Recurrent Flow-Guided Semantic Forecasting

Adam M. Terwilliger, Garrick Brazil, and Xiaoming Liu
Department of Computer Science and Engineering, Michigan State University
{adamtwig, brazilga, liuxm}@msu.edu

Computer Vision Lab

Overview

Problem:

Forecast semantic segmentation masks for arbitrary future frames, using predicted motion of RGB sequence frames.

Motivation:

Reasoning about the future is a crucial component to the deployment of robust and proactive real-world computer vision systems.

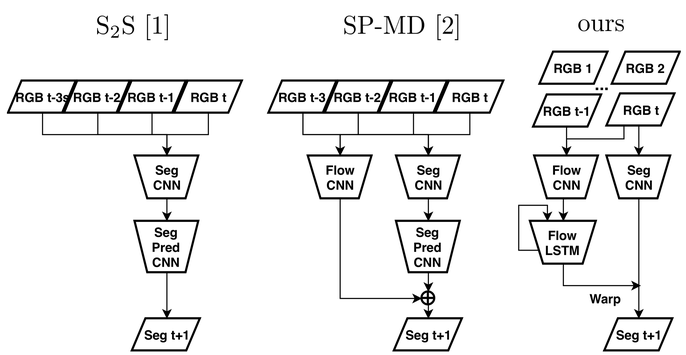
Background:

Extensive prior work in directly modeling future RGB frames and current-frame semantic segmentation, while conversely future frame segmentation is relatively unexplored.

Contributions:

- **learnable warp layer** directly applied to segmentation features
- **convolutional LSTM** to aggregate optical flow features and estimate future optical flow
- collectively, an **effective, efficient, and low overhead network.**

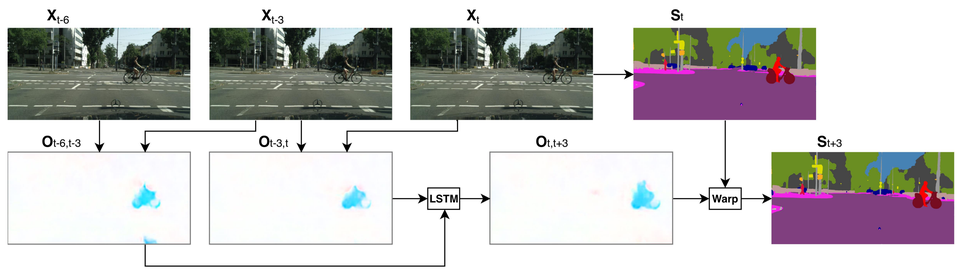
Baseline Comparison



Model	Training (hours)	Testing (sec)	Params (mil)	Training GPUs	Testing Method
S_2S [1]	96 ($8x$)	0.248 ($4.8x$)	≈1.5 (+300%)	1	Single
SP-MD [2]	–	2.182 ($42x$)	≈115 (−95%)	4-8	Sliding
ours	**12**	**0.052**	≈6.0	1	Single

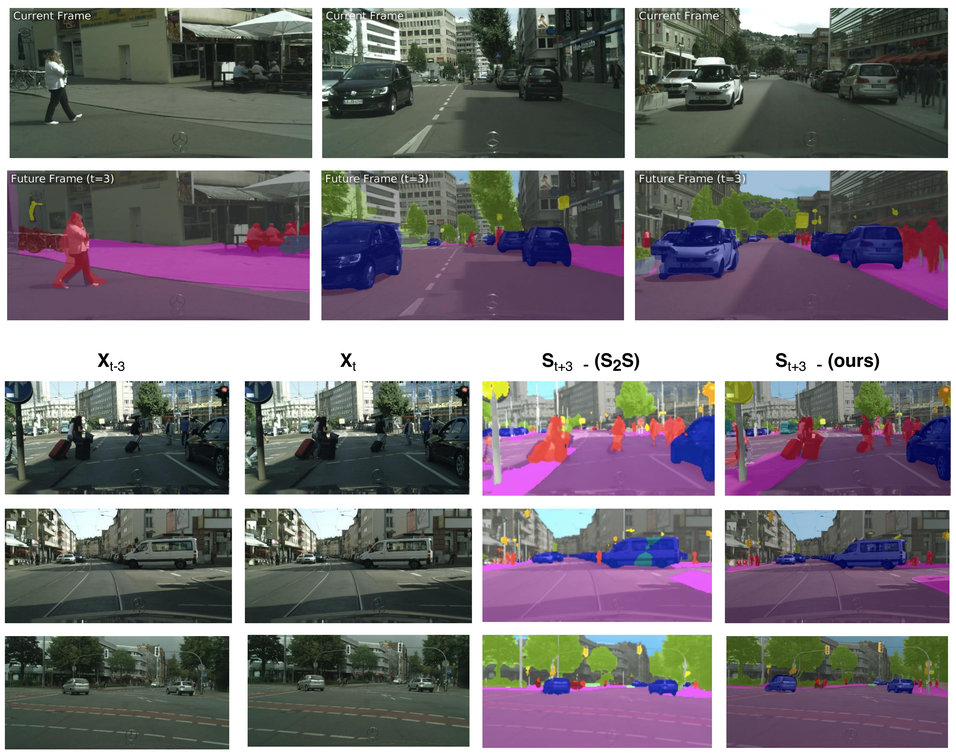
[1]. Luc, P., Neverova, N., Couprie, C., Verbeek, J.J., LeCun, Y.: Predicting deeper into the future of semantic segmentation. ICCV 2017.
[2]. Jin, X., Xiao, H., Shen, X., Yang, J., Lin, Z., Chen, Y., Jie, Z., Feng, J., Yan, S.: Predicting scene parsing and motion dynamics in the future. NIPS 2017.

Methodology



The semantic forecasting framework takes as input a pair of RGB frames, estimates the optical flow, aggregates the flow temporally via a convolutional LSTM, and finally warps current frame segmentation, hence resulting in a highly **flexible** and **modular** design.

Qualitative Results



Experiments

Model	IoU ($t=3$)	IoU-MO ($t=3$)	IoU ($t=9$)	IoU-MO ($t=9$)
Copy last input	49.4	43.4	36.9	26.8
Warp last input	59.0	54.4	44.3	37.0
S_2S	59.4	55.3	47.8	40.8
ours	**67.1**	**65.1**	**51.5**	**46.3**

Model	IoU ($t=1$)	IoU ($t=10$)
Copy last input	59.7	41.3
Warp last input	61.3	42.0
SP-MD†	–	52.6
SP-MD	66.1	**53.9**
ours (c)†	73.0	51.8
ours (C) †	**73.2**	52.5

† - indicates model contained no recurrent fine-tuning.

Ablation Studies

Configuration	IoU ($t=3$)	IoU ($t=10$)
Auto-regressive	64.4	48.7
Single-step	**66.0**	**50.0**

Configuration	IoU ($t=3$)	IoU ($t=10$)
No FlowLSTM	62.3	46.0
FlowLSTM	**66.0**	**50.0**

Step Size	Frames	IoU ($t=3$)
9	7, 16	62.6
3	7, 10, 13, 16	66.0
1	7, 8, ..., 16	**67.1**

Time	Frames	IoU ($t=3$)
2	15, 16	62.4
4	13, 14, 15, 16	67.0
8	7, 8, ..., 16	**67.1**

For details on methodology and ablation experiments please visit the full-length paper using the QR code (left) **or** https://arxiv.org/abs/1809.08318