Conditional Wasserstein Generative Adversarial Networks for Rebalancing Iris Image Datasets

Yung-Hui LI†††, Nonmember, Muhammad Saqlain ASLAM†, Student Member, Latifa Nabila HARFIYA†, and Ching-Chun CHANG†††, Nonmembers

SUMMARY The recent development of deep learning-based generative models has sharply intensified the interest in data synthesis and its applications. Data synthesis takes on an added importance especially for some pattern recognition tasks in which some classes of data are rare and difficult to collect. In an iris dataset, for instance, the minority class samples include images of eyes with glasses, oversized or undersized pupils, misaligned iris locations, and iris occluded or contaminated by eyelids, eyelashes, or lighting reflections. Such class-imbalanced datasets often result in biased classification performance. Generative adversarial networks (GANs) are one of the most promising frameworks that learn to generate synthetic data through a two-player minimax game between a generator and a discriminator. In this paper, we utilized the state-of-the-art conditional Wasserstein generative adversarial network with gradient penalty (CWGAN-GP) for generating the minority class of iris images which saves huge amount of cost of human labors for rare data collection. With our model, the researcher can generate as many iris images of rare cases as they want and it helps to develop any deep learning algorithm whenever large size of dataset is needed. Key words: deep learning, generative adversarial network, iris image generation, machine learning neural networks, signal synthesis

1. Introduction

Generative adversarial networks (GANs), first proposed by [1], have incurred problems such as difficulties in training, the inability of loss functions in generator and discriminator to direct training processes, and the lack of diversity when generating samples. Wasserstein GANs (WGANs) is later introduced in 2017 [2], [3] to successfully solve the training instability in GANs, ensure diversity in generated samples, and clarify the flaws in the training of generator and discriminator networks, thus enabling improvements. Following in the same year, Gulrajani et al. [4] developed an improved training method for WGANs (WGAN-GP), which features gradient penalty, a novel continuity restriction method, to successfully address gradient vanishing and explosion in training. The convergence speed of WGANs and the quality of generated samples were also improved.

Currently, all deep learning algorithms, such as those involved in human face [5]–[7] and [8]–[10] recognition, require large volumes of data to maximize the accuracy of models in training. Therefore, it is very important to collect a large amount of diversified training data for training a robust and accurate deep neural network.

Insufficient training data has undermined the results of the training process [11]. Among iris databases, the CASIA-IrisV4 database released by the Chinese Academy of Sciences is currently the largest [12], but it contains only 20,000 images, which cannot be described as “big” in the deep learning world. Among those 20,000 images, there are even fewer for special cases. A database larger than CASIA-IrisV4 is highly desired for developing deep-learning-based algorithms for iris recognition. However, in order to collect a huge iris database like CASIA-IrisV4, it asks for a large number of cooperative subjects. Moreover, if we would like to ask for diversity in the dataset for some specific purpose, it becomes even harder since such special cases are rare in nature.

The aforementioned phenomenon when the data for special cases and normal cases are not equally represented is brought up as a class imbalance issue. Class imbalance can inhibit predictive performance by its nature [13]. For iris recognition, normal iris images which mean naked eye with pupil and eye with normal size are easy to get segmented correctly. However, most iris segmentation algorithms fail when some abnormal cases happen, for example, iris images with a too small or too big pupil, iris images of the person wearing glasses that induce strong specular reflection in large areas, and iris image with off-axis gazing [14], [15]. The normal iris image samples are shown in the first row of Fig. 1 while the abnormal case iris image samples are shown in the second row of Fig. 1. Such abnormal cases are rare and very hard to collect. Nevertheless, they are highly valuable in terms of algorithm training and fine-tuning. With a large number of such abnormal images, a better deep neural network with high robustness can be trained [16]. Such a model is able to successfully perform iris segmentation as well as other necessary processes for abnormal iris images.

The class imbalance issue dwells in many datasets for real-world applications. One of the most popular approaches to overcome the issue is by resampling the data. Former researcher by [17] uses a one-class learning method by support vector machine (SVM) and under-sampling technique to pre-process the data. This technique shows
satisfactory performance on nine real-life datasets collected from the University of California, Irvine (UCI) machine learning repository (UCI Machine Learning Repository) without harming the majority class. The second approach is based on an oversampling method to increase the number of data in a minority class named the synthetic minority oversampling technique (SMOTE). Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates by [18] significantly outperforms the other methods being compared on the China stock market and accounting research database. The other method that is different from data-level techniques to classify imbalanced data is proposed by [19] and [11].

In the first-mentioned study, the authors employed a renowned statistical logistic regression approach called logistic regression for imbalanced learning-based clustering (LRILC) to tackle the imbalanced learning problem at preprocessing phase. The latter uses a weighted pattern matching approach. In the paper [11], the feature selection task is applied for learning by combining the fireworks algorithm. The paper yields positive results on datasets from the KEEL dataset repository [20]. Despite the acceptable results, the generated data may not be sufficiently realistic since these traditional methods are more concerned with local data information.

Notwithstanding that image collection for special cases is a very difficult task, recent advancements in the research of GAN [21]–[23] open the possibility of a new solution to this problem. In 2018, a method called Iris-GAN [24] has been developed using simple deep convolutional (DC) GAN model. This model is able to generate realistic iris images over 120 epochs of training. The diversity of the output of this finding, however, cannot be controlled. The other work from 2019 called RaSGAN [25] emphasizes the relativistic of the GAN to be more generalized by updating its loss function. The result from this study is reported to have high resemblance with the real iris images. Yet, the method is not focusing on rebalancing the iris images dataset.

We propose a new parametric GAN model with the ability to learn from example images and generate images for specific purposes. Inspired by CWGAN-GP [23], we want to achieve superior results by paying more attention to the true distribution of data instead of local data information. The GAN employed in this study is a combination of a WGAN-GP and an auxiliary classifier network. The addition of an auxiliary classifier network helps WGAN-WP to discern the additional information. CWGAN-GP’s performance is evaluated on 15 different datasets extracted from the UCI Machine Learning repository with outstanding results. Nonetheless, we limit our goal to generate iris images of the desired quality and accurately classify these images under particular network architecture to resolve better address training instability and CWGAN-GP’s limitation of the training effort. However, the architecture of the propose parametric GAN is generic, meaning that it is able to be applied in all kinds of image generation tasks where scarce data problem occurs. In our experimental section, the proposed method was also applied to MNIST handwritten digits datasets [26] and it can generate images with higher quality than those of traditional GANs.

In this study, our main contribution is summarized here:

1. Our approach is able to generate high-quality synthesized iris images, not merely enlarging the quantity.
2. Although we limit our study to generate iris images, our approach works well to reconstruct the MNIST handwritten dataset by adding center loss to the model. This implies that our approach has high generalizability toward the different tasks.
3. From the perspective of industrial importance, we can generate and synthesize an unlimited number of iris images by using our proposed method. Thus, in the future, it can be utilized to train many different iris segmentation models for robust iris detection or recognition.

The rest of the paper is organized as follows. In Sect. 2, we review the related past works about GAN and its variants. In Sect. 3, the architecture and training method of the proposed GAN is described. Section 4 presents the experimental results and discussion. Finally, in Sect. 5, the conclusion of this work is presented.

2. Literature Review

2.1 GAN

GANs [1] are generative networks intended for adversarial learning. The goal of the generative model $G$ is to generate samples that cannot be identified as real or generated by the discriminative model $D$. The goal of $D$ is to successfully distinguish real samples from the samples generated by $G$.

The value function of $D$ and $G$, $V(G, D)$, refers to a 2-player minimax game and is expressed as follows:

$$
\min_G \max_D V(G, D) = 
E_{x \sim P_{data}}[\log D(x)] + E_{z \sim P_z}[\log(1 - D(G(z)))],
$$

(1)
where $P_{\text{data}}$ is the distribution of the real data; $P_x$ is the distribution of the random noise $z$, and $P_z$ is a uniform or Gaussian distribution.

Essentially, the similarity between the two probability distributions is quantified by Jensen–Shannon (JS) divergence. With symmetric and smoother arguments [27], using this divergence as a loss function when the discriminator is optimal is considered as the ground behind GANs’ great success [28]. But the triumph is nevertheless having some adverse effects in the training process.

2.2 WGAN

WGANs [3] feature Wasserstein distances (i.e., earth mover’s distances), which replace the JS divergence in the original GANs. When two distribution patterns do not overlap, the JS divergence can only return a single constant, whereas Wasserstein distances can indicate the distance between the two patterns. The introduction of WGANs addressed the training instability in GANs, eliminated the need to cautiously balance the training levels of the generator and discriminator, mitigated the problems involving crash models, and guaranteed the diversity of generated samples.

The Wasserstein distance:

$$W(P_r, P_g) = \inf_{\gamma \sim \prod(P_r, P_g)} \mathbb{E}(x, y) \|x - y\|,$$

(2)

where $\prod(P_r, P_g)$ represents the aggregation of all the potential joint distributions of $P_r$ and $P_g$. Each potential joint distribution $\gamma$ consists of a real sample $x$ and a generated sample $y$, and the distance between these samples is calculated. The minimum of the expected distance values in all the joint distribution aggregates is the Wasserstein distance.

The target function of a WGAN, which is based on the Kantorovich–Rubinstein duality, is expressed as follows:

$$W(P_r, P_g) = \sup_{\|f\|_{TV} \leq 1} \mathbb{E}_{x \sim P_r}[f(x)] - \mathbb{E}_{z \sim P_g}[f(x)],$$

(3)

assuming that the fitting function of $D$ is a 1-Lipschitz continuous function. For Lipschitz continuity to hold, the following additional restriction is placed on a continuous function for all $x_1$ and $x_2$ in the domain of definition, there exists a constant $K \geq 0$ such that:

$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2|,$$

(4)

In this expression, $K$ is termed the Lipschitz constant of $f$. Specifically, Lipschitz continuity restricts the maximal range of local changes in a continuous function. To fulfill this goal, all the parameters are restricted to the range of $[-c, c]$ through weight clipping, which is a gradient clipping approach.

2.3 WGAN-GP

Although WGANs employ Wasserstein distances to generate value functions superior to those generated through JS divergence, they still exhibit the training instability found in GANs. Weight clipping may lead to gradient explosion when $c$ is extremely large and cause gradient vanishing when $c$ is extremely small, significantly undermining the fitting ability of a WGAN with complex functions. In WGANs-GP [4], weight clipping is replaced with a gradient penalty to achieve Lipschitz continuity.

In practice, the norm of the discriminator’s gradient is penalized with respect to its input. When a function is 1-Lipschitz, its gradient change is no greater than 1. The function is given by:

$$D \leq 1 - \text{Lipschitz} \iff \|\nabla_x D(x)\| \leq 1 \text{ for all } x,$$

(5)

Combining this function with the WGAN target function yields the following:

$$W_{gp}(P_r, P_g) = \left(\max_{D} \mathbb{E}_{x \sim P_r}[D(x)] - \mathbb{E}_{z \sim P_g}[D(x)]\right) - \lambda \int_{0}^{1} \max(0, \|\nabla_x D(x)\| - 1) dx,$$

(6)

where

$$\|\nabla_x D(x)\| \leq 1 \text{ for all } x$$

Practically, integrating over all $x$ values is impossible. Therefore, the gradient change of $x$ as sampled through $P_x$ is determined as 1. Specifically, $P_x$ represents the uniform distribution of the straight line between a random sample of the real sample distribution $P_r$ and a random sample of the generated sample distribution $P_g$. In a WGAN-GP, a gradient closer to 1 is more desirable; any gradient larger than 1 is penalized. Thus, the target function is expressed as

$$W_{gp}(P_r, P_g) = \left(\max_{D} \mathbb{E}_{x \sim P_r}[D(x)] - \mathbb{E}_{z \sim P_g}[D(x)]\right) + \lambda \mathbb{E}_{x \sim P_r}[\|\nabla_x D(x)\| - 1]^2),$$

(7)

where $\lambda$ is the gradient penalty coefficient.

2.4 CGAN and ACGAN

A conditional GAN (CGAN) was developed by [29], as the expansion of the original GAN. Because GANs do not require preliminarily trained models, their generated samples tend to be excessively free. Therefore, restrictions are implemented in CGANs. The conditional variable $y$ is added in both the generator and discriminator. The target function of a CGAN is a two-player minimax game with a conditional probability:

$$\min_{G} \max_{D} V(G, D) = \mathbb{E}_{x \sim P_{\text{data}}(x)}[\log D(x | y)] + \mathbb{E}_{z \sim P(z)}[\log(1 - D(G(z | y)))]$$

(8)

Consequently, conditional image synthesis with an auxiliary classifier GAN (ACGAN) [30] was proposed as a new CGAN framework. Similar to CGAN, the generator of an ACGAN involves input labels and noises. However, an ACGAN features an auxiliary discriminator, which doubles as a classifier. This classifier identifies the types of input
samples and is thus not limited to distinguishing real and generated samples. The target function of an ACGAN is expressed as follows:

$$L_s = E[\log P(S = \text{real} | X_{\text{real}})] + E[\log P(S = \text{fake} | X_{\text{fake}})],$$

and

$$L_c = E[\log P(C = c | X_{\text{real}})] + E[\log P(C = c | X_{\text{fake}})],$$

The goal of the discriminator in the AC-GAN is to maximize the value of $L_s + L_c$, whereas that of the generator is to maximize the value of $L_c - L_s$. The subsequent generator learns a dormant space representation that is independent of the class label, dissimilar to the restrictive GAN.

2.5 CWGAN-GP

The advancement of WGAN-GP is dilated out to CWGAN-GP if both the discriminator and generator are conditioned on some additional extra information $y$ derived from the CGAN. Here, $y$ could be any sort of information and in this investigation, $y$ is the class name. In the discriminator, we link both $P_r$ and $P_g$ with $y$ in the joint shrouded portrayal. In the generator, we additionally connect $y$ with $P(z)$ in a similar portrayal. Officially, with the same gradient penalty coefficient presented as $\lambda$, the target esteem work between the generator and the discriminator is expressed as follows:

$$\min_G \max_D V(G, D) = (E_{x \sim P_r}[D(x | y)] - E_{z \sim P}(D(x | y)]) + \lambda E_{x \sim P}[\|\nabla_x D(x | y)\| - 1]^2],$$

3. Method

3.1 Proposed Method

Conventional classifiers face a challenging issue of handling imbalanced image datasets [31]. A restricted WGAN-GP warrants in-depth research handling the mentioned issue. Inspired by the success of CWGAN-GP, extended development of WGAN-GP by conditioning both of the discriminator and generator with CGAN, we want to adopt a WGAN-GP-based network that is capable of reinforcing the classification performance of imbalanced iris images data. This study, thus, incorporated an auxiliary classifier into a WGAN-GP to enhance the quality of generated samples and enable their classification.

Our proposed method produces increasingly sensible and realistic data while overcoming the problems of training in GAN. Instead of being based on the local information of the datasets, just like WGAN-GP, our proposed method depends on the global information of the real data distribution. However, in contrary with WGAN-GP that does not completely think about the extra restrictive information, our proposed method not just offers a portion of the benefits of WGAN-GP yet in addition draws on the upsides of the auxiliary classifier of CGAN, and accordingly builds the nature of generative engineered data.

In Fig. 2, $D$ and $G$ represent the discriminator and generator, respectively; $C$ represents the classifier; $z$ represents a random noise; $y_1$ represents the generated-image type; $x$ represents the real image; $y_2$ represents the real-image type; $G(z_{y_1})$ represents the generated image (referred as “fake image” in Fig. 2); $D(x)$ represents the identification results from the discriminator, and $C(x)$ represents the classification results from the classifier. The algorithm of the WGAN-GP is divided into the following steps:

1. Combine $z$ and $y_1$ to obtain a vector to input into $G$, thus generating the sample $G(z_{y_1})$.
2. Transmit $x$ and $G(z_{y_1})$ to $D$; use the WGAN-GP algorithm to calculate the Wasserstein distance between $x$ and $G(z_{y_1})$.
3. Fix the parameters of $G$, acquire the loss function of $D$ according to the Wasserstein distance between $x$ and $G(z_{y_1})$, and update the parameters of $D$ accordingly. Repeat this step k times to maximize the accuracy of $D$.
4. Fix the parameters of $D$ and use the calculation result of $G(z_{y_1})$ in $D$ to update the parameters of $G$, to increase the similarity between $G(z_{y_1})$ and $x$.
5. Enter $x$ and $y_2$ into $C$ to classify the sample type of $x$; calculate the softmax loss between the sample type classified by $C$ and the real-image type $y_2$ to update the parameters of $C$.
6. Enter $G(z_{y_1})$ into $C$ to classify its sample type; calculate the softmax loss between the sample type classified by $C$ and the generated-image type $y_1$ to update the parameters of $G$.
7. Repeat these steps until the parameters converge.
3.2 Extension Method

For the MNIST digit image generation, a method for the situation that requires a large number of various types of generated images was developed. Revisiting our algorithm, instead of merely taking the softmax loss calculation, we add a center loss to the softmax loss in Step 5 for updating the parameters of the discriminator. Essentially, softmax loss trains a network to learn image features and output a probability over the $K$ classes for each image in such cases and it is commonly used for multi-class classification [32].

Let $x_i$ and $y_i$ be the input feature and its label corresponding to class $i$, respectively, we have the formal expression of softmax loss as follow:

$$L_S = \frac{1}{N} \sum_i L_{S,i} = \frac{1}{N} \sum_i -\log \left( \frac{e^{y_i}}{\sum_j e^{j}} \right),$$  \hspace{1cm} (12)

where $N$ is the number of images in the training process and $f_j$ denotes the $j$-th ($j \in [1,K]$) element of the class score vector $f$. Nonetheless, center loss [33] was designed to address the issue of softmax loss that only learns separable features and barely discriminative.

$$L_C = \frac{1}{2} \sum_i \|x_i - c_{y_i}\|^2_2,$$  \hspace{1cm} (13)

As formulated in Eq. (13), $c_{y_i}$ denotes the class center of the features that should be updated as the features changed in every mini-batch. With a hyperparameter $\alpha$ that controls the learning rates of the centers, this detail effectively portrays the intraclass varieties. The update is given by

$$c_j^{t+1} = c_j^t - \alpha \Delta c_j^t,$$  \hspace{1cm} (14)

Add Eq. (12) and Eq. (13) to be

$$L = L_S + \lambda L_C,$$  \hspace{1cm} (15)

where we use another scalar $\lambda$ for balancing the two-loss function. Thus, the interclass dispersion and intraclass compactness could be simultaneously increased, which enhanced the robustness of the discriminator. Subsequently, the classification accuracy of the generated samples in the experiment was improved.

4. Experimental Results and Discussion

4.1 MNIST

A training set comprising 55,000 MNIST handwritten digits was employed for digit image generation. The generated images from the original CGAN, the generated images from our proposed method based on only softmax loss, and the generated images from our proposed method based on softmax loss with center loss were compared. The comparative summary in this area is presented in Table 1.

|                           | Wasserstein Auxiliary Classifier | Wasserstein Gradient penalty | Softmax loss | Center loss |
|---------------------------|---------------------------------|-----------------------------|--------------|-------------|
| CGAN                      | +                               | -                           | +            | -           |
| Proposed Method           | +                               | +                           | +            | -           |
| Extension Method          | +                               | +                           | +            | +           |

Fig. 3 MNIST samples generated from the CGAN.

Fig. 4 (a) Generated digit samples using the proposed method based on softmax loss only. (b) Generated digit samples using an extended proposed method based on softmax loss + center loss.

For this comparison experiment, the samples were generated at a resolution of $32 \times 32$ employing a single GTX-1080Ti graphics card for training over 20000 iterations. As shown in Fig. 3, despite the generator successfully learns the desired structure, some of the generated images from CGAN can appear quite obscure. On average, four out of twenty samples we present in the Fig. 3 are oddly shaped and appearing to be similar to the other class. The experimental results revealed that our proposed method yields the most satisfactory results which can be seen in Fig. 4. Our
proposed method is able to generate feasible MNIST digits with less noise left but the distinctness is at a loss compared to the original CGAN, shown in Fig. 4 (a). By adding center loss into the loss function in the extension method exhibits a great improvement to the generated MNIST digits which is shown in Fig. 4 (b). Numbers of different classes are clearly distinguished and the variation inside each class is less significant. In that way, applying the softmax loss + center loss approach is evident to enhance the classifier performance.

4.2 CASIA-Iris-Thousand

For iris image generation, images from the CASIA-Iris-Thousand database [12] are selected as the training data. CASIA-Iris-Thousand is the subset of the CASIA-Iris-V4 database with the major sources of intra-class variations of eyeglasses and specular reflection. This dataset contains 2850 images with oversized iris pupils, 2823 images with undersized pupils, and 2876 with glasses.

A single GTX-1080Ti graphics card was employed for training and 20000 iterations were performed. The samples were generated at a resolution of 128 × 128. Since the gradient penalty turns down the sensitivity of model performance in WGAN, tweaking the hyperparameters is inessential. The generated images from the original GAN, those from WGAN-GP, and those from our proposed method based on WGAN-WP with the addition of an auxiliary classifier were compared. The comparative summary in this area is presented in Table 2.

The generated images example can be seen in Fig’s 5, 6, 7, and 8. In Fig. 5 we can see that the original GAN produces inadequate iris images in quality. We exhibit some generated examples with the unnatural shape of eyes such as blurry eyes images, images with pupils outside the eye, and even images with two pupils within an image. The following generated examples from the implementation of WGAN-GP are shown in Fig. 6. WGAN-GP clearly provides better performance in converging and generating results with less peculiar images.

This result implies that Wasserstein distance outperforms JS divergence in measuring the distance between the probability distribution of real data and the generated ones. Despite the exemplary performance, the variation of the generated iris images is still low. Thus, we use an auxiliary classifier to learn more cases such as iris images with glasses, undersized, and oversized pupils to improve the WGAN-WP. Figure 7 shows that our proposed method yields superior performance which is also able to produce realistic-looking iris images of an abnormal case with or without glasses. Figure 8, on the other hand, shows the ability of our proposed method to produce a natural-looking iris image with both undersized and oversized pupils. In both Fig. 7 and Fig. 8, we can see the iris images are generated with high quality as well. Here we can infer the use of the addition of auxiliary classifier in WGAN-GP as the success of our proposed method.

4.3 Evaluation with Frechet Inception Distance (FID)

The Frechet distance which is also known as Wassertain-2
distance has been used to extend the original Inception Score (IS) [34] into the Frechet Inception Distance (FID) [35]. The FID score measures the difference between two Gaussians (the real and synthetic images) by:

\[
FID = ||\mu_r - \mu_s||^2 + T_r\left(\Sigma_r + \Sigma_s - 2\sqrt{\Sigma_r\Sigma_s}\right),
\]

where \(X_r \sim N(\mu_r, \Sigma_r)\) for the real images, \(X_s \sim N(\mu_s, \Sigma_s)\) for the synthetic images, and \(T_r\) represents the trace of the covariance of the corresponding matrix.

In this evaluation we use the CASIA-Iris-Thousand and train the iris images with and without glasses separately and obtain the overall score of 0.690. The same dataset is also used in [24] with FID score of 42.1. Beside CASIA-Iris-Thousand, they also use IIT-Delhi database which obtain lower FID score of 41.08. Lastly, in [25] they obtain an overall FID score of 39.17 from several dataset including CASIA-Iris-Fake. Since the mentioned works do not provide official code, we present a comparison (see Fig. 9) based on their reported result. Furthermore, we can see that our proposed method has the lowest FID score which means that we have highest similarity between our real and synthetic iris images.

4.4 Model Complexity and Limitations

In this paper, our contribution is to add center loss in CWGAN-GP’s formulation. In that way, we merely need to add a new loss function into the last layer when performing the backpropagation. Therefore, the complexity of our proposed method remains the same as the complexity of the generator and the discriminator of the original CWGAN-GP. Conducting experiments with GAN, it is common to spend much time waiting for the training of the generator and the discriminator. In our case, since we not only train the generator and the discriminator but also the classifier, each experiment using MNIST dataset and CASIA-Iris dataset takes 2 – 3 days of training.

5. Conclusion

Collecting special images for training a deep neural network for specific purposes is often challenging because they are rare in nature and it asks for a large amount of human effort to collect and judge if they are qualified for training. For iris images, normal iris images are easy to collect, but those with misaligned iris locations, oversized or undersized pupils, with glasses, and with excessive reflection are insufficient. In the present study, an auxiliary classifier was added to the WGAN-GP network to generate high-quality images and accurately classify them, thereby yielding realistic and specific images and overcoming the insufficiency in image collection for specific purposes.

For iris image generation, a generative network that generated specific image types with oversized and undersized pupils, as well as eye with or without glasses, was successfully trained, thereby overcoming the insufficiency in iris image of specific types. For MNIST handwritten digit
image generation, the extended CWGAN-GP model developed in this study also generated clearer and more accurate images than the CGAN did.

References

[1] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” arXiv preprint arXiv:1406.2661, 2014.

[2] M. Arjovsky and L. Bottou, “Towards principled methods for training generative adversarial networks,” arXiv preprint arXiv:1701.04862, 2017.

[3] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” International Conference on Machine Learning, pp.214–223, PMLR, 2017.

[4] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved training of wasserstein gans,” arXiv preprint arXiv:1704.00028, 2017.

[5] O.M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015.

[6] Y. Taigman, M. Yang, M.A. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” Proc. IEEE conference on computer vision and pattern recognition, pp.1701–1708, 2014.

[7] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” Proc. IEEE conference on computer vision and pattern recognition, pp.815–823, 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770–778, 2016.

[9] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol.39, 06/04, 2015.

[10] S. Bell, C. Zitnick, K. Bala, and R. Girshick, “Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks,” 12/13, 2015.

[11] N. Sreeja, “A weighted pattern matching approach for classification of imbalanced data with a firework-based algorithm for feature selection,” Connection Science, vol.31, no.2, pp.143–168, 2019.

[12] “Casia iris image database,” ed. C. A. o. S. I. o. A. (CASIA).

[13] S. Wang, L.L. Minku, N. Chawla, and X. Yao, “Learning from data imbalanced data set repository, integration of algorithms and experimental analysis framework,” Journal of Multiple-Valued Logic & Soft Computing, vol.17, 2011.

[14] Q. Wu, B. Zhu, B. Yong, Y. Wei, X. Jiang, R. Zhou, and Q. Zhou, “Clothgan: Generation of fashionable dunhuang clothes using generative adversarial networks,” Connection Science, pp.1–18, 2020.

[15] G. Douzas and F. Baco, “Effective data generation for imbalanced learning using conditional generative adversarial networks,” Expert Systems with applications, vol.91, pp.464–471, 2018.

[16] M. Zheng, T. Li, R. Zhu, Y. Tang, M. Tang, L. Lin, and Z. Ma, “Conditional wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification,” Information Sciences, vol.512, pp.1009–1023, 2020.

[17] S. Minaee and A. Abdolrasahidi, “Iris-gan: Learning to generate realistic iris images using convolutional gan,” 2018.

[18] S. Yadav, C. Chen, and A. Ross, “Synthesizing iris images using rasgan with application in presentation attack detection,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.2422–2430, 2019.

[19] Y. LeCun, C. Cortes, and C. Burges, “Mnist handwritten digit database,” Florham Park, NJ, USA, 2010.

[20] D.M. Endres and J.E. Schindelin, “A new metric for probability distributions,” IEEE Trans. Inf. Theory, vol.49, no.7, pp.1858–1860, 2003.

[21] F. Huszár, “How (not) to train your generative model: Scheduled sampling, likelihood, adversary?,” arXiv preprint arXiv:1511.05101, 2015.

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

[23] A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier gans,” International Conference on Machine Learning, pp.2642–2651, 2017.

[24] S. Guldas, “Efficient learning on imbalanced image set,” International Journal of Computer Sciences and Engineering, vol.6, no.10, pp.121–126, 10/31, 2018.

[25] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “Sphereface: Deep hypersphere embedding for face recognition,” Proc. IEEE conference on computer vision and pattern recognition, pp.212–220, 2017.

[26] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A discriminative feature learning approach for deep face recognition,” European conference on computer vision, pp.499–515, Springer, 2016.

[27] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” arXiv preprint arXiv:1606.03498, 2016.

[28] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” arXiv preprint arXiv:1706.08500, 2017.
Yung-Hui Li  received the B.S. degree from the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, in 1995, the M.S. degree from the Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA, in 1998, and the Ph.D. degree from the Language Technology Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, in 2010. He is currently an Associate Professor with the Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan, and the Director of AI Research Center at Hon Hai Research Institute in Taipei, Taiwan. His current research interests include long-range iris recognition, including extended DoF for iris image acquisition, blur iris image segmentation, iris mask generation, and iris image super-resolution; face recognition; eye-tracking; human–computer interaction; computer vision; pattern recognition; and machine learning. Prof. Li is currently a member of the IEEE Computer Society and Signal Processing Society.

Muhammad Saqlain Aslam  received the B.S. degree in Software Engineering from Government College University, Faisalabad, Pakistan, in 2013, and the M.S. degree in Software Engineering from Beijing Institute of Technology, Beijing, China, in 2017. He is currently pursuing the PhD. Degree in Computer Science in the Department of Computer Science and Information Engineering of National Central University, Taiwan. His current research interests include machine learning, deep learning, biometric recognition, and computer vision.

Latifa Nabila Harfiya  received her B.S. in computer Science from University of Brawijaya, Indonesia, in 2017. She was a student member of artificial intelligent lab in University of Brawijaya. She is now actively pursuing her PhD degree in computer science at Dept. of Computer Science and Information Engineering of National Central University, Taiwan. She is interested in signature recognition, face recognition, and her current focus of area, blood pressure estimation.

Ching-Chun Chang  received the PhD degree from the Department of Computer Science, University of Warwick, UK, in 2019. He received the BBA degree in information management from National Central University, Taiwan, in 2015. He was granted the Marie-Curie fellowship in 2017. He engaged in a short-term scientific mission supported by European cooperation in science and technology actions at the Faculty of Computer Science, Otto von Guericke University Magdeburg, Germany, in 2016. He participated in a research and innovation staff exchange scheme supported by Marie Skłodowska-Curie actions at the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, USA, during 2017. He was a visiting scholar at the School of Computer and Mathematics, Charles Sturt University, Australia in 2018 and at the School of Information Technology, Deakin University, Australia in 2019. He has been a research fellow at Department of Electronic Engineering, Tsinghua University, China since 2020. His research interests include steganography, multimedia security, machine learning, natural language processing, and computational linguistics.