Forecasting Future Instance Segmentation with Learned Optical Flow and Warping

Author: Andrea Ciamarra
Department of Information Engineering
Laboratory: Media Integration and Communication Center – MICC

PhD program in Information Engineering

Abstract

For an autonomous vehicle it is essential to observe the ongoing dynamics of a scene and consequently predict imminent future scenarios to ensure safety to itself and others.

In this paper we investigate the usage of optical flow for predicting future semantic segmentations.

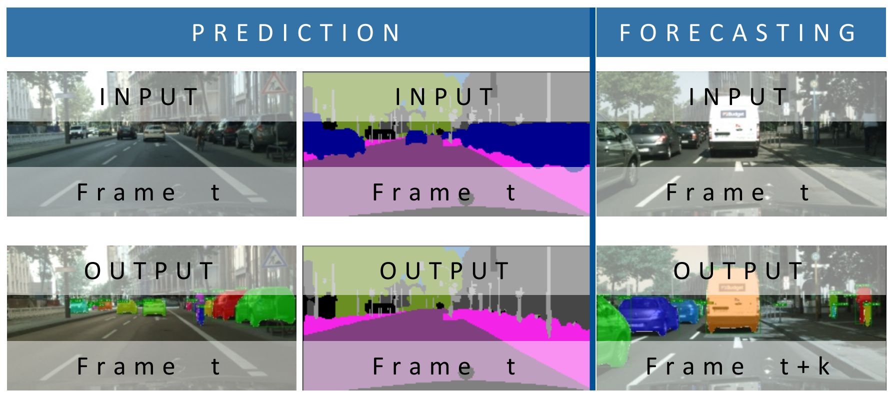
To do so, we design a model that forecasts flow fields autoregressively. Such predictions are then used to guide the inference of a learned warping function that moves instance segmentations on to future frames.

Task

Understanding urban environments from an ego-vehicle perspective is crucial for safe navigation.

This problem can be declined under several points of view, such as detecting entities, understanding the road layout or predicting the future.

In this paper we study how can we rely on optical flow as unique source of information to forecast instance segmentations of moving objects[4].

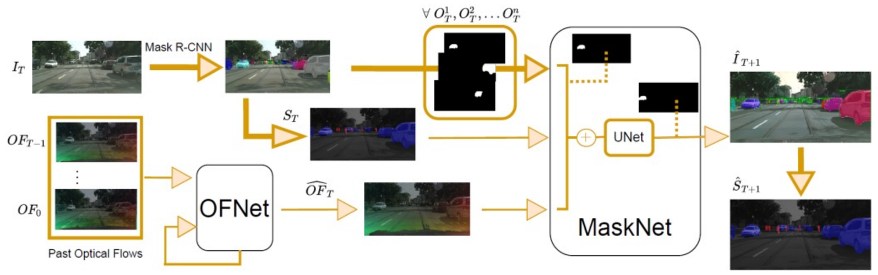


Forecasting Future Instance Segmentations

We break down the problem in two steps:

1. We first design OFNet, an optical flow predictor, which generates future flows;
2. We train a warping neural network, MaskNet, with a dice loss, that propagates current instance segmentations onto future frames, guided by the predicted optical flows and the current semantic segmentation.

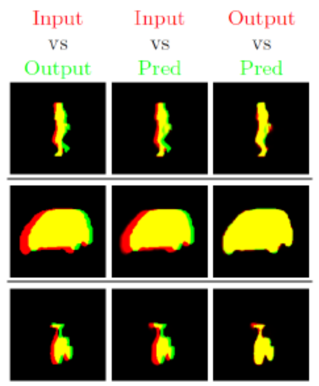
In additional, we generate future semantic segmentation by grouping all object masks together[1][2].



Results

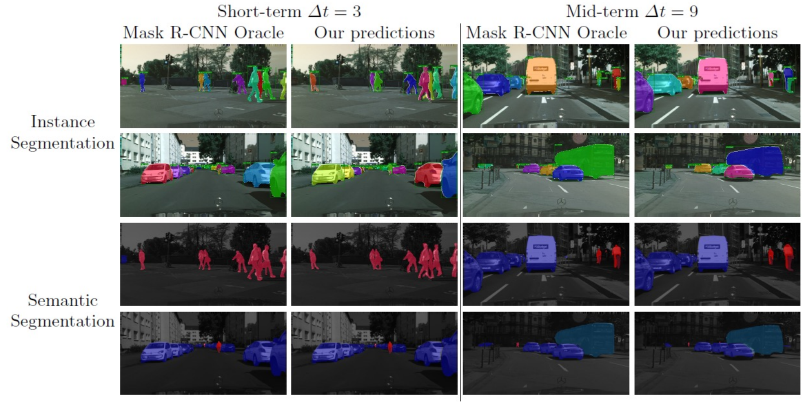
We conduct experiments on the Cityscapes dataset[3] and we test our method on 8 moving objects: person, rider, car, truck, bus, train, motorcycle, bicycle.

Method	Short term $\Delta t = 3$ AP	AP50	IoU	Mid term $\Delta t = 9$ AP	AP50	IoU
Mask RCNN oracle	34.6	57.4	73.8	34.6	57.4	73.8
MaskNet-Oracle	24.8	47.2	69.6	16.5	35.2	61.4
Copy-last segm. [1]	10.1	24.1	45.7	1.8	6.6	29.1
Optical-flow shift [1]	16.0	37.0	56.7	2.9	9.7	36.7
Optical-flow warp [1]	16.5	36.8	58.8	4.1	11.1	41.4
Mask H2	[1] 11.8	25.5	46.2	5.1	14.2	30.5
S2S	-	-	55.4	-	-	42.4
F2F [1]	19.4	39.9	61.2	7.7	19.4	41.2
LSTM M2M	-	-	65.1	-	-	46.3
Deform F2F	-	-	63.8	-	-	49.9
CPConvLSTM	22.1	44.3	-	11.2	25.6	-
F2MF	-	-	67.7	-	-	54.6
PSF[8]	17.8	38.4	60.8	10.0	22.3	52.1
APANet[2]	23.2	46.1	64.9	12.9	29.2	51.4
PFL	24.9	48.7	69.2	14.8	30.5	56.7
MaskNet (Ours)	19.5	40.5	65.9	6.4	18.4	45.5



Recent frameworks, e.g. APANet [2] and PFL, report better results, but take several trainings to capture more information from the past FPNs.

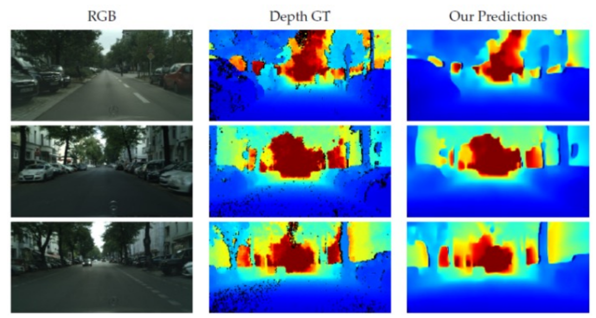
In general our approach is simple and without complex optimization stages can reach competing performances, relying solely on optical flow data.



Depth Forecasting

Our recent updates show that sticking depth map and flow together bring improvements. Furthermore, we can also predict depths beforehand.

Such a model demonstrates how to face with the video understanding task for unobserved frames.



References

[1] Luc, P., Couprie, C., LeCun, Y. and Verbeek, J., 2018. Predicting Future Instance Segmentation by Forecasting Convolutional Features. *arXiv preprint arXiv:1803.11496*.
[2] J.-F. Hu, J. Sun, Z. Lin, J.-H. Lai,W. Zeng, andW.-S. Zheng, "Apanet: Auto-path aggregation for future instance segmentation prediction," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
[3] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B., 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. *arXiv preprint arXiv:1604.01685*.
[4] Ciamarra, A., Becattini, F., Seidenari, L. and Del Bimbo, A., 2022. Forecasting Future Instance Segmentation with Learned Optical Flow and Warping. *arXiv preprint arXiv:2211.08049*.

Vote here for this poster: