International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 05,
2019 ISSN: 1475-7192

Perceptual Faces Completion Using Self-Attention Generative Adversarial Networks

[1]Sarmista Satrusalya, [2]Mihrnarayan Mohanty

Abstract--- *This paper propose method based on self-attention generative adversarial networks (SAGAN) to accomplish the task of image completion wherever completed images become globally and domestically consistent. Using self-attention GANs with contextual and different constraints, the generator will draw realistic images, wherever fine details are generated within the damaged region and coordinated with the entire image semantically. To train the consistent generator, i.e. image completion network, this paper tend to use global and native discriminators wherever the global discriminator is responsible for evaluating the consistency of the whole image, whereas the local discriminator assesses the local consistency by analyzing local areas containing completed regions only. Last but not least, an attentive recurrent neural block is introduced to get the attention map regarding the missing part within the image, which is able to facilitate the subsequent completion network to fill content better. By comparison of the experimental results of various approaches on CelebA data set, our technique shows relatively good results. Traditional Convolutional GANs generate high-resolution details as a function of only spatially local points in lower-resolution feature maps. In SAGAN, details will be generated using cues from all feature locations. Moreover, the discriminator will make sure highly elaborate features in distant parts of the image are consistent with one another. Moreover, recent work has shown that generator conditioning affects GAN performance. Investing this insight, this paper tend to apply spectral normalization to the GAN generator and find that this improves training dynamics.*

Index Terms--- *Attention Mechanism, Image Completion, Semantic Completion, Neural Network, Computer Vision*

I. INTRODUCTION

Image synthesis is an important problem in computer vision. There has been remarkable progress in this direction with the emergence of Self- Attention Generative Adversarial Networks (SAGANs),[1] though many problems remain. As it will be wont to fill the occluded image regions or repair damaged photos, it's aroused widespread interest in a computer vision. This technology has also been extended to different related applications, like video completion. However, it's still a difficult problem because it typically requires a high-level semantic understanding of the scene[2]. It's not only necessary to complete the texture within the image however also understands the semantic anatomy of the scene and also the object is completed. Basically, for image completion tasks, the generator has to understand the data distribution of individual objects within the image and therefore,[3] the overall structure of the scene, similar to a real-life painter will draw one thing well. But structural characteristics and distributions totally different objects are quite different. It's difficult to directly learn distributions of a large range of different objects at once[4]. Usually speaking, learning the distribution of an equivalent kind of object is comparatively simple, like a face image, whose structure is comparatively fixed. Once the probability distribution of the object to be completed is understood, the completed task can become a natural state of being as a painter knows a way to draw the unfinished

Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, sarmisthasatrusalya@soa.ac.in, mihirmohanty@soa.ac.in

Received: 10 Nov 2018 | Revised: 20 Dec 2018 | Accepted: 02 Jan 2019

portrait[5]. Therefore, at this stage, several completion strategies are based on this idea. Here, this paper tends to review a number of the classical methods each within the past and within the present[6].

In this article, this paper tends to propose Self-Attention Generative Adversarial Networks (SAGANs), which introduce a self-attention mechanism into Convolutional GANs[7]. The self-attention module is complementary to convolutions and helps with modeling long-range, multi-level dependencies across image regions[4]. Armed with self-attention, the generator will draw images during which fine details at each location are carefully coordinated with fine details in distant parts of the image. Moreover, the discriminator also can a lot of accurately enforce difficult geometric constraints on the global image structure. Additionally,[8]to self-attention, this paper tend also to incorporate recent insights relating network conditioning to performance. Showed that well-conditioned generators tend to b perform better. This paper tends to propose implementing good conditioning of SAGAN generators using the spectral normalization technique that has previously been applied only to the discriminator[9].

II. RELATED WORK

A large range of literature exists for image inpainting, and because of space limitations, this paper tends to be unable to discuss all of these in detail. Groundbreaking adds that direction includes the same works and references in this[10]. Since our technique is based on generative models and neural networks, this paper is going to review relevant academic researches and technical works below.

II.I. Generative Adversarial Network

GANs was 1st introduced by Goodfellow et al.[11], that trains two adversarial networks at the same time to capture information distribution of input images. Therefore, a typical GANs network consists of a generator and a discriminator, during which the generator tries to find out the optimal transport mapping from inputs to outputs, and therefore,[12] the discriminator judges the standard of its generation. The generator continuously improves the ability to come up with images that may fool the discriminator to determine the artificial as a real one. However, the discriminator keeps up itself to inform whether the image is generated or real. By mutually learning with alternating iterations between the discriminators and generator,[13] a Nash equilibrium is going to be achieved in theory, when the generator will generate images that are visually plausible enough to form the discriminator unable to determine whether the image is artificial. The game between each will be expressed because of the following optimization of the min-max problem[14].

II.II. Attention model

Recently, attention mechanisms became an integral part of models that have to capture global dependencies. Specifically, self-attention.

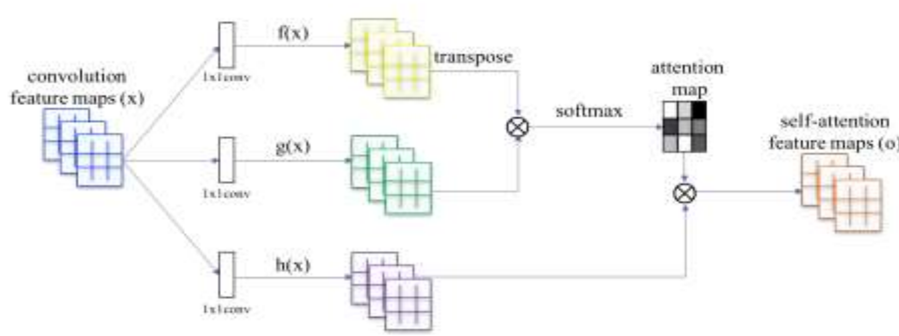

Fig. 1: Self Attention Mechanism

Also referred to as intra-attention, calculates the response at a position in a sequence by progressing to all positions among a similar sequence[15]. Machine translation models may come through progressive results by solely using a self-attention model. Image transformer model to feature self-attention into an autoregressive model for image generation. Wang et al. [16] formalized self-attention as a non-local operation to model the spatial-temporal dependencies in video sequences. In spite of this progress, self-attention has not however been explored within the context of GANs. (ASGAN uses attention over word embedding at intervals an input sequence, however not self-attention over internal model states)[13]. SAGAN learns to expeditiously notice the world, long-range dependencies among internal representations of images.

II.III. Image inpainting

Face inpainting is just one aspect of image restoration technology. Therefore, image restoration technology is extremely helpful for face inpainting. Nowadays, image completion problems will be thought of to be an application of image generation[17]. Particularly in recent years, there is a large range of educational papers on image generation, most of that is supported in the deep GANs framework. And most GANs frameworks are presented with image generation as examples. Like Radford et al. [18] additional developed deep Convolutional GANs (DCGANs) that have certain architectural constraints combining convolution neural network, and demonstrate that they're a powerful candidate for image generation. Kataoka et al. [19] mixed GANs with a special reasonably attention mechanism and generated a lot of real images. Ouyang et al. [20] planned a novel learning design of LSTM conditional generative adversarial networks to generate plausible images from word descriptions.

It is quite intuitive to use image generation technology to image completion. Moreover, within the real image completion task,[21] the data distribution of the image to be completed should be loaded, and therefore, the perfect image generator has naturally acquired this distribution. Several scholars have done the work of image inpainting and completion used AlexNet design because of the encoder with a completely unique channel-wise fully connected layer for feature learning for semantic inpainting[22]. Replacement technique supported the deep generative model that not only will synthesize novel image structures, however also make an improved prediction by using the surrounding image features as a reference throughout network training. In particular, an effective facial completion model using GANs with several losses[23].

II.IV. Visual Attention

Attention-related neural processes are extensively investigated in Neurobiology. Visual attention may be a fascinating aspect: several animals focus on specific elements of their visual inputs. This principle is of nice significance to the sensory system as we'd like to decide on the foremost pertinent a part of the knowledge,[24] instead of using all available data, a large part of that is independent of the response of the system. This is often very the same as the focusing function of a camera, that concentrates on one or two targets only and burs others as a background at a selected moment[25]. The same plan focusing on specific parts of inputs has been applied to any or all possible areas of deep learning, like natural language processing, reasoning, and computer vision. A multi-objects detection model based on attention mechanism. The model is a type of recurrent neural network (RNN) trained with reinforcement learning, which is used to identify the most relevant regions in input images[26].

An attention-based model that automatically learns to annotate the content of images and gazes at salient objects and generates equivalent words in output sequences.

III. APPROACHES

Our technique is predicated on self-attention GANs and consists of two main parts: a generator module and a discriminator module[27]. The generator module consists of an attention long short term memory (LSTM) module and an entire module. The discriminator module is formed from a global discriminator and a local one. Additionally,[28] there's a face semantic segmentation web, referred to as face parsing network that may be a retrained model and remains fixed, for additional ensuring the recently generated face and also the corresponding ground truth a lot of consistent semantically. In practice, it's used to figure the semantic loss between the synthetic and real ones. Next, we'll introduce every module of the complete model in detail[29].

III.I. Attention LSTM Module

It is a recurrent neural network (RNN) based module, wherever this paper tend to use long short-term memory (LSTM) rather than the standard RNN unit. The attention LSTM module finds the area of interest within the input image for repair[30]. These areas are primarily the missing areas and their encompassing structures necessary to complete the network, to help get higher local image restoration results, not only three generated attention maps target missing parts, however also concern about the encompassing face contour, like the opposite eye, the only nose, mouth, and hair in long-distance from the corrupted region. Every unit of the module consists of four parts—a Residual neural network (ResNet) block, a non-local neural block, an LSTM unit and a conv2d operation[31].

III.II. Discriminator Model

Our complete discriminator here consists of a local and a worldwide network, that are learned to differentiate synthesize contents in marked regions and assess whole images shaped by splicing the generated missing dispense with the input image[32]. At the top of the discriminator, two sub-discriminators can decide global and local consistency and facticity respectively, and each of them can output a scalar, that is, the probability of decision-

making to be true. The detailed structure is comparable to, that isn't any longer drawn here. The discriminator is simply helpful within the training stage, however, no longer within the testing stage[33].
The explanation for the existence of discriminators here is to train higher generators. Since just one hole was dug within the original image throughout training, only a local discriminator was used[34]. If you dig two holes, you must use two local discriminators, and then on. However, when there are several holes, too several local discrimination is unreasonable. Furthermore, if dividing the image into four blocks of fixedly and putting in four, local discriminators, this paper are going to encounter the problem of the amount of positive and negative examples mismatching in training. For simplicity's sake here, this paper tend to only dig one hole when training,[35] whose size and position is random. However, the input of local discriminator may be a mounted window larger than the hole.

IV. CONCLUSION

In this paper tends to develop a deep generative design for face completion. The network is predicated on a GANs with an encoder–decoder-like model as a generator. Apart from the damaged image itself as input to the generator, an attention map of the missing a part of the image is added as a region of inputs too, that comes from an attentive Long Short Term Memory (LSTM) module. The proposed model will successfully synthesize semantically valid and visually plausible content for the missing facial key parts. Each qualitative and quantitative experiments show that our model generates the completion results of high perceptual quality and is sort of flexible to handle a range of holes. However, the training model is extremely long and hardware resources-consuming (it needs to take a couple of months to train our model), it's necessary to optimize the energy consumption of the model within the future, hoping to use less hardware and get higher results in a shorter time.

REFERENCES

[1] K. E. Ak, J. Hwee, L. Jo, Y. Tham, and A. A. Kassim, "Attribute Manipulation Generative Adversarial Networks for Fashion Images," Iccv 2019, 2019.
[2] Y. Li, F. Wu, X. Chen, and Z. J. Zha, "Linestofacephoto: Face photo generation from lines with conditional self-attention generative adversarial network," in MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia, 2019.
[3] H. Zhang, B. S. Riggan, S. Hu, N. J. Short, and V. M. Patel, "Synthesis of High-Quality Visible Faces from Polarimetric Thermal Faces using Generative Adversarial Networks," Int. J. Comput. Vis., 2019.
[4] J. Zhang, R. Zhan, D. Sun, and G. Pan, "Symmetry-Aware Face Completion with Generative Adversarial Networks," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019.
[5] C. Wang, H. Zheng, Z. Yu, Z. Zheng, Z. Gu, and B. Zheng, "Discriminative Region Proposal Adversarial Networks for High-Quality Image-to-Image Translation," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018.
[6] L. Horsley and D. Perez-Liebana, "Building an automatic sprite generator with deep convolutional generative adversarial networks," in 2017 IEEE Conference on Computational Intelligence and Games, CIG 2017, 2017.
[7] R. Ma and H. Hu, "Perceptual Face Completion using a Local-Global Generative Adversarial Network," in Proceedings - International Conference on Pattern Recognition, 2018.
[8] 9. Retrieved from aarssenl@queensu.ca Aarssen, L. W., & Crimi, L. (2016). Legacy, leisure and the 'work hard—Play hard' hypothesis. The Open Psychology Journal et al., "Self-conscious emotion and existential concerns: An examination of the effect of shame on death-related thoughts.," 2017.
[9] 9. Retrieved from aarssenl@queensu.ca Aarssen, L. W., & Crimi, L. (2016). Legacy, leisure and the 'work hard—Play hard' hypothesis. The Open Psychology Journal et al., "Playing the objectification game: How women's self-esteem impacts the existential consequences of objectification.," 2018.
[10] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings, 2016.
[11] J. Ho and S. Ermon, "Generative adversarial imitation learning," in Advances in Neural Information Processing Systems, 2016.
[12] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018.

[13] J. Zhao, M. Mathieu, and Y. LeCun, “Energy-based generative adversarial networks,” in 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, 2019.
[14] T. Xu et al., “AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018.
[15] L. Mescheder, S. Nowozin, and A. Geiger, “Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks,” in 34th International Conference on Machine Learning, ICML 2017, 2017.
[16] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie, “Stacked generative adversarial networks,” in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017.
[17] X. Wang et al., “ESRGAN: Enhanced super-resolution generative adversarial networks,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019.
[18] T. Chen, M. Lucic, N. Houlsby, and S. Gelly, “On self modulation for generative adversarial networks,” in 7th International Conference on Learning Representations, ICLR 2019, 2019.
[19] S. Lloyd and C. Weedbrook, “Quantum Generative Adversarial Learning,” Phys. Rev. Lett., 2018.
[20] Z. Deng et al., “Structured generative adversarial networks,” in Advances in Neural Information Processing Systems, 2017.

[21] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, “Perceptual generative adversarial networks for small object detection,” in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017.
[22] I. Durugkar, I. Gemp, and S. Mahadevan, “Generative multi-adversarial networks,” in 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, 2019.
[23] M. Arjovsky and L. Bottou, “Towards principled methods for training generative adversarial networks,” in 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, 2019.
[24] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, “Generative adversarial text to image synthesis,” in 33rd International Conference on Machine Learning, ICML 2016, 2016.
[25] K. Lata, M. Dave, and N. K.N., “Data Augmentation Using Generative Adversarial Network,” SSRN Electron. J., 2019.
[26] B. Dolhansky and C. C. Ferrer, “Eye In-painting with Exemplar Generative Adversarial Networks,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018.
[27] A. Kuefler, J. Morton, T. Wheeler, and M. Kochenderfer, “Imitating driver behavior with generative adversarial networks,” in IEEE Intelligent Vehicles Symposium, Proceedings, 2017.
[28] J. Wu, C. Zhang, T. Xue, W. T. Freeman, and J. B. Tenenbaum, “Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling,” in Advances in Neural Information Processing Systems, 2016.
[29] J. Gauthier, “Conditional generative adversarial nets for convolutional face generation,” Tech. Rep., 2014.
[30] S. Gurumurthy, R. K. Sarvadevabhatla, and R. V. Babu, “DeLiGAN : Generative adversarial networks for diverse and limited data,” in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017.
[31] A. Jaiswal, W. AbdAlmageed, Y. Wu, and P. Natarajan, “CapsuleGAN: Generative adversarial capsule network,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019.
[32] G. Antipov, M. Baccouche, and J. L. Dugelay, “Face aging with conditional generative adversarial networks,” in Proceedings - International Conference on Image Processing, ICIP, 2018.
[33] C. Donahue, J. McAuley, and M. Puckette, “Synthesizing Audio with Generative Adversarial Networks,” in International Conference on Learning Representation (ICLR 2019), 2019.
[34] J. Toutouh, E. Hemberg, and U. M. O’Reilly, “Spatial evolutionary generative adversarial networks,” in GECCO 2019 - Proceedings of the 2019 Genetic and Evolutionary Computation Conference, 2019.
[35] C. Ledig et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017.