# TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2020

# Generative Adversarial Networks (GANs)

# A Timeline of GAN Development

# GANs

$$\Phi^* = \operatorname*{argmax}_{\Phi} \; \min_{\Psi} \; E_{\langle i, y \rangle \sim \tilde{p}_\Phi} \; -\ln P_\Psi(i|y)$$

$$\Phi^* = \operatorname*{argmax}_{\Phi} \; \min_{\Psi} \begin{cases} E_{y \sim \text{pop}} \; -\ln P_\Psi(1|y) \\ + \; E_z \; -\ln P_\Psi(-1|y_\Phi(z)) \end{cases}$$

# Generative Adversarial Nets

# Goodfellow et al., June 2014

The rightmost column (yellow boarders) gives the nearest neighbor in the training data to the adjacent column.

# Unsupervised Representation Learning ... (DC GANS)

## Radford et al., Nov. 2015

# Unsupervised Representation Learning ... (DC GANS)

## Radford et al., Nov. 2015

# Interpolated Faces

[Ayan Chakrabarti, January 2017]

# Conditional GANS

In the conditional case we have a population distribution over pairs $\langle x, y \rangle$.

For conditional GANs we have a generator $p_\Phi(y|x)$ and a discriminator $P_\Psi(i|x,y)$ where $i = 1$ if $y$ the real "label" for $x$ and $-1$ if $y$ is generated from $x$.

$$\Phi^* = \underset{\Phi}{\mathrm{argmax}} \ \underset{\Psi}{\min} \ E_{\langle x, y, i \rangle \sim \tilde{p}_\Phi} \ -\ln P_\Psi(i|x,y)$$

# Image-to-Image Translation (Pix2Pix)
## Isola et al., Nov. 2016

We assume a corpus of "image translation pairs" such as images paired with semantic segmentations.

# Adversarial Discrimination as an Additional Loss

$$\Phi^* = \operatorname*{argmin}_{\Phi}\; E_{(x,y)\sim\text{pop}}\; ||y - y_\Phi(x)||^2 \;+\; \lambda\, \mathcal{L}_{\text{Discr}}(\Phi)$$

$$\mathcal{L}_{\text{Discr}}(\Phi) = \max_{\Psi}\; E_{x,y,i\sim\tilde{p}_\Phi}\; \ln P_\Psi(i|y,x)$$

9

# Discrimination as an Additional Loss

L1 : $\qquad \Phi^* = \text{argmin}_\Phi \ E_{(x,y)\sim\text{pop}} \ ||y - y_\Phi(x)||_1$

cGAN : $\qquad \Phi^* = \text{argmin}_\Phi \ \mathcal{L}_{\text{Discr}}(\Phi)$

L1 + cGAN : $\Phi^* = \text{argmin}_\Phi \ E_{(x,y)\sim\text{pop}} \ ||y - y_\Phi(x)||_1 \ + \ \lambda \, \mathcal{L}_{\text{Discr}}(\Phi)$

# Image-to-Image Translation (Pix2Pix)

## Isola et al., Nov. 2016

# Arial Photo to Map and Back

# Unpaired Image-to-Image Translation (Cycle GANs)

## Zhu et al., March 2017

We have two corpora of images, say images of zebras and unrelated images of horses, or photographs and unrelated paintings by Monet.

We want to construct translations between the two classes.

# Cycle Gans

# Cycle Gans

# Unsupervised Machine Translation (UMT)

## Lample et al, Oct. 2017, also Artetxe et al., Oct. 2017

In unsupervised machine translation the cycle loss is called **back-translation**.

# Feature Alignment by Discrimination
# Text to Speech (Saito et al. Sept. 2017)

Minimum Generation Error (MGE) uses <span style="color:red">perceptual distortion</span> — a distance between the feature vector of the generated sound wave and the feature vector of the original.

<span style="color:red">Perceptual Naturalness</span> can be enforced by a feature discrimination loss.

# Adversarial Discriminative Domain Adaptation

## Tzeng et al. Feb. 2017

A feature discrimination loss can be used to align source and target features.

# Progressive GANs

Progressive Growing of GANs, Karras et al., Oct. 2017

# Progressive GANs

# Early GANs on ImageNet

# BigGans

Large Scale GAN Training, Brock et al., Sept. 2018

This is a class-conditional GAN — it is conditioned on the imagenet class label.

This generates 512 X 512 images without using progressive training.

# StyleGANs

A Style-Based Generator Architecture for Generative Adversarial Networks, Karras et al., Dec. 2018

# StyleGans: Architecture

# StyleGans: Style Transfer

# StyleGans: Noise Variation

# StyleGANS2

StyleGan2 appeared in December of 2019 with significant improvements.

It was demonstrated to work on many classes of images, not just faces.

# Projecting Images into Latent Space

Given an image, can we find a noise vector (a latent vector) that generate a close approximation of the given image. Can we invert the generator?

This appears to be possible with StyleGAN2 but not with the original even though StyleGAN2 produces better images.

By measuring the match between an image $y$ and $g(g^{-1}(y))$ we can determine whether $y$ was generated by StyleGAN2.

# GANs for Pretraining

A main motivation for distribution modeling is to provide pre-trainined models that can be used in downstream tasks.

This has proved very effective in natural language processing.

To date GANs have not provided competative pretraining for downstream applications other than image generation.

# END