http://doi.org/10.35784/iapgos.6809

received: 11.11.2024 | revised: 30.05.2025 | accepted: 09.06.2025 | available online: 27.06.2025

FINE-GRAINED DETECTION AND SEGMENTATION OF CIVILIAN AIRCRAFT IN SATELLITE IMAGERY USING YOLOv8

Ramesh Kumar Panneerselvam, Sarada Bandi, Sree Datta Siva Charan Doddapaneni

Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India

Abstract. *Detection and segmentation of civilian aircraft from satellite imagery has significant importance in applications for air traffic management, surveillance, and defense. Yet, its visual confusions and lack of unification in recognition make it hard. This paper presents that by developing an efficient YOLOv8-based model for aircraft detection, classification, and segmentation within the FAIR1M-2.0 dataset. This proposed methodology involves dataset preprocessing and compatibility adjustments where the backbone used is CSPDarknet53 combining with the C2f module, which provides an efficient multi-scale representation, this happens to be the most critical requirement in distinguishing between among 11 unique categories of aircraft. Including the SAM model helps improve localization precision by achieving more accurate pixel-level segmentation. The present work effectively carried out an accurate classification and described civilian aircraft, containing the enhanced detection and quantification capability appropriate for complex satellite-oriented aircraft analysis. These reasons make the work satisfy the fundamental requirement for very accurate identification and evaluation of aerial images. The approach improves the accuracy and precision of aircraft classification over delicate satellite images, and thus is useful in operations for real-time surveillance and monitoring. Fine-grained classification and segmentation would then be able to effectively capture slight differences between aircraft types, which are now vital to the reliable management of airspaces. This work, therefore sets a good foundation for future development and advancement of high-resolution aerial analysis in diverse operational settings.*

Keywords: aircraft classification, fine-grained recognition, object detection, satellite imagery, segmentation, YOLOv8

PRECYZYJNE WYKRYWANIE I SEGMENTACJA SAMOLOTÓW CYWILNYCH NA ZDJĘCIACH SATELITARNYCH PRZY UŻYCIU YOLOv8

Streszczenie. *Wykrywanie i segmentacja cywilnych samolotów na podstawie obrazów satelitarnych mają kluczowe znaczenie w zarządzaniu ruchem lotniczym, nadzorze oraz obronności. Ze względu na wizualne podobieństwa między różnymi typami samolotów oraz brak standaryzacji w rozpoznawaniu, jest to zadanie trudne. Niniejszy artykuł przedstawia efektywny model oparty na YOLOv8 do wykrywania, klasyfikacji i segmentacji samolotów w zbiorze danych FAIR1M-2.0. Zaproponowana metodologia obejmuje wstępne przetwarzanie danych i dostosowanie do zgodności, w którym wykorzystano CSPDarknet53 jako bazę, połączoną z modułem C2f, co zapewnia efektywną reprezentację wieloskalową – jest to kluczowy element przy rozróżnianiu 11 unikalnych kategorii samolotów. Włączenie modelu SAM poprawia precyzję lokalizacji, pozwalając na dokładniejszą segmentację na poziomie pikseli. Prezentowane badania pozwoliły na dokładną klasyfikację i opisanie cywilnych samolotów, zapewniając ulepszone możliwości wykrywania i analizowania obiektów na obrazach satelitarnych. Takie podejście znacznie zwiększa dokładność i precyzję klasyfikacji samolotów, co czyni je przydatnym w operacjach nadzoru i monitorowania w czasie rzeczywistym. Precyzyjna klasyfikacja i segmentacja umożliwia skuteczne rozróżnianie subtelnych różnic między typami samolotów, co jest istotne dla niezawodnego zarządzania przestrzenią powietrzną. Niniejsza praca stanowi solidną podstawę dla przyszłych badań nad analizą obrazów lotniczych w wysokiej rozdzielczości w różnych kontekstach operacyjnych.*

Słowa kluczowe: klasyfikacja samolotów, rozpoznawanie szczegółowe, wykrywanie obiektów, obrazy satelitarne, segmentacja, YOLOv8

Introduction

Detection, classification and segmentation of aircraft from satellite images are of paramount importance in air traffic management, surveillance, and national security. There exists a greater demand for automated classification systems that could recognize and classify an enormous number of aircraft with greater accuracy due to their higher resolutions available in satellite imagery. Adequate and appropriate classification and segmentation of aircraft increases situational awareness while utilizing all available space effectively, thus making travel by air on time possible.

A huge number of papers discuss airplane classification and segmentation tasks with different approaches and datasets. Azam et al, enhanced aircraft classification accuracy using PCA combined with fusion techniques and convolutional neural network [1]. Castilho et al. employed machine learning methods including support vector machines and classification trees to accurately identify aircraft bleed valve malfunctions in real-time systems [2]. CV et al. focused on deep learning-based instance segmentation of aircraft in aerial images using Detectron2, effectively demonstrating the model's ability to accurately identify and segment aircraft [3]. Studies [4, 7, 8, 10, 11] have highlighted the effectiveness of deep learning models, including CNNs, EfficientNet, and Swin Transformer, in enhancing aircraft classification accuracy by integrating feature fusion and attention mechanisms to boost performance on aerial and SAR images. Gao et al. optimized the Swin Transformer for classifying aircraft based on the MTARSI dataset that serves as an alternative means of feature extraction along with classification [6]. Kang et al. presented ST-Net for classification in high resolution SAR images with very impressive accuracy obtained [12]. Khan et al. introduced a deep learning-improved YOLOv8 algorithm for real-time precise instance segmentation of crown region orchard canopies, demonstrating enhanced accuracy in natural environments [13]. Liu et al aimed to design YOLO-Class-a model inspired from the architecture of YOLO-to detect and classify targets methods referring to aircraft in remote sensing images by satellites so showed that based on YOLO can efficiently be used for it [14]. Studies such as [19, 21] combine spatial attention and component-based analysis to facilitate accurate aircraft identification, utilizing supervised learning and convolutional neural network modules on synthetic aperture radar and radar data to enhance interpretability in complex scenarios. Pandey et al. presented a comprehensive multimodal segmentation approach by combining YOLOv8 with SAM and HQ-SAM models, achieving impressive accuracy in medical imaging segmentation tasks [16]. Poojitha and Nasreen proposed a highly accurate aircraft recognition system using EfficientNet with a focus on fewer parameters, which can further be used in optimizing the parameters of YOLOv8 [17]. Sapkota et al. compared YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments and highlighted the strengths and limitations of each model [18]. Sun et al proposed Scattering Characteristics Analysis Network as an alternative to the scarcity of the number of labeled data for aircraft classification in SAR images [20]. Zhao et al designed a pyramid attention dilated network to enhance detection accuracy for SAR images [29].

Although the account assimilates such recent advances, this task is still quite challenging to differentiate between many classes of aircraft based on satellite imagery because most types of aircraft are visually alike. Researchers [22, 23, 25] investigated scattering characteristics and pyramid attention methods in SAR-based aircraft identification to enhance spatial context recognition and position determination in densely populated areas. Recent studies [27, 28] have shown that YOLOv8 and advanced segmentation frameworks such as Detectron2 and Mask R-CNN perform effectively in instance segmentation tasks, particularly in intricate and real-world settings such as orchards and aerial photography. Aircraft of many types resemble each other so that the model developed has to work with high accuracy in classification and fine-grained segmentation of various types of aircraft in satellite images.

Wang et al. in 2023 also proposed advanced model of YOLOv5 for a more proficient recognition of aircraft targets in SAR images, and work presented here may be extended in developing YOLOv8 [24]. These references are important as they include architectures similar to the one deployed, object detection, and classification in high-resolution images; hence, approaches toward addressing challenges like data scarcity and model optimization, etc., are necessary.

The goal of this research is the development of an aircraft type detector model, that is, a YOLOv8-based detector model. This model uses the FAIR1M-2.0 dataset to classify and segment the particular aircraft types involved. The methodology implemented is filtering the dataset to only include images of aircraft, then converting XML-based bounding box annotations into YOLOv8-compatible .txt format. The YOLOv8 Nano model features a CSPDarknet53 backbone for feature extraction, the C2f module for fusing feature maps, and multiple detection heads for predicting bounding boxes and estimating class probabilities.

1. Literature review

Azam et al. [1] used PCA with feature fusion within a Convolutional Neural Network for the classification of aircraft. The authors combined handcrafted LBP features with deep features extracted from AlexNet and Inception V3, using CNNs. Then, these features are fed to PCA for dimensionality reduction as well as improving the classifier performance. The linear SVM classifier tests the MTARSI set to achieve 96.8% accuracy. This is an indicator of a good feature choice and working; however, its dependence on manually engineered features and the intricacies entailed in feature fusion exposes some challenges as regards handling new data and its adequacy with available computational resources.

CV et al. [3] use deep learning to the instance segmentation of aerial imagery concerning aircraft as the chief tool machinery using the Detectron2. Compared with other types, such models correctly identify and distinguish various sorts of aircraft in complicated aerial images. Results from testing yield an F1 score of 91% on the Aerial Images for Aircraft Detection dataset, meaning the model performs well in several environmental settings and, thus, is an extremely valuable asset for applications in aerial surveillance and military.

Gao et al. [6] utilized the MTARSI dataset in their "Optimizing and Evaluating Swin Transformer for Aircraft Classification" research as an experiment for utilizing Swin Transformer on tasks of aircraft classification. Their effort on trying to optimize the models chosen prominently highlighted training schemes development, resulting in excellent validation accuracy levels of 99.4% in super-resolution-free testing, while the utilization of super-resolution methods attained 99.5% levels. The authors have mentioned that the MTARSI dataset is very simple with limited diversity and perhaps also misannotated. Such a restriction makes the model less capable of generalizing well beyond simple scenarios to complex real-world scenarios and out-of-distribution test samples even though it has excellent near-perfect performance on validation sets. The study thus clearly shows that benefits and challenges when using advanced models like Swin Transformers in niches like aircraft classification come together.

Hassan et al. [9] presents a deep learning-based framework designed for automatic airplane detection in satellite imagery. It accounted for some loss of information from the resizing and includes both the mosaic layer after the output of the detection layer and the split layer before the input. Their proposal framework used faster R-CNN for the detection, and Inception v2 as a feature extractor. The model tested on Google Earth datasets and proved excellent detection accuracy of airplanes. Though a very good detection accuracy and a new strategy towards the challenge of solving in image resizing, adding Faster R-CNN and Inception v2, do make the model more complex. If one applies only the transfer learning technique, its application scenarios may be limited in the cases of different remote sensing images or new datasets.

Kang et al. [12] presented ST-Net: a new approach in the field of classification of high-resolution aircraft in SAR images with a module simulating spatial relationships and semantic interactions between discrete scattering points for variability in SAR imaging as well as aircraft scattering characteristics. Performance is based on Context Attention Excitation that enhances global and semantic representations of information while deleting interference from the background as well as ambiguity during classification. Thus, superior classification performance on the SAR-ACD dataset is achieved with ST-Net as it correctly justifies complete exploitation of its capacity in discriminating useful features and modeling complicated scattering topologies. Its implementation complexity and potential for overfitting to a particular training dataset would also affect its generalisability to various SAR imaging conditions as well as practical applications.

Khan et al. [13] recently published a state-of-the-art YOLOv8 algorithm for real-time segmentation of orchard canopies in the crown region in natural environments. It removes conditions including occlusion and variability in illumination, which should give a quite fair canopy segmentation. The Orchard Canopy Segmentation dataset was utilized for the purpose of results evaluation; it brought forth a mean Intersection over Union (mIoU) of 88%, making it appropriate for agriculture's monitoring and management applications.

Liu et al. [14] developed the improved model named YOLO-Class based on the method of classification and the detection of aircraft over remote sensing imagery by satellite by adopting the YOLOv5 approach. The model employed all the improvements including Representative Batch Normalization, Mish activation function, optimized convolutional modules, and VariFocal loss due to imbalanced data sets. This integration comes with a backbone that improves detection accuracy from 0.608 up to 0.704 and FPS from 36.16 to 39.598. YOLO-Class shows excellent performance on small, dense, or occluded objects but at the same time loses much in terms of complexity and susceptibility to overfitting with specific optimizations, which proves relatively important for broader applicability and generalization over various datasets and scenes in the real world.

Nie et al. [15] specifically develop an adaptive approach termed Adap-EMD for the task of fine-grained aircraft classification in remote sensing. It incorporates FSL techniques that utilize EBAM – an efficient block attention mechanism – and AFMF – an adaptive feature measurement filter – as means of addressing data scarcity. EBAM captures better global feature maps by incorporating channel and spatial correlations; AFMF boosts the accuracy in computing similarity by focusing strongly on key information of the feature map. Adap-EMD performs better than state-of-the-art models on benchmark datasets for few-shot aircraft fine-grained recognition with better performance at a minimal computational overhead. There is a dire lack of fine-grained aircraft data in remote sensing that affects model training; furthermore, few-shot learning is extremely challenging and requires advanced strategies to achieve the highest accuracy.

Pandey et al. [16] suggested a complex multimodal segmentation method for medical images, making use of the YOLOv8 model, and SAM along with HQ-SAM models. The resulting framework increases the accuracy level for segmentation since it takes complementary information obtained from the various modalities; thus, it demonstrates significant improvements in many anatomical structures' recognition within medical images. Experiments were successfully carried out on the dataset Kaggle Medical Segmentation with a fantastic accuracy of 93.7%, which can be applied to clinical cases requiring high accuracy levels.

Poojitha et al. [17] in Aircraft Recognition System, used the EfficientNet B3 model, to classify and recognize the military aircraft with much accuracy. It uses deep learning algorithms and computer vision-based techniques to automatically enhance aviation security by aircraft detection. The proposed model was also much better when compared with the architectures of VGG16, InceptionV3, and ResNet50 and attained the accuracy of 95.76%. This system is excellent for applications requiring high accuracy, real-time operations, with efficient processing and where model complexity demands plentiful amounts of computational resources and long training times because the system is deep and possibly complex and may thereby affect the deployment speed to be used in fast operational scenarios.

Sapkota et al. [18] performed the experiments of YOLOv8 and Mask R-CNN in more complex scenarios of orchard application. According to their results, as though both models represented competitive results, it is claimed that the developed YOLOv8 attained better speed and efficacy under a real-time scenario. After applying the Complex Orchard Segmentation dataset, YOLOv8 attained an mIoU of 85%, and thereby showing that the proper choice of model with the needed specifics is required, especially for application domains within precision agriculture where the right information is timely critical.

Sun et al. [20] proposed in synthetic aperture radar images and tested in high resolution, the Scattering Characteristics Analysis Network for few-shot aircraft classification. A new network, less training data compared to other networks, sets in a novel technique of analyzing scattering characteristics, which are crucial in aircraft-type distinctions. On the SAR images, target discreteness and background interference are two aspects that SCAN perceives to reduce false alarms and enhance class classification accuracy. It demonstrates efficacy on the SAR-ACD dataset-improving the structural correlation between scattering points while suppressing clutter interference. Biases involved in current datasets carry some of which include limitations in few-shot learning for SAR images would challenge the direct applicability and performance evaluation.

Yang et al. [26] proposed the new framework of plant leaf image segmentation with the integration of an advanced DeepLabV3+ model and YOLOv8 architecture. The main objective of this research work is to improve the accuracy of segregation for leaf plants as this is important for analysis and monitoring in the agricultural field. This is achieved through the primitive object detection scheme with the help of YOLOv8 followed by pixel-wise segmentation with an advanced DeepLabV3+.The experiments were carried out using a diversified dataset of plant leaf images, that have been intentionally composed from different species and contexts The evaluation metric shows that the proposed approach reaches an average IoU of 93.6%. The huge improvement with respect to standard methods applied in segmentation demonstrated in the proposed approach testifies to the fact that using YOLOv8 within efficient segmentation models is a clever idea for detailed high-resolution analysis of plant leaf images.

2. Methodology

2.1. Task definition

Ensuring the safety and efficiency of air travel is of utmost importance in the aviation sector. The early and precise identification of aircraft in a variety of operating settings is a critical component of this endeavor. However, the creation of automated solutions is required to expedite this work because human aircraft identification procedures are laborious and prone to mistakes. With a particular focus on the YOLOv8 Nano model, the research intends to use deep learning techniques to construct an effective aircraft identification and segmentation system in order to meet this demand.

Aircraft counting involves the application of computer vision techniques to accurately determine the number of aircraft present in a given scene or image. Utilizing algorithms such as object detection and image segmentation, aircraft counting systems identify and localize individual aircraft within the image. By analyzing the spatial distribution and characteristics of detected aircraft, these systems can provide precise counts in various environments, including airports, airspace, or aerial imagery. Aircraft counting enhances operational efficiency and safety in aviation by facilitating real-time monitoring of air traffic and resource allocation.

The study uses visual clues taken from optical pictures to classify airplanes. The capability of YOLOv8's sophisticated pattern recognition to distinguish between different types of aircraft increases the level of detail in airspace monitoring. Deep learning algorithms and other computer vision techniques are frequently employed for this kind of work. These systems learn to distinguish between passenger jets, cargo planes, helicopters, and other aircraft variants by training models on annotated datasets including photographs of different aircraft types. Aviation stakeholders can gain significant insights from aircraft classification, which facilitates a better knowledge of fleet compositions, airspace usage, and operational patterns.

2.2. Proposed model

Our proposed airplane type identification and segmentation methodology on the FAIR1M-2.0 dataset actually is three-step-preprocessing of the data, followed by the training of YOLOv8 Nano on the models and then evaluation. Data preprocessing includes conversion of the original XML annotations of bounding box coordinates into the traditional YOLOv8 format, that is, .txt format. After that, the dataset filtered images containing different types of aircraft so that our data would be relevant to our classification task. In addition to that, all images are resized to uniform resolution amenable to training YOLOv8 Nano.

It utilizes the YOLOv8 Nano model at the training stage. This model has been selected for an optimal balance between precision and computational requirements. Hence, it is very suitable for such an application with the use of high-resolution satellite imagery. The YOLOv8 Nano model learns by utilizing a preprocessed FAIR1M-2.0 dataset to carry out feature extraction as well as instance detection regarding aircraft. It first of all categorizes the aircraft, after the aircraft are sensed, according to classes like Boeing 737, Boeing 747, Airbus A320, and many more.

It finally evaluates the performance of the model. To measure the accuracy of the model with which different types of aircraft are classified, scores like Precision, Recall, F1-score, and mean Average Precision or mAP are used. At last, a comparison with traditional methods is carried out to point out the progress acquired and to outline possible directions for future improvement. The process flow of the proposed model is shown in Fig. 1.

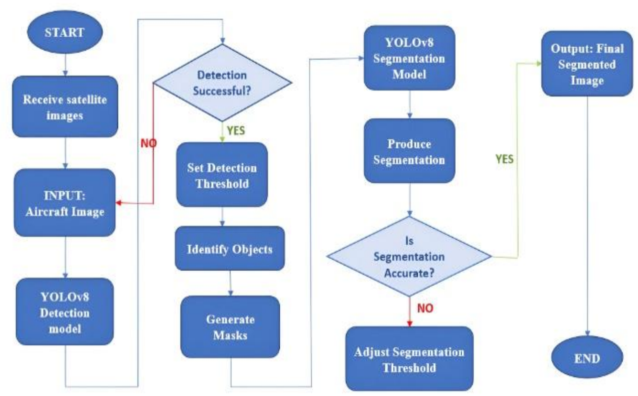

Fig. 1. Process flow model

2.3. Dataset description

The FAIR1M-2.0 dataset is originally created, focusing on high-resolution satellite images particularly designed for tasks requiring detailed object detection and recognition. It comprises 15,000 RGB images sourced from Gaofen satellite by accessing Google Earth. The dataset features rich annotations that use rotated bounding boxes for more accurate depiction of the orientation of different objects and scenes, with 37 subcategories across five major categories: ships, vehicles, airplanes, courts, and roads.

It filtered the FAIR1M-2.0 dataset for aircraft images only. Filtering only for images containing aircraft resulted in a subset of subsets of various categories of aircraft. Besides that, the dataset was well-curated towards 11 specific aircraft types as shown in Fig. 2, which gives better precision in detection and classification effort and effect.

Translating these initial XML-formatted annotations of bounding box coordinates on aircraft instances into standard YOLO format (.txt) will make them compatible with the needs of the YOLOv8 model. Thus, it would be easy to implement inside the Ultralytics YOLOv8 Nano framework. Annotations for other objects which are not aircraft have been excluded to make this dataset narrow to exclusive aircraft identification and classification.

The architecture of the model used for this research is shown in Fig. 3.

Images from the dataset for reference are shown in Fig. 4 and Fig. 5.

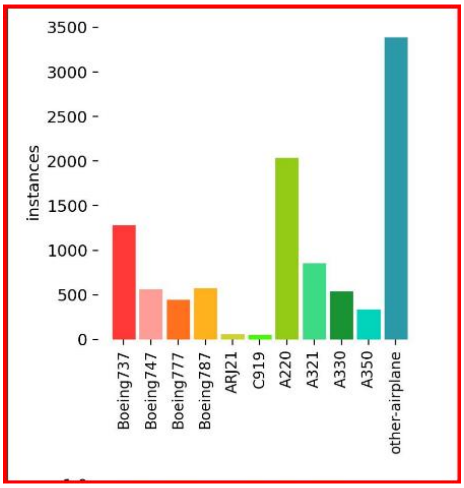

Fig. 2. Aircraft classes

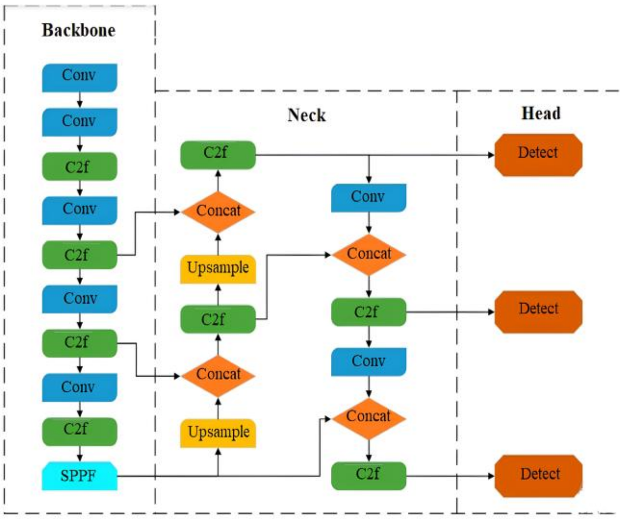

Fig. 3. YOLOv8 architecture [5]

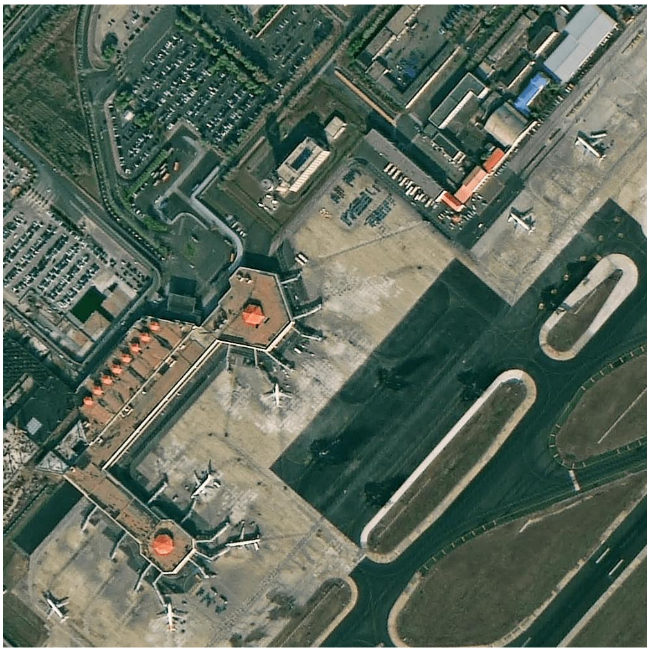

Fig. 4. Image from the FAIR1M-2.0 Dataset [19]

Fig. 5. Image with bright background from FAIR1M-2.0 Dataset [19]

2.4. Data preprocessing

The preprocessing of the dataset is our first step in our approach, making sure it gets put into a proper form to train the YOLOv8 Nano model. The dataset FAIR1M-2.0 is annotated in terms of XML stating coordinates for the bounding boxes of aircraft instances, and these XMLs are turned into .txt, which is common in use when annotating about to use a model like YOLO. Each file of the .txt contains a class label along with normalized coordinates of bounding boxes in such a way that annotation could match up to the specs of the YOLOv8 Nano model.

We then curate that set more into images of merely airplanes belonging to different classes of aircraft, disregarding other object categories such as vessels, cars, parks, and roads. In this way, it will ensure the set only consists of images with regard to aircraft classification. Hence, the images are resized to a uniform resolution often of 640×640 pixels so that standard dimensions for the input to the model can be stipulated. This ensures consistency and that the training of the model is efficient as well as effective.

Other features that improved the high-resolution pixel-level segmentation functionality applied to create the masks included the application of the Segment Anything Model, or SAM. These are quite effective at identifying very fine characteristics within complex images. SAM employs deep learning techniques for automatic generation of high-quality segmentation masks delineating the objects of interest, in this case, aircraft within a satellite image.

The integration of SAM into the workflow is based on improving the segmentation process, so for each aircraft, precise masks were available. Methodology: input satellite images were taken into the SAM framework, where visual information was automatically processed and segmentation masks describing contours of each aircraft were produced. Critical annotations; these masks genuinely enhance the efficacy of next-level segmentation algorithms.

3. YOLOv8 nano model training

The training portion is focused on using YOLOv8 Nano model. The choice has been done with YOLOv8 Nano as it is basically an optimal design well balanced in terms of accuracy with computational efficiency, which makes it very useful for high-resolution satellite image analysis when computational resources are in short supply.

The preprocessed images with their annotations taken as input in the training of the YOLOv8 model. This led to inputs into the model which would detect and classify as many types of aircraft that appear in the images. The CSPDarknet53 components include a neck and head structure to input data from input images. The proposed method outputs bounding boxes for the detected objects along with their objectness scores and class probabilities.

Alongside the typical YOLOv8 loss functions – classification loss, localization loss, and objectness loss, the segmentation model incorporates an extra loss function that is dedicated to evaluating the precision of the predicted segmentation masks. Mask loss might differ from the differences that occur between the masks generated and the real masks presented by SAM. The model refines its prediction based on the differences identified wherein masks closely resemble the real boundaries of aircraft.

During training, the model is designed to predict everything at the same time-that is, object detection's bounding boxes and the segmentation pixel-level masks. The common features existing between these two, namely the detection task and the segmentation task, can be exploited in the dual-task architecture of the model to better facilitate accuracy in these two tasks. It systematically reduces the loss function describing the differences between the bounding boxes and class labels predicted and true bounding boxes, class labels, and segmentation masks. Therefore, it enhances not only the accuracy of recognition but also delimiting different types of aircraft while strengthening the reliability and efficiency of segmentation.

3.1. Evaluation

The last step involves the performance evaluation of YOLOv8. In order to evaluate our model, we used different measuring criteria. We employed various evaluation metrics to measure the model's accuracy and to identify differences when sorting aircraft accurately. This includes Precision, Recall, F1-score, and mean Average Precision (mAP).

- Accuracy is the number of true positives divided by the total number of positive events.
- Recall assesses the degree of correct identification and measures the ratio of correctly identified positive events.
- F1-score is the harmonic mean of Precision and Recall, used as a single metric to weigh the costs and benefits.
- Average Precision(AP) measures the average precision across all classes and provides a balanced measurement of the model's performance.
- Mean Average Precision (mAP) calculates the mean of the average precision scores for each class, serving as a comprehensive metric for assessing overall model effectiveness in multi-class scenarios.

The evaluation of these metrics defines the value of the model and highlights any areas for improvement. We also compared our results with those obtained from known methods to demonstrate the validity of our approach. This comparison reveals whether improvements have been made and identifies areas that may benefit from further research.

Algorithm 1: Generate Segmentation Masks

Input: Dataset D; Detection Model M_a ; Segmentation Model SAM; Threshold T

Output: List of segmentation masks M.

1. Initialize an empty list of masks M.
2. For each image I in the dataset D:
    1. Use the detection model M_a to identify objects in image I, generating a set of predictions in the form of pairs: (B_i, C_i), where B_i is the bounding box and C_i is the class label for each detected object from i=1 to K.
    2. For each bounding box B_i in P:
        1 Extract the region of interest (ROI) R from I based on B_i.
        2 Apply SAM to generate a mask M_i for the ROI R.
        3 Refine the mask M_i by applying the threshold T.
    3. Combine all masks M_i for image I to produce the final mask M for that image.
    4. Append the final mask M to the list M.
    5. Return the list of segmentation masks M.

Generating Segmentation masks is imperative to fine-grained civilian aircraft detection and segmentation in satellite imagery because these masks ensure pixel-level precision needed to differentiate various types of aircraft. Interesting as it may sound, while other algorithms would at best generate only simple bounding boxes, the information captured by the model through segmentation masks would be crucial in picking out subtle shape-based features that distinguish between quite similar types of aircraft. That increased accuracy is particularly important for complex satellite imagery, where aircraft may be very close or partially obscured by other objects. Such masks also allow for post-processing analysis, calculating shape-based metrics for the visual results, making it easier to understand the outcome for end-users.

Algorithm 2: Aircraft Detection and Segmentation

Input: Dataset D with bounding box annotations and segmentation masks; Number of epochs E; Batch size B; Learning rate LR

Output: Trained model M; Detection results R.

1. Initialize the YOLOv8 model M.
2. Split the dataset D into a training set D_{train} and validation set D_{val}.
3. Preprocess images in D_{train} and D_{val}:
    1. For each image I in D_{train}:
        1. Clean and augment I.
        2. Normalize I.
        3. Resize I to the input size (H, W).
        4. Convert annotations A to YOLO format.
    2. For each image I in D_{val}:
        1. Normalize and resize I.
4. For each epoch from 1 to E:
    1. Shuffle D_{train}.
    2. For each batch B in D_{train} :
        1. Load batch images and annotations.
        2. Feed images through model M to generate predictions P.

3. Calculate the loss L between predictions P and the ground truth annotations.
4. Update model weights using backpropagation with learning rate LR.

5. Validate the model M on D_{val}:
   - For each image I in D_{val}:
     1. Feed I through M to obtain predicted bounding boxes B and class labels C_i.
     2. Apply non-maximum suppression to remove duplicate predictions.
     3. Store results R={(I,B,C_i)}.
6. Return the trained model M and detection results R.

This algorithm outlines the training process for a YOLOv8 model on object detection and segmentation tasks with an annotated image dataset using bounding boxes and segmentation masks. First, initialize the model and split the set into training and validation subsets. All images in both subsets are preprocessed, augmented, normalized, resized, and transformed to the appropriate YOLO format. Then the epochs iteration process occurs; data is shuffled and batched accordingly. This is a computation of loss between the predictions of a model and actual annotations of each batch. It updates the weights by backpropagation. After each epoch, it tests the performance of the model using predictions from the model on the validation set, applying non-maximum suppression to remove duplicate predictions ensuring each aircraft is counted only once, and saving the output. Finally, the algorithm outputs the trained model and detection results, hence providing an all-inclusive output of civilian aircraft detection and segmentation.

3.2. Mathematical relations

$$L_{total} = L_{loc} + L_{conf} + L_{class} + L_{seg} \tag{1}$$

$$L_{loc} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y - \hat{y}_i)^2 + \\ +(w - \widehat{w}_i)^2 + \left(h - \hat{h}_i\right)^2] \tag{2}$$

$$L_{conf} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left(C_i - \hat{C}_i\right)^2 + \\ + \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{noobj} \left(C_i - \hat{C}_i\right)^2 \tag{3}$$

$$L_{class} = \sum_{i=0}^{S^2} 1_{i}^{obj} \sum_{c\in classes} \left(p_{i(c)} - \hat{p}_{i(c)}\right)^2 \tag{4}$$

$$L_{seg} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{C} \left(y_{n,c} log\left(\hat{y}_{n,c}\right) + \\ +(1 - y_{n,c}) log(1 - \hat{y}_{n,c})\right) \tag{5}$$

$$Loss = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \ [(x_i - \hat{x}_i) + (y_i - \hat{y}_i)]^2 + \\ +\lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left(C_i - \hat{C}_i\right)^2 + \\ +\lambda_{mask} \sum_{i=0}^{S^2} \mathbb{1}_{ij}^{obj} \ MaskLoss\left(\widehat{M}_i, M_i\right) \tag{6}$$

Recall:

$$R = \frac{TP}{TP+FN} \tag{7}$$

Precision:

$$P = \frac{TP}{TP+FP} \tag{8}$$

F1 Score:

$$F1 = \frac{2.P.R}{P+R} \tag{9}$$

mean Average Precision (mAP):

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{10}$$

Intersection over Union (IoU):

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{11}$$

mean IoU (mIoU):

$$mIoU = \frac{1}{N} \sum_{i=1}^{N} IoU_i \tag{12}$$

where FN – False Negatives, TN – True Negatives, FP – False Positives, TP – True Positives.

4. Experimental setup and result

It is the methodology designed to recognize, classify and segment aircraft using YOLOv8 Nano in PyTorch as well on Google Colab using the dataset FAIR1M-2.0 – this is a total collection of 2550 high-resolution images captured by Gaofen-3 satellite, where the bounding boxes for the aircraft, belonging to the 11 types, have known coordinates.

This model, based on the YOLOv8, depicts it to be very efficient in detecting, classifying, and segmenting a civilian aircraft from the satellite image.

The model is trained over epochs, fine-tuning parameters such as learning rate, batch size, and augmentations to optimize its performance. For every training epoch, forward passes make predictions of aircraft locations and classifications and then back propagate the loss using a custom-built function to penalize localization and classification errors.

Moreover, segmentation accuracy was much improved since the model was capable of distinguishing an outline of an aircraft from a complex background. This approach of performing the task simultaneously had improved detection precision and thus enabled full segmentation.

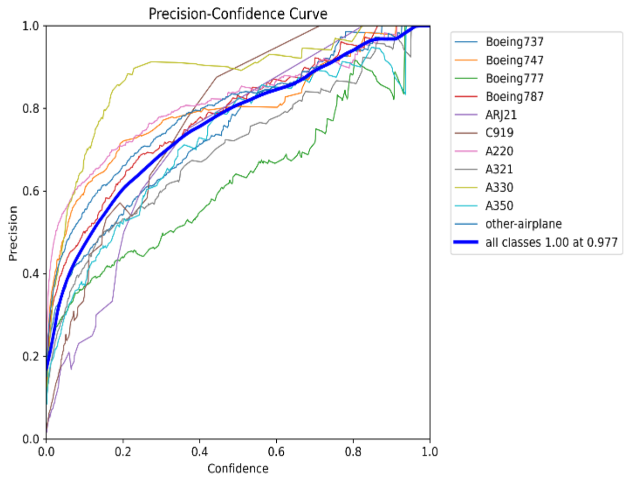

Fig. 6. Precision curve

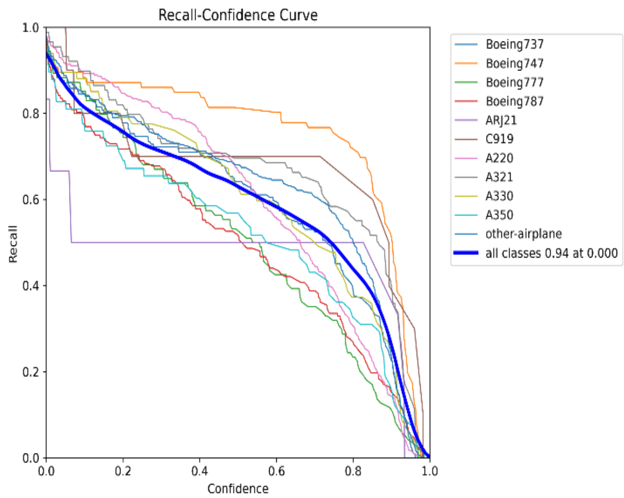

Fig. 7. Recall curve

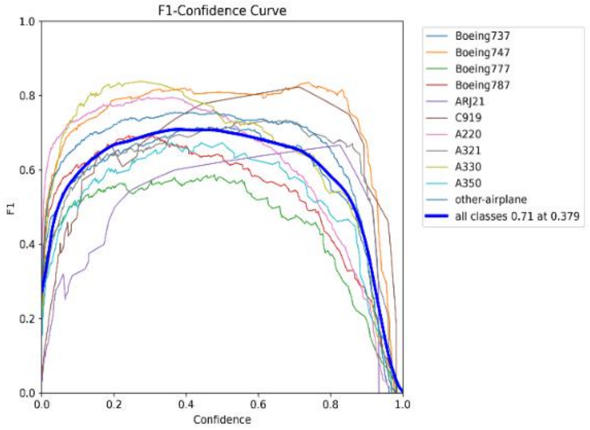

Fig. 8. F1-score curve

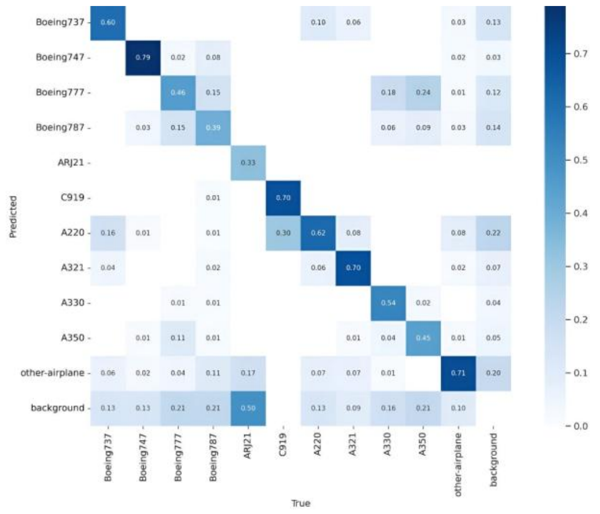

Fig. 9. Confusion matrix

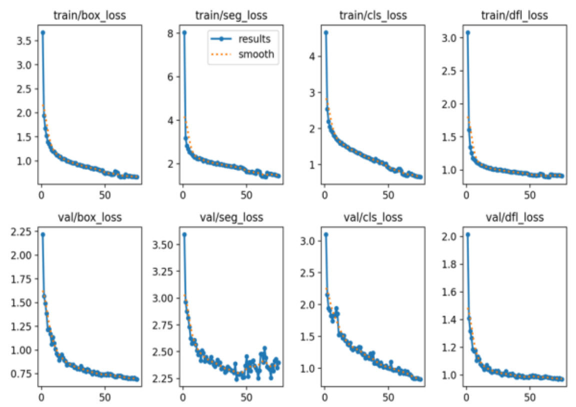

Fig. 10. Plots for loss

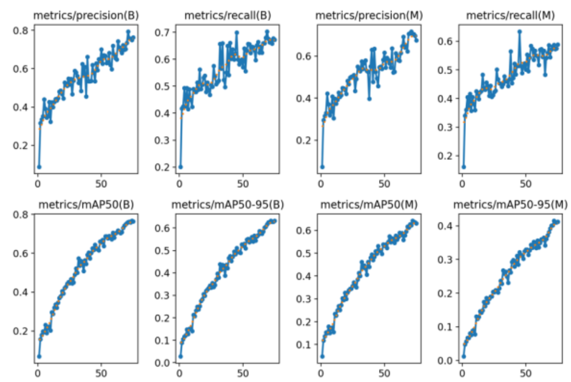

Fig. 11. Plots for mAP

Fig. 12. User interface

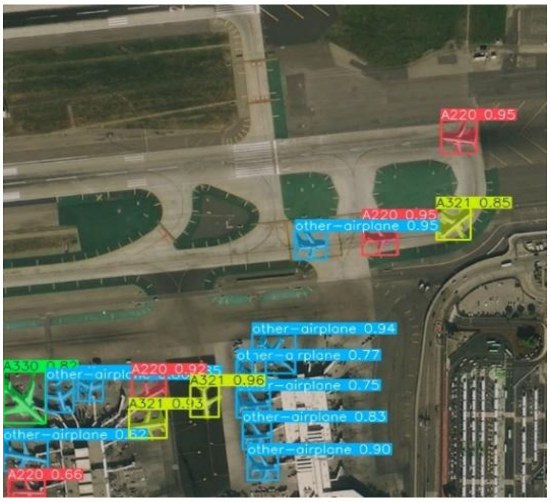

Fig. 13. Detection and segmentation results

The cloud-based GPU infrastructure is used to train the process, making use of Google Colab; the model performance is then tested using 5-10-fold cross-validation, categorizing the dataset into training (70%) and testing and validation (30%). Other performance metrics are illustrated to assess the model's effectiveness under diverse environmental conditions and image variations captured by the Gaofen-3 satellite. As shown in Fig. 6, the precision curve highlights how accurately the model identifies aircraft across different confidence thresholds. The recall performance, as illustrated in Fig. 7, reflects the model's ability to detect all relevant aircraft instances. Fig. 8, as depicted, presents the F1-score curve, showing the balance between precision and recall. Classification accuracy across the various aircraft categories is demonstrated by the confusion matrix in Fig. 9. As illustrated in Fig. 10, the training and validation loss graphs indicate stable learning and effective convergence. Fig. 11 shows metric plots, providing insights into how well the model refines its mask predictions over epochs. To make results accessible and user-friendly, a dedicated interface was developed, as shown in Fig. 12, allowing users to view detection and segmentation outputs interactively. Accurate and consistent localization of all aircraft is clearly demonstrated by the model, as illustrated in Fig. 13.

5. Conclusion and future work

Recognition and segmentation of aircraft types from geospatial imagery attracts a high level of research attention. In this paper, we first tested several datasets of satellite images that have been predominantly utilized to date for object recognition. We concluded that separate datasets either were inappropriate for our area of research or had significant limitations to be effectively applied to our project. Thus, we chose the FAIR1M-2.0 dataset finally for aircraft type recognition. The great diversity and the high-resolution Gaofen-3 satellite image are the primary reasons for selecting the dataset for our research. The FAIR1M-2.0 is a benchmark of 2550 images annotated with accurate bounding box coordinates of 11 specific aircraft types, against which methodologies can be developed to identify the type of aircraft, developed and reviewed. In addition, usage of dataset creates techniques in computer vision, image manipulation, and target identification, leading to creating aerial imageries. Our experiments leveraged a set of cutting-edge detection, classification and segmentation techniques focus on the YOLOv8 Nano model. By our experimental protocols, epochs of training, forwarding and back-propagation and key performance metric computation such as Precision, Recall, F1-score, and mean Average Precision (mAP), the data show that these different methods offer a good starting point for future researchers to perform comparative analysis. Although the YOLOv8 Nano model was established to be efficient in aircraft image classification, in our study, we were limited to using only satellite-based aircraft images. Multi-resolution methodology's effect on aerial imagery and relative importance of each tier of resolution in the image classification process are not analyzed in our study. Further development of these data sets in future research projects may be used for exploring some advanced aircraft recognition methods. Other goals include expanding the dataset to include heterogeneous data, possibly even classes of objects that may exist in a remotely sensed image.

Acknowledgement

We gratefully acknowledge the use of the FAIR1M-2.0 dataset for our research on aircraft detection and segmentation. This dataset, consisting of high-resolution satellite images from the Gaofen-3 satellite, has been instrumental in advancing our study. We appreciate the efforts of the dataset creators for providing a comprehensive and well-annotated resource that significantly contributed to the success of our project.

References

[1] Azam F., et al.: Aircraft classification based on PCA and feature fusion techniques in convolutional neural network. IEEE Access 9, 2021, 161683–161694 [https://dx.doi.org/10.1109/ACCESS.2021.3132062].

[2] Castilho H. M., Nascimento C. L., Loesch Vianna W. O.: Aircraft bleed valve fault classification using support vector machines and classification trees. Annual IEEE International Systems Conference – SysCon. IEEE, 2018 [https://doi.org/10.1109/SYSCON.2018.8369568].

[3] C V A., P K.: Deep Learning-Based Instance Segmentation of Aircraft in Aerial Images using Detectron2, 2023 [https://dx.doi.org/10.2139/ssrn.4485468].

[4] Dästner K., et al.: Classification of military aircraft in real-time radar systems based on supervised machine learning with labelled ads-b data. Sensor Data Fusion: Trends, Solutions, Applications – SDF. IEEE, 2018 [https://dx.doi.org/10.1109/SDF.2018.8547077].

[5] Elhanashi A., et al.: TeleStroke: real-time stroke detection with federated learning and YOLOv8 on edge devices. Journal of Real-Time Image Processing 21(4), 2024, 121 [https://dx.doi.org/10.1007/s11554-024-01500-1].

[6] Gao K., et al.: Optimizing and evaluating swin transformer for aircraft classification: Analysis and generalizability of the mtarsi dataset. IEEE Access 10, 2022, 134427–134439 [https://doi.org/10.1109/ACCESS.2022.3231327].

[7] Guo Q., Wang H., Xu F.: Scattering enhanced attention pyramid network for aircraft detection in SAR images. IEEE Transactions on Geoscience and Remote Sensing 59(9), 2020, 7570–7587 [https://doi.org/10.1109/TGRS.2020.3027762].

[8] Guo Q., et al.: Aircraft target detection from spaceborne SAR image. IEEE International Geoscience and Remote Sensing Symposium – IGARSS 2019. IEEE, 2019 [https://doi.org/10.1109/IGARSS.2019.8898548].

[9] Hassan A., et al.: A deep learning framework for automatic airplane detection in remote sensing satellite images. IEEE Aerospace Conference. IEEE, 2019 [https://dx.doi.org/10.1109/AERO.2019.8741938].

[10] He C., et al.: A component-based multi-layer parallel network for airplane detection in SAR imagery. Remote Sensing 10(7), 2018, 1016 [https://doi.org/10.3390/rs10071016].

[11] Kang Y., et al.: SFR-Net: Scattering feature relation network for aircraft detection in complex SAR images. IEEE Transactions on Geoscience and Remote Sensing 60, 2021, 1–17 [https://doi.org/10.1109/TGRS.2021.3130899].

[12] Kang Y., et al.: ST-Net: Scattering topology network for aircraft classification in high-resolution SAR images. IEEE Transactions on Geoscience and Remote Sensing 61, 2023, 1–17 [https://dx.doi.org/10.1109/TGRS.2023.3236987].

[13] Khan Z., et al.: Deep learning improved YOLOv8 algorithm: Real-time precise instance segmentation of crown region orchard canopies in natural environment. Computers and Electronics in Agriculture 224, 2024, 109168 [https://doi.org/10.1016/j.compag.2024.109168].

[14] Liu Z., Gao Y., Du Q.: Yolo-class: Detection and classification of aircraft targets in satellite remote sensing images based on yolo-extract. IEEE Access 11, 2023, 109179–109188 [https://doi.org/10.1109/ACCESS.2023.3321828].

[15] Nie Y., Bian C., Li L.: Adap-EMD: Adaptive EMD for aircraft fine-grained classification in remote sensing. IEEE Geoscience and Remote Sensing Letters 19, 2022, 1–5 [https://doi.org/10.1109/LGRS.2022.3168581].

[16] Pandey S., Chen K-F., Dam E. B.: Comprehensive multimodal segmentation in medical imaging: Combining YOLOv8 with SAM and HQ-SAM models. IEEE/CVF International Conference on Computer Vision. 2023 [https://dx.doi.org/10.1109/ICCVW60793.2023.00273].

[17] Poojitha K., Nasreen A.: Aircraft recognition system using deep learning based efficientnet. International Journal of Computer Science and Engineering 14(5), 2023, 182–187 [https://dx.doi.org/10.21817/indjcse/2023/v14i5/231405025].

[18] Sapkota R., Ahmed D., Karkee M.: Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments. Artificial Intelligence in Agriculture 13, 2024, 84–99 [https://doi.org/10.1016/j.aiia.2024.07.001].

[19] Sun X., et al.: FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. ISPRS Journal of Photogrammetry and Remote Sensing 184, 2022, 116–130 [https://doi.org/10.1016/j.isprsjprs.2021.12.004].

[20] Sun X., et al.: SCAN: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution SAR images. IEEE Transactions on Geoscience and Remote Sensing 60, 2022, 1–17 [https://doi.org/10.1109/TGRS.2022.3166174].

[21] Thai V.-P., et al.: Detection, tracking and classification of aircraft and drones in digital towers using machine learning on motion patterns. Integrated Communications, Navigation and Surveillance Conference – ICNS. IEEE, 2019 [https://doi.org/10.1109/ICNSURV.2019.8735240].

[22] Wang J., et al.: Integrating weighted feature fusion and the spatial attention module with convolutional neural networks for automatic aircraft detection from SAR images. Remote Sensing 13(5), 2021, 910 [https://dx.doi.org/10.3390/rs13050910].

[23] Wang W., et al.: Aircraft target classification for conventional narrow-band radar with multi-wave gates sparse echo data. Remote Sensing 11(22), 2019, 2700 [https://dx.doi.org/10.3390/rs11222700].

[24] Wang X., et al.: Aircraft target interpretation based on SAR images. Applied Sciences 13(18), 2023, 10023 [https://doi.org/10.3390/app131810023].

[25] Wang X., et al.: SAR Image Aircraft Target Recognition Based on Improved YOLOv5. Applied Sciences 13(10), 2023, 6160 [https://doi.org/10.3390/app13106160].

[26] Yang T., et al.: An approach for plant leaf image segmentation based on YOLOV8 and the improved DEEPLABV3+. Plants 12(19), 2023, 3438 [https://doi.org/10.3390/plants12193438].

[27] Yue B., et al.: MS-Net: A Multi-modal Self-supervised Network for Fine-Grained Classification of Aircraft in SAR Images. arXiv preprint arXiv:2308.14613, 2023 [https://dx.doi.org/10.48550/arXiv.2308.14613].

[28] Zhao Q., Du X., Lu Y.: Aircraft target classification based on CNN. IEEE 11th Sensor Array and Multichannel Signal Processing Workshop – SAM. IEEE, 2020 [https://doi.org/10.1109/SAM48682.2020.9104254].

[29] Zhao Y., et al.: Pyramid attention dilated network for aircraft detection in SAR images. IEEE Geoscience and Remote Sensing Letters 18(4), 2020, 662–666 [https://doi.org/10.1109/LGRS.2020.2981255].

Dr. Ramesh Kumar Panneerselvam
e-mail: rameshkumar@vrsiddhartha.ac.in

Dr. Ramesh Kumar Panneerselvam is currently serving as a associate professor at Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India. With more than 15 years of teaching experience, his research interests span biometrics, image and video processing, computer vision, and the Internet of Things. He has contributed significantly to his field with numerous publications.

https://orcid.org/0000-0003-3952-7072

Eng. Sarada Bandi
e-mail: bandisarada6@gmail.com

Sarada Bandi is a fourth-year B. Tech. student at Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India.
Her academic focus includes cloud computing, AI, machine learning and computer vision. As the Cloud Lead at Microsoft Learn Student Community, she has coordinated workshops on cloud technologies.

https://orcid.org/0009-0004-9279-4043

Eng. Sree Datta Siva Charan Doddapaneni
e-mail: sreedatta07@gmail.com

Sree Datta Siva Charan is a fourth-year B. Tech. student at Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India.
His interests lie in the fields of Internet of Things, machine learning and web programming.

https://orcid.org/0009-0009-1594-8607