International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 12 | Dec 2020 www.irjet.net p-ISSN: 2395-0072

A Review on Image to Image Translation using Generative Adversarial Networks

Sunchit Lakhanpal[1], Akshat Jaipuria[2], Saurav Banerjee[3], Shaurya Pandey[4]

[1,2,3,4]*Dept. of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bangalore, K.A., India*

Abstract – With the advent of Machine Learning and Deep Learning, businesses can save time, costs and manual labor, editing visual content. Generative Adversarial Networks can reconstruct images, complete missing parts and make creative changes, which are otherwise impossible with image editing software. Generative Adversarial Networks can generate images from scratch or from a semantic input to automate the content creation process. We can generate photo-realistic images using sketches or semantic images as input which can be used for creating synthetic training data for visual recognition algorithms and for forensic recognition in criminal identification

Keywords: **Deep Learning, Artificial Intelligence, Generative Adversarial Networks, Image Translation, Review**

LITERATURE REVIEW

In 2014, Ian J. Goodfellow et al. [1] proposed a new framework for estimating generative models via an adversarial process that trains two models, a generative model G and a discriminative model D based on the idea of a minimax two player game. They proposed that there is no requirement of any Markov chains or unrolled approximate inference networks during both training and generation.

In 2016, Guim Perarnau et al. [2] demonstrated to successfully approximate complex data distributions. They evaluated encoders to inverse the mapping of a conditional generative adversarial network (cGAN) which allowed them to reconstruct and modify real images of faces based on arbitrary attributes. They introduced an encoder in a conditional setting within the GAN framework, a model which they called the Invertible Conditional GANs (IcGANs).

In 2017, Augustus Odena et al. [3] introduced new methods for improved training of generative adversarial networks for image synthesis. They constructed a variant of GAN's employing label conditioning that results in 128×128 resolution image samples exhibiting global coherence. They expanded on previous work on image quality assessment to provide two new analyses for assessing the discriminability and diversity of samples from class-conditional image synthesis models.

In 2018, Ting-Chun Wang et al. [4] proposed a new method for synthesizing high resolution photo-realistic images from semantic label maps using conditional generative adversarial networks. In this work, they generated 2048×1024 visually appealing results with a novel adversarial loss, as well as new multi-scale generator and discriminator architectures. Their method significantly outperformed existing methods in both quality and resolution of deep image synthesis.

In 2018, Yongyi Lu et al. [5] proposed to use sketch as a weak constraint, where the output edges do not necessarily follow the input edges. They proposed a joint image completion approach where the sketch provides the image context for completing and generating the output image. Their experiments evaluated on three different datasets showed that their contextual GAN can generate more realistic images than state-of-the-art conditional GAN's on challenging inputs.

In 2018, Philip Isola et al. [6] proposed conditional adversarial networks as a general purpose solution to image-to-image translation problems. These networks not only learn the mapping from input to output image but also learn a loss function to train this mapping. They demonstrated that their approach is effective at synthesizing photos from label maps, reconstructing objects from edge maps, colorizing images, among other tasks.

In 2019, Dingdong Yang et al. [7] proposed a simple yet highly effective method that addressed the mode collapse problem in cGAN. They proposed to explicitly regularize the generator to produce diverse outputs depending on latent codes. They demonstrated the effectiveness of their method on three conditional generation tasks: image-to-image translation, image inpainting and future video prediction. The simple addition of their regularization to existing models led to diverse generations, substantially outperforming previous approaches for multi-modal conditional generation specifically designed in each individual task.

In 2019, Taesung Park et al. [8] proposed spatially adaptive normalization for synthesizing photo-realistic images given an input semantic layout that allows control over both semantic and style. They proposed that normalization layers tend to wash away semantic information. To address this, they used the input layout for modulating the activation in normalization layers through a spatially-adaptive, learned transformation. Their method had advantages over existing approaches regarding both visual fidelity and alignment with input layouts.

In 2019, Hao Tang et al. [9] proposed a novel approach named Multi-Channel Attention SelectionGAN that makes it possible to generate images of natural scenes in arbitrary viewpoints, based on an image of the scene and novel semantic map. The proposed SelectionGAN explicitly utilized the semantic information and consisted of two stages. In the first stage, the condition image and the target semantic map were fed into a cycled semantic-guided generation network to produce initial coarse results. In the second stage, they refined the initial results by using a multi-channel attention selection mechanism.

In 2019, Yunjei Choi et al. [10] proposed that a good image-to-image translation model should learn a mapping between different visual domains while satisfying diversity of generated images and scalability over multiple domains. They proposed StarGAN v2 that tackles both and showed significant improvement over the baselines. Their model can generate images with rich styles across multiple domains.

In 2020, Runtao Liu et al. [11] incorporated a self-supervised denoising objective and an attention module to handle abstraction and style variations that are inherent and specific to sketches. A two-stage translation task was proposed as opposed to existing works. This approach works effectively for not only spatially imprecise and geometrically distorted sketches but also without color and visual details. Their synthesis is sketch faithful and photo-realistic to enable sketch-based image retrieval in practice.

In 2020, Hajar Emami et al. [12] introduced the attention mechanism to the generative adversarial network architecture and proposed a novel spatial attention GAN model (SPA-GAN) for image-to-image translation tasks. SPA-GAN computes the attention in its discriminator and use it to help the generator focus more on the most discriminative regions between the source and target domains, leading to more realistic output images. Qualitative and quantitative comparison against state-of-the-art methods demonstrated the superior performance of SPA-GAN.

In 2020, Subhankar Roy et al. [13] proposed the approach for multi-source domain adaption (MSDA) based on GAN. They proposed to project the image features onto a space where only the dependence from the content is kept, and then re-project this invariant representation onto the pixel space using the target domain and style. In this way, new labeled images can be generated which are used to train a final target classifier. They tested their approach using common MSDA benchmarks, outperforming the state-of-the-art methods. In this work they proposed TriGAN, an MSDA framework which is based on data-generation from multiple source domains using a single generator.

In 2020, Jun-Yan Zhu et al. [14] presented an approach for learning to translate an image from a source domain to target domain in the absence of a paired example. Qualitative results were presented on several tasks where paired training data did not exist, including collection style transfer, object transfiguration, season transfer, photo enhancement, etc. Quantitative comparisons against prior methods demonstrated the superiority of their approach where paired examples were not present.

CONCLUSION:

This paper presents a survey on different methods and experiments on image to photo translation and image-to-image translation using generative adversarial networks. There are several such methods for semantic image to photo translation but still the scope of research is vast in this domain in terms of style transfer techniques and the application of such methods in forensic recognition. From the study of above mentioned methods, we come up with the following conclusion. Early approaches in this field focused on using conditional GAN's (cGAN) [1] and auxiliary classifier GAN (AC-GAN) [2] which were then outperformed by contextual GAN [5] architecture using the sketch as a weak constraint. The conditional GAN's were vulnerable to mode collapse problems [7] and hence spatially adaptive normalization [8] was proposed that had advantages with visual fidelity. SelectionGAN [9] made it possible to generate images based on a semantic map while StarGAN v2 [10] generated images across multiple domains. SPA-GAN [12] could generate more realistic output images against other methods and is considerably lightweight and simpler. Further research [14] presented approaches where paired examples were absent, which can open the scope of further research in the topic where paired examples do not exist. In conclusion, spatially adaptive normalization (SPADE) [8] is currently the most competitive of approaches since it gives much more control to the user in terms of style and semantic. The proposed normalization leads to the first semantic image synthesis model that can produce photo-realistic outputs for diverse scenes including indoor, outdoor, landscape, and street scenes.

REFERENCES

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. (2014) Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2 (NIPS'14). MIT Press, Cambridge, MA, USA, 2672–2680.

[2] Perarnau, G., Weijer, J.V., Raducanu, B., & Álvarez, J.M. (2016). Invertible Conditional GANs for image editing. ArXiv, abs/1611.06355.

[3] Augustus Odena, Christopher Olah, and Jonathon Shlens. (2017) Conditional image synthesis with auxiliary classifier GANs. In Proceedings of the 34th International Conference on Machine Learning – Volume 70 (ICML'17). JMLR.org, 2642–2651.

[4] T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz and B. Catanzaro, "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 8798-8807, doi: 10.1109/CVPR.2018.00917.

[5] Lu Y., Wu S., Tai YW., Tang CK. (2018) Image Generation from Sketch Constraint Using Contextual GAN. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11220. Springer, Cham.

[6] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.

[7] Yang, D., Hong, S., Jang, Y., Zhao, T., & Lee, H. (2019). Diversity-Sensitive Conditional Generative Adversarial Networks. ArXiv, abs/1901.09024.

[8] T. Park, M. Liu, T. Wang and J. Zhu, "Semantic Image Synthesis With Spatially-Adaptive Normalization," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 2332-2341, doi: 10.1109/CVPR.2019.00244.

[9] H. Tang, D. Xu, N. Sebe, Y. Wang, J. J. Corso and Y. Yan, "Multi-Channel Attention Selection GAN With Cascaded Semantic Guidance for Cross-View Image Translation," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 2412-2421, doi: 10.1109/CVPR.2019.00252.

[10] Y. Choi, Y. Uh, J. Yoo and J. -W. Ha, "StarGAN v2: Diverse Image Synthesis for Multiple Domains," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 8185-8194, doi: 10.1109/CVPR42600.2020.00821.

[11] Liu, R., Yu, Q., & Yu, S. (2019). Unsupervised Sketch-to-Photo Synthesis. arXiv: Computer Vision and Pattern Recognition. arXiv:1909.08313v3

[12] H. Emami, M. M. Aliabadi, M. Dong and R. Chinnam, "SPA-GAN: Spatial Attention GAN for Image-to-Image Translation," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2020.2975961.

[13] ROY S, SIAROHIN A, SANGINETO E, SEBE N, RICCI, E. 2020. TriGAN: Image-to-Image Translation for Multi-Source Domain Adaptation. arXiv preprint arXiv:2004.08769.

[14] J. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.