

Deep features for recognizing disguised faces in the wild

Shuang Sha

June 16, 2018

Abstract

Unconstrained face verification is a challenging problem owing to variations in pose, illumination, resolution of image, age, etc. Because of this reason, in this paper, researchers propose a DCNN-based approach for recognizing people under disguises and picking out impostors. They train two different networks on a large dataset comprising of still images and video frames with L2-softmax loss. They fuse features obtained from the two networks and show that the resulting features are effective for discriminating between disguised faces and impostors in the wild.

1. Related Work

The standard framework of face verification pipeline is shown in Figure 1 and consists of the following steps: face detection, facial landmark detection and alignment, feature representation of a face and metric learning.

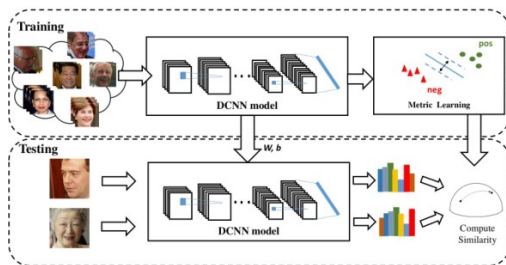


Figure 1. A typical face verification system pipeline. During training, a deep network is trained for classification using a large training dataset. After training the network, a metric learning framework (e.g. triplet embedding) is used to embed the features obtained from the deep CNN into a discriminative subspace. At test time, given two faces, the features from the deep CNN are computed and embedded into the embedding subspace. Finally, a similarity score (e.g. cosine similarity) is calculated between the two embedded features.

1.1. Face Detection

Face detection is the process of localizing all faces present in an image. Typical face detection methods output bounding box coordinates for each face in an image. With the popularity of deep CNNs, several CNN-face detectors have been proposed in recent years [2, 3]. Significant improvements in unconstrained face detection performance have also been supported by the availability of large training datasets e.g. FDDB, WIDER [5]. Both of these datasets contain faces with large variations in pose, illumination, scale, resolution, occlusion etc.

1.2. Landmark Localization

The next step of the framework is fiducial keypoint localization. Such landmarks can include eye centers, nose top, mouth corners, ear lobe tips, chin, etc. These landmarks are used for aligning the detected faces, in other words, to transform a given face into a canonical view of the face without losing the identity information [1].

1.3. Feature Representation

As shown in Figure 1, the next step is to extract features from the aligned face images. Again, deep CNNs are currently the most popular and best performing methods for extracting features. These feature representations are usually learned using large training sets [4]. The requirement for such features is that the features for different images of the same subject should be close (in some metric) and the features for images of different subjects should be far.

1.4. Metric Learning

The fourth step is metric learning, and it is the process of learning a classifier or similarity measure from data and is an important step for enhancing performance of face verification systems.

References

- [1] A. Bansal, R. Ranjan, C. D. Castillo, and R. Chellappa. Deep features for recognizing disguised faces in the wild. In *CVPRW*, pages 10–16, 2018. 1

- [2] D. Chen, G. Hua, F. Wen, and J. Sun. Supervised transformer network for efficient face detection. In *ECCV*, pages 122–138, 2016. [1](#)
- [3] Y. Z. Li, B. Y. Sun, T. F. Wu, and Y. Z. Wang. Face detection with end-to-end integration of a ConvNet and a 3D model. In *ECCV*, pages 420–436, 2016. [1](#)
- [4] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR*, pages 529–534, 2011. [1](#)
- [5] S. Yang, P. Luo, C.-C. Loy, and X. Tang. Wider face: A face detection benchmark. In *CVPR*, pages 5525–5533, 2016. [1](#)