



Relatório Técnico:

Laureani Lima de Jesus

Matheus Santos da Silva

03 de dezembro de 2024

Sumário

Resumo.....	3
Introdução.....	4
Metodologia.....	6
Discussão.....	8
Conclusão e Trabalhos Futuros.....	10
Referências.....	12

Resumo

Este projeto explorou a aplicação de redes neurais convolucionais (CNNs) para a classificação de imagens no dataset CUHK Face Sketch Database (CUFS), categorizando rostos em duas classes: masculino e feminino. O trabalho abrangeu todas as etapas do pipeline de aprendizado profundo, incluindo anotação manual das imagens, pré-processamento para normalização dos valores RGB e ajuste de dimensões (250x200 pixels), além da divisão dos dados em conjuntos de treino (50%), validação (30%) e teste (20%), utilizando uma semente fixa para replicabilidade.

A arquitetura do modelo CNN foi projetada para extrair características visuais relevantes utilizando camadas convolucionais, max pooling e camadas densas para classificação, enquanto técnicas de regularização como dropout foram aplicadas para prevenir overfitting. O treinamento foi realizado com a função de perda de entropia cruzada binária e o otimizador Adam, monitorando métricas como acurácia e F1-Score.

Os resultados indicaram que o modelo foi capaz de aprender padrões discriminantes, com um desempenho satisfatório medido por métricas como F1-Score, curva ROC e AUC-ROC. Entretanto, desafios como variações de iluminação, ângulos complexos e detalhes faciais sutis afetaram a precisão em algumas imagens. A análise qualitativa das classificações incorretas evidenciou a necessidade de melhorias no dataset e na arquitetura do modelo, como o uso de técnicas de aumento de dados e ajustes nos filtros convolucionais.

Este projeto destaca a eficiência de redes convolucionais em tarefas desafiadoras, mesmo com datasets limitados, e oferece aprendizados valiosos para aprimoramentos futuros. As conclusões fornecem uma base sólida para avançar na aplicação de aprendizado profundo em problemas de visão computacional, com implicações práticas e teóricas para a área.

Introdução

As redes neurais convolucionais (CNNs) transformaram a abordagem de problemas complexos em visão computacional, especialmente na classificação de imagens. Sua capacidade de extrair automaticamente características visuais, como bordas, texturas e padrões, as torna particularmente eficazes em tarefas que envolvem reconhecimento e diferenciação de classes. Neste projeto, o foco é a aplicação de um modelo CNN autoral para a classificação de imagens do dataset CUHK Face Sketch Database (CUFS), categorizando-as em duas classes: masculino e feminino, com base em traços faciais.

Originalmente, o dataset CUFS foi concebido para explorar a relação entre fotografias e esboços artísticos de rostos. No entanto, sua utilização neste projeto para a classificação de sexo biológico ilustra a versatilidade das redes convolucionais e do próprio dataset. A tarefa exige um modelo robusto, capaz de aprender características discriminantes em um conjunto limitado de dados, destacando desafios como o desbalanceamento de classes, variações de iluminação e ruído presente nas imagens.

O desenvolvimento deste projeto não se limita à criação e ao treinamento de um modelo CNN. Ele abrange etapas fundamentais do pipeline de aprendizado profundo, como a anotação manual dos dados, o pré-processamento das imagens para normalização e ajuste de dimensões, a divisão do dataset em conjuntos de treino, validação e teste, e, finalmente, a avaliação criteriosa do desempenho do modelo. A análise dos resultados será conduzida por métricas como F1-Score, curva ROC e AUC-ROC, além da identificação de erros de classificação e possíveis limitações do modelo ou do dataset.

Este projeto tem como objetivo não apenas criar uma solução técnica, mas também oferecer insights sobre os desafios e as possibilidades da aplicação de redes convolucionais em

problemas reais. Ao final, as conclusões extraídas e as sugestões de melhorias contribuirão para ampliar o entendimento sobre o uso de aprendizado profundo na classificação de imagens.

Metodologia

Este projeto foi desenvolvido com o objetivo de implementar um modelo baseado em redes neurais convolucionais (CNNs) para a classificação de imagens do dataset CUHK Face Sketch Database (CUFS), categorizando-as em classes de sexo biológico: masculino e feminino. Inicialmente, o conjunto de dados, composto por 188 imagens localizadas na pasta **photos**, foi preparado para treinamento. Essa preparação incluiu a anotação manual das imagens, atribuindo rótulos binários às classes, sendo **0** para masculino e **1** para feminino. Esta etapa demandou uma análise cuidadosa para assegurar a consistência e a precisão dos rótulos.

Após a anotação, realizou-se o pré-processamento das imagens, ajustando todas para uma dimensão uniforme de 250x200 pixels (altura x largura). Além disso, os valores RGB das imagens foram normalizados para o intervalo $[0, 1]$, dividindo cada valor de pixel por 255, a fim de melhorar a convergência do modelo durante o treinamento. Com as imagens preparadas, o conjunto de dados foi dividido em três subconjuntos: 50% para treinamento, 30% para validação e 20% para teste. Para garantir a replicabilidade da divisão, foi utilizada a semente fixa **23**. O conjunto de treinamento foi utilizado para ajustar os pesos do modelo, enquanto o conjunto de validação monitorou o desempenho durante o treinamento, e o conjunto de teste foi reservado exclusivamente para a avaliação final do modelo.

A construção do modelo CNN foi baseada em uma arquitetura autoral, projetada para explorar características relevantes nas imagens. Camadas convolucionais foram utilizadas para extrair padrões visuais, com diferentes tamanhos de filtro e a função de ativação ReLU para introduzir não linearidade. Em seguida, camadas de pooling max foram aplicadas para reduzir a dimensionalidade das características extraídas, preservando os padrões mais importantes. O modelo também incluiu camadas totalmente conectadas para realizar a classificação final,

associando as características aprendidas às saídas binárias. Para evitar sobreajuste, técnicas como Dropout e normalização batch foram empregadas, assegurando maior generalização do modelo.

Durante o treinamento, o modelo foi ajustado utilizando a função de perda de entropia cruzada binária, adequada para problemas de classificação binária. O otimizador Adam foi escolhido devido à sua eficiência no ajuste dos pesos em tarefas de aprendizado profundo. O treinamento foi conduzido em várias épocas, com early stopping implementado para interromper o processo caso o desempenho no conjunto de validação não apresentasse melhorias após algumas iterações consecutivas. Durante esse processo, métricas como acurácia e F1-Score foram monitoradas para avaliar a performance do modelo.

Após o treinamento, a avaliação final foi realizada no conjunto de teste. Métricas como F1-Score, curva ROC e área sob a curva ROC (AUC-ROC) foram utilizadas para medir o desempenho do modelo na distinção entre as classes. Além disso, as imagens classificadas incorretamente foram analisadas para identificar possíveis padrões nos erros cometidos, considerando fatores como ruído, variações de iluminação, limitações do dataset ou características específicas que o modelo não conseguiu capturar adequadamente.

Todo o projeto foi implementado em Python, utilizando bibliotecas como TensorFlow e Keras para a construção e treinamento do modelo, além de Matplotlib e Seaborn para a visualização dos resultados. Através dessa abordagem sistemática, o projeto buscou não apenas desenvolver um modelo eficiente, mas também compreender as limitações e os desafios associados à aplicação de redes convolucionais em tarefas reais de classificação de imagens.



Discussão

Após a conclusão do treinamento e da avaliação do modelo CNN para classificação de imagens baseado no dataset CUFS, alguns insights importantes foram obtidos. Em termos de desempenho, o F1-score apresentou resultados satisfatórios, indicando um bom equilíbrio entre precisão e sensibilidade. Essa métrica foi especialmente útil para avaliar o modelo, dado o potencial desbalanceamento nas classes masculino e feminino. No entanto, foi identificado que em algumas instâncias específicas, a predição falhou, o que sugere que o modelo teve dificuldades com certas características das imagens.

Entre os fatores que podem ter influenciado as classificações incorretas, destaca-se a qualidade do dataset. Imagens com variações de iluminação, ângulos extremos, ou presença de ruídos visuais foram as mais desafiadoras para o modelo. Isso pode ser atribuído ao tamanho limitado do dataset e à ausência de dados suplementares para treinar o modelo com uma maior diversidade de exemplos. Além disso, a homogeneidade nas representações faciais em termos de estilo e pose também pode ter impactado negativamente, restringindo a capacidade do modelo de generalizar para variações menos comuns.

Uma análise detalhada das imagens mal classificadas revelou padrões comuns. Imagens com forte sombra facial ou detalhes sutis nos contornos faciais confundiram o modelo, possivelmente porque essas características não foram bem capturadas pelos filtros convolucionais iniciais. Isso sugere que ajustes na arquitetura, como o uso de mais camadas convolucionais ou filtros com tamanhos diferentes, poderiam melhorar a capacidade do modelo de capturar essas nuances.

Adicionalmente, o uso de técnicas de aumento de dados (data augmentation), como rotações, ajustes de brilho e adição de ruído, poderia fortalecer a robustez do modelo frente a variações naturais das imagens. Outra possibilidade seria incluir camadas adicionais de regularização, como dropout, para evitar overfitting, especialmente considerando o tamanho limitado do dataset.

Por fim, o projeto trouxe algumas surpresas. Apesar da simplicidade relativa da arquitetura proposta, o modelo conseguiu aprender padrões significativos e, em muitos casos, realizou previsões confiáveis. Esse resultado destaca a eficácia das CNNs mesmo em cenários com datasets restritos, desde que o pré-processamento seja rigoroso e a abordagem metodológica seja consistente.

A experiência geral reforça a importância de um planejamento meticuloso no preparo do dataset e no desenho da arquitetura do modelo. Em projetos futuros, seria interessante explorar datasets maiores ou combinados com outras bases de dados, bem como avaliar o impacto do uso de diferentes funções de ativação ou estratégias de otimização.

Conclusão e Trabalhos Futuros

Este projeto demonstrou a aplicabilidade de redes neurais convolucionais (CNNs) no desafio de classificação de imagens, utilizando o dataset CUFS para diferenciar rostos masculinos e femininos. Apesar das limitações relacionadas ao tamanho e à qualidade do dataset, o modelo apresentou desempenho satisfatório, com métricas como o F1-score indicando um equilíbrio razoável entre precisão e sensibilidade. No entanto, a análise revelou que imagens com variações de iluminação, ângulos desafiadores e detalhes faciais sutis representaram um obstáculo significativo, reforçando a necessidade de uma abordagem cuidadosa no preparo e na diversificação dos dados.

Com base nos desafios enfrentados, recomenda-se, em trabalhos futuros, ampliar o dataset para incluir maior diversidade de características faciais, iluminação e poses, o que aumentaria a capacidade do modelo de generalizar para novos cenários. Técnicas de aumento de dados, como rotações, ajustes de brilho e adição de ruídos, também podem ser exploradas para melhorar a robustez do modelo. Além disso, revisões na arquitetura da CNN, como a inclusão de mais camadas convolucionais, diferentes tamanhos de filtros e funções de ativação, poderiam contribuir para capturar padrões mais complexos nas imagens. Estratégias de regularização, como dropout e batch normalization, bem como testes com diferentes algoritmos de otimização, também são caminhos promissores para melhorar o desempenho e evitar o overfitting.

A análise crítica dos erros de classificação evidenciou a importância de compreender as limitações do modelo e do dataset utilizado. Características como iluminação irregular e detalhes faciais complexos demonstraram ser áreas em que o modelo encontrou maior dificuldade, indicando que melhorias no processo de coleta de dados e na arquitetura poderiam mitigar esses problemas. Em suma, este trabalho ofereceu uma experiência enriquecedora na aplicação de aprendizado profundo para problemas de visão computacional, servindo como base para projetos

futuros mais ambiciosos e impactantes, com potencial para avançar o entendimento e a eficiência dessas tecnologias em diferentes contextos.

Referências

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning**. *Nature*, v. 521, p. 436–444, 2015. CUHK Face Sketch Database (CUFS). Disponível em: <https://www.kaggle.com/datasets/arbazkhan971/cuhk-face-sketch-database-cufs>. Acesso em: 3 dez. 2024.

ZHANG, Z.; LIN, Y.; SUN, H. **Face sketch synthesis using deep neural networks**. *IEEE Transactions on Image Processing*, v. 26, n. 1, p. 188–202, 2017.