

CARLOS EDUARDO CICHON HENRIQUES

INFERÊNCIA DE EMOÇÕES UTILIZANDO DETECÇÃO FACIAL E REDES NEURAIIS
CONVOLUCIONAIS

(versão pré-defesa, compilada em 22 de dezembro de 2023)

Trabalho apresentado como requisito parcial à conclusão do Curso de Bacharelado em Ciência da Computação, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Andrey Ricardo Pimentel.

CURITIBA PR

2023

RESUMO

A representação e reconhecimento de emoções no âmbito computacional é um grande problema que ainda não apresenta consenso no mundo acadêmico e justamente por isso, o estudo de emoções e a computação afetiva estão em uma crescente nas últimas décadas. Com o Covid-19 atingindo o mundo todo, diversas pessoas tiveram de migrar do contato pessoal para ambientes virtuais, seja ele se tratando de trabalho ou se tratando de ensino. Com essa migração, softwares de reuniões virtuais e videoconferências ficaram cada vez mais populares e sua utilização cresceu exponencialmente. Apesar de terem sido a solução para o maior problema: a distância, estas ferramentas trouxeram novos desafios, principalmente no âmbito de feedback dos participantes. Com as interações “olho no olho” impedidas, a que sobrou se limitava a visualização de pequenos quadros com imagens dos participantes, e, muitas vezes, nem todos estavam inclusos ali. Desta forma, apresentadores em reuniões e tutores no ensino à distância ficaram sem saber o que os participantes estavam achando de suas apresentações e aulas. Este trabalho busca mitigar este problema provendo uma ferramenta de inferência de emoções baseada na imagem capturada de participantes, utilizando técnicas de reconhecimento facial e redes neurais para prover à organizadores e professores uma noção de como suas apresentações estão sendo recebidas, mesmo sem poder olhar cada um dos participantes presencialmente.

Palavras-chave: Inferência. Emoções. Redes Neurais.

ABSTRACT

The representation and recognition of emotions in the computational field is a major problem that still lacks consensus in the academic world and precisely for this reason, the study of emotions and affective computing have been growing in recent decades. With the Covid-19 pandemic reaching the entire world, many people had to migrate from personal contact to virtual environments, whether it be for working or teaching. With this migration, virtual meeting and video conferencing software has become increasingly popular and its use has grown exponentially. Despite being the solution to the biggest problem: distance, these tools brought new challenges, mainly in terms of feedback from participants. With “eye to eye” interactions prevented, the one left was limited to viewing small frames with images of the participants, and, often, not everyone was included there. In this way, presenters at meetings and distance learning tutors were left without knowing what participants were thinking about their presentations and classes. This work seeks to mitigate this problem by providing an emotion inference tool based on captured images of participants, using facial recognition techniques and neural networks to provide organizers and teachers a sense of how their presentations are being received, even without being able to look at each participant in person.

Keywords: Inference. Emotions. Neural Networks.

LISTA DE FIGURAS

1.1	Contagem de pico de sessões por continente (Google Meet).	9
3.1	Face detectada pelo classificador em cascatadas do OpenCV	15
3.2	Múltiplas faces detectadas pelo classificador em cascatadas do OpenCV	15
4.1	Expressões em classificação. (Gottardo, 2018).	17
5.1	Amostra de imagens contidas na AffectNet. Fonte: Mollahosseini et al. (2017). .	20
5.2	Estrutura da CNN construída	21
6.1	Exemplo de imagens classificadas pela rede.	23
6.2	Matriz de Confusão para a CNN proposta	25
7.1	Tela inicial da aplicação.	27
7.2	Tela para entrada na aplicação	28
7.3	Tela de cadastro.	29
7.4	Tela de visualização de salas.	29
7.5	Tela de cadastro de salas.	30
7.6	Tela para inserção de senha da sala.	30
7.7	Telas de visão do dono da sala.	30
7.8	Tela de visão do participante.	31

LISTA DE TABELAS

2.1	Comparação de rótulos para emoções na literatura	11
5.1	Número de imagens por emoção	20
6.1	Precisão por classe	26

LISTA DE ACRÔNIMOS

UFPR	Universidade Federal do Paraná
EaD	Ensino a Distância
CNN	Convolutional Neural Network
STI	Sistemas Tutores Inteligentes
STA	Sistemas Tutores Afetivos
OpenCV	Open Computer Vision
RGB	Red Green Blue
SBDG	Sistema Gerenciador de Banco de Dados

SUMÁRIO

1	INTRODUÇÃO	9
1.1	OBJETIVO	10
1.2	ESTRUTURA	10
2	EMOÇÕES E COMPUTAÇÃO AFETIVA	11
2.1	ABORDAGEM CATEGÓRICA	11
2.2	ABORDAGEM DIMENSIONAL	12
2.3	REPRESENTAÇÃO NESTE ESTUDO	12
2.4	COMPUTAÇÃO AFETIVA	12
2.5	CONCLUSÃO	13
3	DETECÇÃO FACIAL E REDES NEURAIS	14
3.1	CLASSIFICADOR EM CASCATA	14
3.2	CONEXÃO COM REDES NEURAIS	14
3.3	CNNS	15
3.4	CONCLUSÃO	16
4	ESTADO DA ARTE	17
5	PROPOSTA	19
5.1	BASE DE DADOS	19
5.1.1	Preparo das imagens	19
5.2	CONSTRUÇÃO E TREINAMENTO DA CNN	21
5.2.1	Treinamento da Rede	21
5.3	CONCLUSÃO	22
6	RESULTADOS	23
6.1	MATRIZ DE CONFUSÃO	23
6.2	OUTRAS MÉTRICAS	24
6.2.1	Acurácia	24
6.2.2	Precisão	24
7	PROTOTIPO DE APLICAÇÃO	27
7.1	CONCEITOS E ESTRUTURAS	27
7.2	AUTENTICAÇÃO E USUÁRIOS	28
7.3	SALAS	28
7.4	MOMENTOS	29
7.5	CAPTURA E CLASSIFICAÇÃO DE IMAGENS	31
8	CONCLUSÃO	32
8.1	O ESTUDO	32
8.2	POSSÍVEIS APLICAÇÕES	32

8.3	TRABALHOS FUTUROS	33
	REFERÊNCIAS	34
	APÊNDICE A – CÓDIGOS UTILIZADOS NO PRÉ-PROCESSAMENTO	
	DE IMAGENS	36
A.1	CONVERSÃO PARA ESCALA DE CINZA.	36
A.2	ESPELHAMENTO DE IMAGENS	36
A.3	AJUSTE DE BRILHO DE IMAGENS	37
A.4	AJUSTE DE CONTRASTE DE IMAGENS	37
A.5	ESCOLHA ALEATÓRIA DE IMAGENS	38
	APÊNDICE B – ESPECIFICAÇÕES TÉCNICAS DA MÁQUINA QUE	
	TREINOU A CNN.	39

1 INTRODUÇÃO

A pandemia de Covid-19 trouxe, entre tantos outros, desafios na adaptação dos ambientes de contato social para ambientes virtuais.

Estimava-se que 1.38 bilhões de estudantes fossem impactados mundialmente com o fechamento de escolas (McCarthy, 2020) e além disso foi apontada uma perda de cerca de 8.5 trilhões de dólares para a economia mundial (Nations, 2020). Como forma de mitigar estas dificuldades, tanto o ambiente acadêmico quanto o corporativo tiveram de se adaptar à novas ferramentas, entre elas as plataformas de videoconferência como Google Meet, Zoom, Microsoft Teams e diversas outras. Segundo a Google, a demanda dos usuários pela sua aplicação cresceu cerca de 30 vezes em março de 2020 (Schaevitz, 2020) e em concordância a esse movimento o aplicativo do Zoom na Apple Store foi baixado 2.13 milhões de vezes em um único dia do mesmo mês (CNN, 2020).

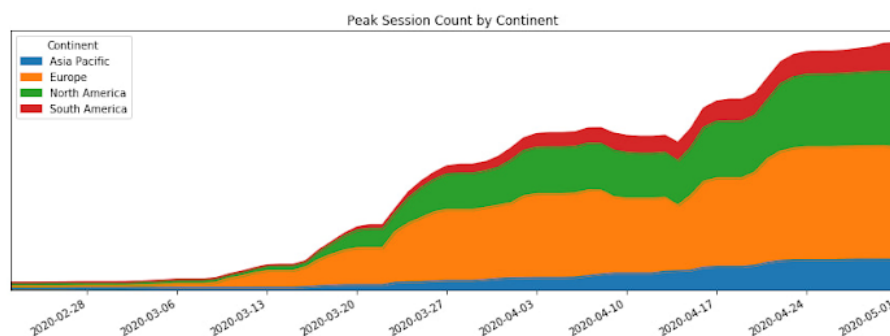


Figura 1.1: Contagem de pico de sessões por continente (Google Meet).

Apesar de ter sido a solução do momento, o EaD (*Ensino a Distância*) apresentou novos problemas. Como é mostrado por (Orhan e Beyhan, 2020) em sua pesquisa, quando questionados sobre o EaD em comparação com o ensino regular, professores frequentemente citam os termos “gerenciamento de turma inefetivo” e “ausência de feedback dos estudantes”. Além disso, ao serem questionados sobre suas percepções do ensino a distância, ele foi citado como “ensino sem sentimento”. Dentro das empresas não foi diferente, (Bergmann et al., 2022) mostra em seu estudo funcionários que usam videoconferências e apontam a relação da dificuldade de saber o humor, atenção e linguagem corporal de pessoas com a diminuição da conexão pessoal entre elas.

Percebe-se então que ambientes virtuais, sejam voltados ao aprendizado ou ao trabalho, são limitados quanto a prover informações sobre sentimento e engajamento de participantes em videoconferências.

1.1 OBJETIVO

A deficiência de feedback para tutores ou apresentadores por parte de ferramentas de reuniões virtuais é o que pauta a construção deste trabalho, que visa, de forma geral, desenvolver uma ferramenta de inferência de emoções a partir de imagens captadas de participantes.

Mais especificamente, propõe-se a utilização de detecção facial e a construção de uma CNN treinada com uma grande base de imagens de emoções já catalogadas em: alegria, surpresa, raiva, desgosto, surpresa, medo, contentamento e neutralidade para atingir o objetivo geral.

1.2 ESTRUTURA

Esse estudo está dividido em oito capítulos.

O capítulo dois mostra uma visão geral sobre o que a literatura diz a respeito e como define estados afetivos, bem como estudos relacionados com o objetivo deste trabalho. O capítulo três busca definir conceitos de detecção facial e também de deep learning e de redes neurais, como a CNN utilizada.

No quarto capítulo, mostra-se o estado da arte para o problema a ser tratado neste trabalho. No quinto, a ferramenta de inferência desenvolvida em resposta ao objetivo proposto é apresentada. Ele descreve sua metodologia: base de dados utilizada, algoritmos de preparação e a construção da rede neural. O sexto capítulo apresenta os resultados obtidos após o desenvolvimento da rede. Já o sétimo capítulo apresenta um protótipo de aplicação desenvolvida para empregar a inferência em um ambiente real.

Por fim, o oitavo e último capítulo agrega toda a informação produzida, resume o que foi alcançado com o trabalho e mostra o atendimento do objetivo geral. Além disso, mostra possíveis aplicações futuras e sugestões de expansões para o trabalho atual.

2 EMOÇÕES E COMPUTAÇÃO AFETIVA

Apesar de não ser o foco principal deste trabalho detalhar e debater as definições e teorias sobre emoções, entender como a comunidade científica vem as estudando e classificando se mostra importante. Portanto, ao longo deste capítulo será introduzida a base conceitual de emoções e sua relação com a Computação Afetiva, área na qual este trabalho se encaixa. Além disso, serão mostrados trabalhos relacionados e possíveis aplicações sobre a inferência de emoções.

Definir emoção ainda é um problema notório entre os estudiosos do tema (Scherer, 2005). Sem um conceito estabelecido em consenso, ainda é difícil atingir avanços em teorias e pesquisas. Apesar disso, os trabalhos das últimas décadas convergiram para duas correntes principais de pensamento e teoria: uma abordagem dimensional e uma abordagem categórica, a qual será utilizada ao longo deste trabalho.

2.1 ABORDAGEM CATEGÓRICA

A abordagem categórica define um conjunto discreto de rótulos para emoções. Sua quantidade e quais são esses rótulos divergem entre a comunidade, Ekman (Ekman, 1992) por exemplo, usa sete rótulos, Plutchik (Plutchik, 1980) oito e Izard (Izard, 1990) dez emoções. Apesar dos diferentes conjuntos, muitos rótulos se repetem na literatura, como mostra a tabela 2.1.

Tabela 2.1: Comparação de rótulos para emoções na literatura

	Ekman	Plutchik	Izard
Raiva	•	•	—
Medo	•	•	•
Alegria	•	•	—
Surpresa	•	•	•
Desgosto	•	•	•
Tristeza	•	•	•
Desprezo	•	—	—
Antecipação	—	•	—
Confiança	—	•	—
Contentamento	—	—	•
Vergonha	—	—	•
Culpa	—	—	•
Interesse	—	—	•

2.2 ABORDAGEM DIMENSIONAL

Pesquisadores mostram que em situações do dia a dia, pessoas expressam estados afetivos não básicos, sutis e mais complexos como pensamento, vergonha e depressão. Tais estados são exibidos por meio de várias expressões corporais e faciais possíveis. Dessa forma, um único conjunto de rótulos discretos como na abordagem categórica pode não refletir a complexidade desses estados. (Gunes et al., 2005)

Para acomodar esses estados afetivos complexos, alguns pesquisadores (Gunes et al., 2005) utilizam modelos dimensionais para emoções. O mais amplamente usado foi proposto por Russel (Russell, 1980), neste modelo, as emoções não são consideradas independentes entre si, mas sim possuem uma representação sistemática em duas dimensões: valência e ativação.

2.3 REPRESENTAÇÃO NESTE ESTUDO

Para a inferência de emoções proposta por este trabalho, optou-se por uma abordagem categórica contendo o seguinte conjunto de rótulos: raiva, medo, alegria, surpresa, desgosto e tristeza, que são os seis rótulos mais frequentemente utilizados na comunidade acadêmica (vide tabela 2.1), além de contentamento e mais um rótulo neutro amplamente utilizado em aplicações computacionais (Gottardo, 2018) para representar um estado “sem emoção”.

A inclusão do rótulo contentamento ao conjunto visa abranger a classe homônima presente na base de dados utilizada (Mollahosseini et al., 2017) que será detalhada posteriormente no capítulo quatro.

2.4 COMPUTAÇÃO AFETIVA

A Computação Afetiva é o termo empregado para a computação que se relaciona, surge de ou deliberadamente influencia emoções (Picard, 2000), sendo assim, é a área que melhor descreve e engloba os objetivos deste trabalho.

Percebe-se nos últimos anos um aumento de softwares e pesquisas nessa área, principalmente aqueles ligados à Sistemas Tutores Afetivos (STA), especializações de Sistemas Tutores Inteligentes (STI), que agregam funcionalidades de adaptação aos estados afetivos dos estudantes (Gottardo, 2018).

Gottardo (2018), por exemplo utiliza uma técnica mista de inferência de estados afetivos baseada em informações físicas (detecção facial) e medições cognitivas para então adaptar o sistema tutor da melhor maneira ao estado afetivo detectado.

Diferente de um sistema tutor, este trabalho não busca realizar adaptações conforme as emoções de estudantes ou participantes, mas pretende detectar de forma rápida e prática as emoções para prover esta informação ao organizador de uma reunião.

2.5 CONCLUSÃO

A partir da contextualização sobre Computação Afetiva, entende-se a importância do crescimento dessa área dentro da computação e as bases que norteiam este trabalho. Além disso, pôde-se ver como as emoções, importantes na computação afetiva e para a sociedade em geral, são representadas cientificamente na literatura e pautar como elas serão abordadas neste estudo.

3 DETECÇÃO FACIAL E REDES NEURAIAS

Apesar do termo “detecção facial” ter ganhado mais popularidade nas últimas década com os avanços na computação, originalmente ele refere-se a uma habilidade inata dos seres humanos, sendo que até mesmo bebês de um até três dias de vida conseguem realizar detecção e até mesmo distinção de faces. (Turati et al., 2006).

A “réplica” deste termo na computação se dá por conta do aumento do desenvolvimento de aplicações que capturam mídias (imagens e vídeos) de usuários e além de tudo realizam processamentos sobre elas. Atualmente, qualquer software de captura ou galeria de fotos, como o Google Fotos (Google LLC, 2015), por exemplo, realiza este tipo de detecção. Neste trabalho, a detecção facial se mostra como parte importante e inicial, já que não seria possível inferir emoções a partir de imagens capturadas sem avaliar se há mesmo faces nessa imagem.

3.1 CLASSIFICADOR EM CASCATA

A técnica de detecção utilizada neste trabalho foi proposta por Viola e Jones (2004). Nela diferentes classificadores sucessivamente mais complexos são combinados em uma estrutura de cascata que aumenta a velocidade da detecção por focar apenas em regiões promissoras de uma imagem. No algoritmo, a imagem é dividida em sessões e todas essas sessões são filtradas (e possivelmente descartadas) por cada um dos classificadores, até chegar no classificador final. Esta abordagem levou ao resultado de uma detecção de face aproximadamente 15 vezes mais rápida do que qualquer outra proposta anteriormente (Viola e Jones, 2004), sendo assim extremamente apropriada para uma aplicação em tempo real, como a construída no presente trabalho.

A implementação dos classificadores em cascata foi obtida através da biblioteca OpenCV (Open Source Computer Vision), em Python, e tem resultados bastante satisfatórios, como o da figura 3.1.

3.2 CONEXÃO COM REDES NEURAIAS

O avanço e popularização das redes neurais nos últimos anos colaborou com o desenvolvimento e aperfeiçoamento em diversos problemas da computação, principalmente quando relacionados com visão computacional, como segmentação semântica, segmentação biomédica, restauração de imagens e reconhecimento de padrões (Koziarski e Cyganek, 2019). O objetivo desta sessão não é explicar como funcionam as redes em si, mas sim mostrar como um de seus tipos foi utilizado neste trabalho e em que contexto.

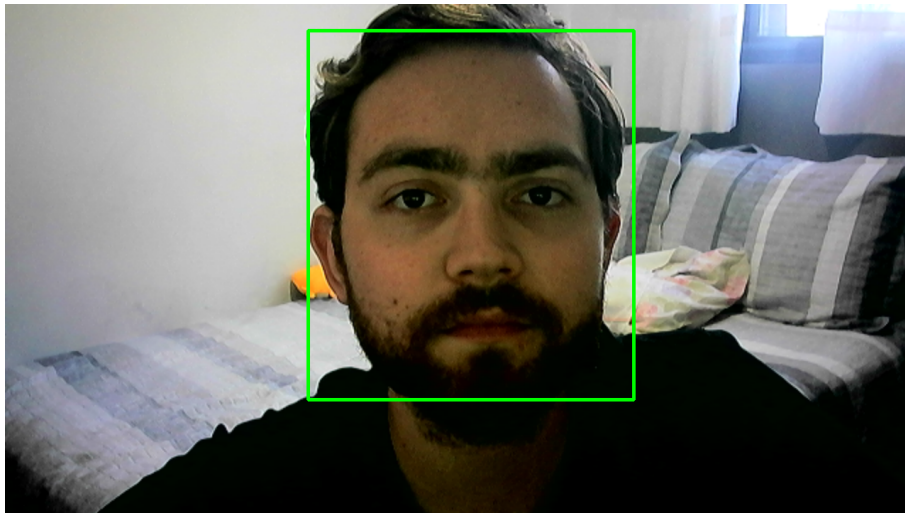


Figura 3.1: Face detectada pelo classificador em cascatadas do OpenCV

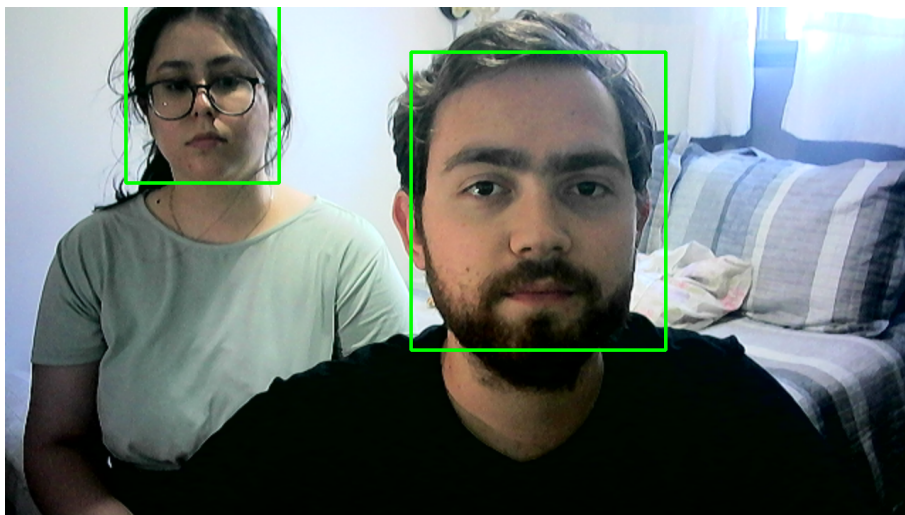


Figura 3.2: Múltiplas faces detectadas pelo classificador em cascatadas do OpenCV

3.3 CNNs

Neste estudo, uma rede neural ocupa o papel principal logo após a detecção facial. Imagens que de fato contenham uma face são introduzidas num modelo previamente treinado (veremos mais sobre no próximo capítulo) que classifica a imagem em uma das emoções previamente citadas.

Tendo em vista que o núcleo do problema a ser resolvido é, em suma, um problema de classificação, a arquitetura escolhida foi a Convolutional Neural Network (Rede Neural Convolutiva). Redes como essa foram primeiramente desenvolvidas com inspiração no funcionamento do córtex visual humano (Hubel e Wiesel, 1968) e são particularmente eficazes na extração e compreensão de padrões em dados bidimensionais, como imagens. O diferencial das CNNs está na incorporação de camadas convolucionais, que aplicam filtros para identificar características locais e hierárquicas em regiões específicas da imagem (Venkatesan e Li, 2018) mostra que o compartilhamento de pesos dentro dos neurônios de uma rede neural ajuda a

identificar padrões dentro de uma imagem, e que camadas convolucionais são um bom meio de atingir isso. Essa construção de rede permite que as CNNs automaticamente aprendam representações relevantes e discriminativas durante o treinamento, fator que é ideal para o objetivo deste trabalho.

3.4 CONCLUSÃO

Diante das bases teóricas apresentadas neste capítulo, torna-se evidente a relevância da detecção facial como um estágio inicial na inferência de emoções. Além disso, destaca-se a eficiência do classificador em cascata em termos de velocidade quando comparado a outras abordagens anteriormente propostas, tornando-o adequado para aplicações em tempo real.

A partir da detecção facial bem-sucedida, o foco se desloca para o treinamento de uma rede neural especializada na tarefa de classificação de emoções. Conforme discutido, as CNNs aparecem como uma escolha particularmente vantajosa, dada a sua alta capacidade de identificar padrões em imagens.

4 ESTADO DA ARTE

Neste capítulo serão apresentadas uma visão geral sobre trabalhos correlatos da área da computação afetiva e em sistemas tutores inteligentes e também as diferenças entre as técnicas utilizadas na literatura e neste estudo.

Gottardo (2018) propõe em sua tese um modelo híbrido de inferência de emoções num ambiente de STIs, sua arquitetura apresenta como característica principal a divisão da inferência em três componentes: cognitivo, físico e fusão. No componente cognitivo, acontecimentos e respostas do usuário no sistema tutor disparam eventos que são avaliados e utilizados para inferência de emoção. No componente físico, a expressão física (facial) do usuário é avaliada e utilizada para a inferência. No componente de fusão, ambas as inferências são utilizadas para formar a classificação final.

Tendo em foco o componente físico da tese de Gottardo (2018), que se assemelha com a proposta que será apresentada neste trabalho, notam-se semelhanças e diferenças. O conjunto de emoções utilizados em ambos os trabalhos é semelhante, Gottardo (2018) utiliza raiva, desgosto, medo, alegria, tristeza, surpresa, desprezo e neutro. No presente trabalho, utiliza-se raiva, desgosto, medo, alegria, tristeza, surpresa, contentamento e neutro, as emoções variantes são desprezo e contentamento. As diferenças começam a aparecer majoritariamente nos métodos de treinamento e classificação, em Gottardo (2018) as imagens passam por um algoritmo de extração de características e estas servem de base para o algoritmo de aprendizagem de máquina para treinamento e classificação conhecido como SVM (Support Vector Machines), não utilizando portanto nenhuma técnica de rede neurais ou camadas convolucionais como as deste trabalho.

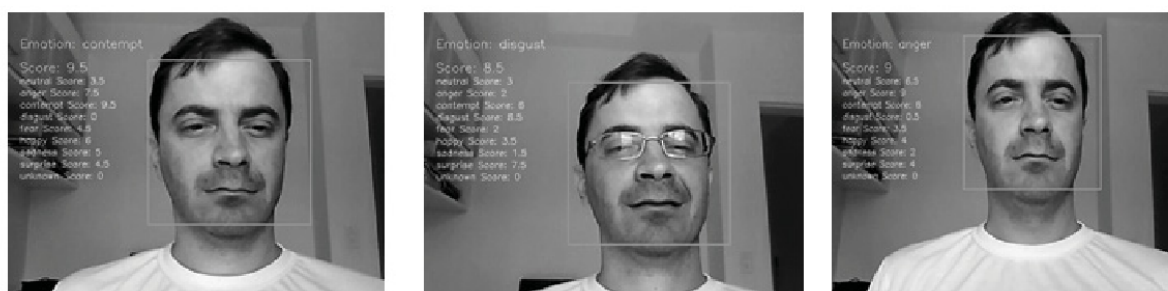


Figura 4.1: Expressões em classificação. (Gottardo, 2018)

Apesar do desenvolvimento deste classificador, Gottardo (2018) não utiliza-o. Os resultados encontrados não foram satisfatórios o suficiente para serem utilizados no restante do trabalho. Neste caso, optou-se pela utilização da ferramenta EmotionAPI que era disponibilizada sem custo pela Microsoft e pela Universidade de Oxford. Atualmente a ferramenta faz parte de um conjunto de software chamada Azure AI Vision e não é mais acessível gratuitamente, assim como seu código. Analisando a construção dos dois trabalhos, este diferencia-se pela maior quantidade de imagens utilizadas no treinamento e pela elaboração da rede neural convolucional

para a classificação das imagens. Além disso, enquanto o trabalho de Gottardo (2018) é voltado para a adaptação de sistemas tutores às emoções de estudantes, este é focado em prover um classificador para emoções de estudantes ou participantes de reuniões virtuais.

5 PROPOSTA

Este capítulo abordará a solução proposta para o cumprimento dos objetivos, tendo como base os conceitos apresentados nos dois últimos capítulos. Ele apresentará a base de dados utilizada, como a detecção facial foi realizada e a construção da CNN para atingir a inferência de emoções a partir da captura de imagens de participantes.

5.1 BASE DE DADOS

Para que o treinamento da CNN pudesse ter sido realizado foi necessário escolher uma base de dados abrangente de imagens de face representando emoções. A base escolhida foi a AffectNet (Mollahosseini et al., 2017) por três motivos principais:

- **Dimensão:** A base de dados AffectNet é composta por aproximadamente 420 mil imagens faciais, todas anotadas manualmente. Essa vasta quantidade de dados proporciona uma rica e abrangente fonte para o treinamento eficaz de redes neurais.
- **Método de Coleta:** A construção da AffectNet envolveu a coleta de imagens de faces provenientes da internet, utilizando três dos principais motores de busca contemporâneos. Foram buscadas 1250 palavras-chave relacionadas a emoções, abrangendo seis línguas diferentes. Esse método de coleta "selvagem" resultou em uma diversidade significativa e cobertura ampla, assegurando que as expressões nas imagens da base sejam predominantemente autênticas, ao invés de serem expressões “atuadas” ou artificiais.
- **Acessibilidade:** O acesso à base de dados AffectNet é facilmente solicitado através do site dos autores, proporcionando uma conveniência adicional para pesquisadores e desenvolvedores interessados em utilizar esse recurso para seus estudos e projetos.

Como mostrado na tabela 5.1, as oito primeiras emoções são aquelas citadas no capítulo 3 e apenas as imagens referentes à elas foram consideradas neste trabalho e no treinamento da CNN, totalizando assim cerca de 287 mil imagens.

5.1.1 Preparo das imagens

Apesar de a base de dados fornecer uma grande quantidade de imagens já anotadas, essas ainda não estão prontas para serem diretamente incorporadas à rede para fins de treinamento. Foi necessário desenvolver programas específicos para organizar as imagens em diretórios de acordo com as emoções, converter as imagens para escalas de cinza e distribuí-las entre diretórios de treinamento, validação e teste. Os códigos implementados para essa finalidade foram escritos em Python e Bash, e estão disponíveis no Apêndice A para referência.

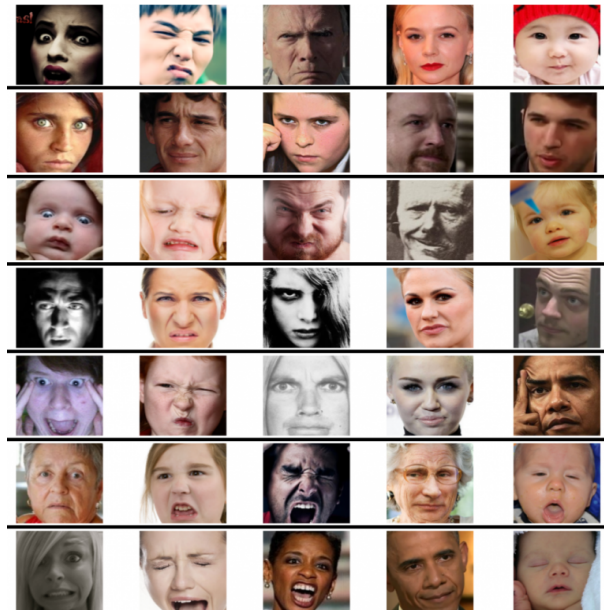


Figura 5.1: Amostra de imagens contidas na AffectNet. Fonte: Mollahosseini et al. (2017)

Tabela 5.1: Número de imagens por emoção

Emoção	Quantidade de Imagens
Neutro	75.374
Alegria	134.915
Tristeza	25.959
Surpresa	14.590
Medo	6.878
Desgosto	4.303
Raiva	25.382
Contentamento	4.250
Nenhum	33.588
Incerto	12.145
Não há face	82.915

Vale ressaltar que, para cada novo ciclo de treinamento, 20% das imagens associadas a cada emoção foram selecionadas aleatoriamente e alocadas para os conjuntos de validação e teste, sendo distribuídas de forma equitativa, com 10% em cada grupo.

Adicionalmente, no decorrer dos experimentos, foram desenvolvidos scripts adicionais para a aplicação de transformações nas imagens, tais como mudanças de brilho, espelhamento e ajuste de contraste. Essas transformações visaram a aplicação da técnica de aumento de dados em classes (emoções) com um número reduzido de imagens. Todos esses códigos suplementares estão documentados no Apêndice A para referência e replicação.

5.2 CONSTRUÇÃO E TREINAMENTO DA CNN

Iniciando com as especificações técnicas, a CNN foi desenvolvida utilizando a biblioteca Keras (Chollet et al., 2015) na linguagem de programação Python. A construção desta rede teve como base diversas implementações de CNNs dedicadas a análise de imagens e expressões emocionais disponíveis na internet. Além disso, foram realizadas adaptações e testes iterativos ao longo de vários treinamentos para otimizar o desempenho da rede. Em sua forma final, a arquitetura da CNN é composta pelas camadas detalhadas na figura 5.2.

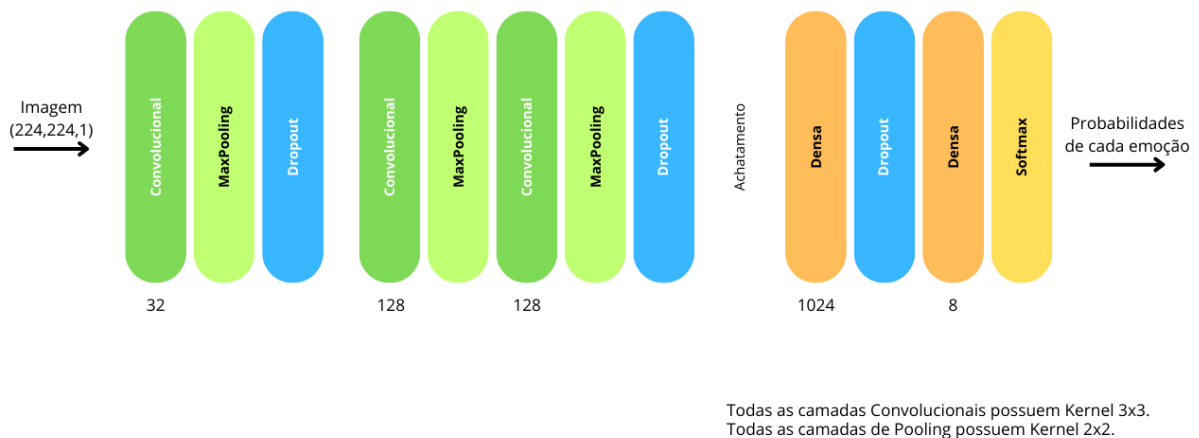


Figura 5.2: Estrutura da CNN construída

Observamos a presença de três camadas convolucionais seguidas por camadas de pooling, essenciais para a síntese e redução da informação, contribuindo, assim, para a diminuição do número de parâmetros. Após cada bloco de camadas convolucionais, são inseridas camadas de Dropout, incorporadas para mitigar o risco de "overfitting". Na sequência, encontram-se as camadas densas, completamente conectadas, que constituem a parte final da rede. Finalizando, uma camada de ativação Softmax é aplicada para gerar a saída da rede em forma de probabilidades associadas a cada emoção.

5.2.1 Treinamento da Rede

O treinamento da rede foi realizado utilizando como conjunto de imagens de entrada a base de dados AffectNet (Mollahosseini et al., 2017). A quantidade de imagens utilizadas foi a previamente mostrada na tabela 5.1, decrescida dos 20% separados para validação e teste.

O período de treino foi realizado em uma máquina pessoal, utilizando uma unidade de processamento gráfica, o que acelerou o processo como um todo. Mesmo assim, cada época,

de um total de 200, levou cerca de 30 minutos para ser concluída, tornando o gasto de tempo com o treinamento um pouco elevado. As especificações técnicas na máquina que realizou o treinamento estão disponíveis no Apêndice B.

Para amenizar este problema de tempo, testes e alterações na CNN foram realizados com número de épocas menores (por volta de 30) para validar as mudanças aplicadas e verificar sua eficiência. Essa rede “menos treinada” foi bastante útil para ver as mudanças em um menor período de tempo.

Outras tentativas de melhoria também foram testadas no treinamento, como o aumento de dados em classes (emoções) que possuíam menos imagens na base de dados, por exemplo contentamento e desgosto, mas não trouxeram resultados expressivos. Assim como a mudança de imagens em escala de cinza para imagens no padrão RGB (Red, Green, Blue) que também não trouxe nenhuma melhoria ao desempenho da rede.

5.3 CONCLUSÃO

Neste capítulo, apresenta-se a fundamentação da proposta de solução para os objetivos deste trabalho. Inicia-se pela explicação da escolha da base de dados, destacando as motivações que nortearam sua escolha, e avança para a definição das camadas da Rede Neural Convolucional utilizada. A implementação dos conceitos e do projeto proposto resultou em uma diversidade de resultados, os quais serão mais precisamente analisados no próximo capítulo.

6 RESULTADOS

Os resultados discutidos nesta seção são referentes aos testes realizados com a CNN construída. Para esses testes, como informado anteriormente, 10% das imagens da base de dados foram separadas e não utilizadas no treinamento. A obtenção das métricas foi realizada comparando as emoções inferidas pela rede para essas imagens com suas emoções anotadas originalmente. A imagem 6.1 mostra alguns exemplos de resultados inferidos pela rede.

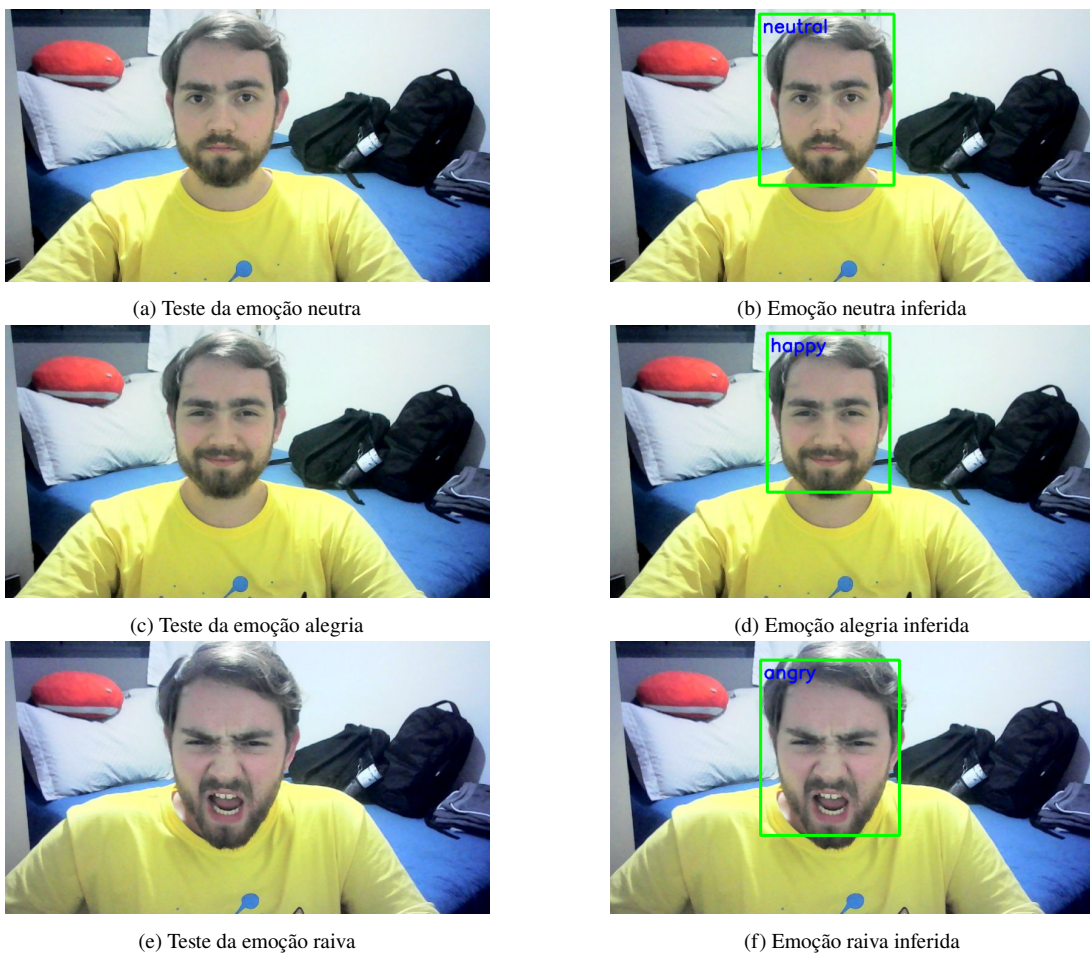


Figura 6.1: Exemplo de imagens classificadas pela rede.

6.1 MATRIZ DE CONFUSÃO

As matrizes de confusão são uma ferramenta fundamental na avaliação de desempenho de modelos de classificação. Elas fornecem uma visão detalhada e estruturada da qualidade das previsões feitas por um modelo, comparando-as com as classes reais dos dados. Cada elemento na matriz representa a contagem de instâncias correspondentes à combinação específica de previsões e rótulos reais. Sua análise permite uma compreensão mais aprofundada do

desempenho do modelo, identificando áreas de força e fraqueza, e é essencial para ajustes e otimizações necessárias na busca por um modelo mais preciso e confiável.

A figura 6.2 mostra a matriz de confusão para a CNN de inferência de emoções. Ao todo foram utilizadas mais de 30 mil imagens, portanto para facilitar a visualização os números foram convertidos em porcentagem arredondadas.

Há muitos pontos interessantes que a imagem apresenta, mas pode-se começar notando a performance de duas emoções: alegria e neutra. As duas mostram mais de 85% de “positivos verdadeiros”, ou seja, imagens que pertenciam a essas classes e foram classificadas como tal após sua passagem pelo modelo. Sendo assim, o modelo se mostra eficiente na classificação dessas emoções. É importante notar também que estas eram as emoções que continham imagens na base de dados, fato que pode ter influenciado a performance.

O segundo fator de destaque são as expressões de desgosto e contentamento. As duas performaram extremamente mal no modelo proposto. Suspeita-se que a causa seja, principalmente, a pequena quantidade de imagens dessas classes na base de dados. Desgosto, por exemplo, tem cerca de 10 vezes menos imagens que a emoção raiva, e essa diferença fica ainda maior quando a comparação é feita com alegria e neutralidade. Foram feitas tentativas para aumentar a quantidade de dados para estas classes, como evidenciadas pelos códigos do Apêndice A, entretanto, essas adições não tiveram o efeito desejado, não aumentando a performance dessas emoções.

Por último, nota-se o desempenho mediano de emoções como raiva, medo, tristeza e surpresa. Este desempenho pode estar relacionado também com o número de imagens dessas emoções, que não estão nem entre os mais baixos, nem entre os mais altos. Além disso, percebe-se que na matriz de confusão, os maiores erros avaliando as emoções citadas acontecem inferindo alegria e neutralidade. Uma possibilidade é que estas duas emoções estejam exercendo uma certa dominância na rede sobre as demais.

6.2 OUTRAS MÉTRICAS

6.2.1 Acurácia

A métrica acurácia mede o quanto a rede está inferindo corretamente as emoções para as imagens de entrada. Neste trabalho ela foi obtida calculando a soma total de quantas imagens foram preditas corretamente em relação ao total de imagens testadas. O valor obtido foi de aproximadamente 72,02%.

6.2.2 Precisão

A precisão no contexto de modelos de previsão tem como objetivo mostrar em quantas das vezes em que ele inferiu determinada emoção ele estava certo. Por exemplo, de 200 vezes em que a classe alegria foi predita, em quantas delas a emoção realmente era alegria? Essa métrica, mostrada em porcentagem para cada uma das classes, está descrita na tabela 6.1

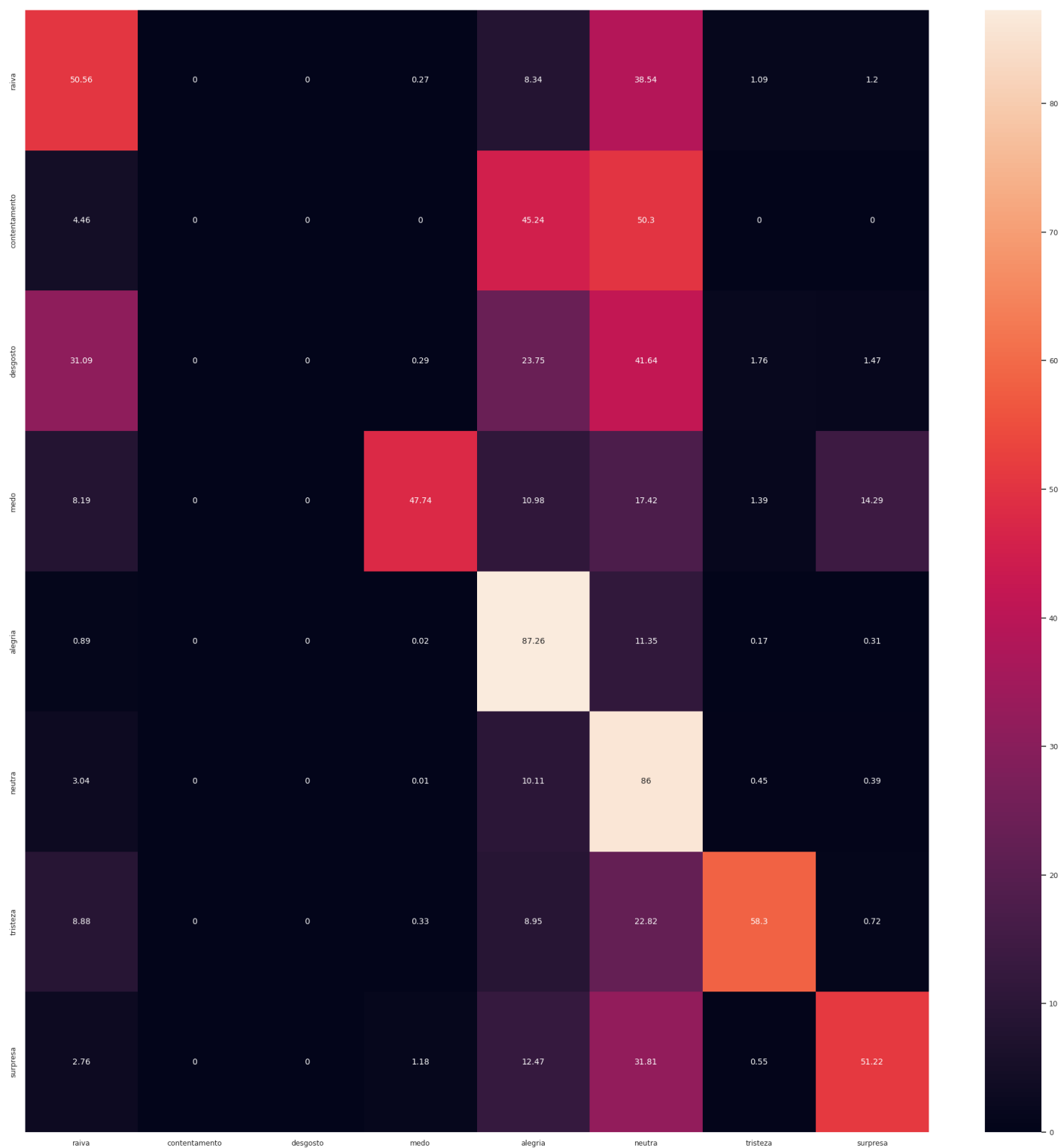


Figura 6.2: Matriz de Confusão para a CNN proposta

Tabela 6.1: Precisão por classe

Emoção	Índices de precisão (%)
Raiva	71
Contentamento	0
Desgosto	0
Medo	86
Alegra	85
Neutra	54
Tristeza	96
Surpresa	73

7 PROTOTIPO DE APLICAÇÃO

Com o objetivo de mostrar como a classificação de emoções pode ser utilizada num contexto real, um protótipo de aplicação web foi desenvolvido para simular o funcionamento real do classificador numa sala de reuniões virtuais. Não é o foco construir uma aplicação pronta, robusta e pronta para ser usada, mas sim uma que permita a visualização do conceito. Este capítulo detalha o seu funcionamento e também aponta como a rede neural de inferência foi integrada ao sistema.

A aplicação web foi desenvolvida majoritariamente em Django, um framework da linguagem de programação Python, voltado especificamente para o desenvolvimento web. A escolha deste framework foi realizada principalmente pela facilidade que seria integrar o modelo de CNN, escrito em Python com uma aplicação web também desenvolvida em Python. Essa integração se mostrou eficaz e realmente fácil de ser implementada, como veremos ao longo do capítulo.

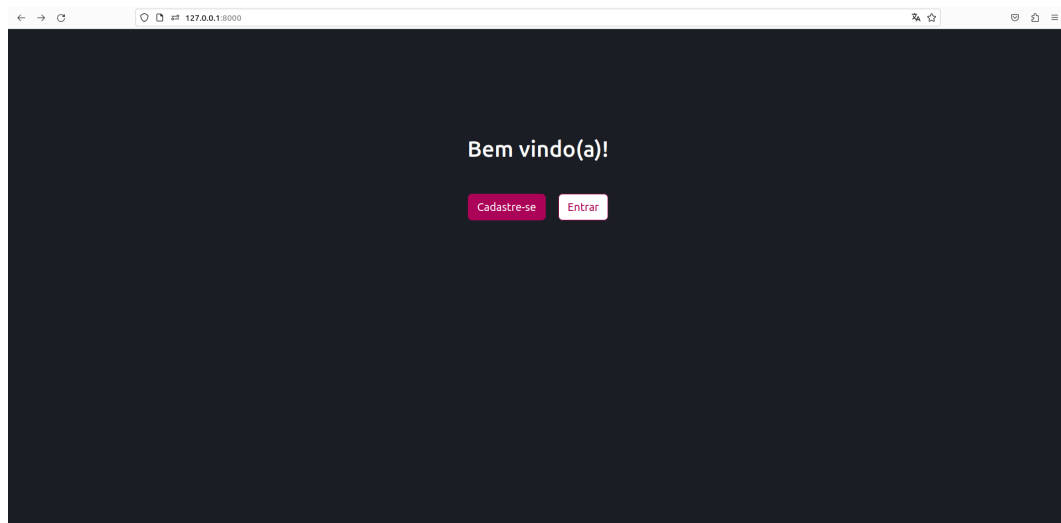


Figura 7.1: Tela inicial da aplicação

7.1 CONCEITOS E ESTRUTURAS

Apesar de ser um protótipo, a aplicação realmente cria um ambiente de salas de reunião virtuais. As imagens dos participantes são realmente capturadas e enviadas para o servidor da aplicação. Tendo isso em vista, o site conta com alguns módulos principais como autenticação de usuários, criação de salas e captura e classificação de imagens dos participantes.

Para tornar os módulos concretos, além do framework Django citado anteriormente, a aplicação também utiliza um banco de dados relacional. Optou-se pelo gerenciador de banco de dados SQLite por ainda se tratar de um protótipo em construção, ainda não sendo necessário

maior robustez, entretanto, a adaptação para outro SGBD (Sistema Gerenciador de Banco de Dados) seria fácil e prática.

Em determinados pontos da aplicação, como na contagem de participantes em uma sala, foi necessária a comunicação entre clientes e não apenas entre cliente e servidor. Para isso, foram utilizadas as bibliotecas Channels e Daphne do próprio Django para a criação de um websocket que transmita estas mensagens entre clientes.

Mas próximas seções, serão detalhadas as funcionalidades da aplicação e serão apresentados também os visuais empregados neste protótipo.

7.2 AUTENTICAÇÃO E USUÁRIOS

Para a utilização do sistema de forma mais personalizada e para que os usuários possam criar salas próprias, foi desenvolvido um sistema de autenticação básico que fornece cadastro de usuários, além das funcionalidades de Entrar e Sair.

O cadastro de usuário solicita apenas informações básicas como nome de usuário, email e senha, necessários para posterior acesso. Uma vez autenticado, um usuário tem acesso à criação de salas e também pode ver as salas que já criou (figura 7.4). A figura 7.2 mostra a tela de Entrar do sistema, já a figura 7.3 mostra a tela de cadastro de usuários.

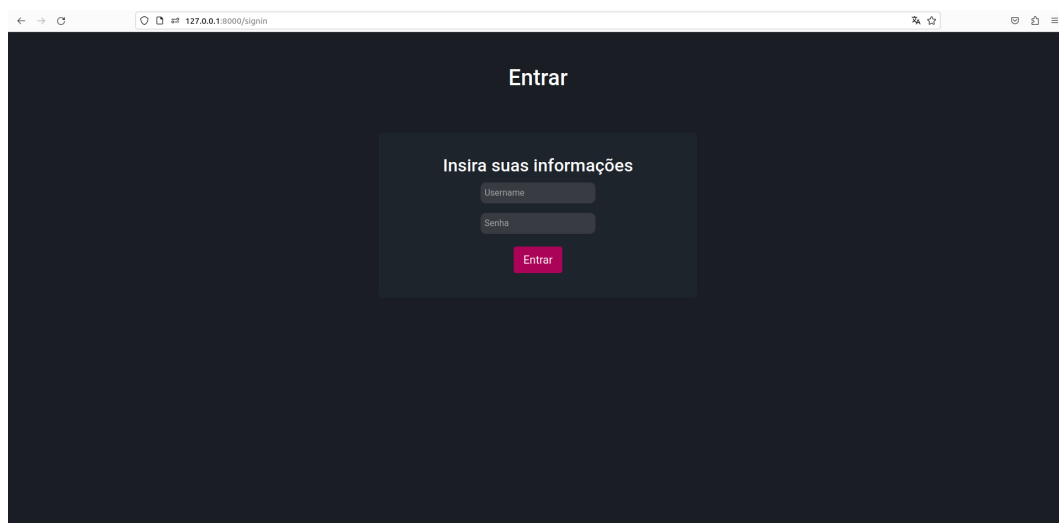


Figura 7.2: Tela para entrada na aplicação

7.3 SALAS

Uma estrutura de salas é disponibilizada aos usuários da aplicação. Usuários autenticados têm a possibilidade de criar salas (7.5) onde outros usuários entrarão para realizar uma reunião, por exemplo. Uma sala é composta por três atributos: um nome descritivo, um código que será utilizado na sua url e uma senha de acesso. Quando o criador (dono) de uma sala a acessa, ela é aberta e outros participantes podem entrar utilizando o mesmo endereço do criador e inserindo a

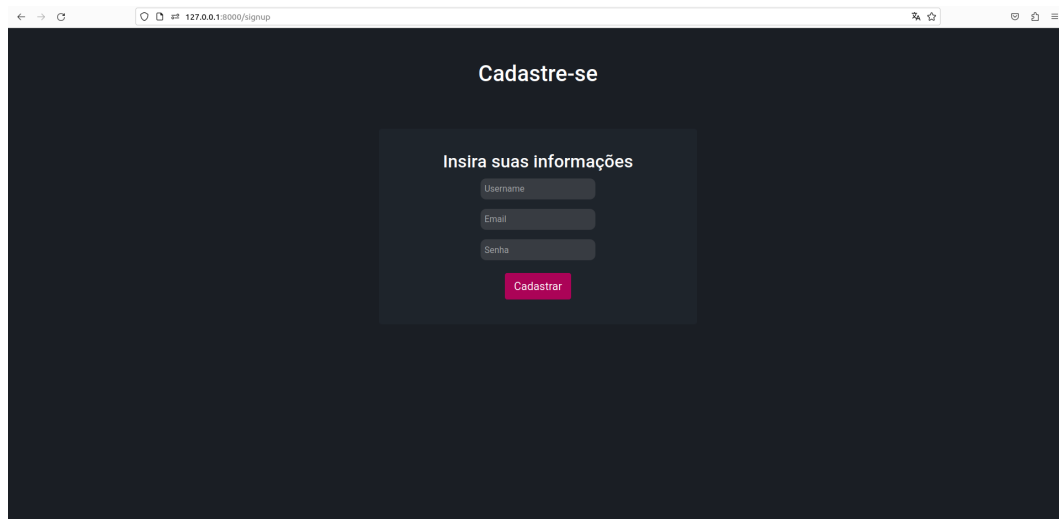


Figura 7.3: Tela de cadastro.

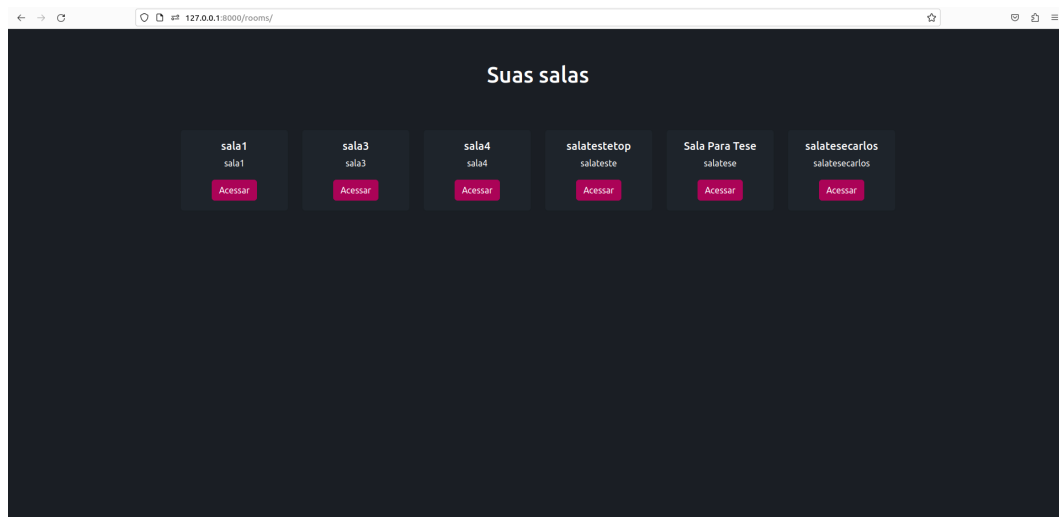


Figura 7.4: Tela de visualização de salas.

senha (figura 7.6). O dono da sala tem disponível a informação de quantos participantes estão conectados à sala (figura 7.7) e a opção de iniciar a reunião. Quando a reunião é iniciada, as imagens dos participantes começam a ser capturadas e classificadas pelo sistema.

Por outro lado, quando um participante regular acessa à reunião, ele pode ver uma pré-visualização da imagem que está sendo capturada dele, como na figura 7.8.

7.4 MOMENTOS

A partir do momento em que uma reunião é iniciada pelo dono da sala, momentos passam a ser criados e registrados num intervalo de cinco segundos. Momentos nada mais são do que entradas num banco de dados com as seguintes informações:

- A sala na qual aquele momento foi capturado;
- Identificador do momento na sala (unidades de cinco segundos após o início);



Figura 7.5: Tela de cadastro de salas.

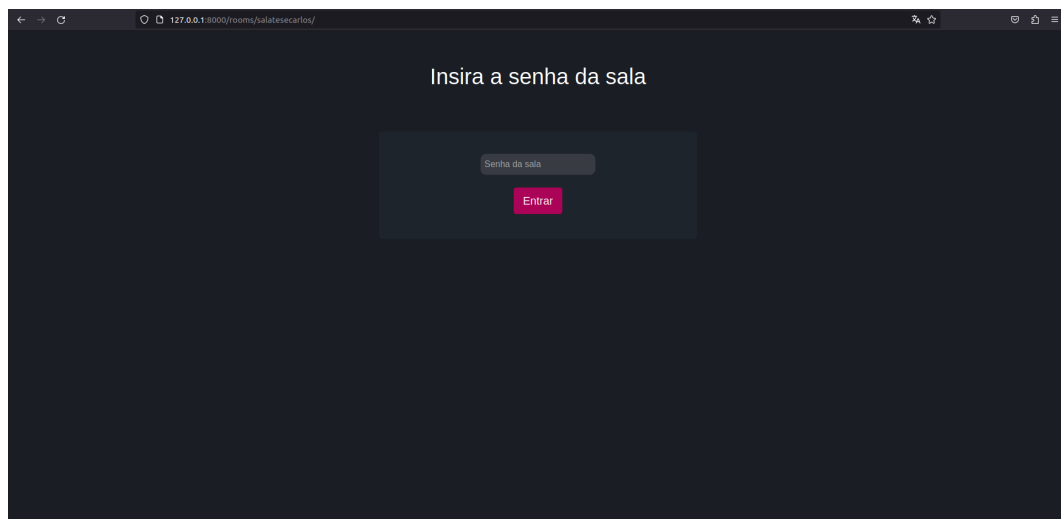
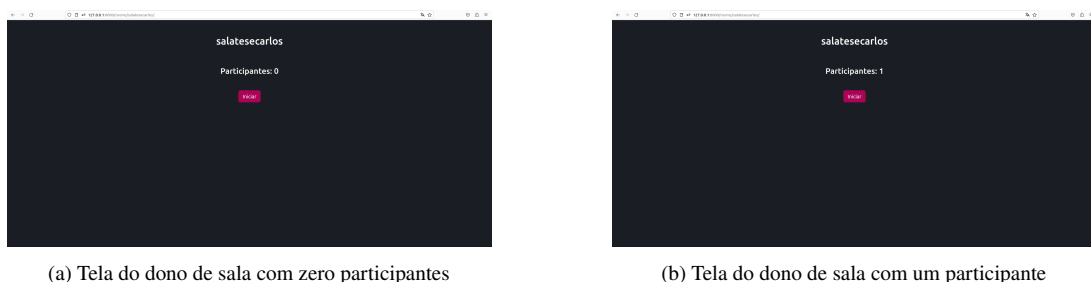


Figura 7.6: Tela para inserção de senha da sala.



(a) Tela do dono de sala com zero participantes

(b) Tela do dono de sala com um participante

Figura 7.7: Telas de visão do dono da sala.

- A identificação do participante naquela sala: apenas um número para indexação, que não identifica o usuário em si.
- A emoção expressada pela imagem capturada do participante naquele momento.



Figura 7.8: Tela de visão do participante.

7.5 CAPTURA E CLASSIFICAÇÃO DE IMAGENS

Assim que a reunião é iniciada, as imagens dos participantes começam a ser capturadas a cada cinco segundos. Estas imagens são codificadas em base64 para facilitar a transferência e então são enviadas junto com outros dados numa requisição ao servidor. No backend do servidor, a imagem é decodificada e é utilizada como entrada no modelo de CNN discutido anteriormente e quando a expressão é inferida é criado um objeto de momento que é salvo no banco de dados.

Esta forma de lidar com as imagens garante dois pontos principais, sendo o primeiro o anonimato, pois não salva as imagens capturadas dos participantes, apenas as processa, e além disso, a expressão obtida da inferência não é atribuída a um usuário em si, apenas a seu índice na sala. O segundo ponto é a sincronia, as imagens tem tempo suficiente para serem enviadas, processadas e geram um objeto de momento. Este objeto é que será posteriormente analisado e não a imagem ou expressão em si.

Do lado do dono da sala, o processo ocorre em complemento ao dos participantes. Instantes após o início da sala e, regularmente a cada cinco segundos, uma requisição é enviada ao servidor. Esta requisição faz uma varredura nos registros de momentos, buscando todos aqueles referentes à sala em questão e ao tempo (unidades de cinco segundos após o início) atual. Essa requisição tem como retorno um número de momentos correspondente ao número de participantes da sala, afinal cada um teve um momento registrado após a última captura. Desta lista de momentos, é selecionada a expressão que mais apareceu e ela é mostrada na tela do dono da sala.

8 CONCLUSÃO

Este capítulo conclui o que foi estudado neste trabalho e resume também tudo aquilo que foi alcançado. Além disso, busca mostrar como o objetivo deste estudo foi cumprido e propõe possíveis continuações para a pesquisa.

8.1 O ESTUDO

Neste trabalho foram introduzidos, primeiramente, conceitos de emoções na literatura. Foi mostrado que as emoções podem ser estudadas de duas maneiras, uma categórica, utilizada aqui, e uma dimensional, que foi descrita, mas não abordada. Nesta abordagem categórica utilizamos um conjunto discreto de oito emoções, sendo elas raiva, alegria, neutra, desgosto, contentamento, surpresa, tristeza e medo.

Logo depois, foram construídas as bases teóricas de detecção facial e redes neurais, mais especificamente as convolucionais e como elas são particularmente eficientes em reconhecimento de padrões em imagens, fator útil para este trabalho.

Com os conceitos de emoções e redes neurais introduzidos, pode-se mostrar a proposta que visa cumprir o objetivo geral deste trabalho: uma ferramenta de inferência de emoções que usa reconhecimento facial e uma CNN. Esta ferramenta foi de fato construída e o objetivo deste estudo foi alcançado. Os resultados foram apresentados em conjunto com uma série de métricas que os validam. Viu-se que emoções como alegria e neutra, as mais populosas da base de imagens, tiveram um índice de acerto bastante alto. Outras emoções tiveram desempenhos bons e medianos e houveram ainda as emoções que performaram de forma ruim e necessitam de atenção.

8.2 POSSÍVEIS APLICAÇÕES

O problema norteador deste trabalho foi a falta de feedback que apresentadores ou tutores notam em reuniões online. Tendo isso em vista, a ferramenta construída aqui pode ser utilizada de duas formas: a primeira é aquela apresentada no capítulo 6, o desenvolvimento de uma aplicação web que integra o classificador de emoções e fornece informações constantes ao apresentador da videoconferência. A segunda aplicação é pura utilização e integração da ferramenta de inferência em aplicações de reuniões já existentes. Tendo acesso a base de dados e utilizando as camadas da rede neural proposta, os resultados devem ser similares aos apresentados neste estudo.

Por outro lado, uma aplicação fora do problema norteador, seria a inclusão do classificador em sistemas tutores inteligentes, ou a construção de um integrado a ele, com o objetivo de adaptar o sistema às emoções que o estudante apresentou. Além disso, a inferência de emoções ainda

pode ser utilizada em pesquisas mercadológicas, apresentação de produtos, receptividade de mídias e em qualquer outro contexto onde as expressões e impressões de uma pessoa importem.

8.3 TRABALHOS FUTUROS

O que foi aqui apresentado apenas arranha a superfície daquilo que a inferência de emoções tem a dar. A primeira possibilidade seria trabalhar com emoções em escala dimensional da mesma forma que a categórica trabalhada aqui, é possível também treinar uma rede neural com dados de valência e ativação, que inclusive a AffectNet também disponibiliza.

Um ponto de dificuldade neste estudo foi a inferência correta de emoções como contentamento e desgosto, cuja base de imagens é extremamente pobre quando comparada com outras emoções. Uma possível continuação de trabalho seria focar em melhorar o resultado destas emoções com diferentes técnicas de visão computacional, ou ainda modificar a CNN apresentada para lidar melhor com elas.

Tratando-se do protótipo proposto, ainda há muito o que ser feito nele. Atualmente a emoção apresentada ao dono da sala é aquela que mais aparece na lista de momentos de mesmo tempo, seria interessante explorar outras técnicas que visem expor uma “emoção” média, ou ainda um “engajamento médio” por parte dos participantes. Quanto aos aspectos técnicos, seria interessante a inclusão de funcionalidades analíticas para os dados de momentos, a implementação de gerência de cadastros de sala e de usuários, melhora em questões de segurança e uma possível inserção de passagem de voz.

Por fim, trabalhos futuros também podem incorporar a análise de outros aspectos, como medições cognitivas, voz e postura na inferência da emoção de um participante.

REFERÊNCIAS

- Bergmann, R., Rintel, S., Baym, N., Sarkar, A., Borowiec, D., Wong, P. e Sellen, A. (2022). Meeting (the) pandemic: Videoconferencing fatigue and evolving tensions of sociality in enterprise video meetings during covid-19. *Computer Supported Cooperative Work (CSCW) volume*, 32(1):347–383.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.
- CNN (2020). Zoom’s massive ‘overnight success’ actually took nine years. <https://edition.cnn.com/2020/03/27/tech/zoom-app-coronavirus/index.html>. Acessado em 30/11/2023.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200.
- Google LLC (2015). Google photos. <https://www.google.com/intl/pt-BR/photos/about/>. Acessado em 02/12/2023.
- Gottardo, E. (2018). *INFERÊNCIA DE ESTADOS AFETIVOS EM AMBIENTES EDUCACIONAIS: PROPOSTA DE UM MODELO HÍBRIDO BASEADO EM INFORMAÇÕES COGNITIVAS E FÍSICAS*. Tese de doutorado, Universidade Federal do Paraná, Curitiba - Paraná. 127 pgs.
- Gunes, H., Schuller, B., Pantic, M. e Cowie, R. (2005). Emotion representation, analysis and synthesis in continuous space: A survey. Em *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*, páginas 827–834, EUA.
- Hubel, D. H. e Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol*, 195(1):215–243.
- Izard, C. E. (1990). Facial expressions and the regulation of emotions. *Journal of Personality and Social Psychology*, 58(3):487–598.
- Koziarski, M. e Cyganek, B. (2019). Impact of low resolution on image recognition with deep neural networks: An experimental study. *International Journal of Applied Mathematics and Computer Science*, 28(4):735–744.
- McCarthy, N. (2020). Covid-19’s staggering impact on global education. <https://www.weforum.org/agenda/2020/03/infographic-covid19-coronavirus-impact-global-education-health-schools/>. Acessado em 30/11/2023.

- Mollahosseini, A., Hasani, B. e Mahoor, M. H. (2017). Affectnet: A new database for facial expression, valence, and arousal computation in the wild. *IEEE Transactions on Affective Computing*.
- Nations, U. (2020). Covid-19 to slash global economic output by \$8.5 trillion over next two years. <https://www.un.org/en/desa/covid-19-slash-global-economic-output-85-trillion-over-next-two-years>. Acessado em 30/11/2023.
- Orhan, G. e Beyhan, O. (2020). Teachers' perceptions and teaching experiences on distance education through synchronous video conferencing during covid-19 pandemic. *Social Sciences and Education Research Review*, 7:8–44.
- Picard, R. W. (2000). *Affective Computing*. The MIT Press.
- Plutchik, R. (1980). *A general psychoevolutionary theory of emotion*, páginas 3–33. Academic Press.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.
- Schaevitz, S. (2020). Three months, 30x demand: How we scaled google meet during covid-19. <https://workspace.google.com/blog/productivity-collaboration/keeping-google-meet-ahead-of-usage-demand-during-covid-19>. Acessado em 30/11/2023.
- Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social Science Information*, 44(4):695–729.
- Turati, C., Cassia, V. M., Simion, F. e Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. *Child development*, 77(2):297–311.
- Venkatesan, R. e Li, B. (2018). *Convolutional Neural Networks in Visual Computing: A Concise Guide*. CRC Press.
- Viola, P. e Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154.

APÊNDICE A – CÓDIGOS UTILIZADOS NO PRÉ-PROCESSAMENTO DE IMAGENS

Os códigos que serão apresentados neste apêndice detalham algumas operações que foram realizadas nas imagens da base dados AffecNet antes de seu uso na CNN construída. É importante detalhar que todos os programas consideram uma estrutura de diretórios onde as imagens de cada emoção estão separadas em diretórios específicos.

A.1 CONVERSÃO PARA ESCALA DE CINZA

```

1 import cv2
2 from os import scandir
3
4 path = "" #caminho para o diretório que contém as imagens.
5 newpath = "" #caminho para o diretório que conterá as novas imagens em cinza.
6
7 subdir = "surprise/" #subdiretório de emoção
8
9 for file in scandir(path+subdir):
10     if file.is_file():
11         #print(path+subdir+file.name)
12         img = cv2.imread(path+subdir+file.name)
13         gray_img = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
14         cv2.imwrite(newpath+subdir+file.name, gray_img)

```

A.2 ESPELHAMENTO DE IMAGENS

```

1 import cv2
2 from os import scandir
3
4 #diretório contendo os subdiretórios de emoções
5 path = ""
6 # subdiretórios de emoções: na forma "emocaol/", "emocao2/"
7 subdirs = []
8
9 for subdir in subdirs:
10     cpath = path+subdir
11     print("Flipping in " + cpath)
12     for file in scandir(cpath):
13         if file.is_file():
14             #print("lendo " + cpath+file.name)
15             img = cv2.imread(cpath+file.name)

```

```

16         flip_image = cv2.flip(img, 1)
17         newfile = cpath + "flip/flipped_" + file.name
18         cv2.imwrite(newfile, flip_image)

```

A.3 AJUSTE DE BRILHO DE IMAGENS

```

1 import cv2
2 from os import scandir
3
4 #diretório contendo os subdiretórios de emoções
5 path = ""
6 # subdiretórios de emoções: na forma "emocao1/", "emocao2/"
7 subdirs = []
8 for subdir in subdirs:
9     cpath = path+subdir
10    print(cpath)
11    for file in scandir(cpath):
12        if file.is_file():
13            #print("lendo " + path+file.name)
14            img = cv2.imread(cpath+file.name)
15            #lowbright_image = cv2.convertScaleAbs(img, beta=-50)
16            bright_image = cv2.convertScaleAbs(img, beta=50)
17            #newlow = cpath + "low/lowbright_" + file.name
18            newb = cpath + "bright/bright_" + file.name
19            #cv2.imwrite(newlow, lowbright_image)
20            cv2.imwrite(newb, bright_image)

```

A.4 AJUSTE DE CONTRASTE DE IMAGENS

```

1 import cv2
2 from os import scandir
3
4 #diretório contendo os subdiretórios de emoções
5 path = ""
6 # subdiretórios de emoções: na forma "emocao1/", "emocao2/"
7 subdirs = []
8 for subdir in subdirs:
9     cpath = path+subdir
10    for file in scandir(cpath):
11        if file.is_file():
12            #print("lendo " + path+file.name)
13            img = cv2.imread(cpath+file.name)
14            lowcon_image = cv2.convertScaleAbs(img, alpha=0.5)
15            con_image = cv2.convertScaleAbs(img, alpha=1.5)
16            newlow = cpath + "low/lowcon_" + file.name

```

```
17         newc = cpath + "con/con_" + file.name
18
19         #print(newfile)
20         cv2.imwrite(newlow, lowcon_image)
21         cv2.imwrite(newc, con_image)
```

A.5 ESCOLHA ALEATÓRIA DE IMAGENS

```
1 EXPS="anger sad surprise happy neutral disgust fear contempt"
2 IDIR="/home/carlos/UFPR/tcc/tcc-carloscichon/affectnet_data_gray"
3 ODIR="/home/carlos/UFPR/tcc/tcc-carloscichon/affectnet_data_gray_ult_test"
4 for EXP in $EXPS; do
5     total=$(ls $IDIR/$EXP | wc -l)
6     #tenpercent=$(( total*10/100))
7     echo $EXP $tenpercent
8     rm $ODIR/$EXP/*
9     for file in $(ls $IDIR/$EXP | shuf -n 100); do
10         cp $IDIR/$EXP/$file $ODIR/$EXP/$file
11     done
12 done
```

APÊNDICE B – ESPECIFICAÇÕES TÉCNICAS DA MÁQUINA QUE TREINOU A CNN

- **Processador:** AMD A8-7600 Radeon R7, 10 Compute Cores 4C+6G.
- **Placa gráfica:** GeForce GTX 1060 6GB
- **RAM:** 12GB (DDR3)
- **Sistema Operacional:** Ubuntu 22.04
- **Tensorflow:** 2.11.0