CycleGAN-based Approach for Masked Face Classification

Tomoya Matsubara and Ahmed Moustafa
Nagoya Institute of Technology, Nagoya, Japan

Keywords: Machine Learning, Image, Pattern Recognition.

Abstract: In this paper, we propose a learning model for not only distinguishing whether a person is wearing masks but also classifying the position of the worn masks (mask on my chin, mask on my chin and mouth). First, the synthesized face masks image dataset used for training the model is generated closer to the real world data by CycleGAN. Then, the presence / absence and position of masks are classified using a machine learning model. Experimental results show that this approach provides excellent performance in classifying the presence/ absence and the position of masks.

1 INTRODUCTION

WHO considers wearing a mask to be one of the solutions to prevent the spread of COVID-19 and keep oneself and others safe. In addition, a recent study by researchers at the University of Edinburgh to understand (Bandiera et al., 2020) the measures to tackle the COVID-19 pandemic revealed the following: Wearing a face mask or other cover that covers the nose and mouth reduces the risk of coronavirus infection. In this regard, an efficient system is much needed that can recognize whether or not people's faces are masked in regulated areas and the position of those masks. Therefore, a large dataset of masked faces is required to detect the presence or absence of masks and the position of the masks and to train deep learning models. In this sense, several large datasets of facial images with virus-related protective masks are available in the literature such as the MAsked FAces dataset (MAFA) (Ge et al., 2017), the Real-World Masked Face Dataset (RMFD2) and a comprehensive masked face recognition dataset (Wang et al., 2020) composed of Masked Face Detection Dataset (MFDD), Real -world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD).

Besides, many people have never worn masks or are not wearing them properly due to bad habits or behavior. We then use the following dataset consisting of images with individual or multiple masked faces to create a detection model that takes into account improperly masked faces. A combination of them for correctly masked face datasets (CMFD), incorrectly masked face datasets (IMFD), and global masked face detection (MaskedFace-Net) (Cabani et al., 2021). In addition, there are three types of incorrectly face datasets (IMFD): nose and mouth masks, mouth and chin only masks, and chin only masks:

However, this is a dataset created by synthesizing fake masks. Therefore, we use CycleGAN (Zhu et al., 2017) to take an approach that brings the mask of this dataset closer to the mask of the real world. In addition, since the data set that can be used to detect a human face mask is relatively small, transfer learning is used to classify the presence or absence of a mask and its position.

The rest of this paper is organized as follows: Section 2 introduces the background and preliminaries of the proposed approach. Section 3 presents the proposed approach. Section 4 presents the data, models, settings and results used in the experiment. Section 5 concludes the paper and points out the future work.

2 PRELIMINARIES

2.1 GAN(Generative Adversarial Networks)

The Generative Adversarial Network (GAN) (Goodfellow et al., 2014) consists of a generative model G and a discriminative model D as shown in Figure 1, and learns the generation of data in which G is indistinguishable from the original data x based on the input noise z. do. The loss function of GAN proposed by Goodfellow et al. is called Adversarial Loss as proposed by Zhu et al., and is expressed by Eq.(1)(Zhu et al., 2017) . The first term of Eq.(1) teaches D to correctly identify the original data x, and

Matsubara, T. and Moustafa, A.
CycleGAN-based Approach for Masked Face Classification.
DOI: 10.5220/0010844100003116
In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022) - Volume 3*, pages 476-483
ISBN: 978-989-758-547-0; ISSN: 2184-433X
Copyright © 2022 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

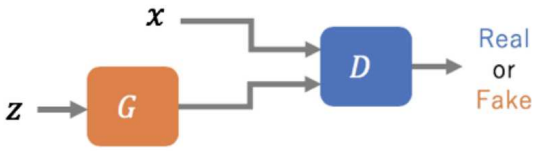

Figure 1: GAN data transition.

the second term teaches D to identify the data generated by G as not the original data. The relationship between G and D here is often expressed by the relationship between the currency counterfeiter and the police who distinguish between the counterfeit currency and the real currency, so that the counterfeiter can produce a counterfeit currency that can deceive the police. Strive, the police will try to distinguish between the fake currency and the real currency. In this way, the counterfeiter and the police compete against each other, and the counterfeiter learns how to make a fake currency that is close to the real thing. Similarly, in GAN, G and D are hostile to each other, allowing G to generate data similar to the original data x. In recent years, models such as CycleGAN and SRGAN that perform super-resolution have been proposed by applying GAN.

$$\mathcal{L}(G,D) = \mathbb{E}_{x\sim Pdata}[log(D(x))] + \mathbb{E}_{z\sim Pz}[1-log(D(z))] \quad (1)$$

2.2 CycleGAN

Pix2pix performs so-called Image-to-Image transformations that generate realistic objects from handwritten edges by acquiring transformation rules for each image pair and unique loss functions between each domain (Isola et al., 2017). Also propose the CycleGAN to perform the Image-to- Image in the framework of unsupervised learning by removing the pair constraint of training data from this pix2pix (Zhu et al., 2017).

CycleGAN (Zhu et al., 2017) is an advanced model of GAN, and is composed of four models, the generative model G_X, G_Y and the discriminative model D_X, D_Y, which correspond to the two datasets X and Y, as shown in Fig.2. Learn the mutual conversion between X and Y, G_X:Y $\rightarrow$ X, G_Y: X $\rightarrow$ Y. Since the mutual conversion here is performed by unsupervised learning, there is no need for pairs between datasets, and the constraints of datasets are relaxed. The transformation obtained by solving this with the following mini-max equation is G_X^*, G_Y^* of Eq.(3) In the work of Zhu et al., the loss function of his CycleGAN is expressed by Eq.(2), $\mathcal{L}_{GAN}$ used in two term in Eq.(2) is expressed by Eq.(4), and $\mathcal{L}_{cyc}$ used in three term is expressed by Eq.(5). Eq.(4) is the GAN loss function, "Adversarial Loss", which two of the Adversarial Loss learn to convert datasets to each other. Eq.(5) is the "Cycle Consistency Loss" for making the transformations consistent, and Eq.(5) teaches it to reduce the difference between the original image and the reconstructed image when reconstructed in two transformations. Fig.2 shows the data transition when trying mutual conversion with a two-dimensional image of apples and mikan as an example of data transition in the CycleGAN, and the image with a hat on the variable shows the converted image. The image with the double hat represents the reconstructed image. Fig.2 shows that Adversarial Loss is used to learn mutual conversion, and Cycle Consistency Loss is used to ensure consistent conversion by G_X and G_Y so that images can be reconstructed.

The expected value of the difference from the original data when G_X and G_Y are similarly applied to the images in both the X and Y domains is recorded as the restoration loss (Zhu et al., 2017). In recent years, applications for information hiding (Chu et al., 2017) and proposals for successfully exchanging human faces (Jin et al., 2017) have also been reported, which is impressive for style conversion between domains under non-pair constraint conditions.

$$\mathcal{L}(G_X,G_Y,D_X,D_Y) = \mathcal{L}_{GAN}(G_X,D_X,X,Y) + \mathcal{L}_{GAN}(G_Y,D_Y,Y,X) + \lambda\mathcal{L}_{CYC}(G_X,G_Y) \quad (2)$$

$$G_X^*,G_Y^* = \arg\min_{G_X,G_Y}\max_{D_X,D_Y}\mathcal{L}(G_X,G_Y,D_X,D_Y) \quad (3)$$

$$\mathcal{L}_{GAN}(G_X,D_X,X,Y) = \mathbb{E}_{x\sim Pdata(x)}[logD_X(x)] + \mathbb{E}_{y\sim Pdata(y)}[log(D_X(1-(G_X(y))))] \quad (4)$$

$$\mathcal{L}_{cyc}(G_X,G_Y) = \mathbb{E}_{x\sim Pdata(x)}[||G_X(G_Y(x))-x||_1] + \mathbb{E}_{y\sim Pdata(y)}[||G_Y(G_X(y))-y||_1] \quad (5)$$

2.3 MaskedFace-Net

Correctly masked face datasets (CMFD), incorrectly masked face datasets (IMFD), and their combination are being used for masked face detection (Masked Face-Net) (Cabani et al., 2021). A realistic masked face dataset is proposed for two purposes: i) detect people whose face is masked or unmasked, ii) the

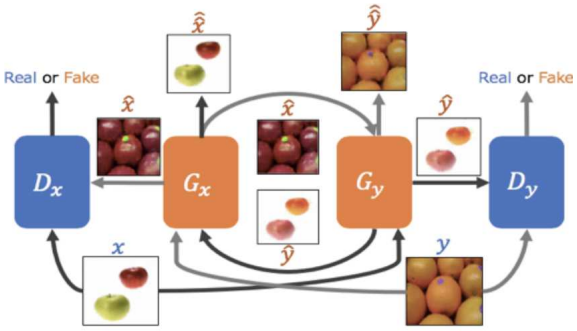

Figure 2: CycleGAN data transition.

mask is properly worn, or Detects incorrectly worn faces (such as airport portals and crowds).

The original facial image dataset is Flickr-Faces-HQ (FFHQ) (Karras et al., 2019), which has been selected as the basis for enhancing the MaskedFace net. In fact, FFHQ contains 70,000 high quality images of human faces in PNG file format. It is published as 1024 x 1024 resolution. The FFHQ dataset offers a variety of things in terms of age, ethnicity, perspective, lighting, and image background. That is, FFHQ contains facial images of all ages, so it also applies to MaskedFace-Net masked facial images.

Such datasets can also be used to detect children in crowds wearing masks below the recommended age limit. MaskedFace-Net composite mask images are created using facial landmarks. For each type of mask-to-face mapping (CMFD, IMFD1, IMFD2, or IMFD3), a subset of 12 face keypoints out of 68 automatically detected landmarks is retained. Then it matches the 12 mask key points. In this way, the mask can fit into a specific area of the face in each case of interest. Therefore, a mask-to-face deformable model was created to generate MaskedFace-Net. In addition, each case of interest can contain up to two keypoints in the mask (out of twelve keypoints), and their positions move randomly around a limited area. In particular, this margin of error can affect the height of the facial mask and bring realism to the generated dataset. Therefore, MaskedFace-Net also includes various placement masks, as shown in Fig3.

Figure 3: Examples of images included in MaskedFace-Net.

Therefore, the resulting Masked Face-Net dataset contains 137,016 masked face images. The proposed Masked Face-Net dataset consists of 49% of properly masked faces (67,193 images) and 51% of incorrectly masked faces (69,823 images). In this latter set, about 80% represent a face with only the mouth and chin masked, 10% with a face with only the nose and mouth masked, and 10% with a face with only the chin masked.

2.4 ResNet

Residual Networks (ResNet) (He et al., 2016) is the 2015 ILSVRC winning model. Deepening the network improves expressiveness and recognition accuracy, but too deep a network makes efficient learning difficult. ResNet does not simply pass the conversion $F(x)$ by some processing block to the next layer like a normal network, but shortcuts the input x to that processing block, and $H(x) = F(x) + x$ is passed to the next layer. The processing unit including this shortcut is called the residual module. In ResNet, the gradient is directly transmitted to the lower layer during backpropagation through the shortcut, and it has become possible to learn efficiently even in a very deep network.It is also used in Highway Networks (Srivastava et al., 2015a)(Srivastava et al., 2015b), but has not improved accuracy in very deep networks.

Fig.4 shows the structure of the Residual module. Fig.4(a) shows the abstract structure of the residual module, and Fig.4(b) shows an example of the residual module actually used, with two 3x3 convolution layers with 64 output channels. It is arranged. To be precise, in addition to the convolutional layer, batch normalization and ReLU, which will be described later, are arranged, and ResNet uses a residual module with the following structure

$$conv - bn - relu - conv - bn - add - relu$$

Here add shows the sum of $F(x)$ and x. Fig.4(c) shows the bottleneck version of the residual module, which is a 1×1 convolution that reduces dimensions, then 3×3 convolutions, and then 1×1 to restore dimensions. By taking the form, a deeper model can be constructed while maintaining the same amount of calculation as in Fig.4(b). In fact, the accuracy of ResNet-50 using the module of Fig.4(c), which has the same number of parameters, is greatly improved compared to ResNet-34 using the residual module of Fig.4(b). Has been reported. Identity function $f(x) = x$ is basically used as a shortcut for the Residual module, but if the number of input channels and the number of output channels are different, zero-padding is used to fill the missing channels with 0. Two patterns of shortcuts, projection, which adjust the number of channels by convolution of , 1×1,

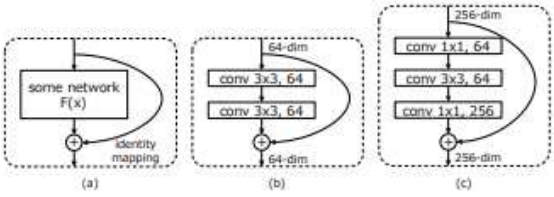

Figure 4: Residual module structure.

are available as options. From these, the zero-padding approach is better because it does not increase the parameters, but projection, which is easy to implement, is often used.

In deep networks, updating the parameters of one layer causes an internal covariate shift in which the distribution of inputs to the next layer changes significantly from batch to batch, resulting in inefficient learning. was there. Batchnormalization (Ioffe and Szegedy, 2015) is a method to stabilize and speed up learning by normalizing this internal covariate shift and allowing each layer to learn independently as much as possible. In ResNet, efficient learning of deep networks is realized by incorporating this batchnormalization in the residual module, and batchnormalization has come to be used as standard in the models after ResNet.

3 PRPPOSED APPROACH

It is said that differences can be identified with high accuracy by using CNN, and attempts have been made to identify gender differences using Grad-CAM, which identifies regions that contribute to discrimination using the weights after learning CNN (Jiang et al., 2020). However, it has not been possible to identify a meaningful area, and the reason for the identification has hardly been explained.

One way to know what shape or pattern the CNN is looking at is to find out what features the CNN filter is looking at. However, since CNN learns to extract more complex features by combining simple features extracted in the shallow layer in the deep layer, it is necessary to express complex features in order to know what CNN itself is looking at. It is necessary to know the multiple simple features used in the above and to consider what kind of features are expressed from them. In addition, since many filters are used in each layer of CNN, it is necessary to think about multiple filters in the same way in multiple layers, and it is not realistic to actually explain the identification process from the filters. In addition, in order to explain the difference between images, it is necessary to know not only the area but also the difference in shape and pattern within the area, and it is not enough to specify the area. Therefore, in order to explain fake masks and real masks, there is a need for a method that can identify the areas involved in their identification and obtain differences in shapes and patterns within the areas.

Therefore, in this study, we propose a method using the hostile generation network (GAN) as a method to know the shape and pattern related to the identification of CNN from the result instead of the process. In order to analyze fake masks and real masks using GAN, it is necessary to use a model that can learn the difference between fake masks and real masks. Therefore, as a model that can learn the difference, there is CycleGAN, which is a model that applies GAN. Since CycleGAN learns mutual conversion between datasets by unsupervised learning, there is no need for data pairs between datasets, and it is possible to learn transformations that have no solution in reality, such as mutual conversion between fake masks and real masks.Specifically, the MaskedFace-Net dataset is used as the ”fake mask” domain and the MAsked FAces dataset (MAFA) (Ge et al., 2017) is used as the ”real mask” domain to create a CycleGAN. Use to perform mask transformation training between domains. Fig.5 shows a schematic diagram. When training data is given with the ”fake mask” domain as X and the ”real mask” domain Y as training data, it is equivalent to optimizing G_Y: X $\rightarrow$ Y and G_X: Y $\rightarrow$ X by CycleGAN. In addition, since each domain plays the role of a teacher domain with each other, it should be considered that it exerts a probabilistic learning effect on a data group such as MaskedFace-Net that is not given clear teacher label data.However, since CycleGAN converts the entire image, the area related to the mask cannot be specified from the difference.

Therefore, in this study, we introduced the following additional loss function into CycleGAN to limit the conversion area.

$$\mathcal{L}_{identity}(G_X, G_Y) = \mathbb{E}_{y\sim Pdata(y)}[||G_Y(y) - y||_1] + \mathbb{E}_{x\sim Pdata(x)}[||G_X(x) - x||_1] \quad (5)$$

In other words, the L1 norm is used to set the loss function so that the distributions of the ”input” and the ”generated image” are close to each other. In other words, by using this ”identity loss”, GAN works to learn the conversion between each domain. Then, we thought that we could change the color and style of the area we wanted to convert and keep the color and style of the area we didn't want to convert. In addition, the following three points were implemented to improve the accuracy of images by CycleGAN.

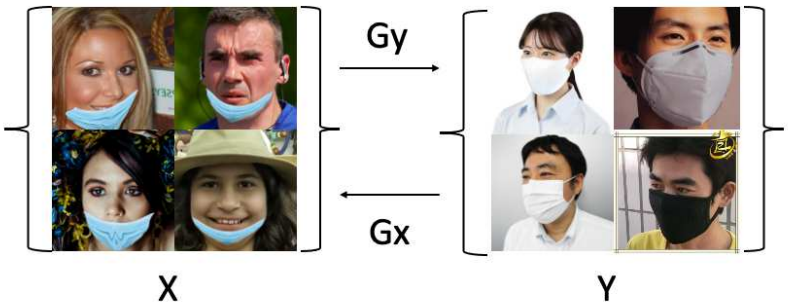

Figure 5: Schematic diagram of mask conversion learning between domains by CycleGAN Relationship between “fake mask” domain X and “real mask” domain Y extracted from MaskedFace-Net G_Y: X → Y and G_X: Learn Y → X using an unsupervised method with no pair constraints.

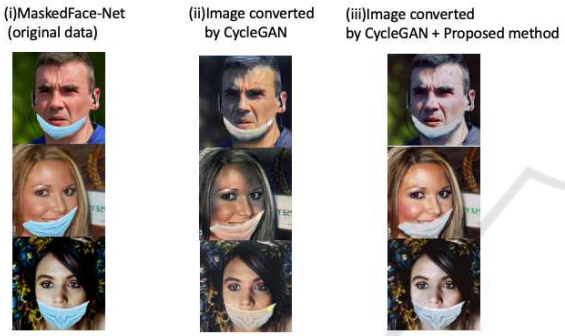

Figure 6: Three types of training data sets. From left i) MaskedFace-Net (original data), ii) Image converted by CycleGAN iii) Image converted by CycleGAN + Proposed method.

- Changed Discriminator loss function to ”MSE” instead of ”BCE”
- Changed cyclerate from ”10” to ”1”
- Learn BatchSize with ”1”

4 EXPERIMENT

Figure 7: An example of test data.

4.1 Training Data

The following three types of training data sets have been prepared this time. i) Dataset that uses MaskedFace-Net as it is, which is the original data. ii) Dataset that Converts MaskedFace-Net to match the real world using only CycleGAN. iii) Dataset that Converts MaskedFace-Net to match the real world using CycleGAN + Proposed Approach.

The images in Fig.6 are (i), (ii), and (iii) explained earlier from the left. The original data on the left is a fake made by compositing the mask, so it feels strange compared to the data in the real world. Also, comparing the center and right images, we think the proposed method is less likely to darken and is applied to more realistic data.

In addition, there are four types of these datasets: Mask, MaskChin, MaskMouthChin, and NoMask. ”Mask” is a mask that covers the entire face, ”MaskChin” is a mask that covers only the chin, ”MaskMouthChin” is a mask that covers only the mouth and chin, and ”NoMask” is a face without a mask.

4.2 Test Data

For the test data, we use the pictures we actually took and the images we searched for on the Internet and SNS. In addition, these test data include a wide variety of masks such as various colored masks and cloth masks. And, there are masked face images at various angles such as facing front and facing sideways, facing up. Fig.7 is an example of such test data.

In addition, there are four types of this dataset: Mask, MaskChin, MaskMouthChin, and NoMask.

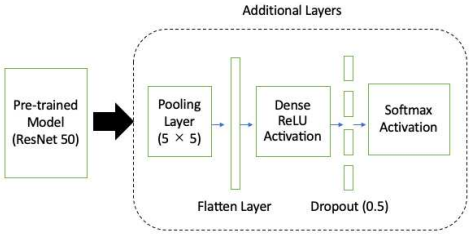

Figure 8: Fine-tuning of ResNet50.

4.3 Fine-tuning Model

Face mask detection is achieved using deep neural networks. However, training deep neural networks is time consuming and expensive due to the high computational power required. To solve these problems, we will use deep learning-based transfer learning here. Transfer learning can transfer the trained knowledge of neural networks to new models. So even if one are training on a small dataset, one can improve the performance of one's new model. There are several pre-trained models such as AlexNet, MobileNet, and ResNet50 trained with 14 million images from the ImageNet dataset (Chowdary et al., 2020). And Fig.8 is a simplified version of the model used this time. ResNet50 has been selected as the pre-trained model for face mask classification. In addition to the Pre-trained Model ResNet50, the last layer has been tweaked by adding five new layers. Newly added layers include a 5x5 pool size average pooling layer, a flattering layer, a 128 neuron ReLU layer, a 0.5 dropout, and a decision layer with a softmax activation function for binary classification.

4.4 Experiment Setting

Three types of training datasets are used. i) MaskedFace-Net (original data), ii) Image converted by CycleGAN, iii) Image converted by CycleGAN + Proposed approach. In addition, each of these three types has the following settings.

- Number of images: 3825
- Number of types: 4
- Image size: 256×256
- Learning: 20 epochs

Also, for the test data, we used the real world data by ourselves, and the settings are as follows.

- Number of images: 1285
- Number of types: 4
- Image size: 256×256

Then, using the deep learning model shown in Fig.5, four classifications are performed: Mask, MaskChin, MaskMouthChin, and NoMask. In addition, use k-fold cross-validation to divide the data into k pieces, use one of them as test data and all the rest as validation data. This time, using ResNet, the evaluation result is obtained by averaging the results of K times. In this experiment, $k = 5$.

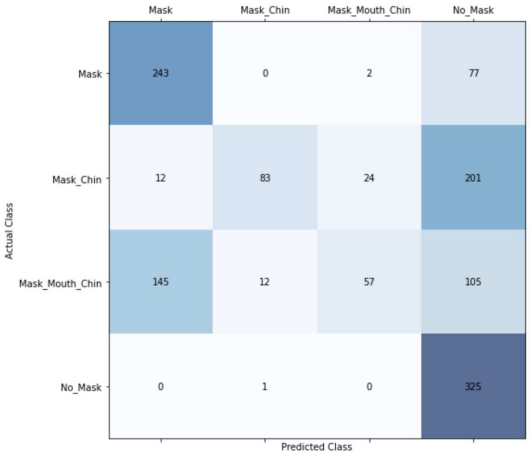

Figure 9: Mixed matrix classified using i)MaskedFace-Net (original data) as training data.

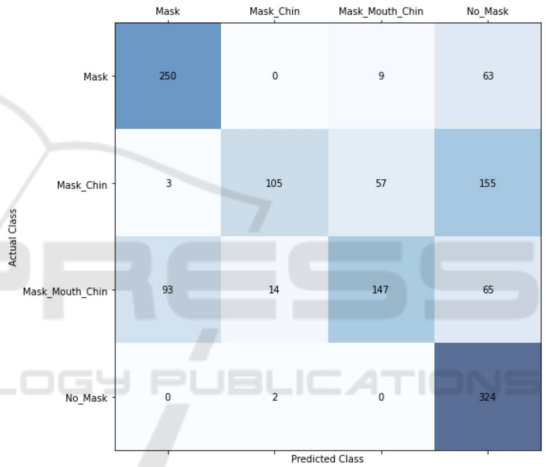

Figure 10: Mixed matrix classified using ii)Image converted by CycleGAN as training data.

4.5 Experiment Result

Here we show that the training dataset varies in classification accuracy. Therefore, a mixed matrix of classification results by each training dataset, Fig.9 to 11 have been created.

In addition, Precision, Recall, and F1 scores for each training data and classification were calculated. Based on these evaluations, we evaluated the classification performance of each case classification. The results are shown in Table 1, Table 2, and Table 3. From Tables 1-3, Precision, Recall, and F1 scores were the highest in all classifications of the proposed method. From this result, it is considered that the mask training data created by synthesis can be adapted to the real-world data by using CycleGAN. Furthermore, we believe that the color space can be stabilized by letting GAN learn the conversion be-

Table 1: Precision of each training data and classification.

Training Data	Mask	MaskChin	MaskMouthChin	NoMask
i)MaskedFace-Net (original data)	0.607	0.864	0.686	0.459
ii)Image converted by CycleGAN	0.722	0.867	0.690	0.553
iii)Image converted by CycleGAN + Proposed	0.920	0.881	0.888	0.655

Table 2: Recall of each training data and classification.

Training Data	Mask	MaskChin	MaskMouthChin	NoMask
i)MaskedFace-Net (original data)	0.754	0.259	0.178	0.996
ii)Image converted by CycleGAN	0.770	0.328	0.460	0.993
iii)Image converted by CycleGAN + Proposed	0.879	0.645	0.695	0.996

Table 3: F1-score of each training data and classification.

Training Data	Mask	MaskChin	MaskMouthChin	NoMask
i)MaskedFace-Net (original data)	0.673	0.399	0.247	0.628
ii)Image converted by CycleGAN	0.748	0.476	0.552	0.694
iii)Image converted by CycleGAN + Proposed	0.899	0.745	0.787	0.790

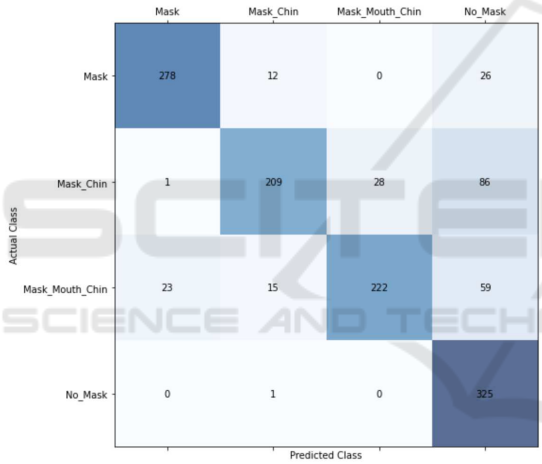

Figure 11: Mixed matrix classified using iii)Image converted by CycleGAN + Proposed as training data.

tween domains using ”identity loss”. As a result, we found that it is possible to classify the presence or absence of masks and their positions by CNN even with composite images.

5 CONCLUSION

We suggested using CycleGAN and ”identity loss” and three points to classify the presence and location of masks. Experimental results have shown that the proposed approach is very accurate in all classifications. In other words, from this result, it is considered that the fake mask data created by synthesis can be brought closer to the actual mask by using CycleGAN. We also believe that even now that the coronavirus is widespread, we can still collect data without having to meet people in person.

However, the classification accuracy of ”Mask Chin” and ”Mask Mouth Chin” was not very good, so it is necessary to work on improving the accuracy by improving each model, verifying with more data, and verifying the reliability of the result with different data. We also want to develop a system that can detect the presence and position of multiple people’s masks at once by combining it with object detection for multiple purposes.

REFERENCES

Bandiera, L., Pavar, G., Pisetta, G., Otomo, S., Mangano, E., Seckl, J. R., Digard, P., Molinari, E., Menolascina, F., and Viola, I. M. (2020). Face coverings and respiratory tract droplet dispersion. *Royal Society open science*, 7(12):201663.

Cabani, A., Hammoudi, K., Benhabiles, H., and Melkemi, M. (2021). Maskedface-net–a dataset of correctly/incorrectly masked face images in the context of covid-19. *Smart Health*, 19:100144.

Chowdary, G. J., Punn, N. S., Sonbhadra, S. K., and Agarwal, S. (2020). Face mask detection using transfer learning of inceptionv3. In *International Conference on Big Data Analytics*, pages 81–90. Springer.

Chu, C., Zhmoginov, A., and Sandler, M. (2017). Cyclegan, a master of steganography. *arXiv preprint arXiv:1712.02950*.

Ge, S., Li, J., Ye, Q., and Luo, Z. (2017). Detecting masked faces in the wild with lle-cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2682–2690.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B.,

Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

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*, pages 770–778.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Jiang, H., Lu, N., Chen, K., Yao, L., Li, K., Zhang, J., and Guo, X. (2020). Predicting brain age of healthy adults based on structural mri parcellation using convolutional neural networks. *Frontiers in neurology*, 10:1346.

Jin, X., Qi, Y., and Wu, S. (2017). Cyclegan face-off. *arXiv preprint arXiv:1712.03451*.

Karras, T., Laine, S., and Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410.

Srivastava, R. K., Greff, K., and Schmidhuber, J. (2015a). Highway networks. *arXiv preprint arXiv:1505.00387*.

Srivastava, R. K., Greff, K., and Schmidhuber, J. (2015b). Training very deep networks. *arXiv preprint arXiv:1507.06228*.

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*.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS