

Real-time Face Detection and Classification from YouTube videos using Matlab, Python & OpenCV

Donald Shenaj

ICT for Internet and Multimedia
Digital Forensics A.Y. 19-20

July 6, 2020

Table of Contents

1 Face Detection

2 Face Classification

3 Final System

4 Conclusions

Face Detection - Viola and Jones algorithm

Object detection method using Haar features, consists of 3 concepts:

- Integral images
- Adaboost
- Cascade classifiers

Main characteristics:

- Fast, Real-time
- Already trained by OpenCV
- Returns bounding boxes

[1] P. Viola, M. Jones, *Rapid Object Detection using a Boosted Cascade of Simple Features*, International Conference on Computer Vision, 2001



Figure: Example of face detected

Face Classification - CNN design

Designed in Matlab by extending the code of lab 3:

- Added convolutional layers
- Added dropout
- Npeople = 13

Will be imported in Python after the training:

- ONNX format
- OpenCV

```
layer_vet=[
    imageInputLayer([64 64 3])

    convolution2dLayer([3 3],64)
    batchNormalizationLayer
    reluLayer();
    maxPooling2dLayer(2,'Stride',2)

    convolution2dLayer([5 5],128);
    batchNormalizationLayer
    reluLayer();
    maxPooling2dLayer(2,'Stride',2)

    convolution2dLayer([8 8],128);
    batchNormalizationLayer
    reluLayer();
    maxPooling2dLayer(2,'Stride',2)

    convolution2dLayer(9, 128, 'Padding','same');
    batchNormalizationLayer
    reluLayer();
    maxPooling2dLayer(2,'Stride',2)
    dropoutLayer(0.25)

    fullyConnectedLayer(Npeople)

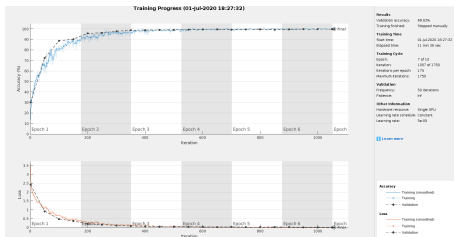
    softmaxLayer();

    classificationLayer()
];
```

Face Classification - CNN Training

Training parameters:

- Epochs = 6
- Batch size = 128
- Accuracy = 98%



Confusion Matrix																
Output Class	Adam Sandler	430 5.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	0 0.0%	0 0.0%	1 0.1%	0 0.0%	1 0.1%	0 0.0%	98.3% 0.7%	
	Alyssa Milano	0 0.0%	430 5.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	98.3% 0.7%	
	Brace Willis	0 0.0%	0 0.0%	388 4.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	7 0.8%	3 0.4%	9 1.1%	0 0.0%	0 0.0%	95.1% 4.9%	
	Danise Richards	0 0.0%	0 0.0%	0 0.0%	533 6.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	7 0.8%	0 0.0%	0 0.0%	0 0.0%	97.2% 2.8%	
	George Clooney	1 0.0%	0 0.0%	0 0.0%	0 0.0%	767 9.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.9% 0.1%
	Gwyneth Paltrow	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	737 8.9%	3 0.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	96.8% 3.2%
	Hugh Jackman	7 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	524 6.3%	4 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	97.9% 2.1%
	Jason Statham	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	234 2.8%	0 0.0%	0 0.0%	1 0.1%	7 0.8%	0 0.0%	0 0.0%	98.3% 1.7%
	Jennifer Love Hewitt	0 0.0%	1 0.1%	1 0.1%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	0 0.0%	434 5.2%	0 0.0%	4 0.5%	0 0.0%	0 0.0%	0 0.0%	98.8% 1.2%
	Lindsay Lohan	0 0.0%	0 0.0%	13 1.6%	0 0.0%	3 0.4%	0 0.0%	3 0.4%	0 0.0%	2 0.2%	1988 24.1%	2 0.2%	0 0.0%	0 0.0%	0 0.0%	98.9% 1.1%
	Mark Ruffalo	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	248 3.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	Robert Downey Jr	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	171 2.1%	0 0.0%	0 0.0%	100% 0.0%
	Will Smith	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	380 4.6%	95.5% 4.5%
	Target Class		97.2% 2.3%	98.8% 3.8%	99.2% 3.2%	99.5% 2.5%	100% 0.0%	98.5% 1.5%	99.1% 0.9%	95.5% 4.5%	99.8% 0.2%	91.9% 8.1%	99.8% 0.2%	100% 0.0%	98.7% 1.3%	
Adam Sandler																
Alyssa Milano																
Brace Willis																
Danise Richards																
George Clooney																
Gwyneth Paltrow																
Hugh Jackman																
Jason Statham																
Jennifer Love Hewitt																
Lindsay Lohan																
Mark Ruffalo																
Robert Downey Jr																
Will Smith																

Figure: Confusion matrix of the CNN

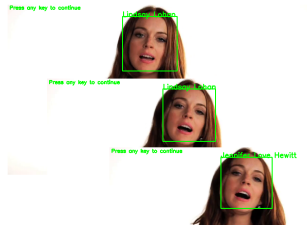
Video Processing

Implemented in python using:

- OpenCV: face detection and classification
- VidGear: YouTube video frames from url

Main features:

- N consecutive frames classification
- Removed false positive faces
- No video tracking



$$N = 3$$



Predicted label: Lindsay Lohan

Detection and Classification

- Limits:
 - Not too accurate
 - Multiface detector
- Improvements:
 - Efficient tracking algorithm
 - Faster face detection