

TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2020

Generative Adversarial Networks (GANs)

A Timeline of GAN Development

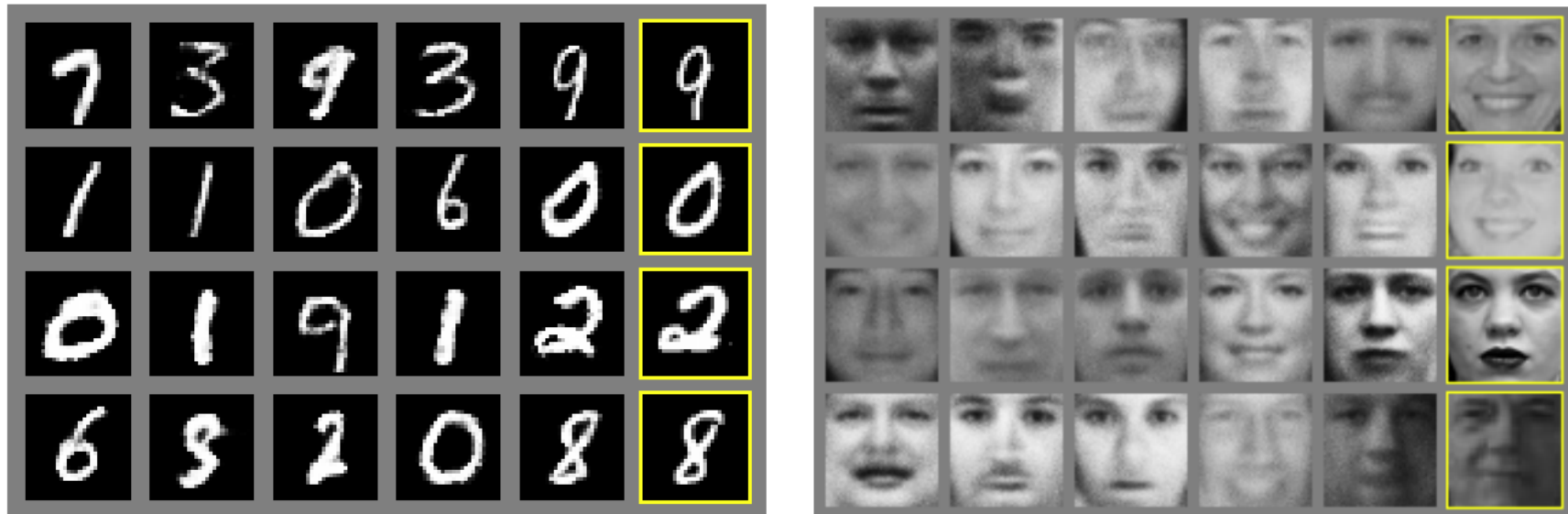
GANs

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

$$\text{gen}^* = \operatorname{argmax}_{\text{gen}} \min_{\text{disc}} \left\{ \begin{array}{l} E_{y \sim \text{pop}} - \ln P_{\text{disc}}(1|y) \\ + E_z - \ln P_{\text{disc}}(-1|y_{\text{gen}}(z)) \end{array} \right.$$

Generative Adversarial Nets

Goodfellow et al., June 2014



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

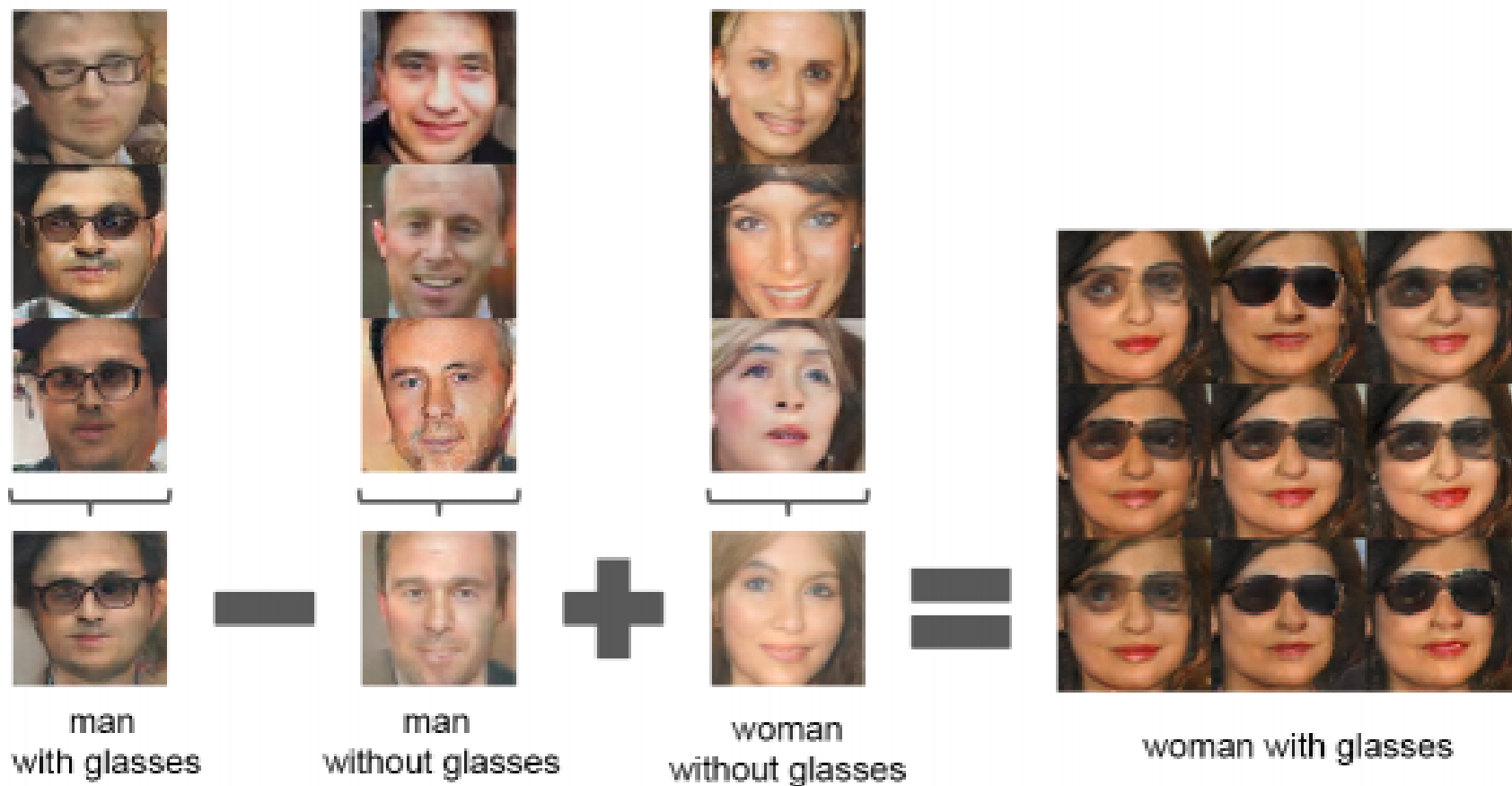
Unsupervised Representation Learning ... (DC GANS)

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 $\langle x, y \rangle$.

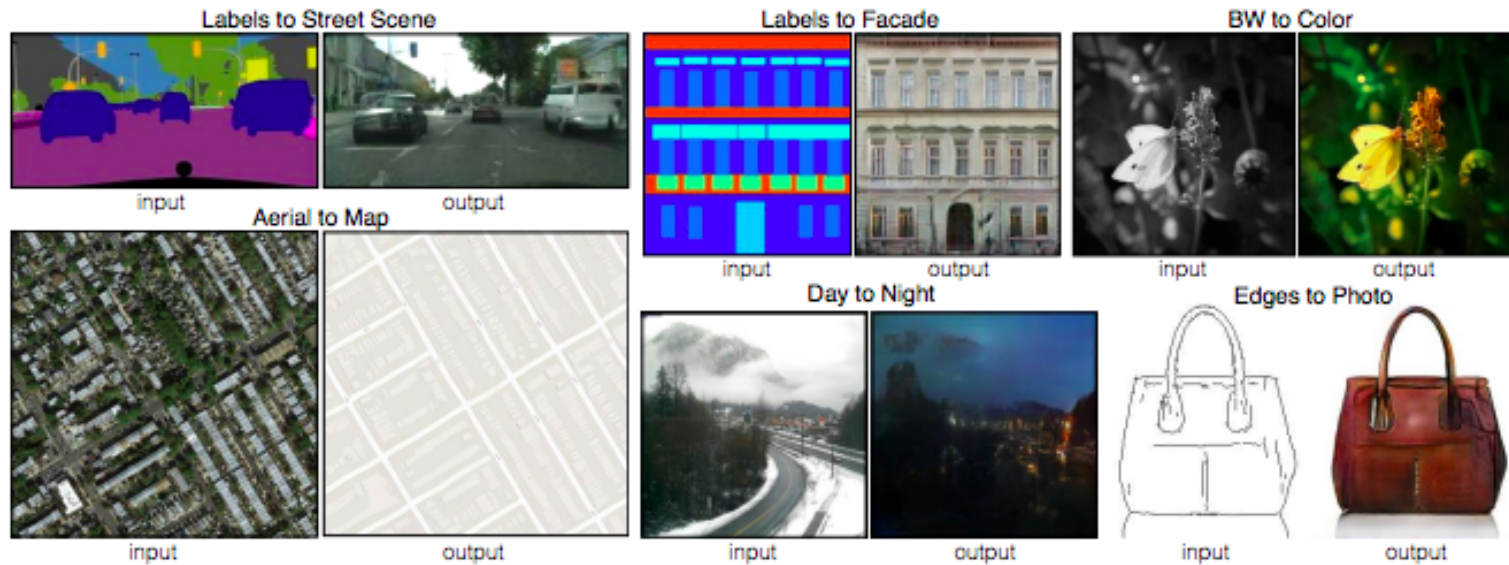
For conditional GANs we have a generator $p_{\text{gen}}(y|x)$ and a discriminator $P_{\text{disc}}(i|x, y)$ where $i = 1$ if y the real “label” for x and -1 if y is generated from x .

$$\text{gen}^* = \underset{\text{gen}}{\text{argmax}} \quad \underset{\text{disc}}{\text{min}} \quad E_{\langle x, y, i \rangle \sim \tilde{p}_{\text{gen}}} \quad - \ln P_{\text{disc}}(i|x, y)$$

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

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

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

Discrimination as an Additional Loss

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

$$\text{cGAN :} \quad \text{gen}^* = \operatorname{argmin}_{\text{gen}} \mathcal{L}_{\text{Discr}}(\text{gen})$$

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

Image-to-Image Translation (Pix2Pix)

Isola et al., Nov. 2016



Aerial Photo to Map and Back

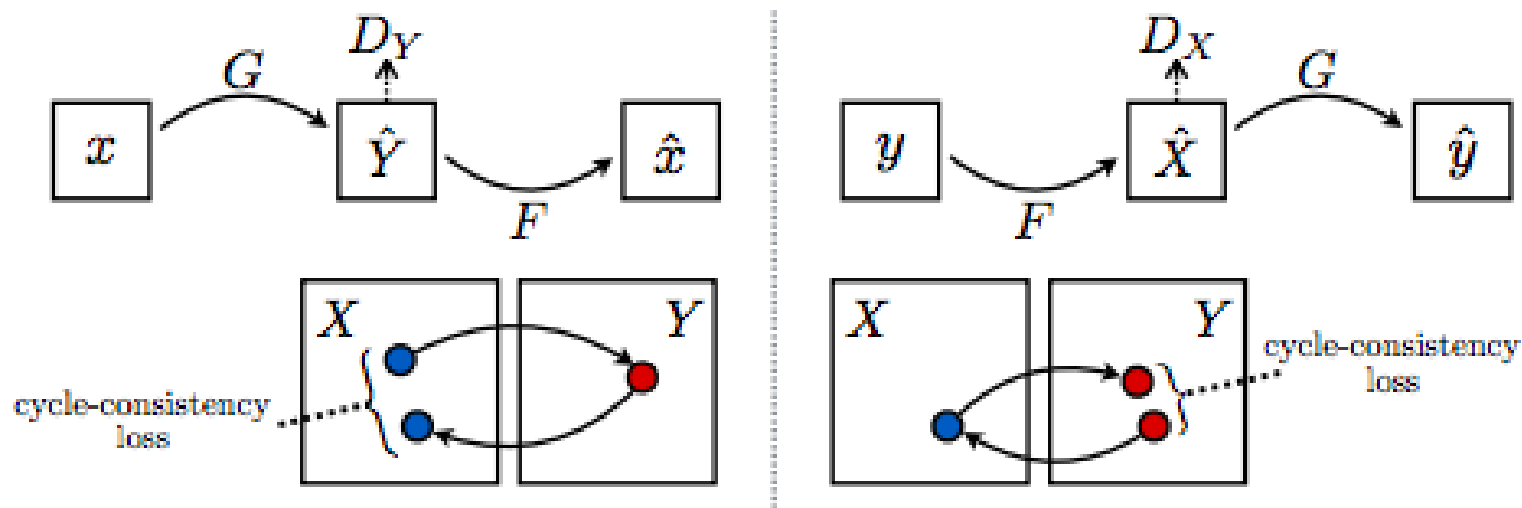


Unpaired Image-to-Image Translation (Cycle GANs)

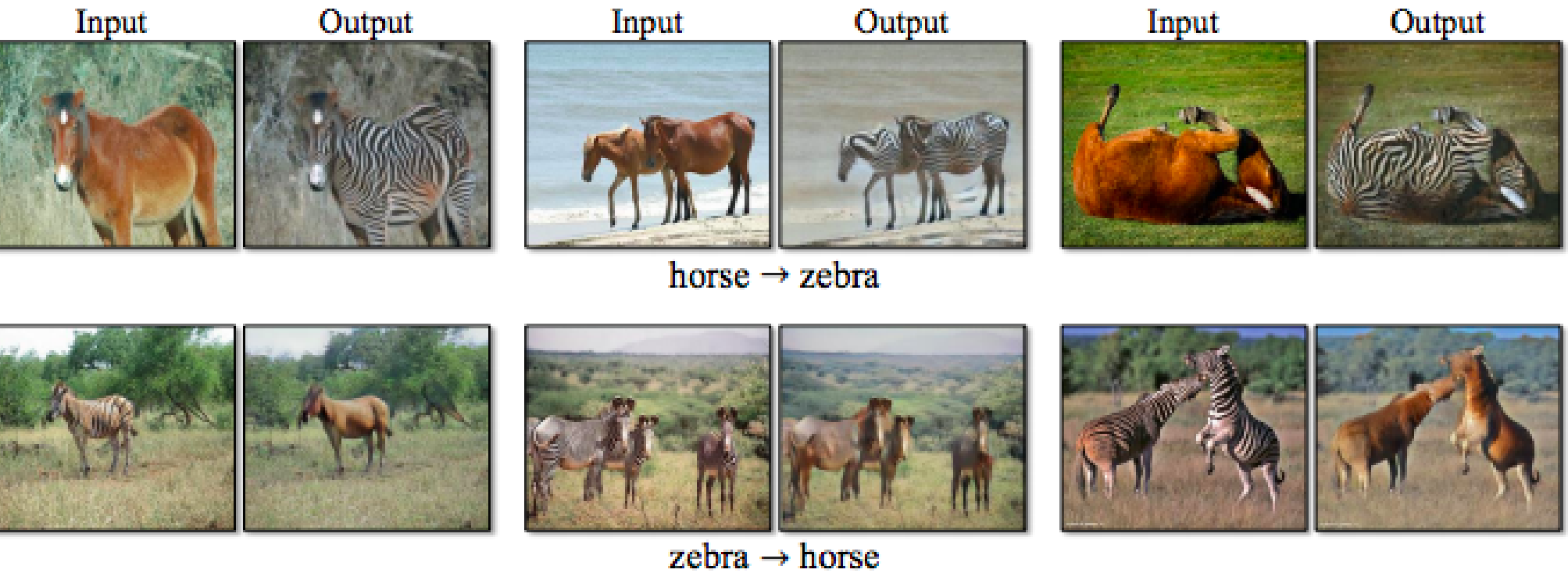
Zhu et al., March 2017

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



Horse → Zebra

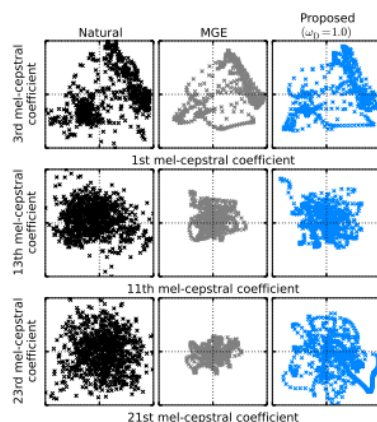
Unsupervised Machine Translation (UMT)

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)

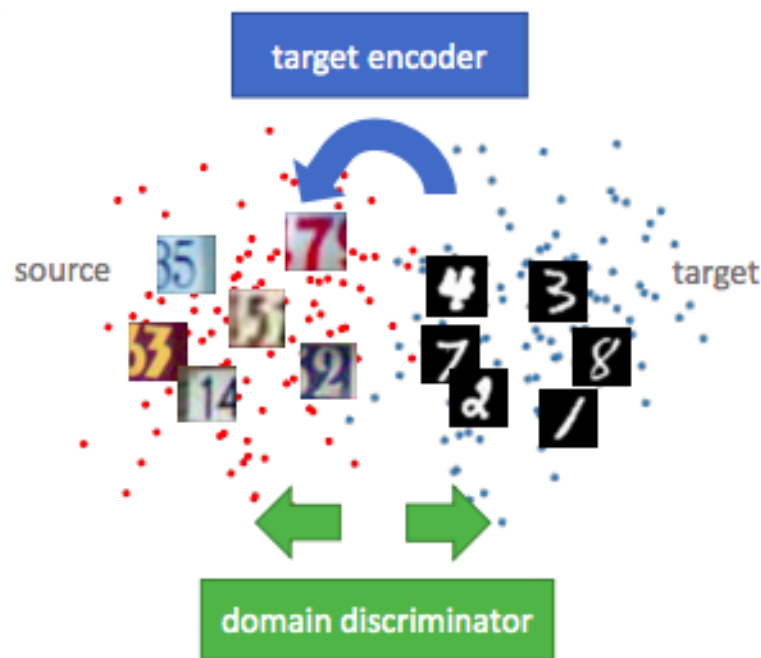


Minimum Generation Error (MGE) uses **perceptual distortion** — a distance between the feature vector of the generated sound wave and the feature vector of the original.

Perceptual Naturalness 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.

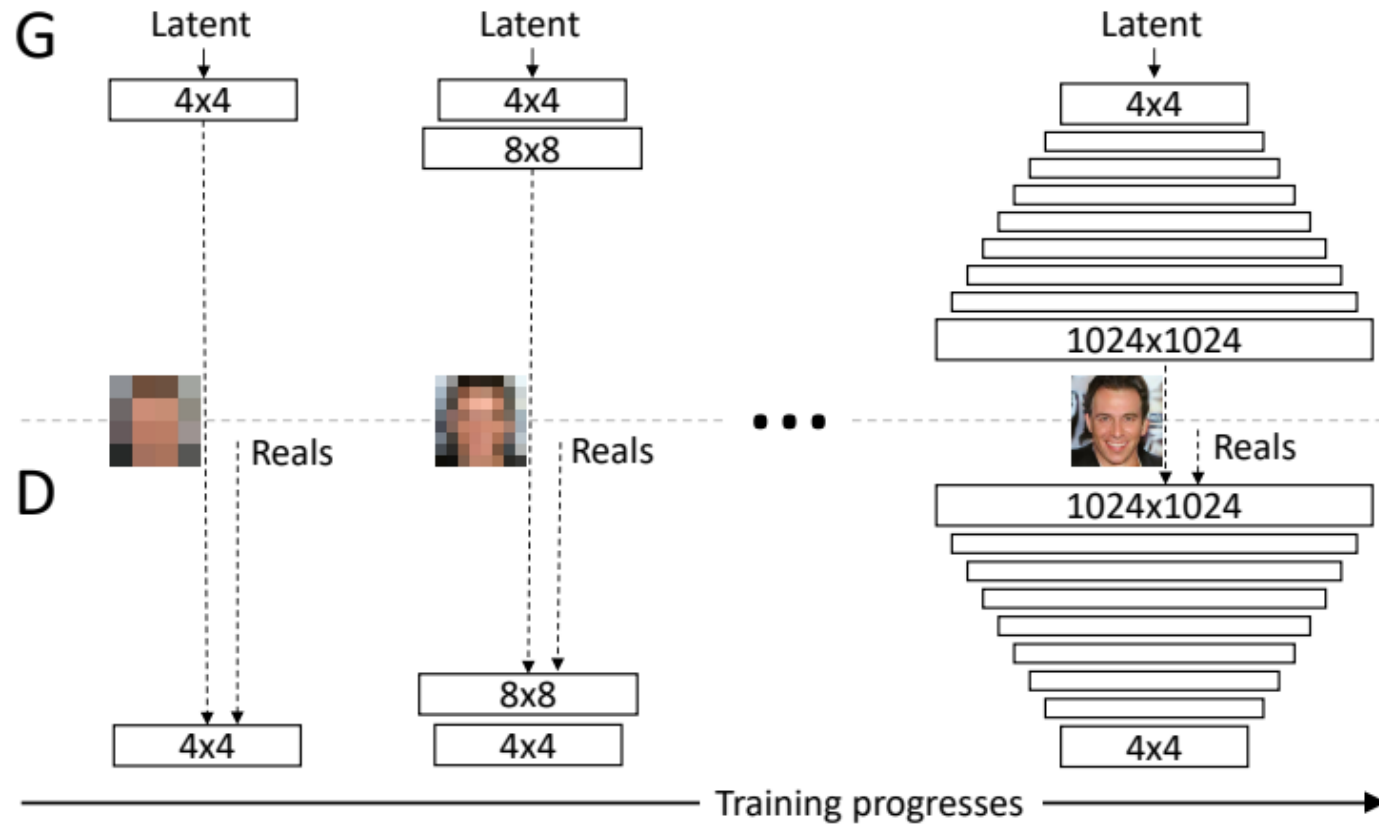
Progressive GANs

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

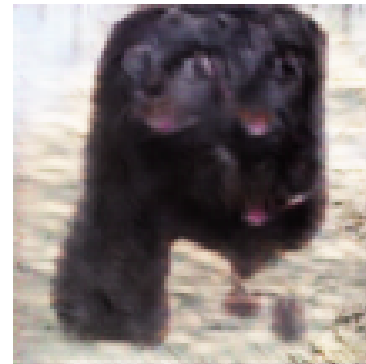
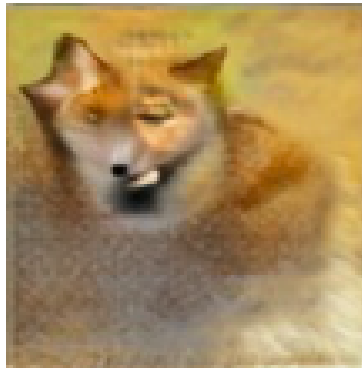
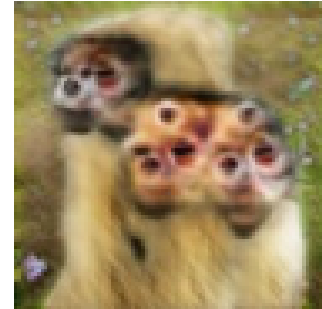
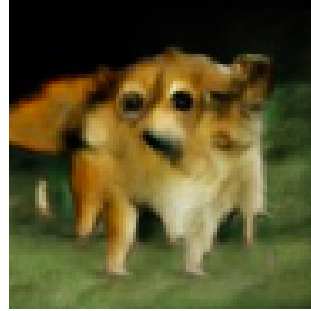
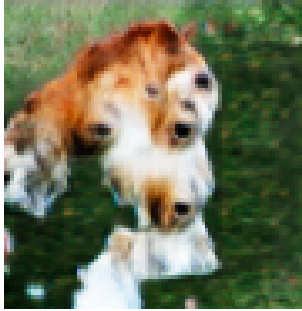


Figure 5: 1024×1024 images generated using the CELEBA-HQ dataset. See Appendix F for a larger set of results, and the accompanying video for latent space interpolations.

Progressive GANs



Early GANs on ImageNet



BigGans

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

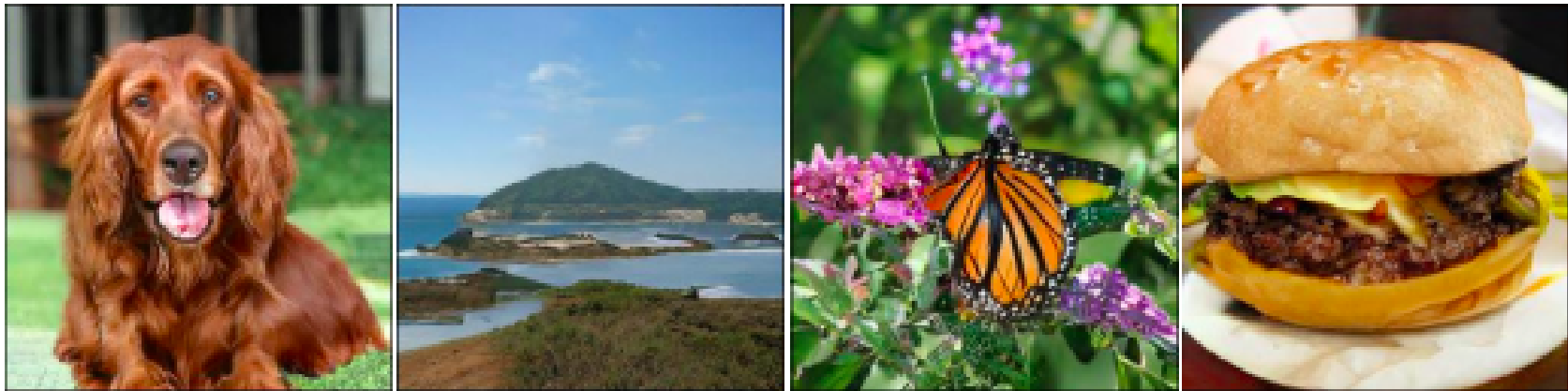


Figure 1: Class-conditional samples generated by our model.

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

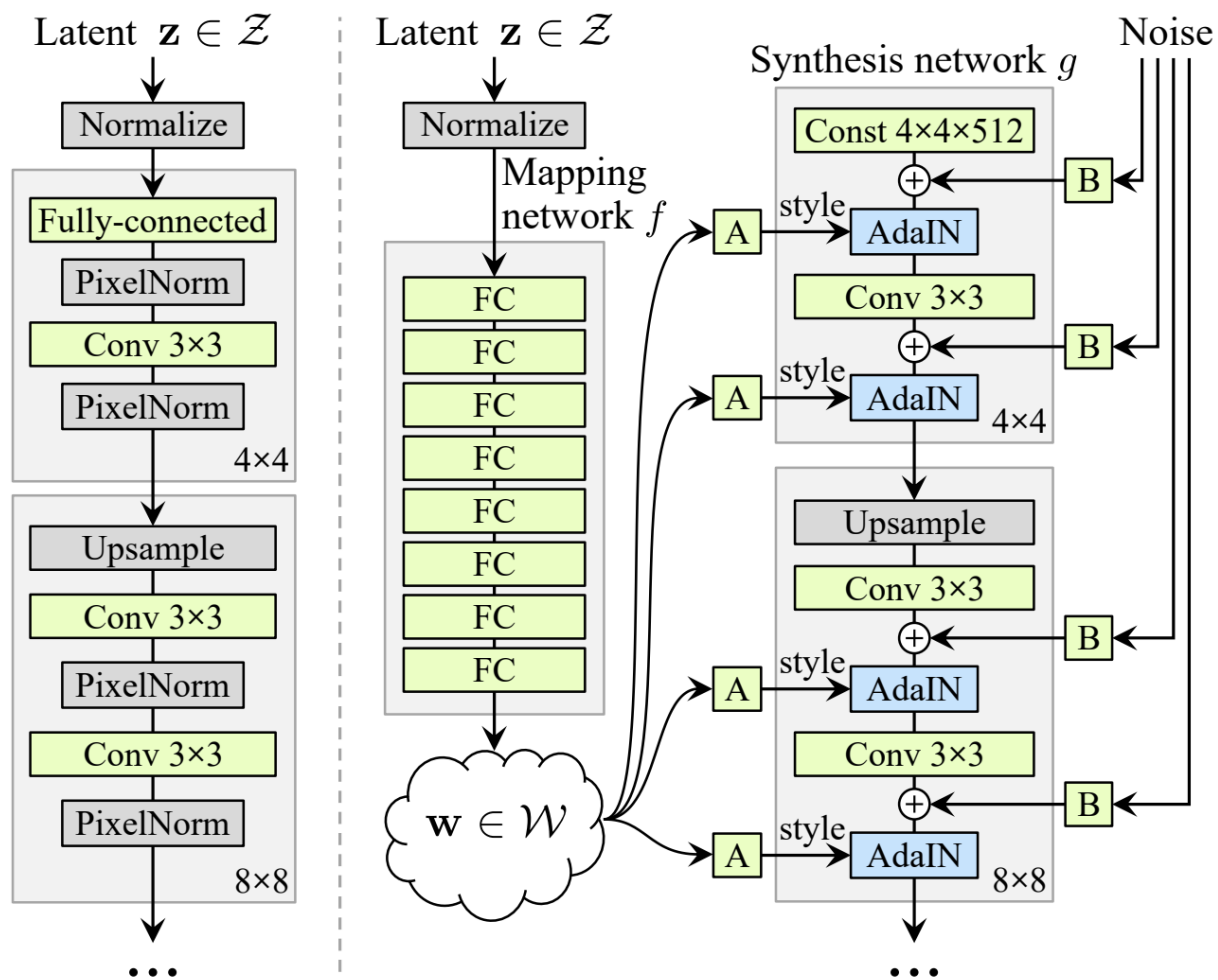
This generates 512 X 512 images without using progressive training.

StyleGANs

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



StyleGans: Architecture



StyleGans: Style Transfer



StyleGans: Noise Variation



StyleGAN2

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-trained models that can be used in downstream tasks.

This has proved very effective in natural language processing.

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

END