

Identificação de emoções em tempo real utilizando Redes Neurais Convolucionais

Luis F. Bertuol Ulisses L. da Silva

Instituto Tecnológico de Aeronáutica

São José dos Campos/SP – 2025

Resumo—A presente proposta de projeto apresenta o desenvolvimento de um sistema para reconhecimento de emoções faciais em tempo real a partir de imagens capturadas pela webcam. O sistema combina a detecção de rostos por meio de filtros HaarCascade, que permitem identificar rapidamente a região de interesse, com a classificação das expressões faciais realizada por uma Rede Neural Convolucional (CNN). O modelo é treinado sobre a base de dados Augmented FER2013, composta por imagens em tons de cinza de rostos humanos rotulados em sete categorias emocionais. Durante o treinamento, são aplicadas técnicas de deep learning, como camadas convolucionais, pooling, dropout, batch normalization e validação cruzada, a fim de melhorar a generalização e reduzir o sobreajuste. O objetivo principal é permitir que o sistema identifique automaticamente a face em frente à câmera e classifique a emoção exibida dentre as opções disponíveis, garantindo funcionamento em tempo real. A pesquisa justifica-se pela ampla aplicabilidade de sistemas de reconhecimento emocional em áreas como interfaces homem-máquina mais empáticas, educação adaptativa, saúde mental, monitoramento de fadiga e segurança. Os resultados esperados incluem não apenas a validação do modelo em condições controladas, mas também a análise de seu desempenho em cenários práticos, com vistas à sua evolução futura para aplicações mais robustas e generalizáveis.

Palavras-chave—redes neurais convolucionais, visão computacional, emoções, reconhecimento em tempo real, deep learning.

I. INTRODUÇÃO

As Redes Neurais Convolucionais (CNNs) consolidaram-se como o estado da arte no processamento de imagens, pois são capazes de extrair características hierárquicas a partir de dados visuais, reconhecendo padrões sutis como bordas, texturas e formas. Aplicadas ao reconhecimento de expressões faciais, as CNNs permitem classificar emoções a partir de imagens em tempo real, alcançando resultados superiores aos métodos tradicionais. Em complemento, técnicas clássicas como o uso de filtros HaarCascade, implementados na biblioteca OpenCV, oferecem uma forma eficiente e de baixo custo computacional para detecção de rostos, servindo como etapa preliminar de localização antes da classificação pela rede neural.

Neste trabalho, propõe-se o desenvolvimento de um sistema para reconhecimento de emoções faciais em tempo real captadas por webcam. O sistema será treinado sobre a base Augmented FER2013, que amplia o conjunto FER2013 por meio de técnicas de data augmentation, tornando-o mais robusto às variações de pose, iluminação e expressão. As imagens, em escala de cinza e resolução fixa, são organizadas

em conjuntos de treino e validação. A CNN será composta por múltiplas camadas convolucionais e densas, empregando recursos como *pooling*, *dropout*, *batch normalization* e técnicas de regularização. O treinamento fará uso de validação cruzada e ajustes de hiperparâmetros, como *early stopping*, *learning rate scheduling*, etc., para aprimorar a capacidade de generalização.

A. Problema

Este trabalho aborda o reconhecimento de emoções faciais em tempo real a partir de imagens capturadas por webcam. A detecção do rosto será realizada por meio de classificadores em cascata do tipo Haar (*HaarCascade*), e a classificação das expressões será feita por uma Rede Neural Convolucional (CNN), treinada a partir da base de dados Augmented FER2013, disponível no Kaggle.

B. Motivação

O reconhecimento automático de emoções faciais constitui um dos desafios mais instigantes da visão computacional e da inteligência artificial contemporânea. A capacidade de um sistema identificar e interpretar estados emocionais humanos em tempo real abre possibilidades em múltiplos domínios, como educação personalizada (feedback em tempo real do engajamento do aluno), monitoramento de saúde, interação homem-máquina mais natural e empática, e até mesmo em sistemas de transporte inteligentes, com mecanismos de segurança capazes de detectar sinais de fadiga ou estresse em motoristas. Nesse contexto, estudar e aprimorar soluções computacionais para detecção de emoções faciais é de grande relevância científica e social.

C. Objetivos

Os objetivos específicos do presente trabalho são:

- detectar, em tempo real, faces na imagem da webcam;
- classificar a emoção exibida dentre um conjunto pré-definido de categorias.

Para alcançar esses objetivos específicos, buscar-se-á atingir os seguintes objetivos gerais:

- treinar e validar uma CNN usando a base *Augmented FER2013* do Kaggle;
- avaliar desempenho por métricas padrão (acurácia, F1 macro, matriz de confusão);
- disponibilizar um protótipo executável, com código e instruções de reprodução.

II. METODOLOGIA

A. Base de dados

Será utilizada a base Augmented FER2013 (derivada da FER2013), disponível publicamente no Kaggle.

As imagens são faces em escala de cinza com resolução 48×48 , rotuladas em sete emoções: *angry*, *disgust*, *fear*, *happy*, *sad*, *surprise* e *neutral*.

Variáveis de entrada: imagens 48×48 em escala de cinza (1 canal).

Variável de saída: rótulo de emoção (7 classes, codificação *one-hot*).

Número de amostras: partindo-se das cerca de $\sim 35,9k$ imagens originais (FER2013), a versão *Augmented* inclui variações sintéticas (rotações, translações, zoom, *flip* horizontal, ruído leve), elevando o número efetivo de exemplos vistos por época. As contagens exatas serão reportadas após o download/organização da base no código-fonte.

B. Organização e divisão dos dados

Os dados estão organizados em pastas por classe no arquivo compactado original, já divididos entre treino e teste. Além disso, será empregada validação cruzada para estimar a variância de desempenho e reduzir viés de partição.

C. Detecção facial (tempo real)

A detecção será feita com *HaarCascade* (disponível na biblioteca OpenCV).

D. Arquitetura da CNN

Será utilizada uma CNN multicamadas, cujo número de blocos será definido ao longo do desenvolvimento do projeto.

Técnicas previstas: Batch Normalization, Dropout, *data augmentation*, *early stopping* (paciente $\approx 15-20$ épocas), *learning rate scheduling* (p.ex., *ReduceLROnPlateau*), e *gradient clipping*.

E. Treinamento

Otimizador: Adam.

Perda: *categorical cross-entropy* com *class weights*.

Métricas: acurácia, F1 macro, precisão/recall por classe.

Batch size/épocas: 100-200 épocas (com *early stopping*).

Callbacks: *ModelCheckpoint* (melhor validação), *EarlyStopping*, *LR scheduler*.

Sementes aleatórias serão fixadas para reprodutibilidade.

F. Avaliação e análise

No conjunto de validação/teste serão avaliados os seguintes aspectos:

- Acurácia global;
- F1 macro e micro;
- Relatório por classe (precisão, recall e F1-score);
- Matriz de confusão;
- Curvas de aprendizado (perda e acurácia por época).

G. Avaliação em tempo real

O código será desenvolvido num arquivo Jupyter Notebook, na linguagem Python (OpenCV + TensorFlow/Keras) que captura vídeo da webcam, detecta faces via *HaarCascade*, classifica emoções com a CNN e exibe a predição sobreposta. A avaliação será testada, também, ao vivo no dia da apresentação do projeto.

H. Repositórios

Os arquivos utilizados, bem como resultados obtidos e demais artefatos produzidos serão armazenados num repositório público do Github, a ser disponibilizado na versão final do projeto.

REFERÊNCIAS

- [1] Goodfellow et al., “Challenges in Representation Learning: A report on three machine learning contests”, Unicer, 2013. Disponível em: <http://arxiv.org/abs/1307.0414>.
- [2] Goodfellow, I., Bengio, Y., Courville, A., *Deep Learning*. MIT Press, Cambridge, Massachusetts, 2016. Disponível em: <http://www.deeplearningbook.org>.
- [3] D. P. Lopez, F. Z. Canal, G. G. Scotton, E. Pozzebon e A. C. Sobieranski, “Uma abordagem computacional em tempo real para reconhecimento de expressão facial humana baseada na extração de recursos de referência”, *SEVENED – Revista de Saúde, Educação e Meio Ambiente*, vol. 4, n. 1, pp. 4–6, 2024. Disponível em: <https://doi.org/10.56238/sevened2024.004-006>.
- [4] N. Mehendale, “Facial emotion recognition using convolutional neural networks (FERC)”, *SN Applied Sciences*, vol. 2, no. 3, pp. 1–8, Feb. 2020. DOI: <https://doi.org/10.1007/s42452-020-2234-1>.
- [5] R. Ratan, “Building an Emotion Detector with LittleVGG”, GitHub repository, [Online]. Available: <https://github.com/rajeevratan84/DeepLearningCV/blob/master/18.2>.