Gustavo Cunha Lacerda - Agosto de 2021

GUARNERA, Luca; GIUDICE, Oliver; BATTIATO, Sebastiano. Deepfake detection by analyzing convolutional traces. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops**. 2020. p. 666-667.

O autor começa falando sobre a importância que as deepfakes stá tendo na mídia, principalmente com os memes e vídeos de famosos feitos por usuários na internet. Com essa popularização surge uma necessidade de que o conteúdo falso possa ser detectado e, com isso, algumas grandes empresas decidiram tomar uma ação contra as deepfakes. O google crinado base de dados para ajudar pesquisadores e o Facebook e Microsoft criaram o Deeppfake Detection Challenge initiative.

Em seguida o artigo fala sobre algumas técnicas para a geração de deepfakes, sendo elas:

1. STARGAN
2. STYLEGAN
3. STYLEGAN2
4. ATTGAN
5. GDWCT

Depois, o autor fala sobre criar uma técnica de detecção de deepfakes usando traços convulsionais. Isso pode ser feito pois ao ser criado, as imagens que passam por uma rede neural que gera deepfakes acaba deixando rastros no processamento das imagens e na compressão a medida que avamça nas etapas da rede. Essa nova técnica consiste num método de analisar a relação de cada pixel e seus vizinhos, encontrando a relação dessas vizinhanças utilizando maximização de expectativa. Assim, após analisar as relações é possível fazer uma classificação utilizando K-NN, SVM e LDA para definir se a imagem é uma deepfake. Com essa nova técnica, a partir do banco de dados CELEBA e os algoritmos de criação de deepfakes citados, o autor conseguiu os seguintes resultados de classificação:

1. STARGAN: 93.17% de precisão máxima utilizando SVM linear e tamanho do kernel 7x7.
2. STYLEGAN: 99.65% de precisão máxima com KNN – K = 3, 5, 7, 9 e tamanho do kernel 4x4
3. STYLEGAN2: 99.81% de precisão máxima com um SVM linear e tamanho do kernel 4x4.
4. ATTGAN: 92.67% de precisão máxima com KNN – K = 3.
5. GDWCT: 88.40% de precisão máxima com KNN – K = 3, 5, 7 e tamanho do kernel 3x3.

O artigo finaliza explicando que essa abordagem se mostrou muito mais precisa do que uma usando VGG-16 (Uma rede neural de reconhecimento de imagens) na mesma base de dados. Assim, a nova técnica criada utilizando EM se mostrou bastante eficiente para detectar deepfakes geradas pelas arquiteturas GANs recentes.