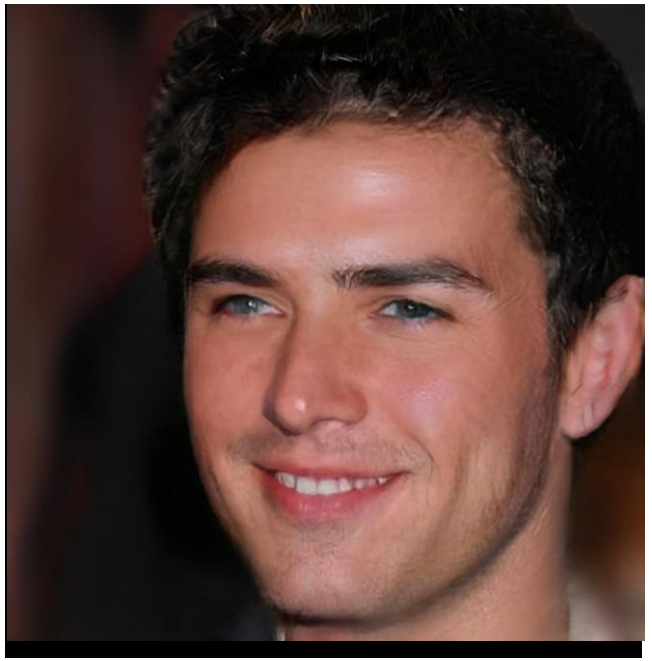
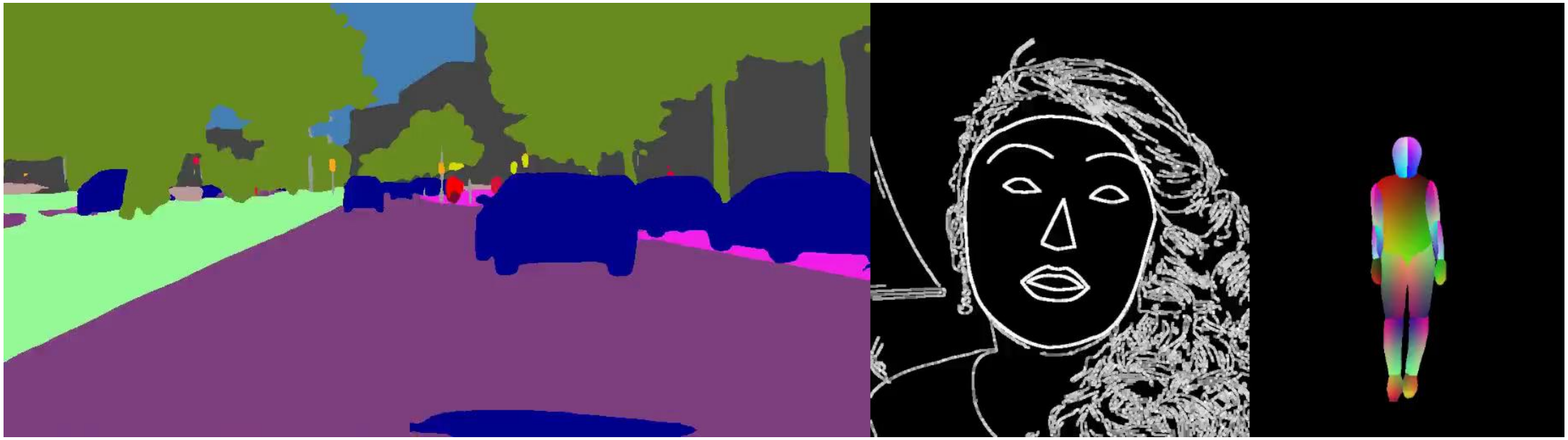
Video-to-Video Translation

Ting-Chun Wang
NVIDIA

Image-to-Image Translation



Video-to-Video Translation



Motivation

- AI-based rendering

Traditional graphics

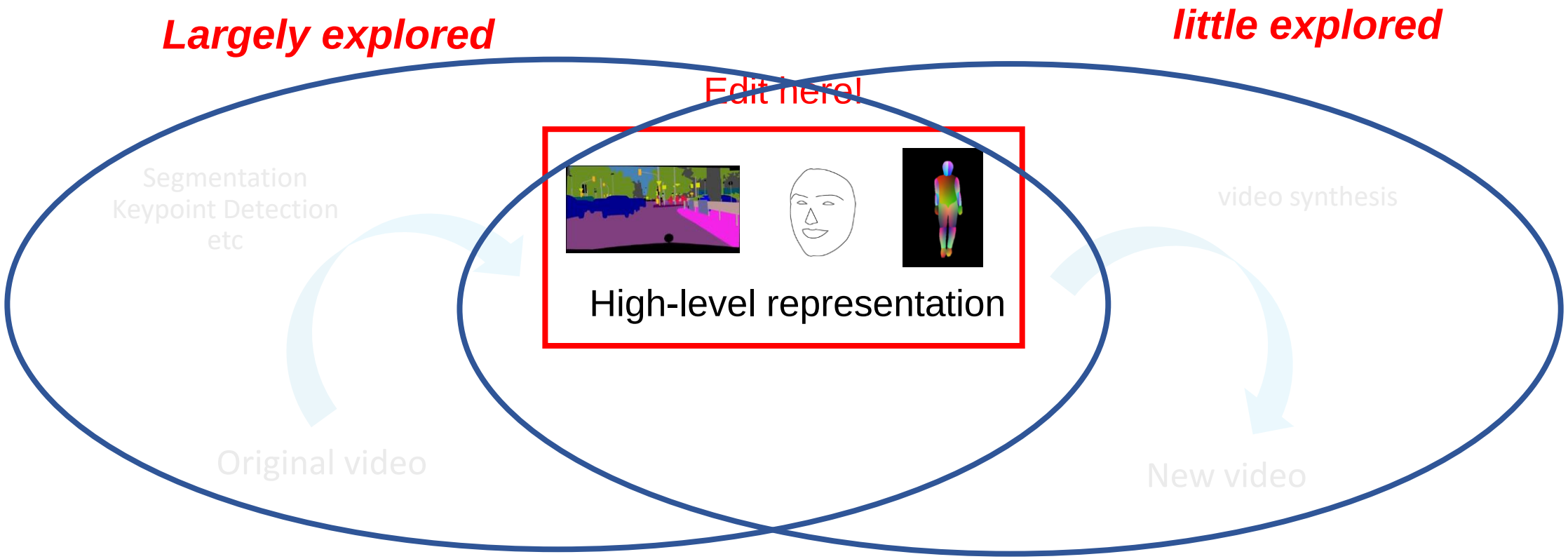
Geometry, texture, lighting

Machine learning graphics

Data

Motivation

- High-level semantic manipulation



Previous Work

Image translation

Flickering

pix2pixHD [2018], CRN [2017], pix2pix [2017]

Unconditional synthesis

Low-res, Short-term

MoCoGAN [2018], TGAN [2017], VGAN [2016]

Video style transfer

Simple smoothing

COVST [2017], ArtST [2016]

Video prediction

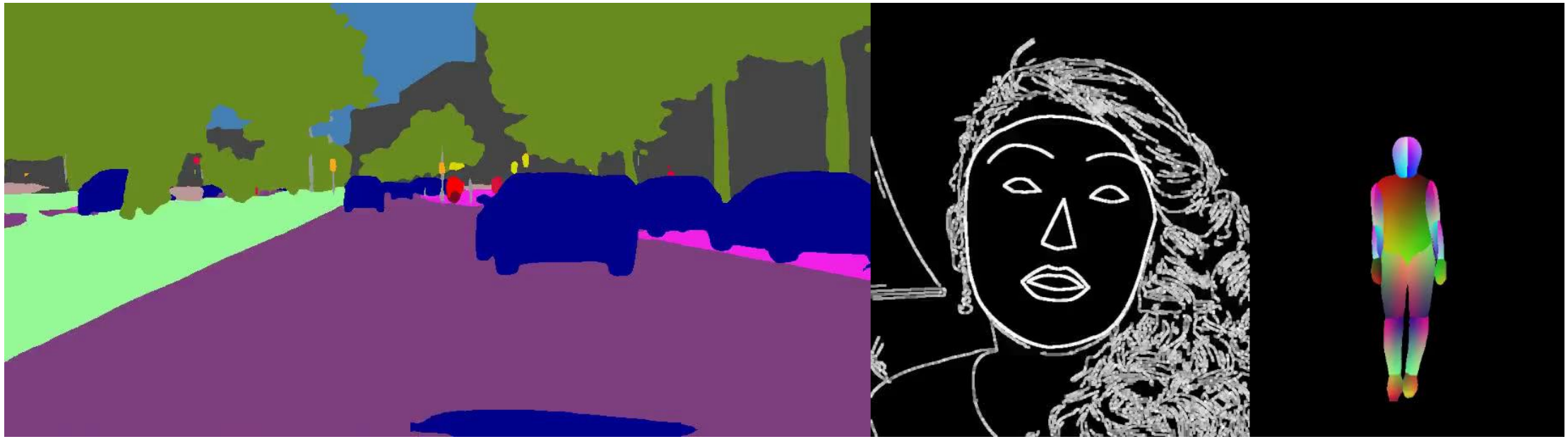
Same domain prediction

MCNet [2017], PredNet [2017]

Previous Work: Frame-by-Frame Result

Video-to-Video Synthesis (vid2vid)

T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz, B. Catanzaro,
“Video-to-Video Synthesis,” NeurIPS 2018.
https://github.com/NVIDIA/vid2vid

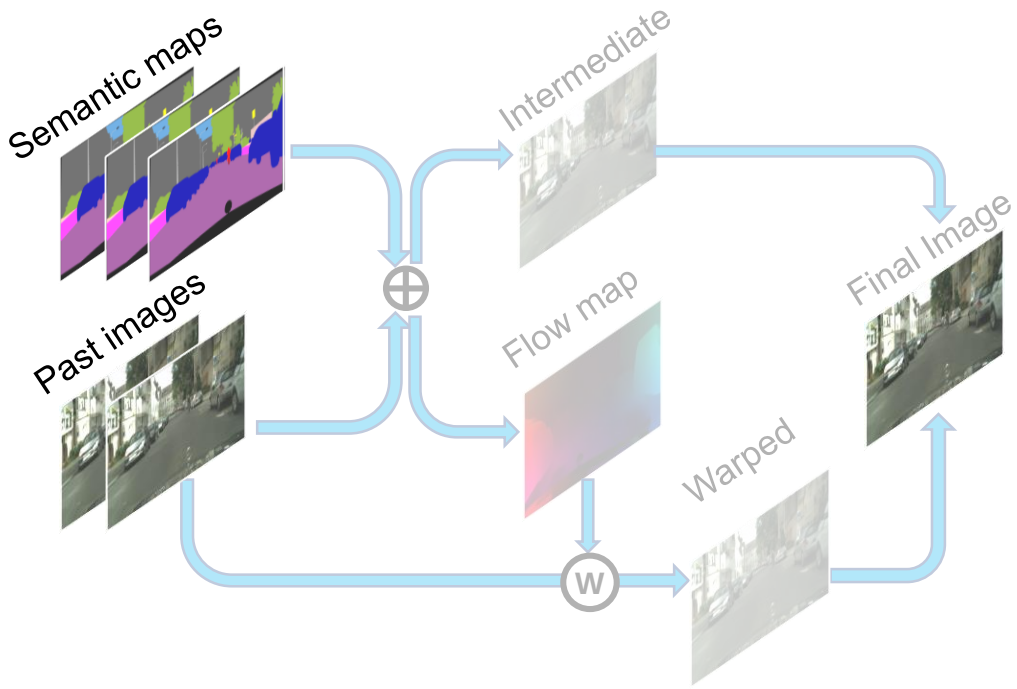


vid2vid

- Sequential generator
- Multi-scale temporal discriminator
- Spatio-temporal progressive training procedure

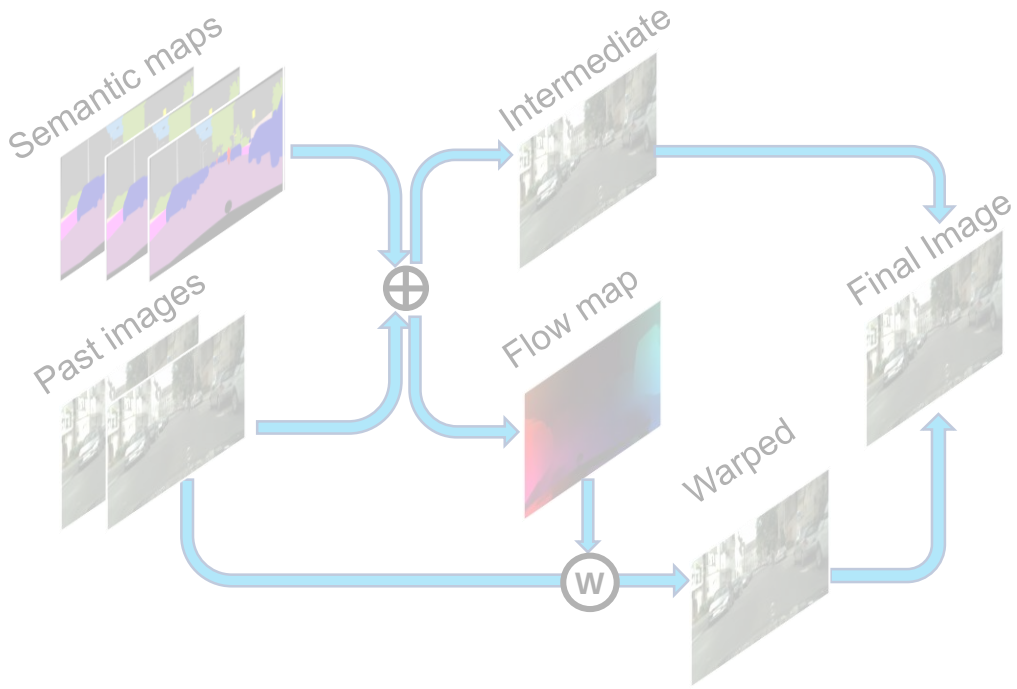
vid2vid

Sequential Generator



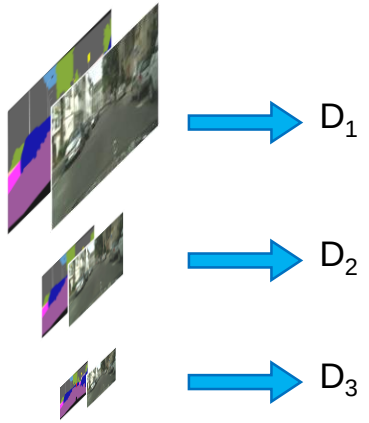
vid2vid

Sequential Generator

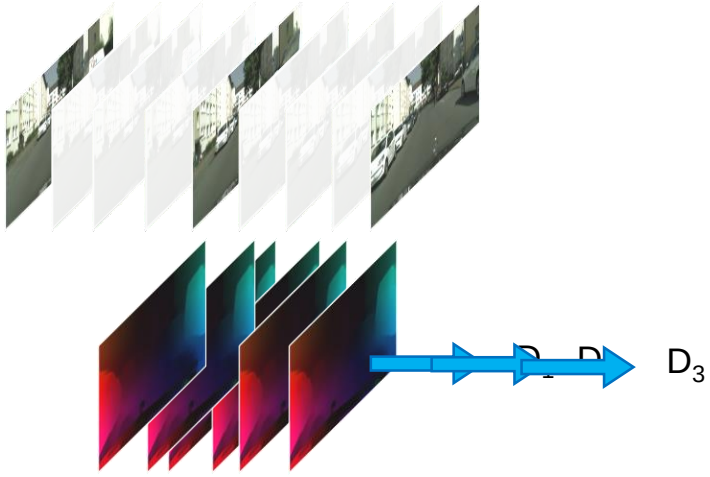


Multi-scale Discriminators

Image Discriminator



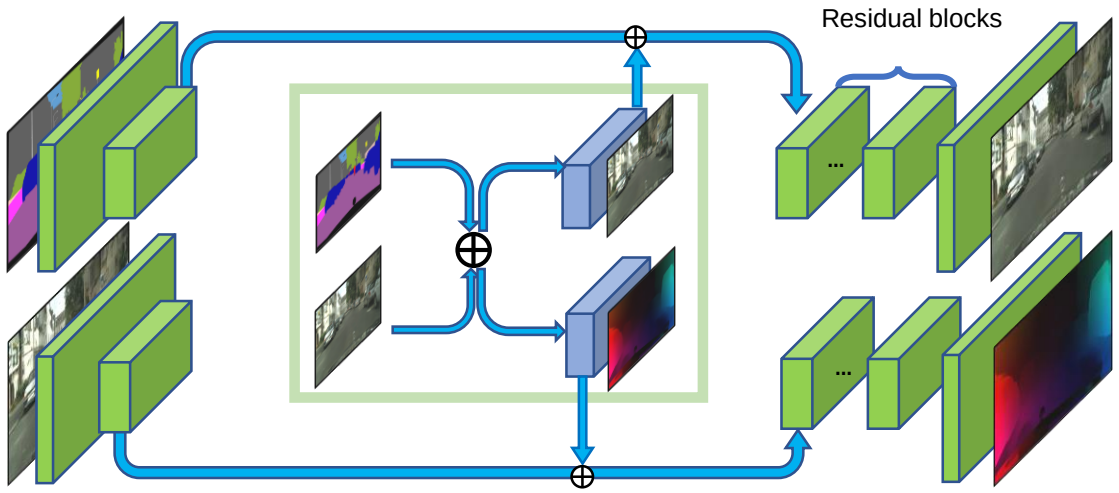
Video Discriminator



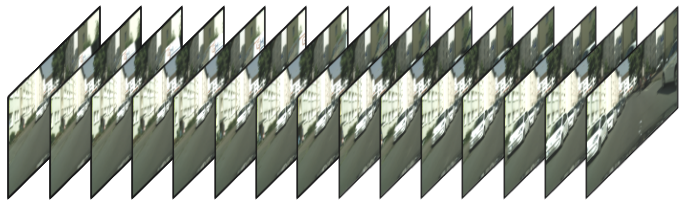
vid2vid

Spatio-temporally Progressive Training

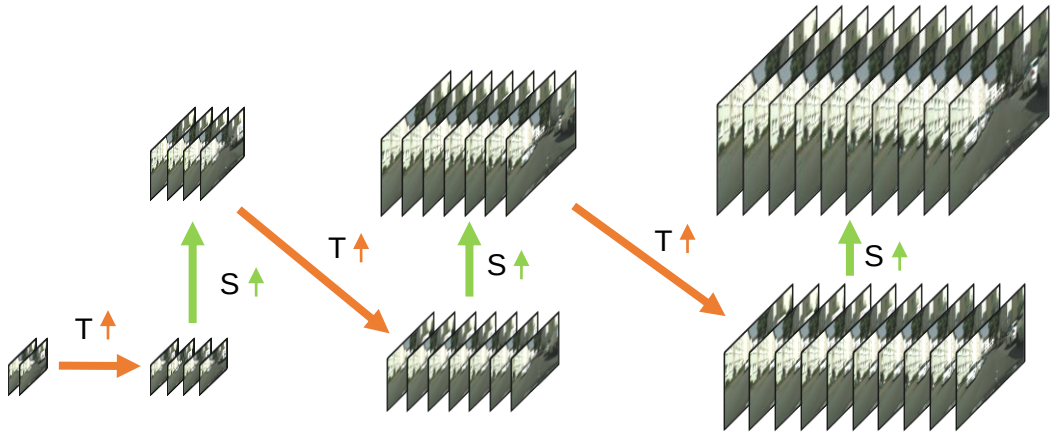
Spatially progressive



Temporally progressive



Alternating training



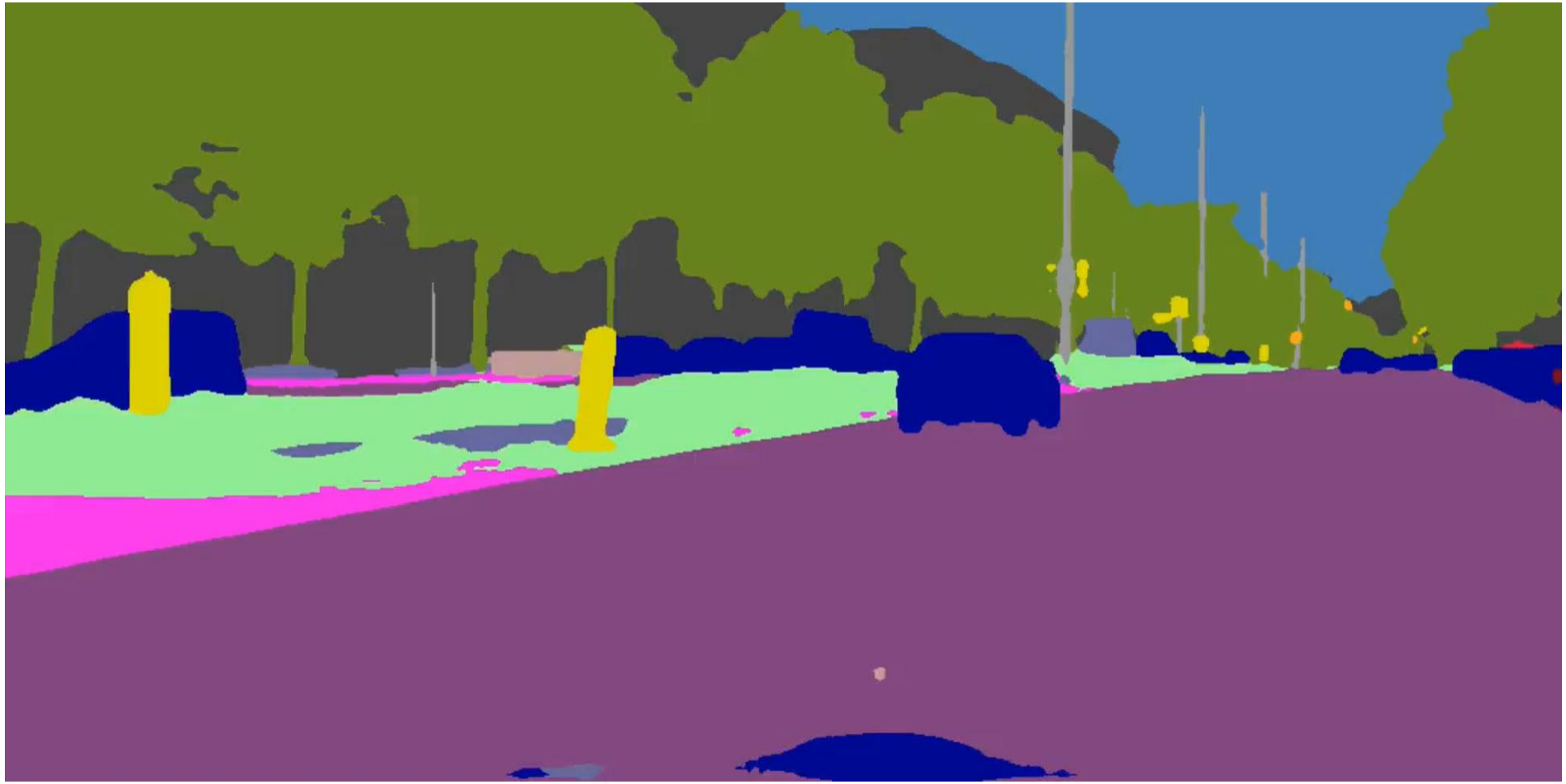
vid2vid Results

- Semantic → Street view scenes
- Edges → Human faces
- Poses → Human bodies

vid2vid Results

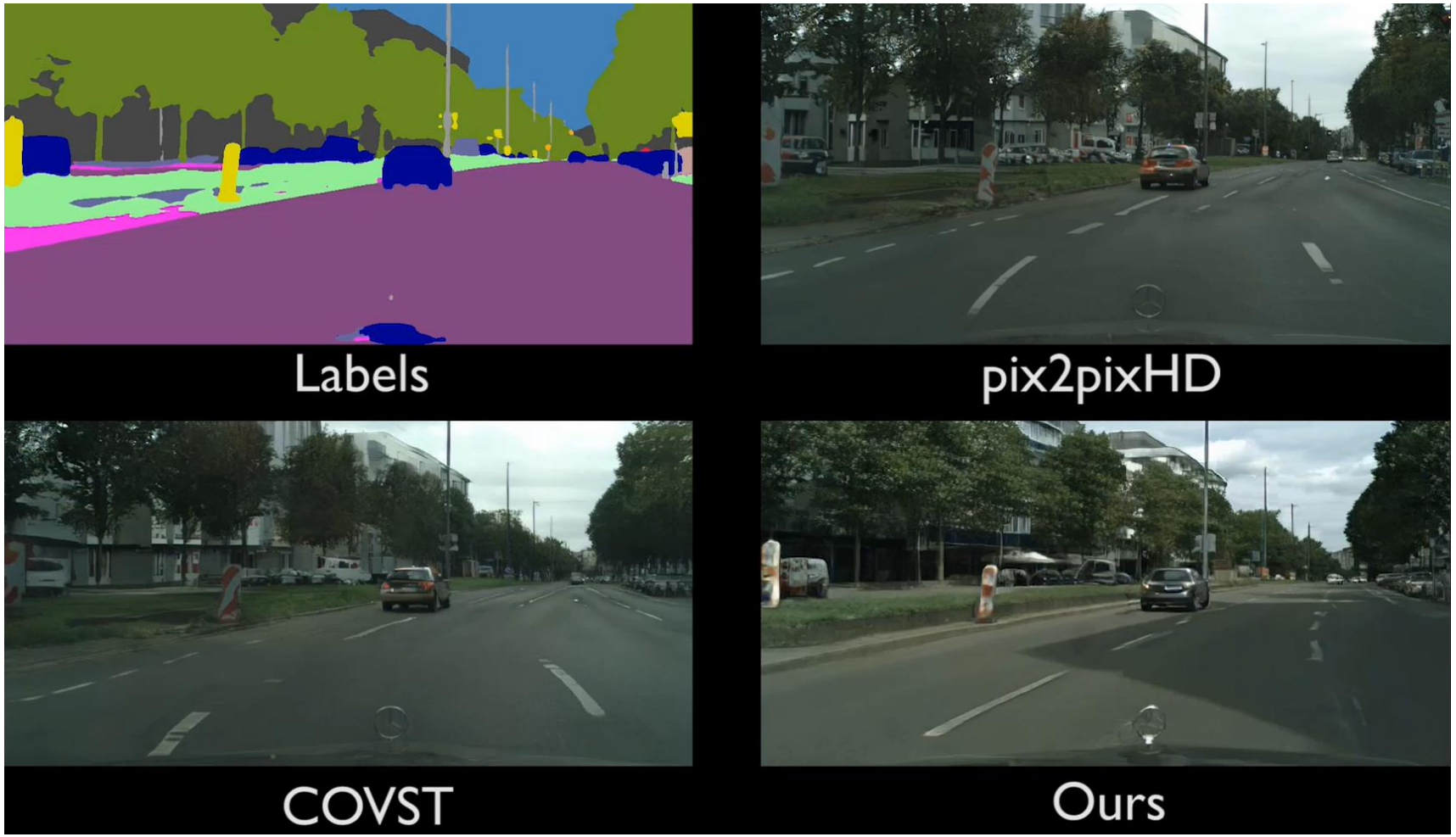
- Semantic → Street view scenes

Street View: Cityscapes



Street View: Cityscapes

Street View: Cityscapes



Street View: Boston

Street View: NYC

Results

- Edges → Human faces

Face Swapping (face → edge → face)

input edges output

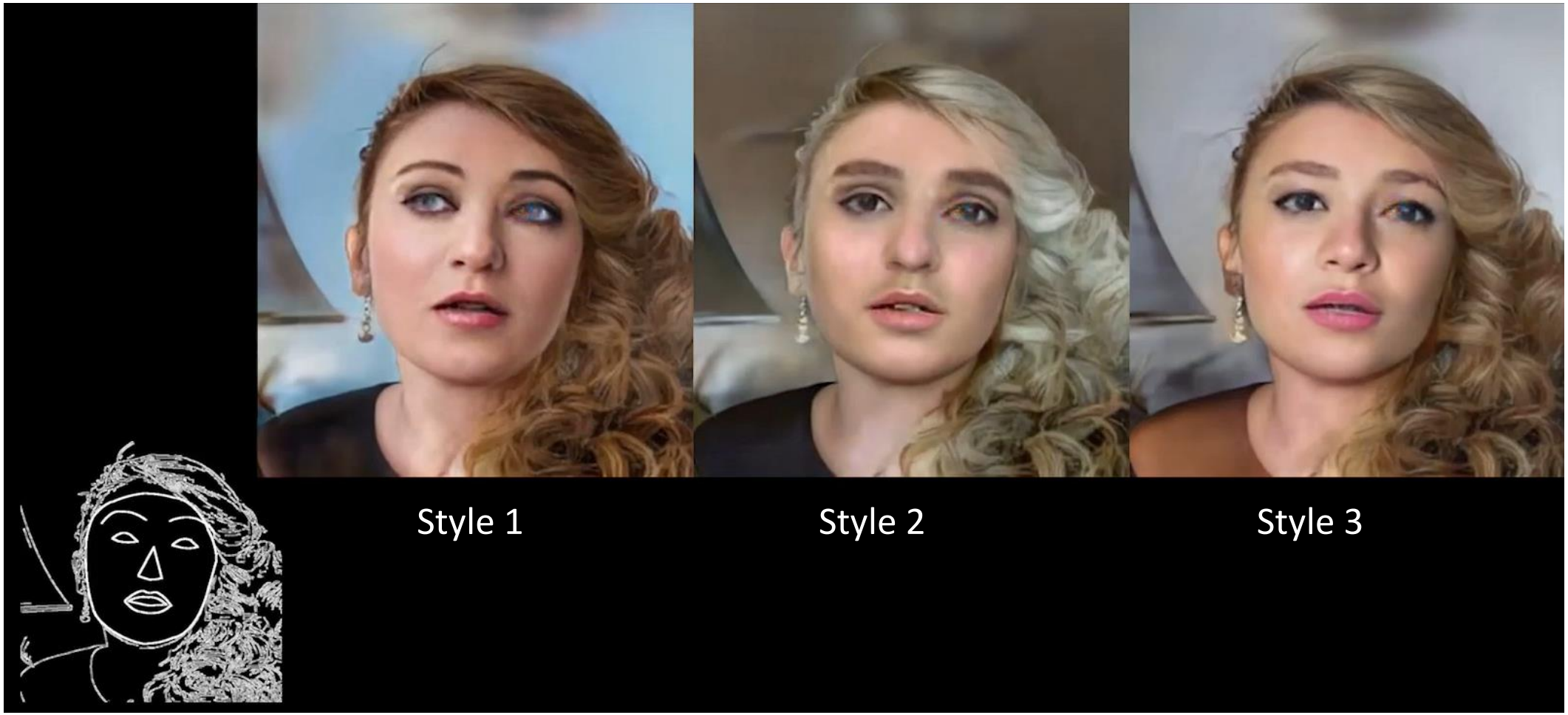
Face Swapping (slimmer face)

input (slimmed) edges (slimmed) output

Face Swapping (slimmer face)

input (slimmed) edges (slimmed) output

Multi-modal Edge → Face



Results

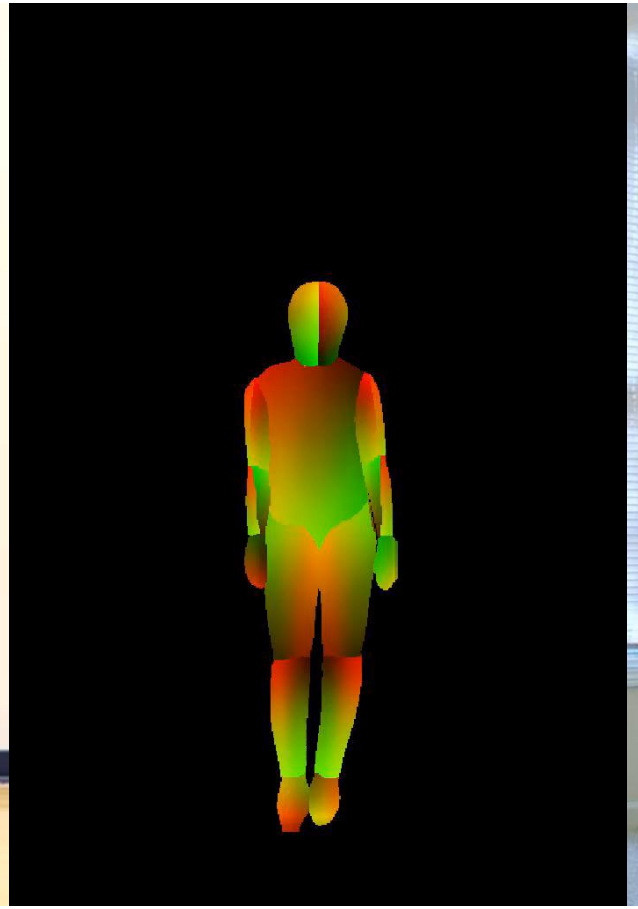
- Poses → Human bodies

Motion Transfer (body → pose → body)

input poses output

Motion Transfer (body → pose → body)

input



poses

output

Motion Transfer (body → pose → body)

input poses output

Motion Transfer (body → pose → body)

input poses output

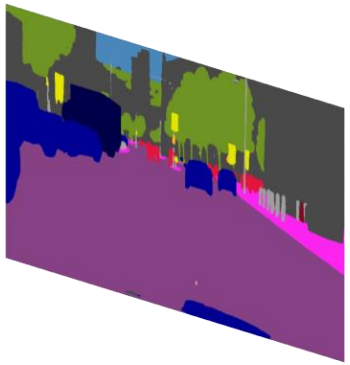
Motion Transfer

Motion Transfer

vid2vid Extensions: Interactive Graphics

User control

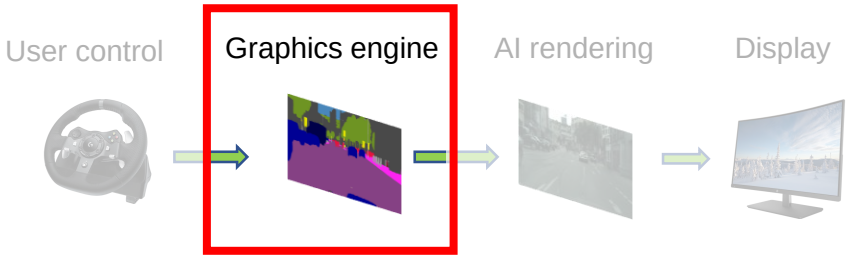
Graphics engine



AI rendering

Display

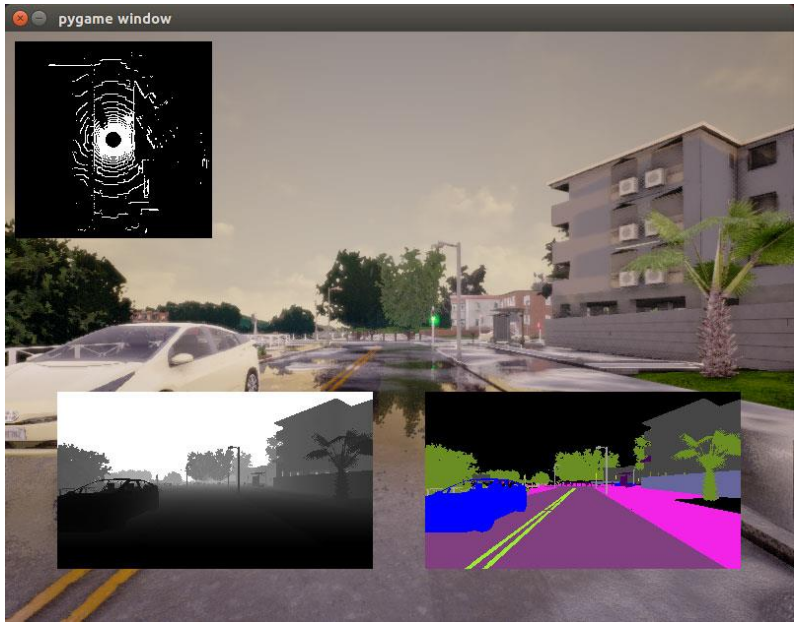
Graphics Engine: CARLA



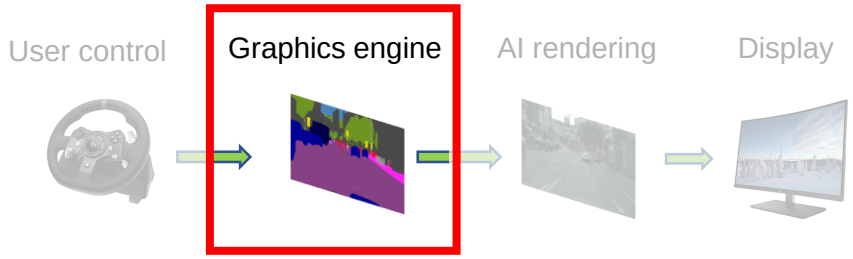
DATA

PROTOBUF messages

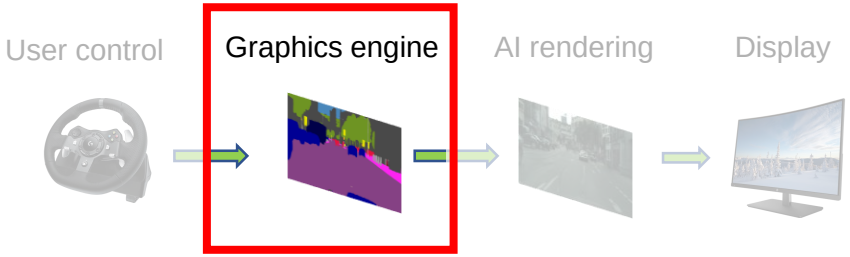
CONTROL

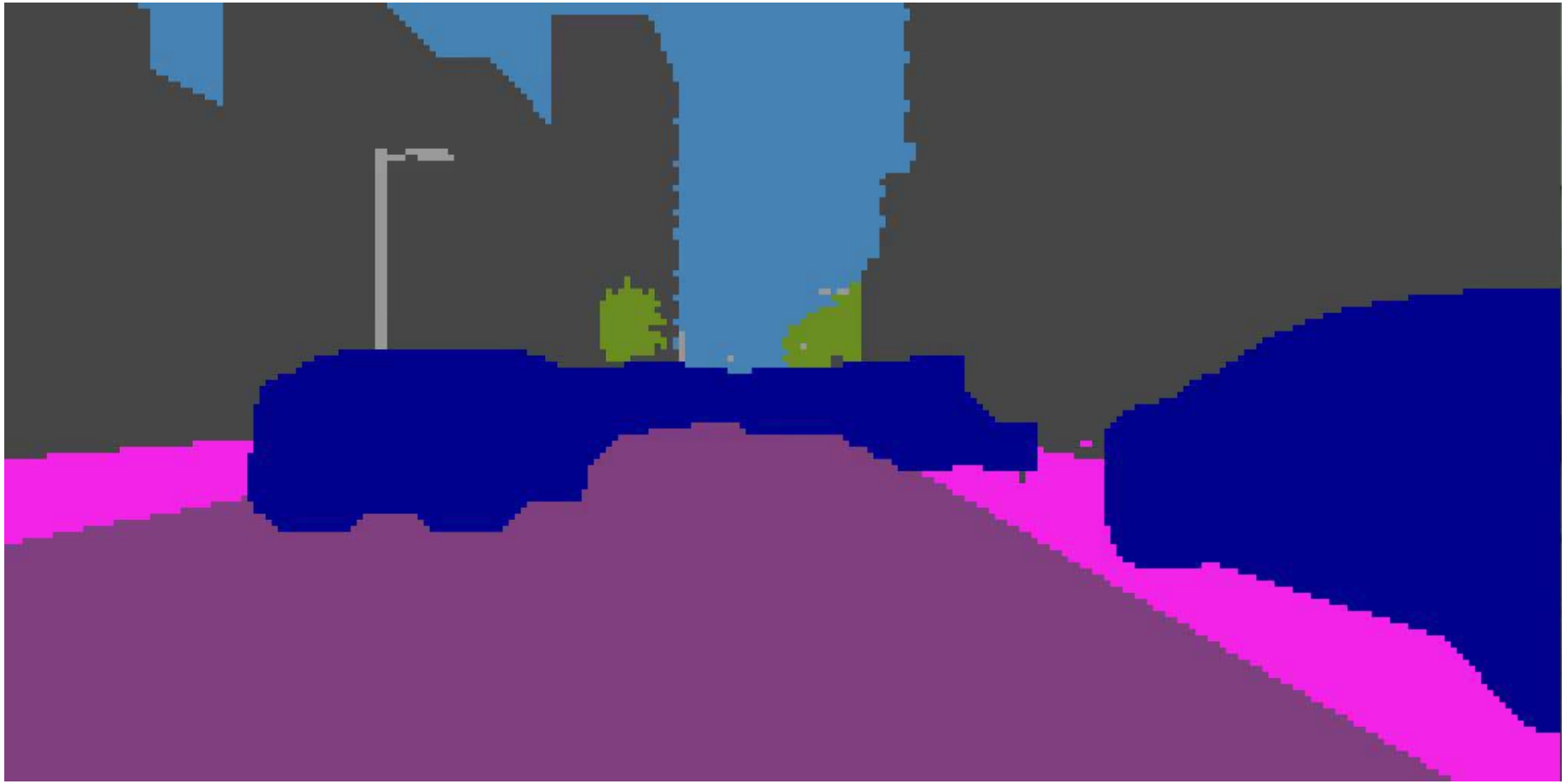


Original CARLA Sequence

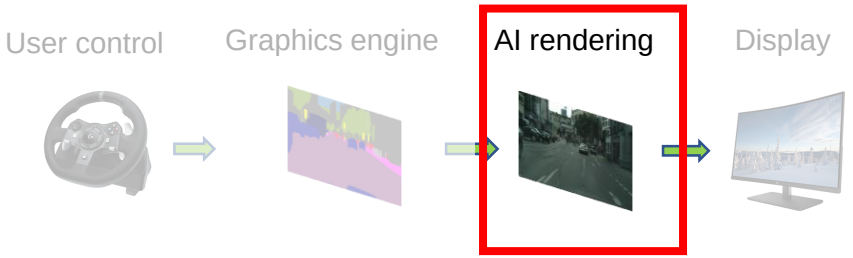


CARLA Semantic Maps

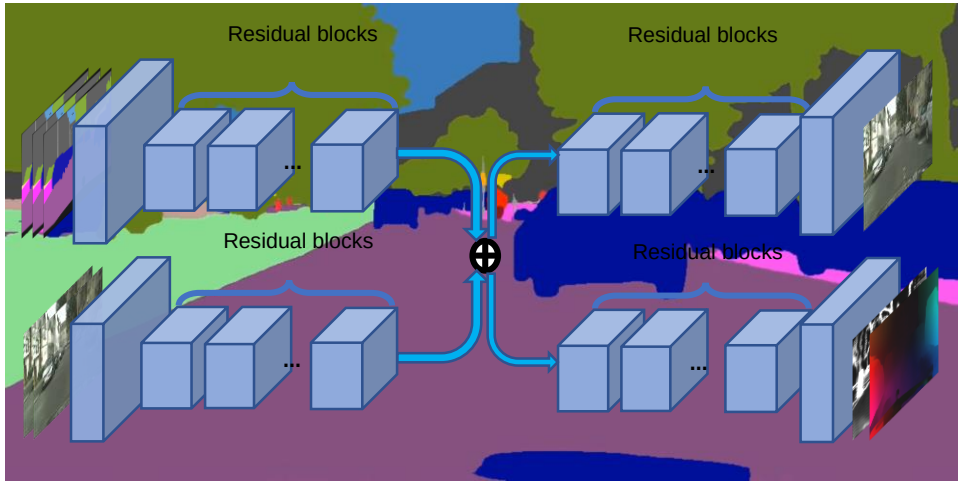


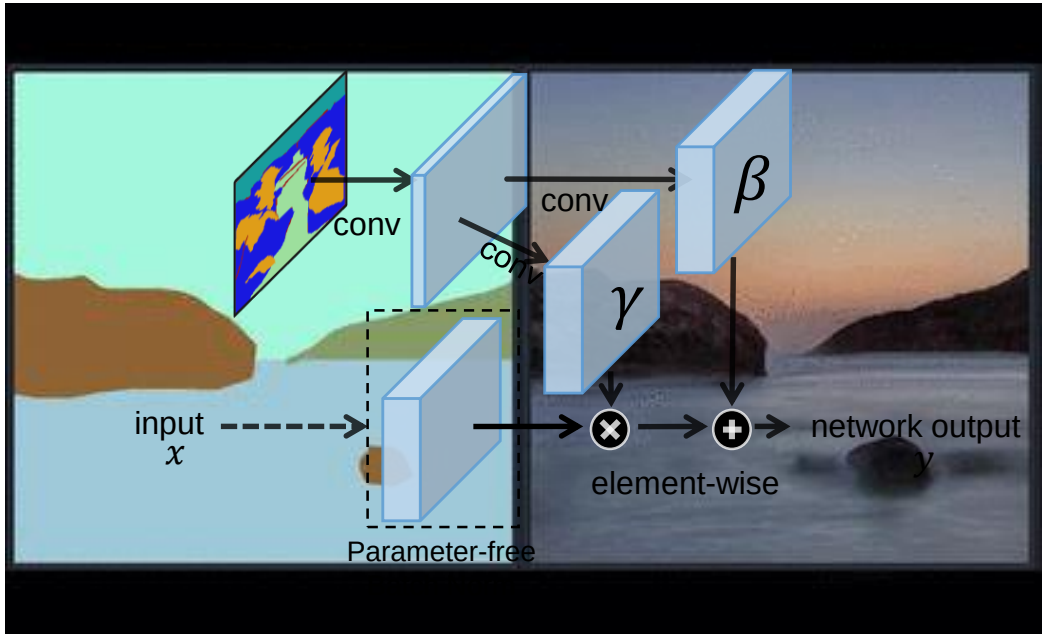


Methodology

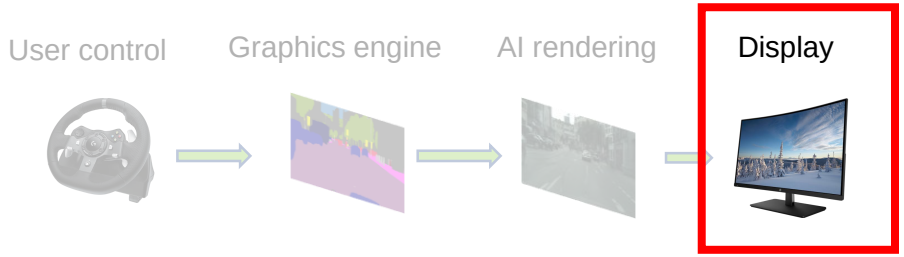


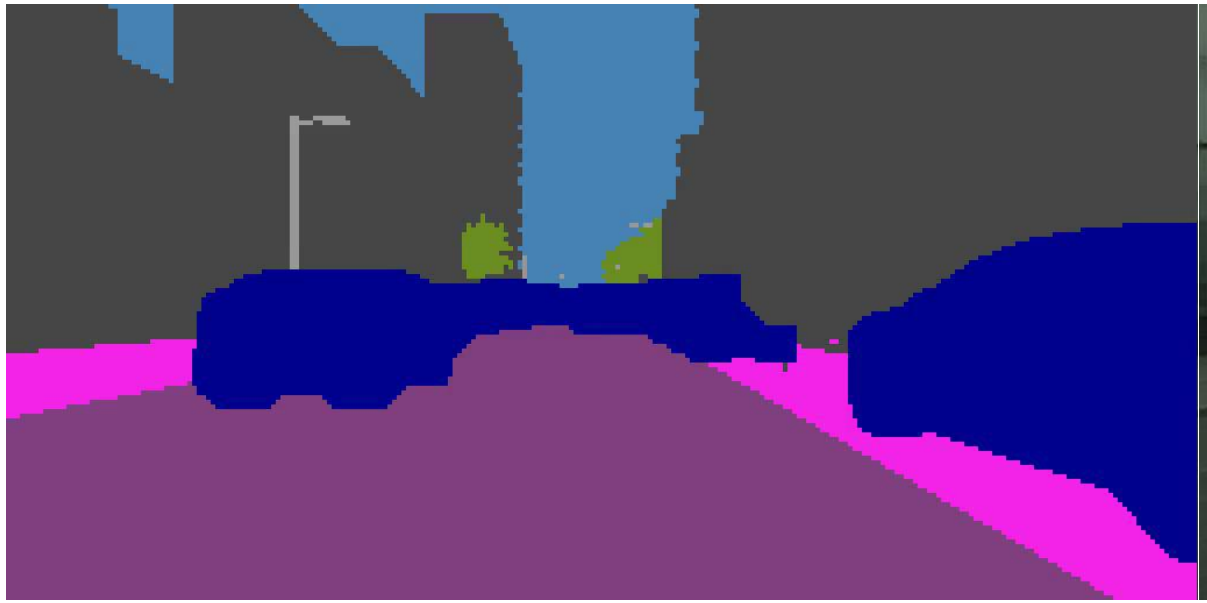
- Combine vid2vid with SPADE





Demo Result





vid2game by FAIR

O. Gafni, L. Wolf, Y. Taigman. “Vid2Game: Controllable Characters Extracted from Real-World Videos,” 2019