МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ

Национальный исследовательский Томский государственный университет
Томский государственный университет систем управления
и радиоэлектроники
Болгарская Академия наук
Академия инженерных наук им. А.М. Прохорова
Международная научно-техническая организация «Лазерная ассоциация»
Всероссийское общество изобретателей и рационализаторов

ИННОВАТИКА-2021

СБОРНИК МАТЕРИАЛОВ

**XVII Международной школы-конференции студентов,
аспирантов и молодых ученых
22–23 апреля 2021 г.
г. Томск, Россия**

Под редакцией А.Н. Солдатова, С.Л. Минькова

Scientific & Technical Translations
STT
ИЗДАТЕЛЬСТВО
Томск – 2021

COMPUTER VISION WITH SEMANTIC SEGMENTATION FOR AUTONOMOUS GROUND VEHICLE

M.J. Mohammed, S.V. Shidlovskiy
National Research Tomsk State University
maryamjasim80@gmail.com

In this work, we implement of semantic segmentation for images to enhance object detections and autonomous ground vehicle for driving in road, also an object understanding and classification. We performed on video of regional roads to verify the feasibility of semantic segmentation. The results of our semantic segmentation determined the road areas in each frame were truly segmented, and that road lane and objects can be categorized.

Keywords: autonomous driving, SegNet, deep learning, object detection, object classification, semantic segmentation.

Introduction

In auto-driving setting, the primary inputs from sensors and cameras such as depth information, proximity measurements, and distance measurement information. The model can use these inputs in to produce optimal driving instructions [1]. The number of training examples required to train the model for arbitrary scenarios is huge. So, it's extremely difficult to interpret or taking a particular action or to make model lead decisions for vehicle behavior. However, in an autonomous driving scenario with image that has high dimensional input, these methods should demonstrate to what extent each pixel contributes to the driving decision. Because of our real-world geometry, this results in smaller or larger continuous patches of single class identities that represent the shape of individual objects, e.g , cars, sidewalks, buildings and pedestrians (Figure 1). It is clear that the product of the semantic segmentation network represents important initial decisions of the analogous system [2]. Decisions are made about the presence or absence of humans or vehicles at this stage to reflect the output as one of the sectors. In order to define a conventional pipeline of engineered features, also categorized not to be handcrafted, by using human knowledge to define the abstraction layer of the overall problem. Elementary feature analysis is fully trained and improved in holistic methods. In this work we propose constructs capable of inferring semantic hashing in real time while achieving better accuracy than other competitors.

Semantic Segmentation

Semantic segmentation gives detailed information of road scenes which can be important for mapping and localization tasks. Pixels classifying in the image to classes can offer gathering points from the scene which are static that can be used in localization. The classes like «Road», «Side-walk» and «Parking» give information that can be directly used for path planning. The deep learning approaches presented a good study that achieve competitive and promising results on segmentation task [3]. Also achieve grate results on the Cityscapes evaluation benchmark compared to the others. Semantic segmentation presented in this work essentially pixel level classification belonging to different pre-defined classes. This task involves associating pixels of the image to belong to individual objects. A difficulty with this task is that the image might contain any number of objects, and thus having a pre-defined number of object instances is not sufficient. Recent work combining CNNs and recurrent neural networks [4], segmenting one object at a time, shows promising results and could be useful for future work in relation to the content of this work. the goal of our task is to assign a class to each pixel of the input image as shown in Figure 2, each mask covers all the objects that has the same class with distinction between different instances.

Fig. 1. Example of ground truth on images

Fig. 2. Result of performing Semantic Segmentation over an image with ten classes

Experimental details

In the task parsing and perception scene semantic segmentation was used. We used a modified SegNet with a pre trained auto driving model proposed by Badrinarayanan et al [5]. SegNet is a deep learning encoder network that has thirteen layers of convolutional network with VGG16 and it had eleven classes, such as cars, road, bikes, etc. The information gives a segment understanding for a scene to recognize the objects. In a dynamic road traffic scene, the pixel for the building class may be more reliable than a pixel from a Cars or road signs or pedestrians in motion estimation and it should be sampled with high probability process.

In order to obtain in real time instance semantic segmentation, we have to execution speed versus segmentation accuracy as most effective segmentation networks, our model is structured as pair of encoder decoder. An encoder CNN detects high level classes such as cars or pedestrians in the input image. A decoder takes this information and emboss it with information from the layers in the lower structure of our encoder, providing a prediction to each pixel in the original input. Figure 3 shows network architecture that used. Pooling layers in the encoder was used to guarantee the degree of translational invariance when detecting the object or a part of object by turn reduce the spatial.

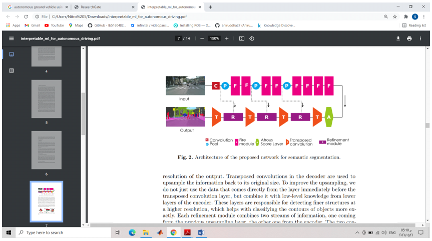

Fig. 3. Proposed architecture network for semantic segmentation

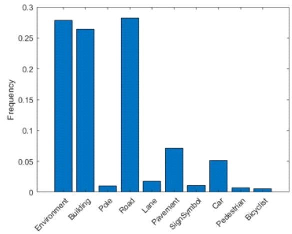

Fig. 4. Statistics of counts the number of pixels by class label

Results

We evaluated the network on the test set using the official evaluation server using Intersection over Union (IoU) where *IoU = area of overlab / area of union*. We achieve 66.1 average of IoU and higher IoU of class is 97.1 and minimum 11.5. Hence the network architecture has good performance like other models like SegNet in [6]. Visual examination of the classes predictions of the showed satisfying results on typical street scene images object in the image segmented looks very sharply (Figure 1). We believe that this is due to the enhanced ability to combine pixel level information in earlier layers in the encoder to upper layers in the decoder by using improvement modules in the architecture.

Conclusions

The presented work designed to explore the road semantic segmentation with SegNet network initialized using VGG-16 weights on the domain. The tests have been carried out on Camvid road datasets. Utilizing this methodology, demonstrated that SegNet based semantic segmentation which is tuned to identify general images can be useful to effectively segment road area on image domain. The evidence from this study suggests that a changed system can improve the accuracy at distinguishing many road scene situations.

The most obvious advantage is efficiency, since road centered crops allow for more training examples of classes and better control of the ratio of positive to negative ratio whilst training mostly involving memory, time and accuracy.

Hence, the detailed analysis of deep learning methods presented in the work offers the potential understanding and ongoing research to improve current methods.

References

1. Sun C. et al. Proximity based automatic data annotation for autonomous driving // IEEE/CAA Journal of Automatica Sinica. 2020. Vol. 7, No. 2. P. 395–404.
2. Zhong J., Lei T., Yao G. Robust vehicle detection in aerial images based on cascaded convolutional neural networks // Sensors. 2017. Vol. 17, No. 12. P. 2720.
3. Yecheng L., Lin B., Xinming H. Road segmentation using CNN and distributed LSTM // 2019 IEEE International Symposium on Circuits and Systems (ISCAS). 2019. P. 1–4.
4. Girshick R., Donahue J., Darrell T. et al. Region-based convolutional networks for accurate object detection and segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016. Vol. 38, No. 1. P. 142–158.
5. Badrinarayanan V., Handa A., Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling // arXiv preprint. 2015. Article No. 1505.07293.
6. Visual Perception Using Monocular Camera [Electronic Resource]. URL https://www.mathworks.com/help/driving/examples/visual-perception-using-monocular-camera.html (date: 21.03.2021).