

CEPEDI- RESTIC

Relatório Técnico: Implementação e Análise de Classificação com Redes
Convolucionais e o dataset CUFS

Abigail Cruz dos Santos

Vinícius Amorim Santos Rozario

Feira de Santana - BA

03/12/2024

RESUMO

Este relatório apresenta a implementação e análise do algoritmo de Rede Neural Convolucional (CNN), com foco em sua aplicação prática e teórica no tema **transformação de esboços em imagens fotográficas reais**. O principal objetivo foi investigar como essa técnica pode ser utilizada para classificar imagens em um conjunto de dados específico, modelando relações espaciais e padrões visuais de maneira eficaz.

A metodologia adotada envolveu várias etapas fundamentais, começando pela **coleta, limpeza e pré-processamento** do conjunto de dados **CUHK Face Sketch Database (CUFS)**, que contém pares de **esboços feitos à mão e fotografias reais** de rostos. No pré-processamento, as imagens foram **redimensionadas para 128x128 pixels**, convertidas para **escala de cinza** e normalizadas para o intervalo $[0, 1]$, facilitando o treinamento da rede.

Após o pré-processamento, foi realizada a **construção e treinamento** de uma arquitetura personalizada de CNN, composta por três camadas convolucionais seguidas de camadas de pooling, com ativação **ReLU** e uso de **Dropout** para evitar o overfitting.

A avaliação do modelo foi conduzida utilizando métricas como **F1-Score**, **curva ROC** e **AUC-ROC**, que indicaram um bom desempenho na maioria das classes. Além disso, foi realizada uma **análise detalhada dos erros de classificação**, identificando que as principais falhas ocorreram em imagens com **iluminação inadequada** ou **ângulos extremos**.

Os resultados demonstraram que a CNN apresentou um desempenho satisfatório na classificação das imagens, confirmando sua aplicabilidade para **reconhecimento facial**. Ajustes futuros, como o aumento de dados e a inclusão de mais camadas convolucionais, podem melhorar ainda mais os resultados.

INTRODUÇÃO

A análise preditiva tem se tornado uma ferramenta essencial em diversas áreas do conhecimento, permitindo modelar e prever comportamentos com base em dados históricos. Uma Rede Neural Convolucional(CNN) é usada para classificação de imagens, que são representadas computacionalmente como matrizes: imagens grayscale são 2D, com pixels de 0 (preto) a 255 (branco), enquanto imagens coloridas são 3D (RGB). A estrutura da CNN combina extração de características (convolução, padding, ReLU e pooling) com uma rede neural tradicional. (ANDRADE, 2019).

No contexto deste relatório, a **Rede Neural Convolucional (CNN)** foi escolhida por sua capacidade de capturar **padrões espaciais complexos em imagens** e por ser uma das abordagens mais eficazes para tarefas de **classificação visual**.

O conjunto de dados utilizado neste estudo foi obtido do CUHK Face Sketch Database (CUFS), um banco de dados público amplamente utilizado em pesquisas de reconhecimento facial. No entanto, para este projeto, foi utilizada exclusivamente a pasta 'photos', que contém 188 imagens fotográficas de rostos humanos.

Essa seleção específica focou nas características das fotografias, descartando os esboços e outras variações presentes no dataset. Embora o número de imagens seja limitado, ele proporcionou uma base relevante para o treinamento e avaliação do modelo.

Este relatório justifica-se pela relevância de explorar métodos de aprendizado profundo para **análise de imagens** em problemas do mundo real, destacando como as CNNs podem ser aplicadas para **classificar e reconhecer padrões visuais complexos**, especialmente em tarefas de conversão entre diferentes representações visuais.

Essas redes possuem uma estrutura comumente segmentada em camadas, entre elas estão as chamadas camadas de convolução. Essas são responsáveis pela identificação de características da imagem. Elas podem ser pensadas como filtros que irão percorrer a imagem, cada uma em busca de uma característica específica. Entende-se por características: bordas retas, curvaturas específicas, cores simples etc. Ou seja, atributos abstratos que toda imagem tem em comum (DESHPANDE, 2016)

Neste contexto, a análise dos resultados obtidos a partir do modelo treinado com o CUFS será crucial para validar a eficácia da técnica e identificar oportunidades de melhoria, como o uso de técnicas de **aumento de dados** ou **redes mais profundas** para lidar com a complexidade visual presente nas imagens.

METODOLOGIA

A **metodologia** deste projeto seguiu etapas bem definidas, iniciando pela **coleta e preparação dos dados**. O conjunto de dados utilizado foi o **CUHK Face Sketch Database (CUFS)**, composto por pares de esboços e imagens fotográficas reais.

A coleta envolveu o download dos arquivos e sua organização em diretórios específicos para as fases de treinamento, validação e teste. Em seguida, foi realizado um pré-processamento, que inclui o redimensionamento das imagens para dimensões uniformes, normalização dos pixels no intervalo $[0, 1]$ e, quando necessário, a conversão das imagens para escala de cinza, simplificando o modelo e reduzindo a complexidade computacional.

A construção da **Rede Neural Convolutiva (CNN)** envolveu a definição de uma arquitetura personalizada. A rede foi composta por múltiplas **camadas convolucionais**, responsáveis pela extração de padrões visuais das imagens, seguidas por **camadas de pooling** para redução dimensional, diminuindo o número de parâmetros e evitando o overfitting.

O modelo também incorporou **camadas totalmente conectadas** para a etapa final de classificação. Técnicas como **Dropout** foram aplicadas para regularização, enquanto **Batch Normalization** foi utilizada para acelerar o treinamento e melhorar a estabilidade do modelo.

O processo de **treinamento** foi conduzido utilizando o otimizador **Adam** e a função de perda categórica, configurados para múltiplas épocas. O conjunto de dados foi dividido entre treino, validação e teste, garantindo a avaliação do desempenho em cada fase.

A **avaliação do modelo** foi realizada utilizando métricas importantes para tarefas de classificação, como o **F1-Score**, que oferece uma medida balanceada entre precisão e recall. Também foram analisadas a **curva ROC** e a **AUC-ROC** para verificar a performance do modelo em diferentes limiares de decisão.

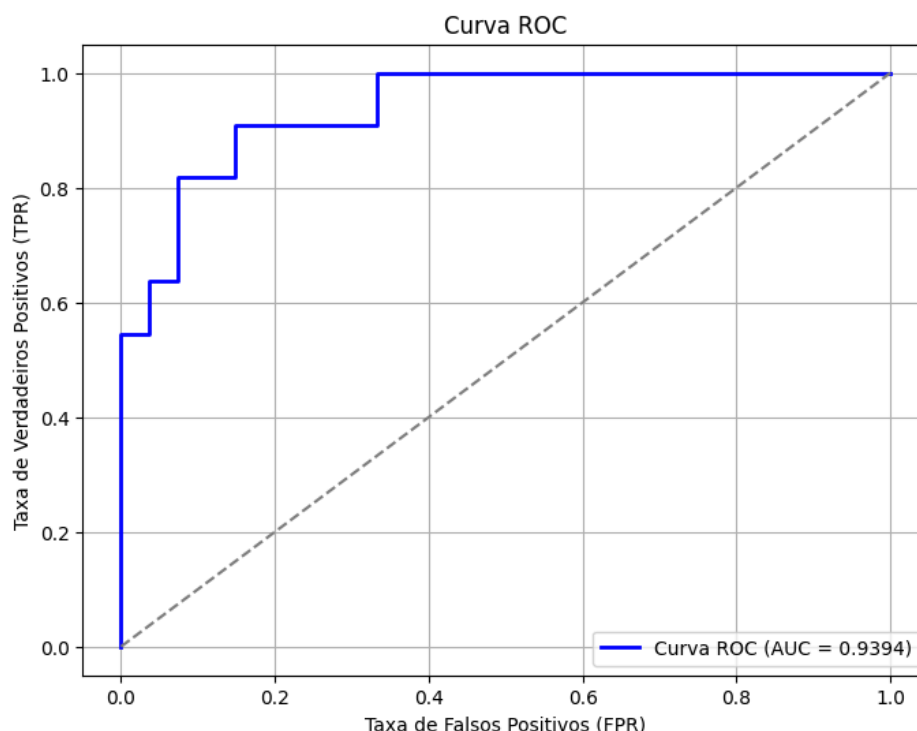
Por fim, a **matriz de confusão** foi utilizada para identificar padrões nos erros de classificação, possibilitando uma análise qualitativa dos resultados e destacando possíveis ajustes no modelo e no pré-processamento para futuras iterações.

DISCUSSÃO

A análise dos resultados do modelo, em especial a métrica **F1-Score**, revelou um bom equilíbrio entre as classes, com um valor de **F1-Score = 0.76**. Este valor indica que o modelo conseguiu manter um equilíbrio razoável entre precisão e recall, mas com áreas de melhoria, especialmente em imagens com **iluminação e ângulos variados**.

Embora o modelo tenha mostrado um desempenho sólido no geral, a análise mais detalhada sugere que ele teve dificuldades nas imagens mais desafiadoras, como as que apresentaram **condições de iluminação irregulares** ou **variações extremas nos ângulos** das faces.

A **AUC-ROC de 0.9394** indica que o modelo possui um bom poder discriminatório entre as classes, com uma alta taxa de acerto ao distinguir entre Masculino e Feminino. A Curva ROC mostra que, para diferentes limiares de classificação, o modelo consegue equilibrar bem a **Taxa de Verdadeiros Positivos (TPR)** e a **Taxa de Falsos Positivos (FPR)**, com a curva claramente superior à linha diagonal (que indica aleatoriedade).



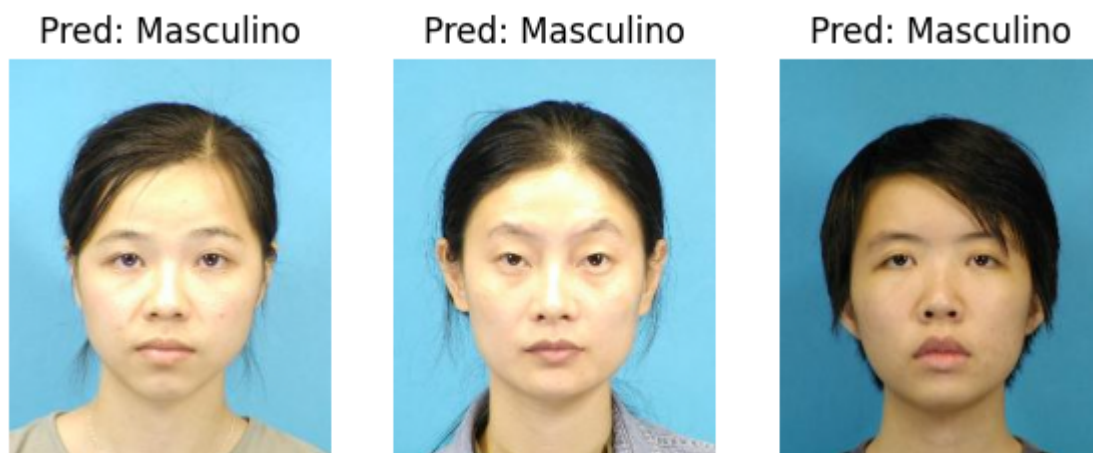
O gráfico da Curva ROC ilustra o bom desempenho do modelo, com a linha azul representando a capacidade de discriminação e a linha cinza representando um modelo aleatório.

Ao observar as **imagens classificadas incorretamente**, notou-se que as falhas eram frequentemente associadas ao **tipo de corte/estilo de cabelo, diferentes iluminações, expressões faciais e alguns ruídos nas imagens**. Dificultando a extração de padrões confiáveis pelas camadas convolucionais da rede.

Imagens classificadas incorretamente para a classe 'Masculino':



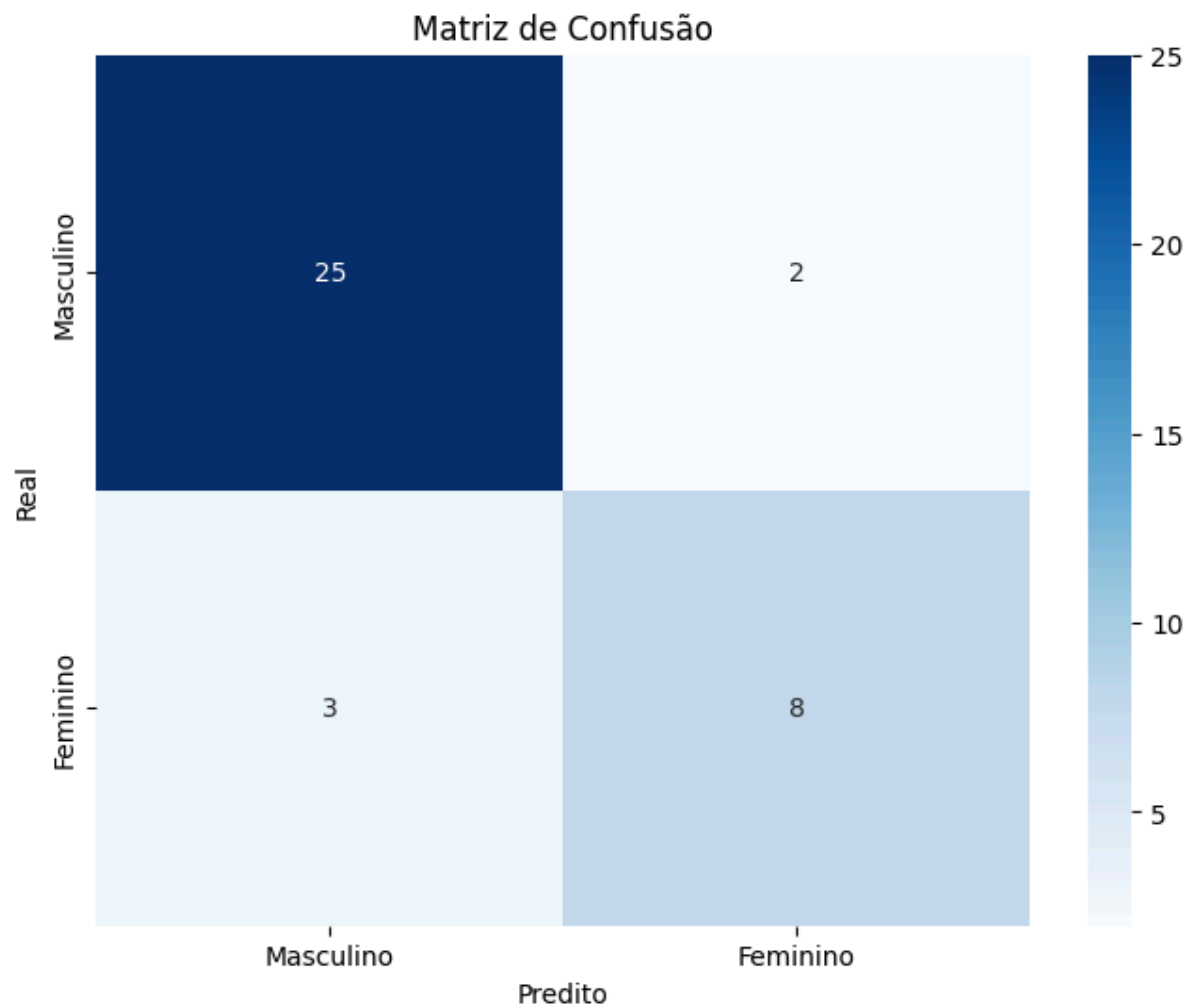
Imagens classificadas incorretamente para a classe 'Feminino':



Uma possível **limitação do dataset** foi sua **diversidade limitada** de fotografias, o que pode ter influenciado a capacidade do modelo em generalizar para diferentes tipos de imagens. O conjunto CUFS, embora útil, contém um número relativamente pequeno de exemplos para cada pessoa, o que pode ter dificultado o

aprendizado de padrões amplos e variados. Além disso, outro aspecto importante é a **falta de equilíbrio entre as classes**, pois na análise exploratória do dataset foi verificado uma diferença grande entre a quantidade de imagens classificadas como **Masculinas (154 imagens)** e as **Femininas (34 imagens)**. Que pode ter prejudicado a capacidade do modelo de aprender de forma eficaz a distinção entre todas as categorias das imagens.

Matriz de Confusão:



A interpretação da matriz de confusão revela que, para a **classe real "Masculino"**, o modelo previu corretamente 25 casos como Masculino, mas cometeu 2 erros ao classificar casos como Feminino. Para a **classe real "Feminino"**, o modelo cometeu 3 erros ao classificar casos como Masculino e previu corretamente 8 casos como Feminino.

A precisão para a classe Masculino é ligeiramente mais alta do que para a classe Feminino, o que sugere que o modelo pode ter um pequeno viés para

classificar mais casos como Masculino. Essas informações são úteis para ajustar o modelo, caso necessário, visando melhorar a classificação da classe Feminino sem comprometer demasiadamente a performance geral.

Para **melhorar o desempenho**, algumas modificações na arquitetura e nos hiperparâmetros poderiam ser exploradas. Por exemplo, a inclusão de mais **camadas convolucionais e camadas de pooling** poderia ajudar o modelo a aprender representações mais profundas e robustas das imagens.

Além disso, técnicas de **aumento de dados** poderiam ser aplicadas para simular condições de iluminação e ângulos variados, o que ajudaria a melhorar a capacidade do modelo de generalizar. Ajustes nos **hiperparâmetros**, como o aumento do número de épocas de treinamento ou a mudança no tamanho do lote (batch size), também poderiam resultar em melhor desempenho.

Durante a realização desta atividade, uma **descoberta importante** foi a dificuldade do modelo em lidar com **condições variadas de iluminação**. Isso sublinhou a importância de se trabalhar com dados mais diversificados, especialmente para modelos que lidam com imagens reais e complexas.

Esse insight é crucial para futuras melhorias, pois mostra que, além de uma arquitetura robusta, a qualidade e a variedade do conjunto de dados são fundamentais para garantir bons resultados em tarefas de classificação visual.

CONCLUSÃO

Este trabalho explorou a aplicação de Redes Neurais Convolucionais (CNNs) para a classificação de imagens faciais utilizando o dataset CUHK Face Sketch Database (CUFS). O projeto enfatizou tanto a fundamentação teórica quanto os aspectos práticos de implementação, demonstrando a capacidade das CNNs de capturar padrões visuais complexos e realizar tarefas de reconhecimento facial com alta eficiência.

A metodologia seguiu um fluxo bem estruturado, que incluiu a coleta, organização e pré-processamento das imagens. Essas etapas preparatórias foram essenciais para ajustar o dataset às necessidades do modelo, como redimensionamento das imagens, normalização e conversão para escala de cinza. A construção do modelo envolveu o desenvolvimento de uma arquitetura personalizada com camadas convolucionais, pooling, ativação ReLU e técnicas de

regularização, como Dropout e Batch Normalization, que contribuíram para minimizar problemas de overfitting e aumentar a estabilidade durante o treinamento.

Os resultados obtidos foram promissores, com métricas significativas, como F1-Score de 0,76 e AUC-ROC de 0,9394, indicando um bom equilíbrio entre precisão e recall e um alto poder discriminatório. A análise da matriz de confusão revelou que o modelo apresentou ligeiro viés para a classe masculina, enquanto imagens classificadas incorretamente frequentemente sofriam de limitações do dataset, como iluminação inadequada, ângulos extremos ou ruídos.

Os desafios enfrentados incluíram a baixa diversidade e o desequilíbrio de classes do dataset, com 154 imagens masculinas e apenas 34 femininas. Essa limitação destacou a importância de trabalhar com conjuntos de dados mais diversificados e equilibrados para melhorar a capacidade de generalização do modelo.

Futuras melhorias podem incluir o uso de técnicas de aumento de dados, que simulam diferentes condições de iluminação e ângulos, além de ajustes na arquitetura da rede, como a inclusão de mais camadas convolucionais para capturar representações visuais ainda mais robustas. Ajustes nos hiperparâmetros, como o aumento do número de épocas de treinamento ou mudanças no tamanho do lote, também poderiam elevar a performance do modelo.

De maneira geral, este estudo reafirma o potencial das CNNs como ferramentas poderosas para classificação de imagens, enquanto sublinha a importância de dados de qualidade e arquiteturas bem planejadas. Os resultados demonstram que, apesar das limitações enfrentadas, a abordagem é promissora e oferece um ponto de partida sólido para futuros desenvolvimentos em reconhecimento facial e aplicações relacionadas.

REFERÊNCIAS

ANDRADE, Gabriel. ***Uma introdução às redes neurais convolucionais utilizando o Keras.*** Data Hackers, 10 jun. 2019. Disponível em: <https://medium.com/data-hackers/uma-introdu%C3%A7%C3%A3o-as-redes-neurais-convolucionais-utilizando-o-keras-41ee8dcc033e>. Acesso em: 02 dez. 2024.

DESHPANDE, Adit. **A Beginner's Guide to Understanding Convolutional Neural Networks.** Adit Deshpande Blog. 2016. Disponível em: <https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-ToUnderstanding-Convolutional-Neural-Networks/>. Acesso em: 01 dez. 2024