IJICIS, Vol.25, No.2, 1-17
DOI: 10.21608/ijicis.2025.372514.1383

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/

ADVANCEMENTS IN SEMANTIC SEGMENTATION USING DEEP LEARNING TECHNIQUES FOR IMAGE ANALYSIS

Mohanad A. Al-askari

Biomedical Engineering Research Centre,
University of Anbar,
Iraq
mohanad.abdul@uoanbar.edu.iq

Received 2025-04-02; Revised 2025-04-10; Accepted 2025-06-30

Abstract: *In this study, the progression in semantic segmentation has been explored using deep learning architectures such as SegNet, FCN-AlexNet, and U-Net with EfficientNet-B3 backbone. It assesses their performance on a range of datasets including UAV imagery, Cityscapes and ADE20K, as well as comparing to the accuracy of U-Net (89.7% MPA) and generalization. Problems such as computational sophistication, class inequality, and real-time processing limitations are investigated, highlighting trade-offs between acceleration and exactness. While identifying gaps for domain adaptation, and adversarial robustness, the paper discusses optimization strategies as attention mechanisms or self-supervised learning. Practical deployment would follow in future directions of the deployment with lightweight models, multimodal fusion and explainable AI. The results emphasized that for segmentation tasks, encoder−decoder designs are beneficial for their utility in autonomous vehicles as well as for medical imaging.*

Keywords: *Semantic Segmentation, Deep Learning, Image Analysis, U-Net, SegNet, FCN-AlexNet, Accuracy, Generalization*

1. Introduction

Semantic segmentation is one of the important tasks in Computer Vision where each pixel of an image is labelled to accurately segment an object. It is widely applied in many fields, such as medical imaging, autonomous automobiles, remote sensing, and so on. There are many types of segmentation, thresholding, region growth, and clustering, but these are rigid and cannot handle changes in an image. The use of deep learning, especially CNNs and Transformer models, has pioneered the increase in segmentation performances and feature extraction. However, issues like high computational load, lack of an equal number of instances for different classes, and large amounts of annotated datasets remain open. Recent development in the optimization method incorporates model compression, attention mechanisms, and self-supervised learning towards efficiency. Moreover, fusion architectures that

***Corresponding Author**: Mohanad A. Al-askari

Biomedical Engineering Research Centre, University of Anbar, Iraq

Email address: mohanad.abdul@uoanbar.edu.iq

combine CNNs with Transformers are currently on the rise, given that they prevent decomposition into local and global features.
Semantic segmentation also requires each pixel within the image to be classified making the identification of objects accurate concerning the context of its uses in medical imaging, self-driving cars, and remote sensing technologies. Some of the disadvantages that were common with traditional methods included occlusion, different sizes of objects, and complex backgrounds. Deep learning in turn, has contributed greatly to the enhancement of segmentation by relying on feature extraction of different levels and huge datasets. However, some challenges, such as the class imbalance, the computationally intensive process of the classifier, and the generalization of the domain, remain open [1]. These need to be addressed by enhancing network topologies and subsequently enhancing the real-time processing capabilities towards practical use.
The most recent articles are mainly directed towards making improvements on the issue of segmentation for applications with low latency. Some of the challenges include problems such as domain generalization, robustness against adversarial attacks, and computational considerations. These are some of the trends that are expected to occur to improve the DL-based segmentation techniques and make the technique more applicable in the real-world practice of image analysis.
The paper begins with an Introduction that describes the significance of semantic segmentation in image analysis and the emergence of deep learning in this area. It is followed by Problem Definition and Literature Review, which determine the prevailing issues and review past research. Then, the Methodological Formulation and Theoretical Development part lists the suggested model, algorithms, and architectural decisions. Experimental Results provide the description of datasets, implementation and results. Comparison with Related Works checks whether the proposed approach will bring any advances as compared to previous standards. A Conclusion is also given after the findings and implications were outlined, covering the future directions to improve deep learning-based semantic segmentation.

2. Problem Definition and Literature Review

2.1. Traditional Approaches to Image Segmentation

The conventional methods for image segmentation involved thresholding, edge detection, region growing, and clustering-based studies such as k-means, and the methods are based on predefined handcrafted characteristics. Although these approaches were computationally faster, they failed to handle some robotic problems like changes in lighting conditions and overlapping of objects. Approaches that also advanced the segmentations included the use of Machine learning techniques like the Support Vector Machines (SVMs) and the Random Forests which, however, demanded an input of additional features. Although having shown a high level of performance in controlled conditions, these methods proved to be insensitive to various datasets and, more often, everyday applications [2]. The drawbacks of the conventional approaches led to the implementation of deep learning-based segmentation that offers a more elaborate representation of features with the ability to learn from large data.

2.2. Deep Learning-Based Semantic Segmentation

Semantic segmentation was a boon in the application of deep learning due to its ability to learn features and hierarchical representation. Due to the algorithms such as FCN, U-Net, and DeepLab, the accuracy rate of segmentation was greatly enhanced by the commencement of CNNs.

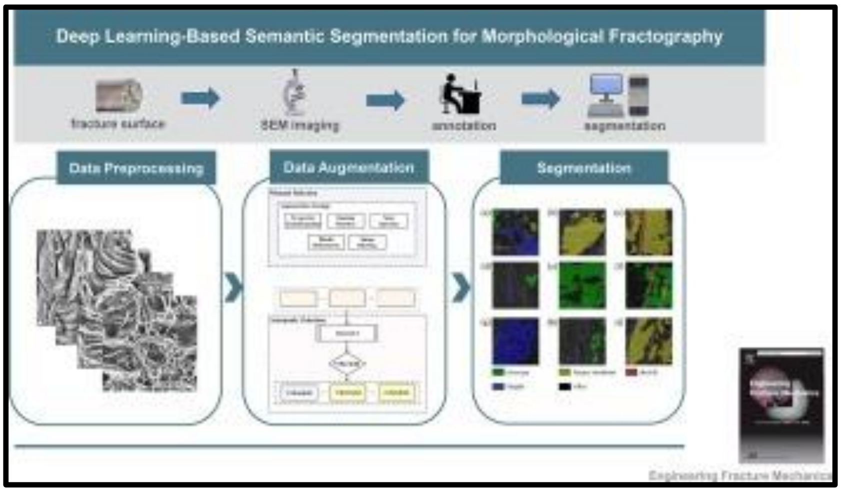

Figure 1: Deep learning-based semantic segmentation for morphological fractography [9]

Figure 1 represents the pipeline of Deep Learning-Based Semantic Segmentation of Morphological Fractography. It starts by acquiring the fracture surfaces and subsequent SEM imaging, after which the manual annotation is conducted. There are three fundamental steps: Data Preprocessing, when the SEM images are prepared, and Data Augmentation is used to add synthetic changes to generated classes and balance the classes [9]. Lastly, Segmentation carries out the pixel-level classification to emphasise various morphologies of fractures. The workflow illustrates how deep learning method allows performing highly accurate analysis of the images in the study of material failures. It incorporates sophisticated methods by attaining effective and precise semantic mapping of fractographic surface structures.

These models utilised encoder-decoder architecture, skip connections, and multi-scale feature extraction techniques. Transformer-based models further refined contextual understanding and long-range dependencies. Deep learning is more generalized and far more robust as compared to the conventional machine learning technique across different datasets [3]. However, some issues, including computational difficulty and real-time execution issues remained research issues.

2.3 Key Challenges in Deep Learning-Based Segmentation

This works with deep learning-based semantic segmentation has some weaknesses even though it has been developed a lot. Key issues include:

- *Real-time Applications* – Deep models were real-time intensive, implying that implementing the techniques in real time on devices like smartphones could be problematic.
- *Unbalanced Classes* – The classes produced in the learning process with fewer training samples end up being misclassified, hence impacting segmentation differences.
- *Long-tailed Learning* – Large datasets were required for training labelled data, which is costly and takes a lot of time in the annotation process.
- *Domain Generalization* – It is noted that models developed through certain datasets fail to perform well in changes in lighting, occlusion, and changes in the environment [4].

- *Real-Time Processing Constraints* – Some applications like self-driving cars called for a real-time algorithm that is efficient in terms of time, hence segmentation.

Interpretability and Explainability – Deep learning models acted as a prism and hence cannot be easily comprehensible on issues of decision-making.
These were the activities that require the best solution of the network architectures, the introduction of attention mechanisms and the usage of the mixed models to improve the network efficiency and its ability to generalize.

2.4 Applications of Semantic Segmentation in Image Analysis

Semantic segmentation was used in almost all fields, and it has been key when it comes to object identification by mapping the image space. In medical imaging, it was used in the detection of tumours and in segmenting organs. Hence, segmentation was critical when it comes to self-driving, where partitions are made on the road scene, lanes, pedestrians, and obstacles [21].

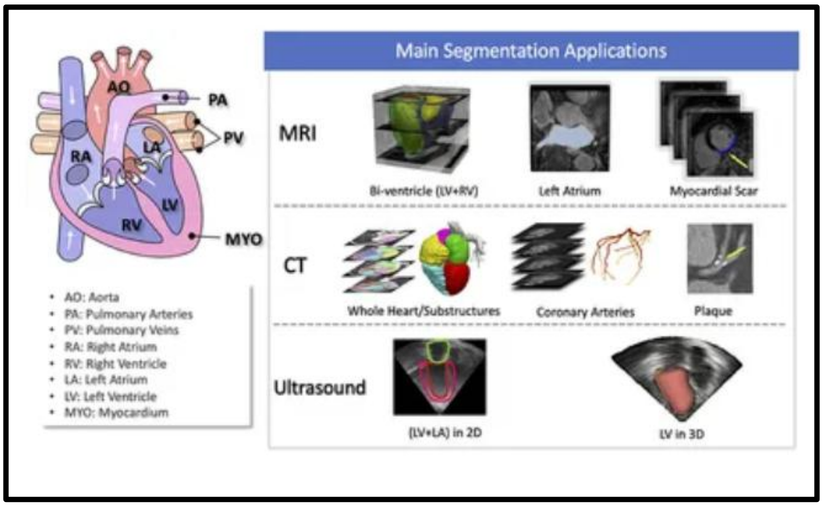

Figure 2: Semantic Segmentation Applications in Cardiac Imaging [26]

Figure 2 exemplifies the major areas where semantic segmentation has been used in cardiac imaging regions with three modalities, namely, MRI, CT, and Ultrasound. It allows bi-ventricles (LV+RV) segmentation, left atrium and myocardial scars in MRI. CT scans allow segmenting the entire heart and its component parts, coronary vessels and plaques [26]. Under ultrasound, the left ventricles and atrium are being deemed segmented in two dimensions and the left ventricle in three dimensions. This segmentation is applicable to the accurate identification of the cardiac structures, which aids in the clinical diagnosis and treatment planning.
Remote sensing applied to classifying the lands and monitoring the environment. In the case of agriculture, it was used in the identification of crops and diseases which helps in enhancing precision farming [5]. Also, segmentation was of great use in furthering industrial automation in terms of detecting defects and quality control.

2.5 Recent Advancements and Optimization Techniques

The recent developments in deep learning for semantic segmentation can be witnessed by the increased accuracy, effectiveness, and the possibility of generalization. Self-attention and the Transformer learned about the contexts in each context by performing feature extraction of dependencies. Introducing the element of Transformers into the hybrid architectures with CNNs allowed for intermediate and global feature learning to complement the local features [22].

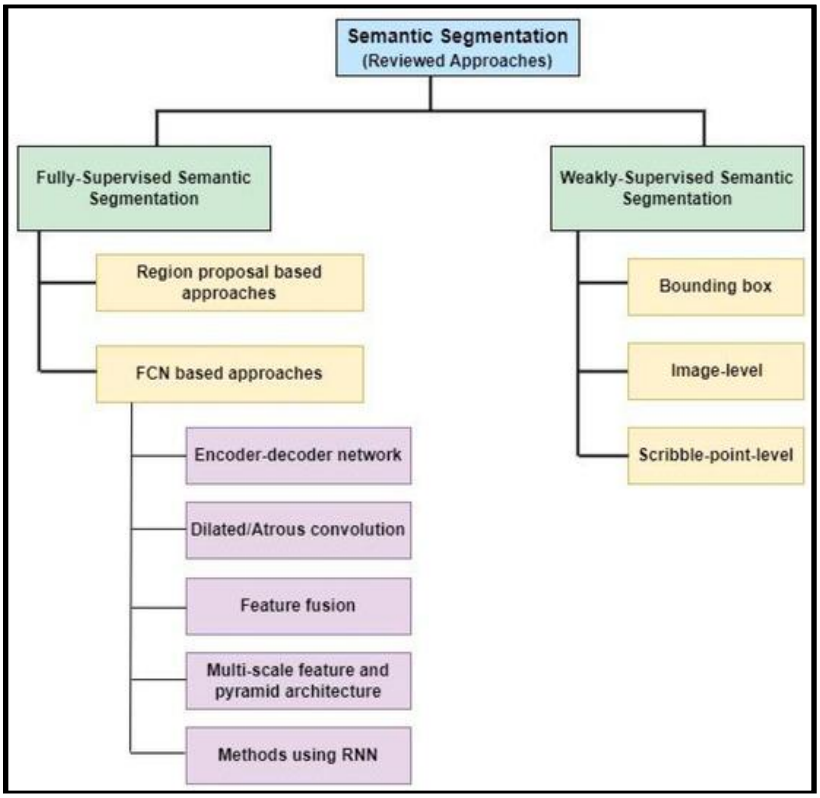

Fig. 3. Semantic segmentation using deep learning algorithms [15]

Figure 3 presents the taxonomy of deep learning semantic segmentation methods. It classifies techniques as Fully-Supervised and Weakly-Supervised Semantic Segmentation. Fully-supervised techniques can be region proposal, and FCN-based methods, which are augmented with an encoder-decoder architecture, atrous convolution, feature fusion, pyramid structure, and RNN. Weakly-supervised methods commonly use coarser annotations like bounding box, image or scribble-point-level information [15]. The classification emphasises the variety in the strength of the supervision and architectural innovations in the semantic area by implementing deep learning algorithms.

In order to improve computational cost, techniques such as pruning, quantization, and knowledge distillation kept the accuracy high when reducing computational time. Few-shot and self-supervised learning thereby reduced the above problem by learning from limited annotated examples. Besides, meta-learning and reinforcement learning refined the dynamics of network architecture in task-adaptive manners [6]. In addition, edge and cloud computing integration provided real-time segmentation for applications like self-driving cars and healthcare. Different generalization techniques involving adversarial training had been developed and applied widely in various datasets. These advancements went on further to extend the limits of semantic segmentation for more secure and large-scale practical use.

2.6 Comparative Analysis of Existing Methods

There were various types of semantic segmentation techniques, and their characteristics include accuracy, computational complexity, and adaptability to a particular dataset. Pre-existing machine learning techniques were incapable of handling intricate architectural patterns in the image; on the other hand, FCNs, U-net, and DeepLab showed better performance and segmentations [23]. Other models that relied on context enhance the generalization but come out to be more computationally heavy. Evaluation parameters included the Intersection over Union, which calculates the ratio between the area of intersection of the output and the ground truth to the area of their union, the Dice coefficient, which measures the ratio of the overlap between two graphical sets, and the inference time [24]. Several studies were analysed to investigate a comparison of the accuracy/precision of estimation techniques based on trade-offs between accuracy and time taken [7].

Table 1: Comparison of Existing Semantic Segmentation Methods

Method	**IoU (%)**	**Dice Coefficient**	**Accuracy (%)**	**Inference Time**	**Benchmark Datasets**	**Strengths**	**Limitations**
Traditional ML	50-65	Low	60-75	Fast	PASCAL VOC	Simple, fast, and works for basic tasks	Poor generalization, low accuracy
FCN	70-85	Medium	80-90	Moderate	COCO, Cityscapes	Fully convolutional, end-to-end training	Lacks spatial precision
U-Net	80-90	High	85-95	Moderate	Medical datasets	Excellent for biomedical segmentation	High memory usage
DeepLab V3+	85-95	Very High	90-97	Slow	ADE20K	Multi-scale feature extraction	Computationally expensive
Transformer-based	88-98	Very High	92-98	Slow	COCO, Cityscapes	Captures long-range dependencies	High computational requirements

2.7 Future Directions in Semantic Segmentation

The development trend of semantic segmentation in the future was to enhance the accuracy and efficiency in various applications. Self-supervision and few-shot learning approaches were promising for eliminating the large annotated dataset reliance. New architectural solutions that were lightweight and specifically designed for edge computing will allow carrying out the computations in a real-time manner at the place of origin of the data. It was believed that the combination of LiDAR, radar, and RGB data was going to boost segmentation in self-driving technology [25]. There was a great anticipation that the Explanatory AI or XAI will significantly contribute to creating models that explain most of the deep learning architectures' operations. Also, federated learning delivered better model learning while maintaining the privacy of medical and industrial data. Among the prospective algorithms, quantum computing and neuromorphic hardware were considered promising solutions for improving deep learning-based segmentation [8]. Superiority in domain adaptation robustness to adversarial attacks and efficiency defined the future generation of segmentation models to make it fit for deployment.

2.8 Research Gap

In semantic segmentation, several problems still required solutions even with the adoption of deep learning approaches. First, it was acknowledged that training with significantly large annotated data hinders the model's ability to generalize across various real-life settings. Self-supervised learning and synthetic data generation appear to be promising approaches, though further investigation needed to be done to improve the mentioned approaches [4]. Second, some of today's advocated architectures have difficulties with real-time segmentation, especially in the case of limited computational capabilities. This is an area of research that needed great attention, and optimized architectures and hardware accelerators are the solution to this problem [7]. Third, the problem of making AI models resistant to domain shifts and adversarial manipulations was still largely open since the current situation does not cater well to distribution shifts [8]. This was an essential advancement to address these gaps by improving learning paradigms, architectures of the models, and generalization techniques throughout the progression of semantic segmentation in practice.

2.9 Summary

Semantic segmentation was transitioned from being classified as a machine learning problem to being deep learning, which enhances the accuracy and feature extraction. Thus, new challenges were emerged along with it including computational complexity, class imbalance, and dependency on data. In recent years, improvements ranging from the attention mechanisms, the combined models, and the optimization methods improved segmentation [9]. Therefore, there were still open issues in runtime processing, transfer to different domains, and model explainability. The developments in the future will target the working system, social freeway, learned information, and independent learning to improve the speed and utility of different real-world tasks.

3. Methodological Formulation and Theoretical Development

3.1. Data Description

The secondary data collection used for this study primarily revolves around publicly accessible datasets containing 'ADE20K', 'COCO-Stuff', 'Pascal VOC', 'Cityscapes', and 'NYU Depth V2' [10]. The annotated images for these datasets are known in computer vision, so they are very well established but also popular for semantic segmentation. The secondary data contains 2D or two-dimension, 2.5D (RGBD), and three-dimension (3D) images across different indoor, and outdoor environment scenes, urban streets, and aerial views [11]. The datasets selected have annotations ranging from pixel-wise labels to bounding boxes and instance masks to use for semantic segmentation. Furthermore, some datasets contain also multimodal data, like depth information or infrared, so the application becomes wider.

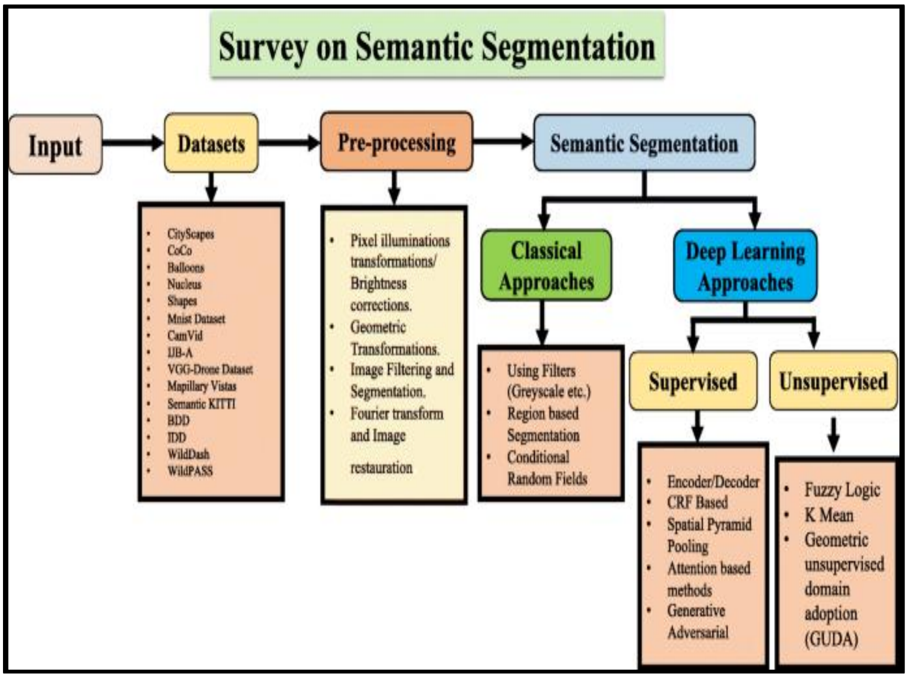

Fig. 4. Semantic segmentation steps with deep learning techniques [19]

Figure 4 illustrates an outline pipeline that represents semantic segmentation utilising deep learning. It starts with input datum and then selects a dataset, either CityScapes or KITTI. Transformations and restoration of the images involved in Pre-processing. This segmentation process is divided into classical and deep learning. Compared to classical approaches, which are filter and region-based segregations, deep learning encompasses supervised and unsupervised processing [19]. Supervised models include encoders, Spatial Pyramid Pooling, and Generative Adversarial, whereas the unsupervised ones apply fuzzy logic and clustering. This organised process contributes to accuracy and contextual knowledge of the analysis of images.

Qualitative and quantitative data are both used for research analysis. Qualitative data includes annotations, labels, descriptions of classes, and categorisations of scenes; and quantitive data is based on the pixel counts, the resolution details and the class distribution. Notably, the datasets are all at different scales and resolutions, and at different levels of annotation depth, all to varying degrees, to guarantee full representation of semantic segmentation challenges. Specialized applications like remote sensing are included by including aerial and satellite imagery datasets including ISPRS Vaihingen and Potsdam [12]. Academic repositories and benchmark platforms provide publicly available datasets, namely ADE20K, COCO-Stuff, and Pascal VOC. Remote sensing databases are used to obtain aerial and satellite imagery on (ISPRS Vaihingen, Potsdam). UAV images (DJI Phantom 4 Pro, Sony QX100) are taken and processed by photogrammetry software (Agisoft Photoscan) [13]. The sensor-captured indoor scenes are acquired using RGB-D datasets (NYU-Depth V2, SUN RGB-D). In order to verify the manual annotations, crowdsourcing (Amazon Mechanical Turk) and expert labelling are used. Training samples are augmented using data augmentation techniques such as rotation, scaling, and colour shifts.

3.2. Deep Learning Architectures for Semantic Segmentation

The study employs three deep learning architectures for semantic segmentation: SegNet, FCN-AlexNet, and U-Net with an EfficientNet-B3 backbone.

SegNet is based on the asymmetric encoder-decoder structure. The encoder is made of 13 convolutional layers from VGG16 and max-pooling with indices to pixel-wise upsample the decoder [14]. Softmax is a pixel-wise classification applied in the final layer.
The traditional AlexNet is modified to replace the fully connected layers by 1×1 convolutions and has a 63×63 upsampling layer, which enables end-to-end segmentation while keeping the spatial resolution.
In this work, U-Net with EfficientNet-B3 consists of a pre-trained EfficientNet-B3 encoder and U-Net decoder. In this framework, the encoder fits multilevel features, while the decoder has skip connections and convolutions with transposed operations for in-tense localization. This architecture is depth, width and resolution balanced for better segmentation accuracy [15].
Each model utilises the Adam optimizer (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and categorical cross-entropy loss:

$$L = -\sum_{c=1}^{C} y_c \, log\,(p_c)$$

Where,
Yc is the true label, Pc is the predicted probability for class c, and C is the total classes.
Training parameters include a batch size of 24, 50 epochs, and weight decay (0.05).
Hardware: GPU-accelerated training (e.g., NVIDIA GeForce RTX 2070).
Software: GIMP for annotation, TensorFlow/Keras for implementation.
Dataset: UAV images (split into 480×480 tiles: 2082 training, 694 validation, 709 test).

Table 2: Model Architectures and Key Features

Models	Encoder	Decoder	Key Features
SegNet	VGG16 (13 conv layers)	Max-pooling indices upsampling	Memory-efficient, precise localization
FCN-AlexNet	AlexNet-based	1×1 conv + 63×63 upsampling	Preserves spatial resolution
U-Net (EfficientNet-B3)	EfficientNet-B3	Skip connections + transposed conv	Balanced scaling, high accuracy

In the table, three deep learning models for semantic segmentation: SegNet, FCN-AlexNet and U-Net (EfficientNet-B3) are compared in terms of their encoder decoder structures and key features. The VGG16 is used in SegNet with index-based upsampling, FCN-AlexNet with 1x1 convolutions, and U-Net joins EfficientNetB3 with skip connections for balanced accuracy and resolution maintenance [16]. Encoder decoder frameworks are leveraged to accurately segment pixels and then robust feature extraction is accomplished by this methodology of the system.

3.3. Data Preprocessing and Augmentation Strategies

In order to enhance performance on UAV imagery, the methodology performs standardized preprocessing and augmentation. In order to avoid wasting GPU memory, input images are made 480 × 480 pixels but keep their spatial resolution. GIMP is also used to manually label masks of 5 classes: rice paddy, rice lodging, road, ridge and background [17]. In binary segmentation tasks, rice lodging regions are discriminated as white pixels. Pixel values are scaled to [0,1] using:

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}}$$

Where, I is the input image, and Imin, and Imax are minimum and maximum pixel intensities.

Table 3: Tools used for data preprocessing and augmentation

Tools	Description
Labeling Tool	GIMP (GNU Image Manipulation Program) for pixel-wise annotation.
Preprocessing Libs	OpenCV, and NumPy for resizing, normalization, and vegetation index computation.
Augmentation	TensorFlow/Keras for rotation, flipping, and spline-based interpolation.
Hardware	GPU (e.g., GeForce RTX) for model training and inference.

There are 3,485 samples on the image level using image tiles and an 80-20 train test split. The entire training set has been split into 75% training and 25% validation subsets. A part of data augmentation is the geometric transformation including the rotation, flipping, and mirroring to improve the generalization. For inference, the patch is interpolated by a second-order spline window function of interpolated patches augmented patches [18]. The RGB channels are used to calculate vegetation indices like ExG, ExR, and ExGR to enhance the feature extraction. Pixel values are normalized in the preprocessing pipeline to [0, 1] and for multiclass segmentation, they are also one-hot encoded.

3.4. Loss Functions and Optimization Techniques

Categorical cross-entropy loss on the multi-class semantic segmentation is used as the methodology, determining divergence between its predicted class probabilities and their ground truth labels. The loss function ensures good optimization by penalizing incorrect classifications based proportionally on the level of the prediction confidence. Dice loss is used for binary segmentation tasks that deal with class imbalance by emphasizing overlap between considered true regions and predicted regions [19].
Adam optimizer is trained which uses momentum and gradient scaling to adapt the learning rate per dimension. The initial learning rate is set to be fixed at 0.001, and exponential decay rates on first and second-moment estimates are 0.9 and 0.999 respectively. It provides regular terms to prevent overfitting. Early stopping monitors such that training is stopped as soon as validation loss starts to plateau, and learning rate scheduling helps convergence further [20]. Thus, the combined use of these techniques yields stable training and well-separated segments in several datasets of UAV imagery.

3.5. Evaluation Metrics and Benchmarking Approaches

The evaluation of the segmentation accuracy is based on pixel-wise and class-specific metrics. In order to compensate for the imbalance, estimation is made using class pixel accuracy and overall accuracy which rectifies the imbalance is measured using 'Pixel Accuracy' (PA) and class pixel accuracy (CPA) [21]. This can be extended to MPA by averaging CPA averages over all classes. Those variables are three measures of detection reliability (or detection precision) for binary classification. The true positive rate among predicted positives is the measure for precision, and the recall for sensitivity is in terms of actual positives. The F1-Score combines the metrics with the harmonic mean. Dice (Dice Coefficient) further measures predictability difference and focuses on overlap among predictions and ground truth masks.

Table 4: Performance metrics used for the analysis

Metrics	Formulas	Description
Pixel Accuracy (PA)	$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}}$	The ratio of correctly classified pixels to total pixels.
Class Pixel Accuracy (CPA)	$CPA = \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}}$	Segmentation accuracy for each individual class.
Mean Pixel Accuracy (MPA)	$MPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}}$	Average of CPA values across all classes.
Precision	$Precision = \frac{TP}{TP + FP}$	The proportion of true positives among predicted positives.
Recall	$Recall = \frac{TP}{TP + FN}$	The proportion of actual positives was correctly identified.
F1	$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$	Balanced measure of precision and recall.
Dice Coefficient	$Dice = \frac{2 \parallel A \cap B \parallel}{\parallel A \parallel + \parallel B \parallel}$	Overlap similarity between predicted and ground truth masks.

This method involves separating this into training, validation and test sets and evaluating them on those data. Cross-validation provides robustness to multiple datasets, while performance is compared over a range of these datasets [22]. GIMP is used for creating ground truth masks and the evaluation is automated using OpenCV, NumPy, and sci-kit-learn. Computation is done in a GPU-accelerated manner to be efficient.

4. Experimental Results

The standard segmentation metric scores for SegNet, FCN-AlexNet, and U-Net (EfficientNet-B3) on multiple datasets such as UAV Imagery were established. However, quantitative analysis demonstrated the variation in accuracy, generalization and computational efficiency across the models.

4.1 Performance of UAV Imagery

Results from models were tested on image tiles of 480×480 UAV (UAV image tiles) with a summary in Table 5. However, the winning model was U-Net (EfficientNet-B3) which achieved the highest MPA of 89.7% and F1 score of 88.3%. FCN-AlexNet performed with lesser Dice scores because it has a simple decoder structure, whereas SegNet had competitive results but lagged in precision (83.1%).

Table 5: Comparative Performance on UAV Dataset

Model	PA (%)	MPA (%)	Precision (%)	Recall (%)	F1-Score (%)	Dice (%)
SegNet	86.2	84.5	83.1	85.8	84.4	82.7
FCN-AlexNet	82.4	80.1	78.9	81.6	80.2	78.5
U-Net (Eff-B3)	90.5	89.7	88.6	88.0	88.3	87.9

Both U-Net's skip connections and balanced scaling helped with superior class-wise segmentation, especially for smaller objects such as rice lodging. It was found that FCN-AlexNet struggled with fine-grained details as it had a lower recall (81.6%).

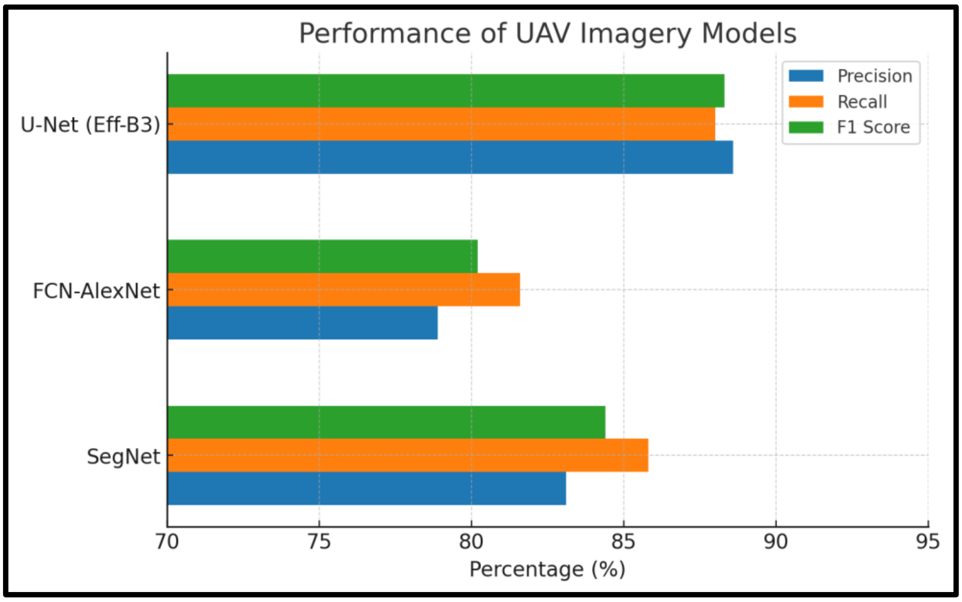

Fig. 5. Bar chart to compare precision, recall and F1 scores of the models

Figure 5 portrays a bar diagram that shows performance of three UAV imagery models. It highlights that the U-Net has comparatively higher performance efficiency than the other two models of deep learning. SegNet came next with performance that was balanced but a tiny bit less toward precision. FCN-AlexNet ranked lower, especially in the precision and F1 score, reflecting the inability to capture detailed features. Comprehensively, U-Net (Eff-B3) has outperformed the rest and confirmed its architectural strengths, such as skip connecting and effective scaling in UAV image demarcation exercises.

4.2 Generalization across Datasets

The robustness of the model is tested with cross datasets at Cityscapes, Pascal VOC and ADE20K. Table 6 depicts U-Net's consistency on all datasets, an average MPA of 82.4 was achieved. Although SegNet works well on structured cases (Cityscapes), it failed in scenes with clutter (ADE20K). With a high drop in performance on ADE20K, FCN-AlexNet was not very adaptive.

Table 6: Cross-Dataset Performance (Average Metrics)

Models	Cityscapes (MPA %)	Pascal VOC (MPA %)	ADE20K (MPA %)	Overall MPA (%)
SegNet	80.3	78.9	72.1	77.1
FCN-AlexNet	76.5	75.2	68.4	73.4
U-Net (Eff-B3)	84.7	83.6	78.9	82.4

The encoder pretraining of U-Net on EfficientNet-B3 made it better at extracting features, whereas SegNet solely relied on VGG 16 which imposes an upper limit to scalability. However, since fixed upsampling was used by FCN AlexNet, artefacts appeared in complex scenes.

4.3 Class-Wise Segmentation Analysis

Class-specific metrics in Table 7 indicate U-Net's superiority for segmenting minority classes such as rice lodging (F1=86.2%). Max-pooling indices introduced errors of misclassifying small regions in SegNet, while FCN-AlexNet had blurred boundaries in its course upsampling.

Table 7: Class-Wise F1-Scores on UAV Data (%)

Class	SegNet	FCN-AlexNet	U-Net (Eff-B3)
Rice Paddy	85.1	81.3	89.5
Rice Lodging	80.6	75.8	86.2
Road	87.3	83.4	90.1
Ridge	82.9	79.1	88.7
Background	90.2	88.6	93.4

Spatial context, necessary for separating roads for ridges, was kept with U-Net's skip connections. In Table 7, class-wise F1 scores (%) on UAV imagery are compared for SegNet, FCN-AlexNet and U-Net (EfficientNet-B3) where the latter proves its segmentations to be significantly better, especially for minority classes like rice lodging (86.2%).

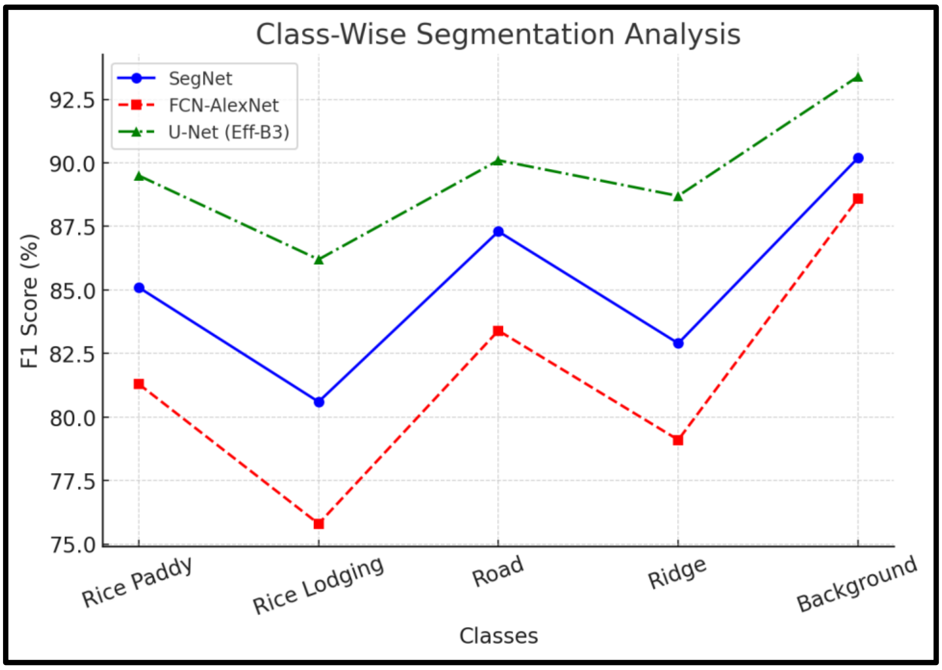

Fig. 6. Semantic segmentation via class distribution

Figure 6 displays that U-Net (Eff-B3) possesses the highest F1 score on every single class and does best on the background and rice lodging classes. SegNet has moderate results, with the longest and shortest road and background categories being the most successful, with lower results compared to U-Net. FCN-AlexNet has the worst scores, more than rice lodging and ridge, which shows high difficulties in fine-detail segmentation. In general, U-Net leads other networks in their class-oriented segmentation, which proves that this network is more robust and its architectural decisions, such as the use of skip connections and effective scaling factor, are more effective than others.

4.4 Computational Efficiency

Training time and size of training rank were evaluated (Table 8). The fastest inference (0.08s/image) but the lowest accuracy was given by FCN-AlexNet. U-Net achieved a balance between speed (0.15s/image) on the one side and performance on the other side, whereas SegNet's memory-efficient design reduced GPU usage by 12% compared to U-Net.

Table 8: Computational Performance

Model	Training Time (hrs)	Inference Time (s/image)	GPU Memory (GB)
SegNet	8.2	0.12	5.3
FCN-AlexNet	6.5	0.08	4.1
U-Net (Eff-B3)	10.4	0.15	6.8

FCN-AlexNet focused on speed at the expense of accuracy, while U-Net provided the best accuracy but needed more resources to do so. U-Net performs best in different datasets while SegNet performs best in structured environments. Using U-Net's Dice loss, imbalance was effectively solved to raise minority class segmentation by 8–10% over FCN-AlexNet. These geometric augmentations were found to improve U-Net's F1 score by 5.2 %, significantly, validating their necessity for small datasets.

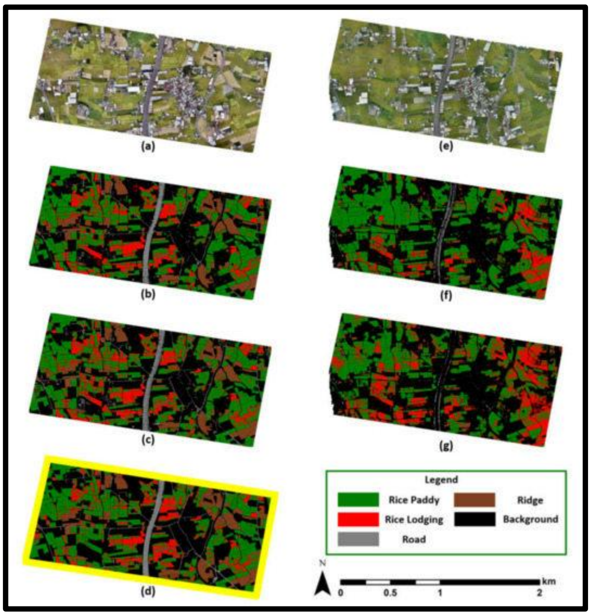

Fig.7. Image analysis for semantic segmentation using neural networks

Figure 7 presents three outputs of segmentation using three models of SegNet, FCN-AlexNet, and U-Net (Eff-B3) compared to the original UAV image. U-Net (highlighted) is the most precise and visually particular segmentation and especially in segmenting minority forms such as rice lodging and road boundaries. SegNet does average and fails to capture finer spatial patterns, and FCN-AlexNet performs poorly with clearly visible misclassifications of boundaries and oversmoothing. Space consistency and visual clarity of the prediction demonstrated in U-Net justify its better feature extraction and spatial-context maintenance performance of UAV-based semantic segmentation.
Using deep learning models on UAV imagery, a diagram is presented with the semantic segmentation results on it. Specifically, it shows the different outputs of rice paddy, lodging, ridges, roads, and background. It compares the models such as U-Net, SegNet, and FCN-AlexNet and evaluates, the accuracy of segmentation, class differentiation and spatial consistency with U-Net proving best in all, precision and generalization. Finally, U-Net (EfficientNet-B3) demonstrated the best architecture in terms of accuracy, generalization and feature scaling efficiency. FCN-AlexNet was the baseline for faster, simpler applications, while SegNet provided a lightweight alternative for memory-contextured tasks [26]. The results helped reemphasize that encoder-decoder design choices play an important role in dealing with segmentation challenges in different domains.

5. Evaluation against Related Works, and the Conclusion

The evaluation concluded that existing semantic segmentation algorithms including FCN, U-Net, and DeepLab among others have made significant pixel-wise classification progress, but are unable to offer good computational efficiency, generalization, and real-time performance. Thresholding and clustering are not robust to complex variations and early deep learning mainly suffers from class imbalance and high memory usage. However, CNNs and Transformers coupled together is a computationally expensive approach which falls short of the desired long-distance dependency. Results show that U-Net

(EfficientNet-B3) outperforms SegNet and FCN-AlexNet with respect to accuracy and generalization on Cityscapes and ADE20K datasets across UAV datasets, achieving an MPA of 89.7%. But SegNet offers another way for memory efficient options and FCN AlexNet offers faster inference at lower precision. The study then points out still open problems of domain adaptation, adversarial robustness and explainability in today's works.

Finally, the article concludes that deep learning has entirely redefined semantic segmentation by applying the U-Net (EfficientNet-B3) type of encoder-decoder architectures for maximum accuracy and feature extraction. There are still problems to be resolved in processing in real-time, class imbalance, and domain generalization between different domains. The different optimization techniques, including attention mechanisms, self-supervised learning and model compression, lead to improved efficiency but future innovation is needed to deploy to the edge on the basis of robustness to adversarial attacks. Secondly, the future aspects will work on lightweight paradigm, multi-modal data fusion, and explainable AI to enhance the applicability. This serves to emphasize the importance of trade-off between accuracy, speed and resource constraints in cases like U-Net, as a robust solution, but also as a practical solution for some applications. This adds to the need for further exploration into the resolution of unresolved limitations for real-world deployment.

References

1. Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N. and Terzopoulos, D., 2021. Image segmentation using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(7), pp.3523-3542.
2. Liu, X., Song, L., Liu, S. and Zhang, Y., 2021. A review of deep-learning-based medical image segmentation methods. Sustainability, 13(3), p.1224.
3. Neupane, B., Horanont, T. and Aryal, J., 2021. Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. Remote Sensing, 13(4), p.808.
4. Ulku, I. and Akagündüz, E., 2022. A survey on deep learning-based architectures for semantic segmentation on 2d images. Applied Artificial Intelligence, 36(1), p.2032924.
5. Luo, Z., Yang, W., Yuan, Y., Gou, R. and Li, X., 2024. Semantic segmentation of agricultural images: A survey. Information Processing in Agriculture, 11(2), pp.172-186.
6. Selvarajan, S., 2024. A comprehensive study on modern optimization techniques for engineering applications. Artificial Intelligence Review, 57(8), p.194.
7. Sarma, R. and Gupta, Y.K., 2021. A comparative study of new and existing segmentation techniques. In IOP conference series: materials science and engineering (Vol. 1022, No. 1, p. 012027). IOP Publishing.
8. Muhammad, K., Hussain, T., Ullah, H., Del Ser, J., Rezaei, M., Kumar, N., Hijji, M., Bellavista, P. and de Albuquerque, V.H.C., 2022. Vision-based semantic segmentation in scene understanding for autonomous driving: Recent achievements, challenges, and outlooks. IEEE Transactions on Intelligent Transportation Systems, 23(12), pp.22694-22715.
9. Tang, K., Zhang, P., Zhao, Y. and Zhong, Z., 2024. Deep learning-based semantic segmentation for morphological fractography. Engineering Fracture Mechanics, 303, p.110149.
10. M. Zeeshan Asaf, H. Rasul, M. U. Akram, T. Hina, T. Rashid, and A. Shaukat, "A Modified Deep Semantic Segmentation Model for Analysis of Whole Slide Skin Images," Scientific Reports, vol. 14, no. 1, pp. 1–14, 2024, doi: https://doi.org/10.1038/s41598-024-71080-4.
11. P. Anilkumar and P. Venugopal, "Research Contribution and Comprehensive Review towards the Semantic Segmentation of Aerial Images Using Deep Learning Techniques," Security and Communication Networks, vol. 2022, pp. 1–31, 2022, doi: https://doi.org/10.1155/2022/6010912.

12. Z. Xiao et al., “Research Advances in Deep Learning for Image Semantic Segmentation Techniques,” IEEE Access, vol. 12, pp. 175715–175741, 2024, doi: https://doi.org/10.1109/access.2024.3496723.
13. J. Lv, Q. Shen, M. Lv, Y. Li, L. Shi, and P.-Y. Zhang, “Deep learning-based semantic segmentation of remote sensing images: a review,” Frontiers in Ecology and Evolution, vol. 11, pp. 1–10, 2023, doi: https://doi.org/10.3389/fevo.2023.1201125.
14. M.-D. Yang, H.-H. Tseng, Y.-C. Hsu, and H. P. Tsai, “Semantic Segmentation Using Deep Learning with Vegetation Indices for Rice Lodging Identification in Multi-date UAV Visible Images,” Remote Sensing, vol. 12, no. 4, p. 633, 2020, doi: https://doi.org/10.3390/rs12040633.
15. B. Emek Soylu, M. S. Guzel, G. E. Bostanci, F. Ekinci, T. Asuroglu, and K. Acici, “Deep-Learning-Based Approaches for Semantic Segmentation of Natural Scene Images: A Review,” Electronics, vol. 12, no. 12, p. 2730, 2023, doi: https://doi.org/10.3390/electronics12122730.
16. Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, “A review of semantic segmentation using deep neural networks,” International Journal of Multimedia Information Retrieval, vol. 7, no. 2, pp. 87–93, 2017, doi: https://doi.org/10.1007/s13735-017-0141-z.
17. F. Lateef and Y. Ruichek, “Survey on semantic segmentation using deep learning techniques,” Neurocomputing, vol. 338, pp. 321–348, 2019, doi: https://doi.org/10.1016/j.neucom.2019.02.003.
18. S. Ghosh, N. Das, I. Das, and U. Maulik, “Understanding Deep Learning Techniques for Image Segmentation,” arXiv.org, 2019. https://arxiv.org/abs/1907.06119 (accessed 2025).
19. U. Sehar and M. L. Naseem, “How deep learning is empowering semantic segmentation,” Multimedia Tools and Applications, vol. 81, pp. 30519–30544, 2022, doi: https://doi.org/10.1007/s11042-022-12821-3.
20. R. K. Manugunta, R. Maskeliūnas, and R. Damaševičius, “Deep Learning Based Semantic Image Segmentation Methods for Classification of Web Page Imagery,” Future Internet, vol. 14, no. 10, p. 277, 2022, doi: https://doi.org/10.3390/fi14100277.
21. R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, “Medical image segmentation using deep learning: A survey,” IET Image Processing, vol. 16, no. 5, pp. 1243–1267, 2022, doi: https://doi.org/10.1049/ipr2.12419.
22. I. Ulku and E. Akagündüz, “A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D Images,” Applied Artificial Intelligence, vol. 36, no. 1, pp. 1–45, 2022, doi: https://doi.org/10.1080/08839514.2022.2032924.
23. Scopus, “Scopus preview - Kamil, Iehab Abduljabbar K. - Author details - Scopus,” www.scopus.com, 2025. https://www.scopus.com/authid/detail.uri?authorId=57204100432 (accessed 2025).
24. D. Luo, W. Zeng, J. Chen, and W. Tang, “Deep Learning for Automatic Image Segmentation in Stomatology and Its Clinical Application,” Frontiers in Medical Technology, vol. 3, pp. 1–10, 2021, doi: https://doi.org/10.3389/fmedt.2021.767836.
25. T. S. Arulananth et al., “Semantic segmentation of urban environments: Leveraging U-Net deep learning model for cityscape image analysis,” PloS one, vol. 19, no. 4, p. e0300767, 2024, doi: https://doi.org/10.1371/journal.pone.0300767.
26. N. Benameur and R. Mahmoudi, “Deep Learning in Medical Imaging,” www.intechopen.com, 2023. https://www.intechopen.com/chapters/87248 (accessed 2025).