Application and Optimisation of GAN in Image Processing

Shiying Luo[a]
Guangdong Polytechnic Normal University, School of Electronic and Information Engineering Guangzhou, Guangdong, China

Keywords: Deep Learning, Generative Adversarial Networks, Image Synthesis, Image Translation, Image Style Transformation.

Abstract: Since the introduction of Generative Adversarial Networks (GAN), it has rapidly emerged in the field of image processing and demonstrated its strong application potential.GAN breaks through the limitations of traditional generative models through the adversarial training mechanism between generators and discriminators, and is able to generate high-quality images without the need for large amounts of labelled data. The continuous optimisation and improvement of GAN models by researchers have driven the rapid development of GAN-based image processing techniques. These improvements are mainly in the fields of image synthesis, image translation and image style transfer. In this paper, we systematically review the research of GAN in the three key areas of image synthesis, image translation and image style transfer, focusing on the contributions of various improved models in terms of performance optimisation and defect overcoming. By comprehensively combing the existing research, this paper aims to summarise the key technological advances of GAN in the field of image processing, reveal the challenges and limitations of the current research, and provide theoretical basis and practical guidance for future research directions.

1 INTRODUCTION

With the rapid development of deep learning technology, Generative Adversarial Networks (GAN) have gradually occupied an important position in the field of computer vision (Goodfellow et al., 2014).The particularity of GANs is reflected in the fact that they can be adversarially trained by generators and discriminators, which enables the generative model to automatically learn the complex distribution of data and generate highly realistic images. This technique has not only accelerated the progress in the field of image synthesis, but has also shown great potential for application in many fields.

GAN consists of two neural networks, the generator and the discriminator. The goal of the generator is to generate fake data as close as possible to the real data in order to deceive the discriminator. And the discriminator is to judge whether the input data is real or generated as accurately as possible, which is a zero-sum game. In image processing, through continuous adversarial training, the generator can gradually generate images that are highly similar to the real data. Therefore, the core advantage of GAN is that it does not require a large amount of labelled data, and can generate high-quality new images only through adversarial training, breaking the limitations of traditional generative models.

In addition, GAN, through adversarial training of generators and discriminators, the generative model is able to automatically learn the complex distribution of data and generate highly realistic images. This technique has not only pushed forward the progress in the field of image synthesis, but also shown great potential for application in the fields of image translation and image style transfer.

This paper focuses on three key areas, namely image synthesis, image style transfer and image translation, and investigates GAN-based image processing methods respectively, aiming to deepen people's understanding of GAN technology in the field of image processing, to provide valuable theoretical insights and practical guidance for researchers in related fields, and to promote the further research and development of GAN technology.

[a] https://orcid.org/0009-0003-0977-9823

Luo, S.
Application and Optimisation of GAN in Image Processing.
DOI: 10.5220/0013700000004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 492-496
ISBN: 978-989-758-765-8

2 DIFFERENT METHODS OF APPLYING GAN

2.1 Image Synthesis

2.1.1 Concepts of Image Synthesis

GAN based image synthesis technique generates brand new images through adversarial training mechanism. In this process, the generator and the discriminator play with each other, and the generator transforms random noise or simple inputs into highly realistic images by learning the distribution of the data. The generator continuously optimises its output to make the generated image closer and closer to the real data, thus gradually improving the image quality during adversarial training.

2.1.2 Improved Methods for Image Synthesis

An optimisation model called LR-GAN was proposed by Yang et al. To a certain extent, it overcomes the problems of distortion that the foreground is prone to when the traditional GAN model synthesises images. It generates the foreground and background of an image through an unsupervised, recursive method, and this recursive structure allows the model to take into account previously generated objects and backgrounds when generating subsequent foreground objects, thus achieving contextual relevance rather than just generating a projection of a particular object. This approach results in synthesised images with higher semantic discrimination (Yang et al., 2017).

Li et al. proposed GLIGEN, an optimised text-to-image generation method. It freezes the weights of the original pre-trained model during the training process and gradually incorporates the new localisation information into the pre-trained model through a gating mechanism, which effectively avoids forgetting old knowledge in the process of learning new knowledge. It is thus able to combine textual descriptions and other conditional inputs (e.g. bounding boxes, key points, etc.) to form a composite input structure. The problem caused by the original text-to-image generation model due to the difficulty of expressing precise concepts such as position and size in the text itself is solved. Experiments on the COCO and LVIS datasets show that GLIGEN significantly outperforms existing supervised layout-to-image generation baseline methods (Li et al., 2023).

Liao et al. proposed SSA-GAN, which is a text-to-image generation method. A simple semantic space-aware block SSA Block is introduced.This module is able to learn semantic masks in text through text-based semantic adaptive transformation and dynamically predicts semantic masks based on currently generated image features. The mask indicates which parts of the image need to be enhanced by textual information and applies the semantic mask to the normalisation parameter, which determines how much textual information should be applied at each pixel position, enabling spatial control over the fusion of textual information and avoiding the problem of conflicting textual and graphical information (Liao et al., 2022).

Johnson et al. developed an end-to-end method for generating images from scene graphs. The model takes as input a scene graph describing objects and their relationships and generates an image corresponding to that graph. The input graph is processed using graph convolution, the scene layout is computed by predicting the bounding boxes and isolation masks of the objects, and the layout is converted into an image with a cascading refinement network. This approach enables explicit reasoning about objects and relationships and generates complex images with many recognisable objects. The problem of difficulty in handling complex sentences with many objects and relations when generating images from text alone is avoided (Johnson, Gupta, & Fei-Fei, 2018).

2.2 Image Translation

2.2.1 Concepts of Image Translation

Image translation is the conversion of one type of image into another type of image, and GAN-based image translation can achieve this conversion through generative adversarial networks. One of the most classical applications is to convert a black and white image into a colour image, or a daytime scene into a nighttime scene.The power of GAN in this regard lies in its ability to capture complex mapping relationships between different image domains.

2.2.2 Improved Methods for Image Translation

Choi et al. proposed StarGAN, an approach that enables image conversion between multiple domains in a single model. While traditional methods require training different models independently for each pair of domains, StarGAN requires only one generator and one discriminator. By introducing target domain labels and auxiliary classifiers, it enables the generator to generate corresponding images based on the target domain labels and to flexibly translate the input images to any desired target domain, thus enabling flexible multi-domain transformation (Choi et al., 2018). [6]

Gao et al. proposed the SketchyCOCO framework and EdgeGAN model. For the first time, the use of hand-drawn sketches as input was proposed compared to text or scene graphs. This approach can become a straightforward and relevant presentation of the user's ideas and is more controllable.

The image generation process is decomposed into two consecutive phases based on the characteristics of the scene sketches; the first phase focuses on foreground generation and for each foreground object instance, EdgeGAN generates the image content separately. The second stage is responsible for background generation, where the pix2pix model is used to generate the background image. By using the generated foreground images as constraints, the network can generate a reasonable background that matches the foreground images, somewhat ameliorating the gaps caused by missing background inputs. This approach is highly robust, but the current version of the SketchyCOCO dataset still suffers from bias in terms of perspective diversity of foreground objects (Gao et al., 2020).

Isola et al. published a generalised solution to the image-to-image translation problem using conditional adversarial networks (cGANs), pix2pix. instead of manually designing a mapping function and a loss function, the method optimises this mapping process by learning a mapping relationship from the input image to the output image, while automatically learning a loss function. pix2Pix's generator uses the U-Net architecture, which preserves the low-level detail information of the input image by adding skip connections so that this information can be used more efficiently in the generation process. The discriminator of the PixPix employs the PatchGAN architecture, a local modelling approach that allows the discriminator to focus on high-frequency structures, thus generating clearer images. However, even though it outperforms traditional methods in terms of realism and clarity, the performance in generating semantic labels is still inferior to traditional L1 regression methods (Isola et al., 2017).

Liu et al. proposed UNIt, an unsupervised, coupled GAN-based image-to-image translation framework.

UNIt proposes a potential space assumption and shares this potential space. It enables images in two domains to be mapped to the same representation in the same potential space, and the model can learn the joint distribution between the two domains by sharing the potential space. And combines VAE and GAN, VAE is used to encode and decode the image to ensure that the input image can be reconstructed efficiently and GAN is used to generate the image in the target domain to ensure that the generated image is realistic. And through cyclic consistency constraints ensure that an image is reduced to the original image when it is translated and then translated back to the original domain (Liu, Breuel, & Kautz, 2017).

Siarohin et al. proposed a Deformable GAN. using Deformable Skip Connections and Nearest-Neighbor Loss to solve the problem of traditional GANs dealing with deformable objects (e.g., human posture changes) due to spatial mismatches between the inputs and outputs leading to the the poor generation effect of traditional GANs. By decomposing the human body into multiple rigid sub-parts and computing local affine transformations for each part, the feature maps in the encoder are spatially deformed according to the pose differences, and the deformed feature maps are passed to the decoder through jump connections. The nearest neighbour loss is then used to replace the traditional pixel-level loss. This approach can tolerate small spatial misalignments while preserving the texture details of the image, resulting in clearer and more realistic images (Siarohin et al., 2018).

Lu et al. proposed Contextual Generative Adversarial Network (Contextual GAN) for generating images from hand-drawn sketches.

Traditional conditional GANs require the generated image to strictly follow the edge information of the input sketch in the sketch-to-image generation task. In contrast, this approach uses the sketch as a weak constraint and redefines the sketch-to-image generation task as an image completion problem, where the sketch provides the context for generating the image. The sketch and the image are stitched together into a joint image and their joint distribution is learnt in the same space. And due to the symmetry of the joint image representation, this model can generate not only images from sketches but also sketches from images without additional training. This approach avoids the problem of losses caused by edges when the input sketches are of poor quality. However, the enhancement of this approach in appearance degrees of freedom may also lead to changes in certain image features (Lu et al., 2018).

2.3 Image Style transfer

2.3.1 Concepts of Image Style Transfer

Image style transfer refers to applying the style (e.g., colours, textures, strokes, etc.) of one image to another, while preserving the structural and semantic information of the content image. The goal is to generate a new image that has both the main structure of the content image and the visual style of the style image. This technique can transform photographs into different artistic styles or transform the style of one image into the style of another. By training the generator and discriminator, the GAN is able to learn the features of the target art style and apply them to the input image to generate output images with various artistic effects.

2.3.2 Improved Methods for Image Style Transfer

Conventional GANs usually struggle to handle multiple styles simultaneously in art style transfer, and especially perform poorly when dealing with collections of styles.The DRB-GAN proposed by Xu et al. dynamically adjusts the parameters of the convolutional layer and the adaptive instance normalisation layer by introducing a dynamic residual block, which takes the style code as a shared parameter. This dynamic adjustment flexibly adjusts the mean and variance in the feature space through learnable parameters, which better matches the feature statistical information between the content image and the style image and effectively avoids artefacts. In addition, DRB-GAN handles multiple styles in the style ensemble through a weighted average strategy, and the ensemble discriminator can maximally ensure that the styles of the final generated images are consistent through the comparison of the feature space (Xu et al., 2021).

Traditional GANs usually rely on second-order statistics (e.g., Gram matrix or mean/variance) as style representations in style transfer, which restricts the full utilisation of the style information and leads to possible local distortions and stylistic inconsistencies in the generated images.

The CAST method proposed by Zhang et al. introduces a multilayer style projector (MSP) to learn the style representation directly from the image features, which enables the MSP to represent the style information in a more comprehensive way by mapping different levels of features to independent style coding spaces. And contrast learning is introduced, which learns the distribution of stylistic features through the comparison between positive sample pairs (stylised images and their enhanced versions) and negative samples (other stylised images), thus improving the discriminative ability and consistency of the stylistic representation (Zhang et al., 2022).

Traditional GANs are usually difficult to achieve flexible control of character attributes (e.g., pose, clothing, texture details, etc.) in character image synthesis.The Attribute-Decomposed GAN proposed by Men et al. uses a pre-trained human body parser to extract the semantic layout of the source image, which can be re-combined to construct a style code by decomposing the character attributes into separate latent codes in the latent space where flexible control of these attributes is accomplished through blending and interpolation operations.

The introduction of the decomposition component coding module and the texture style transfer module, and the global texture coding module enables the model to learn and generate character images with user-specified attributes more efficiently, and better capture and generate complex texture features to produce more realistic results (Men et al., 2020).

3 CONCLUSIONS

GAN has made significant progress in the field of image processing by virtue of its unique adversarial training mechanism and powerful image generation capability. In this paper, we systematically review the research progress and technical optimisation of GAN from three key perspectives: image synthesis, image translation and image style transfer. In image synthesis, researchers have significantly improved the semantic recognition and quality of synthetic images by introducing recursive structures, semantic space-aware modules and scene graphs. In the field of image translation, GAN is able to realise multi-domain image conversion, introduce target domains and separate encoding to solve the problems of input-output spatial misalignment and semantic consistency in traditional methods. In the area of image style transfer, GAN effectively improves the accuracy and consistency of style representation through innovative methods such as dynamic residual block, multi-layer style projector and contrast learning, and achieves flexible processing of multiple styles.

Although GAN has made many advances in the field of image processing, it still faces some challenges. For example, the instability of the training process, the pattern collapse problem, and the dependence on high-quality data limit its wider adoption in practical applications. In the future, researchers can start from the directions of improving training strategies, enhancing model robustness, and exploring unsupervised or weakly supervised learning methods to promote the further development of GAN technology. By systematically combing the research progress of GAN in image processing, this paper aims to provide comprehensive theoretical references and practical guidance for researchers in related fields. At the same time, this paper also points out the limitations of the current research and looks forward to possible future research directions, with a view to providing useful insights for the further development and application of GAN technology.

REFERENCES

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797).

Gao, C., Liu, Q., Xu, Q., Wang, L., Liu, J., & Zou, C. (2020). SketchyCOCO: Image generation from freehand scene sketches. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5174-5183).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Isola, P., Zhu, J. Y., Zhou, T., & 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).

Johnson, J., Gupta, A., & Fei-Fei, L. (2018). Image generation from scene graphs. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1219-1228).

Li, Y., Liu, H., Wu, Q., Mu, F., Yang, J., Gao, J., ... & Lee, Y. J. (2023). GLIGEN: Open-set grounded text-to-image generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 22511-22521).

Liao, W., Hu, K., Yang, M. Y., & Rosenhahn, B. (2022). Text-to-image generation with semantic-spatial aware GAN. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 18187-18196).

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. Advances in neural information processing systems, 30.

Lu, Y., Wu, S., Tai, Y. W., & Tang, C. K. (2018). Image generation from sketch constraint using contextual GAN. In Proceedings of the European conference on computer vision (ECCV) (pp. 205-220).

Men, Y., Mao, Y., Jiang, Y., Ma, W. Y., & Lian, Z. (2020). Controllable person image synthesis with attribute-decomposed GAN. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5084-5093).

Siarohin, A., Sangineto, E., Lathuiliere, S., & Sebe, N. (2018). Deformable GANs for pose-based human image generation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3408-3416).

Xu, W., Long, C., Wang, R., & Wang, G. (2021). DRB-GAN: A dynamic resblock generative adversarial network for artistic style transfer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6383-6392).

Yang, J., Kannan, A., Batra, D., & Parikh, D. (2017). LR-GAN: Layered recursive generative adversarial networks for image generation. arXiv preprint arXiv:1703.01560.

Zhang, Y., Tang, F., Dong, W., Huang, H., Ma, C., Lee, T. Y., & Xu, C. (2022, July). Domain-enhanced arbitrary image style transfer via contrastive learning. In ACM SIGGRAPH 2022 conference proceedings (pp. 1-8).