Hochschule der Medien
Fakultät für Druck und Medien
Computer Science and Media

Generative Data Augmentation

Multi-Agent Diverse Generative Adversarial Networks for Generative Data Augmentation

Dissertation submitted for the degree of Master of Science

Topic:	Generative Data Aufmentation
Author:	Nicolas Reinhart - nr063@hdm-stuttgart.de MatNr. 44100
Version of Date:	March 23, 2025
1. Advisor:	Prof. Dr.-Ing. Johannes Maucher
2. Advisor:	Prof. Dr.-Ing. Oliver Kretzschmar

Abstract

asdf

Contents

List of Figures

List of Tables

1 Introduction and Motivation

Generative Adversarial Networks (GANs) [GPAM+14] and their variants revolutionized the field of computer vision in the year of 2014, enabling advacements in multiple areas of generating data. From *Text to Image Synthesis* [RAY+16], *Image Translation* [IZZE18], *Super Resolution* [LTH+17], *Image Inpainting* [PKD+16], *Style Transfer* [WWR+23] to *Data Augmentation* [SK19], GANs have been used in a variety of applications.

The idea of using GANs for *Generative Data Augmentation* (GDA) has already been applied sucessfully, e.g.: in computer vision [JLR25], [BNI+23] or for creating music [JLY20]. Especially the former survey *A Comprehensive Survey of Image Generation Models Based on Deep Learning* has, along *Variational Auto Encoders* (VAEs), a dedicated focus on GANs. Despite these achievements, in practice, GANs suffer from several challenges, complicating the training and inference process:

- Mode Collapse
- Loss of inter-class Diversity
- Failure to Converge
- Vanishing Gradiants & Unstable Gradiants
- Imbalance between Generator- and Discriminator Model

This thesis investigates the potential of using GANs - specifically *Multi-Agent Diverse Generative Adversarial Networks* (MADGANs) [GKN+18] - for Generative Data Augmentation. MADGANs aim to aid the first two of the afore mentioned in particular: Mode Collapse and Loss of inter-class Diversity. They, along other modifications, "*propose to modify the objective function of the discriminator, in which, along with finding the real and the fake samples, the discriminator also has to correctly predict the generator that generated the given fake sample.*" [GKN+18]. The goal of this adjustment of the discriminator is, that the discriminator has to push the generators towards distinct identifiable modes. While various strategies have been proposed to addrese mode collapse and inter-class diversity MADGANs explicitly enfocre mode seperation by introduction of multiple generators and the adjusted discriminator objective. This makes them particularly promising for GDA, as diverse samples and clear distinction of modes is crucial for training robust classifiers. In their paper, they experimentally show, that their architectural adjustment of GANs is generally capable of giving providing assistance for the first two of the mentioned problems.

The experiments in this work are structured into three major parts. The first set trains and analyses GANs, explicitly MADGANs and *Conditional GANs* (cGANs). Here, the quality of the resulting images during training will be scored by the *Fréchet Inception Distance* (FID) [HRU+18] and the *Inception Score* (IS) [SGZ+16]. The second set uses the afore trained generative models to create images. Images without labels - images originating from MADGANs - will be classified using auxilary classifiers trained using classical data augmentation techniques. The third and most significant set of experiments trains classifiers, using the generated data. For this, stratified classifiers, with differing numbers of real and fake images are trained and evaluated on the respective validation set and their classification performance will be assesed using standard metrics. [1]

All of the above described is executed on the following datasets:

- MNIST [LCB10]
- Fashion MNIST [XRV17]
- CIFAR10 [Kri09]

Aim of the Thesis The general goal is to investigate the robustness of MADGANs and their potential for GDA; and compare their performance against traditional data augmentation techniques, together with data generated by cGANs. Traditional techniques involve altering operation on the training images, such as flipping, rotating, altering their contrasts and adding a small amount of noise to them.

[1] The set of metrics used to asses the quality of the resulting clasifiers is defined in chapter Experiments Setup 5.

2 Related Work

The effectiveness of deep learning models is instrinsically linked to the availability of large and diverse datasets for training. Models with deep and complex architectures require extensive exposure to a wide range of data to learn underlying patterns and generalize well to unseen instances. Insufficient training data can lead to a phenomena called *overfitting*, where a model becomes too specializied to the training data, failing to perfom accurately on previously unencountered data [Yin19].

To mitigate the problem of data scarcity and improve generalization capabilities of deep learning models, data augmentation techniques became indispensable. Data augmentation artificially expands the amounts and diversity of training datasets by creating modified versions of existing data or by generating entirely new instances.

Traditional Data Augmentation

Traditional data augmentation on images typically involves applying various tranformations to existing data. For image based data, augmentations can take a variaty of forms such as [2] :

Geometric Augmentation Modifying the shape, position and perspective: Rotation, Scaling, Flipping, Cropping, Shearing, Perspective Transform.

Photometric Augmentation Altering the pixel values while keeping the spatial structure: Brightness, Contrast, Hue Shift, Noise Injection, Blurring.

Noise-Corruption Augmentation Imitating real-world degradations and distorions of cameras and sensors: Gaussion Noise, Speckle Noise, Salt-and-Pepper Noise.

The sucess of the above mentioned augmentation techniques is established in many papers [PW17], [KSH12], [Yin19], [SK19], [WZZ$^+$13].

The basic GAN framework introduced by Goodfellow and collegues offers a high degree of flexibility and can be adapted for specific augmentation tasks. It can be applied to generate music [DHYY17], speech [LMWN22], text [YZWY17], images [GPAM$^+$14] or other instances of data, e.g. tabular data [XSCIV19].

Especially for image data, *Deep Convolutional GANs* (DCGANs) represent a significant advacement in applying GANs to iamge data augmentation [HFM22]. Their architecture specifically utilizes *Convolutional Neural Network* (CNNs) [LBD$^+$89] in both, the generator and the discriminator. The use of CNNs allows DCGANs to learn hierarchical features from the input images and capture the spatial relationship and structure inherent in images. This leads to the generation of more realistic and coherent synthetic images. A study from Zhau et al. [ZCWD23] applied DCGANs, along their adjusted version of those on multiple dataset, including *Fashion MNIST* and *Cifar10*.

[2] More traditional data augmentation techniques exists, such as Occlusion-Based, Composition-Based, Domain-Specific or Adversarial Augmentation. For the purpose of this work, soley the afore mentioned are discussed in greater detail.

With their experimental setup, they achieved consistent significant improvements over multiple datasets using the DCGAN-architechture, compared to their baseline.

Inherently in the vanilla version of GANs or the DCGANs, the generators role is soley to learn the underlying data distribution of the training samples and produce instances of close resemblance to instances from the training data. This however results in unlabeled samples, not to be beneficial to expand data for a supervised classification task. The introduction of *Conditional Generative Adversarial Networks* (cGANs) allowed to condition the generative process by additional information, such as class labels.

3 Theoretical Background

4 Preliminary Remarks

5 Experiments Setup

6 Experiments Results

Motivation

7 Outlook

repair networks [TFNL22]

8 Conclusion

Data generated by a GAN, may it be a cGAN or a MADGAN, may not fully capturethe distribution characteristics of its training data. Though, generated images do visually appear realistic, they may only partially reflect the statistical characteristics of the original data. This can lead to synthetic images that appear *good* to a human inspector, but may contain amounts of noice that may interfere with a subsequent classifier.

List of References

[BNI+23] Biswas, Angona ; Nasim, MD Abdullah A. ; Imran, Al ; Sejuty, Anika T. ; Fairooz, Fabliha ; Puppala, Sai ; Talukder, Sajedul: *Generative Adversarial Networks for Data Augmentation*. https://arxiv.org/abs/2306.02019. Version: 2023

[DHYY17] Dong, Hao-Wen ; Hsiao, Wen-Yi ; Yang, Li-Chia ; Yang, Yi-Hsuan: *MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment*. https://arxiv.org/abs/1709.06298. Version: 2017

[GKN+18] Ghosh, Arnab ; Kulharia, Viveka ; Namboodiri, Vinay ; Torr, Philip H. S. ; Dokania, Puneet K.: *Multi-Agent Diverse Generative Adversarial Networks*. https://arxiv.org/abs/1704.02906. Version: 2018

[GPAM+14] Goodfellow, Ian J. ; Pouget-Abadie, Jean ; Mirza, Mehdi ; Xu, Bing ; Warde-Farley, David ; Ozair, Sherjil ; Courville, Aaron ; Bengio, Yoshua: *Generative Adversarial Networks*. https://arxiv.org/abs/1406.2661. Version: 2014

[HFM22] Huang, Y. ; Fields, K. G. ; Ma, Y.: A Tutorial on Generative Adversarial Networks with Application to Classification of Imbalanced Data. In: *Statistical Analysis and Data Mining* 15 (2022), Nr. 5, S. 543–552. http://dx.doi.org/10.1002/sam.11570. – DOI 10.1002/sam.11570

[HRU+18] Heusel, Martin ; Ramsauer, Hubert ; Unterthiner, Thomas ; Nessler, Bernhard ; Hochreiter, Sepp: *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*. https://arxiv.org/abs/1706.08500. Version: 2018

[IZZE18] Isola, Phillip ; Zhu, Jun-Yan ; Zhou, Tinghui ; Efros, Alexei A.: *Image-to-Image Translation with Conditional Adversarial Networks*. https://arxiv.org/abs/1611.07004. Version: 2018

[JLR25] Jun Li, Wei Z. Chenyang Zhang Z. Chenyang Zhang ; Ren, Yawei: A Comprehensive Survey of Image Generation Models Based on Deep Learning. In: *Annals of Data Science* 12 (2025), February, 141–170. http://dx.doi.org/10.1007/s40745-024-00544-1. – DOI 10.1007/s40745–024–00544–1

[JLY20] Ji, Shulei ; Luo, Jing ; Yang, Xinyu: *A Comprehensive Survey on Deep Music Generation: Multi-level Representations, Algorithms, Evaluations, and Future Directions*. https://arxiv.org/abs/2011.06801. Version: 2020

[Kri09] Krizhevsky, Alex: *Learning Multiple Layers of Features from Tiny Images*. Technical Report, University of Toronto. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf. Version: 2009

[KSH12] Krizhevsky, Alex ; Sutskever, Ilya ; Hinton, Geoffrey E. ; Pereira, F. (Hrsg.) ; Burges, C.J. (Hrsg.) ; Bottou, L. (Hrsg.) ; Weinberger, K.Q. (Hrsg.): *ImageNet Classification with Deep Convolutional Neural Networks*. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf. Version: 2012

[LBD+89] LeCun, Y. ; Boser, B. ; Denker, J. S. ; Henderson, D. ; Howard, R. E. ; Hubbard, W. ; Jackel, L. D.: Backpropagation Applied to Handwritten Zip Code Recognition. In: *Neural Computation* 1 (1989), Nr. 4, S. 541–551. http://dx.doi.org/10.1162/neco.1989.1.4.541. – DOI 10.1162/neco.1989.1.4.541

[LCB10] LeCun, Yann ; Cortes, Corinna ; Burges, CJ: MNIST handwritten digit database. In: *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist* 2 (2010)

[LMWN22] Li, Xiaomin ; Metsis, Vangelis ; Wang, Huangyingrui ; Ngu, Anne Hee H.: *TTS-GAN: A Transformer-based Time-Series Generative Adversarial Network*. https://arxiv.org/abs/2202.02691. Version: 2022

[LTH+17] Ledig, Christian ; Theis, Lucas ; Huszar, Ferenc ; Caballero, Jose ; Cunningham, Andrew ; Acosta, Alejandro ; Aitken, Andrew ; Tejani, Alykhan ; Totz, Johannes ; Wang, Zehan ; Shi, Wenzhe: *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*. https://arxiv.org/abs/1609.04802. Version: 2017

[PKD+16] Pathak, Deepak ; Krahenbuhl, Philipp ; Donahue, Jeff ; Darrell, Trevor ; Efros, Alexei A.: *Context Encoders: Feature Learning by Inpainting*. https://arxiv.org/abs/1604.07379. Version: 2016

[PW17] Perez, Luis ; Wang, Jason: *The Effectiveness of Data Augmentation in Image Classification using Deep Learning*. https://arxiv.org/abs/1712.04621. Version: 2017

[RAY+16] Reed, Scott ; Akata, Zeynep ; Yan, Xinchen ; Logeswaran, Lajanugen ; Schiele, Bernt ; Lee, Honglak: *Generative Adversarial Text to Image Synthesis*. https://arxiv.org/abs/1605.05396. Version: 2016. – arXiv:1605.05396

[SGZ+16] Salimans, Tim ; Goodfellow, Ian ; Zaremba, Wojciech ; Cheung, Vicki ; Radford, Alec ; Chen, Xi: *Improved Techniques for Training GANs*. https://arxiv.org/abs/1606.03498. Version: 2016

[SK19] Shorten, Connor ; Khoshgoftaar, Taghi M.: A survey on Image Data Augmentation for Deep Learning. In: *Journal of Big Data* 6 (2019), July, Nr. 1, 60. http://dx.doi.org/10.1186/s40537-019-0197-0. – DOI 10.1186/s40537–019–0197–0. – ISSN 2196–1115

[TFNL22] Tanno, Ryutaro ; F.Pradier, Melanie ; Nori, Aditya ; Li, Yingzhen: Repairing Neural Networks by Leaving the Right Past Behind. (2022), 07, S. 19. http://dx.doi.org/10.48550/arXiv.2207.04806. – DOI 10.48550/arXiv.2207.04806

[WWR+23] Wang, Hanyu ; Wu, Pengxiang ; Rosa, Kevin D. ; Wang, Chen ; Shrivastava, Abhinav: *Multimodality-guided Image Style Transfer using Cross-modal GAN Inversion*. https://arxiv.org/abs/2312.01671. Version: 2023

[WZZ+13] Wan, Li ; Zeiler, Matthew ; Zhang, Sixin ; Le Cun, Yann ; Fergus, Rob ; Dasgupta, Sanjoy (Hrsg.) ; McAllester, David (Hrsg.): *Regularization of Neural Networks using DropConnect*. https://proceedings.mlr.press/v28/wan13.html. Version: 17–19 Jun 2013 (Proceedings of Machine Learning Research)

[XRV17] Xiao, Han ; Rasul, Kashif ; Vollgraf, Roland: *Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms*. 2017

[XSCIV19] Xu, Lei ; Skoularidou, Maria ; Cuesta-Infante, Alfredo ; Veeramachaneni, Kalyan: *Modeling Tabular data using Conditional GAN*. https://arxiv.org/abs/1907.00503. Version: 2019

[Yin19] Ying, Xue: An Overview of Overfitting and its Solutions. In: *Journal of Physics: Conference Series* 1168 (2019), feb, Nr. 2, 022022. http://dx.doi.org/10.1088/1742-6596/1168/2/022022. – DOI 10.1088/1742–6596/1168/2/022022

[YZWY17] Yu, Lantao ; Zhang, Weinan ; Wang, Jun ; Yu, Yong: *SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient*. https://arxiv.org/abs/1609.05473. Version: 2017

[ZCWD23] Zhao, Gaochang ; Cai, Zhao ; Wang, Xin ; Dang, Xiaohu: GAN Data Augmentation Methods in Rock Classification. In: *Applied Sciences* 13 (2023), Nr. 9, S. 5316. http://dx.doi.org/10.3390/app13095316. – DOI 10.3390/app13095316

Appendix

Declaration of Academic Integrity

Generative Data Augmentation

Multi-Agent Diverse Generative Adversarial Networks for Generative Data Augmentation.

I hereby declare that I have written this thesis independently. I have properly cited all passages that are taken verbatim or in essence from published or unpublished works of others. All sources and aids used in the preparation of this thesis have been fully acknowledged. Furthermore, this thesis has not been submitted, in whole or in substantial part, to any other examination authority for academic credit.

$Signiture:$ $Place, Date:$