# Generative Models:

## Recent Progresses and Applications

CEDL 2016

# References

- “A path to unsupervised learning through adversarial networks”
  - Soumith Chintala @ FAIR
  - [Soumith]
- “Learning deep generative models”
  - Ruslan Salakhutdinov
  - [Ruslan]

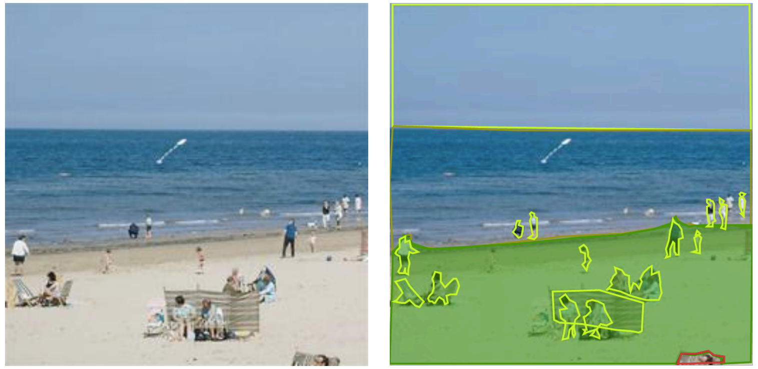
# Machine Learning 101: Recall

Generative, discriminative

Unsupervised, supervised, semi-supervised

Parametric models, non-parametric models

# Tackling Vision Tasks Using Discriminative Models (Usually Supervised)



LabelMe annotations of SUN dataset


Faster R-CNN object detector

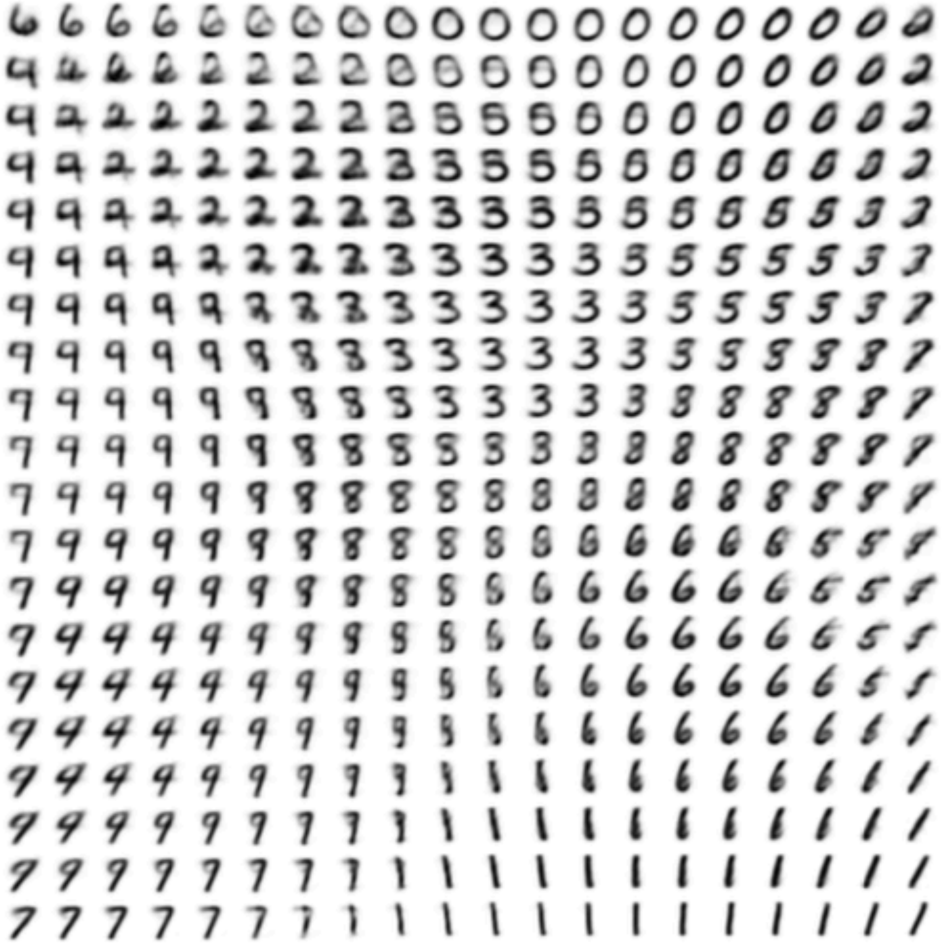
"man in black shirt is playing guitar."

Multimodal RNN
Image captioning

# Unsupervised Learned Generative Models

## Variational Auto-Encoder

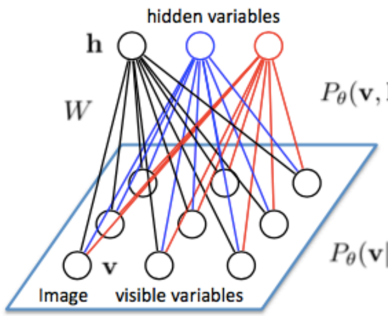
(a) Learned Frey Face manifold



(b) Learned MNIST manifold

# Unsupervised Learning of Image Representations

## Restricted Boltzmann Machines




$$P_\theta(\mathbf{v},\mathbf{h}) = \frac{1}{\mathcal{Z}(\theta)} \exp\left( \overbrace{\sum_{i=1}^{D}\sum_{j=1}^{F} W_{ij} v_i h_j}^{\text{Pair-wise}} + \overbrace{\sum_{i=1}^{D} v_i b_i + \sum_{j=1}^{F} h_j a_j}^{\text{Unary}} \right)$$

$$\theta = \{W, a, b\}$$

$$P_\theta(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^{D} P_\theta(v_i|\mathbf{h}) = \prod_{i=1}^{D} \frac{1}{1+\exp(-\sum_{j=1}^{F} W_{ij} v_i h_j - b_i)}$$

RBM is a Markov Random Field with:

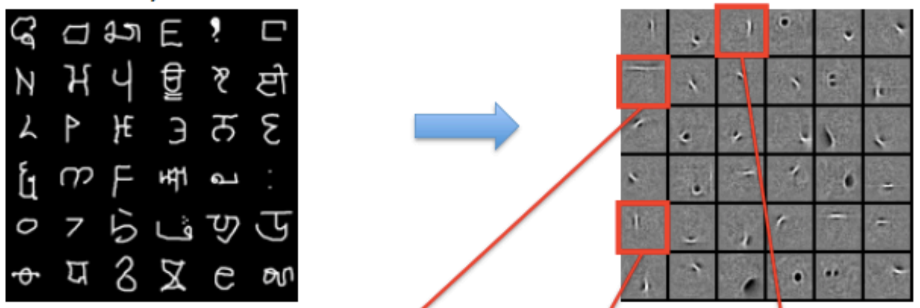
- Stochastic binary visible variables $\mathbf{v} \in \{0,1\}^D$.
- Stochastic binary hidden variables $\mathbf{h} \in \{0,1\}^F$.
- Bipartite connections.

Markov random fields, Boltzmann machines, log-linear models.

## Learning Features

Observed Data
Subset of 25,000 characters

Learned W: "edges"
Subset of 1000 features



Sparse representations

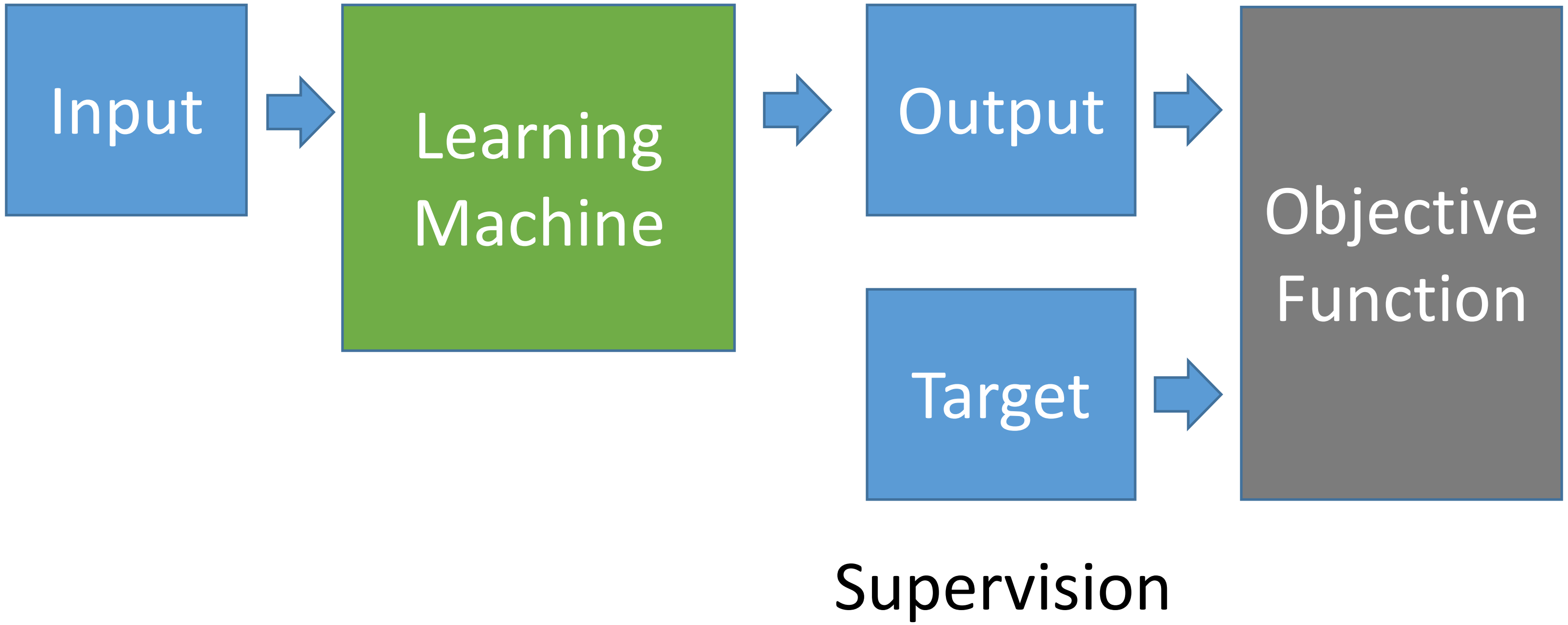
New Image: $p(h_7 = 1|v)$ $p(h_{29} = 1|v)$


$\sigma(x) = \frac{1}{1+\exp(-x)}$

Logistic Function: Suitable for modeling binary images

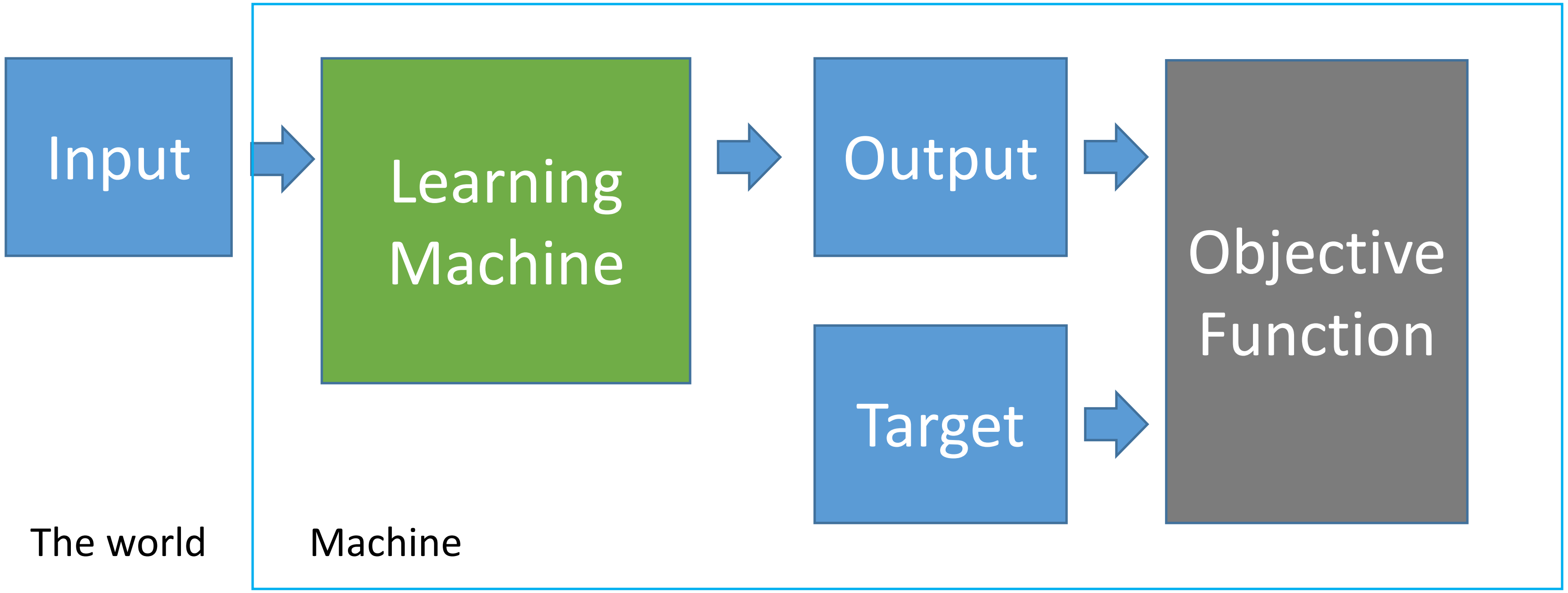
[Ruslan]

# Supervised Learning



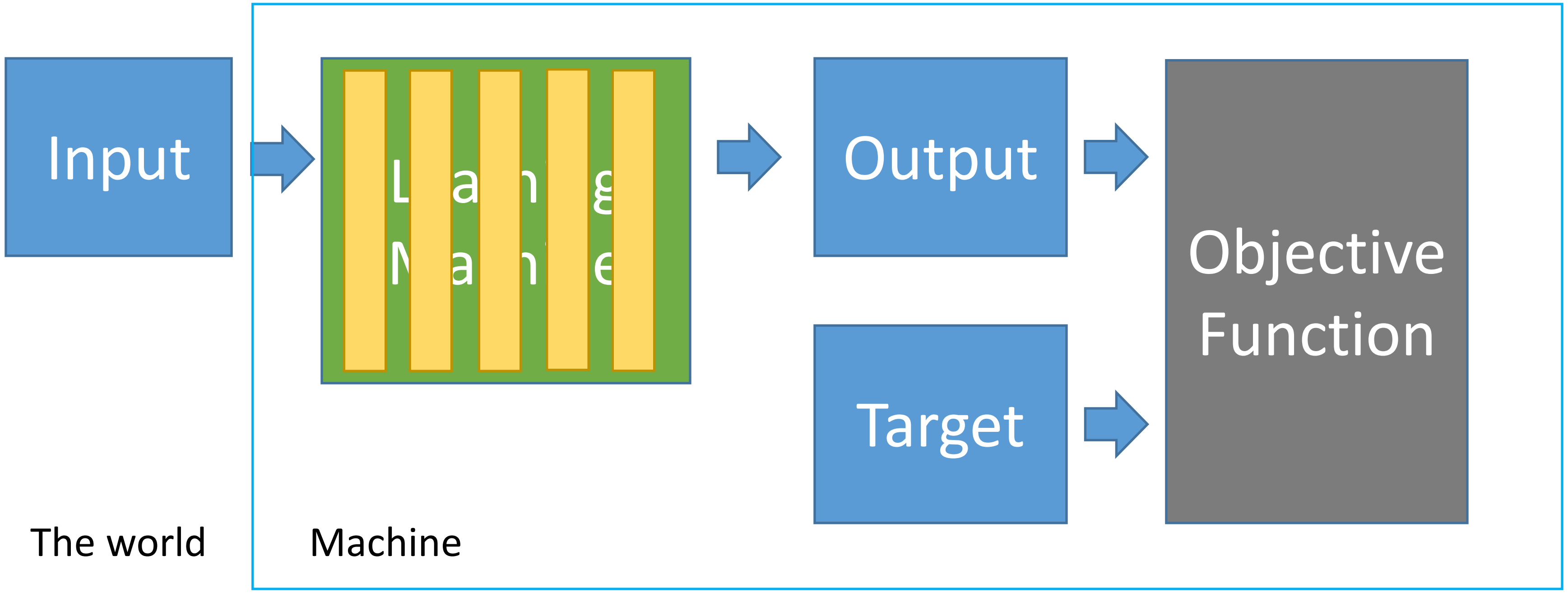

[Soumith]

# Unsupervised Learning



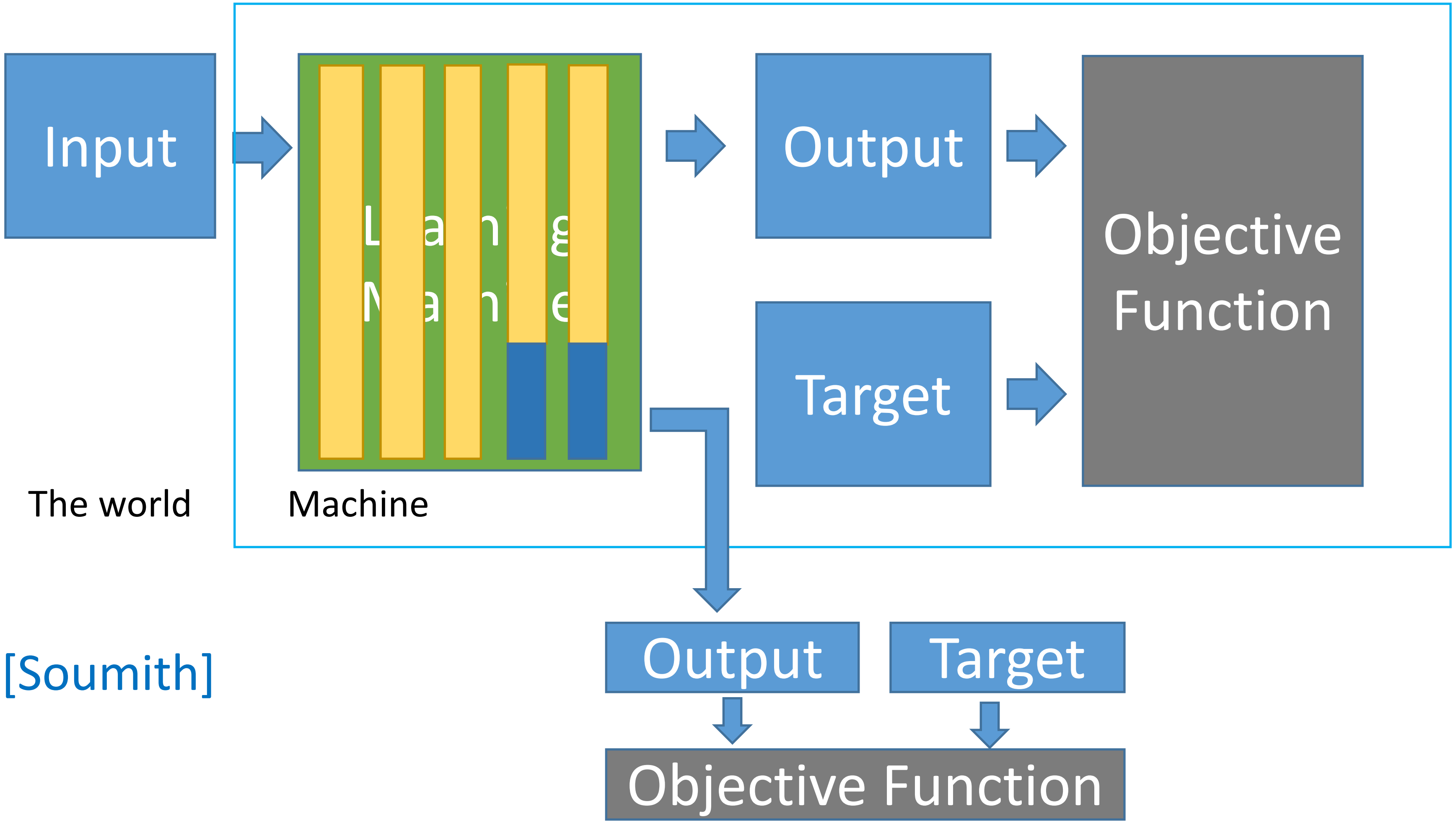

[Soumith]

# Unsupervised Learning




[Soumith]

# Unsupervised Learning

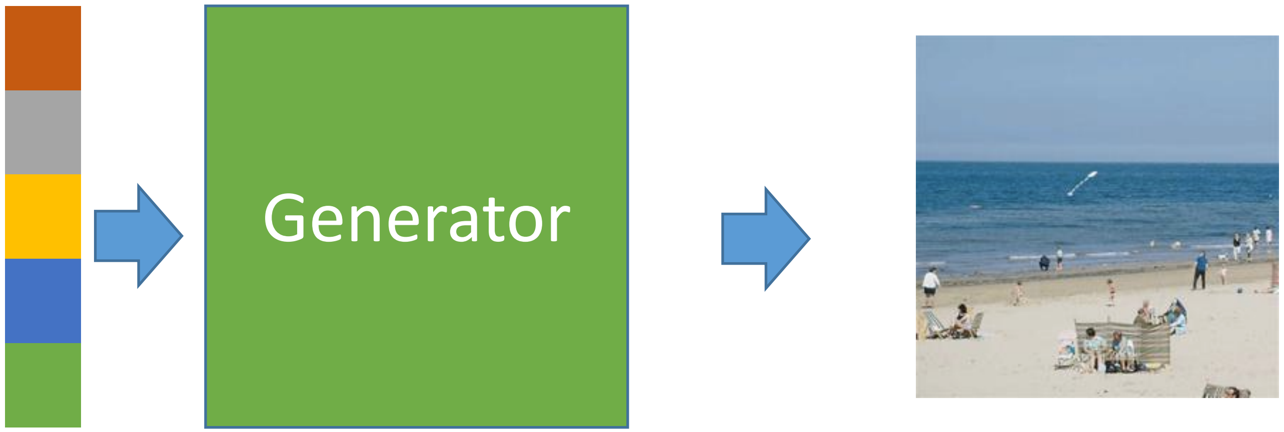



[Soumith]

# Generative Models


[Soumith]

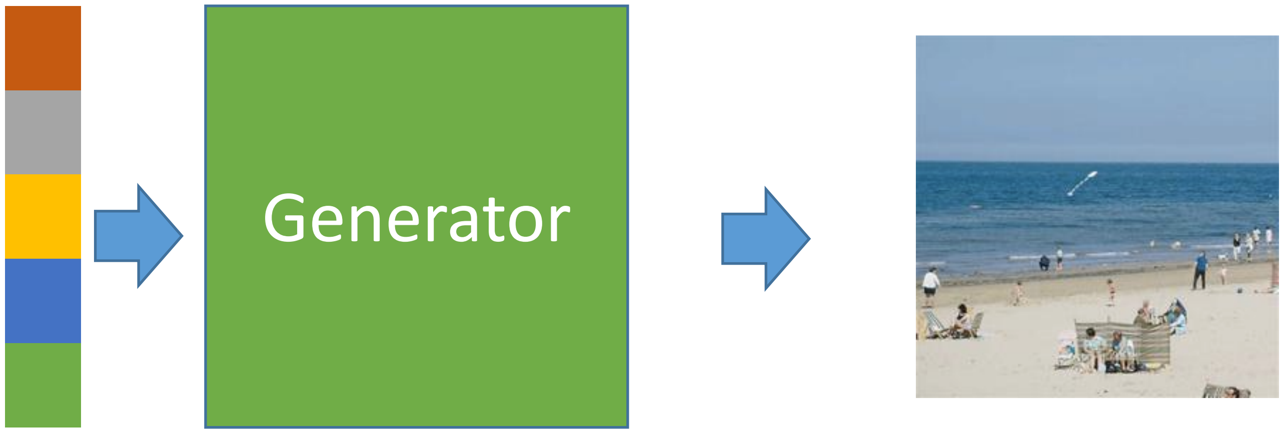
# Generative Models




$x = p(z)$, $z$ is a latent variable

[Soumith]

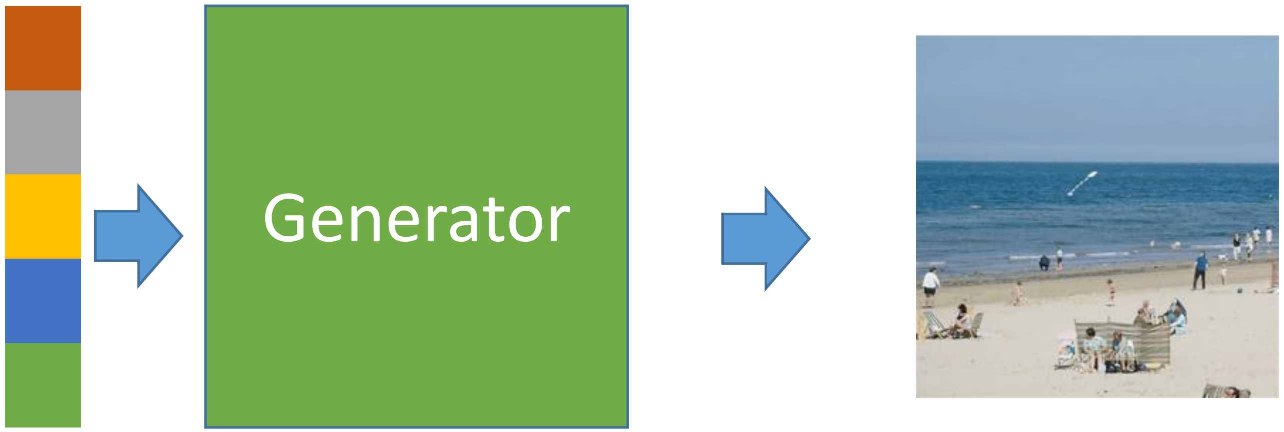
# Generative Models




p(z) is a neural network

[Soumith]

# Unsupervised and Generative




How good?

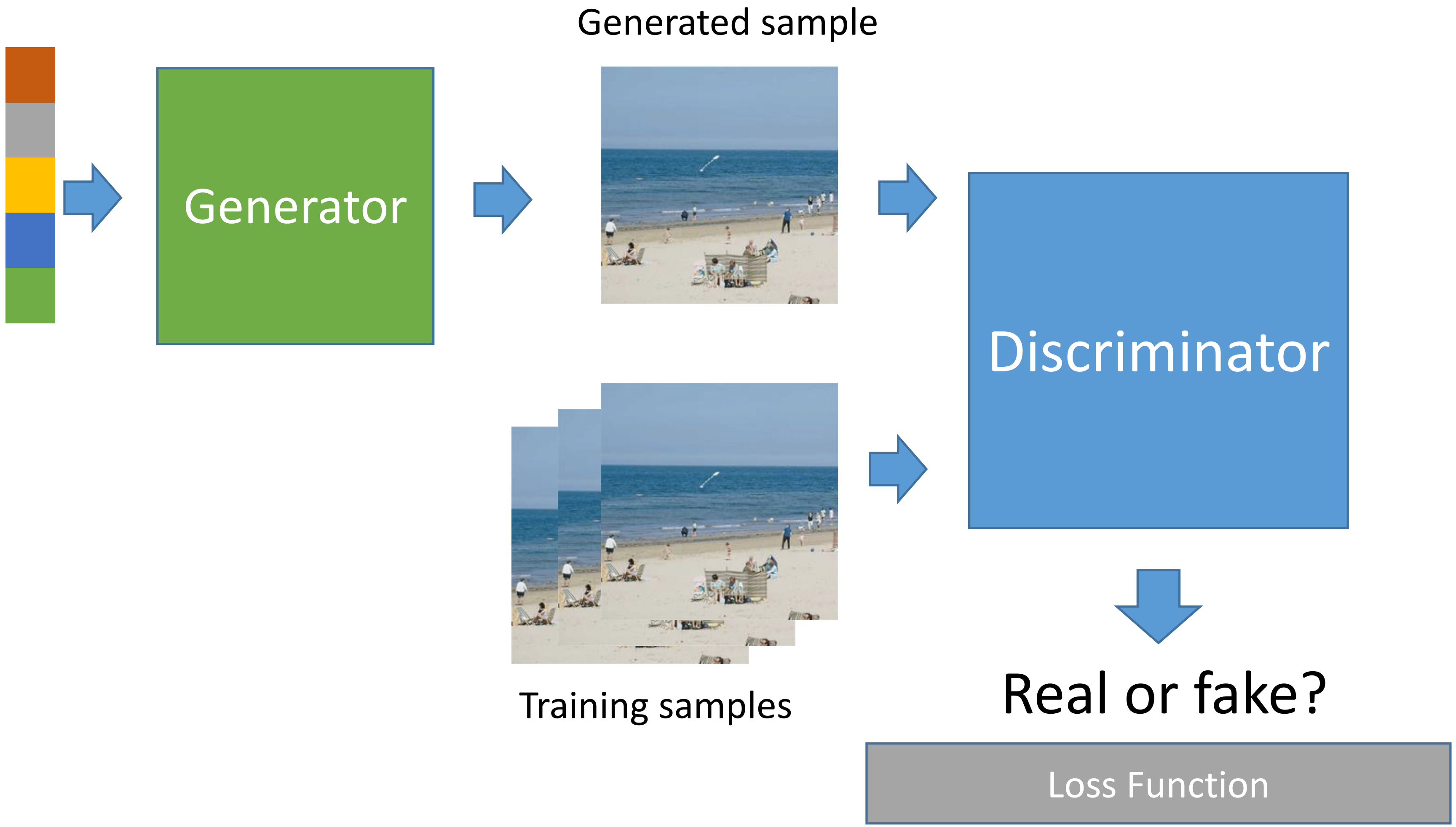
# Generative Adversarial Nets

**Ian J. Goodfellow, Jean Pouget-Abadie,[*] Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair,[†] Aaron Courville, Yoshua Bengio[‡]**
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

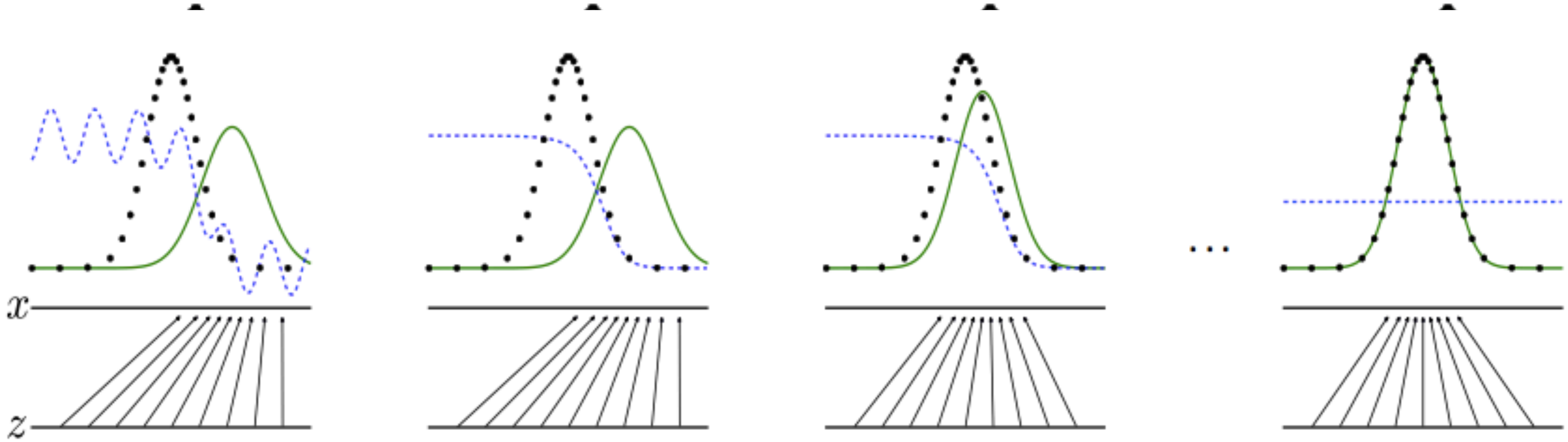
## Abstract

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model $G$ that captures the data distribution, and a discriminative model $D$ that estimates the probability that a sample came from the training data rather than $G$. The training procedure for $G$ is to maximize the probability of $D$ making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions $G$ and $D$, a unique solution exists, with $G$ recovering the training data distribution and $D$ equal to $\frac{1}{2}$ everywhere. In the case where $G$ and $D$ are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

# Generative Adversarial Nets (GANs)



$$\min_G \max_D V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$



**Algorithm 1** Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, $k$, is a hyperparameter. We used $k = 1$, the least expensive option, in our experiments.

---

**for** number of training iterations **do**
  **for** $k$ steps **do**
    • Sample minibatch of $m$ noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
    • Sample minibatch of $m$ examples $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ from data generating distribution $p_{\text{data}}(\boldsymbol{x})$.
    • Update the discriminator by ascending its stochastic gradient:

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

  **end for**
  • Sample minibatch of $m$ noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
  • Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right).$$

**end for**
The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

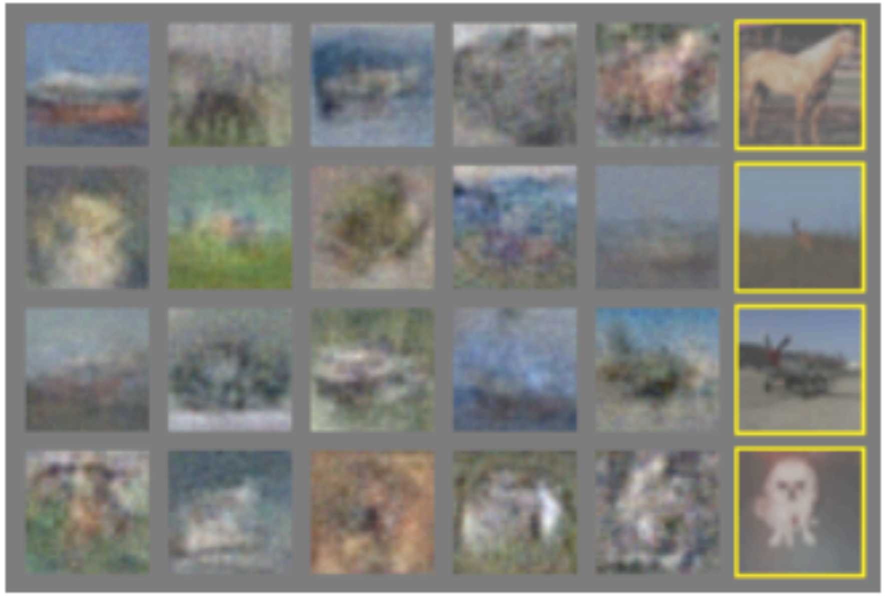
Generated by GAN

a)

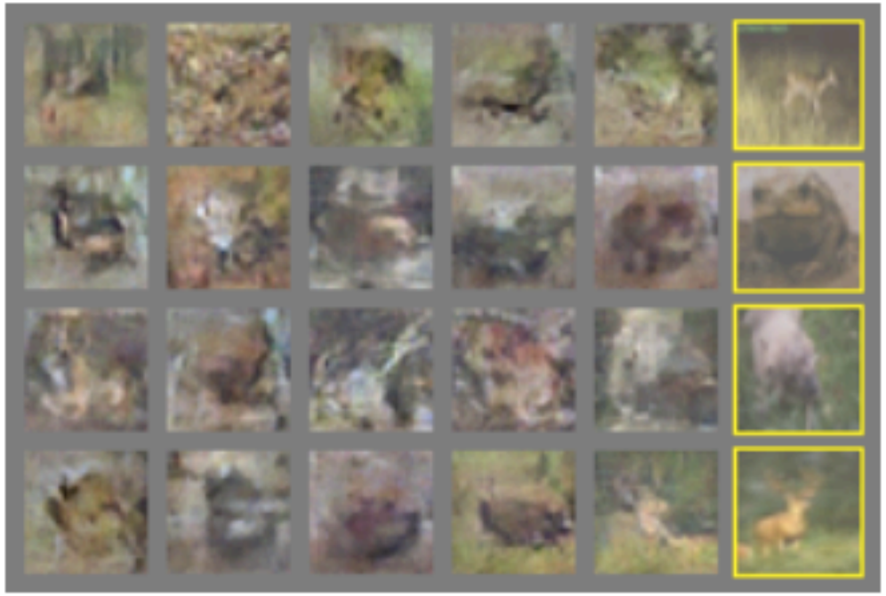
b)

Blurry

Collapsing into
a single point



c)



d)

# Extensions

- DCGAN
- InfoGAN
- Improved techniques

# UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS

**Alec Radford & Luke Metz**
indico Research
Boston, MA
{alec,luke}@indico.io

**Soumith Chintala**
Facebook AI Research
New York, NY
soumith@fb.com

## ABSTRACT

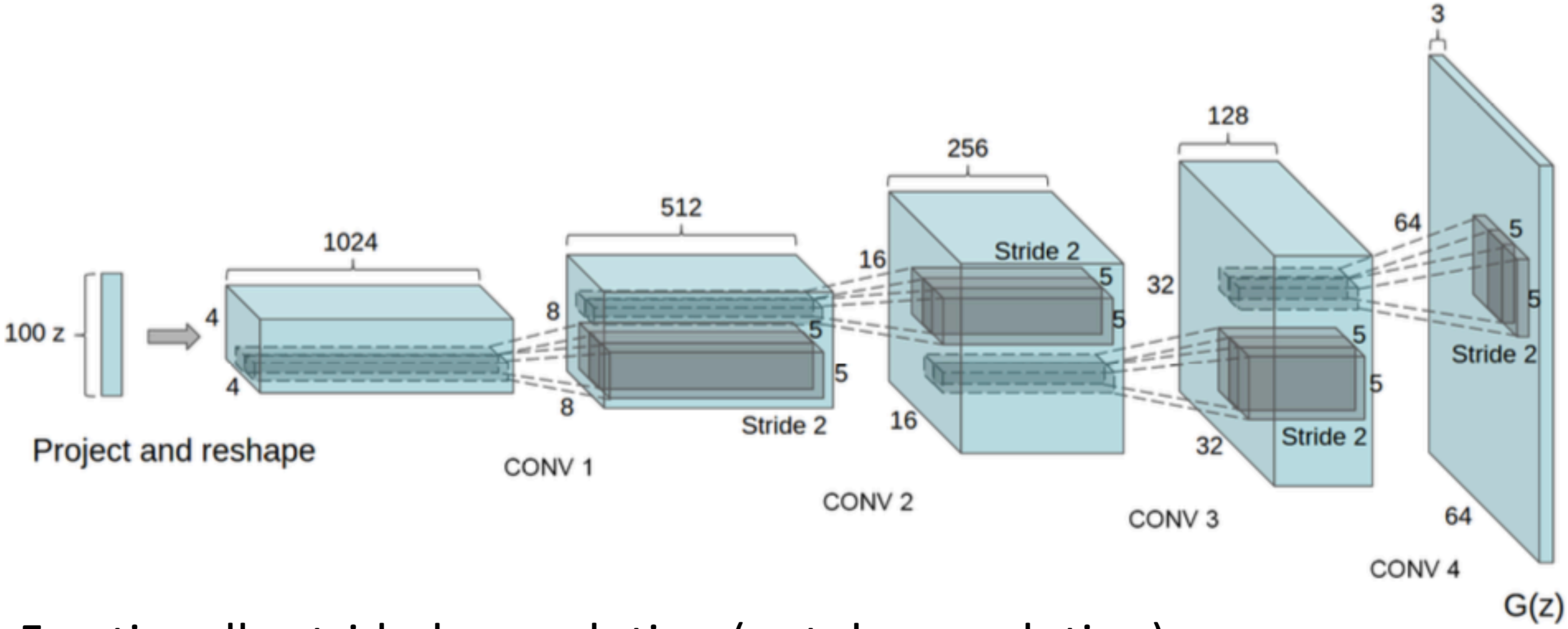
In recent years, supervised learning with convolutional networks (CNNs) has seen huge adoption in computer vision applications. Comparatively, unsupervised learning with CNNs has received less attention. In this work we hope to help bridge the gap between the success of CNNs for supervised learning and unsupervised learning. We introduce a class of CNNs called deep convolutional generative adversarial networks (DCGANs), that have certain architectural constraints, and demonstrate that they are a strong candidate for unsupervised learning. Training on various image datasets, we show convincing evidence that our deep convolutional adversarial pair learns a hierarchy of representations from object parts to scenes in both the generator and discriminator. Additionally, we use the learned features for novel tasks - demonstrating their applicability as general image representations.

# DCGAN

All convolutional
No spatial pooling
Eliminating fully connected
Batch normalization
ReLU for generator
Leaky ReLU for discriminator

Batch normalization not applied to generator output layer and discriminator input layer

DCGAN structure




Fractionally-strided convolution (not deconvolution)

# DCGAN results

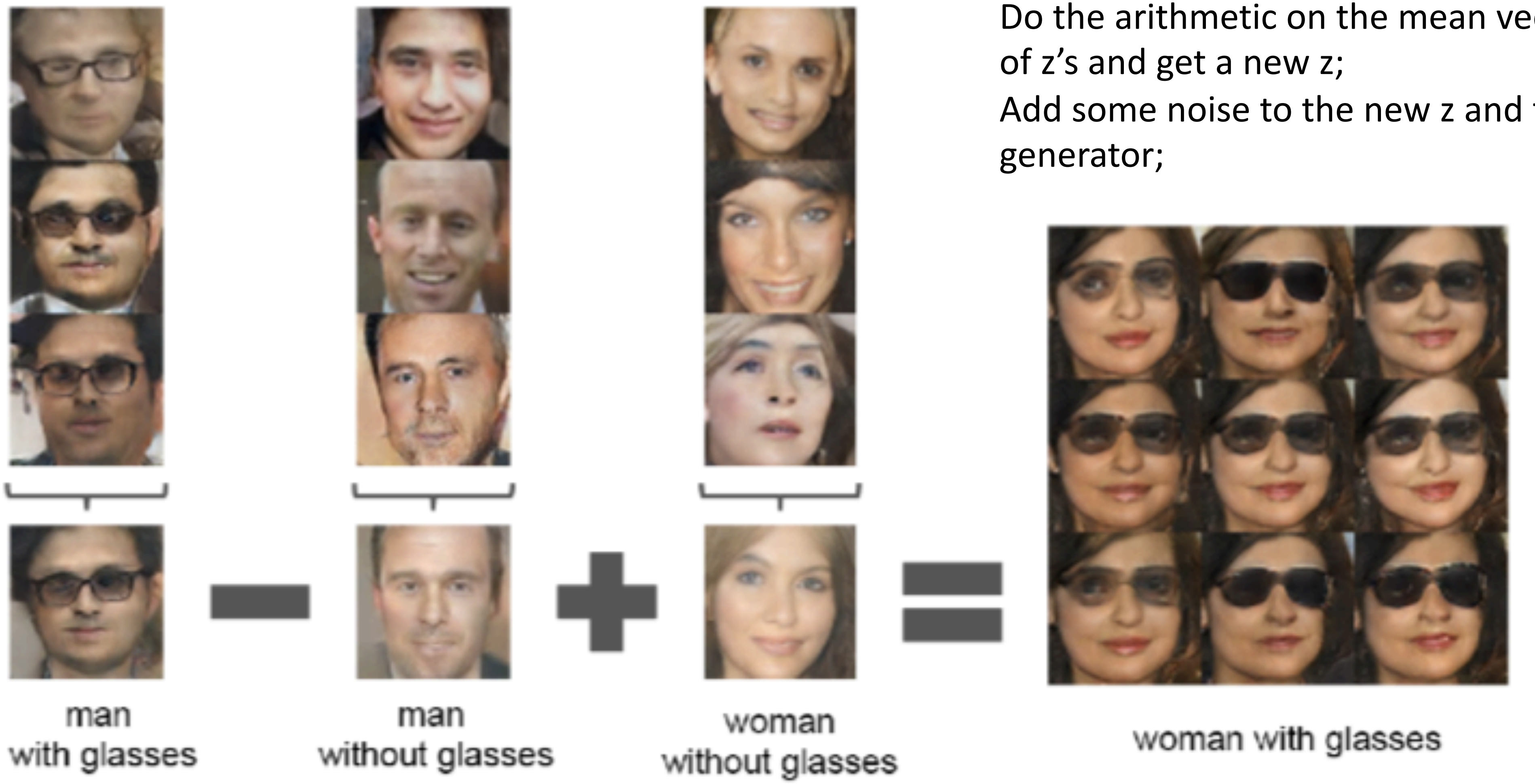
Memorize the training data?

# Walking in the Latent Space

# Visual Arithmetic

Get several z’s and compute their mean;
Feed the mean into the generator will get a mean image

Do the arithmetic on the mean vectors of different groups of z’s and get a new z;
Add some noise to the new z and feed them into the generator;



# InfoGAN

## InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets

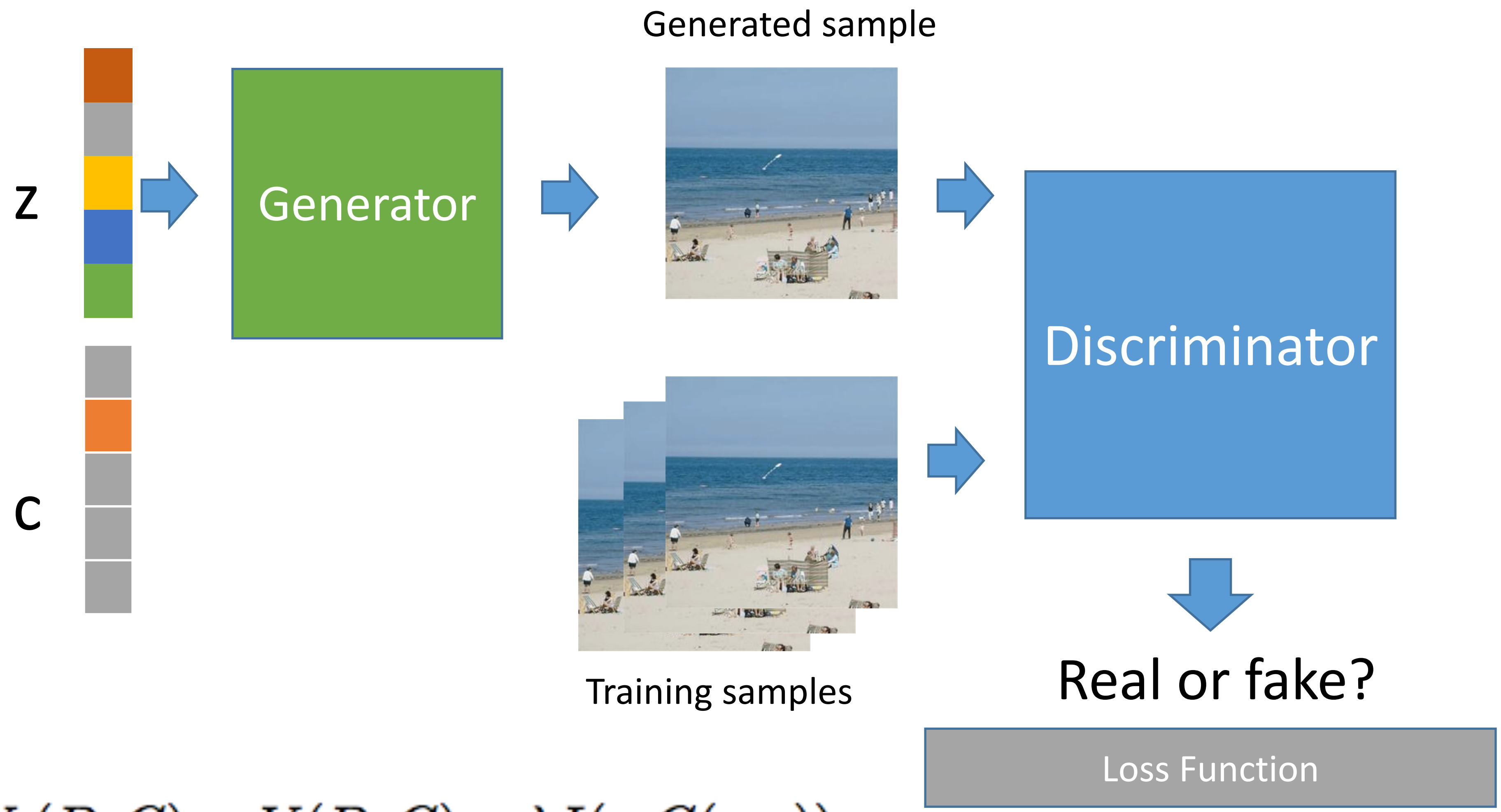
Xi Chen[†‡], Yan Duan[†‡], Rein Houthooft[†‡], John Schulman[†‡], Ilya Sutskever[‡], Pieter Abbeel[†‡]
† UC Berkeley, Department of Electrical Engineering and Computer Sciences
‡ OpenAI

### Abstract

This paper describes InfoGAN, an information-theoretic extension to the Generative Adversarial Network that is able to learn disentangled representations in a completely unsupervised manner. InfoGAN is a generative adversarial network that also maximizes the mutual information between a small subset of the latent variables and the observation. We derive a lower bound of the mutual information objective that can be optimized efficiently. Specifically, InfoGAN successfully disentangles writing styles from digit shapes on the MNIST dataset, pose from lighting of 3D rendered images, and background digits from the central digit on the SVHN dataset. It also discovers visual concepts that include hair styles, presence/absence of eyeglasses, and emotions on the CelebA face dataset. Experiments show that InfoGAN learns interpretable representations that are competitive with representations learned by existing supervised methods.

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

# Disentangling Representations




$$\min_G \max_D V_I(D,G) = V(D,G) - \lambda I(c; G(z,c))$$

(a) Rotation (b) Width

# Improved Techniques for Training GANs

**Tim Salimans**
tim@openai.com

**Ian Goodfellow**
ian@openai.com

**Wojciech Zaremba**
woj@openai.com

**Vicki Cheung**
vicki@openai.com

**Alec Radford**
alec.radford@gmail.com

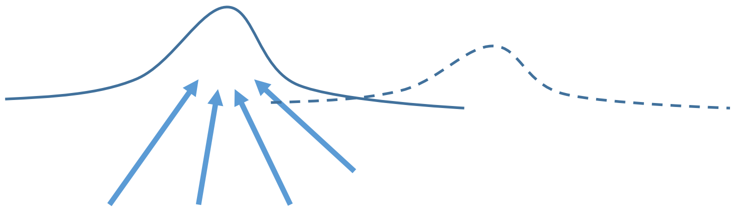
**Xi Chen**
peter@openai.com

## Abstract

We present a variety of new architectural features and training procedures that we apply to the generative adversarial networks (GANs) framework. We focus on two applications of GANs: semi-supervised learning, and the generation of images that humans find visually realistic. Unlike most work on generative models, our primary goal is not to train a model that assigns high likelihood to test data, nor do we require the model to be able to learn well without using any labels. Using our new techniques, we achieve state-of-the-art results in semi-supervised classification on MNIST, CIFAR-10 and SVHN. The generated images are of high quality as confirmed by a visual Turing test: our model generates MNIST samples that humans cannot distinguish from real data, and CIFAR-10 samples that yield a human error rate of 21.3%. We also present ImageNet samples with unprecedented resolution and show that our methods enable the model to learn recognizable features of ImageNet classes.

# Some Tricks

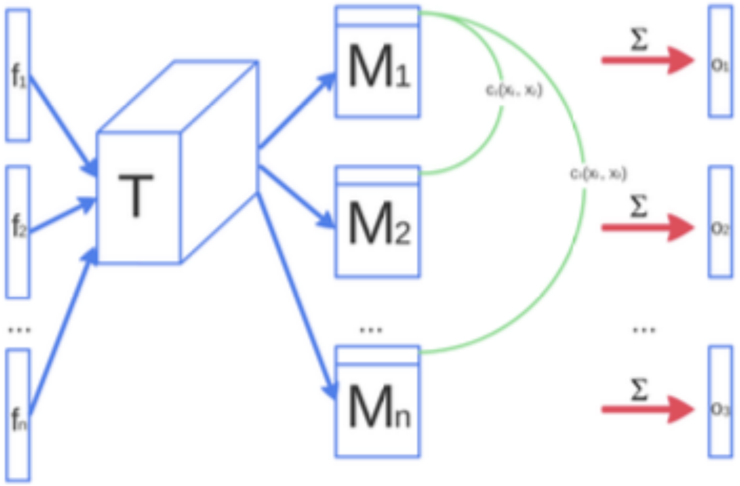
- Feature matching
  - A new objective: matching the expected value of the features on an intermediate layer of the discriminator
- Minibatch discrimination
  - Preventing collapsing to a single point
  - Detect closeness in a minibatch
  - Side information



$$c_b(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-||M_{i,b} - M_{j,b}||_{L_1})$$

$$o(\boldsymbol{x}_i)_b = \sum_{j=1}^{n} c_b(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathbb{R}$$

$$o(\boldsymbol{x}_i) = \left[o(\boldsymbol{x}_i)_1, o(\boldsymbol{x}_i)_2, \ldots, o(\boldsymbol{x}_i)_B\right]$$



# Some Tricks

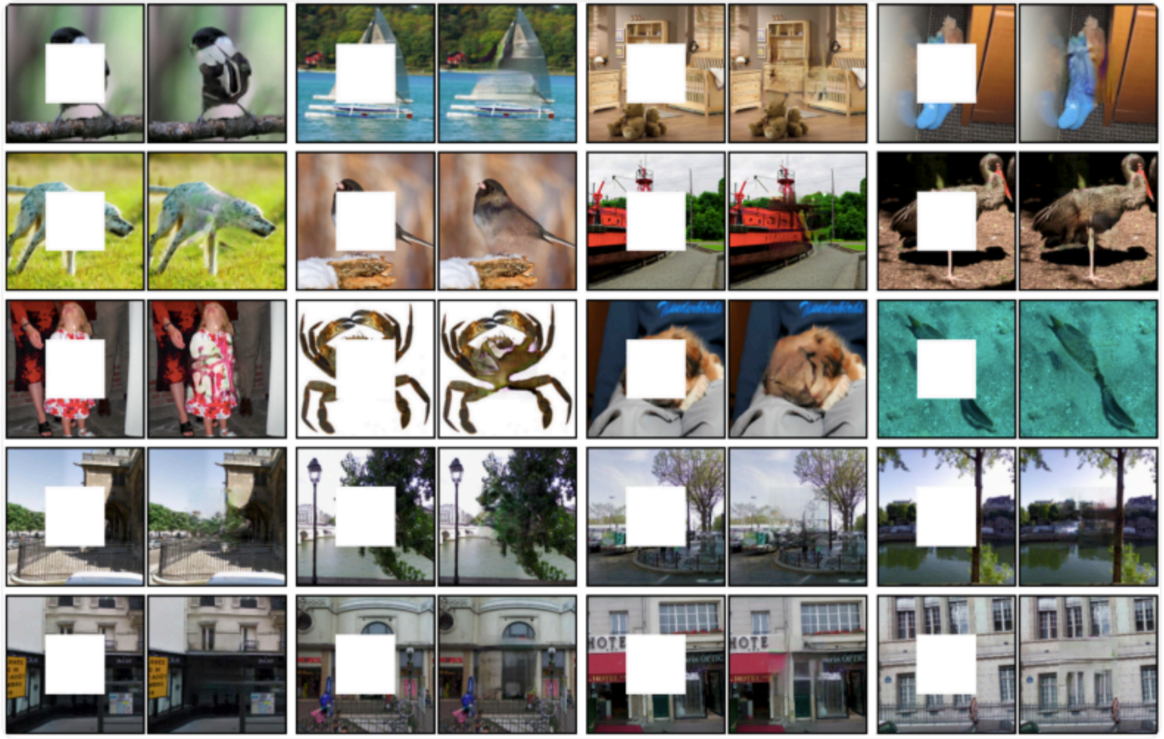
- Historical averaging $||\boldsymbol{\theta} - \frac{1}{t}\sum_{i=1}^{t}\boldsymbol{\theta}[i]||^2$
- One-sided label smoothing

$$D(\boldsymbol{x}) = \frac{\alpha p_{\text{data}}(\boldsymbol{x}) + \beta p_{\text{model}}(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_{\text{model}}(\boldsymbol{x})}$$

- Semi-supervised training
  - Add the (K+1)th class

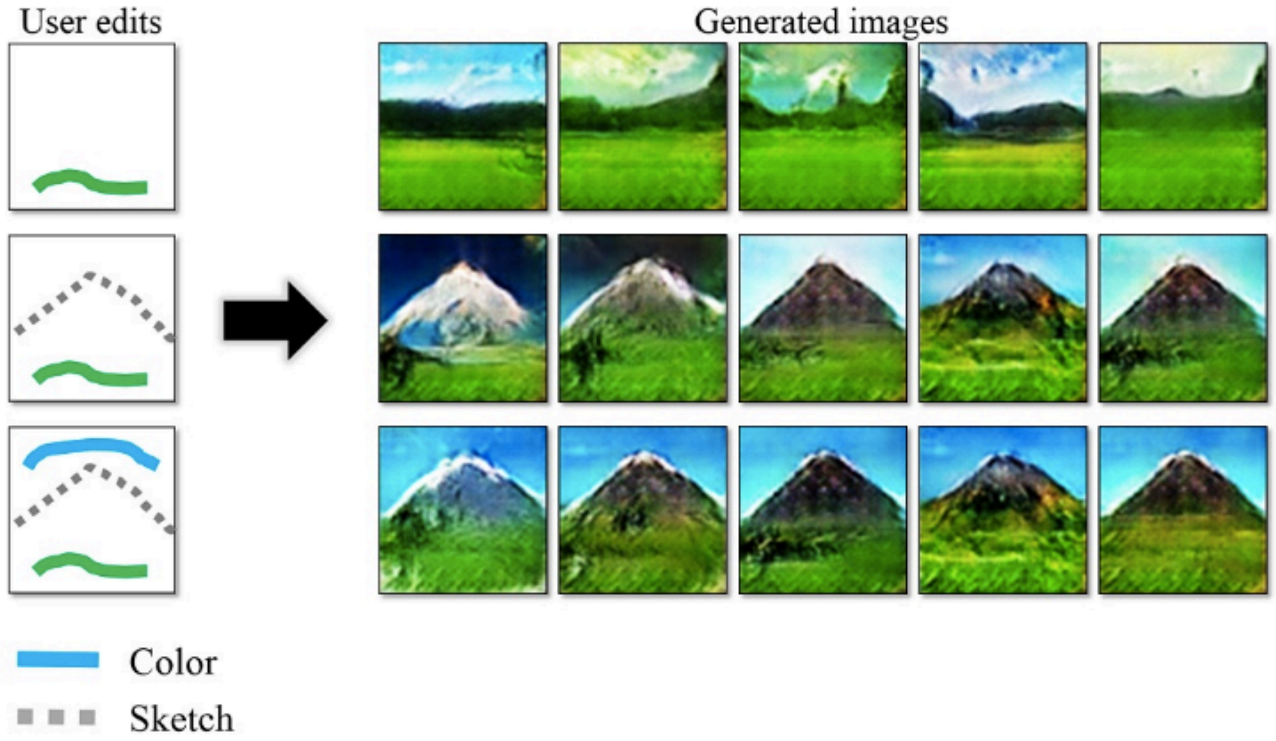
# Applications of GAN-alike Architectures in Computer Vision

- Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell and Alexei A. Efros. **Context Encoders: Feature Learning by Inpainting.** In *CVPR 2016*.



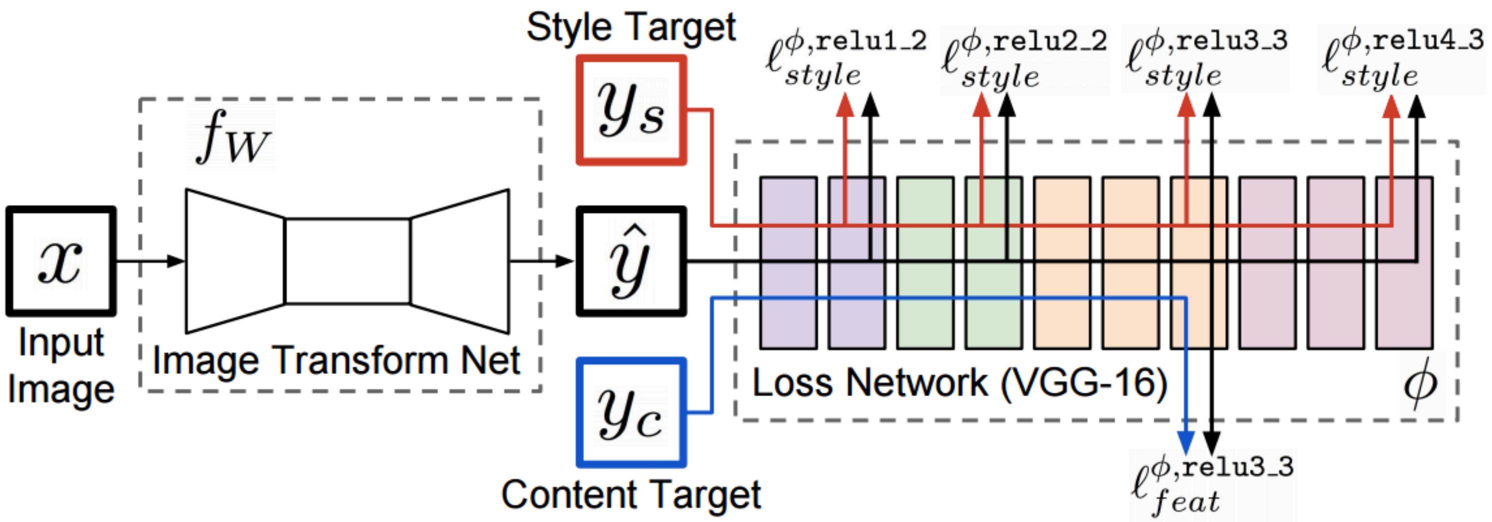
# Applications of GAN-alike Architectures in Computer Vision

- Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman and Alexei A. Efros. "Generative Visual Manipulation on the Natural Image Manifold", in European Conference on Computer Vision (**ECCV**). 2016.



# Applications of GAN-alike Architectures in Computer Vision

- Johnson, Justin and Alahi, Alexandre and Fei-Fei, Li ,“Perceptual losses for real-time style transfer and super-resolution”, ECCV 2016.




# Prof. Bengio’s Talk at ECCV 2016

- Key ingredients for ML towards AI
  - Lots of data
  - Very flexible models
  - Enough computing power
  - Computationally efficient inference
  - Powerful priors that can defeat the curse of dimensionality
- **Prior assumption: compositionality is useful to describe the world around us efficiently**

# Prof. Bengio's Talk at ECCV 2016

- The next big challenge: unsupervised learning
  - Recent progress mostly in supervised DL
  - Real technical challenges for unsupervised DL
  - Potential benefits
    - Exploit tons of unlabeled
    - Answer new questions about the variables observed
    - Regularizer -- transfer learning -- domain adaptation
    - Easier optimization (local training signal)
    - Structured outputs (compositional, joint distributions of many things)
    - Necessary for RL without given model or domain simulator

# Prof. Bengio's Talk at ECCV 2016

- Bypassing normalization constants with generative black boxes
    - Instead of parameterizing p(x), parametrize a machine which generates samples
    - Goodfellow et al. NIPS 2014, GAN) for the case of ancestral sampling in a deep generative net. Variational autoencoders are closely related
        - Also Generative moment matching networks
    - Generative Stochastic networks (Bengio et al., ICML 2014), learning the transition operator of a Markov chain

# Prof. Malik’s Talk at ECCV 2016

- Prediction is a sign of understanding
- Planning actions

- What has been responsible for recent AI success?
  - Big computing
  - Big data
  - Big annotation
  - Big simulation
    - Game scenarios can be simulated, but it’s not so easy in other settings
- Being interested in computer vision tasks where feedback is key to the solution. This is a very natural way to capture “context”