Advanced Cross-Domain Image Synthesis Through Generative Adversarial Architectures

Aryan Kothari, V.R.N.S Nikhil, Namana Rohit,Dr. Suja P
Department of Computer Science and Engineering
Amrita School of Computing, Bangalore
Amrita Vishwa Vidyapeetham, India
Email: bl.en.u4aie22005, bl.en.u4aie22062, bl.en.u4aie22071,p_suja@blr.amrita.edu

***Abstract*—This paper explores the application of GAN-based architectures to cross-domain image synthesis, emphasizing their ability to transform images while preserving semantic details and visual coherence. Using a robust dataset designed for semantic segmentation and scene parsing, we demonstrate how GAN models like CycleGAN and StyleGAN can successfully learn unpaired mappings between image domains. Our approach aims to balance the generation of photorealistic images with the preservation of critical features, resulting in significant advancements in image quality and domain adaptation. Through our experiments, we highlight the effectiveness of these architectures in maintaining style consistency and structural fidelity across varied domains.**

***Index Terms*—Brain Diseases, DNA sequence, DNABert embeddings, deep learning, machine learning, bioinformatics**

I. Introduction

Cross-domain image synthesis has experienced remarkable progress, primarily due to the advancements in Generative Adversarial Networks (GANs). These architectures enable effective transformations between image domains, such as converting sketches into photorealistic images. A key challenge in this field is to maintain semantic preservation, style consistency, and visual coherence during the image translation process. Despite these complexities, GANs like CycleGAN, StarGAN, and StyleGAN have demonstrated success in learning mappings between unpaired data, offering efficient and flexible solutions for various image synthesis tasks.

The adversarial framework, consisting of generator and discriminator networks, facilitates continuous refinement, producing highly realistic images with enhanced detail. These advancements have expanded the scope of GAN-based image synthesis across domains and applications, showcasing their potential in diverse visual tasks. In this paper, we focus on leveraging GANs to improve cross-domain image synthesis, utilizing a dataset optimized for scene understanding, ensuring that the generated images not only transform effectively but also retain high-level structural integrity.

II. Literature Survey

The paper "Image-to-Image Translation: Methods and Applications" by Yingxue Pang et al[1]. provides a detailed overview of image-to-image translation (I2I), which focuses on transforming images from one domain to another while preserving their content. The authors highlight the significant advancements in I2I techniques, which have found applications in various fields such as image synthesis, segmentation, style transfer, and restoration. They categorize I2I methods into two main types: two-domain and multi-domain tasks, discussing key generative models like Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) that underpin these methods. The paper also addresses the challenges faced in the field and outlines future research directions to enhance the effectiveness and quality of I2I applications.

The paper "AttentionGAN: Unpaired Image-to-Image Translation Using Attention-Guided Generative Adversarial Networks"[2] introduces a novel framework for unpaired image-to-image translation that enhances image quality by utilizing attention mechanisms. The proposed AttentionGAN effectively identifies and focuses on the most discriminative foreground objects while preserving the background, addressing the limitations of existing models that often produce visual artifacts and fail to maintain high-level semantics. The architecture includes attention-guided generators that create attention masks to guide the translation process and a discriminator that evaluates only attended regions of the image. Extensive experiments across eight public datasets demonstrate that AttentionGAN generates sharper and more realistic images compared to state-of-the-art methods like CycleGAN and GANimorph, achieving new benchmarks in image quality and detail preservation..

The paper "Unified Generative Adversarial Networks for Controllable Image-to-Image Translation"[3] introduces a novel GAN framework designed for controllable image-to-image translation, allowing the transformation of images from a source to a target domain based on various structural inputs like class labels, object keypoints, and semantic maps. This unified model consists of a single generator and discriminator that utilize both a reference image for appearance information and a controllable structure for structural guidance, thereby enhancing the quality of generated images. The authors propose three innovative loss functions—color loss, controllable structure guided cycle-consistency loss, and self-content preserving loss—to improve the learning process and output fidelity. Additionally, they introduce the Fréchet ResNet Distance (FRD) as a new evaluation metric for assessing the quality of generated images.

The paper introduces a new way to translate images from one

style or domain to another using a pre-trained StyleGAN2 model[4], making the process more efficient than older methods. Instead of building new models for each domain, they modify an existing model to work with new image styles. They also use a method to convert an image into a code that the model can understand, allowing for easier translation between different styles. This approach improves the quality and variety of the translated images while using fewer resources and less time than traditional methods.

The paper "Image Translation based on Attention Residual GAN" proposes an improved GAN architecture called Attention Residual GAN for image-to-image translation tasks[5]. It builds on previous GAN work like DCGAN, which uses convolutional neural networks, and CycleGAN for unpaired image translation. The key innovation is adding an attention-based residual neural network (ResNet) to the generator of the GAN. Specifically, it incorporates ResNet-50 with channel attention mechanisms added to each residual block. This allows the generator to weight different channels of the feature maps, emphasizing important information and suppressing irrelevant details. The authors claim this improves the generator's ability to capture semantic information and reduce artifacts in the translated images. They evaluate their approach on the Facades dataset of building images, using the FID metric to assess image quality. The results show the Attention Residual GAN can produce high quality translated images that preserve details well compared to the ground truth.

The paper "Image Translator : An Unsupervised Image –to-Image Translation Approach Using GAN" presents a literature survey on recent approaches to unsupervised image-to-image translation using generative adversarial networks (GANs)[6]. The authors review several key works in this area, including CycleGAN by Zhu et al., which introduced cycle-consistency loss to enable unpaired training. They discuss improvements like the multi-model mapping approach of Xiong et al., the use of multiple cooperating generators by Nizan et al., and the weighted hybrid discriminator of Lira et al. The survey also covers techniques to enhance image quality and stability, such as the attention mechanism proposed by Emami et al. The authors note that current methods still face challenges with geometric transformations, mode collapse, and preserving fine details. Overall, the literature review provides context on recent advances in GAN-based unsupervised image translation, highlighting both progress and remaining limitations in the field.

The paper "Study on Image-to-image Translation Based on Generative Adversarial Networks" provides a comprehensive overview of the advancements in image-to-image translation techniques utilizing Generative Adversarial Networks (GANs). It highlights key methodologies such as CycleGAN, which enables unpaired image translation through cycle consistency loss, and StarGAN, which facilitates multi-domain transformations using a single model by integrating domain labels. The survey emphasizes the rapid evolution of these technologies and their potential applications in various fields, showcasing the innovative approaches that enhance the capabilities of GANs in generating realistic images across different styles and categories.[7]

The paper "Image to Image Translation in Video Call Using Based StyleGAN Encoder" focuses on various neural network approaches for image translation, particularly in the context of video conferencing. It highlights the effectiveness of Convolutional Neural Networks (CNNs) for background subtraction and real-time processing, as demonstrated by Bouwmans et al. and Babaee et al., who improved segmentation methods through learned parameters and spatial filtering techniques. The survey also discusses Generative Adversarial Networks (GANs), emphasizing their dual-model structure that enables the generation of realistic images while minimizing loss functions across diverse tasks[8], as proposed by Goodfellow et al. Additionally, it examines advancements such as TimeConvNets for capturing facial expressions in real-time and local blur mapping algorithms that enhance image quality. Overall, the survey underscores the transition from traditional methods to more sophisticated neural architectures that facilitate efficient image-to-image translation with reduced latency and improved accuracy in video outputs.

A literature review on paper "Image-to-Image Translation with Conditional Adversarial Networks" highlights their role as a versatile solution for image-to-image translation tasks, effectively learning mappings from input to output images without the need for application-specific algorithms. This approach builds upon the foundational work of Generative Adversarial Networks (GANs), which adaptively learn loss functions that enhance image quality by minimizing perceptual discrepancies rather than relying solely on traditional pixel-wise losses, which often result in blurry outputs. The review discusses various applications of cGANs, including tasks such as photo synthesis from label maps and edge detection, showcasing their ability to generalize across different domains while maintaining high fidelity in generated images. Furthermore, it emphasizes the significance of architectural innovations, such as the U-Net generator and PatchGAN discriminator, which facilitate effective training and improve the quality of outputs by focusing on local structures within images. Overall, the literature underscores the transformative impact of cGANs in computer vision and graphics, paving the way for more sophisticated methodologies in image processing tasks. [9]

The paper by [10] explores the use of generative adversarial networks (GANs) to generate synthetic facial data for children from different ethnicities, focusing on Caucasian and Asian races. The authors compare three image-to-image (I2I) translation methods: pix2pix, CycleGAN, and CUT, to assess their effectiveness in transforming child facial data between these ethnic groups. Using a synthetic dataset of 2,400 Asian children and 2,400 Caucasian children, the study evaluates the models using metrics like FID, PSNR, and Structural Similarity Indexv. The results indicate that pix2pix achieved

the best PSNR (24.41) and SSIM (0.75) scores, suggesting high image quality and similarity to real data, while CUT had the best FID score (26.36), indicating better alignment with the original image distribution. The study emphasizes the importance of diversifying child facial datasets to improve the robustness of face recognition systems and reduce racial bias.StyleGAN2 model, making the process more efficient than older methods. Instead of building new models for each domain, they modify an existing model to work with new image styles. They also use a method to convert an image into a code that the model can understand, allowing for easier translation between different styles.

introduces a method to improve the efficiency of training generative adversarial networks (GANs) for real-time image editing, especially on mobile devices. The authors propose a novel framework, E2GAN, which addresses the high computational cost of using diffusion models for image editing by distilling knowledge from these models into lightweight GANs. They achieve this through three innovations: 1) a base GAN model trained on generalized concepts, which can be fine-tuned for specific tasks, reducing the need for retraining, 2) Low-Rank Adaptation (LoRA) applied only to critical layers to further reduce training time and storage, and 3) a data efficiency strategy that minimizes the training data needed. Experimental results show that E2GAN requires only 1% of trainable parameters, saving up to 33 times the training time and achieving better or comparable performance in terms of FID scores. The method enables real-time image editing (30 FPS) on mobile devices with latency as low as 15.5 ms, making it significantly faster than alternatives like pix2pix and diffusion models
The paper proposes a novel approach to improve Image-to-Image (I2I) translation by introducing a communication channel between the discriminator and generator in a Generative Adversarial Network (GAN). It formulates the GAN as a Partially-observed Markov Decision Process (POMDP) for the generator, addressing the information insufficiency that arises during training. The communication channel, represented by a continuous message vector, helps improve the quality of image translation by guiding the generator with additional information from the discriminator. Experiments on various benchmark datasets such as apple2orange and horse2zebra demonstrate that the proposed model improves both qualitative and quantitative performance, outperforming several baseline models like CycleGAN, MUNIT, and DRIT. The paper reports Kernel Inception Distance (KID) improvements, such as reducing KID from 13.86 to 10.52 for apple2orange translation and achieving notable texture and background consistency in generated image[12]

investigates transferring knowledge from pre-trained GANs to new domains, similar to how pre-trained discriminative models are used in transfer learning. The authors conduct experiments using WGAN-GP architecture on various source (ImageNet, Places, LSUN Bedrooms, CelebA) and target (Flowers, Kitchens, LFW, Cityscapes) datasets. They find that transferring both the generator and discriminator yields the best results, with FID scores improving from 32.87 to 24.35 when using ImageNet as source and LSUN Bedrooms as target. Pre-trained GANs converge faster and perform better, especially with limited target data - requiring 2-5 times fewer images to achieve similar FID scores. Interestingly, dense datasets like LSUN Bedrooms often outperform diverse datasets like ImageNet as source models. The authors also demonstrate successful transfer from unconditional to conditional GANs, with pre-trained models showing higher accuracy (e.g., 65.9% vs 52.7% for 1K images/class) and lower FID scores (77.5 vs 117.3) compared to training from scratch. Human evaluation confirms the superiority of images generated by pre-trained models, with 67% preference for 1K images/class scenario.

III. DataSet Details

The ADE20K dataset is a comprehensive resource designed for semantic segmentation and scene parsing, featuring a diverse collection of high-resolution images that span various indoor and outdoor environments. Each image is meticulously annotated with detailed pixel-wise labels that identify a wide range of object categories and stuff categories, enabling models to recognize and delineate multiple elements within complex scenes, such as vehicles, people, buildings, and natural landscapes. This dataset is particularly valuable for training and evaluating algorithms in tasks like semantic segmentation, where the goal is to classify each pixel according to its corresponding category, and scene understanding, which involves analyzing the layout and relationships of objects within a scene. Additionally, ADE20K supports applications in robotics and autonomous driving, where accurate perception of the environment is crucial for navigation and decision-making. The rich annotations and diverse scenes make ADE20K an essential tool for advancing research in computer vision, providing a robust foundation for developing models that can interpret and generate visual data in real-world scenarios. For more information, you can access the dataset at ADE20K Dataset.

IV. FUTURE ENHANCEMENTS AND CONCLUSION

The integration of the U-Net generator and the PatchGAN discriminator with added Multi-Head Attention produces the desired outcome in image-to-image translation. Skip connections are employed to maintain spatial information during feature extraction, and the discriminator dominates local patches to gauge realism although feature maps are enhanced through attentions. This adversarial setup nicely splits the task of remembering fine detail against gross structural integration and allows the generation of high quality and semantically relevant images. There are directions for improvements in the future; better attention mechanisms, like transformers can be incorporated to capture more global dependencies. Moreover, the prospective for applying the architecture for deploying

on edge devices or real-time applications using model compression methods also remains high. Extending to domains such as medical imaging, satellite, cross-modality (text to image) data, would generalise the use of the model; As such, self-supervised learning would minimize the dependence on labelled data.

REFERENCES

[1] Y. Pang, J. Lin, T. Qin and Z. Chen, "Image-to-Image Translation: Methods and Applications," in IEEE Transactions on Multimedia, vol. 24, pp. 3859-3881, 2022, doi: 10.1109/TMM.2021.3109419.

[2] H. Tang, H. Liu, D. Xu, P. H. S. Torr and N. Sebe, "AttentionGAN: Unpaired Image-to-Image Translation Using Attention-Guided Generative Adversarial Networks," in IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 4, pp. 1972-1987, April 2023, doi: 10.1109/TNNLS.2021.3105725.

[3] H. Tang, H. Liu and N. Sebe, "Unified Generative Adversarial Networks for Controllable Image-to-Image Translation," in IEEE Transactions on Image Processing, vol. 29, pp. 8916-8929, 2020, doi: 10.1109/TIP.2020.3021789.

[4] J. Huang, J. Liao and S. Kwong, "Unsupervised Image-to-Image Translation via Pre-Trained StyleGAN2 Network," in IEEE Transactions on Multimedia, vol. 24, pp. 1435-1448, 2022, doi: 10.1109/TMM.2021.3065230.

[5] M. Zhang, L. He and X. Wang, "Image Translation based on Attention Residual GAN," 2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), Hangzhou, China, 2021, pp. 802-805, doi: 10.1109/ICAICE54393.2021.00156.

[6] Plabon, Silvia & Khan, Mohammad Shabaj & Khaliluzzaman, Md & Islam, Md. (2022). Image Translator: An Unsupervised Image-to-Image Translation Approach using GAN. 10.1109/ICISET54810.2022.9775902.

[7] X. Li, "Study on Image-to-image Translation Based on Generative Adversarial Networks," 2022 2nd International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR), Xi'an, China, 2022, pp. 92-97, doi: 10.1109/ICBAR58199.2022.00025.

[8] S. R. Unni and S. S. L R, "Image To Image Translation in Video Call Using Residual Based StyleGAN Encoder," 2021 Fourth International Conference on Microelectronics, Signals & Systems (ICMSS), Kollam, India, 2021, pp. 1-5, doi: 10.1109/ICMSS53060.2021.9673655.

[9] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[10] https://doi.org/10.48550/arXiv.2308.04232

[11] https://doi.org/10.48550/arXiv.2401.06127

[12] arXiv:2303.03598 [cs.CV]

[13] arXiv:1805.01677 [cs.CV]