Critical Analysis: Generative Adversarial Nets

161189284

Guanzhen Wu

The theory of Generative Adversarial Nets (GAN), which is widely used in various fields, has received extensive attention since the publication of the discussed article. Application areas include image conversion [1], texture learning, and text-to-image conversion [2], Super-Resolution [3], image inpainting [4], image mixing [5], and even data generation in deep learning. As we know, deep learning requires a lot of data as training material, the generative model of GAN can just achieve this function.

The above applications are all based on the model principle of GAN, the confrontation between the generative model *G* and the discriminative model *D*. The generative model *G* needs to generate pictures like the real scene as much as possible, while the discriminant model *D* needs to distinguish fake pictures. This is similar to a max-min game. In the process of GAN application, a large number of ideas in the field of pictures and natural language processing are involved, such as text-to-image conversion [2], semantic segmentation is used to help the model identify and remember specific features and associate them with corresponding labels. Which means you can enter any text to get the corresponding picture with textual meaning. In the application of SR [3], GAN replaces the traditional interpolation method and uses the generated pixels to fill in the blurred parts to achieve super-resolution image generation.

However, although GAN has a wide range of applications in various fields, its model itself still has some problems. Like other deep neural network models, the GAN model is a black box. How to design and train update models to develop high-quality generators is very difficult. In response to the training process and the problem of training collapse, Arjovsky et al. proposed the Wasserstein Generative Adversarial Networks (WGAN) model [6] which improves the stability of learning, get rid of problems like mode collapse, and provide meaningful learning curves useful for debugging and hyperparameter searches. They also explored the impact of weight pruning, regularization and other strategies on the model, which made an improvement compared to the original model version. Other articles also design new loss functions to improve the GAN model efficiency. Overall, the modification of the training strategy as well as the experiments helped the GAN to form a more efficient and accurate generative discriminator. In addition to the problems of algorithms and models, the development of GAN has also brought some scientific and technological ethical issues [7], such as the application of fake face, the generated fake pictures may be used for fraud or slander, how to distinguish and suppress also worth thinking about.

For the remaining problem, the model evaluation of GAN is a major remaining problem [8]. Various evaluation methods are classified into two categories, qualitative and quantitative. Qualitative models can be evaluated by users as a final test (similar to the Turing test) as to whether a person can tell that an image was generated by a machine. This type of evaluation method will make the model too dependent on the existing data, that is, the model is easy to be overfitting. Models with high pass rates tested using this method typically exhibit low diversity in the resulting material. Quantitative evaluation is to judge the internal parameters of the model, such as judging whether the loss function converges, whether the configuration of each hyperparameter is optimal, and so on. But this evaluation method is too abstract, and people cannot directly perceive the generated material and reflect it on the generative model.

I am very interested in the evaluation of GAN models, as in the article discussed, the author Goodfellow uses Average log-likelihood as a quantitative evaluation criterion for generative model G, and Rapid Scene Categorization as a qualitative evaluation criterion for discriminative model D. Since the input of the final model is a noise represented by a matrix, the objective to be optimized of generator should be the loss between generated image and ground truth. A large number of loss functions can be applied to this module. However, how to connect the discriminative method of model D to model G to achieve the best model generation effect, is still left to the researchers to discuss. Besides, the optimal distribution of probabilities of generated image does not necessarily make the picture look more natural and realistic, so how to solve this problem is interesting.

Reference:

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

[2] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X. and Metaxas, D.N., 2017. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 5907-5915).

[3] Zhang, H., Sindagi, V. and Patel, V.M., 2019. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, *30*(11), pp.3943-3956.

[4] Yeh, R.A., Chen, C., Yian Lim, T., Schwing, A.G., Hasegawa-Johnson, M. and Do, M.N., 2017. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5485-5493).

[5] Wu, H., Zheng, S., Zhang, J. and Huang, K., 2019, October. Gp-gan: Towards realistic high-resolution image blending. In *Proceedings of the 27th ACM international conference on multimedia* (pp. 2487-2495).

[6] Arjovsky, M., Chintala, S. and Bottou, L., 2017, July. Wasserstein generative adversarial networks. In *International conference on machine learning* (pp. 214-223). PMLR.

[7] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B. and Anderson, H., 2018. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*.

[8] Ali Borji, Pros and cons of GAN evaluation measures, Computer Vision and Image Understanding, Volume 179,2019,Pages 41-65,ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2018.10.009.