applied sciences

MDPI

Article

Improving Road Segmentation by Combining Satellite Images and LiDAR Data with a Feature-Wise Fusion Strategy

Ozan Ozturk [1], Mustafa Serkan Isik [1], Martin Kada [2] and Dursun Zafer Seker [1,*]

[1] Department of Geomatics Engineering, Civil Engineering Faculty, Istanbul Technical University, Istanbul 34469, Türkiye; oozturk16@itu.edu.tr (O.O.)
[2] Methods of Geoinformation Science, Institute of Geodesy and Geoinformation Science, Technische Universität Berlin, 10553 Berlin, Germany
* Correspondence: seker@itu.edu.tr; Tel.: +90-212-285-3414

Abstract: Numerous deep learning techniques have been explored in pursuit of achieving precise road segmentation; nonetheless, this task continues to present a significant challenge. Exposing shadows and the obstruction of objects are the most important difficulties associated with road segmentation using optical image data alone. By incorporating additional data sources, such as LiDAR data, the accuracy of road segmentation can be improved in areas where optical images are insufficient to segment roads properly. The missing information in spectral data due to the object blockage and shadow effect can be compensated by the integration of 2D and 3D information. This study proposes a feature-wise fusion strategy of optical images and point clouds to enhance the road segmentation performance of a deep learning model. For this purpose, high-resolution satellite images and airborne LiDAR point cloud collected over Florida, USA, were used. Eigenvalue-based and geometric 3D property-based features were calculated based on the LiDAR data. These optical images and LiDAR-based features were used together to train, end-to-end, a deep residual U-Net architecture. In this strategy, the high-level features generated from optical images were concatenated with the LiDAR-based features before the final convolution layer. The consistency of the proposed strategy was evaluated using ResNet backbones with a different number of layers. According to the obtained results, the proposed fusion strategy improved the prediction capacity of the U-Net models with different ResNet backbones. Regardless of the backbone, all models showed enhancement in prediction statistics by 1% to 5%. The combination of optical images and LiDAR point cloud in the deep learning model has increased the prediction performance and provided the integrity of road geometry in woodland and shadowed areas.

Keywords: road segmentation; feature-wise fusion; LiDAR features; satellite images; U-Net; ResNet

check for updates

Citation: Ozturk, O.; Isik, M.S.; Kada, M.; Seker, D.Z. Improving Road Segmentation by Combining Satellite Images and LiDAR Data with a Feature-Wise Fusion Strategy. *Appl. Sci.* **2023**, *13*, 6161. https://doi.org/10.3390/app13106161

Academic Editor: Juan A. Gómez-Pulido

Received: 4 April 2023
Revised: 4 May 2023
Accepted: 16 May 2023
Published: 17 May 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Road segmentation is the process of predicting and classifying road pixels in an image to aid the generation of accurate road network information. Road segmentation using satellite images is a critical tool for efficient traffic management and planning. Proactive measures can be taken to identify potentially dangerous intersections and congestion areas, thereby reducing the risk of traffic-related accidents and improving the overall safety of a road network. Hence, authorities can use road segmentation to monitor changes over time and detect any obstacles or hazardous conditions that could make a route less safe, allowing for constructive action to be taken.

Recently, advances in the field of artificial intelligence have resulted in various state-of-the-art model architectures and their optimizations to be proposed to solve complex segmentation problems [1–3]. Machine learning and deep learning approaches are being applied to improve the quality of road segmentation from remote sensing data [4–7]. Naturally, a proper solution of road segmentation using remote sensing images is a highly

non-linear problem; hence, deep learning-based semantic segmentation solutions are more favorable in this matter [8–11]. Deep learning model development is seen as an epochal development. However, it is strongly data driven, which sometimes makes it difficult to make accurate predictions. The performance of deep learning models is greatly influenced by both the quality and quantity of the data utilized in their training process [12,13]. If the data are not representative of the real-world scenario, this leads to poor performance. A major challenge is the shadow effect, which makes it difficult to segment the road pixels [14]. The resolution and spectral capability are highly interrelated, and they result in false perceptions at the boundary points of the road [15]. The resolution of the satellite image should give high spatial information for extraction. However, accessing high-resolution images can make it costly or impossible to perform qualified road segmentation. There are also a lot of problems with woodlands because of the lack of information about roads. The missing information can be completed using the LiDAR point cloud. Therefore, it may be possible to overcome the limitations of satellite images in areas where the road information is problematic and improve the accuracy of road segmentation by exploiting the additional geometric relations, integrating LiDAR data into deep learning models [16,17].

LiDAR is seen as a valuable data source and it provides useful information that cannot be extracted from optical images. The LiDAR scanner can be mounted on terrestrial, mobile, and airborne platforms. The mobile and terrestrial LiDAR data can be used for the semantic segmentation of road information [18,19], but the applications of these platforms are limited due to their small coverage areas and the difficulties of point cloud measurements increases in challenging areas [20]. To be able to conduct road segmentation in large areas, or even country-wide, the airborne LiDAR platforms become more feasible in this regard, as these platforms can cover cities and countries rapidly. There is substantial research on road segmentation using airborne LiDAR data. In Li et al. [21], using a grid index structure to detect roads in LiDAR point clouds, a morphological gradient was applied on ground points which were filtered using local intensity distribution. Li et al. [22] conducted a study based on similar parameters to the method of Li et al. [21]. After identifying the differences in shape, reflectance, and road width, road centrelines were extracted with local principle component analysis. Then, road networks were extracted using the global primitive grouping method. Hui et al. [23] introduced a novel approach consisting of skewness balancing, rotating neighbourhoods, and hierarchical fusion and optimization to extract roads from point cloud data. Tejenaki et al. [24] implemented a hierarchical method that refined intensity with mean shift segmentation and extraction of road centerlines with a Voronoi diagram. The proposed method improved both road extraction results and water surface detection. Sánchez et al. [16] aimed to distinguish between road and ground points based on intensity constraints. In their proposed method, an improved skewness balancing algorithm was used for the calculation of the intensity threshold. There are deep learning techniques, such as PointNet [25] and PointConvNet [26], that take point clouds as direct input for object classification and segmentation problems. Following the gradual development of these approaches, PointNet++ was published for labeling the road points in the point cloud data [27]. There are, however, still issues to be addressed when relying solely on LiDAR data. For instance, there is no spectral information in LiDAR data which can lead to mixing between objects with similar density and texture such as parking lots. Therefore, the combination of satellite images and LiDAR data can resolve such deficiencies, leading to more accurate segmentation of the road network [22,28,29].

There have been numerous studies conducted on the potential benefits of remote sensing data in segmenting images. However, the combination and integration of different data sources are yet to achieve the same level of attention. Audebert et al. [30] conducted thorough research on integrating multiple sources and models and proposed an approach that combined LiDAR and multispectral images using different fusion strategies. They used an Infrared-Red-Green (IRRG) image and a combination of the Normalized Digital Surface Model (NDSM), Digital Surface Model (DSM), and Normalized Difference Vegetation Index (NDVI) obtained from LiDAR data. In a similar study, Zhang et al. [31] suggested

using high-resolution images and nDSM derived from LiDAR. Their proposed method involves segmenting fused data, classifying image objects, generating a road network, and extracting the road centerline network using a multistage approach, including morphology thinning, Harris corner detection, and least square fitting. Zhou et al. [32] introduced a novel method named the FuNet (Fusion Network) that integrates satellite images with binary road images generated from GPS data collection. This approach, which depends on a multi-scale feature fusion strategy, has been found to be more effective at resolving the road connectivity issue than using solely satellite imagery. Torun and Yuksel [33], on the other hand, proposed a technique that involves combining hyperspectral images with LiDAR data for unsupervised segmentation. They applied a Gaussian filter to the point cloud while performing principle component analysis on the images, which yielded results used to create an affinity matrix. In another study, Gao et al. [34] presented multi-scale feature extraction for 3D road segmentation, combining characteristic features from high-resolution images and LiDAR data.

Although there are findings suggesting the benefits of integrating diverse data sources for multiple applications, existing research indicates that there are still challenges that need to be addressed in this area. The combination of different types of remotely sensed data can be challenging, yet resourceful for segmentation studies considering the complexity of the solution. Specifically, the combination of 2D images and a 3D point cloud that has an irregular data structure requires complex solutions for deep learning architectures. An alternative and easier approach is to extract context information from the point cloud that represents the local 3D shapes and fuse them with the high-level features extracted from 2D optical images [20]. Therefore, in this study, a feature-wise fusion strategy of 2D optical images and point cloud data was proposed to enhance the road segmentation capability of deep learning models in areas where the use of optical images is inadequate due to object blockage and shadow problems. The high-resolution optical satellite images obtained from the Google Maps platform were combined with the contextual feature images derived from the point cloud obtained using airborne LiDAR data. The combination of these two types of data was carried out feature-wise in a deep residual U-Net-based deep learning model. The features generated by different ResNet backbones using only optical satellite images were fused with the geometric features calculated from the LiDAR point cloud before the final convolution layer of the model. This study provides insight into how to combine optical images and point-cloud data to improve the segmentation of roads. The proposed strategy can be implemented in any deep learning model architecture with proper modifications. The motivation of this study is to improve the road segmentation capability of the deep learning models by combining 2D and 3D information properly.

The major objectives of the study were as follows: (1) the use of irregular point cloud data together with the 2D optical images in an end-to-end deep learning model was introduced as a feature-wise fusion strategy; (2) the improvement brought about by the LiDAR data was outlined together with the statistical results of the models and the prediction performance of the proposed fusion strategy was evaluated in areas where the road segmentation is challenging; (3) the consistency of the strategy was evaluated with different ResNet backbones in deep residual U-Net architecture. Finally, the relative importance of optical and LiDAR-derived features was calculated using the best-performing combined model to outline which features contribute to the improvement of road segmentation the most. The intent of this study is not to propose a deep learning architecture, but to provide a combination strategy for optical images and point-cloud data to improve the segmentation of roads. However, the proposed fusion strategy can be implemented in any model architecture with proper modifications. The motivation behind this study was to improve the road segmentation capability of any deep learning model, which has been proven to be successful in this regard, by combining 2D and 3D information properly.

2. Data and Methods

2.1. Study Area and Data Collection

In this study, open-source airborne LiDAR data collected by the U.S. Geological Survey were used together with high-resolution optical satellite images from the Google Maps platform. While the Google Map API service is available to gather satellite images everywhere on Earth, airborne LiDAR is a valuable (and expensive) data source. The U.S. Geological Survey initiated the National Geospatial Program with airborne LiDAR campaigns that cover the United States to improve and deliver topographic information. This program contains different quality completed products from quality Level-0 to Level-3. In order to test the performance of the feature-wise fusion of optical images and LiDAR, only level-1 quality data that are located over major cities are taken into account. In this context, the Florida southeast project that falls within the Florida counties of Broward, Collier, Hendry, Miami-Dade, Monroe, and Palm Beach was chosen as the study area [35].

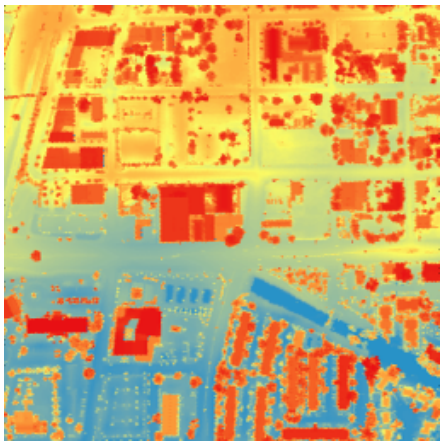
This LiDAR data were collected on June 2018 and published by U.S. Geological Survey with the name of Florida Southeast LiDAR-Block 1. The point cloud was generated at level 1 quality with a source DEM of 0.5 m/pixel. The vertical accuracy of the point cloud is $\pm$10 cm in terms of root mean square error. Nominal pulse spacing is smaller than 0.35 m. The point density is 14.87 points/m^2. It has seven classes, namely ground, low noise, water, bridge decks, high noise, ignored ground, and unclassified. The dataset consists of 526 tiles, each of them covering a 1 km $\times$ 1 km area. Concerning data size, only 393 tiles, which cover the majority of the road network in the area, were used in the study.

After the LiDAR data were obtained, the required satellite images were generated using the Google Map Static API-based tool. This tool was created by [36] and it generates satellite and corresponding mask images randomly or in sequence in the defined region based on latitude and longitude. It also produces a metadata file with which to rectify these satellite images, if required. In this way, the registration of the LiDAR data and the satellite images could be performed. The Google Maps Static API provides images at various levels which correspond to different scales and resolutions on the Earth's surface. To generate images, zoom level 17 was chosen because of its ideal coverage area and pixel resolution. The satellite and mask images were extracted with a dimension of 512 $\times$ 512 pixels. This resulted in an image with a spatial resolution of 1.07 m $\times$ 0.96 m. Consequently, a total of 1426 images were generated, which cover the LiDAR dataset and contain roads.

The optical satellite images and their corresponding masks are geo-referenced in the WGS84 datum. This enables the inter-usability of these images together with other geospatial datasets; in this case, the LiDAR point cloud. As shown in a sample overlay presented in Figure 1, the horizontal coordinates of these datasets match with each other harmoniously for use in road segmentation studies.

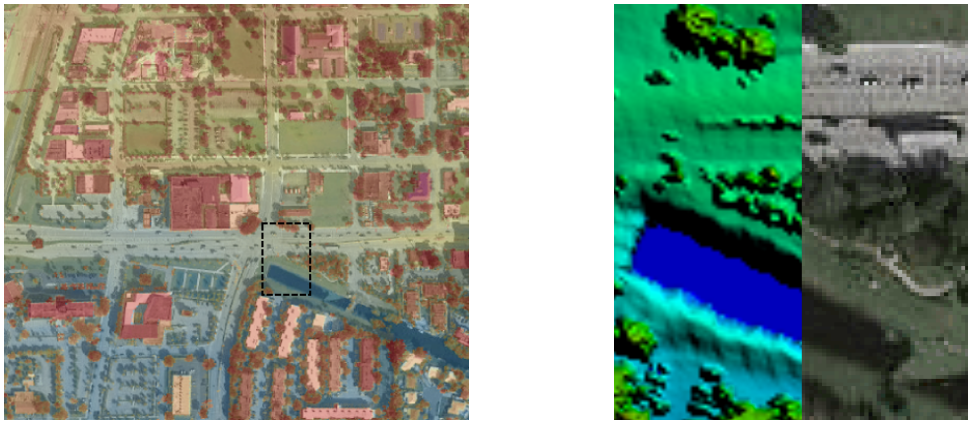


(**a**)



(**b**)

Figure 1. *Cont.*



(**c**) (**d**)

Figure 1. Satellite image (**a**), rendering of point cloud (**b**) and their overlapping image composite (**c**), and comparison of satellite image and LiDAR data side by side for the area of interest enclosed by the black box (**d**).

2.2. LiDAR Feature Extraction

LiDAR is an active sensing system that operates by measuring the round-trip timing of a laser beam from an object in order to determine the distance from the sensor to the target. By analyzing the laser time range in combination with the scan angle and spatial coordinates of the laser scanner, the spatial coordinates of an object as (X, Y, Z) are obtained. Along with spatial data, intensity values, number of returns, point classification values, and GPS times are recorded. However, the complete geometric properties of the targetted objects in the point cloud cannot be solely represented by their spatial information (X, Y, and Z), though they can be characterized by their geometric features. In order to avoid the heavy computational load of using complex point cloud data, semantic information, extracted by generating geometric features, can be used instead.

As presented in Figure 2, feature extraction was carried out in four steps. In the first step, points whose noise levels are high were eliminated from the LiDAR point cloud. After noise removal, outlier detection was performed by removing the points whose average distance from their neighbouring points exceeded a given threshold value. In the next step, the statistical relationship of the points was formed by creating a new data structure based on neighbouring points, similar to the second step. But this time, the neighbouring points were determined using a clean point cloud by applying a k-dimensional tree (*k*-d tree) algorithm in 3D. In the final step, eigenvalues and the 3D geometric features were calculated.

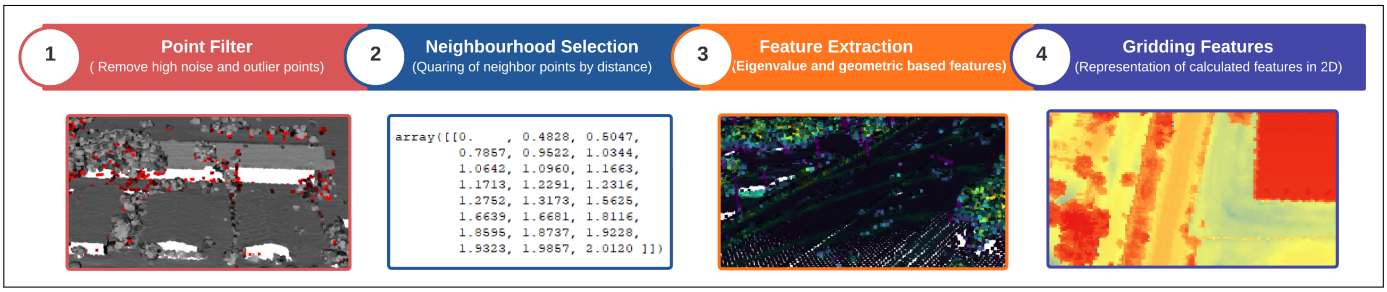


Figure 2. Feature extraction pipeline.

There are various methods available for determining the neighborhood of points in LiDAR data. The neighborhood of points can be constructed using a spherical radius or parameterized based on the number of closest neighbours in 2D or 3D space (i.e., k-NN methods). In this study, k-NN based neighbourhood selection was constructed via the well-known *k*-d tree algorithm, in which the data are partitioned using a binary tree

structure [37]. As a result of the neighbourhood selection, each query point ($P(i)$) and its neighbors were indexed. The distance between the query point and its k neighbours was determined and indexed in order from nearest to farthest.

In the first part of the feature extraction, the features that represent the geometric 3D properties of the neighbourhood were calculated using the height of each query point, i.e., the absolute height H, the height difference between the query point and its neighbouring points $\Delta H_{\text{kNN,3D}}$, the standard deviation of absolute height within neighbouring points $\sigma H_{\text{kNN,3D}}$, local 3D neighbourhood represented by the radius of k nearest neighbours $r_{\text{kNN,3D}}$), and local point density ρ_{3D}.

$$H = |Z| \tag{1}$$

$$\Delta H_{\text{kNN,3D}} \tag{2}$$

$$\sigma H_{\text{kNN,3D}} \tag{3}$$

$$r_{\text{kNN,3D}} = max||D_{\text{kNN,3D}}|| \tag{4}$$

$$\rho_{\text{3D}} = \frac{k+1}{\frac{4}{3}\pi r^3_{\text{kNN,3D}}} \tag{5}$$

where $D_{\text{kNN,3D}}$ is the distance between the query point and its neighbours.

In the other part of the feature extraction, eigenvalue-based features were generated after the neighbourhood was determined. In order to compute these features for a point in the 3D point cloud, the covariance matrix (denoted as **Cov**) needed to be calculated.

$$\mathbf{Cov} = \frac{1}{k+1}\sum_{i=0}^{k}\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)^T \tag{6}$$

In this equation, $\mathbf{X}_i$ is the ith neighbour and $\overline{\mathbf{X}}$ is the geometric centre of the neighbourhood determined by,

$$\overline{\mathbf{X}} = \frac{1}{k+1}\sum_{i=0}^{k}\mathbf{X}_i \tag{7}$$

After the covariance matrix was computed, eigenvalues (w) and eigenvectors (v) were determined. The direction along which the dataset had the maximum variation is indicated by the eigenvector with the largest eigenvalue. These were used for eigenvalues-based feature extraction. First, the eigenvalues, and correspondingly the eigenvectors, were sorted in ascending order as $\lambda_1 > \lambda_2 > \lambda_3$. In order to perform accurate feature extraction, it had to be ensured that the vectors were normalized between 0 and 1 and that the eigenvalues were greater than 0 [38].

The calculated eigenvalue based features were linearity L_λ, planarity P_λ, sphericity S_λ, omnivariance O_λ, anisotropy A_λ, eigenentropy E_λ, sum of eigenvalues $\sum_\lambda$ and change of curvature C_λ.

$$L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1} \tag{8}$$

$$P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1} \tag{9}$$

$$S_\lambda = \frac{\lambda_3}{\lambda_1} \tag{10}$$

$$O_\lambda = \sqrt[3]{\lambda_1\lambda_2\lambda_3} \tag{11}$$

$$A_\lambda = \frac{\lambda_1 - \lambda_3}{\lambda_1} \tag{12}$$

$$E_\lambda = -\sum_{i=1}^{3} \lambda_i \ln(\lambda_i) \tag{13}$$

$$\sum_{\lambda} = \lambda_1 + \lambda_2 + \lambda_3 \tag{14}$$

$$C_\lambda = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \tag{15}$$

The 3D features calculated for each point $P(i)$ were reduced to the gridded horizontal plane in order to be combined with the features calculated from the satellite images. Each tile covered by LiDAR point clouds was divided into 1 m $\times$ 1 m cells in (x, y). Due to one or more $P(i)$ points falling within the same cell, the feature value of each cell is represented by the average value of these points. All gridded LiDAR features can be used directly, but the absolute height requires post-processing to be performed. The elevations derived from LiDAR data represent geographic elevations as absolute heights, in contrast with other features that are characterized by statistical relationships. To determine the distance between objects and the ground in this study, the digital elevation model was subtracted from the height values of points in the point cloud. Finally, the LiDAR feature extraction was completed by clipping the features based on satellite images.

Figure 3 illustrates the rendering of the generated features in blue and red colour for low- and high-impact feature values. In general, these features provide insight into the geometry of objects in the point cloud and their relation with the surrounding objects. For instance, anisotropy represents the uniformity of a point cloud, while linearity can be defined as a measure of linear attributes. Furthermore, absolute height provides a distinction between the roads and other objects above [39]. Together, these features complete the geometric relation between the road and its neighbouring object which is not properly handled in the satellite-only segmentation solutions.

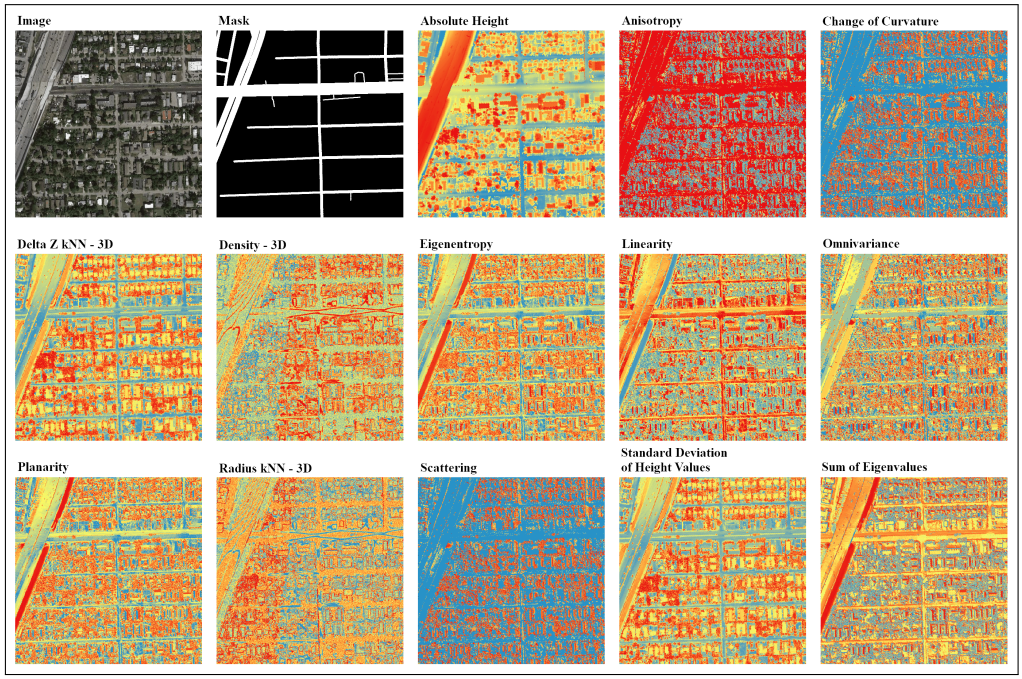


Figure 3. Illustration of gridded feature samples from LiDAR data.

2.3. U-Net Model Structure

The U-Net was initially proposed for segmenting biomedical images [40]. It was formed as the encoder (contracting), bridge, and decoder (expansive) blocks. As its name implies, it is a U-shaped architecture in which the high-resolution features are extracted by down-sampling the input image, followed by up-sampling them to recover the original spatial resolution. These features are concatenated with the up-sampled output to achieve precise localization. High-dimensional feature spaces, created by up-sampling, allow the architecture to feed the context information into higher-resolution layers. In order to achieve a precise output, high-resolution features from the encoder path are concatenated with the up-sampled output. Consequently, pixel-wise predictions can be made.

As with many deep learning models, U-Net is susceptible to vanishing gradient issues. Although deepening the networks is intended to extract the complex information that cannot be obtained in the background, it also tends to reveal the problem of vanishing gradients. The reason for this is that in the back-propagation stage, the gradient of the loss values updates very little to the previous layers, or sometimes does not update at all. Additionally, as the depth of the network increases, it becomes more difficult to optimize and reaches a kind of accuracy saturation, which leads to higher training error rates [41]. Accordingly, He et al. [1] introduced residual learning, called ResNet, which can extract underlying features by increasing the depth. Basically, it is a deep-learning algorithm used to classify images. By integrating shortcut connections into the plain network and copying identity mapping from a shallower model, residual learning will prevent the training error of a deeper model from increasing. The ResNet architecture includes residual connections, which enable the gradients to be carried forward and work out the vanishing gradient problem by going deeper. A ResNet model is identified based on the number of layers it contains. There are five different ResNet models published with 18, 34, 50, 101, and 152 layers. There is an increase in model depth from ResNet 18 to ResNet 152. In order to resolve a highly non-linear problem, important features can be extracted by deepening the model architecture. However, deeper models require high computation costs and hyperparameter tuning is more challenging compared to shallow models. By exploiting ResNet's ability to provide identity mapping between layers, these training problems can be overcome by extracting high-resolution complex features.

Zhang et al. [8] introduced a novel method for segmenting roads by combining the strengths of U-Net and deep residual learning. This cooperation solves the vanishing gradient problem and performs a powerful segmentation. In the case of ResNet, the network can be trained more easily, while in the case of U-Net, the intricacies of the algorithm will be greatly reduced. As part of this approach, called deep residual U-Net architecture, residual blocks are incorporated into the encoder stages. Thus encoders of ResNet are responsible for extracting difficult-to-obtain features from images, while decoders of U-Net are responsible for generating segmentation masks with precise localization.

2.4. Feature-Wise Fusion Strategy

In this study, the deep residual U-Net, described as an end-to-end network with a ResNet backbone, was used. To implement feature-wise fusion for segmentation, the training was carried out in a sequential model where the satellite-based features and LiDAR-based features were both integrated. The model consists of two sequential parts connected to each other at the end of U-Net architecture. This was achieved by forwarding optical images through the encoder and decoder segments of the deep residual U-Net architecture, which makes up the first sequential part of the model, to extract high-level features from satellite images. In the second part, 13 geometric features derived from the LiDAR point cloud are fed to the sequential model without any computation to extract new features from them, as they are already linear features. Before the final convolution block of the model, these geometric features ($512 \times 512 \times 13$) were concatenated with the high-level features ($512 \times 512 \times 16$) along the channel dimension to form a multi-layer feature map with size $512 \times 512 \times 29$. This combined feature map passed to the final

convolution layer and an output layer with *Sigmoid* as an activation function, to predict road pixels (see Figure 4).

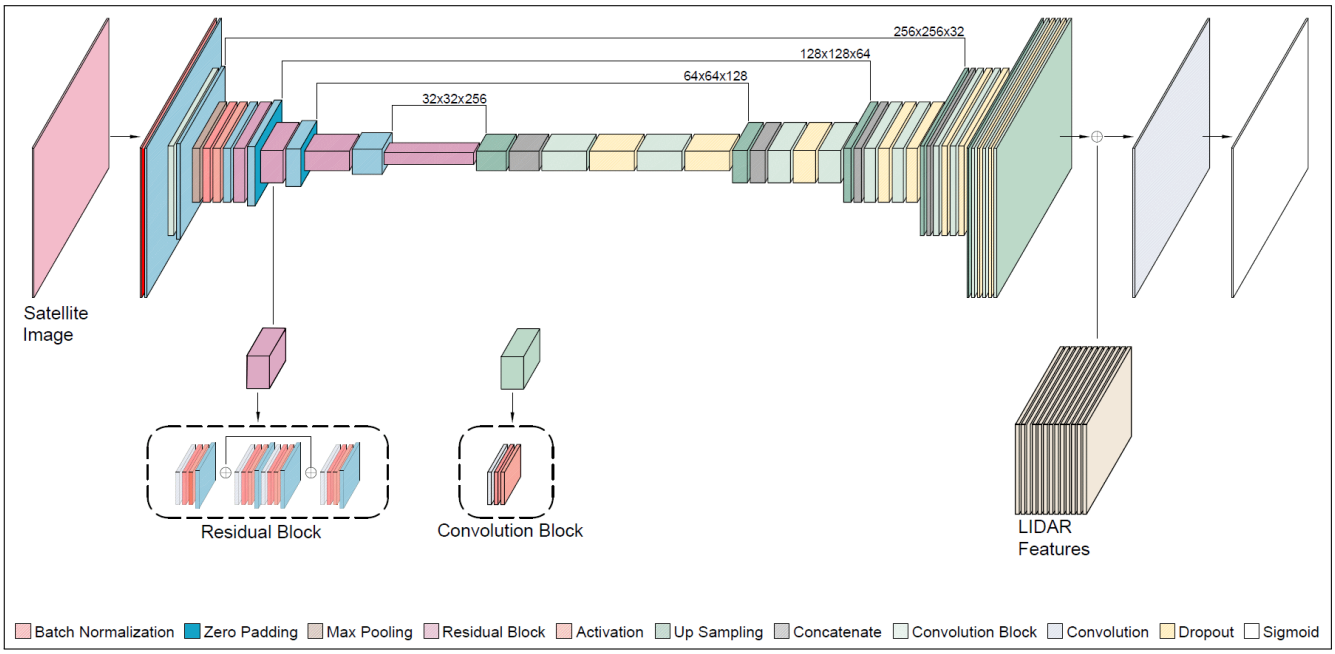


Figure 4. Proposed model structure.

2.5. Model Setup

The deep residual U-Net architecture consists of five down-sample blocks, each with decreasing filter sizes of 256, 128, 64, 32, and 16. The convolution blocks consist of a sequence of convolution, batch normalization, activation, and zero padding layers repeated two times for each block. The convolution layers use a 3×3 kernel size and the ReLU activation function. There is a 4×4 transpose convolution with stride 2×2 in the upsampling stage, which is consistent with the size of the filter. In this stage, the dropout layer was added to each convolution block flow.

The model was trained with 80% of the dataset, randomly split from the entire optical images, LiDAR, and corresponding mask images. The training data were further divided into 75% training and 25% validation data. The optuna framework [42] was used to optimize the hyperparameters, such as the optimizer, learning rate, and dropout rate. Adam was selected as an optimizer and Binary cross entropy was used as a loss function in the model. The learning rate was tested between 10^{-1} and 10^{-5}. The Dropout rate was tested from 10% to 50%. The training process was stopped after 10 epochs if no improvement was observed in the validation accuracy.

2.6. Evaluation Metrics

The trained models were evaluated based on five metrics, namely Precision, Recall, F1-Score, Intersection over Union (IoU), and Cohen's kappa score (κ). The recall is how much of the road that should be predicted is actually a road. F1-Score is the equilibrium level between Precision and Recall. IoU indicates the ratio of the overlapping area to the size of the total area of prediction and the ground truth. All metrics range from 0 to 1 and they can be calculated as

$$\begin{aligned}
\text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \\
\text{Recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \\
\text{F1-Score} &= \frac{2\text{TP}}{2\text{TP}+\text{FP}+\text{FN}} \\
\text{IoU} &= \frac{\text{TP}}{\text{TP}+\text{FP}+\text{FN}} \\
\text{Kappa} &= \frac{2(\text{TP}\times\text{TN}-\text{FN}\times\text{FP})}{(\text{TP}+\text{FP})(\text{FP}+\text{TN})+(\text{TP}+\text{FN})(\text{FN}+\text{TN})}
\end{aligned} \tag{16}$$

where TP, TN, FP, and FN are the true positive, true negative, false positive, and false negative, respectively.

3. Results & Discussion

In this section, the numerical results of the road segmentation carried out via feature-wise fusion of optical images and LiDAR are presented together with the visual comparison of enhancements.

3.1. Experimental Results

The contribution of LiDAR features was clarified by using two training scenarios. In the first scenario, the deep residual U-Net model was trained using satellite images only, while the second scenario includes the feature-wise fusion of optical and LiDAR datasets. In both scenarios, the U-Net architecture is integrated with ResNet-18, ResNet-34, ResNet-50, and ResNet-152 backbones for the feature extraction, leading to eight different model training setups. The results obtained from the analyses are given with the models' prediction metrics in Table 1. The model with the best results for each scenario is indicated in bold.

Table 1. The prediction statistics of all models.

Model No	Input Data	Backbone	Metrics				
			Precision	**Recall**	**F1-Score**	**IoU**	**Kappa**
1	Image	ResNet18	0.828	0.817	0.820	0.711	0.790
2	Image + LiDAR		**0.842**	**0.820**	**0.824**	**0.739**	**0.799**
3	Image	ResNet34	0.825	0.825	0.823	0.717	0.796
4	Image + LiDAR		**0.839**	**0.855**	**0.845**	**0.748**	**0.821**
5	Image	ResNet50	0.854	0.82	0.834	0.735	0.810
6	Image + LiDAR		**0.865**	**0.829**	**0.845**	**0.750**	**0.821**
7	Image	ResNet152	0.862	0.809	0.830	0.732	0.810
8	Image + LiDAR		**0.880**	**0.852**	**0.863**	**0.781**	**0.843**

The results clearly showed that the integration of optical satellite imagery with LiDAR features enhances the performance of road segmentation by 1% to 5% in all models. For each backbone configuration, the scenario with the LiDAR data performed better than satellite image-only training. On the other hand, the ResNet-152 backbone performed better for feature extraction in U-Net architectures, indicating that increasing the depth of the model increased the performance of segmentation as well.

3.2. Discussions

Regarding the improvements in areas where pixel complexity increases, the models trained with the fusion strategy showed better performance compared to image-only solutions, as highlighted with yellow boxes in Figure 5. The contribution of the fusion strategy can be readily observed by examining individual results and visualized road predictions. It can be seen that the fusion helped to complete the road geometry in these areas. The model is able to further segment the road pixels in the presence of tree cover more successfully than when using only satellite images in the deep learning model. The precision value for the individual predictions, where the image-only model exhibits worse performance, is increased up to 5% using the fusion strategy. This indicates that adding LiDAR features increased the number of true positives obtained. Using the ResNet-152, the highest F1-Score (93.7%) was achieved and the lowest (77.3%) when applying the ResNet-18. Road boundary continuity was achieved for the woodlands shown in Figure 5 and was able to increase up to 3% of recall values. To clarify the superiority of the proposed approach, predictions in wooded areas where the road pixels are blocked by trees are presented in Figure 6. Moreover, LiDAR-image fusion increased the performance of prediction in areas where there are shadow effects in the satellite images (see Figure 7). The recall and the F1-score values are significantly higher in these images. Additionally, the shadow factor can be seen to be too complex when relying only on satellite data. The proposed approach is effective at providing reinforcing information. Another problem related to road integrity at circular intersections has been resolved. This study demonstrates that the proposed strategy was perfectly executed in terms of both the quality and quantity of road extraction.

Compared to the recall and IoU values, improved GAN approaches, such as those introduced by Hu et al. [15] and Zhang et al. [43], have demonstrated similar recall and IoU results using only satellite images. Studies utilizing the Vaihingen and Potsdam datasets published by the International Society for Photogrammetry and Remote Sensing (ISPRS) are particularly prominent when studying multiple data sources. These datasets consist of Digital Aerial Images and Digital Surface Models, which classify various features such as surfaces and buildings. To benchmark their studies against previous studies, the researchers evaluated the accuracy of estimating impermeable surfaces formed by roads in these datasets. Their findings indicate that the use of fusion techniques led to an increase in the F1-Score statistics by 1% in the Vaihingen dataset and 2% in the Potsdam dataset when compared to the original U-Net model with ResNet 50 integration, as reported in previous studies by Sun et al. [29] and Audebert et al. [30]. The study utilized models 5-6, which yielded similar results. Furthermore, it was found that integrating GPS data into satellite images improved relative accuracy by 2% Zhou et al. [32], and analysis of the tested models supports this observation in this study. In addition, this study revealed that roads were more completely represented in individual image analysis, a significant finding previously reported in other studies, particularly in woodland and shadow areas. As noted by Zhang et al. [31], while this study has achieved road connectivity completion even on roads with curves, post-processing techniques can be employed to further enhance the quality of results on such roads.

The effectiveness of the fusion strategy approach can be demonstrated by calculating feature importance. In Figure 8, the relative importance of each feature is presented. The optical satellite images were found to be one of the most relevant features, as expected, with 23% importance. The model architecture seems to benefit the most from the change in absolute heights within neighbourhoods, i.e., height difference, and the linearity out of 13 LiDAR features. These two features, individually, have as much importance as the optical images, and together they represent 47% of the model's predictions. The rest of the feature importance is represented by the sum of eigenvalues, standard deviation of height values, planarity, entropy, anisotropy, and absolute height features, each of which has importance less than 10%. Scattering, radius of k nearest neighbours, omnivariance, and changes in curvature features did not contribute to the model's performance at all. Additionally, 1D and 2D geometric features computed with the eigenvalues, such as linearity and planarity,

were found to be effective, while the 3D geometric feature, scattering (or sphericity), was found insignificant. The change in absolute heights, such as height difference and standard deviation of absolute heights, were found to be dominant predictors compared to absolute heights which indicates the physical properties of road geometry are represented better by the statistical relations of physical height values, rather than the height values.

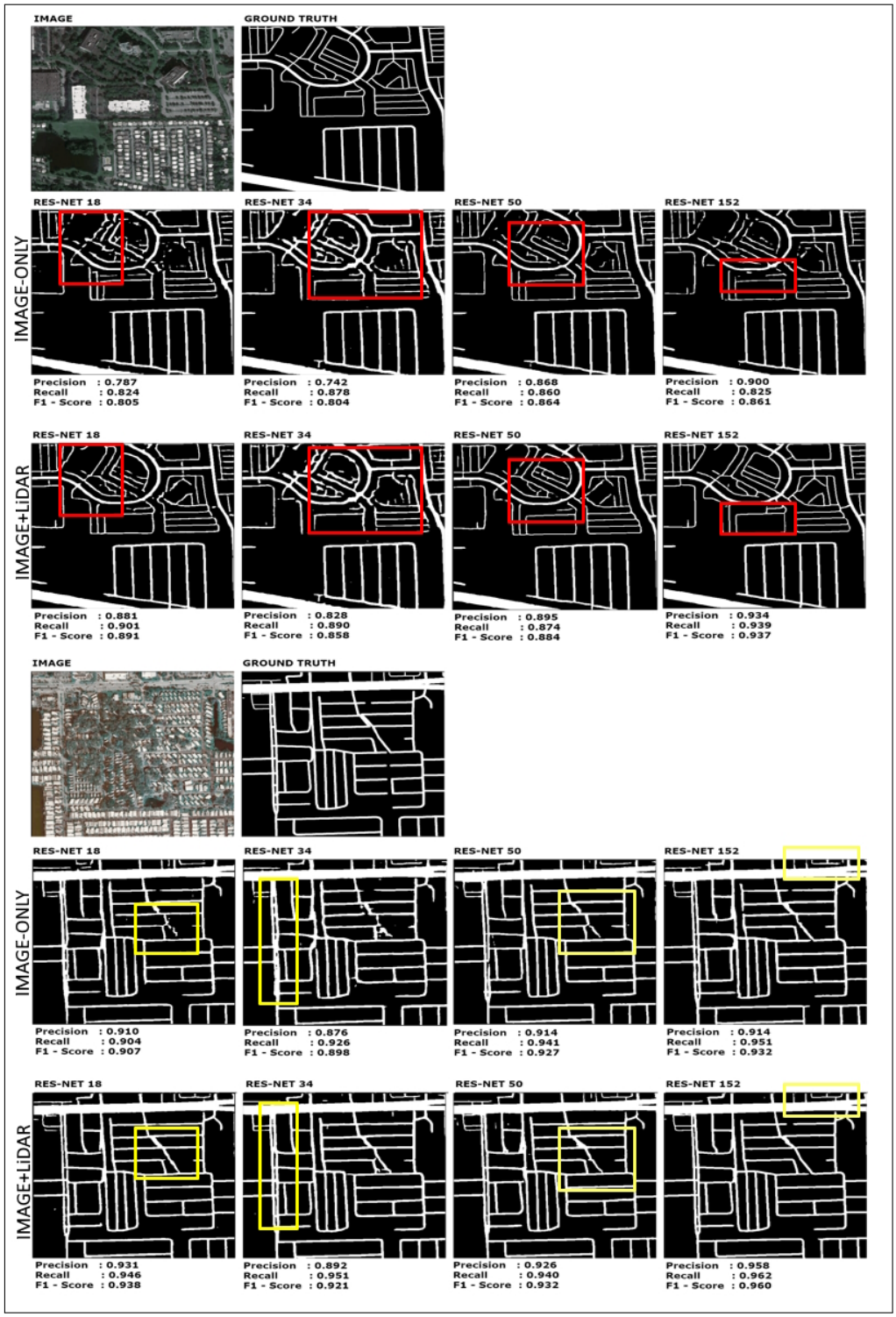


Figure 5. Example of individual prediction samples.

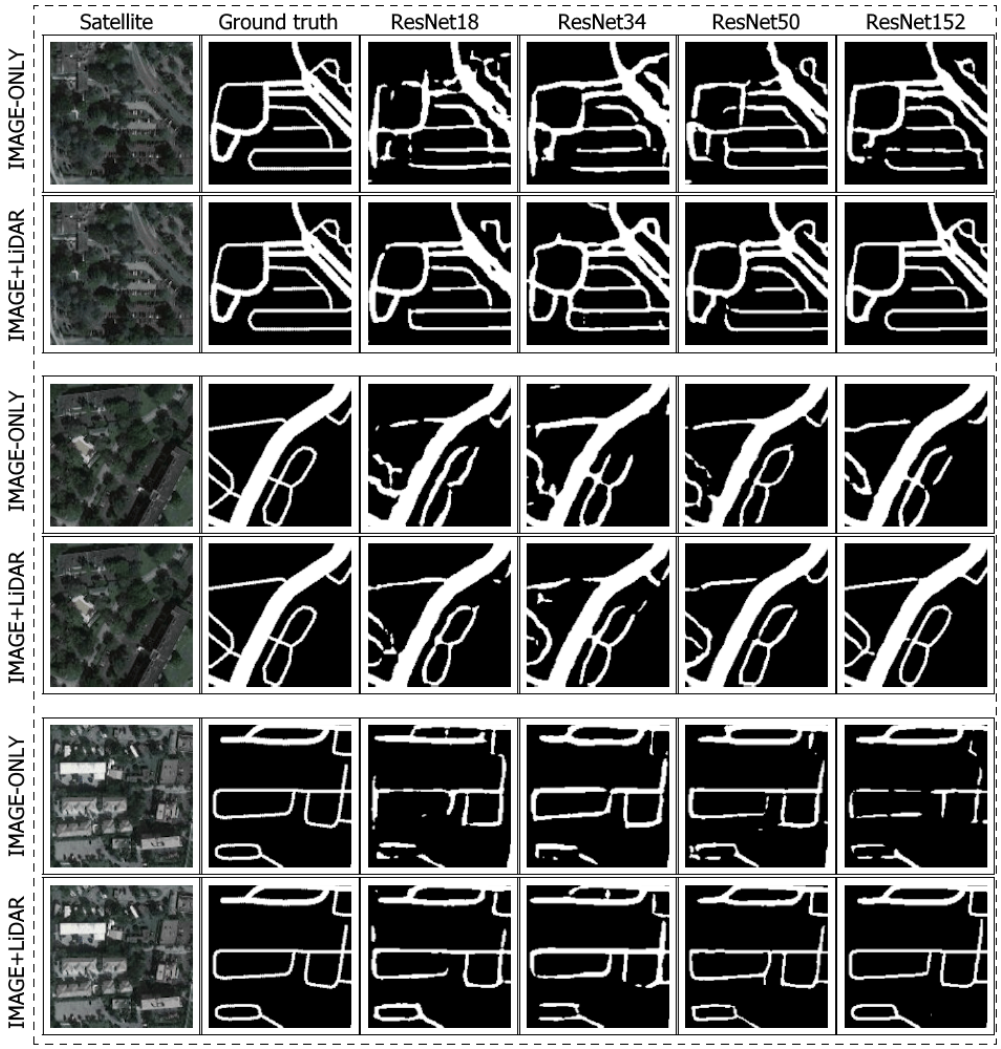


Figure 6. Example of predictions in woodland areas.

In order to highlight the contribution of highly important features to the model's performance, model training was repeated with the same hyper-parameters and backbone (ResNet152) using only the selected feature set (change in absolute heights, optical image, and linearity). The statistical results of the model performance trained with the three most important features are closer to the model trained with the whole feature set (see Table 2). However, this model is inferior compared to the best model achieved, as the other features, whose relative importance is smaller, also contribute to the model's performance at a total of 30%. The use of the three most important features showed better performance compared to the model trained with satellite images only. Even though the feature selection reduced the computational burden by employing a smaller feature set, the contribution of less impacting features cannot be disregarded, as they provide useful information regarding the geometric properties.

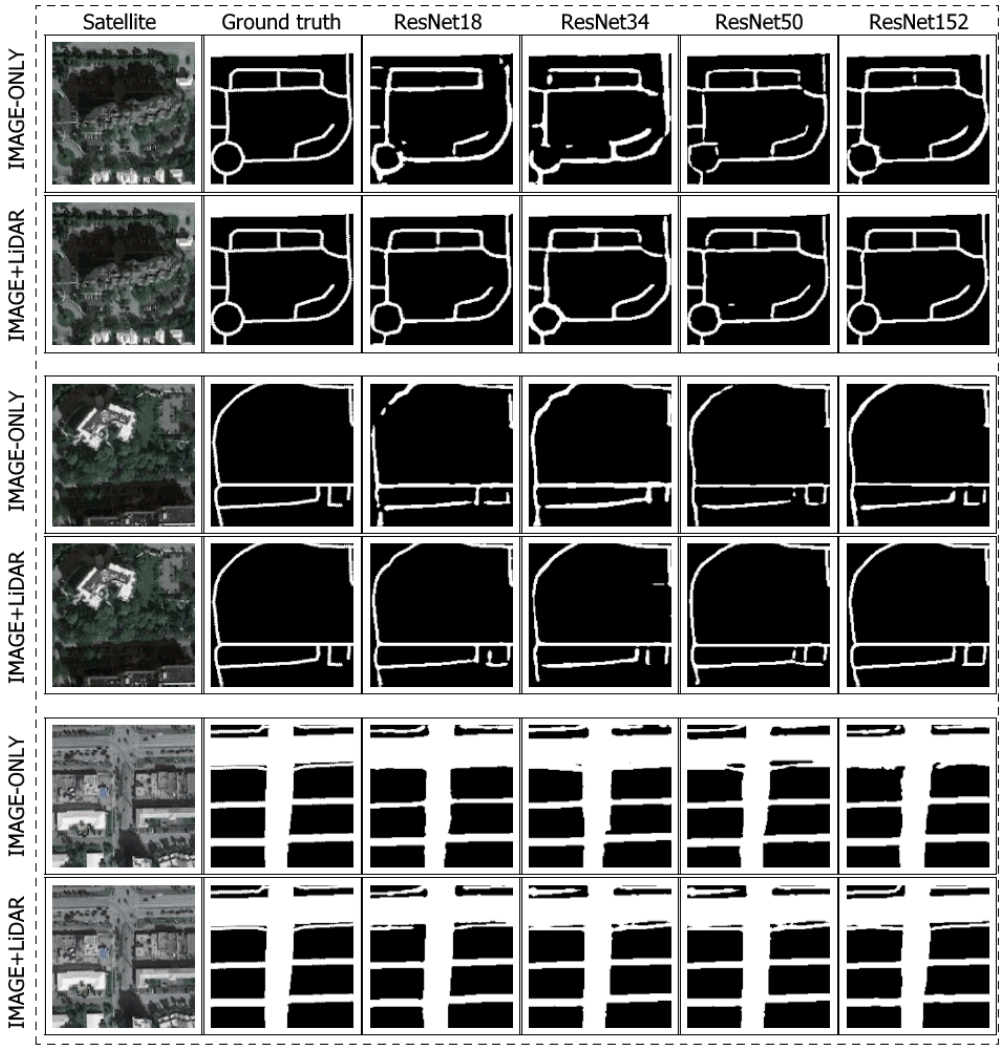


Figure 7. Example of predictions over shadow areas.

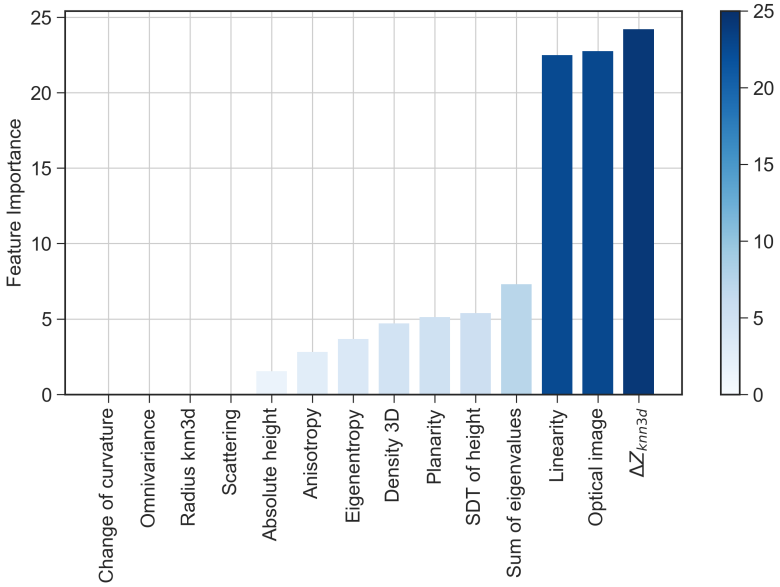


Figure 8. Importance of optical image and LiDAR features.

Table 2. The prediction statistics of model trained with selected features.

Input Data	Backbone	Metrics				
		Precision	Recall	F1-Score	IoU	Kappa
Image	ResNet152	0.862	0.809	0.830	0.732	0.810
Image + LiDAR		0.880	0.852	0.863	0.781	0.843
Selected features		**0.865**	**0.845**	**0.847**	**0.752**	**0.824**

4. Conclusions

In this study, a feature-wise fusion strategy of optical images and point cloud was performed to enhance the road segmentation performance of a deep learning model based on Deep Residual U-Net architecture. In order to compensate for the missing information in optical satellite images that stems from the obstacles and shadow effects in the scenery, we proposed to combine 2D and 3D information to compensate for the absence of depth information in satellite images. For this purpose, high-resolution satellite images and their corresponding road masks, generated over Florida state, were combined with 3D geometric features computed from airborne LiDAR data. In the proposed fusion technique, the optical satellite images were fed into U-Net-based model architecture to generate deep features. Before the final convolution layer, these high-level features were concatenated with the geometric features of the point cloud.

The experimental results validated the effectiveness of the proposed fusion approach for enhancing road segmentation. The combination of 2D and 3D information from satellite images and LiDAR point cloud showed superior results in areas where the satellite image-only models could not predict the road pixels accurately. In challenging areas where the road pixels precluded observation, such as wooded areas and shadows from objects in scenery, the deep learning models trained with the fusion approach predicted the road pixels and the continuity of the road network better than deep learning models trained with satellite images only. The study provided new insight into the relationship between the 2D and 3D features of satellite images and LiDAR data. Moreover, the findings of this study showed the combination of data from various sources can be promising for enhancing the quality of road segmentation. The importance of the features showed that while the optical images have a significant impact on the prediction of road pixels, the contribution of LiDAR features, specifically linearity and height difference within neighbouring points, can lead to a more effective segmentation model development for the extraction of the road network. These 3D features and the geometric relation between the neighbouring points are proven to be significant as the optical images and indicate the importance of contextual information among the point cloud data. It is necessary to further investigate the optimal integration of the most important LiDAR features with the optical satellite images so as to not only capture the geometric properties of the road network efficiently but also to minimize computation costs. It is worth noting that the advancement in the remote sensing data collection techniques and mobility of measuring platforms for LiDAR and laser scanning can ease the generation of accurate and reliable point cloud data. Hence, with these technological developments, the combination of 2D and 3D information from the road networks to increase the performance of deep learning models can be a feasible solution for road extraction studies over challenging areas. The aim of this study is not to exceed the performance of all existing models, but rather to show the combination of optical images and LiDAR can exceed the performance of satellite-only segmentation models. The improved model showed superiority over the problematic areas. In future studies, multi-model fusion strategies will be analyzed to exploit the contribution of LiDAR features that were found to be most effective in road network extraction.

Author Contributions: Conceptualization, O.O.; methodology, O.O. and M.K.; software, O.O. and M.S.I.; investigation, O.O. and M.S.I.; resources, O.O. and M.S.I.; data curation, O.O. and M.S.I.; writing—original draft preparation, O.O. and M.S.I.; writing—review and editing, M.K. and D.Z.S.; visualization, O.O.; supervision, M.K. and D.Z.S.; project administration, O.O. and M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The satellite images are available through Google Maps API services. The LiDAR point clouds can be accessed from the database of the U.S. Geology Survey via https://apps.nationalmap.gov/lidar-explorer/ (accessed on 3 April 2023).

Acknowledgments: The research presented in this article constitutes a part of the first author's Ph.D. thesis study at the Graduate School of Istanbul Technical University (ITU). The first author carried out this research during his visit to TU Berlin with the Scientific and Technological Research Council of Türkiye (TÜBİTAK) 2214/A fellowship program numbered 1059B142000410.

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

References

1. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]
2. Liu, Z.; Tan, Y.; He, Q.; Xiao, Y. SwinNet: Swin Transformer Drives Edge-Aware RGB-D and RGB-T Salient Object Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 4486–4497. [CrossRef]
3. Wang, G.; Zhang, F.; Chen, Y.; Weng, G.; Chen, H. An Active Contour Model Based on Local Pre-Piecewise Fitting Bias Corrections for Fast and Accurate Segmentation. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–13. [CrossRef]
4. Song, M.; Civco, D. Road Extraction Using SVM and Image Segmentation. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 1365–1371. [CrossRef]
5. Wegner, J.D.; Montoya-Zegarra, J.A.; Schindler, K. A Higher-Order CRF Model for Road Network Extraction. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1698–1705. [CrossRef]
6. Grinias, I.; Panagiotakis, C.; Tziritas, G. MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 145–166. [CrossRef]
7. He, L.; Peng, B.; Tang, D.; Li, Y. Road Extraction Based on Improved Convolutional Neural Networks with Satellite Images. *Appl. Sci.* **2022**, *12*, 10800. [CrossRef]
8. Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [CrossRef]
9. Henry, C.; Azimi, S.M.; Merkle, N. Road Segmentation in SAR Satellite Images With Deep Fully Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1867–1871. [CrossRef]
10. Li, Y.; Peng, B.; He, L.; Fan, K.; Tong, L. Road Segmentation of Unmanned Aerial Vehicle Remote Sensing Images Using Adversarial Network with Multiscale Context Aggregation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2279–2287. [CrossRef]
11. Zhang, Z.; Miao, C.; Liu, C.; Tian, Q. DCS-TransUperNet: Road Segmentation Network Based on CSwin Transformer with Dual Resolution. *Appl. Sci.* **2022**, *12*, 3511. [CrossRef]
12. Cira, C.I.; Kada, M.; Ángel Manso-Callejo, M.; Alcarria, R.; Sanchez, B.B. Improving Road Surface Area Extraction via Semantic Segmentation with Conditional Generative Learning for Deep Inpainting Operations. *ISPRS Int. J.-Geo-Inf.* **2022**, *11*, 43. [CrossRef]
13. Sariturk, B.; Seker, D.Z. A Residual-Inception U-Net (RIU-Net) Approach and Comparisons with U-Shaped CNN and Transformer Models for Building Segmentation from High-Resolution Satellite Images. *Sensors* **2022**, *22*, 7624. [CrossRef] [PubMed]
14. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. *Remote Sens.* **2020**, *12*, 1444. [CrossRef]
15. Hu, A.; Chen, S.; Wu, L.; Xie, Z.; Qiu, Q.; Xu, Y. WSGAN: An Improved Generative Adversarial Network for Remote Sensing Image Road Network Extraction by Weakly Supervised Processing. *Remote Sens.* **2021**, *13*, 2506. [CrossRef]
16. Sánchez, J.M.; Rivera, F.F.; Domínguez, J.C.C.; Vilariño, D.L.; Pena, T.F. Automatic extraction of road points from airborne LiDAR based on bidirectional skewness balancing. *Remote Sens.* **2020**, *12*, 2025. [CrossRef]
17. Abdollahi, A.; Pradhan, B.; Sharma, G.; Maulud, K.N.A.; Alamri, A. Improving Road Semantic Segmentation Using Generative Adversarial Network. *IEEE Access* **2021**, *9*, 64381–64392. [CrossRef]
18. Zheng, J.; Yang, S.; Wang, X.; Xia, X.; Xiao, Y.; Li, T. A Decision Tree Based Road Recognition Approach Using Roadside Fixed 3D LiDAR Sensors. *IEEE Access* **2019**, *7*, 53878–53890. [CrossRef]

19. Mi, X.; Yang, B.; Dong, Z.; Chen, C.; Gu, J. Automated 3D Road Boundary Extraction and Vectorization Using MLS Point Clouds. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 5287–5297. [CrossRef]
20. Chen, Z.; Deng, L.; Luo, Y.; Li, D.; Marcato Junior, J.; Nunes Gonçalves, W.; Awal Md Nurunnabi, A.; Li, J.; Wang, C.; Li, D. Road extraction in remote sensing data: A survey. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102833. [CrossRef]
21. Li, Y.; Yong, B.; Wu, H.; An, R.; Xu, H. Road detection from airborne LiDAR point clouds adaptive for variability of intensity data. *Optik* **2015**, *126*, 4292–4298. [CrossRef]
22. Li, Y.; Hu, X.; Guan, H.; Liu, P. An efficient method for automatic road extraction based on multiple features from lidar data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.-ISPRS Arch.* **2016**, *41*, 289–293. [CrossRef]
23. Hui, Z.; Hu, Y.; Jin, S.; Yevenyo, Y.Z. Road centerline extraction from airborne LiDAR point cloud based on hierarchical fusion and optimization. *ISPRS J. Photogramm. Remote Sens.* **2016**, *118*, 22–36. [CrossRef]
24. Tejenaki, S.A.K.; Ebadi, H.; Mohammadzadeh, A. A new hierarchical method for automatic road centerline extraction in urban areas using LIDAR data. *Adv. Space Res.* **2019**, *64*, 1792–1806. [CrossRef]
25. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv* **2016**, arXiv:1612.00593.
26. Wu, W.; Qi, Z.; Li, F. PointConv: Deep Convolutional Networks on 3D Point Clouds. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9613–9622.
27. Ma, H.; Ma, H.; Zhang, L.; Liu, K.; Luo, W. Extracting Urban Road Footprints from Airborne LiDAR Point Clouds with PointNet++ and Two-Step Post-Processing. *Remote Sens.* **2022**, *14*, 789. [CrossRef]
28. Zhou, K.; Ming, D.; Lv, X.; Fang, J.; Wang, M. CNN-based land cover classification combining stratified segmentation and fusion of point cloud and very high-spatial resolution remote sensing image data. *Remote Sens.* **2019**, *11*, 2065. [CrossRef]
29. Sun, Y.; Fu, Z.; Sun, C.; Hu, Y.; Zhang, S. Deep Multimodal Fusion Network for Semantic Segmentation Using Remote Sensing Image and LiDAR Data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [CrossRef]
30. Audebert, N.; Le Saux, B.; Lefèvre, S. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 20–32. [CrossRef]
31. Zhang, Z.; Zhang, X.; Sun, Y.; Zhang, P. Road centerline extraction from very-high-resolution aerial image and LiDAR data based on road connectivity. *Remote Sens.* **2018**, *10*, 1284. [CrossRef]
32. Zhou, K.; Xie, Y.; Gao, Z.; Miao, F.; Zhang, L. Funet: A novel road extraction network with fusion of location data and remote sensing imagery. *ISPRS Int. J.-Geo-Inf.* **2021**, *10*, 39. [CrossRef]
33. Torun, O.; Yuksel, S.E. Unsupervised segmentation of LiDAR fused hyperspectral imagery using pointwise mutual information. *Int. J. Remote Sens.* **2021**, *42*, 6465–6480. [CrossRef]
34. Gao, L.; Shi, W.; Zhu, J.; Shao, P.; Sun, S.; Li, Y.; Wang, F.; Gao, F. Novel framework for 3D road extraction based on airborne LiDAR and high-resolution remote sensing imagery. *Remote Sens.* **2021**, *13*, 4766. [CrossRef]
35. U.S. Geological Survey. 3D Elevation Program LiDAR Point Cloud. 2019. Available online: https://rockyweb.usgs.gov/vdelivery/Datasets/Staged/Elevation/LPC/Projects/USGS_LPC_FL_Southeast_B1_2018_LAS_2019/ (accessed on 1 March 2023).
36. Ozturk, O.; Isik, M.S.; Sariturk, B.; Seker, D.Z. Generation of Istanbul road data set using Google Map API for deep learning-based segmentation. *Int. J. Remote Sens.* **2022**, *43*, 2793–2812. [CrossRef]
37. Bentley, J.L. Multidimensional binary search trees used for associative searching. *Commun. ACM* **1975**, *18*, 509–517. [CrossRef]
38. Weinmann, M.; Jutzi, B.; Hinz, S.; Mallet, C. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 286–304. [CrossRef]
39. Lai, X.; Yang, J.; Li, Y.; Wang, M. A Building Extraction Approach Based on the Fusion of LiDAR Point Cloud and Elevation Map Texture Features. *Remote Sens.* **2019**, *11*, 1636. [CrossRef]
40. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, 18th International Conference, Munich, Germany, 5–9 October 2015*; Springer: Berlin/Heidelberg, Germany, 2015. [CrossRef]
41. He, K.; Sun, J. Convolutional Neural Networks at Constrained Time Cost. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015. [CrossRef]
42. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019. [CrossRef]
43. Zhang, X.; Han, X.; Li, C.; Tang, X.; Zhou, H.; Jiao, L. Aerial Image Road Extraction Based on an Improved Generative Adversarial Network. *Remote Sens.* **2019**, *11*, 930. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.