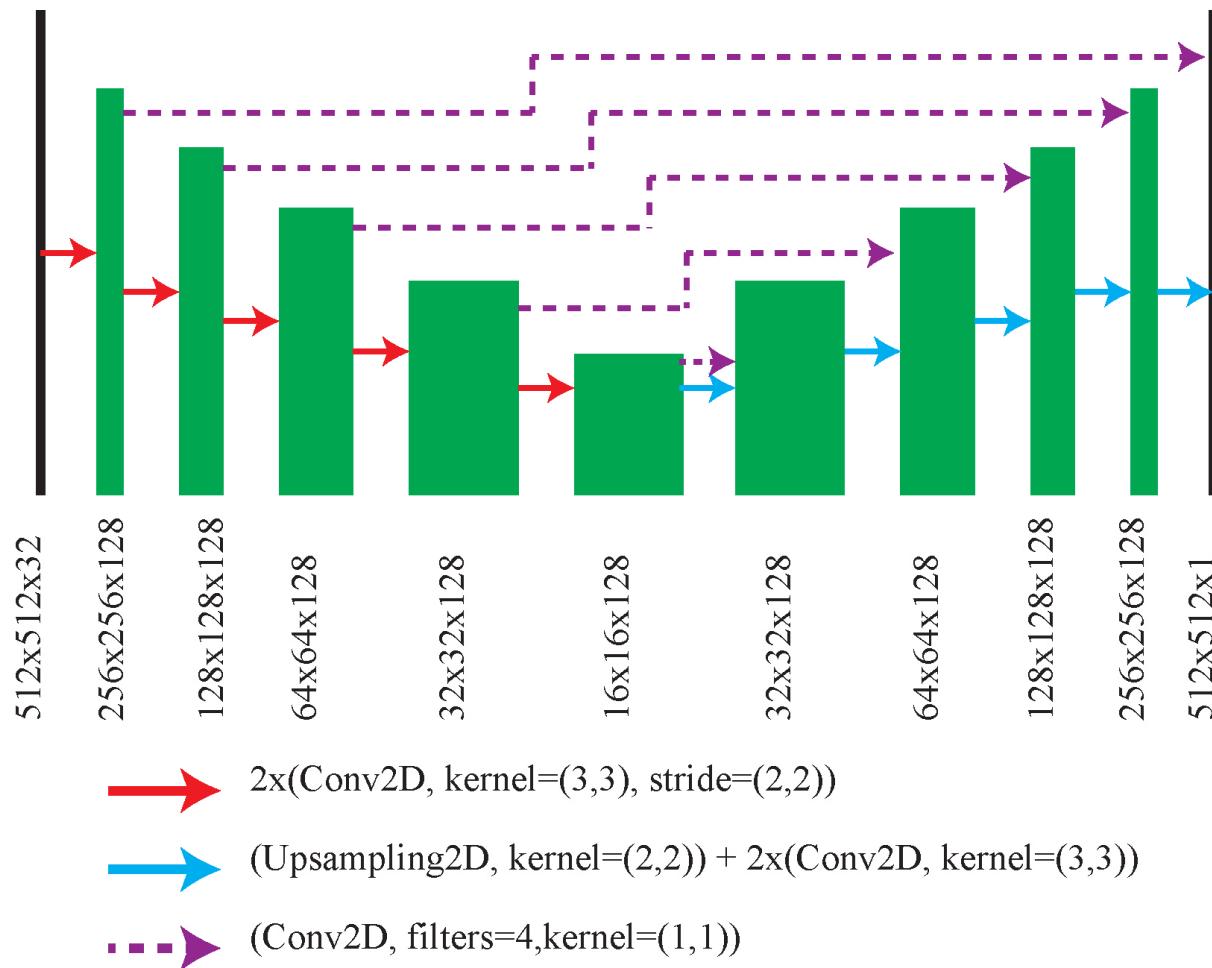


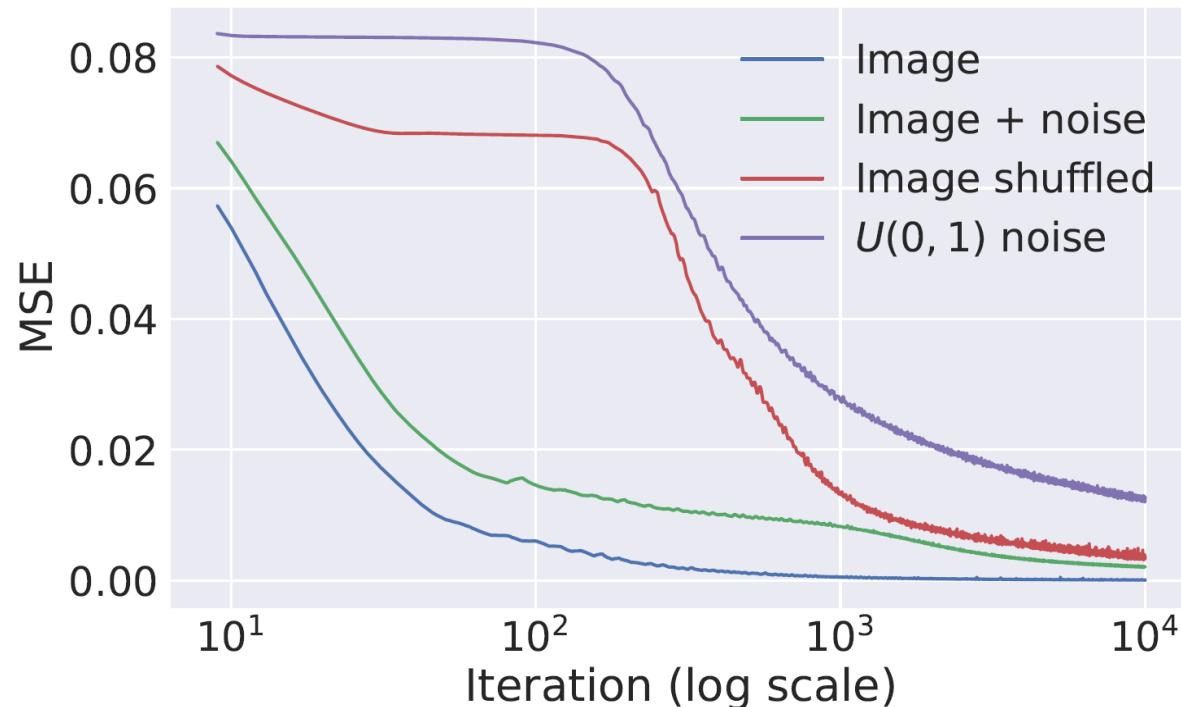
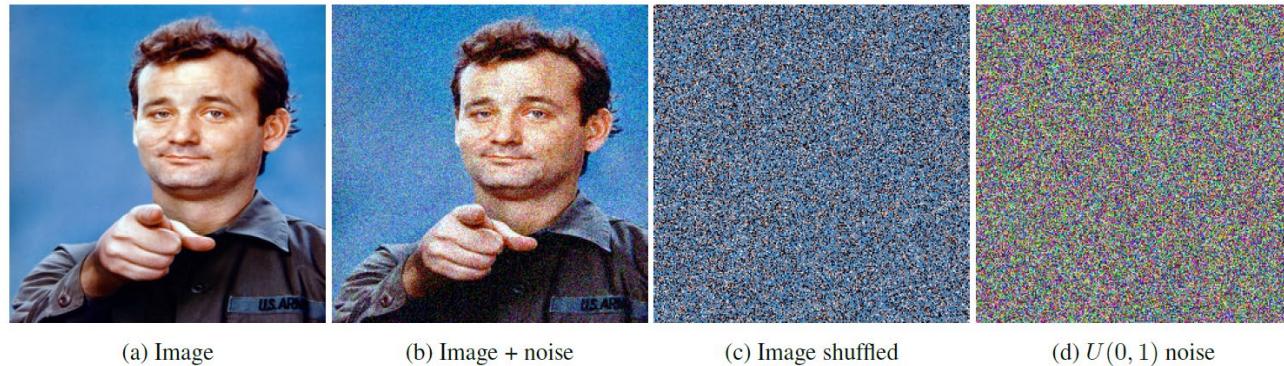
Deep Image Prior and Discriminator Networks

Lecture 8b

Deep Image Prior



Deep Image Prior



Deep Image Prior: Denoising

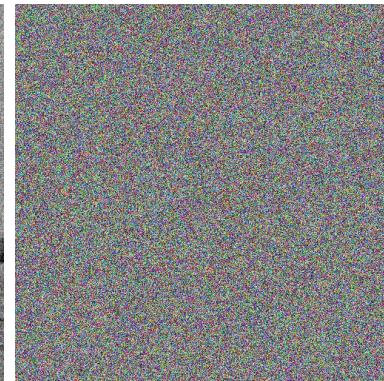
Clean image



Noisy image



Network input



Network structure:

- 5 down-sampling
 - Kernel size=3
 - Number of filters=128
- 5 up-sampling
 - Kernel size=3
 - Number of filters=128
- 5 skip connection
 - Kernel size=1
 - Number of filters=4
- Leaky ReLU activation function

$$\text{Cost function: } C = \left\| \text{network}_{\text{output}} - \text{image}_{\text{noisy}} \right\|^2$$

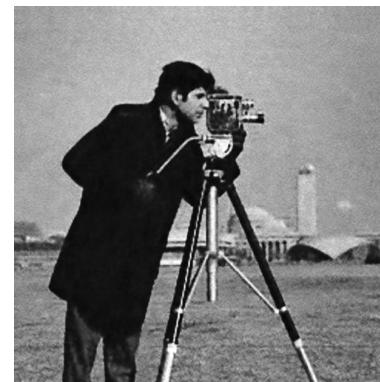
Epoch 100



Epoch 300



Epoch 3000



Epoch 3900



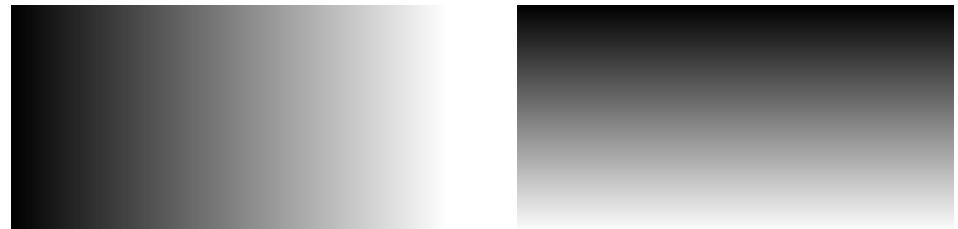
The network reaches to the denoised image at epoch=3000.

Deep Image Prior: Inpainting

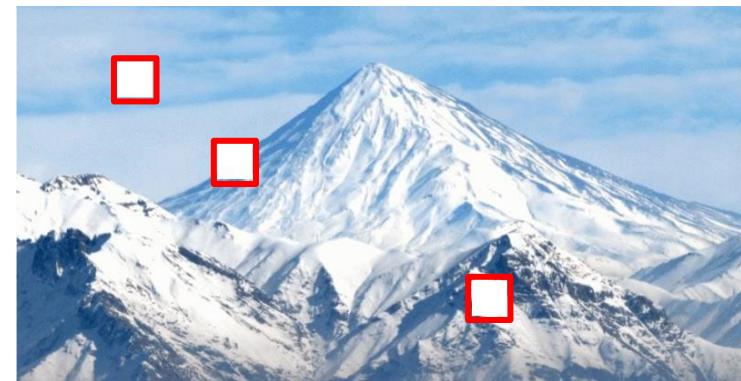
Network structure:

- 5 down-sampling
 - Kernel size=3x3
 - Number of filters in each layer=128
- 5 up-sampling
 - Kernel size=3x3
 - Number of filters in each layer=128
- No skip connection
- Leaky ReLU activation function

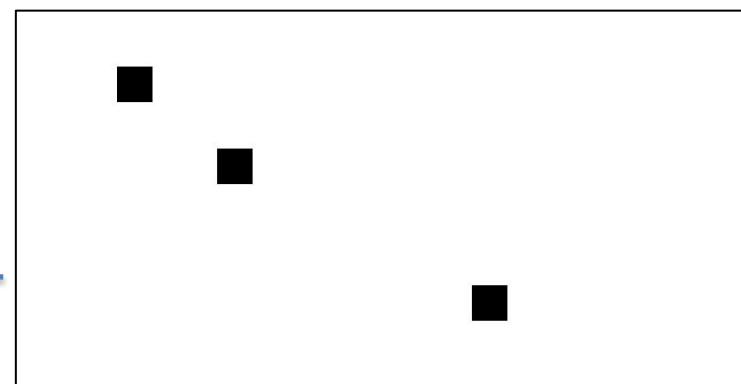
Network input



Masked image



Mask



Cost function:

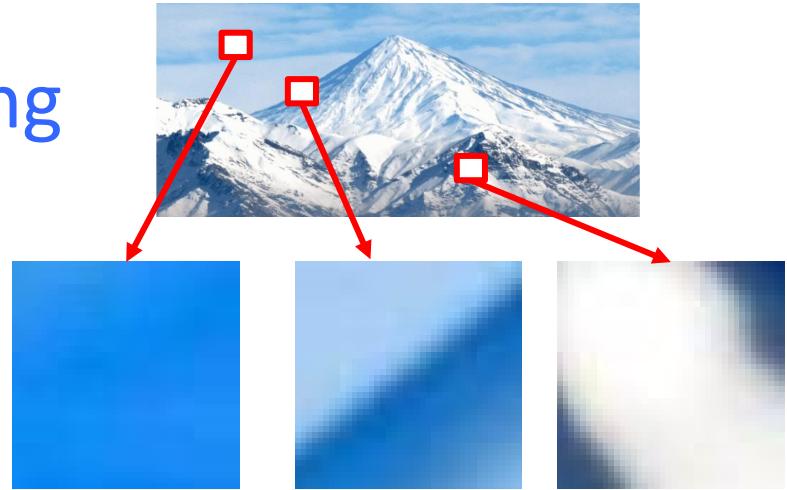
$$C = \left\| (network_{output} - image_{masked}) \square mask \right\|^2$$

INNER
PRODUCT

Diagram illustrating the cost function calculation. The equation shows the squared difference between the network output and the masked image, where the difference is weighted by the mask. A red arrow labeled "INNER PRODUCT" points from the term $(network_{output} - image_{masked})$ to the multiplication symbol \square .

Deep Image Prior: Inpainting

Epoch
1000



Epoch
10000



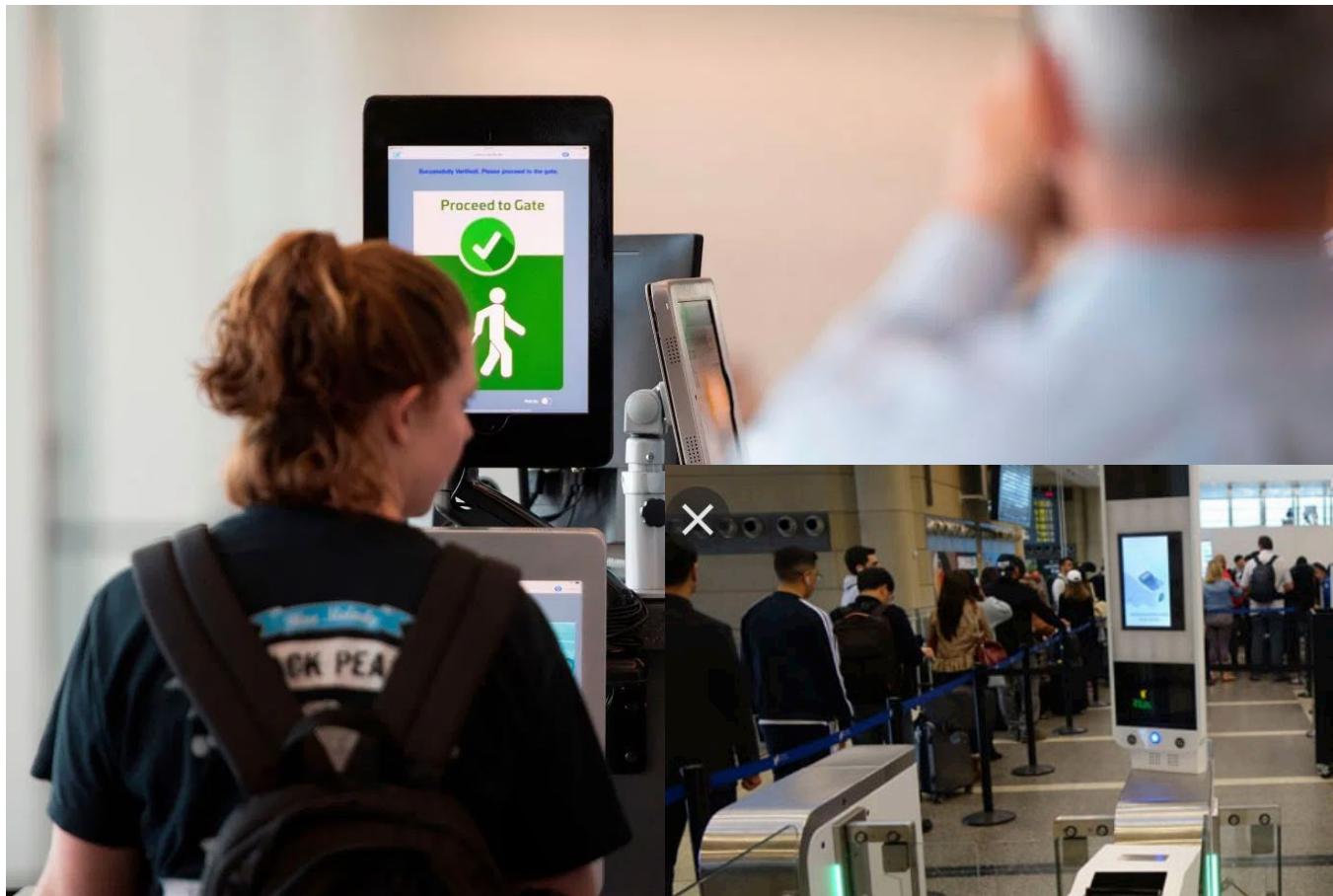
Original
image



Face recognition

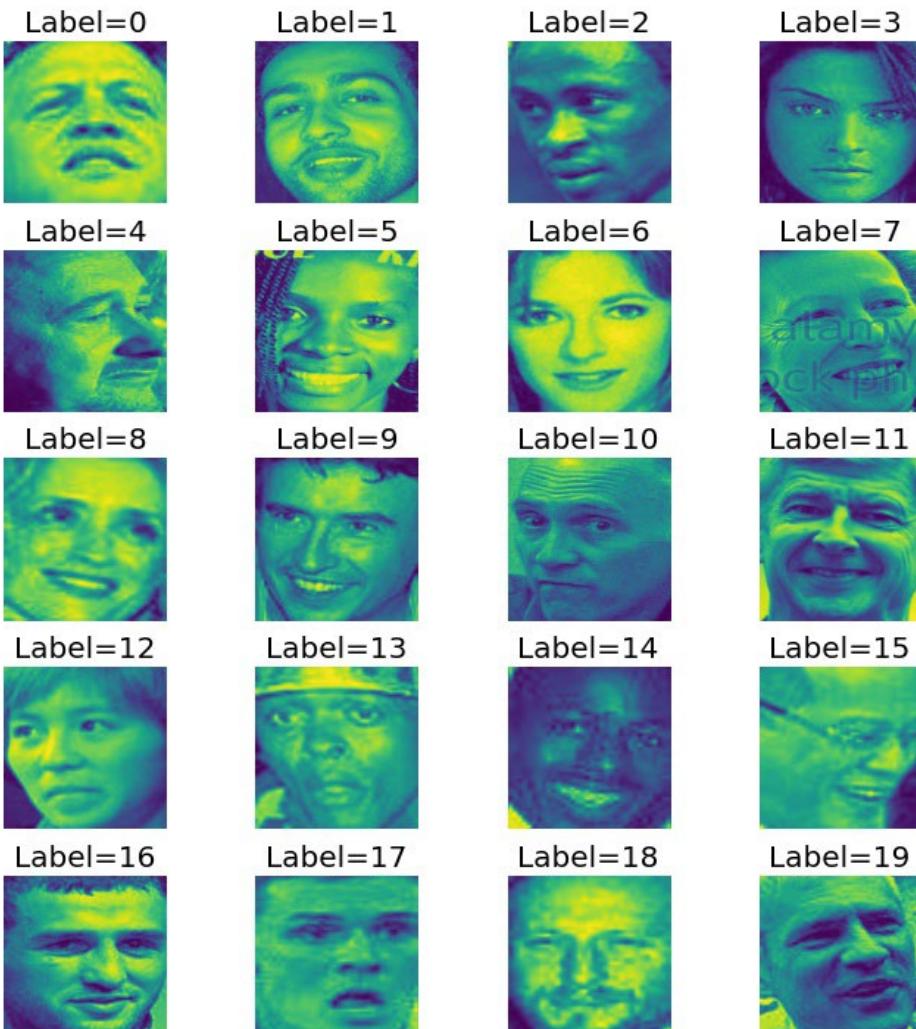


LAX



20 people classification

Part of VGGFace2 dataset



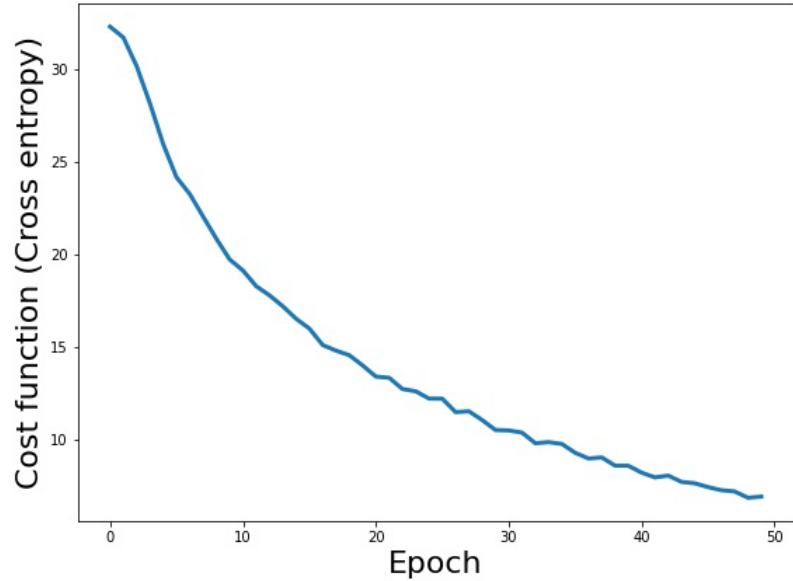
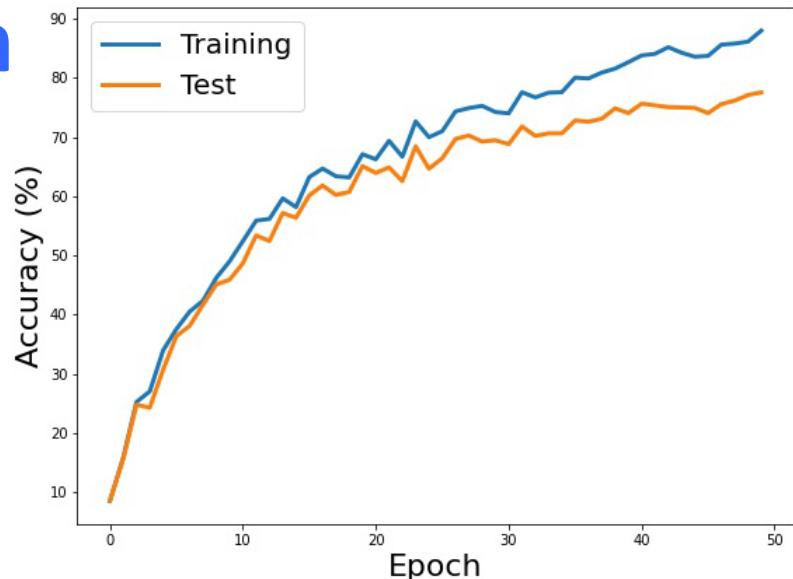
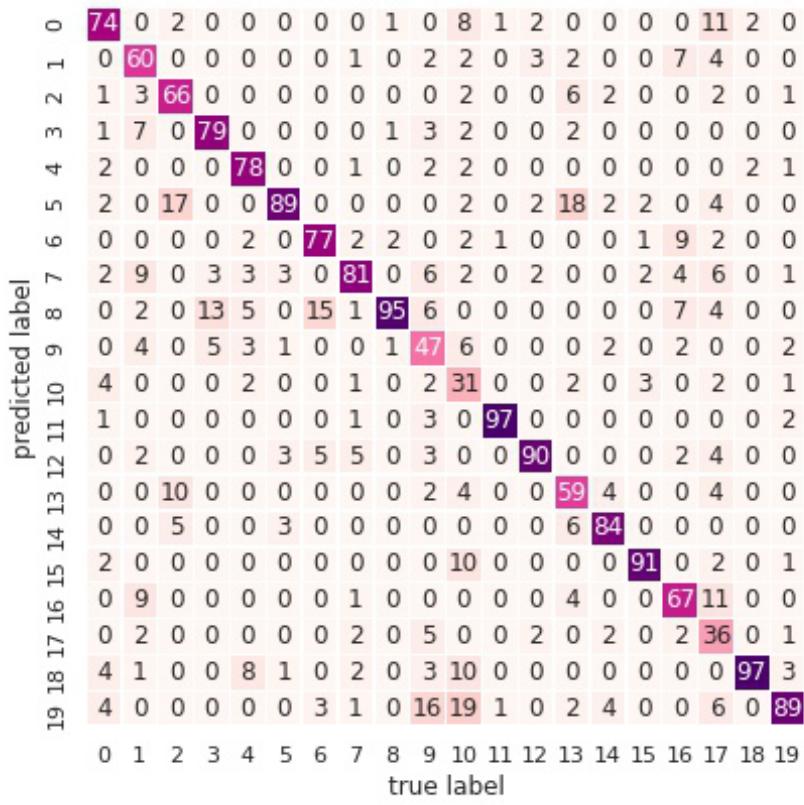
- Face classification of 20 people
- 5535 training samples
- 200 to 400 samples/person
- 1375 test samples
- 20 classes
- Adam optimizer
- Learning rate=0.001
- Cross entropy cost function

Layer	Size out	Activation	Kernel
Input	128x128x1		
Convolution	128x128x32	LeakyReLU	3x3
Maxpool	32x32x32		4x4
Convolution	32x32x64	LeakyReLU	3x3
Maxpool	8x8x64		4x4
Convolution	8x8x128	LeakyReLU	3x3
Maxpool	2x2x128		4x4
Fully connected	512→128	LeakyReLU	
Fully connected	128→20	Softmax	

20 people classification

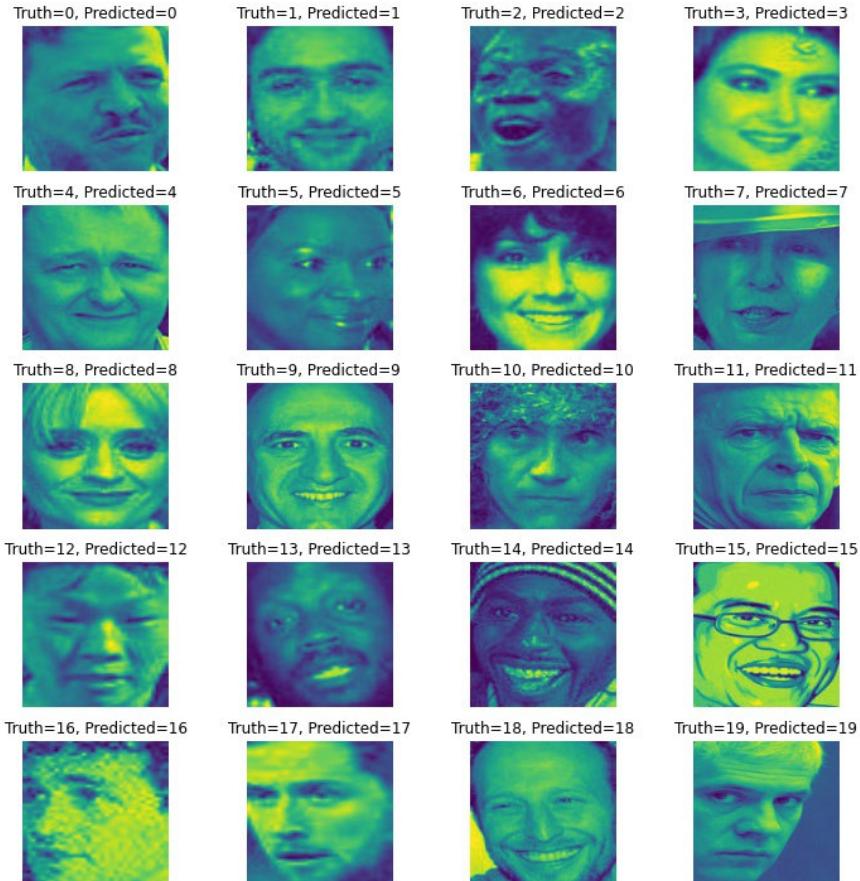
Training accuracy: 87.97%

Test accuracy: 77.53%

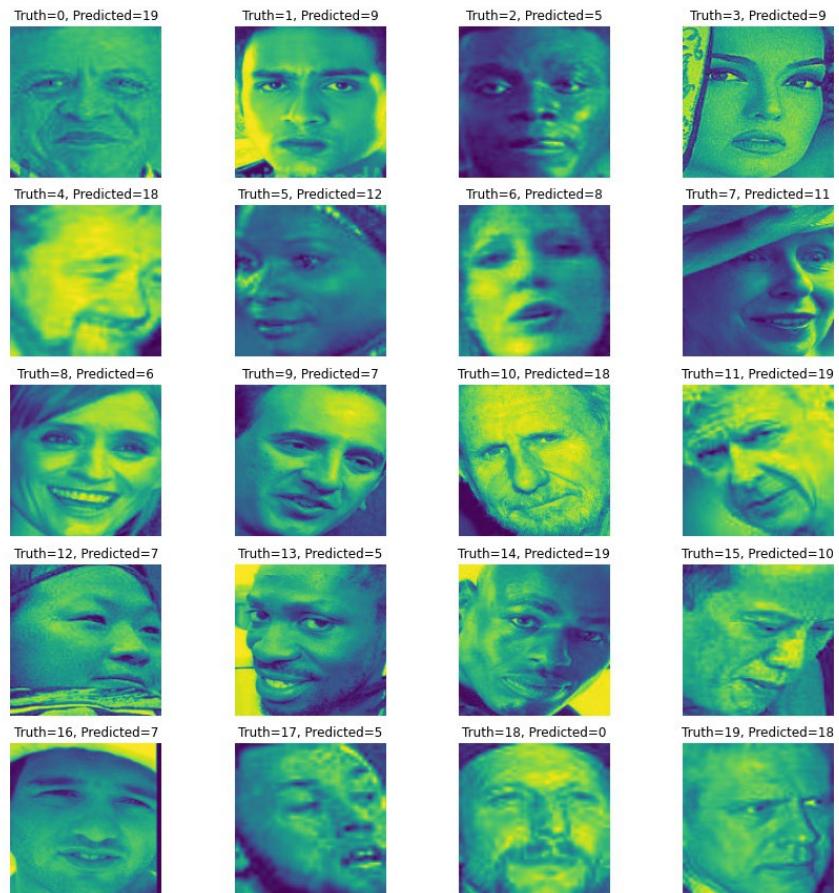


20 people classification

Examples of correct prediction

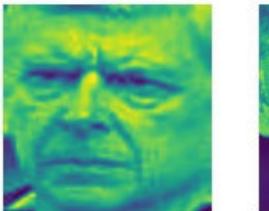


Examples of wrong prediction

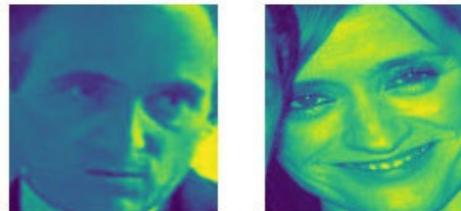


Recognizing one person from others

Label=1



Label=0

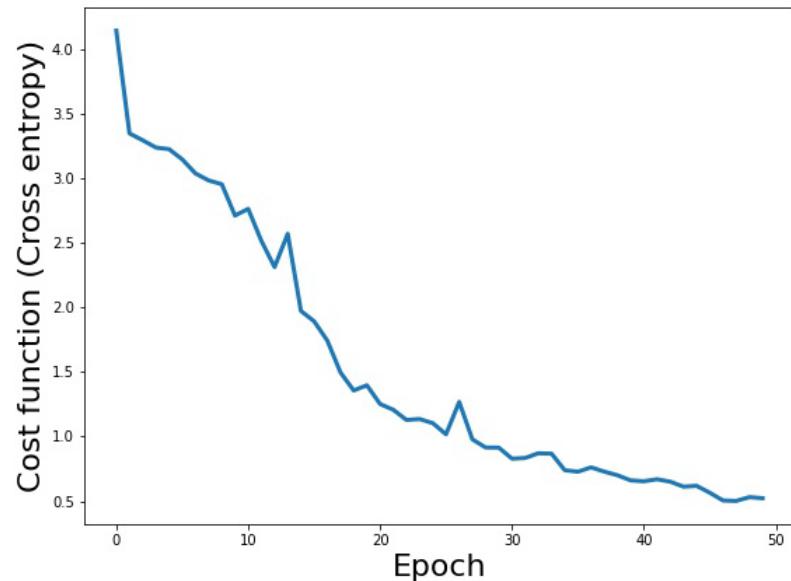
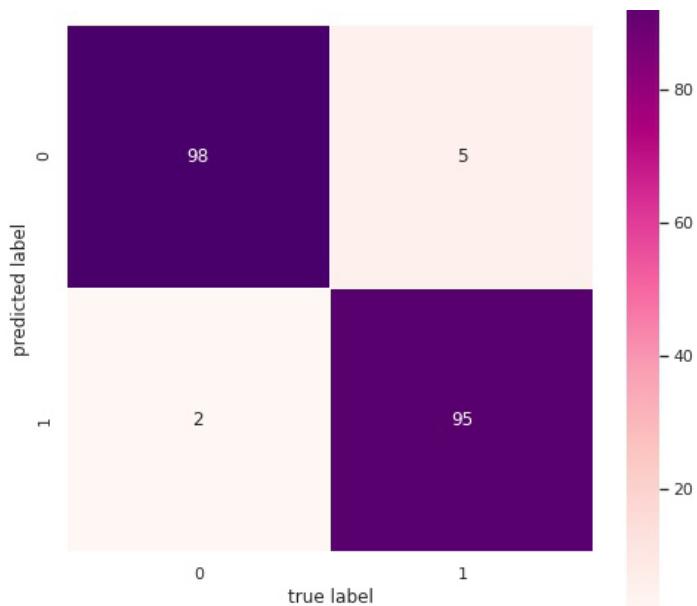
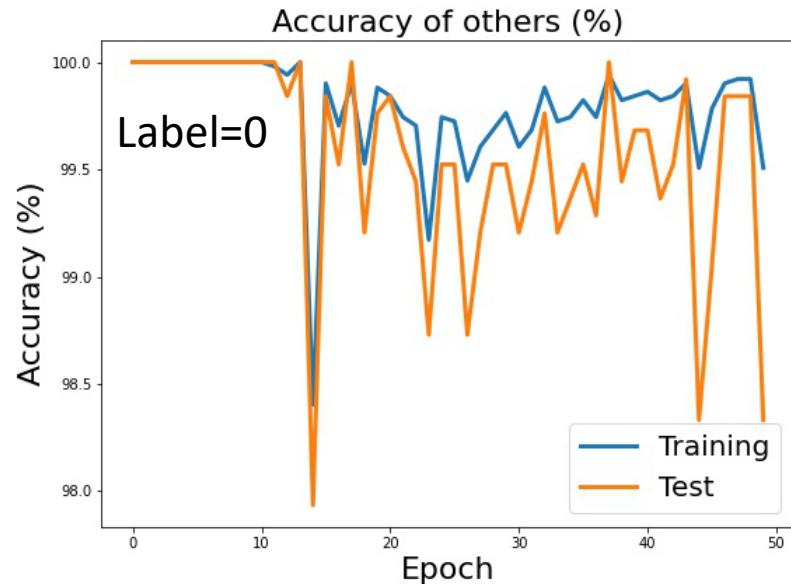
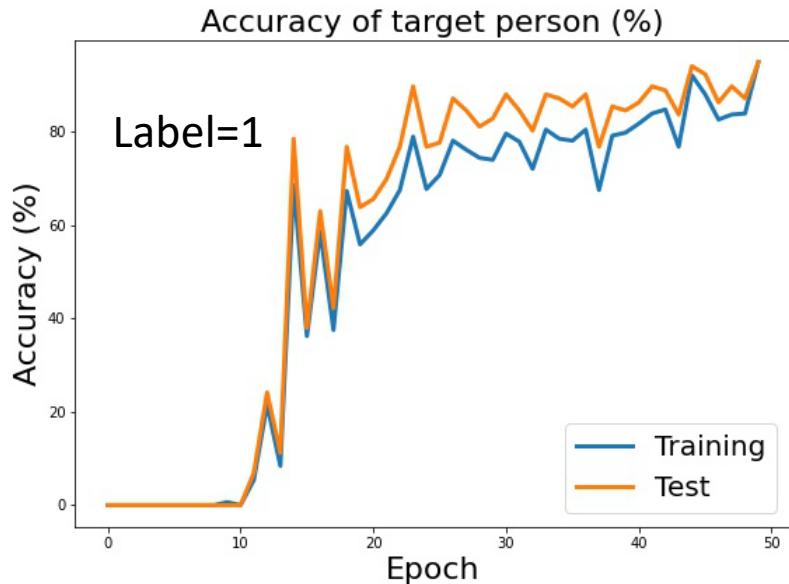


- This person should be recognized among 19 other people.
- The label for this person is 1.
- The label for others is 0.

- Classification with 2 classes
- 5535 training samples
 - Label 0: 5071 samples
 - Label 1: 464 samples
- 1375 test samples
 - Label 0: 1259 samples
 - Label 1: 116 samples
- Adam optimizer
- Learning rate=0.001
- Cross entropy cost function

Layer	Size out	Activation	Kernel
Input	128x128x1		
Convolution	128x128x32	LeakyReLU	3x3
Maxpool	32x32x32		4x4
Convolution	32x32x64	LeakyReLU	3x3
Maxpool	8x8x64		4x4
Convolution	8x8x128	LeakyReLU	3x3
Maxpool	2x2x128		4x4
Fully connected	512→128	LeakyReLU	
Fully connected	128→2	Softmax	

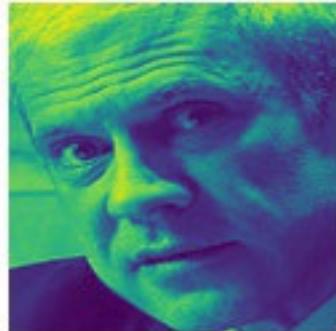
Recognizing one person from others



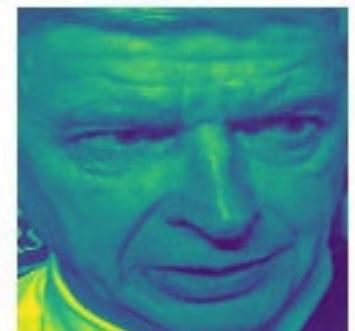
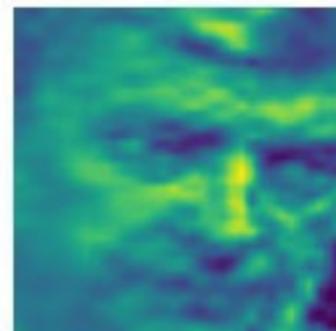
Recognizing one person from others

- Examples of wrong classifications of the test set

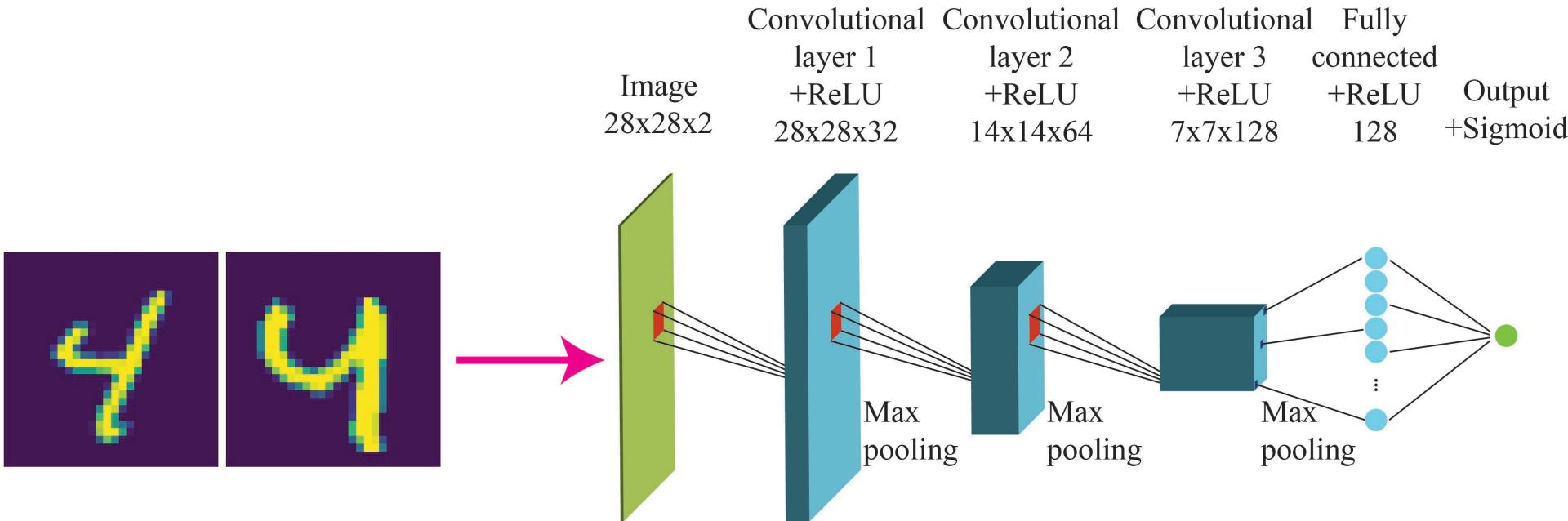
0 predicted as 1



1 predicted as 0

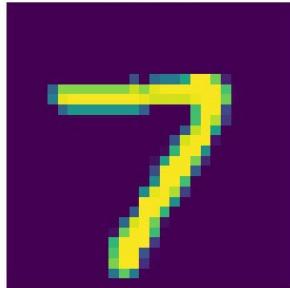


Discriminators

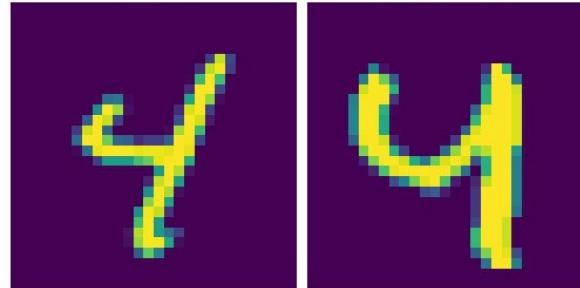


- Pairs match: Output=1
- Pairs do not match: Output=0

Discriminator: Digits

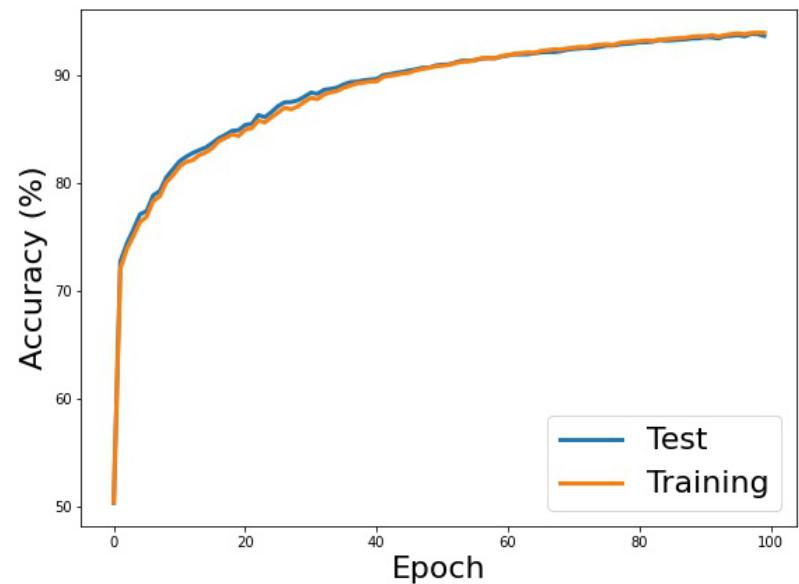


Pairs do not match



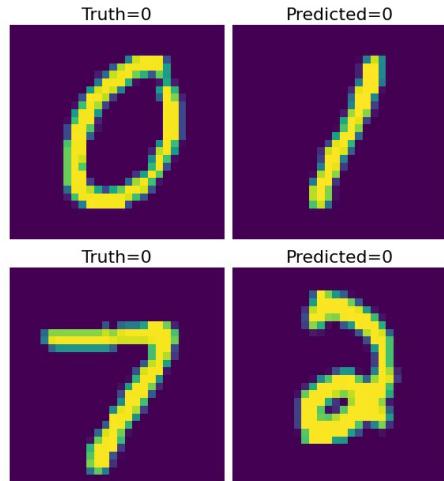
Pairs match

- ADAM optimizer, learning rate=0.00002
- Binary cross entropy
- 100 batch size – 100 epochs
- 120000 training samples
- 20000 test samples
- Activation function: relu and sigmoid (last layer)
- Training accuracy: 94.05%
- Test accuracy: 93.98 %

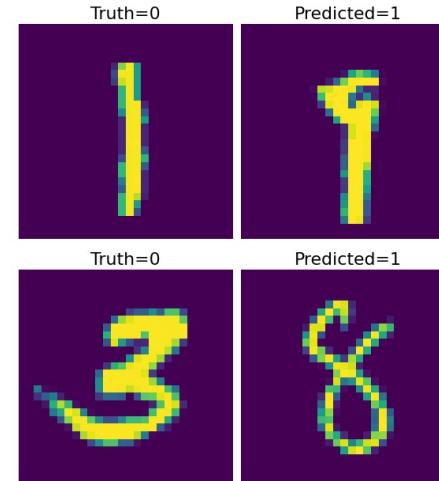


Discriminator: Digits

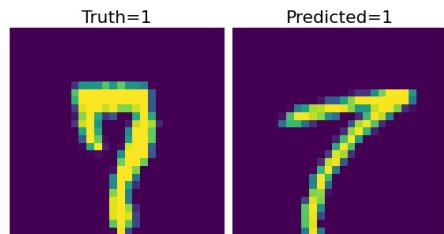
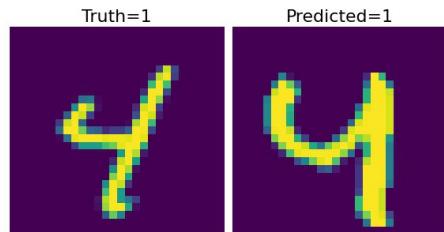
Correctly predicted different



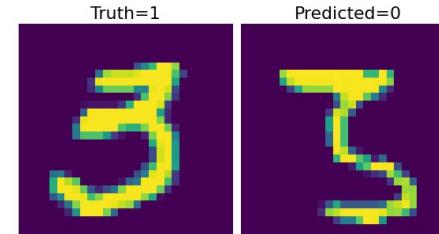
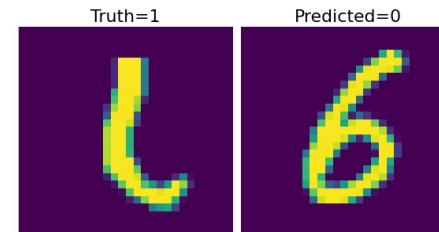
Wrongly predicted same



Correctly predicted same



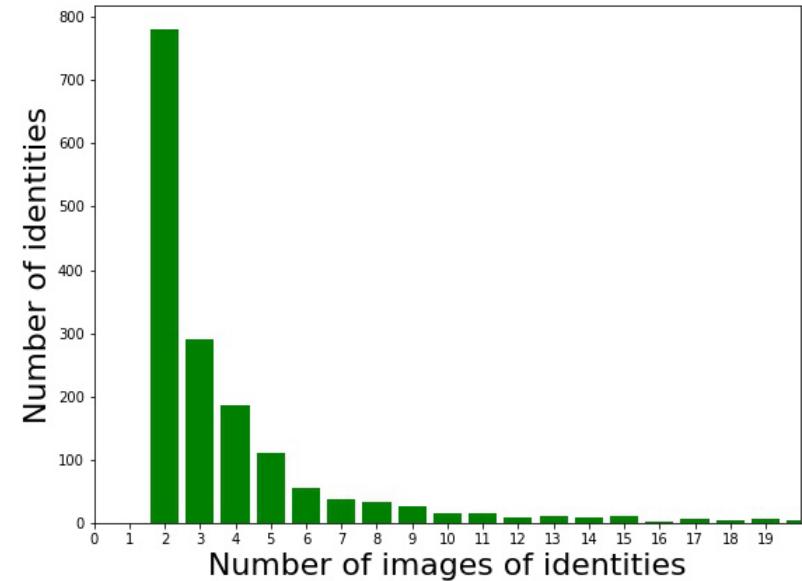
Wrongly predicted different



Discriminator: Faces

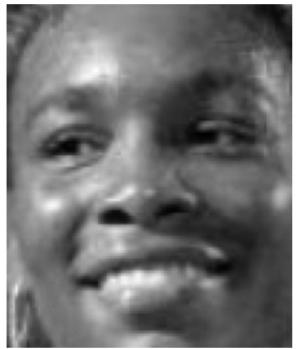
Labeled Faces

Examples of faces



- Total number of identities: 1680
 - Total number of images: 9164
- There is not enough images for each person to train a classifier network**

Discriminator: Faces

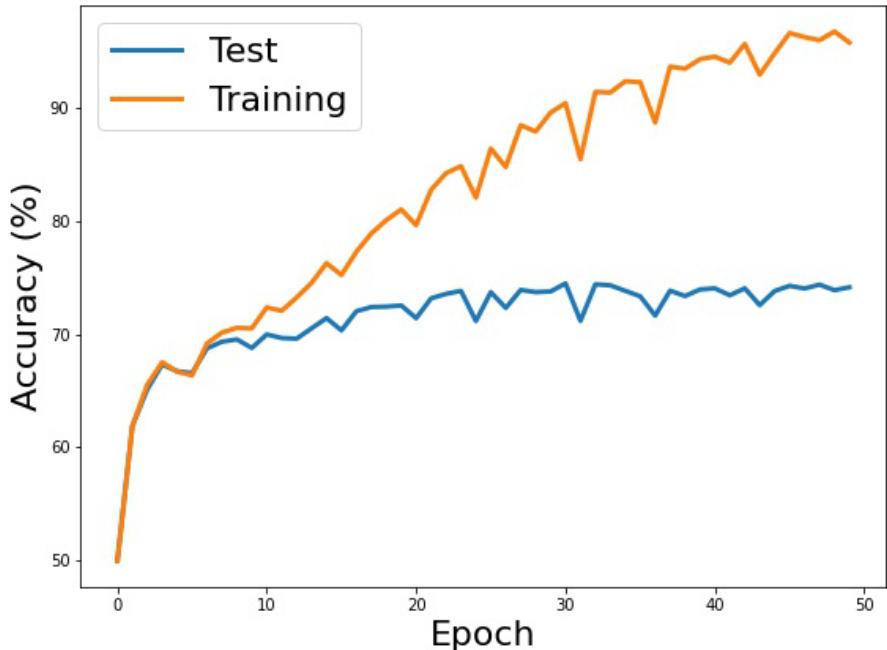


Pairs that do not match



Pairs that match

- ADAM optimizer, learning rate=0.0002
- Binary cross entropy
- Batch size 2– 50 epochs
- 16402 training samples
- 14044 test samples
- Activation function: leaky relu and sigmoid (last layer)
- Training accuracy: 96.43%
- Test accuracy: 74.03 %

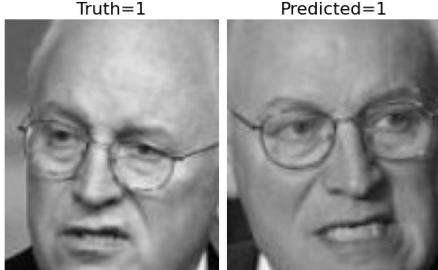


Discriminator: Faces

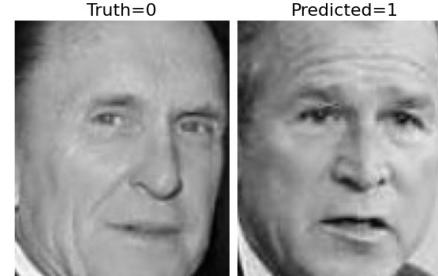
Correctly predicted different



Correctly predicted same



Wrongly predicted same



Wrongly predicted different



Clearview

Mail Online

Hundreds of law enforcement agencies are using a new facial recognition app that can reveal names and addresses of unknown individuals by comparing them to THREE BILLION images scraped from the internet

- More than 600 police departments are said to be using the Clearview AI app
- The app reportedly uses facial recognition AI to match uploaded photos to the three billion images it has in its database

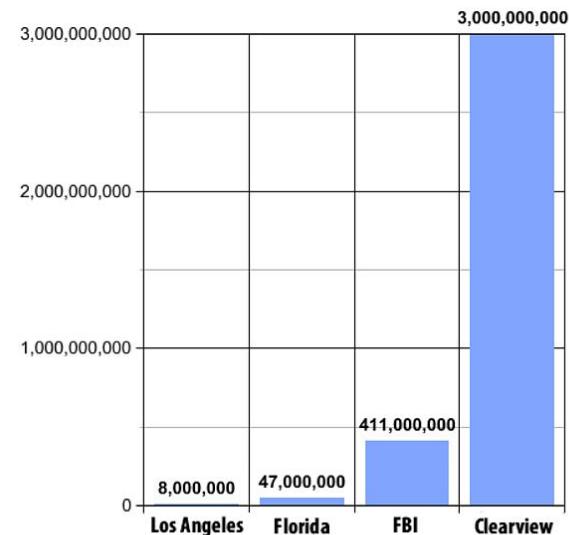
WIRED



Scraping the Web Is a Powerful Tool. Clearview AI Abused It

The facial recognition startup claims it collected billions of photos from sites like Facebook and Twitter. What does the practice mean for the open web?

HOW MANY PHOTOS CAN BE SEARCHED WITH...



Source: Muckrock

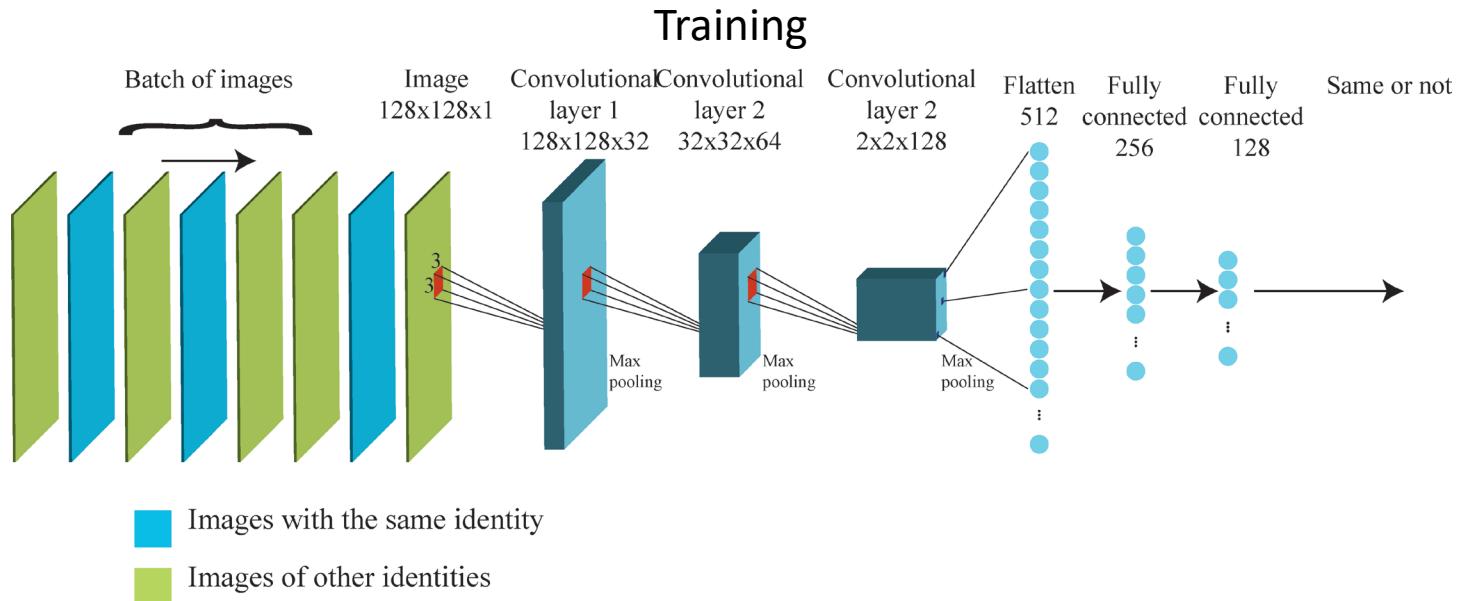
Clearview.ai



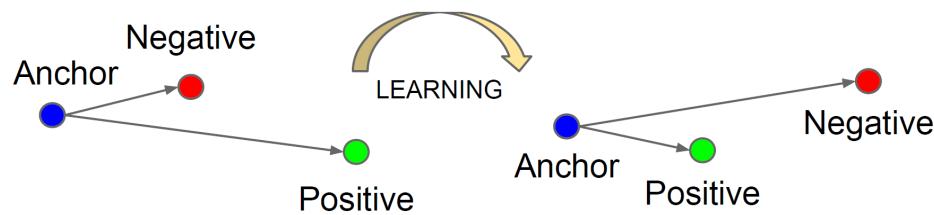
Computer vision for a safer world

[Request Access](#)

Facenet



A minimal number of exemplars of any identity is required in each batch.



Triplet loss:

$$C = \sum_i \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]$$

x_i^a image of a specific person

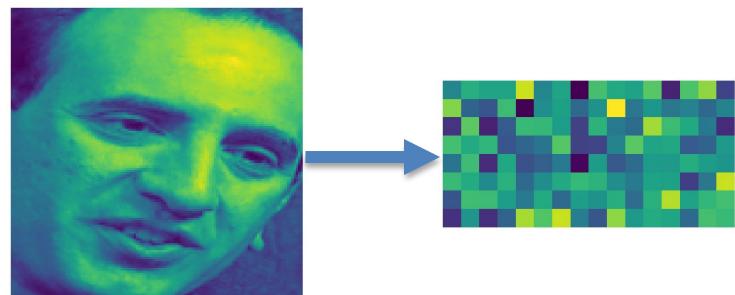
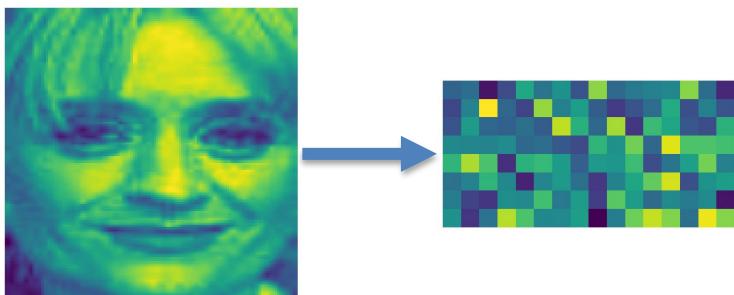
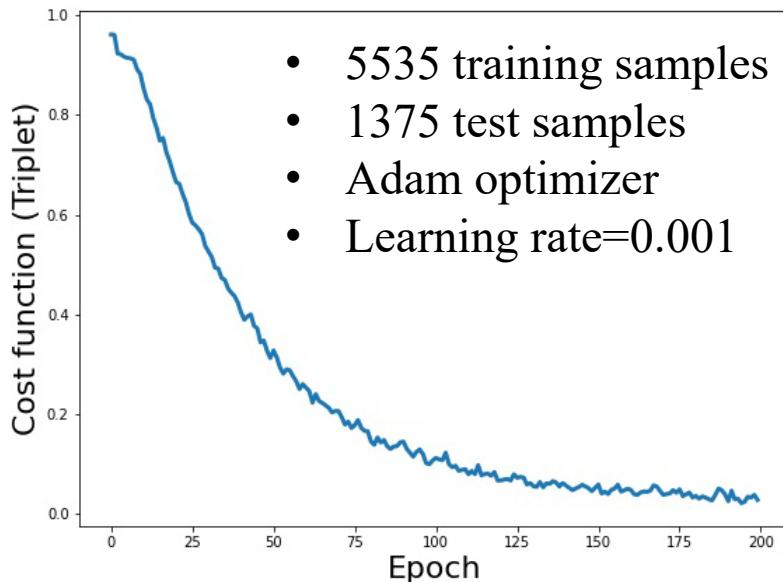
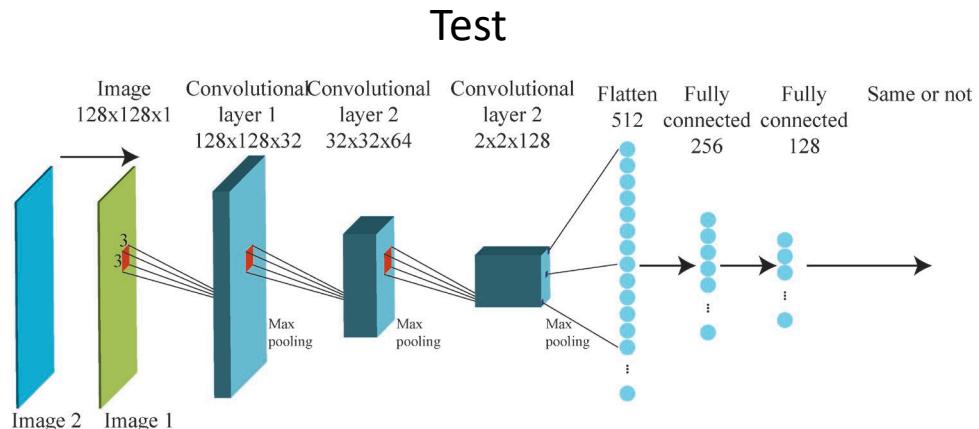
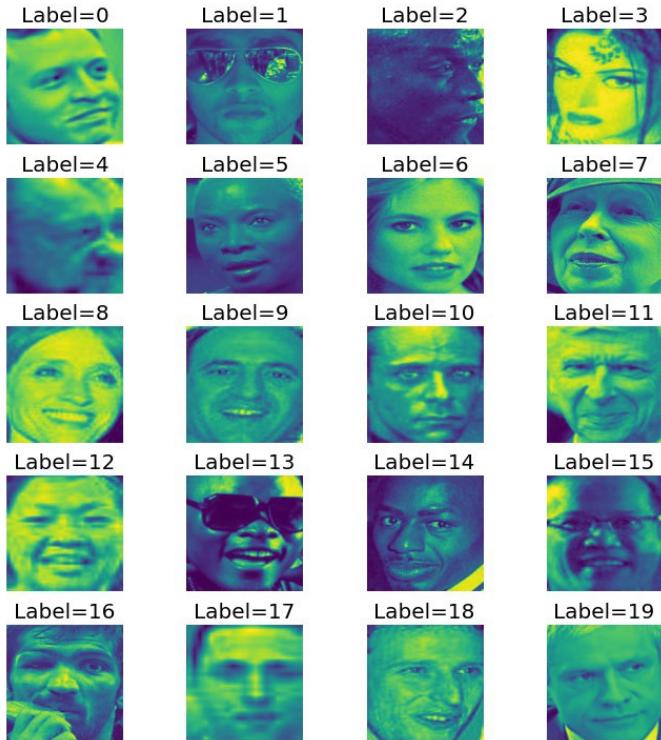
x_i^p other images of the specific person

x_i^n images of any other person

α margin between positive and negative pairs

Facenet: 20 people

Part of VGGFace2 dataset



Facenet: 20 people

- The network maps each image to a vector with the size of 128.
- Two images belong to the same identity if the distance between the corresponding vectors is less than the threshold.

True accept



x_1 and x_2 are two face images

$$\begin{cases} \text{Same identity} & \text{if } \|f(x_1) - f(x_2)\|_2^2 < \alpha \\ \text{Different identities} & \text{if } \|f(x_1) - f(x_2)\|_2^2 > \alpha \end{cases}$$

False reject

