48,55%. Um possível motivo para esses valores é o baixo número de 30 épocas de treinamento.

**Tabela 10 - Base de imagens 2 - Síntese dos experimentos com o LeNet-5**

Experimento	43	44	45	46	47
imagens	Ocular	ocular	ocular	completa	completa
classes	10	20	20	20	20
treina	200	500	700	200	500
teste	100	100	100	100	100
épocas	30	30	30	30	50
Kappa	0,40	0,42	0,48	0,33	0,42
ER	1,35	0,86	0,56	0,91	0,86
Revocação	1	1	1	1	1
Precisão	0,40	0,42	0,48	0,33	0,42
Taxa de Acertos	40,50%	42,35%	48,55%	33,25%	42,15%

## 5.6 Experimentos com AlexNet usando Base de Imagens 2

Os experimentos descritos nesta seção correspondem à rede AlexNet com Base de Imagens 2. Foram realizados os Experimentos 48 a 50, usando imagens de região ocular exclusiva do olho; e os Experimentos 51 e 52, usando imagens da face completa, variando o número de classes e número de amostras de treinamento e teste. Em todos os experimentos, as amostras da fase de treinamento foram distintas das amostras da fase de teste, e o número de épocas de treinamento adotado foi 10.

### Experimentos 48 a 50: Imagens da região ocular exclusiva do olho

No Experimento 48 foram consideradas 10 classes e 200 amostras por classe para treinamento. A taxa de acerto para o teste com 100 amostras por classe foi de 51,50%.

No Experimento 49 considerou-se 20 classes e 500 amostras por classe para treinamento. A taxa de acerto com o número de 100 amostras por classe foi de 48,80%.

No Experimento 50 considerou-se 20 classes e 700 amostras por classe para treinamento. A taxa de acerto com o número de 100 amostras por classe foi de 49,50%.

### **Experimentos 51 e 52: Imagens da face completa**

No Experimento 51 são consideradas 20 classes e uma fase de treinamento com 200 amostras por classe. A taxa de acerto com o número de 100 amostras por classe foi de 51,45%.

No Experimento 52 foram consideradas 20 classes e uma fase de treinamento com 500 amostras por classe. A taxa de acerto com o número de 100 amostras por classe foi de 56,45%.

#### **5.6.1 Síntese dos Experimentos usando AlexNet e Base de Imagens 2**

Na tabela 11 é mostrada uma síntese dos experimentos realizados com a rede AlexNet, onde na primeira linha (experim.) são mostrados os 5 experimentos realizados, numerados de 48 a 52. A segunda linha (imagens) descreve o tipo de imagens utilizada dentre região ocular, face completa (completa). A terceira linha descreve o número de classes relativo ao experimento, que variou entre 10 a 20 classes. A quarta linha refere-se ao número de amostras utilizadas para o treinamento, que variou entre 200 e 700. A quinta linha refere-se ao número de 100 amostras utilizadas para os testes. A sexta linha refere-se ao número de épocas utilizadas durante o treinamento. As sétima, oitava, nona e décima linhas referem-se aos métodos de avaliação usados nos experimentos, índice Kappa, erro relativo (ER), revocação e precisão respectivamente e finalmente na última linha são mostrados os resultados em taxa de acerto nos testes.

Nota-se que a taxa de acerto, tanto para imagens de região ocular exclusiva, quanto para imagens de face completa, foi variou de 48,80% a 56,45%. Essa taxa deve ser possível de ser melhorado com o aumento do número de épocas para o treinamento.

**Tabela 11 – Base de imagens 2 - Síntese dos experimentos com AlexNet**

<b>Experim.</b>	<b>48</b>	<b>49</b>	<b>50</b>	<b>51</b>	<b>52</b>
<b>imagens</b>	<b>Ocular</b>	<b>Ocular</b>	<b>ocular</b>	<b>Completa</b>	<b>completa</b>
<b>classes</b>	<b>10</b>	<b>20</b>	<b>20</b>	<b>20</b>	<b>20</b>
<b>treina</b>	<b>200</b>	<b>500</b>	<b>700</b>	<b>200</b>	<b>500</b>
<b>teste</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>épocas</b>	<b>30</b>	<b>30</b>	<b>30</b>	<b>30</b>	<b>50</b>
<b>Kappa</b>	<b>0,51</b>	<b>0,48</b>	<b>0,49</b>	<b>0,51</b>	<b>0,56</b>
<b>ER</b>	<b>0,94</b>	<b>0,70</b>	<b>0,63</b>	<b>0,73</b>	<b>0,67</b>
<b>Revocação</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
<b>Precisão</b>	<b>0,51</b>	<b>0,48</b>	<b>0,49</b>	<b>0,51</b>	<b>0,56</b>
<b>Taxa de acertos</b>	<b>51,50%</b>	<b>48,80%</b>	<b>49,50%</b>	<b>51,45%</b>	<b>56,45%</b>

## 5.7 Análise dos Resultados

Nesta seção será realizada uma análise geral de todos os experimentos, baseados nos resultados de teste, em que as amostras de treinamento foram distintas das amostras de teste.

Inicialmente, considera-se a tabela de resultados usando a Base de Imagens 1, para a rede Neocognitron, onde os melhores resultados dos experimentos com diferentes tipos de imagens foi reescrita como Tabela 12, a partir da Tabela 4.

**Tabela 12. Neocognitron e Base de Imagens 1**

<b>Experim.</b>	<b>Exp2</b>	<b>Exp7</b>	<b>Exp10</b>
<b>Imagens</b>	<b>ocular</b>	<b>Comp</b>	<b>Acess</b>
<b>Classes</b>	<b>10</b>	<b>5</b>	<b>5</b>
<b>Treina</b>	<b>50</b>	<b>25</b>	<b>25</b>
<b>Teste</b>	<b>50</b>	<b>25</b>	<b>25</b>
<b>Taxa de acertos</b>	<b>42,0%</b>	<b>80,0%</b>	<b>76,0%</b>

Os melhores casos foram do Experimento 7, amostras de face completa, e Experimento 10, amostras de região ocular com acessórios em que a taxa de acertos foi

de 80,0% e 76,0%, respectivamente. Para imagens de região ocular, a maior taxa obtida foi de 42%, valor que provavelmente pode ser melhorado aumentando o número de amostras de treinamento e de teste.

Os resultados do uso da rede LeNet e a Base de Imagens 1, com exceção do experimento 19, estão refeitos como Tabela 13, a partir da Tabela 6.

**Tabela 13. LeNet e Base de Imagens 1**

<b>Experim</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>18</b>	<b>21</b>	<b>22</b>
<b>imagens</b>	<b>ocular</b>	<b>Ocular</b>	<b>ocular</b>	<b>Ocular</b>	<b>comp</b>	<b>Acess</b>	<b>access</b>
<b>classes</b>	<b>10</b>	<b>10</b>	<b>10</b>	<b>13</b>	<b>5</b>	<b>5</b>	<b>7</b>
<b>Treina</b>	<b>50</b>	<b>100</b>	<b>150</b>	<b>195</b>	<b>25</b>	<b>25</b>	<b>49</b>
<b>Teste</b>	<b>50</b>	<b>30</b>	<b>50</b>	<b>65</b>	<b>25</b>	<b>25</b>	<b>21</b>
<b>épocas</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>
<b>Taxa de acertos</b>	<b>60%</b>	<b>80%</b>	<b>66%</b>	<b>76,9%</b>	<b>80%</b>	<b>76%</b>	<b>76%</b>

Para a rede LeNet, a maioria dos experimentos resultaram em valores de taxa de acerto acima de 60%, com exceção do Experimento 19, que teve uma taxa de acerto de 10%. Um possível motivo para essa discrepância em relação ao Experimento 19 deve ser o baixo número de amostras de treinamento e teste.

Os resultados dos experimentos com a rede AlexNet e Base de Imagens 1, exceto dos Experimentos 24 e 25, estão resumidos na Tabela 14, a partir da Tabela 8.

**Tabela 14. AlexNet e Base de Imagens 1.**

<b>Experim.</b>	<b>26</b>	<b>27</b>	<b>29</b>	<b>30</b>	<b>32</b>	<b>33</b>
<b>imagens</b>	<b>Ocular</b>	<b>ocular</b>	<b>comp</b>	<b>comp</b>	<b>Acess</b>	<b>access</b>
<b>classes</b>	<b>10</b>	<b>13</b>	<b>5</b>	<b>7</b>	<b>5</b>	<b>7</b>
<b>treina</b>	<b>150</b>	<b>195</b>	<b>25</b>	<b>49</b>	<b>25</b>	<b>49</b>
<b>Teste</b>	<b>50</b>	<b>65</b>	<b>25</b>	<b>21</b>	<b>25</b>	<b>21</b>
<b>épocas</b>	<b>20</b>	<b>20</b>	<b>20</b>	<b>20</b>	<b>20</b>	<b>20</b>
<b>Taxa de acertos</b>	<b>74%</b>	<b>92,3%</b>	<b>80%</b>	<b>88,8%</b>	<b>96%</b>	<b>88,8%</b>

Para a rede AlexNet e Base de Imagens 1, a maioria dos experimentos tiveram resultados de taxa de acertos acima de 74%, com exceção dos Experimentos 24 e 25, de 50% e 23,3%, respectivamente.

Considerando-se a aplicação das redes na Base de Imagens 2, a Tabela 15 refere-se aos resultados dos experimentos com o Neocognitron, com exceção dos resultados do Experimento 23.

**Tabela 15- Banco de imagens 2 - Síntese dos experimentos com o Neocognitron**

Experi m.	34	35	36	37	38	39	40	41
Imagens	Ocular	Ocular	ocular	ocular	ocular	Ocular	comp	comp
Classes	5	10	20	10	20	10	10	16
Treina	100	100	100	200	200	500	100	200
Teste	40	40	40	100	100	100	40	100
Kappa	0,44	0,45	0,64	0,82	0,49	0,70	0,63	0,44
ER	0,53	1,05	0,43	0,22	0,64	0,37	0,57	0,11
Revocaç ão	0,5	0,93	0,82	1	0,92	0,98	0,90	0,55
Precisão	0,87	0,49	0,77	0,82	0,78	0,71	0,65	0,78
Taxa de acertos	46,50%	46,75%	65,87%	82,00%	49,45%	69,40%	64,20%	47,87%

Nos experimentos com o Neocognitron e Base de Imagens 2, a taxa de acertos teve valores na faixa de 41,80% a 82,00%. A maior taxa de acerto se deu no Experimento 37, de 82,00% de taxa de acerto, em que usando imagens da região ocular exclusiva de um olho, o número de classes foi de 10, e no número de amostras de treinamento por classe foi de 200. O melhor resultado para imagens de face completa foi obtido no Experimento 40, com taxa de acerto de 64,20%, 10 classes e 100 amostras de treinamento por classe. Nota-se que o número de 10 classes é compatível com o tipo de implementação da rede Neocognitron usado, pois quando o número de classes é maior, a taxa de acerto diminui.

Além da taxa de acerto, outros índices foram computados para os experimentos da Base de Imagens 2. Assim, para o Experimento 37, o índice Kappa obtido foi de 0,82, o que significa “excelente”, conforme Tabela 2. Os outros índices obtidos foram o Erro Relativo de 0,22; revocação de 1; e precisão de 0,82. Para o Experimento 40, o índice Kappa foi de 0,63, considerado de significância ‘alta’ conforme tabela 2; o erro relativo foi de 0,57; revocação de 0,90; e precisão 0,65.

Os resultados da rede LeNet-5, quanto a Base de Imagens 2, foi sintetizada na Tabela 16, onde os melhores resultados de 48,55% e 42,15% para taxa de acerto foram obtidos, para as imagens de região ocular exclusiva e face completa, respectivamente.

**Tabela 16 - Base de imagens 2 - Síntese dos experimentos com o LeNet-5**

<b>Experimento</b>	<b>45</b>	<b>47</b>
<b>imagens</b>	<b>Ocular</b>	<b>Completa</b>
<b>classes</b>	<b>20</b>	<b>20</b>
<b>treina</b>	<b>700</b>	<b>500</b>
<b>teste</b>	<b>100</b>	<b>100</b>
<b>épocas</b>	<b>30</b>	<b>50</b>
<b>Kappa</b>	<b>0,48</b>	<b>0,42</b>
<b>ER</b>	<b>0,56</b>	<b>0,86</b>
<b>Revocação</b>	<b>1</b>	<b>1</b>
<b>Precisão</b>	<b>0,48</b>	<b>0,42</b>
<b>Taxa de Acertos</b>	<b>48,55%</b>	<b>42,15%</b>

Para a rede LeNet-5, a taxa de acertos diminuiu de forma geral em relação aos experimentos usando Base de Imagens 1. Uma justificativa para isso seria o número de épocas de treinamento de variou de 20 para 30, da Base de Imagens 1 para a Base de Imagens 2, enquanto o número de amostras de treinamento aumentou significativamente. Esse número de épocas de treinamento pode não ter sido suficiente para a Base de Imagens 2. Os índices Kappa foram 0,48 e 0,42, para imagens de região ocular e face completa, respectivamente, consideradas de significância ‘moderada’ conforme Tabela 2.

Os experimentos com a rede AlexNet e Base de Imagens 2 resultaram em taxa de acertos com poucas variações. Os melhores resultados para imagens de região ocular e face completa, são mostrados na Tabela 17.

**Tabela 17 – Base de imagens 2 - Síntese dos experimentos com AlexNet**

<b>Experim.</b>	<b>50</b>	<b>52</b>
<b>imagens</b>	<b>Ocular</b>	<b>Completa</b>
<b>classes</b>	<b>20</b>	<b>20</b>
<b>treina</b>	<b>700</b>	<b>500</b>
<b>teste</b>	<b>100</b>	<b>100</b>
<b>épocas</b>	<b>30</b>	<b>50</b>
<b>Kappa</b>	<b>0,49</b>	<b>0,56</b>
<b>ER</b>	<b>0,63</b>	<b>0,67</b>
<b>Revocação</b>	<b>1</b>	<b>1</b>
<b>Precisão</b>	<b>0,49</b>	<b>0,56</b>
<b>Taxa de acertos</b>	<b>49,50%</b>	<b>56,45%</b>

Para a rede AlexNet, a taxa de acertos também foi reduzido, em relação à Base de Imagens 1, da mesma forma que no LeNet. O número de épocas de treinamento variou de 20 para 30, em relação à Base de Imagens 1 e Base de Imagens 2, respectivamente, enquanto o número de amostras de treinamento aumentou substancialmente. Esse número de épocas de treinamento pode ter sido insuficiente para um melhor resultado. Os índices Kappa de 0,49 e 0,56, obtidos nos Experimentos 50 e 52, respectivamente, estão na faixa de significância ‘moderada’.

De uma forma geral, os resultados de teste utilizando a Base de Imagens 1 foi melhor em taxa de acerto do que os resultados de teste utilizando a Base de Imagens 2, porém os resultados usando a Base de Imagens 1 teve mais variações nos diversos experimentos provavelmente devido ao número reduzido de amostras de treinamento e teste. Uma síntese das melhores taxas de acertos dos experimentos com a Base de Imagens 1 está resumido na Tabela 18, e com a Base de Imagens 2, na Tabela 19.

Nota-se que para a Base de Imagens 1, a rede AlexNet teve melhores resultados nos três tipos de imagens. As redes Neocognitron e LeNet tiveram resultados coincidentes, com exceção das imagens da região ocular em que Neocognitron teve uma taxa de acerto de 42% enquanto LeNet teve 80%.

**Tabela 18 – Base de imagens 1- Melhores resultados das redes**

	redes	Neocognitron	LeNet	AlexNet
<b>Taxa de Acertos</b>	<b>Região Ocular</b>	<b>42%</b>	<b>80%</b>	<b>92,3%</b>
	<b>Face Completa</b>	<b>80%</b>	<b>80%</b>	<b>88%</b>
	<b>Região Ocular com Acessórios</b>	<b>76%</b>	<b>76%</b>	<b>96%</b>

Para a Base de Imagens 2, a rede Neocognitron teve melhores resultados nos dois tipos de imagens, enquanto que as redes LeNet e AlexNet tiveram um desempenho inferior, com ligeira vantagem do AlexNet em relação a LeNet. Esse resultado considerado apenas de significância ‘moderada’ pelo índice Kappa, provavelmente ocorreu devido ao número de épocas de treinamento, que deveria ser maior pelo número grande de amostras utilizado, tanto para as redes LeNet e AlexNet.

**Tabela 19 – Base de imagens 2- Melhores resultados das redes**

	redes	Neocognitron	LeNet	AlexNet
<b>Taxa de Acertos</b>	<b>Região Ocular</b>	<b>82%</b>	<b>48,55%</b>	<b>49,50%</b>
	<b>Face Completa</b>	<b>64,2%</b>	<b>42,15%</b>	<b>56,45%</b>

Para a rede Neocognitron, apesar do melhor resultado alcançado usando a Base de Imagens 2, ainda deve ser aperfeiçoado o treinamento. Uma possível forma de melhoria no treinamento seria na seleção de melhores atributos para treinamento e para

tal seriam usadas técnicas como filtro, *wrapper* e *embedded*. Na seção 6.1 tais técnicas serão descritas em detalhes.

Uma outra observação, seria quanto ao uso de diferentes tipos de imagens: região ocular incluindo ambos os olhos, região ocular exclusiva de apenas um olho, face completa e região ocular com acessórios. Aparentemente, os diferentes tipos de imagens não mostraram vantagens uma em relação à outra. Na seleção de melhores atributos para treinamento, podem surgir essas vantagens. No caso do Neocognitron, na Base de Imagens 1, o desempenho com imagens da região completa foi melhor em relação a outros tipos de imagens, mas no caso da Base de Imagens 2, o desempenho com imagens da região ocular foi melhor.

## 5.8 Aspectos Técnicos

Utilizou-se como plataforma de hardware para execução da aplicação, um notebook modelo Inspiron 15 da DELL, com processador Intel® Core™ i7; 8GB de memória RAM, 1TB de HD; Placa GPU NVidia GeForce GTX 950M.

Todos os simuladores utilizados nos experimentos rodaram sobre a plataforma do sistema operacional Windows 10 da Microsoft. Além disso os simuladores das redes LeNet-5 e AlexNet utilizaram algumas bibliotecas, entre elas destacam-se:

- NumPy: que suporta vetores e matrizes multidimensionais, possuindo uma larga coleção de funções matemáticas para trabalhar com estas estruturas;
- TensorFlow: biblioteca de código aberto para aprendizado de máquina aplicável a uma ampla variedade de tarefas, neste caso para as tarefas de aprendizado da rede neural.

- Keras: biblioteca de aprendizado profundo que suporta o uso de redes convolucionais e recorrentes ou a combinação de ambas.

Após as configurações necessárias para efetuar o carregamento das imagens da base de dados, deu-se início aos experimentos, onde o tempo médio entre cada experimento varia de acordo com a quantidade de amostras utilizadas nos mesmos. Os experimentos usando a rede convolucional AlexNet, possivelmente pelo fato de possuir mais camadas, consumiu maior tempo de processamento.

Quando o assunto é tempo, a rede Neocognitron teve melhor desempenho em relação às outras redes convolucionais utilizadas nos experimentos, tendo em vista que os estágios de treinamento e reconhecimento dessa rede são realizados separadamente.

Ressalto ainda que, foram feitos vários experimentos, afim de definir o número de épocas de treinamento que alcançasse maior taxa de acurácia e menor taxa de perda.

## **5.9 Considerações Finais**

Neste capítulo foram descritos os experimentos usando as redes Neocognitron, LeNet 5 e AlexNet, onde foram usadas imagens da região ocular da face e imagens da face completa com e sem acessórios (óculos). Ao final dos experimentos de cada rede foi gerada uma síntese dos experimentos com o intuito de fornecer uma visão geral dos experimentos realizados; e no final foi realizada uma análise geral.

## CAPÍTULO 6 - Conclusão

Neste trabalho foram apresentados resultados de desempenho das redes neurais convolucionais Neocognitron, LeNet e AlexNet, em reconhecimento de padrões de imagens faciais. Os experimentos foram realizados em duas etapas com bases de imagens diferentes:

Etapa 1: Experimentos realizados com a Base de Imagens 1, com um conjunto reduzido de amostras, da região ocular incluindo ambos os olhos, face completa, e região ocular com acessórios. Para essa base de imagens, a rede AlexNet alcançou melhores resultados nos três tipos de imagens. As redes Neocognitron e LeNet obtiveram resultados coincidentes, com exceção das imagens da região ocular em que Neocognitron teve a menor taxa de acerto de 42% enquanto LeNet teve 80%.

Etapa 2: Experimentos realizados com a Base de Imagens 2, com um conjunto de imagens maior em relação à Base de Imagens 1, da região ocular exclusiva de um dos olhos, e da face completa. Os experimentos usando a Base de Imagens 2 atingiram um desempenho em taxa de acerto menor em relação à Base de Imagens 1. A rede Neocognitron alcançou um desempenho melhor com essa base de imagens, enquanto que as redes LeNet e AlexNet atingiram desempenhos considerados apenas ‘moderado’ em relação ao índice kappa obtido.

Quanto ao uso de diferentes tipos de imagens: região ocular incluindo ambos os olhos, região ocular exclusiva de apenas um olho, face completa e região ocular com acessórios, os resultados não apresentaram vantagens muito grandes de um tipo de imagem em relação a outros. Como exemplo, no caso do Neocognitron, na Base de Imagens 1, o desempenho com imagens da região completa foi melhor em relação a outros

tipos de imagens, mas no caso da Base de Imagens 2, o desempenho com imagens da região ocular foi melhor.

### **6.1 Propostas de Trabalhos Futuros**

Nota-se que os resultados dos experimentos realizados variaram de acordo com o número de amostras de treinamento, número de classes e base de imagens diferentes. Assim, como forma de evoluir nos experimentos dois trabalhos futuros são propostos:

1) estudo em relação à seleção de amostras utilizadas nos experimentos. Existem basicamente três tipos de metodologias em relação à seleção de amostras ou atributos:

a) filtro, em que consiste de uma forma de pré-processamento das amostras para que o resultado do aprendizado seja otimizado;

b) *wrapper*, em que as amostras ou atributos sejam selecionadas com o auxílio do algoritmo de classificação;

c) *embedded*, em que a seleção de amostras ou atributos é realizada junto com o algoritmo de classificação (Vanaja et al., 2014).

A partir desse estudo, será possível verificar se os tipos de amostras utilizadas nos experimentos influenciam nos resultados.

2) aumento do número de épocas nos treinamentos das redes LeNet e AlexNet, e a verificação do desempenho.

### **6.2: Artigos Publicados ou Submetidos**

Como resultados desta dissertação foram gerados dois artigos sendo o primeiro no WCF 2019 (Workshop de Computação da UNIFACCAMP), com o título:

Redes Convolucionais para Reconhecimento Facial Através de Imagens de Regiões Oculares; e o segundo, submetido ao ICCS2020 (International Conference on Computer Science), com o título: Convolutional Neural Networks and Periocular Region Image Recognition.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Abdel-Hamid, O., Mohamed, A., Jiang, H., Deng, L., Penn, G. and Yu, D. (2014) Convolutional Neural Networks for Speech Recognition, in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*,. vol. 22, no. 10, pp. 1533-1545.
- Agarwal, M., Agrawal, H., Jain, N. and Kumar, M. (2010) Recognition Using Eigen Faces and Artificial Neural Network.. *International Journal of Computer Theory and Engineering*, Vol. 2, No. 4.
- Albuquerque, M.A., Silva, E.L., Barros, K.N.N.O. and Junior, S.F.A.X. (2016). Comparação entre Coeficientes de Similaridade em Aplicação em Ciências Florestais, *Matemática e Estatística em Foco*, v.4, n.2, p.102-114.
- AlexNet. Arquitetura AlexNet. [online] Disponível em: <https://engmrk.com/alexnet-implementation-using-keras/>. Acesso em março 2019.
- Almisreb, A. A., Jami, N. and Din, N. M., (2018). Utilizing AlexNet Deep Transfer Learning for Ear Recognition, *Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, Kota Kinabalu, 2018, pp. 1-5.
- Arbib, M. A., (1995). Introducing the Neuron, the handbook of brain theory and Neural Networks, MIT Cambridge , pp. 4-11,
- Arel, I., Rose, D.C. and Karnowski, T.P. (2010). Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier],.in *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13-18, Nov.
- Barcellos, W., Shitara, N.H., Ferraz, C.T., Queiroga, R.T.V.; Saito, J.H. and Gonzaga, A. (2019). Evaluation of Fine Tuning and Feature Extraction Methods in Biometric

- Periocular Recognition, Proceedings do XV Workshop de Visão Computacional (WVC), pp. 9-11 de setembro de 2019.
- Bengio, Y. and LeCun, Y., (2007). *Scaling Learning Algorithms towards AI*, MIT Press
- Bianchini, A. R., (2001). *Arquitetura de redes neurais para o reconhecimento facial baseado no Neocognitron*, Dissertação de mestrado - Universidade Federal de São Carlos
- Cardoso, A. and Wichert, A., (2010). Neocognitron and the map transformation cascade, *Neural Networks*, vol. 23, no. 1, pp. 74–88,.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales Educational and Psychological Measurement, *Educational and Psychological Measurement* , Vol. 20, no. 1, pp.37–46
- Crihalmeanu, S. and Ross, A., (2012). Multispectral scleral patterns for ocular biometric recognition. *Pattern Recognition Letters*. 33. 1860–1869
- Dalapicola, R.C., Queiroga, R.T.V., Ferraz, C.T., Borges, T.T.N., Saito, J.H., Gonzaga, A. (2019). Impact of facial expressions on the accuracy of a CNN performing periocular recognition. 8th Brazilian Conference on Intelligent Systems (BRACIS).
- Elaiwat, S., Bennamoun, M., Boussaid, F. and El-Sallam, A. (2015). A Curvelet-based approach for textured 3D face recognition. *Pattern Recognition* 10.1016/j.patcog.2014.10.013..
- Ertugrul, O. F., Tekin, R. and Kaya, Y., (2017). Randomized feed-forward artificial neural networks in estimating short-term power load of a small house: a case study. *International Artificial Intelligence and Data Processing Symposium (IDAP)*

Essex – Description of the Collection of Facial Images. Disponível em:

<https://dces.essex.ac.uk/mv/allfaces/>. Acessado em janeiro de 2019.

Fan, D., Shim, Y., Raghunathan, A. and Roy, K, (2015). STT-SNN: A Spin-Transfer-Torque Based Soft-Limiting Non-Linear Neuron for Low-Power Artificial Neural Networks . IEEE Transactions On Nanotechnology, Vol. 14, No. 6, N

Fukushima, K., (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybernet. 36, pp. 193–202,

Fukushima, K., (1982). Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts position, Pattern Recognition, Vol 15, N. 6, pp. 455-469,

Fukushima, K., (1988). Neocognitron: Original Contribution. A Hierarchical Neural Network Capable of Visual Pattern Recognition, -Vol. 1, pp. 119-130

Fukushima, K. and Wake, N. (1991). Handwritten Alphanumeric Character Recognition by the Neocognitron. IEEE Transactions on Neural Networks, 2 (3), 355-365. doi: 10.1109 / 72.97912

Fukushima, K. and Shouno, H. (1998). Neocognitron with improved bend-extractors: Recognition of handwritten digits in the real world. Neural Computing and Applications. 7. 260-272. 10.1007/BF01414887.

Fukushima, K., (2011). Increasing robustness against background noise: Visual pattern recognition by a neocognitron. Neural networks : the official journal of the International Neural Network Society. 24. 767-78. 10.1016/j.neunet.2011.03.017.

- Fukushima, K., Hayashi, I. and Léveillé, J. (2011). Neocognitron Treinado por Vencedor-Matador-Perdedor com Triplo Limiar. In: *Processamento de Informação Neural. ICONIP 2011. Notas de Aula em Ciência da Computação*, vol. 7063. Springer, Berlin, Heidelberg
- Fukushima, K. and Shouno, H., (2015). Deep convolutional network neocognitron: Improved Interpolating-Vector, International Joint Conference on Neural Networks (IJCNN), Killarney, 2015, pp. 1-8.
- Fukushima, K., (2016). Margined Winner-Take-All: New learning rule for pattern recognition, International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, pp. 977-984.
- Gorokhovatskyi, O., (2016). Neocognitron as a tool for optical marks recognition. *IEEE First International Conference on Data Stream Mining & Processing (DSMP)*, 169-172.
- Guo, D. and Liu, Y. (2017) Research on Trim of Multilayer Feedforward Small World Network Based on E-exponential Information Entropy.
- Hayat, S., Kun, S., Tengtao, Z., Yu, Y., Tu, T. and Du, Y., (2018). A Deep Learning Framework Using Convolutional Neural Network for Multi-class Object Recognition 3rd IEEE International Conference on Image, Vision and Computing
- Haykin S., *Redes Neurais, princípios e práticas.* (2001). Porto Alegre: Bookman.
- Hubel, D.H. and Wiesel, T.N., (1977). Functional Architecture of macaque monkey visual cortex, *Proc. Royal Society, London.*
- Huang, F. J. and LeCun, Y. (2006). Large-scale Learning with SVM and Convolutional for Generic Object Categorization, *IEEE Computer Society Conference on*

- Computer Vision and Pattern Recognition (*CVPR'06*), New York, NY, USA, 2006, pp. 284-291.
- Huang, G. B., Ramesh, M., Berg, T. and Learned-Miller, E., (2007). Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49, October,.
- Johnson, R. and Zhang, T., (2015). Semi-supervised Convolutional Neural Networks for Text Categorization via Region Embedding. Advances in neural information processing systems. pp. 919-927. Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1
- Kangin, D.; Kolev, G. and Vikhoreva, A. (2012). Further Parameters Estimation of Neocognitron Neural Network Modification with FFT Convolution. . Journal of Telecommunication, Electronic and Computer Engineering (JTEC), v. 4, n. 2, p. 21-26.
- Kangin D., Kolev G. and Angelov P. (2013). Vehicle Plate Recognition Using Improved Neocognitron Neural Network. In: Mladenov V., Koprinkova-Hristova P., Palm G., Villa A.E.P., Appollini B., Kasabov N. (eds) Artificial Neural Networks and Machine Learning –. Lecture Notes in Computer Science, vol 8131. Springer, Berlin, Heidelberg
- Komyakov, A. A., Nikiforov, M. M., Erbes, V. V; e Cheremisin, V. T. and Ivanchenko, V., I. (2016). Construction of electricity consumption mathematical models on railway transport used artificial neural network and fuzzy neural network,

- IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC), Florence, pp. 1-4
- Krizhevsky, A., Sutskever, I. and Geoffrey E.H., (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems., pages 1106–1114.
- Le, Q. V., Ngiam, J., Chen, Z., Chia, D., Koh, P.W., and Ng, A. Y., (2015). Tiled convolutional neural networks. Advances in Neural Information Processing Systems Conference
- Lecun, Y., Bengio, Y., Bottou, L., and Haffener, P., (1998). Gradient-based learning applied to document recognition. In Proceedings of the IEEE, pp.2278–2324.
- Lecun, Y., Bengio, Y. and Hinton, G. (2015). Deep Learning. Nature Vol 521 pp 436-444
- Lee, T. S., Mumford, D., Romero, R. and Lamme, V. A. F., (1998). The role of the primary visual cortex in higher level vision. Vision Research 38 2429–2454
- Lee, Y.-P. (2017). Multidimensional Hebbian Learning With Temporal Coding in Neocognitron Visual Recognition. IEEE Transactions On Systems, Man, And Cybernetics: Systems
- LeNet-5. Arquitetura clássica da CNN. [online] Disponível em: <https://engmrk.com/lenet-5-a-classic-cnn-architecture/> . Acessado em janeiro de 2019
- Li, E., Xia, J., Du, P., Lin, C. and Samat, A., (2017) Integrating Multilayer Features of Convolutional Neural Networks for Remote Sensing Scene Classification, in IEEE

- Transactions on Geoscience and Remote Sensing,. *vol. 55, no. 10, pp. 5653-5665, Oct. 2017.*
- Lihao, W. and Yanni, D. (2018). A Fault Diagnosis Method of Tread Production Line Based on Convolutional Neural Network..
- Lovell, D. R.; Downs, T. and Tsoi, A. C., (1997). An Evaluation of the Neocognitron, IEEE Transactionson Neural Networks, Vol 8, N. 5
- Mazade, R.E, Eggers, E.D . Mazade, RE, and Eggers, ED . (2016). A adaptação à luz altera a inibição retiniana interna para moldar a sinalização da via retiniana. Journal of Neurophysiology, 115 (6), 2761-2778. doi: 10.1152 / jn.00948.2015
- McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of mathematical biology. 52. 99-115; discussion 73. 10.1007/BF02459570.
- Mikaeil, A. M., Hu, W. and Hussain, S. B. (2018). A Low-Latency Traffic Estimation Based TDM-PON Mobile Front-Haul for Small Cell Cloud-RAN Employing Feed-Forward Artificial Neural Network - ICTON 2018
- Miyazaki, A. and Yamada, T., (1993). A Systematic Synthesis Procedure for Feedforward Neural Networks by Using the GRBF (Generalized Radial Basis Function) Network Technique. Proceedings of 1993 International Joint Conference on Neural Networks
- Mousa, A. and Yusof, Y. (2015). Feature Extraction using Neocognitron Learning in Hierarchical Temporary Memory. IEEE International Conference on Computer, Communication, and Control Technology (I4CT 2015), April 21 - 23 in Imperial Kuching Hotel, Kuching, Sarawak, Malaysia

- Musakulova, Z., Mirkin, E., and Savchenko, E., (2018). Synthesis of the Backpropagation Error Algorithm for a Multilayer Neural Network with Nonlinear Synaptic Inputs, IEEE International Conference on Electrical Engineering and Photonics (EExPolytech), St. Petersburg, pp. 131-135.
- Myler, H. R., (1997). Application of the Neocognitron to Target Identification. in: Proceedings of SPIE - The International Society for Optical Engineering
- Poli, G., Saito, J. H. Mari, J. F. and Zorzan, M. R. (2008). Processing Neocognitron of Face Recognition on High Performance Environment Based on GPU with CUDA Architecture, 20th International Symposium on Computer Architecture and High Performance Computing, Campo Grande, MS, pp. 81-88.
- Saito, J. H.; Carvalho, T. V.; Hirakuri, M.; Saunite, A.; Ide, A. N. and Abib, S. (2005). Using cmu-pie human face database to a convolutional neural network - Neocognitron. - In: Proceedings of European Symposium on Artificial Neural Networks,
- Shen, W. and Wang, W. (2018). Node Identification in Wireless Network Based on Convolutional Neural Network, 14th International Conference on Computational Intelligence and Security (CIS), Hangzhou, pp. 238-241.
- Shi, D., and Tan, C. L., (2000). GA-based supervised learning of Neocognitron, Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN. Neural Computing: New Challenges and Perspectives for the New Millennium, Como, Italy, pp. 559-564 vol.6.

- Sim, T.; Baker, S.; and Bsat, M. (2002). The CMU Pose, Illumination, and Expression (PIE) Database. Proceedings of the Fifth International Conference on Automatic Face and Gesture Recognition (FGR'02), IEEE Computer Society, 2002.
- Srivastava, N. Hinton, G.E., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research
- Steck, J.E. and Dalton, J. S. (1992). Modeling the neurodynamics of a biological neuron using an artificial feedforward neural network. IJCNN Joint International Conference on Neural Networks.
- Sun, Y.; Chen, Y.; Wang, X; and Tang, X. (2014). Deep Learning Face Representation by Joint Identification-Verification, Proc. Adv. Neural Information and Processing Systems, pp. 1988-1996.
- Taigman, Y.; Yang, M.; Ranzato, M.; and Wolf, L. (2014). Deepface: Closing the Gap to Human-Level Performance in Face Verification, Proc. IEEE Conf. Comput. Vision and Pattern Recognition, pp. 1701-1708.
- Teo, M.-Y. and Sim, S.-K. (1995) Training the neocognitron network using design of experiments. Artificial Intelligence in Engineering.. Artificial Intelligence in Engineering 9, 85-94
- Ting, C. and Chuang, K., (1993). An Adaptive Algorithm for Neocognitron to Recognize Analog Images NeuralNetworks. Vol. 6, pp. 285-299
- Tissot, H., Camargo, L. and Pozo, A., (2012). Treinamento de redes neurais feedforward: comparativo dos algoritmos backpropagation e differential evolution. Conference: Brazilian Conference on Intelligent Systems

- Vanaja, S. and Kumar, K.R.(2014). Analysis of Feature Selection Algorithms on Classification: A Survey, *International Journal of Computer Applications* (0975-8887) Vol. 96, n.17, junho 2014, pp. 28-35.
- Zhang, S. Wu,T., Shi, M. and Qi, S. (2018). Construction and Localization of Multi Resolution Map Based on Vision. 10th International Conference on Intelligent Human-Machine Systems and Cybernetics 978-1
- Zhou, E.; Cao, Z.; and Yin, Q. – (2015). Naive-Deep Face Recognition: Touching the Limit of LFW Benchmark or not, *CoRR*, [Online]. Disponível:  
<http://arxiv.org/abs/1501.04690>. Acessado em maio de 2019.
- Wan, L., Li, D., Wee, W. G., Han, C. -Y. and Porembka, D. T., (1991). ANN approach for 2-D echocardiographic image processing— Application of neocognitron model to LV boundary formation, in *Applications of Artificial Neural Networks II*, vol. 1469
- White, BA and Elmasry, MI. (1992) O digi-neocognitron: um modelo de rede neural neocognitron digital para VLSI. *IEEE Transactions on Neural Networks*, 3 (1), 73–85.
- Wijnhoven, R. G. J. and With,P. H. N. (2010). Fast Training of Object Detection Using Stochastic Gradient Descent, 20th International Conference on Pattern Recognition, Istanbul, pp. 424-427.
- Willoughby, C.E, Ponzin, D., Ferrari, S., Lobo, A., Landau, K., e Omid, Y. (2010). Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function – a reviewceo\_23 *Clinical and Experimental Ophthalmology* 2010; 38: 2–11

Yi, D., Lei, Z., Liao, S. and Li, S.Z. (2014). Learning Face Representation from Scratch. CoRR, [Online]. Disponível: <http://arxiv.org/abs/1411.7923>

## APÊNDICE 1 – Base de imagens 1

### Região ocular incluindo ambos os olhos



Figura A-1 – Imagens sobre a região ocular

## APÊNDICE 2 – Base de imagens 1

### Imagens Faciais Completas

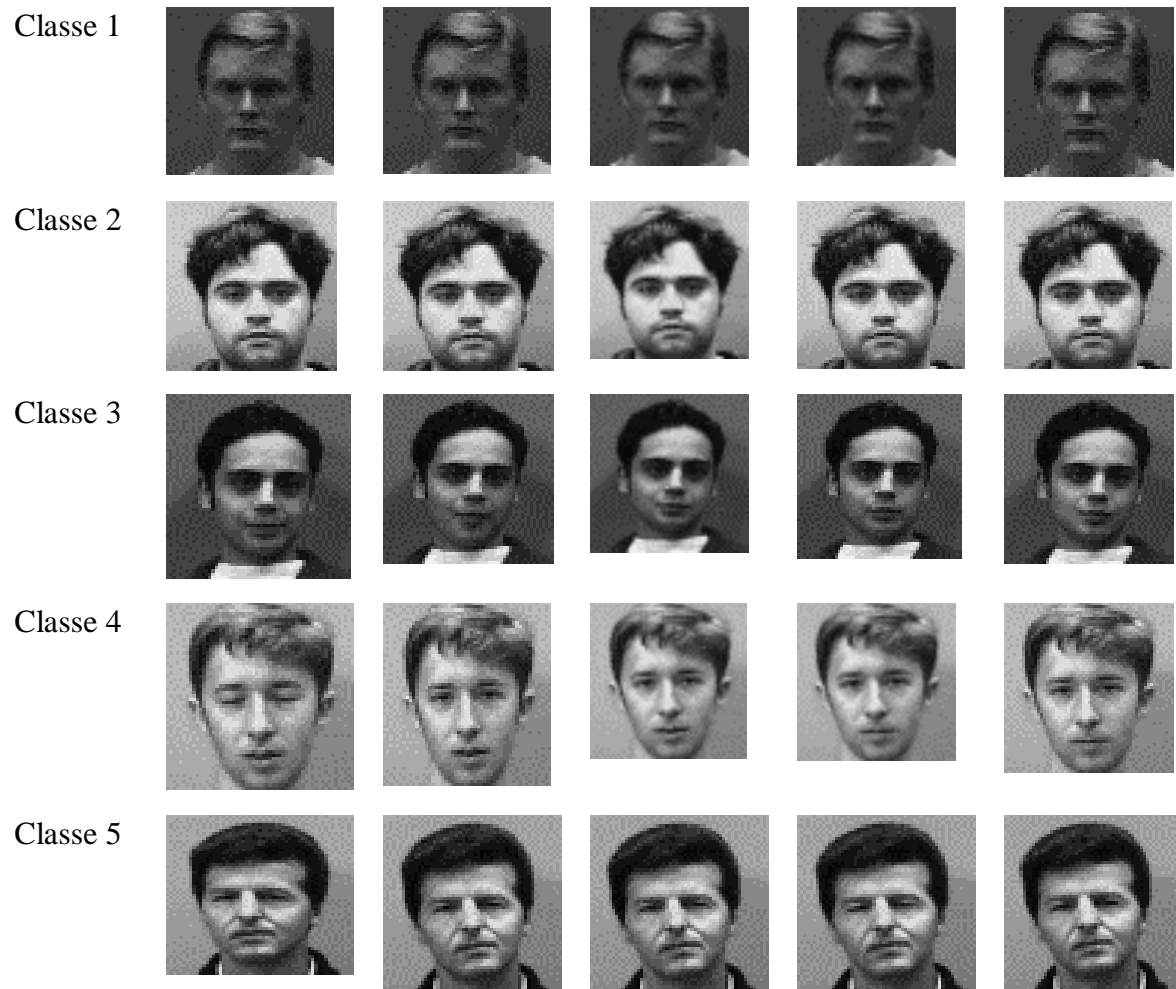


Figura A-2 -Amostras imagens de faces completas

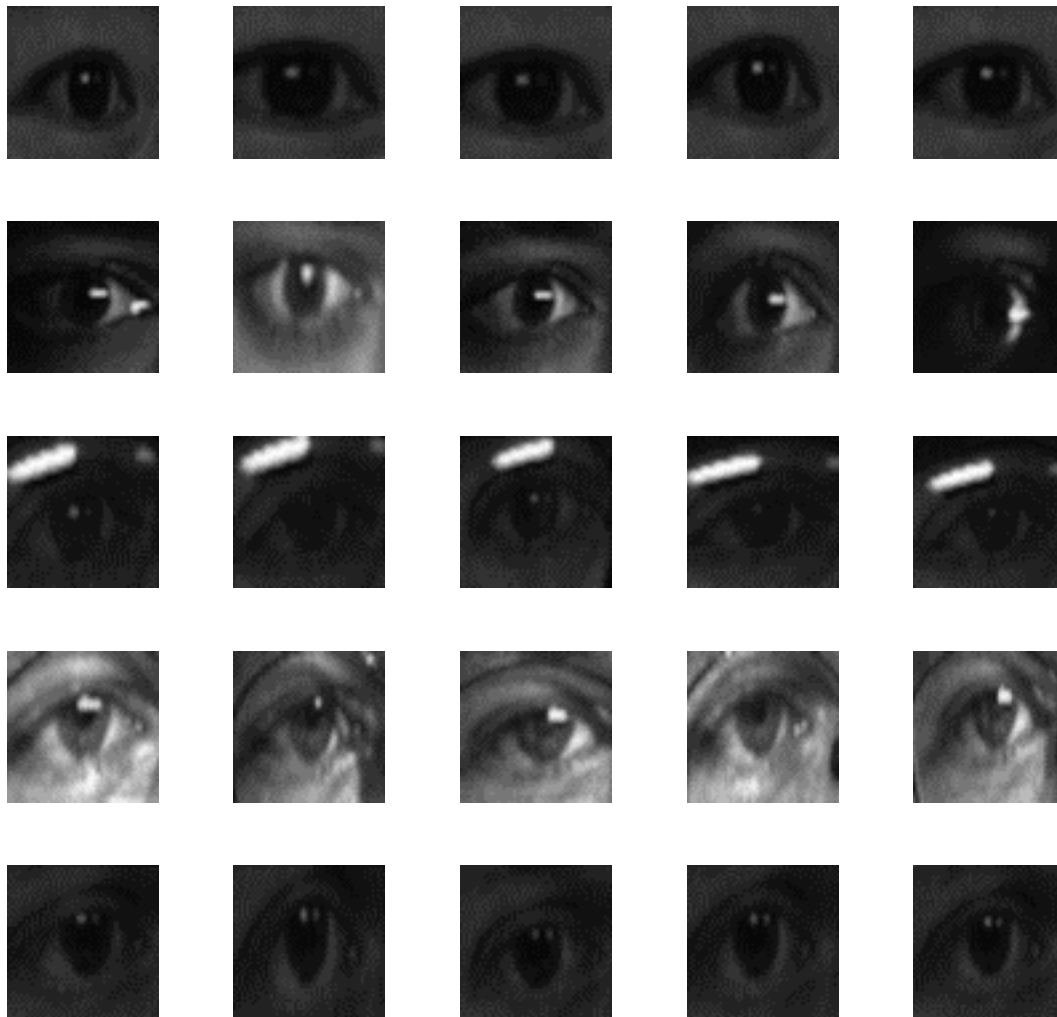
## APÊNDICE 3 – Base de imagens 1

### Região ocular incluindo ambos os olhos e acessórios



Figura A-3 – Amostras de imagens de região ocular da face com acessórios

**APÊNDICE 4 – Base de imagens 2**  
**Imagens da Região Ocular do Olho Direito**



**Figura B-1 – Imagens sobre a região ocular da face com e sem acessórios**

## APÊNDICE 5 – Base de imagens 2

### Imagens Faciais Completas

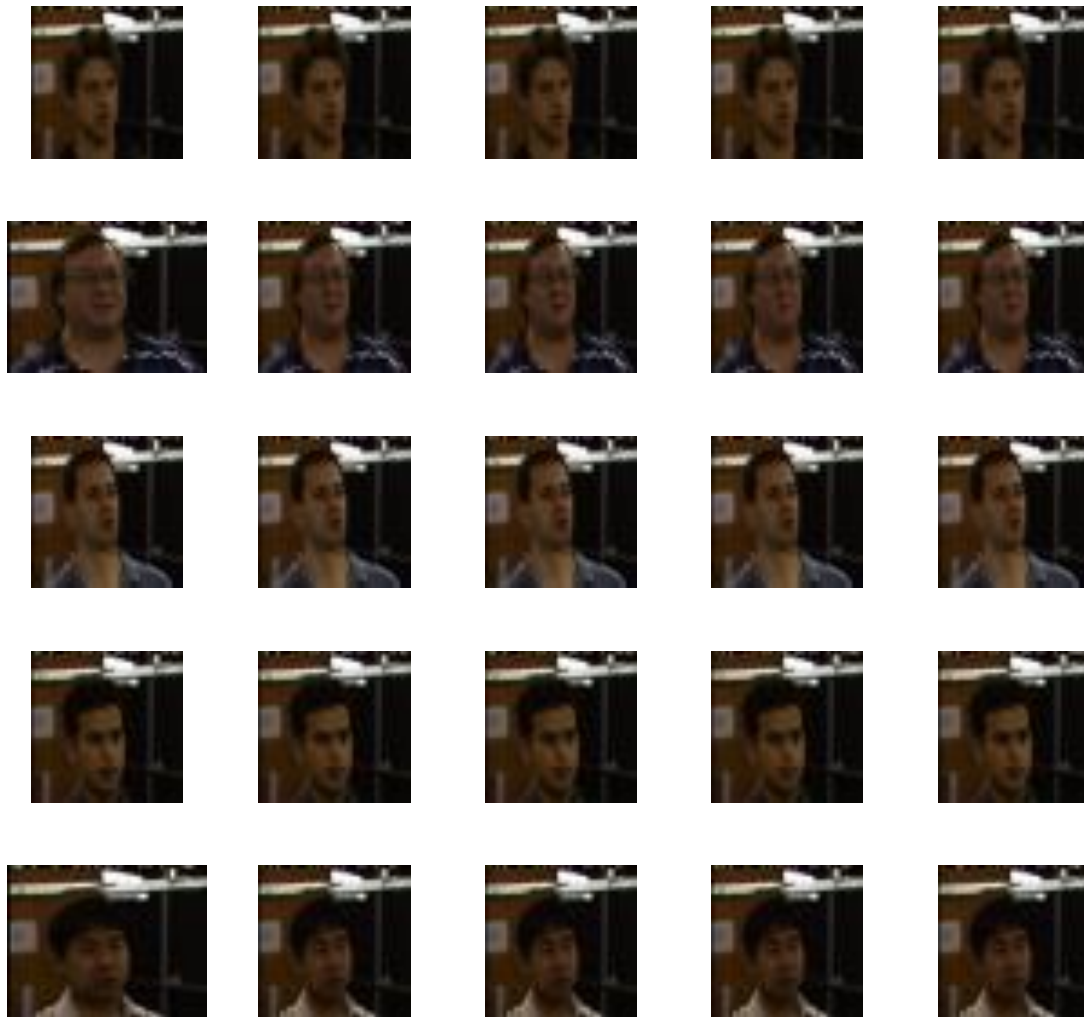


Figura B-2 – Imagens sobre a face completa com e sem acessórios