

ARTICLE

An Uncertainty Quantization-Based Method for Anti-UAV Detection in Infrared Images

Can Wu^{1,2}, Wenyi Tang², Yunbo Rao^{1,2,*}, Yinjie Chen¹, Hui Ding², Shuzhen Zhu³ and Yuanyuan Wang³

¹School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, 610054, China

²State Key Laboratory of Air Traffic Management System, Nanjing, 210000, China

³National Airspace Management Center, Beijing, 100094, China

*Corresponding Author: Yunbo Rao. Email: uestc2008@126.com

Received: 17 October 2024; Accepted: 23 December 2024; Published: 26 March 2025

ABSTRACT: Infrared unmanned aerial vehicle (UAV) target detection presents significant challenges due to the interplay between small targets and complex backgrounds. Traditional methods, while effective in controlled environments, often fail in scenarios involving long-range targets, high noise levels, or intricate backgrounds, highlighting the need for more robust approaches. To address these challenges, we propose a novel three-stage UAV segmentation framework that leverages uncertainty quantification to enhance target saliency. This framework incorporates a Bayesian convolutional neural network capable of generating both segmentation maps and probabilistic uncertainty maps. By utilizing uncertainty predictions, our method refines segmentation outcomes, achieving superior detection accuracy. Notably, this marks the first application of uncertainty modeling within the context of infrared UAV target detection. Experimental evaluations on three publicly available infrared UAV datasets demonstrate the effectiveness of the proposed framework. The results reveal significant improvements in both detection precision and robustness when compared to state-of-the-art deep learning models. Our approach also extends the capabilities of encoder-decoder convolutional neural networks by introducing uncertainty modeling, enabling the network to better handle the challenges posed by small targets and complex environmental conditions. By bridging the gap between theoretical uncertainty modeling and practical detection tasks, our work offers a new perspective on enhancing model interpretability and performance. The codes of this work are available openly at https://github.com/general-learner/UQ_Anti_UAV (accessed on 11 November 2024).

KEYWORDS: Object segmentation; uncertainty quantification; bayesian convolutional neural network

1 Introduction

With the rapid development of UAV technology, it is gradually being integrated into daily life and is widely applied in various fields. Systems with remote control of drones can monitor the location of targets, alert for potential threats, and provide communication in emergency situations [1]. However, the advantages and disadvantages of UAVs need to be carefully scrutinized to assess their practical usability. We must be aware of the potential threats posed by drone intrusions to the public, such as their possible use for explosives, cyber attacks, and unauthorized access to critical facilities. Unauthorized drones pose a danger to civil aircraft and have, on many occasions, led to the disruption of air traffic, causing significant financial losses to airlines. In addition, the ability of drones to collect images without being detected may violate people's right to privacy. Therefore, it is important to monitor the operational status of drones, including their position and trajectory.



Traditional techniques such as radio frequency (RF) [2], radar [3] and acoustic sensors [4] play an important role in anti-UAV detection. However, these sensors perform poorly in localizing long-range targets or in environments with strong noise sources [5]. The challenge with using standard cameras, while providing visual cues to security personnel, is the small size of the drone in the image, which directly impacts detection performance. The rapid development of deep neural networks has led to the widespread use of vision-based models for various detection tasks, which are applicable to a wide range of optical images. Therefore, the use of computer vision algorithms to recognize UAVs has become a key component of anti-UAV systems [6].

Despite notable advancements in UAV detection technologies, existing approaches continue to exhibit significant limitations. Traditional methods, radio frequency, radar, and acoustic sensors, are each constrained by inherent shortcomings [2–4]. RF detection systems are highly dependent on the interception of signals emitted by UAVs, rendering them ineffective against drones operating autonomously or utilizing encrypted communication protocols. Similarly, radar systems, while proficient in detecting objects over long distances, often face challenges in distinguishing UAVs from other small aerial objects, such as birds, due to their limited resolution and a high propensity for false alarms. Acoustic sensors, although capable of identifying drones through distinctive sound signatures, are particularly vulnerable to environmental noise. Conventional optical cameras, though effective for visual detection, exhibit limitations in identifying small, mobile UAVs, particularly at long ranges or under low-light conditions. Infrared (IR) imaging provides a robust alternative by ensuring reliable operation in low-light or nighttime environments, enabling continuous surveillance. Additionally, IR sensors detect the heat signatures of UAVs, facilitating their recognition even in the presence of visual obstructions such as fog, smoke, or adverse weather. These capabilities position IR imaging as a highly effective technology for UAV detection across diverse operational scenarios.

Infrared imaging technology, known for its all-weather capabilities and adaptability, has become a crucial tool in remote anti-UAV monitoring. Infrared-based UAV detection complements other methods, enhancing overall detection performance [7–8]. Infrared target detection can be broadly categorized into general target detection and small target detection [9]. As illustrated in Fig. 1, small target detection differs significantly from general target detection. Infrared small targets are typically imaged from greater distances, with target sizes often smaller than 30×30 pixels. Moreover, drones are not only compact but also easily camouflaged within complex backgrounds, lacking distinctive color or texture features, which poses significant detection challenges. Current single-frame infrared UAV small target detection methods generally fall into two categories: model-driven and data-driven approaches.

Model-based methods for infrared small target detection utilize prior knowledge of targets, backgrounds, or imaging features to construct detection models, typically categorized into filtering techniques, human visual system-inspired approaches, and low-rank models [10–12]. Despite their ability to handle complex backgrounds, these methods are hindered by high computational costs, sensitivity to parameter adjustments, and significant manual tuning requirements. Their performance is often compromised by noise and background clutter, limiting robustness, while the reliance on hyperparameter optimization further reduces generalization across diverse scenarios. With advancements in deep learning and the availability of infrared small target datasets [13], data-driven methods for infrared small target detection have rapidly evolved. Convolutional neural networks (CNNs), known for extracting deep semantic features, have achieved significant success in general target detection. However, the unique characteristics of infrared small targets, such as their limited size and shape, necessitate specialized architectures. To address this, researchers have developed networks tailored for this task. Li et al. [14] introduced a dense nested attention network (DNA Net) to mitigate deep feature loss by enhancing and integrating contextual information. Liu et al. [15] employed a transformer-based self-attention mechanism to capture large-scale feature correlations and distinguish targets from backgrounds. Similarly, Qi et al. [16] proposed a hybrid architecture

combining transformers and CNNs, leveraging a dual-branch structure to capture fine-grained target details while suppressing background interference. These innovations significantly improve detection performance and robustness.

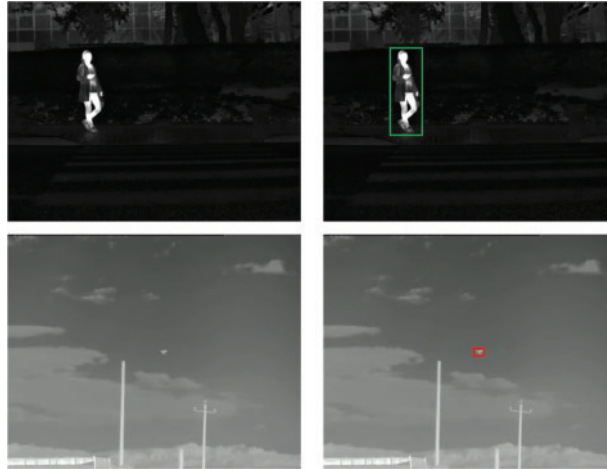


Figure 1: Visual differences between general infrared target detection and unmanned aerial vehicle small target detection

Despite notable progress, existing methods struggle to address performance degradation caused by the small size of targets. To mitigate this, some researchers have reformulated drone object detection as a semantic segmentation task rather than a traditional object detection problem. For instance, Dai et al. [17] proposed an attention-based local contrast network to capture long-range background interactions and implemented infrared small target segmentation using a cross-layer fusion module. However, in complex backgrounds, factors such as background fluctuations, signal noise, and spatial uncertainty of targets complicate detection. To address these challenges, Qu et al. [18] introduced a local entropy operator to quantify image complexity and suppress cloud-like backgrounds by analyzing target-background differences. Nevertheless, this approach overlooks interactions within and across component signals, limiting its effectiveness in processing highly complex scenes.

To develop a single-frame infrared drone target detection method capable of handling complex scenes, we incorporate uncertainty quantification to enhance target saliency by analyzing the uncertainties of targets and backgrounds. Uncertainty is typically classified into data (aleatoric) uncertainty, caused by inherent observational noise such as that from radar or cameras, and model (epistemic) uncertainty, arising from limitations in the model or insufficient data understanding [19]. While aleatoric uncertainty is unavoidable, epistemic uncertainty can be mitigated by acquiring additional data or improving the model. In fact, Uncertainty quantification has shown significant success in fields such as medical imaging [20] and autonomous driving [21], where it guides network learning and supports decision-making based on the reliability of predictions [22]. Despite its potential, uncertainty quantification has received limited attention in infrared drone target detection. Addressing uncertainties related to image blurriness and target similarity in infrared drone detection can significantly enhance model performance, particularly for small target detection in complex scenarios.

The data ambiguity arising from the overlap of drones with trees, buildings, and other objects, along with significant noise in small-target datasets, is collectively referred to as data uncertainty. This uncertainty introduces penalizing factors during model training, contributing to cognitive uncertainty in the learning

process of model. To address this challenge, we propose a novel three-stage drone segmentation framework that leverages Bayesian networks to quantify uncertainty, guiding the segmentation process to focus on high-uncertainty regions and optimizing the model through uncertainty learning. Traditional segmentation models perform well when targets are clearly distinguishable from the background but struggle with complex backgrounds and visually similar targets. The proposed framework addresses this limitation by incorporating uncertainty into the segmentation process. In the first stage, rough predictions and uncertainty estimates are generated for image regions, followed by uncertainty map creation through quantification. In the final stage, the segmentation network prioritizes high-uncertainty areas to refine target detection. This framework is specifically designed to improve segmentation performance in challenging scenarios. The key contributions of this work are outlined as follows:

1. We propose a novel three-stage framework for UAV segmentation that integrates uncertainty quantification to enhance segmentation accuracy and robustness. Unlike traditional encoder-decoder convolutional neural networks (CNNs), our approach extends the architecture into a Bayesian convolutional neural network (BCNN), enabling the simultaneous generation of segmentation maps and probabilistic uncertainty maps. This integration provides a more comprehensive understanding of the confidence of model in its predictions, particularly for infrared UAV scenarios.

2. Uncertainty modeling is applied to infrared UAV target detection for the first time, addressing segmentation challenges posed by complex backgrounds and low-contrast targets. The framework leverages uncertainty visualization to identify and rectify potential segmentation errors, while the generated uncertainty maps serve as auxiliary guidance for the segmentation network, leading to improved segmentation performance and robustness under diverse environmental conditions.

3. Extensive experiments on three public infrared UAV datasets demonstrate the efficacy of the proposed framework. The results show that incorporating uncertainty modeling not only enhances the segmentation accuracy across various segmentation models but also improves the generalization and reliability of the network, making it better suited for real-world UAV detection tasks.

The remainder of this article is organized as follows: [Section 2](#) briefly reviews the research methods related to object segmentation and uncertainty quantification. [Section 3](#) provides a detailed introduction to the proposed three-stage unmanned aerial vehicle target segmentation framework, including methods for uncertainty quantification. [Section 4](#) analyzes the experimental results from two single-frame infrared drone image datasets. Finally, [Section 5](#) summarizes the findings and discusses potential future directions for this research.

2 Related Work

2.1 Object Segmentation

Image segmentation, as a pixel-level classification task, focuses on dense image prediction. With the rapid advancement of Convolutional Neural Networks (CNNs), segmentation methods have evolved from early fully convolutional networks [23] and encoder-decoder frameworks [24] to approaches utilizing dilated or atrous convolutions [25]. Infrared small target detection networks like NUDT-SIRST [14] and MSFFNet [26] incorporate attention mechanisms to enhance spatial and semantic information, improving detection of shallow targets. However, these CNN-based methods are limited by their local receptive fields, reducing their ability to model long-range dependencies in large-scale images. To address this, Transformer-based approaches [15–16] have been introduced, utilizing hierarchical self-attention mechanisms to capture correlations between distant image features effectively. This network has significantly improved detection performance compared to previous detection methods. However, most of these deep learning models focus

on the one-to-one mapping from images to segmentation labels. In reality, the inherent uncertainty and the impact of low-contrast drone targets or similar targets in weak sampling environments can lead to ambiguity in infrared drone target segmentation tasks.

Unlike previous work, we have designed a novel three-stage drone segmentation framework that extends the encoder decoder convolutional neural network to a Bayesian convolutional neural network to quantify the uncertainty information in the image. We use this information to guide the segmentation network to focus on highly uncertain regions and optimize them by learning uncertainty auxiliary models.

2.2 Uncertainty Quantification

Existing deep learning models perform well on various tasks, but in point estimation neural networks, the probability obtained by the softmax layer is often mistakenly interpreted as confidence, resulting in the model generating unreasonably high confidence for points far from the training data [27]. In addition, there is a lack of attention to uncertainty issues in infrared drone target detection tasks. Quantifying the blurriness in infrared drone images and the uncertainty caused by drone like targets can provide more accurate adversarial decisions [28]. In fact, since bayesian neural networks are composed of probability distributions of parameters instead of parameter point estimation, they can quantify the cognitive uncertainty of neural networks [29]. However, in practice, inferring the posterior distribution of weights requires marginalization or integration across the entire parameter space, making it extremely difficult to perform precise inference on bayesian neural networks. Instead, approximate methods based on Monte Carlo sampling and variational inference are commonly used. One of the most popular and simple methods is Monte Carlo dropout, which activates the dropout operation during both the training and testing phases of the model, effectively using different sets of random weights for each network evaluation to quantify uncertainty by evaluating any differences in the results [22]. Dropout is a commonly used method to avoid overfitting of neural networks [30]. In [27], the author proves that using dropout during neural network training can approximate bayesian inference without changing the model architecture and without incurring heavy computational costs. Based on this principle, Monte Carlo dropout technology uses dropout to train the model and keep it active during the inference phase, as described in Fig. 2. Perform multiple forward propagation on a given input query, with each forward propagation randomly using a different dropout mask to generate different predictions. The above process is similar to bayesian neural networks, which can obtain predicted distributions.

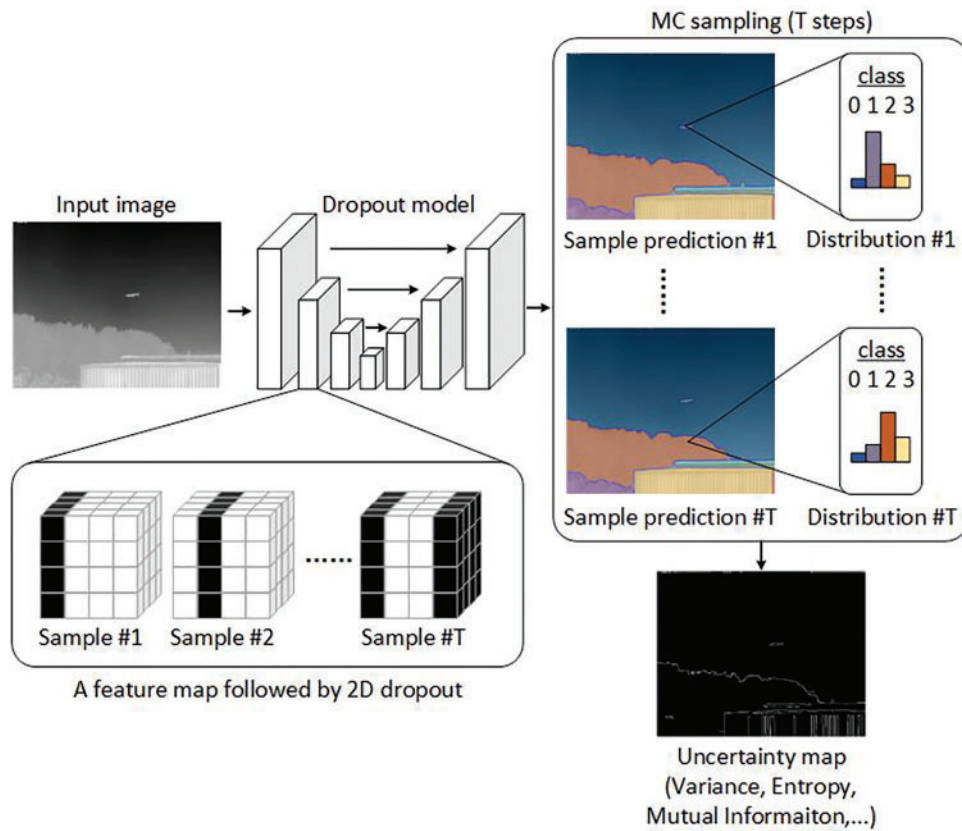


Figure 2: Description of Monte Carlo dropout paradigm

3 Proposed Methods

3.1 Overview of the Proposed Methods

The method proposed is divided into three steps for segmenting unmanned aerial vehicles (as shown in Fig. 3):

Monte Carlo dropout for obtaining segmentation samples: In the first step, the encoder based DNN is combined with Monte Carlo dropout for semantic segmentation training, and Monte Carlo dropout is activated during the test phase to obtain multiple segmentation samples. In addition, averaging the multiple segmented samples obtained can yield the initial segmentation results.

Uncertainty quantification: In the second step, uncertainty quantification is performed on the segmented samples to obtain a global heatmap containing uncertainty values for each pixel, and arbitrary uncertainty maps and cognitive uncertainty maps are generated separately.

Global enhancement: In the third step, a deep neural network for global segmentation is trained by combining the uncertainty map from the previous step with the global feature map to obtain the final segmentation result. Using uncertain information can make the network focus on high uncertainty areas, improve overall segmentation accuracy, and ultimately enhance the quality of semantic segmentation.

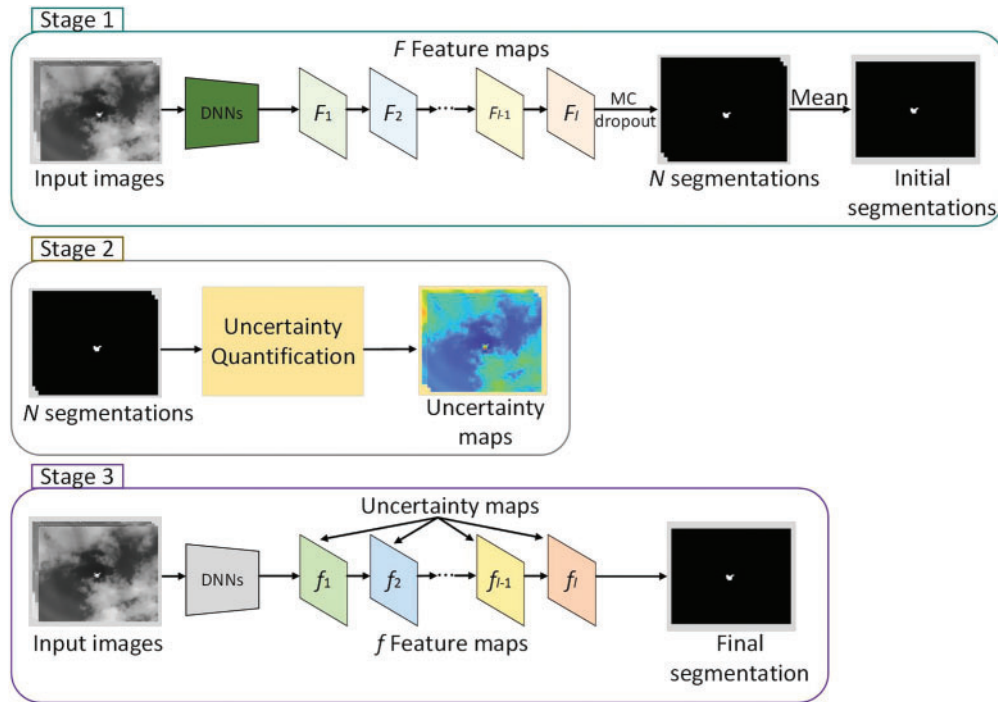


Figure 3: Overview of the proposed network architecture. The infrared UAV image is first processed by an encoder-based deep neural network to generate multiple (N) segmentation samples via Monte Carlo dropout. Uncertainty quantification is then applied to produce uncertainty maps, which are fed back into the network to guide learning and refine the final segmentation output

3.2 Uncertainty Quantification

Bayesian methods are widely used for uncertainty quantification by assigning a prior distribution to neural network parameters and computing the posterior distribution based on training data [31]. For a dataset sample $X = x_1, \dots, x_N$ and $Y = y_1, \dots, y_N$, the random output of bayesian network is represented as $f^W(x)$, with the likelihood $p(y|f^w(x))$. Bayesian inference calculates the posterior distribution $p(\omega|X, Y)$ of the weights ω given a prior $p(\omega)$. The posterior distribution for the model parameters is provided in Eq. (1).

$$p(\omega | X, Y) = \frac{p(Y | X, \omega)p(\omega)}{p(Y | X)} \quad (1)$$

For a given test sample x^* , Eq. (2) presents the posterior distribution of the parameters along with the predictive distribution derived by marginalizing over the model parameters.

$$p(y^* | x^*, X, Y) = \int p(y^* | x^*, \omega) p(\omega | X, Y) d\omega \quad (2)$$

For the input infrared uav image, the model derived through the inference of Eq. (2) is capable of performing a weighted average of the posterior distribution for each pixel in the image, a process commonly referred to as Bayesian Model Averaging (BMA). Nevertheless, the underlying rationale for Eq. (2) stems from the inherent challenges associated with the analytical computation of the parameter posterior distributions.

$$p(Y | X) = \int_{\omega} p(Y | X, \omega)p(\omega) d\omega \quad (3)$$

Eq. (3) represents the likelihood term in Eq. (1), which necessitates marginalizing ω by integrating over the entire weight space. However, given that deep neural networks often contain millions of parameters, performing such comprehensive integration is computationally infeasible.

Due to the large computational complexity of bayesian methods, it is impractical to directly quantify model uncertainty. To address this, Monte Carlo dropout has become a widely recognized statistical approach for approximating Bayesian models to estimate uncertainty in predictions [27]. Building upon this, Kwon et al. [28] developed a method tailored for classification tasks, which decomposes the variation in predicted probabilities into components representing aleatoric and epistemic uncertainties. In this framework, $G = \{(x_i, y_i)\}_{i=1}^N$ denotes a dataset of independent and identically distributed random variables, where $x_i \in R$ and $y_i \in \{e_1, \dots, e_J\}$ correspond to the i -th input and output, respectively. Here, N represents the sample size, J indicates the total number of classes, e_j is a one-hot encoded label vector, and ω signifies the neural network's parameters learned during training.

$$\begin{aligned} \text{Var}_{p(y|x,G)}(y) &= \int_{\Omega} \text{Var}_{p(y|x,G)}(y) p(\omega | G) d\omega + \\ &\int_{\Omega} \{E_{p(y|x,\omega)}(y) - E_{p(y|x,G)}(y)\}^{\otimes 2} p(\omega | G) d\omega \end{aligned} \quad (4)$$

In Eq. (4), the first term of the summation reflects arbitrary uncertainty, capturing the inherent randomness in the output, while the second term denotes cognitive uncertainty, representing the variability in the input model data.

To integrate uncertainty quantification into infrared UAV image detection model, we employ a bayesian convolutional neural network (BCNN) as the core framework. The BCNN extends conventional convolutional neural networks by introducing the Monte Carlo (MC) dropout mechanism, treating the model parameters as distributions rather than fixed values. In the encoder stage, feature extraction is performed using convolutional layers interspersed with dropout layers, which maintain a fixed dropout probability of 15% during both training phases. This enables the model to approximate bayesian inference by performing N times forward passes over the same input image, generating N predictions of distribution. The N predictions are then utilized to compute uncertainty maps [28], distinguishing between aleatoric uncertainty, which captures inherent noise in the data, and epistemic uncertainty, which reflects limitations in the knowledge of model due to finite training data. Specifically, the uncertainty maps are incorporated into the global enhancement phase as auxiliary feature maps. This is achieved by the feature fusion in uncertainty guiding encoder structure. By embedding this additional information, the encoder is guided to focus on regions with high uncertainty, improving the segmentation accuracy in challenging scenarios, such as low-contrast or occluded regions in infrared UAV images. This integration not only enhances the interpretability and robustness of the segmentation results but also preserves the modularity of the architecture, allowing the method to be easily adapted to other encoder-decoder-based networks.

3.3 Uncertainty Guiding Encoder

For conventional image segmentation, the network treats every pixel in the image equally. In fact, images have rich contextual information, and the difficulty of segmentation varies greatly between different regions. In infrared uav segmentation tasks, drones with complex background areas are often difficult to segment. In Section 3.2, we extended the deep neural network to a bayesian network using Monte Carlo dropout technique, sampled its output, and obtained the uncertainty maps using uncertainty quantization method. In the uncertainty map, regions with higher values indicate that the network has a high degree of uncertainty in the image region, corresponding to greater segmentation difficulty.

To enable the network learn the varying difficulty of segmenting different regions, we propose an uncertainty guiding encoder structure, as illustrated in Fig. 4. This innovative framework integrates uncertainty information into the feature extraction process of deep neural networks, allowing the segmentation model to dynamically focus on regions with high uncertainty. The structure begins by transforming the uncertainty map, derived through Monte Carlo dropout-based bayesian inference, into feature blocks using a series of convolutional and activation layers. These feature blocks are then processed using max pooling and average pooling to retain both the most prominent uncertainty features and the overall uncertainty distribution across the image. The pooled outputs are concatenated along the channel dimension and further refined through convolutional operations to generate global uncertainty features. These global features are subsequently fused with specific feature maps extracted from the encoder. By combining the global uncertainty features with the outputs of encoder, the network is empowered to simultaneously prioritize uncertain regions at both local and global scales. This novel approach not only enhances segmentation accuracy in challenging scenarios, such as occluded or low-contrast regions, but also preserves the modularity and adaptability of the network, making it a flexible component for a variety of encoder-decoder architectures. Integrating uncertainty directly into the encoding process, rather than treating it as a post-processing step, underscores the innovative nature of this design and its potential to significantly improve segmentation robustness and interpretability.

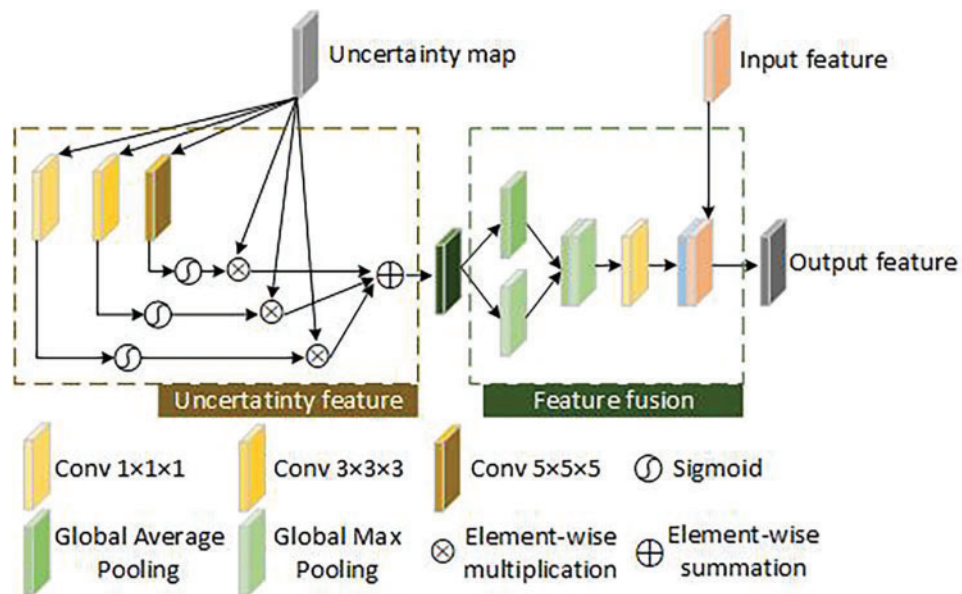


Figure 4: An uncertainty guiding encoder structure is proposed, consisting of an uncertainty feature extraction module and an uncertainty feature fusion module. It can adapt to various encoder based network structures

4 Experimental Analysis and Results

In addition, this paper conducts comparative experiments on multiple encoder-based segmentation networks to demonstrate the importance of introducing uncertainty learning and a priori constraints in the infrared UAV segmentation task.

4.1 Dataset Analysis

1) **Anti-UAV Test-dev Dataset [6]:** To date, two versions of the dataset have been released. The former contains 100 high-quality infrared and RGB video sequences, and the UAVs in the dataset are characterized by multi-scale and cross-scene. Among them, the UAV images have complex backgrounds, including clouds and urban buildings. The latter version discards the RGB video sequences and expands the infrared data. The second version of the dataset contains more complex scenes, including ocean, forest, and mountain scenes. In addition, small objects and UAV-like targets often appear in the images, making it easy for the UAV to be submerged in the background.

2) **Infrared Dim-Small Aircraft Detection Dataset [32]:** This dataset aims to address the lack of authenticity in simulation data and the scarcity of measured data samples in the field of infrared target detection and recognition. It is designed for low altitude weak aircraft target detection applications and integrates a set of test datasets using one or more fixed wing unmanned aerial vehicle targets as detection objects through methods such as shooting and data processing in an off-site environment. The dataset covers various scenarios such as the sky and the ground.

3) **Multi-Sensor Drone Detection Dataset [33]:** This dataset focuses on the relatively understudied challenge of UAV detection using thermal infrared cameras. It includes data captured by thermal infrared cameras, visible light cameras, and microphones. Notably, the dataset provides the infrared UAV target image sources essential for this research, as well as data on other aerial objects that could be misclassified as UAVs, such as birds, airplanes, and helicopters.

The above three datasets are all used for frame extraction of video sequences, with a frame extraction interval of 10. Additionally, data augmentation was applied to all three datasets using three methods: cropping, concatenation, and rotation. The total number of images used in the experiment is 15,124 images from the Anti-UAV Test-dev Dataset, 10,725 images from the Multi-sensor Drone Detection Dataset, and 9944 images from the Infrared Dim-small Aircraft Detection Dataset. The movements of uavs in the dataset exhibit considerable diversity, with their trajectories dispersed throughout the entire field of view. Fig. 5 presents statistical data on the scale distribution of drones across the three datasets, calculated from the width and height of the images. For the training, validation, and test sets, data samples were randomly selected from the raw data to ensure that the scale distributions remained consistent across all sets.

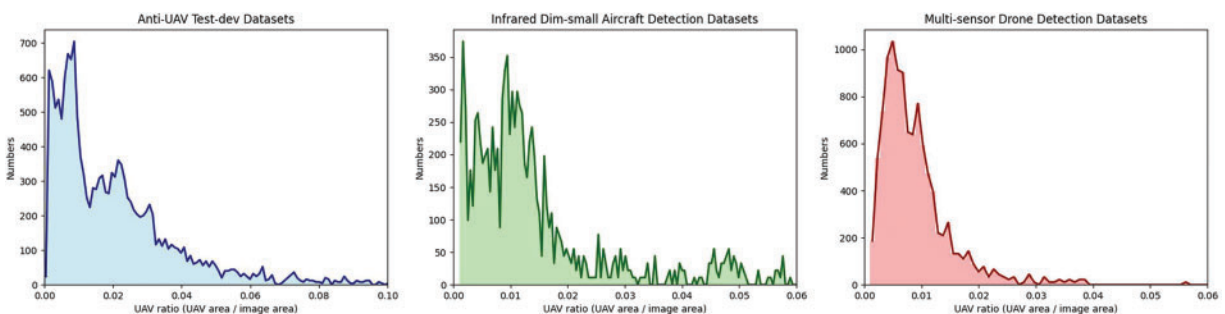


Figure 5: The scale distribution of the three datasets used in the experiment

The datasets used for evaluation are diverse and representative of real-world scenarios, including the Anti-UAV Test-dev Dataset, the Multi-sensor Drone Detection Dataset, and the Infrared Dim-small Aircraft Detection Dataset. These datasets not only cover a wide range of environmental conditions and drone movements but also feature a variety of scales, trajectories, and potential false positives, such as birds

and airplanes. The proposed method is rigorously tested against these models and datasets to ensure its robustness, scalability, and adaptability across different scenarios. The inclusion of models ranging from traditional CNN-based architectures to modern hybrid approaches provides a holistic benchmark.

4.2 Implementation Details

a) Uncertainty network: In the experiment, the uncertainty network is initially trained with a learning rate of 0.01 for a total of 20 epochs. Following the training phase, the model performs multiple segmentation predictions to obtain 50 Monte Carlo samples. From these samples, an uncertainty map and a priority map are derived for each image. The prior map is utilized as a prior label to compute the prior loss, while the uncertainty map serves as a guide for refining the segmentation network during training.

b) Segmentation network: To train the segmentation network, the proposed model is optimized using the SGD optimizer. Across all datasets, the learning rate is set to an initial value of 0.001, with a weight decay parameter of $1e-4$. The training process employs a batch size of 8 and spans 20 epochs. The implementation is conducted within the PyTorch framework and utilizes an NVIDIA A100 GPU for computational efficiency.

c) Comparative methods: We evaluate effectiveness of the proposed method and superiority using a comprehensive set of seven benchmark models, which are widely recognized in semantic segmentation tasks. The models include DeepLabv3 [34], a state-of-the-art architecture known for its atrous spatial pyramid pooling; U-Net [35], a classical encoder-decoder model frequently used in medical and environmental image segmentation; SegNet [24], which emphasizes efficient upsampling for semantic segmentation; BiSeNetv2 [36], designed for real-time applications with a focus on balancing accuracy and speed; HRNet [37], known for maintaining high-resolution representations throughout the network; and DDRNet [38], a dual-resolution network that integrates multi-scale features effectively. Additionally, the Transformer-CNN hybrid model [15] is employed, combining the strengths of convolutional neural networks and transformers for capturing both local and global dependencies.

d) Evaluation metrics: In this paper, the infrared UAV target detection task is formulated as a semantic segmentation problem to effectively delineate drone targets from the background. To rigorously evaluate the performance of the proposed method, the experiments utilize two widely recognized evaluation metrics: Intersection over Union (IoU) and Normalized Intersection over Union (nIoU). IoU measures the overlap between the predicted segmentation and the ground truth, providing a robust assessment of segmentation accuracy. Meanwhile, nIoU normalizes the IoU score to account for imbalances in class distribution, ensuring a fair evaluation of the performance of model, especially in datasets where drone targets occupy significantly smaller portions of the image compared to the background.

4.3 Results and Analysis on the Anti-UAV Test-Dev Dataset

1) Quantitative comparison: Table 1 summarizes the infrared UAV detection accuracy of seven models on the Anti-UAV Test-dev Dataset. Leveraging uncertainty quantification for segmentation training, the proposed uncertainty-guided training model surpasses the base model, with most models achieving optimal IoU and nIoU values in infrared UAV detection. For datasets like Anti-UAV, which are small and realistic, uncertainty-guided training effectively captures deep multi-scale and high-resolution features while focusing on global and local high-uncertainty regions, leading to improved detection performance and reduced false alarm rates. However, models like SegNet and BiSeNetv2, relying on ImageNet-pretrained classification backbones, tend to prioritize objects with ImageNet-like distributions over the Anti-UAV data, limiting their performance. Additionally, uncertainty-guided training provides minimal benefit in improving their detection accuracy. The Transformer-CNN model, while balancing missed detection and false alarm rates

through its self-attention mechanism, faces challenges with uncertainty-guided training due to its higher complexity, which hinders performance gains.

Table 1: Quantitatively compare the IoU and nIoU values of different methods on the Anti-UAV Test-dev Datasets. UQ indicates the use of the uncertainty quantification method proposed in this article to guide network training, with the best results displayed in bold

Model	Optimizer	Methods	Baseline	Baseline+UQ
Deeplabv3	Adam	IoU	57.34	60.58
		nIoU	62.45	64.21
Unet	RMSprop	IoU	59.55	63.89
		nIoU	61.23	65.27
SegNet	SGD	IoU	54.62	55.35
		nIoU	52.55	53.28
BiSeNet v2	Adam	IoU	52.33	53.79
		nIoU	54.66	54.71
HRNet	Adam	IoU	58.64	58.67
		nIoU	57.42	58.58
DDRNet	Adam	IoU	56.53	59.88
		nIoU	57.22	60.33
Transformer-CNN	Adam	IoU	56.21	55.84
		nIoU	58.34	59.22

2) Visual comparison: Fig. 6 illustrates detection results from seven models on the Anti-UAV dataset. The limited feature representation of SegNet and BiSeNet v2 leads to a higher omission rate, visually reflected in fewer detected pixels compared to the reference label. In contrast, the other four models achieve improved detection through uncertainty-guided training, aligning with the quantitative results in Table 1. Notably, the proposed method enhances segmentation performance, particularly for closely connected objects. However, SegNet and BiSeNet v2 still exhibit issues like adhesion or pixel loss in their visual outputs.

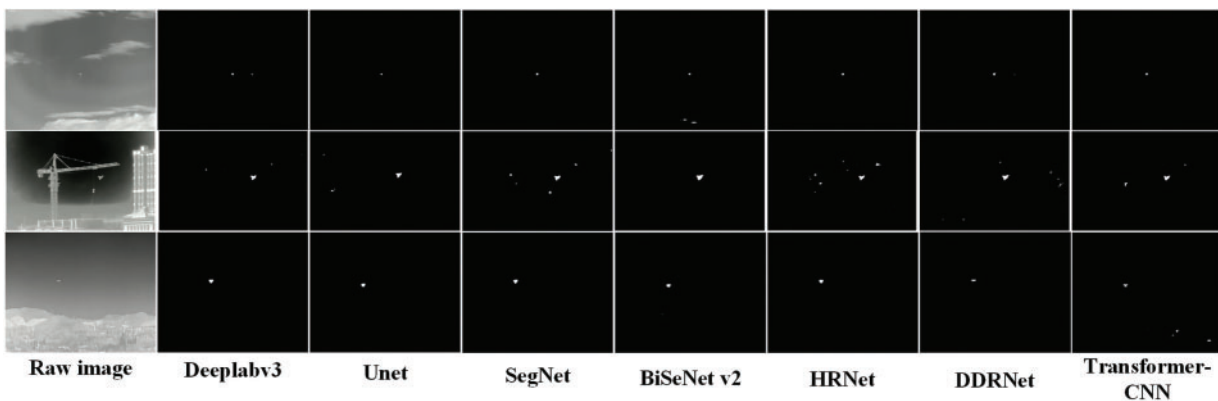


Figure 6: Visualization results of various methods in the Anti-UAV Test-dev Dataset

4.4 Results and Analysis on the Infrared Dim-Small Aircraft Detection Dataset

Table 2 presents the quantitative detection performance on the Infrared Dim-Small Aircraft Detection Dataset, with corresponding visualizations in Fig. 7. The results show a clear performance improvement due to the application of the proposed uncertainty quantification method during model training. By integrating uncertainty maps into the training process, most models outperform the base model in detecting infrared small aircraft, achieving notable gains in IoU and nIoU, which confirms the effectiveness of this approach in enhancing detection accuracy. However, it is noteworthy that not all models benefit equally from uncertainty-guided training, as the performance of DDRNet remains limited on this dataset.

Table 2: Quantitatively compare the IoU and nIoU values of different methods on the Infrared Dim-small Aircraft Detection Dataset. UQ indicates the use of the uncertainty quantification method proposed in this article to guide network training, with the best results displayed in bold

Model	Optimizer	Methods	Baseline	Baseline+UQ
Deeplabv3	Adam	IoU	45.22	47.96
		nIoU	41.21	45.33
Unet	RMSprop	IoU	42.14	47.74
		nIoU	45.60	48.11
SegNet	SGD	IoU	33.57	32.95
		nIoU	34.12	34.31
BiSeNet v2	Adam	IoU	41.35	42.37
		nIoU	40.18	43.55
HRNet	Adam	IoU	40.08	45.20
		nIoU	45.88	49.66
DDRNet	Adam	IoU	47.69	44.55
		nIoU	46.10	44.25
Transformer-CNN	Adam	IoU	46.33	48.90
		nIoU	45.64	47.31

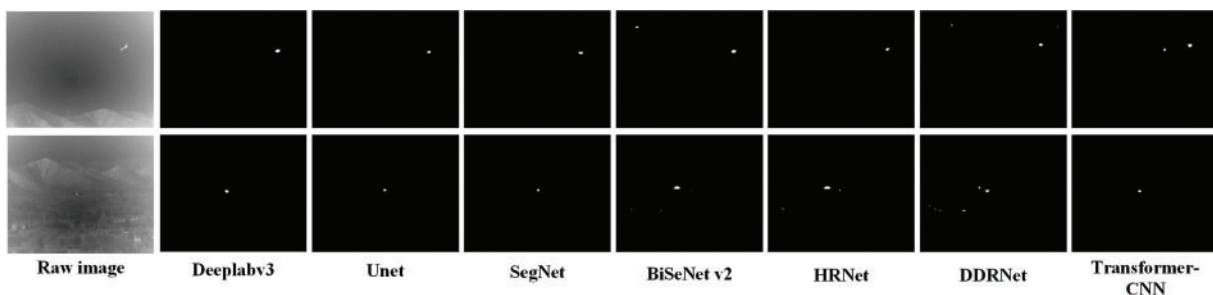


Figure 7: Visual examples of detection results using various methods on the Infrared Dim-Small Aircraft Detection Dataset

The uncertainty guided training method shows distinct advantages on the Infrared Dim-Small Aircraft Detection Dataset with its limited and realistic data. It effectively captures deep multi-scale and high-resolution features while prioritizing regions of high global and local uncertainty. This approach enhances model generalization and robustness, improves detection performance, and significantly lowers the false detection rate, enabling more accurate identification of small aircraft targets in complex infrared image environments.

4.5 Analysis of Data Volume

This section evaluates whether the proposed method overly depends on dataset size using the Multi-sensor Drone Detection Dataset. Two sub-experiments are conducted: first, 5000 high-quality images are manually selected for training, and second, a subset of 500 images is randomly chosen from these to create a smaller sample dataset. The final detection output is defined as the minimum bounding rectangle of the optimal segmentation result, with IoU and nIoU used as quantitative metrics for comparison.

In the previous experiments, a longitudinal comparison between datasets was conducted. As shown in Table 2, the Infrared Dim-Small Aircraft Detection Dataset exhibited lower IoU and nIoU values compared to the Anti-UAV dataset. This is attributed to its lower resolution of 256×256 , which limits feature distinction and contextual representation. Additionally, the reference labels of Anti-UAV dataset align more closely with actual objects, enhancing detection accuracy.

Furthermore, a comparison of data volumes was performed, with the Multi-sensor Drone Detection Dataset having a 10:1 ratio of large to small samples. Despite only a 2% improvement in overall IoU and nIoU, the proposed uncertainty quantification method remains effective. Results from the smaller dataset also demonstrate strong performance, highlighting its practical applicability.

The visualization of detection results for the comparison methods is presented in Fig. 8, with the quantitative detection performance on the Multi-sensor Drone Detection Dataset summarized in Table 3. The proposed method continues to enhance network segmentation performance. However, complex backgrounds, object variations, and limited image quality have led to a decline in detection performance across all models. Notably, SegNet demonstrates particularly poor results on underfitting samples in the small-sample dataset.

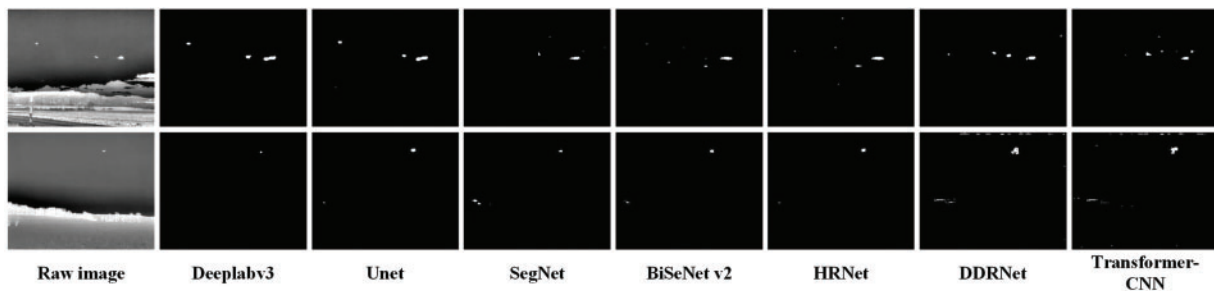


Figure 8: Visualization examples of detection results using various methods on small-scale samples selected from Multi-sensor Drone Detection Dataset

Table 3: Quantitatively compare the IoU and nIoU values of different methods on small-scale samples selected from Multi-sensor Drone Detection Dataset. UQ indicates the use of the uncertainty quantification method proposed in this article to guide network training, with the best results displayed in bold

Model	Optimizer	Methods	Baseline	Baseline+UQ
Deeplabv3	Adam	IoU	38.97	40.63
		nIoU	40.23	41.12
Unet	RMSprop	IoU	40.15	41.26
		nIoU	39.65	41.42
SegNet	SGD	IoU	32.33	31.79
		nIoU	30.22	29.55
BiSeNet v2	Adam	IoU	41.21	41.89
		nIoU	39.63	41.32
HRNet	Adam	IoU	40.81	42.34
		nIoU	40.27	41.35
DDRNet	Adam	IoU	40.87	41.26
		nIoU	39.78	40.26
Transformer-CNN	Adam	IoU	40.32	41.25
		nIoU	39.83	40.20

4.6 Ablation Study

This section details the ablation experiments conducted on three infrared UAV target datasets to assess the effectiveness of the proposed uncertainty-guided training approach. The analysis focuses on evaluating how the number of segmentation slices utilized for uncertainty quantification influences model performance. Additionally, the contribution of the uncertainty guiding encoder to overall detection accuracy is systematically examined. To ensure experimental consistency, U-Net is employed across all experiments and trained using the uncertainty bootstrapping technique, with identical parameter configurations applied throughout.

1)Comparative Analysis of Detection Performance with Varying Numbers of Segmentations Guided Training: [Table 4](#) presents a comparison of IoU and nIoU values for U-Net trained with different numbers of segmentations guided. The results indicate that the optimal performance, in terms of both IoU and nIoU, is consistently achieved across all three datasets when $N = 50$. This analysis highlights the effectiveness of segmentation slice uncertainty bootstrap training in enhancing model performance. Notably, this approach significantly mitigates the loss of fine details in small targets and addresses challenges associated with weak feature representation, thereby improving overall detection accuracy.

2)Evaluation of the Contribution of the Uncertainty Guiding Encoder: The impact of incorporating the Uncertainty Guiding Encoder (UGE) is analyzed by comparing IoU and nIoU values across two configurations, as summarized in [Table 5](#). The results reveal that the inclusion of the UGE module (U-Net + UGE) consistently enhances model performance on all three datasets. This analysis underscores the effectiveness of the proposed uncertainty guiding encoder. It highlights the critical role of this module in improving feature extraction and enabling efficient fusion of uncertainty maps, thereby demonstrating its importance in the uncertainty bootstrap framework.

Table 4: Effects of uncertainty guided training on U-Net using different numbers of segmentation. The best results are shown in bold

No.	Baseline	<i>N segmentations</i>	AUTD		IDADD		MSDD	
			IoU	nIoU	IoU	nIoU	IoU	nIoU
1	U-Net	0	59.55	61.23	42.14	45.60	40.15	39.65
2	U-Net	30	59.66	61.45	43.56	45.62	40.58	41.05
3	U-Net	40	60.12	62.98	45.87	46.87	41.02	41.28
4	U-Net	50	63.89	65.27	47.74	48.11	41.26	41.42
5	U-Net	60	63.07	64.86	46.82	47.86	41.25	41.38

Table 5: Effects of incorporating the uncertainty guiding encoder on uncertainty map processing in U-Net. The best results are shown in bold

No.	Baseline	<i>UGE</i>	AUTD		IDADD		MSDD	
			IoU	nIoU	IoU	nIoU	IoU	nIoU
1	U-Net	✓	63.89	65.27	47.74	48.11	41.26	41.42
2	U-Net		60.03	61.09	42.20	45.67	40.28	39.87

4.7 Generalization Analysis

This section highlights the evaluation of the generalization capability of the proposed method, leveraging the Anti-UAV Test-dev Dataset as the benchmark. To comprehensively assess performance, two prominent models, U2-Net [39] and Swin-Unet [40], were employed, trained using the proposed uncertainty quantification based approach. A subset of 2000 images was selected for test, allowing for a detailed comparison of the generalization abilities of models under varying training data conditions. As shown in Table 6, the experimental results demonstrate a clear performance improvement for models trained on 8000 samples compared to those trained on only 4000 samples. This observation underscores the effectiveness of incorporating uncertainty-guided training, as the additional data enables the models to better capture intricate patterns in the dataset and adapt to diverse scenarios.

Table 6: Quantitative comparison of model performance with varying training data volumes. Bold values indicate the best results. UQ represents uncertainty-guided training

Model	Train/Num	Test/Num	IoU	nIoU
U2-Net	AUTD/4000	AUTD/2000	47.23	48.56
U2-Net + UQ	AUTD/4000	AUTD/2000	51.56	49.47
U2-Net	AUTD/8000	AUTD/2000	52.67	54.12
U2-Net + UQ	AUTD/8000	AUTD/2000	63.56	67.28
Swin-Unet	AUTD/4000	AUTD/2000	49.24	48.57
Swin-Unet + UQ	AUTD/4000	AUTD/2000	52.78	53.64
Swin-Unet	AUTD/8000	AUTD/2000	55.89	56.47
Swin-Unet + UQ	AUTD/8000	AUTD/2000	68.21	66.89

The results also highlight the robustness of the proposed approach in handling the challenges associated with small target sizes, noise, and complex backgrounds, which are common in infrared UAV detection. By leveraging the uncertainty bootstrapping method, the models exhibit enhanced learning of nuanced features, resulting in significantly improved detection accuracy. Metrics IoU and nIoU demonstrate notable gains, further validating the ability of method to generalize across varying data distributions. Moreover, the experiments reveal the scalability of the uncertainty-guided framework, as the performance gap between models trained on 4000 and 8000 samples indicates that the method effectively utilizes additional data to refine predictions and reduce errors.

5 Conclusion

In this study, we propose an innovative method for infrared UAV target detection designed to address the limitations of traditional detection techniques in long-range and noise-prone environments. Specifically, the challenge of identifying small UAV targets, which are often obscured by complex backgrounds and difficult to distinguish using standard imaging systems, is redefined as a semantic segmentation problem. To enhance detection performance, our approach integrates uncertainty quantification, enabling the effective differentiation of target features while accounting for uncertainties present in both the target and background. The proposed method employs a three-stage UAV segmentation framework. First, Bayesian convolutional neural networks are utilized to perform initial image segmentation and generate uncertainty estimates. These outputs are subsequently refined through uncertainty quantification techniques, resulting in more accurate segmentation maps that enhance target saliency. Experimental results demonstrate the superiority of this approach, with significant improvements observed in key metrics such as Intersection over Union (IoU) and normalized IoU (nIoU), especially in challenging infrared UAV detection scenarios. This research highlights the potential of uncertainty-guided frameworks to overcome the inherent difficulties in detecting small, complex targets in infrared imagery. Future work could explore the integration of more advanced network architectures with uncertainty modeling to further improve detection accuracy under increasingly demanding conditions. Additionally, the real-time application of this framework and its adaptation to other detection tasks offer promising avenues for advancing UAV detection systems in dynamic, real-world environments.

Acknowledgment: I sincerely thank all my fellow students in the lab for your academic support and encouragement, which have filled my research journey with warmth and motivation.

Funding Statement: This research was supported by the Science and Technology Project of Sichuan (Grant No. 2024ZHCG0170), the National Key Research and Development Program of China, “Key Technologies for Instrumentation and Control System Program Security Based on Blockchain” (Project No. 2024YFB3311000), States Key Laboratory of Air Traffic Management System (Grant No. SKLATM202202), and the Chengdu Science and Technology Project (Grant No. 2022-YF05-00068-SN).

Author Contributions: The authors confirm contribution to the paper as follows: propose ideas and design experiments: Can Wu, Wenyi Tang; review of relevant literature: Yunbo Rao; data collection: Yinjie Chen, Hui Ding; draft manuscript preparation: Can Wu; analyze experimental data: Shuzhen Zhu, Yuanyuan Wang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: 1. Data openly available in a public repository: The data that support the findings of this study are openly available in Anti-UAV410 at <https://github.com/HwangBo94/Anti-UAV410> (accessed on 22 December 2024). 2. Data openly available in a public repository: The data that support the findings of this study are openly available in DroneDetection Thesis/Drone-detection-dataset: First release at <http://dx.doi.org/10.5281/zenodo.5500576> (accessed on 22 December 2024). 3. Data openly available in a public repository: The data that support the

findings of this study are openly available in a dataset for infrared image dim-small aircraft target detection and tracking under ground/air background at <https://www.scidb.cn/en/detail?dataSetId=720626420933459968> (accessed on 22 December 2024).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Nex F, Remondino F. UAV for 3D mapping applications: a review. *Appl Geomatics*. 2014 Nov;6(1):1–15. doi:10.1007/s12518-013-0120-x.
2. Xiao Y, Zhang X. Micro-UAV detection and identification based on radio frequency signature. In: 2019 6th International Conference on Systems and Informatics (ICSAI); 2019; Shanghai, China. p. 1056–62.
3. Klare J, Biallowons O, Cerutti-Maori D. UAV detection with MIMO radar. In: 2017 18th International Radar Symposium (IRS); 2017; Prague, Czech Republic. p. 1–8.
4. Yang B, Matson ET, Smith AH, Dietz JE, Gallagher JC. UAV detection system with multiple acoustic nodes using machine learning models. In: 2019 Third IEEE International Conference on Robotic Computing (IRC). Naples, Italy; 2019. p. 493–8.
5. Taha B, Shoufan A. Machine learning-based drone detection and classification: state-of-the-art in research. *IEEE Access*. 2019;7:138669–82. doi:10.1109/ACCESS.2019.2942944.
6. Jiang N, Wang K, Peng X, Yu X, Wang Q, Xing J, et al. Anti-UAV: a large-scale benchmark for vision-based UAV tracking. *IEEE Trans Multimedia*. 2023;25:486–500. doi:10.1109/TMM.2021.3128047.
7. Wu X, Hong D, Chanussot J. UIU-Net: U-Net in U-Net for infrared small object detection. *IEEE Trans Image Process*. 2023;32(3):364–376. doi:10.1109/TIP.2022.3228497.
8. Huang B, Chen J, Xu T, Wang Y, Jiang S, Wang Y, et al. SiamSTA: spatio-temporal attention based siamese tracker for tracking UAVs. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW); 2021; Montreal, BC, Canada. p. 1204–12.
9. Zhu H, Ni H, Liu S, Xu G, Deng L. TNLRS: target-aware non-local low-rank modeling with saliency filtering regularization for infrared small target detection. *IEEE Trans Image Process*. 2020;29:9546–58. doi:10.1109/TIP.2020.3028457.
10. Kim S, Yang Y, Lee J, Park Y. Small target detection utilizing robust methods of the human visual system for IRST. *J Infrared Millim Te*. 2006;30(9):994–1011. doi:10.1007/s10762-009-9518-2.
11. Xu Y, Wan M, Zhang X, Wu J, Chen Y, Chen Q, et al. Infrared small target detection based on local contrast-weighted multidirectional derivative. *IEEE Trans Geosci Remote Sens*. 2023;61:1–16. doi:10.1109/TGRS.2023.3244784.
12. Liu T, Yang J, Li B, Xiao C, Sun Y, Wang Y, et al. Nonconvex tensor low-rank approximation for infrared small target detection. *IEEE Trans Geosci Remote Sens*. 2022;60:1–18. doi:10.1109/TGRS.2022.3230051.
13. Wang H, Zhou L, Wang L. Miss detection vs. false alarm: adversarial learning for small object segmentation in infrared images. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019; Seoul, Republic of Korea; p. 8508–17.
14. Li B, Xiao C, Wang L, Wang Y, Lin Z, Li M, et al. Dense nested attention network for infrared small target detection. *IEEE Trans Image Process*. 2023;32(3):1745–58. doi:10.1109/TIP.2022.3199107.
15. Liu F, Gao C, Chen F, Meng D, Zuo W, Gao X. Infrared small and dim target detection with transformer under complex backgrounds. *IEEE Trans Image Process*. 2023;32:5921–32. doi:10.1109/TIP.2023.3326396.
16. Qi M, Liu L, Zhuang S, Liu Y, Li K, Yang Y, et al. FTC-Net: fusion of transformer and CNN features for infrared small target detection. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2022;15:8613–23. doi:10.1109/JSTARS.2022.3210707.
17. Dai Y, Wu Y, Zhou F, Barnard K. Attentional local contrast networks for infrared small target detection. *IEEE Trans Geosci Remote Sens*. 2021 Nov;59(11):9813–24. doi:10.1109/TGRS.2020.3044958.

18. Qu X, Chen H, Peng G. Novel detection method for infrared small targets using weighted information entropy. *J Syst Eng Electronics*. 2012 Dec;23(6):838–42. doi:10.1109/JSEE.2012.00102.
19. Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision? In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017); 2017 Jul 21–26; Honolulu, HI, USA. doi:10.48550/arXiv.1703.04977.
20. Lambert B, Forbes F, Doyle S, Dehaene H, Dojat M. Trustworthy clinical AI solutions: a unified review of uncertainty quantification in Deep Learning models for medical image analysis. *Artif Intell Med*. 2024;150(16):102830. doi:10.1016/j.artmed.2024.102830.
21. Su S, Li Y, He S, Han S, Feng C, Ding C, et al. Uncertainty quantification of collaborative detection for self-driving. In: 2023 IEEE International Conference on Robotics and Automation (ICRA); 2023; London, UK; 2023. p. 5588–94. doi:10.1109/ICRA48891.2023.10160367.
22. Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, et al. A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Inf Fusion*. 2021;76(1):243–97. doi:10.1016/j.inffus.2021.05.008.
23. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017 Apr;39(4):640–51. doi:10.1109/TPAMI.2016.2572683.
24. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017 Dec;39(12):2481–95. doi:10.1109/TPAMI.2016.2644615.
25. Wu H, Zhang J, Huang K, Liang K, Yu Y. FastFCN: rethinking dilated convolution in the backbone for semantic segmentation. 2019. doi:10.48550/arXiv.1903.11816.
26. Tong X, Su S, Wu P, Guo R, Wei J, Zuo Z, et al. MSAFFNet: a multiscale label-supervised attention feature fusion network for infrared small target detection. *IEEE Trans Geosci Remote Sens*. 2023;61:1–16. doi:10.1109/TGRS.2023.3279253.
27. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. 2016. doi:10.48550/arXiv.1506.02142.
28. Kwon Y, Won JH, Kim BJ, Paik MC. Uncertainty quantification using bayesian neural networks in classification: application to biomedical image segmentation. *Computat Stat Data Analy*. 2020;142(2):106816. doi:10.1016/j.csda.2019.106816.
29. Zhao E, Zheng W, Li M, Sun H, Wang J. Infrared small target detection using local component uncertainty measure with consistency assessment. *IEEE Geosci Remote Sens Lett*. 2022;19:1–5. doi:10.1109/LGRS.2022.3221088.
30. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929–58. doi:10.5555/2627435.2670313.
31. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. 2017. doi:10.48550/arXiv.1612.01474.
32. Hui B, Song Z, Fan H, Zhong P, Hu W, Zhang X, et al. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background. *China Scient Data*. 2020;5. doi:10.11922/csdata.2019.0074.z.
33. Svanström F, Alonso-Fernandez F, Englund C. A dataset for multi-sensor drone detection. *Data Brief*. 2021;39(4):107521. doi:10.1016/j.dib.2021.107521.
34. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: The 15th European Conference; 2018 Sep 8–14; Munich, Germany. p. 833–51.
35. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015); 2015; Munich, Germany. p. 493–8.
36. Yu C, Gao C, Wang J, Yu G, Shen C, Sang N. BiSeNet V2: bilateral network with guided aggregation for real-time semantic segmentation. *Int J Comput Vis*. 2021 Sep;129(11):3051–68. doi:10.1007/s11263-021-01515-2.
37. Sun K, Xiao B, Liu D, Wang J. Deep high-resolution representation learning for human pose estimation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019; Long Beach, CA, USA. p. 5686–96.

38. Pan H, Hong Y, Sun W, Jia Y. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes. *IEEE Trans Intell Transp Syst.* 2023 Mar;24(3):3448–60. doi:10.1109/TITS.2022.3228042.
39. Qin X, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M. U2-Net: going deeper with nested U-structure for salient object detection. *Pattern Recognit.* 2020 Oct;106(11):107404. doi:10.1016/j.patcog.2020.107404.
40. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-Unet: unet-like pure transformer for medical image segmentation. In: *Computer Vision-ECCV 2022 Workshops; 2023 Feb; Cham: Springer Nature Switzerland. Vol. 13803.* doi:10.1007/978-3-031-25066-8_9.