DOI: 10.59981/ijoa.2025.8155
https://ejournal.brin.go.id/ijoa/

# Comparative Performance of U-Net CNN in Multi-Class Aircraft Segmentation and Classification Using Polygon and Bounding Box Annotations

**Rivilyo Mangolat Rizky Sitanggang[1], Wa Ode Dianita Putri Suaiba Dani[1], Bambang Setiadi[2], and Yanif Dwi Kuntjoro[1]**

[1]Republic of Indonesia Defense University, Bogor, Jawa Barat 16810, Indonesia
[2]National Research and Innovation Agency, Bandung, Jawa Barat 40135, Indonesia
e-mail: bambang.setiadi@brin.go.id

Received: 08-10-2024 Accepted: 21-04-2025 Publishied: 05-10-2025

**Abstract**

Recent advancements in deep learning have revolutionized image processing tasks such as segmentation and classification. This study investigates the performance of U- Net-CNN models in multi-class aircraft segmentation and classification using polygon and bounding box annotations. Military aircraft classification is crucial for defense applications, as it aids in rapid and accurate decision-making during critical missions. This study investigates how these annotation methods affect training time, segmentation accuracy, and classification performance in multi-class segmentation and classification tasks involving military aircraft. The research compares polygon and bounding box methods to evaluate their effectiveness in capturing object details and computational efficiency. While polygon annotations achieved superior precision with a mean test accuracy of 0.987 and lower loss of 0.041, bounding boxes excelled in computational efficiency. Future research should expand datasets and explore additional annotation techniques to further generalize these findings.

***Keywords***: *U-Net CNN, Multi-Class Classification, Polygon Annotations, Bounding Box Annotations, Image Segmentation.*

## Nomenclature

| | | |
|---|---|---|
| $c$ | = | The convolution result |
| $a_{u+i,v+j}$ | = | The component of the input matrix or image |
| $k_{i+1,j+1}$ | = | The element of the convolution kernel |
| $b_q$ | = | Bias |
| $pool$ | = | Pooled feature |
| $Conv$ | = | The convolution result |
| $f(x)$ | = | Activation function |
| $Softmax_i$ | = | The softmax output |
| $e^i$ | = | Exponential |
| $m_t$ | = | First moment (momentum) |
| $v_t$ | = | Second moment (RMSprop-like) |
| $\beta_1$ | = | Exponential decay rate for first moment ($m_t$) |
| $\beta_2$ | = | Exponential decay rate for second moment ($v_t$) |


e-ISSN 2597-7873 (online) 

| | | |
|---|---|---|
| $g_t$ | = | Gradient |
| $\hat{m}_t$ | = | Bias-corrected first moment estimate |
| $\hat{v}_t$ | = | Bias-corrected second moment estimate |
| $\in$ | = | Small constant |
| $n$ | = | Learning rate |
| $w_t$ | = | Weights |
| $Loss$ | = | Loss value |
| $y_i$ | = | True label |
| $\hat{y}_i$ | = | Prediction probability |

## 1. Introduction

Advancements in remote sensing technology and computer vision have significantly enhanced applications in areas such as aerial surveillance, air traffic monitoring, and satellite imagery analysis (Maria et al., 2021; Salakhutdinov, 2015; Sternberg, 1983; Morgan et al., 2020). However, accurately identifying, segmenting, and classifying objects in complex aerial images remains a primary challenge. These challenges arise due to the intricate and often cluttered nature of aerial imagery, which demands precise object recognition techniques to achieve reliable results. Despite progress, there remains a critical gap in defining robust methods for military aircraft classification—a task vital for strategic defense operations.

Deep learning models, particularly U-Net and Convolutional Neural Networks (CNNs), have demonstrated remarkable effectiveness in addressing these challenges (Sun et al., 2019; Girshick et al., 2014). However, there is a lack of comprehensive studies comparing the impact of annotation methods on these models, especially when applied to military aircraft classification. U- Net, with its encoder-decoder architecture, excels in achieving high-precision segmentation, while CNNs are highly effective in object classification tasks following segmentation. Nevertheless, the performance of these models is strongly influenced by the data annotation methods employed during training. Polygon and bounding box annotations are the two primary labeling techniques used in image analysis. Polygon annotations provide detailed boundaries that closely match an object's shape, enabling the capture of intricate visual details. In contrast, bounding box annotations offer a simpler and less precise representation of objects. This study investigates how these annotation methods affect training time, segmentation accuracy, and classification performance in multi-class segmentation and classification tasks involving military aircraft.

This study aims to evaluate the performance of U-Net-CNN models trained using polygon and bounding box annotations by examining their influence on training time, segmentation accuracy, and classification performance. The results will provide valuable insights into how these annotation methods impact the precision and computational efficiency of automated systems, particularly in military applications. By addressing this gap, the study aims to enhance the reliability of deep learning models for real-world defense scenarios and inform the development of efficient image analysis pipelines.

## 2. Methodology

This section outlines the dataset used for training, validation, and testing of the U-Net-CNN model, along with the data preprocessing steps involved. The segmentation dataset comprises over 2,500 satellite images, which have been both augmented and annotated using polygon and bounding box methods. For the classification task, the dataset includes more than 8,000 samples from MTARSI, covering four classes: Attacker, Bomber, Carrier, and Fighter (Ling et al., 2019). These images exhibit variability in lighting, angles, and background complexity—factors that can influence classification performance. Challenges in the sample images include the presence of occlusions, overlapping objects, and inconsistent resolutions, which necessitate robust preprocessing.

The images were treated as static, with each frame representing a distinct and independent observation. Augmentation techniques, including random rotations, scaling, and flipping, were applied to enhance model generalization by introducing diversity in lighting, angles, and structural details. These preprocessing steps ensured balanced and comprehensive data distribution, allowing for meaningful comparisons between the two annotation methods. This setup supports an in-depth investigation into the sequence of operations—first identification through classification, then segmentation—to evaluate U- Net-CNN performance.

## 2.1. Annotations

**Polygon Annotations**

Polygon annotation is a data labeling method in image analysis and computer vision that enables the precise definition of object boundaries. As shown in Figure 1, this method involves marking a series of connected points along the object's contour, forming a polygon that closely follows its actual shape. This allows for the depiction of more complex and detailed shapes compared to the simpler, less precise bounding box annotation (Ling et al., 2019).

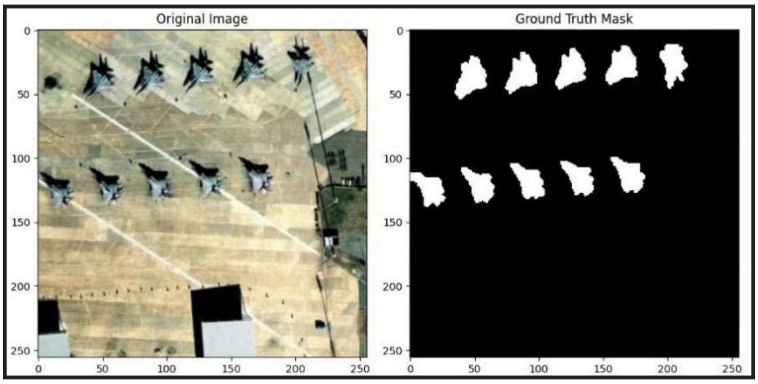


**Figure 1:** Polygon Annotation.

Polygon annotation is widely used in applications that require high-precision segmentation, such as object recognition and shape analysis (Yang et al., 2023). By recording the coordinates of each point, this method allows machine learning models to recognize objects based on intricate shapes, such as curved edges, sharp angles, and protruding parts, and spatial features, including varying textures, gradients, and spatial relationships among closely packed objects. It is particularly beneficial for objects with irregular shapes or those that are closely spaced, as it accurately traces the object's boundaries without capturing irrelevant background areas.

**Bounding Box Annotations**

Bounding box annotation is a widely used data labeling method in image analysis and computer vision for defining the area surrounding an object using a rectangular box (Zheng et al., 2023). As shown in Figure 2, this method involves placing a box around the object and recording the coordinates of the top-left and bottom-right corners. Its simplicity and efficiency in defining objects make bounding box annotation one of the most popular techniques in image analysis.

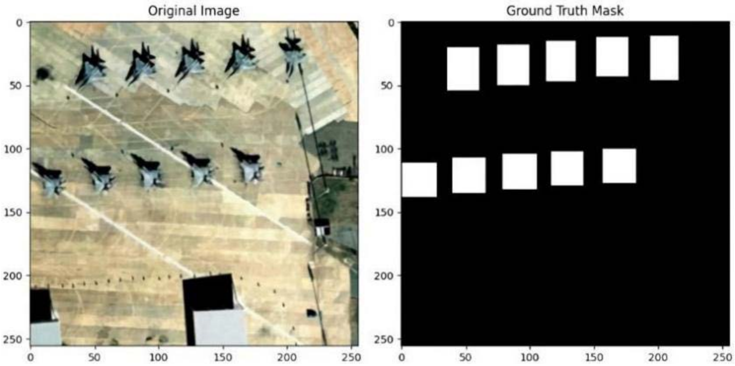


**Figure 2:** Bounding Box Annotation.

## 2.2. U-Net Architecture

Initially developed for biomedical image segmentation, the U-Net architecture as illustrated in Figure 3 is tailored to improve the accuracy of object segmentation in images (Ronneberger et al., 2015). In this study, U-Net performs segmentation using input images of size 256×256×3 and produces segmentation masks of size 256×256×1. The learning rate of 0.001 was chosen for its ability to balance fast convergence and stability during training, ensuring effective optimization while avoiding overshooting minima. The model is configured with the Adam optimizer for adaptive learning and binary cross-entropy as the loss function to handle the segmentation task. To assess segmentation precision, accuracy is used as the evaluation metric. The architecture of U-Net consists of four encoders that extract features from the input image, a bottleneck layer that preserves abstract feature representations, and four decoders that reconstruct the spatial information from the extracted features.

The U-Net-CNN model has been successfully applied in various fields, including object segmentation in medical imaging (e.g., tumor detection) and object recognition in satellite imagery for disaster analysis. Initially, the model identifies objects through classification before proceeding to segmentation, ensuring that all relevant objects are detected and appropriately labeled. In the military context, U-Net-CNN can be employed to segment aircraft from complex backgrounds, such as scenes with overlapping objects, variable lighting, and noisy environments. These challenges arise due to the high variability in aerial imagery, making robust segmentation essential for mission-critical operations. Rapid and accurate decision-making is achieved by providing precise visual outputs, which enable operators to act on detailed and reliable information in high-stakes scenarios.

A confidence threshold of 0.5 is applied in this study to filter out predictions with lower probabilities, ensuring that only outputs exceeding this threshold are considered valid. This threshold prevents the inclusion of ambiguous results, thereby reducing false positives and increasing the reliability of the system for defense applications. By evaluating two labeling methods (polygon and bounding box) and three different batch sizes (2, 4, and 8), this study comprehensively analyzes the trade-offs between segmentation accuracy and computational resource utilization, offering practical insights into the performance of the U-Net-CNN model.

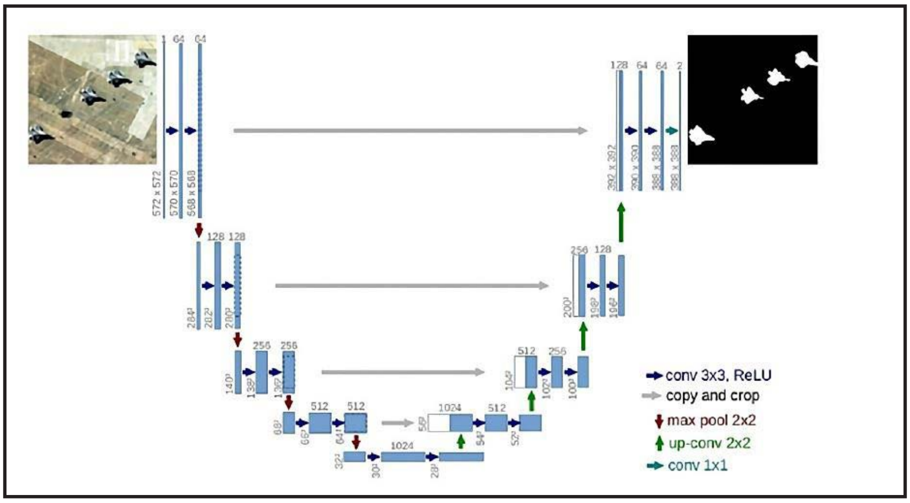


**Figure 3:** U-Net Architecture.

## 2.3. CNN Architecture

The Convolutional Neural Network (CNN) model utilized in this study consists of several key layers, starting with convolutional layers for feature extraction and ending with fully connected layers for final classification (Krizhevsky et al., 2012).

$$c_{i,j} = \left(\sum_{u=0}^{n-1}\sum_{v=0}^{n-1} a_{u+i,v+j} \times k_{i+1,j+1}\right) + b_q \quad (1)$$

$$pool_{x,y} = Max(Conv_{x,y}, Conv_{x+1,y}, Conv_{x,y+1}, Conv_{x+1,y+1}) \quad (2)$$

The convolutional layers process input images of size (256, 256, 3), representing the images' resolution and color channels. A 3 × 3 kernel is utilized in the convolutional layers, as it strikes a balance between computational efficiency and the ability to capture spatial features. The output of a convolutional layer is calculated using Equation (1), while Equation (2) defines the Max Pooling function, where $i$ and $j$ represent the row and column indices, respectively. In these equations, n denotes the kernel height, $c_{i,j}$ is an element in the input matrix, $k_{i,j}$ is an element in the kernel matrix, and $b_q$ is the bias associated with the $q$-th kernel. Following the convolution and pooling stages, the data are flattened and passed through fully connected layers. The first fully connected layer comprises 128 units, using the ReLU activation function (Equation (3)) to capture complex patterns in the data by mapping input values to non-negative outputs. This activation introduces non-linearity, enabling the model to learn intricate patterns. The final output layer contains four units, corresponding to the four target classes (Attacker, Bomber, Carrier, and Fighter), and applies the softmax activation function (Equation (4)) to generate class probabilities, which represent the likelihood of each input belonging to a specific class.

$$f(x) = max(0, x) = \begin{cases} x, & jika\ x \geq 0 \\ 0, & jika\ x < 0 \end{cases} \tag{3}$$

$$Softmax_i = \frac{e^i}{(\sum_{j=1}^{class} e^j)} \tag{4}$$

Let $x$ represent the input image value. The ReLU activation function, denoted as $f(x)$, maps $x$ to a non-negative output, introducing non-linearity into the network. Each convolutional layer is followed by a MaxPooling2D layer with a 2 × 2 pool size, which reduces the data's spatial dimensions while preserving critical features. After passing through the convolution and pooling stages, the data are flattened and connected to the fully connected layers. The first fully connected layer consists of 128 units and utilizes the ReLU activation function to capture complex patterns in the data. This is followed by an output layer with a number of units equal to the target classes, employing the softmax activation function to convert the output into class probabilities.

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \\ \widehat{m_t} &= \frac{m_t}{1 - \beta_1^t} \\ \widehat{v_t} &= \frac{v_t}{1 - \beta_2^t} \\ w_t &= w_{t-1} - n \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon} \end{aligned} \tag{5}$$

$$Loss = -\sum_{i}^{class} Y_i \times Log(\widehat{y}) \tag{6}$$

The model is compiled using the Adam optimizer, which adaptively adjusts the learning rate based on past gradients and their squared values. In Equation (5), $\beta_1$ and $\beta_2$ represent the exponential decay rates for the moving averages of the gradient and the squared gradient, respectively. $\beta_1$ controls how much weight is given to recent gradients, while $\beta_2$ controls the smoothing of the squared gradients to stabilize updates. The learning rate η determines the step size for weight updates and is set to 0.001 in this study, balancing fast convergence with optimization stability. The optimizer improves training efficiency but does not guarantee higher accuracy in a shorter time for all tasks. Categorical cross-entropy is used as the loss function to measure prediction errors, as shown in Equation (6), and accuracy is selected as the evaluation metric to assess the model's performance in classifying the input images.

## 3. Result and Analysis

### 3.1 Training Performance

A detailed analysis was conducted to assess the segmentation and classification outcomes of the U-Net-CNN model, addressing its performance in the context of polygon and bounding box annotation methods. Specifically, the study examined how these annotations influenced training accuracy, loss convergence, and segmentation precision. The dataset used for this study consisted of aircraft images, divided into 70% for training, 20% for validation, and 10% for testing. However, validation metrics, including accuracy and loss curves, have been incorporated into the results to provide evidence of the model's reliability. The U-Net-CNN model was configured with four encoder and decoder layers, a bottleneck layer for feature abstraction, ReLU activation functions, the Adam optimizer with a learning rate of 0.001, and binary cross-entropy as the loss function to optimize segmentation performance.

To investigate the impact of different annotation methods on model performance, the training parameters were set with a batch size of 4 (as illustrated in Figure 4). Loss and accuracy metrics were employed to assess the model's convergence and classification accuracy, and the results were presented through epoch curve graphs to provide a detailed visualization of the training process.

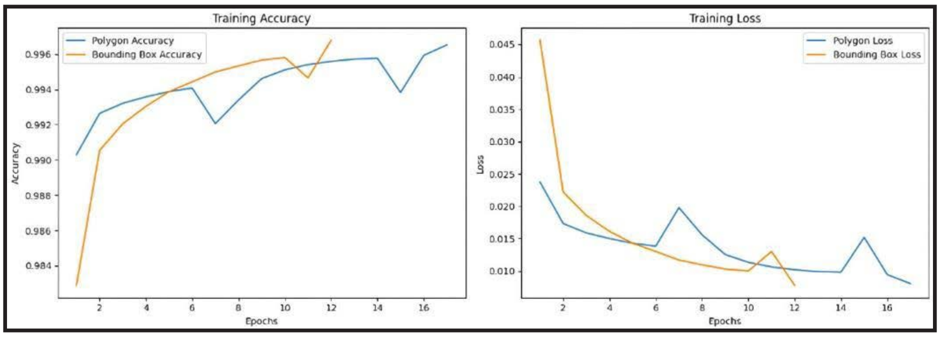


**Figure 4:** Training Performance (batch size: 4).

The experimental results revealed that the model utilizing polygon annotations exhibited faster stabilization of loss and accuracy metrics compared to the model trained with bounding box annotations. The training loss curve (Figure 3-1) indicates fewer fluctuations and a downward trend. The accuracy curve demonstrates consistent improvement, underscoring the effectiveness of polygon annotations in providing detailed boundary information. These results highlight the importance of annotation techniques in influencing learning dynamics and improving model performance, particularly for complex object features. Further validation through statistical analysis or larger datasets is required to generalize these findings.

These findings emphasize the critical role of annotation methods in complex object segmentation and classification tasks. The superior performance of polygon annotations, especially in the context of military applications, suggests that the choice of annotation technique can significantly influence the accuracy and reliability of automated aircraft recognition systems. This study thus provides valuable insights into how detailed labeling can enhance the capabilities of deep learning models in remote sensing and defense related image analysis.

### 3.2 Segmentation Performance

The performance evaluation of segmentation was conducted by comparing the results of the U-Net-CNN model trained with different annotation methods—polygon and bounding box—across various batch sizes (2, 4, and 8) over 100 initial epochs. Training was terminated early if there were no significant changes in the model's performance to optimize computational resources. The metrics used to assess the segmentation performance included test loss, test accuracy, mean Intersection over Union (IoU), and mean Dice Coefficient, providing a comprehensive analysis of the model's accuracy and precision in delineating aircraft within the images.

**Table 1:** Segmentation Performance

| Annotations | Batch Size | Test Loss | Test Accuracy | Mean IOU | Mean Dice Coefficient |
|---|---|---|---|---|---|
| Polygon | 2 | 0.047 | 0.988 | 0.766 | 0.856 |
| | 4 | 0.041 | 0.987 | 0.749 | 0.843 |
| | 8 | 0.039 | 0.987 | 0.732 | 0.829 |
| Bounding Box | 2 | 0.091 | 0.980 | 0.793 | 0.872 |
| | 4 | 0.078 | 0.981 | 0.802 | 0.879 |
| | 8 | 0.069 | 0.977 | 0.754 | 0.842 |

Table 1 presents the segmentation performance metrics for the U-Net-CNN model trained using both polygon and bounding box annotations. Different batch sizes were selected to evaluate how varying computational loads and data grouping affected the performance of each annotation method. Notably, bounding box annotations demonstrated a higher performance in terms of mean IoU and mean Dice Coefficient, achieving values of 0.802 and 0.879, respectively, when trained with a batch size of 4. For polygon annotations with a batch size of 2, the results showed variation compared to other configurations due to the smaller data group leading to more frequent updates during training, which may have affected model convergence. These results suggest that while bounding box annotations may offer an edge in certain segmentation metrics, polygon annotations provide a more reliable and stable performance overall, especially for tasks requiring detailed object boundaries.

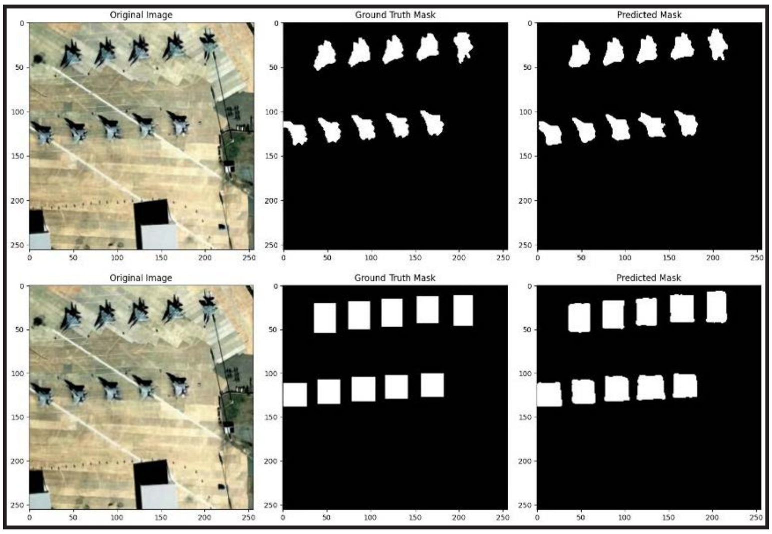


**Figure 5:** Polygon Segmentation Output (Top) and Bounding Box Segmentation Output (Bottom).

The differences in segmentation output are visually represented in Figure 5, which illustrates the segmentation results of polygon annotations (top) and bounding box annotations (bottom). The figure clearly shows that polygon annotations capture the intricate contours of the aircraft more effectively than bounding box annotations, which tend to approximate object boundaries with less precision.

Overall, the results indicate a trade-off between the two annotation methods. Polygon annotations yield superior performance in terms of test loss and test accuracy, demonstrating their effectiveness in producing more precise segmentations. On the other hand, bounding box annotations excel in mean IoU and mean Dice Coefficient metrics, suggesting that they may provide a more generalized representation of object boundaries. This comprehensive evaluation highlights the importance of selecting an appropriate annotation method based on the specific requirements of the segmentation task, particularly in applications where the balance between accuracy and computational efficiency is crucial.

### 3.3 Classification Performance

Satellite imagery is increasingly used to classify military aircraft through Convolutional Neural Networks (CNNs). This study uses Maxar Technologies' satellite images to classify aircraft into four categories: Attacker, Bomber, Carrier, and Fighter. Bounding box annotations, presented first in Table 2, serve as a baseline due to their simplicity and common use in object detection. In contrast, Table 3 highlights the superior performance of polygon annotations, which provide detailed object boundary representations, resulting in higher classification accuracy and precision. This comparison evaluates the strengths of each annotation method for military aircraft classification.

**Table 2:** CNN Classification Performance (Bounding Box).

| Classification Batch Size | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| 8 | 0.361 | 0.13 | 0.361 | 0.191 |
| 16 | 0.826 | 0.844 | 0.826 | 0.829 |
| 32 | 0.603 | 0.651 | 0.603 | 0.55 |
| 64 | 0.607 | 0.654 | 0.607 | 0.555 |

**Table 3:** U-Net CNN Classification Performance (Polygon).

| Classification Batch Size | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| 8 | 0.512 | 0.398 | 0.425 | 0.251 |
| 16 | 0.901 | 0.862 | 0.843 | 0.899 |
| 32 | 0.832 | 0.792 | 0.721 | 0.631 |
| 64 | 0.846 | 0.754 | 0.781 | 0.667 |

The performance of the CNN model for classifying military aircraft was evaluated across various batch sizes. A batch size of 16 consistently demonstrated superior results compared to 8, 32, and 64, suggesting an optimal balance between training stability and model generalization. The U-Net CNN model, using polygon annotations, further enhanced performance, achieving significantly higher accuracy, precision, recall, and F1-score compared to the bounding box approach. This improvement highlights the importance of polygon-based representations in capturing more relevant visual details for accurate classification of military aircraft.

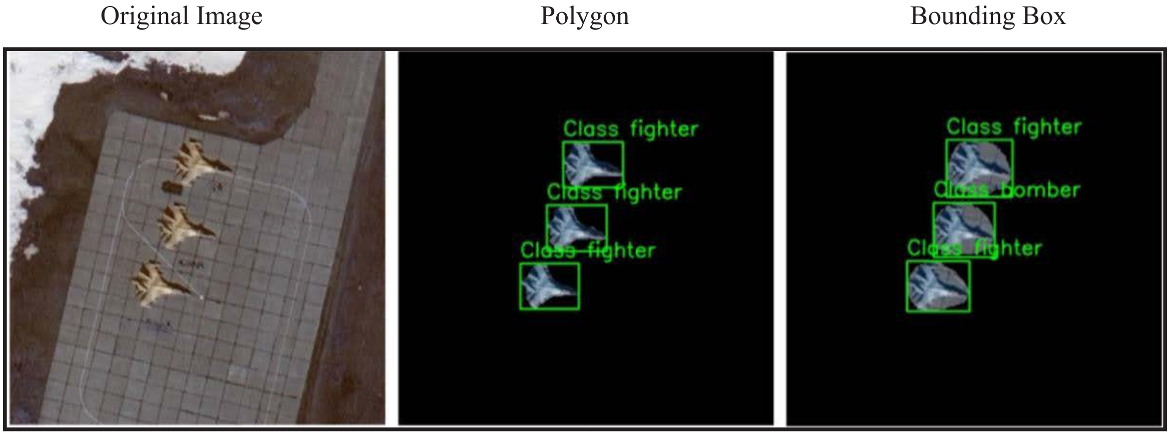


**Figure 6:** Satellite image of three Russian Su-35 (Fighter) landing at Anadyr Airport, northeastern Russia. Satellite image © 2019 Google Earth, Maxar Technologies.

Figure 6 further illustrates the advantage of polygon annotations over bounding boxes in terms of segmentation accuracy. The satellite image depicts three Russian Su-35 fighters landing at Anadyr Airport, and the segmentation results reveal the superior capability of polygon annotations to trace object contours more precisely. This level of detail is particularly beneficial for objects with complex shapes or protruding parts, as polygons can adapt to the exact boundaries of the object. Conversely, bounding boxes often result in less accurate segmentations, especially in areas where the object has irregular shapes, leading to errors such as over-segmentation or under-segmentation.

Overall, these findings highlight the limitations of bounding box annotations, which, while providing a general indication of an object's location, may lack the precision required for accurate classification. The analysis in Tables 3-2 and 3-3 demonstrates that polygon annotations not only improve the segmentation process but also enhance the model's ability to classify military aircraft effectively. This study emphasizes the critical role of choosing the appropriate annotation method, particularly in defense-related applications where precise object recognition is paramount for decision-making

## 4. Conclusions

This study demonstrates that polygon labeling significantly outperforms bounding box labeling when applied in a U-Net-CNN model for aircraft segmentation and classification. The results reveal that polygon labeling provides more precise and detailed segmentation, contributing to enhanced model performance. Although bounding box annotations showed some advantages in metrics such as mean Intersection over Union (IoU) and mean Dice Coefficient, polygon annotations consistently delivered superior results, achie ving a lower test loss and higher test accuracy. Specifically, the U-Net-CNN model utilizing polygon annotations reached an accuracy of 0.901, markedly higher than the 0.826 achieved with bounding box annotations.

These findings underscore the importance of choosing the appropriate labeling method to improve segmentation and classification performance, especially in applications requiring high precision, such as military and surveillance contexts. The superior performance of polygon labeling suggests that it captures more relevant visual details, enabling the model to make more accurate predictions.

However, to further validate and generalize these findings, future research should focus on expanding the dataset to include a wider variety of aircraft types and environmental conditions. Additionally, investigating different labeling techniques and their impact on model generalization across diverse scenarios will be crucial. This research provides a solid foundation for optimizing annotation methods in deep learning models, opening avenues for advancements in remote sensing and object recognition technologies.

## Contributorship Statement

Rivilyo Mangolat Rizky Sitanggang developed the research idea, collected and analyzed the data, and wrote the manuscript. Wa Ode Dianita Putri Suaiba Dani collected the data, analyzed the results, and contributed to writing the methodology and results sections. Bambang Setiadi supervised the research, provided academic guidance, and contributed to the revision of the manuscript. Yanif Dwi Kuntjoro supervised the research, provided technical and conceptual guidance, and contributed to the review and revision of the manuscript.

## References

Maria, N., Sadia, S., & Khurram, K. (2021). Role of deep learning in brain tumor detection and classification (2015 to 2020). Comput Med Imaging Graph.

Salakhutdinov, R. (2015). Learning deep generative models. Annual Review of Statistics and Its Application, 2(1), 361-385.

Sternberg, S. R. (1983). Biomedical image processing. Computer, 16(01), 22-34.

Morgan, F. E., Boudreaux, B., Lohn, A. J., Ashby, M., Curriden, C., Klima, K., & Grossman, D. (2020). Military applications of artificial intelligence. Santa Monica: RAND Corporation.

Sun, Y., Wang, H., Xue, B., Jin, Y., Yen, G. G., & Zhang, M. (2019). Surrogate-assisted evolutionary deep learning using an end-to-end random forest-based performance predictor. IEEE Transactions on Evolutionary Computation, 24(2), 350-364

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

Ling, H., Gao, J., Kar, A., Chen, W., & Fidler, S. (2019). Fast interactive object annotation with curve-gcn. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5257-5266).

Yang, F., Hu, L., Liu, X., Huang, S., & Gu, Z. (2023). A large-scale dataset for end-to-end table recognition in the wild. Scientific Data, 10(1), 110.

Zheng, D., Li, S., Fang, F., Zhang, J., Feng, Y., Wan, B., & Liu, Y. (2023). Utilizing bounding box annotations for weakly supervised building extraction from remote- sensing images. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-17.

Marrable, D., Barker, K., Tippaya, S., Wyatt, M., Bainbridge, S., Stowar, M., & Larke, J. (2022). Accelerating species recognition and labelling of fish from underwater video with machine-assisted deep learning. Frontiers in Marine Science, 9, 944582.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

Francis, A., Sidiropoulos, P., & Muller, J. P. (2019). CloudFCN: Accurate and robust cloud detection for satellite imagery with deep learning. Remote Sensing, 11(19), 2312.

Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS. pp. 1106–1114 (2012).

Wu, Z. (2019). Muti-type Aircraft of Remote Sensing Images: MTARSI [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3464319.