3D Mask-Based Shape Loss Function for LIDAR Data for Improved 3D Object Detection

R. Park and C. Lee
Dept. of Electric and Electronic Engineering, Yonsei University, Republic of Korea

Keywords: LIDAR, 3D Modelling, Shape Loss, Objection Detection, Autonomous Driving, Adaptive Ground ROI Estimation.

Abstract: In this paper, we propose a 3D shape loss function for improved 3D object detection for LIDAR data. As the LiDAR (Light Detection And Ranging) sensor plays a key role in many autonomous driving techniques, 3D object detection using LiDAR data has become an important issue. Due to inaccurate height estimation, 3D object detection methods using LiDAR data produce false positive errors. We propose a new 3D shape loss function based on 3D masks for improved performance. To accurately estimate ground ROI areas, we first apply an adaptive ground ROI estimation method to accurately estimate ground ROIs and then use the shape loss function to reduce false positive errors. Experimental shows some promising results.

1 INTRODUCTION

In autonomous driving techniques, object detection is a key element (Simony, 2018; Shi, 2019; Lang, 2019). Although vision-based object detection methods have several advantages in terms of cost and flexibility (Wang, 2021; Bochkovskiy, 2020; Wang, 2022), they tend to produce errors under poor conditions such as backlighting, dark scene, and sudden illumination changes (Xu, 2020; Jeong, 2021; Xu, 2020). On the other hand, LiDAR-based 3D object detection methods provide more reliable performance under those challenging conditions.
However, the LiDAR-based methods, which use the entire point cloud (PC), also showed some limitations in real-time processing(Shi, 2019). Since the MV3D method was proposed (Chen, 2017), many researchers have studied 3D object detection methods using BEV(Bird's Eye View) (Yang, 2018; Simony, 2018). However, converting 3D information of LiDAR data to 2D BEV, some features were lost, which may produce some errors. When BEV images are produced, the height information is permanently lost. From the BEV images, the ground ROI (region of interest) is estimated. Since the goal is to estimate 3D boxes of targets (cuboids), the height is estimated as the average height values of the PC sample points within the cuboid. Fig. 1 illustrates this procedure. Fig. 1(a) is a point cloud and Fig. 1(c) shows a BEV image with 2D bounding boxes. Fig. 1(b) is the estimated 3D object cuboids. However, this procedure tends to produce many false positive (FP) errors. Fig. 2 shows such false positive errors of the complex YOLO algorithm (Simony, 2018).

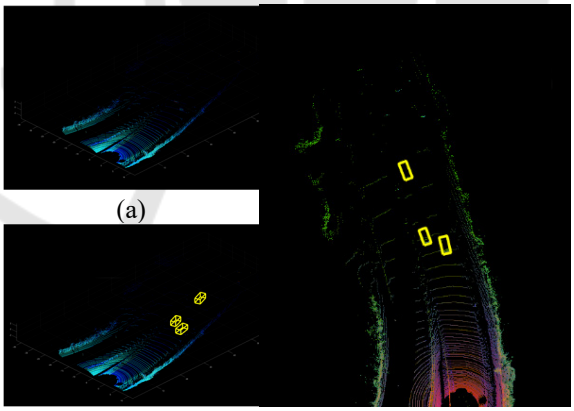


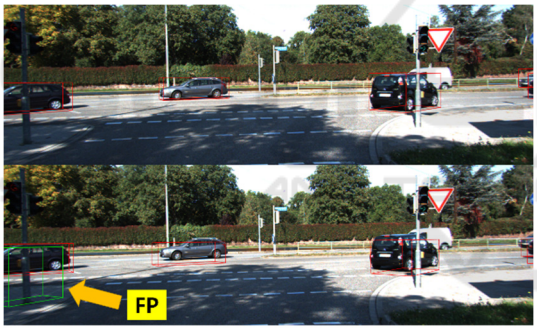
Figure 1: Examples of LiDAR data. (a) point cloud, (b) 3D object cuboid, (c) BEV image with 2D bounding boxes.

In order to reduce this kind of false positive error, we propose to use 3D shape masks to compute a 3D shape loss function for improved 3D object detection for LIDAR data.

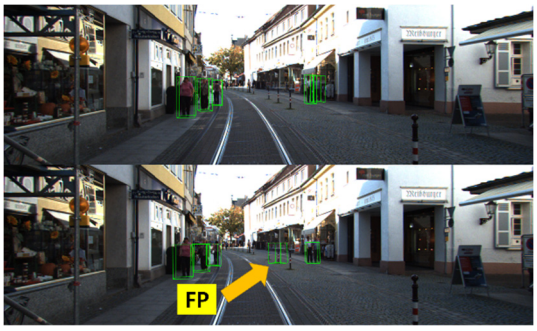
Park, R. and Lee, C.
3D Mask-Based Shape Loss Function for LIDAR Data for Improved 3D Object Detection.
DOI: 10.5220/0011966800003479
In *Proceedings of the 9th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2023)*, pages 305-312
ISBN: 978-989-758-652-1; ISSN: 2184-495X

(a)

(b)



(c)



(d)

Figure 2: Examples of false positive errors of complex YOLO (top: ground truth, bottom: false positive errors of the complex YOLO algorithm). The red cuboids represent cars whereas the green cuboids represent pedestrians.

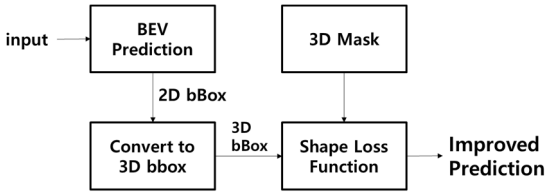


Figure 3: Flowchart of the proposed method.

Figure 4: Incorrect 3D localization examples (top: ground truth, bottom: outputs of the complex YOLO algorithm).

2 PROPOSED METHOD

Fig. 3 shows a flowchart of the proposed method. First, we apply an adaptive ground ROI estimation method, which produces a more accurate ground ROI estimation. Then, we estimate a target cuboid. Finally, we use 3D masks to compute a shape loss function.

2.1 Adaptive Ground ROI Estimation

We observed that some errors were caused by inaccurate estimation of ground ROIs in BEV images. Fig. 4 shows some inaccurate estimations of ground ROIs when the Complex YOLO algorithm (Simony, 2018) was used. To better estimate ground ROIs, we propose an adaptive ground ROI estimation method (Fig. 5). First, we estimate an initial ROI and then search neighbour areas to produce an improved estimation using a ground prediction algorithm (Pingel, 2013). With this adaptive ground ROI estimation method, the incorrect localization errors (incorrect ground ROI estimation) were noticeably reduced as can be seen in Fig. 9.

2.2 3D Shape Loss Function Based on 3D Masks

In conventional methods, the cuboid height is estimated based on the average value of LiDAR samples (z-direction). However, this estimation method may produce some erroneous results. In particular, it may produce some false positive errors as can be seen in Fig. 2.

In order to solve this problem, we propose to use 3D masks for the three major objects: car, pedestrian, and cyclist. Fig. 6 shows the 3D masks used in this paper. Using the LiDAR points within a candidate cuboid, we computed the shape loss function as follows:

$$loss_{shape} = \frac{1}{N}\sum_{i=1}^{N} \min_{k}(\|p_i - \{\text{reference point cloud}\}\|)$$

where p_i is the i-th point of the candidate cuboid, $\{\text{reference point cloud}\}$ is a set of the 3D mask points, N is the number of points of the candidate cuboid. Fig. 7 shows the histogram of the shape loss function of the three 3D masks.

In order to normalize the values of the shape loss function, we used the following normalization function so that the range is between 0 and 1:

$$loss_{shape}^{normalized} = 1 - sigmoid(loss_{shape}).$$

Fig. 8 shows the graph of the normalization function.

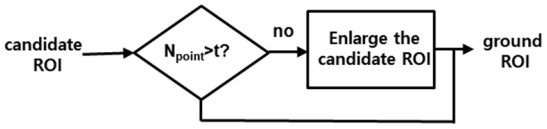


Figure 5: Adaptive ground ROI estimation (N_p: number of points with the candidate ROI).

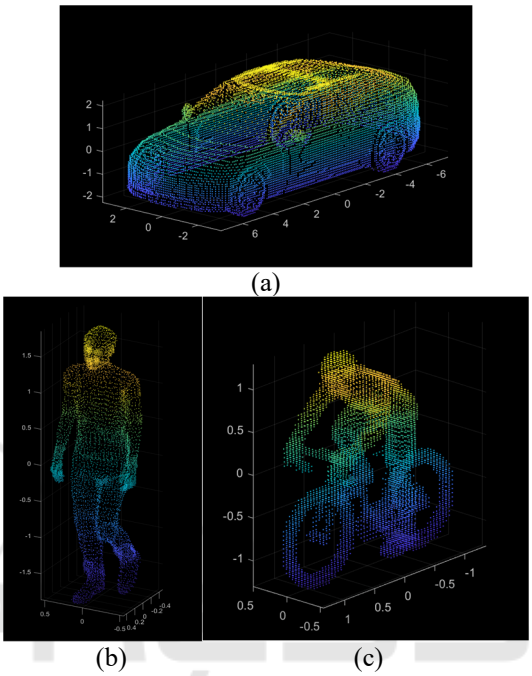


Figure 6: 3D masks. (a) car, (b) pedestrian, (c) cyclist.

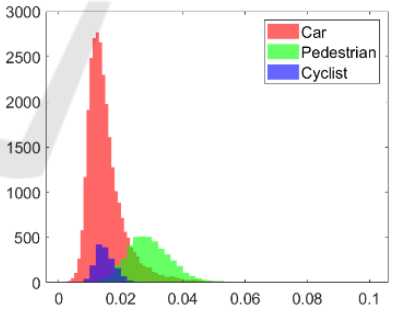


Figure 7: Histogram of the shape loss function of the three 3D masks.

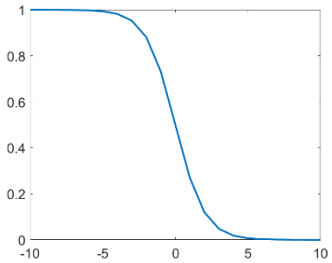


Figure 8: Normalization function.

Figure 9: Improved localization (ground ROI estimation) of the proposed adaptive ground ROI estimation method (top: ground truth, middle: outputs of the complex YOLO algorithm, bottom: improved ground ROI estimation of the proposed adaptive ground ROI estimation method.

Figure 10: Improvement performance of the proposed method that uses the shape loss function with reduced false positive errors (top: ground truth, middle: outputs of the complex YOLO algorithm, bottom: proposed method.

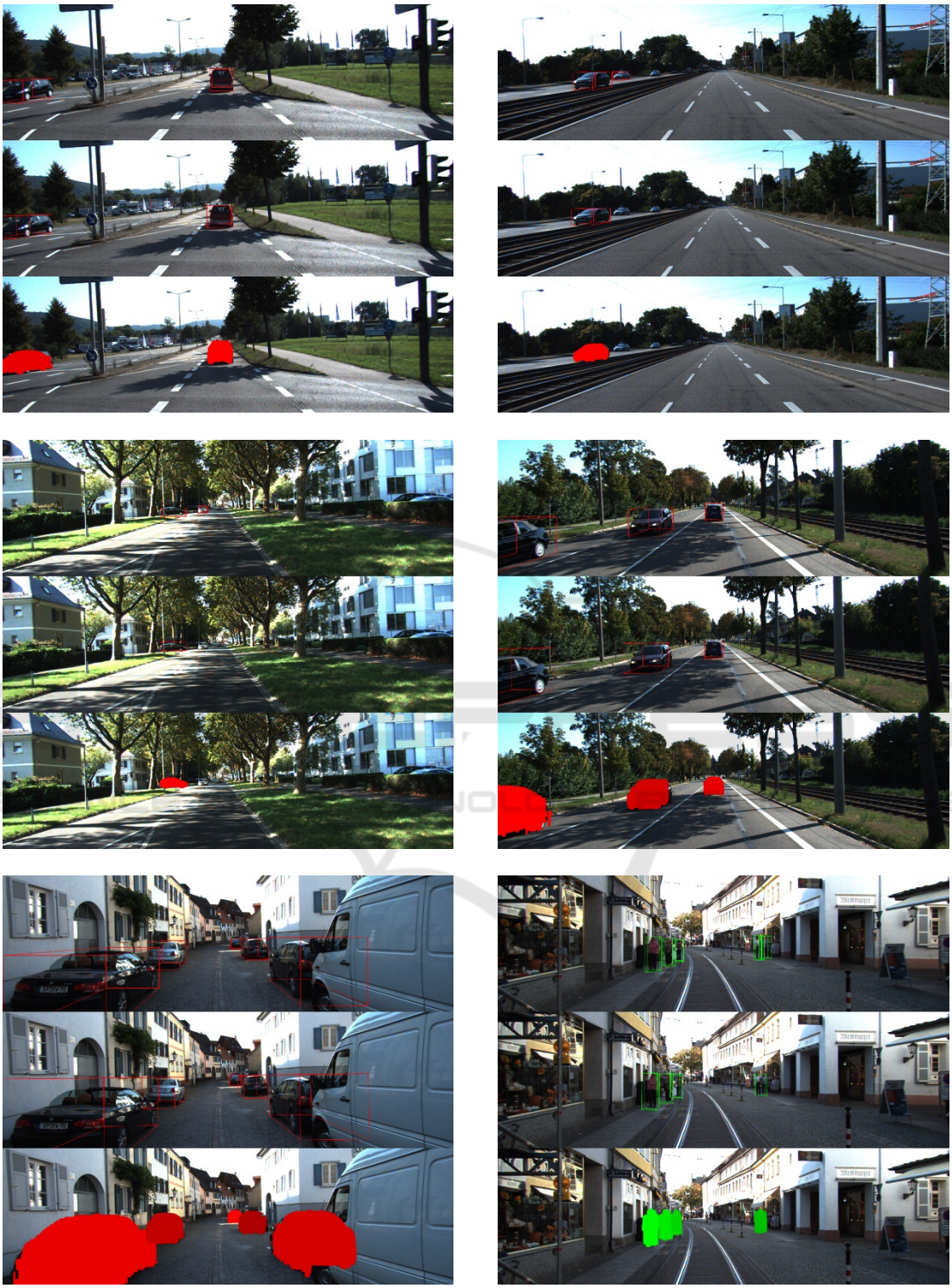


Figure 11: Improvement instance segmentation of the proposed method.

3 EXPERIMENTAL RESULTS

3.1 Dataset and Results

Experiments were performed using the KITTI Dataset, which is widely used in 3D object detection research, was used (Geiger, 2012).

The KITTI Dataset consists of 7481 images with 9 classes. We used 3 major classes (car, pedestrian, and cyclist) for autonomous driving. We used 70% of the KITTI dataset for training and the remaining 30% for validation. Although the complex YOLO (YOLOv2) was used, the prediction model was designed using YOLOv4 (Bochkovskiy, 2020). Tables 1-2 show the performance comparison between the proposed algorithm and the complex YOLO algorithm in terms of AOS (Average Orientation Similarity) (Geiger, 2012) and AP (Average Precision) (Everingham, 2010; Geiger, 2012). It is noted that both AP and AOS metrics consider the result is correct if a predicted box overlaps by at least 50% with a ground truth bounding box. Thus, the metrics of Tables 1-2 may not fully reflect more accurate bounding box estimations of the proposed method.

Fig.10 shows some prediction output images obtained by applying the proposed method and the complex YOLO algorithm. Green cuboids represent people, and red cuboids represent vehicles. It can be seen that the proposed algorithm noticeably reduced false positive errors.

Table 1: Performance comparison (AOS).

Model	Car	Pedestrian	Cyclist
complex YOLO	0.729	0.406	0.573
Proposed	0.730	0.418	0.579

Table 2: Performance comparison (AP).

Model	Car	Pedestrian	Cyclist
complex YOLO	0.780	0.413	0.582
Proposed	0.782	0.425	0.588

3.2 Instance Segmentation

Since the proposed method based on 3D masks can produce accurate 3D boundaries, we can generate accurate instance segmentation, whereas the conventional methods can only produce 3D bounding boxes (cuboids) that provide approximate 3D locations of target objects. Fig. 11 shows some instance segmentation results of the proposed method.

4 CONCLUSIONS

The LiDAR sensor can provide important information for 3D object detection in autonomous driving methods. Using the LiDAR sensor, one can overcome the reliability issues of vision-based objection methods. However, 3D object detection methods based on BEV images of LiDAR data have some other problems such as inaccurate ground ROI estimation and false positive errors. We propose to use a 3D shape loss function based on 3D masks for three major targets. Although experimental results show some promising results, one can improve the performance by using more diverse 3D masks.

ACKNOWLEDGEMENTS

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2020R1A2C1012221).

REFERENCES

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Chen, X., Ma, H., Wan, J., Li, B., & Xia, T. (2017). Multi-view 3d object detection network for autonomous driving. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 1907-1915).

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2), 303-338.

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? the KITTI vision benchmark suite. In 2012 IEEE conference on computer vision and pattern recognition (pp. 3354-3361).

Jeong, Y. (2021). Predictive lane change decision making using bidirectional long shot-term memory for autonomous driving on highways. IEEE Access, 9, 144985-144998.

Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., & Beijbom, O. (2019). Pointpillars: Fast encoders for object detection from point clouds. In Proceedings of

the IEEE/CVF conference on computer vision and pattern recognition (pp. 12697-12705).

Pingel, T. J., Clarke, K. C., & McBride, W. A. (2013). An improved simple morphological filter for the terrain classification of airborne LIDAR data. ISPRS Journal of Photogrammetry and Remote Sensing, 77, 21-30.

Shi, S., Wang, X., & Li, H. (2019). Pointrcnn: 3d object proposal generation and detection from point cloud. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 770-779).

Simony, M., Milzy, S., Amendey, K., & Gross, H. M. (2018). Complex-yolo: An euler-region-proposal for real-time 3d object detection on point clouds. In Proceedings of the European Conference on Computer Vision (ECCV).

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2021). Scaled-yolov4: Scaling cross stage partial network. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13029-13038).

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696.

Xu, Z. F., Jia, R. S., Liu, Y. B., Zhao, C. Y., & Sun, H. M. (2020). Fast method of detecting tomatoes in a complex scene for picking robots. IEEE Access, 8, 55289-55299.

Xu, Z. F., Jia, R. S., Sun, H. M., Liu, Q. M., & Cui, Z. (2020). Light-YOLOv3: fast method for detecting green mangoes in complex scenes using picking robots. Applied Intelligence, 50(12), 4670-4687.

Yang, B., Luo, W., & Urtasun, R. (2018). Pixor: Real-time 3d object detection from point clouds. In Proceedings of the IEEE conf. on CVPR (pp. 7652-7660).