Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 26 (2023), No. 1, pp. 117–131
DOI : 10.47974/JSMS-952

Attendance monitoring of masked faces using ResNext-101

Sushil Kumar Mahapatra *
Binod Kumar Pattanayak †
Department of Computer Science & Engineering
Siksha 'O' Anusandhan (Deemed to be Univesity)
Bhubaneswar
India

Bibudhendu Pati §
Department of Computer Science
Ramadevi Women's University
Bhubaneswar
India

Abstract

SARC virus, Coronavirus, Ebola and bird flu have all caused pandemics in the last few decades. Most of these diseases spread through the air when someone coughs, sneezes or even talks. The government makes citizens wear masks. Furthermore, all academic activities are conducted in virtual mode as a result of this predicament, making taking attendance of pupils difficult while they are wearing masks on their faces. To overcome this issue, the proposed work will track a student's attendance while using ResNext-101 in a virtual classroom setting. ResNext-101, a deep learning technique, is used on masked faces in this work and it is a good model for accurately detecting masked faces. By using the Gaussian data augmentation approach, the outcome reveals a level of accuracy of 51.70 percent with a loss of 1.9452.

Subject Classification: *68T07, 68U10.*

Keywords: *Attendance monitoring, Masked face, Resnet-50, ResNext-101, Teaching learning method, Deep learning.*

1. Introduction

Due to technological advancements, the security of any organization's gadgets is at risk of fraud, tampering and hacking among other

**E-mail:* `mohapatrasushil@gmail.com (Corresponding Author)`
† *E-mail:* `binodpattanayak@soa.ac.in`
§ *E-mail:* `patibibudhendu@gmail.com`

things[1], [2]. Several authentication procedures are employed to check the authenticity of the user to avoid this problem. Radio Frequency Identification (RFID), fingerprint, facial recognition and iris recognition are examples of authentication technologies[3]. Face recognition and iris recognition, together with machine learning and deep learning algorithms have been demonstrated to be the most accurate in terms of accuracy due to their exceptional level of authenticity[4]. Face recognition algorithms based on deep learning algorithms such as ResNet and Facenet have been presented by various academics in recent years[5], [6]. Some researchers have also developed a different method for detecting faces. However, all of these studies are based on full-face photographs. Because there are more features on a full face image, the accuracy level of face recognition is higher[7]. However, if a person wears a mask all of the time due to a pandemic-like circumstance, it is extremely difficult to track the identity of a person with limited features because the face is partly hidden[8], [9].

In the education system, attendance is a very vital aspect of the teaching-learning process. A genuine attendance leads to a proper effective learning system and hence can enhance the student's academic outcome. Several types of research have been proposed to enhance academic outcomes by using different teaching-learning processes[10], [11]. Apart from the traditional teaching-learning process, nowadays education system transforms itself from a regular mode of class to a virtual model of the class by using different Internet of Things (IoT) based gadgets[12]. But the institutions as well as the instructor faces many difficulties while conducting the online classes such as false attendance, false network issue, slow internet connection and so on. Many types of research are proposed to sort out these problems utilizing IoT as it can enhance the tracking capability utilizing different sensors and different software-based solutions. In the virtual model of class, generally, the students are logged in to a specified learning management system and their attendance is recorded while the class is in progress without the intervention of the instructor using a face recognition system installed in the Learning Management System.

Students may or may not wear a mask in virtual class mode. The institution or the person teaching the class cannot force the pupils to wear or not wear the mask in this case. As a result, developing a model that can recognize faces with or without a face mask is a difficult undertaking[13]. The model that has been shown to work best for full-face photographs may not work as well for masked images. According to prior research, a deep neural network requires more features to achieve a higher level of accuracy[14]. However, masked faces with a masked mouth, nose and some elements of the cheek have a limited number of features with which to train the model[15]. Furthermore, several aspects of the forehead are obscured when a person wears a hat or a headscarf. As a result, this suggested effort employs ResNext-101, a deep learning-based architecture, to address such circumstances and challenges[16]. We also used the transfer learning method to train the model for faster convergence and accuracy in this study. In the training phase, the settings are fine-tuned on the faces without the mask. The model is then put to the test with masked faces.

1.1 *Contribution*

To detect the masked faces in this suggested work, a deep learning approach called ResNext-101 is used. Both unmasked and masked faces are used to test the suggested model. Because there is only one masked face dataset available with a small number of masked faces, this suggested model employs a data augmentation strategy using image flipping and the inclusion of Gaussian noise to increase accuracy and reduce loss[17].

The rest of the paper is laid out as follows. In the related work area, an in-depth assessment of the various facial recognition techniques has been undertaken. The materials and procedures used to recognize the masked faces are discussed in the following section. Following that, the outcome is presented, followed by a commentary in the following section. Finally, in the concluding section, a conclusion is reached on the proposed model.

2. Related Works

Face recognition techniques have proven to be a dominant security system in recent years. As a result, it attracts an increasing number of researchers to work in this sector. Researchers typically recognize important features of a face and match them to an existing database to determine a person's genuine identity and permit admission to the facility. As a result, the extraction of key features is the most difficult task. The researchers used a variety of algorithms, including Machine Learning (ML) and Deep Learning (DL) techniques, to extract such properties. Different ML and DL approaches for face recognition that have been previously developed are covered in this section.

Principal Component Analysis (PCA) and K-Nearest Neighbors were utilized in the early days of face recognition. For face detection, the highest eigenvalues of the eigenvectors are determined. This can also be used to reduce the size of a dataset by deleting unnecessary data[18]. When the dataset is small and the number of features that may be extracted is small, however, these models fail to attain a reasonable degree of accuracy. Some scholars have used Linear Discriminant Analysis (LDA) for facial recognition [19]. This model is superior since it demands data variance within the same group as well as a substantial variance between these groups. Some researchers have proposed a face recognition model based on Support Vector Machines in a different way (SVM). The hyper-plane is defined in SVM by combining the features of the two faces. The Cambridge OR1 database is employed in their suggested model. Different facial expressions and positions can be found in this database. In the same hierarchy, an adaboost method was applied to improve the performance of the classifier[20]. Adaboost has the advantage of employing weighted training data as well as a weighted voting method to improve the classifier's performance. Recently, considerable research has been carried out to detect blurred faces using the Convolutional Neural Network (CNN) [21]. The researchers the Trunk Branch Ensemble CNN model (TBE-CNN) in their proposed approach to obtain a holistic representation of an image for CNN training. The CNN's ability to learn from obscured insensitive features will be improved as a result of this. A similar strategy, presenting a new state-of-the-art model for huge datasets that used the weak classifier to rank the data that would be delivered to the annotators[22]. Because CNN requires a big training dataset for face recognition, UmmeAiman et. al. devised a method for generating the training dataset by applying Gaussian and Poisson noise to an existing dataset[23]. According to their research, utilizing this network model resulted in a significant increase in accuracy. Facenet, a pretrained network model, is used to recognize faces from a real-time video[24]. In that approach, Fnet is used to transform video to images and then derived euclidean space from the images to distinguish faces. Muhammad Zeeshan Khan and colleagues suggested a CNN-based face recognition system for tracking student attendance in a smart classroom[25]. This data is then transferred through the Internet of Things (IoT) devices. On the LFW dataset, they compared their model to the scale-invariant features transform and speed up robust features. However, all of the prior attempts at facial recognition have relied on full-face photos. Only a few attempts were made with masked faces. To identify the obstructed face, researchers utilized two methods. They're either restringing the occluded part of the face or discarding it from the image [26]. Bagchi et. al. offer a closed point algorithm for registering a person's face and creating a 3D representation from an occluded face[27].

PCA was also utilized to reconstruct the occluded portion. Some authors also proposed a statistical approach for restoring the obscured section from a 3D model[28]. There may be an incorrect reconstruction of the obscured part during the facial image restoration process. As a result, incorrect face recognition occurs. In another approach, the authors have offer an SVM-based discard occlusion procedure in which they divide the entire face image into small areas and use SVM to try to detect the occluded section, after which the remaining part of the face is produced using a mean-based weighted matrix[29]. Hariri et. al., proposed a model for the masked face last year. To extract the features in their model, they used the VGG-16 and BAG of feature algorithms[26]. In their model, the features maps for the cropped eye and forehead are considered using a quantization pooling method in VGG-16.

3. Methods and Materials

In the previous few decades, various approaches for recognizing human faces have been used. However, in a recent pandemic-like circumstance, we were required to wear face masks to avoid contamination. Face recognition as an authentication method, on the other hand, suffers the most because the algorithm is developed for the entire face. When the face is hidden behind a mask, the features are severely reduced and the current face recognition system fails to recognize the faces[8]. In this study, a pre-trained ResNext-101 is utilized to detect and recognize limited-featured face photos with greater accuracy than earlier research in the field.

3.1 *Method*

Nowadays, the transfer learning method is used to shorten training time and attain a high level of accuracy. A model that has been trained and built using another face recognition task is applied to the present model via the transfer learning approach[30]. We utilized ResNext-101 to train our model in this method. Initially, the model is trained using a full-face picture database and the hyper-parameters associated with it are fine-tuned to produce better results. The trained model is then applied to the masked-faced picture database to obtain the desired outcome.

3.2 *Transfer Learning*

In transfer learning, a pre-trained model that was previously used for one task is now used for another[31]. Because the dataset for the masked face is so little that deep learning techniques can't be used, a transfer learning technique can be used, in which the model is built on a big dataset. Because there are more unmasked datasets available, the initial model is built using these datasets and the parameters are adjusted accordingly. Furthermore, the masked faces databases are used to test this pre-trained model.

3.3 *ResNext-101 Architecture*

The ResNext-101 architecture is a straightforward version of the VGG-Nets and ResNets architectures[32], [33]. The transformation approach, in which some processes are aggregated, is used in this proposed architecture to leverage the strategy of the split, transformation and merger[16]. The same topologies are clustered in between residual blocks in his architecture, which share the upper parameters. This architecture includes a new concept, cardinality, which refers to the magnitude of a set of transformations and is symbolized by the letter 'c.' A few factors are necessary due to the inclusion of cardinality. Figure 1 depicts a ResNext-101 block with the cardinality value of 32. The aggregated transformation is described by equation 1.

$$F(x) = \sum_{i=1}^{c} T_i(x) \quad (1)$$

Where $T_i(x)$ is the set of transformation.

In this proposed study, ResNext-101 is utilized to obtain weights from ResNext. MS1M is used to train the ResNext at first and subsequently, VGGFace2 is used to train its parameters.

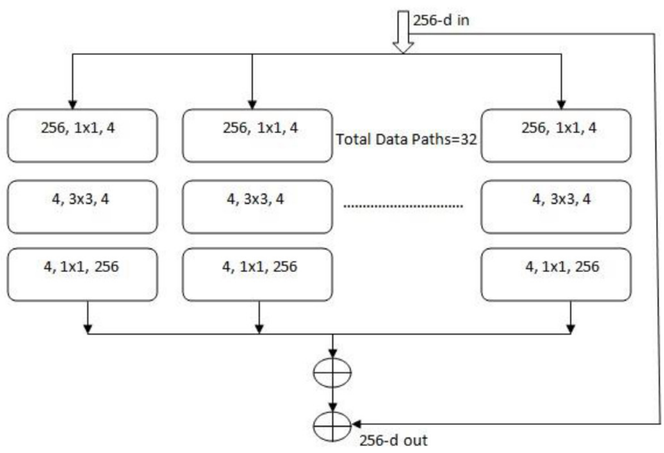

Figure 1

Cardinality 32 ResNext-101 block

3.4 *Model Parameter and Dataset*

This facial recognition attendance system was created using a Jupyter notebook and Pytorch 1.2.0 version library on a Windows 10 operating system. A GP104 GPU is also used.

The real-world masked face recognition dataset (RMFRD) is employed in this suggested study [8]. We chose the RWMF recognition dataset out of the three available in this dataset: Real-world masked face recognition, synthesized masked faces and RWMF verification dataset. Because this dataset only contains 5000 masked faces and approximately 9000 unmasked faces, it is insufficient for training. So, to expand the masked face dataset, we added some Gaussian and Poisson noise to the original dataset to generate new data. As a result, we were able to build a dataset of 12000 masked faces of 1130 people at the end of the process. In addition, we selected photos with at least 12 images of masked and unmasked faces. The complete dataset has been divided into two parts. One is for training, which is kept at 67 percent of the whole dataset and the other is for testing, which is kept at 33 percent. All of the photos in the collection are reduced to a standard size of 128x128 pixels. Furthermore, these photos are randomly flipped throughout the training process to maximize data augmentation.

3.5 *Parameter Tuning*

Different parameter optimizers were examined to adjust the hyper-parameter and it was discovered that SGD with momentum is the best candidate for unmasked faces, while Adam is the best candidate for the masked face dataset in terms of accuracy[34], [35]. The parameter optimization takes into account SGD, SGD with momentum, Adagrad, RMSprop and Adam optimizer. Table 1 shows a comparison of different optimizers in terms of accuracy. Table 2 lists the test parameters for the Masked and Unmasked datasets.

4. Result and Discussion

In this part, performance matrices like the FI score, Precision and Recall are utilized to assess the proposed model's performance. This study also includes a comparison of our model to other models in the same field. The model is first trained on the unmasked dataset, which yields a better result than the masked dataset. Because the unmasked dataset has a larger number of photos, it performs well in terms of accuracy and loss. The result, however, is quite close to other state-of-the-art facial

Table 1

Comparative analysis of optimizer

Optimizer	SGD	SGD with Momentum	Adagrad	RMSprop	Adam
Accuracy	0.9337	0.9364	0.8963	0.9505	0.9771

Table 2

Hyper-parameter for testing

Hyper parameter	Unmasked	Masked
Optimizer	SGD with Momentum	Adam
Epoch	20	25
Batch Size	128	32
Dropout	0.4	0.3
Loss	Cross-Entropy	Cross-Entropy
Learning rate	0.0025	0.0018

recognition models like CNN and ResNet-5[36], [37]. However, comparing these models is unfair because they employed a different dataset for their model, whereas ours is based on the masked faces image dataset. Table 3 lists some of the additional hyper-parameters that were chosen.

A drop-out layer is added to the model to keep it from over fitting, resulting in an accuracy level of 83.7 percent. Table 4 shows the results of the performance matrices.

We used the masked dataset to test our model after it had trained using the aforesaid hyper-parameter and produced an acceptable result. However, the end outcome was not adequate. A better recognition rate could not be produced as the number of features in the faces decreases to

Table 3

Additional hyper-parameter for testing

Hyper parameter	Unmasked
SGD with momentum	0.8
Children for Pretrained Network	14
Decay rate	0.12
Step Size	12

Table 4

Performance metrics of the model using unmasked dataset

Precise Score	0.9162
Recall Score	0.9143
FI Score	0.913

Table 5

Hyper-parameter for masked faces

Hyper-parameter	Masked
Batch Size	32
Learning rate	0.0018
Decay rate	0.1
Step size	14

the occlusion of the face and the number of masked face photos necessary to train the model decreases. So we used Adam to improve our model and the hyper-parameters were fine-tuned once more to achieve a desirable level of accuracy and reduced loss. Table 5 lists the hyper-parameters for the masked faces.

With the above parameter, the model can achieve an accuracy level of 51.73% and a loss of 1.9452. The performance matrices are given in table 6.

4.1 *Comparative Analysis of Masked and Unmasked Dataset*

A comparison of results in terms of loss and accuracy for both masked and unmasked faces is discussed in this section. The unmasked pictures' loss vs. epoch graphs are displayed in Figure 2, while the accuracy vs. epochs graphs are shown in Figure 3. The loss vs. epochs plots indicate that the loss decreases as the number of epoch's increases, indicating that the model is learning from the inputs. The loss curve for masked data,

Table 6

Performance metrics of the model using the masked dataset

Precise Score	0.5164
Recall Score	0.5123
FI Score	0.5048

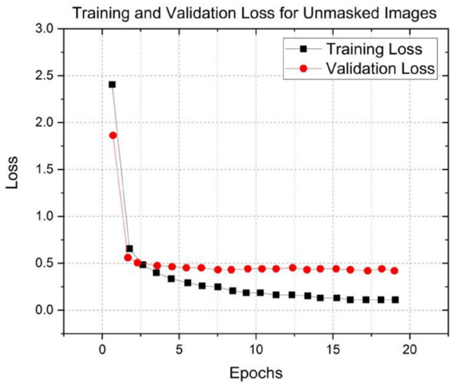

Figure 2
Training loss and validation loss for unmasked image dataset

however, shows that as the number of masked faces decreases, the number of features decreases due to the occluded section of the face and so the loss cannot reach its minimum. Though the validation curve approaches a minimal value, it is still high when compared to the unmasked faces' loss graph. Figure 5 displays a good curve of accuracy vs. epochs for the unmasked picture dataset. However, the accuracy drops to 51.73 percent for the masked faces dataset. As more data is required to train a deep neural network, the accuracy level falls as expected. To improve accuracy, the parameters must be fine-tuned and a larger dataset of masked faces should be included in the model's training. A comparative analysis of the accuracy level is presented in Figure 6 to validate the result.

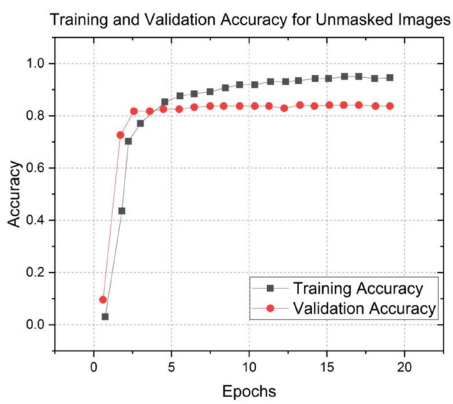

Figure 3
Training accuracy and validation accuracy for unmasked image dataset

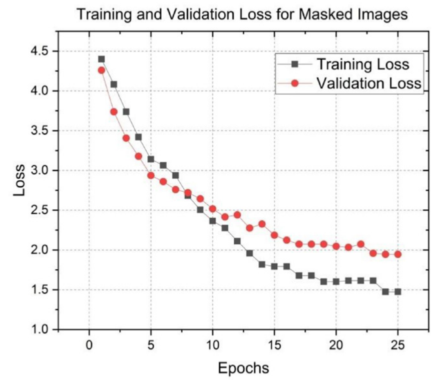

Figure 4

Training loss and validation loss for masked image dataset

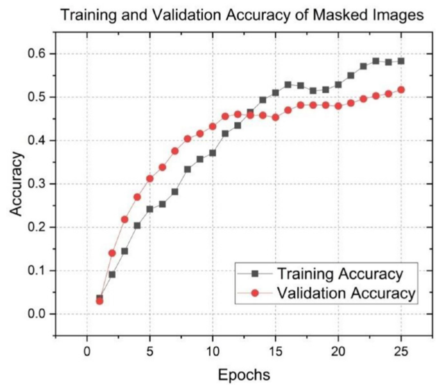

Figure 5

Training accuracy and validation accuracy for masked image dataset

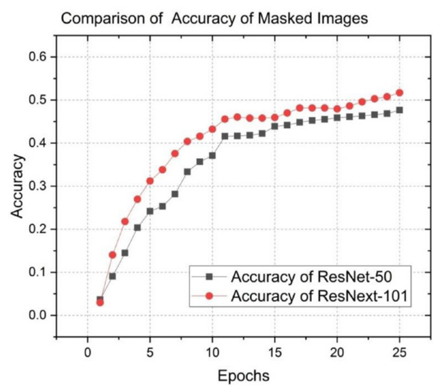

Figure 6

Comparison of accuracy for masked using ResNet-50 and ResNext-101

5. Conclusion

In recent years, airborne infections have been on the rise. Microorganisms that can travel via the air can infect anyone in a matter of minutes. Face masks have been made mandatory for all citizens to prevent the spread of deadly diseases. The current security system, which relies on face recognition, is unable to detect veiled faces. As a result, there is a huge demand for a new face recognition system. The ResNext-101 deep learning algorithm is proposed in this paper for both masked and unmasked face recognition. With a small dataset, the model's outcome indicates a high level of accuracy. This model can also be evaluated with a huge dataset and other data augmentation techniques. For masked faces, several deep learning and machine learning approaches can be utilized to test the capability of them in the future.

References

[1] Robertson, D. J. (2018). Face recognition: Security contexts, super-recognizers and sophisticated fraud. *The Journal of The United States Homeland Defence and Security Information Analysis Center (HDIAC)*, *5*(1), 6–10.

[2] Sahasrabudhe, S., Shaikh, N., & KishoriKasat (2020). Internationalisation of higher education - Necessity to adapt to new forms of engagement for ensuring sustainability?, *Journal of Statistics and Management Systems,* 23:2, 431-444, DOI: 10.1080/09720510.2020.1736328.

[3] Al-Naji, F. H., & Zagrouba, R. (2020). A survey on continuous authentication methods in Internet of Things environment. *Computer Communications*.

[4] De Marsico, M., Galdi, C., Nappi, M., & Riccio, D. (2014). Firme: Face and iris recognition for mobile engagement. *Image and Vision Computing*, *32*(12), 1161–1172.

[5] Jose, E., Greeshma, M., Haridas, M. T., &Supriya, M. H. (2019). Face recognition based surveillance system using facenet and mtcnn on jetson tx2. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 608–613.

[6] Lu, Z., Jiang, X., &Kot, A. (2018). Deep coupled resnet for low-resolution face recognition. *IEEE Signal Processing Letters*, *25*(4), 526–530.

[7] Jain, A. K., & Li, S. Z. (2011). *Handbook of Face Recognition* (Vol. 1). Springer.

[8] Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., & Pei, Y. (2020). Masked face recognition dataset and application. *ArXiv Preprint ArXiv:2003.09093*.

[9] Sahoo, K.K.,Muduli, K.K.,Luhach, A.K., & Poonia, R.C., (2021). Pandemic COVID-19: An empirical analysis of impact on Indian higher education system, *Journal of Statistics and Management Systems*, 24:2, 341-355, DOI: 10.1080/09720510.2021.1875571

[10] Mahapatra, S. K., Pattanayak, B. K., & Pati, B. (2021). Flip Learning: A Novel IoT-Based Learning Initiative. In *Intelligent and Cloud Computing* (pp. 59–67). Springer.

[11] Ramlowat, D. D., & Pattanayak, B. K. (2019). Exploring the internet of things (IoT) in education: A review. *Information Systems Design and Intelligent Applications*, 245–255.

[12] Khan, S., Al-Dmour, A., Bali, V., Rabbani, M.R., & Thirunavukkarasu K. (2021) Cloud computing based futuristic educational model for virtual learning, *Journal of Statistics and Management Systems*, 24:2, 357-385, DOI: 10.1080/09720510.2021.1879468

[13] Ibitoye, O. (2021). A Brief Review of Convolutional Neural Network Techniques for Masked Face Recognition. *2021 IEEE Concurrent Processes Architectures and Embedded Systems Virtual Conference (COPA)*, 1–4.

[14] Mehdipour Ghazi, M., & Kemal Ekenel, H. (2016). A comprehensive analysis of deep learning based representation for face recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 34–41.

[15] Deng, J., Guo, J., An, X., Zhu, Z., & Zafeiriou, S. (2021). Masked face recognition challenge: The insight face track report. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1437–1444.

[16] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1492–1500.

[17] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1–48.

[18] Gottumukkal, R., & Asari, V. K. (2004). An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters*, *25*(4), 429–436.

[19] Lu, J., Plataniotis, K. N., & Venetsanopoulos, A. N. (2005). Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. *Pattern Recognition Letters*, *26*(2), 181–191.

[20] Guo, G., Li, S. Z., & Chan, K. (2000). Face recognition by support vector machines. Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), 196–201.

[21] Freund, Y., Iyer, R., Schapire, R. E., & Singer, Y. (2003). An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, *4*(Nov), 933–969.

[22] Ding, C., & Tao, D. (2017). Trunk-branch ensemble convolutional neural networks for video-based face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 1002–1014.

[23] Aiman, U., &Vishwakarma, V. P. (2017). Face recognition using modified deep learning neural network. *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–5.

[24] F. Schroff, D. Kalenichenko and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[25] Khan, M. Z., Harous, S., Hassan, S. U., Khan, M. U. G., Iqbal, R., & Mumtaz, S. (2019). Deep unified model for face recognition based on convolution neural network and edge computing. *IEEE Access*, *7*, 72622–72633.

[26] Hariri, W., & Farah, N. (2020). Efficient Graph-based Kernel using Covariance Descriptors for 3D Facial Expression Classification. *Proceedings of the 1st International Conference on Intelligent Systems and Pattern Recognition*, 7–11.

[27] Bagchi, P., Bhattacharjee, D., & Nasipuri, M. (2014). Robust 3D face recognition in presence of pose and partial occlusions or missing parts. *ArXiv Preprint ArXiv:1408.3709*.

[28] Drira, H., Amor, B. B., Srivastava, A., Daoudi, M., & Slama, R. (2013). 3D face recognition under expressions, occlusions and pose variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(9), 2270–2283.

[29] Priya, G. N., & Banu, R. W. (2014). Occlusion invariant face recognition using mean based weight matrix and support vector machine. *Sadhana*, *39*(2), 303–315.

[30] Torrey, L., & Shavlik, J. (2010). Transfer learning. In Handbook of research on machine learning applications and trends: Algorithms, methods and techniques (pp. 242–264). IGI global.

[31] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359.

[32] Jun, H., Shuai, L., Jinming, S., Yue, L., Jingwei, W., &Peng, J. (2018). Facial expression recognition based on VGGNet convolutional neural network. *2018 Chinese Automation Congress (CAC)*, 4146–4151.

[33] Targ, S., Almeida, D., & Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. *ArXiv Preprint ArXiv:1603.08029*.

[34] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*.

[35] Liu, C., & Belkin, M. (2018). Accelerating sgd with momentum for over-parameterized learning. *ArXiv Preprint ArXiv:1810.13395*.

[36] Mandal, B., Okeukwu, A., & Theis, Y. (2021). Masked Face Recognition using ResNet-50. *ArXiv Preprint ArXiv:2104.08997*.

[37] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). *Deep Face Recognition*.