Age Group Recognition from Face Images using a Fusion of CNN- and COSFIRE-based Features

Demajo, Lara Marie; Guillaumier, Kristian; Azzopardi, George

Published in:
APPIS '19: Proceedings of the 2nd International Conference on Applications of Intelligent Systems

DOI:
10.1145/3309772.3309784

Publication date:
2019

Age Group Recognition from Face Images using a Fusion of CNN- and COSFIRE-based Features

Lara Marie Demajo
University of Malta
lara.m.demajo.15@um.edu.mt

Kristian Guillaumier
University of Malta
kristian.guillaumier@um.edu.mt

George Azzopardi
University of Groningen
g.azzopardi@rug.nl

ABSTRACT

Automatic age group classification is the ability of an algorithm to classify face images into predetermined age groups. It is an important task due to its numerous applications such as monitoring, biometrics and commercial profiling. In this work we propose a fusion technique that combines CNN- and COSFIRE-based features for the recognition of age groups from face images. Both CNN and COSFIRE are trainable approaches that have been demonstrated to be effective in various computer vision applications. As to CNN, we use the pre-trained VGG-Face architecture and for COSFIRE we configure new COSFIRE filters from training data. Since recent literature suggests that CNNs deliver the highest accuracy rates within such problems, the hypothesis which we want to investigate in this work is whether combining CNN and COSFIRE approaches together will improve results. The proposed fusion technique using stacked Support Vector Machine (SVM) classifiers, and trained and tested with the FERET data set images has shown that, indeed, CNN- and COSFIRE-based features are complimentary as their combination reduces the error rate by more than 25%.

KEYWORDS

Age classification, feature extraction, COSFIRE, VGG-Face, face images, CNN, trainable filters, FERET, FG-NET

ACM Reference Format:
Lara Marie Demajo, Kristian Guillaumier, and George Azzopardi. 2019. Age Group Recognition from Face Images using a Fusion of CNN- and COSFIRE-based Features. In *2nd International Conference on Applications of Intelligent Systems (APPIS 2019), January 7–9, 2019, Las Palmas de Gran Canaria, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3309772.3309784

1 INTRODUCTION

Age estimation from images is the attempt to determine the age of an individual through visual information. According to a crowd-sourced experiment in [22], the Mean Absolute Error (MAE) of human age estimations from visual appearance is 4.7 years, concluding that age estimation is quite challenging also for humans [25]. Automatic age classification can be a strong benefit to database search problems such as person identification systems. Forensic art, monitoring and biometrics [30] are other major fields that advantageously utilize age estimation and offer superior safety amongst us like ensuring that younger children have no access to prohibited internet pages, and that vending machines refuse to sell alcohol and cigarettes to people underage [21]. The marketing sector can also benefit and enhance its success rate by showing viewers more relevant commercials. Moreover, through automatic age recognition, devices can upgrade their user interface intelligence and adapt better to the user.

Automatic age group classification from face images is possible through the identification of significant facial changes such as craniofacial growth and skin deformation. For this project, the used images are annotated with the genuine age of the person. The age groups considered are 0-3, 4-7, 8-13, 14-22, 23-35, 36-47, 48-59 and 60+. These non-overlapping age groups start with smaller ranges within the initial age groups and increase over the older age groups since younger ages experience greater changes [22]. Low image quality, facial expressions and facial poses are all challenges for the task of age estimation [17]. Moreover, internal factors such as genetics, gender and race cause people to age differently whereas external factors such as the presence of facial hair or glasses, and cosmetic treatments may hide the real age of a person [16].

In this work, we propose a fusion technique of features constructed from the output of a Convolutional Neural Network (CNN) and that of Combination of Shifted Filter Responses (COSFIRE) for age classification from face images. Both CNN and COSFIRE [6] approaches are trainable in that they learn feature detectors from training data. In particular, we use the VGG-Face pretrained CNN [32, 37] and for COSFIRE we configure a new set of feature detectors. We fuse both methods with a stacked SVM classifier and apply the proposed approach on the benchmark FERET and FG-NET data sets. In this work we deal with cropped face images and we assume that no person in the images has undergone cosmetic treatments.

The rest of the paper is organized as follows. Section 2 includes an account of state-of-the-art methods for age classification. Section 3 provides a detailed explanation of the proposed technique. Section 4 contains the experiments carried out to evaluate the method, followed by a discussion in Section 5 and conclusions in Section 6.

2 RELATED WORK

Existing methods for age recognition via face images are categorized in six main approaches [24], which we list below. The first five are based on hand-crafted feature descriptors and are usually followed by an age prediction method such as classification, regression[1] or hybrid, whilst the last approach is a trainable one that learns features automatically from the training data.

[1]Refer to [28] for detailed information on regression.

APPIS 2019, January 7–9, 2019, Las Palmas de Gran Canaria, Spain

ACM ISBN 978-1-4503-6085-2/19/01…$15.00
https://doi.org/10.1145/3309772.3309784

Anthropometric models: These models use the geometric relationships between facial components. They describe the craniofacial growth but not the skin aging [26].

Active appearance models: This seminal approach encodes object shapes and appearances for template matching. They are well suited for specific age estimation [15, 23].

AGing PattErn Subspace (AGES): This method constructs a subspace using a sequence of an individual's face images ordered in time. It predicts the age through the subspace projection that best reconstructs the query image [19].

Aging manifold learning: This represents face images as low-dimensional manifolds that capture data distribution and geometric structure important for age prediction [18].

Appearance models: This method uses image filtering and a local coding schema in order to extract facial features that model both the face shape and skin information [23].

Yang and Ai [39] use an appearance model called LBP to summarize a given local texture patch by an LBP histogram (LBPH). Chi-square distance is used to find similarity between the LBPH and the optimal reference histogram, and binary classification through a binary tree is performed using AdaBoost. Similarly, Gunay and Nabiyev [20] extract spatial histograms of LBP from face regions and concatenate them into an image descriptor. The Minimum Distance, Nearest Neighbour and k-NN classifiers are used to predict the final age class. Local Matched Filter Binary Pattern is another appearance model proposed by Ouloul et al.[29], which detects face areas containing wrinkles and then extracts the necessary data to construct a histogram, later used to train a hybrid classifier. An SVM first divides the facial features into distinct age groups and then an SVR is used for specific age estimation. Iqbal et al. [24] proposed Directional Age-Primitive Pattern (DAPP), which is an improved configuration of LBP. DAPP characterizes aging information through histograms but avoids patches that have no contribution to aging to escape the accumulation of unnecessary bins in the feature-histogram. Classification is done using an SVM with RBF kernel function.

Deep learning: This approach learns deep neural networks to generate problem-specific models [36]. Recently, more work has been published showing the use of CNNs in age recognition, following the deep learning approach. A deep learning algorithm determines features from images without any hand-crafting and classifies them using class scores.

The method D2C, proposed by Li et al. [27], stands for deep cumulative and comparative learning. The CNN employed in that work is an AlexNet but includes a cumulative hidden layer and a comparative ranking layer. The cumulative layer allows the model to learn from the faces with adjacent ages, whereas the comparative layer performs pair-wise comparative operations, specifying who is older when given two faces. Belver et al. [12] and Anand et al. [1] conducted other studies, further inspiring this project. Both studies analyze the performance of features obtained from different pretrained deep networks. Both cases bypass any fine-tuning, treating CNNs as generic feature extractors. Both works concluded that pretrained CNNs achieve better performance on age estimation when used as feature extractors rather than end-to-end. Similarly, the face recognition pretrained VGG-Face deepnet was used by Qawaqneh et al. [35] in their age group estimation study, showing good results. The final three fully-connected layers were swapped with four new ones but nothing was changed in the other layers.

COSFIRE is a trainable filter approach whose selectivity is determined from a specified pattern of interest. COSFIRE filters are nonlinear, in that they only respond when all parts of the preferred local patterns are present. Conceptually, COSFIRE filters share some architectural properties with CNNs. In their basic form COSFIRE filters have two-layer architectures where they are characterized by a convolutional first-layer, followed by a rectification linear unit (ReLU) and the second layer combines certain responses from the convolutional response maps by a highly nonlinear function. In [9], it was also shown that this basic architecture can be extended to a multi-layer hierarchical approach as well. COSFIRE filters have been used in many applications such as vascular bifurcations detection [5], traffic sign detection and recognition, shape description [7], contour detection [4, 10], delineation of vessel-like structures [11], and gender recognition [3]. In [2], COSFIRE is used along with SURF descriptors, achieving very good results on gender recognition. Therefore, this motivated us to use COSFIRE to solve age group recognition, but this time in combination with a CNN, which is also a trainable computer vision approach.

Hence, the aim of this work is to investigate the complementarity of CNN- and COSFIRE-based features for the application at hand.

3 METHODOLOGY

Figure 1 illustrates a high-level schematic overview of the proposed system. It starts off by passing the labelled images through a preprocessing phase. The preprocessed images are then used by the feature extractors to obtain feature vectors, which are used to train and test the stacked SVM classifier for the prediction of age groups.

3.1 Pre-processing

Each image is preprocessed through face detection and rescaling. The Viola-Jones algorithm[2] [38] was chosen due to its popularity and significant precision. The implementation proposed in [3] is used for face localization and the image is then cropped to keep only the detected face. The image is then resized to a square so that each image attains the same dimensions. If no face is detected by the Viola-Jones algorithm, the query image is ignored. This leaves a resulting set of 12,743 equally sized images with the face as the focal point. The resulting set of face images is subdivided into two groups in a stratified manner; 70% (8,921) training images and 30% (3,822) test images.

3.2 Feature extraction

3.2.1 ***Using VGG-Face.*** We use MatConvNet's pretrained VGG-Face[3], which provides all weight values of the CNN, making it readily available to use without any need of retraining, signifying less computational costs and no overfitting risks. Before inputting the images into the CNN, each image is resized and normalized, by subtracting the mean of the training images, to become a 224×224×3 dimensional image, as required by VGG-Face. Each image is then fed to the neural network for its transformation through a series of layers until evolving into a single dimension feature vector. The

[2]Refer to [40] for an introduction on the Viola-Jones algorithm.
[3]MAT-file & scripts at https://tinyurl.com/y7w5uaal [32].

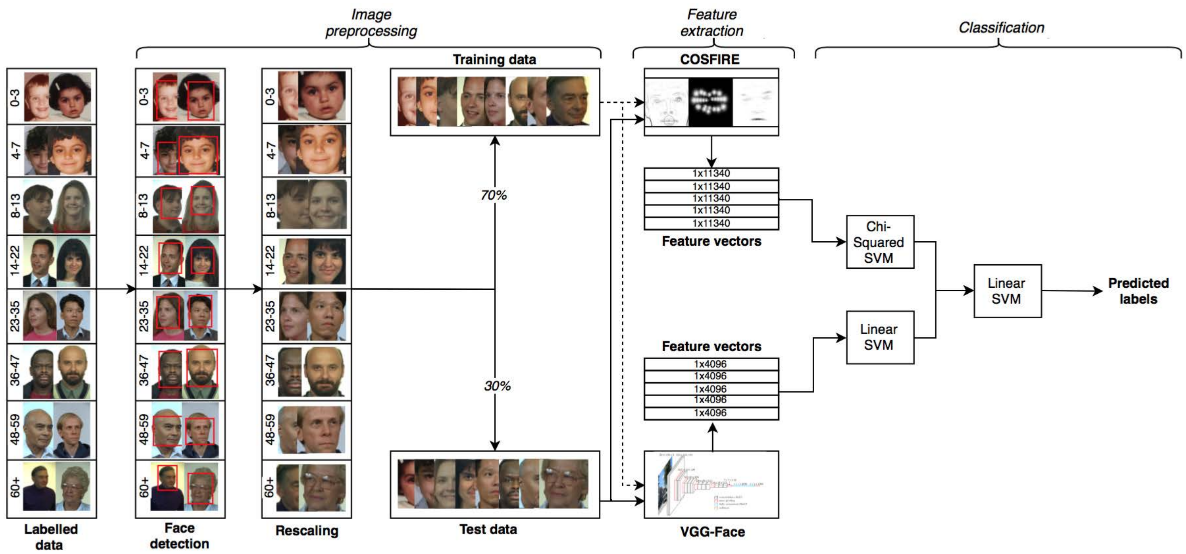

Figure 1: Pipeline of the proposed system.

image is passed through a number of convolutional layers that make use of 3×3 filters and rectification non-linearity (ReLU), and max-pooling layers to get transformed into $7 \times 7 \times 512$. The final three fully-connected layers compute the class scores, where the first two layers have 4096 elements each whilst the last contains 2622 elements. Through transfer learning [31] we ignore the final layer and take the penultimate layer. This layer outputs a 4096-element vector, which we use as the descriptor of a given face image.

3.2.2 ***Using COSFIRE.*** For feature extraction using the COSFIRE approach, we use the implementation of Azzopardi et al.[4] [3, 8] that was applied to gender recognition. Despite gender recognition being very similar to age group classification, the COSFIRE application still required some modifications to adapt to our 8-group age classification problem. Where applicable, most of the parameters were kept the same.

Before COSFIRE can be used as a feature extractor, the appropriate filters must be configured. This is done by using some randomly selected training images, where for each image, a random point of interest is automatically chosen and the local pattern around it is used as a prototype from which we extract features to configure a COSFIRE filter. In order to have sufficient diversity we take equal number of n training images for each age group (we only consider 6 age groups when using FERET since it has no images for the first two groups). This results in the configuration of $6n$ COSFIRE filters.

For the feature extraction, we apply all $6n$ COSFIRE filters to each image. Therefore, the strongest COSFIRE responses for a filter are achieved where the local patterns are similar to the prototype that was used for its configuration. As seen in Figure 2, a spatial pyramid is implemented, where level 0 considers the response map as one tile, level 1 uses 2×2 tiling and level 2 uses a grid of 4×4. We form a feature descriptor by taking the maximum values of all COSFIRE filters in each of the 21 $(1 + (2 \times 2) + (4 \times 4))$ tiles. We normalize to unit length the maximum COSFIRE filter responses for each tile. Using this descriptor a given image is represented by $21(6n)$ elements where n is the number of configured COSFIRE filters per age category.

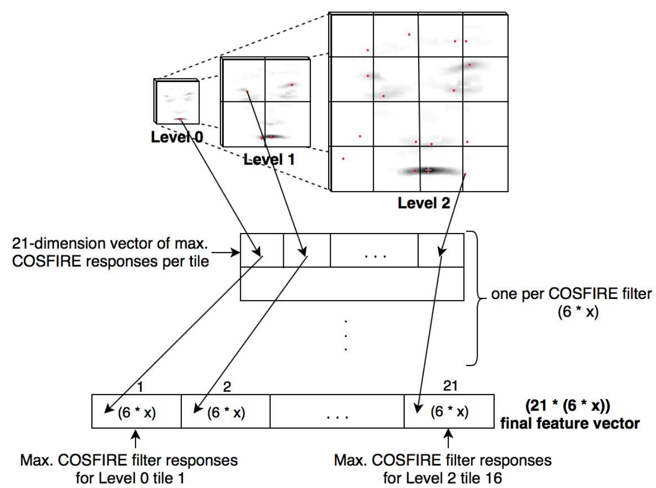

Figure 2: Forming a COSFIRE feature vector using a 3-level spatial pyramid that gives $(1+(2\times2)+(4\times4) =)$ 21-dimensional vectors for each COSFIRE filter. These are then concatenated forming a $(21 \times (6n))$ dimensional feature vector given 6 age groups and n being the number of COSFIRE filters configured per age category.

[4]Scripts for gender recognition with COSFIRE filters is available at https://tinyurl.com/kj54rx4 .

3.3 Stacked Classification

We use a stacked classification approach, motivated by [14], to fuse the CNN- and COSFIRE-based features. For the 4096-element CNN-based vectors we use a one-versus-one SVM[5] with a linear kernel and for the $21(6n)$-element COSFIRE-based vectors we use a one-versus-one SVM with the following chi-squared kernel k:

$$k(x, y) = 1 - \sum_i \gamma \, \frac{(x[i] - y[i])^2}{\frac{1}{2}(x[i] + y[i])} \quad (1)$$

where γ represents the weight[6] that is applied to the values in the feature vector according to the spatial pyramid level they are coming from, making use of spatial information encoded within the descriptor.

For the six age categories, each of the two one-versus-one SVM models give an output vector of 15 ($\frac{6 \times 5}{2}$) values. We concatenate these two sets of 15 values to form a new descriptor of 30 values and use the resulting training vectors to learn a stacked SVM classifier with a linear kernel. The predicted label vectors are then compared to the ground truth labels to generate the performance reports.

4 EVALUATION

4.1 Data sets

Considering that both our feature extractors require supervised learning, the FERET data set[7] of facial images is used as the primary data set. It contains images of individuals of different race and gender, aged between 10 and 80 years old, including variations in terms of pose, lighting and expression, and the possible presence of facial hair and eye glasses. For this project, we consider a total of 23,126 images. The data set also provides the subject and image metadata. Since FERET does not include any images of individuals aged between 0-9, the first two age groups are not used when using this data set alone. Besides FERET, the FG-NET Aging Database[8] is also used. It contains 1,002 images of different multi-race individuals aged between 0 and 69, hence including all defined eight age groups.

The ground truth age of the individual within each image is determined by subtracting the subject's date of birth from the date of capture of the image. The age is then quantized into the eight categories mentioned above.

4.2 Method of Evaluation

As a norm to system testing, the used data set was split in a stratified manner into a training subset (70% of the data) and a test subset (30%). Given that we are dealing with a sufficiently large data set, it is highly likely that the resulting training and test subsets are good representation of the population. With regards to the classifier evaluation, the common approach of computing the True Positives (TP), False Positives (FP) and False Negatives (FN) from the predicted age groups is used. We use these quantities to compute the accuracy, precision, recall and F-measure, Eq. 2-5.

$$Accuracy = \frac{total\ TP}{total\ number\ of\ images} \quad (2)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (3)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (4)$$

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Precision measures the classifier's exactness whilst recall measures the classifier's completeness. F-measure is the harmonic mean of precision and recall. Apart from the macro average recall, precision and F-measure, and micro average accuracy, the system is further evaluated using a confusion matrix, presenting both the class distribution within the data and the classification break down.

5 EXPERIMENTS

Each experiment below was carried out three times in order to compensate for the random stratified split of training and test images, and finally we report the average performance. Note that for each experiment the same images were used for both methods to ensure fair comparisons.

5.0.1 ***Image Rescaling.*** Here, we evaluate empirically both CNN-based and COSFIRE-based methods with the resolution of the input images, and it turns out that for both methods the best performance is achieved when we rescale the images to 128×128 pixels.

5.0.2 ***CNN Transfer Learning.*** Through transfer learning, some of the final fully-connected layers of VGG-Face are ignored. In this experiment, we investigate which of the 4096-dimensional fully-connected layer gives the best feature vectors. This experiment was carried out using Bootstrap Aggregating (Bagging) with 5 bags of 300 samples per class during training, and a linear SVM for classification. The 15th layer (the penultimate layer) of VGG-Face gives the highest recall and accuracy, Table 1. This confirms that this layer generates feature vectors that contain more information than the 14th layer.

Table 1: Results based on different CNN output layers.

CNN final layer	Mean Macro Recall	Global Accuracy
14th	*0.908*	*0.857*
15th	*0.921*	*0.877*

5.0.3 ***COSFIRE Filter Configuration.*** The length of the COSFIRE-based descriptor depends on the number of COSFIRE filters used. The best results are achieved when we configure 90 COSFIRE filters per age group meaning that for six categories, the spatial pyramid descriptors results in 11,340-element vectors ($21 \times 6 \times 90$). Performance decreases when increasing or decreasing the number of COSFIRE filters.

5.0.4 ***SVM Kernel.*** For this experiment, we compare the results of both methods obtained with SVMs using linear kernels and SVMs using Chi-squared kernels. Bagging was not used. The Chi-squared kernel has a major impact on the COSFIRE-based descriptors, as it makes use of the spatial information embodied in its decriptors, Table 2. The lack of such information within VGG-Face's descriptors

[5]SVM scripts imported from LiBSVM library [13].
[6]$\gamma = 1$ for level-2 pyramid responses, $\gamma = 0.5$ for level-1 pyramid responses and $\gamma = 0.25$ for level-0 pyramid responses.
[7]Available at https://tinyurl.com/ycr4ueyr [33, 34].
[8]Acquired by request to fgnet.aging@gmail.com.

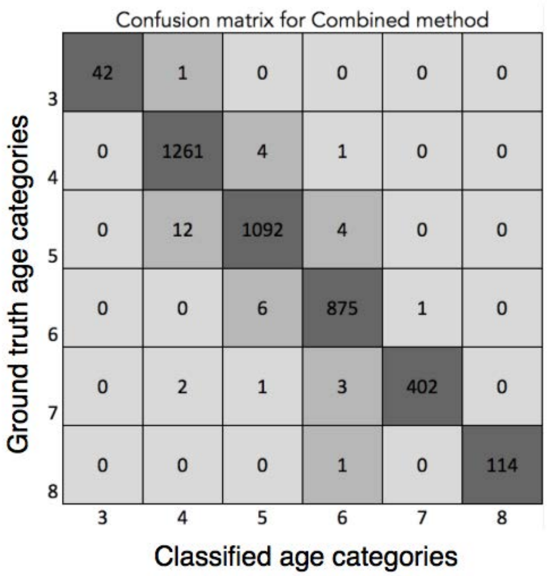

Figure 3: Confusion matrix for the fusion technique.

explains the very small improvement made. Infact, for VGG-Face, the weighting of the Chi-squared kernel was not applied.

Table 2: Results of different SVM kernels.

	Mean Macro Recall		Global Accuracy	
	VGG-Face	**COSFIRE**	**VGG-Face**	**COSFIRE**
Linear	*0.971*	*0.862*	*0.972*	*0.886*
Chi-Squared	*0.976*	*0.900*	*0.975*	*0.922*

5.0.5 ***Fusing Methods***. We take the best parameter values obtained from the above experiments and rerun the system by fusing the descriptors with a stacked classification approach, as explained in Section 3.3. Table 3 reports the performance results of the proposed fusion method compared to the standalone VGG-Face and COSFIRE pipelines. This experiment was carried out five times and the averages were taken as final values. The mean macro precision and F-score, along with the global accuracy of the fusion method surpass those of both individual methods. Figure 3 depicts the confusion matrix of the proposed fusion method and notably is the fact that most of the errors are due to misclassifications in the adjacent age categories. This is reasonable as many people with ages close to class boundaries can also be easily misclassified by humans.

Table 3: Results of the proposed fusion approach on the FERET data set.

	VGG-Face	**COSFIRE**	**Fusion method**
Mean Macro Recall	*0.971*	*0.900*	*0.966 (0.02*)*
Mean Macro Precision	*0.978*	*0.936*	*0.986 (0.005*)*
Mean Macro F-Score	*0.974*	*0.918*	*0.977 (0.01*)*
Global Accuracy	*0.972*	*0.922*	*0.978 (0.008*)*

* Standard deviation of the results obtained from five experiments.

5.1 Using FG-NET

We further analyzed the proposed fusion method by merging the images of FERET to those of FG-NET, which include all the eight age groups in order to create a bigger and a more diverse data set. We use the same preprocessing and feature extraction steps, as mentioned above. The 4096-dimension VGG-Face and $(21 \times 8 \times 90)$ 15,120-dimension COSFIRE descriptors of all images are fed into the stacked SVM classification model. Table 4 demonstrates that the mean macro recall, precision and F-measure are much lower than the global micro accuracy. This is because the first three age groups only contain a few images, leading to a standard deviation of 0.25 between the age group recalls. Hence, whilst the first three metrics are affected by this variation, the global accuracy is not.

Table 4: Results of the fusion method for the combined FG-NET and FERET data sets.

	Fusion method
Mean Macro Recall	*0.807*
Mean Macro Precision	*0.871*
Mean Macro F-Score	*0.838*
Global Accuracy	*0.932*

5.2 Discussion

The results obtained by the above experiments conclude that the best performance of each method is achieved when using an image resize scale of 128×128 pixels, an SVM with a chi-squared kernel for the COSFIRE-based descriptors and linear SVMs for the CNN-based descriptors and the stacked classifier, and using the penultimate layer for the output of VGG-Face whilst configuring 90 filters per age category for COSFIRE. Using these parameter values, VGG-Face achieves an 97.5% accuracy with a recall of 97.6% whereas COSFIRE achieves an 92.2% accuracy with a recall of 90%.

Table 5 shows the result of the proposed combined VGG-Face and COSFIRE method as opposed to some of the existing age classification methods in the literature. Our method achieves a much higher accuracy, proving its applicability to age classification.

As shown in Table 3, we can conclude that VGG-Face is a better feature extractor than COSFIRE in age classification. This is probably because VGG-Face was pretrained on a very large and diverse data set of face images, resulting in more effective features. On the other hand, COSFIRE configures filters on the data set at hand by selecting random local patterns from training images. The improvement in the results, although minor, demonstrates that COSFIRE may provide complementary features to the CNN-based descriptor. In future work, we will configure COSFIRE filters from bigger data sets in order to allow for more variability. Moreover, we will investigate a ranking approach to use only the most effective COSFIRE filters and discard the rest.

Further future work may include the investigation of replacing the linear filters of CNNs with non-linear COSFIRE filters. We expect that one of the benefits of such a network would be robustness to adversial attacks, something which current CNNs tend to suffer from.

Table 5: Comparison of our results with those of published methods.

Method	Dataset	Age Groups	Accuracy	Reference
DAPP	ADIENCE	8	63.3%	[24]
LBPH + Real AdaBoost	FERET	3	92.1%	[39]
Spatial LBP Histograms	FERET	6	80%	[20]
Pretrained VGG-Face + new layers	ADIENCE	8	59.9%	[35]
Proposed combined method	FERET	8	**97.8%**	Ours
	FERET + FG-NET	8	**93.2%**	Ours

6 CONCLUSIONS

We showed that both VGG-Face and COSFIRE serve as good feature extractors for age classification even when using challenging images. While the method that is only based on VGG-Face descriptors performs better than the standalone COSFIRE-based approach, the proposed fusion method that combines both descriptors achieves the highest accuracy rates, outperforming those of existing methods in the literature.

REFERENCES

[1] Abhinav Anand, R Donida Labati, Angelo Genovese, E Munoz Ballester, Vincenzo Piuri, and Fabio Scotti. 2017. Age Estimation Based on Face Images and Pre-trained Convolutional Neural Networks. In *CISDA*. IEEE.

[2] George Azzopardi, Antonio Greco, Alessia Saggese, and Mario Vento. 2018. Fusion of domain-specific and trainable features for gender recognition from face images. *IEEE Access* 6 (2018), 24171–24183.

[3] George Azzopardi, Antonio Greco, and Mario Vento. 2016. Gender recognition from face images with trainable COSFIRE filters. In *Advanced Video and Signal Based Surveillance, 2016 13th IEEE International Conference on*. IEEE, 235–241.

[4] George Azzopardi and Nicolai Petkov. 2012. A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model. *Biological cybernetics* 106, 3 (2012), 177–189.

[5] George Azzopardi and Nicolai Petkov. 2013. Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters. *Pattern Recognition Letters* 34, 8 (2013), 922–933.

[6] George Azzopardi and Nicolai Petkov. 2013. COSFIRE: A Brain-Inspired Approach to Visual Pattern Recognition. In *International Workshop on Brain-Inspired Computing*. Springer, 76–87.

[7] George Azzopardi and Nicolai Petkov. 2013. A shape descriptor based on trainable COSFIRE filters for the recognition of handwritten digits. In *International Conference on Computer Analysis of Images and Patterns*. Springer, 9–16.

[8] George Azzopardi and Nicolai Petkov. 2013. Trainable COSFIRE filters for keypoint detection and pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 2 (2013), 490–503.

[9] George Azzopardi and Nicolai Petkov. 2014. Ventral-stream-like shape representation: from pixel intensity values to trainable object-selective COSFIRE models. *Frontiers in computational neuroscience* 8 (2014), 80.

[10] George Azzopardi, Antonio Rodríguez-Sánchez, Justus Piater, and Nicolai Petkov. 2014. A push-pull CORF model of a simple cell with antiphase inhibition improves SNR and contour detection. *PLoS One* 9, 7 (2014), e98424.

[11] George Azzopardi, Nicola Strisciuglio, Mario Vento, and Nicolai Petkov. 2015. Trainable COSFIRE filters for vessel delineation with application to retinal images. *Medical image analysis* 19, 1 (2015), 46–57.

[12] C Belver, I Arganda-Carreras, and Fadi Dornaika. 2016. Comparative Study of Human Age Estimation Based on Hand-Crafted and Deep Face Features. In *International Workshop on Face and Facial Expression Recognition from Real World Videos*. Springer, 98–112.

[13] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2 (2011), 27:1–27:27. Issue 3. Software available at http://www.csie.ntu.edu.tw/čjlin/libsvm.

[14] Jin Chen, Cheng Wang, and Runsheng Wang. 2009. Using stacked generalization to combine SVMs in magnitude and shape feature spaces for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 47, 7 (2009), 2193–2205.

[15] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. 2001. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence* 23, 6 (2001), 681–685.

[16] A Deepa and T Sasipraba. 2017. Age estimation in facial images using histogram equalization. In *Advanced Computing (ICoAC), 2016 Eighth International Conference on*. IEEE, 186–190.

[17] Eran Eidinger, Roee Enbar, and Tal Hassner. 2014. Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security* 9, 12 (2014), 2170–2179.

[18] Yun Fu, Ye Xu, and Thomas S Huang. 2007. Estimating human age by manifold analysis of face pictures and regression on aging features. In *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 1383–1386.

[19] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. 2007. Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence* 29, 12 (2007), 2234–2240.

[20] Asuman Gunay and Vasif V Nabiyev. 2008. Automatic age classification with LBP. In *Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on*. IEEE, 1–4.

[21] Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang. 2008. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing* 17, 7 (2008), 1178–1188.

[22] Hu Han, Charles Otto, and Anil K Jain. 2013. Age estimation from face images: Human vs. machine performance. In *Biometrics (ICB), 2013 International Conference on*. IEEE, 1–8.

[23] Md Tauhid Bin Iqbal, Byungyong Ryu, Gihun Song, and Oksam Chae. 2016. Positional Ternary Pattern (PTP): An edge based image descriptor for human age recognition. In *Consumer Electronics (ICCE), 2016 IEEE International Conference on*. IEEE, 289–292.

[24] Md Tauhid Bin Iqbal, Mohammad Shoyaib, Byungyong Ryu, M Abdullah-Al-Wadud, and Oksam Chae. 2017. Directional Age-Primitive Pattern (DAPP) for Human Age Group Recognition and Age Estimation. *IEEE Transactions on Information Forensics and Security* (2017).

[25] Manavpreet Kaur, Rakesh K Garg, and Sanjeev Singla. 2015. Analysis of facial soft tissue changes with aging and their effects on facial morphology: A forensic perspective. *Egyptian Journal of Forensic Sciences* 5, 2 (2015), 46–56.

[26] Young Ho Kwon et al. 1994. Age classification from facial images. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*. IEEE, 762–767.

[27] Kai Li, Junliang Xing, Weiming Hu, and Stephen J Maybank. 2017. D2C: Deep cumulatively and comparatively learning for human age estimation. *Pattern Recognition* 66 (2017), 95–105.

[28] Peter Meer, Doron Mintz, Azriel Rosenfeld, and Dong Yoon Kim. 1991. Robust regression methods for computer vision: A review. *International journal of computer vision* 6, 1 (1991), 59–70.

[29] Imad Ouloul, Karim Afdel, and Abdellah Amghar. 2016. Age Estimation using Local Matched Filter Binary Pattern. (2016).

[30] Suvarna E Padme and PS Desai. 2015. Estimation of Age from Face Images. *International Journal of Science and Research (IJSR)* 4, 12 (2015).

[31] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2010), 1345–1359.

[32] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. 2015. Deep Face Recognition.. In *BMVC*, Vol. 1. 6.

[33] P Phillips, H Moon, S Rizvi, and P Rauss. 2000. The FERET Evaluation Methodology for Face Recognition Algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22 (2000), 1090–1104.

[34] P J Phillips, J Wechsler, and P Rauss. 1998. The FERET database and evaluation procedure for face recognition algorithms. In *Image and Vision Computing J*, Vol. 16. 295–306.

[35] Zakariya Qawaqneh, Arafat Abu Mallouh, and Buket D Barkana. 2017. Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model. *arXiv preprint arXiv:1709.01664* (2017).

[36] Rasmus Rothe, Radu Timofte, and Luc Van Gool. 2015. Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 10–15.

[37] K. Simonyan and A. Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*.

[38] Paul Viola and Michael Jones. 2001. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision & Pattern Recognition, CVPR 2001.*, Vol. 1. IEEE.

[39] Zhiguang Yang and Haizhou Ai. 2007. Demographic classification with local binary patterns. In *International Conference on Biometrics*. Springer, 464–473.

[40] Cha Zhang and Zhengyou Zhang. 2010. A survey of recent advances in face detection. (2010).