

Projeto Demonstrativo 6: Reconhecimento de Faces

Pedro Henrique Luz de Araujo Raphael Soares Ramos

Universidade de Brasília

{pedrohluzaraujo, raphael.soares.1996}@gmail.com

6 de dezembro de 2018

- 1 Introdução
 - Motivação
 - Revisão de técnicas
 - Dataset
- 2 Extrator fixo
 - Modelo
 - Experimentos
 - Resultados
- 3 Pipeline de Detecção e Reconhecimento
 - Datasets
 - Pipeline
 - Previsões e Resultados
- 4 Conclusão

Introdução

Reconhecimento de Faces



The image is a composite illustrating a facial recognition search process. The background is a photograph of four men outdoors. Overlaid on the image are several UI elements:

- Found:** A label with a white bounding box around a man wearing a red hat and sunglasses.
- Searching.....:** A label with three white bounding boxes around the other three men in the photo.
- Search Results Window:** A dark grey window titled "You (and 15 other people) say this is Henrik Andersen (B)". It lists "Also known as... (4 alternatives)" and shows "Possible matches (opens in a new window)" with three small thumbnail images.
- Google Search Interface:** A screenshot of a Google search page. The search bar contains the text "Google Search". Below it, there's a "Web" section with "Results 1 - 10 of about 881,000". A "Personalization Engine" overlay is visible on the right side of the search results.

- As imagens de face pertencem a mesma pessoa?
- A face na imagem pertence a um indivíduo específico?
- O rosto pertence a alguma pessoa de um conjunto de indivíduos específicos?

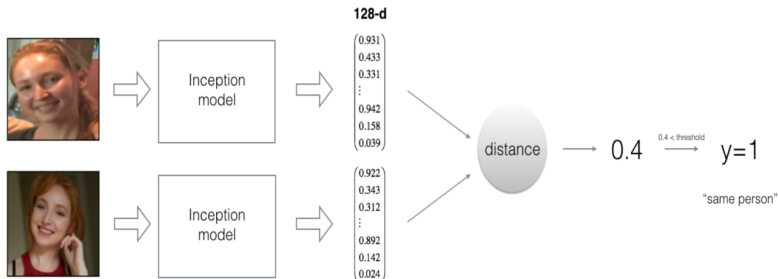
Eigenfaces [Turk e Pentland, 1991]

- Redução de dimensionalidade de imagens de face.
- Reconhecimento se dá por medição de distância no espaço de *eigenfaces*.



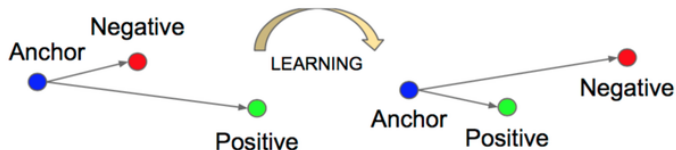
Deepface [Taigman et al. 2014]

- Duas redes convolucionais profundas com pesos compartilhados.
- Modelo aprende a distanciar faces de pessoas diferentes.



Facenet [Schroff et al. 2015]

- Também aprende *embedding* de face por rede neural profunda.
- Usa *triplet loss*



$$L = \sum_i^N \left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 + \alpha \quad (1)$$

Labeled faces in the wild [Huan et al. 2008]

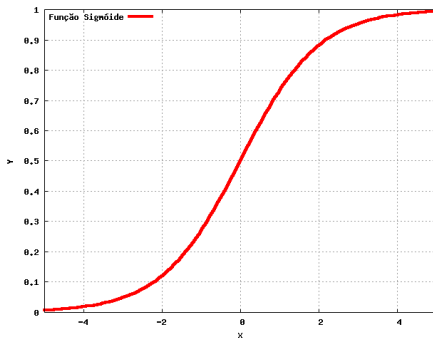
- 13233 imagens.
- 5749 indivíduos, 1680 com mais de uma imagem no conjunto.
- Rostos obtidos pelo Viola-Jones.



Extrator fixo

O modelo

- Vetores de imagem computados por rede Xception treinada no ImageNet.
- Dois vetores de imagens são dados como entrada a uma camada completamente conectada com *dropout* seguida de uma unidade sigmóide.
- Classifica se as imagens são da mesma pessoa ou não.



Hiperparâmetros:

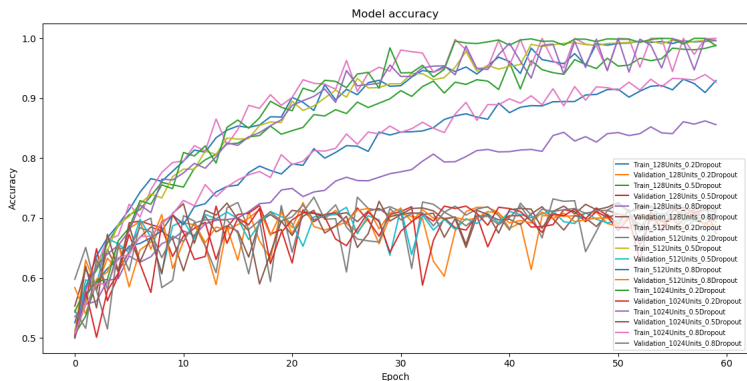
- Número de unidades:
 - 1 128
 - 2 512
 - 3 1024
- Probabilidade de dropout:
 - 1 0.2
 - 2 0.5
 - 3 0.8
- Modo de combinação de imagens:
 - 1 Adição
 - 2 Concatenação
 - 3 Subtração
 - 4 Produto interno
 - 5 Produto elemento a elemento

- 2200 pares de treinamento.
- 1000 pares de validação.
- Minimização da entropia cruzada binária:

$$\text{cross-entropy} = -(y \log(p) + (1 - y) \log(1 - p)). \quad (2)$$

Resultados

Acurácia no treino e validação para diferentes valores de número de unidades e chance de dropout:

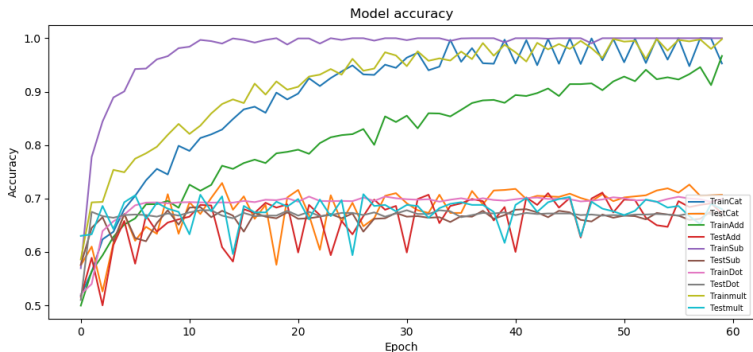


Resultados na primeira etapa de experimentos:

Modelo	Acurácia
Units_128_Drop_0.2	0.718
Units_128_Drop_0.5	0.721
Units_128_Drop_0.8	0.719
Units_512_Drop_0.2	0.735
Units_512_Drop_0.5	0.718
Units_512_Drop_0.8	0.726
Units_1024_Drop_0.2	0.72
Units_1024_Drop_0.5	0.728
Units_1024_Drop_0.8	0.722

Resultados

Acurácia no treino e validação para diferentes modos de combinação de entradas:



Resultados na segunda etapa de experimentos:

Modelo	Acurácia
concat	0.729
add	0.711
subtract	0.684
dotProduct	0.678
multiply	0.708

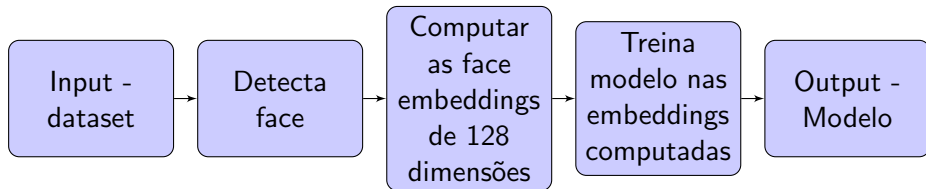
Pipeline de Detecção e Reconhecimento

Dataset da turma:

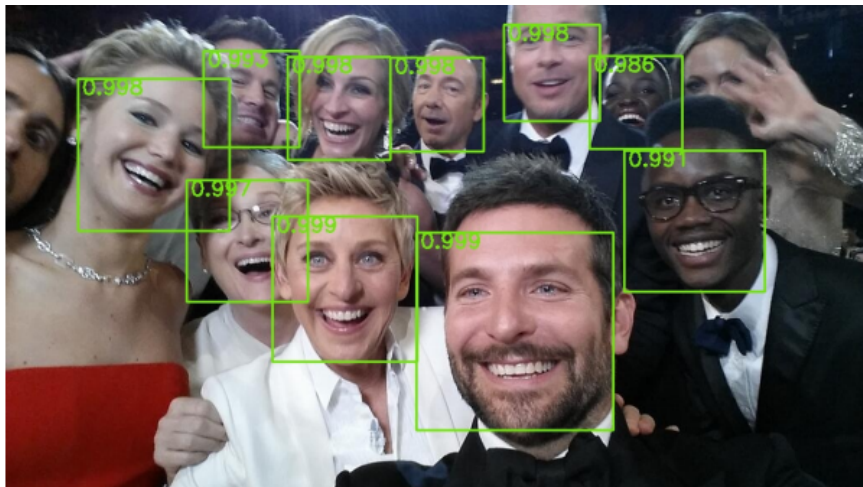
- 6 classes: Felipe, Livia, Natália, Pedro, Rafael, Raphael e Unknown.
- Total de 36 imagens, sendo que todas as classes possuem 6 imagens, com exceção do Rafael (4 imagens) e Natália (2 imagens).

O dataset LFW é o mesmo utilizado anteriormente, tirando o fato que é feito a detecção da face, antes do reconhecimento.

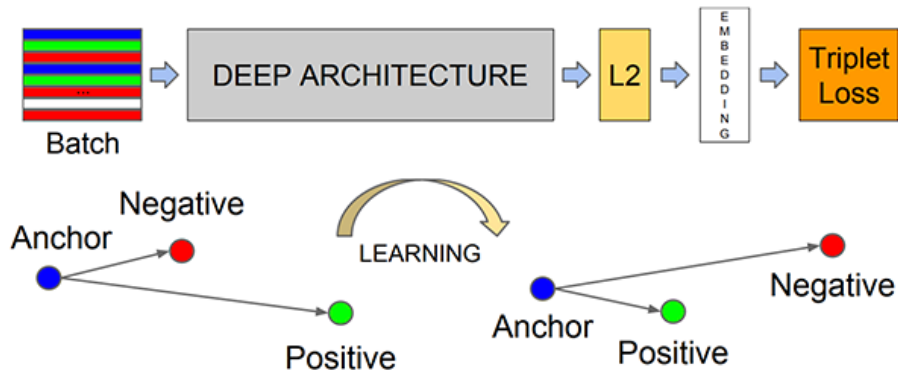
Pipeline



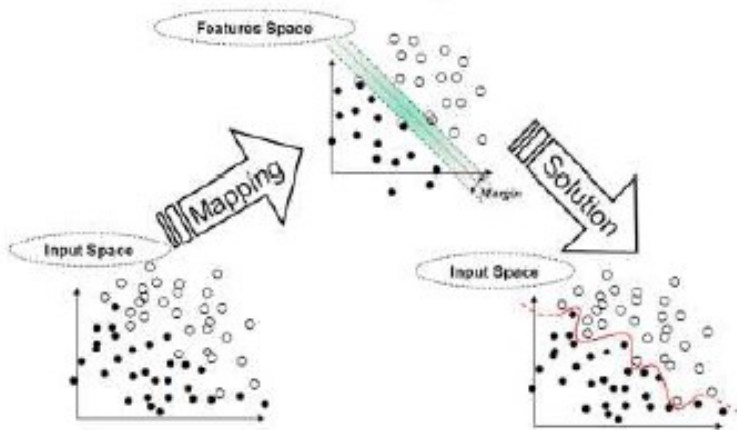
Detecção de Faces



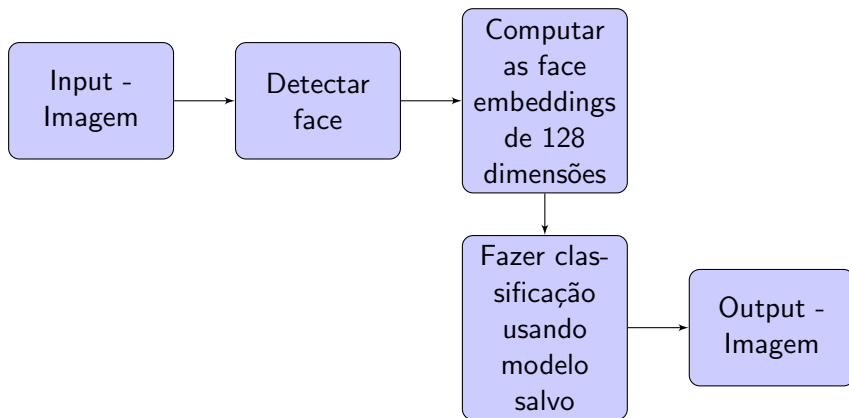
Cálculo das face embeddings



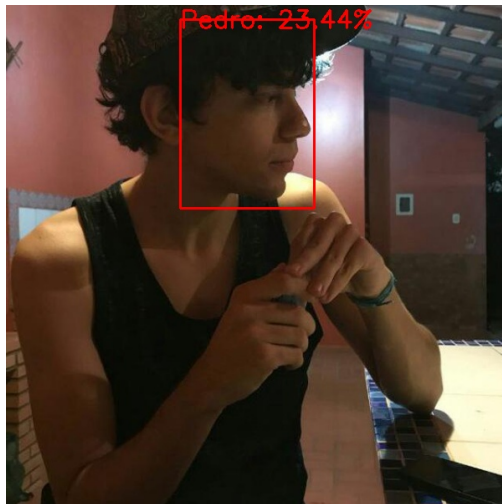
The SVM algorithm



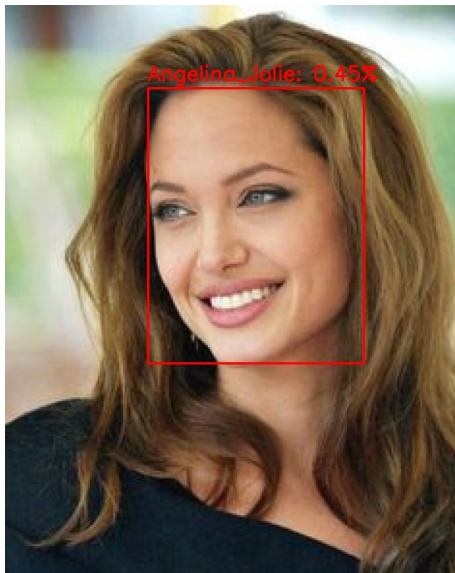
Previsão de novas imagens - Pipeline



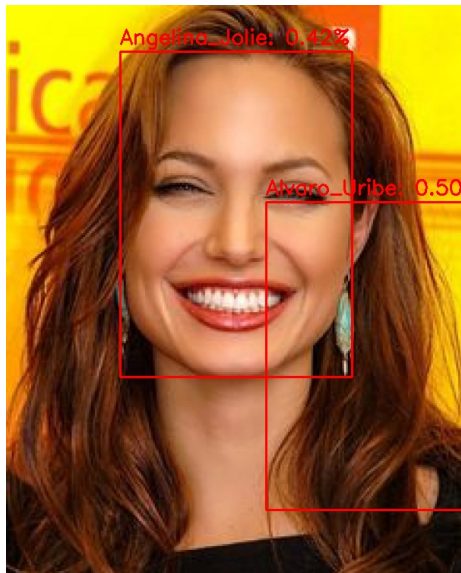
Resultados



Resultados



Resultados



Conclusão

- O modelo apresentado usando rede convolucional pré-treinada como extrator de características de imagens superou a baseline do conjunto de validação, alcançando 73.5% de acurácia, contra os 50% de baseline.
- Este modelo apresenta como vantagem grande rapidez de treinamento, simplicidade do modelo e desnecessidade de gerar características manualmente.
- O pipeline construído para detecção e reconhecimento de faces mostrou-se muito bom para a detecção de faces nas imagens testadas, embora não tenha apresentado bons resultados para nenhum dos dois datasets abordados.
- Como trabalho futuro é interessante avaliar o último modelo tratado no mesmo dataset usado na parte 1.

References I



Matthew A Turk e Alex P Pentland.

Face recognition using eigenfaces.

In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, pages 586–591.



Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, e Lior Wolf.

Deepface: Closing the gap to human-level performance in face verification.

In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014.



Ilorian Schroff, Dmitry Kalenichenko, e James Philbin.

Facenet: A unified embedding for face recognition and clustering.

In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015.



Gary B Huang, Marwan Mattar, Tamara Berg, e Eric Learned-Miller.

Labeled faces in the wild: A database for studying face recognition in unconstrained environments.

In: Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition, 2008.