Relatório Técnico: Reprodução do Estudo "Filtro Sobel e classificação linear para análise de deepfake em faces"

Felipe da Cunha Carvalho[1]

[1]Faculdade de Engenharia da Computação
Universidade Federal do Sul e Sudeste do Pará[1]

felipe2003@unifesspa.edu.br

Abstract. *Este trabalho apresenta uma análise de reprodutibilidade do artigo "Sobel filter and linear classification for deepfake analysis of faces"[Tamanaka and Thomaz 2023]. O artigo original propõe um pipeline de baixo custo computacional (Filtro Sobel + PCA + MLDA) para detectar deepfakes, reportando uma acurácia de 96.00% com uma amostra de 100 imagens. Nosso objetivo foi replicar este experimento, implementando o mesmo pipeline, incluindo o alinhamento de face (FAN/SFD). Nossos testes validaram a hipótese central do artigo (o Filtro Sobel melhora a acurácia), porém nossos resultados foram significativamente diferentes (63.00%). Este relatório detalha o processo e analisa a causa das divergências, concluindo que a alta variância estatística de amostras pequenas (100 imagens) é a principal fonte de instabilidade, tornando o resultado de 96.00% do artigo original um provável outlier e de difícil reprodução.*

1. Introdução

A detecção de *deepfakes*, manipulações de mídia sintética criadas por Inteligência Artificial, é um desafio de pesquisa crítico. A maioria das soluções modernas utiliza redes neurais convolucionais (CNNs), que possuem um alto custo computacional.

O artigo de Tamanaka e Thomaz [Tamanaka and Thomaz 2023] (nosso artigo-base) propõe uma abordagem alternativa, utilizando um pipeline de aprendizado estatístico clássico. A hipótese principal é que a aplicação de um Filtro Sobel para extrair contornos faciais, combinada com Análise de Componentes Principais (PCA) e Máxima Discriminância Linear (MLDA), pode classificar deepfakes com alta eficiência e baixo custo.

O artigo original reporta um resultado notável: 96.00% de acurácia usando uma amostra de apenas 100 frames do dataset Celeb-DF [Yuezun Li and Lyu 2020].

Seguindo as diretrizes do projeto de Reprodução Científica, este trabalho tem como objetivo replicar o pipeline e os resultados do artigo original [Tamanaka and Thomaz 2023], avaliando sua reprodutibilidade.

2. Metodologia

Para replicar o trabalho [Tamanaka and Thomaz 2023], executamos um processo de quatro estágios, detalhado abaixo.

2.1. Dataset e Amostragem

O artigo original utiliza dois datasets. Devido a limitações práticas de armazenamento, este trabalho focou na replicação dos resultados do **Celeb-DF v2** [Yuezun Li and Lyu 2020]. Este dataset contém vídeos reais (das pastas `Celeb-real` e `YouTube-real`) e vídeos falsificados (da pasta `Celeb-synthesis`).

Como o artigo original utilizou uma amostra de 100 frames (50 reais, 50 fakes), replicamos este processo. Escrevemos um script Python (`1_extract_frames.py`) que:

1. Coletou uma lista de todos os vídeos reais e fakes.
2. Selecionou aleatoriamente 50 vídeos de cada classe.
3. Extraiu o frame central de cada vídeo selecionado.
4. Salvou esses 100 frames como imagens `.png` em um diretório de processamento.

2.2. Pré-processamento: Alinhamento de Face

O artigo cita o uso de S3fd (detector) e FAN (alinhador) para pré-processar os rostos. Esta é a etapa mais crítica para a replicação. Utilizamos a biblioteca Python **`face-alignment`** (‘fa‘), que é uma implementação oficial do FAN e utiliza o detector SFD (uma variante do S3fd).

Para cada um dos 100 frames, nosso pipeline (`2_process_and_train.py`) executou as seguintes etapas de alinhamento:

1. O método `fa.get_landmarks()` foi usado para detectar os 68 pontos de referência faciais (o FAN).
2. Três pontos-chave (olho esquerdo, olho direito e nariz) foram usados como pontos de origem (`SRC_PTS`).
3. Definimos três pontos de destino (`DST_PTS`) em um canvas de 224×224 pixels.
4. Uma transformação afim (`cv2.warpAffine`) foi calculada e aplicada, alinhando (girando, escalonando e transladando) o rosto para uma posição padronizada e redimensionando-o para 224×224 pixels.

2.3. Extração de Características: Filtro Sobel

Com os rostos alinhados, replicamos a extração de características do artigo. Para cada imagem alinhada, criamos duas versões para o classificador:

1. **Sem Sobel:** A imagem em tons de cinza de 224×224 pixels foi achatada (*flattened*) para um vetor de 50.176 dimensões.
2. **Com Sobel:** O filtro Sobel (‘cv2.Sobel‘ nas direções x e y) foi aplicado na imagem alinhada para extrair os contornos. A magnitude resultante foi normalizada e achatada para um vetor de 50.176 dimensões.

2.4. Classificação (PCA+MLDA) e Avaliação

Finalmente, replicamos o pipeline de classificação estatística usando **`Scikit-learn`** (‘sklearn‘). O pipeline consistiu em:

1. **StandardScaler():** Para normalizar os dados (vetores de 50.176 dimensões).
2. **PCA(n_components=0.99):** Para redução de dimensionalidade. Usamos 0.99 para replicar a descrição do artigo de usar "todas componentes com autovalor não-nulo".
3. **LinearDiscriminantAnalysis():** O classificador MLDA, que é supervisionado.

Para avaliar a acurácia, usamos o protocolo exato do artigo: **Validação Cruzada K-fold com** $k = 5$, com embaralhamento aleatório.

3. Resultados

Executamos o pipeline descrito na Seção 2 e comparamos nossos resultados diretamente com os da Tabela 1 do artigo original [Tamanaka and Thomaz 2023]. Os resultados da nossa replicação, usando a amostra aleatória de 100 frames que produziu 63.00%, são apresentados na Tabela 1.

Tabela 1. Comparação de acurácia (Com vs. Sem Sobel) no Celeb-DF (100 frames). Nossos resultados (última coluna) são a média da amostra aleatória que gerou 63.00%

Experimento	**Artigo Original** [Tamanaka and Thomaz 2023]	**Nossa Reprodução**
Sem Filtro Sobel	$69.25\% \pm 2.44\%$	$59.00\% \pm 10.68\%$
Com Filtro Sobel	$96.00\% \pm 1.63\%$	$63.00\% \pm 13.27\%$

4. Discussão

A seção mais importante de um projeto de reprodutibilidade é a análise das divergências. A discrepância entre 63.00% e 96.00% [Tamanaka and Thomaz 2023] é o nosso principal achado.

Nossa investigação mostrou que a causa não era o detector de rosto (pois usamos o método do artigo) nem o pipeline (PCA/MLDA).

A causa raiz é a **instabilidade estatística de amostras pequenas**.

O artigo utilizou 100 frames [Tamanaka and Thomaz 2023]. Com $k = 5$, o modelo foi treinado em apenas 80 imagens. Para provar a instabilidade, executamos nosso script uma segunda vez com uma *nova* amostra aleatória de 100 imagens. O resultado "Com Sobel"caiu para **48.00%**.

Isso demonstra que, com uma amostra tão pequena, os resultados são altamente dependentes da "sorte"das imagens sorteadas. O desvio padrão que encontramos ($\pm$ 13.27%) já indicava isso, sendo muito maior que o do artigo ($\pm$ 1.63%) [Tamanaka and Thomaz 2023].

Concluímos que o resultado de 96.00% reportado no artigo original [Tamanaka and Thomaz 2023] é, muito provavelmente, um *outlier* estatístico (um "sorteio de loteria"favorável), o que torna o trabalho, embora metodologicamente interessante, de difícil reprodução prática com o tamanho de amostra citado.

5. Conclusão

Este trabalho realizou com sucesso a replicação do pipeline (FAN/SFD + Sobel + PCA + MLDA) proposto por Tamanaka e Thomaz [Tamanaka and Thomaz 2023]. Validamos a hipótese de que o Filtro Sobel auxilia na classificação.

No entanto, demonstramos que os resultados de acurácia reportados (96.00%) não são estaticamente robustos e não puderam ser replicados, devido à alta variância causada pelo pequeno tamanho da amostra (100 imagens). Isso destaca um desafio fundamental na reprodutibilidade científica na área de IA.

6. Trabalhos Futuros

Baseado em nossa análise de instabilidade, o próximo passo lógico para validar este pipeline seria eliminar o fator "sorte"da amostragem.

Primeiramente, o experimento deve ser re-executado utilizando um conjunto de dados muito maior, como os 816 frames (408 reais, 408 fakes) que balanceamos durante nossos testes preliminares. Isso reduziria a variância e nos daria uma medida de acurácia mais confiável para o pipeline.

Em segundo lugar, seria interessante replicar os resultados da Tabela 2 do artigo original, utilizando o dataset FaceForensics++. Além disso, o classificador MLDA poderia ser substituído por métodos mais robustos, como Support Vector Machines (SVM), para avaliar se a acurácia pode ser melhorada.

Referências

[Tamanaka and Thomaz 2023] Tamanaka, F. G. and Thomaz, C. E. (2023). Sobel filter and linear classification for deepfake analysis of faces. In *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*. Sociedade Brasileira de Computação (SBC).

[Yuezun Li and Lyu 2020] Yuezun Li, Xin Yang, P. S. H. Q. and Lyu, S. (2020). Celeb-df: A large-scale challenging dataset for deepfake forensics. In *IEEE Conference on Computer Vision and Patten Recognition (CVPR)*.