Inter Forensics
CONFERÊNCIA INTERNACIONAL DE CIÊNCIAS FORENSES 2023
BRASÍLIA / BRASIL
28 a 31 de Agosto/2023
BRASÍLIA - BRASIL

Detecção de *DeepFakes* Utilizando Sistemas de Reconhecimento Facial

Paulo M. G. I. Reis[1*], Rafael O. Ribeiro[1]
[1] Instituto Nacional de Criminalística, Brasília, Distrito Federal
**Autor; e-mail: paulo.pmgir@pf.gov.br*

RESUMO

Trata-se de método para detectar vídeos *DeepFake* com troca de rosto, usando uma Rede Neural Convolucional Profunda treinada para reconhecimento facial. Foram obtidos excelentes resultados no conjunto de dados Celeb-DF (v2), e nas bases desafio do DFGC [1,2].

Palavras-chave: deepfake, faceswap, troca de rosto, autenticidade, verificação de edição.

Introdução

Vídeos *DeepFake* com troca de rosto têm impacto social significativo, especialmente quando retratam pessoas politicamente expostas ou estão relacionados a crimes relevantes. Propõe-se um método de detecção que utiliza vídeos legítimos da pessoa retratada como referência, se a pessoa for conhecida ou acessível para coleta de padrões.

Objetivos

Descrever método baseado em escores de similaridade facial para detecção de *DeepFake.*

Métodos

Extraem-se N imagens faciais da pessoa retratada em um vídeo questionado, assim como 50 imagens de faces de vídeos autênticos (cerca de 10 vídeos) da mesma pessoa para usá-las como material de referência. Usando uma rede neural convolucional profunda treinada para reconhecimento facial (ArcFace), calculam-se os escores de similaridade entre todas as combinações possíveis de pares de imagens de faces questionadas e de referência, gerando cerca de 500xN pontuações. Os valores máximo, mínimo e a mediana das pontuações (S_{max}, S_{min} e S_{med}, respectivamente) compõem um vetor F de entrada em um classificador SVM (Figura 1).

Resultados e Discussão

Empregando N=50, foi obtido um AUC de 0,994 utilizando os subconjuntos de treinamento e teste da base de dados Celeb-DF v2, superando as abordagens mais robustas publicadas contra essa base [3]. Contra 17 bases publicadas da DFGC, utilizando N=1 (tais bases são limitadas a apenas um quadro por vídeo questionado), foi obtido um AUC médio de 0,891, sendo o maior AUC de 0,958 e o menor de 0,464. O menor AUC corresponde a edição grosseira por cópia e colagem, o que explica o menor desempenho. Contra apenas edições desafiadoras nas demais bases o AUC médio sobe para 0,920.

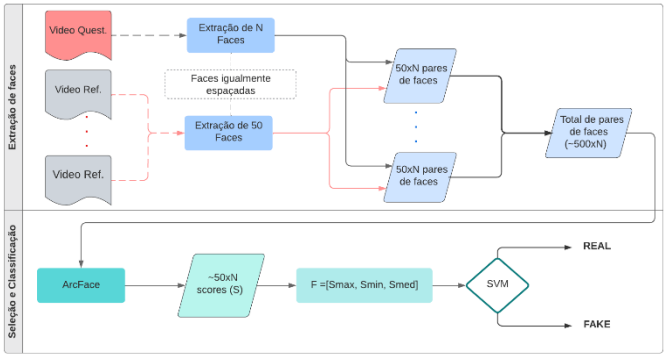

Figura 1. Método proposto em dois estágios.

Conclusão

O método proposto detecta *DeepFakes* com troca de rosto com eficácia, superando as abordagens mais robustas na base Celeb-DF v2, e mostra-se eficaz mesmo quando há apenas um quadro disponível do vídeo questionado, conforme avaliado nas bases da DFGC. Trabalhos futuros incluem validação cruzada com outras bases.

Referências

[1] Li, Yuezun, et al. “Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics”. 2020 IEEE/CVF (CVPR).

[2] B. Peng et al., "DFGC 2021: A DeepFake Game Competition," 2021 IEEE (IJCB), Shenzhen, China.

[3] Van-Nhan, et al. “High Performance DeepFake Video Detection on CNN-Based with Attention Target-Specific Regions and Manual Distillation Extraction”. Applied Sciences, vol. 11, no 16, Aug. 2021.