CS4098 Seminar

Report

Generative Adversarial Networks

Submitted in partial fulfillment of
the requirements for the award of the degree of

Bachelor of Technology
in
Computer Science and Engineering

Submitted by

Roll No	Names of Students
B140598CS	Vrushabh Jambhulkar

तमसो मा ज्योतिर्गमय

Department of Computer Science and Engineering

NATIONAL INSTITUTE OF TECHNOLOGY CALICUT

Calicut, Kerala, India – 673 601

Contents

List of Figures

1

Introduction

Generative Adversarial Networks(GANs) are neural networks that are trained in an adversarial manner to generate data mimicking some dataset or distribution[6]. The framework uses an adversarial process in which simultaneously train two models:

- Generative Model: a model that generates samples to match the data distribution.
- Discriminative Model: a model that learns to determine whether a sample is from the generative model distribution or the data distribution.

The generative model is trained to maximize the probability of discriminative model to give wrong answer. This network framework uses a minimax two-player strategy[7]. These two models compete with each other, where the generator is learning to produce more and more realistic samples, and the discriminator is learning to get better and better at distinguishing generated data from real data. The aim is to generate samples to be indistinguishable from real data[6].

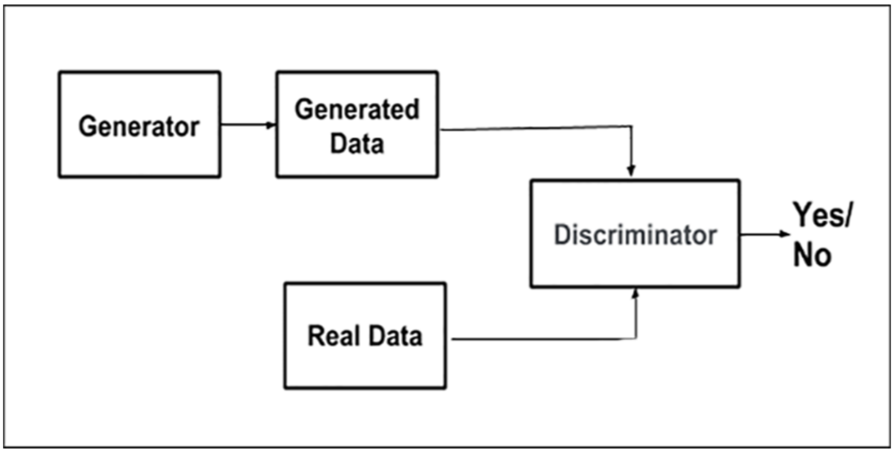

FIGURE 1.1: Structure of GANs[1]

In the procedure [figure:1.1], the generator is input with data along with random noise to produce a fake data. Then the output of the generator is then feed to the discriminator processes this data and tries to indicate that the input is very likely to be fake. The loss of the discriminator is back-propagated to generator to increase its accuracy. Using game theory analysis techniques, it can be proved that there is an equilibrium to this game, between the generator and the discriminator, where the generator makes data that looks identical to the training data and the discriminator assigns probability 1/2 to every input being real or fake[7] The whole idea behind GAN is if you have a generator algorithm a generator that takes its own output as inputs it will do what you want.

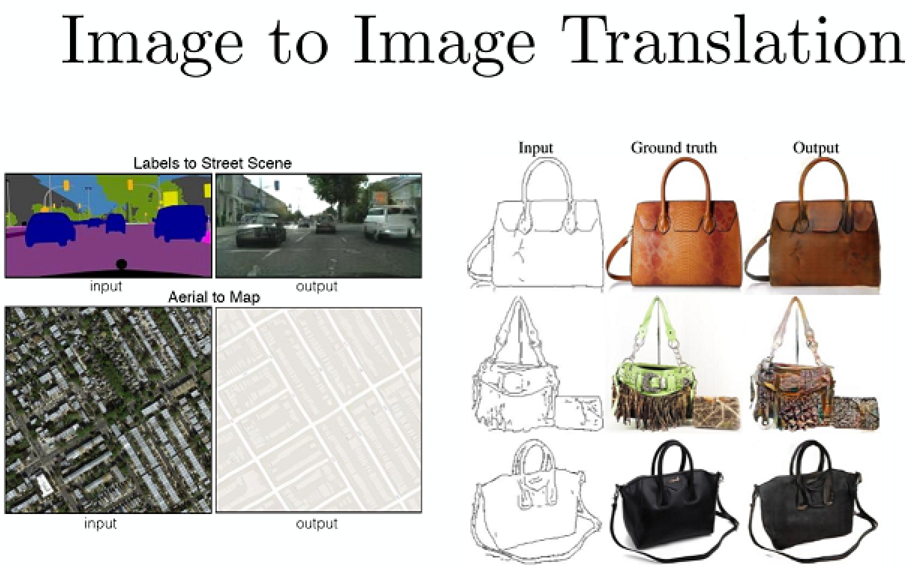

FIGURE 1.2: Example of GANs[2]

2

Motivation

There are many different algorithms and networks that can be used for solving complex problems. Many of these networks use partition functions and Markov Chain Monte Carlo methods, but have trivial estimations or approximation. Other alternatives criteria, that do not approximate or bound the log-likelihood, such as score matching and noise-contrastive estimation (NCE)[7], but they require learned probability density which can sometimes be even impossible to derive one. Some techniques do not involve dening a probability distribution explicitly, but rather train a generative machine to draw samples from the desired distribution. They tend to use the Markov chain model for sampling.

However, the Generative Adversarial Neural networks are better than other algorithms. They do not use probability distributions to discriminate or fit the data, rather they use itself to discriminate generated data from samples a xed noise distribution. They do not use Markov Chain model for sampling. They create a better understanding and help tackle Complex problemsy[8]. They not only generate a Model, but also generate additional samples based on the inputs which can be used to further train the model more effectivel.

3

Seminar Content

Generative Adversarial Networks have advanced since 2014, and have been used in many applications to improve their accuracy in solving complex problems and creating general models to be applied on similar problems. Some of these application have been discussed in the seminar.

WaveNet: A Generative Model for Raw Audio[9]

WaveNet is a generative model developed by the DeepMind. This model was able to generate speech which mimics any human voice and which sounds more natural than the best existing Text-to-Speech systems, reducing the gap with human performance by over 50%. Generating speech with computers is largely based on *Concatenative TTS*, where a very large database of short speech fragments are recorded from a single speaker and then recombined to form complete utterances. This makes it difficult to modify the voice without recording a whole new database. This difficulty help to create an improved model *Parametric TTS*, where all the information required to generate the data is stored in the parameters of the model, and the contents and characteristics of the speech can be controlled via the inputs to the model, but they do not sound highly natural[3]. With the introduction of GANs, DeepMind created WaveNet which was able to model raw waveform of audio signals directly. The model needs to process 16,000 samples per second. For training the model, the input is a real waveform recorded from human speakers given as a sample data set to the discriminator of the GAN and to the generator model. The generator model is also provided with noise along with samples to generate different results[9]. These results are then discriminated and fed back into the input and a new prediction for the next step is made. Wave net was able to mimic US English and Mandarin Chinese, which sounded more realistic than concatenative and parametric [figure:3.1].

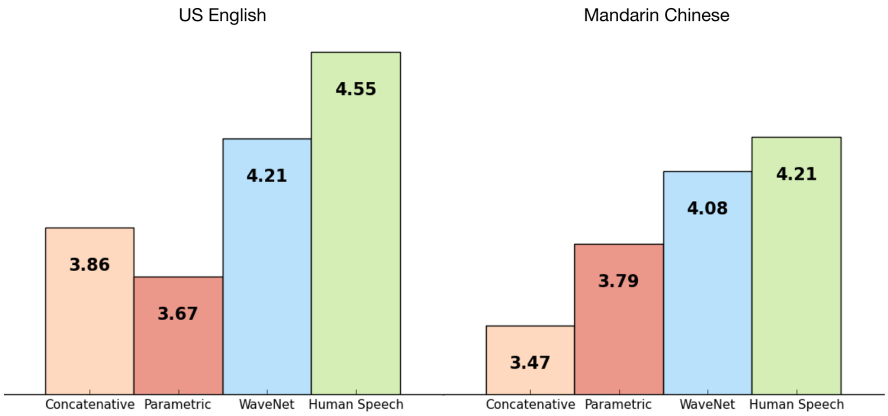

FIGURE 3.1: Comparison of different models with human speech[3]

Photo-Realistic Single Image Super Resolution Using a Generative Adversarial Network[4]

In Image Processing, a highly challenging task of estimating a high resolution (HR) image from its low-resolution (LR) counterpart is referred to as super-resolution (SR). In the past years, there have been many developments in the single image super resolution techniques using deep convolution neural nets, but they still suffer in recovering the fine texture details during up-scaling. This paper[4] uses Generative Adversarial Networks for image super resolution. They use a never ending or changing loss function which consists of an adversarial loss and a content loss. The adversarial loss pushes our solution to the natural image manifold using a discriminator network that is trained to differentiate between the super-resolved images and original photo-realistic images. In addition, we use a content loss motivated by perceptual similarity instead of similarity in pixel space. Our deep residual network is able to recover photo-realistic textures from heavily down sampled images on public benchmarks. An extensive mean-opinion-score (MOS) test shows hugely signicant gains in perceptual quality using SRGAN. The MOS scores obtained with SRGAN are closer to those of the original high-resolution images than any other methods [figure: 3.2]. They were able to achieve 4x up-scaling factor.

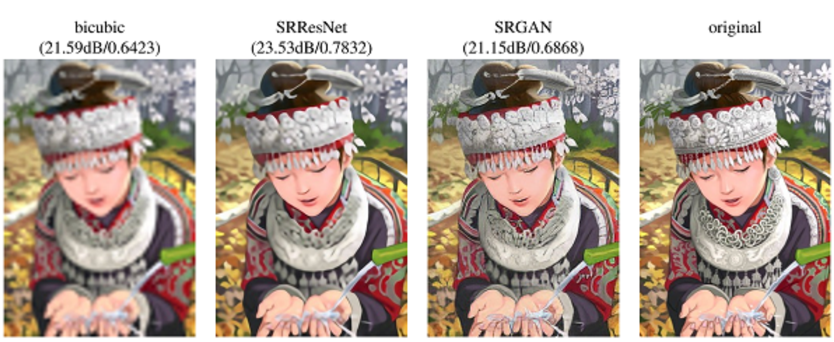

FIGURE 3.2: Comparison of different super resolution methods[4]

Video Imagination from a Single Image with Transformation Generation[5]

Video Imagination is synthesizing imaginary videos from single static image, which maybe easy for humans but not for computers. This requires synthesized videos to be diverse and plausible. Major problems during the process high dimensionality of pixel space and the ambiguity of potential motions. To overcome these problems, Baoyang Chen et. al[5]. proposed a framework that creates imaginary videos using transformation generation. The generated transformations are applied to the original image in a novel volumetric merge network to reconstruct frames in imaginary video. Through sampling different latent variables, our method can output different imaginary video samples. The framework is trained in an adversarial way with unsupervised learning. They were able to create five-frame videos based on a single image supplied to them [figure: 3]. They used a new method of evaluation called relative image quality assessment (RIQA). They also used 3 datasets, two artificial video datasets with simple motion and one natural scene video dataset with complex motions.

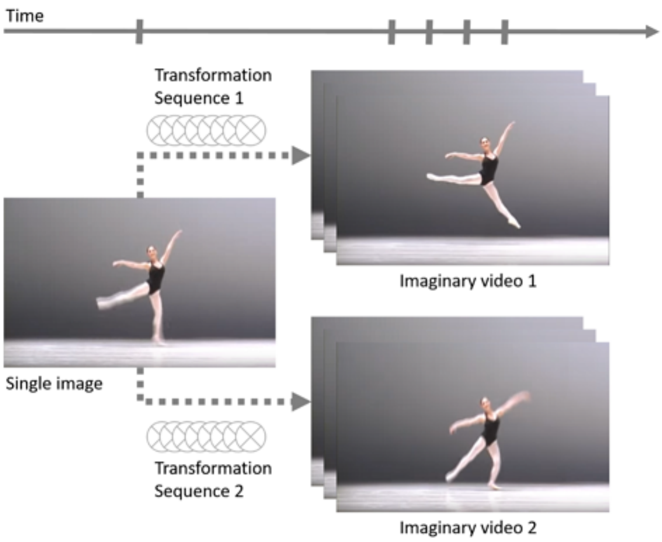

Figure 3.3: Synthesizing multiple imaginary videos from one single image[5]

4

Conclusion

GANs have definitly improved the deep neural networks and complex problem solving. The advantages are that Markov chains are never needed, only back-propagation is used to obtain gradients, no inference is needed during learning, and a wide variety of functions can be incorporated into the model. Adversarial models may also gain some statistical advantage from the generator network not being updated directly with data examples, but only with gradients flowing through the discriminator. This means that components of the input are not copied directly into the generators parameters. Another advantage of adversarial networks is that they can represent very sharp, even degenerate distributions, while methods based on Markov chains require that the distribution be somewhat blurry in order for the chains to be able to mix between modes. Semi-supervised learning features is obtained from the discriminator or inference net which could improve performance of classifiers when limited labeled data is available. Efficiency improvements are found as the training could be accelerated greatly by divising better methods for coordinating Generator and Discriminator or determining better distributions to samples from during training.
This demonstrates the viability of the adversarial modeling framework, suggesting that these research directions could prove useful.

Bibliography

[1] Vrushabh Jambhulkar. Structure of gans, 2017. Note: This figure was made by Vrushabh Jambhulkar, it is unpublished.

[2] Generative adversarial networks(slides - ian goodfellow), 2016. URL `www.iangoodfellow.com/slides/2016-12-04-NIPS.pdf`.

[3] Google DeepMind. Wavenet: A generative model for raw audio, 2016. URL `https://deepmind.com/blog/wavenet-generative-model-raw-audio/`.

[4] Christian Ledig, Lucas Theis, Ferenc Huszr, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. *ArXiv Preprint arXiv:1609.04802*.

[5] Baoyang Chen, Wenmin Wang, Jinzhuo Wang, and Xiongtao Chen. Video imagination from a single image with transformation generation. *ArXiv Preprint arXiv:1706.04124*.

[6] AYLIEN. An introduction to generative adversarial networks (with code in tensorflow), 2015. URL `http://blog.aylien.com/introduction-generative-adversarial-networks-code-tensorflow/`.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. URL `http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf`.

[8] Hitchhikers guide to generative adversarial networks (gans), 2017. URL `https://anthillinside.talkfunnel.com/2017/13-hitchhikers-guide-to-generative-adversarial-networ`.

[9] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *ArXiv Preprint arXiv:1609.03499*.