Iraqi Journal of Intelligent Computing and Informatics (IJICI)
Vol. 4, 2. December 2025, pp. 103~112
ISSN: 2791-2868, DOI: 10.52940/ijici.v4i2.112

A Comprehensive Review Cross-Domain Image Translation: A Framework Using Generative Adversarial Networks and Variational Autoencoders

Sahar Jabbar Mohammed[1], Salah Aldarraj[2]
[1,2]Department of Computer Science, University of Basrah, Basrah, Iraq

Article Info

Article history:

Received May 11, 2025
Revised Jun 20, 2025
Accepted July 7, 2025

Keywords:

GANs
VAEs
Deep Convolutional GANs
Framework

ABSTRACT

Within the extensive array of image generative models, two models are particularly notable: Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN). Generative Adversarial Networks (GANs) can generate realistic images; nevertheless, they are prone to mode collapse and lack straightforward methods for obtaining the latent representation of an image. Conversely, VAEs do not encounter these issues; yet, they frequently produce images that are less realistic than those generated by GANs. This article elucidates that the absence of realism is partly attributable to a prevalent overestimate of the dimensionality of the natural image manifold. To address this issue, we propose a new framework that integrates VAE with GAN in a unique and complementary manner, resulting in an auto-encoding model that retains the features of VAEs while creating images of GAN quality. We assess our methodology using both qualitative and quantitative analyses across five image datasets.
We introduce a comprehensive learning system that integrates a deep convolutional GAN network with a variational autoencoder network. Initially, we identified a technique that addresses the issue of images generated by GANs typically being unclear and distorted. In this scenario, the integration of GAN with VAE may be a more advantageous option.

Corresponding Author:

Sahar Jabbar Mohammed
Department of Computer Science, Basrah University, Basrah, Iraq
Email: sahar.jabbar@uobasrah.edu.iq

1. INTRODUCTION

In machine learning, generative models have become a transforming technique allowing the synthesis of intricate data distributions over fields including picture production and translation. These models seek to replicate the fundamental data distribution to generate outputs with real-world data likeness [1]. Since Cycle Generative Adversarial Networks (CGANs) has been released, numerous methods have been proposed which try to address various problems from different perspectives [2]. Generative models have transformed disciplines such computer vision, where realistic picture generation and domain translation are crucial tasks, by mapping latent representations to high-dimensional outputs. Especially, its uses cover creative media, medical imaging, and data augmentation, providing creative responses to problems needing synthetic but realistic data [3].

For many real-world uses, it is challenging to create a large enough, high-quality labelled dataset. There are a number of factors that can contribute to this, including: (i) the possibility of noisy data and the associated computational or financial costs of denoising or curating it; (ii) the impracticality, prohibitiveness, or infeasibility of obtaining samples from rare classes or events due to the large observation times required; and (iii) the possibility of class imbalance or loss of diversity due to implicit biases in the collected data [4]. Via utilising an adversarial training framework, GAN methodically improves the quality of visual perception by increasing picture fidelity, narrowing the difference between the two over time [5]. Recent development in neural networks has greatly improved the generalization ability of image domain transfer algorithms, in which two parallel lines of development prevails: using Neural Style Transferand using Generative Adversarial Networks (GAN) [6].

Popular generating models include stable diffusion models, variational autoencoders (VAEs), and generative adversarial networks (GANs). Anomaly detection and image production are only two of the many potential applications for these systems' ability to learn data representations and generate new information in creative ways. It is easier to generate a range of outputs when employing VAEs since they rapidly encode input into a latent space using a probabilistic framework. Unexpected training dynamics and mode collapse are two of the downsides of GANs' adversarial training strategy, but overall, it has changed the game when it comes to making high-quality, realistic images. In order to address some of the issues with VAEs and GANs, Stable Diffusion models employ recurrent refinement techniques to generate images that are both highly detailed and coherent in terms of semantics. Despite their popularity, these models have limitations that render them useless in various contexts. Although stable diffusion models are computationally and computationally intensive, they produce excellent results. Conversely,VAEs often struggle to generate differentiable images. Training stability and variation are challenges for GANs [7]. With vast amounts of visually impaired data, most of it is noise-related. Critical image processing tasks including object detection, document digitization, and image recognition are all severely impacted by noise, which lowers picture quality and utility. Picture denoising is an essential first step in many image processing pipelines. Using classic image-denoising algorithms often leads to blurred images and the loss of high-frequency components since these algorithms struggle to maintain image properties.

By integrating denoising capabilities with stability properties of GANs, a new hybrid architecture is proposed to generate denoised images while preserving the primary significant features observed in binary images. By means of integrating the strengths of the two models, we can get greater denoising performance crucial for binary images and therefore enhance the quality of digital documents [8]. A subset of image translation known as "exemplar-based image translation" seeks to produce an image with an exemplar image style while preserving the input image's content. There are several promising uses for these techniques, including scene transformation, face editing, and style transfer [9]. Unsupervised image translation by resolving the drawbacks of current techniques and putting forth a novel framework that makes use of generative priors to enhance performance in a variety of fields [10].

2. Background

Deep learning is one cutting-edge artificial intelligence (AI) technique that has gained popularity recently [11]. Multiple-layer neural networks are utilized to analyze different kinds of data. These are some crucial details regarding CNNs in particular and deep learning in general. The input layer, convolutional layers, pooling layers, fully connected layers, and output layer are some of the layers that make up a CNN. Each layer performs a variety of tasks, including feature extraction and categorization [12].

"Deep learning" is a robust subfield of machine learning that models complex patterns in data using multi-layered neural networks. Because it can yield cutting-edge results in a range of applications, including computer vision and pictures, it has garnered a lot of interest.

Through trying to translate images across multiple domains, deep learning has fundamentally altered image-to-image translation. Conventional methods may require matching examples of source and destination images to comprehend the mapping across domains. Deep learning algorithms, like GANs, may be able to accomplish this operation without supervision, therefore pairing data is not required [13].

2.1 GENERATIVE ADVERSARIAL NETWORKS (GANs)

Introduced in 2014, GANs play a crucial role in deep generative models that do not require supervision. They engage in a minmax game as a generator and discriminator. The generator produces realistic samples, and the discriminator determines whether or not they are real [14]. In a dynamic adversarial process, generative adversarial networks (GANs) produce synthetic data [15]. This class of deep learning models is known for this ability. In a GAN, the two main components are the generator and the discriminator. The generator creates fake data samples, which the discriminator then checks against real data samples to see if they are accurate [16]. This method is known as adversarial training, and it involves training the generator and discriminator at the same time. Both the discriminator and the generator are constantly working to enhance their abilities; the discriminator is trying to detect genuine images more accurately, while the generator is trying to make more realistic ones. Because of this rivalry, both networks keep getting better over time [17].

The Role of GAN Architecture in Map Generation Understanding. A robust approach for creating fresh data samples that mimic a specific dataset is (GANs). Getting good results when making maps online from aerial photos relies heavily on the design of GANs [18]. By improving the overall performance of the, these loss functions show how important it is to use custom loss functions when training generative models to convert images accurately.

The authors guarantee structural integrity and genuine seasonal aspects in the generated images out of meticulously developing these loss functions [19].

Challenges (GANs) Present a number of obstacles that might reduce their effectiveness, despite the fact that they have utterly transformed generative modeling. Some of the most significant problems with conventional GANs are as follows:

When a GAN's generator only generates a small subset of the possible outputs, a problem known as "mode collapse" occurs, and the model is unable to adapt to the wide variety of training data. The quality of the generative model is negatively affected because of the lack of variety in the generated samples. The training dynamics of GANs are notoriously unstable. Importantly, the generator and discriminator must be in harmony with one another; otherwise, convergence will be weak and the outcomes will be subpar. Consistent performance across training runs is difficult to obtain due to this volatility. Because of this, the discriminator provides very little helpful feedback to the generator, making it impossible for it to learn. Hyperparameter Sensitivity: Tuning learning rates and network topologies, two examples of hyperparameters, is commonly necessary for GANs. Because of this sensitivity, training can become more complicated and results from diverse studies can be hard to reproduce. Traditional GANs fail to offer a meaningful loss indicator that is both transparent and corresponds well with the quality of the samples generated. Because of this, hyperparameter adjustment, debugging, and evaluating the model's performance during training can be somewhat tough. The problem of discriminator overfitting arises when the discriminator gets too good at telling the difference between real and false samples, which leaves the generator without any relevant gradient information. In turn, this can make the generator less effective during training [20]. When it comes to computer vision tasks involving domain transformation, image translation (GANs) are incredibly effective. The term "image-to-image translation" describes a method for editing individual pixels in a picture, for as changing a character's expression or the background. With the advent of GANs, which can produce high-quality images using features learnt from training data, this task has witnessed tremendous progress [21].

GANs find use in a wide range of contexts, including: The term "style transfer" describes the process of altering an image's aesthetic while preserving its original content. Dividing a picture into smaller pieces for more manageable processing and analysis is known as image segmentation [22]. Architecture of Generative Adversarial Network is shown in Figure 1[23].

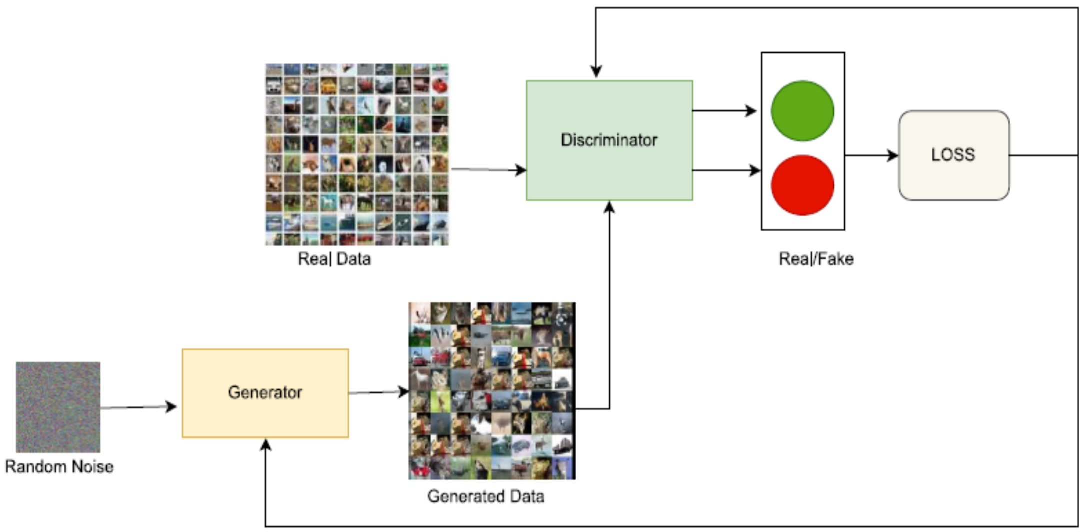

Figure 1. Architecture of Generative Adversarial Network

2.1.1. Conditional (cGANs)

Conditional Generative Adversarial Networks (cGANs) are a type of GAN that incorporate conditional information into the generative process. Unlike standard GANs, which generate images purely from random noise, cGANs take an additional input — such as class labels or conditional images — to guide the generator in producing more controlled outputs [24].

A cGAN consists of two competing neural networks:

Generator (G): Takes =both random noise and conditional input (such as an image or label) to produce a realistic output.

Discriminator (D): Receives the real and generated images, along with the conditional input, and learns to distinguish between authentic and fake samples [25]. Equation is the aim function 1 [26]. Conditional (cGAN) architecture is depicted in Figure 2 [27].

$$G_{min} D_{max} V(D,G) = E_{x\sim pdata(x)}\left[\log D(x|y)\right] + E_{z\sim pz(z)}\left[\log(1 - D(G(z|y)))\right] \quad (1)$$

Where y: serves as a class label (conditional information), G(z|y): is the generator's output, conditioned on y, which aims to generate samples that resemble x given the label y, and D(x|y): is the discriminator's probability estimate that x is an original sample given the condition y.

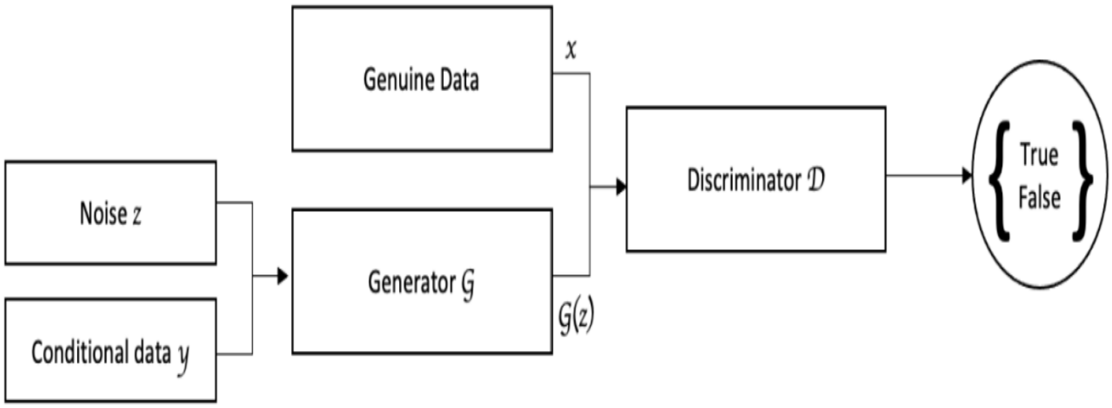

Figure 2. An architecture of conditional generative adversarial network

2.1.2 Deep convolutional (DCGANs)

Are a subset of GANs that use convolutional layers to enhance picture generating stability and performance. They were created to address the mode collapse and inadequate convergence issues that arise during GAN training [28]. Strong artificial intelligence frameworks that can generate visuals are called Deep Convolutional Generative Adversarial Networks, or DCGANs.

The DCGANs integrating the concepts of Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) to improve the output quality of images. Integrating these two uses the processing and image recognition capabilities of CNNs [29]. The discriminator and generator networks compete with one another in the DCGAN, which is a minimax, zero-sum game. The discriminator network, which attempts to distinguish between real and phony images, can b0e tricked by the generator's increasing ability to produce realistic images. A Z dimensional latent vector, where Z is a vector located in the Z-dimensional subspace known as the latent space, is fed through a sequence of convolutional layers by the generator to create an image tensor [30].

DCGANs follow an all-convolutional network architecture by having the discriminator and generator train their own spatial upsampling and downsampling. To make DCGANs much more reliable and trainable, several methods — which discriminate using real photos — such as batch normalization, rectified linear units in the generator, and leaky rectified linear units in the discriminator — have been upgraded. They are thus being extensively studied and incorporated into contemporary GAN designs [31].

2.1.3 Cycle generative adversarial networks (Cycle GANs)

CycleGAN: Two translation procedures are used in this model:

A generator network (Enx) and a decoder (Dey) are required for forward translation (X → Y), which converts images from domain X to domain Y.

Reverse translation (Y → X): This method converts images from domain Y to domain X using a decoder (Dex) and an alternative generator (Eny). The translation process's integrity is maintained by this cycle consistency [32].

Each Cycle GAN generator includes: Stride: 2; Kernel size: 3 Layers of Convolution: While extracting features, these layers make it easier to downsample input images. A robust foundation for unsupervised image translation is established by the Cycle GAN model's dual translation methods and cycle consistency.

Via ensuring that a picture that has been altered from one domain to another may be returned to its initial state, Cycle GAN works on the principle of cycle consistency. Cycle GAN assumes a shared latent space between domains, which may lead to domain-specific information being included in the translated image in order to preserve cycle consistency [33].

To guarantee that translated images may be returned to their original domain, one such model is Cycle GAN, which includes a cycle-consistency loss [34]. Ensure uniformity Another way to think of cycle consistency is as a kind of regularization. An increase in cycle consistency loss and wasteful information loss can result from excessive hallucinations and mode collapse in generators, which are prevented by the CycleGAN architecture.

Unusable artifacts Despite its advantages, cycle regularity has drawbacks. Pixel-by-pixel cycle consistency is preserved. Even when information loss is necessary during translation, it maintains that there is no loss of information and assumes a one-to-one relationship between the two picture domains, Cycle GAN architecture is depicted in Figure 3 [35].

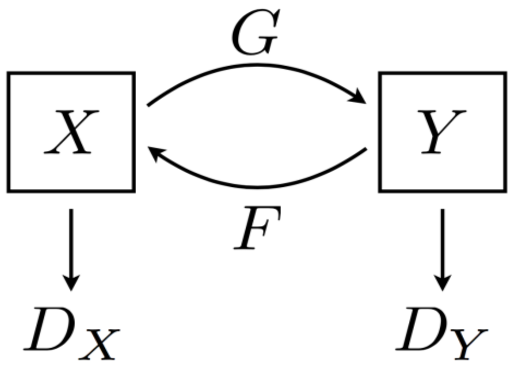

Figure 3: CycleGAN architecture.

2.2 Variational autoencoders (VAEs)

Are neural networks that learn lower-dimensional input representations through an unsupervised approach. An encoder converts the data into this representation, or the latent space, and a decoder uses the encoded data to recreate the original data. Variational autoencoders (VAEs) differ from vanilla autoencoders in that they learn a continuous latent space by transforming the latent space model into a probability distribution [36]. Advantages of Pixel Space Compared to VAEs Because VAEs must predict high-frequency details when trained directly on pixel-space, image blur results [37].

The key insight of VAEs is to learn the latent distribution of data in such a way that new meaningful samples can be generated from it. This approach led to tremendous research and variations in the architectural design of VAEs, nourishing the recent field of research known as unsupervised representation learning. In this article, we provide a comparative evaluation of some of the most successful, recent variations of VAEs [38].

VAEs also have two modules: an encoder and a decoder, although they are not competitors in this case. The decoder has been trained to use the few important variables that the encoder finds to characterise the attributes of the input data in order to recreate the original data. Modifications to VAE yield outcomes that are comparable to those of GANs, including the Vector Quantised Variational Autoencoder (VQ-VAE-2) (Razavi et al. 2019). The versatility of VAEs enables their application to progressively more intricate generation tasks [39].

Machine learning has relied heavily on VAEs since its inception. There remain many unresolved problems regarding their theoretical properties despite their widespread use. In this paper, PAC-Bayesian theory is used to construct the statistical guarantees for VAEs. The first PAC-Bayesian bound is initially obtained using the distribution that produces the data for posterior distributions conditioned on individual samples. Upper constraints on the distance between the input and regenerated distributions are then established using this approach, together with generalisation guarantees for the reconstruction loss of the VAE [40]. Architecture of Autoencoder is depicted in Figure 4 [41].

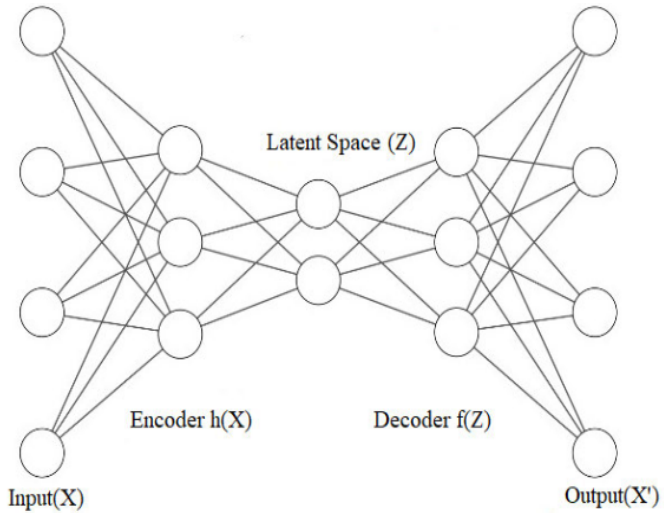

Figure 4: Architecture of Autoencoder

2.3. Hybrids VAE and GAN

The complementing strengths and constraints of GANs and VAEs have driven academics to investigate hybrid architectures using the best of both worlds. Hybrid models seek to mix the strong latent representation capabilities of VAEs with the sharp image quality of GANs by including VAEs with GANs. Many of the flaws of stand-alone models these hybrids solve.

Usually leveraging the encoder-decoder structure of VAEs to generate a structured latent space, VAE-GAN hybrids then feed GANs from this ordered latent space. This integration lets the hybrid model create outputs with variety and excellent quality. For example, by offering more useful latent representations and thereby minimizing mode collapse , the VAE's probabilistic architecture helps stabilize GAN training.
Their capacity to maximize a integrating loss function integrating the adversarial loss of GANs and the reconstruction loss of VAEs is one clear benefit of VAE-GAN hybrids. This mix guarantees the interpretability and diversity of the latent space while yet producing reasonable images from the model. Medical imaging, artistic content development, and e-commerce where both quality and variability are crucial are just a few of the many disciplines where these hybrids find use.

VAE-GAN hybrids have certain difficulties even with their potential. Because of the architectural complexity, training these models may be computationally costly. Furthermore under continuous study is the design of an efficient loss function that strikes a compromise between adversarial integrity and reconstruction accuracy. Still, the combination of VAEs with GANs shows a good path for developing generative modelling [42].

To solve problems, we suggest a Pyramid-VAE-GAN (PVG) network. To describe complex high-dimensional prior distributions of pictures, our model encodes latent variables using a variational autoencoder (VAE) backbone [43]. VAEs are generally easy to train, but the generated results have low quality due to imperfect measures such as the squared error. On the other hand, GANs generate samples with higher quality, but they suffer from training instability. In order to improve the training process and the quality of the generated samples, some researchers suggested hybrid VAE- GAN models [44].

We confrm that the hybrid model takes advantages of both the high coverage of VAE and the high fdelity of GAN simultaneously. Another advantage of the hybrid model is that it is more stable than the standalone GAN. Tis can be observed from the fact that the coverage and energy metric values of the fve independent trials of hybrid models show less dispersion compared to those from the standalone GANs [45].

3. Applications

Key Applications of GANs:
Cross-Domain Image Translation – Maps images between domains while preserving their structure.
Style Transfer – Ensures stylistic features of a target reference image are reflected while maintaining original content. Image Editing – Modifies textures and colors without altering structural elements.

Super-Resolution & Colorization –Enhances image resolution and applies accurate color mappings [46].

Applications of VAEs :
Image Synthesis–VAEs are used for generating realistic images across domains such as fashion, art, and medical imaging. Representation Learning.
VAEs help disentangle underlying factors of variation in data, leading to more interpretable and controllable representations.
Semi-Supervised Learning – VAEs are applied in scenarios where labeled data is scarce, leveraging their ability to model data distributions. Data Augmentation – VAEs generate synthetic data to improve the robustness of machine learning models [47].

This paper presents a GAN-based super - resolution model for enhancing medical images. The proposed method is evaluated across multiple medical imaging modalities, including:
Retinal Fundoscopy Images – DRIVE and STARE datasets. Brain MRI Scans – BraTS dataset. Skin Cancer Dermoscopy Images – ISIC dataset. Cardiac Ultrasound Images – CAMUS dataset.

The architecture improves image resolution while preserving fine details, which is crucial for accurate diagnosis in medical imaging [48]. Its applications span a variety of image-to-image translation tasks, including: Semantic label to street scene (e.g., translating urban segmentation maps into realistic street images).

Face to cartoon (transforming real human faces into cartoon versions).
Profile to frontal face (addressing face frontalization challenges using limited paired data) [49]. The key applications highlighted include: Automated map creation: Converting satellite images into human-readable maps without manual intervention. Urban planning and navigation: Enhancing maps for ride-sharing, delivery services, and driverless cars (e.g., Uber, Tesla) [50].

The key applications include:
Geospatial Image Generation: Using conditional transformations to generate realistic maps and satellite imagery for urban planning and remote sensing. Handwritten Character Synthesis: Generating Chinese handwritten characters for training OCR systems with large vocabulary datasets [51].

4. Related work

Table 1. Related work

Year	Authors	Paper title	Paper objective
2020	Xuerong Xiao, Swetava Ganguli, Vipul Pandey[4]	VAE-Info-cGAN: Generating Synthetic Images by Combining Pixel-level and Feature-level Geospatial Conditional Inputs	The model generates precise synthetic data, allowing for specific properties modification, making it useful for geographic applications like census maps and vector maps. The accuracy of the model VAE-Info-cGAN was evaluated using the Average Percentage Normalized Deviation (APND) metric. The results for the test set were: CRM Generation: 0.53% HCRM Generation: 0.98%
2020	Antoine Plumerault, Herve Le Borgne, Celine Hudelot[1]	AVAE: Adversarial Variational Auto Encoder	the AVAE paper's objectives focus on creating a novel generative model that combines the strengths of VAEs and GANs, addresses the limitations of existing models, and demonstrates its effectiveness through comprehensive evaluations on image datasets. Datasets Used:- (LSUN Bedroom, Celeb A,FFHQ,SVHN)
2024	Jeongik Cho , Adam Krzyzak[3]	Efficient integration of perceptual variational autoencoder into dynamic latent scale generative adversarial network	the paper focuses on improving generative models by effectively combining perceptual VAE with DLSGAN, aiming for enhanced performance in data generation and inversion while maintaining computational efficiency. Dataset used (FFHQ,AFHQ)
2022	Jiayi Chen, Wei Song[28]	GAN-VAE: Elevate Generative Ineffective Image Through Variational Autoencoder	The paper "GAN-VAE: Elevate Generative Ineffective Image Through Variational Autoencoder" enhances image generation using deep learning techniques, combining Generative Adversarial Networks with Variational Autoencoders to reduce noise, optimize computational resources, and address limitations of VAEs and GANs. Dataset used (fashion MNIST , celebA) Accuracy : SSIM(0.949), IS (31.863)
2023	Chengchao Wang , Guohong Gao, Jianping Wang , Yingying Lv , Qian Li , Zhiyu Li , Xueyan Zhang , And Haoyu WU[5]	GCT-VAE-GAN: An Image Enhancement Network for Low-Light Cattle Farm Scenes by Integrating Fusion Gate Transformation Mechanism and Variational Autoencoder GAN	The GCT-VAE-GAN paper presents a novel image enhancement network for low-light cattle farm scenes, integrating GCT, VAE, and GAN techniques. It addresses environmental challenges, features feature fusion, and optimizes performance with a joint loss function. Experiments demonstrate resilience. used a self-constructed dataset collected from a cattle farm in Yuanyang County, China. Accuracy: PSNR: 17.61 , SSIM: 0.521,NIQE: 3.15,LOE: 282

4. DISCUSSION

Examined is the basis for picture translation between domains. Cross-domain image translation is comprehensively investigated in this paper using variational autoencoders (VAEs) and generative adversarial networks (GANs). Typically, the following topics were discussed:

Building Generative Models According to the studies, "Generative models, specifically GANs and VAEs, have revolutionised machine learning by enabling the synthesis of complex data distributions." This tendency is particularly noticeable in computer vision tasks that are essential for many applications, such creating and analysing pictures.

Challenges to the present approach: Despite its advancements since its inception, CycleGAN's use remains restricted in many industries due to the task-specific nature of many of the current techniques. A general-purpose framework that integrate GANs and VAEs may be able to get beyond these restrictions, according to the study. The framework is a two-branch architecture, according to the authors, for learning a disentangled latent space to provide robust representations. Through tackling the problem of preserving semantic integrity during picture translation with a GAN-based pipeline, our method guarantees photorealistic outcomes. Unique characteristics: The adaptive attention approach is used to generate a multi-modal loss function by combining adversarial, perceptual, and cycle-consistency losses.

The model's cross-domain generalisation is enhanced by this method, enabling zero-shot translation to other domains without the need for more training. Results and Analysis. The suggested approach is demonstrated to perform better than the most advanced technologies on a number of benchmarks. Potential Use: The suggested architecture's potential applications in autonomous driving, medical imaging, and creative stylisation are examined in this paper.

Qualitative and quantitative indicators that support the conclusions include user surveys and measures such as FID, LPIPS, and SSIM. It may be able to close the gap between task-specific models and real-world use cases by improving generalisability across domains, perhaps creating new opportunities for useful applications. While showcasing the advancements and possible uses of combining GANs and VAEs across many domains, this study emphasises the necessity of a flexible framework in cross-domain picture translation.

5.CONCLUSION

The research provides significant theoretical and practical advancements, especially at the convergence of generative modeling and visual translation. This is an analysis of its importance:

The study introduces an innovative hybrid framework that integrates (VAEs) for reliable latent representation with (GANs) for superior image production. This approach enhances generative models by offering a more equitable resolution to enduring trade-offs, such as image quality against interpretability in latent space.

Multi-Modal Losses for Semantic Consistency: The training framework is enhanced through the integration of adversarial, perceptual, and cycle-consistency losses. This enhances the theoretical underpinning for the multi-objective optimization of deep generative models.

Technical and Practical Consequences Enhanced Domain Adaptability: The lack of task-specific fine-tuning augments the model's cross-domain generalization in applications such as digital restoration, exemplar-guided creation, and unsupervised translation.

The proposed hybrid methodology enhances image denoising and translation by mitigating the limitations of traditional GANs, which suffer from mode collapse and unstable training, as well as VAEs, which sometimes yield confusing outcomes. This improvement is particularly significant for applications such as medical imaging, document digitization, and geospatial analysis that utilize low-resolution or noisy images.

The assessment of the system across five datasets demonstrates its robustness and generalizability, essential attributes for implementing AI in practical applications.

High fidelity and clarity are crucial for life preservation in retinal scans, brain MRIs, and ultrasound imaging, with research emphasizing the practical advantages of enhanced diagnostic image quality in these procedures. This initiative possesses transdisciplinary and societal importance.

Urban Planning and Autonomous Systems: The model facilitates automated, high-precision image synthesis for geospatial applications, including satellite-to-map conversion and autonomous vehicle navigation.

REFERENCES

[1] A. Plumerault, H. L. Borgne, and C. Hudelot, "AVAE: Adversarial Variational Auto Encoder," 2022 26th International Conference on Pattern Recognition (ICPR), pp. 8687–8694, Jan. 2021, doi: 10.1109/icpr48806.2021.9412727.

[2] H. Hoyez, C. Schockaert, J. Rambach, B. Mirbach, and D. Stricker, "Unsupervised Image-to-Image translation: A review," Sensors, vol. 22, no. 21, p. 8540, Nov. 2022, doi: 10.3390/s22218540.

[3] J. Cho and A. Krzyzak, "Efficient integration of perceptual variational autoencoder into dynamic latent scale generative adversarial network," Expert Systems, vol. 41, no. 10, May 2024, doi: 10.1111/exsy.13618.

[4] X. Xiao, S. Ganguli, and V. Pandey, "VAE-Info-cGAN," Proceedings of the 13th ACM SIGSPATIAL International Workshop on Computational Transportation Science, pp. 1–10, Nov. 2020, doi: 10.1145/3423457.3429361.

[5] C. Wang et al., "GCT-VAE-GAN: An image enhancement network for Low-Light cattle farm scenes by integrating Fusion Gate Transformation Mechanism and variational Autoencoder GAN," IEEE Access, vol. 11, pp. 126650–126660, Jan. 2023, doi: 10.1109/access.2023.3328923.

[6] J. Li, "Twin-GAN -- Unpaired Cross-Domain Image Translation with Weight-Sharing GANs," arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1809.00946.

[7] S. Vivekananthan, "Comparative Analysis of Generative Models: Enhancing Image Synthesis with VAEs, GANs, and Stable Diffusion," arXiv (Cornell University), Aug. 2024, doi: 10.48550/arxiv.2408.08751.

[8] Tirel, L., Ali, A. M., & Hashim, H. A "Novel Hybrid Integrated Pix2Pix and WGAN Model with Gradient Penalty for Binary Images Denoising." https://arxiv.org/html/2407.11865v1.

[9] T. Ma, B. Li, W. Liu, M. Hua, J. Dong, and T. Tan, "CFFT-GAN: Cross-Domain Feature Fusion Transformer for Exemplar-Based Image Translation," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 2, pp. 1887–1895, Jun. 2023, doi: 10.1609/aaai.v37i2.25279.

[10] S. Yang, L. Jiang, Z. Liu, and C. C. Loy, "Unsupervised Image-to-Image Translation with Generative Prior," arXiv (Cornell University), Jan. 2022, doi: 10.48550/arxiv.2204.03641.

[11] J. Schmidhuber, “Annotated history of modern AI and deep learning,” arXiv (Cornell University), Jan. 2022, doi: 10.48550/arxiv.2212.11279.

[12] Z. K. Maseer, R. Yusof, N. Bahaman, S. A. Mostafa, and C. F. M. Foozy, “Benchmarking of machine learning for anomaly based intrusion detection Systems in the CICIDS2017 dataset,” IEEE Access, vol. 9, pp. 22351–22370, Jan. 2021, doi: 10.1109/access.2021.3056614.

[13] R. Kumar, R. Dabral, and G. Sivakumar, “Learning Unsupervised Cross-domain Image-to-Image Translation using a Shared Discriminator,” Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 256–264, Jan. 2021, doi: 10.5220/0010184102560264.

[14] A. Alotaibi, “Deep Generative Adversarial Networks for Image-to-Image translation: a review,” Symmetry, vol. 12, no. 10, p. 1705, Oct. 2020, doi: 10.3390/sym12101705.

[15] J. Chen and W. Song, “GAN-VAE: Elevate generative ineffective image through variational autoencoder,” 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), pp. 765–770, Aug. 2022, doi: 10.1109/prai55851.2022.9904067.

[16] S. Kim, J. Lee, K. Jeong, J. Lee, T. Hong, and J. An, “Automated door placement in architectural plans through combined deep-learning networks of ResNet-50 and Pix2Pix-GAN,” Expert Systems With Applications, vol. 244, p. 122932, Dec. 2023, doi: 10.1016/j.eswa.2023.122932.

[17] A. Gonzalez-Garcia, V. De Weijer Joost, and Y. Bengio, “Image-to-image translation for cross-domain disentanglement,” arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1805.09730.

[18] Y. Fu, S. Liang, D. Chen, and Z. Chen, “Translation of aerial image into digital map via discriminative segmentation and creative generation,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–15, Sep. 2021, doi: 10.1109/tgrs.2021.3110894.

[19] F. Zhang and C. Wang, “MSGAN: Generative Adversarial Networks for Image Seasonal Style Transfer,” IEEE Access, vol. 8, pp. 104830–104840, Jan. 2020, doi: 10.1109/access.2020.2999750.

[20] Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

[21] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, “STARGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation,” arXiv (Cornell University), Jan. 2017, doi: 10.48550/arxiv.1711.09020.

[22] M. Safayani, B. Mirzapour, H. Aghaebrahimian, N. Salehi, and H. Ravaee, “Unpaired Image-to-Image Translation with Content Preserving Perspective: A Review,” arXiv (Cornell University), Feb. 2025, doi: 10.48550/arxiv.2502.08667.

[23] R. Mehmood, R. Bashir, and K. J. Giri, “Deep Generative Models: a review,” Indian Journal of Science and Technology, vol. 16, no. 7, pp. 460–467, Feb. 2023, doi: 10.17485/ijst/v16i7.2296.

[24] X. Zhao, H. Yu, and H. Bian, “Image to image translation based on differential image Pix2Pix model,” Computers, Materials & Continua/Computers, Materials & Continua (Print), vol. 77, no. 1, pp. 181–198, Jan. 2023, doi: 10.32604/cmc.2023.041479.

[25] A. A. Et. Al, “CONDITIONAL GANS AS a SOLUTION TO IMAGE-TO-IMAGE RENDERING PROBLEMS,” INFORMATION TECHNOLOGY IN INDUSTRY, vol. 9, no. 2, pp. 1106–1111, Apr. 2021, doi: 10.17762/itii.v9i2.458.

[26] Y. Xu, J. Yang, and M. Sawan, “Multichannel synthetic preictal EEG signals to enhance the prediction of epileptic seizures,” IEEE Transactions on Biomedical Engineering, vol. 69, no. 11, pp. 3516–3525, May 2022, doi: 10.1109/tbme.2022.3171982.

[27] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, “Generative adversarial networks: introduction and outlook,” IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 4, pp. 588–598, Jan. 2017, doi: 10.1109/jas.2017.7510583.

[28] J. Chen and W. Song, “GAN-VAE: Elevate generative ineffective image through variational autoencoder,” 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), pp. 765–770, Aug. 2022, doi: 10.1109/prai55851.2022.9904067.

[29] J. Jenkins and K. Roy, “Exploring deep convolutional generative adversarial networks (DCGAN) in biometric systems: a survey study,” Discover Artificial Intelligence, vol. 4, no. 1, May 2024, doi: 10.1007/s44163-024-00138-z.

[30] A. Gautam, M. Sit, and I. Demir, “Realistic river image synthesis using deep generative adversarial networks,” Frontiers in Water, vol. 4, Feb. 2022, doi: 10.3389/frwa.2022.784441.

[31] Y. Shin, J. Yang, and Y. H. Lee, “Deep Generative Adversarial Networks: applications in musculoskeletal imaging,” Radiology Artificial Intelligence, vol. 3, no. 3, p. e200157, Mar. 2021, doi: 10.1148/ryai.2021200157.

[32] J. Yoo, H. Eom, and Y. S. Choi, “Image-To-Image translation using a Cross-Domain Auto-Encoder and decoder,” Applied Sciences, vol. 9, no. 22, p. 4780, Nov. 2019, doi: 10.3390/app9224780.

[33] H. Kazemi, S. Soleymani, F. Taherkhani, S. M. Iranmanesh, and N. M. Nasrabadi, “Unsupervised Image-to-Image translation using Domain-Specific variational information bound,” arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1811.11979.

[34] A. Almudévar, T. Mariotte, A. Ortega, and M. Tahon, “Unsupervised Multiple Domain Translation through Controlled Disentanglement in Variational Autoencoder,” arXiv (Cornell University), Jan. 2024, doi: 10.48550/arxiv.2401.09180.

[35] T. Wang and Y. Lin, “CycleGAN with Better Cycles,” arXiv (Cornell University), Aug. 2024, doi: 10.48550/arxiv.2408.15374.

[36] A. Sarkar, Z. Yang, and S. Cooper, “Controllable Level Blending between Games using Variational Autoencoders,” arXiv (Cornell University), Jan. 2020, doi: 10.48550/arxiv.2002.11869.

[37] T. Luhman and E. Luhman, “High fidelity image synthesis with deep VAEs in latent space,” arXiv (Cornell University), Jan. 2023, doi: 10.48550/arxiv.2303.13714.
[38] A. Asperti, D. Evangelista, and E. L. Piccolomini, “A Survey on Variational Autoencoders from a Green AI Perspective,” SN Computer Science, vol. 2, no. 4, May 2021, doi: 10.1007/s42979-021-00702-9.
[39] S. Molnár and L. Tamás, “Variational autoencoders for 3D data processing,” Artificial Intelligence Review, vol. 57, no. 2, Feb. 2024, doi: 10.1007/s10462-023-10687-x.
[40] S. D. Mbacke, F. Clerc, and P. Germain, “Statistical Guarantees for Variational Autoencoders using PAC-Bayesian Theory,” arXiv (Cornell University), Jan. 2023, doi: 10.48550/arxiv.2310.04935.
[41] R. Wei, C. Garcia, A. El-Sayed, V. Peterson, and A. Mahmood, “Variations in Variational Autoencoders - A Comparative evaluation,” IEEE Access, vol. 8, pp. 153651–153670, Jan. 2020, doi: 10.1109/access.2020.3018151.
[42] D. Cai, “Enhancing capabilities of generative models through VAE-GAN integration: A review,” Applied and Computational Engineering, vol. 67, no. 1, pp. 239–246, Jul. 2024, doi: 10.54254/2755-2721/67/2024ma0070.
[43] H. Tian, L. Zhang, S. Li, M. Yao, and G. Pan, “Pyramid-VAE-GAN: Transferring hierarchical latent variables for image inpainting,” Computational Visual Media, vol. 9, no. 4, pp. 827–841, Jul. 2023, doi: 10.1007/s41095-022-0331-3.
[44] M. Akbari and J. Liang, “Semi-Recurrent CNN-based VAE-GAN for sequential data generation,” arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1806.00509.
[45] S. M. Park, H. G. Yoon, D. B. Lee, J. W. Choi, H. Y. Kwon, and C. Won, “Topological magnetic structure generation using VAE-GAN hybrid model and discriminator-driven latent sampling,” Scientific Reports, vol. 13, no. 1, Nov. 2023, doi: 10.1038/s41598-023-47866-3.
[46] Z. Lin, J. Chen, X. Ma, C. Li, H. Zhang, and L. Zhao, “Unsupervised content and style learning for multimodal cross-domain image translation,” Scientific Reports, vol. 14, no. 1, Nov. 2024, doi: 10.1038/s41598-024-80508-w.
[47] S. Bond-Taylor, A. Leach, Y. Long, and C. G. Willcocks, “Deep Generative Modelling: a comparative review of VAEs, GANs, normalizing flows, Energy-Based and autoregressive models,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 7327–7347, Sep. 2021, doi: 10.1109/tpami.2021.3116668.
[48] W. Ahmad, H. Ali, Z. Shah, and S. Azmat, “A new generative adversarial network for medical images super resolution,” Scientific Reports, vol. 12, no. 1, Jun. 2022, doi: 10.1038/s41598-022-13658-4.
[49] P. Zhang, J. Bao, T. Zhang, D. Chen, and F. Wen, “Semi-Supervised Image-to-Image Translation using Latent Space Mapping,” arXiv (Cornell University), Jan. 2022, doi: 10.48550/arxiv.2203.15241.
[50] V. Ingale, R. Singh, and P. Patwal, “Image to Image Translation : Generating maps from satellite images,” arXiv (Cornell University), Jan. 2021, doi: 10.48550/arxiv.2105.09253.
[51] F. Makhmudkhujaev, J. Kwon, and I. K. Park, “Controllable image dataset construction using conditionally transformed inputs in generative adversarial networks,” IEEE Access, vol. 9, pp. 144699–144712, Jan. 2021, doi: 10.1109/access.2021.3122834.