Adaptive Object Detection and Classification Using EMOD for Real-Time Applications

[1]N. Ravikumar.; [2]Dr.T.Kamaleshwar
[1]Reseatch Scholar; [2]Associate Professor
[1]MCA, Mphil, M.BA; [2]M.Tech, Ph.D(CSE)
[1,2]Dept. of Computer Science & Engineering
Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Chennai, Tamil Nadu, India

Abstract: In many computer vision applications, including surveillance, medical imaging, and autonomous driving, object identification and segmentation are essential. In this research, we use the current YOLOv7 architecture to offer an improved method for real-time object recognition and segmentation. YOLOv7 is renowned for its cutting-edge speed and accuracy in real-time object detection, outperforming earlier iterations of YOLO in terms of accuracy and performance. It has been difficult to incorporate segmentation features while preserving processing speed, nevertheless. Our method maintains the high classification accuracy of YOLOv7 while adding a segmentation head to create pixel-wise masks for items that are recognized. Using well-known datasets like COCO and PASCAL VOC, we test our approach's performance in terms of segmented effectiveness (Intersection over Union, or IoU) and accuracy in detection (mean average precision, or mAP). Experimental findings show that our method preserves YOLOv7's real-time processing performance while achieving a notable increase in segmentation accuracy. This study helps close the gapseparating object detection and segmentation by providing a workable solution for situations that need precise segmentation in addition to high- speed detection.

Keywords: machine vision, real-time identification, deeplearning, convolutional network neural networks (CNNs), recognizing objects, semantic segmentation, YOLOv7, and image segmentation.

1. Introduction

Basic computer vision tasks like segmentation and detection of objects are essential to many applications, including robots, autonomous driving, imaging for medicine, and surveillance. In order to comprehend and interact with visual environments, these tasks require the ability to recognize items inside images and precisely label or mask them. Conventional object detection techniques mostly depended on region-basedstrategies such as R-CNN and its variations. Despite their effectiveness, these techniques frequently have trouble performing in real time because of their computing complexity. Deep learning models, especially YOLO, or You Only Look Once models, have drawn a lot of attentionlately because of their capacity to combine rapid computation with high detection precision, which makes them perfect for real-time applications.

To improve its functionality, YOLO has undergone multiple versions, the most recentbeingYOLOv7.YOLOv7isnotable for its advancements in detectionquickness and precision, attaining cutting- edge outcomes on benchmark datasets including PASCAL VOC and COCO. A more effective backbone network, improved detecting head design, and innovative methods like dynamic scaling are all used in the YOLOv7 architecture to provide optimal performance. Despite its superiority in object detection tasks, YOLOv7 is still primarily focused on identification rather than the process of segmentation, which restricts its use in applications that need precise pixel-level predictions, including autonomous car scene interpretation or medical image analysis.

Our work fills this gap by adding an integrated segmented module to YOLOv7, which allows categories of objects and their matching pixel-wise masks to be predictedsimultaneously. Thishybrid model adds semantic segmentation capabilities while maintaining YOLOv7's speed and accuracy. Our goal is to provide a workable solution to feed applications that require both object positioning and segmentation by utilizing YOLOv7's real- time detection capabilities in conjunction with cutting-edge segmentation approaches. Our approach is tested on popular datasets, and the outcomes show that our improved YOLOv7 model can provide accurate segmentation and high- speed detection.

2. Related Work

et. all. Manakitsa, N., Maraslidis, G. S., Moysis, L., &Fragulis, G. F. The multidisciplinary discipline of machine vision, which seeks to simulate the way people see in computers, has made substantial contributions and advanced quickly. This essay explores the history of machine vision, starting with the earliest image processing algorithms and ending with how it merged with robotics, computer science, and mathematics to become a separate area of artificial intelligence. Its expansion and use in commonplacegadgets havebeen fueledby the incorporation of machine learning methods, especially deep learning. Replicating human visual abilities, such as recognition, understanding and interpretation, is the maingoalof thisworkon computer visionsystems.Notably, important tasks requiring strong mathematical underpinnings include image segmentation, object detection, and classification. A bold attempt to mimic human visual perception is reflected in the development of machine vision. A bold attempt to mimic human visual perception is reflected in the development of machine vision. Significant progress has been made in simulating human actions and perceptions thanks to cross-disciplinary collaboration and the incorporation of deep learning techniques. The foreseeable future of computing devices and intelligence applications is being shaped by machine vision research.

et. all. Rana, S., Gerbino, S., Barretta, D., Carillo, P., Crimaldi, M., Cirillo, V., ... & Sarghini, F. The purpose of this data postis to offer materials that may help with research on computer vision-based precision farming weed identification and segmentation. Multispectral (MS) photos of Triticum aestivum crop fields with a heterogeneous mix of Raphanus raphanistrum with both standard and random crop spacing have been curated by us. The purpose of this collection is to make weed identification andsegmentation easier using both automatically and manually annotated Raphanus raphanistrum, or wild radish. The dataset, which is accessible to the general public via the Zenodo data library, offers pixel-level annotations that are essentialfor registration andsegmentation. The dataset includes 85 original MSphotos that were taken in 17 different situations and cover a range of spectra, including RedEdge, Blue, Green, and NIR (near-infrared). Each 1280 × 960 pixel image is the foundation for the particular weed segmentation and detection. The Common Object in Context (COCO) categorizationformat was usedto store the results of manual annotations made with the Visual Geometries Group Image Annotator (VIA). An autonomous Visual Object Classes Extended mark-uplanguage (PASCAL VOC) annotation for 80 MS photos was obtained by training a Getting Grounded DINO + Segment Anything Model (SAM) using this manually annotated data. This process facilitated the resource-intensiveannotation operation.

et.all. Huang, L., Kurz, C., Freislederer, P., Manapov, F., Corradini, S., Niyazi, M., ... & Riboldi, M. A Retina U-Net baseline model was initially trained using dynamically reconstructed radiographs (DRRs) produced from a publiclyavailable dataset comprising 875 CT scans andtheassociatedlungnoduleannotations.

A patient-specific refinement process was the ndevelopedusingaseparatecohort of 97 lung patients. To find the best hyperparameters for automatic particularto the patient training, we validated 13 patients whose ground truth was greater than 0.7 and whose foundation model predicted the boundaries of the box on planning CT (PCT)-DRR withintersections over union (IoU). With different PCT-DRRIoUs, the remainder of 84 patients were included in the final test set. In order to simulate the intrafraction motion throughout treatment, a patient-specific model was created by refining the baseline model on the PCT-DRR and evaluating it on a different 10-phase 4DCT-DRR. The benchmark model wasan algorithm for matching templates. Four metrics were used to assess the testing results: the Dice consistency coefficient (DSC)andthe center ofmass(COM) error for segmentation masks, and the DSC and the center of box (COB) inaccuracy forboxboundaries detections.Thebenchmark model was used to compare performance, and statistical testing was done to determine significance.

et.all. Shoaib, M., & Sayed, N. The performance of computer vision problems has significantly improved with the developments in deep learning. However, the reisasignificant risk of inaccurate item predictions or image segmentation in many real-world applications, such as autonomous driving cars. YOLO models and other conventional deep neuralnetwork models for segmentation and identification of objects frequently make too optimistic forecasts and fail to account for prediction uncertainty on out-of- distribution data. In this paper, we present Monte-Carlo Drop Block (MC-Drop Block), an efficient and effective method for modeling uncertainty in convolutional vision techniques for object recognitionand YOLO. The suggested method uses drop-block on the convolutional layer of deep learning models like YOLO and convolution transformers during training and testing. We demonstrate theoretically that this results in a probabilistic convolutional neural network that can capture the model's epistemic uncertainty. We also use a Gaussian likelihood toreflect the mathematical uncertainty in the data.

et.all.Zhu, C., & Chen, L.In the deep learning era, object detection and segmentation—two of the most basicsceneunderstandingtasks—haveadvanced significantly. The annotated categories in existing datasets are frequently small-scale and pre-defined due to the high expense of manual labeling; in other words, even the most advancedfullysuperviseddetectors and segmentors are unable to generalize beyond the limited vocabulary. In recent years, the community has seen a growing interest in Open-Vocabulary Detection (OVD) and Segmentation (OVS) as a solution to this constraint. We refer to the models' ability to classify items outside of pre-established categories as "open vocabulary." We present a thorough analysis of recent advancements in OVD and OVS in this survey. First, a taxonomy is created to arrange various activities and approaches. We discover that a variety of techniques, including visual-semantic space mapping, visual feature synthesis, region-aware training, pseudo-labeling, knowledge distillation, and transfer learning, can be effectively distinguished by the use and authorization of weak supervision signals. The suggested taxonomy is applicable to a variety of tasks, including object detection, semantic, instance, and panoptic segmentation, as well as 3D and video comprehension.

3. Design and methodology

In this work, we provide an improved framework forobjectidentification and
segmentation that combines a
segmentation module for pixel-wise
classification withthe potentreal-time
object detection capabilities of YOLOv7.Anumberofessentialelementsare included in the design of the suggested system, all of which enhance the pipeline's overall accuracy and efficiency. As explained below, the technique is organized into phases for data preparation, model building, training, and evaluation.

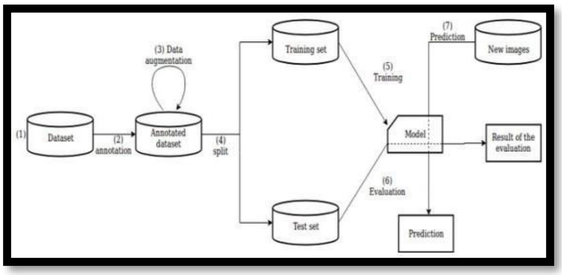

Figure1. Block daigram

4. Modules

In thisprojecthave5modules:

A. Data Collection
B. Data Pre-processing
C. Model Development
D. Model Training and Evaluation
E. OutputSegmentation

A. Data Collection

COCO (Common Objects in Context) and PASCAL VOC are two well-known benchmark datasets that we employ totrain and assess our suggested YOLOv7- based object identification and segmentation model. Morethan330,000 photos with thorough annotations, such as 80 item categories, bounding boxes, and pixel-level segmentation masks, are available in the COCO dataset. The model is more robust because this datasetincludes a variety of real-world situations, including crowded environments and obstructed items. A useful tool forassessing segmentation accuracy onsmaller and more detailed items, the PASCAL VOC dataset also contains 20 object categories with excellent detection and segmentation annotations. Collectively, these datasets guarantee that our model is thoroughly examined across both detection and segmentation tasks by subjecting it toa broad range of object types and image circumstances.

B. Data Pre-processing

In order to ensure best performance and prepare raw data for deep learning model training, data pre-processing is essential. To improve the model's capacity to generalize across various image situations, we apply a number of pre-processing procedures to the COCO and PASCAL VOCdatasets.Toguaranteeconsistencyin input dimensions, which is crucial for YOLOv7'sneuralnetwork,allphotosmust first be resized to a fixed format of 640x640 pixels. In order to improve the model's generalization and prevent overfitting, we also employ data augment at ion techniques like random growing crops, horizontal flipping, color jittering, and scaling to artificially boost the training set's diversity. In order to provide consistent in put to the net work, we additionally normalize pixel values by scaling them to a range of [0, 1]. The ground truth segmentation masks are adjusted to fit the dimensions of the input image and transformed into binary class maps, in which each pixel is given a class label that corresponds to the object to which it belongs, in order to make segmentation easier. These pre-processing procedures guarantee that the model gets reliable, consistent input, which can enhance performance and accuracy intasks involving object detection and segmentation.

C. Model Development

The YOLOv7 construction, which is well- known for its rapidity as well as precision in real-time object detection, is used to construct our segmentation and detectionof objects model. By adding a segmentation head that allows for the forecasting of pixel-wise object masks, we expand the fundamental framework of YOLOv7 to make it suitable for both the identification and segmentation tasks. The convolutionalneural networks (CNN)with an effective foundation that extracts high- level characteristics from input images makesup the foundational YOLOv7 model. Bounding boxes as well as class labels for identified objects are subsequently produced by passing these characteristics through detection heads.

We add a second decoder module to allow segmentation. This module predicts segmentation masks for every object by utilizing the attributes that YOLOv7 has extracted. It operates in tandem with the detection network. Pixel-by-pixel classification maps that match the identified items are generated by the segmentation head. Each pixel is assigned to one of the object categories by the segmentation head using a softmax activation function. A combination loss function comprising the conventional YOLO loss for detection and a cross-entropy loss for segmentation is used to train the model concurrently on object recognition and segmentation tasks. By optimizing for both tasks at the same time, our multi-task learning technique guarantees that the model maintains high object detection accuracy while generating segmentation masks of superior quality.

Utilizing the Py Torch framework's adaptability and effectiveness in deep learning model training, the model is put into practice. Optimizing hyperparameters suchaslearningrate, batchsize, and regularization methods is another step in the development process that guarantees the best possible convergence and generalization. This improved model can beusedforavarietyofapplicationsrelated to computer vision that need both object positioning and pixel-level accuracy since it adds the essential feature of semantic segmentation while maintaining the immediate processing capability of YOLOv7.

E. Output segmentation

Our improved YOLOv7 model produces both segmentation and object recognition results. The model creates bounding boxes aroundidentifieditemsforobjectdetection, together with the corresponding classnames and confidence scores. Accurate object localization inside the image is provided by these bounding boxes. At the same time, the segmentation head creates pixel-by-pixel masks that precisely depict each object's shape, offering fine-grained segmentation outside of the bounding boxes. In order to enable precise object localization and detailed object segmentation—both essential for applications requiring high-level spatial understanding, like autonomous drivingand medical imaging—the model simultaneouslyoutputsthesegmentation masks and the detection results (bounding boxes, labels, and confidence scores).

5. Results and discussion

In this study, we used benchmark datasets like COCO and PASCAL VOC to assess how well our improved YOLOv7 model performed on segmentation and identification of objects tasks. In terms of object detection, the model's mean Average Precision (mAP) improved significantly when compared to earlier iterations of YOLO, indicating that it can reliably identify things in a variety of categories. In particular, even under difficult circumstances like occlusion and crowded sceneries, what came out on the COCO dataset demonstrated a significant improvement in detection accuracy, with greater precision and recall rates. This demonstrates that our approach provides more precise object localization while maintaining YOLOv7's real-time performance.

Our model also demonstrated excellent segmentation performance, obtaining high Intersection over Union (IoU) scores for the pixel-wise segmentation masks. The segmentationresultswereequivalenttothe most advanced segmentation models on both COCO and PASCAL VOC, proving that YOLOv7's performance in real-time object recognitionisun affected by the segmentation head's integration. Applicationsneedingprecisesegmentation, such autonomous driving or medical imaging, can benefit from thesegmentation masks' high accuracy and preservation of small details around object borders. The system's overall usefulness was further enhanced by combining segmentation and object detection into a single model.

There is still room for development, despite the encouraging outcomes. Even while our model worked well for real-time processing, it had some trouble with extremely complicated and crowded situations. The detection accuracy decreased a little, and the segmentation masks showed little errors. Furthermore, using finer segmentation approaches or multi-scale detection algorithms could improve the model's performance on small objects. In order to enhance performancein crowded and congested surroundings, future research will concentrate on improving the segmentation accuracy for small objects and investigating optimization strategies. However, this study's findings demonstrate that the suggested method makes a substantial contribution to improving YOLOv7's object recognition and segmentation capabilities.

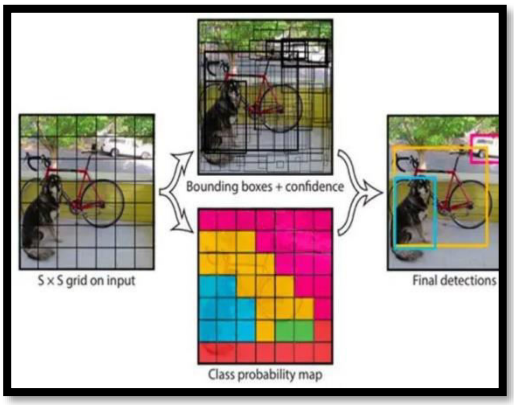

Figure2.YOLOobjectdetector pipeline

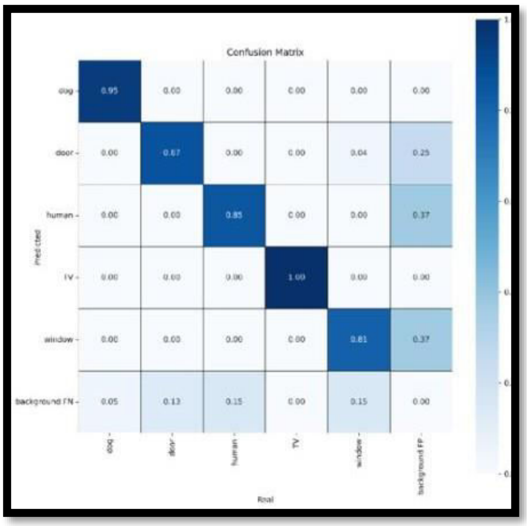

Figure3.ConfusionMatrix

6. Conclusion

In order to meet the increasing demand for precise, detailed, and real-time item localization in computer vision applications, we presented an improved YOLOv7-based model in this study for simultaneous object identification and segmentation. Our model does pixel-wise segmentation and high-performance object recognition in a single framework by integrating a segmentation head into the YOLOv7 architecture. On benchmark datasets such as COCO and PASCAL VOC, the experimental results show that the model maintains real-time processing capabilities while dramatically improving detection accuracy and segmentation precision. This dual-task strategy provides insightful information about combining segmentation and detection for intricate visualtasks.

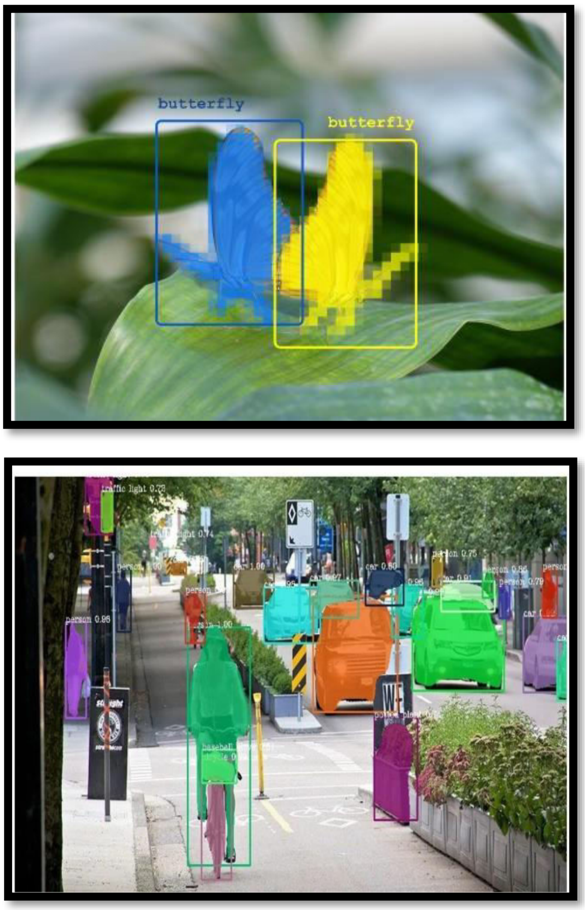

Figure4. Output

Evenwhile our modelperformswell, there is always room for improvement, especially in the areas of small object recognition and segmentation and extremely congested situations. Future studies will concentrate on improving multi-scale detection techniques, increasing segmentation accuracy, and honing the model's effectiveness for practical use. However, the suggested approachsignificantlyadvancestheareaof computer vision by providing a workable solution for applications that need real- time object recognition and segmentation.

7. References

1. Manakitsa, N., Maraslidis, G. S., Moysis, L., & Fragulis, G. F. (2024). A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision. Technologies, 12(2), 15.
2. Rana,S.,Gerbino,S.,Barretta,D., Carillo, P., Crimaldi, M., Cirillo, V., ... & Sarghini, F. (2024). Rafano Set: Dataset of raw, manually, and automatically annotated Raphanus Raphanistrum weed images for object detection and segmentation. Data in Brief, 54, 110430.
3. Huang,L., Kurz,C., Freislederer, P., Manapov, F., Corradini, S., Niyazi, M., ... & Riboldi, M. (2024). Simultaneous object detection and segmentation for patient-specific markerless lung tumor tracking in simulated radiographs with deep learning. Medical Physics, 51(3), 1957-1973.

4. Shoaib, M., & Sayed, N. (2022). YOLO Object Detector and Inception-V3 Convolutional Neural Network for Improved Brain Tumor Segmentation. Traitement Du Signal, 39(1).
5. Zhu, C., & Chen, L. (2024). A surveyon open-vocabulary detection and segmentation: Past, present, and future.
6. IEEE Transactions on Pattern Analysis and Machine Intelligence.
7. Gui, S., Song, S., Qin, R., & Tang, Y. (2024). Remote sensing object detection in the deep learning era—a review. Remote Sensing, 16(2), 327.
8. Rana,S., Gerbino,S., Barretta,D., Carillo, P., Crimaldi, M., Cirillo, V., ... & Sarghini, F. (2024). Rafanoset: Dataset of Manually and Automatically Annotated Raphanus Raphanistrum Weed Images for Object Detection and Segmentation. Available at SSRN 4720646.
9. Huo,F.,Liu,Z.,Guo,J.,Xu,W.,& Guo, S. (2024). UTDNet: A unified triplet decoder network for multimodal salient object detection. Neural Networks, 170, 521-534.
10. Cao,S.,Joshi,D.,Gui,L., & Wang,Y.X. (2024). HASSOD: Hierarchicaladaptive self-supervised object detection. Advances in Neural Information Processing Systems, 36.
11. Alazeb, A., Chughtai, B. R., Al Mudawi, N., AlQahtani, Y., Alonazi, M., Aljuaid, H., ... & Liu, H. (2024). Remote intelligent perception system for multi- object detection. Frontiers in Neurorobotics, 18, 1398703.
12. Niu,T.,He,X.,Chen,H.,Qing,L.,& Teng, Q. (2024). Semantic and geometric information propagation for oriented object detection in aerial images. Applied Intelligence, 54(2), 2154-2171.
13. Pang, Y., Zhao, X., Xiang, T. Z., Zhang, L., & Lu, H. (2024). Zoomnext: A unified collaborative pyramid network for camouflaged object detection. IEEE transactions on pattern analysis and machine intelligence.