Gustavo Lacerda – Setembro 2021

ROSSLER, Andreas et al. Faceforensics++: Learning to detect manipulated facial images. In: **Proceedings of the IEEE/CVF International Conference on Computer Vision**. 2019. p. 1-11.

O trabalho inicia discorrendo sobre como a computação gráfica tem sido usada para difamar pessoas. Principalmente através da face que é o principal foco por algumas razões: ser a o centro da comunicação humana e reconstruir e localizar faces ser um assunto bem desenvolvido na área de visão computacional. As manipulações faciais podem ser divididas em duas categorias distintas: manipulação de expressões faciais e manipulação da identidade visual.

Visto o problema e a evolução das deepfakes este trabalho propõe a criação de uma base de dados contendo vários deepfakes feitas utilizando técnicas diversas. Foi escolhido, então, as seguintes técnicas de computação gráfica para compor os itens do dataset: Face2Face (Manipulação de expressões faciais) e FaceSwap (Troca de características faciais) baseadas nas abordagens NeuralTextures e Deepfake. Ademais, o autor propõe um benchmark que considera os quatro métodos de manipulação para análise de um detector de manipulações faciais.

A maior contribuição do trabalho é o dataset nomeado FaceForenciss++. Ele tem a proposta de ser um conjunto de dados de larga escala para treinar as técnicas de falsificações de imagens faciais. Para isso, primeiramente, foram selecionados 1.000 vídeos da internet contendo 509.914 imagens. Com a seleção de vídeos feitas foram feitas manipulações com os métodos de FaceSwap, Face2Face e NeuralTExtures e Deepfakes, além de os vídeos serem feitos com diferentes qualidades para simular um conteúdos que passe pela rede através de alguma rede social ou fórum. Foi escolhido o codec de compressão H.264, bastante popular entre a internet, sendo um codec muito utilizado por redes sociais por ser compatível com html5 e o YouTube. O dataset foi dividido e em treino, validação e teste, contendo, respectivamente, 720, 140 e 140 vídeos.

Foi feito um estudo de como é a performance humana na detecção de falsificação. Foram usadas 204 pessoas com a maioria sendo estudante de Ciência da Computação. Os resultados demonstraram cerca de 68,69% de precisão em vídeos brutos, 66,57% em vídeos de alta qualidade e 58,73% em vídeos de baixa qualidade. A NeuralTexture foi a técnica que mais conseguiu enganar os seres humanos, que mesmo em baixa qualidade atingiram menos de 40% de precisão.

Para os testes de detecção automáticos, foi, primeiro, utilizado o FaceTraking para cortar apenas o rosto das imagens, que seriam usados como entrada nos algoritmos de classificação. Foram escolhidos diversos algoritmos de classificação, entre eles o SVM que ganhou o DeepFake Detection Challenge e CNNs como InceptionNet e XceptionNet. Comparando todos os quatro métodos em diferentes qualidades, o autor nota que todos conseguiram uma precisão alta em vídeos com qualidade bruta e precisão razoável em qualidade alta. O XceptionNet conseguiu os melhores resultados gerais nos testes conseguindo 99,26% 95,73% e 81% nos tested de qualidade bruta, alta

qualidade e baixa qualidade respectivamente. Todos os resultados mostraram também uma baixa precisão em NeuralTextures baseados em GANs.

Por fim o trabalho conclui que apesar dos resultados impressionantes da criação de falsificação de imagens o trabalho demonstra que é possível criar detectores treinados para classificar imagens manipuladas.