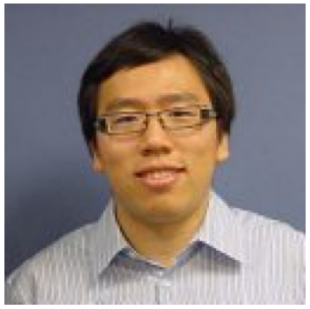
# Deep Face Recognition

## Tutorial at SIBGRAPI 2018


**Iacopo Masi***
**USC ISI**



Yue Rex Wu
USC ISI

Tal Hassner
Open University of Israel

Prem Natarajan
USC ISI


USC University of Southern California

THE OPEN UNIVERSITY OF ISRAEL

# Face Matching

## Deep Face Recognition: A tutorial

SIBGRAPI 2018

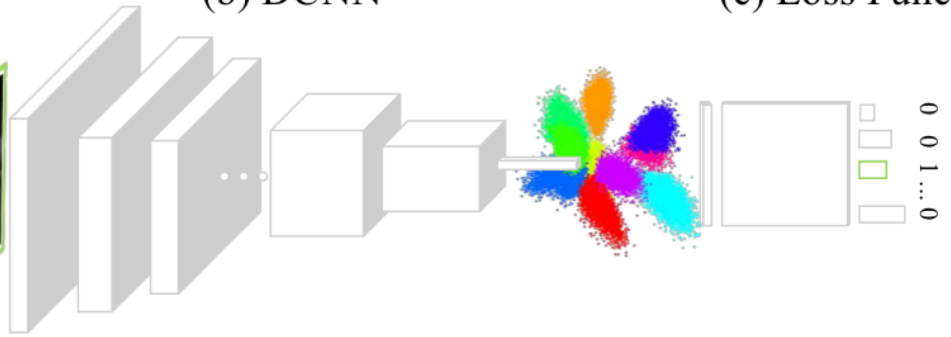
# Deep Face Recognition Pipeline

## (I) Training Phase

(a) Training set with identity labels

**Face Preprocessing**

(b) DCNN




(c) Loss Function

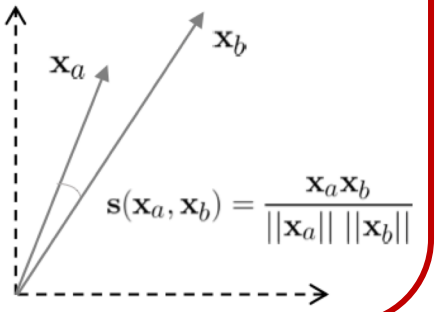
Maps image intensity into a feature

Discriminate between subjects (**classification**)

## (II) Testing Phase



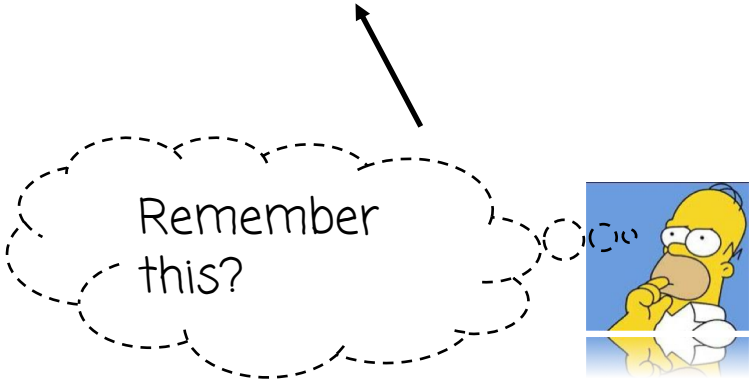

$x_a$ $x_b$



# How to actually perform recognition

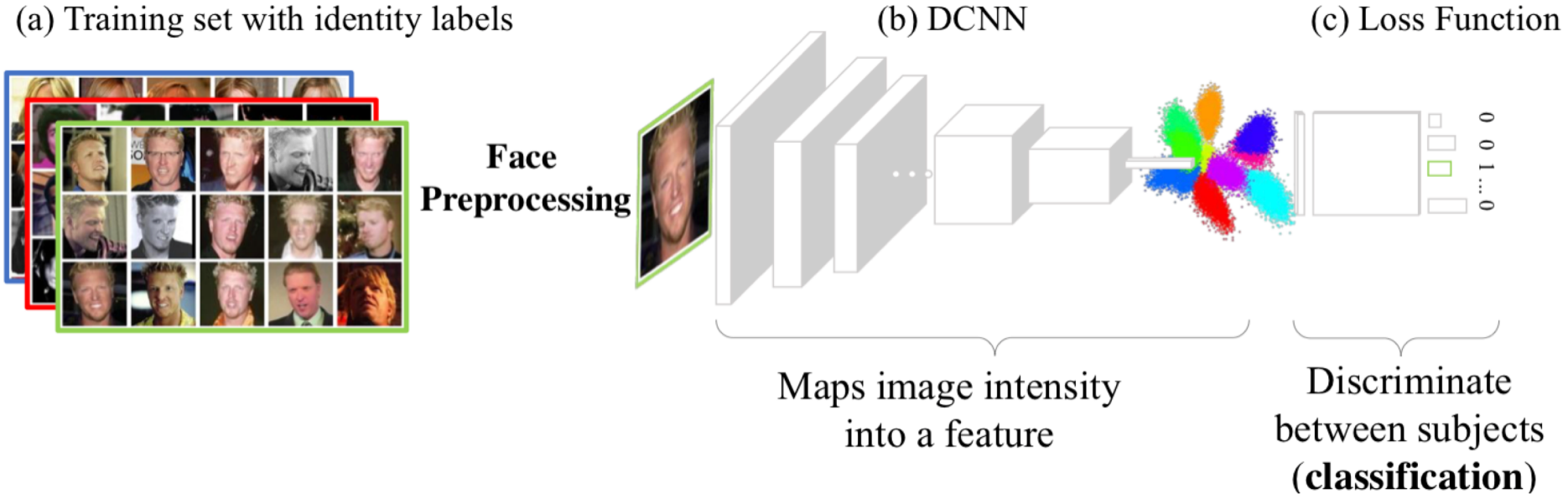
OK, now we have a network trained for face recognition, what do we do?

## Basic Transfer Learning Recipe



# Basic Transfer Learning Recipe:

- Train a ConvNet on a set of subjects for face recognition
- The layer prior the classification (feature embedding) layer will learn discriminative face representations
- At test time, use the feature embedding as a discriminative face descriptor



# Matching Face Descriptors


- Chop off the classification layer
- Extract the activations from the layer prior the classification layer
- Use these activations as feature descriptors:
  - If trained with SoftMax+CE, use **cosine distance or correlation**
  - If trained w/ a loss enforcing euclidean margins, use L2 distance.
  - If features are normalized to have unit L2 norm at test time, then cosine distance ~ L2 distance

# Augmentation at test-time

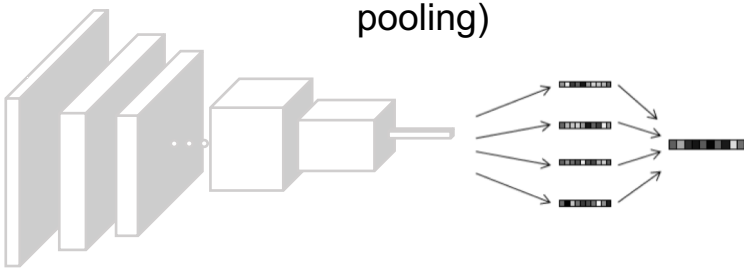
**Key Idea:**

- **Training time**: Add random transformation in the pixel space (acts as noise)
- **Testing**: Marginalize over the noise

Input Image

**Add noise**

**Marginalize over it** (usually average pooling)



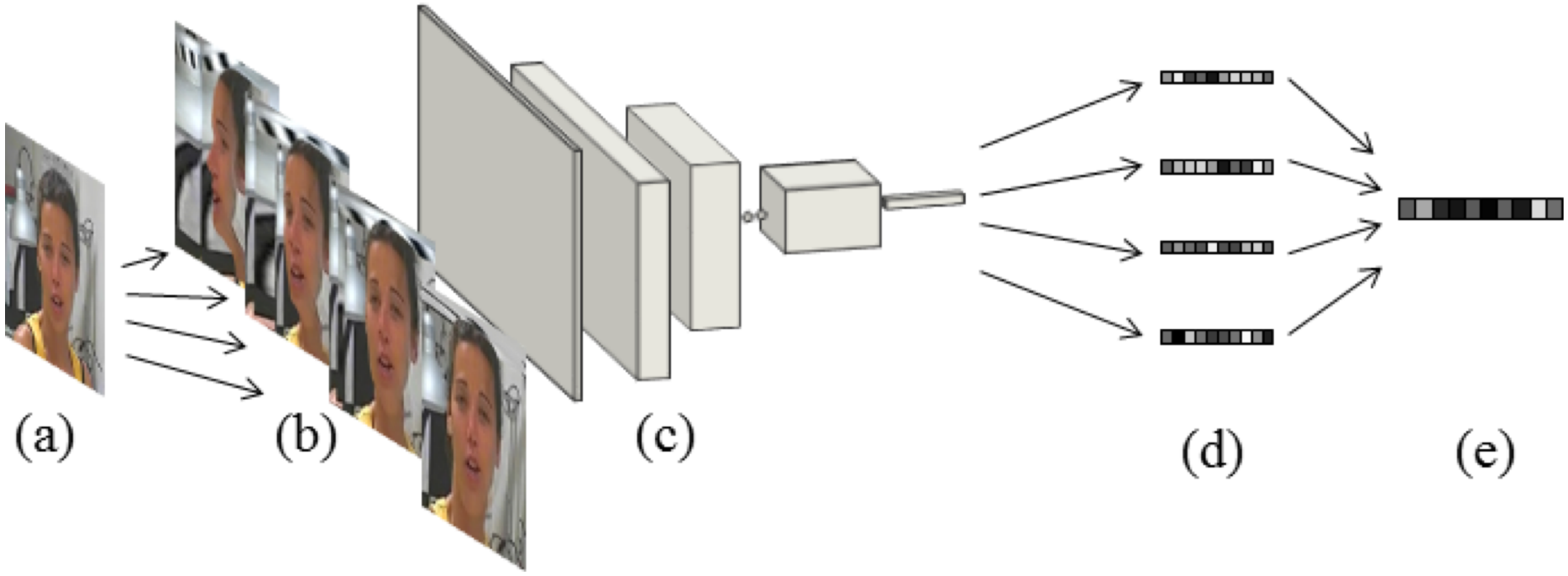
Perturbations (leaving the class unchanged)
Usually transformation used in object recognition:

- Flipping
- Multiple Crops
- Color Jittering etc...

# Face-Specific Augmentation at test-time

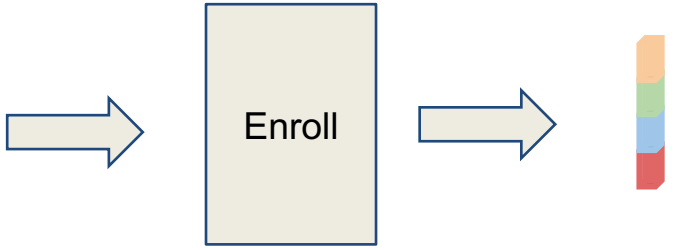
**Key Idea:**

- The image does contain a face. Why not perturbing the input with face-specific transformations?
- **Testing**: Averaging synthesized images and in-plane aligned images



# How to enroll with multiple media?

IJB-A brings up the question: *"How do we build a model of a subject given multiple heterogeneous media?"*

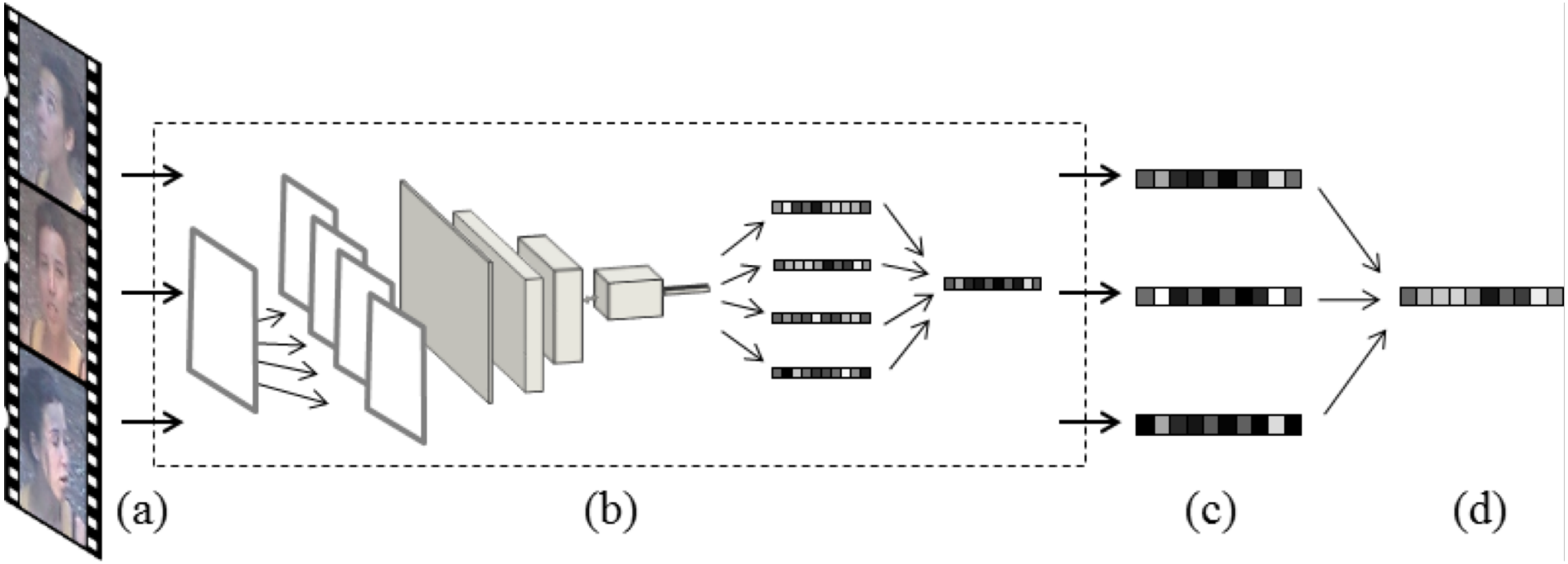



- Raw media (still images, videos)
- Order of megabytes
- Multiple faces

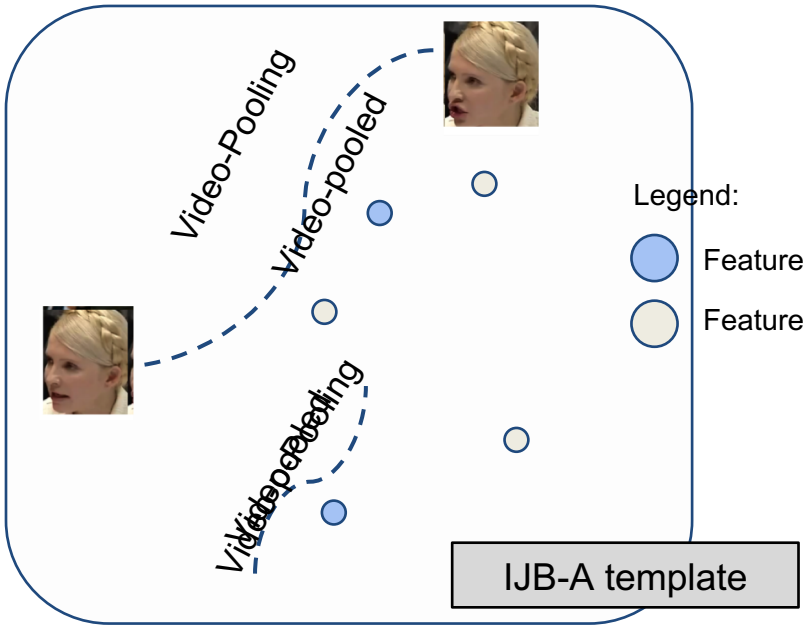
- Compact template
- Fixed size (**8KB!**)
- (ResNet encoding is 2048 floating point)
- **Invariant** to the media or video length!

# Video Pooling

We can use the same rationale in case we have videos

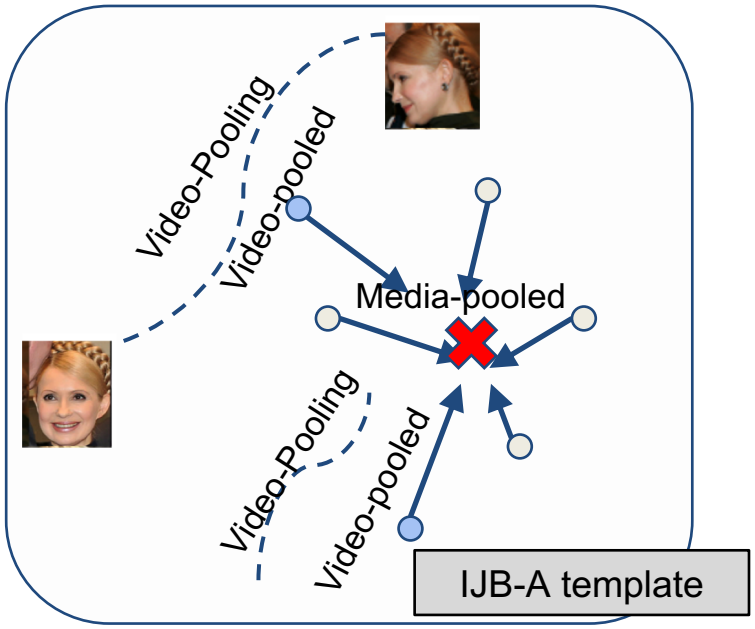


# Video and Media Pooling




Video-Pooling

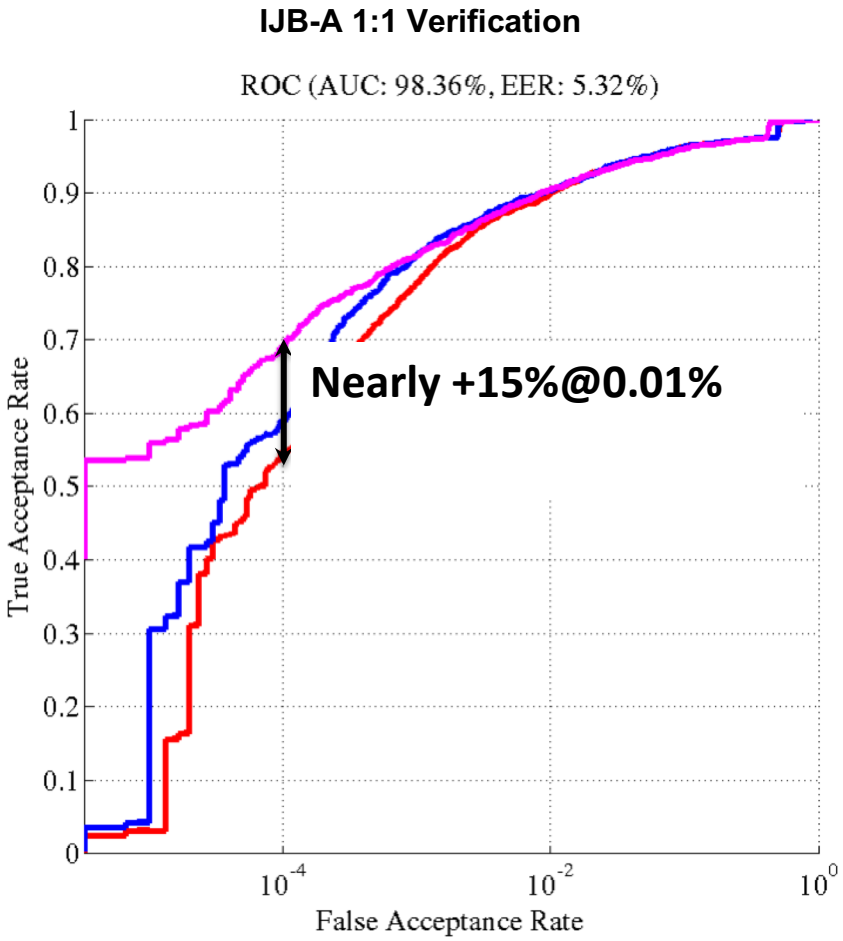
[Masi, Tran, Hassner *et al.* ECCV'16]




Media-Pooling

[Crosswhite *et al.* FG'17]

# Video and Media Pooling



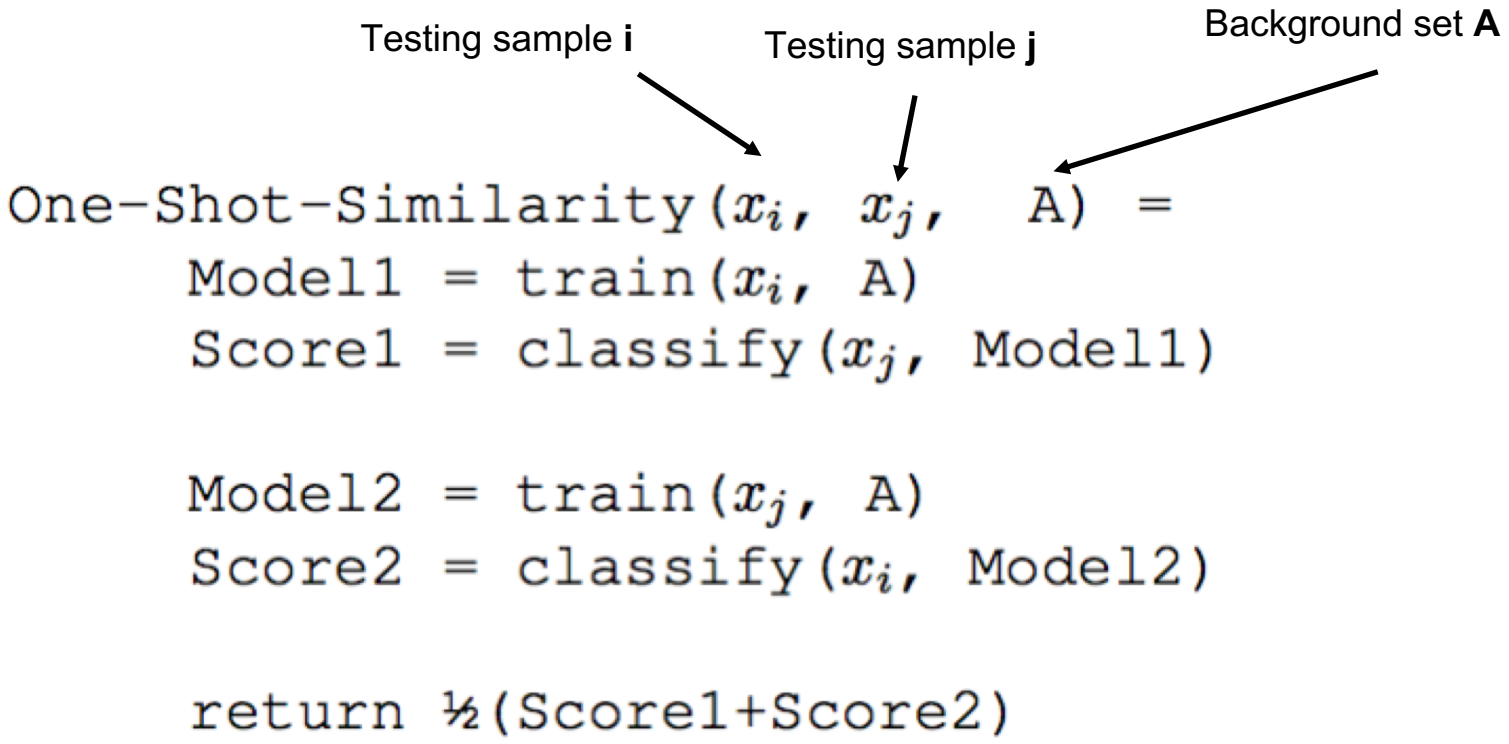

**Legend**

- Media-pooling
- Video-pooling + score fusion
- Score fusion

# Lazy learning: One-Shot Similarity Kernel

**Lazy Learning Philosophy:**

- Postone all (or part of) the training at test time.
- Learn a classifier online, for each sample.

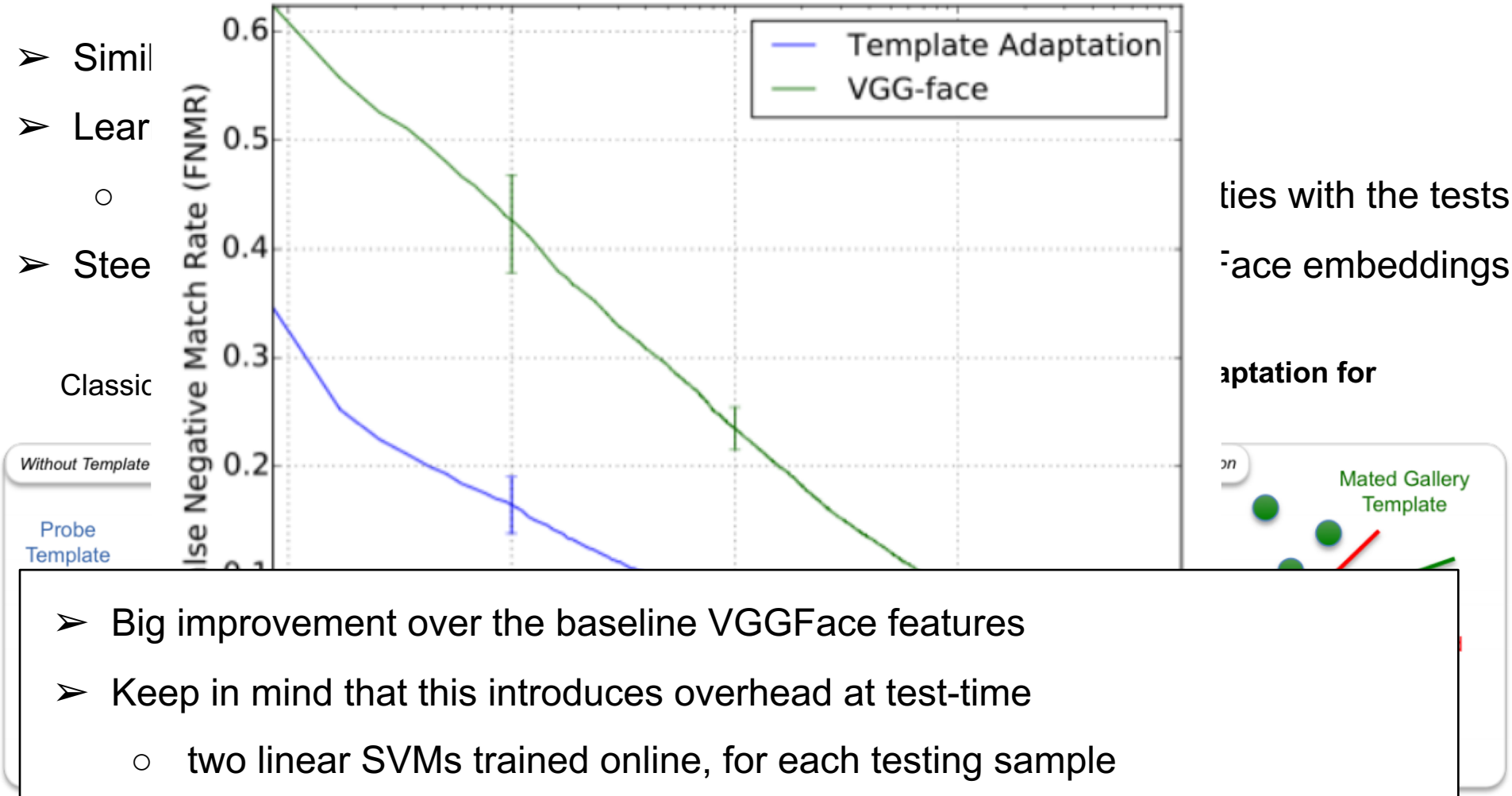



[Wolf *et al.* ICCV'09]

# Lazy learning: Template Adaptation

- Simil
- Lear
  - ties with the tests
- Stee
  - ace embeddings

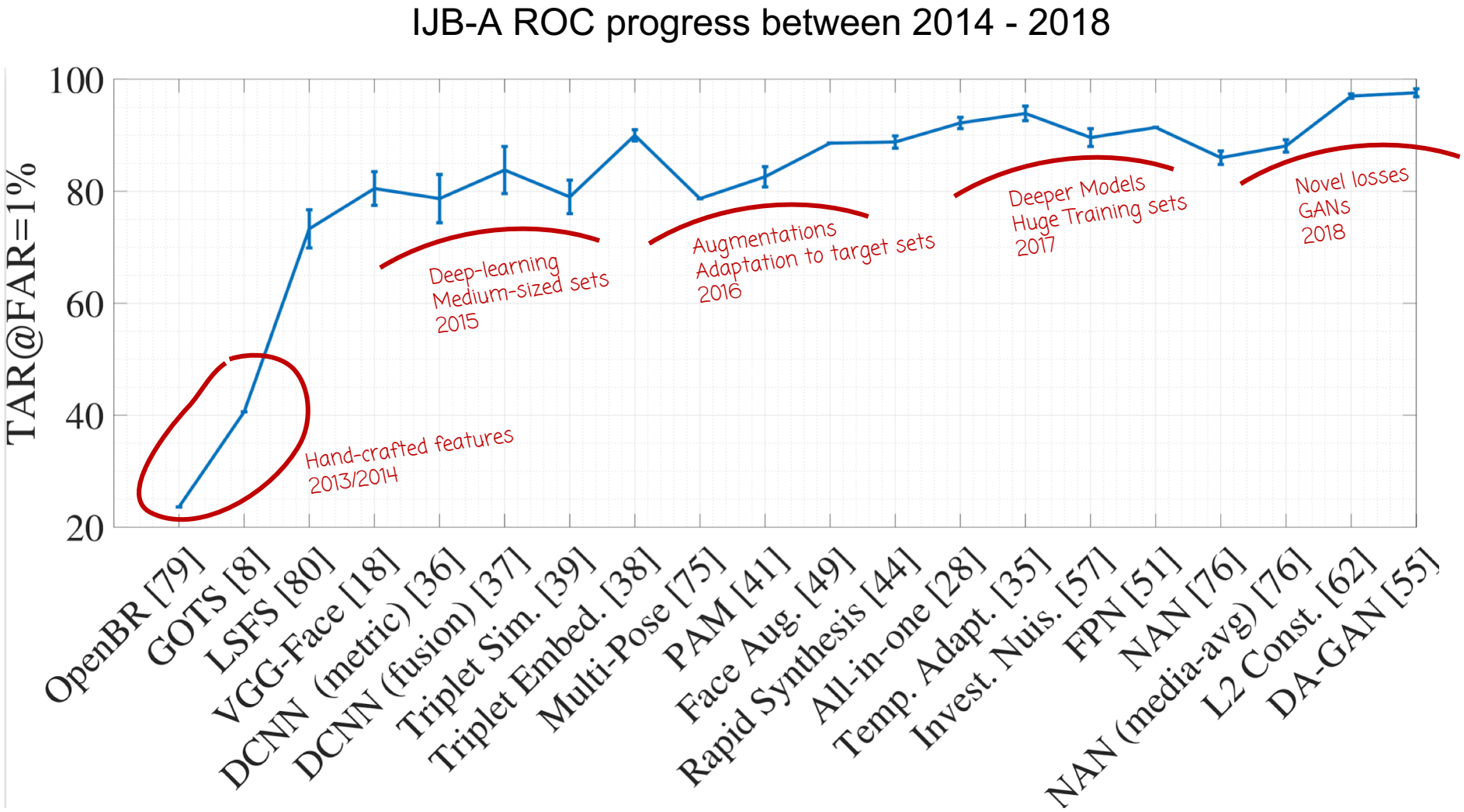
Classic aptation for




- Big improvement over the baseline VGGFace features
- Keep in mind that this introduces overhead at test-time
  - two linear SVMs trained online, for each testing sample

[pictures from Crosswhite et al. FG'17]

# How much all this contributed?



# Is Face Recognition solved once again?

- In IJB-A, “templates” are defined manually. How to create the templates automatically?
  - **Video-based** Face Recognition
  - There is a need for a better synergy between face recognition and multiple person tracking. Are those the same thing, actually?
- Automatic self-organization of a large corpus of unlabeled faces:
  - *“If we are given an unlabeled corpus of data, such as videos and images, with the scope of **clustering** same identities, what is the best way to proceed? Simply train a DCNN offline and then use standard clustering methods to discover novel subjects? Or explore the myriad of unlabeled data we have?”*
- FAIR representation learning for face recognition

# The End

# Thanks for your attention!

# References


1. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the devil in the details: Delving deep into convolutional nets,” in *BMVC*, 2014.
2. I. Masi, Y. Wu, T. Hassner, P. Natarajan “Deep Face Recognition: a Survey”, SIBGRAPI18
3. I. Masi, S. Rawls, G. Medioni, and P. Natarajan, “Pose-Aware Face Recognition in the Wild,” in *CVPR*, 2016.
4. J.Yang, P.Ren, D.Zhang, D.Chen ,F.Wen, H.Li, and G.Hua, “Neural aggregation network for video face recognition,” in *CVPR*, July 2017. [77] L. Wolf, T. Hassner, and Y. Taigman, “The one-shot similarity kernel,” in *CVPR*, 2009.
5. P. J. Phillips, P. Grother, and R. Micheals, “Evaluation methods in face recognition,” in *Handbook of Face Recognition*. Springer, 2011, pp. 551–574.
6. K. Kim, Z. Yang, I. Masi, R. Nevatia, and G. Medioni, “Face and body association for video-based face recognition,” in *WACV*, 2018.
7. L. Wolf, T. Hassner, and Y. Taigman, “The one-shot similarity kernel,” in *CVPR*, 2009.
8. N. Crosswhite, J. Byrne, C. Stauffer, O. Parkhi, Q. Cao, and A. Zisser- man, “Template adaptation for face verification and identification,” in *AFGR*, 2017.