2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

A Comparative Study on Segmentation Techniques for Context-Aware Safe Landing of UAVs

Miguel S. Soriano-Garcia[1], Julio De La Torre-Vanegas[1], Diego Mercado-Ravell[2] and Israel Becerra[3]

Abstract— As the use of Unmanned Aerial Vehicles (UAVs) in various tasks within human-inhabited environments becomes increasingly common, critical aspects such as emergency landing need to be addressed. The use of deep learning has become widely adopted to provide context-sensitive solutions, where semantic segmentation has shown promising results. Therefore, this paper presents a comparative study that aims at evaluating several candidate semantic segmentation algorithms, offering a guide for an appropriate selection of the segmentation module in a UAV safe landing task in unstructured urban environments. More specifically, a comparison was made between three prominent segmentation models, U-Net, SegFormer, and MANet, using the Semantic Drone Dataset. First, the models were evaluated using the original 24 classes of the dataset. SegFormer performed slightly better than the other algorithms tested. In a second, more critical experiment, the classes were grouped into six risk levels based on the Specific Operations Risk Assessment (SORA) framework. To address class imbalance and prioritize high-risk categories, a weighted Cross-Entropy loss was employed, assigning higher penalties to misclassifications in critical risk levels. In this setup, MANet achieved the best results, showing its ability to adapt to risk-based classification and capture important features. Later, the three models were optimized to prove that even complex architectures can be enhanced for real-time inference. The optimization included converting the model to TensorRT format and using FP16 precision, which reduced the models' size by at least 30%. Finally, the optimized U-Net, the largest model, was tested on a NVIDIA Jetson Orin Nano platform achieving real-time inference. This shows that heavy models can run on embedded hardware like the Jetson Orin Nano, making them viable for safe, autonomous UAV applications.

I. Introduction

Unmanned Aerial Vehicles (UAVs), combined with deep learning models, are being used in many tasks such as traffic monitoring, object detection and tracking, agricultural surveillance, wildlife monitoring, search and rescue, and more [1]. One of the main uses is in landing, where tasks include finding the best landing zones, landing on moving objects, or delivering packages [2]. As UAVs are used in more situations, it is very important that they can land on their own safely and accurately in complex, unstructured, and dynamic scenarios. This matters even more in emergencies like in system failures, lost GPS, strong winds, sudden battery drops, extreme weather conditions, or even human errors. These issues can force a UAV to abort its mission and land immediately. Hence, autonomous real-time identification of safe landing zones that acknowledge the risk of accidents based on the context of the scene is crucial to minimize the risks of crash or collateral damage [3].

Therefore, it is evident that landing is one of the most critical phases during UAV operation, and the accuracy and success of this step often determine whether the mission is completed or not [4]. Given the high stakes of urban operations, regulatory frameworks such as the Specific Operations Risk Assessment (SORA) have been developed to standardize risk mitigation, ensuring safe integration of UAVs in populated areas [5]. Many recent approaches use deep learning and computer vision to help UAVs find safe landing zones (SLZ) by analyzing terrain characteristics such as slope, texture, and nearby obstacles [6], where simulation tools such as ViVa-SAFELAND [7] can help evaluate proposed approaches in a realistic and safe manner.

There are several ways that safe landing methods leverage deep learning techniques to propose solutions. For example, in scenarios where GPS signals are unreliable due to obstacles such as tall buildings, [8] proposes a cost-effective solution that combines GPS with computer vision techniques based on convolutional neural networks (CNN). Their system, implemented on low-cost hardware, is capable of detecting humans in real time and adjusting landing zones dynamically. Alternatively, [9] focuses specifically on human crowds, introducing an algorithm that generates density maps to estimate human presence and avoids causing harm during emergency landings. By applying tracking techniques such as Kalman Filters, the UAV can monitor and select safe landing spots even when people move unpredictably.

In particular, different semantic segmentation strategies have shown promising results. For example, [10] proposes a method that generates continuous safety maps from UAV images, assigning safety scores to each area rather than using a simple safe/unsafe classification, which is useful during emergency landings. In [11], the authors replace traditional visual markers with a segmentation-based method capable of recognizing surrounding objects and extracting the largest safe region for landing using contour detection, while operating

[1] Center for Research in Mathematics CIMAT AC, campus Zacatecas, Calle Lasec y Andador Galileo Galilei, Manzana 3, Lote 7 Quantum Ciudad del Conocimiento, Zacatecas, 98160, Zacatecas, Mexico. M.S.S-G: miguel.garcia@cimat.mx, J.D-V.: julio.delatorre@cimat.mx

[2] Center for Research and Advanced Studies CINVESTAV-IPN, campus Guadalajara, Av. del Bosque 1145, Zapopan, 45017, Jalisco, Mexico. diego.mercado@cinvestav.mx

[3] Investigadores por México at Center for Research in Mathematics CIMAT AC, calle Jalisco s/n, Valenciana, Guanajuato, 36023, Guanajuato, Mexico. israelb@cimat.mx

efficiently on limited hardware. In [12] urban delivery scenarios are targeted and lightweight segmentation models are used with single RGB images to reduce energy consumption and sensor requirements, identifying safe landing areas without human intervention. Similarly, [13] focuses on enabling autonomous landings in unknown areas, essential for Beyond Visual Line of Sight operations, by applying segmentation techniques that allow UAVs to identify landing spots without relying on operators.

Despite these advances, safety remains a critical concern, particularly during emergency scenarios where system failures or environmental hazards require instantaneous and reliable landing decisions. To address this, various works have proposed methods focused on risk assessment and autonomous emergency landing. For example, [14] introduces the concept of Emergency Landing (EL) as a formal mitigation strategy within urban air operations, highlighting the need for real-time monitoring systems to detect failures and ensure safe landings in the event of critical errors. Similarly, [15] presents a vision-based approach using semantic segmentation and risk mapping, where urban elements captured by an onboard RGB camera are classified and evaluated in terms of potential danger to people and property, offering valuable support during emergency landings.

However, regardless of whether semantic segmentation is applied at the object detection level or in any other alternative way, such as the identification of regions related to the risk level, all these solutions include the semantic segmentation task as an interchangeable module within a broader pipeline, which can be implemented with different segmentation algorithms. Therefore, this paper presents a comparative study that aims at evaluating different state-of-the-art candidate semantic segmentation algorithms, offering a guide for an appropriate selection of the segmentation module in the UAV safe landing context. Furthermore, the resulting implementations of the compared algorithms are then optimized to run on an embedded platform.

As mentioned above, there are a variety of recent semantic segmentation methods to choose from. Although newer models have been developed [16], [17], especially those based on transformers, most of their progress has focused on tasks such as mask classification or unifying segmentation objectives, rather than directly improving pixel-level semantic segmentation performance. With this in mind, the models tested in our comparison range from well-established CNN-based models to more recent Transformer-based architectures that are capable of performing pixel-level classification. More specifically, we tested three models: U-Net [18], which remains the de facto algorithm in many image segmentation tasks, SegFormer [19], which has a more recent Transformer-based architecture, and MANet [19], which includes attention mechanisms to improve performance.

Outlining, this work addresses the challenge of choosing and implementing semantic segmentation systems for UAVs on computationally limited platforms. Building on existing multi-category classification and risk assessment, we focus on two practical aspects often overlooked in the literature:

- A comparative evaluation of selected semantic segmentation models in urban environments.
- Optimized implementation of the models to allow them to run on an embedded platform like a Jetson Nano, demonstrating their feasibility under real-world hardware constraints.

An overview of the proposed pipeline for this work is presented in Fig. 1.

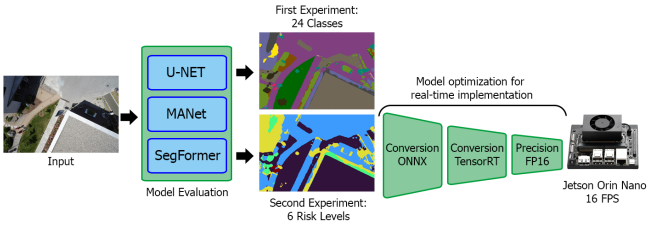

Fig. 1: Overview of the proposed workflow, from aerial image input, through model comparison (U-Net, SegFormer, MANet) and SORA-based risk classification, to model optimization (TensorRT/FP16) and final deployment.

The remainder of this work is organized as follows. Section II presents the comparison methodology, elaborating on the models, datasets, and evaluation metrics used. Section III presents the results of the comparative study, and Section IV describes the optimization of the models to be implemented on embedded hardware. Finally, Section V provides the conclusions and future work.

II. Methods

The development of secure systems for UAVs remains an active research field, particularly in equipping them with vision capabilities and semantic understanding of their environment. To address this challenge, semantic segmentation models have become key tools, enabling UAVs to interpret the context in which they operate. In the following, some of the most widely used semantic segmentation models are tested by applying them specifically to aerial images of complex urban environments. Using two evaluation metrics, we analyze their performance and effectiveness with this type of imagery, allowing us to identify which models are better suited for navigation and recognition tasks in urban contexts. The components on which the comparative study is built are presented below.

A. Semantic Segmentation Models

U-Net The U-Net [18] model is a convolutional neural network architecture originally designed for the segmentation of biomedical images, although its effectiveness

has made it suitable for various computer vision tasks. Its U-shaped structure consists of two main parts: an encoder and a decoder. The encoder uses convolutional blocks to progressively reduce spatial resolution while extracting hierarchical features. The decoder restores spatial resolution, enhancing information through skip connections that link corresponding feature maps from the encoder. These connections help preserve fine details and improve localization accuracy. The key advantage of U-Net is its ability to achieve precise segmentation even with limited training data, due to its symmetric design and multiscale feature integration.

SegFormer SegFormer [19] is a segmentation architecture that combines a Transformer-based encoder (MiT - Mix Transformer) with a lightweight multilayer perceptron (MLP) decoder. The hierarchical design generates multiscale representations through an encoder that produces features at four resolution levels. It employs efficient self-attention blocks and feedforward layers (Mix-FFN) with 3×3 convolutions to maintain positional information without requiring explicit embeddings. The decoder uses a linear fusion scheme with MLPs to effectively combine multiscale features, followed by a final MLP layer to produce the segmentation mask. Key advantages include lower computational cost compared to conventional Transformers, the ability to process variable-sized images without retraining, and strong performance on objects at different scales.

MANet MANet [20] is an architecture designed to improve semantic segmentation by efficiently capturing contextual dependencies through a combination of attention mechanisms. The model introduces a linear complexity kernel attention module, significantly reducing the computational cost of traditional dot-product attention while maintaining high performance. This mechanism allows the network to focus on relevant spatial regions without excessive memory or processing overhead. Additionally, MANet incorporates channel attention to adaptively reweight feature maps, enhancing discriminative feature learning. The backbone of the architecture is based on ResNeXt-101, which extracts multiscale local features, while a hierarchical self-attention mechanism aggregates long-range contextual dependencies across different levels. By integrating multiscale feature fusion and attention modules, MANet effectively balances fine-grained details with global contextual information, leading to improved segmentation accuracy.

B. Aerial-view Urban Dataset

The Semantic Drone Dataset [21] is used for training and testing. This data set is a high-resolution aerial image collection designed to advance autonomous drone navigation and landing safety through a semantic understanding of urban environments. Captured from a nadir perspective at altitudes between 5-30 meters, this data set consists of 400 images with 40169 labeled objects belonging to 24 different classes including person, vegetation, grass, water, dog, car, etc., all with a resolution of 6000×4000 pixels. Each image is accompanied by pixel-level semantic segmentation masks as ground truth.

C. Evaluation metrics

It is important to use multiple evaluation metrics when analyzing the performance of a model, as a high score in one metric does not always mean that the model is overall accurate or reliable. In this work, we use accuracy and Intersection over Union (IoU) as key metrics to evaluate the models more completely.

To correctly compute these metrics, we first define the main classification outcomes. True Positives (TP) occur when the model correctly predicts a positive result; True Negatives (TN) refer to correctly predicted negative cases; False Positives (FP) occur when the model incorrectly labels a negative case as positive; and False Negatives (FN) occur when a positive case is incorrectly predicted as negative.

The accuracy metric measures the overall correctness of a model by calculating the proportion of correctly predicted cases, both positive and negative, out of all predictions. In image segmentation, accuracy reflects how much of the total area has been correctly classified, regardless of class. Although it provides a general indication of performance, it may be less informative in cases with class imbalance. The expression to compute the accuracy Acc is

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}. \tag{1}$$

The Intersection over Union (IoU) metric quantifies how much two sets overlap by comparing the size of their intersection to the size of their union. In image segmentation tasks, IoU is used to assess how well the predicted regions of the model match the ground truth. A higher IoU means a better match between the predicted and actual object boundaries. The metric is calculated using the following expression.

$$IoU = \frac{TP}{TP + FP + FN}. \tag{2}$$

III. Comparative Study

This study evaluates the performance of various models in the semantic segmentation of aerial images from the Semantic Drone Dataset. The objective is to improve autonomous drone navigation and improve landing safety by enabling a semantic understanding of urban environments. In the following, training details and results of the study itself are provided.

A. Training

The training was carried out on a machine equipped with an NVIDIA RTX 4080 GPU with 16 GB of memory, an Intel Core i9 processor at 3.2 GHz and 64 GB of RAM, using PyTorch. For each epoch, the model was trained and then evaluated by computing accuracy and IoU on both training and validation sets.

Each epoch took approximately 80 seconds for both parts. The data set was divided into 306 images for training, 54 for validation, and 40 for testing. All models were trained for 100 epochs using the Adam optimizer with a learning rate of $1e$-3 and a weight decay of $1e$-4. The loss function used for training was Cross Entropy. To improve generalization and reduce overfitting, data augmentation techniques were applied, including horizontal flip, vertical flip, grid distortion, brightness and contrast adjustment, and Gaussian noise. All images and their corresponding segmentation masks were resized to 1056×704 pixels using nearest-neighbor interpolation, ensuring consistency in input dimensions and label alignment.

B. Comparative evaluation

Two experimental studies were conducted. In the first study, the models were trained to segment using all 24 classes available in the dataset. In the second study, the number of classes was redefined as 6 risk levels based on the criteria set by the SORA framework. These levels range from 0 (lowest risk) to 5 (highest risk), reflecting the degree of potential hazard associated with each class in the context of drone landing in urban environments.

Table I presents the performance comparison of the evaluated models U-Net, SegFormer, and MANet on the Semantic Drone Dataset using the 24 original classes. In terms of accuracy, SegFormer achieved the highest score with 0.87, closely followed by U-Net and MANet, both with 0.86. Regarding IoU, SegFormer also led with 0.47, while MANet and U-Net obtained 0.46 and 0.44, respectively. These results suggest that SegFormer offers slightly better segmentation performance in multiclass settings, likely due to its Transformer-based architecture, which improves global context understanding. U-Net performed competitively, highlighting its robustness in dense pixel-wise tasks; however, it is worth mentioning that U-Net has the largest number of parameters (6.6 million). However, MANet, with the same number of parameters as SegFormer (2.9 million), also showed strong performance, indicating a good balance between model complexity and segmentation quality. Overall, while SegFormer showed marginally better results, all three models demonstrated solid performance for semantic segmentation of aerial urban scenes. Fig. 2 shows a visual comparison of the segmentation results obtained by the different models using the 24 original classes from the Semantic Drone Dataset. These qualitative results complement the quantitative metrics presented in Table I, providing a clearer view of how each model handles complex urban scenes. As seen in the figure, SegFormer produces more defined object boundaries and better class separation, particularly in areas with multiple overlapping elements. U-Net also delivers consistent results, but tends to struggle with fine details in smaller objects. MANet offers a balance between both, with reasonably accurate segmentation in most cases. These visual outcomes align with the numerical results, supporting the conclusion that SegFormer slightly outperforms the other models in detailed multi-class segmentation, although all three architectures show strong potential for semantic understanding in drone-based applications.

TABLE I: Comparison of model performance on the Semantic Drone Dataset using all 24 original classes.

Model	Accuracy	IoU	Parameters
U-Net	0.86	0.44	6632135
SegFormer	0.87	0.47	2929943
MANet	0.86	0.46	2929943

TABLE II: Mapping to SORA Risk Levels.

Risk Level	Risk Color	Original Class
Level 0		unlabeled, dirt, grass, gravel, ar-marker
Level 1		paved-area, vegetation
Level 2		rocks, pool, roof, fence, pole
Level 3		water, wall, window, door, bicycle, tree, obstacle
Level 4		dog, car, conflict
Level 5		person

As part of a second experiment, the models were evaluated using a reduced set of six risk levels, derived from the original 24 classes based on SORA. This reclassification, detailed in Table II, aggregated the original classes into risk levels from 0 (lowest risk) to 5 (highest risk). This aggregation, however, revealed a significant class imbalance in the dataset. The new distribution (risk 0: 32.15%, risk 1: 43.8%, risk 2: 9.39%, risk 3: 12.84%, risk 4: 0.75%, risk 5: 1.07%) heavily favors majority classes, while critical high-risk classes 4 and 5 represent less than 2% of the data combined. To address this, we implemented a weighted cross-entropy [22] loss function, which assigns higher penalties to misclassifications in these rare but critical categories.

1) Initial weight assignment: Each class c receives a weight w_c inversely proportional to its relative frequency f_c, i.e.,

$$w_c = \frac{1}{f_c}. \tag{3}$$

2) Normalization: Weights are normalized by their total sum for numerical stability, that is,

$$w_c = \frac{w_c}{\sum_{i=0}^{5} w_i}. \tag{4}$$

The final loss function becomes

$$L = -\sum_{c=0}^{5} w_c \cdot \mathbb{I}_{y=c} \log(\hat{y}_c), \tag{5}$$

where $\hat{y}_c$ is the predicted probability and $\mathbb{I}_{y=c}$ is 1 when the true class is c and 0 otherwise. This approach forces the model to pay more attention to rare but critical cases (risks 4 and 5), improving the overall

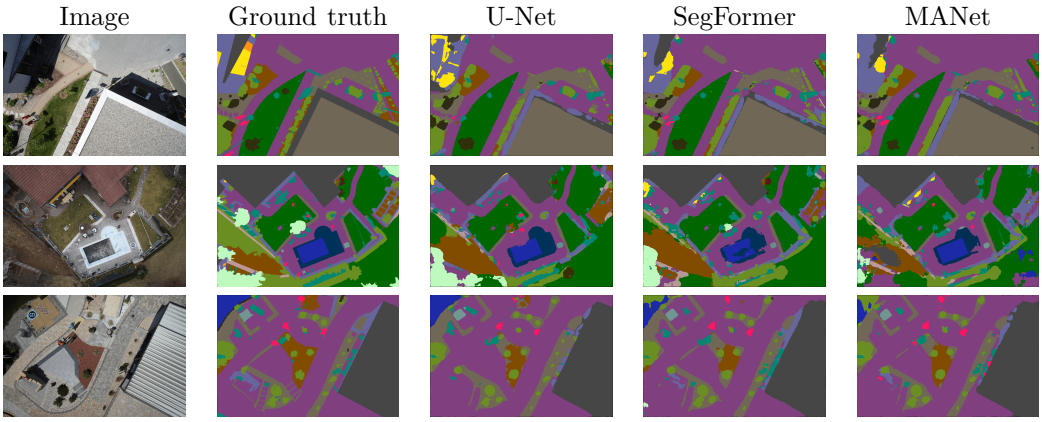

Fig. 2: Visual comparison of segmentation results from different models using the 24 original classes from the Semantic Drone Dataset. Each example shows the ability of the models to identify and distinguish various urban objects relevant to autonomous drone navigation and scene understanding.

predictive capability without compromising accuracy in the majority classes.

The results, shown in Table III, demonstrate consistent performance across all models. U-Net achieved an accuracy of 0.86 and an IoU of 0.57, maintaining its position as the model with the most parameter intensiveness (6.6 million parameters). MANet showed slightly better accuracy (0.87) while matching SegFormer's IoU performance (0.59), with both lightweight models using only 2.9 million parameters. The comparable IoU scores between SegFormer and MANet (0.59) suggest that the reduction in class complexity helped these more efficient models achieve similar segmentation quality to U-Net in distinguishing risk regions, despite U-Net's larger complexity. These results indicate that risk level-based segmentation provides a practical balance between model complexity and performance for safety-critical UAV applications, with MANet emerging as particularly promising due to its combination of the highest accuracy (0.87) and parameter efficiency. Fig. 3 shows the confusion matrix of the MANet model, the best performing model. The matrix demonstrates a remarkable ability to correctly classify different risk categories. The majority classes (0 and 1) show high accuracy rates of 92.56% and 87.37%, respectively, indicating that the model properly handles the most frequent categories in the data set. A notable aspect is the excellent performance in high-risk classes (4 and 5), achieving precision above 98.5%. This result is particularly relevant considering these are the most critical classes during landing missions, but also they are

TABLE III: Model performance on the 6 risk levels.

Model	Accuracy	IoU	Parameters
U-Net	0.86	0.57	6629670
SegFormer	0.86	0.59	2925574
MANet	0.87	0.59	2925574

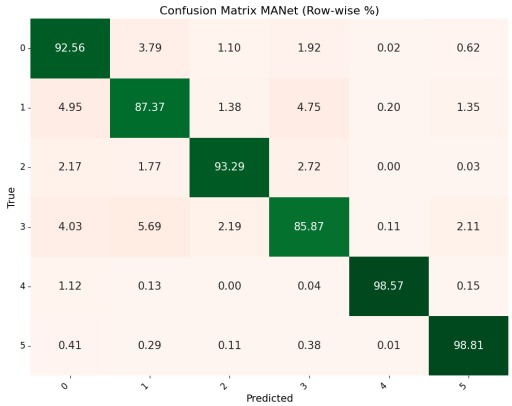

Fig. 3: Row-normalized confusion matrix for MANet model (in %) showing prediction accuracy across the six risk categories. Values on the diagonal represent correct classifications, while off-diagonal elements show misclassifications patterns.

the least represented categories in the dataset (0.75% and 1.07%, respectively), and demonstrates the effectiveness of the implemented weighted cross-entropy technique for handling class imbalance. The minimal presence of confusion between neighbor risk classes indicates that the model has adequately learned the distinctive features of each risk category. This behavior is especially valuable in UAV safety applications, where incorrect classification between very different risk categories could have more severe consequences than errors between similar risk categories. Fig. 4 presents a visual comparison of the segmentation results obtained by the models when using the 6 risk levels. These results demonstrate how effectively each model can identify and differentiate areas according to their assigned risk category. As shown in the figure, the U-Net model produces a clearer and more consistent segmentation of the high- and low-risk zones, in agreement with its higher accuracy and

IoU scores reported in Table III. MANet also shows strong performance, especially in distinguishing medium- to high-risk regions, while SegFormer, although slightly behind in visual clarity, still maintains good separation across most risk levels. This visual evidence supports the findings of the quantitative results and highlights the practical value of risk-based segmentation to improve the safety and decision-making of autonomous drone landings in complex environments.

IV. Model Optimization

In this section, we describe the selected embedded hardware, followed by the optimization procedure.

A. Embedded Hardware

The Jetson Orin Nano is a high-performance computing platform developed by NVIDIA, specifically designed to run artificial intelligence applications, particularly for embedded applications. It delivers significant computing power thanks to its GPU based on the NVIDIA Ampere architecture and its 6-core ARM CPU. It is aimed at addressing tasks such as computer vision, robotics, video analytics, and more, making it ideal for projects that require real-time processing and low power consumption.

Deep learning models can be optimized using TensorRT, a tool from NVIDIA that accelerates inference. TensorRT converts trained models into an optimized format that reduces both execution time and resource usage. During this process, techniques such as layer fusion, lower-precision quantization, and computational graph optimization are applied. The result is a lighter and faster model, well-suited for deployment in real-time on devices like the Jetson Orin Nano, which can be embedded in mobile platforms such as drones.

B. Optimization procedure

All models were optimized to ensure that reducing their size and complexity did not affect their performance. This process significantly improved inference times, making the models fast enough to run in real time on a NVIDIA Jetson Orin Nano while maintaining accuracy close to the original versions. The models were first trained in PyTorch. To optimize them, we converted them to ONNX (Open Neural Network Exchange) as an intermediate step. This allowed us to later transform them into TensorRT format. We also used FP16 (16-bit floating point) precision to shrink the model size without losing accuracy.

The optimized models delivered similar performance as the original versions when evaluated, as shown in Table IV (compared to Table III). The U-Net model achieved the most significant size reduction, shrinking by 39.45% from its original 25.6 to 15.5 MB. The other models also showed substantial reductions, SegFormer decreased from 11.44 to 7.69 MB (32.89% smaller), while MANet went from 11.44 to 7.68 MB (32.98% smaller).

Given that the optimized U-Net remained the largest and most parameter heavy model, we specifically tested its performance on the Jetson Orin Nano board to establish an upper bound on the expected processing times for the optimized models. The results showed a notable improvement in inference speed, while the original U-Net model ran at 5 frames per second (FPS), the optimized version achieved 16 FPS. This threefold increase in processing speed shows that the optimized models are suitable for real-time operation on embedded devices that can be carried by a medium-sized UAV, such as the Jetson Orin Nano.

TABLE IV: Optimized models performance on the 6 risk levels, where the '+' symbol means optimized version.

Model	Accuracy	IoU	Reduction(%)
U-Net+	0.86	0.57	39.45
SegFormer+	0.86	0.59	32.78
MANet+	0.87	0.59	32.86

V. Conclusions and Future Work

This work addressed the challenge of implementing semantic segmentation systems for UAVs on platforms with limited computing resources, which is key to achieving safe and autonomous landings in real-time during emergency situations. Hence, a comparison was made between three segmentation models (U-Net, SegFormer, and MANet) using the Semantic Drone Dataset. The experimental results revealed an interesting change in model performance between the different classification schemes. Although SegFormer demonstrated slightly better performance when segmenting the original 24 classes, the transition to 6 risk levels based on the SORA framework showed that MANet emerged as the most effective model. Contrary to initial expectations about the capacity of the model, MANet achieved the highest accuracy among all architectures, outperforming both U-Net and SegFormer in this scenario focused on safety, despite having fewer parameters.

The practical deployment potential was confirmed by optimizing the three models. In particular, the U-Net was tested on an NVIDIA Jetson Orin Nano board. The optimization process, which involved conversion to TensorRT and FP16 precision reduction, successfully compressed the model size from 25.6 to 15.5 MB while tripling the inference speed from 5 to 16 FPS. This performance enhancement meets the essential requirements for real-time processing during autonomous flights and landings.

In future work, we will evaluate these optimized segmentation models within complete, context-aware autonomous landing strategies. This integration bridges the gap from the pixel-level risk classification to the practical selection of safe landing zones. This exact challenge is addressed in [23], which proposes a complete strategy to select the best safe landing areas from semantic-risk maps. We will further aim to evaluate this combined approach in real-time embedded on a UAV under a

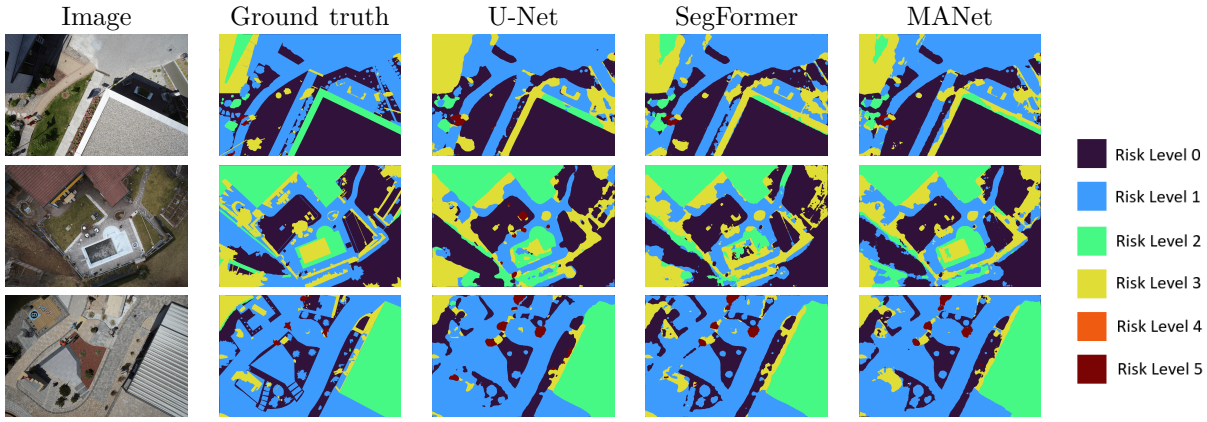

Fig. 4: Visual comparison of segmentation outputs based on the 6 risk levels, presented as a heatmap, where blue regions are safer and red are riskier. The examples illustrate how each model identifies areas with different risk levels, supporting safer context-aware decision-making for autonomous drone landing.

robot-in-the-loop scheme.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research Global ONRG, Award No. N62909-24-1-2001.

References

[1] O. K. Pal, M. S. H. Shovon, M. Mridha, J. Shin, In-depth review of ai-enabled unmanned aerial vehicles: trends, vision, and challenges, Discover Artificial Intelligence 4 (1) (2024) 97.

[2] A. V. R. Katkuri, H. Madan, N. Khatri, A. S. H. Abdul-Qawy, K. S. Patnaik, Autonomous uav navigation using deep learning-based computer vision frameworks: A systematic literature review, Array 23 (2024) 100361.

[3] L. Bartolomei, Y. Kompis, L. Teixeira, M. Chli, Autonomous emergency landing for multicopters using deep reinforcement learning, in: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2022, pp. 3392–3399.

[4] L. Xin, Z. Tang, W. Gai, H. Liu, Vision-based autonomous landing for the uav: A review, Aerospace 9 (11) (2022) 634.

[5] Jarus guidelines on specific operations risk assessment (sora) v2.0. joint authorities for rulemaking of unmanned systems (2024).

[6] M. S. Alam, J. Oluoch, A survey of safe landing zone detection techniques for autonomous unmanned aerial vehicles (uavs), Expert Systems with Applications 179 (2021) 115091.

[7] M. S. Soriano-García, D. A. Mercado-Ravell, Viva-safeland: a new freeware for safe validation of vision-based navigation in aerial vehicles, arXiv preprint arXiv:2503.14719 (2025).

[8] D. Safadinho, J. Ramos, R. Ribeiro, V. Filipe, J. Barroso, A. Pereira, Uav landing using computer vision techniques for human detection, Sensors 20 (3) (2020) 613.

[9] J. González-Trejo, D. Mercado-Ravell, I. Becerra, R. Murrieta-Cid, On the visual-based safe landing of uavs in populated areas: a crucial aspect for urban deployment, IEEE Robotics and Automation Letters 6 (4) (2021) 7901–7908.

[10] S. Abdollahzadeh, P.-L. Proulx, M. S. Allili, J.-F. Lapointe, Safe landing zones detection for uavs using deep regression, in: 2022 19th conference on robots and vision (CRV), IEEE, 2022, pp. 213–218.

[11] H. Y. Putranto, A. N. Irfansyah, M. Attamimi, Identification of safe landing areas with semantic segmentation and contour detection for delivery uav, in: 2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), IEEE, 2022, pp. 254–257.

[12] J. Kinahan, A. F. Smeaton, Image segmentation to identify safe landing zones for unmanned aerial vehicles, CEUR Workshop Proceedings 3105 (2021) 235–247.

[13] H. S. Dhami, D. Ignatyev, A. Tsourdos, Semantic segmentation based mapping systems for the safe and precise landing of flying vehicles, IFAC-PapersOnLine 55 (22) (2022) 310–315.

[14] J. Guérin, K. Delmas, J. Guiochet, Certifying emergency landing for safe urban uav, in: 2021 51st Annual IEEE/IFIP international conference on dependable systems and networks workshops (DSN-W), IEEE, 2021, pp. 55–62.

[15] J. A. Loera-Ponce, D. A. Mercado-Ravell, I. Becerra, L. M. Valentin-Coronado, Risk assessment for uav autonomous landing in urban environments using semantic segmentation, in: Ibero-American Conference on Artificial Intelligence, Springer, 2024, pp. 197–208.

[16] S. Shahabodini, M. Mansoori, F. Bayatmakou, J. Abouei, K. N. Plataniotis, A. Mohammadi, The missing point in vision transformers for universal image segmentation, preprint arXiv:2505.19795 (2025).

[17] J. Jain, J. Li, M. T. Chiu, A. Hassani, N. Orlov, H. Shi, Oneformer: One transformer to rule universal image segmentation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 2989–2998.

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

[19] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, Segformer: Simple and efficient design for semantic segmentation with transformers, Advances in neural information processing systems 34 (2021) 12077–12090.

[20] R. Li, S. Zheng, C. Zhang, C. Duan, J. Su, L. Wang, P. M. Atkinson, Multiattention network for semantic segmentation of fine-resolution remote sensing images, IEEE Transactions on Geoscience and Remote Sensing 60 (2021) 1–13.

[21] D. Ninja, Visualization tools for semantic drone dataset, visited on 2025-07-13 (jul 2025).
URL https://datasetninja.com/semantic-drone

[22] Y. S. Aurelio, G. M. De Almeida, C. L. de Castro, A. P. Braga, Learning from imbalanced data sets with weighted cross-entropy function, Neural processing letters 50 (2) (2019) 1937–1949.

[23] J. de la Torre-Vanegas, M. Soriano-Garcia, I. Becerra, D. Mercado-Ravell, Vision-based risk aware emergency landing for uavs in complex urban environments, arXiv preprint arXiv:2505.20423 (2025).