Engineering Science & Technology
https://ojs.wiserpub.com/index.php/EST/

UNIVERSAL WISER PUBLISHER

Research Article

Uncertainty-Aware Semantic Segmentation via Decision Boundary Modeling and *Confusion Zone* Refinement

Antoni Mestre[1*], **Franccesco Malafarina**[2]

[1]VRAIN Institute, Universitat Politècnica de València, Camino de Vera s/n, Valencia, 46022, Spain
[2]Università degli Studi del Sannio, Piazza Guerrazzi 1, Benevento, 82100, Italy
E-mail: anmesgas@vrain.upv.es

Received: 29 June 2026; **Revised:** 3 August 2026; **Accepted:** 3 August 2026

Abstract: Deep learning has significantly improved the performance of semantic segmentation in high-resolution imagery and computer vision tasks. However, standard segmentation models treat all pixel-wise predictions equally, despite the fact that predictions close to the decision boundary are inherently uncertain. This limitation often leads to unstable classifications and increased false positives or false negatives in complex scenes. In this paper, we propose a practical uncertainty-aware semantic segmentation refinement framework that explicitly handles decisionboundary ambiguity through an empirical routing mechanism through a *Confusion Zone* mechanism. Instead of forcing a binary decision for all pixels, predictions whose probability scores fall within a predefined interval [τ_{low}, τ_{high}] are marked as uncertain and redirected to a secondary refinement module that exploits additional spatial features and local context before assigning a final label. The proposed two-stage architecture allows the segmentation model to treat ambiguous predictions differently from confident ones, improving robustness near object boundaries and visually complex regions. Experiments on benchmark remote sensing segmentation datasets demonstrate that incorporating the *Confusion Zone* mechanism improves segmentation quality, achieving up to 0.851 IoU with SegFormer and consistent gains of 2-3 percentage points across U-Net, DeepLabV3+, and SegFormer backbones. These results suggest that explicitly routing decision-boundary predictions for local refinement is a simple and practical strategy for improving the reliability of deep semantic segmentation systems.

Keywords: emantic segmentation, uncertainty-aware segmentation, decision-boundary routing, *confusion zone*, remote sensing

1. Introduction

Semantic segmentation is a fundamental task in computer vision that aims to assign a semantic label to every pixel in an image. In recent years, deep learning approaches have significantly advanced the state of the art in pixel-level scene understanding. Early convolutional architectures such as UNet [1] and DeepLab [2] established the foundation for modern segmentation pipelines by enabling end-to-end learning of dense spatial representations. More recently, transformer-based architectures such as SegFormer [3] and Mask2Former [4] have further improved segmentation performance by modeling long-range dependencies and multi-scale contextual information. In addition, large-scale foundation models such as the Segment Anything Model (SAM) [5] have demonstrated unprecedented generalization

DOI: https://doi.org/10.37256/est.81202610368

capabilities, enabling segmentation across diverse visual domains with minimal task-specific adaptation.

Despite these advances, most semantic segmentation pipelines rely on a simple deterministic decision rule applied to probabilistic model outputs. Typically, the network produces a probability map representing the likelihood that each pixel belongs to a given class, and a fixed threshold is applied to convert these probabilities into discrete predictions. While this approach is computationally efficient and widely adopted, it implicitly assumes that predictions across the probability spectrum can be treated uniformly. In practice, however, predictions that lie close to the decision boundary often reflect inherent uncertainty in the learned representation. Such uncertainty commonly arises near object boundaries, in visually heterogeneous regions, or when the model encounters ambiguous patterns in the input data.

The role of uncertainty in deep learning has received increasing attention in recent years. A variety of approaches have been proposed to quantify predictive uncertainty in neural networks, including Bayesian inference methods, ensemble-based techniques, and uncertainty-aware loss functions [6]-[8]. These methods aim to capture both epistemic uncertainty, which reflects model uncertainty due to limited knowledge, and aleatoric uncertainty, which arises from inherent noise in the data. Recent studies have also proposed frameworks for systematically evaluating uncertainty estimation in segmentation systems [9]. While these approaches provide valuable information about prediction confidence, they typically focus on estimating uncertainty rather than incorporating it directly into the segmentation decision process.

From a decision-theoretic perspective, treating uncertain predictions in the same way as highly confident ones may lead to unstable segmentation outcomes and increased misclassification rates. Pixels with probability values close to the classification threshold often correspond to regions where the model lacks sufficient evidence to make a reliable decision. Ignoring this ambiguity may therefore propagate errors across the segmentation map, particularly in visually complex scenes where boundaries and textures are difficult to distinguish.

In this paper, we propose an uncertainty-aware semantic segmentation framework that explicitly models ambiguity near the decision boundary. Inspired by decision-boundary modeling approaches in binary classification, we introduce a mechanism referred to as the *Confusion Zone*. Instead of forcing an immediate binary decision, predictions whose probability values fall within a predefined interval $[\tau_{\text{low}}, \tau_{\text{high}}]$ are treated as uncertain. These ambiguous predictions are redirected to a secondary refinement module that incorporates additional contextual information before assigning the final label. By separating confident predictions from uncertain ones, the proposed approach allows the segmentation system to handle ambiguous regions in a more robust and structured manner.

The proposed framework is model-agnostic and can be integrated with different semantic segmentation architectures without requiring substantial modifications to the underlying network. By explicitly incorporating decisionboundary uncertainty into the segmentation pipeline, the method improves robustness in challenging image regions while preserving the efficiency of standard deep learn-ing workflows.

Unlike conventional uncertainty-aware segmentation approaches that primarily estimate predictive uncertainty as a scalar quantity, the proposed framework introduces a decision-boundary modeling strategy that explicitly partitions predictions into confident and ambiguous regions. This enables a structured inference process in which uncertain predictions are not merely identified but actively redirected for additional processing. In this sense, the *Confusion Zone* transforms uncertainty from a passive measure into an operational decision-routing mecha-nism within the segmentation pipeline.

It is important to note that the proposed method is not intended as a new probabilistic theory of uncertainty estimation. Rather, it is presented as a practical engineering strategy for incorporating decision-boundary uncertainty into an existing segmentation workflow. The novelty of the approach lies in the explicit operational use of an uncertainty interval as a routing mechanism, which separates confident predictions from ambiguous ones and applies a lightweight local refinement procedure only to the latter. This design choice aims to improve segmentation reliability while preserving compatibility with standard segmenta-tion backbones and avoiding the computational cost of stochastic uncertainty-estimation methods.

The main contributions of this work are threefold. First, we formulate a simple decision-boundary-aware routing mechanism for semantic segmentation, in which predictions close to the classification threshold are explicitly separated from confident predictions. Second, we implement this mechanism as a lightweight two-stage engineering strategy, where uncertain pixels are refined using local spatial and probability-map context. Third, we provide an empirical evaluation showing that this practical refinement strategy can improve segmen-tation performance across different backbone architectures, without requiring substantial modifications to the original models.

While the proposed framework is general in design, this study focuses on remote sensing imagery as a representative high-resolution domain characterized by complex spatial structures and ambiguous object boundaries.

2. Related work

2.1 *Deep learning for semantic segmentation*

Semantic segmentation has experienced substantial progress with the emergence of deep learning methods. Early convolutional neural network architectures such as Fully Convolutional Networks (FCN) [10], U-Net [1], and SegNet [11] established the foundations of modern segmentation pipelines by enabling end-to-end learning for dense pixel-wise prediction. These encoder-decoder architectures leverage hierarchical feature extraction and multi-scale contextual information to accurately delineate objects and structures in images. Later works introduced improvements in contextual modeling and multi-scale reasoning, such as PSPNet [12] and DeepLab [2], which incorporate spatial pyramid pooling and atrous convolution to better capture global scene context.

More recent research has explored architectures capable of preserving highresolution representations throughout the network. For instance, HRNet [13] maintains multiresolution features in parallel, allowing detailed spatial information to be retained during feature extraction. Such approaches have demonstrated improved performance in scenarios where precise boundary delineation is required.

Transformer-based architectures have further improved segmentation performance by capturing long-range dependencies and global contextual information. Models such as SegFormer [3] introduce lightweight transformer encoders combined with efficient decoding mechanisms, achieving strong performance while maintaining computational efficiency. Similarly, universal segmentation architectures such as Mask2Former [4] unify semantic, instance, and panoptic segmentation under a single transformer-based framework. In addition, hierarchical vision transformers such as Swin Transformer [14] have shown strong performance across multiple vision tasks, including semantic segmentation.

In parallel, the development of large-scale foundation models has introduced a new paradigm for segmentation. The Segment Anything Model (SAM) [5] illustrates the potential of large pretrained models to generalize segmentation capabilities across diverse visual domains with minimal supervision. Such models highlight the growing trend toward general-purpose segmentation systems that can operate across heterogeneous datasets and tasks.

2.2 *Uncertainty in deep learning*

Uncertainty estimation has become an important research direction in deep learning, particularly for applications where prediction reliability is critical. Bayesian approaches such as Monte Carlo dropout [15] and deep ensembles [16] have been widely used to estimate epistemic uncertainty in neural networks. These techniques approximate Bayesian inference and allow models to quantify predictive confidence without substantially modifying the underlying architecture.

More recent work has explored improved uncertainty-aware training objectives and calibration techniques. For example, uncertainty-aware loss functions such as U-CE aim to explicitly model prediction uncertainty during training, improving segmentation reliability in ambiguous regions [6]. In addition, recent frameworks such as ValUES provide systematic methodologies for evaluating uncertainty estimation in semantic segmentation systems [9]. Comprehensive surveys have also highlighted the importance of uncertainty quantification in deep learning, outlining various approaches for modeling epistemic and aleatoric uncertainty in vision tasks [7], [8].

Another important line of research focuses on model calibration and confidence estimation in deep neural networks. Studies have shown that modern neural networks tend to produce overconfident predictions, motivating the development of calibration methods such as temperature scaling and probabilistic confidence modeling [17]. These methods aim to improve the reliability of predictive probabilities, which is particularly relevant when model outputs are used in downstream decision-making processes.

Despite these advances, most existing research focuses primarily on estimating uncertainty rather than integrating it directly into the decisionmaking process of segmentation pipelines. In many practical systems, uncertainty estimates are used only for posthoc analysis or visualization, while the final segmentation still relies on deterministic thresholding of probability maps.

2.3 *Uncertainty-aware decision mechanisms*

A smaller body of research has investigated methods for incorporating uncertainty into the decision process of machine learning models. Building on prior work on text classification, where ambiguous probability scores within a Confusion Zone were routed to a sentiment-analysis module before final classification [18], the present work extends this decision-routing paradigm to semantic segmentation by introducing an explicit decision-boundary routing strategy. These approaches often aim to identify ambiguous predictions and treat them differently from highly confident outputs. In segmentation tasks, uncertain predictions frequently occur near object boundaries or in visually complex regions where class distinctions are difficult to determine.

Recent studies have explored adaptive thresholding, confidence-based filtering, and uncertainty-driven refinement strategies to address this challenge. For instance, confidenceaware segmentation frameworks attempt to refine predictions in low-confidence regions or incorporate additional contextual cues to improve reliability. However, many of these approaches rely on global uncertainty measures or additional stochastic sampling procedures, which may increase com-putational complexity and inference time.

The proposed approach is related to confidence-based refinement methods, which typically use low-confidence predictions to trigger post-processing, filtering, or additional correction steps. However, there are three specific differences. First, the *Confusion Zone* is defined explicitly around the decision boundary, rather than by using a generic low-confidence score over the entire probability range. Second, the mechanism separates the segmentation process into three decision regions: confident background, confident target, and ambiguous predic-tions requiring refinement. Third, the refinement stage is applied selectively only to pixels within this intermediate region, using local image and probability-map neighborhoods. Therefore, the method should be understood as a lightweight decision-routing strategy rather than as a general-purpose uncertainty quantification framework.

In contrast, the approach proposed in this work focuses on explicitly modeling the decision boundary of segmentation predictions. By introducing a *Confusion Zone* around the classification threshold, the proposed framework identifies ambiguous predictions and redirects them to a secondary refinement module for additional analysis. This mechanism enables the segmentation system to handle uncertain regions in a structured and computationally efficient manner while preserving compatibility with existing deep learning architectures.

3. Methodology

3.1 *Proposed architecture*

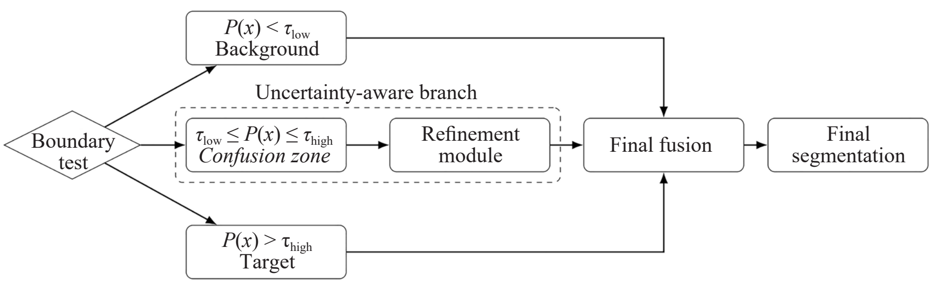

Figure 1. Workflow of the proposed uncertainty-aware segmentation approach. Predictions outside the confidence interval are directly classified, while pixels inside the *Confusion Zone* are processed by a refinement module before the final fusion stage produces the final segmentation output

The proposed architecture follows a two-stage uncertainty-aware segmentation workflow. In the first stage, an input image is processed by a segmentation backbone, such as U-Net or SegFormer, to generate a pixel-wise probability map. Instead of directly thresholding all predictions, the method evaluates whether each probability score falls into a confident or ambiguous region. Predictions outside the uncertainty interval are directly assigned to the background or target class, whereas predictions within the interval $[\tau_{\mathrm{low}}, \tau_{\mathrm{high}}]$ are considered part of the *Confusion Zone*. These uncertain predictions are then passed to a refinement module that exploits additional local context and spatial consistency before producing the final segmentation output. Figure 1 summarizes the complete workflow.

The first stage consists of a deep semantic segmentation network that produces a pixel-wise probability map. In this work, we adopt an encoderdecoder architecture based on U-Net due to its effectiveness in dense prediction tasks and its ability to preserve spatial information through skip connections. The encoder extracts hierarchical feature representations from the input image, while the decoder reconstructs a full-resolution segmentation map.

Given an input image I, the segmentation backbone generates a probability map $P(x)$ where each pixel represents the likelihood of belonging to the target class. Instead of applying a single global threshold to obtain binary predictions, the proposed approach introduces a *Confusion Zone* around the decision boundary.

Pixels whose probability values fall within a predefined interval $[\tau_{\mathrm{low}}, \tau_{\mathrm{high}}]$ are classified as uncertain and are forwarded to a secondary refinement module. Predictions outside this interval are directly assigned to their corresponding class labels.

The refinement module analyzes additional spatial context to resolve ambiguous predictions. Specifically, it considers local neighborhood consistency and contextual features extracted from the segmentation probability map. This step allows the model to refine predictions in regions where the backbone network exhibits low confidence, such as object boundaries or visually heterogeneous areas.

The final segmentation map is obtained by combining the confident predictions from the backbone network with the refined labels produced by the uncertainty refinement module. This two-stage architecture enables the system to explicitly handle decision-boundary uncertainty while maintaining the computational efficiency of standard segmentation pipelines.

3.2 *Datasets*

To evaluate the proposed segmentation framework, we employ several publicly available datasets commonly used for building detection and semantic segmentation in remote sensing imagery. These datasets provide high and very high spatial resolution imagery with annotated building footprints.

The SpaceNet dataset [19] contains high-resolution satellite imagery acquired by DigitalGlobe satellites with detailed building footprint annotations across several geographic locations. In particular, SpaceNet 1 provides imagery with a spatial resolution of approximately 0.5 meters and thousands of labeled building instances.

The INRIA Aerial Image Labeling Dataset [20] consists of high-resolution aerial images (0.3 m spatial resolution) covering multiple cities across Europe and the United States. The dataset includes large-scale building footprint annotations and has become a widely used benchmark for building segmentation tasks.

In addition, SpaceNet 7 [21] provides multi-temporal satellite imagery across more than 100 locations worldwide, enabling the study of segmentation robustness across different geographic regions and environmental conditions.

To increase the diversity of the training data, images from these datasets are combined and reorganized into training sets. While SpaceNet and SpaceNet 7 contribute to training data diversity across geographic regions and spatial resolutions, quantitative evaluation is conducted exclusively on the INRIA dataset, which provides a standardized benchmark with established evaluation protocols for building segmentation in high-resolution aerial imagery.

3.3 *Baseline segmentation model*

The baseline segmentation model used in this work follows an encoder-decoder architecture based on the U-Net framework [1]. U-Net has been widely adopted for semantic segmentation tasks due to its ability to capture both global contextual information and finegrained spatial details.

The architecture consists of two main components: an encoder and a decoder. The encoder progressively extracts hierarchical features through convolutional and pooling layers, reducing spatial resolution while increasing the number of feature channels. The decoder then reconstructs the segmentation map through upsampling layers, combining high-level semantic information with fine spatial details via skip connections between encoder and decoder layers.

Given an input image I, the segmentation network produces a probability map:

$$P(y_i \mid x_i)$$

where x_i represents the input pixel and y_i denotes the predicted class label. The output probability map indicates the likelihood that each pixel belongs to the target class.

3.4 *Confusion zone for decision-boundary uncertainty*

Standard segmentation pipelines convert probability maps into binary predictions using a fixed threshold, typically set to 0.5. However, predictions near the decision boundary are inherently uncertain and frequently correspond to visually ambiguous regions such as object boundaries or heterogeneous textures.

To address this limitation, we introduce a *Confusion Zone* mechanism that explicitly models decision-boundary uncertainty. Instead of forcing a binary decision for all pixels, predictions whose probabilities fall within an intermediate interval are considered uncertain and processed separately.

Formally, the segmentation decision is defined as:

$$y = \begin{cases} 0 & P(x) < \tau_{\text{low}} \\ 1 & P(x) > \tau_{\text{high}} \\ U & \tau_{\text{low}} \le P(x) \le \tau_{\text{high}} \end{cases}$$

where τ_{low} and τ_{high} define the boundaries of the *Confusion Zone* and U represents uncertain predictions.

The interval $[\tau_{\text{low}}, \tau_{\text{high}}]$ should be interpreted as a validation-guided design parameter rather than as a universal uncertainty bound. In binary segmentation settings, a threshold around 0.5 is commonly used as the decision boundary. The proposed *Confusion Zone* therefore defines a symmetric margin around this boundary in order to identify predictions for which the model output is neither clearly associated with the background nor with the target class. The width of this interval controls the trade-off between refinement coverage and computational selectivity: wider intervals route more pixels to the refinement module, while narrower intervals restrict the refinement stage to only the most ambiguous predictions.

Pixels assigned to the *Confusion Zone* correspond to regions where the segmentation model lacks sufficient confidence to produce a reliable classification.

3.5 *Refinement of uncertain predictions*

Pixels classified within the *Confusion Zone* are redirected to a secondary refinement stage designed to exploit local spatial context and improve classification reliability.

For each uncertain pixel, a fixed-size neighborhood window (e.g., 15 × 15) is extracted from both the original input image and the corresponding probability map generated by the segmentation backbone. These inputs are concatenated and processed by a lightweight convolutional refinement network composed of two convolutional layers followed by a sigmoid activation function.

Formally, for a pixel x_i belonging to the *Confusion Zone*, the refinement module computes a corrected probability value:

$$\hat{P}(x_i) = R\left(I_{\mathcal{N}(i)}, P_{\mathcal{N}(i)}\right)$$

where $I_{N(i)}$ and $P_{N(i)}$ denote the local neighborhood patches of the input image and probability map, respectively, and $R(\cdot)$ represents the refinement function implemented by the convolutional module.

The refinement module is trained after the segmentation backbone has been trained. In this stage, the backbone parameters are kept fixed a nd t he trained model is used to generate probability maps for the training images. Pixels whose predicted probabilities fall within the *Confusion Zone* are then selected as refinement samples. For each selected pixel, a local image patch and the corresponding local probability-map patch are extracted and concatenated as input to the refinement module.

The target label for each refinement sample is obtained from the groundtruth segmentation mask. The refinement network is optimized using binary cross-entropy loss, computed only for the pixels routed to the *Confusion Zone*. This training strategy ensures that the refinement module specializes in resolving ambiguous decision-boundary cases, rather than relearning the full segmentation task. At inference time, the same routing rule is applied: confident pixels are directly assigned to their corresponding class, while only uncertain pixels are processed by the refinement module.

By incorporating local spatial consistency and contextual cues, this module resolves ambiguous predictions that typically occur near object boundaries or visually complex regions. This two-stage process enables the model to treat confident and uncertain predictions differently, improving the robustness of the overall segmentation pipeline.

3.6 *Evaluation metrics*

To evaluate the performance of the proposed method, we employ the Intersection over Union (*IoU*) metric, which is commonly used in semantic segmentation tasks. *IoU* measures the overlap between the predicted segmentation and the ground truth:

$$IoU = \frac{TP}{TP + FP + FN}$$

where TP represents true positives, FP false positives, and FN false negatives.

In addition, precision, recall, and F1-score are also reported to provide a comprehensive evaluation of segmentation performance.

4. Experimental evaluation

This section evaluates the effectiveness of the proposed uncertainty-aware segmentation framework. We first describe the datasets and experimental protocol, followed by implementation details and quantitative performance results.

4.1 *Datasets and experimental protocol*

Following the data configuration described in Section 3.2, the backbone models were trained on the combined multi-source dataset, and quantitative evaluation was performed on the INRIA Aerial Image Labeling Dataset.

The proposed approach was evaluated using publicly available datasets commonly employed in high-resolution aerial image segmentation tasks.

In particular, experiments were conducted on the INRIA Aerial Image Labeling Dataset, which contains high-resolution aerial imagery (0.3 m spatial resolution) collected from several urban areas across Europe and the United States. Each image is annotated with detailed building footprint masks, enabling precise evaluation of semantic segmentation algorithms.

Following standard practice in aerial image segmentation, the dataset was divided into training and validation subsets. To increase the diversity of spatial contexts and improve computational efficiency during training, the images were divided into smaller patches. This patch-based strategy allows the model to learn both local spatial patterns and contextual information while increasing the number of training samples.

4.2 *Implementation details*

All experiments were implemented using the PyTorch deep learning framework. To evaluate the proposed uncertainty-aware mechanism across different segmentation architectures, we conducted experiments using three representative back-bone models: U-Net, DeepLabV3+, and SegFormer. These models were selected because they represent different design paradigms in semantic segmentation, including encoder-decoder convolutional networks and transformer-based archi-tectures.

Each model was trained in a fully supervised setting using RGB image patches as input. The networks produce a pixel-wise probability map indicating the likelihood that each pixel belongs to the target class. Training was performed using the Adam optimizer with an initial learning rate of 10^{-4}, and the binary cross-entropy loss function was used for optimization.

To improve model generalization and reduce overfitting, standard data augmentation techniques were applied during training. These included horizontal and vertical flipping as well as random rotations of the input patches.

The proposed *Confusion Zone* mechanism was applied as a post-processing stage on top of the probability maps generated by each backbone model. The thresholds defining the uncertainty interval were empirically set to

$$\tau_{\text{low}} = 0.4,\ \tau_{\text{high}} = 0.6.$$

These thresholds were selected empirically using the validation set and were not tuned on the test results. The objective was not to derive an optimal theoretical uncertainty bound, but to define a moderate decision-boundary interval that captures ambiguous predictions while avoiding excessive routing of already confident pixels. The sensitivity of the method to this choice is analyzed in the ablation study. When applying the method to other datasets, the same threshold values should not be assumed to be universally optimal. Instead, they should be reestimated on a validation subset, taking into account the calibration of the underlying backbone, the class imbalance of the segmentation task, and the proportion of pixels routed to the refinement module. Nevertheless, the ablation study indicates that a moderate interval around the standard 0.5 decision boundary provides a robust and computationally efficient starting point for high-resolution building segmentation tasks.

Pixels with probability values within this interval were considered uncertain and subsequently processed by the refinement module described in Section 3. Predictions outside this interval were directly assigned to their corresponding class labels.

This experimental setup allows the evaluation of the proposed uncertaintyaware strategy independently of the underlying segmentation architecture.

4.3 *Ablation study*

To assess the contribution of individual components and the sensitivity of the proposed mechanism to the uncertainty interval, we conduct an ablation study using U-Net as the backbone.

Table 1 reports the effect of removing the refinement module (i.e., uncertain pixels are directly assigned to the majority class of their neighborhood) and of varying the threshold interval $[\tau_{\text{low}}, \tau_{\text{high}}]$.

Table 1. Ablation study on the INRIA dataset (U-Net backbone)

Configuration	*IoU*	F1
U-Net (baseline)	0.783	0.814
U-Net + CZ, no refinement	0.794	0.823
U-Net + CZ, [0.3, 0.7]	0.808	0.836
U-Net + CZ, [0.4, 0.6] (ours)	**0.812**	**0.840**
U-Net + CZ, [0.45, 0.55]	0.805	0.832

The results indicate that the refinement module provides a measurable contribution beyond the routing mechanism alone. The interval $[\tau_{\text{low}} = 0.4,\ \tau_{\text{high}} = 0.6]$ yields the best performance, suggesting that a moderate uncertainty band captures the most ambiguous boundary pixels without over-routing confident pre-dictions. This also suggests that the selected thresholds are not intended to encode a dataset-independent theoretical notion of uncertainty. Rather, they provide an empirically effective operating point for the evaluated remote sensing setting. In datasets with different calibration properties, spatial resolutions, or class distributions, the *Confusion Zone* interval should be selected through the same validation-based procedure.

This result also confirms that the threshold interval acts as a practical design parameter rather than as a theoretically derived uncertainty estimator. A very wide interval routes too many pixels to the refinement module, including predictions that are already sufficiently reliable, whereas a very narrow interval fails to capture a substantial portion of boundary ambiguity. The selected inter-val therefore represents an empirical trade-off between refinement coverage and preservation of confident predictions.

4.4 *Results*

To assess the effectiveness of the proposed uncertainty-aware segmentation strategy, we compare the performance of several baseline segmentation architectures with both mainstream uncertainty-aware methods and the proposed *Confusion Zone* mechanism.

Table 2 reports the segmentation performance obtained by three commonly used architectures: U-Net, DeepLabV3+, and SegFormer. For the U-Net backbone, we include comparisons with representative uncertainty-aware approaches, namely Monte Carlo Dropout and the uncertainty-aware U-CE loss. Results are reported as mean and standard deviation over three independent runs to provide an indication of robustness.

For each architecture, we report the baseline performance as well as the performance obtained when incorporating the proposed *Confusion Zone* refinement stage. The results show that the proposed method consistently improves segmentation performance across all backbones and achieves competitive or superior re-sults compared to uncertainty-aware baselines, while maintaining a lightweight inference scheme.

Table 2. Performance and efficiency comparison between baseline segmentation models and their uncertainty-aware variants

Method	*IoU*	Prec.	Rec.	F1	Time (ms)
U-Net	0.783 ± 0.004	0.831 ± 0.005	0.802 ± 0.004	0.814 ± 0.004	18
U-Net + MC Dropout	0.791 ± 0.005	0.836 ± 0.006	0.809 ± 0.005	0.821 ± 0.005	142
U-Net + U-CE	0.795 ± 0.003	0.840 ± 0.004	0.812 ± 0.003	0.825 ± 0.004	19
U-Net + CZ (ours)	**0.812** ± 0.003	0.851 ± 0.004	0.831 ± 0.003	0.840 ± 0.003	**23**
DeepLabV3+	0.801 ± 0.004	0.852 ± 0.004	0.821 ± 0.004	0.834 ± 0.004	31
DeepLabV3+ + CZ (ours)	**0.829** ± 0.003	0.871 ± 0.003	0.851 ± 0.004	0.859 ± 0.003	**36**
SegFormer	0.822 ± 0.003	0.872 ± 0.004	0.843 ± 0.003	0.855 ± 0.003	27
SegFormer + CZ (ours)	**0.851** ± 0.002	0.891 ± 0.003	0.872 ± 0.003	0.880 ± 0.002	**33**

The results show that the proposed uncertainty-aware mechanism consistently improves segmentation performance across different backbone architectures. In particular, the incorporation of the *Confusion Zone* leads to improvements of approximately 2-3 percentage points in IoU across all tested models.

These improvements are more pronounced in regions where the baseline models exhibit lower confidence, such as object boundaries or areas with heterogeneous textures. By identifying predictions close to the decision boundary and processing them through a refinement stage, the proposed framework reduces both false positives and false negatives.

Another relevant observation is that the proposed approach is largely architectureagnostic. The performance improvements are observed not only for convolutional architectures such as U-Net and DeepLabV3+, but also for transformer-based models such as SegFormer. This suggests that the *Confusion Zone* mechanism can be integrated with a wide range of segmentation backbones without requiring modifications to their internal architecture.

Regarding computational overhead, the *Confusion Zone* mechanism adds a negligible inference cost relative to the backbone models (approximately 5-6 ms per patch), in contrast to MC Dropout, which requires multiple forward passes and incurs a substantially higher cost (~ 142 ms).

To further characterize the behavior of the proposed mechanism, we report the Expected Calibration Error (ECE) for the U-Net backbone: the baseline model achieves an ECE of 0.087, while U-Net + CZ reduces it to 0.071, suggesting that redirecting uncertain predictions to the refinement stage also yields bettercalibrated outputs as a secondary effect.

These results should be interpreted as evidence of the usefulness of a simple uncertainty-aware engineering refinement strategy, rather than a sevidence of a fundamentally new uncertainty estimation method. The main advantage of the proposed approach is its ability to introduce an explicit decision-boundary refinement step with limited computational overhead and without changing the internal architecture of the segmentation backbone.

5. Conclusions and future work

This paper presented an uncertainty-aware framework for semantic segmentation that explicitly addresses prediction ambiguity near the decision boundary. Instead of applying a single global threshold to all pixel-wise probabilities, the proposed approach introduces a *Confusion Zone* mechanism that identifies uncertain predictions and redirects them to a secondary refinement module. This strategy allows the segmentation system to treat confident and ambiguous predictions differently, improving the robustness of the final segmentation results.

The experimental evaluation shows that incorporating the *Confusion Zone* mechanism leads to consistent improvements across different segmentation backbones. In particular, the proposed framework achieves moderate but stable performance gains in Intersection over Union (*IoU*) and F1-score when compared to baseline architectures such as U-Net, DeepLabV3+, and SegFormer. These improvements suggest that explicitly modeling uncertainty around the classification threshold can reduce misclassification errors in challenging regions, such as object boundaries or visually heterogeneous areas.

Another important characteristic of the proposed method is its architectureagnostic design. The *Confusion Zone* mechanism operates as a post-processing stage that can be integrated with different segmentation networks without modifying their internal structure. This makes the approach compatible with a wide range of deep learning architectures, including both convolutional and transformer-based models.

Future work will explore several directions to further improve the proposed framework. First, adaptive strategies for determining the *Confusion Zone* thresholds could be investigated, allowing the uncertainty interval to dynamically adjust based on model confidence or dataset characteristics. Second, more advanced refinement modules based on attention mechanisms or contextual feature aggregation could be incorporated to better resolve ambiguous predictions. In addition, future work will extend the evaluation of the proposed approach to general-purpose computer vision benchmarks in order to assess its generalization capabilities across different visual domains. Finally, the integration of the proposed uncertainty-aware mechanism with large-scale foundation models for segmentation represents an interesting direction for future research.

The main limitation of the proposed approach is that the *Confusion Zone* is defined through fixed empirical thresholds. Consequently, the method should not be interpreted as a theory-based uncertainty modeling framework, but as a practical decision-boundary refinement mechanism. Although the empirical results indicate that this simple strategy is effective in the evaluated setting, future work should investigate adaptive threshold selection, calibration-aware routing criteria, and more principled uncertainty measures for defining the ambiguous region.

Overall, the results indicate that incorporating a simple decision-boundary routing and refinement mechanism into the segmentation pipeline can provide practical reliability gains, particularly in ambiguous boundary regions.

Acknowledgments

The authors declare that no specific funding was received for this work.

Conflict of interest

The authors have no competing interests to declare. No ethical issues are associated with the present study.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, 2015, pp. 234-241.

[2] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 2018.

[3] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. Alvarez, and P. Luo, “SegFormer: Simple and efficient design for semantic segmentation with transformers,” in Advances in Neural Information Processing Systems (NeurIPS), 2021.

[4] B. Cheng, I. Misra, A. Schwing, A. Kirillov, and R. Girdhar, “Masked-attention mask transformer for universal image segmentation,” in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[5] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al., “Segment anything,” in 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023.

[6] S. Landgraf, M. Hillemann, K. Wursthorn, and M. Ulrich, “U-CE: Uncertainty-aware cross-entropy for semantic segmentation,” in ISPRS Annals of the Photogrammetry Remote Sensing and Spatial Information Sciences, 2024, pp. 129-136.

[7] K. Zou, Z. Chen, X. Yuan, X. Shen, M. Wang, and H. Fu, “A review of uncertainty estimation and its application in medical imaging,” *Meta-Radiology*, vol. 1, no. 1, pp. 100003, 2023.

[8] A. Valiuddin, R. J. G. van Sloun, C. Viviers, P. H. N. De With, and F. van der Sommen, “A review of bayesian uncertainty quantification in deep probabilistic image segmentation,” *arXiv*, 2024. [Online]. Available: https://doi:org/10.48550/arXiv.2411.16370. [Accessed August 2, 2026].

[9] K.-C. Kahl, C. T. Lüth, M. Zenk, K. Maier-Hein, and P. F. Jaeger, “ValUES: A framework for systematic validation of uncertainty estimation in semantic segmentation,” in Proceedings of the International Conference on Learning Representations (ICLR), 2024.

[10] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431-3440.

[11] 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, 2017.

[12] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[13] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, et al., “Deep high-resolution representation learning for visual recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3349-3364, 2021.

[14] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021.

[15] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning,” in Proceedings of the 33rd International Conference on Machine Learning, 2016, pp. 1050-1059.

[16] B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and scalable predictive uncertainty estimation using deep ensembles,” in Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 6405-6416.

[17] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On calibration of modern neural networks,” in Proceedings of the 34th International Conference on Machine Learning, 2017, pp. 1321-1330.

[18] A. Mestre, F. Malafarina, J. Fons, M. Albert, M. Gil, and V. Pelechano, “SentimentEnriched AI for toxic speech detection: A case study of political discourses in the valencian parliament,” in Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025), 2025, pp. 555-561.

[19] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, “SpaceNet: A remote sensing dataset and challenge series,”

arXiv, 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1807.01232. [Accessed August 2, 2026].
[20] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark," in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017, pp. 3262-3265.
[21] A. Van Etten, D. Hogan, J. Martinez Manso, J. Shermeyer, N. Weir, and R. Lewis, "The multi-temporal urban development SpaceNet dataset," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 6398-6407..