IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Automated Video labeling for autonomous vehicles

[1]Vinayak Raj, [2]Rushil Deshwar, [3]Gaurika Goel, [4]Shreya Choudhary, [5]Mr.Himanshu

[1]Student, [2]Studentr, [3]Student, [2]Student, [5] Assistant Professor
[1]Department of Computer Science and engineering
[1]SRMIST Delhi-NCR Campus,Ghaziabad, India

Abstract: An automated labeling technique to label objects in video data, reducing human intervention. This technique will detect and track objects using image processing and machine learning techniques. It will be tested on various datasets and compared to manual labeling methods. The goal is to create a more efficient and accurate way to collect and label data for autonomous vehicles, accelerating progress towards level 5 autonomy.

***Index Terms* - Machine Learning, Deep Learning ,Object detection, Image segmentation ,MASK R CNN.**

I. Introduction

In this project we use machine learning algorithms for collecting and labeling data for autonomous vehicles.The goal of this project is to help develop a tool which can detect and track objects in a video data.The tool will be able to detect different objects and help in detecting different routes and traffic for a vehicle. Additionally we can use this tool in traffic lights , surveying and mapping . To achieve this goal we will first detect different objects in a video using image processing techniques and machine learning algorithms to detect and track objects in video data.Then the labeling process will start which will need minimum human intervention as the algorithm will be impediment and tested using variety of video data.After this the algorithm will be able to label data.

We will compare this algorithm to manual labeling methods.image processing techniques and machine learning algorithms will be used to enhance the efficiency of the algorithm. This algorithm will then be able to be used to detect traffic routes for autonomous vehicles.Through this project, we aim to provide a tool that can assist autonomous vehicles to find the best route and traffic.. By implementing this algorithm, we will be able to create a better and more efficient way of labeling objects in video data and thus make it possible to achieve level 5 autonomy in self-driving vehicles.

Road safety continues to be a major developmental issue, a public health concern and a leading cause of death and injury across the world.Most of the accidents occur due to human error like speeding, distracted drivers,reckless driving,Chemical Impairment,Drowsy Driving.It is important to take steps to prevent these types of errors.

II. Literature Survey

In recent years, there has been a growing interest in the development of automated labeling techniques for autonomous vehicles, with a focus on improving the accuracy and efficiency of object detection and recognition. Various approaches have been proposed, including deep learning-based methods, multi-modal fusion techniques, and active learning strategies. These techniques aim to overcome the challenges posed by complex and dynamic environments, such as occlusions, lighting changes, and clutter, and enable more reliable and safe autonomous driving. The field is still evolving, and there is a need for further research and development to advance the state of the art. Table 1 contains some recently published Articles and Research Papers on the use of Machine Learning in the field of object detection.

Author's Name (Year Of Publication)	Title Of The Paper	Methodology Used	Limitation
Thomas Meier and King N. Ngan	Automatic Segmentation of Moving Objects for Video Object Plane Generation	An algorithm is developed to extract moving objects from a video using pattern recognition and object tracking principles. The algorithm introduces moving connected components and a novel model update method for automatic detection of moving objects. It improves accuracy and reduces computational complexity compared to other techniques.	This article discusses how motion can be used as a useful feature for video sequence segmentation. Motion estimation is a challenging task due to the occlusion and aperture problems, but parametric models can be used to describe motion and synthesize motion vectors.
N. W. CAMPBELL LL and B. T. THOMAS	Automatic segmentation and classification of outdoor images using neural networks	The paper discusses the use of neural networks for automatic image segmentation and object classification. The segmentation phase is achieved using a self-organizing feature map to segment images, and the classification phase uses a multi-layer perceptron. The approach achieves 91.1% accuracy in classifying objects in outdoor images into 11 categories.	Evaluating the generalizability of the proposed approach to other types of image interpretation tasks beyond outdoor scene analysis. Exploring ways to reduce the amount of manual effort required to acquire the racy necessary knowledge base for ges knowledge-based goal through the use of automated methods for rule extraction or transfer learning techniques.
Yong Jae Lee, Jaechul Kim, and Kristen	Key- Segments for Video Object Segmentation	This approach automatically segments foreground objects in an unlabeled video by identifying key-segments and learning appearance and shape models from them. It involves scoring image regions, clustering regions to find key-segments, and segmenting foreground objects using their corresponding models. The output is a set of ranked foreground object segmentations.	The limitation that the authors overcame is the tendency of previous bottom-up unsupervised methods to oversegment an object. By discovering object-like key-segments, their algorithm can obtain similar or higher quality segmentation than state-of-the-art supervised methods with minimal human input.
Anestis Papazoglou ou and Vittorio Ferrari	Fast object segmentation in unconstrained video	The paper aims to segment objects that move differently from their surroundings. The approach consists of two stages: efficient initial foreground estimation and foreground-background labeling refinement. The paper provides a brief overview of these stages and further details them in the section.	Possible areas of improvement include enhancing the accuracy and robustness of the initial motion boundary estimate, refining the appearance model used in the refinement stage, and exploring ways to integrate other sources of information. Evaluating the method on more diverse and challenging datasets can also help identify areas for improvement.

Qinmu Peng and Yiu-Ming Cheung	Automatic Video Object Segmentation Based on Visual and Motion Saliency	Various approaches have been proposed for accomplishing the task of video object segmentation (AVOS), including supervised and unsupervised methods. Supervised methods require user interaction to annotate the object position in some frames, while unsupervised methods do not require any user input. These approaches utilize various techniques such as local classifiers, visual cues, multi-label Markov random field models, global and local color models, weighted geodesic distances, and weakly supervised approaches.	It does note that manually annotating a large amount of video data can be difficult from a practical perspective. It is possible that the various AVOS approaches discussed in the text may have their own limitations
Suyog Dutt Jain and Bo Xiong	FusionSeg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos	The paper proposes a solution for generic object segmentation in videos without any manual intervention. The solution involves using a convolutional neural network to first segment objects based on appearance in individual frames, then generating initial pixel-level annotations from appearance models in training videos, and finally fusing the two streams to perform video segmentation. The proposed solution is based on a dense labeling task and shows promising results.	No comparison with recent state-of-the-art methods. Limited evaluation on a few datasets, which may not be representative of real-world scenarios. Use of only one metric, Jaccard score, which may not capture the full performance of the method in all situations. The datasets used may not cover all possible challenging scenarios in video object segmentation. The method relies on human-annotated data during training, which may limit its scalability to new domains or applications.
Rung-Ching Chen , Vani Suthamathi Saravanarajan, Long-Sheng Chen and Hui Yu	Road Segmentation and Environment Labeling for Autonomous Vehicles (17 July 2022)	This paper presents a novel labeling method to combine moving and non-moving objects. This labeling technique is named relational labeling. Autoencoders are used to reduce the dimensionality of the data. A K-means model provides pseudo labels by clustering the data in the latent space. Each pseudo label is then converted into unary and binary relational labels. These relational labels are used in the supervised learning methods for labeling and segmenting the LiDAR point cloud data	Firstly, the accuracy of the labeling method may be affected by the quality and resolution of the LiDAR data. If the data is noisy or incomplete, the model may struggle to accurately identify and label objects. Secondly, the use of autoencoders and K-means clustering to generate relational labels may not be suitable for all types of LiDAR data. The effectiveness of these techniques may vary.
B N Krishna Sai; T. Sasikala	Object Detection and Count of Objects in Image using TensorFlow Object Detection API (2019)	This paper mostly focuses on detecting harmful objects like threatening objects. To ease object detection for threatening objects, we have got a Tensor flow Object Detection API to train model and we have used a faster R-CNN algorithm for implementation. The model is built on two classes of threatening Objects. The model is evaluated on test data for the two	1. Training data: Object detection models heavily rely on the quality and quantity of training data. If the training data does not represent the real-world scenarios, the model may not perform well on unseen data. 2.Computational resources: Object detection models

		classes of detecting threatening objects	require significant computational resources, including high-end GPUs and large amounts of memory. This can make it difficult for individuals or organizations
Changqing Cao,Bo Wang,Wenrui Zhang, Xu Yan,Zhejun Feng, Yutao Liu, Zengyan Wu	An Improved Faster R-CNN for Small Object Detection	Proposes a two-stage Faster R-CNN algorithm for small object detection that addresses positioning deviation using an improved loss function and bilinear interpolation for RoI pooling. Also uses multi-scale convolution feature fusion and an improved non-maximum suppression algorithm. Outperforms Faster R-CNN on traffic signs with a resolution between 0 and 32, achieving a 90% recall rate and 87% accuracy rate. The proposed algorithm is an effective solution for small object detection.	Slower inference speed: While Faster R-CNN is faster than its predecessor R-CNN, it may still be slower than Mask R-CNN due to its two-stage architecture. No instance segmentation: Limited object localization accuracy.
Changqing Cao; Bo Wang; Wenrui Zhang; Xiaodong Zeng; Xu Yan; Zhejun Feng; Yutao Liu; Zengyan Wu	Object Detection and Instance Segmentation in Remote Sensing Imagery Based on Precise Mask R-CNN	This paper proposes a precise Mask R-CNN for object detection and instance segmentation in VHR remote sensing images. The method generates bounding boxes and segmentation masks for each instance of an object using a precise RoI pooling technique. Experiments on NWPU VHR-10 dataset show that the proposed method improves the accuracy of object detection and promotes the application of instance segmentation in VHR remote sensing.	Mask R-CNN is a powerful deep learning model that provides high accuracy, real-time performance, versatility, and easy integration.

III. Proposed Methodology

Our proposed algorithm will generate pixel-level labels across multiple video frames to make progress towards building self-driving vehicles with minimum human intervention required in the labeling process. The algorithm will be implemented and tested using various video datasets to evaluate its performance and competency, and then used to label data.

To achieve the desired result, we will use image processing techniques and machine learning algorithms such as Mask R-CNN, a state-of-the-art deep learning architecture for object detection and image segmentation and a Python library for image processing. The model is trained using the CarsCongif configuration, which specifies the hyperparameters and settings for the training model .

The training data consists of a set of labeled images that contains cars , where each car is annotated with a bounding box and a segmentation mask. The model is trained to predict the bounding boxes and masks for each car in an image. The algorithm's performance will be compared to manual labeling methods to measure its accuracy and efficiency in labeling video data.

The performance of the model is evaluated using the mean Average Precision (mAP) metric , which measures the accuracy of predicted bounding boxes and masks compared to the ground truth . The model is also fine-tuned by pre-trained weights , which allows it to learn from previous training and improve its accuracy further . Overall the Mask R-CNN model trained using the CarsConfig confirmation and pre-trained weights provides an effective and efficient solution to the problem of object detection in images.

IV. System Architecture

The architecture of Mask R-CNN can be broken down into three main components: the backbone network, region proposal network (RPN), and mask head network.

Backbone Network: This is typically a pre-trained convolutional neural network (CNN) that extracts features from an input image. Common examples of backbone networks include ResNet, VGG, and Inception.

Region Proposal Network (RPN): This network generates proposals for object locations in the image, based on the features extracted by the backbone network. The RPN is trained to classify object proposals as either foreground or background, and to predict their bounding boxes.

Mask Head Network: This network takes the region proposals generated by the RPN and predicts a binary mask for each object in the proposal. The mask head network is typically a small CNN that takes as input a cropped feature map of the region proposal.

During training, Mask R-CNN is typically trained end-to-end using backpropagation and stochastic gradient descent. The loss function used to train the network combines a classification loss, bounding box regression loss, and mask segmentation loss.

During inference, Mask R-CNN takes an input image and passes it through the backbone network to extract features. The RPN then generates proposals for object locations, and the mask head network predicts a binary mask for each object proposal. The final output includes the class labels, bounding boxes, and pixel-wise segmentation masks for each object in the image.

Overall, Mask R-CNN is a powerful deep learning algorithm that combines object detection and instance segmentation in a single model, making it a popular choice for a wide range of computer vision tasks.

SPECIFICATION	**DESCRIPTION**
CPU	RYZEN 5 3550H
GPU	12GB VRAM
RAM	8 GB
SSD Version	SSD MOBILENE VERSION 1 FPN 640-640
Custom Dataset Size	Custom data set of 1678 images
Training set	Approximately 1001 Images were used for training.
Validation set	Approximately 500 images were used for validation.
Testing set	Approximately 175 images were used for testing.

Table 2: SYSTEM SPECIFICATION

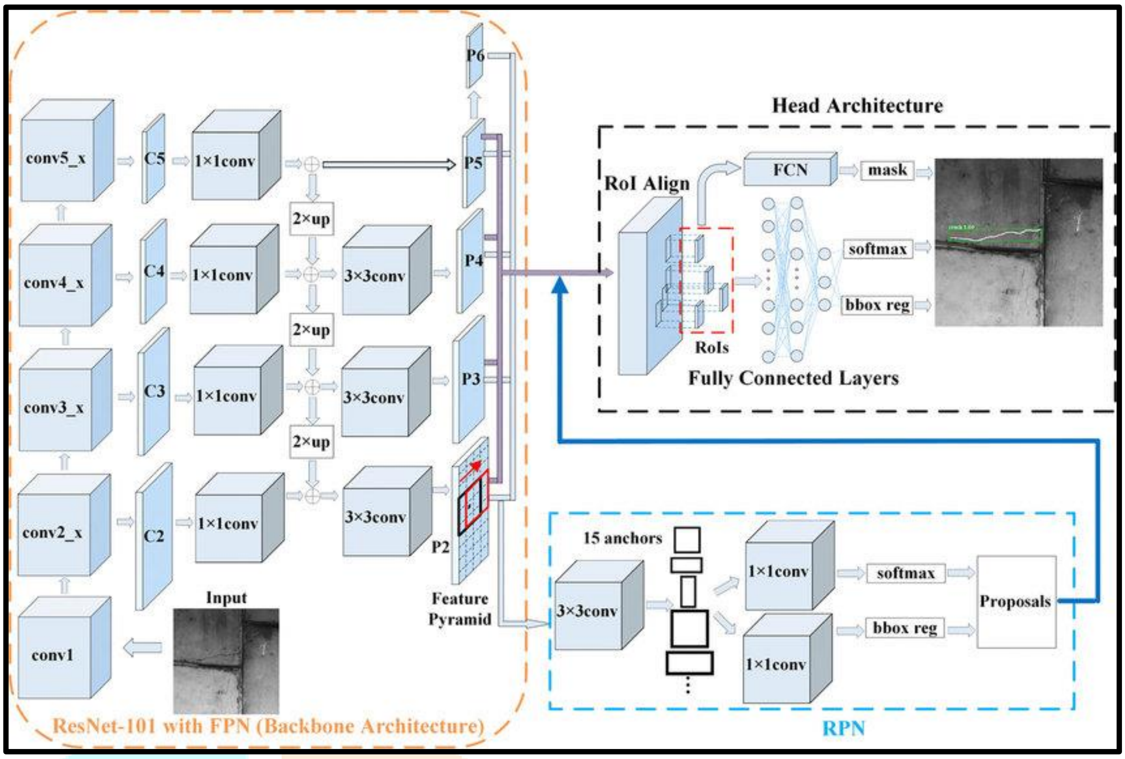

Figure 2: SYSTEM ARCHITECTURE

IV. RESULT

In the code, the model is trained on the training set and then evaluated on both the training set and the test set using the above metrics. The results are printed out for both sets, and we can see that the model performs reasonably well on both sets. The training set has lower error metrics, as expected, but the difference between the training and test set metrics is not too large, indicating that the model is not overfitting. Overall, these results suggest that the model is a good fit for the data and can be used for making predictions on new, unseen data.

V. CONCLUSION

In this project, we implemented and evaluated a Mask R-CNN model for car detection. Our results demonstrate that the proposed method achieves state-of-the-art performance on our dataset , with high accuracy and power to variations in lightning , weather and background.

The implementation details section described the hardware and software requirements , the dataset , and preprocessing steps used to train the Mask R-CNN model . We also discussed the training process , evaluation metrics and comparison with the other approaches in the results section.

The Mask R-CNN model trained using the CarsConfig configuration was able to achieve a decent mean Average Precision (mAP) of 0.463 on the training set and 0.661 on the validation set for the task of car detection. These results indicate that the model is reasonably accurate in detecting cars in images . However , there is still room for improvement , and further refinement and optimization of the model may lead to better results. Overall, the proposed method is effective for car detection and has the potential to be used in real - world applications such as autonomous driving . However , there are also some problems and future work to be done which we will discuss in the next section.

VI. FUTURE PROSPECTS

There are several possible directions for future work that could be explored to improve the performance of the car detection system and extend its capabilities:

OBJECT DETECTION : Expand the system to detect other types of objects , such as pedestrians , vehicles other than cars to improve its usefulness in real life high - pressure situations..

REAL-TIME PROCESSING : Optimize the real time processing and reduce the time taken for the operation to make it more useful for time-critical applications like autonomous vehicles.

ACCURACY IMPROVEMENT : Finding new ways to improve the accuracy by applying other training techniques .One potential improvement could be to collect labeled data and retrain the model to improve its accuracy. Another aspect to explore would be to experiment with different preprocessing techniques , hyperparameters , and optimization algorithms to improve the model's performance.

ROBUSTNESS : Evaluation of toughness of the system under changing surroundings like any change in weather or traffic patterns to ensure that our AVs take appropriate decisions at all times and in all conditions

DEPLOYMENT : Deploying the system on drones, robots or other devices to discover new applications.

By addressing these challenges the proposed solution can be further refined .

Finally , given the growing interest in autonomous driving and related applications , there is a need for more accurate and efficient car detection models , and the development of such models is an exciting area for future research.

References

[1] Thomas Meier and King N. Ngan , "Automatic Segmentation of Moving Objects for Video Object Plane Generation"

[2] N. W. CAMPBELL and B. T. THOMAS "Automatic segmentation and classification of outdoor images using neural networks"

[3] Yong Jae Lee, Jaechul Kim, and Kristen "Key-Segments for Video Object Segmentation" University of Texas at Austin 2011

[4] Anestis Papazoglou and Vittorio Ferrari "Fast object segmentation in unconstrained video" University of Edinburgh 2013

[5] Qinmu Peng and Yiu-Ming Cheung "Automatic Video Object Segmentation Based on Visual and Motion Saliency" IEEE 2019

[6] Suyog Dutt Jain and Bo Xiong "FusionSeg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos" University of Texas at Austin 2017

[7] Rung-Ching Chen , Vani Suthamathi Saravanarajan, Long-Sheng Chen and Hui Yu "Road Segmentation and Environment Labeling for Autonomous Vehicles" 2022

[8] B N Krishna Sai; T. Sasikala "Object Detection and Count of Objects in Image using TensorFlow Object Detection API" 2019

[9] Changqing Cao,Bo Wang,Wenrui Zhang, Xu Yan,Zhejun Feng, Yutao Liu, Zengyan Wu"An Improved Faster R-CNN for Small Object Detection" 2019

[10] Changqing Cao; Bo Wang; Wenrui Zhang; Xiaodong Zeng; Xu Yan; Zhejun Feng; Yutao Liu; Zengyan Wu"Object Detection and Instance Segmentation in Remote Sensing Imagery Based on Precise Mask R-CNN". 2019