*Annals of Emerging Technologies in Computing (AETiC)*
*Vol. 8, No. 3, 2024*

*Research Article*

# Optimization of Road Detection using Semantic Segmentation and Deep Learning in Self-Driving Cars

**Mohammed Sameeh Hammoud[1] and Sergey Lupin[2,*]**

[1]National Research University of Electronic Technology, Russia
hammoudmsh93@gmail.com
[2]National Research University of Electronic Technology, Russia
lupin@miee.ru
*Correspondence: lupin@miee.ru

Received: 20th October 2023; Accepted: 17th May 2024; Published: 1st July 2024

**Abstract: Robust and accurate road detection is an essential part of Automatic Driver Assistance Systems (ADAS). Self-driving Cars have the capability to revolutionize the way we travel, making transportation safer, more effective and more available to all. With the ability to navigate roads without human intervention, self-driving cars can reduce the number of accidents caused by human error, eliminate the demand for drivers to be behind the wheel and make it easier for people who can't drive, such as the elderly or disabled people to get around. In addition, self-driving cars can enhance traffic flow by reducing congestion and optimizing routes, eventually saving time and reducing emissions. As technology continues to advance, self-driving cars are poised to transform the transportation industry and change the way we think about mobility. In this work, a convolutional neural network-based deep learning to achieve road detection based on image segmentation to be applied in self-driving cars. In addition to our proposed network, multiple experiments were conducted to investigate the impact of different deep-earning architectures on performance. A public dataset called the KITTI road dataset is used to train and validate the model. The images were down-sampled from 1224x370 to 256x256. We compared our model's performance with the performance of popular deep learning architectures such as Unet and LinkNet A transfer learning technique is used while training the models based on network weights trained on the famous dataset ImageNet, including popular architectures such as ResNet, VGG, SeresNet and EfficientNet. The results show that our model achieves an F1-score of 0.9909, outperforming Unet and LinkNet architectures. In the second place, the best results were obtained based on Unet and ResNet50 with an F1-score of 0.9904.**

**Keywords:** *Advanced Driver Assistance Systems; Computer vision; Deep Learning; Image segmentation; Mobile Robots; Navigation; Road Detection; Self-Driving Cars*

## 1. Introduction

Road recognition is an essential task of self-driving cars, enabling the vehicle to understand and navigate its surroundings. Street recognition allows the robot to drive on unstructured streets. In addition, driver assistance systems are to be provided to prevent human error and accidents. Road recognition systems use a combination of sensors, cameras and algorithms to identify and track the road surface, lane markings and other important road features. This information is then used to control the vehicle's steering, acceleration and braking, allowing it to navigate traffic and safely avoid obstacles.

Roads can be either urban (structured) or rural (unstructured). In urban areas, streets and their elements have a specific structure, so street recognition, in this case, is straightforward. In contrast, unstructured roads are difficult to recognize because they have no defined structure. Road detection is highly hardware-dependent. The system can include a single camera, a stereo camera, or multi-spectral

Mohammed Sameeh Hammoud and Sergey Lupin, "Optimization of Road Detection using Semantic Segmentation and Deep Learning in Self-Driving Cars", Annals of Emerging Technologies in Computing (AETiC), Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 51-63, Vol. 8, No. 3, 1st July 2024, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2024.03.004, Available: http://aetic.theiaer.org/archive/v8/v8n3/p4.html.

sensors. The hardware defines two types of road detection, either based on monocular or based on multiple views (stereo/radar).

Machine learning is a subfield of AI that develops algorithms allowing computers to learn from data and make decisions. Deep learning is a subfield of machine learning that uses artificial neural networks with multiple layers to extract patterns from complex data. The key difference between machine learning and deep learning is the complexity and structure of the algorithms used. Additionally, ML requires a features set, while DL extracts and learns from features. Deep learning requires more data for training, and it can learn hierarchical data representations.

With advances in machine learning and computer vision technology, road recognition systems are becoming increasingly accurate and reliable, paving the way for a future where self-driving cars will be a common sight on our roads. In deep learning, artificial neural networks are trained to recognize patterns and make data-based decisions. It uses algorithms and architectures inspired by the structure and function of the human brain to learn from large amounts of data. Deep learning has proven itself in a different task, such as image and speech recognition, natural language processing and autonomous vehicles. Image segmentation divides an image into multiple segments or regions, each corresponding to a different object or part of the image. The goal of image segmentation is to simplify the image so that it can be analyzed and understood. Image segmentation based on deep learning has two types: semantic segmentation and instance segmentation. Semantic segmentation Maps each pixel in an image to a label or category based on the object or feature to which it belongs. For example, in a street scene, the semantic segmentation would label each pixel as belonging to a car, a pedestrian, a building, or a road. This type of segmentation is useful in object detection and scene understanding. Instance segmentation assigns a label to each pixel in an image and distinguishes between individual instances of the same object or feature. For example, in a street scene, instance segmentation would not only identify each pixel as belonging to a car but would also distinguish between different cars in the scene. This type of segmentation is useful for tasks such as object tracking and counting.

The first deep learning model in image segmentation – Unet [1], aimed to identify objects within an medical image. It was originally proposed in 2015 by researchers at the University of Freiburg in Germany. Unet consists of an encoder network that down-samples the input image and a decoder network that up-samples the output to the original image size. The encoder and decoder networks are linked by skipping connections that allow the model to retain spatial information from previous layers. This architecture is commonly used in medical imaging applications, such as identifying tumours in MRI scans. Good performance is achieved even with a small data set.

EfficientNet is a CNN-based model, which achieves State of the art Architecture (SOTA) in terms of performance on a range of computer vision tasks, comprising image classification, object detection and segmentation. These models are highly efficient in terms of computational resources while maintaining high accuracy. This is attained via using a combination of techniques such as compound scaling, which comprises scaling the network properties such as the width, depth and its resolution in a balanced way and using a novel architecture called the "swish" activation function.

MobileNet [2] is also a CNN-based model that is designed for mobile and embedded devices with limited computational resources. They are designed and optimized specifically for efficient inference on devices with low memory and processing power, such as smartphones and Internet of Things (IoT) devices. MobileNet achieves this by using depth-wise separable convolutions, which separate the spatial and channel-wise convolutions, reducing the computational complexity by reducing the parameters number. This makes MobileNet models much smaller and faster than traditional convolutional neural networks while maintaining high accuracy on various computer vision tasks like image classification, object detection and segmentation. EfficientNet and MobileNet models have been pre-trained on large datasets such as ImageNet and can be fine-tuned on specific tasks with relatively small amounts of data. Attention Gates (AGs) with the ability to focus on some parts of images (or specific classes more than others) were proposed [3]. This model learns how to highlight the useful features for a specific task and suppress irrelevant ones. This architecture combined with CNN like Unet can achieve minimal computational cost while increasing the performance. AGs are highly efficient for organ identification. LinkNet [4] allows models to be trained with a smaller number of parameters and extensive calculations to be carried out efficiently. LinkNet outperforms SegNet, ENet, Dilation 8/10 and Deep-Lab CRF. Moreover, it is applicable in embedded systems (it was tested on NVIDIA TXI and NVIDIA Titian X).

The major contributions of our work are summarized as follows: 1) Proposing a deep learning model based on Unet based on RGB images with a combination of Dice-Loss and Focal Loss. 2) Investigation of the performance of using transfer learning based on UNet and LinkNet architecture. Comparing the proposed model's performance with multiple models, achieving an F1-score of 0.9908. This article is structured as follows: A review of literature is covered in Section2. While Section3 covers the implementation details, including dataset and evaluation metrics. lastly, the obtained results and their discussion are covered in Section 4.

## 2. Related Works

Several studies were proposed to achieve road segmentation. They are different with the input they supply to their algorithms. The input can be images [5-6], point clouds [7-8], or combination of them [9-14].

A unified architecture was proposed to perform segmentation, classification and detection in real-time [5]. All these tasks used a shared encoder (joint training), resulting in a faster inference time of 42.48ms for three tasks. Their model is scalable and is capable of dealing with different image sizes. The detection decoder achieves a better speed-accuracy ratio using Mask-RCNN and fast regression design from YOLO with the size-adjusting Region of Interest (ROI) alignment of Faster-RCNN were used to achieve better speed-accuracy ratio. Instead of using only one image as input, two images were used as input to a hybrid mode of CNN and distributed Long short-term memory (LSTM) [15], one for near-range segmentation with the resized frame to 600x160 and the other for far-range segmentation with cropped to 600x160 in the center. This approach enhanced the feature extraction and processing faster than CNN. Both images were provided in parallel to two instances of the model. The model has 348,801 parameters with an inference time of 16 ms for one image. The model's accuracy is comparable with the existing solutions, but with fewer calculations and less time, making it suitable for limited-resource devices. Despite CNN's high accuracy and ease of implementation in a graphical processing unit (GPU), it requires high memory(parameter-heavy) and computational power. Additionally, it demands a considerable amount of data to obtain high accuracy. Instead of providing the image to CNN, researchers provided irregular superpixels as input to CNN [6]. This model is applicable in real-world tasks with less complexity than traditional CNN since it uses irregular superpixels. Complexity was reduced using CRF to refine the superpixels touching the road boundary.

Visual images suffer from visual noises, such as illumination changes, blurry images, ambiguous appearance and overexposure. To overcome this issue, a Fast, Fully Convolutional Network (FCN) was proposed based on Light Detection and Ranging (LiDAR) data to achieve pixel-wise semantic segmentation [7]. It achieved high accuracy in any light conditions in real-time using GPU-accelerated platforms by transforming point cloud data into images. Despite its high accuracy, it is still less than image-based CNN-LSTM performance [15]. A similar study, Progressive Adaptation-aided Road Detection (PLARD), was proposed using LiDAR [8]. It takes benefits from both visual and LiDAR information by applying two models. The first one (data space adaption) transforms the LiDAR data space into a visual data space, making detecting roads easier. The second one (feature space adaptation) uses a cascade fusion structure to transform between LiDAR feature space and the visual features space.

Further improvements were introduced using hybrid data, such as images and point clouds [9-14]. An end-to-end semantic segmentation network was proposed for road segmentation [9]. They used an image and LiDAR fusion model where the data was fused in the decoder part instead of the encoder. The precision was increased using a proposed method (pyramid projection), which improves the map generation of the multi-scale LiDAR obtained from LiDAR point clouds. Additionally, alongside the data fusion model, a multi-path refinement network was used to get better road detection performance. A hybrid model, "Siamese Network" based FCN-8s, uses RGB images and projected point clouds into sparse images, where each input is processed in a separate branch [11]. Their extracted features are fused before pooling layers to increase the performance. Compared to [6], this mode did not classify road edge areas well. Another research called SNE-RoadSeg+ used RGB and depth images as input [12]. It consists of a lightweight model (SNE+) to achieve accurate surface normal estimation and a data fusion Deep Convolutional neural network (DCNN) (RoadSeg+) to achieve a trade-off between accuracy and efficiency using deep supervision on the intermediate layers of the network based on the model pruning. Different DCNNs with SNE/SNE+ embedded layers were evaluated, With SNE+ outperforming the ones with SNE. SNE-RoadSeg+ outperforms all other free-space detection approaches. RoadSeg+ with SNE+ embedded outperforms all other DCNNs, with an Intersection over union OR Jaccard index (IoU) increment of around 1–11%. With TensorRT, it provides real-time processing on resource-limited embedded computing. Similarly, a USNet network [13] used two-light sub-networks to learn features from the input (RGB and depth) instead of using cross-modal feature fusion. This model decreased inference time while maintaining high accuracy, with the ability to work with Frames Per Second (FPS) >=43. The fusion of models is guided using an uncertainty-aware fusion module. Another research, siamese FCN [10], segments roads based on three inputs: RGB image, contour maps and location priors. The first two images (RGB image, grey contours) are supplied simultaneously to two networks sharing their parameters. The last one helps to improve the performance. Compared to FCN [7], their proposed network learns more representative features and even learns 30% faster utilizing road boundaries. Researchers proposed an end-to-end CNN-based deep learning called 3D-DEEP [14] to perform road segmentation using two types of data: RGB images and three-dimensional information (point clouds, Disparity maps) projected in a two-dimensional plane. Based on BiSeNet [16], the spatial and context information are considered to achieve high performance. Instead of using raw point clouds with images, synchronized three-dimensional information with images was used to reduce complexity and increase performance simultaneously. 3D-DEEP achieved good performance based on point cloud data, while stereoscopic vision performed badly.

Traditional road detection uses handcrafted features, such as Local Binary Patterns (LBP), Scale Invariant Feature Transform (SIFT) and Histogram of Oriented Gradient (HOG), amongst others, or filter banks. Despite these algorithms detecting robust features to scale and rotation, and they are very specific for images. The feature descriptors are different between images of the road in sunlight and rainy weather. Also, their prediction is highly noisy, which, in turn, will be supplied to the graphical models: Conditional random forest (CRF) or Markov fields (MF). CRFs are computationally intensive and slow and have errors. These errors are due to unplanned predictions smoothing. On the contrary, CNN has high accuracy and is easy to implement on a Graphical Processing Unit (GPU), but it requires high memory(parameter-heavy) and computational power. Additionally, it demands a considerable amount of data to obtain high accuracy. Image-based models are more sensitive to visual noise. On the contrary, cloud-based is a better choice than image-based. Higher accuracy can be achieved using combination of images and point-cloud.

## 3. Methods and Methodology

### 3.1. Dataset

As a benchmark, we used a public dataset called KITTI road dataset[1] [17]. It consists of two folders, one for training with 290 images and the other for testing with 289 images recorded in real-life scenarios. The images are of size 1245x375 or 1224x370. In the training part, there are visible light images, with corresponding masks represented by manual annotation of the images. These masks are available for two different road terrain types. It contains the category URBAN ROAD including Urban Marked (UM), Urban Unmarked (UU) and Urban Multiple Marked (UMM). The lane where the vehicle is currently driving is available only for category UM, as shown in Figure 3 in the middle column. Since we are concerned with road segmentation, we dropped that lane masks. Some samples from the dataset are illustrated in Figure 1, where the columns represent the image, the mask for the lane where the car is moving and the full road respectively.

[1] https://www.cvlibs.net/datasets/kitti/eval_road.php

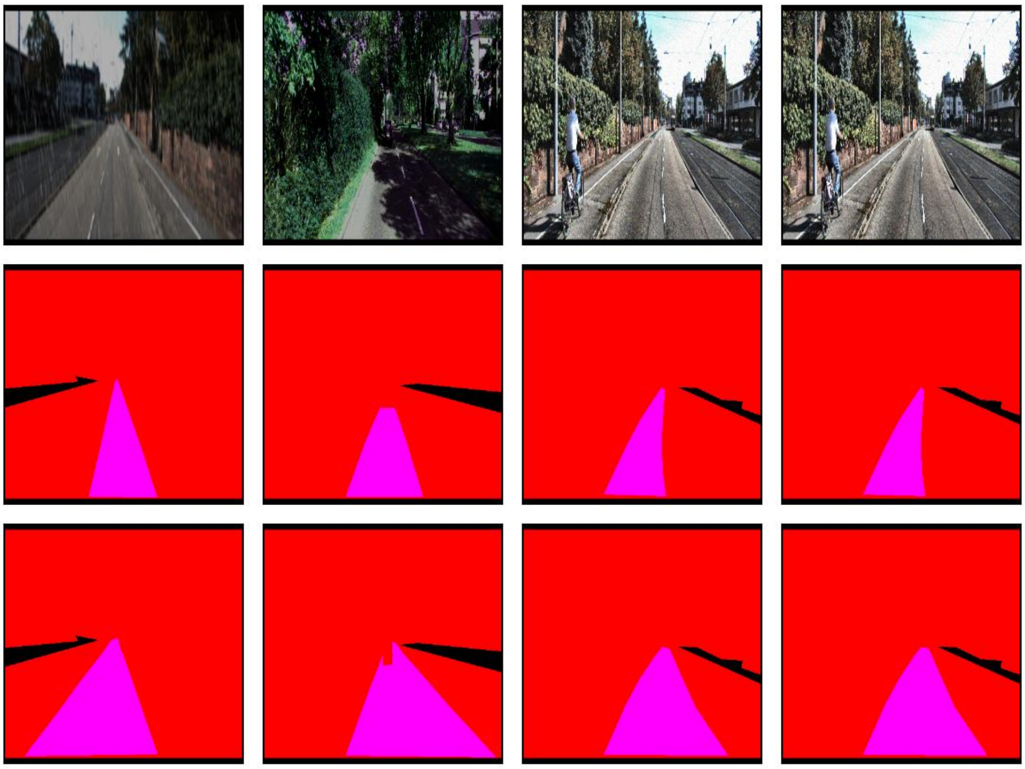

**Figure 1.** Samples from the dataset, where columns represent the image, lane mask and road mask

### 3.2. Implementation Details

We conducted several experiments to improve our machine learning model. To reduce training time, we used transfer learning techniques by utilizing known models with pre-trained weights on the well-known ImageNet dataset. We tested two models, namely LinkNet and Unet, with different networks including ResNet, VGG, and mobileNet. Furthermore, we developed our own network architecture and trained it from scratch to enhance the accuracy of our model.

First of all, the images are down-sampled to a size of 256x256. The proposed architecture in this paper follows Unet architecture but with some modifications, as shown in Figure 2. The encoder part is 5 stages of convolution blocks. Each one includes two convolution layers 3x3 with rectified linear units (ReLU) as an activation function. A max-pooling layer follows them to down-sample the information. We added some drop layers to avoid over-fitting. The decoder up-samples the output to the original image size. The encoder and decoder networks are linked by skipping connections that allow the model to retain spatial information from previous layers. We trained the models for 100 epochs and a batch size of 128. The initial weights of the models are initialized using the "He normal" method. The data is split into three parts: training, validation and testing in a ratio of 0.7:0.2:0.1. As a loss function for training, we used a combination of Diceloss and focal loss function, as shown in Equation 1, Equation 2. Focal loss [18] is helpful since we are working with imbalanced data, so focal loss gives more weight to hard-detected classes (the minority class). We saved the weight for the best model based on the best-acquired validation loss. we used an early stopping technique with patience of 20 epochs.

Python[2] is used as a programming language, while TensorFlow[3], Keras and segmentation models application programming interface (API)[4] are used for model implementation. Additionally, OpenCV[5] is used to perform image processing. The source code of the project will be available here[6].

[2] https://www.python.org/

[3] https://www.tensorflow.org/

[4] http://github.com/qubvel/segmentation_models/

[5] https://opencv.org/

[6] https://github.com/Hammoudmsh/Road-using-Segmentation-Semantic-Segmentation-for-Self-Driving-Car.git

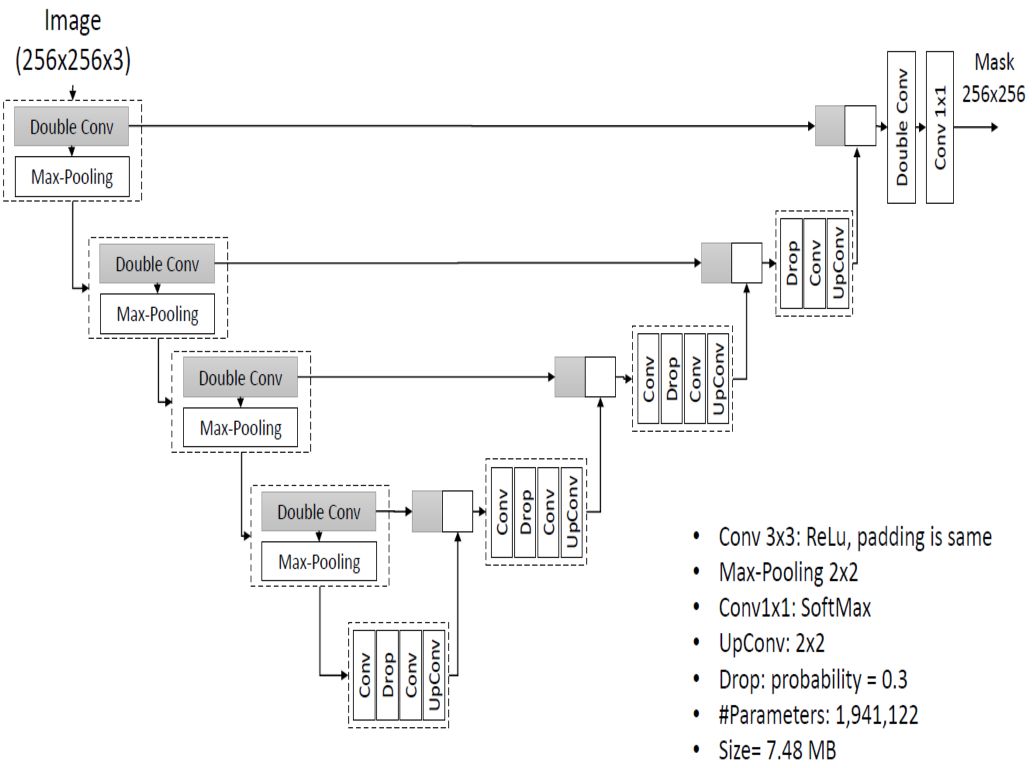

**Figure 2.** The structure of the proposed deep learning model

### 3.2.1. Data Augmentation Process

3.2.1.1. Data augmentation

Data augmentation is a technique used to increase the amount of data, for example, in the case of images, and involves operations like resizing, cropping, zooming, and blurring. As mentioned previously, the KITTI dataset is small, so we used data augmentation to increase the data. In our work, the following pipeline of methods are applied, using [19]:

- Gaussian noise with p=0.5
- flip horizontally with probability p = 0.5
- random rotation
- Colour Jitter
- RGB shifting
- Random brightness
- histogram equalization via Contrast Limited Adaptive Histogram Equalization (CLANE)
- adding random rain

The values for all parameters not mentioned in these transformations use the default values from [19]. Some examples of these transformations can be noticed in Figure 1, for example, rotation in the last image.

3.2.1.2. Input pre-processing

Since we are concerned with road segmentation, we pre-processed the masks to be suitable for our task. All unused classes(masks) were cancelled, as shown in Figure 3. As a pre-processing step, images were enhanced using CLANE. CLANE is a technique utilized in image processing to enhance images' contrast. CLANE is a modified version of the classical histogram equalization method that adapts to the local contrast of an image. In CLANE, the image is split into small areas called tiles and histogram equalization is applied to each tile separately. This helps prevent over-enhancement of the image and preserves local contrast.

3.2.1.3. Quantitative evaluation

As evaluation metrics, we used three metrics, represented by F1 score, Average precision (AP), Precision (PRC), Precision (REC), False Positive Rate (FPR), False Negative Rate (FNR) and IoU, where TP, FP, TN, FN stand for True Positive, False Positive, True Negative and False Negative, respectively.

$$FL(p_t) = -\alpha(1 - p_t)^{\gamma} log\ (p_t) \tag{1}$$

$$Loss = DiceLoss + 2 * FocalLoss \tag{2}$$

$$IoU(P,G) = \frac{P \cap G}{P \cup G} \tag{3}$$

$$F1 - score = 2 * \frac{PRC*REC}{PRC + REC} \tag{4}$$

$$FPR = \frac{FP}{TP*FP} \tag{5}$$

$$FNR = \frac{FN}{TP * FN} \tag{6}$$

$$AP = \frac{TP+TN}{TP+FP+TN+ FN} \tag{7}$$

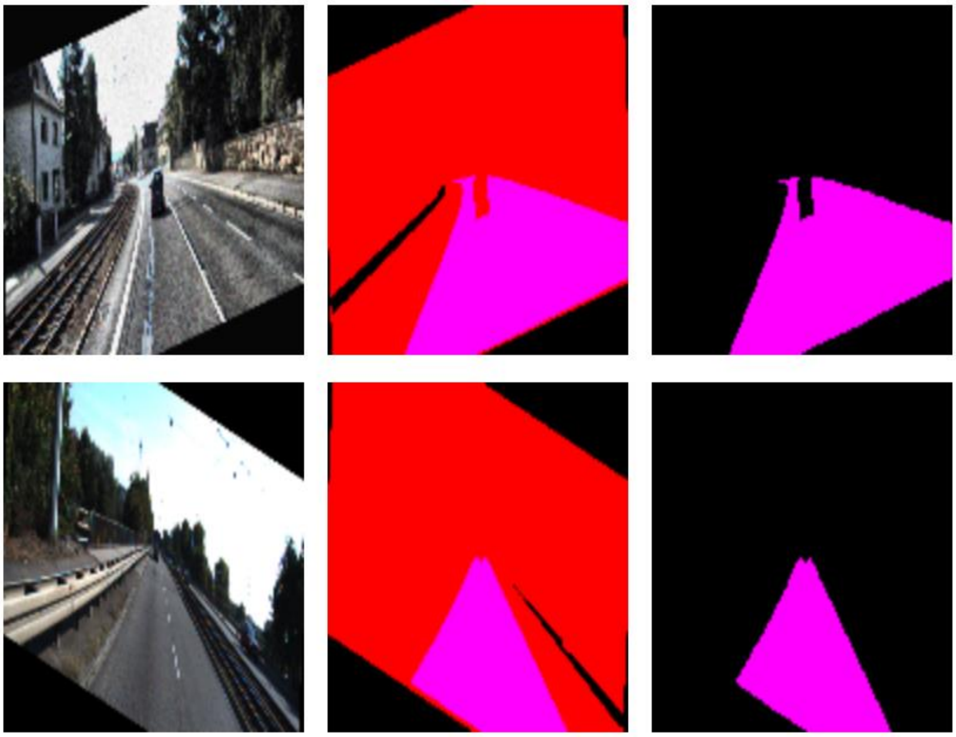

**Figure 3**. Dataset pre-processing

## 4. Results and Discussion

In this section, multiple experiments were conducted. We trained different models based on different architectures, such as LinkNet, Unet and our proposed architecture. The first two architectures used weights from famous networks, such as ResNet, SeresNet, vgg16 and mobilenetv2. The evaluation results for different networks are listed in Table 1. As listed in Table 1, the best performance is obtained from our models with F1-score of 0.9909. In the second place, UNet and LinkNet with resnet50 as backbone achieved F1-score of 0.9841 and 0.9905, respectively.

**Table 1**. The evaluation of trained models on the testing data (testing split) from KITTI dataset, where the best results is bolded and underlined.

| | Backbone | Loss | IoU | Fl-score | Precision |
|---|---|---|---|---|---|
| LinkNet | resnet18 | 0.057 | 0.9354 | 0.9661 | 0.9802 |
| | resnet34 | 0.0601 | 0.9291 | 0.9626 | 0.9782 |
| | resnet50 | 0.0296 | 0.9691 | 0.9841 | 0.9905 |
| | seresnet18 | 0.0431 | 0.9515 | 0.9749 | 0.9854 |
| | seresnet34 | 0.0454 | 0.9527 | 0.9755 | 0.9857 |
| | vgg16 | 0.0555 | 0.937 | 0.967 | 0.9814 |
| | mobilenetv2 | 0.036 | 0.9646 | 0.9818 | 0.9892 |
| UNet | resnet18 | 0.0217 | 0.9771 | 0.9884 | 0.9931 |
| | resnet34 | 0.0192 | 0.9781 | 0.9888 | 0.9934 |
| | resnet50 | 0.017 | 0.9813 | 0.9905 | 0.9944 |
| | seresnet18 | 0.0201 | 0.9784 | 0.989 | 0.9934 |
| | seresnet34 | 0.0195 | 0.9796 | 0.9896 | 0.9939 |
| | vgg16 | 0.0199 | 0.9774 | 0.9885 | 0.9934 |
| | mobilenetv2 | 0.0238 | 0.976 | 0.9878 | 0.9928 |
| Ours | - | 0.0163 | 0.9822 | **0.9909** | 0.9946 |

Despite that the performance of our architecture is close to Unet, Unet is more complex and requires extra memory compared to our architecture.

The visual performance of the proposed model is illustrated on Figure 4. to Figure 6. , where they represent the images from UM, UMM and UU environment from KITTI dataset (TESTING dataset). As noticed, the model successfully detected the roads in different scenarios, but it fails in road images with shadow, such as images 73 and 74 in Figure 4. Similarly, there is wrong detection in images 30 and 60 in Figure 5.

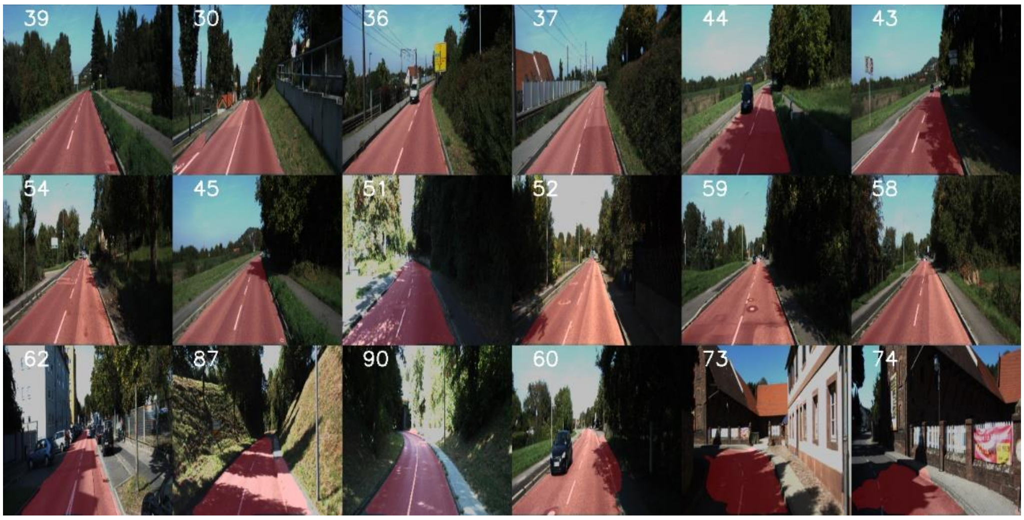

**Figure 4.** UM from TESTING dataset in KITTI

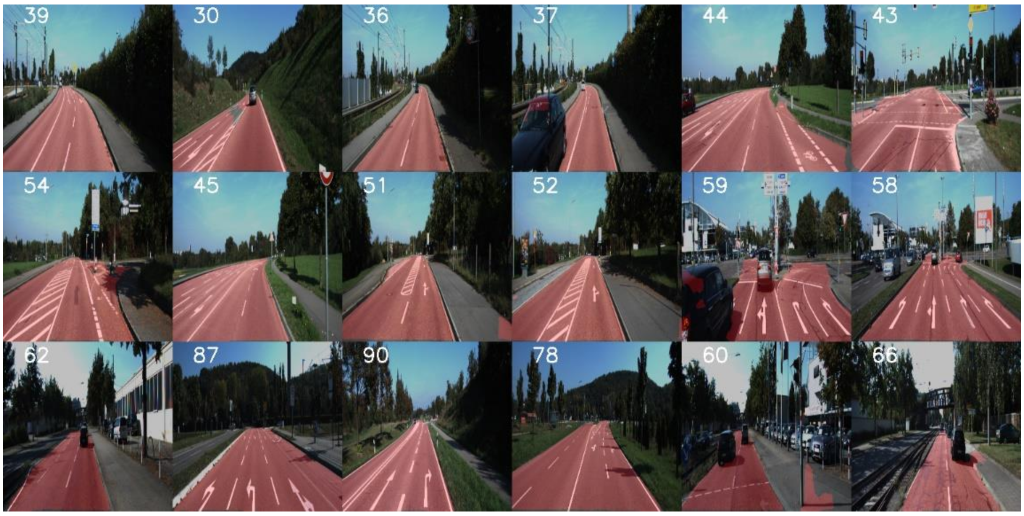

**Figure 5.** UMM from TESTING dataset in KITTI

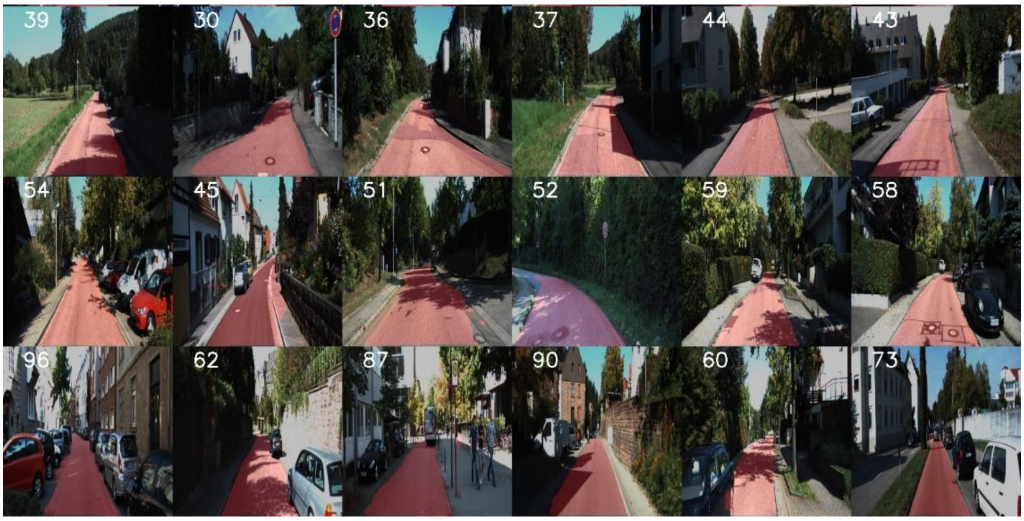

**Figure 6.** UU from TESTING dataset in KITTI

Both Table 2 and Table 3 show the evaluation results of our project on four types of roads: UM, UMM, UU, and Urban roads. We tested the performance of our model on the testing data that we split. We plan to submit our model to the company server to obtain metrics, which are not downloadable. However, testing on the testing data provides insights into the model's performance.

**Table 2.** The Comparison of related works on different parts of KITTI dataset (UM and UMM)

| | UM | | | | | | UMM | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MaxF | AP | PRC | REC | FPR | FNR | MaxF | AP | PRC | REC | FPR | FNR |
| Road Detection using Siamese Network [11] | **91.03** | **84.64** | **89.98** | **92.11** | **4.67** | **7.89** | **93.68** | **89.74** | **93.48** | **93.87** | **7.2** | **6.13** |
| superpixelo N Nc RF [6] | 83.22 | 72.94 | 77.11 | 90.39 | 12.23 | 9.61 | 90.96 | 84.63 | 87.86 | 94.29 | 14.32 | 5.71 |
| Image-LiDARData Fusion [9] | 93.29 | 91.11 | 93.53 | 92.49 | - | - | 95.05 | 94.01 | 95.47 | 95.02 | | |
| SPRAYIT [20] | 88.14 | 91.24 | - | - | - | - | 89.69 | 93.84 | - | - | - | - |
| FCN; [21] | 89.36 | 78.8 | - | - | - | - | 94.09 | 90.26 | - | . | : | - |
| HybridCRF [22] | 90.99 | 85.26 | - | - | - | - | 91.95 | 86.44 | - | - | . | - |
| FTP [23] | 91.2 | 90.6 | - | - | - | ~ | 92.98 | 92.89 | - | - | - | - |
| Upconv [24] | 90.48 | 88.2 | - | - | - | - | 93.89 | 92.62 | - | - | - | - |
| LoDNN [7] | 92.75 | 89.98 | - | - | - | - | 96.05 | 95.03 | - | - | - | - |
| MultiNet [5] | 93.99 | 93.24 | - | - | - | - | 96.15 | 95.36 | - | - | - | - |
| StixelNet IT [25] | 94.05 | 85.85 | - | - | . | - | 96.22 | 91.24 | - | - | - | - |
| RBNet [26] | 94.77 | 91.42 | \|- | - | - | - | 96.06 | 93.49 | - | - | - | - |
| LidCamNet [14] [27] | 95.62 | 93.54. | - | - | . | - | 97.08 | 95.51 | 97.28 | 96.88 | - | - |
| NF2CNN [14] | 96.09 | 88.4 | - | - | - | - | 97.77 | 93.31 | - | - | - | - |
| PSPNet [28] | 95.62 | 92.95 | - | - | - | - | 96.95 | 95.38 | - | - | - | - |
| PLARD [8] [14] | 96.34 | 93.43 | - | - | - | - | 97.53 | 95.61 | 97.75 | 97.79 | - | - |
| PLARD+ [8] | 97.05 | 93.53 | - | - | - | - | 97.77 | 95.64 | - | - | - | - |
| Multipurpose Decoder Deconvolution Network | 93.99 | - | - | - | - | . | 96.15 | - | - | - | - | - |
| Neural Network plus Plane [29] | 90.5 | - | - | - | - | - | 91.34 | - | - | . | - | - |
| Superpixelsc RF with global shape prior [30] | 83.73 | - | - | - | - | - | 87.96 | - | - | - | - | - |
| Graph Based Road Estimation [31] | 85.43 | - | - | - | - | - | 88.19 | - | - | - | - | - |
| Histogram - Based Joint Boosting [32] | 83.68 | - | - | - | - | - | 88.73 | - | - | - | - | - |
| Structured RF [29] | 76.43 | - | - | - | - | - | 90.77 | - | - | - | - | - |
| 3D-DEEP [14] | 95.35 | 93.5 | 93.5 | 93.5 | 93.5 | 93.5 | 93.5 | 95.76 | 97.01 | 97.01 | 97.01 | 97.01 |

**Table 3.** The Comparison of related works on different parts of KITTI dataset (UU and Urban Road) (values in percent)

| | UU | | | | | | Urban Road | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Max F | AP | PRE | REC | FPR | FNR | Max F | AP | PRE | REC | FPR | FNR_ |
| Road Detection using Siamese Network [11] | 88.02 | 75.58 | 86.91 | 89.16 | 4.37 | 10.84 | 91.51 | 85.79 | 90.82 | 92.21 | 5.13 | 7.79 |
| superpixelo N Nc RF [6] | 80.02 | 67.93 | 77.56 | 82.64 | 7.79 | 17.36 | 85.97 | 77.81 | 82.04 | 90.31 | 10.89 | 9.69 |
| Image-LiDARData Fusion [9] | 92.21 | 91.62 | 93.25 | 94.2 | | | 93.98 | 92.23 | 94.06 | 93.9 | - | - |
| SPRAYIT [20] | 82.71 | 87.19 | - | - | - | - | - | - | - | - | - | - |
| FCN; [21] | 86.27 | 75.37 | - | - | - | - | - | - | - | - | - | - |
| HybridCRF [22] | 88.53 | 80.79 | - | - | - | - | - | - | - | - | - | - |
| FTP [23] | 80.62 | 88.93 | - | - | - | - | - | - | - | - | - | - |
| Upconv [24] | 91.89 | 89.44 | - | - | - | - | - | - | - | - | - | - |
| LoDNN [7] | 92.29 | 90.35 | - | - | - | - | - | - | - | - | - | - |
| MultiNet [5] | 93.69 | 92.55 | - | - | - | - | - | - | - | - | - | - |
| StixelNet IT [25] | 93.4 | 85.01 | - | - | - | - | - | - | - | - | - | - |

| | UU | | | | | | Urban Road | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Max F | AP | PRE | REC | FPR | FNR | Max F | AP | PRE | REC | FPR | FNR_ |
| RBNet [26] | 93.21 | 89.18 | - | - | - | - | - | - | - | - | - | - |
| LidCamNet [14] [27] | 94.54 | 92.74 | - | - | - | - | - | - | - | - | - | - |
| NF2CNN [14] | 95.47 | 86.98 | - | - | - | - | - | - | - | - | - | - |
| PSPNet [28] | 95.86. | 92.73 | - | - | - | - | - | - | - | - | - | - |
| PLARD [8] [14] | 95.95 | 95.25 | - | - | - | - | - | - | - | - | - | - |
| PLARD+ [8] | | | | | | | | | | | | |
| Multipurpose Decoder Deconvolution Network | 93.69 | - | - | - | - | - | - | - | - | - | - | - |
| Neural Network plus Plane [29] | 85.55 | - | - | - | - | - | - | - | - | - | - | - |
| Superpixelsc RF with global shape prior [30] | 80.78 | - | - | - | - | - | - | - | - | - | - | - |
| Graph Based Road Estimation [31] | 84.14 | - | - | - | - | - | - | - | - | - | - | - |
| Histogram - Based Joint Boosting [32] | 74.19 | - | - | - | - | - | - | - | - | - | - | - |
| Structured RF [29] | 76.07 | - | - | - | - | - | - | - | | | | |
| 3D-DEEP [14] | 94.67 | 93.04 | 94.23 | 95.12 | 1.9 | 4.88 | 0.9602 | 0.94 | 0.9568 | 0.9635 | 0.0239 | 0.0365 |

## 5. Conclusion

Road detection is an essential task navigation for autonomous cars and mobile robots. Accurate detection is vital in different scenarios and different environments. In this work, an image segmentation based on deep learning was built to detect the road accurately. The proposed model and different architectures were trained and tested on the KITTI dataset. Transfer learning using Unet and LinkNet, with pertained weights from ImageNet, was used. Additionally, we propose a network architecture that performs comparable to Unet and LinkNet. The model achieved an F1-score of 0.9909, while LinkNet based on resenet50 and Unet based on resnet50 achieved 0.9841 and 0.9905, respectively. The suggested model architecture is appealing, however, it necessitates a longer training time. On the other hand, the transfer learning approach can attain comparable performance with less training time. The only drawback is that the model is more intricate, which leads to increased inference time.

One limitation of this work is that it does not show model performance in real-time to show the inference time and memory consumption. In future work, more architectures will be investigated. Moreover, the model will be tested on the KITTI test dataset. We will investigate using key point-based image matching for known areas with road segmentation to increase performance. Finally, we will be developing a model to achieve a trade-off between accuracy and inference time.

## References

[1] Ronneberger Olaf, Philipp Fischer and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation”, in *Proceedings of the Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference*, 5-9 October, 2015, Munich, Germany, Online ISBN: 978-3-319-24574-4, Print ISBN: 978-3-319-24573-7, pp. 234–241, Published by Springer, Cham, DOI: 10.1007/978-3-319-24574-4_28, Available: https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28.

[2] Debjyoti Sinha and Mohamed El-Sharkawy, “Thin MobileNet: An Enhanced MobileNet Architecture”, in *Proceedings of the 2019 IEEE 10th Annual Ubiquitous Computing*, Electronics *& Mobile Communication Conference (UEMCON)*, 10-12 October 2019, New York, NY, USA, E-ISBN: 978-1-7281-3885-5, Print ISBN: 978-1-7281-3886-2, pp. 0280-0285, Published by IEEE, DOI: 10.1109/UEMCON47517.2019.8993089, Available: https://ieeexplore.ieee.org/abstract/document/8993089.

[3] Chen Li, Yusong Tan, Wei Chen, Xin Luo, Yuanming Gao *et al.*, "Attention Unet++: A Nested Attention-Aware U-Net for Liver CT Image Segmentation", in *Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, 25-28 October 2020, E-ISBN: 978-1-7281-6395-6, Print ISBN: 978-1-7281-6396-3, E-ISSN: 2381-8549, Print ISSN: 1522-4880, Published by IEEE, pp. 345-349, DOI: 10.1109/ICIP40778.2020.9190761, Available: https://ieeexplore.ieee.org/document/9190761.

[4] Chaurasia Abhishek and Eugenio Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation", in *Proceedings of the IEEE Visual Communications and Image Processing (VCIP)*, 10-13 December 2017, Petersburg, FL, USA, E-ISBN: 978-1-5386-0462-5, Print ISBN: 978-1-5386-0463-2, pp. 1–4, Published by IEEE, DOI: 10.1109/VCIP.2017.8305148, Available: https://ieeexplore.ieee.org/document/8305148.

[5] Teichmann Marvin, Michael Weber, Marius Zoellner, Roberto Cipolla and Raquel Urtasun, "Multinet: Real-time joint semantic reasoning for autonomous driving", in *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV)*, 26-30 June 2018, Changshu, China, E-ISBN: 978-1-5386-4452-2, Print ISBN: 978-1-5386-4453-9, pp. 1013-1020, Published by IEEE, DOI: 10.1109/IVS.2018.8500504, Available: https://ieeexplore.ieee.org/document/8500504.

[6] Zohourian Farnoush and Josef Pauli, "Coarse-to-Fine Semantic Road Segmentation Using Super-Pixel Data Model and Semi-Supervised Modified CycleGAN", *Journal of Image and Graphics*, December 2022, Print ISSN: 2301-3699, Online ISSN: 2972-3973, pp. 132-144, Vol. 10, No. 4, Published by World Scientific, DOI: 10.18178/joig.10.4.132-144, Available: https://www.joig.net/index.php?m=content&c=index&a=show&catid=79&id=305.

[7] Caltagirone Luca, Samuel Scheidegger, Lennart Svensson and Mattias Wahde, "Fast LIDAR-based road detection using fully convolutional neural networks", in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV)*, 11-14 June 2017, Los Angeles, CA, USA, E-ISBN: 978-1-5090-4804-5, Print ISBN: 978-1-5090-4805-2, pp. 1019-1024, Published by IEEE, DOI: 10.1109/IVS.2017.7995848, Available: https://ieeexplore.ieee.org/document/7995848.

[8] Chen Zhe, Jing Zhang and Dacheng Tao, "Progressive lidar adaptation for road detection", in I*EEE/CAA Journal of Automatica Sinica*, May 2019, E-ISSN: 2329-9274, Print ISSN: 2329-9266, Vol. 6, No. 3, pp. 693 - 702, Published by IEEE, DOI: 10.1109/JAS.2019.1911459, Available: https://ieeexplore.ieee.org/document/8707128.

[9] Liu Huafeng, Yazhou Yao, Zeren Sun, Xiangrui Li, Ke Jia *et al.*, "Road segmentation with image-LiDAR data fusion in deep neural network", *Multimedia Tools and Applications*, 27 July 2019, Online ISSN: 1558-0016, Print ISSN: 1524-9050, Vol. 79, No. 12, pp. 35503–35518, Published by Springer, DOI: 10.1007/s11042-019-07870-0, Available: https://link.springer.com/article/10.1007/s11042-019-07870-0.

[10] Wang Qi, Junyu Gao and Yuan Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection", in *IEEE Transactions on Intelligent Transportation Systems*, 4 October 2017, Online ISSN: 1558-0016, Print ISSN: 1524-9050, Vol. 19, No. 1, pp. 230-241, Published by IEEE, DOI: 10.1109/TITS.2017.2749964, Available: https://ieeexplore.ieee.org/document/8058005.

[11] Liu Huafeng, Xiaofeng Han, Xiangrui Li, Yazhou Yao, Pu Huang *et al.*, "Deep representation learning for road detection using Siamese network", *Multimedia Tools and Applications*, 15 September 2020, E-ISSN: 1573-7721, Vol. 78, pp. 24269–24283, Published by Springer, DOI: 10.1007/s11042-018-6986-1, Available: https://link.springer.com/article/10.1007/s11042-018-6986-1.

[12] Wang Hengli, Rui Fan, Peide Cai and Ming Liu, "SNE-RoadSeg+: Rethinking depth-normal translation and deep supervision for freespace detection", in *Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, 27 September 2021 - 01 October 2021, E-ISBN: 978-1-6654-1714-3, Print ISBN: 978-1-6654-1715-0, pp. 1140-1145, Published by IEEE, DOI: 10.1109/IROS51168.2021.9636723, Available: https://ieeexplore.ieee.org/document/9636723.

[13] Chang Yicong, Feng Xue, Fei Sheng, Wenteng Liang and Anlong Ming, "Fast road segmentation via uncertainty-aware symmetric network", in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, Philadelphia, PA, USA, 23-27 May 2022, E-ISBN: 978-1-7281-9681-7, Print ISBN: 978-1-7281-9682-4, pp. 11124-11130, DOI: 10.1109/ICRA46639.2022.9812452, Available: https://ieeexplore.ieee.org/document/9812452.

[14] Hernandez Álvaro, Suhan Woo, Héctor Corrales, Ignacio Parra, Euntai Kim *et al.*, "3D-DEEP: 3-Dimensional Deep-learning based on elevation patterns for road scene interpretation", in *Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, USA, 2020, 19 October 2020 - 13 November 2020, E-ISBN: 978-1-7281-6673-5, Print ISBN: 978-1-7281-6674-2, E-ISSN: 2642-7214, print ISSN: 1931-0587, pp. 892-898, Published by IEEE, DOI: 10.1109/IV47402.2020.9304601, Available: https://ieeexplore.ieee.org/document/9304601.

[15] Yecheng Lyu, Lin Bai and Xinming Huang, "Road Segmentation using CNN and Distributed LSTM", in *Proceedings of the 2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, Sapporo, Japan, 26-29 May 2019, Print ISBN: 978-1-7281-0397-6, print ISSN: 2158-1525, pp. 1-5, Published by IEEE, DOI: 10.1109/ISCAS.2019.8702174, Available: https://ieeexplore.ieee.org/document/8702174.

[16] Yu Changqian, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu *et al.*, "Bisenet: Bilateral segmentation network for real-time semantic segmentation", in *Proceedings of the ECCV 2018: 15th European Conference*, Munich, Germany, 8-14 September 2018, ISBN: 978-3-030-01260-1, pp. 334 - 349, Published by Springer-Verlag, DOI: 10.1007/978-3-030-01261-8_20, Available: https://dl.acm.org/doi/10.1007/978-3-030-01261-8_20.

[17] Fritsch Jannik, Tobias Kuehnl and Andreas Geiger, "A new performance measure and evaluation benchmark for road detection algorithms", in *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, The Hague, Netherlands, 06-09 October 2013, E-ISBN: 978-1-4799-2914-6, Print ISSN: 2153-0009, E-ISSN: 2153-0017, pp. 1693-1700, Published by IEEE, DOI: 10.1109/ITSC.2013.6728473, Available: https://ieeexplore.ieee.org/document/6728473.

[18] Lin Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He and Piotr Dollár, "Focal loss for dense object detection", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 01 February 2020, E-ISSN: 1939-3539 , Print ISSN: 0162-8828, Vol. 42, No. 2, pp.318 - 327, Published by IEEE, DOI: 10.1109/TPAMI.2018.2858826, Available: https://ieeexplore.ieee.org/document/8417976.

[19] Buslaev Alexander, Iglovikov Vladimir I, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin *et al.*, "Albumentations: fast and flexible image augmentations", *Information*, 24 February 2020, ISSN: 2078-2489, Vol. 11, No. 2, Published by MDPI, DOI: 10.3390/info11020125, Available: https://www.mdpi.com/2078-2489/11/2/125.

[20] Kühnl Tobias, Franz Kummert and Jannik Fritsch, "Spatial ray features for real-time ego-lane extraction", in *Proceedings of the 2012 15th International IEEE Conference on Intelligent Transportation Systems*, Anchorage, AK, USA, 16-19 September 2012, E-ISBN: 978-1-4673-3063-3, Print ISBN: 978-1-4673-3064-0, Online ISBN: 978-1-4673-3062-6, pp. 288-293, Published by IEEE, DOI: 10.1109/ITSC.2012.6338740, Available: https://ieeexplore.ieee.org/document/6338740.

[21] Teodoro Vincent Frémont and Denis Fernando Wolf, "Exploiting fully convolutional neural networks for fast road detection", in *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, 16-21 May 2016, E-ISBN: 978-1-4673-8026-3, pp. 3174-3179, Published by IEEE, DOI: 10.1109/ICRA.2016.7487486, Available: https://ieeexplore.ieee.org/document/7487486.

[22] Xiao Liang, Ruili Wang, Bin Dai, Yuqiang Fang, Daxue Liu *et al.*, "Hybrid conditional random field based camera-LIDAR fusion for road detection", *Information Sciences*, March 2018, E-ISSN: 0020-0255, Vol. 432, No. 2, pp.543-558, Published by IEEE, DOI: 10.1016/j.ins.2017.04.048, Available: https://ieeexplore.ieee.org/document/8417976 .

[23] Laddha Ankit, Kocamaz Mehmet Kemal, Luis E Navarro-Serment and Martial Hebert, "Map-supervised road detection", in *Proceedings of the 2016 IEEE Intelligent Vehicles Symposium (IV)*, Gothenburg, Sweden, 19-22 June 2016, E-ISBN: 978-1-5090-1821-5, Print ISBN: 978-1-5090-1822-2, pp. 118-123, Published by IEEE, DOI: 10.1109/IVS.2016.7535374, Available: https://ieeexplore.ieee.org/document/7535374.

[24] Oliveira Gabriel L, Wolfram Burgard and Thomas Brox, "Efficient deep models for monocular road segmentation", in *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, South Korea, 09-14 October 2016, E-ISSN: 2153-0866, E-ISBN: 978-1-5090-3762-9, Print ISBN: 978-1-5090-3763-6, pp. 4885-4891, DOI: 10.1109/IROS.2016.7759717, Available: https://ieeexplore.ieee.org/document/7759717.

[25] Garnett Noa, Shai Silberstein, Shaul Oron, Ethan Fetaya, Uri Verner *et al.*, "Real-time category-based and general obstacle detection for autonomous driving", in *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, 22-29 October 2017, E-ISSN: 2473-9944, Print ISSN: 1931-0587, E-ISBN: 978-1-5386-1034-3, print ISSN: 978-1-5386-1035-0, pp. 198-205, Published by IEEE, DOI: 10.1109/ICCVW.2017.32, Available: https://ieeexplore.ieee.org/document/8265242.

[26] Chen Zhe and Zijing Chen, "Rbnet: A deep neural network for unified road and road boundary detection", in *Proceedings of the Neural Information Processing: 24th International Conference, ICONIP 2017*, Guangzhou, China, 14-18 November 2017, Online ISBN: 978-3-319-70087-8, Print ISBN: 978-3-319-70086-1, pp. 677-687, Published by Springer, DOI: 10.1007/978-3-319-70087-8_70, Available: https://link.springer.com/chapter/10.1007/978-3-319-70087-8_70.

[27] Caltagirone Luca, Mauro Bellone, Lennart Svensson and Mattias Wahde, "LIDAR–camera fusion for road detection using fully convolutional neural networks", *Robotics and Autonomous Systems*, January 2019, ISSN: 0921-8890, Vol. 111, pp. 125-131, Published by Elsiever B.V., DOI: 10.1016/j.robot.2018.11.002, Available: https://www.sciencedirect.com/science/article/abs/pii/S0921889018300496 .

[28] Zhao Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang and Jiaya Jia, "Pyramid scene parsing network", in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21-26 July 2017, Print ISSN: 1063-6919, E-ISBN: 2380-7504, E-ISBN: 978-1-5386-0457-1, Print ISBN: 978-1-5386-0458-8, pp. 6230-6239, Published by IEEE, DOI: 10.1109/CVPR.2017.660, Available: https://ieeexplore.ieee.org/document/8100143.

[29] Kontschieder Peter, Samuel Rota Bulo, Horst Bischof and Marcello Pelillo, "Structured class-labels in random forests for semantic image labelling", in *Proceedings of the 2011 International Conference on Computer Vision*, Barcelona, Spain, 6-13 November 2011, Print ISSN: 1550-5499, E-ISBN: 2380-7504, E-ISBN:978-1-4577-1102-2, Print ISBN:978-1-4577-1101-5, pp. 2190-2197, Published by IEEE, DOI: 10.1109/ICCV.2011.6126496, Available: https://ieeexplore.ieee.org/abstract/document/6126496.

[30] Chen Xiaozhi, Kaustav Kundu, Yukun Zhu, Andrew G Berneshawi, Huimin Ma *et al.*, "3D object proposals for accurate object class detection", *Advances in Neural Information Processing Systems*, 1 May 2018, E-ISSN: 1939-3539 , Print ISSN: 0162-8828, Vol. 40, No. 5, pp. 1259 - 1272, Published by IEEE, DOI: 10.1109/TPAMI.2017.2706685, Available: https://ieeexplore.ieee.org/document/7932113.

[31] Shinzato Patrick Y, Denis F Wolf and Christoph Stiller, "Road terrain detection: Avoiding common obstacle detection assumptions using sensor fusion", in *Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings*, Dearborn, MI, USA, 08-11 June 2014, Print ISSN: 1931-0587, E-ISBN: 978-1-4799-3638-0, pp. 687-692, Published by IEEE, DOI: 10.1109/IVS.2014.6856454, Available: https://ieeexplore.ieee.org/document/6856454.

[32] Vitor Giovani B, Alessandro C Victorino and Janito V Ferreira, "Comprehensive performance analysis of road detection algorithms using the common urban Kitti-road benchmark", in *Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings*, Dearborn, MI, USA, 08-11 June 2014, E-ISBN: 978-1-4799-3638-0, Print ISSN: 1931-0587, pp. 19-24, Published by IEEE, DOI: 10.1109/IVS.2014.6856616, Available: https://ieeexplore.ieee.org/document/6856616.