DeepFace:

Closing the Gap to Human-Level Performance in Face Recognition

Problem: one-shot learning (not enough dataset from single person)

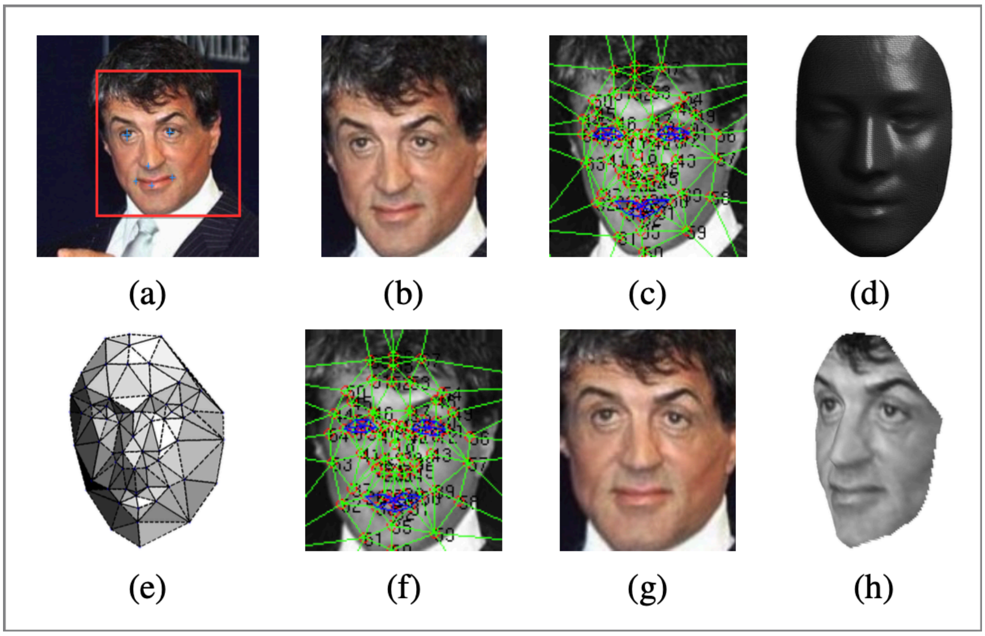
DeepFace: 3D Face Alignment + DNN Architecture

—> Closely approaching human-level performance

1. 3D Face Alignment

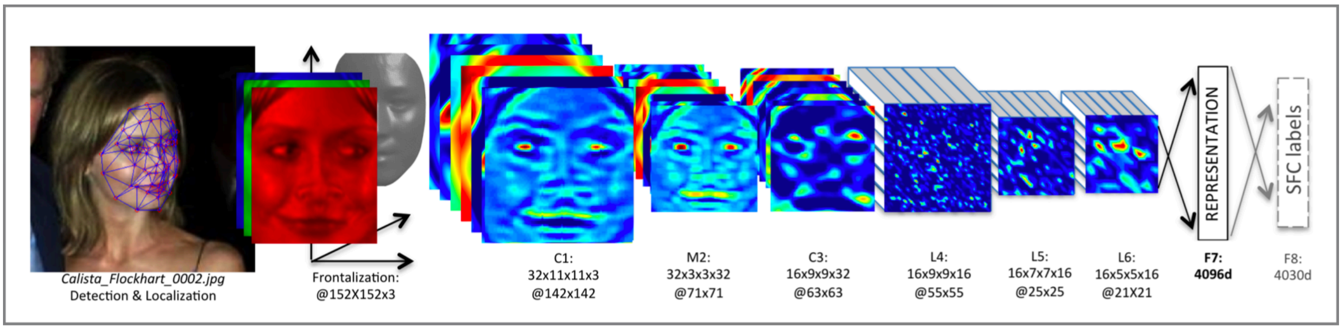
- "Unconstrained" scenario (non-planarity, non-rigid face expressions)
- 3D Modeling (frontalization)

1. 2D Crop: 6 fiducial points ==> 2D similarity transformation
2. 3D Alignment: warp the 2D-aligned crop to the 3D shape



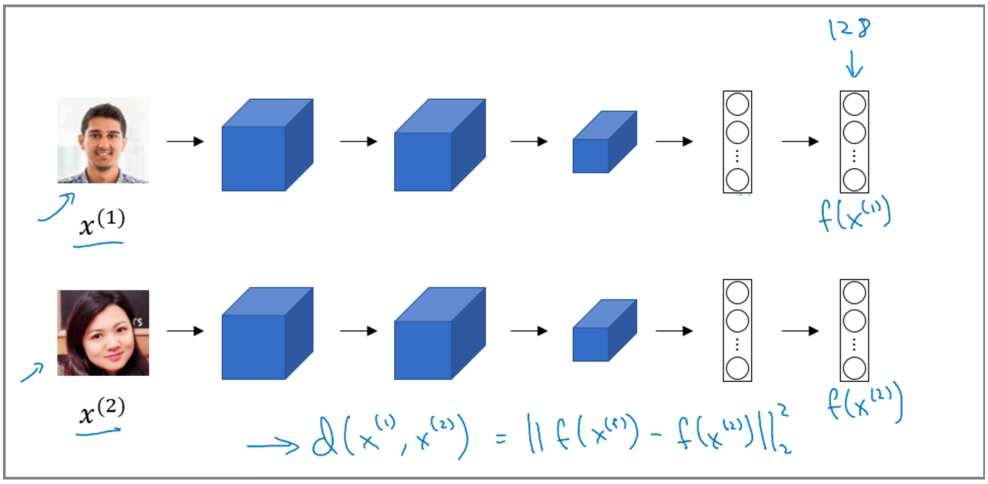
(a) Detected face (b) 2D crop (e) Transformed into 3D image plane
(g) Final frontalized crop

2. Representation



DNN Architecture

1. <Pre-processing> Single front-end (C1-M2-C3): several layers of pooling would cause detailed facial structure and micro-texture loss
2. Locally connected layers: apply a filter bank (different location, different local statistics)
3. Fully connected layers: capture correlations between features in distance
4. Softmax Classifier



Siamese Network

- Pass multiple inputs to mulitple networks with the same architecture and parameters
- Enable training for only the two topmost layers (top layer: new logistic unit)
- <u>Similarity function</u>
  - $d(i_1, i_2)$: degree of difference between two images
  - If $d(i_1, i_2) \leq \tau$, then the faces are the same.

3. Further Studies

- Face Recognition Task

1. 빠른 Computation을 요구하는 application이 다수
2. 타 image recognition task와 달리 얼굴 내 미세한 feature difference를 감지할 수 있어야 한다.
3. One-shot learning: 하나의 target에 대한 다양한 이미지셋이 존재하지 않는다. 하지만 feature를 손상시키는 augmentation을 하면 안되므로 training data 수집에 큰 어려움을 겪는다. 위 논문에서는 unconstrained conditions(다양한 포즈, 자세)에서 frontalization을 수행함으로써 양질의 training set을 얻고자 하였다.

- 생각해 볼 내용

1. 논문의 Siamese Network에서는 Similarity function으로 Manhattan distance 사용하였다. 만약 Euclidean이나 Weighted Chi-squared distance를 사용한다면 성능이 어떻게 달라질까
2. Triplet Loss, MoCo 을 이용한 Unsupervised learning도 가능하지 않을까