
UNIVERSIDADE FEDERAL DE OURO PRETO
DEPARTAMENTO DE CIÊNCIA DE COMPUTAÇÃO
PROCESSAMENTO DIGITAL DE IMAGENS

Projeto de Pesquisa

DETECÇÃO DE EMOÇÕES HUMANAS EM CÂMERAS DE
VIGILÂNCIA

Alunos:

Eduardo Lopes Amaro,

Bárbara Neves,

Marina Lima,

Lucas Sena,

Resumo

Este trabalho tem o intuito de explorar as diversas formas de detectar e reconhecer emoções humanas por meio de câmeras de vigilância. Utilizamos uma abordagem de acordo com a revisão bibliográfica, para criar e adaptar uma Rede Neural Convolucional (CNN) com o objetivo de processar e analisar as imagens. A CNN possui uma alta taxa de sucesso em reconhecer as emoções, com isso, decidimos explorar essa técnica para treinar nossos modelos e obter a rede que nos forneça uma melhor acurácia na classificação das emoções e um menor erro. Após treinar nosso modelo, carregamos um vídeo como entrada e realizamos o reconhecimento das faces e o pré-processamento da imagem para obter a classificação das emoções.

1 Introdução

As expressões faciais são formas de comunicação não verbal amplamente utilizadas no cotidiano dos seres humanos. Elas externam, consciente ou inconscientemente, nossas manifestações em relação aos estímulos sociais a que somos submetidos. Seres humanos e a maioria dos mamíferos utilizam expressões faciais para demonstrar estados emocionais. Com a popularização dos sistemas de detecção de faces, os sistemas de detecção de expressões faciais vêm se popularizando. As possibilidades de utilização destes sistemas e seus impactos na atual sociedade são inúmeros. [Ekman & Friesen \(2003\)](#) [Libralon \(2014\)](#)

Com base na literatura, podemos classificar as expressões faciais de uma pessoa em seis diferentes emoções universais, felicidade, raiva, nojo, medo, tristeza e a emoção neutra. [Khan \(2018\)](#) Durante o processo de identificação e classificação das emoções, são comumente utilizadas técnicas de processamento de imagens atrelada a inteligência artificial para auxiliar nas principais etapas de pré-processamento, extração de recursos e classificação. [Revina & Emmanuel \(2018\)](#)

Hoje vemos uma crescente demanda por sistemas que utilizam a detecção de emoções faciais em câmeras de vigilância e, consequentemente pesquisas utilizando tais técnicas que nos proporciona uma maior acurácia na classificação e um menor custo computacional.

Em detrimento dessas informações, o intuito do nosso projeto é utilizar as diferentes abordagens já realizadas para detecção de emoções faciais em câmeras de vigilância e as respectivas técnicas de processamento de imagens e inteligência artificial utilizadas para fornecer uma maior acurácia e um menor custo computacional.

2 Estado da Arte

Esta seção descreve, com o apoio das principais referências bibliográficas, aspectos e questões relevantes ao tema proposto. A base para a realização do projeto inclui temas relacionados à extração de características multimodais e ao reconhecimento de atividades humanas de alto nível.

2.1 Reconhecimento de emoções em vídeo utilizando redes neurais artificiais

O artigo [Behling et al. \(2019\)](#) teve como objetivo analisar o desempenho de redes de aprendizado supervisionado para a classificação de imagens individuais ou frames de vídeo de entrada em diferentes classes de emoções previamente treinadas.

Buscando fornecer embasamento e indícios da necessidade da ferramenta desenvolvida, foram utilizadas diversas técnicas com base na literatura para implementação do reconhecimento de emoções. Para realizar a detecção das faces foi utilizado o detector de faces do OpenCV e para extrair as características foi utilizado o extrator de características do OpenFace, trazendo bons resultados.

Na execução dos testes foram utilizados dois tipos de aprendizados de máquinas supervisionados, CNN (Rede Neural Convolucional) e SVM (Máquinas de Vetores de Suporte), e apresentado os pontos positivos e negativos de cada um. As redes CNNs obtiveram melhores métricas de assertividade na classificação de emoções, porém podem levar muito mais tempo para serem treinadas. Já as redes SVMs, se mostraram muito mais eficientes para o treinamento, no entanto, conforme o tamanho do conjunto de dados fique maior a rede tem dificuldades para conseguir gerar hiperplanos que gerem menos erros.

2.2 Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks

O artigo [Khan \(2018\)](#) tem como objetivo classificar a expressão facial de uma pessoa. A classificação é dada por seis emoções universais, felicidade, raiva, nojo, surpresa, medo, tristeza, mais a neutra.

Podemos dividir a implementação em 4 passos: localização facial, demarcação facial e detecção, extração de características e classificação de emoção.

O banco de dados selecionado foi o Karolinska Directed Emotional Faces (KDEF) que consiste em um conjunto de 4900 imagens de 70 pessoas - 35 homens e 35 mulheres, cada um mostrando as 6 expressões básicas e 1 expressão neutra e cada expressão foi fotografada de diferentes ângulos.

Para a extração de características foi utilizado os dados do passo anterior detecção de olhos (12 pontos), detecção de nariz (9 pontos), detecção de sobrancelhas (10 pontos) e por fim a detecção de lábios (18 pontos).

Para determinar qual a emoção que está sendo expressada foi treinada uma rede neural com *back-propagation* tendo resultado de 95% para feliz e de 84-86% para raiva, tristeza, nojo e medo.

2.3 A Survey on Human Face Expression Recognition Techniques

O artigo [Revina & Emmanuel \(2018\)](#) concentra-se em várias técnicas de FER que possuem três etapas principais : pré-processamento, extração de recursos e classificação. Além disso, o artigo discute sobre as vantagens de diferentes técnicas de FER e a análise de desempenho de tais técnicas. É importante ressaltar que nesse artigo apenas técnicas de FER baseadas em imagens foram escolhidas para revisão de literatura.

Sobre o pré-processamento de imagens alguns diferentes tipos de processos são mencionados como a amostragem descendente de Bessel para a redução da imagem do rosto, protegendo os aspectos e também a percepção da imagem original. O filtro Gaussiano é usado para redimensionar as imagens de entrada, fornecendo suavidade às imagens. Outra técnica também presente é a normalização que pode ser projetada para redução da iluminação e variações da face das imagens. Já a equalização de histograma é uma importante técnica para FER porque melhora a distinção da imagem através do aumento do contraste das imagens do rosto.

O processo de extração de recursos é o próximo estágio do sistema FER. A extração de recursos é encontrar e descrever recursos positivos de interesse em uma imagem para processamento posterior. Em processamento de imagem a extração de recursos de visão computacional é um estágio significativo, enquanto ele identifica a mudança da representação gráfica para a representação implícita de dados. Então esta representação de dados pode ser usada como uma entrada para a classificação. O artigo também destaca os cinco tipos de métodos de extração , tais como método baseado em recurso de textura, método baseado em borda, global e método baseado em características locais, método baseado em características geométricas e método baseado em patch. Ainda de acordo com o artigo , os descritores baseados em recursos de textura são mais úteis do que os outros já que extraem os recursos de textura relacionados à aparência, fornecendo recursos importantes para a FER.

E por fim, tem-se o último estágio do reconhecimento de expressão facial em que o classificador categoriza a expressão como sorriso, tristeza, surpresa, raiva, medo, nojo e neutro. Dentre as

várias técnicas de classificação citadas para FER, o classificador SVM oferece melhor precisão de reconhecimento e melhor classificação.

2.4 Reconhecimento de expressões faciais através de análise facial computadorizada utilizando o kinect v2

No artigo [Aires et al. \(n.d.\)](#), o autor levanta que traduzir informações através da comunicação não se restringe ao intercâmbio de palavras, tal complexidade se dá baseada em fatores como meio cultural, vivência, formação educacional e emoções. Tendo em vista o obstáculo, a Computação Afetiva abarca aplicações com enfoque em identificar padrões que possam vir a se tornar meios de o sistema se expressar e compreender expressões em uma junção de mente e corpo. Uma vez que o panorama está exposto, foi desenvolvido um Sensor Kinect que é capaz de detectar expressões faciais do rosto humano, tal conjunto emocional e gestual permite o reconhecimento não somente do gestual, mas também a entonação do que se fala.

Para o ser humano é inata a percepção do seu próprio estado emocional, sendo assim capaz de compor padrões de comportamento distintos para cada ocasião, já nos sistemas computacionais as emoções são compreendidas através de mecanismos de controle que se relacionam a experiencia emocional, onde as bases são medições automatizadas.

A Computação Afetiva deduz que computadores sejam capazes de aferir e deduzir aspectos faciais de seres humanos e a partir de tal ponto sintetiza, se necessário, expressões emocionais humanas. Baseando-se no reconhecimento da emoção, o modelo OCC compreende que três aspectos são avaliados como eventos, ação dos agentes e aparência. Já o método de inferência híbrida é a mistura entre o método de prognóstico e diagnóstico.

Para que o sistema é capaz de reconhecer o ambiente e quantos rostos existem no cenário por método matemático que consiste em algumas etapas que constituem na detecção de faces da imagem, alinhamento fácil e por fim o sistema cadastra um novo usuário caso necessário após a identificação do mesmo. O desenvolvimento foi realizado em etapas, sendo a primeira delas a criação de funções básicas, depois a implementação do rastreo facial, verificação dos pontos faciais e, por último, a comparação das variáveis demonstrativas. A partir daí foram desenvolvidas ferramentas de funcionamento do Sensor Kinect, assim sendo possível configurar a câmera RGB

do mesmo.

É importante ressaltar que tal sistema identifica a experiência emocional para aferir o estado emocional do usuário especificamente nos estados de alegria e tristeza, as percepções são confiáveis quando se trata de uma foto, entretanto se tornam imprecisos quando o usuário desvia o olhar ou rotaciona a cabeça. Os autores afirmam que o uso dos recursos do Sensor Microsoft Kinect v2, unido às ferramentas de sua API, fornecida pelo SDK, é possível determinar o estado afetivo de um usuário.

2.5 Automatic Emotion Detection Model from Facial Expression

No artigo “[modelo de detecção automática de emoção da expressão facial](#)” (n.d.) projeta-se uma estrutura automatizada para detecção de emoções usando expressão facial. Para o sistema de reconhecimento facial considera múltiplas variáveis como cor, orientação, expressão, postura e textura do rosto humano. Neste estudo utilizou-se um quadro de streaming ao vivo, que foi processado através do método de extração de características de Gabor e rede neural.

Já para detectar a emoção nos recursos faciais extraídos utilizou-se a técnica de análise de componente principal e também agrupamentos de diferentes expressões faciais com suas respectivas emoções.

E finalmente, para determinar as expressões faciais separadamente usou-se um vetor de características obtido através de classificadores. Ressalta-se também que a estrutura proposta lida apenas com imagens faciais frontais, obtendo-se um resultado falso positivo para faces de pessoas que não expressavam nenhum tipo de emoção.

2.6 Video Game Review using Facial Emotion Recognition

O artigo [Dharshni et al. \(2020\)](#) tem com foco obter um *feedback* de um jogo eletrônico através das reações faciais produzidas em tempo real enquanto desbrava o jogo. Utilizando uma *webcam* grava as reações do jogador para ter um melhor retorno de o quanto foi bom ou ruim a experiência de jogo.

O conjunto de dados foi usado o FER2013 com rostos rotulados com uma das seis expressões básicas mais a neutra (0 = Zangado, 1 = Nojo, 2 = Medo, 3 = Feliz, 4 = Triste, 5 = Surpresa, 6

= Neutro). As imagens, que estão em escala de cinza de 48 x 48 pixel, foram divididas em 80:20 resultando em 28.109 para o conjunto de treinamento e 7178 para o de teste.

Para a detecção do rosto em cada quadro do vídeo foi utilizado o classificador *HAAR Cascade*. Posteriormente a imagem é tratada sendo convertida em tons de cinza e redimensionada para a resolução de 48x 48, forma como o classificador foi treinado.

Para a classificação das emoções o sistema proposto usou *e CNN (Convolutional Neural Network)*, pois segundo o autor as CNN é popularmente usado com vídeos, imagens 2D, espectrogramas, radares de abertura sintética.

A abordagem utilizada se aproxima muito das filmagens de câmeras de segurança que na sua maioria seria em tons de cinzas e imagens pequenas.

2.7 Protótipo De Reconhecimento De Expressões Faciais Utilizando Inteligência Artificial

Um outro artigo também referência na temática do trabalho é do [Gomes & Tavares \(2021\)](#) através do projeto: "Protótipo de Reconhecimento de Expressões Faciais Utilizando Inteligência Artificial", que foi publicado na Revista Conectus: tecnologia, gestão e conhecimento. Tem como principal objetivo criar um protótipo de aplicação que realiza o reconhecimento facial e detecta diferentes emoções de tal forma que seja possível utilizá-lo no mercado de varejo como uma ferramenta para constatar a satisfação do cliente com a aquisição do produto. Uma proposta bem diferente em relação às oferecidas pelo mercado. É possível então através da ideia proposta saber a idade aproximada e o nível de satisfação do cliente com o produto através da análise das expressões faciais dele. Todos os dados para processamento são capturados através das câmeras de vídeo local. Obteve-se como o resultado uma alta confiabilidade com 90 por cento para o gênero e idade e 70 por cento na detecção de emoções. De acordo com o autor, a Jetson Nano foi considerado o dispositivo ideal para essa prototipagem. Além do dispositivo, utilizou-se um algoritmo de detecção de objetos baseado em rede neural chamado Yolo que foi responsável pela detecção de faces e análise de dados juntamente com outros algoritmos.

2.8 Reconhecimento Facial em Imagens Capturadas por Câmeras Digitais de Rede

O artigo: Reconhecimento facial em imagens capturadas por câmeras digitais de rede de [Okabe & Carro \(2015\)](#) aborda do reconhecimento facial através de imagens capturadas por câmeras de segurança utilizando a biblioteca OpenCV em conjunto com outras bibliotecas como a JavaCV para a detecção e reconhecimento facial onde foram explorados os algoritmos de HaarCascade, Eigenface, Fisherface e LBPH. O autores implmentaram um prototipo com o objetivo de indentificar individuos em meio a um grupo de pessoas. Os testes foram realizados com alunos da Faculdade de Informática de Presidente Prudente/Unoeste . Os experimentos foram conduzidos no sentido de testar a capacidade de detecção e reconhecimento das funções embutidas no OpenCV. Outro parâmetro de teste foi a adoção ou não de filtros utilizados para preparar as imagens. O sistema proposto utilizando as bibliotecas e também rotinas de controle de câmera Axis apresentou um bom desempenho e resultados satisfatorios.

3 Metodologia

A elaboração do projeto foi realizada em três etapas: uma revisão da literatura, a adaptação de uma base de dados de imagens de expressões faciais já catalogadas em grupos específicos e a implementação de uma rede neural convolucional utilizando a linguagem python para reconhecer as emoções humanas.

3.1 Revisão da Literatura

A revisão da literatura foi realizada em cima de trabalhos e artigos acadêmicos, sobre reconhecimento de emoções, processamento de imagens e redes neurais artificiais, publicados por diversas universidades. A pesquisa bibliográfica também abrangeu as documentações das ferramentas utilizadas nesse projeto. As pesquisas foram realizadas nas bibliotecas digitais, das quais a Universidade Federal de Ouro Preto disponibiliza acesso aos alunos para fins de pesquisa.

3.2 Ferramentas

Para esse trabalho utilizamos a base de dados FER2013 que possui um bom número de imagens já catalogadas em conjunto de treino e um conjunto de teste. Como apresentado nos artigos correlatos, utilizaremos as sete classes para representar as emoções básicas e o neutro, sendo elas: Raiva, Nojo, Medo, Surpresa, Felicidade, Tristeza e Neutro.

Para a detecção das faces utilizamos o OpenCV baseado em Cascata de Haar, ele possibilita a detecção de rostos em imagens estáticas e em fluxos de vídeo em tempo real.

Utilizamos a API Keras do TensorFlow para criar e treinar nossos modelos, ela possui uma interface simples e consistente para criação de modelos modulares e compostos.

3.3 Desenvolvimento

Para conseguirmos encontrar o reconhecimento facial ou alguma característica específica foi necessário resolver alguns problemas relacionados entre si. Primeiro, deve-se encontrar um rosto na imagem, em seguida, deve-se focar individualmente em cada rosto encontrado, mesmo se estiver rotacionado ou com iluminação ruim. O terceiro passo demanda ser capaz de identificar aspectos únicos no rosto que possam diferenciá-los de outros grupos, tais como elevação das sobrancelhas, formato da boca, etc. E finalmente, comparar dados aspectos do rosto com todos os grupos pré-classificados de modo a obter o grupo com o qual este rosto mais se assemelha.

4 Resultados

Foram criados cinco modelos com diferentes redes neurais convolucionais, em cada modelos foram realizadas as etapas de pré-processamento da imagem, treinamento e teste para verificar a acurácia alcançada. Nesta seção vamos apresentar os resultados obtidos por meio de gráficos da evolução do treinamento, a matriz de confusão gerada e o teste realizado.

4.1 Modelo 1

Para o primeiro modelo utilizamos a arquitetura VGG, definindo 256 como tamanho do lote para atualização dos pesos e 100 épocas para treino. Para realizarmos a convolução, utilizamos

a Conv2D com função de ativação Relu. Para reduzir a resolução de entrada ao longo de suas dimensões espaciais (altura e largura), utilizamos o MaxPooling2D obtendo o valor máximo sobre uma janela de entrada para cada canal de entrada. E utilizamos a função SoftMax para classificação entre as classes distintas.

Com o modelo treinado conseguimos alcançar uma acurácia de 61.60 e obtivemos um Erro de 1.46. Durante as etapas de treinamento, as seguintes melhoras foram obtidas e representadas no gráfico da Figura 1. A matriz de confusão obtida para a classificação de cada umas das emoção foi apresentada na Figura 2. E para realizar os testes utilizamos uma imagem que possui varias fotos com diferentes emoções e a sua respectiva classificação como apresentado na Figura 3.

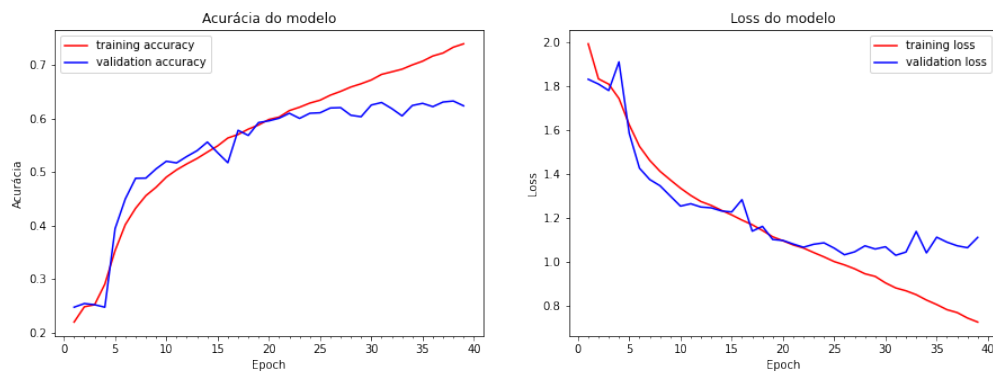


Figura 1: Gráfico de melhora nas etapas de treinamento

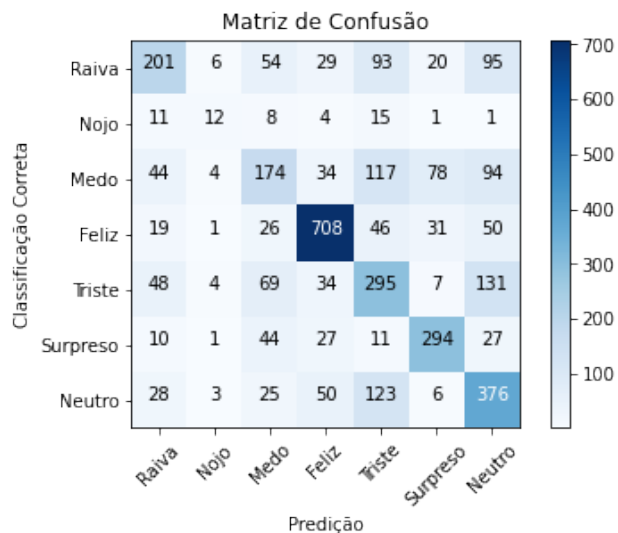


Figura 2: Matriz de confusão das emoções



Figura 3: Teste do modelo

4.2 Modelo 2

Para o segundo modelo utilizamos a arquitetura VGG, definindo 16 como tamanho do lote para atualização dos pesos e 100 épocas para treino. Para realizarmos a convolução, utilizamos a Conv2D com função de ativação Elu. Para reduzir a resolução de entrada ao longo de suas dimensões espaciais (altura e largura), utilizamos o MaxPooling2D obtendo o valor máximo sobre uma janela de entrada para cada canal de entrada. E utilizamos a função SoftMax para classificação entre as classes distintas.

Com o modelo treinado conseguimos alcançar uma acurácia de 64.30 e obtivemos um Erro de 1.21. Durante as etapas de treinamento, as seguintes melhoras foram obtidas e representadas no gráfico da Figura 4. A matriz de confusão obtida para a classificação de cada umas das emoção foi apresentada na Figura 5. E para realizar os testes utilizamos uma imagem que possui varias fotos com diferentes emoções e a sua respectiva classificação como apresentado na Figura 6.

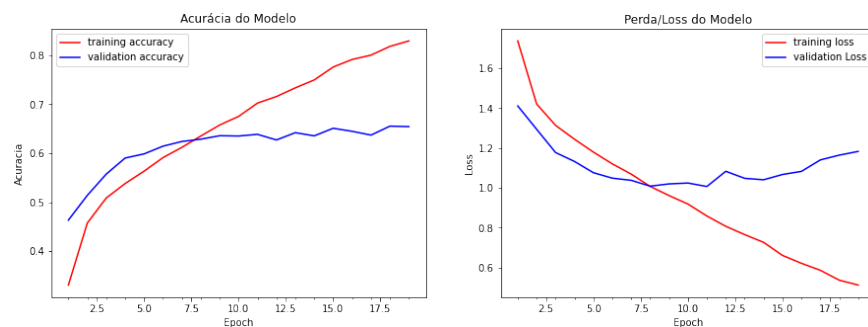


Figura 4: Gráfico de melhoria nas etapas de treinamento

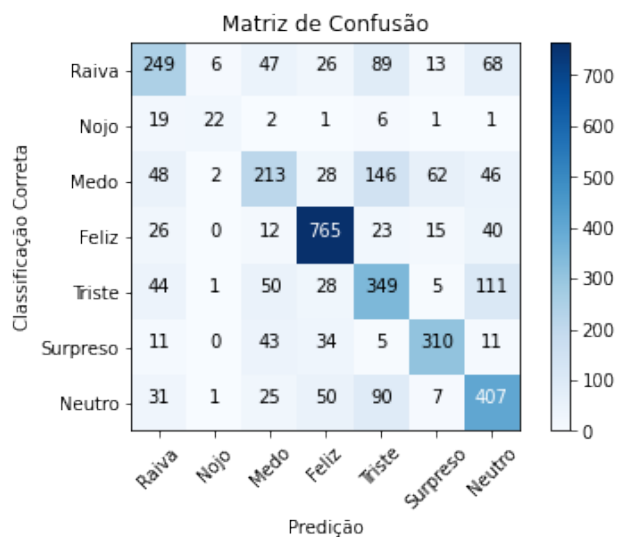


Figura 5: Matriz de confusão das emoções



Figura 6: Teste do modelo

4.3 Modelo 3

Para o terceiro modelo utilizamos a arquitetura VGG, definindo 64 como tamanho do lote para atualização dos pesos e 100 épocas para treino. Para realizarmos a convolução, utilizamos a Conv2D com função de ativação Relu. Para reduzir a resolução de entrada ao longo de suas dimensões espaciais (altura e largura), utilizamos o MaxPooling2D obtendo o valor máximo sobre uma janela de entrada para cada canal de entrada. E utilizamos a função SoftMax para classificação entre as classes distintas.

Com o modelo treinado conseguimos alcançar uma acurácia de 61.40 e obtivemos um Erro de 1.21. Durante as etapas de treinamento, as seguintes melhoras foram obtidas e representadas no gráfico da Figura 7. A matriz de confusão obtida para a classificação de cada umas das emoção foi apresentada na Figura 8. E para realizar os testes utilizamos uma imagem que possui varias fotos com diferentes emoções e a sua respectiva classificação como apresentado na Figura 9.

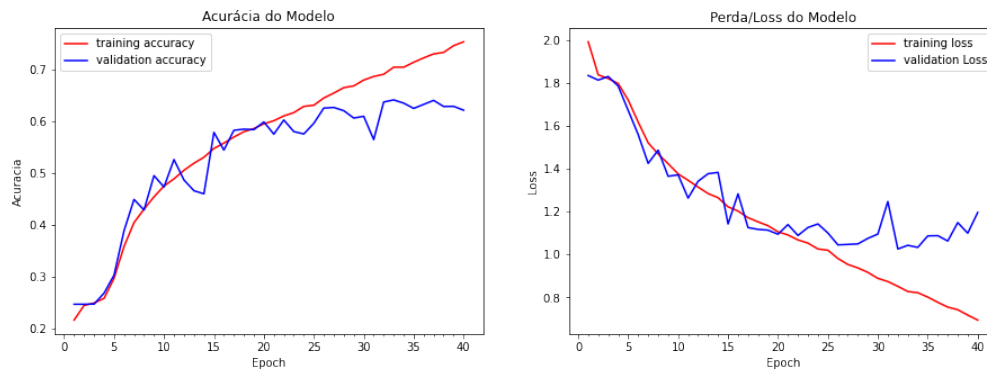


Figura 7: Gráfico de melhora nas etapas de treinamento

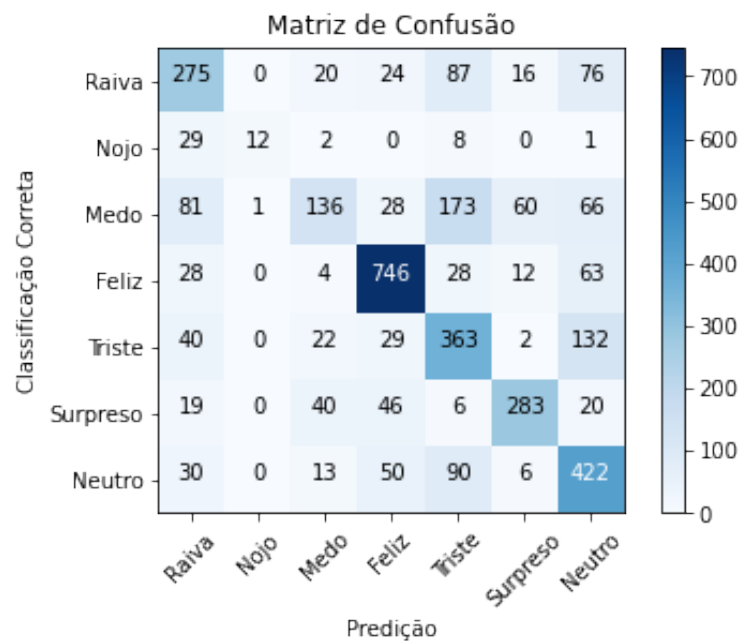


Figura 8: Matriz de confusão das emoções



Figura 9: Teste do modelo

4.4 Modelo 4

Para o quarto modelo utilizamos a arquitetura VGG, definindo 64 como tamanho do lote para atualização dos pesos e 100 épocas para treino. Para realizarmos a convolução, utilizamos a Conv2D com função de ativação Relu. Para reduzir a resolução de entrada ao longo de suas dimensões espaciais (altura e largura), utilizamos o MaxPooling2D obtendo o valor máximo sobre uma janela de entrada para cada canal de entrada. E utilizamos a função SoftMax para classificação entre as classes distintas.

Com o modelo treinado conseguimos alcançar uma acurácia de 60.76 e obtivemos um Erro de 1.22. Durante as etapas de treinamento, as seguintes melhoras foram obtidas e representadas no gráfico da Figura 10. A matriz de confusão obtida para a classificação de cada umas das emoção foi apresentada na Figura 11. E para realizar os testes utilizamos uma imagem que possui varias fotos com diferentes emoções e a sua respectiva classificação como apresentado na Figura 12.

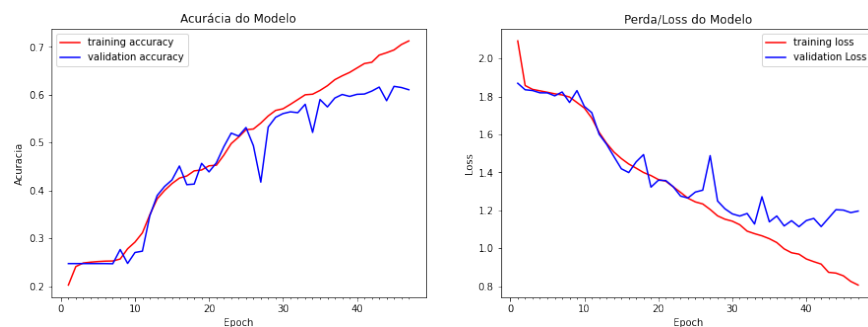


Figura 10: Gráfico de melhora nas etapas de treinamento

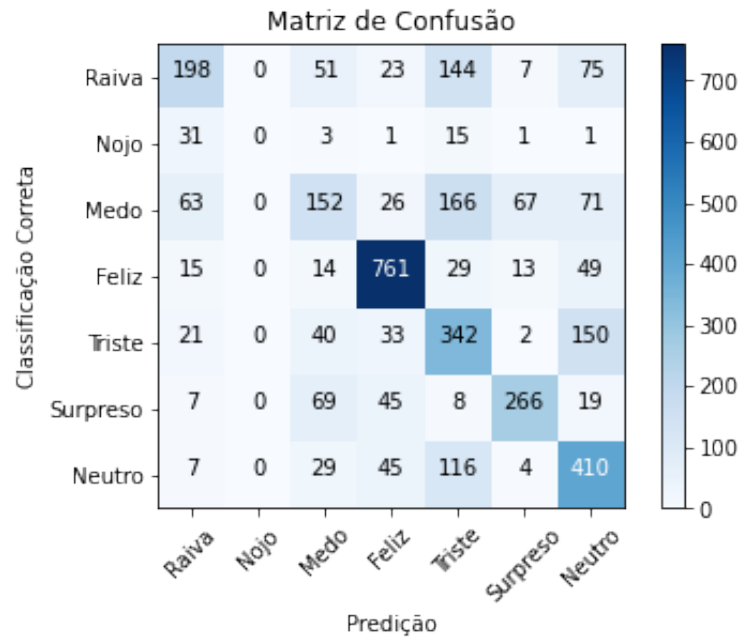


Figura 11: Matriz de confusão das emoções



Figura 12: Teste do modelo

4.5 Modelo 5

Para o quinto modelo utilizamos a arquitetura VGG, definindo 256 como tamanho do lote para atualização dos pesos e 100 épocas para treino. Para realizarmos a convolução, utilizamos a Conv2D com função de ativação Relu. Para reduzir a resolução de entrada ao longo de suas dimensões espaciais (altura e largura), utilizamos o MaxPooling2D obtendo o valor máximo sobre

uma janela de entrada para cada canal de entrada. E utilizamos a função SoftMax para classificação entre as classes distintas.

Com o modelo treinado conseguimos alcançar uma acurácia de 59.87 e obtivemos um Erro de 1.56. Durante as etapas de treinamento, as seguintes melhoras foram obtidas e representadas no gráfico da Figura 13. A matriz de confusão obtida para a classificação de cada umas das emoção foi apresentada na Figura 14. E para realizar os testes utilizamos uma imagem que possui varias fotos com diferentes emoções e a sua respectiva classificação como apresentado na Figura 15.

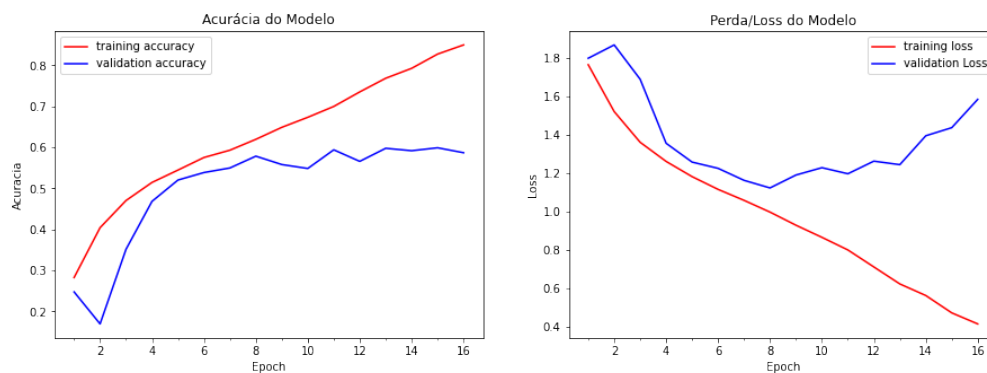


Figura 13: Gráfico de melhoria nas etapas de treinamento

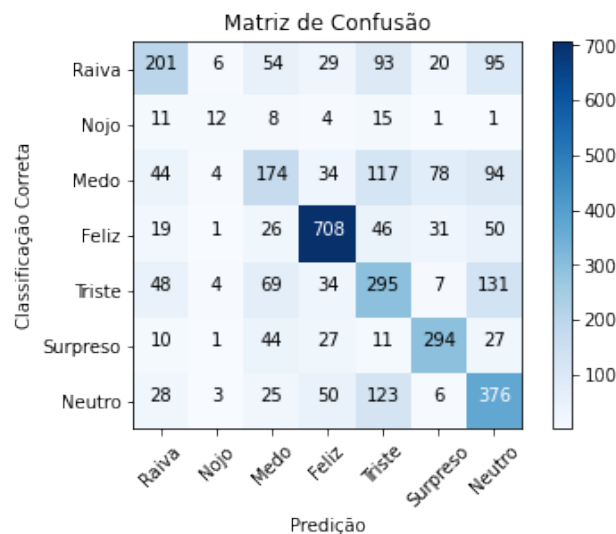


Figura 14: Matriz de confusão das emoções

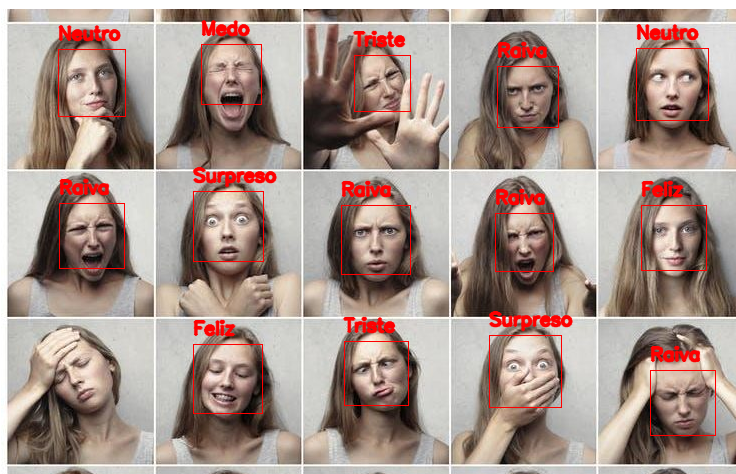


Figura 15: Teste do modelo

4.6 Teste do modelo do Detector de Emoções em Webcam e Video

Com os nossos modelos já treinados, selecionamos a opção que possui a melhor acurácia para testar nosso detector de emoções. Em um primeiro momento, utilizamos a webcam do computador para captar o video como entrada e tiramos uma foto em um determinado momento para verificar se a emoção é classificada corretamente. No segundo momento, carregamos um video já gravado como entrada e realizamos a detecção das emoções frame a frame para verificar se ela é classificada corretamente. Conseguimos obter ótimos resultados na classificação das emoções.



Figura 16: Medo



Figura 17: Neutro

Figura 18: Teste do detector de emoções na Webcam



Figura 19: Triste



Figura 20: Raiva

Figura 21: Teste do detector de emoções em Video

5 Conclusão

Neste trabalho foi desenvolvido um software que realiza a detecção de emoções humanas por meio de câmeras de vigilância, além de discriminar as mesmas entre diferentes pessoas no vídeo. O algoritmo se mostrou uma eficiente ferramenta na detecção e classificação das emoções.

Optamos por utilizar a rede neural convolucional, por apresentar um algoritmo de aprendizado profundo que pode captar uma imagem de entrada e atribuir importâncias como pesos e vieses a vários aspectos e objetos da imagem e ser capaz de diferenciar umas das outras. Outro ponto importante que levamos em consideração, foi que a CNN exige um menor pré-processamento em comparação a outros algoritmos de classificação e têm a capacidade de aprender sozinha esses filtros ou características.

O uso de rede neural convolucional apresentou ótimos resultados, com precisão superior a 60 por cento na classificação de todas as classes. No conjunto de 35 mil imagens, obtivemos apenas 1.21 por cento de erro nas imagens analisadas. Além disso, o uso de redes convolucionais possibilita um rápido processamento e viabiliza o treinamento de um grande conjunto de imagens.

Para trabalhos futuros pretendemos treinar nosso próprio modelo de forma que ele seja mais adequado ao cenário atual, utilizando uma quantidade maior de imagens e buscando aplicar outras arquiteturas. E dado a dificuldade encontrada no reconhecimento facial de algumas imagens, vamos buscar outras técnicas para melhorar a identificação como a Facial Landmarks (FACS).

Referências

- Aires, G. V., Santos, I. M., Brandão, P. S., et al. (n.d.). Reconhecimento de expressões faciais através de análise facial computadorizada utilizando o kinect v2.
- Behling, A., et al. (2019). Reconhecimento de emoções em vídeo utilizando redes neurais artificiais.
- Dharshni, R. D., SreedeviM., G., SreelakshmiM., G., Raj, S., & NeethuM., S. (2020). Video game review using facial emotion recognition. *International Journal of Engineering Research and*, 9.
- Ekman, P., & Friesen, W. (2003, 01). Unmasking the face: A guide to recognizing emotions from facial clues.
- Gomes, G. O., & Tavares, J. L. (2021). Protótipo de reconhecimento de expressões faciais utilizando inteligência artificial. *Revista Conectus: tecnologia, gestão e conhecimento*, 1(2).
- Khan, F. (2018). Facial expression recognition using facial landmark detection and feature extraction via neural networks. *CoRR*, abs/1812.04510. Retrieved from <http://arxiv.org/abs/1812.04510>
- Libralon, G. L. (2014). *Modelagem computacional para reconhecimento de emoções baseada na análise facial* (Unpublished doctoral dissertation). Universidade de São Paulo.
- modelo de detecção automática de emoção da expressão facial. (n.d.). In *2016 conferência internacional sobre controle de comunicação avançada e tecnologias de computação (icaccct)*. doi: 10.1109/ICACCCT.2016.7831605
- Okabe, R. K., & Carro, S. A. (2015). Reconhecimento facial em imagens capturadas por câmeras digitais de rede. In *Colloquium exactarum*. issn: 2178-8332 (Vol. 7, pp. 106–119).
- Revina, I. M., & Emmanuel, W. S. (2018). A survey on human face expression recognition techniques. *Journal of King Saud University-Computer and Information Sciences*.