자율 주행 차를 위한 고효율 시맨틱 분할에 관한 연구

Gankhuyag Ganzorig, 박창규, 민경원
KETI

{gnzrg25, chgpark, minkw}@keti.re.kr

High efficient semantic segmentation for autonomous vehicle

Ganzorig Gankhuyag, Changyu Park, KyoungWon Min
KETI

Abstract

The Image segmentation algorithm has significantly improved by deep learning. However, high accuracy trades off huge computational costs and is therefore not suitable for the limited resource of autonomous driving vehicle. We propose a deep learning framework for image segmentation that is efficient enough for an autonomous driving vehicle environment while maintaining high accuracy and speed. Our proposed work achieves high segmentation accuracy at Cityscape public dataset and our private test dataset. At the same time, the frame rate is high enough for deployment in autonomous vehicles.

Ⅰ. Introduction

Autonomous driving systems based on visual recognition input are 2D/3D object recognition, object positioning, and semantic segmentation. Modern autonomous vehicle driving systems require multiple visual cameras, Radar, LiDAR, and other types of sensors. Therefore, they use powerful hardware systems with GPU to compute all visual perceptions by using a neural network. The recent deep learning algorithm of various perception algorithms illustrated promising good performance[1]. Therefore, all deep learning algorithms deployed on these systems must meet the straight requirements for GPU memory and speed. However, semantic segmentation architecture requires more computation power compared to other visual perception algorithms.

The semantic segmentation is to label each pixel in the digital image into a given semantic class (e.g., road, street marking, vehicle, pedestrian, or sidewalk). Semantic segmentation requires a different type of network scale for proper labeling. Furthermore, the high-level feature of classes is predicted from the higher scaled layer of the network, and the low-level feature of each pixel is classified from the low scaled layers of the network. This trade-off of the network resolution is solved by applying skip-connections from the lower layer to the higher layer. The first introduced fully convolutional network (FCN)[2] for image segmentation, end to end trained neural network is applied to classify each pixel in an image with skip connections. Then U-Net[3] was introduced, fairly symmetrical U-shaped encoder-decoder architecture, combined with skip connections at different resolution layers, which still serves as baseline network architecture for semantic segmentation. In recent years, the use of a better CNN backbone for semantic segmentation has given good results. For example, Densenet[4] and HarDNet[5]. Also, insert the additional modules such as attention modules[6] to capture more precise information for features from the different scaled network layer.

We proposed semantic segmentation architecture that uses HarDNet as the backbone and encode-decoder architecture. In addition, we added an attention model[6] to our network architecture to improve the accuracy. For the limited resource of GPU, we used the TensorRT framework to reduce the memory bandwidth and increase the inference speed.

Ⅱ. High efficient semantic segmentation architecture

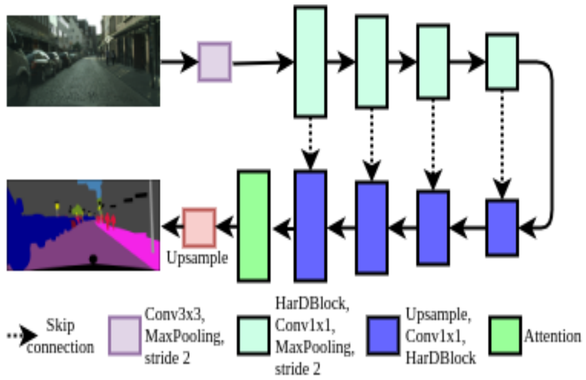

Fig. 1. The proposed semantic segmentation architecture.

We use HarDNet (i.e., Harmonic Densenet) as the model backbone of our proposed network. HarDNet improved and sparsified Densenet. Regarding the memory traffic for model design. It reduces the shortcuts of Densnet to increase the speed and weighting the channel width of the layer to get parameter efficiency and reduce the loss of accuracy.

This design achieves a 30% inference time reduction compared to Densenet. Additionally, FCHarDNet reaches the state of the art in image segmentation on the Cityscapes dataset. Therefore, we use HarDNet as the model backbone for our proposed work.

We add an attention network[6] in our semantic segmentation model. After up-sampling to the same scale, the attention network performs.

Ⅲ. Simulation result

Our proposed network is compared to the FCHardNet segmentation network. As far as we know, it is the currently best network designed for semantic segmentation for low latency.

We evaluated the proposed network on the Cityscapes public data set and our KETI semantic segmentation private dataset (16 different semantic classes). We achieved a mean IoU of 77.1% at Cityscapes and 76.1% at KETI semantic segmentation dataset. Hence the network architecture can outperform the FCHardNet at both datasets, shown in Table 1 and Table 2.

Table 1: mIoU of proposed work on Cityscape val dataset.

	mIoU
FCHarDNet70	76.0 %
Proposed	77.1 %

Table 2: mIoU of proposed work on KETI semantic segmentation dataset.

	mIoU
FCHarDNet70	75.2 %
Proposed	76.1 %

To reduce the memory usage and improve the inference speed, we use the TensorRT framework. TensorRT is a well-known framework to optimize the inference network. Table 3 shown the inference speed and memory usage of our proposed work.

Table 3: Inference speed (GPU: GeForce RTX 2080TI, Image size: 1024x2048, TensorRT: FP16)

	Speed (FPS)	Memory usage (MB)	mIoU (%)
with TensorRT	102.5 fps	1049 MB	77.1%
without TensorRT	56.4 fps	3910 MB	77.1%

IV. Conclusion

We have presented a high efficient semantic segmentation algorithm for autonomous driving vehicle. Our proposed method achieved the mean IoU of 77.1% at the Cityscape dataset and the mean IoU of 76.1% in the KETI segmentation dataset with high inference speed, shown in Fig 2 and Fig 3.

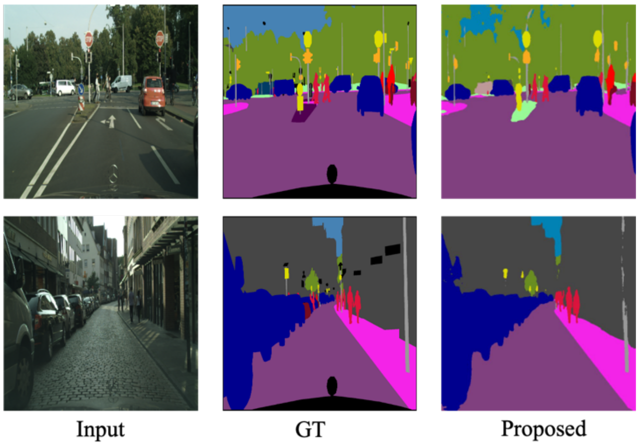

Fig 2. Segmentation result of proposed work on Cityscape val set.

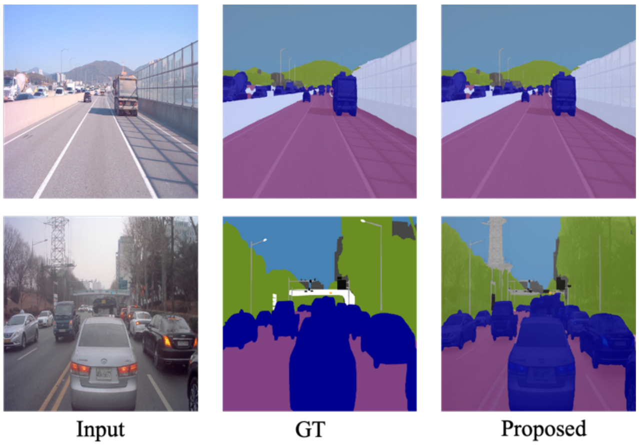

Fig 3. Segmentation result of proposed work on KETI segmentation val dataset.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2019-0-00399, Development of A.I. based recognition, judgement and control solution for autonomous vehicle corresponding to atypical driving environment).

References

[1] Miglani, et all., “Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges.”, Vehicular Communications, 2019.

[2] Long, J., et all., “Fully convolutional networks for semantic segmentation.”, CVPR, pp. 3431-3440, 2015.

[3] Ronneberger, et all., “U-net: Convolutional networks for biomedical image segmentation.”, MICCAI, pp. 234-241, 2015.

[4] Iandola, et all., “Densenet: Implementing efficient convnet descriptor pyramids.”, arXiv preprint arXiv:1404.1869, 2014.

[5] Chao, P., Kao, et all., “Hardnet: A low memory traffic network.”, CVPR, pp. 3552-3561, 2019

[6] Fu, J., et all., “Dual attention network for scene segmentation.”, CVPR, pp. 3146-3154, 2019