High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs

Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, Bryan Catanzaro

Paper Reading Li Na

Synthesizing high-resolution photo-realistic images from semantic label maps

Existing problem

- (1) the difficulty of generating high-resolution images with GANs
- (2) the lack of de- tails and realistic textures in the previous high-resolution results

Improving Photorealism and Resolution

- a coarse-to- fine generator
- a multi-scale discriminator architecture
- a robust adversarial learning objective function

Coarse-to-fine generator

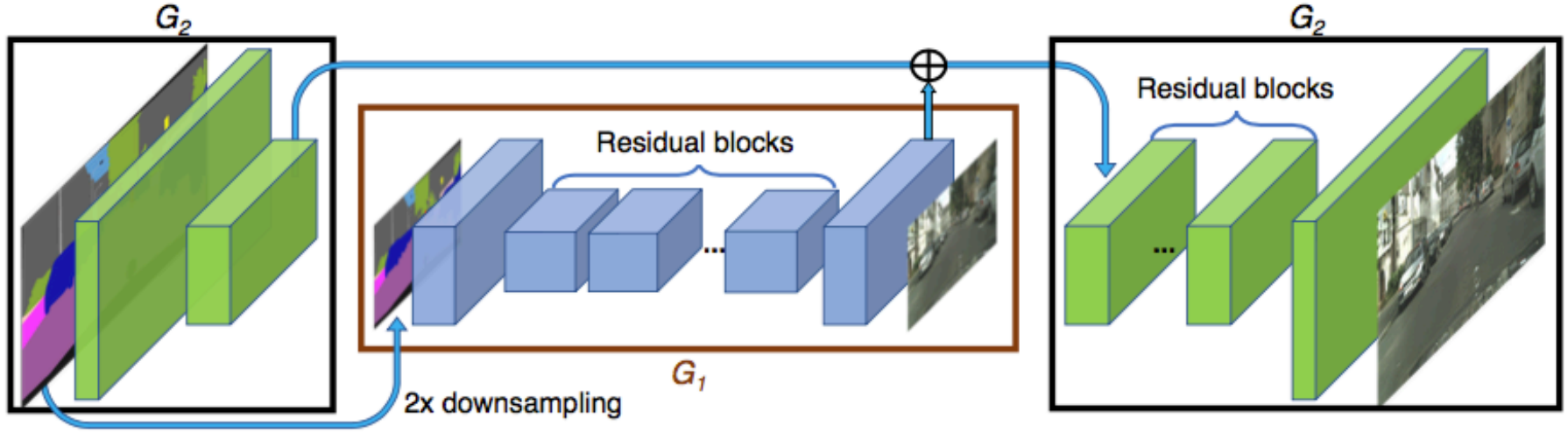


Figure 2: Network architecture of our generator. We first train a residual network G_1 on lower resolution images. Then, another residual network G_2 is appended to G_1 and the two networks are trained jointly on high resolution images. Specifically, the input to the residual blocks in G_2 is the element-wise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

Multi-scale discriminators

$$\min_{G} \max_{D_1, D_2, D_3} \sum_{k=1,2,3} \mathcal{L}_{\text{GAN}}(G, D_k)$$

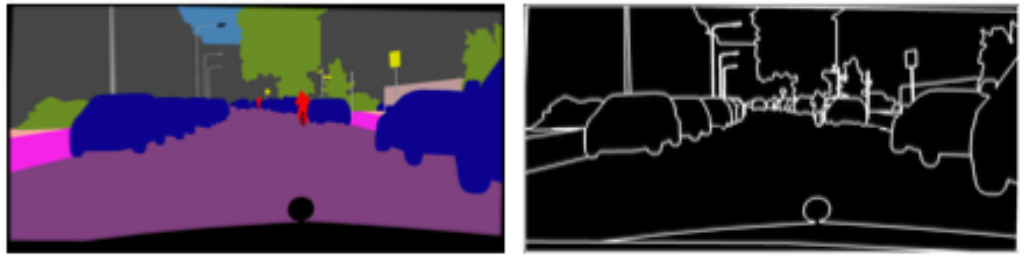
Improved adversarial loss

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

Full objective: GAN loss + feature matching loss

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

Using Instance Maps



(a) Semantic labels (b) Boundary map

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.

(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.

Learning an Instance-level Feature Embedding

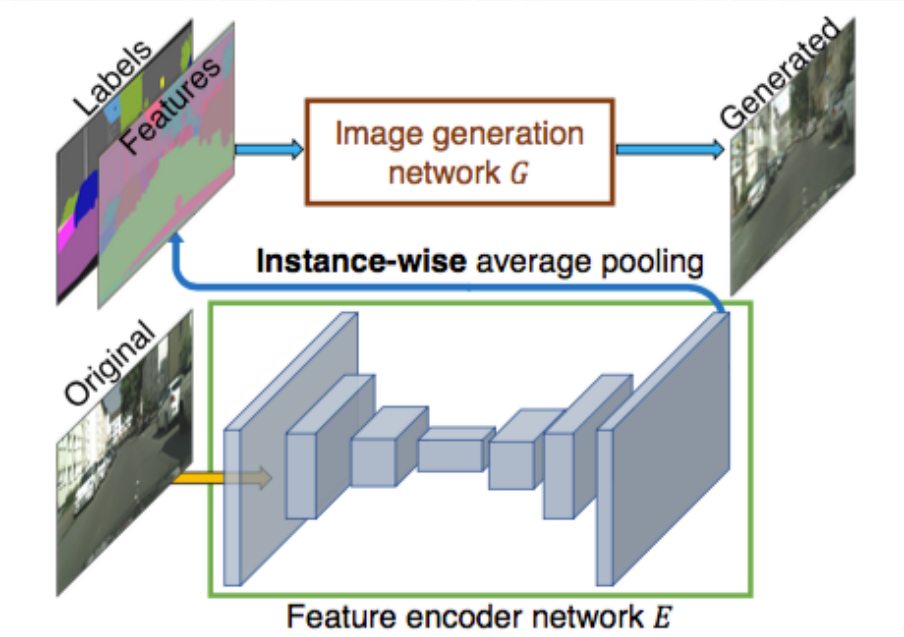


Figure 5: Using instance-wise features in addition to labels for generating images.

Quantitative Comparisons

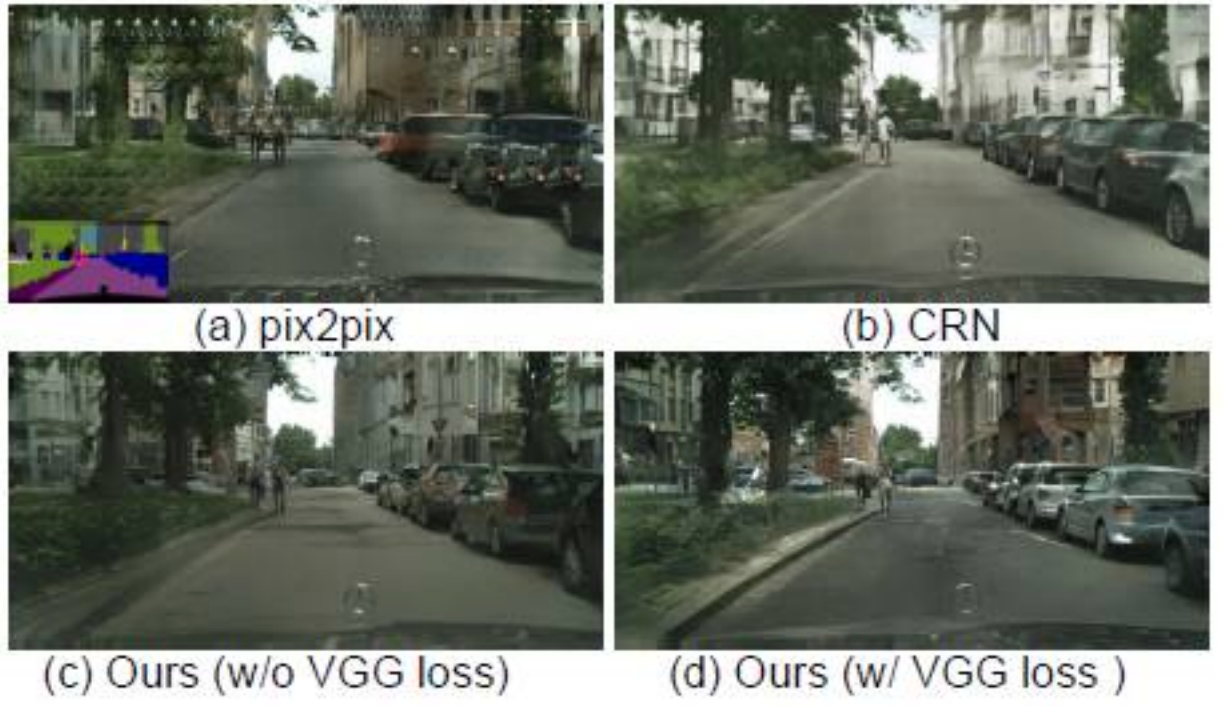
- pixel-wise accuracy
- mean intersection-over-union (IoU)

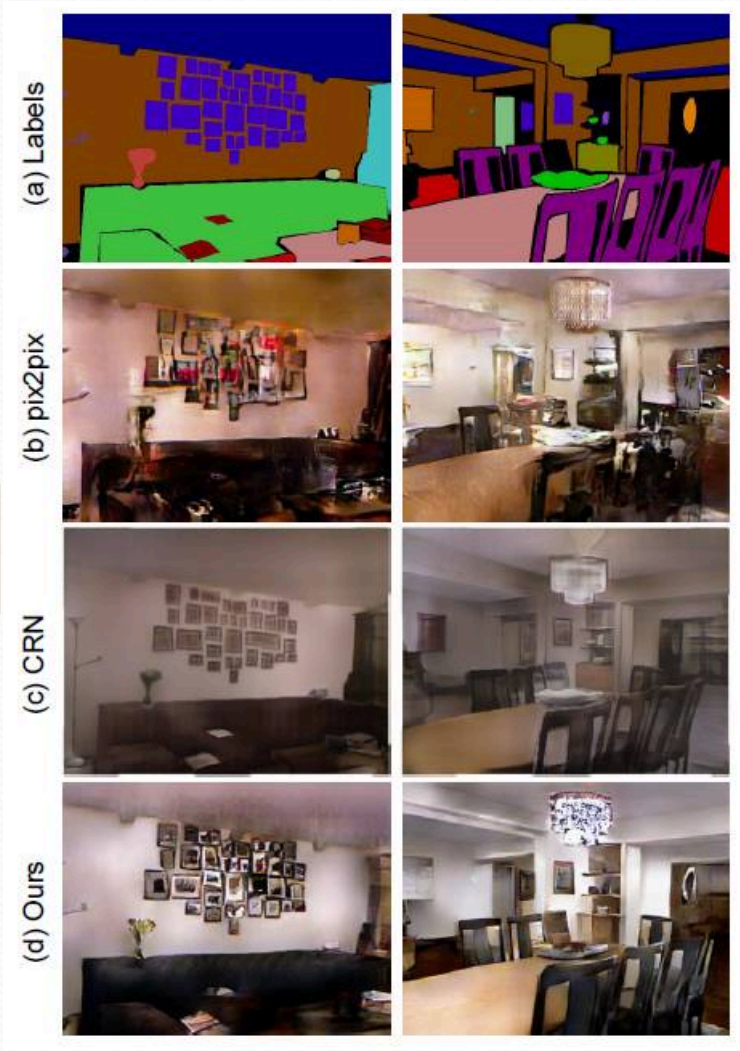
Cityscapes dataset

	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

	pix2pix [21]	CRN [5]
Ours	93.8%	86.2%
Ours (w/o VGG)	94.6%	85.2%

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





NYU indoor RGBD dataset

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