remote sensing

MDPI

Article

Focal Combo Loss for Improved Road Marking Extraction of Sparse Mobile LiDAR Scanning Point Cloud-Derived Images Using Convolutional Neural Networks

Miguel Luis R. Lagahit [1,2,*] and Masashi Matsuoka [1,2]

[1] Department of Architecture and Building Engineering, Tokyo Institute of Technology, Tokyo 152-8550, Japan
[2] Tokyo Tech Academy for Super Smart Society, Tokyo Institute of Technology, Tokyo 152-8550, Japan
* Correspondence: lagahit.m.aa@m.titech.ac.jp

Abstract: Road markings are reflective features on roads that provide important information for safe and smooth driving. With the rise of autonomous vehicles (AV), it is necessary to represent them digitally, such as in high-definition (HD) maps generated by mobile mapping systems (MMSs). Unfortunately, MMSs are expensive, paving the way for the use of low-cost alternatives such as low-cost light detection and ranging (LiDAR) sensors. However, low-cost LiDAR sensors produce sparser point clouds than their survey-grade counterparts. This significantly reduces the capabilities of existing deep learning techniques in automatically extracting road markings, such as using convolutional neural networks (CNNs) to classify point cloud-derived imagery. A solution would be to provide a more suitable loss function to guide the CNN model during training to improve predictions. In this work, we propose a modified loss function—focal combo loss—that enhances the capability of a CNN to extract road markings from sparse point cloud-derived images in terms of accuracy, reliability, and versatility. Our results show that focal combo loss outperforms existing loss functions and CNN methods in road marking extractions in all three aspects, achieving the highest mean F1-score and the lowest uncertainty for the two distinct CNN models tested.

Keywords: low-cost mobile mapping; modified loss function; image segmentation

Citation: Lagahit, M.L.R.; Matsuoka, M. Focal Combo Loss for Improved Road Marking Extraction of Sparse Mobile LiDAR Scanning Point Cloud-Derived Images Using Convolutional Neural Networks. *Remote Sens.* **2023**, *15*, 597. https://doi.org/10.3390/rs15030597

Academic Editors: Wataru Takeuchi, Hirokazu Yamamoto, Sayaka Yoshikawa and Naoyuki Hashimoto

Received: 2 December 2022
Revised: 12 January 2023
Accepted: 17 January 2023
Published: 19 January 2023

1. Introduction

Road markings, such as lane lines and crossing marks, are features made of highly retro-reflective materials that are painted on the road. In complex urban road networks, these markers provide the necessary information for reliable routing and collision prevention [1]. As such, they must be correctly depicted in their virtual counterparts, such as those found in high-definition (HD) maps. HD maps are centimeter-level three-dimensional (3D) maps that support autonomous vehicles (AVs); self-driving cars that better localize themselves in their environments forecast occurrences outside the reach of their sensors and enhance path planning in complicated traffic scenarios [2–4].

Due to the surface characteristics of road markings, they return relatively high-intensity values during light detection and ranging (LiDAR) scanning. This unique feature enables effective road marking extraction using point clouds [5]. This made LiDAR a viable alternative to cameras that have shown poor performance in low light [6].

A common method for automatically extracting road markings from point clouds is to project them to a 2D plane in a top-down or bird's-eye-view (BEV) manner, using intensity as pixel values, and then making use of 2D image-based techniques, such as thresholding [1,6]. Additionally, by converting 3D point clouds to 2D images, the computational complexity of the automated procedure significantly decreases [1]. To reduce the effects of varying intensity, Cheng et al. used dynamic intensity thresholding, a method that combines scan–angle–rank-based intensity correction and large-size high-pass filtering. However, it still proved insufficient for inconsistencies in intensity brought by uneven road

surfaces (e.g., cracks and potholes) [7]. To improve road marking visibility, Pan et al. combined conditional Euclidean clustering, Otsu thresholding, and statistical outlier remover (SOR) to remove unwanted noise and retain only points with high intensity and density. Unfortunately, due to subsequent filtering, the proposed method struggles with short and broken geometries, such as dashed center lines [8]. To compensate for small changes in intensity caused by large differences in elevation, Soilan et al. used adaptive thresholding. It relies on the laser beam angle to compute several masks grouped by angle similarity. However, when compared to preceding thresholding methods, it showed the equivalent performance in terms of extracting road markings [9].

All of these prove that traditional image processing is inadequate in instances where road marking intensity is inconsistent, the road surface varies, and the contrast to the surrounding surface is minimal [10]. This led to the exploration of deep learning techniques, such as semantic segmentation via convolutional neural networks (CNNs), on point cloud-derived images to extract road markings. Wen et al. demonstrated that U-Net, a full CNN, can extract road markings in scenarios where there are varying intensities, point densities, and poor distinctions with the surrounding surface [10]. Lagahit et al. have also shown that U-Net successfully extracts road markings at varying point densities using different sample sizes when projecting the point cloud to an image [11]. Lagahit et al. made use of transfer learning (i.e., a deep learning method that retrains a previously trained model on a more task-specific dataset) on multiple scales and proved that it outperforms traditional image processing as well as a single-trained U-Net in some cases [12]. Lastly, Ma et al. took it one step further by modifying the structure of U-Net and introducing capsule networks for improved extractions [13].

However, in the aforementioned studies, the point clouds used to extract road markings are highly dense, reaching around 0.5 to 1.0 points per square centimeter [10–13]. These are obtained through the use of mobile mapping systems (MMSs). MMSs employ survey-grade cameras and LiDAR sensors integrated with global navigation satellite system(s) (GNSS) and inertial measurement unit(s) (IMU) or through simultaneous localization and mapping (SLAM) for positioning and georeferencing [14]. Unfortunately, MMSs can be quite expensive [15], making it costly to frequently deploy for road marking extraction in tasks such as roadway monitoring and HD map updates.

Together with the increasing interest in the development of AVs, this has increased the investigations toward the use of lower-cost MMSs to acquire reliable spatial data [15]. However, low-cost LiDAR sensors produce sparser point clouds at single and aggregated sweeps, resulting in projected BEV images with numerous no-value pixels—no corresponding point in the point cloud—and hardly any target road marking pixels. This extreme data imbalance is demonstrated to have a negative effect on the performances of existing deep-learning methods. Lagahit et al. and Lagahit et al. trained two different CNN models, U-Net and Fast-SCNN, both trained with cross-entropy loss as a benchmark. Both displayed little to no detection of pixels in the road marking class [16,17].

As a solution, Lagahit et al. and Lagahit et al. explored the use of a more suitable loss function to boost the CNN model's extraction performance [16,17]. Loss functions help guide model training by measuring the difference between predicted and reference images and adjusting the weights of a neural network [18,19]. Their findings show that shifting to a weighted focal loss function allowed the model to significantly improve predictions on a sparse target feature class [16,17]. The switch in loss function has also been proven to strengthen the model performance for dense point cloud-derived images. Wen et al. and Ma et al. obtained better results by using the semantic segmentation metric, intersection-over-union, as their loss function [10,13]. Despite these recent advances, no work has been done to develop a loss function suitable for extracting features from sparse point cloud-derived images, such as road markings.

In this work, we propose a modified loss function, focal combo loss, which aims to further improve road marking extraction on sparse point cloud-derived images in terms of (1) accuracy—obtaining the highest F1-score; (2) reliability—attaining the highest

minimum resulting accuracy after factoring in uncertainty from multiple trials; and (3) versatility—consistently achieving the best performance on different CNN models with completely distinct structures. To accomplish this goal, (1) a comprehensive comparison of existing related loss functions was conducted; (2) a thorough analysis of the loss function parameters was performed; and (3) an extensive comparison with existing CNN methods was conducted.

2. Materials and Methods

Our proposed loss function, focal combo loss, will be applied to the CNN used for road marking extraction, as shown in Figure 1. As an overview of the methodology, the sparse point cloud collected from low-cost mobile LiDAR scanning will be converted to intensity images to serve as training and testing inputs for CNN-based road marking extraction. Furthermore, as an extension of the concept, a classified point cloud can be generated by coupling these classified images with a depth image from the same point cloud.

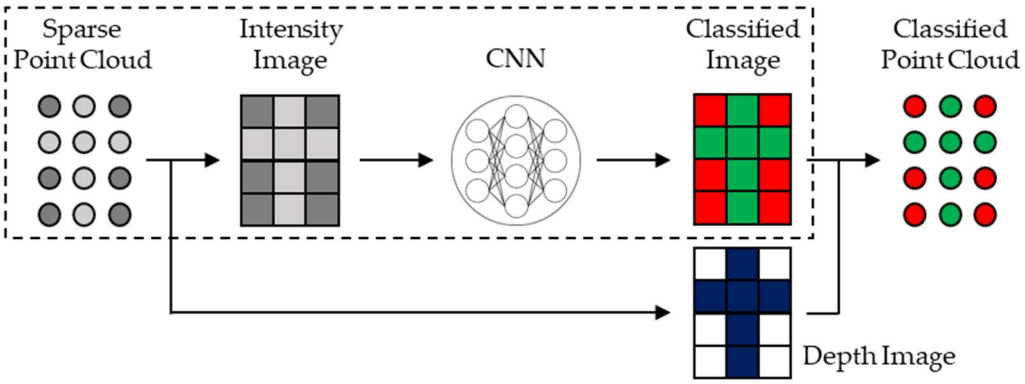


Figure 1. Road marking extraction procedure.

2.1. Dataset Preparation

The sparse point cloud used in this experiment was captured using a Velodyne Puck (VLP-16) LiDAR sensor mounted in front of a small vehicle tilted 45 degrees downward from horizontal. As shown in Figure 2, it was temporarily attached to the vehicle's large windscreen, since there was no available space in front for a permanent installation. This enabled the scanning to obtain more ground surface points.

Figure 2. The small vehicle used for LiDAR scanning, with the LiDAR sensor encircled in red.

As shown in Figure 3, the scanning was carried out on roadways within the East Area of the Ookayama Campus, Tokyo Institute of Technology. The roads contained an assortment of road marking features but mainly consisted of lane lines and crosswalks. As a result, the target road markings were restricted to only these types.

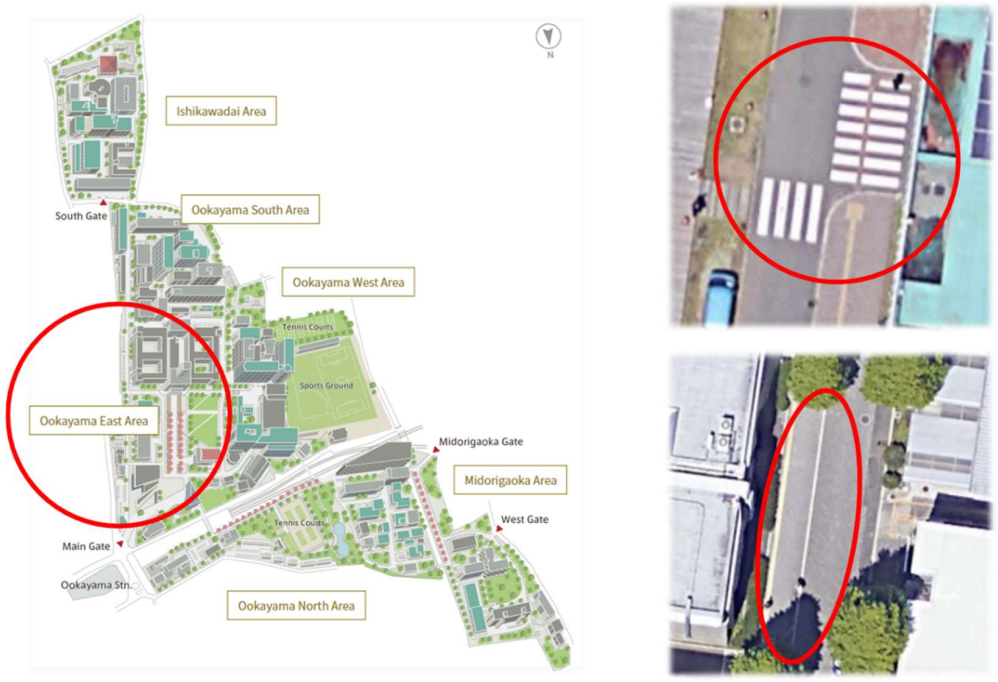


Figure 3. Encircled in red are the scanning location (**left**) taken from the Tokyo Tech website and road marking types visible in the area (**right**) taken from Google Earth.

The point clouds were then filtered to retain only points on the ground and in a portion in front of the vehicle, removing any unwanted points, such as vegetation and overhead structures, as illustrated in Figure 4. After which it was projected top-down into a 2D plane with a 1 cm square grid and average intensity as pixel values, yielding 2048 by 512-pixel BEV intensity images.

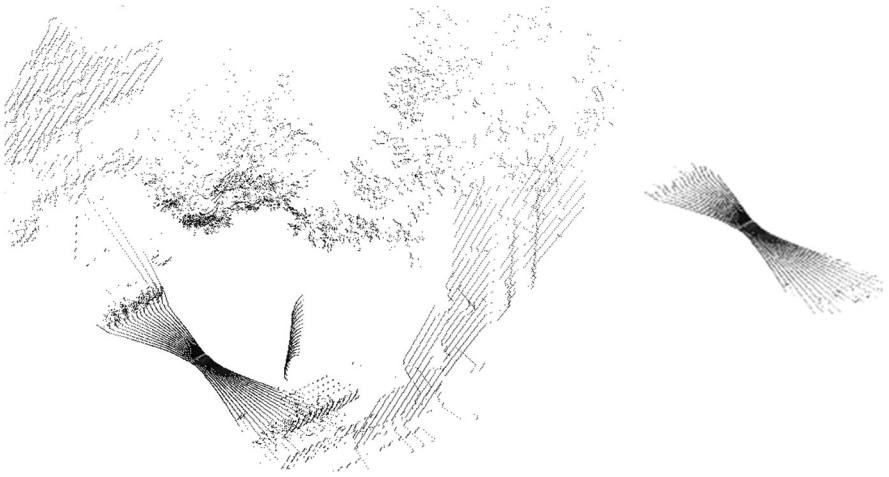


Figure 4. Sample sparse point cloud before (**left**) and after (**right**) filtering.

The intensity images were manually labeled into three classes to be used as training data for the CNN: 'black', which has no point cloud value, 'road marking', which is the target road marking, and 'others' which are all remaining pixels. A sample labeled image can be seen in Figure 4. These, along with the BEV intensity images, make up the training, validation, and testing datasets. Around 90% of all the processed images were used as training and validation data, which were further densified using the data augmentation method of multi-orientation flipping [20]. The remaining were used as testing data to assess the model performance in road marking extractions.

In this paper, the resulting predicted images will be dilated using a 3×3 kernel solely for visualization purposes. As shown in Figure 5, this is done to increase the interpretability of the results since the target features on the images are too small. Because of this, there may also be minor changes, but nothing significant enough to alter the outcome.

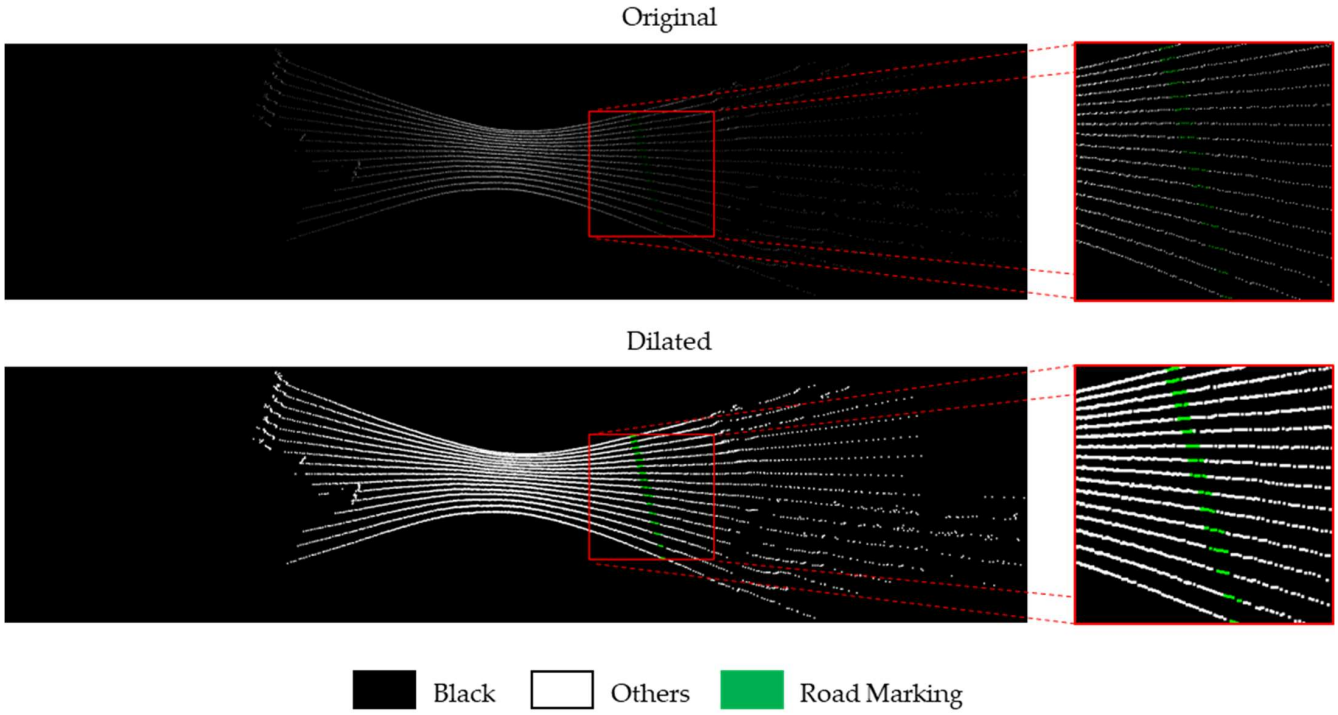


Figure 5. Sample image of the dataset (**top**) and dilated counterpart for visualization (**bottom**). The red square indicates the location of the detected road marking.

2.2. Convolutional Neural Network Models

To test for versatility, our proposed loss function will be applied to two differently structured convolutional neural networks, the popular and robust U-Net and a more lightweight fast-SCNN.

U-Net is a CNN that was originally designed for biomedical image segmentation. Its structure, as seen in Figure 6, consists of a symmetrically connected succession of convolution layers with pooling operators in the first half and upscaling operators in the latter half for better and more precise predictions [21]. It is a popular CNN model that is now widely used in varying scientific fields, including the extraction of road markings from point cloud-derived images. It has also surpassed other CNN models in terms of accuracy by achieving the highest F1 score, reaching up to 90% in extracting road markings [10–13,16].

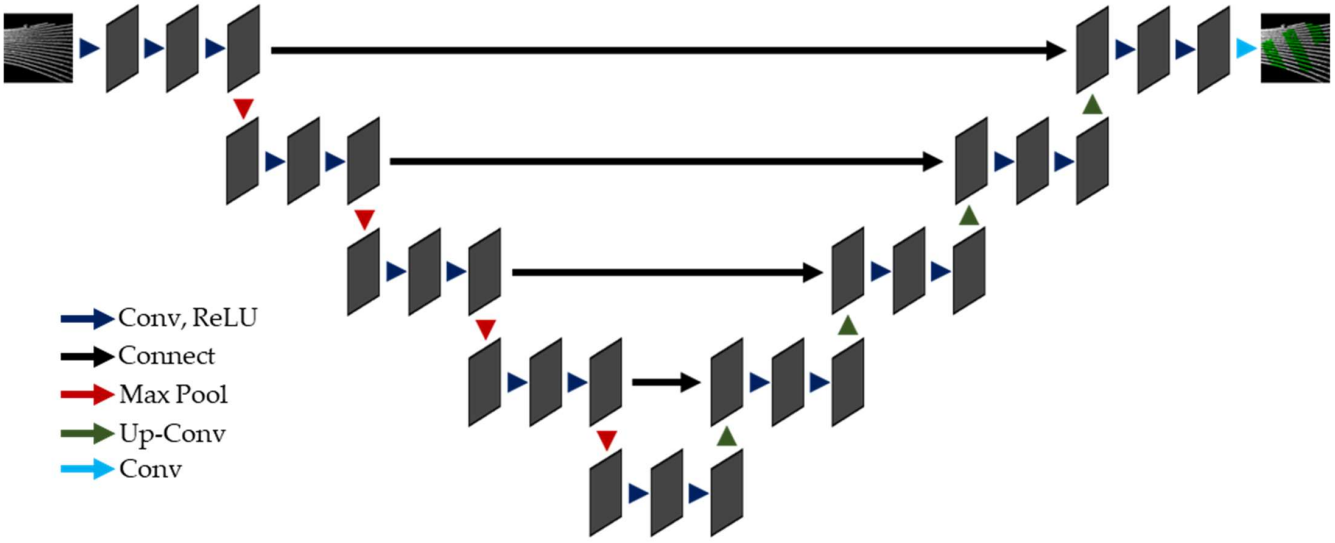


Figure 6. U-Net structure.

FSCNN is a lighter CNN model that achieves real-time segmentation by utilizing popular techniques, such as pyramid pooling, inverted residual bottlenecks, and feature fusion, as seen in Figure 7 [22]. Recently, it has been explored for road marking extraction due to its suitability for low-cost applications, but its results were still inferior to U-Net's under the same conditions. Nonetheless, it has been shown to achieve an F1-score of more than 70% at prediction speeds of 0.2 s [16,17].

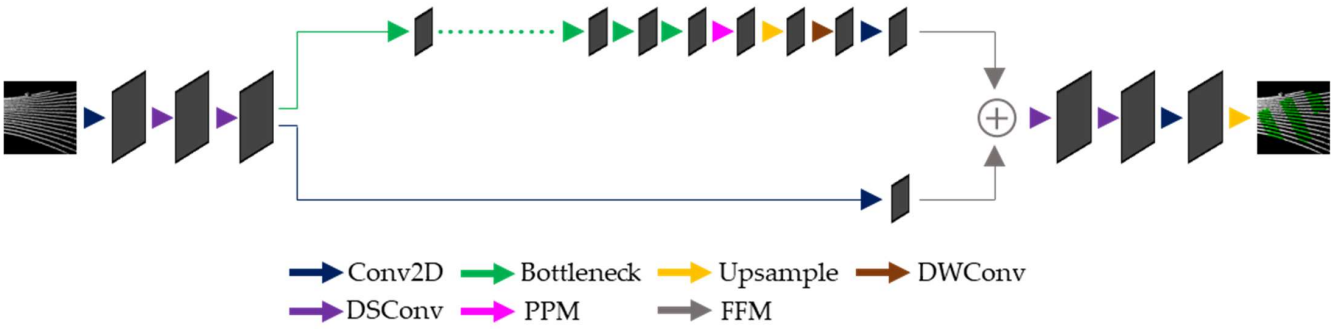


Figure 7. Fast-SCNN structure.

2.3. Loss Functions

In deep learning, the loss function:

$$J(W) = k - \hat{k}, \tag{1}$$

is used to compute the difference between the predicted results ($\hat{k}$) and the actual labels (k) after going through an activation function. The resulting loss value is then minimized by backpropagating through the network layers to modify the network's weights and biases through an optimization function:

$$W^{(k+1)} = W^{(k)} - \frac{\partial}{\partial W^{(k)}} J(W). \tag{2}$$

This is done by adding the gradient of the loss function in the opposite direction to the previous weights [18,19]. By applying the appropriate loss function when training, the CNN could be guided to make better predictions.

To begin with, a commonly used loss function is the cross-entropy loss,

$$Cross\ Entropy\ Loss = -\log(P_t) \text{ where } P_t = \begin{cases} p & if\ \hat{k} = 1 \\ 1-p & otherwise \end{cases}. \tag{3}$$

It calculates the difference between two probability distributions for a set of events [23]. However, cross-entropy loss fails for imbalanced datasets, where certain classes greatly outnumber the other classes.

To address the issue of imbalanced datasets, predetermined class weights (w) can be introduced to cross-entropy loss:

$$Weighted\ Cross - Entropy\ Loss = -wlog(P_t). \tag{4}$$

where the weights are computed by taking the inverse ratio of the number of pixels in that class ($\#samples_i$) multiplied by the total number of classes ($\#classes$), with the total number of pixels in the image ($\#samples$) [24].

$$Class\ Weight = \frac{\#samples}{\#classes \cdot \#samples_i} \tag{5}$$

Further improving the performance on imbalanced datasets, focal loss,

$$Weighed\ Focal\ Loss = -w(1-P_t)^{\gamma} log(P_t), \tag{6}$$

introduced a modulating term $(1-P_t)^{\gamma}$ to penalize the contribution of easier classes and enabled the CNN model to concentrate more on learning harder classes [25]. The focal loss combined with class weights significantly boosts the capability of a CNN model in detecting sparse features in datasets with extreme class imbalance.

Another loss function that handles class imbalance well is dice loss,

$$Dice\ Loss = 1 - \frac{2TP + c}{2TP + FP + FN + c}, \tag{7}$$

Dice loss originates from the dice coefficient, or F1-score, evaluation metric modified with the addition of a smoothing constant to prevent gradient explosion. It makes use of the metric's property as a harmonic mean to obtain a more accurate measure of the difference in correct overlaps even on harder features [26].

Leveraging this property, combo loss,

$$Combo\ Loss = \alpha\ (Modified(-wlog(P_t))) + (1-\alpha)\left(1 - \frac{2TP}{2TP + FP + FN}\right), \tag{8}$$

takes the weighted sums of the dice loss and a modified cross-entropy loss as another approach to further improve the classification of poorly represented features [27]. However, it was demonstrated that using the proposed modified cross-entropy loss yielded the same F1-score as using the cross-entropy loss. Hence, in this paper, the non-modified version of cross-entropy will be used instead.

Returning to dice loss, a modulating term, in the form of an exponent factor $\left(\frac{1}{\beta}\right)$, is introduced in focal dice loss,

$$Focal\ Dice\ Loss = 1 - \left(\frac{2TP}{2TP + FP + FN}\right)^{\frac{1}{\beta}}. \tag{9}$$

Similar to focal loss, it aimed to direct dice loss into giving more focus to poorly segmented classes [28].

Our proposed loss function is a modified version of the combo loss,

$$\textit{Focal Combo Loss} = \alpha \left(-w(1-P_t)^{\gamma} log(P_t)\right) + (1-\alpha)\left(1 - \left(\frac{2TP}{2TP+FP+FN}\right)^{\frac{1}{\beta}}\right), \quad (10)$$

in which we take the weighted sum of the weighted focal combo loss and the focal dice loss, both developed to better focus on harder-to-classify features. This will be highly advantageous, given the scarcity of features in sparse point cloud-derived images.

2.4. Model Training

To speed up the training of multiple models, training was done on two computers. Tsubame 3.0, a supercomputer at the Tokyo Institute of Technology with four 16 GB GPU and 256 GB of RAM, was used to train U-Net models. A desktop computer with an 11 GB GPU and 48 GB of RAM was used to train fast-SCNN models.

Parameter limitations were set based on the computer with lower processing capacity to ensure a fair comparison. Because of these constraints, the batch size was limited to 16, and images for training and validation were downscaled to a quarter within the network. In addition, an Adam optimizer and a learning rate of 0.0001 was used.

To account for uncertainty, models were trained in three batches, each with fixed seeds, when training multiple models with changing loss functions. A total of 100 epochs were performed for each trial, using the model version at an epoch with the lowest loss value for road marking extraction.

3. Results

3.1. Generated Dataset

Table 1 lists the breakdown of pixels per class of the generated datasets used in the experiment. The overwhelming number of 'black' pixels makes it clear that there is an extreme class imbalance. Additionally, we can see that the target class 'road marking' is still considerably smaller than the 'others' class, which is already scant. This can have a significant impact on a CNN's capability to correctly detect such sparse features.

Table 1. Dataset statistics.

Dataset	Number of Images	Number of Pixels per Class		
		Black	Others	Road Marking
Training	1000	99.03%	0.88%	0.09%
Validation	100	99.04%	0.76%	0.20%
Testing	100	99.03%	0.94%	0.03%

3.2. Resulting Images of Extracted Road Markings

Figures 8 and 9 show the sample cropped predictions of road marking extractions using our proposed loss function, focal combo loss, in comparison to other related existing loss functions using the U-Net and fast-SCNN models, respectively. Projected versions of the classified images are shown alongside them, with only those with corresponding point cloud values retained.

Cross-entropy loss fails to classify road marking pixels for U-Net and does not classify anything at all for fast-SCNN. The addition of weights and a modulating term, in case of focal loss, proved to enhance model performance reflected by a partially visible depiction of the target road marking. For dice loss and focal dice loss, although they fared poorly, the results in the two models were completely different. In U-Net, dice and focal dice loss barely caught any road markings and showed a focused misclassification in a different region. In fast-SCNN, dice and focal dice loss were able to correctly classify all road markings but overreached and included a large portion of the surrounding pixels. Nonetheless, only combo loss and our proposed focal combo loss produced good representations of

road markings. However, it is important to notice that only our proposed focal combo loss prediction achieved a cleaner result when only pixels with point cloud values were retained.

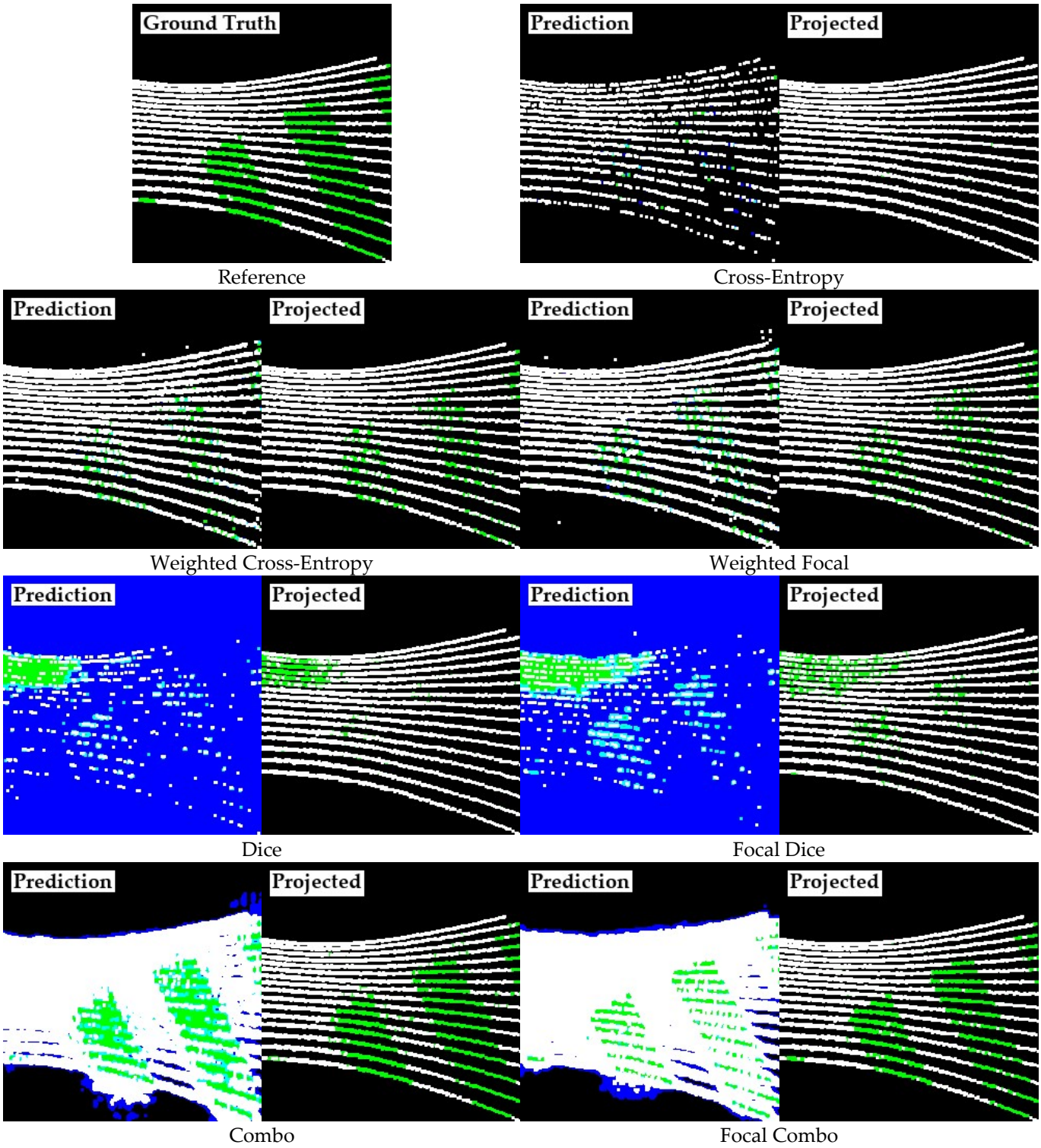


Figure 8. Sample prediction results and projected counterparts using U-Net.

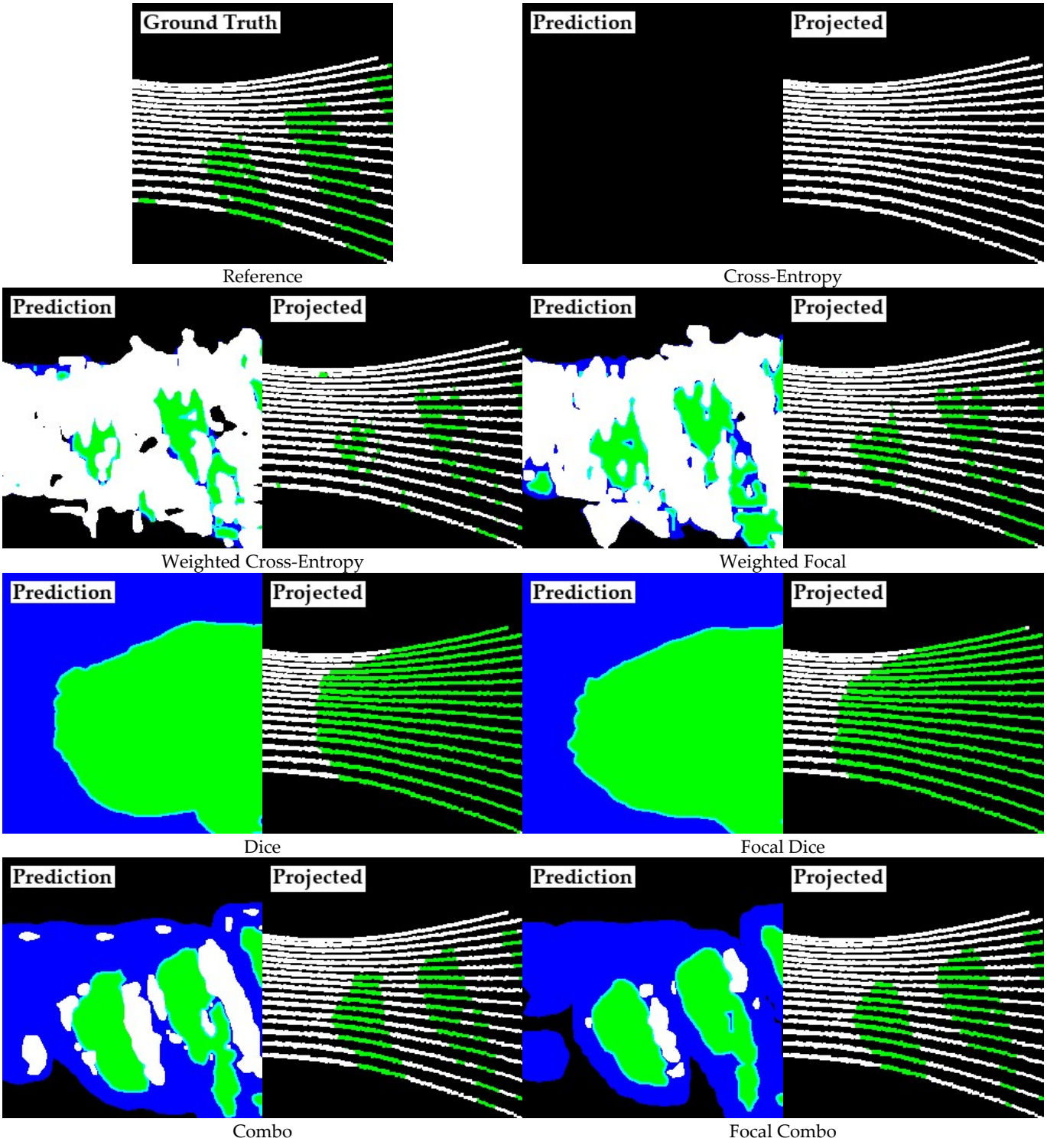


Figure 9. Sample prediction results and projected counterparts using fast-SCNN.

3.3. Assessment Criteria

Road marking extraction was evaluated through the following criteria obtained from the calculated confusion matrix between the predicted and actual image: recall, precision, F1-score, and intersection-over-union (IoU), as shown in Equations (11)–(14). Recall means that pixels in the reference image correspond correctly to those in the predicted image. Precision, on the other hand, means that pixels in the predicted image correspond correctly to those in the reference image. High values in both of those criteria indicate that the

extraction was successful. Taking both into account, the F1-score, also known as the dice coefficient, is the harmonic mean of precision and recall. The harmonic mean tends to favor smaller values; it is a reliable criterion for uneven precision and recall. As a final measure, the IoU, also known as the Jaccard index, is also included. It divides the correctly overlapping pixels by the union of the reference and the prediction, making the resulting value nearer the minimum of precision and recall [29]. Tables 2 and 3 shows the initial assessment results based on these criteria.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}, \tag{11}$$

$$Precision = \frac{True\ Positive}{True\ Positive\ +\ False\ Positive}, \tag{12}$$

$$F1_{score} = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{13}$$

$$IoU = \frac{True\ Positive}{True\ Positive + False\ Positive + False\ Negative} \tag{14}$$

Table 2. Initial results for U-Net in (%) for the road marking class.

	Loss	Recall	Precision	F1-Score	IoU
[23]	Cross Entropy	7.3 ± 4.5	**67.2 ± 6.1**	12.8 ± 7.6	6.9 ± 4.2
[24]	Weighted Cross Entropy	46.9 ± 5.4	42.7 ± 5.2	44.3 ± 0.6	**28.4 ± 0.5**
[25]	Weighted Focal (γ = 1)	52.5 ± 6.0	37.9 ± 5.7	**44.0 ± 5.9**	28.3 ± 4.9
[26]	Dice	3.0 ± 5.0	1.2 ± 1.4	0.7 ± 0.8	0.3 ± 0.4
[28]	Focal Dice (β = 3)	8.3 ± 14.1	5.4 ± 6.2	1.3 ± 1.8	0.7 ± 0.9
[27]	Combo (α = 0.50)	**96.2 ± 2.7**	10.5 ± 2.2	18.8 ± 3.4	10.4 ± 2.1
	Focal Combo (γ = 1, β = 3, α = 0.25)	95.3 ± 1.3	9.9 ± 0.8	18.0 ± 1.3	9.9 ± 0.8

Table 3. Initial results for fast-SCNN in (%) for the road marking class.

	Loss	Recall	Precision	F1-Score	IoU
[23]	Cross Entropy	0.0 ± 0.0	- - -	- - -	0.0 ± 0.0
[24]	Weighted Cross Entropy	55.4 ± 12.2	**6.4 ± 0.6**	**11.4 ± 1.0**	**6.0 ± 0.6**
[25]	Weighted Focal (γ = 5)	52.7 ± 8.8	5.9 ± 0.4	10.5 ± 0.7	5.6 ± 0.4
[26]	Dice	56.8 ± 46.9	3.3 ± 1.0	4.6 ± 1.2	2.4 ± 0.6
[28]	Focal Dice (β = 1.5)	57.4 ± 48.4	3.4 ± 0.7	4.8 ± 2.5	2.5 ± 1.3
[27]	Combo (α = 0.25)	61.5 ± 10.4	5.2 ± 0.4	9.6 ± 0.8	5.0 ± 0.4
	Focal Combo (γ = 5, β = 1.5, α = 0.25)	**66.6 ± 1.7**	5.5 ± 0.8	10.1 ± 1.3	5.3 ± 0.7

For the succeeding tables, rows with a gray background contain our results and values in bold refer to the highest mean score in the corresponding evaluation criteria.

3.4. Assessment after 'Black' Pixel Omission

As a special property of sparse point cloud-derived images, misclassifications in the 'black' pixels can be omitted in the assessment computations, as these pixels correspond to no value when projected as a point cloud, as shown in Figure 10. Since 'black' pixels occupy most of the images, this has a significant impact on the evaluation.

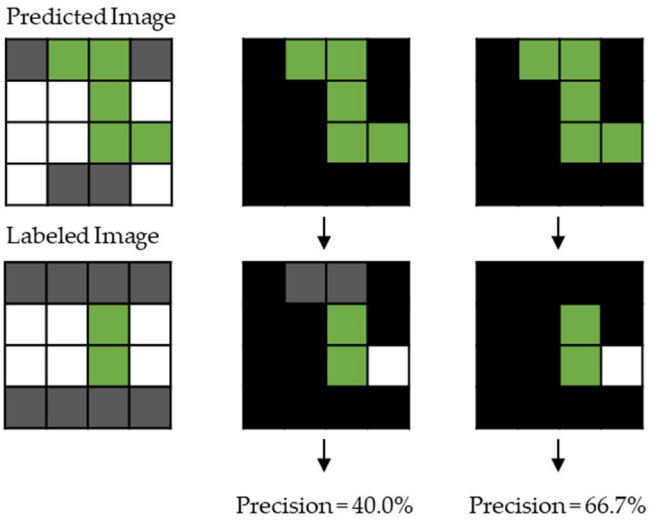


Figure 10. Rethinking assessment computation for sparse point cloud-derived images.

In Tables 4 and 5, after removing misclassifications in the 'black' pixels, we can see a tremendous increase in the evaluation results. Our proposed loss function, focal combo loss, achieved the highest mean F1-score in both models, outperforming all other loss functions. Combo loss came close, but focal combo loss proved to be more reliable by having the smallest uncertainty and the highest minimum value for both the F1-score and IoU.

Table 4. Results after the 'black' pixel omission for U-Net in (%) for the road marking class.

	Loss	Recall	Precision	F1-Score	IoU
[23]	Cross Entropy	7.3 ± 4.5	**97.1 ± 0.2**	13.3 ± 8.0	25.3 ± 15.2
[24]	Weighted Cross Entropy	46.9 ± 5.4	95.6 ± 0.5	62.8 ± 4.7	50.7 ± 4.9
[25]	Weighted Focal (γ = 1)	52.5 ± 6.0	94.3 ± 1.7	67.3 ± 55.2	5.2 ± 7.9
[26]	Dice	3.0 ± 5.0	2.2 ± 1.9	1.8 ± 2.7	0.9 ± 1.4
[28]	Focal Dice (β = 3)	8.3 ± 14.1	14.3 ± 16.7	3.9 ± 6.3	2.1 ± 3.3
[27]	Combo (α = 0.50)	**96.2 ± 2.7**	76.4 ± 6.7	84.9 ± 3.1	73.9 ± 4.6
	Focal Combo (γ = 1, β = 3, α = 0.25)	95.3 ± 1.3	77.4 ± 2.7	**85.4 ± 1.1**	**74.5 ± 1.7**

Table 5. Results after the 'black' pixel omission for fast-SCNN in (%) for the road marking class.

	Loss	Recall	Precision	F1-Score	IoU
[23]	Cross Entropy	0.0 ± 0.0	- - -	- - -	- - -
[24]	Weighted Cross Entropy	55.4 ± 12.2	80.9 ± 10.6	64.6 ± 6.5	48.2 ± 6.9
[25]	Weighted Focal (γ = 5)	52.7 ± 8.8	**85.2 ± 3.3**	64.9 ± 6.7	48.7 ± 7.7
[26]	Dice	56.8 ± 46.9	41.3 ± 5.2	37.7 ± 28.7	26.3 ± 19.3
[28]	Focal Dice (β = 1.5)	57.4 ± 48.4	43.4 ± 10.0	40.0 ± 32.9	28.8 ± 23.7
[27]	Combo (α = 0.25)	61.5 ± 10.4	78.2 ± 4.1	68.6 ± 7.6	**53.7 ± 6.8**
	Focal Combo (γ = 5, β = 1.5, α = 0.25)	**66.6 ± 1.9**	71.4 ± 6.9	**68.9 ± 4.2**	52.9 ± 4.8

3.5. Analyzing the Weighted Sum Combinations

The parameters γ and β, which were used in the focal combo loss, were derived from the γ and β of the best-performing weighted focal loss and focal dice loss. Parameter α, on the contrary, determines which of the loss functions in the weighted sum had the strongest influence on achieving the best road marking extraction results. From Tables 6 and 7, we can derive that focal dice loss had a greater influence on both models.

Table 6. Results of analyzing parameter α for U-Net in (%) for the road marking class.

α	Recall	Precision	F1-Score	IoU
0.25	**95.3 ± 1.3**	77.4 ± 2.7	**85.4 ± 1.1**	**74.5 ± 1.7**
0.50	66.7 ± 19.3	79.0 ± 5.1	70.9 ± 8.1	55.3 ± 10.1
0.75	23.4 ± 10.5	**85.7 ± 7.8**	36.2 ± 13.4	22.7 ± 10.5

Table 7. Results of analyzing parameter α for fast-SCNN in (%) for the road marking class.

α	Recall	Precision	F1-Score	IoU
0.25	**66.6 ± 1.9**	**71.4 ± 6.9**	**68.9 ± 4.2**	**52.9 ± 4.8**
0.50	68.0 ± 3.9	69.0 ± 10.7	68.0 ± 4.4	51.9 ± 5.1
0.75	60.0 ± 21.7	70.5 ± 10.2	61.9 ± 11.1	45.7 ± 11.3

4. Discussion

We now compare the results of our proposed focal combo loss to those of existing methods on road marking extractions from point cloud-derived BEV images using CNNS, after comparing them to the results of related existing functions.

Figure 11 clearly shows that our proposed loss function, focal combo loss, still manages to outperform existing CNN methods that have been used for road marking extractions on point cloud-derived BEV images. Table 8 backs this up by showing that it has the highest F1-score and IoU among the competition. More importantly, we can see that a lighter model, fast-SCNN with a focal combo loss, could best even the other methods that used the U-Net model.

Table 8. Results in comparison to existing methods in (%) for the road marking class.

	Method	Recall	Precision	F1-Score	IoU
[11]	U-Net + Cross Entropy	7.3 ± 4.5	97.1 ± 0.2	13.3 ± 8.0	25.3 ± 15.2
[12]	U-Net + Cross Entropy (Transfer Learning)	6.7 ± 1.1	**97.9 ± 0.9**	12.5 ± 2.0	30.2 ± 1.5
[13]	U-Net + Weighted Focal	52.5 ± 6.0	94.3 ± 1.7	67.3 ± 55.2	5.2 ± 7.9
[10]	U-Net + IoU	8.4 ± 14.5	- - -	- - -	2.4 ± 4.2
	U-Net + Focal Combo	**95.3 ± 1.3**	77.4 ± 2.7	**85.4 ± 1.1**	**74.5 ± 1.7**
[14]	Fast-SCNN + Weighted Focal	52.7 ± 8.8	**85.2 ± 3.3**	64.9 ± 6.7	48.7 ± 7.7
	Fast-SCNN + Focal Combo	**66.6 ± 1.7**	71.4 ± 6.9	**68.9 ± 4.2**	**52.9 ± 4.8**

Overall, we can see that the inability of focal combo loss to produce high precision values compared to cross-entropy loss is a limiting factor. This means that in its attempt to focus on harder features, it tends to prioritize minimizing the difference to the actual labeled image, even if it exaggerates predictions at the target class's bounds. Furthermore, it can be seen that it is unnecessarily the better-performing loss function that should be given more weight in the weighted sum. Focal dice loss performed far less than the weighted focal loss, but having a stronger influence on it caused the best performance in our proposed focal combo loss.

In addition, as mentioned in the overview of the methodology, in Figure 12 the extracted road markings can be combined with depth images to generate classified sparse point clouds as an extension. Figure 8 shows that this procedure can produce promising results for both CNN models trained with focal combo loss. However, it is clear that U-Net provides better geometry than fast-SCNN, but given that fast-SCNN has shown prediction speeds of 0.2 s, it shows potential for achieving real-time road marking extraction.

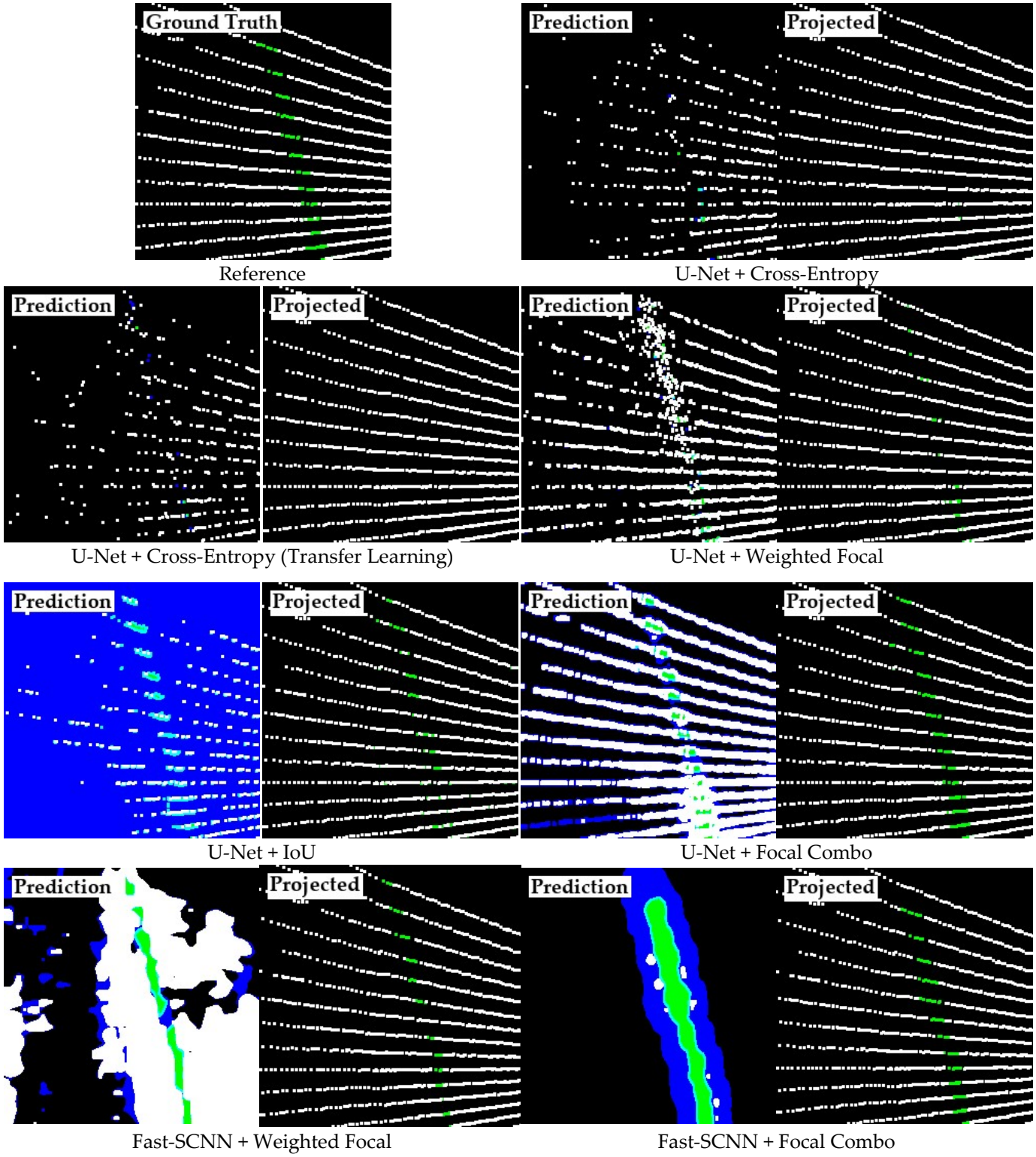
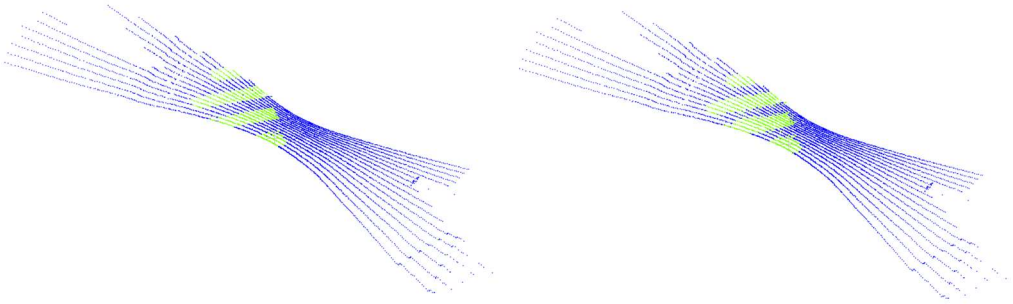


Figure 11. Sample prediction results using existing methods in road markings and their projected counterparts.



U-Net + Focal Combo Loss Fast-SCNN + Focal Combo Loss

Figure 12. Sample generated classified sparse point cloud.

5. Conclusions

In this work, we proposed a modified loss function, focal combo loss, to improve the automatic extraction of road markings from sparse mobile point cloud-derived BEV images using CNNs. It is an attempt to provide expensive alternatives by using data obtained from lower-cost LiDAR sensors rather than survey-grade LiDAR sensors typically installed on a MMS. This can be a practical approach for tasks involving multiple and frequent deployments and involving road markings, such as road monitoring and HD map updates.

Focal combo loss was able to outperform all related existing loss functions as well as existing CNN methods used for road marking extraction on three fronts. First is accuracy, with U-Net and fast-SCNN achieving the highest mean F1-score values of 85.4 and 68.9, respectively. Second is reliability, factoring in uncertainty values of ± 1.1 and ± 4.2 and yielding minimum values of 84.3 and 64.7, respectively, which are the highest among all other minima. The third and final front is versatility, which is reflected in the previous two fronts by consistently being the best in both models.

In future work, we intend to implement our method on more complex road environments with more diverse road markings. Modifications of the other components of the deep learning framework will also be explored, such as the modification of the CNN structure and the training procedure, to further improve road marking extractions on sparse point cloud-derived images. Other feature extraction applications, such as targeting vehicles, pedestrians, and cyclists from a different projection perspective, will also be investigated. This could provide additional support for advocating the use of low-cost sensors for mobile mapping.

Author Contributions: Conceptualization, M.L.R.L.; Methodology, M.L.R.L.; Formal analysis, M.L.R.L.; Writing—original draft, M.L.R.L.; Writing—review & editing, M.L.R.L.; Supervision, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by Tokyo Institute of Technology's WISE Program for Super Smart Society. A special thanks go out to Zongdian Li of Sakageuchi-Tran Laboratory, Department of Electrical and Electronics Engineering, Tokyo Institute of Technology, who assisted and provided expertise during the data collection.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ma, L.; Li, Y.; Li, J.; Wang, C.; Wang, R.; Chapman, M. Mobile Laser Scanned Point-Clouds for Road Object Detection and Extraction: A Review. *Remote Sens.* **2018**, *10*, 1531. [CrossRef]
2. Chiang, K.; Zeng, J.; Tsai, M.; Darweesh, M.; Chen, P.; Wang, C. Bending the Curve of HD Maps Production for Autonomous Vehicle Applications in Taiwan. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8346–8359. [CrossRef]
3. Seif, H.; Hu, X. Autonomous Driving in the iCity–HD Maps as a Key Challenge of the Automotive Industry. *Engineering* **2016**, *2*, 159–162. [CrossRef]

4. Liu, R.; Wang, J.; Zhang, B. High Definition Map for Automated Driving: Overview and Analysis. *J. Navig.* **2020**, *73*, 324–341. [CrossRef]
5. Kashani, A.; Olsen, M.; Parrish, C.; Wilson, N. A Review of LIDAR Radiometric Processing: From Ad Hoc Intensity Correction to Rigorous Radiometric Calibration. *Sensors* **2015**, *15*, 28099–28128. [CrossRef] [PubMed]
6. Chen, S.; Liu, B.; Feng, C.; Vallespi-Gonzales, C.; Wellington, C. 3D Point Cloud Processing and Learning for Autonomous Driving. *IEEE Signal Process. Mag.* **2021**, *38*, 68–86. [CrossRef]
7. Cheng, M.; Zhang, H.; Wang, C.; Li, J. Extraction and Classification of Road Markings Using Mobile Laser Scanning Point Clouds. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1182–1196. [CrossRef]
8. Pan, Y.; Yang, B.; Li, S.; Yang, H.; Dong, Z.; Yang, X. Automatic Road Marking Extraction, Classification and Vectorization from Mobile Laser Scanning Data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W13*, 1089–1096. [CrossRef]
9. Soilan, M.; Riveiro, B.; Martinez-Sanchez, J.; Arias, P. Segmentation and Classification of Road Markings using MLS Data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *123*, 94–103. [CrossRef]
10. Wen, C.; Sun, X.; Li, J.; Wang, C.; Guo, Y.; Habib, A. A Deep Learning Framework for Road Marking Extraction, Classification and Completion from Mobile Laser Scanning Point Clouds. *ISPRS J. Photogramm. Remote Sens.* **2019**, *14*, 178–192. [CrossRef]
11. Lagahit, M.; Tseng, Y. Using Deep Learning to Digitize Road Arrow Markings from LIDAR Point Cloud Derived Images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *43*, 123–129. [CrossRef]
12. Lagahit, M.; Tseng, Y. Road Marking Extraction and Classification from Mobile LIDAR Point Clouds Derived Imagery using Transfer Learning. *J. Photogramm. Remote Sens.* **2021**, *26*, 127–141.
13. Ma, L.; Li, Y.; Li, J.; Yu, Y.; Junior, J.M.; Goncalves, W.N.; Chapman, M. Capsule-Based Networks for Road Marking Extraction and Classification from Mobile LiDAR Point Clouds. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1981–1995. [CrossRef]
14. Elhashash, M.; Albanwan, H.; Qin, R. A Review of Mobile Mapping Systems: From Sensors to Applications. *Sensors* **2022**, *22*, 4262. [CrossRef]
15. Masiero, A.; Fissore, F.; Guarnieerri, A.; Vettore, A.; Coppa, U. Development and Initial Assessment of a Low Cost Mobile Mapping System. *R3 Geomat. Res. Results Rev.* **2020**, *1246*, 116–128.
16. Lagahit, M.; Matsuoka, M. Boosting U-Net with Focal Loss for Road Marking Classification on Sparse Mobile LIDAR Point Cloud Derived Images. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *5*, 33–38. [CrossRef]
17. Lagahit, M.; Matsuoka, M. Exploring FSCNN + Focal Loss: A Faster Alternative for Road Marking Classification on Mobile LIDAR Sparse Point Cloud Derived Images. In Proceedings of the 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 3135–3138.
18. Jadon, S. A Survey of Loss Functions for Semantic Segmentation. In Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, Vina del Mar, Chile, 27–29 October 2020.
19. Wang, Q.; Ma, Y.; Zhao, K.; Tian, Y. A Comprehensive Survey of Loss Functions in Machine Learning. *Ann. Data Sci.* **2022**, *9*, 187–212. [CrossRef]
20. Shorten, C.; Khoshgoftaar, T. A survey on Image Data Augmentation for Deep learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
21. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
22. Poudel, R.; Liwicki, S.; Cipolla, R. Fast-SCNN: Fast Semantic Segmentation Network. *arXiv* **2019**, arXiv:1902.04502.
23. Kohonen, T.; Barna, G.; Chrisley, R. Statistical Pattern Recognition with Neural Networks: Benchmarking Studies. In Proceedings of the IEEE 1988 International Conference on Neural Networks, San Diego, CA, USA, 24–27 July 1988; Volume 1, pp. 61–68.
24. Picek, S.; Heuser, A.; Jovic, A.; Bhasin, S.; Regazzoni, F. The Curse of Class Imbalance and Conflicting Metrics with Side-Channel Evaluations. *IACR Trans. Cryptogr. Hardw. Embed. Syst.* **2019**, *1*, 209–237.
25. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
26. Sudre, C.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalized Dice Overlap as a Deep Learning Loss Function for highly Unbalanced Segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2017; Volume 10553, pp. 240–248.
27. Taghanaki, S.; Zheng, Y.; Zhou, S.K.; Georgescu, B.; Sharma, P.; Xu, D.; Comaniciu, D.; Hamarneh, G. Combo Loss: Handling Input and Output Imbalance in Multi-Organ Segmentation. *Comput. Med. Imaging Graph.* **2019**, *75*, 24–33. [CrossRef] [PubMed]
28. Wang, P.; Chung, A.C. Focal Dice Loss and Image Dilation for Brain Tumor Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11045, pp. 119–127.
29. Tharwat, A. Classification Assessment Methods. *Appl. Comput. Inform.* **2018**, *17*, 168–192. [CrossRef]