



Relatório: Implementação e Análise de Classificação com Redes Convolucionais no dataset CUFS

RESIDÊNCIA TECNOLÓGICA
RESTIC36

ANA CAROLINA SANTOS DE SOUZA

BRENDA BARBOSA DE OLIVEIRA

IMPLEMENTAÇÃO E ANÁLISE DE CLASSIFICAÇÃO DE
GÊNERO COM REDES CONVOLUCIONAIS NO DATASET
CUFS

RESUMO

Este relatório tem como objetivo descrever a preparação, processamento e análise para o desenvolvimento de um modelo de classificação de gênero. Para isso, o conjunto de imagens do *CUHK Face Sketch Database (CUFS)* foi utilizado no treinamento de uma rede convolucional simples. Desse modo, a partir de uma arquitetura eficiente, é possível classificar automaticamente os indivíduos das imagens de acordo com o gênero.

INTRODUÇÃO

A classificação de imagens é uma tarefa fundamental em muitos sistemas de inteligência artificial, sendo amplamente utilizada em aplicações como reconhecimento facial, diagnóstico médico e identificação de objetos em fotografias. Um dos maiores desafios nesse tipo de problema é a extração eficiente de padrões visuais significativos que permitem distinguir diferentes classes, elementos, animais ou objetos de imagens. Redes Neurais Convolucionais (CNNs) surgiram como uma abordagem poderosa para superar esses desafios, sendo projetadas especificamente para lidar com os dados estruturados em grade, como imagens.

As redes neurais convolucionais (CNNs) são arquiteturas de aprendizado profundo especialmente projetadas para trabalhar com dados visuais. Elas possuem características que as tornam ideais para a tarefa de classificação de imagens, como reconhecimento de bordas, texturas, formas e padrões complexos com precisão. Ao identificar pequenos padrões como curvas que formam outros padrões como um rosto, as redes neurais convolucionais permitem classificar e distinguir elementos em uma foto similarmente ao processo humano. Essa abordagem torna as CNNs especialmente adequadas para tarefas que exigem diferenciação de padrões visuais sutis, como é o caso da classificação de rostos.

Neste projeto, o dataset CUHK Face Sketch Database (CUFS), geralmente empregado para mapear fotografias de rostos para esboços artísticos, foi utilizado com foco na classificação das imagens de rostos fotográficos com base no gênero (masculino ou feminino). Desse modo, a solução baseada no modelo de redes convolucionais também incluiu a preparação, normalização e pré processamento das imagens para treinamento e classificação dos indivíduos. A análise abrange todas as etapas do processo, desde a preparação dos dados até a avaliação do modelo, destacando os desafios enfrentados e os resultados obtidos.

METODOLOGIA

O dataset CUFS inclui 188 imagens coloridas de dimensões 250 x 200 pixels de rostos do banco de dados estudantil da *Chinese University of Hong Kong* (CUHK) obtidas no diretório *photos*. Para a preparação e processamento dos dados as bibliotecas *keras*, *tensorflow*, *numpy* *scikit-learn* e *matplotlib* foram utilizadas.

Sendo o título de cada uma das imagens iniciado por “f” para imagens de pessoas do gênero feminino e “m” para indivíduos do gênero masculino, a obtenção do rótulos para cada classe foi obtida através dessa organização do dataset. Dessa maneira, cada imagem foi analisada manualmente para codificação de rótulos adequados para a arquitetura da CNN, com 0 representando masculino e 1 representando feminino.

Devido ao reduzido número de amostras, algumas técnicas de *data augmentation*, principalmente rotações, foram aplicadas ao conjunto de treinamento de modo a aumentar a diversidade dos dados e evitar o *overfitting*, totalizando ao final 940 imagens no conjunto. Em seguida, para garantir que as imagens estivessem no formato ideal para o treinamento do modelo, todas as imagens foram redimensionadas para 250x200 pixels para uniformidade. Também foi realizado o processo de normalização dividindo os valores RGB por 255, para obter uma melhor distribuição dos valores dos pixels no intervalo entre 0 e 1 e melhor desempenho no treinamento.

O dataset de 940 imagens foi dividido em 50% para treinamento, 30% para validação e 20% para teste, utilizando a seed 23 para garantir replicabilidade. Além disso, a distribuição das classes nos conjuntos foi realizada de acordo com a proporção dos gêneros no conjunto de imagens inicial. A distribuição das classes no conjunto de treino, validação e testes é explicitada na Tabela 1.

	Quantidade no conjunto de treino	Quantidade no conjunto de validação	Quantidade no conjunto de teste
Masculino	355	201	134
Feminino	135	54	54

Tabela 1 - Distribuição das classes nos modelos de treino, validação e testes.

Assim, o modelo de rede neural convolucional implementado tem a seguinte estrutura:

- **Camada de Entrada:** Conv2D com 32 filtros, kernel 3x3, função de ativação ReLu, Dropout de 20% e seguida por uma operação de Max Pooling 2x2 para redução da dimensionalidade. Os dados de entrada nesta camada tem formato (250, 200, 3).
- **Camadas Convolucionais:**
 - Conv2D com 64 filtros, kernel 3x3, função de ativação ReLU, Dropout de 20% e Max Pooling 2x2.
 - Conv2D com 128 filtros, kernel 3x3, e função de ativação ReLU, Dropout de 20% e Max Pooling 2x2..
 - Conv2D com 256 filtros, kernel 3x3, e função de ativação ReLU, Dropout de 20% e Max Pooling 2x2..
 - Conv2D com 512 filtros, kernel 3x3, e função de ativação ReLU, Dropout de 20% e Max Pooling 2x2..
- **Camada Densa:** Uma camada densa com 512 neurônios, ReLU e Dropout de 20%.
- **Camada de Saída:** Finalmente, a saída com 1 neurônio e ativação sigmoid para classificação binária.

O modelo foi compilado com a função de perda *Binary Crossentropy*, otimizador: Adam, com taxa de aprendizado inicial de 0.001. Também a métrica de eficiência adotada para dizer o quanto a rede acerta a cada época foi a acurácia.

Dessa maneira, o modelo foi treinado por 30 épocas com um batch size de 16, utilizando o conjunto de validação para monitorar o desempenho e aplicando *Early Stopping* com paciência de 5 épocas. Estes hiperparâmetros e seus valores foram ajustados e definidos para o melhor desempenho no treinamento do modelo de aprendizagem. Finalmente, as métricas, F1-score, curva ROC e a área sob a curva ROC (AUC-ROC) foram analisadas para fornecer uma visão detalhada do desempenho do modelo.

DISCUSSÃO

O treinamento apresentou convergência após 22 épocas, com a acurácia de validação estabilizando em 85,63%. A perda no conjunto de validação foi de 0.38, valor que diminuiu consistentemente indicando uma boa generalização. A priori, o percentual de classificações corretas em 85% representa um desempenho satisfatório na diferenciação dos gêneros em imagens dado o limitado número de amostras para aprendizagem. Na tentativa de prever o gênero de uma imagem do conjunto de testes, uma imagem do conjunto foi selecionada e submetida ao modelo para classificação do gênero. Assim, a classificação obtida do gênero foi correta.

O gênero predito da imagem é: Masculino

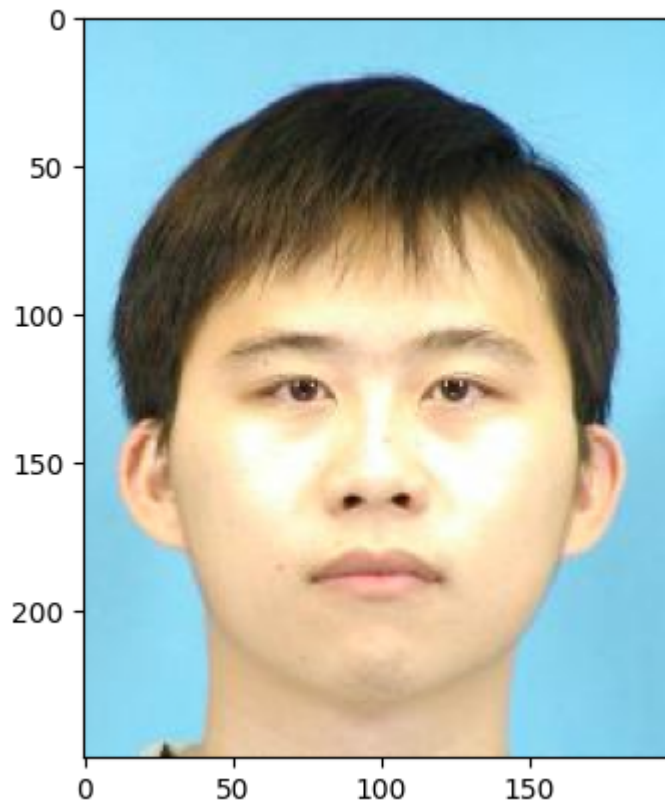


Figura 1 - Imagem do conjunto de testes usada para avaliar a predição realizada pelo modelo.

Para analisar o treinamento do modelo e seus desempenho durante o processo, é preciso observar o comportamento das curvas de acurácia do conjunto de treinamento e do conjunto de validação ao longo das épocas exposto na Figura 2. Também a curva das perdas do modelo no conjunto de treino e validação em função das épocas é apresentada na figura 3. Dessa maneira, o certo alinhamento entre as curvas indica uma boa performance do modelo. Entretanto, alguns espaços maiores entre as curvas indicam um possível overfitting, no qual o modelo pode se tornar cada vez mais especializado nos dados de treinamento. Enquanto a curva de acurácia do conjunto de treino cresce, a de validação decresce levemente em alguns momentos, o que sugere que o modelo possivelmente está se ajustando ao conjunto de treinamento.

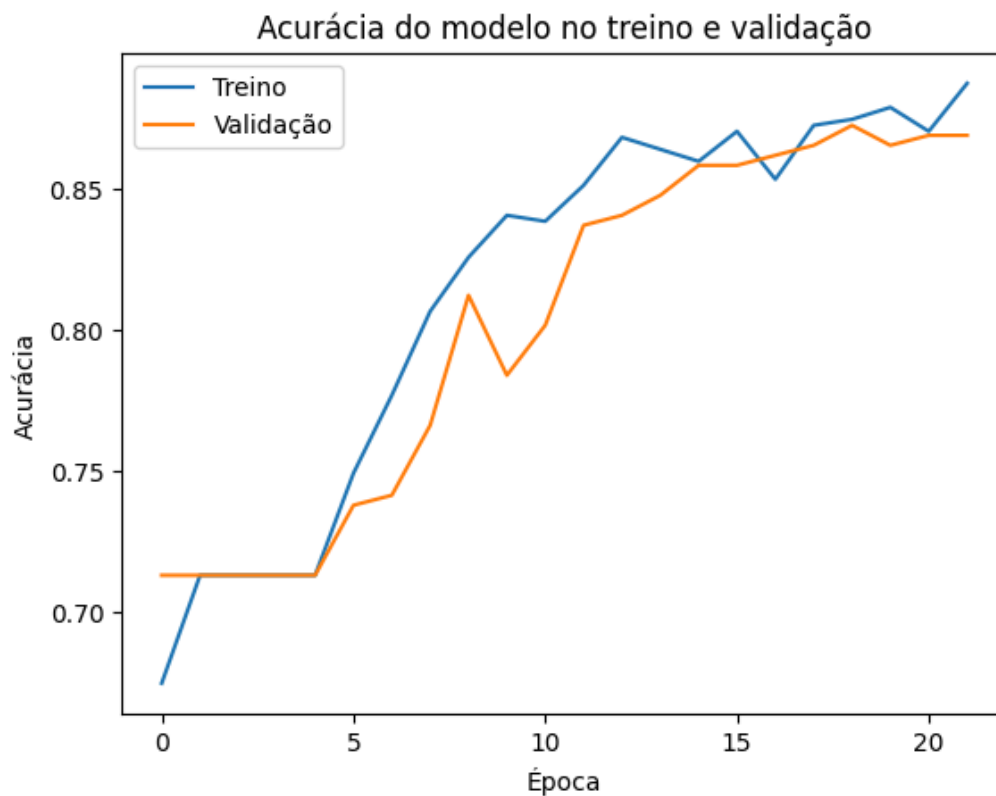


Figura 2 - Gráfico da curva de acurácia nos conjuntos de treino e validação em função das épocas

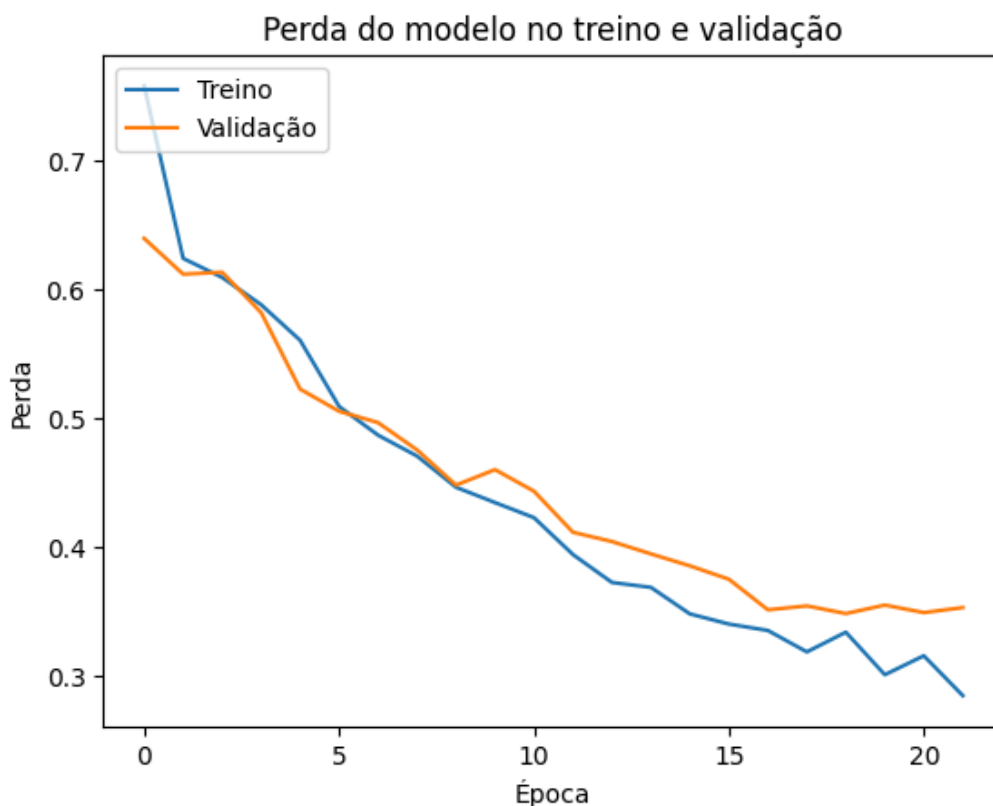


Figura 3 - Gráfico da curva de perdas nos conjuntos de treino e validação em função das épocas.

Para compreender e avaliar melhor o desempenho do modelo, é importante considerarmos demais métricas como F1-score, a curva ROC e a área sob a curva ROC (AUC-ROC). O valor de F1-score de 0.6824 é considerado moderado e aceitável, principalmente diante do desbalanceamento entre as classes expostas na Tabela 1. Dessa maneira, o modelo ainda enfrenta dificuldades em capturar os padrões das duas

classes de forma equilibrada enviesando a classificação para a classe masculina. Tal comportamento é corroborado pela matriz de confusão na Figura 4.

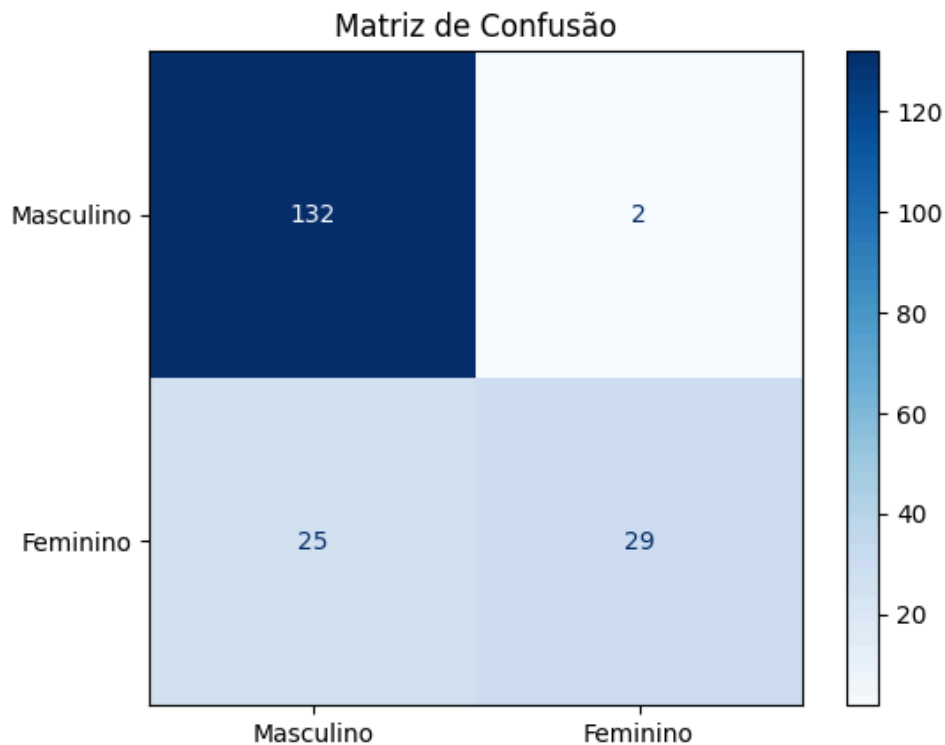


Figura 4 - Matriz de confusão das classes masculino e feminino no conjunto de testes.

Além disso, durante a análise das imagens classificadas incorretamente, notamos que o modelo teve maior dificuldade com imagens que apresentavam bruscas variações na luminosidade bem como falta de certa granulação ou baixa resolução reduziram a capacidade do modelo de capturar os padrões faciais necessários para a diferenciação entre as classes. Também, algumas imagens exibiam características andróginas ou pouco marcantes, dificultando a identificação.

Desse modo, o dataset apresentou desafio para a construção do modelo devido a pequena quantidade de imagens, apenas 188 inicialmente, o desequilíbrio nas proporções entre as classes (Masculino e feminino) e a baixa resolução das imagens

que podem ter influenciado nos resultados. Para melhorar o desempenho, técnicas mais avançadas de data augmentation, incluindo simulações mais realistas, como distorções geométricas e variações de cor, podem ajudar a melhorar a robustez do modelo. Também, o modelo pode ser redefinido adicionando mais camadas convolucionais e densas bem como explorando mais números de filtros e tamanhos de batch padrões mais complexos podem ser capturados, aumentando a taxa de aprendizagem do modelo. A aplicação da técnicas de regularização L2 poderia reduzir o risco de overfitting.

CONCLUSÃO E TRABALHOS FUTUROS

Neste projeto, exploramos o uso de redes neurais convolucionais (CNNs) para classificar imagens do dataset CUHK Face Sketch Database (CUFS) de acordo com o genero. A abordagem exigiu uma preparação cuidadosa dos dados, que incluiu normalização e técnicas de data augmentation para melhorar a performance do modelo.

O modelo alcançou uma acurácia de 85% e um F1-score de 0.6824, indicando um bom desempenho, mas com espaço para melhorias. Tais resultados foram influenciados por limitações e condições desafiadoras como dataset pequeno e com classes desbalanceadas, baixa resolução das imagens, características faciais que impactaram a capacidade de generalização do modelo e aumentaram a sensibilidade da rede a pequenos desvios nos padrões visuais.

Os resultados reforçam a importância de treinamento bem estruturado e destacam os desafios inerentes a conjuntos de dados pequenos e com possíveis limitações de qualidade. A implementação prática consolidou o entendimento sobre o papel de arquiteturas convolucionais no reconhecimento de padrões visuais, bem como a aplicação de métricas como F1-score, curva ROC e AUC para avaliação de desempenho. Em trabalhos futuros, deve-se considerar a expansão do dataset, correção do desbalanceamento entre as classes e melhorias na arquitetura do modelo, explorando arquitetura como a ResNet com maior capacidade de extração de características.

Referências: Liste todas as fontes consultadas.

REFERÊNCIAS

CUHK. *CUHK Face Sketch Database (CUFS)*. Disponível no material fornecido na trilha de aprendizado.

TENSORFLOW. TensorFlow Documentation. Disponível em: <https://www.tensorflow.org>. Acesso em: 28 nov. 2024.

KERAS. Keras Documentation. Disponível em: <https://keras.io>. Acesso em: 28 nov. 2024.

SCIKIT-LEARN. Scikit-learn Documentation. Disponível em: <https://scikit-learn.org>. Acesso em: 28 nov. 2024.

TOWARDS DATA SCIENCE. Artigos sobre CNNs e treinamento de modelos em datasets pequenos. Disponível em: <https://towardsdatascience.com>. Acesso em: 26 nov.. 2024.

Material didático e notebooks. *Exemplos e referências para manipulação de datasets e redes convolucionais*. Fornecidos no contexto da trilha de aprendizado.

