International Journal of Computer Research
Volume 23, Number 2
ISSN: 1535-6698

DEEP FEATURES COMBINED WITH HAND-CRAFTED FEATURES FOR FACE RECOGNITION

Alessandra Lumini[1,*], Loris Nanni[2], and Stefano Ghidoni[2]
[1]DISI, Università di Bologna, Cesena, Italy
[2]DEI, University of Padua, Padua, Italy

ABSTRACT

Most of recent advances in the field of face recognition are related to the use of a convolutional neural network (CNN) and the availability of very large scale training datasets. Unfortunately, large scale public datasets are not available to most of the research community, which therefore can hardly compare with big companies. To overcome this drawback, in this work we suggest to use an already trained CNN and we perform a study in order to evaluate the representation capability of its layers. Most of previous face recognition approaches based on deep learning use a CNN self-trained on a very large training set, taking one on the last intermediate layer as a representation and adding a classification layer trained over a set of known face identities to generalize the recognition capability of the CNN to a set of identities outside the training set. The idea is that the representation capabilities of the last one of two layers of a deep trained CNN is higher than traditional handcrafted features. In this work, starting from a CNN trained for face recognition, we study and compare the representation capability of several different layers in CNNs (not only the last ones) showing that they contain more accurate information about the face image than to believe. The proposed system extracts learned features from different layers of a CNN and uses them as a feature vector for a general purpose classifier. Moreover, we study the independence of the different sets of features used and between learned and handcrafted features, showing that they can be exploited to design an effective ensemble.

The proposed approach gains noticeable performance both in the FERET datasets, with the highest performance rates published in the literature, and the Labeled Faces in the Wild (LFW) dataset where it achieves good results. The MATLAB source of our best ensemble approach will be freely available at https://www.dei.unipd.it/node/2357 “+Pattern Recognition and Ensemble Classifiers”

Keywords: face recognition, similarity metric learning, deep learning, shallow descriptors

[*] Corresponding author: alessandra.lumini@unibo.it

1. INTRODUCTION

Face recognition has been an area of intense research since the 1960s (Zhou et al. 2014), and the great interest about it is justified by the growing number of applications ranging from biometric security to criminal identification, from access management to human machine interaction, from photo album management in social networks to digital entertainment. Many innovative applications making use of this technology are continuously being developed at a rapid pace. Such applications can be categorized into three classes according to the goal of the face recognition task: face verification, where the aim the authentication of an individual to assert his/her identity; face identification, where the aim is to find a correspondence in a database of faces; face tagging, where the aim is to label face images based on identification. The face recognition problem consists in comparing two images of faces and determining whether both images frame the same person or not. The typical face recognition pipeline consists of four steps: face detection, face alignment, feature extraction and classification, where feature extraction is the crucial step. Most conventional face recognition techniques based on hand-crafted features such as Local Binary Patterns (LBP) (Ahonen et al. 2006), Local Phase Quantization (LPQ) (Chan et al. 2013) or Patterns of Oriented Edge Magnitudes (POEM) (Vu 2013)(Nanni et al. 2013) perform well when facial images are captured in optimal (controlled) conditions; but their performance is quickly degraded when facial images are captured in the wild. Unfortunately, faces appearing in most applications, like social networks and digital entertainment, are acquired in uncontrolled conditions: they usually exhibit dramatic pose, expression and illumination variations and often a low image quality. The main difficulty of face identification consists in separating the specific features carrying information on the identity from the huge mass of features expressing other characteristics. It is still an open problem to find an ideal feature set for face recognition, robust under any acquisition setup. In the last few years, a new class of methods has been proposed, based on convolutional neural networks (CNN), and referred to as "*deep methods*" in opposition to "*shallow methods*" which are based on hand-crafted features. Deep methods learn their features during a training phase, and the set of learned features are more robust than handcrafted ones in detecting complex intra-personal variations.

Deep learning is a real breakthrough in the field of face recognition: a CNN model can not only characterize large data variations but also learn a compact and discriminative feature representation that can be generalized to dataset that were not involved in training. Deep learning has a great advantage over shallow methods in both identification and recognition. The deep features learned by trained CNN models are highly discriminative in performing large-scale face identification.

The first precursor paper in this area was proposed in 2005 (Chopra et al. 2005) and employed a convolutional neural network to learn a metric between face images. The deep learning approach was so dominant that after a decade of study, researchers (Taigman et al. 2014) have finally closed the "gap to human-level performance in face verification": DeepFace has achieved 97.25% accuracy on the LFW dataset, which is very close to human level accuracy (97.53%) using an ensemble of CNNs to find a good numerical representation of the face. Afterwards, many other deep learning approaches (Lu and Tang 2014)(Sun et al. 2014b)(Sun et al. 2014a)(Sun et al. 2015) have significantly outperformed previous shallow methods. For example, the approach based on Gaussian Processes and multi-source training sets in (Lu and Tang 2014) has achieved 98.52% accuracy on the LFW dataset, which is better than human performance.

Even if the LFW dataset is the de-facto benchmark for face recognition in-the-wild, some researchers have pointed out (Zhou et al. 2015) some limitations existing between big training set and recognition performance. During the history of LFW benchmark, the largest performance improvements have been gained the last few years by deep learning techniques trained from huge datasets (from ~10000 samples in (Cao et al. 2013) to ~4,000,000 images in (Taigman et al. 2014)). The best performance using a training set of less than 10,000 images with deep learning was lower than 85%. Since Deep Learning approaches require millions of images for training, their results on benchmarks cannot be directly compared with approaches obtained using a testing protocol based on a few training samples.

The approach presented in this paper can be referred to as a "*transfer learning*" method. Unlike *shallow approaches*, it is not based on a representation of the face image by means of handcrafted descriptors only. It is also different from deep methods, since it is not based on a supervised deep neural networks specifically trained for this face recognition problem, i.e., to minimize the distance between features of the same identity while simultaneously decreasing intra-personal variations.

The system presented here is an evolution of the approach in (Lumini et al. 2016) where preliminary results about transfer learning were discussed as a direction for future research. The approach presented in (Lumini et al. 2016) is a *shallow* method based on a combination of handcrafted local image descriptors. The system is based on a combination of different preprocessing techniques and of several handcrafted feature extractors: then, similarly to this work, the classification is performed by an ensemble of classifiers. Moreover, in (Lumini et al. 2016) preliminary results about the combination of "learned" and "handcrafted" features were discussed.

In this paper we further evaluate the idea of performing "*transfer learning*" from an already trained CNN, analyzing the layer of the network which is most suited for face representation. The feature extraction step, obtained by convolving the input face image with a CNN and extracting the response of several different layers, is inserted into a well-tested framework consisting in face detection and cropping, frontalization, feature extraction and classification. The components used in each step have been already tested and tuned in (Lumini et al. 2016) and demonstrated good performance both in the FERET and LFW datasets. In this work we test the proposed system using different set of "learned" features, which have been obtained from the internal representation of a deep method, specifically a Convolutional Neural Network (CNN) trained for the face recognition problem.

The resulting fusion with the handcrafted features proposed in (Lumini et al. 2016) obtains, to the best of our knowledge, the highest mean accuracy ratings on the FERET datasets and very good results on the LFW dataset.

2. The Proposed Approach

The method proposed in this work is an evolution of the approach presented in (Lumini et al. 2016) where learned features are employed instead of handcrafted descriptors. The general schema of the approach is shown in Figure 1 and consists of the following steps:

- *Face detection and crop*: once the precise position of the face image is detected according to the approach in (Hassner et al. 2015a) the resulting face is tight cropped and aligned according to eye position;
- *Frontalization:* recent experiments (Lumini et al. 2016) demonstrated the importance of frontalization for precise face recognition also in presence of pose changes; in this work the approach proposed in (Hassner et al. 2015a) is used to synthesize frontal views of faces from the detected face;
- *Feature extraction:* feature extraction is performed using learned features obtained taking the response to the input face image of one intermediate layer of a CNN. Several experiments are reported to evaluate the best combination of layers;
- *Feature Transformation*: before classification the dimensionality of each descriptor is reduced via Principal Component Analysis (PCA) (Duda et al. 2000);
- *Classification:* a general-purpose classifier is trained on each reduced feature vector. The final decision is then determined according to the sum rule by summing up the scores/similarity values (SIM_i) obtained from each classifier. In this work, the simple angle distance is used in the FERET datasets, where the aim is identification. SML classifier (Cao et al. 2013) is used on the LFW dataset, where the aim is to verify a given match.

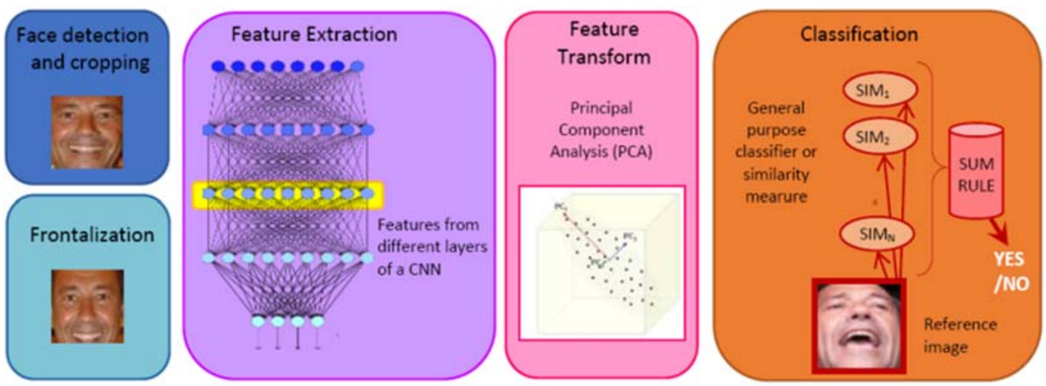

Figure 1. Schema of the proposed face recognition ensemble.

2.1. Hard Frontalization (HF)

Hard Frontalization (HF) (Hassner et al. 2015b) is a technique that uses a unique 3D geometrical shape to obtain a frontal view from a set of face pictures acquired in the wild from different angles. This is in contrast to the other approaches that employ a different 3D shape for each person considered. While the last approach aims at a more accurate reconstruction, the first one assumes that the difference between a "standard" 3D shape and the real one can be neglected for the purpose of frontalization.

Images acquired in the wild are first processed to extract faces. Each face is rescaled to a standard size, and a set of 49 facial features are detected in the sample. Such features are exploited to estimate the 3×4 projection matrix P describing the camera pose under which the face was framed, i.e., the rototranslation between the camera and the framed face. The P

matrix describes the geometrical association between each pixel and the portion of the face it represents. In other words, by knowing P it is possible to understand which part of the face is represented by each pixel found in the image, and it is also possible to project such pixel onto the standard 3D face model considered in the HF algorithm, thus associating a color to each 3D location. The model itself is supposed to have a plane of symmetry, therefore some parts of the model that are not seen are filled with the color of the symmetrical locations; for example, if the left half of the mouth is seen in the image, the right half will be completed exploiting the symmetry.

The HF algorithm generates the frontalized view starting from the textured 3D model described above following a four step process. In the first step, a frontal synthetic view of the model is obtained by projecting the 3D model using a camera matrix whose rotation matrix and translation vector define a frontal projection that is used as a reference coordinate system. The second step generates the frontal pose synthesis by projecting the facial features from the 3D model onto such reference system. Step 3 deals with visibility estimation, which depends on the projection of the reference 3D model onto the given view. Finally, in step 4 the detection problems introduced by conditional soft-symmetry are tackled; this is done based on a standard representation based on LBP (Local Binary Pattern) and an SVM classifier.

2.2. Feature Extraction

In this paper we extract “learned” features from a CNN already trained for face recognition. However, we do not rely only on the data provided by the last layer of the CNN, as it is usually the case. Rather, we consider the information provided by other layers throughout the network (deep layers). The layers of a CNN can be considered as a set of features that are automatically learned during the training phase, and whose characteristics depend on the depth of the layer itself. Those layers that are closer to the input data, process information coming from a small neighborhood of pixels, and extract low-level, local features. Conversely, layers that are far from input data are made of nodes getting input that has passed through several processing steps, and depends on a larger set of pixels: this leads to the conclusion that such layers provide as output high-level, global features. The transition from local to global features can be seen as a gradual process that is the consequence of the peculiar scheme of the connections among nodes in a CNN, that makes it particularly suited for processing 2D data, as it is the case of images.

The features extracted from different stages of a CNN are used in the same way as it usually happens with hand-crafted features: the feature vectors become the input of a classifier – a set of SVMs in our case – that is trained to solve the face classification problem. This structure, composed of a previously trained classifier connected to a second classification stage for changing the problem to be solved, is known as transfer learning. It is particularly convenient when dealing with Deep Neural Networks (DNNs), because it allows to skip the training phase of such networks, which is computationally very intensive and requires a huge number of samples. Instead, the training of a set of SVMs requires smaller datasets and reduced computational effort.

The transfer learning scheme described above has been applied considering different combinations for the layers of the CNNs to be used as features. This enables to investigate the representation capabilities at various depths of a convolutional network, and the dependencies among their information representation. In this study, the CNN presented in (Parkhi et al. 2015) was considered: it is a VGG- Very-Deep-16 CNN architecture whose models, trained on a very large face collection, are freely available for downloading.

2.3. Feature Transform

The features described above represent a high amount of data that could cause the system to fall into the curse of dimensionality. To cope with this problem, dimensionality reduction methods have been applied. The best performance was achieved using PCA, Principal Component Analysis (Duda et al. 2000), which is a common approach. This technique generates a projection of the original space onto a reduced number of directions in order to maximize the variance of the projected vectors. In our experiments, the orthogonal basis used for projecting the features expresses 99% of the input variance. When the classifier is chosen to be an SML, the first 300 components are selected, as suggested in (Lumini et al. 2016).

2.4. Classification

The descriptors previously detailed are processed using a separate distance function or classifier (depending on the problem being addressed) for each feature. Such functions are then combined by sum rule to obtain the final classification output. This technique was selected because it does not require a deep analysis of the uncertainty space of the ensemble classifiers, as it was performed in (Fernández-Martínez and Cernea 2015).

The similarity function chosen for comparing faces in the experiments on identification (run on the FERET datasets) are angle distance. The angle distance α between two vectors v_1 and v_2 is evaluated as:

$$\alpha = \sin^{-1} \frac{v_1 \times v_2}{\|v_1\|\|v_2\|},$$

and represents the size of the angle defined by the two directions defined by the vectors. A different function is used for the experiments aimed at verifying given matches run on the LFW dataset. In this case, a general purpose binary classifier (Similarity Metric Learning or SML (Cao et al. 2013) in our experiments) is used to distinguish between good and bad match. SML is a novel regularization framework proposed for learning similarity metrics for unconstrained face verification. The similarity function between the images x_i, x_j is defined as:

$$f_{M,G}(x_i, x_j) = s_G(x_i, x_j) - d_M(x_i, x_j)$$

where $s_G(x_i, x_j)$ and $d_M(x_i, x_j)$ are a weighed similarity and a weighed distance, respectively. The weight matrices G and M are learned from the training set with the goal of being robust to large intra-personal variations.

3. EXPERIMENTS

3.1. Datasets

The performance of the proposed approach was assessed on the FERET (Phillips et al. 2000) and LFW (Huang et al. 2007) datasets. The FERET dataset was collected in the context of FacE REcognition Technology (FERET) program; it is made of five datasets acquired in different time periods, under different weather conditions – the gallery set Fa (1196 images), and four datasets used for testing:

- Fb: 1195 samples acquired in the same day as Fa, using the same camera and under similar lighting conditions;
- Fc: 194 samples taken in the same day as Fa, but with a different camera and under different lighting conditions;
- Dup1: 722 samples acquired within one year since the acquisition of Fa;
- Dup2: 234 samples acquired more than one year after the acquisition of Fa.

FERET proposes a standard evaluation protocol that requires each test image to be compared against all the images in the gallery set. In our experiments we modified the images by aligning all the faces using the true eye positions and cropping the images to a fixed size of 110×110 pixels. Some samples from the FERET databases are reported in Figure 2.

Some samples from the FERET databases are reported in Figure 2.

Figure 2. Samples from the FERET database.

The LFW (Labeled Faces in the Wild) database addresses the problem of unconstrained face recognition. It is made of more than 13000 internet images of 5749 celebrities, 1680 of which appear in two or more images. It represents a very difficult testbed because images were acquired in a totally uncontrolled way, and there is no control over imaging system, lighting, image quality and appearance of the subjects, since images of the same person at different ages are present. The database is divided into two views: the first one shall be used for training and testing (supports 10-fold validation), while view 2 is meant for benchmarking. Our experiments are designed to follow the testing protocol and dataset subdivision guidelines proposed by the authors of the database.

Figure 3. Samples from the LFW database.

Table 1. Accuracy obtained by our ensemble as a variation of the CNN layers for tight cropped images

Layers		FERET recognition rate				LFW accuracy
Level	**Dimensionality**	**Fb**	**Fc**	**Dup1**	**Dup2**	
30	100352	98.66	**100**	84.21	86.32	92.70
31	100352	99.33	98.97	90.3	89.74	93.07
32	25088	99.16	99.48	90.58	89.74	92.12
33	4096	98.66	98.97	91	**91.88**	92.15
34	4096	98.66	98.97	90.72	**91.88**	93.00
35	4096	98.66	98.97	89.89	**91.88**	92.82
36	4096	98.66	**100**	89.61	91.45	92.88
37	2622	97.49	98.97	87.67	87.18	92.30
[33 34]	8192	**98.74**	99.48	**90.86**	91.45	**93.43**
[36 37] (Lumini et al. 2016)	6718	98.33	99.48	89.06	91.03	93.22

The performance of our algorithms was measured by means of recognition rate for the FERET dataset, and accuracy for the LFW dataset, defined as the ratio between correct classification results (true positives and true negatives) and the total population. Some samples from the LFW database are reported in Figure 3.

Some samples from the LFW database are reported in Figure 3.

3.2. Results

The first experiment is aimed at evaluating the importance of the frontalization step: starting from results published in (Lumini et al. 2016), where the CNN outputs of the 37th and 36th fully-connected layers were used for describing the images, we evaluate the recognition performance on the LFW dataset.

The recognition accuracy reported using frontalization is 93.22, while it drops to 92.82 without frontalization. Therefore, all the following experiments will be carried out maintaining the frontalization step.

The second experiment is aimed at comparing the representation capabilities of several different layers of the VGG-Very-Deep-16 CNN proposed in (Parkhi et al. 2015). In the first column of Table 1 the layers used for representation purposes are reported (the presence of two numbers denotes the concatenation of two layers); the second column is the dimensionality of the feature vector, and the remaining columns report the recognition accuracies. The last row of Table 1 also reports the result of the fusion of the last two layers, already published in (Lumini et al. 2016). The experiments are carried out using the complete approach described in Figure 1 (including the steps of frontalization and feature transformation). The classifiers used in these experiments are SML for LFW and the angle distance for the FERET datasets. The best result is obtained when the CNN outputs of the 33th and 34th layers are used for describing the images.

All the results reported in Table 1 were obtained using a tight cropping, i.e., cropping faces so that the background is minimally involved in classification. For the sake of completeness we also report in Table 2 performance obtained using larger images (in Figure 4 an example of different cropping sizes is reported). Large crop include portions of hairs and clothes that can be considered as "soft biometrics", useful to improve the recognition rate. In order to confirm this hypothesis we also tested the performance obtained by a large crop image where the face was removed (see Figure 2.d) obtaining a surprisingly high accuracy of 75.52%.

Figure 4. Sample of different cropping from the LFW database: (a) original image, (b) tight crop (c) large crop, (d) face removed.

Table 2. Accuracy obtained by our ensemble as a variation of the CNN layers for large cropped images

Layers		**FERET recognition rate**				**LFW accuracy**
Level	**Dimensionality**	**Fb**	**Fc**	**Dup1**	**Dup2**	
30	100352	98.83	98.97	69.81	60.68	93.87
31	100352	99.33	97.94	89.75	85.47	95.53
32	25088	99.67	98.97	91.27	87.61	94.72
33	4096	**100**	**99.48**	94.6	**94.02**	96.03
34	4096	**100**	98.97	94.6	93.59	96.95
35	4096	99.92	98.97	93.21	91.88	96.40
36	4096	99.92	**99.48**	92.94	92.74	**96.93**
37	2622	99.75	**99.48**	90.03	88.46	96.50
[33 34]	8192	**100**	**99.48**	**95.01**	**94.02**	96.85
[36 37] (Lumini et al. 2016)	6718	99.92	**99.48**	92.11	91.88	96.75

Table 3. Comparison among the proposed ensemble with the state-of-the-arts approaches

Methods		FERET recognition rate					LFW accuracy
Reference	Year	Fb	Fc	Dup1	Dup2	Avg	
(Ahonen et al. 2004)	2004	93.0	51.0	61.0	50.0	63.8	---
(Zhang et al. 2005)	2005	94.0	97.0	68.0	58.0	79.2	---
(Deng et al. 2005)	2005	96.3	99.5	78.8	77.8	88.1	---
(Zhang et al. 2007)	2007	97.6	99.0	77.7	76.1	87.6	---
(Tan and Triggs 2007)	2007	98.0	98.0	90.0	85.0	92.8	---
(Xie et al. 2010)	2010	99.0	99.0	94.0	93.0	96.3	---
(Yang et al. 2012)	2012	99.7	99.5	93.6	91.5	96.07	---
(Vu 2013)	2013	99.7	100	94.9	94.0	97.2	86.2
(Nanni et al. 2013)	2013	98.7	100	94.6	93.6	96.7	76.9
(Chai et al. 2014)	2014	99.9	100	95.7	93.1	97.17	---
(Nowak and Jurie 2007)	2007	---	---	---	---	---	73.9
(Wolf et al. 2008)	2008	---	---	---	---	---	78.5
(Pinto et al. 2009)	2009	---	---	---	---	---	79.35
(Li et al. 2013)	2013	---	---	---	---	---	84.08
(Arashloo and Kittler 2013)	2013	---	---	---	---	---	79.08
(Simonyan et al. 2013)	2013	---	---	---	---	---	87.47
(Cao et al. 2013)	2013	---	---	---	---	---	88.5[1]
(Li and Hua 2015)	2015	---	---	---	---	---	88.97
(Arashloo and Kittler 2014)	2015	---	---	---	---	---	95.89
(Li et al. 2015)	2015	---	---	---	---	---	91.10
(Juefei-Xu et al. 2015)	2015	---	---	---	---	---	87.55
(Lumini et al. 2016)	2016	99.2	100	94.6	94.0	97.0	91.7
HERE	-	98.74	99.48	90.86	91.45	95.13	93.43
HERE+(Lumini et al. 2016)	-	99.58	100	97.37	96.15	98.27	93.32
2×HERE+(Lumini et al. 2016)	-	99.58	100	97.78	97.44	98.7	93.65
HERE + (Lumini et al. 2016) + (Cao et al. 2013)	-	---	---	---	---	---	93.97
2 × HERE+(Lumini et al. 2016)+(Cao et al. 2013)	-	---	---	---	---	---	94.08

[1] Obtained using the source code shared by the authors of (Cao et al. 2013) and the testing protocol described in this work (which is slightly different from the one used in (Cao et al. 2013)).

The third experiment, reported in Table 3, is a comparison with the state-of-art for both the FERET and LFW datasets for methods not based on outside training data. In Table 3 the best approach tested in this work is denoted by HERE (i.e., the one reported in last line of Table 1 and related to tiny cropped images and to the layers [33 34]). The last four rows of Table 3 report the weighed fusion of HERE and some of the best shallow methods proposed in the literature. Examining Table 3, it is clear that the system performance has significantly increased in the last few years: the proposed system gains very good performance in both the datasets. The fusion between the "learned features" proposed in this work and the handcrafted features of (Lumini et al. 2016) further improves both approaches obtaining one of the best recognition performance ever published for the FERET databases, and very valuable results in LFW too. When different approaches are combined, before the fusions, their scores are normalized to have zero mean and standard deviation 1. The methods 2 × X + Y means that the methods are combined with weighted sum rule where the weight of X is 2.

Our deep transfer learning approach does not achieve performance comparable with the state of the art of deep learning methods (e.g., see (Zhou et al. 2015)). However, we use the CNN only for extracting the features from the images, and only the standard training set of LFW is used for training SML. Interestingly, in the FERET dataset (high quality frontal images) the hand crafted features work better than the features extracted by CNN, but their fusion nevertheless permits to boost the performance.

CONCLUSION

In this work we studied the representation capability of convolution neural networks using intermediate layers of an already trained CNN for extracting features for the face recognition problem. Our experiments, carried out considering two of the most used benchmark databases in this field, show that not only the last two layers, but also several different internal layers in CNNs contain accurate information about the face image.

The proposed approach gains noticeable performance both in the FERET dataset, with the highest performance rates published in the literature, and the Labeled Faces in the Wild (LFW) dataset, where it achieves good results.

In the LFW dataset the approach proposed here, combined with the method in (Lumini et al. 2016), obtains a 93.65% accuracy, which can be further improved to 94.08% considering also the method in (Cao et al. 2013). The only approach which outperforms our method, without using outside training data, is (Arashloo and Kittler 2014) with 95.89% accuracy, but the authors do not share their source code, therefore results are not easily reproducible.

Another important aspect to be analyzed in face recognition approaches is the dimension of the cropping for the face image: our experiments demonstrate that there is a noticeable performance gap between loosely cropped and tightly cropped images. In this work we observe that even if a tight crop produces a performance drop, it is fairer for pure face recognition, since it allows to base the recognition task only on the face region, discarding possible information in the contours. Unfortunately, results reported in the literature using the LFW or FERET benchmark do not always clearly explain which kind of crop was used, therefore a fair comparison is not possible.

In the future, we plan to experiment CNN not specifically trained for the face recognition task (i.e., object recognition, scene classification, etc.) in order to evaluate the degree of independence of such sets of features and their ability to work with different classification problems.

REFERENCES

Ahonen, T., et al., (2004). Face Recognition with Local Binary Patterns. In *European Conference on Computer Vision*., pp. 469–481.

Ahonen, T., Hadid, A. and Pietikainen, M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(12), pp. 2037–2041.

Arashloo, S. R. and Kittler, J. (2014). Class-Specific Kernel Fusion of Multiple Descriptors for Face Verification Using Multiscale Binarised Statistical Image Features. *Information Forensics and Security, IEEE Transactions on*, 9(12), pp.2100–2109.

Arashloo, S. R. and Kittler, J. (2013). Efficient processing of mrfs for unconstrained-pose face recognition. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*., pp. 1–8.

Cao, Q., Ying, Y. and Li, P. (2013). Similarity Metric Learning for Face Recognition. *Computer Vision (ICCV), 2013 IEEE International Conference on*, pp. 2408–2415. Available at: http://ieeexplore.ieee.org/ lpdocs/epic03/wrapper.htm?arnumber=6751410.

Chai, Z., et al., (2014). Gabor ordinal measures for face recognition. *IEEE Transactions on Information Forensics and Security*, 9(1), pp. 14–26.

Chan, C. H., et al., (2013). Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(5), pp. 1164–1177.

Chopra, S., Hadsell, R. and LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*., pp. 539–546.

Deng, W., Hu, J. and Guo, J. (2005). Gabor-eigen-whiten-cosine: a robust scheme for face recognition. In *Analysis and Modelling of Faces and Gestures*., Springer, pp. 336–349.

Duda, R. O., Hart, P. E. and Stork, D. G. (2000). *Pattern Classification*,

Fernández-Martínez, J. L. and Cernea, A. (2015). Exploring the uncertainty space of ensemble classifiers in face recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, *29*(03), p. 1556002.

Fernández-Martínez, J. L. and Cernea, A. (2014). *Numerical Analysis and Comparison of Spectral Decomposition Methods in Biometric Applications*, Available at: http:// www.worldscientific.com/doi/abs/10. 1142/S0218001414560011.

Hassner, T., et al., (2015a). Effective Face Frontalization in Unconstrained Images. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Hassner, T., et al., (2015b). Effective Face Frontalization in Unconstrained Images. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. Available at: http:// www.openu.ac.il/home/hassner/ projects/ frontalize.

Huang, G., et al., (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *University of Massachusetts Amherst Technical Report 07*, 49(07-49), pp. 1–11. Available at: http:// citeseerx.ist.psu.edu/viewdoc/download?

doi= 10.1.1. 122.8268&rep=rep1&type=pdf\nhttps://www.cs.umass.edu/~elm/papers/lfw.pdf.

Juefei-Xu, F., Luu, K. and Savvides, M. (2015). Spartans: Single-sample Periocular-based Alignment-robust Recognition Technique Applied to Non-frontal Scenarios. *Image Processing*, *24*(12).

Li, H., et al., (2015). Eigen-pep for video face recognition. In *Computer Vision--ACCV 2014*., Springer, pp. 17–33.

Li, H., et al., (2013). Probabilistic elastic matching for pose variant face verification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3499–3506.

Li, H. and Hua, G. (2015). Hierarchical-PEP Model for Real-world Face Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*., pp. 4055–4064.

Lu, C. and Tang, X. (2014). Surpassing Human-Level Face Verification Performance on LFW with GaussianFace. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*., pp. 3811–3819. Available at: http://arxiv.org/abs/1404.3840.

Lumini, A., Nanni, L. and Braham, S. (2016). Ensemble of texture descriptors and classifiers for face recognition. Applied Computing and Informatics, to appear 2016 doi: 10.1016/j.aci.2016.04.001.

Nanni, L., et al., (2013). Ensemble of Patterns of Oriented Edge Magnitudes Descriptors For Face Recognition. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV).*

Nowak, E. and Jurie, F. (2007). Learning visual similarity measures for comparing never seen objects. In *IEEE Conference on Computer Vision and Pattern Recognition*., pp. 1–8.

Parkhi, O. M., et al., (2015). Deep face recognition. *Proceedings of the British Machine Vision*.

Phillips, P. J., et al., (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10), pp. 1090–1104. Available at: http:// ieeexplore. ieee.org/lpdocs/epic03/wrapper.htm?arnumber=879790.

Pinto, N., DiCarlo, J. J. and Cox, D. D. (2009). How far can you get with a modern face recognition test set using only simple features? *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, pp. 2591–2598.

Simonyan, K., et al., (2013). Fisher Vector Faces in the Wild. In *Procedings of the British Machine Vision Conference 2013*. pp. 8.1–8.11. Available at: http://www.bmva.org/bmvc/2013/Papers/paper0008/index.html.

Sun, Y., Wang, X. and Tang, X. (2014a). Deep Learning Face Representation by Joint Identification-Verification. In NIPS. pp. 1–9. Available at: http://arxiv.org/abs/1406.4773.

Sun, Y., Wang, X. and Tang, X. (2014b). Deep Learning Face Representation from Predicting 10,000 Classes. In CVPR. pp. 1–9. Available at: http:// arxiv.org/abs/1406.4773.

Sun, Y., Wang, X. and Tang, X. (2015). Deeply learned face representations are sparse, selective, and robust. Cvpr.

Taigman, Y., et al., (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. *Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 8. Available at: http://www.cs.tau.ac.il/~ wolf/papers/deepface_11_01_2013.pdf.

Tan, X. and Triggs, W. (2007). Fusing Gabor and LBP Feature Sets for Kernel-Based Face Recognition. *Analysis and Modeling of Faces and Gestures*, 4778, pp. 235–249. Available at: http://eprints.pascal-network.org/ archive/ 00003663/.

Vu, N. S. (2013). Exploring patterns of gradient orientations and magnitudes for face recognition. *IEEE Transactions on Information Forensics and Security*, *8*(2), pp. 295–304.

Wolf, L., Hassner, T. and Taigman, Y. (2008). Descriptor based methods in the wild. *Science*, 6, pp. 1–14. Available at: http://hal.inria.fr/inria-00326729/\nhttp://citeseerx.ist.psu.edu /viewdoc/download?doi=10.1.1.150.2875&rep=rep1&type=pdf.

Xie, S., et al., (2010). Fusing local patterns of gabor magnitude and phase for face recognition. *IEEE Transactions on Image Processing*, *19*(5), pp. 1349–1361.

Yang, M., et al., (2012). Monogenic Binary Coding: An Efficient Local Feature Extraction Approach to Face Recognition. *Information Forensics and Security, IEEE Transactions on*, *7*(6), pp. 1738–1751.

Zhang, B., et al., (2007). Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition. *IEEE Transactions on Image Processing*, *16*(1), pp. 57–68.

Zhang, W., et al., (2005). Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition. *Proceedings of the IEEE International Conference on Computer Vision*, *I*, pp. 786–791.

Zhou, E., Cao, Z. and Yin, Q. (2015). Naive-Deep Face Recognition: Touching the Limit of LFW Benchmark or Not? *CoRR*, abs/1501.0. Available at: http://arxiv.org/abs /1501.04690.

Zhou, H., et al., (2014). Recent advances on singlemodal and multimodal face recognition: A survey. *IEEE Transactions on Human-Machine Systems*, *44*(6), pp. 701–716.