

# Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation

Jadson Crislan Santos Costa  
jcsc@ic.ufal.br

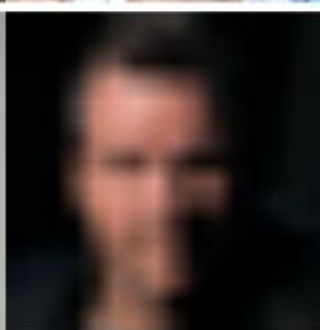
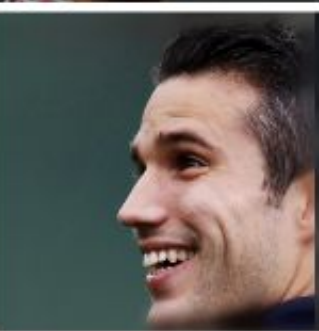


# 01

## Artigo

# Introduction

- ▲ Generative Adversarial Networks (GANs), uma classe de rede neural de geração de imagens.
- ▲ Propõe uma nova arquitetura de StyleGAN introduzindo um encoder na arquitetura.



# Metodologia

## Face Frontalization

- Utiliza as bases de dados CelebA-HQ e FFHQ.
- Espelha a imagem e força o algoritmo a encontrar a imagem frontalizada tanto a partir da original quanto a partir da imagem espelhada, para a rede neural aprender apenas um padrão de imagem frontal.

## Conditional Image Synthesis

- Utiliza uma base de dados de CelebA-HQ, AFHQ Cat e AFHQ Dog.
- Para cada imagem aplica um filtro de "esboço a lápis".
- Em seguida, aplica o método de simplificação de esboço, e assim, resulta em imagens que lembram esboços feitos à mão.

# Resultados: Face Frontalization



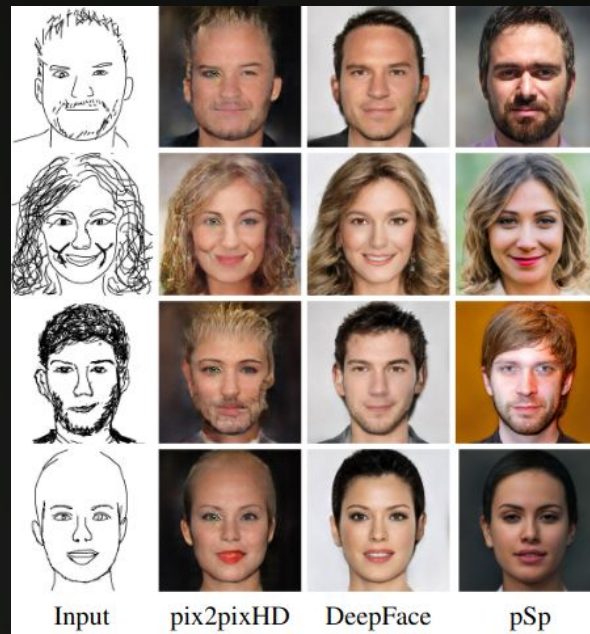
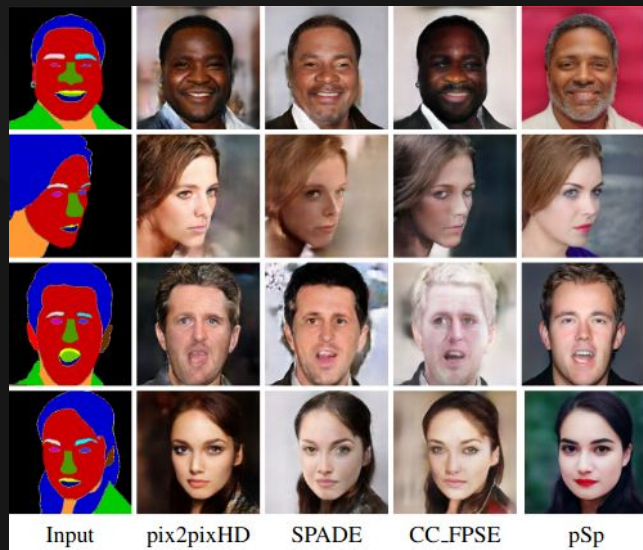
| Method | Similarity                | Runtime |
|--------|---------------------------|---------|
|        | 90° - 70° - 50° - 30°     |         |
| R&R    | 0.34 - 0.56 - 0.66 - 0.7  | 1.5s    |
| pSp    | 0.32 - 0.52 - 0.60 - 0.63 | 0.1s    |



# Resultados: Conditional Image Synthesis

| Task         | p2p    |
|--------------|--------|
| Segmentation | 94,72% |
| Sketch       | 93,34% |

Percepção humana  
(8400 pessoas)





# 02

## Projeto



# Descrição do projeto



## Problema

Imagens e documentos borrados.



## Pré processamento de imagem

Diminuir ruídos e simplificar a imagem.



## Data augmentation

Aumentar a base de dados.



## Aplicar o algoritmo

Criação de uma rede neural encoder.

# 1. Problema

- ▲ Documentos com marca de café.
- ▲ Documentos e imagens antigas.
- ▲ Documentos amassados.
- ▲ Extração de OCR.

## 2. Pré processamento de imagem

- ▲ Transformar imagem para escala de cinza
- ▲ Filtro gaussiano (3\*3)
- ▲ Filtro laplaciano
- ▲ Inverter as cores
- ▲ Limiarização
- ▲ Mudar o tamanho da imagem para 172 x 360

There are several classic spatial filters for reducing or eliminating noise from images. The mean filter, the median filter and the closing filter are frequently used. The mean filter is a lowpass or smoothing filter that replaces each pixel with the neighborhood mean. It reduces the image noise but blurs the image edges. The median filter calculates the median of the pixel neighborhood for each pixel, thereby reducing the blurring effect. Finally, the opening closing filter is a mathematical morphology operation that combines the same number of erosion and dilation morphological operations to eliminate small objects from images.

The main goal was to train a neural network in a supervised learning scheme to obtain a clean image from a noisy one. In this particular case, it was much easier to obtain a noisy image from a clean one than to clean a subset of noisy images.

There are several classic spatial filters for reducing or eliminating noise from images. The mean filter, the median filter and the closing filter are frequently used. The mean filter is a lowpass or smoothing filter that replaces each pixel with the neighborhood mean. It reduces the image noise but blurs the image edges. The median filter calculates the median of the pixel neighborhood for each pixel, thereby reducing the blurring effect. Finally, the opening closing filter is a mathematical morphology operation that combines the same number of erosion and dilation morphological operations to eliminate small objects from images.

The main goal was to train a neural network in a supervised learning scheme to obtain a clean image from a noisy one. In this particular case, it was much easier to obtain a noisy image from a clean one than to clean a subset of noisy images.

## Exemplo:

There are several classic spatial filters for reducing or eliminating noise from images. The mean filter, the median filter and the closing filter are frequently used. The mean filter is a lowpass or smoothing filter that replaces each pixel with the neighborhood mean. It reduces the image noise but blurs the image edges. The median filter calculates the median of the pixel neighborhood for each pixel, thereby reducing the blurring effect. Finally, the opening closing filter is a mathematical morphology operation that combines the same number of erosion and dilation morphological operations to eliminate small objects from images.

The main goal was to train a neural network in a supervised learning scheme to obtain a clean image from a noisy one. In this particular case, it was much easier to obtain a simulated noisy image from a clean one than to clean a subset of noisy images. The process for obtaining simulated noisy images was as follows: This process requires images of the background and the acquisition forms, which were obtained by printing.

### 3. Data augmentation

- ▲ Criar imagens com ruídos (gauss, salt and pepper, poisson, speckle).
- ▲ Espelhamento das imagens.
- ▲ Rotacionar as imagens.
- ▲ Adicionar novas imagens fora do padrão da base.
- ▲ Base de dados vai de 288 imagens para 2592

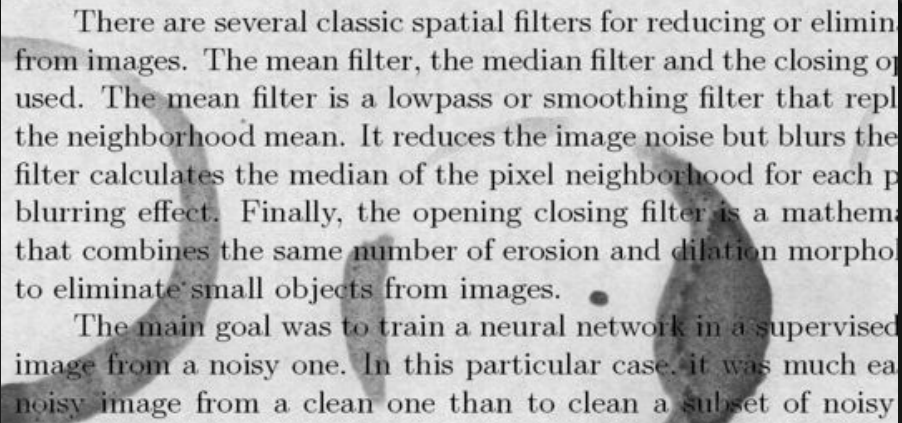
There are several classic spatial filters for reducing frequency noise from images. The mean filter, the closing opening filter are frequently used. The mean smoothing filter that replaces the pixel values with the average of the pixel values in the neighborhood. It reduces the image noise but blurs the image edges. The median filter that replaces the pixel value with the median of the pixel neighborhood for each pixel, which reduces the blurring effect. Finally, the opening closing filter that combines the same number of morphological operations in order to eliminate small features.

The main goal was to train a neural network in a obtain a clean image from a noisy one. In this particular case, it was easier to obtain a simulated noisy image from a clean image than from a subset of noisy images. The process for obtaining simulated noisy images follows this scheme: This process requires images of objects (in this case, rectangles) of the acquisition forms, which were obtained from a clean image.



## 4. Aplicar o algoritmo

- ▲ Saida  $540 * 258 = 139,320$ .
- ▲ 5 camadas
- ▲ 168,047,480
- ▲ Loss: BinaryCrossentropy
- ▲ Optimizer: ADAM
- ▲ Early Stop
- ▲ Rodou por 119 epocas
- ▲ Loss: 0.0868 treino
- ▲ Loss: 0.0951 teste



There are several classic spatial filters for reducing or eliminating noise from images. The mean filter, the median filter and the closing operation are commonly used. The mean filter is a lowpass or smoothing filter that replaces each pixel with the neighborhood mean. It reduces the image noise but blurs the image. The median filter calculates the median of the pixel neighborhood for each pixel, which reduces the blurring effect. Finally, the opening closing filter is a mathematical morphology operation that combines the same number of erosion and dilation morphological operations to eliminate small objects from images.

The main goal was to train a neural network in a supervised learning task to clean an image from a noisy one. In this particular case, it was much easier to clean a noisy image from a clean one than to clean a subset of noisy

There are several classic spatial filters for reducing or eliminating noise from images. The mean filter, the median filter and the closing operation are commonly used. The mean filter is a lowpass or smoothing filter that replaces each pixel with the neighborhood mean. It reduces the image noise but blurs the image. The median filter calculates the median of the pixel neighborhood for each pixel, which reduces the blurring effect. Finally, the opening closing filter is a mathematical morphology operation that combines the same number of erosion and dilation morphological operations to eliminate small objects from images.

The main goal was to train a neural network in a supervised learning task to clean an image from a noisy one. In this particular case, it was much easier to clean a noisy image from a clean one than to clean a subset of noisy



03

# Criação de uma API

- ▲ JadsonKris/Encoder ([github.com](https://github.com))
- ▲ Cada chamada demora aproximadamente 10 segundos para ser executada (i5-7200u).
- ▲ Melhorou a performance da api Tesseract para OCR.
- ▲ Modelo ocupa mais de 300mb no formato h5.
- ▲ Mais de 20 segundos para carregar o modelo.

# Referências

1. Richardson, Elad, et al. "Encoding in style: a stylegan encoder for image-to-image translation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.



Obrigado!