Hardware and Software supporting Physiological Measurement (HSPM-2022)

Hardware and Software Supporting Physiological Measurement Workshop

October 27-28, 2022

Konstanz, Germany

Edited by

Ángel Serrano Alarcón and Andrei Boiko

Generative Adversarial Networks: Project Relevant Overview

Rodion Kraft[1][0000-0003-3428-8201], Natividad Martínez Madrid[1] [0000-0003-1965-9481] and Ralf Seepold[2] [0000-0002-5192-2537]

Abstract. Generating synthetic data is a relevant point in the machine learning community. As accessible data is limited, the generation of synthetic data is a significant point in protecting patients' privacy and having more possibilities to train a model for classification or other machine learning tasks. In this work, some generative adversarial networks (GAN) variants are discussed, and an overview is given of how generative adversarial networks can be used for data generation in different fields. In addition, some common problems of the GANs and possibilities to avoid them are shown. Different evaluation methods of the generated data are also described.

Keywords -- Generative Adversarial Networks, Synthetic Data, Machine Learning.

I. Introduction

Nowadays, data is significant in the machine learning community as it is difficult to get valuable data because of privacy laws. Generating synthetic data can help to improve machine learning tools to have more data to train on and to generate better results in classification or other tasks. Generative adversarial networks (GAN) are well-known and established ways to generate data. It is commonly used for image generation. In this work, we will discuss the problems of GANs and which variants of GANs are already used to generate synthetic data for different aspects. This work will discuss GANs, their problems, and the variants that exist to avoid common problems. Then some examples of how GANs are used in the medical field and a short overview of the evaluation metrics will be given.

II. Generative Adversarial Networks

The training of the GAN is based on two networks that are competing. The generator tries to generate values that are synthetic but look real. On the other hand, the discriminator tries to identify the values it gets from the generator as real or fake. The discriminator learns how real data looks like from the real data sample which it get for the training. In every iteration the generator gets feedback from the discriminator, if the generated data is recognized as real or fake [1-7].

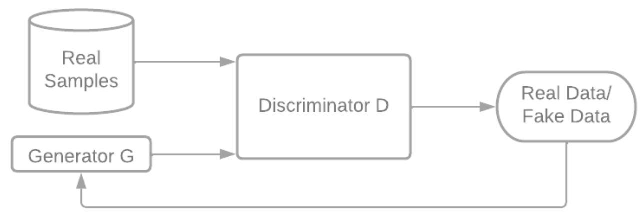

Fig. 1. Structure of a GAN

A. Problems of GANs

A common problem of generative adversarial networks is that the generator may generate samples that the discriminator does not detect as fake. So the generator can remember the characteristics and generate only similar samples where the distribution is insufficient. This is called a mode collapse. Another problem is that if the discriminator is too good, the generator does not get enough information to adjust the procedure. This also can lead to the discriminator and the generator can fail to cover, and so they cannot generate correct results [1,3,5,8,10].

B. Wasserstein GAN

The Wasserstein GAN uses the Wasserstein distance instead of the traditional Jenson-Shannon Divergence for evaluating the real and synthetic distribution. The use of the Wasserstein Distance leads to more stable training of the GAN [3,11].

C. Anonymization Through Data Synthesis using GANs

The difference between a conditional GAN and an ADS-GAN is that the ADS-GAN optimizes each patient's conditioning set and generates new data based on them. On the other hand, the set of variables is predetermined in a conditional GAN. Moreover, to overcome the known limitations of GAN Yoon et al. used a Wasserstein GAN with Gradient Penalty [3].

D. Other GANs Variants

In order to improve the GAN and eliminate its problems, many different GAN versions were developed. For example, there is the Deep convolutional GAN, Cycle GAN, Private Aggregation of Teacher Ensemble GAN, RedGan, METGAN, XGAN, and StarGAN [3,4,8].

[1] Reutlingen University, Alteburgstr. 150, 72762 Reutlingen, Germany
{rodion.kraft,natividad.martinez}@student.reutlingen-university.de,
[2] HTWG Konstanz, Alfred-Wachtel-Str. 8, 78462 Konstanz, Germany
ralf.seepold@htwg-konstanz.de

III. Generative Adversarial Networks in the medicine field

In this section, some examples will show there GANs are used in medicine.

A. Dual-Discriminator Fourier Acquisitive GANs

As a classification of disorders based on optical coherence tomography(OPC) with deep learning needs a lot of data. Tajmirriahi et al. extended a GAN model to generate OPC images. By adding a dual discriminator and a dual-adversarial GAN with a Fourier acquisitive, the model can learn spatial and frequency domain characteristics and create higher-resolution images so that more images of realistic diseases like diabetic macular edema can be used for the classification models to learn. This allows the GAN to multitask learning of spatial and Fourier information of the OCT images. This also leads to an increase in stability and prevents the mode collapse of the GAN. The authors described that with this method, the generated data follow the original data pattern [1].

B. X-ray Images from Point Clouds using Conditional GANs

For simulation X-ray images from point clouds, Haiderbhai et al. used conditional GANS to generate x-ray images. They used a depth camera and other sensors to create point cloud images of hands, and then they generated the images with the GAN. These generated pictures should be a prediction of the real X-ray images. They used the Root Mean Squared Error, the Structural Similarity Index Metric, and the Perak Signal Noise Ratio for validation. In all experiments with different point cloud densities, the results were similar, but the points in the cloud were the better the results were, even if only minimally better [2].

C. Generating ECG Data

Harada et al. changed the generator to an RNN based on LSTMs to generate time-series data. The discriminator is also an LSTM layer with an average pooling layer. [5] Another paper tries to use the basic GAN to generate ECG data, but before training the model, Piacemtino et al. transform the time series into images with a data arrangement algorithm so that no information is missing and the standard use of GAN can be continued [6].

D. Using GANs for protecting private data

One key aspect of using GAN in the medical field is to protect the private data of patients but still allow researcher to have enough data for there projects. GAN can be used to generate synthetic data by using dataset of real data. One way to anonymize the generated data is to add noise to the learning process of the GANs generator during the learning process. An other way is to add the noise directly to the data. This process is also called Differential privacy(DP). A popular DP method is the Laplacian mechanism.[12] The anonymized synthetic data can then be shared between different institutions for different research project without compromising the privacy of the patients. And as more data is available more successful results can be accomplished.

IV. GAN Evaluation Metrics

Kolmogorov-Smirnov test, Pearson Chi-square test, Mean Absolut Error, Mean Sum of Squared Differences, image similarity index measure, Fréchet inception, and learned perceptual image patch similarity are the most commonly used metrics to evaluate the similarity of synthetic and real images. The main issue tested with the metrics is if the distribution of the generated data is similar to the original distribution of the data [1,3,4,7]. If the generated data has a relation in between itself it is useful to use the Persons correlation coefficient which can detect a relations between data [13].

For evaluation generated synthetic time series by GANs the Maximum Mean Discrepancy is used often. This method measures the distance between probability distributions of different data samples [14,15]. Another way to evaluate the results of GAN is to use the generated synthetic data in classification models and compare these results with the classification of the real data. If the results should the similar in the best case. Naveed et al. compared also forecasting models with real and synthetic data to evaluate the used GAN [16].

V. Acknowledgment

This research was partially funded by (a) Carl Zeiss Foundation and the MORPHEUS-Project “Non-invasive system for measuring parameters relevant to sleep quality” (project num-ber: P2019-03-003); and (b) by the German Academic Exchange Service – DAAD within the project "Portable system for detecting and alerting driver fatigue" supported (project number 57562477).

VI. References

1. M. Tajmirriahi, R. Kafieh, Z. Amini and V. Lakshminarayanan, "A Dual-Discriminator Fourier Acquisitive GAN for Generating Retinal Optical Coherence Tomography Images," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-8, 2022, Art no. 5015708, doi: 10.1109/TIM.2022.3189735.
2. M. Haiderbhai, S. Ledesma, N. Navab and P. Fallavollita, "Generating X-ray Images from Point Clouds Using Conditional Generative Adversarial Networks," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 1588-1591, doi: 10.1109/EMBC44109.2020.9175420
3. J. Yoon, L. N. Drumright and M. van der Schaar, "Anonymization Through Data Synthesis Using Generative Adversarial Networks (ADS-GAN)," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 8, pp. 2378-2388, Aug. 2020, doi: 10.1109/JBHI.2020.2980262.
4. I. Horvath et al., "METGAN: Generative Tumour Inpainting and Modality Synthesis in Light Sheet Microscopy," 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022, pp. 3230-3240, doi: 10.1109/WACV51458.2022.00329.
5. S. Haradal, H. Hayashi and S. Uchida, "Biosignal Data Augmentation Based on Generative Adversarial Networks," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 368-371, doi: 10.1109/EMBC.2018.8512396.
6. Piacentino E, Guarner A, Angulo C. Generating Synthetic ECGs Using GANs for Anonymizing Healthcare Data. Electronics. 2021; 10(4):389. https://doi.org/10.3390/electronics1004038
7. O. O. Nedrejord, V. Thambawita, S. A. Hicks, P. Halvorsen and M. A. Riegler, "Vid2Pix - A Framework for Generating High-Quality Synthetic Videos," 2020 IEEE International Symposium on Multimedia (ISM), 2020, pp. 25-26, doi: 10.1109/ISM.2020.00010.

8. Garrett Wilson and Diane J. Cook. 2020. A Survey of Unsupervised Deep Domain Adaptation. ACM Trans. Intell. Syst. Technol. 11, 5, Article 51 (October 2020), 46 pages. https://doi.org/10.1145/3400066
9. Juan Ye, Pakawat Nakwijit, Martin Schiemer, Saurav Jha, and Franco Zambonelli. 2021. Continual Activity Recognition with Generative Adversarial Networks. ACM Trans. Internet Things 2, 2, Article 9 (May 2021), 25 pages. https://doi.org/10.1145/3440036
10. Sudhanva Bhat and Enrique Hortal. 2021. GAN-Based Data Augmentation For Improving The Classification Of EEG Signals. In The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021). Association for Computing Machinery, New York, NY, USA, 453–458. https://doi.org/10.1145/3453892.3461338
11. Gulrajani, Ishaan, Ahmed, Faruk, Arjovsky, Martin, Dumoulin, Vincent and Courville, Aaron Improved Training of Wasserstein GANs. (2017). , cite arxiv:1704.00028Comment: New CIFAR-10 and LSUN image generation experiments
12. S. Imtiaz, M. Arsalan, V. Vlassov and R. Sadre, "Synthetic and Private Smart Health Care Data Generation using GANs," 2021 International Conference on Computer Communications and Networks (ICCCN), 2021, pp. 1-7, doi: 10.1109/ICCCN52240.2021.9522203.
13. Hazra, D.; Byun, Y.-C. SynSigGAN: Generative Adversarial Networks for Synthetic Biomedical Signal Generation. Biology 2020, 9, 441. https://doi.org/10.3390/biology9120441
14. C. Zhang, S. R. Kuppannagari, R. Kannan and V. K. Prasanna, "Generative Adversarial Network for Synthetic Time Series Data Generation in Smart Grids," 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2018, pp. 1-6, doi: 10.1109/SmartGridComm.2018.8587464.
15. V. Avkhimenia, T. Weis and P. Musilek, "Generation of Synthetic Ampacity and Electricity Pool Prices using Generative Adversarial Networks," 2021 IEEE Electrical Power and Energy Conference (EPEC), 2021, pp. 225-230, doi: 10.1109/EPEC52095.2021.9621365.
16. M. H. Naveed, U. S. Hashmi, N. Tajved, N. Sultan and A. Imran, "Assessing Deep Generative Models on Time Series Network Data," in IEEE Access, vol. 10, pp. 64601-64617, 2022, doi: 10.1109/ACCESS.2022.3177906