

pix2pixHD: High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs

Ting-Chun Wang¹, Ming-Yu Liu¹, Jun-Yan Zhu², Andrew Tao¹, Jan Kautz¹, Bryan Catanzaro¹
NVIDIA¹, University of California Berkeley²



Outline

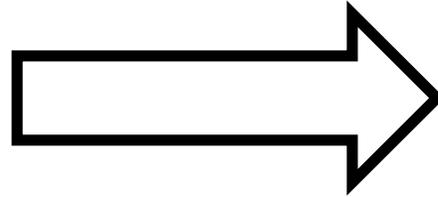
- Introduction
- Related work
- Method
- Results
- Applications
- Conclusion

Outline

- Introduction

Introduction

Semantically
editable
input image

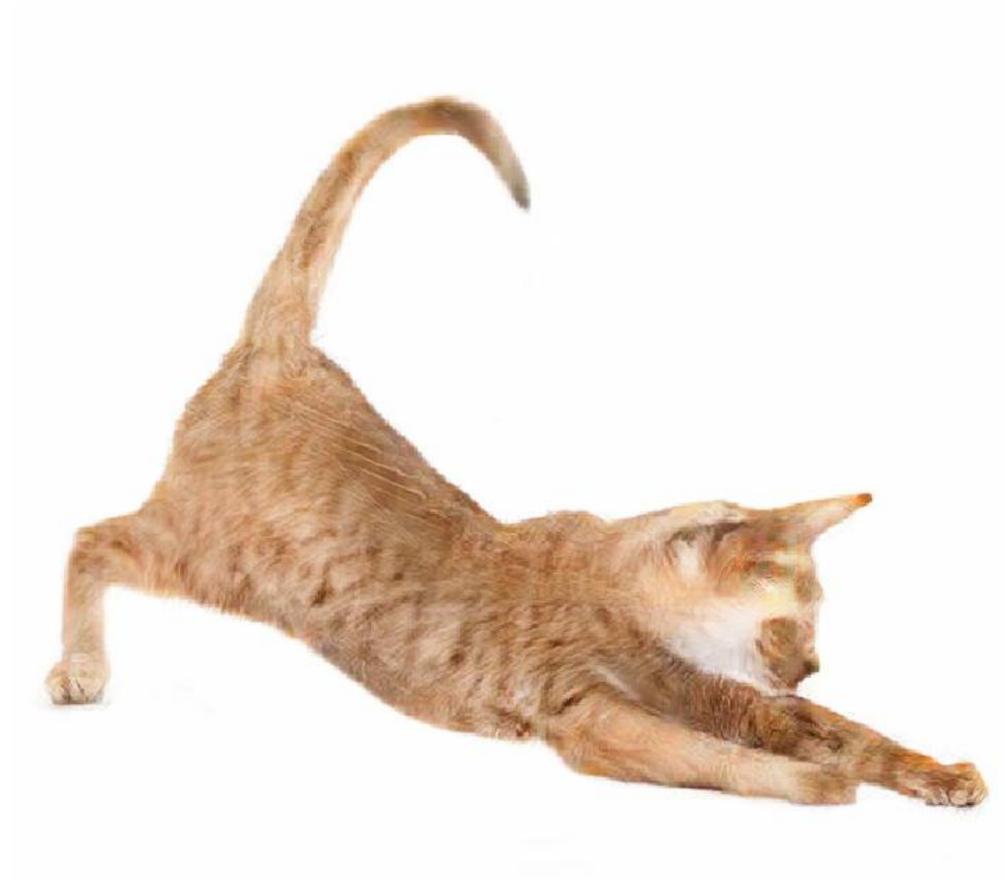
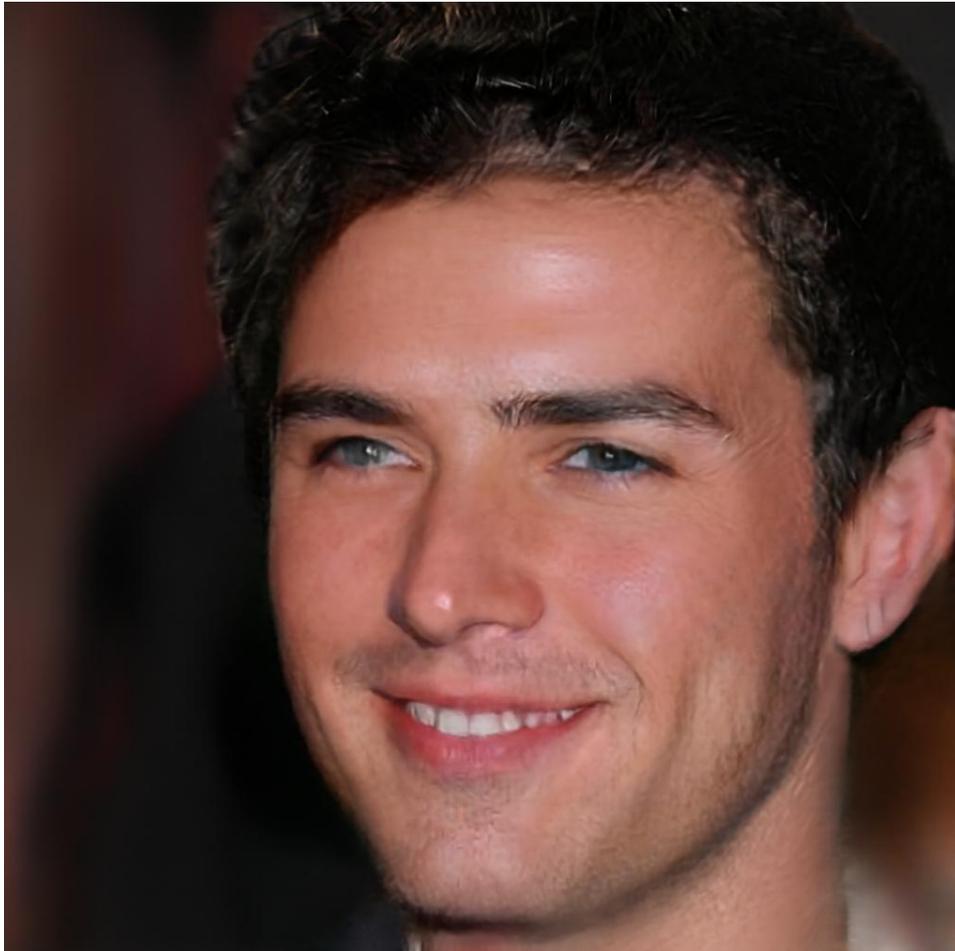


Photorealistic
output
image

Introduction



Introduction



Outline

- Related work

Generative Adversarial Networks (GANs)

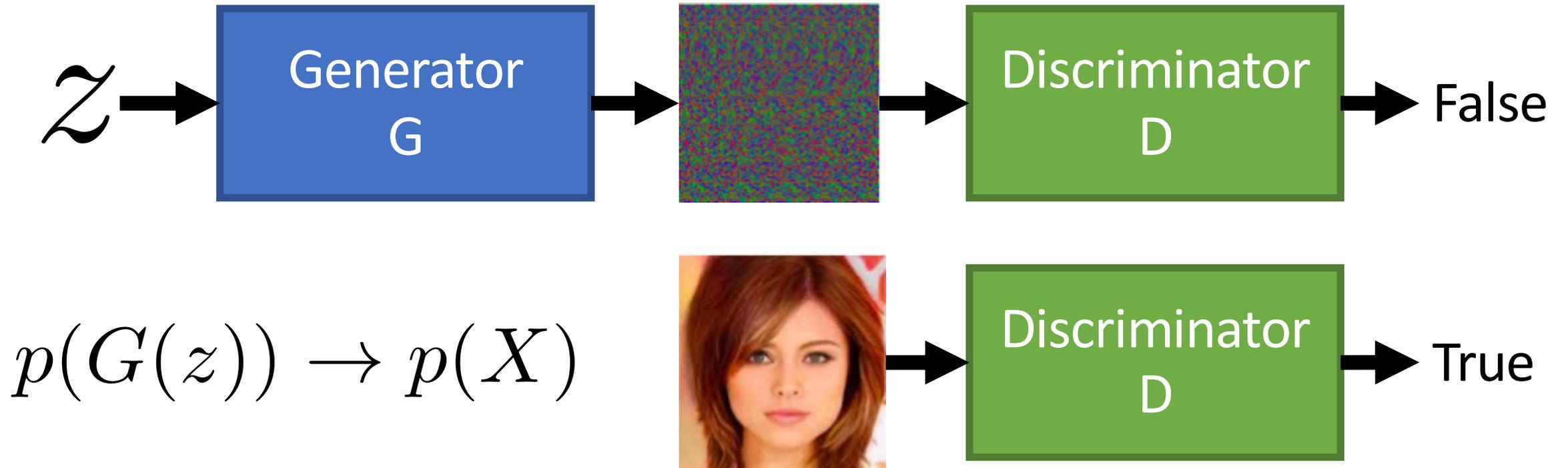


Image-to-Image Translation Framework

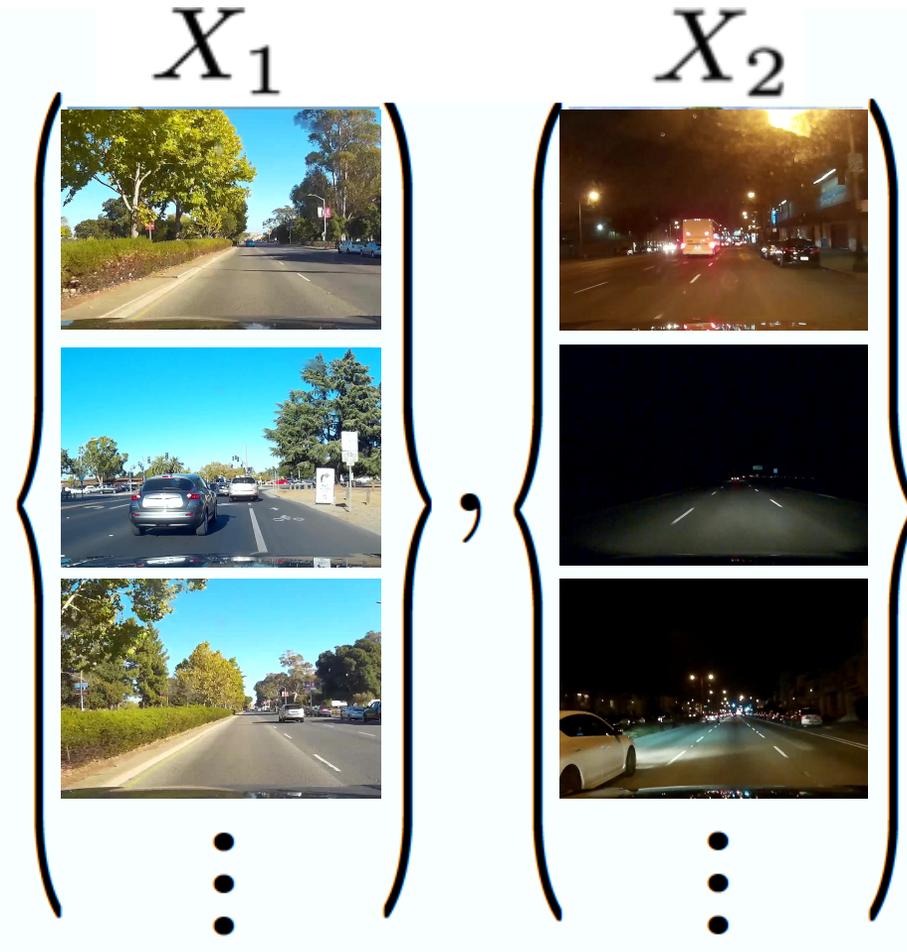
	Supervised	Unsupervised
Unimodal	pix2pix, CRN, SRGAN	DiscoGAN, CycleGAN, UNIT, DTN, DualGAN, StarGAN,
Multimodal	pix2pixHD, BicycleGAN	MUNIT, Augmented CycleGAN

Supervised vs Unsupervised

Supervised

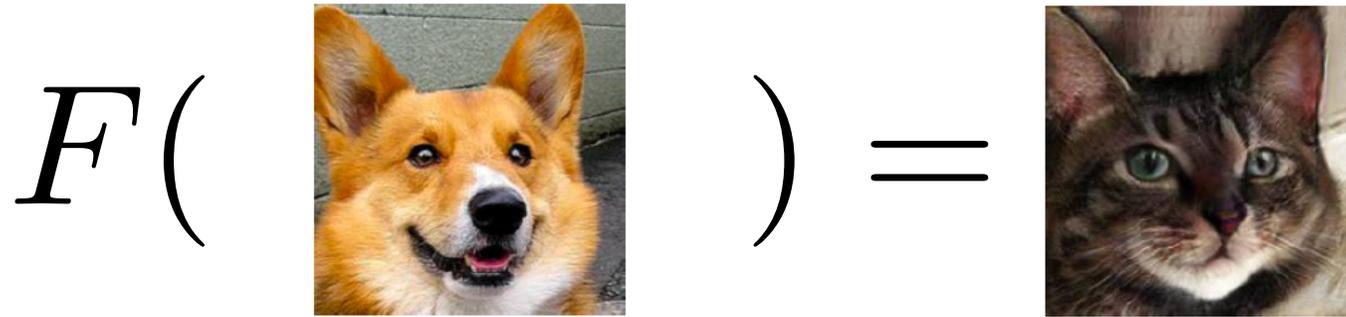


Unsupervised

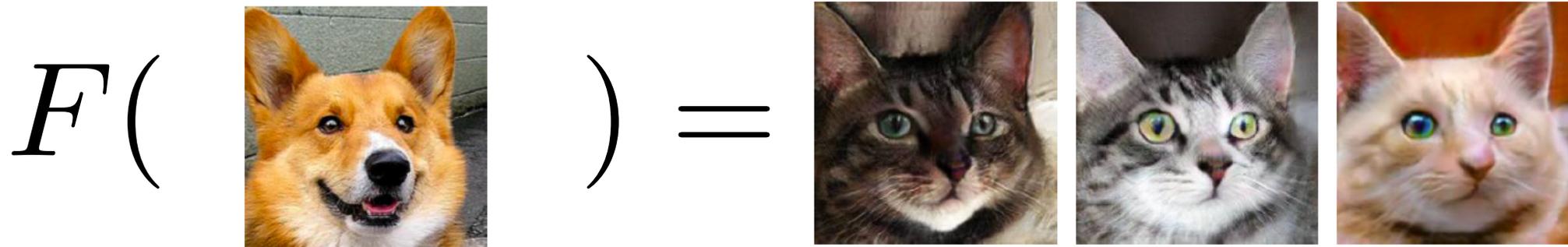


Unimodal vs Multimodal

Unimodal $p(Y|X) = \delta(F(X))$



Multimodal $p(Y|X) = F(X, S)$

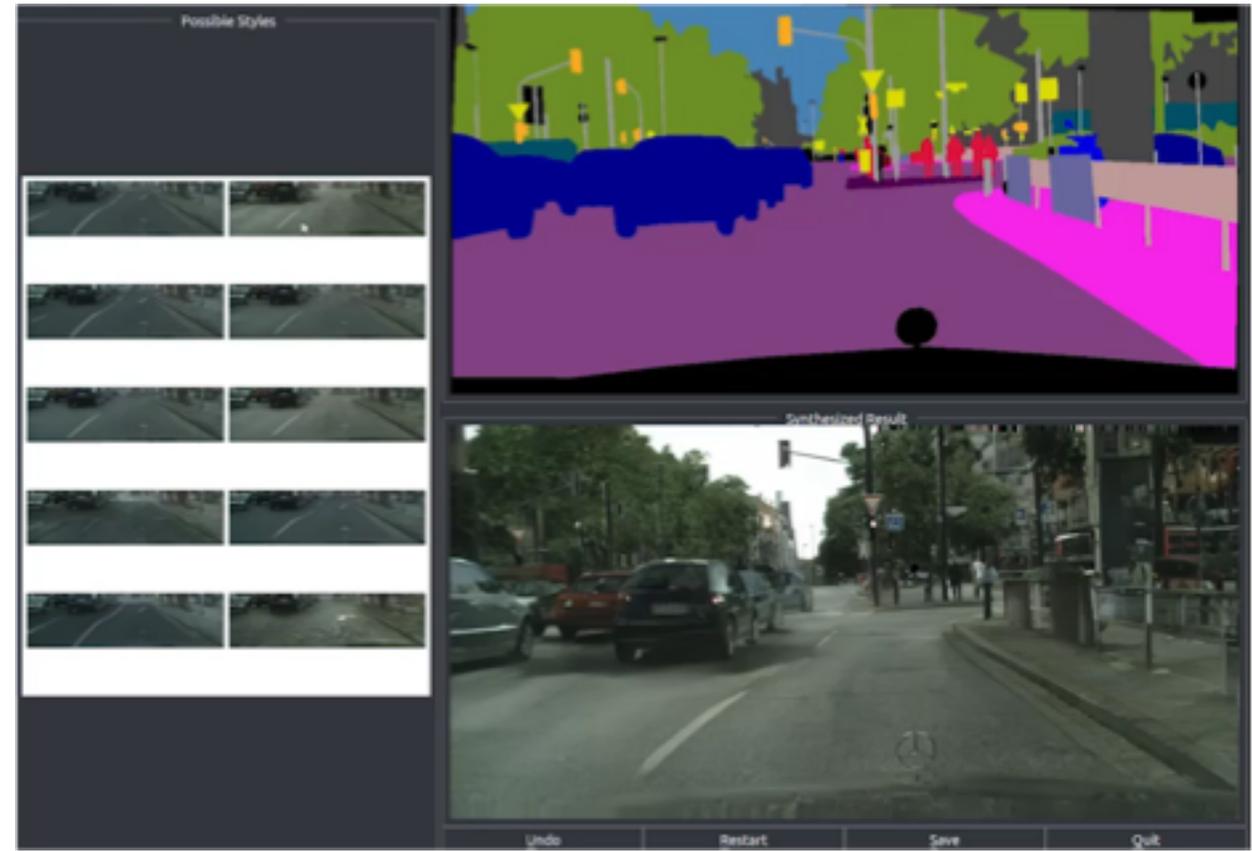


Related Work

High Resolution Image



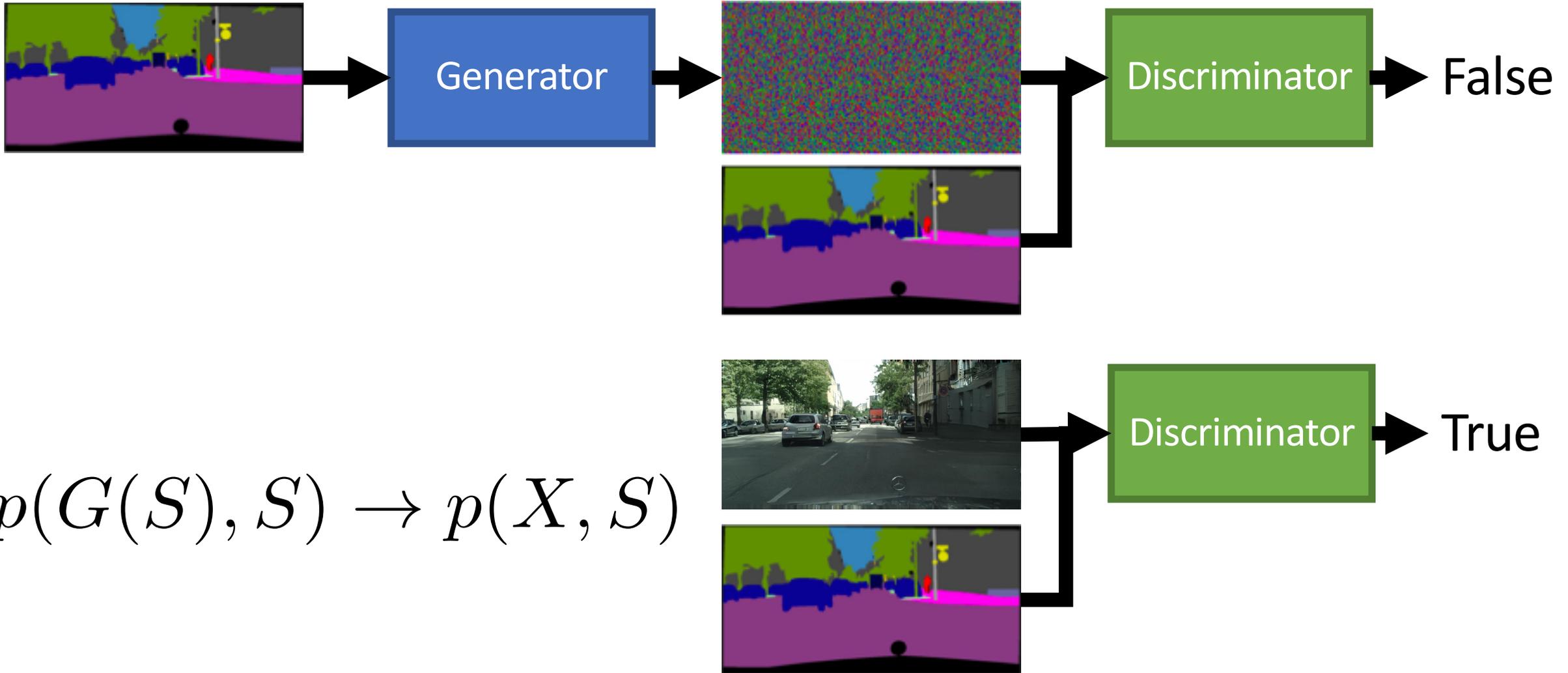
Semantic Manipulation



Outline

- Method

Joint Distribution Learning



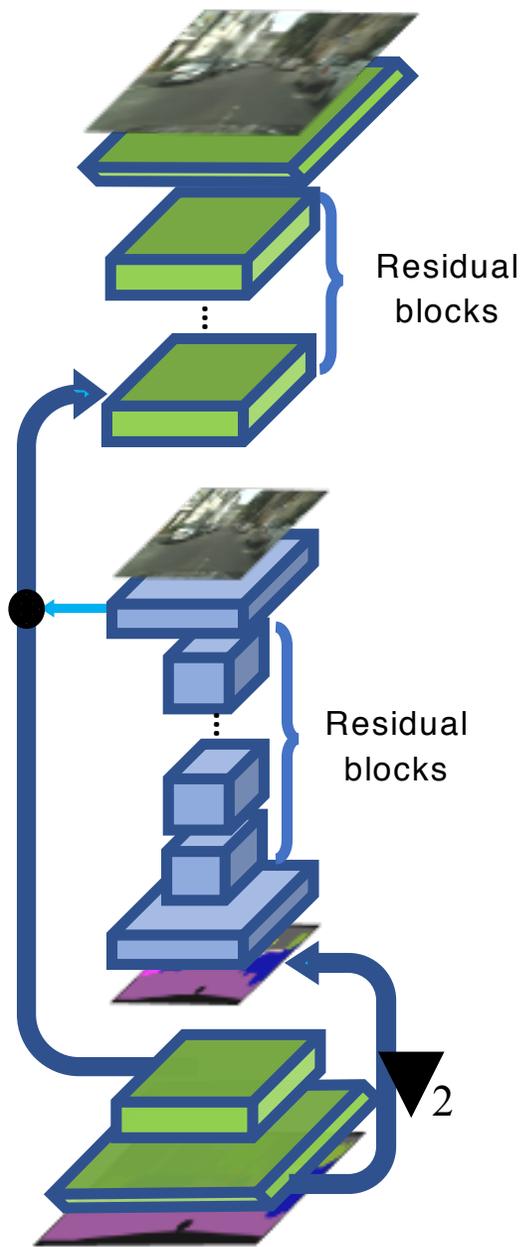
Our Method

- Extending to high resolution
- Using instance-level segmentation maps

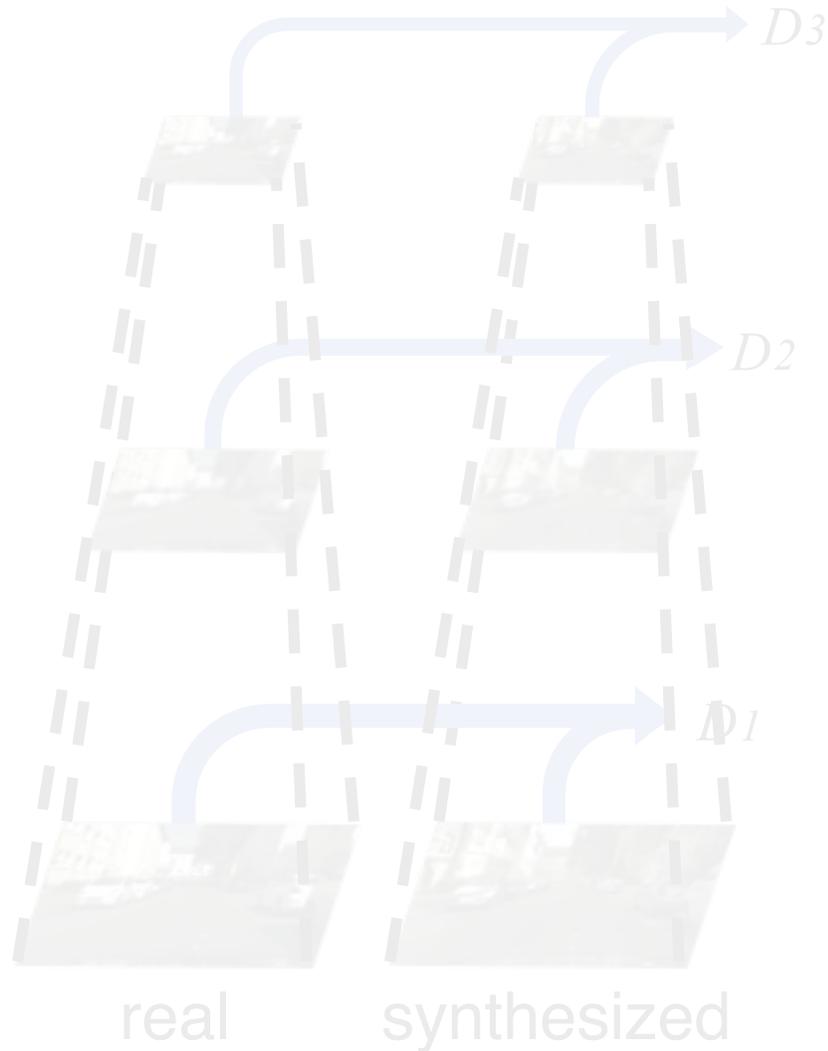
Our Method

- Extending to high resolution
 - New generator
 - New discriminator
 - New objective function

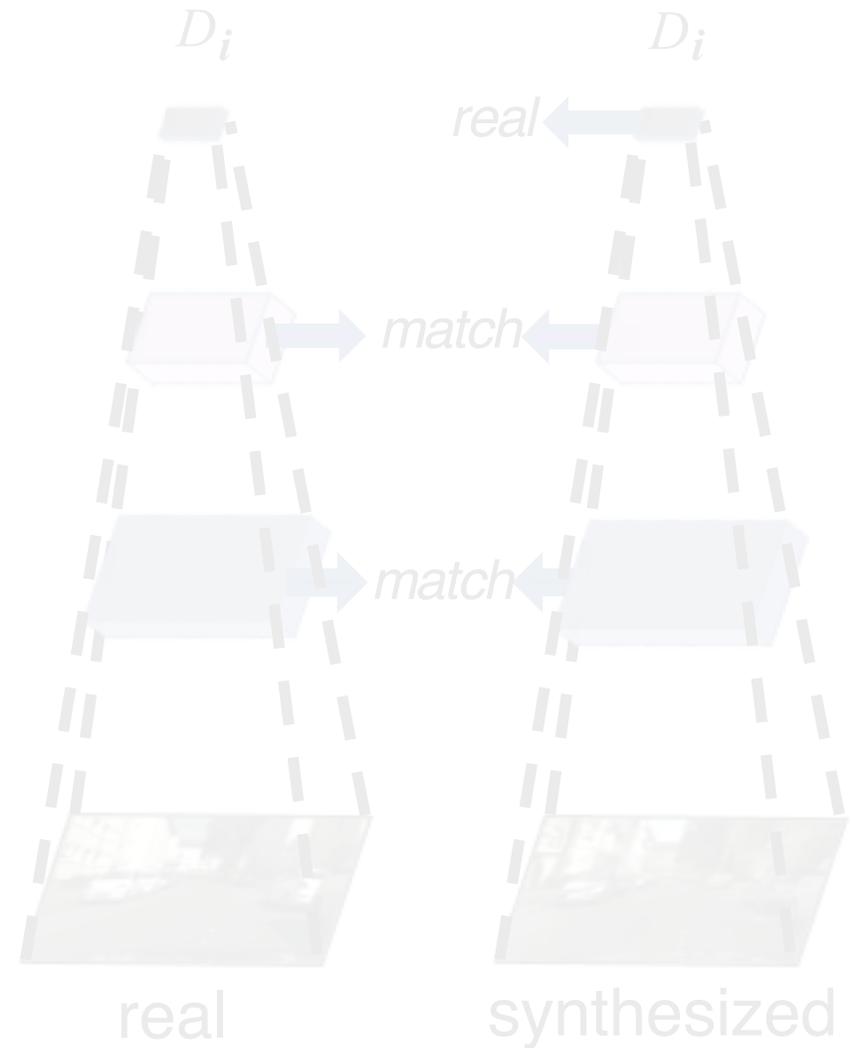
Coarse-to-fine Generator



Multi-scale Discriminators



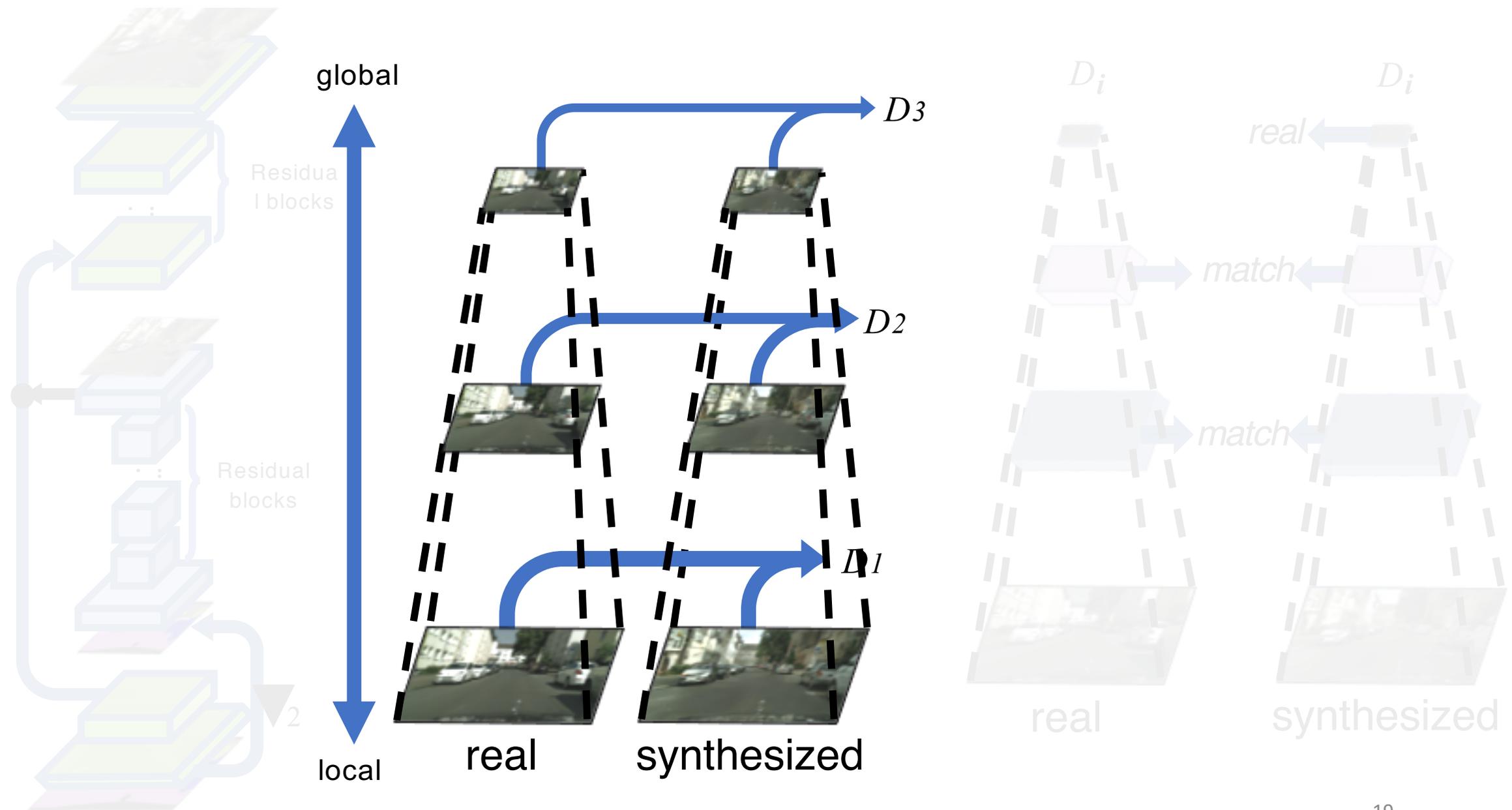
Robust Objective



Coarse-to-fine Generator

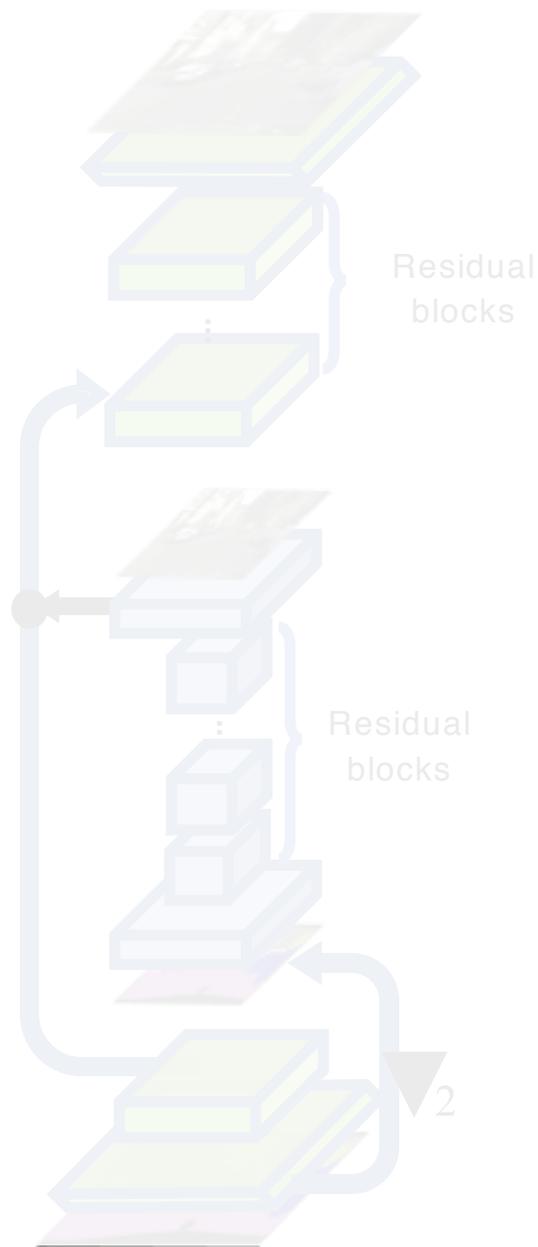
Multi-scale Discriminators

Robust Objective

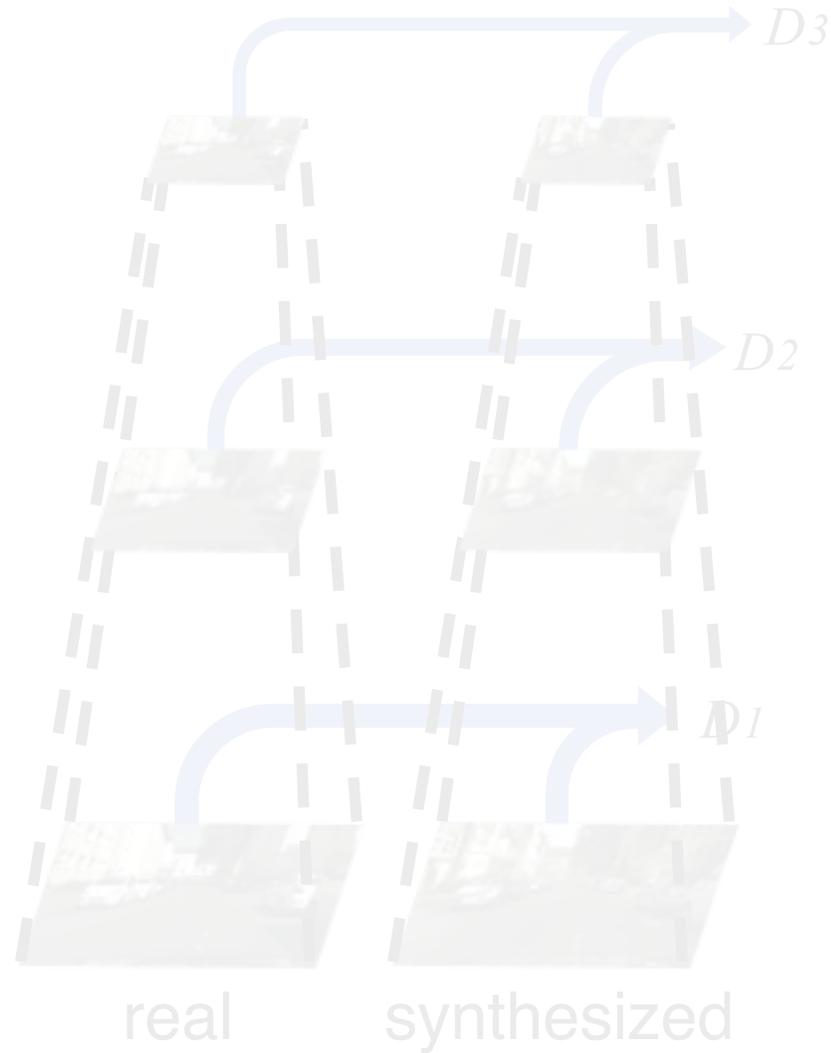


*Similar ideas in Durugkar et al. 2016, Iizuka et al. 2017, Zhang et al. 2017

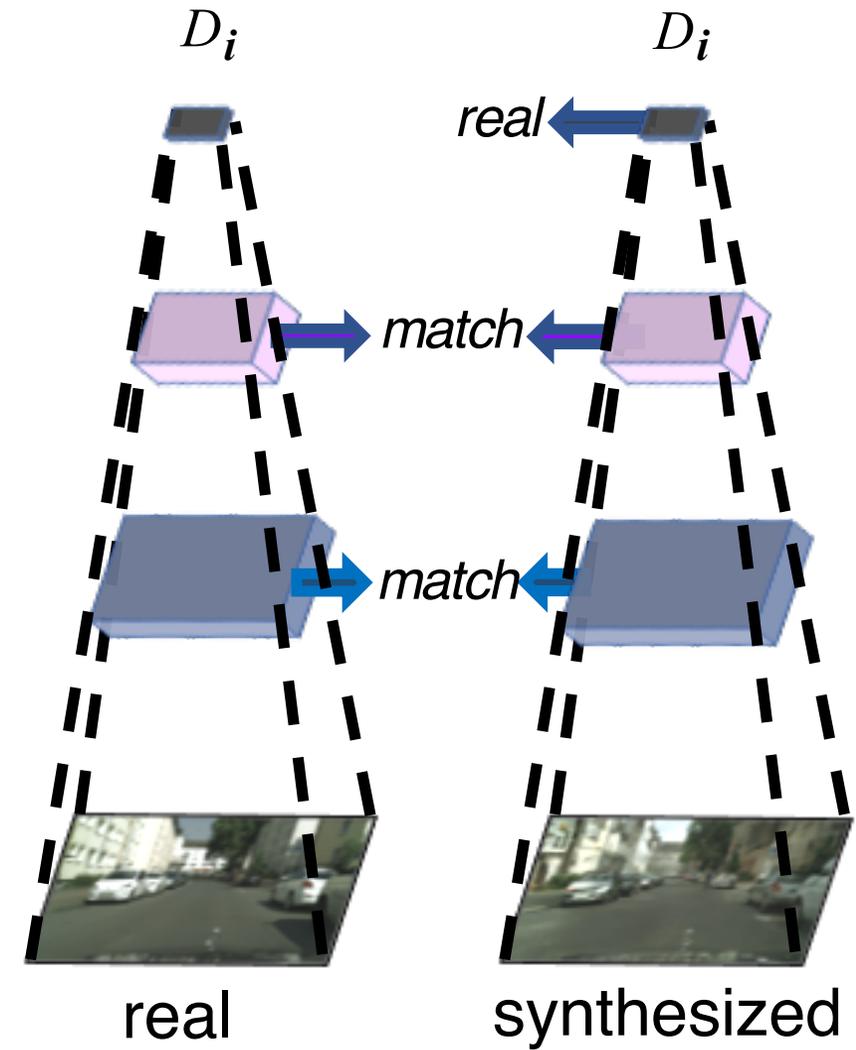
Coarse-to-fine Generator



Multi-scale Discriminators



Robust Objective



Our Method

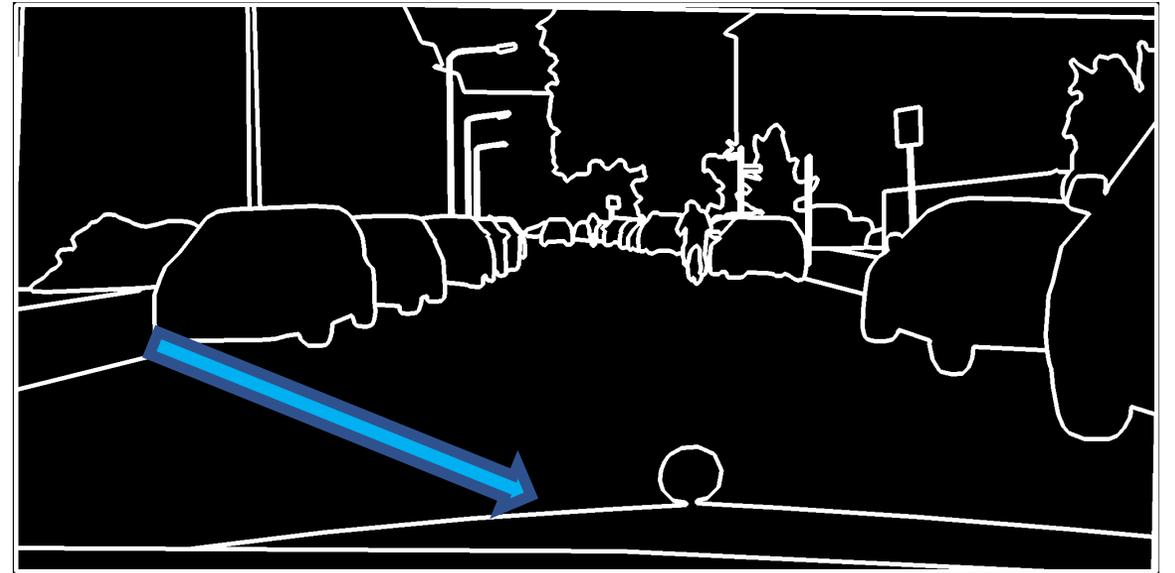
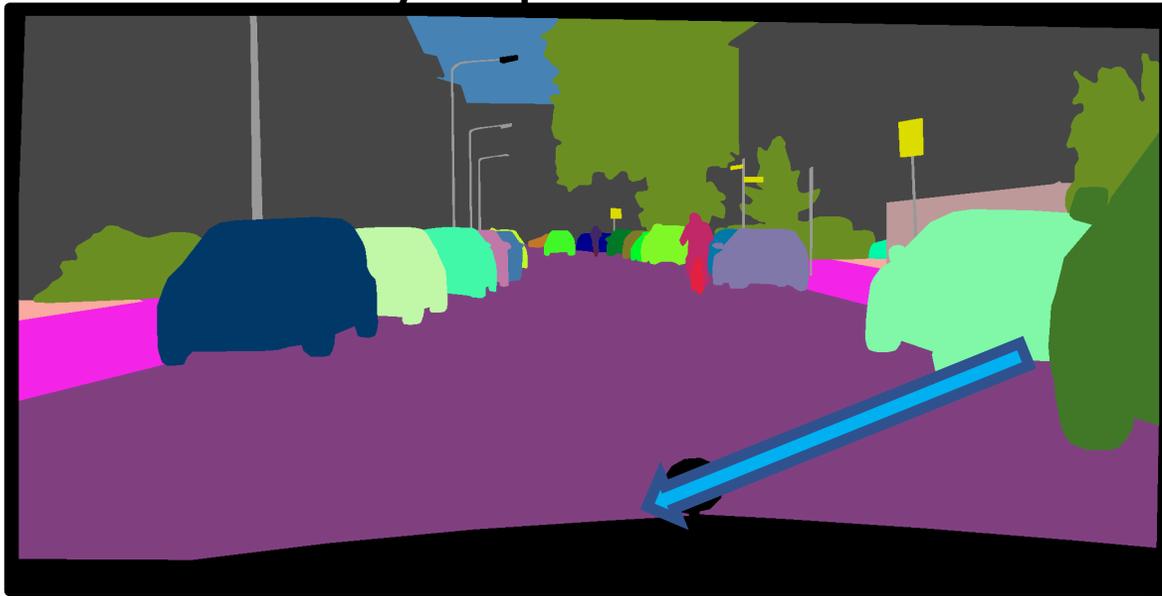
- Using instance-level segmentation maps
 - Boundary improvement
 - Multi-modal results using feature embedding

Our Method

- Boundary improvement

Our Method

- Boundary improvement

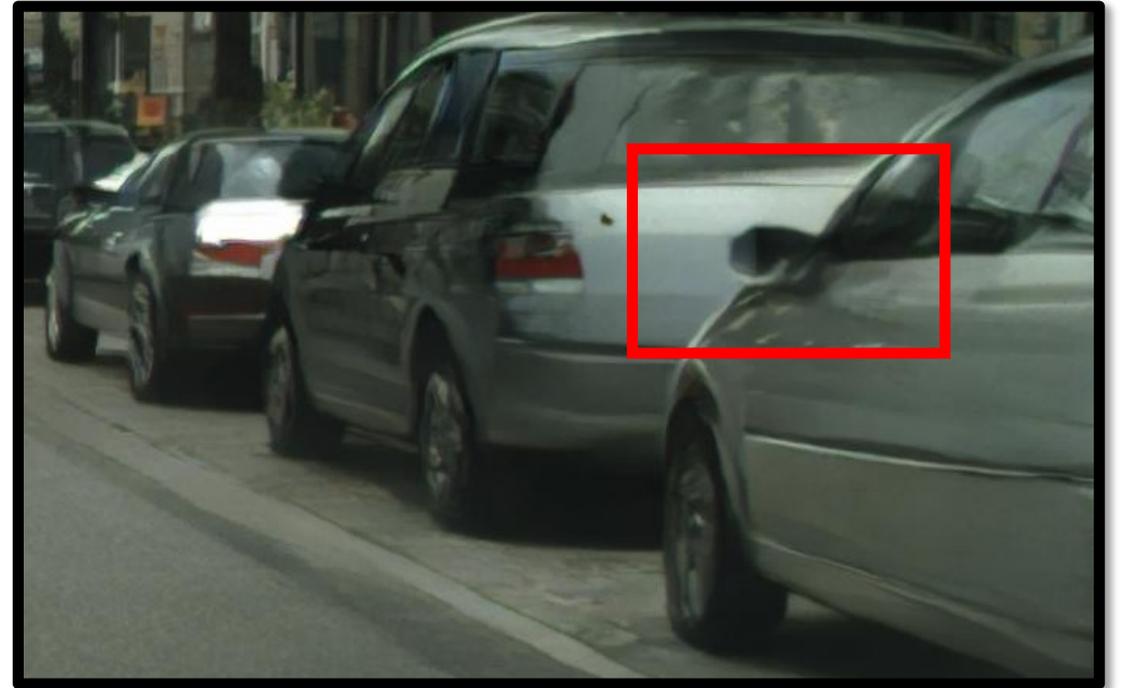


Our Method

- Boundary improvement



without instance maps



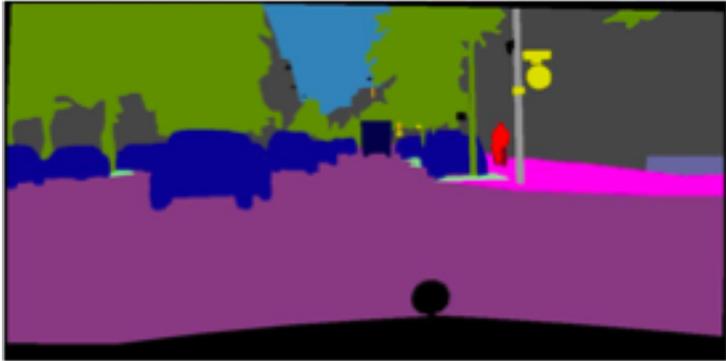
with instance maps

Our Method

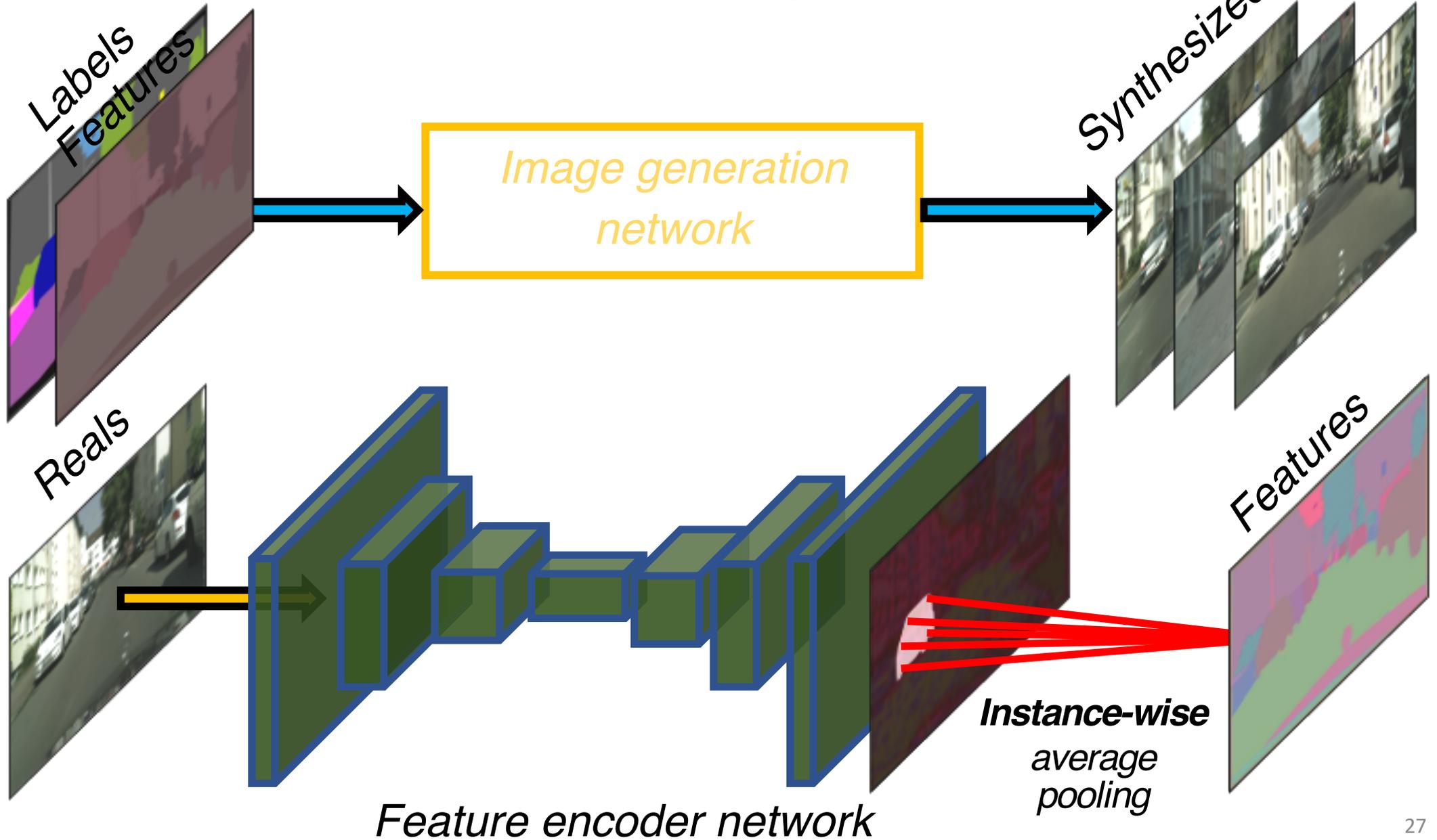
- Multi-modal results using **feature embedding**

Our Method

- Multi-modal (one-to-many) results



Feature Embedding Scheme



Outline

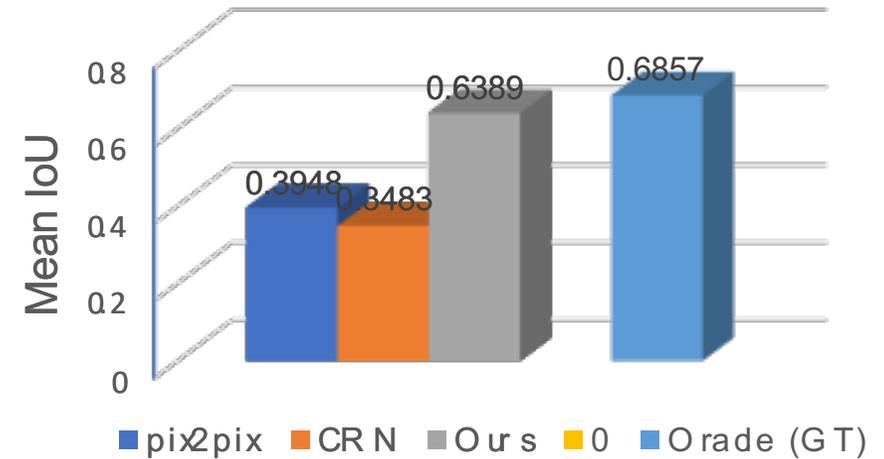
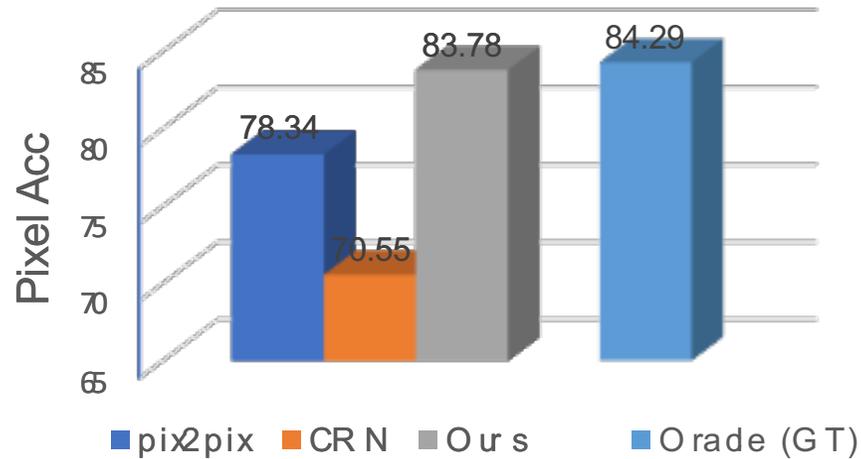
- Results

Results

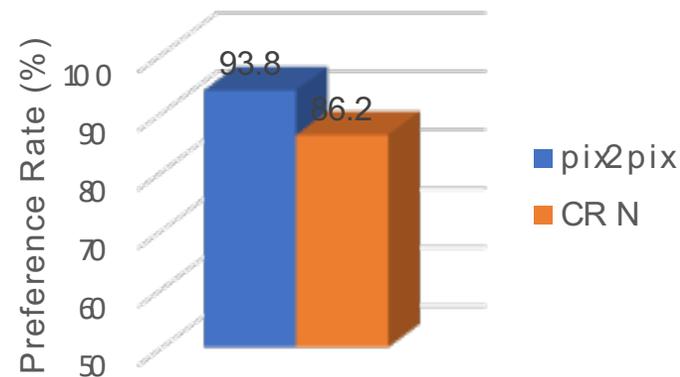
- Comparisons with
 - pix2pix [Isola et al. 2017]
 - CRN [Chen and Koltun 2017]
- Datasets
 - Cityscapes [Cordts et al. 2016]
 - NYU [Silberman et al. 2012]
 - ADE20K [Zhou et al. 2017]
 - Helen Face [Smith et al. 2013]
 - CelebA-HQ [Karras et al. 2017]

Results

- Quantitative comparisons (Cityscapes)
 - Semantic segmentation scores



- Subjective scores



Results

- Qualitative comparisons



Semantic Map



pix2pix



CRN



Ours

Results on NYU dataset



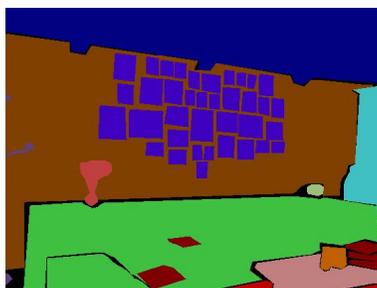
pix2pix



CRN



Ours



Results on NYU dataset



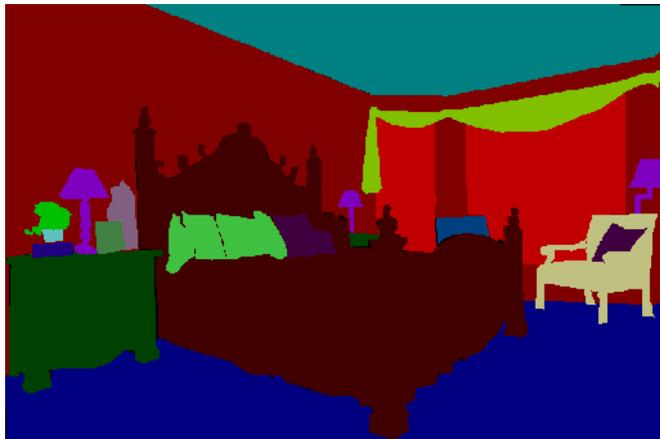
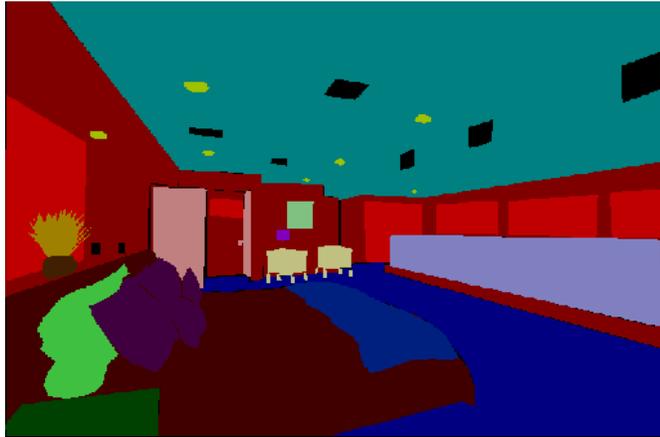
pix2pix

CRN

Ours



Results on ADE20K dataset



Labels

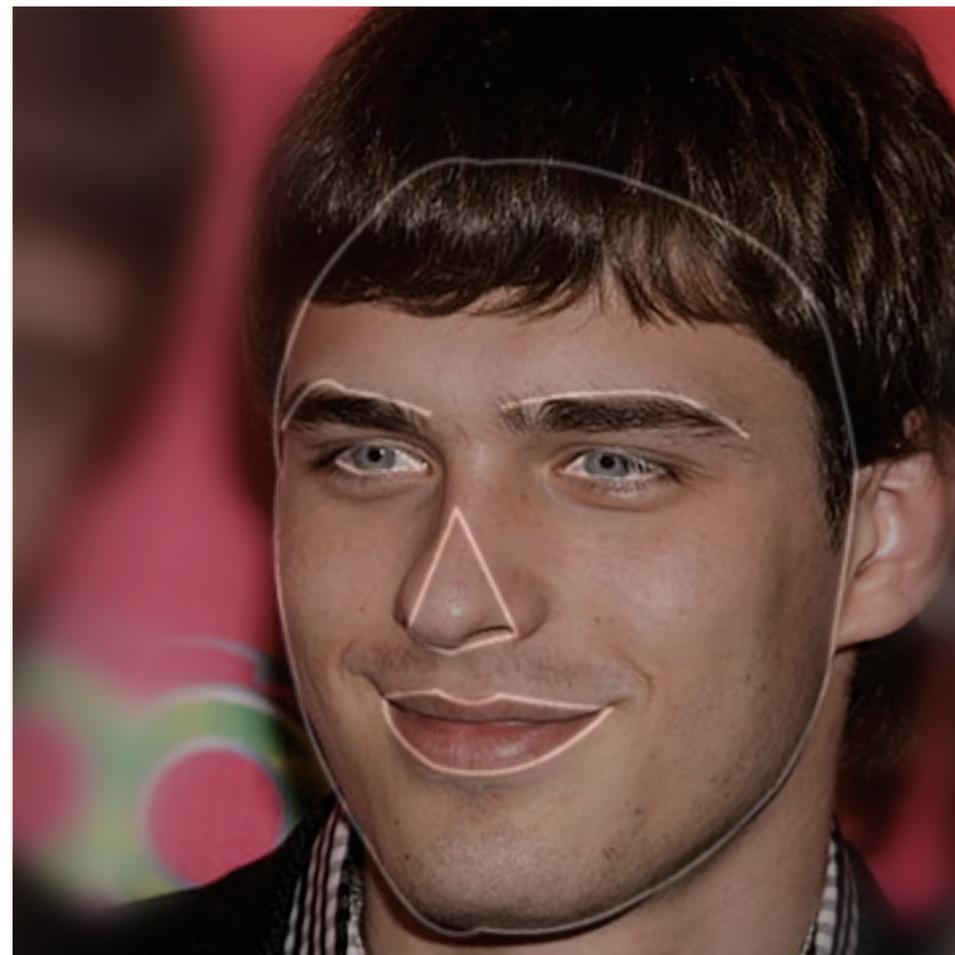
Ours

Ground truth

Results on CelebA-HQ

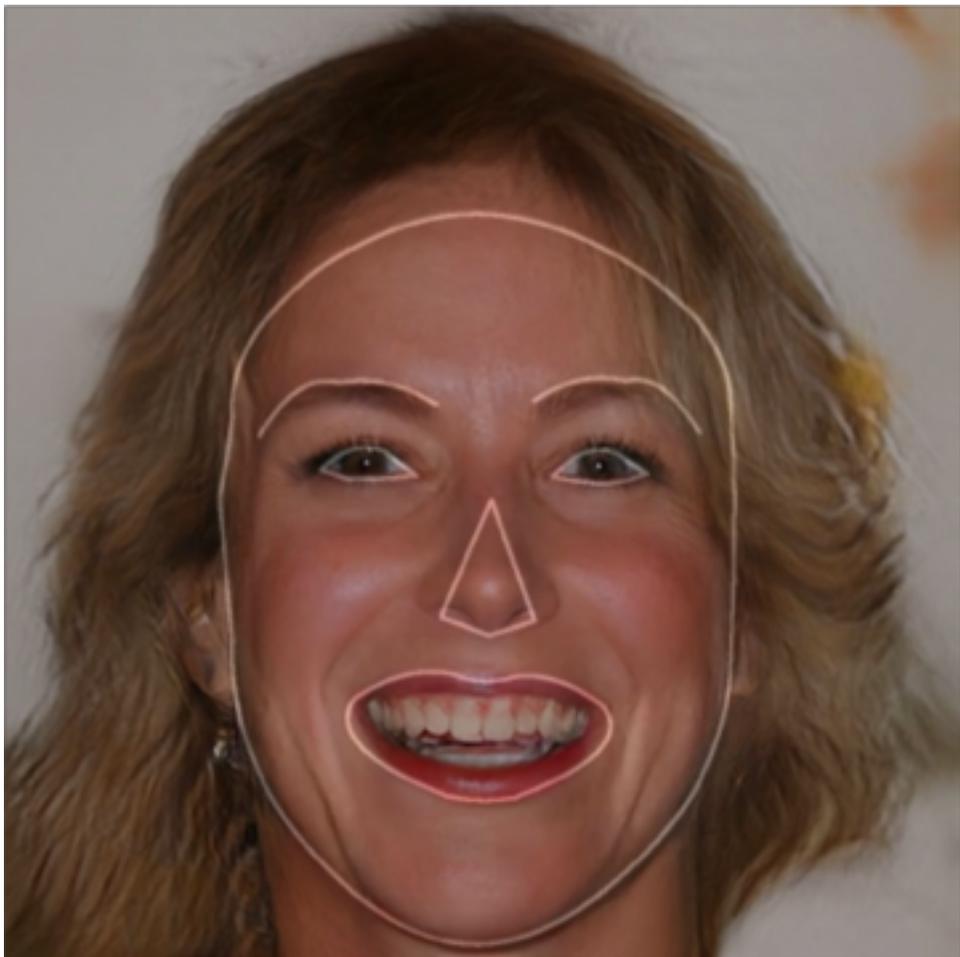


Synthesized



Ground truth

Results on CelebA-HQ



Synthesized

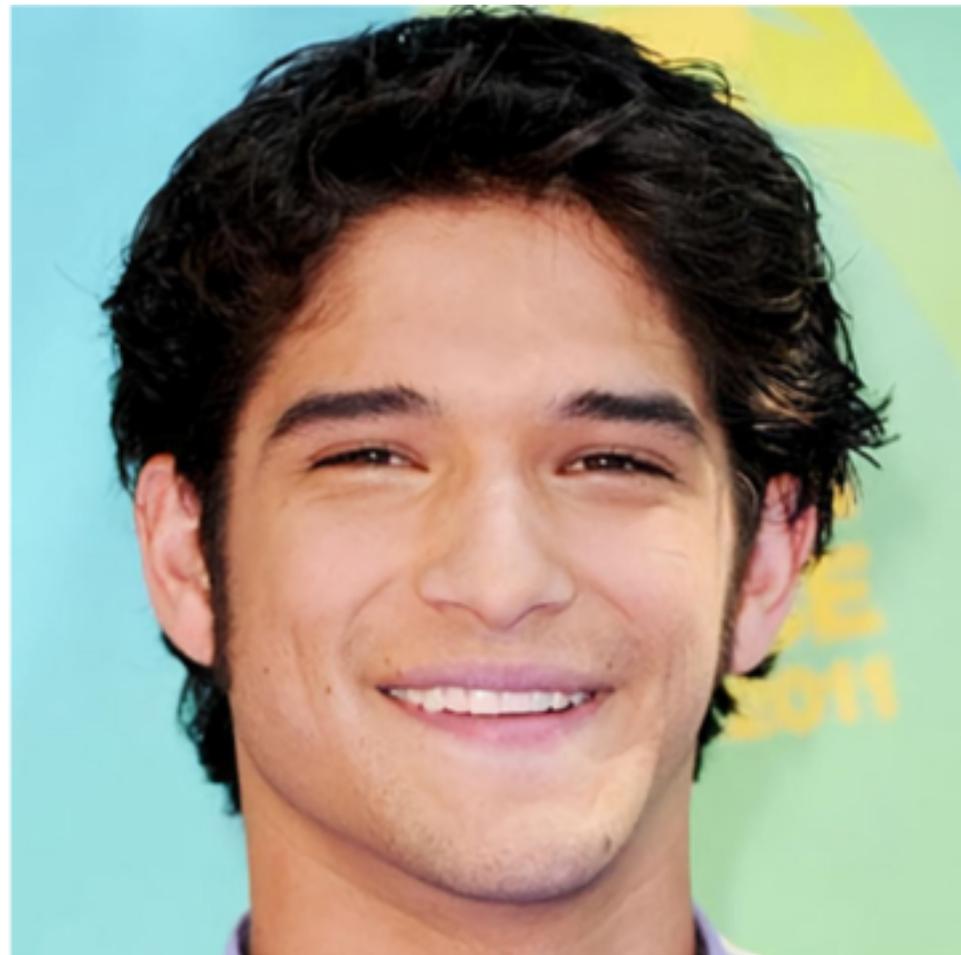


Ground truth

Results on CelebA-HQ



Synthesized



Ground truth

Results on CelebA-HQ



Synthesized

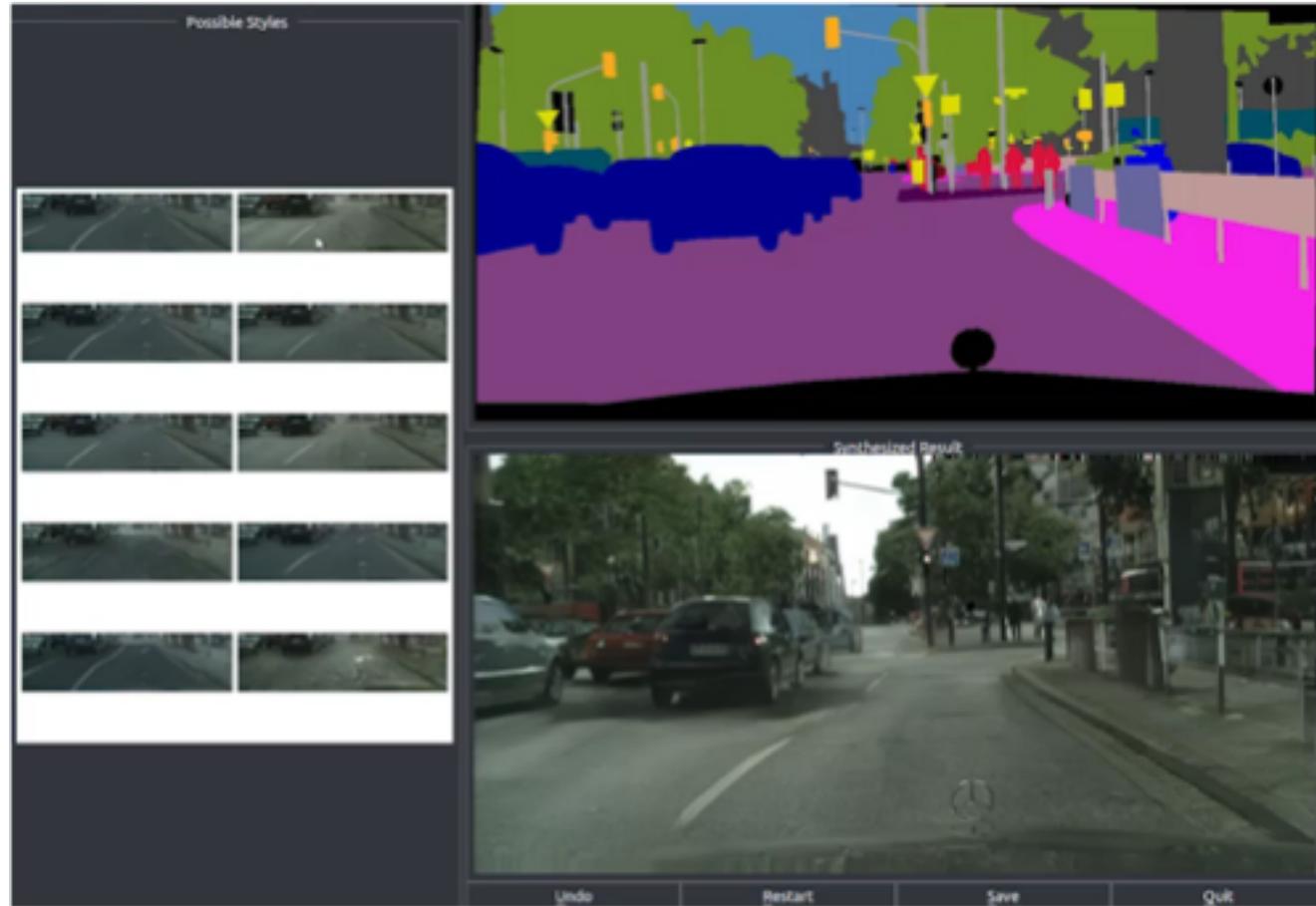


Ground truth

Outline

- Applications

Applications: style changing



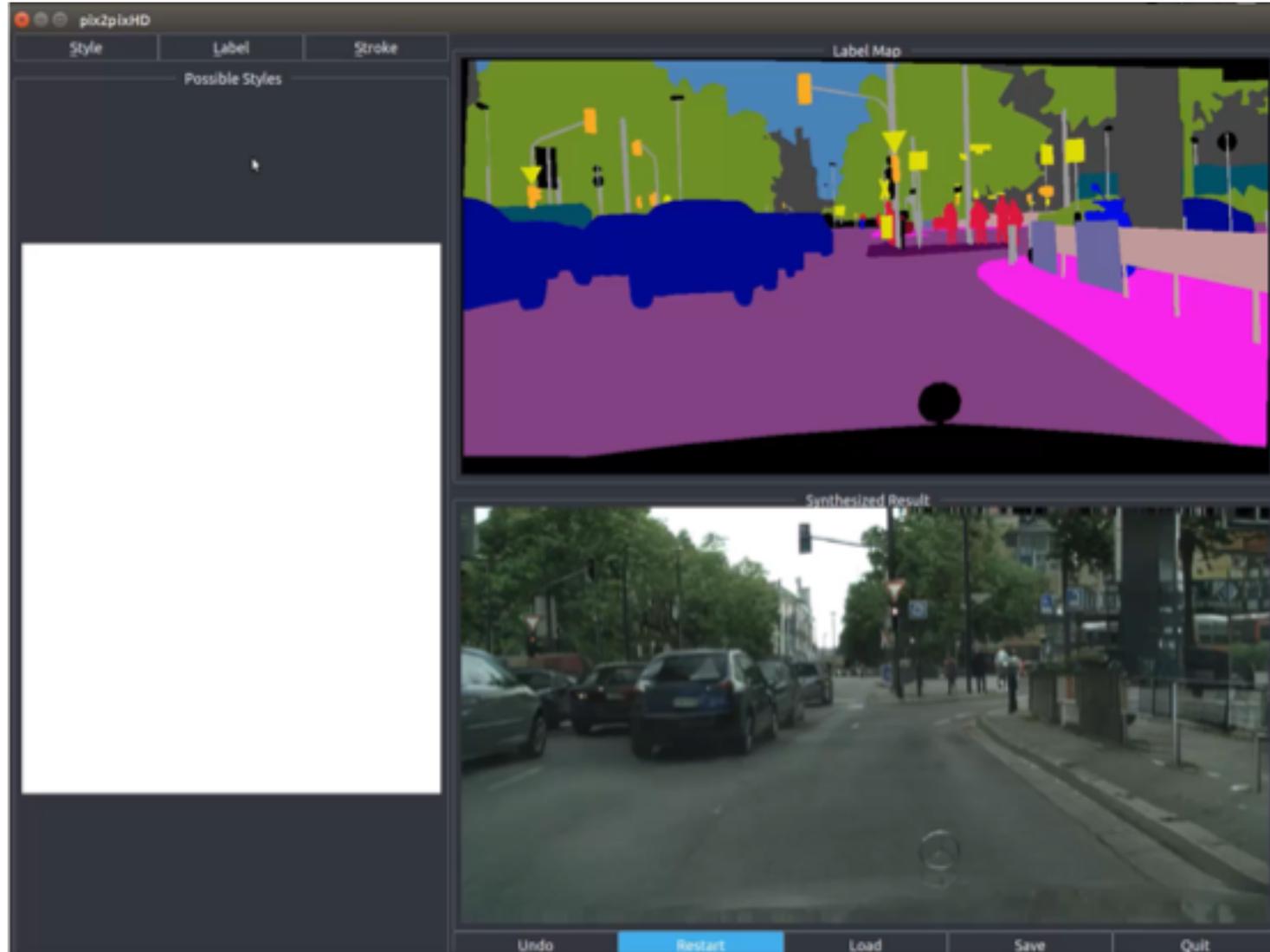
Applications: style changing



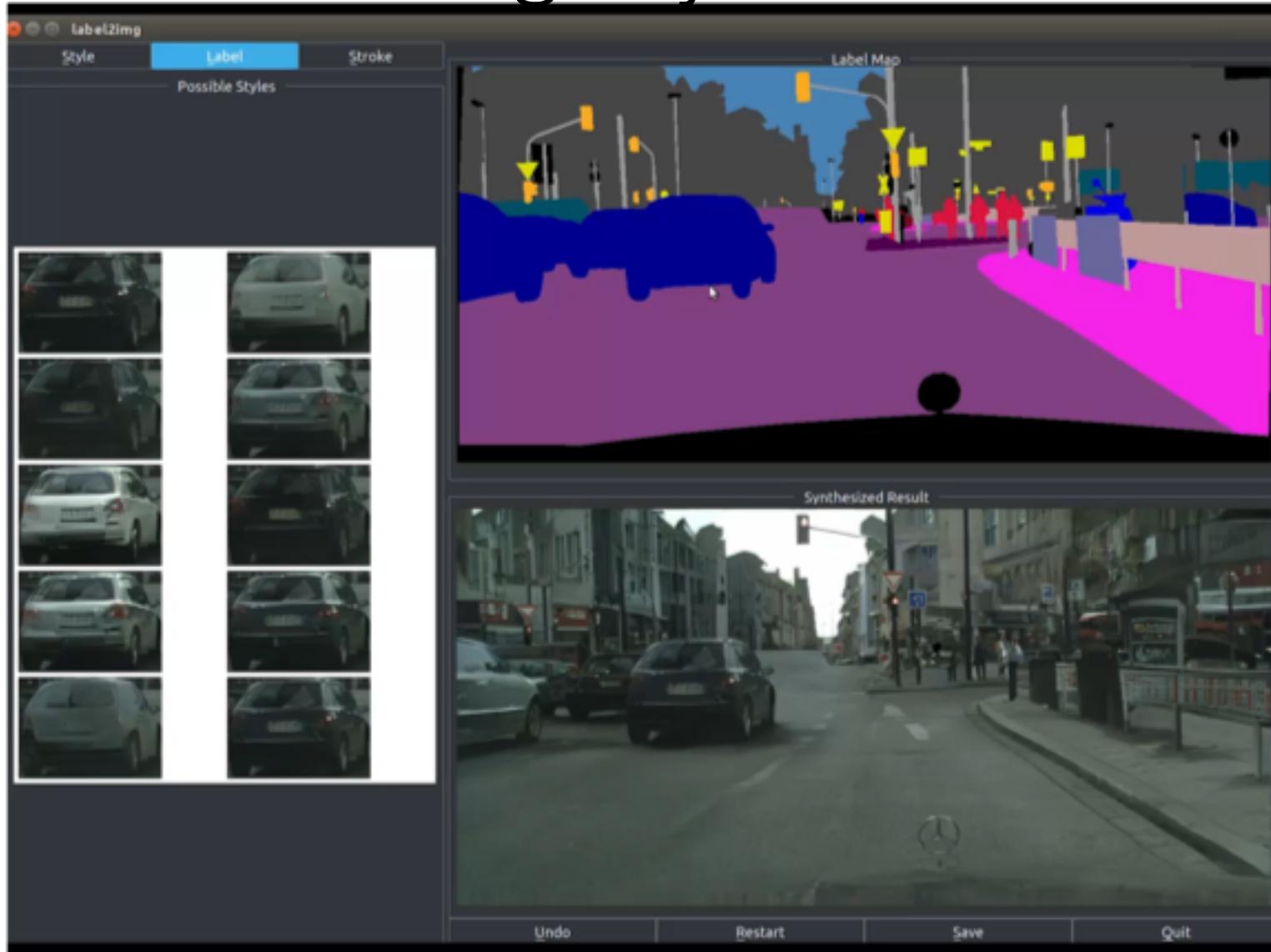
Applications: style changing



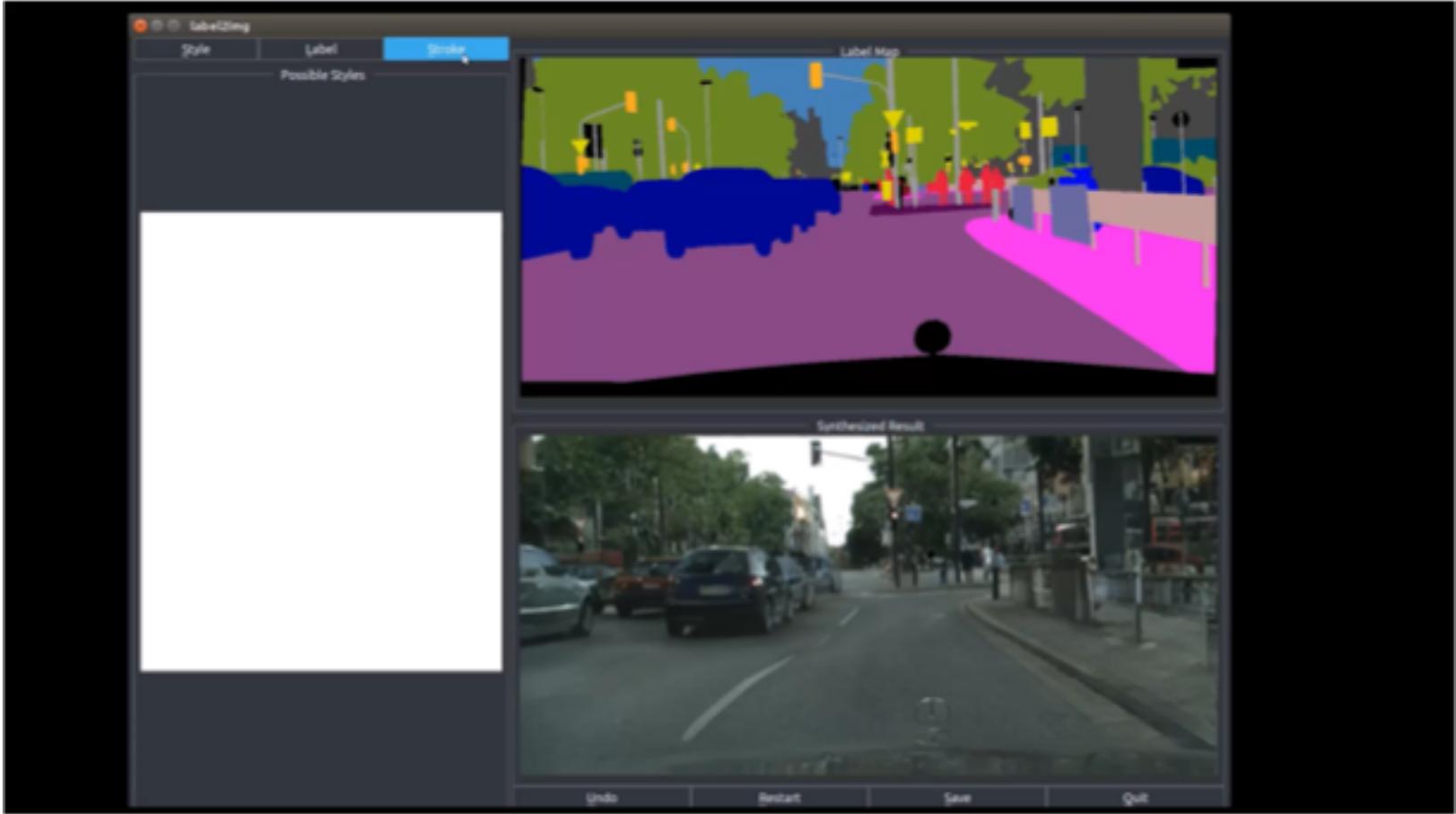
Applications: label changing



Applications: adding objects



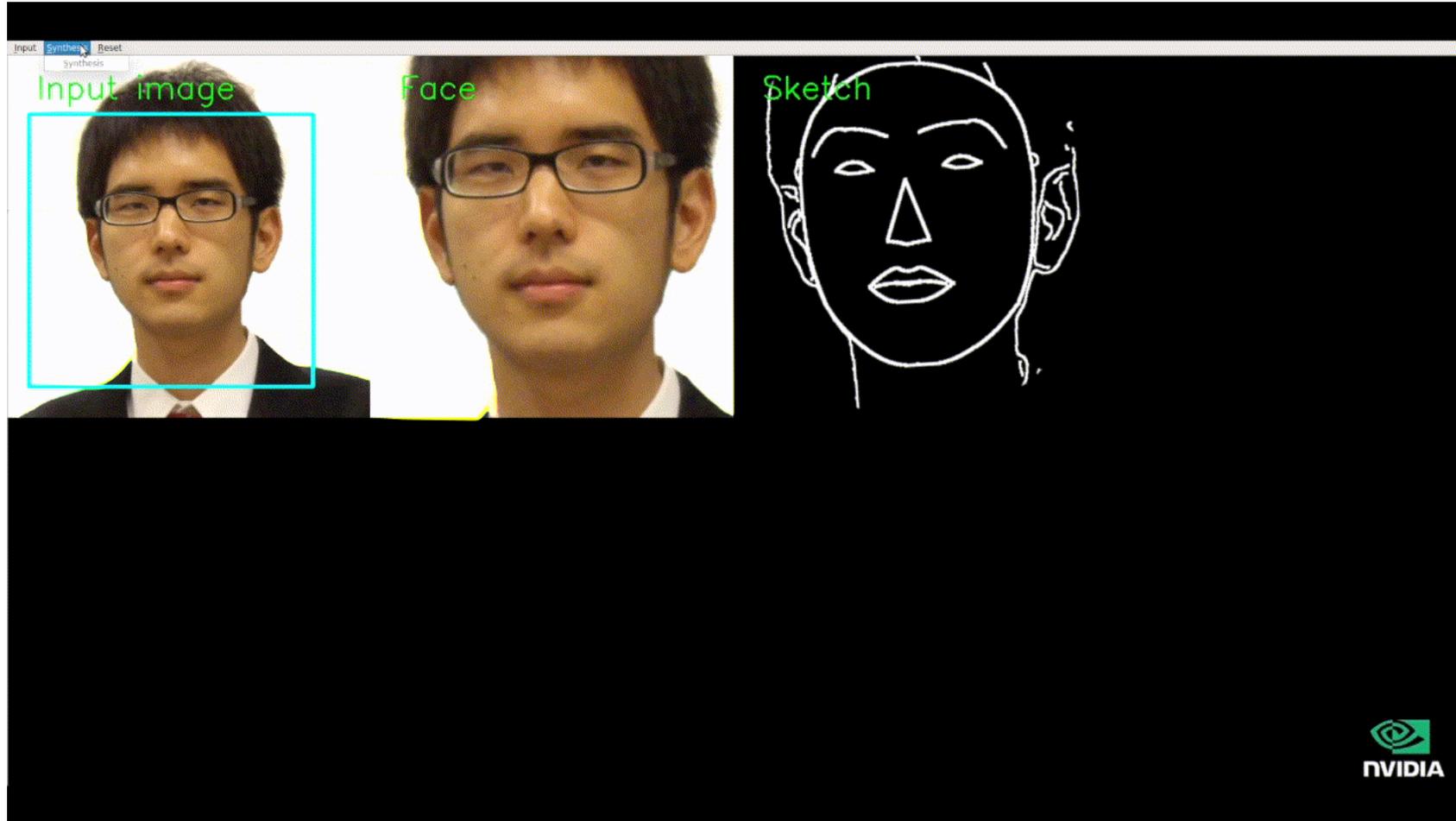
Applications: adding strokes



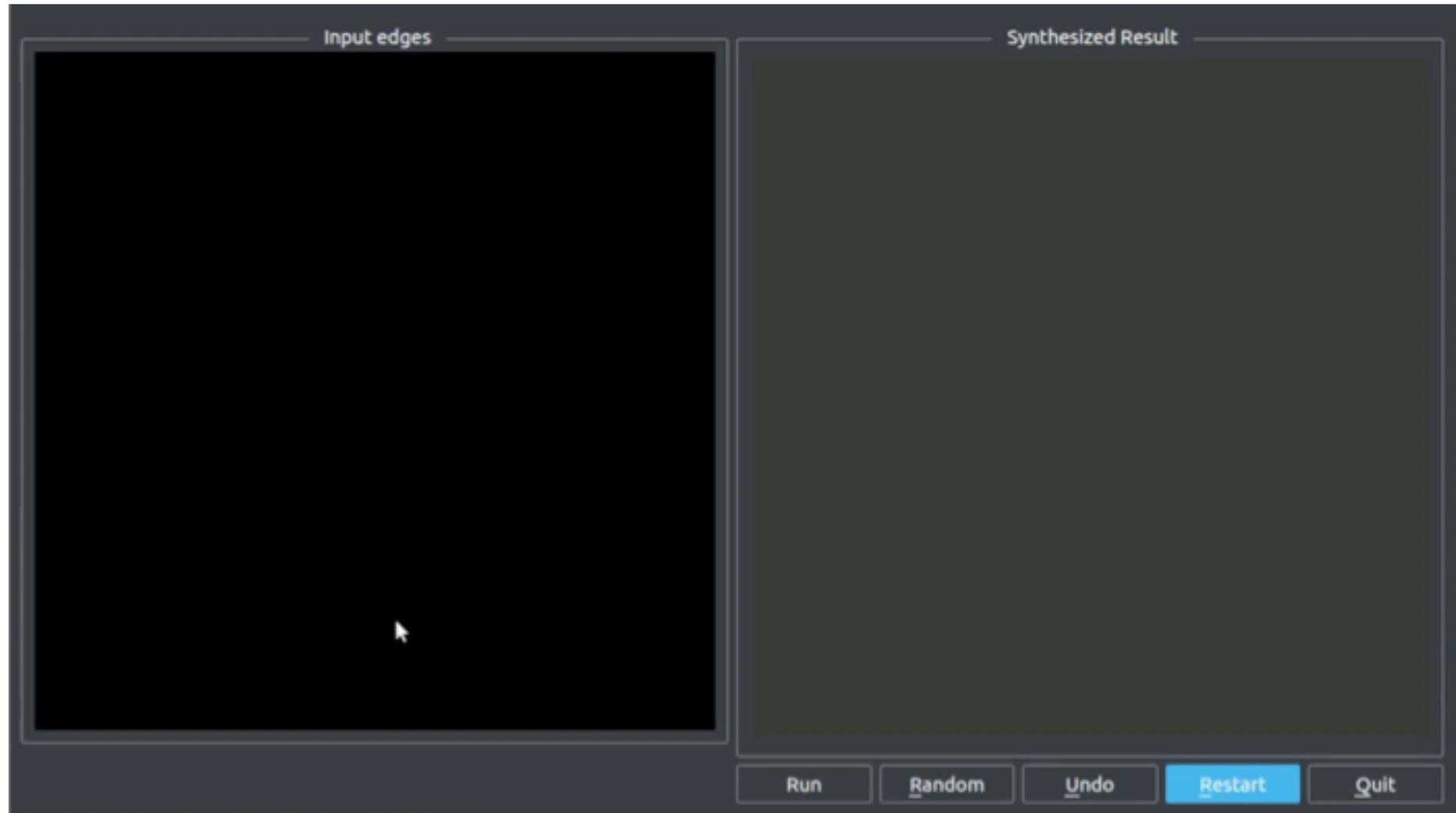
Applications: adding strokes



Live Demo in the NVIDIA Booth



Live Demo in the NVIDIA Booth



Extension: vid2vidHD



Extensions: Videos



$$\mathbf{x}_t = (1 - \mathbf{m}_t) \odot \mathbf{w}_{t-1} + \mathbf{m}_t \odot \mathbf{h}_t$$

Extension: vid2vidHD



Paper and code will be released soon!

Extension: vid2vidHD



Paper and code will be released soon!

Extensions: Videos

- edge2face



Outline

- Conclusion

Conclusion

- We present a GAN based framework that can
 - Synthesize high-res realistic images



Conclusion

- We present a GAN based framework that can
 - Synthesize high-res realistic images
 - Generate multi-modal results



Thank you!

Project: <https://tcwang0509.github.io/pix2pixHD/>
Code: <https://github.com/NVIDIA/pix2pixHD>

