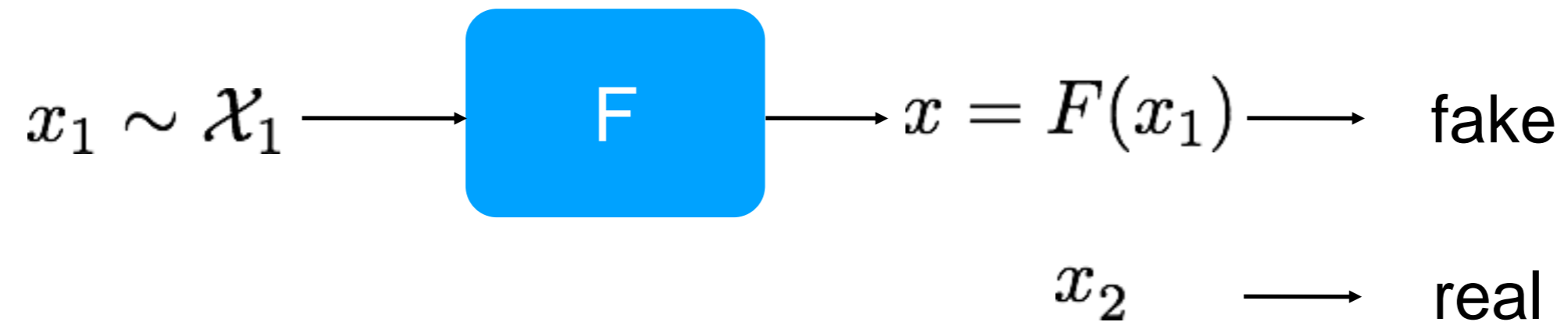


pix2pixHD:  
High-Resolution Image Synthesis and  
Semantic Manipulation with Conditional  
GANs

Ting-Chun Wang  
NVIDIA

# Supervised Image-to-image Translation



- Supervisedly relating  $x = F(x_1^{(i)})$  to  $x_2^{(i)}$
- Isola et al (CVPR'17): Learning a joint distribution  
Discriminator sees both input and output images

$$\max_F E_{p_{\mathcal{X}_1}} [\log(D(x_1, F(x_1)))]$$

# pix2pix

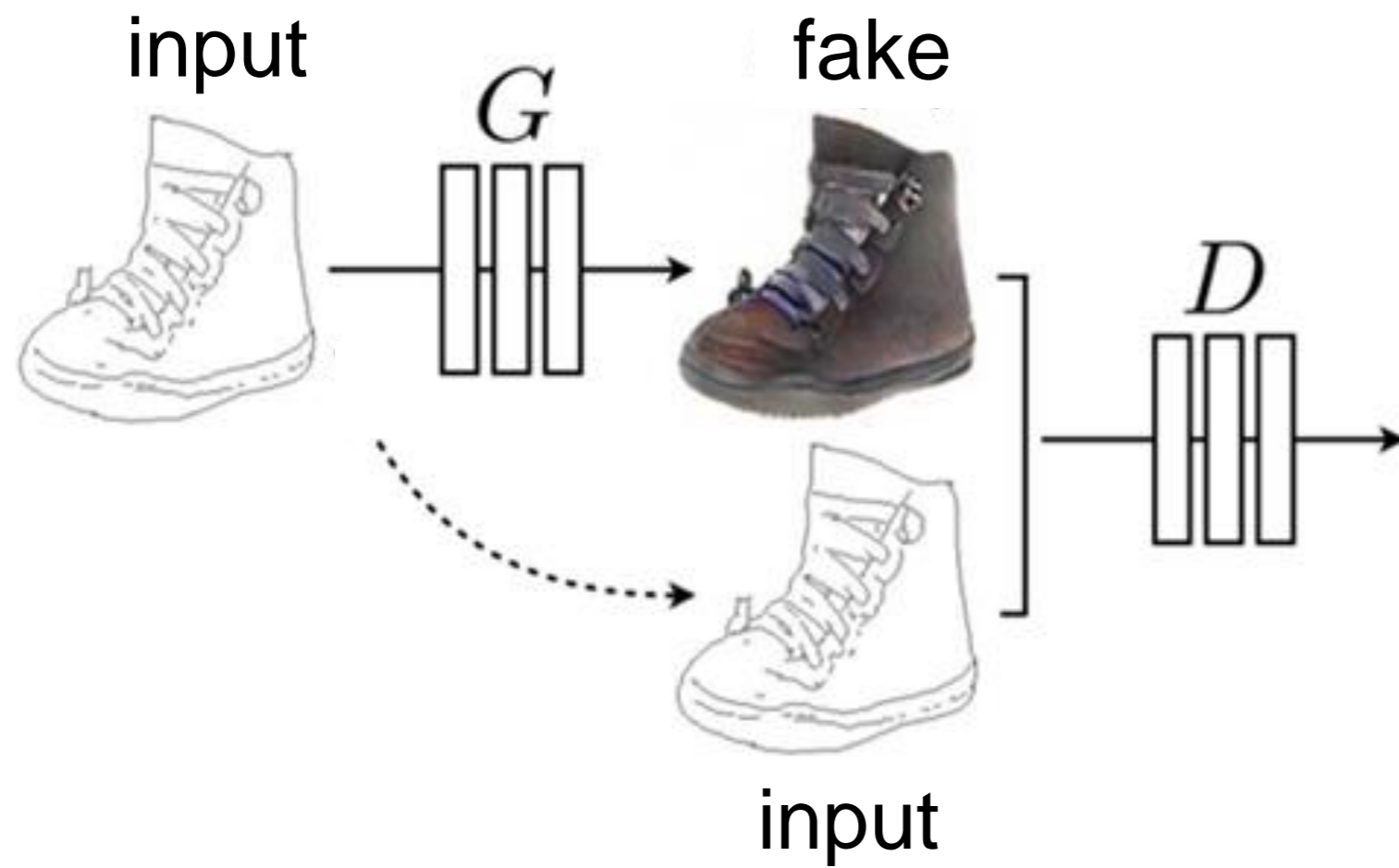
- Discriminator Training

input



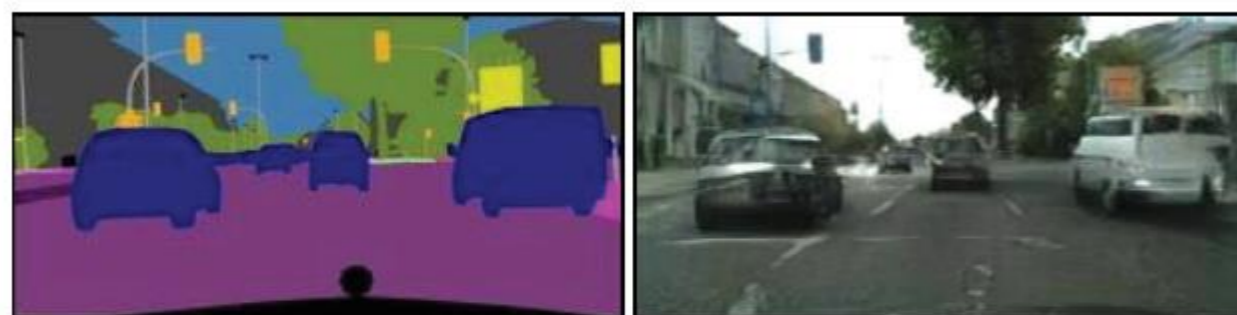
# pix2pix

- Generator Training



# pix2pix

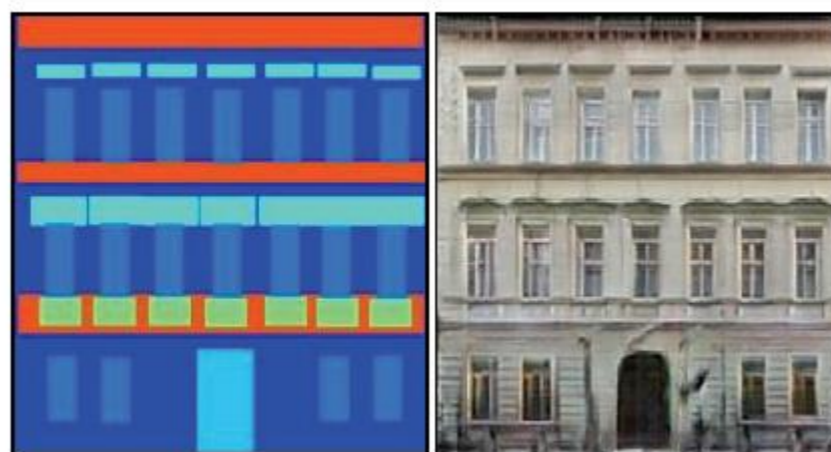
Labels to Street Scene



input

output

Labels to Facade



input

output

BW to Color



input

output

Aerial to Map



input

output

Day to Night



input

output

Edges to Photo



input

output

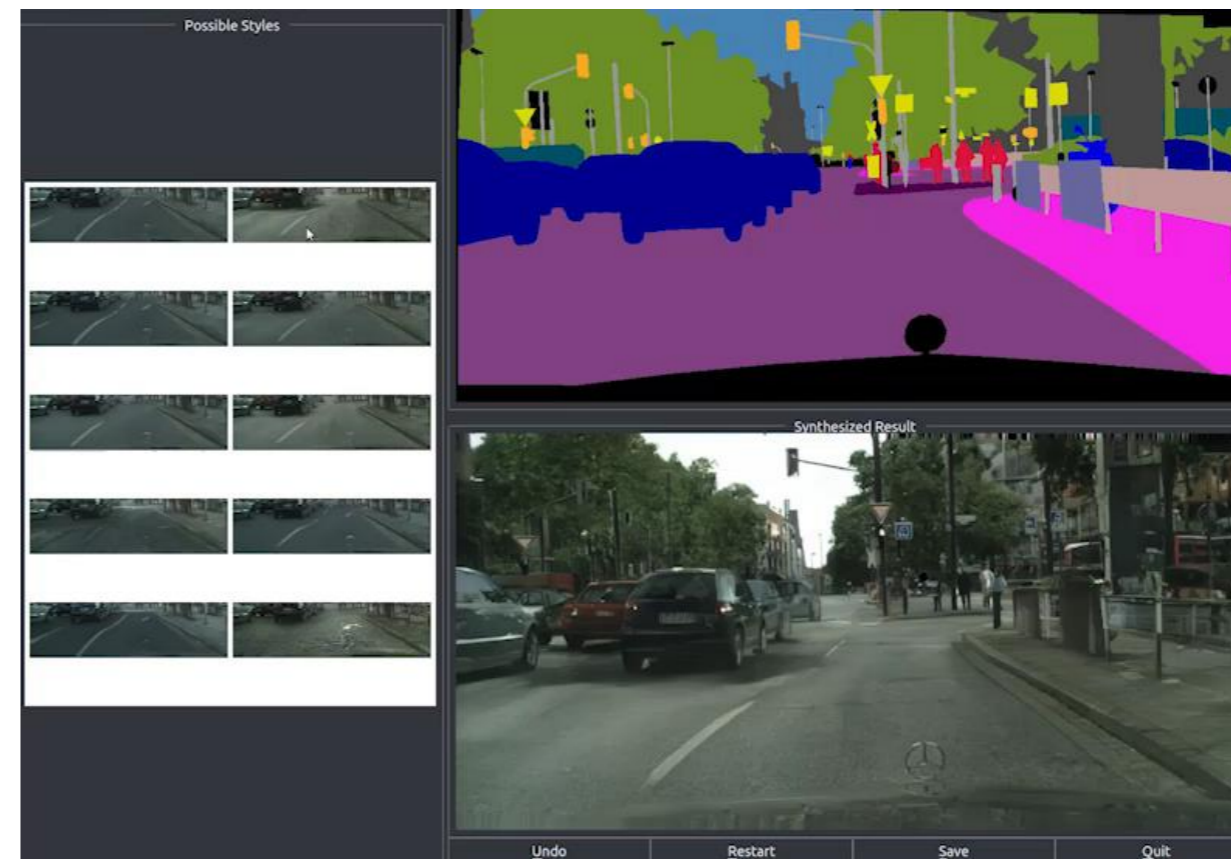
# pix2pixHD

T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro, "High-resolution Image Synthesis and Semantic Manipulation with Conditional GANs," CVPR 2018  
<https://github.com/NVIDIA/pix2pixHD>

## High Resolution Image



## Semantic Manipulation



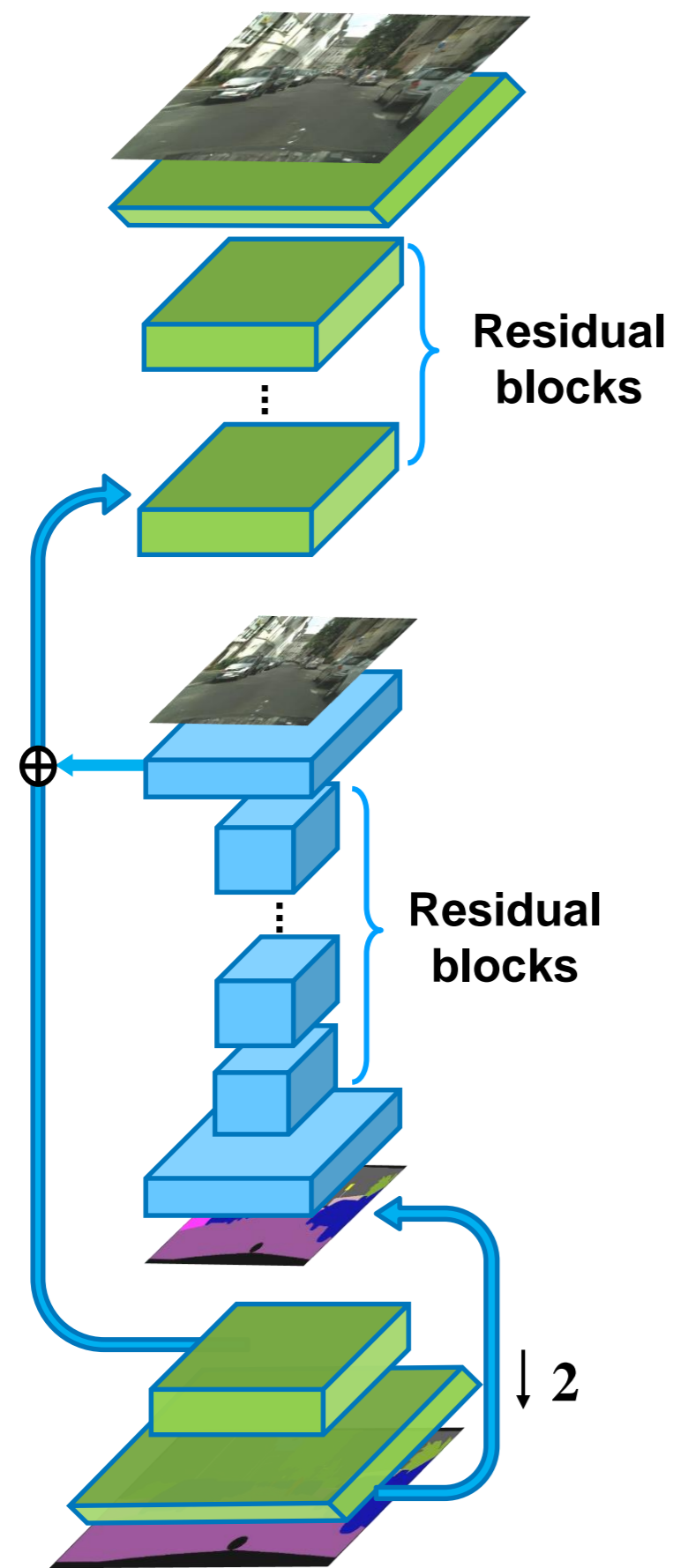
# pix2pixHD

- Extend pix2pix to high resolution (256  $\rightarrow$  2k)
  - Coarse-to-fine generator
  - Multi-scale discriminator
  - Robust objective function

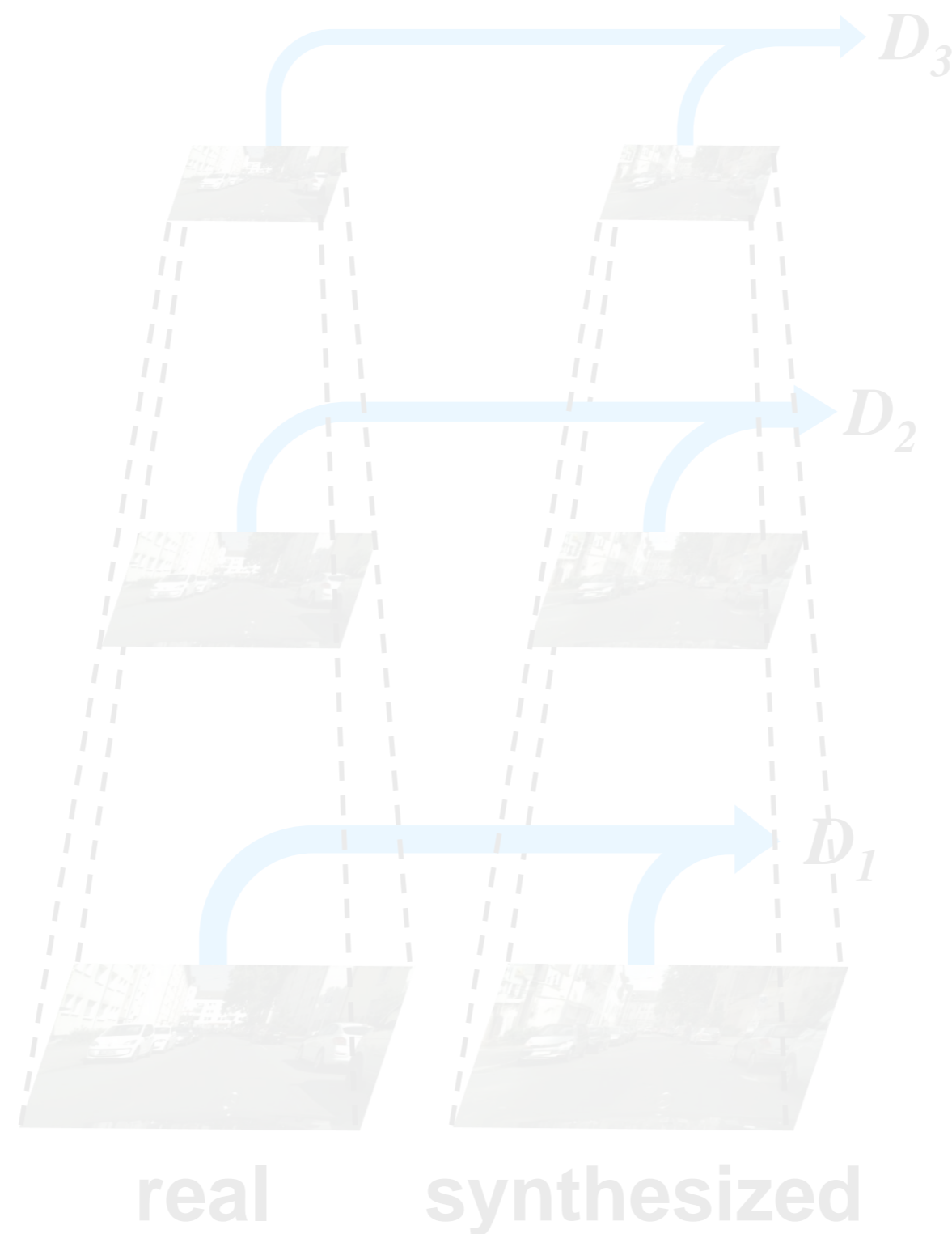




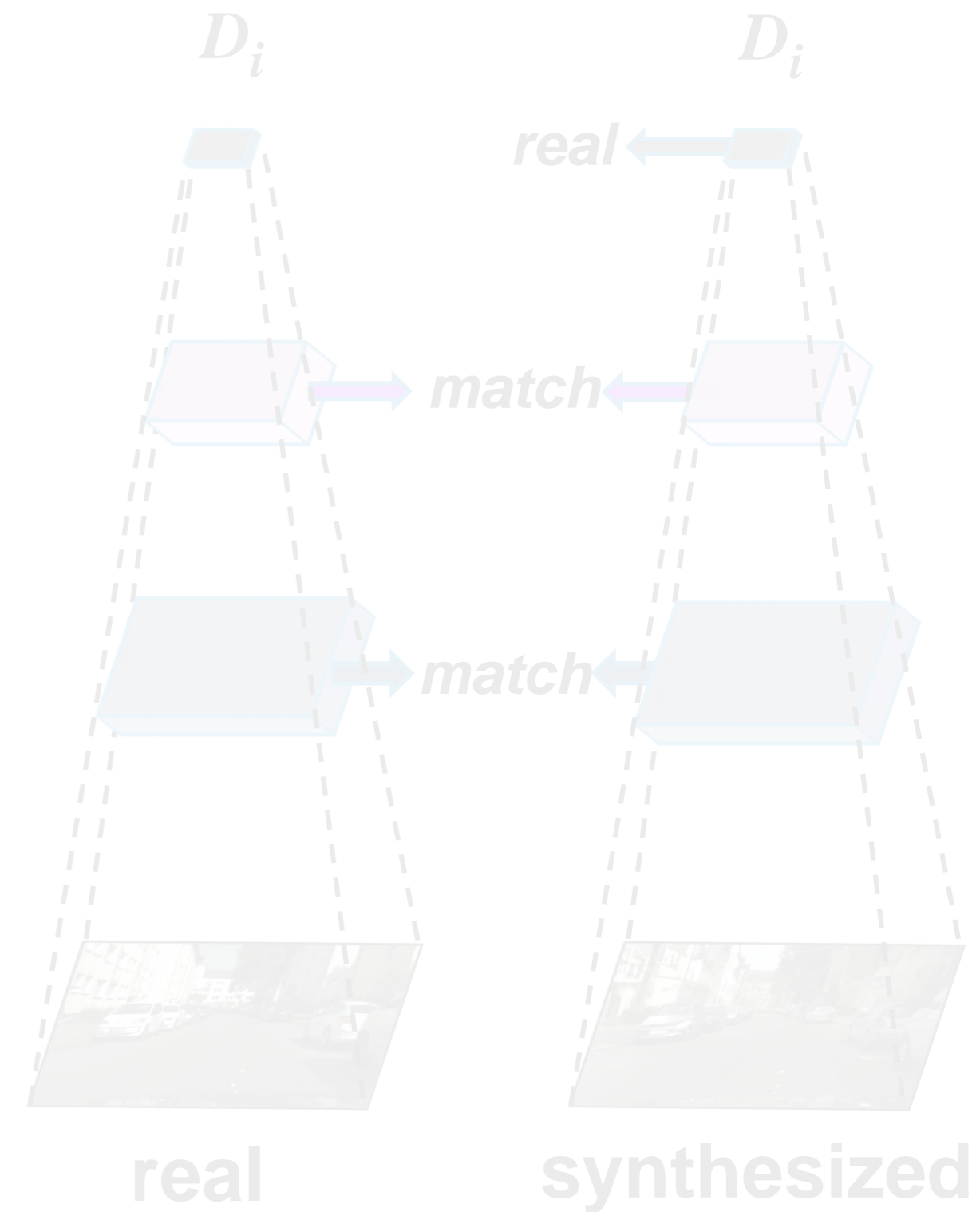
# Coarse-to-fine Generator



# Multi-scale Discriminators



# Robust Objective

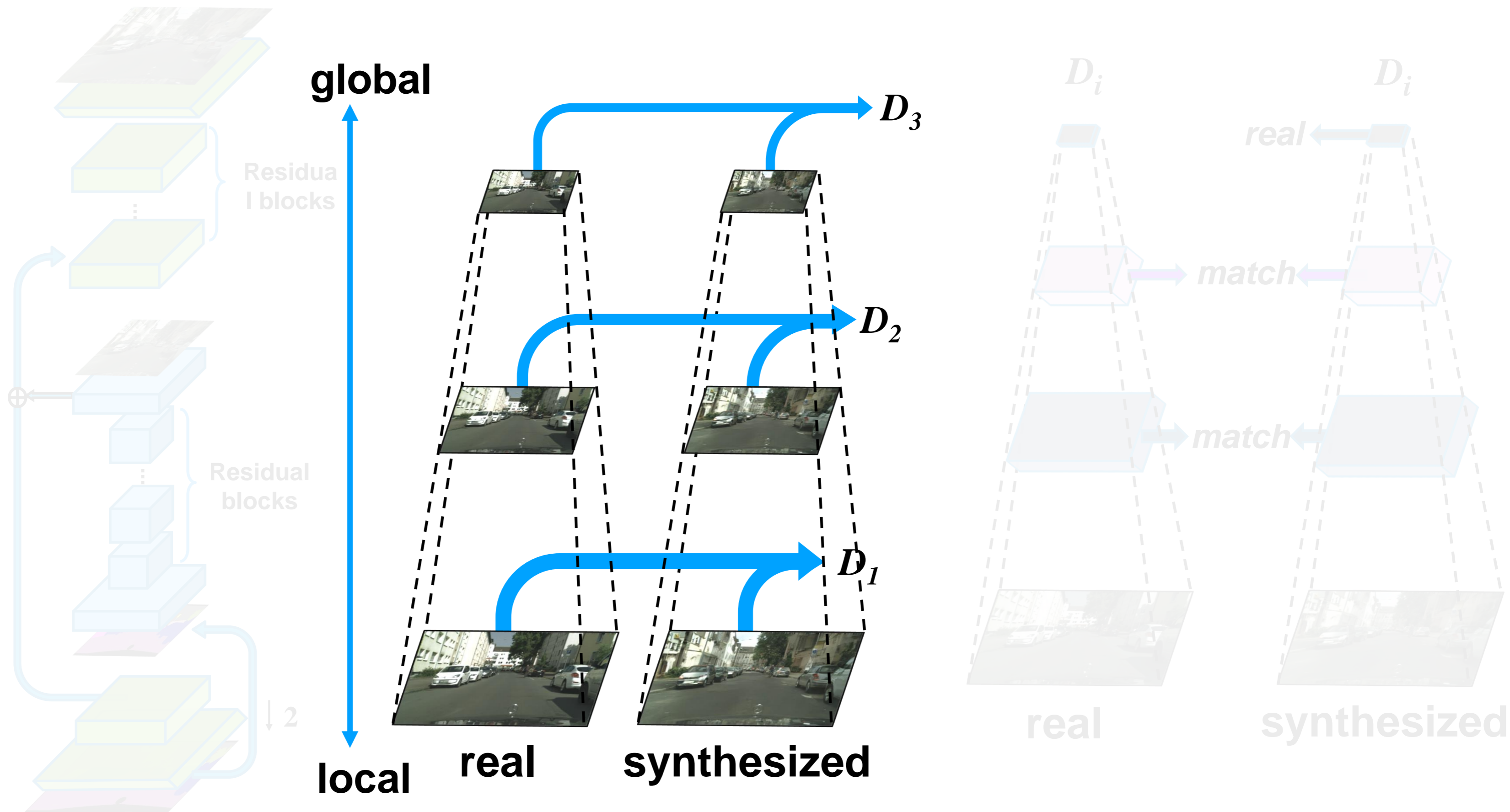


\*Similar ideas in Denton et al. 2015, Huang et al. 2017, Chen et al. 2017, Zhang et al. 2017

# Coarse-to-fine Generator

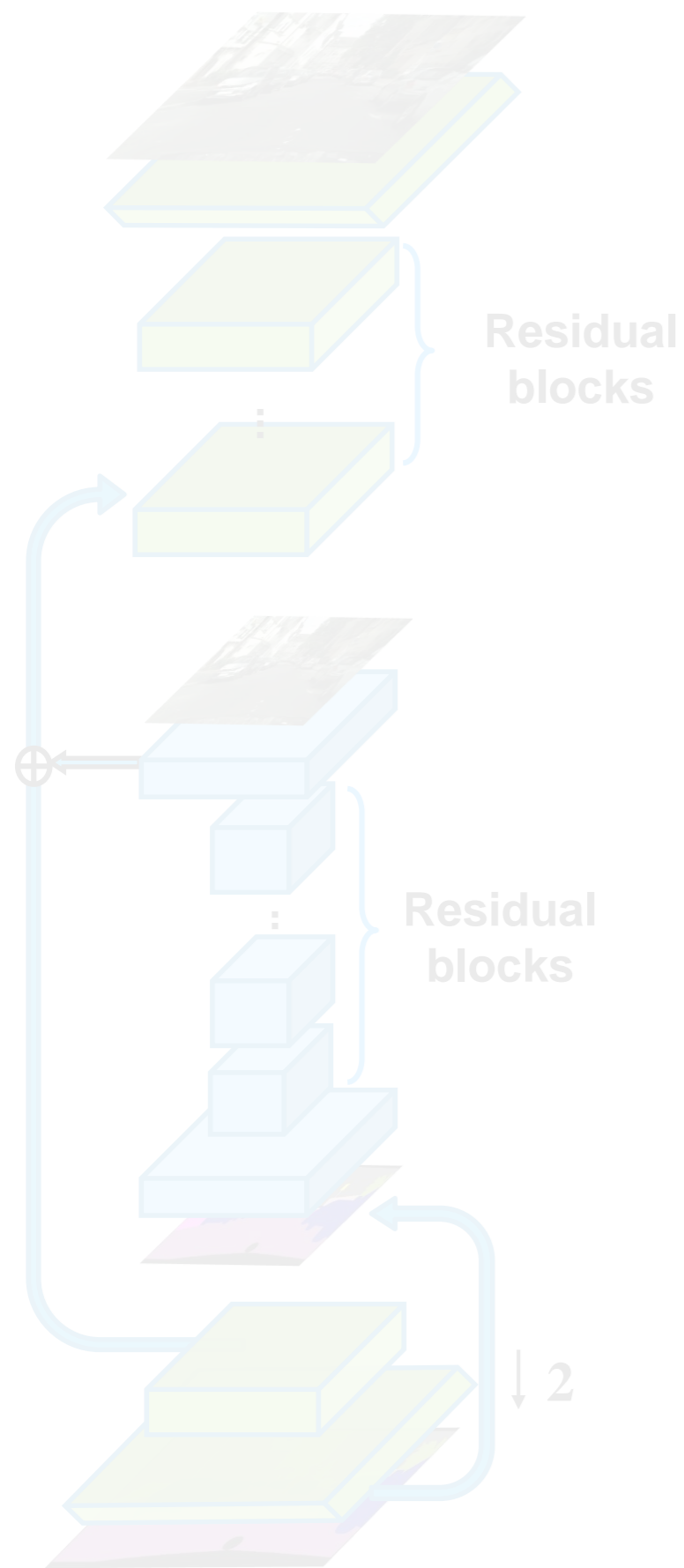
# Multi-scale Discriminators

# Robust Objective

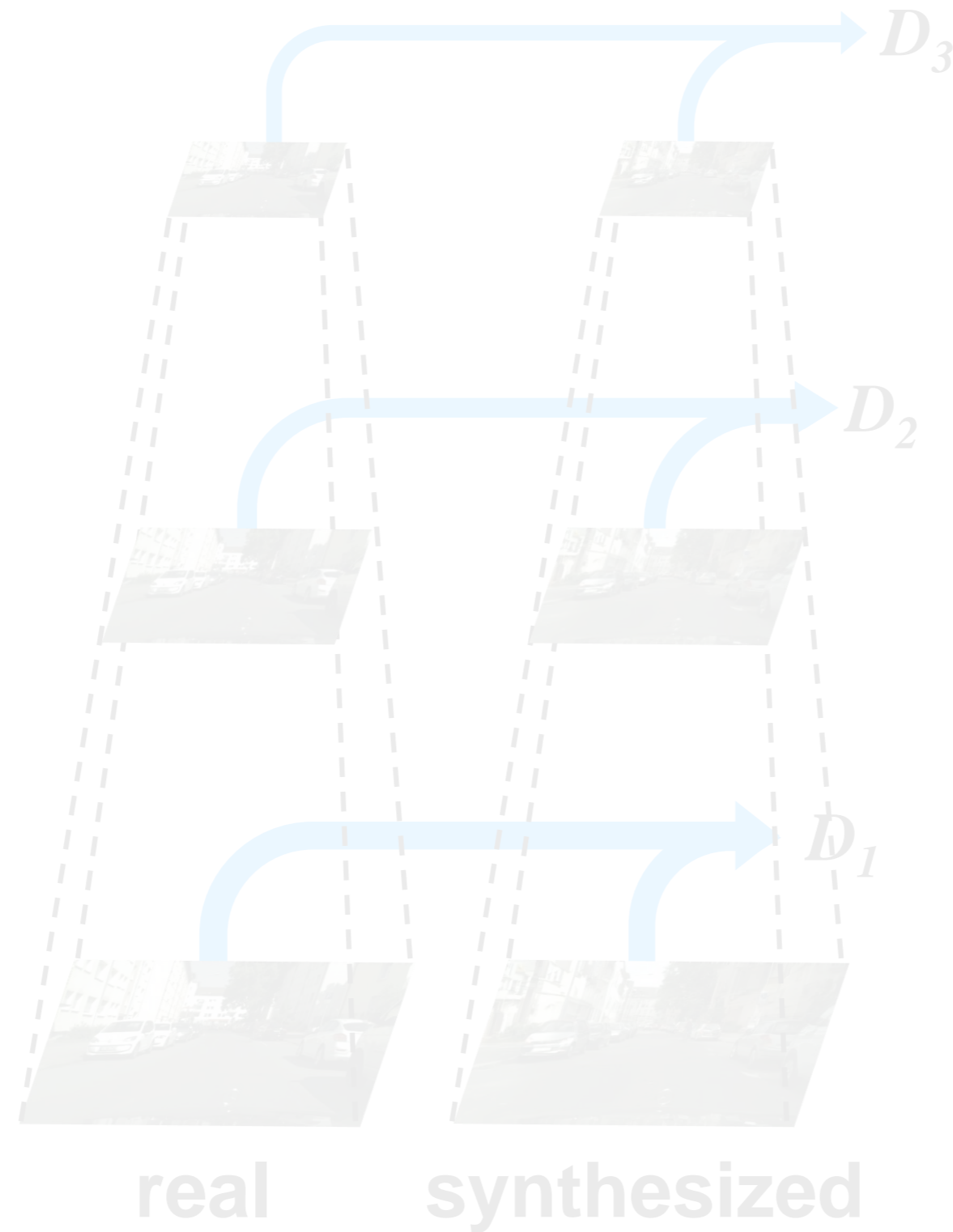


\*Similar ideas in Durugkar et al. 2016, Iizuka et al. 2017, Zhang et al. 2017

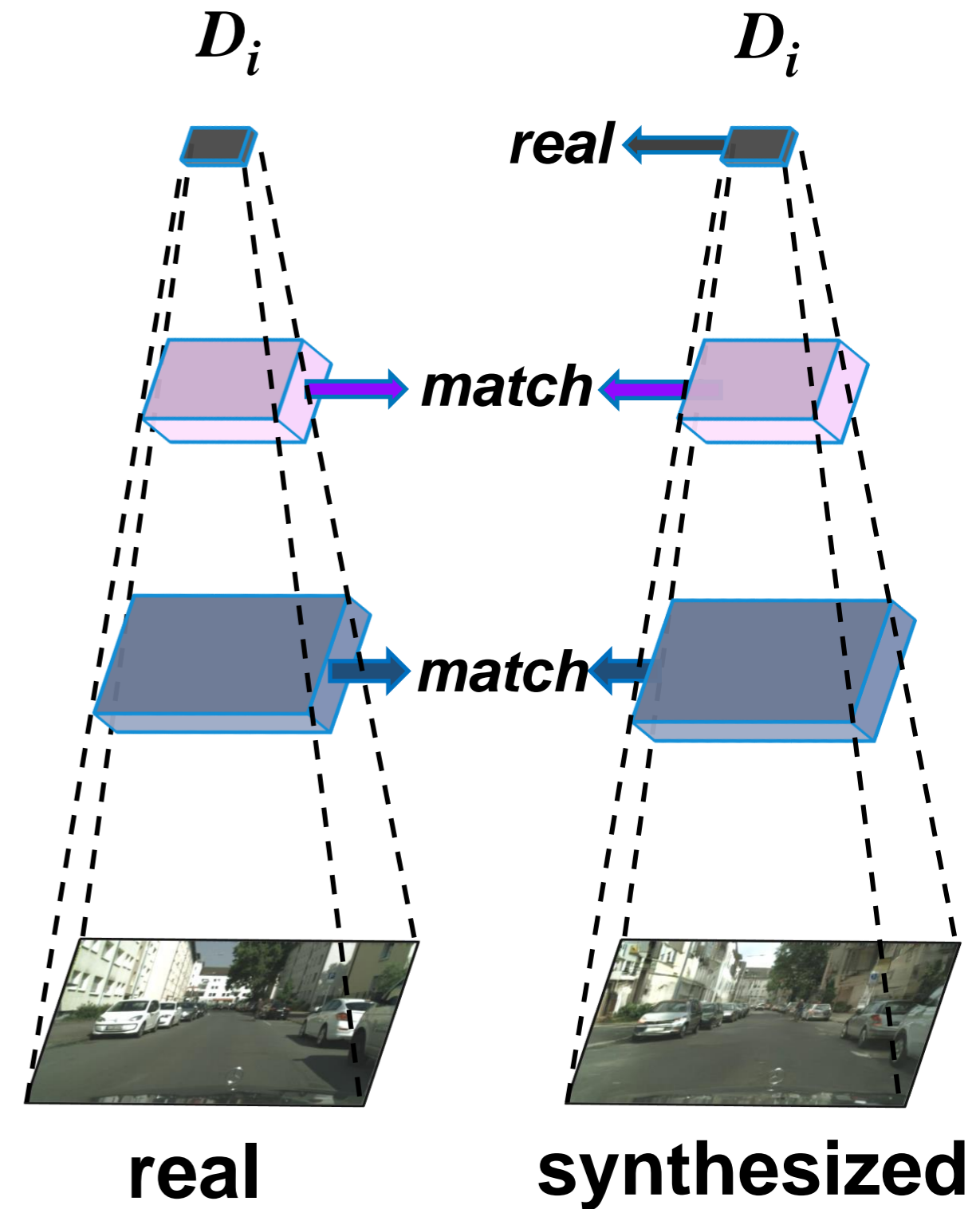
# Coarse-to-fine Generator



# Multi-scale Discriminators



# Robust Objective



\*Similar ideas in Larsen et al. 2016

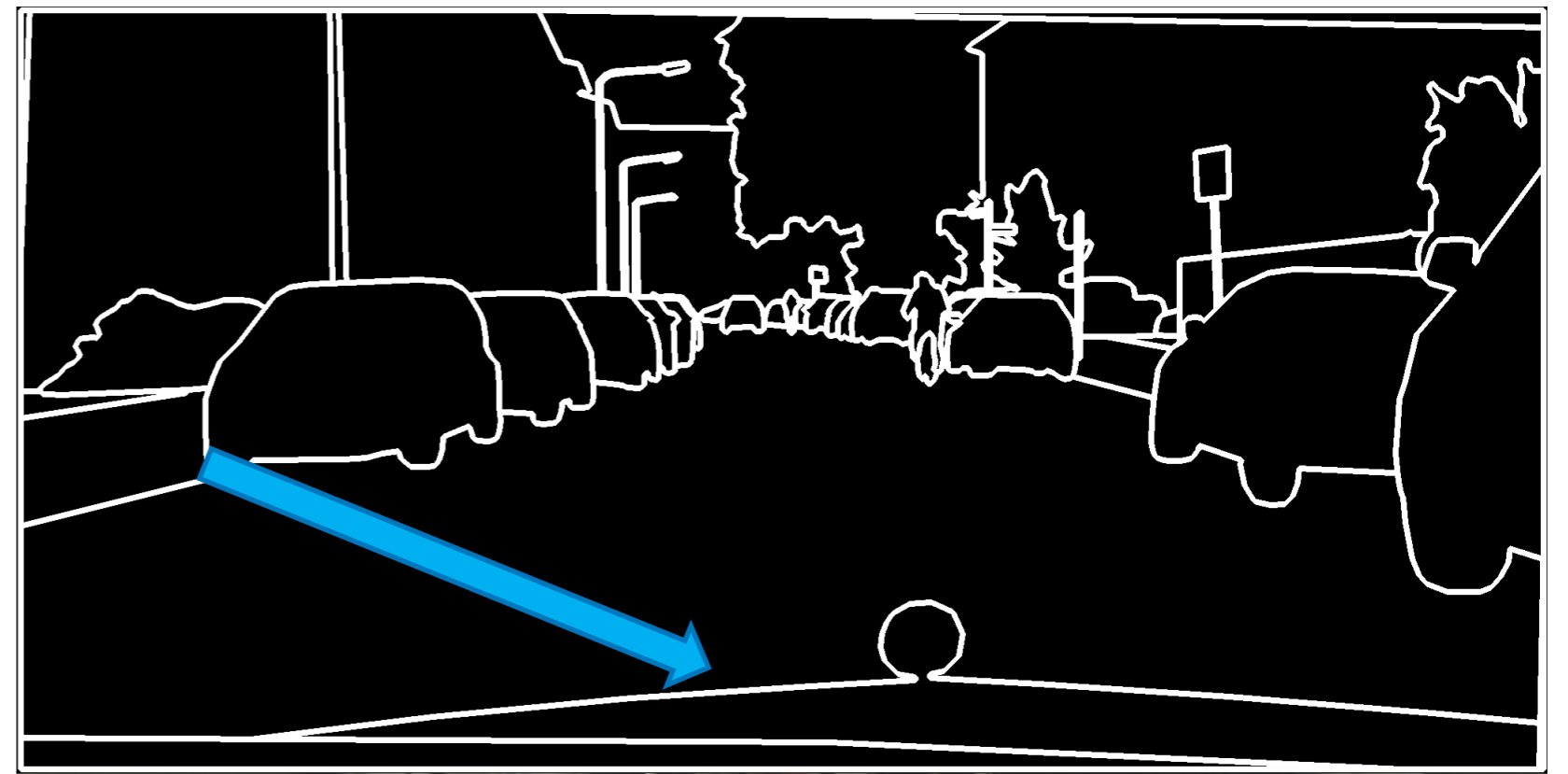
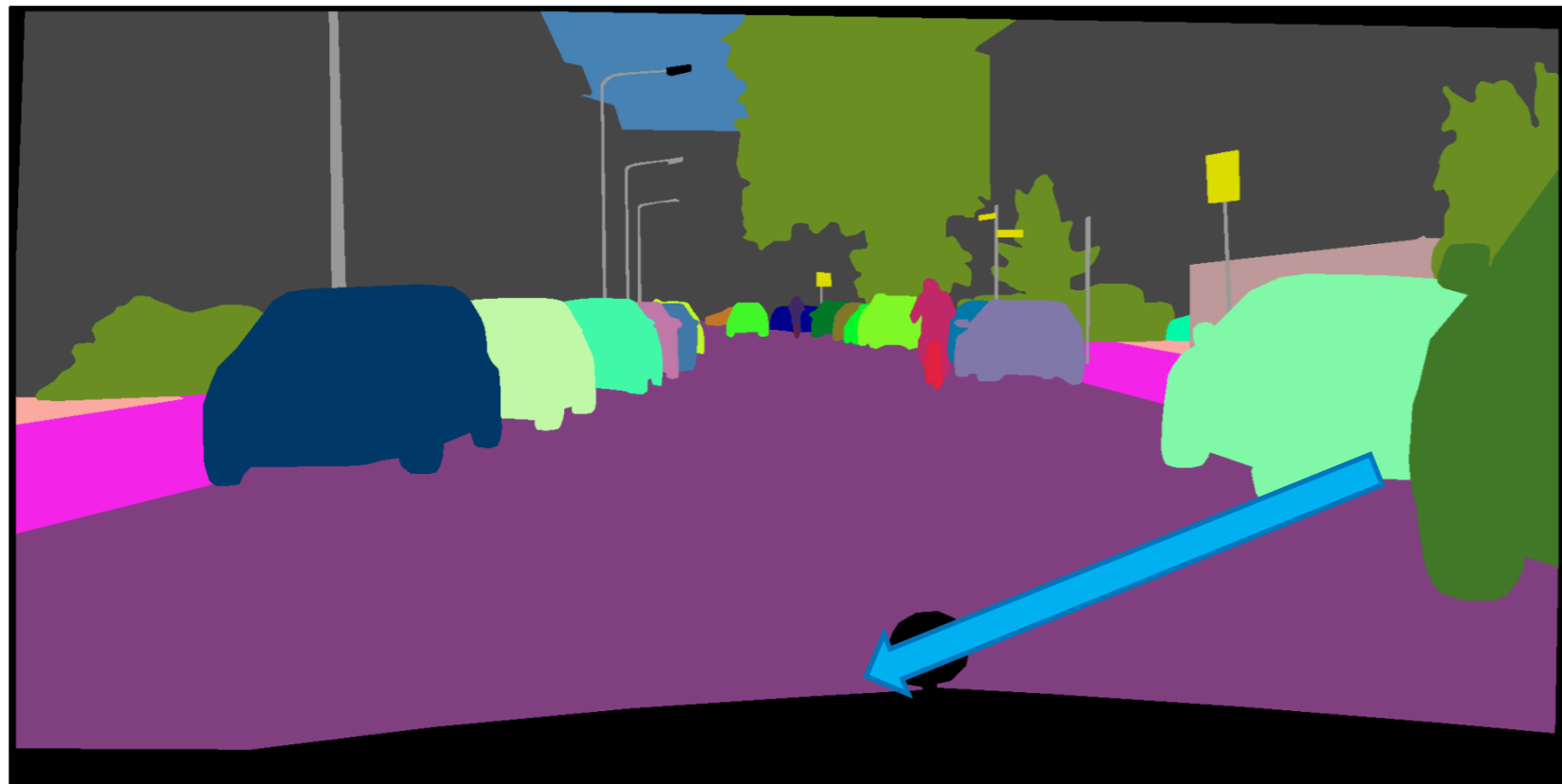
# pix2pixHD

- Extending pix2pix to high resolution
- Using instance-level segmentation maps
  - Boundary improvement
  - Multi-modal results using feature embedding

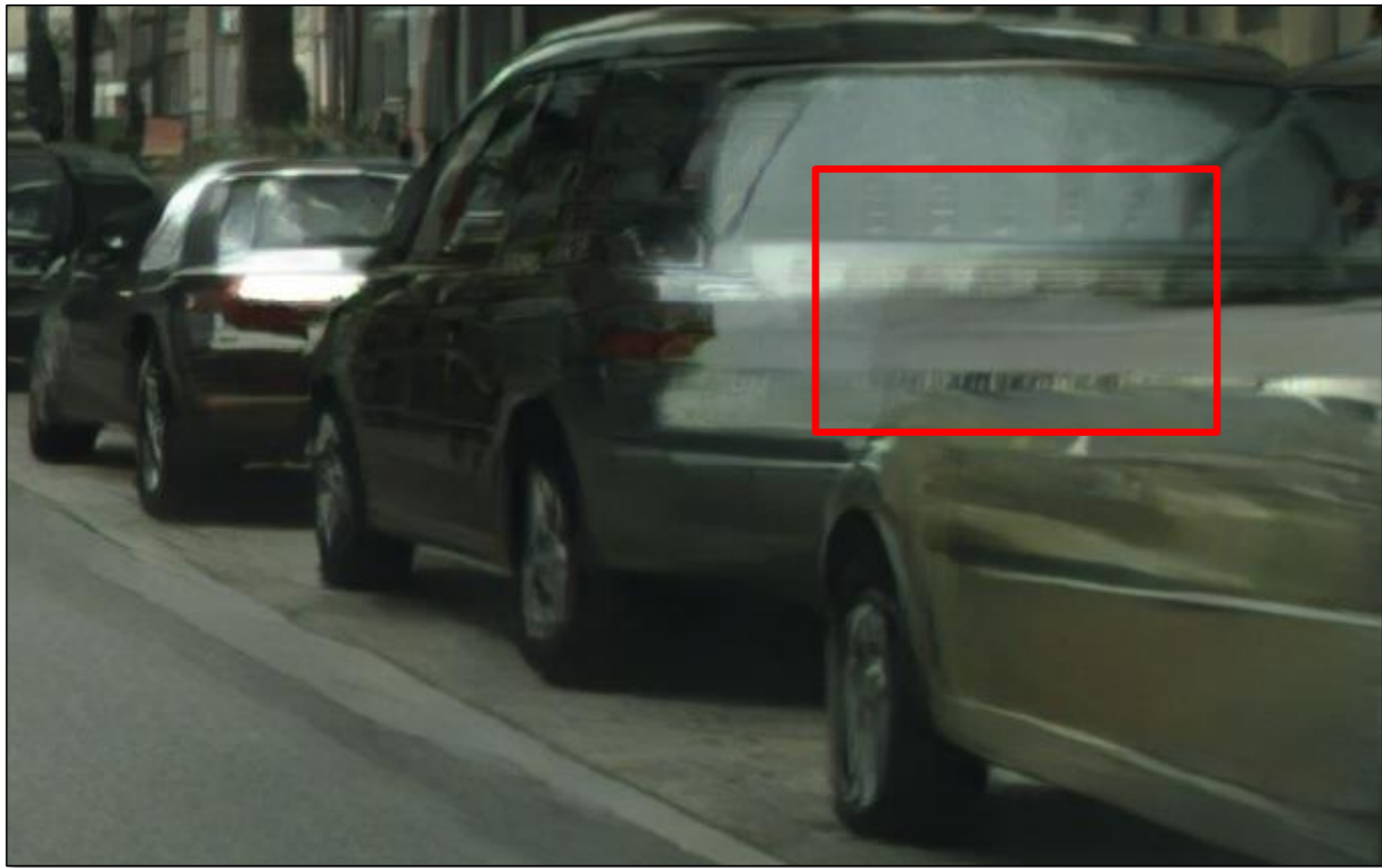
# pix2pixHD

- Extending pix2pix to high resolution
- Using instance-level segmentation maps
  - **Boundary improvement**
  - Multi-modal results using feature embedding

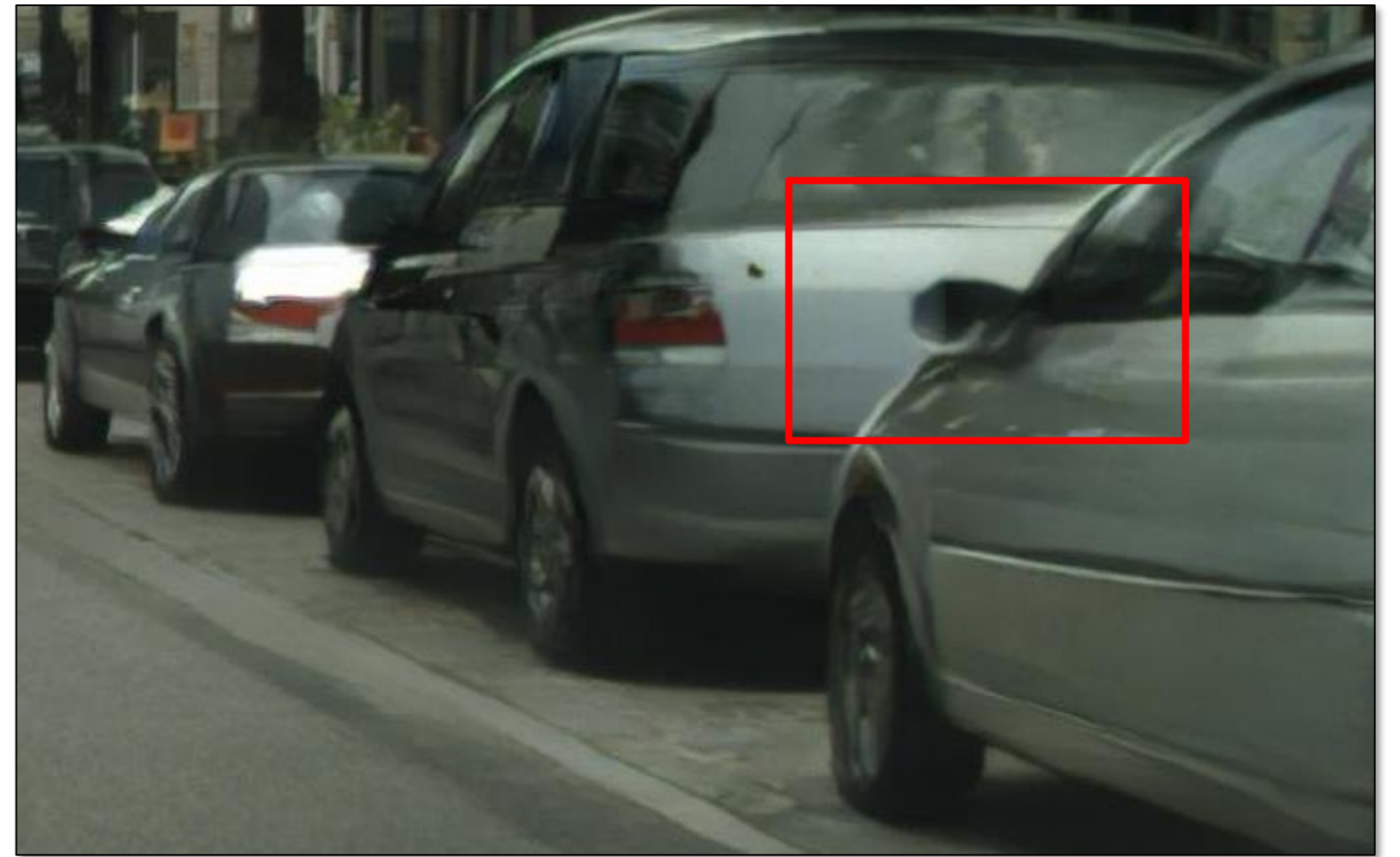
# pix2pixHD: Boundary Improvement



# pix2pixHD: Boundary Improvement



**without instance maps**



**with instance maps**

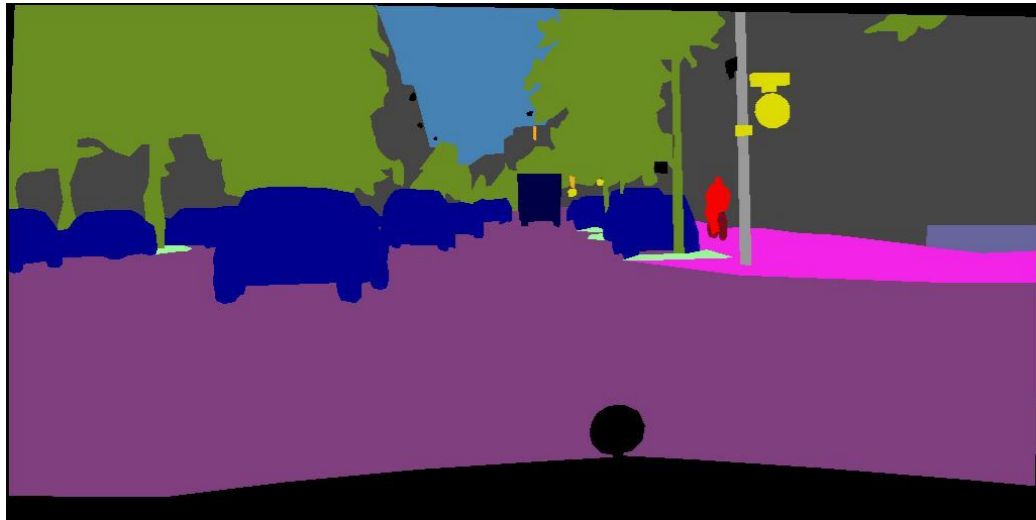
# pix2pixHD

- Extending pix2pix to high resolution
- Using instance-level segmentation maps
  - Boundary improvement
  - **Multi-modal results using feature embedding**

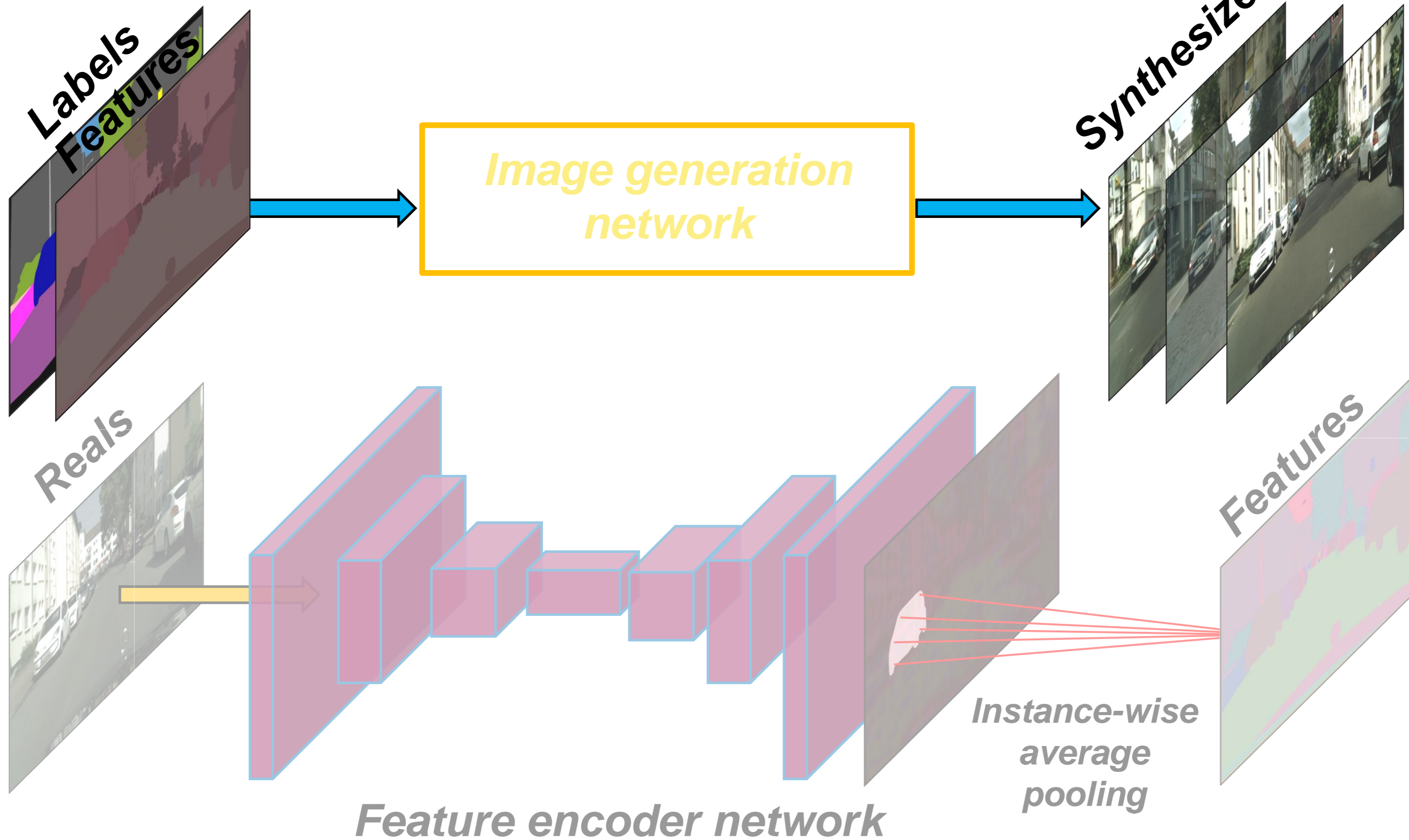


# pix2pixHD: Multi-Modal Results

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



# *Feature Embedding: Trifairange*



# pix2pixHD: Comparisons



Semantic Map



pix2pix



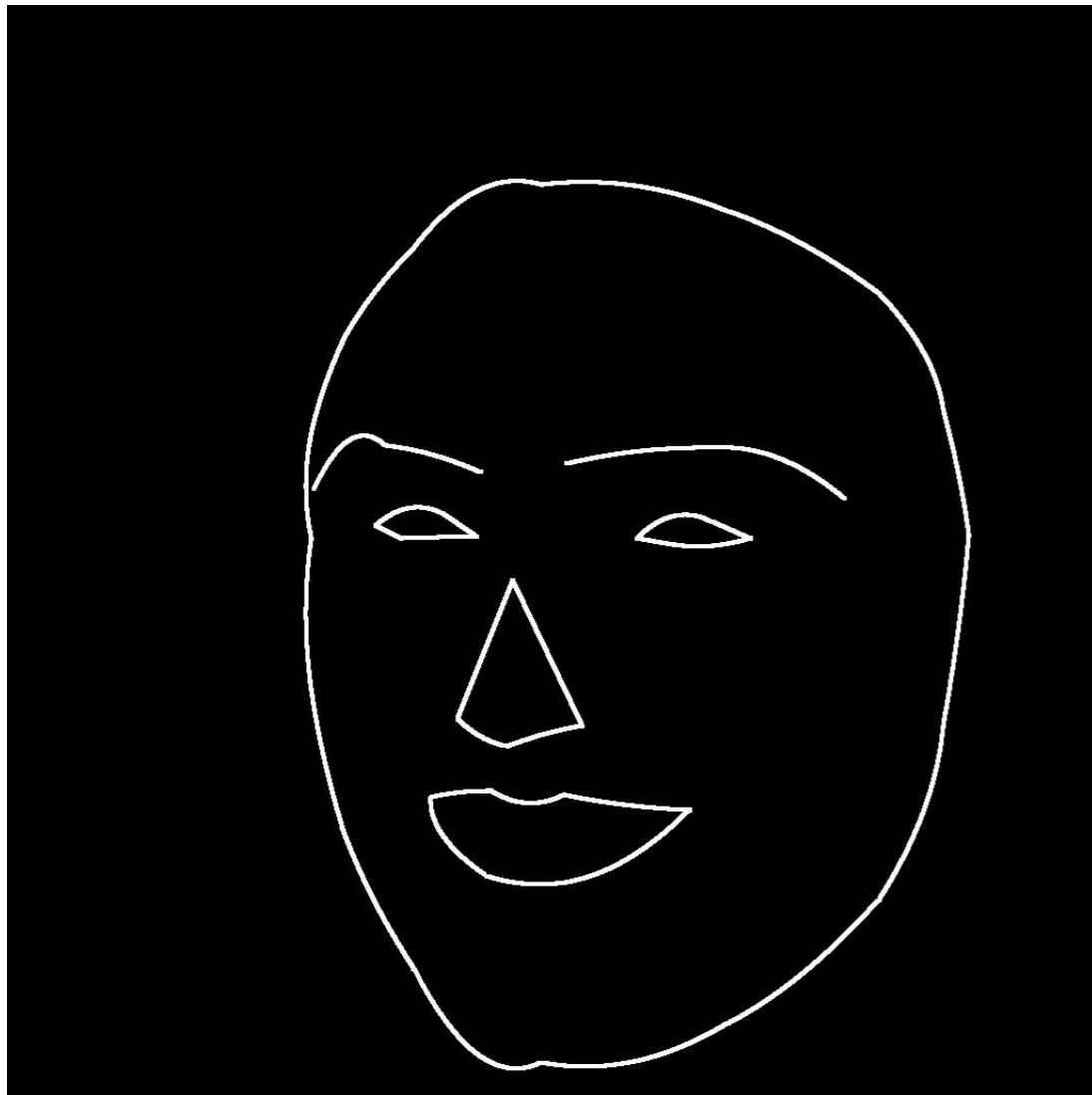
CRN



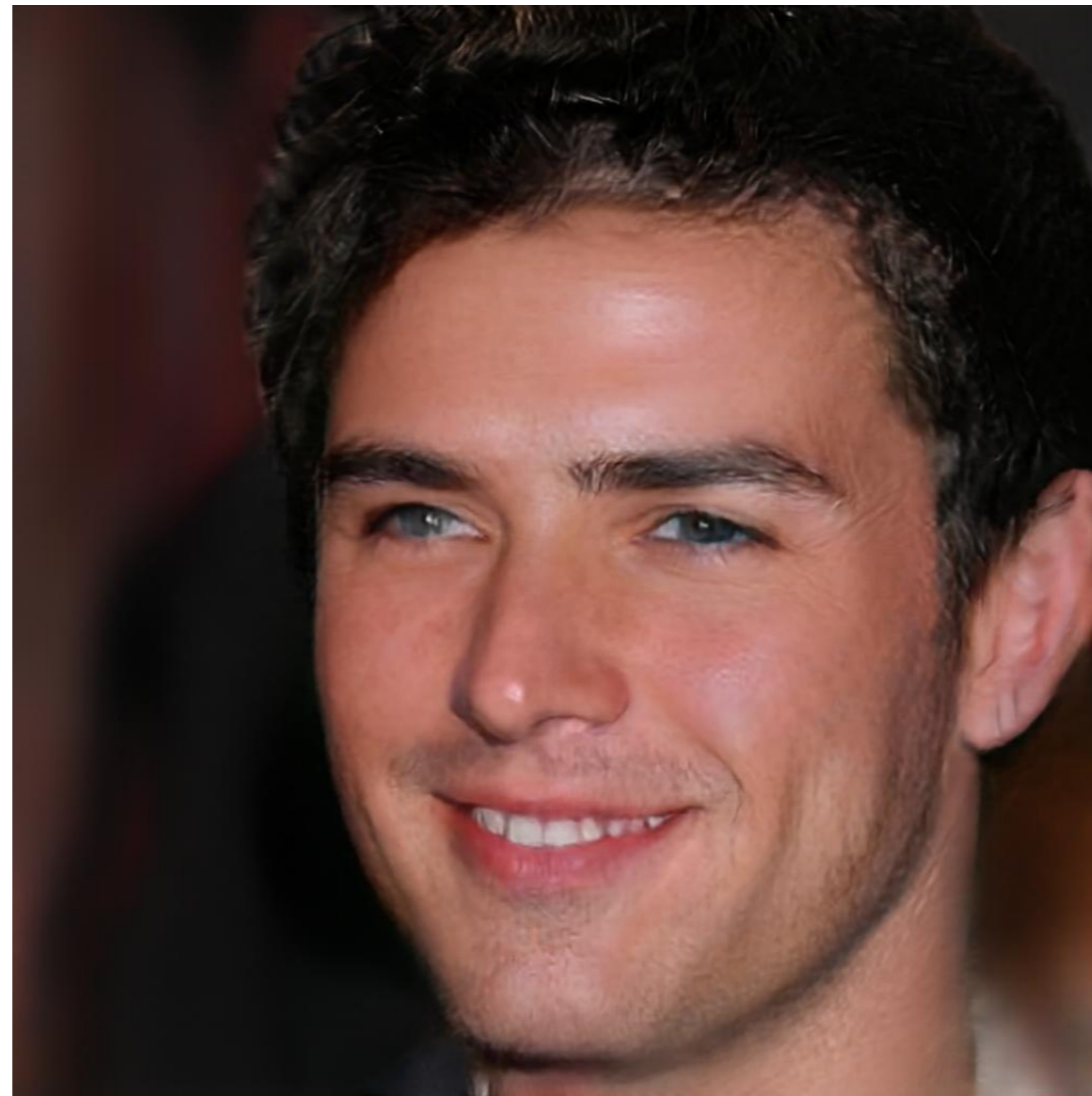
Ours

# pix2pixHD: Face Results

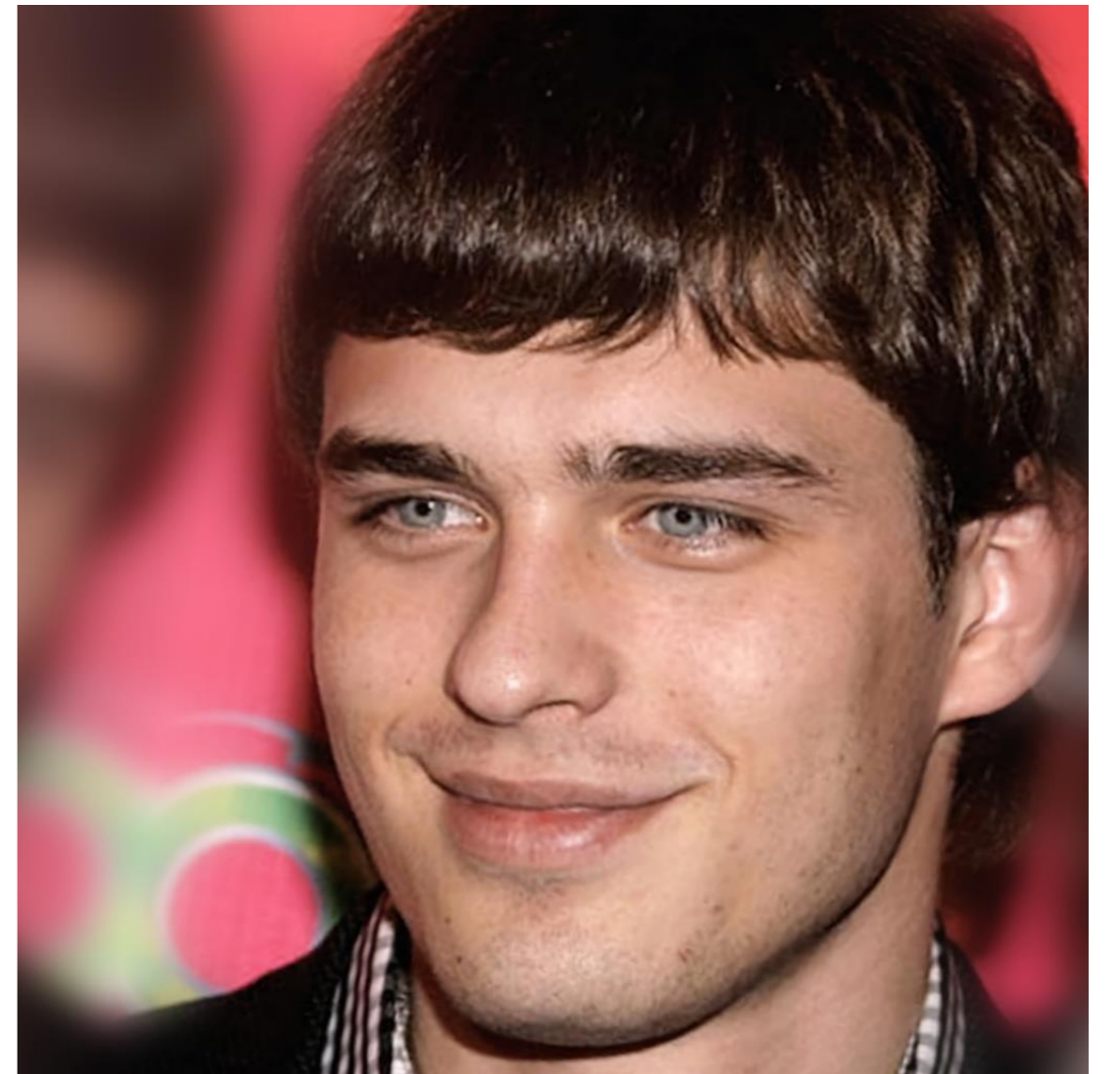
- CelebA-HQ



**Edges**



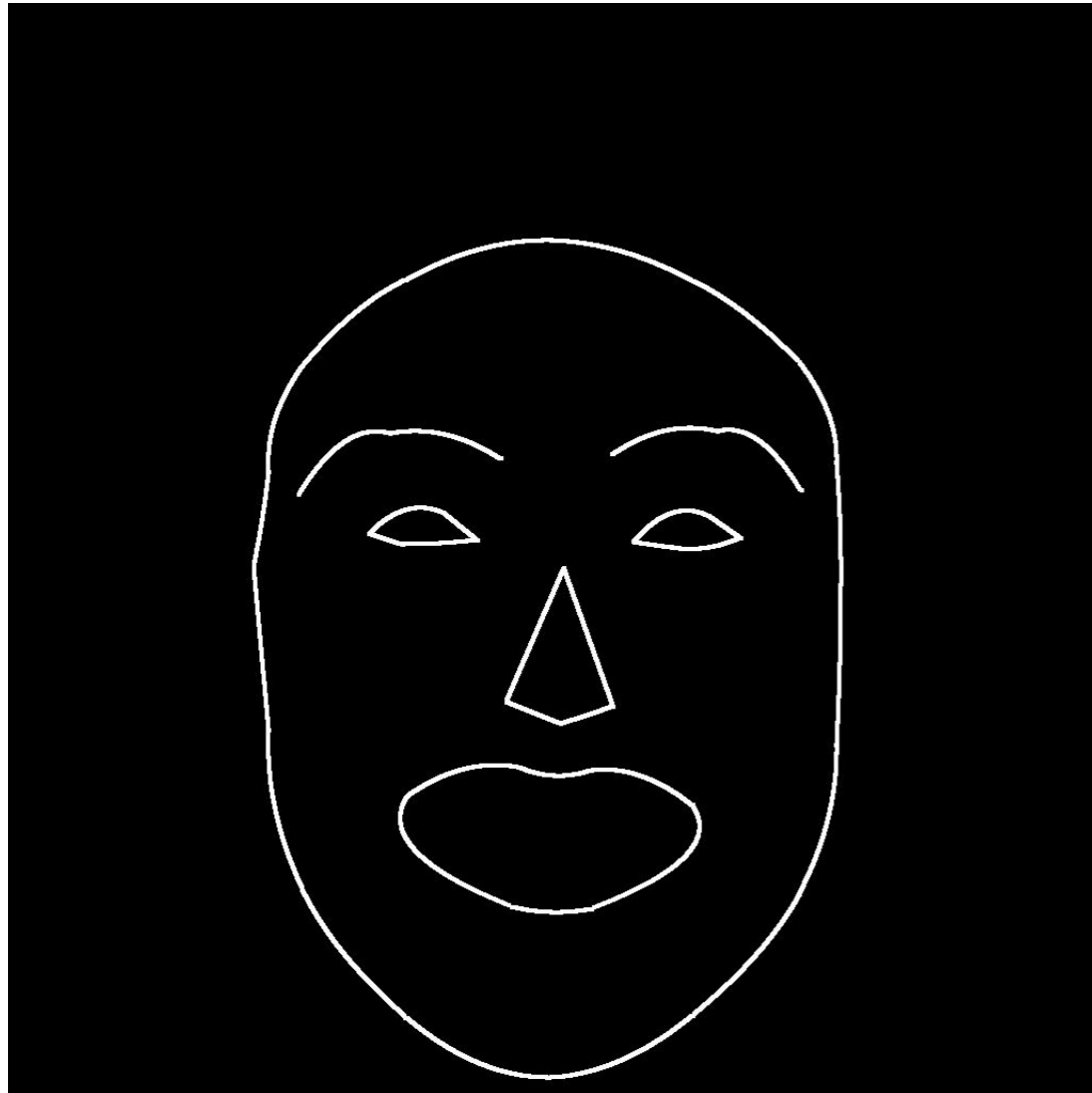
**Synthesized**



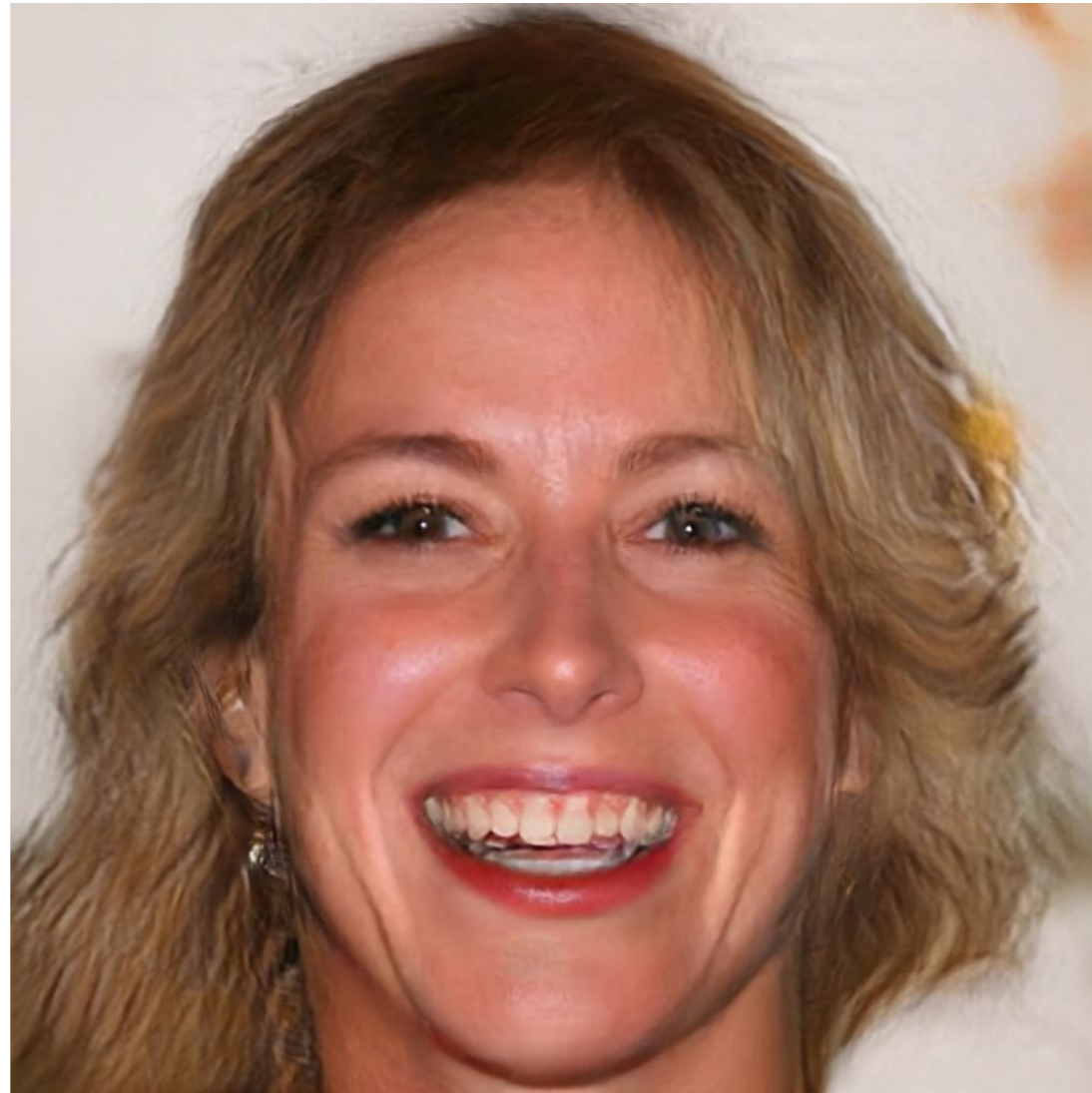
**Ground truth**

# pix2pixHD: Face Results

- CelebA-HQ



**Edges**



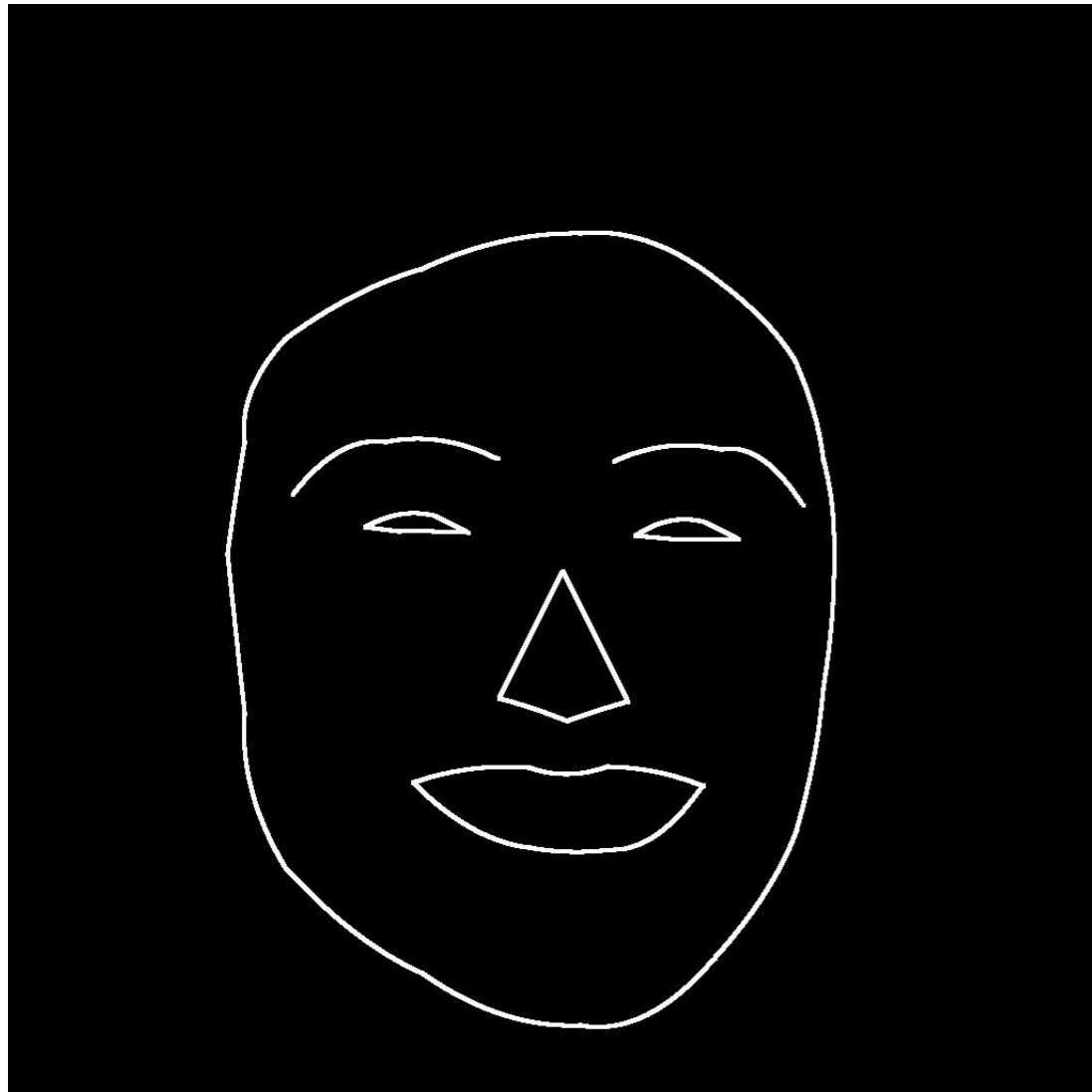
**Synthesized**



**Ground truth**

# pix2pixHD: Face Results

- CelebA-HQ



**Edges**



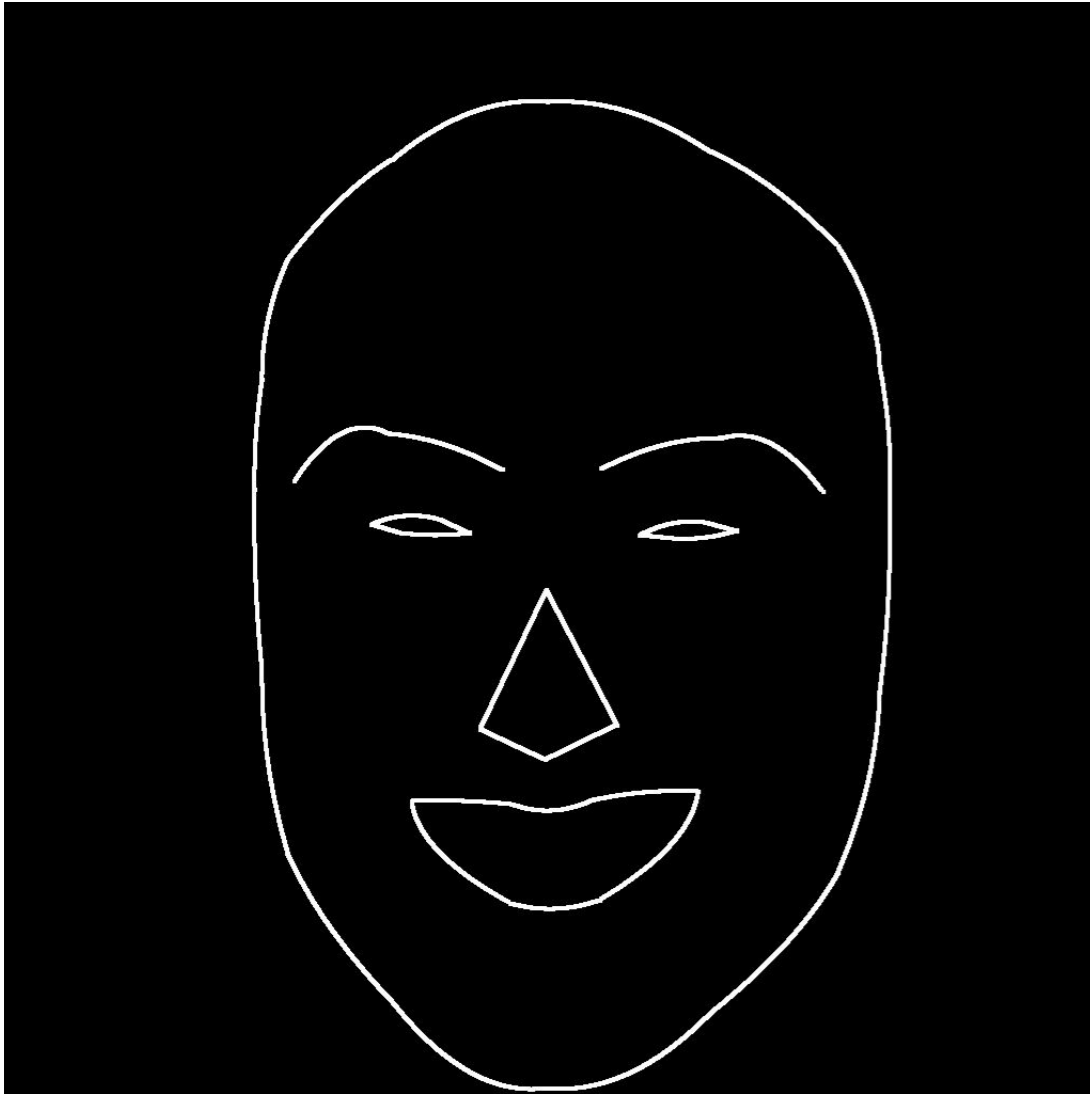
**Synthesized**



**Ground truth**

# pix2pixHD: Face Results

- CelebA-HQ



**Edges**

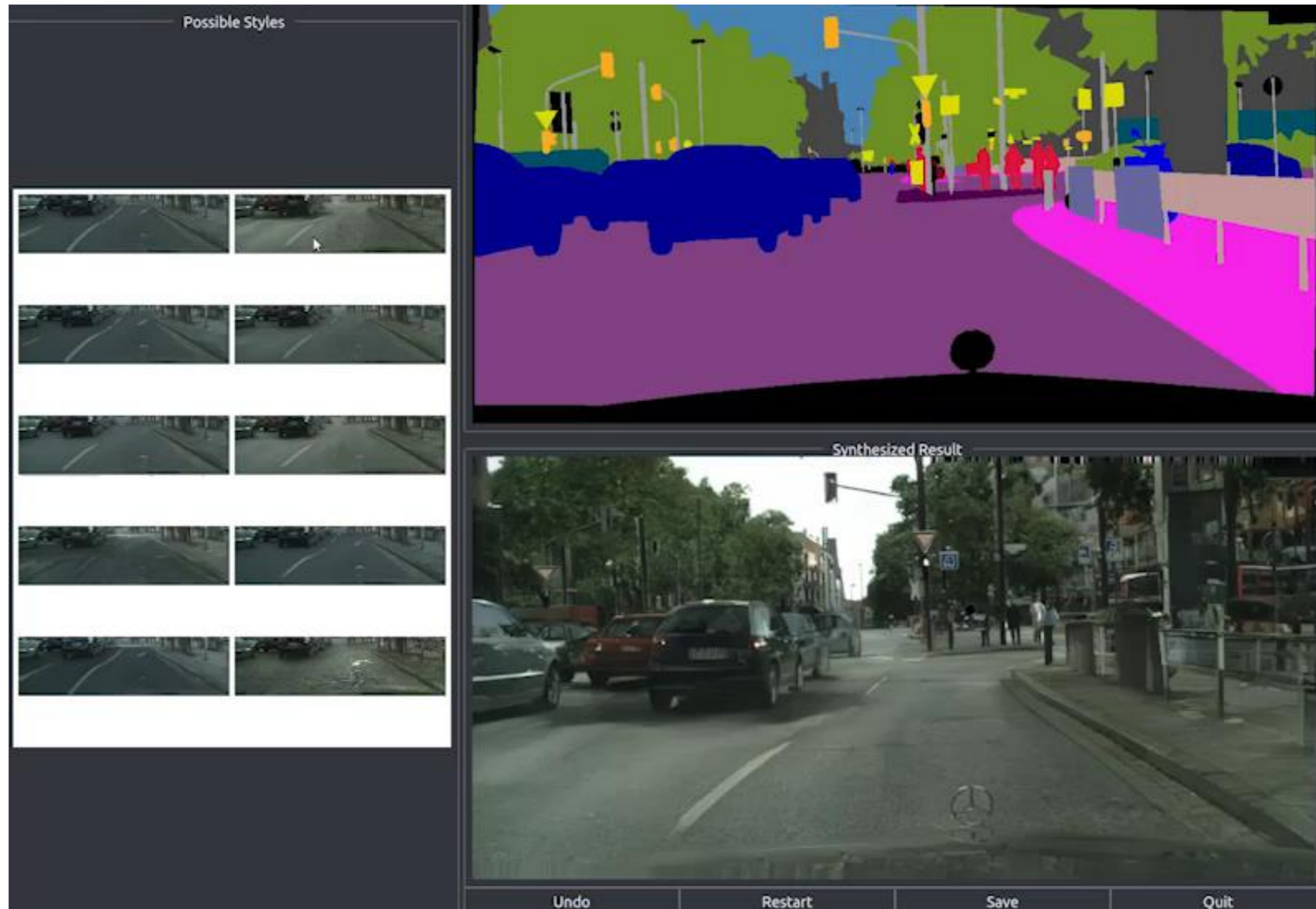


**Synthesized**



**Ground truth**

# pix2pixHD Applications: Style Changing

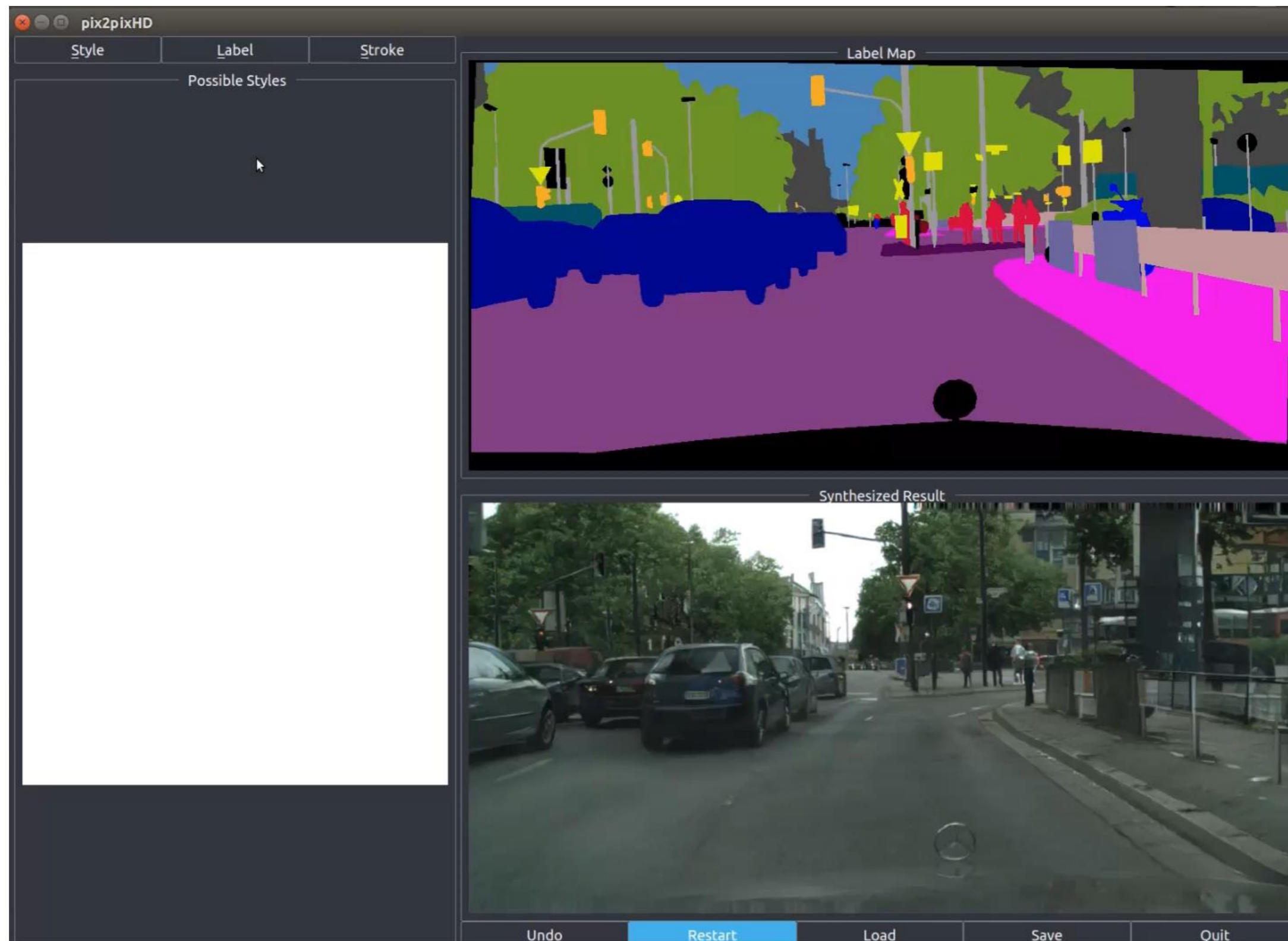




# pix2pixHD Applications: Style Changing



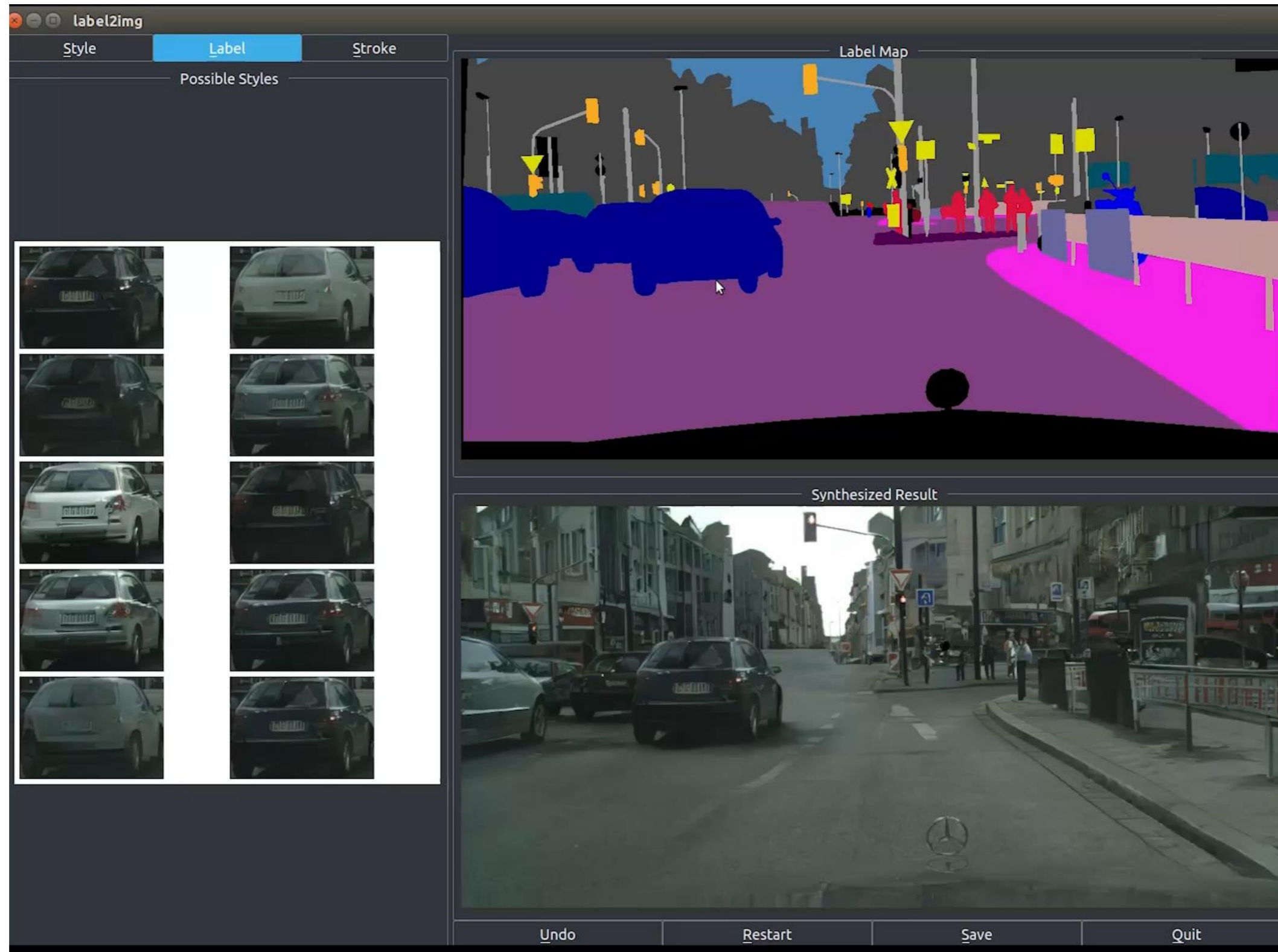
# pix2pixHD Applications: Label Changing



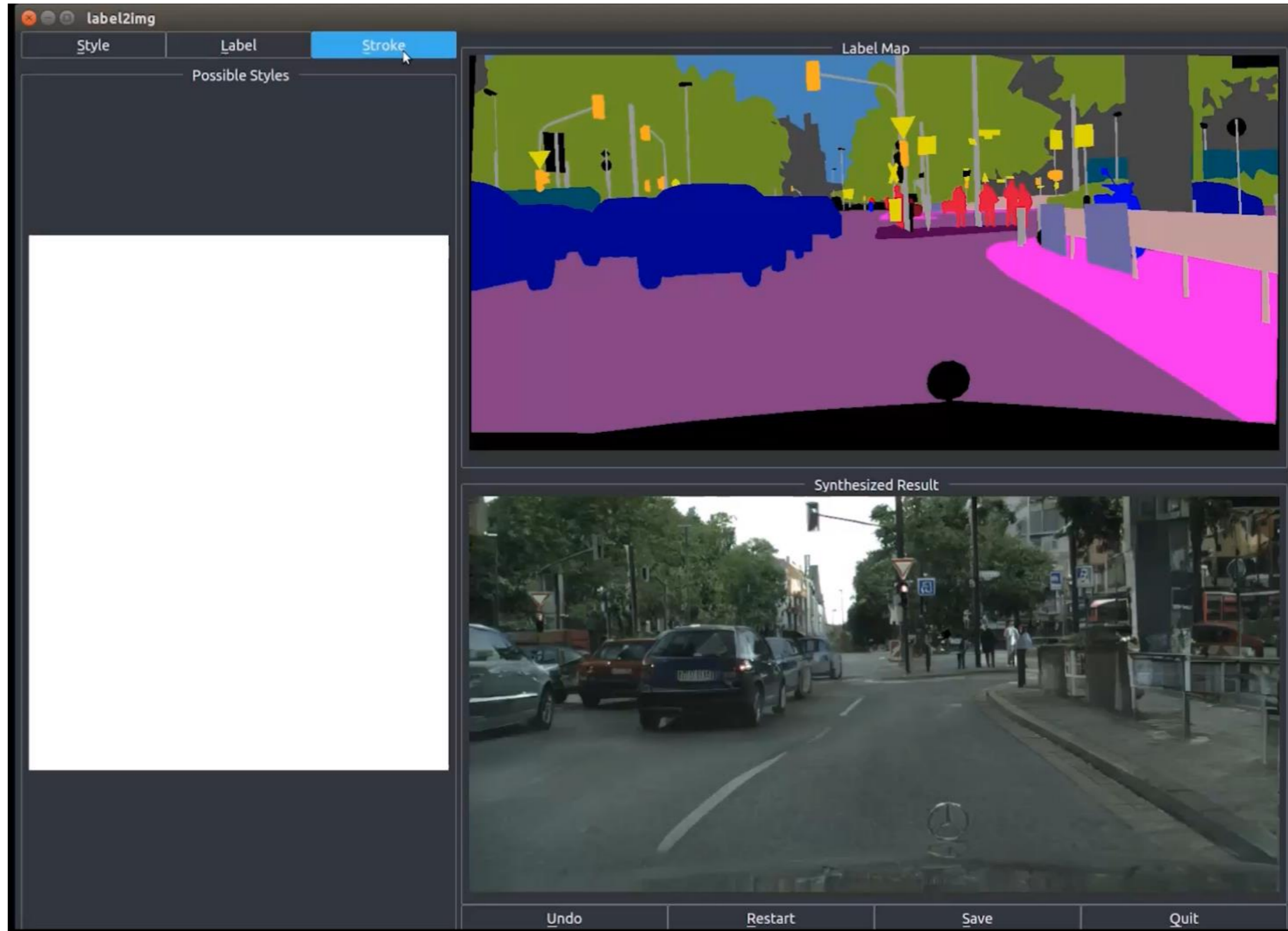
# pix2pixHD Applications: Label Changing



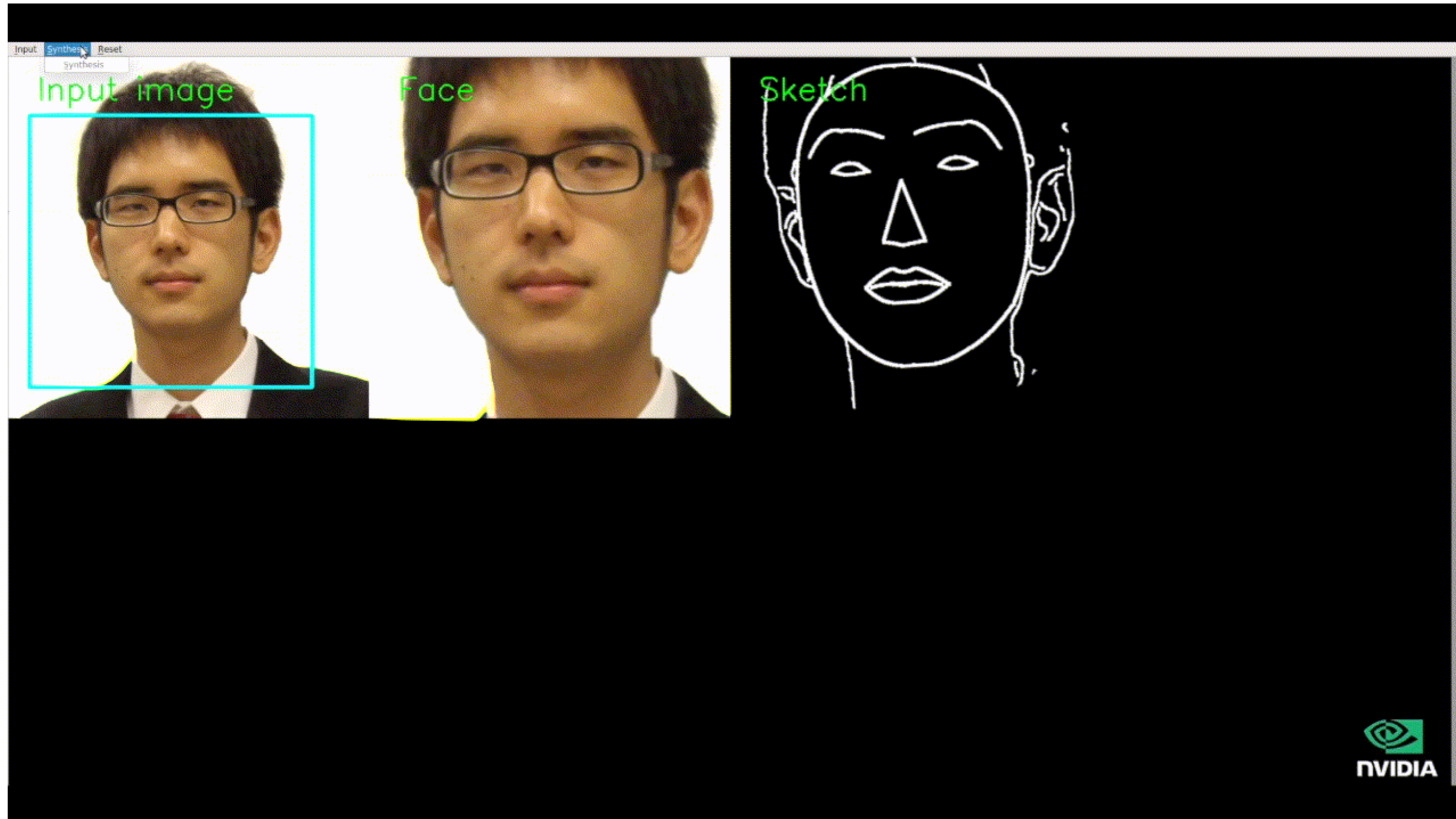
# pix2pixHD Applications: Adding Objects



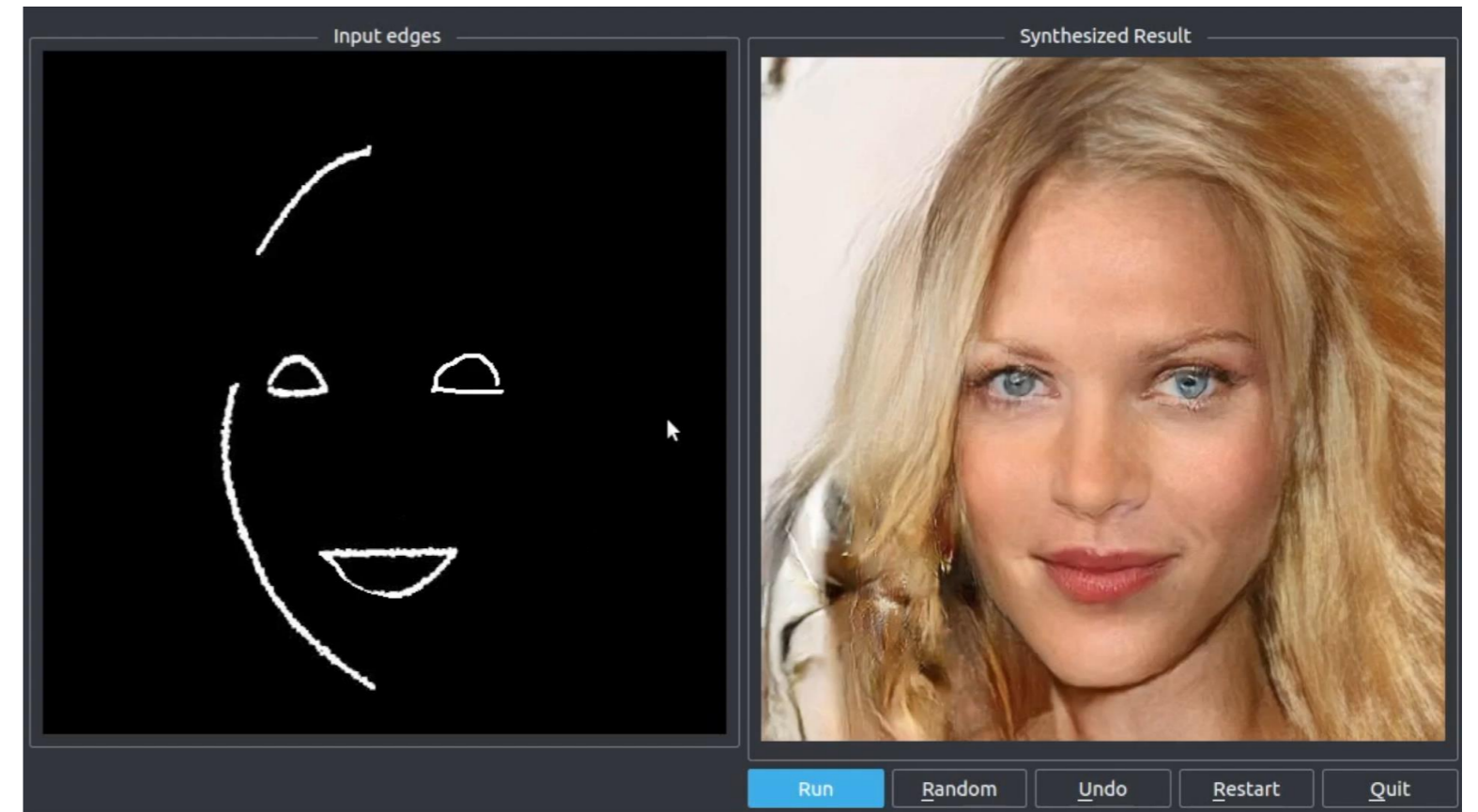
# pix2pixHD Applications: Adding Strokes



# pix2pixHD Extensions



# pix2pixHD Extensions



Video credit: Mario Klingemann