

Uso de Redes Neurais GANs (Generative Adversarial Network) para geração de retratos falados.

Fabiano Borgem Victor,
Orientador: Prof. Dr. Henrique Dezani
E-mail: fabiano.victor@fatec.sp.gov.br

Resumo: Este artigo consiste em gerar imagens utilizando uma estrutura de rede neural denominada GAN (Generative Adversarial Network), que tem a capacidade de aprender a gerar suas próprias imagens de um determinado assunto, visualizando alguns exemplos de imagens. A rede neural é formada por duas redes com papéis diferentes, uma delas será responsável por gerar imagens e a outra será responsável por validar se as imagens geradas pela primeira são imagens válidas ou imagens sem sentido. O treinamento é dado por um conjunto de imagens de exemplo do objeto a ser gerado. Como resultado, foi possível criar novas imagens diferentes das utilizadas como exemplos no banco de imagens. Usando este tipo de rede neural, há uma ampla gama de possibilidades, como gerar fotos, músicas, textos, poemas, entre outras coisas. Na prática, a rede será capaz de imitar tudo o que lhe for ensinado.

Palavras-chave: Aprendizado de Máquina, Rede Neural Gan, Geração de Retratos Falados.

Abstract: *This article consists of generating images using a neural network structure named GAN (Generative Adversarial Network). It has the ability to learn how to generate its own images of a given subject, within just a few examples of similar images. The neural network is formed by two networks with different roles, one of them will be responsible for generating images and the other is responsible for validating whether the images generated by the first one are valid images or meaningless images. The training is given by a set of example images of the object to be generated. As a result, it was possible to create new images different from those used as examples in the image bank. Using the neural network, there is a wide range of possibilities, such as generating photos, music, texts, poems, among other things. In practice, the network will be able to imitate anything that is taught to it.*

Keywords: *Machine Learning, Neural Gan Network, Generation of Spoken Portraits.*

1 Introdução

Atualmente são constantemente desenvolvidas máquinas para o auxílio humano. Elas aperfeiçoam características físicas como potência e velocidade, reduzindo o tempo de trabalho ou aumentando a força. Ainda que auxiliem os seres vivos, uma máquina não substitui completamente um humano pois em algumas atividades e ou situações ainda há a necessidade de um raciocínio ou tomada de decisão mais humanizada.

O aprendizado de máquina foi criado como um meio de imitar a forma que um ser humano aprende e com isso realizar atividades que atualmente só pessoas são capazes. Uma das técnicas criadas ao longo dos anos foram as redes neurais, inspiradas no cérebro humano, ou seja, contendo espécies de neurônios e sinapses.

Alguns tipos de redes usadas com frequência são as para classificação de padrões e regressão. As de classificação possuem a capacidade de identificar e rotular padrões, como por exemplo olhar para uma foto e identificar se ela é de um cachorro ou de um gato, já as de regressão possuem a capacidade de prever um valor futuro com uso de

valores passados, como por exemplo olhar a movimentação passada de preços de ativos na bolsa de valores e prever qual o preço futuro desse ativo.

1.1 Redes Neurais Artificiais

Segundo OSCAR LANDMANN, 05/05/2022, na prática, as redes replicam nas máquinas a maneira como o ser humano pensa. Ou seja, elas permitem que as máquinas aprendam e entendam padrões, adaptem-se a mudanças e imitem processos de pensamento humano em situações que precisam de respostas ou soluções.

Possíveis atividades que uma rede é capaz de realizar:

- Classificação - Consiste na capacidade de interpretação que os seres humanos possuem como por exemplo, visualizar uma imagem e reconhecer objetos ou padrões que a mesma possui.
- Predição(Regressão) - É a capacidade de predição que os seres humanos possuem, como por exemplo prever qual o próximo movimento financeiro visualizando movimentos anteriores em uma bolsa de valores.

1.2 Convolução

Uma forma de extrair ou transformar características importantes da imagem, por meio de multiplicação de matrizes, onde os pixels de uma imagem são multiplicados por uma matriz especial também conhecida como kernel ou filtro.

Segundo Rodrigo Fill e Paulo Sestini, Podemos utilizar filtros para detectar padrões e objetos em imagem, um bom exemplo disso é o de detecção de bordas e contornos em imagens. Há vários tipos de filtro para isso, sendo que um dos mais famosos é o filtro de Sobel usado para detecção de bordas.

1.3 Rede Neural Convolutacional

É uma Rede Neural Artificial que faz uso da convolução para um pré-processamento ou pós-processamento em imagens para extração ou geração de características de uma forma mais otimizada.

Desta maneira a rede precisará atuar apenas nas características mais importantes da imagem, acelerando em muitas vezes o processamento.

1.4 Redes Adversárias Generativas (GANs)

Segundo o Deep Learning Book, são arquiteturas de redes neurais profundas compostas por duas redes colocadas uma contra a outra (daí o nome “adversárias”). Esta é uma das arquiteturas mais recentes e mais fascinantes em Deep Learning, ela possui a capacidade de gerar padrões que a rede aprende, por exemplo, se ela aprende o que é um cachorro ela então terá a capacidade de gerar uma foto de um cachorro.

As GANs foram introduzidas em um artigo de Ian Goodfellow e outros pesquisadores da Universidade de Montreal, incluindo Yoshua Bengio, em 2014. Referindo-se às GANs, o diretor de pesquisa de IA do Facebook, Yann LeCun, chamou o treinamento adversário de “a ideia mais interessante nos últimos 10 anos em Machine Learning”.

O potencial das GANs é enorme porque elas podem aprender a imitar qualquer distribuição de dados, ou seja, as GANs podem ser ensinadas a criar mundos estranhamente semelhantes aos nossos em qualquer domínio: imagens, música, fala, prosa. Elas são artistas robóticos, em certo sentido, e sua produção é impressionante – até comovente.

Em uma virada surreal, a Christie's (famosa rede inglesa de leilões) vendeu um retrato de \$432.000 gerado por uma GAN, com base no código-fonte aberto escrito por Robbie Barrat, de Stanford. Como a maioria dos artistas de verdade, ele não viu nada do dinheiro, que foi para a empresa francesa Obvious.

2 Justificativa

O trabalho se justifica pela importância de se utilizar as GANs para criar uma imagem (retrato falado) o mais semelhante possível à imagem do suspeito de um crime ou alvo de investigação, a partir de características desse indivíduo passada por testemunhas, auxiliando assim na sua captura e punição com mais agilidade.

3 Objetivo(s)

Objetivo Geral:

Criar uma ferramenta que faz uso de Inteligência Artificial para geração de retratos falados com base em características previamente informadas por testemunhas.

Objetivos específicos:

- Contribuir para a otimização de investigações
- Entender possibilidades que a rede neural GAN proporciona
- Gerar faces com apenas algumas informações
- Facilitar a identificação de um possível suspeito

4 Fundamentação Teórica

A ferramenta consiste na criação de duas redes neurais com diferentes papéis, são eles:

- **Discriminativa** - Identificar em uma foto a existência das características informadas pelo usuário;
 - **Generativa** - Gerar imagens com uso das características informadas pelo usuário;
- Elas terão as seguintes estruturas:

4.1 Discriminativa

A rede discriminativa com base em um banco de dados, irá aprender a identificar em fotos cores de cabelos, tamanhos de cabelo, cores de olhos e tonalidades de peles, para assim ensinar a rede generativa a gerar fotos que possuem as mesmas características aprendidas.

4.1.1 Treinamento

Consiste na entrada de uma imagem válida junto às suas características dentro da rede, onde a mesma irá testar as informações e entregará uma saída, se a saída for diferente da esperada para uma imagem válida, então será feita uma retropropagação, para o ajuste da rede, para que a mesma teste corretamente na próxima execução.

4.2 Generativa

A rede gerativa com base no aprendizado da discriminativa, irá aprender a gerar fotos com as características visuais pretendidas.

4.2.1 Treinamento

Consiste na entrada de números aleatórios junto às suas características informadas pela testemunha dentro da rede, onde a mesma irá testar as informações e gerar uma imagem utilizando elas, a saída será testada pela rede discriminativa, se a saída for diferente da esperada para uma imagem válida, então será feita uma retropropagação, para o ajuste da rede, para que a mesma teste corretamente na próxima execução.

4.3 Rede Neural GAN

A combinação das redes generativa e discriminativa forma a Rede Neural nomeada de GAN.

5 Trabalhos Similares

O que é deepfake? Inteligência artificial é usada pra fazer vídeo falso

O deepfake é uma tecnologia que usa inteligência artificial (IA) para criar vídeos falsos, porém “realistas”, de pessoas fazendo ou falando coisas que na realidade elas nunca fizeram ou falaram. A técnica que permite fazer as montagens de vídeo já gerou desde conteúdos pornográficos com celebridades até discursos fictícios de políticos influentes. Circulam agora debates sobre a ética e as consequências da tecnologia, para o bem e para o mal.

O termo deepfake apareceu em dezembro de 2017, quando um usuário do Reddit com esse nome começou a postar vídeos de famosas com conteúdo explícito. Com softwares de deep learning, ele substituiu nos vídeos os rostos originais das pessoas

pelo rosto de famosos. Os casos mais populares foram os das atrizes Gal Gadot e Emma Watson. A expressão deepfake logo passou a ser usada para indicar uma variedade de vídeos editados com machine learning e outras capacidades da IA.

6 Metodologia

Por via de IA foi buscado um meio de gerar faces e assim auxiliar e otimizar investigações policiais, o tipo de pesquisa utilizado foi a exploratória. Baseado em imagens publicadas na plataforma Kaggle de bancos de dados para aprendizado de máquina, serão analisadas 200000 imagens para compreender e aprender padrões e suas características.

7 Desenvolvimento

7.1 Rede Neural Discriminativa:

A rede discriminativa irá possuir uma camada de entrada, uma oculta e por fim uma de saída. A mesma fará uso também de uma rede convolucional para pré-processamento das imagens.

7.2 Rede Neural Generativa:

Por sua vez, a rede generativa irá possuir uma estrutura similar a discriminativa, porém com o número de entradas e saídas diferentes, e fará uso também de uma rede convolucional para um pós-processamento de imagens.

7.3 Rede Neural GAN:

A rede neural GAN, consistirá na combinação da rede neural generativa com a discriminativa, sendo que a generativa será responsável por gerar as imagens, enquanto a discriminativa irá ser responsável pela validação das imagens geradas para assim detectar imagens reais e imagens sem sentido visual.

7.4 Banco de dados:

Para treino da rede, serão utilizadas diversas faces encontradas no banco de imagens CelebFaces Attributes baixado na plataforma Kaggle.

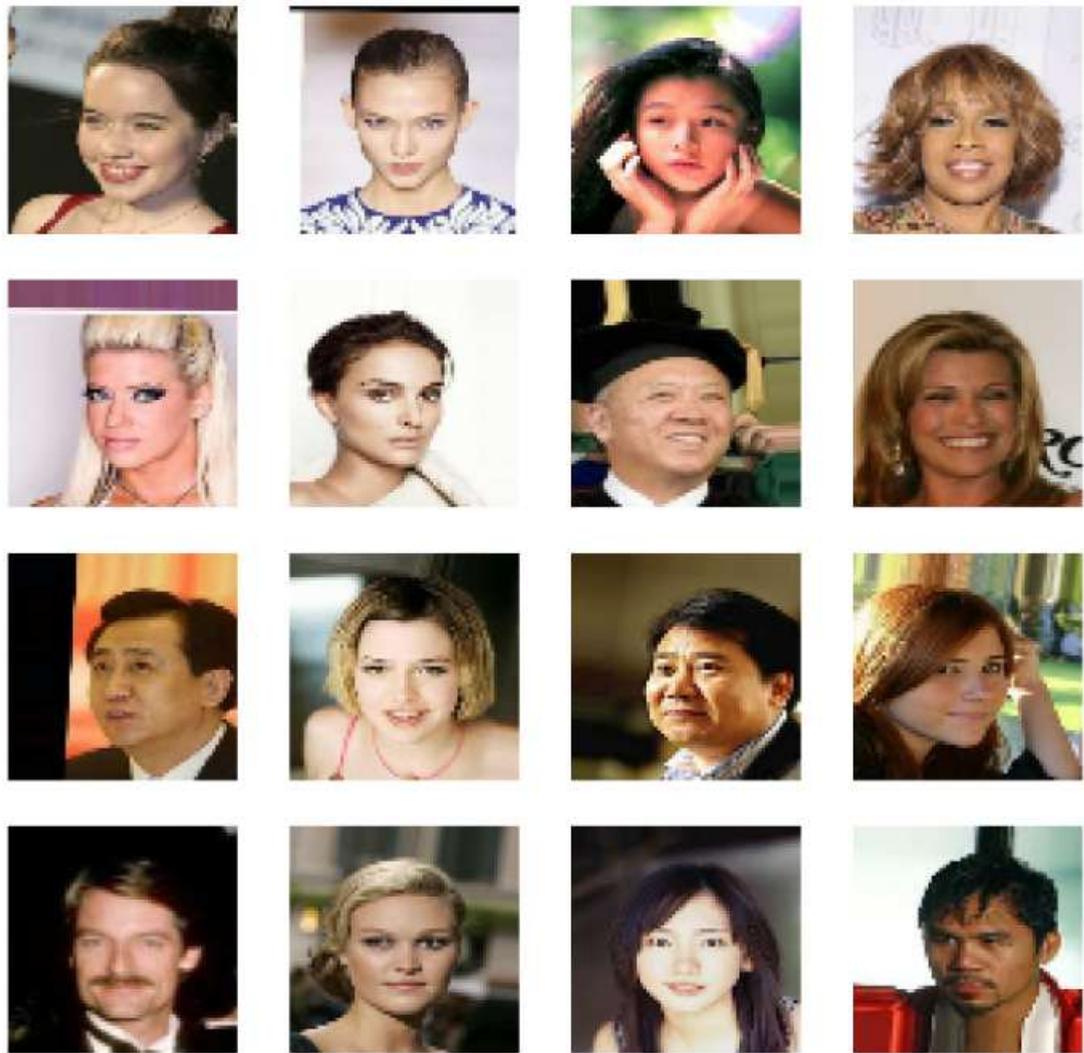


Figura 1. Amostras **CelebFaces Attributes** fonte: <https://www.kaggle.com/datasets/jessicali9530/celeba-dataset>

7.5 Treinamento:

O Treinamento consiste em ensinar a rede neural discriminativa a identificar se uma imagem é válida com base no banco de dados entregue para a mesma, e assim a discriminativa ganhará a capacidade de treinar a rede neural generativa para gerar imagens semelhantes às do banco de dados informado.

8 Resultados e discussões

Após 5000 épocas de treinamento da rede GAN, foi possível gerar faces com uma taxa de erro de 0,0145 na rede discriminativa e 0,9856 na generativa, e assim obter imagens que nunca existiram, puramente criadas por um computador com base na experiência que a rede obteve com o treinamento da mesma.

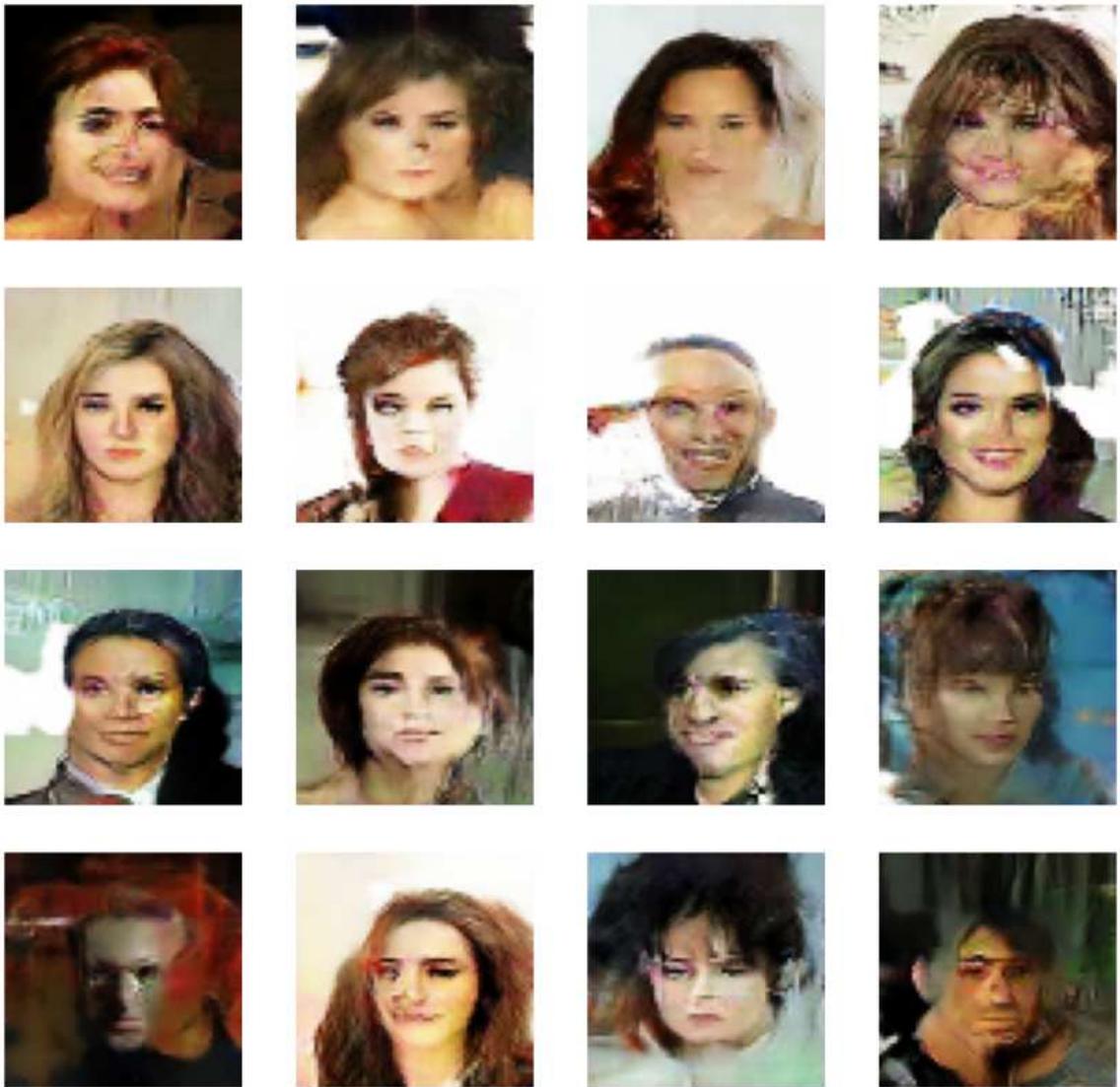
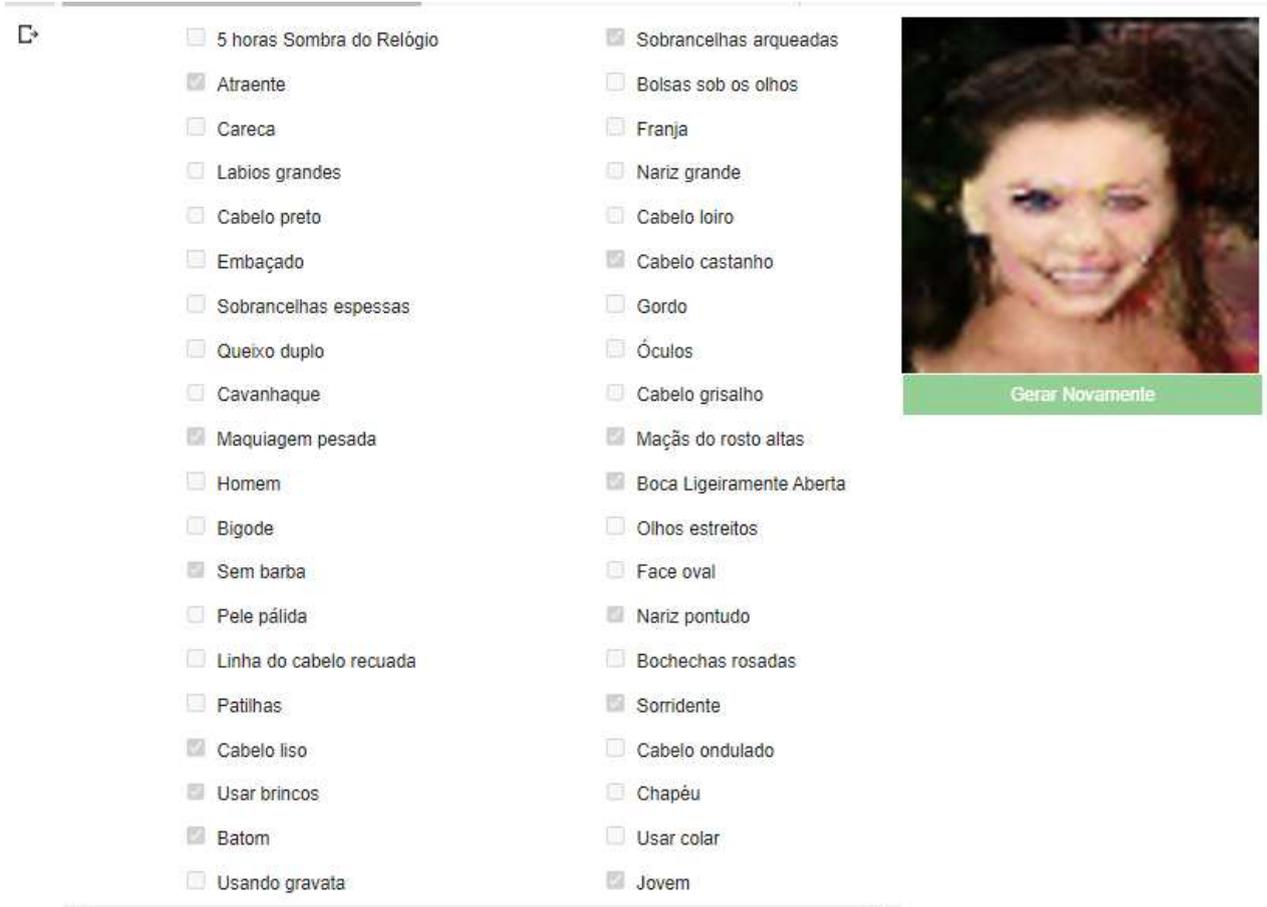


Figura 2. Amostras geradas após treinamento da Rede Neural GAN. Fonte: Próprio Autor

9 Conclusão

Com uso da rede podemos notar um leque imenso de possibilidades, como geração de fotos, músicas, textos, poemas, entre outras coisas. Na prática, a rede irá conseguir imitar qualquer coisa que é ensinado a ela, assim se tornando uma ferramenta poderosa para geração de retratos falados.



<input type="checkbox"/> 5 horas Sombra do Relógio	<input checked="" type="checkbox"/> Sobrancelhas arqueadas
<input checked="" type="checkbox"/> Atraente	<input type="checkbox"/> Bolsas sob os olhos
<input type="checkbox"/> Careca	<input type="checkbox"/> Franja
<input type="checkbox"/> Labios grandes	<input type="checkbox"/> Nariz grande
<input type="checkbox"/> Cabelo preto	<input type="checkbox"/> Cabelo loiro
<input type="checkbox"/> Embaçado	<input checked="" type="checkbox"/> Cabelo castanho
<input type="checkbox"/> Sobrancelhas espessas	<input type="checkbox"/> Gordo
<input type="checkbox"/> Queixo duplo	<input type="checkbox"/> Óculos
<input type="checkbox"/> Cavanhaque	<input type="checkbox"/> Cabelo grisalho
<input checked="" type="checkbox"/> Maquiagem pesada	<input checked="" type="checkbox"/> Maças do rosto altas
<input type="checkbox"/> Homem	<input checked="" type="checkbox"/> Boca Ligeiramente Aberta
<input type="checkbox"/> Bigode	<input type="checkbox"/> Olhos estreitos
<input checked="" type="checkbox"/> Sem barba	<input type="checkbox"/> Face oval
<input type="checkbox"/> Pele pálida	<input checked="" type="checkbox"/> Nariz pontudo
<input type="checkbox"/> Linha do cabelo recuada	<input type="checkbox"/> Bochechas rosadas
<input type="checkbox"/> Patilhas	<input checked="" type="checkbox"/> Sorridente
<input checked="" type="checkbox"/> Cabelo liso	<input type="checkbox"/> Cabelo ondulado
<input checked="" type="checkbox"/> Usar brincos	<input type="checkbox"/> Chapéu
<input checked="" type="checkbox"/> Batom	<input type="checkbox"/> Usar colar
<input type="checkbox"/> Usando gravata	<input checked="" type="checkbox"/> Jovem

Gerar Novamente

Figura 3. Ferramenta desenvolvida para geração de retratos falados, desenvolvida com uso da rede. Fonte: Próprio Autor.

10 Referências

CABRAL, Isabela. **O que é deepfake? Inteligência artificial é usada para fazer vídeo falso?** Inteligência artificial é usada para fazer vídeo falso: Inteligência artificial é usada para fazer montagem com famosos e políticos, substituindo rostos e vozes em vídeos realistas; a novidade traz preocupações éticas. O que é deepfake? Techtudo, p. 1, 28 jul. 2018. Disponível em: <https://www.techtudo.com.br/noticias/2018/07/o-que-e-deepfake-inteligencia-artificial-e-usada-para-fazer-videos-falsos.ghtml>. Acesso em: 20 mar. 2022.

Capítulo 54 - Introdução às Redes Adversárias Generativas (GANs - Generative Adversarial Networks). Disponível em: <https://www.deeplearningbook.com.br/introducao-as-redes-adversarias-generativas-gans-generative-adversarial-networks/>

CelebFaces Attributes (CelebA) Dataset. Disponível em: <https://www.kaggle.com/datasets/jessicali9530/celeba-dataset>. Acesso em: 6 jun. 2022.

LANDMANN, OSCAR. **O que são Redes Neurais?** Disponível em: <https://opencadd.com.br/o-que-sao-redes-neurais>. Acesso em: 7 jul. 2022.

RANGEL, R. F. **Visão Computacional — O que é convolução?** Disponível em: <https://medium.com/turing-talks/vis%C3%A3o-computacional-o-que-%C3%A9-convolu%C3%A7%C3%A3o-ad709f7bd6b0>. Acesso em: 14 jun. 2022.

ROMERO, John. **Introdução às Redes Adversárias Generativas (GANs – Generative Adversarial Networks):** Você pode não achar que os programadores são artistas, mas a programação é uma profissão extremamente criativa. É a criatividade baseada em lógica. – John Romero. Introdução às Redes Adversárias Generativas, Deep Learning Book, p. 1, 7 out. 2020. Disponível em: [https://www.deeplearningbook.com.br/introducao-as-redes-adversarias-generativas-gans-generative-adversarial-networks/#:~:text=Redes%20Advers%C3%A1rias%20Generativas%20\(GANs\)%20s%C3%A3o.o%20nome%20%E2%80%9Cadvers%C3%A1rias%E2%80%9D](https://www.deeplearningbook.com.br/introducao-as-redes-adversarias-generativas-gans-generative-adversarial-networks/#:~:text=Redes%20Advers%C3%A1rias%20Generativas%20(GANs)%20s%C3%A3o.o%20nome%20%E2%80%9Cadvers%C3%A1rias%E2%80%9D). Acesso em: 20 mar. 2022.