Generating images using a Deep Convolutional GANs

Author: Kai Deng
Supervisor: Serhiy Yanchuk

A thesis submitted in partial fulfilment of the requirements for the degree of
MSc Mathematical Modelling and Machine Learning

School of Mathematical Sciences,
University College Cork,
Ireland

September 2024

Declaration of Authorship

This report is wholly the work of the author, except where explicitly stated otherwise. The source of any material which was not created by the author has been clearly cited.

Date: 26/07/2024

Signature: Kai Deng

Acknowledgements

I would like to thank...

Abstract

A brief overview of the thesis...

Contents

List of Figures

Chapter 1

Introduction

Generative Adversarial Networks (GANs) have emerged as a powerful class of generative models that revolutionize generative modeling by framing it as a game between two networks: a generator network that produces synthetic data from noise and a discriminator network that distinguishes between the generated data and real data [1]. These networks, introduced in 2014, have found applications in various fields, including materials science, radiology, and computer vision [2], [3], [4]. For example, CycleGAN has been effectively applied in the medical field, notably in medical imaging tasks. It has enhanced liver lesion classification through GAN-based synthetic medical image augmentation, surpassing traditional data augmentation methods in sensitivity and specificity [5]. Abdal et al. (2019) demonstrated the effectiveness of StyleGAN in tasks such as image deformation, style transfer [6]. GANs have gained significant attention in the computer vision community due to their ability to generate data without explicitly modeling the probability density function [3].

The purpose of this paper is to gain a comprehensive understanding of Generative Adversarial Networks (GANs). To achieve this, I will utilize the Animal Faces-HQ dataset, which comprises 16,130 high-quality images with a resolution of 512×512 pixels, to train a standard GAN model specifically designed for generating realistic cat pictures.

The Structure of this thesis.

Chapter 2

Related Work

Deep generative models

Deep generative models, such as Deep Boltzmann Machines (DBMs), are a significant area of research in machine learning. These models are characterized by providing parameterized specifications of probability distribution functions and are typically trained by maximizing the log-likelihood function [7]. DBMs have shown success in extracting deep hierarchical representations of input data, although they often face challenges due to intractable likelihood functions, necessitating multiple approximations of the likelihood gradient [8].

One approach to addressing the challenges in training deep generative models is through techniques like deep tempering, which leverages properties of models like Deep Belief Networks (DBNs) to enhance ergodicity by sampling from deeper levels of the latent variable hierarchy [9]. Additionally, methods like reweighted wake-sleep have been proposed to improve the generative performance of deep models by capturing high-level abstractions and enhancing generalization capabilities [10].

Furthermore, the training of DBMs can be optimized by centering binary variables, which has been shown to improve generative performance and stabilize learning processes [11]. Additionally, the use of mean-field inference in Gaussian Restricted Boltzmann Machines has been explored to meet the increasing demand for analyzing computational algorithms for RBMs in various fields [12].

In conclusion, the development and training of deep generative models like DBMs involve addressing challenges related to intractable likelihood functions, gradient approximations, and model optimization techniques. By leveraging properties of these models, exploring novel training algorithms, and optimizing learning processes, researchers aim to enhance the generative performance and generalization capabilities of deep generative models.

Generative Stochastic Networks

Generative machines have emerged as a solution to the challenges associated with training Deep Boltzmann Machines (DBMs) [13]. One notable example is Generative Stochastic Networks (GSNs), which have been developed to generate samples without explicitly representing the likelihood function. GSNs can be trained using exact backpropagation, eliminating the need for approximate methods required by DBMs. These networks are based on learning the transition operator of a Markov chain to estimate the data distribution. Furthermore, the concept of generative machines has been extended by eliminating Markov chains in generative stochastic networks, enhancing their efficiency and effectiveness [14].

Generative Neurosymbolic Machines represent another advancement in generative models, combining distributed and symbolic representations to support structured symbolic components and density-based generation. This approach leverages the benefits of both types of representations to enhance the overall performance of generative models.

In summary, generative machines like GSNs and Generative Neurosymbolic Machines offer innovative solutions for generating samples without explicitly modeling the likelihood function, thereby overcoming the complexities associated with training DBMs. These advancements in generative models pave the way for more efficient and effective sample generation in machine learning applications.

Variational Autoencoders

Variational Autoencoders (VAEs) have gained significant attention in the field of deep learning due to their ability to learn latent representations of complex data. Kingma and Welling, along with Rezende et al., introduced a stochastic back-propagation rule that enables training VAEs by back-propagation through a Gaussian distribution [15]. This approach pairs a generative network with a discriminative model to perform approximate inference, allowing VAEs to learn to encode data into a low-dimensional latent space and decode it back to the original data [16].

However, VAEs face limitations in modeling discrete data as they require back-propagation through hidden units, which poses challenges in handling such data types effectively [15]. Despite this limitation, VAEs have been extensively used to represent high-dimensional complex data by learning a low-dimensional latent space in an unsupervised manner [17].

In summary, VAEs offer a powerful framework for deep generative modeling, enabling the creation of latent representations of data that can be decoded back to the original form. While they excel in modeling continuous data, challenges persist in effectively modeling discrete data due to the nature of back-propagation through hidden units.

Generative Stochastic Networks

Noise Contrastive Estimation (NCE) is a technique used in training generative models by distinguishing between data and noise samples [18]. It has been applied in various fields such as speech recognition and language modeling due to its ability to handle large vocabularies efficiently [19]. NCE addresses the computational challenges posed by traditional methods like softmax by transforming the estimation problem into a binary classification task [20]. By discriminating between observed data and artificially generated noise, NCE enables the estimation of non-normalized models effectively [21].

One of the key limitations of NCE is the need to evaluate and backpropagate two probability densities, one for the noise distribution and the other for the model distribution [20]. Despite this limitation, NCE has proven to be a highly effective approach for unsupervised representation learning using deep networks [22]. The method has also been extended to flow models for energy-based models, where the update is based on noise contrastive estimation [23].

In the context of training large vocabulary neural language models, NCE has been shown to be a sampling-based technique that offers speed improvements [24]. Researchers have explored different strategies to enhance the training algorithms, including investigating noise contrastive estimation and diagonal contexts for further speed improvements [25]. Additionally, the use of noise-contrastive estimation has been linked to self-supervised tasks in state-of-the-art methods [26].

In conclusion, Noise Contrastive Estimation (NCE) is a valuable technique for training generative models efficiently, particularly in scenarios with large vocabularies. While it has some limitations related to the evaluation of multiple probability densities, researchers continue to explore and extend NCE to address various challenges in unsupervised representation learning.

Predictability minimization

Predictability minimization is a technique that involves training two neural networks competitively to encourage the hidden units of the neural network to be independent of each other. Unlike Generative Adversarial Networks (GANs), where the competition is the sole training criterion, predictability minimization uses a regularizer to promote independence among the hidden units [27].

In the context of predictive coding, predictability minimization aims to reduce prediction errors by minimizing the mismatch between incoming sensations and predictions established through experience. This process involves deviance processing, which is part of an inference process where prediction errors are minimized at different levels of the auditory hierarchy [28].

Regularization techniques play a crucial role in predictability minimization by penalizing model complexity to prevent overfitting. These methods typically involve balancing prediction errors on training data against regularization to optimize the model's performance [29]. Regularization functions often impose constraints on the model to ensure smooth transitions and prevent over-parameterization, thus aiding in minimizing prediction errors [30].

Overall, predictability minimization, through the use of regularization techniques and competitive training of neural networks, aims to enhance the independence of hidden units and reduce prediction errors by optimizing model complexity and promoting smoother transitions within the model.

Chapter 3

Theoretical Background

Generative Adversarial Nets

Chapter 4

Results

This is my results chapter...

Chapter 5

Discussion

In this chapter I provide a discussion and concluding remarks...

Appendix A

Code

An appendix can be used for many things, including relevant code...

Bibliography

[1] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans. 2017.

[2] Y. Jiang. Applications of generative adversarial networks in materials science. *Materials Genome Engineering Advances*, 2, 2024.

[3] Y. Xin, E. Walia, and P. Babyn. Generative adversarial network in medical imaging: a review. *Medical Image Analysis*, 58:101552, 2019.

[4] S. Kazeminia, C. Baur, A. Kuijper, B. Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay. Gans for medical image analysis. *Artificial Intelligence in Medicine*, 109:101938, 2020.

[5] M. Frid-Adar, I. Diamant, E. Klang, M. M. Amitai, J. Goldberger, and H. Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, 2018.

[6] R. Abdal, Y. Qin, and P. Wonka. Image2stylegan: how to embed images into the stylegan latent space? *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[7] H. Xu and Z. Ou. Joint stochastic approximation learning of helmholtz machines. 2016.

[8] K. Zhang and X. Chen. Large-scale deep belief nets with mapreduce. *Ieee Access*, 2:395–403, 2014.

[9] G. Desjardins. Deep tempering. 2014.

[10] J. Bornschein. Reweighted wake-sleep. 2014.

[11] J. Melchior. How to center binary deep boltzmann machines. 2013.

[12] C. Takahashi and M. Yasuda. Mean-field inference in gaussian restricted boltzmann machine. *Journal of the Physical Society of Japan*, 85:034001, 2016.

[13] G. Alain, Y. Bengio, L. Yao, J. Yosinski, É. Thibodeau-Laufer, S. Zhang, and P. Vincent. Gsns: generative stochastic networks. *Information and Inference a Journal of the Ima*, 5:210–249, 2016.

[14] L. Pan, D. Zhang, J. Moksh, L. Huang, and Y. Bengio. Stochastic generative flow networks. 2023.

[15] P. Munjal, A. Paul, and N. Krishnan. Implicit discriminator in variational autoencoder. 2019.

[16] M. Sidulova. Conditional variational autoencoder for functional connectivity analysis of autism spectrum disorder functional magnetic resonance imaging data: a comparative study. *Bioengineering*, 10:1209, 2023.

[17] X. Bie, L. Girin, S. Leglaive, T. Hueber, and X. Alameda-Pineda. A benchmark of dynamical variational autoencoders applied to speech spectrogram modeling. 2021.

[18] B. Damavandi, S. Kumar, N. Shazeer, and A. Bruguier. Nn-grams: unifying neural network and n-gram language models for speech recognition. 2016.

[19] F. Liza and M. Grzes. Improving language modelling with noise contrastive estimation. *Proceedings of the Aaai Conference on Artificial Intelligence*, 32, 2018.

[20] M. Labeau and A. Allauzen. An experimental analysis of noise-contrastive estimation: the noise distribution matters. 2017.

[21] T. Matsuda and A. Hyvärinen. Estimation of non-normalized mixture models and clustering using deep representation. 2018.

[22] P. Awasthi, N. Dikkala, and P. Kamath. Do more negative samples necessarily hurt in contrastive learning? 2022.

[23] R. Gao, E. Nijkamp, D. Kingma, Z. Xu, A. Dai, and Y. Wu. Flow contrastive estimation of energy-based models. 2020.

[24] W. Chen, D. Grangier, and M. Auli. Strategies for training large vocabulary neural language models. 2016.

[25] P. Baltescu and P. Blunsom. Pragmatic neural language modelling in machine translation. 2015.

[26] O. Chehab, A. Gramfort, and A. Hyvärinen. The optimal noise in noise-contrastive learning is not what you think. 2022.

[27] M. Li. Scaling distributed machine learning with the parameter server. 2014.

[28] F. Lecaignard, O. Bertrand, G. Gimenez, J. Mattout, and A. Caclin. Implicit learning of predictable sound sequences modulates human brain responses at different levels of the auditory hierarchy. *Frontiers in Human Neuroscience*, 9, 2015.

[29] H. Liu, K. Verspoor, D. Comeau, A. MacKinlay, and W. Wilbur. Optimizing graph-based patterns to extract biomedical events from the literature. *BMC Bioinformatics*, 16, 2015.

[30] B. Wang, J. Basart, and J. Moulder. Linear and nonlinear image restoration methods for eddy current nondestructive evaluation. pages 791–798, 1998.