ABSTRACT

KEY WORDS: Semantic Segmentation, Fully-Convolutional Neural Networks

Over the past 100 years, Automobile Industry has seen steady innovation with the help of Technology on par with other fields. Self-driven cars have taken the world by storm and its efficiency has been proved to be better than calculated. Despite its popularity, its yet to be tested and green-lit for Indian roads, and that is the motivation behind this project. This demonstration is a scaled down approach to the world of Computer vision that deals in Vehicle Automation.

The central processing unit of this project is NVIDIA Jetson Nano; it is a compact AI compute Module which can be used to learn and create intuitive Deep Learning based applications.

The main 3 objectives of the project are:

- To Apply Convolutional neural networks to perform object detection.
- To Train the application and increase scalability using Transferred Learning technique to make it work in native environment.
- To Experiment with Fully-Convolutional semantic segmentation networks on a live camera stream.

TABLE OF CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

AI	Artificial Intelligence
FCN	Fully convolutional network
ARM	Advanced RISC Machines
CUDA	Compute Unified Device Architecture
GPU	Graphic processing unit
SDK	Software development kit
DNN	Deep neural network
VGG	Visual Geometry Group
LIDAR	Light Detection and Ranging

CHAPTER 1

INTRODUCTION

A self-driven car is a concept that is being actively researched and tested for usage on the road. Furthermore, as time passes, the machine learning techniques required to build such a vehicle have become increasingly accessible to the general public. With a variety of libraries and software for designing and training neural networks accessible, as well as affordable but powerful tiny computers on the market, developing a self-driving vehicle can be researched.

The 3 pillars this project relies primarily on are:

- Image Classification: Classify the objects within an image.
- Object Detection: Classify and detect the object(s) within an image with bounding box(es).
- Semantic Segmentation: Classify the object class for each pixel within an image.

Image recognition is used in semantic segmentation, but classifications are done at the pixel level rather than the entire image. Convoluting a pre-trained image recognition backbone transforms the model into a Fully Convolutional Network (FCN) capable of per-pixel labelling.

“ NVIDIA’s Jetson Nano with quad-core ARM CPU, 4GB of RAM, and most importantly 128 CUDA cores GPU is used that would allow to run neural networks in real-time. **“**

The NVIDIA® Jetson Nano Developer Kit is a compact, powerful computer that's perfect for learning about AI. Jetson Nano paves the way for robotics and edge deployment of deep learning

for real-time image classification, object identification, segmentation, audio processing, and more. It's the ideal platform for beginning AI projects and studying popular machine learning frameworks like PyTorch and TensorFlow for learners, makers, and developers.

Today, driver-less cars are a reality after constant research and development.
Still, there are many challenges in designing a fully autonomous algorithm for driver-less cars.

The primary challenges faced are:

1. Road conditions: Road conditions could be highly uncertain and vary hugely. In some cases, there are smooth and marked broad highways. In other cases, road conditions are deteriorated.

2. Weather conditions: Weather conditions play another spoilsport. There could be a sunny and clear weather or rainy and stormy weather. Autonomous cars should work in all sorts of weather conditions.

3. Traffic conditions: Autonomous cars would have to get onto the road where they would have to drive in all sorts of traffic conditions. They would have to drive with other autonomous cars on the road, and at the same time, there would also be a lot of humans.

4. Accident Liability: The most important aspect of autonomous cars is accidents liability.

5. Radar Interference: Autonomous cars use lasers and radar for navigation. The lasers are mounted on roof top while the sensors are mounted on the body of the vehicle.

CHAPTER 2

Literature survey

Real-Time Semantic Segmentation for Autonomous Driving

The majority of semantic segmentation research focuses on improving accuracy,with computationally efficient solutions receiving less attention. The majority of effective semantic segmentation algorithms contain unique optimizations that aren't scalable, and there's no way to compare them in a systematic fashion.

This research examines various segmentation algorithms For autonomous driving and proposes a real-time segmentation benchmark system.

The basic goal is to do pixel-by-pixel categorization of the image in order to understand the scenario. However, several features of semantic segmentation, such as computing efficiency, have applications such as autonomous driving.

Fully-Convolutional networks covers the core corpus of work on deep learning-based semantic segmentation.

Semantic Segmentation using Deep Neural Networks

Object recognition is not required for segmentation, it is a step deeper than object recognition. Humans, in particular, may segment images without even understanding what the items are (for example, there may be multiple unknown things in satellite photography or medical X-ray scans, but they can still be segmented within the image for future examination).

Performing segmentation without knowing the exact details of all objects is a important aspect of the visual processing process, and it can be used to improve or complement existing computer vision algorithms. Image segmentation is one of the most difficult topics in computer vision.

Unlike image classification or object detection, segmentation does not require prior knowledge of the visual concepts. To be more straightforward, an object classification will only classify objects for which it has specific labels, such as a horse, an automobile, or a house.

NVIDIA Computer Vision

Computer vision is a branch of science that allows smart cameras to acquire, process, analyze, and interpret images and movies. For example, a vehicle's driver assistance system use cameras and other sensors to not only display, but also sense what's in front of and behind it in order to detect and classify regions or points of interest inside an image frame.

In this scenario, computer vision serves as a safety feature, assisting the driver in navigating around road debris, other vehicles, animals, and pedestrians. Similarly, farmers may use computer vision-enabled devices to identify weeds and where crops are growing well across a broad area in order to boost productivity. Artificial intelligence and, more specifically, deep learning, a sort of machine learning modeled after the brain, are used in today's computer vision jobs.

Computer vision models based on deep learning enable devices to perform and adapt like a human expert while requiring substantially less input.

The majority of computer vision techniques start with a model, or a mathematical algorithm, that has been trained on large amounts of data to do a given task. The following are a few examples of frequent techniques:

1. Classification
2. Detection
3. Segmentation
4. Image Synthesis

CHAPTER 3

METHODOLOGY

3.1 PROPOSED FRAMEWORK

The primary intentions of the project are:

- Applying CNN to perform object detection.
- Training the application and increase scalability using Transferred Learning technique to make it work in native environment.
- Experimenting with fully-convolutional semantic segmentation networks on a live camera stream.

NVIDIA TensorRT is the DNN library being used for efficiently deploying neural networks onto the embedded Jetson platform, improving performance and power efficiency .
The Transfer Learning with PyTorchis used for training DNNs in Jetson module.

The steps involving setting up Jetson and building the project are as follows:

- Interfacing Jetson Nano using Jetpack SDK.
- A camera which allows to interact with objects is required..
  In the demonstration, a Logitech C270 HD webcam is used.
- Pre-trained models are used for experimenting with the test data and gather inferences.
- Nvidia GPU Cloud provides models as per the requirements through Docker Container which is deployed onto Jetson Nano.

- Coding and developing image recognition and object detection applications using python by loading pre-trained models is demonstrated.

- Further, Transferred Learning concept is used to train pre-existing models to increase scalability and adapt to native environment.

- Based on in-depth analysis and inferences, developing applications to perform semantic segmentation using image data is demonstrated.

3.2 IMAGE PROCESSING

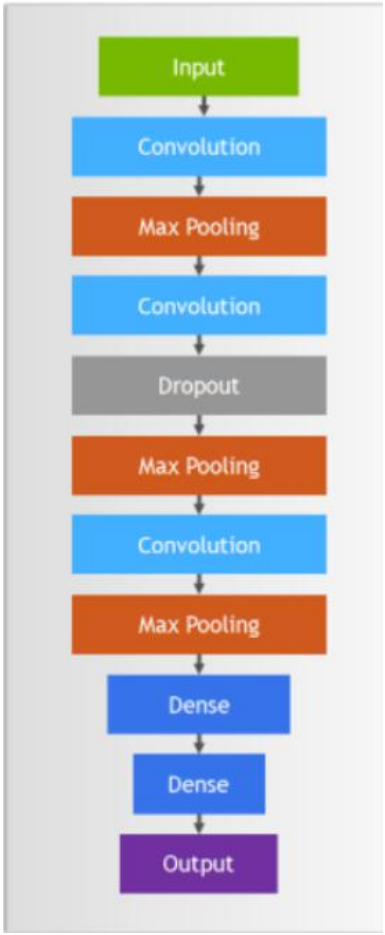
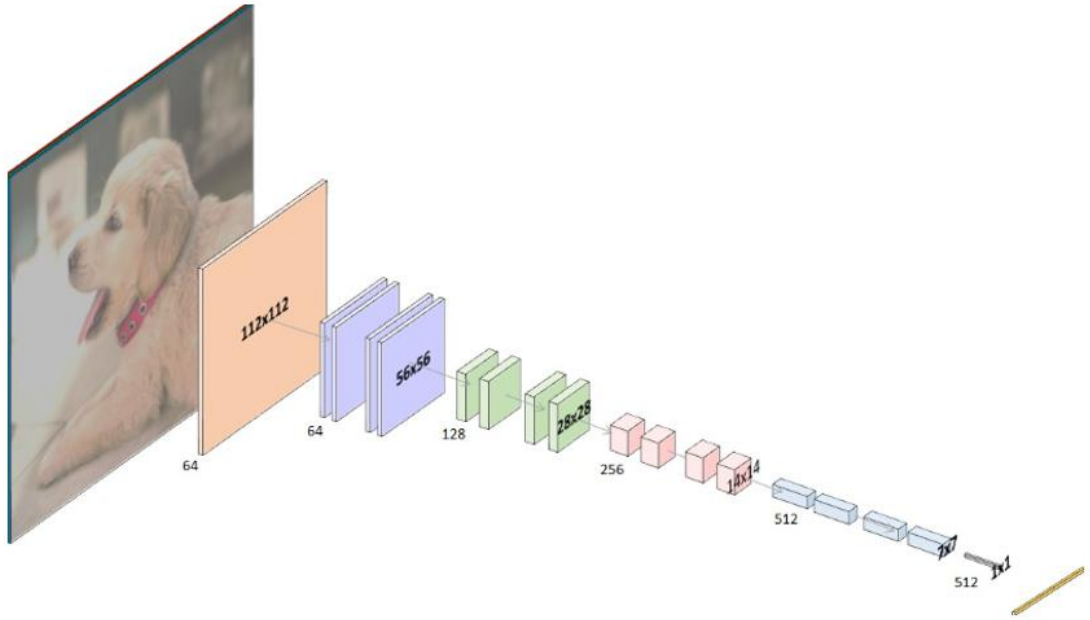


Image processing by neural network

The video inputs are images presented at a fast rate. Images are matrices with integer values that represent distinct color intensities. Attempting to solve these formerly unsolvable challenges with semantic image segmentation with the introduction of Deep Learning algorithms and increased processing capacity.

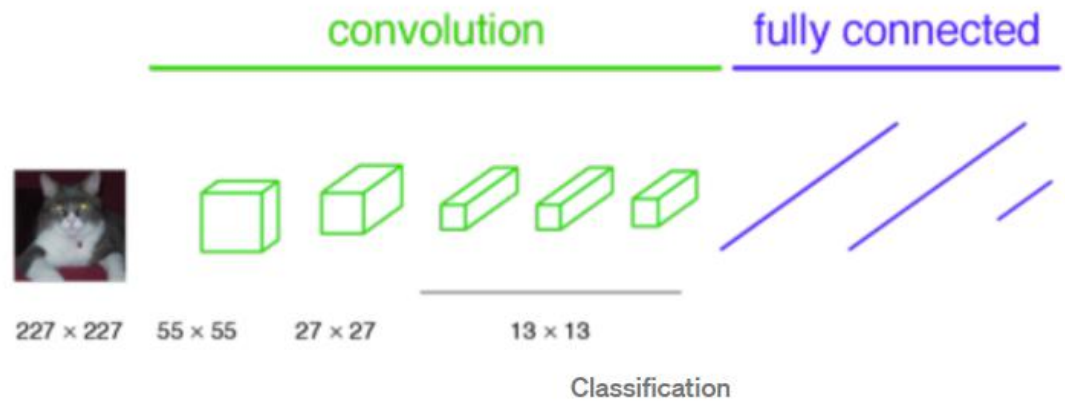
Its high accuracy, however, comes at the expense of higher computing expenses, making it unsuitable for embedded sensors in self-driving automobiles.

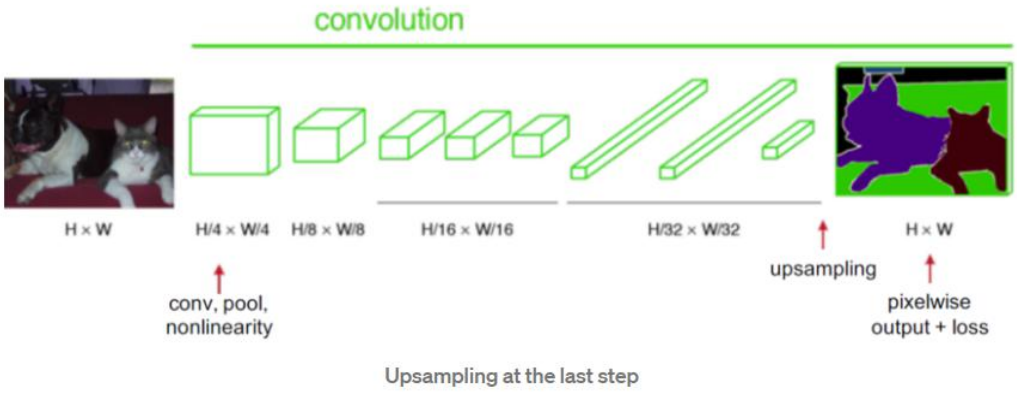


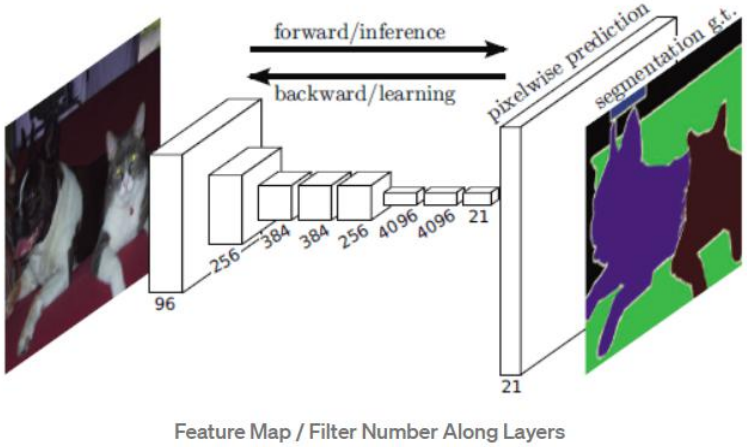
Pre-Trained model architecture

3.3 Image Classification to Semantic Segmentation

In classification, an input image is downsized and goes through the convolution layers and FC layers, and outputs one predicted label for the input image.

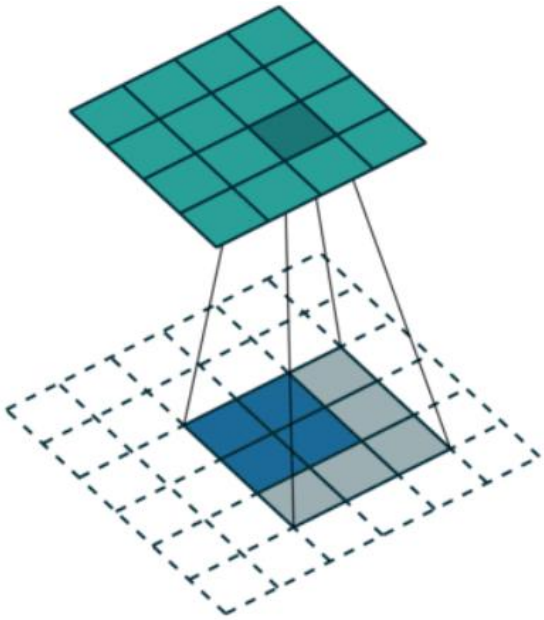






3.3.1 Up-sampling Via Deconvolution

Convolution is a process getting the output size smaller. Thus, the name, deconvolution, is coming from when we want to have up-sampling to get the output size larger.

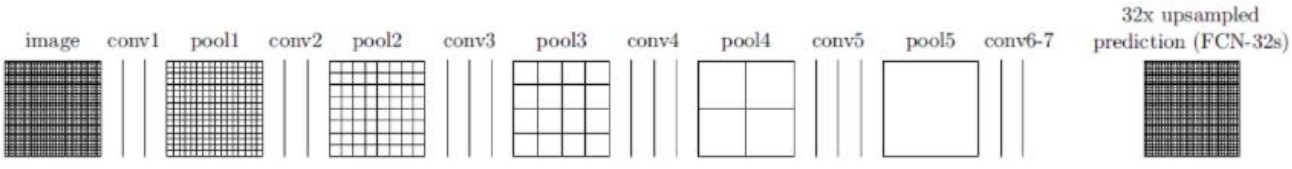


Upsampling Via Deconvolution (Blue: Input, Green: Output)

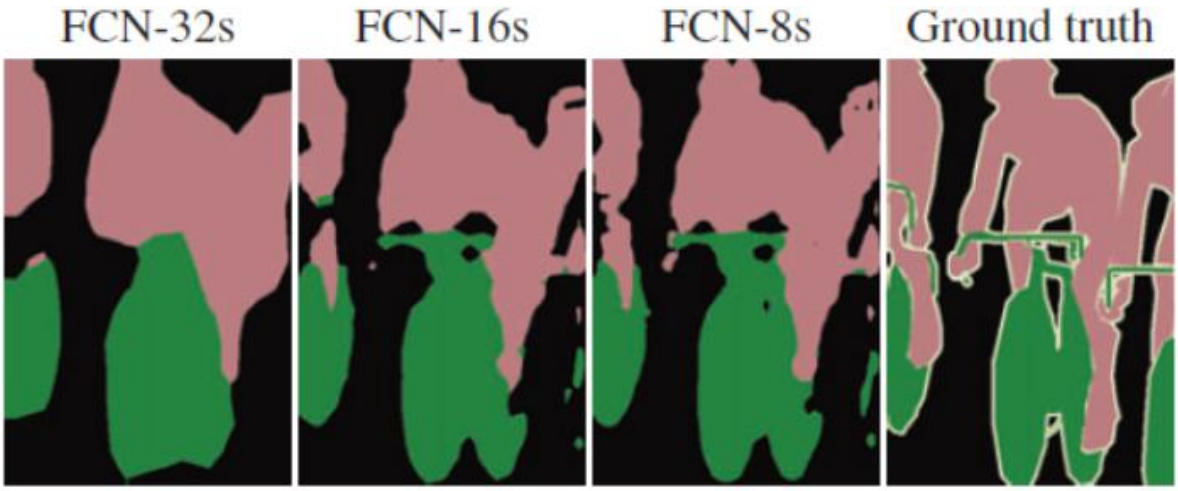
3.4 Element-wise addition

After going through conv7 as below, the output size is small, then 32× up-sampling is done to match the size of input and output. But it makes the output label map rough and is called **FCN-32s.**

FCN-32s result is very rough due to loss of location info while FCN-8s has the best result.



FCN-32s

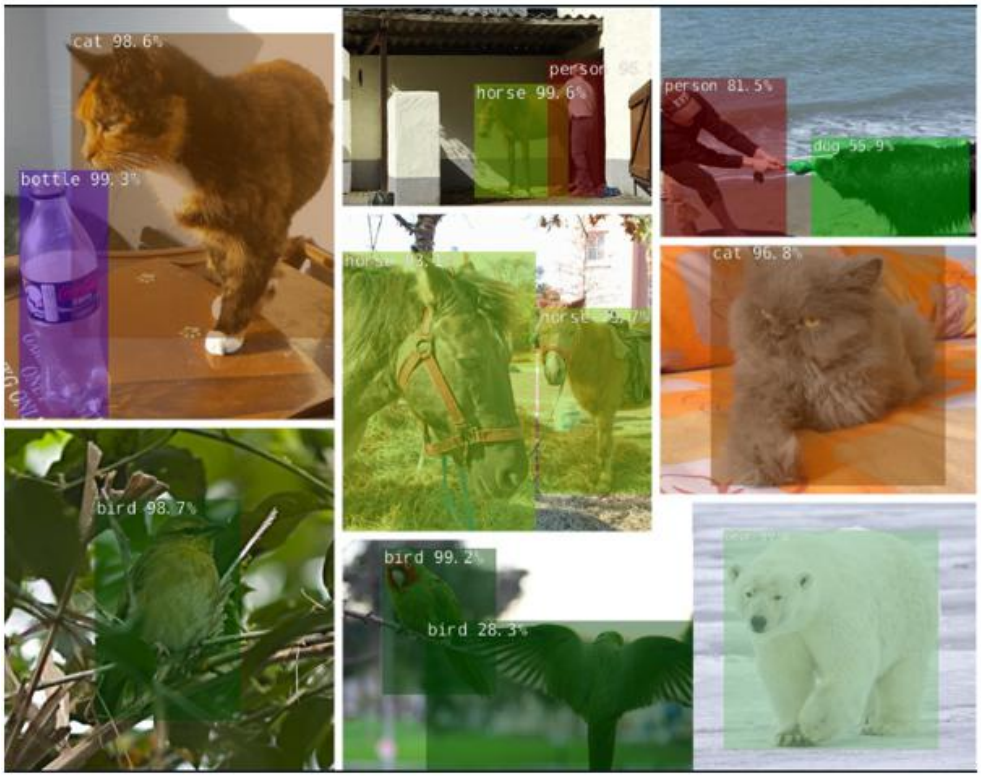


Comparison with different FCNs

The low resolution output due to down-scaling is usually managed by employing skip-connections from lower layers to the output, which increases resolution at layers near the output. The Fully-Convolutional Network (FCN), which currently serves as a blueprint for most subsequent techniques, is the first to implement skip-connection. The sole difference between these systems is how object level information is encoded and how decoding these classifications into pixel-exact labels is performed.

Then, using a transposed convolution, FCN added information to upper layers that came from lower layers. Alternative techniques to link to the lower layers improved the FCN design. e.g. Pooling procedures for dilated convolutions can be avoided by accessing lower-layer pooling layers, employing better ways to integrate lower-level information, or foregoing pooling operations altogether.

3.5 RESULTS



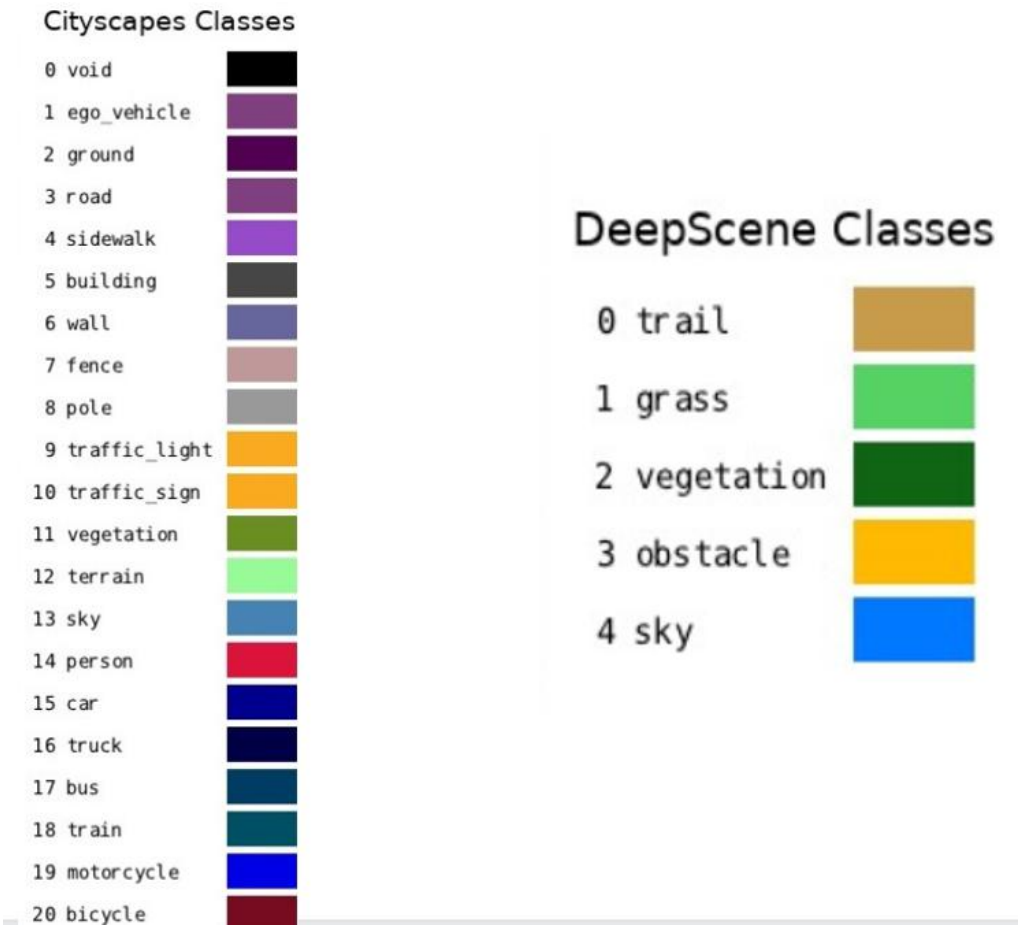
Expected image classification and object detection outcome

Expected semantic segmentation outcome

Semantic Segmentation Implementation

Final Outcome Demonstration

3.5.1 Classes and Models



Each class is represented by a separate integer in the segmented data, which is an image. Unlabeled pixels are set to 255, while labeled pixels range from 0 to 18, denoting different classifications (road, car, pedestrian, sign, train, etc.). It pre-processes the data by setting unlabeled pixels to 19, resulting in 20 distinct classes represented by consecutive integers, which is a more common representation of segmented data.

Semantic Segmentation Models

- **Cityscape**: assessing the performance of vision algorithms for major tasks of semantic urban scene understanding.

- **DeepScene**: It consists forest tracks,trees etc.
- **Multi-Human Parsing (MHP)**: Provides labeling of body parts and clothing.
- **Pascal VOC**: It contains various people, animals, vehicles, and household objects.
- **SUN RGB-D**: Indoor objects and scenes in office spaces and homes.

3.5.2 Prediction Accuracy

	pixel acc.	mean acc.	mean IU	f.w. IU
FCN-32s-fixed	83.0	59.7	45.4	72.0
FCN-32s	89.1	73.3	59.4	81.4
FCN-16s	90.0	75.7	62.4	83.0
FCN-8s	**90.3**	**75.9**	**62.7**	**83.2**

Dataset	Resolution	Accuracy(FCN-32)	Frame Rate
Cityscape	512*256	83%	48 FPS
DeepScene	576*320	96%	26 FPS

CHAPTER 4

Conclusion and Future work

People believe that developing a self-driving car is not a difficult task, yet driving is one of the more complex actions that humans engage in on a daily basis. Following a set of road rules isn't enough to drive like a person, because unforeseen scenarios have to be dealt with, react to weather conditions, and make decisions that go against the rules to prevent harming human life. One of the most actively investigated fields in the age of artificial intelligence is self-driving cars in the automobile industry. Many global corporations are devoting significant human and financial resources to the development of hardware and software to serve this purpose. To accomplish complete autonomous driving, several components, such as LIDAR's, cameras, sensors, and the algorithms that operate behind them, must function closely together.

In autonomous driving, image segmentation is critical. The training datasets are the most significant data in the development of self-driving cars since without them, no such car would be able to grasp and make estimations. To avoid mishaps, self-driving automobiles require proper pixel knowledge of their surroundings. Object identification using bounding boxes is not capable of providing the pixel-perfect recognition necessary.

Semantic segmentation is a difficult task when using computer vision to interpret images. While doing picture segmentation, two key challenges are faced: maintaining high standards of consistency inaccuracy and precision level.

To overcome these obstacles and achieve excellent AI model performance at scale, using an image labeling tool is the best option because it provides simplicity and efficiency while delivering high-quality training data, allowing you to save time, money, and improve the annotation process' efficiency.

So far this project has achieved autonomous segmented vision solely in the software division. Applying this concept on real environment and testing its efficiency requires adequate research. A 360° view of the surroundings along with accurate calibration is required, involving inputs from the hardware division.

REFERENCES

1. **Mostafa Gamal Badawy,Cairo University; Moemen Abdelrazek,Cairo University;**
   A Comparative Study of Real-Time Semantic Segmentation for Autonomous Driving,
   The 14th IEEE Embedded Vision Workshop, CVPR 2018 At: Salt Lake City, Utah, USA
   https://www.researchgate.net/publication/324866024_A_Comparative_Study_of_Real Time_Semantic_Segmentation_for_Autonomous_Driving

2. **Yanming Guo, Yu Liu, Theodoros Georgiou & Michael S. Lew,**
   A review of semantic segmentation using deep neural networks,
   International Journal of Multimedia Information Retrieval
   https://link.springer.com/article/10.1007%2Fs13735-017-0141-z

3. ***NVIDIA GPU Cloud***
   https://www.youtube.com/playlist?list=PL5B692fm6--up8j7qWID9cluY24vky3Xw

4. ***Getting Started with AI on Jetson Nano***
   https://courses.nvidia.com/courses/course-v1:DLI+S-RX-02+V2/about

5. ***NVIDIA Computer Vision***
   https://developer.nvidia.com/computer-vision

6. ***FCN — Fully Convolutional Network (Semantic Segmentation)***
   https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1