A Preliminary Study of Attribute-aware Semantic Segmentation for Pedestrian Understanding

Mahmud Dwi Sulistiyo[1, 2], Yasutomo Kawanishi[1], Daisuke Deguchi[1], Ichiro Ide[1], Hiroshi Murase[1]

[1]Graduate School of Informatics, Nagoya University, Japan, [2]School of Computing, Telkom University, Indonesia

1. MOTIVATION

Semantic segmentation is widely developed in many researches for various purposes. In recent years, several studies have been conducted using various approaches for better classifying a number of distinct objects. Furthermore, semantic information is beneficial for segmentation and segmentation results help more in recognizing semantics. However, only class of each region is considered as the semantic information in traditional semantic segmentation methods. We consider attributes of objects as such additional information. For instance, since the silhouette of a pedestrian is strongly related to the pedestrian's facing orientation, it can help orientation recognition of the pedestrian. We are aiming to collaborate between semantic segmentation and attribute recognition for better results in scene understanding. As the first step, we utilize semantic segmentation results for pedestrian orientation recognition.

2. EXISTING WORK

In this study, we use SegNet [1], which is a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. The encoder contains convolutional layers followed by max-pooling and sub-sampling layer, while the decoder consists of layers same as in the encoder but in reverse order. The model is trained end-to-end using stochastic gradient descent.

We also considered a related work that had been conducted using PDC 2014 dataset [2]. It is based on the "Daimler Mono Pedestrian Classification" dataset [3] that was extended for body direction classification. It contains 8 orientations i.e. North, Northeast, East, Southeast, South, Southwest, West, and Northwest, respectively.

3. PROPOSED FRAMEWORK

We proposed a framework consisting 3 main steps: semantic segmentation, pedestrian detection, and attribute recognition. In semantic segmentation, an image is segmented into pixel-wise class labels. The pixels corresponding to pedestrians are being traced in the pedestrian detection step. Then, a trained attribute recognition model is applied to the detected pedestrians.

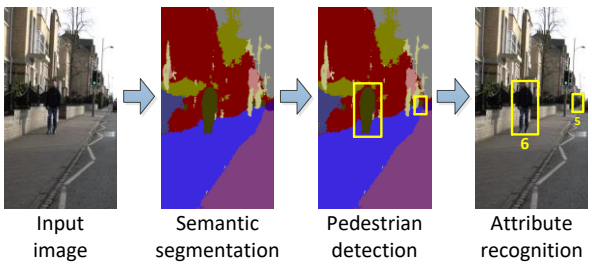

Fig. 1. Main process flow in the proposed framework.

Table 1 Evaluation of preliminary experiment

Task	Performance
Semantic segmentation	Global acc.: 85.8%, Class acc.: 57.3%, Mean IoU: 45.2%
Pedestrian orientation (CNN)	PDC 2014: 93.1% (train), 81.8% (test)
	CamVid: 67.6% (37 testing images)
Pedestrian orientation (HOG+SVM)	PDC 2014: 96.2% (all 11,562 images)
	CamVid: 51.35% (37 testing images)

The whole concept of this research is intended to understand the pedestrian conditions to provide more visual semantic information. The system will finally be informed about where people commonly walk, for what purposes, and whether those people would be obstacles for the driving path or not.

4. PRELIMINARY RESULT

Preliminary experiments were conducted using existing datasets. We re-trained and re-tested the VGG16 model using CamVid dataset [4] for the semantic segmentation. It contains 367 training and 233 testing images with 11 object class labels. For pedestrian's facing orientation recognition, we utilized CNN and HOG+SVM. PDC dataset [2] and a few number of pedestrians extracted from CamVid dataset [4] were used in this task. PDC dataset contains a total of 11,562 images. We used 90% of it as training set and the rest 10% for testing, while all additional CamVid images were used for testing. The overall results are shown in Table 1. Unfavorable results on extracted CamVid images are due to inappropriate cropping, lack of lighting, and low image resolution.

5. CONCLUSION

Based on the preliminary results, we conclude that our proposed framework seems working and gives sufficient result. The current result was not bad, but optimizing the training process as well as dataset enhancement is necessary to obtain a better performance. In addition, we consider creating a hybrid system to better recognize pedestrian attributes using semantic segmentation techniques.

References

(1) V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation," in IEEE Trans. PAMI, 2017.

(2) J. Tao and R. Klette, "Part-based RDF for Direction Classification of Pedestrians, and a Benchmark," ACCV workshop, 2014.

(3) S. Munder and D.M. Gavrila, "An Experimental Study on Pedestrian Classification," IEEE Trans. PAMI, vol. 28, no. 11, pp. 1863-1868, 2006.

(4) G.J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic Object Classes in Video: A High-definition Ground Truth Database," Pattern Recognition Letters vol. 30, no. 2, pp. 88-97, 2009.