Face Detection using Deep Learning

Akshay Khadse Ashish Sukhwani Raghav Gupta Soumya Dutta

Indian Institute of Technology
Bombay

April 28, 2017

Outline

Outline

Face Detection as problem

- Pictures: Convenient medium of expression
- Retrieving and Organisation of large collection is a challenge
- Well established techniques for geotagging and timestamping.
- More contextual queries: An issue drawing increased attention
- Eg: Photos taken with a particular person.

Outline

Relevance and Motivation

- Faces: Not unique rigid object
- Inter-personal variations due to identity, genetics etc
- Intra-personal variations due to expressions, aging, facial hair etc
- Active research since two decades have resulted in well established techniques
- These methods fail to detect faces with different angle, partial faces.

Outline

Outline

Haar Cascade

- Haar feature based Cascade Classifers
- Haar features are extracted from images
- Construct feature vector of each image
- Adaboost: Selecting from large Number of features
- Final Classifier: Weighted sum of individual classifier at each step
- Works very well for images with frontal faces
- Not satisfactory for images with other views, multiple faces

Outline

Preprocessing

- Main paper referred [2] describesfinetuning of AlexNet for multiview face detection
- AFLW dataset: 21k images with 24k face annotations
- Randomly sampling sub windows of the images
- IOU with ground truth to increase number of examples
- IOU $>$ 50%: Positive example, IOU $<$ 30%: Negative example
- 200k positive examples, 20 million negative examples
- Image size $227 \times 227 \times 3$.

Outline

Training

- AlexNet: 5 Convolutional Layers, Fully connected Layers
- Batch Size 128: 32 Positive examples, 96 Negative Examples
- Trained for 50K iterations
- Final detector can be using Region based or sliding window approach
- This method uses sliding window approach due to its less complexity
- Paper describes comparison of this method with R-CNN
- AlexNet trained by reshaping output of fully connected layer for training SVM classier
- A bounding box regression unit was trained
- Both methods also trained on PASCAL VOC 2013 objects

Outline

Results

- Face trained AlexNet performs better than PASCAL VOC trained AlexNet
- Significant improvement of R-CNN due to bounding box regression
- Performance of RCNN not as good as previous method
- Reasons for inferior performance
  - Selective search leading to missing out some of face regions.
  - Imperfection in bounding box regression.

Outline

Data Collection and Pre-processing

- Same AFLW dataset used
- Dataset augmented in same way as described in [2]
- Total 24k positive and 120k negative
- Images reduced to 128×128
- All images converted to grayscale

Outline

Network Architecture

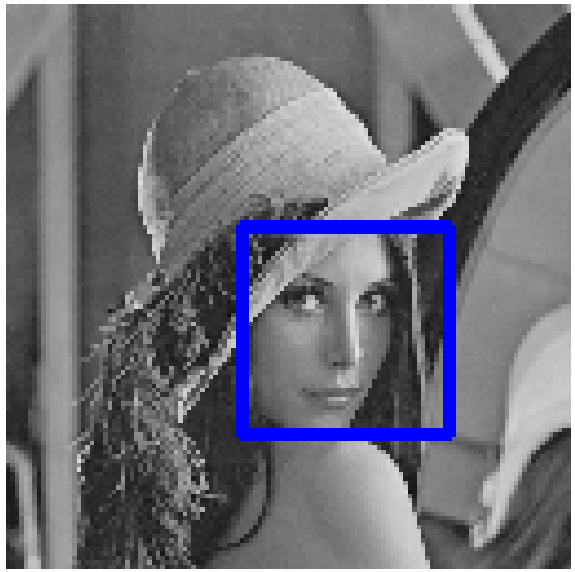
- CONV LAYER 1(Activation - ReLU Filter - $18 \times 18 \times 1 \times 60$ Bias - 60 Stride - 1×1)
- MAXPOOL LAYER1(Filter - 2×2 Stride - 2×2)
- CONV LAYER 2(Activation - ReLU Filter - $12 \times 12 \times 60 \times 30$ Bias - 30 Stride - 1×1)
- MAXPOOL LAYER2(Filter - 2×2 Stride - 2×2)
- CONV LAYER 3(Activation - ReLU Filter - $6 \times 6 \times 30 \times 15$ Bias - 15 Stride - 1×1)
- MAXPOOL LAYER3(Filter - 2×2 Stride - 2×2)
- FC LAYER 1(Activation - ReLU Weights - 3840×4096 Bias - 4096)
- FC LAYER 2(Activation - ReLU Weights - 4096×256 Bias - 256)
- FC LAYER 3(Activation - Sigmoid Weights - 256×1 Bias - 4096)

Outline

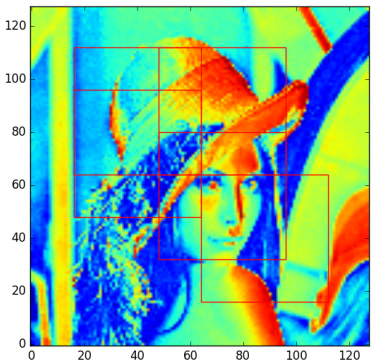
Training

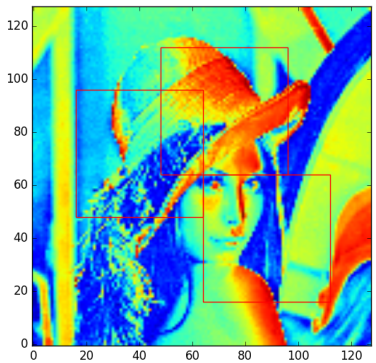
- Batch size 120, 50K iterations
- Originally Cross entropy with regularisation was used as cost
- Cost showed decreasing trend, accuracy finally setlled on 75%.
- Weights at end of training were very small
- Modified cost function: Multiplying error on positive by a constant $\geqslant$ 3.5
- Training accuracy of $\geqslant$ 85% was achieved
- Changing regularization parameter did not affect training accuracy
- 3 models with accuracies 88.3333%, 91.6667%, 94.2857% were saved for testing.

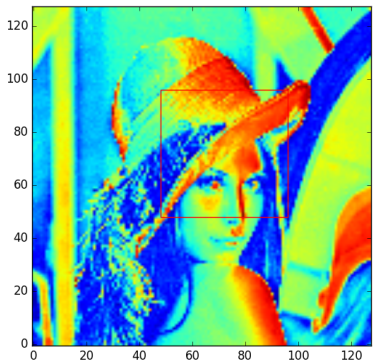
Results



(a) Haar Cascade Classifer for Image with single face



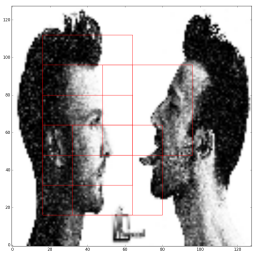


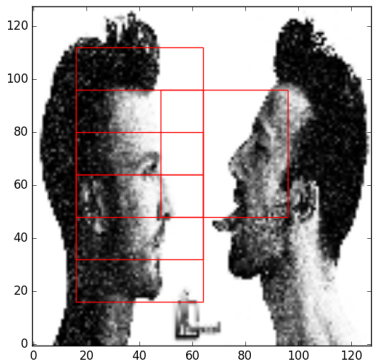


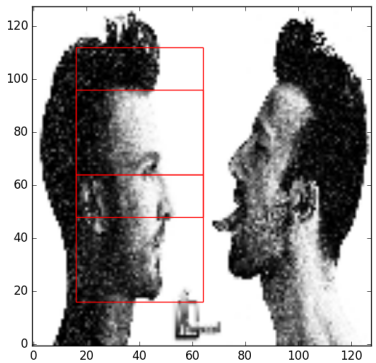
(b) Our face detector performance-model 1 to 3 (left to right)

Results

(a) Haar Cascade Classifer for Image with two faces



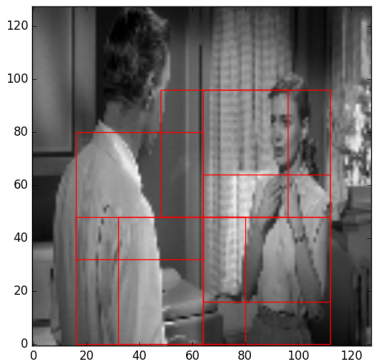


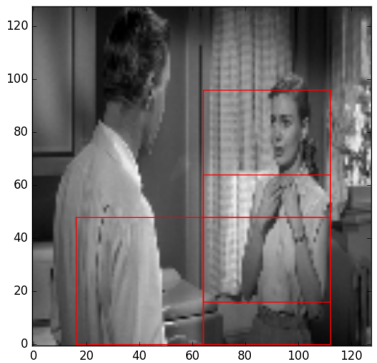


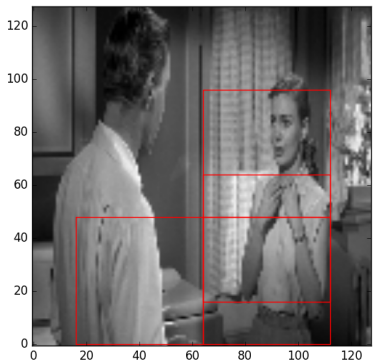
(b) Our face detector performance-model 1 to 3 (left to right)

Results

(a) Haar Cascade Classifer for Image with two faces

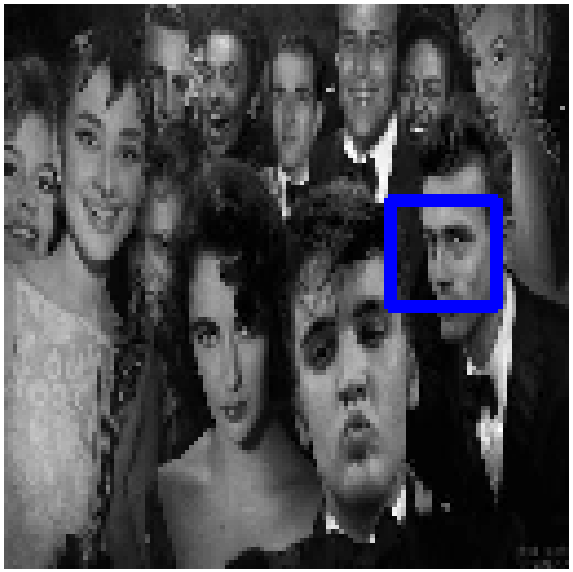




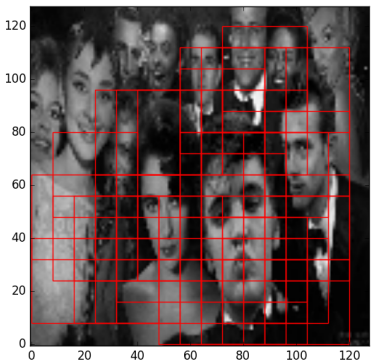


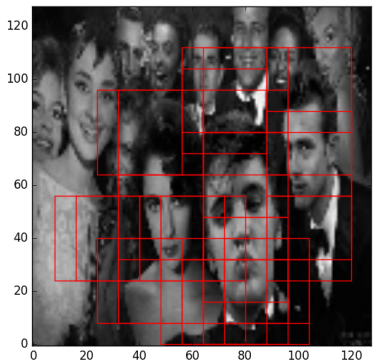
(b) Our face detector performance-model 1 to 3 (left to right)

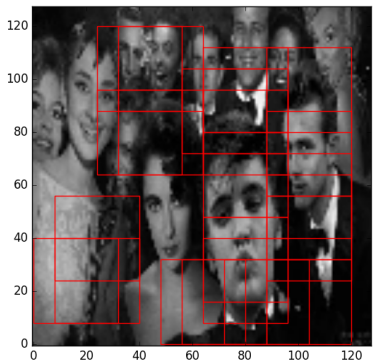
Results



(a) Haar Cascade Classifer for Image with two faces







(b) Our face detector performance-model 1 to 3 (left to right)

Conclusions

- A Deep Convolutional Neural Network was designed and using Tensorflow to detect faces from images
- A comparative study of this face detector was carried out against Haar cascade.
- Results significantly superior to the the Haar Cascade Classifier were achieved.
- Network is much simple than one described in paper[2]
- Many false positives are an issue.

Timeline

- Upto midterm-Collection and Pre-processing of data-initial architecture of the CNN
- Experimented with different network architectures
- Changing the cost function to increase training accuracy
- Comparison of Haar Cascade Classifier and our face detector
- Tested the models on some test images
- Implemented PCA for face detection at a preliminary level.

References I

- Jones, Michael, and Paul Viola. "Fast multi-view face detection." Mitsubishi Electric Research Lab TR-20003-96 3 (2003): 14.
- Farfade, Sachin Sudhakar, Mohammad J. Saberian, and Li-Jia Li. "Multi-view face detection using deep convolutional neural networks." Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. ACM, 2015.
- Zhang, Zhanpeng, et al. "Facial landmark detection by deep multi-task learning." European Conference on Computer Vision. Springer International Publishing, 2014.