JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A RESEARCH STUDY ON GAN AND ITS APPLICATION IN COMPUTER VISION

Mrs. Jayalakshmi. V[1], Mrs. Deepa. G[2]
[1]Assistant Professor, Department of Computer Application,
DRBCCC Hindu College, Chennai, Tamilnadu, India.vkjlakshmi@gmail.com
[2]Assistant Professor, Department of Computer Application,
DRBCCC Hindu College, Chennai, Tamilnadu, India. nikideepa19@gmail.com

Abstract

Computer vision is one of the hottest areas of research in deep learning. The advent of Generative Adversarial Networks (GAN) offers new methods and models of computer vision. The idea of GANs using ancient machine learning algorithms in conditions of feature / characteristic learning and picture recognition / generation. The GAN is used not only in image generation and style transfer, but also in areas such as text, voice and video processing. However, GAN still has some problems such as model collapse and out-of-control training. This paper delves into the theoretical foundations of GAN and examines the recently developed GAN model in comparison to the traditional GAN model. GAN applications in computer vision include data expansion, domain transfer, high quality sample generation, and image recovery. Introducing the latest advances in research by GAN in the area of artificial intelligence (AI) -based security attacks and defenses. The future development of GAN in computer vision is also at the end of this paper, discussing possible applications of AI in computer vision.

Keywords: Generative Adversarial Network, GAN Types, Image Generation, Video Generation.

I.Introduction

Generative Adversarial Network is one of today's most intriguing computer science learning. An adversarial process is used to train two models at the same time. A generator learns to create realistic images, whereas a discriminator learns to distinguish between real and fake images. [1] If the goal of AI is to mimic human intelligence, the biggest challenge is creativity. Generative models are one of the most popular models today. [2] Colorized version, super-resolution, and artistic generation are just a few of the applications for generative models. The advantage of using a simulation model over a real model is that it can be faster to compute.Image-to-image translation can help designers draw and be more creative by using GANs to generate accurate representations of new medical images. Furthermore, even though we only have just one hundred image data and need more, for data augmentation GAN can be used.

II. Types of GAN

1.DCGAN

Deep Convolutional Generative Adversarial Network architecturefollows a few guidelines in particular. [3]Strided convolutions (discriminator) and fractional-strided convolutions are used to replace any pooling layers (generator). Both the generator and the discriminator use batchnorm. Fully connected hidden layers are omitted for deeper architectures. All layers except the output use [3] ReLU activation in the generator, which uses tanh. For all layers, LeakyReLU activation is used in the discriminator.

In computer vision applications, supervised learning with convolutional networks (CNNs) has been widely adopted. Unsupervised learning with CNNs, on the other hand, has gotten less press. We hope that research will contribute to closing the gap between CNN success in supervised and unsupervised learning. By introducing a new class of CNNs known as DCGANs that have architectural constraints and show that they are a good candidate for unsupervised learning, then by exhibiting convincing evidence that in our deep convolutional adversarial pair learns a hierarchy of characterizations from object parts to scenes both in the generator and discriminator by training on various image datasets. By applying the learned features to new tasks, demonstrating their utility as general image representations.

In both general computer vision and image processing, unsupervised representation learning is a well-studied problem. A classic approach to unsupervised representation learning is clustering the data (K-means) and leveraging the clusters for improved classification total score. [4] The clustering method of image patches can be used to understand powerful image representations in the case of images. [5] Another common method is to train auto-encoders to encode and decode an image into a compact code and decode the code to reconstruct the image as accurately as possible.

2.STYLE GAN

A style transfer-inspired alternative generator architecture for generative adversarial networks the new architecture provides intuitive, scale-specific control of the synthesis while also enabling an automatically learned, unsupervised separation of high-level attributes (e.g., posture and identification when trained on face images) and dynamical difference in the image features (e.g., freckles, hair). The new generator improves on the current state-of-the-art in terms of traditional allocation quality measures, gives rise to substantially better linearization properties, and better disentangles underlying features of variation. We have two new automated methods for quantifying interpolation quality and disentanglement that can be applied to any generator architecture. Finally, it will present a new dataset that is both diverse and high-quality.

The input latent code is embedded in an intermediate latent space, which has a significant impacton how the network's factors of variation are represented. The probability density of the training data must be followed by the input latent space, andresults in unavoidable entanglement. As a result, our intermediate latent space i s unrestricted and can be disentangled. Even though previous methods for estimating the degree of latent space disentanglement are not directly applicable, two new auto mated metrics for quantifying these aspects of the generator: perceptual path length and linear separability. When compared to traditional generator architecture, our generator admitsa more linear, less entangled strea m of data.

3. CycleGAN

Image-image translation is a type of vision and graphics problem in which the goal is to figure out how to map two images together using a training set of aligned image pairs tocompare an input image to an output image , regrettably, paired training data will not be available for many tasks. In the absence of paired examples, we present a method for learning to translate an image from one domain(X) to another(Y). Our g oal is to use an adversarial loss to learn a mapping G: X Y such that the distribution of images from G(X) is distinct from the distribution Y. We pair it with an inverse mapping F: Y X and introduce a cycle consistency loss to enforceF(G(X)) X because this mapping is highly under-constrained (and vice versa).Several tasks, such as collection style transfer, object transfiguration, season transfer, photo enhancement, and others, are presented with qualitative results in the absence of paired training data. The superiority of our approach is demonstrated by quantitative comparisons against several prior methods.

4. Conditional GANS

To avoid the difficulty of estimating many insolvable probabilistic computations, generative adversarial nets have recently been introduced an alternative framework for training generative models. Adversarial nets have the advantages of not requiring markov chains, only using back propagation algorithm to obtain gradients, requiring no implication during learning, and being able to easily integrate a number of different factors and interactions into the framework. It can also generate state-of-the-art log-likelihood estimates and feasible samples, as shown in [8]. There is no control over the modes of data generated in an unconditioned generative model. It is necessary to direct the data generation process by conditioning the model on additional information. Conditioning could be based on class labels, a portion of the data for inpainting, or even data from different modality.

If both the generator as well as the discriminator have been conditioned on certain multiple data collection y, including such class labels or information from other methodologies, generative adversarial nets can be extended to a conditional model. Conditioning can be done by feeding y into the discriminator and generator as an additional input layer.The previous input noises pz (z) and y are combined into a common hidden representation in the generator, and the adversarial training framework provides a lot of flexibility in constructing this hidden representation.

5. Progressive Growing GAN (PGGANS)

Our main contribution is a GAN training methodology in which we start with low-resolution images and gradually increase the resolution by layering the networks. Rather than learning all scales at once, this incremental approach allows the training to first explore the image distribution's large-scale structure before moving on to finer scale detail.The generator and discriminator networks that are mirror images of one another and grow in lockstep. Throughout the training process, all existing layers in both networks are trainable. We gradually introduce new layers into the networks, avoiding abrupt shocks to the already well-trained, lower-resolution layers.

We notice that progressive training has a number of advantages. Because there is less class information and fewer modes early on, the generation of smaller images is much more stable. By gradually increasing the res olution, we are posing a much simpler question as opposed to the ultimategoal of discovering a mapping fro m latent vectors to, say, 10242 images. This strategy is conceptually similar to that of [10]. In practise, it sta bilises the training to the point where we can synthesise megapixel-scale images reliably with WGAN-GP loss [11] and even LSGAN loss [12].Another advantage is the shorter training period. Most iteration are done at lower resolutions with progressively growing GANs, and comparable result quality can be obtained up to 2–6 times faster, depending on the final output resolution.

III.Applications of GAN

1. Image super-resolution:

Image super-resolution is the process of converting a low-resolution image into a high-resolution image. SRGAN [13] is the first image super-resolution framework that can derive photorealistic natural images with a 4x upscaling factor. Several other super-resolution frameworks ([14], [15]) have also been developed to achieve better results. The recently developed Best Buddy GAN [16] is used for single-frame super-resolution (SISR) tasks along with previous works ([17], [18]).

2. Image editing:

Image editing involves deleting or modifying some aspects of an image. For example, images taken in bad weather or heavy rain lack visual quality and require manual intervention to remove anomalies such as raindrops and dust particles that degrade image quality. The authors of IDCGAN [19] used GAN to address the issue of frame DE excitation. Image changes may include changing the color, adding smile, size, etc. in some aspects of the image.([20], [21])

3. HighResolution Face Imaging:

High resolution face imaging is another area of image processing where GAN excels. Face generation and attribute manipulation using GAN can be broadly divided into fully synthetic face generation, facial feature or attribute manipulation, and facial component transformation.

4. Synthetic face generation:

Synthetic face generationrefers back to theintroduction of artificialpicture of the face of those who do now no longer exist in actual life. ProGAN[22], defined with inside the preceding segment validated the technology of practicalsearchingpicture of human faces. Since then there werenumerous works which use GANs for facial picturetechnology ([24], [25]). StyleGAN[23] that isa completely unique generative opposed community added through Nvidia researchers in December 2018. The number one intention of StyleGAN is to generate excessivegreat face picturewhich areadditionallynumerous in nature. To acquire this, the authors used strategiesalong withthe usage of a noise mapping community, adaptive instance normalization and revolutionary developing much like ProGAN to provide very high resolution images.

5.VideoGeneration and Prediction

Video composition using GAN can be divided into three main categories. (a) Unconditional video generation (b) Conditional video generation (c) Video prediction

(a) Unconditional video generation: For unconditional video generation, the output of the GAN model is not conditioned on the input signal. The output video produced by these frameworks is usually of poor quality due to the lack of information provided as video conditions during the training phase. The author of VGAN [26] first used GAN for video generation. These generators consist of two CNN networks, a 3D spatiotemporal convolutional network that captures moving objects in the foreground, and a 2D spatial convolutional model that captures a static background. The two independent outputs of the generator are combined to create the generated video, which is sent to the discriminator to determine if it is real or fake.

(b)Conditional video generation: Audio-to-video compositing involves the task of generating synchronized video frames based on audio / audio input. Jalalifar et al. [27] used LSTMs and CGANs for voice-tuned generation of speaking faces. The LSTM network learns to extract and predict the location of facial features from voice features. Using a set of extracted landmarks, cGANproduces synchronized facial images with accurate lip synchronization. Vougioukasetal. [28] usedGAN to generate a video of talking head. Video frame generation relies on audio clips containing still images and sounds of people, not depends on removing intermediary features.

(c)Video Prediction: Video Prediction is the ability to predict future video frames based on the context of the sequence of previous frames. The network was trained using hostile training methods and an image gradient differential loss function. MCNet [29] performs the task of video image prediction by solving the temporal and spatial dynamics of video. The EncoderDecoder convolutional neural network is used to model video content, and the Convolutional LSTM is used to model the temporal dynamics or motion of video. Thus, predicting the next frame is as easy as using the detected motion features to convert the extracted content.

IV.Conclusion:

This paper introduced the current GAN models and their applications in various disciplines. Today, machines produce perfect images, and it is becoming increasingly difficult to distinguish machine-generated images from the original images. These GAN technologies are used in a variety of areas, including medicine and health care, astronomy, biology, entertainment (cartoon character generation, games), facial aging, and clothing translation. The popularity stems from the ability to learn highly non-linear correlations between latent space and data space. As a result, large amounts of unlabeled data that remain closed to supervised learning can be used. Discussed a lot about GAN aspects, including theory, applications,types. I believe this study will do thatsupporting scholars and industry researchers from different disciplines to fully understand GAN and its potential application.

References:

[1].Ian Goodfellow, Jean Pouget-Abadie, M. Mirza, Bing Xu, David Warde-Farley, SherjilOzair, Aaron C. Courville and YoshuaBengio. “Generative Adversarial Nets.” *NIPS* (2014).

[2].Ian Goodfellow. “NIPS 2016 Tutorial: Generative Adversarial Networks.” *ArXiv abs/1701.00160* (2017).

[3].Nair, Vinod and Hinton, Geoffrey E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807– 814, 2010

[4].Coates, Adam and Ng, Andrew Y. Learning feature representations with k-means. In Neural Networks: Tricks of the Trade, pp. 561–580. Springer

[5].Vincent, Pascal, Larochelle, Hugo, Lajoie, Isabelle, Bengio, Yoshua, and Manzagol, Pierre-Antoine. Stacked denoisingautoencoders: Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research, 11:3371–3408, 2010

[6].Zhao, Junbo, Mathieu, Michael, Goroshin, Ross, and Lecun, Yann. Stacked what-where autoencoders.arXiv preprint arXiv:1506.02351, 2015.

[7].Rasmus, Antti, Valpola, Harri, Honkala, Mikko, Berglund, Mathias, and Raiko, Tapani. Semisupervised learning with ladder network.arXiv preprint arXiv:1507.02672, 2015

[8].Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets.In NIPS’2014.

[9].Conditional Image Synthesis with Auxiliary Classifier GANs ***Augustus Odena, Christopher Olah, Jonathon Shlens*** *Proceedings of the 34th International Conference on Machine Learning*, PMLR 70:2642-2651, 2017.

[10].Photographic Image Synthesis With Cascaded Refinement Networks ***Qifeng Chen, VladlenKoltun***; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1511-1520

[11].Improved Training of Wasserstein GANs Ishaan Gulrajani1⇤ , Faruk Ahmed1, Martin Arjovsky2, Vincent Dumoulin1, Aaron Courville1,3 1 Montreal Institute for Learning Algorithms 2 Courant Institute of Mathematical Sciences

[12].K.B. Mao, Y. Ma, X.L. Tan, X.Y. Shen, G. Liu, Z.L. Li, J.M. Chen, L. Xia,Global surface temperature change analysis based on MODIS data in recent twelve years,Advances in Space Research,Volume 59, Issue 2,2017,Pages 503-512,ISSN 0273-1177,

[13] Christian Ledig, Lucas Theis, FerencHuszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, AlykhanTejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. 2017. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. arXiv:1609.04802 [cs.CV]

[14] Yuning Jiang and Jinhua Li. 2020.Generative Adversarial Network for Image Super-Resolution Combining Texture Loss. Applied Sciences 10, 5 (2020). https://doi.org/10.3390/app10051729

[15] Xintao Wang, Ke Yu, Shixiang Wu, JinjinGu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2019. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Computer Vision – ECCV 2018 Workshops, Laura Leal-Taixé and Stefan Roth (Eds.). Springer International Publishing, Cham, 63–79.

[16] Wenbo Li, Kun Zhou, Lu Qi, Liying Lu, Nianjuan Jiang, Jiangbo Lu, and JiayaJia. 2021. Best-Buddy GANs for Highly Detailed Image Super-Resolution. arXiv:2103.15295 [eess.IV]

[17] Xining Zhu, Lin Zhang, Lijun Zhang, Xiao Liu, Ying Shen, and Shengjie Zhao. 2020. GAN-Based Image Super-Resolution with a Novel Quality Loss. Mathematical Problems in Engineering 2020 (18 Feb 2020), 5217429.https://doi.org/10.1155/2020/5217429

[18] Thang Vu, Tung M. Luu, and Chang D. Yoo. 2019. Perception-Enhanced Image Super-Resolution via Relativistic Generative Adversarial Networks. In Computer Vision – ECCV 2018 Workshops, Laura Leal-Taixé and Stefan Roth (Eds.). Springer International Publishing, Cham, 98–113.

[19] He Zhang, VishwanathSindagi, and Vishal M. Patel. 2020. Image De-Raining Using a Conditional Generative Adversarial Network. IEEE Transactions on Circuits and Systems for Video Technology 30, 11 (2020), 3943–3956. https://doi.org/10.1109/TCSVT.2019.2920407

[20] GuimPerarnau, Joost van de Weijer, Bogdan Raducanu, and Jose M. Álvarez. 2016. Invertible Conditional GANs for image editing. arXiv:1611.06355 [cs.CV]

[21] Peiye Zhuang, OluwasanmiKoyejo, and Alexander G. Schwing. 2021. Enjoy Your Editing: Controllable GANs for Image Editing via Latent Space Navigation. arXiv:2102.01187 [cs.CV]

[22] TeroKarras, TimoAila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive Growing of GANs for Improved Quality, Stability, and Variation. CoRR abs/1710.10196 (2017).arXiv:1710.10196 http://arxiv.org/abs/1710.10196

[23] TeroKarras, Samuli Laine, and TimoAila. 2018. A Style-Based Generator Architecture for Generative Adversarial Networks. CoRR abs/1812.04948 (2018).arXiv:1812.04948 http://arxiv.org/abs/1812.04948

[24] TING ZHANG, WEN-HONG TIAN, TING-YING ZHENG, ZU-NING LI, XUE-MEI DU, and FAN LI. 2019. Realistic Face Image Generation Based on Generative Adversarial Network.In 2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing.303–306.https://doi.org/10.1109/ICCWAMTIP47768.2019.9067742

[25] Zhixin Zhang, Xuhua Pan, Shuhao Jiang, and Peijun Zhao. 2020. High-quality face image generation based on generative adversarial networks. Journal of Visual Communication and Image Representation 71 (2020), 102719.https://doi.org/10.1016/j.jvcir.2019.102719

[26] Carl Vondrick, HamedPirsiavash, and Antonio Torralba. 2016. Generating Videos with Scene Dynamics. In Proceedings of the 30th International Conference on Neural Information Processing Systems (Barcelona, Spain) (NIPS'16). Curran Associates Inc., Red Hook, NY, USA, 613–621.

[27] Seyed Ali Jalalifar, HoseinHasani, and Hamid Aghajan. 2018. Speech-Driven Facial Reenactment Using Conditional Generative Adversarial Networks. arXiv:1803.07461 [cs.CV]

[28] Konstantinos Vougioukas, Stavros Petridis, and Maja Pantic. 2019. End-to-End Speech-Driven Realistic Facial Animation with Temporal GANs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

[29] Ruben Villegas, Jimei Yang, Seunghoon Hong, Xunyu Lin, and Honglak Lee. 2018. Decomposing Motion and Content for Natural Video Sequence Prediction. arXiv:1706.08033 [cs.CV]