DOI: 10.36558/rsc.v13i3.8501

# Deepfakes na Era da Desinformação: Uma Análise Comparativa de Algoritmos de Aprendizado Profundo

## Deepfakes in the Era of Disinformation: A Comparative Analysis of Deep Learning Algorithms

Gustavo S. Rodrigues
Instituto Federal de Minas Gerais
CEP: 34590-390, Sabará,
MG, Brasil
gustavosr_13
@outlook.com

Carlos A. Silva
Instituto Federal de Minas Gerais
CEP: 34590-390, Sabará,
MG, Brasil
carlos.silva
@ifmg.edu.br

## ABSTRACT

With the rapid advancement of artificial intelligence, the creation and manipulation of images and videos have become increasingly common and accessible to the general public. As a result, deepfakes have been widely employed in the spread of misinformation and the dissemination of fake news. Recognizing deepfakes as a significant element in this new information age, this article aims to analyze three significant deep learning algorithms for the generation of deepfakes. The results indicate that Deepfacelab requires nearly double the processing time compared to FaceSwap and First Order Motion, yet it delivers superior quality in the generated results.

## CCS Concepts

•**Computing methodologies** → **Machine learning algorithms; Image processing;**

## Keywords

Deep Learning; DeepFake; Machine Learning Algorithms; Image Processing

## RESUMO

Com o rápido avanço da inteligência artificial, a criação e manipulação de imagens e vídeos se tornaram cada vez mais comuns e acessíveis para o público em geral. Como resultado, as *deepfakes* têm sido amplamente empregadas na propagação de desinformação e na disseminação de notícias falsas. Reconhecendo as *deepfakes* como um elemento marcante nessa nova era da informação, este artigo busca analisar três algoritmos de aprendizado profundo significativos para a geração de *deepfakes*. Os resultados indicam que o *Deepfacelab* requer quase o dobro do tempo de processamento em comparação com o *FaceSwap* e o *First Order Motion*, no entanto, oferece uma qualidade superior nos resultados gerados.



## Palavras-chave

Aprendizagem Profunda; Deepfake; Algoritmos de Aprendizado de Máquina; Processamento de Imagem

## 1. INTRODUÇÃO

*Deepfake*, uma expressão originada da fusão entre "Aprendizado Profundo" e "Falsificação", se refere a conteúdos fabricados artificialmente por meio de inteligência artificial. Nesse contexto, são produzidos vídeos nos quais imagens de uma pessoa-fonte [15] são superpostas a um vídeo de uma pessoa-alvo, resultando na modificação das ações e discursos da pessoa-alvo, com base no conteúdo da pessoa-fonte. Em linhas gerais, esses métodos necessitam de uma grande quantidade de dados de imagens e vídeos, tornando figuras públicas e celebridades alvos frequentes de *deepfakes*, devido à sua constante presença nos meios de comunicação. A raiz conceitual dos *deepfakes* não é algo recente. Para compreender sua origem, é importante analisar os primeiros estudos acadêmicos que estabeleceram seus princípios. Em 1997, os autores do artigo [3] criaram um programa inovador e único que, fundamentalmente, automatizou tarefas que, até então, eram realizadas exclusivamente por alguns estúdios de cinema.

O programa denominado *Video Rewrite* utiliza imagens já existentes para criar vídeos inéditos nos quais uma pessoa "aparenta" proferir palavras que não foram originalmente pronunciadas durante as filmagens originais. Essa tecnologia foi inicialmente desenvolvida com o propósito de dublar filmes, permitindo a adaptação da sequência de um filme de modo a sincronizar os movimentos labiais dos atores com o áudio correspondente. Décadas depois, essa mesma tecnologia serviu como base para a criação dos *deepfake* modernos. O primeiro vídeo *deepfake* surgiu em 2017, no qual o rosto de uma personalidade famosa foi substituído pelo de um ator de conteúdo adulto [22]. Na literatura, existem numerosos exemplos dessas manipulações visuais em contextos políticos [2], em plataformas de mídia social [7], na indústria do entretenimento [20], entre outros.

Com a crescente disseminação da internet, tornou-se cada vez mais comum que as pessoas busquem informações em redes sociais, como o Facebook, e em plataformas de compartilhamento de vídeos, como o Youtube [1]. No entanto, lamentavelmente, esse aumento no acesso à informação tem sido acompanhado pelo surgimento e propagação de notícias falsas, que podem ter consequências significativas tanto para os indivíduos quanto para a sociedade como um todo. Observa-se uma tendência de aumento no uso de *deepfakes* em situações como pornografia de vingança, bullying, criação de vídeos falsos para influenciar decisões judiciais, sabotagem política, propagação de propaganda terrorista, chantagem, manipulação do mercado financeiro e disseminação de notícias falsas [11].

Pesquisas que abordam a criação e detecção de *deepfakes* têm despertado considerável interesse na comunidade científica nos últimos anos. É comum encontrar estudos na literatura que se concentram em duas abordagens principais: reencenação, que envolve a alteração da expressão facial, boca, pose, olhar e corpo, e substituição, na qual se substitui o rosto de uma pessoa-alvo [13]. Alguns autores procuram classificar os métodos de criação e detecção de *deepfakes*, como [19], que categorizam com base na síntese facial, troca de identidade, manipulação de expressões e atributos faciais.

A discussão sobre os impactos sociais desses métodos é abrangente e aborda diversas perspectivas, incluindo mídia, sociedade, leis e regulamentações. Por exemplo, um estudo realizado por [9] investiga os *deepfakes* sob múltiplos ângulos, analisando seu impacto nas esferas midiáticas, sociais e jurídicas, refletindo a preocupação com as implicações amplas dessas tecnologias na sociedade.

Neste estudo, temos como objetivo aprofundar a compreensão acerca dos métodos de geração de *deepfakes*, analisando a funcionalidade de três algoritmos específicos, nomeadamente, *Deepfacelab*, *Faceswap* e *First Order Motion*. Para alcançar esse propósito, a estrutura do artigo é organizada da seguinte forma: A seção 1 apresenta uma introdução ao tópico abordado no trabalho. Na seção 2, são discutidos trabalhos relevantes da literatura relacionados à área de pesquisa. A metodologia utilizada é descrita na seção 3. Os algoritmos de aprendizado profundo, que constituem a base para a abordagem de *deepfakes*, são apresentados na seção 4, e a análise dos resultados obtidos é realizada na seção 5. Por fim, na seção 6, são apresentadas as conclusões finais e perspectivas para futuras pesquisas.

## 2. REVISÃO BIBLIOGRÁFICA

Os *deepfakes* representam uma inovação tecnológica fascinante que desafia as fronteiras da realidade e da ficção. Tal tecnologia têm o poder de alterar drasticamente a paisagem da informação e do entretenimento, levantando questões fundamentais sobre a autenticidade das imagens e vídeos que consumimos diariamente. Esse fenômeno, baseado em técnicas de aprendizado profundo, tem implicações profundas na esfera da mídia e da comunicação, exigindo uma avaliação crítica e aprimoramento das ferramentas de detecção para mitigar potenciais riscos e usos indevidos.

O artigo [10] se destaca como um dos primeiros estudos acadêmicos a introduzir uma implementação de código aberto de tecnologia *deepfake*.Esta estratégia fundamenta-se nas Redes Adversárias Generativas (GANs, ou *Generative Adversarial Networks* em inglês) com o propósito de efetuar a substituição do rosto de uma pessoa por outro em vídeos. Os autores destacam a importância de diversos fatores, tais como a quantidade de dados acessíveis, a metodologia de treinamento e a calibração meticulosa dos parâmetros, os quais exercem um impacto direto na qualidade dos vídeos gerados.

No estudo de [5], a aplicação de *deepfake* teve como objetivo a transferência de movimentos entre duas pessoas em uma performance de dança. A realização bem-sucedida desse processo destacou o notável potencial dessa tecnologia, sugerindo a possibilidade de sua expansão para outras áreas, como animação e cinema.

Com o objetivo de tornar a tecnologia mais acessível e demonstrar seu potencial, os autores de [18] criaram um modelo intuitivo voltado para o público não especializado. Eles disponibilizaram uma rede neural treinada para reconhecimento de imagens. O algoritmo desenvolvido possibilitou a manipulação de imagens para a criação de vídeos ou *gifs* animados de forma altamente convincente.

Em [14], os autores realizaram uma avaliação de algoritmos de *deepfake*, revelando a relevância do treinamento progressivo para a substituição de rostos em alta resolução. Os resultados indicaram que a aplicação de técnicas de estabilização de pontos de referência (*landmarks*) aprimora os resultados, minimizando oscilações perceptíveis e outras instabilidades temporais que podem surgir ao trabalhar com imagens de alta resolução.

Em 2020 um grupo de pesquisadores desenvolveu um *framework* que viabiliza a substituição de rostos em imagens ou vídeos, com o propósito de aprimorar o acesso de outros estudiosos nesta área [16]. Os pesquisadores optaram por desenvolver um método de alto desempenho por conta própria, a fim de simplificar a aprendizagem e manipulação dessa tecnologia, proporcionando uma ferramenta mais versátil e facilmente adaptável a diversos contextos. Ao final deste trabalho, os autores realizaram um comparação da solução desenvolvida com outras soluções disponíveis, obtendo um excelente resultado ao final da comparação.

É de suma importância debater o potencial impacto que as *deepfakes* podem ter na sociedade. Esse debate foi o ponto central de um estudo conduzido por [15], no qual os autores se dedicaram a apresentar diversos métodos para a detecção de *deepfake*, estimulando uma reflexão ampla acerca dos desafios, tendências futuras e possíveis direções nessa área. Além disso, um trabalho de pesquisa descrito em [21] ressaltou o impacto das *deepfakes* na confiabilidade das informações contidas em propagandas veiculadas pelos meios de comunicação tradicionais, assim como no discurso público, que acaba por criar um ambiente permeado de desconfiança e desinformação. Os autores sublinham que a ausência de regulamentações que limitem o uso indevido da inteligência artificial na internet pode acarretar consequências socioeconômicas significativas no futuro.

## 3. METODOLOGIA

O presente estudo aborda uma pesquisa exploratória centrada no tópico de *deepfake*, com foco em três algoritmos relevantes que são empregados na criação dessa tecnologia. Especificamente, os algoritmos selecionados são o *Deepfacelab* [16], *Faceswap*[1], e *First Order Motion* [18]. A seleção destes três algoritmos foi baseada na análise de diversos

[1]https://faceswap.dev

critérios. Foram priorizados parâmetros como a facilidade de implementação, que requer um nível mais baixo de conhecimento técnico para utilização e configuração. Esses algoritmos são amplamente adotados por criadores de conteúdo audiovisual, acadêmicos e empresas, e contam com uma documentação *online* abrangente que auxilia no manuseio e aprendizado. Outro aspecto considerado foi a adaptabilidade do sistema às especificações de *hardware* disponíveis, pois alguns algoritmos demandam requisitos de *hardware* mais avançados, o que pode tornar sua utilização mais complexa. Inicialmente foi realizado um levantamento bibliográfico em bases de dados científicas como *Web of Science*, *IEEE Xplore*, Google Acadêmico, *arXiv* entre outros, além dos inúmeros repositórios do GitHub, sobretudo referendado por [12], o qual aplica análise de conteúdo qualitativo contextual para explorar os repositórios mais populares do GitHub e contas do YouTube que ensinam "como fazer *deepfake*".

Após estabelecer a base teórica e compreender o estado-da-arte, o passo subsequente envolveu a criação de uma base de dados unificada, que serviria como entrada para os três algoritmos sob análise. Essa base de dados foi obtida a partir de [17] e consiste em mais de meio milhão de imagens editadas, originárias de mais de 1.000 vídeos. Em seguida, os algoritmos foram implementados e as simulações computacionais foram conduzidas. O sistema utilizado para essas tarefas foi baseado em um processador Intel i7 10700F, equipado com uma placa de vídeo Zotac RTX 3080 12GB e 16GB de memória RAM DDR4 a 3200MHz. Assim, com o propósito de apresentar resultados e informações pertinentes relacionadas aos métodos de aprendizado profundo, a seção 4 oferece uma descrição dos algoritmos adotados neste estudo.

# 4. DESENVOLVIMENTO

A compreensão da operação dos procedimentos de criação de *deepfakes* é de extrema relevância para estimular uma análise mais abrangente sobre o tópico em questão. Portanto, nas subseções seguintes, serão introduzidos os três algoritmos empregados neste estudo.

## 4.1 Deepfacelab

O sistema *Deepfacelab* [16] representa uma plataforma de código aberto desenvolvida para realizar substituições de rostos de alta qualidade por meio da tecnologia *deepfakes*. Este projeto no ambiente do Github acumula mais de 3.000 ramificações (repositórios derivados que compartilham configurações e visibilidade com o repositório original) e possui um impressionante número de 14.000 estrelas, que pode ser considerado como uma métrica de sua popularidade. O sistema é notável por sua facilidade de uso e oferece um equilíbrio entre desempenho e acessibilidade.

Atualmente, o sistema disponibiliza duas diferentes arquiteturas, a saber, a DF (*DeepFake*) e a LIAE (*Lightly Improved Auto Encoder*), cada uma com distinções significativas relacionadas a parâmetros como interpretação facial com ou sem transformação, bem como a precisão e fidelidade em relação aos dados de origem. A arquitetura DF demonstra um desempenho superior quando as faces de origem e destino possuem semelhanças notáveis em termos de características e cores. Em contraste, a arquitetura LIAE é mais tolerante em relação a transformações e se adapta de maneira mais flexível aos rostos de destino, incluindo formatos e cores diferentes.

Com base nessas arquiteturas, o sistema disponibiliza os modelos Quick96, SAEHD e AMP. Para os propósitos deste estudo, optamos por utilizar o modelo SAEHD devido à sua capacidade de selecionar qualquer uma das arquiteturas, tornando-o o modelo mais versátil entre os três disponíveis. O sistema opera por meio de uma fase de conversão que envolve um "codificador" e um "decodificador de destino", com uma camada intermediária entre eles. A abordagem proposta pela arquitetura LIAE é ilustrada na Figura 1.

No que se refere à extração de características da imagem de origem, o algoritmo emprega um sistema de referência facial baseado em calor 2DFAN [4], que atua em conjunto com uma rede de segmentação facial de granulação fina, denominada TernausNet [8], para obter a segmentação facial.

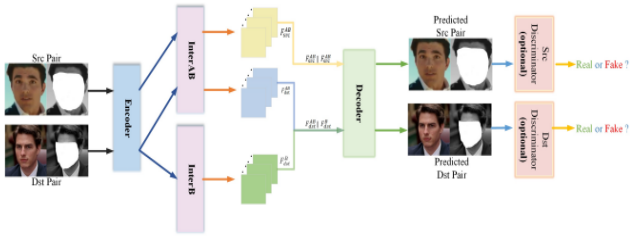


Figura 1: Arquitetura do *Deepfacelab* [16].

## 4.2 Faceswap

O Faceswap [6] é uma tecnologia de *deepfake* que facilita a substituição de dois rostos por meio de uma rede neural profunda. Quando o rosto de uma pessoa é substituído por um rosto proveniente de uma fonte de código aberto, o resultado torna-se irreconhecível em relação ao rosto original. De acordo com [23], o rosto gerado mantém as características faciais originais, como expressões e a tonalidade da pele, e essa capacidade se estende igualmente a vídeos.

Um exemplo do funcionamento do Faceswap é descrito em detalhes em [23] e é ilustrado na Figura 2.

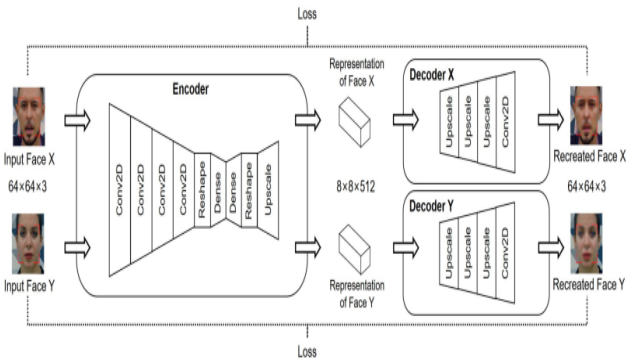


Figura 2: *Faceswap* [23].

As entradas $X$ e $Y$ referem-se a exemplos de faces de código aberto. No procedimento, a face de entrada passa por um codificador, que a converte em um vetor de representação. Esse vetor é então processado por um decodificador

correspondente, que o transforma de volta em uma face recriada. A perda é calculada durante a retropropagação, representando a discrepância entre as faces de entrada e as faces regeneradas. Durante o processo de troca, o modelo efetua uma transição no decodificador para produzir as imagens trocadas.

### 4.3 First Order Motion

Diferentes domínios de interesse, como a produção de filmes, fotografia, *e-commerce*, podem se beneficiar de diversas aplicações resultantes da geração de vídeos por meio da animação de objetos em imagens estáticas. A animação de imagens envolve a síntese de vídeos, combinando a aparência extraída de uma imagem de origem com os padrões de movimento derivados de um vídeo de referência. Problemas que se enquadram nessa categoria são frequentemente abordados na literatura por meio da ênfase na representação do objeto e pelo uso de técnicas de computação gráfica que pressupõem conhecimento sobre o modelo específico do objeto a ser animado.

A abordagem proposta pelo *First Order Motion* consiste em empregar modelos generativos profundos que utilizam um conjunto de pontos-chave autoaprendidos, em conjunto com transformações afins locais, que é a razão para o nome "primeira ordem". Esses elementos permitem modelar movimentos complexos. Além disso, o *First Order Motion* introduz um gerador com capacidade de reconhecimento de oclusão. Esse gerador adota automaticamente uma máscara de oclusão para identificar regiões do objeto que não estão aparentes na imagem original, as quais devem ser inferidas com base no contexto [18].

Na Figura 3, podemos observar a representação visual do modo de operação do algoritmo *First Order Motion*.

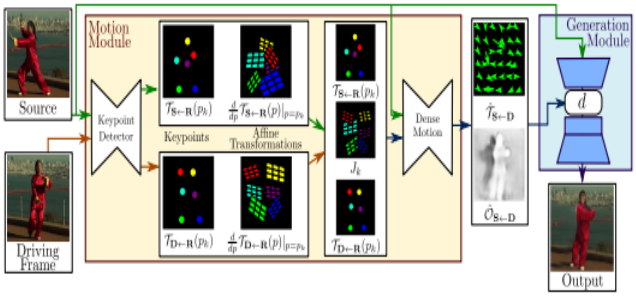


Figura 3: *First Order Motion* [18].

Conforme afirmado em [18], o método pressupõe a existência de uma imagem de origem, denotada como $S$, e um quadro (*frame*) $D$ de um vídeo de referência. Um detector de pontos-chave não supervisionado é empregado para extrair o movimento de primeira ordem, que compreende pontos-chave esparsos e transformações afins locais em relação ao quadro de referência $R$. Uma densa rede de movimento utiliza essa representação de movimento para gerar o fluxo ótico denso $\tau_S \leftarrow D$ de $D$ para $S$ e o mapa de oclusão $\mathcal{O}_S \leftarrow D$. A imagem de origem, juntamente com as saídas da rede de movimento denso, são então utilizadas pelo gerador para criar a imagem de destino.

## 5. DISCUSSÃO DOS RESULTADOS

Além do amplo levantamento bibliográfico que embasa este estudo, buscamos a compreensão do funcionamento dos principais algoritmos de geração de *deepfakes* por meio de simulações computacionais utilizando um conjunto de dados da literatura.

Para a realização dos testes, utilizamos a base de dados Faceforensic++[2]. Essa base de dados consiste em uma coleção de 1.000 vídeos originais que foram sujeitos a manipulações com quatro técnicas automatizadas de alteração facial: *Deepfakes*, *Face2Face*, *Faceswap* e *NeuralTextures*. Uma parcela desses dados, equivalente a 100 GB de vídeos, foi separada para os testes dos algoritmos empregados neste estudo.

Na Tabela 5, é apresentado o tempo despendido por cada algoritmo, levando em conta o treinamento, o número máximo de iterações e os requisitos mínimos de *hardware* necessários para a execução de cada algoritmo.

| Alg. | Tam. base | Mod. | Tempo(s) | Iter. | Requisito |
|---|---|---|---|---|---|
| *DeepFaceLab* | 100Gb | SAEHD | $8,28 \times 10^4$ | 100k | GPU CUDA 3.5, $\geq$ 16Gb ram, CPU $\geq$ 8 nuc., 8 Gb de VRAM |
| *First Order Motion Model* | 100Gb | - | $4,02 \times 10^4$ | 100k | GPU CUDA |
| *Faceswap* | 100Gb | - | $4,84 \times 10^4$ | 100k | GPU CUDA 3.5, 8 Gb de VRAM |

Ao analisar os resultados obtidos, fica evidente uma disparidade significativa entre os tempos de execução dos diferentes modelos de algoritmos ao realizar a mesma tarefa com a mesma base de dados. Notavelmente, o *Deepfacelab*, quando empregando o modelo SAEHD, exigiu quase um dia completo, totalizando 23 horas de processamento, em contraste com os algoritmos *First Order Motion* e *Faceswap*, que demandaram aproximadamente 11 horas e 13,5 horas, respectivamente.

Nesta pesquisa, exploramos o *Deepfacelab* com o modelo SAEHD, notando que, embora este modelo seja mais exigente em termos de recursos de *hardware* e requer um tempo de treinamento e execução mais prolongado, ele demonstra a capacidade de proporcionar resultados superiores, com imagens finais mais nítidas. Em trabalhos futuros, outros modelos podem ser considerados, como o *Quick96*, que representa uma opção alternativa para treinamento dos dados e oferece a vantagem de tempos de treinamento e execução mais curtos em comparação com o modelo SAEHD.

Após a conclusão dos testes computacionais que abrangeram os três algoritmos, o *Deepfacelab*, embora tenha demonstrado tempos de treinamento e execução notavelmente mais longos em comparação com os outros dois algoritmos, se destacou ao gerar *deepfakes* de qualidade visual superior. As produções resultantes exibiram menos artefatos, como serrilhados, e apresentaram movimentos mais suaves em partes críticas dos vídeos, como a boca e os olhos, conforme evidenciado nas Figuras 8 e 9.

[2] `https://github.com/ondyari/FaceForensics/blob/master/dataset/README.md`

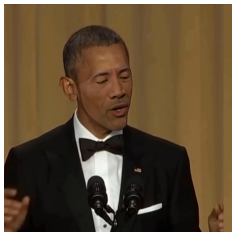

Figura 4: *Faceswap.*

Figura 5: *Deepface.*

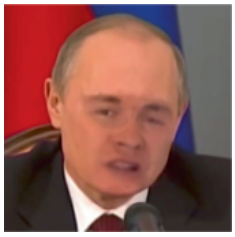

Figura 6: *FirstOrder.*

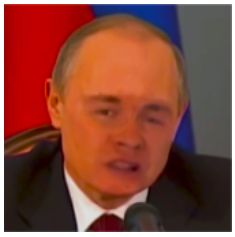

Figura 7: *Deepface.*

Figura 8: *FirstOrder.*

Figura 9: *Deepface.*

## 6. CONCLUSÃO

O termo *deepfakes* engloba uma série de algoritmos que se baseiam em inteligência artificial e têm a capacidade de criar vídeos fraudulentos de maneira convincente, com o objetivo de enganar os espectadores. Embora tenham aplicação no entretenimento, como em filmes e vídeos virais, os *deepfakes* possuem igualmente o potencial de serem explorados para a disseminação de informações falsas, difamação de indivíduos e a causa de danos substanciais. É imperativo que a sociedade esteja alerta quanto aos perigos relacionados às tecnologias de *deepfake* e colabore na busca por soluções para identificar e enfrentar sua propagação.

Neste estudo, realizou-se uma análise exploratória desta tecnologia, onde se abordaram e compararam os algoritmos *Deepfacelab*, *Faceswap* e *First Order Motion*. O objetivo foi avaliar suas capacidades, visando a uma compreensão mais aprofundada de seu funcionamento e à apresentação de uma discussão significativa sobre as *deepfakes*. Em sua essência, é possível afirmar que alguns desses modelos, embora imponham maiores exigências em termos de recursos de hardware e tempo computacional, têm o potencial de proporcionar resultados finais de maior qualidade. Como parte de trabalhos futuros, planeja-se estabelecer uma métrica de desempenho específica para avaliar a qualidade dos vídeos e imagens gerados por meio de abordagens de aprendizado profundo. Como evidenciado anteriormente no estudo de [14], foi demonstrado ser factível realizar uma comparação entre algoritmos de *deepfake*. Neste trabalho, uma métrica de comparação foi estabelecida para conduzir a pesquisa, sendo o pioneiro em alcançar resultados convincentes na troca de rostos em vídeos de alta resolução, incluindo o domínio de megapixels e além.

## 7. AGRADECIMENTOS

Agradecemos aos professores Glauco Douglas Moreira - (Chefe do Setor de Tecnologia da Informação do IFMG-Sabará) e Jean Nunes Ribeiro Araújo (Pesquisador do ORC-SLab@UFMG) pela assistência e comentários que aprimoraram o manuscrito e ao Instituto Federal de Minas Gerais campus Sabará pelo apoio financeiro fornecido para a realização deste trabalho.

## 8. REFERÊNCIAS

[1] K. E. Anderson. Getting acquainted with social networks and apps: combating fake news on social media. *Library Hi Tech News*, 35(3):1–6, 2018.

[2] M. Appel and F. Prietzel. The detection of political deepfakes. *Journal of Computer-Mediated Communication*, 27(4):zmac008, 2022.

[3] C. Bregler, M. Covell, and M. Slaney. Video rewrite: Driving visual speech with audio. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 353–360, 1997.

[4] A. Bulat and G. Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In *Proceedings of the IEEE international conference on computer vision*, pages 1021–1030, 2017.

[5] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros. Everybody dance now. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5933–5942, 2019.

[6] Deepfakes. Faceswap. https://github.com/deepfakes/faceswap, 2017. Acessado: 03-04-2023.

[7] T. Fagni, F. Falchi, M. Gambini, A. Martella, and M. Tesconi. Tweepfake: About detecting deepfake tweets. *Plos one*, 16(5):e0251415, 2021.

[8] V. I. Iglovikov and A. A. Shvets. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *ArXiv*, abs/1801.05746, 2018.

[9] S. Karnouskos. Artificial intelligence in digital media: The era of deepfakes. *IEEE Transactions on Technology and Society*, 1(3):138–147, 2020.

[10] P. Korshunov and S. Marcel. Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685*, 2018.

[11] M.-H. Maras and A. Alexandrou. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos. *The International Journal of Evidence & Proof*, 23(3):255–262, 2019.

[12] A. McCosker. Making sense of deepfakes: Socializing ai and building data literacy on github and youtube. *New Media & Society*, page 14614448221093943, 2022.

[13] Y. Mirsky and W. Lee. The creation and detection of

deepfakes: A survey. *ACM Computing Surveys (CSUR)*, 54(1):1–41, 2021.

[14] J. Naruniec, L. Helminger, C. Schroers, and R. M. Weber. High-resolution neural face swapping for visual effects. In *Computer Graphics Forum*, volume 39, pages 173–184. Wiley Online Library, 2020.

[15] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi. Deep learning for deepfakes creation and detection. *arXiv preprint arXiv:1909.11573*, 1:1–19, 2019.

[16] I. Perov, D. Gao, N. Chervoniy, K. Liu, S. Marangonda, C. Umé, M. Dpfks, C. S. Facenheim, L. RP, J. Jiang, et al. Deepfacelab: Integrated, flexible and extensible face-swapping framework. *arXiv preprint arXiv:2005.05535*, 2020.

[17] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner. Faceforensics: A large-scale video dataset for forgery detection in human faces, 2018.

[18] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe. First order motion model for image animation. *Advances in Neural Information Processing Systems*, 32, 2019.

[19] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia. Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64:131–148, 2020.

[20] B. Usukhbayar and S. Homer. Deepfake videos: The future of entertainment. http://dx.doi.org/10.13140/RG.2.2.28924.62085, mar 2020.

[21] C. Vaccari and A. Chadwick. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1):2056305120903408, 2020.

[22] P. Yu, Z. Xia, J. Fei, and Y. Lu. A survey on deepfake video detection. *Iet Biometrics*, 10(6):607–624, 2021.

[23] B. Zhu, H. Fang, Y. Sui, and L. Li. Deepfakes for medical video de-identification: Privacy protection and diagnostic information preservation. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 414–420, 2020.