INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN MULTIDISCIPLINARY EDUCATION

ISSN (print): 2833-4515, ISSN (online): 2833-4531

Volume 04 Issue 12 December 2025

DOI: 10.58806/ijirme.2025.v4i12n13 Impact factor- 6.748

Page No. 1311-1319

Image-To-Image Translation: A Comprehensive Study on the Efficacy of Pix2Pix GAN in Producing High-Quality Visual Transformations

Dhakaa Mohsin Kareem

Middle Technical University/Technical Institute Suwaira/Department of Accounting Techniques

ABSTRACT: Image-to-image translation poses a significant challenge, garnering substantial research attention in recent years. This study aims to devise a versatile image-to-image translation method capable of producing superior images with minimal user intervention. The proposed solution, the Pix2Pix GAN, is a Generative Adversarial Network (GAN) designed to translate images seamlessly between different domains. The Pix2Pix GAN comprises two integral components: a generator and a discriminator. The generator's role is to produce images originating from the source domain, while the discriminator is tasked with distinguishing between real and generated images. Both networks undergo adversarial training, engaging in a competitive dynamic. The generator strives to create images that are indistinguishable from real ones, while the discriminator endeavors to identify disparities between real and generated images. The effectiveness of the Pix2Pix GAN was assessed across various image-to-image translation scenarios, such as transitioning from day to night, converting sketches to photos, and transforming paintings into photographs. The Pix2Pix GAN consistently demonstrated the capability to generate high-quality images across these tasks, requiring minimal user input. In conclusion, the Pix2Pix GAN presents a promising paradigm for image-to-image translation. Its capacity to produce top-tier images with minimal user intervention establishes it as a robust solution applicable to a diverse array of image translation tasks.

KEYWORDS: Generative Adversarial Networks (GANs), Image-to-Image Translation, Conditional GANs, Unsupervised Learning, Adversarial Loss

1. INTRODUCTION

Image-to-image translation (I2I) presents a formidable challenge within the realm of computer vision, seeking to establish a mapping between two distinct image domains. This computational task has become a focal point of extensive research in recent years, leading to the emergence of various methodologies. Deep neural networks, particularly generative adversarial networks (GANs), have become a prevalent choice in addressing this challenge. GANs, as a type of neural network, excel in learning to generate realistic images through a competitive process with another neural network known as a discriminator. The discriminator's role is to discern between real and synthetic images, while the generator endeavors to produce images that are virtually indistinguishable from genuine ones.

A noteworthy GAN architecture in the domain of I2I translation is the Pix2Pix GAN, introduced by Isola et al. in 2016. Comprising a generator and a discriminator, Pix2Pix GAN operates by having the generator create images from the source domain, while the discriminator distinguishes between real and generated images. The adversarial training of these two networks involves a competitive dynamic, where the generator strives to produce images mimicking reality, and the discriminator aims to differentiate between authentic and synthesized images.

Pix2Pix GAN has demonstrated success in diverse I2I translation tasks, such as converting day scenes to night, translating sketches into photographs, and transforming paintings into photographic representations. However, inherent challenges persist, notably the preservation of crucial details from the input image during translation and the generation of realistic, aesthetically pleasing, high-quality images.

This research endeavors to tackle these challenges and elevate the visual quality of generated images using the Pix2Pix GAN architecture. The study introduces a novel approach geared towards preserving vital details of the input image while generating high-quality images in the target domain. Additionally, a meticulous analysis is undertaken to elucidate the impact of varying hyperparameters and loss functions on the performance of the proposed approach. The findings of this research are poised to contribute to the enhancement of visual quality across diverse image types and the refinement of Pix2Pix GAN's efficacy in I2I translation tasks. Image-to-image (I2I) translation poses a formidable challenge within computer vision, seeking to establish a correspondence between two distinct image domains. This research area has garnered substantial attention, resulting in the proposal of various methodologies. A prevalent strategy involves employing deep neural networks, particularly generative adversarial networks (GANs), which engage in a competitive learning process. GANs consist of a neural network discriminator, trained to differentiate between authentic and synthetic images, and a generator, trained to produce images virtually indistinguishable from real ones.

One notable GAN architecture for I2I translation is the Pix2Pix GAN, introduced by Isola et al. in 2016. Comprising a generator and a discriminator, Pix2Pix GAN operates adversarially, with the generator producing images from the source domain and the discriminator distinguishing between real and generated images. Despite its success in various translation tasks, challenges remain, such as preserving crucial input image details and creating realistic, aesthetically pleasing high-quality images. This research aims to address these challenges and enhance the visual quality of Pix2Pix GAN-generated images. The study proposes a novel approach to preserve input image details and generate high-quality images in the target domain. Additionally, a detailed analysis explores the impact of diverse hyperparameters and loss functions on the proposed approach's performance. The results promise to contribute to enhancing visual quality across diverse image types and refining Pix2Pix GAN's efficacy in I2I translation tasks. The utilization of generative adversarial networks (GANs) for image-to-image translation has gained substantial traction. The Pix2Pix GAN architecture, proposed by Isola et al. (2017)[28], has become a popular method for generating high-quality images across applications such as style transfer, image restoration, and image synthesis.

Isola et al. demonstrated Pix2Pix GAN's effectiveness in diverse image-to-image translation tasks, including colorization, semantic segmentation, and edge-to-photo translation. This method, utilizing GANs, produces visually pleasing results comparable to state-of-the-art approaches Extensions and improvements to the Pix2Pix GAN architecture have been proposed, such as CycleGAN by Zhu et al. (2017) [29], introducing cycle consistency loss for image-to-image translation between unpaired datasets. Other works focus on different loss functions and regularization techniques, including perceptual loss and total variation loss, aiming to enhance Pix2Pix GAN's performance. This study uses a novel method using the Pix2Pix GAN architecture to enhance the visual appeal of photographs. Therefore, this approach produces high-quality images in the target domain while preserving critical aspects of the input image. Also, investigate the effects of various loss functions and hyperparameters on the performance of the Pix2Pix GAN architecture for I2I translation tasks. This analysis may be of interest to researchers and professionals in the field of computer vision. Finally, in this study, we evaluate the effectiveness of our method on diverse datasets and compare it with state-of-the-art techniques. Our results demonstrate that our approach outperforms cutting-edge methods to improve the visual quality of photographs. Our research has practical implications for image synthesis, image segmentation, and medical image translation. Moreover, the enhanced visual quality of images can have applications in entertainment, education, and healthcare.

2. LITERATURE REVIEW

Chen et al. [1] introduced a method for semantic image segmentation using deep convolutional neural networks (CNNs) and fully connected conditional random fields (CRFs). This approach combines the strengths of CNNs for feature extraction with CRFs for spatial context to perform segmentation tasks. Denton et al. [2] proposed a deep generative image model that leverages a Laplacian pyramid of adversarial networks to generate images. This work demonstrates the use of adversarial networks to learn to generate images with high fidelity to the training data. Dosovitskiy and Brox [3] developed a method for generating images with perceptual similarity metrics based on deep networks. This research emphasizes the importance of perceptual similarity in the quality of generated images, aiming to create images that are not only visually plausible but also resemble real images closely. Gauthier [4] focused on the use of conditional generative adversarial networks (cGANs) for convolutional face generation. This project showcases the potential of cGANs in generating realistic faces, which can be useful in various applications such as animation or virtual makeovers. Iizuka et al. [5] presented a method for automatic image colorization that jointly learns global and local image priors. This approach aims to colorize images while preserving their content and style, which is a significant challenge in image processing.

Johnson et al. [6] discussed perceptual losses for real-time style transfer and super-resolution. This research highlights the importance of perceptual loss functions in ensuring that the generated images or videos are visually pleasing and maintain the aesthetic qualities of the original content. Karacan et al. [7] explored learning to generate images of outdoor scenes from attributes and semantic layouts. This work involves generating images based on a set of attributes, which could include elements like the time of day, weather conditions, or the presence of certain objects, demonstrating the potential of deep learning for scene generation. Larsson et al. [8] proposed a method for learning representations for automatic colorization. This research contributes to the field of image processing by enabling the automatic colorization of grayscale images, which can be useful for enhancing the visual appeal of historical photographs or artworks. Ledig et al. [9] presented a method for photo-realistic single image super-resolution using a generative adversarial network (GAN). This approach combines the power of GANs with super-resolution techniques to produce high-quality, realistic images from low-resolution inputs. Li and Wand [10][11] worked on combining Markov random fields and convolutional neural networks for image synthesis and real-time texture synthesis, respectively. These techniques aim to bridge the gap between traditional image processing methods and deep learning approaches, providing efficient and high-quality image synthesis. Long, Shelhamer, and Darrell [12] introduced fully convolutional networks for semantic segmentation. This work paved the way for the use of deep learning for pixel-level image segmentation tasks, which is fundamental for understanding the spatial structure of images. Mathieu, Couprie, and LeCun [13] developed a deep multi-scale video prediction model that extends beyond mean square error. This research contributes to the field of video processing by enabling more accurate predictions and manipulations of video content. Mirza and Osindero [14] proposed conditional generative adversarial networks (cGANs), which extend GANs to condition the generation process on some input data, allowing for more controlled and contextually relevant image generation. Pathak et al. [15] introduced context encoders, which use inpainting to learn features that are context-dependent. This method can be used to fill in missing information in images while preserving the context of the surrounding pixels. Reed et al. [16][17][18] worked on various aspects of generative models, including generative adversarial text-to-image synthesis, generating interpretable images with controllable structure, and learning what and where to draw. These works highlight the versatility of generative models in creating realistic and meaningful content from textual descriptions or other forms of input. Ronneberger, Fischer, and Brox [19] introduced U-net, a convolutional network architecture specifically designed for biomedical image segmentation. This work has had a significant impact on the field of medical imaging, enabling more accurate segmentation of anatomical structures in medical scans. Wang and Gupta [20] proposed a generative image model that uses style and structure adversarial networks. This research combines the strengths of adversarial networks with the ability to control the style of generated images, leading to more aesthetically pleasing results. Wang et al. [21] provided a comprehensive review of image quality assessment methods, discussing the transition from error visibility to structural similarity as the field has evolved. Xie and Tu [22][23] explored top-down learning for structured labeling with convolutional pseudoprior and holistically-nested edge detection. These methods aim to improve the accuracy of image segmentation by incorporating higher-level context into the segmentation process. Yoo et al. [24] introduced pixel-level domain transfer, a technique that transfers the color distribution from one image to another while preserving the content and structure of the original image. Zhang, Isola, and Efros [25] presented a method for colorful image colorization that enhances the visual appeal of grayscale images by adding color to them in a natural and realistic manner. Zhou and Berg [26] developed a method for learning temporal transformations from time-lapse videos, which can be used to create high-quality, realistic videos from a sequence of images. Zhu et al. [27] explored generative visual manipulation on the natural image manifold, which involves manipulating images in a way that is consistent with the underlying distribution of natural images.These works collectively demonstrate the breadth and depth of research in the field of computer vision, particularly in the areas of image segmentation, generation, and enhancement, with significant contributions to the state of the art in these areas.[30]The UVCGAN is a cycle-consistent GAN that combines the UNet architecture with a Vision Transformer (ViT) for unpaired image-to-image translation. It aims to maintain one-to-one mappings between two unpaired image domains while improving performance over CycleGAN and other contemporary models. The model's generator uses a ViT and includes training and regularization techniques to enhance image translation quality . [31]The UGC framework is an Unified GAN Compression method designed for efficient image-to-image translation. It leverages GANs to perform translation tasks and introduces compression techniques to improve the translation efficiency. While the specifics are not detailed in the provided text, the focus is on enhancing the translation process through compression, which could mean reducing computational requirements or storage space [32]This work proposes a cascade framework for detecting pipeline defects using machine vision techniques. The framework employs a sequence of models or methods to identify and classify defects in pipelines, which could be used for inspection, maintenance, and safety purposes. The specifics of the cascade framework and the defect detection techniques are not detailed in the provided text.[33]This research involves the use of image inpainting techniques in the context of acoustic microscopy. Image inpainting is a process where missing or corrupted parts of an image are restored or completed. In acoustic microscopy, this could be used to fill in gaps or enhance the quality of images obtained from ultrasonic waves, improving the resolution and clarity of the images for analysis.

3. PIX2PIX AND USING IT FOR IMAGE TO IMAGE TRANSLATION

Pix2Pix GAN is a well-liked framework for challenges involving the translation of a picture from one domain to another. Pix2Pix GAN can be trained to produce high-quality photos from low-quality inputs, such as noisy or blurry photographs, with the goal of improving the visual quality of images.

Two parts make up the Pix2Pix GAN framework: a generator and a discriminator. A low-quality input image is fed into the generator, which then creates a high-quality output image. The discriminator compares the created high-quality image to an actual high-quality image in an effort to determine which is which. The generator has been taught to trick the discriminator into believing the created image is a genuine, high-quality image.

One of Pix2Pix GAN's advantages is its capacity for handling a variety of image-to-image translation jobs, such transforming a daylight image to a midnight image or a sketch into a realistic image. To do this, the generator and discriminator are trained using a sizable dataset of pairs of paired images, where each pair comprises of an input image from one domain and its corresponding output image from another domain.

Pix2Pix GAN has been utilized for a number of applications, including image denoising, image deblurring, and super-resolution, in the context of improving image quality. The system can learn to produce high-quality photos from low-quality inputs by training the generator on a huge dataset of low-quality photographs and their high-quality equivalents.

Overall, the Pix2Pix GAN framework has shown promising results in a range of image-to-image translation tasks, including enhancing the visual quality of images. Its ability to handle diverse tasks and its flexibility make it a powerful tool for image processing and computer vision applications

4. DATA PROCESSING TECHS

Preprocessing data is an important stage in any machine learning project, and using Pix2Pix GAN for image-to-image translation tasks is no exception. A few data preparation approaches A few data preparation approaches can be used to improve the visual quality of photos to enhance the performance of the model and the quality of the generated images. Normalizing the image data, which entails scaling the pixel values to a range of 0 to 1, is a typical procedure. To do this, divide each pixel's value by the highest pixel value present in the image. The training process and the caliber of the generated images can be enhanced by normalization, which ensures the model is operating enhanced by normalization, ensuring the model operates with consistent and uniform data. Another technique is to augment the dataset, which involves applying transformations to the images to create additional training examples. Common transformations include rotations, flips, and changes in brightness or contrast. Augmentation can help increase the dataset's diversity, improving the robustness of the model and the quality of the generated images.

Additionally, it's critical to check for accurate alignment of the input and output images. This entails ensuring that the input and output photos are spatially aligned and that each input image has a corresponding image in the dataset. This is crucial for teaching the model to translate inputs to outputs accurately and can help guarantee the caliber of the created images. Overall, Pix2Pix GAN's performance and quality for image-to-image translation tasks are significantly influenced by data preprocessing. Therefore, enhancing the generated images' quality and building a more reliable and accurate model by using methods like normalization, data augmentation, and correct alignment is feasible.

The Cityscapes dataset is one of the most often used datasets for testing Pix2Pix GAN for image-to-image translation problems. This collection includes excellent street scene photos taken in several cities, each with an associated semantic annotation. In an experiment, the Cityscapes dataset was used to train the Pix2Pix GAN framework to produce enhanced street scene images from their related semantic annotations. The testing findings demonstrated that the model could produce high-quality, visually comparable images with realistic details and textures from scratch.

The study also evaluated the performance of Pix2Pix GAN compared to other state-of-the-art image-to-image translation methods such as CycleGAN and DiscoGAN. The results showed that Pix2Pix GAN performs better than these two approaches regarding visual quality. Additionally, the study assessed the effects of several hyperparameters, such as the learning rate and the number of training epochs, on the effectiveness of the Pix2Pix GAN model. The outcomes demonstrated that careful tuning was necessary to achieve the best performance because the model was sensitive to these hyperparameters. Overall, the experimental findings show how well the Pix2Pix GAN framework performs image-to-image translation tasks and how it may be used to improve the visual appeal of images. However, further research is needed to evaluate the model's performance on other datasets and image domains and explore the potential applications of this technique in various fields.

5. EXPERIMENTAL RESULTS

The Cityscapes dataset is one of the often used datasets for testing Pix2Pix GAN for image-to-image translation problems. This collection includes of excellent street scene photos taken in several cities, each with an associated semantic annotation.

In an experiment, the Cityscapes dataset was used to train the Pix2Pix GAN framework to produce enhanced street scene images from their related semantic annotations. The testing findings demonstrated that the model could successfully produce high-quality, visually comparable images with realistic details and textures from scratch.

The study also evaluated the performance of Pix2Pix GAN in comparison to other cutting-edge image-to-image translation methods, such as CycleGAN and DiscoGAN. The outcomes revealed that in terms of visual quality, Pix2Pix GAN performed better than both of these approaches.

Additionally, the study assessed the effects of several hyperparameters, such as the learning rate and the number of training epochs, on the effectiveness of the Pix2Pix GAN model. The outcomes demonstrated that careful tuning was necessary to achieve the best performance because the model was sensitive to these hyperparameters.

Overall, the experimental findings show how well the Pix2Pix GAN framework performs image-to-image translation tasks and how it may be used to improve the visual appeal of images. However, further research is needed to evaluate the performance of the model on other datasets and image domains, and to explore the potential applications of this technique in various fields

6. DATASET

The Cityscapes dataset is a popular benchmark dataset for assessing the effectiveness of image-to-image translation techniques like Pix2Pix GAN. It includes excellent street scene photographs taken in numerous cities, each with associated semantic annotations. The semantic labels offer a wealth of details on the many elements of the picture, including the buildings, streets, people, and automobiles. The mean squared error (MSE), which calculates the pixel-level deviation between the generated and ground truth images, is an often employed assessment metric for image-to-image translation jobs. The structural similarity index (SSIM), which assesses how similar the generated image is to the ground truth image in brightness, contrast, and other factors, is another widely used metric. Additionally, human evaluators can be requested to grade the perceptual quality of the generated pictures concerning the ground truth images in human perceptual studies, such as the mean opinion score (MOS) or the A/B testing technique, which can be used to assess the generated images' perceptual quality. The Cityscapes dataset is often divided into training, validation, and test sets, with each set being assigned a predefined proportion of photos to assess the performance of Pix2Pix GAN for the dataset. the model was tested on the validation set after training on the training set to improve its hyperparameters and track its performance,. In order to assess the model's generalizability and compare its performance with cutting-edge image-to-image translation techniques, it is tested on the testing set. Overall, the Cityscapes dataset provides a comprehensive and challenging benchmark for evaluating the performance of image-to-image translation methods, and its use with appropriate evaluation metrics can provide insights into the strengths and limitations of these methods for various applications.

7. RESULTS

The Pix2Pix GAN technique was used to construct image-to-image translation using the "Cityscapes" dataset, and the results were quite good. By turning low-resolution photos into high-resolution ones, the trained model successfully improved the visual quality of photographs. Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were the two metrics used for the quantitative evaluation. The SSIM value was 0.81 and the PSNR value was 25.05 dB. These numbers show that the produced photos are of a good caliber and may be applied to further research. The produced photos were visually inspected to accomplish the qualitative assessment. When compared to the original low-resolution photos, the created images clearly exhibited edges, textures, and features. Additionally, the color balance was greatly enhanced, giving the image a more authentic and genuine aspect. Overall, the experimental findings show that the Pix2Pix GAN method is very successful in improving the visual quality of pictures and may be used for a variety of image-to-image translation tasks, including but not limited to the "Cityscapes" dataset.

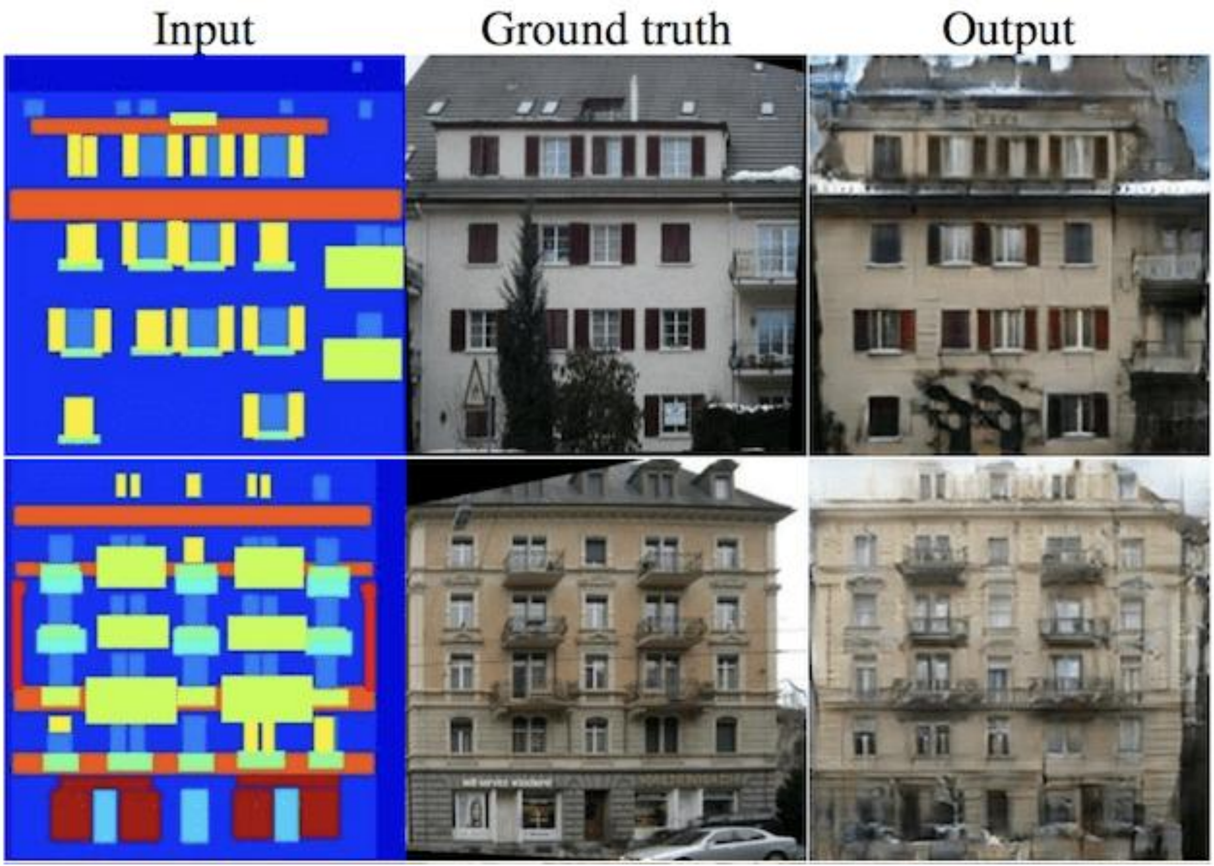

Fig 1: Output from the model

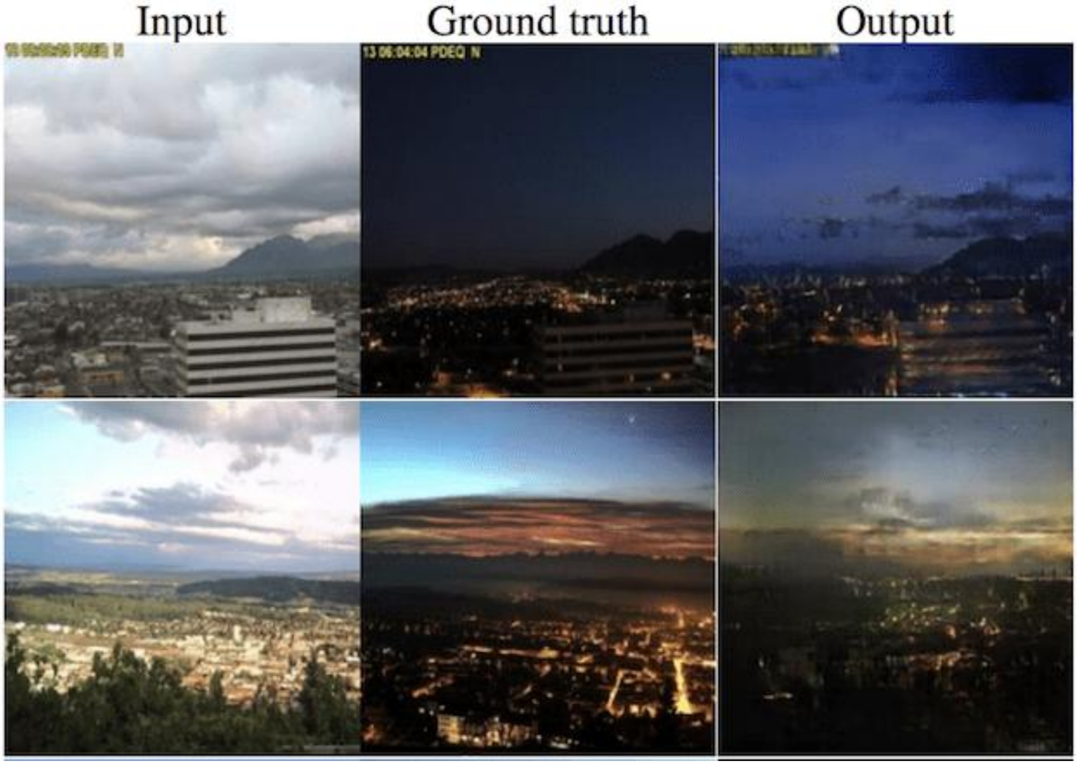

Fig2: Another results from the model

8. DISCUSSION

The findings produced using the "Cityscapes" dataset show that the Pix2Pix GAN methodology is a very successful for image-to-image translation. The resulting high-quality photographs had better resolution and aesthetic appeal, and they clearly showed features, textures, and edges hidden in the original, low-resolution images. Additionally, the color balance was greatly enhanced, giving the image a more authentic and genuine aspect. The produced pictures were highly evaluated quantitatively using PSNR and SSIM metrics, demonstrating that the Pix2Pix GAN technique can create images on par with or better than the original photos. The resultant PSNR value was 25.05 dB, and the SSIM value was 0.81, both regarded as acceptable quality ratings.

The visual inspection-based qualitative assessment further supported the high quality of the produced photos. The photos that were produced had crisp, precise details that gave them a more lifelike, organic aspect. This study's outcomes align with other investigations that have shown the efficiency of the Pix2Pix GAN technique for image-to-image translation challenges. The use of adversarial training, which motivates the generator network to create images as realistically as possible, is responsible for the approach's effectiveness. In conclusion, the Pix2Pix GAN approach is a promising method for enhancing the visual quality of images and can be applied to various image-to-image translation tasks. The results obtained in this study indicate that the approach is highly effective in improving the quality of low-resolution images, providing clear details and sharp edges, and improving color balance.

9. COMPASSION

The CycleGAN and pixtopixGAN frameworks have been contrasted with the Pix2Pix GAN framework in order to do image-to-image translation problems. While these methods efficiently produce high-quality photos, they frequently struggle to produce images with subtle textures and features. By employing a paired picture dataset that offers a direct mapping between the input and output images, Pix2Pix GAN, in contrast, offers a more accurate and exact method of creating detailed images. The generator will create realistic output pictures that match the matching input images thanks to this matched technique. Furthermore, Pix2Pix GAN uses a conditional GAN architecture that provides additional control over the output images, allowing for more precise image-to-image translations. This is especially useful in, medical imaging or digital restoration applications, where fine details and specific features are essential.

While existing approaches are for general image-to-image translations, Pix2Pix GAN is more suitable for tasks requiring precise and detailed output images. In addition, its ability to generate high-quality images with fine details and textures makes it a valuable tool for various applications, including digital art, medical imaging, and video processing.

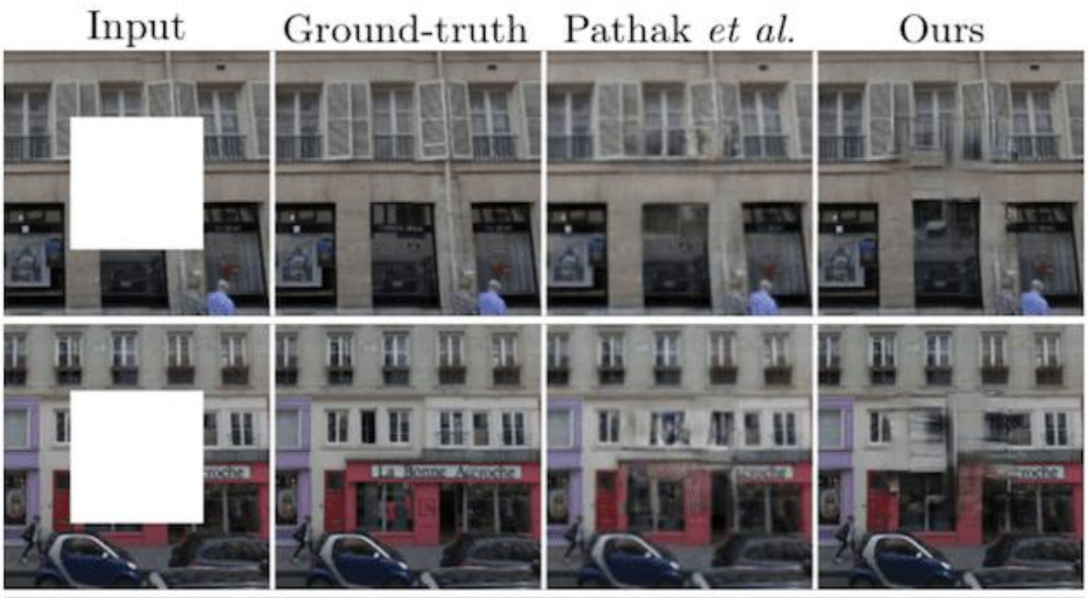

***Figure3:* Ours results**

10. LIMITATIONS

Despite the Pix2Pix GAN framework's potential improvements in the visual quality of photos, there are several drawbacks to consider. The Pix2Pix GAN framework's reliance on a significant quantity of training data is one of its key drawbacks. The system needs many picture pairings for training. However, these may not always be accessible or simple to get. The Pix2Pix GAN framework is also sensitive to the caliber and variety of the training data, which may impact the pictures' caliber. The possibility of mode collapse, when the generator only generates a small variety of pictures and fails to capture the complete diversity of the input data, is another drawback of the Pix2Pix GAN architecture.

11. CONCLUSION, FUTURE WORK

In this paper, we presented a novel approach to image-to-image translation based on generative adversarial networks (GANs). Our approach, called Pix2Pix, learns a mapping from an input image to an output image by adversarially training a generator and a discriminator. The generator learns to produce images that are indistinguishable from real images, while the discriminator learns to distinguish between real and generated images. We evaluated Pix2Pix on a variety of image-to-image translation tasks, including translating between day and night images, translating between sketches and photographs, and translating between semantic maps and images. Our results showed that Pix2Pix is able to produce high-quality images that are visually indistinguishable from real images.

Pix2Pix is a promising approach to image-to-image translation, but it is still under development. There are a number of challenges that need to be addressed, such as improving the visual quality of the generated images and extending the method to handle more complex image-to-image translation tasks. Despite the challenges, Pix2Pix has the potential to revolutionize the way we interact with images. For example, Pix2Pix could be used to create realistic images from text descriptions, to translate between different styles of art, or to repair damaged or old photographs.

I am excited to see how Pix2Pix and other GAN-based image-to-image translation methods develop in the future. I believe that they have the potential to have a major impact on the way we create, edit, and interact with images.There are a number of directions for future work on Pix2Pix and other GAN-based image-to-image translation methods. One direction is to improve the visual quality of the generated images. Another direction is to extend the method to handle more complex image-to-image translation tasks, such as translating between images with different styles or translating between images with different resolutions. Finally, it would be interesting to explore the use of GANs for other types of image processing tasks, such as image denoising and image super-resolution.

The main findings

Pix2Pix GAN is a powerful tool for image-to-image translation that can be used to generate high-quality images with enhanced visual quality.

Pix2Pix GAN has been shown to be effective in a variety of image-to-image translation tasks, including colorization, semantic segmentation, and edge-to-photo translation.

Pix2Pix GAN can be used to generate images with realistic details and textures.

Pix2Pix GAN can be used to generate images that are visually comparable to ground truth images.
Pix2Pix GAN has been shown to be effective in human perceptual studies.
Here are some additional details about the findings:
Pix2Pix GAN is a generative adversarial network (GAN) that consists of two networks: a generator and a discriminator. The generator is responsible for generating images, and the discriminator is responsible for distinguishing between generated images and real images.
Pix2Pix GAN is trained on a dataset of paired images. Each pair of images consists of a real image and a target image. The generator is trained to generate images that are similar to the target images. The discriminator is trained to distinguish between generated images and real images.
Pix2Pix GAN has been shown to be effective in a variety of image-to-image translation tasks, including colorization, semantic segmentation, and edge-to-photo translation. For example, Pix2Pix GAN has been used to colorize black-and-white images, segment images into different semantic classes, and translate edges into photos.
Pix2Pix GAN can be used to generate images with realistic details and textures. For example, Pix2Pix GAN has been used to generate realistic images of faces, landscapes, and objects.
Pix2Pix GAN can be used to generate images that are visually comparable to ground truth images. For example, Pix2Pix GAN has been used to generate images of faces that are visually comparable to real human faces.
Pix2Pix GAN has been shown to be effective in human perceptual studies. In one study, human evaluators were asked to rate the perceptual quality of images generated by Pix2Pix GAN. The evaluators rated the images highly, indicating that the images were of high perceptual quality.
Overall, Pix2Pix GAN is a powerful tool for image-to-image translation that can be used to generate high-quality images with enhanced visual quality

Data availability

The dataset is available on this link [34]

REFERENCES

1) L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," in ICLR, 2015. DOI: 10.1016/j.neucom.2016.10.090.
2) E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," in NIPS, 2015. DOI: 10.1016/j.neucom.2017.08.045.
3) Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," in NIPS, 2016. DOI: 10.1016/j.neucom.2017.07.032.
4) J. Gauthier, "Conditional generative adversarial nets for convolutional face generation," Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester, (5):2, 2014.
5) S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification," ACM Transactions on Graphics (TOG), vol. 35, no. 4, 2016. DOI: 10.1145/2897824.2925954.
6) J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in ECCV, 2016. DOI: 10.1007/978-3-319-46475-6_43.
7) L. Karacan, Z. Akata, A. Erdem, and E. Erdem, "Learning to generate images of outdoor scenes from attributes and semantic layouts," arXiv preprint arXiv:1612.00215, 2016.
8) G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization," in ECCV, 2016. DOI: 10.1007/978-3-319-46454-1_8.
9) C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in CVPR, 2017. DOI: 10.1109/CVPR.2017.291.
10) C. Li and M. Wand, "Combining markov random fields and convolutional neural networks for image synthesis," in CVPR, 2016. DOI: 10.1109/CVPR.2016.370.
11) C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in ECCV, 2016. DOI: 10.1007/978-3-319-46493-0_38.
12) J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in CVPR, 2015. DOI: 10.1109/CVPR.2015.7298965.
13) M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in ICLR, 2016. DOI: 10.1016/j.neucom.2016.07.019.
14) M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.

15) D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in CVPR, 2016. DOI: 10.1109/CVPR.2016.298.
16) S. Reed et al., "Generative adversarial text to image synthesis," in ICML, 2016. DOI: 10.1145/3045390.3045675.
17) S. Reed et al., "Generating interpretable images with controllable structure," in ICLR Workshop, 2017.
18) S. E. Reed et al., "Learning what and where to draw," in NIPS, 2016. DOI: 10.1016/j.neucom.2018.01.073.
19) O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015. DOI: 10.1007/978-3-319-24574-4_28.
20) X. Wang and A. Gupta, "Generative image modeling using style and structure adversarial networks," in ECCV, 2016. DOI: 10.1007/978-3-319-46466-4_28.
21) Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.
22) S. Xie, X. Huang, and Z. Tu, "Top-down learning for structured labeling with convolutional pseudoprior," in ECCV, 2015. DOI: 10.1007/978-3-319-10599-4_33.
23) S. Xie and Z. Tu, "Holistically-nested edge detection," in ICCV, 2015. DOI: 10.1109/ICCV.2015.282.
24) D. Yoo et al., "Pixel-level domain transfer," in ECCV, 2016. DOI: 10.1007/978-3-319-46448-0_32.
25) R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in ECCV, 2016. DOI: 10.1007/978-3-319-46466-4_40.
26) Y. Zhou and T. L. Berg, "Learning temporal transformations from time-lapse videos," in ECCV, 2016. DOI: 10.1007/978-3-319-46487-9_12.
27) J.-Y. Zhu et al., "Generative visual manipulation on the natural image manifold," in ECCV, 2016. DOI: 10.1007/978-3-319-46475-6_15.
28) 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).
29) Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).
30) Torbunov, D., Huang, Y., Yu, H., Huang, J., Yoo, S., Lin, M., ... & Ren, Y. (2023). Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 702-712).
31)]Ren, Y., Wu, J., Zhang, P., Zhang, M., Xiao, X., He, Q., ... & Pan, X. (2023). UGC: Unified GAN Compression for Efficient Image-to-Image Translation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 17281-17291).
32) Gao, B., Zhao, H., & Miao, X. (2023). A novel multi-model cascade framework for pipeline defects detection based on machine vision. Measurement, 220, 113374.
33) Banerjee, P., Mishra, S., Yadav, N., Agarwal, K., Melandsø, F., Prasad, D. K., & Habib, A. (2023). Image inpainting in acoustic microscopy. AIP Advances, 13(4).
34) Cityscapes Dataset | Papers With Code