

Classifying Synthetic Facial Images: A Quantitative and Qualitative Approach with AI

Classificação de imagens faciais sintéticas: Uma análise quantitativa e qualitativa usando IA

Fernanda Goyo Tamanaka^{1*}, Lucas Fontes Buzuti¹, Carlos Eduardo Thomaz¹

Abstract: Through artificial intelligence (AI) techniques, synthetic content is generated, which is an increasingly critical issue for digital security. This paper aims to classify real and synthetic facial images using a state-of-the-art AI model to analyze, both quantitatively and qualitatively, its performance and the most discriminating features across the various data transformations used for this purpose in current literature. In this scenario, the study utilized three public benchmark datasets: two for generating fake images (Stable Diffusion and StyleGAN2) and one for real images (FFHQ), employing the Xception architecture with the Grad-CAM++ viewer. Overall, the results indicate that the central region of the image is the most discriminative one, whereas random cropping produced the worst performance. Future work aims to explore other explainable Artificial Intelligence (XAI) methods, as well as other training and validation protocols, such as cross-testing between synthetic databases.

Keywords: Xception — Grad-CAM++ — synthetic faces — image classifying

Resumo: Por meio de técnicas de inteligência artificial (IA), são criados conteúdos sintéticos. Sendo um problema crescente na segurança digital, este artigo propõe classificar imagens faciais reais e sintéticas usando um modelo de IA do estado-da-arte para analisar, tanto quantitativa quanto qualitativamente, o desempenho dele e as características mais discriminantes nas diferentes transformações de dados utilizadas para este fim na literatura corrente. Para isto, este estudo utilizou três bases de dados públicas de referência na área, sendo duas que geram imagens fakes (Stable Diffusion e StyleGAN2) e uma real (FFHQ), utilizando a arquitetura Xception com o visualizador Grad-CAM++. De modo geral, os resultados indicam que a parte central da imagem é a mais discriminante, ao passo que a aleatoriedade do recorte apresentou o pior desempenho. Como trabalhos futuros, espera-se explorar outros métodos de explainable Artificial Intelligence (XAI), além de outros protocolos de treinamento e validação, como o teste cruzado entre as bases sintéticas.

Palavras-Chave: Xception — Grad-CAM++ — faces sintéticas — classificação de imagens

¹ Departamento de Engenharia Elétrica, FEI, Brasil

*Corresponding author: fegoyo@fei.edu.br

DOI: <http://dx.doi.org/10.22456/2175-2745.150810> • Received: 09/10/2025 • Accepted: 03/11/2025

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introdução

Por meio de técnicas de inteligência artificial (IA), são criados conteúdos sintéticos [1, 2] que envolvem a manipulação de textos, áudios, vídeos e/ou imagens, utilizando técnicas de machine learning (ML) e deep learning (DL), especificamente, as Redes Generativas Adversariais (GANs) [3].

Diante desse cenário que engloba um problema crescente na segurança digital, empresas como Google e Facebook promoveram competições públicas de identificação de conteúdo fake, em que o modelo que melhor predizia a classificação era o vencedor [4]. Como consequência, houve um aumento exponencial nas pesquisas voltadas para o desenvolvimento de métodos eficazes de classificação dessa tecnologia [5, 6]. Embora existam conteúdos fakes que abrangem desde textos até imagens [1, 2], esta pesquisa focará na classificação de

imagens faciais reais e fakes, por permitirem melhor entendimento e controle dos atributos a serem estudados.

Neste contexto, apesar de existirem trabalhos de classificação [5, 6, 1], até o momento, não foi possível encontrar um que tenha analisado quantitativa e qualitativamente as transformações de dados. Com isso, este artigo propõe classificar imagens faciais reais e sintéticas usando um modelo de IA do estado-da-arte para analisar, tanto quantitativa quanto qualitativamente, o desempenho dele e as características mais discriminantes nas diferentes transformações de dados utilizadas para este fim na literatura corrente. A identificação de transformações de dados ótimas e a análise das características faciais críticas para a classificação fornecem uma orientação prática para o desenvolvimento de sistemas mais eficazes e interpretáveis de detecção de imagens falsas.

Este estudo está dividido em cinco seções, conforme

segue: A Seção 2 apresenta os principais trabalhos relacionados com o objetivo deste trabalho. A Seção 3 descreve a metodologia para classificar as imagens faciais. Resultados e discussão estão descritas na Seção 4. Por fim, a Seção 5 apresenta os achados e os direcionamentos para a pesquisa futura.

2. Trabalhos Relacionados

No estudo de [5], os autores usaram quatro bases de dados de vídeos: Faceforensics++ [7], Celeb-DF [8], DFDC [9] e UADFV [10] para classificá-los usando a segmentação das regiões faciais (olhos, nariz, boca, resto da face e toda a face).

Com exceção do DFDC, os resultados mostraram que os classificadores de aprendizado profundo, principalmente o Xception [11], apresentaram um desempenho significativo com AUC superior a 91% ao analisar toda a face e as regiões dos olhos e da boca. A combinação dessa análise com o Grad-CAM revelou que os principais erros de detecção ocorrem quando as áreas dos olhos ou da boca estão fechadas.

Diferentemente de [5], [6] trabalharam apenas com bases de imagens faciais, sendo o pioneiro a comparar imagens geradas por texto (Stable Diffusion) [12] e imagens sintéticas produzidas pela GAN [3] (StyleGAN [13], StyleGAN2 [14] e StyleGAN3 [15]) com diferentes imagens reais do Flickr-Faces-HQ Dataset (FFHQ) [13], para identificar qual é a melhor técnica de transformação de dados (Center Crop, Random Crop, Random Resize Crop e Resize) para discriminar as imagens nos diferentes modelos estudados: Xception [11], VGG16 [16], ResNet-50 [17], Mobilenet [18] e ViT [19]. Os resultados mostraram que o **Center Crop** foi a melhor técnica para classificar as imagens, apresentando uma acurácia de, pelo menos, 99% em todos os modelos analisados.

As transformações utilizadas por [6] têm como vantagem a otimização do desempenho computacional por focarem em pontos específicos, porém dependem da qualidade da imagem e são sensíveis à posição do recorte.

3. Materiais e Métodos

A integração dos trabalhos do estado-da-arte de [6] e [5] possibilitou investigar, tanto quantitativa quanto qualitativamente, como a rede neural Xception [11] classificou as imagens como reais ou sintéticas, além de identificar suas características mais discriminantes, utilizando o Grad-CAM++ [20].

O uso da Grad-CAM++ [20] em particular se justifica por ser um método de explainable Artificial Intelligence (XAI) que destaca de forma holística as regiões da imagem associadas à tomada de decisão do modelo, além de ser independente da arquitetura de IA utilizada.

Para realizar este trabalho, foi utilizado um servidor o seguinte framework computacional:

- Sistema Operacional Linux, arquitetura x86_64
- CPU, Intel(R) Xeon(R) Gold 5118 2.30GHz

- GPU, 1x Nvidia V100 16Gb
- RAM, 188Gb

3.1 Bases de dados

Esta pesquisa utilizou três bases de imagens públicas disponíveis no github de [6]¹ e que podem ser encontrados alguns exemplos na Figura 1:

- Flickr-Faces-HQ Dataset (FFHQ): Desenvolvido pela NVIDIA [13], o FFHQ é uma base de dados com 70.000 faces extraídas do site Flickr. Essas faces são imagens distintas de alta qualidade com resolução de 1024x1024 pixels, abrangendo uma diversidade de idades, etnias, fundos de imagens e acessórios.
- Stable Diffusion (Stable): Desenvolvido por [12], o Stable Diffusion utiliza o modelo de difusão latente (*Latent Diffusion Models*, ou LDM) para gerar imagens fakes de alta qualidade com resolução de 512x512 pixels, a partir de textos. O treinamento do Stable Diffusion é baseado, principalmente, no uso da base de dados LAION [21].
- StyleGAN2 (GAN2): Criado por [14] para gerar imagens fakes, principalmente de faces, com resoluções de até 1024x1024 pixels. Esses modelos foram treinados com bases de dados reais, como FFHQ [13], CelebA-HQ [22] e LSUN [23].

3.2 Quantificação da detecção das imagens

A Figura 2 ilustra o arcabouço computacional utilizado para classificar se a imagem é real ou fake, considerando as três bases de imagens (FFHQ, Stable e GAN2) que foram separadas em dois conjuntos de imagens distintos (FFHQStable) e (FFHQGAN2) para avaliação das métricas, dado a diferença na abordagem da geração do conteúdo sintético entre o Stable e o GAN2 e, conseqüentemente, na qualidade da imagem.

Dentro de cada conjunto, existem duas bases diferentes: a primeira denominada **Normal** não foi aplicado o pré-processamento e, portanto, teve todo o conteúdo preservado. Já a segunda, denominada **Recortada**, foi aplicado o pré-processamento utilizando o software estado-da-arte Openface [24] através da técnica **Crop**, para recortar apenas as faces das imagens, porém assegurando que elas não fossem totalmente escuras e que aparecesse a face. Tanto na **Normal** como na **Recortada** as imagens foram redimensionadas para 512x512 pixels. Com isso, as imagens puderam ser comparadas.

Na base **Normal**, a preservação do conteúdo original da imagem possibilitou a aplicação de diferentes técnicas de transformação dos dados para analisar se o fundo é relevante para a classificação. Por outro lado, na base **Recortada**, o foco é verificar se o recorte facial isola as características mais discriminantes.

Esses dois conjuntos de imagens (FFHQStable) e (FFHQGAN2), são descritas a seguir:

¹https://github.com/LucaCorvito/RealFaces_w_StableDiffusion

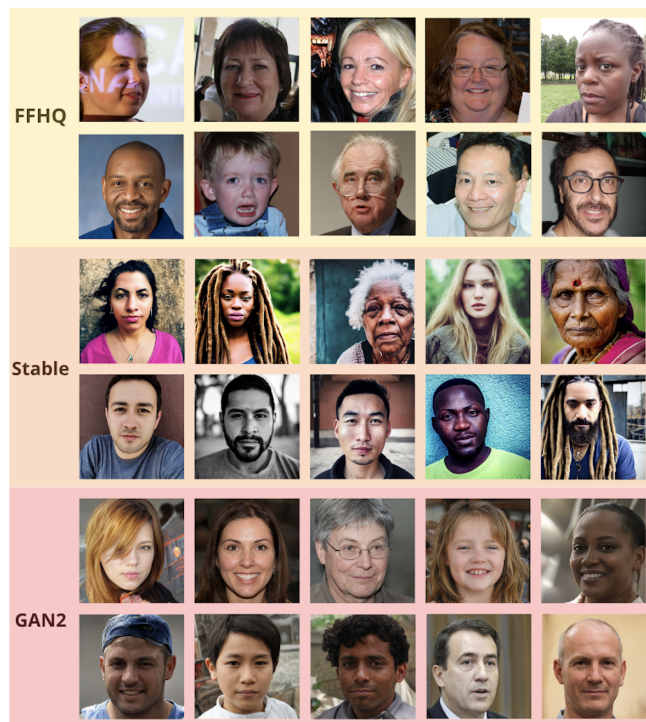


Figure 1. Exemplos de figuras presentes em cada base de dados.

- **Conjunto de Imagens 1 (FFHQStable):** A base **Normal** contém 25.830 imagens reais da base FFHQ e 25.369 imagens falsas geradas com Stable Diffusion [12], com textos [6] que detalharam a iluminação e o tipo de câmera. Já a base **Recortada** contém a mesma quantidade de imagens reais da base **Normal**, porém foram excluídas 2.569 imagens fakes que ficaram totalmente escuras ou não aparecem a face, totalizando 23.261 imagens sintéticas.
- **Conjunto de Imagens 2 (FFHQGAN2):** A base **Normal** contém 7.000 imagens reais da base FFHQ e 6.873 imagens falsas geradas com StyleGAN2 [14]. Já a base **Recortada** apresenta a mesma quantidade de imagens sintética da base **Normal**, porém foi excluída uma imagem real por ter ficado totalmente escura, totalizando 6.999 imagens reais.

Após o pré-processamento, as imagens passaram por um processo de transformação dos dados e normalização antes de serem escolhidas aleatoriamente visando à proporção entre reais e sintéticas de cada conjunto de imagens usado nos experimentos na proporção de 70%, 15% e 15% para, respectivamente, treinar, validar e testar o modelo Xception [11]. As quantidades de imagens por conjunto estão apresentadas nas Tabelas 1 e 2.

No processo de transformação dos dados para 224x224 pixels, **Center Crop** refere-se ao recorte do centro da imagem; **Random Crop** realiza o recorte de um ponto aleatório da imagem; **Random Resize Crop** recorta um ponto aleatório da

imagem, reduzindo sua resolução e, por fim, o **Resize** reduz a resolução da imagem.

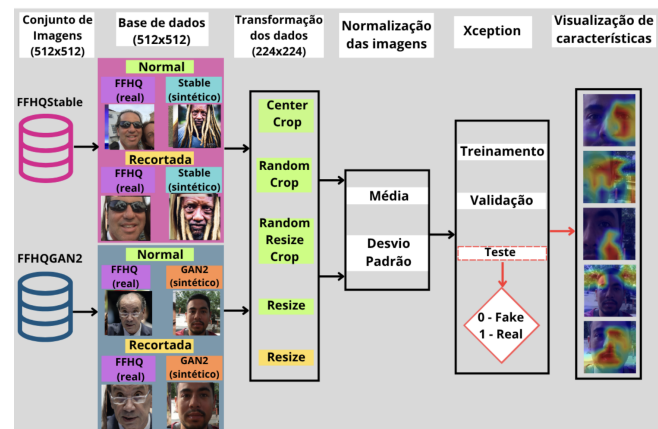


Figure 2. Arcabouço computacional.

Table 1. Quantidade de imagens de treinamento, validação e teste - Conjunto de Imagens 1 (FFHQStable).

Base de dados	Treinamento	Validação	Teste	Total
Normal	35.839	7.680	7.680	51.199
Recortada	34.377	7.366	7.348	49.091

Table 2. Quantidade de imagens de treinamento, validação e teste - Conjunto de Imagens 2 (FFHQGAN2).

Base de dados	Treinamento	Validação	Teste	Total
Normal	9.711	2.081	2.081	13.873
Recortada	9.711	2.081	2.080	13.872

Com relação ao treinamento, não foi usada uma versão do modelo Xception pré-treinada. Além disso, não foram adicionadas camadas ao final do modelo, sendo apenas removida a fully connected padrão de 1000 neurônios e modificada para 1 neurônio, para resolver o problema binário em questão de classificação de imagens. Utilizando o Pytorch, o modelo de classificação para o conjunto de imagens FFHQStable foi treinado com 200 épocas e para o conjunto FFHQGAN2 foram utilizadas 400 épocas, ambos usando o otimizador AdaGrad, com batch size de 32, taxa de aprendizado de $1e-6$ e regularização (L2) de 0,01. O ponto de parada desses treinamentos foi em função da análise de convergência conjunta das curvas de treinamento e validação, com o recurso de early stopping somente usado no segundo caso.

A validação do modelo Xception foi realizada analisando as métricas da acurácia, precisão, revogação e F1-score das imagens de teste usando uma função sigmoide, com valores variando entre 0 e 1, em que 0,5 é o valor limiar. Nesse caso, valores acima de 0,5 significam que a imagem foi classificada como real, enquanto valores abaixo desse limiar original mostram que a imagem foi considerada fake.

Com o auxílio do Grad-CAM++ [20] foi possível visualizar, nas imagens de teste, quais foram as características que o modelo usou para fazer essa distinção, considerando todas essas transformações. Além disso, realizou-se uma análise das métricas, destacando o F1-score para cada classificação, tendo em vista que esse é um indicador que mensura a assertividade geral do modelo na classificação das imagens.

4. Resultados e Discussão

Esta seção explana os resultados obtidos a partir da metodologia explicada anteriormente na seção 3.

Table 3. Resultados do Conjunto de Imagens 1 (FFHQStable).

Técnica	Acurácia	Precisão	Recall	F1 score
Center Crop	0.9947	0.9962	0.9935	0.9948
Random Crop	0.9339	0.9702	0.8999	0.9337
Random Resize Crop	0.9641	0.9773	0.9527	0.9649
Resize	0.9861	0.9842	0.9889	0.9866
Crop + Resize	0.9729	0.9644	0.9781	0.9712

Table 4. Resultados do Conjunto de Imagens 2 (FFHQGAN2).

Técnica	Acurácia	Precisão	Recall	F1 score
Center Crop	0.8241	0.8348	0.8207	0.8277
Random Crop	0.7049	0.8715	0.5005	0.6358
Random Resize Crop	0.7669	0.7190	0.8982	0.7987
Resize	0.7838	0.8423	0.7134	0.7725
Crop + Resize	0.8284	0.8256	0.8449	0.8351

De modo geral, observando as Tabelas 3 e 4, o **Center Crop** e o **Crop + Resize** foram as técnicas com melhor desempenho geral, permitindo inferir que a preservação do conteúdo central da imagem foi importante para a classificação, considerando que houve a remoção das características distintas, presentes nas outras partes da imagem. Por outro lado, em ambas as tabelas, **Random Crop** teve o pior desempenho, indicando que a aleatoriedade na seleção da região recortada pode interferir negativamente no resultado do modelo.

As Figuras 3 e 4 ilustram os resultados de alguns exemplos utilizados no teste dos conjuntos de imagens. Cada figura apresenta dois exemplos de imagens provenientes de suas respectivas bases, acompanhadas de suas transformações correspondentes. Em cada imagem, há uma moldura nas cores vermelha ou verde que indica, respectivamente, erro ou acerto na classificação realizada pelo modelo. Além disso, sobrepõe-se um mapa de calor que ilustra as regiões de maior atenção do modelo: quanto mais vermelha a área, maior a importância atribuída à classificação e quanto mais azul, menor a atenção dedicada àquela região.

A análise do Grad-CAM++ revelou, em ambas as figuras, que as técnicas **Center Crop** e **Crop + Resize** focaram em algumas características faciais, como olhos e boca, para classificar melhor as imagens. Ao passo que as demais técnicas

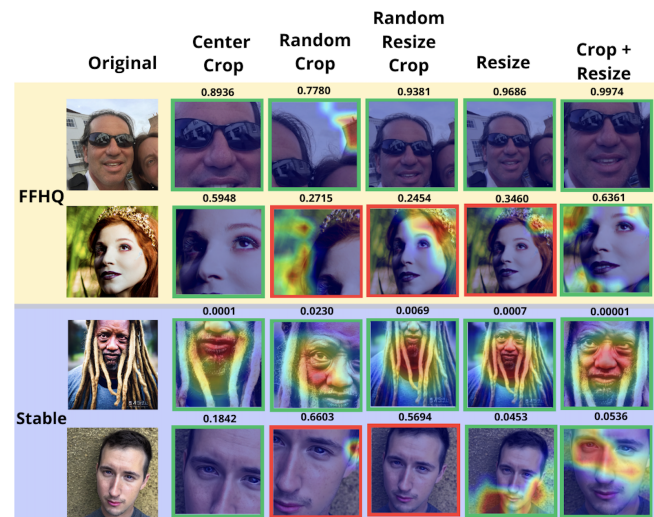


Figure 3. Resultados de alguns exemplos do Conjunto de Imagens 1 (FFHQStable).

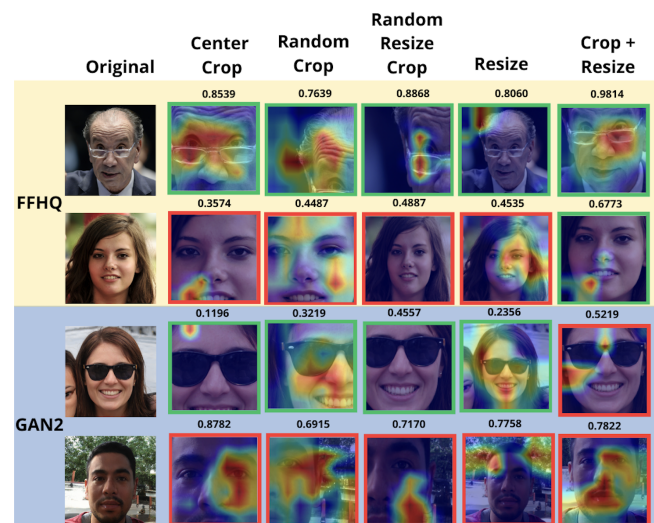


Figure 4. Resultados de alguns exemplos do Conjunto de Imagens 2 (FFHQGAN2).

consideraram também outras partes da imagem, incluindo o fundo. Já as imagens com coloração uniformemente azul indicam que o modelo distribuiu igualmente a importância entre as regiões da imagem durante a classificação.

Table 5. Resultados do teste cruzado do Conjunto de Imagens 1 (FFHQStable).

Técnica	Acurácia	Precisão	Recall	F1 score
Center Crop	0.4491	0.4821	0.8612	0.6182
Random Crop	0.3749	0.4174	0.5235	0.4645
Random Resize Crop	0.4776	0.4975	0.8909	0.6385
Resize	0.4318	0.4692	0.7420	0.5749
Crop + Resize	0.5145	0.5272	0.8619	0.6542

Diferentemente das Tabelas 3 e 4, as Tabelas 5 e 6 apresentam os resultados dos testes cruzados realizados com

Table 6. Resultados do teste cruzado do Conjunto de Imagens 2 (FFHQGAN2).

Técnica	Acurácia	Precisão	Recall	F1 score
Center Crop	0.5137	0.5142	0.9972	0.6785
Random Crop	0.4863	0.5005	0.8982	0.6428
Random Resize Crop	0.5060	0.5107	0.9108	0.6669
Resize	0.5132	0.5140	0.9972	0.6783
Crop + Resize	0.5106	0.5128	0.9766	0.6725

ambas as bases de dados. Enquanto na Tabela 5 foram utilizados os pesos do melhor treinamento/validação obtido com a FFHQGAN2, na Tabela 6 aplicaram-se os pesos do melhor treinamento da FFHQStable.

Em ambos os testes cruzados, observou-se uma queda considerável no desempenho geral, permitindo inferir que o modelo teve dificuldades em generalizar os dados. Esse comportamento pode estar relacionado à natureza distinta das bases: enquanto a GAN2 foi derivada de imagens reais, a Stable foi gerada a partir de descrições textuais, conforme explicado na seção 3.1. Por isso, ambas apresentam artefatos específicos que as diferenciam, respectivamente, como a falta de brilho na pupila e a manipulação no fundo da imagem.

5. Conclusão e Trabalhos Futuros

Os resultados indicam que, dentre as técnicas estudadas, o **Center Crop** e **Crop + Resize** foram os que melhor discriminaram os conteúdos das imagens, permitindo inferir que a preservação das regiões centrais da face foi importante para a classificação, como ilustrado no Grad-CAM++. Ao passo que a técnica **Random Crop** apresentou o pior desempenho, indicando que a aleatoriedade na seleção da região recortada pode interferir negativamente no resultado do modelo, usando informações que não descrevem a face propriamente.

Em ambos os testes cruzados, observou-se uma queda considerável no desempenho geral, permitindo inferir que o modelo estado da arte de aprendizado supervisionado Xception apresentou dificuldades para classificar as imagens reais e sintéticas, especialmente estas últimas.

Como trabalhos futuros, espera-se aplicar o mesmo protocolo de treinamento e validação (incluindo teste cruzado) com outros modelos. Além disso, será explorada uma comparação quantitativa entre os achados obtidos com o Grad-CAM++ e outros métodos XAI, com o objetivo de identificar quais técnicas oferecem explicações mais robustas e interpretáveis para a detecção de imagens sintéticas. Dessa forma, será possível ter uma intuição mais profunda sobre o processo de aprendizado do modelo e sua generalização para novos modelos generativos. Adicionalmente, serão avaliadas a eficiência e as áreas de interesse (AOIs) desses modelos de IA em comparação com seres humanos, por meio de um experimento de percepção visual.

Agradecimentos

Este estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de financiamento 001.

Contribuição dos Autores

Fernanda Goyo Tamanaka: Writing – review & editing, Validation, Methodology, Conceptualization, Data analysis. Lucas Fontes Buzuti: Conceptualization, Methodology, Validation. Carlos Eduardo Thomaz: Writing – review & editing, Methodology, Conceptualization, Project administration, Supervision, Funding acquisition, Validation.

Referências

- [1] GUPTA, P. et al. Generative ai: A systematic review using topic modelling techniques. *Data and Information Management*, Elsevier, p. 100066, 2024.
- [2] KAUSHAL, A.; KUMAR, S.; KUMAR, R. A review on deepfake generation and detection: bibliometric analysis. *Multimedia Tools and Applications*, Springer, p. 1–41, 2024.
- [3] GOODFELLOW, I. et al. Generative adversarial nets. In: GHAHRAMANI, Z. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. v. 27. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.
- [4] MITRA, A. et al. A machine learning based approach for deepfake detection in social media through key video frame extraction. *SN Computer Science*, Springer, v. 2, n. 2, p. 1–18, 2021.
- [5] TOLOSANA, R. et al. Deepfakes detection across generations: Analysis of facial regions, fusion, and performance evaluation. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 110, p. 104673, 2022.
- [6] PAPA, L. et al. On the use of stable diffusion for creating realistic faces: from generation to detection. In: *IEEE. 2023 11th International Workshop on Biometrics and Forensics (IWBF)*. [S.l.], 2023. p. 1–6.
- [7] ROSSLER, A. et al. Faceforensics++: Learning to detect manipulated facial images. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2019. p. 1–11.
- [8] LI, Y. et al. Celeb-df: A large-scale challenging dataset for deepfake forensics. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2020. p. 3207–3216.
- [9] DOLHANSKY, B. et al. The deepfake detection challenge (dfdc) preview dataset. *arXiv preprint arXiv:1910.08854*, 2019.

- [10] LI, Y.; CHANG, M.-C.; LYU, S. In *ictu oculi: Exposing ai created fake videos by detecting eye blinking*. In: *IEEE. 2018 IEEE international workshop on information forensics and security (WIFS)*. [S.l.], 2018. p. 1–7.
- [11] CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1251–1258.
- [12] ROMBACH, R. et al. High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2022. p. 10684–10695.
- [13] KARRAS, T.; LAINE, S.; AILA, T. A style-based generator architecture for generative adversarial networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2019. p. 4401–4410.
- [14] KARRAS, T. et al. Analyzing and improving the image quality of stylegan. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2020. p. 8110–8119.
- [15] KARRAS, T. et al. Alias-free generative adversarial networks. In: *Proc. NeurIPS*. [S.l.: s.n.], 2021.
- [16] SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [17] HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.
- [18] HOWARD, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
- [19] DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [20] CHATTOPADHAY, A. et al. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In: *IEEE. 2018 IEEE winter conference on applications of computer vision (WACV)*. [S.l.], 2018. p. 839–847.
- [21] SCHUHMANN, C. et al. Laion-400m: Open dataset of clip-filtered 400 million image-text pairs. *arXiv preprint arXiv:2111.02114*, 2021.
- [22] KARRAS, T. et al. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2018.
- [23] YU, F. et al. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015.
- [24] BALTRUŠAITIS, T.; ROBINSON, P.; MORENCY, L.-P. Openface: an open source facial behavior analysis toolkit. In: *IEEE. 2016 IEEE winter conference on applications of computer vision (WACV)*. [S.l.], 2016. p. 1–10.