International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 2, April 2022, pp. 2066~2078
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i2.pp2066-2078

Gender recognition from unconstrained selfie images: a convolutional neural network approach

Saddam Bekhet[1], Abdullah M. Alghamdi[2], Islam Taj-Eddin[3]
[1]Quantitative Methods Department, Faculty of Commerce, South Valley University, Qena, Egypt
[2]Management Information System Department, Collage of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[3]Information Technology Department, Faculty of Computers and Information, Assiut University, Assiut, Egypt

Article Info

Article history:

Received May 4, 2021
Revised Sep 18, 2021
Accepted Oct 10, 2021

Keywords:

CNN
Deep learning
Gender recognition
selfie images
Soft biometrics
Transferred learning

ABSTRACT

Human gender recognition is an essential demographic tool. This is reflected in forensic science, surveillance systems and targeted marketing applications. This research was always driven using standard face images and hand-crafted features. Such way has achieved good results, however, the reliability of the facial images had a great effect on the robustness of extracted features, where any small change in the query facial image could change the results. Nevertheless, the performance of current techniques in unconstrained environments is still inefficient, especially when contrasted against recent breakthroughs in different computer vision research. This paper introduces a novel technique for human gender recognition from non-standard selfie images using deep learning approaches. Selfie photos are uncontrolled partial or full-frontal body images that are usually taken by people themselves in real-life environment. As far as we know this is the first paper of its kind to identify gender from selfie photos, using deep learning approach. The experimental results on the selfie dataset emphasizes the proposed technique effectiveness in recognizing gender from such images with 89% accuracy. The performance is further consolidated by testing on numerous benchmark datasets that are widely used in the field, namely: Adience, LFW, FERET, NIVE, Caltech WebFaces and CAS-PEAL-R1.

Corresponding Author:

Saddam Bekhet
Quantitative Methods Department, Faculty of Commerce, South Valley University
Qena-Safaga Rd, Qena Governorate, Egypt
Email: saddam.bekhet@svu.edu.eg

1. INTRODUCTION

Automated soft biometrics identification has attracted a great deal of attention in the past era. This was primarily related to the uprising dependence on surveillance systems that produces enormous volumes of data that need to be examined [1], [2] in an off-line manner. Furthermore, it is possible to obtain soft biometrics without subject participation from low quality videos/images, making them highly valuable [2]. In addition, people normally utilize these soft biometric to identify each other [1]. Finally, in today's social era, images are described digitally while vast users are interested in their semantic content. However, it is difficult to find correspondences between the digital and semantic level [3]; this is where gender, as a soft biometric, plays a very important role, due to its close relation to the semantic level [4].

The human facial area is a core demographic attribute that effectively helps in gender recognition [4] and has been extensively studied since 1990 [5]. The gender recognition issue is always considered a

two-class problem in which a standard input query face image is analyzed and assigned to either male or female class [6]. Broadly, facial pictures are presumably the most widely utilized soft biometric to identify individuals [7]. Thus, numerous field specialists depended on facial examination for gender recognition. This was fully experimented besides outer face region landmarks, i.e., hair and beard shape, that positively assisted the recognition process [8]. Conclusively, the aforementioned research style has always been the common work scenario for the standard gender recognition work from facial photos. Sample standard face images are shown in Figure 1 for illustrative purpose. A diverse sample of standard face images depicted from the CAS-PEAL-R1 [9], FERET [10], LFW [11] and NIVE [12] datasets that depicts full faces in controlled environment. The images were shot over a textured background, mostly cinematic lighting conditions and suitable viewing angles. All of these factors help to recognize gender using non-complex facial features.

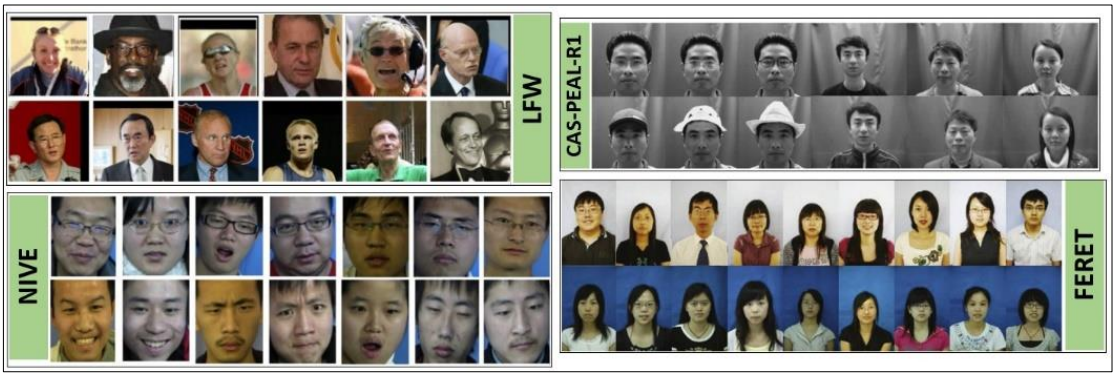

Figure 1. Sample standard face images

However, in the recent years gender recognition became more challenging, especially when dealing with unconstrained real-world photos [13] which depicts full/partial occluded frontal facial images that depicts extreme viewing angles. This situation is best illustrated in the contemporary selfie photos, which are hard to analyze using the systematic facial-based approaches [14]. This is due to their variable resolution, extreme occlusion, and deformation, induced by the real-life situations they were recorded in. The case becomes harder when it is required to perform a specific demographic analysis on such selfie images, because they are not photographed in the best capturing conditions, i.e., neither a textured background nor a controlled lighting. Despite recent advances in machine learning (ML) methods, the issue is still problematic, mainly due to ML sensitivity to variations in the facial area and the infinite changes [15] that could be noticed in selfie images, i.e., tattoos and emojis.

This research paper's main contribution is introducing a solid convolution neural network (CNN) network model that identifies human gender from a selfie image. The introduced deep learning approach uses a transferred learned knowledge harvested from the ImageNet dataset [16] (1M images). This is beside to new group of kernels and features acquired from the selfie image [17] dataset itself. The use of transferred learning in conjunction with newly acquired function maps, repays for the unconstrained dynamic nature of selfie photos. This shapes the deep learning model to better identify gender compared to state-of-art approaches. Eventually, the obtained research outcomes present a robust benchmark for selfie images gender recognition utilizing deep learning approaches and leaves a space for any future enhancements.

The remainder of this research paper is structured as follows: section 2.1 covers relevant background work on selfie photos as a recent topic in the field of computer vision (CV). Section 2.2 presents and addresses relevant literature work on gender recognition. The contributed CNN model is presented in section 3. The experiments and testing protocols are discussed in the section 4. The paper is eventually concluded in the section 5.

2. LITERATURE BACKGROUND

2.1. Selfie images

Since the proliferation of smart-phones and the invasion of social media websites, e.g., Facebook, and Instagram, selfie images became a major aspect in our life. Selfies are self-portraits usually taken and shared on social media, as a means for people to introduce/express themselves to the world [17]. The

dramatic rise in selfies gave it a big-data dimension and forced its presence as a newly formed CV research area. Furthermore, conventional computer vision approaches were not able to tackle selfies successfully. This is due to two major causes: i) their non-standard method of recording makes them often vulnerable towards severe occlusion of facial/body landmarks and ii) their big-data aspect, which makes hand-engineered features costly to extract and may not extrapolate in such quantities [3]. In addition, these selfies may depict a facial side-view with artificial effects or stamped emojis, i.e., cartoon moustache. These issues contribute to the toughness of gender recognition from such selfie photos. A collection of selfie photos that visualizes part of the aforementioned issues are illustrated in Figure 2.

Figure 2. A group of sample selfie photos [17] that shows the dataset common problems, such as occlusion, artificial effects, and side-face/partial-face view

To further quantify the selfie dataset problems, it was processed using the famous Viola and Jones [18] face detector. This is to further highlight its poor performance under the standard facial recognition algorithms (common core part for gender recognition algorithms). The maximum reached detecting accuracy was 32% compared to 100% on FERET, CAS-PEAL-R1 and LFW, as shown in Table 1 (based on a subset of each dataset). Noting that, FERET, CAS-PEAL-R1 and LFW and are fully standard datasets widely used for facial analysis and depicts standard frontal faces. For demonstration purpose, Figure 3 shows the face detection result on the 16 images depicted earlier in Figure 2, where the detection performance is very poor on such unconstrained dataset, as only 5 successful attempts

Table 1. Face detection accuracy on multiple facial analysis/verification benchmark datasets compared to the selfie images dataset

Dataset	Face detection accuracy (%)
LFW [11]	100
FERET [10]	100
NIVE [12]	31
Caltech WebFaces [19]	92.8
CAS-PEAL-R1 [9]	100
Adience [20]	93.5
Selfie images	32

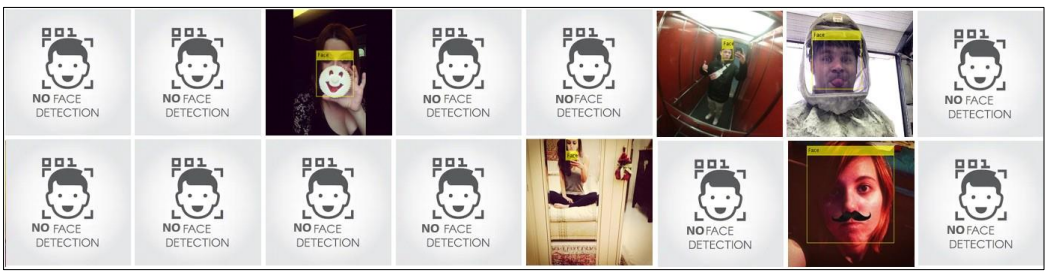

Figure 3. Face detection results on the 16 sample selfie images from Figure 2

From a researcher perspective, selfies were almost never seen in CV work, as they were mostly connected to psychology research studies [21]. From the CV viewpoint, selfies were analyzed with respect to

various demographic attributes, i.e., Asian, senior, and young, where HOG and SIFT, features were used identify these attributes. However, the achieved performance was limited, i.e., <38% and <34% accuracy for gender identification using HOG and SIFT respectively [17].

Selfie images are, without a doubt, a modern-day global craze that is extremely sophisticated and special case in computer vision field, dignified to be analyzed. This will facilitate unlocking the underlying profit of such massive quantities of available selfies (»24-billion photos [4]). Furthermore, their unconstrained nature contributes to build a better gender recognition system. This may help to enhance current surveillance systems in circumstances where surveillance footage is being examined to identify a particular suspect with certain demographic attributes that are related to the gender.

2.2. Gender recognition from facial images

The traditional historic approach of gender recognition requires features extraction from standard face images [10], [22], followed by merging these features to build a heuristic model to inspect the gender [23]. Broadly, gender recognition techniques can be grouped into either i) geometric or ii) appearance-based strategies [24], as illustrated in Figure 4. The first strategy intends to portray anthropometric estimates, for example, confront width/length and distance between the eyes [25]. The second strategy depicts the surfaces of facial skin where furrows and bulges are present. In general, both methods are on 2D hand-crafted features. However, 2D features depict some limitations. This is related to the dynamic environmental and experimental factors as illumination condition and head pose. Recently, there have been several attempts to use 3D features [25] to pump the effectiveness of current gender recognition techniques. However, such techniques require the affordability of 3D imaging devices or the construction of 3D face models, which is expensive in either case, and even inapplicable for the case of selfie images that depict other visible body parts in contrast to the facial area. Later, a different approach was taken considering the popularity of using SVM as a strong and reliable classifier. SVM was attempted over various popular face-applied descriptors, i.e., low plasticity burnishing (LBP) [10], LBP and HOG [25], fisher vectors [25] or even raw-pixel data [20]. Though, the ability of SVM-based approaches saturated to a very good levels, but the bulk of the work used standard full facial images in ideal/near-ideal conditions. This is clearly not the case of the targeted selfie photos that are far away of any near-ideal conditions.

Figure 4. Classification of classic gender recognition techniques. Inner images credits are for [26], [27]

Recently, there were numerous attempts to use CNN to recognize gender [28]–[30] from "in the wild" images. However, the results did not cross an accuracy borderline of 88% in the best case [29]. Moreover, those in "in the wild" images were often clean; with respect to viewing conditions and depict attention from the subjects being photographed [20]. Also, the bulk of previous CNN models requires special pre-processing for the input source image, such as cropping and face alignment [26]. Such pre-processing was not contributive for unconstrained images and added an extra computational load on the system performance [31].

Conclusively, hand-crafted features utilize the available domain-specific knowledge towards more accuracy performance. However, this type of features requires robust initial pre-processing, e.g., face alignment [20]. This adds an additional computational load [31], and any drop in this initial pre-processing affects the overall system performance. Furthermore, the current SVM-based techniques are incapable of fully handling "in the wild"/unconstrained images. Thus, their performance will be worse with selfie images, due to their extreme and true unconstrained imaging nature, as illustrated in the previous section. However, CNN models might have promising results in inspecting gender from selfies, where the network might learn features that belongs to the full image not only the facial area. This includes features from the internal (eyes/nose/mouth) and external (head/ears/chin) face zone. Moreover, the CNN may make use of contextual features that are not considered as a part of the face zone, i.e., breast size, shoulders' size, clothing style, number of tattoos and any other features the network might learn during training.

3. RESERCH METHOD

CNNs is a breakthrough advance in artificial intelligence that changed the entire related CV research [1], [32]–[34]. These CNNs use comparatively less pre-processing than other methods that utilizes feature extraction and description. A typical CNN network is entirely responsible for creating its own filters for unsupervised learning. This is in contrast with other conventional algorithms that rely on selective pre-processed hand-crafted features. Furthermore, the availability of enormous, labelled archives of data, i.e., ImageNet, enabled CNNs to learn and infer valuable features' formations, which rendered a non-precedent breakthrough in various pattern recognition tasks that even exceeded human performance [2]. In addition, using a bit of network reshaping, these inferred representations and/or features could be surprisingly reused and applied to a newly different problem. This is applicable if exists a new dataset smaller than the source dataset that were utilized during the initial CNN model training [35]. In spite of the transferred learning, yet still some significant modifications and full training to be carried out to the original network model to suit the new task. The new parameterization and training phase might yet still prolong to multiple weeks, where the new network need to iterate through every example of the training data to learn new features, remap previous ones and adjust the last fully connected group of layers [35]. Thus, the primary goal of this research study is to use the CNN capacity in transferred learning. Hence, we introduce the network architecture depicted in Figure 5.

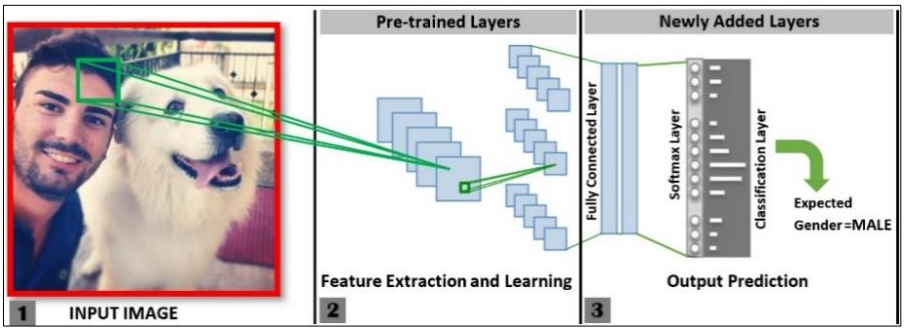

Figure 5. An outline structure of the proposed gender recognition deep learning model

The deep learning core operation is applying a series of convolutional operations on the input query images followed by aggregating layers to produce the resultant feature maps. The convolution step is illustrated in (1):

$$(\mathrm{F} \circledast \mathrm{V})(\mathrm{i},\mathrm{j}) = \sum\sum \mathrm{V}(\mathrm{m},\mathrm{n})\mathrm{F}(\mathrm{i}-\mathrm{m},\mathrm{j}-\mathrm{n}) \quad (1)$$

where F is the input query image and V represents the 2-dimensional convolution matrix, and $\circledast$ acts as the convolution operator that rolls over the input image. Mathematically, the transferred learning process is formulated through identifying a source domain problem with specific data in this way:

$$B_S = \{(w_{S_1}, k_{S1}), \ldots, (w_{Sn_S}, k_{Sn_S})\} \quad (2)$$

The data instance $w_{Si} \in W_S$ and the associated class label is $k_{Si} \in K_S$. The target domain data is defined as $B_T = \{(w_{T_1}, k_{T1}), \ldots, (w_{Tn_T}, k_{Tn_T})\}$, where $w_{Ti} \in W_T$ is the data instance and $k_{Ti} \in K_T$ is the associated class label. However, commonly:

$$0 \leq n_T \ll n_S \quad (3)$$

where n_S are the target data (selfie images), which does not exist in equal quantity as n_T, ImageNet in this scenario. Transfer learning targets enhancing the knowledge of a predictive function $h_T(\cdot)$ of a specific target domain problem B_T considering the knowledge extracted from the initial source domain problem B_S using the learning task T_S. However, the following constrains further defines the task:

$$\begin{cases} B_S \neq B_T \\ T_S \neq T_T \end{cases} \quad (4)$$

The function $h(\cdot)$ represents the objective predictive function that is learned from the associated training pairs of data $\{w_i, k_i\} \equiv \{feature, label\}$, where $w_i \in W$ and $k_i \in K$. Hence, the feature space is expressed by W, and the label space is expressed by K. The introduced gender recognition CNN architecture is inspired by the well-known AlexNet [36]. AlexNet is a kind of feed forward CNNs that was fully trained over the ImageNet dataset. The standard AlexNet network embodies eight major layers (five convolution and three fully connected). This network was able to classify query images into thousand various object categories, as food, plants, materials, and numerous animals. Thus, the network earned sufficient knowledge of dense and various feature representations, which makes it an excellent starting core for the introduced gender recognition CNN model.

The original AlexNet require an input RGB images of dimensionality 224×224. This was altered to 227×227 during the data augmentation phase, in accordance with the new selfie images dimensionality. Furthermore, the AlexNet contains an output layer of thousand neurons, corresponding to the individual ImageNet [16], [36] internal object classes. Hence, as a part of the rectification step the final three layers were fine-tuned to fit for the tackled gender classification problem. This was done by replacing these layers with a full connection (fc) layer plus a SoftMax layer and an output layer (binary classification, i.e., males/females). This fine-tuning step enables the network to infer the required bias and specifics of the selfie data, which contributes towards robust recognition performance. Additionally, to counteract overfitting, due to the selfie dataset limited size; two dropout layers (randomly 50%) were added to the network. Figure 6 shows the implementation details of the entire CNN network layers after adjustments. The following section describes the selfie dataset characteristics and summarizes the experimental findings with a detailed investigation.

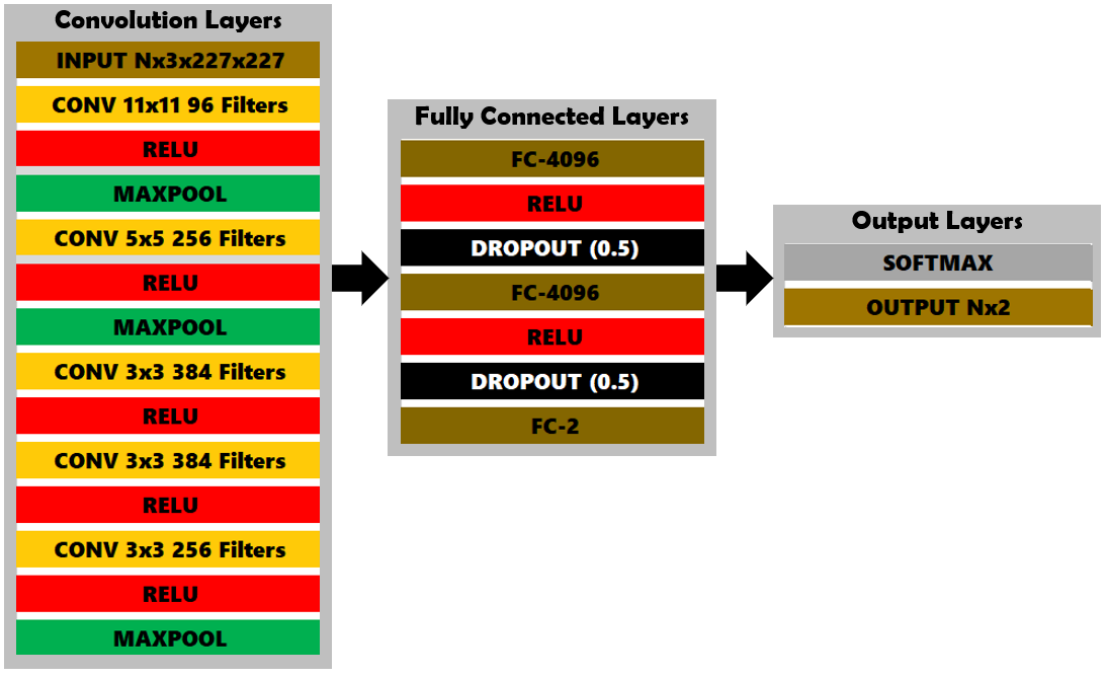

Figure 6. The proposed selfie images gender recognition CNN internal implementation details

4. RESULTS AND DISCUSSION

The performance of the proposed CNN model for gender recognition in unconstrained selfie images is investigated in this section. Section 4.1 introduces the selfie dataset characteristics. Section 4.2. presents the steps and details of network training stage, besides the experimental part and its findings in section 4.3.

4.1. Dataset

The experiments in this research were performed using the only and first selfie images dataset [17] (as far as we are aware). The selfie dataset consists of 46,836 pictures that was selected from an 85 k pictures harvested from the web. Every image within the dataset contains one face image, wherever all the multiple faces pictures were excluded by the creators of the dataset. (However, during our experiments multiple faces images were found, but they only represent 0.25% of the total dataset size and did not affect any of the

reported results). The dataset is annotated with thirty-six completely different attributes scattered across many classes, such as gender, age, race, face-shape, accessories, and many other attributes. Figure 2 shows a diverse collection of selfie photographs that belongs to this dataset.

4.2. Network training stage

The randomized translation phase is required to alleviate the data positional bias, as most of the selfie photos are mostly centered. Furthermore, the selection of ± 30 range reduces the effect to $\sim$10% of each image size, which is a standard procedure [37]. These pre-processing steps are regularly applied to the entire dataset at runtime to fictitiously expand its volume using label-preserving techniques [38] with minimum computational/memory load.

Throughout the network training phase, a gradient descent of type stochastic was used with ten examples as the patch-size. The momentum and weight decay values were set to 9×10^{-1} and 4×10^{-4} respectively. The selection of such small value for the weight decay is essential for a correct learning process, as it minimizes the network's training error. The entire experiments and results were performed with an Intel Core i7 computer occupied with 16 gigabytes of working RAM. The training duration lasted for 480 hours ($\equiv$20 days) to loop through the whole training dataset (original+augmented=$\sim$150k) and reshape the network intrinsic parameters according to the newly data. The following section introduces and analyses the proposed CNN model performance regarding the gender recognition task in unconstrained images.

4.3. Network performance

This section introduces the procedure in setting up experiments to evaluate the proposed network model and analyzing the obtained results. The mostly common experimental data setup was followed, the dataset was split randomly by considering 70% of the data to training stage and 30% to validation stage. Because the selfie dataset has no previously established test-set split, the aforementioned split does not affect any of the findings. Furthermore, the accuracy metric [39] is adopted (5), as commonly used in literature to present the final results.

$$Accuracy = \frac{\sum_{i=1}^{N} I_k(\widehat{y_i})}{\sum_{i=1}^{N} I_k(y_i)} \quad (5)$$

Where the count of test-set categories is K, while N is the testing queries number. $I_k(y)$ is a binary function, which assigns to one when $y = k, I_k(y) = \text{one}$, otherwise assigns to zero. $\widehat{y_i}$ and y_i are the predicted and true labels consecutively of the i^{th} sample respectively. In addition, the validation-loss metric is also adopted in the experiments. This is to further provide an additional consolidation measure about the presented model performance. This metric quantifies how good the presented network model is generalizing to unseen data, (6) shows the loss function. Where $Y_{predicted}$ is the network prediction associated with the ground truth values Y_{true}.

$$LogLoss = \sum_{i=1}^{N} log\left(\left(Y_{true}, Y_{predicted}\right)^2\right) \quad (6)$$

The network achieved 89% accuracy on the selfie dataset with 5^{-1} log-loss, as shown in Figure 7. Practically, this is a good result knowing that the human ability to identify gender is 95% [23]. Such result confirms that the learned features from the ImageNet repository are quite sufficient to generalize to the selfie data. However, a large share of this result is related to the additional feature-maps that were reformed throughout the model training stage.

The introduced CNN model's output is also contrasted to five benchmark baselines in order to highlight its solid performance. The chosen benchmarks represent the current research directions in gender recognition work: Namely i) hand crafted-based approaches and ii) CNN-based approaches. The three-hand crafted-based approaches are dense HOG3D [17], dense SIFT [17] and a numerous group of features extracted from beard, moustache, forehead, face shape and cheek that are fused through a series of AdaBoost classifiers. For CNN based techniques they are expressed by the modern Google TensorFlow teachable machine [40] that is powered by Google TPUs and MXNET CNN that represent the stat-of-art in facial analysis work [41].

The results depicted in Figure 8 confirms the proposed CNN model robust performance, as it per-formed better than hand crafted based approaches with $54.6 \pm 2.8\%$ and $1.05 \pm 0.6\%$ for the CNN based approach. Furthermore, in order to highlight and confirm the effectiveness of transferred learning, the experiment was repeated using GoogLeNet [37]. GoogLeNet has a totally different structure compared to the AlexNet model, i.e., Inception Modules [37], and also had learned different feature representations

considering the ImageNet data. Similarly, GoogLeNet had to go through the same tweaking as the proposed CNN model, i.e., layers adjustment and training. The network scored an accuracy score of 87.9%, as depicted in Figure 9, which is comparable to the proposed AlexNet based model. This result confirms the benefit of transferred learned and highlights the power of deep networks to recognize gender from those very non-standard selfie images.

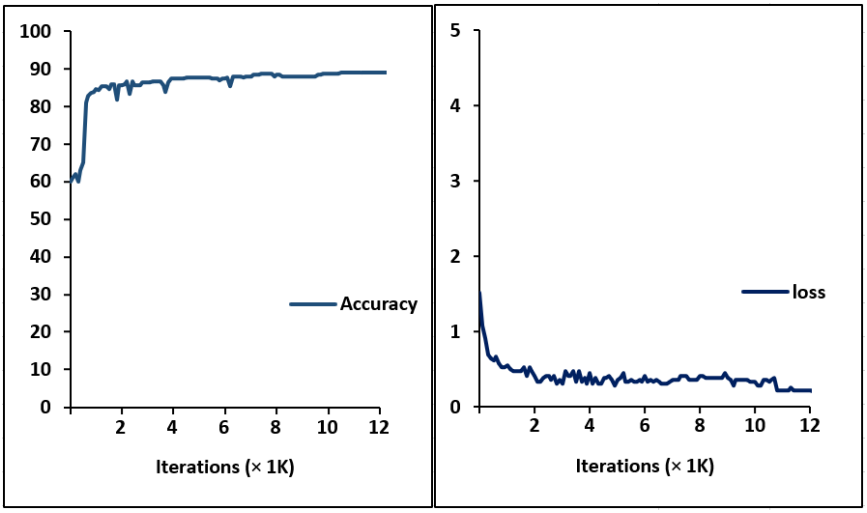

Figure 7. The performance of the introduced network model across the first 12k epochs

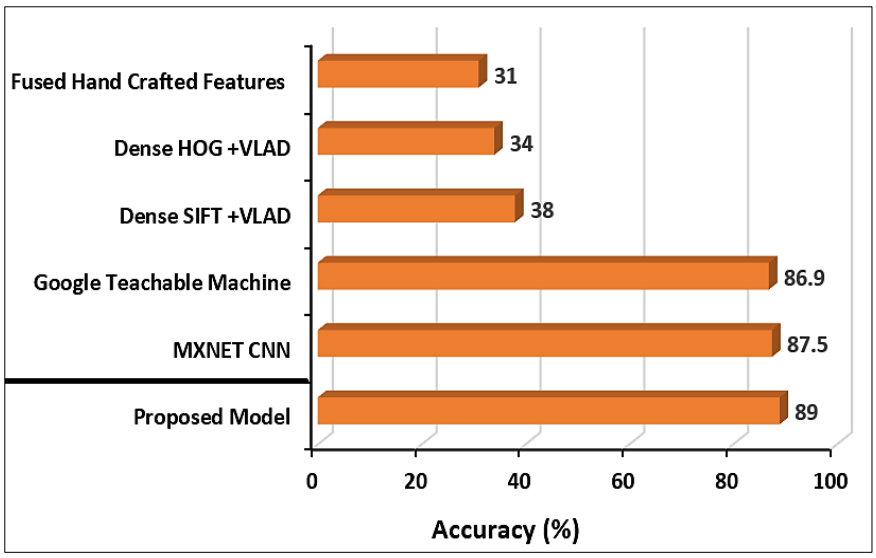

Figure 8. A comparison of the proposed CNN gender recognition model's performance versus hand-crafted baselines and CNN-based baselines

Furthermore, the proposed network model's accuracy performance was confirmed using seven additional popular benchmark facial-analysis datasets. The first dataset: CAS-PEAL-R1 [9], is a well-known Chinese facial dataset, which contains a collection of 99594 images captured from 1040 individuals (445 females and 595 males). The second: FERET [10], is a common face analysis dataset that contains 14126 images from 1199 subjects and another duplicate set of 365 images as well. The third: NIVE [12], is a facial expression examination dataset, captured from a collection of 215 test subjects mimicking various facial expressions. The fourth: LFW [11], is a face verification public dataset that contains 13233 images collected from 5749 different subjects. The fifth: Caltech [19] embodies a total of 10524 image faces that were collected from the internet. The sixth: Adience [20], is composed of 26580 images that belong to 2284

different subject. It is widely used for gender recognition and age estimation. It depicts an extreme variety in terms of gender, including a large quantity of children and a lot of images with very low resolution. The last dataset is an artificially synthesized version of the public MS-Celeb [42] dataset (47917 image). These datasets all show male and female figures with large intra/inter class variations that consolidates the reached result. This experiment was performed following the recommended cross-fold validation setting for each dataset as described in its respective paper.

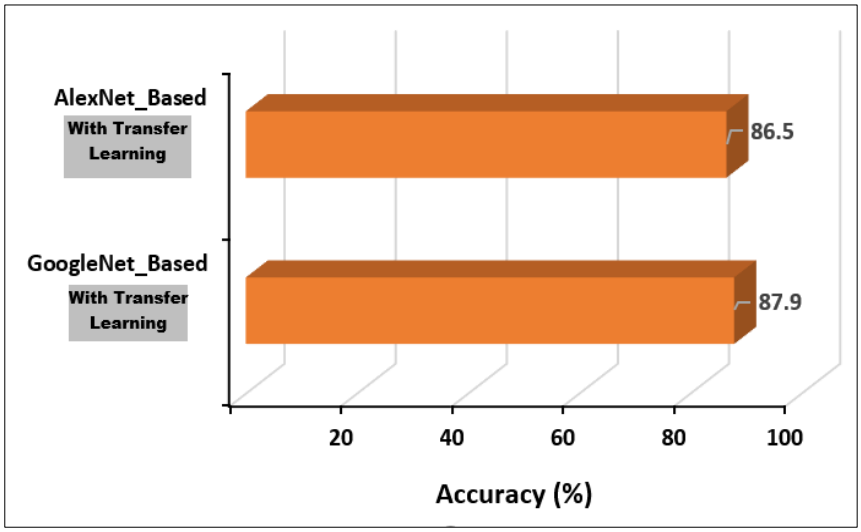

Figure 9. The proposed CNN gender recognition model performance using AlexNet versus GoogLeNet

The proposed model gender recognition performance on the aforementioned group of benchmark datasets is depicted in Figure 10. The result shows an average accuracy of 95.01±3.1% over LFW, FERET, NIVE, MS-Celeb, CAS-PEAL-RI, Adience and Caltech WebFaces datasets. Finally, in order to have a better understanding of internal network learning, Figure 11 shows a visualization of the first layer learnt convolutional kernels. The figure reflects the network ability to learn various constructive components of images such as edges and color blobs. This is plus a rich group of orientation/frequency filters, which contributes to detecting input images features.

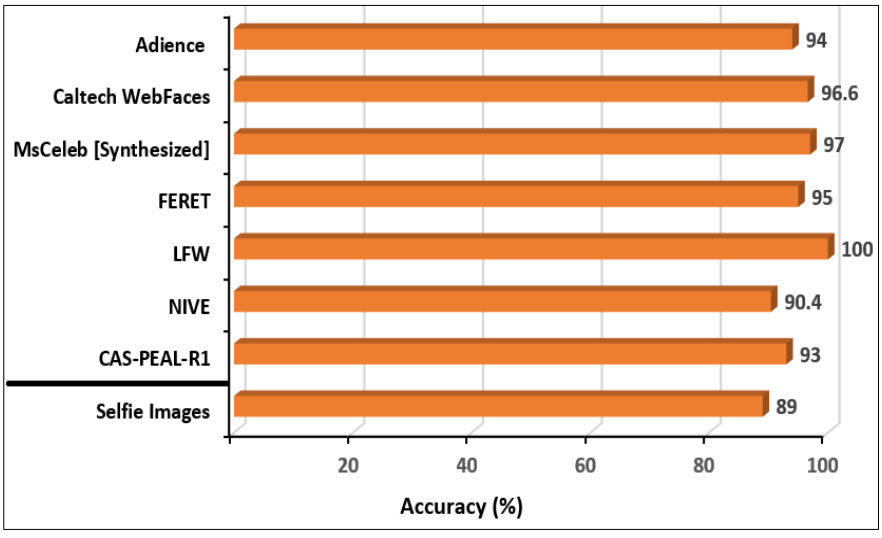

Figure 10. The performance of the introduced CNN model on seven public datasets. The selfie dataset is added to the figure for comparison

Figure 11. Initial convolutional kernels learned by the first convolutional layer versus the final *fc* Layer

Regarding the introduced CNN model qualitative performance, Figure 12 shows a random set of selfie photos that were classified using their gender ground-truth annotation through the introduced network model. The result shows the robust performance of the introduced deep learning CNN model, as it has learnt a rich group of features to inspect the gender in these unconstrained selfie photos. However, in some limited cases, the network cannot classify the type of input query image. For example, in Figure 12, the image indexed at position 1×8 (row×column) was mistakenly classified as female. This was due to the inevitable occlusion in the picture caused by the horse subject. The same occlusion issues apply to the image indexed at location 3×3, where most of the subject facial/body features are occluded.

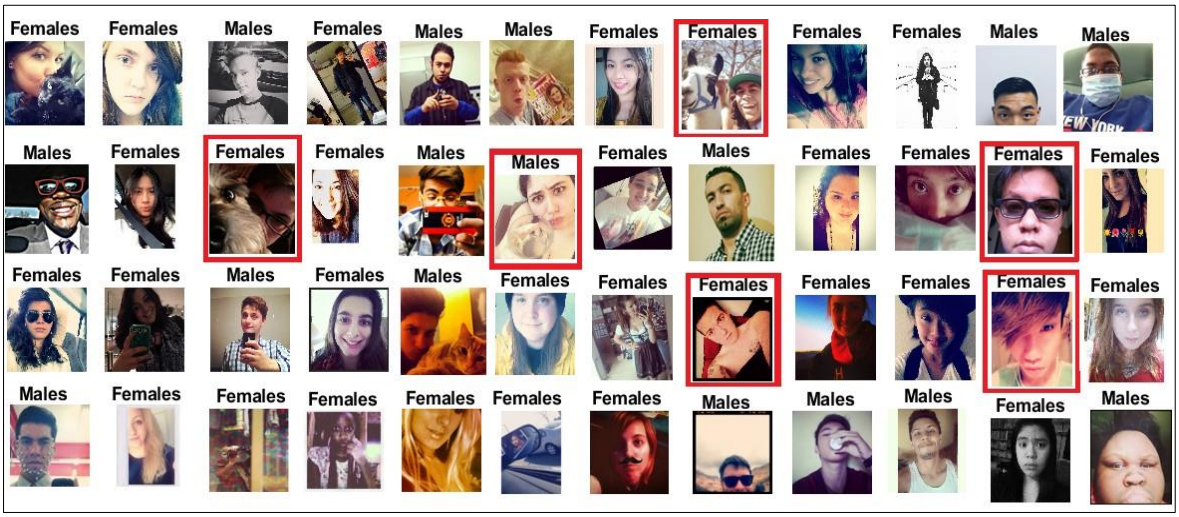

Figure 12. An example images from the selfie dataset that were classified using the introduced deep learning network model. Wrong predictions are marked by a red square

4.4. Data pre-processing effect

The full past results were performed on the selfie dataset without any attempted pre-processing (it is commonly to perform pre-processing in related literature). This is related to the CNN capability to use images at their raw format. However, this section investigates the effect of data pre-processing on the accuracy performance of the proposed network model. Practically, there are various pre-processing steps that could be applied, but only detection-based pre-processing is adopted for this experiment. The famous Viola and Jones face detector [18] is adopted, due to its robustness. Furthermore, the experiment was repeated with respect to the upper-body image only, to further investigate the effect of body-features on the recognition accuracy. The same CNN model was tested on a subset of the selfie dataset, which has been processed using Viola and Jones face detector. The model accuracy was 87.5% for facial images and 88.8% regarding upper-body images, as shown in Figure 13. This result emphasizes the importance of the facial area and its richness with features, followed by the upper body importance to identify gender.

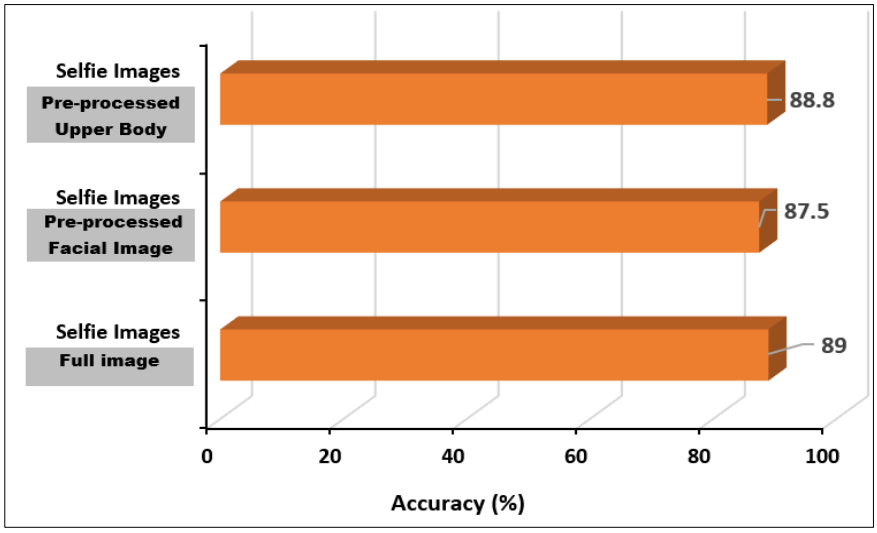

Figure 13. The effect of data pre-processing on gender recognition performance using the selfie dataset

5. CONCLUSION

An effective deep learning approach for human gender recognition from selfie images was introduced in this paper. Selfie images are captured by normal users for themselves any time/where in real-life environment, which makes them highly unconstrained. The presented CNN model achieved 89% accuracy considering the selfie dataset and 100% on other common benchmark datasets, i.e., LFW. This performance was achieved by harvesting a knowledge transferred from the ImageNet dataset to enrich the network performance. However, the network had to include extra layers and conduct complete training epochs that lasted for more than two weeks even with such transferred learning. Following such exhaustive training the network had learnt numerous image components, i.e., edge patterns and color blobs, which are extremely important in detecting features of input images and correctly classifying its respective gender. The paper also conducted an extra analysis to determine the importance of various selfie image components, e.g., cropped face and upper body only.

Future work will mainly include addressing three problems. The first is improving the accuracy through constructing other different CNN models. The second is minimizing the required resources utilized by the proposed CNN approach. The final is expanding the analysis of the selfie images dataset to other soft biometrics and facial attributes.

REFERENCES

[1] S. Bekhet and H. Alahmer, "A robust deep learning approach for glasses detection in non-standard facial images," *IET Biometrics*, vol. 10, no. 1, pp. 74–86, Jan. 2021, doi: 10.1049/bme2.12004.

[2] T. J. Yu, C. P. Lee, K. M. Lim, and S. F. A. Razak, "AI-based targeted advertising system," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 13, no. 2, pp. 787–793, Feb. 2019, doi: 10.11591/ijeecs.v13.i2.pp787-793.

[3] Y. Su and C.-C. J. Kuo, "On extended long short-term memory and dependent bidirectional recurrent," *Neurocomputing*, vol. 356, no. 1, pp. 151–161, 2019.

[4] V. K. Verma, S. Srivastava, T. Jain, and A. Jain, "Local invariant feature-based gender recognition from facial images," in *Soft Computing for Problem Solving*, Springer Singapore, 2019, pp. 869–878.

[5] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "Sexnet: A neural network identifies sex from human faces," in *NIPS*, 1990.

[6] N. A. Binti Mat Kasim, N. H. Binti Abd Rahman, Z. Ibrahim, and N. N. Abu Mangshor, "Celebrity face recognition using deep learning," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 12, no. 2, pp. 476–481, Nov. 2018, doi: 10.11591/ijeecs.v12.i2.pp476-481.

[7] C.-Y. Hsu, L.-E. Lin, and C. H. Lin, "Age and gender recognition with random occluded data augmentation on facial images," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11631–11653, Mar. 2021, doi: 10.1007/s11042-020-10141-y.

[8] F. Juefei-Xu, E. Verma, P. Goel, A. Cherodian, and M. Savvides, "DeepGender: Occlusion and low resolution robust facial gender classification via progressively trained convolutional neural networks with attention," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2016, pp. 136–145, doi: 10.1109/CVPRW.2016.24.

[9] Wen 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, Jan. 2008, doi: 10.1109/TSMCA.2007.909557.

[10] B. Li, X.-C. Lian, and B.-L. Lu, "Gender classification by combining clothing, hair and facial component classifiers," *Neurocomputing*, vol. 76, no. 1, pp. 18–27, Jan. 2012, doi: 10.1016/j.neucom.2011.01.028.

[11] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[12] S. Wang *et al.*, "A natural visible and infrared facial expression database for expression recognition and emotion inference," *IEEE Transactions on Multimedia*, vol. 12, no. 7, pp. 682–691, Nov. 2010, doi: 10.1109/TMM.2010.2060716.
[13] A. Rattani, R. Derakhshani, and A. Ross, "Introduction to selfie biometrics," in *Selfie Biometrics*, Springer International Publishing, 2019, pp. 1–18.
[14] S. Ravidas and M. A. Ansari, "Deep learning for pose-invariant face detection in unconstrained environment," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 1, pp. 577–584, Feb. 2019, doi: 10.11591/ijece.v9i1.pp577-584.
[15] M. Nimbarte and K. Bhoyar, "Age invariant face recognition using convolutional neural network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 4, pp. 2126–2138, Aug. 2018, doi: 10.11591/ijece.v8i4.pp2126-2138.
[16] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.
[17] M. M. Kalayeh, M. Seifu, W. LaLanne, and M. Shah, "How to take a good selfie?," in *Proceedings of the 23rd ACM international conference on Multimedia*, Oct. 2015, pp. 923–926, doi: 10.1145/2733373.2806365.
[18] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.
[19] W. Setiawan, M. I. Utoyo, and R. Rulaningtyas, "Reconfiguration layers of convolutional neural network for fundus patches classification," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 10, no. 1, pp. 383–389, Feb. 2021, doi: 10.11591/eei.v10i1.1974.
[20] 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, Dec. 2014, doi: 10.1109/TIFS.2014.2359646.
[21] T. M. Schneider and C.-C. Carbon, "Taking the perfect selfie: Investigating the impact of perspective on the perception of higher cognitive variables," *Frontiers in Psychology*, vol. 8, Jun. 2017, doi: 10.3389/fpsyg.2017.00971.
[22] K. Slimani, M. Kas, Y. El Merabet, Y. Ruichek, and R. Messoussi, "Local feature extraction based facial emotion recognition: a survey," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, pp. 4080–4092, Aug. 2020, doi: 10.11591/ijece.v10i4.pp4080-4092.
[23] A. A. Moustafa, A. Elnakib, and N. F. F. Areed, "Optimization of deep learning features for age-invariant face recognition," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 2, pp. 1833–1841, Apr. 2020, doi: 10.11591/ijece.v10i2.pp1833-1841.
[24] A. Boukhalfa, A. Abdellaoui, N. Hmina, and H. Chaoui, "LSTM deep learning method for network intrusion detection system," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 3, pp. 3315–3322, Jun. 2020, doi: 10.11591/ijece.v10i3.pp3315-3322.
[25] W. Zhang, M. L. Smith, L. N. Smith, and A. Farooq, "Gender recognition from facial images: two or three dimensions?," *Journal of the Optical Society of America A*, vol. 33, no. 3, pp. 333–344, Mar. 2016, doi: 10.1364/JOSAA.33.000333.
[26] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," *ACM Transactions on Intelligent Systems and Technology*, vol. 7, no. 3, pp. 1–42, Apr. 2016, doi: 10.1145/2845089.
[27] S. Kumar, S. Singh, and J. Kumar, "A study on face recognition techniques with age and gender classification," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, May 2017, pp. 1001–1006, doi: 10.1109/CCAA.2017.8229960.
[28] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2015, pp. 34–42, doi: 10.1109/CVPRW.2015.7301352.
[29] M. Duan, K. Li, C. Yang, and K. Li, "A hybrid deep learning CNN–ELM for age and gender classification," *Neurocomputing*, vol. 275, pp. 448–461, Jan. 2018, doi: 10.1016/j.neucom.2017.08.062.
[30] M. J. Fadhil, M. N. Hawas, and M. A. Naji, "Architecture neural network deep optimizing based on self organizing feature map algorithm," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 9, no. 6, pp. 2538–2546, Dec. 2020, doi: 10.11591/eei.v9i6.1935.
[31] Y. Akbulut, A. Sengur, and S. Ekici, "Gender recognition from face images with deep learning," in *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, Sep. 2017, pp. 1–4, doi: 10.1109/IDAP.2017.8090181.
[32] S. Bekhet and A. Ahmed, "An integrated signature-based framework for efficient visual similarity detection and measurement in video shots," *ACM Transactions on Information Systems*, vol. 36, no. 4, pp. 1–38, Oct. 2018, doi: 10.1145/3190784.
[33] S. Bekhet, M. H. Alkinani, R. Tabares-Soto, and M. Hassaballah, "An efficient method for Covid-19 detection using light weight convolutional neural network," *Computers, Materials and Continua*, vol. 69, no. 2, pp. 2475–2491, 2021, doi: 10.32604/cmc.2021.018514.
[34] P. C. Nissimagoudar, A. V Nandi, A. Patil, and G. H. M., "AlertNet: Deep convolutional-recurrent neural network model for driving alertness detection," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3529–3538, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3529-3538.
[35] H. Asil and J. Bagherzadeh, "A new approach to image classification based on a deep multiclass AdaBoosting ensemble," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, pp. 4872–4880, Oct. 2020, doi: 10.11591/ijece.v10i5.pp4872-4880.
[36] N. Azida Muhammad, A. Ab Nasir, Z. Ibrahim, and N. Sabri, "Evaluation of CNN, Alexnet and GoogleNet for fruit recognition," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 12, no. 2, pp. 468–475, Nov. 2018, doi: 10.11591/ijeecs.v12.i2.pp468-475.
[37] D. A. Jasm, M. M. Hamad, and A. T. Hussein Alrawi, "Deep image mining for convolution neural network," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 20, no. 1, pp. 347–352, Oct. 2020, doi: 10.11591/ijeecs.v20.i1.pp347-352.
[38] S. Bekhet and A. M. Alghamdi, "A comparative study for video classification techniques using direct features matching, machine learning, and deep learning," *Journal of Southwest Jiaotong University*, vol. 56, no. 4, pp. 745-757, 2021, doi: 10.35741/issn.0258-2724.56.4.63.
[39] D. Mohammad, I. Aljarrah, and M. Jarrah, "Searching surveillance video contents using convolutional neural network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 2, pp. 1656–1665, Apr. 2021, doi: 10.11591/ijece.v11i2.pp1656-1665.
[40] "Teachable machine." https://teachablemachine.withgoogle.com/. (Accessed: May 2021)
[41] "A flexible and efficient library for deep learning," *MXNET CNN library*, 2021. https://mxnet.apache.org/. (Accessed: May 2021)
[42] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Illumination-robust face recognition based on deep convolutional neural networks architectures," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 18, no. 2, pp. 1015–1027, May 2020, doi: 10.11591/ijeecs.v18.i2.pp1015-1027.

BIOGRAPHIES OF AUTHORS

Saddam Bekhet (BSc'05, MSc'10, PhD'16) is an Assistant Professor at South Valley University, Qena, Egypt. Received his PhD degree in Computer Science at University of Lincoln, UK in 2016, MSc (2010), BSc (2005) in Computer Science from the University of Assuit, Egypt. Saddam's current research focus on content-based video/image analysis, human identification, soft-biometrics, pattern recognition and deep learning applications for visual data. He can be contacted at email: saddam.bekhet@svu.edu.eg.

Abdullah M Alghamdi (BSc'02, MSc'09, PhD'17) is an Assistant Professor at Imam Abdulrahman bin Faisal University (IAU), Dammam, Saudi Arabia. Received his Ph.D. (2017) degree in E-learning Information Technologies from the University of Lincoln, UK, MSc (2009) from SJU, USA in Computer information Science 2009, BSc (2002) in Education in Computer from IAU, Dammam, Saudi Arabia. Alghamdi's current research focuses on human-computer-interaction, human identification, deep learning, ICT in education, User experience, and acceptance. He can be contacted at email: amghamdi@iau.edu.sa.

Islam Taj-Eddin (PhD'07) is a Lecturer at the Information Technology department, Faculty of Computers and Information, Assiut University. He received his Ph.D., from the City University of New York. He has published over two dozen refereed research papers related to Information Technology, E-learning, Web-Based Education, Technology for special needs users, Software Engineering and Algorithms. He can be contacted at email: itajeddin@aun.edu.eg.