



## RESEARCH ARTICLE - ENGINEERING

### Generating and Improving a Dataset of Masked Faces Using Data Augmentation

Waleed Ayad<sup>1\*</sup>, Siraj Qays<sup>1</sup>, Ali Al-Naji<sup>2</sup>

<sup>1</sup> Electrical Engineering Technical College, Middle Technical University, Baghdad, Iraq

<sup>2</sup> School of Engineering, University of South Australia, Mawson Lakes, SA 5095, Australia

\* Corresponding author E-mail: [bbc0054@mtu.edu.iq](mailto:bbc0054@mtu.edu.iq)

| Article Info.   | Abstract  |
|---|---|
| <p><i>Article history:</i></p> <p>Received<br/>23 December 2022</p> <p>Accepted<br/>12 March 2023</p> <p>Publishing<br/>30 June 2023</p>                                  | <p>Before the spread of the COVID-19 virus in 2020, modern face recognition systems performed excellently, but then the wearing of masks was imposed by countries on their population, which led to a noteworthy decrease in the discriminatory ability of those systems, where they had been trained on large-scale datasets of unmasked faces and not available large-scale masked faces datasets that time. To contribute to addressing the shortage of large-scale data sets that consist of people wearing masks, a developed method has been presented to create simulated masks and overlay them on faces in two main steps. The first step was to detect, align and crop the faces of unmasked faces datasets in a dataset and then apply simulated masks on the faces utilizing the dlib-ml library. This method was used to generate a dataset for masked faces (CASIA-mask). The second step used five techniques of data augmentation with the generated dataset. To evaluate the masked dataset and data augmentation, an accuracy of 96.4% was achieved by training one of the latest and most important facial recognition systems, FaceNet, on the masked dataset. The same system also achieved excellent results of 97.71% when trained on CASIA-mask and data augmentation together.</p> |
| <p>This is an open-access article under the CC BY 4.0 license (<a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>)</p> |   |
| <p>Publisher: Middle Technical University</p>   |   |
| <p><b>Keywords:</b> COVID-19; Masked Face Dataset; Data Augmentation; Face Recognition.</p>   |   |

## 1. Introduction

The state-of-art and most well-known systems for face recognition, clustering, or verification [1–5] recorded interesting results in the last years. Datasets, the size of which has heightened exponentially in previous years, are utilized for training these facial recognition systems. Practically, the number of images in face recognition datasets for diversified persons heightened to thousands in 2007, such as the database released in 2007 by Huang et al. [6] called Labeled Faces in the Wild (LFW), which includes 13,233 face images. Then, in 2014, the number of dataset images up to a few hundred thousand images, such as the dataset proposed by Yi et al. called CASIA-WebFace [7], which include 494,414 images. In 2015, Zivew et al. proposed a database named CelebA containing 200K images for 10K people. After that, the number of images of datasets doubled dramatically, reaching millions of images in 2021, such as the WebFace42M dataset [9], which consists of 42M images, and the WebFace260M, which includes 260M images. Generally, the databases mentioned have been released in the previous decades, and those interested in this field made great efforts to gather and clean them. Some face recognition models can work very well, and they could even detect errors if found in the annotations (see [1]). However, these models have been tested and trained by unmasked facial databases such as the LFW dataset [6] and the CASIA-WebFace dataset [7]. In early 2020, the Covid-19 virus propagation in most countries prompted those countries to impose health restrictions on the population. Wearing a mask on the face is considered one of the most critical health restrictions. It hides most of the main facial features, as it hides the cheeks, nose, and mouth. Hiding many basic facial features due to wearing a mask led to a remarkable decrease in discriminatory ability. There are currently many types of facial masks, all of which hide the facial features behind them, and they are diverse in color, fabric, texture, and shape (see Fig. 1).

Remarkably, the success of the latest face recognition algorithms works mainly depends on training recent complex deep learning models for face recognition [1–5] on huge face recognition datasets such as CASIA-WebFace [7] and CelebA [8]. However, most data sets of masked faces [10-13] that were proposed after the spread of the COVID-19 pandemic are smaller compared to the databases consisting of faces of people not wearing masks. Currently, the largest large-scale datasets consisting of images of people wearing masks are [11,12], [14] to our knowledge.

In this paper, the lack of datasets consisting of images of people wearing masks is addressed by employing an automatic method proposed where images of people wearing masks are created from images of unmasked people by overlaying a simulated mask on the unmasked faces. In this method, there are several types of face masks (14 masks) of diverse shapes, colors, and textures have been utilized (see Fig. 1). Where these masks are used with the proposed method by randomly choosing one of them and placing it on the unmasked face in the image. The proposed method can be used to generate a dataset of masked faces from all datasets of unmasked faces. Specifically, the proposed method has been used to generate a dataset of masked face images from images of unmasked faces from the CASIA-WebFace dataset [7].

| Nomenclature & Symbols |                                    |       |                                    |
|------------------------|------------------------------------|-------|------------------------------------|
| Dlib-ml                | Dlib Machine Learning Library      | ROI   | Region Of Interest                 |
| LFW                    | Labelled Faces in the Wild Dataset | RMFRD | Masked Facial Recognition Data Set |

After the spread of the Covid-19 virus, there has been an urgent need to release or collect datasets consisting of images of people wearing masks for training the modern facial recognition systems [1–5] on visible facial features only because some facial features have been covered after wearing the mask, where covering parts of the face with the mask led to a remarkable reduction of the discriminatory ability of the systems recognizing faces that were trained on databases of people who do not wear masks.

The proposed method was used for generating a developed dataset of masked faces, which can be utilized for training the latest face recognition systems [1-5], as it was used to train a face recognition model called FaceNet [1]. Practically, the FaceNet [1] facial recognition system achieved good performance in recognizing the faces of people wearing masks compared to previous works [10], [12], [14] when trained on a dataset of faces of masked persons developed using the proposed method. There was an additional improvement in the performance of the face recognition system FaceNet [1] when it was trained using the proposed developed dataset for masked people with data augmentation.

In this paper, there are three essential contributions are the following:

- Proposing a fully automatic approach to generate images of people wearing masks from images of unmasked people by overlying simulated masks on their faces using Dlib-Machine Learning Library Dlib-ml [15,16].
- Developing a dataset of people wearing masking from a dataset consisting of unmasked people images.
- Showing an improvement in the performance of the latest face recognition systems in recognizing the faces of people wearing masks when trained on the proposed developed masked faces dataset.
- Showing better improvement in modern face recognition systems when training them on the proposed developed masked faces data set and data augmentation.

The presentation of this paper is organized as follows: Section 2 presents the related work. Section 3 describes the masked Face Dataset and data augmentation. Section 4 shows the experiment and results. Finally, the conclusion is provided in Section 5.



Fig. 1. Data set of mask images for this research contains fourteen different masks

## 2. Related Work

Since the beginning of 2020, the Covid-19 virus has spread in most countries of the world, which prompted most countries of the world to impose the wearing of masks to limit the spread of the epidemic. Wearing a mask led to a decrease in the discriminatory ability to recognize masked faces because the modern face recognition systems were trained on data sets consisting of unmasked persons. Then the researchers collected data sets consisting of real images of people wearing masks [10], [12, 13]. However, its problem is the small number of images in it compared to data sets of unmasked faces such as [7,8], as there is hardness in gathering a real database of large-scale masked faces due to the shortage of images of masked people for different races, genders, ages, and type of masks.

The authors of [12] were among the first to release a dataset that contains real images of people wearing masks on their faces named the Masked Facial Recognition Data set (RMFRD), where they gathered 5,000 real images of 525 people wearing masks and 90,000 real images for the same people but without a mask. A study by Anwar et al. [2] presented a dataset of real images collected from websites called the MFR2 data set, which includes 269 images of 53 people, each of whom had photos with a mask and images without a mask at an average of five images. Deng et al. [11] released a test data set consisting of 6,964 real images of people wearing masks and 13,928 real images of people without masks for 5,964 persons. This dataset is not yet available due to legal issues with the data. Zhu et al. [13] proposed a test data set consisting of 60,926 images of people with and without masks but containing only 3,211 images of people wearing masks for 862 people.

The latest facial recognition systems need to be trained on data sets consisting of hundreds of thousands or even millions of images of faces, but the data sets of masked faces that have been collected have been of insufficient size to train face recognition systems due to their small size compared with the databases of unmasked faces. Researchers have made many efforts to find a solution to this problem. Currently, the generation of simulated masks on people's faces in images is considered one of the most successful and easiest available solutions, as this method was used to generate a data set consisting of images of people wearing masks from a data set of unmasked faces.

The organizers of [12] utilized an algorithm based mainly on the Dlib-ml library [15,16] to place simulated masks on the faces of unmasked people. Where they released a dataset for masked faces named SMFRD consisting of 500,000 images of masked faces for 10,000 persons, where they used their approach to generate simulated masks on two datasets (LFW [6] and CASIA-WebFace [7]). The work of Anwar et al. [10] introduced an open-source tool named Mask-The-Face to set simulated masks on the faces of people who are not wearing masks. Mask-The-Face is based on the facial feature detector from the dlib-ml library for estimation of the angle of inclination of a person's face and identifying six basic features to placing the simulated mask on it. The organizers of [17] released a dataset of artificial occluded face recognition named Webface\_OCC containing 804,704 images for 10,575 people. The Webface\_OCC data set was gathered by collecting two data sets, the first is original CASIA-WebFace for unmasked faces, and the second is CASIA-WebFace for artificially covered faces. The CASIA-WebFace dataset of artificially occluded faces is obtained by placing special obstructive objects, such as glasses and masks, on the faces of unmasked persons in the original CASIA-WebFace dataset of uncovered faces. The work of Ding et al. [18] introduced a data augmentation approach to train the data, whereby they used the proposed approach to automatically generate simulated masks on unmasked faces in the LFW dataset. Their study applied an algorithm named Delaunay-triangulation that was used to detect diverse features in the face and place masks depending on the detected places on the face. The authors in [14] presented an approach for developing facial datasets consisting of people wearing masks where they created simulated masks and placed them on the faces of people not wearing masks in the images. Their proposed approach is based on a program called SparkAR-Studio, a program developed by programmers at Facebook, which they used to put simulated masks on the faces of unmasked people in images. Using their proposed approach, they generated two data sets of masked faces, the first consisting of 445,446 images generated from the CASIA-WebFace data set of unmasked faces and the second consisting of 196,254 images generated from the CelebA data set of unmasked faces.

In the fourth section of this paper, the method that has been proposed to generate masks on faces for creating a masked data set and the method of data augmentation for the created masked dataset has been compared with the rest of the other methods [10], [12], [14] in terms of accuracy of verification and testing. This is the primary factor that discriminates the proposed approach from the rest of the competitors.

### 3. Masked Face Dataset and Data Augmentation

In this paper, a method has been presented to generate simulated masks on the faces of people who did not wear masks in the images. Using this method, a developed dataset called CASIA-mask has been generated, which is a data set of people wearing masks, after which the FaceNet [1] facial recognition system has been trained on the proposed dataset CASIA-mask. Then, data augmentation was used with the proposed masked dataset together to train the FaceNet face recognition system [1]. The proposed approaches were used to generate the proposed data set, and then data augmentation was used with the generated data set, as detailed in this section.

#### 3.1. Dataset

Most of the data sets that consist of faces of people who do not wear masks can be used in applying the proposed method to generate data sets of masked faces. In this paper, a large-scale and freely available unmasked face dataset called CASIA-WebFace containing 494,414 images of 10,575 persons was selected to generate the dataset of images of masked faces.

#### 3.2. Dataset generation

In this paper, a data set of masked faces was generated from a data set of unmasked faces. There were several steps taken to generate a data set consisting of the faces of people wearing simulated masks called CASIA-mask, as illustrated in Fig. 2. In the beginning, if the face is existing in the image, it will be detected, aligned and cropped using Multitask Cascading Convolutional Neural Networks (MTCNN) [19], which detect five main points in the human face: the right and left eyes, the mouth, and the right and left corners of the nose from the images in the CASIA-WebFace data set [7] for unmasked faces. As a result of using MTCNN to detect, align and crop existing images from the original CASIA-WebFace dataset, a dataset named CASIA-WebFace-Aligned was generated consisting of 471,669 images of 10,567 people (about 96% of the images in the CASIA-WebFace dataset).

The final step was to place simulation masks on the faces of people in the images of the CASIA-WebFace-Aligned dataset, where the simulation mask was placed on the areas of the face that needed to be covered to generate the proposed masked dataset. Therefore, there is a need to detect more details of the face, as the five areas of the face that MTCNN detects are not sufficient, and there is a need to discover more details in the face. As a result of this, the dlib-ml library was used in this step for face detection due to its ability to detect many details of the face compared to MTCNN. The dlib-ml library can detect 68 basic points in the human face, as shown in Fig. 3, which was used to accurately determine the location of the simulation mask. A specific range (48 to 68) of the 68 key points of the dlib-ml library was selected as a region of interest (ROI) [20–23]. Finally, one of the fourteen masks was randomly selected and resized to fit over the chosen ROI of the person's face. The proposed masked dataset (CASIA-mask) generated following the mentioned steps consisted of 418,978 images from 10,567 persons because some faces from the CASIA-WebFace-Aligned dataset could not be detected using the dlib-ml library.

#### 3.3. Data augmentation

Data augmentation is the methods and techniques that are applied to increase the quantity of data by adding slightly modified copies of the original data or artificial data that is generated from the original data. It can help in enhancing the generalization ability of the network and prevent overfitting. Data augmentation can increase facial recognition systems' accuracy by training them on the original data set images and the images that were modified using data augmentation techniques [24,25]. In this paper, five types of data augmentation techniques have been used: flip [26], brightness [27], crop [28], rotate [29], and Impulse Noise [30], as shown in Fig. 4. About the flip, only the horizontal flip was used randomly, the brightness was used with a random range between 70% to 130%, and the crop was used with a random area of 15×15. The rotation was used with a randomly changing angle between 60 and -60 degrees, and the Impulse Noise (Pepper Noise) was used with a threshold of 240.

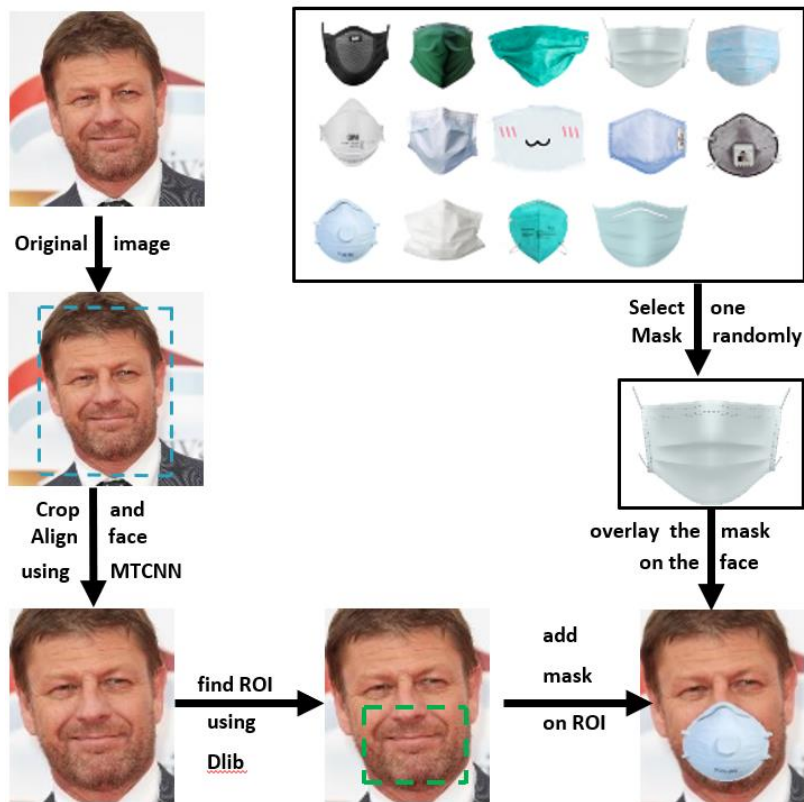


Fig. 2. The steps for placing simulated masks on the faces of unmasked people in photos to generate images containing people wearing simulated masks

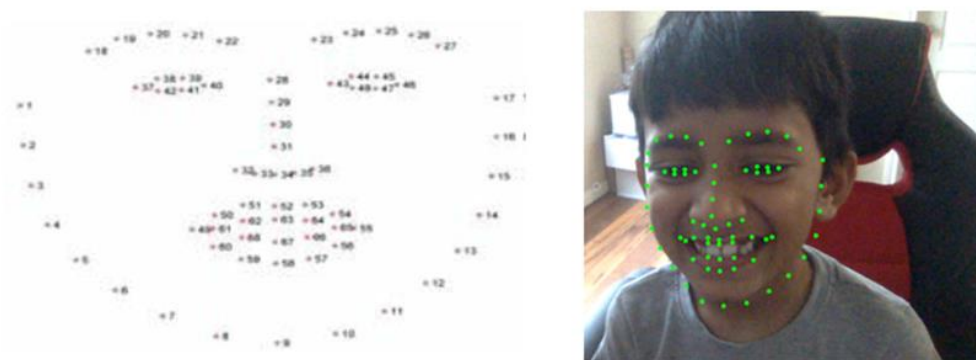


Fig. 3. Detect 68 basic points in the human face using dlib-ml library

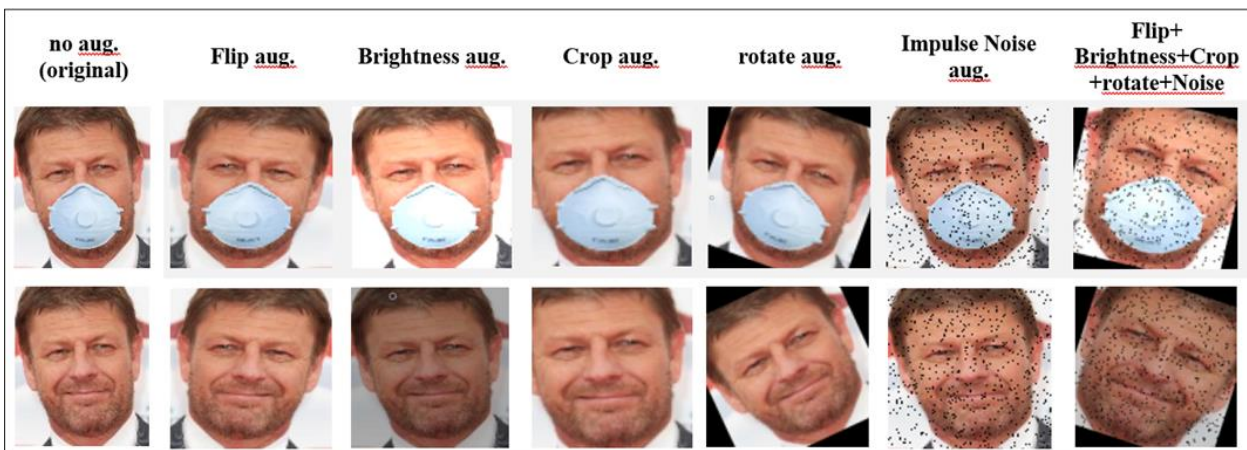


Fig. 4. Samples of images before and after applying the data augmentation techniques used, where the first row is for an image from the unmasked faces dataset (CASIA-WebFace), and the second is for an image from the proposed masked dataset

## 4. Results and Discussion

### 4.1. Experimental setups

This paper used an end-to-end open-source machine learning platform named TensorFlow to implement the deep learning network and trained it on the large-scale dataset. In practice, the GPU version of TensorFlow was utilized for training on large-scale datasets with 100 epochs of training. The experiments were performed on the Windows 10 OS with 2.30GHz Intel Core i7-11800H and Nvidia GeForce RTX 3070 laptop GPU.

### 4.2. Training and verification dataset & data augmentation

The generated CASIA-mask dataset containing 418,978 images of 10,567 persons was combined with the generated CASIA-WebFace-Aligned data set consisting of 471,669 images of 10,567 unmasked persons as a unified training dataset so that the total number of images was 890,647 images. For evaluation, the LFW dataset [6] was applied for validation testing, where the results were reported by utilizing the standard experiential protocol from view 2, which consisted of ten subsets, each subset containing 300 face pairs of the same matches and 300 face pairs of non-same matches. Then the previous steps were reperformed, but five types of data augmentation were used together, including flip, brightness, crop, rotate, and impulse noise (as shown in the last column of Fig. 4) on the images in the two datasets CASIA-mask and CASIA-WebFace-Aligned. Where, at each epoch during training, for each image in the two datasets CASIA-mask and CASIA-WebFace-Aligned, a different modified image is generated using the five types of data augmentation together (they are not saved and used only during training), and the modified images that are used It was generated during each training epoch are combined with the two datasets CASIA-mask and CASIA-WebFace-Aligned to form a unified training set. The total number of images was twice that of the two datasets CASIA-mask and CASIA-WebFace-Aligned when combining the modified images resulting from the data augmentation (890,647 images at each training epoch) with the two datasets CASIA-mask and CASIA-WebFace-Aligned, so the total number of images was 1,781,294 images at each training epoch.

### 4.3. Face Recognition System

To assess the efficiency of the proposed masked dataset and data augmentation in increasing the efficiency of face recognition systems trained on it, one of the most important and famous modern face recognition systems, called FaceNet [1], was used. The proposed approach of generating the masked dataset and data augmentation was compared with competing approaches [10], [12], [14] in terms of the accuracy of validation and testing of the generated dataset (masked dataset) and using five techniques of data augmentation together: flip, brightness, crop, rotate, and Impulse Noise (see Table 1). For face recognition model training (FaceNet) [1], inception-ResNet V1 [31] was applied as the backbone network trained with softmax loss, the batch size set to 96 with Adam optimizer, and a learning rate of 0.1 with 100 epochs. Interesting results were achieved compared to previous works [10], [12], [14] when training the FaceNet face recognition system [1] using a masked dataset and CASIA-WebFace-Aligned. The results were then improved using five types of data augmentation with masked dataset and CASIA-WebFace- Aligned to train the FaceNet facial recognition system [1]. The accuracy results are illustrated in Table 1.

Table 1. The accuracy results of the face verification On LFW with Facenet

| Train set  | accuracy |
|--|----------|
| CASIA-WebFace-train+masks [14]                               | 88.06    |
| CelebA-train+masks [14]                                      | 96.22    |
| SMFRD [12]   | 95       |
| MFR2 [10]  | ≈97%     |
| CASIA-mask+ CASIA-WebFace- Aligned (ours)                    | 96.4     |
| CASIA-mask+ CASIA-WebFace- Aligned+ data augmentation (ours) | 97.71    |

Using data augmentation with the two datasets CASIA-mask and CASIA-WebFace-Aligned when training the FaceNet face recognition system [1] enhanced the network's generalization ability, preventing overfitting and increasing the accuracy. However, the use of data augmentation with two datasets, CASIA-mask and CASIA-WebFace-Aligned, led to an increase in the training time spent for each epoch due to the doubling of the number of images when using data augmentation. where the average time spent for training in each epoch with two data sets CASIA-mask and CASIA-WebFace-Aligned is about 120 minutes while the average time spent training in each epoch with the data augmentation and the two datasets CASIA-mask and CASIA-WebFace-Aligned was about 158 minutes.

For the generated dataset, there was a failure to detect the face using MTCNN in some images, which led to the generation of the CASIA-WebFace-Aligned data set with 4% fewer images than the CASIA-WebFace data set. And there was a failure to detect points 48-68 out of 68 base points for the face using the dlib-ml library in some images, which led to the generation of the CASIA-mask dataset for masked faces with some images that are 11% less than the CASIA-WebFace-Aligned dataset, and thus the CASIA-mask dataset for masked faces 15% smaller than the CASIA-WebFace dataset in terms of the number of face images.

## 5. Conclusion

In this paper, a developed method was proposed to create simulated masks and place them on the faces of unmasked people in the images and to generate data sets for masked faces from data sets for unmasked faces. Based on the proposed method, a masked dataset named CASIA-mask containing 418,978 images was generated, which was combined with the generated CASIA-WebFace-Aligned dataset containing 471,669 images. The acceptable results were achieved when they were used to train the FaceNet face recognition system. The results were also significantly improved when five types of data augmentation were used together with two datasets, including masked dataset and CASIA-WebFace-Aligned to train the FaceNet face recognition system.

### Acknowledgement

I would like to express my deepest appreciation to Middle Technical University, Electrical Engineering Technical College-Baghdad, Iraq, for the encouragement and support for disseminating the engineering research paper.

## Reference

- [1] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 07-12-June, pp. 815–823, 2015.
- [2] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep Hypersphere Embedding for Face Recognition," Proc. IEEE Conf. on computer vision and pattern recognition, vol. 47, no. 5, pp. 212–220, 2017.
- [3] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," Deep Learning in Visual Computing, 2015.
- [4] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 1, pp. 4685–4694, June 2019.
- [5] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 1701–1708, 2014, doi: 10.1109/CVPR.2014.220.
- [6] G. B. Huang, M. Mattar, T. Berg, E. L. Labeled, R. Images, and E. Learned-miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," work. faces in'Real-Life'Images Detect. alignment, Recognit., 2008.
- [7] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning Face Representation from Scratch," arXiv Prepr. arXiv preprint arXiv:1411.7923, Nov 2014.
- [8] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," IEEE Int. Conf. on Computer Vision, vol. 2015 Inter, pp. 3730–3738, 2015.
- [9] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du, J. Zhou, "WebFace260M: A Benchmark Unveiling the Power of Million-Scale Deep Face Recognition," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 10487–10497, 2021.
- [10] A. Anwar and A. Raychowdhury, "Masked Face Recognition for Secure Authentication," arXiv Prepr. arXiv2008.11104, pp. 1–8, August 2020.
- [11] J. Deng, J. Guo, X. An, Z. Zhu, and S. Zafeiriou, "Masked Face Recognition Challenge: The InsightFace Track Report," Proc. IEEE Int. Conf. Comput. Vis., pp. 1437–1444, October 2021.
- [12] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, Wu, H., P. Yi, K. Jiang, N. Wang, Y. Pei, and H. Chen, "Masked Face Recognition Dataset and Application," arXiv Prepr. arXiv2003.09093, pp. 1–3, March 2020.
- [13] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Guo, J. Lu, and D. Du, "Masked Face Recognition Challenge: The WebFace260M Track Report," arXiv Prepr. arXiv2108.07189, August 2021.
- [14] T. Mare et al., "A realistic approach to generate masked faces applied on two novel masked face recognition data sets," arXiv Prepr. arXiv2109.01745, October 2021.
- [15] M. Asif and D. Kamthania, "Face Detection and Recognition with Mask," IEEE Int. Conf. Autom. Logist. ICAL, vol. 10, pp. 121–126, October 2021.
- [16] D. E. King, "Dlib-ml: A machine learning toolkit," J. Mach. Learn. Res., vol. 10, pp. 1755–1758, 2009.
- [17] B. Huang, Huang, B., Z. Wang, G. Wang, K. Jiang, K. Zeng, Z. Han, X. Tian, and Y. Yang, "When face recognition meets occlusion: A new benchmark," ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 1, pp. 1–5, Mar 2021.
- [18] F. Ding, P. Peng, Y. Huang, M. Geng, and Y. Tian, "Masked Face Recognition with Latent Part Detection," 28th ACM Int. Conf. Multimed, pp. 2281–2289, October 2020.
- [19] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Process. Lett., vol. 23, pp. 1499–1503, 2016.
- [20] T. H. Huang, K. Y. Cheng, and Y. Y. Chuang, "A collaborative benchmark for region of interest detection algorithms," 2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009, pp. 296–303, 2009.
- [21] Y. Liu, Z. G. Li, and Y. C. Soh, "Region-of-interest based resource allocation for conversational video communication of H.264/AVC," IEEE Trans. Circuits Syst. Video Technol., vol. 18, pp. 134–139, 2008.
- [22] H. Mliki, S. Dammak, and E. Fendri, "An improved multi-scale face detection using convolutional neural network," Signal, Image Video Process., vol. 14, pp. 1345–1353, 2020.
- [23] W. Sun, H. Zhao, and Z. Jin, "A visual attention based ROI detection method for facial expression recognition," Neurocomputing, vol. 296, pp. 12–22, 2018.
- [24] X. Wang, K. Wang, and S. Lian, "A Survey on Face Data Augmentation," arXiv:1904.11685v1 [cs.CV] 26 Apr 2019.
- [25] Z. Pei, H. Xu, Y. Zhang, M. Guo, and Y. Yee-Hong, "Face recognition via deep learning using data augmentation based on orthogonal experiments," Electron., vol. 8, no. 10, pp. 1–16, 2019, doi: 10.3390/electronics8101088.
- [26] H. Ben Fredj, S. Bouguezzi, and C. Souani, "Face recognition in unconstrained environment with CNN," Vis. Comput., vol. 37, no. 2, pp. 217–226, 2021, doi: 10.1007/s00371-020-01794-9.
- [27] V. Uchoa, K. Aires, R. Veras, A. Paiva, and L. Britto, "Data Augmentation for Face Recognition with CNN Transfer Learning," Int. Conf. Syst. Signals, Image Process., vol. 2020-July, pp. 143–148, 2020, doi: 10.1109/IWSSIP48289.2020.9145453.
- [28] M. Wang and W. Deng, "Deep face recognition: A survey," Neurocomputing, vol. 429, pp. 215–244, 2021, doi: 10.1016/j.neucom.2020.10.081.
- [29] X. Peng, Z. Tang, F. Yang, R. S. Feris, and D. Metaxas, "Jointly Optimize Data Augmentation and Network Training: Adversarial Data Augmentation in Human Pose Estimation," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 2226–2234, 2018, doi: 10.1109/CVPR.2018.00237.
- [30] J. H. Lee, M. Z. Zaheer, M. Astrid, and S. I. Lee, "SmoothMix: A simple yet effective data augmentation to train robust classifiers," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2020-June, pp. 3264–3274, 2020, doi: 10.1109/CVPRW50498.2020.00386.
- [31] M. Längkvist, L. Karlsson, and A. Loutfi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," Pattern Recognit. Lett., vol. 42, pp. 11–24, 2014.