Gender Recognition from Facial Images using Convolutional Neural Network

Shubham Mittal
Ambedkar Institute of Advanced Communication Technologies and Research
New Delhi, India
shubhammittl@gmail.com

Shiva Mittal
Ajay Kumar Garg Engineering College
Ghaziabad, India
shivamttl@gmail.com

***Abstract*— Gender is one of the facial attributes whose identification serves crucial step in systems including smart billboards, emotion recognition and surveillance systems. This paper explores a deep learning based solution for automatic detection of gender from face images of a well-balanced dataset. The solution involves a transfer learning framework where the knowledge is reused out of a deep learning model which performed quite well in classification task of other domain. We investigate the reusability of Visual Geometry Group-16 (VGG16), an off-the-shelf Convolutional Neural Network (CNN) model which is pre-trained upon large dataset of natural images. Fine-tuning a segment of the CNN upon the moderately sized dataset yields better performance than obtained from the state-of-the-art approaches when implemented upon the public LFW-Gender dataset.**

Keywords*— *face recognition, gender recognition, deep learning, image classification, neural network

I. INTRODUCTION

Automatic detection of gender from facial images has gained huge attention during recent years. This can be attributed to the increased acquisition of facial biometrics from online platforms as well as offline sources such as webcam, CCTV and smartphone cameras. Recognizing gender from facial images has vast applications [1] including, visual surveillance where a specific gender may be restricted, smart billboards, custom user-computer interfaces, and precursor to more complex tasks such as biometric facial recognition and context specific emotion recognition.

Facial gender recognition task poses various challenges for machines, particularly while dealing with low-quality, misalignment and occlusion in images [2],[3]. A vast literature exists upon the proposed methods to achieve increasingly better performance besides such constraints, some of which are summarized next.

Pixel intensity distributions may differ among the two genders' faces and such differences have been deployed to train binary classifiers [4], [5]. Zhou et al. [6] presented a novel method which processes grey-level information using principle component analysis (PCA) and genetic algorithm (GA) with prior pre-processing steps and train a neural network upon resulting features. Local binary pattern (LBP) histograms [7] have been proposed to be used for classification based on textural differences in faces [8],[9]. Several works [10],[11],[12] investigated into the geometrical differences between male and female features and calculate facial fiducial distances. Histogram of oriented gradients (HOG) [13] have been proposed to be used as discriminatory features for explaining the gender-based facial shape differences even in presence of illumination changes [14]. Azzopardi et al. [15] proposed the usage of trainable COSFIRE (Combination of Shifted Filter Responses) filters [16], which are trainable shape detectors performing efficiently in several other computer vision tasks as well [17]. A fusion of such trainable features with domain specific speeded up robust features (SURF) [18] has been proposed as a robust method against face variations [19]. Yet another fusion strategy [20] demonstrated COSFIRE filters aiding CNNs leading to better performance. Another work [21] achieved better classification performance using clustering and incremental learning with support vector machines (SVMs) [22].

Several facial image datasets have been used in the literature for the task of gender recognition. Those include AR [23], XM2VTS [24], FERET [25], BioID [26], PIE [27], FRGC [28], MORPH-2 [29], LFW [30], CAS-PEAL-R1 [31], Multi- PIE [32] and Adience [33] datasets, most of which contain images captured in constrained environments, but with variations in terms of pose, age, expression, etc. These datasets are not customized for the task of gender recognition and the performance of proposed methods in past have seldom been reported upon any standard dataset.

The main contribution of this paper is the improved test-set performance, upon a well-balanced public dataset of facial images specifically segregated for the task of gender recognition, using a more convenient paradigm of deep learning – the transfer learning, which reuses the pre-trained network for new classification tasks by partial or complete adaption to the given dataset. The resultant CNN model outperforms the existing baseline methods significantly as tested upon the four-fold structure of the public dataset. In the following, we briefly review the deep learning strategy. Experimental details are given in Section III. Section IV explains the experimental results. Finally, Section V concludes the paper.

II. DEEP LEARNING STRATEGY

Deep learning is becoming state of the art method for classification problems in various fields including medicine, structural health monitoring and maritime target tracking [34],[35],[36],[37]. It is an amazing paradigm of machine learning where the representations for the problem are learnt automatically through a hierarchical structure of layers [38] involving several operations such as convolution, sub-sampling and non-linear activation. Features developed within CNNs prove much more optimal than the handcrafted ones, provided the networks are trained with sufficient data samples.

978-1-7281-0899-5/19/$31.00 ©2019 IEEE

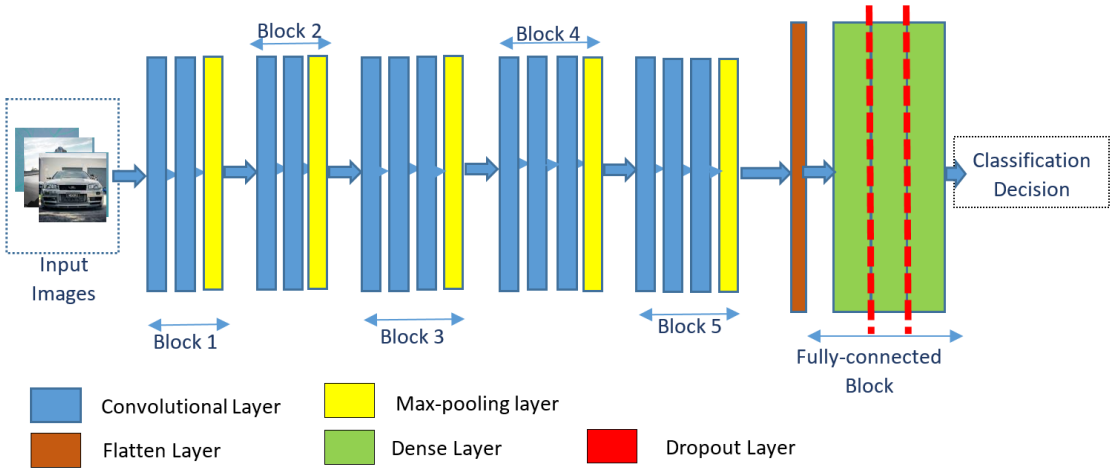

Fig. 1. Horizontally laid down view of VGG-16 CNN. Relu-activation and zero-padding layers have been omitted for brevity of representation. Input-side represents the bottom and classification decisions are made at the top of the hierarchy.

While working out with the solutions to a classification problem using deep CNN, designer may be hindered by the paucity of training samples, since deep nets require millions of samples to converge to a generalized solution to the problem. Further difficulty is the selection of optimal hyper-parameters such as choice of gradient optimizer, learning rate, etc. Such challenges may be overcome using the technique of transfer learning, which aims to reuse the capability of CNN which has been previously trained upon a large set of images from a different domain but achieved considerable performance upon the dataset.

III. Experimentation

The entire experimentation was done on a machine equipped with hexa-core central processing unit (CPU), graphics processing unit (GPU) having 8GB graphics memory and random access memory (RAM) of 16 GB. The software was implemented in Keras [39] running upon TensorFlow-backend [40]. The pre-trained weights of the CNN model used in the work was available within Keras API.

A. Dataset Used

This work used public LFW Gender dataset [41] which is a subset of LFW (Labelled Faces in the Wild) dataset [30] designed for the problem of face recognition in unconstrained environments. The LFW Gender dataset contains total of 5810 images of size 200×200 pixels containing equal number of male and female faces. One of the key saliencies of the dataset is that it has predefined 4-folds structure such that each fold contains train, validate and test subsets with no repetition of a person across the subsets. Table 1 details the distribution of images across the folds.

TABLE I. Distrubution of facial images across 4 folds

Fold Number	Train subset	Validate subset	Test subset
1	1886	1556	2368
2	2296	1284	2230
3	1932	1550	2328
4	2108	1302	2400
Total images= 5810			

B. Transfer Learning

This work re-uses the pre-trained VGG-16 [42] network developed by Visual Geometry Group from University of Oxford, which performed outstanding in ImageNet Challenge 2014 for classification of huge dataset [43] into 1000 classes. As shown in Fig. 1, VGG-16 hierarchical architecture is quite deep with 5 convolutional blocks and a fully connected network. The hierarchy of features grows from bottom to top as they develop towards more abstract representations from simple lines and edges to complex shapes and objects. In other words, bottom layers are more generalized while top-most layers are specific to the recognition problem.

In order to retrain VGG-16 upon facial images, its pre-trained convolutional base up to 'block 5' was truncated with a small neural network consisting of 1024 densely connected neurons followed by a softmax layer containing 2 logits for inference. The dropout layer [44] with 50% probability of retention of units during training was used within the network to avoid possible overfitting. The network's image-input size was modified to accept 50×50 pixels-sized images.

C. Dataset Augmentation

The training images from each split were subjected to augmentation on-the-fly during each iteration of gradient optimization. Each image was randomly rotated within the range of [0, 20] degrees, flipped about the vertical axis and zoomed in or out by ten per cent. Rotation operation was inspired by previous work [45], where it was argued that deliberate addition of misaligned images result in robust classifier.

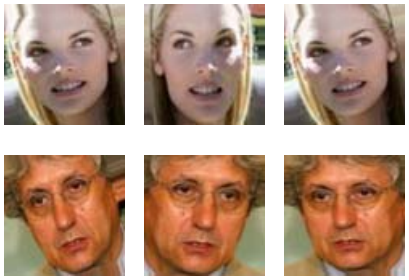

Fig. 2. Specimens of resultant images after applying augmentation

Figure 2 shows the result of applying the stated augmentations on some of the training images. Such augmentation operations increase the sample size and ensure

that the network learns upon a slightly different version of the facial image on every iteration mitigating the chances of overfitting.

D. Training

The modified VGG-16 architecture was retrained in two steps. First, the network weights of entire convolutional base were 'frozen' so that the newly truncated dense network may adapt to the new dataset. In the second step, the hence obtained network is fine-tuned to update the weights of last two convolutional blocks namely, 'block4' and 'block5' so that the network learns the high-level features such as shapes for the problem of facial gender recognition. The procedure is summarized by figure 3.

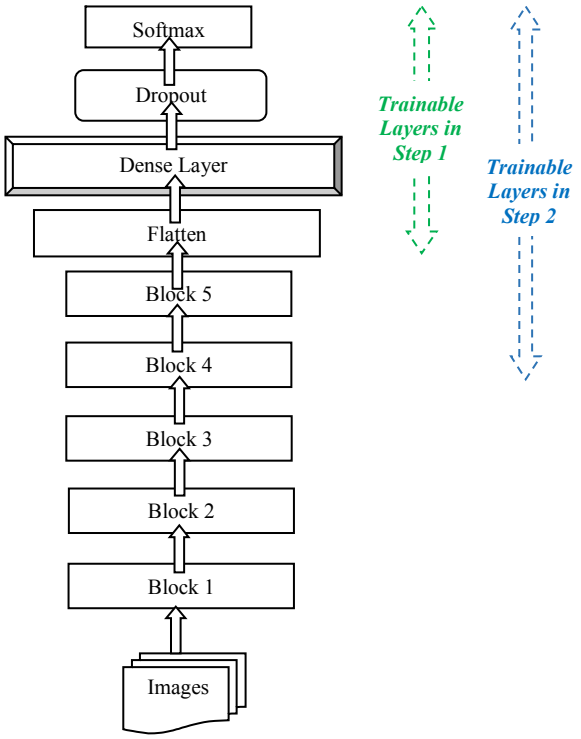

Fig. 3. Modified VGG-16 network for classification of facial images

This work used Adam [46] optimizer for gradient optimization of weights of the network. A moderate learning rate of 0.01 was used when entire convolutional base was kept frozen, while a much lower learning rate of 0.0001 was used while fine-tuning the upper blocks so that the amount of update is gradual rather than distorting the pre-trained weights entirely and overfitting is inhibited. The images were fed to the network in mini-batches of 32 images shuffled out of augmented training subset. After each epoch of iterating entire training subset, the model performance was evaluated on validation subset. To further address the potential overfitting, technique of early stopping was deployed, where the model training was terminated if the performance didn't improve for five consecutive epochs.

Figure 4 depicts the performance trend upon the validation sets of the four folds. The accuracy quickly reaches around 80 percentage point within a few epochs when a small dense neural network is trained upon the bottleneck features of the pre-trained convolutional base.

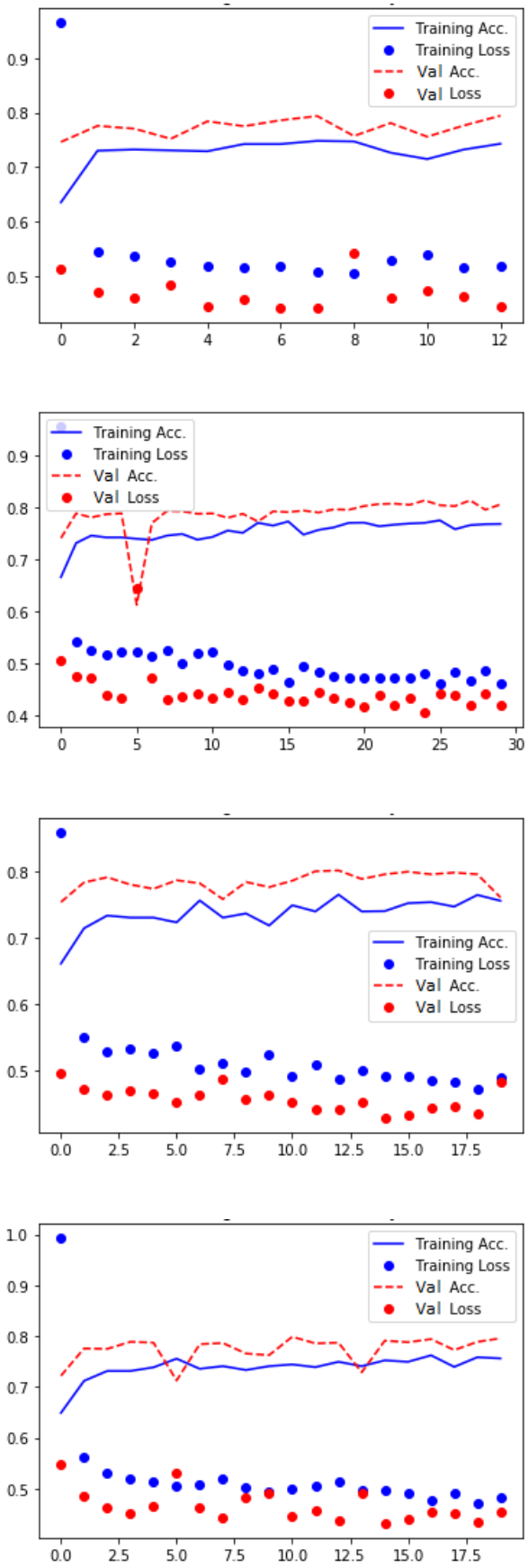

Fig. 4. Performance trends upon four-folds (fold number 1 to 4) while training of truncated dense network.

Validation set performance is improved further to reach more than 91 percentage point when the hence-obtained model is fine-tuned using training images with prior unfreezing of last two convolutional blocks. Figure 5 shows the trend of improvements obtained.

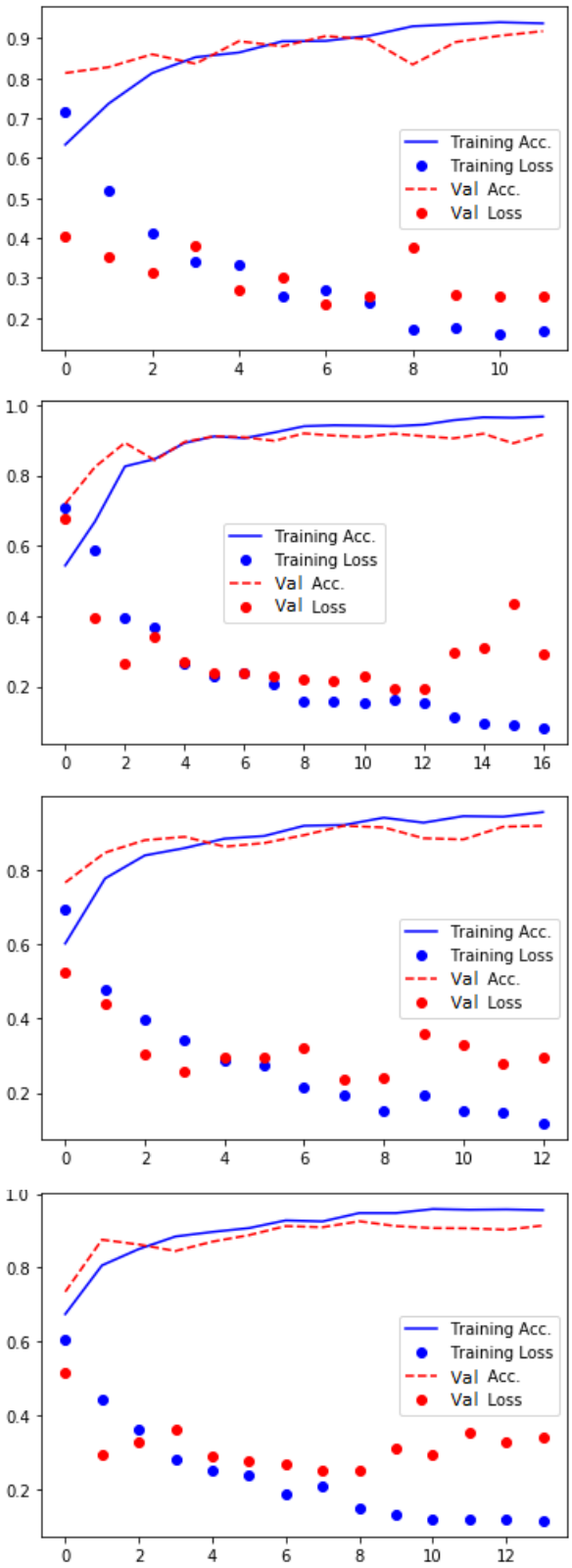

Fig. 5. Performance trends upon four-folds (fold number 1 to 4) while fine-tuning the model parameters till block-4.

IV. Results

A. Test-set performance

After obtaining the fine-tuned modified VGG-16 network trained upon subset of each fold, it was evaluated upon the held-out test subset. Table 2 gives the accuracy achieved by the classifier upon testing subsets from each fold.

TABLE II. TEST SET PERFORMANCE OF FINE-TUNED NETWORK

Fold number	Number of test set images	Accuracy
1	2368	91.34 %
2	2230	91.79 %
3	2328	91.58 %
4	2400	91.04 %

The average accuracy across the four folds was obtained to be (91.4375 ± 0.0780) %.

B. Comparison with baseline performance

The baseline performance for the dataset was established in previous work [41] by experimenting with several combinations of features and classifiers. Candidate features included raw pixel values, eigen or principle component analysis (PCA) [47], [48], linear discriminant analysis (LDA) and random projections. Classifiers included k-Nearest Neighbours (KNN), support vector machines (SVM) [22] with linear and radial basis function (RBF) kernels, and deep CNN. We mention in Table 3 the top-performing cases for each type of classifier as reported in previous work [41].

TABLE III. PERFORMANCE COMPARISON OF BASELINE CLASSIFERS WITH PROPOSED CLASSIFIER

Classifier	kNN	Linear SVM	RBF-SVM	Deep CNN (from scratch)	*Proposed Deep CNN (finetuned)*
Features	PCA features	PCA features	PCA feature-s	Raw pixel values	*Raw pixel values*
Average Test-set accuracy	76.51%	83.88%	86.11 %	87.95 %	*91.44 %*

Figure 6 gives the comparison graph between the classifier performances.

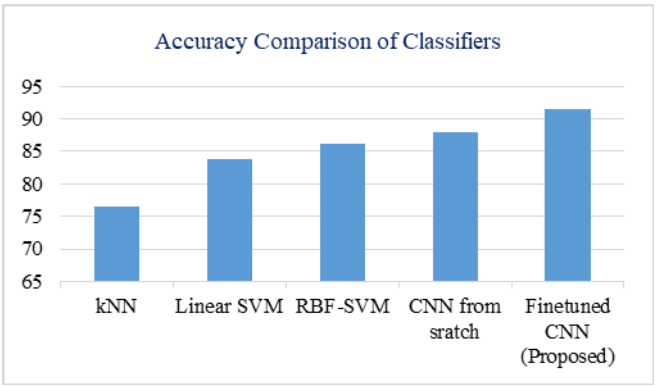

Fig. 6. Comparison graph of proposed vs baseline classifiers

As evident from comparison, test set accuracy improves significantly by 3.5 percentage points against the previously best performing deep CNN. This increment in performance can be attributed to fine dataset augmentation and transfer learning approach where an already mature classifier is adapted to the new recognition task.

V. Conclusions and Future Work

This work explored the paradigm of transfer learning for training a deep learning model upon a small-size dataset of facial images for gender recognition. A pre-trained CNN namely VGG-16 performing fairly good in natural image classification task has been adapted towards the task of facial gender recognition. This is achieved by modifying the topmost classification sub-block of the network to output two probability scores for prediction of both genders. The modified network is fine-tuned in steps with moderate to low learning rate to obtain significantly better test-set performance than the baseline methods previously reported on the same dataset, including CNN trained from scratch. Significant improvement using pre-trained CNN indicate that the weights of a mature CNN upon a problem are more apt to obtain a classifier than that obtained by training a network upon the new dataset entirely from scratch. Future work will be carried out to explore ensemble of such fine-tuned networks for the recognition task.

References

[1] C.-B. Ng, Y.-H. Tay, and B.-M. Goi, "A review of facial gender recognition," Pattern Analysis and Applications, vol. 18, no. 4, pp. 739–755, 2015.

[2] S. Gutta and H. Wechsler, "Gender and ethnic classification of human faces using hybrid classifiers," Proceedings of the International Joint Conference on Neural Networks, vol. 6, pp. 4084–4089, 1999.

[3] C. B. Ng, Y. H. Tay, and B. M. Goi, "Recognizing human gender in computer vision: A survey," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7458 LNAI, pp. 335–346, 2012.

[4] B. Moghaddam and M. Yang, "Learning gender with support faces," IEEE Transactions on Pattern Analysis and Machine Inteltigence, vol. 24, no. 5, pp. 707–711, 2002.

[5] S. Baluja and H. A. Rowley, "Boosting sex identification performance," International Journal of Computer Vision, vol. 71, no. 1, pp. 111–119, 2007.

[6] Y. Zhou and Z. Li, "Real-time gender recognition based on eigen-features selection from facial images," IECON Proceedings (Industrial Electronics Conference), pp. 1025–1030, 2016.

[7] T. Ojala and M. Pietikaeinen, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, 2005.

[8] Z. Yang and H. Ai, "Demographic classification with local binary patterns," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 4642 LNCS, pp. 464–473, 2007.

[9] C. Shan, "Learning local binary patterns for gender classification on real-world face images," Pattern Recognition Letters, vol. 33, no. 4, pp. 431–437, 2012.

[10] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5305 LNCS, no. PART 4, pp. 504–513, 2008.

[11] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3476–3483, 2013.

[12] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Extensive facial landmark localization with coarse-to-fine convolutional network cascade," Proceedings of the IEEE International Conference on Computer Vision, pp. 386–391, 2013.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. I, pp. 886–893, 2005.

[14] V. Singh, V. Shokeen, and B. Singh, "Comparison Of Feature Extraction Algorithms For Gender Classification From Face Images," International Journal of Engineering Research and Technology, vol. 2, no. 5, pp. 1313–1318, 2013.

[15] G. Azzopardi, A. Greco, and M. Vento, "Gender recognition from face images with trainable COSFIRE filters," 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2016, no. August, pp. 235–241, 2016.

[16] G. Azzopardi and N. Azzopardi, "Trainable COSFIRE filters for keypoint detection and pattern recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 2, pp. 490–503, 2013.

[17] G. Azzopardi, L. Fernandez-Robles, E. Alegre, and N. Petkov, "Increased generalization capability of trainable COSFIRE filters with application to machine vision," Proceedings - International Conference on Pattern Recognition, vol. 0, pp. 3356–3361, 2016.

[18] H. Bay, E. Andreas, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF) Herbert," European Conference on Computer Vision, vol. 3951, no. September, pp. 404–417, 2006.

[19] G. Azzopardi, A. Greco, A. Saggese, and M. Vento, "Fusion of Domain-Specific and Trainable Features for Gender Recognition from Face Images," IEEE Access, vol. 6, pp. 24171–24183, 2018.

[20] F. Simanjuntak and G. Azzopardi, "Fusion of CNN- and COSFIRE-Based Features with Application to Gender Recognition from Face Images," Advances in Intelligent Systems and Computing, vol. 943, pp. 444–458, 2020.

[21] I. Dagher and F. Azar, "Improving the SVM gender classification accuracy using clustering and incremental learning," Expert Systems, vol. 36, no. 3, pp. 1–17, 2019.

[22] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[23] A. M. Martinez and R. Benavente, "The AR face database," CVC Technical Report 24, 1998.

[24] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," Proceedings of the Second International Conference on Audio and Video-based Biometric Person Authentication (AVBPA'99), pp. 1–6, 1999.

[25] P. Jonathon Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090–1104, 2000.

[26] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, "Robust face detection using the Hausdorff distance," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 2091, pp. 90–95, 2001.

[27] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) database," Proceedings - 5th IEEE International Conference on Automatic Face Gesture Recognition, FGR 2002, pp. 53–58, 2002.

[28] P. J. Phillips et al., "Overview of the face recognition grand challenge," Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. I, pp. 947–954, 2005.

[29] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, pp. 341–345, 2006.

[30] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in Workshop on Faces in "Real-Life" Images: Detection, Alignment, and Recognition, 2008.

[31] W. Gao et al., "The CAS-PEAL large-scale chinese face database and baseline evaluations," IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans, vol. 38, no. 1, pp. 149–161, 2008.

[32] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," Image and Vision Computing, vol. 28, no. 5, pp. 807–813, 2010.

[33] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170–2179, 2014.

[34] A. S. Kumar and E. Sherly, "A convolutional neural network for visual object recognition in marine sector," 2017 2nd International Conference for Convergence in Technology, I2CT 2017, vol. 2017-January, pp. 304–307, 2017.

[35] W. Yang, Y. Si, D. Wang, and B. Guo, "Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine," Computers in Biology and Medicine, vol. 101, pp. 22–32, 2018.

[36] Z. Fan, Y. Wu, J. Lu, and W. Li, "Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network," arXiv preprint, no. 1802.02208, pp. 1–9, 2018.

[37] M. Leclerc, R. Tharmarasa, M. C. Florea, A. C. Boury-Brisset, T. Kirubarajan, and N. Duclos-Hindié, "Ship Classification Using Deep Learning Techniques for Maritime Target Tracking," 2018 21st International Conference on Information Fusion, FUSION 2018, pp. 737–744, 2018.

[38] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[39] F. Chollet, "Keras: Deep Learning library for Theano and TensorFlow," https://keras. io/k, 2015. .

[40] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," Software available from tensorflow.org, 2016. [Online]. Available: http://arxiv.org/abs/1603.04467.

[41] A. Jalal and U. Tariq, "The LFW-gender dataset," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, vol. 10118 LNCS, pp. 531–540.

[42] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in International Conference on Learning Representations 2015, 2015, pp. 1–14.

[43] K. L. and L. F.-F. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, "ImageNet: A Large-Scale Hierarchical Image Database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[44] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," Journal of Machine Learning Research, vol. 15, pp. 1929–1958, 2014.

[45] M. Mayo and E. Zhang, "Improving face gender classification by adding deliberately misaligned faces to the training data," 2008 23rd International Conference Image and Vision Computing New Zealand, IVCNZ, 2008.

[46] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in International Conference on Learning Representations 2015, 2015, pp. 1–15.

[47] I. Jolliffe, "Principal component analysis. Springer Series in Statistics," Encyclopedia of Statistics in Behavioral Science, p. 487, 2002.

[48] H. Abdi and L. J. Williams, "Principal Component Analysis," Wiley Interdisciplinary Reviews: Computational Statistics, 2010.