Incremental Training for Face Recognition

Martin Winter and Werner Bailer

25th International Conference on MultiMedia Modeling,
Thessaloniki, Greece on January 10th, 2019

Talk - outline

- Problem statement
- Incremental training approach
- Application to Auto-Training for unknown persons
- Experiments / results

Problem statement

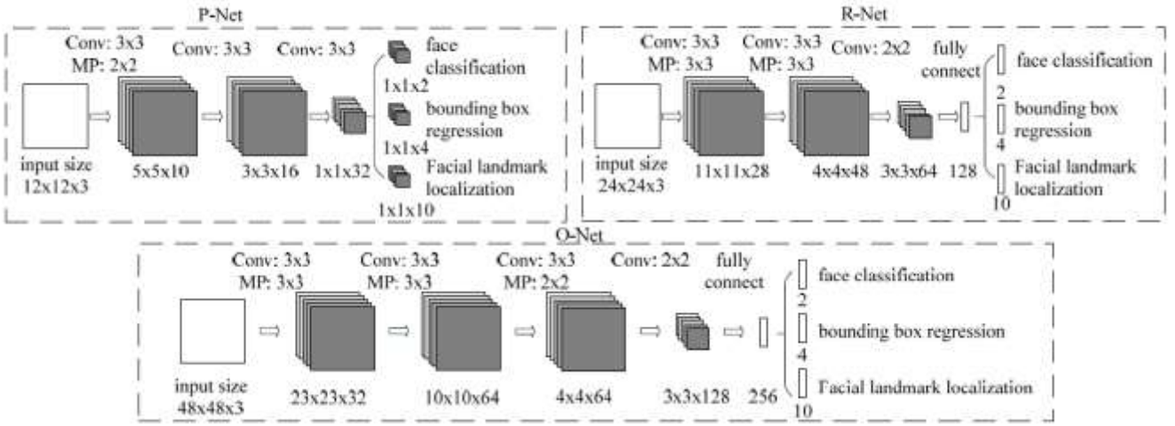
- **Identifying persons is a crucial task in video analysis**
  - broadcast content/archive documentation
  - media monitoring, open source intelligence
  - processing of user generated content
- **Annotating training data is costly**
  - large databases cover internationally known celebrities etc.
  - for many applications less known people are of interest, for which it is not easy to find samples on the web
  - resources for collecting and annotating samples are limited, identification must work with few examples

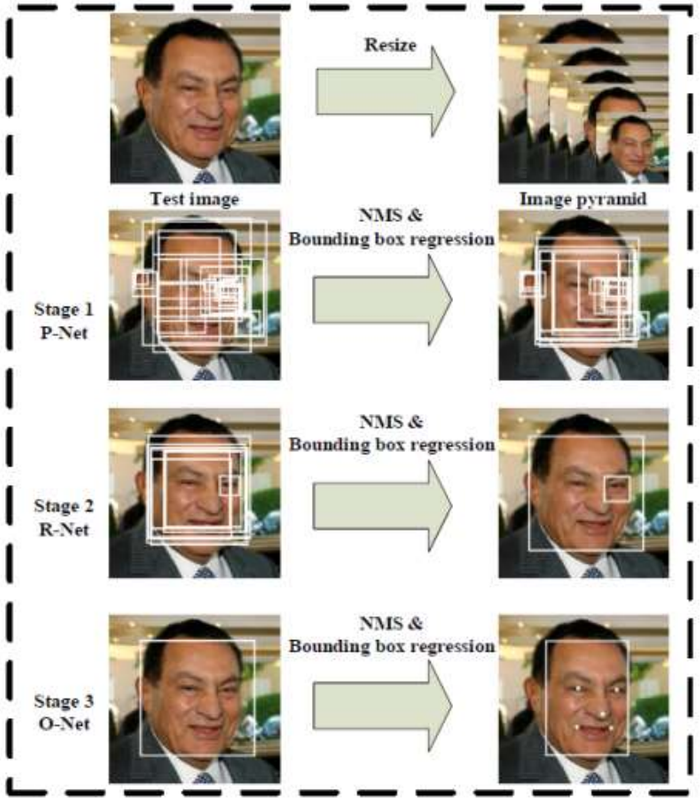
Problem statement

- **Persons of interest not known in advance**
  - add quickly new persons to be identified
  - add additional samples of a person not reliably detected in some cases
- **Analyse only once**
  - re-running (total) analysis is costly
  - it shall be possible to find people added later in already annotated content and link them to names

Incremental training approach

- **1. Multiple Face Detection using Multi-task cascaded CNNs[3]**
  - deep learning based approach, multi-scale detection
  - GPU support using TensorFlow framework





[3] Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks; K. Zhang et al.; IEEE Signal Processing Letters (SPL), vol. 23, no. 10, pp. 1499-1503, 2016

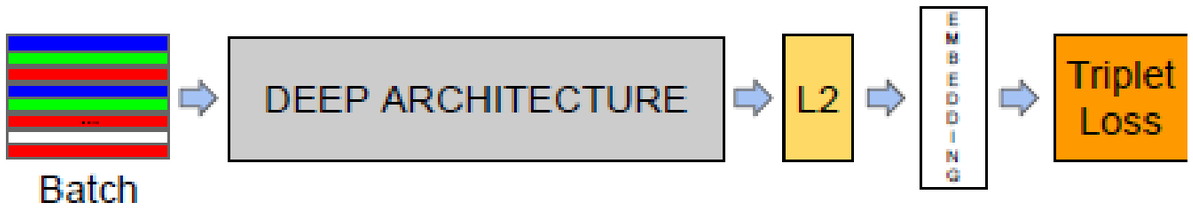
Incremental training approach

- 2. Alignment (normalization, canonical faces)

from FaceNet: A Unified Embedding for Face Recognition and Clustering,
Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015

Incremental training approach

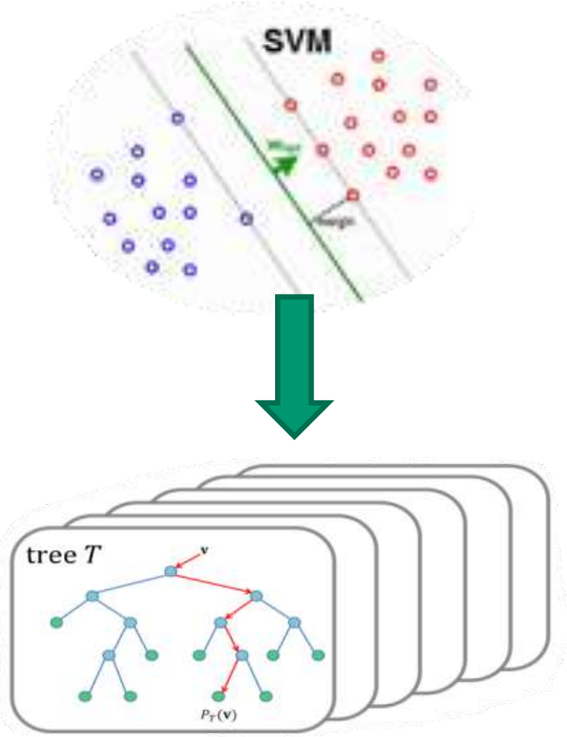
- **3. Face feature generation – using FaceNet[2]**
  - add FaceNet as feature extractor to the pipeline
  - convolutional neural net trained on large database
  - Inception ResNet v1 (http://arxiv.org/abs/1602.07261)
  - trained by stochastic gradient descent / backpropagation
  - very robust to compression artefacts / 128d – vector per norm. face



[2] FaceNet: A Unified Embedding for Face Recognition and Clustering,
Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015

Incremental training approach

- **4. Substitute SVM by fast, incremental (online) classificator**
  - Our choice: “(Online) Random Forest (ORCF) [1]”
  - Random Forests
    - many applications e.g.
      - density estimation, manifold learning semi-supervised learning….
    - based on ensemble of binary decision trees
    - simple node-decisions, majority-voting of leafs
    - best trade-off between accuracy and runtime
    - globally optimized classifier!
  - “online” is crucial for (fast, interactive) adding of „new“ faces!



[1] Saffari, A.; Leistner, C.; Santner, J.; Godec, M. & Bischof, H. On-line Random Forests Proceedings of the 3rd IEEE ICCV Workshop on On-line Computer Vision, 2009

Experiments / Results

- **Reference Evaluation (FaceNet – comp.)**
  - based on “Labelled Faces in the Wild” (LWF)[5]
  - proposed training-evaluation splitting

https://www.nist.gov/itl/iad/image-group/color-feret-database

99.8% CCR FaceNet-reference setting, SVM

95.0% our incremental approach with ORCF

- Note: Training time is ~10-20ms per each sample for both“
  - But + 1 face (incremental) ➔ 26s for entire LFW (org.)

<u>but max./const. cost <20ms (for our approach)!!!</u>

[5] Erik Learned-Miller, Gary B Huang, Aruni RoyChowdhury, Haoxiang Li, and Gang Hua. Labeled faces in the wild: A survey. In Advances in face detection and facial image analysis, pages 189-248. Springer, 2016.

Experiments / Results
Incremental learning on LFW database

- **Performance with minimal number of training images**
  - simulates minimalized user labelling
  - 3 - 4 images > 74% CCR / for only one train-image 27.6% CCR!
  - Intel i7-4790 CPU,3.60GHz, 16GB RAM, NVIDIA GeForce GTX 970

Nr. Training Img.		Nr. Test Img.	Accuracy		Classification-Time per sample	
used	total		Ours	[21]	Ours	[21]
1	1680	7484	27.59%	0.00%	1.37 ms	179.90 ms
2	1802	5804	68.00%	79.00%	0.82 ms	49.06 ms
3	1830	4903	74.16%	95.90%	0.59 ms	18.63 ms
4	1692	4293	86.05%	99.00%	0.45 ms	7.12 ms
5	1555	3870	88.68%	99.20%	0.40 ms	4.07 ms

Auto-training for unknown persons

- **Auto-training key-properties**
  - provides grouping of unknown / unlabelled faces
  - labelling can be done later (in a post processing step)
- **Requirements**
  - classification with few training samples is possible (proven)
  - ORCF provides reliable „confidences“ (c_f) for classification results
    - crucial for identification of first novel face!

Auto-Train algorithm

Auto-training for unknown persons
Algorithm outline

- **Classification is applied on each image!**
- **c_f > $Conf_{min}$ ➔ face reliably classified**
  - selected(!) features are stored for later retraining
  - or ongoing „improvement“ of classifier in FFS
- **Selection of features for FaceFeatureStore (FFS):**
  - $Conf_{min} < c_f < ConfStore_{min}$ ➔ face only classified
  - $ConfStore_{min} < c_f < Conf_{high}$
    - feature added to FFS, will improve classifier!
    - features near „decision boundary“ contribute a lot (similar to idea of boosting)
  - $Conf_{high} < c_f$
    - faces very reliably classified & no information gain in classifier

Auto-training for unknown persons algorithm outline

- **c_f < $Conf_{min}$ (low confidence classification)**
  - ➔ training of new face (person) or heavily distorted view
    - e.g. caused by viewpoint-change, lighting change, elapsed time (age) etc.
  - Note: **$Conf_{min}$** is very a crucial parameter!
    - (too) low: weak classifiers, flashlights trigger new faces, strongly different faces considered as matching
    - (too high): no robustness to short time distortions, gradual transitions etc.
    - tracking information helps in the case of videos

Auto-training for unknown persons algorithm outline

- **New face training strategy**
  - create candidate classifier (temporary updated)
    - multi-presentation for faster adaptation
  - check „feasibility“ of updated classifier
  - candidate classifier is only accepted if
    - confidence of new face is high $c_f > Conf_{high}$ OR
    - ratio between first/second match is high ($c_{f,first} / c_{f,\ second} > r_{min}$)
  - (if number of faces exceeds model ➔ re-training with FFS)

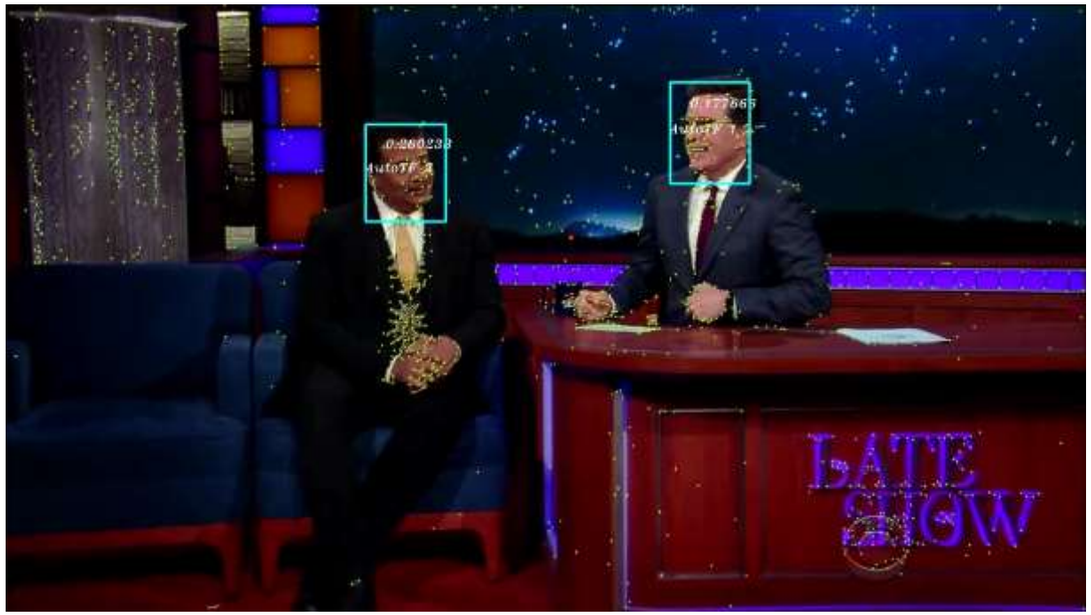
Auto-training for unknown persons algorithm outline

- **Classifier balancing (adaptation strategy for classifier)**
  - if new features are added
    - either auto-training (new face) or face feature store extension (reliable det. face)
  - focusing on samples yet not well classified
  - iteratively presenting „meaningful“ (not well classified) FFS-samples to get better (adjusted) classifier until convergence criterion is met

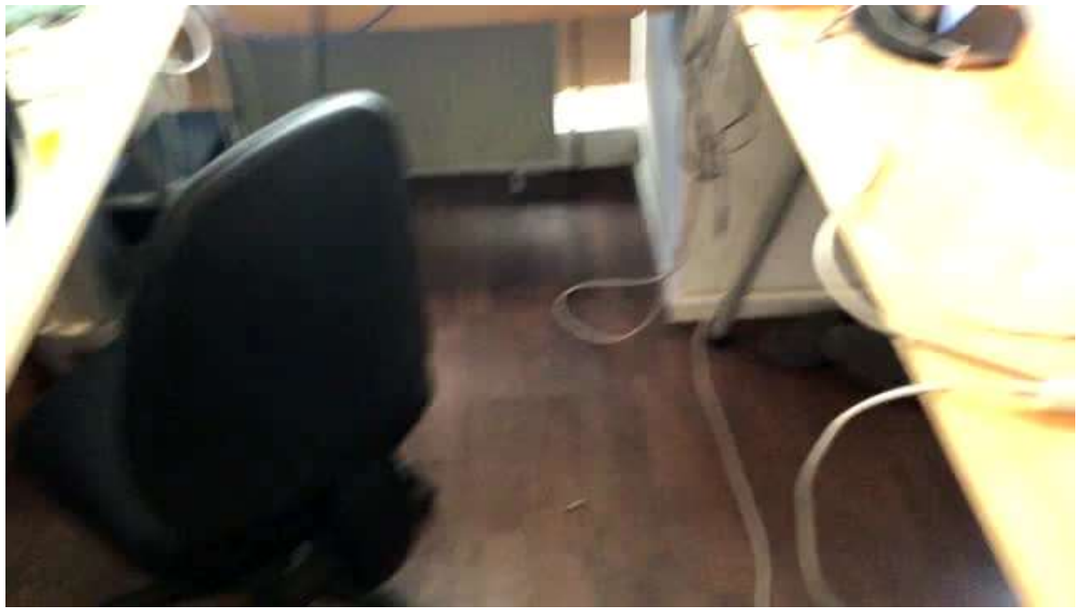
$$(c_f > Conf_{high}) \;\&\; (c_{f,first} / c_{f,\,second} > r_{min})$$

  - shuffle features and classes (randomized) avoiding ordering effects
  - max. number of iterations to limit balancing time

Experiments / results



Videos\20181126_095551_25_Faces_SVN_12903_PartialCut.avi



Videos\TAO_ClassificationAfterAutoTrain.avi

Summary

- **New approach for fast, incremental training of faces**
  - based on s.o.a. CNN-based detection and feature extraction
  - Introduce online-learning capability using online random forest for training faces as they arrive ➔ no need for re-training!
  - training with a few samples (even a single) one is possible
  - converges quickly to comparable (s.o.a.) correct classification rate on several DBs
- **New approach for automatic training of unknown faces**
- **Future / outlook**
  - Runtime improvement (C++, real-time)
  - avoid auto train “slipping” effect, large scale classifier optimization, …

Werner Bailer E-Mail: werner.bailer@joanneum.at

Martin Winter E-Mail: martin.winter@joanneum.at

Institute for Information and Communication Technologies

www.joanneum.at/digital