Supplementary Material

1 OURS V2 NETWORK

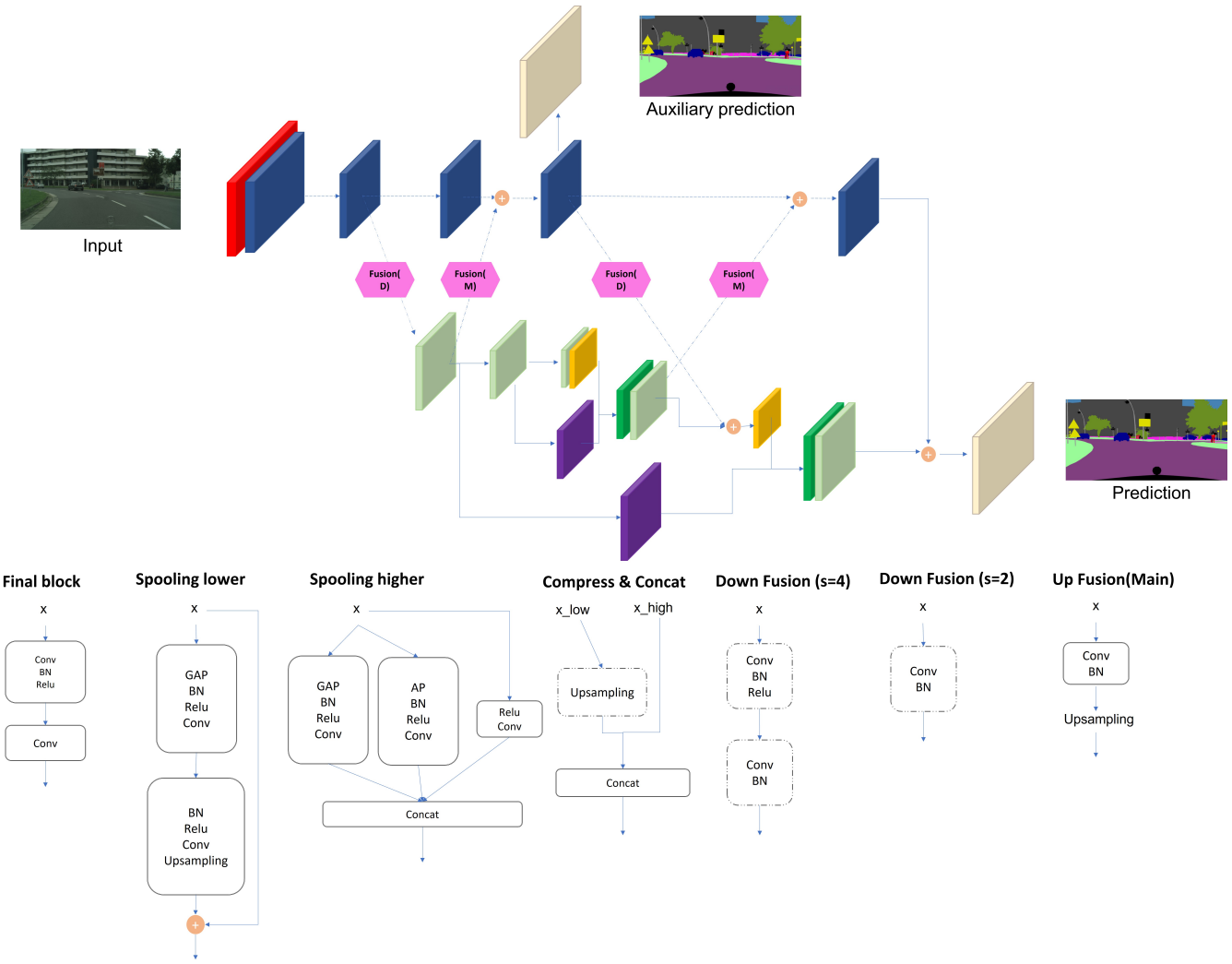

Figure S1. This is Ours_v2 network architecture. Each color represents a different layer: Red - two sequential convolution layers(stride is 2); Blue - Residual Block(He et al., 2016); Beige - Final layer; Light green - HardNet Block; Dark green - compress & concat layer; Purple - S-pooling higher; Yellow - Spooling lower; Fusion(D) and Fusion(M)- Down fusion and Up fusion. The stride and upscale depend on resolution of feature maps. AP - Average pooling, BN - Batch normalization, GAP - Global Average Pooling. All convolution layers have 3x3 kernel size.

The overall architecture is similar to the Our network used for the crossing street navigation system. Differences are: 1) Main branch that consists of several Residual block 2) HardNet block in the down branch and 3) S-pooling compared to Ours network. It achieved the best mIOU on validation Cityscapes dataset with fewer parameters (vs. Ours network). Compared to a network without the S-pooling, the network with S-pooling identifies details of features(mIOU: 74.12 to 75.93). However, the 8 sequential layers of the default HardNet block and the S-pooling from the block resulted in increased number of flops

due to the average pooling so that the FPS and the Net score decreased in spite of fewer parameters. Ours v2 can be considered as a base network to add more layers for different tasks such as instance segmentation due to the highest mIOU results.

REFERENCES

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778