review: GAN's application on computer vision-2

Huang Daoji

June 14, 2020

1 GAN for high resolution generation

While generative adversarial networks(GAN)[1] can generate realistic images, they fail to do so at a larger scale. Some of the reasons they fail are

- a) the gradient direction becomes more random in high dimension space due to less overlap of the training and generated distributions,
- b) high dimensionality also make the discriminator more easily to differentiate the training and generated images, and
- c) due to memory constraints, large resolution leads to smaller mini-batches, which further compromises training stability.

One way to tackle this issue is to focus on gradient directions. Other metrics have been proposed, including least squares[2] and Wasserstein distance[3], in addition to the original Jensen-Shannon divergence. These methods are useful but not enough to alleviate GAN's instability at large scale.

An orthogonal approach aims to change the way GAN is being trained. The key observation is that instead of training all layers from scratch, GAN can be trained step by step, whether it be training a hierarchy of GANs, or training one layer at once. Denton et al.[4] inserted a hierarchy of GANs into the Laplacian pyramid, while Huang et al.[5] further introduced a pretrained encoder to match intermediate representations. StackGAN[6] and Wang et al. 2017[7] both used a two-stage generation scheme to enable high-resolution synthesis. ProGAN[8] further trains GAN in smaller steps: layer by layer, without pre-configured layers. Several tricks are applied to stabilize training: smooth fade-in of new layers, explicitly concatenating standard deviation, and more normalizations, including equalized learning rate, pixelwise normalization. ProGAN achieved more realistic results even using smaller mini-batches and trained faster compared to ones trained from scratch.

A seemingly more brute-forcing way to train GAN for high-resolution synthesis is to train them at large scale. In BigGAN[9], it has been shown that GAN's performance simply increases as channel numbers and batch sizes grow larger. GAN behaviors at large scale are also studied, including failure when truncation trick applied, mode collapse which only studied at a smaller scale. Several tradeoffs have been found when tackling these issues: the tradeoff between fidelity and variety, between image quality and training stability.

2 style transfer

Given a content and a style image, style transfer aims to synthesize an output image which combines the content and style from reference images. A general approach[10, 11, 12, 13] is to encode input images to some latent space and synthesize output from these latent vectors. Inspired by AdaIn[14], various methods[15, 16] combine these vectors via aligning the distribution of pixels by adjusting the scale and variance of activation in a neural network. Lu et al.[17] proposed a closed-form solution using technics from optimal transport. Also, due to lack of paired training data, style transfer methods often adapt their own approach to enforcing the separation of content and style, like VGG loss[18], minimizing cycle loss[19], or GAN loss[20, 21].

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 2672–2680, Curran Associates, Inc., 2014.

[2] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2813–2821, 2017.

[3] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning* (D. Precup and Y. W. Teh, eds.), vol. 70 of *Proceedings of Machine Learning Research*, (International Convention Centre, Sydney, Australia), pp. 214–223, PMLR, 06–11 Aug 2017.

[4] E. L. Denton, S. Chintala, a. szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," in *Advances in Neural Information Processing Systems 28* (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds.), pp. 1486–1494, Curran Associates, Inc., 2015.

[5] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie, "Stacked generative adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1866–1875, 2017.

[6] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *ICCV*, 2017.

[7] T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8798–8807, 2018.

[8] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *ArXiv*, vol. abs/1710.10196, 2018.

[9] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *International Conference on Learning Representations*, 2019.

[10] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, "Joint discriminative and generative learning for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[11] H.-J. Chen, K.-M. Hui, S.-Y. Wang, L.-W. Tsao, H.-H. Shuai, and W.-H. Cheng, "Beautyglow: On-demand makeup transfer framework with reversible generative network," pp. 10034–10042, 06 2019.

[12] T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu, and L. Lin, "Beautygan: Instance-level facial makeup transfer with deep generative adversarial network," in *MM '18*, 2018.

[13] W. Wu, K. Cao, C. Li, C. Qian, and C. C. Loy, "Disentangling content and style via unsupervised geometry distillation," *ArXiv*, vol. abs/1905.04538, 2019.

[14] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *ICCV*, 2017.

[15] P. Zhu, R. Abdal, Y. Qin, and P. Wonka, "Sean: Image synthesis with semantic region-adaptive normalization," 2019.

[16] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[17] M. Lu, H. Zhao, A. Yao, Y. Chen, F. Xu, and L. H. Zhang, "A closed-form solution to universal style transfer," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5951–5960, 2019.

[18] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414–2423, 2016.

[19] Y. Lu, Y.-W. Tai, and C.-K. Tang, "Conditional cyclegan for attribute guided face image generation," *ArXiv*, vol. abs/1705.09966, 2017.

[20] M. Wang, G.-Y. Yang, R. Li, R.-Z. Liang, S.-H. Zhang, P. M. Hall, and S.-M. Hu, "Example-guided style-consistent image synthesis from semantic labeling," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[21] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. Van Gool, "Pose guided person image generation," in *Advances in Neural Information Processing Systems*, pp. 405–415, 2017.