Title

MSc Research Project
Data Analytics

Kapil Lalwani
Student ID: x21123292

School of Computing
National College of Ireland

Supervisor: Prof. Jorge Basilio

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Kapil Lalwani
Student ID:	x21123292
Programme:	MSc in Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Prof. Jorge Basilio
Submission Due Date:	15/12/2022
Project Title:	Intelligent Transportation System for Lane Detection using Focus-based Instance Segmentation
Word Count:	5174
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	30th January 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Intelligent Transportation System for Lane Detection using Focus-based Instance Segmentation

Kapil Lalwani
x21123292

Abstract

Lane detection is an essential feature for developing intelligent transportation systems that can prevent road accidents due to undetected lane changes. The study will propose a method for reducing road accidents by using computer vision robotics to detect lanes for self-driving cars. The comprehensive aim of the research is to determine one such intelligent transportation system that can use focus-based instance segmentation for detecting lanes and improving autonomous vehicle transportation. This research proposes a technique for detecting traffic lanes using focus-based instance segmentation for vehicle detection on the road and highlighting lanes with drivable space. The proposed network uses a loss-trained hourglass model. Additionally, this model is trained multiple times using the same loss function to achieve competitive accuracy and a high proportion of false-positive results. The information comes from the German Traffic Sign Detection Benchmark (GTSDB), which is public data.

keywords - fast region-based convolutional neural network, focus-based image segmentation, hough transform, hourglass model

1 Introduction

The emphasis of industry and academic computer vision and robotics research is on fully autonomous cars. The purpose of each research is to get a better understanding of the vehicle's surroundings via the use of different control and sensor modules. That environmental awareness is completed by camera-based lane recognition, which enables automobiles to be put accurately in the lane. Traditional lane recognition algorithms depend on heuristics and handcrafted characteristics, while fully autonomous driving relies on precise camera-based lane detection (Li et al.; 2019). Following the identification of line segments, post-processing methods are utilized to locate group segments and erroneous lane instructions for the remaining lanes. The pixel hierarchy feature may aid in the improvement of algorithms and the modeling of contextual information when selecting contextual characteristics for lane marker recognition.

Figure 1 displays the objective of our proposed technique, which predicts drive area and lanes from segmented RGB input pictures and differentiates regions into individual instances using embedding characteristics recovered by the suggested network. Furthermore, the proposed network is trained from start to finish, and the network size can be adjusted by adding another loss function model based on the computing power of the target system.

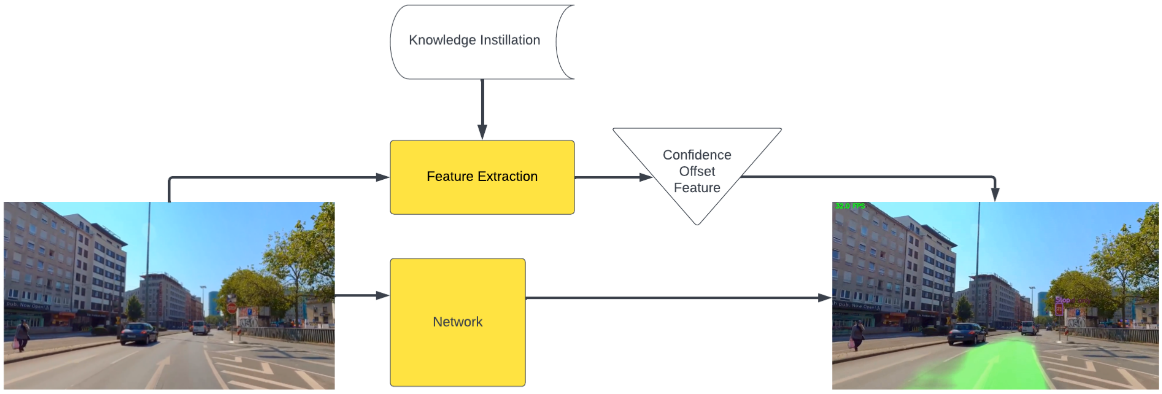

Figure 1: Overview: Drive Area and Lane Detection

The choice of topic is due to the increasing contribution of road accidents to global death numbers, the prevalence of poor structuring and designing of roads, and the fact that reckless driving has become eminent. Nearly 1.35 million accidental annual deaths take place globally. It represents traffic accident risk in underdeveloped and developing nations due to road accidents. Considering an economic perspective, road accident costs are projected to be more than 4 billion US dollars. Increasing the number of autonomous vehicles rather than manual vehicles can help to reduce the number of accidents. Autonomous cars contribute to avoiding road accidents by combining different modules such as lane detection, adaptive cruise control, and driver awareness. Because onboard visual sensors are available, computer vision methods are chosen for lane detection. Traditionally, lane detecting methods were focused on conventional characteristics (Mamidala et al.; 2019). However, it did not provide satisfactory performance regarding road curvature challenges and blurred lane lines.

The goal of this paper is to find intelligent transportation systems that use focus-based instance segmentation to find lanes efficiently.

The following are the objectives of the paper based on creating an intelligent transportation system:

1. To develop the implementation of lane detection systems facilitating automatic driving, technology integrated with artificial intelligence, sensor perception networks, intelligent networks, and computer vision is needed.

2. To evaluate the establishment of a Faster Region-based Convolutional Neural Network (Faster R-CNN) road track detection system for identifying concrete tracks.

3. To implement an instance segmentation technique for end-to-end road detection.

Therefore, the research intends to identify the answer to the following question:

1. How well does Focus-based Instance Segmentation (FIS) spot arrow and curved traffic lanes to improve transportation network efficiency?

This research's primary contribution is to help understand the positive impact of lane detection techniques that can prevent road accidents. Proper leave detection will help autonomous vehicles position the lane and avoid the risk of intruding into other lanes. The research will provide a solution to imply a transformation view in the image before fitting the curve, depending on a fixed transformation matrix. To output a transformation coefficient, a neural training network requires an image input optimized with a loss function structured for the lane fitting issue. These elements' evaluation represents specific findings about plane segmentation and branch output densely using multi-task and branch architecture for lane detection problems. It enables the inference of arbitrary lane numbers. The network parameters for estimating perspective transformations facilitate robust lane fitting for road layout changes.

The limitation associated with the paper is due to the existing gap in the literature regarding the effectiveness of the lane detection system for autonomous and human-driven vehicles. Further, the research will imply model-based methods using dynamic programming and support vector machines using geometric parameters for better lane detection. It is efficient to incorporate manual feature extraction to eliminate time complexity and high computational leads. However, implementing these methods in real life is challenging since they only emphasize providing higher accuracy in lane detection for the simulation environment. To encourage real-world implementation, the systems required for a better application of improving lane detection should be capable of incorporating valuable data with less render time. This led to the development of lightweight CNN techniques for real-time plane detection (Zhang et al.; 2018). E-net, another encoder-decoder structure model, employed self-attention dilation and distillation convolution to create a lightweight model with a quicker rendering duration. Simultaneously, another novel, lightweight, fast structure for lane detection is Dual Attention-based Dense SU-net (DSUNet), which offers significant advantages in reducing the render time. Convolutional neural network (CNN) methods have some drawbacks because they do not account for how poor road conditions affect the performance of models and autonomous vehicles. The lack of communication pathway maintenance and the creation of a good road network to incorporate existing networks of length detection and allow independent vehicle introduction in developing countries is challenging due to financial restrictions. Therefore, low-income countries have difficulties solving such issues. Beyond the constraints of the proposed lightweight CNN techniques for estimating lanes on bad and structured roads regardless of road condition, it is critical to visualize the model elements.

The paper is structured in the following manner: Section 1 will provide an introduction. Section 2 will discuss related work regarding existing lane detection techniques. Section 3 will be a concise methodological overview of the research. Section 4 will evaluate the design specifications underpinning the implementation of the lane detection techniques. Section 5 will provide the course of action for implementing the proposed linked detection solution. Section 6 will critically evaluate and summarize the proposed solution in the form of experimental research outputs. The last section, 5, will conclude the paper and provide inferences for future work.

2 Related Work

The model-based method considers the road lane lines as the simultaneous geometric technique for fitting them to obtain model parameters. Model-based methods utilize specific templates or parameters for representing lines, assuming parabolic curves or straight lines. The model-based technique is more efficient in noise removal because of its highlevel processing rather than pixel-based processing.

2.1 Model- and feature-based methods

An intelligent transportation system for lane detection is primarily based on visual sensors. These have become significant for capturing road scenes on the smart card through cameras in front of the vehicle. Visual sensors use a broad-spectrum response range that visualizes infrared rays, which are not visible to the bare human eye. The model and feature-based methods are used to increase computer speed (Cao et al.; 2019). Also, using Bayesian classifiers to build probability generation models from pixel points on the road surface helps to figure out where the road ends and where the vehicle is going.

The feature-based method segregates the lane from the actual road condition based on color as well as each feature of the lane (Zhang et al.; 2018). This feature-based model uses a two-dimensional linear filtering technique for the detection of right and left lanes, which will eliminate noise interferences during the detection process and conserve features of the lane from a distant perspective. The feature waste method is efficient in creating an intelligent transportation system that identifies road lane markings with low-level characteristics. It provides clear color representation in image processing that gathers feature information about lane markings. The effectiveness of color representation can be improved by developing an adaptive technique for lane marking. However, color representation cannot portray lane marking characteristics, therefore making its use in integration with other non-color characteristics (Li et al.; 2018) possible.

2.2 Lane detection system

The development of a lane detection warning system is a module of safe driving assistance that offers safety through early warnings to the driver if the autonomous cars have been distracted from the identified lane. The purpose of such a system is to notify vehicle riders to avoid or reduce lane departure incidents. It uses relevant sensors to obtain road data close to the vehicle and analyze the information for the state of the driving vehicle, notifying the time established in the system and the warning threshold. The system characterized as supporting staying in the correct lane while driving uses a camera for detecting lane markings (Navarro et al.; 2017). It has an instrument cluster indicating that it is on standby and that lane lines on both sides are identified. The lane detection technique is based on computer vision and image resizing. In this context, the lane line detection technique initially uses image acquisition. Lane detection based on the camera is an essential technique for environment perception, helping vehicles locate the correct lane (Xing et al.; 2020).

The system simultaneously prevents image and noise processing. This will aid in regionally dividing lane images based on variations that take into account changes in road lighting and eventual road conditions. Further, the use of image processing algorithms will help in lane image enhancement. The valid region image is pre-processed using image sharpening and smoothing to improve image quality and feature extraction of the lane line. It enhances the big image using algorithms of exponential transformation, increasing contrast between lines in the lane and making other images darker. Another algorithm transforms an image into one with lighter, more vague lane lines. The following algorithm uses image equalization but provides a still-wave aspect. In contrast, the dark channel prior-based algorithm helps improve image contrast and color with unclear lino in the lane (Romera et al.; 2017). The last is using a retinex-based algorithm to improve the contrast between lane lines and other parts without enhancing the color properly.

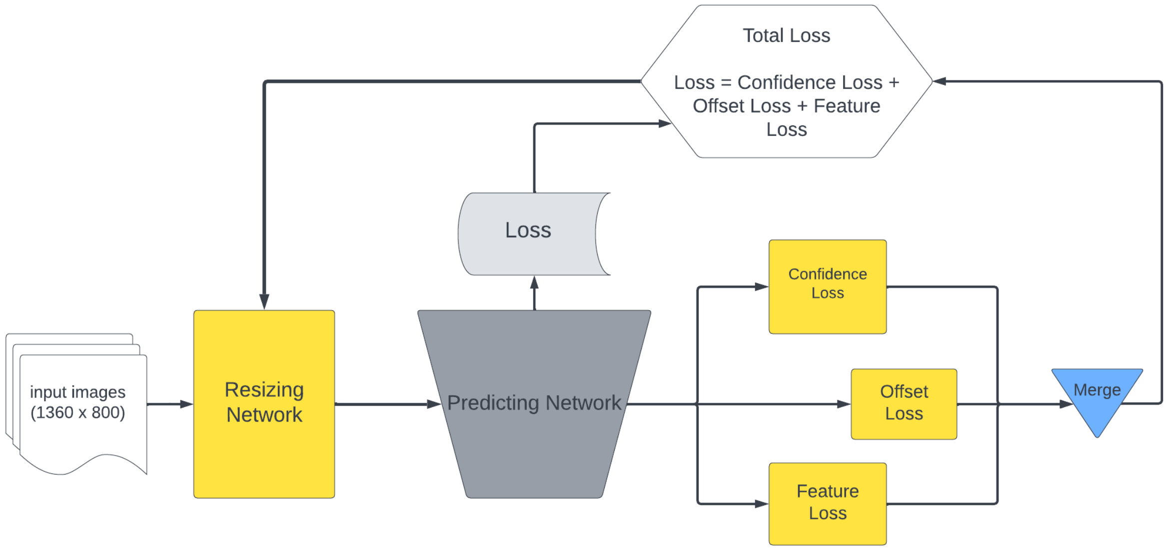

Figure 2: Hourglass Model with Loss Function

Figure 2 shows the network design. 1360 x 800 RGB image input. The network resizes it. The resizing network lowers the original image's resolution to 64 x 32 pixels using convolution layers; the prediction network uses its output. The prediction network employs four hourglass modules. All hourglass modules utilize the same loss function. Customers may determine the amount of hourglass modules to use depending on processing capacity after training. Each network is explained below.

2.3 Segmentation network

Simultaneously, the cinematic segmentation network, based on deep convolution and neural networks, classifies various pixel points in the image while aggregrating similar kinds of pixel points. This helps in the distinction between various target points in the image (Wang et al.; 2019). These algorithms and techniques have aided in the improvement of the land detection warning system, which can improve the combination of several sensors for better information about roads and their conditions, allowing vehicles to be directed in different directions. Further, it encourages 3D detection algorithms or methods for improving reliability to detect the correct lanes. Moreover, the classification of adverse road conditions and weather using particular algorithms is appropriate for lane line detection routes along with deep learning methods, which should be beneficial with deep learning methods developed based on traditional image processing algorithms (Hu et al.; 2020).

2.4 Random sample consensus (RANSAC) algorithm

Convolutional Neural Network (CNN) technology combined with the RANSAC algorithm begins lane detection from the image's edge. Dual-View CNN is another excellent technique that utilizes top-view and front-view concepts to remove non-club-shaped structures and prevent false detection. The difficulty in such cases arises at the final stage, when generating binary lane segments for various lane instances must be disentangled (Tang et al.; 2021). Overcoming this challenge required approaches to be applied to post-processing techniques depending on heuristics. Such heuristic techniques are vulnerable to robustness issues and computationally costly because of roadside variation. Another issue is the multi-class segmentation challenge, where each lane creates its own class, leading to the network output containing the disentangled binary map (Wu et al.; 2021). Such a method restricts the detection of fixed and predefined lane numbers. Proper lane instance estimation makes it easier to convert into a parametric description for each. Popular models of curve fitting algorithms are clothoid, spline, and cubic polynomials. Converting the image into a bird's-eye view will provide computational efficiency (Neven et al., 2018). This can be estimated in the original picture using an inverse transformation matrix. Consecutively, lane points near the horizon can be estimated to infinity, adversely impacting lane fitting.

2.5 Hough transform

It is possible to create a reliable model for lane identification by using driving data in conjunction with vertical spatial properties. The data exchange block and the characteristic merging block are effective in allowing the detection of occluded and unclear lane lines. The Hough transform can be used in image-processing computer language platforms to identify lane markers, which is then followed by noise filtering (Shashidhar et al.; 2022). Hough transform may also be used to build a set of hyperbolas that correspond to the lane's borders and enhance the performance of an eyesight lane identification technique. It is possible to recognize approaching autos on the road using both feature-based and appearance-based methods. Most vehicle detection systems employ either optical flow or frame or background removal (Li et al.; 2019). However, background subtraction is today's most widely used method for detecting vehicles. Moreover, tail light pairs can be used for detecting vehicles during the night, whereas other methods include the Haar filter and texture descriptor.

Several researchers established an architecture for building in a new layer employing factorized convolutions and residual collections. Deep convolutional networks that multi-task help identify the existence of the target and its geometric characteristics. A recurrent neural layer helps manage the spatial distribution of observational cues per structure for visual identification associated with objects whose structure or shape is problematic to feature directly (Yang et al.; 2020). In an emergency, the modern era proposed Internet of Things IoT for traffic systems, where specific sensors are integrated into its intelligent traffic system using a microcontroller.

2.6 Faster Region-based Convolutional Neural Network (Faster R-CNN)

Faster R-CNN algorithms are currently known for making the most progress in object detection accuracy and execution efficiency. The identification of objects is critical in this lane detection system, and the Faster R-CNN algorithm has been used effectively for this. This algorithm has three major steps: region proposal generation, feature extraction, and classification as discussed by (Zou and Song; 2018). In this case, the feature vector is used to assign each region proposal for object classification, and a number of classes are linked to each other to build different aspects of objects. The R-CNN network is composed of three major modules: the first is associated with the selective search algorithm, the second with a fixed predefined size to effectively extract an image or object (Girshick; 2015) . This project has successfully implemented all three of these modules.

3 Methodology

The suggested framework, Focus-based Instance Network, is shown in detail in Figure 2. It produces traffic line points and differentiates distinct cases. The resizing network receives the 1360 $\times$ 800 input size and scales it down to 64 x 32. From the resized input, the prediction network extracts characteristics to create outputs; many hourglass modules are linked in sequence. Each hourglass module has output branches that create three types of output. These results allow us to discriminate between similar situations and estimate where traffic line points will be. By clipping a few modules, users may construct a lighter model without further training or alterations since all hourglass modules share the same output branch (Zagoruyko and Komodakis; 2016). To avoid performance differences between full-size and short-clipped models, the knowledge distillation approach is used. In all hourglass modules, the knowledge distillation component and three outputs are used to construct loss functions. Back-propagation is performed using a weighted sum of all loss functions, with individual loss functions.

3.1 Architecture

The intended network design is shown in Figure 2. 1360 $\times$ 800 is the size of the input RGB image. The network receives it for resizing. The input of the prediction network is the output of the resizing network, which uses a sequence of convolution layers to lower the original image's resolution to 64 $\times$ 32 pixels. The prediction network in this study utilizes four hourglass modules (Zhu et al.; 2021). All hourglass modules are trained simultaneously using the same loss function. Following the training phase, users may decide how many hourglass modules to use according on their computer's processing power without further instruction. Each network is thoroughly described in the sections that follow.

3.1.1 Shrinking Network

To conserve memory and inference time, the resizing network shrinks the size of the input picture. First, the RGB input picture size is 1360 x 800. Three convolutional layers make up this network. Filter size is 33, stride is 2, and padding is 1, which are the settings for all convolution layers. Following each convolution layer, batch normalization (Ioffe and Szegedy; 2015) is used. Finally, this network produces resized output with a size of 64 x 32. The component layers are detailed in Table 1.

3.1.2 Forecasting Network

The resizing network's output goes to the prediction component. Exact positions on traffic lines and embedding properties, such as segmentation, are predicted in this section. This network comprises many hourglass modules. Several skip-connections transmit information from several scales to deeper layers. There are three types of bottlenecks: identical, downward, and upward. The same bottleneck produces output and input of the same size. For downsampling in the encoder, we use the down bottleneck, with its first layer swapped out for a convolution layer using a 3 x 2 x 1 stride and padding filter. For the upsampling bottleneck in the upsampling stages, the transposed convolution layer with a filter size of 3, stride of 2, and padding of 1 is used. Each output branch is comprised of three convolution layers and produces a 64 x 32 grid. Output branches forecast confidence values about the presence, offset, and embedding characteristics of every cell in the output grid. Table 1 contains the network's specifics. A deeper network with superior performance may serve as a teaching network (Newell et al.; 2016). As a result, we may anticipate improved performance from short-clipped networks when applying knowledge distillation approaches. Each output branch has a distinct channel of confidence as 1, offset as 2, embedding as 4 and uses a unique loss function to achieve its specific aim.

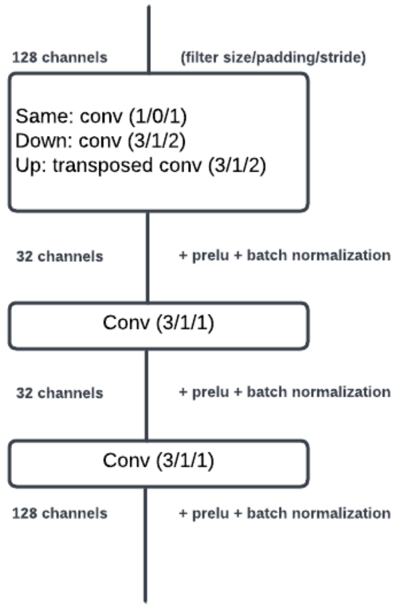

Figure 3: Bottle-neck layers

3.2 Loss calculation

Each hourglass module's output is subjected to one of four training-specific loss functions. The next sections describe each loss function in depth. According to Table 1, this same output branch creates a 64-grid, with each cell containing the projected values from seven channels, including the embedding feature, offset, and value of 1 for each channel (4 channels). The confidence value decides if there are drivable areas; the offset value specifies the precise location of the key points anticipated by the confidence value; and the integration feature is used to separate key points into distinct instances. Therefore, each cell of the output grid is subjected to three loss functions, except the distillation loss function. Each encoder's distillation layer has an individually optimized distillation

Table 1: Specifics of the proposed network

Type	Layer	Output size
Input data		3 x 512 x 256
Resizing	Conv+Prelu+bn	32 x 256 x 128
	Conv+Prelu+bn	64 x 128 x 64
	Conv+Prelu+bn	128 x 64 x 32
Encoder	Bottle-neck(down)	128 x 32 x 16
	Bottle-neck(down)	128 x 16 x 8
	Bottle-neck(down)	128 x 8 x 4
	Bottle-neck(down)	128 x 4 x 2
	Bottle-neck	128 x 4 x 2
(Distillation layer)	Bottle-neck	128 x 4 x 2
	Bottle-neck	128 x 4 x 2
	Bottle-neck	128 x 4 x 2
Decoder	Bottle-neck(up)	128 x 8 x 4
	Bottle-neck(up)	128 x 16 x 8
	Bottle-neck(up)	128 x 32 x 16
	Bottle-neck(up)	128 x 64 x 32
Output branch	Conv+Prelu+bn	64 x 64 x 32
	Conv+Prelu+bn	64 x 64 x 32
	Conv	C x 64 x 32

loss function for extracting the instructor network's knowledge (Zhu et al.; 2021). Each anticipated value and characteristic is described in the sections that follow.

3.2.1 Confidence Loss

Each cell's confidence value is predicted by the confidence output branch. If the cell has a critical landmark, this confidence value is quite high; otherwise, it is rather low, often around 0. The one-channel outcome of such confidence branch is passed to the subsequent hourglass module. Existence loss and nonexistence loss are the two components of confidence loss (Ko et al.; 2021). Existence reduction is applied to cells containing important points, while non-existence loss is used to decrease the confidence score of each depth cell. Due to the quick convergence of cells far from key points, this training method focuses on cells nearer the important points.

3.2.2 Offset Loss

Focus-based Instance Segmentation (FIS) predicts the precise position of each output cell's key points based on the offset branch. Each cell's output is between 0 and 1, where the higher the number, the higher the cell's relative position. For the purposes of this work, 8 input pixels are associated with a single cell. For instance, the key point would actually be 4 pixels from the cell's edge if the anticipated offset value was 0.5. The offset branch consists of two channels for forecasting x- and y-axis offsets. When figuring out offset loss (Ko et al.; 2021), cells without critical points are left out because they don't have any ground truth.

3.2.3 Feature Loss

Once the network has been trained, the error function brings features together when each cell corresponds to the identical instance and scatters features when cells correspond to distinct instances. Using a simple distance-based clustering method, we can separate important points into unique instances. In this work, we consider two instances to be identical if their embedding characteristics are within a specific distance. In this trial, the feature count is set at 4, however this size has no discernible impact on performance.

3.2.4 Distillation Loss

When more hourglass units are stacked, performance is improved. In this way, the whole hourglass module may serve as a teacher network, and we anticipate that short-clipped networks, which are less resource-intensive than the teacher network, would exhibit superior performance after the use of a knowledge distillation technique as proposed by (Zagoruyko and Komodakis; 2016).

The end-to-end approach is used to train the whole network, and the overall loss is calculated as the sum of the aforementioned four individual loss components.

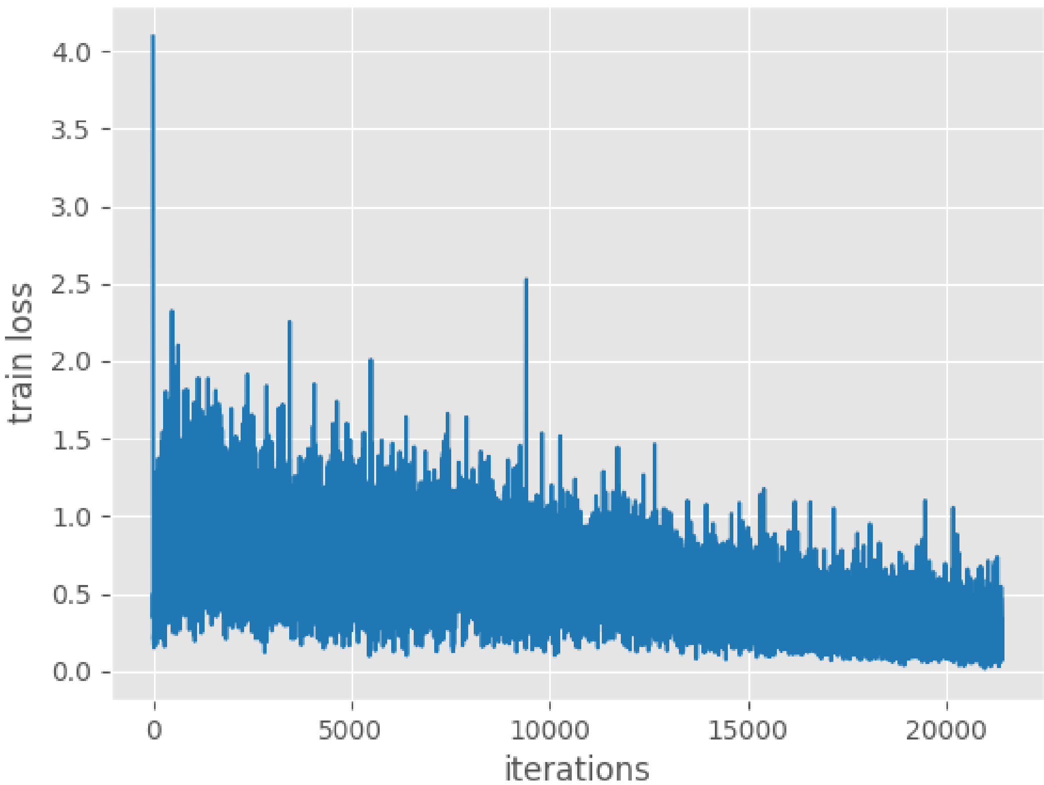

Figure 4: Loss function training values

Figure 4 Figure 4 depicts the train loss values attained after about twenty thousand cycles of training, resulting in a decreased loss value.

4 Design Specification

This section will focus on the methods employed in the study of lane detection utilizing computer vision and robotics. There is a brief overview of the project, as well as a list of sources and some approaches. In order to create intelligent transportation systems that can reduce car accidents caused by unnoticed lane changes, lane detection is a crucial component. This linear parameter space combines findings from computer vision and robotics research to construct a model with improved lane change prediction accuracy.

Faster R-CNN, the most extensively utilized state-of-the-art variant of the R-CNN family (Zou and Song; 2018), was published for the first time in 2015. This section explains the principles of modern object detection and elaborates on the technical specifics of the pipeline that the Faster R-CNN uses to recognize objects. Fast R-CNN is independently trained. CNNs trained on ImageNet are used to seed the main network for this job, which is subsequently tweaked for object detection.

Typically, the development of the Region-based Convolutional Neural Network (R-CNN) family occurred in terms of computing efficiency, merging the various training phases, decreasing test time, and enhancing performance. These networks typically include a region proposal method to produce the locations or boundries of potential objects in the picture, a feature generating stage to get the features of these objects, classification layer that determines the class this item belongs to, as well as a regression layer to improve the accuracy of the object boundries' coordinates.

Lane detection is an essential feature for developing intelligent transportation systems that can prevent road accidents due to undetected lane changes (Girshick; 2015). Computer vision and robotics research and design specifications are combined in this linear parameter space to build a model for better performance at predicting lane changes. This article discusses related research works as well as the design specifications themselves, along with some methods applied in the evaluation phase of the project.

Table 2: Frames Per Second on GPU (google colab) for FIS

Method	**frames per second**
SCNN (Pan et al.; 2018)	28
PINet (Ko et al.; 2021)	33
FIS	32

5 Implementation

In this part, we test Faster R-CNN using public dataset and discusses implementation specifics and assessment methodologies. For both training and testing, all input pictures are scaled to 512 x 256 pixels and normalized from RGB values of 0 to 255 to values of 0 to 1 before being given to the proposed network. The German Traffic Sign Detection Benchmark (GTSDB) dataset utilized to evaluate the proposed technique gives x-axis readings of traffic lines based on fixed y-axis values. The annotation technique sparsely annotates certain traffic lines along the horizontal line. Using linear regression on the original data, we create extra annotations per 10 pixels along the x-axis to fix this issue. Various data enhancement techniques, including shadowing, noise addition, flip, rotation, translation, and intensity modification, are also used.

Furthermore, the dataset contains a large number of picture frames, but the data is unbalanced. The GTSDB dataset's testing set, for instance, has scenes labeled as "normal," "night," and "crossroad," among others, and the quantities of category frames vary considerably. We To tackle this problem, we collect hard data that exhibits poor loss values during training and raise the hard data selection ratio. The idea is comparable to the practice of "hard negative mining".

Visual Studio Code is used to write source code. In the training phase, around 100 to 200 epochs are used to train our network. The thresholds and coefficients used in each batch are based on extensive experimentation, and there are six photos in each batch. In the next section, we display the precise values for the hyper-parameters. A spline curve fitting approach (Lu et al.; 2011) is used to create smooth curve lines, and a focus-based segment is used to forecast the precise empty space on the road.

5.1 Dataset

Our suggested network, FIS, is trained with the German Traffic Sign Detection Benchmark (GTSDB) dataset (Fan; 2021). Table 3 provides a summary of the dataset. The GTSDBdataset contains simply the highway area and fewer impediments. To assess FIS's performance, we consulted the official assessment source codes, and we will go through the datasets and metrics we used in a little more depth below.

Table 3: Dataset Summary

Property	**Value**
Dataset	GTSDB
Train	600
Test	300
Resolution	1360 x 800
Type	highway, regional roads

The GTSDB dataset is composed of 900 training sets, 600 of which are allocated to the training set and 300 to the testing set. The GTSDB dataset's primary assessment measure is accuracy, which is calculated by the equation based on the average percentage of right points.

$$accuracy = \sum_{clip} \frac{C_{clip}}{S_{clip}}$$

where C_{clip} is the fraction of the picture clip for which the trained module made an accurate prediction and S_{clip} is the fraction of the clip for which the ground-truth prediction was right. The following expressions also give information on the rates of false negatives (FN) and false positives (FP):

$$FP = \frac{F_{pred}}{N_{pred}}$$

$$FN = \frac{M_{pred}}{N_{gt}}$$

F_{pred} signifies mistakenly predicted lanes, N_{pred} predicted lanes, M_{pred} missing lanes, and N_{gt} ground-truth lanes.

6 Evaluation

The Focus-based Instance Segmentation (FIS) network is investigated under various conditions and compared to segmentation models SCNN and PINet. In terms of road area detection, FIS yields comparable results to other cutting-edge techniques.

6.1 Experiment / Case Study 1

Table 4 displays the complete assessment findings for the GTSDB dataset. The high accuracy and low false-positive rate achieved by the Focus-based Instance Segmentation (FIS) network are achieved without the use of pre-trained weights or any additional datasets. The rate of false-negatives likewise shows a respectable value. The performance of the full-size network is superior to that of the short-clipped networks, but there are some small changes. The minimum distance required to differentiate each occurrence is 0.08 meters.

Table 4: Evaluation Results

Method	**Accuracy**	**FP**
SCNN (Pan et al.; 2018)	96.53%	0.0617
PINet (Ko et al.; 2021)	95.81%	0.0310
FIS	96.71%	0.0585

6.2 Experiment / Case Study 2

Figure 5 illustrates instances on arrow road classifications. Using confidence maps, we compare our technique to various semantic segmentation-based methods. The arrow category comprises of certain road markings that might be mistaken with traffic lines; FIS separates several road markers from traffic boundaries. Although the FIS approach uses the whole input size, it works well.

6.3 Experiment / Case Study 3

This case study examines the detection of curved lane roads by the proposed FIS network, which, despite its imprecision, is able to detect curves along the road line as well as traffic on the road Figure 6.

6.4 Discussion

In Case Study 1, which examines the accuracy and false positives of the proposed model, it is discovered that the system's accuracy is marginally higher than that of other modern segmentation networks. In contrast to Point Instance Network (PINet) (Ko et al.; 2021), which trains several hourglass models at the same time using the same loss functions, the number of false positives that are produced by the proposed Focus-based Instance Segmentation (FIS) method may be said to have a greater scope. The second case study compares the images that were produced on the arrow-signed road lanes. The comparison is between the Spatial Convolutional Neural Network (SCNN) (Pan et al.; 2018), the PINet, and FIS. It was discovered that the proposed model is able to detect road signs

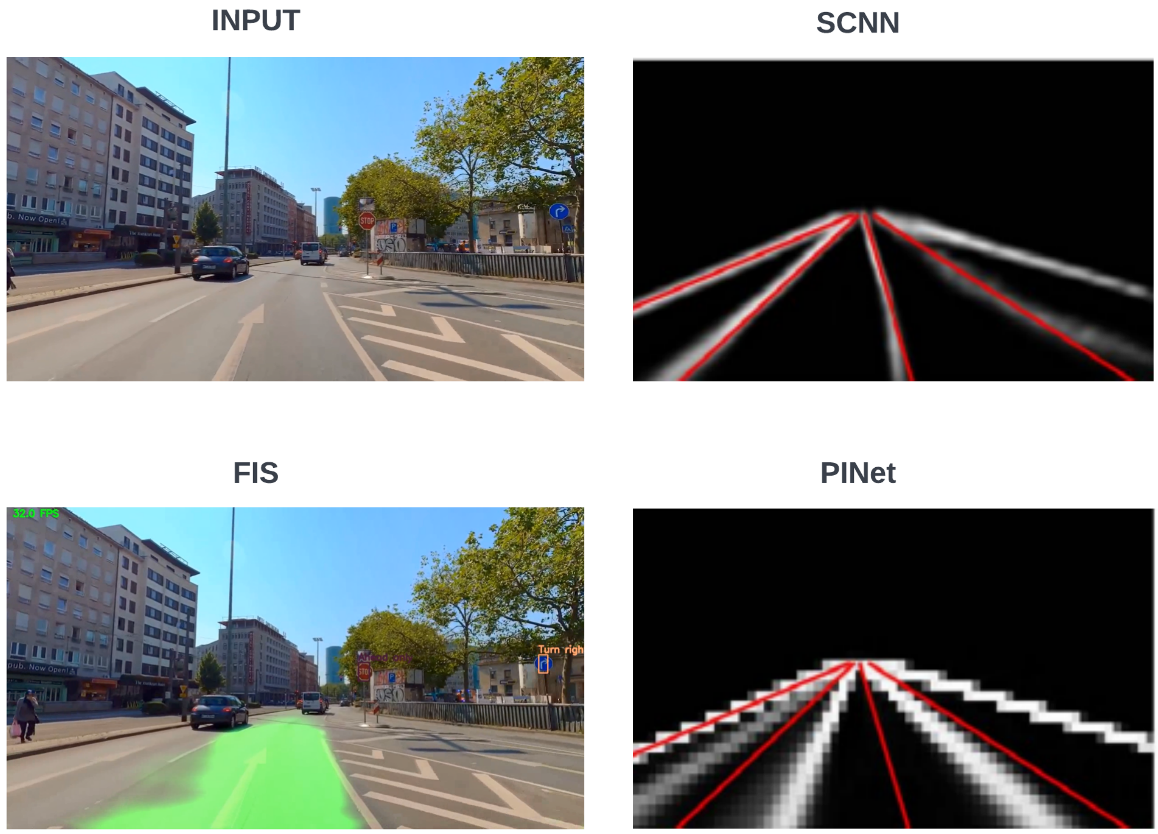

Figure 5: Comparison: Arrow Road Lane

Figure 6: Output: Curved Road Lane

reasonably well compared to other networks, in spite of using a single loss function for approximately twenty thousand training iterations.

Case study 3 demonstrates the output of the proposed network in the form of curving road lanes, where it is determined to be suitable and effective for the given circumstances. Overall, FIS delivers anticipated and equivalent results in terms of accuracy, although it may be tweaked to incorporate additional loss training, which can raise false positives and be used to advertise the technology.

7 Conclusion and Future Work

In this research, we presented a lane recognition algorithm, Focus-based Instance Segmentation (FIS), and examined its performance on arrow-signed highways and curving roads. Also suggested is the use of Fater Region-based Convolutional Neural Network (Faster R-CNN) as the training method and picture segmentation for road item identification. Some prior algorithms predicted pixel-level classifications, but FIS identifies all the drivable region inside a traffic lane. FIS is comprised of hourglass modules and loss functions applied to hourglass modules. FIS achieves good performance, a reduced frequency of false positives, as well as frames per second in the GTSDB dataset; the lower false positive rate ensures the safety management of autonomously driving automobiles since incorrectly predicted lanes occur rarely. Specifically, FIS works better than other approaches in arrow-marked lanes, and it also performs well when curved road lanes are considered and the vehicle remains in the same curved lane (the vehicle is not changing the lane). For curving lanes, the FIS network does not function well, especially when a vehicle is switching lanes. We considered a few case studies to demonstrate this. Therefore, we conclude that the overall performance of FIS is excellent since it takes into account a variety of road lane types, including curved lanes, arrow-signed lanes, and straight lanes.

This project is part of a long-term research theme in computer vision that will require many projects over time to yield findings in the area of intelligent transportation systems. Future research can utilize multiple hourglass models and simultaneously train them with loss functions in order to achieve more accurate results. Also, more research can be done on how to make tighter turns and narrower lanes in cities so that self-driving cars will be more reliable when they hit the market.

Acknowledgement

Prof. Jorge Basilio is sincerely thanked for his calm, on-target input throughout the duration of this project, as well as his sincere guidance and comprehension of the endeavor.

References

Cao, J., Song, C., Song, S., Xiao, F. and Peng, S. (2019). Lane detection algorithm for intelligent vehicles in complex road conditions and dynamic environments, *Sensors* **19**(14): 3166.

Fan, S. (2021). Road sign detection and recognition system based on multi-layers convolutional neural network model trained with german traffic sign recognition benchmark, *2021 4th World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM)*, pp. 395–399.

Girshick, R. (2015). Fast r-cnn, *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448.

Hu, J., Xiong, S., Sun, Y., Zha, J. and Fu, C. (2020). Research on lane detection based on global search of dynamic region of interest (droi), *Applied Sciences* **10**(7): 2543.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift, *International conference on machine learning*, pmlr, pp. 448–456.

Ko, Y., Lee, Y., Azam, S., Munir, F., Jeon, M. and Pedrycz, W. (2021). Key points estimation and point instance segmentation approach for lane detection, *IEEE Transactions on Intelligent Transportation Systems* .

Li, D., Zhao, D., Zhang, Q. and Chen, Y. (2019). Reinforcement learning and deep learning based lateral control for autonomous driving [application notes], *IEEE Computational Intelligence Magazine* **14**(2): 83–98.

Li, L., Luo, W. and Wang, K. C. (2018). Lane marking detection and reconstruction with line-scan imaging data, *Sensors* **18**(5): 1635.

Lu, S., Wang, K., Li, Y. and Chen, Y. (2011). Method of boundary detection based on b-spline curve fitting, *Proceedings of the 30th Chinese Control Conference*, pp. 3248–3253.

Mamidala, R. S., Uthkota, U., Shankar, M. B., Antony, A. J. and Narasimhadhan, A. (2019). Dynamic approach for lane detection using google street view and cnn, *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, IEEE, pp. 2454–2459.

Navarro, J., Deniel, J., Yousfi, E., Jallais, C., Bueno, M. and Fort, A. (2017). Influence of lane departure warnings onset and reliability on car drivers' behaviors, *Applied ergonomics* **59**: 123–131.

Newell, A., Yang, K. and Deng, J. (2016). Stacked hourglass networks for human pose estimation, *European conference on computer vision*, Springer, pp. 483–499.

Pan, X., Shi, J., Luo, P., Wang, X. and Tang, X. (2018). Spatial as deep: Spatial cnn for traffic scene understanding, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

Romera, E., Alvarez, J. M., Bergasa, L. M. and Arroyo, R. (2017). Erfnet: Efficient residual factorized convnet for real-time semantic segmentation, *IEEE Transactions on Intelligent Transportation Systems* **19**(1): 263–272.

Shashidhar, R., Arunakumari, B., Manjunath, A., Ahuja, N. J., Hoang, V. T., Tran-Trung, K. and Belay, A. (2022). Computer vision and the iot-based intelligent road lane detection system., *Mathematical Problems in Engineering* .

Tang, J., Li, S. and Liu, P. (2021). A review of lane detection methods based on deep learning, *Pattern Recognition* **111**: 107623.

Wang, W., Wang, M., Li, H., Zhao, H., Wang, K., He, C., Wang, J., Zheng, S. and Chen, J. (2019). Pavement crack image acquisition methods and crack extraction algorithms: A review, *Journal of Traffic and Transportation Engineering (English Edition)* **6**(6): 535–556.

Wu, L., Lin, X., Chen, Z., Lin, P. and Cheng, S. (2021). Surface crack detection based on image stitching and transfer learning with pretrained convolutional neural network, *Structural Control and Health Monitoring* **28**(8): e2766.

Xing, Y., Lv, C., Wang, H., Cao, D. and Velenis, E. (2020). An ensemble deep learning approach for driver lane change intention inference, *Transportation Research Part C: Emerging Technologies* **115**: 102615.

Yang, W., Zhang, X., Lei, Q., Shen, D., Xiao, P. and Huang, Y. (2020). Lane position detection based on long short-term memory (lstm), *Sensors* **20**(11): 3115.

Zagoruyko, S. and Komodakis, N. (2016). Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer, *arXiv preprint arXiv:1612.03928* .

Zhang, Y., Liang, W., He, H. and Tan, J. (2018). Perception of vehicle and traffic dynamics using visual-inertial sensors for assistive driving, *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, IEEE, pp. 538–543.

Zhu, Y., Xu, X., Ji, Y., Shen, F., Tao Shen, H. and Lu, H. (2021). Graph convolutional hourglass networks for skeleton-based action recognition, *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6.

Zou, J. and Song, R. (2018). Microarray camera image segmentation with faster-rcnn, *2018 IEEE International Conference on Applied System Invention (ICASI)*, pp. 86–89.