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

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

Outline

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

Outline

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

Introduction

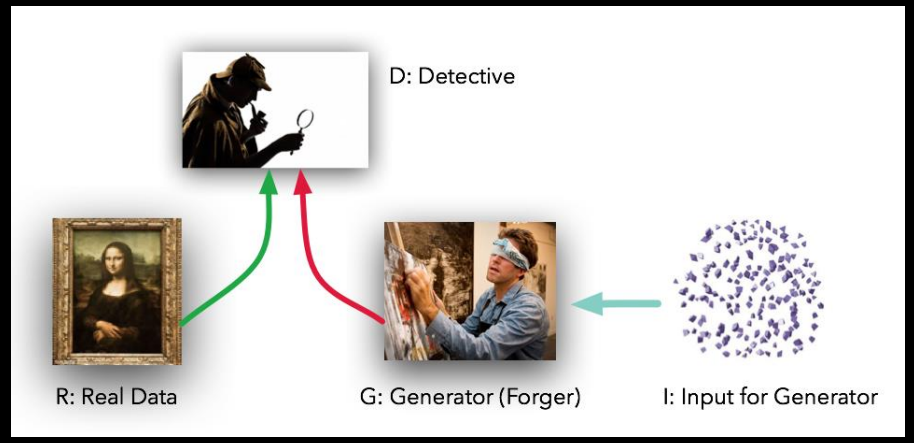
Introduction

Outline

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

Related Work

Generative Adversarial Network (GAN)

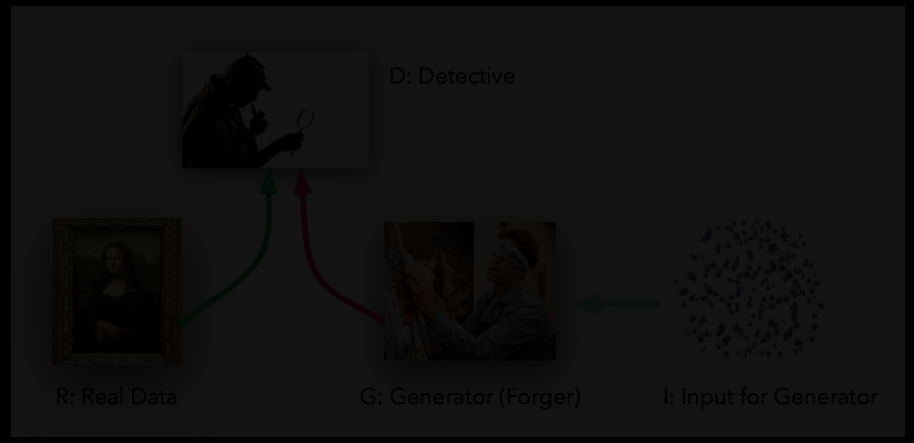


Goodfellow et al. [2014] Radford et al. [2015] Arjovsky et al. [2017]

Figure from: https://medium.com/@devnag/

Related Work

Generative Adversarial Network (GAN)



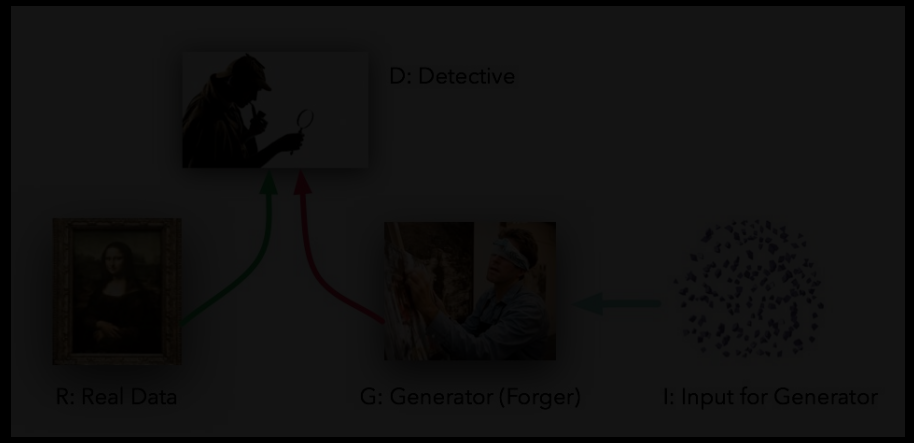
Goodfellow et al. [2014] Radford et al. [2015] Arjovsky et al. [2017]

Image-to-Image Translation

Johnson et al. [2016] Isola et al. [2017]

Related Work

Generative Adversarial Network (GAN)



Goodfellow et al. [2014] Radford et al. [2015] Arjovsky et al. [2017]

Image-to-Image Translation

Johnson et al. [2016] Isola et al. [2017]

Cascaded Refinement Network (CRN)

Chen and Koltun [2017]

Related Work

Generative Adversarial Network (GAN)

Goodfellow et al. [2014] Radford et al. [2015] Arjovsky et al. [2017]

Image-to-Image Translation

Johnson et al. [2016] Isola et al. [2017]

Cascaded Refinement Network (CRN)

Chen and Koltun [2017]

Deep Visual Manipulation

Zhu et al. [2016] Zhang et al. [2017]

High Resolution Image

Low resolution

Semantic Manipulation

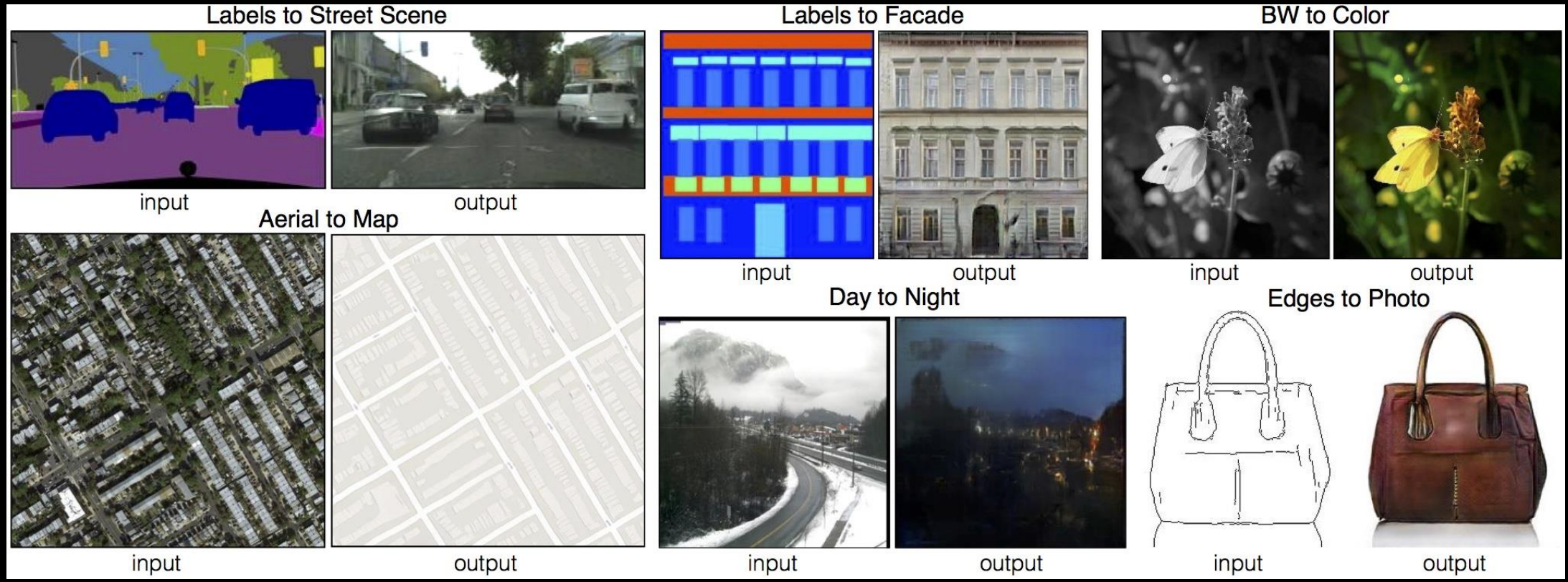
Lack fine details

Outline

- Introduction
- Related work
- Method
  - Baseline method
  - Our method
- Results
- Applications
- Conclusion

Baseline Method

- pix2pix



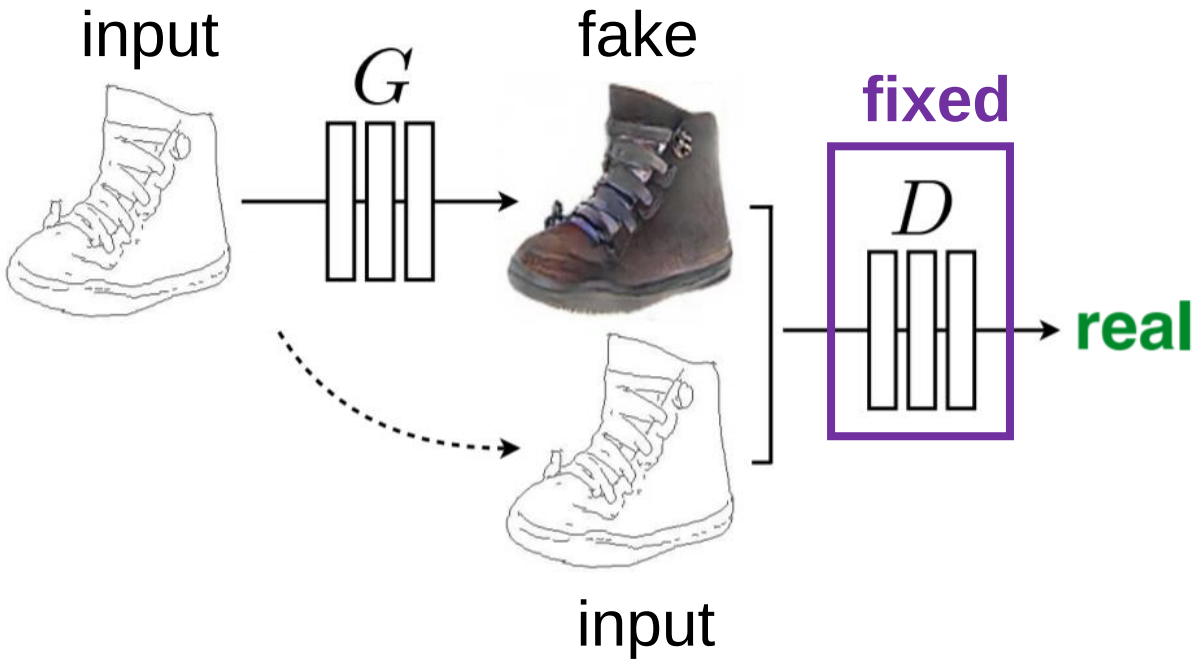
Baseline Method

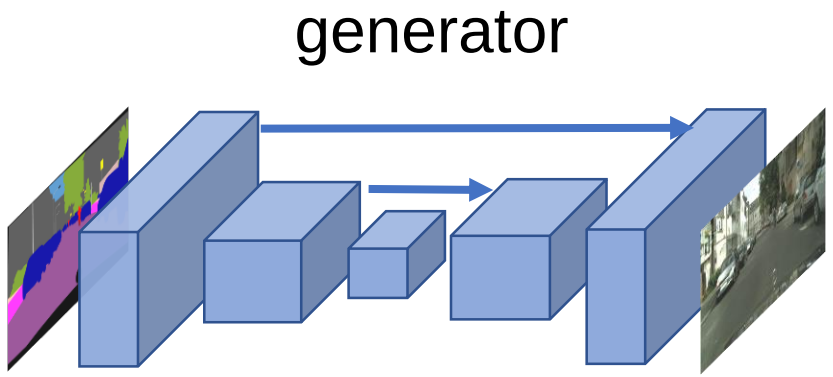
- pix2pix: training discriminator

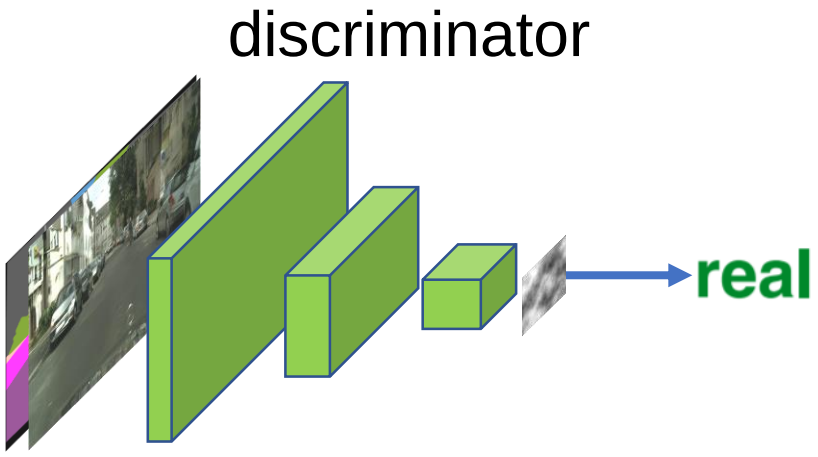
input

Baseline Method

- pix2pix: training generator







Outline

- Introduction
- Related work
- Method
  - Baseline method
  - Our method
- Results
- Applications
- Conclusion

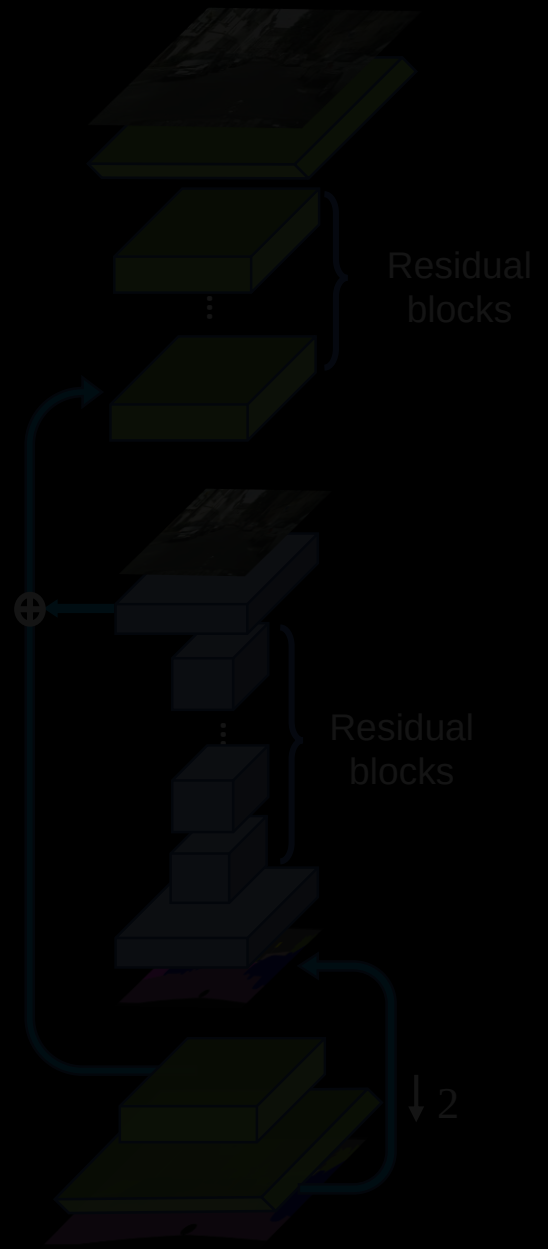
Our Method

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

Our Method

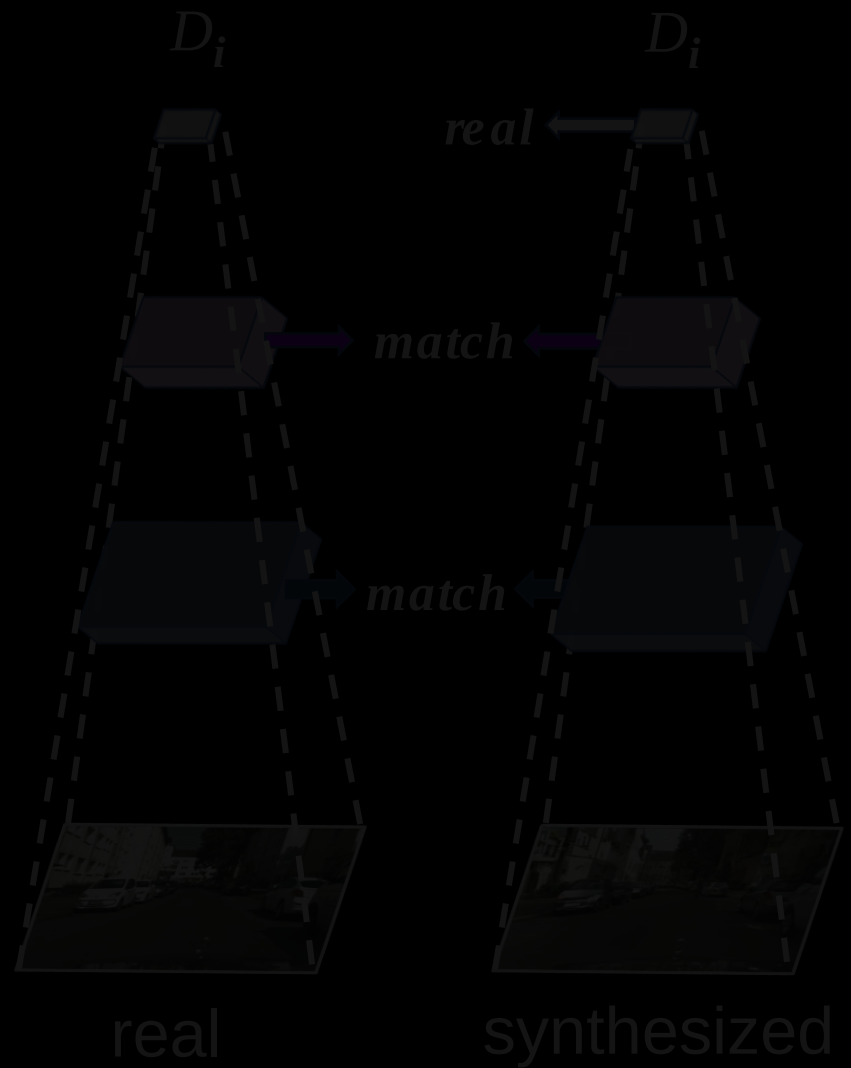
- Extending to high resolution
  - Generator design
  - Discriminator design
  - Objective function
- Using instance-level segmentation maps

Coarse-to-fine Generator

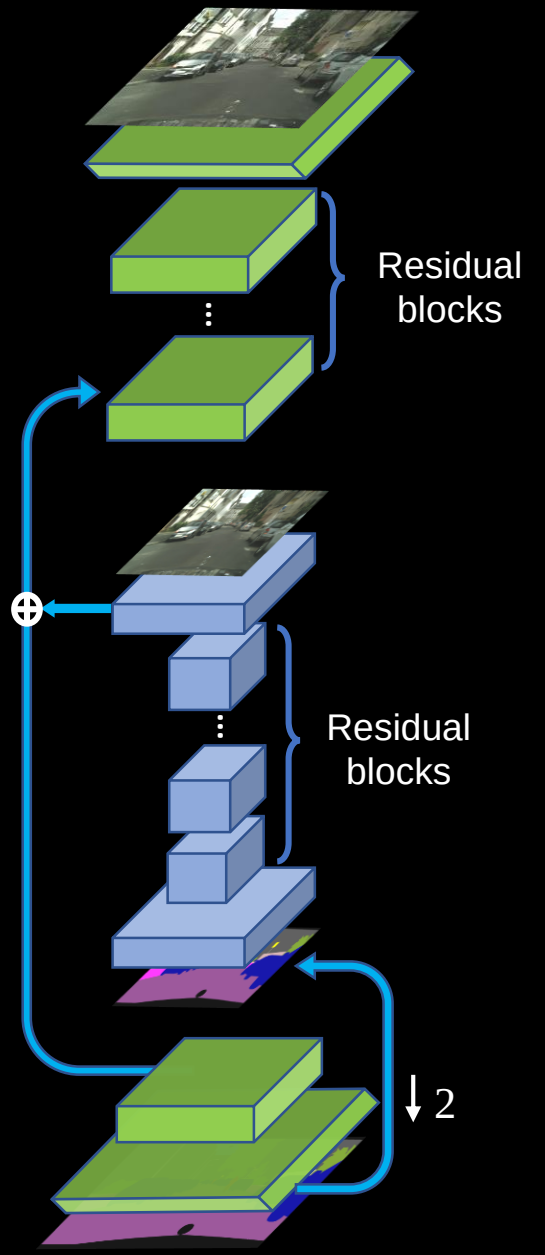


Multi-scale Discriminators

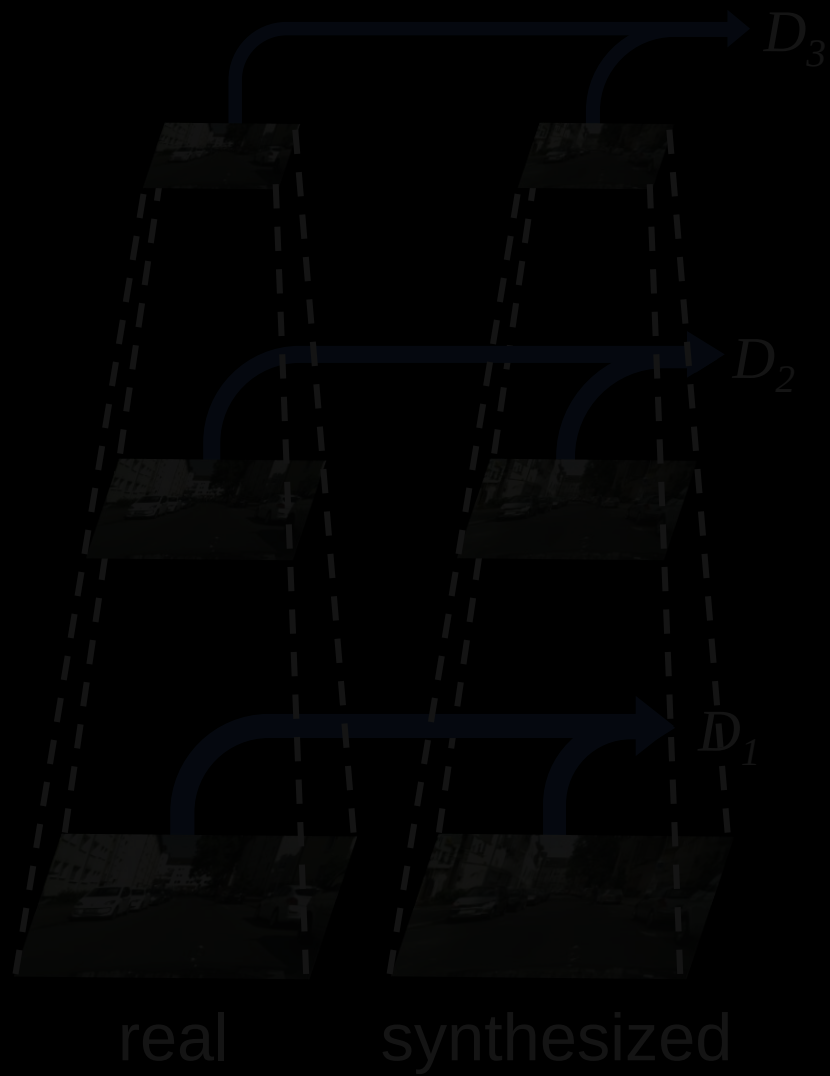
Robust Objective



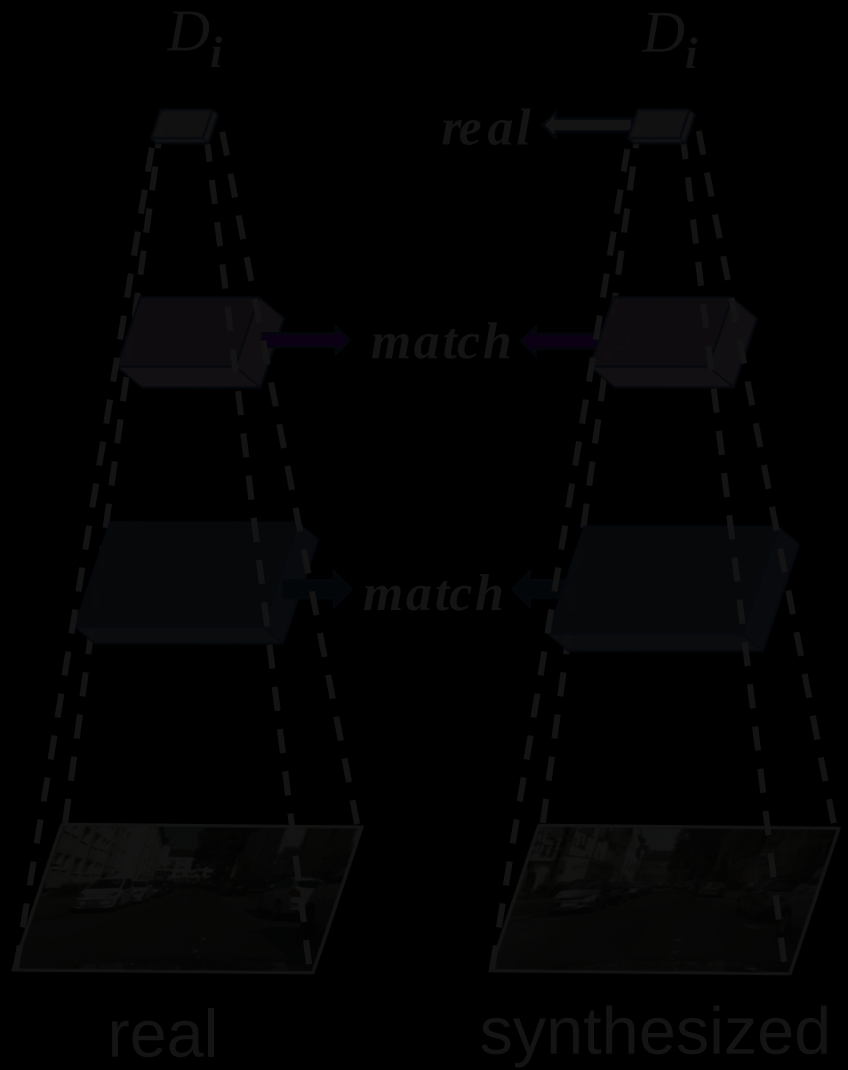
Coarse-to-fine Generator



Multi-scale Discriminators



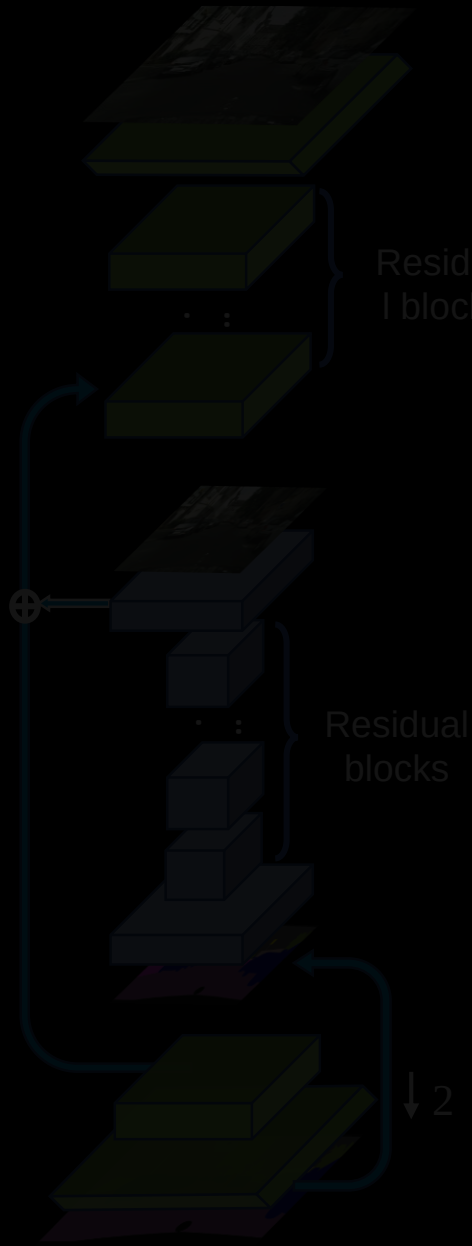
Robust Objective

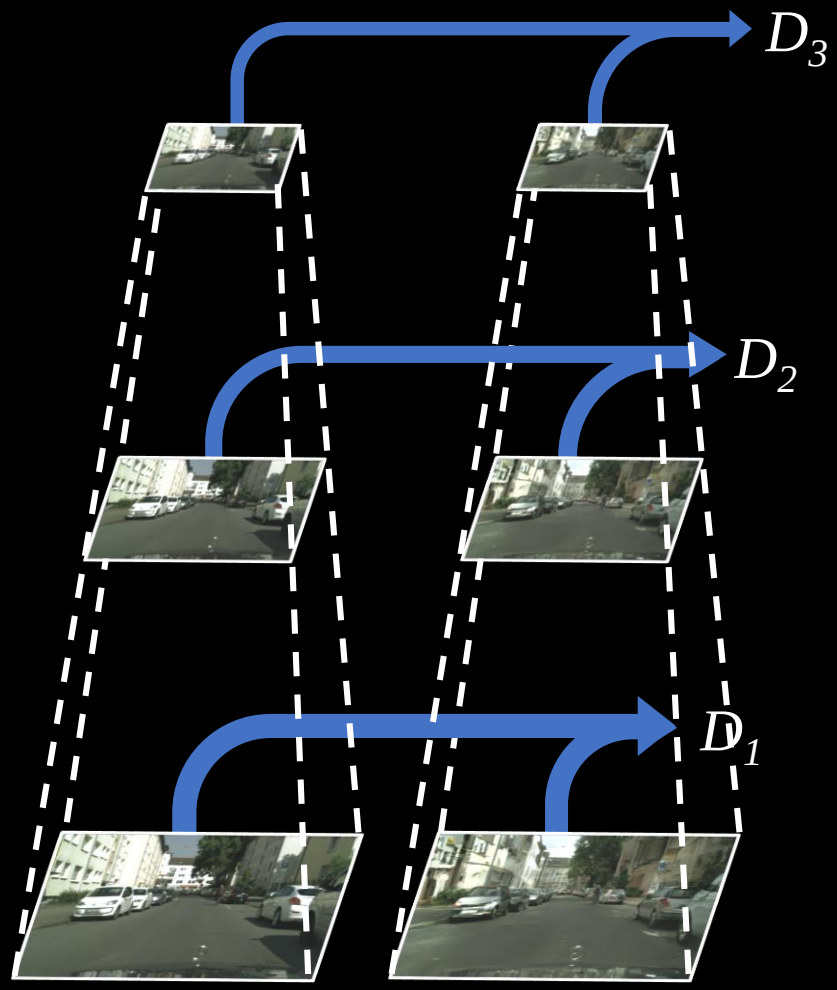


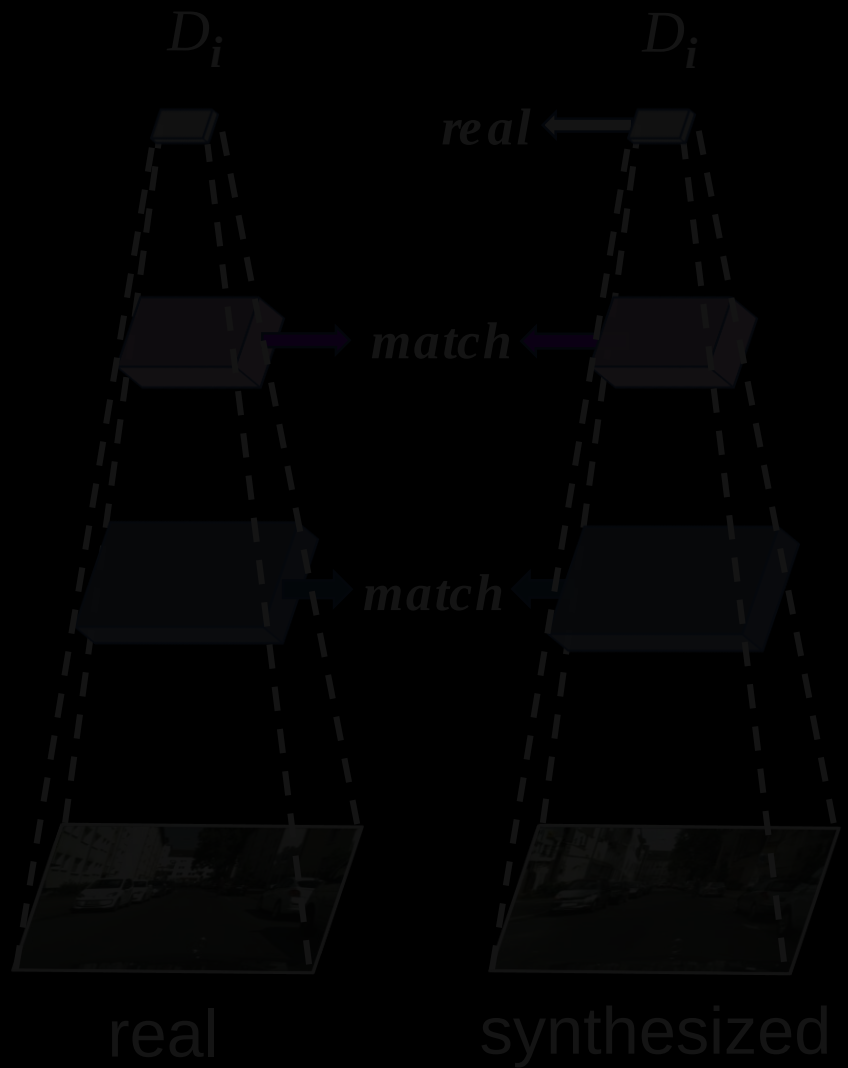
Coarse-to-fine Generator

Multi-scale Discriminators

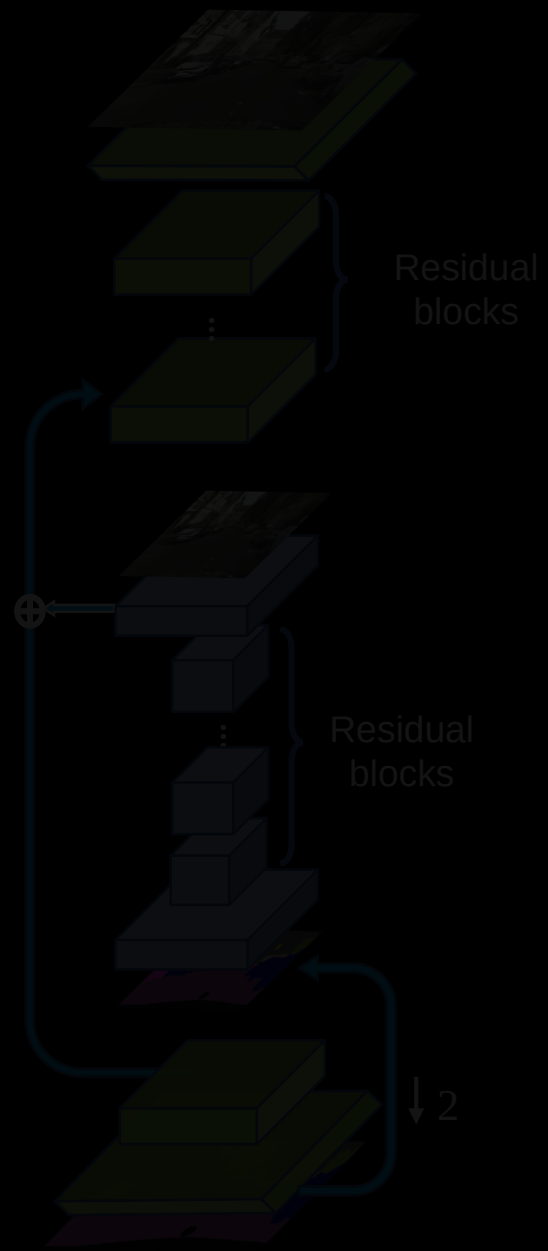
Robust Objective



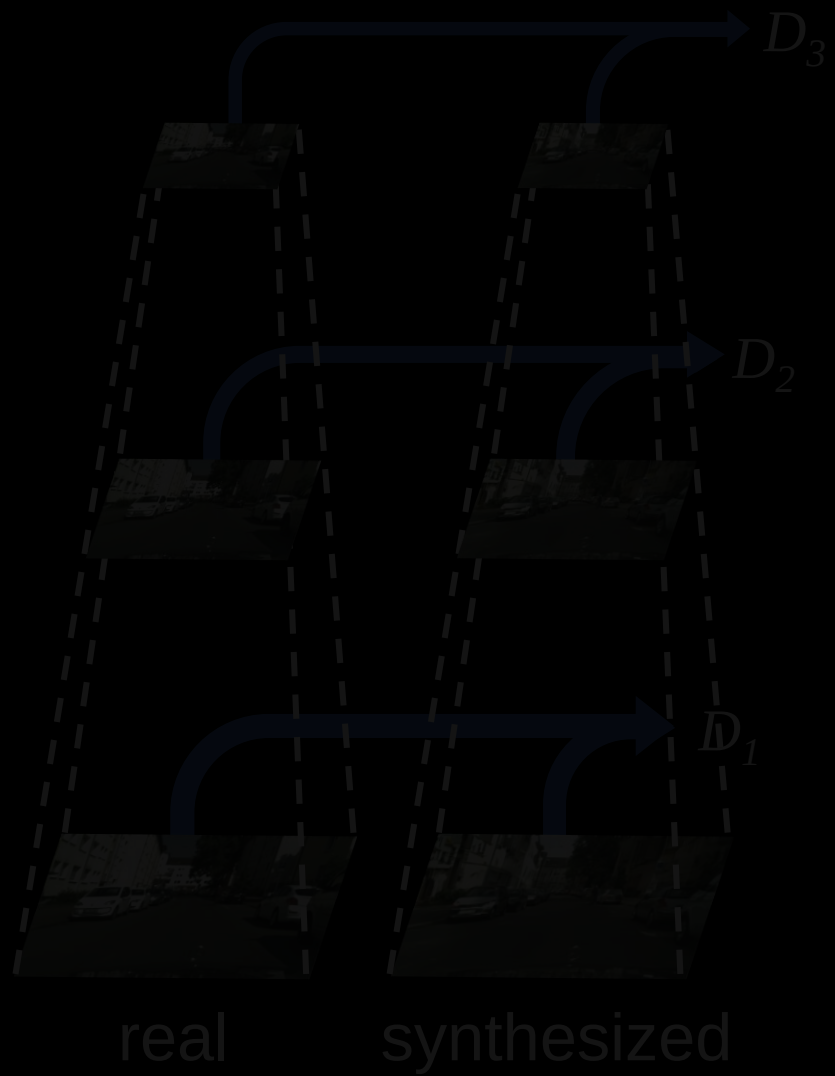




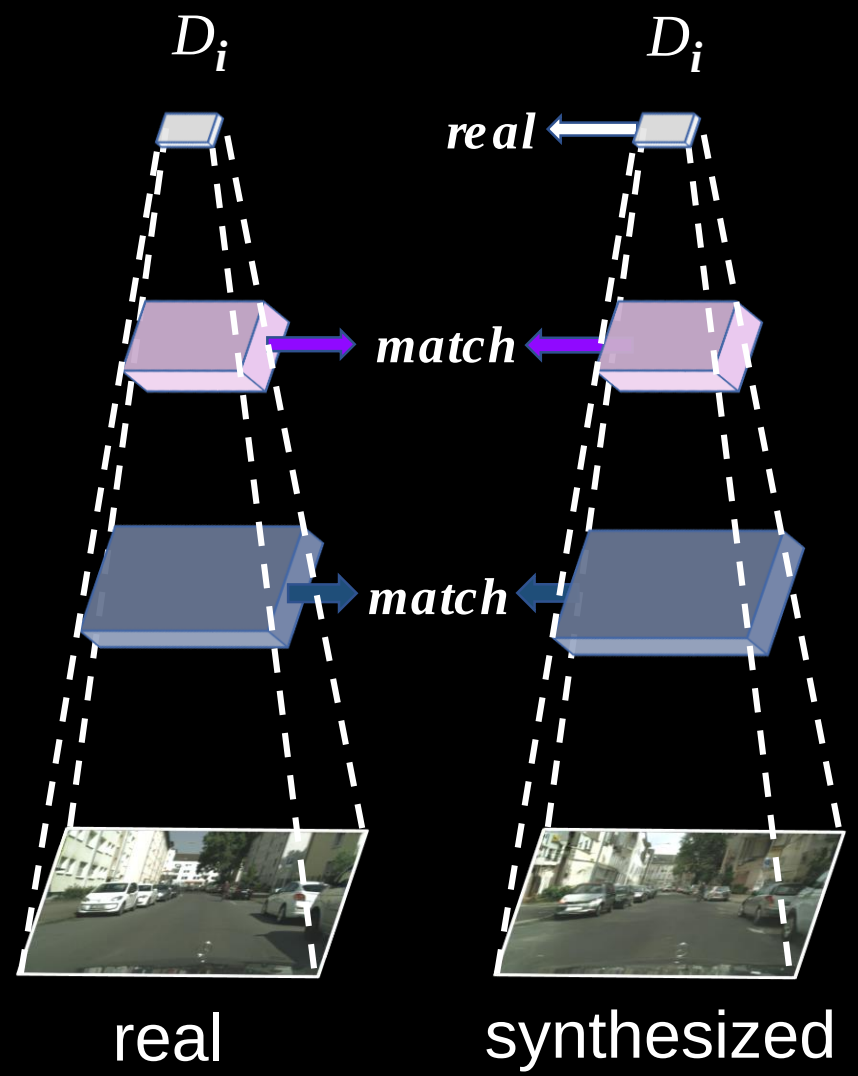
Coarse-to-fine Generator



Multi-scale Discriminators



Robust Objective



Our Method

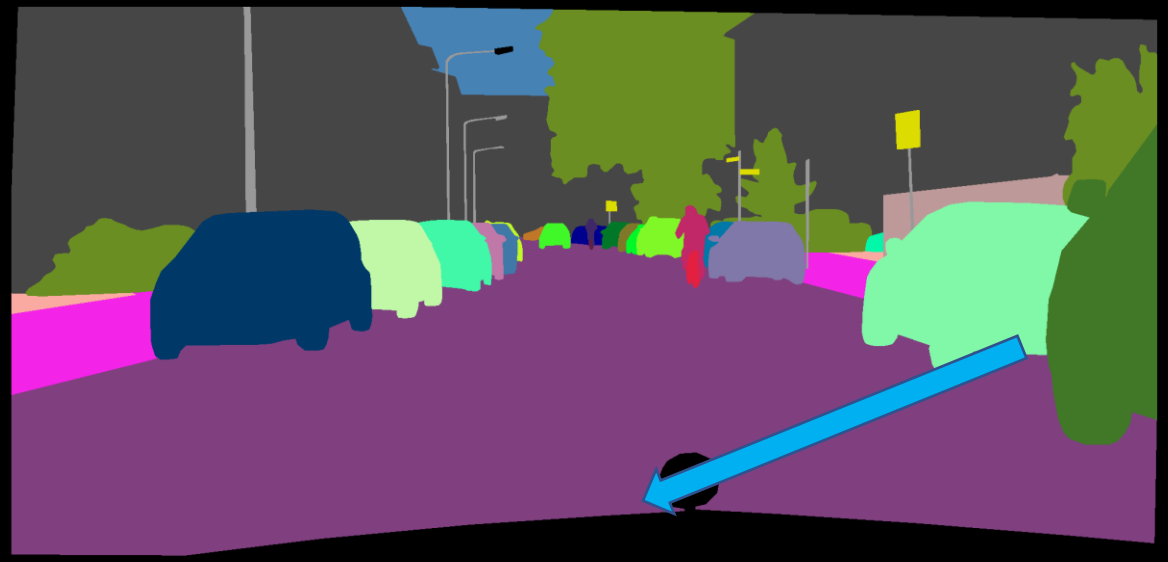
- Extending to high resolution
- Using instance-level segmentation maps
  - Boundary improvement
  - Multi-modal results using feature embedding

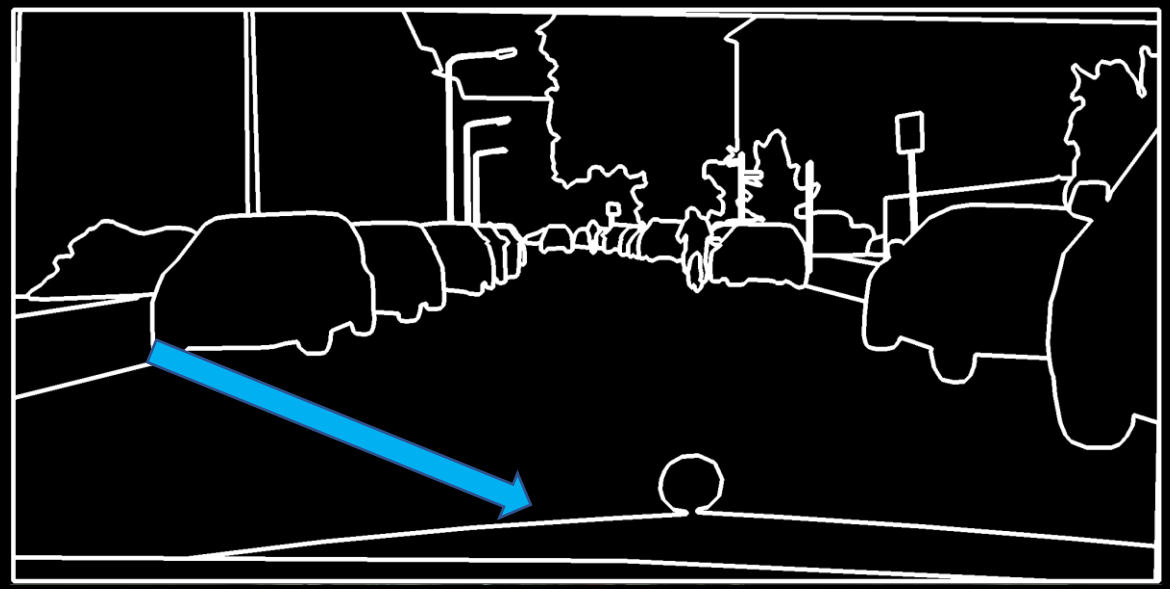
Our Method

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

Our Method

- Boundary improvement

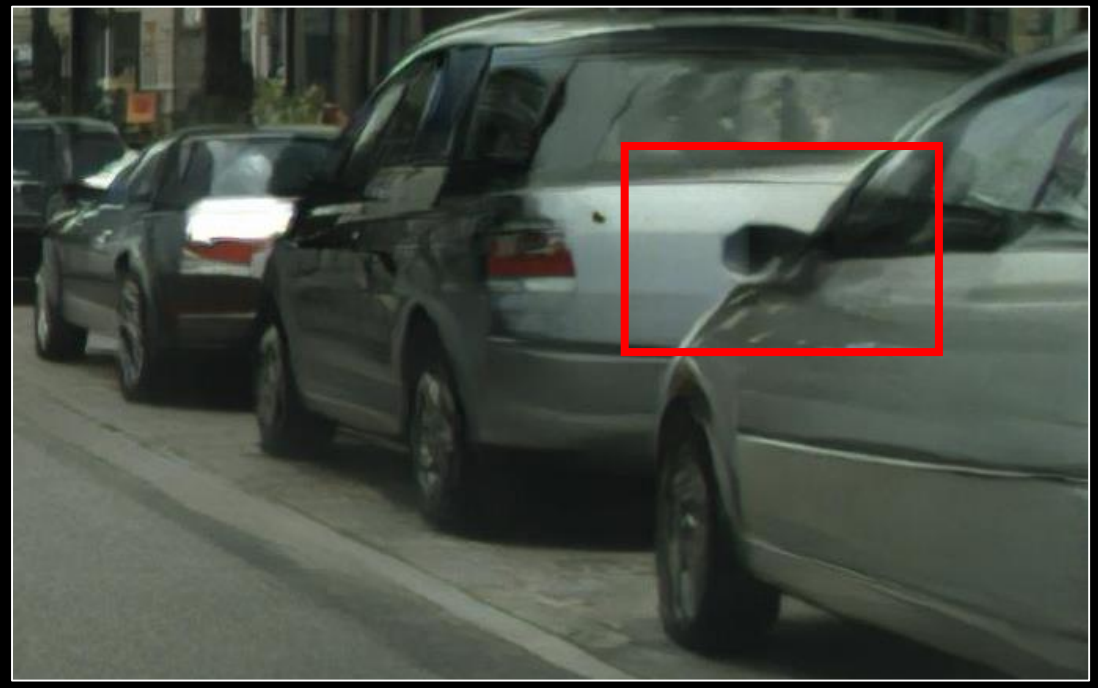




Our Method

- Boundary improvement

without instance maps



with instance maps

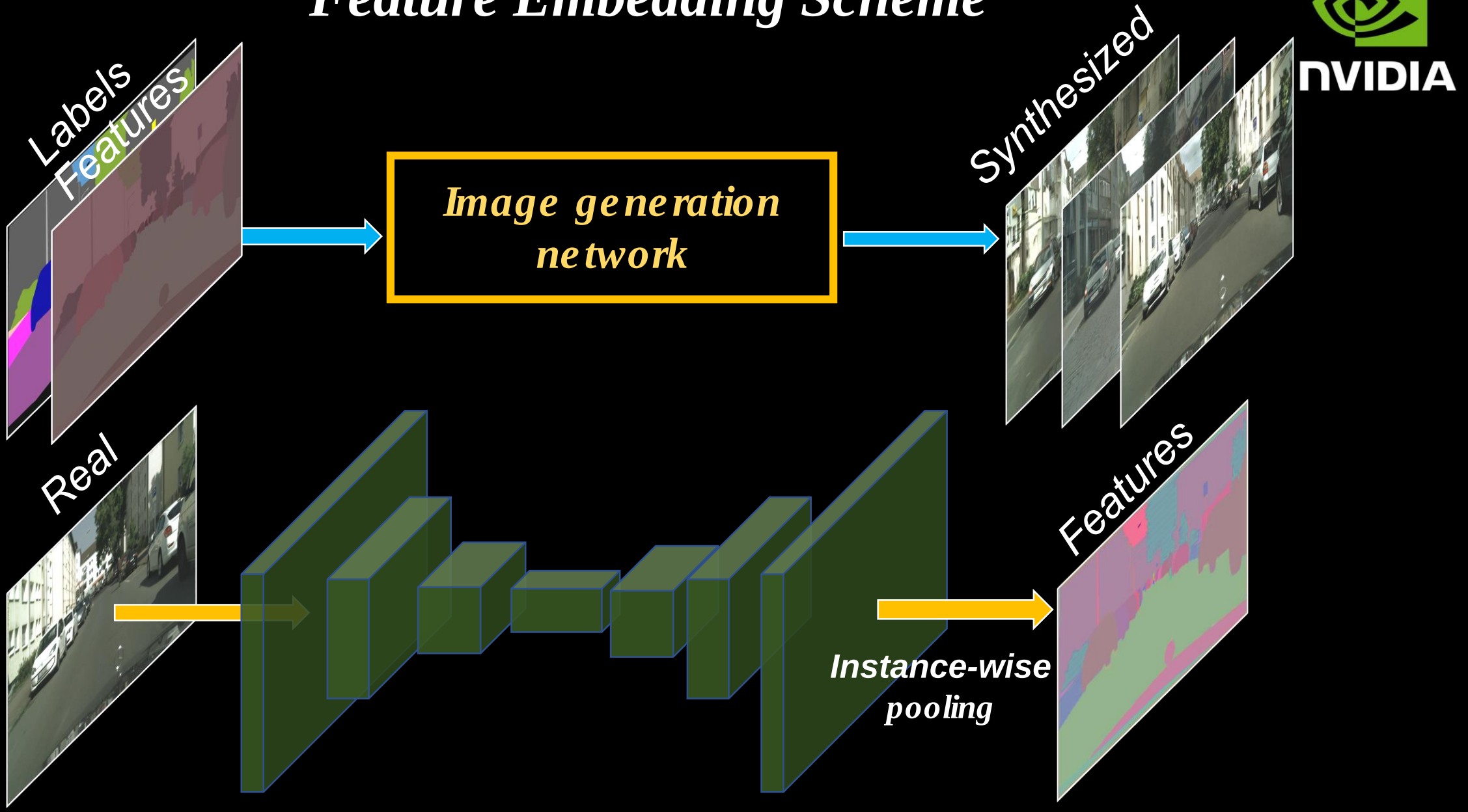
Our Method

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

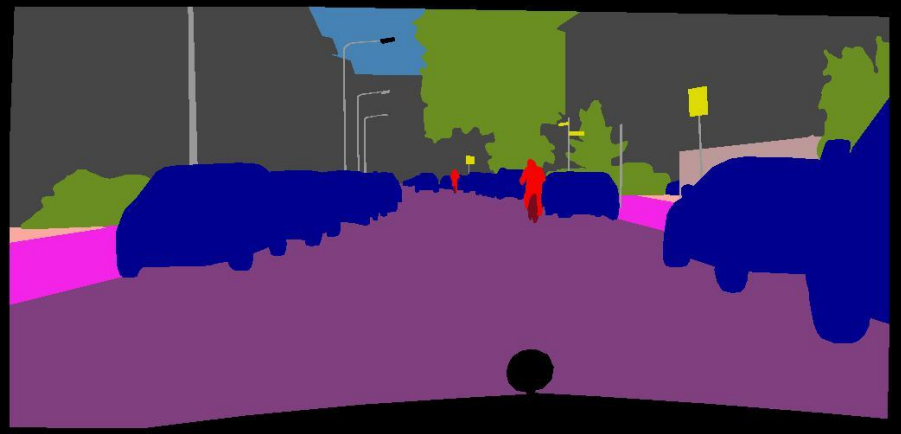
Our method

- Multi-modal results

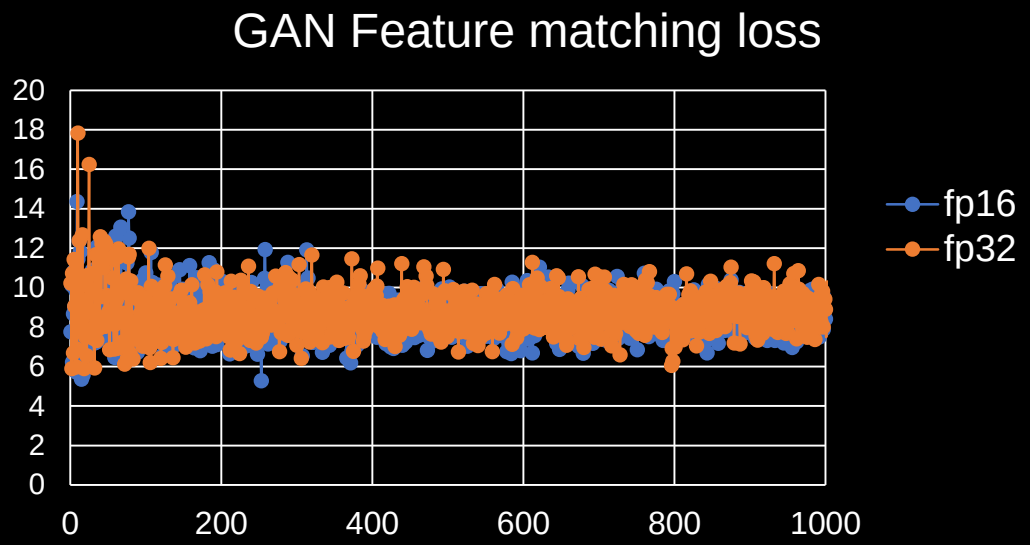
Feature Embedding Scheme



Mixed-Precision Training



input



fp16 (mixed-precision) result

fp32 result

Outline

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

Results

- Comparisons with
  - pix2pix [Isola et al. 2017]
  - CRN [Chen and Koltun 2017]
- Datasets
  - Cityscapes
  - NYU
  - ADE20K
  - Helen Face
  - CelebA-HQ

Results

- Quantitative comparisons (Cityscapes)
  - Semantic segmentation scores

	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

  - Subjective user study scores

	pix2pix [21]	CRN [5]
Ours	93.8%	86.2%

Results

- Qualitative comparisons

Semantic Map

pix2pix

CRN

Ours

Results

- NYU

pix2pix

CRN

Ours

Results

- NYU

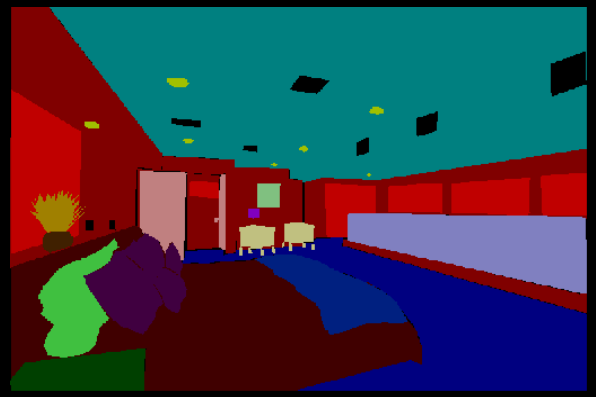
pix2pix

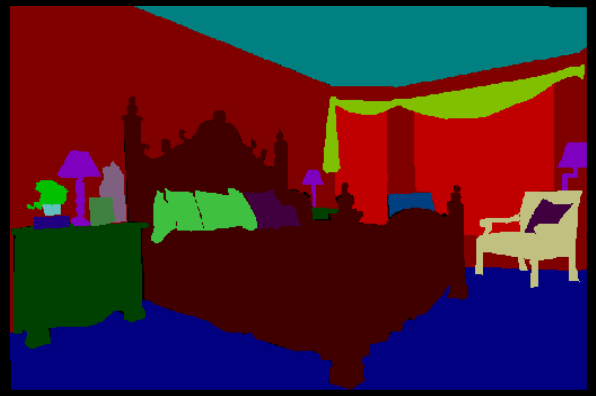
CRN

Ours

Results

- ADE20K



Labels | Ours | Ground truth

Results

- Helen Face: multi-modal results




Changing skin colors

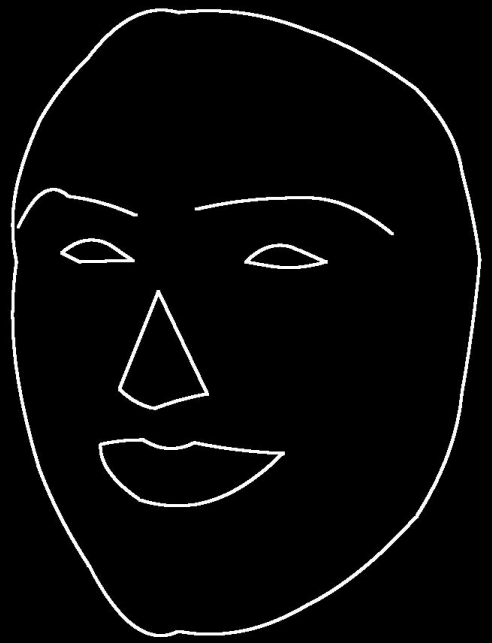
Results

- Helen Face : multi-modal results

Adding eyebrows / beards

Results

- CelebA-HQ



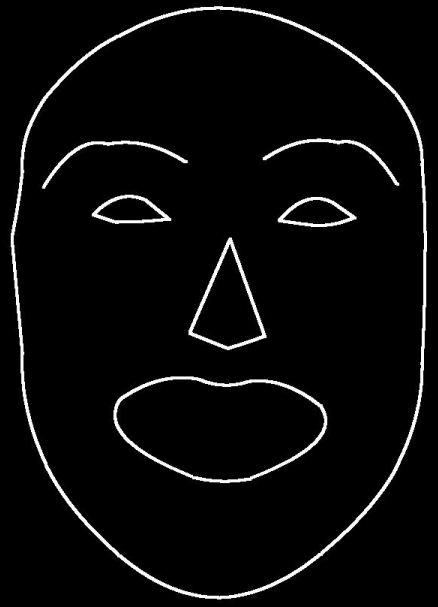
Edges

Synthesized

Ground truth

Results

- CelebA-HQ



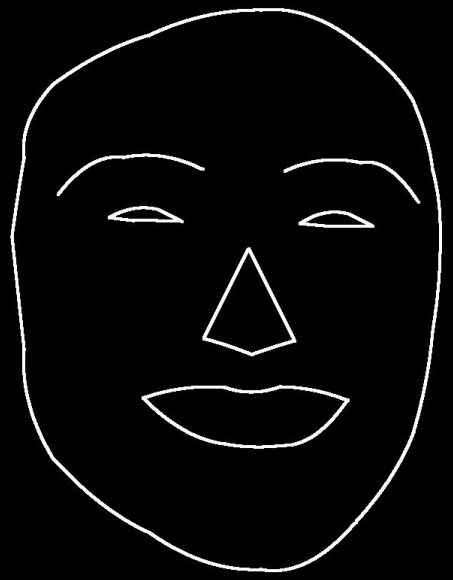
Edges

Synthesized

Ground truth

Results

- CelebA-HQ



Edges

Synthesized

Ground truth

Results

- CelebA-HQ

Edges

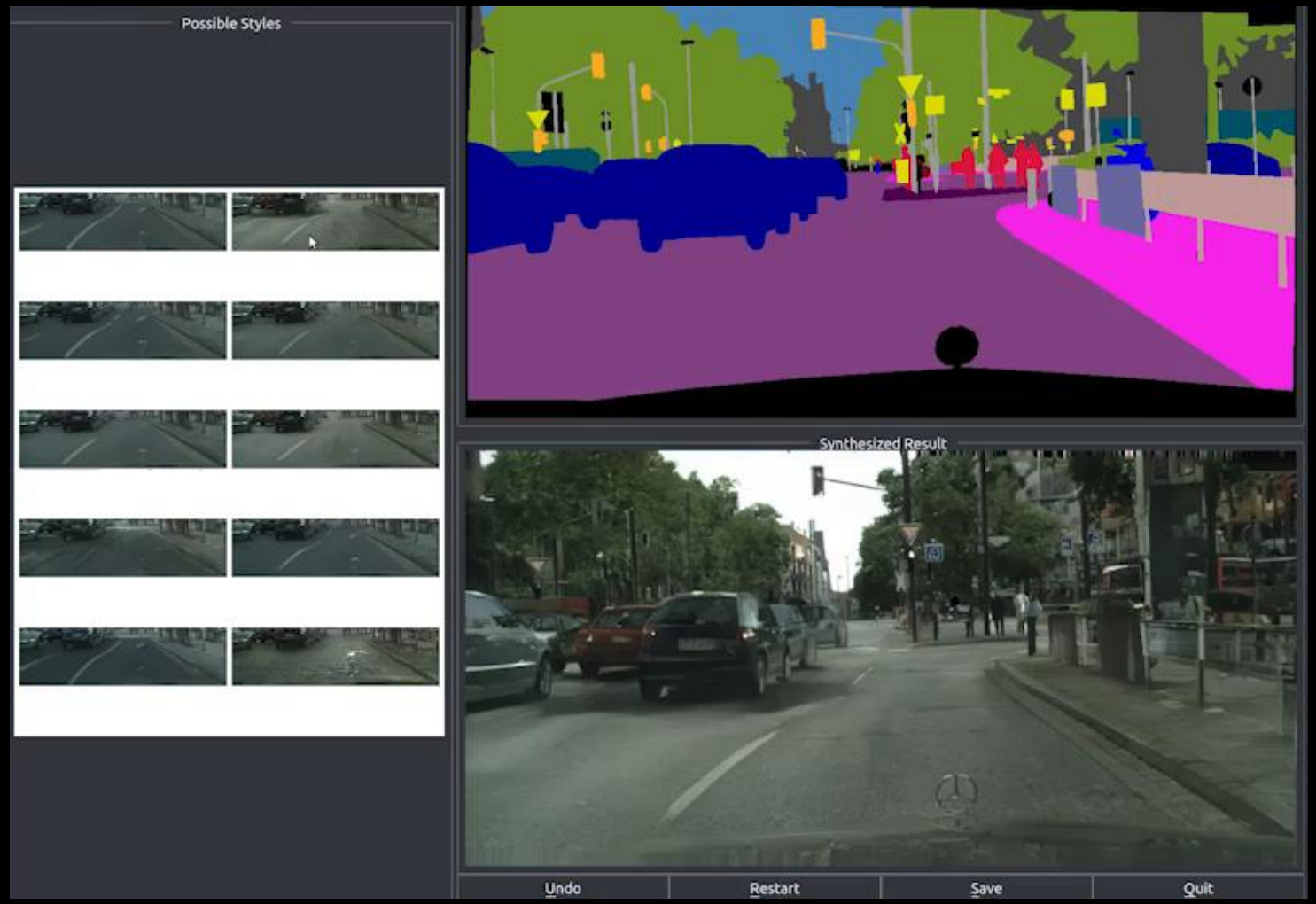
Synthesized

Ground truth

Outline

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

Applications: style changing

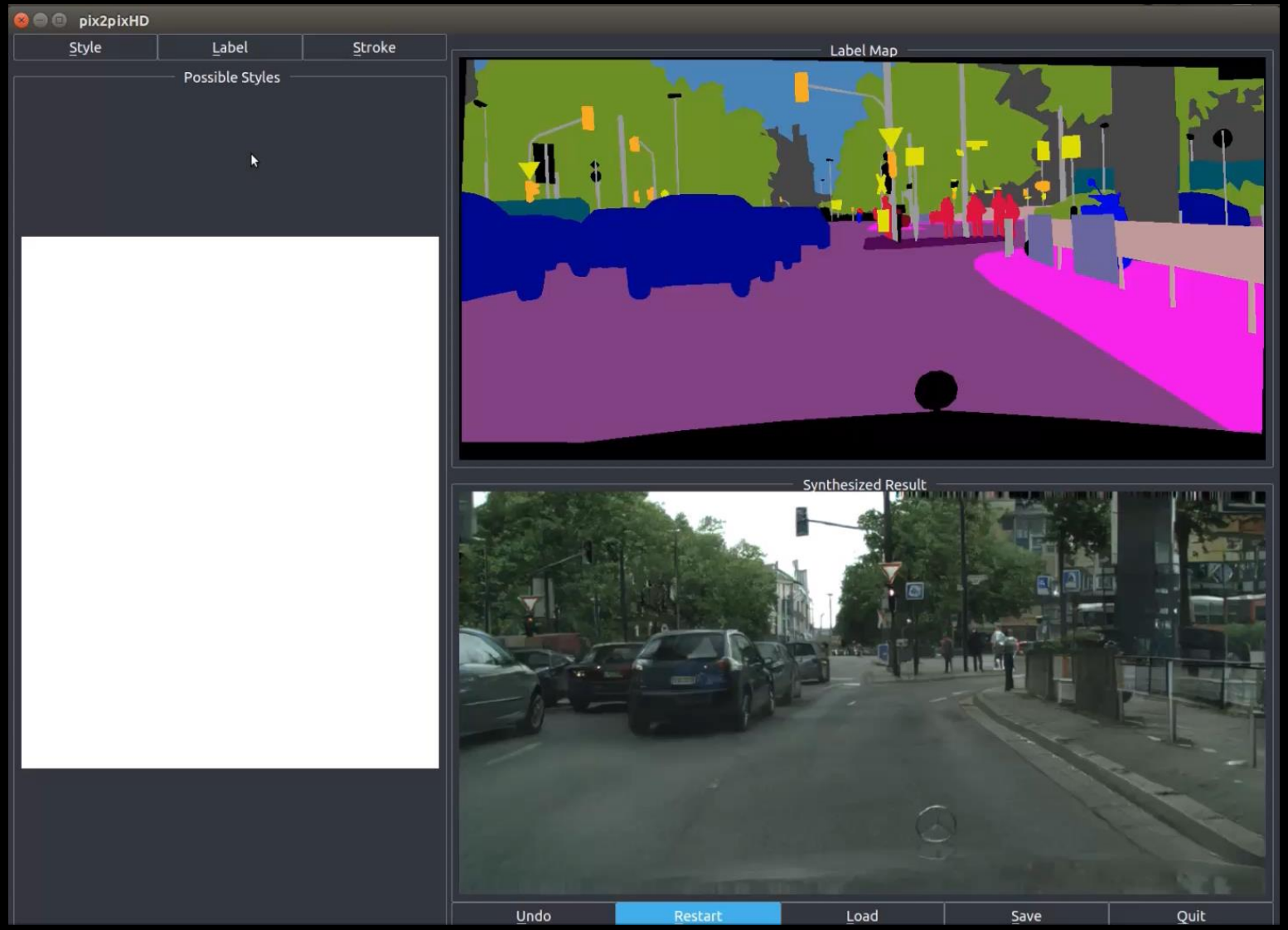


Applications: style changing

Applications: style changing

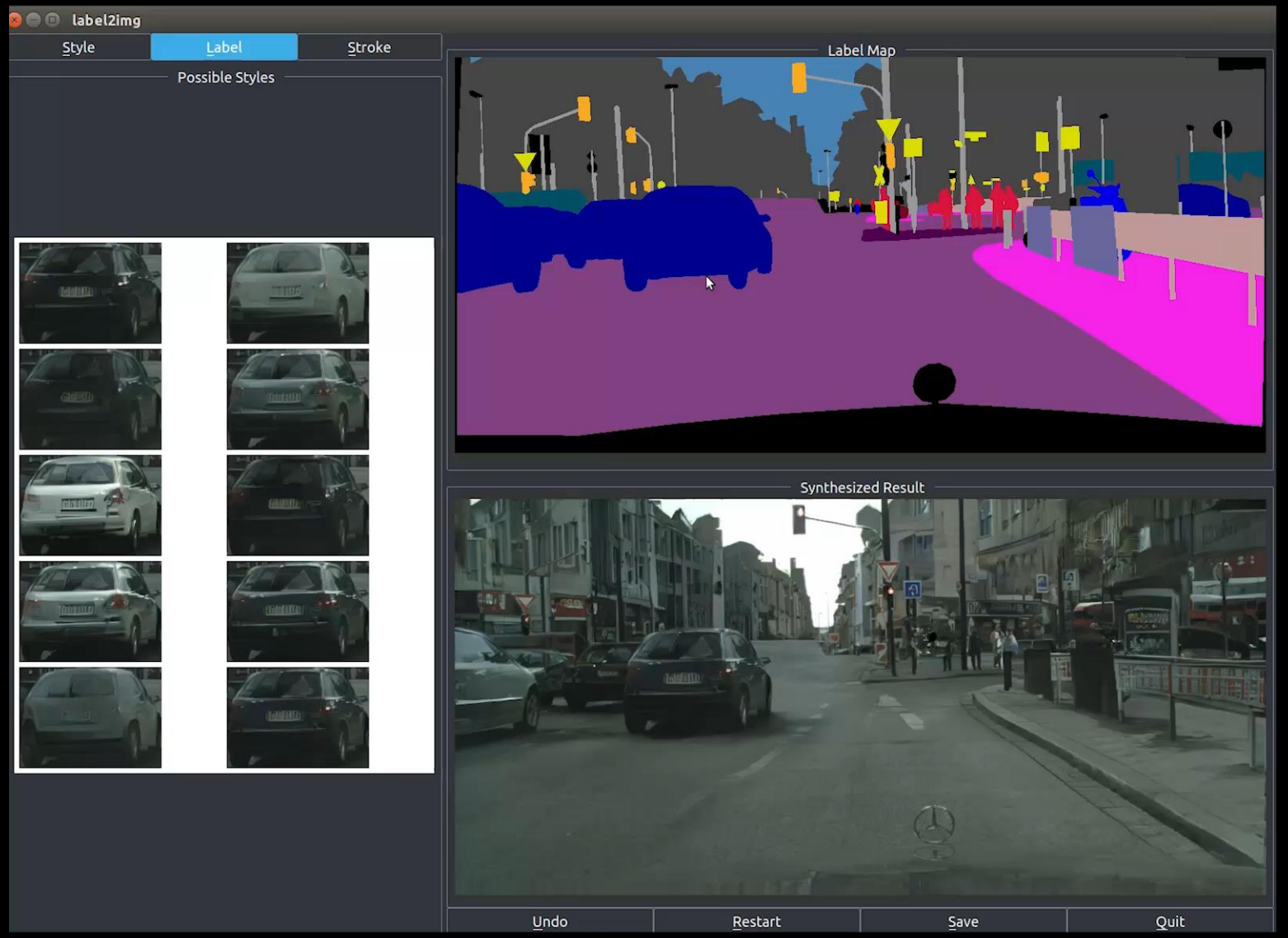
Applications: style changing

Applications: label changing

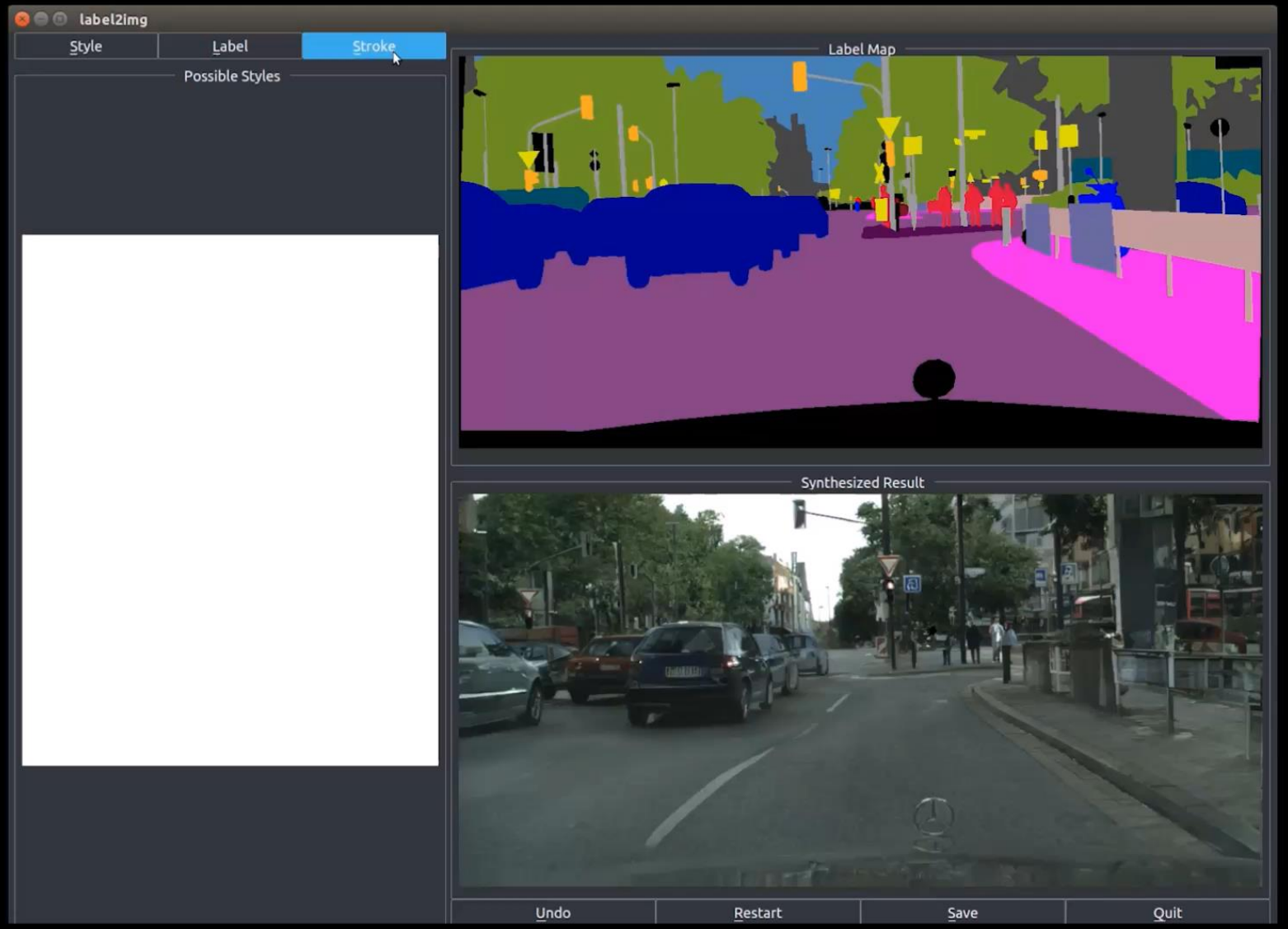


Applications: label changing

Applications: adding objects



Applications: adding strokes



Applications: adding strokes

Outline

- Introduction
- Related work
- Method
- Results
- Applications
- 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-model results

Acknowledgements

- We thank the following people for helpful comments
  - Taesung Park, Phillip Isola, Tinghui Zhou, Richard Zhang, Rafael Valle and Alexei A. Efros

Thank you!

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

Results

- Quantitative comparisons
  - Generator choice
    - Semantic segmentation scores

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

    - Subjective user study

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

Results

- Quantitative comparisons
  - Discriminator choice
    - Semantic segmentation scores

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

    - Subjective user study
      - 69.2 %

Training details

- LSGAN, Adam solver
- Feature embedding:
  - K=10 for K-means
  - 3-dimentional feature vector
- Training:
  - 1024x512 resolution: 12G memory GPU
  - 2048x1024 resolution:
    - FP32: NVIDIA Quadro M6000 GPU (24G)
    - FP16: NVIDIA Volta V100 GPU (16G)