This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2804278, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Improved Boundary Equilibrium Generative Adversarial Networks

YANCHUN LI[1], NANFENG XIAO[1], AND WANLI OUYANG[2], (Senior Member, IEEE)
[1]School of Computer Science and Engineering, South China University of Technology, Guangzhou, 510006, China (e-mail: 201510105258@mail.scut.edu.cn)
[2]School of Electrical and Information Engineering, University of Sydney, Sydney, Australia (e-mail: wanli.ouyang@gmail.com)

Corresponding author: Yanchun Li (e-mail: 201510105258@mail.scut.edu.cn).

This work is supported by the National Natural Science Foundation of China under Grant 61573145, the Public Research and Capacity Building of Guangdong Province of China under Grant 2014B010104001 and the Basic and Applied Basic Research of Guangdong Province of China under Grant 2015A030308018.

ABSTRACT Boundary equilibrium generative adversarial networks (BEGANs) can generate impressively realistic face images, but there is a trade-off between the quality and the diversity of generated images. Based on BEGANs, we propose an effective approach to generate images with higher quality and better diversity. By adding a second loss function (a denoising loss) to the discriminator, the discriminator can learn more useful informations about the distribution of real images. Naturally, the ability of discriminator in distinguishing between real and generated images is improved, which further guides the generator to produce more realistic images to confuse the discriminator. We also find that using technique of batch normalization in BEGANs architecture can improve the diversity of generated images. By using batch normalization and adding a denoising loss to the objective of discriminator, we achieve comparative generations on CIFAR-10 and CelebA datasets. In addition, we evaluate the effect of several techniques on BEGANs framework through "Inception-Score", a measure which has been found to correlate well with human assessment of generated samples.

INDEX TERMS generative adversarial networks (GANs); boundary equilibrium generative adversarial networks (BEGANs); deep generative model; image generation.

I. INTRODUCTION

Generating realistic-looking images has been a longstanding goal in machine learning. Deep models were found to be effective for this goal. In recent years, Variational Auto-encoders (VAEs) [1, 2] and Generative Adversarial Networks (GANs) [3] are the two most prominent ones and have shown their effectiveness. In this paper, we focus on GAN-based approaches.

A typical GAN usually simultaneously trains two models: a generative model $G(z)$ to synthesize samples given some random source z, and a discriminative model $D(x)$ to differentiate between real and synthesized samples. GANs can produce visually appealing images, usually regarded as the best but so far no good way to quantify this [3]. Goodfellow et al. [4] first proposed Generative Adversarial Networks, analyzed the theory of GANs and explained the learning process based on a game theoretic scenario in 2014. And then GANs have achieved impressive results in many specific tasks, such as image generation [5, 6], image super-resolution [7], image to image translation [8, 9], video prediction[10], text generation [11] and text to image synthesis [12].

In practice, GANs also have been known to be unstable in the training stage and easily suffer from modal collapse, in which just one image is learned [13]. Many recent works focus on stabilizing the training process via analyzing the objective functions of GANs. McGANs [14] used mean and covariance feature matching as objective function. Loss-Sensitive GANs [15] learned a loss function which can quantify the quality of generated samples and used this loss function to generate high-quality images. Energy Based GANs [16] (EBGANs) were proposed as a class of GANs that aimed to model the discriminator as an energy function. Auto-encoder was used as the discriminator for the first time in [16]. Least Squares GANs [17] adopt the least square loss function for the discriminator. More recently, Wasserstein GANs (WGANs) [18] used Earth Mover Distance as an objective for training GANs, and Ishaan Gulrajani et al. [11] found that applying Earth Mover Distance with gradient penalty as loss function can make Wasserstein GANs [18] converge faster and generate images with higher-quality. Boundary equilibrium generative adversarial networks (BE-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2804278, IEEE Access

GANs) [19], a simple and robust architecture, optimized the lower bound of the Wasserstein distance between auto-encoder loss distributions of real and synthesized samples.

In this paper, we address the problem of training stability and quality of generated images. We propose to augment the objective of auto-encoder (discriminator) with an additional loss function, so that the auto-encoder can learn more about real-data distribution and the ability of auto-encoder in distinguishing between real-data and generated-data is more powerful, which implicitly guides the generator to produce more realistic data.

In summary, our contributions are as follows:

- We propose BEGANs with denoising loss, a simple, easy to implement but effective method to improve the quality of generated images. With experiments on CelebA dataset, we show the effectiveness of denoising loss on reducing the noise-like regions and improving the quality of generated face images.
- We demonstrate that adding denoising loss to the discriminator can improve the training and converging stability, theoretically, this method can be applied in any model, which employs auto-encoder as discriminator in GANs.
- We empirically show the effectiveness of batch normalization on improving the diversity of generated images. By adding denoising loss and batch normalization, we generate higher quality and diversity face images over BEGANs.
- We evaluate the effectiveness of several techniques on BEGANs framework through “Inception-score", a measure which has been found to correlate well with human assessment of generated samples.

II. BACKGROUND

As our methods based on BEGANs [19], in this section, we first introduce the objective functions of BEGANs, then according to the experimental results, we illustrate the shortage of BEGANs. Next, we briefly review two techniques related to our method, i.e., adding denoising loss to discriminator and batch normalization.

A. BOUNDARY EQUILIBRIUM GENERATIVE ADVERSARIAL NETWORKS

Denote the discriminator by D and the generator by G. BEGANs [19] used auto-encoder as the discriminator, as in [16]. BEGANs are simple and robust architectures with an easy way to control the balance between the discriminator and the generator[19]. BEGANs matched the auto-encoder loss distributions of real and generated data by optimizing the Wasserstein distance. In the following, we first introduce the lower bound of Wasserstein distance of auto-encoders and the objective of BEGANs, and then analyze the limitations of BEGANs.

Let $L(v) = \|D(v) - v\|_1$ be the L_1 loss of auto-encoder, let μ_1 and μ_2 be, respectively, the distributions of real-data and generated-data auto-encoder losses. Let $\gamma \in \Gamma(\mu_1, \mu_2)$ be the set all of couplings of μ_1 and μ_2. And let $m_1 \in R$ and $m_2 \in R$ be their mean respectively. The definition of Wasserstein distance is:

$$W_1(\mu_1, \mu_2) = \inf_{\gamma\in\Gamma(\mu_1,\mu_2)} \mathbf{E}_{(x_1,x_2)\sim\gamma}[\|x_1 - x_2\|_1] \quad (1)$$

By applying Jensen’s inequality to Eq.(1), the lower bound of $W_1(\mu_1, \mu_2)$ is:

$$\inf \mathbf{E}[\|x_1 - x_2\|]_1 \geq \inf \|\mathbf{E}[x_1 - x_2]\|_1 = \|m_1 - m_2\|_1 \quad (2)$$

Where $\inf$ denotes the infimum. Eq.(2) implies that the lower bound of $W_1(\mu_1, \mu_1)$ is $\|m_1 - m_2\|_1$. To maximize the distance between real and generated data, the only two solutions of $\|m_1 - m_2\|_1$ is to have $m_1 \to 0, m_2 \to \infty$ or $m_1 \to \infty, m_2 \to 0$. BEGANs [19] chose $m_1 \to 0, m_2 \to \infty$, as minimizing m_1 is equivalent to reconstructing real data. The whole objective function of BEGANs [19] was defined as follow:

$$\begin{cases} L_D(x, z) = L(x) - k_t * L(G(z)) \\ L_G(z) = L(G(z)) \\ k_{t+1} = k_t + \lambda_k(\gamma \cdot L(x) - L(G(z))) \end{cases} \quad (3)$$

where $L(x) = \|D(x) - x\|_1$ is the auto-encoder L_1 loss of real data, and $L(G(z)) = \|D(G(z)) - G(z)\|_1$ is the auto-encoder L_1 loss of generated data. The variable $k_t \in [0, 1]$ controls the emphasis of generator losses when training the discriminator and $k_0 = 0$. $\gamma = L(G(z))/L(x)$ maintains the balance between the auto-encoder loss of real-data and generated-data. γ is also an indicator of diversity with small values meaning less diversity. λ_k is the learning late of k, which is 0.001 in experiments. The auto-encoder in BEGANs reconstructs images and discriminates real images from generated images simultaneously. BEGANs [19] also proposed an approximate measure of convergence: $M_{global} = L(x) + |\gamma \cdot L(x) - L(G(z))|$, where $|\gamma \cdot L(x) - L(G(z))|$ is the absolute value of $\gamma \cdot L(x) - L(G(z))$. We adopt this measure in experiments.

B. THE SHORTAGE OF BEGANS

Despite BEGANs made some progress on image quality and measuring convergence, there are still many problems that need to be improved. As show in Fig.1[19], at low values of $\gamma = 0.3$, the generated images looks uniform with many noise-like regions, while at high values ($\gamma = 0.7$), the diversity of the generated images increases but the quality declines. Another shortage of BEGANs is that the generator cannot learn the low-probability features. For example, BEGANs almost cannot generate older people faces, and cannot generate glasses even with highest diversity value $\gamma = 1$(We performed the program with $\gamma = 1$ and produced 12800 images with the trained model. In the 12800 generated images, we observed no glasses and hardly old faces). This point was also supported by Berthelot et al, as they stated “However we did not see glasses, we see few older people"[19]. To reduce the noise-like regions in the generated images, we add a denoising loss to the discriminator and to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2804278, IEEE Access

improve the diversity we introduce batch normalization. In the following, we briefly review these two methods. Note that some noise-like regions on Fig.1 are marked with red circles for highlight.

FIGURE 1. Images generated by BEGANs $\gamma \in \{0.3, 0.5, 0.7\}$. Note: some noise-like regions are marked with red circles at $\gamma = 0.3$.

C. DENOISING AUTO-ENCODER

The denoising auto-encoder [20], is trained to minimize the following denoising criterion:

$$L_{DAE} = \mathbf{E}[L(x, r(N(x)))] \tag{4}$$

Where $N(x)$ is a stochastic corruption of x and the expectation in the right of Eq.(4) is over the training distribution and the corruption noise source. For easy mathematical calculation, usually apply squared loss and Gaussian noise corruption, which means $L_{DAE} = \|r(N(x)) - x\|_2^2$. According to Alain and Bengio [21], a suitably trained[1] denoising auto-encoder can estimate some local characteristics of the data density, such as the first derivative (score) and second derivative of the log-density and the local mean. They further showed that when the denoising auto-encoder has been suitably trained, the quantity L_{DAE} denoising reconstruction loss $\|r(N(x)) - x\|_2^2$ assessed the score of the data density, up to a multiplicative constant, which is illustrated by the following Eq.(5):

$$\|r(N(x)) - x\|_2^2 \propto \frac{\partial \log p(x)}{\partial x} \tag{5}$$

D. BATCH NORMALIZATION

First introduced by Ioffe and Sergey [22], Batch normalization was proposed to alleviate the internal covariate shift by incorporating a normalization step and a scale and shift step before the nonlinearity in each layer. For batch normalization, only two parameters per activation are added, and they can be updated with back-propagation. Batch normalization enjoys several merits, such as fast training, better performance, and low sensitivity to initialization. For further details on batch normalization, please refer to [22]. In this paper, we confirm the state that batch normalization can improve the performance of BEGANs [19], especially in improving the diversity of generated images, which will be demonstrated in experiment section.

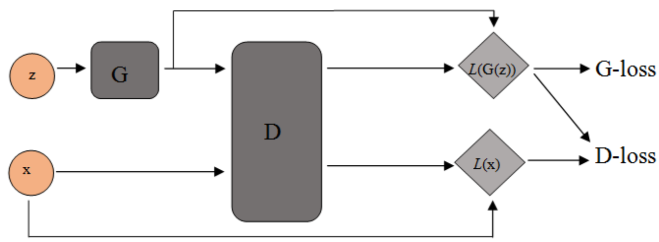
[1] With infinite training data, perfect minimization and isotropic Gaussian noise of some standard deviation σ

III. METHOD AND ARCHITECTURES

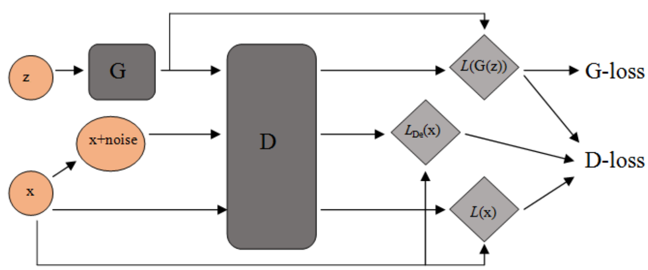
As BEGANs devise auto-encoder as discriminator, a natural idea is to add denoising loss to the discriminator, so that the discriminator can learn much useful information about real-data distribution. The generator of our models is trained as BEGANs, while the discriminator of our models is trained using the following loss function:

$$L_D(x, z) = L(x) - k_t * L(G(z)) + \lambda_{noise} \cdot L_{DAE} \tag{6}$$

where $L_{DAE} = \|D(x + noise) - x\|_2^2$ is the denoising loss of discriminator. λ_{noise} is the weighting coefficient of denoising loss, which controls the proportion of denoising loss during training the discriminator. All experiments in this paper use $\lambda_{noise} = 2$, which we found to work well on CIFAR-10 dataset and CelebA dataset. According to Eq.(6), the objective loss function of discriminator includes optimizing the lower bound of the Wasserstein distance between real-data and fake-data auto-encoder loss, and optimizing the denoising loss between real-data and the corruption noise source. Other symbols have the same meaning as that in Eq.(3). The dataflow of our models is illustrated in Fig.2(b). Compared with BEGANs (Fig.2(a)), our model adds corrupted real images as extra inputs and an denoising loss to the discriminator loss function (Fig.2(b)). Therefore, the discriminator encodes all images (including real images, noisy real images, and generated images), distinguishes real images from generated images, as well as denoises the corrupted real images.



(a) BEGANs



(b) Ours

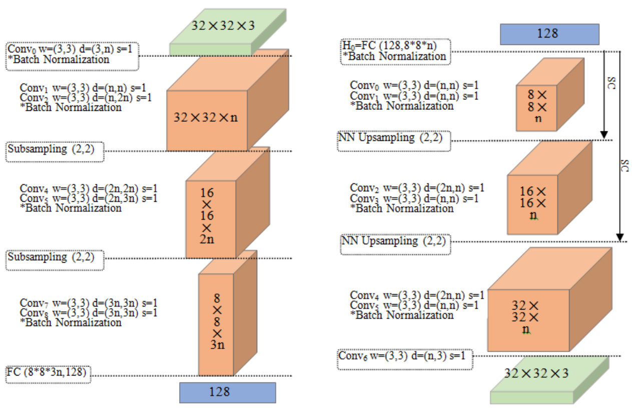
FIGURE 2. Dataflow chart of two models. Compared with BEGANs (a), ours model (b) adds corrupted real image as input and an auto-encoder denoising loss to the discriminator loss function

Adding denoising signal to generative adversarial networks was inspired by denoising feature matching (DFM) introduced by Wardefarley et al [23], which adding denoising feature loss to the generator, while the denoising feature was trained on real data by an independent fully connected

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2804278, IEEE Access

denoising auto-encoder neural networks. DFM [23] showed the effectiveness of denoising signal but it was time consuming because it needed to train three networks simultaneously and one of which was a deep fully connected network. For BEGANs architecture, adding denoising signal is easy and only need to add corrupted real data as extra inputs to the discriminator (Fig.2(b)). For the denoising reconstruction error L_{DAE} estimates the log-density gradient of real data distribution, it can improve the discriminator ability to distinguish real images and generated images and then implicitly guide the generator produce higher quality images. In addition, adding denoising loss to the discriminator can improve the stability of training and convergence.

The models' architectures are shown in Fig.3. As the principal purpose of this paper is to improve the performance of BEGANs, we follow the BEGANs architectures and only add batch normalization [22] to the second convolution layer of each block in some experiments. Both the discriminator and generator are deep convolutional neural networks, using 3×3 convolutions with exponential linear units [24] (ELUs) applied at their outputs. Down-sampling is implemented as sub-sampling with stride 2 and up-sampling is done by nearest neighbor as in [19]. We observe no improvement for replacing up-sampling with transpose convolution, which was usually used in typical GANs. All networks were trained with the Adam [25] optimizer with $\beta_1 = 0.5$ and an initial learning rate of 0.0001, decayed by a factor of 2 when the measure of convergence stalls. In all experiments, we employed isotropic Gaussian corruption noise with $\sigma = 1$. We also conducted experiments with annealing σ towards 0 (as also performed in [26]), however an annealing strategy did not perform better than fixed level noise.



(a) Encoder (b) Decoder/Generator

FIGURE 3. Network structures of the generator and discriminator. "Conv $w = (k, k), d = (m_1, m_2), s = S$" denotes a convolutional layer with $k \times k$ kernel, m_1 input filters, m_2 output filters and stride = S. "*Batch Normalization" means that the layer is followed by a batch normalization layer,"*" show the difference from BEGANs. "$FC(N_1, N_2)$" denotes a fully-connected layer with N_1 input nodes and N_2 output nodes. "SC" denotes skip connection, which means concatenating the first decoder tensor H_0 and the output of up-sampling. The activation layers are omitted.

IV. EXPERIMENTS

In this section, we demonstrated that our method can produce high quality images with stable training process and convergence on diversity datasets. We first verified the effectiveness of denoising loss on reduce noise-like regions and batch normalization on improving the diversity of generated images on CelebA dataset [27]. By adding these two methods to BEGANs, we produced high quality and diversity face images. Then by a series of ablation experiments on CIFAR-10 dataset [28], we presented the effects of several techniques trough "Inception score"[29], which was often used to measure quality and diversity of generated images. On CIFAR-10 dataset, we also demonstrated the effectiveness of denoising loss on improving the training and convergence stability.

A. CELEBA

We trained our models on the large-scale face dataset CelebA under alignment [27] to demonstrate the effectiveness of our method. We first used opencv2 to detect the face in the image and then resize to 64×64 and 128×128, so the image was more concentrated on the face and the effectiveness was more obvious. After preprocessed, there were about 150k celebrity images remained, with high degree of variability in view point, skin color, face pose, hair color, hair style, age, gender and so on.

To fairly evaluate the performance of denoising loss, we performed two group comparison testes with various γ on the same model [19](original BEGANs' model) and implementation details. The difference of each group was whether the objective function of the discriminator had denoising loss. The experimental results are shown in Fig.4. We observe that the generated images of BEGANs with denoising loss have far fewer noise-like regions than that of BEGANs, and at $\gamma = 0.3$ (Fig.4(b)) the generated images of BEGANs with denoising loss are smoothness and almost noiseless, while there are some noise-like regions on face images of BEGANs (Fig.4(d)).

Another advantage of adding denoising loss is the model can converge to a more stable and lower value. In Fig.5, we compare the convergence of the two models, which shows that adding denoising loss to the BEGANs can improve the convergence stability.

We performed experiments on BEGANs with $\gamma = 1$ (The highest diversity according to [19]), in which we did not see glasses and only saw few older people as in [19]. To improve the diversity of generated images, we applied batch normalization to the second convolution of each block of BEGANs, and the architecture is illustrated in Fig.3. The results are shown in Fig.6(b). The generated face samples have various viewpoint, outlook, expressions, genders, skin colors, hairstyle, age and glasses. On the fourth row of Fig.6(b), we highlight some samples with attributes failed or rarely generated in BEGANs, which are glasses, the older, beards and bangs from left to right. For comparison, we also displayed some results without batch normalization in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2804278, IEEE Access

(a) BEGANs with denoising loss on CelebA $\gamma = 0.5$ (64×64)

(b) BEGANs with denoising loss on CelebA $\gamma = 0.3$ (64×64)

(c) BEGANs on CelebA $\gamma = 0.5$ (64×64)

(d) BEGANs on CelebA $\gamma = 0.3$ (64×64)

FIGURE 4. Random samples of BEGANs with or without denoising loss on CelebA dataset $\gamma \in \{0.3, 0.5\}$. Note: the noise-like regions are marked with red circles for highlight in Fig(d).

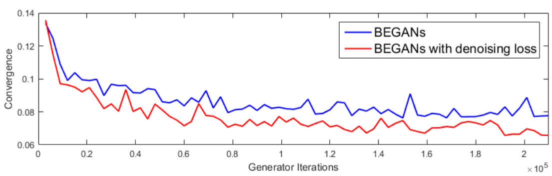
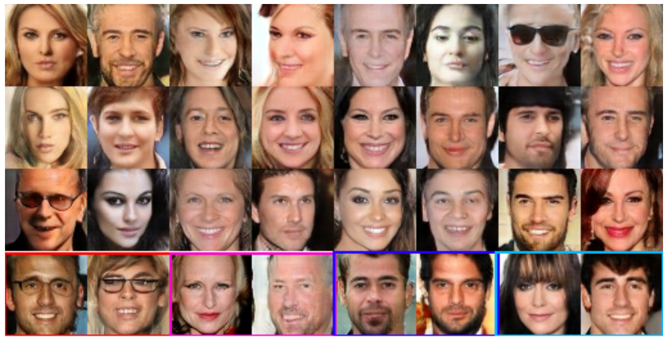


FIGURE 5. CelebA convergence of BEGANs with or without denoising loss.

Fig.6(a). Please note that in these two experiments, we also added denoising loss to the loss function of discriminator.

To further demonstrate the robustness of our method, we trained our model with $\gamma = 1$ at resolution 128×128. Some representative generated samples are presented in Fig.7(b). Higher resolution images still maintain coherency and diversity. The generated face images (Fig.7(b)) of our method are various on face shape, age, decorations (glasses), which indicates the effectiveness of our method in high-resolution image generation. We marked old face with red box, glasses with yellow box and visor-like with blue box for highlight. For comparison, we also displayed BEGANs [19] results in Fig.7(a). Note that these were trained on different datasets so direct comparison was difficult.

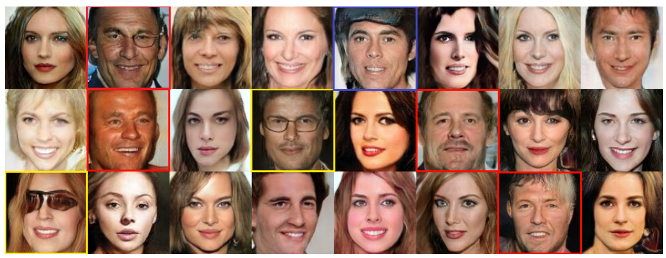
(a) Without batch normalization on CelebA $\gamma = 1$ (64×64)



(b) With batch normalization on CelebA $\gamma = 1$ (64×64)

FIGURE 6. Random 64×64 samples comparison. In the fourth row of (b), we highlight glasses, the older, beards and bangs from left to right.

(a) BEGANs (128×128)



(b) Ours $\gamma = 1$ (128×128)

FIGURE 7. Representative 128×128 samples comparison. In figure (b), we highlight old face, glasses and visor-like with red box, yellow box and blue box respectively.

B. CIFAR-10

CIFAR-10[28] is a small, well studied dataset containing 60,000 color images with resolution 32×32. We used this dataset to study the effectiveness of several techniques and to show the effectiveness of denoising loss in stable training and stable convergence, as well as to examine the visual quality of samples that can be achieved.

A series of ablation experiments were performed and all the results are presented in Table 1, which listed the "Inception score" [29] and the generated samples of different models on BEGANs [19] framework. The results in Table 1 indicate that all the techniques have effects, with the most significant effect of 'BN' ('BN' is the abbreviation of batch normalization), from 5.62 to 6.25. Adding denoising loss to discriminator also has effect on improving the "Inception score", from 6.25 to 6.53. Using all the techniques, we achieve Inception score of 7.05, a little higher than Salimans et al. [29] using unsupervised network (6.86). Training model with $\gamma = 0.7$ leads to a decline in performance, from 6.53 to 5.51. Note that in Table 1, "Baseline" used BEGANs[19] model with technique of skip connections ("SC" for abbreviation), "+BN" used our model illustrated in Fig.3 and did not add denoising loss ("DE" for abbreviation) to discriminator, "-SC+DE emloyed our model and added "DE" but removed skip connections , "Ours" used all the above mentioned techniques, "+HA" added historical averaging to "Our method", All the above models trained with $\gamma = 1$, "$\gamma = 0.7$" meant using "Our method" but trained models with $\gamma = 0.7$.

TABLE 1. Table of Inception scores for samples generated by various models.

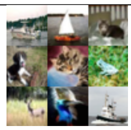
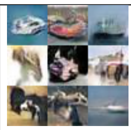
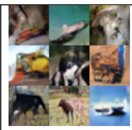
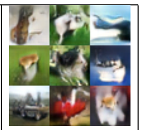
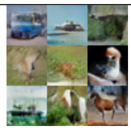
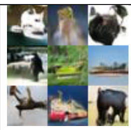
samples				
models	Real data	Baseline	+BN	-SC+DE
score	11.24	5.62	6.25	6.30

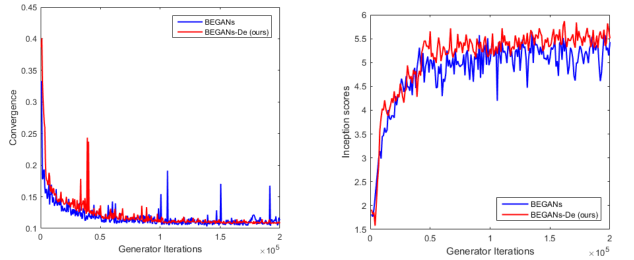
TABLE 2. Continued Table 1

samples			
models	Ours	+HA	$\gamma = 0.7$
score	6.53	7.05	5.51

Another advantage of argument denoising loss to the objective of discriminator is that it can improve training and convergence stability. To demonstrate this, we performed two expreiments on BEGANs architecture with or without denoising and plotted the measure of convergence [19] and "Inception scores" [29] over the course of training iterations (see Fig.8). As can be seen from Fig.8, adding denoising loss can make the BEGANs converge to a more stable state (Fig.8(a)) and a better final score (Fig.8(b)).

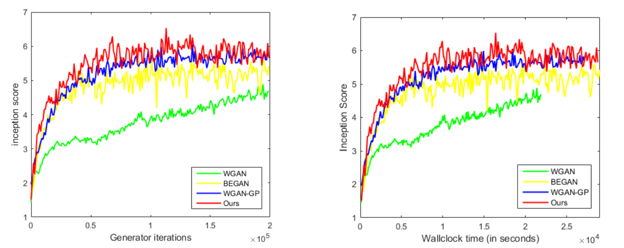
We also compared our model with the recently proposed method WGAN with weight clipping [18], WGAN with gradient penalty [11], BEGANs [19], and plotted Inception scores over the course of training (see Fig.9). Our models significantly outperformed BEGANs on stability and image quality, achieves slightly higher "Inception scores" than WGAN with gradient penalty. We also plotted the Inception scores over time (in terms of wall-clock time) and observed that our method had almost the same convergence rate as WGAN with gradient penalty. Note: the results of WGAN and WGAN-GP were performed the programs provided by Ishaan Gulrajani et al.[2], and all the programs in this subsection were performed on a single Nvidia GeForce GTX 1070 GPU.

[2]https://github.com/igul222/improved_wgan_training



(a) CIFAR-10 convergence (b) "Inception score" over generator iterations

FIGURE 8. CIFAR-10 convergence or "Inception score" over generator iterations for two models



(a) CIFAR-10 "Inception score" (b) Wall-clock time

FIGURE 9. CIFAR-10 "Inception score" over generator iterations and wall-clock time for four models: WGANs with weight clipping, WGAN with gradient penalty, BEGANs and BEGANs with denoising loss.

TABLE 3. Inception scores of different models

Method(unsupervised)	score
ALI [13] (in [23])	5.34
BEGANs[19]	5.62
Improved GANs (-L+HA)[29]	6.86
Ours	7.05
DFM[23]	7.72
WGAN-GP ResNet[11]	7.82

Table 3 shows the Inception scores of some lately similar works on models trained entirely unsupervised. Our score is higher than other GANs techniques exception of Denoising Feature Matching [23] (DFM) and WGANs with gradient penalty [11]. It is necessary to note that Denoising Feature Matching [23] used an additional network to train the denoising feature and WGANs with gradient penalty [11] used deep residual networks to improve their performance. Employing deep residual networks in our framework is a possible avenue for future work.

V. CONCLUSION

We have proposed a useful method to improve the performance of BEGANs, but a better theoretical grounding regarding the auto-encoder combined with the equilibrium concept is a necessary direction for future work, including choosing other varieties of auto-encoders such as Variational Auto-Encoders [1] (VAEs), more grounded criteria for assessing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2804278, IEEE Access

mode coverage and mass misassignment.

We introduced a simple and effective way to improve the performance of BEGANs. We have shown that adding a denoisng loss to the discriminator and applied batch normalization can significantly improve the quality and diversity of generated images. On CIFAR-10, we also compared our method with recent works and demonstrated that the stability of our method can comparative with WGANs with gradient penalty. Although we only performed our method on BEGANs framework, our method can be generalized to any GANs of employing auto-encoder as discriminator.

REFERENCES

[1] D. P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013)

[2] D. J. Rezende, S. Mohamed, D. Wierstra, Stochastic backpropagation and approximate inference in deep generative models, Eprint Arxiv 1278âĂŞ1286 (2014)

[3] Goodfellow I. NIPS 2016 Tutorial: Generative Adversarial Networks[J] (2016)

[4] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley,S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: International Conference on Neural Information Processing Systems, 2672âĂŞ2680 (2014)

[5] E. Denton, S. Chintala, A. Szlam, R. Fergus, Deep generative image models using a laplacian pyramid of adversarial networks, in: International Conference on Neural Information Processing Systems, 1486âĂŞ1494.(2015)

[6] A. Radford, L. Metz, S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Computer Science (2015).

[7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta,A. Aitken, A. Tejani, J. Totz, Z. Wang, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network arXiv:1609.04802 (2016)

[8] P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, arXiv preprint arXiv:1611.07004 (2016)

[9] T. Kim, M. Cha, H. Kim, J. K. Lee, J. Kim, Learning to discover cross-domain relations with generative adversarial networks, arXiv preprint arXiv:1703.05192.(2017)

[10] M. Mathieu, C. Couprie, Y. Lecun, Deep multiscale video prediction beyond mean square error, in:International Conference on Learning Representations (2016)

[11] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville, Improved training of wasserstein gans, arXiv preprint arXiv:1704.00028.(2017)

[12] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, H. Lee, Generative adversarial text to image synthesis, arXiv preprint arXiv:1605.05396 1060âĂŞ1069 (2016)

[13] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, A. Courville, Adversarially learned inference, arXiv preprint arXiv:1606.00704 (2016)

[14] Y. Mroueh, T. Sercu, V. Goel, Mcgan: Mean and covariance feature matching gan, arXiv preprint arXiv:1702.08398.(2017)

[15] G. J. Qi, Loss-sensitive generative adversarial networks on lipschitz densities, arXiv preprint arXiv:1701.06264(2017)

[16] J. Zhao, M. Mathieu, Y. Lecun, Energy-based generative adversarial network, arXiv preprint arXiv:1609.03126(2016).

[17] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, S. P. Smolley, Least squares generative adversarial networks, arXiv preprint ArXiv:1611.04076 (2016)

[18] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan, arXiv preprint arXiv:1701.07875 (2017)

[19] D. Berthelot, T. Schumm, L. Metz, Began: Boundary equilibrium generative adversarial networks, arXiv preprint arXiv:arXiv:1703.10717 (2017)

[20] P. Vincent, H. Larochelle, Y. Bengio, P. A. Manzagol, Extracting and composing robust features with denoising autoencoders, in: International Conference on Machine Learning, 1096âĂŞ1103 (2008)

[21] G. Alain, Y. Bengio, What regularized auto-encoders learn from the data generating distribution, Journal of Machine Learning Research 15 (1) 3563âĂŞ3593.(2014)

[22] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, 448âĂŞ456 (2015)

[23] D. Wardefarley, Y. Bengio, Improving generative adversarial networks with denoising feature matching, in: International Conference on Learning Representations (2017)

[24] D. A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), Computer Science (2015)

[25] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, Computer Science (2014)

[26] C. K. Snderby, J. Caballero, L. Theis, W. Shi, F. Huszr, Amortised map inference for image super-resolution (2016)

[27] Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, in: Proceedings of the IEEE International Conference on Computer Vision, 3730âĂŞ3738 (2015)

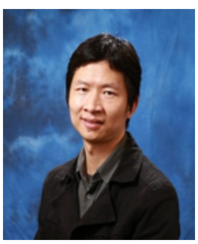
[28] A. Krizhevsky, G. Hinton, Learning multiple layers of features from tiny images (2009)

[29] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, in: Advances in Neural Information Processing Systems, 2234âĂŞ2242 (2016)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2804278, IEEE Access

YANCHUN LI Yanchun Li obtained the B.E.E., M.S.E.E. degrees in computer science all from the College of Information Engineering of Xiangtan University, Hunan, China. She is now a Ph.D. student in the School of Computer Science and Engineering, South China University of Technology. Her research interests include computer vision and image processing.

NANFENG XIAO obtained the B.E.E. degree from the Department of automatic control and computer of Huazhong University of Science and Technology, received the M.S.E.E. and Ph.D. degrees in engineering from Northeastern University, China and Yokohama National University, Japan respectively. He is now a professor and Ph.D. supervisor of South China University of Technology. His research interests include deep learning and artificial intelligence.



WANLI OUYANG obtained Ph.D. degrees from the Department of Electronic Engineering, the Chinese University of Hong Kong. He is now a senior lecturer in the School of Electrical and Information Engineering, University of Sydney. His research interests include deep learning and its application to computer vision and pattern recognition, image and video processing.