Generative Adversarial Neural Networks

Seminaar Presentation

Vrushabh Jambhulkar
B140598CS

October 25, 2017

Outline

Introduction: Supervised Learning VS Unsupervised Learning

Supervised Learning

- Algorithm that analyzes the training data and produces a function, which can be used for mapping new examples.
- Dataset: labelled data
- Examples: Classification algorithms, Regression algorithm

Unsupervised Learning

- Algorithm that generates a function to describe the hidden structure from *Unlabeled data*
- Dataset: unlabelled data
- Examples: Clustering algorithm, Generative models

Generative Model VS Discriminative Model[gmv09]

Generative Model

- It models how the data was generated in order to categorize a data.

Discriminative Model

- It does not care about how the data was generated, it simply categorizes a given data.

Generative Adversarial Networks (a.k.a GANs)

- Consists of two models that are trained simultaneously:
  - **Generative Model**: a model that generates samples to match the data distribution to fool the discriminative model.
  - **Discriminative Model**: a model that learns to determine whether a sample is real or fake
- Created by Ian Goodfellow[GPAM+14] in 2014
- Semi-supervised learning
- Two-player minimax game
- Trained until generated samples are indistinguishable from real data

Generative Adversarial Networks (a.k.a GANs)

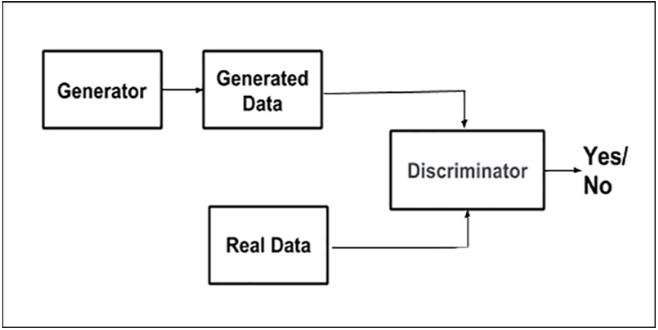

Figure: Generative Adversarial Networks Structure

Motivation

Problems with other Algorithms[GPAM+14]:

- Require learned probability density functions, which is not always possible.
- Use Markov chain model which is highly time consuming.
- Do not give better understanding of complex problems.

Applications: WaveNet

WaveNet: A Generative Model for Raw Audio[vdODZ+]

- Created by DeepMind at Google
- Generate speech which mimics any human voice and sounds more realistic.
- For Text to Speech
- Previous models[Dee16]:
    - **Concatenative TTS(Text To Speech)**:
        - Use very large database of short speech fragments are recorded from a single speaker and then recombined to form complete audio wave.
        - But they have difficulty in modifying voice without creating a new database.
    - **Parametric TTS**:
        - 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

WaveNet: A Generative Model for Raw Audio

GAN based model

- Model raw waveform of audio signals directly.
- Human speaker's audio waveform is given as a sample dataset to the discriminator of the GAN and to the generator model.
- Mimic US English and Mandarin Chinese.
- More realistic [Figure: 2].

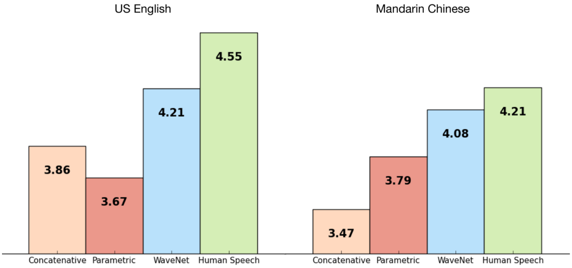

Figure: Comparison of different models for TTS(text to speech)

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

- By Christian Ledig et. al.
- Producing a High resolution Image from Low resolution is Image super resolution.
- Many implementation using deep neural network, but they suffer recovering the fine texture details.
- Using **GANs**, they use perceptual loss function which consists of an adversarial loss and a content loss.
    - **Adversarial loss**: uses natural image manifold using a discriminator network that is trained to dierentiate between the super-resolved images and original photo-realistic images.
    - **Content loss**: uses perceptual similarity instead of similarity in pixel space.

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

- they able to recover photo-realistic textures at higher resolutions [Figure: 3].
- they achieved 4x up-scaling factor.

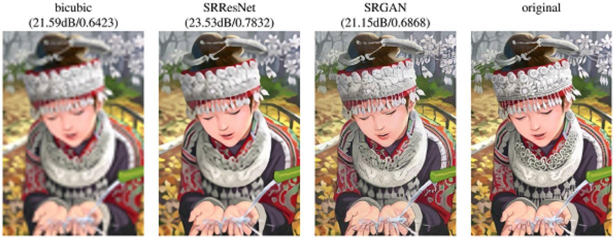

Figure: Comparison of dierent super resolution methods

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

- By Baoyang Chen et. al.
- **Video Imagination**: synthesizing imaginary videos from single static image
- Major problems faced by other algorithms:
  - High dimensionality of pixel space
  - Ambiguity of potential motions
- Applied transformations on the image in a volumetric merge network to reconstruct frames in imaginary video.
- Trained the network in an adversarial way with unsupervised learning.
- Created five-frame videos from a single image.[Figure: 4]

Video Imagination from a Single Image with Transformation Generation

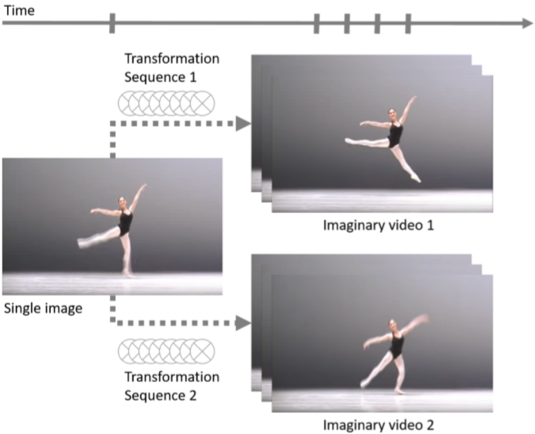

Figure: Video Imagination from a Single Image with Transformation Generation

Conclusion[GPAM+14]

- Markov chains are never needed.
- Semi-supervised learning features is obtained from the discriminator or inference net.
- Eciency improvements: training could be accelerated greatly by divising better methods for coordinating G and D.
- Understand and tackle complex problems.
- Generates model as well as additional samples based on inputs.

References I

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

[Dee16] Google DeepMind.
Wavenet: A generative model for raw audio, 2016.

[gmv09] What is the difference between a generative and discriminative algorithm?, 2009.

[GPAM+14] 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*

References II

in Neural Information Processing Systems 27, pages 2672–2680. Curran Associates, Inc., 2014.

[LTH+] 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.

[vdODZ+] 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.