The masked face recognition with ArcFace

Zhengyu Li 23020211153947, Zekun Ai 23020211153914*, Linglu He 23320211154220*, Dongmei Ma 23020211153901*, Ying Xu 23320211154256*

School of Informatics, Xiamen University
zhengyu19981001@163.com, {aizekun,helinglu,madongmei,wagmy}@stu.xmu.edu.cn

Abstract

Over the last twenty years, there have seen several outbreaks of different corona-virus diseases across the world. These outbreaks often led to respiratory tract diseases and have proved to be fatal sometimes. Currently, we are facing an elusive health crisis with the emergence of COVID-19 disease of the corona-virus family. One of the modes of transmission of COVID-19 is airborne transmission. This transmission occurs as humans breathe in the droplets released by an infected person through breathing, speaking, singing, coughing, or sneezing. Hence, public health officials have mandated the use of face masks which can reduce disease transmission by 65% . For face recognition programs, commonly used for security verification purposes, the use of face mask presents an arduous challenge since these programs were typically trained with human faces devoid of masks but now due to the onset of Covid-19 pandemic, they are forced to identify faces with masks. Hence, this paper investigates the same problem by developing a deep learning based model capable of accurately identifying people with face-masks. In this paper, the authors train a InceptionResNetV1 based architecture that performs well at recognizing masked faces. The outcome of this study could be seamlessly integrated into existing face recognition programs that are designed to detect faces for security verification purposes.

Introduction

As humans continue to make technological advances, the need to secure devices that house both our private and official matters is paramount. Some of the traditional methods achieved this feat by using ID cards, passwords, passphrases, and puzzles. However, with the rapid advancement in the field of deep learning and high performance computing, the usage of human biometrics such as the face, voice and fingerprints are deemed ubiquitous in the modern day security verification programs.(Minaee et al. 2019) The widespread use of these human biometrics relates to their uniqueness, making it difficult to replicate them. (Han et al. 2004) Similarly, face recognition programs allow a quicker yet efficient framework for identification of an individual.(Petrescu 2019)

From 2020, the virus COVID-19, which has recently spread worldwide, has seriously affected people's daily life. Medical experts around the world pointed out that wearing a mask can effectively prevent the respiratory transmission of the virus (Chu et al. 2020) , so wearing a mask has become a common way for people to protect themselves. However, masks block out a lot of information about a person's face, which presents a new challenge to our already mature face recognition technology. The most recent studies on the effects of the face mask on face recognition accuracy have indicated a significant degradation in the accuracy of these models(Damer et al. 2020) , (Neto et al. 2021) , (Ngan et al. 2020) , (Jeevan et al. 2022) .

The new scenario imposed by the pandemic situation motivated the development of face recognition approaches. Based on InceptionResNetV1, we propose a supervised DCNN(deep convolutional neural network) pipeline with ArcFace loss (Deng et al. 2019) , which is capable of identifying people even when they are wearing masks. As a more complicated and powerful version of sigmoid loss, ArcFace has a clear geometric interpretation due to the exact correspondence to the hypersphere's geodesic distance and can help recognize people out of the training set. In our experiment, the proposed model is trained and test on the Masked dataset, which is generated from MS1MV2 by using the MaskTheFace tool (Anwar and Raychowdhury 2020) .Our contributions can be summarized as follows:

•A new marked dataset is generated based on original face-recognition dataset by using data augmentation.

•Design and train a new masked face recognition model based on InceptionResNetV1 with the dataset created by ourselves, which can recognize static facial images with masks.

Related work

Computer vision is one of the most successful research areas in the field of machine learning, and face recognition is one of the most important applications of computer vision. In 2015, FaceNet (Schroff, Kalenichenko, and Philbin 2015) was proposed based on the GoogleNet-24 framework and achieved an accuracy of 99.63% on the LFW database (Huang et al. 2008) . Parkhietal. (Parkhi, Vedaldi, and Zisserman 2015) worked on the VGGFace and achieved an accuracy of 98.95% on the LFW database. After Microsoft Re-

*These authors contributed equally.

search applied the ResNet (He et al. 2016) to image classification with an excellent performance, a series of architectures were designed relied on ResNet. For example, Caoetal. (Cao et al. 2018) used ResNet50 architecture to access face recognition performance in their work. What's more, similar to our work, InceptionResNetV1 (Szegedy et al. 2017) and SE-ResNeXt-101 (Hu, Shen, and Sun 2018) are two common DCNN models for object or facial recognition. InceptionResNetV1 performed an accuracy of 99.65% on the LFW dataset, and SE-ResNeXt-101 achieved an accuracy of 97.4% on ImageNet dataset (Deng et al. 2009) .

Recognizing faces with occlusion is a variant of the facial recognition problem. Simple face recognition algorithms become limited (Hariri 2021) when the intention is to recognize faces when people wear hats, eyeglasses, masks as well as other objects that can occlude part of the face while leaving others unaffected in the images.

A number of approaches were proposed to identify people when they are wearing sunglasses, scarves and other items. There were few studies done in masked face recognition as these were relatively rare cases until 2020. However, the outbreak of COVID-19 worldwide makes recognition of faces with masks an imminent need to improve the current technology. In the following, we classify a set of works for masked face recognition and briefly discuss the methods.

Dataset

Recently, lots of works focus on generating masked face images and constructing related datasets. Anwar et.al. (Anwar and Raychowdhury 2020) present an opensource tool to mask faces and create a large dataset of masked faces. Wang et.al. (Wang et al. 2020) propose three types of masked face datasets. Du et.al. (Du et al. 2021) use 3D face reconstruction to add the mask templates on the UV texture map of the non-masked face. Cabani et.al. (Eyiokur, Ekenel, and Waibel 2021) provide a fine-grained dataset for detecting people wearing masks and those not wearing masks. However, existing tools usually transform the whole mask to generate masked faces, resulting in unrealistic generation effect to a certain extent. In addition, there lacks a specialized dataset to verify the performance of face recognition models.

Occlusion Recovery Method

Many Occlusions recovery method recover unmasked faces for feature extraction based on Generative Adversarial Network. Anwar and Raychowdhury (Anwar and Raychowdhury 2020) firstly detect mask object and then synthesizes the affected region with fine details while retaining the global coherency of face structure. Ge et.al. (Ge et al. 2020) proposes identity-diversity inpainting to facilitate occluded face recognition. However, the reconstructed faces are synthetic and their reliability depends on the quality of the data, the network and the training process. In addition, the process of removing the mask noticeably increases the computation time.

Occlusion-robust Method

In order to avoid a complicated recovering process, these methods aim to produce direct occlusion-robust face feature embedding from masked face images. Hariri (Hariri 2021) propose a method based on occlusion removal and deep learning-based features from eyes and forehead regions. Li et.al. (Li et al. 2021) integrate a cropping-based approach with Convolution Block Attention Module. Nevertheless, this causes an important drop of information when dealing with unmasked faces, so it is not suitable for applications mixing both use cases.

Combined Method

Combined method tackles the problem training the face recognition network with a combination of masked and unmasked faces (Anwar and Raychowdhury 2020) . They combine the VGG dataset (Schroff, Kalenichenko, and Philbin 2015) with augmented masked faces and train the model following the original pipeline described in FaceNet (Cao et al. 2018) . This way, the model learns to distinguish when a face is wearing a mask and to trust more in the features of the upper half of the face, but still extracts information from the whole face. On the other hand, Geng et.al. (Geng et al. 2020) define two centers for each identity which correspond to the full-face images and the masked face images respectively. They use Domain Constrained Ranking for forcing the feature of masked faces getting closer to its corresponding full-face center and vice-versa. Combined method not only avoid complicated training, but also consider the full-face information. Thus, this paper adopts combined methods to handle this problem.

Method

Problem Definition

We consider the problem of facial recognition of subjects who wear masks.To solve this problem ,we aim at increasing the accuracy of the face recognition network when dealing with masked faces.In this section, we introduce the proposed method based on masked faces.

Dataset Creation

For the method we proposed,there is a need of masked face datasets.Therefore,We need to create our own dataset in the project.

Answar and Raychowdhury present a tool MaskTheFace (aqeelanwar 2020) for masking faces in images. It uses a dlib based face landmarks detector to identify the face tilt and six key features of the face necessary for adjusting and applying a mask template. Based on the face tilt, corresponding mask template is selected from the library of mask. The template mask is then transformed based on the six key features to fit perfectly on the face. MaskTheFace can be used to convert any existing face dataset to masked-face dataset.,and supports different types and color of masks.

In this work, we use the tool MaskTheFace to generate the masked version of the dataset.We add surgical blue,surgical gree,white N95,white KN-95,and black cloth masks.Some examples of the generated faces are shown in Figure 1.

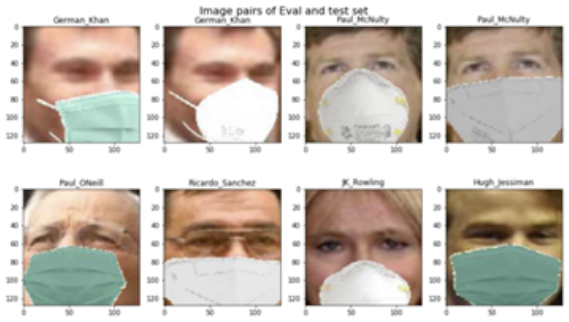

Figure 1: Sample images after applying MaskTheFace to LFW dataset

Masked Face Recognition Model

In this section, we introduce the masked face recognition model,as figure2. The model uses a deep convolutional network InceptionResnetv1 to extract face features and employ the ArcFace (Deng et al. 2019) loss that directly reflects what we want to achieve in face recognition.Namely,we strive for an embedding, from an image into a feature space (Schroff, Kalenichenko, and Philbin 2015) .

InceptionResnetv1 is based on GoogLeNet style Inception models (Szegedy et al. 2015) .The network was designed with computational efficiency and practicality in mind,so that inference can be run on individual devices including even those with limited computational resources,especially with low-memory footprint.Considering the practicability of model and the efficiency of training,we choose this network to extract feature.

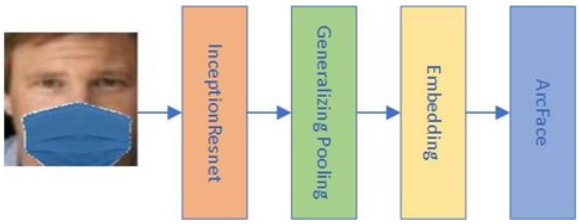

Figure 2: Masked face recognition model

Loss Function

The main challenges in feature learning for face recognition is the design of appropriate loss functions that enhance discriminative power.In this work,We use Arcface (Deng et al. 2019) as the loss function.There are two reasons: it use a softmax-loss-based methodology,which does not require an exhaustive training-data-preparation stage; and it has been proven to be the approach the reports the best results for the original face recognition task.

Experiment

Training details

The backbone architecture to be trained for masked face recognition in this paper is the InceptionResNetV1 (nscozzaro 2019) . The selection of this architecture was motivated by its high performance. Based on the experience of previous related work, we set ArcFace s to 45 and Arcface m to 0.4 in the experiment. Consider GPU memory limitations and model learning efficiency, experiments that trained the backbone from scratch used a mini-batch size of 128. The models in this paper are implemented using Pytorch1.10.0 (Paszke et al. 2019) and ran on one Linux machine with one Nvidia 2080Ti 10 gb GPU. All the experiments were trained with Stochastic Gradient Descent (SGD) with an initial learning rate of 1e-1. The learning rate is decreased by a factor of 10 at 5, 10, 15, 20 iterations. The networks expect 118x118x3 sized images and produce 512-d embeddings, which is a common dimension size for embedding representation in deep learning due to its Highly representative and high computational efficiency. normalization and data augmentation will be done before fitting the data into the model to handle the vanishing gradient and overfitting.

Datasets

In this project, CASIA (Yi et al. 2014) , VGGFace2 (Cao et al. 2018) and LFW (Huang et al. 2012) dataset are used as the raw data in our project with the detailed information illustrated in Table 1.

Dataset	Number of identities	Number of images
CASIA	10,575	494,414
VGGFace2	9,131	3,310,000
LFW	5,749	13,233

Table 1: The raw datasets used

However, the facial images in the original datasets are of low quality and MaskTheFace might fail to detect the faces correctly. Those images are not processed. Nevertheless, as multiple masks are chosen to add to the images, the actual number of output images in our datasets is comparable to or even larger than the original datasets as shown in Table 2. The images from the processed CASIS and VGGFace2 are used for training and the images from the processed LFW will be split into validation and test set.

Dataset	Number of identities	Number of images	Usage
CASIA	10,575	492,832	Training
VGGFace2	8,631	2,024,897	Training
LFW	5,749	64,811	Validation&Test

Table 2: Summary of the final dataset used in the project

Result

Our experiments compare the result with different architects and loss function through the training, validation and testing processes. As shown in Table 3, the best model is attained with InceptionResNetV1 and ArcFace.

Model	Architect	Loss function	Pre-trained	Accuracy
Model1	Inception-ResNetV1	ArcFace	Yes	95.85%
Model2	Inception-ResNetV1	Triplet loss	Yes	94.28%
Model3	SE-ResNeXt-101	ArcFace	No	93.51%

Table 3: Test result of models with different configurations

All images in the validation set are paired with another image within the same class which becomes Validation set 1. All images in the validation set are paired again with another image from a different class which becomes Validation set 2. The same process is applied to the test set as well so it will also contain pairs of images from the same person and pairs of images from different people.

The validation sets are used to evaluate and visualize the capability of our model in distinguishing if two images are from the same person or different people with sample plots shown in Figure 3. The blue area represents the histogram of the Euclidean distance distribution between output embedding of our model for each pair in Validation set 1 while the orange area represents the same information but for Validation set 2. This could be used to visualize the capability of the model in recognizing the pairs of images from the same person and pairs of images from different people. This plot is used in the experiments and evaluation of the performance of our model.

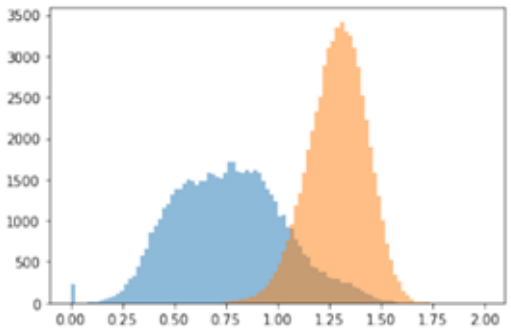

Figure 3: Distribution of two Validation sets

Our experiments mainly discuss the model of Inception-ResNetV1 with ArcFace Loss.

Figure 4 shows the result of training loss reduction. We choose a multistep decay scheduler, an equally spaced five-step learning rate with values 0.1, 0.01, 0.001, 0.0001 and 0.00001(reduce learning rate at the end of epoch 5, 10, 15, 20). The training loss is dropped a lot after every learning rate decay.

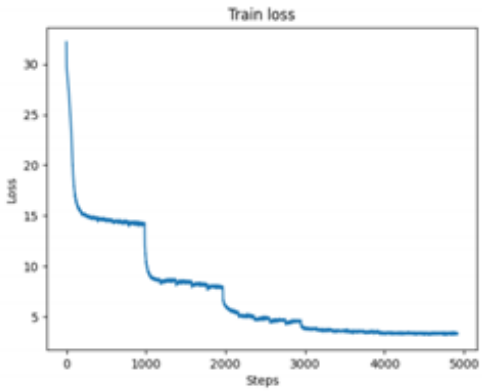

Figure 4: Training loss result of InceptionResNetV1 with ArcFace Loss

Conclusion

To summarize this project of a masked face recognition, We implemented our DCNN pipeline for embedding learning and found that the DCNN architecture, InceptionResNetV1, along with ArcFace loss, could achieve the highest 95.85% accuracy in our experiments.

References

Anwar, A.; and Raychowdhury, A. 2020. Masked face recognition for secure authentication. *arXiv preprint arXiv:2008.11104*.

aqeelanwar. 2020. Masktheface. https://github.com/aqeelanwar/MaskTheFace.

Cao, Q.; Shen, L.; Xie, W.; Parkhi, O. M.; and Zisserman, A. 2018. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, 67–74. IEEE.

Chu, D. K.; Akl, E. A.; Duda, S.; Solo, K.; Yaacoub, S.; Schünemann, H. J.; El-harakeh, A.; Bognanni, A.; Lotfi, T.; Loeb, M.; et al. 2020. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The lancet*, 395(10242): 1973–1987.

Damer, N.; Grebe, J. H.; Chen, C.; Boutros, F.; Kirchbuchner, F.; and Kuijper, A. 2020. The effect of wearing a mask on face recognition performance: an exploratory study. In *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 1–6. IEEE.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Deng, J.; Guo, J.; Xue, N.; and Zafeiriou, S. 2019. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4690–4699.

Du, H.; Shi, H.; Liu, Y.; Zeng, D.; and Mei, T. 2021. Towards NIR-VIS Masked Face Recognition. *IEEE Signal Processing Letters*, 28: 768–772.

Eyiokur, F. I.; Ekenel, H. K.; and Waibel, A. 2021. A Computer Vision System to Help Prevent the Transmission of COVID-19. *arXiv preprint arXiv:2103.08773*.

Ge, S.; Li, C.; Zhao, S.; and Zeng, D. 2020. Occluded face recognition in the wild by identity-diversity inpainting. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(10): 3387–3397.

Geng, M.; Peng, P.; Huang, Y.; and Tian, Y. 2020. Masked face recognition with generative data augmentation and domain constrained ranking. In *Proceedings of the 28th ACM International Conference on Multimedia*, 2246–2254.

Han, Y.; Ryu, C.; Moon, J.; Kim, H.; and Choi, H. 2004. A study on evaluating the uniqueness of fingerprints using statistical analysis. In *International Conference on Information Security and Cryptology*, 467–477. Springer.

Hariri, W. 2021. Efficient masked face recognition method during the covid-19 pandemic. *arXiv preprint arXiv:2105.03026*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.

Huang, G.; Mattar, M.; Lee, H.; and Learned-Miller, E. G. 2012. Learning to align from scratch. In *Advances in neural information processing systems*, 764–772.

Huang, G. B.; Mattar, M.; Berg, T.; and Learned-Miller, E. 2008. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

Jeevan, G.; Zacharias, G. C.; Nair, M. S.; and Rajan, J. 2022. An empirical study of the impact of masks on face recognition. *Pattern Recognition*, 122: 108308.

Li, Y.; Guo, K.; Lu, Y.; and Liu, L. 2021. Cropping and attention based approach for masked face recognition. *Applied Intelligence*, 51(5): 3012–3025.

Minaee, S.; Abdolrashidi, A.; Su, H.; Bennamoun, M.; and Zhang, D. 2019. Biometrics recognition using deep learning: A survey. *arXiv preprint arXiv:1912.00271*.

Neto, P. C.; Boutros, F.; Pinto, J. R.; Damer, N.; Sequeira, A. F.; and Cardoso, J. S. 2021. FocusFace: Multi-task Contrastive Learning for Masked Face Recognition. *arXiv preprint arXiv:2110.14940*.

Ngan, M. L.; Grother, P. J.; Hanaoka, K. K.; et al. 2020. Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face recognition accuracy with face masks using post-COVID-19 algorithms.

nscozzaro. 2019. facenet-pytorch. https://github.com/timesler/facenet-pytorch,2019.

Parkhi, O. M.; Vedaldi, A.; and Zisserman, A. 2015. Deep face recognition.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32: 8026–8037.

Petrescu, R. V. 2019. Face recognition as a biometric application. *Journal of Mechatronics and Robotics*, 3: 237–257.

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

Szegedy, C.; Ioffe, S.; Vanhoucke, V.; and Alemi, A. A. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; and Rabinovich, A. 2015. Going deeper with convolutions, CoRR abs/1409.4842. *URL http://arxiv. org/abs/1409.4842*.

Wang, Z.; Wang, G.; Huang, B.; Xiong, Z.; Hong, Q.; Wu, H.; Yi, P.; Jiang, K.; Wang, N.; Pei, Y.; et al. 2020. Masked face recognition dataset and application. *arXiv preprint arXiv:2003.09093*.

Yi, D.; Lei, Z.; Liao, S.; and Li, S. Z. 2014. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*.