# High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs

Wenjie Niu

2018.08.22

# Indroduction

The method of pix2pix suggest that adversarial training might be unstable and prone to failure for high-resolution image generation tasks.

This paper presents a new method for synthesizing high-resolution(2048 × 1024) photo-realistic images from semantic label maps using conditional generative adversarial networks (conditional GANs).

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Related Work

- Generative adversarial networks
- Image-to-image translation
- Deep visual manipulation

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Method

The objective of the generator G is to translate semantic label maps to realistic-looking images, while the discriminator D aims to distinguish real images from the translated ones.

The paper improves the pix2pix framework by using a coarse-to-fine generator, a multi-scale discriminator architecture, and a robust adversarial learning objective function.

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Method

- Coarse-to-fine generator

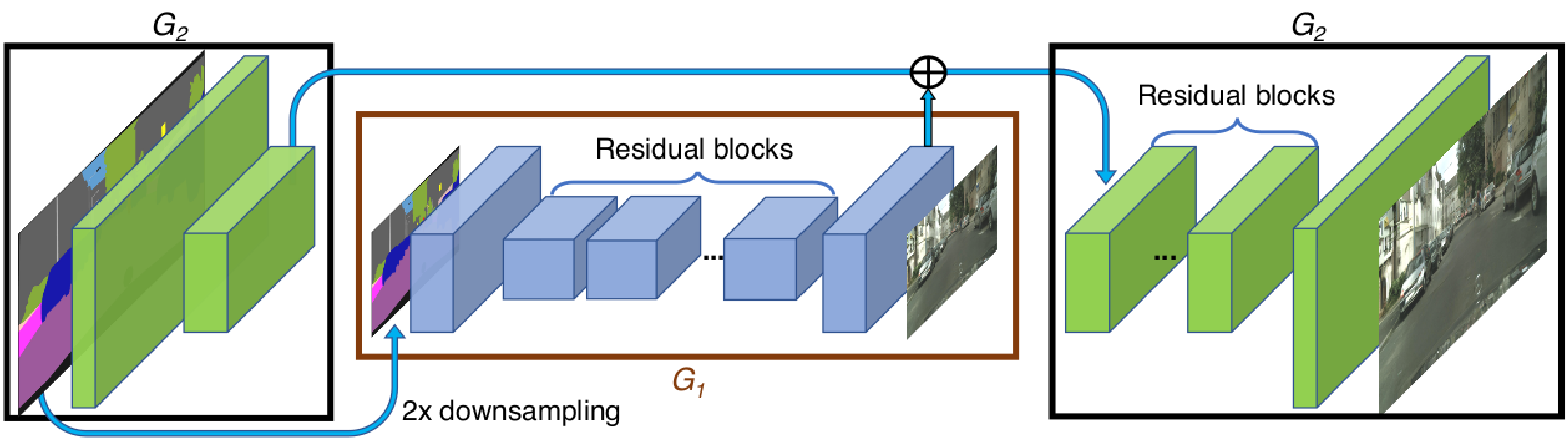
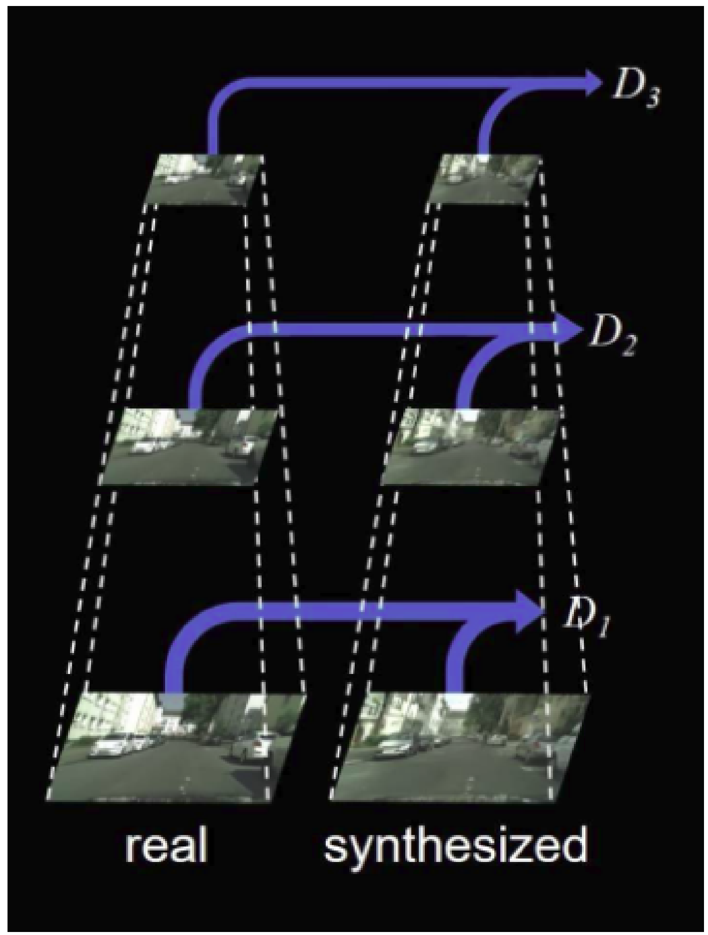



Figure 2: Network architecture of our generator. We first train a residual network $G_1$ on lower resolution images. Then, another residual network $G_2$ is appended to $G_1$ and the two networks are trained jointly on high resolution images. Specifically, the input to the residual blocks in $G_2$ is the element-wise sum of the feature map from $G_2$ and the last feature map from $G_1$.

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Method

- Multi-scale discriminators




T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Method

- Improved adversarial loss

The feature matching loss $\mathcal{L}_{\text{FM}}(G, D_k)$ is then calculated as:

$$\mathcal{L}_{\text{FM}}(G, D_k) = \mathbb{E}_{(\mathbf{s},\mathbf{x})} \sum_{i=1}^{T} \frac{1}{N_i} [||D_k^{(i)}(\mathbf{s}, \mathbf{x}) - D_k^{(i)}(\mathbf{s}, G(\mathbf{s}))||_1],$$

Our full objective combines both GAN loss and feature matching loss as:

$$\min_G \left( \left( \max_{D_1, D_2, D_3} \sum_{k=1,2,3} \mathcal{L}_{\text{GAN}}(G, D_k) \right) + \lambda \sum_{k=1,2,3} \mathcal{L}_{\text{FM}}(G, D_k) \right)$$

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Method

- Using Instance Maps

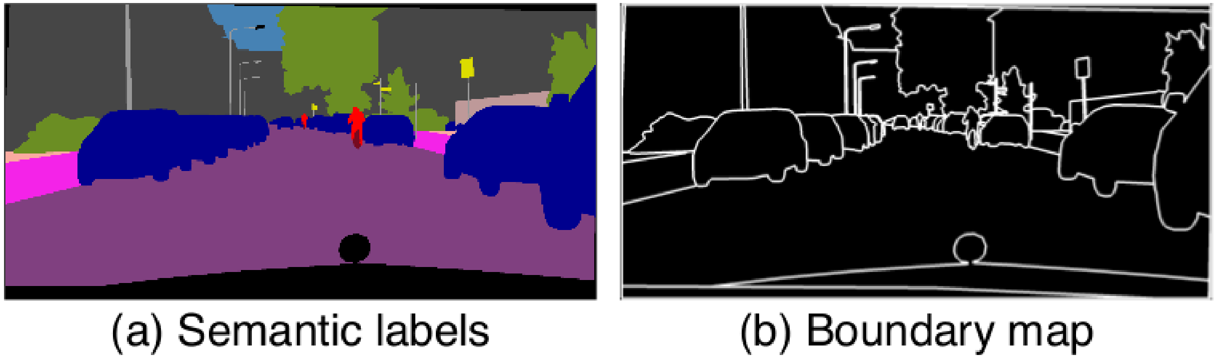
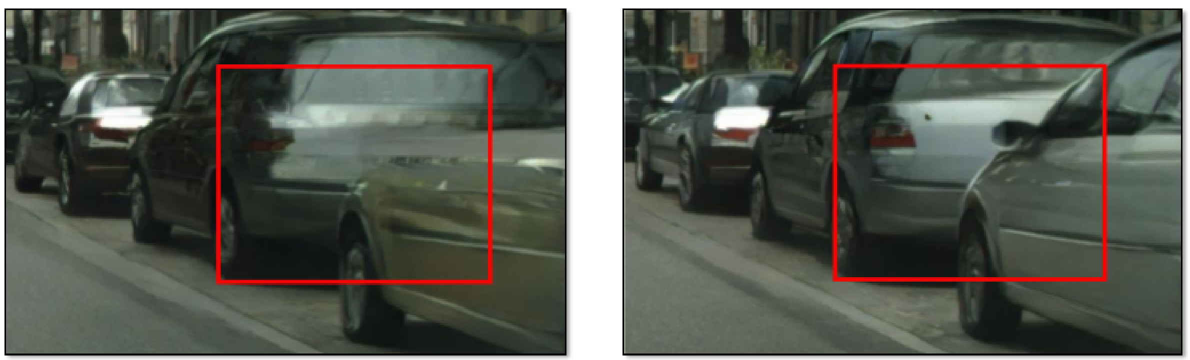



Figure 3: Using instance maps: (a) a typical semantic label map. Note that all connected cars have the same label, which makes it hard to tell them apart. (b) The extracted instance boundary map. With this information, separating different objects becomes much easier.

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Method

- Using Instance Maps



(a) Using labels only (b) Using label + instance map

Figure 4: Comparison between results without and with instance maps. It can be seen that when instance boundary information is added, adjacent cars have sharper boundaries.

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Results

- Quantitative Comparisons

| | pix2pix [21] | CRN [5] | Ours | Oracle |
|---|---|---|---|---|
| Pixel acc | 78.34 | 70.55 | **83.78** | 84.29 |
| Mean IoU | 0.3948 | 0.3483 | **0.6389** | 0.6857 |

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Analysis

| | U-Net [21,43] | CRN [5] | Our generator |
|---|---|---|---|
| Pixel acc (%) | 77.86 | 78.96 | **83.78** |
| Mean IoU | 0.3905 | 0.3994 | **0.6389** |

| | single D | multi-scale Ds |
|---|---|---|
| Pixel acc (%) | 82.87 | **83.78** |
| Mean IoU | 0.5775 | **0.6389** |

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Analysis

(a) pix2pix

(b) CRN

(c) Ours (w/o VGG loss)

(d) Ours (w/ VGG loss )

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

# Q&A

T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.