Video to Video Synthesis

Elit Cenk Alp
01-2020

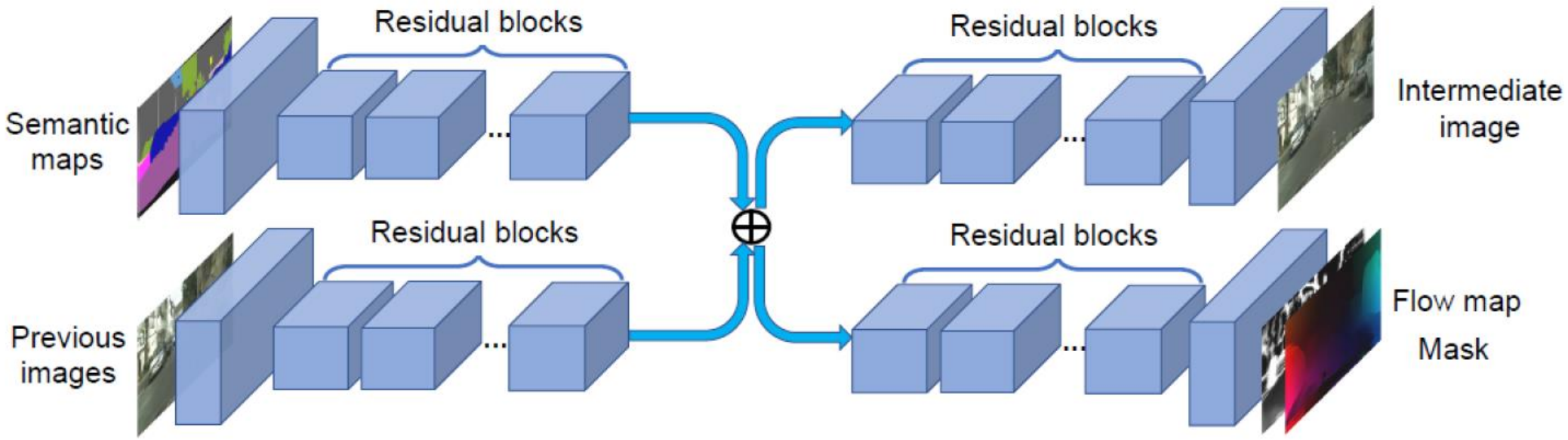
Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

Introduction

- Goal is to learn a mapping function from an input source video to an output photorealistic video.
- Image-to-image synthesis problem, is a popular topic, the video-to-video synthesis problem is less explored in the literature.
- A new video-to-video synthesis approach has been proposed under the framework of generative adversarial learning.
- High resolution, photorealistic, temporally consistent video results have been achieved on a wide variety of input formats, including segmentation masks, sketches, and poses.
- Model is capable of synthesizing 2K resolution videos of street scenes up to 30 seconds long, which significantly advances the state-of-the-art of video synthesis

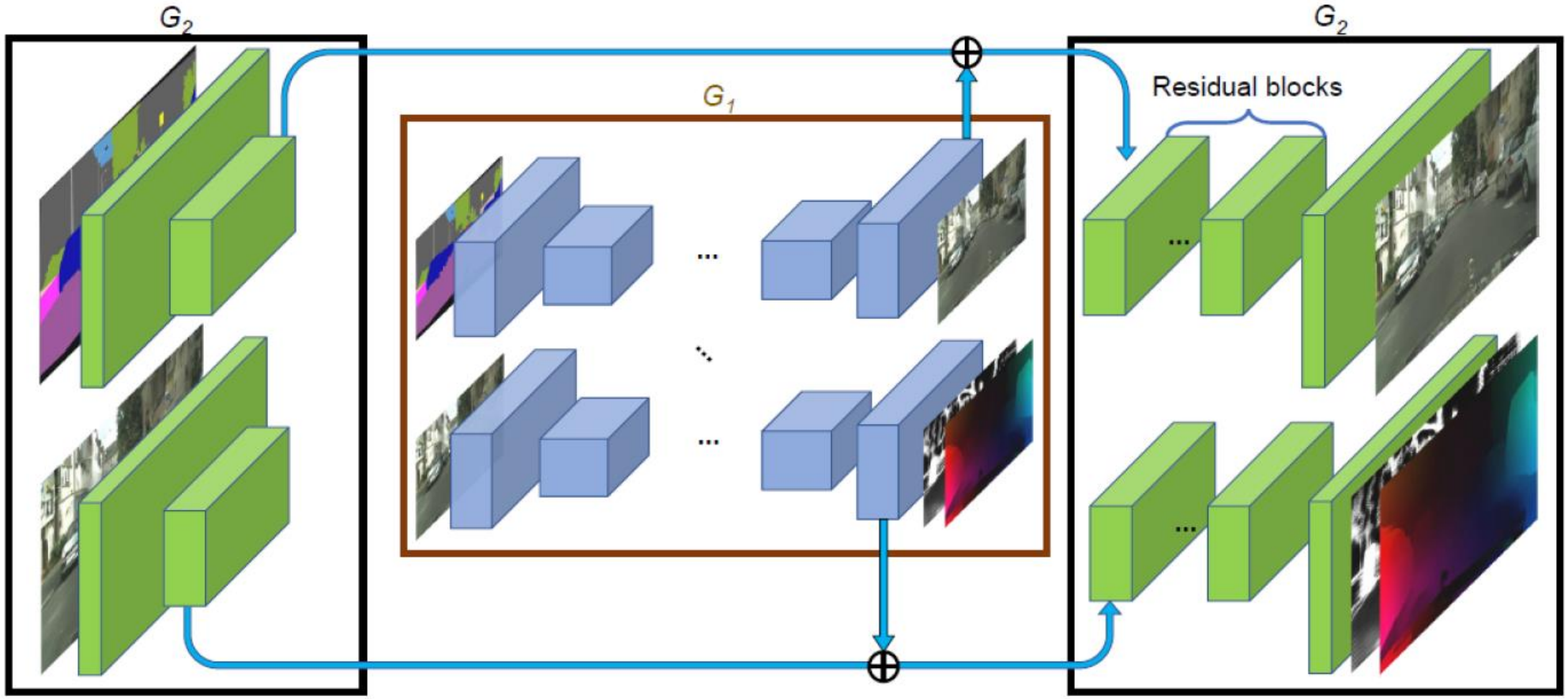
Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

Generators- Low-Res Network



Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

Generators- High-Res Network



Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

Discriminators

- The multi-scale patch GAN architecture is adopted.
- Different amounts of actual/generated sequences are subsample to generate different inputs to temporal discriminators.
- In the finest scale, the original sequence K consecutive frames are taken as input.
- In the next scale, the video is subsampled with a K factor (i.e. skipping every K - 1 intermediate frame), then consecutive K frames are taken as input in this new sequence.
  - It is made for 3 scales and this has been shown to help ensure both short-term and long-term consistency.
- We do this for up to 3 scales in our implementation, and found that this helps us ensure both short-term and long-term consistency.

Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

Experiments

- Semantic Labels → Cityscapes Street Views

Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

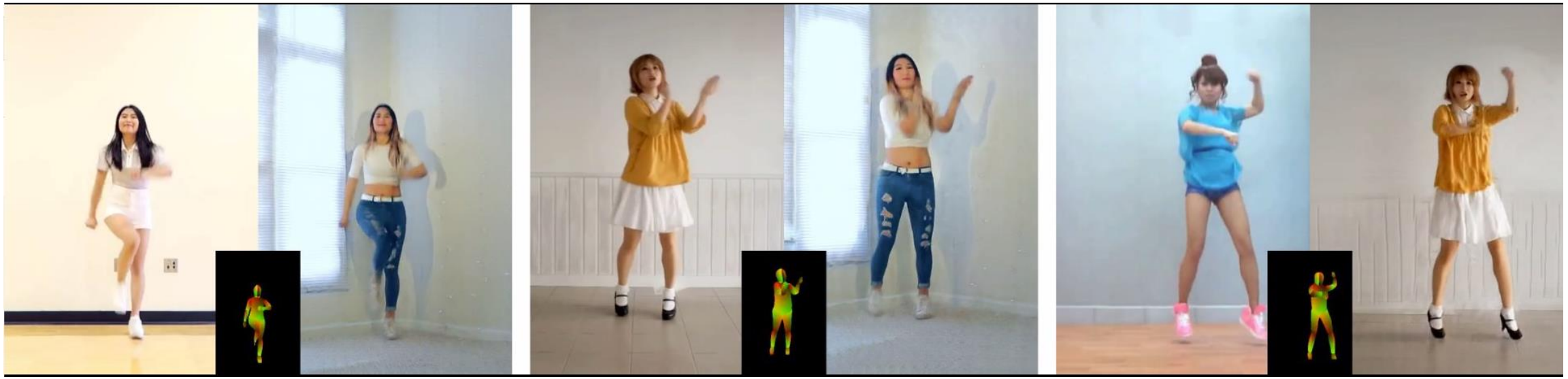
Experiments

- Face → Edge → Face

Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

Experiments

- Body → Pose → Body



Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

Experiments

- Frame Prediction

Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

Results

Video generation score comparision on Cityscape dataset

Fréchet Inception Dist	I3D	ResNeXt
pix2pixHD [12]	5.57	0.18
COVST [13]	5.55	0.18
vid2vid	4.66	0.15

Human Preference Score	Short seq.	Long seq.
vid2vid/pix2pixHD	0.87/0.13	0.83/0.17
vid2vid /COVST	0.84/0.16	0.80/0.20

Video prediction score comparision on Cityscape dataset

Fréchet Inception Dist	I3D	ResNeXt
PredNet [5]	11.18	0.59
MCNet [14]	10.00	0.43
vid2vid	3.44	0.18

Human Preference Score	
vid2vid/ PredNet	0.92/0.08
vid2vid / MCNet	0.98/0.02

Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018

Referance

- Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Guilin Liu and Andrew Tao and Jan Kautz and Bryan Catanzaro. Video-to-Video Synthesis. Conference on Neural Information Processing Systems (NeurIPS), 2018