

Do We Really Need to Collect Millions of Faces for Effective Face Recognition?

Iacopo Masi^{,1}, Anh Tuan Tran^{*,1}, Tal Hassner^{*,2,3},
Jatuporn Toy Leksut¹ and Gerard Medioni¹*

1. Institute for Robotics and Intelligent Systems, USC, CA, USA
2. Information Sciences Institute, USC, CA, USA
3. The Open University of Israel, Israel

* Denotes equal authorship

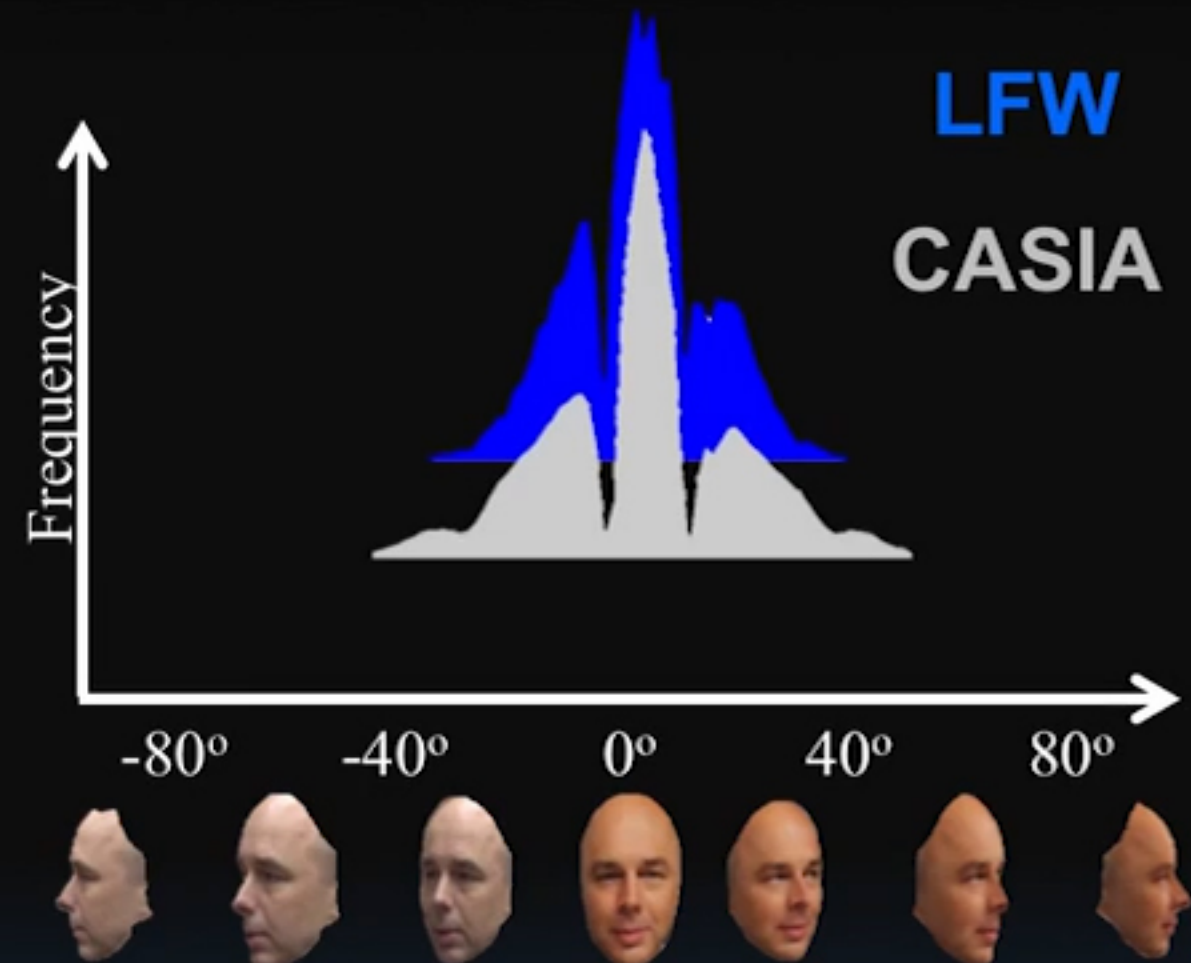


Motivation

Intra-subject Variations

Method	# train img.	subj.	img. / subj.
DeepFaces'14 (FB)	4 m	4,030	1k
VGG Face'15	2.6 m	2,622	1k
Face Net'15 (Google)	200 m	8 m	25
Fusion'15 (FB)	500 m	10 m	50
MegaFace'16	1.02 m	690.5 k	1.5

Pose (yaw) Variations



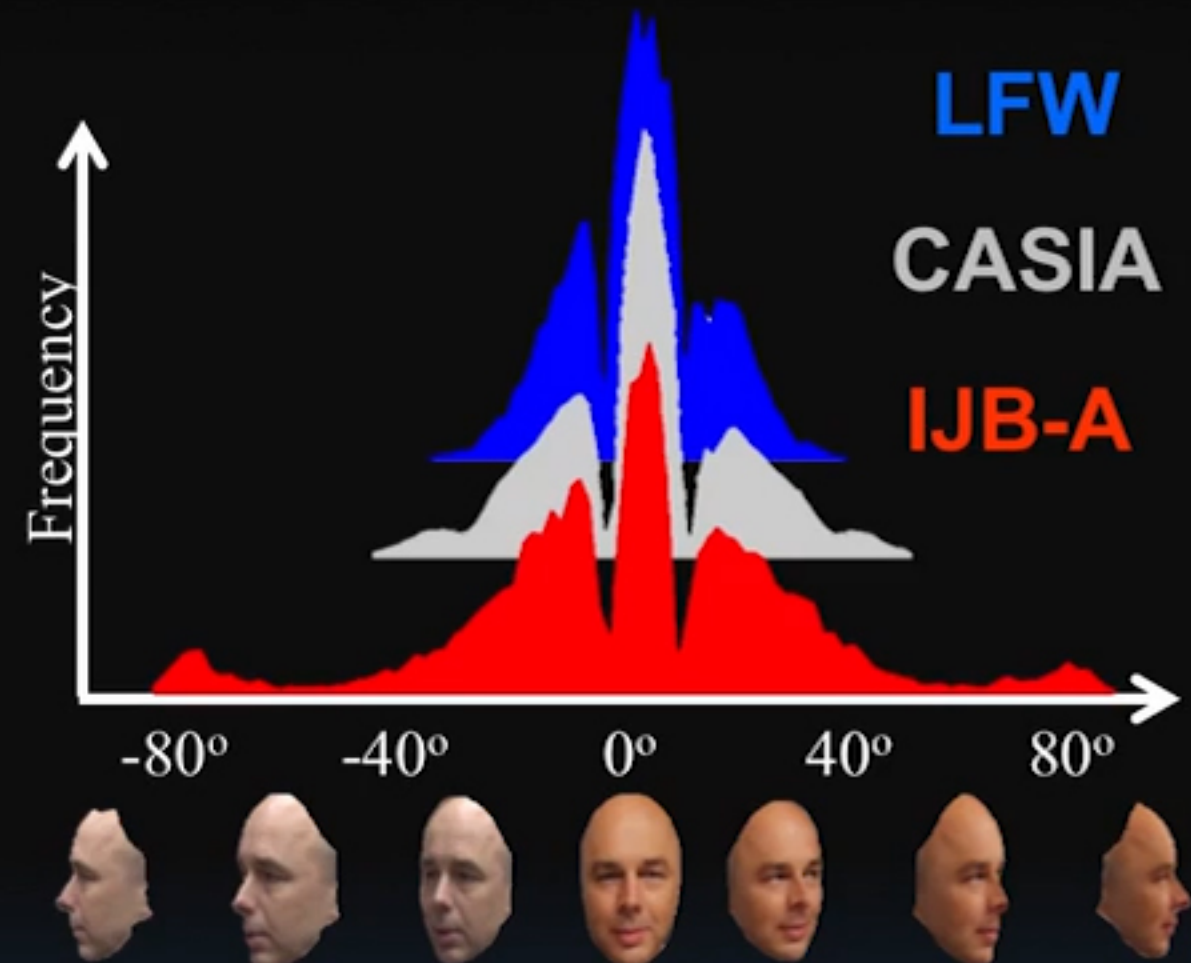
Even with lots of resources, it's hard to ensure sufficient intra-subject and pose variations

Motivation

Intra-subject Variations

Method	# train img.	subj.	img. / subj.
DeepFaces'14 (FB)	4 m	4,030	1k
VGG Face'15	2.6 m	2,622	1k
Face Net'15 (Google)	200 m	8 m	25
Fusion'15 (FB)	500 m	10 m	50
MegaFace'16	1.02 m	690.5 k	1.5

Pose (yaw) Variations




Even with lots of resources, it's hard to ensure sufficient intra-subject and pose variations



The two keys to successful face recognition

1. During training: Learn the variability of same-subject appearances

Increase training set intra-subject appearance variations



2. During testing: Make same subjects easier to compare

Reduce test set intra-subject appearance variations



Domain (face) specific data augmentation

Increasing appearance variability in the training set



3D pose



Domain (face) specific data augmentation

Increasing appearance variability in the training set



3D pose



3D shape



Domain (face) specific data augmentation

Increasing appearance variability in the training set



3D pose



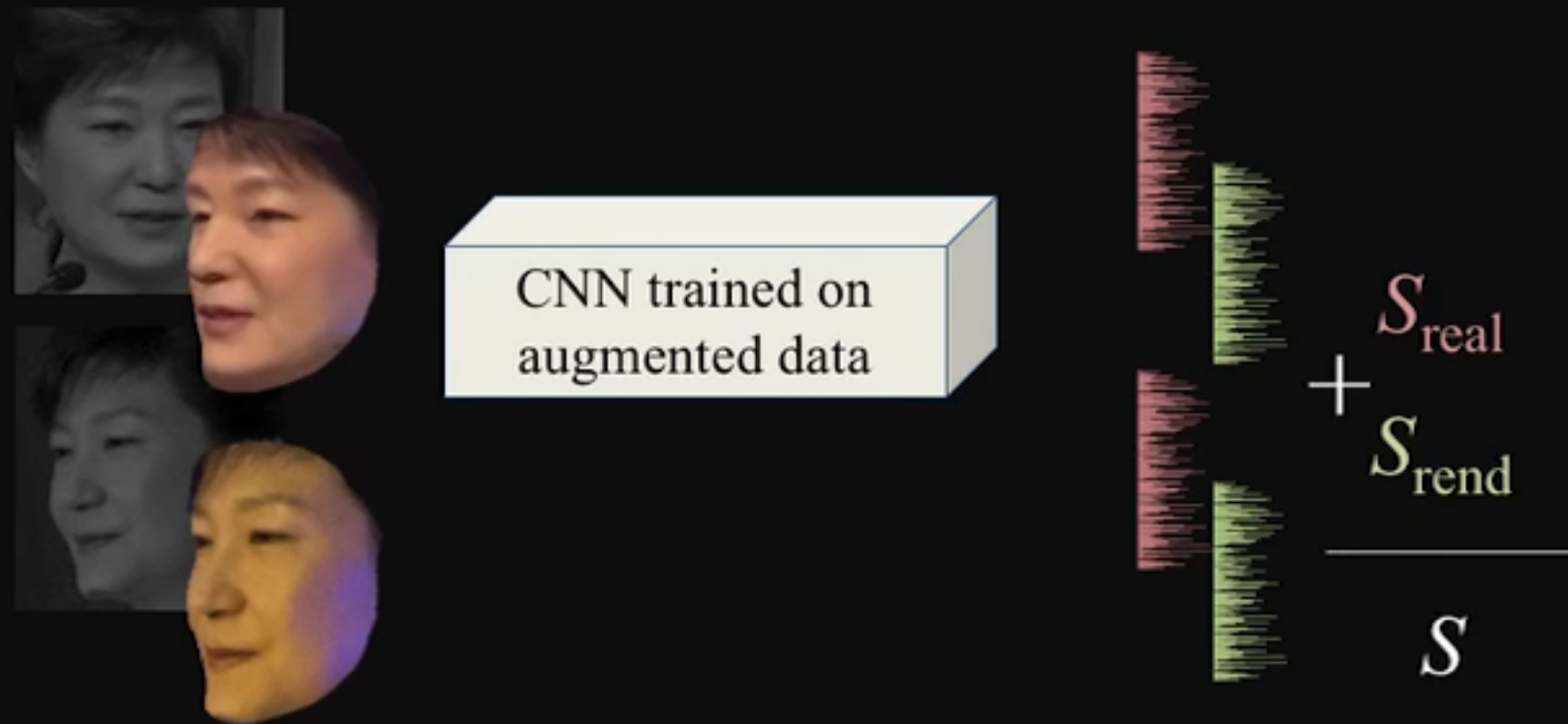
3D shape



Expression



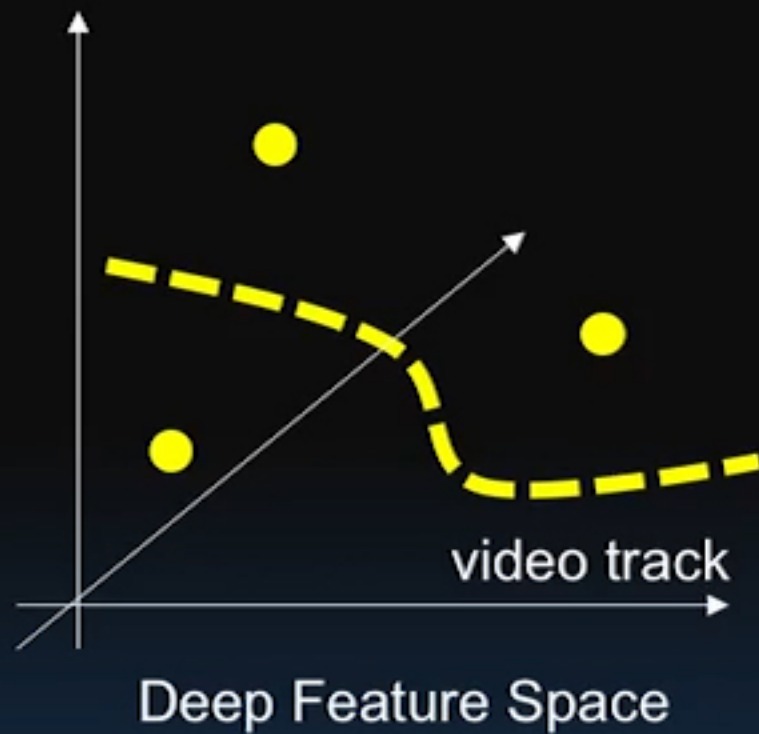
Reducing appearance variability in the test set



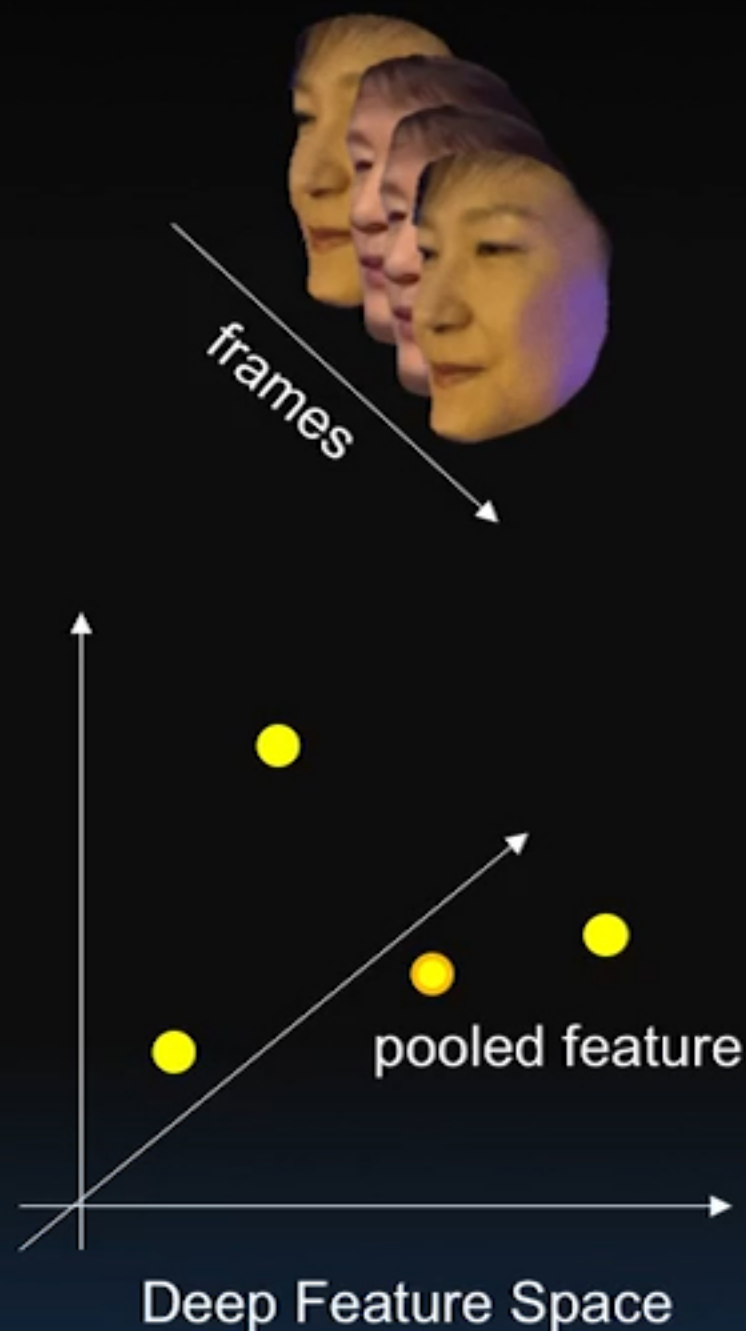
Reducing appearance variability in the test set



- In case we have multiple frames from videos in a template (set of images)



Reducing appearance variability in the test set



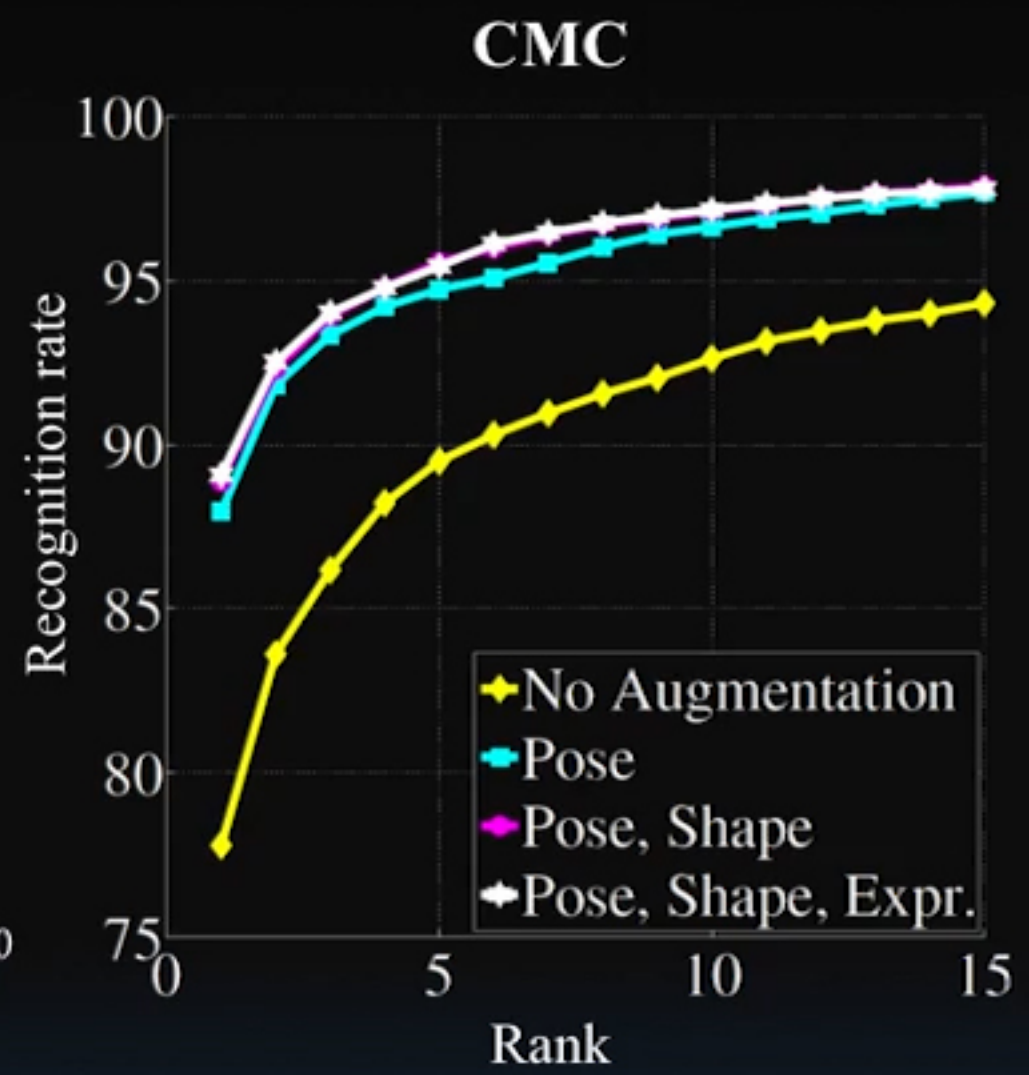
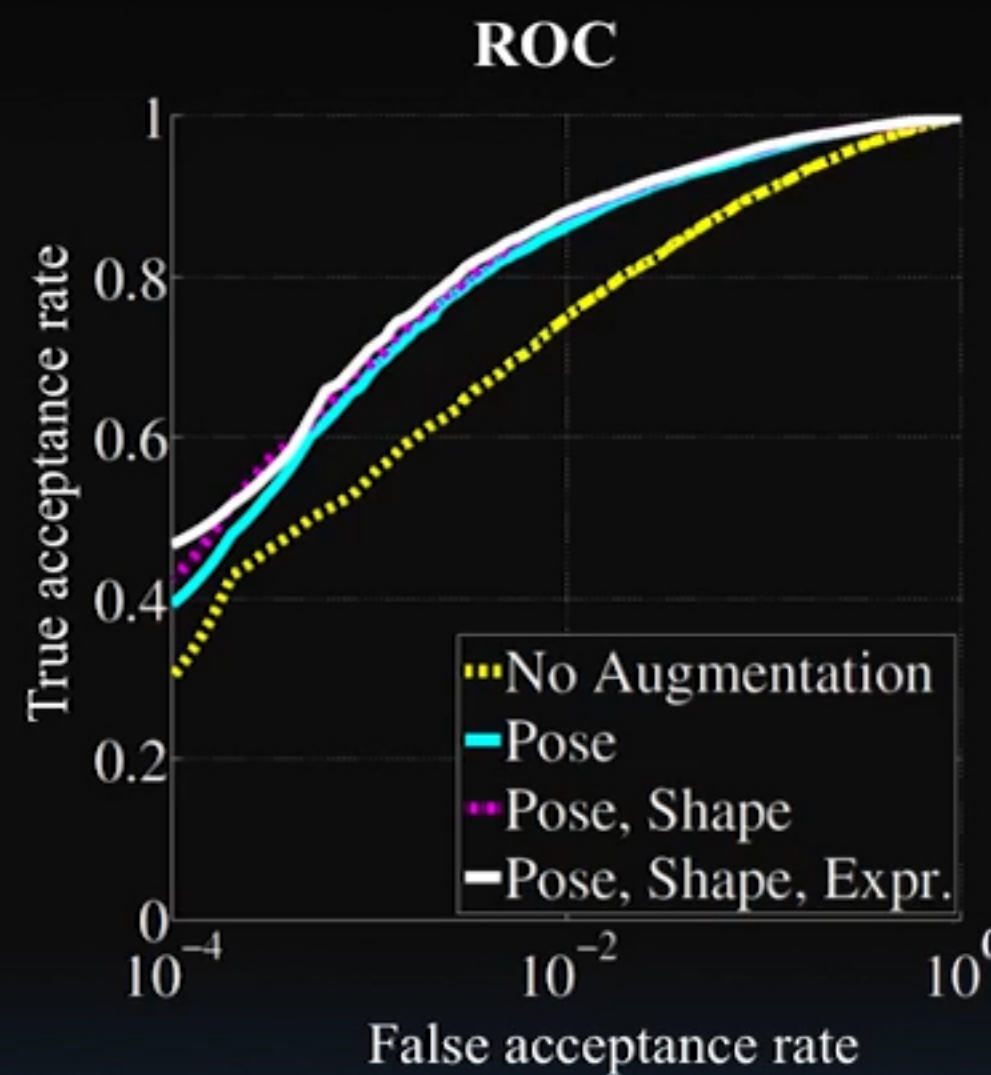
- In case we have multiple frames from videos in a template (set of images)
- Each video track is pooled across frames in the feature space with average
- Pair-wise similarity scores are then pooled with Soft-Max operator

$$\mathbf{s}^* = \frac{\sum_{i=1}^N \mathbf{s}_i \exp(\alpha \mathbf{s}_i)}{\sum_{i=1}^N \exp(\alpha \mathbf{s}_i)}$$

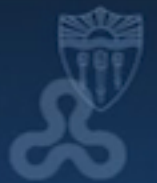


What does this do to performance?

Example: IJB-A



Training better CNNs with less effort using domain (face) specific data augmentation!!!



Do We Really Need to Collect Millions of Faces for Effective Face Recognition?

Iacopo Masi^{,1}, Anh Tuan Tran^{*,1}, Tal Hassner^{*,2,3},
Jatuporn Toy Leksut¹ and Gerard Medioni¹*



Come see us at the poster (S-4B-09) or visit our webpage for more information, code and results

Thank you!

