High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs

By Clayton Barham

Contents

- Motivation
- Pix2pix Baseline
- Improving Photorealism and Resolution
- Using Instance Maps
- Learning an Instance Level Feature Embedding
- Results
- Interactive Object Editing

Motivation

- Synthesize high resolution, photorealistic images from semantic label maps using conditional GANs
  - Pix2pix – Low resolution
  - Perceptual loss – Lacking detail, repeated patterns

- Allow objects in the image to be added, removed, or edited

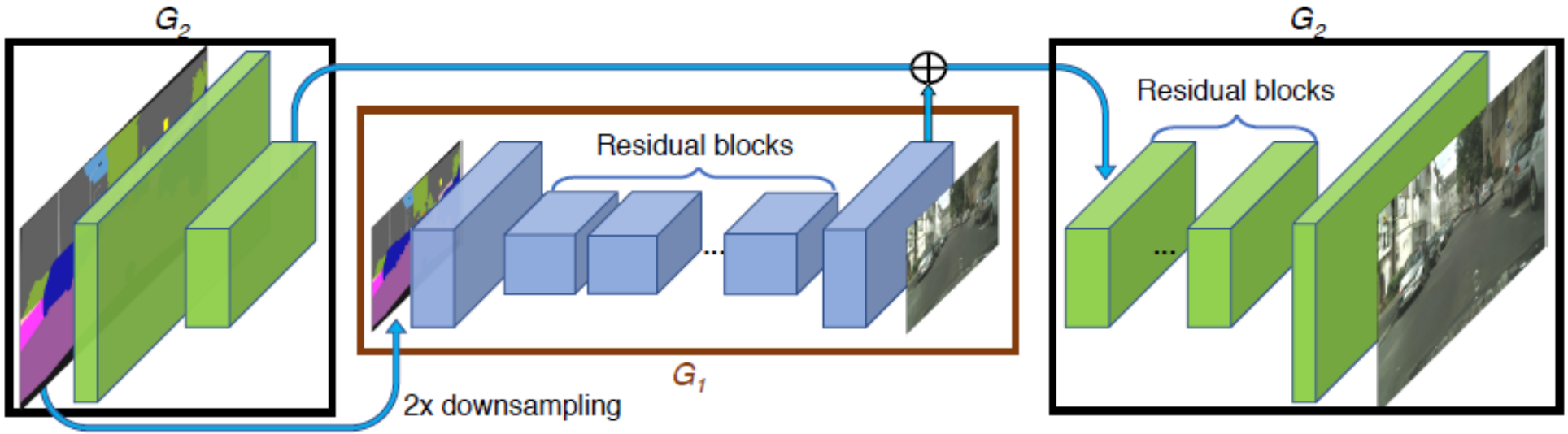
pix2pix Baseline



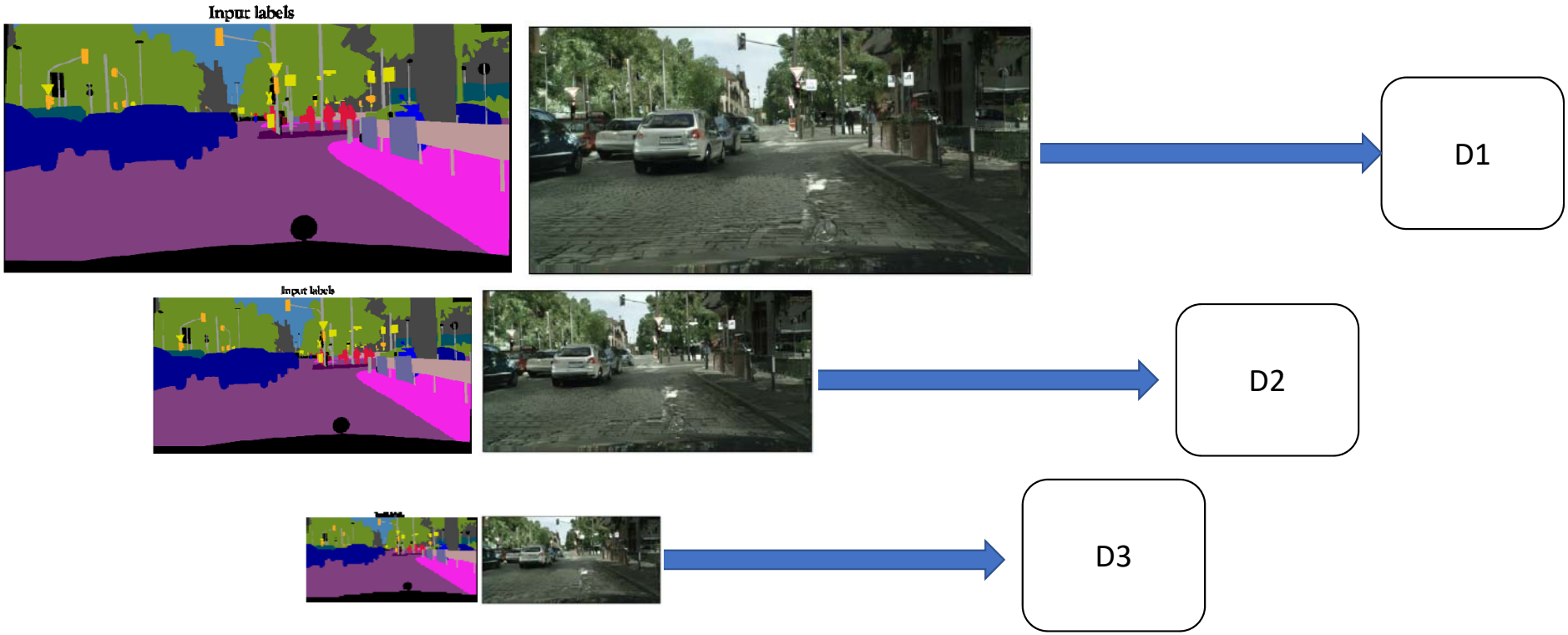
Blurry Quality

Visual Artifacts

Improving Photorealism and Resolution: Coarse to Fine Generator



Improving Photorealism and Resolution: Multi-Scale Discriminators



Improving Photorealism and Resolution: Improved Adversarial Loss and Perceptual Loss

$$\mathbb{E}_{(\mathbf{s},\mathbf{x})}[\log D(\mathbf{s},\mathbf{x})] + \mathbb{E}_{\mathbf{s}}[\log(1 - D(\mathbf{s}, G(\mathbf{s}))].$$

GAN Loss

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

Feature Matching Loss

$$\lambda \sum_{i=1}^{N} \frac{1}{M_i} [||F^{(i)}(\mathbf{x}) - F^{(i)}(G(\mathbf{s}))||_1]$$

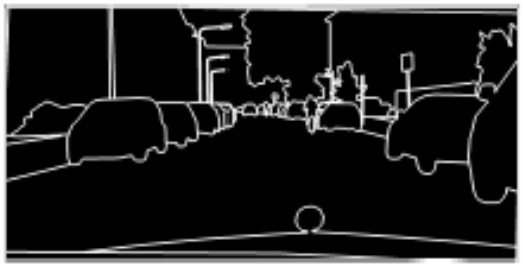
Perceptual Loss

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

Using Instance Maps



(a) Semantic labels



(b) Boundary map

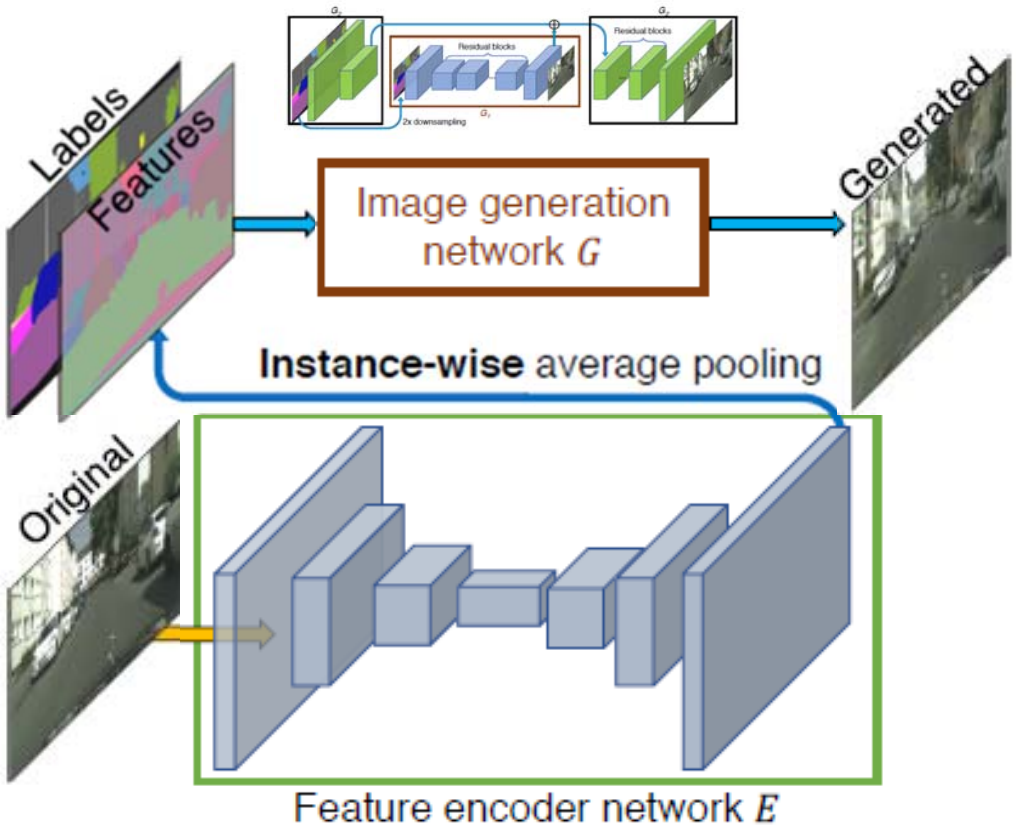


(a) Using labels only



(b) Using label + instance map

Learning an Instance Level Feature Embedding



Learning an Instance Level Feature Embedding

- Encoder-Decoder network learns feature values
- Average Pooling layer standardizes feature values for each instance
- Feature vectors concatenated with other input to generator

Learning an Instance Level Feature Embedding







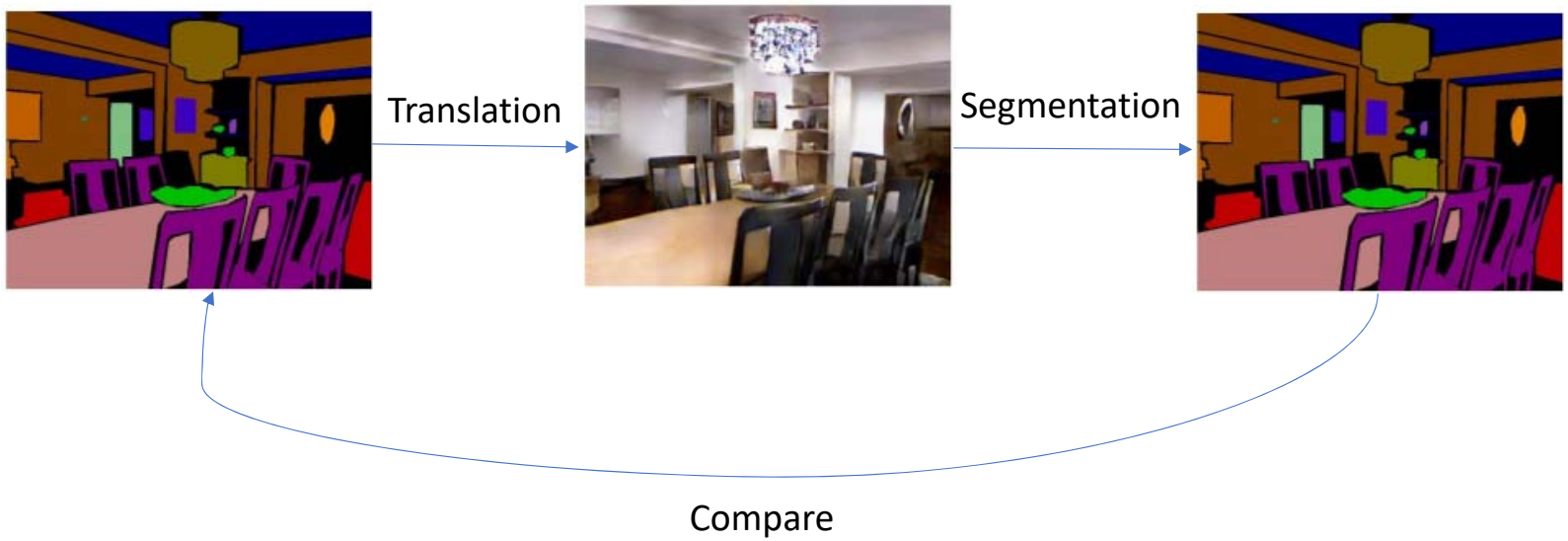




Learning an Instance Level Feature Embedding

- Perform K-Means clustering on per class basis
- Each cluster corresponds to one ‘style’
- Make centroid of each cluster exemplar of that style
- Substitute centroid values to edit object styles

Results: Quantitative Comparison



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

Results: Perceptual Study



pix2pix



CRN

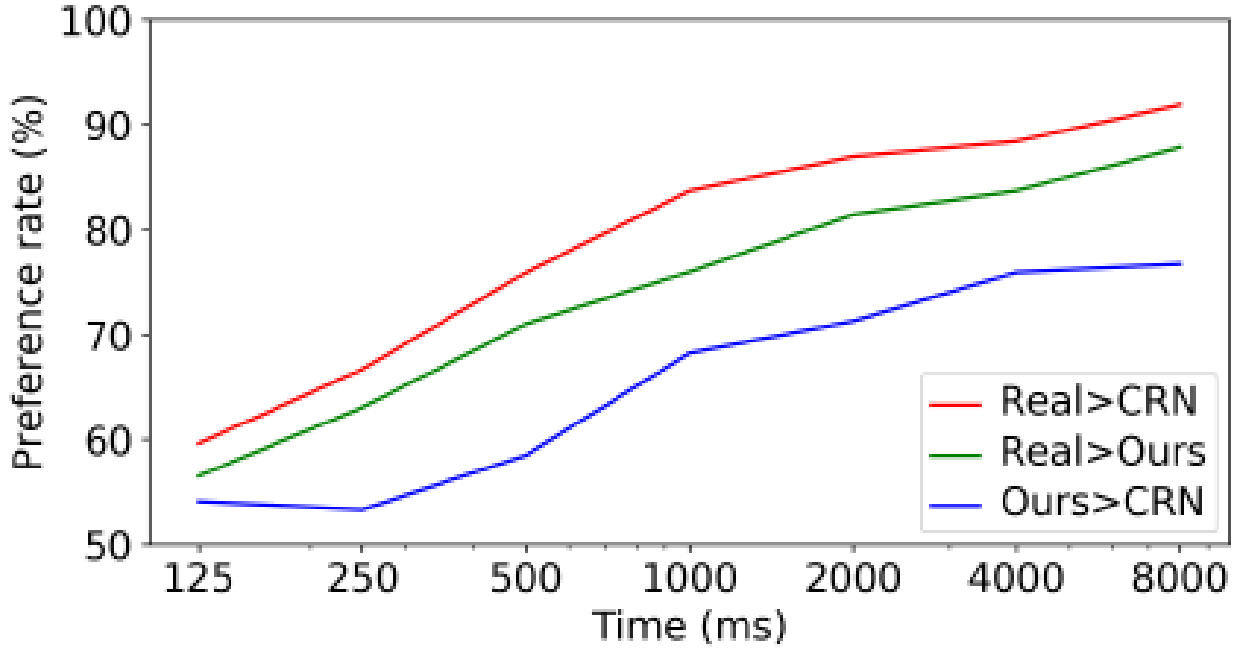


Coarse to Fine

Results: Human Perceptual Study – Unlimited Time

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

Results: Human Perceptual Study – Limited Time



Results: Generator Comparison

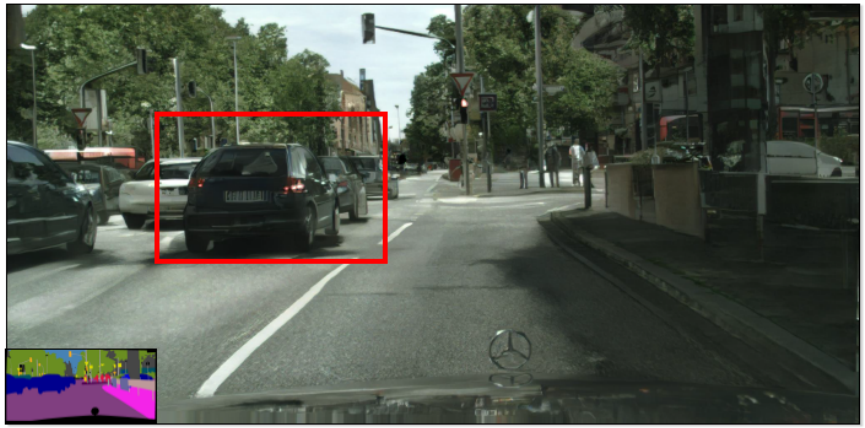
	U-Net [21, 43]	CRN [5]	Our generator
Pixel acc (%)	77.86	78.96	**83.78**
Mean IoU	0.3905	0.3994	**0.6389**

	U-Net [21, 43]	CRN [5]
Our generator	80.0%	76.6%

Results: Discriminator Comparison

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

Interactive Object Editing



(a) Synthesized result



(b) Application: Change label types



(c) Application: Edit object appearance

Thank you