GAN-BASED DATA AUGMENTATION AND ANONYMIZATION FOR MASK CLASSIFICATION

Mustafa Çelik[1, 2], Ahmet HaydarÖrnek[1, 3]

[1]Huawei Turkey R&D Center, Istanbul, Turkey
[2]Department of Computer Engineering Faculty of Computer and Informatics, Istanbul Technical University, Istanbul, Turkey
[3]Department of Electrical and Electronics Engineering, Konya Technical University, Konya, Turkey

ABSTRACT

Deep learning methods, especially convolutional neural networks (CNNs), have made a major contribution to computer vision. However, deep learning classifiers need large-scale annotated datasets to be trained without over-fitting. Also, in high-data diversity, trained models generalize better. However, collecting such a large-scale dataset remains challenging. Furthermore, it is invaluable for researchers to protect the subjects' confidentiality when using their personal data such as face images. In this paper, we propose a deep learning Generative Adversarial Networks (GANs) which generates synthetic samples for our mask classification model. Our contributions in this work are two-fold that the synthetics images provide. First, GANs' models can be used as an anonymization tool when the subjects' confidentiality is matters. Second, the generated masked/unmasked face images boost the performance of the mask classification model by using the synthetic images as a form of data augmentation. In our work, the classification accuracy using only traditional data augmentations is 93.71 %. By using both synthetic data and original data with traditional data augmentations the result is 95.50 %. It is shown that the GAN-generated synthetic data boosts the performance of deep learning classifiers.

Keywords

Convolutional Neural Network, Data Anonymization, Data Augmentation, Generative Adversarial Network, Mask classification

1. INTRODUCTION

In the last few years, wearing a mask had vital importance because of the pandemic that affects the whole people. It has become mandatory to wear a mask to avoid the disease. However, people are not sensitive to wearing a mask. For this reason, mask classifiers detect whether a person wears a mask is necessary for a public area to warn people and provide a healthy environment. Although the traditional machine learning methods are used in image classification work, it requires a huge amount of preprocessing and feature extraction steps. Over the last decade with the popularity of deep learning, image classification models have produced remarkable performance without spending effort on preprocessing and feature extraction.

One of the crucial issues in deep learning-based classifiers is the dataset size. Small-sized datasets may cause over-fitting and produce poor generalization on the test set. In many realistic cases like mask classification, we have limited datasets because obtaining labeled data is costly and time-consuming. Moreover, collecting personal data (e.g., face and face with mask) is more challenging due to subjects' confidentiality.

David C. Wyld et al. (Eds): NLP, MLTEC, CLBD, SEAPP, NeTIoT, VLSIE, ITCS, ARIA, DMS - 2021
pp. 191-199, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112315

To solve the overfitting problem, dropout [7], layer normalization [1], batch normalization [9] methods have been implemented. However, the models underperform in testing data because of the small-sized dataset. Researchers try to overcome this problem by using traditional data augmentation techniques. These techniques which include operations such as zooming, cropping, rotating, flipping, and scaling, are the standard methods that improve the performance of the classifiers (e.g., mask classifier). However, the images that are proliferated by using traditional augmentation techniques are fundamentally highly correlated and have a similar distribution to the original images.

In this work, we propose to use the generative adversarial network (GAN) [5] model to proliferate synthetic masked and unmasked face images which provide an additional form of data augmentation and an effective tool for data anonymization. The model consists of two networks (e.g., generator, discriminator), one network generates synthetic images and the other one discriminates between original and synthetic images.

The contributions of this work are the following:

- Generating synthetic masked and unmasked face images using GANs to boost the accuracy of the mask classification model.
- Anonymization of the real-world images to protect subjects' confidentiality.

2. Related Work

Although there are plenty of studies [6, 8, 16, 18] that use deep learning-based approaches, these studies are based on clinical data and aim to diagnose the COVID-19 after the subjects have been already infected. The advantage of the deep learning approach can be used to develop a prevention system against the pandemic such as a model that detects whether people wear a mask or not.

Authors in [13] developed a combined deep learning and machine learning model which used deep learning to extract features and support vector machines to classify the samples whether wear a mask or not. In [19], researchers proposed a deep learning-based face mask-wearing condition identification method that consists of pre-processing, face detection crop, super-resolution, and mask-wearing identification steps. On common camera devices, [11] proposed an edge computing-based mask detection model which provides a real-time performance.

GANs are a promising approach that syntheses images [5]. Over the last decades, GANs have gained an extreme reputation in computer vision. Different types of GANs have been proposed to generate quite realistic natural images [4, 10, 14, 17, 20, 21]. Also, GAN-based models have been used to generate synthetic samples, especially in medical imaging [2, 3]. To the best of our knowledge, there is no existing literature on GAN-based synthetic masked face images generation as a form of data augmentation and anonymization.

3. Materials and Method

3.1. Materials

The images used in this study were taken by Huawei M2150 Camera which can send images via FTP protocol. By creating a monitoring setup at the Huawei entrance a dataset including real-world images was obtained. To train and test our deep learning model, 18400 images and 1178 images were used, respectively. The summary of the material can be seen in Table 1.

Table 1. The Properties of The Camera and Dataset

Effective pixels	2560 (H) x 1920 (V)
CPU	Hi3516D
Effective pixels	2560 (H) x 1920 (V)
CPU	Hi3516D
Frame Rate	30 FPS
Computing Power	1 TOP
Video Encoding Format	H.265/H.264/MJPEG
Intelligent Analysis	Face and Person Detection
Dataset	19578 images

3.2. Generating Synthetic Images

The key point of training a CNN network (e.g., mask classification) is the size of the training dataset. A large-sized dataset boosts the accuracy of the model. To enlarge the dataset we augmented the dataset in two different approaches: 1) classic augmentation methods which include operations such as zooming, cropping, rotating images, etc. 2) generating synthetic masked and unmasked face images with the help of generative models which use the data samples.

3.2.1. Classic Data Augmentation

The CNN model which has hundreds of parameters need a large-sized dataset to be trained. When building models with such networks that have multiple layers and there is a limited number of data, it is possible to face an over-fitting problem. The basic approach to solve the over-fitting is the classic data augmentation techniques [12]. These techniques include image transformations such as zooming, cropping, rotating, translation, scaling, flipping, and so on.

3.2.2. Generative Adversarial Network for Image Synthesis

Recently, GANs are popular frameworks that generate synthetic samples. It aims to learn the distribution of a dataset (e.g., masked/unmasked face images) in order to generate new images based on the learned distribution.

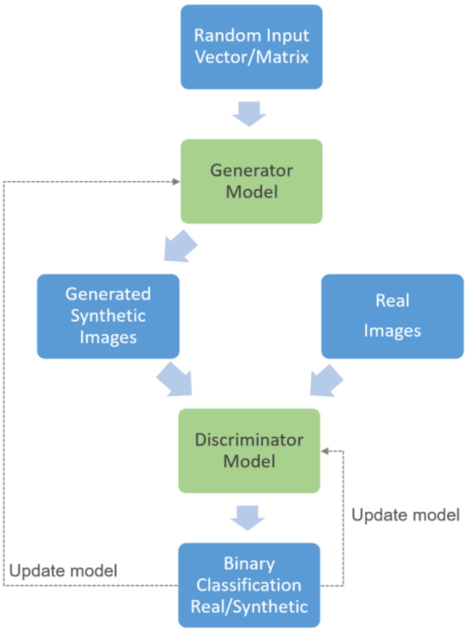
Random Input Vector/Matrix
Generator Model
Generated Synthetic Images
Real Images
Discriminator Model
Update model
Binary Classification Real/Synthetic
Update model

Figure 1. GAN Architecture Overview

While there exist many different types of approaches used in generative modeling, a GAN uses the approach shown in Fig. 1. There are two different neural networks called Generator (G) and Discriminator (D). The generator model generates synthetic samples (e.g., images) by using the given random input vector. The discriminator model tries to detect whether a given sample is real or synthetic. The training process follows each other, the discriminator model is trained a few epochs, then the generator is trained a few epochs, the process repeats till both the generator and discriminator model get better.

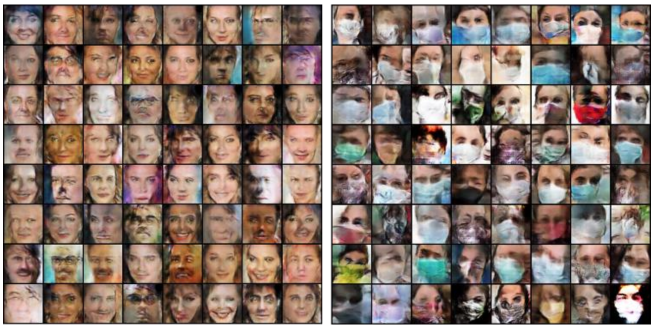

Figure 2. Masked and Unmasked Synthetic Images Generated by GAN Model

GANs are highly sensitive to hyperparameters, activation functions, and regularization. Table 2 shows the parameters of the models.

Table 2. Hyperparameters Used For Training The Model

Parameter Name	Value
Image size	64
Batch size	128
Learning rate	0.0001
Epochs	100
The activation function of discriminator's output-layer	sigmoid
The activation function of discriminator's middle-layers	Leaky ReLu
The activation function of the generator's output-layer	hyperbolic tangent
The activation function of the generator's middle-layer	ReLu

3.2.2.1. Generator Network

Generator network uses a random number vector or matrix as which is used as a seed to generate synthetic samples (e.g., image). It takes a 128x1x1 shaped tensor and converts it to a 3x64x64 images. In order to do the conversion, a deep convolutional GAN architecture is used (Fig. 3).

The middle layer of the architecture uses the ReLu Activation function [15]. In the output layer of the activation function, the hyperbolic tangent function (tanh) is used.

Generator Training Steps:

- The generator generates a batch of synthetic images.
- The synthetic images are given to the discriminator model.
- The discriminator model calculates the loss for the synthetic images.
- The weights of the generator model are adjusted with the help of the loss value which is calculated by the discriminator model.

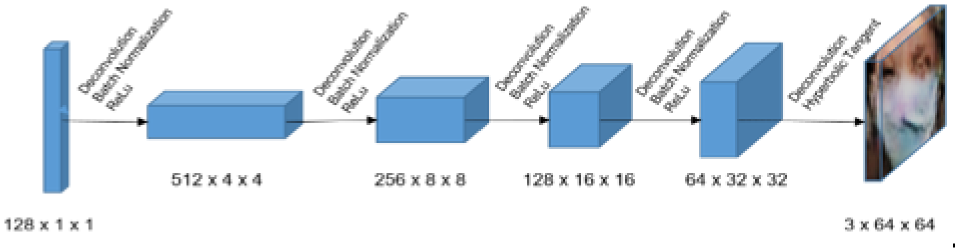

Figure 3. Generator Architecture

3.2.2.2. Discriminator Network

The Discriminator network uses CNN. It takes an image as input and classifies it as a real or synthetic image. As an input 3x64x64 image is given to the network. Discriminator gives an output of a single number between 0 and 1 which is a probability of the image being real. The architecture of the discriminator is shown in Fig. 4. After each layer Leaky ReLu activation is used, except the output layer uses the Sigmoid function. Batch normalization is applied after each middle layer conversion. Also, the discriminator model which is a binary classification model can use binary cross-entropy loss function for evaluation.

Discriminator Training Steps:

- It is expected that the discriminator model gives 0 if the given input image is generated by the generator model. If the output is 1, it means that the given image is from the real dataset which means the image is not generated by the generator model.
- A batch of real images is given to the discriminator model with the label of 1. Discriminator calculates the loss for real images.
- A batch of synthetic images is given to the discriminator model with the label of 0. Discriminator calculates the loss for synthetic images.
- The loss values of the synthetic and real images are added, and an overall loss value is calculated.
- The weight of the discriminator model is updated by using the overall loss of the whole input images.

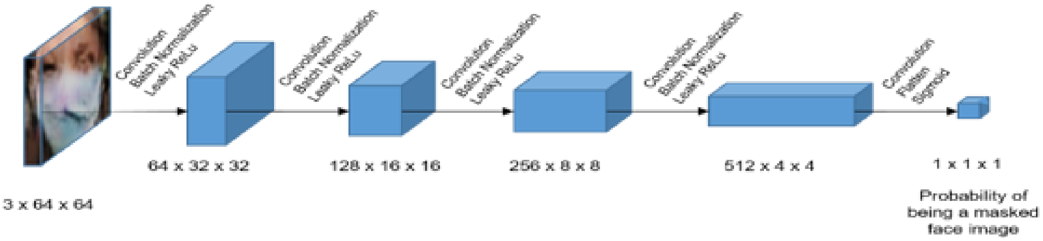

Figure 4. Discriminator Architecture

4. EXPERIMENTS

To generate new face images with and without a mask, a GAN model was used and 1000 pieces images with mask 1000 pieces images without mask were generated. Two different scenarios were created to compare the effects of GAN-generated images on the training performance (i) training set was used to train the model and tested (Fig. 5) (ii) training set with GAN-generated images was used to train the model and tested (Fig. 6).

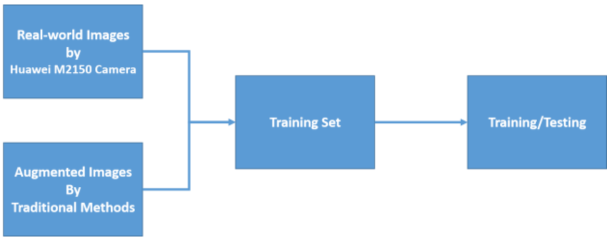
Real-world Images by Huawei M2150 Camera
Augmented Images By Traditional Methods
Training Set
Training/Testing

Figure 5. Scenario 1. The real-world images and images augmented by traditional methods are used to train the ResNet18 model.

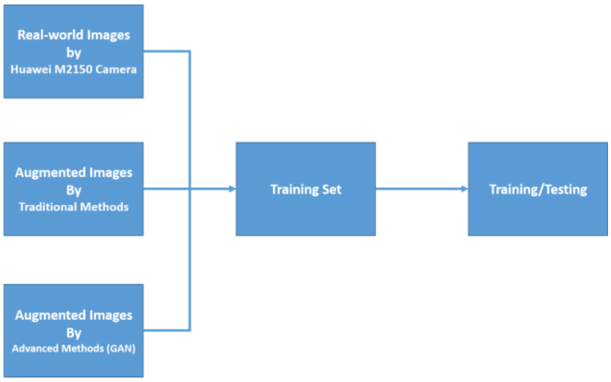

Figure 6. Scenario 2. The real-world images , images augmented by traditional methods, and images augmented by the GAN are used to train the ResNet18 model.

The ResNet18 pre-trained model was trained with a training set for the first scenario. For the second scenario, 1000 pieces images with mask 1000 pieces without masks were generated, and the ResNet18 pre-trained model was trained. The data information can be seen in Table 3}.

Table 3. Scenarios - Training Dataset

Training Dataset	**Scenario 1**	**Scenario 2**
With mask	9200	9200
Without mask	9200	9200
GAN-generated with mask	-	1000
GAN-generated without mask	-	1000
Total training data	18400	20400

5. Results

After training parts were completed, trained 2 models were evaluated with the same testing dataset including 950 images with mask, and 228 images without a mask. The results can be seen in Table 4.

Table 4. All Results

All Results	**Scenario 1 (%)**	**Scenario 2 (%)**
Sensitivity	75.00	88.15
Specificity	98.21	97.28
Accuracy	93.71	95.50

According to the first scenario, 933 of 950 images with masks and 171 of 228 images without masks were correctly classified by the ResNet18 model. Therefore, the model achieved 98.21% specificity, 75% sensitivity, and 93.71% accuracy.

According to the second scenario, 924 of 950 images with masks and 201 of 228 images without masks were correctly classified by the ResNet18 model. Therefore, the model achieved 97.28% specificity, 88.15% sensitivity, and 95.50% accuracy.

6. DISCUSSION

Training a deep learning model without data augmentation generally causes overfitting because small datasets cannot be generalized by deep learning models. Data augmentation methods are categorized into traditional and advanced methods. Traditional data augmentation methods are already implemented by deep learning frameworks such as Tensorflow and Pytorch. Although traditional methods increase model performances advanced methods help models to achieve more performances.

In this study, we show how advanced methods increase the model performance by creating 2 different scenarios. Whereas, only traditional methods were used and achieved 93.71% accuracy in the first scenario, traditional and advanced (GAN-generated) methods were used together and achieved 95.50% accuracy in the second scenario.

It is shown that the GAN-generated synthetic data boosts the performance of the classifier. In future studies, we will be working on generating different images to train deep learning models.

7. CONCLUSION

In this work, we propose a GAN model to generate synthetic masked and unmasked images to increase classification performance and provide data anonymization.

By creating 2 different scenarios, how advanced methods increase the model performance was shown. While traditional methods achieved 93.71% accuracy, traditional and GAN methods together achieved 95.50% accuracy.

With the development of new GAN models, we will generate new images and extend our dataset to train our models without overfitting and provide more privacy.

ACKNOWLEDGMENT

This study was supported by the ”Epidemic Prevention System” project of Huawei Turkey R&D Center.

REFERENCES

[1] J. L. Ba, J. R. Kiros, and G. E. Hinton. Layer normalization. arXiv preprint arXiv:1607.06450, 2016.

[2] A. Chartsias, T. Joyce, M. V. Giuffrida, and S. A. Tsaftaris. Multimodal mr synthesis via modality-invariant latent representation. IEEE transactions on medical imaging, 37(3):803–814, 2017.

[3] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abra`moff, A. M. Mendonc¸a, and A. Campilho. End-to-end adversarial retinal image synthesis. IEEE transactions on medical imaging, 37(3):781–791, 2017.

[4] E. Denton, S. Chintala, A. Szlam, and R. Fergus. Deep generative image models using a laplacian pyramid of adversarial networks. arXiv preprint arXiv:1506.05751, 2015.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. Advances in neural information processing systems, 27, 2014.

[6] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055, 2020.

[7] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.

[8] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, and L. Xia. Serial quantitative chest ct assessment of covid-19: a deep learning approach. Radiology: Cardiothoracic Imaging, 2(2):e200075, 2020.

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

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

[11] X. Kong, K. Wang, S. Wang, X. Wang, X. Jiang, Y. Guo, G. Shen, X. Chen, and Q. Ni. Real-time mask identification for covid-19: an edge computing-based deep learning framework. IEEE Internet of Things Journal, 2021.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25:1097–1105, 2012.

[13] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic. Measurement, 167:108288, 2021.

[14] M. Mirza and S. Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.

[15] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In Icml, 2010.

[16] Q. Ni, Z. Y. Sun, L. Qi, W. Chen, Y. Yang, L. Wang, X. Zhang, L. Yang, Y. Fang, Z. Xing, et al. A deep learning approach to characterize 2019 coronavirus disease (covid-19) pneumonia in chest ct images. European radiology, 30(12):6517–6527, 2020.

[17] A. Odena, C. Olah, and J. Shlens. Conditional image synthesis with auxiliary classifier gans. In International conference on machine learning, pages 2642–2651. PMLR, 2017.

[18] Y. Oh, S. Park, and J. C. Ye. Deep learning covid-19 features on cxr using limited training data sets. IEEE transactions on medical imaging, 39(8):2688–2700, 2020.

[19] B. Qin and D. Li. Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19. Sensors, 20(18):5236, 2020.

[20] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.

[21] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. Semantic image in painting with deep generative models. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5485–5493, 2017.