July 25th, 2020

GANs:
the story so far

Vikram Voleti

PhD student - Mila, University of Montreal **Prof. Christopher Pal**

Visiting Researcher - University of Guelph **Prof. Graham Taylor**

voletiv.github.io

@ (virtual) Summer Symposium on AI Research, India

Generative Adversarial Networks (GANs)

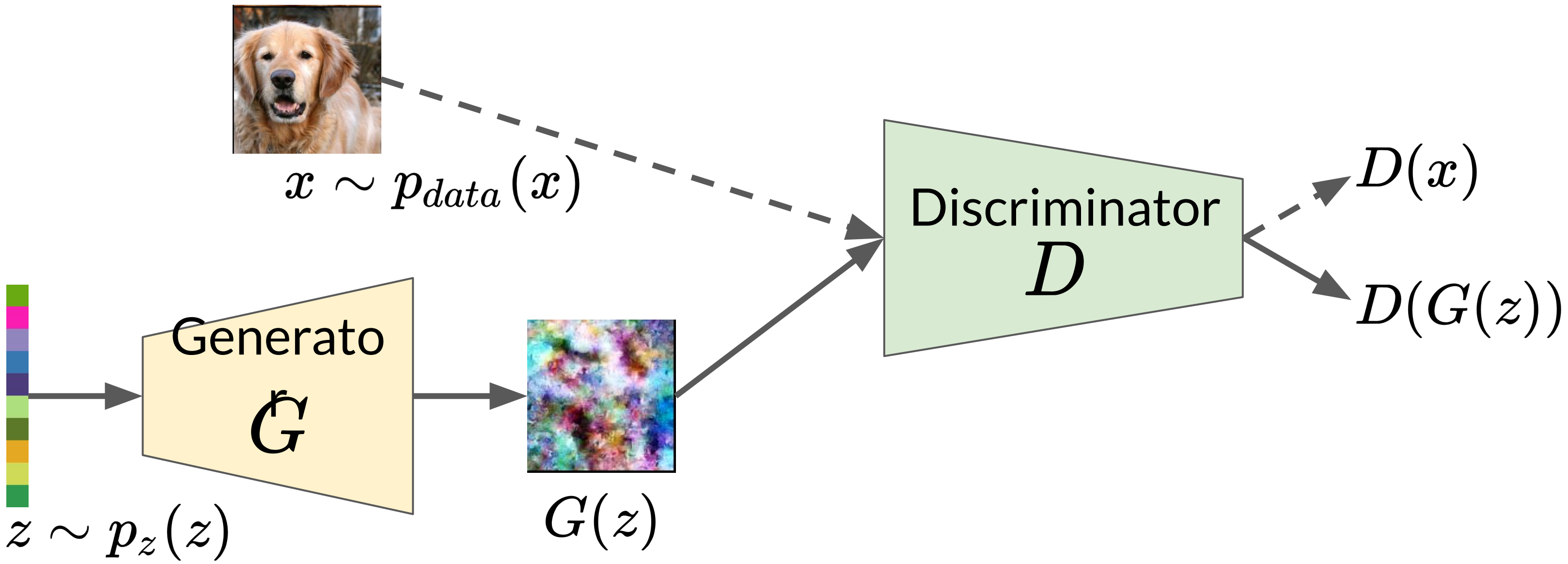
Goal : model the real data distribution

Neural networks!

Generator v/s Discriminator

https://arxiv.org/abs/1406.2661

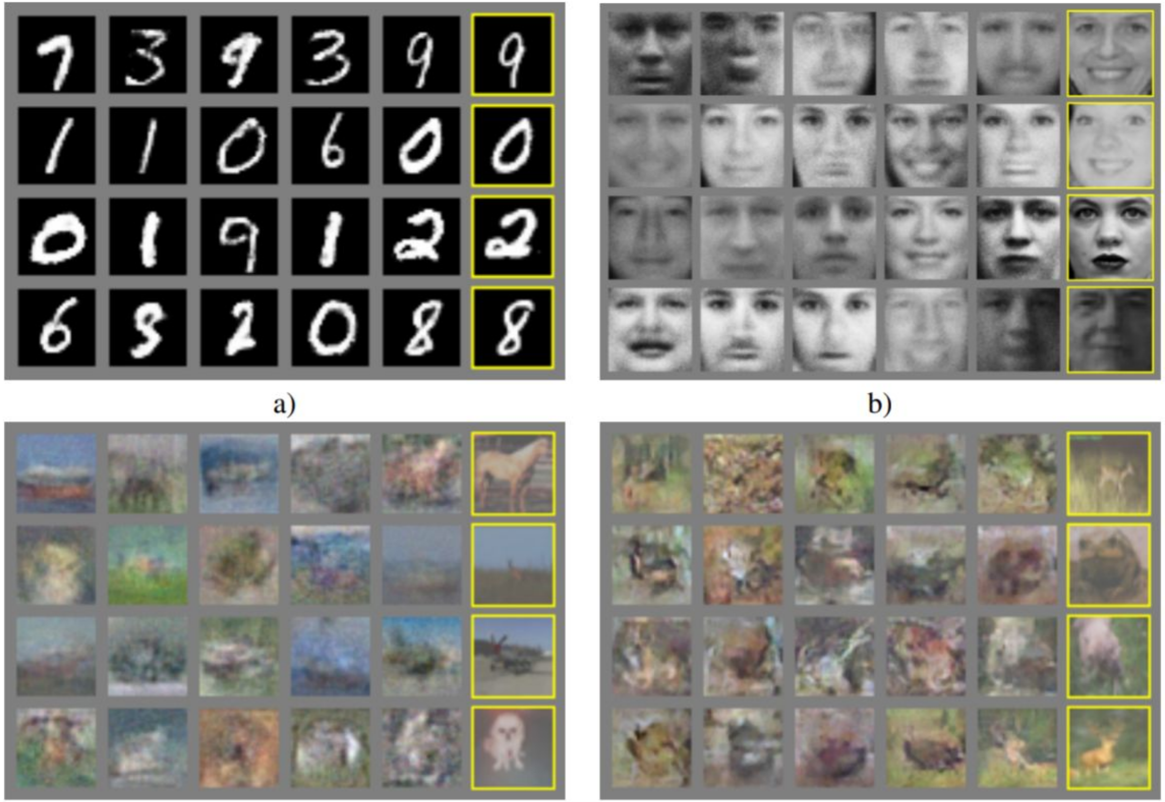
GAN: Generative Adversarial Networks



$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

https://arxiv.org/abs/1406.2661

GAN: Generative Adversarial Networks

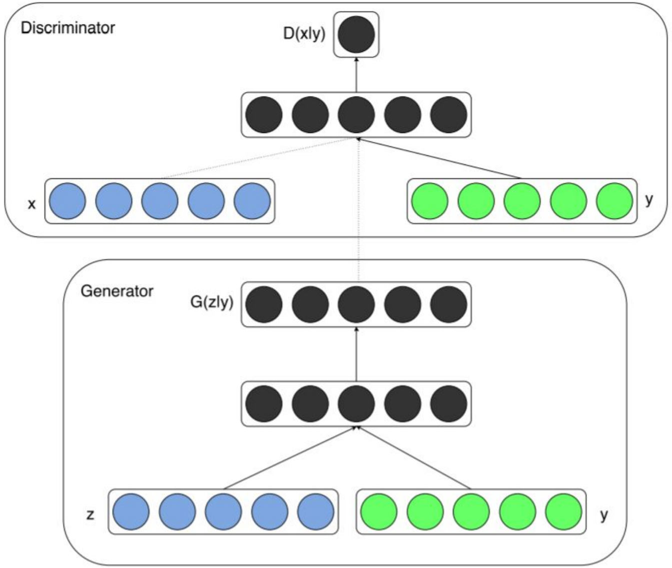


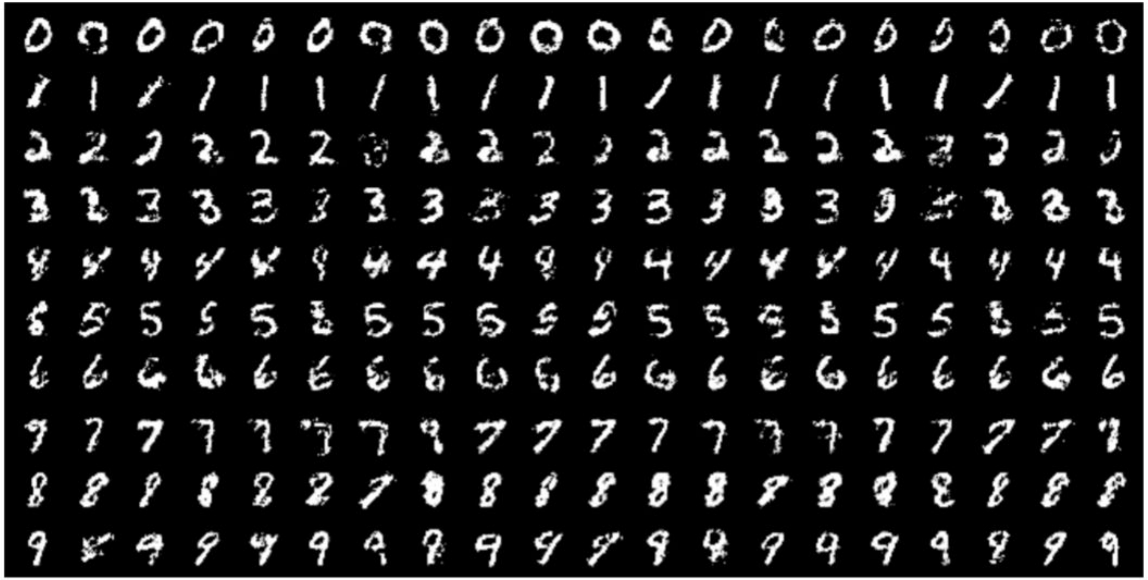
a) b) c) d)

https://arxiv.org/abs/1406.2661

cGAN : Conditional GAN

- Adds class-conditioning to G and D

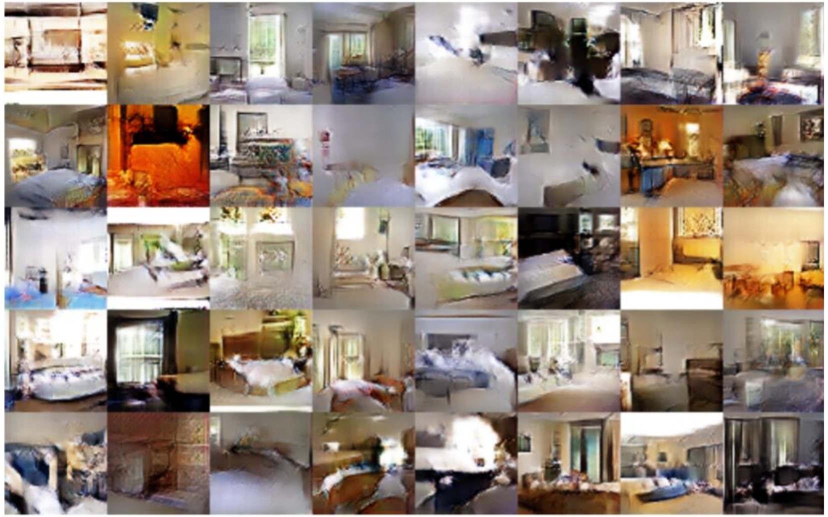


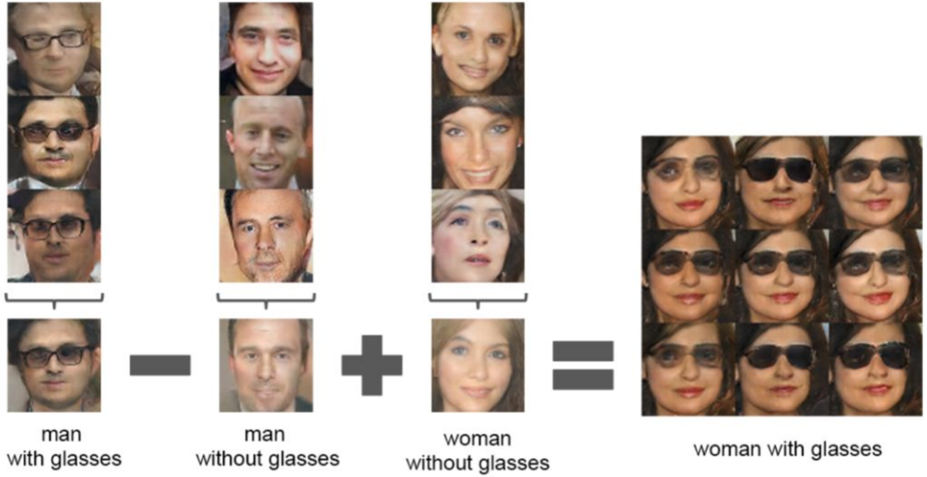


Conditional Generative Adversarial Nets https://arxiv.org/pdf/1411.1784.pdf

DCGAN : Deep Convolutional GAN

- Replaces all FC and pooling layers with convolutional layers

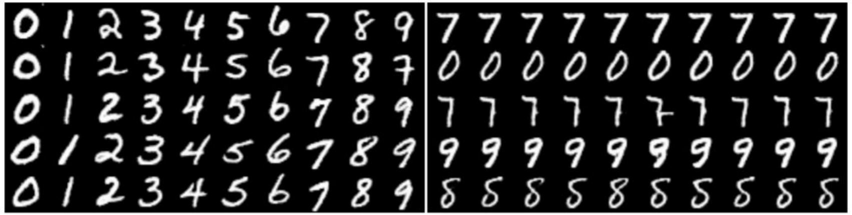




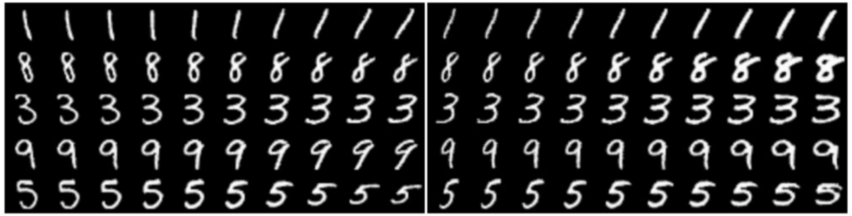
Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks https://arxiv.org/abs/1511.06434

InfoGAN

- Improves conditional dependence by maximizing mutual information between latent code and generated image



(a) Varying c_1 on InfoGAN (Digit type) (b) Varying c_1 on regular GAN (No clear meaning)



(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation) (d) Varying c_3 from -2 to 2 on InfoGAN (Width)

(a) Azimuth (pose) (b) Elevation

(c) Lighting (d) Wide or Narrow

InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets https://arxiv.org/abs/1606.03657

ALI : Adversarially Learned Inference / BiGAN : Bidirectional GAN

- Discriminates on the joint distribution of noise and sample

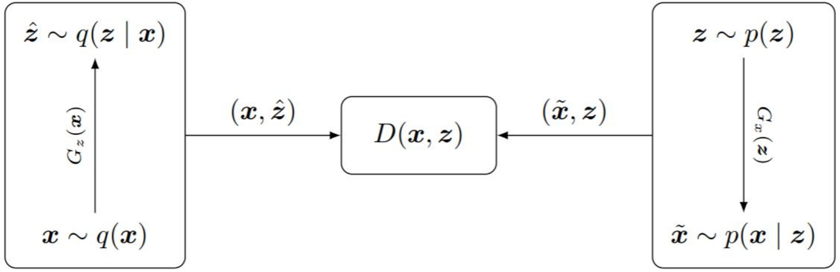


Figure 1: The adversarially learned inference (ALI) game.

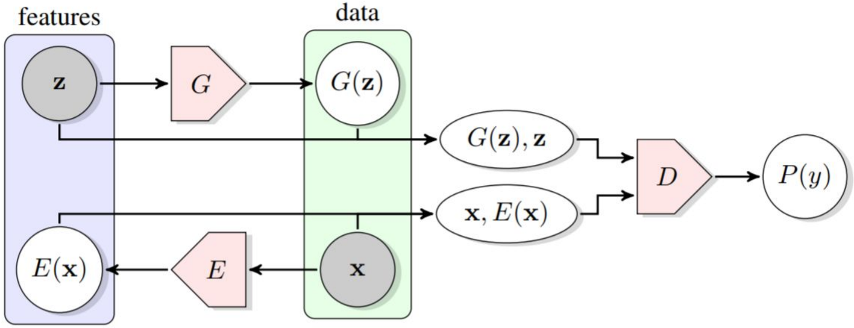
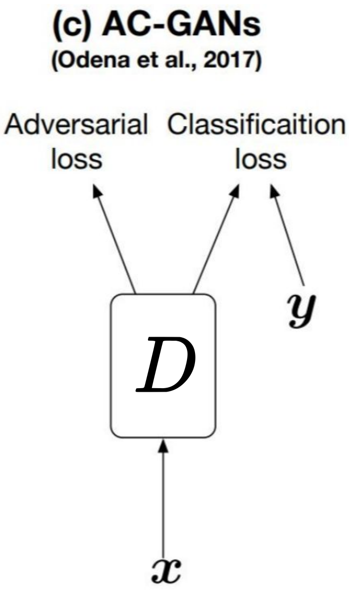


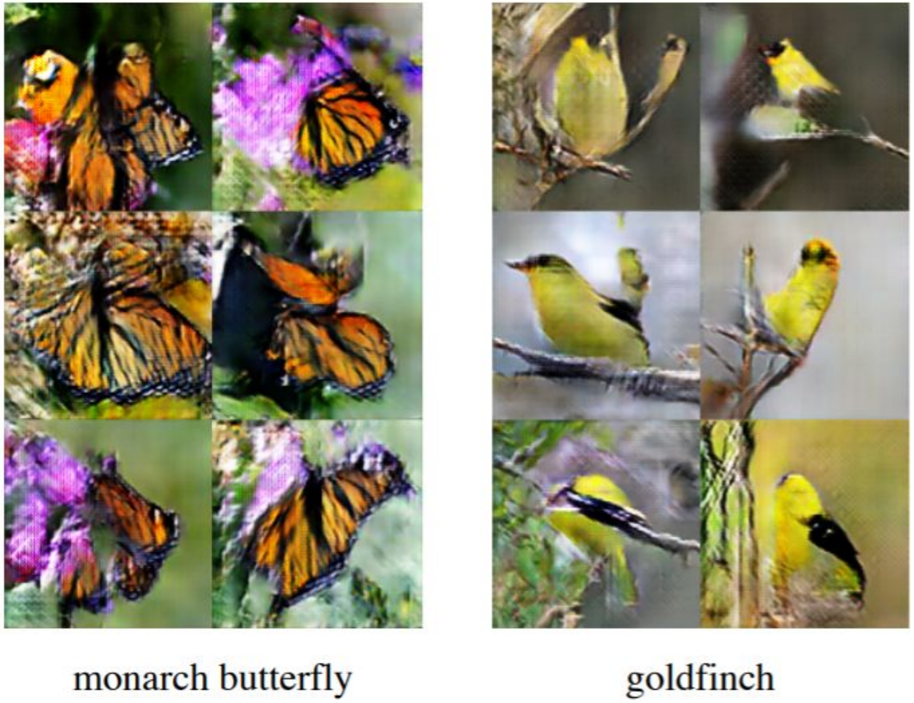
Figure 1: The structure of Bidirectional Generative Adversarial Networks (BiGAN).

(ICLR 2017) Adversarially Learned Inference https://arxiv.org/abs/1606.00704 (ICLR 2017) Adversarially Learned Inference https://arxiv.org/abs/1606.00704

AC-GAN: Auxiliary Classifier GANs

- Added auxiliary classifier to D
- Attempted ImageNet 128x128

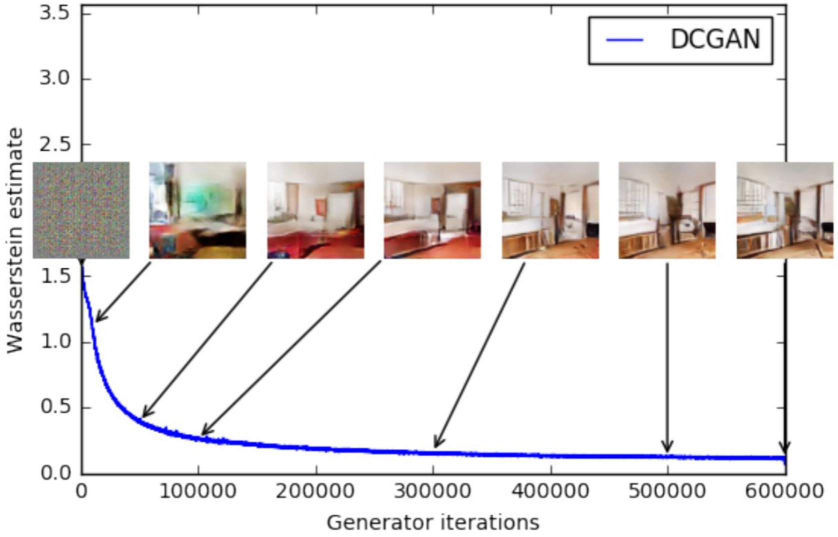




Conditional Image Synthesis with Auxiliary Classifier GANs https://arxiv.org/abs/1610.09585

WGAN : Wasserstein GAN

- Replaced JS divergence with Wasserstein distance - improved stability
- Notion of Lipschitzness for stability in GANs



Wasserstein GAN https://arxiv.org/abs/1701.07875

WGAN-GP : Wasserstein GAN with Gradient Penalty

- Constrains the Lipschitz by constraining the gradient norm of D's output (instead of weight clipping as was done in WGAN)

Algorithm 1 WGAN with gradient penalty. We use default values of $\lambda = 10$, $n_{\text{critic}} = 5$, $\alpha = 0.0001$, $\beta_1 = 0$, $\beta_2 = 0.9$.

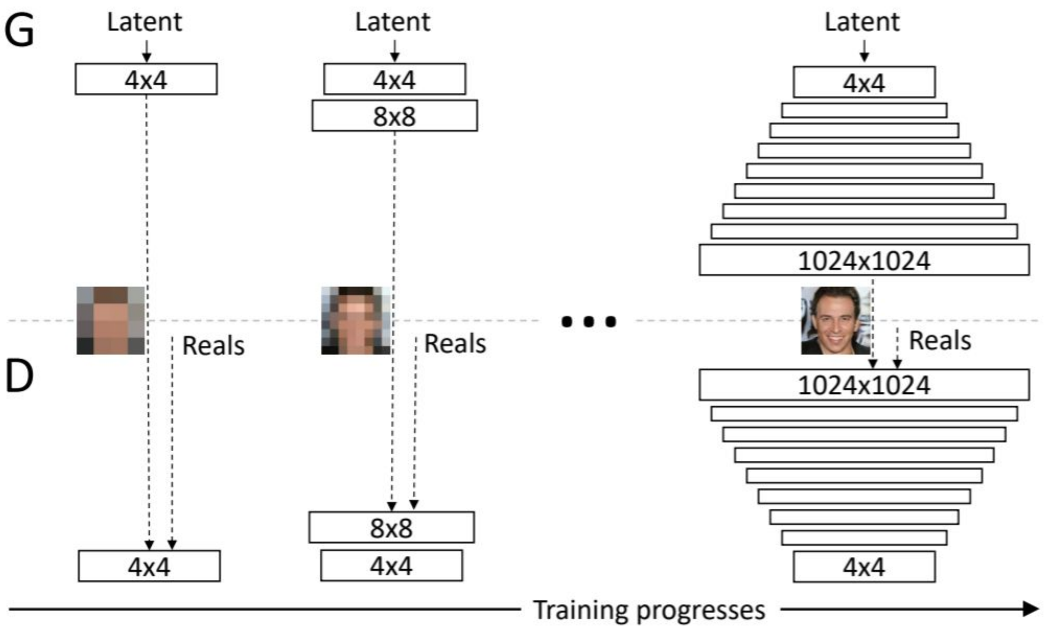
Require: The gradient penalty coefficient λ, the number of critic iterations per generator iteration n_{critic}, the batch size m, Adam hyperparameters α, β_1, β_2.
Require: initial critic parameters w_0, initial generator parameters θ_0.

while θ has not converged **do**
  for $t = 1, ..., n_{\text{critic}}$ **do**
    for $i = 1, ..., m$ **do**
      Sample real data $\boldsymbol{x} \sim \mathbb{P}_r$, latent variable $\boldsymbol{z} \sim p(\boldsymbol{z})$, a random number $\epsilon \sim U[0, 1]$.
      $\tilde{\boldsymbol{x}} \leftarrow G_\theta(\boldsymbol{z})$
      $\hat{\boldsymbol{x}} \leftarrow \epsilon \boldsymbol{x} + (1 - \epsilon)\tilde{\boldsymbol{x}}$
      $L^{(i)} \leftarrow D_w(\tilde{\boldsymbol{x}}) - D_w(\boldsymbol{x}) + \boxed{\lambda(\|\nabla_{\hat{\boldsymbol{x}}} D_w(\hat{\boldsymbol{x}})\|_2 - 1)^2}$
    end for
    $w \leftarrow \text{Adam}(\nabla_w \frac{1}{m} \sum_{i=1}^m L^{(i)}, w, \alpha, \beta_1, \beta_2)$
  end for
  Sample a batch of latent variables $\{\boldsymbol{z}^{(i)}\}_{i=1}^m \sim p(\boldsymbol{z})$.
  $\theta \leftarrow \text{Adam}(\nabla_\theta \frac{1}{m} \sum_{i=1}^m -D_w(G_\theta(\boldsymbol{z})), \theta, \alpha, \beta_1, \beta_2)$
end while

Improved Training of Wasserstein GANs https://arxiv.org/abs/1704.00028

ProGAN : Progressive GAN

- Progressively trains to higher resolutions
- First to produce 1024x1024 images



Progressive Growing of GANs for Improved Quality, Stability, and Variation https://arxiv.org/abs/1710.10196

SNGAN : Spectral Normalization GAN

- Introduces Spectral Normalization to constrain the Lipschitz constant of D

Our *spectral normalization* normalizes the spectral norm of the weight matrix W so that it satisfies the Lipschitz constraint $\sigma(W) = 1$:

$$\bar{W}_{\mathrm{SN}}(W) := W/\sigma(W). \tag{8}$$

Spectral Normalization for Generative Adversarial Networks https://arxiv.org/abs/1802.05957

BigGAN

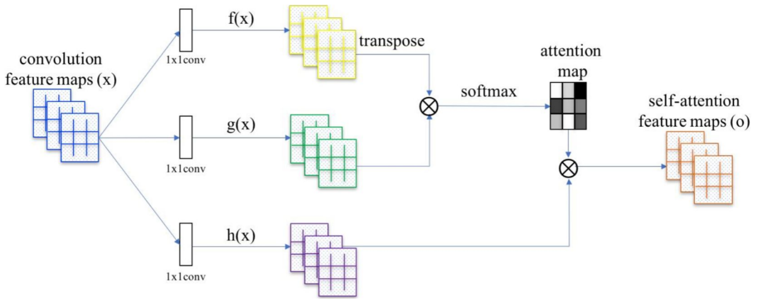
- SOTA results on condition high-res image generation from ImageNet
- Bumped up batch size using TPUs

Inception score (128x128) **166.3** from 52.52, FID **9.6** from 18.65

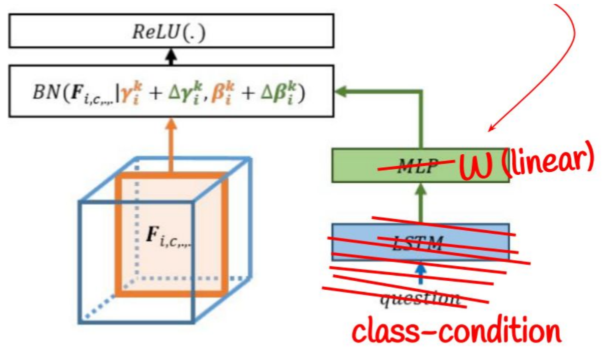
Large Scale GAN Training for High Fidelity Natural Image Synthesis https://arxiv.org/abs/1809.11096

BigGAN

Builds on top of SAGAN



Uses Conditional Batch Norm



Uses hinge loss

$$L_D = -\mathbb{E}_{(x,y)\sim p_{data}}[\min(0, -1 + D(x,y))] - \mathbb{E}_{z\sim p_z, y\sim p_{data}}[\min(0, -1 - D(G(z),y))],$$
$$L_G = -\mathbb{E}_{z\sim p_z, y\sim p_{data}} D(G(z),y),$$

Large Scale GAN Training for High Fidelity Natural Image Synthesis https://arxiv.org/abs/1809.11096

BigGAN

SUMMARY:

Batch size 8x

Width↑ 50%

Shared embeddings w/ linear projection

Hierarchical latent space

Orthogonal Regularization

Batch	Ch.	Param (M)	Shared	Hier.	Ortho.	Itr $\times 10^3$	FID	IS
256	64	81.5	SA-GAN Baseline			1000	18.65	52.52
512	64	81.5	✗	✗	✗	1000	15.30	58.77(±1.18)
1024	64	81.5	✗	✗	✗	1000	14.88	63.03(±1.42)
2048	64	81.5	✗	✗	✗	732	12.39	76.85(±3.83)
2048	96	173.5	✗	✗	✗	295(±18)	9.54(±0.62)	92.98(±4.27)
2048	96	160.6	✓	✗	✗	185(±11)	9.18(±0.13)	94.94(±1.32)
2048	96	158.3	✓	✓	✗	152(±7)	8.73(±0.45)	98.76(±2.84)
2048	96	158.3	✓	✓	✓	165(±13)	8.51(±0.32)	99.31(±2.10)
2048	64	71.3	✓	✓	✓	371(±7)	10.48(±0.10)	86.90(±0.61)

Table 1: Fréchet Inception Distance (FID, lower is better) and Inception Score (IS, higher is better) for ablations of our proposed modifications. *Batch* is batch size, *Param* is total number of parameters, *Ch.* is the channel multiplier representing the number of units in each layer, *Shared* is using shared embeddings, *Hier.* is using a hierarchical latent space, *Ortho.* is Orthogonal Regularization, and *Itr* either indicates that the setting is stable to 10^6 iterations, or that it collapses at the given iteration. Other than rows 1-4, results are computed across 8 different random initializations.

Large Scale GAN Training for High Fidelity Natural Image Synthesis https://arxiv.org/abs/1809.11096

BigGAN

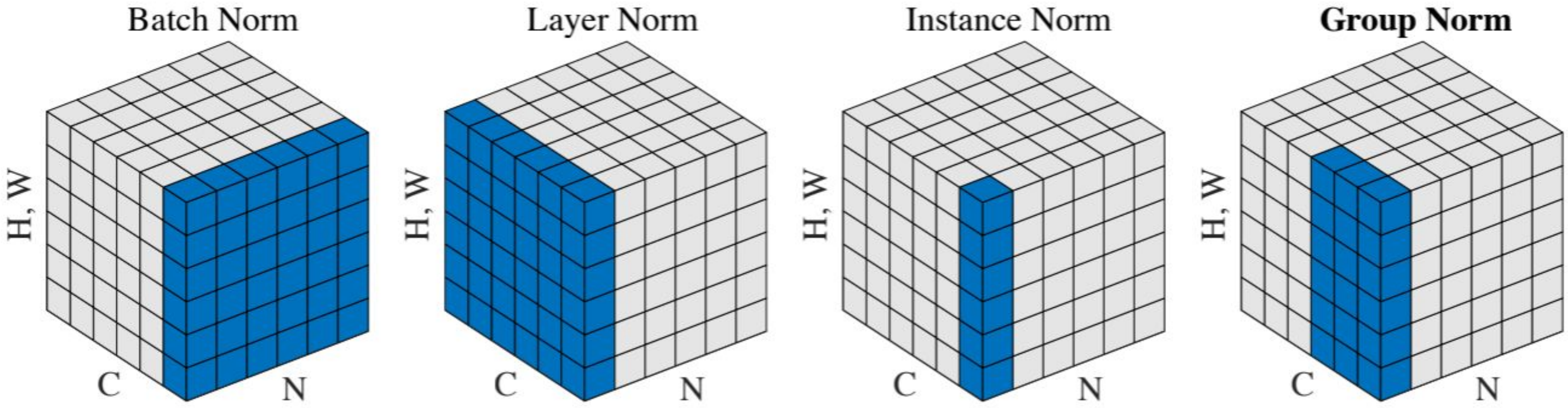
- Appendix contains important info!

Latent interpolations

Weird examples

Large Scale GAN Training for High Fidelity Natural Image Synthesis https://arxiv.org/abs/1809.11096

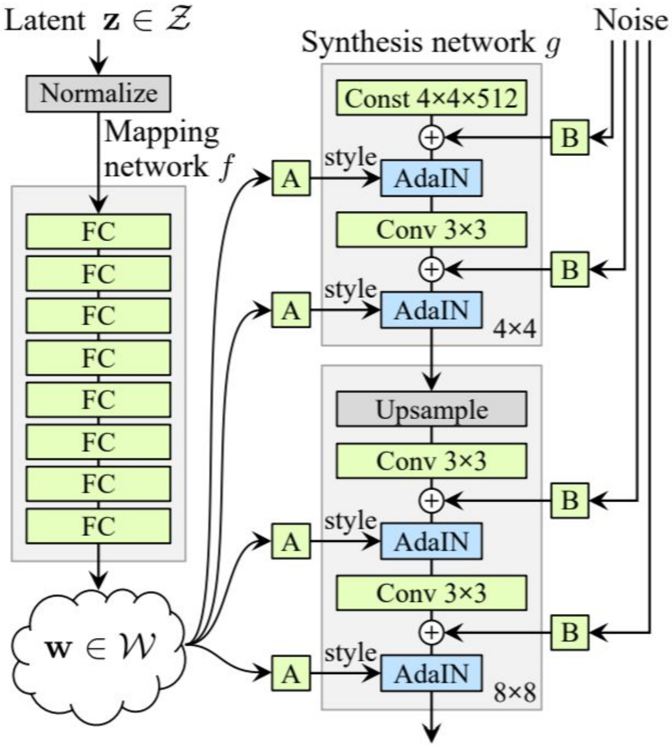
Normalization techniques



Group Normalization https://arxiv.org/abs/1803.08494

StyleGAN

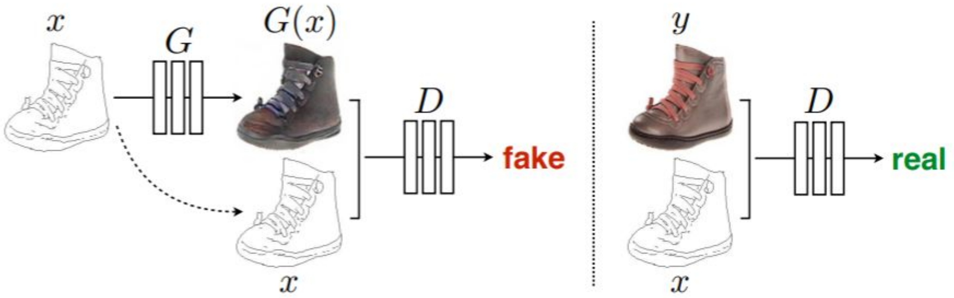
- Photo-realistic 1024x1024 images



A Style-Based Generator Architecture for Generative Adversarial Networks https://arxiv.org/abs/1812.04948

pix2pix

- Paired image-to-image translation



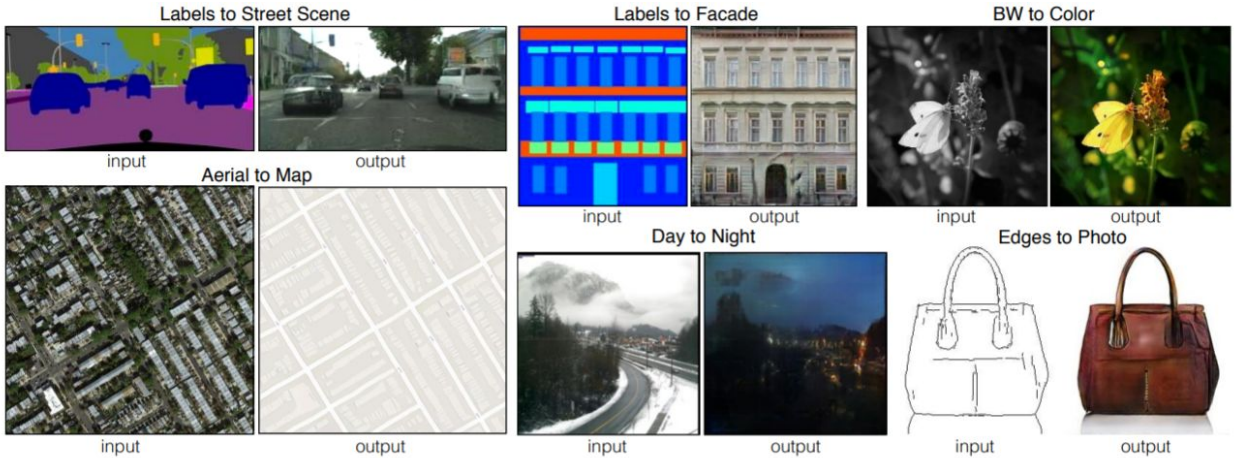
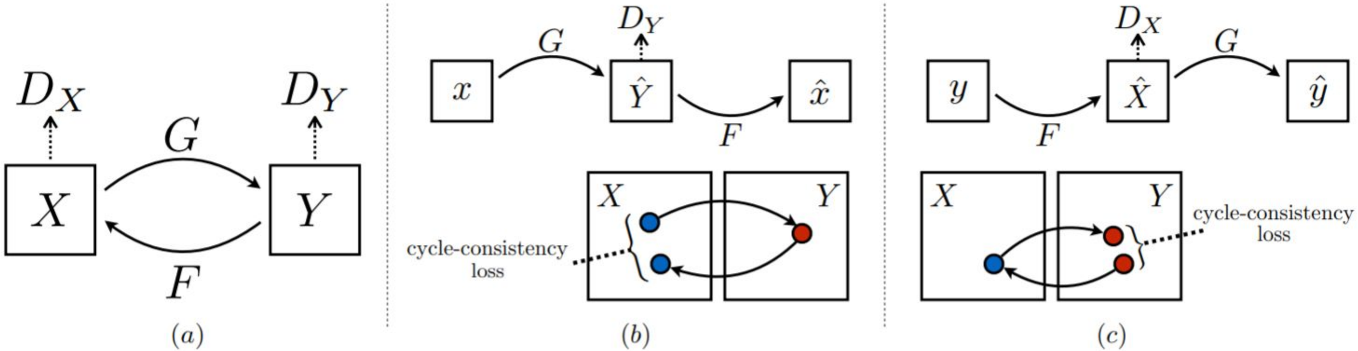
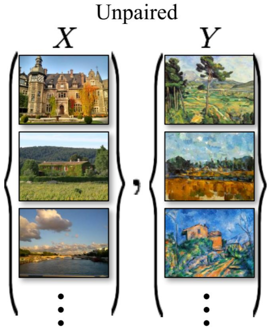


Image-to-Image Translation with Conditional Adversarial Networks https://arxiv.org/abs/1611.07004

CycleGAN

- Unpaired image-to-image translation

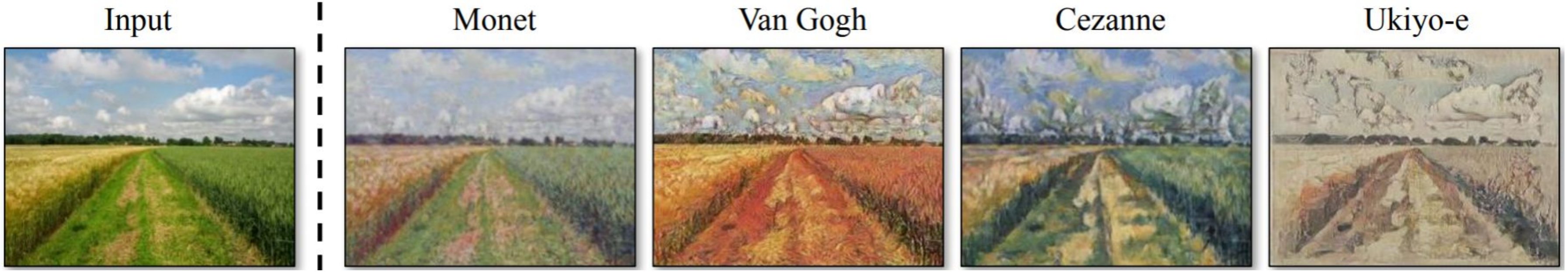


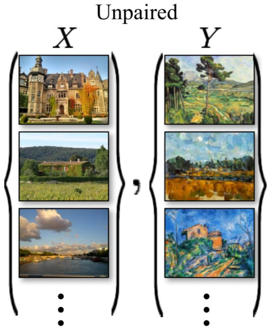


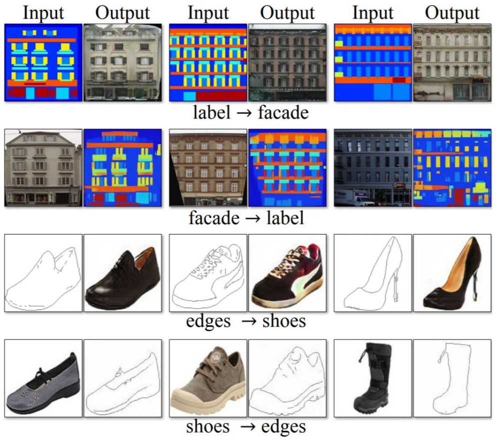
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks https://arxiv.org/abs/1703.10593

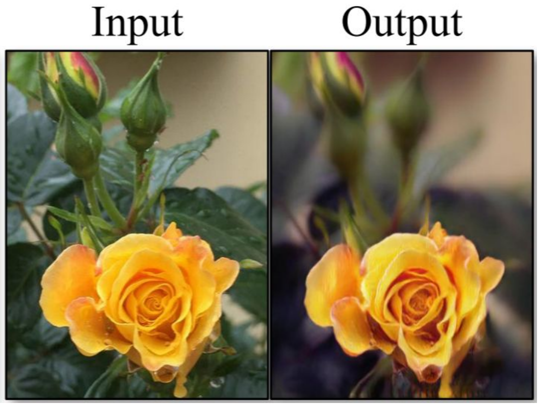
CycleGAN

- Unpaired image-to-image translation





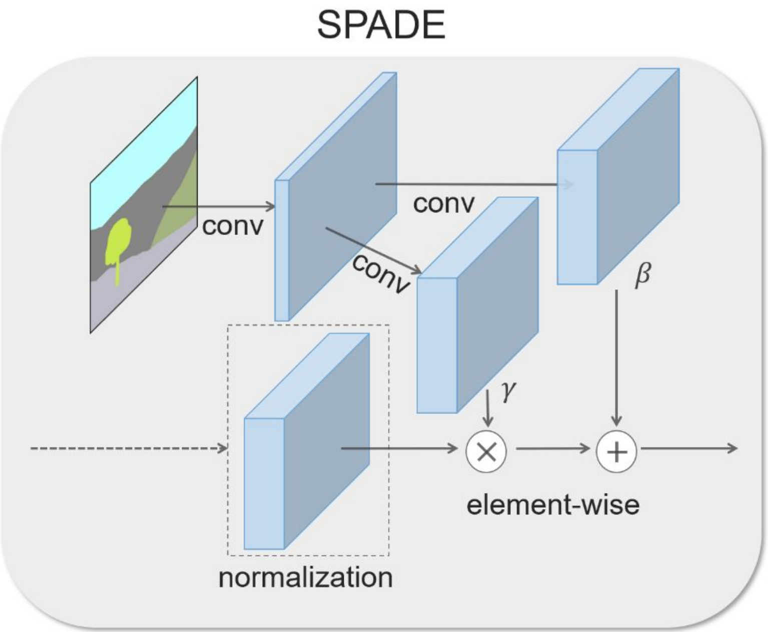


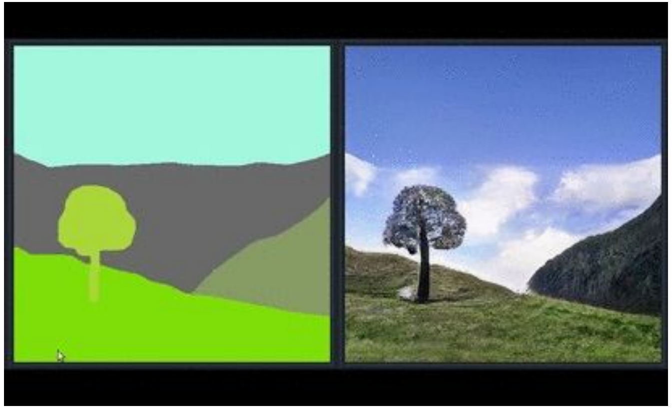


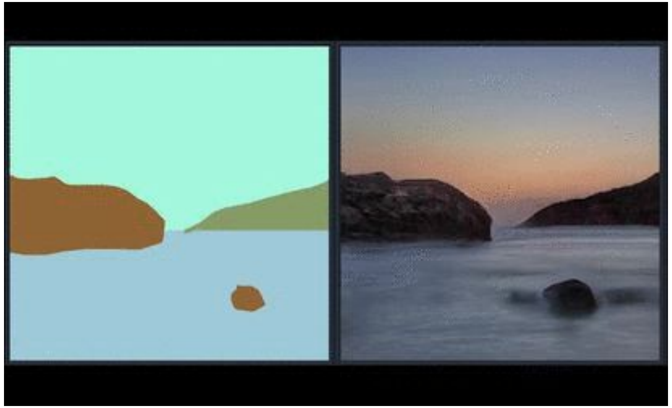
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks https://arxiv.org/abs/1703.10593

GauGAN / SPADE

- Element-wise re-normalization





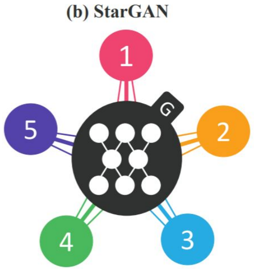


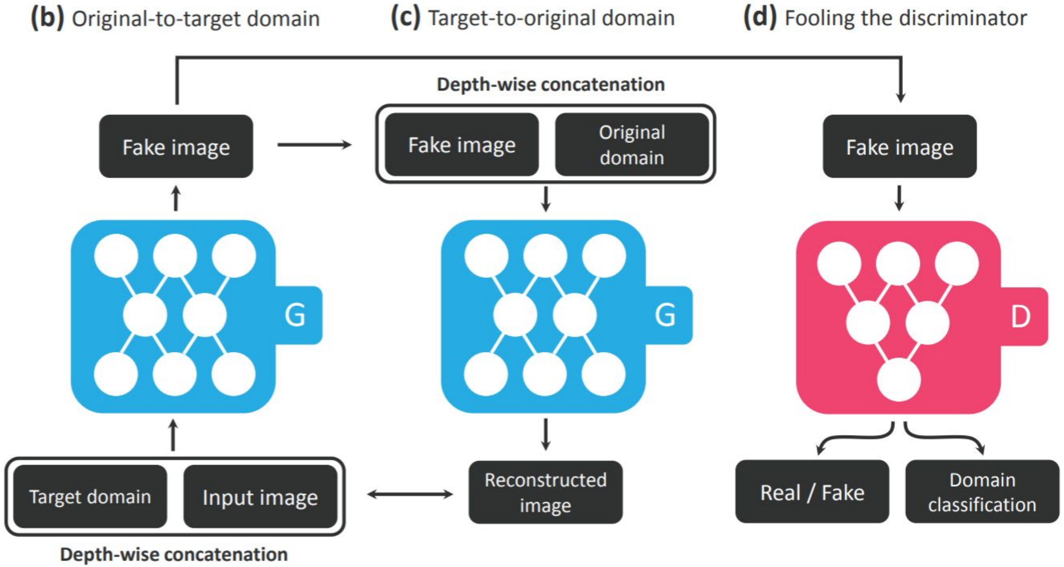
http://nvidia-research-mingyuliu.com/gaugan/

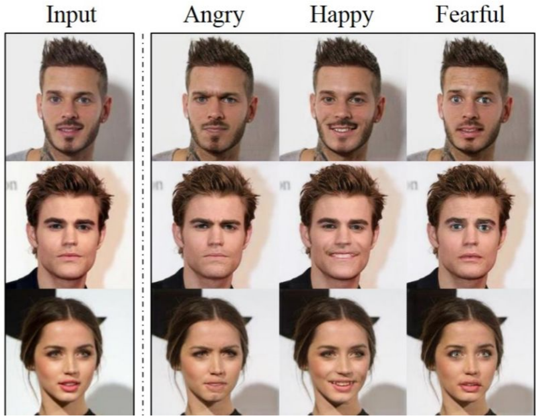
Semantic Image Synthesis with Spatially-Adaptive Normalization https://arxiv.org/abs/1903.07291

StarGAN

- Multi-domain image-to-image translation







StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation https://arxiv.org/abs/1711.09020

Relativistic GANs

- Changes the GAN formulation to a relative score

$$L_D^{RSGAN} = -\mathbb{E}_{(x_r, x_f)\sim(\mathbb{P},\mathbb{Q})} \left[\log(\mathrm{sigmoid}(C(x_r) - C(x_f)))\right].$$

$$L_G^{RSGAN} = -\mathbb{E}_{(x_r, x_f)\sim(\mathbb{P},\mathbb{Q})} \left[\log(\mathrm{sigmoid}(C(x_f) - C(x_r)))\right].$$

256x256 images of cats

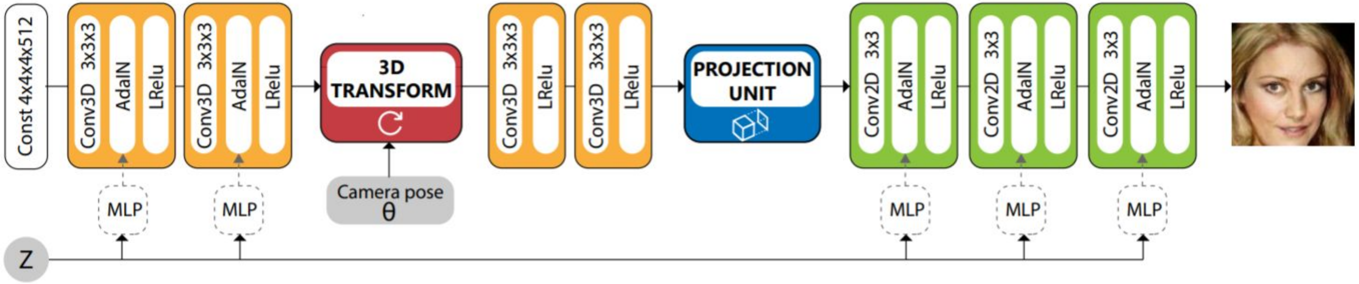
WGAN-GP

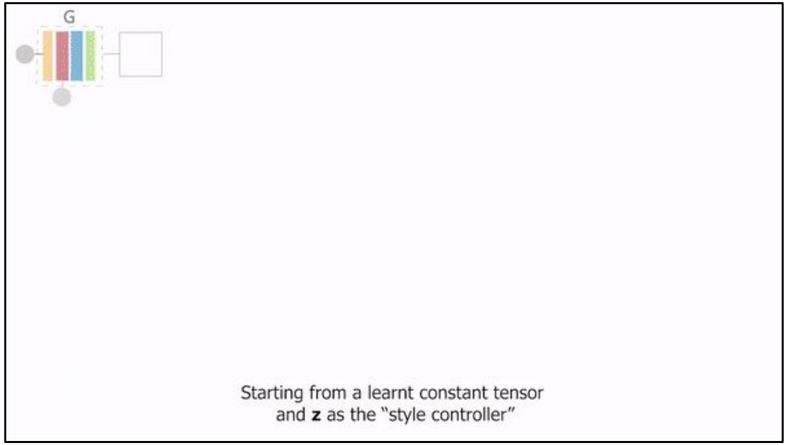
Relativistic GAN

The relativistic discriminator: a key element missing from standard GAN https://arxiv.org/abs/1807.00734

HoloGAN

- Incorporates 3D geometry into the model



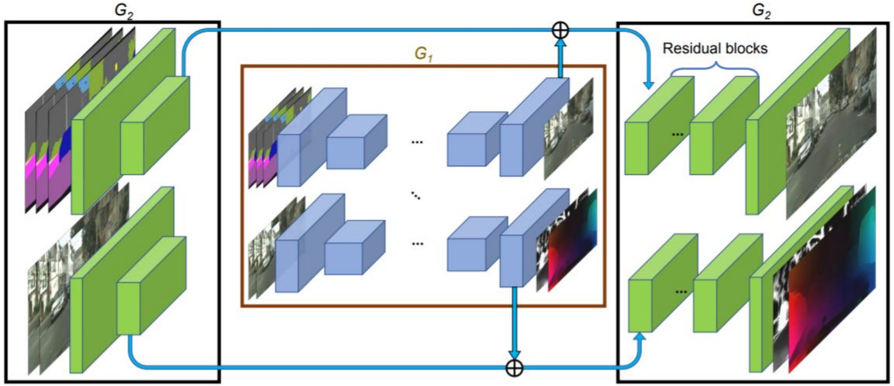


https://www.monkeyoverflow.com/#/hologan-unsupervised-learning-of-3d-representations-from-natural-images/

HoloGAN: Unsupervised learning of 3D representations from natural images https://arxiv.org/abs/1904.01326

Vid2Vid

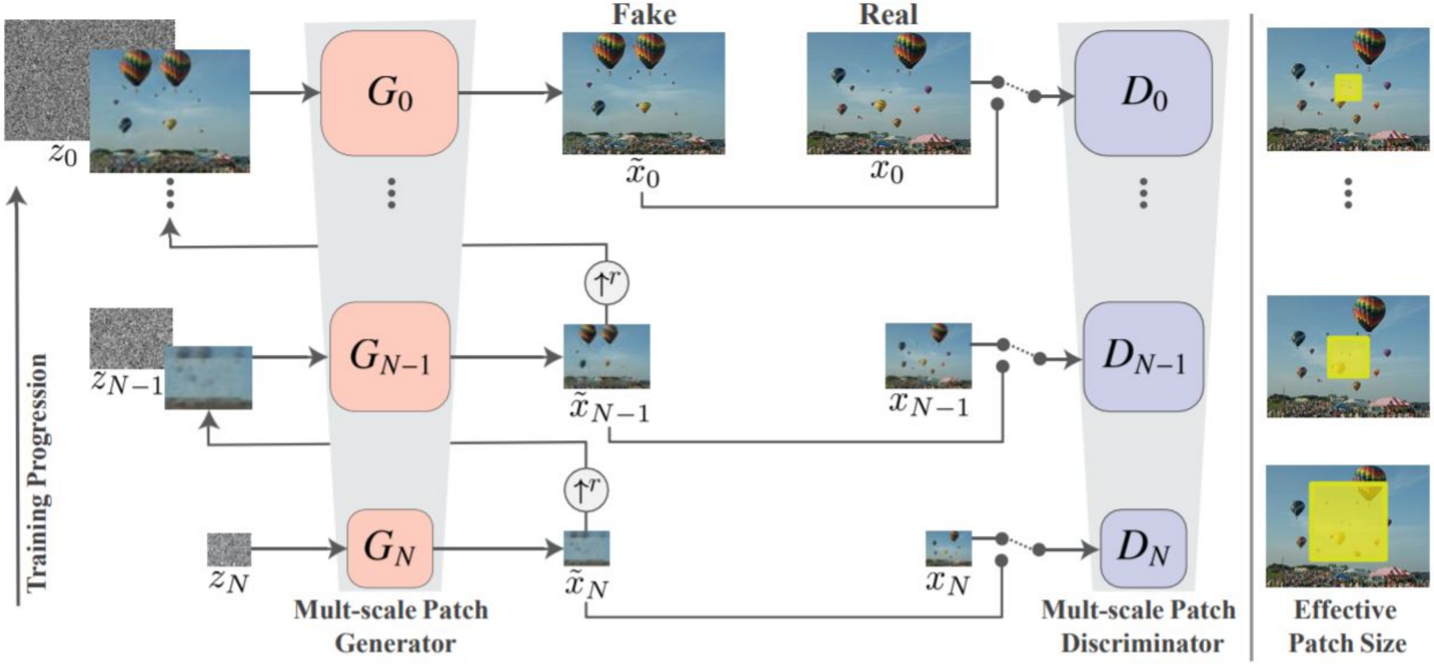
- Multi-scale video-to-video translation



Video-to-Video Synthesis https://tcwang0509.github.io/vid2vid/

SinGAN

- Single-image multi-scale generation



SinGAN: Learning a Generative Model from a Single Natural Image https://arxiv.org/abs/1905.01164

Visualizing climate change

- Uses Cycle-GANs

Visualizing the Consequences of Climate Change Using Cycle-Consistent Adversarial Networks https://arxiv.org/abs/1905.03709

Thank you!

voletiv.github.io