Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks

Emily Denton[1*], Soumith Chintala[2*], Arthur Szlam[2], Rob Fergus[2]

[1]New York University
[2]Facebook AI Research
*Denotes equal contribution

December 16, 2015

Overview

- Parametric generative model of natural images
- Difficult to generate large natural images in one shot, but we can exploit their multi-scale structure
- We combine the power of generative adversarial networks (GAN) with a multi-scale image representation (Laplacian pyramid)

Generative modelling of natural images

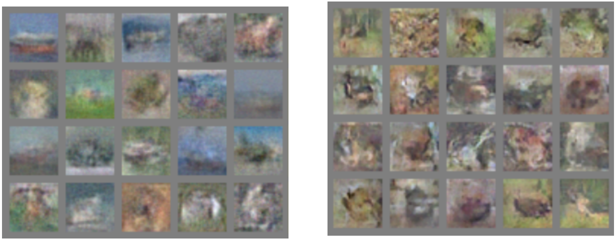
- Have access to $x \sim p_{data}(x)$ through training set
- Want to learn a model $x \sim p_{model}(x)$
- Want p_{model} to be similar to p_{data}
  - Samples drawn from p_{model} reflect structure of p_{data}
  - Samples from true data distribution have high likelihood under p_{model}

Why do generative modeling?

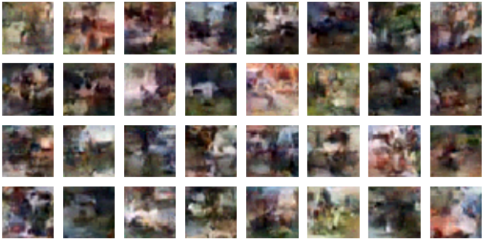
- Unsupervised representation learning
  - Can transfer learned representation so discriminative tasks, retrieval, clustering, etc.
- Train network with both discriminative and generative criterion
  - Very little labeled data
  - Regularization
- Understand data
- Density estimation
- ...

CIFAR-10 samples from other models

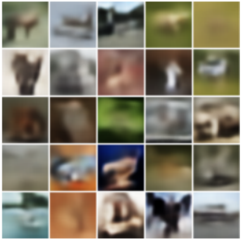
Goodfellow et al. (2014):



Sohl-Dickstein et al. (2015):

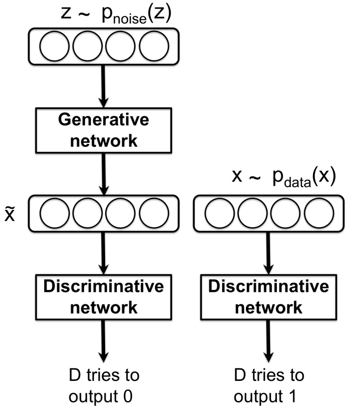


Gregor et al. (2015):



Generative adversarial networks (Goodfellow et al., 2014)

- Generative model G: captures data distribution
- Discriminative model D: trained to distinguish between real and fake samples , defines loss function for G



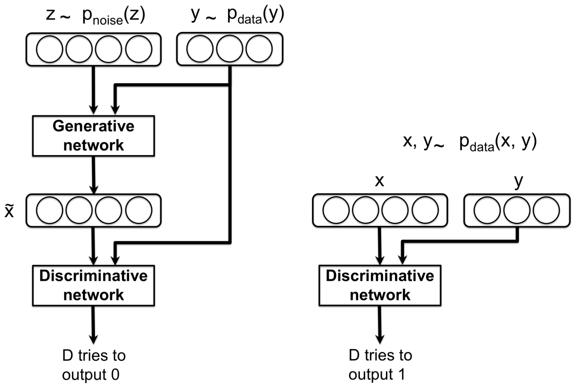
Generative adversarial networks

- D is trained to estimate the probability that a sample came from data distribution rather than G
- G is trained to maximize the probability of D making a mistake

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

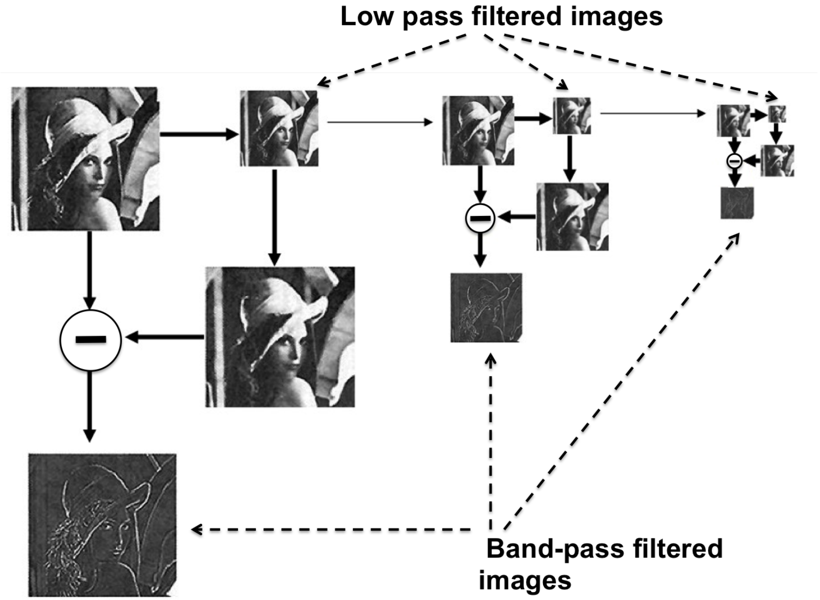
Conditional generative adversarial networks (CGAN)

- Condition generation on additional info **y** (e.g. class label, another image)
- *D* has to determine if samples are realistic given **y**

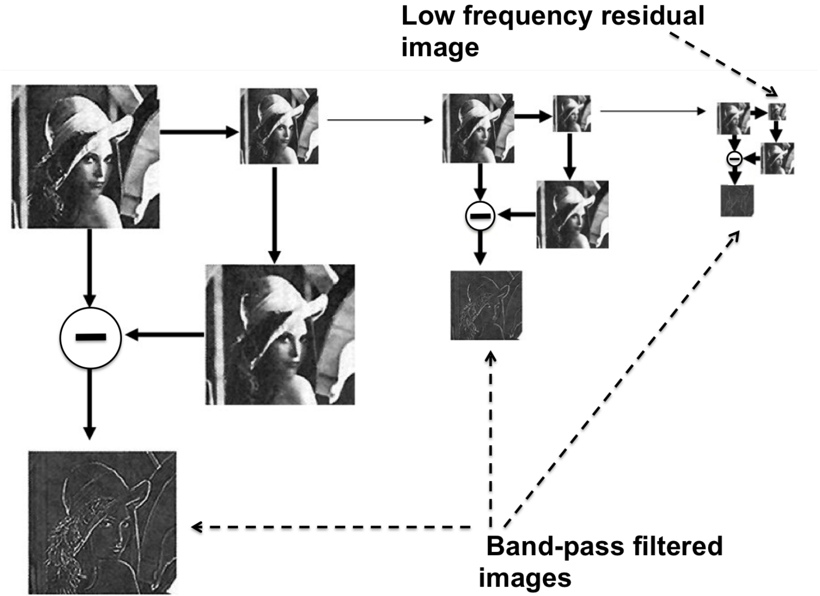


[Mirza and Osindero (2014); Gauthier (2014)]

Laplacian pyramid (Burt & Adelson, 1983)

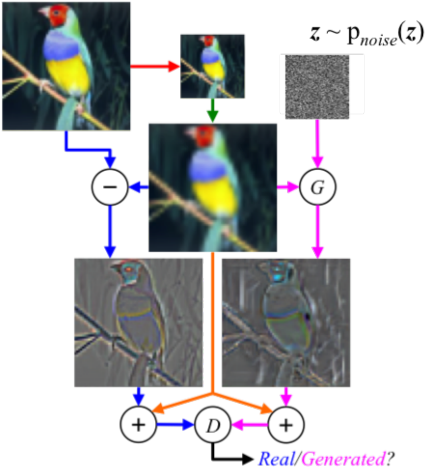


Laplacian pyramid (Burt & Adelson, 1983)



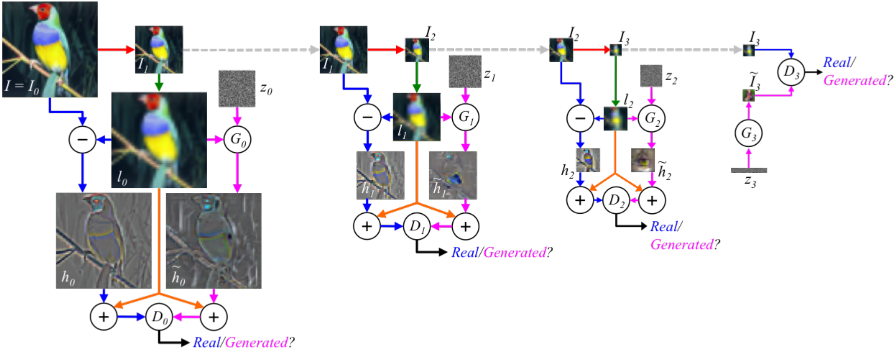
Training procedure

- Train conditional GAN for each level of Laplacian pyramid
- G learns to generate high frequency structure consistent with low frequency image

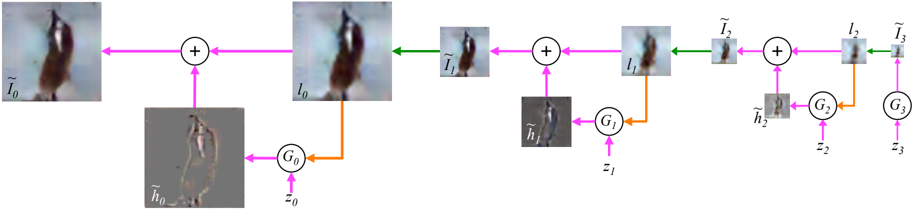


Training procedure

Each level of Laplacian pyramid trained independently

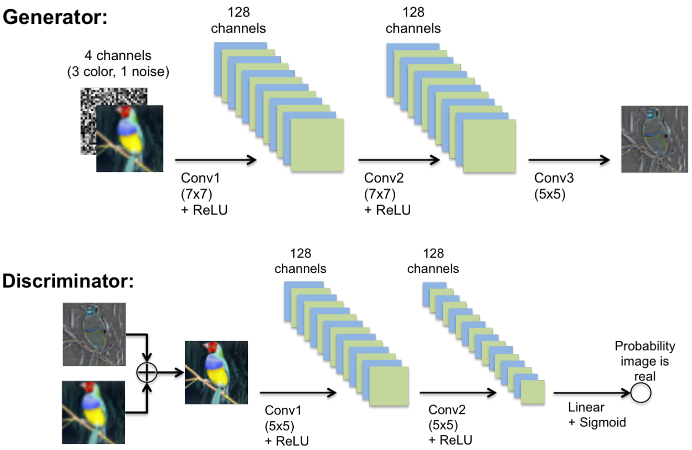


Sampling procedure



CIFAR-10

- Small dataset 32x32 images of objects, 50k images, 10 classes

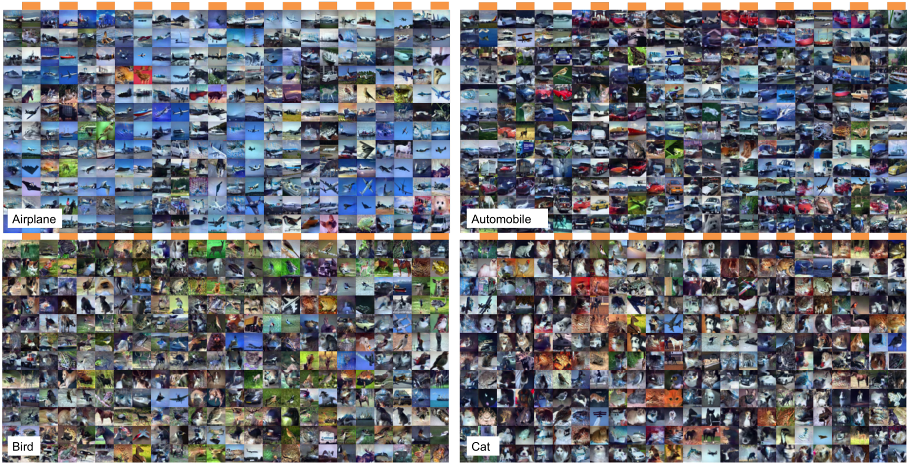


CIFAR-10 ship samples

CIFAR-10 horse samples

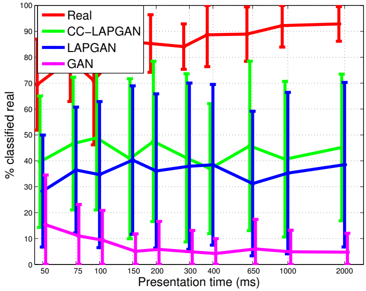
CIFAR-10 nearest neighbours (pixel space)

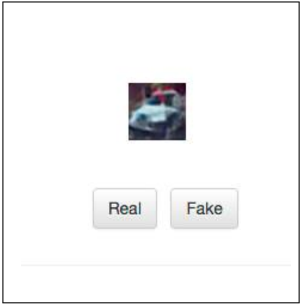
CIFAR-10 nearest neighbours (nn feature space)



CIFAR-10 human evaluations

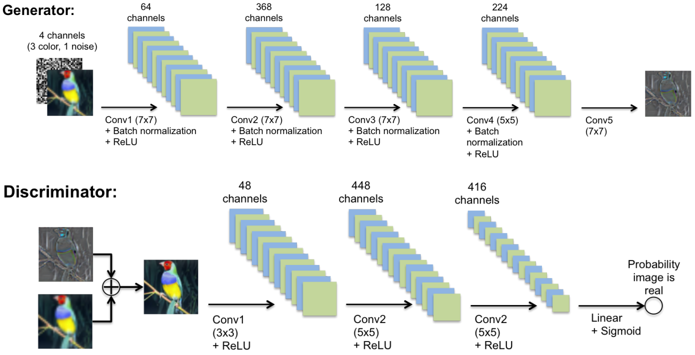
- Humans randomly presented with real or generated image and asked to determine if real of fake
- Humans think LAPGAN generations are real ∼40% of the time



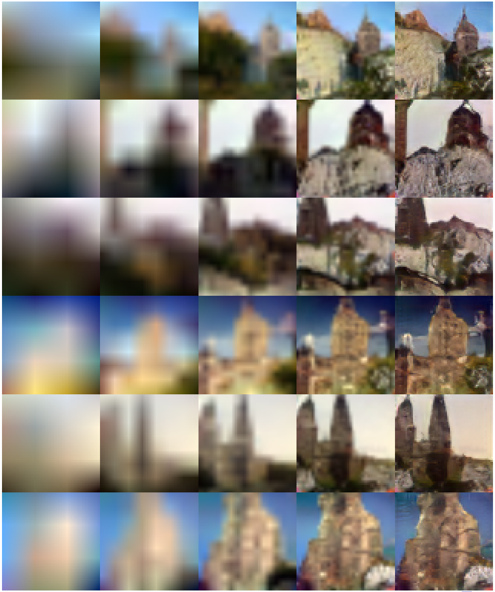


LSUN

- Large dataset of scenes, ∼10M images, 10 classes.



LSUN coarse-to-fine chain



LSUN church samples

LSUN tower samples

LSUN variability

Recent developments in GAN training

- Radford, Metz and Chintala (2015) propose several tricks to make GAN training more stable
  - http://arxiv.org/pdf/1511.06434v1.pdf

- Future work: apply same tricks to training of LAPGAN model to potenitally improve samples and produce higher resolution images

Conclusion

- Proposed a simple generative model that can produce decent quality samples of natural images
- Potential to be used as a decoder in autoencoder framework for unsupervised learning
- GAN framework is difficult to train, no clear objective function to track
- Code & demo: http://soumith.ch/eyescream

The End

Code & demo: http://soumith.ch/eyescream