Extended Project Qualification | Level 3

An exploration of the Generative Adversarial Network and its Applications within Deep Learning.

James Keywood | Kingston Grammar School

James Keywood
3-23-2022

Contents

Abstract

The Generative Adversarial Network is a deep learning algorithm used to generate images from a set of given input. It uses an unsupervised and generative technique, and has applications within design, image editing and computer vision.

The purpose of this project is to bring light to this important and rapidly growing field within machine learning. The GAN model has the capability to generate high resolution pictures when given an input of example images. Some models have furthered this, by only requiring a text input, yet still resulting in a high-quality output. This broad range of application is the reason for the GAN model's recent success and progression within deep learning.

When researching for this project, I tried to understand where the GAN model sits within artificial intelligence, machine learning and deep learning. After this, I researched how the model actually works, how it has developed since its creation in 2014 and the various times it has been applied in the last decade. This research has led me to the belief that the GAN model has played a big part in machine learning in recent years and will be even bigger, or at least act as a stepping stone, in the future of deep learning algorithms.

Introduction

The GAN model and Deep Learning

Machine Learning is the science that is "concerned with the question of how to construct computer programs that automatically improve with experience".[1] The Generative Adversarial Network, or GAN, model sits within a subset of Machine Learning, known as Deep Learning, which is no exception to the goal of constructing an intelligent program. Although Machine Learning and Deep Learning both fall under Artificial Intelligence, they focus purely on learning, mainly from data, whereas AI is about solving, reasoning, and learning in general.[2] John McCarthy, an early AI computer scientist at Stanford, described Artificial Intelligence as "the science and engineering of making intelligent machines".[3] The Generative Adversarial Network, or GAN, can be used to produce a program that improves with increased data and time. What is unique about the GAN model however is its ability to produce exemplar data from a dataset of examples. This is extremely useful in the modern day due to the ever-increasing need for and importance of data, and datasets.

The Importance of Data

The amount of data collected across the globe is growing exponentially. Data is becoming more meaningful and important, which breaks new ground for machine learning algorithms, especially deep learning. Algorithms are moving out of research labs into production due to the growing amount of data that needs to be processed. For example, substantial amounts of sensor readings and hyperspectral images of plants can be used to identify drought conditions and to gain insights into when and how stress impacts plant growth and development and in turn how to counterattack the problem of world hunger.[4] This is just one example of the countless, important applications of data made a reality by deep learning. However, the process of creating labelled training data is usually the most time-consuming and expensive part of creating a deep learning algorithm. Learning

[1] (Mitchell 1997)
[2] (Kersting 2018)
[3] (McCarthy 2012)
[4] (Kersting 2018)

to play video games may require hundreds of hours of training experience and/or expensive computing power.[5] This, again, is where the GAN model shines. It does not necessarily require labelled data, instead it can be used to create more examples of the dataset given, or even label unlabelled data within a dataset. The GAN model is playing an increasingly significant role within the deep learning and machine learning field.

What is the GAN model?

In my dissertation I will be exploring what the GAN model is and how it has been developed and used since its creation in 2014 by Ian Goodfellow. This includes its many variations that have been created since 2014, and its vast uses and applications in the modern day. The algorithm used in a GAN is quite unique since it requires no supervision, which will be covered in detail further in this dissertation. Furthermore, it uses two networks which compete against each other to produce a realistic output. In short, a generator produces an output which is shown to the discriminator with other real examples, and a discriminator determines whether it is real or fake. The discriminator tells the generator how to make the output more realistic until the discriminator is tricked and labels the generator's output as real. This output would be a realistic output like the exemplar inputs.

How the GAN model can be applied

The GAN model has many applications, but most fall into creating exemplar data from a given dataset. Some examples of this are generating high-resolution versions of input images, creating new and artistic images, sketches, or paintings, and translating images across domains, such as day to night or summer to winter.[6] Along with this, an increasingly popular model known as the StackGAN can be used to create images from given text, which will be a focal point in this dissertation. This is due to the unique aspect of this model, being that the input and output data are different data types. The GAN model and its development in the past decade has shifted the direction in which deep learning is heading and has helped the progression of machine learning.

Motivations for this project

This subject area and topic, being machine learning and the GAN model, are in my eyes extremely important not only now, but also looking into the future. My interest in machine learning stems from my links to computer science in school but also in my free time. This model is able to produce very impressive results with little resources, showing its efficient and well thought out design. This is exactly what motivated me to choose this project and research the topic.

Research Review

Overview

In this dissertation, I will be exploring what a Generative Adversarial Network, or GAN, is and its applications in the past decade. Furthermore, I will be covering what Machine Learning is and how it is linked to Deep Learning, along with why it has been so successful in the past few years. I took a logical approach to researching this field, and therefore began by reading into what machine learning is and how this learning algorithm fits into it. This gave me some overall knowledge surrounding the field of machine learning and allowed me to understand the scale of what I was looking into. Next, I began to read into the GAN algorithm in itself. This includes how the model works when it was created and who by. After this I began to develop this idea by researching how

[5] (Kersting 2018)

[6] (Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs) 2019)

the model has evolved since its birth in 2014, and what advances researchers in the field have been able to make. Finally, I began to investigate its applications over the last decade and how influential it has been to the field of machine learning and deep learning as a whole. The research should be conducted from a range of sources to avoid any weighted opinions, since many companies use this model in their product and therefore promote it in their research articles. This reading will help me to consider the importance of this algorithm in the current day, and its place within deep learning in the future.

Machine Learning and Deep Learning

The initial research that I conducted was focused on finding out what Machine Learning actually is and how it ties in with Deep Learning and the GAN model. Along with this, I looked into the difference between Artificial Intelligence and Machine Learning. This research was done so that I fully understand what kind of algorithm the GAN is and what its goals now and in the future. The first resource I used was the website www.frontiersin.org[7] which I found really helpful in understanding this, due to its highly factual approach to the topic. Furthermore, this source consists of peer-reviewed free access scientific journals which makes it an exceptionally reliable source to take research from. However, the main issue I found when using this source was the amount of detail used in the articles, which meant that I had to skip over some information which I did not understand. Overall, however, this was an especially useful source. Next, I used a couple sources to try to find some quotes from famous researchers in Artificial Intelligence to attempt to define the difference between AI and ML. Initially I found a quote from Professor John McCarthy from the website jmc.stanford.edu[8], which was helpful in defining AI. The next source I used was the book ***Machine Learning***[9] which was useful in finding a good definition for Machine Learning. Both of these sources come from well-known machine learning researchers, being John McCarthy and Tom Mitchell. This perhaps increases the reliability due to the credibility of these authors. However, both of these definitions are quotes, and have aged since, thus may be less useful in my dissertation than first thought.

How the GAN model works

In the next part of my dissertation, I will begin to focus on what the GAN model actually is and how it works. I mainly used an article from www.machinelearningmastery.com[10] which I found extremely useful due to the amount of depth that it went into when describing how the model works. Not only did this source cover the GAN model, but it also covered all the context needed to understand it. This included supervised/unsupervised learning and discriminative/generative modelling which is essential to understand when talking about the GAN's structure. This article references a broad range of sources which adds to its reliability, whilst also being up to date. To further this knowledge, I ventured to find some quotations and in-depth knowledge for this topic, which led me to the book ***Machine Learning: A Probabilistic Perspective***[11] which was published in 2012. This gave me incredibly good definitions of supervised/unsupervised learning. The only issue with this source is the fact that it was made without the GAN model as the focus, although these definitions do not change so it is not really an issue. Next, I used the book ***Pattern Recognition and Machine Learning***[12] to try and understand training data and its relationship with generative/discriminative

[7] (Kersting 2018)
[8] (McCarthy 2012)
[9] (Mitchell 1997)
[10] (Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs) 2019)
[11] (Murphy 2012)
[12] (Bishop 2006)

models. This was also helpful, and I found a lot of useful information surrounding the topic. However, some of the information may be quite dated so I only took the most essential information. Along with this, I used the book ***NIPS 2016 Tutorial: Generative Adversarial Networks***[13] which gave a really good and up to date explanation of the GAN model and why they are defined as adversarial. Finally, I found another similar explanation in the book ***Deep Learning***[14] which focuses more on the training process of the GAN which again was really helpful. All of these sources will be reliable for my dissertation since they are from GAN research papers and will therefore not contain much bias if any, which is immensely helpful.

Evolution of the GAN model

I continued my research onto the topic of the development of the GAN and its evolution since its creation in 2014. This began at www.frontiersin.org[15] which proved to be an extremely useful source yet again. It went through the development of the GAN since the first model and described each step of the process in great detail. It covers many different models which are key in the evolution of the GAN since 2014. This source has proved to be very credible yet went to a degree of detail which was too high for my understanding at some points which was challenging and made this slightly less useful. Along with this, I visited www.towardsdatascience.com[16] to further my knowledge of Deep Convolutional Neural Networks. I found it particularly useful for finding a definition and an explanation at my level of understanding. After reading this source I went to www.machinelearningmastery.com[17] again to look at a specific model of GAN which was a big player in the model's development. This was the Semi-Supervised GAN, which took the model in a slightly different direction by adding a sense of supervision into the previously unsupervised learning model. This has been used a lot since its birth in 2016, only 2 years after the initial creation of the GAN. This was an extremely helpful article and covered all the necessary detail. It is also a reliable source which references many scholarly articles and papers. To further my knowledge of the semi-supervised model I looked into the book ***Improved Techniques for Training GANs***[18] which was helpful in finding key definitions and uses of the model. Furthermore, I found a good explanation of the semi-supervised learning model with the book ***Semi-Supervised Learning with Generative Adversarial Networks***[19]. Both these two sources are reliable due to their educational nature and proved to be very helpful. Penultimately, I looked into the StackGAN, a focus for this dissertation, and found out how it worked and what made it different from other algorithms. The original paper named ***StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks***[20] was especially useful due to the in-depth explanation given. It went into great detail about the text to photo-realistic image process, which makes use of a stacked GAN. This uses multiple layers of processing to break up the steps of generating a high-definition image. Even though this source was exceptionally reliable seeing as it is a scientific research paper, I found a lot of the content overwhelming and too complex for my level and was unable to include it. Finally, I researched the Wasserstein Generative Adversarial Network and used the early 2017 paper ***Wasserstein Generative Adversarial Networks***[21] useful since it explained the concept to a

[13] (Goodfellow, NIPS 2016 Tutorial: Generative Adversarial Networks 2016)
[14] (Goodfellow, Bengio and Courville, Deep Learning 2016)
[15] (Lan, et al. 2020)
[16] (Moolayil 2020)
[17] (Brownlee, How to Implement a Semi-Supervised GAN (SGAN) From Scratch in Keras 2019)
[18] (Salimans, et al. 2016)
[19] (Odena 2016)
[20] (Zhang, et al. 2016)
[21] (Arjovsky, Chintala and Bottou 2016)

satisfactory level. Unfortunately, this source offered lots of information that I could not understand and therefore I could not include everything covered by this paper. The sources I used for this section of my dissertation were extremely helpful at broadening my understanding of the topic, but some went into too much detail for my understanding.

Applications in the past decade

Finally, I began to look into the applications of the GAN model in the past decade. I started by looking at the project www.thispersondoesnotexist.com[22] which highlights the GANs extraordinary uses. The website is able to generate high-definition, portrait faces of people that have never actually existed in real life, all due to the StyleGAN2 model. This was a major help to research since it showed the type of model which is used in actual applications. After this I looked at www.frontiersin.org[23] again, which explained the use of the GAN model in image processing. The article covered super resolution, in which an input of an image of low resolution is turned into a high-resolution output. This has many applications in the real world and was helpful to read about. It also covered image inpainting which reconstructs missing or damaged parts of images or videos, which is also a widely used application of the GAN algorithm. This source was helpful and reliable, yet again. Next, I looked at www.machinelearningmastery.com[24] to research a wider range of applications. This varied from generating examples for image datasets to text-to-image translation (text2image). This again proved to be a reliable source due to the education purpose. Finally, I read the paper called ***StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks, 2017***[25] again which had a lot of information about the applications of the StackGAN. This was helpful and was also reliable since it is a scientific paper from well-known machine learning researchers. The majority of these sources proved to be extremely helpful, reliable, and greatly beneficial to my understanding.

Discussion

The Generative Adversarial Network

Overview

Generative Adversarial Networks, or GANs, are a deep-learning-based generative model.[26] The GAN model architecture involves two sub-models: a generator model and a discriminator model. The generator is used to create new, plausible examples from the domain. The discriminator on the other hand is used to determine and classify whether examples are real, from the domain, or fake, from the generator. GANs are “based on a game theoretic scenario in which the generator network must compete against an adversary”.[27] The generator can be seen as an art counterfeiter, which is trying to make fake paintings. On the contrary, the discriminator can be seen as the police, trying to determine what paintings are real and which ones are counterfeit. For the generator to beat the discriminator, it must produce paintings completely indistinguishable from the real paintings. Both sub-models learn from each other and drive one another to become better at their own functions, thus improving results.

[22] (Wang n.d.)
[23] (Lan, et al. 2020)
[24] (Brownlee, 18 Impressive Applications of Generative Adversarial Networks (GANs) 2019)
[25] (Zhang, et al. 2016)
[26] (Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs) 2019)
[27] (Goodfellow, Bengio and Courville, Deep Learning 2016)

Supervised vs. Unsupervised Learning

Many machine learning algorithms apply a model which makes a prediction, also known as predictive modelling. This requires training data to train the model, which consists of input variables and output class labels. During the learning process, some predictions made by the model may be corrected, which ensures outputs are more like expected outputs. This correction of the model is referred to as a supervised form of learning, or supervised learning.[28] Kevin Murphy, an Artificial Intelligence scientist states that "In the predictive or supervised learning approach, the goal is to learn a mapping from inputs x to outputs y, given a labeled set of input-output pairs".[29] A diagram of this approach can be seen in appendix 1.1.

On the other hand, however, some models only utilise the input variables, and are not given the output labels. Along with this there is no form of correcting here since the model is not predicting anything. This lack of correction is referred to as an unsupervised form of learning, or unsupervised learning.[30] This can be seen in appendix 1.2. Kevin Murphy also states that to "find interesting patterns in the data"[31] is the main goal of unsupervised learning. He also described unsupervised learning as "much less defined" since the model is "not told what kind of patterns to look for" much unlike supervised learning where the predicted value can be compared against the given output label.

Discriminative vs. Generative Modelling

Predictive modelling is used for tasks that involve classification, thus predicting a class label from a set of input variables. Expanding on this, classification is referred to as discriminative modelling. It is referred to as this because the model must discriminate, or in other words choose or decide what class a given example belongs to.[32] This can be seen in appendix 1.3.

On the contrary, unsupervised models which are given input variables may be able to create or generate new examples based on the given data set. These models are known as generative, due to their ability to generate, which is shown in appendix 1.4. In fact, a good generative model may be able to generate new examples that are not just plausible, but indistinguishable from real examples from the problem domain.[33] Professor Christopher Bishop states that "it is possible to generate synthetic data points in the input space"[34], in other words, it is possible to create new input variables, or examples.

GANs as a Two Player Game

Generative modelling may be described as unsupervised, but the GAN model has a clever property that almost acts as an exception to this, since the training of a GAN model can be seen as supervised learning. This is due to the fact that the two models within a GAN, the generator, and the discriminator, are trained together, simultaneously. As stated previously, the generator produces samples, which are presented to the discriminator along with real examples from the given data set. From here the discriminator must classify whether these images are real or fake. After this stage, the discriminator learns how to improve at classifying images between real and fake for the next sample. More importantly, the generator is also updated based on how well, or not, the generated samples

[28] (Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs) 2019)
[29] (Murphy 2012)
[30] (Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs) 2019)
[31] (Murphy 2012)
[32] (Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs) 2019)
[33] (Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs) 2019)
[34] (Bishop 2006)

fooled the discriminator.[35] Due to the way the GAN model works, the two sub-models can be seen to compete against each other, and therefore are adversarial.

The generator and discriminator are both rewarded and penalized depending on how well they did at their own job. If they are rewarded, no changes are made to model parameters. Whereas if they are penalized, model parameters are changed. At some point, a perfect generator would produce samples identical to the given dataset, and the discriminator will have a 50% chance of guessing whether these indistinguishable samples are real or fake. This point does not have to be reached however to have a successful generator. A diagram of how this all ties together can be seen in appendix 1.5.

Development of the GAN

Creation of the GAN

The GAN model architecture was first described in the 2014 paper "Generative Adversarial Networks" by Ian Goodfellow. At first glance the GAN is a great generative model, but the original algorithm had many problems. One of the first of these is an issue known as the vanishing gradient problem. This is caused when the discriminator is, in essence, too good, and does not give enough feedback to the generator, which greatly slows improvement. Along with this, the model could not diversify very easily and only worked when a small scope was applied. General difficulty in training the model also meant that obtaining high quality results was very time consuming.[36] Due to these listed issues, many different optimization methods have been attempted since 2014. Furthermore, theories and articles related to increasing the results and functionality of GANs have been made in an endless stream. The result of this is the creation of many new GAN-based models being proposed, mainly to improve the stability and quality of the generated results.

Deep Convolutional Generative Adversarial Networks (DCGAN)

Almost a year after its creation, the model was standardized to increase stability, and is known as the Deep Convolutional Generative Adversarial Network, also known as the DCGAN. This was formalized by Alec Radford in his 2015 paper on the model. Deep convolutional networks are typically used to identify patterns in images and videos. For this reason, it is extremely useful for a GAN model, due to the common use of images and videos in GAN applications. The DCGAN model has played a crucial part in the development of the GAN, since "most GANs today are at least loosely based on the DCGAN architecture".[37] This model led the way for the many more GAN variations to come in the next decade.

The DCGAN is made possible due to Deep Convolutional Neural Networks (DCCN), which is based on an abstracted version of how human vision and recognition works. Our brains receive signals from the retina about the visuals being perceived from the external world. Initially, edges are detected, then edges are used to help detect curves, which is then used to detect more complex patterns such as shapes[38]. This hierarchical nature of neural activity is also used in a DCCN. A DCCN is made up of three parts: convolution, pooling, and repetition. Convolution is used through matrix manipulation to carry out the process described above, by taking an image and visualising it through edges, then curves and complex shapes. A demonstration of convolution and matrix manipulation can be seen in appendix 2.1. Furthermore, pooling is used to reduce the number of parameters and computation in

[35] (Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs) 2019)
[36] (Lan, et al. 2020)
[37] (Goodfellow, NIPS 2016 Tutorial: Generative Adversarial Networks 2016)
[38] (Moolayil 2020)

the network, by taking many values and turning them into one. This dual use of convolution and pooling are what makes a convolutional neural network particularly useful. Finally, repetition of the process is what creates a deep convolutional neural network, by adding layers in the network.

Semi-supervised Generative Adversarial Networks (SGAN)

The semi-supervised form of learning refers to a problem where a predictive model is required and there are few labeled examples, but many unlabelled examples[39]. When it comes to the GAN model, this kind of case is used in image classification, which may consist of a large dataset of examples, where only a small fraction are labelled. The model learns from this small set of labelled examples and can then begin to label the larger dataset. Tim Salimans, a researcher in the field, had a goal to "improve the effectiveness of generative adversarial networks for semi-supervised learning"[40]. The SGAN can be seen as an extension of the GAN architecture when addressing semi-supervised learning problems[41], and was proposed in June of 2016.

This results in the GAN being trained in two modes: unsupervised and supervised. Initially, the discriminator is trained the same way as a traditional GAN model, predicting whether examples are real or fake. This is the unsupervised training. This kind of training allows the model to learn useful extraction capabilities from a large unlabelled dataset[42]. After this, supervised training is used, and the discriminator is trained to predict class labels of real images. This also allows the model to use the extracted features and apply class labels.

The product is a classifier model that can achieve state-of-the-art results on standard problems when trained on very few labeled examples. Furthermore, the training process can also result in better quality images output by the generator model.[43] Augustus Odena in his 2016 paper titled "Semi-Supervised Learning with Generative Adversarial Networks" demonstrated how a GAN-trained classifier can outperform a standalone DCGAN model with the MNIST handwritten digit recognition task, shown in appendix 3.1. Tim Salimans also achieved these kinds of results with the MNIST dataset in his 2016 paper titled "Improved Techniques for Training GANs". This can be seen in appendix 4.1.

Stacked Generative Adversarial Networks (StackGAN)

To be able to synthesise high-quality images from text descriptions is a challenging feat in the field of computer vision. Simply adding more upsampling layers into the GAN model will result in training instability and nonsensical outputs. Previous to this model, text-to-image approaches could roughly reflect the meaning of the given descriptions. However, high detail and vivid object parts could not be generated. At the time of proposal in late 2016, the StackGAN was able to generate 256x256 photo-realistic images conditioned on text descriptions.[44] This achievement is due to the decomposition of the complex problem described above which is used in the StackGAN model. The process of training the StackGAN is like how human painters draw, by decomposing the problem into two subproblems: drawing the baseline image and then adding the further details.

The StackGAN consists of two main components: the Stage-1 GAN and the Stage-2 GAN. The Stage-1 GAN is used to sketch the "primitive shape"[45] and basic colouring of the object given by the text

---

[39] (Brownlee, How to Implement a Semi-Supervised GAN (SGAN) From Scratch in Keras 2019)
[40] (Salimans, et al. 2016)
[41] (Brownlee, How to Implement a Semi-Supervised GAN (SGAN) From Scratch in Keras 2019)
[42] (Brownlee, How to Implement a Semi-Supervised GAN (SGAN) From Scratch in Keras 2019)
[43] (Brownlee, How to Implement a Semi-Supervised GAN (SGAN) From Scratch in Keras 2019)
[44] (Zhang, et al. 2016)
[45] (Zhang, et al. 2016)

description. The background layout is drawn from a random value, producing a low-resolution image. The Stage-2 GAN then corrects the defects in the low-resolution image given by the Stage-1 GAN and completes details by reading the text description again, producing a high-resolution photo-realistic image.[46] There is another component, however, known as the Conditioning Augmentation technique, which is used to encourage smoothness and variation in the results. This is required due to the limited amount of training data, in the form of text-image pairs, for the Stage-1 GAN. The result of this is a lack of variation from the generator. This Conditioning Augmentation technique is used to encourage small, random changes during the generation process to increase the diversity of synthesized images. The StackGAN decomposes the complex problem of creating high resolution images from text descriptions, and its success can be seen in appendix 5.1.

Wasserstein Generative Adversarial Networks (WGAN)

The WGAN was proposed by Martin Arjovsky, Soumith Chintala and Léon Bottou in their paper titled "Wasserstein Generative Adversarial Networks" in January of 2017. This model is primarily focused on improving the stability of learning. The WGAN is concerned with unsupervised learning[47] and attempts to fix one of the main problems that occurs when trying to train GANs known as mode collapse. This is done by training the generator with an alternate method to usual training.

Mode collapse occurs when the generator only produces a single type of output, or a small range of outputs. This is usually due to problems during the training process. An example of this would be if the generator found a type of data that is easily able to fool the discriminator and thus keeps generating that one type of output. The WGAN attempts to fix this issue by changing how the loss function within the GAN is determined[48]. The loss function is a representation of the distance between the generated data and the real data. The discriminator is changed or replaced with a 'critic' which returns a score of realness of the given image, in contrast to predicting whether images are real or fake. This adaption to the discriminator appears to make the loss function have a more direct correlation to the quality of the images created by the generator. This stops the generator from converging, and mode collapse from occurring whilst also increasing generator quality.

Applications of the GAN

Overview

The GAN model is widely used in image generation. Some of examples of this are face images, flower images, animal images and so on. Along with this, the algorithm can be used to create artistic creations, not only real-life images. Due to the extensive research and development in the past decade, the GAN model has evolved into the progressively growing GAN, also called proGAN.[49] A result of this, for example, is the imaging capability of the GAN improving from a resolution of 32x32 pixels to an HD resolution of 1920x1080 pixels.[50] Not only does this improve the quality of results from a GAN, but simultaneously increases the number of applications that can utilise its impressive function. This includes uses in image processing, such as increasing the resolution of a low-quality image, repairing corrupted photographs, or completely changing the theme of a picture. Furthermore, it can be used to generate images such as high-quality human faces, realistic photographs, and further examples of an image dataset. An example of this success can be seen in appendix 6.1. Finally, one of the most compelling applications of the GAN is its ability to translate

[46] (Zhang, et al. 2016)
[47] (Arjovsky, Chintala and Bottou 2016)
[48] (Arjovsky, Chintala and Bottou 2016)
[49] (Lan, et al. 2020)
[50] (Lan, et al. 2020)

text into images using a StackGAN, as described earlier in this dissertation. The applications and stability of the GAN model has grown significantly since its creation in 2014 and will continue to grow into the next decade.

Image Processing

One of the most effective applications of the GAN model is image super resolution, in which high-resolution images can be generated from an input consisting of a low-resolution image. In 2016 the Super Resolution GAN (SRGAN) was proposed by Christian Ledig. This was the first application of the GAN in the super resolution task. The SRGAN takes a blurred low-resolution image as input and can output a clear image with high resolution. In this case the discriminant model in the SRGAN determines whether images are "true high-resolution" or a "high-resolution image converted from a low-resolution image".[51]

Another impressive application of the GAN is known as image inpainting, which refers to the process of reconstructing missing or damaged parts of images and videos. This involves image editing and generation and is a process of "artificially filling a region where information on the image is missing according to certain rules".[52] Deep learning has strong learning abilities and can learn advanced features from images. When using a GAN, an image is input to the generator with missing parts. The generator uses this input image to generate a new complete image. The discriminator judges whether this image is realistic enough and feedbacks to the generator. Through continuous training for optimization, the generator can produce a complete image that is sufficiently realistic. Then this image is put back into the original image or video and the inpainting is complete.

A further application within image processing is referred to as image-to-image translation. This process can be used to translate images from one domain to another. Examples of this include the translation of painting to photograph, apples to oranges, summer to winter, or even black and white photographs to colour.[53] This was achieved by Philip Isola, et al, in their 2016 paper titled "Image-to-image translation with Conditional Adversarial Networks". Examples of their results can be seen in appendix 10.1 and 10.2.

Image Generation

One of the most effective applications of the GAN model is to generate new images. As described in Ian Goodfellow's original 2014 paper, GANs can create "new plausible examples"[54] for the MNIST handwritten digit dataset for example. Furthermore, the 2017 paper "Progressive Growing of GANs for Improved Quality, Stability and Variation" by Tero Karras, it was demonstrated that the generation of plausible realistic photographs of human faces is possible.[55] This can be seen in appendix 7.1. Adding to this, the project known as This Person Does Not Exist, created by Philip Wang, is a notable example of realistic human faces produced by a GAN model. This project uses a StyleGAN2, also created by Tero Karras in December of 2019. It has been used to generate thousands of portraits of people that have never existed in real life. Examples of this can be seen in appendix 8.1 and 8.2. The website is also not standalone but is also accompanied by many other variations which produce images such as real estate boards, cats and more.

---

[51] (Lan, et al. 2020)
[52] (Lan, et al. 2020)
[53] (Brownlee, 18 Impressive Applications of Generative Adversarial Networks (GANs) 2019)
[54] (Brownlee, 18 Impressive Applications of Generative Adversarial Networks (GANs) 2019)
[55] (Karras, et al. 2017)

On the other hand, realistic photographs are another product of GAN application[56]. Andrew Brock, et al, showed this in the 2018 paper “Large Scale GAN Training for High Fidelity Natural Image Synthesis”. It demonstrated the generation of synthetic photographs with the BigGAN technique, which are practically indistinguishable from real photographs. This can be observed in appendix 9.1 and 9.2.

As described earlier in this discussion, generating photo-realistic images from text is an important problem within image generation, with tremendous applications.[57] These applications include photo-editing, computer-aided design and more. Recently, GANs have shown promising results in synthesising real-world images.[58] When conditioned correctly on given text descriptions, conditional GANs are capable of generating images that are highly related to text meanings.

Conclusion

In this dissertation, I was able to explore the Generative Adversarial Network in great detail, covering how it works, how it has developed and how it can be applied. Initially, I began to research and understand how the algorithm actually works, such as the use of unsupervised learning and generative modelling. Explaining the GAN as a two-player game is a perfect use of abstraction, and I think that it is key to understanding the model.

In the next section of this dissertation, I explored how the GAN has been developed since its creation in 2014. This consists of models such as the DCGAN, which I believe to be one of the most important advances to the GAN model made since 2014. Furthermore, models such as the SGAN and the WGAN proved to be especially useful for certain applications within deep learning. The StackGAN however stood out from the others due to its unique application, of converting text to images, and the very promising results.

Finally, I found many applications of the GAN model, contained within image processing and generation. The GAN model proved to have a vast amount of application within these fields, such as image super resolution which can take a low-resolution image and increase said resolution. This is an outstanding use of the model. Furthermore, image inpainting can be used to reconstruct missing or damaged parts of images and videos, another particularly important problem within image and video editing. Image-to-image translation can also be used to convert the context of an image, such as from day to night or summer to winter. This has great application within photography and image processing. Applications within image generation also include generating new example images for a dataset, which is a simple but powerful use. The generation of realistic images or portraits, using a BigGAN or StyleGAN2 respectively, is another important achievement in the field of deep learning. Finally, I explored the applications of text to image generation with the StackGAN model, which has uses within design and computer vision.

This project has clearly proven to me that the Generative Adversarial Network has a broad range of applications within computer vision and has been extremely important in the advancements of deep learning within the past decade. This algorithm can easily be applied in fields such as designing image mock-ups for clients, generating art from a text description, or even large-scale image

[56] (Brock, Donahue and Simonyan 2018)
[57] (Zhang, et al. 2016)
[58] (Zhang, et al. 2016)

recognition. Looking into the future, I think that this algorithm will play a big part in the field of machine learning or will at least be a stepping stone for other networks and algorithms to come.

Evaluation

Within this project, I believe that I successfully covered the majority of the relevant subject fields within the topic of the GAN model. Furthering this, I think I was able to maintain an unbiased view when exploring the field and kept focus on the topic during research.

However, if I were to do this project again, I would change a few things. Firstly, I would attempt to get professional opinions in the field, such as an interview. Specifically, I would like to contact Ian Goodfellow or Tero Karras, both experts in this field of machine learning. I think this would be extremely helpful for both the reader and I to understand the topic with more detail. Furthermore, I think that it would be great to get professional statements and opinions to improve the depth of this discussion.

Along with this, I think that in the research stage of this project I heavily used a handful of sources. For a few topics in the discussion, I only used one or two sources instead of taking a bit of information from many. If I repeated the process, I would try to vary the sources more to reduce bias and take in a more rounded view or explanation of certain topics.

Finally, I would like to improve the case study of the StackGAN, since I think that it is an extremely interesting project and needs more attention and time in this discussion. Perhaps if I repeated this project, I could look in depth at other projects like this such as the BigGAN, which produces high resolution images of everyday objects.

Overall, I think that my project was successful at covering the broad topic of the Generative Adversarial Network and informs the reader of the topic to a conclusive extent. However, I think that it fails to do the subject area justice and a lot more can be discussed about it.

Bibliography

Arjovsky, Martin , Soumith Chintala, and Léon Bottou. 2016. *Wasserstein Generative Adversarial Networks.* Institute of Electrical and Electronics Engineers.

Bishop, Christopher. 2006. *Pattern Recognition and Machine Learning.* Springer.

Brock, Andrew, Jeff Donahue, and Karen Simonyan. 2018. *Large Scale GAN Training for High Fidelity Natural Image Synthesis.*

Brownlee, Jason. 2019. *18 Impressive Applications of Generative Adversarial Networks (GANs).* 14 June. Accessed November 9, 2021. https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/.

—. 2019. *A Gentle Introduction to Generative Adversarial Networks (GANs).* 17 June. Accessed November 11, 2021. https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/.

—. 2019. *How to Implement a Semi-Supervised GAN (SGAN) From Scratch in Keras.* 24 July. Accessed November 8, 2021. https://machinelearningmastery.com/semi-supervised-generative-adversarial-network/.

Goodfellow, Ian. 2016. *NIPS 2016 Tutorial: Generative Adversarial Networks.*

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning.* MIT Press.

Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei Efros. 2016. *Image-to-Image Translation with Conditional Adversarial Networks.*

Karras, Tero, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. *Progressive Growing of GANs for Improved Quality, Stability, and Variation.*

Kersting, Kristian. 2018 . *Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines.* 19 November. Accessed October 20, 2021. https://www.frontiersin.org/articles/10.3389/fdata.2018.00006/full.

Lan, Lan, Lei You, Zeyang Zhang, Zhiwei Fan, Weiling Zhao, Nianyin Zeng, Yidong Chen, and Xiaobo Zhou. 2020. *Generative Adversarial Networks and Its Applications in Biomedical Informatics.* 12 May. Accessed November 2, 2021. https://www.frontiersin.org/articles/10.3389/fpubh.2020.00164/full.

McCarthy, John. 2012. *What is AI?* 25 March. Accessed October 20, 2021. http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html.

Mitchell, Tom. 1997. *Machine Learning.* McGraw Hill.

Moolayil, Jojo John. 2020. *A Layman's Guide to Deep Convolutional Neural Networks.* 23 May. Accessed February 28, 2022. https://towardsdatascience.com/a-laymans-guide-to-deep-convolutional-neural-networks-7e937628605f.

Murphy, Kevin Patrick. 2012. *Machine Learning: A Probabilistic Perspective.* MIT Press.

Odena, Augustus. 2016. *Semi-Supervised Learning with Generative Adversarial Networks.*

Salimans, Tim, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. *Improved Techniques for Training GANs.*

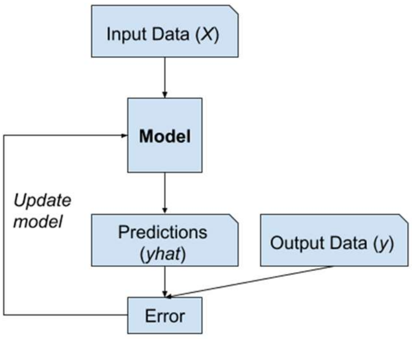
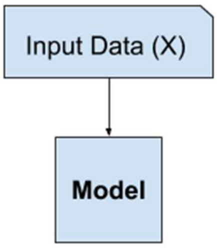
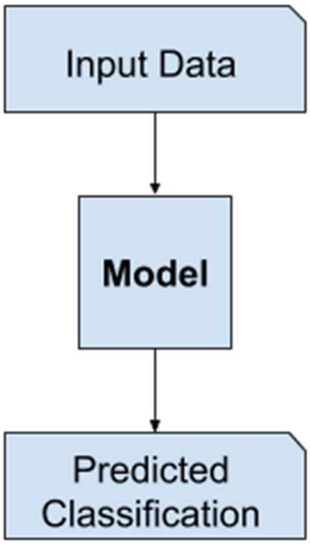
Wang, Phillip. n.d. *This Person Does Not Exist.* Accessed 03 15, 2022. https://thispersondoesnotexist.com/.

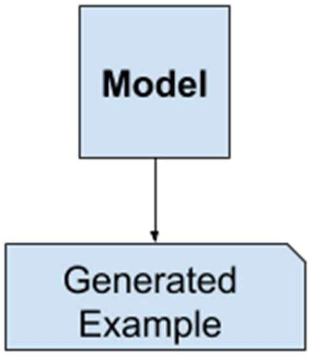
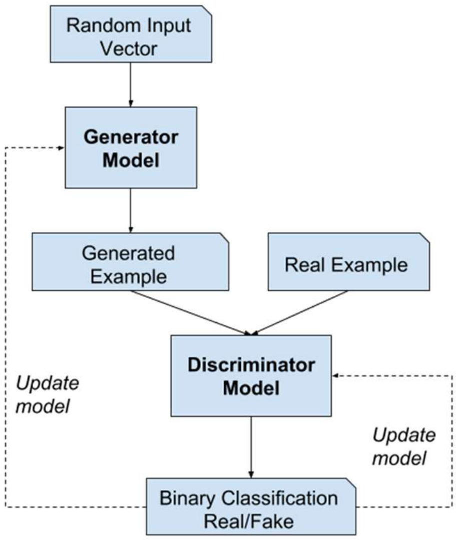
Zhang, Han, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris Metaxas. 2016. *StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks.* Institute of Electrical and Electronics Engineers.

Appendices

1. Diagrams to show the different structures within the GAN

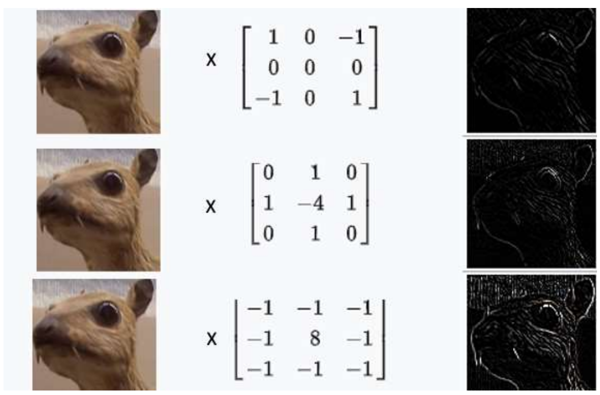
(Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs) 2019)

Number	Figure	Description
1.1		Example of Supervised Learning
1.2		Example of Unsupervised Learning
1.3		Example of Discriminative Modelling

1.4		Example of Generative Modelling
1.5		Example of the Generative Adversarial Network Model Architecture

2. Imagery to explain Deep Convolutional Neural Networks

(Moolayil 2020)

Number	Figure	Description
2.1		A demonstration of how matrix manipulation is used in convolution.

3. Tables of results comparing a CNN to a SGAN

(Odena 2016)

<table>
<tr><th>Number</th><th>Figure</th><th>Description</th></tr>
<tr><td>3.1</td><td>
<table>
<tr><th>EXAMPLES</th><th>CNN</th><th>SGAN</th></tr>
<tr><td>1000</td><td>0.965</td><td>0.964</td></tr>
<tr><td>100</td><td>0.895</td><td>0.928</td></tr>
<tr><td>50</td><td>0.859</td><td>0.883</td></tr>
<tr><td>25</td><td>0.750</td><td>0.802</td></tr>
</table>
</td><td>Table of results comparing classification accuracy of a CNN and SGAN on MNIST.</td></tr>
</table>

4. Table of results to compare the SGAN with other popular models

(Salimans, et al. 2016)

<table>
<tr><th>Number</th><th>Figure</th><th>Description</th></tr>
<tr><td>4.1</td><td>
<table>
<tr><th>Model</th><th colspan="4">Number of incorrectly predicted test examples for a given number of labeled samples</th></tr>
<tr><th></th><th>20</th><th>50</th><th>100</th><th>200</th></tr>
<tr><td>DGN [21]</td><td></td><td></td><td>333 ± 14</td><td></td></tr>
<tr><td>Virtual Adversarial [22]</td><td></td><td></td><td>212</td><td></td></tr>
<tr><td>CatGAN [14]</td><td></td><td></td><td>191 ± 10</td><td></td></tr>
<tr><td>Skip Deep Generative Model [23]</td><td></td><td></td><td>132 ± 7</td><td></td></tr>
<tr><td>Ladder network [24]</td><td></td><td></td><td>106 ± 37</td><td></td></tr>
<tr><td>Auxiliary Deep Generative Model [23]</td><td></td><td></td><td>96 ± 2</td><td></td></tr>
<tr><td>Our model</td><td>1677 ± 452</td><td>221 ± 136</td><td>93 ± 6.5</td><td>90 ± 4.2</td></tr>
<tr><td>Ensemble of 10 of our models</td><td>1134 ± 445</td><td>142 ± 96</td><td>86 ± 5.6</td><td>81 ± 4.3</td></tr>
</table>
</td><td>Table of results comparing classification accuracy of other GAN models to SGAN on MNIST</td></tr>
</table>

5. Figure to show how the StackGAN uses two GAN models.

(Zhang, et al. 2016)

<table>
<tr><th>Number</th><th>Figure</th><th>Description</th></tr>
<tr><td>5.1</td><td>
</td><td>Samples generated by the StackGAN from text descriptions.</td></tr>
</table>

6. Evolution of the GAN model from 2014 to 2017

(Brownlee, 18 Impressive Applications of Generative Adversarial Networks (GANs) 2019)

Number	Figure	Description

6.1		Progression in the Capabilities of GANs from 2014 to 2017.

7. Realistic human photographs using a GAN model

(Karras, et al. 2017)

Number	Figure	Description
7.1		1024x1024 human face images generated.

8. This Person Does Not Exist portrait images

(Wang n.d.)

Number	Figure	Description
8.1		A portrait image produced by the StyleGAN2 model.

8.2		A portrait image produced by the StyleGAN2 model.

9. Realistic photographs produced by a BigGAN

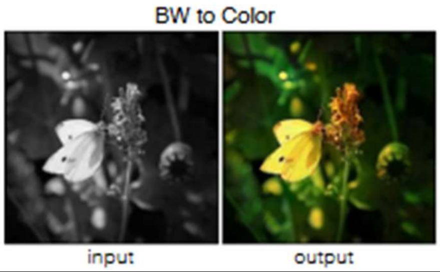
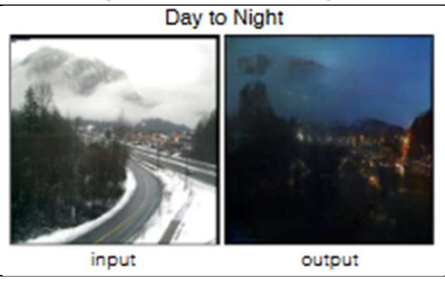
(Brock, Donahue and Simonyan 2018)

Number	Figure	Description
9.1		Realistic 256x256 synthetic photographs generated with BigGAN.
9.2		Realistic 512x512 synthetic photographs generated with BigGAN.

10. Image to Image translation examples

(Isola, et al. 2016)

Number	Figure	Description

10.1		A black and white colour photograph which has been converted to colour using a Conditional Adversarial Network.
10.2		A photograph taken in the day which has been converted to a night time image using a Conditional Adversarial Network.