SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V10I10P104

Volume 10 Issue 10, 40-48, October 2023

Original Article

SegMatic: A Deep Neural Network Learning Model for Semantic Segmentation

K.G. Suma[1], Gurram Sunitha[2], J. Avanija[3]

[1]*School of CSE, VIT-AP, Amaravati, Andhra Pradesh, India.*
[2,3]*Department of AI & ML, School of Computing, Mohan Babu University, Tirupati, Andhra Pradesh, India.*

[2]*Corresponding Author : gurramsunitha@gmail.com*

Received: 05 August 2023 Revised: 19 September 2023 Accepted: 09 October 2023 Published: 31 October 2023

Abstract *- Semantic segmentation enables vehicles to accurately identify and categorize objects in their surroundings, such as pedestrians, other cars, road signs, and obstacles. Semantic segmentation, object detection, and deep learning have emerged as critical pillars, enabling machines to perceive and understand the visual world with unprecedented precision. This paper introduces SegMatic, a novel deep-learning model specifically tailored to address the unique demands of autonomous vehicles. SegMatic is a novel deep model for semantic segmentation and precise object detection. This model harnesses the power of deep learning to transform raw images into pixel-wise semantic maps, providing detailed insights into object boundaries and category-specific regions. SegMatic employs a two-stage approach. It uses a modified U-Net as the first stage to extract feature maps. Mask R-CNN is used as the second stage for post-processing. Experiments are conducted on the Pascal VOC 2012 dataset. SegMatic outperformed traditional models with remarkable precision and pixel accuracy scores. It achieved superior results in both semantic segmentation and object detection. This success is evidenced by achieving a mIoU of 94.6 and PA of 95.7 across various object categories. These results substantiate the significance of SegMatic's contributions to computer vision and deep learning.*

Keywords *- Deep learning, Image segmentation, Object detection, U-Net, Mask R-CNN.*

1. Introduction

Accurate object detection and localization are paramount in computer vision [1, 2]. They enable machines to identify and precisely position objects within images. Underpinning applications range from autonomous vehicles to medical imaging, with safety, diagnostics, and decision-making implications. Precise object detection is intricately intertwined with semantic segmentation [3]. It is the semantic understanding of images that empowers object detection models to not only identify objects but also to delineate them with remarkable precision.

Semantic segmentation involves labelling each pixel in an image with a specific class. This task is essential because it enables machines to gain a more profound visual understanding of the data they process [4]. Thus allowing them to decipher not just the presence of objects in an image and their precise boundaries and spatial distribution.

Semantic segmentation has witnessed unprecedented advancements in deep learning [5, 6]. This progress, coupled with the symbiotic relationship between semantic segmentation and accurate object detection, has ignited innovations that transcend the boundaries of traditional computer vision. Deep learning is instrumental in underpinning the accuracy and efficacy of semantic segmentation [7]. CNNs, with their inherent capacity to hierarchically extract features from images, have propelled the boundaries of what can be achieved in pixel-level object recognition.

Accurate segmentation is crucial for understanding the visual content of images, enabling applications such as object recognition, scene understanding, and image-based navigation [8]. However, achieving high accuracy in semantic segmentation remains a complex problem. The impeding factors are object occlusion, varying lighting conditions, the presence of fine-grained details, etc.

This research proposes a novel framework, "SegMatic", that leverages deep learning to achieve accurate object detection and image localisation. First, a modified U-Net architecture forms the backbone of the semantic segmentation module. This U-Net, enriched with attention mechanisms and skip connections, excels in capturing intricate object boundaries and multi-scale features. Second, Mask R-CNN is used as the post-processing technique. Mask

R-CNN extends the model's capabilities by offering fine-grained object detection. Together, these components provide a comprehensive framework that exemplifies the significance of semantic segmentation.

2. Literature Review

Deep learning has revolutionized image segmentation by enabling the automated extraction of intricate object features [9]. Thus, improving precision reduces the need for handcrafted features or manual intervention [10]. Its ability to learn complex hierarchical patterns from data has unlocked new possibilities [11]. Medical imaging, autonomous systems [12], and countless applications reliant on accurate and efficient image segmentation were developed using deep learning [13-16].

Neural networks are a fundamental tool used for semantic segmentation [17]. While not focused on object detection, FCNs set the stage for subsequent advancements in pixel-wise labelling. Many later models, including those for object detection and localization, adopted FCN-inspired architectures for semantic segmentation [18]. Mask R-CNN is a pioneering model that seamlessly combines instance segmentation with object detection [19]. It has set a new standard, for instance, segmentation and object localization, achieving appreciable results on various benchmarks. U-Net is a widely recognized architecture for semantic segmentation, especially in biomedical image analysis [20]. Its encoder-decoder structure, featuring skip connections, has influenced the design of various models, making it highly relevant to both semantic segmentation and object detection.

Visual Transformers represent a novel approach to object detection and localization, applying the transformer architecture to images [21, 22].

3. Proposed Model

The proposed SegMatic model is a custom-designed architecture that addresses the intricate tasks of semantic segmentation object detection (Figure 1). This innovative framework conjoins the capabilities of semantic segmentation object detection. Semantic segmentation is accomplished through a modified U-Net architecture.

The integration of Mask R-CNN facilitates object detection as a powerful post-processing technique. The semantic segmentation segment employs an encoder-decoder structure with attention mechanisms. The post-processing phase operates Mask R-CNN's backbone, Region Proposal Network (RPN), Regions of Interest (ROI) alignment, heads for classification, and Bounding Box Regression (BBR).

The proposed model represents a seamless integration of semantic segmentation and object detection with the aid of Mask R-CNN as a post-processing technique. The semantic segmentation component captures multi-scale features. It has the potential to delineate object boundaries with precision. The incorporation of Mask R-CNN offers fine-grained object detection, localization, and instance-specific segmentation. This enables a deeper understanding of complex visual scenes. The proposed framework is targeted towards efficient object detection while maintaining high standards of semantic segmentation.

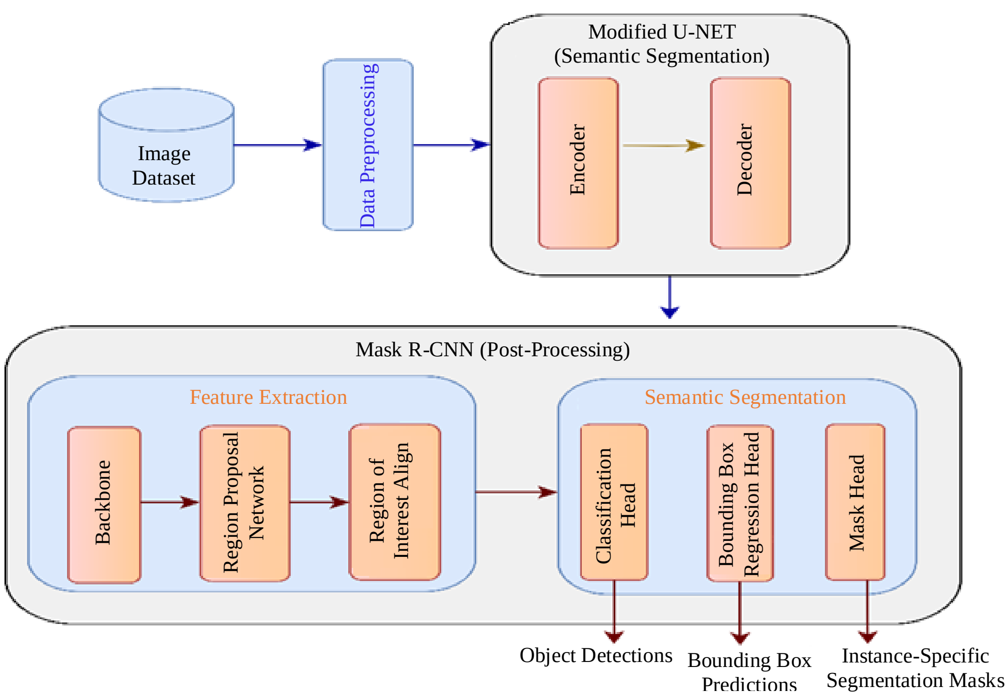


Fig. 1 Proposed "SegMatic" deep model for semantic segmentation and object detection

3.1. Semantic Segmentation (U-Net)

The mathematical model for the U-Net-based semantic segmentation involves convolutional layers, skip connections, and activation functions (ReLU). Given an input feature map X with dimensions HxW representing image size, C represents the number of channels. Let Y denote the output feature map of U-Net.

$$Y = ReLU(X \times K + b) \tag{1}$$

K represents the convolutional kernel with learnable weights; b is the bias term. The loss function for semantic segmentation is cross-entropy loss.

3.2. Mask R-CNN (Post-Processing)

Mask R-CNN involves several components.

1. RPN
2. ROI Align
3. Classification Head
4. BBR Head
5. Mask Head

The mathematical models of each component are presented in the following sections.

3.3. Region Proposal Network (RPN)

RPN uses a set of convolutional layers to generate region proposals (bounding box coordinates) and assigns scores to these proposals. The mathematical model includes the computation of scores and the generation of proposals based on anchor boxes.

Let Y be the feature map from the backbone network, with size $HxWxC$, where HxW is the image size, and C is the number of channels. Let its size and aspect ratio characterize each region proposal. The RPN component performs two main tasks.

3.3.1. Generate Region Proposals (Bounding Boxes)

Figure 2 presents the algorithm for generating region proposals. A position-sensitive score si is computed using a convolutional layer for each anchor box ai in A.

Sigmoid activation is applied to s_i to obtain a probability score σ (s_i). σ (s_i) indicates whether the anchor box contains an object. BBR offsets Δb_i are calculated for each anchor box using another convolutional layer. Finally, region proposals R_i are generated based on anchor boxes a_i and their associated bounding box regression offsets.

3.3.2. Classification of Region Proposals

The algorithm for the classification of region proposals is presented in Figure 3. The classification score is assigned to vector c_i for each region proposal R_i for each class using a convolutional layer. Softmax activation is then applied to c_i to obtain the class probabilities. The computation of s_i, p_i, Δb_i, and c_i involves convolutional layers with learnable weights (W_s, W_b, W_c). The convolutional layers effectively process the input feature map Y to produce the required scores, probabilities, and bounding box regression offsets. Final region proposals R_i are obtained by applying Non-Maximum Suppression (NMS) to retain a subset of high-scoring proposals while discarding redundant ones.

Parameters:

Y = Feature Map from Backbone
|Y| = HxWxC
A = Set of Regions Proposals
N = |A|
S = HxW = Spatial Dimensions of Feature Map
B = (x, y, w, h) = Number of BBR Branches
P = Number of Classes
S_i = Position-Sensitive Score
σ (s_i) = Probability Score of Class i
Δb_i = BBR offsets
R_i = Predicted Region Proposal
W_s = Learnable Weights
Σ = Sigmoid Function

Algorithm:

1. For each a_i in A
2. $s_i = Conv(Y, W_S)$
3. $p_i = \sigma(s_i)$
4. $\Delta b_i = Conv(Y, W_b)$
5. $R_i = AnchorBoxTransform(a_i, \Delta b_i)$
6. end for

Fig. 2 Algorithm for generating regions proposals

Parameters:

c_i = Vector of Classification Scores
R_i = Predicted Region Proposal
$P(c_i | R_i)$ = Class Probability

Algorithm:

1. For each R_i in R
2. $c_i = Conv(Y, W_C)$
3. $p(c_i|R_i) = softmax(c_i)$
4. end for

Fig. 3 Algorithm for classification of regions proposals

3.4. ROI Align

It uses bilinear interpolation to extract features from ROIs in the feature maps produced by the RPN. Let F be the feature map produced by RPN, with dimensions $H_f \times W_f \times C_f$, where $H_f \times W_f$ is the size, and C_f is the number of channels.

Let *R* be an ROI specified by its coordinates *(x, y, w, h)*, where *(x, y)* is the top-left corner, and *(w, h)* is the ROI size. Let *P* be the output feature map from ROI Align for the specified ROI. The mathematical model for ROI Align can be defined as follows.

3.4.1. ROI Proposal Transformation

- Transform ROI coordinates from the original image space to feature map space. This transformation involves scaling and quantization, aligning ROI with the feature map's spatial dimensions.
- Calculate the size of each bin in output feature map *P*. It is done by dividing the ROI size by the corresponding P size.
- Divide ROI into a grid of equally sized bins based on the calculated bin sizes.

3.4.2. Bilinear Interpolation

- For each bin in the grid, perform bilinear interpolation to extract features from input feature map *F*. Bilinear interpolation considers the four nearest neighbour values in *F* to compute the interpolated value for each bin.
- The interpolated values are collected for each bin, resulting in a set of features corresponding to ROI in the input feature map.

3.4.3. Output Feature Map

The extracted features from all bins form *P*. Dimensions of *P* are determined by the number of bins in the grid. The bilinear interpolation step is crucial in preserving spatial information and accurately aligning features from *F* with ROI. It ensures that elements are sampled at sub-pixel locations within the ROI. Thus contributing to the object detection capabilities of the model.

Parameters:

$|F|_{1d} = |H_p \times W_p \times C_p|$

$|P|_{3d} = H_p \times W_p \times C_p$

$|C| = |Z| = |S|$ = Number of Class Labels

Z = Raw Class Scores for Each of C Classes

W_{cls} = Weight Matrix

b_{cls} = Bias Vector

S = Class Scores

S_i = Probability of ROI Belonging to Class i

Algorithm:

1. F = Flatten (P)
2. $Z = F \times W_{cls} + b_{cls}$
3. $S_i = \frac{e^{Zi}}{\sum_{j=1}^{C} e^{Zj}}$

Fig. 4 Algorithm for classification head

3.5. Classification Head

It assigns class labels to objects within ROIs. The mathematical model involves matrix multiplications and softmax calculations (Figure 4).

Parameters:

W_{reg} = Weight Matrix of FCL

b_{reg} = Bias Vector of FCL

O = Raw Predictions of BB Offsets

$|O|$ = B

G = Ground Truth BB Offsets

$|F|_{1d} = H_p \times W_p \times C_p$

$|P|_{3d} = |H_p \times W_p \times C_p|$

L = Smooth L1 Loss

Algorithm:

1. F = Flatten (P)
2. $O = F \times W_{reg} + b_{reg}$
3. $S_i = \frac{e^{Zi}}{\sum_{j=1}^{C} e^{Zj}}$
4. $\text{SmoothL1}(x) = \begin{cases} 0.5x^2, \text{if } |x| < 1 \\ |x| - 0.5, \text{otherwise} \end{cases}$
5. $L(O, G) = \sum_{i=1}^{B} \text{SmoothL1}(O_i, G_i)$

Fig. 5 Algorithm for BBR head

Parameters:

M = Number of Class Labels

S = Predicted Segmentation Mask for a Specific Class

$|S|$ = H_p x W_p x M

G = Ground Truth Binary Segmentation Mask for a Specific Class

W_{mask} = Weight Matrix of FCL

b_{mask} = Bias Vector of FCL

L = Binary Cross Entropy Loss

Algorithm:

1. S = Conv (P, W_{mask}, b_{mask})
2. $p_i = \sigma(s_i)$
3. $L(S, G) = -\frac{1}{N}[G_i \times \log(P_i) + (1 - G_i) \times \log(1 - P_i)]$

Fig. 6 Algorithm for mask head

3.6. BBR Head

This head predicts the bounding box coordinates (offsets from anchor boxes) for the objects within ROIs. FCL is used for BBR. It uses the Smooth L1 Loss. The mathematical model BBR head is presented in Figure 5. Smooth L1 is computed independently for each component of the bounding box.

3.7. Mask Head

It predicts instance-specific segmentation masks. It uses convolutional layers and binary cross-entropy loss for mask prediction. The output feature map is passed through convolutional layers to predict the segmentation mask for each class.

Convolutional layers are parameterized by weights W_{mask} and biases b_{mask}. Sigmoid activation is then applied to the expected mask tensor S to obtain pixel-wise probabilities. Finally, binary cross-entropy loss L between P and G is calculated. L is computed for each pixel independently and averaged over N pixels. The mathematical model is presented in Figure 6.

4. Results and Discussion

4.1. Fine-Tuned Deep Model

SegMatic deep model for deep semantic segmentation and object detection is a carefully designed framework that leverages the power of deep networks to achieve precise and robust results.

The model comprises several components, each with its specific role and configuration. The initial phase of the model is the input layer, which takes in the pre-processed image data. This ensures that the model receives consistent and well-prepared data for processing.

The feature extraction phase captures hierarchical features from the input image. In this phase, U-Net serves as the backbone. The backbone's layers and filters are configured to balance computational efficiency and feature representation (Figure 7 and Figure 8).

The model splits into two main branches following feature extraction to perform semantic segmentation and object detection simultaneously. The first branch focuses on semantic segmentation. It consists of a series of convolutional layers with skip connections. It captures multi-scale contextual information.

The output layer employs a softmax activation function to assign class probabilities to each pixel. Thus generating a semantic segmentation mask.

The second branch, for object detection, includes multiple sub-components. RPN is responsible for proposing potential object regions (Figure 9). It generates region proposals based on predefined anchor boxes. It computes objectness scores and bounding box regressions. RPN is configured with anchor scales and aspect ratios that are carefully chosen based on the dataset's characteristics.

Following RPN, the model incorporates ROI Align layers. They crop and align feature maps for each proposed region. This step ensures that object features are accurately localized within their respective bounding boxes. ROI Align mitigates the quantization errors commonly associated with ROI pooling. The classification head consists of FCL followed by a softmax activation function to assign class labels to objects within ROIs (Figure 10). BBR head predicts BB coordinates as offsets from the anchor boxes.

Table 1. Dataset description

Item	Description
# Classes	20
# Images	11530
# Training Images	5717
# Validation Images	5823
# Test Images	1449
Environment	Multiple Instances, Objects, Camera Angles, Scale Variations
Annotations	Labels, Bounding Boxes, Pixel-wise Semantic Segmentation

The model includes NMS as a post-processing step to filter and refine object detections. Combining these techniques ensures that the model outputs accurate and reliable object detections. Figure 11 presents the loss functions used to fine-tune the components of the deep model.

Ultimately, the model generates an output image showcasing profound semantic segmentation results, with each class colour-coded for clarity and precise object detection with bounding boxes, class labels, and confidence scores.

4.2. Dataset

Pascal VOC 2012 is a widely used benchmark dataset in computer vision (Table 1). It contains extensive images and annotations for multiple object classes and pixel-wise semantic segmentation masks. The dataset includes 20 object classes, including common categories such as person, car, dog, cat, bicycle, bus and more.

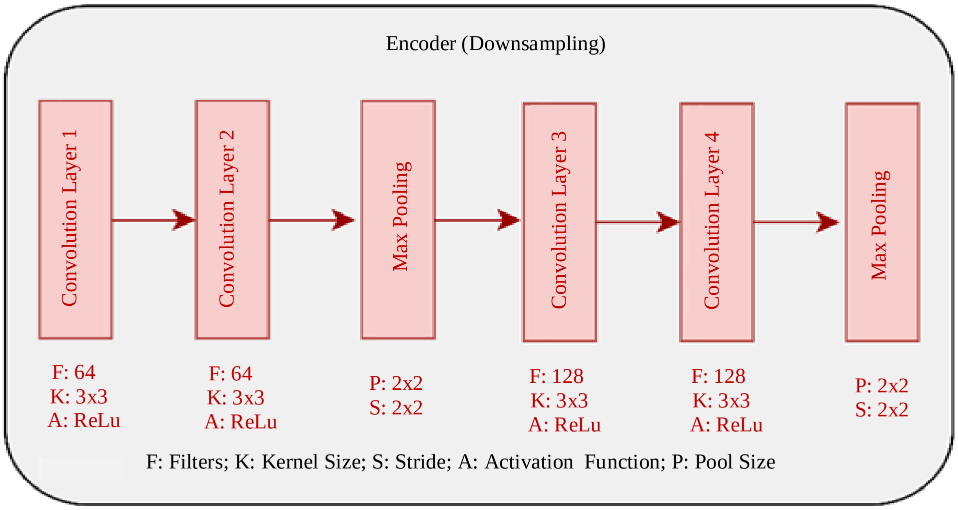


Fig. 7 Proposed encoder model

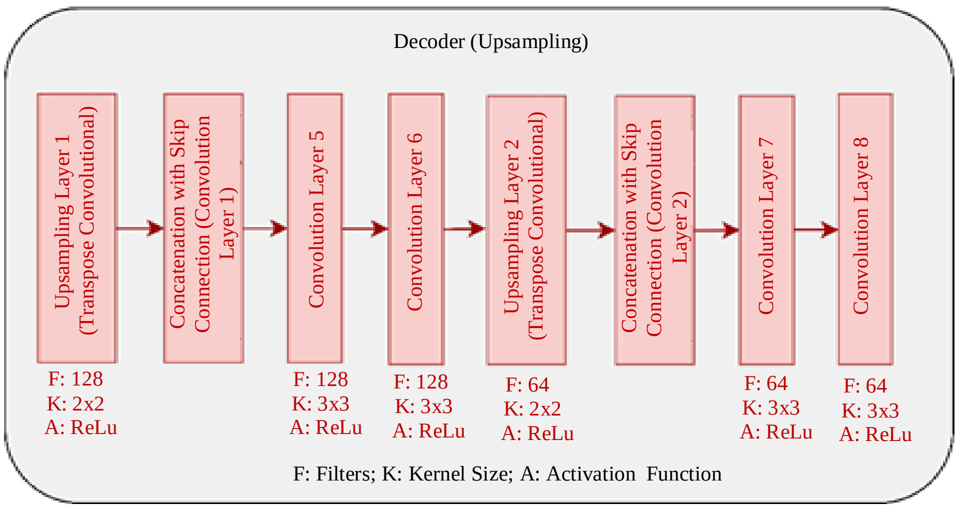


Fig. 8 Proposed decoder model

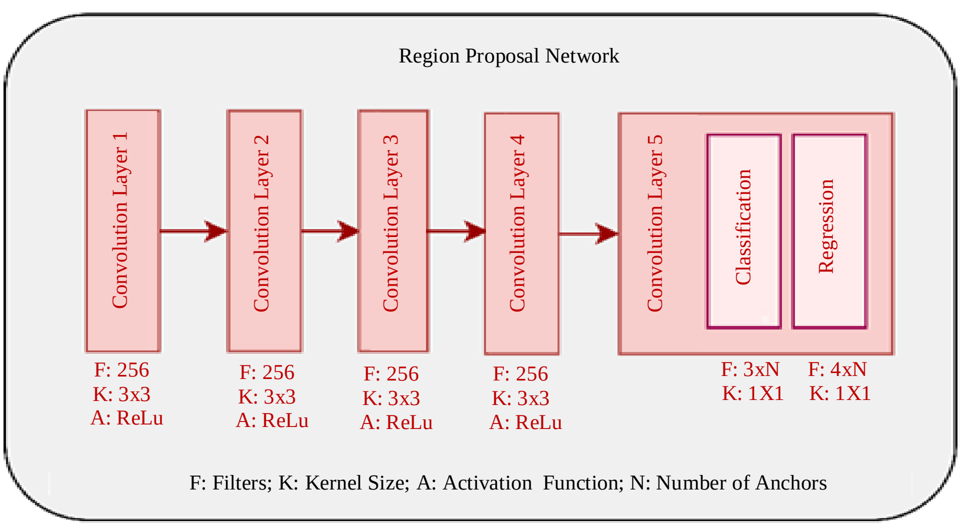
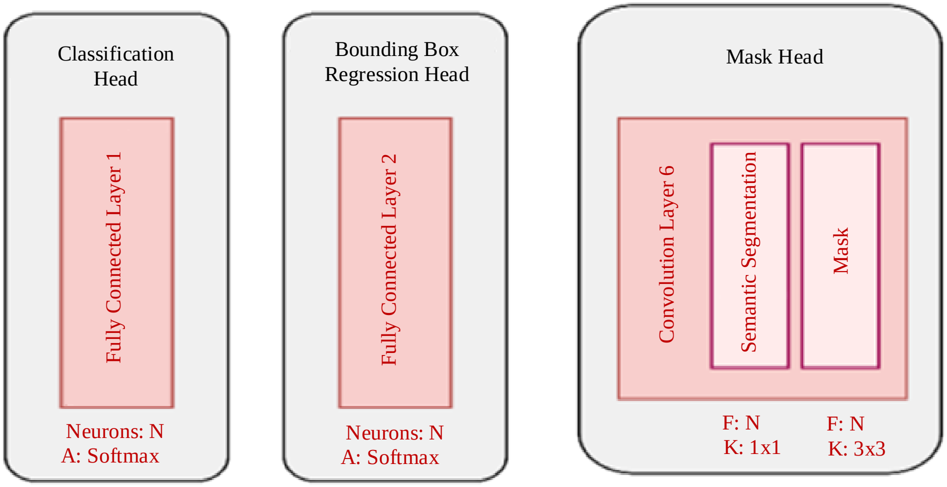


Fig. 9 Proposed region proposal network



F: Filters; A: Activation Function; N: Number of Class Labels

Fig. 10 Proposed classification head, BBR head, mask head models

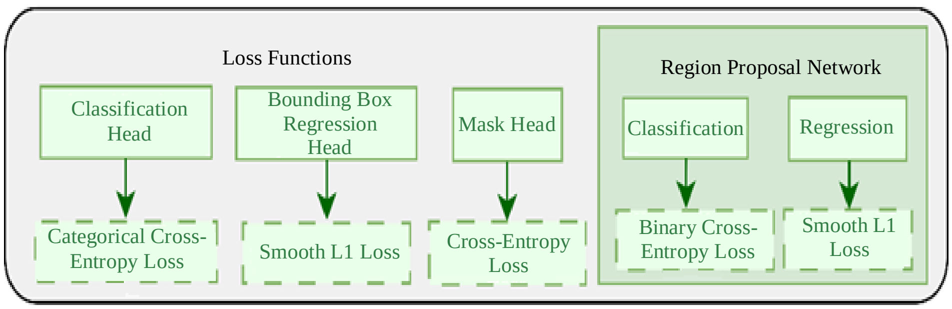


Fig. 11 Loss functions

(a) Input Image (b) Semantic Segmentation Map (c) Object Detection

Green-Grass; Purple-Road; Red-Object; Blue-Motor Bike

Fig. 12 Samples of experimentation

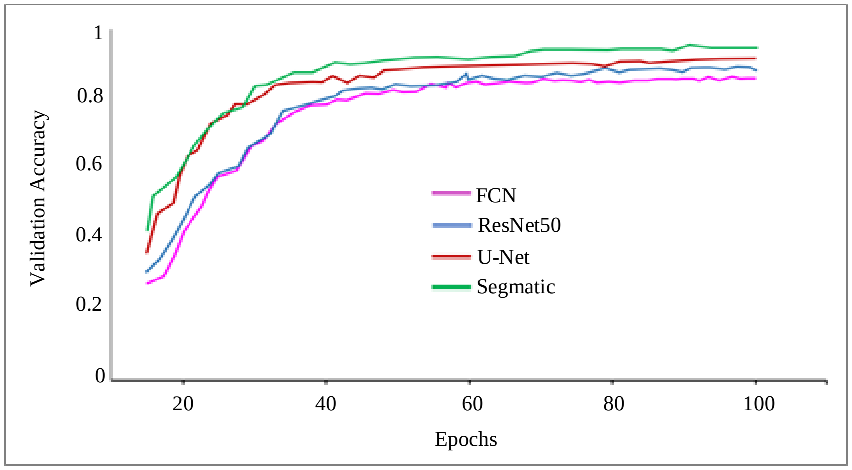


Fig. 13 Performance of deep models

Table 2. Performance of deep models

Model	Accuracy	Precision	F1-Score	mIoU	PA
FCN	87.2	86.4	81.2	74.2	78.1
ResNet-50	89.7	93.1	94.8	93.5	92.6
U-Net	92.3	92.8	93.3	92.8	94.3
SegMatic	95.8	93.5	93.9	94.6	95.7

4.3. Experimental Results

The segMatic model utilizes a custom backbone network and a decoder with specialized post-processing techniques for accurate object detection. The proposed model is evaluated against three baseline models – FCN (classic semantic segmentation model based on fully convolutional layers); ResNet-50; which is used as a backbone architecture; and U-Net (well-known architecture featuring a contracting and an expansive path for semantic segmentation). The models' hyperparameters, including learning rates, batch sizes, and optimization techniques, are fine-tuned to ensure they converge effectively during training.

The results of experiments on the dataset are evaluated in terms of crucial evaluation metrics. It is a critical metric for assessing the quality of object detection. Figure 12 presents the sample input images, semantic segmentation maps and object detections. Table 2 and Figure 13 shows the evaluation results.

SegMatic operates on raw images from the Pascal VOC dataset, adding considerable complexity to the segmentation model. It works well mainly when contrasted with the other approaches that rely on smaller ROIs. Nevertheless, the proposed method consistently outperformed compared alternatives across various metrics. It achieved superior results in both semantic segmentation and object detection. This success is evidenced by achieving a mIoU of 94.6 and PA of 95.7 across various object categories. The performance of the proposed model can be attributed to several factors.

The proposed model features a custom-designed architecture optimized for deep semantic segmentation. Traditional models, like FCN or ResNet, have been initially designed for different tasks. The proposed model is purpose-built for accurate object detection. This tailored architecture allows it to extract and represent features in a way that is more conducive to semantic segmentation.

The proposed model incorporates specialized semantic segmentation modules that handle complex object boundaries, intricate textures, and diverse object shapes. These modules enabled the model to capture fine-grained details in object segmentation, contributing to higher accuracy and PA.

The proposed model employs the advanced post-processing technique Mask R-CNN to refine object boundaries and eliminate false positives. This enhanced the overall quality of segmentation masks, resulting in improved PA**.**

5. Conclusion

Semantic segmentation empowers machines to distinguish objects, comprehend their spatial relationships, and interpret complex scenes. Object detection is an indomitable force in applications ranging from autonomous navigation to medical diagnostics. SegMatic is an efficient deep model designed for accurate and reliable object detection and classification. Its architecture includes a powerful backbone, semantic segmentation, object detection, and post-processing. The segMatic model represents a remarkable demonstration of how machines understand images. It is all about recognizing objects, understanding their positions, and using the power of deep learning to do this exceptionally well. It allows machines to learn and become better at understanding what they see. SegMatic, with its holistic approach, has the potential to enhance object detection precision alongside semantic segmentation accuracy.

References

[1] Zhengxia Zou et al., "Object Detection in 20 Years: A Survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257-276, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Esmaeil Mirmahdi, and Omid Ghorbani Shirazi, "Installation of Suitable Sensors for Object Detection and Height Control on Combine Harvester," *SSRG International Journal of Mechanical Engineering*, vol. 8, no. 5, pp. 12-19, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ashish Kumar Gupta et al., "Salient Object Detection Techniques in Computer Vision-A Survey," *Entropy*, vol. 22, no. 10, pp. 1-49, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Hongshan Yu et al., "Methods and Datasets on Semantic Segmentation: A Review," *Neurocomputing*, vol. 304, pp. 82-103, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[5] Indrabayu et al., "Various Obstacles Detection Systems Using Single Shot Multi-Box Detector (SSD) for Autonomous-Driving Vehicles," *International Journal of Engineering Trends and Technology*, vol. 71, no. 5, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] E. Parimalasundar et al., "Investigation of Efficient Multilevel Inverter for Photovoltaic Energy System and Electric Vehicle Applications," *Electrical Engineering & Electromechanics*, no. 4, pp. 47-51, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] D. Raja, and M. Karthikeyan, "Plant Disease Detection and Classification Based on Rat Swarm Optimization Using Deep Learning Approach," *International Journal of Engineering Trends and Technology*, vol. 71, no. 7, pp. 42-52, 2023. [CrossRef] [Publisher Link]

[8] Yujian Mo et al., "Review the State-of-the-Art Technologies of Semantic Segmentation Based on Deep Learning," *Neurocomputing*, vol. 493, pp. 626-646, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Shervin Minaee et al., "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523-3542, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Swarnendu Ghosh et al., "Understanding Deep Learning Techniques for Image Segmentation," *ACM Computing Surveys*, vol. 52, no. 4, pp. 1-35, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] Foivos I. Diakogiannis et al., "ResUNet-A: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94-114, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Manasa R., K. Karibasappa, and Rajeshwari J., "Autonomous Path Finder and Object Detection Using an Intelligent Edge Detection Approach," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 9, no. 8, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Parimalasundar Ezhilvannan et al., "Analysis of the Effectiveness of a Two-Stage Three-Phase Grid Connected Inverter for Photovoltaic Applications," *Journal of Solar Energy Research*, vol. 8, no. 2, pp. 1471-1483, 2023. [Google Scholar] [Publisher Link]

[14] Jian-Hua Shu et al., "An Improved Mask R-CNN Model for Multiorgan Segmentation," *Mathematical Problems in Engineering*, vol. 2020, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Kitsuchart Pasupa et al., "Evaluation of Deep Learning Algorithms for Semantic Segmentation of Car Parts," *Complex & Intelligent Systems*, vol. 8, no. 5, pp. 3613-3625, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Xiangyang Xu et al., "Crack Detection and Comparison Study Based on Faster R-CNN and Mask R-CNN," *Sensors*, vol. 22, no. 3, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Panqu Wang et al., "Understanding Convolution for Semantic Segmentation," *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA, pp. 1451-1460, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[18] Alex Dominguez-Sanchez et al., "A New Dataset and Performance Evaluation of a Region-Based CNN for Urban Object Detection," *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, pp. 1-8, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[19] Kaiming He et al., "Mask R-CNN," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961-2969, 2017. [Google Scholar] [Publisher Link]

[20] Nitigya Sambyal et al., "Modified U-Net Architecture for Semantic Segmentation of Diabetic Retinopathy Images," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1094-1109, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Mohammad Hesam Hesamian et al., "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," *Journal of Digital Imaging*, vol. 32, pp. 582-596, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Fahad Lateef, and Yassine Ruichek, "Survey on Semantic Segmentation Using Deep Learning Techniques," *Neurocomputing*, vol. 338, pp. 321-348, 2019. [CrossRef] [Google Scholar] [Publisher Link]