e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

( Peer-Reviewed, Open Access, Fully Refereed International Journal )

Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

# GAN-INDUCED BIAS IN DATA AUGMENTATION

**Kumpatla Mahit Venkat Gautam**[*1], **Gandivalasa Keerthi Tej**[*2], **Seshabhattar Srivastav**[*3], **Anjali**[*4], **S Abhishek**[*5]

[*1,2,3]Amrita School Of ComputingAmrita Vishwa VidyapeethamAmritapuri, India.

[*4,5]Dept. Of Computer Science And EngineeringAmrita School Of Computing
Amrita Vishwa Vidyapeetham Amritapuri, India.

DOI : https://www.doi.org/10.56726/IRJMETS45565

## ABSTRACT

The use of data augmentation techniques, particularly those powered by Generative Adversarial Networks (GANs),has proven crucial for improving machine learning models' performance in a variety of fields. However, as these techniques become more widely used, it is crucial to examine the possible sources of bias that can be introduced. This work explores the important problem of bias in data augmentation, especially when using GANs, providing a thorough analysis of the difficultiesand consequences. The purpose of this research is to look atsituations when this bias is negligible and has no negative effects on performance. This carries out tests to measure bias in different GAN-based DA configurations. The developed methodology is based on the results to assess if GAN-based DA can effectively enhance a given dataset. Based on our attempts to minimize bias,the onward suggestion for mitigating it is the implementation of GAN-based DA.

**Keywords:** Generative Adversarial Network, Machine Learning, Data Augmentation, Configurations.

## I. INTRODUCTION

By expanding the variety and volume of training data,data augmentation techniques have become essential toolsfor machine learning practitioners, allowing them to improvemodel performance. Generative Adversarial Networks (GANs) are one of these methods that have become well-known for their capacity to produce synthetic data that closely resembles real-world data. GAN-based data augmentation has demonstrated efficacy in a range of applications, including natural language processing and computer vision, providing state-of-the-art outcomes in several fields. When the dataset has asmall sample size the dimension of the data is greater thanthe number of occurrences. On the other hand, it is hardto extract more actual extra data from a tiny dataset. If wehave enough data, we can extract a smaller sample of datafrom it [1]. Although GANs have shown to be very effectiveat producing realistic and diverse data, they also highlight a significant but sometimes disregarded issue: the introductionand amplification of biases within augmented datasets. In lightof the fact that machine learning models will soon have asignificant influence on actual decision-making processes, it is critical to recognize and resolve bias in data augmentation. A thorough investigation of bias in GAN-based data augmentation illuminates the complex interactions between generative models and the moral and equitable principles that guide machine learning. Our study aims to reveal the complexstructure of bias in data augmentation, especially when GANs are involved. Our goal is to clarify how biases might enter the augmentation process, the effects of these biases on model behavior, and the moral and societal ramifications of using these models in real-world settings. This exploration is timely,as the machine-learning community grapples with challenges related to transparency, fairness, and accountability in algorithmic decision-making. Biases in augmented data can lead to unequal treatment, reinforce stereotypes, and compromise the ethical foundations of AI systems. In this context, it becomes imperative to not only understand the potential sources of biaswithin GAN-based augmentation but also to propose strategiesand best practices for its detection, mitigation, and ethical deployment. Using case studies from the real world, they analyzethe subtle aspects of bias in GAN-based data augmentationand investigate approaches and policies that support equality, openness, and justice in AI applications. In the process of creating more moral and just artificial intelligence systems,we hope to further the ongoing conversation on responsible AI development and provide academics, practitioners, and policymakers with the tools they need to successfully negotiatethe intricacies of bias in data augmentation.

## II. LITERATURE REVIEW

This work [2] discusses the machine learning, generative adversarial networks, or GANs, have become a valuable tool withthe ability to produce synthetic data for a range of uses. Their use in data augmentation has drawn a lot of interest lately, especially when it comes to the creation of images. It draws attention to the advancements achieved in the application of GANs for data augmentation and emphasizes how generative models may be used to handle issues with complexity andlack of data. This work [3] states that Deep Convolutional Neural Networks (CNNs) are trained on augmented data using Generative Adversarial Networks (GANs), which have become a mainstay in the creation of synthetic data. Although there is no denying that GANs have great promise, earlier research has brought to light some serious difficulties, chief among them being the diversity and realism of the synthetic data that is produced. This study examined recent developments that tackle these issues by using creative synthetic data sampling techniques that leverage the knowledge gained from GANtraining to improve the effectiveness and precision of deep CNNs.

This work [4] defines that a particular problem in the medical imaging arena is the lack of manually annotated data needed to train machine learning algorithms. When dealing with three-dimensional (3D) medical imaging, manually defining diseased areas at the pixel level is a tedious and time- consuming process that frequently calls for the assistance of qualified experts. Because of this, medical imaging-supervised machine learning algorithms often work with small quantities of labeled data even while large volumes of unannotated data are available. Unlabeled data frequently includes important information, although it is less immediately usable. This work [5] proposes that based on the idea that digital twins are representations of actual physical systems, the NVIDIA Omniverse platform has become a ground-breaking solution for unified 3-D production pipelines. This technique makesuse of the Internet of Things (IoT) to gather data from many sources, such as sensors and simulations of finite-element models. The quality of this data, which represents the reactions of physical systems, varies greatly. Acquiring high-fidelity (HF) data might be expensive, but it provides accurate system answers. Low-fidelity (LF) data, on the other hand, is less expensive but cannot provide the required degree of precision. In order to provide precise digital system responses, the discipline of multi-fidelity data fusion (MDF), which focuses on integrating plentiful LF data and restricted HF data, has become more important. This work [6] says that the huge displacement energy, GaN is recognized for having a relativelyhigh radiation hardness; nonetheless, high-energy irradiation particles have the ability to introduce different kinds of defectsand change the equilibrium defect concentrations that are created during development. Defect pairs, complex defect structures, and spot defects can all arise as a result of these high-energy particles. Knowing the nature of these flaws and how they affect radiation-damaged GaN characteristics is criti- cal. Using first-principles computations, the paper investigatesthe 21 defect pairs in GaN generated by two-point defects in this setting. Six unique structural configurations with various defect-defect distances are taken into consideration for each defect pair; this makes a total of 126 structural variants amongthe 21 defect pairs.

## III. METHODOLOGY

### A. Data Augmentation

An essential method in deep learning and machine learning, data augmentation greatly improves the performance and durability of models in a variety of fields. It entails manipulating the original data in a number of ways, creating additional samples, in order to artificially increase the size of a training dataset. When resources are few, data augmentation has shown to be extremely beneficial in reducing overfitting, enhancing generalization, and alleviating the lack of labeled data [7].

The main goal of data augmentation is to increase the adaptability of machine learning models to changes in the input data. Datasets are sometimes unbalanced by nature, with significant elements appearing in disparate orientations, sizes, or lighting conditions in real-world circumstances. Through data augmentation, models can become more resilient to these kinds of modifications by including these differences during training. To improve the model’s performance on unseen data,for example, flipping, rotating, or resizing photographs might aid in object recognition from different perspectives [8].

### B. Generative adversarial networks

The notion of Generative Adversarial Networks, or GANs, has become a groundbreaking idea in the fields of artificial intelligence and machine learning. A generator and a discriminator are the two main parts of a GAN as mentionedin Fig. 1. Neural networks make up these parts; for picture production, deep convolutional networks are usually used. Thediscriminator's job is to discern between actual data and data produced by the generator, while the generator attempts to provide as realistic data samples as it can, such as photographs.The adversarial training mechanism of GANs is the mainnovelty [9]. In a game of cat and mouse, the discriminator and generator are trained concurrently. While the discriminatoraims to improve its ability to discern between created and actual data, the generator tries to produce data that is more and more lifelike. It is this antagonistic dynamic that propels both component's advancement. The generator takes as input random noise and outputs data samples that should ideally be identical to genuine data. This is frequently an iterative process where the generator keeps improving its output. Generating excellent text, photos, music, and other content has proven to be a great accomplishment for GANs [10].

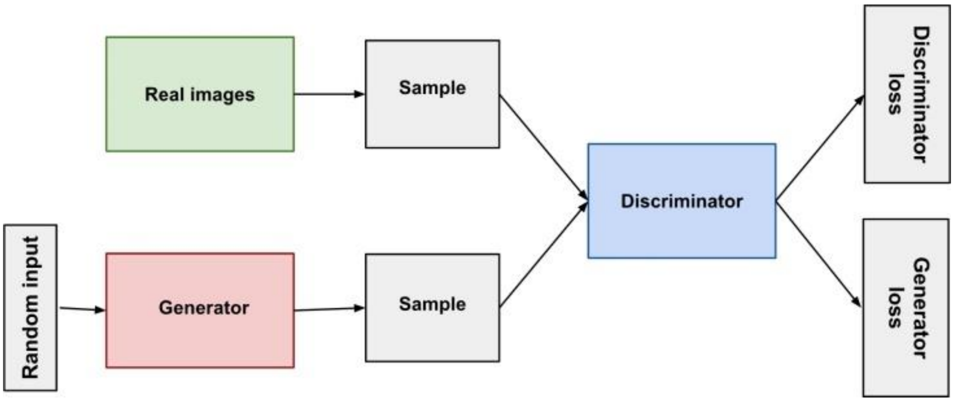

**Fig. 1.** Input for GAN

### C. GAN-based Data Augmentation

Generative Adversarial Networks (GANs) are becoming more than simply a tool for producing lifelike synthetic data; they are revolutionizing the field of data augmentation. GAN- based data augmentation is a cutting-edge method that makes use of GANs' capabilities to add value to training datasets, improving machine learning models' performance and gen- eralizability. The method offers a sophisticated resolution to the ongoing problem of inadequate or unbalanced training data [11]. A generator network that generates synthetic data samples and a discriminator network that attempts to discern between real and synthetic data comprise the initial phase of training a GAN. With time, the generator becomes better at producing realistic data that closely resembles the original dataset's features. The generator is used to create more data samples once the GAN has been trained. The size and variety of the initial training dataset can be considerably increasedby including these artificial examples in it. Machine learning models are then trained on this enhanced dataset [12].

## IV. EXPERIMENTATION

### A. Data Bias Measurements

The ideas of data diversity and bias are inversely connected.To the best of our knowledge, no one has come up with a theoretical definition for data variety. We think that bias is inherited in the data, even if the definition of a classifier'sbias was established. Certain aspects, such as gender traits, could go unnoticed if we just utilize basic data variance to quantify data diversity. Therefore, they ought to be given more consideration when gauging diversity, yet face features are nota reliable indicator of gender [13]. The density of each label group may also be used to describe variety in labeled data. All GAN versions, with the exception of conditional GAN, cannot produce labeled false data; thus, labeling the generated data is a must for computing density. Unlabeled data may be grouped using the clustering technique, and the density can be calculated using the diameters of the clusters and the number of occurrences. However, because it is challenging to calculatethe area of clusters, we can barely employ CURE [14] tohave irregularly shaped clusters if data cannot be efficiently clustered in fixed-axes ellipses. Additionally, it is difficult to calculate the density of a group from many sub-clusters when using the BFR [15] or k-means [16] method since inter-clusterdistance needs to be taken into account.

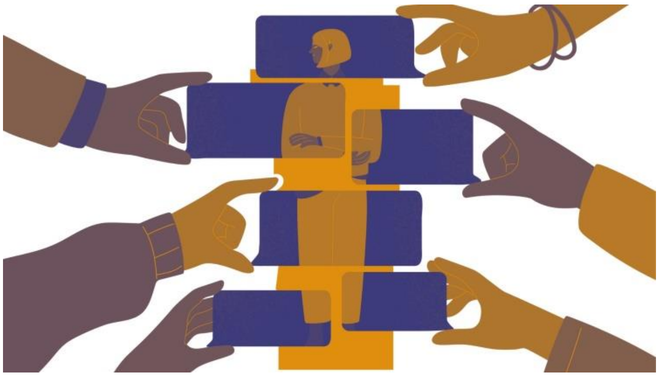

**Fig. 2.** Data Bias Measurements

### B. Experiments on GAN's Variants

Initially, we supplemented with mixed-class data illustratingthe majority of GAN variations are only capable of producing a small number of classes, meaning that their biases are too great to sample various classes fairly [17]. Since the other versions—aside from conditional GAN—cannot producepictures based on the labels provided, we first organize the data according to the labels, and then feed the variants withthe appropriate groups. And in all of the tests that follow,we employ the same training strategy for the remaining four types. The MNIST dataset is for the experiment, initially concentrating on the photographs of a single digit in orderto rapidly assess the quality of the produced data [18].

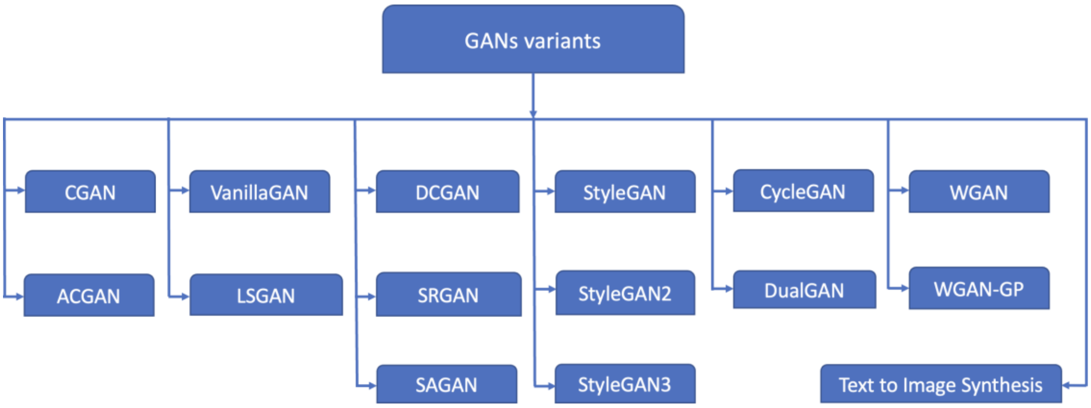

**Fig. 3.** Different types of GAN Variants

## V. RESULTS AND ANALYSIS

The study looks at many bias mitigation techniques that try to undo the bias that was created. These tactics include post- processing modifications, adversarial training, and resamplingmethods. The findings suggest that these tactics can success- fully lessen model prediction bias, providing a road map for achieving justice and equity in AI applications. The particular use case, the type of bias present, and the intended trade-offs between fairness and model performance may all influence the approach selection. Responsible AI development necessitates a thorough understanding of the ethical implications of introducing or perpetuating bias. This includes ensuring that the data augmentation process adheres to ethical guidelines and safeguards, thus preventing biases from seeping into AI systems. Addressing these ethical considerations is essential to building AI systems that are transparent, accountable, and equitable.

It is crucial to recognize that this research has limitations and that the study of bias in GAN-based data augmentation is a developing topic. More research is required to improve bias assessment measures and provide reliable evaluation techniques since bias identification and measurement are still difficult tasks. Furthermore, as biases can be introduced earlierin the process, it is crucial to address them in the GAN training data. Future work can concentrate on expanding the area of AI fairness, producing cutting-edge techniques for mitigating prejudice, and standardizing ethical guidelines for GAN-based data augmentation. By using the creative potentialof GANs to improve machine learning skills, these initiatives will help develop AI systems that put justice, transparency,and accountability first.

## VI. CONCLUSION

In this research, the biases of the original training datacan unintentionally be amplified when using GAN-based data augmentation. While the goal of adding synthetic data is to diversify the dataset, it's possible that this can unintentionally inherit and spread biases from the training set of the GAN. This issue highlights the importance of exercising cautionwhen training GANs and using augmentation techniques. Pre- dictions made by machine learning models based on skewed augmented data may unjustly benefit or disfavor particular classes or demographic groupings. This emphasizes how cru- cial it is to deal with bias in both training data and data augmentation processes. The strategies can effectively reduce bias in model predictions, promoting fairness and equity in AIapplications. The particular use case and the type of bias at play may influence the approach selection. This study has brought attention to the moral issues related to GAN-based data augmentation. It is crucial to make sure that enhanced data is impartial and morally sound. To stopbias from being unintentionally introduced or from continuingto exist during the data augmentation process, rules and safe- guards must be put in place. The goal of ethically progressing AI is intrinsically tied to the quest for machine learning models and bias-free data augmentation. We need to be conscious of the possible repercussions of prejudice in AI systems and work toward accountability, justice, and transparency at every level of development. We can create AI systems that support justice, inclusiveness, and equitable results by detecting and reducing bias in augmented data. We can also leverage the creative potential of GANs to further develop these systems' capabilities. The goal of this research is to shape a future in which artificial intelligence (AI) supports a diverse and equitable society by adding to the continuing conversationon the development of AI that respects and promotes theprinciples of fairness and ethical usage.

## VII. REFERENCES

[1] Mengxiao Hu and Jinlong Li. Exploring bias in gan-based data augmentation for small samples. arXiv preprint arXiv:1905.08495, 2019.

[2] David Liu and Nathan Hu. Gan-based image data augmentation, 2020.

[3] Binod Bhattarai, Seungryul Baek, Rumeysa Bodur, and Tae-Kyun Kim. Sampling strategies for gan synthetic data. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2303–2307. IEEE, 2020.

[4] Christopher Bowles, Roger Gunn, Alexander Hammers, and Daniel Rueckert. Gansfer learning: Combining labelled and unlabelled data for gan based data augmentation. arXiv preprint arXiv:1811.10669, 2018.

[5] Lixue Liu, Xueguan Song, Chao Zhang, and Dacheng Tao. Gan-mdf: An enabling method for multifidelity data fusion. IEEE Internet of Things Journal, 9(15):13405–13415, 2022.

[6] He Li, Menglin Huang, and Shiyou Chen. First-principles exploration of defect-pairs in gan. Journal of Semiconductors, 41(3):032104, 2020.

[7] Emilija Strelcenia and Simant Prakoonwit. A survey on gan techniques for data augmentation to address the imbalanced data issues in credit card fraud detection. Machine Learning and Knowledge Extraction, 5(1):304–329, 2023.

[8] David C Hoaglin, Frederick Mosteller, and John W Tukey. Exploring data tables, trends, and shapes. John Wiley & Sons, 2011.

[9] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. Generative adversarial networks: An overview. IEEE signal processing magazine, 35(1):53–65, 2018.

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Gen- erative adversarial networks. Communications of the ACM, 63(11):139– 144, 2020.

[11] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. IEEE transactions on knowledge and data engineering, 35(4):3313–3332, 2021.

[12] Hadi Mansourifar, Lin Chen, and Weidong Shi. Virtual big data for gan based data augmentation. In 2019 IEEE International Conference on Big Data (Big Data), pages 1478–1487. IEEE, 2019.

[13] Andras Rozsa, Manuel Gu¨nther, Ethan M Rudd, and Terrance E Boult. Are facial attributes adversarially robust? In 2016 23rd International Conference on Pattern Recognition (ICPR), pages 3121–3127. IEEE, 2016.

[14] Sudipto Guha, Rajeev Rastogi, and Kyuseok Shim. Cure: An efficient clustering algorithm for large databases. ACM Sigmod record, 27(2):73– 84, 1998.

[15] Paul Bradley, Johannes Gehrke, Raghu Ramakrishnan, and Ramakrish- nan Srikant. Scaling mining algorithms to large databases. Communi- cations of the ACM, 45(8):38–43, 2002.

[16] Edward W Forgy. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. biometrics, 21:768–769, 1965.

[17] Hung-Yu Tseng, Lu Jiang, Ce Liu, Ming-Hsuan Yang, and Weilong Yang. Regularizing generative adversarial networks under limited data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7921–7931, 2021.

[18] Bo Zhao, Bo Chang, Zequn Jie, and Leonid Sigal. Modular generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV), pages 150–165, 2018.