

Image detection and segmentation using YOLO v5 for surveillance

Mohanapriya S^{1,2}, Mohana Saranya S¹, Kumaravel T¹, Sumithra P¹

¹Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India

²mohanapriyas.cse@kongu.edu

Abstract. Segmentation an advancement of object detection where bounding boxes are placed around object in object detection whereas segmentation is used to classify every pixel in the given image. In Deep Learning, YOLOv5 algorithm can be used to perform segmentation on the given data. Using YOLOv5 algorithm objects are detected and classified by surrounding the objects with the bounding boxes. Compared to the existing algorithms for segmentation, YOLOv5 algorithm has improved time complexity and accuracy. In this paper YOLOv5 algorithm is compared with the existing CNN algorithm.

Keywords: deep learning, object detection, segmentation, YOLOv.

1. Introduction

1.1. Deep learning

Artificial Intelligence is the one which is used for learning from any unsupervised data that doesn't contain predetermined labels[1]. Making computers think and behave like humans is the idea behind Artificial Intelligence. Deep learning part of the Artificial Intelligence is used to carry out many complicated tasks like recognition of patterns, object detection, segmentation like semantic and instance segmentation. All the functionalities of Deep Learning are carried with huge amount of data which ranges from Gigabytes to Terabytes. With the help of deep learning valuable information can be retrieved from large amount of data [2].

1.2. Object detection

Object detection is crucial in computer vision, automated vehicles, and industrial automation and other applications. It's difficult to detect items in real time. Deep learning outperforms traditional target detection in object detection. This paper provides an object detection and segmentation system that is conceptually simple, versatile, and broad. The method proposed in this paper successfully detects objects in images and videos while producing high-quality segmentation for each instance also. To detect objects, the YOLO algorithm requires only one forward propagation across a neural network. The You Only Look Once model family is a collection of end-to-end deep learning models for quick object detection. The newest algorithm in the YOLO family is YOLOv5. This paper discusses various channel characteristics. It also improves the detecting approach so that more detailed feature

information may be preserved. The final findings suggest that the upgraded YOLO V5 approach enhances performance.[3]

1.3. Segmentation

Segmentation an advancement of object detection is used to detect and classify objects in the image. Instance segmentation is applied in many real time applications like self driving cars, agriculture, medical systems etc. CNN one of the important object detection frameworks is used for detecting objects in the image. All the object detection and segmentation frameworks were developed based on this CNN algorithm. One such detection algorithm is the YOLOv5 algorithm. YOLOv5 algorithm is proved to be the state-of-the-art algorithm for segmentation of objects in the image.[4]

2. Literature review

Image segmentation has gotten a lot of attention recently as one of the successful applications of categorising objects and applying masks for the item present in the image. The study addresses over a hundred deep learning-based segmentation algorithms proposed through 2019 and includes the most recent research on instance segmentation [5]. We present a thorough examination and analysis of several elements of these approaches, including training data, network architecture selection, loss functions, training state, and major contributions. We give a comparison of the performance of the approaches under consideration, as well as many obstacles and possible future directions for deep learning-based instance segmentation models. YOLO established a single unified architecture for breaking go picture into bounding boxes and calculating class probabilities for each box, in comparison to object identification approaches that came before it, such as R-CNN. As a result, YOLO was able to execute significantly faster and with greater precision. It may also properly anticipate artwork. [6]

Object Detection aims to construct a general object recognition network, complex degradation methods including noise, blurring, rotating, and cropping of images were applied. The model's generalisation and robustness were improved by employing degraded training sets during training. The study found that the model's generalisation and resilience when used on damaged images were weak when trained on standard sets. After training the model with damaged images, average accuracy increased. It was demonstrated that the wide degenerative model outperformed the conventional model in terms of average accuracy for degraded images.

The YOLO Network Model says an improved network model is developed and a new network structure known as YOLO-R has been proposed to boost the network's capacity to extract information from superficial pedestrian characteristics by including pass through layers into the original YOLO network. The INRIA data collection's test set had been used to assess YOLO v2 and YOLO-R network models. Compared to YOLO v2 network model, YOLO-R network model performs better. The real-time performance criterion was met when the detecting frame rate increased to 25 frames per second.

Solder Joint Recognition and Detection in Automotive Door Panels, a solder joint recognition method based on the YOLO algorithm that gives the kind and location of solder joints in real time for automobile door panels. In order to more easily identify tiny patch crossings, this study applies the YOLO approach, which employs staggered forecasts, expecting on many size highlight guides, and merging the expectation outcomes to form the final conclusion. The proposed YOLO approach successfully locates solder connections in real time. This increases the productivity of the production line and is crucial for the flexible and real-time welding of vehicle door panels.

Though many works have been proposed to address the problem of object detection and segmentation, still a research gap available to improve accuracy in this area. This paper focuses to use YOLO v5 algorithm for object detection and segmentation to improve that gap.

3. Proposed work

3.1. YOLO V5 algorithm

YOLO algorithm is known for its high performance and quick time requirements. It is one of the most popular deep convolutional neural methods for object segmentation. The PyTorch framework is used in YOLOv5. It is the most recent version of the YOLO object recognition model, which was created with the help of 58 open source contributors throughout time. Other deep neural networks may be used to detect things as well. One of them is the Mask-RCNN [9], which is designed to handle the problem of instance segmentation in computer vision machine learning. Mask-RCNN is more exact, but it takes longer to process. YOLO and Mask R-CNN models give results of high recall and precision for detecting a ball sports. The YOLOv5 network is used in this paper since it is a good and quicker detector with excellent levels of performance. Other architectures, such as the MaskRCNN, may be able to achieve comparable detection results while providing more exact object positioning.

3.2. Working principle of YOLO V5

In YOLO algorithm, the image is divided into ‘n’ grids of equal size. In each grid, the object contained in that grid is detected and localized. The grids are responsible for predicting coordinates of bounding box according to their cell coordinates. Prediction in this way greatly reduces computation of detection and recognition but it leads to duplicate predictions. This issue is dealt with maximal suppression in YOLO

YOLO eliminates the bounding boxes that have probability score very minimum. This is done by seeing the score of each decision and finding out which one is the largest. After finding the largest value, YOLO eliminates all the bounding boxes which have the highest IoU value with the current bounding box which have the high value. The above step is repeated until the target bounding box is obtained in Figure 1.

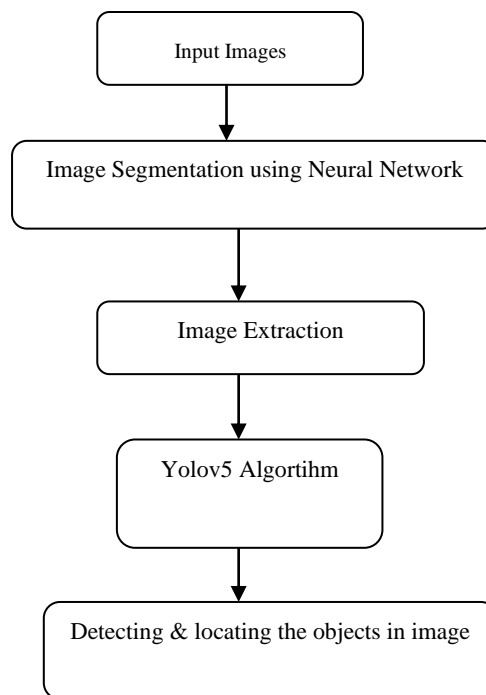


Figure 1. Flow chart of proposed work.

3.3. Model backbone

From the raw photos provided, Model Backbone is used to extract significant characteristics. To extract highly useful data from an input image, Cross Stage Partial Networks (CSP) can be employed as the backbone.

3.4. Model neck

The fundamental purpose of a model neck is to build feature pyramids. Models generalise easily to objects of various sizes thanks to feature pyramids. The ability to recognise the same thing in various sizes and scales is helpful. Models that employ feature pyramids perform well on unobserved data. PANet is utilised in Yolo V5 as a neck to get feature pyramid.

3.5. Model head

The last detecting step is carried out by the model Head. It applies anchor boxes to features and creates a final output vector with bounding boxes, an objectness score, and a class likelihood.

3.6. Object detection using neural network

Using a neural network classifier with a feed-forward, one hidden layer network and back propagation as the learning method, an object detection algorithm is built. The definition of efficient object characteristics, which are utilised to train the classifier, is a crucial component of this system.

4. Results and discussion

4.1. Dataset

A COCO dataset of nearly 10-20 lakhs that has already been trained by using predefined functions is used to assess the proposed work. The dataset images are frame-by-frame trained. From the COCO dataset, we took 5000 images for testing. The MS COCO dataset offers a sizable dataset for object recognition and instance segmentation, both of which were used to test several deep learning techniques. Figure 2 demonstrate an example input image and output from the dataset.



Figure 2. Input and output image.

4.2. Accuracy

Accuracy is used to measure how the model performs for different classes of objects. It is the ratio between total number of correct predictions to the total number of predictions made. The Yolo V5 and CNN algorithms' degrees of accuracy are displayed in Table 1.

Table 1. Accuracy level (%) of Yolo V5 and CNN.

No. of images	Time Complexity of Yolo V5(ms)	Time Complexity of CNN(ms)
1000	0.22	1.36
2000	0.14	2.20
3000	0.40	2.207
4000	0.10	1.94

Figure 3 shows the accuracy level comparison of Yolo V5 and CNN. In the figure, we can see that Yolo V5 performs better than CNN.

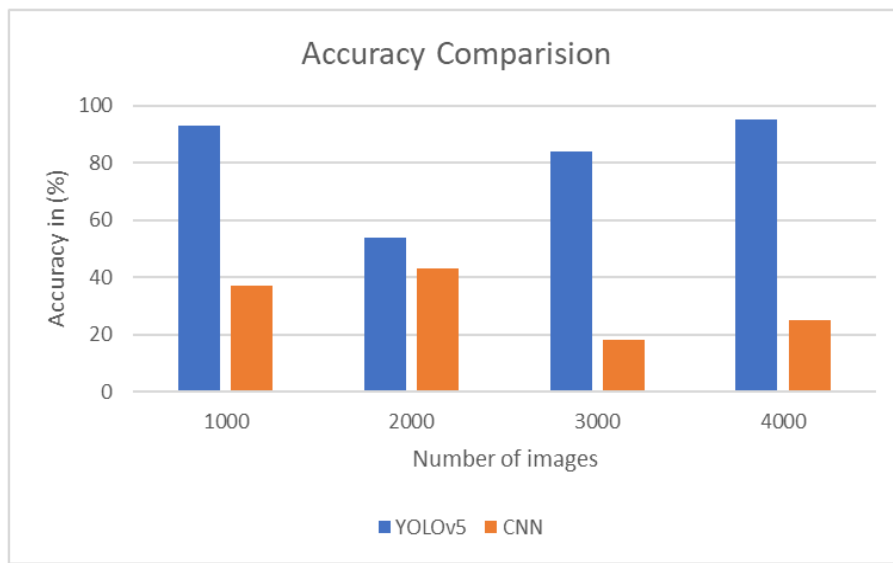


Figure 3. Accuracy level comparison of Yolo V5 and CNN.

4.3. Time complexity

The amount of time taken to run an algorithm is known as the Time complexity. It is known as the Computational Complexity. It can be measured in ms. The Time complexity of the Yolo V5 and CNN algorithms are shown in Table 2.

Table 2. Time complexity (%) of Yolo V5 and CNN.

No. of images	Accuracy of Yolo V5(%)	Accuracy of CNN (%)
1000	93	37
2000	54	43
3000	84	18
4000	95	25

The comparison of time complexity between Yolo V5 and CNN is shown in Figure 4. The figure shows.

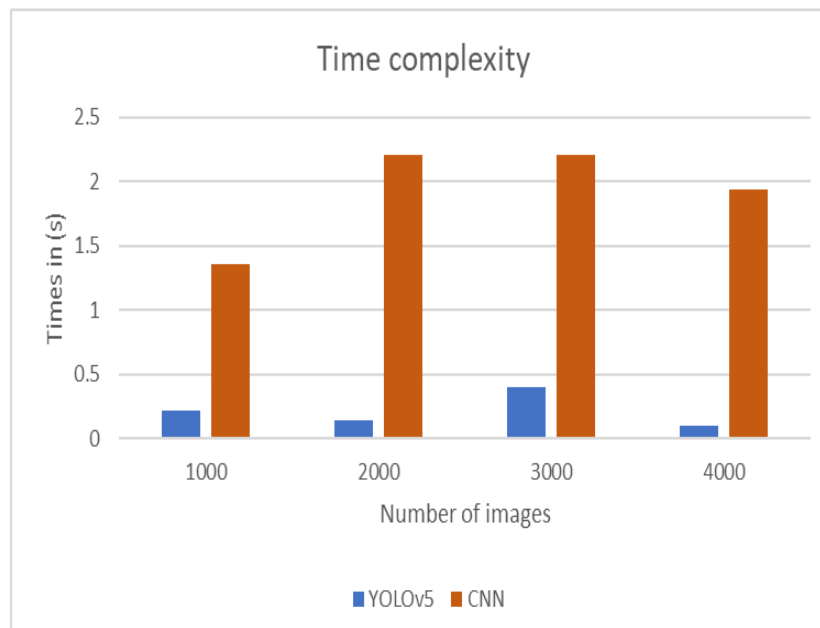


Figure 4. Time complexity comparison of Yolo V5 and CNN.

5. Conclusion and future work

The segmentation technique an advancement of object detection is used to detect and classify pixels in the image. The YOLOv5 method, which is based on deep learning and excels at object detection, has been made available. YOLOv5 significantly reduces time complexity and improves segmentation accuracy when compared to earlier state-of-the-art algorithms. As a result, YOLOv5 is a superior option for identifying things and determining objects in the image.

References

- [1] Sathishkumar, V. E., Cho, J., Subramanian, M., & Naren, O. S. (2023). Forest fire and smoke detection using deep learning-based learning without forgetting. *Fire Ecology*, 19(1), 1-17.
- [2] Subramanian, M., Cho, J., Sathishkumar, V. E., & Naren, O. R. (2023). Multiple types of Cancer classification using CT/MRI images based on Learning without Forgetting powered Deep Learning Models. *IEEE Access*.
- [3] Kogilavani, S. V., Sathishkumar, V. E., & Subramanian, M. (2022, May). AI Powered COVID-19 Detection System using Non-Contact Sensing Technology and Deep Learning Techniques. In *2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS)* (pp. 400-403). IEEE.
- [4] Shanmugavadeivel, K., Sathishkumar, V. E., Kumar, M. S., Maheshwari, V., Prabhu, J., & Allayear, S. M. (2022). Investigation of Applying Machine Learning and Hyperparameter Tuned Deep Learning Approaches for Arrhythmia Detection in ECG Images. *Computational & Mathematical Methods in Medicine*.
- [5] Krishnamoorthy, N., Prasad, L. N., Kumar, C. P., Subedi, B., Abraha, H. B., & Sathishkumar, V. E. (2021). Rice leaf diseases prediction using deep neural networks with transfer learning. *Environmental Research*, 198, 111275.
- [6] Easwaramoorthy, S., Sophia, F., & Prathik, A. (2016, February). Biometric Authentication using finger nails. In *2016 international conference on emerging trends in engineering, technology and science (ICETETS)* (pp. 1-6). IEEE.