# Conditional Generative Adversarial Networks

Jia-Bin Huang

Virginia Tech

ECE 6554 Advanced Computer Vision

# Today's class

- Discussions
- Review basic ideas of GAN
- Examples of conditional GAN
- Experiment presentation by Sanket

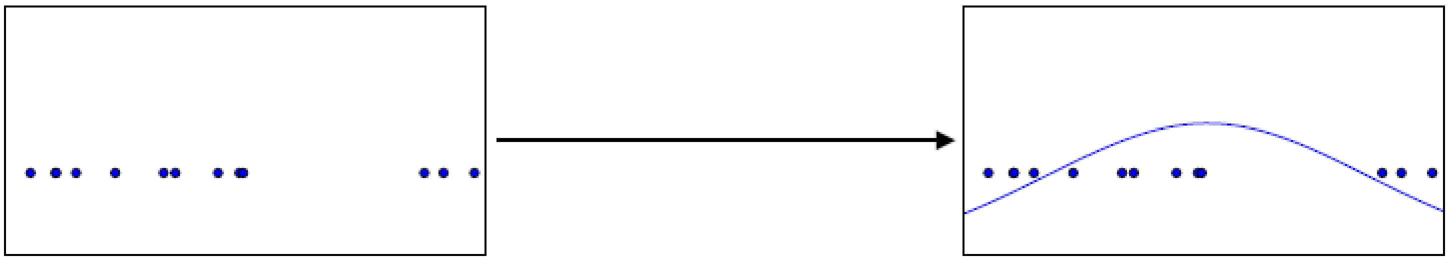
# Why Generative Models?

- Excellent test of our ability to use high-dimensional, complicated probability distributions
- Simulate possible futures for planning or simulated RL
- Missing data
    - Semi-supervised learning
- Multi-modal outputs
- Realistic generation tasks

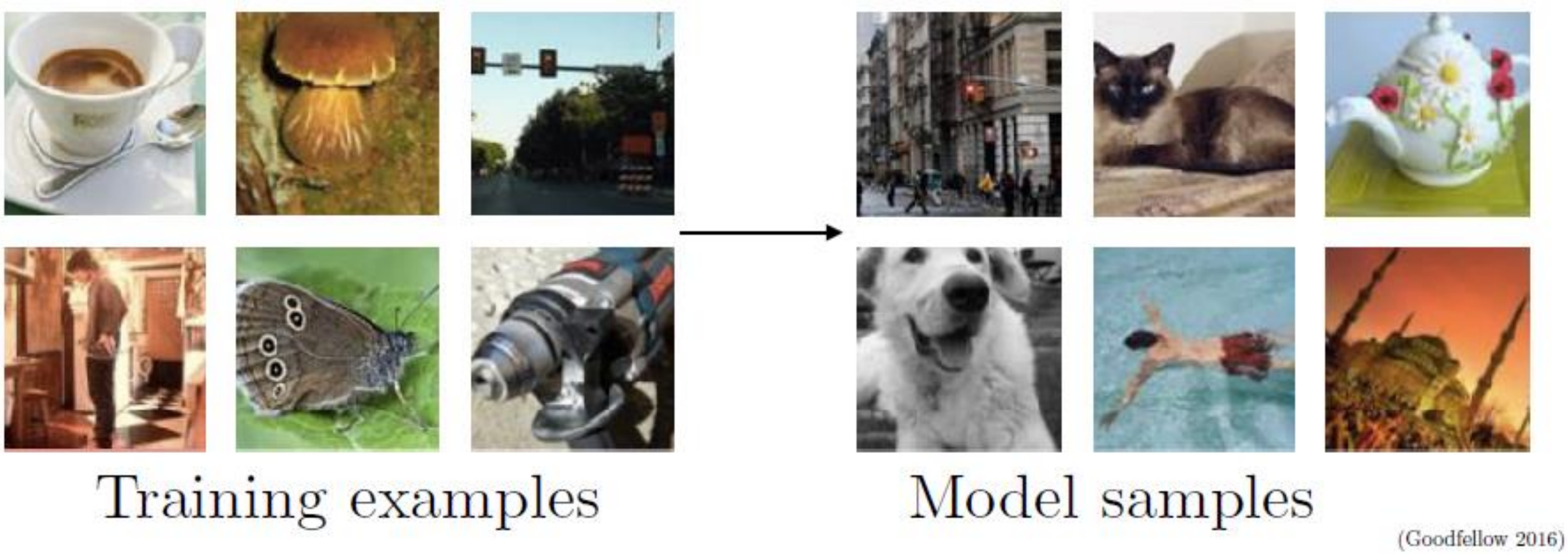
(Goodfellow 2016)

# Generative Modeling

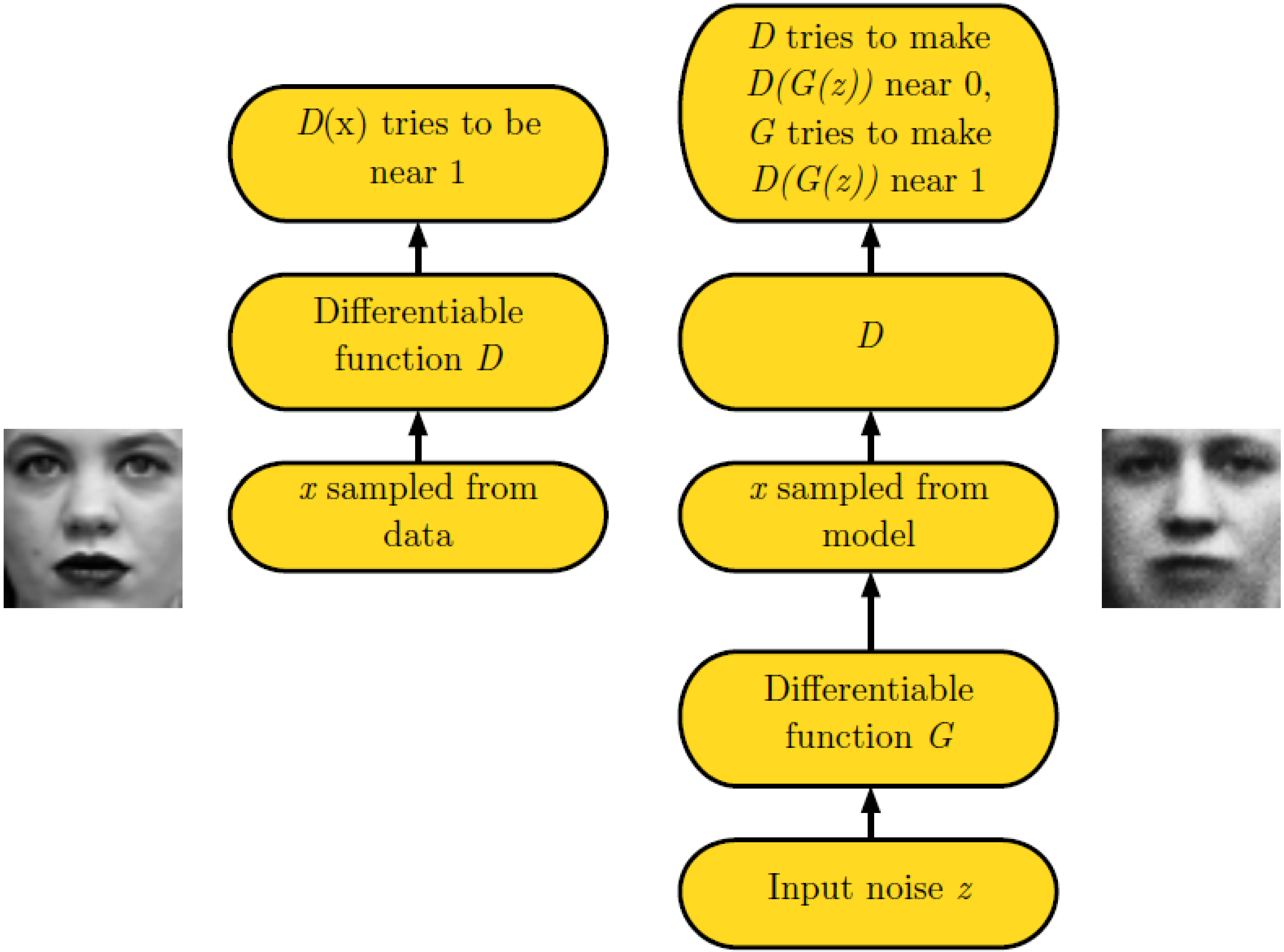
- Density estimation



- Sample generation



# Adversarial Nets Framework




(Goodfellow 2016)

# Training Procedure

- Use SGD-like algorithm of choice (Adam) on two minibatches simultaneously:
  - A minibatch of training examples
  - A minibatch of generated samples
- Optional: run *k* steps of one player for every step of the other player.

(Goodfellow 2016)

# Minimax Game

$$J^{(D)} = -\frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}} \log D(\boldsymbol{x}) - \frac{1}{2}\mathbb{E}_{\boldsymbol{z}} \log\left(1 - D\left(G(\boldsymbol{z})\right)\right)$$
$$J^{(G)} = -J^{(D)}$$

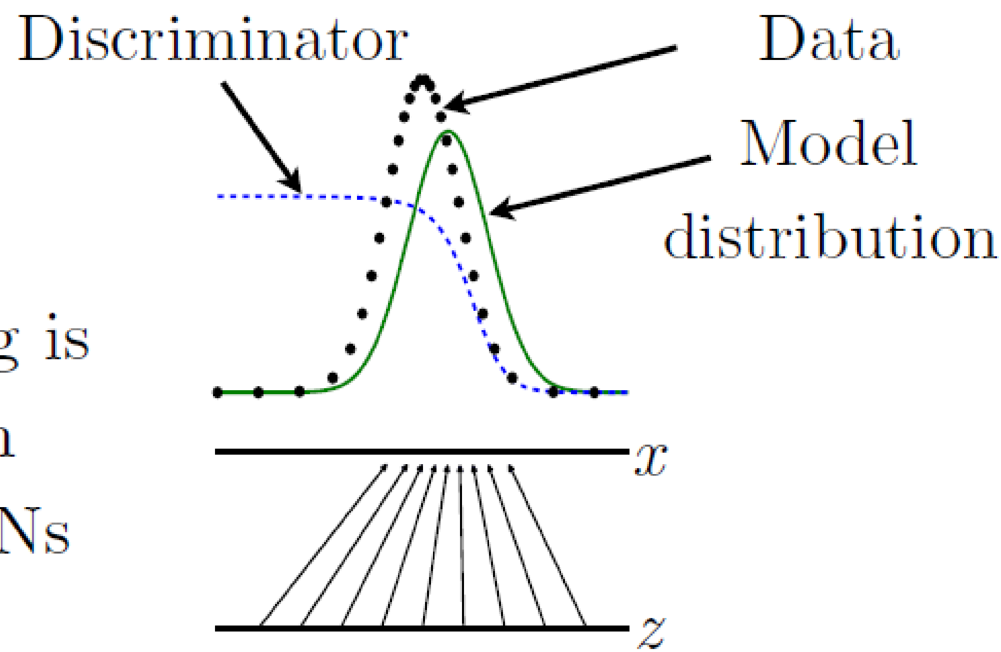
- Equilibrium is a saddle point of the discriminator loss
- Resembles Jensen-Shannon divergence
- Generator minimizes the log-probability of the discriminator being correct

(Goodfellow 2016)

# Discriminator Strategy

- Optimal discriminator

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




Estimating this ratio using supervised learning is the key approximation mechanism used by GANs

(Goodfellow 2016)

# Non-Saturating Game

$$J^{(D)} = -\frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}} \log D(\boldsymbol{x}) - \frac{1}{2}\mathbb{E}_{\boldsymbol{z}} \log\left(1 - D\left(G(\boldsymbol{z})\right)\right)$$

$$J^{(G)} = -\frac{1}{2}\mathbb{E}_{\boldsymbol{z}} \log D\left(G(\boldsymbol{z})\right)$$

- Equilibrium no longer describable with a single loss
- Generator maximizes the log-probability of the discriminator being mistaken
- Heuristically motivated; generator can still learn even when discriminator successfully rejects all generator samples

(Goodfellow 2016)

# Review: GAN

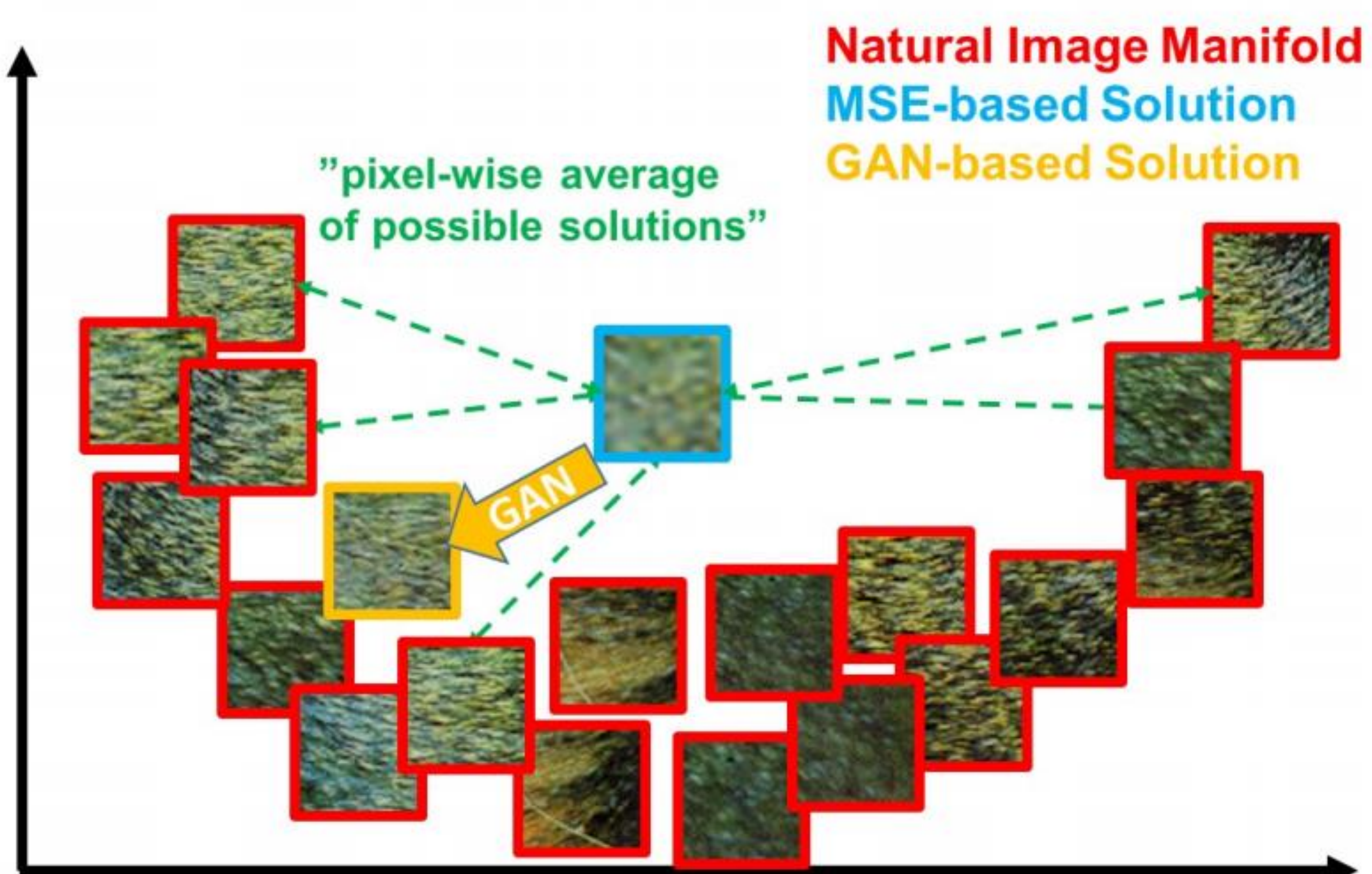
- GANs are generative models that use supervised learning to approximate an intractable cost function

- GANs can simulate many cost functions, including the one used for maximum likelihood

- Finding Nash equilibria in high-dimensional, continuous, nonconvex games is an important open research problem

# Conditional GAN

- Learn P(Y|X)




[Ledig et al. CVPR 2017]

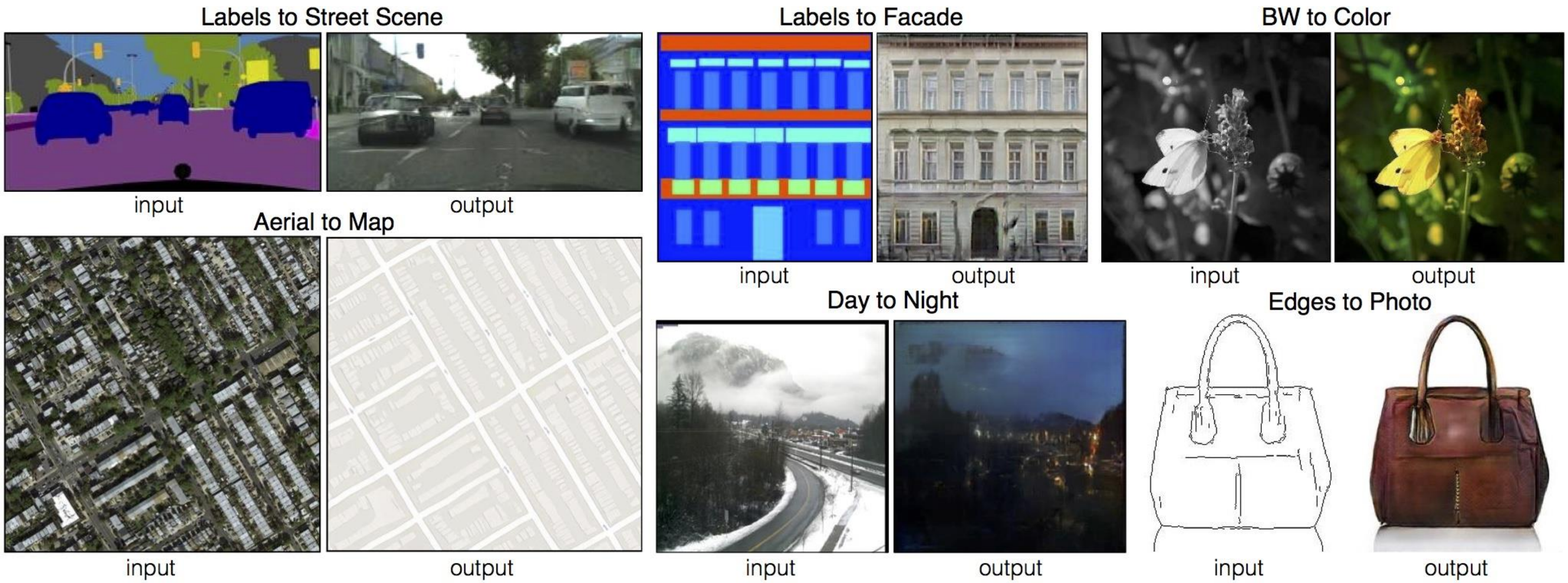
# Image Super-Resolution

bicubic (21.59dB/0.6423) | SRResNet (23.53dB/0.7832) | SRGAN (21.15dB/0.6868) | original

- Conditional on low-resolution input image

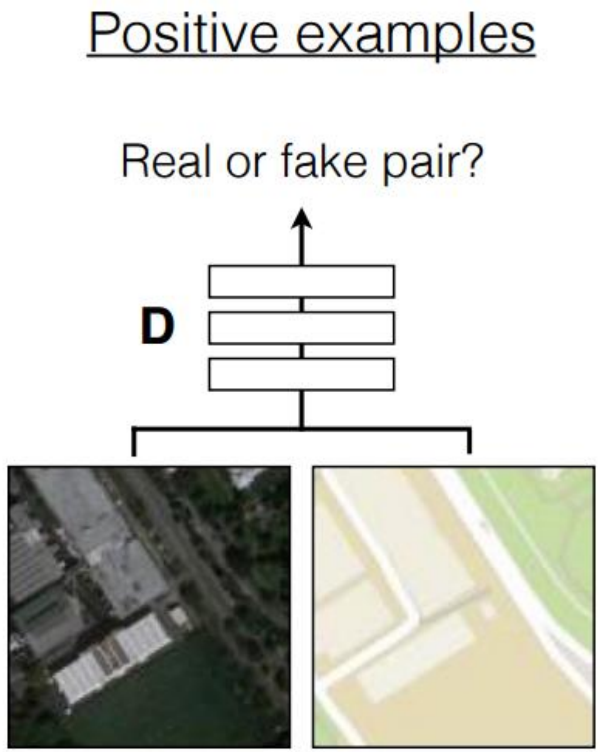
[Ledig et al. CVPR 2017]

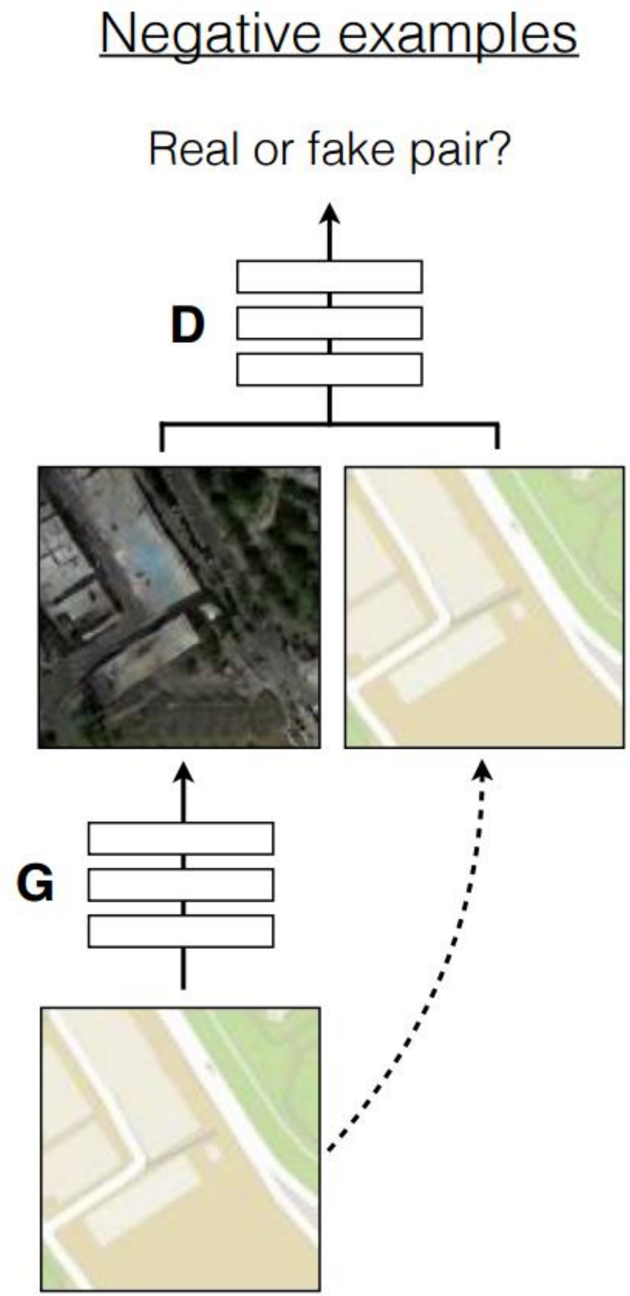
# Image-to-Image Translation




- Conditioned on an image of different modality
- No need to specify the loss function

[Isola et al. CVPR 2017]



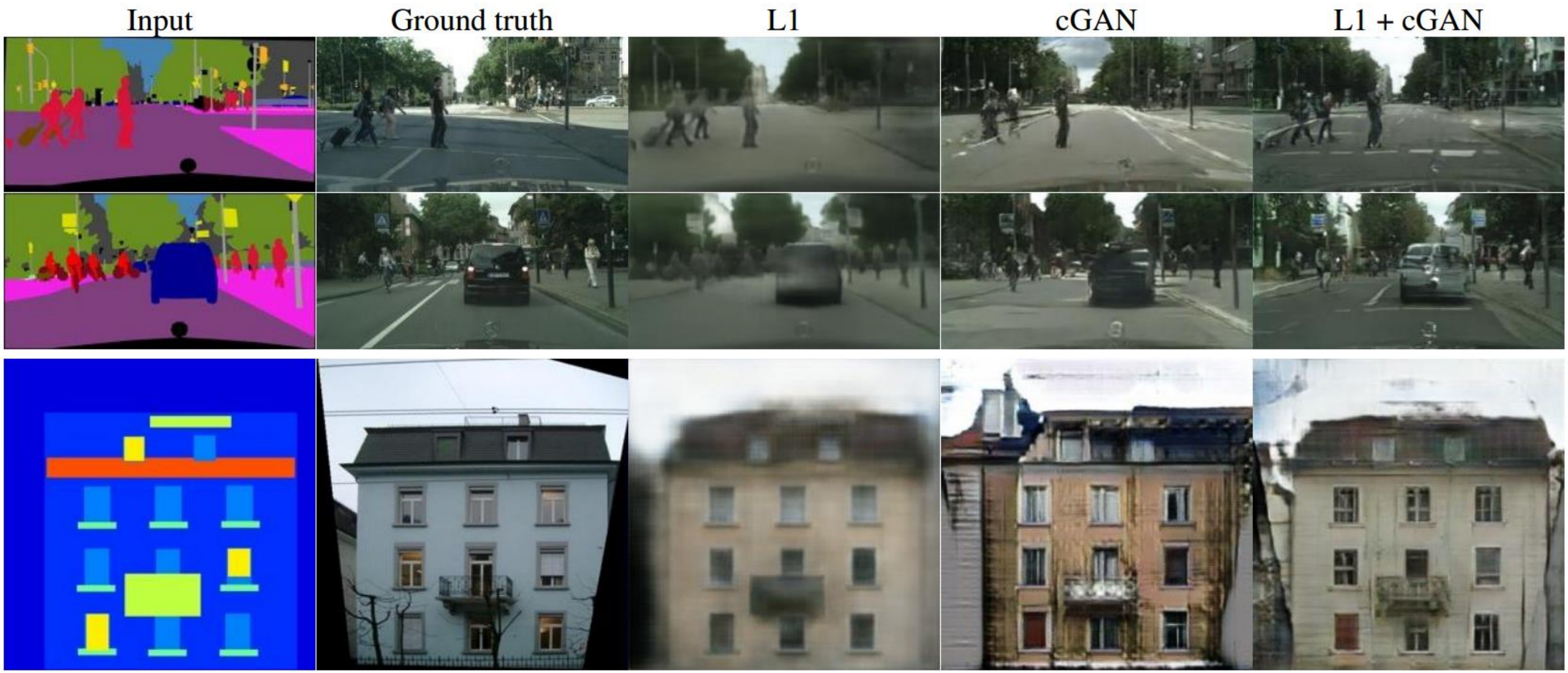




**G** tries to synthesize fake images that fool **D**

**D** tries to identify the fakes

[Isola et al. CVPR 2017]

# Label2Image

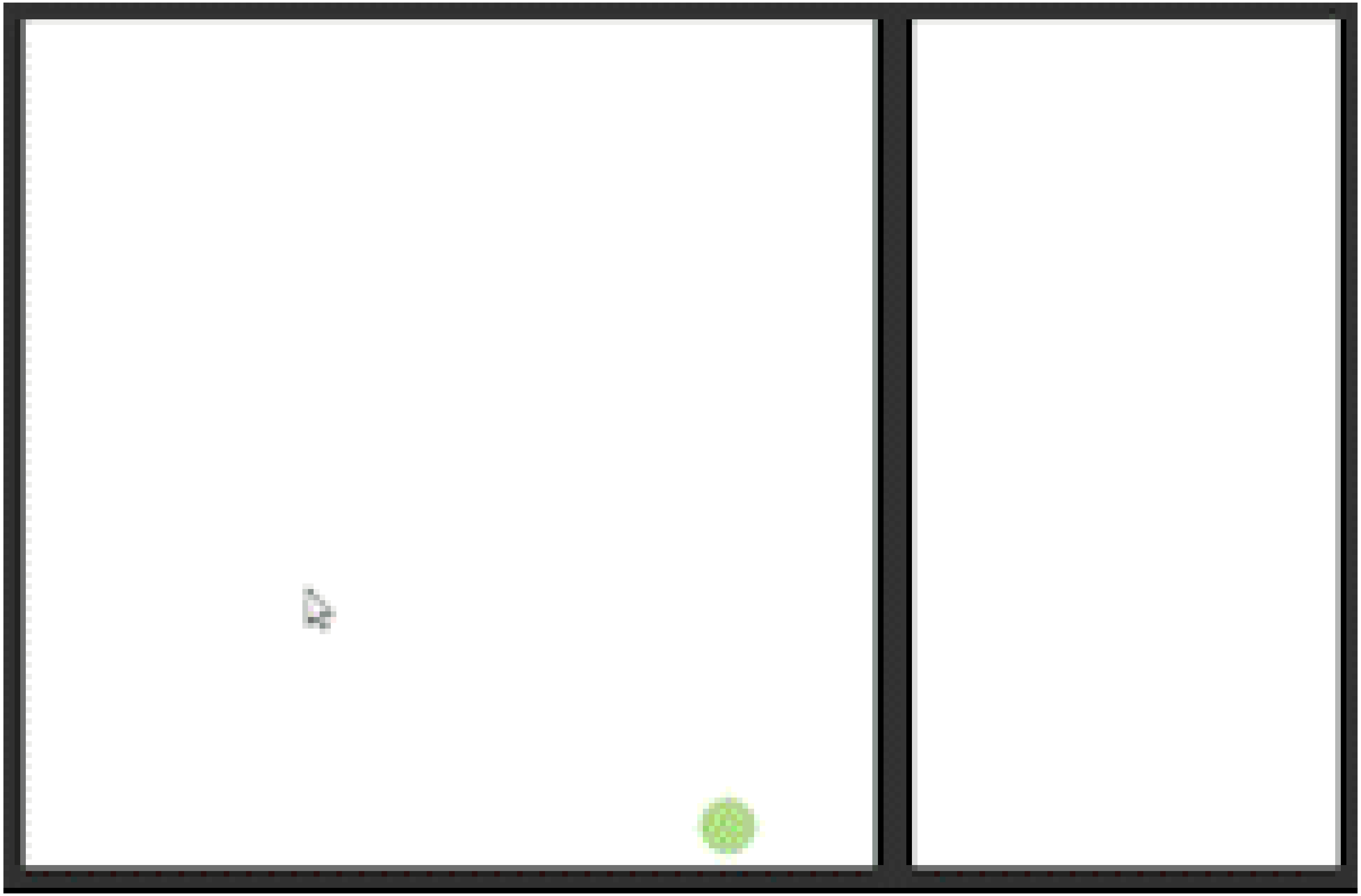



[Isola et al. CVPR 2017]

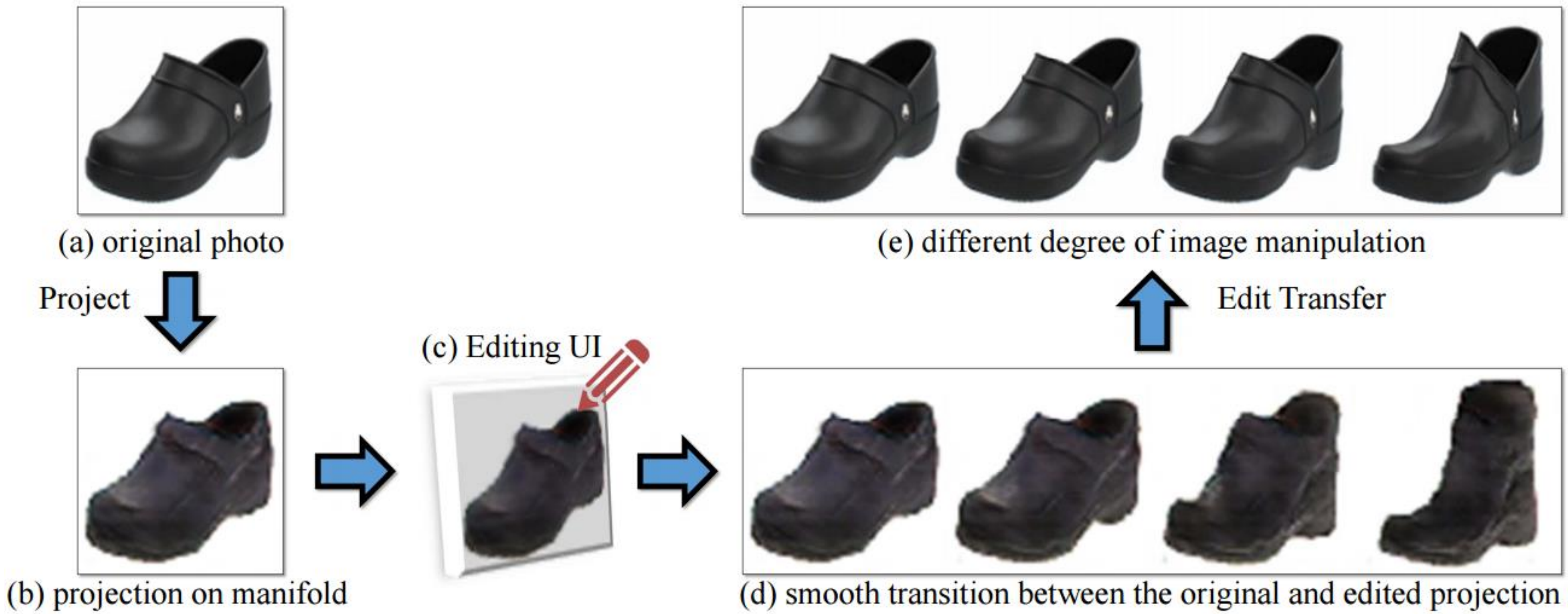
# Edges2Image


[Isola et al. CVPR 2017]

# Generative Visual Manipulation



[Zhu et al. ECCV 2016]




$$z^* = \arg\min_{z \in \mathbb{Z}} \Big\{ \underbrace{\sum_g \| f_g(G(z)) - v_g \|^2}_{\text{data term}} + \underbrace{\lambda_s \cdot \| z - z_0 \|^2}_{\text{manifold smoothness}} + E_D \Big\}$$

[Zhu et al. ECCV 2016]

# Text2Image

this small bird has a pink breast and crown, and black primaries and secondaries.

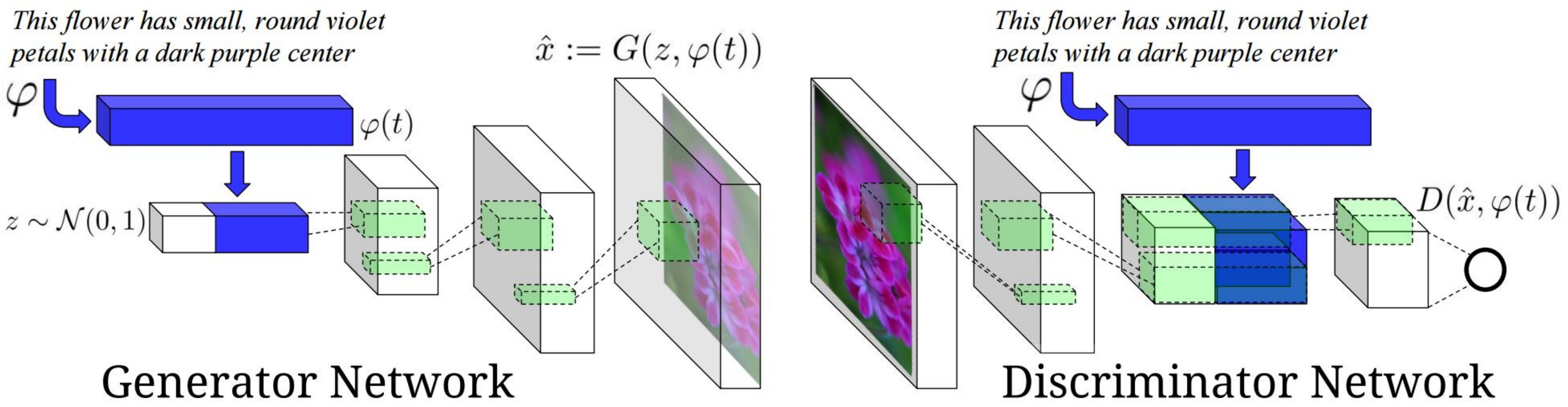
this magnificent fellow is almost all black with a red crest, and white cheek patch.

the flower has petals that are bright pinkish purple with white stigma

this white and yellow flower have thin white petals and a round yellow stamen

[Reed et al. ICML 2016]

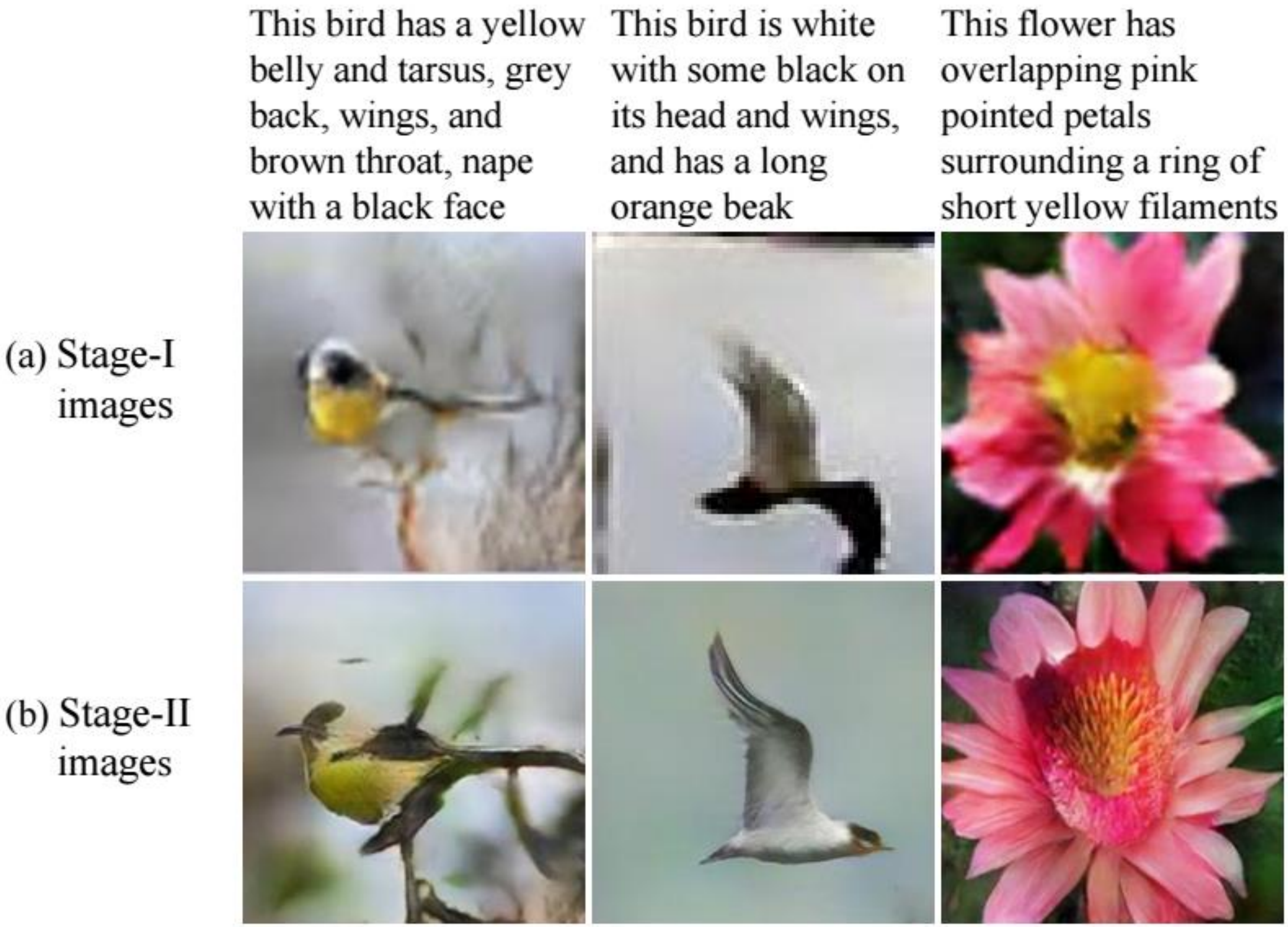
# Text2Image




- Positive samples:
  - real image + right texts
- Negative samples:
  - fake image + right texts
  - Real image + wrong texts

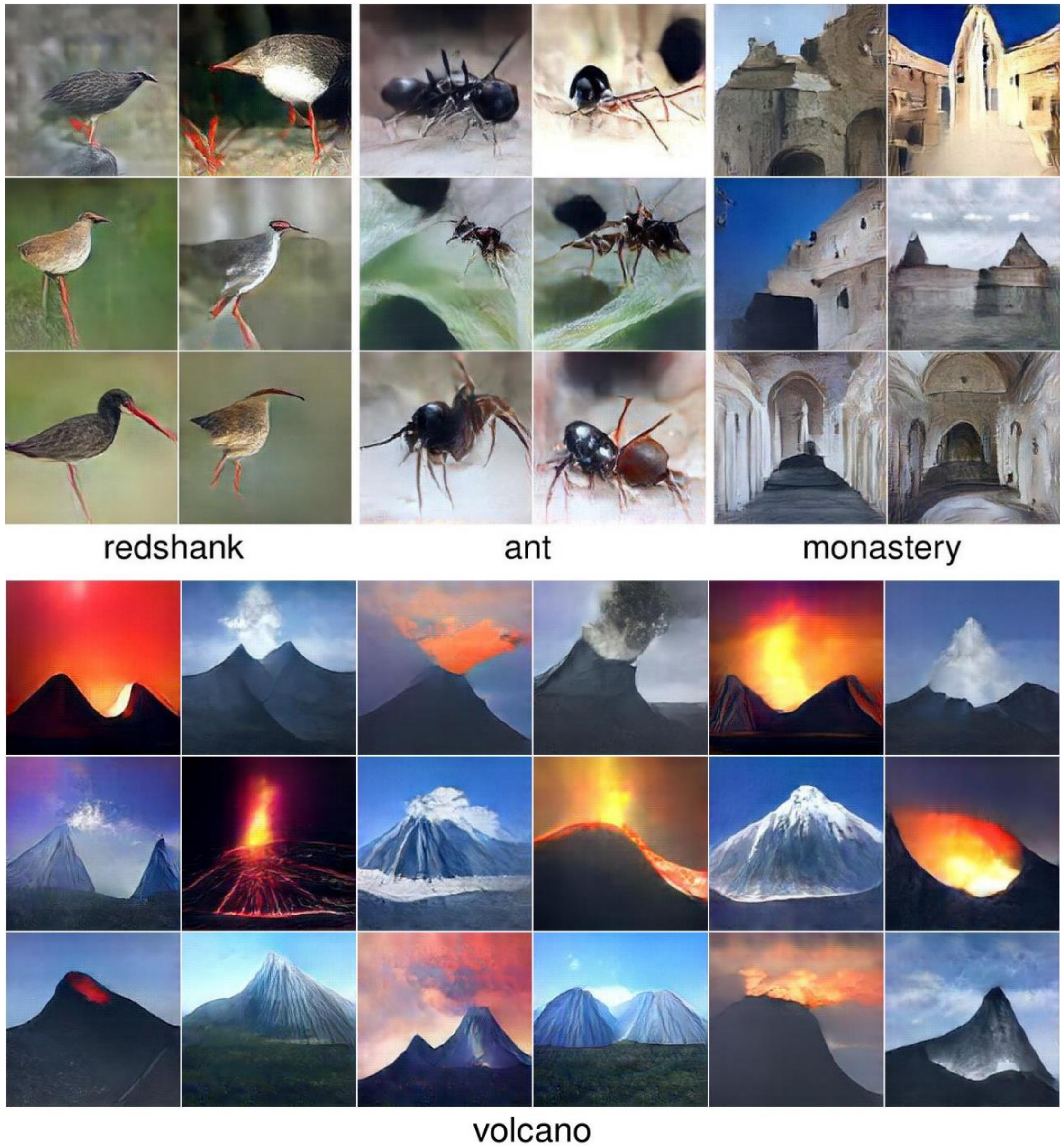
[Reed et al. ICML 2016]

# StackGAN



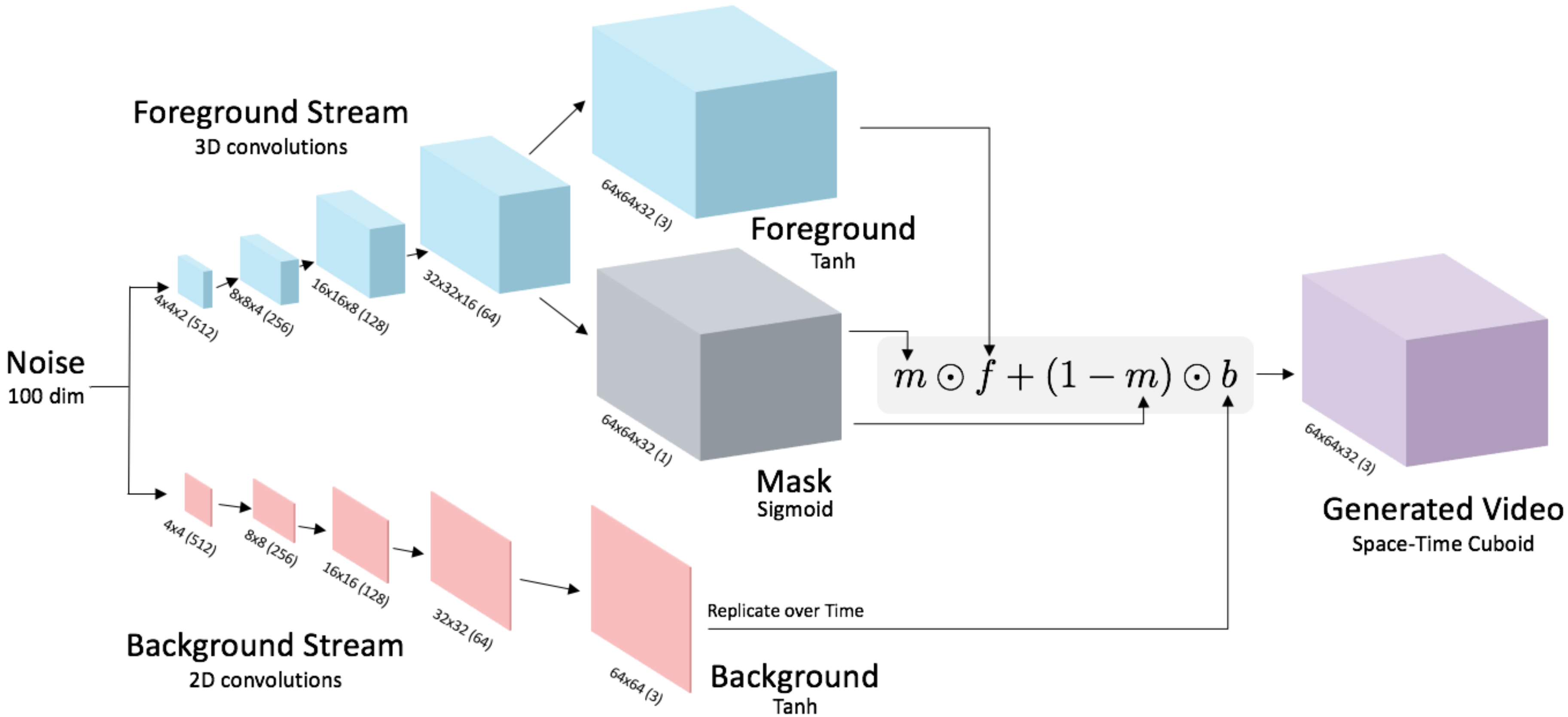

[Zhang et al. 2016]

# Plug & Play Generative Networks




[Nguyen et al. 2016]

# Video GAN



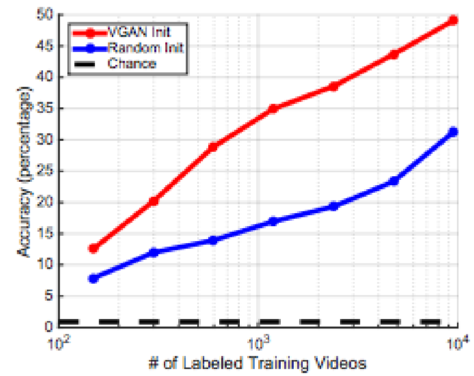

- Videos http://web.mit.edu/vondrick/tinyvideo/

[Vondrick et al. NIPS 2016]

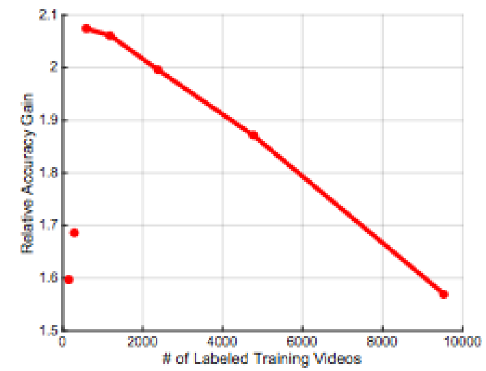
# Generative Modeling as Feature Learning

| Method | Accuracy |
|---|---|
| Chance | 0.9% |
| STIP Features [10] | 43.9% |
| Temporal Coherence [3] | 45.4% |
| Shuffle and Learn [42] | 50.2% |
| VGAN + Random Init | 36.7% |
| VGAN + Logistic Reg | 49.3% |
| **VGAN + Fine Tune** | **52.1%** |
| ImageNet Supervision [37] | 91.4% |

(a) Accuracy with Unsupervised Methods



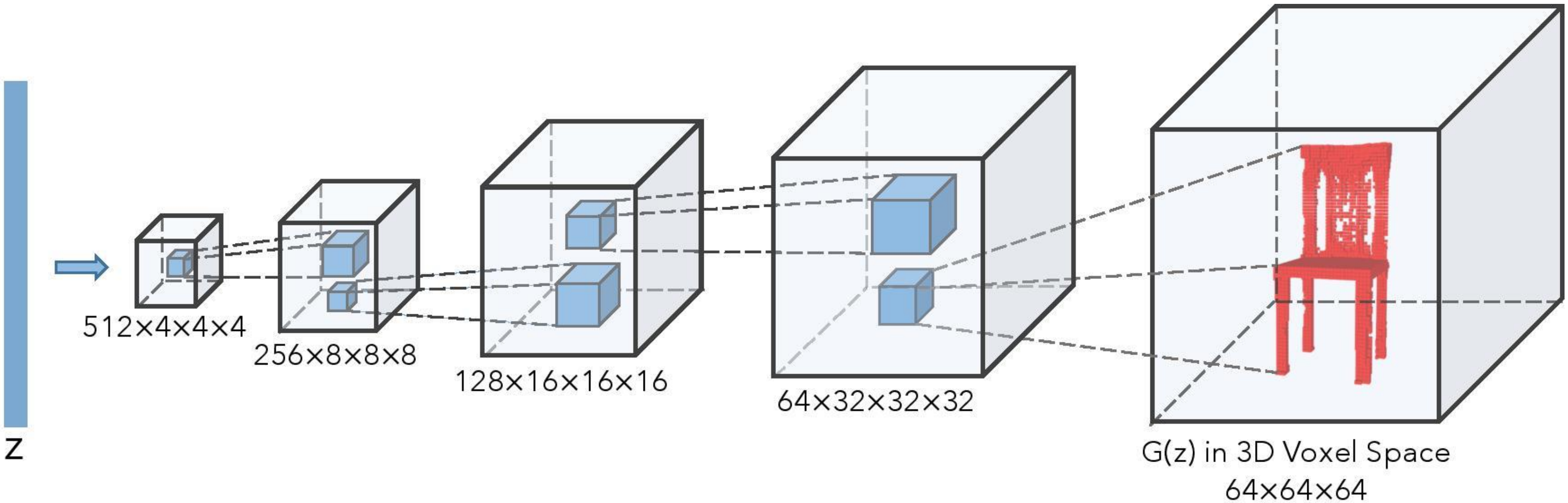

(b) Performance vs # Data

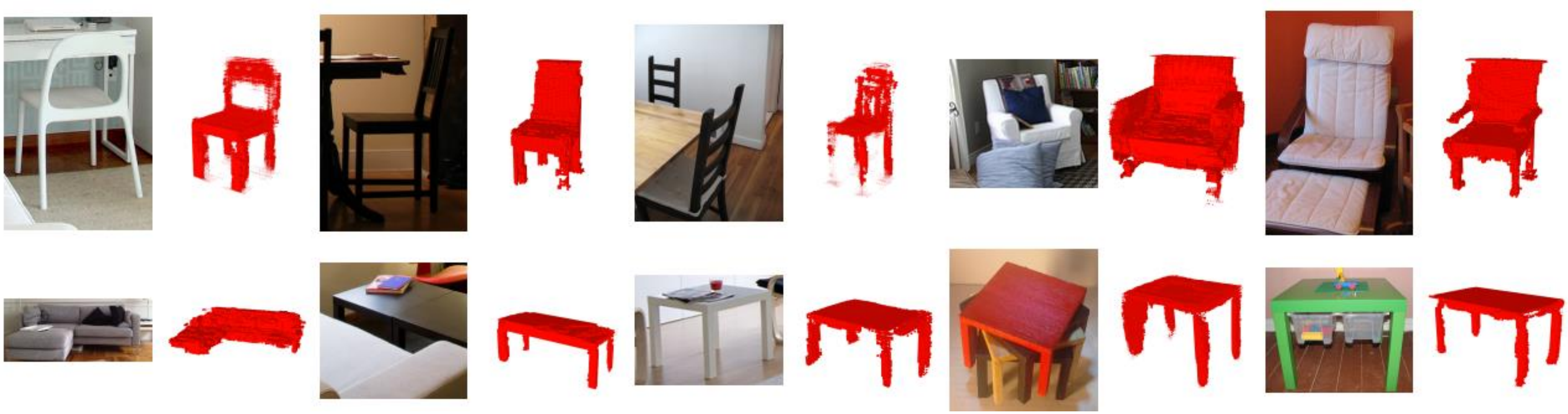



(c) Relative Gain vs # Data

[Vondrick et al. NIPS 2016]

# Shape modeling using 3D Generative Adversarial Network



[Wu et al. NIPS 2016]

# Things to remember

- GANs can generate sharp samples from high-dimensional output space
- Conditional GAN can serve as general mapping model X->Y
  - No need to define domain-specific loss functions
  - Handle one-to-many mappings
  - Handle multiple modalities