AUSUD BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ JITSA

Research Article

U-Net-Based detection of road and lane markings from high-resolution ımages

Oğuzhan KATAR[1,*]
[1]Software Engineering, Technology of Faculty, Fırat University, Elazığ, Turkey
*Correspondence: okatar@firat.edu.tr

DOI: 10.51513/jitsa.1172992

Abstract: With technological advancements in hardware, many autonomous systems are now utilized in daily life. Autonomous vehicles, designed for safe transportation, rely on sensors and cameras for dynamic environmental control. Accurate detection and classification of road and lane markings from high-resolution images are crucial for ensuring safe and efficient operation of autonomous vehicles, enhancing their performance in dynamic environmental control, and ensuring secure travel. To process the image data received from their cameras and transform it into meaningful information, artificial intelligence-based approaches are highly effective. In this study, ResNet-based U-Net method is proposed that can automatically detect and classify areas of road and lane markings from high-resolution images. A publicly available dataset was customized for the model's training, validation, and testing phases. The pre-processing phase designed to include high-resolution images in the training of segmentation model is explained. Dataset samples are split into 70% training, 20% validation, and 10% testing. The early stopping function was implemented during the training phases for four different U-Net models with ResNet architectures in the encoder network. The numerical data of the training and validation phases, which were carried out in accordance with the multi-class semantic segmentation method, were shared. The test phase revealed that the ResNet-101 U-Net model achieved the highest mean Intersection over Union (mIoU) value, with a rate of 86.75%. With this method, the classification and detection of road and lane markings areas can help the dynamic environment control of autonomous vehicles.

Keywords: Autonomous vehicles, semantic segmentation, U-Net, auto-detection

Yüksek çözünürlüklü görüntülerden yol ve şerit işaretlerinin U-Net tabanlı tespiti

Özet: Donanımdaki teknolojik gelişmelerle birlikte artık günlük hayatta birçok otonom sistem kullanılmaktadır. Güvenli ulaşım için tasarlanan otonom araçlar, dinamik çevre kontrolü için sensörlere ve kameralara güveniyor. Yüksek çözünürlüklü görüntülerden yol ve şerit işaretlerinin doğru tespiti ve sınıflandırılması, otonom araçların güvenli ve verimli çalışmasını sağlamak, dinamik çevre kontrolündeki performanslarını artırmak ve güvenli seyahat sağlamak için çok önemlidir. Kameralarından alınan görüntü verilerini işlemek ve anlamlı bilgilere dönüştürmek için yapay zekaya dayalı yaklaşımlar oldukça etkilidir. Bu çalışmada, yüksek çözünürlüklü görüntülerden yol ve şerit işaretlerinin alanlarını otomatik olarak algılayabilen ve sınıflandırabilen ResNet tabanlı U-Net yöntemi önerilmiştir. Modelin eğitim, doğrulama ve test aşamaları için herkese açık bir veri kümesi özelleştirildi. Segmentasyon modelinin eğitiminde yüksek çözünürlüklü görüntüleri içerecek şekilde tasarlanan ön işleme aşaması açıklanmaktadır. Veri kümesi örnekleri% 70 eğitim,% 20 doğrulama ve% 10 teste bölünmüştür. Erken durdurma işlevi, kodlayıcı ağında ResNet mimarisine sahip dört farklı U-Net modeli için eğitim aşamalarında uygulandı. Çok sınıflı anlamsal segmentasyon yöntemine uygun olarak gerçekleştirilen eğitim ve validasyon aşamalarının sayısal verileri paylaşıldı. Test aşaması, ResNet-101 U-Net modelinin% 86,75 oranıyla Birleşim üzerinden en yüksek ortalama Kesişme (mIoU) değerine ulaştığını ortaya koydu. Bu yöntemle yol ve şerit işaretleme alanlarının sınıflandırılması ve tespiti, otonom araçların dinamik ortam kontrolüne yardımcı olabilir.

Anahtar Kelimeler: Otonom araçlar, anlamsal bölütleme, U-Net, otomatik tespit

ORCID: 0000-0002-5628-3543
Received 09.09.2022; accepted 24.04.2023
Peer review under responsibility of Bandirma Onyedi Eylul University.

1. Introduction

The transportation story, which started with the invention of the wheel in 3500 BC, has continued with various vehicles to today (Kaushal et al. 2012). With the industrial revolution, steam vehicles pioneered a great change in the field of transportation. The development process, which continues with internal combustion engines, has gained great momentum today with high-tech vehicles such as autonomous vehicles, flying cars, and high-speed trains (Winner and Wachenfeld, 2016). Autonomous vehicles are vehicles that can travel without human intervention. Autonomous vehicles, which automatically detect roads, lanes, and objects by processing data received from many sensors and cameras with the help of artificial intelligence, perform dynamic environmental control during their travels (Milakis, 2019). Due to the size and complexity of the data processed per unit of second, computers with high processing capacity are used in autonomous vehicles.

With the development of hardware technology, artificial intelligence has become popular. Artificial intelligence-based systems, which are used in many areas such as health, education, defense industry, and transportation, can make decisions without the need for any human intervention (Zhang and Lu, 2021). The main purpose of the development of such systems is to minimize the human factor. Many negatives occur in the transportation sector due to fatigue, carelessness, and similar human factors. Mistakes such as lane crossing and incorrect overtaking can cause the death of drivers and other people inside the vehicle (Stanton and Salmon, 2009). The widespread use of autonomous vehicles, thanks to artificial intelligence-based solutions, can help make travel more reliable and efficient. However, infrastructure and superstructure needs on these roads must be met (Henschke, 2020). Because one of the most important parameters for autonomous vehicles is road images taken with the help of cameras. Autonomous systems cannot provide the expected efficiency if the lane markings are below the expected standards or if they are absent.

The image data from the cameras alone is not enough for autonomous vehicles to decide. These data need to be processed and transformed into information. Classical methods and artificial intelligence-based solutions are frequently used for the processing of image data. Deng and Wu (2018) proposed a lane detection method based on the Hough Transform. Road images are grayed out by pre-processing steps. Canny edge detection operator is used to obtain road marks information from grayscale images. The Hough transform based on pole angle and pole radius constraints is used to obtain the double edges of the road marks. The results show that mark detection is fast and accurate using the double-edged subtraction method. He et al. (2004) proposed a road area detection algorithm based on color images. This algorithm consists of two modules, which estimate the boundaries according to the density image and detect the path areas from the full-color image. In the first module, using the color components of the road surfaces, the mean and variance of the Gaussian distribution for the left and right road boundaries are calculated. The second module effectively extracts the path area and reinforces the boundaries that best fit the path extraction result. Tests using real road images confirmed the effectiveness of the proposed method. Yadav et al. (2017) proposed a road segmentation approach constructed by a combination of deep convolutional neural networks with a color line model based on a conditional random field (CRF) framework. In this technique, convolutional neural networks learn the texture of the road, while the pattern of colored lines allows adaptation to changing lighting conditions. The researchers, who completed the test processes on publicly available datasets, examined precision and recall values as performance evaluation metrics, as they were only interested in the detection of paths. The method they proposed reached 93.31% precision and 94.99% recall. Dewangan and Sahu (2021) proposed a deep learning-based approach to segment road and non-road areas. This approach uses U-Net, SegNet, and FCN-32 models with fully convolutional network architecture. Experimental studies were carried out on a publicly available dataset. This dataset contains 101 images with a resolution of 960×720 pixels (px). The number of images was increased to 909 using various data augmentation techniques. Models were trained under the same conditions for 100 iterations. When the test predictions were examined, it was stated that the U-Net model reached an average IoU value of 94% and produced better results compared to other models. Li et al. (2021) proposed a model called Lane-DeepLab based on semantic segmentation to detect multi-class lane lines in unmanned driving scenarios. The proposed model is based on the DeepLabv3+ architecture. The encoder-decoder network of the model has been redesigned to better detect lane lines. The model is trained for 25 epochs using crossentropy loss and a learning rate of 0.005. As a result of the tests performed on publicly available dataset samples, it was observed that the model reached an average IoU value of 79.23%. The labelling processes carried out to collect real images of traffic and to indicate road areas in these images are very costly in terms of time and resources. For this reason, the number of publicly available datasets for related studies is limited. Among the publicly available datasets used in studies in the literature are datasets such as Cityscapes, KITTI, and CamVid (Cordts et al, 2016; Geiger et al, 2013; Brostow et al, 2009).

In the existing literature, studies on road and lane detection can be classified into two categories: those focused on determining road areas and those focused on determining lane marking areas. However, the number of studies that can simultaneously detect both road and lane markings is limited. A major limitation of such studies, which require a multi-class segmentation process, is the need for input images to be at a fixed resolution. Nevertheless, given the diversity of camera equipment used today, limiting the image resolution reduces the generality and scope of the studies. To address this issue, this study proposes a method for detecting road and lane markings on high-resolution images on a pixel-by-pixel basis without any fixed resolution requirement. The proposed method aims to achieve high performance with minimal training resources by utilizing pre-trained convolutional neural networks in the encoder network of the U-Net architecture. The rest of this paper is organized as follows. Section 2 contains information about the method proposed for this study, the used dataset, the deep learning model, and the performance metrics used in semantic segmentation studies. The numerical values obtained during the training phase of the model and the test outputs are given in Section 3. The conclusion part of the study is in Section 4.

2. Material and Methods

A system has been designed for automatic detection of road and lane markings areas by the deep learning model on the images taken from the cameras placed in the vehicles. The block diagram of this system is given in Figure 1.

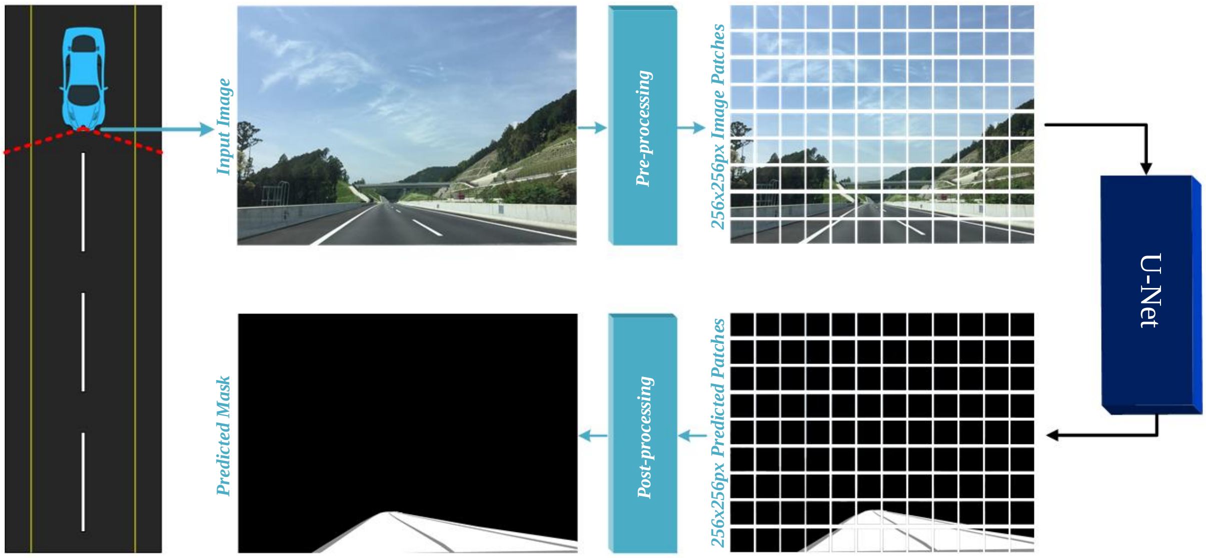


Figure 1. A block representation of the proposed method in this study.

In the proposed method, the image giving as input to the U-Net model is segmented on a pixel-by-pixel basis in the output layer. The U-Net architecture is equipped with skip connections between the contracting and expanding paths, enabling the network to retain spatial information during the segmentation process. This feature is particularly beneficial for detecting road and lane markings, as these markings may exhibit intricate shapes and textures necessitating meticulous segmentation. Additionally, the U-Net architecture possesses a relatively low parameter count compared to alternative segmentation architectures, resulting in computational efficiency and suitability for real-time applications. As a result, it is ensured that driverless vehicles can automatically detect road and lane markings areas. In such systems, which aim to minimize the human factor, a zero margin of error is expected. Therefore, deep learning models with high success rates are indispensable. One of the most critical components in achieving deep models with high success rates is the datasets used for model training, validation, and testing. Unlike classification studies, segmentation studies require mask images in addition to the original images, which denote the areas containing meaningful information in the original images. Creating mask images manually for high-resolution images is very costly. For this reason, the number of publicly available datasets is limited.

2.1. Dataset

The Mapillary Vistas dataset was used in the training, validation, and testing stages of this study (Neuhold et al., 2017). The dataset contains high-resolution images. The dataset includes traffic images captured from all over the world in different weather, season, and daylight conditions. These images were captured from different imaging devices, including mobile phones, tablets, action cameras, and professional cameras. In segmentation studies, the objects in the images are distinguished from the others due to their pixel values. For this reason, labeling and mask images were created for high-resolution dataset samples by the researchers. Labeling is performed in a dense and fine-grained style by using polygons for delineating individual objects. In the labeling process, classes with different pixel values are defined for 66 objects. These pixel classes and their distribution in the dataset are shown in Figure 2.

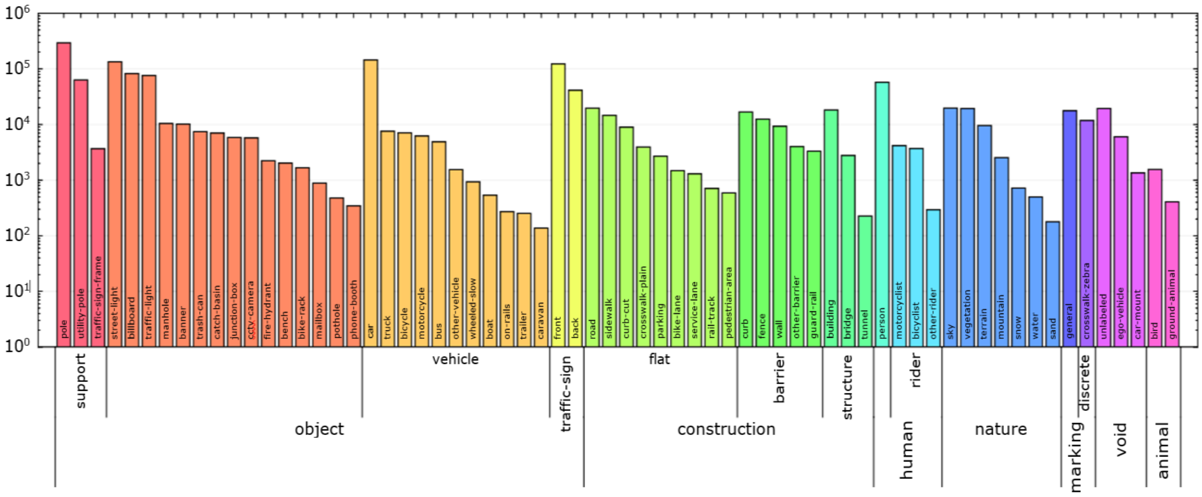


Figure 2. Class distribution in the dataset (Neuhold et al., 2017).

The dataset samples do not have a fixed resolution value. The images in the dataset have resolution values in the range of 640×480px and 6528×3680px. Randomly selected samples of the original images and mask images from the dataset are shown in Figure 3.

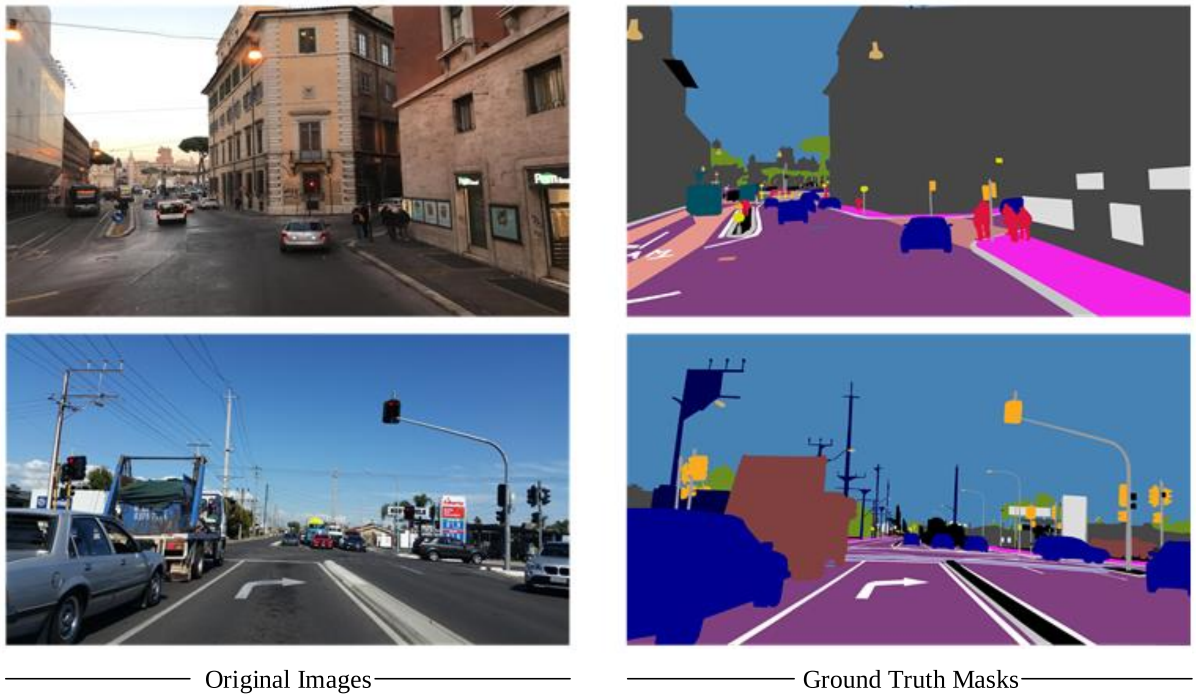


Figure 3. The dataset samples.

2.2. Pre-processing

There are pixel values corresponding to 66 different objects in the mask images. In this study, since it is desired to detect only road and lane markings, the pixel values of other objects should be removed from the mask images. With the help of the pixel-based value change process, the road areas are assigned as ‘1’, the lane markings areas as ‘2’ and the pixel values of all other objects are assigned as ‘0’. In Figure 4, the result obtained after applying pixel-based value change to the ground truth mask image is given.

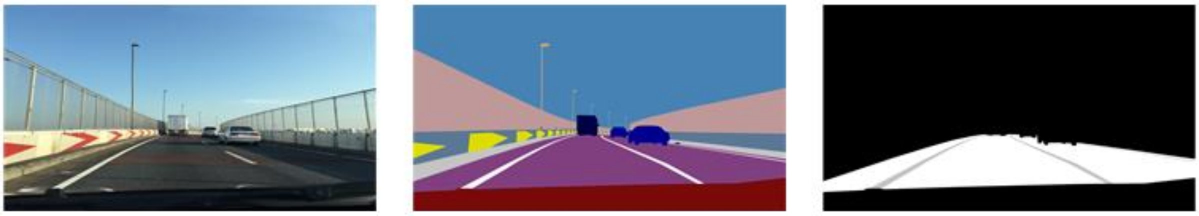

(a) Original Image (b) Ground Truth Mask (c) Pre-processed Mask

Figure 4. Pixel-based value changing method.

All of the mask images have been adapted to the multi-class semantic segmentation concept thanks to the pixel-based value change process. Mask images are converted into 8-bit single-channel and '.png' data format. In this way, the area occupied by the mask images in the memory has been reduced. Since the resolutions of the images in the dataset are high and unequal, they cannot be used directly in model training. There are hardware limitations, and the input size of the model must be a constant value. To overcome this problem, there are two alternatives, resizing and image cropping. The loss of information experienced when resizing high-resolution images is enormous. For this reason, resizing is not preferred in semantic segmentation studies. The default input size of the U-Net model is 256×256px. Thanks to the image cropping method, sub-images can be created equal to the default input size of the U-Net model. Starting from the first pixel, the first sub-image is obtained, covering 256 frames on the x-axis and 256 frames on the y-axis. This process is then repeated, advancing until it covers the last pixel of the image. If the size of the original image is not an exact multiple of 256×256, the padding method is used. In this way, high-resolution images can be included in model training without loss of information. The specified method is also applied for mask images. In the sub-mask images obtained, images that do not contain any road or lane markings areas may occur. Therefore, images with a pixel value of only '0' and which do not contain any meaningful objects for our study may cause problems in model training. To avoid these disadvantages, mask images consisting of only '0' and their corresponding original images were filtered out and were not included in the training, validation, and testing stages of the model. The flowchart of the proposed cropping and filtering approach is shown in Figure 5.

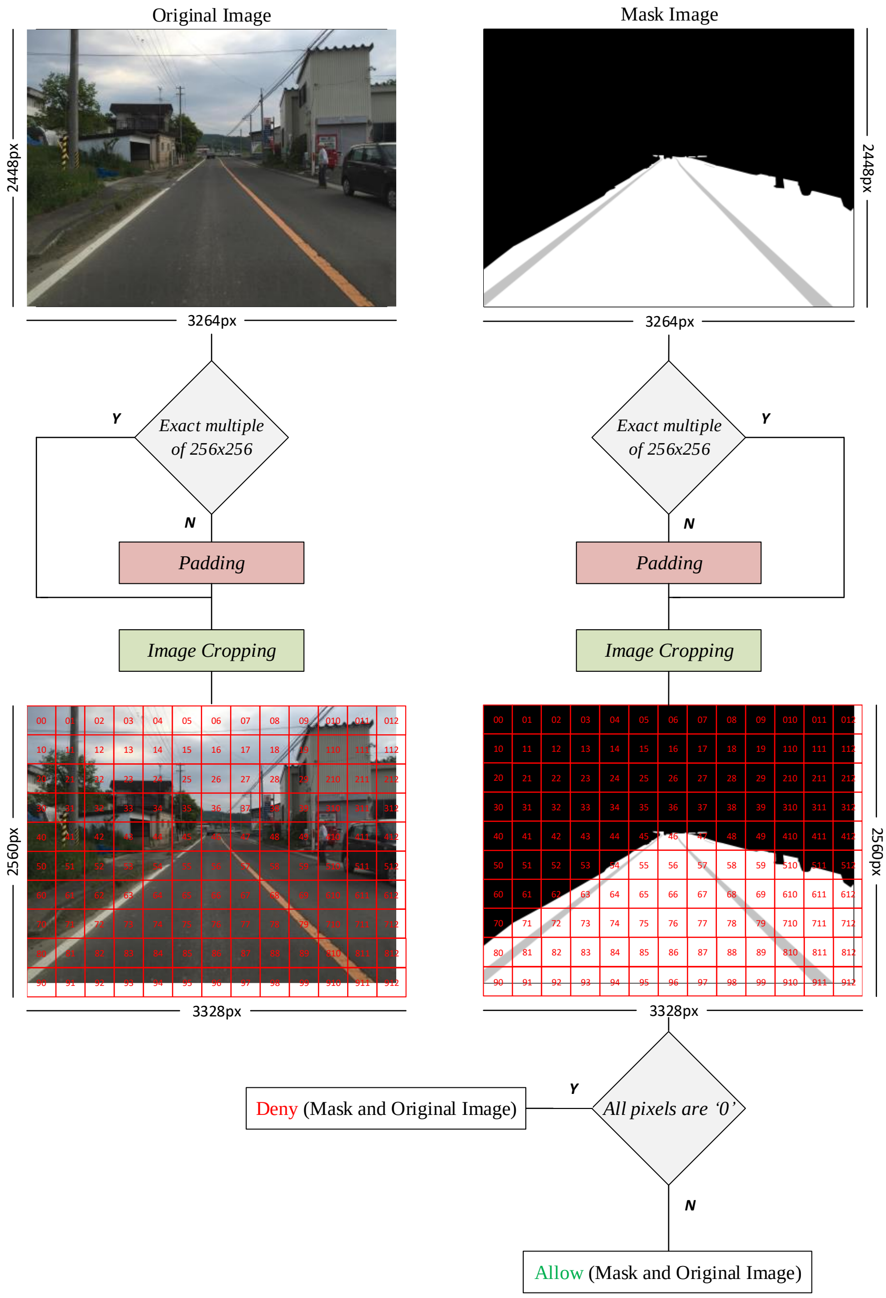


Figure 5. The proposed cropping and filtering approach.

2.3. U-Net

The U-Net is an architecture for convolutional neural networks that was initially introduced for the purpose of performing image segmentation tasks, wherein the objective is to assign a class label to each pixel in an image (Ronneberger et al., 2015). The U-Net architecture derives its name from its "U" shape.

It comprises two main components: an encoding path and a decoding path. A block representation of the U-Net architecture is given in Figure 6.

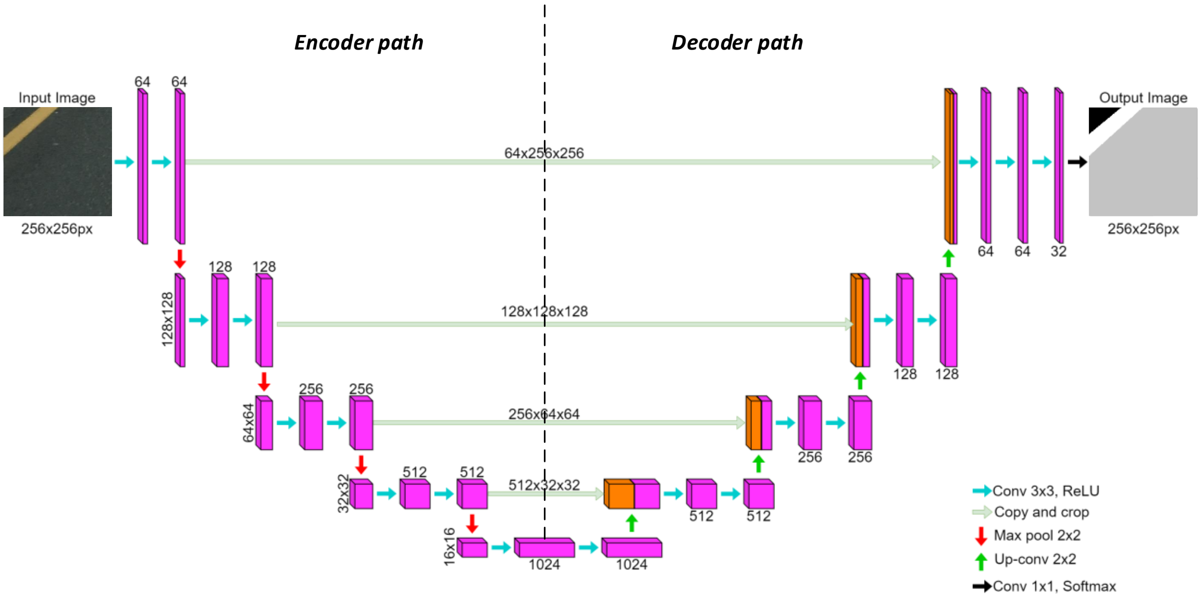


Figure 6. The U-Net architecture.

The encoder network comprises several layers, each of which performs a specific function in the feature extraction process. The first layer is typically a convolutional layer, which applies a set of filters to the input image to produce a set of feature maps. The filters in this layer are designed to capture low-level features, such as edges and corners, that are present in the input image. In addition to the convolutional layers, the encoder network also incorporates max pooling operations. Max pooling is a downsampling technique that reduces the spatial dimensions of the feature maps, while retaining their important features. Max pooling is typically applied after each convolutional layer, and reduces the spatial dimensions of the feature maps by a factor of two. This results in a smaller feature map with fewer parameters, which speeds up the computation and reduces the risk of overfitting. The output of the encoder network is a set of high-level feature maps that capture the most important features of the input image. These feature maps are then passed to the decoder network, which forms the expanding path of the U-Net model.

The decoder network is an essential component of a segmentation model that recovers the spatial dimensions of the feature maps created by the encoder network and generates a segmentation map of the input image. It achieves this through transposed convolutional layers that increase the spatial dimensions of the feature maps. The first layer is a transposed convolutional layer, which applies a set of filters to the encoder network's feature maps and produces new feature maps with a larger spatial dimension. Subsequent layers apply filters to the previous layer's feature maps, capturing increasingly local features of the input image. The decoder network includes skip connections that combine the encoder and decoder feature maps to access both the local and global features of the input image. The decoder network's output is a segmentation map, which is a pixel-wise classification of the input image, assigning each pixel a label corresponding to an object class.

2.4. Post-processing

The input layer resolutions of the deep learning models must be a fixed value. For this reason, it is not possible to input an image with a different resolution to the relevant models. However, considering the existence of different technologies, devices, and standards used in daily life, it is not a logical approach to expect only a predefined fixed value as input. To overcome all these difficulties, it is necessary to divide the images into 256×256px pieces and give them as input to the model with their names according to the row-column structure. The 256×256px images obtained at the output of the deep learning model should be combined sequentially. In this way, the detection of road and lane markings areas on high-resolution images can be performed automatically without being subject to limitations due to any resolution value. An approach to obtaining the predicted mask corresponding to the original image by combining the 256×256px sub-images predicted by the U-Net model is shown in Figure 7.

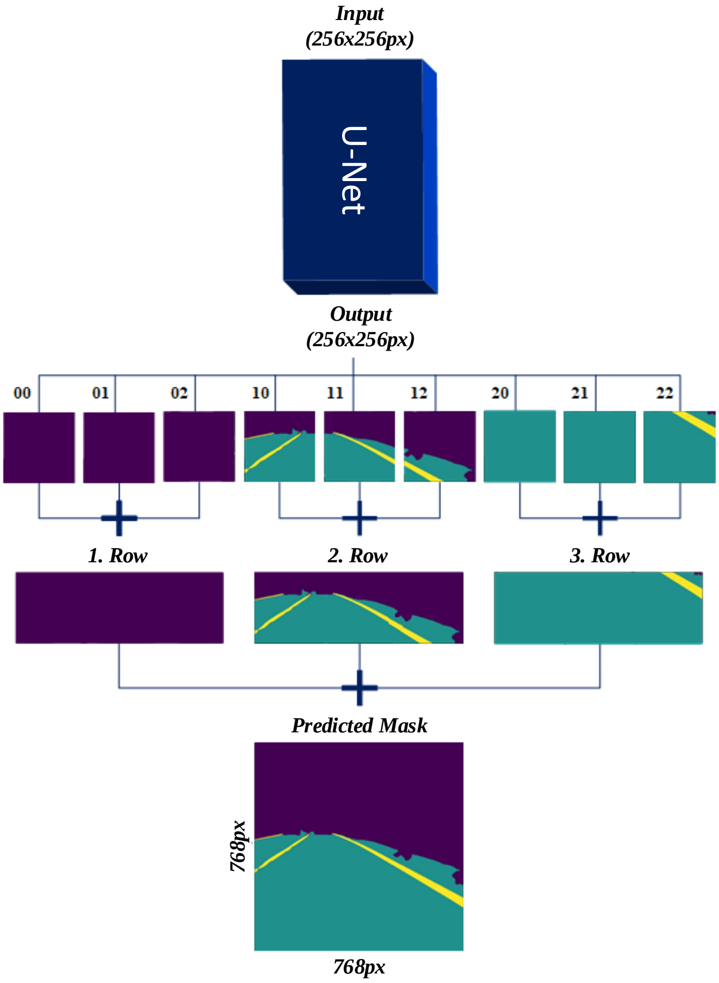


Figure 7. An approach to obtaining the predicted mask.

2.5. Performance metrics

Semantic segmentation is a computer vision technique that involves partitioning an image into multiple segments, each of which corresponds to a specific object or region within the image. In the context of semantic segmentation, each pixel in the predicted segmentation mask can be classified as either a True Positive (TP), False Positive (FP), True Negative (TN), or False Negative (FN) based on its correspondence with the ground truth segmentation mask. Table 1 provides a summary of these terms and the corresponding situations they represent.

Table 1. Summary of performance evaluation terms

Terms	Description
TP	A pixel is considered a TP if it is correctly classified as belonging to a specific object class by the predicted segmentation mask, and this classification is also correct in the ground truth segmentation mask.
FP	A pixel is considered a FP if it is incorrectly classified as belonging to a specific object class by the predicted segmentation mask, but this classification is not present in the ground truth segmentation mask.

TN	A pixel is considered a TN if it is correctly classified as not belonging to a specific object class by the predicted segmentation mask, and this classification is also correct in the ground truth segmentation mask.
FN	A pixel is considered a FN if it is incorrectly classified as not belonging to a specific object class by the predicted segmentation mask, but this classification is present in the ground truth segmentation mask.

These terms are used to compare the predicted segmentation masks generated by the model with the ground truth masks provided by human annotators. Some commonly used metrics for evaluating the performance of semantic segmentation models are as follows.

- Pixel Accuracy (PA) measures the proportion of correctly classified pixels in the image. It is a simple and intuitive metric that provides a quick estimate of the segmentation model's overall accuracy, but it does not take into account the importance of different object classes.
- Precision (Pre) and Recall (Rec) are commonly used in classification tasks but can also be applied to semantic segmentation. Precision measures the proportion of true positive pixels in the predicted segmentation mask, while recall measures the proportion of true positive pixels in the ground truth segmentation mask that are correctly identified by the predicted mask.
- Intersection over Union (IoU) measures the overlap between the predicted segmentation mask and the ground truth mask for each object class in the image. IoU is computed as the ratio of the intersection of the two masks to their union, and values range from 0 (no overlap) to 1 (perfect overlap).
- Mean Intersection over Union (mIoU) is the average IoU across all object classes in the image. It provides an overall estimate of the segmentation model's accuracy, taking into account the performance across all object classes.
- Dice Coefficient (DC) measures the similarity between the predicted and ground truth segmentation masks. DC values range from 0 to 1, with higher values indicating better segmentation accuracy. The DC is often used in conjunction with IoU to evaluate segmentation models

IoU and DC are two commonly used performance metrics in semantic segmentation studies. Both metrics are used to evaluate the similarity between the predicted segmentation mask and the ground truth mask. While both metrics measure the similarity between the predicted and ground truth masks, they differ in their focus. IoU gives more weight to the size of the intersection relative to the size of the union, whereas the DC gives equal weight to both the intersection and the total number of pixels in the masks. The IoU metric tends to penalize more than the DC. Therefore, the DC tends to measure closer to the average performance, whereas the IoU score measures closer to the worst-case performance. Due to this feature, the DC is effectively used as a loss parameter in model training.

The mathematical equations required for the calculation of performance metrics used in segmentation studies are presented in Table 2.

Table 2. The mathematical equations of performance metrics

Metrics	**Equation**
PA	$\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$
Pre	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$

Rec	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
IoU	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$
mIoU	$\frac{(\mathrm{IoU1}+\mathrm{IoU2}+\cdots+\mathrm{IoUn})}{\mathrm{n}}$
DC	$\frac{2\times \mathrm{TP}}{(\mathrm{TP}+\mathrm{FP})+(\mathrm{TP}+\mathrm{FN})}$

3. Experimental Results

The results of the U-Net models trained for automatic detection of road and lane markings areas in high-resolution images are presented in this section. In addition, various evaluation metric values of experimental findings are shown in the following sections.

3.1. Experimental setups

In this study, the images are randomly divided into 70% training, 20% validation, and 10% testing. The training set was used to train semantic segmentation model on a large amount of data, allowing it to learn to recognize and segment different classes of objects within images. Four different models were compiled for this study: U-Net with ResNet-18 encoder, U-Net with ResNet-34 encoder, U-Net with ResNet-50 encoder, and U-Net with ResNet-101 encoder. The models and their hyperparameters utilized in this study are presented in Table 3.

Table 3. The models and hyperparameters

Model	Input Size	Weights	Learning Rate	Optimizer	Loss Function	Total Parameters
ResNet-18 U-Net	256×256×3	ImageNet	0.0001	Adam	DiceLoss	14,340,860
ResNet-34 U-Net	256×256×3	ImageNet	0.0001	Adam	DiceLoss	24,456,444
ResNet-50 U-Net	256×256×3	ImageNet	0.0001	Adam	DiceLoss	32,561,404
ResNet-101 U-Net	256×256×3	ImageNet	0.0001	Adam	DiceLoss	51,605,756

The models underwent training and validation processes, utilizing an early stopping function. The primary objective of this function is to mitigate overfitting, which transpires when a model becomes excessively intricate and begins memorizing the training data instead of learning generalized patterns that can be extrapolated to new data. The function monitors the validation loss during the training process, and terminates the training process if the validation loss stops decreasing or starts increasing. The early stopping function is optimized to activate after 15 consecutive epochs, where the monitored loss value displays no decrease. Following the completion of the training and validation phases, the final weights of each model were recorded in the '.hdf5' format. Subsequently, the models with the recorded weights were tested using test images, and the optimal model was determined by evaluating the attained performance metrics. The schematic representation of the training, validation, and testing phases of the models is illustrated in Figure 8.

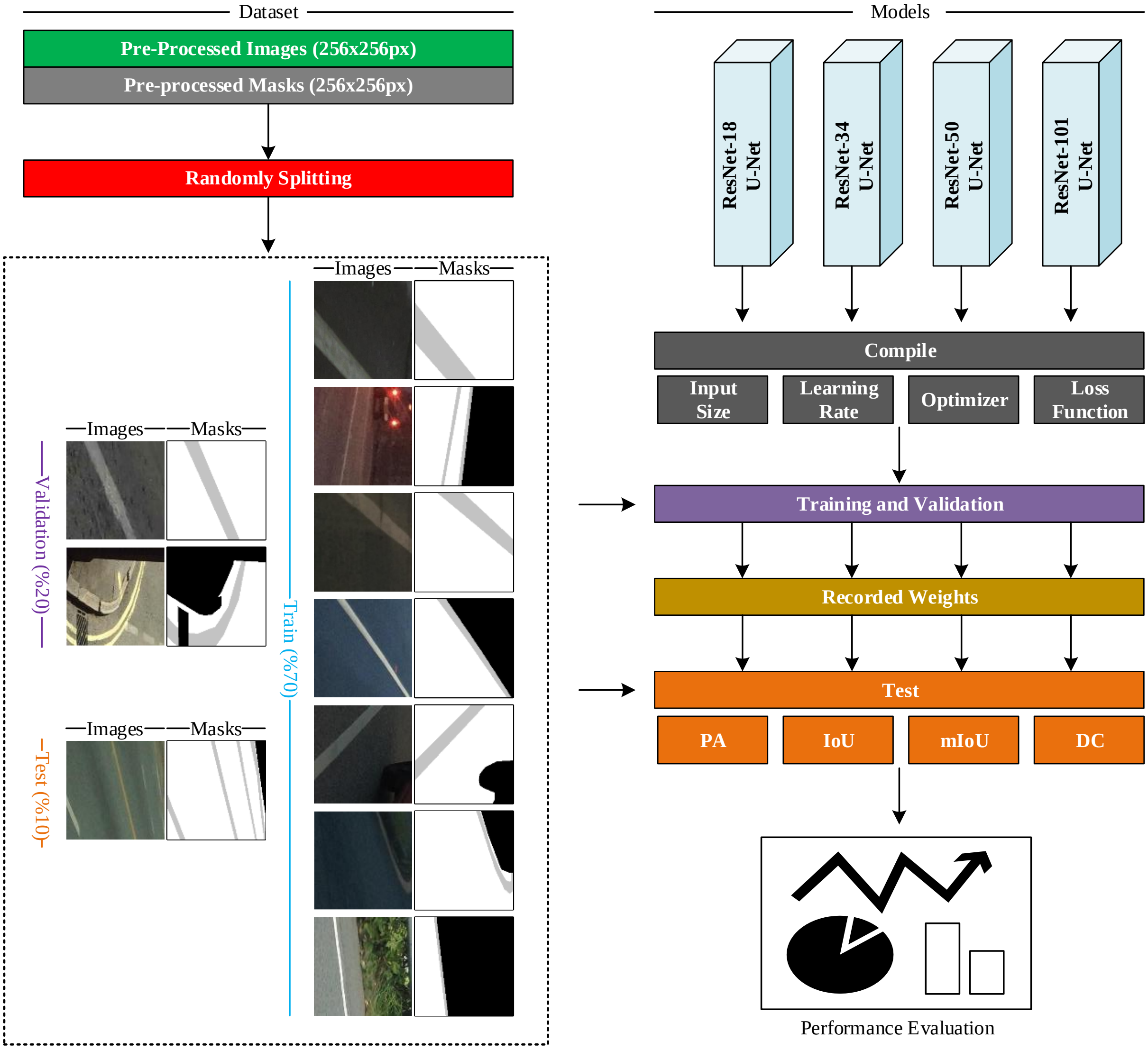
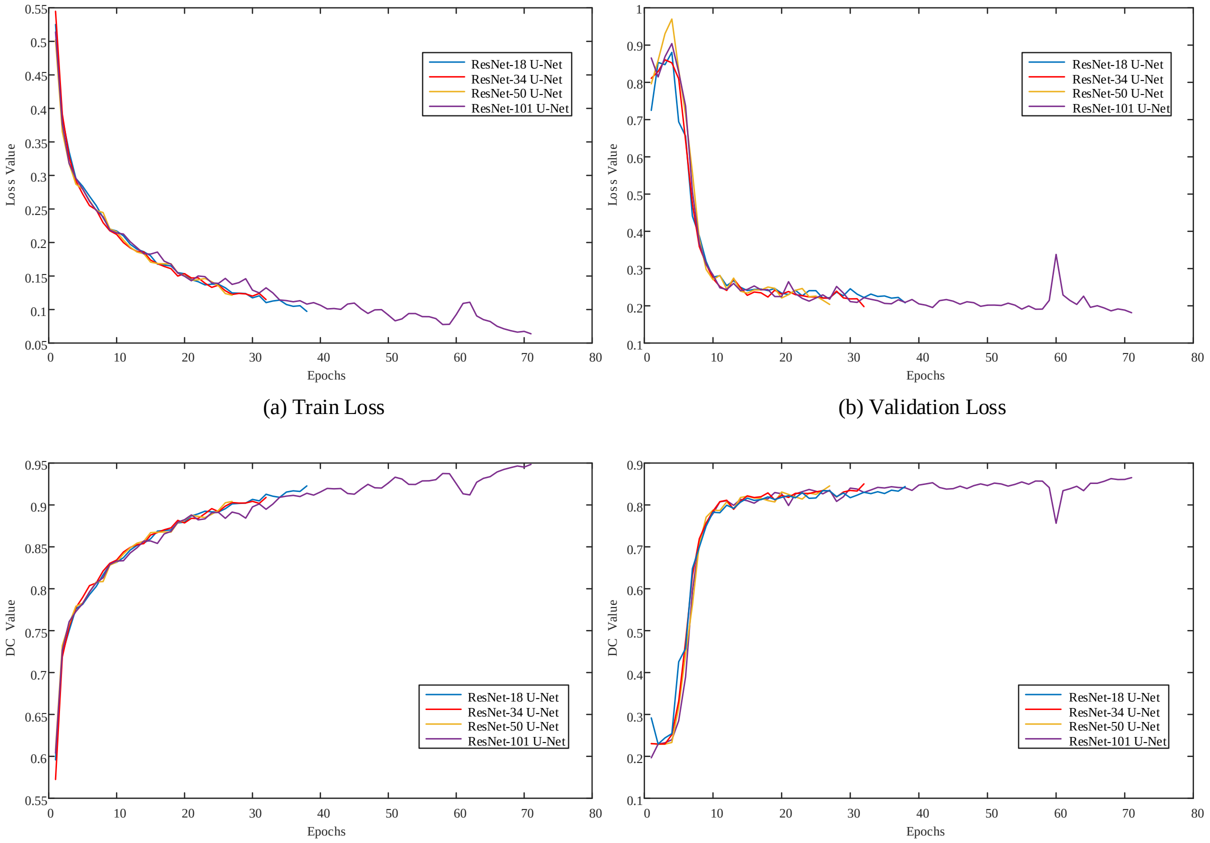


Figure 8. The schematic representation of the training, validation, and testing phases

3.2. Results

Accurate and efficient segmentation of road and lane markings is a critical task for the development of computer vision systems in the field of autonomous driving. To accomplish this task, we trained U-Net models with four different additional variants using the ResNet architecture. These backbones are widely recognized for their ability to extract meaningful features from images, which is essential for accurate segmentation. The performance graphs of the models obtained during the training and validation stages are given in Figure 9.



(a) Train Loss

(b) Validation Loss

(c) Train DC

(d) Validation DC

Figure 9. Performance Graphs (a) Train Loss, (b) Validation Loss, (c) Train DC, (d) Validation DC

Among all models, the ResNet-101 U-Net exhibited the lowest validation loss and the highest validation DC, indicating its superior performance in segmenting road and lane areas. The ResNet-101 U-Net model achieved a validation loss of 0.1715 and a validation DC of 0.8853, outperforming the other three models in terms of accuracy and robustness. To assess the effectiveness of the proposed method, we evaluated the four models on a separate set of test images that were not used for training or validation. The ResNet-18 U-Net model achieved a mIoU of 85.37%, while the ResNet-34 U-Net model achieved a slightly higher mIoU of 86.02%. The ResNet-50 U-Net model had an mIoU of 85.71%, and the ResNet-101 U-Net model achieved the highest mIoU value of 86.75%. Figure 10 presents the class-based IoU value distribution for each of these models on the test images.

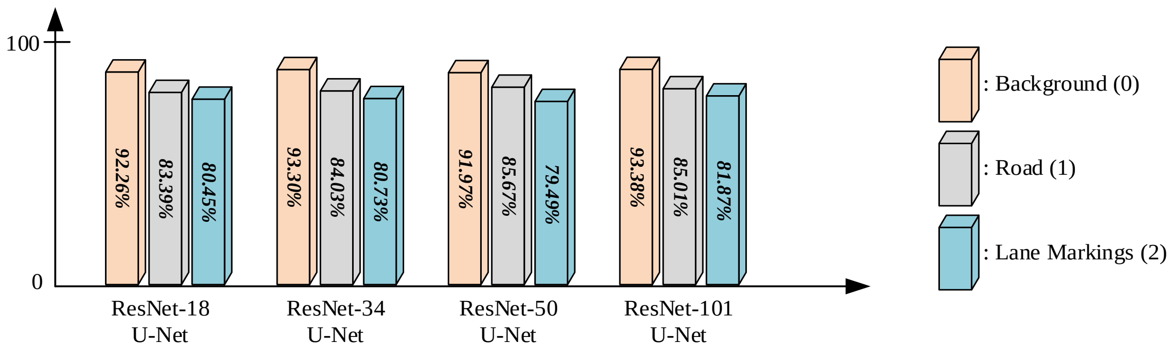
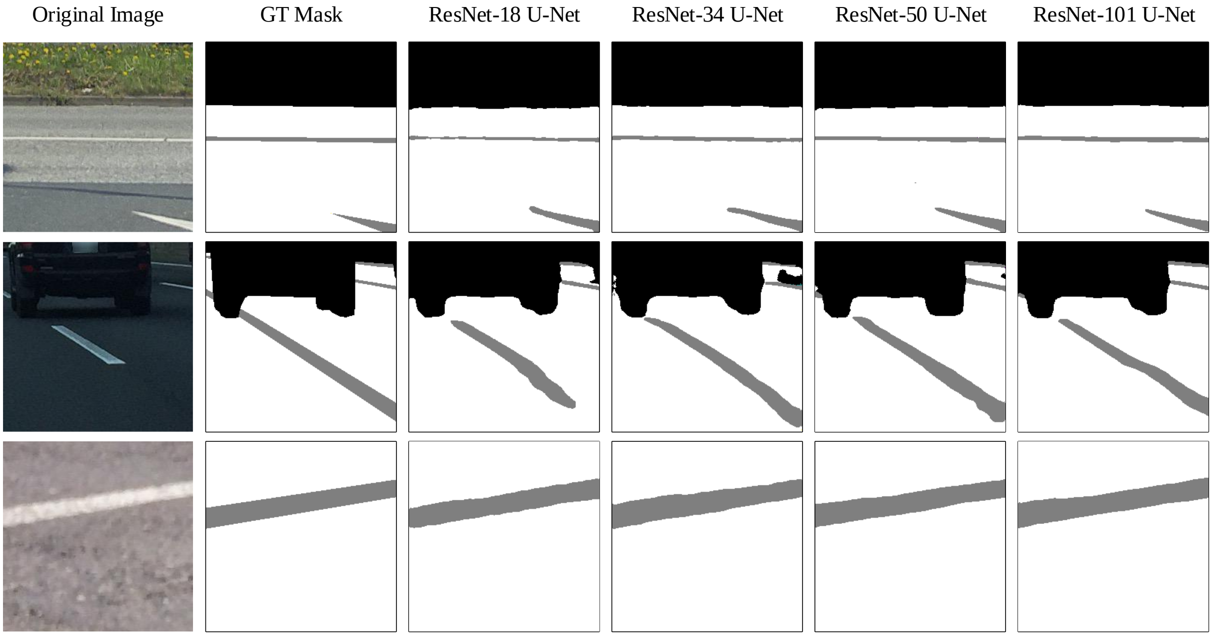


Figure 10. The class-based IoU value distribution.

Figure 11 shows the predicted segmentation mask for each model on randomly selected test images. By visually comparing these results, the segmentation performance of each model in real-world scenarios can be better understood.



Original Image GT Mask ResNet-18 U-Net ResNet-34 U-Net ResNet-50 U-Net ResNet-101 U-Net

Figure 11. The predicted segmentation mask on test images.

4. Discussion

The development of computer vision systems for autonomous vehicles is a rapidly growing research area. With advancements in hardware technology, there is an increasing number of studies focused on providing vision to autonomous systems. In this study, we propose a ResNet-101 U-Net model for accurately segmenting roads and lane markings from street-level images. The results obtained from our experiments indicate that the proposed model performs well in accurately segmenting both the road and lane markings. Furthermore, the resolution of the input image does not significantly impact the performance of the proposed method. Figure 12 illustrates the prediction made by the proposed model on test image with resolutions different from 256×256 pixels.

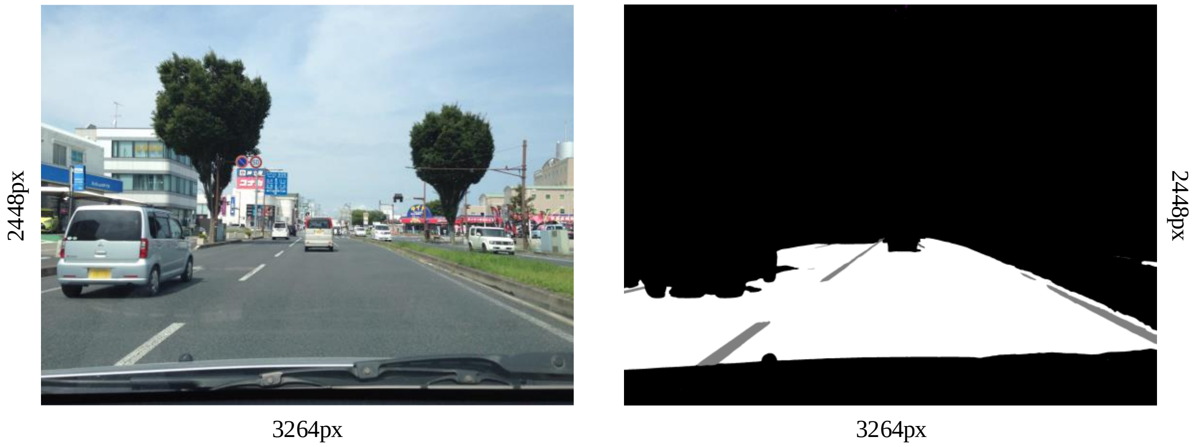


Figure 12. The predicted segmentation mask on high-resolution image.

To the best of our knowledge, this is one of the first studies to use this model for this task. Table 4 provides details for a hand-curated selection of research studies on computer vision for autonomous vehicles. Cheng et al. (2023) proposed a lane line detection algorithm based on instance segmentation. The proposed method optimizes the RepVgg-A0 network structure. Experimental results shows that the algorithm achieves an accuracy (Acc) value of 96.70% on the TuSimple dataset. Liu et al. (2022) proposed a Lane-GAN network for lane line detection that is robust to blurred images in complex road environments. Their model yielded a 96.56% Acc rate. Dewangan et al. (2021) proposed an approach based on U-Net for segmenting road and non-path areas from images. The experimental results, obtained using the Camvid dataset, show that U-Net outperforms other models with a score of 94.00% for both mIoU and DC. Das et al. (2021) proposed a method for road boundary estimation using deep learning-based semantic segmentation, without prior knowledge of road markings. The method employs a DeepLab architecture with different types of backbone networks and handles class imbalance using weighted loss contribution. The method is evaluated using the 'ICCV09DATA' dataset. The method achieved the Acc of 0.9596.

Table 4. The research studies on computer vision for autonomous vehicles.

Study	Dataset	Default Input Size	Number Of Classes	Method	Performance
Cheng et al. (2023)	TuSimple	368×640	2 (Line and background)	Custom Model	Acc=96.70%
Liu et al. (2022)	TuSimple and CULane	1280×720	2 (Line and background)	Lane-GAN	Acc=96.56%
Dewangan et al. (2021)	Camvid	960×720	2 (Road and non-road)	U-Net	mIoU=94.00%
Das et al. (2021)	ICCV09DATA	256×256	3 (Road, road boundary and background)	DeepLabV3+	Acc=95.96% Pre=94.53% Rec=93.69%
This study	Mapillary Vistas	-	3 (Road, Lane markings and background)	ResNet-101 U-Net	mIoU=86.75%

In the literature studies, binary segmentation studies were commonly conducted, such as distinguishing between road or background, and lane markings or background. As a result, high performance in these studies is typically achieved. However, a fixed input size is usually required in all proposed methods. When an image of a resolution other than this value is inputted, no action can be taken. In this study, an image cropping and merging method is proposed to enable segmentation without requiring a fixed resolution value.

The advantages of our method can be summarized as follows:

- The image cropping and merging method reduces resource usage.
- End-to-end multi-class segmentation can be performed without input size limitations.
- The pixel-based classification feature provides more precise detection of road and lane markings.
- The filtering function used during the training phase enables higher success rates to be achieved in a shorter amount of time.

The primary limitation of the proposed method is the high cost associated with creating new datasets for use in the training, validation, and testing phases of the model. In future studies, we intend to validate and enhance the model's generalization ability by utilizing a hybrid dataset obtained from public sources.

Contribution Rate Statement

All research and writing steps belong to the corresponding author.

Acknowledgment

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest Statement

No conflict of interest was declared by the author.

References

Kaushal, P., Vatsa, D., Gupta, S., and Raj, R. (2022). Historical Analysis of Wheel and Diving into Future of Wheel Made with Additive Manufacturing. *Recent Trends in Industrial and Production Engineering*, 95-106. doi:10.1007/978-981-16-3330-0_8

Winner, H., and Wachenfeld, W. (2016). Effects of autonomous driving on the vehicle concept. *Autonomous Driving*, 255-275. doi:10.1007/978-3-662-48847-8_13

Milakis, D. (2019). Long-term implications of automated vehicles: An introduction. *Transport Reviews*, 39(1), 1-8. doi:10.1080/01441647.2019.1545286

Zhang, C., and Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23, 100224. doi:10.1016/j.jii.2021.100224

Stanton, N. A., and Salmon, P. M. (2009). Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Safety Science*, 47(2), 227-237. doi:10.1016/j.ssci.2008.03.006

Henschke, A. (2020). Trust and resilient autonomous driving systems. *Ethics and Information Technology*, 22(1), 81-92. doi:10.1007/s10676-019-09517-y

Deng, G., and Wu, Y. (2018). Double lane line edge detection method based on constraint conditions hough transform. *17th International symposium on distributed computing and applications for business engineering and science (DCABES)*, 107-110.

He, Y., Wang, H., and Zhang, B. (2004). Color-based road detection in urban traffic scenes. *IEEE Transactions on intelligent transportation systems*, 5(4), 309-318.

Yadav, S., Patra, S., Arora, C., and Banerjee, S. (2017). Deep CNN with color lines model for unmarked road segmentation. *2017 IEEE International Conference on Image Processing (ICIP)*, 585-589.

Dewangan, D. K., and Sahu, S. P. (2021). Road detection using semantic segmentation-based convolutional neural network for intelligent vehicle system. *Data engineering and communication technology*, 629-637. doi:10.1007/978-981-16-0081-4_63

Li, J., Jiang, F., Yang, J., Kong, B., Gogate, M., Dashtipour, K., and Hussain, A. (2021). Lane-deeplab: Lane semantic segmentation in automatic driving scenarios for high-definition maps. *Neurocomputing*, 465, 15-25. doi:10.1016/j.neucom.2021.08.105

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. *IEEE conference on computer vision and pattern recognition*, 3213-3223.

Geiger, A., Lenz, P., Stiller, C., and Urtasun, R. (2013). Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11), 1231-1237. doi:10.1177/0278364913491297

Brostow, G. J., Fauqueur, J., and Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2), 88-97. doi:10.1016/j.patrec.2008.04.005

Neuhold, G., Ollmann, T., Rota Bulo, S., and Kontschieder, P. (2017). The mapillary vistas dataset for semantic understanding of street scenes. *IEEE international conference on computer vision*, 4990-4999.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, 234-241.

Cheng, W., Wang, X., and Mao, B. (2023). Research on Lane Line Detection Algorithm Based on Instance Segmentation. *Sensors*, 23(2), 789. doi: 10.3390/s23020789

Liu, Y., Wang, J., Li, Y., Li, C., and Zhang, W. (2022). Lane-GAN: A Robust Lane Detection Network for Driver Assistance System in High Speed and Complex Road Conditions. *Micromachines*, 13(5), 716. doi: 10.3390/mi13050716

Dewangan, D. K., and Sahu, S. P. (2021). Road detection using semantic segmentation-based convolutional neural network for intelligent vehicle system. *Data Engineering and Communication Technology: Proceedings of ICDECT*, 629-637.

Das, S., Fime, A. A., Siddique, N., and Hashem, M. M. A. (2021). Estimation of road boundary for intelligent vehicles based on deepLabV3+ architecture. *IEEE Access*, 9, 121060-121075.