

Aluno: Rafael Tatsuki Takagi
Orientador: Prof. Associado Aparecido Nilceu Marana

Utilização de Redes Neurais para a criação de um Gerador e Editor de Faces

Sumário

- Introdução
- Fundamentação Teórica
- Materiais e Tecnologias
- Desenvolvimento
- Conclusão

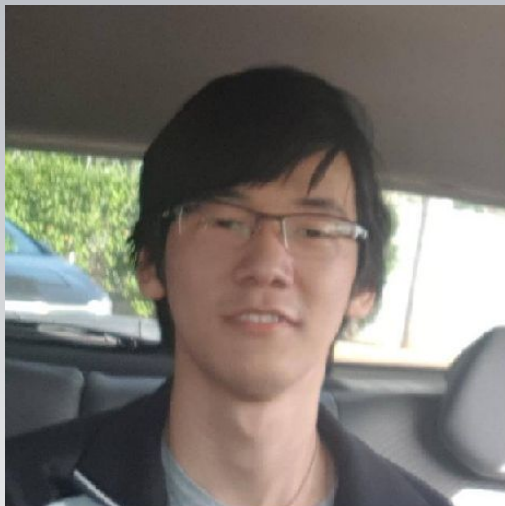
1

Introdução

Problemas

Uma pessoa pode facilmente criar uma imagem mental de uma face a partir de algumas informações de sua composição.

Para um computador isso é um desafio considerável pois este consegue “enxergar” e “interpretar” apenas um conjunto de números a partir de uma imagem.



O que um humano vê

```
ff d8 ff e1 00 16 45 78 69 66 00 00 4d 4d 00 2a .....Exif..MM.*
00 00 00 08 00 00 00 00 00 00 ff ed 00 84 50 68 .....Ph
6f 74 6f 73 68 6f 70 20 33 2e 30 00 38 42 49 4d .....otoshop 3.0.8BIM
04 04 00 00 00 00 00 67 1c 02 28 00 62 46 42 4d .....g..(.bFBM
44 30 31 30 30 30 61 61 36 30 33 30 30 30 30 37 D01000aa60300007
63 31 34 30 30 30 38 62 32 32 30 30 30 30 39 c1400008b2200009
65 32 33 30 30 30 66 36 32 34 30 30 30 30 65 e230000f6240000e
35 32 36 30 30 30 32 61 33 38 30 30 30 30 31 52600002a3800001
32 33 64 30 30 30 37 33 33 66 30 30 30 30 65 23d0000733f0000e
62 34 31 30 30 30 30 30 36 65 30 30 30 30 00 b410000006e0000.
ff e2 02 1c 49 43 43 5f 50 52 4f 46 49 4c 45 00 ....ICC_PROFILE.
01 01 00 00 02 0c 6c 63 6d 73 02 10 00 00 6d 6e .....lcms....mn
74 72 52 47 42 20 58 59 5a 20 07 dc 00 01 00 19 trRGB XYZ .....
00 03 00 29 00 39 61 63 73 70 41 50 50 4c 00 00 (...).9acspAPPL..
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
00 00 00 00 00 00 00 00 f6 d6 00 01 00 00 00 00 .....
d3 2d 6c 63 6d 73 00 00 00 00 00 00 00 00 00 .....-lcms.....
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
00 00 00 00 00 0a 64 65 73 63 00 00 00 fc 00 00 .....desc.....
00 5e 63 70 72 74 00 00 01 5c 00 00 00 0b 77 74 .^cptt...\.wt
70 74 00 00 01 68 00 00 00 14 62 6b 70 74 00 00 pt...h....bkpt..
01 7c 00 00 00 14 72 58 59 5a 00 00 01 90 00 00 .|....xXYZ.....
00 14 67 58 59 5a 00 00 01 a4 00 00 00 14 62 58 ..gXYZ.....bX
```

O que um computador vê

Objetivo Geral

Desenvolver uma ferramenta para possibilitar a criação e edição de faces a partir de parâmetros dados pelo usuário, como largura da face, posição dos olhos e sobrancelhas e expressões faciais.

Tal ferramenta não deverá exigir um alto poder computacional, devendo rodar sem travamentos em um computador pessoal com componentes razoáveis e sem placa de vídeo dedicada.

Objetivos Específicos

- Estudar Redes Neurais e Auto-Encoders;
- Estudar processamento de imagens digitais;
- Desenvolver um auto-encoder para a criação e edição de faces;
- Obter um banco de imagens de faces que podem ser utilizadas;
- Treinar o auto-encoder desenvolvido;
- Avaliar o desempenho do auto-encoder e a qualidade das imagens geradas por ele.

2

Fundamentação Teórica

Inteligência Artificial

Russell e Norvig (2009) definem Inteligência Artificial como o campo que estuda “agentes inteligentes”: qualquer dispositivo que observa seu ambiente e realiza ações que maximizam suas chances de sucesso em alcançar determinado objetivo.

Aprendizado de Máquina

Aprendizado de Máquina é um ramo de pesquisa da Inteligência Artificial que aborda problemas de maneira diferente, treinando um modelo com base em uma série de exemplos, com um algoritmo sendo usado para modificá-lo e otimizá-lo quando preciso (ALPAYDIN, 2009).

A hierarquia de conceitos possibilita ao computador aprender conceitos complexos a partir de outros mais simples. Por este motivo chamamos essa aproximação de Aprendizado Profundo (GOODFELLOW; BENGIO; COURVILLE, 2016).

PCA - Análise de Componente Principal

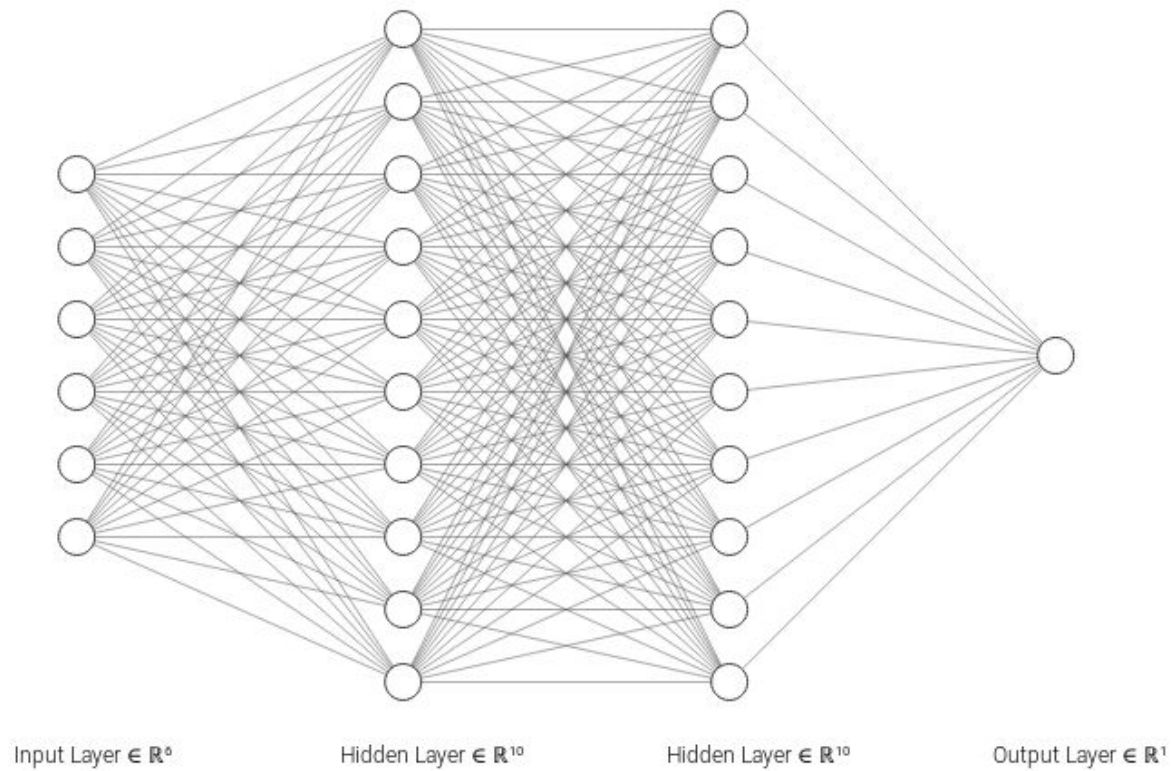
É uma técnica estatística utilizada para a extração de informações a partir de um conjunto de observações.

O PCA é definido de forma que o primeiro componente principal possui a maior variância possível no conjunto inicial, e cada componente seguinte possui a maior variância sob a restrição de ser ortogonal a todos os componentes anteriores.

Redes Neurais

Uma Rede Neural Artificial é um modelo computacional que é baseado em como o cérebro humano processa informação (KARN, 2016).

A arquitetura de uma rede neural é baseada no funcionamento de neurônios artificiais que recebem estímulos de entrada e ativam se estes estímulos forem significantes o suficiente, repassando o estímulo processado para outros neurônios conectados.



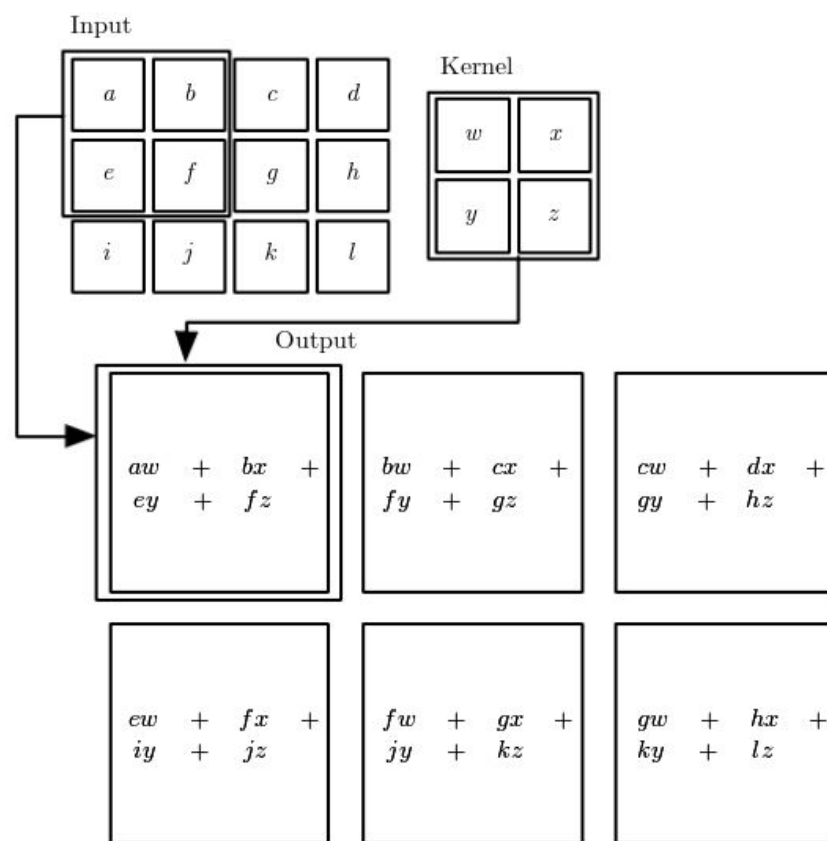
Rede Neural

Redes Neurais Convolucionais

São um tipo especializado de redes neurais para topologias de grade. Como seu nome sugere, esse tipo de rede neural utiliza uma operação linear denominada convolução em pelo menos uma de suas camadas.

Redes Neurais Convolucionais

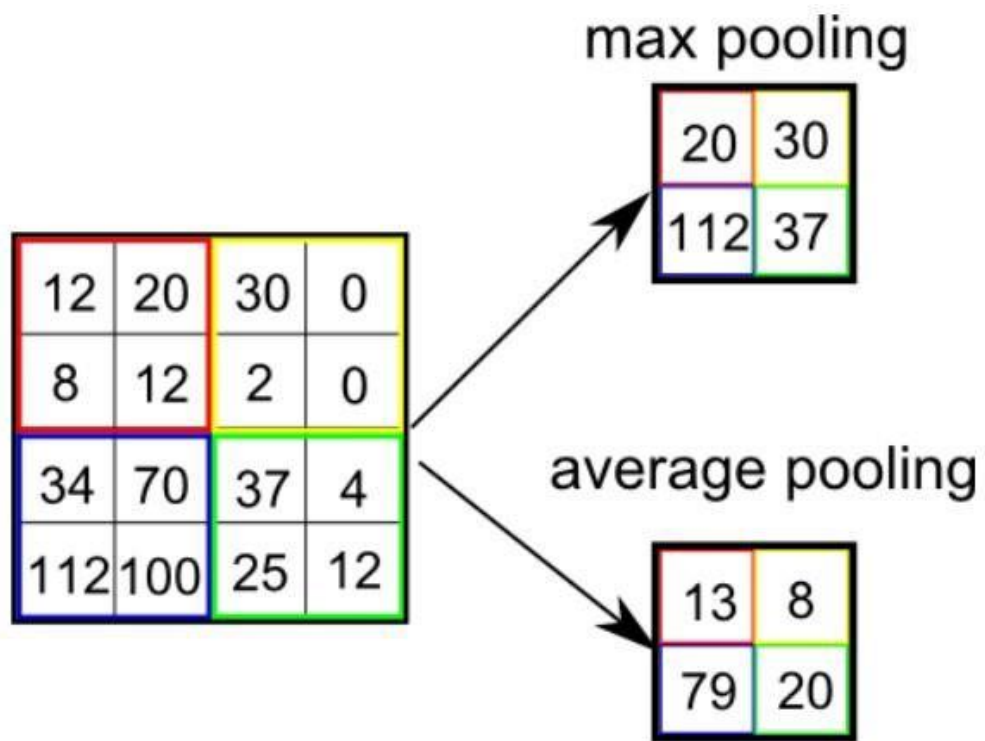
Na camada de convolução cada neurônio possui um *kernel* associado que será aplicado em sua entrada. Caso a entrada seja uma imagem, é comumente utilizado uma função de ativação ReLu nas saídas para eliminar valores negativos.



Convolução

Redes Neurais Convolucionais

Geralmente as redes neurais convolucionais possuem uma camada de *pooling* logo após a camada de convolução para diminuir o poder computacional necessário para processar as saídas.

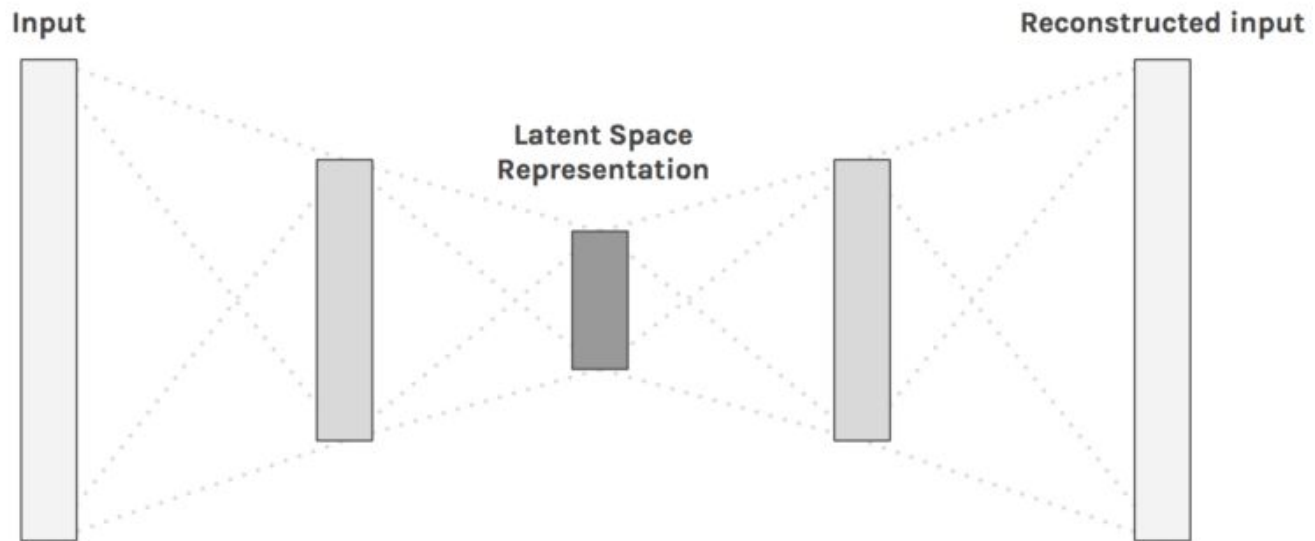


Pooling

Auto-Encoder

Um auto-encoder é um tipo de Rede Neural Artificial que é basicamente constituído pela união entre um encoder e um decoder.

O principal uso de um auto-encoder é de aprender como comprimir dados em código e reconstruir os dados originais a partir do código com o mínimo de perda após a passagem dos dados por um gargalo: o espaço latente (LADJAL; NEWSON; PHAM, 2019).



Auto-Encoder

3

Métodos e Materiais

Conjunto de Treinamento

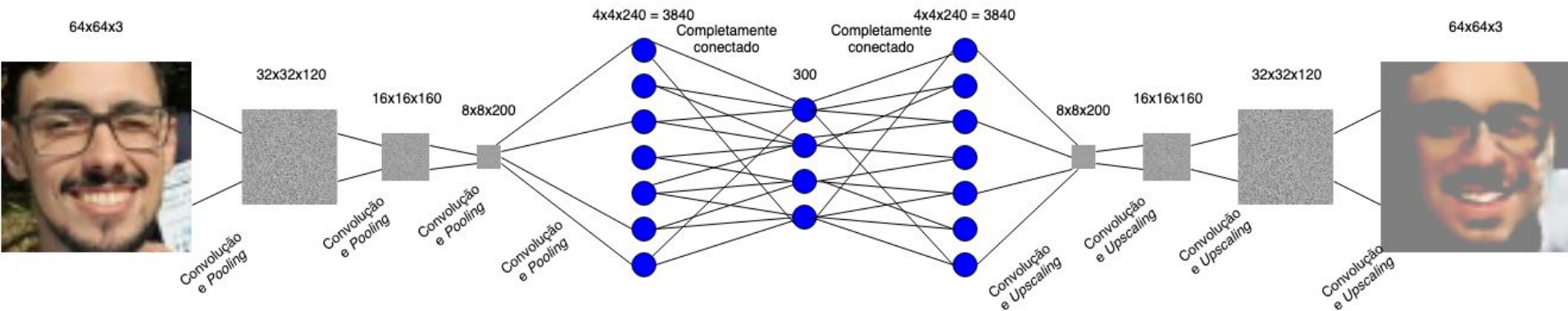
Famous Birthdays (famousbirthdays.com): não é um conjunto de dados formal, porém possui imagens de milhares de celebridades. Cada dia do ano mostra as 48 celebridades mais famosas nascidas no dia, para um potencial de $366 \times 48 = 17520$ imagens. Como cada celebridade aparece apenas uma vez, a variação de faces dentre as imagens é alta.

Auto-Encoder

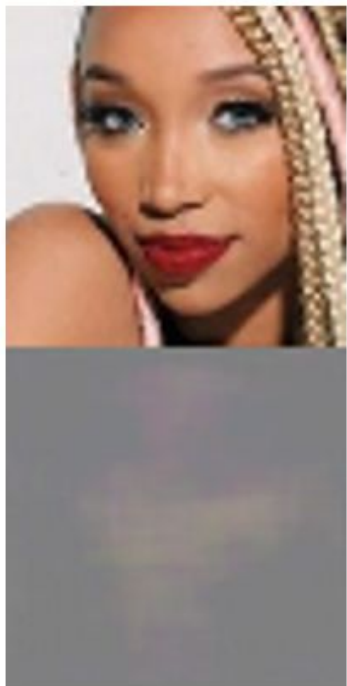
Possui duas funções neste trabalho:

1. Diminuir a quantidade de dimensões da imagem para evitar que o PCA descarte muitas características e também para que o custo de processamento seja menor;
2. Ao passar as imagens pelas camadas de convolução a rede neural passa a ter um conhecimento espacial de como os pixels se relacionam entre si.

Foi definido que o auto-encoder teria quatro camadas de encoder e quatro camadas de decoder após alguns testes.



O Auto-Encoder



Modelo gerado na época 2



Modelo gerado na época 182000

Modelos gerados pelo auto-encoder

Encoder e PCA

Após passado o conjunto de treinamento pelo auto-encoder, este gera um modelo que o encoder usa para codificar todo o conjunto em um vetor denso.

O PCA é então aplicado neste vetor denso para a obtenção dos autovetores e autovalores.

TensorFlow

TensorFlow (ABADI et al., 2015) é uma biblioteca de código aberto multiplataforma para computação numérica e aprendizado de máquina utilizando grafos de fluxo de dados. Ele reúne uma gama de modelos e algoritmos de aprendizado de máquina e aprendizado profundo.

- ❑ Constituído por uma base em C++ e uma interface de mais alto nível em Python;
- ❑ Se baseia na ideia de grafos, onde o usuário utiliza o TensorFlow para construir grafos onde em cada nó uma operação matemática é realizada e as arestas são tensores comunicados entre os nós.

FaceRecognition

Face Recognition é uma biblioteca em código aberto de reconhecimento facial que não necessita do uso do TensorFlow, o que o torna mais leve e menos exigente de poder computacional. A biblioteca se encontra disponível em <https://github.com/ageitgey/face_recognition>.

4

Desenvolvimento

Tratamento do Conjunto de Treinamento

Uma normalização das imagens do conjunto de treinamento foi necessária visto que:

- As imagens do site Famous Birthdays possuíam faces em diferentes posições;
- Algumas imagens eram em escala de cinza;

A normalização se baseou na eliminação de imagens em escala de cinza do conjunto, do reposicionamento das faces nas imagens e o redimensionamento destas.

Após o tratamento das imagens, 12095 sobreviveram para serem utilizadas posteriormente.

Tratamento do Conjunto de Treinamento

Para determinar se uma imagem estava em escala de cinza foi aplicada uma função na imagem.

$$R_{norm} = \frac{R - \mu_R}{\delta_R}$$

$$G_{norm} = \frac{G - \mu_G}{\delta_G}$$

$$B_{norm} = \frac{B - \mu_B}{\delta_B}$$

$$I_{sat} = (R_{norm} - G_{norm})^2 + (R_{norm} - B_{norm})^2 + (G_{norm} - B_{norm})^2$$

A imagem é considerada como estando em escala de cinza e descartada caso I_{sat} for menor que 0,04.



Imagem Original



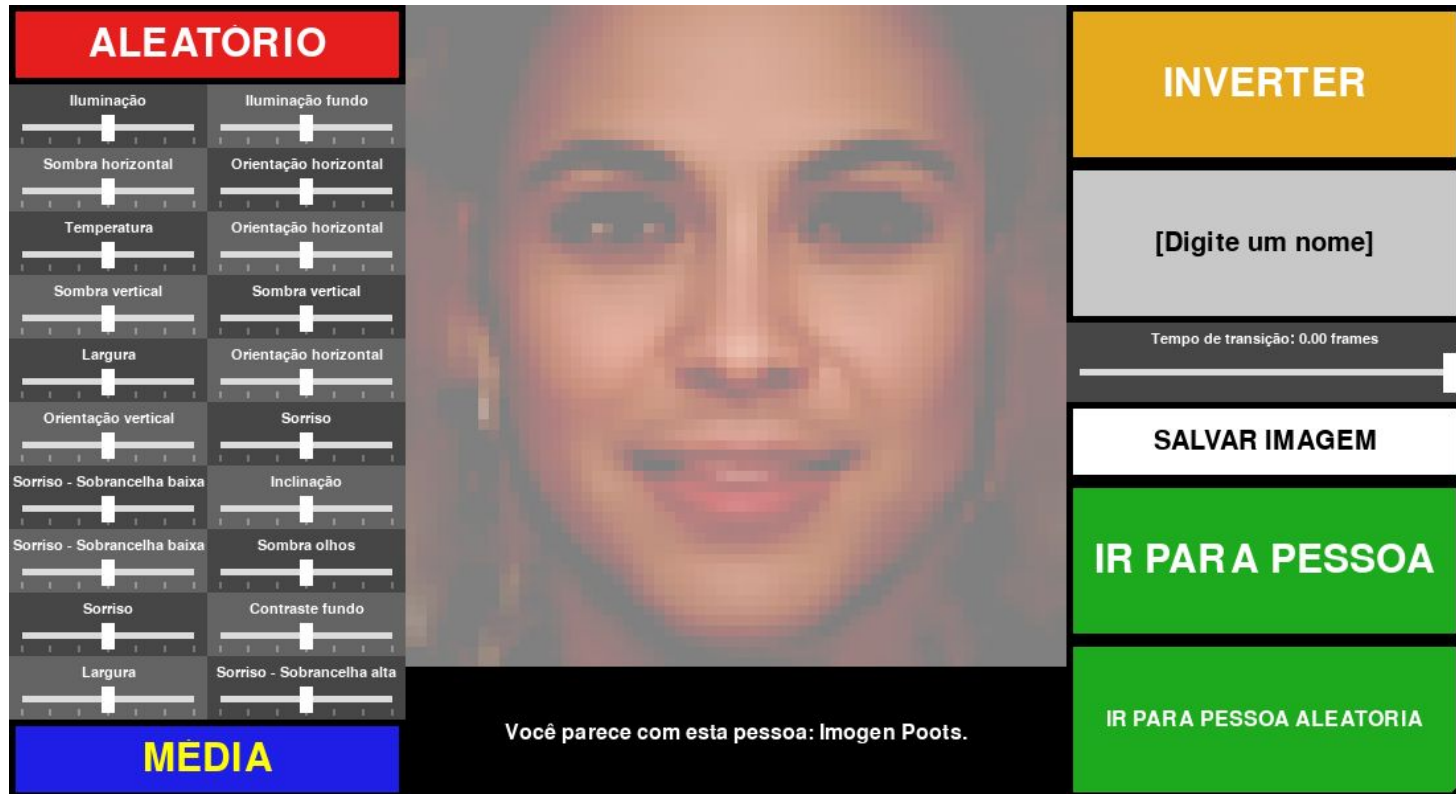
Imagem Tratada

Imagem antes e depois do tratamento

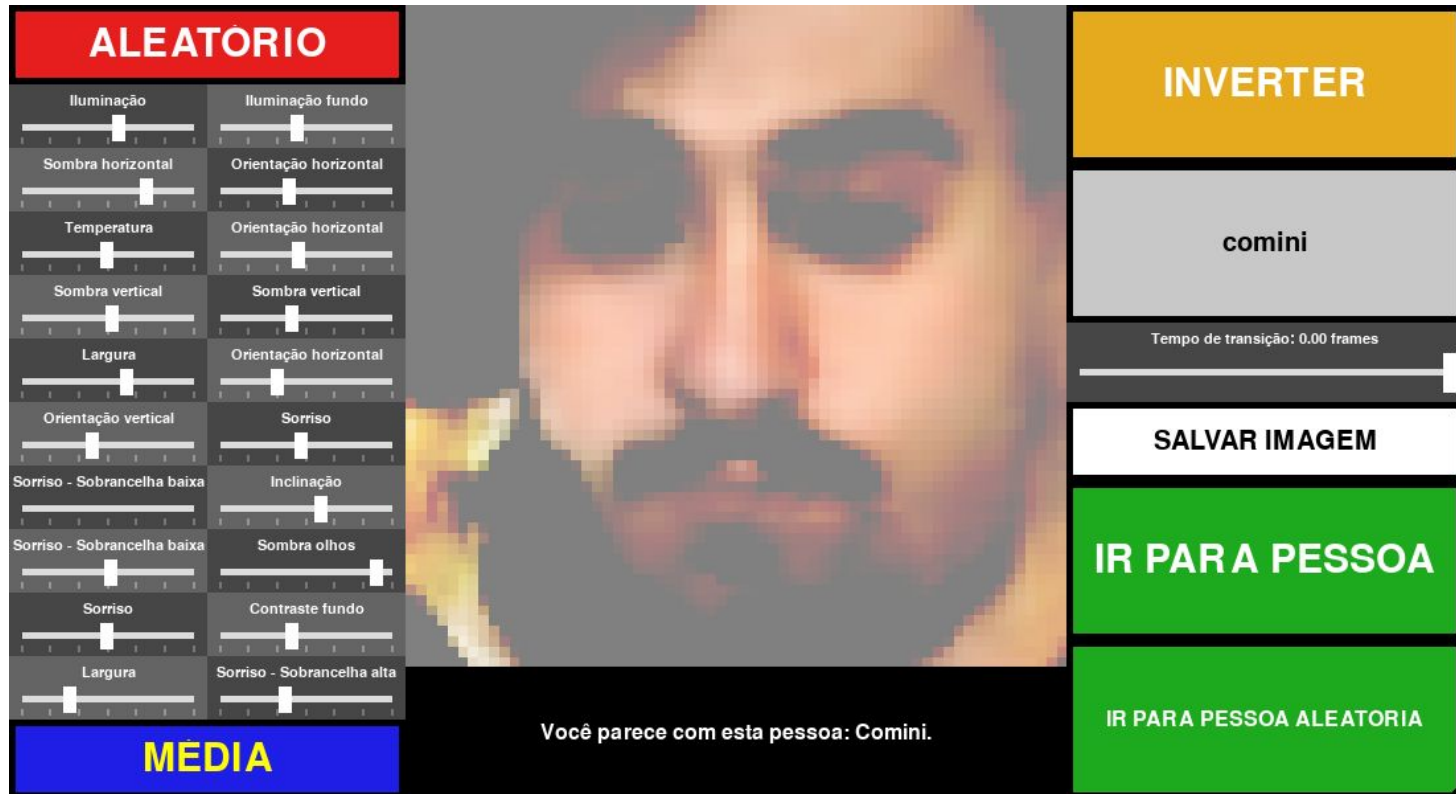
Auto-Encoder, Encoder e PCA

O auto-encoder foi treinado com lotes de 200 imagens por época durante 182000 épocas, com a taxa de perda flutuando entre 0,62 e 0,69.

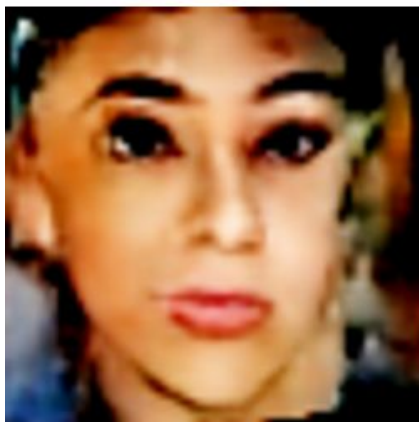
O modelo gerado pelo auto-encoder foi logo utilizado para codificar todo o conjunto em um vetor denso o qual foi aplicado o PCA para a obtenção dos autovetores e autovalores.



Imagens da aplicação final



Imagens da aplicação final



Faces geradas por meio do botão “ALEATÓRIO”

5

Conclusão

Conclusão

- A aplicação teve sucesso em gerar e editar faces;
- Vários componentes obtidos pelo PCA são responsáveis por alterar as mesmas características;
- Percebe-se uma certa quantidade de ruído nas imagens, possivelmente por causa da baixa quantidade de dimensões utilizadas;
- Percebe-se também que há um viés para a geração de faces com traços mais femininos, possivelmente pelo motivo de que o conjunto de treinamento possuía mais imagens de mulheres.

Trabalhos Futuros

- Utilização de outros bancos de dados para o treinamento;
- Utilização de outras técnicas como as GANs;
- A não utilização de um auto-encoder, aplicando o PCA diretamente nas imagens.

Referências

ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANÉ, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCKE, V.; VASUDEVAN, V.; VIÉGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Software available from tensorflow.org. Disponível em: <<https://www.tensorflow.org/>>.

Referências

ALPAYDIN, E. Introduction to Machine Learning. [s.n.], 2009. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=TtrxCwAAQBAJ&oi=fnd&pg=PR7&dq=machine+learning+introduction&ots=T5elKJ-7rR&sig=PSu1GHskI1jQwvBmce_4mdecQc4#v=onepage&q=introduction&f=false>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. MIT Press, 2016. Disponível em: <<http://www.deeplearningbook.org>>.

KARN, U. A Quick Introduction to Neural Networks. 2016. Disponível em: <<https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks>>.

LADJAL, S.; NEWSON, A.; PHAM, C. A pca-like autoencoder. CoRR, abs/1904.01277, 2019. Disponível em: <<http://arxiv.org/abs/1904.01277>>.

RUSSELL, S. J.; NORVIG, P. Artificial Intelligence: A Modern Approach. 3. ed. [S.l.]: Prentice Hall, 2009.

“

“É meio que uma lei da natureza. O objetivo que alguém almeja raramente pode ser alcançado por um caminho simples.”

- Konosuke Matsushita