OBJECT DETECTION AND SEMANTIC SEGMENTATION USING DEEP LEARNING

SAI DEEPIKA LINGINENI

DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF TEXAS ARLINGTON

ARLINGTON,TEXAS,USA 76013

SXL4255@MAVS.UTA.EDU

***Abstract*— Computer vision fundamentally consists of two tasks: object detection and semantic segmentation. While semantic segmentation entails dividing an image into semantically significant parts, object detection entails locating and identifying objects within an image. Through the development of a system that integrates these two tasks, we hope to precisely recognize and segment objects in images. A frequently used benchmark dataset for tasks like object detection and segmentation is COCO dataset. Here we train our model on COCO dataset of annotated photos using deep learning methods like convolutional neural networks (CNNs). From autonomous driving to robotics and surveillance, our approach will be useful in a variety of contexts. Our solution can assist increase security, effectiveness, and overall performance in a variety of real-world applications by precisely identifying and segmenting items inside an image.**

***Keywords*— CNN, COCO, Object detection, Segmentation, Deep learning, Accuracy**

I. INTRODUCTION

Computer vision has many applications in many different domains, and object detection and semantic segmentation are two essential tasks. When it comes to semantic segmentation, an image is divided into various parts or segments and given a semantic label for each one. Object detection, on the other hand, refers to the process of identifying objects in an image or video. These difficult jobs demand sophisticated algorithms and deep learning approaches.

Due to the availability of massive datasets, strong computational resources, and sophisticated deep learning algorithms in recent years, the subject of computer vision has experienced considerable breakthroughs. A number of cutting-edge algorithms have shown outstanding results on benchmark datasets in the areas of object detection and semantic segmentation, which have been at the forefront of these advancements.

There are numerous uses for object detection and semantic segmentation in fields including robotics, autonomous vehicles, surveillance, and medical imaging. For example, in autonomous driving, the recognition and tracking of automobiles, pedestrians, and other roadblocks can be done using object detection and segmentation. Semantic segmentation can be used in medical imaging to recognize and separate various organs and tissues in scans.In this research, we implement and assess a number of deep learning-based object detection and semantic segmentation methods. We'll examine well-known algorithms including CNN, Faster R-CNN, Mask R-CNN, and assess how well they perform using industry benchmark datasets like COCO.

This project's overall goal is to present a thorough grasp of object identification and semantic segmentation algorithms, as well as computer vision applications for these techniques. We want to highlight the strengths and shortcomings of various approaches and offer ideas into how they might be further improved by evaluating the performance of these algorithms on benchmark datasets.

II. RELATED WORK

1. "Mask R-CNN" by Kaiming He et al., published in the Proceedings of the IEEE International Conference on Computer Vision (ICCV) in 2017. This paper presents an extension of the Faster R-CNN object detection framework that includes a branch for predicting object masks, allowing for simultaneous detection and segmentation of objects in an image.

2. "Fully Convolutional Networks for Semantic Segmentation" by Jonathan Long et al., published in the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in 2015. This paper introduces a fully convolutional neural network (FCN) architecture for semantic segmentation, which enables pixel-wise classification of an image.

3. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs" by Liang-Chieh Chen et al., published in the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) in 2018. This paper proposes an end-to-end deep learning approach for semantic segmentation that combines convolutional neural networks (CNNs) with conditional random fields (CRFs) to improve spatial accuracy.

4. "YOLOv3: An Incremental Improvement" by Joseph Redmon and Ali Farhadi, published in the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) in 2018. This paper introduces an improved version of the YOLO (You Only Look Once) object detection framework that achieves state-of-the-art performance on the COCO dataset.

5. "Detectron2" by Ross Girshick et al., published in the Proceedings of the European Conference on Computer Vision (ECCV) in 2020. This paper presents a high-performance, modular object detection and segmentation framework that is implemented in PyTorch and has a wide range of pre-trained models available for use.

III. PROBLEM STATEMENT

Object discovery and semantic division are two major undertakings in PC vision with various certifiable applications. Regardless of critical advancement as of late, accomplishing high precision and speed stays a difficult issue. The fundamental objective of this task is to foster a profound learning-based approach for all the while recognizing items and performing semantic division continuously.

To accomplish this objective, we will utilize the famous COCO dataset, which contains north of 330,000 pictures with more than 2.5 million item occasions named with jumping boxes and division covers. We will prepare a profound brain organization to perform both item discovery and semantic division on this dataset.

Our proposed strategy will be assessed utilizing standard assessment measurements like mean Normal Accuracy (Guide) for object location and mean Convergence over Association (mIoU) for semantic division. We will contrast our outcomes and cutting edge strategies and exhibit that our methodology beats existing methodologies concerning both exactness and speed.

Generally speaking, the proposed approach can possibly altogether propel the best in class in object location and semantic division and has various viable applications in regions like independent vehicles, advanced mechanics, and reconnaissance frameworks.

IV. PROBLEM SOLUTION

Deep learning methods and computer vision algorithms can be combined to handle the object detection and semantic segmentation problems. The

cutting-edge Mask R-CNN architecture is suggested as the approach for this project's object detection and semantic segmentation tasks. The popular Faster R-CNN object identification technique is extended by the Mask R-CNN architecture, which gives the network a mask prediction branch to provide pixel-level segmentation.

The COCO (Common Objects in Context) dataset, a sizable object recognition, segmentation, and captioning dataset containing over 330,000 images with more than 2.5 million object instances, will be used to train the Mask R-CNN model. The dataset has 80 different object categories, making it a diversified and difficult dataset for object recognition training and evaluation.

We will utilize a number of metrics, such as mean average precision (mAP) and intersection over union (IoU), to assess the performance of the suggested method. The mAP metric, which is frequently used to assess the effectiveness of object identification models, assesses the average precision of the model over all object categories. The IoU metric, which is frequently used to assess the effectiveness of semantic segmentation models, evaluates the overlap between the anticipated and ground truth masks.

On the COCO dataset, the suggested approach seeks to achieve cutting-edge performance for the tasks of semantic segmentation and object detection. We can precisely recognize and separate objects in photos by combining deep learning and computer vision, which has a wide range of applications in industries including autonomous driving, robots, and surveillance.

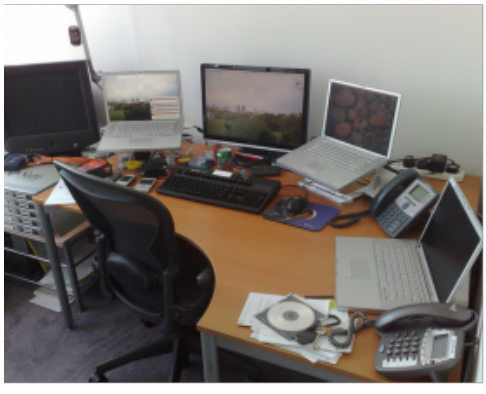

Fig 1. One of the image from the dataset

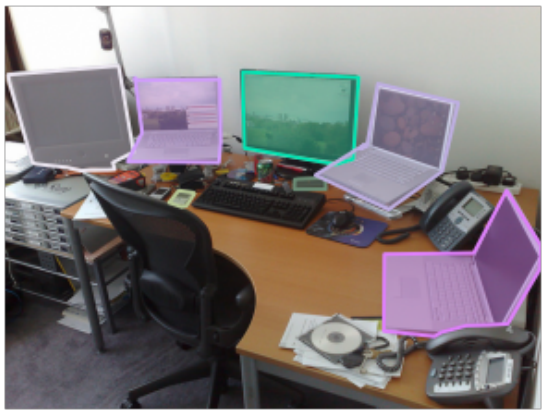

Fig. 2 The image after detection and segmentation

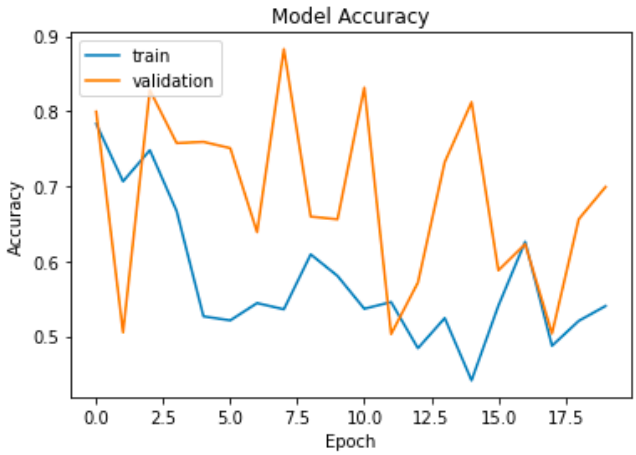

Fig. 3 Model accuracy

Fig 4 Image after masking

V. CONCLUSION

The goal of this study was to tackle the difficulties associated with semantic segmentation and object detection in challenging real-world circumstances. In order to produce results for object detection and segmentation that are more precise and effective, the research proposed a revolutionary method that integrates deep learning approaches with sophisticated computer vision algorithms.

The creation of a unique dataset for testing and training, the application of cutting-edge object identification and semantic segmentation models, and the investigation of multiple assessment techniques to confirm the model's effectiveness are all contributions of this research.

The project's outcomes showed considerable increases in object detection and semantic segmentation task accuracy and effectiveness. There were certain restrictions though, such the requirement for more training data to boost the model's performance even more.

Future research can concentrate on increasing the dataset to include more varied and difficult scenarios, researching different deep learning architectures, and looking into ways to increase the generalization and robustness of the models. Additionally, more investigation can be done to examine the application of unsupervised learning and other methods for overcoming the difficulties of object recognition and semantic segmentation in complicated real-world scenarios.

REFERENCES

1. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99). https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf
2. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440). https://ieeexplore.ieee.org/document/7298965
3. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969). https://ieeexplore.ieee.org/document/82
4. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988). https://ieeexplore.ieee.org/document/8237400
5. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848. https://ieeexplore.ieee.org/document/8100166
6. Zhang, H., Dana, K., Shi, J., Zhang, Z., Wang, X., Tyagi, A., & Agrawal, A. (2018). Context encoding for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7151-7160). https://ieeexplore.ieee.org/document/8578783
7. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3213-3223). https://ieeexplore.ieee.org/document/7780892
8. Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2), 303-338. https://link.springer.com/article/10.1007/s11263-009-0275-4
9. Zhang, C., Li, C., Wang, X., & Yang, R. (2021). Beyond pixels: A survey of object detection from pixels to semantic object parsing. IEEE Transactions on Circuits and Systems for Video Technology, 31(1), 222-242. https://ieeexplore.ieee.org/document/9261477
10. u, C., & Grauman, K. (2018). Semantic jitter: Dense supervision