International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 6, December 2025, pp. 5327~5335
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i6.pp5327-5335

Enhancing semantic segmentation with a boundary-sensitive loss function: a novel approach

Ganesh R. Padalkar, Madhuri B. Khambete
Department of Electronics and Telecommunications, MKSSS's Cummins College of Engineering for Women, Pune, India

Article Info

Article history:

Received Sep 12, 2024
Revised Jul 19, 2025
Accepted Sep 14, 2025

Keywords:

Boundary refinement networks
Conditional random fields
Convolutional neural networks
Markov random fields
Pyramid scene parsing network
Regions with convolutional neural networks
Semantic boundary detection dataset

ABSTRACT

Semantic segmentation is crucial step in autonomous driving, medical imaging, and scene understanding. Traditional approaches leveraging manually extracted pixel properties and probabilistic models, have achieved reasonable performance but suffer from limited generalization and the need for expert-driven feature selection. The rise of deep learning architectures has significantly improved segmentation accuracy by enabling automatic feature extraction and capturing intricate object details. However, these methods still face challenges, including the need for large datasets, extensive hyperparameter tuning, and careful loss function selection. This paper proposes a novel boundary-sensitive loss function, which combines region loss and boundary loss, to enhance both region consistency and edge delineation in segmentation tasks. Implemented within a modified SegNet framework, the approach proposed in the paper is evaluated with the semantic boundary dataset (SBD) dataset using standard segmentation metrics. Experimental results indicate improved segmentation accuracy, substantiating to proposed method.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Ganesh R. Padalkar
Department of Electronics and Telecommunications, MKSSS's Cummins College of Engineering for Women
Cummins College Rd, Karvenagar, Pune, Maharashtra 411052, India
Email: ganesh.padalkar@cumminscollege.in

1. INTRODUCTION

Semantic segmentation is a crucial step of giving a label to image pixels from predefined categories [1]. It has numerous applications [2]–[7]. Traditional methods relied on extracted features, combined with probabilistic models conditional random fields (CRFs) [8] and Markov random fields (MRFs) [9]. However, such approaches struggle with generalization, particularly when dealing with variations in lighting, object appearance, and orientation across datasets [10]. Additionally, hand-crafted features are fixed, time-consuming to select, and ineffective for novel or unseen data [11]. While probabilistic models help model spatial dependencies, they face difficulties handling non-linear, high-dimensional relationships, making them inefficient for large-scale, high-resolution images [12]. Due to their local focus, traditional methods often fail to capture the broader image context, leading to suboptimal segmentation, especially for fine details or complex object boundaries [13].

The emergence of deep architectures has shifted performance for obtaining semantic segmentation to higher level by making possible to infer both low- and high-level features automatically [1]. Deep architectures avoid humanly deciding and extracting the features, enhancing scalability across diverse domains [14]. The advancement models [1], [3], [5], [6], [15], [16] has led to improved segmentation precision and efficiency. However, training these models presents challenges such as the need for large datasets, extensive hyperparameter tuning, and high computational resources. Researchers continue to explore ways to address these challenges, including methods that achieve accurate segmentation with limited data [17] and strategies for optimizing training efficiency [18].

Development of semantic segmentation model is possible by training a multi-layer networks [19]. The model learns hierarchical features, with the primary objective of optimizing parameters—such as weights and biases—to generalize well on unseen data [20]. Backpropagation computes gradients based on the loss function, guiding parameter adjustments to minimize prediction errors [10], [21]. The effectiveness of model training depends on multiple factors, including network architecture, data quality, and optimization strategies [20]. Loss functions, in particular, are critical, as they directly influence parameter updates and model performance [22].

Semantic segmentation requires not only accurate region classification but also precise delineation of object boundaries. Clearly defined boundaries are essential for distinguishing adjacent regions and improving object detection accuracy [23], [24]. Additionally, boundaries often exist between similar regions, making them difficult to segment accurately. Addressing these challenges is essential for improving deep learning models' ability to capture both regional and boundary information effectively.

The research on semantic boundary segmentation in view refining object boundaries and segmentation in complex scenarios led to introduce a boundary-aware deep learning model [25] that improved segmentation accuracy, especially at object boundaries. A lightweight network [26] that leverages boundary-aware learning to boost performance on datasets like Cityscapes and ADE20K. A semi-supervised approach that adapts to structured output spaces, improving boundary delineation with limited labeled data [27]. The researchers focused on a multi-scale fusion network [28] and semantic hierarchy-aware model [29], which enhances boundary precision by utilizing hierarchical relationships between object classes.

This research aims to enhance semantic segmentation by introducing a boundary-sensitive loss function, which integrates region loss and boundary loss. Traditional training methods predominantly emphasize region loss, which, while effective for overall segmentation, may overlook precise boundary alignment. Incorporating boundary loss into the training process enables models to refine their predictions, ensuring more accurate segmentation of object edges. This boundary-aware approach enhances performance in complex scenes where boundary details are critical for accurate segmentation [30]–[32]. This research work proposes a novel boundary-sensitive loss function designed to improve both semantic region and boundary segmentation. By jointly optimizing for region classification and boundary alignment, our approach enhances segmentation accuracy and delineation precision. The model evaluation is performed using publically available image dataset, demonstrating that our boundary-aware loss function significantly improves segmentation performance compared to conventional loss functions [33], [34].

The paper is structured as: Section 2 provides an extensive literature. Section 3 outlines the proposed method. Section 4 is dedicated for the results and discussion. Section 5 concludes with summary of contributions and future research directions.

2. COMPREHENSIVE THEOROTICAL BASIS

Semantic segmentation assigns each pixel with a label leaned form the ground truth data, aiming for accurate segmentation. Research in this area directed to enhance segmentation performance. Pixel-based segmentation techniques include thresholding-based methods, region growing, region splitting and merging, watershed transform, graph cuts, and mean-shift [35]–[37]. However, results from pixel-based and block-based methods may not always be sufficient for complex computer vision applications. To address this, semantic segmentation can be achieved through both supervised and semi-supervised methods. The algorithms can be trained on extracted features and corresponding labels to perform effective semantic segmentation [38], [39].

The transition from traditional methods to deep architecture-based semantic segmentation is driven by the superior accuracy and efficiency that they can achieve in simplifying the contents in the image. Deep networks have revolutionized semantic segmentation by enabling accurate pixel-level classification in images [40]. Key architectures in this field include fully convolutional networks (FCNs), and variants like FCN-8, U-Net and DeepLab [6]. SegNet [31], [32] another effective architecture, incorporates Bayesian inference for handling uncertainty, demonstrating high accuracy on the CamVid dataset [7]. Other influential models in the field include PSPNet, Mask R-CNN, HRNet, and EfficientNet, each contributing to advancements in semantic segmentation through deep learning techniques [15].

In deep architectures, feature extraction involves processing input images through multiple layers to extract high-level features. These features propagated through the network to generate the final segmentation map. Standard feature propagation methods may lose critical details, particularly around object boundaries, leading to inaccurate segmentation results. Some approaches involve estimating boundaries as a separate task but introduction of boundary-aware features, enhancing the preservation of edges and improving overall segmentation quality [1], [15]. The authors demonstrate improved performance on several benchmark datasets. Bischke *et al.* [41] proposes an approach, where the network simultaneously learns to perform segmenting the building footprints and predicting the boundary edges. Incorporating boundary information improves assist model to mark closely packed buildings and their boundaries, resulting in better overall performance. The approach was tested on large-scale datasets, showing significant improvements over traditional methods [41].

The convolutional layers are inevitable part of convolutional neural networks (CNNs) [42]. Though the layers extract the information on every stage of the network, these can blur or lose fine details at the edges between different regions or objects, which can lead to inaccuracies in segmenting objects with complex boundaries [42]. Edge-aware convolutions are designed to address this limitation by specifically incorporating information about edges or boundaries during the convolution process [43]. They preserve edge details. Boundary refinement networks (BRNs) are advanced neural network architectures designed to improve the object boundary delineation [44]. Such networks specifically focus on refining the edges and boundaries of segmented objects to produce clearer and more precise segmentation results [44].

Pixel classification accuracy depends on capturing fine details at object edges [45]. This limitation becomes particularly problematic in images with complex scenes, where objects have intricate boundaries, or when different classes are closely packed together. Boundary-based loss functions specifically address this issue by emphasizing the importance of correct boundary prediction [45]. These loss functions are designed to penalize errors near object edges more heavily, thereby encouraging the network to focus on learning precise boundary representations [3]. This approach leads to sharper and more accurate segmentation outputs, which are crucial in applications like medical imaging, where precise boundary identification is essential for tasks like tumor detection or organ segmentation [3]. The method is tested on urban scene datasets and shows superior performance compared to traditional methods [46]. The authors demonstrate usefulness of the technique by making its implementation in medical applications where precise boundary detection is vital [3].

3. PROPOSED METHOD

In Semantic segmentation, boundary pixel segmentation is a critical task. It determines the precise separation between objects and background or neighboring objects. Boundary pixels are difficult to classify accurately due to the similarity of pixel properties in adjacent regions, which makes difficult to learn distinction between object edges and regions. To overcome this problem, boundary pixel segmentation is addressed separately by [47], [48]. Deep architecture models have been developed to classify boundary pixels [49]–[51]. This is challenging problem as the number of boundary pixels are very less compared to object region pixels. In proposed method, boundary sensitive loss function is included. This loss function enhances model's ability to distinguish adjacent objects, improving segmentation performance, especially in complex scenarios.

3.1. Training of deep network with loss function

For semantic segmentation when deep network is used, they are trained using training dataset. Deep network extract features and learns patterns during the training. Based on this learning class of each pixel is predicted. The error between predicted class label and actual class label is expected as minimal as possible. To compute error between predicted labels and actual label loss functions are used. The cross-entropy loss is calculated using (1).

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(p_{i,c}\right) \tag{1}$$

where, N and C are samples and classes in dataset respectively; $y_{i,c}$ is actual label; p_i is the probability given. The SoftMax provides predicted probabilities for each class. The function $J\ (Ypred,\ Y)$ where J is the loss function computes the loss associated with predicted and actual labels. The gardient descent algorithm optimizes the network parameters to get minimum loss. Our method proposes novel loss function which includes boundary loss along with region loss. Details of proposed methods are given.

3.2. Block diagram of proposed method

The flow of the input and updating of network parameters by calculating the loss at different stages is shown in Figure 1. The proposed method uses SegNet network as a backbone. Pretrained SegNet architecture is used to implement proposed method. This architecture is trained using semantic boundary dataset (SBD) which contains region ground truth (R-GT) and boundary ground truth (B-GT) as given Figure 2. The dataset of 1,026 images has 15 classes. In Figure 1, Image I is provided to network, R-GT is region ground truth of I, B-GT is boundary ground truth of the of I, RO is the region segmentation (RO) of the I, B-Map is boundary map of the region segmentation.

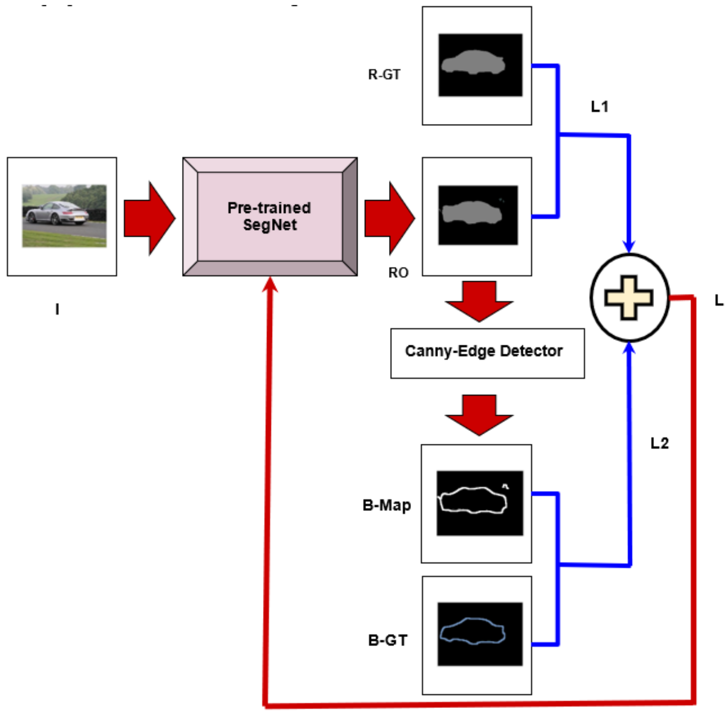

Figure 1. The block diagram

Figure 2. Sample images of R-GT and B-GT from the SBD dataset

3.3. Network training

The pretrained network is trained using SBD dataset [10] which contain labelled ground truth for region (R-GT) as well as boundary (B-GT). Loss between predicted region pixel labels and R-GT labels ($L1$) is calculated using (2). For predicted image map and R-GT having height (H) and width (W) and object classes C:

$$L1 = -\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{c=1}^{C} R-GT^{(i,j,c)}\log\left(RO^{(i,j,c)}\right) \quad (2)$$

here, (i,j) is pixel coordinates and c are object classes.

Edge map of predicted labelled image is obtained using Canny edge detector [52], [53] as shown in Figure 3. The loss between the predicted semantic boundary map [54], [55] (referred to as B-Map) and the ground truth boundary map (B-GT) is calculated using (3). This loss, denoted as $L2$, measures how accurately the model predicts object boundaries. The prediction and ground truth maps both have a height (H), width (W), and represent C object classes. The loss $L2$ is shown in (3):

$$L2 = -\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{c=1}^{C} B - GT^{(i,j,c)} \log\left(B - Map^{(i,j,c)}\right) \quad (3)$$

This equation ensures that the model is penalized more when it fails to accurately predict boundaries of objects in the image. To guide the model in learning both region segmentation and boundary prediction, a total loss (L) is used. This is combination of $L1$ (region loss) and $L2$ (boundary loss) as shown in (4):

$$L=L1+L2 \quad (4)$$

3.4. Training setup and hyperparameter configuration

The SBD [55] is used for training and evaluating the SegNet deep architecture [7]. This dataset, derived from the PASCAL VOC dataset [56], includes detailed boundary annotations for 21 classes, including “background” class. For training the SegNet [7] 1,026 images, for validation 290 images, and for testing 136 images are utilized from the dataset. The proposed method for semantic segmentation was trained on an HP Z4 Tower Workstation with an Intel Xeon W-2245 processor, 32 GB DDR4 RAM, an NVIDIA RTX A5000 GPU, and a 1 TB HDD, using MATLAB R2014a. For hyperparameter tuning, the SegNet [7] model is configured with starting learning rate of 0.01, piecewise learning rate schedule with a drop period of 10 epochs and a drop factor of 0.1, batch size of 1 (stochastic gradient descent), a total of 50 epochs, momentum of 0.9, and validation frequency of 10 epochs.

3.5. Evaluation metrics

This work evaluates the contribution of proposed loss function for both semantic region segmentation and boundary detection tasks. To ensure a comprehensive assessment, several widely recognized metrics are employed, including global accuracy [56], mean accuracy [56], and (mean IoU) [57]. Additionally, weighted intersection over union (weighted IoU) [58] is incorporated to assign higher importance to underrepresented classes, ensuring fair evaluation across different object categories. The mean boundary F-Score (mean BFScore) [59] is particularly crucial for assessing boundary accuracy, as it quantifies how well the predicted segmentation aligns with object boundaries, capturing fine details often overlooked by standard region-based metrics. These metrics collectively provide a holistic evaluation.

4. RESULTS AND DISCUSSION

The performance of integrated loss function with region loss and the proposed boundary-sensitive loss function is shown in Table 1. The table highlights accuracy, IoU, and MeanBFScore, for both methods. Table 2 provides an overall summary of how the segmentation model performs when using only the region loss compared to the proposed boundary-sensitive loss. From the results it is evident that with a single deep network along with a boundary-sensitive loss, improves semantic segmentation overall. By focusing on object boundaries, it makes the segmentation more accurate. This approach also performs better than several other methods that rely on multiple networks, making it both effective and efficient.

Table 1. Comparison of semantic segmentation using only region loss and proposed boundary-sensitive loss

Sr. No	Object class	Accuracy		IoU		MeanBFscore	
		Only region loss	Proposed boundary-sensitive Loss	Only region loss	Proposed boundary-sensitive loss	Only region loss	Proposed boundary-sensitive loss
1.	Background	0.977	0.996	0.608	0.621	0.472	0.479
2.	Aeroplane	0.189	0.667	0.159	0.548	0.210	0.471
3.	Bicycle	0.239	0.279	0.223	0.267	0.034	0.039
4.	Bird	0.163	0.428	0.045	0.164	0.352	0.519
5.	Bus	0.147	0.184	0.131	0.136	0.155	0.146
6.	Car	0.516	0.160	0.401	0.139	0.176	0.116
7.	Cat	0.006	0.246	0.005	0.228	0.047	0.132
8.	Chair	0.314	0.143	0.193	0.084	0.277	0.276
9.	Cow	0.036	0.002	0.008	0.001	0.184	NaN
10.	Dog	0.052	0.040	0.049	0.040	0.001	0.046
11.	Horse	0.011	0.026	0.010	0.025	0.076	0.100
12.	Motorbike	0.029	0.112	0.022	0.097	0.087	0.096
13.	Person	0.259	0.149	0.222	0.145	0.264	0.132
14.	Sofa	0.105	0.064	0.101	0.062	0.080	0.043
15.	TV monitor	0.082	0.057	0.079	0.055	0.044	0.007
16.	Average values	0.208	0.237	0.150	0.174	0.164	0.186

Table 2. Comparison of semantic segmentation with region loss and proposed boundary-sensitive loss

Sr. No	Metric	Only region loss	Proposed boundary-sensitive loss
1.	Global accuracy	0.5979	0.6090
2.	Mean accuracy	0.1524	0.1745
3.	Mean IoU	**0.1107**	0.1286
4.	Weighted IoU	0.3817	0.3908
5.	Mean BF score	**0.2907**	0.3362

5. CONCLUSION

The experiment demonstrates the significant impact of incorporating boundary loss ($L2$) alongside region loss ($L1$) in semantic segmentation. When the model is trained solely with region loss, it performs well in segmenting large and frequent objects but struggles with small objects, leading to poor boundary delineation and lower IoU and boundary accuracy. In contrast, when both region and boundary losses are included, the model achieves better segmentation performance, particularly in preserving object boundaries. The improvements are most evident for objects with well-defined edges, as reflected in higher IoU and boundary F-scores. However, the segmentation of smaller objects remains challenging, likely due to class imbalance and the complexity of detecting fine details.

The results align with recent research findings, reinforcing the idea that boundary-aware models enhance segmentation accuracy. This study demonstrates that balancing region and boundary losses contributes to more detailed and precise segmentation, making it particularly useful in advanced applications, where boundary precision is critical. The proposed approach introduces a novel loss function that improves segmentation accuracy. Future research should focus on addressing the limitations observed in small object segmentation by exploring adaptive weighting strategies to mitigate class imbalance. Expanding the model to handle multi-class segmentation in more complex scenes is another avenue for improvement.

ACKNOWLEDGMENTS

The authors would like to acknowledge MKSSS's Cummins College of Engineering for Women, Pune, for providing laboratory facilities equipped with GPU resources and for supporting this research work through institutional affiliation.

FUNDING INFORMATION

The authors state that no funding was involved in supporting this research work.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ganesh R. Padalkar	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Madhuri B. Khambete	✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to animal use has been complied with all the relevant national regulations and institutional policies for the care and use of animals.

DATA AVAILABILITY

The data supporting the findings of this study are based on the Semantic Boundaries Dataset (SBD), which was accessed by the authors when the original repository was active. Currently, the dataset is available through the following sources:

- Papers with code SBD page:
  https://paperswithcode.com/dataset/sbd
- PyTorch TorchVision Library:
  https://docs.pytorch.org/vision/main/generated/torchvision.datasets.SBDataset.html

REFERENCES

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3431–3440.

[2] I. Papadeas, L. Tsochatzidis, A. Amanatiadis, and I. Pratikakis, "Real-time semantic image segmentation with deep learning for autonomous driving: a survey," *Applied Sciences*, vol. 11, no. 19, p. 8802, Sep. 2021, doi: 10.3390/app11198802.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.

[4] N. S. D'Souza *et al.*, "Deep sr-DDL: Deep structurally regularized dynamic dictionary learning to integrate multimodal and dynamic functional connectomics data for multidimensional clinical characterizations," *NeuroImage*, vol. 241, p. 118388, Nov. 2021, doi: 10.1016/j.neuroimage.2021.118388.

[5] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *arXiv preprint arXiv:1412.7062*, pp. 778–790, 2014.

[6] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

[7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[8] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, and T. Darrell, "Hidden conditional random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1848–1852, Oct. 2007, doi: 10.1109/TPAMI.2007.1124.

[9] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, pp. 105–112, doi: 10.1109/ICCV.2001.937505.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[11] H. R. Roth, L. Lu, C. L. Farago, and E. S. L. H. G. (H). Li, "Deep learning for medical image analysis: a comprehensive review," *Medical Image Analysis*, vol. 54, pp. 215–234, 2019.

[12] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4, no. 4. Springer, 2006.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[15] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[17] B. Xie, S. Li, M. Li, C. H. Liu, G. Huang, and G. Wang, "SePiCo: semantic-guided pixel contrast for domain adaptive semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 9004–9021, Jul. 2023, doi: 10.1109/TPAMI.2023.3237740.

[18] X. Cao, H. Zou, J. Li, X. Ying, and S. He, "OBBInst: remote sensing instance segmentation with oriented bounding box supervision," *International Journal of Applied Earth Observation and Geoinformation*, vol. 128, p. 103717, Apr. 2024, doi: 10.1016/j.jag.2024.103717.

[19] L. Bottou, "Stochastic gradient descent tricks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7700 LECTU, pp. 421–436, 2012, doi: 10.1007/978-3-642-35289-8_25.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[21] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, Nov. 2012, doi: 10.1109/MSP.2012.2205597.

[22] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *3rd International Conference on Learning Representations (ICLR)*, 2015.

[23] A. Arnab, S. Jayasumana, S. Zheng, and P. H. S. Torr, "Higher order conditional random fields in deep neural networks," in *European conference on computer vision*, 2016, pp. 524–540.

[24] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *European conference on computer vision*, 2020, pp. 173–190.

[25] H. Wu, Y. Li, L. Chen, X. Liu, and P. Li, "Deep boundary-aware semantic image segmentation," *Computer Animation and*

Virtual Worlds, vol. 32, no. 3–4, 2021, doi: 10.1002/cav.2023.
[26] Q. Liu, H. Su, M. El-Khamy, and K.-B. Song, "DeepGBASS: deep guided boundary-aware semantic segmentation," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pp. 1525–1529, doi: 10.1109/ICASSP43922.2022.9747892.
[27] W. Huang and F. Zhang, "Semi-supervised semantic segmentation with structured output space adaption," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10095628.
[28] T. Jiang, Y. Jin, T. Liang, X. Wang, and Y. Li, "Boundary corrected multi-scale fusion network for real-time semantic segmentation," *Proceedings - International Conference on Image Processing, ICIP*, pp. 1886–1890, 2022, doi: 10.1109/ICIP46576.2022.9897907.
[29] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "Semantic hierarchy-aware segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 10, pp. 11119–11135, Apr. 2024, doi: 10.1109/TPAMI.2023.3332435.
[30] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. Ben Ayed, "Boundary loss for highly unbalanced segmentation," in *International conference on medical imaging with deep learning*, 2019, pp. 285–296.
[31] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with Atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
[32] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
[33] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: gated shape CNNs for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 5229–5238.
[34] Z. Chen, A. V Oikonomou, and P. Fieguth, "Segmentation-aware convolutional networks using local attention masks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 29, pp. 5038–5047.
[35] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, May 2002, doi: 10.1109/34.1000236.
[36] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001, doi: 10.1109/34.969114.
[37] J. McQueen, "Some methods for classification and analysis of multivariate observations," in *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.
[38] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[39] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.
[40] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. G. Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1205–1220, 2020.
[41] B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel, "Multi-task learning for segmentation of building footprints with deep neural networks," in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 1480–1484, doi: 10.1109/ICIP.2019.8803050.
[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1–14.
[43] L. Yan, J. Huang, H. Xie, P. Wei, and Z. Gao, "Efficient depth fusion transformer for aerial image semantic segmentation," *Remote Sensing*, vol. 14, no. 5, p. 1294, Mar. 2022, doi: 10.3390/rs14051294.
[44] X. Zhang, Y. Wei, J. Feng, Y. Yang, and T. S. Huang, "Adversarial complementary learning for weakly supervised object localization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2018, pp. 1335–1344, doi: 10.1109/CVPR.2018.00145.
[45] T. Chen, L. Lin, X. Wu, N. Xiao, and X. Luo, "Learning to segment object candidates via recursive neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 5827–5839, Dec. 2018, doi: 10.1109/TIP.2018.2859025.
[46] M. Cordts *et al.*, "The cityscapes dataset for semantic urban scene understanding," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 3213–3223, doi: 10.1109/CVPR.2016.350.
[47] T. Lu, D. Ming, X. Lin, Z. Hong, X. Bai, and J. Fang, "Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network," *Remote Sensing*, vol. 10, no. 9, p. 1496, Sep. 2018, doi: 10.3390/rs10091496.
[48] Y. Lu, R. Jiao, S. Dai, and H. Zhang, "Boundary-aware feature propagation for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Oct. 2022, pp. 1240–1249.
[49] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 636–644, doi: 10.1109/CVPR.2017.75.
[50] Y. Wang, Z. Jin, S. Zhao, and X. Chen, "359 Boundary-aware convolutional neural network for skin lesion segmentation in clinical images," *Journal of Investigative Dermatology*, vol. 141, no. 5, p. S63, May 2021, doi: 10.1016/j.jid.2021.02.381.
[51] X. Huang, Y. Li, and C. C. Loy, "Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6834–6842, 2022.
[52] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.
[53] R. Gonzalez and R. Woods, *Digital image processing*, 3rd ed. New Jersey: Prentice-Hall, 2008.
[54] T. Pavlidis, "A thinning algorithm for discrete binary images," *Computer Graphics and Image Processing*, vol. 13, no. 2, pp. 142–157, Jan. 1980, doi: 10.1016/S0146-664X(80)80037-2.
[55] A. Bozorgpour, Y. Sadegheih, A. Kazerouni, R. Azad, and D. Merhof, "DermoSegDiff: A boundary-aware segmentation diffusion model for skin lesion delineation," 2023, pp. 146–158, doi: 10.1007/978-3-031-46005-0_13.
[56] M. Everingham *et al.*, "The PASCAL visual object classes (VOC) challenge," *International Journal of Computer Vision manuscript*, vol. 88, no. 2, pp. 303–338, 2012.
[57] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2117–2125.
[58] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," 2017, pp. 379–387, doi: 10.1007/978-3-319-67389-9_44.
[59] S. Xie and Z. Tu, "Holistically-nested edge detection," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 1395–1403, doi: 10.1109/ICCV.2015.164.

BIOGRAPHIES OF AUTHORS

Ganesh R. Padalkar received his M.E. degree from Shivaji University, Kolhapur, Maharashtra, India in 2009. He is working as assistant professor at MKSSSS' Cummins College of Engineering for Women, an autonomous institute affiliated with Savitribai Phule Pune University, Pune, Maharashtra, India. He is also a research candidate at Savitribai Phule Pune University, Pune. Ganesh R. Padalkar is the author of 6 technical publications. His research interests include image processing and deep learning. He is a life member of the Indian Society for Technical Education. He can be contacted at email: ganesh.padalkar@cumminscollege.in.

Madhuri B. Khambete received her PhD degree in 2009 in the area of image processing from the Savitribai Phule Pune University, Pune, Maharashtra, India. She is serving as a director with the MKSSS's Cummins College of Engineering for Women, Karvenagar, Pune. She also is a PhD guide in the same research center. Dr Madhuri B. Khambete is the author of over 24 technical publications. Her research interests include image processing, computer vision and pattern recognition. She can be contacted at email: principal@cumminscollege.in.