



UEPB

**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS ANTÔNIO MARIZ – CAMPUS VII
CENTRO DE CIÊNCIAS EXATAS E SOCIAIS APLICADAS
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

RICARDO DE SOUSA FARIAS

**APLICAÇÃO PARA DETECÇÃO E RECONHECIMENTO DE EXPRESSÕES
FACIAIS COM REDES NEURAS CONVOLUCIONAIS**

**PATOS - PB
2019**

RICARDO DE SOUSA FARIAS

**APLICAÇÃO PARA DETECÇÃO E RECONHECIMENTO DE EXPRESSÕES
FACIAIS COM REDES NEURAIS CONVOLUCIONAIS**

Trabalho de Conclusão de Curso apresentado ao Curso em Bacharelado em Ciência da Computação da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Ciência Computação.

Área de concentração: Visão Computacional.

Orientador: Profa. Dra. Jannayna Domingues Barros Filgueira.

**PATOS - PB
2019**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

F224a Farias, Ricardo de Sousa.
Aplicação para detecção e reconhecimento de expressões faciais com redes neurais convolucionais [manuscrito] / Ricardo de Sousa Farias. - 2019.
64 p. : il. colorido.
Digitado.
Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências Exatas e Sociais Aplicadas , 2019.
"Orientação : Profa. Dra. Jannayna Domingues Barros Filgueira , Coordenação do Curso de Computação - CCEA."
1. Redes neurais convolucionais. 2. Visão computacional.
3. Detecção de face. I. Título

21. ed. CDD 005.1

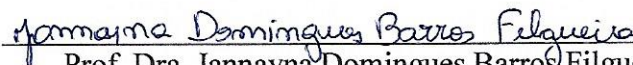
Ricardo de Sousa Farias

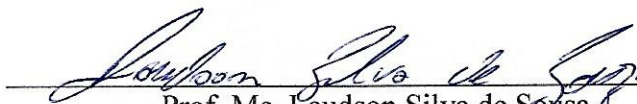
**APLICAÇÃO PARA DETECÇÃO E RECONHECIMENTO DE EXPRESSÕES FACIAIS
COM REDES NEURAIIS CONVOLUCIONAIS**


Trabalho de Conclusão de Curso apresentado ao
Curso de Bacharelado em Ciências da
Computação da Universidade Estadual da
Paraíba, em cumprimento à exigência para
obtenção do grau de Bacharel em Ciência da
Computação.

Aprovado em 25/11/2019

BANCA EXAMINADORA


Prof. Dra. Jannayna Domingues Barros Filgueira
(Orientador)


Prof. Me. Laudson Silva de Sousa
(Examinador)


Prof. Esp. Fábio Júnior F. da Silva
(Examinador)

Ao meu pai, à minha mãe, à minha irmã e amigos, pelo apoio, confiança, incentivo e amizade. DEDICO.

AGRADECIMENTOS

Primeiramente a Deus, pela saúde, determinação e força para superar as dificuldades no decorrer de minha vida, e por sempre atender aos meus pedidos, obrigado por mais essa vitória, concluo aqui mais uma etapa de minha vida.

Aos meus pais, Manoel Farias de Sousa e Ines Maria de Sousa Farias, por todo amor, amizade, presença, conselhos. Por tudo que fizeram por mim no decorrer da minha vida.

A minha irmã, Rita, por sempre acreditar em mim.

Aos meus familiares, amigos e colegas de classe e de ônibus pelos momentos de amizade e apoio, em especial aos meus amigos Fabio, Kaique, Hoffmann, João Paulo, Luís, Paulo, Juninho, Pedro, e ao meu primo Jonas por sempre me ajudar.

A minha orientadora, Jannayna, por todo incentivo, amizade e paciência.

Aos professores do Curso de Computação da UEPB, que contribuíram ao longo desses anos, por meio das disciplinas, conversas, ensinamentos e conselhos, como também aos meus professores do pré-escolar, ensino fundamental e ensino médio que contribuíram para pessoa que sou hoje.

Aos funcionários da UEPB, pela presteza e atendimento quando nos foi necessário.

Aos motoristas que nos transportavam até a universidade pelo os serviços prestados.

A todos que contribuíram de alguma forma em minha vida. Agradeço.

“Se sentir que chegou ao seu limite, lembre-se do motivo pelo qual você cerra os punhos, lembre-se porque resolveu trilhar este caminho e permita que essa memória o carregue além de seus limites.”

(All Might)

RESUMO

As técnicas de visão computacional são capazes de extrair das imagens parâmetros que são utilizados para caracterizar as emoções de uma pessoa por meio das expressões faciais. Neste aspecto, esta pesquisa apresenta o desenvolvimento de uma aplicação capaz de reconhecer as expressões emitidas por usuários utilizando Haar Cascade para detectar face e Rede Neural Convolutiva para reconhecer as expressões faciais. Para isso, a imagem da face dos usuários é capturada com auxílio de uma webcam do computador ou vídeo gravado, além do vídeo é possível reconhecer as expressões faciais em imagens. A aplicação consegue reconhecer as expressões, feliz, medo, raiva, triste, nojo, surpresa e neutra. Para treinar a Rede Neural Convolutiva foi utilizado a base de imagens Fer2013 e as bases JAFFE, CK+ e FacesDB para realização dos testes. Como também foi realizado um teste de vídeo em tempo real com 12 voluntários.

Palavras-Chave: Redes Neurais Convolutivas. Visão computacional. Detecção de face.

ABSTRACT

Computer vision techniques are capable of extracting from images parameters that are used to characterize a person's emotions through facial expressions. In this regard, this research presents the development of an application capable of recognizing the expressions emitted by users using Haar Cascade to detect face and Convolutional Neural Network to recognize facial expressions. For this, the image of the user's face is captured with the aid of a computer webcam or recorded video, in addition to the video it is possible to recognize facial expressions in images. The application can recognize the expressions happy, fear, anger, sad, disgust, surprise and neutral. To train the Convolutional Neural Network was used the Fer2013 image base and the bases JAFFE, CK + and FacesDB to perform the tests. As well as a real time video test with 12 volunteers.

Keywords: Convolutional Neural Networks. Computer vision. Face Detection.

LISTA DE FIGURAS

Figura 1 –	Expressões Faciais: neutra, feliz e triste	17
Figura 2 –	Expressões Faciais: raia e medo	17
Figura 3 –	Expressões Faciais: surpresa e nojo	18
Figura 4 –	Áreas que engloba a visão computacional	20
Figura 5 –	Esquema de um Sistema de visão computacional	22
Figura 6 –	Reconhecimento de objetos	24
Figura 7 –	Haar Feature	25
Figura 8 –	Detectando face com técnicas computacionais	27
Figura 9 –	Esquema de um sistema de reconhecimento de expressão facial	28
Figura 10 –	Neurônio artificial	30
Figura 11 –	Arquitetura de uma CNN demonstrando a divisão da extração de características e classificação	32
Figura 12 –	Aplicando o Kernel em uma imagem	33
Figura 13 –	Criando um mapa de características a partir do Kernel	35
Figura 14 –	Como funciona o max pooling	37
Figura 15 –	Exemplo de expressões da base Fer2013	39
Figura 16 –	Arquitetura Mini Xception	41
Figura 17 –	Base de Imagens JAFFE	43
Figura 18 –	Expressões do tipo feliz, base FacesDB	43
Figura 19 –	Expressões do tipo triste, base CK+	44
Figura 20 –	Esquema reconhecendo expressões faciais	45
Figura 21 –	Reconhecendo a expressão neutra em tempo real	46
Figura 22 –	Classificando as expressões triste, surpreso, raiva e feliz	52
Figura 23 –	Variação da expressão neutra, detectando corretamente e de forma errada	53
Figura 24 –	Reconhecendo a expressão feliz e errando devido a aparição dos dentes	54

LISTA DE GRÁFICOS

Gráfico 1 – Quantidade de imagens de cada expressão da base Fer2013	40
Gráfico 2 – Precisão expressão neutra	48
Gráfico 3 – Precisão expressão feliz	48
Gráfico 4 – Precisão expressão triste	49
Gráfico 5 – Precisão expressão raiva	49
Gráfico 6 – Precisão expressão medo	50
Gráfico 7 – Precisão expressão surpresa	51
Gráfico 8 – Precisão expressão nojo	51
Gráfico 9 – Comparando os resultados da expressão neutra	55
Gráfico 10 – Comparando os resultados da expressão feliz	56
Gráfico 11 – Comparando os resultados da expressão triste	56
Gráfico 12 – Comparando os resultados da expressão raiva	57
Gráfico 13 – Comparando os resultados da expressão medo	58
Gráfico 14 – Comparando os resultados da expressão surpresa	58
Gráfico 15 – Comparando os resultados da expressão nojo	59

LISTA DE TABELAS

Tabela 1 – Quantidade de expressões capturadas pela aplicação	53
---	----

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Objetivos	14
1.1.1	Objetivo geral	14
1.1.2	Objetivos específicos	15
1.2	Justificativa	15
2	REVISÃO BIBLIOGRÁFICA	16
2.1	Emoção básicas nas expressões faciais	16
2.2	Computação afetiva	18
2.3	Imagem digital	19
2.4	Visão computacional	19
2.5	Extração de características	22
2.6	Reconhecimento de padrões em imagens	23
2.7	Detecção de face	24
2.8	Reconhecimento de expressões faciais	27
2.9	Redes Neurais Artificiais	29
2.10	Redes Neurais Convolucionais	31
2.10. 1	Camada de convolução	32
2.10. 2	Camada de pooling	36
2.10. 3	Camada totalmente conectadas	37
2.11	Trabalhos relacionados	37
3	METODOLOGIA	39
3.1	Bases de imagens	42
3.1.1	JAFEE	43
3.1.2	FacesDB	43
3.1.3	CK+	44
3.2	Captura de vídeo em tempo real e gravado	44
4	RESULTADOS E DISCUSSÕES	47
4.1	Testes com imagens	47
4.2	Testes com voluntários	52
4.3	Análise dos testes realizados	55
5	CONCLUSÃO	60

5.1	Trabalhos futuros.....	60
	REFERENCIAS.....	61
	ANEXO A – TERMO DE AUTORIZAÇÃO DE IMAGEM	64

1 INTRODUÇÃO

As emoções humanas podem ser percebidas de diferentes formas. A face é uma das maneiras que se consegue entender o sentimento que o ser humano está sentido em determinado momento. Isso acontece pelos movimentos musculares que alteram a face de acordo com o sentimento naquele dado momento. Cada emoção possui movimentos musculares singulares que são utilizados para identificar a emoção, sendo possível conhecer facilmente quando uma pessoa está feliz ou triste. Mas, para isso é necessário que as emoções sejam representadas corretamente pelas expressões faciais.

As expressões faciais é a forma mais significativa e forte que ocorre na comunicação para que possa conhecer o estado emocional de uma pessoa durante a comunicação (PARADA, 2017). É uma forma de comunicação não-verbal que complementa a comunicação verbal.

Em um diálogo é possível entender as emoções que acontecem em determinados momentos pelas expressões faciais, seja quando acontece algo engraçado ou triste. As emoções acontecem de forma inconsciente, sendo possível notar a mudança da emoção de acordo com as expressões.

Ekman e Friesen (1976), classifica sete tipos de expressões, como expressões básicas universais: alegria, raiva, tristeza, medo, nojo, surpresa e neutro. Para determinar essas expressões, Ekman (2011), pesquisou por vários anos em diferentes tipos etnias, utilizando até mesmo povos que não tinham acesso a nenhum tipo de comunicação com o mundo exterior, para comprovar que essas expressões são universais e que em todos os povos as expressões são mesmas.

Desta forma, estudos vêm sendo feitos para que as máquinas possam entender as emoções por meio das expressões faciais e classifica-las. Mas, para uma máquina ainda é considerado complexo reconhecer expressões (SANTIAGO, 2017). Com base nisso a computação tem buscado formas que auxiliem nesse processo.

A computação afetiva busca analisar e entender as emoções que as pessoas sentem quando utilizam um determinado produto. Segundo De Sousa *et al.*, (2016), a computação afetiva se divide em duas perspectivas, inserir emoções humanas em máquinas e reconhecer emoções humanas por máquinas na interação entre humano e computador.

Sendo assim, a visão computacional busca auxiliar nesta tarefa. Por meio de imagens, extrai informações e faz com que uma máquina possa entender um determinado cenário, utilizando métodos computacionais. Essas informações podem variar em diversos tipos como cor, textura, formas e padrões. Portanto, algoritmos são capazes de capturar faces humanas e reconhecer expressões por meio dos padrões que uma determinada expressão possui, sendo possível classificá-la. Mas entender as emoções não é nada fácil, pesquisas vêm sendo feitas trazendo técnicas que auxiliem nessa tarefa, mas ainda há muito a evoluir.

Esta pesquisa tem o objetivo de desenvolver uma aplicação que auxilie e automatize, por meio da visão computacional, o reconhecimento de expressões faciais, utilizando técnicas de detecção de face e reconhecimento de características faciais, com o intuito de diminuir o tempo de análise dos dados obtidos por meio de vídeo. Desta forma, foram analisadas expressões feitas pelo usuário em vídeo gravados ou em tempo real. Para isto, utilizou-se a base de imagens Fer2013. As imagens disponíveis nesta base foram submetidas ao treinamento utilizando Redes Neurais Convolucionais.

Para validar a aplicação, dois tipos de testes foram realizados. O primeiro foi utilizando 3 bases de imagens, e o segundo utilizou-se de 12 voluntários que reproduziam as 7 expressões básicas, a aplicação reconhecia e armazenava as mesmas em um diretório específico de cada expressão.

1.1 Objetivos

Com base na proposta desta pesquisa, foram definidos os seguintes objetivos.

1.1.1 Objetivo geral

Desenvolver uma aplicação para reconhecimento e classificação de expressões faciais contidas em imagens e vídeos em tempo real ou gravado.

1.1.2 Objetivos específicos

Para alcançar o objetivo geral, foram necessários os seguintes objetivos específicos.

- Realizar estudo bibliográfico sobre expressões faciais e visão computacional;
- Treinar rede neural convolucional para reconhecer expressões faciais;
- Desenvolver uma aplicação para reconhecer as expressões de imagens obtidas a partir de vídeos em tempo real por meio da webcam ou vídeos gravados;
- Analisar a precisão da visão computacional da aplicação.

1.2 Justificativa

O reconhecimento de expressões faciais é aplicado em diferentes áreas. Por meio dela é possível entender as emoções como de felicidade, raiva, tristeza ou nojo, etc. As técnicas de visão computacional podem ser aplicadas para automatizar o processo de entendimento das emoções. Dessa forma, esta pesquisa implementou uma aplicação para reconhecer emoções a partir das expressões faciais presentes em vídeo em tempo real ou gravado, buscando otimizar suas análises.

A análise das expressões podem trazer impactos positivos, como por exemplo, em testes de usabilidade para entender a emoção de um usuário ao utilizar um determinado software, em locais de venda, como um restaurante, para analisar as expressões de clientes em relação a uma refeição, nos objetos, como carros para analisar o comportamento do motorista no trânsito, entre várias outras aplicações.

Esta pesquisa visa contribuir para o reconhecimento de expressões faciais, classificando as expressões contidas em imagens e vídeos aplicando técnicas de visão computacional.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo são apresentados conceitos fundamentais para o desenvolvimento da pesquisa.

2.1 Emoções básicas nas expressões faciais

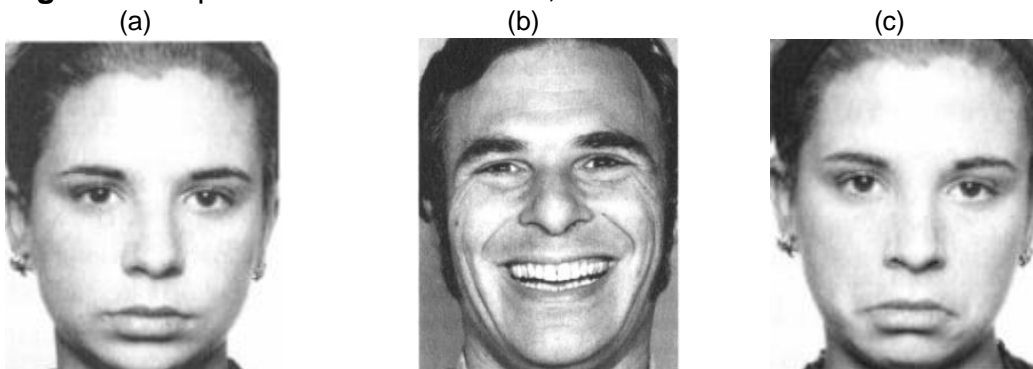
As emoções vêm sendo amplamente estudadas tanto no campo da psicologia como da computação, entre outros. Alguns biólogos consideram o sistema nervoso e hormonal os responsáveis pela as emoções (SANTIAGO, 2017).

Humanos expressam emoções diariamente, essas emoções podem ser percebidas de diferentes formas e uma delas são as expressões faciais. De acordo com Libralon (2014), tanto humanas como boa parte dos animais mamíferos utilizam expressões faciais para demonstrar estados emocionais. Exemplo de emoções que são percebidas pela as expressões faciais é a felicidade por meio do sorriso, uma característica marcante que representa alegria. As expressões faciais são as formas mais significativas e fortes que ocorre na comunicação para que possa conhecer o estado emocional de uma pessoa durante a comunicação (PARADA, 2017). Segundo Panzeri (2017), as expressões faciais são uma forma de comunicação não-verbal pois são uma troca de informações a qual não se utiliza de um vocábulo para se comunicar.

Segundo Ekman e Friesen (1976), as emoções mais comuns conhecidas como emoções básicas se dividem em sete, são elas: neutra, alegria, tristeza, raiva, medo, surpresa e nojo. Essas expressões são consideradas expressões universais. O autor estudou as emoções em diferentes etnias pelo mundo para assim então definir as expressões básicas. Cada expressão citada possui várias características, desde movimentações musculares como alterações nas sobrancelhas olhos, nariz, boca e bochecha, essas movimentações caracterizam uma determinada expressão. Conforme Ekman (2011), a expressão Neutra tem como características músculos relaxados e sobrancelhas e boca normais, como ilustrado na Figura 1 (a). Na expressão que representa alegria possui bochechas altas mostrando o contorno, olhos apertados, boca aberta mostrando o sorriso, sobrancelhas relaxadas como é exibido na Figura 1 (b). Nas expressões que representa tristeza as características que a descreve são cantos da boca para baixo, olhos caídos, movimento intenso

nas sobrancelhas, curvatura nas pálpebras, lábio inferior empurrado para cima, como é demonstrado na Figura 1 (c).

Figura 1 - Expressões Faciais: neutra, feliz e triste



Fonte: Ekman (2011).

A Raiva pode ser descrita com sobrancelhas baixas unidas com os cantos direcionados ao nariz, boca aberta mostrando os dentes ou boca fechada lábio contra lábio, afinamento do lábio, olhar fixo, pálpebra inferior flexionada e superior levantada como ilustra a Figura 2 (a). Na expressão que representa o Medo as características que descrevem podem ser pálpebras inferiores estendidas e pálpebras superiores erguidas e o resto da face inexpressivo ou olhos tensos e lábios estendidos como é demonstrado na Figura 2 (b).

Figura 2 - Expressões Faciais: raiva e medo

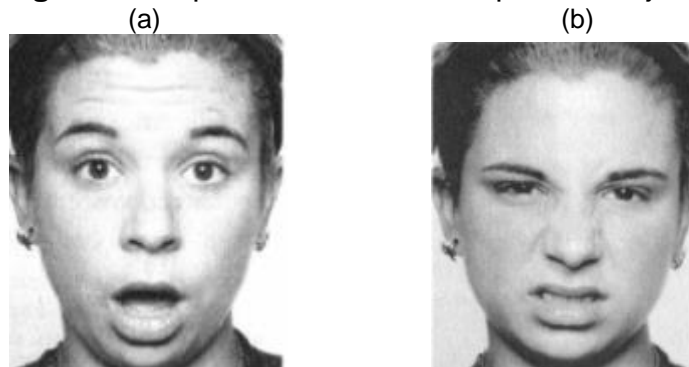


Fonte: Ekman (2011).

Na expressão de Surpresa são características olhos arregalados, pálpebras erguidas, sobrancelhas levantadas e lábios abertos ou não, como é ilustrado na Figura 3 (a). O Nojo tem como características enrugamento do nariz, sobrancelhas baixas, pálpebras superiores não erguidas, bochechas erguidas empurrando as

pálpebras inferiores para cima, canto da boca enrijecido e um pouco erguido como é ilustrado na Figura 3 (b).

Figura 3 - Expressões Faciais: surpresa e nojo



Fonte: Ekman (2011).

As expressões faciais acima representam as emoções consideradas básicas, onde cada emoção traz consigo alterações musculares faciais, que varia de emoção para emoção.

2.2 Computação afetiva

A computação afetiva é um subcampo da Inteligência Artificial (IA) que surgiu dentro da Ciência da Computação e aborda como tornar as máquinas aptas a aprender, logo está relacionado ao Aprendizado de Máquina. E um dos objetivos da computação afetiva é fazer com que a máquina possa reconhecer as expressões humanas e a entendê-las (RÄTSCH, 2004 apud DE JESUS *et al.*, 2018).

A computação afetiva se divide em duas perspectivas: inserir emoções humanas em máquina e reconhecer emoções humanas por máquinas na interação entre humano e computador (DE SOUSA *et al.*, 2016). O que pode trazer melhorias no campo da Interface Homem-Máquina em testes de usabilidades em *software* quando se trata de feedback do usuário. E uma das formas que a máquina possa reconhecer expressões humana é através da análise facial. A partir disso, pode-se identificar a emoção que representa o estado emocional do usuário. Sendo assim será um estado considerado como positivo ou negativo, que este usuário terá ao usar um determinado *software* (COSTA; DE SOUSA; PIRES, 2015).

2.3 Imagem digital

Toda e qualquer visualização é uma imagem, seja objetos, paisagens ou até mesmo pensamentos. Uma imagem consiste em uma função bidimensional, $f(x, y)$, que possui x e y como coordenadas do plano e a amplitude de f em alguma coordenada (x, y) é conhecida como intensidade ou nível de cinza da imagem nesse ponto. Se x , y e os valores de intensidade de f forem valores finitos e discretos, classifica como imagem digital (GONZALEZ; WOODS, 2009).

Imagem digital, consiste em imagem, que por meio eletrônico pode-se ser processada, transferida, impressa, armazenada, entre outras.

Imagens digitais são compostas por pontos chamados de pixels, cada pixel representa uma informação. Imagens com grandes resoluções contém muitos pixels e uma boa qualidade, ou seja, se uma imagem tem muitos pixels ela terá mais informações, o que implica também diretamente no tamanho da imagem. Caso a imagem tenha poucos pixels logo ela terá uma qualidade baixa dependendo do tamanho da resolução.

Uma das primeiras aplicações das imagens digitais ocorreu na indústria dos jornais na década de 1920, as imagens eram enviadas por cabos submarinos reduzindo o tempo de envio de mais de uma semana para 3 horas entre as cidades de Londres e Nova York (GONZALEZ; WOODS, 2009).

Atualmente, as imagens digitais podem ser produzidas de diversas formas, exemplo disto são as imagens capturadas por sensores como as câmeras fotográficas como também as criadas pelo computador que são chamadas de imagens sintéticas.

Existem várias formas de se utilizar imagens seja ela digital ou sintética, exemplo disso são publicações feitas em redes sociais como também a criação de filmes, exames médicos, animes, modelagem 3D, entre outros.

2.4 Visão computacional

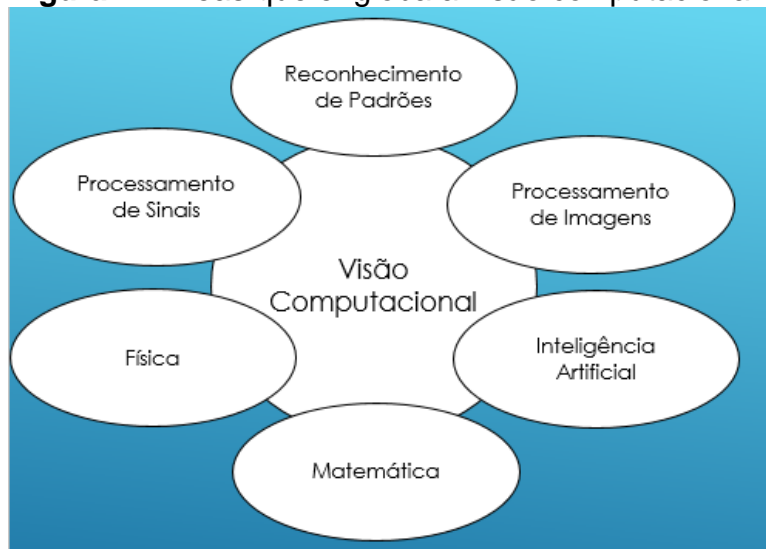
O olho humano é responsável por 1 dos 5 sentidos, a visão. A partir dele consegue-se distinguir cores, texturas, formas, tamanhos entre outras coisas. Conforme De Milano e Honorato (2010), o olho humano consegue perceber e interpretar objetos em uma imagem de forma muito rápida. Isso acontece no córtex

visual do cérebro, uma das partes mais complexas no sistema de processamento do cérebro. Sendo assim a visão computacional busca-se assimilar o olho humano, onde máquinas possam conseguir fazer o que a visão humana é capaz de fazer por meio de *hardwares* e *softwares*. Barielle (2018), fala que a visão computacional pode ser vista como um complemento da visão humana.

De acordo com De Milano e Honorato (2010), por volta da década de 70 foram realizados os primeiros trabalhos com a visão computacional, juntamente com a inteligência artificial. A definição de visão computacional ocorreu na década de 80 na obra *Computer Vision* de Ballard e Brown em 1982, onde foi definido como a ciência que estuda e desenvolve tecnologias que faz com que a máquina possa enxergar e extrair características por meio de imagens capturadas por sensores como scanners, câmeras fotográficas, entre outros. A partir das imagens capturadas é possível processar, extrair e reconhecer informações. Sendo assim é possível fazer com que a máquina passe a entender o que compõe uma determinada imagem (BARIELLE, 2018).

A visão computacional envolve diferentes áreas, como a matemática, física entre outras como se pode observar na Figura 4.

Figura 4 - Áreas que engloba a visão computacional



Fonte: Própria (2019).

Contudo, existem diversas bibliotecas e linguagens de programação que abstrai os conceitos matemáticos e físicos, implicando na facilitação para desenvolvimento de *softwares* nessa área, (BARIELLE, 2018).

Conforme Pereira (2018), cada ser humano usa cerca de dois terços do seu cérebro para o processamento visual. Sendo assim para os computadores fazer algo simples similar a nossa visão como os reconhecimentos de vários objetos demanda um alto processamento. A aplicação das técnicas da visão computacional está em constante crescimento, sendo bastante utilizada na área da medicina no auxílio do diagnóstico e tratamento de doenças, na segurança pública com o monitoramento de pessoas, na robótica, entres outras aplicações.

Backes e Sá Júnior (2016), definem que um sistema de visão computacional é constituído por várias fases:

- **Aquisição:** responsável pela captação das imagens, ou seja, tenta simular a função dos olhos. Os dispositivos que cumprem esse papel são os scanners, filmadoras, máquinas fotográficas, etc.
- **Processamento de imagens:** responsável por “melhorar” a imagem, isto é, retirar ruídos, salientar bordas, suavizar a imagem etc. Essa etapa pode ser um fim em si mesma ou ter o propósito de fornecer uma imagem mais adequada para as próximas fases. É importante salientar que essa fase compreende tanto o que usualmente se denomina “pré-processamento”, como rotação da imagem, equalização de histograma, etc. Quanto processamentos mais complexos, como, por exemplo, filtragens e aplicação de operadores morfológicos.
- **Segmentação:** responsável por particionar a imagem em regiões de interesse. Por exemplo, em uma imagem de paisagem, pode-se estar interessado apenas na porção que representa o céu, ou a vegetação, ou apenas o lago, ou algum cisne nesse lago etc.
- **Extração de características/Análise de imagens:** responsável por obter um conjunto de características do objeto de interesse. Em outras palavras, essa fase é responsável por encontrar uma codificação numérica que represente determinada imagem, como uma espécie de “impressão digital” (analogia imperfeita) que permita identificá-lo.
- **Reconhecimento de padrões:** responsável por classificar ou agrupar as imagens com base em seus conjuntos de características. Por exemplo, uma foto de uma única laranja, saber-se que aquele objeto pertence à classe “laranja” com base em atributos como cor, rugosidade da casca, formato, tamanho etc. É importante salientar que o objeto visto não é igual às laranjas

vistas no passado, mas apenas similar (na verdade, segundo a filosofia, a igualdade é um conceito teórico que não existe na natureza). No entanto, mesmo com essa limitação consegue-se classificá-lo corretamente na maioria dos casos. A Figura 5 ilustra um esquema de visão computacional simples.

Figura 5 - Esquema de um Sistema de visão computacional



Fonte: Backes e Sá Junior (2016 com adaptações).

Conforma a Figura 5 acima, há 5 passos importantes para que haja um sistema de visão computacional. Cada passo foi descrito anteriormente, sendo o primeiro passo adquirir a imagem, posteriormente os demais passos são aplicados tendo como resultados o reconhecimento do objeto.

2.5 Extração de características

Todo sistema de visão computacional possui a etapa extração de característica, essa é uma etapa fundamental que consiste na extração de informações necessárias para que haja o reconhecimento padrões sendo possível classificar ou reconhecer um determinado objeto.

Segundo Santiago (2017), a extração de características se inicia a partir de um conjunto de dados e cria valores derivados que devem ser informativos e não redundantes. O conceito de características é genérico. Pode ser classificado como

características pontos, bordas ou objetos. Para definir quais características escolher, depende do problema que se quer resolver.

Segundo Barelli (2018), existem quatro características principais que são usadas para a classificação:

- Características de aspecto - informações como cores e texturas.
- Características dimensionais - informações sobre o tamanho do objeto de interesse, a área, o perímetro e o diâmetro.
- Características inerciais - informações sobre os momentos, o centro geométrico e as formas geométricas envolventes de um objeto de interesse.
- As características topológicas - informações que não variam quando o objeto de interesse é movido, rotacionado ou sofre distorções na largura ou altura.

2.6 Reconhecimento de padrões em imagens

O reconhecimento de padrões em imagens consiste na classificação a partir de características em imagens como texturas, cores, formas, entre outros. Através do reconhecimento de padrões, os sistemas de visão computacional conseguem identificar determinados objetos. Humanos e máquinas utilizam características particulares para diferenciar e reconhecer objetos (BARIELLE, 2018).

No dia a dia consegue-se reconhecer vários padrões. Exemplo disso é o modo que se diferencia um computador de um lápis rapidamente pela forma dos objetos, o modo que se diferencia o sinal do semáforo nas ruas pela distinção pela cor. Isso tudo acontece pela forma que capturado as informações e em seguida fazer-se a classificação às associando a uma categoria ou classe pertencente.

Segundo Barelli (2018), uma classe é definida como um conjunto de padrões que possuem características em comuns. No semáforo sabe-se que o sinal vermelho está associado para que os pedestres possam atravessar e o sinal verde para parar. As máquinas ainda não têm a mesma facilidade de reconhecimento de padrões em relação ao nosso cérebro mais já é possível identificar e classificar diversos objetos, como ilustrado na Figura 6.

Figura 6 - Reconhecimento de objetos



Fonte: Própria

Conforme pode ser observado na Figura 6, são reconhecidos 3 objetos por meio da visão computacional, cada objeto reconhecido possui uma classe que por meio de seus padrões, a técnica de visão computacional reconhece os mesmos.

Existem vários algoritmos que realizam classificações de objetos, como o *Support Vector Machine (SVM)*, *Convolutional Neural Network (CNN)*, *K-nearest Neighbors (KNN)*, entre outros. Esses algoritmos são classificadores que utilizam das características dos objetos para fazer o Aprendizado de Máquina.

2.7 Detecção de faces

A face contém informações poderosas que são utilizadas para criação de ferramentas de aplicação de segurança, entretenimento, entre outras (SANTIAGO, 2017). Um passo importante para fazer o reconhecimento de expressões faciais consiste na detecção facial, fazer com que a máquina reconheça o que é uma face em uma determinada imagem. A detecção de face ocorre por meio da utilização de técnicas computacionais para analisar a imagens e afirmar se existe ou não uma face, se existir é retornado à sua localização (GOUVEIA; PAIVA, 2009).

Um sistema de detecção de face tem a função de imitar as capacidades naturais do ser humano para reconhecer características faciais em diferentes ambientes e associa-las a informações já armazenadas na memória (DA FONSECA, 2016). A detecção da face se dá pela a detecção de padrões que toda face humana possui, como os olhos, nariz e boca a partir disso é possível identificar se há uma face em uma imagem.

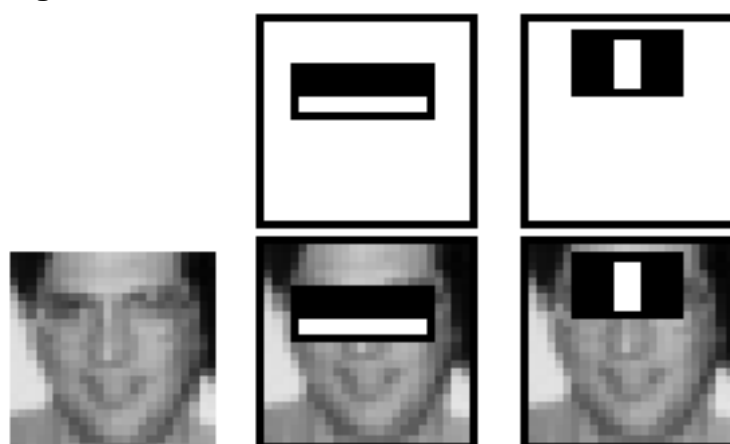
Detectar faces em imagens é uma tarefa considerada difícil, pois depende de condições como iluminação, fundo com detalhes que sobrepõe parcialmente ou totalmente a face que deverá ser localizada, entre outros problemas que podem ocorrer (GOUVEIA; PAIVA, 2009). A partir de uma face detectada é possível extrair sua característica, essas características formam um padrão que é utilizado para que haja o reconhecimento da expressão facial.

Diversas técnicas são utilizadas para a detecção de face e extração de características como por exemplo Haar Cascade e o Histograma de Gradientes Orientados (HOG).

Haar Cascade foi descrito por Viola e Jones (2001), consiste em um método para detecção de objetos, este método tem como característica sua rapidez, mesmo com o passar dos anos este método ainda é muito utilizado principalmente quando se trata de face. Ele consiste em um classificador Cascade que utiliza de um grande número de imagens positivas e negativas, se forem faces, as imagens positivas são as que contêm face e as negativas as que não contêm face.

Para treinar o classificador Cascade se utiliza de um algoritmo de Aprendizado de Máquina chamado AdaBoost que gera um arquivo treinado do tipo *XML*. Mas para isso, conforme Barielle (2018), é necessário extrair as características do objeto em questão onde se utiliza de máscaras chamadas Haar Features. Os Haar Features são similares a filtros onde segmenta as imagens através de operações que extrai características do objeto onde a combinação de linhas, bordas e centros formam o classificador. Na Figura 7 apresenta-se algumas dessas máscaras aplicadas em uma imagem que contém uma face.

Figura 7 - Haar Feature



Fonte: Viola e Jones (2001).

Cada Haar Feature varia a intensidade, esses retângulos demonstrados na Figura 6 calculam a soma dos pixels brancos menos a soma dos pixels preto, com isso cada valor resultante do cálculo é uma característica da face presente na imagem. Conforme Barielle (2018), para determinar as características é necessário a variação de luminosidade, pois as máscaras capturaram essas variações em diferentes amplitudes e direções, dessa forma o objeto passa a ser detectado.

O HOG foi descrito por Dalal e Triggs (2005). É um descritor de características e a sua utilização faz com que a máquina possa selecionar partes interessantes de uma imagem. Satya Mallick (2016), descreve um descritor de características como uma forma de representar uma imagem de maneira simplificada, com intuito de extrair informações úteis e descartar informações irrelevantes. A vários tipos descritores de características como exemplo o de formas, cores, texturas, entre outros.

O HOG se utiliza de métodos de detecção de bordas como o *sobel* com objetivo de extrair informações referentes à orientação das arestas existentes em uma imagem (PANCERI *et al.*, 2015).

O HOG consiste em dividir uma imagem em células, cada célula possui pixels, o intuito é gerar a partir do pixel um histograma da direção dos gradientes. Tendo isso em vista, a imagem original passa a ser descrita por esse procedimento não tendo grandes variações decorrentes de diferenças de luminosidade (MATTOS, 2017).

Conforme Cruz (2014), há quatro passos para gerar o descritor: cálculo do gradiente em cada pixel, agrupamento dos pixels em células, agrupamento das células em blocos e obtenção do descritor.

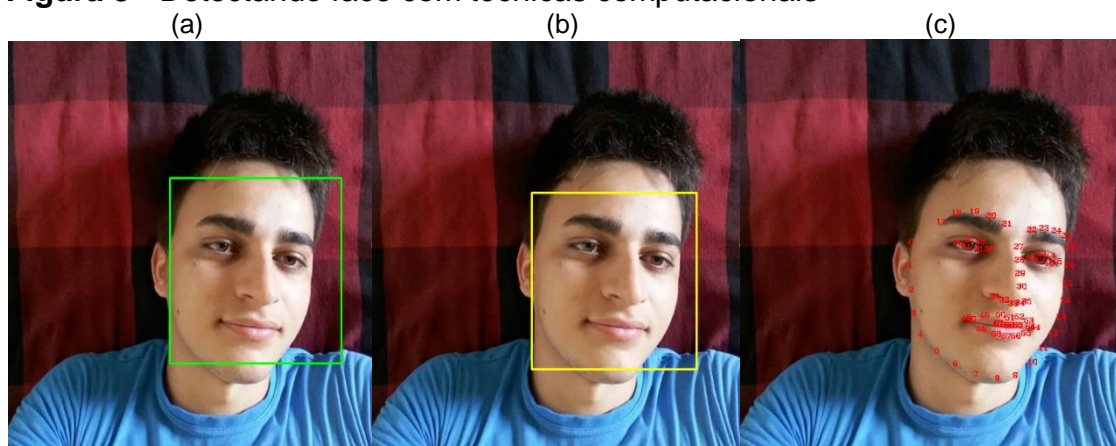
Conforme Gonçalves (2017), os gradientes calculados são indicadores de objetos em determinadas posições, isso ocorre devido a variação de orientação dos gradientes que são vetores, fazendo com que a forma do objeto seja determinada. Devido a isto a quantidade de falsos positivos acaba sendo reduzida, dependendo dos parâmetros utilizados.

Landmarking - após a face detectada é possível identificar pontos faciais, esses pontos indicam várias partes interessantes das face onde cada ponto tem uma numeração é chamado de Landmarks, sendo possível identificar determinadas partes da face como por exemplo os olhos. Tendo isso em vista, os mapeamentos dos pontos podem ser utilizados na identificação de expressões faciais. Wagner

(2017), define o Landmarking como um processo de marcação de pontos de interesse podendo ser utilizados em imagens 2D e 3D.

Na Figura 8 (a) é ilustrado a detecção de face utilizando Haar Cascade, (b) HOG e (c) os Landmarks marcando os 68 pontos faciais.

Figura 8 - Detectando face com técnicas computacionais



Fonte: Própria (2019).

É possível notar na Figura 8, a marcação da face com um quadrado, isto representa que a máquina conseguiu identificar uma face contida na imagem como também marcou 68 pontos faciais.

2.8 Reconhecimento de expressões faciais

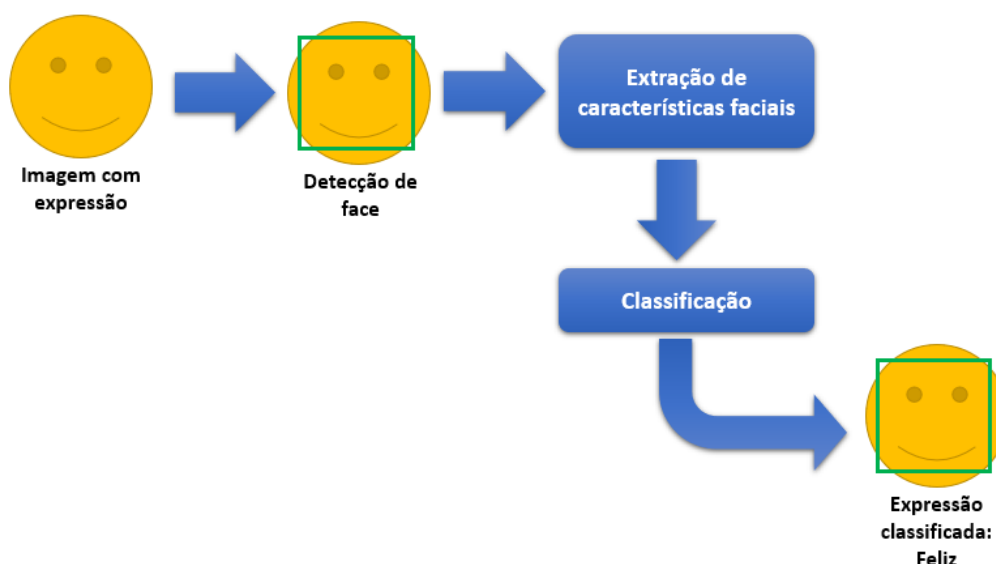
O reconhecimento de expressões faciais pode ser definido como a identificação de emoções através de uma imagem que contenha uma face humana expressando alguma emoção. As expressões variam de acordo com sentimento emitido em um determinado momento, por exemplo, se o sentimento for felicidade, provavelmente a expressão facial será o sorriso. Conforme Santiago (2017), para que um humano reconheça expressões faciais é necessário identificar características na face, essas características acontecem com alterações em algumas partes da face como mudanças na boca, olhos, nariz, bochechas, sobrancelhas, entre outras, o que acaba gerando mudanças nas suas posições relativas. Portanto, para uma máquina identificar expressões faciais contida em uma face através das características faciais é necessário utilizar técnicas de visão computacional para identificar essas características e as extrair-las.

Para a compreensão das emoções é necessário que a máquina além de detectar a informação emocional, possa também armazenar, processar, construir e manter um modelo emocional do usuário (LIBRALON, 2014). Devido a isto, o reconhecimento de expressões faciais acaba se tornando uma tarefa difícil para uma máquina.

Para que um humano reconheça expressões faciais é necessário identificar características na face, essas características acontecem com alterações em algumas partes da face como mudanças na boca, olhos, nariz, bochechas, sobrancelhas, entre outras. Conforme Santiago (2017), alguns métodos são baseados em sequência de imagens ou imagens de vídeo são usados para que haja o reconhecimento de expressões faciais.

Com base nisso o reconhecimento de expressões faciais tem gerado interesses em diversas áreas como visão computacional, processamento de sinais, reconhecimento de padrões e interação homem-computador (IHC) (COSSETIN, 2015). Na Figura 9 ilustra-se os passos básicos de um sistema de reconhecimento de expressões faciais.

Figura 9 - Esquema de um sistema de reconhecimento de expressão facial



Fonte: Própria (2019).

Como pode-se observar na Figura 9, é necessário que a imagem contenha uma face, esta face será analisada, tendo assim a face detectada e posteriormente reconhecida a expressão facial por meio das características faciais presentes.

Conforme apresentado em Zhang *et al.*, apud Cosseti (2015), um sistema automático de reconhecimento de expressões precisa resolver os seguintes problemas:

- Detecção e localização da face em uma cena;
- Extração de características da face;
- Redução de dimensionalidade;
- Classificação da expressão.

2.9 Redes Neurais Artificiais

O cérebro humano de acordo com Finocchio (2014), é composto por aproximadamente 100 bilhões de neurônios. Cada neurônio está conectado a aproximadamente 100 outros através de sinapses, formando, juntos, uma grande rede, chamada rede neural biológica. Uma rede neural artificial, como o próprio nome já diz, é formada por vários neurônios artificiais que tentam simular a rede neural biológica. Tem como intuito resolver problemas complexos que demanda um alto processamento de forma eficiente. Segundo Haykin (2001), a rede neural simula a maneira que é realizada uma tarefa no cérebro, essa rede funciona a partir de componentes eletrônicos ou por simulação da programação em computador digital.

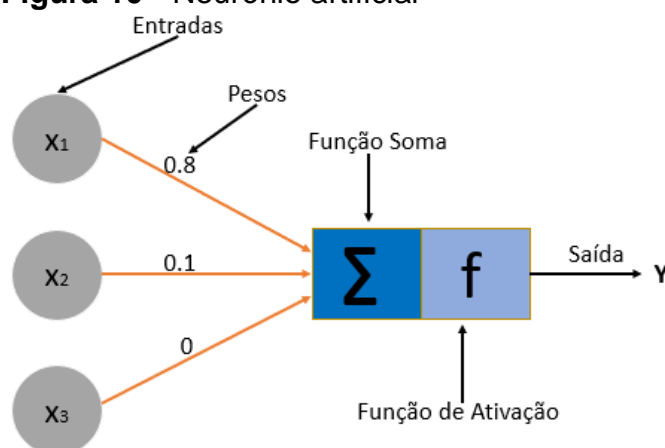
Uma das principais características de uma rede neural artificial é habilidade de aprender de acordo com as entradas passadas. Conforme Santos (2017), as Redes Neurais possuem a habilidade de receber várias entradas ao mesmo tempo e distribuí-las organizadamente, as informações armazenadas pela rede são compartilhadas por todas suas unidades de processamento.

Por exemplo o cérebro humano consegue reconhecer alguma pessoa seja pela face ou pelo o biofísico, mas para isto o nosso cérebro tem que primeiro já ter visto essa pessoa para que possa reconhecer, nesse momento acontece o aprendizado e isso acontece de forma muito rápida. De acordo com Haykin (2001), o cérebro reconhece um rosto familiar em uma cena não familiar em cerca de 100 a 200 ms. O mesmo acontece nas Redes Neurais Artificiais, para reconhecer uma face é preciso ter entradas que sirvam de treinamento para aprendizagem da rede, com isto mesmo uma face estando um pouco diferente da entrada como tendo por exemplo uma barba ou óculos é possível reconhecer a face em questão.

O procedimento para realizar o processo de aprendizagem é chamado de aprendizagem, cuja função é modificar os pesos sinápticos da rede de uma forma ordenada para alcançar um objetivo de projeto desejado (HAYKIN, 2001).

Um neurônio é algo indispensável para que haja operação na rede neural, pois é uma unidade de informação (HAYKIN, 2001). Na Figura 10 é ilustrado um neurônio artificial.

Figura 10 - Neurônio artificial



Fonte: Haykin (2001, com adaptações).

Como pode ser observado na Figura 10, Haykin (2001), identifica três elementos básicos do modelo neural:

- Um conjunto de sinapses uns caracterizados por um peso.
- Um somador para soma os sinais de entradas.
- Uma função de ativação para restringir a amplitude e saída de um neurônio.

Os neurônios de uma rede neural artificial tentam imitar o comportamento dos neurônios do cérebro humano ao acumular os impulsos resultantes de entrada ou de axônios de outros neurônios, até alcançar o limite que é estabelecido pela função de ativação (GUEDES, 2017).

As funções de ativação são funções não-lineares conectadas ao final da estrutura de um neurônio artificial, também são inspiradas biologicamente e definem a saída com base nos dados de entrada e o limiar de ativação (MIYAZAKI, 2018). Existem várias funções de ativações a algumas delas são Unidade Linear Retificada (ReLU), Sigmóide, Tangente Hiperbólica (TanH), Linear, Softmax.

2.10 Redes Neurais Convolucionais

Redes Neurais Convolucionais - (*Convolutional Neural Network - CNN*) tem como foco a visão computacional, tem esse nome devido a camada de convolução e possui arquitetura de aprendizado profundo (*Deep Learning*). Diferente de uma rede neural tradicional ela tem como característica aprender com as características extraídas de uma imagem, ou seja, utiliza imagens e como entradas os pixels com informações importantes. Já a rede neural tradicional quando usada em imagens tem como entradas todos os pixels da imagem mesmo não tendo informações importantes, logo há mais processamento.

Conforme Ebernan e Krohling (2018), na rede neural convolucional consegue-se armazenar semelhanças entre pixels vizinhos da imagem devido ao processo de convolução. No treinamento são capazes de aprenderem com mudanças ou transformações nas imagens que pode ser representações invariantes a escala, translação, rotação e transformações (JARRETT *et al.*, 2009) apud RIGHETTO, 2016). Santos (2017), define que uma CNN tem na arquitetura:

- Mapeamento de características, ou campos receptivos locais, como um dos pontos de maior similaridade com as Redes Neurais biológicas que garante uma maior robustez a distorções locais.
- Pesos compartilhados que permitem uma redução dos parâmetros livres e invariância geométrica.
- Sub-amostragem temporal ou espacial, o que reduz o tamanho total dos mapas de características a cada camada, chegando assim na última camada apenas como valores unidimensionais, o que nesse ponto em diante as torna equivalentes a uma rede neural perceptron de múltiplas camadas (MLP).

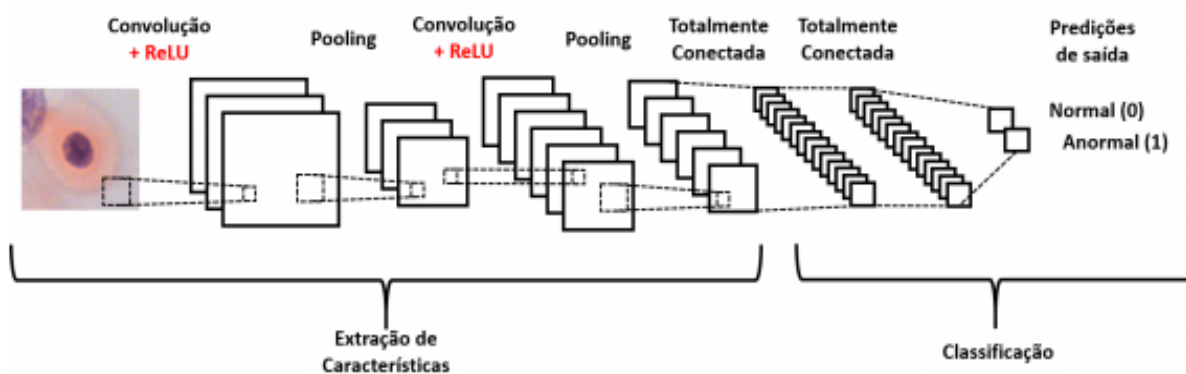
Geralmente as Redes Neurais Convolucionais possuem diversos tipos de camadas, como a camada de convolução, sub-mostragem, normalização de contraste e camadas completamente conectadas. A organização acontece por estágios, cada estágio pode ter uma ou mais camadas de convolução em sequência, que segue uma camada de sub-mostragem e opcionalmente por uma camada de normalização (SANTOS, 2017).

A CNN possui uma hierarquia que busca a estrutura em relação ao reconhecimento de uma imagem, onde os pixels formam arestas, a partir das áreas

fazem padrões, esses padrões descrevem objetos, que por fim descreve as cenas (AREL *et al.*, 2010 apud RIGHETTO, 2016).

Dois blocos básicos de uma CNN são o extrator de característica e classificador. A camada do extrator de características é a camada convolutiva e o *pooling*. Conforme Ebernan e Krohling (2018), é na camada de convolução que ocorre a extração de características da imagem, a camada de *pooling* realiza a subamostragem das imagens e a interpretação das características extraídas por meio da camada totalmente conectada. Na Figura 11 é demonstrado a arquitetura de uma Rede Neural Convolutiva.

Figura 11 - Arquitetura de uma CNN demonstrando a divisão da extração de características e classificação



Fonte: ARAÚJO *et al.*, (2017).

Pode ser observado na Figura 11, que na arquitetura há uma divisão onde ocorre a extração de características e classificação. A parte referente a entrada da imagem, camada de convolução e camada de *pooling* pertence a etapa de extração de característica, após isso nas camadas totalmente conectadas ocorre a classificação da imagem de entrada, identificando se a célula é anormal ou célula normal.




2.10.1 Camada de convolução




A convolução consiste em fazer cálculos matemáticos que envolve soma e multiplicação de matrizes para que possa ocorrer alteração de determinados pixels. É a camada fundamental de uma CNN. Segundo Santos (2017), essa camada é constituída por vários neurônios, cada um responsável por aplicar um *Kernel* em

uma parte específica da imagem, com isso é deslizado o *Kernel* bidimensional sobre as camadas de entrada e calculado o produto escalar entre as camadas do *Kernel* e as da imagem.

Basicamente uma imagem é uma matriz, para saber quantos elementos existem na imagem é feito a multiplicação dos pixels da horizontal pela vertical. O *Kernel* é uma outra matriz que serve para a aplicação de efeito o que gera uma nova imagem além de ser utilizado nesse processo de aprendizado é bastante utilizado no dia a dia, exemplo é quando aplicado o efeito de desfoque em uma imagem, ou seja, um processamento na imagem. Na Figura 12 é demonstrado a aplicação de alguns *Kernel* sobre uma imagem. Para a aplicação do efeito é multiplicado a matriz imagem pela matriz *Kernel*.

Figura 12 - Aplicando o Kernel em uma imagem

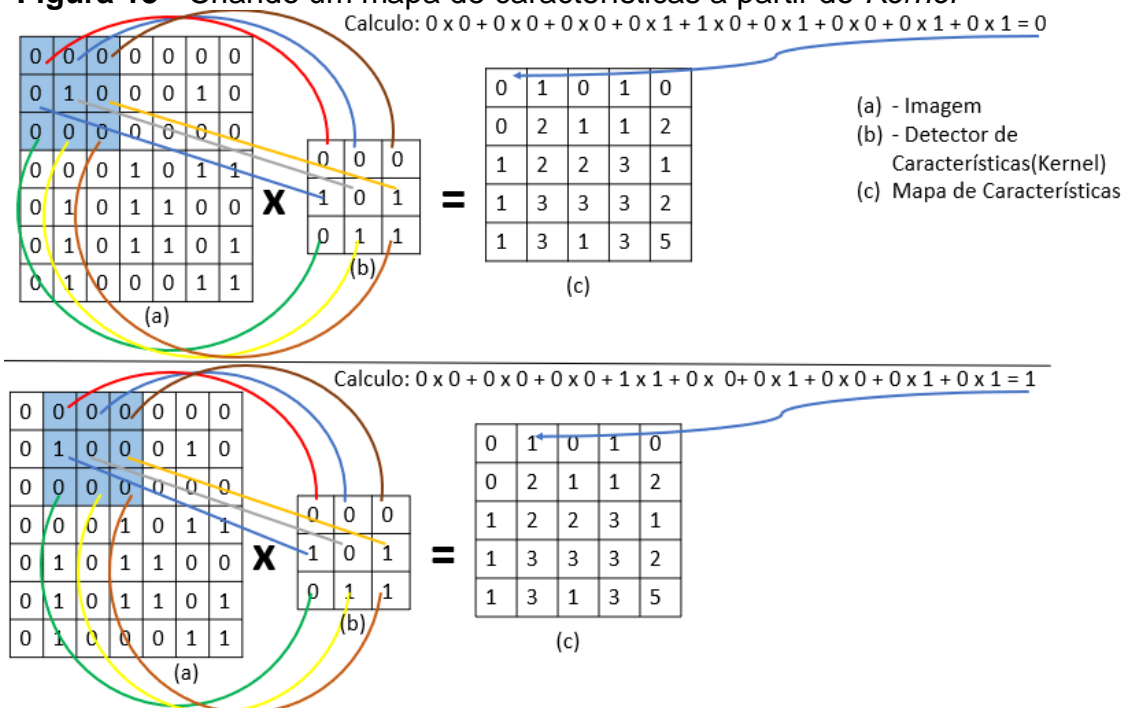
Operação	Kernel	Imagem Resultado									
Identity	<table border="1"> <tr><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> </table>	0	0	0	0	1	0	0	0	0	
0	0	0									
0	1	0									
0	0	0									
Emboss	<table border="1"> <tr><td>-2</td><td>-1</td><td>0</td></tr> <tr><td>-1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>2</td></tr> </table>	-2	-1	0	-1	1	1	0	1	2	
-2	-1	0									
-1	1	1									
0	1	2									
Outline	<table border="1"> <tr><td>-1</td><td>-1</td><td>-1</td></tr> <tr><td>-1</td><td>8</td><td>-1</td></tr> <tr><td>-1</td><td>-1</td><td>-1</td></tr> </table>	-1	-1	-1	-1	8	-1	-1	-1	-1	
-1	-1	-1									
-1	8	-1									
-1	-1	-1									

Bottom Sobel	<table border="1"> <tbody> <tr> <td>-1</td> <td>-2</td> <td>-1</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td>1</td> <td>2</td> <td>1</td> </tr> </tbody> </table>	-1	-2	-1	0	0	0	1	2	1	
-1	-2	-1									
0	0	0									
1	2	1									
Right Sobel	<table border="1"> <tbody> <tr> <td>-1</td> <td>0</td> <td>1</td> </tr> <tr> <td>-2</td> <td>0</td> <td>2</td> </tr> <tr> <td>-1</td> <td>0</td> <td>1</td> </tr> </tbody> </table>	-1	0	1	-2	0	2	-1	0	1	
-1	0	1									
-2	0	2									
-1	0	1									
Top Sobel	<table border="1"> <tbody> <tr> <td>1</td> <td>2</td> <td>1</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td>-1</td> <td>-2</td> <td>-1</td> </tr> </tbody> </table>	1	2	1	0	0	0	-1	-2	-1	
1	2	1									
0	0	0									
-1	-2	-1									

Fonte: Própria (2019).

Como pode ser observado na Figura 12 de acordo com os valores que são passados na matriz *Kernel* é possível criar alterações na imagem, podendo como exemplo destacar bordas de uma imagem. Na CNN, o *Kernel* serve para extrair características importantes de uma imagem. Com isto a CNN pode analisar e extrair as características automaticamente. Na Figura 13 é apresentado como é criado um mapa de características por meio da aplicação do *Kernel*.

Figura 13 - Criando um mapa de características a partir do *Kernel*



Fonte: Própria (2019).

Como pode ser observado na Figura 13, possui uma matriz 7x7 que representa uma imagem com 49 pixels. Para ocorrer a extração das características é multiplicado e somado a parte que está destacada pelo o *Kernel* assim tendo o resultado que representa a área em destaque e preenchendo o mapa de características. Após o cálculo de toda a matriz, consegue-se como resultado o mapa de características, ou seja, uma nova imagem com dimensões menores.

Após o mapa de características ser criado, uma função de ativação chamada *ReLU* é usada, essa função consiste em substituir os valores negativos do mapa de característica por zero, ou seja, se o valor for maior ou igual a zero não há alterações. Sendo assim na convolução aplica vários detectores de características o que acaba criando vários mapas de características, logo mais processamento e maior o número de características extraídas.

É por meio dos filtros que os mapas de características são gerados, isso ocorre por meio da região denominada campo receptivo local. Há também pesos compartilhados entre os neurônios, os pesos fazem com que o filtro aprenda padrões da entrada (HAFEMANN, 2014 apud RIGHETTO, 2016). Segundo Araújo *et al.*, (2017) esses filtros recebem como entrada um arranjo 3D também chamado de volume, caso a imagem seja colorida cada pixels possui três canais que são o RGB (Vermelho, Verde e Azul).

Conforme Ujjwal (2016) apud Araújo *et al.*, (2017), há três parâmetros que tem como função controlar o tamanho do volume resultante da camada convolucional: profundidade (*depth*), passo (*stride*) e *zero-padding*. GUEDES (2017), Define *depth*, *stride* e *zero-padding* como:

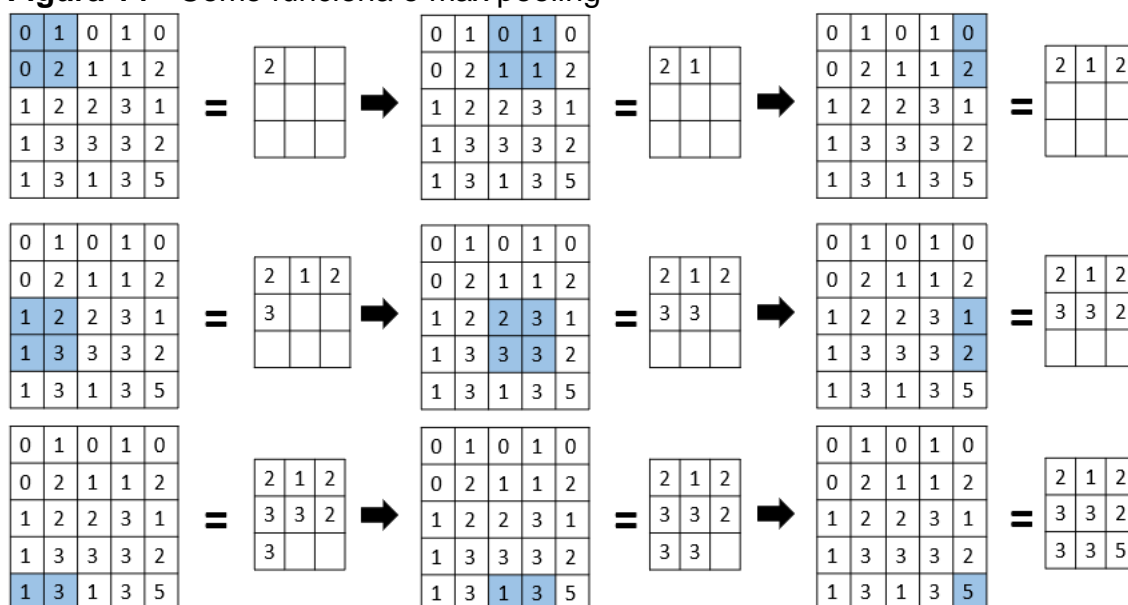
GUEDES (2017), define profundidade, passo e zero-padding como:

- *Depth*: É o número de neurônios que serão estimulados pela mesma parte do campo visual (campo receptivo) ao longo da dimensão da profundidade.
- *Stride*: É o número utilizado para definir o espaçamento da sobreposição dos campos receptivos dos neurônios de colunas diferentes com relação à profundidade.
- *Zero-padding*: Determina qual tamanho do preenchimento com zeros que será adicionado às bordas da entrada.

2.10.2 Camada pooling

Conforme Araújo, Carneiro e Silva (2017) apud Myazaki (2018), após camada de convolução é utilizado a camada *pooling*, que tem como objetivo reduzir a dimensão da camada de entrada o que acaba reduzindo o custo computacional como também evita o *overfitting*. *Overfitting* é a memorização dos dados utilizado gerando bons resultados na classificação, mas quando utilizado outro conjunto de dados sua generalização acaba se tornando ineficiente.

Nessa camada é destacado ainda mais as características, nesta etapa é utilizando o método de *max pooling* que é o mais comum, funciona um pouco diferente quando comparado na Figura 13, pois ao invés de pegar pixel a pixel, é utilizado o *strid* criando uma nova matriz com os valores maiores, desta forma são descartados os valores desprezíveis, gerando uma invariância a pequenas mudanças e distorções de locais. Na Figura 14 é ilustrado o funcionamento do *max pooling*.

Figura 14 - Como funciona o max pooling

Fonte: Própria (2019).

Como pode-se observar na Figura 14, é preenchida a matriz conforme cada passo de acordo com os maiores valores, gerando uma matriz menor.

2.10.3 Camada totalmente conectadas

A camada totalmente conectada tem esse nome devido aos neurônios anteriores estarem conectados com os próximos neurônios. É nessas camadas que acontece a propagação do sinal por meio da multiplicação ponto a ponto, como também se utiliza de uma função de ativação (ARAÚJO *et al.*, 2017). É a camada que também faz parte das Redes Neurais comuns, tem como função a conexão das camadas sem utilizar pesos compartilhados (HAFEMANN, 2014 apud RIGHETTO, 2016). Após as características serem extraídas, as camadas totalmente conectadas utilizam dessas características para fazer a classificação.

2.11 Trabalhos relacionados

O trabalho de Mendonça (2018), foi treinado uma Rede Neural Convolutiva com a biblioteca Keras em conjunto com o Tensorflow para o reconhecimento de expressões faciais em tempo real para auxiliar em testes de usabilidade. Utilizou-se uma máquina virtual dedicada do *Google Colab* com suporte ao Tensorflow em

GPU. A Rede Neural Convolutiva detecta 4 tipos de expressões faciais: (alegria, desgosto, raiva e surpresa) para isto foi realizado um treinamento com apenas a base de dados JAFFE, que possui 213 imagens, descartando as imagens que tinham a expressão de medo e tristeza, resultando assim em 150 imagens, com isso foi feito um redimensionadas para 64x64 pixels devido ao baixo do número de imagens e para melhorar os dados na classificação foram feitos cortes de 48x48 pixel, onde cada imagem foi cortada 16 vezes aumentando a base. A arquitetura da CNN utilizada foi a LeNET, utilizou-se 80% do conjunto de dados para o treinamento e 20% para o conjunto de validação. Obtendo uma precisão de 62%. Foi desenvolvido um módulo de vídeo onde se detecta a expressão usando a técnica de Viola-Jones e em seguida reconhece a expressão.

O trabalho de Oliveira e Jaques (2013), apresenta um sistema computacional que classifica 6 emoções básicas: raiva, medo, repulsa, surpresa, alegria e tristeza, pela webcam do computador. Essas imagens capturadas são classificadas por uma Rede Neural. Antes da classificação é realizada a aplicação de métodos de visão computacional e processamento de imagens como a detecção de faces e das características faciais. Foram utilizadas imagens estáticas e vídeos contando 624 exemplos para treinar e testar, onde foi considerado três fontes para construção do classificador. Obteve-se como resultado nos experimentos uma taxa de 63,33% do sistema e a rede neural isolada obteve uma taxa de reconhecimento de 89,87%.

O trabalho de De Sousa *et al.*, (2016), utilizou do SDK do Kinect versão 1.8 que possui seis expressões faciais em Unidades de Animação (UA's) que são unidade de ações que serve para ilustrar a face humana, com elas é possível modelar expressões faciais. Utilizou-se as imagens capturada pelo sensor do Kinect para aplicar o modo de reconhecimento das expressões e a técnica de Viola-Jones para detectar a face presente na imagem, para implementação foi utilizado os recursos da biblioteca SDK do Kinect. Sendo assim foi possível classificar as expressões do usuário. As expressões que a aplicação classifica são alegria, surpresa, tristeza e raiva. Para os experimentos foi selecionado 20 voluntários onde cada um simulou 20 expressões, 5 expressões de cada tipo. Obteve como resultado de 67,9%.

3 METODOLOGIA

Para tornar possível o reconhecimento de expressões faciais, esta pesquisa segue alguns passos. Inicialmente é feita a aquisição das imagens da base de imagem Fer2013. Algumas imagens desta base são apresentadas na Figura 15. Vale ressaltar que esta base é amplamente utilizada em diversos trabalhos que tratam de reconhecimento de expressões faciais.

Figura 15 - Exemplo de expressões da base Fer2013

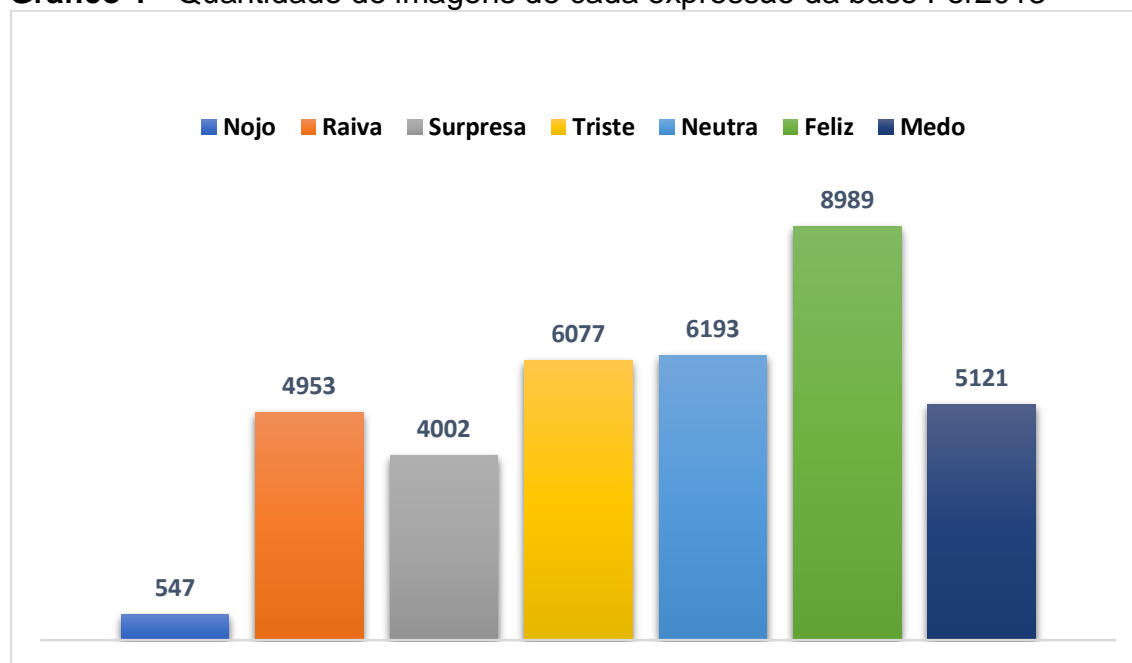


Fonte: Arriaga, Plöger e Valdenegro (2017).

Na Figura 15, as faces contidas nas imagens se encontram em escala de cinza, segundo a descrição da base de imagens Fer2013¹, as imagens possuem dimensões de 48x48 pixels. Na base de imagem, os rostos se encontram centralizados, as faces possuem várias faixas etárias, contendo 7 tipos de emoções que se dividem nas categorias (0 = Raiva, 1 = Nojo, 2 = Medo, 3 = Felicidade, 4 = Tristeza, 5 = Surpresa, 6 = Neutro).

As imagens se encontram no formato .CSV, que contém duas colunas, "emoção" e "pixels". A coluna emoção contém um código numérico que varia de 0 a 6. A coluna pixels contém uma sequência de valores para cada imagem. O conjunto de dados possui 35.888 imagens. O Gráfico 1 a seguir ilustra a quantidade de imagens que cada expressão possui.

¹ Disponível em: <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>. Acesso em: 10 mar. 2019.

Gráfico 1 - Quantidade de imagens de cada expressão da base Fer2013

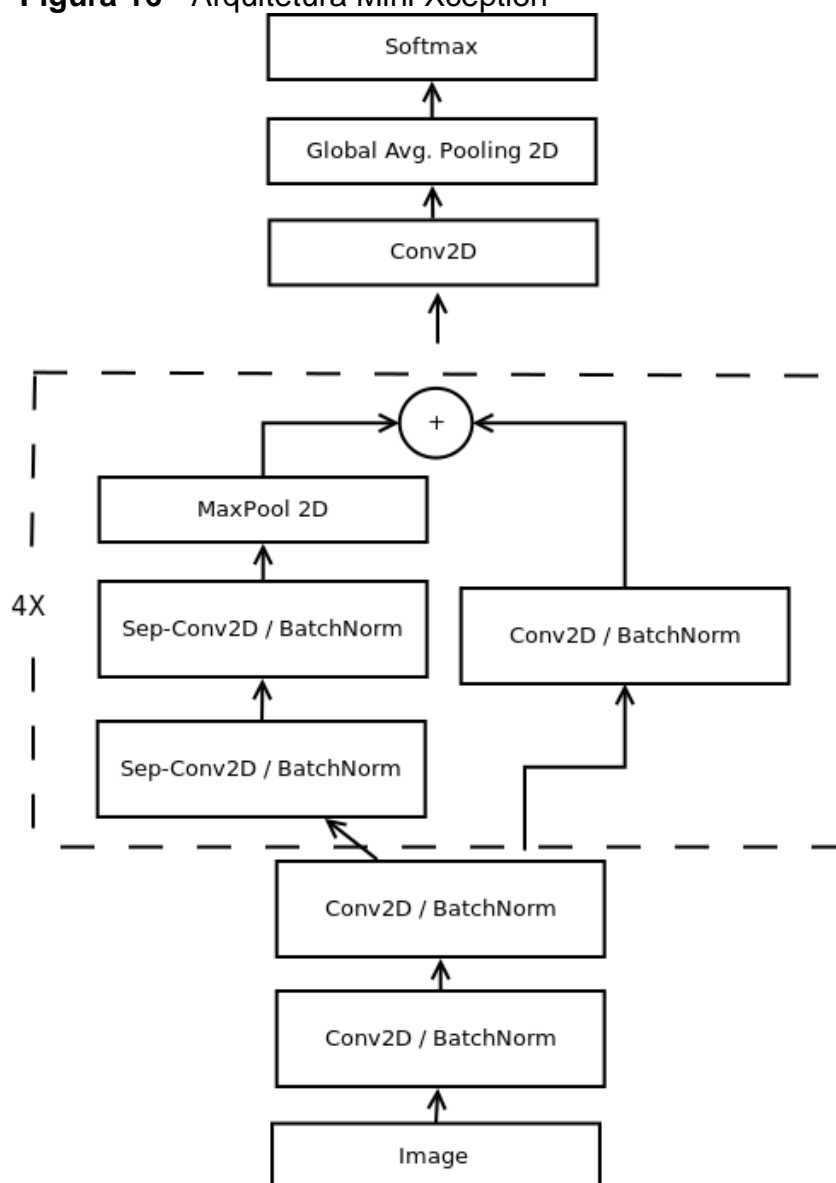
Fonte: Própria

Observando o Gráfico 1, pode-se notar que as expressões do tipo feliz possuem o maior número, com 8989 imagens, seguidas por neutra (6198), triste (6077), medo (5121), raiva (4953), surpresa (4002) e nojo com uma quantidade muito menor em relação as demais com 547 imagens.

No processo de carregamento dos dados é convertido a sequência de pixels em imagens com dimensões de 48x48 pixels.

A arquitetura de rede neural convolucional utilizada foi a Mini Xception proposto por Arriaga, Plöger e Valdenegro (2017), por ter um custo computacional menor devido ser pequena, mas, eficiente para esse conjunto de imagens, tendo bons desempenho. Na Figura 16 é apresentado a arquitetura. Utilizou-se uma rede pré-treinada. É um reaproveitamento de código, podendo assim aproveitar pesos de uma rede já treinada.

A biblioteca Keras disponibiliza uma rede pré-treinada da arquitetura Xception. Sendo assim, suas primeiras entradas estão configuradas de modo que aprendem coisas básicas essenciais que todas CNNs utilizam como bordas, formas, cores, entre outros.

Figura 16 - Arquitetura Mini Xception

Fonte: Arriaga, Plöger e Valdenegro (2017)

De acordo com a Figura 16, que o bloco central é repetido 4 vezes, tendo como características o uso de módulos residuais e convoluções separáveis em profundidade. Para treinar a CNN utiliza-se técnicas como aumento de dados, regularizador de kernel, normalização em lote, pool médio global e convolução profundamente separável. O aumento de dados aplica transformações em imagens para aumentar a quantidade de imagens para serem treinadas, exemplo de transformações são: cortes, zoom, rotação, reflexão, entre outros. O regularizador de kernel é a aplicação de pesos para otimização das camadas. A normalização em lote pela normalização, faz com que o processo de treinamento seja mais rápido, pois aplica transformação que deixa o desvio padrão de ativação próximo a um e a

ativação média próximo a zero. O pool médio global, reduz cada mapa de recursos em um valor escalar, assumindo a média de todos os elementos no mapa de recursos. Convolução profundamente separável, consiste em duas camadas diferentes: convoluções em profundidade e convoluções em ponto. As convoluções separáveis em profundidade reduzem o cálculo em relação às convoluções padrão, reduzindo o número de parâmetros.

Após os processamentos na imagem para o treinamento, o próximo passo foi detectar a face presente na imagem, após isto, foi extraído as características para em seguida serem utilizadas para treinar o modelo.

A rede neural convolucional recebe as imagens em lote para seu treinamento, onde cada lote contém 32 imagens. Foi definido um total de 110 épocas para o treinamento que resulta em um modelo com as possibilidades de 7 emoções diferentes, ou seja, 7 classes que podem pertencer aos rostos presentes nas imagens da base de dados, dimensionadas por 48x48 pixels. A CNN extrai por si só as características, tendo isso em vista, foi gerado o modelo com base na CNN, portanto já será possível utilizar para classificar as imagens com expressões.

A aplicação recebe as imagens ou vídeos captados pela câmera ou arquivos e as classifica, verificando se a expressão contida na imagem é neutra, alegre, triste, raiva, medo, nojo ou surpresa.

Para os testes foi utilizado um computador com processador Intel Core i3 da 5 geração 2.00GHz com 8Gb de RAM sem placa de vídeo dedicada e uma webcam de 0,9MP que veio de fábrica no dispositivo.

Os testes se dividem em dois tipos:

- 1 - Teste com imagens utilizando três bases de expressões faciais;
- 2 - Teste com 12 voluntários que expressão as emoções em frente a webcam do computador, e a aplicação identifica qual é aquela expressão.

3.1 Base de imagens

Neste subtópico será descrito três bases de imagens utilizadas nesta pesquisa para testar a aplicação de reconhecimento de expressões. Sendo as bases JAFFE, FacesDB e CK+.

3.1.1 JAFFE

É um banco de imagens de faces com sete expressões faciais: neutro, alegria, triste, raiva, medo, nojo e surpresa. Esta base possui um total de 213 imagens em escala de cinza com dimensões 256x256 pixels de mulheres japonesas que foram tiradas no Departamento de Psicologia da Universidade de Kyushu, como pode ser visto na Figura 17.

Figura 17 - Base de Imagens JAFFE



Fonte: Modificada a partir do banco de dados JAFFE (MICHAEL *et al.*, 2010).

3.1.2 FacesDB

A base de imagens FacesDB de acordo com a descrição no site², foi criada com o intuito de facilitar pesquisas de animações para análise e sintaxe do rosto e expressões, contem imagens de 22 homens e 16 mulheres, sendo um total 38 indivíduos com idade entre 20 a 50 anos. Algumas imagens desta base são ilustradas na Figura 18.

Figura 18 - Expressões do tipo feliz, base FacesDB



Fonte: Modificada a partir do banco de dados FacesDB.

² Disponível em: app.visgrafimpa.br/database/faces/. Acesso em: 15 set. 2019

Como pode ser observado na Figura 18, as imagens se encontram em colorido e possui um fundo preto, na figura é demonstrada algumas faces com a expressão feliz.

3.1.3 CK+

Segundo Santiago (2017), o Cohn-Kanade possui duas versões CK e CK+, a CK possui apenas imagens na escala de cinza e inclui 486 sequências de 97 indivíduos. Possuem sequência que começa com a expressão neutra e segue até o ápice da expressão e foram rotuladas com a emoção representada. A outra versão a CK+ é uma versão estendida da CK, possui imagens na escala de cinza e coloridas bem como imagens com expressões representadas e espontâneas. Na Figura 19 é demonstrado algumas imagens da base que representa a expressão raiva.

Figura 19 - Expressões do tipo triste, base CK+



Fonte: Modificada a partir do banco de dados CK+ (JUCEY *et al.* 2010).

3.2 Captura de vídeo em tempo real e gravado

Nesta etapa foi desenvolvido uma aplicação de captura de vídeo capaz de identificar e registrar as expressões dos usuários, salvando o nome da expressão, o tempo de quando iniciou e o tempo de quando acabou.

A detecção de face ocorre por meio do Haar Cascade, algoritmo que se encontra na biblioteca OpenCV. O OpenCV é uma biblioteca *open source*,

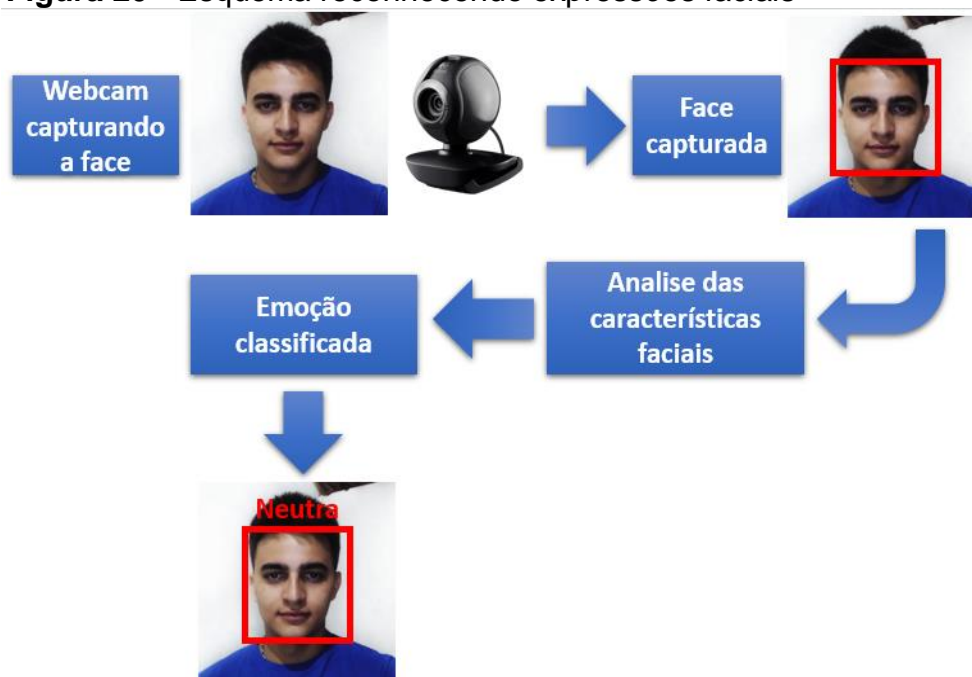
amplamente conhecida e utilizada para reconhecimento de objetos, detecção de faces, processamento de imagens, entre outras. O Haar Cascade tem como características a rapidez e leveza para fazer a detecção, podendo assim ser executado em computadores que possui um *hardware* mais modesto.

A aplicação foi desenvolvida na linguagem de programação Python na IDE Pycharm, utilizando a biblioteca Tensorflow, Keras, OpenCV e Tkinter. Foi criado módulo de captura de vídeo em tempo real e módulo de vídeo gravado.

Ambos detectam as expressões e tem como saída a face marcada com um “quadrado verde” junto ao nome da expressão. Após a classificação, é retornado à emoção identificada.

A Figura 20 exibe o esquema de processos realizados pela aplicação em tempo real: captura do frame contendo a face, detecção da face, análise da face e detecção das características faciais e saída, e a expressão classificada. Inicia-se pela obtenção de imagens capturadas de usuários em frente ao computador por uma webcam. As imagens capturadas são processadas por meio de métodos de visão computacional conseguindo a localização da face. Após a face detectada, são analisadas as características faciais como movimentos musculares que variam de acordo com a expressão, para posteriormente classificar a emoção.

Figura 20 - Esquema reconhecendo expressões faciais



Fonte: Própria.

Diante disso, obtive a aplicação de reconhecimento de expressões faciais. Na opção de vídeo em tempo real ou vídeo gravado, o usuário digita o nome do teste para assim criar um arquivo.CSV que armazenará informações em que hora aconteceu aquela expressão, quando terminou, para entender a sua duração. Podendo ser utilizado posteriormente em uma análise. O nome do teste também é utilizado para criar o diretório e nomear as imagens que estão sendo salvas quando ocorre a detecção da emoção, ou seja, a cada emoção detectada é salvo no diretório que pertence a aquela expressão, se a expressão for raiva, será salvo o frame da expressão raiva na pasta raiva, que se encontra no diretório com nome fornecido quando se iniciou o teste. Na Figura 21 ilustra o reconhecimento de uma expressão em um vídeo em tempo real.



Fonte: Própria

A Figura 21 ilustra o reconhecimento da expressão do tipo neutra, bem como a interface gráfica da aplicação com suas funcionalidades.

4 RESULTADOS E DISCUSSÕES

Este capítulo apresenta os resultados alcançados nos testes realizados com as bases de imagens e com os voluntários. Por fim, apresenta uma análise dos mesmos, comparando os resultados obtidos em ambos os testes.

Os testes se dividem em dois tipos, um teste com imagens utilizando 3 bases de expressões faciais ambas apresentadas no capítulo anterior e o outro teste consiste na utilização de voluntários que demonstram expressões faciais em frente a webcam. Aplicação identifica qual é a expressão e salva nos diretórios específicos que a expressão foi classificada.

4.1 Testes com imagens

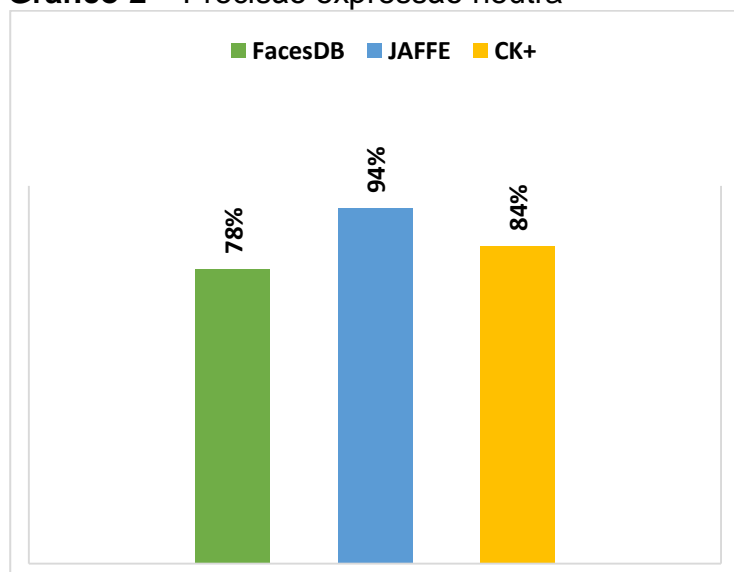
Os testes utilizaram-se de três bases distintas, a JAFFE, FacesDB e CK+. Foram distribuídas as imagens das bases em 7 diretórios.

A base JAFFE contém 213 imagens em escala de cinza, sendo elas, feliz (29 imagens), tristeza (33 imagens) raiva (30 imagens), nojo (29 imagens), medo (30 imagens), surpresa (30 imagens), e neutro (32 imagens).

A base FacesDB possui imagens com resoluções maiores que a base JAFFE e CK+, as imagens se encontram coloridas. Cada diretório possui 36 imagens de cada expressão, totalizando 256 imagens.

A base CK+ foram utilizadas 898 imagens sendo, feliz (69 imagens), tristeza (28 imagens), raiva (41 imagens), nojo (59 imagens), medo (25 imagens), surpresa (83 imagens), e neutro (593 imagens). Vale ressaltar que em ambas bases, cada imagem possui apenas uma face.

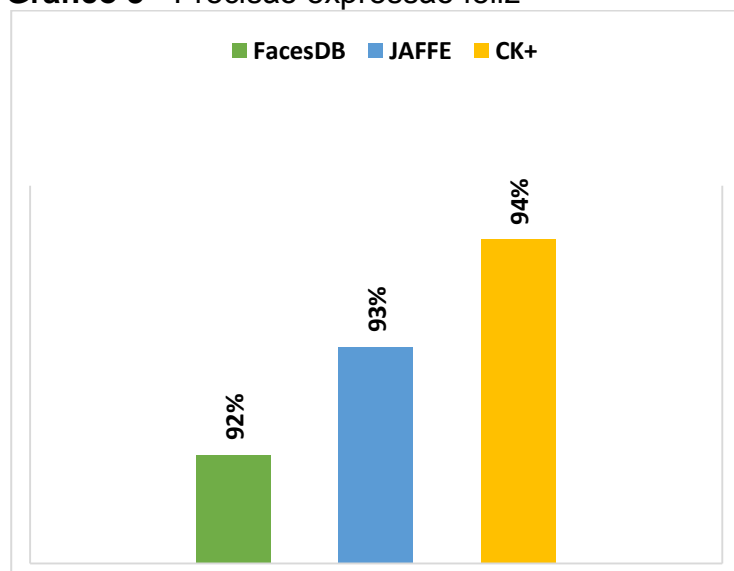
Posteriormente as imagens foram submetidas ao reconhecimento de expressões. No Gráfico 2 é ilustrado o reconhecimento da expressão neutra.

Gráfico 2 – Precisão expressão neutra

Fonte: Própria.

No Gráfico 2, a base JAFFE atingiu o valor de 94% de expressões identificadas corretamente, a base CK+ teve como acerto 84% e FacesDB 78%.

No reconhecimento da expressão feliz a base que se sobressaiu foi a CK+ com 94% de classificação da expressão corretamente, JAFFE com 93% e FacesDB com 92% como é demonstrado no Gráfico 3.

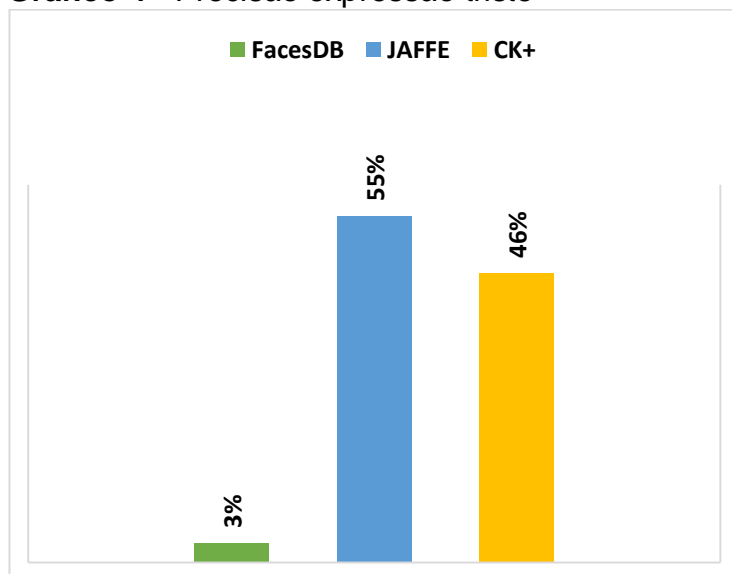
Gráfico 3 - Precisão expressão feliz

Fonte: Própria.

No Gráfico 3 a diferença apresentada foi muito pequena, sendo assim uma diferença de 1% entre as bases Ck+ e JAFFE e 2% entre as CK+ e FacesDB.

Já no Gráfico 4 ilustra o resultado da expressão do tipo triste, a base JAFFE possui os resultados mais satisfatórios tendo como taxa de acerto de 55%, em seguida a CK+ com 46% e FacesDB com apenas 3%.

Gráfico 4 - Precisão expressão triste

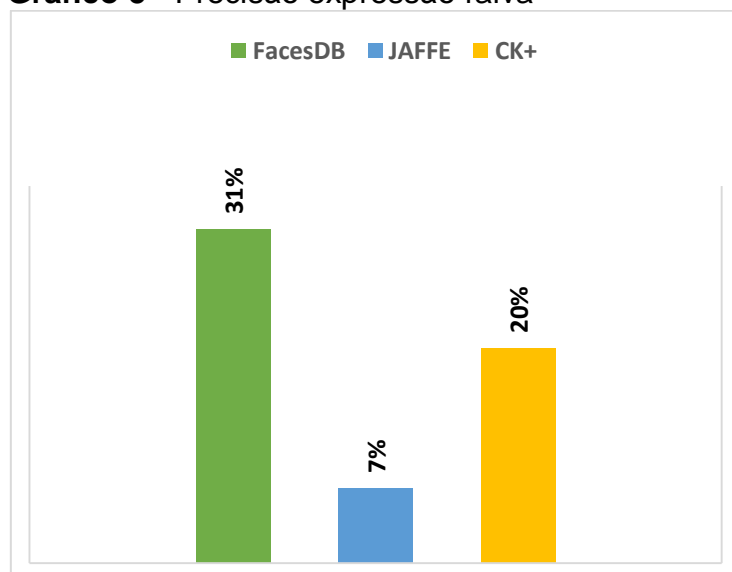


Fonte: Própria.

A taxa de acerto nas imagens que representa a expressão triste, teve uma decaída em relação a taxa de acertos das expressões apresentadas anteriores.

Nas imagens que possuem a expressão raiva, a base de imagem FacesDB se sobressaiu com 31% de acerto, CK+ com 20% e JAFFE com 7%. Como ilustra no Gráfico 5.

Gráfico 5 - Precisão expressão raiva

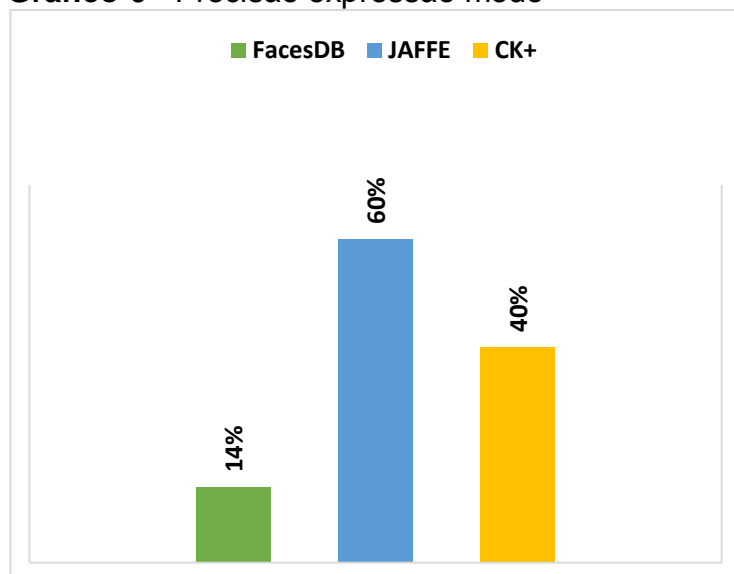


Fonte: Própria.

Pode-se observar que em ambas bases de imagens a taxa de acerto não chegou a 50%, ficando entre 31% e 7%.

No Gráfico 6 ilustra que a base JAFFE teve como taxa de acerto 60%, Ck+ com 40% e FacesDB com 14%.

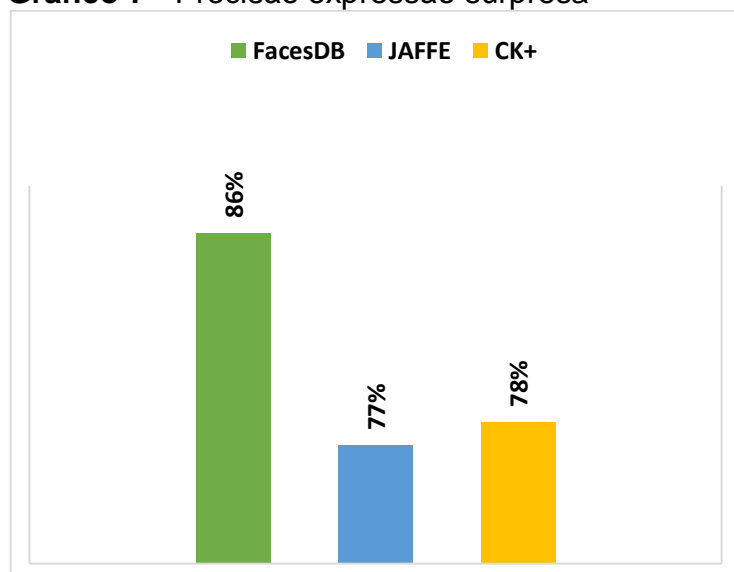
Gráfico 6 - Precisão expressão medo



Fonte: Própria.

Apenas na base JAFFE os resultados foram superiores a 50%. A base CK+ foi inferior em relação a base JAFFE com diferença de 20%, e JAFFE ficou muito a baixo em comparação com as demais.

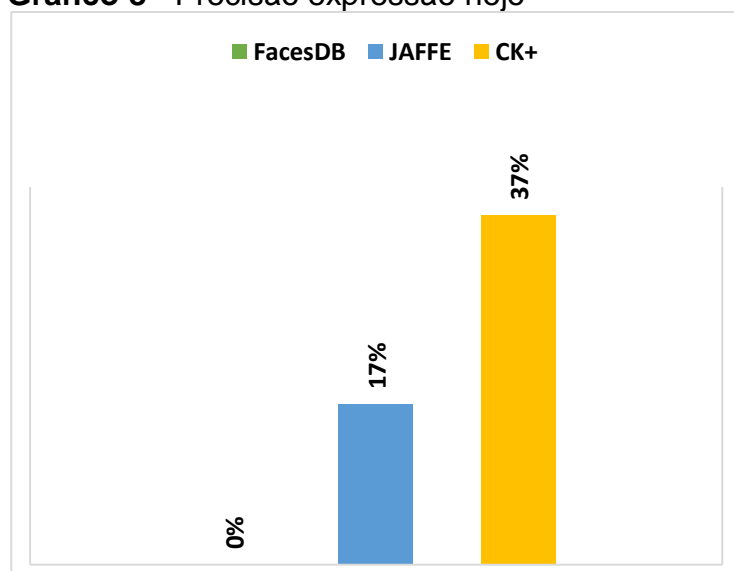
No Gráfico 7 é ilustrado a taxa de acerto das expressões de surpresa tendo a base FacesDB com 86% de taxa de acerto, JAFFE com 77% e CK+ com 78%.

Gráfico 7 - Precisão expressão surpresa

Fonte: Própria.

Sendo assim, a diferença entre a FacesDB e CK+ foi de 8% e FacesDB e JAFFE foi de 9%.

Nas imagens que representam a expressão de nojo, apenas as bases, JAFFE com 17% e CK+ com 37% conseguiram ter a expressão identificada como ilustra no Gráfico 8.

Gráfico 8 - Precisão expressão nojo

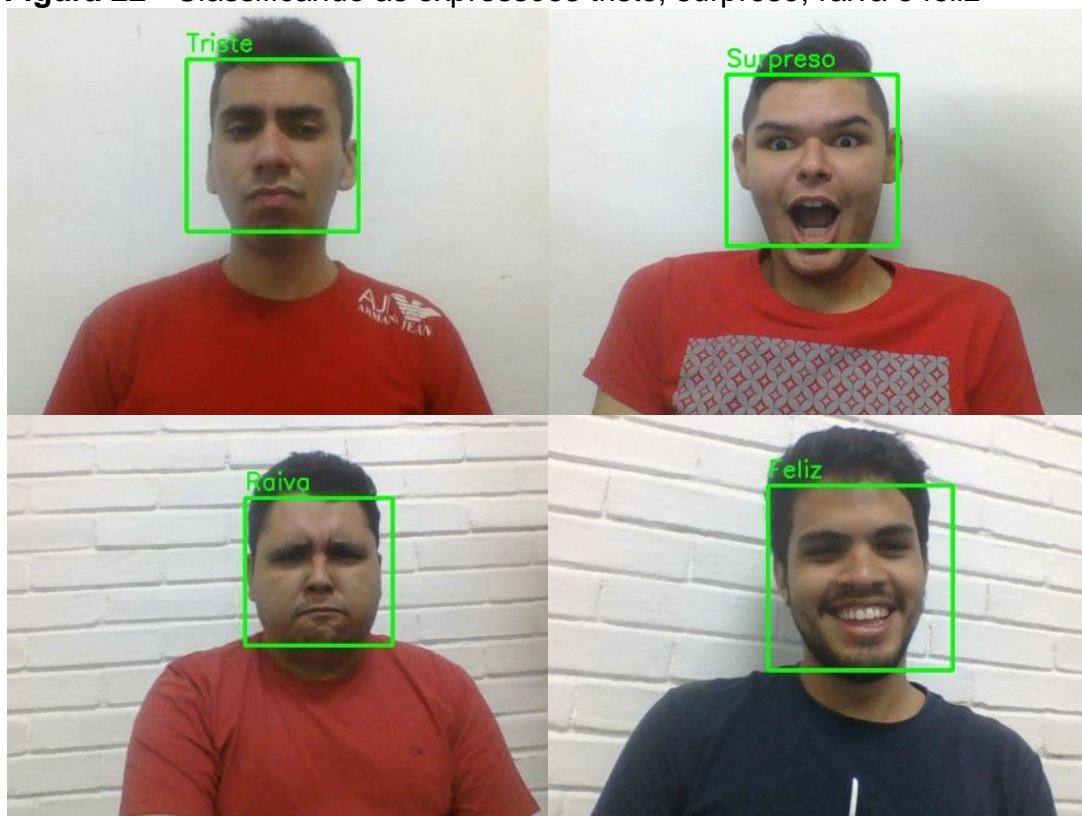
Fonte: Própria.

Na expressão Nojo a base FacesDB não obteve nenhuma classificação, e as demais bases que tiveram a expressão classificada, teve uma taxa de acerto inferior a 50%.

4.2 Testes com voluntários

Os testes foram realizados com 12 voluntários. Os mesmos tentavam reproduzir as setes expressões, para assim o algoritmo reconhecer as expressões que acontecia em determinado momento. Para isso foi selecionado 5 imagens de cada expressão, onde os voluntários olhavam para elas e usavam como inspiração ou tentavam reproduziam as mesmas. Na Figura 22 exibe a classificação de quatro expressões emitidas por voluntários classificadas corretamente.

Figura 22 - Classificando as expressões triste, surpreso, raiva e feliz



Fonte: Própria.

Foi contabilizado um total de 3770 expressões capturadas. Em alguns casos, alguns voluntários sentiram dificuldades de reproduzir expressões mesmo tendo as imagens para auxiliar.

Os resultados obtidos nos testes em tempo real são apresentados na Tabela 1, demonstrando a expressão, o total de expressão classificada e taxa de acerto. Foi observado que quando a emoção emitida pelos os voluntários não forem distinguíveis o suficiente, as expressões capturadas na aplicação são classificadas como emoção neutra, por isso valor da taxa de acerto da expressão neutra foi de

21% devido a esse problema. Por outro lado, se você utilizar a aplicação e expressar a emoção neutra, a aplicação irá classificar corretamente. O problema ocorre na variação de expressão que acaba classificando erroneamente. Essa variação é notada na Figura 23.

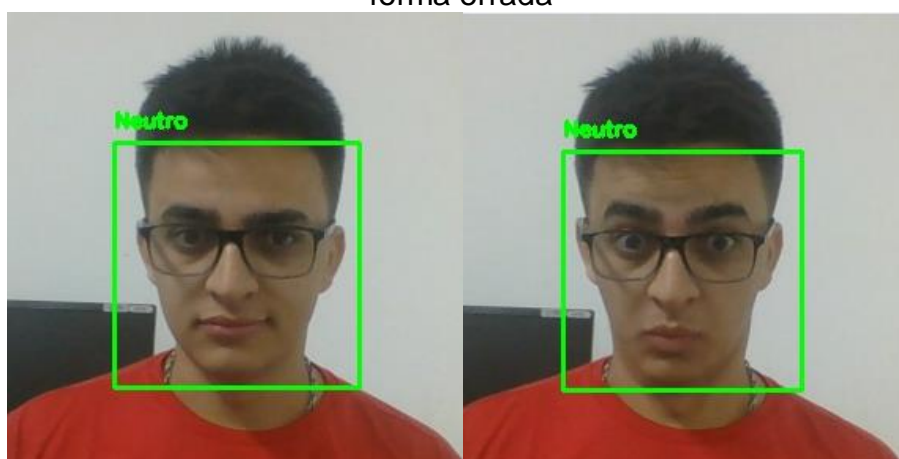
Tabela 1 - Quantidade de expressões capturadas pela aplicação

Emoção	Amostra	Precisão no reconhecimento correto
Raiva	786	30%
Nojo	131	66%
Feliz	346	57%
Triste	789	35%
Medo	533	39%
Surpresa	61	97%
Neutra	1124	21%

Fonte: Própria.

As expressões raiva, triste, medo e neutro houve uma taxa de acerto inferior a 50%, já surpreso, feliz e nojo ficou a cima de 50% a taxa de acerto. Isso aconteceu devido as variações de uma expressão para outra acaba ocorrendo erros de classificação.

Figura 23 - Variação da expressão neutra, detectando corretamente e de forma errada



Fonte: Própria.

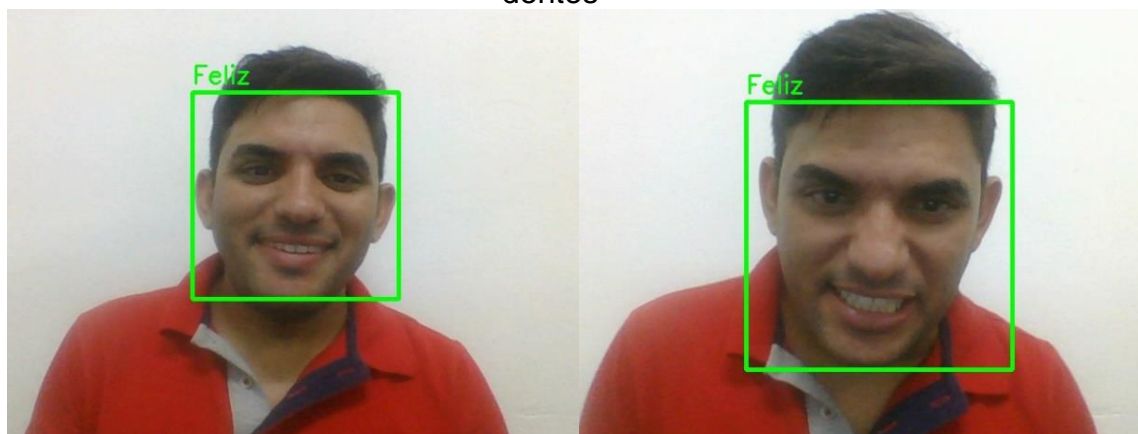
As imagens que representa surpresa, dos 12 voluntários apenas 5 conseguiram expressar e ser detectada, os demais não foram contabilizados nenhuma expressão do tipo surpresa. Foi observado que algumas vezes mesmo expressando a emoção surpresa, não detectava, mas quando detectava, a taxa de

acerto era superior em relação as demais, sem muitos de erros de confundir a expressão com outra. Pois só detectava a expressão surpresa quando as características musculares eram bem descritivas. Sendo que essa taxa de acerto foi por meio das emoções que foi classificadas corretamente, desconsiderando que houve voluntários que nenhuma emoção surpresa foi detectada.

Na expressão nojo também houve dificuldades para os voluntários expressarem, foi detectado uma pequena quantidade de expressões do tipo nojo, em apenas um voluntário não foi possível detectar a expressão. A quantidade de imagens classificadas como nojo foi superior apenas comparando com emoção surpresa.

Na expressão feliz aconteceu o mesmo problema de variação de emoção, quando os voluntários mostravam os dentes não expressando a emoção feliz, o algoritmo se confundia e classificavam como feliz. Na Figura 24 ilustra a emoção feliz sendo classificada corretamente, como também o voluntario mostra os dentes, o algoritmo se confunde e detecta como feliz.

Figura 24 - Reconhecendo a expressão feliz e errando devido a aparição dos dentes



Fonte: Própria.

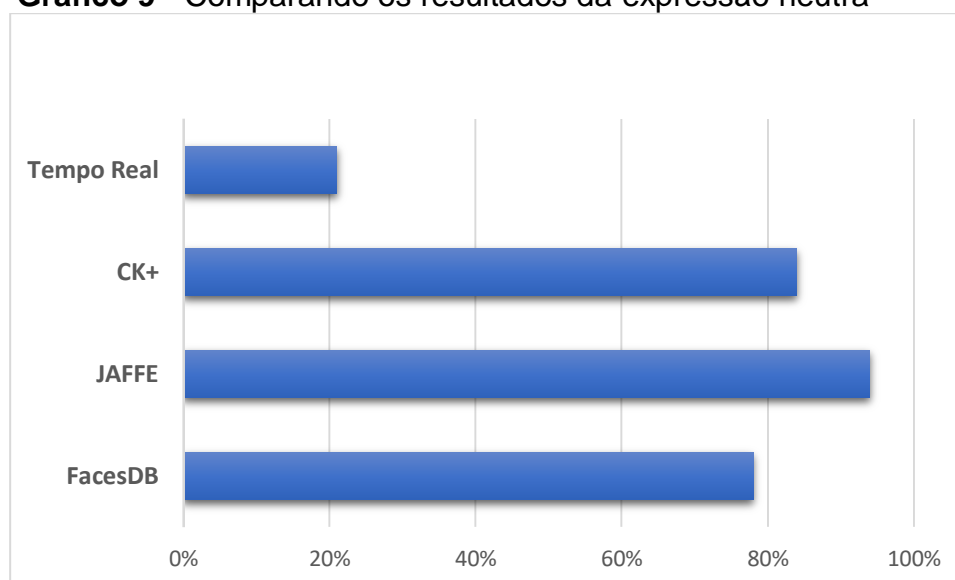
Observa-se na Figura 24, quando o voluntario mostrava os dentes de modo que parecia ser um sorriso, isso acabava confundindo o reconhecimento da expressão, tendo como resultado uma classificação errada da expressão. As emoções de neutro e feliz quando expressas corretamente a aplicação reconhecia facilmente, mas o problema que ocorre é na variação da expressão em alguns casos.

4.3 Análise dos testes realizados

Neste subtópico será discutido a taxa de acerto das expressões em tempo real comparando com as taxas de acerto das bases de imagens, será discutido cada expressão.

Nos testes com imagens das bases, a emoção neutra teve uma taxa média de acerto de 85,3%, como pode ser visto no Gráfico 9. Já nos testes com voluntários em vídeos em tempo real a taxa de acerto caiu bastante, ficou na casa dos 20%.

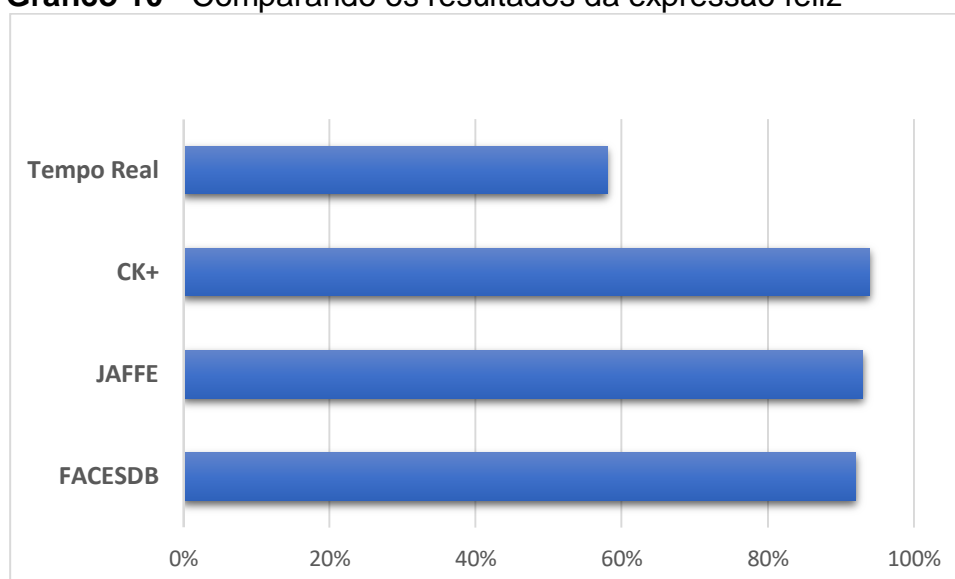
Gráfico 9 - Comparando os resultados da expressão neutra



Fonte: Própria.

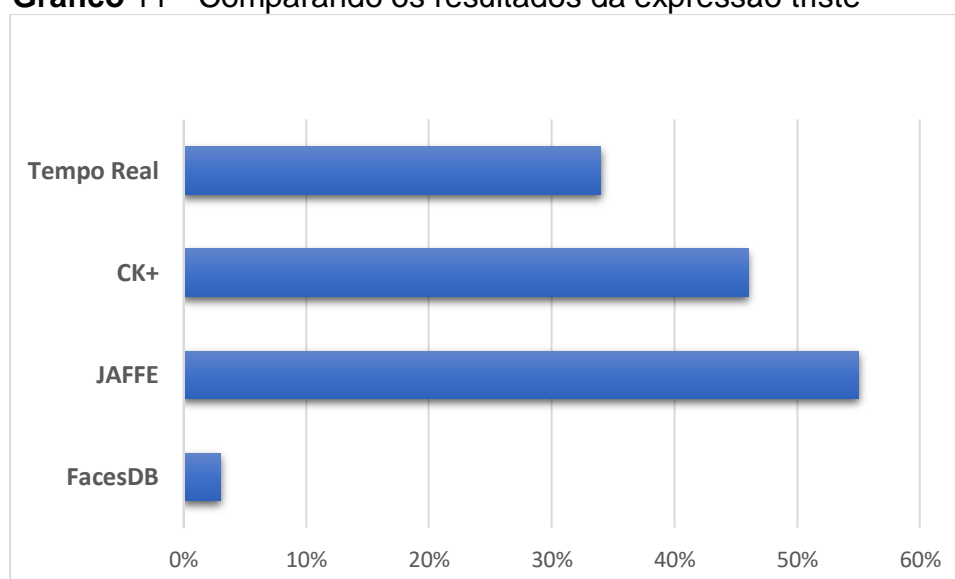
Pode-se notar no Gráfico 9, que em imagens estáticas, a taxa de acerto tem uma precisão maior em relação a vídeos em tempo real, devido que não ocorre problemas de variações devido ao movimento da face.

No Gráfico 10 é demonstrado a comparação dos resultados da emoção feliz, as bases ficaram 93% de taxa média acerto, em tempo real foi de 57%, podendo considerar como resultado muito satisfatório nas bases de imagens, no vídeo em tempo real, ficou acima dos 50%.

Gráfico 10 - Comparando os resultados da expressão feliz

Fonte: Própria.

Na emoção que representa a tristeza, os testes em tempo real tiveram um resultado superior comparando com os dois gráficos apresentados anteriormente, onde os testes em tempo real sempre ficavam para trás em relação as bases de imagens. No Gráfico 11 é apresentado os resultados desta comparação, podendo ser notado que base FacesDB teve uma taxa de acerto inferior a 10%, em tempo real com 35%, CK+ com 46% e JAFFE com 55%.

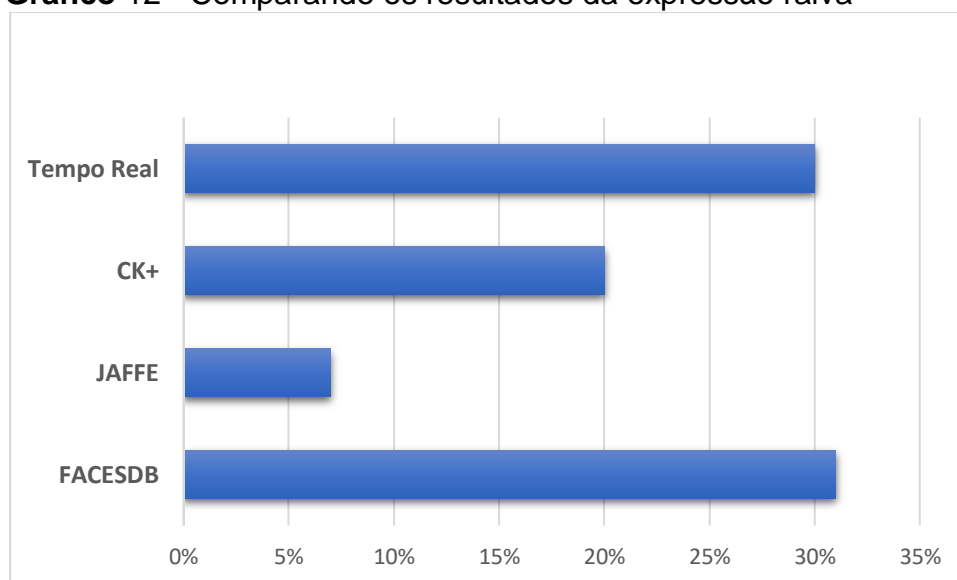
Gráfico 11 - Comparando os resultados da expressão triste

Fonte: Própria

Em contrapartida os resultados da emoção tristeza, a taxa de acerto das bases de imagens foi inferior comparando com os resultados da expressão neutra e feliz. Ficando a baixo dos 60%.

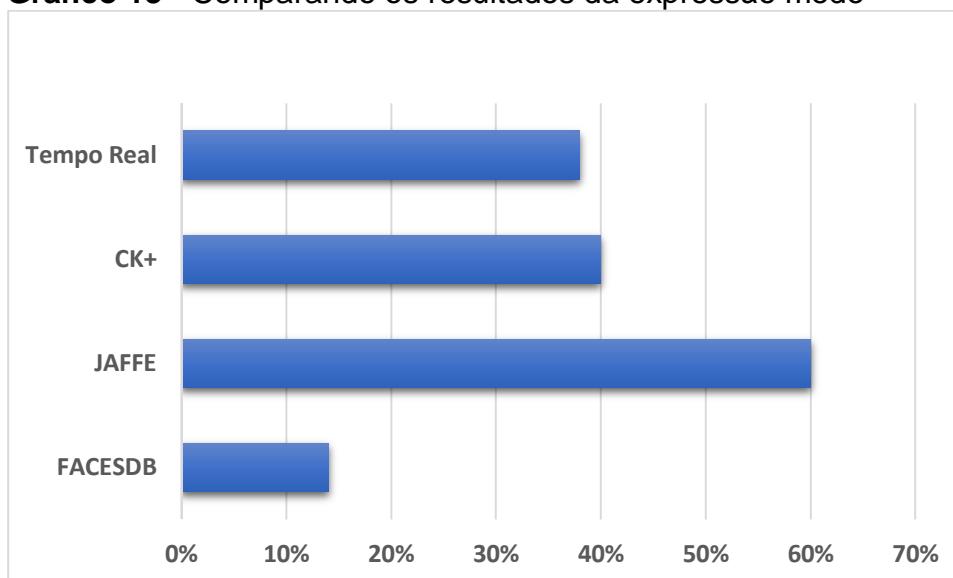
Nos resultados da emoção raiva, os resultados das bases de ficaram a baixo dos 35% de taxa de acerto, sendo mais inferior que os resultados da expressão triste. Mas em contrapartida foi na base FacesDB que a taxa de acerto foi superior com 31%, em seguida o teste em tempo real com 30%. Nesta emoção o teste em tempo real se sobressaiu em relação as bases de imagens JAFFE e CK+, tendo um desempenho inferior apenas em relação a base FacesDB. Como é possível notar no Gráfico 12.

Gráfico 12 - Comparando os resultados da expressão raiva



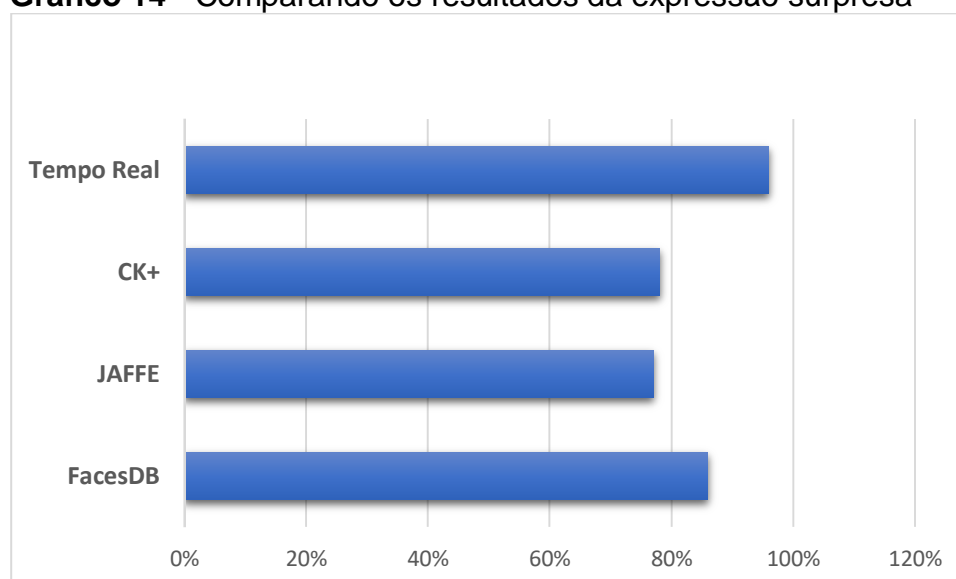
Fonte: Raiva.

Na emoção medo os resultados com maior taxa de acerto foram de 60% na base de imagens JAFFE, CK+ com 40% e em seguida o teste em tempo real com 39% e FacesDB com 14%. Sendo que o teste em tempo real ficou atrás apenas da base de imagens JAFFE se tratando da taxa de acerto, sendo superior a CK+ e JAFFE. Como é ilustrado no Gráfico 13.

Gráfico 13 - Comparando os resultados da expressão medo

Fonte: Própria.

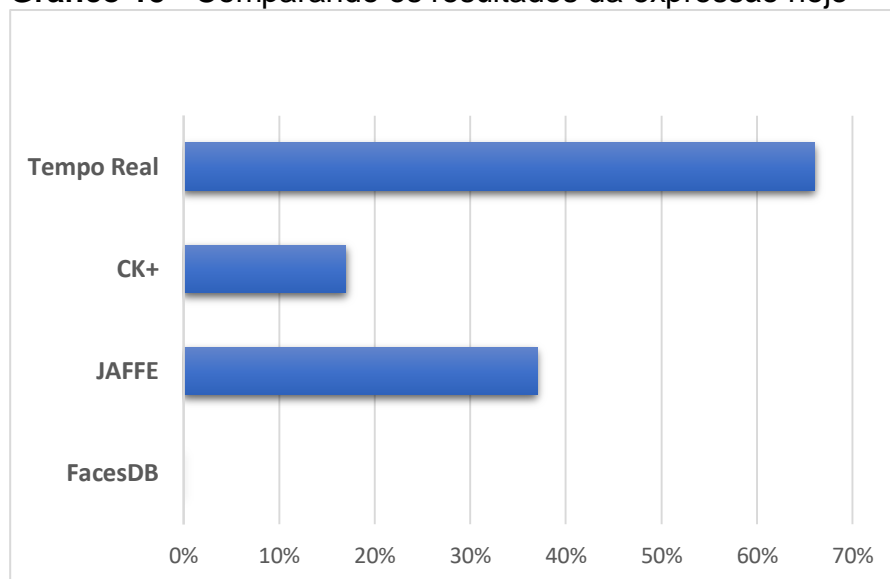
Na emoção surpresa como foi explicado anteriormente no teste em tempo real com alguns voluntários, nenhuma expressão deste tipo foi captada pela aplicação. Sendo que a taxa de acerto apresentada no Gráfico 14 a seguir foi levado em consideração apenas os voluntários que conseguiram expressar e ser reconhecida pela aplicação, tendo como resultado 97% sendo superior que os resultados das 3 bases de imagens, FacesDB com 86%, CK+ 78% e JAFPE com 77%.

Gráfico 14 - Comparando os resultados da expressão surpresa

Fonte: Própria.

Por fim a emoção nojo, teve como taxa de acerto 66% no teste em tempo real sendo muito superior aos testes com as imagens das bases, constando que a emoção nojo teve o pior resultado nas imagens das bases de forma geral, tendo a CK+ com 37%, a JAFFE com 17% e a base FacesDB, nenhuma imagem do tipo nojo foi reconhecida. Como ilustra no Gráfico 15.

Gráfico 15 - Comparando os resultados da expressão nojo



Fonte: Própria.

5 CONCLUSÃO

Nesta pesquisa foram alcançados os objetivos específicos propostos. Desta forma, foi realizando um estudo bibliográfico sobre expressões faciais e visão computacional. Como também foi treinado uma Rede Neural Convolutiva para reconhecer expressões faciais. Posteriormente a Rede Neural Convolutiva foi utilizada para o desenvolvimento da aplicação para reconhecer as expressões de imagens obtidas a partir de vídeos em tempo real por meio da webcam ou vídeos gravados.

Sendo assim, apresentamos nesta pesquisa, os resultados do desenvolvimento de uma aplicação para reconhecimento de expressões faciais, levando em consideração o reconhecimento de sete expressões faciais (neutro, feliz, raiva, surpresa, tristeza, medo e nojo).

Verificou-se que a Rede Neural Convolutiva consegue reconhecer as sete expressões faciais. Nas imagens, o reconhecimento de expressões faciais tende a ter desempenho superior, em relação aos vídeos em tempo real, devido a ocorrência da movimentação do usuário. Apesar dos problemas de confundir uma emoção com outra, leva-se em consideração que a webcam utilizada pode ter comprometido os resultados, já que possuía uma resolução baixa, como também iluminação das salas onde ocorreram os testes, que foram feitos no turno da noite na própria universidade.

5.1 Trabalhos Futuros

Como trabalho futuro, pode-se utilizar a junção de várias bases de imagens para aumentar o número de imagens na fase de treinamento para que haja possibilidade de se obter uma taxa de acerto com uma precisão maior em relação à qual foi apresentada nos testes. Como também abre a possibilidade de criar uma base própria de imagens para a utilização.

Além disso, pode-se utilizar um computador que possua uma configuração superior, utilizando da GPU, para realizar os testes e o treinamento, como também a utilização de uma webcam com maior resolução para entender como a aplicação se comporta.

REFERÊNCIAS

- ARAÚJO, Flávio H. D. *et al.*, **Redes Neurais Convolucionais com Tensorflow: Teoria e Prática**. SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. III Escola Regional de Informática do Piauí. Livro Anais-Artigos e Minicursos, v. 1, p. 382-406, 2017.
- ARRIAGA, Octavio; PLÖGER, Paul; VALDENEGRO, Matias. **Real-time convolutional neural networks for emotion and gender classification**. ArXiv preprint arXiv:1710.07557, 2017.
- BACKES, André Ricardo; SÁ JUNIOR, Jarbas Joaci de Mesquita. **Introdução à visão computacional usando Matlab**. Alta Books Editora, 2016.
- BARIELLE, Felipe. **Introdução à visão computacional**. Casa do Código, 2018. 256 p.
- COSSETI, Marcelo Júnior. **Reconhecimento De Expressões Faciais Utilizando Redução De Dimensionalidade Para Estratégia De Classificação Um-Contra-Um**. 2015. 137 f. Dissertação (Mestrado em Informática) - Pontifícia Universidade Católica do Paraná, Curitiba, 2015.
- COSTA, Saulo William S.; DE SOUSA, Ailton Lopes; PIRES, Yomara. **Computação Afetiva: Uma ferramenta para avaliar aspectos afetivos em aplicações computacionais**. In: VI Encontro Anual de Tecnologia da Informação, 7. 2015. Frederico Westphalen. **Anais...** Frederico Westphalen: Encontro Anual de Tecnologia da Informação, 2015. p. 186-290.
- CRUZ, Juliano Elias Cardoso. **Reconhecimento de objetos em imagens orbitais com o uso de abordagens do tipo descritor-classificador**. 2014. 107 f. Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014.
- DA FONSECA, Fernando Otávio Gomes. **Detector de faces utilizando filtros de características**. 2016. 111 f. Dissertação (Mestrado em Engenharia Elétrica e de Telecomunicações) - Universidade Federal Fluminense, Niterói, 2016.
- DALAL, Navneet; TRIGGS, Bill. Histograms of oriented gradients for human detection. In: **international Conference on computer vision & Pattern Recognition (CVPR'05)**. IEEE Computer Society, 2005. p. 886-893.
- DE JESUS, Augusto Batista *et al.* **Protótipo de reconhecimento de expressões faciais com computação afetiva na educação**. Instituto Federal do Tocantins. Campus Palma. Jornal de iniciação científica e extensão, 2018.
- DE MILANO, Danilo; HONORATO, Luciano Barrozo. **Visão computacional**. Universidade Estadual de Campinas - (Unicamp), Faculdade de Tecnologia, 2010.

DE OLIVEIRA, Eduardo; JAQUES, Patrícia Augustin. Classificação de emoções básicas através de imagens capturadas por webcam. **Revista Brasileira de Computação Aplicada**, v. 5, n. 2, p. 40-54, 2013.

DE SOUSA, Ailton Lopes *et al.*, FOURFACE: Uma ferramenta de reconhecimento de expressões faciais. In: VII Encontro Anual de Tecnologia da Informação, 7. 2016. Frederico Westphalen. **Anais...** Frederico Westphalen: Encontro Anual de Tecnologia da Informação, 2016. p. 185-192.

EKMAN, Paul. **A linguagem das emoções**. São Paulo: Lua de Papel, 2011. 288 p.

Ekman, P; Friesen, W. V. **Pictures of Facial Affect. Consulting Psychologists Press**, 1976.

FINOCCHIO, Marco Antonio Ferreira. **Noções de redes neurais artificiais**. Apostila. Universidade Tecnológica Federal do Paraná, 2014.

GUEDES, André Bernardes Soares. **Reconhecimento de Gestos usando Redes Neurais Convolucionadas**. 2017. Monografia (Bacharelado em Engenharia de Software) - Universidade de Brasília, Brasília, 2017.

Gonzalez, Rafael.C. Woods; Richard E. **Processamento Digital de Imagens**. São Paulo: Pearson Universidades, 2009. 624 p.

GONÇALVES, Luís Pedro Nabais. **Reconhecimento de Gestos para Interação com Robô Móvel em Ambiente Pediátrico**. 2017. 74 f. Dissertação (Mestrado em Engenharia Eletrotécnica e de Computadores) - Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal, 2017.

GOUVEIA, W. R; PAIVA, M. S. V. **Detecção de Faces Humanas em Imagens Coloridas Utilizando Redes Neurais Artificiais**, 2009.

HAYKIN, Simon. **Redes Neurais: Princípios e Prática**. Porto Alegre: Bookman, 2001. 900 p.

LIBRALON, Giampaolo Luiz. **Modelagem computacional para reconhecimento de emoções baseada na análise facial**. 2014. 220 f. Tese (Doutorando em Ciências de Computação e Matemático Computacional) - Universidade de São Paulo, São Carlos, 2014.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): **A complete expression dataset for action unit and emotion-specified expression**. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.

Kanade, T., Cohn, J. F., & Tian, Y. (2000). **Comprehensive database for facial expression analysis**. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.

MENDONÇA, Thomas Dillan Baltazar. **Sistema de reconhecimento de expressões faciais para classificação de emoções de usuários em sistemas computacionais**. 2018. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade Federal do Ceará, Russas, 2018.

Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba. **Coding Facial Expressions with Gabor Wavelets**, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998).
<http://doi.org/10.1109/AFGR.1998.670949>
Open access content available at: <https://zenodo.org/record/3430156>

PANCERI, João Antonio Campos *et al.*, RECONHECIMENTO FACIAL BASEADO EM HOG E PCA: UMA COMPARAÇÃO QUANTO À INVARIÂNCIA À ILUMINAÇÃO. **Revista Ifes Ciência**-ISSN 2359-4799, v. 1, n. 1, 2015.

PANCERI, João Antonio Campos. **Reconhecimento de Expressões Faciais Baseado em Active Appearance Model**. 2017. 77 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Espírito Santo, Brasil 2017.

PARADA, Daniel Mauricio Perdernera. **Reconhecimento de expressões faciais compostas em imagens 3D: ambiente forçado vs ambiente espontâneo**. 2017. 62 f. Dissertação (Mestrado em Informática) - Universidade Federal do Paraná, Curitiba, 2017.

PEREIRA, Tiago. **O QUE É VISÃO COMPUTACIONAL?** 2018. Disponível em: <http://datascienceacademy.com.br/blog/o-que-e-visao-computacional>. Acesso em: 31 mar. 2019.

RIGHETTO, Guilherme. **O uso da rede neural convolucional como extrator de características aplicado ao problema de identificação de escritores**. 2016. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade Tecnológica Federal do Paraná, Campo Mourão, 2016

SANTIAGO, Hemir da Cunha. **Reconhecimento de expressões faciais utilizando estimação de movimento**. 2017. 141 f. Tese (Doutorando em Ciência da Computação) - Universidade Federal de Pernambuco, Recife, 2017.

SANTOS, Carlos Alexandre Silva dos. **Reconhecimento de imagens de marcas de gado utilizando redes neurais convolucionais e máquinas de vetores de suporte**. 2017. 136 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Pampa, Alegrete, 2017.

VIOLA, Paul; JONES, Michael. **Rapid Object Detection using a Boosted Cascade of Simple Features**. CVPR (1), v. 1, p. 511-518, 2001.

WAGNER, Felipe Rocha. **Análise Antropométrica Semiautomática em Imersão para Pesquisa e Diagnóstico Clínico de Síndromes Dismórficas**. 2017. Dissertação (Mestrado em Computação Aplicada) - Universidade do Vale do Rio dos Sinos, São Leopoldo, 2017.

ANEXO A – TERMO DE AUTORIZAÇÃO DE IMAGEM.**TERMO DE AUTORIZAÇÃO PARA USO DE IMAGENS (TCFV)
(FOTOS E VÍDEOS)**

Eu, _____, **AUTORIZO** o Ricardo de Sousa Farias, coordenador(a) da pesquisa intitulada: Aplicação para Detecção e Reconhecimento de Expressões Faciais com Redes Neurais Convolucionais a fixar, armazenar e exibir a minha imagem por meio de foto com o fim específico de inseri-la nas informações que serão geradas na pesquisa, aqui citada, e em outras publicações dela decorrentes, quais sejam: revistas científicas, jornais, congressos, entre outros eventos dessa natureza.

A presente autorização abrange, exclusivamente, o uso de minha imagem para os fins aqui estabelecidos e deverá sempre preservar o meu anonimato. Qualquer outra forma de utilização e/ou reprodução deverá ser por mim autorizada, em observância ao Art. 5º, X e XXVIII, alínea “a” da Constituição Federal de 1988.

O pesquisador responsável Ricardo de Sousa Farias, assegurou-me que os dados serão armazenados em seu computador pessoal, sob sua responsabilidade, por 5 anos, e após esse período, serão destruídas.

Assegurou-me, também, que serei livre para interromper minha participação na pesquisa a qualquer momento e/ou solicitar a posse de minhas imagens.

Ademais, tais compromissos estão em conformidade com as diretrizes previstas na Resolução Nº. 466/12 do Conselho Nacional de Saúde do Ministério da Saúde/Comissão Nacional de Ética em Pesquisa, que dispõe sobre Ética em Pesquisa que envolve Seres Humanos.

Patos, ____ de ____ de ____.

Assinatura do participante da pesquisa

Assinatura do pesquisador responsável