MFSCNN: APPENDING A MASKED BRANCH TO FAST-SCNN TO IMPROVE ROAD MARKING EXTRACTION ON SPARSE MLS POINT CLOUD-DERIVED IMAGES

Miguel Luis R. Lagahit * and Masashi Matsuoka

Department of Architecture and Building Engineering, Tokyo Institute of Technology, Tokyo, Japan
Tokyo Tech Academy for Super Smart Society, Tokyo Institute of Technology, Tokyo, Japan
* lagahit.m.aa@m.titech.ac.jp

KEY WORDS: Low-Cost LiDAR, Mobile Mapping, Road Marking Extraction, Point Cloud-Derived Images, Deep Learning

ABSTRACT:

With the rise of self-driving cars, an increasing number of vehicles are equipped with low-cost light detection and ranging (LiDAR) sensors that could potentially serve as a massive mobile mapping resource, particularly for jobs that require multiple and frequent scanning, such as maintaining dynamic high-definition maps or digital twins. However, low-cost LiDAR sensors produce sparser point clouds during scanning which can make deep learning techniques for the automatic retrieval of features difficult like extracting road markings. In this work, we aim to improve the performance of a convolutional neural network (CNN) model for road marking extraction from sparse mobile LiDAR scanning (MLS) point cloud-derived images. We propose the modification of the Fast-SCNN model structure by adding a 2D convolution branch with masking in the feature fusion step: MFSCNN. To retain speed we only use MFSCNN to boost model training and still utilize Fast-SCNN for inference. Our results indicate potential, with a 4.6% increase in mean f1-score and an 8% decrease in uncertainty for the road marking class after multiple trials. Additionally, this research aims to support and increase research interest in lower-cost LiDARs for mobile mapping.

1. INTRODUCTION

1.1 Background

Successful extractions of road markings from dense point cloud-derived images using convolutional neural networks (CNN) have been demonstrated in many works, such as in Wen et al. (2019) and Lagahit and Tseng (2020) among others. These works take advantage of the road marking's reflective property. Which provides strong intensity values during light detection and ranging (LiDAR) scanning, clearly distinguishing them from other features. Recently, this approach has extended to sparse point cloud-derived images from low-cost mobile LiDAR scanning (MLS) as an alternative to expensive mobile mapping systems (Lagahit and Matsuoka, 2023). This was done in an attempt to utilize low-cost LiDAR sensors onboard self-driving vehicles as a mobile mapping resource for updating and making digital twins or high-definition (HD) maps more dynamic.

However, since points clouds generated by low-cost LiDARs during MLS are sparse, the road markings become poorly represented and hardly identifiable. This situation makes it challenging for CNN models to extract the desired road marking features. One way of addressing this issue is to tweak the deep learning framework to be more suitable for detecting harder-to-classify features. Lagahit and Matsuoka (2013), tackled the loss function, which aids guide the model during training, and showed promising results. However, other aspects of the framework have yet to be explored, such as the structure of the CNN itself.

Currently, there are already a multitude of CNN models with varying structures available, U-Net and Fast-SCNN to name a few (Poudel et al., 2019; Ronneberger et al., 2015). Both of these models have already demonstrated potential in extracting road markings from sparse MLS point cloud-derived images (Lagahit and Matsuoka, 2023). It is worth noting, however, that Fast-SCNN, which was built for real-time segmentation, has shown 15x quicker prediction speeds than U-Net. Given that one of the envisioned applications is map updating, speed would be an advantageous feature. Unfortunately, Fast-SCNN still provides poorer segmentation accuracy than that of U-Net.

Anchoring on Fast-SCNN's speed, we propose a modification to its structure in an early attempt to improve its classification capabilities. As was tackled in its paper, we will introduce an additional 2D convolutional branch in the feature fusion procedure. But, we will also be adding a masking procedure to this branch in order to retain only regions with corresponding point cloud values in order to control misclassifications and strengthen extractions in such areas.

1.2 Objective

The goal of this study is to improve road marking extraction on sparse MLS point cloud-derived images by proposing a Masked-Fast-SCNN (MFSCNN), a modified version of the Fast-SCNN model, for boosting model training. Moreover, the following has been done in support of the proposed model: (1) the extractions are compared to those of Fast-SCNN, (2) different cases of the additional branch have been analyzed, (3) multiple kernel sizes of the masking procedure have been investigated, (4) results of the masking procedure implementation in Fast-SCNN have been tested, and (5) prediction speeds have been observed.

2. METHODOLOGY

2.1 Dataset Gathering and Preparation

The dataset for this study was collected using a low-cost Robosense 16-channel LiDAR, positioned in front of a vehicle tilted 45 degrees down, inside the North Ookayama Area of Tokyo Institute of Technology. The point clouds were initially restricted to an area in front before projecting top-down to a 2D plane with a ground resolution of 1 by 1 centimeter and a size of 2048 by 512 pixels. The roadways had several road markings but mainly lane lines and crosswalks. Moreover, because the original images depict sparse features, all subsequent sparse MLS point

cloud-derived images shown on this paper have been dilated with a 3x3 kernel for better visualization, as shown in Figure 5-1.

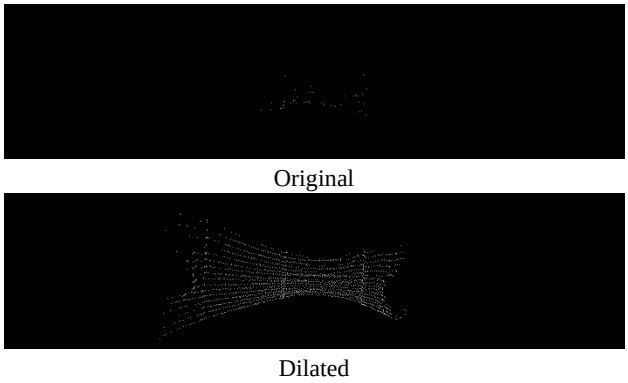


Figure 2-1. Sample sparse MLS point cloud-derived image using intensity as pixel values. (Top) Original image and (Bottom) dilated version for visualization purposes.

To train and test the model a manually labelled dataset, assisted by intensity thresholding, was created for this study. It is divided unto three class categories: (1) 'black' which are the black pixels with no point cloud value, (2) 'others' which are the white pixels with non-road marking features, and (3) 'road marking' which are green pixels with the target road markings.

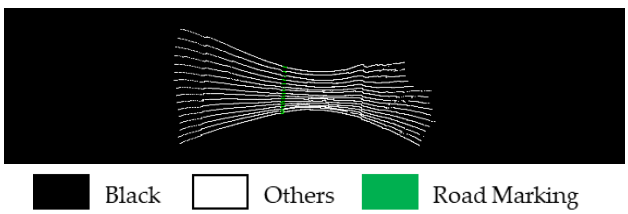


Figure 2-2. A sample labeled sparse MLS point cloud-derived image showcasing the three classes.

The dataset statistics are shown in Table 2-1. A total of 1200 intensity and labeled image pairs were produced after undergoing simple augmentation (e.g. flipping). A distribution of roughly 80%, 10%, and 10% are used for training, validation, and testing datasets, respectively. Looking at the pixel distribution, it is evident that the 'black' class dominates the dataset while the target 'road marking' class is barely present.

Table 2-1. Dataset statistics.

Dataset	Number of Images	Number of Pixels per Class		
		Black	Others	Road Marking
Training	1000	99.13%	0.84%	0.03%
Validation	100	99.10%	0.86%	0.04%
Testing	100	99.16%	0.80%	0.04%

2.2 CNN Structure and Model Training

The base structure of the proposed MFSCNN is based on the Fast-SCNN model, indicated by the black arrows in Figure 2-3. Fast-SCNN was designed for real-time segmentation utilizing techniques like inverted residual bottlenecks among others (Poudel et al., 2019). This paper attemps to make use of Fast-SCNN's speed and improve its accuracy by appending an additional masked 2D convolutional branch to control misclassifications, indicated by the blue arrows in Figure 2-3. The idea stems from the paper of Lagahit and Matsuoka (2013), wherein misclassifications in areas with no corresponding point cloud values are removed.

To further support the model's capability in detecting sparsely represented road markings, Combo loss will be used for the loss function. The loss function takes the difference between initial predictions and the labels and uses it, after going through an activation function and in an optimizer, to adjust the weights of the model. Combo loss, is a loss function that takes the weighted sum of two loss functions to take advantage of their individual properties in order to improve the model's capability in detecting harder features such as our target road marking on sparse MLS point-cloud derived images (Taghanaki et al., 2019).

$$\text{Combo Loss} = \alpha\,(Modified\ Cross - Entropy\ Loss) + (1-\alpha)\,(Dice\ Loss), \quad (1)$$

The proposed method was implemented using python on a computer with an 11th Gen Intel i7 processor, 32 GB of RAM, and an NVIDIA GeForce RTX 3060 Laptop GPU. During training the following hyperparameters were set: batch size of 16, a learning rate of 1×10^{-4}, and an Adam optimizer. Furthermore, each models were trained three times to determine uncertainty, using a fixed seed value for each trial for good comparison in the different CNN structures. Finally, 100 epochs were used for all trials, using the model with lowest loss value.

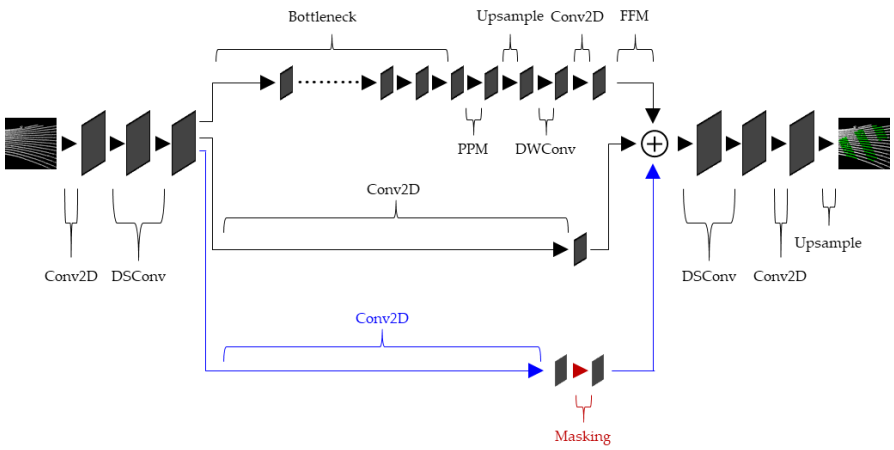


Figure 2-3. MFSCNN structure.

2.3 Assessment

The performance of MFSCNN to conduct road marking extraction on sparse MLS point cloud-derived images will be assessed using precision and recall, which are computed from the confusion matrix when comparing the CNN predictions to its corresponding image labels, as shown in Equations 2 and 3. In addition, when precision and recall values are far apart, the f1-score, which is the harmonic mean of precision and recall, will be used to act as a final evaluation criterion, as shown in Equation 4. Because the harmonic mean leans toward the smaller value among the inputs, it also serves as a reliable criterion for properly evaluating the classified images (Powers, 2011).

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}, \quad (2)$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}, \quad (3)$$

$$\text{F1}_{\text{score}} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

As Lagahit and Matsuoka (2023) have observed, since the pixels in the black class regions do not reflect any point cloud value they can be omitted in the assessment computations, as depicted in Figure 2-4. This is important because of the severe class imbalance at hand. Huge numbers of misclassifications in the black regions will greatly influence the resulting evaluation, and failure to remove them can be misleading to the intended final output, which is a classified sparse point cloud.

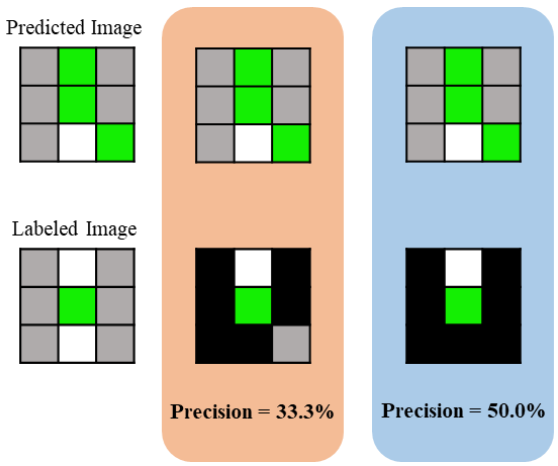


Figure 2-4. Precision value differences when computed without (orange box) and with (blue box) the removal of misclassifications in the black region.

3. RESULTS AND DISCUSSION

3.1 Comparison with Fast-SCNN

In Figure 3-1, we show some selected sample results among multiple trials to showcase the performance of MFSCNN in comparison to that of Fast-SCNN and the reference labeled image. We also present the raw prediction results as well as the projected results, where misclassifications in the black regions are removed to highlight and depict only those road marking pixels with corresponding point cloud values. We can observe that there are cases where MFSCNN can depict road marking geometry far better off than Fast-SCNN.

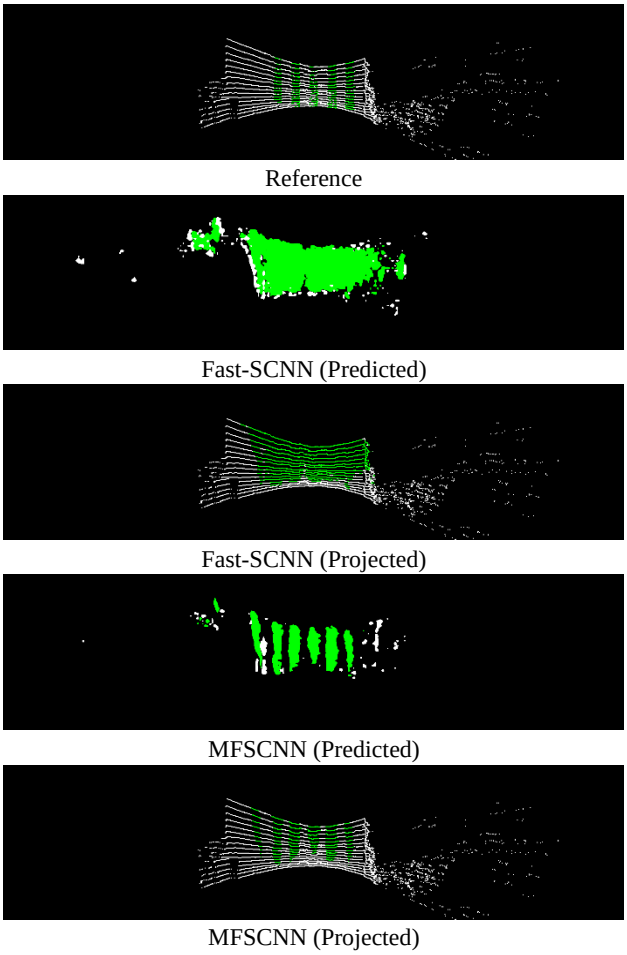


Figure 3-1. Sample predictions of Fast-SCNN and MFSCNN.

Table 3-1 shows the corresponding numerical evaluation of the resulting performance of MFSCNN as compared to Fast-SCNN in road marking extraction from sparse MLS point cloud-derived images. The resulting evaluations of the raw predictions and their projected counterparts are also shown side by side, revealing a roughly 40% difference in the mean f1-scores, emphasizing the importance of excluding misclassifications in the black regions in the resulting evaluations. Looking at the projected evaluation, we could see an increase in precision but a decrease in recall. Due to the contradiction, we could take a look at the f1-score, and we can see that MFSCNN not only gains a 3.8% increase in mean f1-score but also gains an 8% decrease in uncertainty, resulting in more accurate and dependable road marking extractions.

Table 3-1. Evaluation results for the road marking class of MFSCNN as compared to Fast-SCNN (%).

Model	Recall	Predicted		Projected	
		Precision	F1-Score	Precision	F1-Score
Fast-SCNN	**59.3 ± 9.3**	4.2 ± 1.0	7.8 ± 1.7	40.3 ± 10.2	46.7 ± 9.8
MFSCNN (Kernel=5)	56.8 ± 10.6	**5.2 ± 0.4**	**9.5 ± 3.4**	**46.4 ± 0.5**	**50.5 ± 1.8**

3.2 Analyzing the Additional Branch

In this sub-chapter, we compare the results of our proposed additional branch as compared to adding and making use of the main branch of Fast-SCNN as well as their masked versions to justify the proposed additional branch and see the significance of the masking procedure. Figure 3-2, shows the trial cases and Figure 3-3 and Table 3-2 shows the corresponding results. We will refer to the additional Conv2D branch as '+Ghost' and the additional branch with multiple procedures as '+Main'. Furthermore, an additional '-Mask' will be placed in the name if the masking procedure is in place.

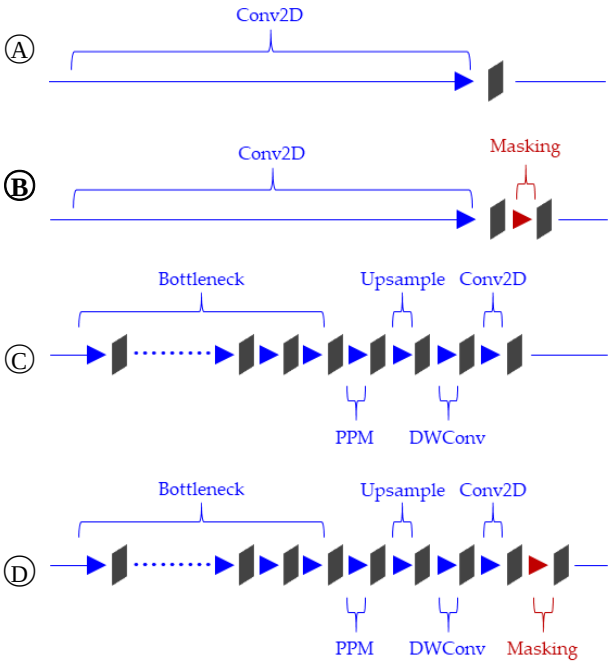


Figure 3-2. Ⓐ and Ⓒ depicts the +Ghost and +Main branches while Ⓑ and Ⓓ depict their corresponding masked versions.

From the sample projected predictions below, it is still clear that no matter the additional branch, they all perform relatively poorly for lane lines that illustrate extreme class imbalance. However, taking a look at the numerical evaluations of multiple trials we can see that an additional '+Ghost-Mask' branch provides the highest precision and recall values. This follows that it would also achieve the best mean f1-score. Also, after factoring in uncertainty it has the highest minimum f1-score, highlighting its consistency and thus its capability to increase the dependability of the CNN in its predictions.

Table 3-2. Evaluation results of MFSCNN for the road marking class as compared to other additional branch cases (%).

Branch	Recall	Precision	F1-Score
+Ghost	52.1 ± 22.1	46.0 ± 7.1	46.0 ± 7.8
+Ghost-Mask (MFSCNN)	**56.8 ± 10.6**	**46.4 ± 0.5**	**50.5 ± 1.8**
+Main	50.9 ± 22.1	37.1 ± 10.8	39.4 ± 1.2
+Main-Mask	52.4 ± 3.6	42. 2 ± 0.4	46.8 ± 1.3

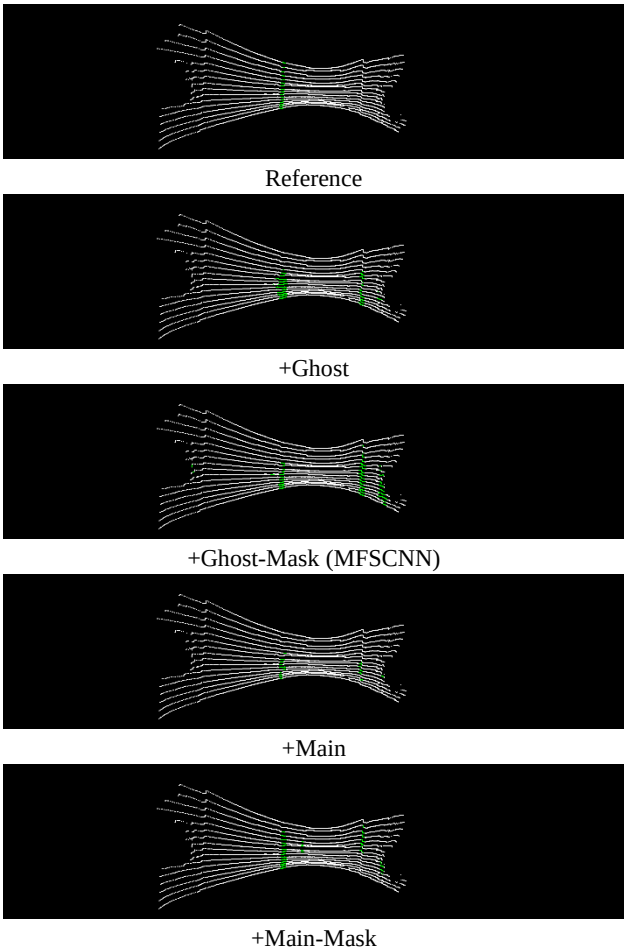


Figure 3-3. Sample projected predictions of MFSCNN and other additional branch cases.

3.3 Analyzing Masks at Varying Dilation Kernel Sizes

In this sub-chapter, we investigate the effects of different kernel sizes when dilating the image mask used in the masking procedure. In the previous results shown, a kernel size of 5x5 was used, as such we will be exploring one step smaller and larger kernel sizes of 3x3 and 7x7, respectively. As seen in Table 3-3, all kernel sizes outperform the resulting mean f1-score of Fast-SCNN by 0.6% to 4.6%. Moreover, the resulting uncertainty also proves to be better by 2% to 8% in all cases. It is also interesting to see that as the kernel size increase so does the resulting mean f1-score, however the difference in improvements did get smaller meaning that at a certain kernel size larger than 7x7, MFSCNN's performance could decline. Visually, larger kernels also tend to hinder overreaching misclassifications, where neighboring pixels are misidentified as the target feature, as seen in Figure 3-4.

Table 3-3. Evaluation results of MFSCNN for the road marking class at varying kernel sizes of the masking procedure (%).

Kernel Size	Recall	Precision	F1-Score
3	70.1 ± 14.7	37.7 ± 11.0	47.3 ± 7.8
5	**56.8 ± 10.6**	46.4 ± 0.5	50.5 ± 1.8
7	55.7 ± 12.7	**49.4 ± 5.7**	**51.3 ± 2.2**

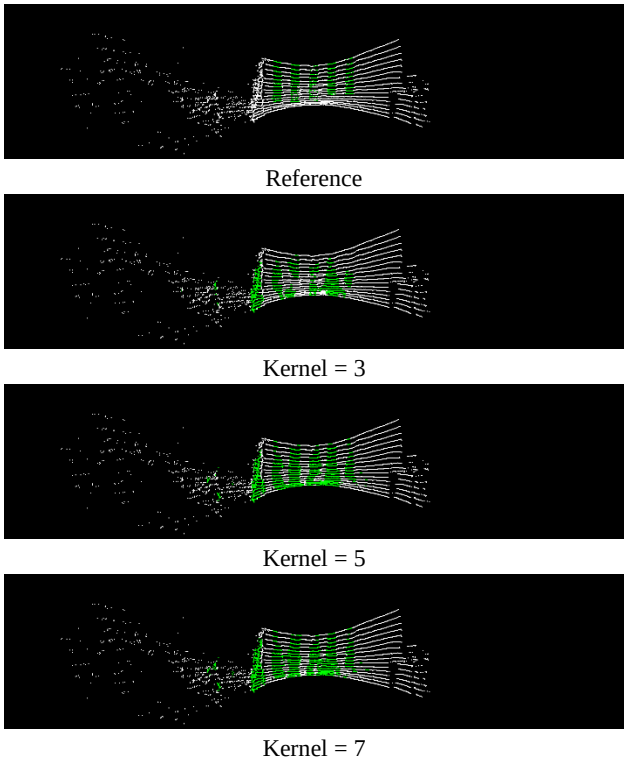
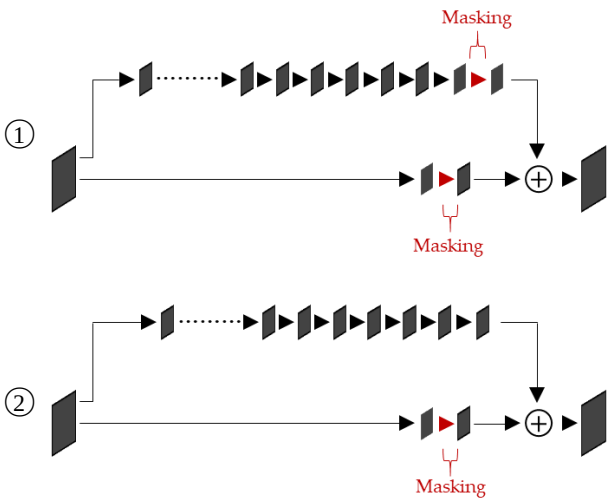


Figure 3-4. Sample projected predictions of MFSCNN at varying kernel sizes of the masking procedure.

3.4 Comparison with Masking Fast-SCNN

In this sub-chapter, for a fairer comparison, we take a look at the effects of the masking procedure in Fast-SCNN. Similar to the naming convention in the additional branches, we call the branch with the 2D convolution 'ghost', the branch with the multiple procedures 'main', and simply 'both' for the two branches. Figure 3-5 illustrates the cases where the masking procedure has been implemented. Like the proposed additional branch all of them are implemented before fusing.



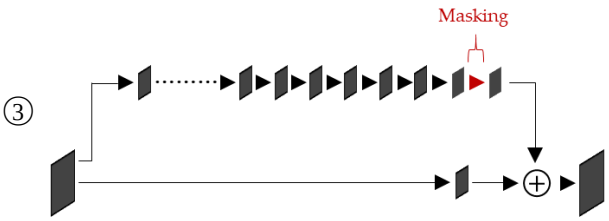


Figure 3-5. The masking procedure has been implemented in the Fast-SCNN model before the end of the feature fusion step in the cases: ① of both the main and ghost branches, ② only the ghost branch, and ③ only the main branch.

Figure 3-6 show the sample projected prediction results of doing the masking procedure within Fast-SCNN and Table 3-4 show the corresponding numerical assessment. All resulting mean f1-score fall behind that of both Fast-SCNN and MFSCNN. This implies that masking alone deteriorates performance and instead should be added as an additional factor to improve and control the misclassifications of Fast-SCNN.

Table 3-4. Evaluation results of Fast-SCNN for the road marking class with the masking procedure (%).

Branch	Recall	Precision	F1-Score
Both	70.1 ± 46.1	25.5 ± 30.9	18.3 ± 7.2
Ghost	65.5 ± 33.4	34.0 ± 28.2	32.5 ± 33.9
Main	88.8 ± 11.2	15.8 ± 9.3	25.7 ± 11.8

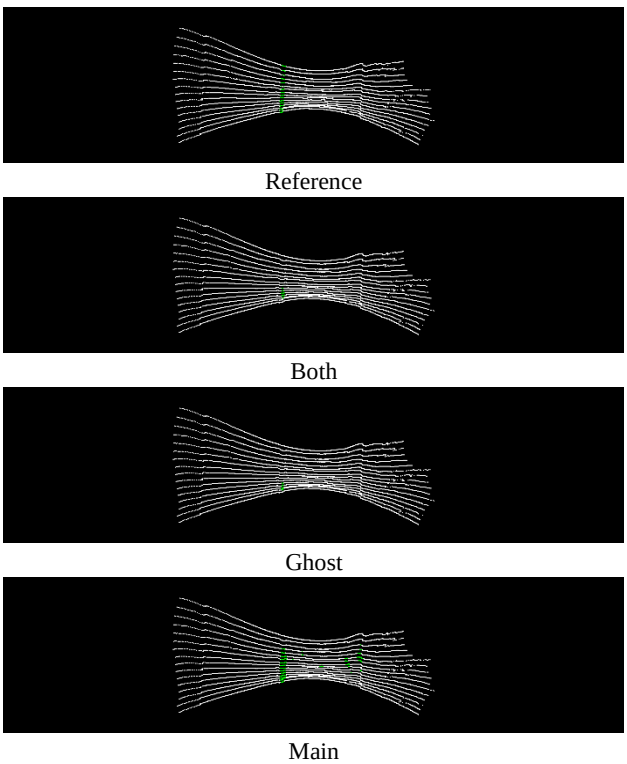


Figure 3-6. Sample projected predictions of Fast-SCNN with the masking procedure.

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-829-2023

3.5 Prediction Speeds

In this sub-chapter, we observe the prediction speeds of Fast-SCNN and MFSCNN, to support the use of Fast-SCNN during the inference step. For this test, a tensor with the same size as the input image in our dataset is done 5 times for the models to get our prediction speeds. Unfortunately, MFSCNN is 1.5x slower than Fast-SCNN, as shown in Table 3-5. So, in attempting near real-time or real time predictions, Fast-SCNN is still recommended but it is advised to be trained with MFSCNN.

Table 3-5. The resulting prediction speeds of MFSCNN in comparison to Fast-SCNN (seconds).

Model	Speed
Fast-SCNN	0.182 ± 0.003
MFSCNN	0.280 ± 0.003

4. CONCLUSION

In this paper, we have attempted to improve road marking extraction from sparse MLS point cloud-derived images through our proposed MFSCNN, which adds a 2D convolution branch with masks to Fast-SCNN. Our results have shown that at varying kernel sizes, our proposed model was able to produce a maximum of 4.6% increase in mean f1-score and an 8% decrease in uncertainty after multiple trials. Extensive analysis has also shown MFSCNN outperformed varying additional branch cases as well as varying masking implementations on Fast-SCNN. However, due to the additional branch MFSCNN became slower as compared to Fast-SCNN, so using MFSCNN only to boost model training and using Fast-SCNN at the inference step is recommended to retain speed but improve accuracy as was demonstrated. Overall, along with previous works of using more suitable loss functions during training, little by little, this work contributes to the improvements in road marking extraction from sparse MLS point cloud-derived images for the goal of supporting the utilization of lower-cost LiDAR alternatives as a practical approach for mobile mapping tasks. In future work, further modifications in the CNN structure would be explored such as the addition of image processing procedures and the reduction of branches.

ACKNOWLEDGMENTS

The research was supported by the WISE-SSS program of Tokyo Institute of Technology. A special thanks go to Zongdian Li of Sakageuchi-Tran Laboratory, Department of Electrical and Electronics Engineering, Tokyo Institute of Technology, who assisted and provided expertise during data collection.

REFERENCES

Lagahit, M.L.R., Matsuoka, M. 2023. Focal Combo Loss for Improved Road Marking Extraction of Sparse Mobile LiDAR Scanning Point Cloud Derived Images using Convolutional Neural Networks. Remote Sensing, 15, 597.

Lagahit, M.L.R., Tseng, Y.H., 2020. Using Deep Learning to Digitize Road Arrow Markings from LIDAR Point Cloud Derived Images. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 127-141. doi.org/10.5194/isprs-archives-XLIII-B5-2020-123-2020.

Poudel, R., Liwicki, S., Cipolla, R. 2019. Fast-SCNN: Fast Semantic Segmentation Network. arXiv.

Powers, D.M.W. 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. International Journal of Machine Learning Technology 2:1, pp. 37-63. Retrieved January 4, 2022, from the arXiv database.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Computing Research Repository. Retrieved January 4, 2022, from the arXiv database.

Taghanaki, S., Zheng, Y., Zhou, S.K., Georgescu, B., Sharma, P., Xu, D., Comaniciu, D., Hamarneh, G. 2019. Combo Loss: Handling Input and Output Imbalance in Multi-Organ Segmentation. Computerized Medical Imaging and Graphics, 75, 24-33.

Wen, C., Sun, X., Li, J., Cheng, W., Guo, Y., Habib, A., 2019. A deep learning framework for road marking extraction, classification, and completion from mobile laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, vol.147 pp. 178-192. doi.org/10.1016/j.isprsjprs.2018.10.007.