

Research Article Infrared Thermal Image Gender Classifier Based on the Deep ResNet Model

Alyaa J. Jalil ¹ and Naglaa M. Reda²

¹Department of Computer Science, Faculty of Computer Science and Information Technology, Basrah University, Basrah, Iraq ²Department of Mathematics, Faculty of Science, Ain Shams University, Cairo, Egypt

Correspondence should be addressed to Alyaa J. Jalil; alyaa.jalil@uobasrah.edu.iq

Received 3 March 2022; Revised 3 June 2022; Accepted 9 June 2022; Published 8 July 2022

Academic Editor: Francesco Bellotti

Copyright © 2022 Alyaa J. Jalil and Naglaa M. Reda. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gender classification from human face images has attracted researchers over the past decade. It has great impact in different fields including defense, human-computer interaction, surveillance industry, and mobile applications. Many methods and techniques have been proposed depending on clear digital images and complex feature extraction preprocessing. However, most recent critical real systems use thermal cameras. This paper has the novelty of utilizing thermal images in gender classification. It proposes a unique approach called IRT_ResNet that adopts residual network (ResNet) model with different layer configurations: 18, 50, and 101. Two different datasets of thermal images have been leveraged to train and test these models. The proposed approach has been compared with convolutional neural network (CNN), principal component analysis (PCA), local binary pattern (LBP), and scale invariant feature transform (SIFT). The experimental results show that the proposed model has higher overall classification accuracy, precision, and F-score compared to the other techniques.

1. Introduction

Extracting humans' traits and personalities automatically has attracted researchers over decades. It helps in different life fields. The eruption of Internet and smart devices allowed developers and engineers to embed different sensors around the users. For example, smartphones and smart watches are equipped with different sensors, such as gyroscopes, acidometers, cameras, and temperature sensors. The output of these sensors can reveal different hidden information on users' trails and habits [1].

It has been shown that human face reveals traits, ethnicity, gender, age, and feelings [2]. It is a challenging task for computer vision researchers. Gender detection from facial images has applications in surveillance and humancomputer interaction systems (HCI). Human faces supply serious visual information for perceiving gender [3]. This information can be utilized in different fields, such as online marking and advertisement [4], user authentication [5], security surveillance, language converting, online family protection, and image searching [6]. In these systems, a camera is required to capture human faces and analyze them for extracting useful information. This operation can be performed online and offline depending on the computational capability of the devices [7]. To start the analysis, face detection algorithms are required to extract faces from videos or images. Many techniques and algorithms have been proposed for face detection process [4, 8]. Subsequently, the output of this step is utilized for further analysis to extract, classify, or predict different information from these faces. Nevertheless, many issues encounter the imaging process, such as nighttime that converts the imaging devices' output into low quality images without detecting information from it. In addition, illumination in the daytime impacts the detection process. Thermal imaging has emerged to tackle these issues.

Two main types of thermal imaging have spread in the past few years: near infrared imaging (NIR) and far infrared or thermal imaging (FIR). In these types, the camera device can detect and record the thermal distribution over objects that produce heat, such as human body. Subsequently, this distribution is recorded according to its value with different colors in the image. This technology has been utilized for face detection and recognition [9, 10]. It has also shown a great potential for face detection and recognition in nighttime and dark environment [11]. However, it is possible to extract other traits and personalities from faces detected in these images. Such information enhances the security and surveillance applications of NIR cameras.

In this work, a new face thermal image gender classifier (IRT_ResNet) is proposed. The model utilizes ResNet 101 CNN that consists of 101 layers. 3366 thermal images have been leveraged for training and testing purposes. The model has been compared to normal CNN, PCA, local binary pattern, and scale invariant feature transform. IRT_ResNet model overcomes these models in the accuracy of gender classification of faces in thermal images.

The rest of this paper is organized as follows: In Section 2, related work that has been conducted on gender classification and age detection in images utilizing machine learning algorithms and techniques is overviewed. Section 3 introduces the IRT_ResNet proposed model. Section 4 overviews the experiments that have been conducted and the comparison results. The paper conclusion and future suggestions are given in Section 5.

2. Related Work

Gender detection and recognition from face images have gained great interest from researchers over the past decade. The eruption in machine learning algorithms and their applications in image processing improved the detection of different human properties and personal traits from images [12–14]. One of the important traits is the gender of the faces detected in the images [15]. To classify the faces into their genders, many machine learning methods and techniques have been leveraged. The following subsection overviews these techniques.

2.1. Machine Learning in Gender Classification. Using supervised machine learning algorithms in image classification can be divided into two main classes: feature extraction as preprocessing and raw data usage for classification. In the first class, the developer and the designer are required to extract features from the face images and to feed them into a machine learning classifier. For example, the authors in [16] utilized a combination of shifted filter responses (COSFIRE) [17] to extract features from points of interest in the face images. In their method, a collection of Gabor filters was used. The outputs of the COSFIRE trained filters were fed into a SVM model for classification. GENDER-FERET dataset [18] has been used. Approximately 470 images for training and testing have been utilized. An accuracy of 93.7% has been reported. In [19], the authors proposed an agegender classifier that can be trained with a small number of images. The proposed method extracted textures and shape information from faces based on canny edge detection method. Different areas of the face, such as nose and mouth,

have been subsequently shaped with the edge detection. Neural network model has been trained for the classification process. An accuracy of 94% has been reported. In [20], the authors extracted several features from face images, such as rectangular features, local binary pattern, and wavelet coefficients. After that, AdaBoost [21] classifier was trained and compared to SVM and PCA algorithms. A combination of three datasets has been used to obtain a total of 4245 images. An accuracy of more than 99% has been reached.

Wavelet and local binary pattern have been also utilized for feature extraction in [22]. Minimum distance classifier has been trained on the FERET dataset, and an accuracy that acceded 99% has been reported. In [23], the authors combined fuzzy rules with face shapes and textures to train and create their model. An accuracy of 85% has been reported on FERET dataset. In [24], the author extracted geometrical features from the images and utilized these features with PCA of the facial extracted features. Subsequently, a near neighbor classifier has been trained to reduce the complexity of the system. The author attempted to predict the ages and classify the gender of the faces in the images. In [25], facial features have been utilized for age and gender classification. The lips in each image were extracted and utilized for the classification process. A multistage SVM model has been trained with the extracted features to classify the image into child, adult, and old classes. In [26], the authors attempted to segment the faces into six different segments: hair, background, lips, eyes, skin, and mouth. Subsequently, probability maps were assigned to these segments. These maps have been leveraged as features to train a random forest classifier. The authors trained the model and tested it on four different public datasets: Adience, LFW [27], FEI [28], and FERET. The accuracy reported on these datasets was 91.4, 93.9, 93.7, and 100, respectively. All proposed methods in this class require an image preprocessing step, feature extraction and selection. This process can fail with glasses and hats that can cover facial features.

In the second class of gender classification based on machine learning algorithms, features are not extracted. Face images are fed into the classifiers as raw pixel data. Subsequently, the classifier model attempts to extract features in its layers. Moreover, the raw data diminutions are reduced in each layer of the model. In [29], the authors generated a method to classify and detect the gender of faces in images utilizing CNN. Five different layers of convolution filters, pooling, and flattering have been implemented. A public dataset, UTKFace [30], has been used for training, testing, and validation. 16K images and 2K images have been used for training and testing, respectively. Images has been leveraged to change the face orientation in the images to reduce the overfitting impact. An accuracy of 90% has been reported. In [31], CNN has been utilized with three convolutional layers, each followed by a rectifying and pooling layer. In the end, two fully connected networks with 512 nodes have been used for age detection and gender classification. The Adience dataset has been leveraged for the training and testing of the module. 26K images were contained in the dataset of more than 2K humans. The CNN module with this sum of images requires massive computation for training. To train it, Amazon GPUs have been used with more than 1.5K cores. The model obtained 87% accuracy in gender classification. This method generated a complex model that requires massive computing for the training process. Another method that generates a complex model and requires massive computing for the training process is introduced in [32]. However, after training the model, no farther image preprocessing is required.

Recently, human identification in smartphone applications plays an important role in different perspectives such as login permission and sign-up certificates. So accurate gender classification algorithms may increase the accuracy of smartphone applications and reduce their complexity. In [33], the researchers proposed new approach of rotation invariant for classifying genders depending on human face images, based on improved local binary pattern (ILBP). That is because the disadvantage of LBP, extracting spatial structure information and local contrast. ILBP solves some factors such as the sensitivity of noise and rotation, in addition to the low discriminative features by using a modern theory for binary patterns categorization. The feature vector is extracted for images based on ILBP. After that, the classifier of Kullback-Leibler divergence is used for classifying gender.

2.2. Gender Classification from Thermal Images. Thermal infrared images capture the temperature distribution over the muscles and vessels over the human body. This distribution of temperature can be utilized as a facial feature in face images to detect faces, classify genders, and detect other trait personalities of the humans [34]. To classify the gender in thermal infrared images, machine learning can be leveraged as in the normal face images. However, the features in these images are harder to extract and to classify. A comparative study of Haar wavelet and local binary pattern for facial texture feature extraction of thermal face images has been carried out in [35]. The thermal images have been preprocessed and cropped to reduce their sizes [36]. Subsequently, a vector of wavelet coefficients has been extracted and combined with the local binary pattern of the image pixels. The output of this process has been fed into two machine learning classifiers: an artificial neural network model and a minimum distance classifier. An accuracy of 95% has been recorded for this method. In [37], the authors attempted to detect texture facial features of thermal images based on AdaBoost and Haar algorithms. Subsequently, complex Gaussian distribution has been utilized to detect the relation between these features. Results have shown that facial features can be detected easily in thermal images. Reference [38] proposed a machine learning algorithm for face detection in thermal infrared images. The method leverages Haar for feature extraction. In [39], a hybrid model of face gender classification based on normal and thermal images has been proposed. Features from normal images and temperature texture have been combined to create feature vector for the classifier. In [40], a combination of visible normal face images and thermal images has been used for gender classification. These methods were complex since two types of images were required.

Although, natural gender classification is used in many different real applications, thermal cameras are now widely used due to their limitless potentials. Current critical systems operate on thermal images for surveillance, skin temperature screening, security, and military applications [40, 41]. To reduce feature extraction and selection complexity, CNN has been implemented on thermal infrared images recently. Reference [42] applied CNN for gender classification on RGB-D-T dataset. Three experimental scenarios have been implemented. CNN accuracy has exceeded that of LBP, HOG, and moment invariants. In [43], CNN with thermal images has been used for liveness of faces detected in images. CNN has been compared to neural network and SVM. The authors of [5] leveraged thermal infrared images with CNN algorithm for security authentication applications. They claimed that the proposed algorithm overcomes other authentication algorithms in dark places.

This work differs from the above surveyed works in two main aspects. First, the proposed method utilizes thermal infrared images from different sources with variable sizes. Second, it utilizes ResNet 101 CNN that consists of 101 layers to enhance the feature extraction process. For the best of our knowledge, this is the first work that utilizes ResNet model in the area of thermal images. ResNet is the winner in ImageNet 2015 image recognition competition [44] and becomes a breakthrough in image processing since.

3. The Proposed Model

The proposed model (IRT_ResNet) adopts ResNet deep convolutional neural network for infrared thermal images. In neural network, with the addition of new layers to the model, the accuracy of the model increases. However, with the addition of new layers, the training process or the backpropagation step to train the model becomes harder and the accuracy befits saturated or degrading. ResNet tackles this issue by the skipping connection process between the stacked convolutional layers in the model. Figure 1(a) shows a normal stacked layer of CNN, where the output of a layer is the input of the next layer, as a connected network. However, in Figure 1(b), it can be recognized that the process inputs to the next stacked layer which is the summation of the output of the first convolutional layer and the real input of the old layer. The skipping connection step reduces the impact of vanishing gradient issue that reduces the accuracy of the network [45]. Moreover, this allows stacking a number of layers in CNN models and reduces the training time and computation. The block diagram of the proposed model is depicted in Figure 2.

3.1. The Convolutional Layer. The most important component of any convolutional neural network architecture is the convolutional layer [46]. It contains a set of convolutional filters (also called kernels); to get output feature map, it must get convolved with the image (N-dimensional metrics).

Kernel is a grid of discrete numbers or values; each value represents the kernel's weight. All these weights are assigned random values in the start of the training process of a CNN



FIGURE 1: CNN and ResNet comparison.



FIGURE 2: Block diagram of the IRT_ResNet proposed model.

model. Then, in each training period, the weights are modified, and the kernel is learned for extracting significative features. Convolution operation is used to understand the CNN input. In other classical neural networks, the input is in a vector format, whereas in CNN, the input is a multichannel image (e.g., three channels in RGB image, single channel for grayscale image).

The following example explains how the feature map is constructed using the convolution operation. Let the input be an image of 4×4 dimension (Figure 3(a)) and a 2×2 kernel with randomly initialized weights (Figure 3(b)). Then, the convolution operation takes the kernel and slides it all over the image (4×4) horizontally as well as vertically. In the method, the dot product between the input image and the kernel is taken by multiplying the corresponding values of them and then summing up all these values to generate one scalar value in the output feature map. This process stops when the kernel can no longer slide further.

Figure 4 illustrates the stages of the process more clearly. The 2×2 kernel values (shown in light blue color) are multiplied by those in the same-sized region (shown in yellow color) within the 4×4 input image. The resulting values are summed up to obtain a corresponding entry (shown in deep blue) in the output feature map at each convolution step.

And the final output feature map after completing nine stages of convolution operation will be as follows: .fx1

3.2. The Batch Normalization Layer. In order to reduce the sensitivity to network initialization and speeding up the convolutional neural network training, batch normalization

Advances in Human-Computer Interaction

1	0	-2	1	0	1
-1	0	1	2	-1	2
0	2	1	0		
1	0	0	1		
(a)				(1))

FIGURE 3: The input image with the kernel: (a) 4×4 image; (b) kernel of size 2×2 .



FIGURE 4: The first 4 stages of convolution operation.

layers are used between convolutional layers and nonlinearities, as in [47]. The batch normalization operation normalizes the elements x_i of the input by first calculating the mean μ_B and the variance σ_B^2 over the space, time, and observation dimensions for each channel independently. Then, it calculates the normalized activation as follows:

$$\widehat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \dot{\epsilon}}},\tag{1}$$

where ϵ is a constant that improves numerical stability when the variance is very small.

To allow for the possibility that inputs with zero mean and unit variance are not optimal for the operations that follow batch normalization, the batch normalization operation further shifts and scales the activation using the transformation:

$$y_i = \widehat{\gamma} \cdot x_i + \beta, \tag{2}$$

where the offset β and the scale factor γ are learnable parameters that are updated during network training.

3.3. Rectified Linear Unit Layer. A Rectified Linear Unit (ReLU) layer performs a threshold operation for each element of the input, where any value less than zero is set to zero.

This operation is evaluated by the following formula:

$$f(x)\begin{cases} x, & x > 0\\ 0, & x \le 0. \end{cases}$$
(3)

The proposed ResNet 101 model consists of 101 layers, as shown in Figure 5. Each layer consists of two convolution filters stacked together. Subsequently, a skipping connection is added after two layers. Max pooling is added in the first layer, and average pooling is added in the last layer. A fully connected network is utilized at the end of these 101 layers.



FIGURE 5: ResNet 101 architecture.





Parameter	#images	#male images	#female images	Images' size
D1	459	314	145	256 × 336
D2	2907	2483	424	256 × 336



FIGURE 7: Dataset males and females' samples (from D1 and D2).



FIGURE 8: Training time comparison between the three networks.

It worth mentioning that the final fully connected stage has two main binary outputs to classify the input images as males and females. Figure 6 presents a flowchart for illustrating the processes of both the training and the recognition stages.

4. Experiments and Results

To train and test the proposed IRT_ResNet classifier for thermal images, infrared thermal image dataset was required. Many researchers attempted to create thermal image datasets for machine learning applications [46]. However, in this work, two datasets have been utilized to evaluate the accuracy of the proposed model. The first dataset (D1) is found in [48] that contains 461 images. The second dataset (D2) is a larger one that consists of approximately 2907 thermal images and can be found in [49]. The description of the used datasets is shown in Table 1. Figure 7 shows samples of male and female images of both datasets. Three different networks of IRT_ResNet have been constructed and trained on these datasets. All the networks consist of the same steps in each layer; however, the number of layers differs. The first network consists of 18 layers, the second has 50 layers, and the third contains 101 layers.

In the preprocessing phase, we examine each image manually and exclude unsuitable ones. Then, we classify the

N 11	Accuracy					
Model name	D1 dat	taset	D2 da	taset	Both da Men Women Average Men Women Average Men Women	atasets
	Men	94%	Men	100%	Men	98%
IRT_ResNet 18	Women	88%	Women	82%	Women	92.9%
_	Average	91%	Average	91%	Average	95.6%
IRT_ResNet 50	Men	100%	Men	100%	Men	94.7%
	Women	94%	Women	94%	Women	96.4%
	Average	97%	Average	100% Men 82% Women 91% Average 100% Men 94% Women 97% Average 100% Men 94% Women	Average	95.6%
	Men	100%	Men	100%	Men	100%
IRT_ResNet 101	Women	94%	Women	94%	Women	98.2%
	Average	97%	Average	97%	Average	99%

TABLE 2: The IRT_ResNet models' accuracy comparison.

TABLE 3: The Performance measures of the proposed IRT_ResNet models.

Net name	Metrics	Data 1 ratio (%)	Data 2 ratio (%)	All data ratio (%)
	Precision	96	92	96.6
IDT DecNet 19	Recall	94	Data 2 ratio (%) 92 99 95.7 95.6 97 98.6 97.8 97.8 97.8 99 94.3 99 94.3 96.6 96.7	93.8
IKI_Kesivet 16	F-score	95.3	95.7	95.2
	Overall accuracy	95.4	95.6	95.2
	Precision	96	97	94.5
IPT DecNet 50	F-score95.395.7Overall accuracy95.495.6Precision9697Recall8798.6F-score91.597.8Overall accuracy91.997.8	96		
IKI_Kesivet 50	F-score	91.5	97.8	95.3
	Overall accuracy	91.9	Data 2 ratio (%) 92 99 95.7 95.6 97 98.6 97.8 97.8 97.8 99 94.3 96.6 96.7	95.2
	Precision	85	99	97.3
IRT RecNet 101	Recall	97.7	Partice (%) Data 2 Fails (%) 96 92 94 99 95.3 95.7 95.4 95.6 96 97 87 98.6 91.5 97.8 91.9 97.8 85 99 97.7 94.3 90.9 96.6 90.2 96.7	97
IKI_Kesivet 101	F-score	90.9	96.6	97.2
	Overall accuracy	90.2	96.7	97.2

TABLE 4: Comparisons of different techniques.

Model	Male (%)	Female (%)	All (%)
IRT_ResNet 101	100	94	97
PCA	73	69	70
CNN	60	60	60
Local binary pattern	56	52	53
Scale invariant feature transform	61	61	61

selected images as males and females. Finally, we assign a unique identifier to each image, using odd numbers for females and even numbers for males in order to facilitate the recognition stage.

For the three networks, 10% of dataset images have been devoted for the recognition stage. The other 90% of images have been entered into the models to create a dataset of features that are used in both training and validating stages. These images have been divided into 40% for training and 60% for testing. MATLAB has been used for coding the three versions of the proposed IRT_ResNet. The training time and the accuracy of these models have been recorded for each dataset separately.

Moreover, other different models have been implemented to compare the output with the proposed IRT_ResNet model, CNN model with five layers, and three normal neural network models with feature selection based on PCA, scale invariant feature transform, and local binary pattern feature extraction algorithms. These algorithms have shown high accuracy in feature extraction of visual normal images as mentioned in the "Related Work." The neural network model used for classification is a fully connected network with approximately 86K input features, one hidden layer, and an output layer of one neuron.

Our performance experiments go through two phases. In the first phase, the three IRT_ResNet models are compared in terms of their accuracy and training time on both datasets. In the second phase, a comparison between IRT_ResNet and the other three models is made in terms of their accuracy. The following subsections describe these two experimental phases.

4.1. *IRT_ResNet Performance Measure.* Figure 8 shows the time required to train three IRT_ResNet models. As an observation, with the increase of the layers in the models, the training time increases. This is due to the increase of the number of variables that require tuning for each new added layer. In addition, it can be noticed that the training time increases with the addition of more training data since the training loops' size will grow.



Female



Female

Male FIGURE 9: Samples of IRT_ResNet classification results.

Table 2 shows the accuracy comparison between the three studied models. It can be seen from the table that with the addition of more layers from 50 to 101, the accuracy of the model rises to 99%. However, the ratio of average accuracy is almost the same when increasing the model layers from 18 to 50. This motivated us to select ResNet 101 as the main model for the gender classification in this work. Finally, enlarging the size of the training data further enhanced the accuracy of the proposed model.

Furthermore, four other performance metrics, namely, labeled precision rate, recall rate, F-score, and overall accuracy, have been applied. They are evaluated by the following formulas [33, 47]:

$$Precision = \frac{true \text{ positive}}{true \text{ positive} - \text{ false positive}},$$
 (4)

$$Recall = \frac{true \text{ positive}}{true \text{ positive} - false negative},$$
 (5)

$$F - \text{score} = 2 \cdot \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}},$$
(6)

Overallaccuracy

Table 3 presents the results of the metrics for the IRT ResNet 18, 50, and 101.

4.2. Other Models' Comparison. IRT_ResNet has been compared with other models which have shown high accuracy in feature extraction of visual normal images as surveyed in the "Related Work." These models are the CNN model with five layers and three normal neural network models with feature selection based on PCA, scale invariant feature transform (SIFT), and local binary pattern (LBP) feature extraction algorithm. For unifying the experimental environment and the used programming language, we have also coded all of the compared methods mentioned above.

Table 4 shows the classification comparison of these techniques for males and females' images. From the

recorded results, it is clear that the IRT ResNet model obtained 100% accuracy for male classification and 94.11% for female classification, which overcomes the accuracy of the other classifiers. The reason that the females' classifier accuracy is less than that of the male's classifier is that the face's structure of some males with long hair and other traits may look like that of the females even for human eyes. Figure 9 presents examples of the classification process of the proposed model, with their results according to the ambiguity of the input images.

5. Conclusion

Utilizing thermal images in gender classification is a new direction in computer vision research. In this paper, an infrared imaging gender classifier called IRT_ResNet has been proposed. Three models with different number of convolutional neural network layers have been programed and tested. These models consist of different numbers of 2D convolutional filtering layers (18, 50, and 101) with the same structure. Two different datasets have been leveraged in this work. The first dataset consists of 461 infrared thermal images. The second dataset consists of 2907 images. Both datasets have been utilized to train the models. The comparison between the three models has shown that the classification accuracy increases from 96% to 99% when raising the number of layers from 18 to 101. However, no enhancement has been recorded when the number of lavers increased from 18 to 50. We conclude that ResNet 101 is sufficient for the classification process. In addition, four other efficient models of different machine learning classifiers with feature extraction preprocessing have been coded, trained, tested, and compared to the IRT_ResNet classifier. The results prove that the proposed model outperforms others and is more accurate for males than for females, reaching 100%, while its precision and F-score exceed 97%. Another conclusion is that IRT_ResNet achieves the same accuracy percentage for both used datasets regardless of their different sizes.

In the future, we will attempt to utilize IRT_ResNet model for age detection of faces in infrared thermal images to be employed in security and surveillance applications in dark and nighttime. Furthermore, it is planned to study the effect of increasing the size of tested datasets. Moreover, infrared imaging exertions can be added to upgrade the imaging capabilities of smartphone cameras to this type of imaging.

Data Availability

The data used to support the findings of this study were supplied by (1) IR can be made freely available and can be accessed from Index of/downloads/TD_IR_E (http://tdface.ece. tufts.edu/downloads/TD_IR_E/ and (2) Sciebo can be made freely available and can be accessed from IThermalfaceDatabase (https://rwth-aachen.sciebo.de/s/AoSNdkGBRCtWIzX).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- M. Masoud, Y. Jaradat, M. Ahmad, and I. Jannoud, "Sensors of smart devices in the Internet of everything (IoE) era: big opportunities and massive doubts," *Journal of Sensors*, vol. 2019, Article ID 6514520, 26 pages, 2019.
- [2] Z. Yang and H. Ai, "Demographic classification with local binary patterns," in *International Conference on Biometrics*, pp. 464–473, Springer, Berlin, Germany, 2007.
- [3] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, "Fundamental concepts of convolutional neural network," *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, Springer, Berlin, Germany, 2020.
- [4] E. Hjelmås and B. K. Low, "Face detection: a survey," Computer Vision and Image Understanding, vol. 83, no. 3, pp. 236–274, 2001.
- [5] M. Sayed and F. Baker, "Thermal face authentication with convolutional neural network," *Journal of Computer Science*, vol. 14, no. 12, pp. 1627–1637, 2018.
- [6] N. Kumar, P. Belhumeur, and S. Nayar, "Facetracer: a search engine for large collections of images with faces," in *European Conference on Computer Vision*, pp. 340–353, Springer, Berlin, Germany, 2008.
- [7] W. Zhang, L. Xu, P. Duan, W. Gong, Q. Lu, and S. Yang, "A video cloud platform combing online and offline cloud computing technologies," *Personal and Ubiquitous Computing*, vol. 19, no. 7, pp. 1099–1110, 2015.
- [8] H. Hatem, Z. Beiji, and R. Majeed, "A survey of feature base methods for human face detection," *International Journal of Control and Automation*, vol. 8, no. 5, pp. 61–78, 2015.
- [9] B. Zhang, L. Zhang, D. Zhang, and L. Shen, "Directional binary code with application to PolyU near-infrared face database," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2337–2344, 2010.
- [10] S. Z. Li, R. Chu, S. C. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–639, 2007.
- [11] T. Bourlai, N. Kalka, D. Cao et al., "Ascertaining human identity in night environments," in *Distributed Video Sensor Networks*, pp. 451–467, Springer, Berlin, Germany, 2011.
- [12] M. Yu, X. Tang, Y. Lin et al., "An eye detection method based on convolutional neural networks and support vector machines," *Intelligent Data Analysis*, vol. 22, no. 2, pp. 345–362, 2018.

- [13] C. Ventura, D. Masip, and A. Lapedriza, "Interpreting CNN models for apparent personality trait regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, 2017.
- [14] A. Gangwar, V. González-Castro, E. Alegre, and E. Fidalgo, "AttM-CNN: attention and metric learning based CNN for pornography, age and child sexual abuse (CSA) detection in images," *Neurocomputing*, vol. 445, no. 1, pp. 81–104, 2021.
- [15] S. A. Khan, M. Nazir, S. Akram, and N. Riaz, "Gender classification using image processing techniques: a survey," in *Proceedings of the 2011 IEEE 14th International Multitopic Conference*, Karachi, Pakistan, 2011.
- [16] G. Azzopardi, A. Greco, and M. Vento, "Gender recognition from face images with trainable COSFIRE filters," in *Proceedings of the 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance*, Colorado Springs, CO, USA, 2016.
- [17] G. Azzopardi, L. Fernandez Robles, E. Alegre, and N. Petkov, "Increased generalization capability of trainable cosfire filters with application to machine vision," in *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR)*, Cancún, México, 2016.
- [18] "Gender Feret," http://mivia.unisa.it/database/gender-feret.
- [19] K. B. R. K. Ramesha, K. B. Raja, K. R. Venugopal, and L. M. Patnaik, "Feature extraction based face recognition, gender and age classification," *International Journal on Computer Science and Engineering*, vol. 2, no. 18, pp. 14–23, 2010.
- [20] R. Verschae, J. Ruiz-del-Solar, and M. Correa, "Gender classification of faces using adaboost Iberoamerican segmentation," *Symmetry*, vol. 11, no. 6, p. 770, 2019.
- [21] R. E. Schapire, "A brief introduction to boosting," International Joint Conference on Artificial Intelligence, vol. 99, pp. 1401–1406, 1999.
- [22] I. Ullah, M. Hussain, H. Aboalsamh, G. Muhammad, A. M. Mirza, and B. George, "Gender recognition from face images with dyadic wavelet transform and local binary pattern," in *International Symposium on Visual Computing*, pp. 409–419, Springer, Berlin, Germany, 2012.
- [23] P. Moallem and B. Somayeh Mousavi, "Gender classification by fuzzy inference system," *International Journal of Advanced Robotic Systems*, vol. 10, 2013.
- [24] H. H. K. Tin, "Gender and age estimation based on facial images," International Journal: ACTA TECJNICA NAPO-CENSIS Electronics and Telecommunications, vol. 52, 2011.
- [25] S. Ghosh and S. Bandyopadhyay, "Gender classification and age detection based on human facial features using multi-class SVM," *British Journal of Applied Science & Technology*, vol. 10, no. 4, pp. 1–15, 2015.
- [26] K. Khan, M. Attique, I. Syed, and A. Gul, "Automatic gender classification through face segmentation," *Symmetry*, vol. 11, no. 6, p. 770, 2019.
- [27] 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, pp. 7–49, Technical Report; University of Massachusetts, Amherst, MA, USA, 2007.
- [28] Centro Universitario da FEI, "FEI face database," 2019, http:// www.fei.edu.br/~cet/facedatabase.html.
- [29] R. R. Nair, R. Madhavankutty, and S. Nema, "Automated detection of gender from face images," *International Research Journal of Engineering and Technology*, vol. 6, 2019.
- [30] A. Das, A. Dantcheva, and F. Bremond, "Mitigating bias in gender, age and ethnicity classification: a multi-task convolution neural network approach," in *Proceedings of the*

European Conference on Computer Vision, Munich, Germany, 2018.

- [31] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Boston, MA, USA, 2015.
- [32] E. Shervan Fekri, "Gender classification in human face images for smart phone applications based on local texture information and evaluated kullback-leibler divergence," *Traitement du Signal*, vol. 36, no. 6, pp. 507–514, 2019.
- [33] A. Berg, Detection and Tracking in Thermal Infrared Imagery, Linköping University, Sweden, 2016.
- [34] M. Kopaczka, J. Nestler, and D. Merhof, "Face detection in thermal infrared images: a comparison of algorithm-and machine-learning-based approaches," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 518–529, Springer, Berlin, Germany, 2017.
- [35] A. Seal, S. Ganguly, D. Bhattacharjee, M. Nasipuri, and D. Kumar Basu, "A comparative study of human thermal face recognition based on haar wavelet transform (HWT) and local binary pattern (LBP)," 2013, https://arxiv.org/abs/1309. 1009.
- [36] E. Mostafa, R. Hammoud, A. Ali, and A. Farag, "Face recognition in low resolution thermal images," *Computer Vision and Image Understanding*, vol. 117, no. 12, pp. 1689–1694, 2013.
- [37] R. F. Ribeiro, J. Maria Fernandes, and A. J. R. Neves, "Face detection on infrared thermal image," in *Proceedings of the SIGNAL 2017 the Second International Conference on Advances in Signal, Image and Video Processing*, pp. 38–42, Barcelona, Spain, 2017.
- [38] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, "Face recognition by fusing thermal infrared and visible imagery," *Image and Vision Computing*, vol. 24, no. 7, pp. 727–742, 2006.
- [39] D. T. Nguyen and K. Park, "Body-based gender recognition using images from visible and thermal cameras," *Sensors*, vol. 16, no. 2, p. 156, 2016.
- [40] Z. Wu, M. Peng, and T. Chen, "Thermal face recognition using convolutional neural network," in 2016 International Conference on Optoelectronics and Image Processing, Warsaw, 2016.
- [41] J. Seo and I. J. Chung, "Face liveness detection using thermal face-CNN with external knowledge," *Symmetry*, vol. 11, no. 3, p. 360, 2019.
- [42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016, https://arxiv.org/abs/1512. 03385.
- [43] Y. Hu, A. Huber, J. Anumula, and S.-C. Liu, "Overcoming the vanishing gradient problem in plain recurrent networks," 2018, https://arxiv.org/abs/1801.06105.
- [44] Y. Cimtay and G. N. Yilmaz, "Gender classification from eye images by using pretrained convolutional neural networks," *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, vol. 14, pp. 39–44, 2021.
 [45] S. Ioffe and C. Szegedy, "Batch normalization: accelerating
- [45] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, https://arxiv.org/abs/1502.03167.
- [46] M. Kowalski and A. Grudzień, "High-resolution thermal face dataset for face and expression recognition," *Metrology and Measurement Systems*, vol. 25, no. 2, 2018.
- [47] 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.

- [48] "Thermal image dataset," http://TDface.eca.tufts.edu/ download/, 2018.
- [49] "Second thermal images dataset," http://rwth-aachen-scibo. de/s/AoSNdkGBRtWlzX, 2020.