Michael James Smith's rhyperealistic paintings

Part 3 – Image Editing with GANs

Levent Karacan

Computer Vision Lab, Hacettepe University

HACETTEPE UNIVERSITY COMPUTER VISION LAB

HACETTEPE UNIVERSITY

Works will be presented

- Deep Convolutional Generative Adversarial Networks(DCGAN)
- Image Editing on Learned Manifold(iGAN)
- Conditional Generative Adversarial Networks(cGAN)
  - Image Generation from Text (Text2Im)
  - Stacked Generative Adversarial Networks(StackGAN)
  - Location and Description Conditioned Image Generation(GAWWN)
  - Image to Image Translation(pix2pix)
  - Image Generation from Semantic Segments and Attributes(AL-CGAN)**(Our work)**
  - Unpaired Image to Image Translation(CycleGAN)
- Neural Face Editing

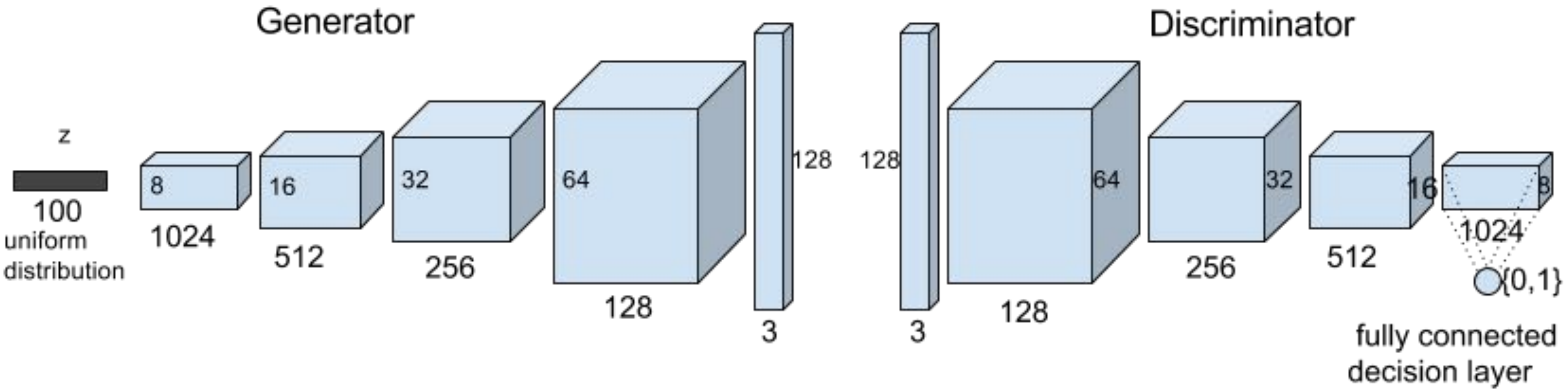
Generative Adversarial Networks(GAN)

Goodfellow vd. 2014(GAN); Radford vd. 2015(DCGAN)

- G tries to generate fake images that fool D.
- D tries to identify fake images.

$$\mathcal{L}_{GAN}(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{x\sim p_{data}(x),z\sim p_z(z)}[\log(1 - D(x,G(z)))]$$

$$G^* = \min_G \max_D \mathcal{L}_{GAN}(G,D)$$



DCGAN

- Cats

Source: https://github.com/aleju/cat-generator

DCGAN

- Animes

Source: https://github.com/jayleicn/animeGAN

DCGAN

- Album covers

Source: https://github.com/jayleicn/animeGAN

DCGAN

- Flowers

DCGAN

- Faces

Image Editing on Learned Manifold(iGAN) Zhu vd. 2016

- An image editing method that aims to find projection $\boldsymbol{z}$ of input image $\boldsymbol{x}$.

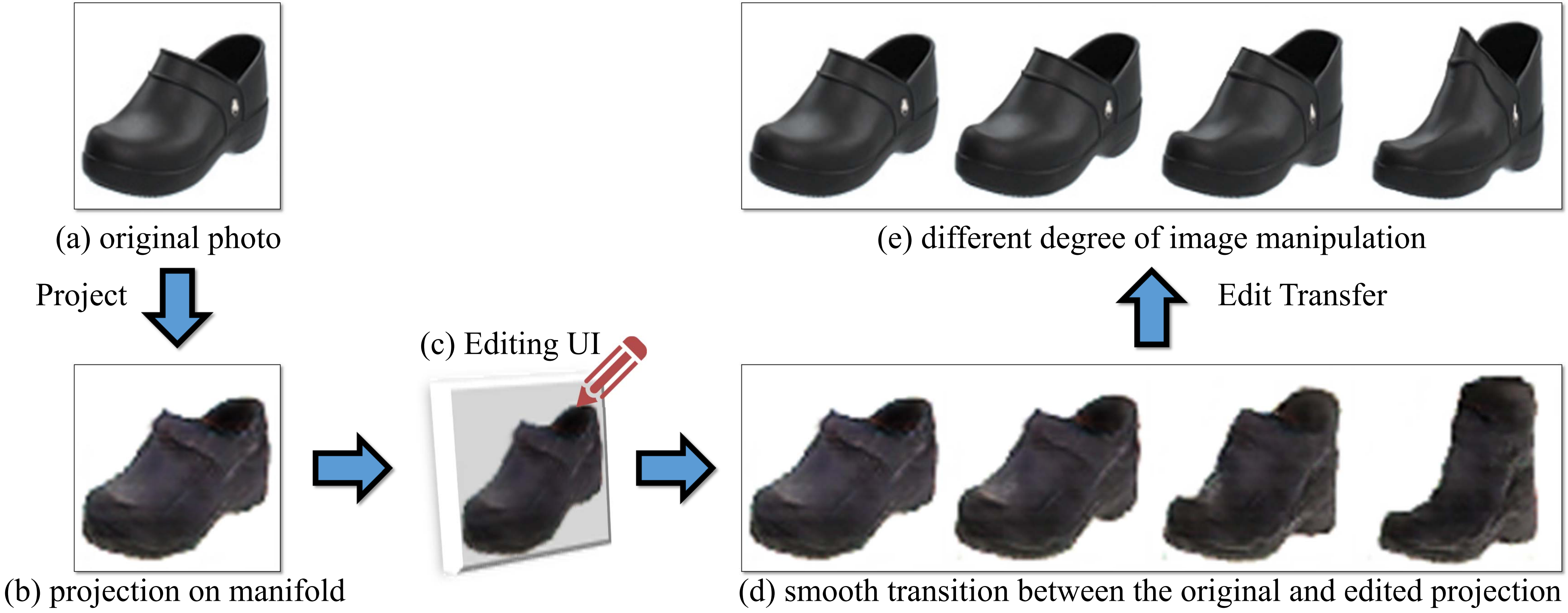


Image Editing on Learned Manifold(iGAN)

Zhu vd. 2016

- Find $\boldsymbol{z}$ that generates the input image $\mathbf{x}$ using generator network .

$$S\big(G(z_1), G(z_2)\big) \approx \|z_1 - z_2\|^2$$

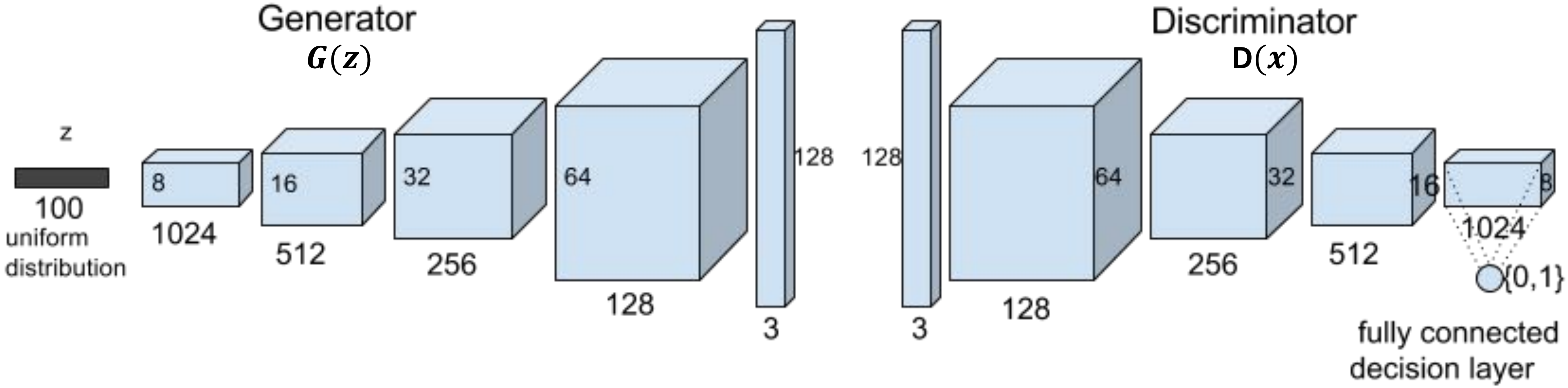


Image Editing on Learned Manifold(iGAN)

Zhu vd. 2016

- Images generated from DCGAN trained on shirt image dataset.

$$S\big(G(z_1), G(z_2)\big) \approx \|z_1 - z_2\|^2$$

random jittering

linear interpolation

Image Editing on Learned Manifold(iGAN)

Zhu vd. 2016

$$S\big(G(z_1), G(z_2)\big) \approx \|z_1 - z_2\|^2$$

$$\mathcal{L}(x_1, x_2) = \|C(x_1) - C(x_2)\|^2$$

- Projection via optimization. L-BFGS-B method.

$$z^* = arg\min_{z \in \tilde{\mathbb{Z}}} \mathcal{L}(G(z), x^R)$$

- Projection via feedforward network.

$$\theta_P^* = arg\min_{\theta_P} \sum_n \mathcal{L}(G(P(x_n^R; \theta_P), x_n^R))$$

- Hybrid method.

Image Editing on Learned Manifold(iGAN)

Zhu vd. 2016

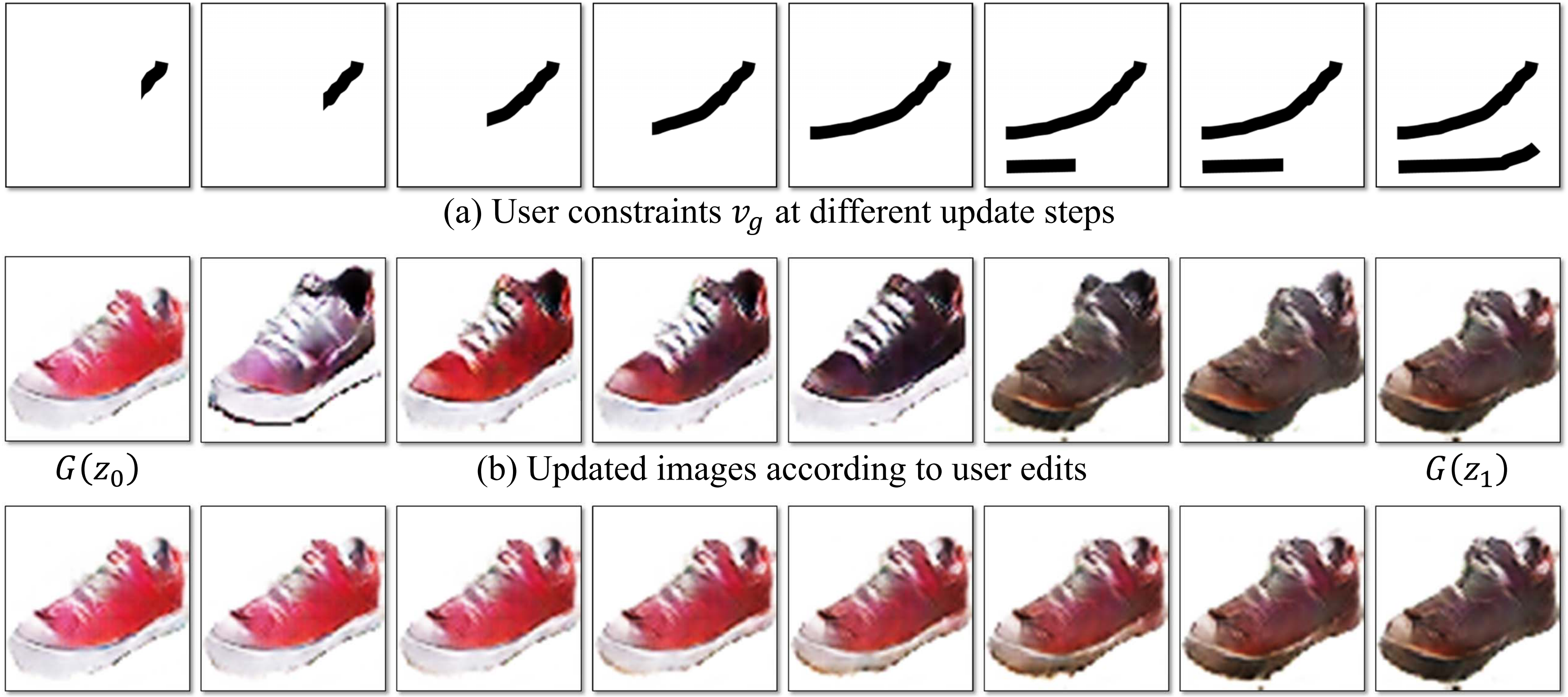
Image Editing on Learned Manifold(iGAN)

Zhu vd. 2016

$$S\big(G(z_1), G(z_2)\big) \approx \|z_1 - z_2\|^2$$

$$z^* = \min_{z \in \mathbb{Z}} \{ \sum_g \left\| f_g(G(z)) - v_g \right\|^2 + \lambda_s . \|z - z_0\|^2$$

g: Color, shape and warping constraints for image editing.



(a) User constraints v_g at different update steps

$G(z_0)$ (b) Updated images according to user edits $G(z_1)$

(c) Linear interpolation between $G(z_0)$ and $G(z_1)$

Image Editing on Learned Manifold(iGAN)

Zhu vd. 2016

Edit Transfer

- A dense correspondence algorithm to estimate both the geometric and color changes induced by the editing process.

Image Editing on Learned Manifold(iGAN)

Zhu vd. 2016

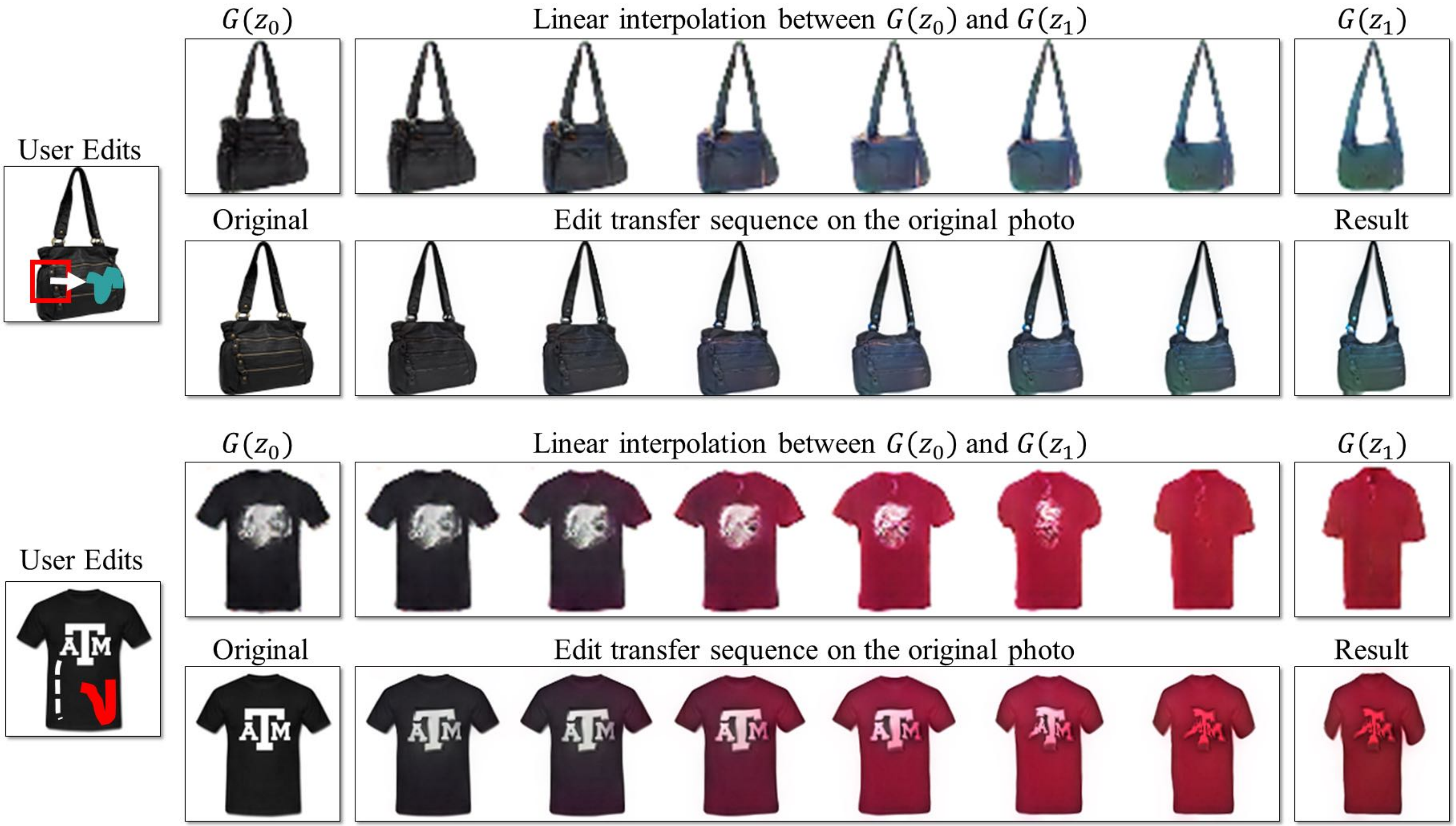
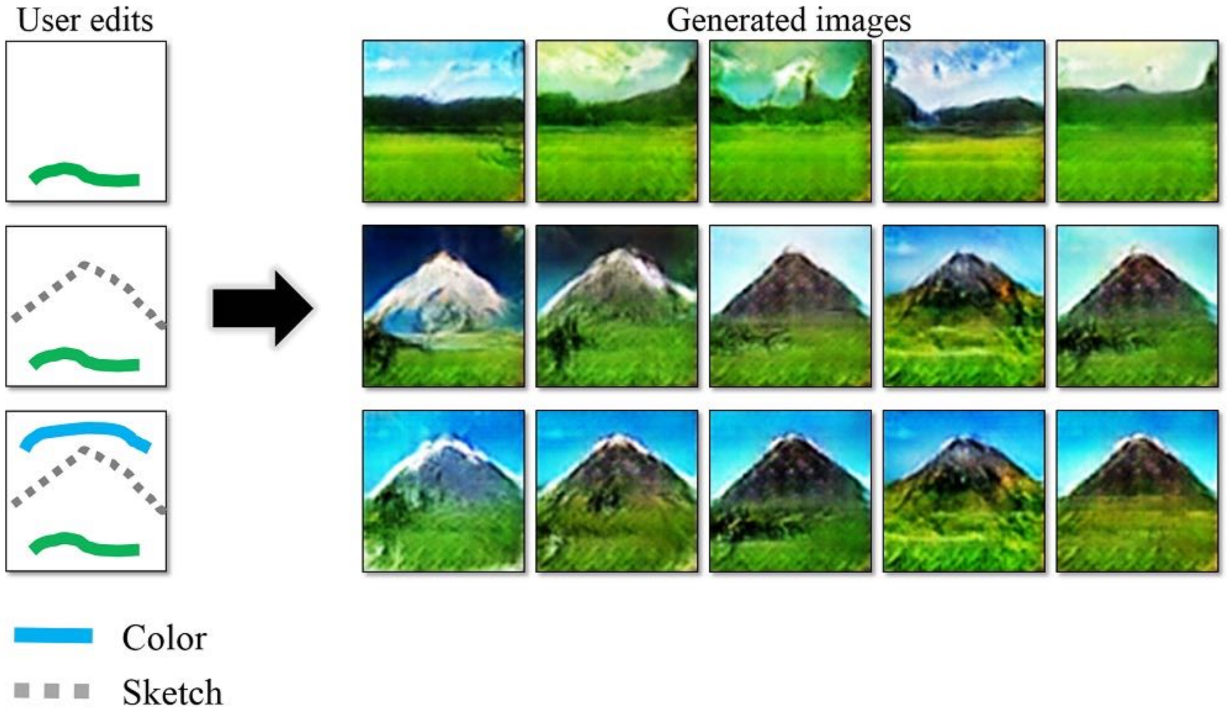


Image Editing on Learned Manifold(iGAN)

Zhu vd. 2016



Conditional Generative Adversarial Networks(cGAN) Mirza vd. 2014

- Concatenate condition information $\boldsymbol{x}$ to noise vector $\boldsymbol{z}$ and introduce to discriminator.

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + \mathbb{E}_{x\sim p_{data}(x),z\sim p_z(z)}[\log(1-D(x,G(x,z)))]$$

$$G^* = \min_G \max_D \mathcal{L}_{cGAN}(G,D)$$

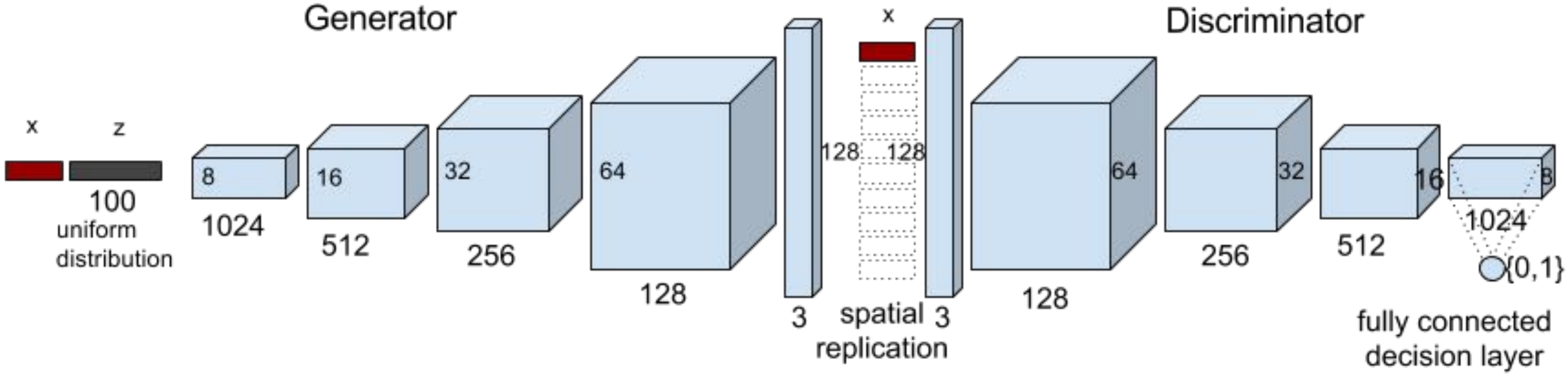


Image Generation from Text(Text2Im) Reed vd. 2016

- Discriminator network tries to classify real image and wrong text as well as real/fake image with right text.
- Condition: Text description embedding.
- CUB bird dataset(11788 images from 200 categories), Oxford-102 flower dataset(8189 images from 102 categories).

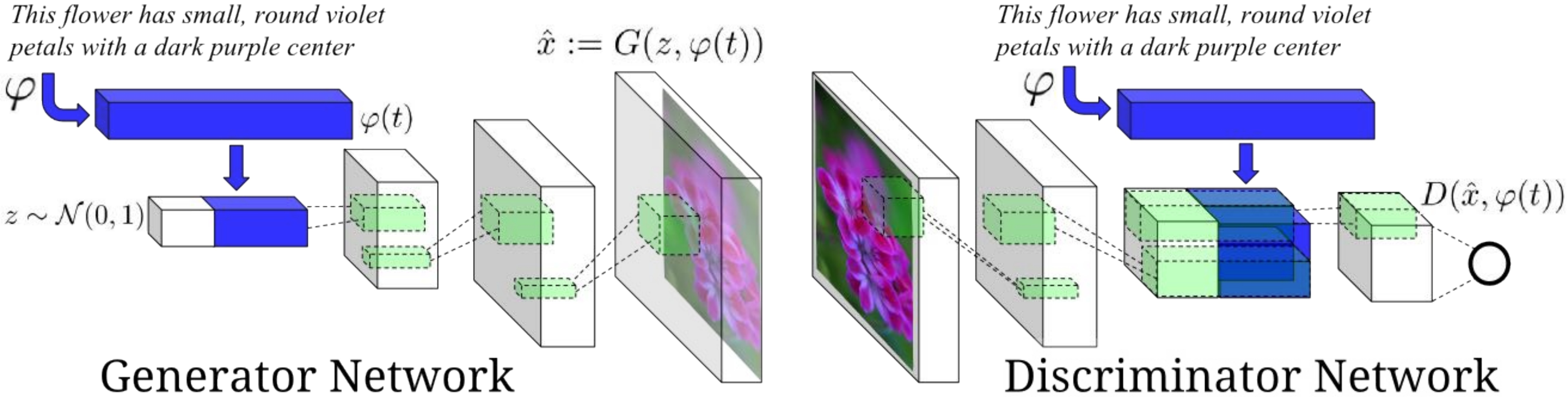


Image Generation from Text(Text2Im) Reed vd. 2016

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

Text descriptions (content)

Images (style)

The bird has a **yellow breast** with **grey** features and a small beak.

This is a large **white** bird with **black wings** and a **red head**.

A small bird with a **black head and wings** and features grey wings.

This bird has a **white breast**, brown and white coloring on its head and wings, and a thin pointy beak.

A small bird with **white base** and **black stripes** throughout its belly, head, and feathers.

A small sized bird that has a cream belly and a short pointed bill.

This bird is **completely red**.

This bird is **completely white**.

This is a **yellow** bird. The **wings are bright blue**.

Image Generation from Text(Text2Im) Reed vd. 2016

"Blue bird with black beak → "This bird is completely red with black wings"

Image Generation from Text(Text2Im) Reed vd. 2016

“Small blue bird with black wings.” → **“Small yellow bird with black wings”**

Image Generation from Text(Text2Im) Reed vd. 2016

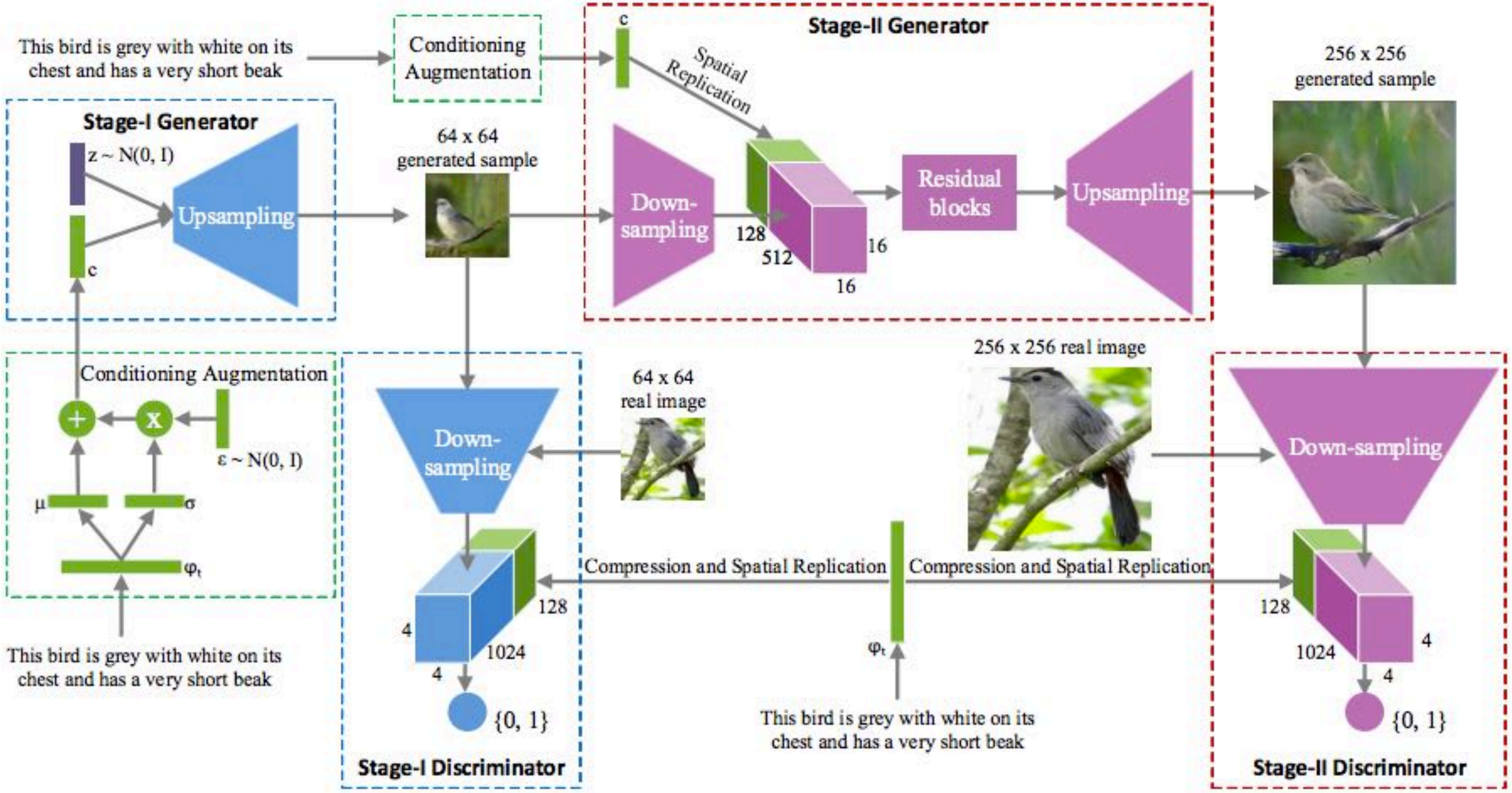
“This bird is bright.” → “This bird is dark”

Stacked Generative Adversarial Networks(StackGAN)

Han vd. 2016

- There are 2 stages.
- Stage-I GAN: Generates low resolution images.
  - Conditioning Augmentation
  - Regularization term is added to generator.

  $D_{KL}(\mathcal{N}(\mu(\varphi_t) \parallel \mathcal{N}(0, I))$
- Stage-II GAN: Generates high resolution detailed images.
  - Noise vector is not used.



Stacked Generative Adversarial Networks(StackGAN) Han vd. 2016

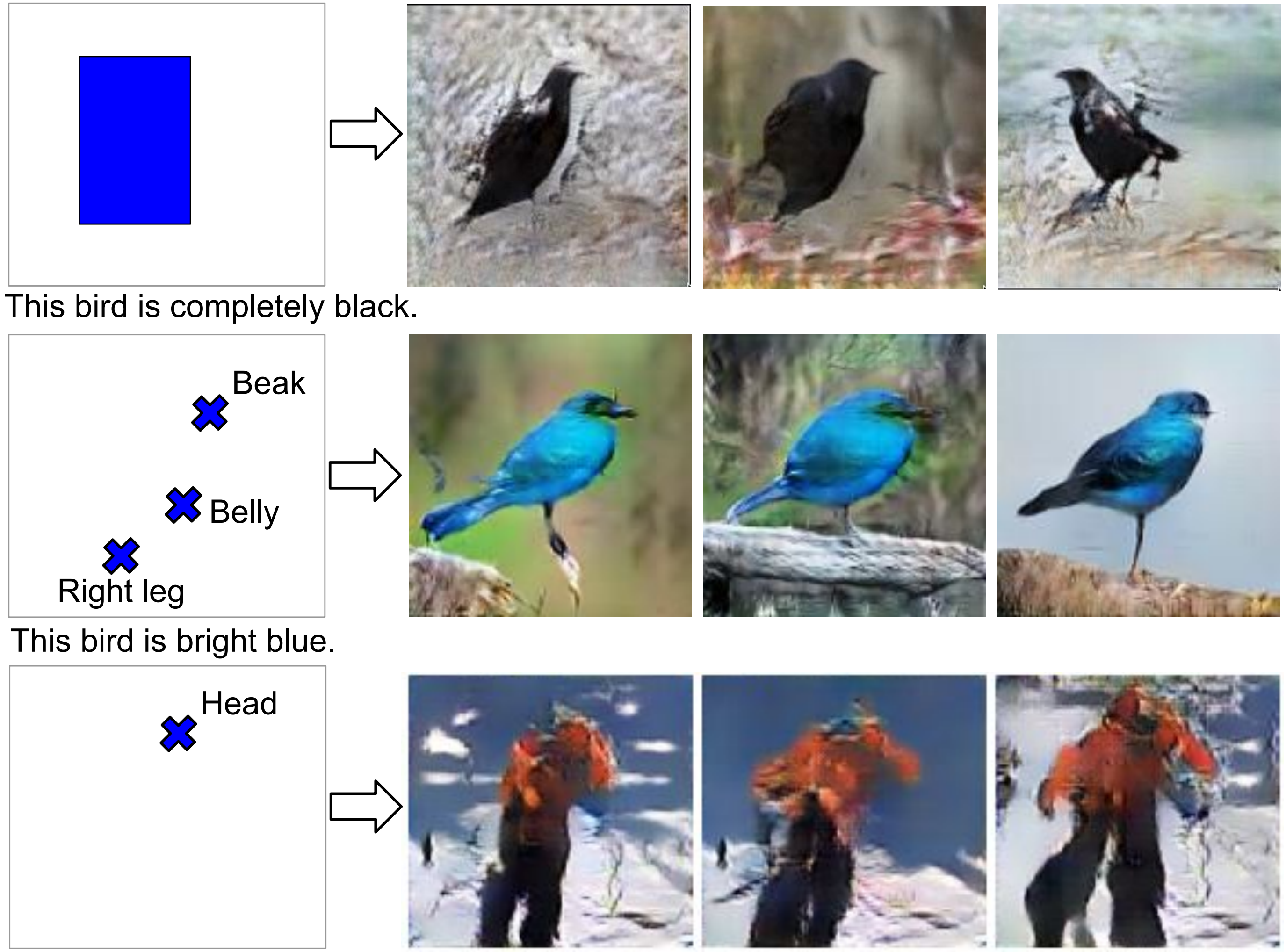
Stacked Generative Adversarial Networks(StackGAN) Han vd. 2016

Stacked Generative Adversarial Networks(StackGAN) Han vd. 2016

Stacked Generative Adversarial Networks(StackGAN) Han vd. 2016

Location and Description Conditioned Image Generation(GAWWN)

Reed vd. 2016



This bird is completely black.

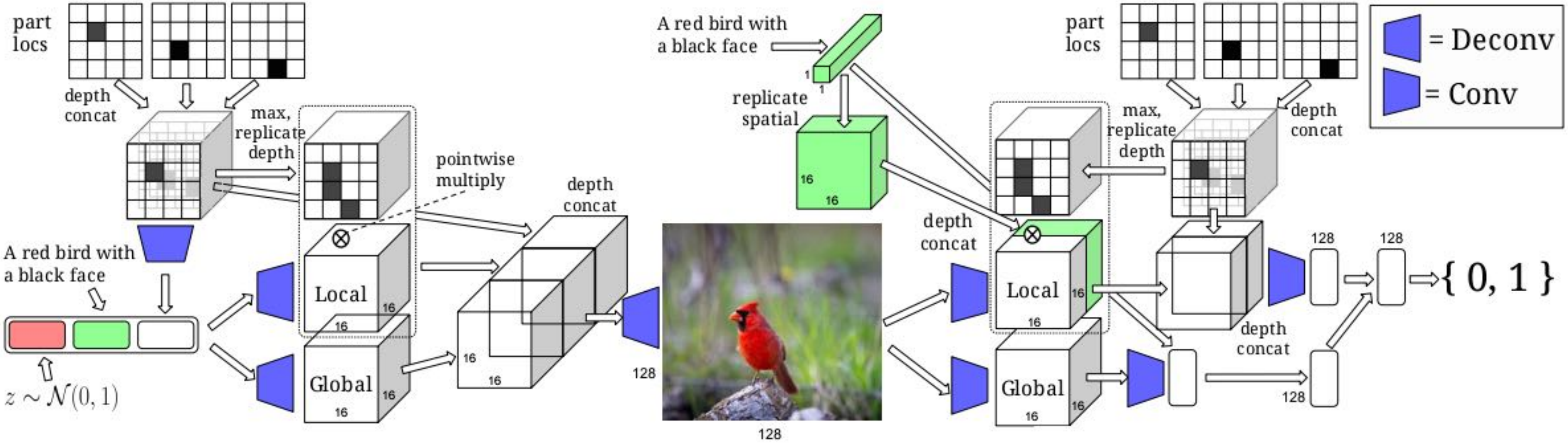
This bird is bright blue.

a man in an orange jacket, black pants and a black cap wearing sunglasses skiing

Location and Description Conditioned Image Generation(GAWWN)

Reed vd. 2016

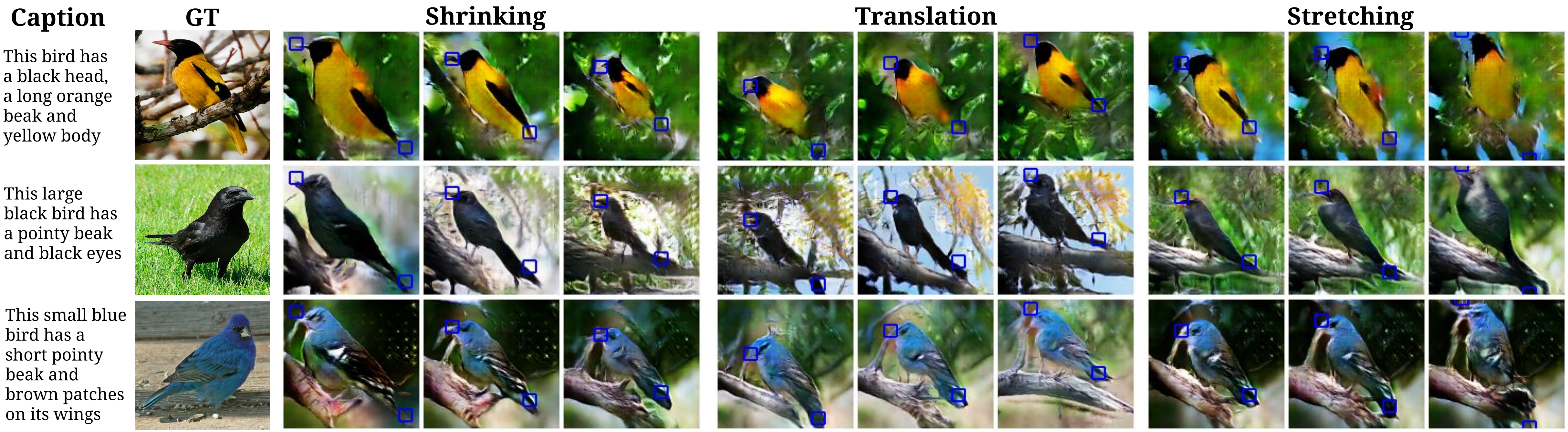
- Keypoint conditioned architecture.



Location and Description Conditioned Image Generation(GAWWN)

Reed vd. 2016

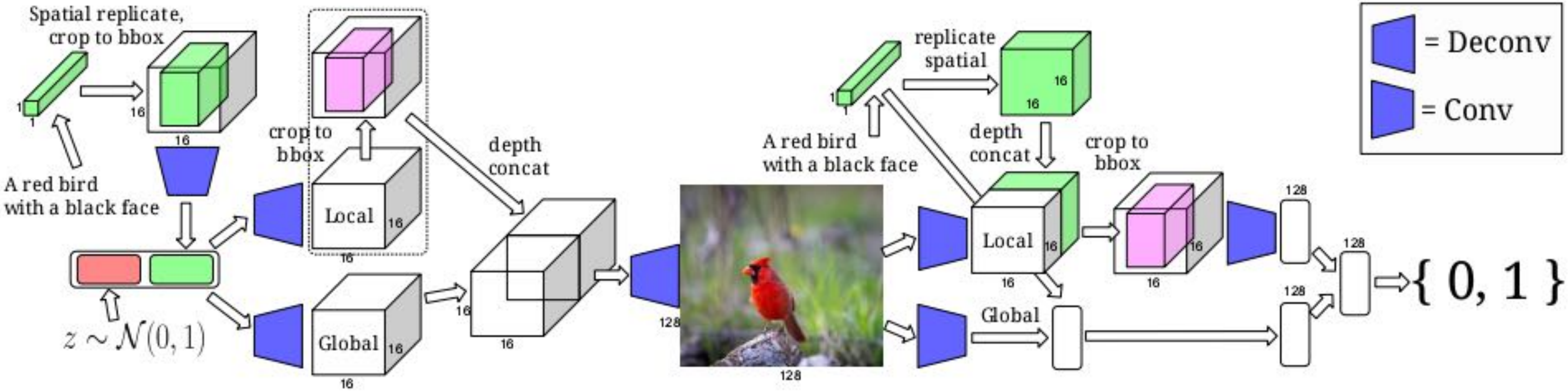
- Keypoint conditioned architecture.



Location and Description Conditioned Image Generation(GAWWN)

Reed vd. 2016

- Bounding box conditioned architecture.



Location and Description Conditioned Image Generation(GAWWN)

Reed vd. 2016

- Bounding box conditioned architecture.

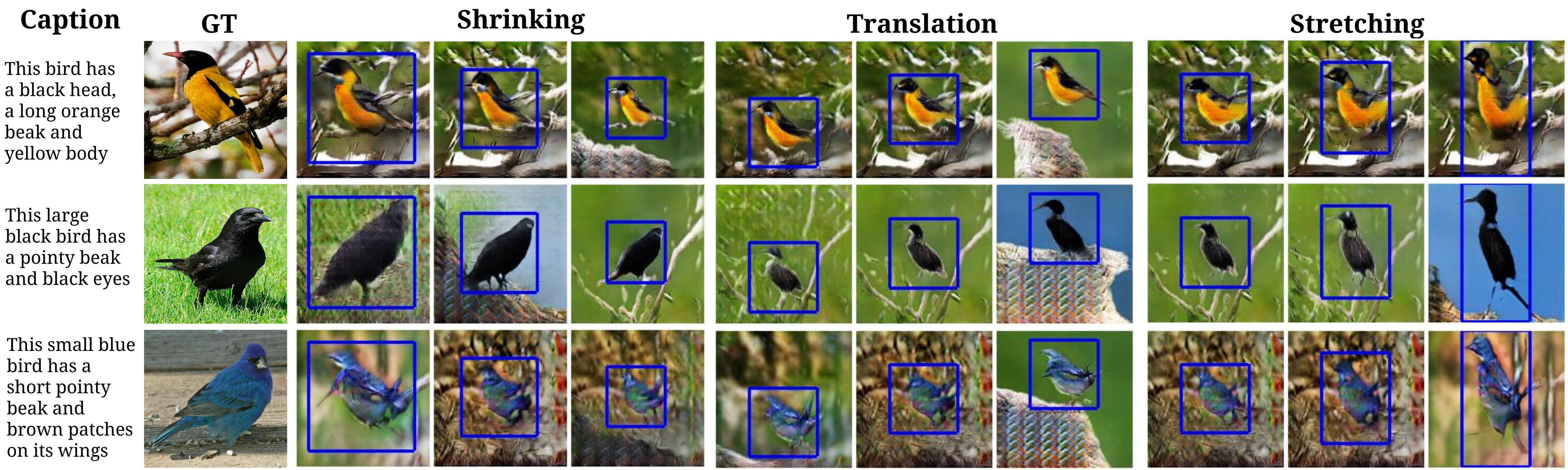


Image to Image Translation(pix2pix) Isola vd. 2017

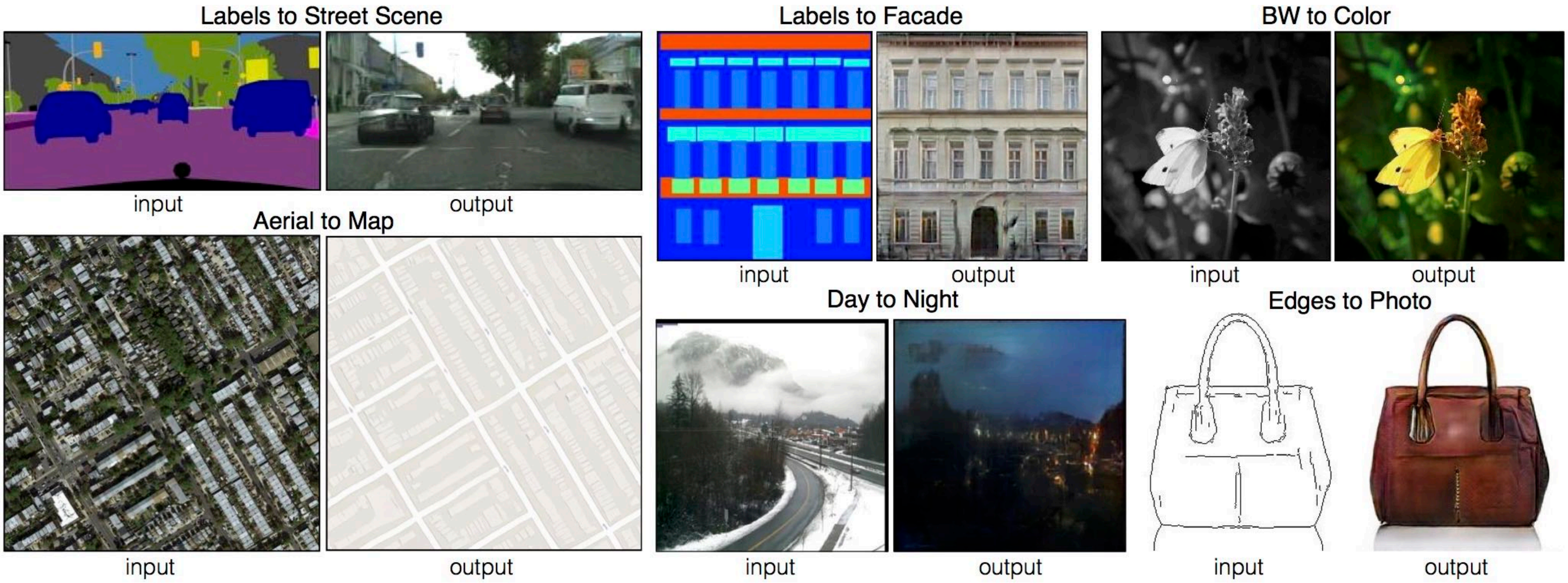
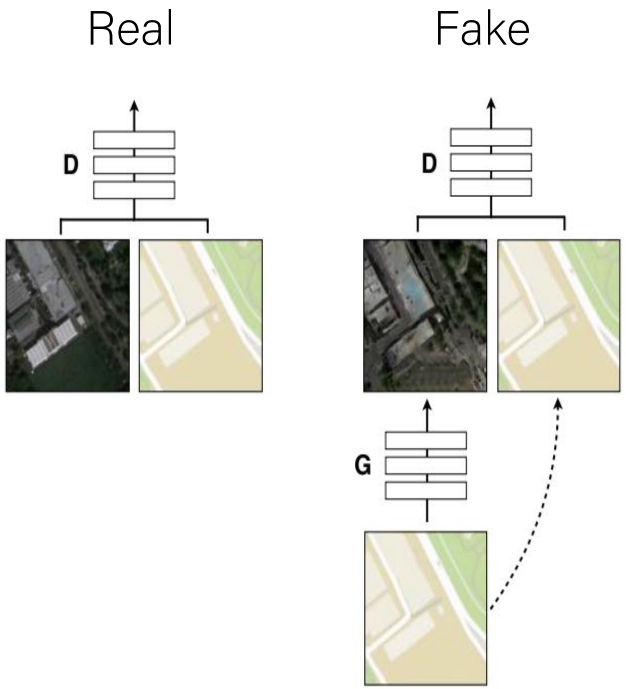
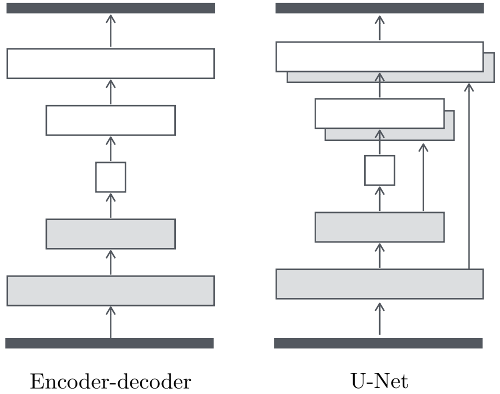


Image to Image Translation(pix2pix) Isola vd. 2017

$$G^* = arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_{L1}(G)$$





- G tries to generate fake images that fool D.
- D tries to identify fake images.

Adversarial Loss

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)] +$$

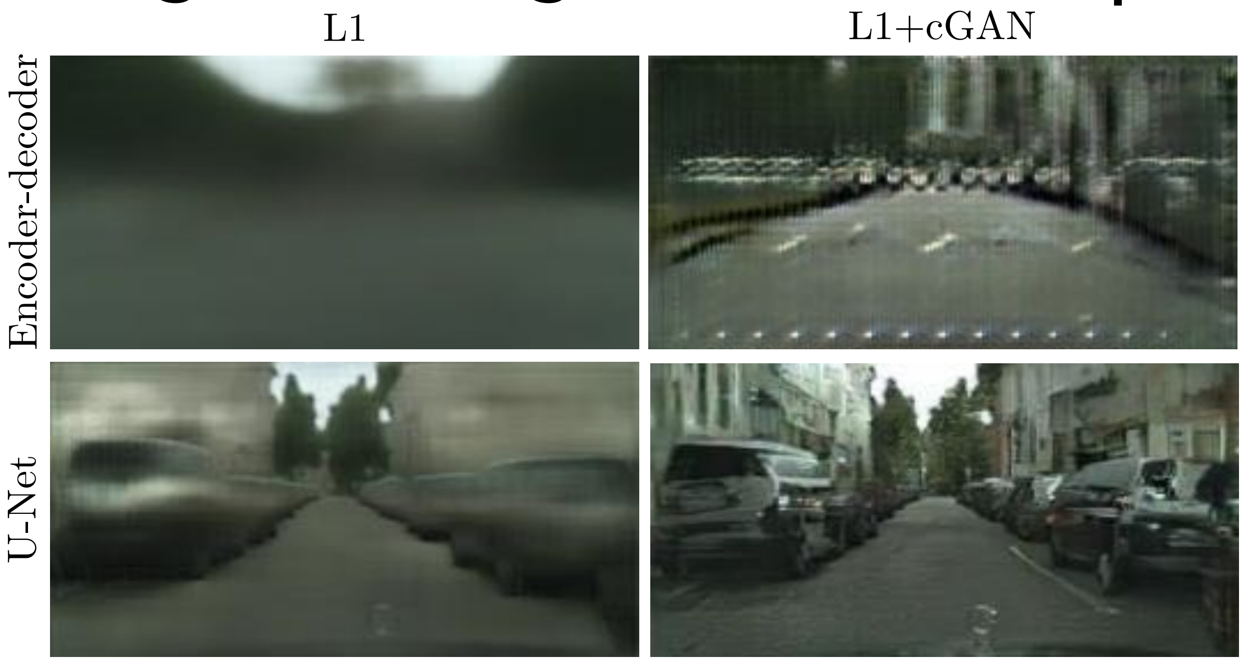
$$\mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D(x, G(x)))]$$

L1 Loss

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y\sim p_{data}(x,y), z\sim p_z(z)}[\|y - G(x,z)\|_1]$$

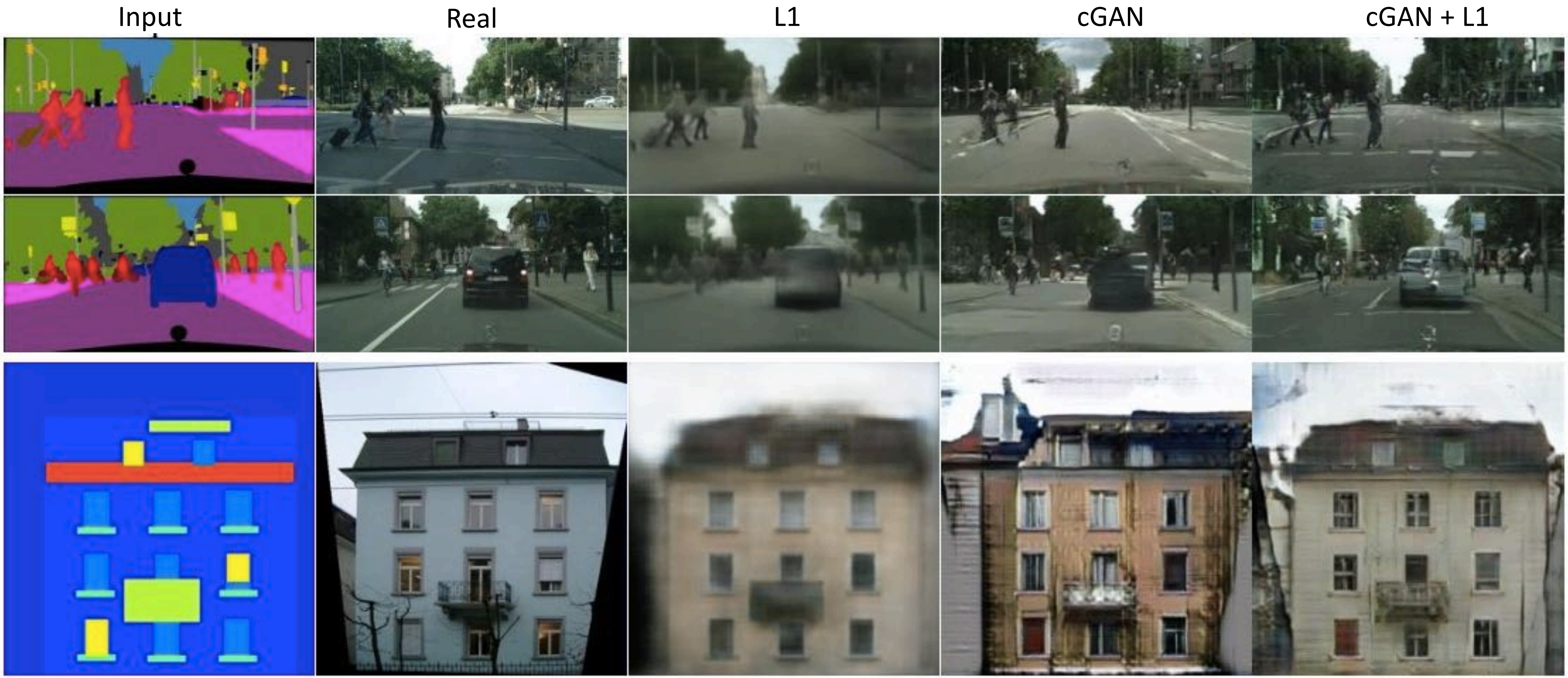
- Noise vector is removed,.Instead dropout is used to provide stochasticity.
- Skip connections on Generative model
- PatchGAN is proposed for dicriminator instead of pixel GAN.

Image to Image Translation(pix2pix) Isola vd. 2017



- U-Net provides to include low-level features to be used yo generate more realistic images.
- PatchGAN provides to generate sharper images. .

Image to Image Translation(pix2pix) Isola vd. 2017



Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A. “Image-to-image translation with conditional adversarial networks.”. In *CVPR 2017*.

.

Image to Image Translation(pix2pix) Isola vd. 2017

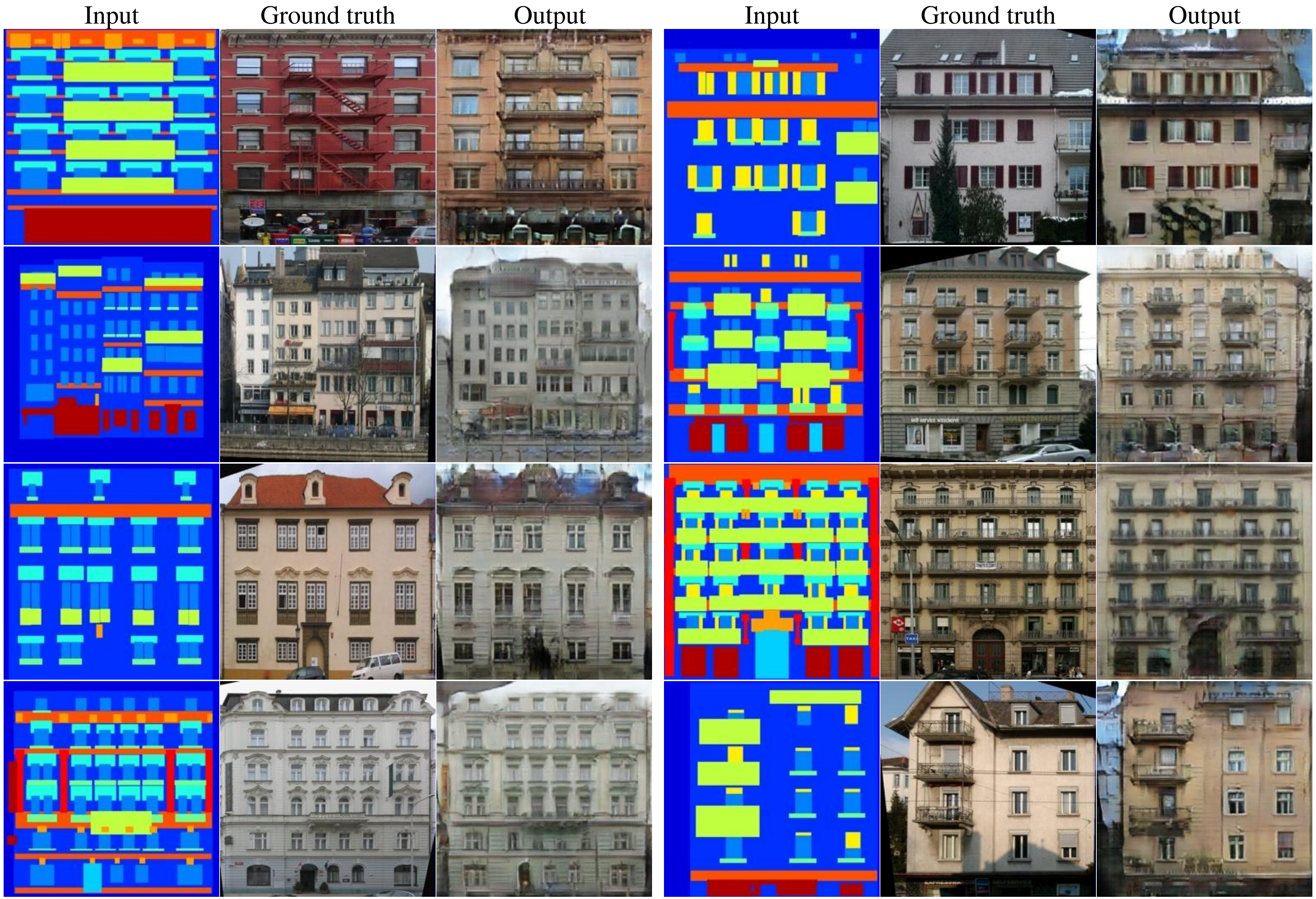


Image to Image Translation(pix2pix) Isola vd. 2017

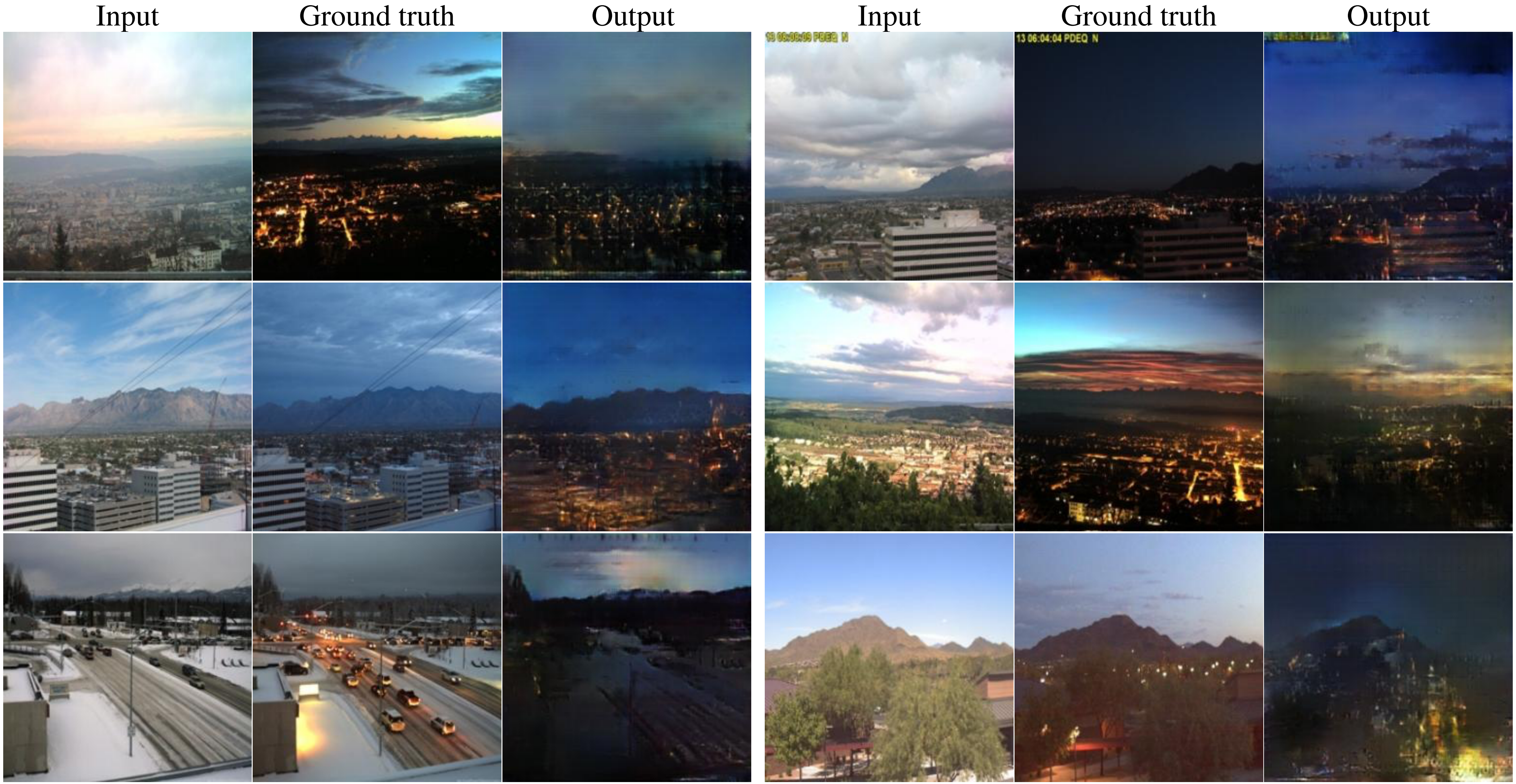
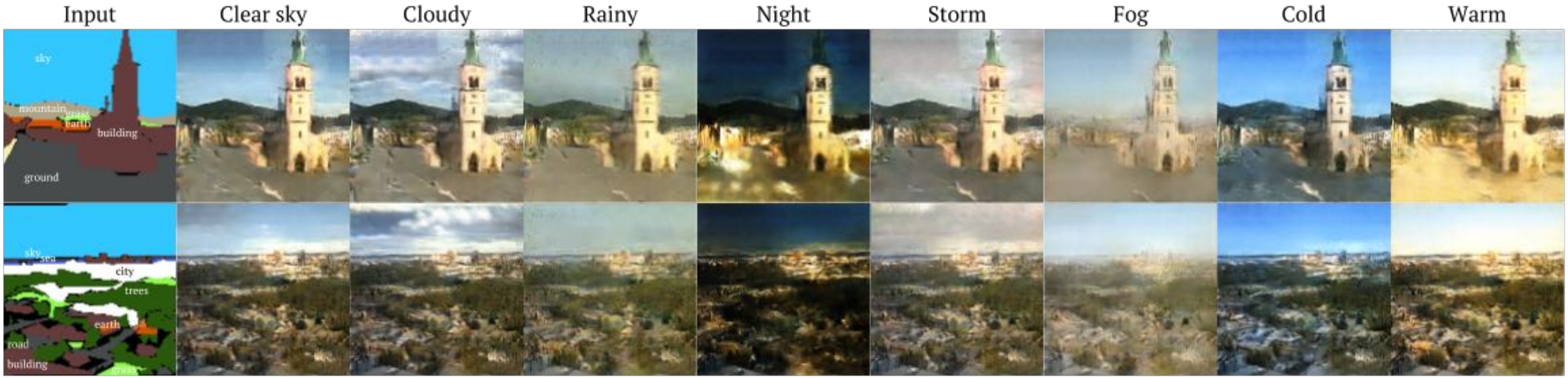


Image to Image Translation(pix2pix) Isola vd. 2017

Image to Image Translation(pix2pix) Isola vd. 2017

Image to Image Translation(pix2pix) Isola vd. 2017

Attribute and Layout Conditioned Image Generation(AL-CGAN) Our work

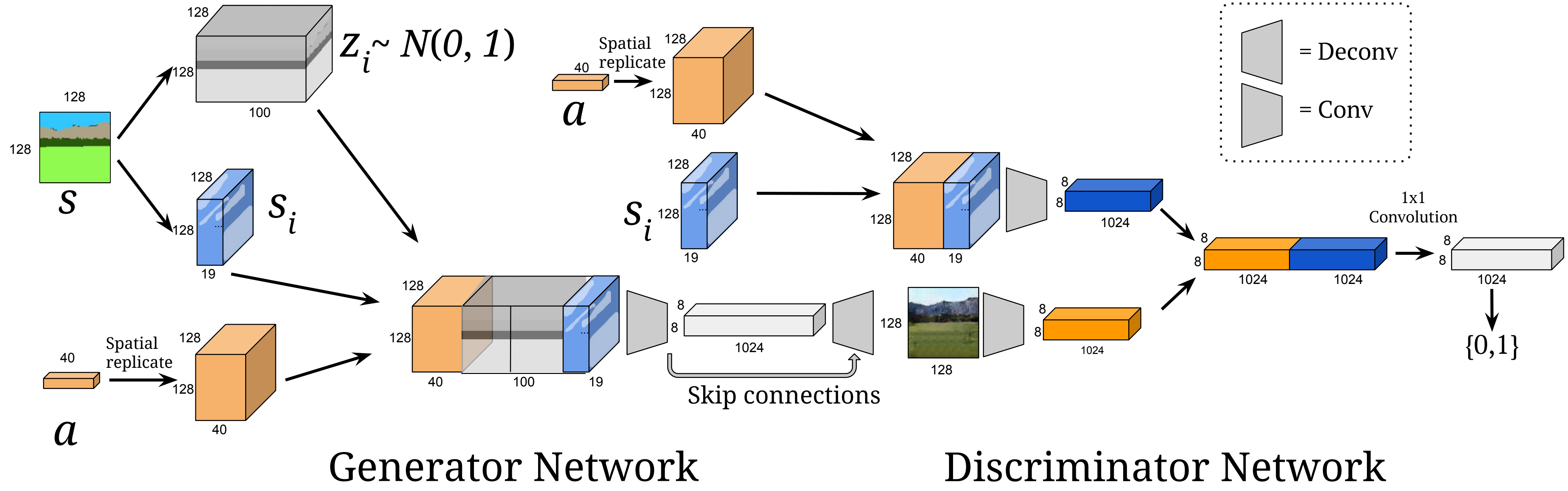


Attribute and Layout Conditioned Image Generation(AL-CGAN) Our work

$$\mathcal{L}_{cGAN}(G, D) \\ = \mathbb{E}_{x,s,a\sim p_{data}(x,s,a)}[\log D(x, s, a)] + \mathbb{E}_{s,a\sim p_{data}(s,a),z\sim p_z(z)}[\log(1 - D(x, G(z, s, a)))]$$

$$\min_G \max_D \mathcal{L}_{cGAN}(G, D)$$

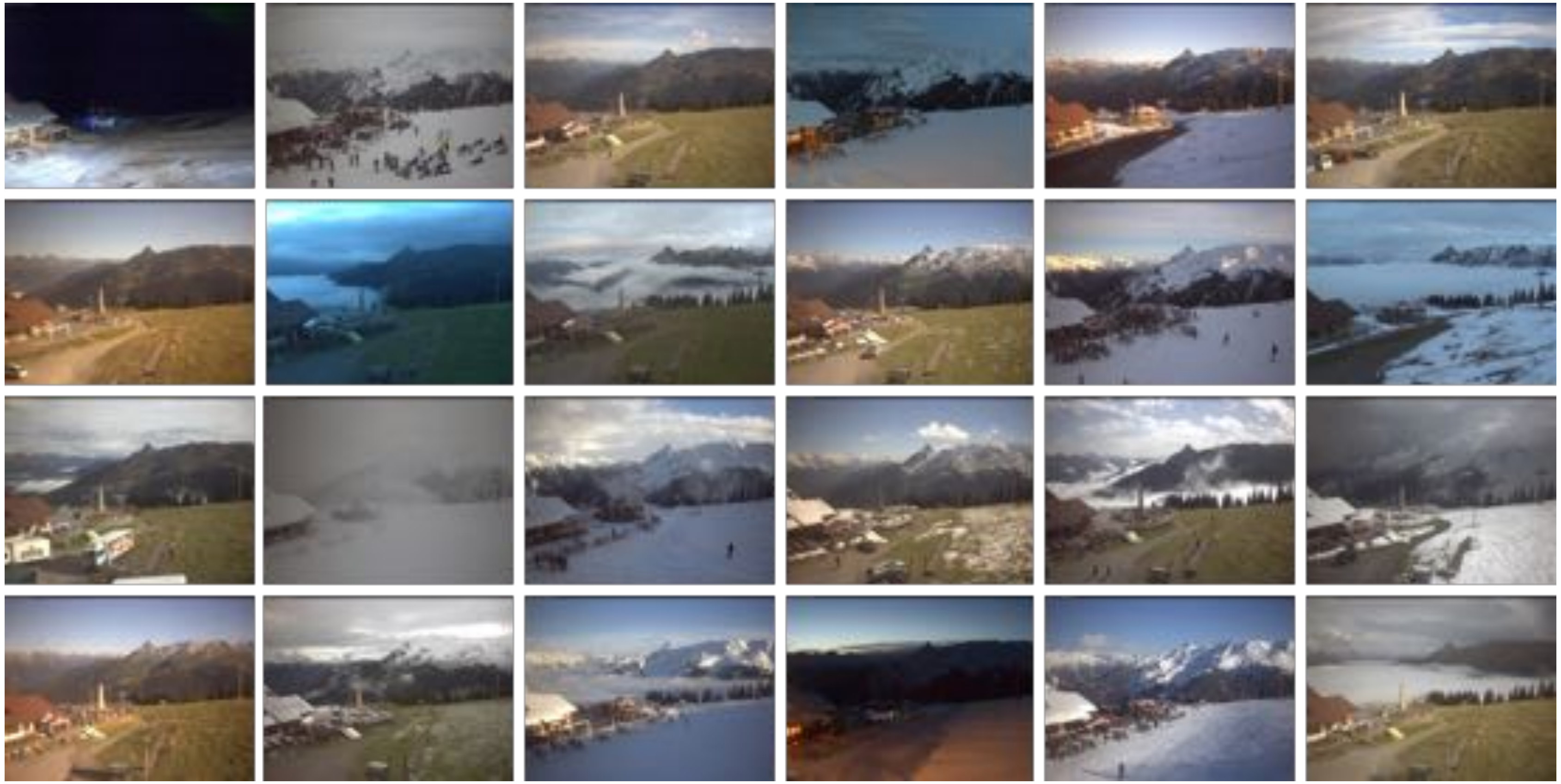
- The noise vectors z are specific to the semantic layout.
- This provides the diversity in generated samples.



Dataset Laffont vd. 2014

- Transient Attribute Dataset
  - 8571 outdoor images from 101 web cams located in different places.
  - 40 dimensional transient attributes for each image.

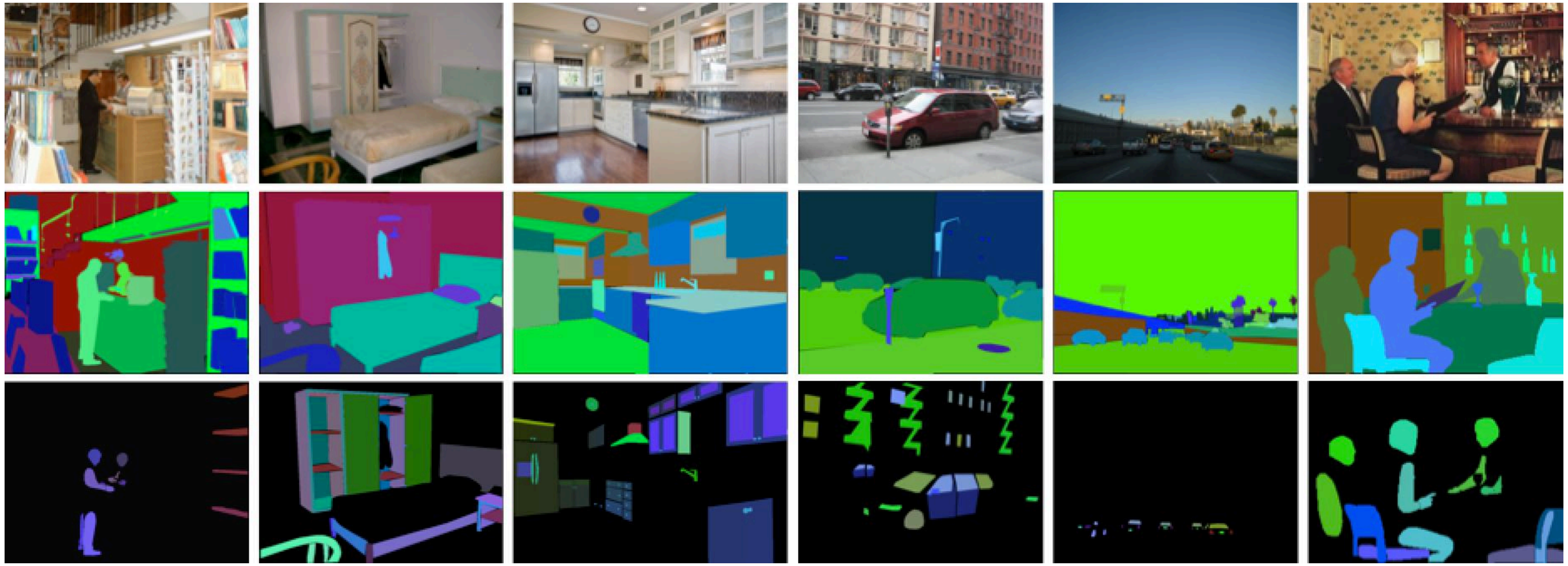
  - We annotate semantic layouts of 101 scenes with predefined 18 categories e.g. sky, tree, building, mountain, etc.



P.-Y. Laffont, Z. Ren, X. Tao, C. Qian, and J. Hays, “Transient attributes for high-level understanding and editing of outdoor scenes,” *ACM Transactions on Graphics*, vol. 33, no. 4, 2014 .

Dataset Zhou vd. 2017

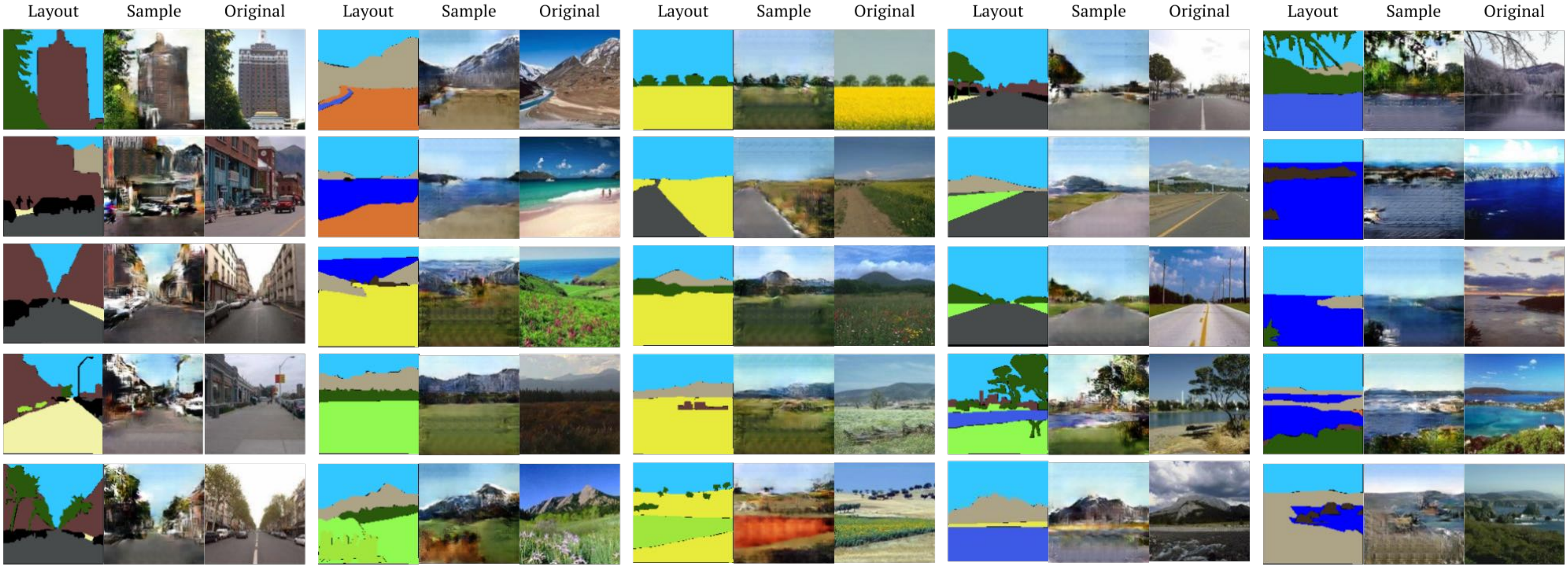
- ADE20K
  - 22210 indoor and outdoor scenes with semantically labeled layouts.



- We selected 9201 outdoor scenes according to predefined 18 categories.
- We predicted transient attributes for each image using a deep transient model.

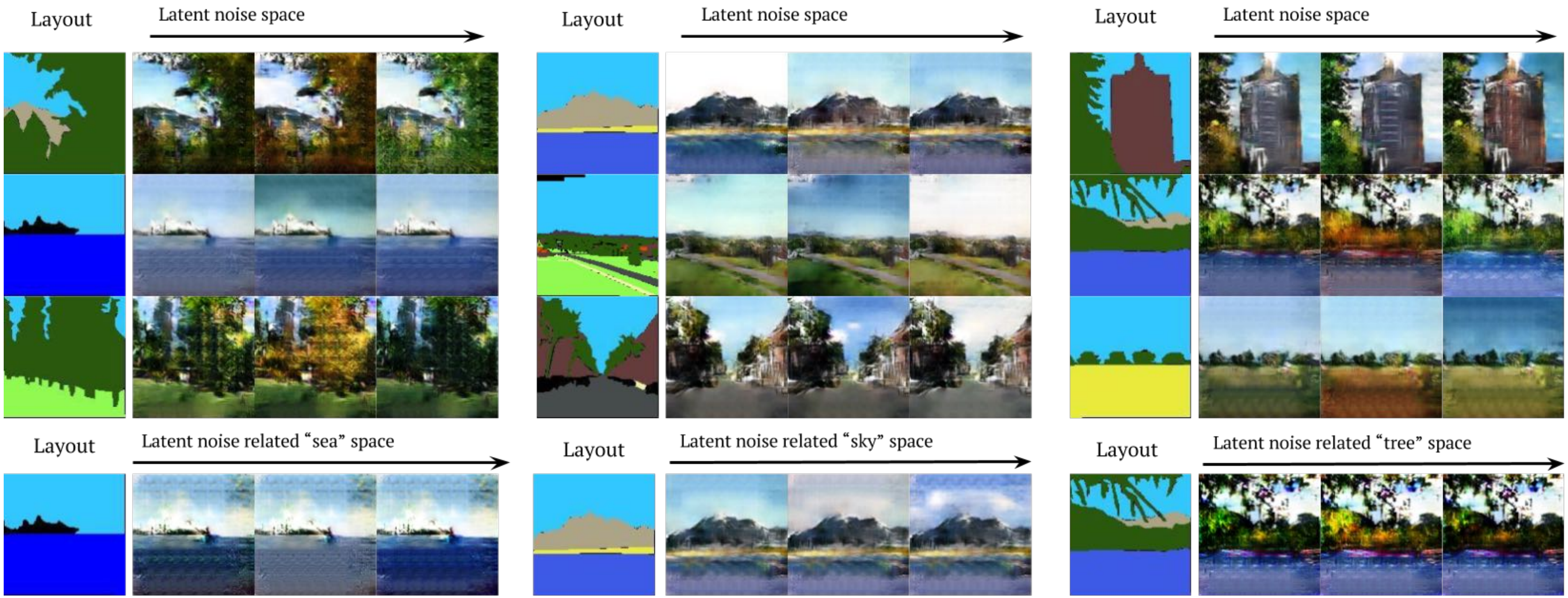
R. Baltenberger, M. Zhai, C. Greenwell, S. Workman, and N. Jacobs. A Fast Method for Estimating Transient Scene Attributes. In *WACV 2016*.

Attribute and Layout Conditioned Image Generation(AL-CGAN) Our work



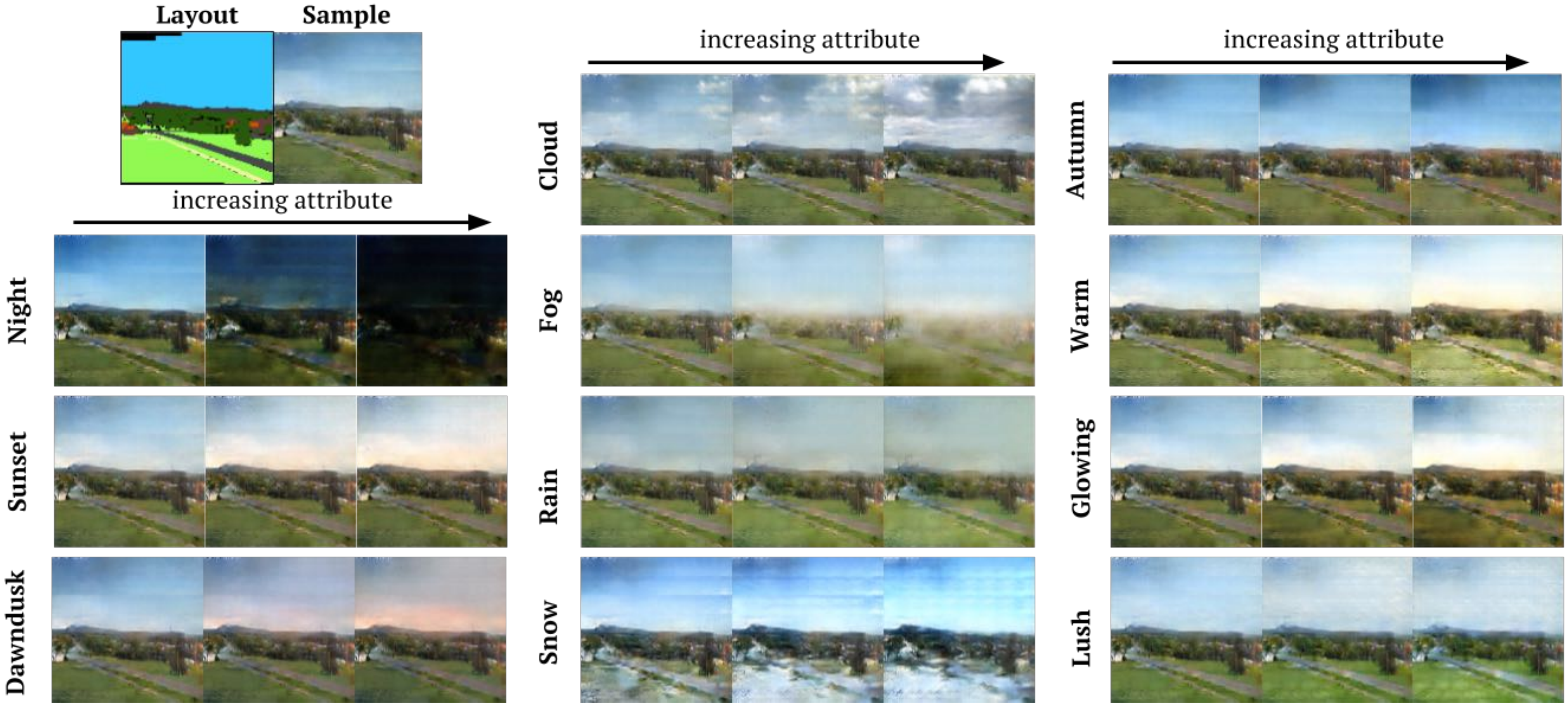
Attribute and Layout Conditioned Image Generation(AL-CGAN) Our work

- Diversity

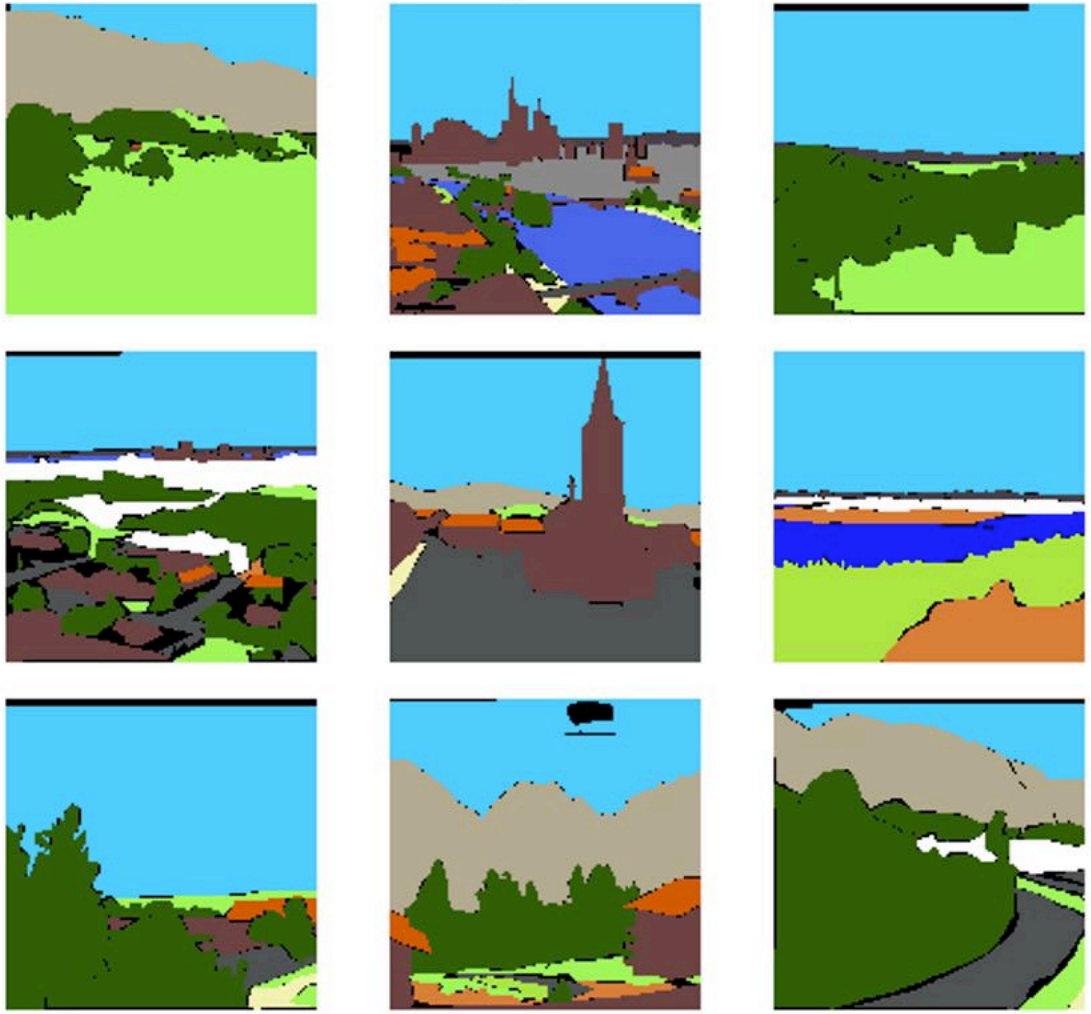


Attribute and Layout Conditioned Image Generation(AL-CGAN) Our work

- Diversity by transient attributes.

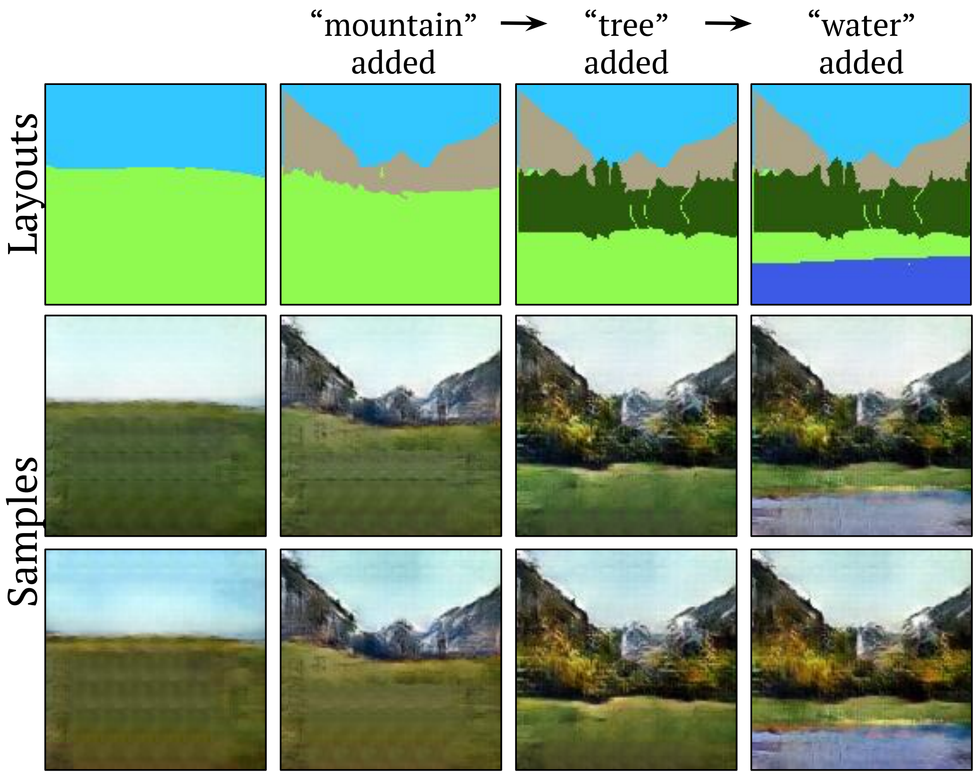


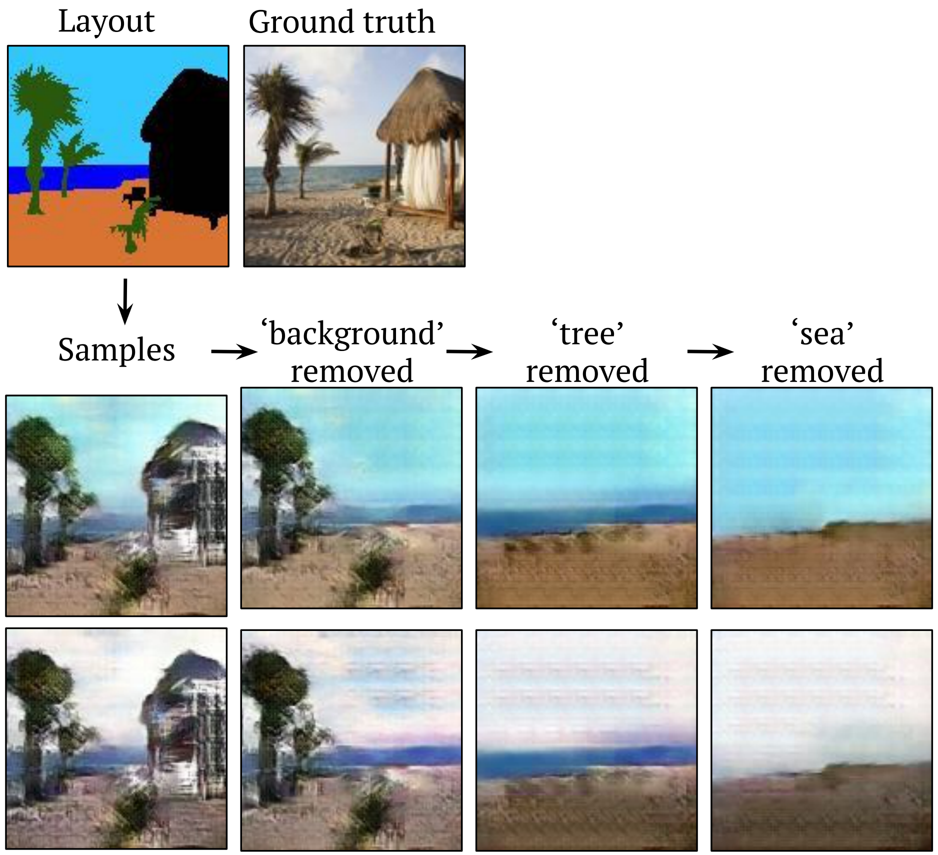
Layouts



Attribute and Layout Conditioned Image Generation(AL-CGAN) Our work

- Object adding / subtracting.

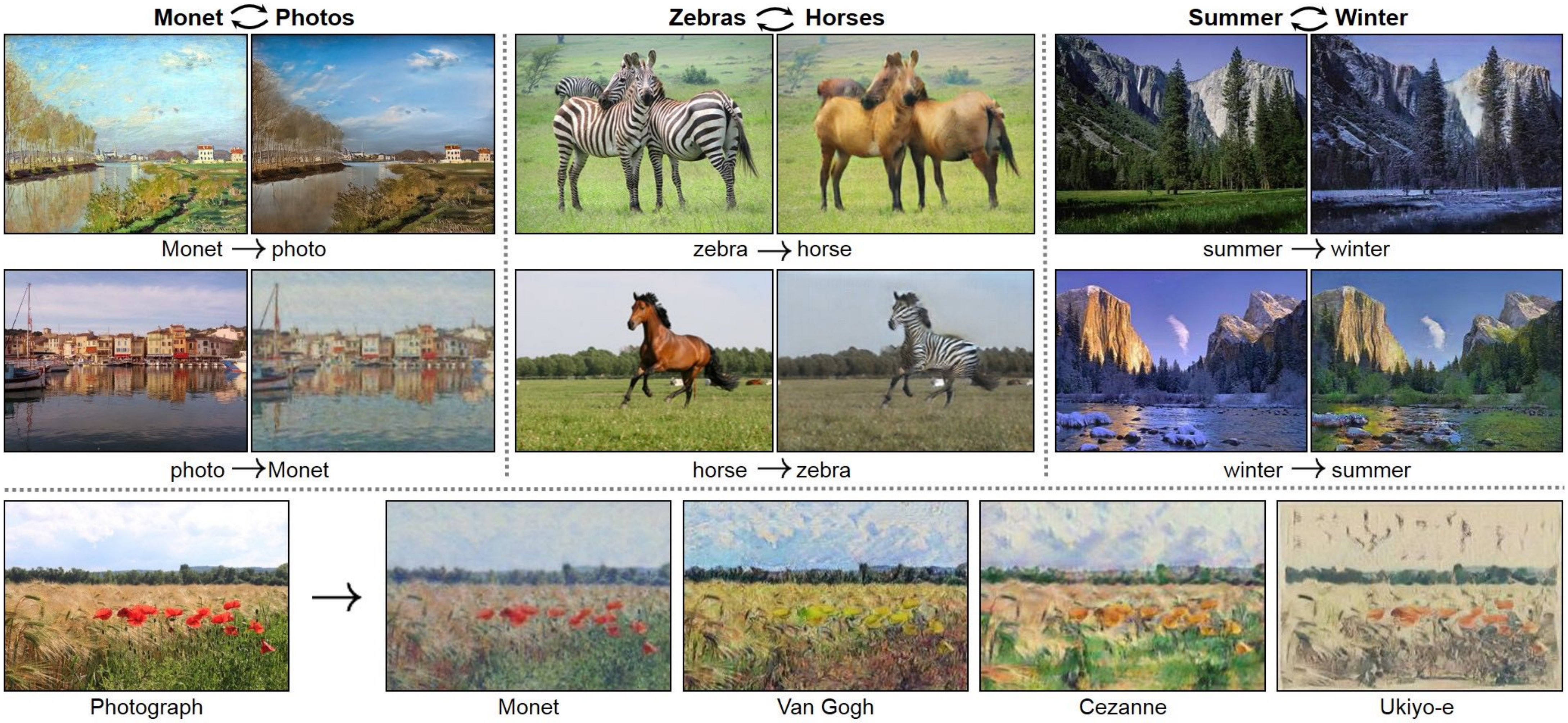




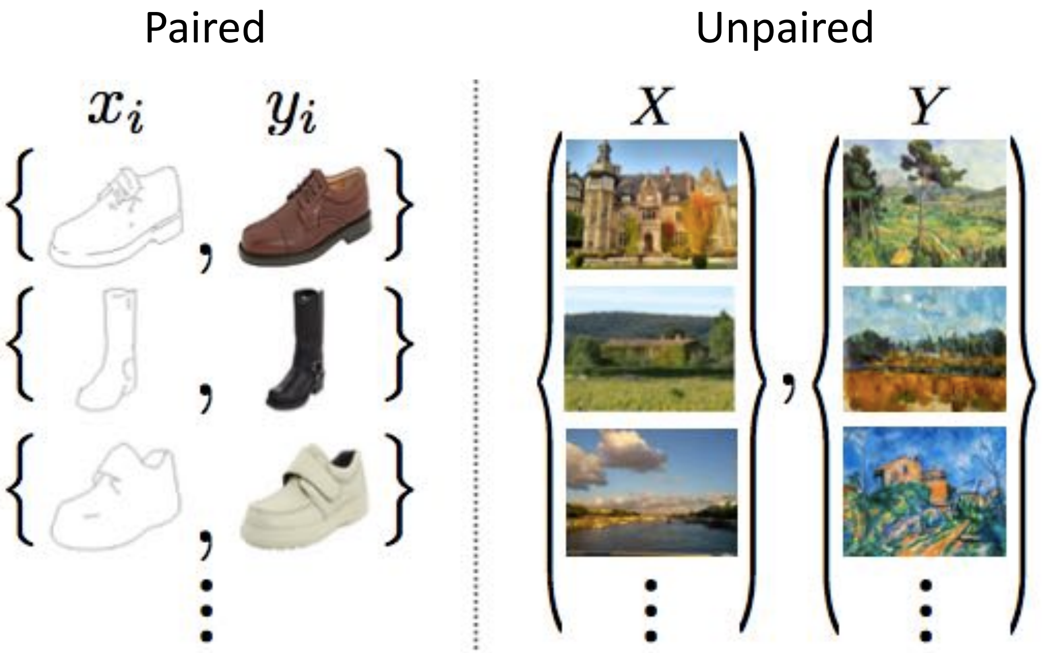
Object Adding

AL-CGAN vs pix2pix

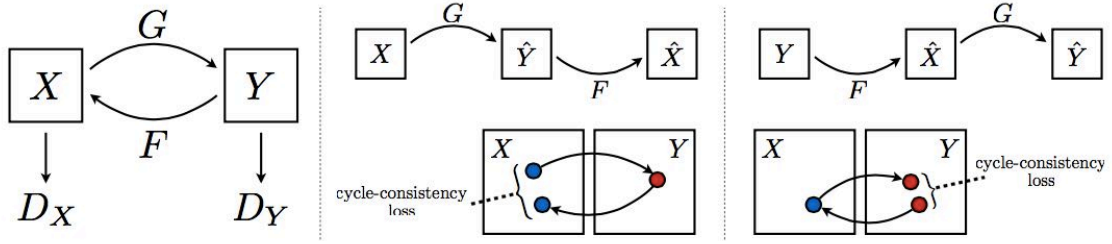
Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017



Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017



$\{x_i, y_i\}_{i=1}^{N} \in X$ $\{x_i\}_{i=1}^{N} \in X$ $\{y\}_{i=1}^{M} \in Y$



Cycle Consistency

"if we translate, e.g., a sentence from English to French, and then translate it back from French to English, we should arrive back at the original sentence. "

$G: X \to Y$ F: $Y \to X$

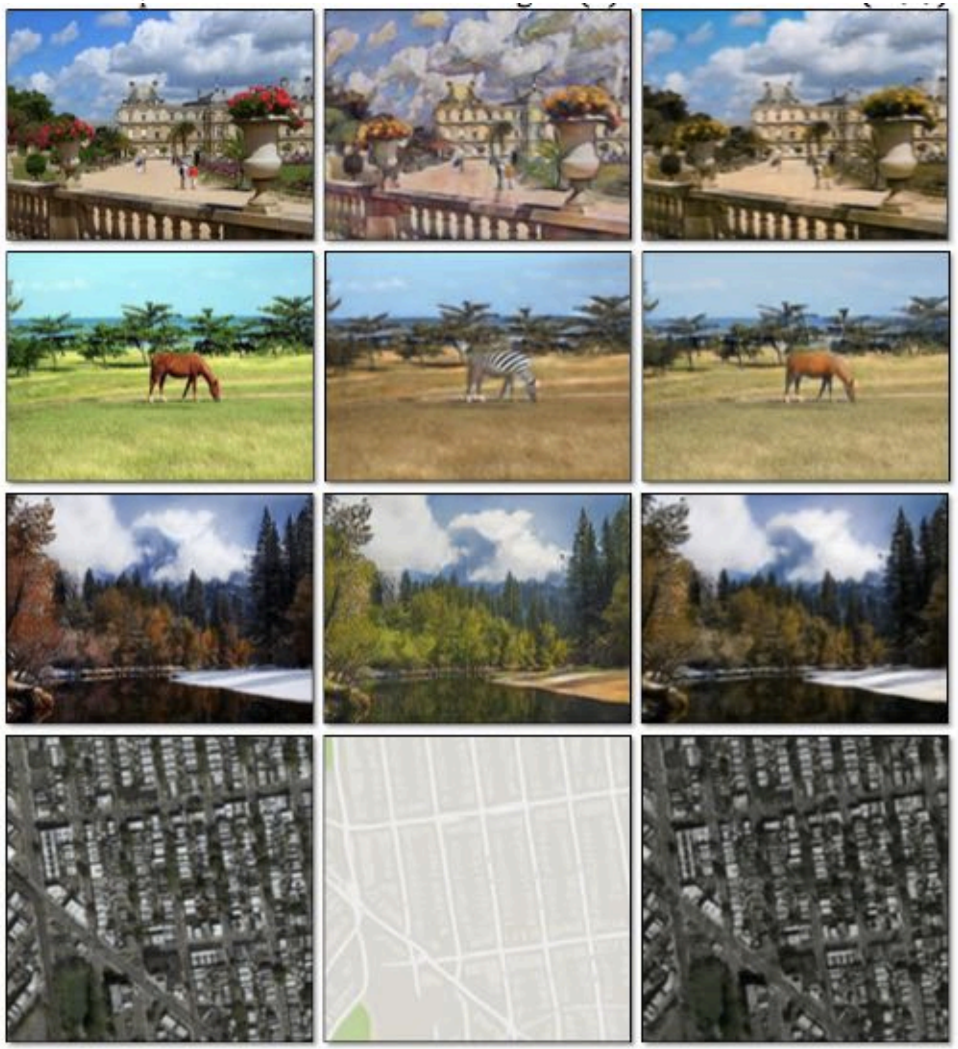
$F(G(x)) \approx x$ G$(G(y)) \approx y$

$$\mathcal{L}_{cyc}(G,F) = \mathbb{E}_{x\sim p_{data}(x)}\left[\left\|F\big(G(x)\big) - x\right\|_1\right] + \mathbb{E}_{y\sim p_{data}(y)}\left[\left\|G\big(F(y)\big) - y\right\|_1\right]$$

$$\mathcal{L}(G,F,D_X,D_Y) = \mathcal{L}_{GAN}(G,D_Y,X,Y) + \mathcal{L}_{GAN}(F,D_x,Y,X) + \lambda\mathcal{L}_{cyc}(G,F)$$

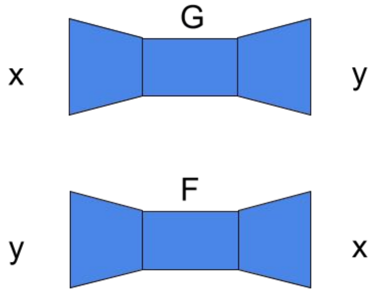
Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017

Input Generated Reconstruction



- Two encoder-decoder networks are jointly trained.

$$FoG: X \to X \text{ ve } GoF: Y \to Y$$

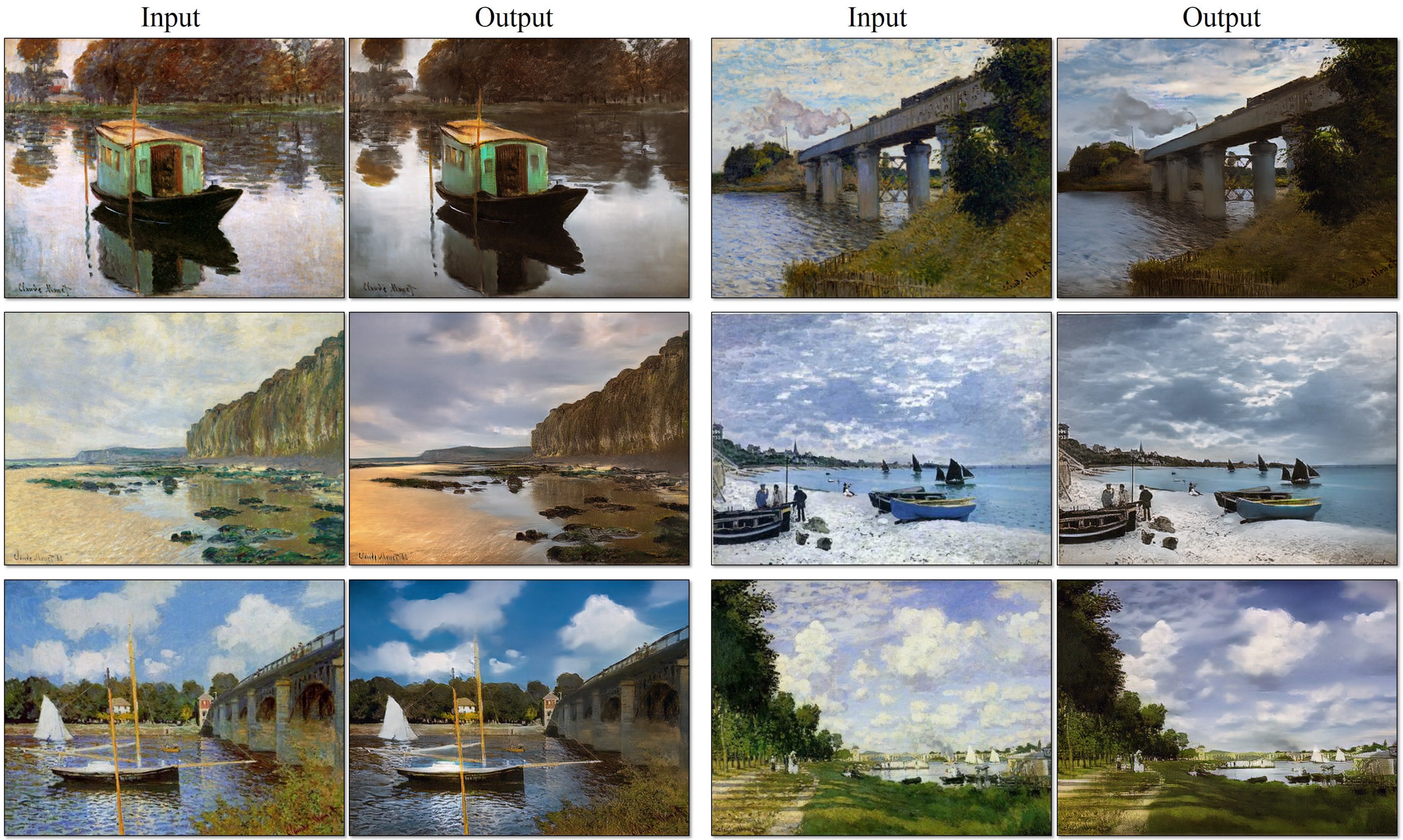


- 70×70 PatchGAN, which try to classify whether 70 × 70 overlapping image patches are real or fake is used.

- Adversarial training.

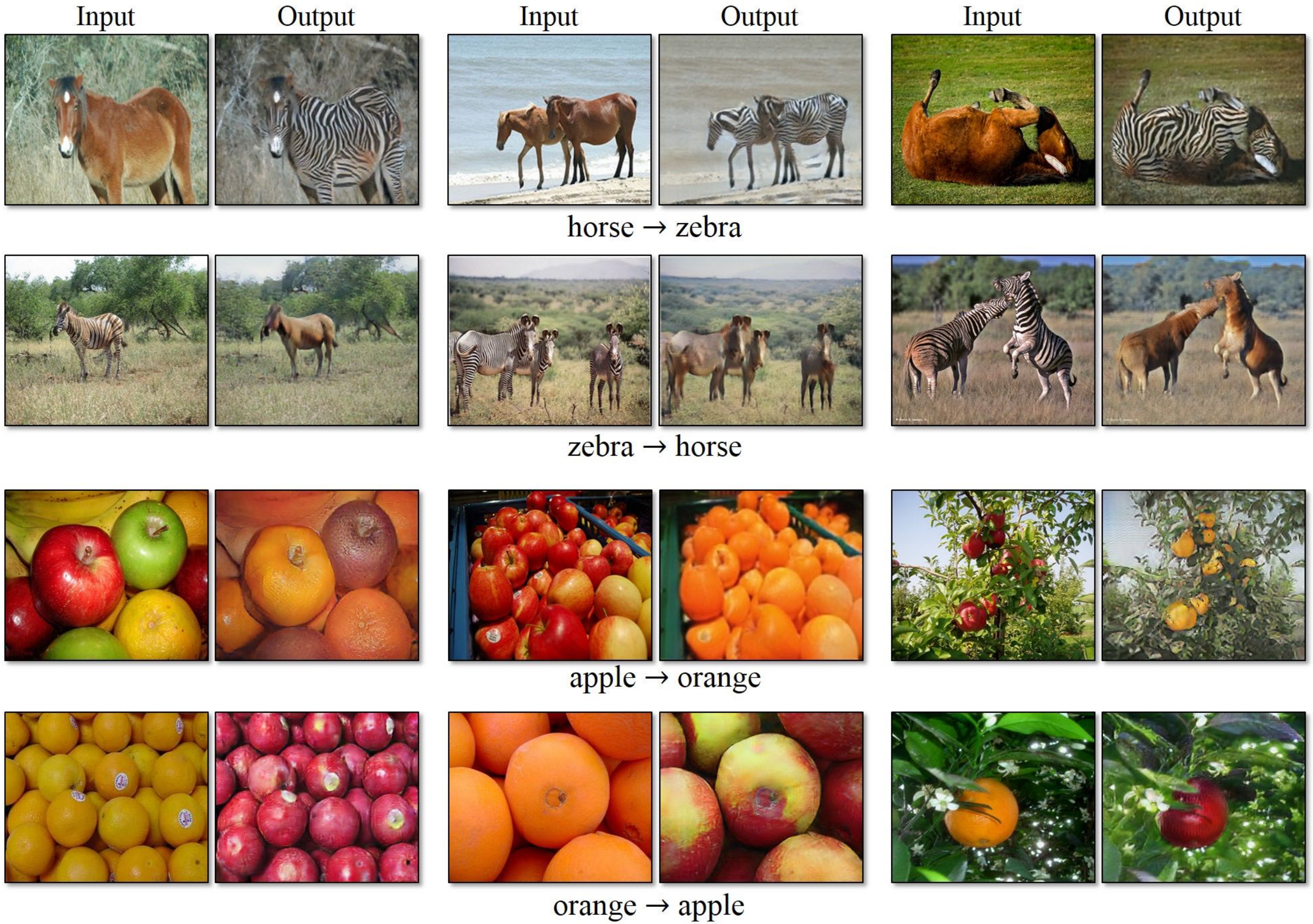
Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017

Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017



Unpaired Image to Image Translation(CycleGAN)

Zhu vd. 2017



Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017

Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017

Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017

Source: https://github.com/tatsuyah/CycleGAN-Models

Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017

Source: https://github.com/tatsuyah/CycleGAN-Models

Unpaired Image to Image Translation(CycleGAN) Zhu vd. 2017

A failure case

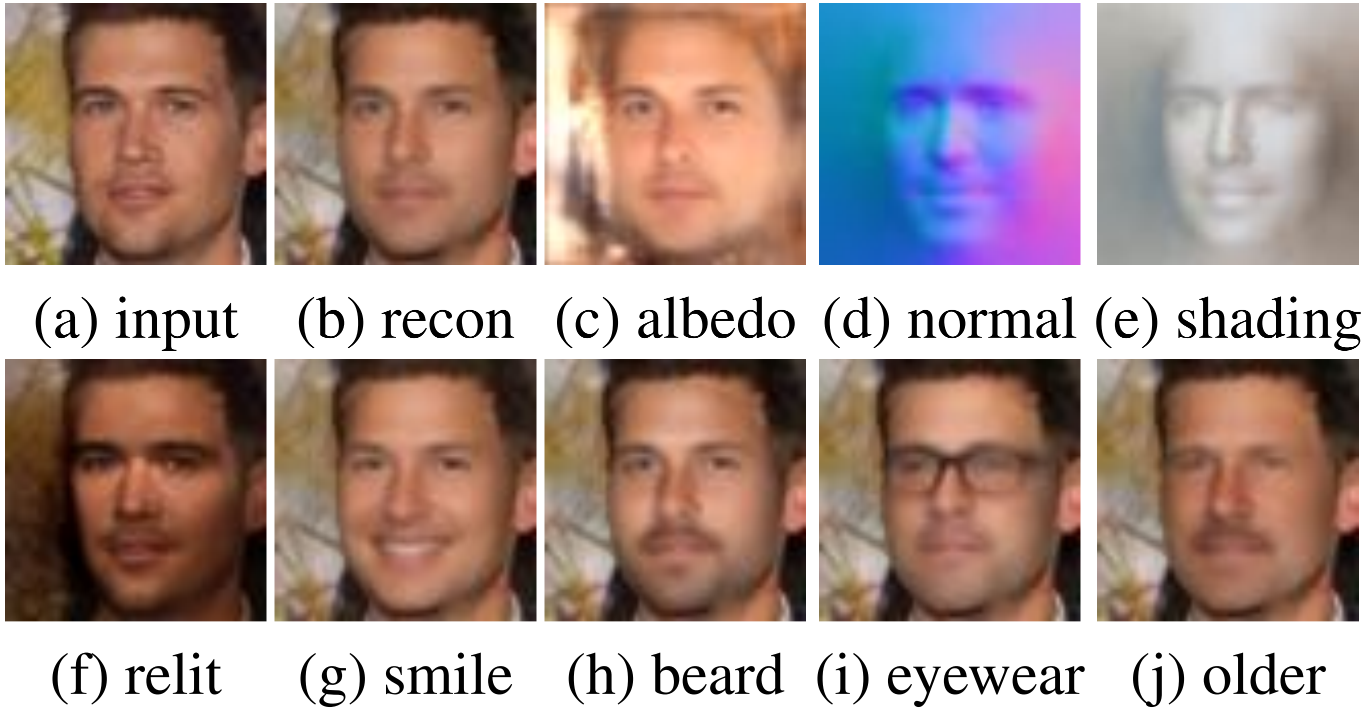
Neural Face Editing with Intrinsic Image Disentangling Shu vd. 2017

- An end-to-end GAN that infers a face-specific disentangled representation of intrinsic face properties.
  - Shape
  - Albedo
  - Lighting
  - Alpha matte
- A given face image I_{fg} is the result of a rendering process: $f_{rendering}$

$$I_{fg} = f_{rendering}(A_e, N_e, L)$$

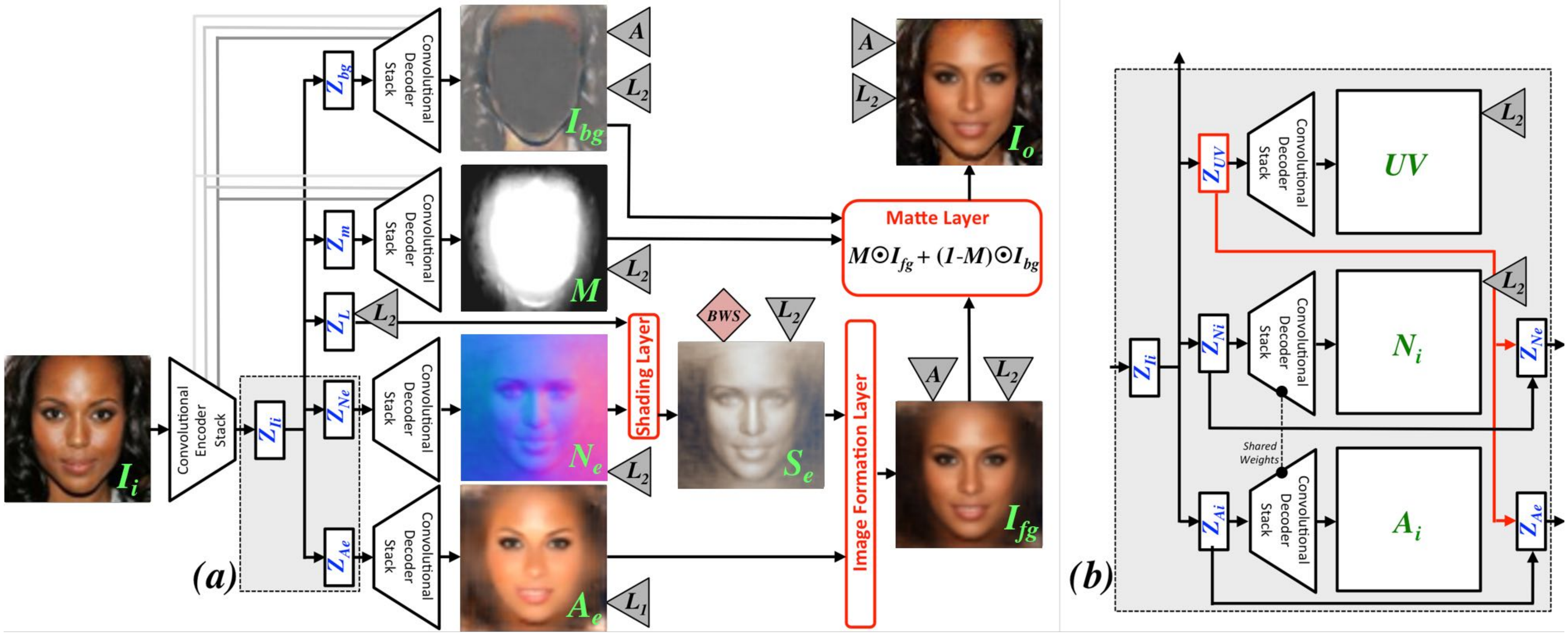
$$I_{fg} = f_{image-formation}(A_e, S_e) = \mathrm{A_e} \odot S_e$$

$$S_e = f_{shading}(N_e, L)$$



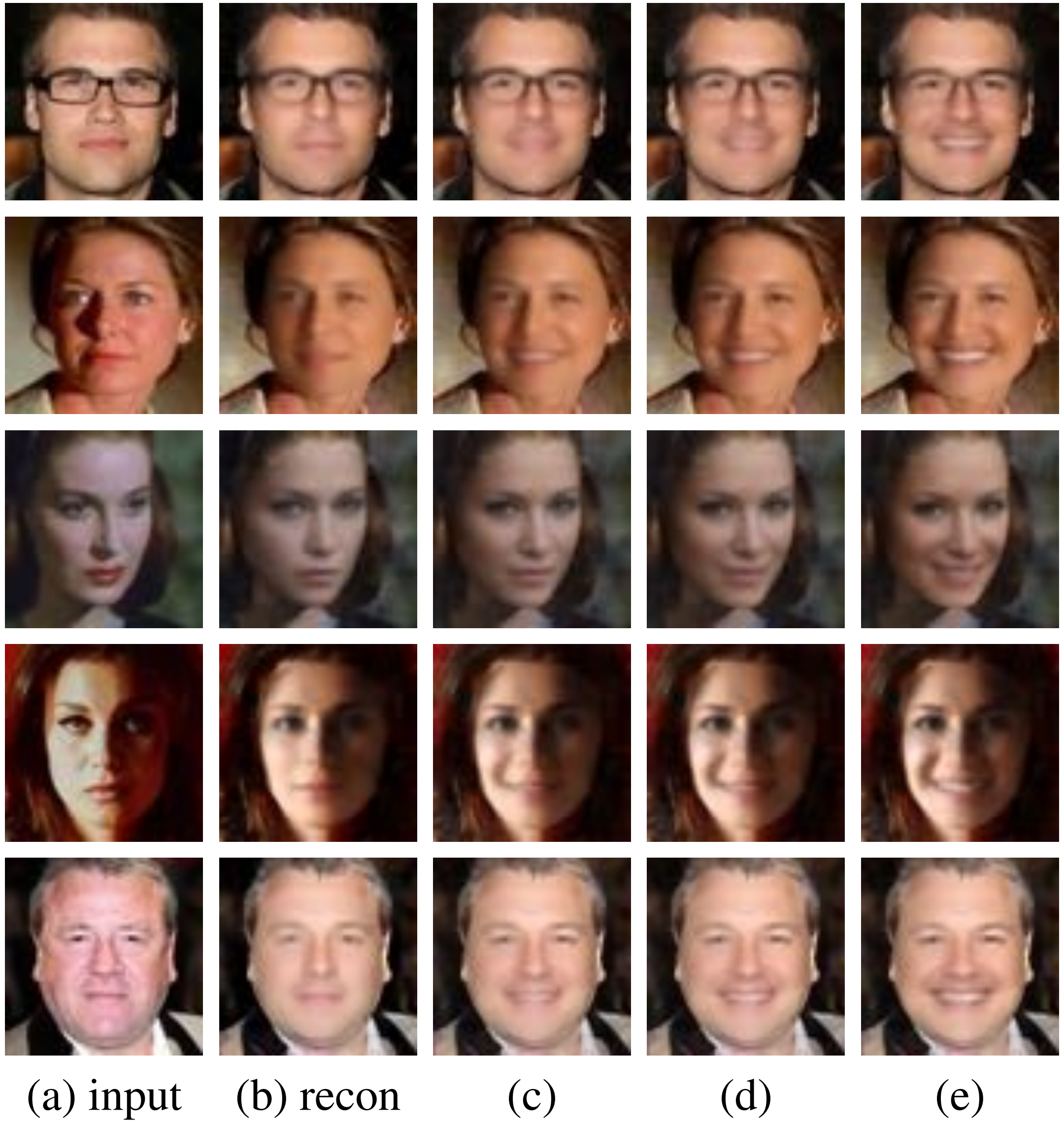
Neural Face Editing with Intrinsic Image Disentangling Shu vd. 2017

$I_{fg} = f_{rendering}(A_e, N_e, L)$ $\quad$ $I_{fg} = f_{image-formation}(A_e, S_e) = A_e \odot S_e$ $\quad$ $S_e = f_{shading}(N_e, L)$



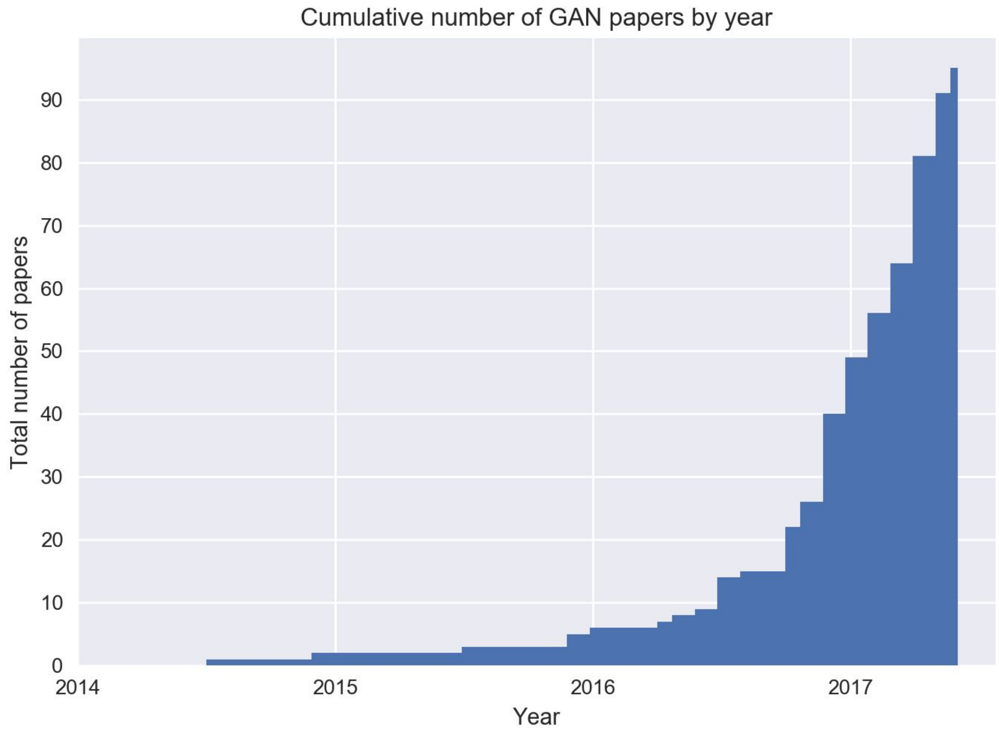
Neural Face Editing with Intrinsic Image Disentangling Shu vd. 2017

Smiling



Aging

Conclusion



- Every week, new GAN papers are coming out.
- Very active topic in Machine Learning and Computer Vision.
- Adversarial loss started to be used for different problems in new papers in premier conferences.
- It has big potential for other areas.