Semi-Supervised Learning with Conditional GANs for Blind Generated Image Quality Assessment

Xuewen Zhang[1], Yunye Zhang[2], Wenxin Yu[1]✉, Liang Nie[1], Zhiqiang Zhang[3], Shiyu Chen[1], and Jun Gong[3]

[1] Southwest University of Science and Technology, Sichuan Province, China
[2] University of Electronic Science and Technology of China, Sichuan Province, China
[3] Hosei University, Japan
[4] Beijing Institute of Technology, China
yuwenxin@swust.edu.com

Abstract. Evaluating the quality of images generated by generative adversarial networks(GANs) is still an open problem. Metrics such as Inception Score(IS) and Fréchet Inception Distance(FID) are limited in evaluating a single image, making trouble for researchers' results presentation and practical application. In this context, an end-to-end image quality assessment(IQA) neural network shows excellent promise for a single generated image quality evaluation. However, generated image datasets with quality labels are too rare to train an efficient model. To handle this problem, this paper proposes a semi-supervised learning strategy to evaluate the quality of a single generated image. Firstly, a conditional GAN(CGAN) is employed to produce large numbers of generated-image samples, while the input conditions are regarded as the quality label. Secondly, these samples are fed into an image quality regression neural network to train a raw quality assessment model. Finally, a small number of labeled samples are used to fine-tune the model. In the experiments, this paper utilizes FID to prove our method's efficiency indirectly. The value of FID decreased by 3.32 on average after we removed 40% of low-quality images. It shows that our method can not only reasonably evaluate the result of the overall generated image but also accurately evaluate the single generated image.

Keywords: Generative adversarial networks · Image quality assessment · Generated image

1 Introduction

Generative Adversarial Networks(GANs)[3] have made a dramatic leap in synthesizing images. However, how to evaluate the single generated image or how to provide a quality score for it is still an open problem. In this case, the absence of single-image evaluation methods brings the following challenges. Firstly, a large number of low-quality images in the results are hard to filter out, which hinders the practical application of image synthesis. Secondly, it makes trouble for

researchers' results presentation. Since the inherent instability of GANs brings too much uncontrollable content to generated images, it is challenging to conduct supervised learning to capture the distortion. If someone wants to create a specific dataset for supervised learning, two questions are inevitable. On the one hand, the samples need to include enough generated images with various distortion types and various categories(Just like ImageNet[2]). It is difficult because so many GAN models need to be considered, which burdens the image collection progress. On the other hand, labeling these enormous numbers of images with a precise score is time-consuming. A bad generated image may due to the lack of authenticity instead of traditional degradation such as blur, low-resolution, or white noise. In contrast, authenticity distortion is challenging to define.

Recently, generated image quality assessment frequently focuses on the distribution of the features. Metrics such as Inception Score(IS)[15] and Fréchet Inception Distance(FID)[6] are employed to evaluate images base on the feature distribution. They may properly evaluate the overall quality, but neither can evaluate a single generated image. Besides, some full-reference metrics such as Peak Signal-to-Noise Ratio(PSNR) and Structural Similarity(SSIM) are limited because the reference images are frequently unobtainable in many image synthesis studies. For example, text-to-image synthesis[13] and image-to-image style transfer[24]. Although reference images exist in some special generation tasks, such as inpainting[12], one of our goals is to have a general quality assessment approach. Intuitively, using a deep neural network to learn the mapping from image features to the quality score is a simple but useful solution. However, the generated image dataset with the quality label is challenging to obtain, as mentioned above. Therefore, finding an effective training strategy to solve the problem caused by the scarcity of datasets is the key to using DNN to achieve GIQA.

In this context, we discover that conditional GANs[10] can constrain the generator's output with additional inputs. In the image-to-image transformation task[7], we observed that the output image is directly affected by the input image. Motivated by it, this paper hypothesizes that the more ground truth information the input image contains, the better the output image's quality, holding other network parameters constant. Based on this hypothesis, this paper proposes a semi-supervised method to train an image quality assessment(IQA) model for generated images. Our core idea is to train an image-to-image conditional GAN(CGAN) to produce images that include a quality label automatically. These samples can be used as the training data of the image quality prediction model. In detail, quality of the generated images by G can be controlled by the conditional input image. Therefore, we can obtain enough training samples with controllable quality while preserving the characteristics of the generated image (unstable and random), as shown in Fig.1. Furthermore, this condition is utilized as the quality label of these generated images. It makes the training of the IQA model for a single generated image at a low cost. For this paper, the main contributions are as follows.

1.A semi-supervised learning method for a single generated image quality assess-

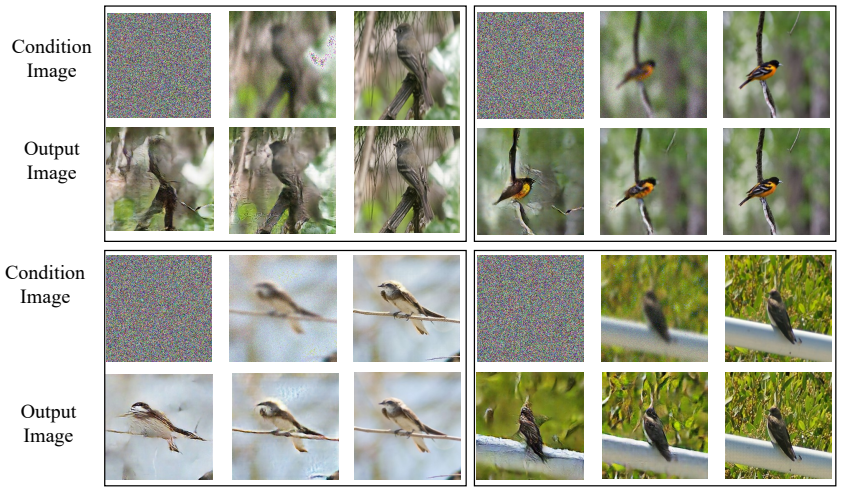

Fig. 1. Conditional GAN is utilized to generate images samples whose qualities are controllable. If other parameters are constant, the quality of the output image is directly influenced by the condition.

ment is proposed combining with the conditional GAN. We call it painter GAN which can produce a large number of quality-controllable generated images.
2.We propose a image filtering technique to remove low-quality generated images by evaluating the quality of a single image.
3.Extensive experiments are conducted to prove that our method can solve the evaluation problem of single-generated images to some extent.

2 Related Work

The evaluation of generated images is achieved by calculating the feature distribution and comparing it with a real image set. Inception Score(IS)[15] measures the image results from two aspects: the statistics level's recognizability and diversity. However, IS inherently has controversy[1], which leads to a lack of confidence in theory and practical applications. Fréchet Inception Distance(FID)[6] calculates the similarity of the feature distribution between the generated results and ground truth. These two methods measure the overall quality instead of the image itself. Therefore, methods that focus on a single image(such as PSNR, SSIM, etc.) are utilized to complete the evaluation system. These solutions directly compare the pixel-level difference with reference images. These kinds of methods are not flexible because reference images frequently unavailable in many studies. Besides, they difficultly match the human subjective perception in some cases. The existing no-reference image quality assessment methods(IQA) are mainly based on hand-craft features or learning-based features. Hand-crafted feature approaches utilize Natural Scene Statistic (NSS) models to capture distortion,

which be used as the feature of the regressing model to achieve quality score[11, 14]. Learning-based methods usually adopt well-designed network structures to map image features to scores[8, 9, 23, 16]. [8] uses GANs to generate a illusory reference and then uses a phantom reference image to evaluate the target image. [9] utilize a Siamese network to learn the degradation distortion from rank order. In [23], they firstly concatenate the meta-learning method to blind image quality assessment to achieve small-sample learning. [16] employed a hyper-network to learn a specific quality representation for each input image and achieve a surprising result.

There are few works for generated images quality assessment. Gu et al.[4] first proposed GIQA to predict the scores of generated images based on CNNs. They saved the intermediate generated images before the GAN converged and used the number of iterations as the label of the quality to obtain a large number of training samples. However, due to the instability in the training process of GANs, the number of iterations is difficult to indicate the quality of the generated image accurately. In the experimental part, we will further compare our method with their work. Our previous work[22, 21] evaluated the generated images using NSS and DNN-based methods combined with specific datasets, respectively. However, these methods rely heavily on the dataset itself. In this paper, a semi-supervised learning method is proposed to solve this problem by inputting various real images without labels and learning various quality representations using conditional GANs and CNNs.

3 Our Proposal

Annotating quality scores for large numbers of generated images is a strenuous task. To train a quality evaluation model without a large-scale labeled dataset, this paper proposed a semi-supervised strategy. As it's shown in Fig.2, a conditional GAN(we metaphor it as a painter) is employed to generate images with different but quality-controllable image samples with unlabeled images. These images are fed into a convolution neural network(CNN) to know what a good or bad image is. Finally, a small number of images with quality labels are used to fine-tune the CNN.

3.1 The Painter GAN

As mentioned above, our purpose is to address the problem that there are not enough generated images with quality labels to train an image quality assessment(IQA) model. The solution is that we use a conditional GAN to get a large number of generated images as the training samples, and the input condition can play a role of score label. In order to better explain our proposal, we metaphor a generator network as a painter, and his ability is to paint images of what he sees. If we assume that the painter's painting ability is constant, we can control the quality of his paintings by controlling what he can see. Therefore, we can get many quality-controllable paintings to teach a kid what painting is good or

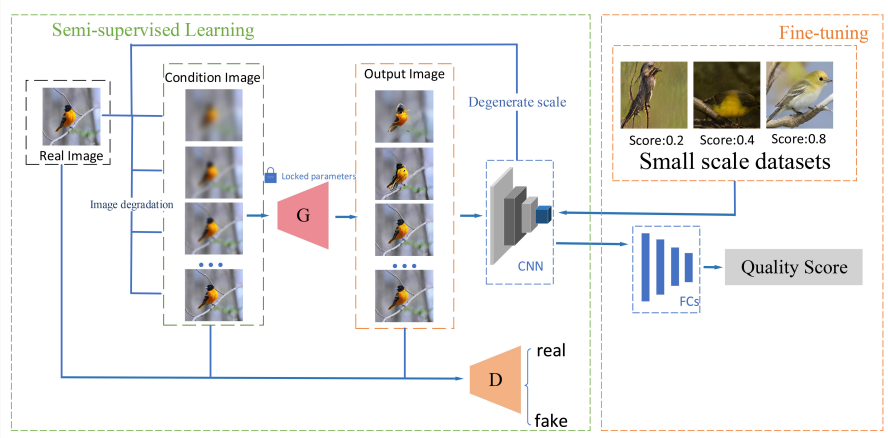

Fig. 2. The architecture of our method. The input of the generator is degraded images, while it outputs samples with stepped quality-level. As the condition image's degradation intensity increases, G will gradually generate low-quality generated images because the parameters of G are locked. These images with corresponding degradation intensity are sufficient training samples for CNN.

bad. Seriously, the painter in our paper is a conditional generator network for image-to-image tasks, and his paintings are the generated images. The kid is the generated IQA model we want to train.

Condition GANs[10] input conditions to both generator and discriminator to control the generated images. Its optimization goal display as follows.

$$\begin{aligned} \min_G \max_D V(G,D) = \mathbb{E}_{y\sim p_{data}(y)}(logD(y|c))+ \\ \mathbb{E}_{x\sim p_x(x)}(log(1-D(G(x|c)))) \end{aligned} \tag{1}$$

where x, y denote input and ground truth, respectively. The conditions c is seen by both generator and discriminator, which control results in customers' aspects.

In this paper, we use the down-sampled images as the conditions, and its original image is the ground truth. For each real image I_{real}, using average-pooling operations to reduce the information.

$$I_c = average\_pooling(I_{real}) \tag{2}$$

Subsequently, the down-sampled image and its original image are paired, and G aims to convert the down-sampled image into the original image. Following pixel-to-pixel [7], the objective includes adversarial loss and L1 loss.

$$L_{GAN}(G,D) = \mathbb{E}_{I_c,I_{real}}[logD(I_c,I_{real})] + \mathbb{E}_{I_c,z}[log(1-D(I_c,G(I_c,z)))] \tag{3}$$

$$L_1(G) = \mathbb{E}_{I_c,I_{real},z}(||I_{real} - G(I_c,z)||_1) \tag{4}$$

$$Obj = \arg\min_G \max_D L_{GAN}(G,D) + \lambda L_1(G) \tag{5}$$

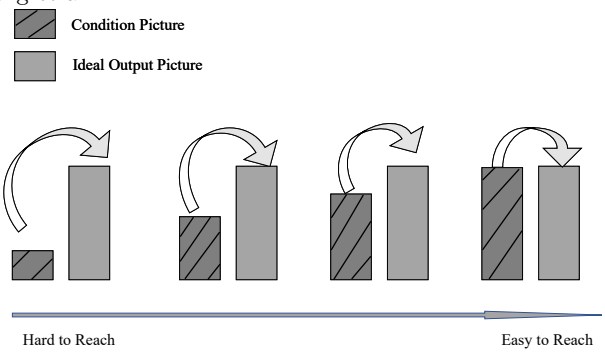

Fig. 3. An illustration of the condition image to the ideal output image in the image-to-image transformation task, the less information the input condition contains about the real image (such as contour or segmentation image), the more difficult it is for the generator to produce the ideal image. In the contrast, if the input condition image is the ideal output image itself, everything becomes easy. For the generator, the cases on the right produce better output than the left one when keeping other parameters unchanged (See Fig.1). Thus, the condition image can be regarded as the quality label of input images.

where G is generator, D is discriminator, I_c is the down-sampled image, I_{real} is the real image, and z denotes random noises. Keeping the network training hyperparameters unchanged, we can determine the quality scores s of the generated image by the degree of image downsampling.

$$s = \frac{W_{I_c} \times H_{I_c}}{W_{I_{real}} \times H_{I_{real}}} \tag{6}$$

where W_{I_c} and H_{I_c} are the width and height of the down-sampled image. Similarly, $W_{I_{real}}$ and $H_{I_{real}}$ are the height and width of the real image.

In the design of this paper, the goal of conditional GAN is to generate the original image based on the down-sampled image(similar to super-resolution). If the condition contains less information, it will be more difficult to restore(See Fig.3), and results tend to have more image artifacts. Therefore, the size of the down-sampled condition image implies the amount of information, which can influence the quality of the generated image.

3.2 Quality Evaluation Model

According to the strategy in Sec.3.1, a large number of samples with different qualities can be produced by the generator while adjusting the parameter of the down-sample operation. Therefore, an semi-supervised generated image quality assessment model(semiGIQA) is achieved through CNNs and FCs. CNNs aim to extract the features of images, while FCs map these features to the quality latent space.

$$s^* = f(I_g; \theta) \tag{7}$$

$$\mathbb{L} = \mathbb{E}(||s^* - s||_1) \tag{8}$$

where I_g denotes input image, f is the quality prediction model that includes the convolution layers and fully connected layers, θ is the model's parameters, s is defined in Equation(6). According to the semi-supervised learning strategy, s also represent the quality label of sample in the small generated image quality assessment dataset. And L_1 loss is utilized to fine-tune the quality score regression model.

3.3 Evaluation and Optimization with Score

Our semiGIQA model can evaluate the quality of a single generated image without reference images, which is more flexible than methods such as IS, FID, or PSNR. The mean score of n generated images is defined as $mean = \frac{1}{n}\sum f(x_i;\theta)$.

Because every image's subjective score is obtained, it's possible to optimize the results by filtering out low-quality images. To not destroy the diversity, we filter out images with low scores in each category. Experiments show that FID decreased, and the image's subjective quality is effectively improved after the screening operation.

4 Experimental Results

To prove the method's effectiveness, we select a series of GANs-based generated images, used our method to predict quality scores, and display its superiority compare to FID, IS, or PSNR. Similar to our method, GIQA[5] is also a method for evaluating a single generated image. They saved the intermediate images before the GAN converged, and used the number of iterations as the label of the quality to obtain a large number of training samples. In the experiment, we will compare our method with the GIQA baseline.

4.1 Datasets

To simplify the process, this paper chooses GANs for the text-to-image synthesis tasks[19, 13, 18, 25] and Caltech-UCSD Birds-200-2011(CUB) dataset[17] to verify the effectiveness of the proposed method. The reasons are as follows. Firstly, compared to the image-to-image conversion task, the text-to-image synthesis task is more complex, so the image quality in the results varies greatly, which is more conducive to verify our method. Secondly, our semi-supervised process requires a small-scale generated images dataset with quality labels. To the best of our knowledge, the only public generated IQA dataset MMQA[20] is based on the CUB dataset. CUB dataset contains 11788 images of birds in 100 different categories. All the images of this dataset are used for unsupervised learning without image quality labels. MMQA contains 5000 generated images by the GANs of text-to-image synthesis. 12 observers give each image quality scorers on 7 different aspects. This dataset is used to fine-tune our network, of which 80% is used for training and 20% is used for testing.

4.2 Implementation Details

All our experiments are implemented on PyTorch with two NVIDIA 1080Ti GPUs. First, each image in the CUB dataset is down-sampled 6 times with different sizes and then pairs them with the original image and marks according to formula (6). Therefore, there are 70,728 pairs used to train image-to-image GANs. Eventually, 70,728 generated images can produce with down-sampled images and automatically obtain their quality labels according to formula (6), and the structure and parameters of the GAN follow[7]. These samples are then used to train the image quality prediction network, of which 80% of images are used for training, and 20% of images are used for testing to prevent overfitting. Finally, the 5,000 images with artificial quality labels in MMQA were used to fine-tune the network. We randomly crop the images to 224×224, perform random flips, set the batch size to 64, and use the Adam optimizer with a 0.9 of momentum and a 0.0005 weight decay to train 10 epochs. Other parameters of IQA model follow [16].

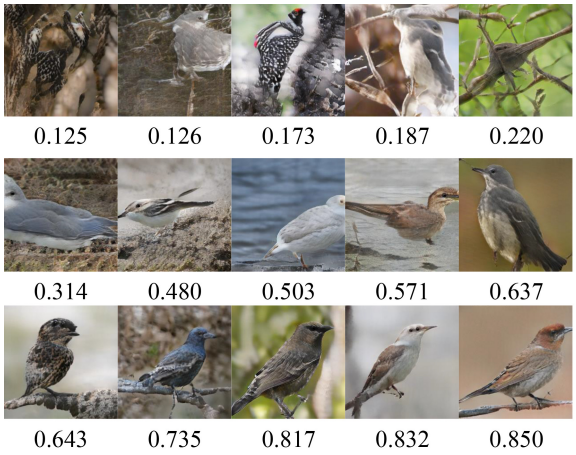

Fig. 4. Example of quality score prediction for single generated images.

Table 1. Our method is used as a supplementary evaluation standard for IS and FID to make a comprehensive quantitative evaluation.

	GAN-CLS	StackGAN++	AttnGAN	DM-GAN
IS↑	2.93	4.08	4.32	4.70
FID↓	174.73	26.85	23.16	15.31
PSNR	-	-	-	-
ours↑	0.27	0.47	0.50	0.51

4.3 Supplementary of Evaluation System

We choose four popular text-to-image synthesis methods, GAN-CLS[13], StackGAN++[19], AttnGAN[18], DM-GAN[25]. As shown in Table.1, our method

evaluates 4 methods (each method produces 30,000 generated images with the same condition). The evaluation trends of ours, GIQA[5], IS, and FID is identical, which matches the actual performance of these models. However, FID in low-quality images is a little biased so that it's not fair for low quality results(174.73 in GAN-CLS) and IS has some controversies. In contrast, our method is smoother, while evaluating images from human subjective perception is more intuitive. PSNR hardly works because it's difficult to find reference images in text-to-image synthesis. Besides, as shown in Fig.4, our method is efficient in predicting scores for a single image, flexible and applicable. In contrast, IS and FID fail because they require a sufficient number of images to extract feature distributions. Therefore, our method is reasonably utilized as a supplement to the evaluation system,which shows that our method can not only reasonably evaluate the result of the overall generated image, but also accurately evaluate the single generated image(See Fig.4).

Table 2. FID score of generated images after filter out low score images. The high-quality images from the top 100% to the top 60% are preserved.

GAN models	Top Percentage	1.0(vanilla)	0.9	0.8	0.7	0.6
GAN-CLS[13]	Random	174.73	174.61	174.30	173.70	173.05
	GIQA[5]	174.73	174.44	173.74	172.88	172.82
	ours	174.73	**172.25**	**170.27**	**168.83**	**167.92**
StackGAN++[19]	Random	26.85	26.51	26.68	26.90	27.24
	GIQA[5]	26.85	25.65	25.11	24.89	24.84
	ours	26.85	**23.96**	**22.28**	**21.24**	**20.95**
AttnGAN[18]	Random	23.16	23.18	23.38	23.7	23.91
	GIQA[5]	23.16	22.86	22.57	22.3	22.13
	ours	23.16	**21.02**	**19.49**	**18.63**	**18.38**
DM-GAN[25]	Random	15.34	15.48	15.61	15.86	16.22
	GIQA[5]	15.34	14.95	14.89	14.87	**15.06**
	ours	15.34	**14.57**	**14.46**	**14.80**	15.46

4.4 Optimization of Generated Results

Compared with FID, our method is available for a single image, which makes results optimization possible. To verify the optimization strategy's rationality, we employ four popular text-to-image synthesis models to generate 30,000 images and filter out low-scoring images. A reasonable assumption is that after removing low-quality images, the overall quality of the rest images is better than before. Therefore, with this assumption, we can get images of high quality. As shown in Table.2, we gradually remove 10% of low-quality images and calculate the FID score of the remaining images. The decrease in FID means an improvement in the overall quality. As a comparison, the same number of images are randomly removed at each step to get the FID score of the remaining images. It also shows that filtering images randomly rarely reduces FID stably and may even increase it. Besides, we use the same strategy in the results of GIQA[5], but its effect is not as evident as our method.

Our method has achieved promising results except for DM-GAN, we analyze the reasons as follows. We adopted a strategy of screening low-quality images proportionally. As shown in Tab.1, the results in DM-GAN have the best quality, the high-quality images stay a large proportion of the results (maybe greater than 70%). Therefore, a lot of high-quality images will be removed, which will lead to an increase in FID. In contrast, the overall result becomes better in other GAN-based results because more low-quality images are removed.

4.5 Discussion

In this section, we will further discuss the motivation and justification of the proposed approach. GANs generally have two well-known problems; Instability and model collapse. The key of our method is to utilized CGANs to produce quality-controllable image samples. On the one hand, GAN's instability is reflected in the results and expressed in the training process. Therefore, it is not reasonable to directly use the number of training iterations as the standard for evaluating the generated images(GIQA adopts it). The experimental results also prove that, as shown in Table.2, GIQA cannot well screen out low-quality results. On the contrary, by observing the output of condition GAN(See Fig.1), we find that sufficient conditions lead to stable results. In our task, GAN only needs to fill the gap between the given conditions and the ground truth, which allows us to control the generation of GANs with enough conditions. It is not a cheating method because our task is image quality evaluation rather than image generation. On the other hand, GAN often faces model collapse, which destroys the diversity of generated results. Fortunately, our approach avoids this problem; As mentioned earlier, we input sufficient conditions into GAN. Since the variety of the output is actually dependent on the diversity of the input conditions, the variety of generated images in our model can be guaranteed. Our semi-supervised learning strategy can learn quality representations from a high-quality unlabeled image, so the evaluation model's generalization can be improved by introducing various types of images.

5 Conclusion

The quality assessment algorithms for a single image usually rely heavily on the image dataset with quality labels. For generated images, it is tough to collect various generated images with quality labels. Therefore, this paper proposes a semi-supervised learning approach that allows models to learn quality representations from single unlabeled images. Firstly, we utilized conditional GAN to produce quality-controllable samples without any labels automatically. After these samples are fed into an IQA model, a small number of labeled images are used to fine-tune the model. The comprehensive experiment proves the effectiveness of the method. Subjectively, the prediction quality scores by our method are consistent with human perception. Objectively, our approach can improve the quality of overall results (according to FID) by accurately filtering low-quality images.

6 Acknowledgements

This research is supported by Sichuan Science and Technology Program (No.2020YFS0307, No.2020YFG0430, No.2019YFS0146), Mianyang Science and Technology Program(2020YFZJ016).

References

1. Barratt, S.T., Sharma, R.: A note on the inception score. CoRR **abs/1801.01973** (2018), http://arxiv.org/abs/1801.01973
2. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
3. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. arXiv preprint arXiv:1406.2661 (2014)
4. Gu, S., Bao, J., Chen, D., Wen, F.: Giqa: Generated image quality assessment. In: European Conference on Computer Vision. pp. 369–385. Springer (2020)
5. Gu, S., Bao, J., Chen, D., Wen, F.: GIQA: generated image quality assessment. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J. (eds.) Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XI. Lecture Notes in Computer Science, vol. 12356, pp. 369–385. Springer (2020)
6. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA. pp. 6626–6637 (2017)
7. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
8. Lin, K., Wang, G.: Hallucinated-iqa: No-reference image quality assessment via adversarial learning. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. pp. 732–741. IEEE Computer Society (2018)
9. Liu, X., van de Weijer, J., Bagdanov, A.D.: Rankiqa: Learning from rankings for no-reference image quality assessment. In: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017. pp. 1040–1049. IEEE Computer Society (2017)
10. Mirza, M., Osindero, S.: Conditional generative adversarial nets. CoRR **1411.1784** (2014), http://arxiv.org/abs/1411.1784
11. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
12. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2536–2544 (2016)
13. Reed, S.E., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative adversarial text to image synthesis. In: Balcan, M., Weinberger, K.Q. (eds.) Proceedings of the 33nd International Conference on Machine Learning, ICML 2016. JMLR Workshop and Conference Proceedings, vol. 48, pp. 1060–1069. JMLR.org (2016)

14. Saad, M.A., Bovik, A.C., Charrier, C.: Blind image quality assessment: A natural scene statistics approach in the DCT domain. IEEE Trans. Image Process. **21**(8), 3339–3352 (2012)
15. Salimans, T., Goodfellow, I.J., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain. pp. 2226–2234 (2016)
16. Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., Zhang, Y.: Blindly assess image quality in the wild guided by a self-adaptive hyper network. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020. pp. 3664–3673. IEEE (2020)
17. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset (2011)
18. Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., He, X.: Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. pp. 1316–1324. IEEE Computer Society (2018)
19. Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., Metaxas, D.N.: Stackgan++: Realistic image synthesis with stacked generative adversarial networks. IEEE Trans. Pattern Anal. Mach. Intell. **41**(8), 1947–1962 (2019)
20. Zhang, X., Yu, W., Jiang, N., Zhang, Y., Zhang, Z.: Sps: A subjective perception score for text-to-image synthesis. In: 2021 IEEE International Symposium on Circuits and Systems (ISCAS). pp. 1–5. IEEE (2021)
21. Zhang, X., Zhang, Y., Zhang, Z., Yu, W., Jiang, N., He, G.: Deep feature compatibility for generated images quality assessment. In: International Conference on Neural Information Processing. pp. 353–360. Springer (2020)
22. Zhang, Y., Zhang, X., Zhang, Z., Yu, W., Jiang, N., He, G.: No-reference quality assessment based on spatial statistic for generated images. In: International Conference on Neural Information Processing. pp. 497–506. Springer (2020)
23. Zhu, H., Li, L., Wu, J., Dong, W., Shi, G.: Metaiqa: Deep meta-learning for no-reference image quality assessment. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020. pp. 14131–14140. IEEE (2020)
24. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
25. Zhu, M., Pan, P., Chen, W., Yang, Y.: DM-GAN: dynamic memory generative adversarial networks for text-to-image synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019. pp. 5802–5810. Computer Vision Foundation / IEEE (2019)