ORIGINAL ARTICLE

3D lidar point-cloud projection operator and transfer machine learning for effective road surface features detection and segmentation

Heyang Thomas Li[1] · **Zachary Todd[1]** · **Nikolas Bielski[1]** · **Felix Carroll[1]**

Abstract

The classification and extraction of road markings and lanes are of critical importance to infrastructure assessment, planning and road safety. We present a pipeline for the accurate segmentation and extraction of rural road surface objects in 3D lidar point-cloud, as well as a method to extract geometric parameters belonging to tar seal. To decrease the computational resources needed, the point-clouds were aggregated into a 2D image space before being transformed using affine transformations. The Mask R-CNN algorithm is then applied to the transformed image space to localize, segment and classify the road objects. The segmentation results for road surfaces and markings can then be used for geometric parameter estimation such as road widths estimation, while the segmentation results show that the efficacy of the existing Mask R-CNN to segment needle-type objects is improved by our proposed transformations.

Keywords Machine learning · Object segmentation · Object classification · Geometric parameter estimation · Road segmentation · Deep neural network

1 Introduction

Due to the rise in the development of mobile mapping systems (MMS) and mobile laser scanning (MLS) systems, research in the areas of lane level feature extraction and infrastructure assessment automation has grown quickly in popularity. The classification and extraction of road markings and lanes are of critical importance to effective road infrastructure assessment and planning to improve road safety. Road expenditure towards New Zealand state highways (SH) was estimated to be 2.8 billion in [1]. There are many road hazards for drivers [2], melding MMS and MLS systems coupled with modern computational capability has the potential for cost-efficient compliance and capital planning by automating the characterization of a rural road environment from lidar and camera images.

1.1 Mobile laser system

Previous studies have been developed using image data gathered by a vehicle-mounted camera [3,4], yet these methods collect data sensitive to the variability in environmental conditions. MLS systems use light detecting and ranging sensors (lidar) to collect point-clouds of their surrounding environments as detailed high definition (HD) maps. Unlike image-based methods, MLS systems are invariant to variability in lighting and environmental conditions [5,6]. These HD maps have been used for the extraction and classification of road features such as road markings, drivable regions, road signs and others present in a vehicles MLS defined surrounding environment [7,8], as well as the development of methods for automated infrastructure assessment [9–11]. For example, Jung et al. [12] uses an MLS system to extract drivable regions, road markings and road lines. Their steps were classifying the drivable regions using the distance from the MLS as a threshold and then using intensity to classify the road markings and road lines.

1.2 Road surface and marking segmentation

It is known that an accurate extraction of road surface allows for accurate extraction of road markings [13,14]. Prior

✉ Heyang Thomas Li
thomas.li@canterbury.ac.nz

1 School of Mathematics and Statistics, College of Engineering, University of Canterbury, Ilam, Christchurch 8041, New Zealand

research efforts have focused on finding effective means for both road surface and road marking extraction with much effort spent on the use of intensity-based methods, as road markings greater reflective capability allows for normalized intensity values to show some contrast between markings and road seal. However, the variability in the contrast between road markings and road seal, reflective intensity and point density values is a constant challenge in the use of MLS data [15,16]. For example, Jung et al. proposed a constrained Random Sampling and Consensus algorithm (RANSAC) to transform extracted road surfaces as 2D images before applying techniques to extract and classify road markings. Yu et al. mitigated issues that arise when using a global threshold measure by segmenting road surfaces into distinct patches with their own threshold measures before applying a spatial density filter to further mitigate noise. Although these methods reduced the effects of this variability, they are still not invariant. Many recent research efforts are using deep learning frameworks for road surface and marking extraction. For example, Kim et al. [17] proposed an extreme learning machine for faster computation of image denoising for road marking and edge detection. Tian et al. developed a modified Fast Region-Based Convolutional Neural Network (Fast-RCNN) to generate regions of interests (RoI) containing road markings. A deep convolutional neural network and a finite state machine were applied to aerial images for road surface extraction [18]. Recently Kruz et al. applied a wavelet-enhanced fully convolutional network with HD multiview aerial images for surface extraction. However, all of these methods are influenced by variability in weather conditions and environmental events due to their reliance on image data.

Deep pixel-wise classification methods such as SegNet [19] and ApesNet [20] have been developed for road scene understanding and segmentation tasks, yet have not specifically targeted the road surface and marking extraction. Rui et al. applied U-net, a deep pixel-wise encoder–decoder network, for the segmentation of road crack instances, receiving state of the art results. Yet do not apply this method to segment road surface or its markings. Only recently, Wen et al. proposed a deep learning framework that uses U-net for surface extraction. This was invariant to the variation that intensity threshold methods face, but used standard techniques for the extraction of road markings once the surface instance was predicted. Although deep pixel-wise classification shows promise in its application to the extraction of both road surfaces and road markings from MLS generated 2D projected images, it is a research area with limited contribution to date.

1.3 Road environment geometric parameter estimation

An MLS system's ability to capture HD information of a road environment has gained interest in the ability to automate the estimation of the detected object's geometric characteristics. Recently, Husain et al. used MLS data to detect and estimate morphological parameters of trees in urban road environments. Street trees were first detected in the MLS system-generated point-cloud before being thinned by slicing on each x, y and z axis to reduce computations. After the instance of street trees were detected and processed morphological parameters such as diameters and heights were computed [9]. Other methods follow the steps of detection, classification, processing and estimation of geometric characteristics. Voxelization segmentation and nearest neighbour search were used to detect road highway overhead structures, before being classified using a density-based spatial clustering of applications with noise algorithm followed by the extraction of structural clearance information [10].

Only a few methods have been applied to estimate geometric characteristics of road surfaces using MLS data. Post-vegetation-filtered point-clouds were geo-referenced and extracted from each other to estimate geometric differences and damages of road surface slopes in [21]. Holgado-Barco et al. first used human interaction to set angular thresholds in a road function to extract a road platform before partitioning it into 1m segmented cross sections. Points encapsulating the rectangular plane cross-sections were then adjusted using principal component analysis (PCA) which were then used in a linear transformation to obtain a new coordinate system. Longitudinal and transversal parameters of each road platform were then derived by calculating the eigenvalues and eigenvectors of each transformed cross-section's covariance matrix [11].

The point-cloud-based estimation of the width and volume belonging to rural road surfaces is an under-researched area. This is the same in application to the rural road tar seal component. Only recently Yadav et al. proposed a method that estimates the geometric parameters for road width, road centreline, longitudinal slope and cross slope in a road surface environment. Their method represented the road surface boundary as piece-wise linear segments of a best-fit polynomial, before estimating the road surface width with refined road boundary points divided into n connected road segments along the trajectory of the road boundary [22]. The main disadvantage of recent methods is that they run numerous operations on the point-cloud data to detect, segment and classify the targeted objects before basic calculations are taken to estimate geometric parameters. These operations can be computationally expensive across numerous point-clouds. The computations in Holgado-Barco et al. proposed method are taken on the road surface platform, yet this relies on the

continued user interaction to set angular thresholds for efficient capturing of segments. Yadav et al. method for road surface characterization faced the difficulty of removing the flat level surface surrounding the targeted road segmented, which resulted in a road boundary that captured a road surface segment larger than desired. Another disadvantage in previous methods is that the geometric characterization of the upper layer of chip seal present in point-cloud data has yet to be achieved. We believe that the potential of deep pixel classification shows promise for the segmentation of rural road environment object instances. These accurately segmented instances would allow for the efficient computation of geometric parameters belonging to road surface objects and road tar seal.

In this paper, we introduce a three-part pipeline for the accurate segmentation, extraction and estimation of geometric parameters belonging to road surfaces. Firstly, a framework is introduced for efficient pre-processing and augmentation of point-cloud data. Secondly, a deep learning framework that combines post-processing steps with deep pixel classification for accurate road surface object segmentation and extraction is proposed. In this second framework, the difficulty that the existing Mask Region-Based Convolutional Neural Network (Mask R-CNN) faces in segmenting object instances that are long and thin needle-shaped is resolved through our proposed post-processing steps [23]. In this step, the issue that Yadav et al. faced is also resolved due to the potential of deep pixel-wise classification for segmentation. Thirdly, a framework for the computationally efficient estimation of geometric parameters belonging to road surfaces is proposed using prior inferred instance predictions, including an ablation study in Sect. 4.3. Lastly, a discussion is provided supporting the potential of our introduced three-part pipeline to achieve efficient and effective automation of road surface infrastructure assessment as an improvement from the existing Mask R-CNN.

2 Dataset

This section explains how the dataset was captured, the format the dataset is in post-capture, and the pre-processing steps used to obtain the sample populating. The section then goes into detail about the three types of road objects being looked at (road surfaces, markings, and seal) and then describes the steps to obtain the augmentation population and which data is being excluded.

2.1 Dataset capture

A series of raw 3D point-clouds of different state highways was captured by a mobile mapping vehicle (MMV) mounted with an MMS using lidar (pictured in Fig. 1).
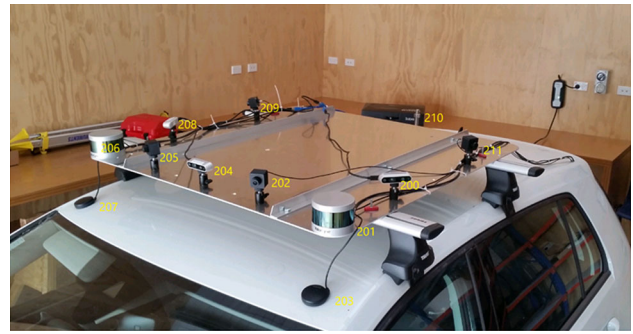
Fig. 1 Mobile mapping vehicle used in the gathering of point-cloud data. The sensors and additional equipment included lidars (201 and 206), GPS (203 and 207), depth cameras (200, 204 and 208), cameras (202, 205, 209 and 201) and the power supply (210)

2.2 Point-cloud pre-processing

Previous methods mentioned used pre-processing steps in order to capture relevant information from the point-cloud, reduce computational complexity during the learning stage and save on memory space [5,6,8,24]. We implemented a similar process to the point-cloud partitioning pre-processing steps used by Soilan et al. with the points-cloud partitions overlapping by roughly 3%.

We projected point-cloud points onto a 2D plane capturing intensity and height information in RGB channels which we then saved a 1280x1280 image. The red colour channel represents the intensity, this was calculated by averaging the intensity of all the points within the pixel. The intensity was then capped to the 99th percentile and normalized. The same process was applied to the blue colour channel using the height of the points instead of the intensity and capping at the 95th percentile. The green colour channel is an average of the red and blue colour channels.

Figure 2 shows the before and after of the prepossessing, with the road marking being redder than the road surface and the trees being bluer than the rest of the image.

2.3 Annotations

VGG Image Annotator (VIA) [25] was used to create ground truth masks of the road surface and road marking instances for each image. Each of these ground truth masks was generated by annotating the images. The ground truth objects were annotated by encapsulating the relevant image segments within polygons, to ensure that only the pixels relevant to an object were included. A Boolean matrix M with the same dimensions as the image is assigned to each object with each $M_{ij} \in M$ being marked true if it is contained inside the polygon.

(a) Point-cloud   **(b)** Image
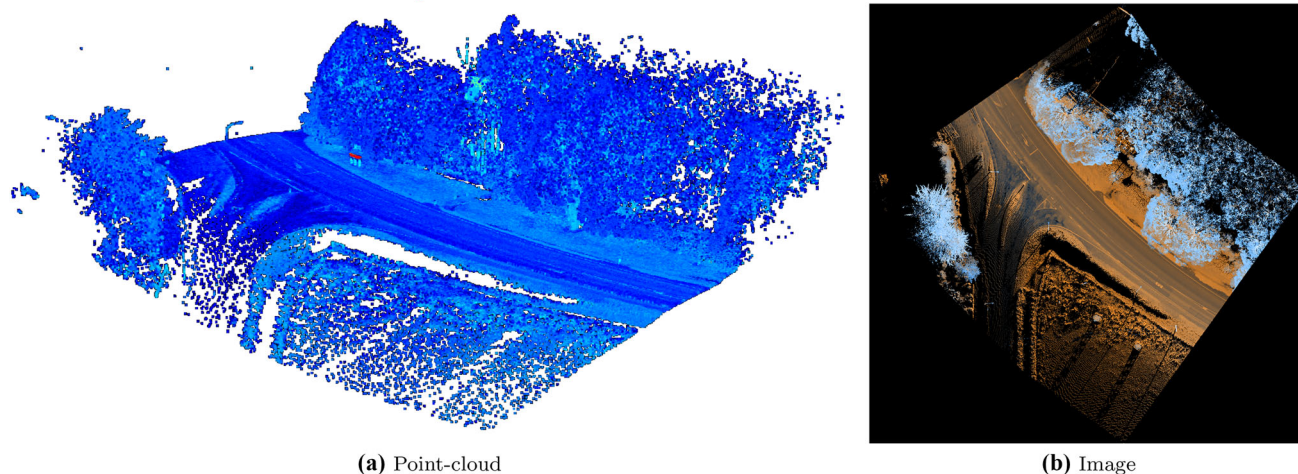
Fig. 2 Road environment can clearly be seen in the point-cloud data. The image has intensity in the red colour channel, height in the blue colour channel and an average of the two in the green colour channel

2.4 Class definitions and components

As the point-cloud was captured along New Zealand state highways the class definitions assigned during the annotation stage were consistent with part two of the Manual Of Traffic Signs And Markings (MOTSAM) policy [26,27]. Typically, a rural road environment includes objects static in nature such as the road's material, markings, road signs and vegetation. Ma et al. [28] described five categories of on-road object inventory consisting of; road surfaces, road markings, driving lines, road cracks, and road manholes. Additional class names outside of MOTSAM standards were assigned for seal instance segmentation, as well as for the drivable area of the road surface. In cases where a road marking included gaps between striped white bars, we included this gap in the object's instance, as we believe this is part of the object instance.

We have split the classes in this paper into three road object categories:

1. Road Surface; defined as the area in the road environment that lies within the outer bounds of the side markings.
2. Road Surface Markings; consisting of three classes that exist within the Road Surface; drivable lane, driving Line and hatched area.
3. Road Seal; defined as the area in the road environment that is the surface layer of chip seal laid when the road was built.

A road surface instance is the m x m binary matrix M where each value M_{ij} is contained within the outer bounds of the road surface's driving line markings on the y-axis, and whose maximum and minimum x-coordinate values equal to that of the x-axis.

The three surface marking classes are; drivable lane, driving Line, and hatched areas. The densities of classes in the images differ according to location and class count (as shown in Fig. 3 and Table 1). Driving lines are markings that partition the road surface into drivable lanes. These include both the continuous white edge lines located at the outer boundaries of the road surface and the centrelines. Drivable lanes exist as the space between driving lines. They are the densest object class with the highest class count in the dataset (as shown in Table 1). Driving lines are dense nearer to the centre of the augmented images with some sparsity caused by the inclusion of edge and centre lines.

The seal is defined as the area along with a road environment where the road is coated with chip seal. In the image, the seal instance can be seen along with the road's environment due to its low reflective intensity and changing height values.

2.5 Data exclusions

The mapping vehicle started and ended its data collection in an urban environment before traversing rural areas. When processed as 2D images, the urban roads made up 13.8% of the dataset before data augmentations. The urban road included a large variety of road marking and include additional road objects on top of the road surface such as roundabouts and crossing islands. These objects significantly increase the variation on the road surface. Considering this and that information on urban areas not being relevant to the task of rural road surface level feature extraction, the urban images were excluded.

(a) Surface **(b)** Seal Heatmap

(c) Drivable Lane **(d)** Driving Line **(e)** Hatched Area
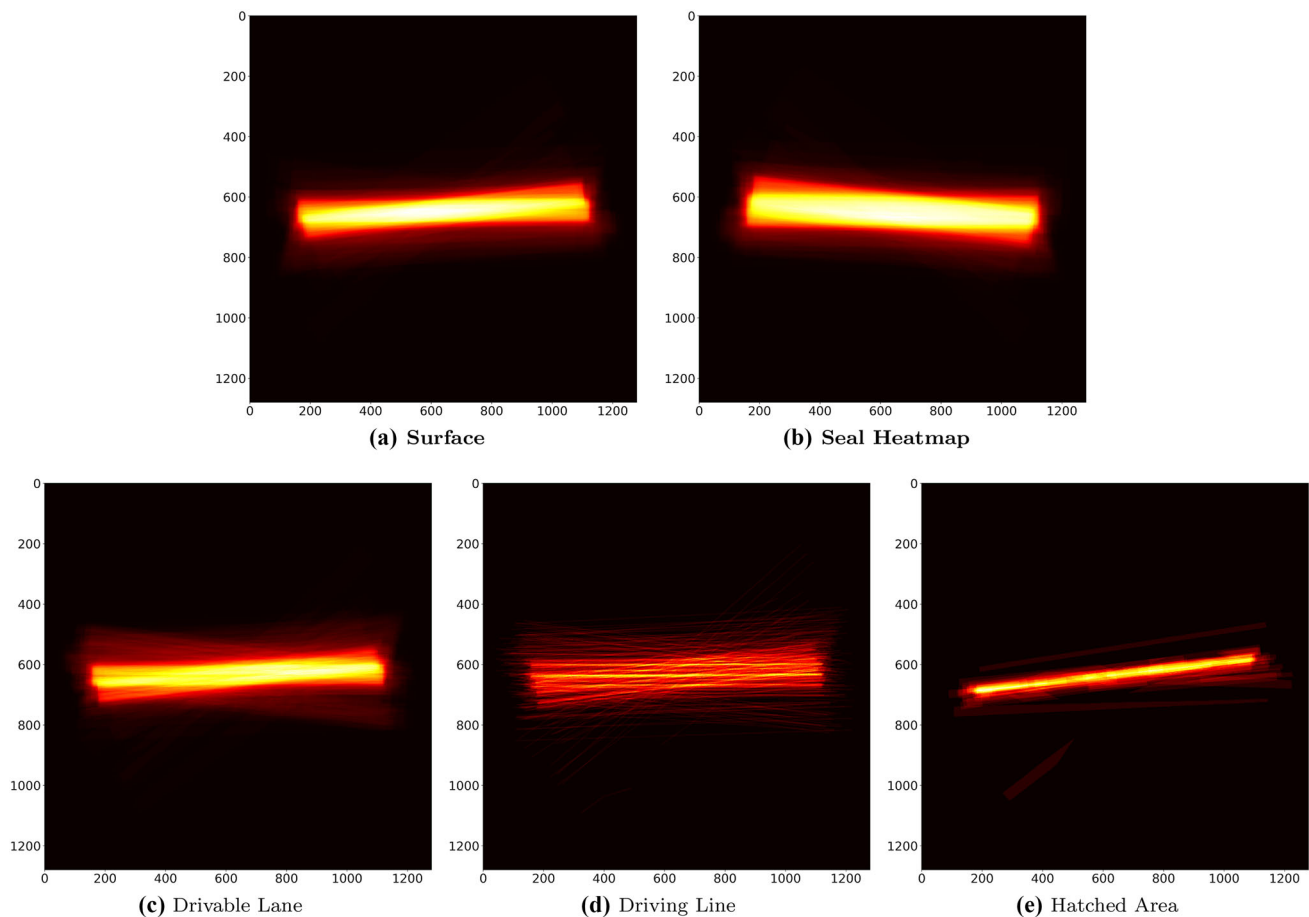
Fig. 3 Road feature heatmaps post 65 degree rotation. The lighter the pixel is in the heatmap the larger number of labelled instances intersect with that pixel

Table 1 Class counts in dataset after augmentations

Set	Drivable Lane	Driving Line	hatched Area
Train	1456	995	175
Validation	476	330	51
Test	448	327	58
Total	2420	1652	284

2.6 Data augmentations

To increase the dataset size, each image went through a series of augmentations. Before the augmentation, the images were rotated 65 degrees as the majority of the surface mask were oriented between 60 and 70 degrees. The rotation increased the relative squareness of the masks, which enables the masks to capture more of the annotated polygon as lines that are not strictly horizontal or vertical result in more jagged edges to mask. The set of augmented applied to the image were a 180 degrees rotation, a horizontal flip and a 180 degrees rotation followed by a horizontal flip. These steps increased the dataset size by 400% to 772 images. The only issues with these augmentations are that some road markings are direction-dependent (e.g. hatched areas) adding features that the algorithm needs to extract which do not exist in the physical environment (e.g. backwards hatched areas).

The dataset is then split into training, validation and testing sets as 60%, 20%, and 20% splits, respectively. Using image augmentation to increase the sample populations size exposes the Mask R-CNN algorithm to a larger amount of examples to train, validate and test.

3 Methods

This section explains the Mask R-CNN algorithm and how it is used for segmentation, localization and classification. The section also goes into detail about the specific pre-processing for the road surfaces, markings and seal object and the uses of PCA and affine transformation to increase size-relevant features. The section also describes the apparatus used to run the various methods used.

3.1 Mask region-based convolutional neural network

Mask R-CNN is an extension to Fast RCNN [29] [30] that uses pixel-wise binary classification and alignment for object instance segmentation in parallel to the Fast RCNN's branch for bounding box recognition. Using the same first stage of fast RCNN, mask R-CNN takes the region proposal network (RPN) and predicts a binary mask for each Region of Interest (RoI) which is then passed through an alignment layer for more accurate pixel-to-pixel segmentation. In this paper, we applied Mask R-CNN with a Resnet-101 backbone architecture. Pre-trained weights from the COCO-2016 dataset were used for transfer learning [31]. The extension to Fast RCNN outputs a binary mask for each RoI and kth class where there are K classes in total. A sigmoid function is then applied to each pixel using an average binary cross-entropy loss l_{mask}. The loss function for Mask R-CNN adds l_{mask} onto the loss function for Fast RCNN, as follows:

$$l = l_{class} + l_{box} + l_{mask} \tag{1}$$

l_{class}, l_{box} are same as they are in Fast RCNN, while l_{mask} is computed on RoIs that are associated with the ground truth k^{th} class and is defined as the average binary cross-entropy loss;

$$l_{mask} = \frac{-1}{m^2} \sum_{1 \leq i, j \leq m} \Big[M_{ij}^k log(\hat{M}_{ij}^k)$$
$$+ (1 - M_{ij}^k) log(1 - \hat{M}_{ij}^k) \Big] \tag{2}$$

where M is a Boolean m x m matrix for the true mask, M_{ij} is the label of $(i, j) \in M$. $\hat{M}_{ij}^k$ denotes the predicted value of (i, j) in the mask that Mask R-CNN learns for the ground-truth class k.

3.1.1 GPU computational equipment

The GPU used to train the models was an Nvidia Titan Xp graphics card with 12GB memory running with CUDA 10.1 and an Nvidia driver version 418.13. The Mask R-CNN implementation uses the Matterport Mask R-CNN repository [32].

3.2 Proposed framework

After pre-processing each task is completed as a separate branch, as shown in Fig. 4. An ensemble method is used for the surface and marking extraction tasks, where Mask R-CNN is used again with post-processing steps for better feature extraction. Surface seal extraction and measurement

is achieved with a single use of Mask R-CNN and its sub-framework for the width estimation stage.

3.2.1 Road surface

To separate the instance of the surface from its surrounding noise, we used Mask R-CNN for binomial instance classification and segmentation between the surface and background. Images were downsized to 1024 x 1024 due to hardware memory limitations. A learning rate of 0.001 was used with a weight decay rate of 0.0001. The mini-mask shape was set to 400 by 400. Training RoIs per image was set to 200 with a detection minimum confidence of 80%. The anchor sizes for the RoIs pooling stage were 50, 150, 300, 500 and 800.

3.2.2 Road surface markings

Deep pixel-wise methods such as the methods we are using rely on training images where the instance needs to be segmented has a resolution as high as possible relative to available memory capability as it can so that the potential for extracted features during convolutions is maximized. The pixel-width of the road markings in each image were 1–2 pixels wide and only took up 5% of the image information which made it difficult for the algorithm to utilize the road marking information without up-sampling as the pixel width was too small for feature extraction. Also, up-sampling alone required a larger amount of processor memory to store the images during training. To solve these issues PCA was used to transform the image to increase the pixel count of the road markings [33].

To mitigate these issues during training, the following augmentation steps were taken of the training, test and validating data set (as demonstrated in Fig. 5.).

(a) Detecting the surface segment. The surface segment is detected using Mask R-CNN trained for surface segmentation.
(b) Fitting a PCA with two components on the predicted surface segment.
(c) Calculating the PCA's bounding box. This is then used to aid in the affine transformation in the next step.
(d) Inverting the PCA bounding box coordinates back to the image space with an affine transformation to map the predicted masks to the transformed image and increase the relevant features.

Another issue that presented itself was that when Mask R-CNN extracted features for each of the different classes, changes in the geometric shape of objects caused by the curvature in the road directly negatively affecting Mask R-CNN's performance. In order to mitigate this multiple PCAs
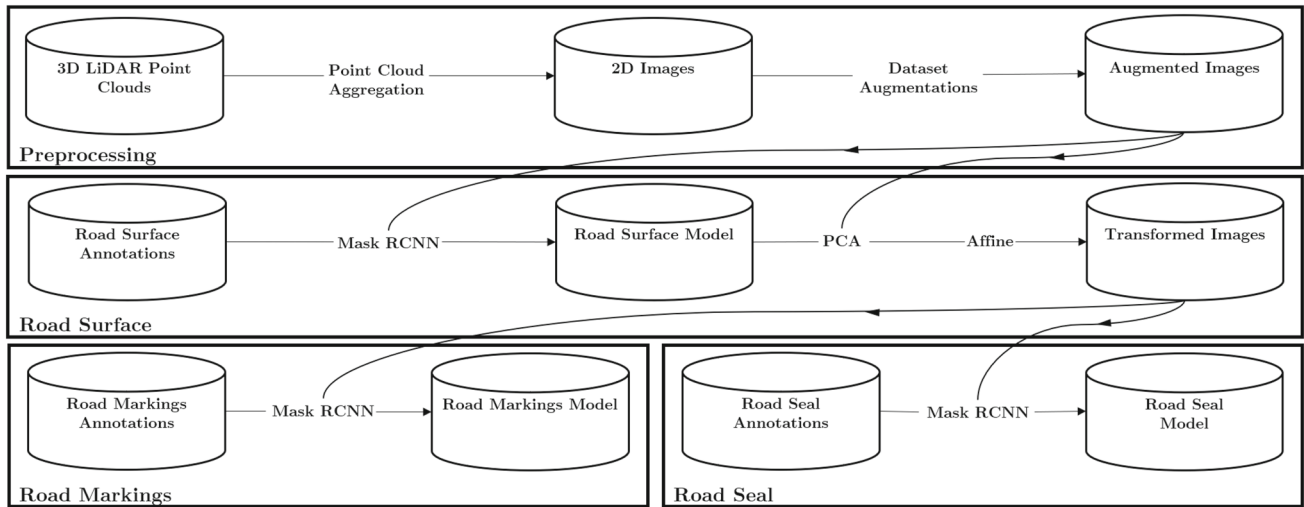
Fig. 4 Proposed framework

(a) Surface Example   **(b)** PCA   **(c)** PCA's Bounding Box   **(d)** Affine transformation
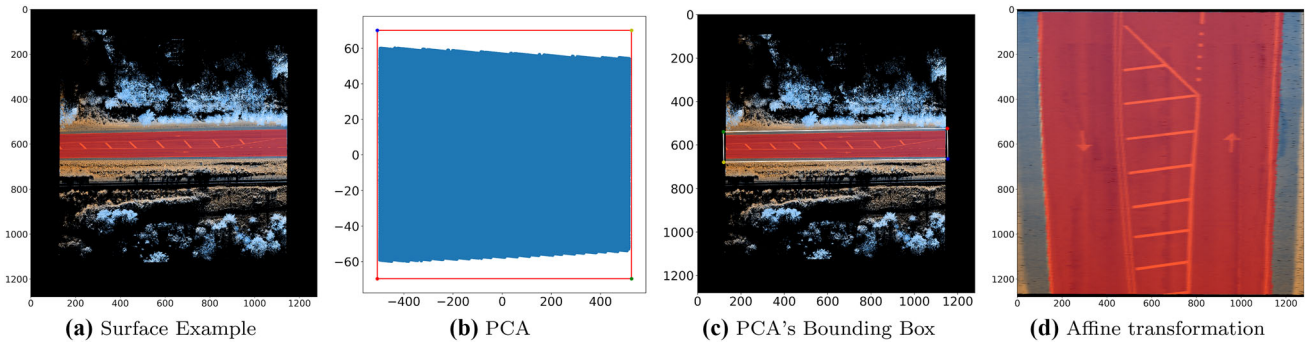
Fig. 5 PCA process

were taken at step (b) in Fig. 5 where a PCA was taken on the first PCA mask resulting in two PCA masks, followed by another PCA on each of the two PCA masks in order to derive at four distinct PCA masks for the detected instance. Each of these distinct PCA masks are then inputted into step (c–d) in Fig. 5 where the output of each is inputted into Mask R-CNN.

This step-wise framework manages to both increase road marking's pixel count in the image and remove the effect of road direction curvature as well as retain the segmented surface in the PCA inverted images.

3.2.3 Road surface configurations

The surface markings multinomial classification and segmentation task involved the three surface marking classes mentioned earlier plus the background. A learning rate of 0.001 was used with a weight decay rate of 0.0001. The mask shape was set to 28 x 28 and mini-mask shape set to [200, 200]. The anchor sizes were 25, 75, 150, 300 and 600 with ratios 0.2, 0.4, 0.6, 0.8 and 1.

3.2.4 Road seal

Mask R-CNN was used for binomial classification and instance segmentation of the tar seal as shown in Fig. 4. The mask's shape was often irregular along with the outer bounds of the seal due to tar seal's liquidity and wear, unlike with the road surfaces which often had a flat-sided shape. This made the mini-mask configurations of higher importance for successful pixel-wise segmentation for seal than for surfaces.

Road seal was treated as a binomial classification task like road surfaces with road seal and background as the two classes. The boundary between the road seal and the road shoulder encompasses a relatively small part of the image, as this boundary is of a similar size to the road markings we used the same PCA process used for the road markings to enlarge the road seal boundary. The setting we used for the Mask R-CNN model where: learning rate of 0.001, weight decay rate of 0.0001 and anchor sizes used for RoI pooling were 50, 150, 300, 500 and 800 with ratios 1.2, 1.8 and 2.5.

(a) Straight Rural SH **(b)** Curved Rural SH **(c)** Straight Urban SH **(d)** Narrowing Urban SH
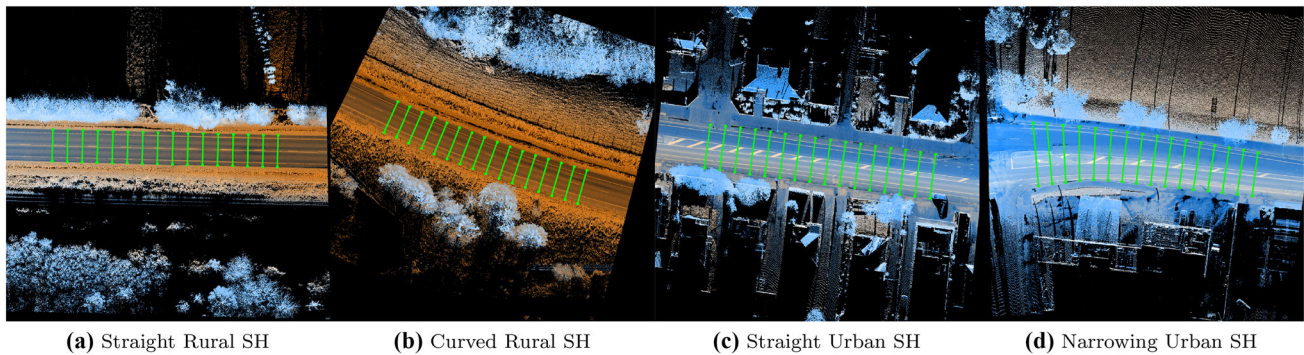
Fig. 6 Ground Truth Seal Widths Examples. Widths in green and $n = 16$ and $r = 150$)

Algorithm 1 Widths Estimation

    Inputs: M (road mask),
          n (number of road widths) and
          r (length from boundary)
    Requires: $2r > range(\{y | (x, y) \in M\})$
    Output: w (road widths)

1: $M_{pca}, M_{comp}, M_{mean} \leftarrow PCA(M)$
2: $step \leftarrow \frac{1}{n}(range(\{x \mid (x, y) \in M_{pca}\}) - 2r)$
3: **for** $i \in \{min(M_{pca}) + kstep \mid k \in [0, n-1]\}$ **do**
4:     $L \leftarrow \{i \leq x < i + 2r \mid (x, y) \in M_{pca}\}$
5:     $L_{pca}, L_{com}, L_{mean} \leftarrow PCA(L)$
6:     $L_{aff} \leftarrow median(L_{pca}) \cdot L_{comp} + L_{mean}$
7:     $M_{aff} \leftarrow L_{aff} \cdot M_{comp} + M_{mean}$
8:     $w \leftarrow w \cup \{M_{aff}\}$

3.3 Width estimation

Our proposed algorithm works by first taking a mask M and performing a PCA on it to output its mask M_{PCA}, its component M_{comp} and mean M_{mean}. M_{PCA} is then partitioned across its x-axis where the number of partitions n is determined after subtracting r length from the boundary. A second PCA L_{PCA} with its component L_{comp} and mean L_{mean} is then taken centred around each n point with a window of r length either side. Finally, each of the points are readjusted to the mask space by first performing a local affine transformation using L_{comp} and L_{mean}, and then performing an affine transformation using M_{comp} and M_{mean}. The result is then added to the set of widths w. Limitations of this are that the correct selection of r is needed to ensure that L_{PCA}'s x range is larger than its y range which results in the selected widths to rotate and select a point near the middle of the surface of the road. An example of the ground truth width measurements with $n = 16$ is provided in Fig. 6.

4 Experiments and results

This section explains the experimentation carried out to test the proposed method and how these experiments are evaluated. This section then describes the experiment outcomes.

The proposed framework was evaluated using the 20% test population after data augmentation. The test population post data augmentation was used rather than sample population as the steps required to create data augmentation are inevitable and not bias towards the predicted results of their structure (i.e. predicting on a non-sample population image is not dependent on first having a ground truth mask for that image). The performance of the localization and segmentation branches for the road surface, road surface markings and tar seal tasks are evaluated independently before results are outlined for geometric parameter estimation tasks.

4.1 Localization and segmentation statistical evaluation methods

For evaluating the performance of the tasks of instance detection and segmentation two common metrics were used. The Intersection over Union (IoU) metric (also known as the Jaccard Index) calculates the spatial overlap of predicted and true m x m binary matrices in proportion to their spatial union. The IoU metric is used for binary m x m matrices that are either the predicted RoIs for localization, or the predicted masks for segmentation.

For example; Let $\hat{X}^n$ be the predicted m x m binary matrix for the nth input image Img_n and true binary m x m matrix of Img_n being X_n. The $IoU(n)$ is then determined by:

$$IoU(n) = \frac{\left| \hat{X}^n \cap X^n \right|}{\left| \hat{X}^n \cup X^n \right|} \tag{3}$$

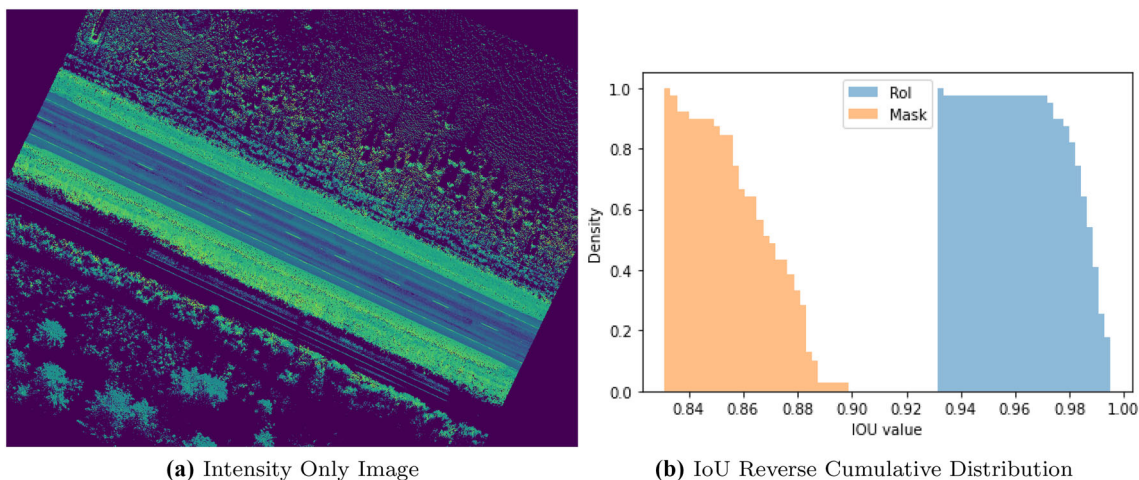**(a)** Intensity Only Image   **(b)** IoU Reverse Cumulative Distribution

Fig. 7 An example of the image being used and the test results using only the point-cloud intensity information

4.2 Classification statistical evaluation methods

To evaluate the performance of our model's classification of the different road markings the following metrics were used.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = \frac{2Recall \cdot Precision}{Recall + Precision} \quad (7)$$

where the predicted class types are compared to the ground truth class types. TP, TN, FP and FN represent the amount of true positive, true negative, false positive and false negative classification of the road object being tested.

4.3 Ablation study

This subsection explains why some design decisions and how they may impact the experiment results.

4.3.1 Road surface colour channel selection

The pipeline above uses three colour channels for the images. However, the mask R-CNN algorithm does not require this and using less colour channels would decrease the computation resource needed, as it would decrease the dimensionality of input images. The choice of using three colour channels instead of one is a trade-off between computation complexity and the inclusion of potentially useful image information.

The results in Fig. 7 show that the IoU results are good with all test RoIs having greater than 93% predicted IoU and all test masks having a greater than 83% IoU despite only

using one colour channel. However, there is a gap between the prediction results RoIs IoU and the Mask IoU and because road surface are rectangular and similar to their bounding boxes using additional data and adding colour channel maybe increase the quality of the predicted masks. It follows that if there is possible improvement for road surfaces classes than fine-grain problems such as markings and seal may also have an improvement.
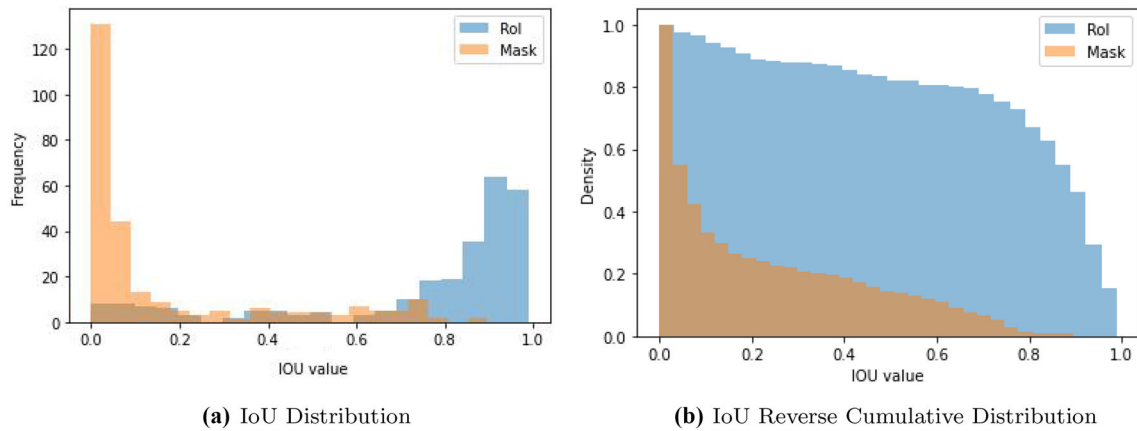
4.3.2 Road markings pre-processing

The main goals of the road markings pre-processing is to improve the results of the road markings via increase the relative size of the markings with in the image. The idea being that a more prominent object relative to the background of an image should be more visible to the Mask R-CNN algorithm. The down side of using our pre-processing pipeline is it requires additional computational steps that maybe not significant improve the prediction.
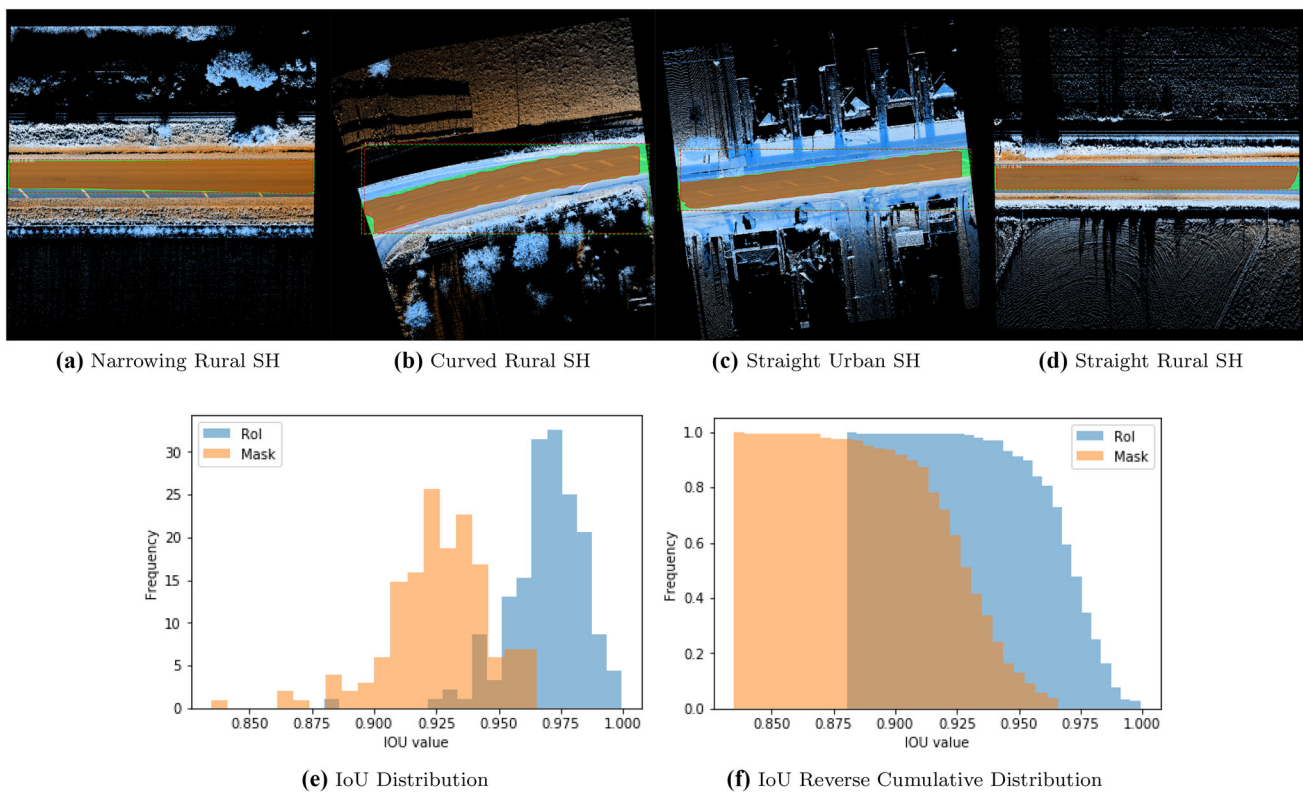
The result in Fig. 8 show that the RoIs IoU is much better than the mask IoU, with 60% of the marks having less than 10%. The relative good result in the RoI can be attributed to the bounding box size of each marking as they span the length of the image thereby having a large amount of information to base the RoI prediction on. The poor mask result are like due to the small size of the markings as they are only 1–2 pixels wide. This shows that implementing a pipeline to increase the prominence of the marking may increase marks IoU.

4.4 Road surface results

Road surface predicted masks median IoU was 93.13%, where at an IoU threshold of 90%, 90% of the test sample is included. Increases in the acceptance threshold past this point results in a steep decrease in the amount of test popula-

(a) IoU Distribution

(b) IoU Reverse Cumulative Distribution

Fig. 8 Test IoU scores for RoIs and masks without using road markings pre-processing

(a) Narrowing Rural SH   **(b)** Curved Rural SH   **(c)** Straight Urban SH   **(d)** Straight Rural SH


(e) IoU Distribution

(f) IoU Reverse Cumulative Distribution

Fig. 9 Road surface test results: **a–e** examples of predicted masks for road surface instances compared to ground truth masks, **e–f** test IoU scores for RoIs and masks

tion being accepted (as shown in Fig. 9e–f). The localization result improves upon this by having 100% of the test population being accepted up to 92.5% IoU acceptance threshold with a steep decline in test population accepted rate at 95% IoU. Our proposed framework predicts masks for road surface instances successfully across road surfaces with varying spatial characteristics (as shown in Fig. 9a–d), where predicted masks accounted for road surface instances that were either tapered or curved in shape throughout the MMS trajectory. One issue with the road surface predicted masks was that the ends of the predicted masks were enclosed within the ground truth masks. In instances where the road was curved, the predicted masks would also contain wavy bounds. This effect can be minimized by having overlapping images, allowing for a larger acceptance threshold to be implemented without decreasing the amount of validation population being selected.
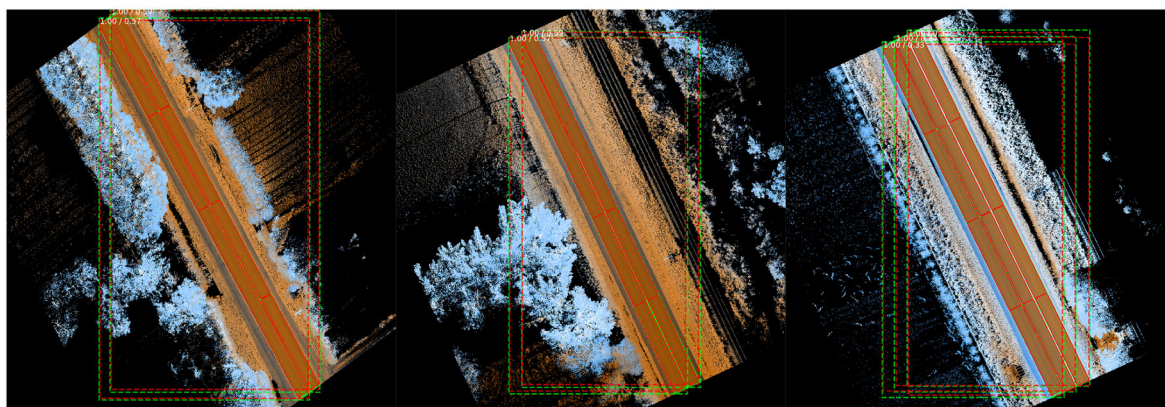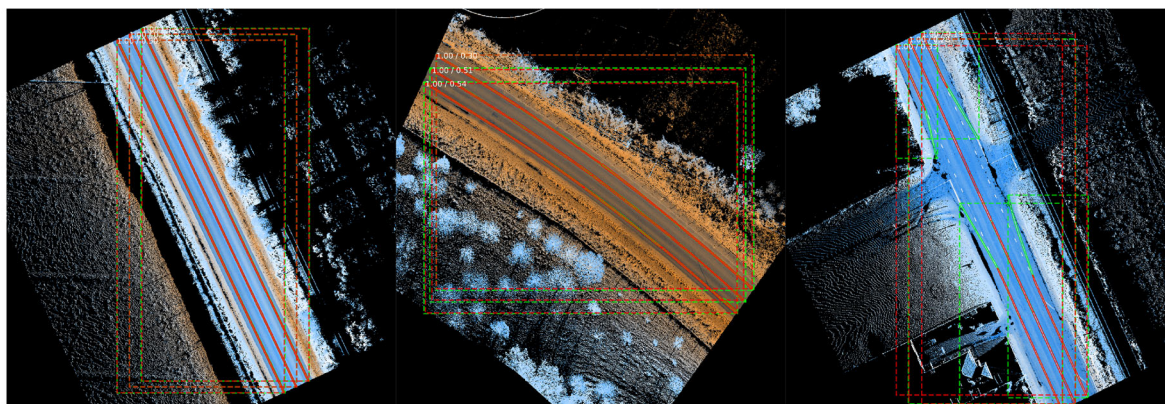
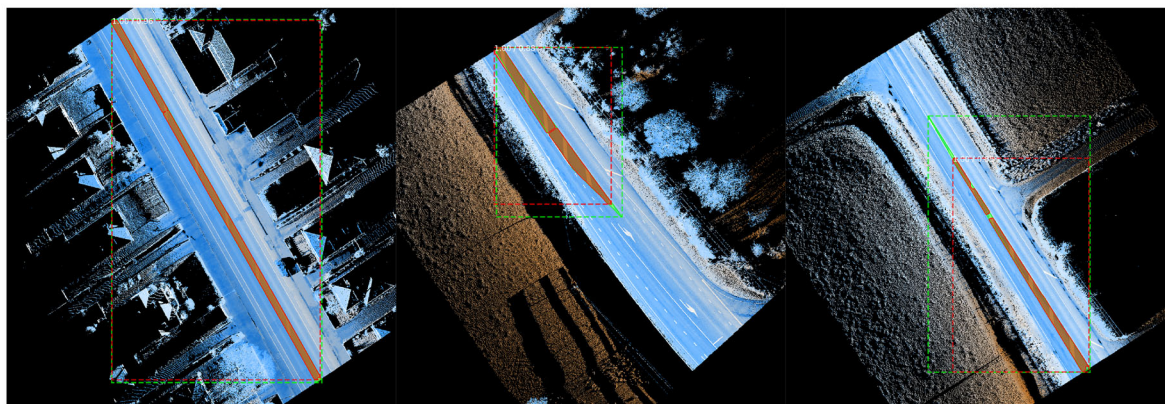**(a)** Best Drivable Lane Result **(b)** Typical Drivable Lane Result **(c)** Poor Drivable Lane Result

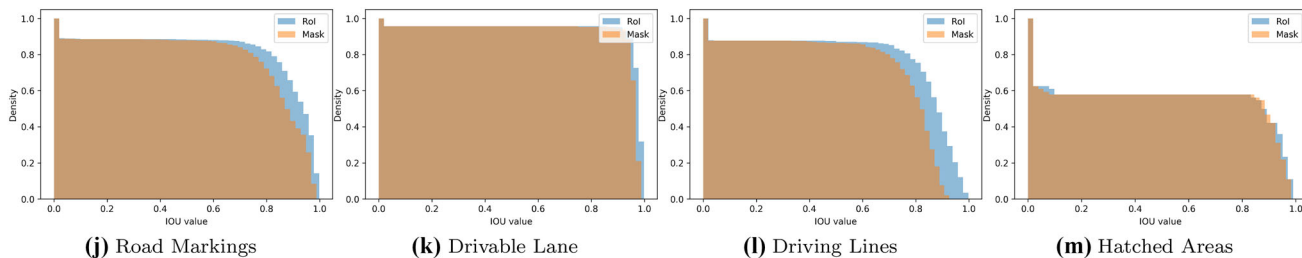**(d)** Best Driving Line Result **(e)** Typical Driving Line Result **(f)** Poor Driving Line Results

(g) Best Hatched Area Result **(h)** Typical Hatched Area Result **(i)** Poor Hatched Area Result

(j) Road Markings **(k)** Drivable Lane **(l)** Driving Lines **(m)** Hatched Areas

Fig. 10 Road marking segmentation results. In (**a**–**i**) the green and red represent the ground truth and predicted mask respectively. **j**–**m** are reverse cumulative IoU distributions

Table 2 Road Markings Evaluation Metrics at 90% IoU

Class	Accuracy	Precision	Recall	F1
Drivable Lane	70.44%	69.05%	84.15%	75.86%
Driving Line	75.52%	91.82%	58.41%	71.40%
Hatched Area	78.54%	5.33%	6.90%	6.01%

Table 3 Road Markings Confusion Matrix at 90% IoU

	Drivable Lane	Driving Line	Hatched Area	Unclassified
Drivable Lane	377	12	24	35
Driving Line	104	191	32	0
Hatched Area	21	4	4	29

4.5 Road markings results

The classification results show that our proposed framework performs best at classifying Driving Lines with a precision score of 91.82% (as shown in Table 2). Average classification accuracy across classes was 77.78%.

Figure 10j–m demonstrates the reverse cumulative distribution of IoU scores for predicted RoIs and masks for the aggregated road markings class and the road markings classes individually. The median IoU score for the predicted RoIs across all classes is greater than 95%. The median RoIs IoU for drivable lanes was the best among the three classes with 97%. Using either 80% or 90% as the RoIs IOU acceptance threshold would result in 97% and 96% of the test population being accepted respectively. The predicted RoIs for Driving Lines had a median IoU of 96%, using IoU acceptance thresholds of 80% and 90% results in 84% and 77% of the test population being accepted respectfully. The results for hatched areas received the lowest IoU scores among the three classes with a median of 57% and at IOU thresholds of 80% and 90% having 57% and 50% of the test population, respectively.

Our method also scored highly in the instance segmentation task across drivable lane and driving line but has some difficulty with hatched areas. (Fig. 10j–m). The median IoU score for predicted masks across all classes was 71.34%. The median IoU scores for predicted masks for drivable lane, driving line and hatched areas were 97.5%, 89%, and 57%, respectfully. Our method performed best at segmenting instances of drivable lane with 94% of the test population being accepted at a threshold of 90%. Second to this, driving lines instances were segmented with 75% of test population belonging to an IoU acceptance threshold of 80% and 31% belonging to 90%. Hatched areas segmentation results performed the worst out the three classes with 57% and 43% of

test population being accepted at IoU thresholds of 80% and 90%, respectively.

The classification results for road markings reflect the overall performance of their segmentation and locations, as shown in Tables 2 and 3. The hatched areas perform the worst out of the three road markings classes with accuracy, precision and recall of 78.54%, 5.33% and 6.90%, respectively. The high accuracy in comparison to the precision and recall is because of the high true negative rate, which more highlights the strengths of the drivable lane and driving lines results rather than the hatched areas results. The drivable lane and driving lines classes perform the best out of the three classes. However, driving lines have a larger proportion of its test population misclassified as a drivable lane.

4.6 Road seal results

Results show that our proposed framework performs well with localizing tar seal instances with a median IoU for RoIs prediction of 95.45% and at an acceptance threshold of 90% results in 95% of predicted RoIs being accepted (Fig. 11e–f). Upon localization, our method performs well on the segmentation of tar seal instances with a median IoU of 92.81% for predicted masks. The predicted masks at an IoU threshold of 90% results in 86.6% of the test population being accepted.

4.7 Geometric parameter estimation results

The results in Fig. 12 show that our proposed method effectively outputs a set of widths for predicted tar seal mask instances. Lines 1 to 9 in Fig. 12d (counting from the left) allowed for changes in road surface shape along its trajectory to be included in each of the estimated widths making estimations flexible to such changes. Figure 12d also shows that the width calculation can only perform as well as the predicted mask as lines 5 and 6 have ends that are over the pavement.

5 Discussion

This section discusses the results from previous sections, including the following sub-sections: Ablation study to compare our added PCA approach to improve the mAP of Mask-RCNN's IoU; Road surface discussion in comparison with prior works; Road surface marking discussion on finding object instance of 1–2 pixels wide; Road seal and geometric estimations to discuss some applications, and our method have certain benefits and limitations.
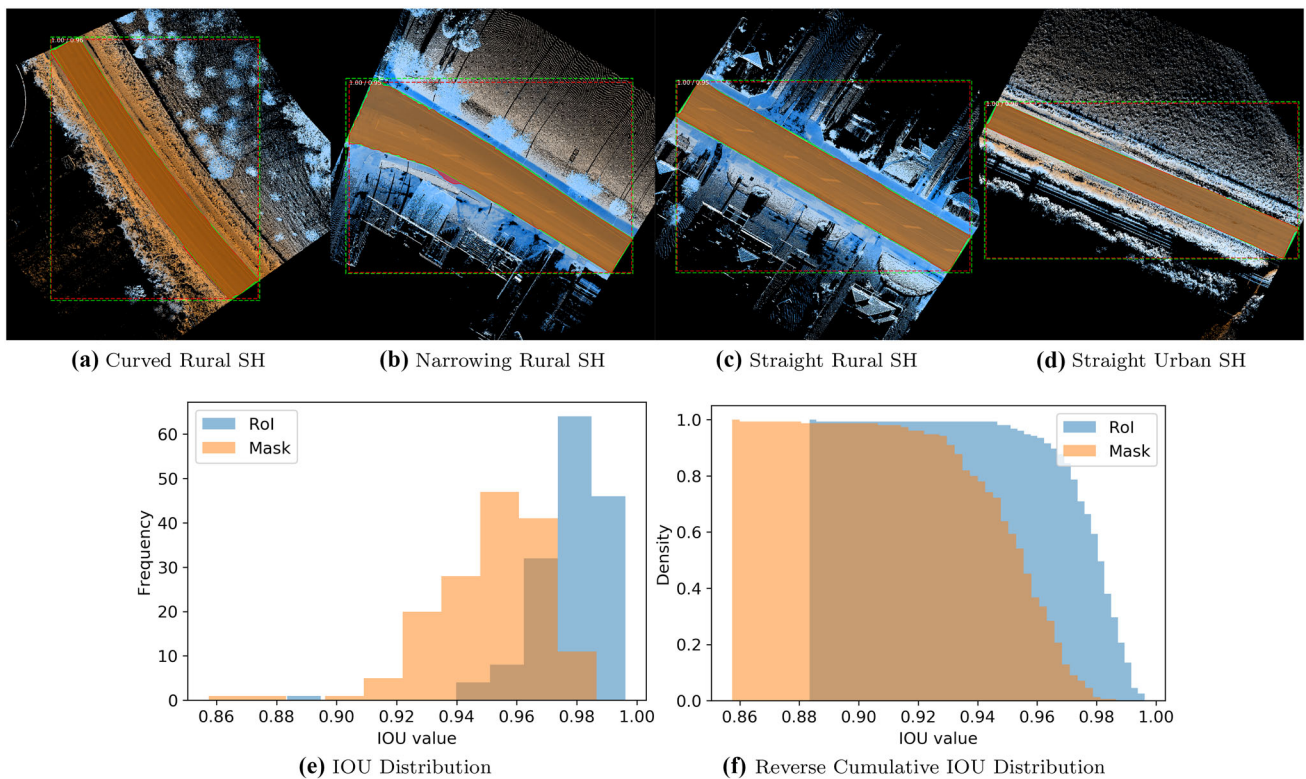
(a) Curved Rural SH **(b)** Narrowing Rural SH **(c)** Straight Rural SH **(d)** Straight Urban SH

(e) IOU Distribution **(f)** Reverse Cumulative IOU Distribution

Fig. 11 Road seal test results: **a–e** examples of predicted masks for road surface instances compared to ground truth masks, **e–f** test IoU scores for RoIs and masks
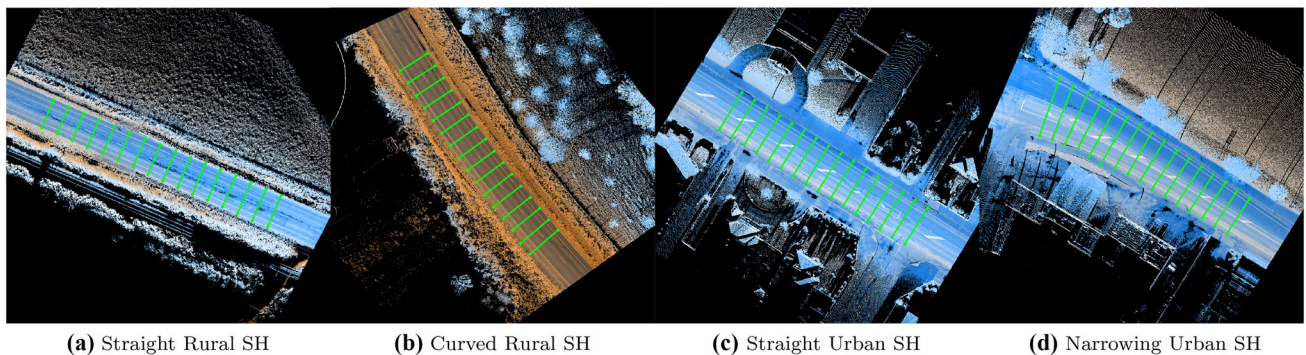

(a) Straight Rural SH **(b)** Curved Rural SH **(c)** Straight Urban SH **(d)** Narrowing Urban SH

Fig. 12 Predicted seal widths examples. Widths in green and $n = 16$ and $r = 150$)

5.1 Ablation study

5.2 Road surface colour channel selection

The results in Table 4 show that the including the height values in the images improves the Mask IoU results. With 0% of the masks having a greater than 90% IoU when not including the height values and 96.% of the masks when using the height in the images. The trade-off for this increase in mask IoU is that the RoI results start decreasing more rapidly over 95% IoU when including the heights. This shows the exist-ing Mask R-CNN's accuracy improved with our inclusion of height colour channel.

5.2.1 Road markings pre-processing

The results in Table 5 show a significant improvement. Improving from 16.2 to 55.0% at 50% IoU and from none to 53.6% at 90% IoU. The existing Mask R-CNN's accuracy improved with our inclusion of both height colour channel and PCA affine transformation, especially in accurately seg-ment the needle type objects such as 1–2 pixel wide road line markings.

Table 4 Comparison of including height in colour channels

Metric	Without Height	With Height
mAP @.75 Mask IoU	100.0	100.0
mAP @.90 Mask IoU	0	96.4

Table 5 Comparison of the inclusion of the PCA pipeline

Metric	Without PCA	PCA
mAP @.50 Mask IoU	16.2	55.0
mAP @.75 Mask IoU	11.4	55.0
mAP @.90 Mask IoU	0.0	53.6

5.3 Road surface

Our proposed framework segments and extracts road surface instances without the issue of including the flat level surface surrounding faced by Yadav et al. , as well as removes their method's need to first extract planar ground surfaces before extracting the targeted road surface. Unlike Rui et al. and Wen et al. , our deep pixel-wise classification approach eliminates the need to rely on relative raised curb locality to segment targeted instances, thus also removing the need for post-processing on the segmented surface to extract the target instance. The proposed framework does not rely on strong differentiation between tar seal and grass/dirt banks like Yadav et al. and can accurately extract surface instances regardless of surrounding details of the road surface environment (see Fig. 9). Road surfaces of different shapes along their trajectory such as curvature and taper are accounted for due to our proposed processing pipeline.

An issue with using IoU metric in the application of width estimation is that it is biased towards high mean IoU relevant to the instance segmentation goal. This is because in this application the accuracy of the predicted mask's endpoints is more important for estimating the seal measurements than the IoU of the predicted mask.

5.4 Road surface markings

The road marking results show that needle-like objects can be segmented using the Mask R-CNN algorithm. In particular, this is demonstrated by the derivable line class where the IOU is particularly high and the width of each instance is only 1–2 pixels wide. However, without a large enough population to train on, the segmentation results for the hatched areas are worse compared to the driving lanes which occur at least twice per image and the driving lines which occur at least thrice per image. Classifying the marking objects have mixed results with the hatched areas having more unclassified results then correctly classed results. In comparison

with methods such as Jung et al. [12], our approach does not depend on factors such as road width consistency, road marking quality or weather. This is because the quality of results is dependent on the ground truth mask set and the section of the road in the dataset.

5.5 Road seal and geometric estimations

The seal result performs as well as the road surface result from localization and segmentation perspective. Thereby resulting in the benefits of not needing a flat level surface surrounding faced, not requiring the need to first extract planar ground surfaces before extracting the seal, and not relying on relative raised curb locality to a segment. It also suffers from the same limitations as the road surface method with IoU not being the best method for width estimation, which is more visible in seal segmentation as there is a thin gap between the edge of the seal segmentation result and the ground truth. The main limitation to the width estimation is that the accuracy of the width location in proportion to the scale ranges. In this case, the images are 1024x1024 pixels width post prediction which allows the width point to be around 5cm to its real width location. The 5cm resolution limitation would improve by increasing the image size during the prepossessing stage. However, this would increase computational power needed to create the associated model and more lidar points gathered in the data acquisition process.

6 Conclusion

In this paper, we demonstrate an improved approach for localizing, segmenting and classifying road objects such as surface, seal, and marking using 3D lidar point-clouds. We achieved this improvement from the existing Mask R-CNN method, by aggregating the relevant 3D lidar point-clouds information into a 2D image and further filtered this image down using PCA and affine transformations. The Mask R-CNN algorithm was then applied to obtain the classes location and segments of the road objects. We also showed that combining pre-processing with Mask R-CNN can be used to segment needle-like objects. Future work would include both improving the results and exploring additional methods to extract geometric parameters. The results could be improved via increasing the dataset size, position of lidars on the data collection apparatus, and increasing the available computational resources. Future work could also focus on geometric parameter estimation for road volume and slope estimation, as well as comparing existing surface extraction methods on our dataset.

Declarations

**Conflicts of interest** The authors have no other conflicts of interest to disclose.

References

1. Chow, M., Chen, T.: Benefits and costs of different road expenditure activities. Science **11**, 50 (2017)
2. NZTA: Traffic control devices manual (2008). https://www.nzta.govt.nz/resources/traffic-control-devices-manual/
3. Yuan, C., Chen, H., Liu, J., Zhu, D., Xu, Y.: Robust lane detection for complicated road environment based on normal map. IEEE Access **4**, 2169–3536 (2018). https://doi.org/10.1109/ACCESS.2018.2868976
4. Lei, G., Sun, J., Xiao, Z., Zhang, F., Wu, J.: Combining cnn and mrf for road detection. Comput. Electr. Eng. **70**, 895–903 (2017). https://doi.org/10.1016/j.compeleceng.2017.11.026
5. Kheyrollahi, A., Breckon, T.: Automatic real-time road marking recognition using a feature driven approach. Mach. Vis. Appl. **23**, 123–133 (2012). https://doi.org/10.1007/s00138-010-0289-5
6. Danescu, R., Nedevschi, S.: Detection and classification of painted road objects for intersection assistance applications. In: Conference Record-IEEE Conference on Intelligent Transportation Systems, pp. 433–438 (2010). https://doi.org/10.1109/ITSC.2010.5625261
7. Guan, H., Li, J., Yu, Y., Wang, C., Chapman, M., Yang, B.: Using mobile laser scanning data for automated extraction of road markings. ISPRS J. Photogramm. Remote Sens. **87**, 93–107 (2014). https://doi.org/10.1016/j.isprsjprs.2013.11.005
8. Soilán, M., Riveiro, B., Martínez-Sánchez, J., Arias, P.: Segmentation and classification of road markings using mls data. ISPRS J. Photogramm. Remote Sens. **123**, 94–103 (2017). https://doi.org/10.1016/j.isprsjprs.2016.11.011
9. Husain, A., Vaishya, R.C.: Detection and thinning of street trees for calculation of morphological parameters using mobile laser scanner data. Remote Sens. Appl. Soc. Environ. **13**, 375–388 (2018). https://doi.org/10.1016/j.rsase.2018.12.007
10. Gargoum, S., Karsten, L., El-Basyouny, K., Koch, J.: Automated assessment of vertical clearance on highways scanned using mobile lidar technology. Autom. Constr. **95**, 260–274 (2018). https://doi.org/10.1016/j.autcon.2018.08.015
11. Holgado-Barco, A., Gonzalez-Aguilera, D., Arias, P., Martínez-Sánchez, J.: An automated approach to vertical road characterisation using mobile lidar systems: Longitudinal profiles and cross-sections. ISPRS J. Photogramm. Remote Sens. **96**, 28–37 (2014). https://doi.org/10.1016/j.isprsjprs.2014.06.017
12. Jung, J., Bae, S.-H.: Real-time road lane detection in urban areas using LiDAR data. Electronics **7**(11), 276 (2018). https://doi.org/10.3390/electronics7110276
13. Wang, H., Luo, H., Wen, C., Cheng, J., Li, P., Chen, Y., Wang, C., Li, J.: Road boundaries detection based on local normal saliency from mobile laser scanning data. IEEE Geosci. Remote Sens. Lett. **12**, 1–5 (2015). https://doi.org/10.1109/LGRS.2015.2449074
14. Kumar, P., McElhinney, C.P., Lewis, P., McCarthy, T.: Automated road markings extraction from mobile laser scanning data, Elsevier. Int. J. Appl. Earth Obs. Geoinf. **32**, 125–137 (2014)
15. Jung, J., Che, E., Olsen, M., Parrish, C.: Efficient and robust lane marking extraction from mobile lidar point clouds. ISPRS J. Photogramm. Remote Sens. **147**, 1–18 (2019). https://doi.org/10.1016/j.isprsjprs.2018.11.012
16. Yang, M., Wan, Y., Liu, X., Xu, J., Wei, Z., Chen, M., Sheng, P.: Laser data based automatic recognition and maintenance of road markings from mls system. Elsevier J. Opt. Laser Technol. **107**, 192–203 (2018). https://doi.org/10.1016/j.optlastec.2018.05.027
17. Kim, J., Kim, J., Jang, G.-J., Lee, M.: Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection. Elsevier Int. Neural Netw. Soc. J. Neural Netw. **87**, 2555 (2016). https://doi.org/10.1016/j.neunet.2016.12.002
18. Wang, J., Song, J., Chen, M., Yang, Z.: Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine. Int. J. Remote Sens. **36**, 3144–3169 (2015). https://doi.org/10.1080/01431161.2015.1054049
19. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for image segmentation. Science (1998). https://doi.org/10.17863/cam.17966
20. Wu, C., Cheng, H.-P., Li, S., Helen, H.L., Chen, Y.: Apesnet: a pixel-wise efficient segmentation network for embedded devices. IET Cyber Phys. Syst. Theory Appl. **1**, 78–85 (2016). https://doi.org/10.1049/iet-cps.2016.0027
21. González-Jorge, H., Arias, P., Puente, I., Martínez, J.: Surveying of road slopes using mobile LiDAR. In: International Association for Automation and Robotics in Construction (IAARC), (2012). https://doi.org/10.22260/isarc2012/0015
22. Yadav, M., Singh, A.K., Lohani, B.: Computation of road geometry parameters using mobile LiDAR system. Remote Sens. Appl. Soc. Environ. **10**, 18–23 (2018). https://doi.org/10.1016/j.rsase.2018.02.003
23. Microsoft: satellite images segmentation and sustainable farming (2018). http://www.microsoft.com/developerblog/2018/07/05/satellite-images-segmentation-sustainable-farming/
24. 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. **147**, 178–192 (2019). https://doi.org/10.1016/j.isprsjprs.2018.10.007
25. Dutta, A., Zisserman, A.: The VIA annotation software for images, audio and video, arXiv preprint arXiv:1904.10699
26. NZTA: Manual of traffic signs and markings (motsam) part 1: traffic signs (2010). https://www.nzta.govt.nz/resources/motsam/part-1/motsam-1.html
27. NZTA: Manual of traffic signs and markings (motsam)-part 2: markings (2010). https://www.nzta.govt.nz/resources/motsam/part-2/motsam-2.html

28. 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. MDPI J. Remote Sens. **10**(10), 9–17 (2018). https://doi.org/10.3390/rs10101531

29. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnnCite arxiv:1703.06870 Comment: open source; appendix on more results

30. Girshick, R.: Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV), IEEE, (2015). https://doi.org/10.1109/iccv.2015.169

31. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C.: A survey on deep transfer learning (2018). arXiv:1808.01974

32. Abdulla, W.: Mask R-CNN for object detection and instance segmentation on keras and tensorflow, https://github.com/matterport/Mask_RCNN (2017)

33. Abass, H.H., Al-Salbi, F.M.M.: Rotation and scaling image using PCA. Comput. Inf. Sci. (2016). https://doi.org/10.5539/cis.v5n1p97
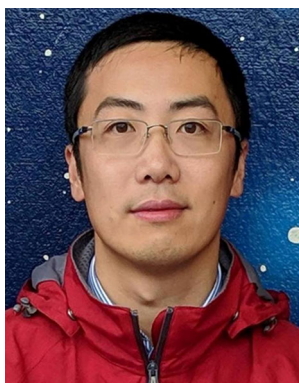
Dr. Heyang Thomas Li is a Lecturer in Statistics and Data Science at University of Canterbury (UC), School of Mathematics and Statistics. His current research areas involve image processing and classification, 3D projective geometry, phylogenetic trees, classification trees, and deep learning applications. He currently leads the UC Spatial And Image Learning (SAIL) group, working on research in collaboration with NZ Transport Agency, Christchurch City Council, Christchurch Airport and Stats NZ. UC SAIL group is kindly supported by the School, UC R&I and KiwiNet.

Zachary Todd received a Master of Applied Data Science degree from the University of Canterbury, New Zealand, in 2018. Currently, he is a PhD Candidate in Computational and Applied Mathematics. He is also working as a research assistant in the Spatial And Image Learning group, University of Canterbury. His research interests include Machine Learning, Digital Image Processing, and their applications to road asset management.

Nikolas Bielski received a Master degree in Applied Data Science from the University of Canterbury, New Zealand in 2018, before joining the Spatial And Image Learning group, Department of Mathematics and Statistics, University of Canterbury as a Research Assistant until 2019. Since then, he has been working as an Analyst at the New Zealand Defence Force and is starting his PhD in Computational and Applied Mathematics. His research interests include Graph representation learning with a focus on attention for coordination and exploration in dynamic multi-agent systems.

Felix Carroll received the Master of Applied Data Science degree with distinction from the University of Canterbury, New Zealand, in 2018, before joining the University's Spatial and Image learning group. Since then, he has joined Statistics New Zealand as a Statistical Analyst in 2019 and is currently a Data analyst for Te Hiringa Hauora/Health Promotion Agency. He is also working as a hired researcher at the Spatial and Image learning group, University of Canterbury. His research interest includes machine learning and digital image processing.