BOOSTING U-NET WITH FOCAL LOSS FOR ROAD MARKING CLASSIFICATION ON SPARSE MOBILE LIDAR POINT CLOUD DERIVED IMAGES

M. L. R. Lagahit*, M. Matsuoka

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

KEY WORDS: Mobile Mapping, LIDAR Point Cloud, Road Marking, Image Classification, Deep Learning, Focal Loss

ABSTRACT:

Road markings play an important role in vehicular navigation. It helps provide sufficient information for safe driving and smooth traffic flow. As such, with the rise of digital maps such as High-Definition (HD) maps, which are used by autonomous vehicles or self-driving cars, they must be well represented in their digital counterparts. However, survey-grade mobile mapping systems are expensive and thus open the idea of using lower-cost/level LIDAR sensors for mapping. Unfortunately, using such sensors provide sparser point clouds. This work aims to propose a method that successfully classifies road markings on sparse mobile LIDAR point cloud-derived images using UNET trained with focal loss. Results have shown successful road marking classification with a 94.68% increase in recall and a maximum 49.39% increase in F1-score. Adjusting precision by removing the insignificant class ("black") further increases the resulting F1-score to 82.74%. Extending the method produces a classified point cloud by combining the classified image with a depth image. This research also aims to help aid boost the research on lower-cost/level sensors for mobile mapping purposes.

1. INTRODUCTION

1.1 Background

Road markings are an essential component in navigating the roadway. They provide the necessary information to guide a vehicle on how it should act to ensure safety and a good flow of traffic. For example, road arrows can provide driving directions and pedestrian cross marks can indicate possible non-vehicular interactions. As such, it must be properly and accurately delineated in virtual representations of our society, like High-Definition (HD) Maps. HD Maps are centimeter-level 3D digital maps that are used by autonomous vehicles (or self-driving cars) and also act as a source of urban inventory (Liu et al, 2020).

There are already many existing types of research on how to automatically extract road markings from raw/processed mobile sensor data. One of which is using deep learning in classifying road markings from LIDAR point cloud-derived images. When projected as top-down view images, these dense point clouds provide clear and detailed representations of the road and everything on it. As such, deep learning techniques on images such as using CNN, the U-Net model to be specific, produce excellent classification results. (Lagahit and Tseng, 2021) (Lagahit and Tseng, 2020) (Wen et al, 2019)

However, in most cases, the LIDAR sensor used for mobile mapping is of survey-grade quality. This means that the scanned point clouds are highly dense, reaching up to 1 point per centimeter cube. But, these high-level sensors can be quite expensive and thus make HD map map-making and updating extremely costly. This problem opens up the idea of exploring the use of lower-cost LIDAR sensors in HD map mapping such as those onboard autonomous vehicles.

Unfortunately, these low-cost sensors produce way sparser point clouds, which causes a poor representation of road markings and other features on the road. This can become a problem that affects the resulting classification accuracy of deep learning methods, such as that of extraction from point-cloud derived images.

1.2 Objective

The research aims to improve road marking classification of deep learning methods on sparse mobile LIDAR point cloud-derived images. This is done by proposing the use of a weighted focal loss as the loss function in training a CNN model, in this case of U-Net, as compared to that of the cross-entropy loss which is the most commonly used loss function. As an extension, a classified point cloud will be generated using the classified point cloud derived-image together with a generated depth map. Finally, this work also aims to help advance research in using low-cost/level sensors for mobile mapping and HD map mapping purposes.

2. METHODOLOGY

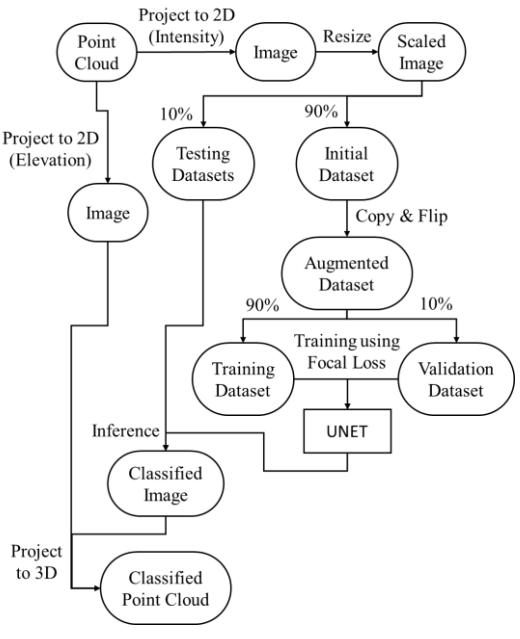

Figure 1. Proposed Workflow

The workflow for road marking classification on mobile LIDAR sparse point cloud-derived images as well as the extended generation of a classified point cloud is shown in Figure 1. Each part will be further explained in the succeeding sub-chapters.

2.1 Data Gathering

The point cloud scanning was done using a small autonomous vehicle with a Velodyne Puck (VLP-16) as the LIDAR sensor. The sensor is located in front of the vehicle tilted at 45-degrees downward from the horizontal. It was driven inside Tokyo Institute of Technology's Ōokayama campus which has roadways containing an assortment of road markings and roadway features (traffic signs, etc.).

2.2 Point Cloud to Image

Each LIDAR scan was geometrically filtered beforehand. This meant that only a fixed portion of the scanning, which is an area in front of the vehicle, is retained. This was done to reduce the number of unnecessary points (e.g. buildings, vegetation, etc.) in road marking classification, which only focuses on the roadway.

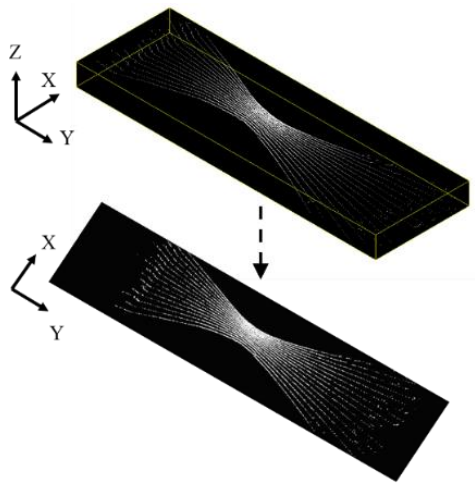

Figure 2. Point Cloud Projection to Top-Down Image.

Then, the point cloud was projected to a top-down or birds-eye-view image. This meant that Z or elevation values are disregarded and the point cloud is viewed flat from above, much like an aerial orthophoto. The corresponding ground resolution of the images is 1 cm x 1 cm, resulting in an image size of 2048 pixels x 512 pixels. Intensity values of the point cloud were used for the pixel values. When multiple points are contained in one pixel, the average intensity value of the points were used.

2.3 Annotation and Augmentation

(a)

(b)

(c)

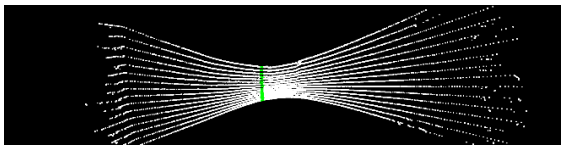

Figure 3. Image Annotation Process: (a) Intensity Filtering, (b) Manual Cleaning, and (c) Labelling.

An annotated dataset was generated for this work. The images were initially filtered by a certain intensity range to isolate road markings as much as possible. Then, the images were manually cleaned to remove non-road marking pixels that were not filtered out in the preceding step. Finally, the identified road markings were used to label the road markings in the original image.

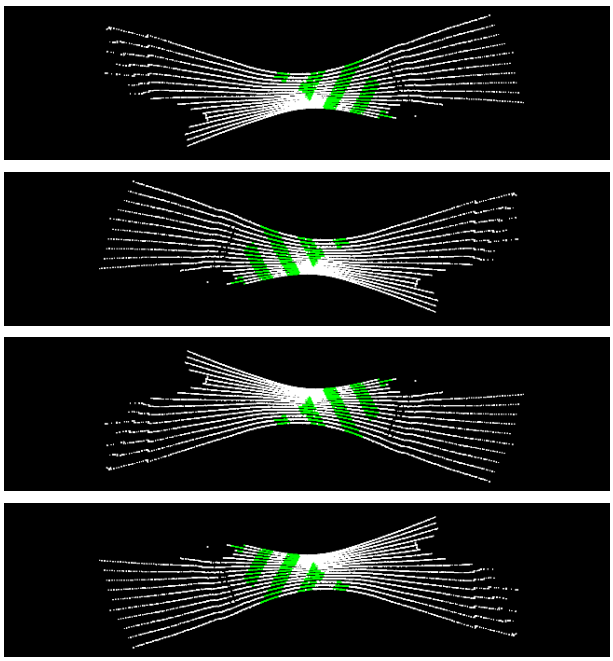

Figure 4. Image Augmentation by Flipping

The annotated dataset will be split, 90% of which will be fed to the neural network model and 10% will be used for testing and accuracy assessment. To densify this sub-dataset for training, augmentation by flipping has been done. The image was copied and flipped in horizontal, vertical, and both horizontal and vertical. In total, the original number of images was increased by three times.

2.4 U-Net and Focal Loss

U-Net is a convolutional neural network (CNN) originally developed for biomedical image segmentation (Ronneberger, 2015). It is a popular CNN model that is now widely used in varying fields of research. It has also been used to extract and classify road markings from point cloud-derived images. (Lagahit and Tseng, 2021) (Lagahit and Tseng, 2020) (Wen et al, 2019)

Focal Loss is an extended or improved version of the Cross-Entropy Loss. It aims to solve the imbalance between classes by providing weights. (Lin et al, 2017) In this way, classes that can be easily misclassified, like the background which holds an assortment of features, can be given less focus when training.

		Class		
Model	**Loss Function**	**Black**	**Ground**	**Road Marking**
A	Cross Entropy	None	None	None
B	Focal	10%	10%	80%
C		1%	1%	98%
D		0.1%	0.1%	99.8%
E		0.01%	0.01%	99.98%
F		0.001%	0.001%	99.998%

Table 1. Model Class Weights

Table 1 shows six U-Net models and their corresponding loss function and class weights used in training. There are 3 classes for the images: (1) "Black" which are pixels that have no value, (2) "Road Marking" which are pixels that are road marking features, and (3) "Ground" which are pixels that are non-road marking features. Since the main target for this work are road markings, it has been assigned the bulk of the weight and the remaining is equally distributed to the other classes. We can see that in this work the difference between the range of weights per model is by the power of 10.

2.5 Model Training

The python implemented UNET model was trained in a computer with an 11th Gen Intel i7 processor, 32 GB of RAM, and an NVIDIA GeForce RTX 3060 Laptop GPU. Due to these conditions, batch size was limited to 16, and images for training and validation were downscaled by a quarter of the original size. 90% of the augmented dataset was used for training and 10% for validation. A maximum of 100 epochs were used for all trials.

2.6 Assessment

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}, \quad (1)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}, \quad (2)$$

$$F1_{score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

The remaining 10% of the original dataset was used to test the trained UNET model. The equations above were used to assess the resulting classification. Recall is the proportion of actual positive cases that are correctly predicted as positive, precision is the proportion of predicted positive cases that are correctly predicted as positive, and the F1-Score or Dice Coefficient is the harmonic mean between precision and recall (Powers, 2011).

3. RESULTS AND DISCUSSION

3.1 Resulting Classified Images

Because the original images are too large and contain very sparse features, making it difficult to interpret, the resulting classified images have been dilated for visualization purposes (dilation has not been used in the method itself), as shown in Figure 5. The green, white, black pixels correspond to road marking, ground, and black (none) classification, respectively.

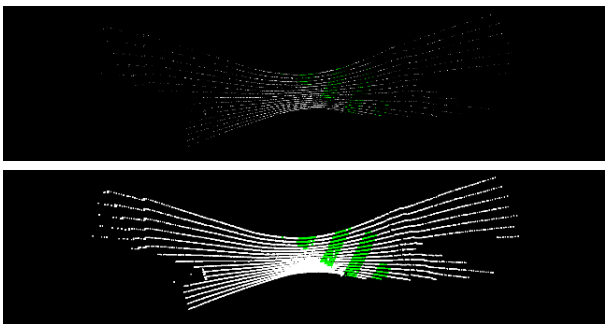

Figure 5. (Top) Original and (Bottom) Dilated Images

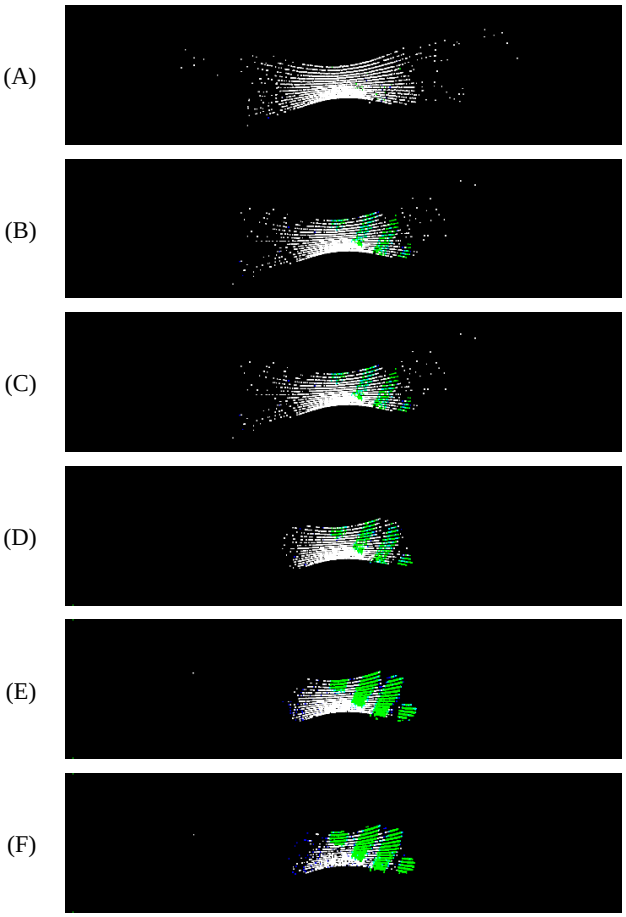

Figure 6. Sample Inference Results

Figure 6 shows sample resulting classified images based on the trained U-Net models. In hindsight, we can see that as the class weight for road marking increases so does it become clearer and better represented on the image. However, considering that the same weight has been given to the ground and black classes, it is interesting to see that as the weights for these classes decrease the number of correctly classified ground pixels decreases as well, but in contrast, the number of black pixels increases. The ratio between the number of pixels may be a factor when given weights to multiple classes are the same. In addition, based on the reference annotated image shown in Figure 5, it has shown that the detected features (non-black) are "focused" or concentrated around the target class (road markings), and features far from it are disregarded.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-5-2022-33-2022 | © Author(s) 2022. CC BY 4.0 License.

3.2 Classification Results

Figure 7. Class Legends

Figure 7 shows the legend, the colors representing the feature classes, that will be used in the graphical representations of the assessment criteria in the succeeding figures.

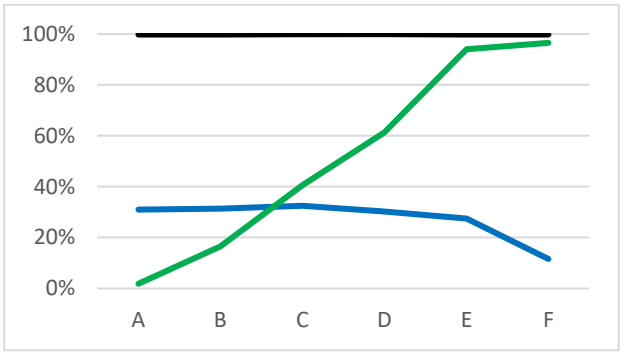

Figure 8. Recall

	Black	Ground	Rd. Marking
A	99.86%	30.96%	1.86%
B	99.81%	31.42%	16.58%
C	99.94%	32.50%	40.55%
D	99.96%	30.20%	61.32%
E	99.85%	27.42%	94.01%
F	99.85%	11.59%	96.54%

Table 2. Recall Values

For recall, which considers the reference or actual image, there is a continuous improvement for the target road marking class, as the given weights increase, for the U-Net models trained with focal loss. It was able to attain a maximum increase of 94.68% as compared to the U-Net model trained with cross-entropy loss. Unfortunately, it is not the case for the ground class which continuously deteriorated and decreased by a maximum of 19.37%. The results for the black class relatively remain unchanged with a maximum increase of 0.1% and a maximum decrease of 0.05%.

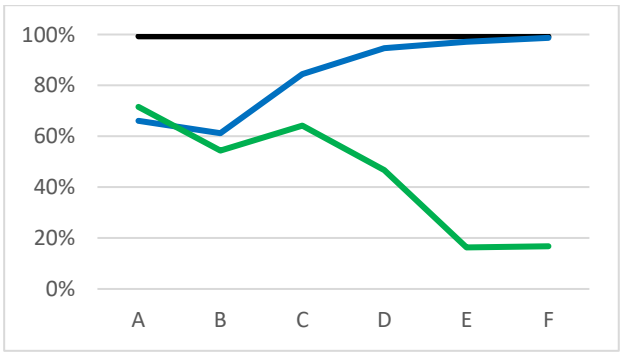

Figure 9. Precision

	Black	Ground	Rd. Marking
A	99.33%	66.12%	71.56%
B	99.33%	61.22%	54.37%
C	99.35%	84.43%	64.20%
D	99.34%	94.70%	46.70%
E	99.34%	97.21%	16.32%
F	99.21%	98.69%	16.79%

Table 3. Precision Values

For precision, which considers the resulting predicted image, it has shown an inverse of the results of recall for the road marking and ground classes. The target road marking class deteriorated by a maximum of 55.24% and the ground class improved by a maximum of 32.57%. To investigate this drastic decrease in precision of the target road marking class we can take a look at its misclassifications.

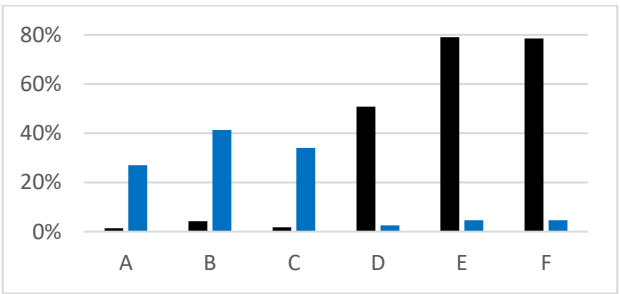

Figure 10. Road Marking Misclassifications (Precision)

	Black	Ground
A	1.42%	27.02%
B	4.28%	41.35%
C	1.74%	34.06%
D	50.78%	2.52%
E	79.05%	4.63%
F	78.52%	4.69%

Table 4. Road Marking Misclassification Values (Precision)

From Figure 10, it can be observed that, from the U-Net model trained with a focal loss and a weight of 99.8%, as the weights increase so do the “black” misclassified pixels. In addition, it can also be seen that “black” class misclassifications take up more than half of the predicted pixels. This causes the drastic decrease in precision values shown in the previous figure. Given that the “black” class has no corresponding point cloud value and that the end product is a point cloud it can be removed in computing precision values.

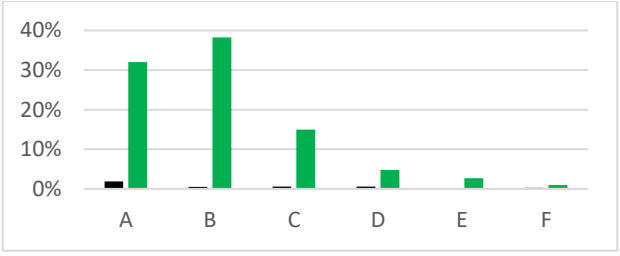

Figure 11. Road Marking Misclassifications (Precision)

	Black	Rd.Marking
A	1.89%	31.99%
B	0.54%	38.24%
C	0.57%	15.00%
D	0.55%	4.75%
E	0.08%	2.71%
F	0.29%	1.02%

Table 5. Road Marking Misclassification Values (Precision)

Moreso, we can also observe the misclassifications on the ground class. It can be seen from Figure 11 that most of the misclassifications are from the target road marking class. However, we can see a continuous decrease in the misclassifications as the weights used in the focal loss for the target class increase. The misclassified road markings decreased to a minimum value of 1.02%.

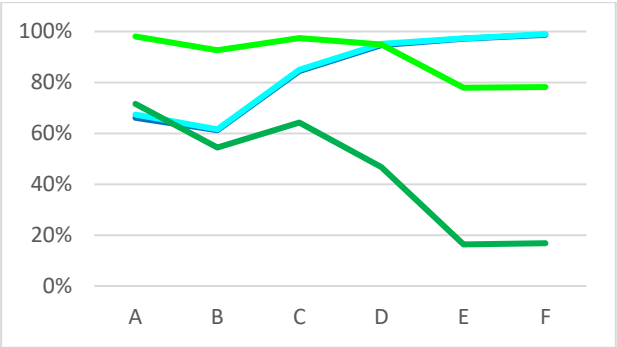

Figure 12. Precision and Adjusted Precision

	Ground	Adjusted Ground	Rd.Marking	Adjusted Rd.Marking
A	66.12%	67.39%	71.56%	98.05%
B	61.22%	61.55%	54.37%	92.70%
C	84.43%	84.91%	64.20%	97.36%
D	94.70%	95.22%	46.70%	94.88%
E	97.21%	97.29%	16.32%	77.90%
F	98.69%	98.98%	16.79%	78.17%

Table 6. Precision and Adjusted Precision Values

As was mentioned previously, since the "black" class did not correspond to a point cloud value it can be removed in the computation of precision values. In Figure 12, the adjusted precision values have been shown alongside the original precision values. Since the "black" misclassifications are minimal in the ground class there was not much difference in the adjusted precision values. However, it has shown tremendous improvement in the precision values of the target road marking class. It was able to achieve an increase of a minimum and maximum of 26.49% and 61.58%, respectively. Although the precision values still decrease as the given weights to the focal loss increase, the range of its deterioration minimizes.

Finally, the F1-Score, which represents the harmonic mean of precision and recall, shows that using focal loss increases the overall results of road marking classification, as shown in Figure 13. Initially, the road marking class gained a minimum and maximum improvement of 22% and 49%, respectively. But, using the adjusted precision, caused a continuous increase in F1-Score snd was able to achieve a maximum increase of 82.74%. For the ground class, initially and after using the adjusted precision, it has improved by a maximum of around 5% but it has also shown a maximum deterioration of around 21%.

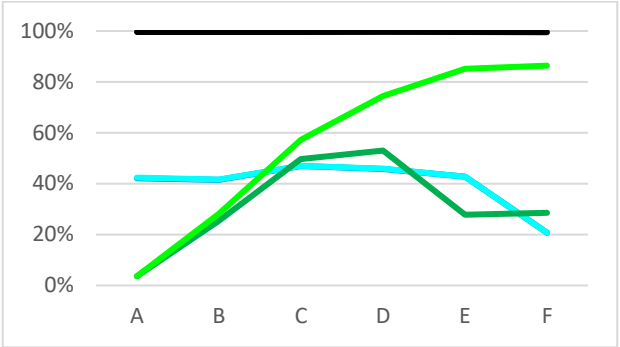

Figure 13. F1-Score

	Black	Ground	Adjusted Ground	Rd. Marking	Adjusted Rd.Marking
A	99.59%	42.18%	42.43%	3.63%	3.65%
B	99.57%	41.53%	41.60%	25.41%	28.13%
C	99.64%	46.94%	47.01%	49.71%	57.25%
D	99.65%	45.79%	45.86%	53.02%	74.49%
E	99.59%	42.78%	42.78%	27.81%	85.20%
F	99.53%	20.75%	20.75%	28.60%	86.39%

Table 7. F1-Score

In direct comparison with previous works that used U-Net in dense point-cloud derived images; (Wen et al, 2019) achieved F1-Scores of 74.42% and 56.42% on the TUM-MLS and their highway dataset, respectively, and (Lagahit and Tseng, 2021) achieved an F1-Score of 86.66% on their data set. This shows that the proposed improvement in the methodology was able to attain F1-Scores from sparse point cloud-derived imagery that was comparable or even better than those that made use of dense point cloud-derived imagery under the same U-Net model.

3.3 Post-Processing: Project to 3D.

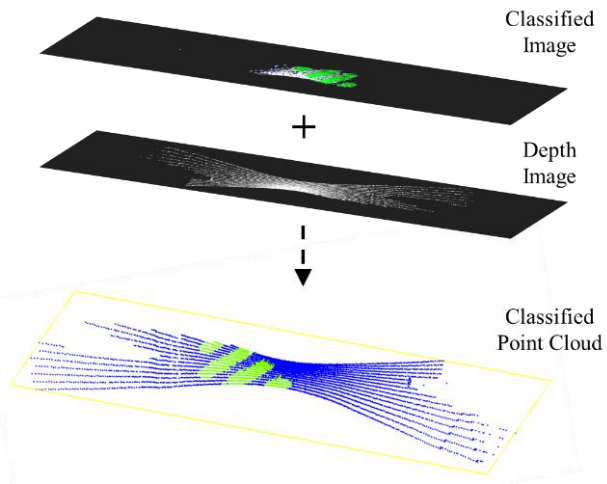

Figure 14. Point Cloud Generation

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-5-2022-33-2022 | © Author(s) 2022. CC BY 4.0 License.

As an extension for post-processing. The road marking pixels in the resulting classified image can be coupled with a generated depth map, created by using Z or elevation values instead of intensity when projecting the point cloud to an image, to generate a classified point cloud. This is done by projecting the classifications unto the depth map and using the elevation values and the position of the pixel in the image as Z, X, and Y values, respectively. If positioning sensors (GNSS, IMU, etc.) and sensor calibration parameters are available, direct georeferencing should be possible.

Furthermore, using this method of point cloud generation supports the method of adjusting the precision values by removing the "black" misclassification in the computation, as was done in the preceding steps of the process. If we project the black classifications onto the depth map, which has the information of which pixels have corresponding values of the point cloud, it will simply be discarded and thus irrelevant when computing precision values.

4. CONCLUSION

Overall, it can be seen that the proposed method of using focal loss with assigned weights in training U-Net to improve the classification accuracy for road markings from sparse mobile LIDAR point cloud-derived images has been successful. A huge improvement in recall from 2.86% to 96.54%, and in F1-score from 3.65% to 86.39%. Its F1-score results on sparse point cloud-derived imagery have also proved to be comparable and even better than recent works on using U-Net on dense point cloud-derived imagery. In addition, given the extension of classified point cloud generation and with more enhancements, this method can be used to realize automatic classification of road markings for mobile mapping units with low-level LIDAR sensors.

ACKNOWLEDGEMENTS

The research has been supported by the WISE-SSS program of Tokyo Institute of Technology. A special thanks to Zongdian Li of Sakageuchi-Tran Laboratory, Department of Electrical and Electronics Engineering, Tokyo Institute of Technology for providing his assistance and expertise during the data gathering.

REFERENCES

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.

Lagahit, M.L.R., Tseng, Y.H., 2021. Road Marking Extraction and Classification from Mobile LIDAR Point Clouds Derived Imagery Using Transfer Learning. CSPRS Journal of Photogrammetry and Remote Sensing, vol.26 pp. 127-141. doi.org/10.6574/JSPRS.202109_26(3).0001.

Lin, T.Y., Goyal, P., Girshick, R.B., He, K., Doll, P., 2017. Focal Loss for Dense Object Detection. Computing Research Repository. Retrieved January 4, 2022, from the arXiv database.

Liu, R., Wang, J., Zhang, B., 2020. High Definition Map for Automated Driving: Overview and Analysis. Journal of Navigation, 73(2), 324-341. doi:10.1017/S0373463319000638.

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.

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.

APPENDIX

Additional sample resulting classified images based on the trained models have been included here.

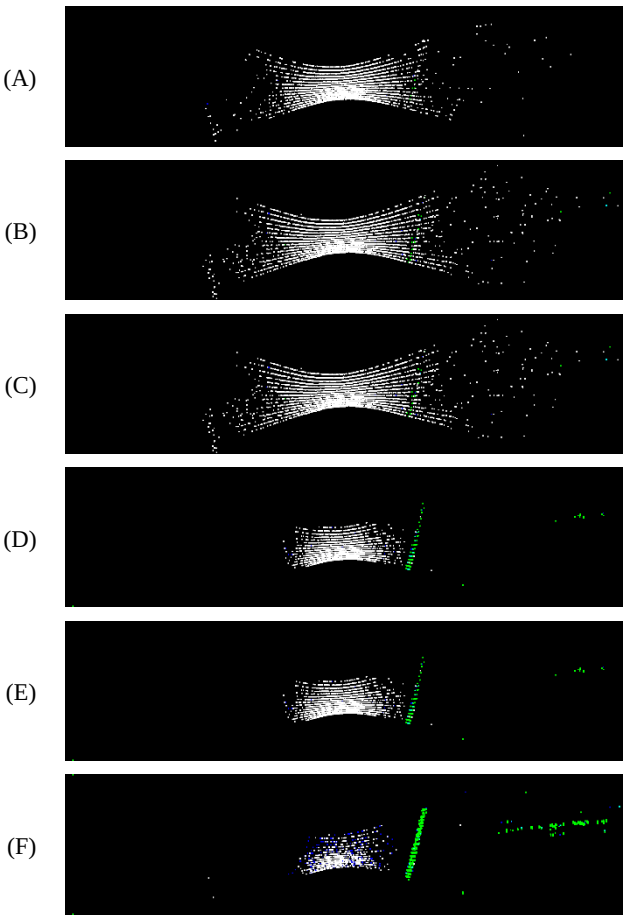

Figure 15. Sample Inference Results 2