Momentum Adapt: Robust Unsupervised Adaptation for Improving Temporal Consistency in Video Semantic Segmentation During Test-Time

Tampere University
NOKIA

Amirhossein Hassankhani[1], Hamed Rezazadegan Tavakoli[2], Esa Rahtu[1]
Tampere University[1], Nokia Technologies[2]

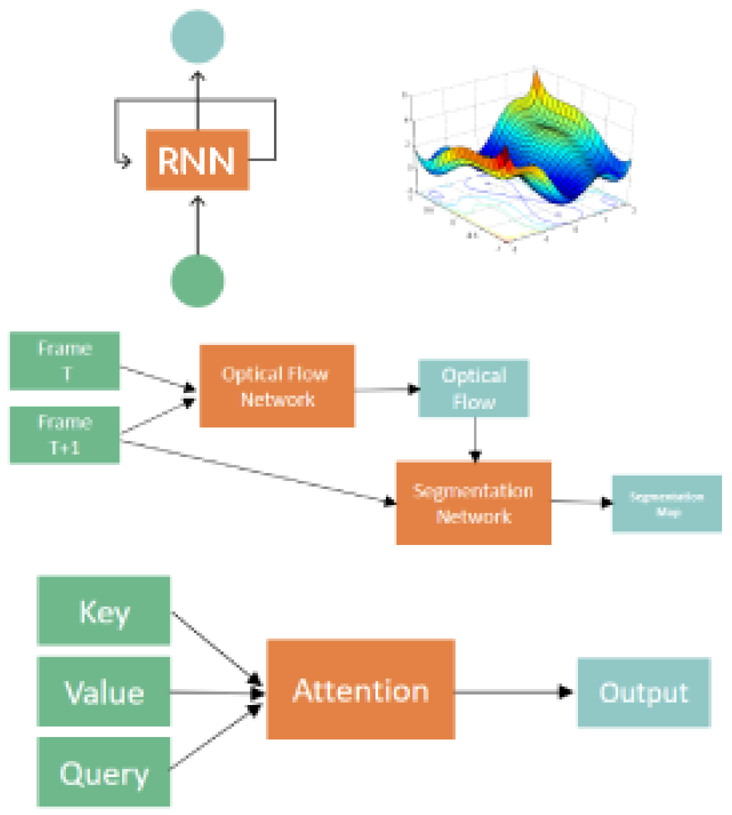
BMVC 2023

Motivation

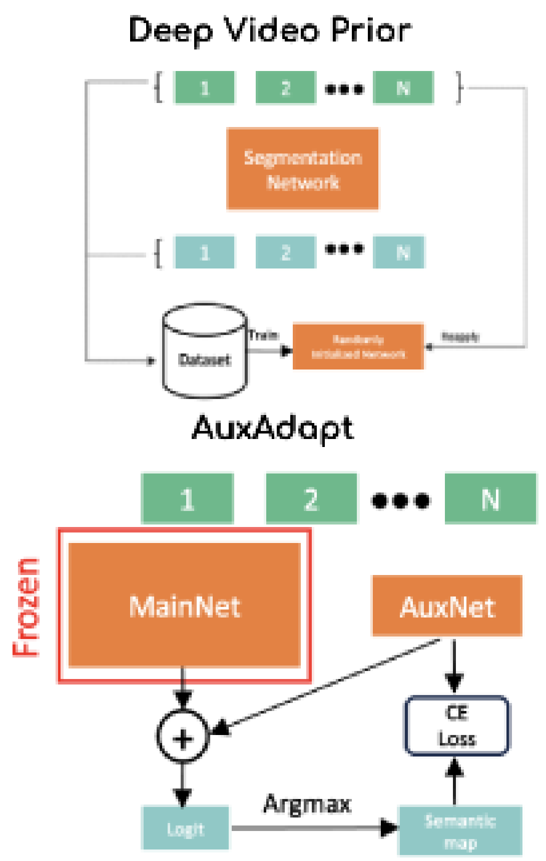
In many situations, the semantic network might miss a particular object because of imperfect conditions (e.g., noise, occlusion. etc.), which could have been avoided by considering the information from past frames. These misdetections can have catastrophic consequences in safety-sensitive applications like self-driving cars.

Previous Works

Changing Model Design



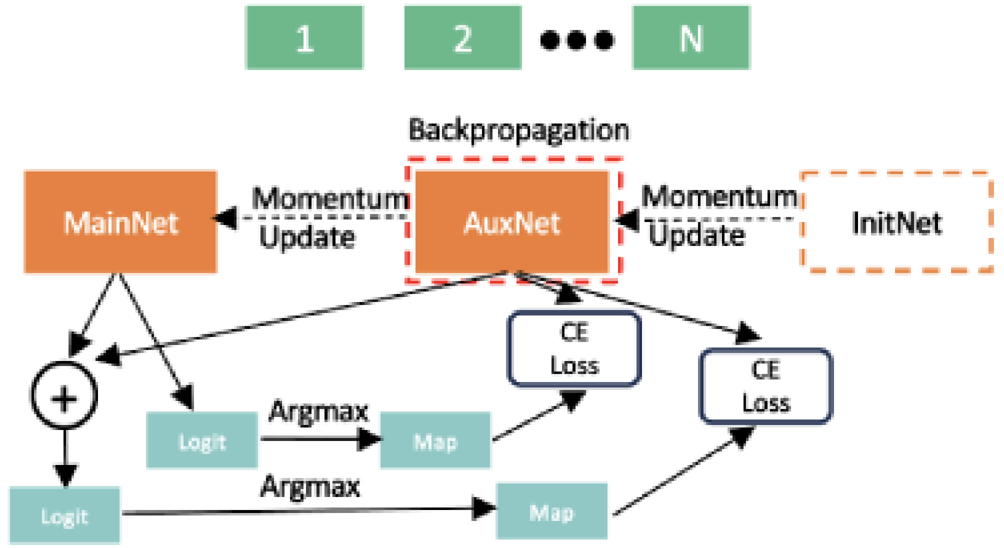
Fine-tuning Pre-trained Model



Ours

Momentum Adapt

Inspired by AuxAdapt, Momentum Adapt is an online unsupervised adaption method that improves the segmentation model's temporal consistency. Momentum Adapt has two identical networks. AuxNet is updated by back-propagation and momentum updates from the initial network. MainNet is updated using momentum updates from AuxNet.



Momentum Update are computed from source network to target network as follows:

$$\theta_t \leftarrow m\theta_t + (1-m)\theta_s$$

Experiments

Performance Metrics

Two methods can have identical pixel metrics (mIoU, Accuracy, etc.) but have different temporal consistency. Additionally, only considering temporal metrics is not useful. For example, the network can output fixed prediction having perfect temporal consistency. For this reason, we use both metrics together.

For the pixel-related metric, mIoU is used where the target is the ground truth label. For measuring temporal consistency, mIoU is also used. However, the target is the previous frame prediction's warped version (using FlowNet2).

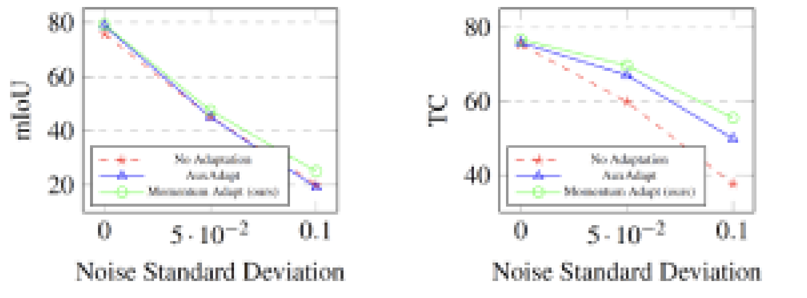
Normal Experiments

Adaptation	Base Network	TC	mIoU
SceneNet RGB-D Dataset			
No Adaptation	HRNetV2-w18s	53.44	44.32
AuxAdapt		57.45	45.00
Momentum Adapt (ours)		**59.47**	**48.03**
No Adaptation	HRNetV2-w18	56.88	45.75
AuxAdapt		61.98	48.48
Momentum Adapt (ours)		**65.20**	**51.15**
No Adaptation	HRNetV2-w48	59.17	54.47
AuxAdapt		45.97	2.32
Momentum Adapt (ours)		**65.87**	**60.3**
No Adaptation	SegFormer-b5	60.61	56.46
AuxAdapt		**64.15**	**58.80**
Momentum Adapt (ours)		63.87	57.19
Cityscapes Dataset			
No Adaptation	Unet-s5-d16	64.02	66.98
AuxAdapt		65.74	67.16
Momentum Adapt (ours)		**67.85**	**67.52**
No Adaptation	HRNetV2-w18s	73.08	72.20
AuxAdapt		77.07	72.80
Momentum Adapt (ours)		**77.70**	**74.09**
No Adaptation	HRNetV2-w18	75.35	75.84
AuxAdapt		78.82	75.85
Momentum Adapt (ours)		**79.27**	**76.74**
No Adaptation	HRNetV2-w48	76.31	77.12
AuxAdapt		78.95	77.46
Momentum Adapt (ours)		**79.13**	**78.19**
No Adaptation	DeepLabV3-r50-d8	75.64	76.91
AuxAdapt		78.92	76.67
Momentum Adapt (ours)		**79.16**	**77.67**

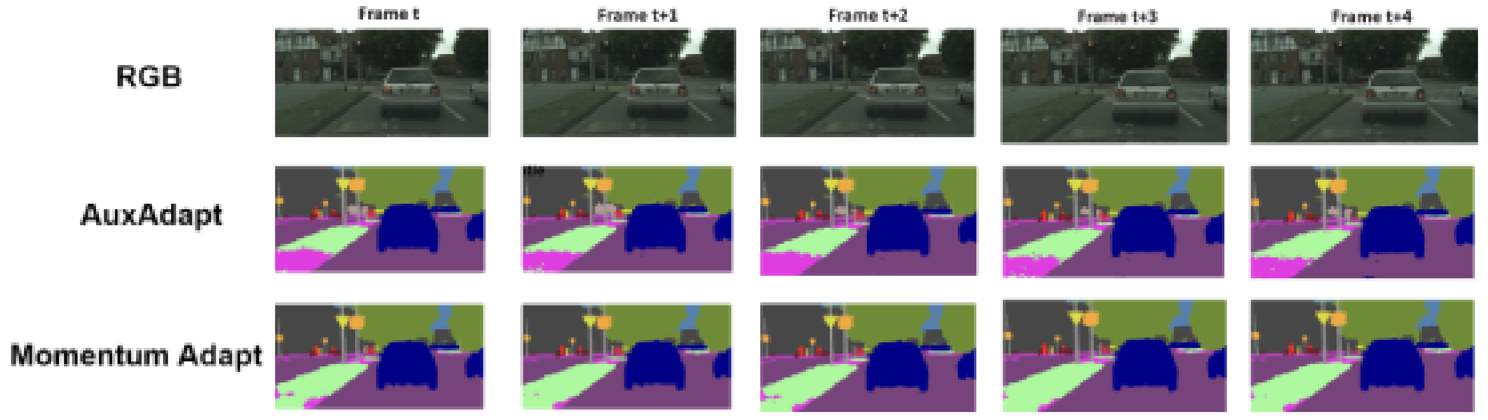
Domain Shift Experiments (KITTI->Cityscapes)

Method	TC	mIoU
FCN-r101-d8 (AuxNet and MainNet) w/o Adaptation	62.17	55.37
w/ AuxAdapt	67.65	58.86
w/ Momentum Adapt (ours)	**69.64**	**59.47**
HRNetV2-W18 | (AuxNet and MainNet) w/o Adaptation	64.42	60.15
w/ AuxAdapt	70.32	62.74
w/ Momentum Adapt (ours)	**71.57**	**64.21**
DeepLabV3Plus-r50-d8 (AuxNet and MainNet) w/o Adaptation	64.03	59.66
w/ AuxAdapt	71.31	63.97
w/ Momentum Adapt (ours)	**72.15**	**65.76**
PSPNet-r101-d8 | (Auxnet and MainNet) w/o Adaptation	66.94	61.20
w/ AuxAdapt	72.00	63.46
w/ Momentum Adapt (ours)	**72.78**	**65.97**

Noise Experiments



Conclusion



In addition to quantitive metrics, qualitative comparison shows that our method, with minimial extra computation, outperforms AuxAdapt and significantly improves the performance of the base network.

The main disadvantage of adaptation during test time is the computation time. Further optimization is needed for these algorithms to be used in real-time settings.