DEEP LEARNING FRAMEWORK FOR IMAGE-TO-IMAGE TRANSLATION USING GANS

Dr. Shweta Kaushik, Associate Professor, Department of Information Technology, ABES Engineering College, Ghaziabad, UP, INDIA.
Arsh Srivastava, Student, Department of Information Technology, ABES Engineering College, Ghaziabad, UP, INDIA
Azam Ali, Student, Department of Information Technology, ABES Engineering College, Ghaziabad, UP, INDIA
Devendra Kumar Nagar, Student, Department of Information Technology, ABES Engineering College, Ghaziabad, UP, INDIA

ABSTRACT
This paper aims of exploring the capabilities of GAN's Network, a type of machine learning demonstrated by Ian Goodfellow and associate in 2014. GANs, produced by a generator and a split system, are designed for generative tasks, creating new data samples similar to existing ones. It involves understanding the architecture and working of GANs. While the generator network creates new information, while the discriminator network create data as real or pirate. These networks are trained together in a game-theoretic framework, improving over time. The research work involves practical implementation of GANs using the MNIST dataset. It comprises 28x28 pixel black and white images of handwritten number. The goal is to train a GAN to generate new handwritten digits indistinguishable from the real ones. The Implementation of GANs Application for image-to-image translation, specifically transforms satellite images into maps[16]. This is achieved using a type of GAN known as Pix2Pix, familiar on blend of images (satellite images and their corresponding maps). The given work provides a comprehensive exploration of GANs, from understanding their architecture to implementing them for practical applications. It works as a valuable opportunity for anyone interested in the field of generative models and their applications.

Keywords: Deep learning, Generative Adversarial Networks, Image translation, Medical image, Segmentation.

Introduction
GANs, introduced in 2014 by Ian Goodfellow and colleagues [1], have significantly impacted unsupervised learning[17][18] in machine learning. This section outlines the fundamental concept of GANs, emphasizing their unique architecture comprising neural networks, namely the generator and discriminator [1]. Generator aims to create information resembling real data from a random noise vector, while the discriminator works as a paired classifier comparison between real and generated sets of the data [1]. The training of these networks in a game-theoretic framework leads to continuous improvement in their performance [1].
The project progresses to a practical implementation using the MNIST dataset [2], consisting of 28x28 pixel black and white images of handwritten number. The objective is to train a GAN[24] to generate new handwritten numbers. This are virtually very similar from real ones, showcasing the capability of GANs in generating realistic data [2]. This stage highlights the hands-on application of GANs in a specific context, demonstrating their potential for creative data generation.
The final stage explores a specific application of GANs—image-to-image adaption [3]. Using Conditional GANs and Co-Variational Autoencoders, the project focuses on transforming satellite images into maps [3]. The Pix2Pix GAN architecture is employed, which is trained on pairs of images to grasp the knowledge of mapping from input data to outcomes [3]. This application showcases the versatility of GANs[24] in solving real-world problems, bridging the gap between different types of data.

The project offers a thorough exploration of GANs, providing valuable insights into their architecture, practical implementation, and real-world applications [4][5]. This section highlights the holistic approach of the project, covering various aspects of GANs from theory to application. Real-world examples, such as image-to-image translation[19], demonstrate the broader impact and versatility of GANs in solving complex problems.
Concluding the project, it is emphasized that this exploration serves as a stepping stone for anyone interested in the field of generative models and their applications [4][5]. This project lays the foundation for further studies in generative models, providing a comprehensive understanding of GANs and inspiring future research in the evolving field of machine learning.

Literature

The most significant Generative Adversarial Networks [1] by Goodfellow et al. is the first study that presents the idea of GANs and their art and technology. A system of two neural networks competing with one another in a zero-sum game context is used to build Generative Adversarial Networks, a subset of artificial intelligence algorithms utilized in unconquered machine learning. With many realistic features, this technology can produce images that appear, to human observers, to be at least somewhat legitimate.
The use of GANs in medical image classification and segmentation[23] is covered in the second publication, "Systematic Review of Generative Adversarial Networks[24] for Medical Image Classification and Segmentation [7]" by Jeong et al. It offers insights into how GANs are actually implemented in practice using the MNIST dataset. The article provides a thorough analysis of current research on the use of GANs in medical imaging. It also talks about the difficulties and possible paths forward in this area.
Image-to-image translation[26] is one of the many uses of Generative Adversarial Networks (GANs) that are reviewed in the third publication by Dash et al., "A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines-From Medical to Remote Sensing [8]". An extensive overview of GANs and their uses across a range of fields is given in this work. It goes on the fundamental ideas of GANs, their varied varieties, and their uses in diverse industries.
The application of GANs[24] in anomaly detection is examined in the fourth publication, "Applications of Generative Adversarial Networks in Anomaly Detection: A Systematic Literature Review [9]" by Sabuhi et al. An organized review of the literature on GAN applications for anomaly detection is presented in this work. Future research directions are provided along with a discussion of the benefits and drawbacks of employing GANs for anomaly detection.
The Gui et al.'s fifth publication, "A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications [10]," offers a thorough analysis of the many GAN techniques. An extensive analysis of GAN theory,
algorithms, and applications is presented in this study. It talks about the creation and advancement of GANs, theoretical research, and practical uses in a range of industries.
The final publication by Brophy et al. examines GAN variations created for time series-related applications and is titled "Generative adversarial networks in time series: A survey and taxonomy [11]." A taxonomy and overview of GANs in time series are presented in the work. It talks about the difficulties and possible fixes when using GANs with time series data.

Table 1: Comparative analysis

Author, Year	Summary/Lingistic Feaures used	DataSet	Model

Wenhan Yang, 2020 [6]	The evolution of single image deraining techniques, spanning from traditional modelbased approaches to contemporary data-driven methodologies and potentially even more advanced strategies.	Garg and Nayar's study of video deraining in 2004.	various types of networks such as convolutional neural networks, recurrent neural networks, and generative adversarial networks, indicating the use of deep learning models and techniques in the field of single-image deraining
Z Wang, 2019 [4]	It provides an extensive overview and classification of the applications of Generative Adversarial Networks (GANs) within the realm of computer vision.	MNIST, CIFAR-10 dataset.	Deep Convolutional GANs (DCGANs), Wasserstein GANs (WGANs), and Progressive GANs (PGANs)
J Han, 2018 [12]	Adversarial training's growing importance in AI, especially emotional systems, is outlined, presenting key algorithms and potential future research directions for affective computing and sentiment analysis advancement.	Itrelated to affective computing and sentiment analysis.	It perform conventional models and nondeep neural networks on two benchmark databases for sentiment analysis
Yu Li, 2019 [13]	AsymGAN tackles unbalanced data in image translation with an extra variable, enabling controlled mapping and stability. Proven effective on diverse datasets, it advances unpaired image translation.	Cityscapes and Helen	Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.
Phillip Isola, 2018 [3]	Conditional adversarial networks enable diverse image-to-image translation tasks without manually engineered loss functions or mappings, allowing for broad applicability and ease of adoption.	The CMP Facade Database provided by the Center for Machine Perception at the Czech Technical University in Prague.	Pix2pix, which is a conditional generative adversarial network (cGAN).

2.1 IoT sensors in agriculture

GANs are generative models that acquire the knowledge of a mapping, G: $z \rightarrow y$, from a random noise vector z to an output image y [1]. On the other hand, G: $\{x, z\} \rightarrow y$ is the mapping that conditional GANs[25] learn given an observed picture (x) and a random noise vector (z). An adversarially trained discriminator, D, is trained to detect the generator's "fakes" as accurately as possible. The generator G is trained to produce outputs that cannot be distinguished from "real" images. The objectives of this project are as following:

- The first objective is to understand the basics of GANs, including the roles and consequence of the generator and the discriminator networks.
- The next objective is to implement a GAN[24] on the MNIST dataset. This will involve training a GAN to generate new handwritten digits that are very similar from the original set of the data.
- The next objective is to apply GANs to a more complex task - image-to-image translation[26] with satellite images. The goal is to transform satellite images into maps[16] using techniques like Conditional Adversarial Networks and Co-Variational Autoencoders.
- Understanding Advanced level of GANs: Conditional GANs, Basic architecture and role of Generator and Discriminator in CGANs[15].
- Apply GANs to a more complex task - image-to-image translation[26] with satellite images. The goal is to transform satellite images into maps using techniques like Conditional Adversarial Networks and Co-Variational Autoencoders.

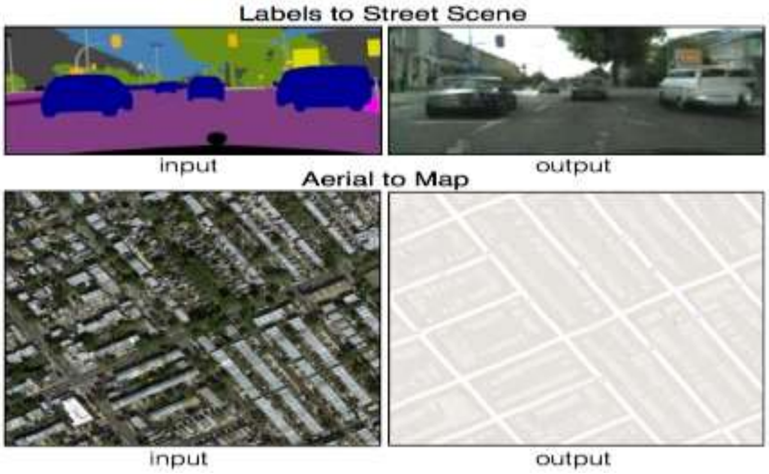

Figure 1: Training a conditional GAN to map edges. The discriminator learns to classify between fake and real tuples. The generator learns to fool the discriminator.

As the generator and the discriminator attempt to concurrently optimize the loss function by minimizing the generator's loss and maximizing the discriminator's loss, GANs[25] objective loss function is also known as min-max loss. Another way to classify the GAN's loss function is as follows:

2.1.1 Generator Loss

Following training, an input image is supplied to the generator, which creates an image. This generated image[20] is then passed to a discriminator, which classifies it as either real or fake. When the discriminator successfully classifies the generated image[21] using a given function, the generator penalizes the discriminator. The generator uses the discriminator to determine its loss.

minEz pNoise [log(1˜D(G(z))]

2.1.2 Discriminator Loss

After being trained, the discriminator network attempts to identify whether images produced by the generator network are real or fake. If it is unable to do so, it penalizes itself for incorrectly identifying real photos as fake or fake images (made by the generator) as real using the specified function.

maxEx pReal [logD(x)] + Ez pNoise [log(1˜D(G(z)))]

- The likelihood that the discriminator will accurately distinguish between a generated image(20, 21, 22) and a genuine image is log(D(x)).
- The Discriminator can accurately identify a sample produced by the Generator network as phony by optimizing log(1 -D(G(z))).

This section emphasizes the various real-world uses for image-to-image conversion[19] systems and provides examples of how effective they are. Three essential topics are covered in each subsection: concrete applications, evaluation measures, and benchmark datasets. The adaptability and significance of image-to-image conversion[26] systems are made clear by looking at these elements. These systems are vital in many fields, from improving satellite imaging to supporting artistic expression and urban planning. By means of a thorough examination of reference datasets and evaluation criteria, scholars determine the efficacy and suitability of these tools, which in turn aids in their improvement and extensive use in real-world scenarios.

3.1. Dataset

To carry out image-to-image conversion[19] tasks, a number of benchmark datasets for image synthesis tasks are available. These datasets vary in terms of picture counts, quality, resolution, complexity, and diversity. They also enable researchers to explore a wide range of real-world applications, including the analysis of urban scenes, cartoon faces, facial features, and semantic applications. Table 2 provides an overview of the chosen benchmark datasets.

Table 2. Various Dataset with Application

Dataset	Model	Metric Name	Metric Value	Application
ImageNet[14]	ResNet-152	Top 1 Accuracy	81.6%	computer vision applications
Cifar-100	CNN36	Percentage correct	36.07%	Image classification and computer vision
Cifar-100	CNN39	Percentage Correct	42.68%	It refine the model's architecture
MNSIT	PixelCNN	Accuracy	98.6%	realistic handwritten digits
Cifar-10	NAT-M4	Accuracy	90-99%	image classification tasks using CNNs
SVHN	WRN28-10	Percentage error	0.99%	digit recognition
Stanford Cars	EffNet-L2	Accuracy	95.96%	computer vision applications
Fashion-MNSIT	SAM	Accuracy	96.41%	Image classification
ModelNet	CAD	Accuracy	90%	object classification, retrieval, and segmentation

3.2. Practical Application

Image-to-image conversion[19] techniques can be applied to several difficulties in computer vision and graphics. With this method, an image is altered to get a desired result. There are two primary types of translation: multimodal translation, which creates numerous outputs from a single image (e.g., diverse artistic styles), and cross-domain translation, which transfers the image to a different visual context (e.g., summer to winter scene). There are many uses for image-to-image translation[26], but this section will concentrate on four common uses: super-resolution, object transfiguration, style transfer, and medical image processing.

3.2.1 Satellite Image to Map

The generator mimics real data by producing new data, such as photographs. As this is going on, the discriminator serves as a critic, attempting to separate the created from the genuine. The generator becomes more adept at producing outputs that are realistic as a result of this competition. Consider mapping towns using satellite pictures using GANs. The discriminator, trained on genuine maps, would attempt to distinguish between an image that is generated and one that is real, while the generator would take a satellite image and create an image (figure 2) that looked like a map.

The graphic illustrates the transformative capability of Generative Adversarial Networks (GANs) in converting satellite images into maps, though it doesn't delve into the intricacies of GAN[25] design or training procedures. It showcases the potential of GANs to bridge the gap between visual data and meaningful representations. The efficacy of this translation process hinges upon the complexity of the GAN architecture and the quality of the training dataset. By discerning patterns and generating corresponding outputs, GANs offer a promising avenue for enhancing spatial understanding and analysis. The success of the translated map[16] depends on the GAN's ability to learn and replicate real-world features accurately.

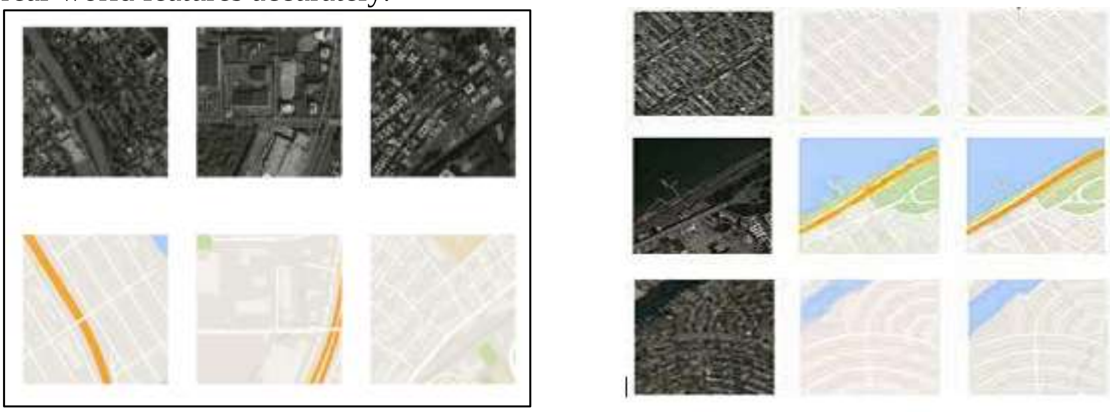

Figure 2. Satellite Image to Map

Conclusion

This project has provided a comprehensive exploration of Generative Adversarial Networks (GANs), from their foundational concepts to practical applications. Beginning with an overview of GAN[24] architecture and training dynamics, the project delved into hands-on implementation using the MNIST dataset, showcasing GANs' ability to generate realistic handwritten digits. Furthermore, the project extended its scope to image-to-image translation[19], demonstrating the application of GANs in transforming satellite images into maps. By employing techniques like Conditional Adversarial Networks and Co-Variational Autoencoders, the project exemplified the versatility of GANs in solving real-world problems. Throughout the exploration, various literature reviews and real-world examples emphasized the significance of GANs[24] across diverse domains, from medical imaging to anomaly detection. With a focus on understanding advanced concepts like Conditional GANs and analyzing GANs' loss functions, this project serves as a valuable resource for individuals interested in generative models and their practical applications. By bridging theoretical insights with practical implementations, this project lays the groundwork for further studies in the evolving field of machine learning, inspiring future research endeavors and advancements in generative model technologies

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," arXiv:1406.2661 [cs.CV], 2014.
[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 19986.
[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," arXiv:1611.07004 [cs.CV], 2016.

[4] Z. Wang, X. Liu, L. Li, Y. Gao, and P. Zhang, "Recent Progress on Generative Adversarial Networks (GANs): A Survey," IEEE Access, vol. 7, pp. 36322-36333, 2019.
[5] Z. Wang, X. Liu, L. Li, Y. Gao, and P. Zhang, "A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 4, pp. 3313-3332, 2023.
[6] W. Yang, R. T. Tan, S. Wang, Y. Fang, and J. Liu, "Single image deraining: From model-based to data-driven and beyond," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
[7] J. J. Jeong, A. Tariq, T. Adejumo, H. Trivedi, J. W. Gichoya, and I. Banerjee, "Systematic Review of Generative Adversarial Networks (GANs) for Medical Image Classification and Segmentation," Journal of Digital Imaging, vol. 35, pp. 137–152, 2022.
[8] A. Dash, J. Ye, and G. Wang, "A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines - From Medical to Remote Sensing," arXiv:2110.01442 [cs.LG], 2021.
[9] M. Sabuhi, M. Zhou, C.-P. Bezemer, and P. Musilek, "Applications of Generative Adversarial Networks in Anomaly Detection: A Systematic Literature Review," IEEE Access, vol. 9, pp. 161003 - 161029, 2021.
[10] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 4, pp. 3313 - 3332, 2023.
[11] E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative adversarial networks in time series: A survey and taxonomy," arXiv:2107.11098 [cs.LG], 2021.
[12] J. Han, Z. Zhang, and B. Schuller, "Adversarial training in affective computing and sentiment analysis: Recent advances and perspectives," IEEE Computational Intelligence Magazine, vol. 14, no. 2, pp. 68–81, 2019.
[13] Yu Li, Sheng Tang, Yongdong Zhang, Jintao Li, Shuicheng Yan: "Asymmetric GAN for Unpaired Image-to-Image Translation, 2019".
[14] Pierre Foret, Ariel Kleiner, Hossein Mobahi, Behnam Neyshabur: "Sharpness-Aware Minimization for Efficiently Improving Generalization".
[15] Salakhutdinov, R.; Hinton, G. "Deep boltzmann machines." In Proceedings of the Artificial Intelligence and Statistics, Clearwater, FL, USA, 16–19 April 2009; pp. 448–455.
[16] Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via atten tion transfer," in International Conference on Learning Representations (ICLR), 2017.
[17] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proceedings of the IEEE International Conference on Computer Vision, 2017.
[18] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in Proceedings of the IEEE International Conference on Computer Vision, 2017.
[19] Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X.; Metaxas, D.N. Stackgan: "Text to photo-realistic image synthesis with stacked generative adversarial networks." In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5907–5915.
[20] A. E. Dirik, S. Bayram, H. T. Sencar, et al., "New features to identify computer generated images," in Image Processing, 2007. ICIP 2007. IEEE International Conference on, 4, IV–433, IEEE (2007).
[21] R. Wu, X. Li, and B. Yang, "Identifying computer generated graphics via histogram features," in Image Processing (ICIP), 2011 18th IEEE International Conference on, 1933–1936, IEEE (2011).

[22] B. Mader, M. S. Banks, and H. Farid, "Identifying computer generated portraits: The importance of training and incentives," Perception 46(9), 1062–1076 (2017).
[23] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3431–3440, 2015.
[24] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. "Generative adversarial nets." In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
[25] Hertzmann, A.; Jacobs, C.E.; Oliver, N.; Curless, B.; Salesin, D.H. "Image analogies." In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; pp. 327–340.
[26] Mo, S.; Cho, M.; Shin, J. Instagan: "Instance-aware image-to-image translation". arXiv 2018, arXiv:1812.10889.