MU	<b>ROAD SEGMENTATION IN SEGNET</b>
STAFA KEMAL AMBAR	A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY
ROAD SEGMENTATION	by MUSTAFA KEMAL AMBAR
IN SEGNET	In Partial Fulfillment of the Requirements for the Degree of Master of Science in Electrical and Electronic Engineering
NEU 2019	
	NICOSIA, 2019

# **ROAD SEGMENTATION IN SEGNET**

# A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

by MUSTAFA KEMAL AMBAR

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Electrical and Electronic Engineering

NICOSIA, 2019

### Mustafa Kemal AMBAR: ROAD SEGMENTATION IN SEGNET

# Approval of Director of Graduate School of Applied Sciences

# **Prof. Dr. Nadire CAVUS**

# We certify this thesis is satisfactory for the award of the degree of Masters of Science in Electrical and Electronics Engineering

**Examining Committee in Charge:** 

Assist. Prof. Dr. Sertan Serte

Committee chairman, Department of Electrical and Electronics Engineering, NEU

Assist. Prof. Dr. Ayşegül Eren

Department of Basic Sciences and Humanities,CIU

Assist. Prof. Dr. Umar Özgünalp

Supervisor, Department of Electrical and Electronics Engineering, CIU

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Signature: Date:

# ACKNOWLEDGEMENTS

Most importantly, I might want to offer my genuine thanks to my supervisor, Assist Prof Dr. Umar Özgünalp for constant help and priceless advices in my exploration and planning of this report. His understanding, inspiration and information have been of incredible incentive for me. I additionally might want to thank my friend Niyazi Şentürk, for his priceless direction, advices, and remarks to my exploration. Last yet not the least, I might want to thank my friends for the help, accommodating talks and joint effort.

To my grandmother...

### ABSTRACT

The purpose of the study is to offer some insights into the process in Semantic segmentation to image analysis tasks for classification. It is an attempt to improve concurrent research, while addressing the advantages and disadvantages of various approaches. In this thesis, applied research is to perform road segmentation with the end goal that the drivable area can be perceived for lane detection. Standard road segmentation approaches are generally using image data from cameras as input, which is exposed to the wide variety of light conditions while segmenting road area. That makes road segmentation difficult. In this work, KITTI dataset has been used. This dataset is aiming for intelligent vehicles applications. In this thesis, road segmentation has been achieved using SegNet (A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation). SegNet, is a deep-learning network designed for advanced driving assistance systems. The proposed approach, is combining Inverse Perspective Mapping (IPM), and SegNet for road segmentation. Additionally, algorithm performs well along with the encouraging results. The custom detection approach was tested on the KITTI object detection benchmark, and was able to successfully detect road and no road conditions. The experimental studies presented so far provide evidence that, pixel-wise semantic segmentation of road images dependent on existing effective designs; namely SegNet has successfully performed road segmentation in algorithm. Detailed discussion and insights into future work are stated, to develop a full picture of the study, additional studies will be needed in the area of neural network and semantic segmentation, which could also be interesting to concentrate for both academia and industry. Currently, pixel-wise road segmentation accuracy for the proposed approach is estimated as 88.73%.

*Keywords*: Neural network; semantic segmentation; road segmentation; inverse perspective mapping; segnet.

# ÖZET

Bu çalışmanın amacı, anlamsal bölümleme sürecindeki süreçle ilgili bazı sınıflandırma için görüntü analiz görevlerini tanıtmaktır. Çeşitli yaklaşımların avantaj ve dezavantajlarını ele alırken eş zamanlı araştırmayı geliştirme girişimidir. Bu tez çalışmasında, uygulamalı araştırma, sürülebilir alanın şerit tespiti için algılanabileceği hedefiyle yol bölümlendirmesi yapmaktır. Standart vol bölümlendirme yaklaşımları genellikle yol alanını segmentlere ayırırken çok çesitli ışık koşullarına maruz kalan girdi olarak kameralardan gelen görüntü verilerini kullanılması, yol bölümlemesini zorlaştırır. Bu çalışmada KITTI veri seti kullanılmıştır. Bu veri seti akıllı araç uygulamaları için kullanılmaktadır. Bu tez çalışmasında, SegNet (Görüntü Bölümlemesine Yönelik Derin Bir Evrimsel Kodlayıcı-Seti) kullanılarak yol bölümlemesi sağlanmıştır. SegNet, gelişmiş sürüş yardım sistemleri için taşarlanmış derin öğrenen bir ağdır. Önerilen yaklaşım, Ters Perspektif Haritalama (IPM) ve yol segmentasyonu için SegNet'i birleştiriyor. Ek olarak, algoritma cesaret verici sonuçlarla birlikte iyi bir performans sergiliyor. Özel algılama yaklaşımı, KITTI nesne algılama kriterinde test edildi ve yol ve yol olmayan koşullarını başarıyla tespit edebildi. Şimdiye kadar sunulan deneysel çalışmalar, yol görüntülerinin piksel şeklinde anlamsal bölümlemesinin mevcut etkili tasarımlara bağlı olduğuna dair kanıtlar sunmaktadır; yani SegNet, algoritmada yol bölümlendirmesini başarıyla gerçekleştirmiştir. Çalışmanın tam bir resmini geliştirmek için, hem akademi hem de sektörel çalışmalar için ilginç olabilecek, sinir ağı ve anlamsal bölümleme alanında ek çalışmalara ihtiyaç duyulacak ayrıntılı tartışma ve gelecekteki çalışmalara yer verilmiştir. Şu anda, önerilen yaklaşım için piksel bazında yol bölümleme doğruluğu% 88,73 olarak tahmin edilmektedir.

Anahtar Kelimeler: Yapay sinir ağı, semantik segmentasyon, yol segmentasyonu, ters perspektif haritalama, segnet

# TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
ABSTRACT	iv
ÖZET	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xi

# **CHAPTER 1: INTRODUCTION**

1.1 Thesis Problem	1
1.2 The Aim of the Thesis	1
1.3 Overview of the Thesis	2

# **CHAPTER 2 : LITERATURE REVIEW**

2.1 Literature Review	3
2.1 Literature Review	3

# **CHAPTER 3: RELATED WORK**

3.1 Neural Network	7
3.2 Deep Learning	10
3.3 Convolutional Neural Network	11
3.3.1 AlexNet	12
3.3.2 ZFNet	12
3.3.3 GoogleNet	13
3.3.4 VGGNet	13
3.3.5 ResNets	13
3.3.6 DenseNet	13

3.4 Intelligent Vehicles	13
3.5 Road Segmentation	14
3.6 Semantic Segmentation	16

# **CHAPTER 4 : MATERIALS AND METHODS**

4.1 Inverse Perspective Mapping	19
4.2 Image Datasets	21
4.3 Visual Geometry Group	23
4.4 Segnet	25

# **CHAPTER 5 : RESULT**

5.1 RESULT	27
------------	----

# **CHAPTER 6 : CONCLUSION AND DISCUSSION**

6.1 Conclusion	35
6.2 Discussion	35
6.3 Future Work	36

REFERENCES	37	7
------------	----	---

# LIST OF TABLES

Table 5.1: Comparing of Datasets Performances in IPM	34
Table 5.2: Comparing of Datasets Performances in Road Images	34

# LIST OF FIGURES

Figure 3.1: Information layers	8
Figure 3.2: Layers frameworks	9
Figure 3.3: A simple deep neural architecture	10
Figure 3.4: CNN picture classification pipeline	11
Figure 3.5: Road segmentation	14
Figure 3.6: Semantic segmentation principle	16
Figure 4.1: IPM method	20
Figure 4.2: IPM projected images	20
Figure 4.3: A typical road scene with IPM Method	21
Figure 4.4: An example image its thing annotations in coco	22
Figure 4.5: Overview of our rgb-d reconstruction and semantic annotation framework	22
Figure 4.6: Architecture of VGG	24
Figure 4.7: An illustration of the segnet architecture	25
Figure 4.8: SegNet predictions on urban and highway scenes	26
Figure 5.1: Graph of training and loss	27
Figure 5.2: Summary of IPM images performance in algorithm	28
Figure 5.3: Graph of KITTI base images	29
Figure 5.4: Graph of KITTI base images loss	29
Figure 5.5: Summary of KITTI images performance in algorithm	30
Figure 5.6: Graph of KITTI IPM images	31
Figure 5.7: Graph of KITTI IPM images loss	31

Figure 5.8: Applying IPM in KITTI road images	32
Figure 5.9: Summary of graphs	33

# LIST OF ABBREVIATIONS

<b>3G</b> :	Third Generation
ACM:	Association of Computing Machinery
ADAS:	Advanced Driving Assistance Systems
ADS:	Automated Driving Systems
AL:	Artificial Intelligent
ANN:	Artificial Neural Network
<b>BPNNS:</b>	Back Propagational Neural Structures
CNN:	Convolutional Neural Network
DARPA:	Defense Advanced Research Projects Agency
DSRC	Dedicated Short-Range Communications
DNN:	Deep Neural Network
FCL:	Fully Connected Layer
FCN:	Fully Convolutional Network
GPRS:	General Packet Radio Service
IETF:	Internet Engineering Task Force
ILSVRC:	The ImageNet Huge Scale Visual Acknowledgment Challenge

IP:	Internet Protocol
IPM:	Inverse Perspective Mapping
ITS:	Intelligent Transportation Systems
KITTI:	Karlsruhe Institute of Technology and Toyota Technological Institute
MANET:	A Mobile Ad Hoc Network
MLP:	Multi-Layered Perceptron
MLT:	Multi Learning Task
NEMO:	Network Mobility
NN:	Neural Network
ReLU:	Rectified linear Non-Linearity
SGD:	Stochastic Gradient Descent
SIGMOD:	Special Interest Group on Management of Data
SPARC:	Supporting Peoples Access To Reliable Care
VANET:	A Vehicular Ad Hoc Network
VGG:	Visual Geometry Group

#### **CHAPTER 1**

#### **INTRODUCTION**

The introduction part presents an overview about the stated problem, aiming to familiarize its readers to related terms and concepts about the topic.

## **1.1 Thesis Problem**

IPM is an approach to convert input images to bird's eye view using basic geometry. IPM removes perspective mapping, thus, output images become more appropriate for the data augmentation. In this work, first, images from KITTI dataset are used in IPM to extract Birds eye view of the scene. Then, the extracted images are feed into the SegNet for semantic segmentation. The organization of the algorithm has four main blocks: Input (an image), inverse perspective mapping of images, resizing of images and resizing of IPM images (so that, the input image size becomes compatible with SegNet), Training and testing in SegNet. The ground truth images provided in the KITTI dataset provides the road and no-road pixel positions for the algorithm. Segmentation of road is difficult due to wide variety of scenarios, due to cars, obstacles, shadows, saturation. IPM, also, removes a lot of noise sources such as sky, buildings, and etc. which becomes out of range in the IPM. These challenges are tried to be minimized by using IPM. The test vehicle is a vehicle equipped for getting test images. The main road is defined as a roadway that the test vehicle is traveling (in this thesis complete road area is segmented). The algorithm utilized MATLAB script for estimating IPM and utilizing SegNet.

#### 1.2 The Aim of the Thesis

This study focuses to develop deep learning that exploiting semantic segmentation patch training in VGG-16 datasets on neural network (SegNet could be initialized based on other architectures as well). Firstly, information about Inverse Perspective Mapping, and Neural Networking are gathered. Purpose of the thesis is segmentation of road using with SegNet and Inverse Perspective mapping images.

## **1.3 Overview of the Thesis**

The remaining of the thesis is organized as follows. Chapter 2, reviews some of the basic concepts and introduce works that are either inspiring or related to researches. Chapter 3, presents the related work for semantic image segmentation, deep learning, and neural network. It shows how we generate semantic segmentation in our thesis, how does the evaluation metric work and the experimental results. In chapter 4, there are materials and methods which has been using during thesis and details of technical background. Chapter 5, presents experimental results of the thesis and, finally, Chapter 6 concludes the thesis and discusses future work.

#### **CHAPTER 2**

#### LITERATURE REVIEW

In chapter 2 related research studies that focus on neural networks, semantic segmentation, road segmentation are reviewed. All researches gives with published years and authors.

#### 2.1 Literature Review:

There have been a number of studies involving about deep learning approach, as reported in (Yisheng et al., 2015) had been proposed and utilized on the accurate and timely traffic flow information is important for deployment of intelligent transportation systems. Existing traffic flow prediction methods mainly use shallow traffic prediction models. They proposed a novel deep learning based traffic flow prediction method which considers the spatial and temporal correlations. They were using auto encoders as building blocks to represent traffic flow features for prediction first time. Over the past years, inverse perspective mapping has been successfully applied to several problems in the field of Intelligent Transportation Systems. (Olivera et al., 2014) consists of mapping images to a new coordinate system where perspective effects are removed and that effects facilitates road and obstacle detection and also assists in free space estimation. In their research they proposed a strong solution based on the use of multimodal sensor fusion. Data from a laser range finder is fused with images from the cameras, so that the mapping is not computed in the regions where obstacles are present. This improves the effectiveness of the algorithm and reduces computation time. (Xiaolei et al., 2017); had proposed a convolutional neural network based method that learns traffic as images and predicts large-scale, network-wide traffic speed with high accuracy. In their method a CNN is applied to the image following two consecutive steps: abstract traffic feature extraction and network-wide traffic speed prediction. The method with four current algorithms, namely, ordinary least squares, k-nearest neighbors, artificial neural network, and three deep learning architectures, namely, stacked autoencoder, recurrent neural network, and long-short-term memory network. The proposed method outperforms other algorithms by an average accuracy improvement of 42.91% within an acceptable execution time. The CNN can train the model in a reasonable time and, thus, is suitable for large-scale transportation networks.

Advanced Driver Assistance System (ADAS) functionalities present in modern vehicles provide traffic improvement, comfort and safety for drivers, pedestrians, and the environment through the application of electronic, control and software. A digital camera is one of the technologies used in ADAS functions that can be installed in front of the vehicle. However, the problem of perspective effects can be seen in this domain, once the image acquired do not represents the real image due to distortions and errors in the processing stage. The method of Inverse Perspective Mapping (IPM) supports this demand by removing image distortions caused by perspective effect, generating new image coordinates in real-time. (Rodrigo, et al., 2010) presents the IPM and the implementation of a function to identify the distance from an object removing the perspective effect in the HSV colormap. They created uses a digital camera installed in the windshield to gather images, manipulates them, calculates parameters of the camera and uses a IPM to remove the perspective effect from plane images of the roads to calculate the distance of the vehicle ahead. After two years, (Chin, et al., 2012) had been proposed a method to detect obstacles around the vehicle by optical flow computation based on inverse perspective mapping. A method to improve is limited on a trajectory of the vehicle which moves along with a straight line. this situation will inconsistent in the left side and in the right side of images whic causes mistakes in obstracle detection. For solving this problem they proposed a using the trajectory of vehicle to calculate the center of turning circle. And the center of turning circle using to adjust the inconsistent with optical flow values. After their improvement, the optical flow values can keep consistence even if the trajectory is along with an arc line. Urban lane detection is an essential task for intelligent vehicles system. (Wang, et al., 2014) proposed to an approach of lane detection algorithm based on Inverse Perspective Mapping. In their research, firstly using overall optimal threshold method to obtain binary image for reducing noise; secondly using Inverse Perspective Mapping to transform binary image space to top view space; then using k-means clustering algorithm to analysis linear discriminant for reducing interference affect; finally, fitting lane discontinuous on the top view space according road models. In order to reduce the influence of road edge and distant scenery in lane marking detection, this algorithm removed the sky portion of the image which made the roads image area accounted for the majority. (Oord, et al., 2016) presented a deep neural network that sequentially predicts the pixels in an image along the two spatial dimensions. Modeling the distribution of natural images is a problem in deep learning. In their research models the discrete probability of the raw pixel values and encodes the complete set of addiction in the

image. Also include fast two-dimensional recurrent layers and an effective use of residual connections in deep recurrent networks. During the same years (Jeong and Ayoung., 2016) proposed an adaptive Inverse Perspective Mapping (IPM) algorithm to obtain accurate bird's-eye view images from the sequential images of forward looking cameras. Using motion derived from the monocular visual simultaneous localization and mapping (SLAM), shows that the proposed approaches can provide stable bird's-eye view images, even with large motion during the drive. IPM to accurately transform camera images to bird'seye view images by using motion information for the an adaptive model have been used. (Badeinarayanan, et al., 2016) presented a novel and practical deep fully convolutional neural network architecture for semantic pixel-wise segmentation termed SegNet. This core trainable segmentation engine consists of an encoder network, a corresponding decoder network followed by a pixel-wise classification layer. The architecture of the encoder network is topologically identical to the 13 convolutional layers in the VGG16 network. In this research compared proposed architecture with the widely adopted FCN and also DeepLab-LargeFOV, DeconvNet architectures. This comparison revealed accuracy achieving good segmentation performance. They performed a controlled benchmark of SegNet and other architectures on both road scenes and SUN RGB-D indoor scene segmentation tasks. Their model provided good performance with memory-wise as compared to other architectures. For applications such as understanding a scene, how the visual cues are spatially distributed in an image becomes essential for successful analysis. (Li, et al., 2017) research worked on the network of deep neural networks by accounting for the structural cues in the visual signals. In particular, two kinds of neural networks have been proposed. Firstly, a multitask deep convolutional network, which simultaneously detects the presence of the target and the geometric attributes of the target with respect to the region of interest. Secondly, a recurrent neuron layer is adopted for structured visual detection. Both the networks are demonstrated by the practical task of detecting lane boundaries in traffic scenes. The multitask convolutional neural network provides auxiliary geometric information to help the subsequent modeling of the given lane structures. In their research, the recurrent neural network automatically detects lane boundaries including those areas containing no marks, without any explicit prior knowledge or secondary modeling. An autonomous vehicle must be taught to read the road like a human driver for better controlling the vehicle, so it is important to efficiently detect the road are. (Yang, et al., 2017) focused on the vanishing point detection and its application in inverse perspective mapping (IPM) for road

marking understanding in their researched. In their researched proposed a fast and accurate vanishing point detection method for different types of roads. Vanishing point detection approach gains a better performance than some state-of-the-art methods in terms of accuracy and computation time. Classifying traffic signs is an necessary part of Advanced Driver Assistant Systems. This requires that the traffic sign classification model accurately classifies the images and consumes as few CPU cycles as possible to immediately release the CPU for other tasks. (Aghdam, et al., 2017) in their research, proposed a new method for creating an optimal ensemble of ConvNets with highest possible accuracy and lowest number of ConvNets. It reduces the number of arithmetic operations 88 and 73% compared with two state-of-art ensemble of ConvNets. In their design ConvNets reduces the number of the multiplications 95 and 88% the classification accuracy drops only 0.2 and 0.4% compared with these two ensembles. In their research also proposed a new method for finding the minimum additive noise which causes the network to incorrectly classify the image by minimum difference compared with the highest score in the loss vector. Deep learning has received significant attention recently as a solution to many problems in the area of artificial intelligence. Among several deep learning architectures, convolutional neural networks (CNNs) demonstrate good performance in the applications of object detection and recognition. In general, the process of lane detection consists of edge extraction and line detection. A CNN can be used to enhance the input images before lane detection by excluding noise and obstacles that are irrelevant to the edge detection result. However, training conventional CNNs requires considerable computation and a big dataset.(Kim, et al., 2017) proposed a new learning algorithm for CNNs using an extreme learning machine (ELM). The ELM is a fast learning method used to calculate network weights between output and hidden layers in a single iteration and thus, can dramatically reduce learning time while producing accurate results with minimal training data. A conventional ELM can be applied to networks with a single hidden layer.

#### **CHAPTER 3**

#### **RELATED WORK**

In chapter 3 you may find information about related work and research areas that this thesis focuses on.

#### 3.1 Neural Network

A neural network is an artificial-intelligence processing method within a computer that allows selflearning from experience. Neural networks are a set of algorithms that are designed to recognize patterns. They interpret sensory data through a kind of machine perception, labeling or clustering raw input. The patterns they recognize are numerical, contained in vectors, into which all realworld data, be it images, sound, text or time series, must be translated. Neural networks can also extract features that are fed to other algorithms for assembling and classification. Deep neural networks is an increasingly important area in applied as components of larger machine-learning applications involving algorithms for learning, classification and regression. Neural systems are normally are designed as layers. Layers are composed of various interconnected 'nodes' which contain an 'imposition work'. Examples are fed into to the system through the 'input layer', which conveys to at least one 'hidden layers' the place the real handling is done by means of an arrangement of weighted 'connections'. A multilayer perceptron (MLP) is a class of feedforward artificial neural network. A MLP consist of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that a nonlinear activation function. MLP utilizes uses a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable. Notwithstanding the path that there are a wide extent of sorts learning rules utilized with neural networks, seeming attentive with one it is called the delta rule. It is routinely used by the most remarkable class of ANNs called 'backpropagation neural networks (BPNNs). Backpropagation abbreviated structure for the retrogressive spread of bumble. Delta rule, correspondingly equivalently comparatively likewise with various sorts of backpropagation 'learning' is a supervised methodology that occurs

with each cycle or 'age' (for example the system is given another data plan) through a forward origin stream of yields, and the retrogressive mistake impelling of weight changes.(Thierry, *et al.*, 2019).



Figure 3.1: Information Layers.

In the figure 3.1, leftmost layer in this network is called the input layer, and the neurons within the layer are called input neurons. The farthest right or output layer contains the output neurons, or, as in this case, a single output neuron. The middle layer is called a hidden layer, since the neurons in this layer are neither inputs nor outputs. The term "hidden" represents to '' not an input or an output". The network above has just a single hidden layer, but some networks have multiple hidden layers.



Figure 3.2: Layers Frameworks.

Reasonably confusingly, and for historical reasons, such multiple layer networks are sometimes called multilayer perceptions or MLPs, despite being made up of sigmoid neurons, not perceptions. The design of the input and output layers in a network is often straightforward. In particular, it's not possible to sum up the design process for the hidden layers with a few simple rules of thumb. Instead, neural networks researchers have developed many design heuristics for the hidden layers, which help people get the behavior they want out of their networks. For example, such heuristics can be used to help determine how to trade off the number of hidden layers against the time required to train the network. Such networks are called feedforward neural networks. This means there are no loops in the network - information is always fed forward, never fed back. However, there are other models of artificial neural networks in which feedback loops are possible. These models are called recurrent neural networks. The idea in these models is to have neurons which fire for some limited duration of time, before becoming quiescent. That firing can stimulate other neurons, which may fire a little while later, also for a limited duration. That causes still more neurons to fire, and so over time, get a cascade of neurons firing. Loops does not cause problems in such a model, since a neuron's output only affects its input at some later time, not instantaneously. Recurrent neural networks have been less influential than feedforward networks, in part because the learning algorithms for recurrent networks are less powerful. But recurrent

networks are still extremely interesting. Networks are much closer in spirit to how our brains work than feedforward networks. And it's possible that recurrent networks can solve important problems which can only be solved with great difficulty by feedforward networks (Martin *et al.*, 2019).

### **3.2 Deep Learning**

Deep learning is a kind of Artificial Intelligence that prepares a computer to perform for example, perceiving discourse, recognizing images. Rather than sorting out information to go through predefined conditions, deep learning sets up fundamental parameters about the information and trails. Computer to learn without anyone else by perceiving examples utilizing numerous layers of handling. Deep learning is utilized to order images, perceive discourse, identify items and portray content. Recently, learning has modified the field of computer-based intelligence, for computer vision explicitly. In this approach, an artificial neural network (ANN) is prepared, regularly in an administered way utilizing backpropagation. Deep learning uses a plan with various layers of trainable parameters and has demonstrated momentous execution in intelligence and knowledge applications deep neural networks (DNNs) are readied from beginning to end by using improvement figuring normally subject to backpropagation. The multi-layer neural plan in the primate has impelled masters to concentrate on the significance of non-direct neural layers instead of using shallow frameworks with various neurons. As shown Fig. 3.3 depicts Neural Network plan with a couple hidden layers which separate complex features through progressive layers of neurons equipped with non-straight, differentiable commencement abilities to give a fitting stage to the backpropagation computation.



Figure 3.3: A simple deep neural architecture (Amirhossein et al., 2018)

A great deal of computational power is expected to tackle deep learning issues as a result of the iterative idea of deep learning calculations, their multifaceted nature as the quantity of layers

increment, and the huge volumes of information expected to prepare the systems. The dynamic idea of deep learning strategies – their capacity to constantly improve and adjust to changes in the basic data design – presents an extraordinary chance to bring increasingly powerful conduct into investigation. More prominent personalization of investigation is one possibility, to improve precision and execution in applications where neural systems have been utilized for quite a while. Through better calculations and all the more figuring force, can include more noteworthy .While the present focal point of deep learning methods is in uses of intellectual registering, there is additionally incredible potential in progressively customary investigation applications, for instance, time arrangement examination (Dan *et al.*, 2012).

#### **3.3 Convolutional Neural Network**

Convolutional Neural Network is a class of Neural Systems that have demonstrated in many areas. ConvNets have been effective in recognizing faces, road signs and traffic signs useful for robots and self-driving vehicles. ConvNets, consequently, are a significant instrument for most Artificial Intelligent professionals today.



Figure 3.4: CNN image classification pipeline (Waseem and Zenghui, 2017)

The Convolutional Neural Network in Figure 3.4 is comparable in engineering to the first LeNet and has four different main layers: input, convolutional layer, fully connected layer and classification layer. (The first LeNet was utilized for character recognition). As apparent from the figure above, on accepting an input image as info, Rectified Linearity Unit (ReLU), Pooling or Fully Connected Layer. Conv Net is included a rehashed grouping of convolutional and pooling layers, trailed by thick layers. As the system is prepared, an input picture is feeded into the convolutional layers, at that point into pooling layers, which subsample the picture, decreasing its size, and summing it up further as it enters the consequent convolutional layers. This procedure iterates until a completely associated (thick) layer is achieved which is a huge straight mix of all contributions to a decided number of yields. For the last output layer, the output is a set of qualities called "logits", which state to unscaled scores for every one of the classes. These are then feeded into the softmax layer, a non-direct capacity used to express the class probabilities as a standardized arrangement of probabilities. For a given picture with a known characterization, an error likelihood is estimated for each class. Afterward back-proliferates this error as a misfortune capacity to limit over various cycles. Inevitably the system loads at each layer are tuned to 'see' different qualities of classes through this progressive structure of layers. (Yoon, 2014).CNNs realize pictures piece by piece, pieces that it searches for are called features. By discovering element coordinates in generally similar positions in two pictures, CNNs show signs of improvement at seeing comparability than entire picture coordinating plans. Each element resembles a smaller than expected picture—a little two-dimensional exhibit of qualities. Highlights coordinate normal parts of the pictures. At the point when given another picture, the CNN doesn't know precisely where these highlights will coordinate so it attempts them every position in the image in each conceivable position. In computing the match to a component over the entire picture, we make it a channel. The math used for this task is called convolution, from which Convolutional Neural Systems take their name. LeNet which was one of the absolute first convolutional neural systems. Also, some other designs are as described below:

#### **3.3.1** Alex Net

Alex Net contained eight layers; the first five were convolutional layers, some of them followed by max-pooling layers, and the last three were fully connected layers. It used the non-saturating ReLU activation function.

#### 3.3.2 ZF Net

It is design the hyper parameters to an enhancement for Alex net. ZFNet adjusts the layer hyper parameters such as filter size or stride of the AlexNet and successfully reduces the error rates.

#### 3.3.3 GoogLe Net

Google Net is a pretrained convolutional neural network that is 22 layers deep. Google Net network to perform a new task using transfer learning. It achieved a top-5 error rate of 6.67% this was highest performance for convolutional neural network architectures.

## 3.3.4 VGG Net

VGG net is a convolutional neural network model for image and object recognition dataset can be achieved using a conventional ConvNet architecture with substantially increased depth.

### 3.3.5 ResNets

A residual network is an artificial neural network of a kind that builds on constructs known from pyramidal cells in the cerebral cortex. Residual networks do this by utilizing skip connections, or short-cuts to jump over some layers.

## 3.3.6 Dense Net

is a new CNN architecture that reached results on classification datasets (CIFAR, SVHN, ImageNet) using less parameters.it can be deeper than the usual networks and still be easy to optimize.is one of the latest neural networks for visual object recognition. (Andrew, 2013).Convolutional Neural Networks have been around since early 1990s Convolutional neural networks could tackle became more and more interesting.

#### **3.4 Intelligent Vehicles**

Intelligent vehicles is a field of robotics which has developed in the last 20–25 years. Today, there are approximately 800 million vehicles in the world and this number is upon to two fold in the accompanying 10 years. Intelligent vehicles are described as a vehicle redesigned with acknowledgment thinking and initiating devices that engage the robotization of driving errands. For instance, safe way following, obstacle of slower traffic, following the vehicle ahead, assessing and keeping up a vital separation from risky conditions, and choosing the course. The general motivation of structure sharp vehicles has been to make motoring progressively secure, and progressively supportive and capable. This test has incited the progression of a working examination space with an authoritative goal of robotizing the common errands that individuals perform while driving Advancement of intelligent vehicles in practical situations. In intelligent

vehicles technology has an electronic system that distinguishes possibly unsafe conditions and reacts by either notifying the driver in due time (Claus et al., 2005). Pilot vehicles can communicate to with each other and make it possible to repair in the prosperity and adequacy of the road structure. For traffic condition information, an emergency braking, and sharing of road location, (DSRC) have been set up from the vehicle to vehicle to help getting information from ahead. For example, vehicle-to-vehicle applications and using any kind of media which are 3G, general parcel radio administration (GPRS),Wi-Fi, WiMax, M5, DSRC, satellite...etc. are still in starting periods. A Vehicular Ad Hoc Network (VANET) is a kind of convenient off the network (MANET) to give exchanges to moving at self-assertive, vehicles will as a rule move in a distribute with manner (Ljubo et al., 2001).

#### **3.5 Road Segmentation**

Recent years witnessed, a growing research interest in automated driving systems (ADS) and advanced driver assistance systems (ADAS). As one of the essential modules, road segmentation perceives the surroundings, detects the drivable region and builds an occupancy map (Lyu et al., 2019). A drivable region is a connected road surface area that is not occupied by any vehicles, pedestrians, cyclists or other obstacles. In the ADS workflow, road segmentation contributes to other perception modules and generates an occupancy map for planning modules. Therefore, an accurate and efficient road segmentation is necessary. Road segmentation is a basic task to allow portable robots to navigate. The utilization of vision sensors has a great deal of importance for the road location techniques. The majority of the vision-based approaches depends on organized roads, where highlights, for example, path markings and characterized road limits. Road signs are intentionally organized to help human drivers. Road signs use a great deal of well-described shapes, tones, and models. Regularly this errand is performed by example coordinating strategies, for instance, image cross-correlation, neural networks, or support vector machines since the possible course of action of road signs is compelled and especially described. The road scene understanding may be tended to in a sudden manner, dependent upon the availability of an adroit system. For instance, number of ways, road conditions, geometry, detectable quality, road signs, traffic-light status traffic conditions or system.



Figure 3.5: Road Segmentation (Yecheng et al., 2019)

Road scene understanding, is using data base to utilizing various sensors joined with programmed thinking. The information that is controlled by the system itself which could be made open to the vehicles consolidates definite geometry of the way/road; road signs, status of traffic lights. For example, street conditions, traffic conditions, bike. For road recognition, vehicle models have been outfitted with path recognition and following approaches have focused on recognizing direction markings and structure in the world. For instance, left and right direction markings, the invariance of road width, or the extensively used dimension road assumption. The issue of path following in parkway circumstances is fundamentally a tackled issue. Road signs use a many of well-portrayed shapes and models. The signs are set at solid statures and positions in association with the road. Along these lines examining road signs and shading area schemes declaration occurs. Normally this task is performed by precedent planning frameworks, for instance, image cross-relationship, and neural networks since the possible course of action of road signs is obliged and particularly described. (Gabriel et al., 2016). There are two principles to the road identification algorithms. Firstly, start with process the direction of each image pixel and a versatile casting is connected to discover the vanishing point. Secondly, the principal estimation of the street area applying a prevailing edge recognition technique. Lately, developing investigation plot is seen in automated driving systems (ADS) and advanced driver assistance systems (ADAS) are one of the basic modules road division sees the environment, identifies the drivable area and constructs a map. A drivable area is related street surface zone that is not involved by any vehicles, cyclists or different obstructions. In the ADS work process, street segmentation adds to other observation modules and creates an inhabitance map for modules. Consequently, an exact and accurate road segmentation is essential for vehicles, trucks, and transports. The future of the intelligent vehicles is not completely known, and is most likely going to move in different paths. In any case, as sensor equipped vehicles become logically ordinary, grow opportunities to consolidate sensor-obliging advances with roadways and into vehicles. It will be alluring to extend the radar cross-section of little vehicles, bicycles, to make them more straightforward for radar-arranged automobiles to recognize. Radar-reflecting labels, would cause smaller vehicles to rise even more clearly (Yecheng et al., 2018).

#### **3.6 Semantic Segmentation**

Semantic segmentation, refer to the way toward distribute a semantic name (for example vehicle, road) to every pixel of an image. Semantic segmentation is a testing task that is naming a class with pixel of objects or every locale for assumes a critical issue in image understanding and basic for image investigation assignments. The previous methodologies utilized for semantic segmentation were based on classifiers, though deep learning strategies permitted exact and a lot faster division. Semantic segmentation provides image order, object identification, and limits for restrictions.



Figure 3.6: Semantic Segmentation principle (Jonathan *et al.*, 2015)

Semantic segmentation works spread the movement of new techniques updates of existing networks, and their strategy in new application zones. The significant convincing usage of convolutional neural network which made by LeCun is accomplishment of noteworthy convolutional neural networks (CNNs) has drawn in semantic segmentation. The designed structure which is named as LeNet5 reads postal area, digits, and highlights at different regions in the image. Afterward, discharged a huge deep convolutional neural network (Alex-Net) which is viewed as a standout amongst the most compelling distributions in the field. Alex-Net is a deeper

and more complex compared to the Le-Net, and used to learn complex items and article chains of importance. Zeiler and Fergus introduced the ZF-Net, which is a adapt of the Alex-Net structure. They proposed a method of imaging highlight maps at any layer in the system demonstrate. This procedure utilizes a multi-layered convolutional system to extend the element actuations back to the information pixel space. Multilayer perceptron (MLP) is a scale neural networks which in the network exhibit reliant on little containing various totally connected layers with nonlinear activation another cornerstone neural network architecture is called Google-Net. It diminished the amount of trainigat each layer subsequently and thus, saving computational cost and time. Practically identical producers proposed a consider refereed BN-Initiation for making, arranging, and performing reasoning with Group Systematization procedure. Further presented two new modules Commencement V2 and Beginning stage V3 with some basic changes (using system decline procedure and factorizing convolutions) of their past module. Afterwards, channel link phase of the beginning design with left over associations so as to build effectiveness and execution. A module named is Xception by Chollet (Chollet et al., 2018) proposed. R-CNN first uses particular pursuit to remove a vast amount of item recommendations and afterward figures. CNN highlights for every one of them. At last, it arranges every locale utilizing the class-explicit direct SVMs. CNN structures which are for the most part expected for image order, R-CNN can address progressively convoluted assignments, for example, object detection and image recognition, and it even combines one imperative reason for the two fields. Additionally, R-CNN can be based over any CNN benchmark structures, for example, Alex Net, VGG, Google Net, and ResNet. Secondly; Fully Convolutional Network (FCN) takes in a mapping from pixels to pixels, without separating the locale proposition. The FCN organization is an expansion of the traditional CNN. The primary thought is to make the established CNN take as info subjective estimated images. The limitation of CNNs to acknowledge and create marks just for explicit estimated inputs originates from the completely associated layers which are fixed. In spite of them, FCNs just have convolutional and pooling layers which enable them to make expectations on self-assertive measured inputs. FCN is that by spreading through a few rotated convolutional layers, the goals of the yield highlight maps are down tested. In this way, the immediate expectations of FCN are ordinarily in low goals, bringing about generally item limits. Further developed FCN-based methodologies have been proposed to address this issue, including Seg Net, Deep Lab-CRF, and Widened. (Jonathan et al., 2015). An objectless-attentive semantic segmentation system (OA-Seg) utilizing two structures,

object recommendation plan (OPN) and Convolutional Structure (FCN) made skips forward in imperative learning based semantic segmentation FCN structures have changed into the standard in semantic segmentation; FCN coverts the technique sort out into completely convolutional structure and conveys a probability map for duty of optional size (Yue et al., 2016). FCN recovers the spatial information from the down sampling layers by adding up sampling layers to standard convolution plan. They depicted a skip configuration (hidden layer) that join semantic information from an input layer with appearance segmentation. The basic idea was to re-sketcher and change request appear (image plot) to get competently from whole image of data and whole image .Along these lines stretching out these plans to segmentation and improving the strategy with multi-targets layer mixes (Yuhang *et al.*, 2016).

#### **CHAPTER 4**

#### **MATERIALS AND METHODS**

This chapter introduces semantic segmentation in general and covers the theoretical concepts, terminologies and components that are relevant for the thesis.

### 4.1 Inverse Perspective Mapping

Inverse perspective mapping (IPM) is the method consists of mapping images to a new coordinate system where perspective effects are removed. Over the past years, inverse perspective mapping has been successfully applied to several problems in the field of Intelligent Transportation Systems. Inverse Perspective mapping (IPM) which is used in vision based road estimations algorithms as a pre-processing component. IPM uses information from the camera's position and orientation towards the road to produce a bird's eye view image, where perspective effects are removed. The correction of perspective allows much more efficient and robust road detection, lane marker tracking, or pattern recognition algorithms to be implemented. IPM has been employed in many other Advanced Driver Assistance Systems (ADAS), free space estimation, pedestrian detection, and obstacle detection (Nastaran and José, 2017). IPM has been utilized not just to identify the vehicle's situation as for the road, but likewise in numerous other applications. IPM produce bird's-eye view images that remove point of view impact by utilizing data about camera parameters and the connection between the camera and the ground IPM network for better road understanding. Experimental studies (Robert et al., 2007) proposed that drivers use some notable focuses to control the vehicle. For example, road limits, road markings, and even trenches and tire tracks left by past vehicles, seem to combine into the vanishing point in the image space. IPM, which maps the image data from the image facilitate network to this present reality arrange network and structures a 3-D orthographic perspective out and about image. By building up a suitable facilitate network, can make the real road plane relate to the plane in the 3-D world arrange network, along these lines acquiring an inverse perspective on the road surface.



Figure 4.1: IPM Method (Nastaran and José, 2017)

As shown in figure 4.1, IPM change to for the geometric analysis of traffic image and create two sorts of techniques to quantify traffic stream, whose outcomes are more instinctive and precise than the customary strategies. Accordingly, as to grow better intelligent transportation systems (ITS), it is fundamental to display a strategy that not exclusively can take out the point of view from the images, yet in addition is fit for delivering an image from the first image, from which real traffic data can be effectively and precisely separated. It likewise utilized the angle administrator to separate edge data of path markings.





As shown in figure 4.2, in any case use for perspective effects is a good method when the video is gotten under moving conditions. Although this present the truth is 3D, and the image which is taken from the camera has a spot with the 3D condition and mapped to the 2D world by the camera. This method constantly recommends to lose a few information, and the need to search for huge features in this 2D condition that may be used to infer 3D properties of the watched articles when video is used for traffic examination and traffic the board. (Yong *et al.*, 2007).



Figure 4.3: A typical road scene with IPM Method (Oliveira et al., 2015)

The IPM technique consists of transforming the images by mapping the pixels to a new reference frame where the perspective effect is corrected. This reference frame is usually defined on the road plane, so that the resulting image becomes a top view of the road. Figure 4.3, shows an example of a road scene and represents the image produced using IPM.

#### **4.2 Image Datasets**

Datasets consisting primarily of images or videos for tasks such as object detection, facial recognition, and multi-label classification. Many datasets are useful especially in intelligent vehicles systems and road segmentations. Datasets aiming for vehicle applications are as follows; CamVid dataset which is considered as the first with semantically which rotates around understanding for semantic urban road scenes. Secondly useful data set is KITTI image datasets which utilized in different computer vision assignments, for example, stereo vision, optical flow, 2D/3D object zone and following; PASCAL VOC is among widely recognized and widely utilized

image dataset in critical learning semantic segmentation; CIFAR contains up to 60,000 images, which has 10 and 100 classes of discreet  $32 \times 32$  images (Jia et al., 2009). Image-net dataset contains more than 14 million fix images, Seg-Track v2 is a video division image-dataset with remarks various things at each packaging, Microsoft product called COCO is a collection of images of complex conventional scenes contains essential basic things ,COCO dataset is considerably good dataset (Jordi et al., 2017).



Figure 4.4: An example image its thing annotations in COCO (Holger et al., 2018)

In Fig 4.3 as shown, indoor condition datasets; NYUDv2 (Nathan et al., 2012) is composed of RGB-D images and video segments from an accumulation of indoor scenes, CornellRGB-D (Yung et al., 2016) contains checked office and home scene point fogs, Scan Net (Angela et al., 2019) incorporates into overabundance of 1500 scenes anno-tasted with 3D camera stance, surface entertainments, and semantic segmentation.



**Figure 4.5:** Overview of our RGB-D reconstruction and semantic annotation framework (Angela *et al.*, 2019)

As shown in Fig 4.5, Stanford 2D-3D (Iro et al., 2017) contains regularly enrolled datasets include accounts of spaces for places driven scene understanding. Datasets; RGB-D Article v2 (Liefeng et al., 2011) containing 2500 images of principal contradicts in 51 courses of action, YouTube Dataset fuses 126 records. Datasets for outside condition; Microsoft Cambridge (Jamie et al., 2011) includes 591 genuine open air scene photos of 21 object classes. The dataset improvement is work intensive, so for the experts the most handy and utilitarian methodology is to utilize existing basic datasets which are delegate for the space to issue. Some datasets have wound up being basic and all things considered utilized by specialists to separate their work and others utilizing standard estimation for evaluation. Dataset choice id a beginning of research and it is an attempting errand. They are giving explicit information, for instance, condition nature, the proportion of classes, planning/testing samples, image objectives, and best shows achieved till date by utilizing semantic segmentation models.

#### 4.3 Visual Geometry Group

Visual Geometry Group (VGG) is a convolutional neural network model for image and object recognition which achieved very good performance on the ImageNet dataset. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. VGGNet utilizes numerous  $3\times3$  convolution in the succession that can coordinate the effect of bigger open fields, for example  $5\times5$  and  $7\times7$  (Karen and Andrew, 2015). It is designed on the core idea that deeper networks are better networks. Though they provide a higher level of accuracy, they have an inherently larger number of parameters (~140M) and use a lot more memory. Visual Geometry Group (VGG) has typically 16-19 layers depending on the particular VGG configuration. In this network smaller filters are used, but the network was built to be deeper then convolutional neural networks.



Figure 4.6: Architecture of VGG (Mircea et al., 2014)

As shown in figure VGG-16 is that instead of having so many hyper parameters use a much simpler network and it have conv layers that are just  $3\times3$  filters with a stride of 1, and with the same padding. In all maxpooling layers use  $2\times2$  filters with a stride of 2.The first two layers are convolutional layers with  $3\times3$  filters, and in the first two layers use 64 filters so end up with a  $224\times224\times64$  volume because using same convolutions height and width are the same. For this reason, this (CONV64)  $\times2$  represents that have 2conv layers with 64 filters. The filters are always  $3\times3$  with stride of 1 and they're always implemented with the same convolutions. Firstly, a pooling layer which will reduce height and width of a volume: it goes from  $224\times224\times64$  down to  $112\times112\times64$ .Secondly, a couple more conv layers. because of using the same convolutions, a new dimension will be  $112\times112\times112\times128$ .Thirdly, a pooling layer is added so new dimension will be  $56\times56\times128$  two-convolutional layers with 256 filters. Finally  $7\times7\times512$  into Fully connected layer (FC) with 4096 units, and in a softmax output one of a 1000 classes(Andrej and Li, 2014).It was demonstrated that the representation depth is beneficial for the classification accuracy, and that state-of-the-art performance on the ImageNet challenge dataset can be achieved using a conventional ConvNet architecture with substantially increased depth.

# 4.4 Segnet

SegNet is a deep encoder-decoder architecture for multi-class pixel wise segmentation for semantic pixel-wise segmentation to relating decoder network search by a pixel-wise layers. The design of the encoder network is topologically low from the 13 convolutional layers in the VGG16 organizer. SegNet, is designed to be a core segmentation engine for pixel-wise semantic segmentation. It is primarily motivated by road scene understanding applications which require the ability to model roads, buildings, cars, and understand the spatial-relationship between different classes such as road and side-walk. In typical road scenes, the majority of the pixels belong to large classes such as road, and building. Hence, the network must produce smooth segmentations. The engine must also have the ability to delineate moving and other objects based on their shape despite their small size. Hence it is important to retain boundary information in the extracted image representation. From a computational perspective, it is necessary for the network to be efficient in terms of both memory and computation time during inference (Vijay et al., 2016).



Figure 4.7: An illustration of the SegNet architecture (Vijay et al., 2016)

SegNet has an encoder network and a corresponding decoder network, followed by a final pixel wise classification layer. As shown in figure 4.5 in the encoder, convolution and max pooling are performed. There are 13 convolutional layers in VGG-16. While doing 2×2 max pooling, the corresponding max pooling indices (locations) are stored. On the other hand, in the decoder, up sampling and convolution are performed and there is also a softmax classifier for each pixel. During up sampling, the max pooling indices at the corresponding encoder layer are recalled to up

sample. Finally, a K-class softmax classifier is used to predict the class for each pixel. These part maps are then convolved with a trainable decoder channel bank to convey thick component maps. The decoder contrasting with the first encoder conveys a multi-channel feature map the way that its encoder data has 3 channels (RGB) (Karen and Andrew, 2014).



Figure 4.8: SegNet predictions on urban and highway scenes (Brostow et al., 2009)

SegNet, a deep convolutional network architecture for semantic segmentation to design an efficient architecture for road scene understanding which is efficient both in terms of memory and computational time. SegNet is more efficient since it only stores the max-pooling indices of the feature maps and uses them in its decoder network to achieve good performance.

# **CHAPTER 5**

#### RESULTS

This chapter presents the results from the different implementations. Graphical results showing the during training and validation are presented for the KITTI road detection tasks. Qualitative image results and results from benchmarking are also presented for the KITTI road detection task and inverse perspective mapping

# 5.1 Results

The process of training the network for classification is presented in Figure 5.1.In these figures, blue and orange lines are for the results of the test set and black lines are the results of the validation set. The algorithm, checks accuracy and the prediction loss for every 50 iterations in the validation set. The figure shows the accuracy percentage of the network predictions from the training data set. Minimum error present in Figure 5.1 is on the training set occurs after 10 and 15 epochs respectively, at which time the state of the network is saved. Evaluating the network on the test set results in 0, 16 prediction loss. The network attempt to make a clear prediction close to the road and no-road. In the graphs blue line represents for training and black line for validation accuracy's a result, this system trained with an accuracy of 88,73%. On the other hand loss is approximately 0,16.



Figure 5.1: Graph of Training and Loss

Simulation results showed in Figure 5.2 that clearly system can understand road and no-road sections. Image results are as follows.

Input Image	Traning Image	Binary Segmentation	Segmentation Image

Figure 5.2: Summary of IPM Images Performance in Algorithm.

This section presents the results from the road detection implementation on the KITTI (Karlsruhe Institute of Technology) data set. KITTI is a design of a car made by Toyota technological institute of Chicago. Advantage of using their autonomous driving platform is that the implementation could be tested on the KITTI road detection test set both for qualitative measures and using the KITTI road detection benchmark. To test the capabilities of the method even further, the network was applied to 300 images from the KITTI dataset. In this test, the accuracy is estimated as 96, 23% and the loss is estimated as 0, 11.Figure 5.3 and 5.4 present a quantitative result of the performance.



Figure 5.3: Graph of the accuracy tests on the images from KITTI dataset.



Figure 5.4: Graph of the loss tests on the images from KITTI dataset.

Simulation results showed in Figure 5.5 that clearly system can understand road and no road sections.Example test images and their results are shown below.

Input Image	Traning Image	Binary Image

Figure 5.5: Summary of the performance on KITTI dataset.

Comparison to previous work has been done using, KITTI road data sets and put them into IPM perceptions. After training and testing the proposed method, results have been obtained in the following relevant 300 IPM images 83, 40% iteration. Figure 5.6 and 5.7 present a quantitative result on the performance.



Figure 5.6: Graph of KITTI IPM images.



Figure 5.7: Graph of KITTI IPM images Loss.

At below Figure 5.8 shows that KITTI images on IPM scheme;



Figure 5.8: Applying IPM in KITTI Road Images.

Figure 5.9 presents a set of images from the test set with predicted classifications by the network used for semantic segmentation. The resulting segmentation correctly distinguish between road and no-road, the network struggles to make clear and stable classifications of labels. Qualitative results of the segmentation network on test images from the KITTI data set. Consisted of images which are superimposed on the original images represent the semantic labels extracted by the network. The class correspondence is assumed to be self-explanatory.



Figure 5.9: Summary of Graphs.

The results for all methods compared together are shown in Tables. The performance averaged over the data set sizes are shown in Table 5.1. In Table 5.2 is the performance for each active learning iteration given for datasets, accuracy, and loss. The method achieves higher performance than baseline the following active learning iteration.

Dataset model	Image	Accuracy (%)	Loss Accuracy (%)
	Numbers		
IPM Road Image(our model)	500	88,73	0,13
Fused CRF	300	79,24	0,16
KITTI Road on IPM	300	83,40	0,10
RES3D-Velo	300	78,34	0,15

Table 5.1: Comparing of Datasets Performances in IPM

 Table 5.2: Comparing of Datasets Performances in Road Images.

Dataset model	Accuracy (%)	Loss Accuracy (%)
KITTI (our model)	96,23	0,11
DDN	93,25	0,13
GWVS	89,04	0,17
HIM	91,06	0,15

#### **CHAPTER 6**

#### **CONCLUSION AND DISCUSSION**

The sixth chapter begins by laying out the findings of this study. It will then discuss the findings of the study. Recommended future work on the topic is also discussed.

### 6.1 Conclusion

The presented study was designed to investigate the road segmentation with SegNet on IPM domain. The tests have been carried out on both the Cityscapes, and KITTI road datasets. Utilizing this methodology, demonstrated that SegNet based semantic segmentation which is tuned to identify general images can be useful to effectively segment road area on IPM image domain. The evidence from this study suggests that a changed system can improve the accuracy at distinguishing road and no-road situations. To measure the precision, the Cityscapes dataset of 500 images were utilized. Hence, the detailed analysis of deep learning methods presented in the thesis offers the potential for in depth understanding and continuing research to improve current methods.

### **6.2 Discussion**

In this thesis, in the 500 test images, the algorithm, created semantic segmentation results based on the two labels: road and no-road. In this thesis, first, 500 test images with their corresponding reference images are first used for training. As provided in the literature review; distinguishing obstacle around the vehicle by optical flow calculation dependent on IPM (Inverse Perspective Mapping) applying optical flow on IPM domain is regarded as less challenging and stated methodology had been proposed and utilized on the semantic segmentation of the road area Yuhang *et al.*, 2016). The present thesis, proposes Segnet dataset performance on road segmentation. A methodology of road recognition calculation based on IPM first transforms input image into the bird's eye view, where borders of a road becomes parallel. Previous research shows that a real-time road detection method based on reorganized LIDAR data (RES3D-VELO) gives less accuracy than segnet for segmentation to road successfully (Valente *et al.*, 2016).One of the more important finding emerged from this study is that the DDN architecture achieves the highest accuracy to time for inference ratio. But this method spent large amount of memory on CPU and its time consuming.

The choice of using the Segnet as a feature extractor in the model architecture was made, since it offers a good trade-off between time for inference and accuracy, and which agrees well with the roads to be detected in the KITTI data set. The implementation presented in this thesis using the training images from the KITTI data set and road IPM images datasets. The most obvious advantage is efficiency, since road centered crops allow for more training examples of classes and better control of the ratio of positive to negative ratio whilst training. Mostly involving memory, time and accuracy.

#### 6.3 Future Work

In this thesis, concentrated to two major aspects: semantic segmentation and inverse perspective mapping using prior information. The tested dataset contains 500 images with their ground truths. In future tests, it might be possible to broaden the dataset either by utilizing the current fine ground truths in the BSD or producing ground truth for new images. For semantic segmentation utilizing training data, made it possible to adjust parameters (weights, and biases) in the SegNet network. The use of training parameters in these networks has resulted in performance gains and this effect can be observed in analysis of various models. However, less attention has been paid to smaller and more efficient models for semantic segmentation such as road image understanding (Huang *et al.*, 2018). For future work, it is desired to be focused on smaller and but efficient, for tasks such as road image understanding compared to the competing architectures.

### REFERENCES

- Andrej, K., & Li, F., (2014). Deep visual-semantic alignments for generating image descriptions. *CoRR*, doi: abs/1412.2306.
- Angela, D., Angel, C., Manolis, S., Maciej, H., Thomas, F., & Matthias, N. (2019). Scannet: Richly-annotated 3d reconstructions of indoor scenes. *In Proceedings of the CVPR*, doi: https://code.google.com/archive/p/scanvnet/.
- Badrinarayanan, V., Handa, A., & Cipolla, R. (2015). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling. *Computer Science*, doi: http://arxiv.org/pdf/1505.04366.pdf.
- Gabriel,O.,Wolfram,B.,Thomas,B.(2016).Efficient deep methods for monocular road segmentation. International Conference on Intelligent Robots and Systems.doi: https://doi.org/10.1177/0278364917710542.
- Graves, Alex and Schmidhuber, J<sup>•</sup>urgen., (2009). Offline handwriting recognition with multidimensional recurrent neural networks. *In Advances in Neural Information Processing Systems* doi: https://code.google.com/archive/p/cuda-convnet/.
- Gregor, Karol, Danihelka, Ivo, Graves, Alex, and Wierstra, Daan (2015). DRAW: A recurrent neural network for image generation. *Proceedings of the 32nd International Conference on Machine Learning*. doi: 10.1145/3313866.
- He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun,Jian (2015). Deep residual learning for image recognition. *ArXiv preprint*, doi: 1512.03385.
- Holger, C., Jasper, U., Vittorio, F. (2018) COCO-stuff: thing and stuffclasses in con- text, *IEEE/CVF*. Conference on Computer Vision and Pattern Recognition CVPR 2018, 1209-1218.
- Iro, A., Sasha, S., Amir, Z., Silvio, S., (2017). Joint 2D-3D-semantic data for in- door scene understanding, *CoRR*, doi:abs/1702.01105.
- Jamie,S., John, W., Antonio, C., Carsten, R. (2006). Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmen- tation, in: Proceedings of the European Conference on Computer, *Vision*, doi: 10.1007/11744023\_1.

- Jason, Y., Jeff, C., Yoshua, B., Hod, L. (2014). How transferable are features in deep neural networks *Advances in Neural Information Processing Systems*, *27*, 3320–3328.
- Jeffrey, D., Greg, C., Rajat, M., Kai, C., Matthieu, D., Quoc, L., Mark, M., Marc, R., Andrew, S., Paul, T., Ke, Y., Andrew, N, (2012) Large scale distributed deep networks. *In NIPS*, doi: https://code.google.com/archive/p/cuda-convnet/.
- Jia, D., Wei,D., Richard, S., Li-Jia, L., Kai, L., Li ,F. (2009). Imagenet: A large-scale hierarchical image database, in: Proceedings of the IEEE Conference on Com-puter Vision and Pattern Recognition, CVPR, doi: 3297001.3297037.
- Jonathan, L., Evan, S., Trevor, D., (2015). Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition, doi: 1411.4038.
- Jordi, T., Federico, P., Sergi, C., Pablo, A., Alexander, H., Van, G. (2017). The 2017 davis challenge on video object segmentation *arXiv preprint*, doi: 1704.00675.
- Karen, S & Andrew, Z. (2014). Very deep convolutional networks for large-scale image recognition, *CoRR*, doi: abs/1409.1556.
- Karen,S., & Andrew, Zisserman.,(2014). Very deep convolutional networks for large-scale image recognition, *CoRR*, doi: abs/1409.1556.
- Karen,S., Andrea, V., Andrew, Z. (2013). deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint*, doi: 1312.6034.
- Karen,S.,&Andrew,Z.,2015.Very Deep Convolutional Networks for Large-Scale Visual Recognition, doi: 1409.1556 2015.
- Kristen, G., & Suyog, J., (2014), Supervoxel-consistent foreground propagation in video, In Proceedings of the European Conference on Computer Vision, Springer, doi: 10.1007/978-3-319-10593-2\_43.
- Krizhevsky, A., Sutskever, I., & Hinton, G.E., (2012). Imagenet classification with deep convolutional neural networks. *In Advances in Neural Information Processing Systems*, doi: https://code.google.com/archive/p/cuda-convnet/.

- Liefeng, B., Dieter, F., Kevin, L., & Xiaofeng, R. (2011). A large Scale Hierarchical Multi-view RGB-D Object Dataset in Proceedings IEE International Conference on Robotics and Automation doi:10.1109/ICRA.2011.5980382.
- Martin, B., David, B., Dania, H., Alistair, K., & Sebastian, O. (2019). Learning, planning, and control in a monolithic neural event inference architecture. *Computer Science*, doi: https://doi.org/10.1016/j.neunet.2019.05.001.
- Miguel, O., Vitor, S., & Angel, S. (2014). Multimodal inverse perspective mapping. *Information Fusion, doi: 10.1016*/j.inffus.2014.09.003.
- Nastaran, E., & José, M., (2017). A New Strategy of Detecting Traffic Information Based on Traffic Camera: Modified Inverse Perspective Mapping, *Journal of Electrical Engineering & Technology*, doi: http://doi.org/11.5271/JEETIU.2017.10.2.984.
- Nathan, S., Derek, H., Pushmeet, K., & Rob., F.(2012). Indoor segmentation and support inference from RGBD images, In Proceedings of the European Conference on Computer Vision, Springer, doi:7576LNCS.
- Robert, B., Hendric, W., Van,A, Jan, B., Jessie, T., Joachim, M. (2007). A vision based row detection system for sugar beet, *computers and electronics in agriculture* doi:10.1016/j.compag.2007.07.006.
- Thierry, B., Sajid, J, Maryam, S.,Soon, J. 2019). Deep neural network concepts for background subtraction: A systematic review and comparative evaluation.doi: https://doi.org/10.1016/j.neunet.2019.04.024.
- Vijay, B., Alex, K., Roberto, C., (2015) Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *Institute of Electrical and Electronics Engineers*, doi: 10.17863/CAM.17966.
- Vijay, Badrinarayanan., Ankur, H., Roberto, C., (2015). Segnet: A deep convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling. *Computer Science*, doi: http://arxiv.org/pdf/1505.04366.pdf.
- Yecheng, L., Lin, B., & Xinming, H. (2019). Road Segmentation Using CNN and Distributed LSTM. arXiv, doi:1808.04450v2.

- Yecheng, L, Lin, B., & Xinming, H. (2018). Real-time road segmentation using lidar data processing on an fpga. In Circuits and Systems IEEE International Symposium .doi: arXiv:1711.02757v1.
- Yong, S., Fan, Y., & Runsheng, W. (2007). Color space selection for moving shadow elimination. International Conference on Image and Graphics, doi: 10.1109/ICIG.2007.54.
- Yue, W., Hong, L., Jun, L., & Yun, F. (2016). Objectness-aware semantic segmentation, in: Proceedings of the ACM on Multimedia Conference. Association for Computing Machinery, doi: https://doi.org/10.1016/j.imavis.2018.09.010.
- Yuhang, W., Jing, L., Yong, L., Junjie, Y., & Hanging, L. (2016). Objectness-aware Semantic Segmentation. Proceedings of the 24th ACM international conference on Multimedia, ,doi: 10.1145/2964284.2967232.
- Yung, H. ,Shao, C., & Yen, K. (2016). Efficient Surface Detection for Augmented Reality on 3D
   Point Clouds In Proceedings of the Advances in Neural Information Processing
   Systems. Association for Computing Machinery, doi: 10.1145/2949035.2949058
- Zhang, Y., Chen, G., Yu, D., Yao, K., Khudanpur, S., & Glass, J. (2016). Highway long shorter memory RNNs for distant speech recognition. *In Proceedings of the International Conference on Acoustics,Speech and Signal Processing,*, doi: https://arxiv.org/abs/1510.08983.