ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 8, August 2021

Analysis and Implementation of Computer Vision Techniques for Autonomous Driving

[1] Riya Chougule, [2] Amruta Bhalerao, [3] Sanika Abhyankar, [4] Kajal Pompatwar, [5] Jitendra Chavan

[1][2][3][4] Department of Information Technology, MMM College of Engineering, Pune, India
[5] Assistant Professor, Department of Information Technology, MMM College of Engineering, Pune, India

***Abstract*---- Autonomous driving technology is one of the fastest-growing technologies in these years. Self-driving vehicles help to reduce the number of road accidents happening every year. Various organizations, startups, and researchers are working on this technology. Due to the advances in the field of machine learning in past decades, a major push is received in the field of computer vision and various techniques like object detection, semantic segmentation, etc. are developed. A Large number of open-source datasets are available for training autonomous driving systems. The review intends to provide a deep survey about different computer vision techniques, architecture, and various datasets used for Autonomous driving technology.**

***Keywords*— Self-driving vehicles; Machine learning; Computer vision; Dataset, semantic segmentation**

I. INTRODUCTION

According to human behavior, driving is one of the most difficult tasks performed by human beings. Many accidents are incurred just by impaired driving. Just by following the rules is not enough to drive a car for us [1]. Because we react to different unexpected situations or weather conditions and make decisions according to them for self-driving. To avoid this, autonomous driving came into the picture. An autonomous car or self-driving car does not require any human involvement and drives safely by automatically sensing its surrounding environment. Autonomous vehicles depend on different sensors like GPS, lidar, radar, high-resolution cameras, etc. to study the environment.

Researchers and engineers are using computer vision technology to build an autonomous driving system to overcome road obstacles while driving by keeping passengers safe [8]. Even a minor defect in this system can cause fatal accidents. To eliminate human touch, computer vision technology uses different algorithms to outline a map of its surroundings. Autonomous driving is one of the most inspiring applications of computer vision technology. Computer vision is a set of techniques used to interpret image-based data. It uses an artificial neural network to build a system that can understand the most important information in the scene and filter out the noise. In deep learning, artificial neural network algorithms are used to imitate the human brain, which is used to give decision-making power to the model [8]. In this paper, we are studying different computer vision techniques which are used to build the model and detect objects for an autonomous driving system with the help of different datasets based on our project. Also, we gathered **the Indian dataset by actually capturing the photographs** of Indian streets, traffic lights, cars, pedestrians, bicycles, sky, trucks, etc. Thus, we will be testing our model for image segmentation for Indian streets and making a video(series of images) as a real-world application.

The flow of the paper is mentioned in the image below:

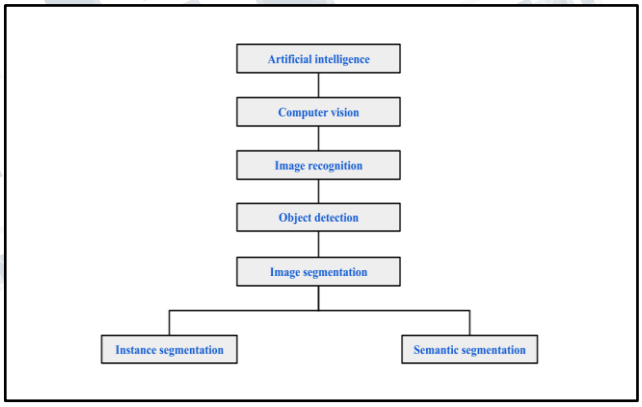

Fig. 1. Computer Vision Techniques

II. ARTIFICIAL INTELLIGENCE

Artificial Intelligence(AI) means combining the working and thinking of the human brain to make a powerful decision-making machine. As machines are becoming incredibly strong in performing difficult tasks as per human requirements. Right now, machines are taking human orders for the completion of tasks. What if the machine will not require you to take orders and perform tasks automatically? This is the power of artificial

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 8, August 2021

intelligence. Artificial Intelligence opens a new way for many industries and businesses. A lot of companies are using computer vision to increase their productivity. Basically, computer vision is used by machines so that they can work like the human eye and make decisions according to what they see.

III. COMPUTER VISION

Computer Vision works as the eyes of Self-driving vehicles to describe how the world around it looks like. It enables the system to see and retrieve meaningful information from digital images, videos, or any form of input [9].

IV. IMAGE RECOGNITION

Image recognition is a part of computer vision used to identify different parameters like the place, weather, obstacles, etc. in an image. The main role of this in our system is to understand the parameters and context of an image like a human does. Image recognition uses a neural network for image identification [13]. It tells the machine to interpret what they see in the given input image and categorize them according to their classes. In image recognition, models are trained to take an input image and output the labeled image describing it. Image recognition is an important machine learning task. It is used for other computer vision techniques such as object detection, image segmentation, etc. It is used to improve user's experience by using visual search, automated image, and video tagging and for marketing purposes.

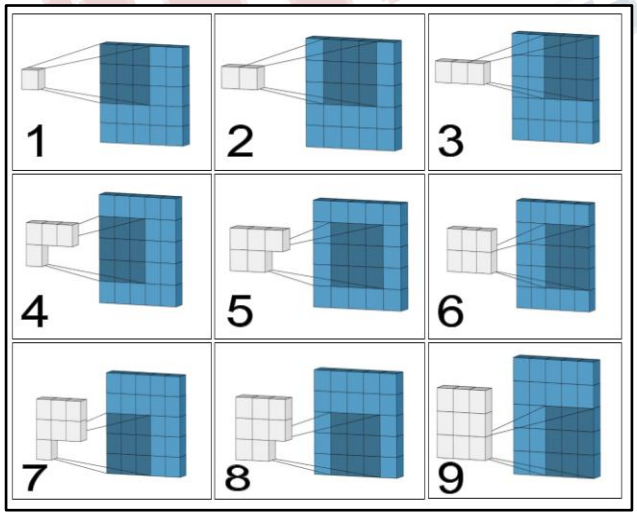

Fig. 2. Image Recognition

Let's see how an image is stored mathematically. For this purpose, we will consider that we have an image of size 5 x 5(blue box shown in image) and a filter of size 3 x 3(white box shown in image). So, the filter is plotted on the upper left corner of the image(box "1" shown in the image), and then we will calculate the **dot product** of the values and store the result. Then the filter will slide to the next location(box "2" shown in the image).

Here the interval at which we can slide is called a **stride**. In the image above, we have used a stride as 1, which means we are sliding the filter over the image with 1 interval and storing the corresponding dot product in a matrix that stores the information of that particular image. We can compose this process into a formula :

The size of the output image to be obtained = (N-F) /stride + 1

Where **N**: Size of the input image
F: Size of the filter

For example, consider the above-mentioned image, here, the size of the input image is 5 x 5, and the size of the filter is 3 x 3 with stride as 1,
N = 5
F = 3
Stride = 1
The size of output image obtained = (5-3)/1+1
= 3
Thus, we will get an output image with the size 3 x 3.
As our input image was of the size 5 x 5, and the output image we obtained is of the size 3 x 3. We can observe that the size of the image is reduced. To avoid the reduction of the size, we can use techniques like 0 paddings.

V. OBJECT DETECTION

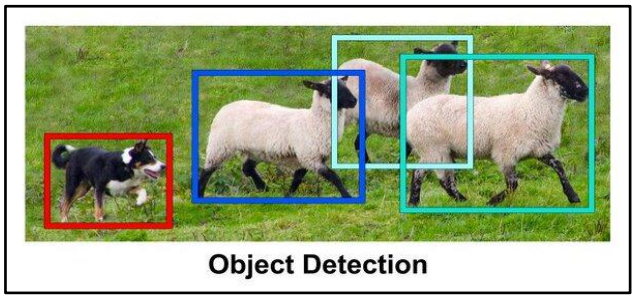

Fig. 3. Object Detection

Object Detection is one of the key factors for computer vision to understand the scene. It is still a challenge today

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 8, August 2021

to precisely determine an object from a background where similarly shaped objects are present in large numbers [1]. Image Segmentation is the initial step in many computer vision approaches and clarifies the problem by collecting the pixels logically. The process of determining the locations of the objects in the image can be recognized by object detection and image segmentation. Deep Learning approaches engage much attention in machine learning in the given modern years, and an application about image segmentation is carried out to help drive unmanned vehicles. Self Driving Cars require a deep knowledge of their surroundings. To prop up this, camera frames are used to recognize the road, pedestrians, sidewalks at pixel-level accuracy [2]. In this project, we are carrying out image segmentation using u-net architecture and Berkeley deep drive dataset [4].

VI. IMAGE SEGMENTATION

Image segmentation is a computer vision function that separates a digital image into various parts [6]. It is an important image processing technique, and it seems everywhere if we want to examine what is inside the image [1]. In an era where cameras and other devices increasingly need to see and elucidate the world around them, image segmentation has become an essential technique for teaching devices how to understand the world around them. Image segmentation generally follows pre-processing before image pattern recognition, image feature extraction, and image compacting [8]. So, we have to evaluate the reason and see what we can do to improve it.

Types of Image Segmentation : i) Instance segmentation ii) Semantic segmentation

i. INSTANCE SEGMENTATION

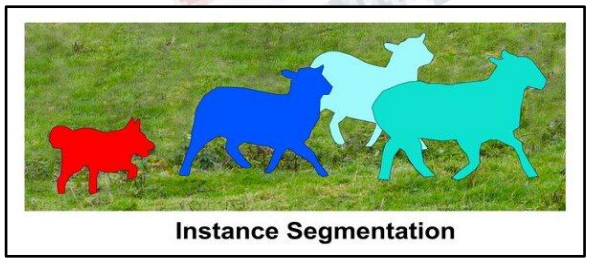

Fig. 4. Instance Segmentation

Instance segmentation is a technique [3] that identifies every pixel belonging to an instance of the object. It detects each individual object of interest in the image. All pixels corresponding to each instance (or object) of a class are given unique values.

In the figure, you can see there are two classes, dog and sheep, and every object belonging to the same class is represented using a different color i.e. all the sheep have a unique color.

ii. SEMANTIC SEGMENTATION

Semantic segmentation is the process of taking an image and labeling each pixel in that image with a certain class [11]. All pixels corresponding to a class are given the same pixel values.

In the figure, you can see there are two classes, dog and sheep, and every object belonging to the same class is represented using the same color i.e. all the sheep have the same color [5].

VII. U-NET

Apart from classification problems, a convolutional neural network(CNN) is applied to different computer vision techniques like image segmentation, object detection, image recognition, etc. U-Net architecture is expanded by making few changes in the CNN architecture. Therefore, here's a function of a neuron is taking an input of (x1 - xn) with their corresponding weights (w1 - wn), a bias (b), and the activation function f applied to the weighted sum of inputs.

$$\mathbf{f(}\ \mathbf{b} + \sum_{i=1}^{n} \quad x_i \quad w_i\mathbf{)}$$

We have used U-Net architecture because it yields better output for segmentation[18]. U-net is called so because it's U-shaped. U-Net is an architecture for semantic segmentation. It consists of a covenant path and an extensive path. It is more fortunate than conventional models, in terms of architecture and in terms of pixel-based image segmentation established from convolutional neural network layers. It's even productive with limited dataset images. The presentation of this architecture was first noticed through the analysis of biomedical images. As it's commonly known, the dimension depletion process in the height and width that we apply throughout the convolutional neural network, that is, the pooling layer is applied in the form of a dimension grown in the second half of the model.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 8, August 2021

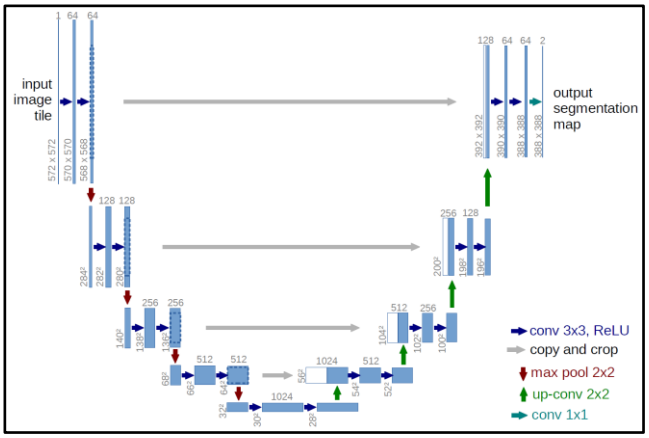

Fig. 5. UNet Architecture

The covenant path comes after the typical architecture of a convolutional network. It contains the periodic application of two 3x3 convolutions also known as unpadded convolutions, each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with pace 2 for downsampling [11]. At each downsampling step, we increase twice the number of feature methods. Each step in the extensive path consists of an upsampling of the feature map followed by a 2x2 convolution also known as up-convolution that halves the number of feature techniques, a succession with the correspondingly cropped attribute map from the decreasing path, and two 3x3 convolutions, each followed by a ReLU. The cropping is essential due to the loss of border pixels in each convolution. At the last layer, a 1x1 convolution is used to map each 64-component feature vector to the required number of classes. The overall network has 23 convolutional layers.

VIII. DATASETS

While choosing an appropriate dataset for our project, we went through a whole lot of datasets. The table below gives a comparison between all the datasets and their features like the size of the dataset, classes used, data provided, and the number of images.

Sr. No.	Dataset	Size	Classes	Data Provided	Instances
1.	Berkeley Deepdrive	9.1 GB	Bus, traffic light, traffic sign, person, bike, truck, motor, car, train, rider	Video, Image, GPS, Codes	100,000 diverse images with pixel-level annotation
2.	KITTI	12 GB	Car, van, truck, pedestrian, the person sitting, cyclist	Video, Image, LiDAR, GPS, Codes	7481 training images and 7518 test images
3.	NuScenes	31 GB	Pedestrian (Standing, wheelchair, etc.), bus, animal, cyclist	Video, Image, LiDAR, GPS, Codes	1,200,000 camera images
4.	Waymo	2 TB	Vehicles, pedestrian, cyclist	Video, Image, LiDAR, GPS, Codes	10,000 camera images
5.	Apolloscape	1.22 TB	Small vehicles, big vehicles, pedestrians, riders, motorcycle	Video	200k image frames

Table 1. Comparison of Datasets

We have chosen the BDD dataset for the following reasons:

- Berkeley DeepDrive(BDD) Dataset has numerous labeled segmentation data from various cars, in multiple cities, and different weather conditions [14].
- It has 1,00,000 images and labels of segmentation, detection, tracking, and lane lines.
- This dataset includes 2D bounding boxes interpreted on object categories(classes) like cars, traffic lights, trains, traffic signs, sky, bus, truck, bike, motor, bicycle, fence, pedestrians, train, streets, etc.
- This dataset also includes video sequences that contain GPS locations, IMU data, and timestamps.
- It consists of more than 100k driving videos (40 seconds each) recorded at various times in a day, seasons, and driving scenarios.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 8, August 2021

IX. RESULT

Below is the accuracy table of our model which is trained using UNet architecture. It gives the semantic segmented image as output. We trained our model on BDD [14] (Berkeley DeepDrive) dataset and the highest accuracy received is 88%. In this paper, we have studied different computer vision techniques used for self-driving systems. After training the model on the Berkeley Deepdrive dataset, we are focusing to test our model on a **self-generated dataset** which we create by capturing the images of Indian roads through mobile phones.

epoch	train_loss	valid_loss	acc	time
0	1.400255	0.982005	0.674887	00:13
1	1.341155	0.941101	0.738613	00:13
2	1.274747	0.786095	0.760430	00:13
3	1.167290	0.685745	0.833505	00:13
4	1.047046	0.611362	0.839518	00:13
5	0.976607	0.606744	0.845238	00:13
6	0.886535	0.478893	0.876563	00:13
7	0.827535	0.492752	0.872245	00:13
8	0.764293	0.438786	0.886581	00:13
9	0.722190	0.439871	0.885276	00:13

```
Better model found at epoch 0 with acc value: 0.6748870611190796.
Better model found at epoch 1 with acc value: 0.7386133670806885.
Better model found at epoch 2 with acc value: 0.7604303956031799.
Better model found at epoch 3 with acc value: 0.8335053324699402.
Better model found at epoch 4 with acc value: 0.8395175933837891.
Better model found at epoch 5 with acc value: 0.8452381491661072.
Better model found at epoch 6 with acc value: 0.8765628933906555.
Better model found at epoch 8 with acc value: 0.8865813612937927.
Loaded best saved model from /content/wandb/run-20210318_192616-1bon2gu3/files/bestmode
```

Fig. 6. Accuracy table

X. CONCLUSION

The evolution of self-driving cars with varying levels of autonomy to fully autonomous vehicles is yet to be made. In this proposed system, first, we went through multiple real-world datasets and shortlisted five of them. After analyzing the five datasets, we decided to use the best suitable dataset for our project, Berkeley deep drive(BDD). BDD contains around 1,27,000 images, we performed pre-processing tasks on the images and later created the prototype [14]. We are using U-Net architecture for the segmentation of images to identify different objects present on the road. This technique takes us to the insight of understanding of the object(s) in the image. Then we trained the model with maximum accuracy of 88% and tried to achieve higher accuracy. Currently, we are working on how we can Indianize our model and use it for Indian streets with good accuracy. Also, we are focusing on preparing a video format(series of images) of the segmented images.

REFERENCES

[1] "Semantic Object Segmentation in Tagged Videos via Detection" by Yu Zhang, Xiaowu Chen, Jia Li, Chen Wang, Changqun Xia, Jun Li, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, Issue:7, 20 July 2017.

[2] "Small Object Sensitive Segmentation of Urban Street Scene with Spatial Adjacency Between Object Classes" by Dazhou Guo, Liang Zhu, Yuhang Lu, Hongkai Yu, Song Wang, IEEE Transactions on Image Processing, Vol. 28, No. 6, June 2019.

[3] "Vehicle Instance Segmentation From Aerial Image And Video Using a Multitask Learning Residual Fully Convolutional Network" by Lichao Mou, Xiao Xiang Zhu, IEEE Transactions on Geoscience and Remote Sensing, Vol. 56, Issue: 11, Nov. 2018.

[4] Di Feng, Christian Haase-Schutz, Lars Rosenbaum, Heinz Hertlein, Claudius Glaser, Fabian Timm, Werner Wiesbeck, Klaus Dietmayer, "Deep multi-model object detection & semantic segmentation for autonomous driving: datasets, methods & challenges", IEEE Transaction on Intelligent Transportation Systems, Vol. 22, Issue: 3, March 2021.

[5] Wei Zhou, Julie Stephany Berrio, Stewart Worrall, Eduardo Nebot, "Automated Evaluation of Semantic Segmentation Robustness for Autonomous Driving", Vol. 21, Issue: 5, May 2020.

[6] "Semantic Instance Segmentation for Autonomous Driving" by Bert De Brabandere Davy Neven Luc Van Gool ESAT-PSI, KU Leuven, In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops.

[7] Saurabh Chaudhury, Indrani Pal, "A New Approach for Image Segmentation with MATLAB" Int'l Conf. on Computer & Communication Technology IEEE 2010

[8] Debalina Barik, Manik Mondal "Object Identification For Computer Vision using Image Segmentation" IEEE 2010 2nd International Conference on Education Technology and Computer

[9] Niall O'Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh (2019). "Deep Learning vs. Traditional Computer Vision" Vol. 943.

[10] https://www.analytixlabs.co.in/blog/what-is-image-segmentation/

[11] https://analyticsindiamag.com/my-experiment-with-unet-building-an-image-segmentation-model/

[12] https://www.allaboutcircuits.com/technical-articles/two-dimensional-convolution-in-image-processing/
[13] https://deepomatic.com/en/what-is-image-recognition
[14] https://bdd-data.berkeley.edu
[15] https://www.anaconda.com
[16] https://www.fast.ai
[17] https://deeplizard.com/
[18] https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47