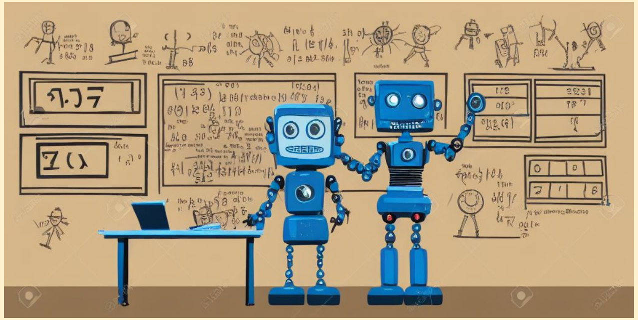


Advanced Workshop on Machine Learning

Lecture 4: Generative Adversarial Networks

Danylo Vashchilenko

2022

Agenda

Part 1

1. Generator Task
2. Discriminator Task
3. Dataset: FFHQ
4. Case Study: StyleGAN

Part 2

5. Assignment 4: FakeNet
6. Case Study: Stable Diffusion
7. Case Study: CLIP
8. Case Study: SRGAN

Part 1

Generator and Discriminator Tasks

(Image by a Stable Diffusion model)

Generator Task

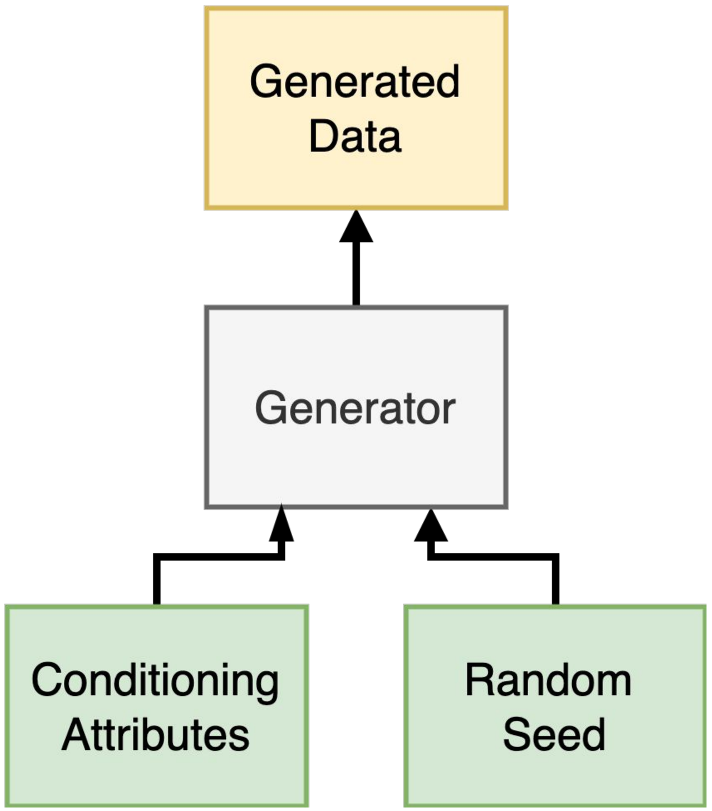
Generator learns a data distribution from the training dataset.

Generator is deterministic.

Attributes:

- Classes
- Text
- Images

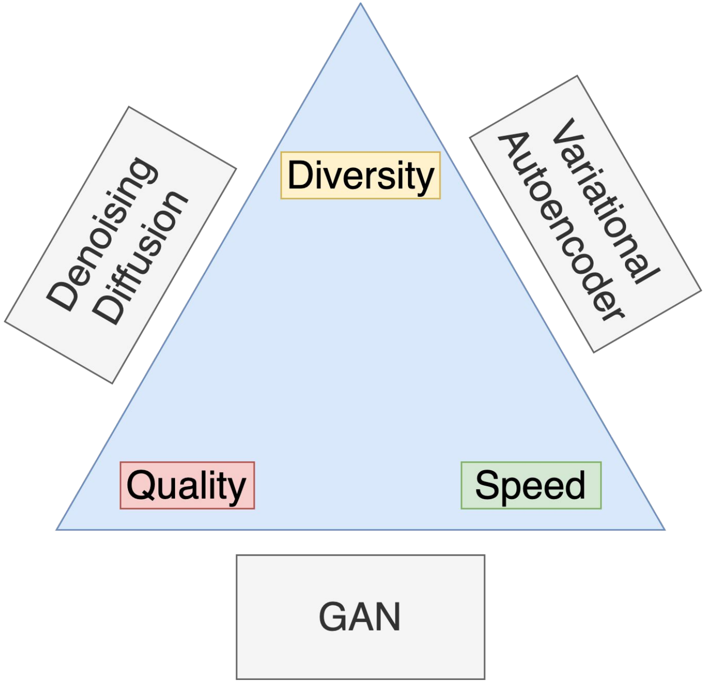
In this lecture: image examples.



Generator Architectures

Trade-offs:

- GANs lack diversity
- Diffusion Models are slow
- VAEs lack quality



Xiao, Z., et al. (2022)

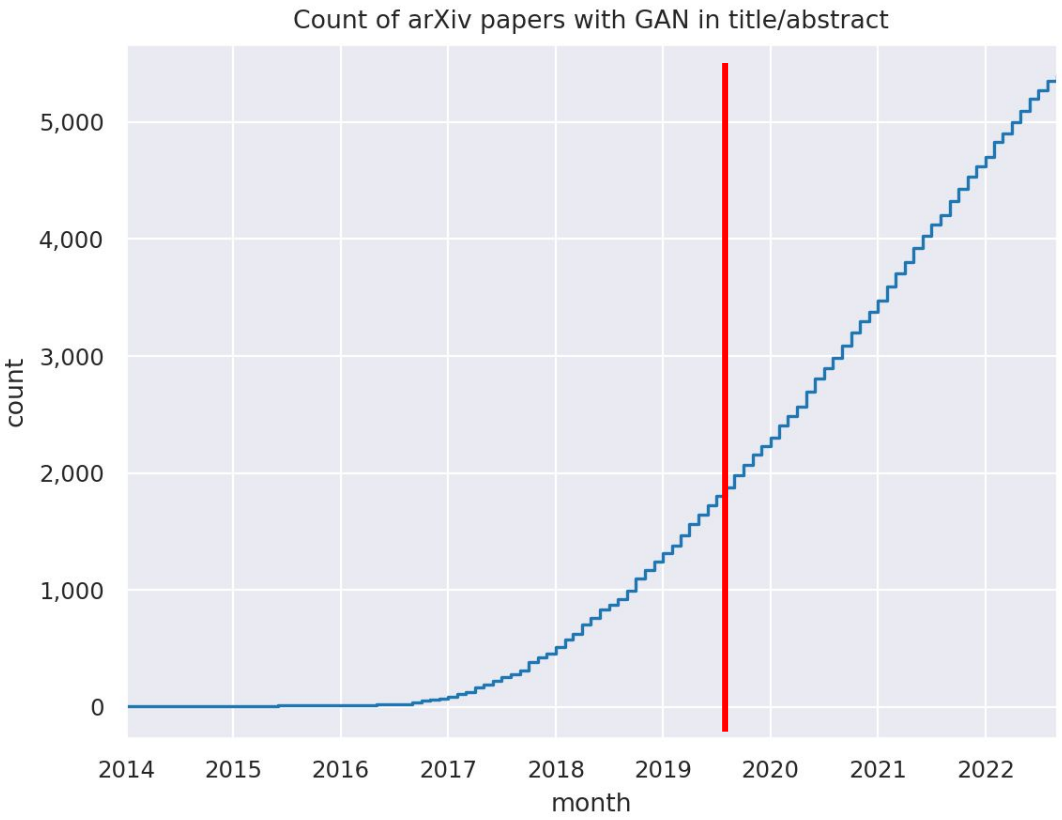
GAN Zoo

Papers on arxiv.org

Filtered by:

- Title mentions GAN
- Abstract mentions GAN

1200 new papers in 2021
5000+ papers in total

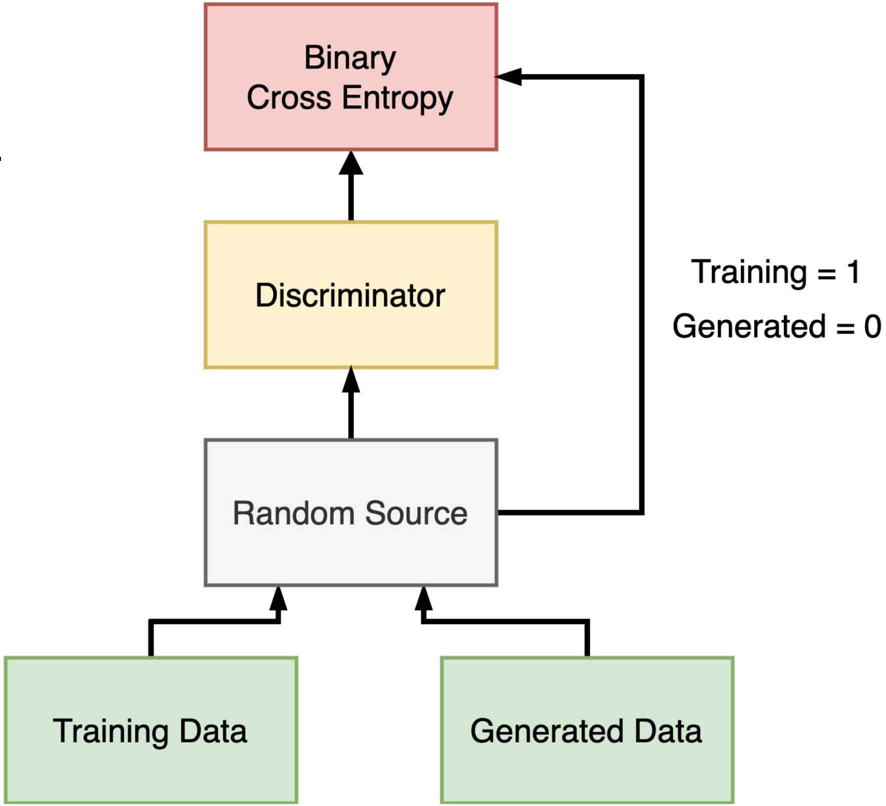


See: GANArxiv.ipynb

Discriminator Task

True label depends on the source of the input data.

Random Source block is omitted in future slides.



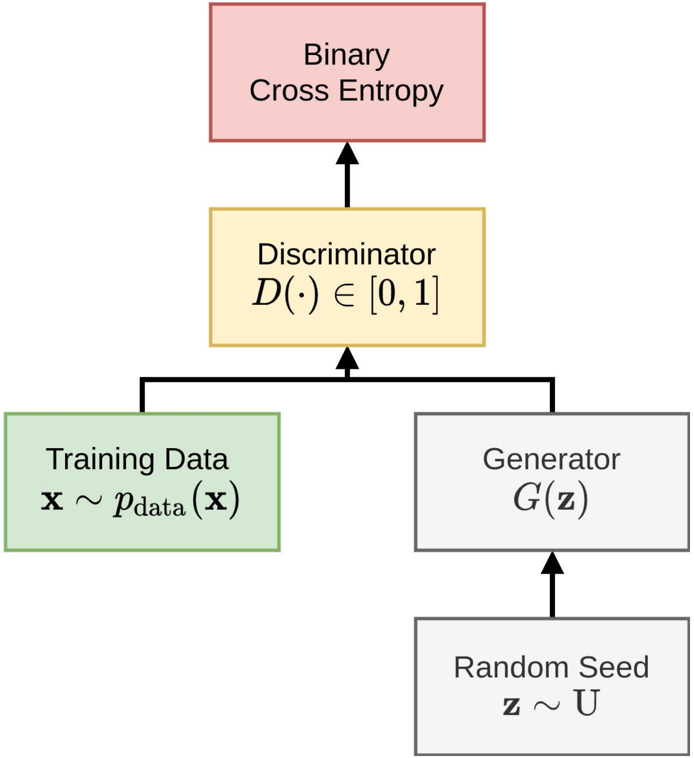
Goodfellow, I., et al. (2014)

Discriminator Training

Discriminator given a batch of examples from training data and the generator's output.

Cost is the average binary cross-entropy across the batch:

$$C_D = -\frac{1}{m}\sum_{i=1}^{m}\left[\log D(\mathbf{x}^{(i)}) + \log\left(1 - D(G(\mathbf{z}^{(i)}))\right)\right]$$



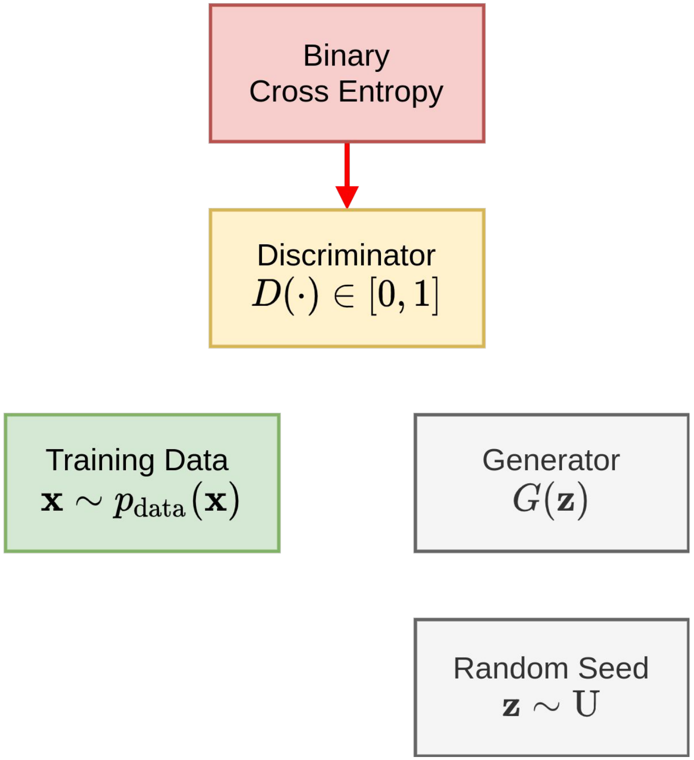
Goodfellow, I., et al. (2014)

Discriminator Training

Discriminator given a batch of examples from training data and the generator's output.

Cost is the average binary cross-entropy across the batch:

$$C_D = -\frac{1}{m}\sum_{i=1}^{m}\left[\,\log D(\mathbf{x}^{(i)}) + \log\left(1 - D(G(\mathbf{z}^{(i)}))\right)\,\right]$$



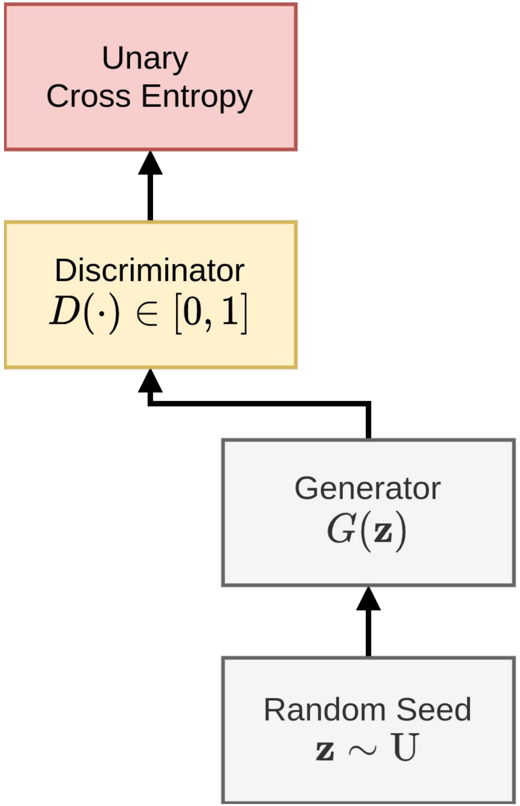
Goodfellow, I., et al. (2014)

Generator Training

The discriminator is given a batch from the generator, so the label = 0.

The generator's cost goes down when the discriminator is wrong:

$$C_G = -\frac{1}{m} \sum_{i=1}^{m} \left[\log \left(D(G(\mathbf{z}^{(i)})) \right) \right]$$



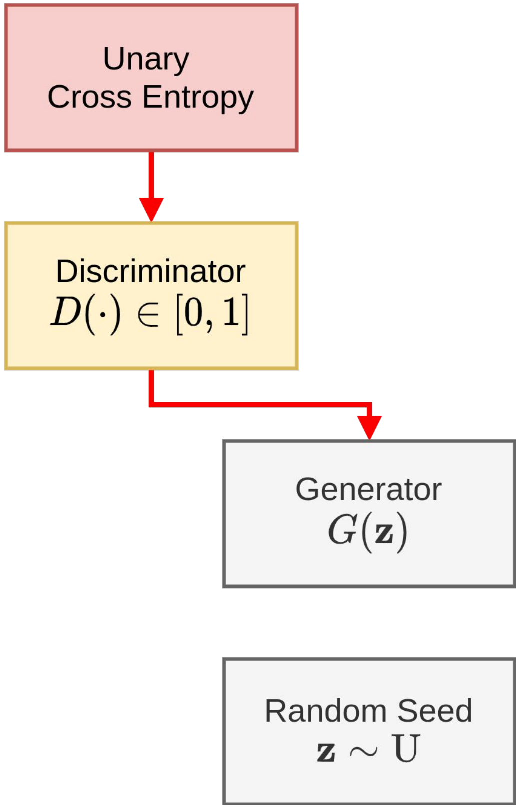
Goodfellow, I., et al. (2014)

Generator Training

The discriminator is given a batch from the generator, so the label = 0.

The generator's cost goes down when the discriminator is wrong:

$$C_G = -\frac{1}{m}\sum_{i=1}^{m}\left[\,\log\left(D(G(\mathbf{z}^{(i)}))\right)\right]$$



Goodfellow, I., et al. (2014)

Training Procedure

1. Sample a batch of training and generated data
2. Train the discriminator for several steps

3. Sample a batch of noise
4. Train the generator

5. Repeat steps 1-4 until convergence.

Training one component at a time

Goodfellow, I., et al. (2014)

Combined Training

$$V(D,G) = \underbrace{\mathbb{E}_{\mathbf{x}}\left[\log D(\mathbf{x})\right]}_{\text{Rewards true positives}} + \underbrace{\mathbb{E}_{\mathbf{z}}\left[\log(1 - D(G(\mathbf{z})))\right]}_{\text{Rewards true negatives}}$$

$$\min_{G}\max_{D} V(D,G)$$

$$G(\mathbf{z}) \rightarrow p_{\text{data}}(\mathbf{x})$$

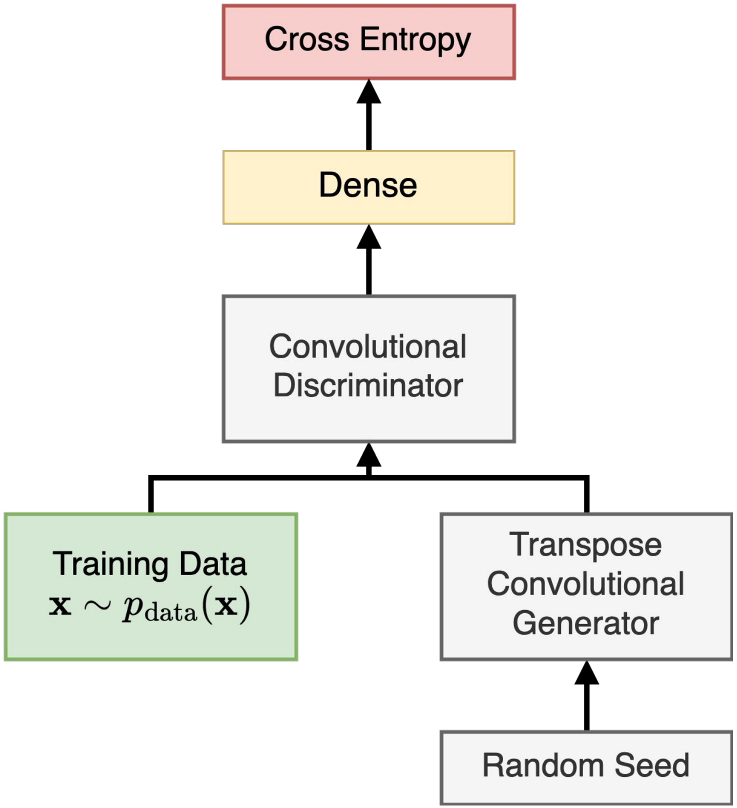
$$D(G(\mathbf{z})) \rightarrow 0.5$$

Goodfellow, I., et al. (2014)

Case Study: Deep Convolutional GAN

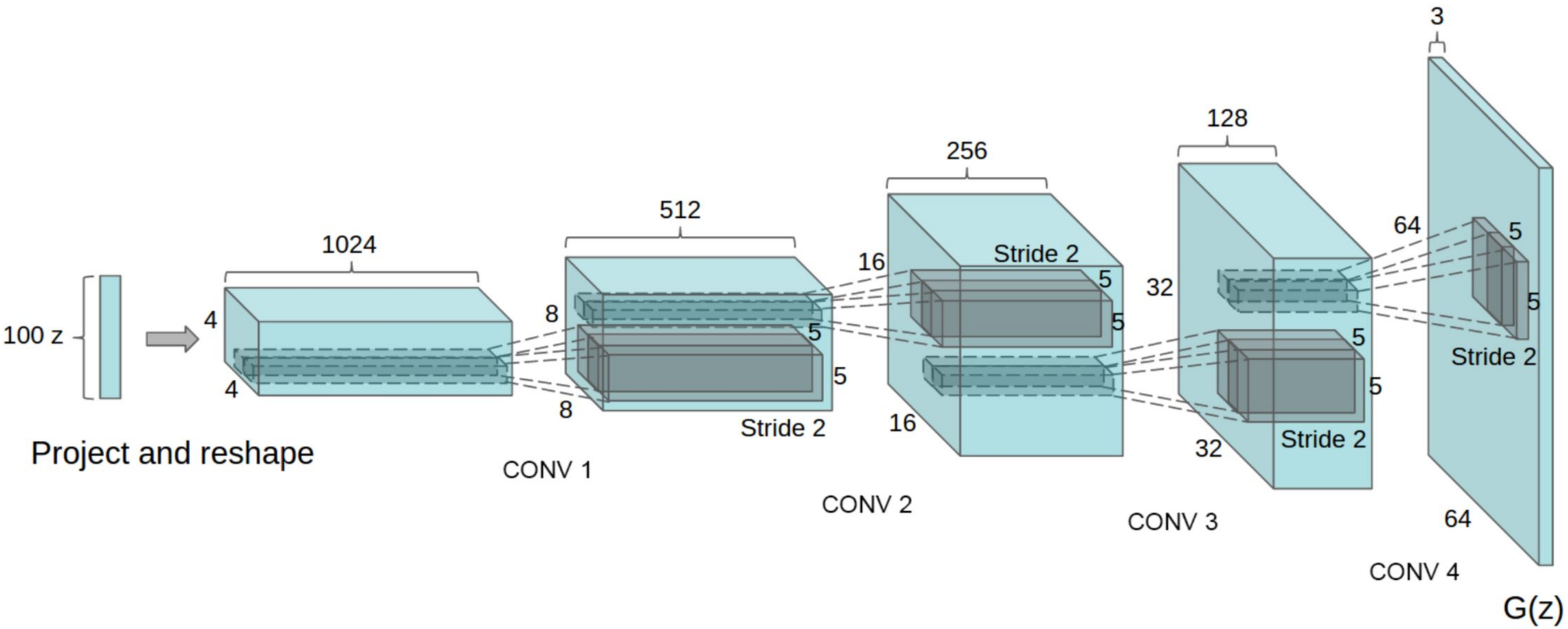
Discriminator
Regular convolutions

Generator
Transpose convolutions



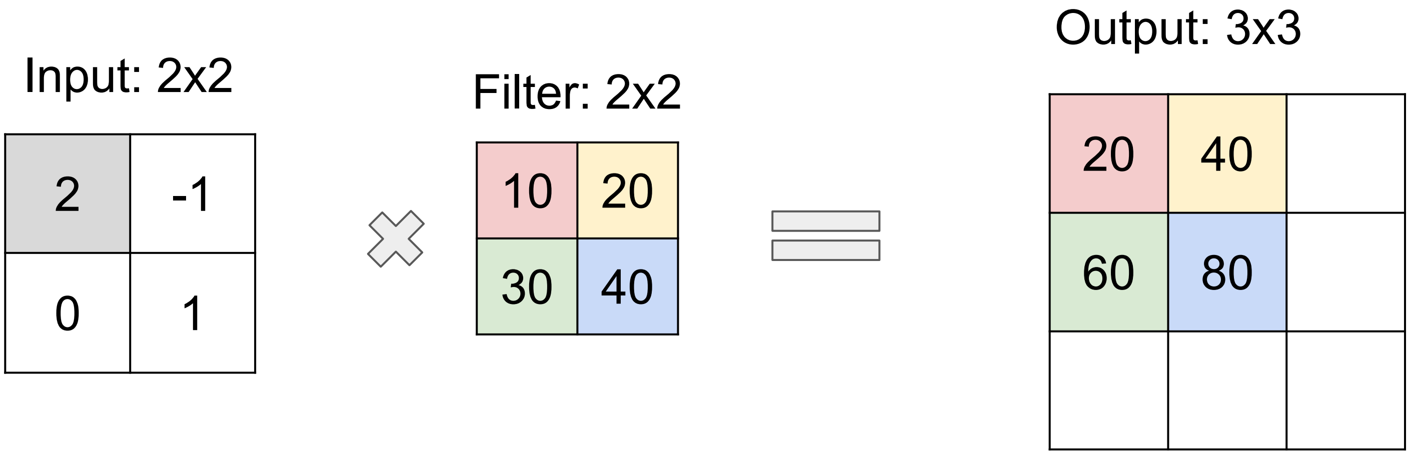
Radford, A., et al. (2016)

DCGAN: Generator

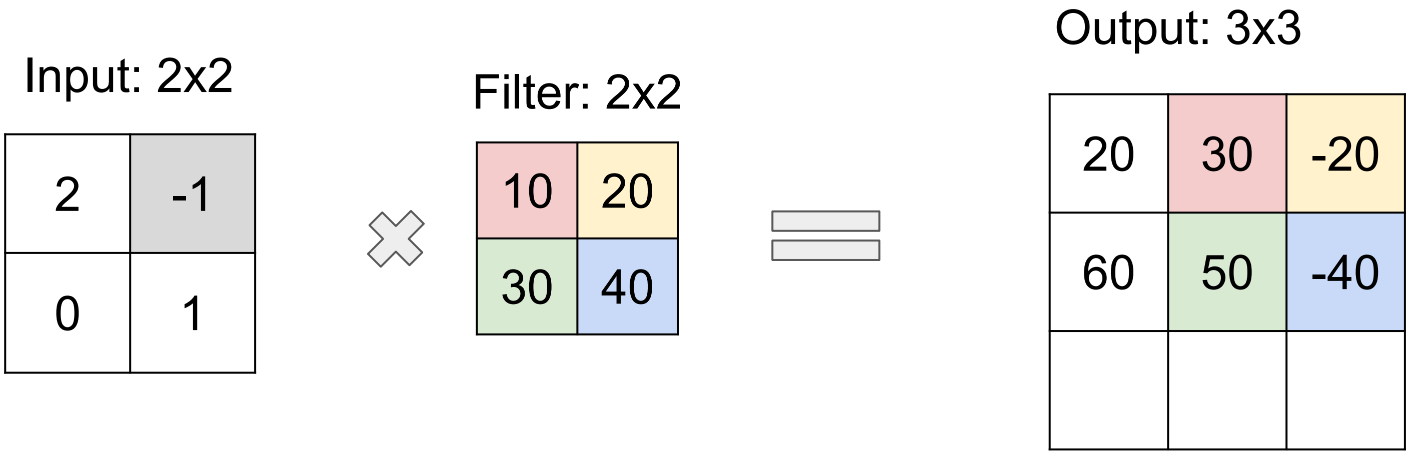


Radford, A., et al. (2016)

Transpose Convolution



Transpose Convolution



Transpose Convolution

Input: 2x2

2	-1
0	1

Filter: 2x2

10	20
30	40

Output: 3x3

20	30	-20
70	70	-40
0	0	

Transpose Convolution

Input: 2x2

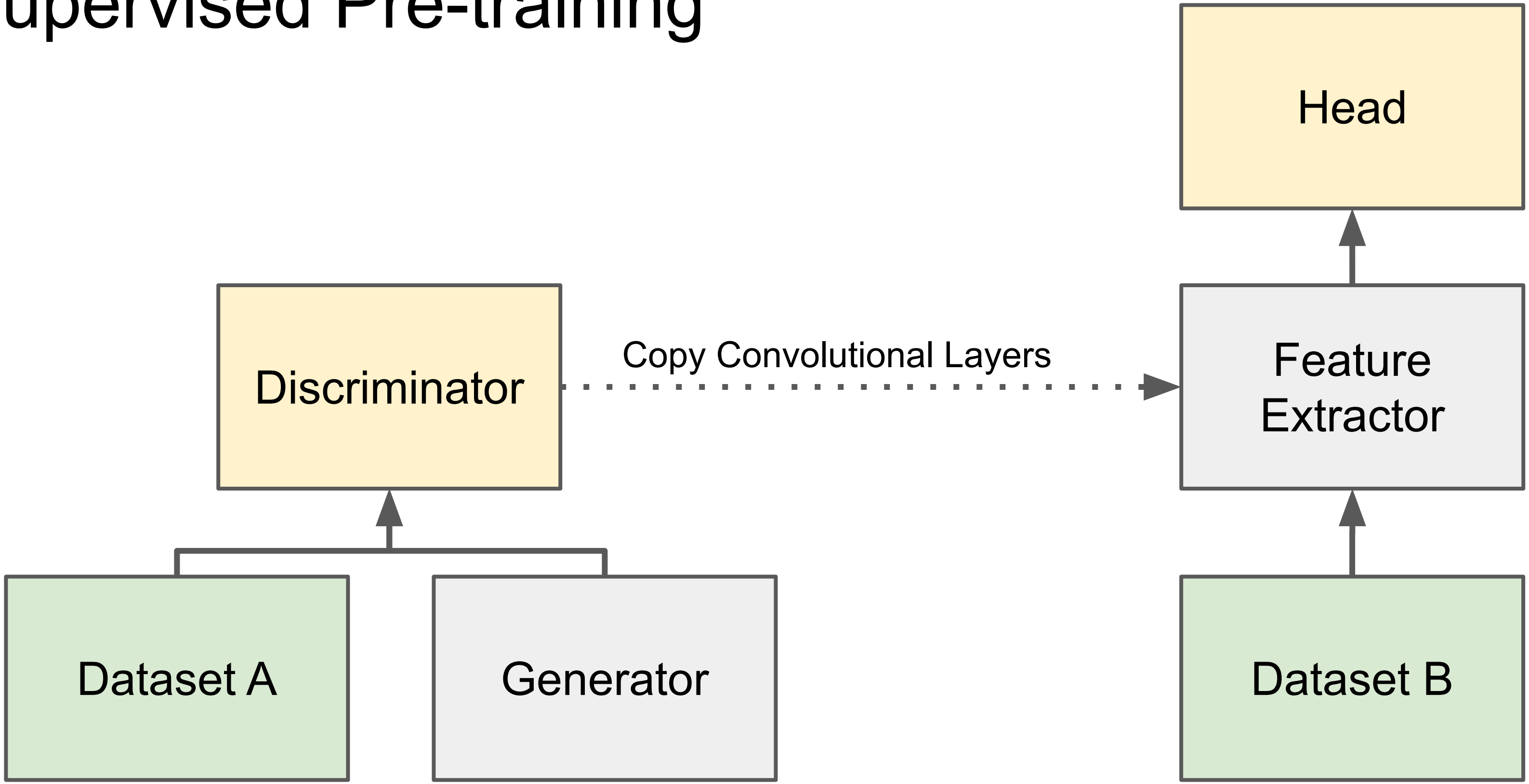
Filter: 2x2

Output: 3x3

$$\begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 10 & 20 \\ 30 & 40 \end{bmatrix} = \begin{bmatrix} 20 & 30 & -20 \\ 70 & 80 & -30 \\ 0 & 30 & 40 \end{bmatrix}$$

Also called fractionally-strided convolution.
Linear algebra details: https://arxiv.org/pdf/1603.07285.pdf

Unsupervised Pre-training



Radford, A., et al. (2016)

DCGAN: Latent Space

Interpolation between **seed vectors z** of the first and the last image.

Radford, A., et al. (2016)

https://thisxdoesnotexist.com/

Selected Models and Examples

BigGAN - Brock, A., et al. (2019)

StyleGAN2 - Karras et al (2019)

Ryan Hover

https://arthurfindelair.com/thisnightskydoesnotexist/

Dataset: Flickr-Faces HQ (FFHQ)

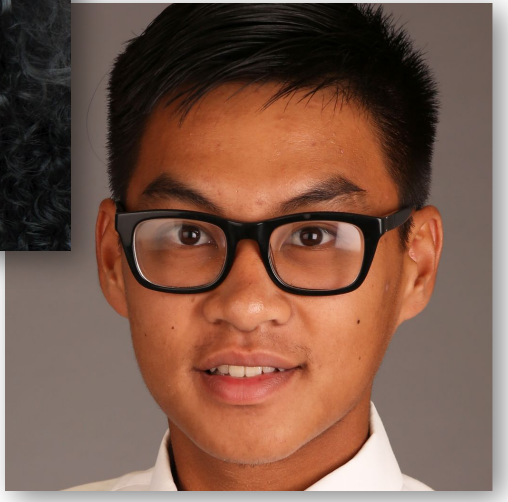
Dimensions

70K images

1024 x 1024

Github

https://github.com/NVlabs/ffhq-dataset



Source: Karras, T., et al. (2018)

NVIDIA Research Lab

Models

ProgressiveGAN, 2018

StyleGAN, 2019

StyleGAN V2, 2020

StyleGAN V2 - ADA, 2020

StyleGAN V3, 2021

Datasets

Flicker Faces HQ

CelebA HQ

Case Study: StyleGAN

NVIDIA Research Labs, 2020

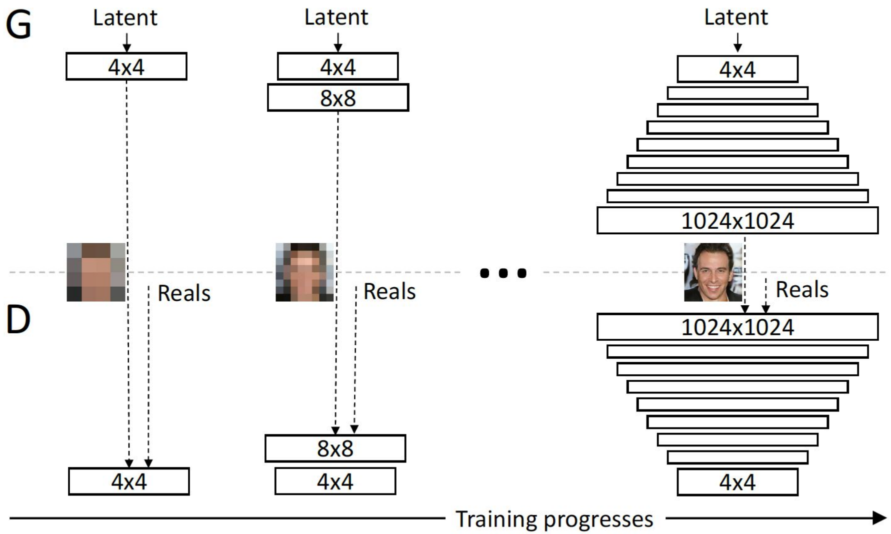
Available pre-trained on

Faces: FFHQ

Animals: AFHQ

CIFAR10

https://github.com/NVlabs/stylegan3



(Karras, T., et al., 2018)

Source: Training Generative Adversarial Networks with Limited Data (Karras, T., et al., 2020)

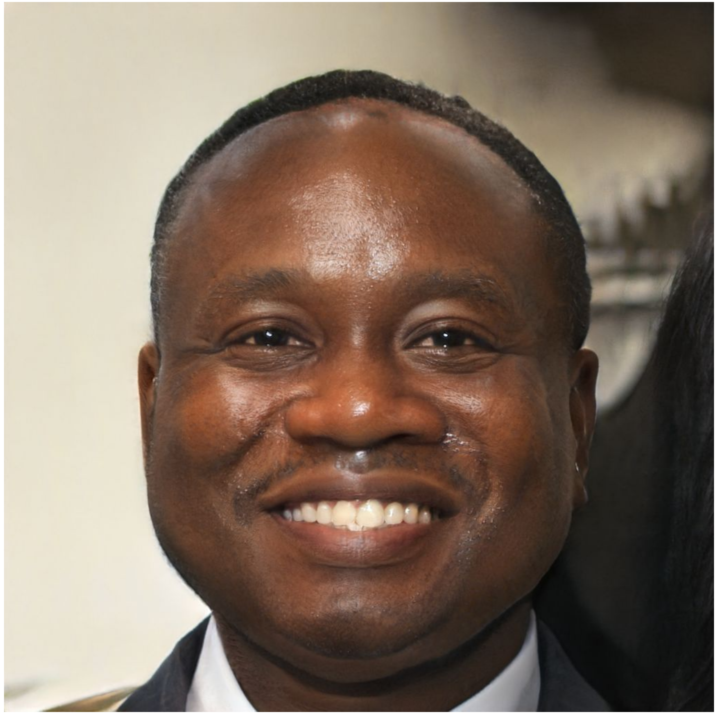
StyleGAN: Faces

Source: Training Generative Adversarial Networks with Limited Data (Karras, T., et al., 2020)

StyleGAN2: Faces

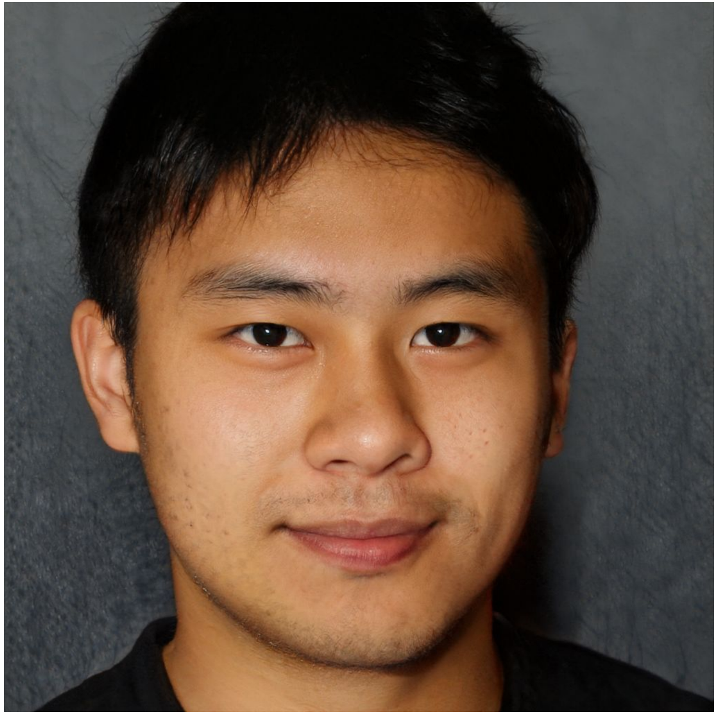
Source: Training Generative Adversarial Networks with Limited Data (Karras, T., et al., 2020)

StyleGAN2: Faces



Source: Training Generative Adversarial Networks with Limited Data (Karras, T., et al., 2020)

StyleGAN2: Faces



Source: Training Generative Adversarial Networks with Limited Data (Karras, T., et al., 2020)

StyleGAN2: Wildlife



https://nvlabs-fi-cdn.nvidia.com/stylegan2-ada/videos/interpolations-afhqwild.mp4

Source: Training Generative Adversarial Networks with Limited Data (Karras, T., et al., 2020)

Part 2

FakeNet, Diffusion, Contrastive Learning

(Image by a Stable Diffusion model)

Dataset: FakeNet

Dimensions

224 x 224 x 3

~12K triplets

Triplets

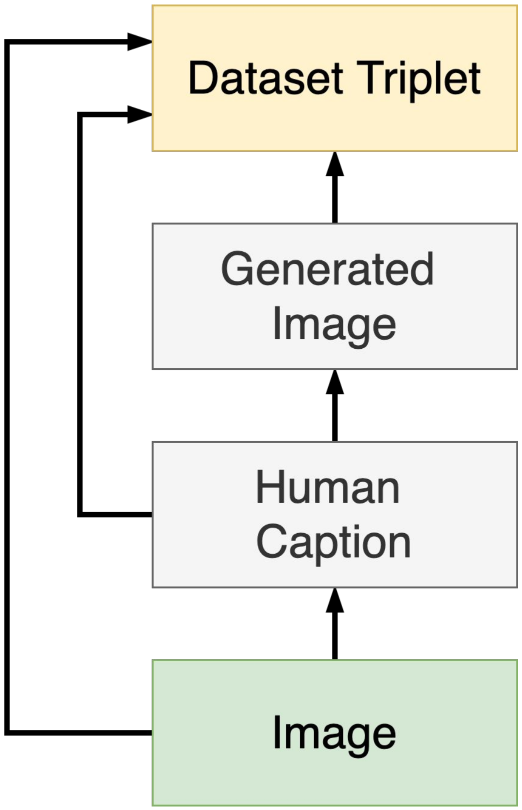
Original image

English caption by human

Generated image from caption

Task

Determine which 1 of 2 images is generated



FakeNet: Examples

A pile of oranges sitting on top of each other.

Original

Generated

FakeNet: Examples

A close of a small bird perched on a tree limb.

Original

Generated

FakeNet: Examples

The Big Ben clock tower sitting under a dark sky.

Original

Generated

FakeNet: Examples

Small bathroom with red curtains on the outside of the shower.

Original

Generated

FakeNet: Examples

Four brown bears are walking through a stream.

Original

Generated

FakeNet: Examples

The cat is sleeping underneath the clock.

Original

Generated

FakeNet: Examples

a man plays baseball on a field with grass.

Original

Generated

Baseline Models

Baseline Description
Very simple architectures
Trained to convergence
No regularization
No fine-tuning

Other considerations
Development time
Training time
Parameter count

Text conditioning?
Pre-training?

Accuracy = 0.70 is a reasonable baseline.

		binary accuracy (t=0.5) on score
backbone	**input_type**	
vanilla cnn	**single image**	0.9398
	paired images	0.8967
mobile net	**paired images**	0.7532
	single image	0.7427

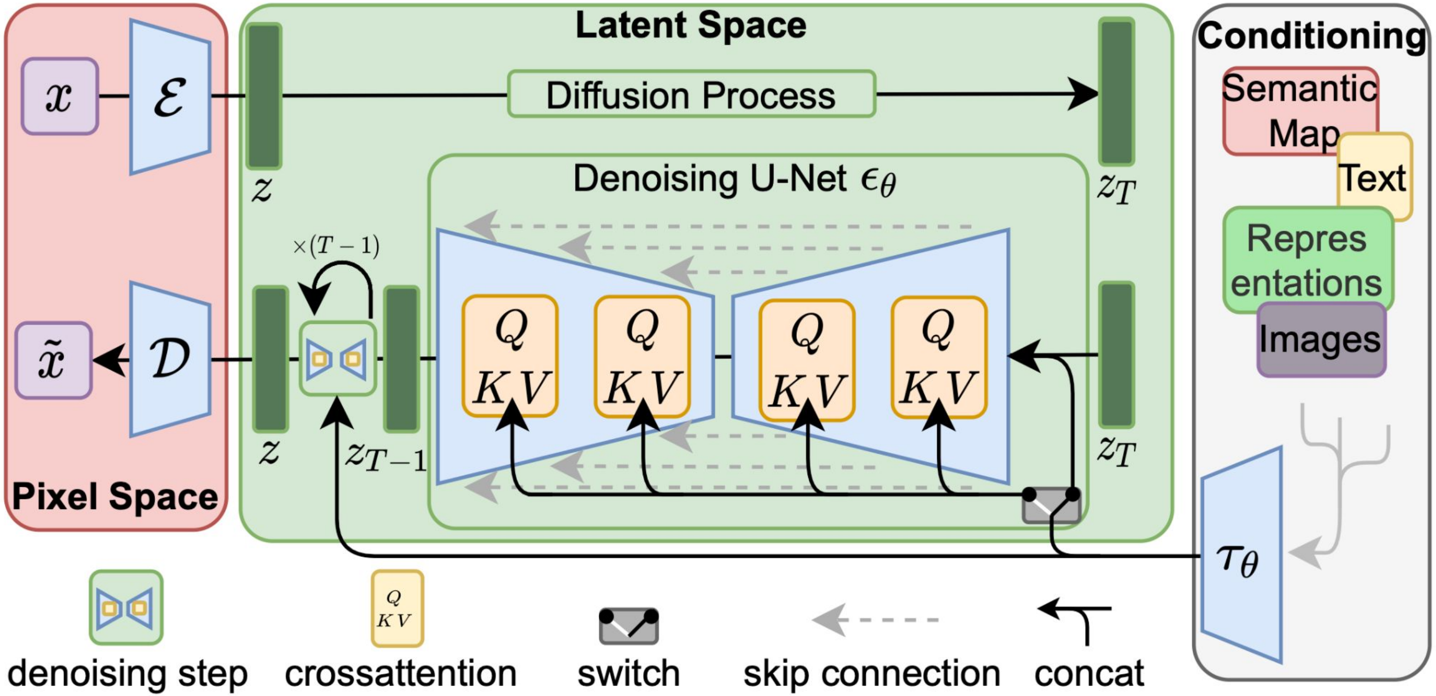
Demo

AssignmentTemplate.ipynb

Assignment 4

1. Task: discriminate between real and fake images
    a. Probability of image A being real
2. Grading:
    a. 10 points if Accuracy (t = 0.5) on Score > 0.70
    b. 10 points proportional to the percentile in class
3. Deliverables:
    a. Jupyter notebook
    b. Saved model definition
    c. Saved model parameters
    d. Predicted labels for the Score segment
4. Due by EOD 10/29

Stable Diffusion



Demo

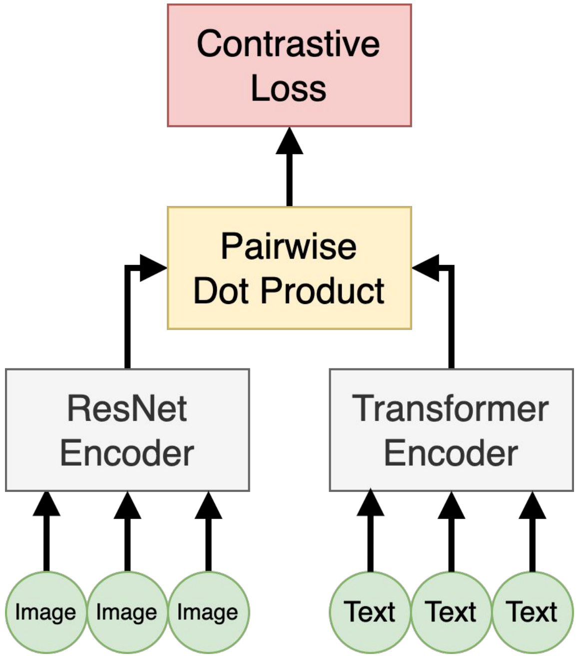
StableDiffusion.ipynb

CLIP: Contrastive Image-Language Pre-training

Dataset:
(Image + English caption)
400 million pairs
scraped from Internet

Model
Encoders create embeddings
Contrastive loss:

- (1) rewards similarity of matching pairs
- (2) penalizes similarity of mismatching pairs



Source: Learning Transferable Visual Models From Natural Language Supervision (Radford, A., et al., 2021)

CLIP: Contrastive Image-Language Pre-training

Model

~400M parameters

- ~100M in text encoder
- ~300M in image encoder

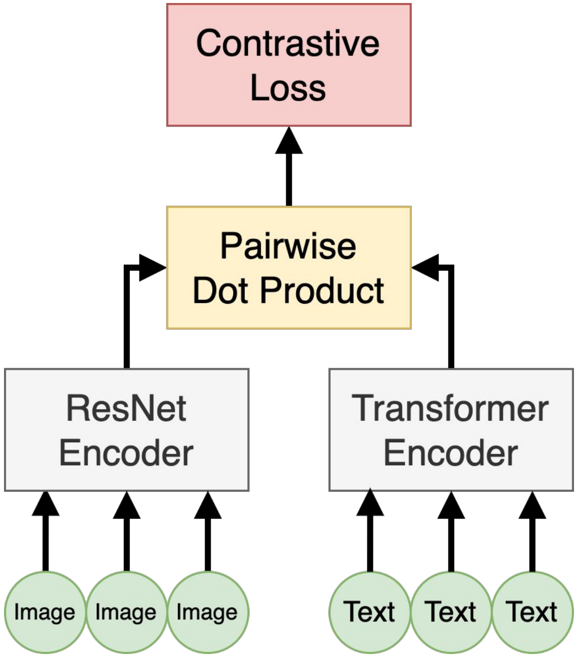
Training

Nvidia V100

256 GPUs

12 days

With AWS prices: ~$90K



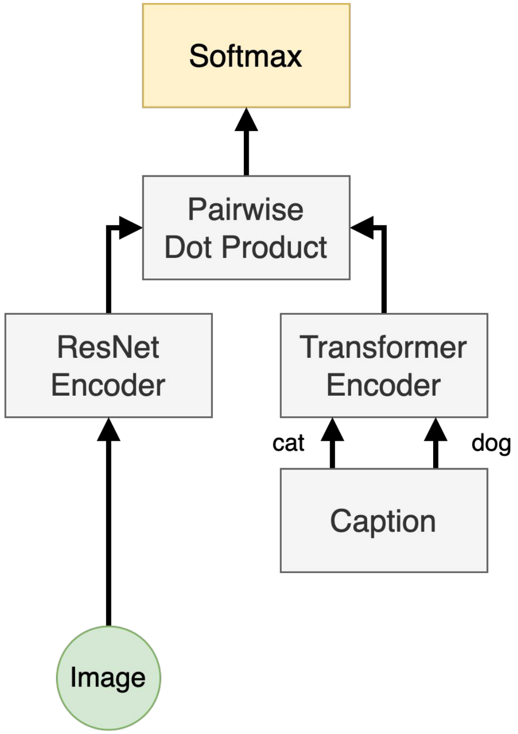
Source: Learning Transferable Visual Models From Natural Language Supervision (Radford, A., et al., 2021)

CLIP: Zero-shot Prediction

Example: image classification

1. Create per-class English captions
2. Pass similarity scores to softmax

No training!

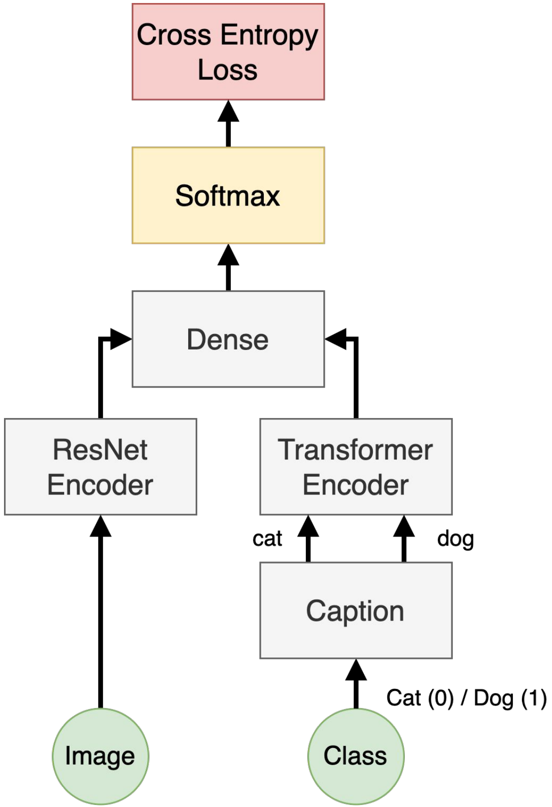


Source: Learning Transferable Visual Models From Natural Language Supervision (Radford, A., et al., 2021)

CLIP: Transfer Learning

Example: image classification

1. Copy the encoder's weights
2. Create per-class English captions
3. Use embeddings as features
4. Optionally, fine-tune the encoders

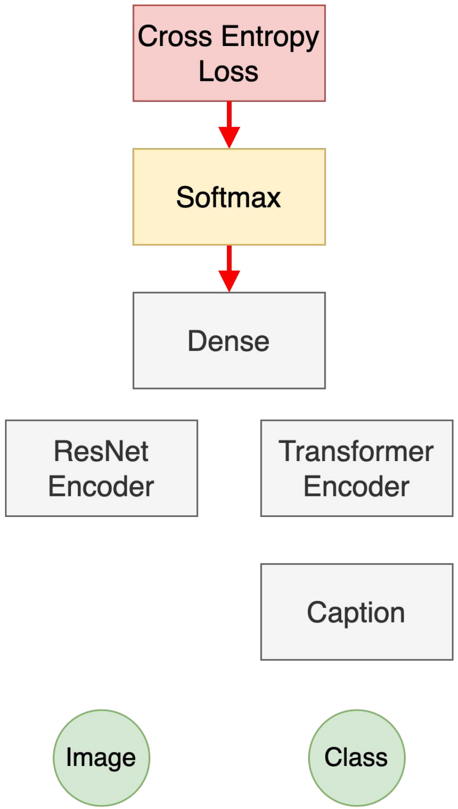


Source: Learning Transferable Visual Models From Natural Language Supervision (Radford, A., et al., 2021)

CLIP: Transfer Learning

Example: image classification

1. Copy the encoder's weights
2. Create per-class English captions
3. Use embeddings as features
4. Optionally, fine-tune the encoders



Source: Learning Transferable Visual Models From Natural Language Supervision (Radford, A., et al., 2021)

Demo

CLIP.ipynb

Case Study: SuperResolution GAN

Supervised Task

Given a low-resolution image, reconstruct the high-resolution image.

Prior Art

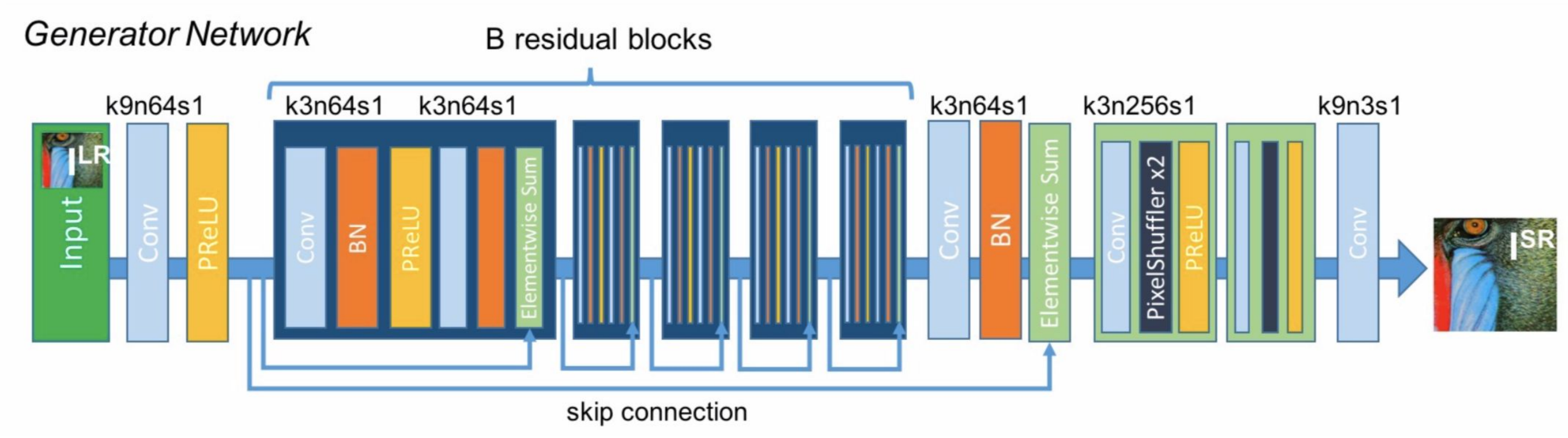
Residual Transpose CNN with MSE loss has shown great performance on this task but the reconstructions are blurry.

Bicubic

Original

Ledig, C., et al. (2017)

SRGAN: Generator CNN



Similar to ResNet, but uses Transposed Convolutions to increase the spatial dimensions.

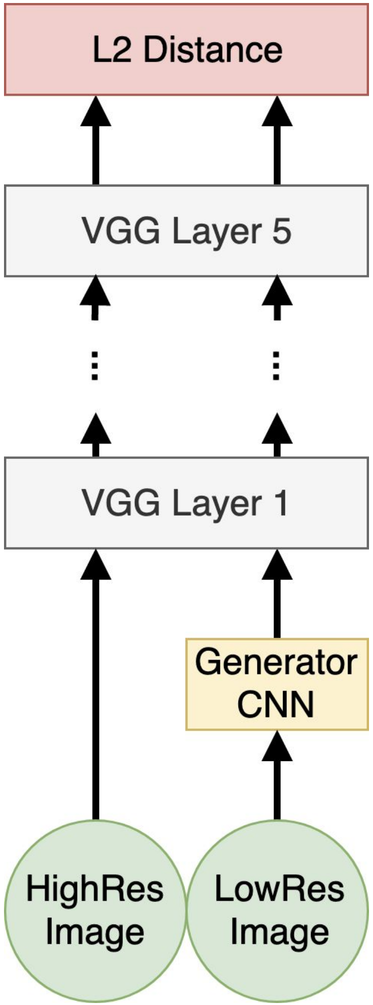
Ledig, C., et al. (2017)

SRGAN: VGG Loss

Comparison of reconstructed vs the original image is done using pre-trained VGG19.

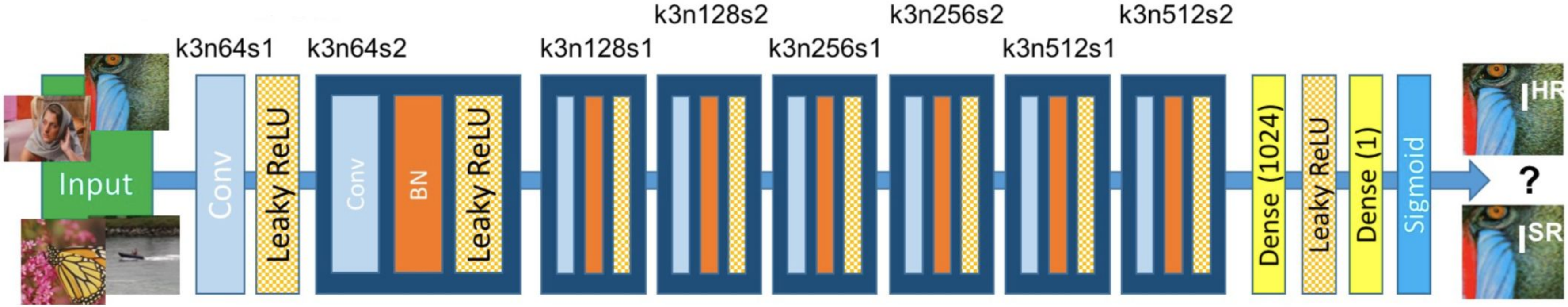
Both images go through several convolutional layers of VGG19, and the final activation volumes are compared.

The loss is the L2 distance between the activation volumes of the reconstruction vs the original image.



Ledig, C., et al. (2017)

SRGAN: Discriminator CNN



1. Fully-convolutional CNN
2. MLP Classifier
3. Binary Cross Entropy.

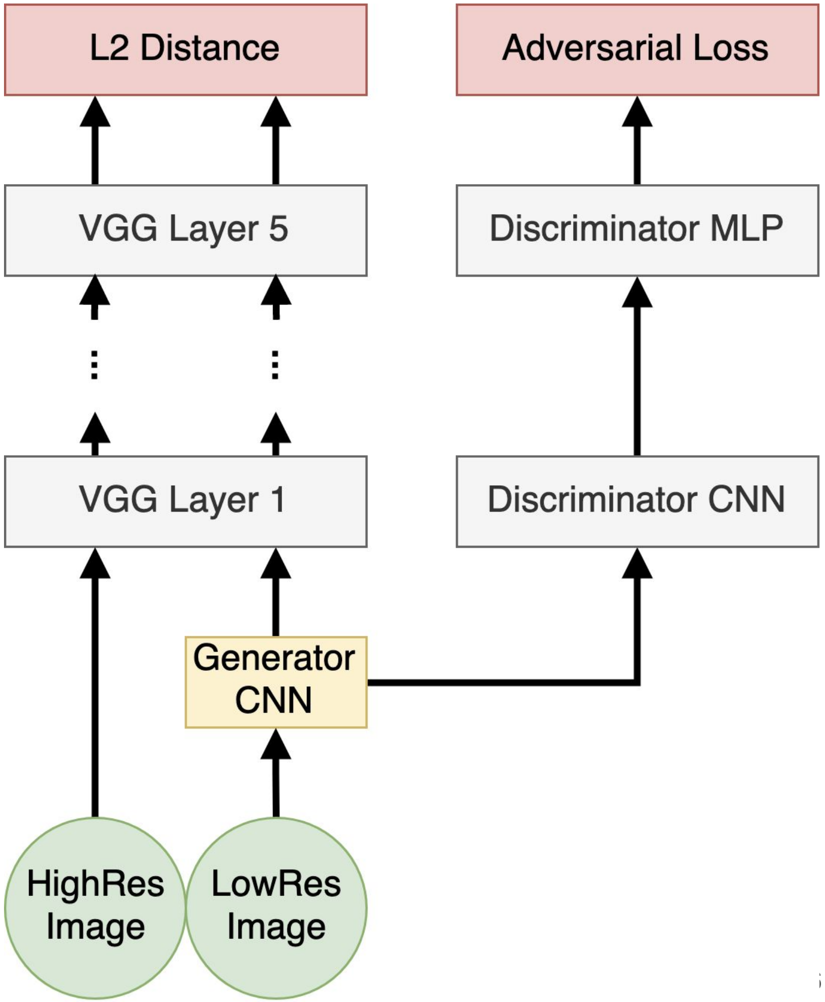
Ledig, C., et al. (2017)

SRGAN: Forward Pass

Dual Loss
Weighted sum of Adversarial Loss and VGG Activation Loss

Authors find the AL+VGG combo to work better than individual losses or MSE.

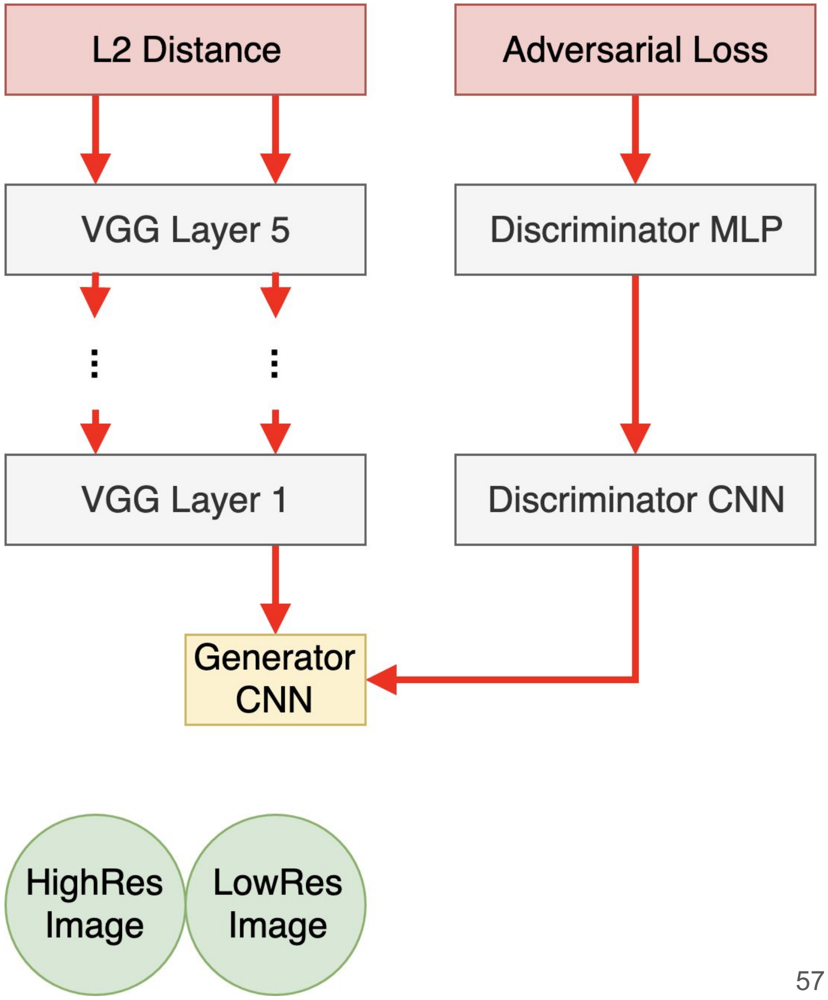
Ledig, C., et al. (2017)



SRGAN: Backprop

Dual Loss
Weighted sum of Adversarial Loss and VGG Activation Loss

Authors find the AL+VGG combo to work better than individual losses or MSE.



Ledig, C., et al. (2017)

SRGAN: Examples

Bicubic

MSE

VGG+AL

Original

Ledig, C., et al. (2017)

Bicubic

MSE

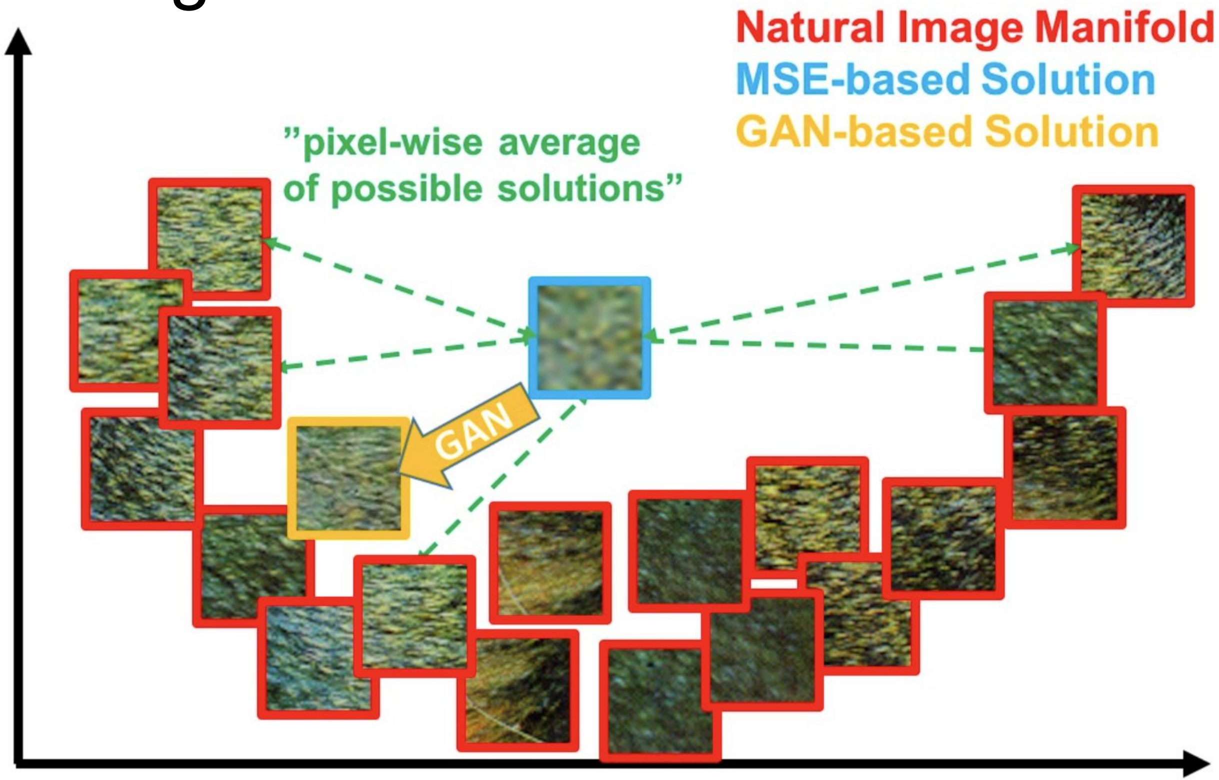
Original

VGG+AL

Ledig, C., et al. (2017)

SRGAN: What's wrong with MSE?

A mean is often a bad approximator of any of the samples from multi-modal distributions.



Ledig, C., et al. (2017)

References

1. A Style-Based Generator Architecture for Generative Adversarial Networks (Karras, T., et al., 2018)
2. Deep Learning Face Attributes in the Wild (Ziwei, L., et al., 2015)
3. Generative Adversarial Nets (Goodfellow, I., et al. 2014)
4. Generative Adversarial Text to Image Synthesis (Redd, S., et al., 2016)
5. High-Resolution Image Synthesis with Latent Diffusion Models (Robin, R., et al., 2022)
6. Large Scale GAN Training for High Fidelity Natural Images Synthesis (Brock, A., et al., 2019)
7. NIPS 2016 Tutorial: Generative Adversarial Networks (Goodfellow, I., 2016)
8. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network (Ledig, C., et al., 2017)
9. Progressive Growing of GANs for Improved Quality, Stability, and Variation (Karras, T., et al., 2018)
10. Training Generative Adversarial Networks with Limited Data (Karras, T., et al., 2020)
11. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks (Radford, A., Metz, L., 2016)

Thank you

Q&A Time

(Image by a Stable Diffusion model)