

Relatório Técnico: Previsão de Rostos

Residentes: Saulo Alves e Thailane Carmo

Data de entrega: 17/11/2024

1. Resumo

Este trabalho teve como finalidade desenvolver e avaliar modelos de redes neurais convolucionais (CNNs) para a distinção entre rostos masculinos e femininos, utilizando imagens do banco de dados CUHK Face Sketch Database. Foram realizadas análises com imagens coloridas, imagens convertidas para preto e branco, além da aplicação de técnicas de aumento de dados (data augmentation). Métricas como Acurácia, F1-Score e AUC-ROC foram empregadas para a comparação dos modelos. Os resultados demonstraram que tanto o aumento de dados quanto a conversão do conjunto de dados para preto e branco contribuíram significativamente para a melhoria do desempenho do modelo.

2. Introdução

A classificação de imagens é um desafio central na visão computacional, com aplicações que vão desde o reconhecimento facial até diagnósticos médicos. O objetivo geral dessa tarefa é categorizar imagens em classes pré-definidas com base em padrões visuais. No entanto, devido à alta dimensionalidade das imagens e à complexidade das características visuais, métodos tradicionais de aprendizado de máquina frequentemente apresentam limitações em precisão e escalabilidade.

Neste contexto, as redes neurais convolucionais (CNNs) surgiram como uma abordagem altamente eficaz para problemas de classificação de imagens. Inspiradas na estrutura do córtex visual humano, as CNNs são projetadas para identificar automaticamente padrões hierárquicos, desde bordas simples a estruturas mais complexas, como formas e texturas. Elas utilizam camadas convolucionais para extração de características, camadas de pooling para redução de dimensionalidade e camadas densas para tomada de decisão, sendo especialmente adequadas para lidar com imagens de alta resolução.

Além disso, as CNNs têm a capacidade de generalizar eficientemente em grandes volumes de dados e são robustas a variações nas imagens, como iluminação, ângulo e escala. Com o suporte de técnicas avançadas, como ajuste de hiperparâmetros e aumento de dados, essas redes podem obter desempenhos superiores em tarefas desafiadoras, como a diferenciação de características sutis entre classes.

O presente trabalho utiliza CNNs para a classificação de imagens de rostos, com ênfase na distinção entre rostos masculinos e femininos. Essa tarefa não apenas avalia a eficácia das redes em identificar padrões faciais, mas também investiga o impacto de diferentes pré-processamentos de imagem e ajustes de hiperparâmetros no desempenho dos modelos.

3. Metodologia

Inicialmente, as imagens foram normalizadas, padronizando todos os pixels entre 0 e 1. Em seguida, o conjunto de dados foi dividido em conjuntos de treino, teste e validação, utilizando o parâmetro stratify para garantir a proporcionalidade das classes.

A arquitetura desenvolvida foi simples, composta por uma camada de entrada e quatro camadas de convolução e pooling. Ao final da arquitetura, foram adicionadas uma camada Flatten e uma camada densa com função sigmoid. O modelo foi compilado utilizando o otimizador Adam, e a métrica de perda escolhida foi a binary cross-entropy.

Os hiperparâmetros foram determinados por meio de grid search, utilizando a biblioteca Keras Tuner. Os melhores parâmetros foram selecionados após 30 testes para treinar as redes. Em uma das execuções, realizamos um teste com imagens em preto e branco, aplicando os mesmos parâmetros utilizados para as imagens originais. Além disso, implementamos o early stopping como um callback para evitar que o modelo fosse treinado por mais tempo do que o necessário.

O processo de avaliação começou com a análise da acurácia e da perda, seguido pela utilização de F1-Score, curva ROC e área sob a curva ROC para avaliar o desempenho do modelo. Também foram examinados os rostos que apresentaram predições incorretas em busca de padrões nos erros. A metodologia foi aplicada de forma uniforme nos três testes realizados (imagens originais, preto e branco e com aumento de dados).

4. Resultados

O F1-Score indicou um bom equilíbrio entre as classes nos testes com imagens em preto e branco e com data augmentation. Contudo, com as imagens originais, mesmo utilizando os melhores parâmetros encontrados durante o grid search, o desempenho foi insatisfatório.

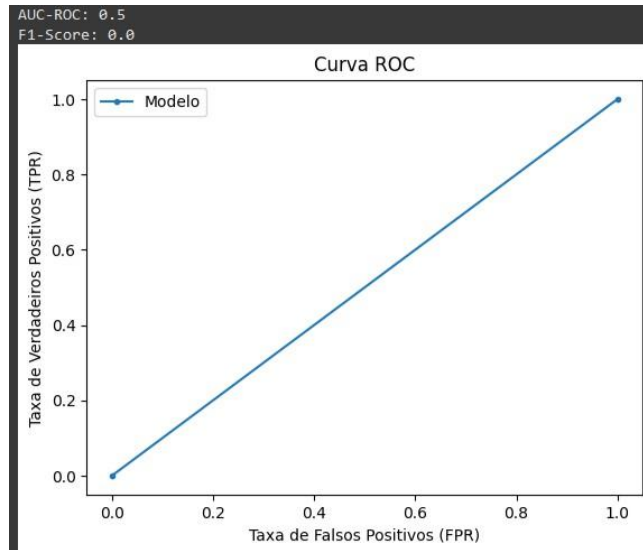


Figura 1: Imagens originais

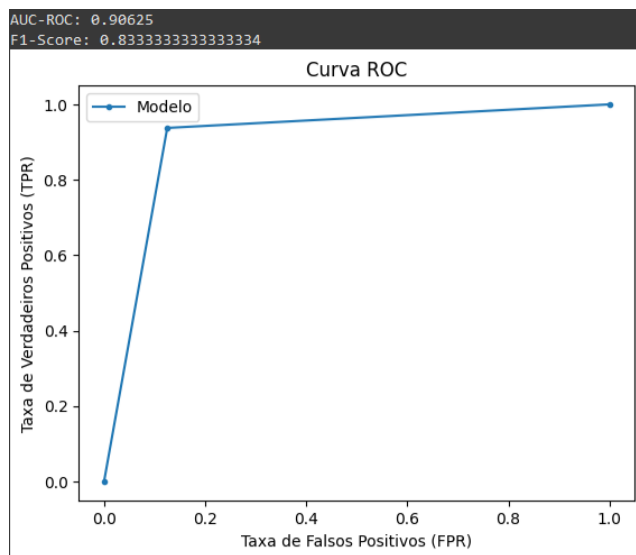


Figura 2: Imagens em Preto e Branco

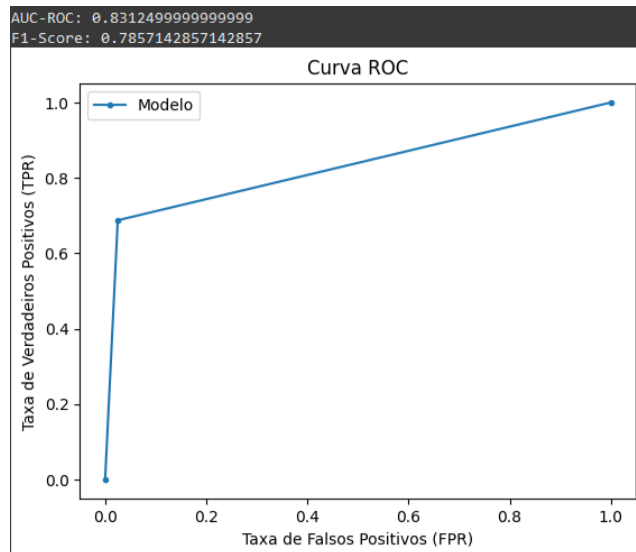


Figura 3: Imagens com Data Augmentation

Os resultados obtidos com as imagens em preto e branco foram os melhores, o que me surpreendeu, visto que esperava que o Data Augmentation produzisse um desempenho significativamente superior, mesmo em comparação com as imagens em preto e branco.

Algumas imagens apresentavam borrões, como o caso de uma moça piscando o olho, ou estavam desfocadas. Embora esses exemplos estivessem entre os erros, eram poucos no conjunto de dados. A maioria das classificações incorretas envolveu principalmente mulheres, possivelmente devido ao desbalanceamento entre as classes. Cabelos amarrados poderiam criar a impressão de cabelo curto masculino, e rostos com características mais quadradas foram erroneamente identificados como homens. Além disso, devido à sua minoria no conjunto, crianças (garotas) também foram classificadas incorretamente. Acredito que aprimorar a estrutura das camadas, talvez com a inclusão de dropout ou batch normalization, poderia melhorar o desempenho; no entanto, optei por manter a arquitetura simples e explorar diferentes formas de utilizá-la.

Uma surpresa durante o desenvolvimento ocorreu ao adicionar o grid search, pois, ao contrário do esperado, algumas métricas pioraram, especialmente aquelas relacionadas aos dados com imagens originais. Essa situação foi bastante preocupante, mas, como os resultados foram positivos nos outros casos, mantive essa abordagem no código.



Figure 4: Erros

6. Conclusão

O projeto evidenciou a eficácia das redes neurais convolucionais na classificação de gênero em imagens faciais, empregando técnicas como conversão para escala de cinza, data augmentation e ajuste de hiperparâmetros. Os resultados, avaliados por métricas como F1-Score e AUC-ROC, demonstraram que o modelo foi capaz de identificar padrões relevantes, com ênfase em informações espaciais e texturais, enquanto a cor se revelou menos determinante para essa tarefa. O uso de data augmentation também foi fundamental para aumentar a robustez do modelo.

Para aprimorar o projeto, planejo utilizar bases de dados maiores e mais diversificadas, além de explorar arquiteturas mais avançadas, como a ResNet ou EfficientNet. Além disso, métodos de pré-processamento mais sofisticados, como equalização de histograma, e a aplicação de técnicas de interpretação, como o Grad-CAM, podem potencializar o desempenho do modelo e proporcionar uma compreensão mais aprofundada dos padrões usados na classificação.

7. Referências

CLARK, Steven. Matplotlib: Visualization with Python. 2003. Disponível em: <https://matplotlib.org/>.

KAGGLE. CUHK Face Sketch Database (CUFS). Disponível em: <https://www.kaggle.com/datasets/arbazkhan971/cuhk-face-sketch-database-cufs>.

HUNTER, John D. Matplotlib: A 2D Graphics Environment. Computing in Science Engineering, v. 9, n. 3, p. 90-95, 2007.

PYTORCH TEAM. TorchVision: Models, Datasets, and Transforms. Disponível em: <https://pytorch.org/vision/>.