A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition

Mostafa Mehdipour Ghazi, Hazım Kemal Ekenel
Istanbul Technical University, SiMiT Lab
ekenel@itu.edu.tr

IEEE Computer Society Workshop on Biometrics 2016

Outline

- **Motivation**
- **Face Representation with Convolutional Neural Networks (CNNs)**
  - VGG-Face Network
  - Lightened CNN
- **Experiments and Results**
  - Face Occlusion
  - Illumination Variations
  - Pose Variations
  - Misalignment
- **Conclusion**

Motivation

- Deep learning based face recognition
  - Increasingly applied for face recognition with promising results
  - Their capability to handle individual appearance variations has not been thoroughly assessed

Illumination

Pose

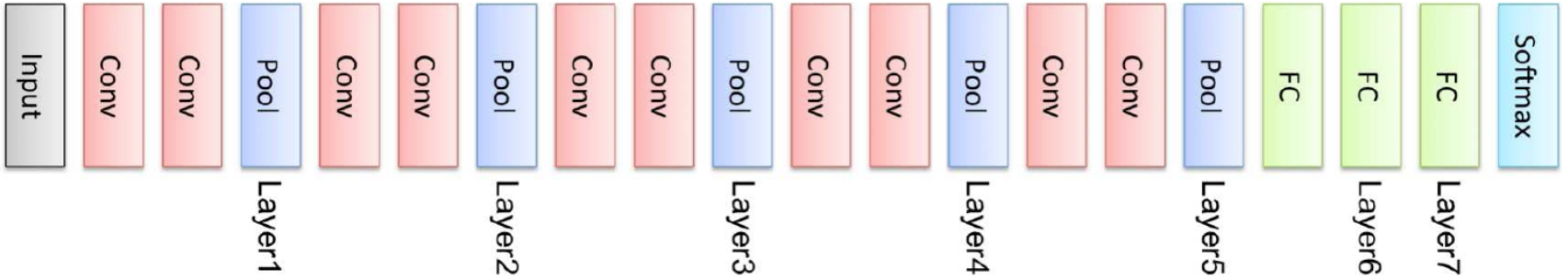
Occlusion

Alignment

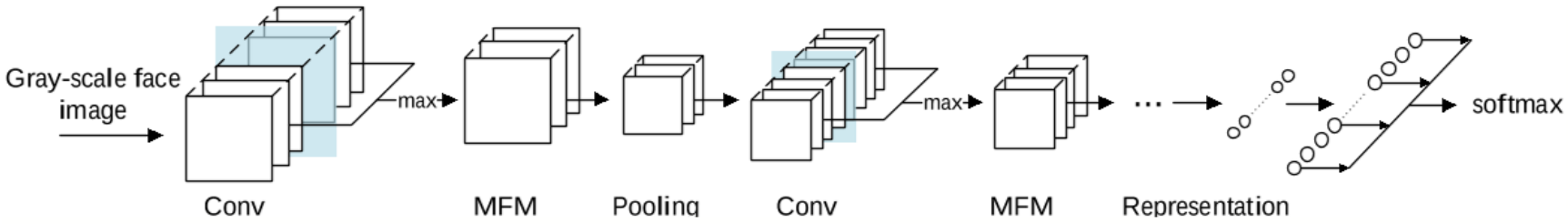
Face Representation

- Higher layer portion of learned weights from pre-trained CNNs

- VGG-Face with a very deep architecture



- Lightened CNN is a computationally efficient network



Datasets

- AR Face Database for Face Occlusion

- CMU PIE Database for Illumination Variations

- Extended Yale Dataset B for Illumination Changes

- Color FERET Database for Pose Variations

- The FRGC Database for Misalignment

Implementation

- Using the Caffe deep learning framework
- Face alignment with respect to eye centers
- Feature extraction from the first fully-connected (FC) layers
  - VGG-Face: 224x224 color image → 4096-D feature set in FC6/FC7
  - Lightened CNN: 128x128 grey image → 512-D feature set in FC1
- Classification using the nearest neighbors with cosine distance

Experiments on AR Database

- Frontal faces of subjects participated in two sessions
- **Training:** One image per subject with neutral expressions from the first session → 110 images/subjects
- **Testing:** Two images per subject per session, one wearing a pair of sunglasses (upper face occlusion) and one wearing a scarf (lower face occlusion) → 110 images per session

Testing Set	VGG-Face		Lightened CNN
	FC6	FC7	
Sunglasses Session 1	33.64	**35.45**	5.45 (A)
Scarf Session 1	86.36	**89.09**	12.73 (A)
Sunglasses Session 2	**29.09**	28.18	7.27 (B)
Scarf Session 2	**85.45**	83.64	10.00 (A)

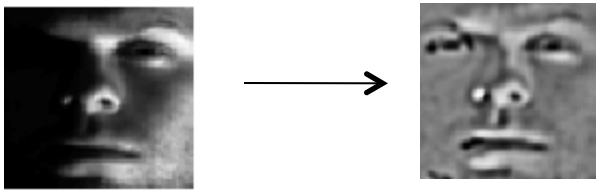
Experiments on CMU PIE Database

- Facial images from 68 subjects
- **Training:** One frontally illuminated face per subject → 68 images
- **Testing:** 20 images with varying illumination per subject → 1360 images

	VGG-Face		Lightened CNN
	FC6	FC7	
Accuracy	**93.16**	92.87	20.51 (A)

Experiments on Extended Yale B

- Facial images of 38 subjects under 64 illumination variations
- **Training:** Subset 1 with the lighting angles <12° → 266 images
- **Testing:** Subset 2 (20°–25°), subset 3 (35°–50°), subset 4 (60°–77°), subset 5 with angles > 77°
- Pre-processing subsets 4 and 5



Testing Set	VGG-Face		Lightened CNN
	FC6	FC7	
Subset 2	**100**	**100**	82.43 (A)
Subset 3	88.38	**92.32**	18.42 (B)
Subset 4	46.62	**52.44**	8.46 (B)
Subset 5	13.85	**18.28**	4.29 (B)
Preprocessed Subset 4	71.80	**75.56**	26.32 (A)
Preprocessed Subset 5	73.82	**76.32**	24.93 (A)

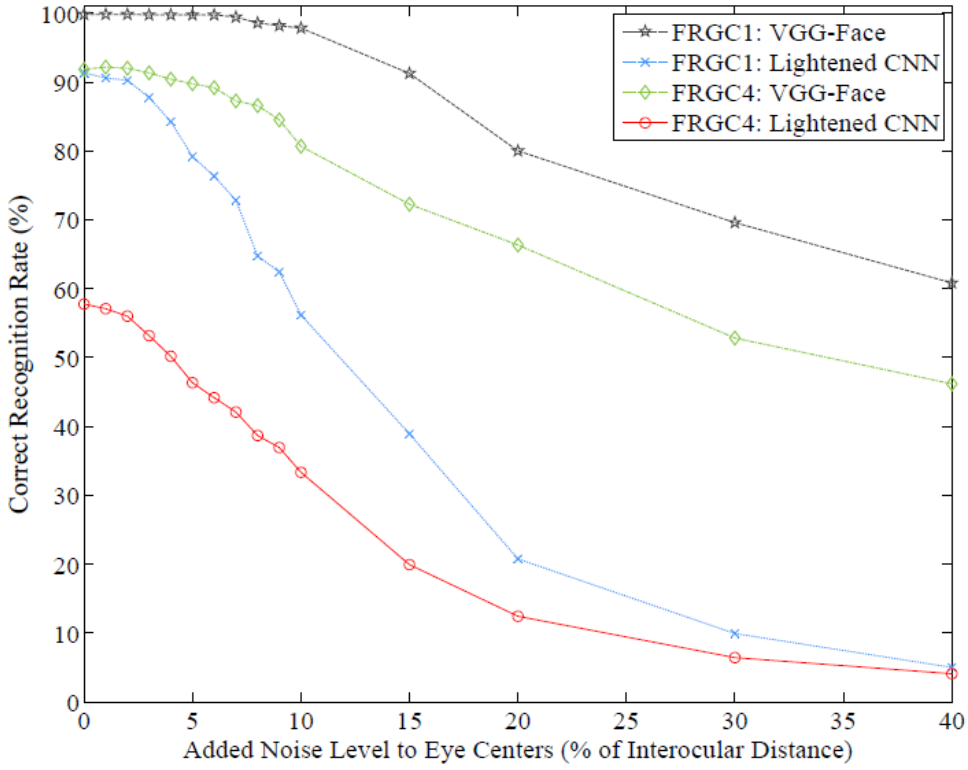
Experiments on FERET Database

- Facial images of 994 subjects with 13 different poses
- **Training:** Regular frontal image set → 994 images
- **Testing:** Quarter left and quarter right (22.5°), half left and half right (67.5°), profile left and profile right poses (90°)

Testing Set	VGG-Face		Lightened CNN
	FC6	FC7	
Quarter Left	**97.63**	96.71	25.76 (A)
Quarter Right	**98.42**	98.16	26.02 (A)
Half Left	**88.32**	87.85	6.08 (B)
Half Right	**91.74**	87.85	5.98 (A)
Profile Left	40.63	**43.60**	0.76 (B)
Profile Right	43.95	**44.53**	1.10 (B)

Experiments on FRGC Database

- Frontal faces in controlled and uncontrolled environments and two sessions
- **Training:** Fall 2003 subsets
- **Testing:** Spring 2004 subsets



Experiments with Facial Bounding Box Extension

- All images are aligned and cropped into an extended facial patch to include all parts of the head (ears, hair, and chain)

Training Set	Testing Set	VGG-Face (FC6)	Previous
AR Neutral Set 1	AR Sunglasses Set 1	44.55	33.64
AR Neutral Set 1	AR Scarf Set 1	93.64	86.36
AR Neutral Set 1	AR Sunglasses Set 2	39.09	29.09
AR Neutral Set 1	AR Scarf Set 2	91.82	85.45
CMU PIE Train	CMU PIE Test	97.72	93.16
Ext. Yale Set 1	Ext. Yale Set 2	100	100
Ext. Yale Set 1	Ext. Yale Set 3	94.52	88.38
Ext. Yale Set 1	Ext. Yale Set 4	56.58	46.62
Ext. Yale Set 1	Ext. Yale Set 5	27.56	13.85

Conclusion

- A comprehensive evaluation of deep representations under various conditions: pose, illumination, occlusion, and misalignment
- Deep representations are not able to achieve state-of-the-art results against pose, illumination, and occlusion. Such variations should be taken into account during training or preprocessing methods should be employed along with pre-trained models
- Deep representations are robust to misalignment. Facial feature localization errors of up to 10% of the interocular distance are tolerated
- The VGG-Face model is shown to be more transferable compared to the Lightened CNN model

Thank You!

Hazım Kemal Ekenel
ekenel@itu.edu.tr