Novel Vehicle Detection and Segmentation Techniques for Autonomous Vehicles

Rajeev Kumar, *Dept. of computer science & engineering, Institute of engineering & Technology Lucknow, Email:* *rajeevkumar95980@gmail.com*

Dr. M. Brindha, *Dept. of computer science & engineering, National Institute of Technology Tiruchirappalli,Email:* brindham@nitt.edu

Abstract*: This paper presents a new approach to detecting and segmenting vehicles with the help of deep learning techniques. The solution uses convolutional neural networks (CNNs) constructed with TensorFlow and U-net to reliably identify and highlight automobiles in a variety of urban settings. This model uses powerful CNN layers to extract detailed characteristics from photos anda customized segmentation module to refine vehicle boundaries. This dataset contains photos taken under various traffic conditions (such as heavy traffic, light traffic, and no traffic) and with various types of vehicles (cars, trucks, buses, and people).This study offers a novel approach employing CNNs and the U-Net architecture, implemented using TensorFlow, to effectively recognize and separate cars in different traffic circumstances. This study highlights the efficiency of utilizing TensorFlow to create robust CNN-based models for vehicle recognition and segmentation. This solution combines TensorFlow -based CNNs to improve real-timevehicle detection accuracy and contribute to the evolution of intelligent transportation systems.***Keywords*: Vehicle Detection, Vehicle Segmentation, Convolutional Neural Networks, TensorFlow, U-Net.**

I.INTRODUCTION

As cities expand and traffic increases, controlling automobiles and guaranteeing road safety becomes more difficult. Vehicle identification and segmentation are critical technologies for improving traffic management, enabling autonomous driving, andincreasing road safety. Detecting vehicles are connected with employing simple procedures that frequently struggle to survive with the complex and shifting conditions of real-world traffic. CNN's have come out as an effective and accurate solution to these issues as technology has become more advanced. CNNs are a positive creation of deep learning models capable of learning from vast quantities of data and making accurate predictions. This approach to vehicle detection and segmentation employs TensorFlow based CNNs, showing how something solves advances in real-time processing and detection accuracy evolution while improving the matrix structure of intelligent and easy transportation systems for safer and more efficient potency solutions. The proposed work presents an accurate method for vehicle recognition and segmentation based on CNNs algorithm with TensorFlow. The primary breakthrough is the combination of powerful CNNs and an overarching complete segmentation technique to reliably detect and outline the borders of cars. This approach trained the model on a wide range of photos to guarantee that it works effectively in a variety of settings, from overcrowded city streets to varying lighting conditions and wider situations.

CNNs automatically train to extract meaningful features from big datasets, which improves detection accuracy and robustness. This study presents a novel technique for vehicle recognition and segmentation based on CNNs, which is implemented using TensorFlow. This trained the model using a dataset that included a variety of traffic circumstances, vehicle kinds, and radiance scenarios. Project imposes processing requirements while retaining best accuracy by looking into efficientnetwork architectures and incorporating techniques like model trim and estimate. The approach, which strengths CNNs and the UNet architecture, as well as TensorFlow frameworks, aims to provide an accurate and efficient solution for vehicle detection and segmentation.

I. *Literature Review*

Geng K et al. obtained major advances in semantic segmentation with Fully Convolutional Networks (FCNs), and Ronneberger et al. with U-Net. FCNs provide comprehensive training for pixel-wise predictions, whereas U-Net's encoder-decoder structure with skip links allows for exact localization and context integration. Zhu et al. developed a hybrid strategy named Mask R-CNN, that enhances Faster R-CNN by adding a branch for predicting segmentation masks, attaining state-of-the-art performance in object instance segmentation, including cars. Zhu et al. proposed a model that combines YOLO and U-Net to accomplish robust vehicle recognition and segmentation.Advances in hardware, such as GPUs, and big annotated datasets, These datasets present a variety of tough circumstances, which contribute to the development of robust models. Vehicle recognition and segmentation have applications beyond autonomous driving, such as traffic monitoring, surveillance, and road safety analysis.

II. *PROPOSED ALGORITHM*

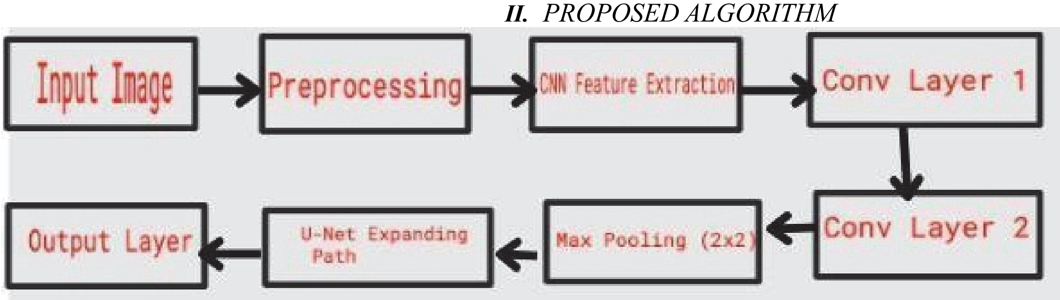

Fig 1. Flow chart for Methodology.

Figure 1 illustrates the proposed approach for vehicle detection which integrates Convolutional Neural Networks (CNNs) for gathering features with the U-Net architecture for precise segmentation. Initially, the input image is preprocessed, which includes resizing and normalization. The preprocessed image is then passed on with a CNN to extract hierarchical features, which are then used by the U-Net's expanding path to capture contextual data via down sampling. The bottleneck processes the most abstract information before being unsampled in the U-Net's expanding route, where accurate localization is accomplished by concatenating with related encoder layers. The result of this layer provides the final recognition map, recognizing vehicle locations.

A. Vehicle Detection

To find the exact location of the vehicles which required high real time performance, intelligent transport systems widely used vehicles from the long era detection technology . This Vehicle detection[1],[7] & [9] advantages the strengths of CNNs[2] andU-Net architectures, providing a powerful tool for numerous applications. CNN-based methods work as in real-time detection with accurate feature ancestry, while U-Net provides detailed segmentation capabilities. presenting Hybrid (u-Net & CNN's)

[3] models which combine both approaches to further detection performance, making them logical to the development of intelligent transportation systems and automated vehicles.

B. Vehicles Segmentation

These technologies enable precise pixel-level classification, crucial for recognizing Plate number, autonomous driving, traffic signal monitoring, and smart city infrastructure. The continuous advancement in segmentation techniques[9] accused even greater accuracy and efficiency, paving the way for most intelligent and working systems. Segmentation is the algorithm[5] which is used to define the required area for developing algorithms like CNN's [6] and U-Net. For improving accuracy and efficiency, deep learning techniques like CNN's and U-Net are used. [2]

C. CNN’s Algorithm

Convolutional Neural Network (CNN)[8] is a modified form of artificial neural networks (ANN) that is mostly used to extractattributes from grid-like matrix datasets. For example, consider visual datasets such as photos or movies in which data patterns play a vital role. CNNs [3] involve a sequence of mathematical operations which transform the input data through layers of convolutions and pooling to function-rate features and make accurate predictions.Figure 2 model shows use of a deep learning framework developed for object detection, which includes a CNN to extract key features from input photos. These features then go into a RPN, which generates possible regions in the image that could contain objects. Proposed region will be defined by a fully connected layer, which handles both segmentation and regression. The classification layer determines the item labeling, while the regression layer adjusts the bounding box coordinates for precise localization of observed objects. This integrated technique allows fast and precise object detection within images.

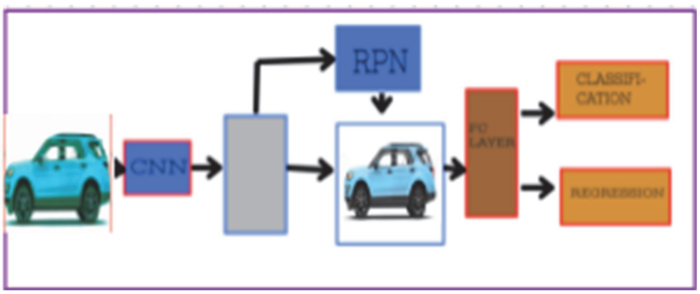

Fig. 2. CNN’s FlowChart Diagram.

The operation can be described mathematically as follows:

$$(I*K)(i,j)= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(i+m,j+n)*K(m,n) \qquad (1)$$

where, I is the input image. K is the filter (kernel) of size M×N. (I*K) (i,j) (I*K)(i, j) (I*K) (i, j) is the result of the convolution operation at position (i, j).

D. TensorFlow

TensorFlow is a complete open-source machine learning platform which is created by Google in collaboration with a huge number of enthusiastic open-source contributors which is freely concluded by machine with deep learning. TensorFlow is scalable and versatile enough to work on both data centers and mobile phones. It can run on both single and numerous machines in a distributed environment. CNNs are a popular and powerful approach to many image-based classification tasks.

ISBN : 978-81-978522-2-0

$$P(i,j) = \max_{0 \le m < M,\ 0 \le n < N} I(si+m,\ sj+n) \qquad (2)$$

where, P (i, j): Pooled feature map . I: Input feature map . si and sj: Stride for the pooling operation. M×N: Size of the poolingwindow.

E. U-Net

The powerful convolutional neural network named U-Net architecture is designed for biomedical image segmentation but is widely used in various segmentation tasks. It consists of an encoder (contracting path) and a decoder (expanding path) with skip connections that concatenate corresponding layers from the encoder and decoder. This encoder network includes 4 encoder blocks. In U-Net Each block contains two convolutional layers with a kernel size of 3*3 and valid padding, followed by a Relu activation function as defined in (3). This is fed as an input to a max pooling layer with a kernel size of 2*2. With the max pooling layer, the Model has halved the spatial dimensions learned, In this way we can reduce the computation cost of training the model [4]. Figure 3 shows the working of the U-Net model to segment an image precisely using a layer and pooling method.

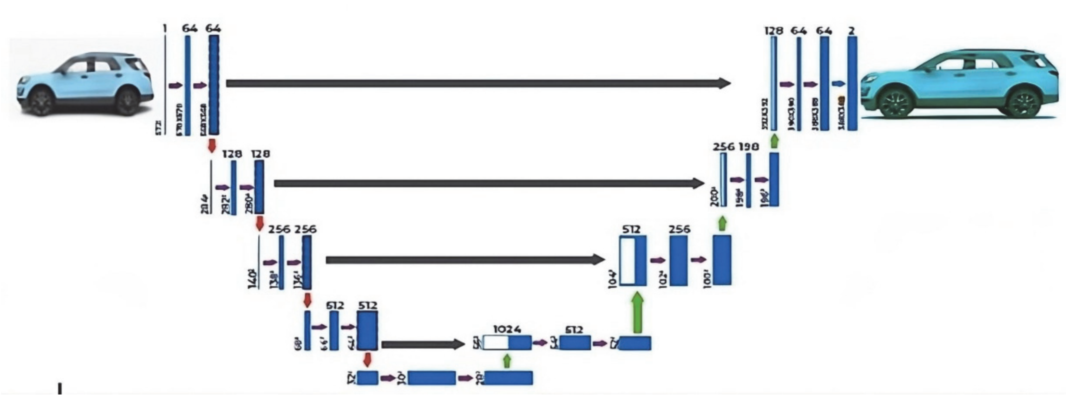

Fig. 3 U-Net Flow State

$$f(x)=\max(0,x)f(x) \qquad (3)$$

where, x shows a functional variable.

ALGORITHM 1(U-NET)

Algorithm 1(U-Net) vehicle detection system uses CNN for feature extraction and U-Net for segmentation. To extract hierarchical features, the input image is preprocessed and then passed through a succession of convolutional layers. These features are processed in the U-Net architecture's contracting path to collect context, then upsampled in the expanding path forexact localization. The final detection map pinpoints vehicle locations with great accuracy and precision.

INPUT LAYER

- Input: 3D tensor III with dimensionsH×W×C.

Contracting Path

(Encoder)

1. Initialize an empty list convs to store outputs of convolutional layers.
2. Set initial input to the original image.
3. Loop over 4 layers:

   I. Layer lll:

   a. Apply two convolutional layers:
      - Filters: num_filters×2l
      - Kernel size: 3×3 Activation: ReLU
      - Padding: Same

   b. Store output in convs list.

   c. Apply max pooling operation:

   d. Pool size: 2×2

Bottleneck

4. Apply two convolutional layers:
   - Filters: num_filters×2
   - Kernel size: 3×3
   - Activation: ReLU
   - Padding: Size

Expanding Path (Decoder)

5. Loop over 4 layers in reverse order
6. Layer lll:
   a. Apply upsampling operation to double spatial dimensions.
   b. Apply convolutional layer:
   c. Filters: num_filters×2l
   d. Kernel size: 2×2
   e. Activation: ReLU`
   f. Padding: Same
   g. Concatenate upsampled feature map with corresponding feature map from convs list.
   h. Apply two convolutional layers:
   i. Filters: num_filters×2l
   j. Kernel size: 3×3
   k. Activation: ReLU
   l. Padding: Same

Output Layer

7. Apply final convolutional layer:
   - Kernel size: 1×11
   - Activation: Sigmoid for binary segmentation, Softmax for multi-class segmentation.

III. TRAINING AND TESTING

The testing phase extracts the model's performance in their validated set. which is not useful during training of a model. The Training is a process which is used for splitting to train a validated dataset .This type of set is used to train a model allowing itto learn from data and also adjust parameters to minimize accepting error. The training and testing phase combined are crucialto developing an accurate vehicle detection model. The mathematical concepts and architecture offer understanding of the process, ensuring the model accuracy in real-time world application.

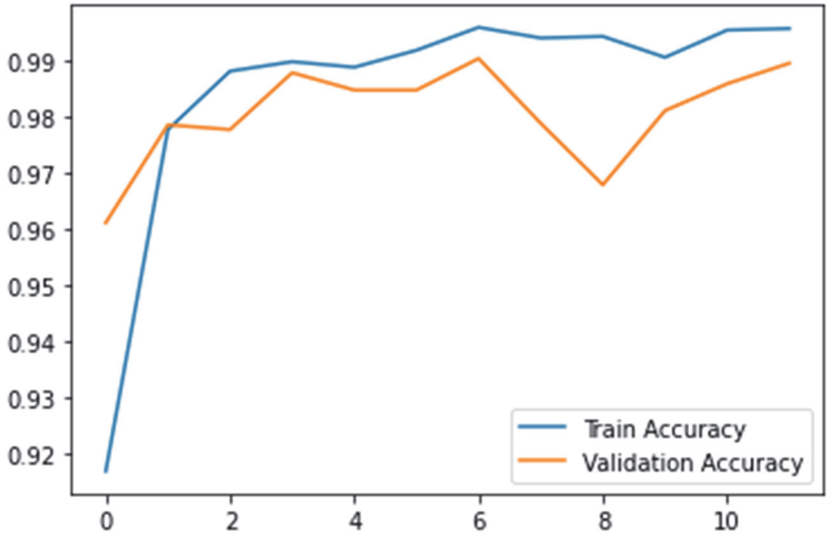

Fig. 4. Accuracy Overall Train values at each validation for datasets Vehicles-Image-Detection-Data

Figure 4 shows that with an increase in the number of iterations, Accuracy varies according to testing performance . Based on this graph since there is not much difference after 5 iterations, the number of iterations is fixed as 10 for the project.

IV. RESULT

Dataset: This dataset is available in Kaggle Datasets [10]and contains 2 Different sections, first is Non-vehicles and another represents vehicles images. along with pixel-level segmentation masks for each image. Both contain 17,762 Images. The dataset was divided into two categories: training and validation sets using a 7033/1759 split. as well as another 7174/1794.

Methodology: In this model create CNN's along with U-Net architecture for obtaining most accurate and efficient vehicles detection and segmentation. The CNN's basically perform to capture and detect the image quality the same as U-Net is combined with CNN performing the feature extraction strengths. after combining both creates a highly efficient vehicle defective tool. Tensor-flow is used for obtaining detective accuracy. The combined model demonstrates high efficiency and accuracy in detecting and segmenting vehicles in diverse conditions as well in heavy to light traffic. Figure 5 shows the final output of the model to validate to display analysis to vehicles in traffic.

Result Accuracy:

Using CNN's the model is less accurate and After combining CNN's, U-Net and Tensorflow create a most accurate and efficient model which is 99.04% accurate. Also this model precise the vehicles highly and distinguish between different vehicles.

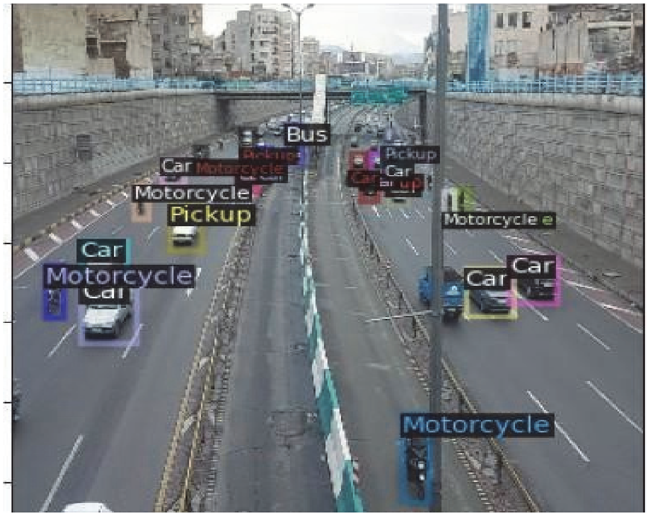

Fig. 5. CNN's and U-Net combination

IV. CONCLUSION

Implemented in TensorFlow to propose a vehicle identification and segmentation technique. High accuracy, most efficient, and continuous performance of the model in detecting and segmenting vehicles can benefit intelligent transport systems and automated vehicles. The model exhibits strong generalization ability after being trained on a wide range of datasets. Reframing the network design to lower computing expenses will be the focus of future research. Continuous efforts are underway to improve and broaden the uses of this research, which offers a useful tool for enhancing traffic management and road safety.

Future research will focus on enhancing performance and discovering new applications for intelligent transportation systems. This solution combines TensorFlow -based CNNs to improve real-time vehicle detection accuracy and contribute to the evolution of intelligent transportation systems.

V. ACKNOWLEDGEMENT

This publication is an outcome of the R&D work undertaken in the project under TiHAN faculty fellowship of NMICPS technology innovation hub on autonomous navigation foundation being implemented by the department of science & technology national mission on interdisciplinary cyber-physical systems (DST NMICPS) at IIT Hyderabad.

VI. REFERENCES

[1] Zhao X, Sun P, Xu Z, Min H (2020) "Fusion of 3D LIDAR and camera data for object detection in autonomous vehicle applications". IEEE Sensors J 20(9):4901–4913.

[2] Geng K, Dong G, Yin G (2020) "Deep dual-modal traffic objects instance segmentation method using camera and LIDAR data for autonomous driving". Remote Sens 12(20):3274

[3] Kaiming H , Georgia G, Piotr D, Girshick R (2018) "Mask R-CNN. arXiv:1703.06870v3" [cs.CV].

[4] Ronneberger et al. (2015) “Encoder-Decoder Methods for Semantic Segmentation: Efficiency and Robustness Aspects”.IEEE Sensors.

[5] Zhu et al. (2018) “Integrates detection and segmentation frameworks for robust performance Granada,” Spain, pp 126–133

[6] Krizhevsky et al. (2012) “Hierarchical feature learning from data for object detection”.IEEE Sensors.

[7] Sun J, Wan C, Cheng J, Yu FL, Liu J (2017) Retinal image quality classification using fine-tuned CNN. In: International workshop on fetal and infant image analysis (FIFI) / 4th international workshop on ophthalmic medical image analysis. Granada, Spain, pp 126–133

[8] Rezaei M, Ravanbakhsh E, Namjoo E, Haghighat M (2019) “Assessing the effect of image quality on SSD and faster R-CNN networks for face detection. In: 27th Iranian conference on electrical engineering”.Mashhad, Iran, pp 1589–1594.

[9] Bolya D, Zhou C, Xiao F, Lee YJ (2019) YOLACT: “real-time instance segmentation. In: IEEE/CVF international conference on computer vision”. Seoul, South Korea, pp 4493–4497

[10] “Kaggle Vehicles-Image-Detection-Data” (Dataset).