The Visual Computer
https://doi.org/10.1007/s00371-024-03686-8

RESEARCH

Advanced semantic segmentation of aircraft main components based on transfer learning and data-driven approach

Julien Thomas[1] · Boyu Kuang[1] · Yizhong Wang[1] · Stuart Barnes[1] · Karl Jenkins[1]

Accepted: 9 October 2024
© The Author(s) 2024

Abstract

The implementation of Smart Airport and Airport 4.0 visions relies on the integration of automation, artificial intelligence, data science, and aviation technology to enhance passenger experiences and operational efficiency. One essential factor in the integration is the semantic segmentation of the aircraft main components (AMC) perception, which is essential to maintenance, repair, and operations in aircraft and airport operations. However, AMC segmentation has challenges from low data availability, high-quality annotation scarcity, and categorical imbalance, which are common in practical applications, including aviation. This study proposes a novel AMC segmentation solution, employing a transfer learning framework based on a sophisticated DeepLabV3 architecture optimized with a custom-designed Focal Dice Loss function. The proposed solution remarkably suppresses the categorical imbalance challenge and increases the dataset variability with manually annotated images and dynamic augmentation strategies to train a robust AMC segmentation model. The model achieved a notable intersection over union of 84.002% and an accuracy of 91.466%, significantly advancing the AMC segmentation performance. These results demonstrate the effectiveness of the proposed AMC segmentation solution in aircraft and airport operation scenarios. This study provides a pioneering solution to the AMC semantic perception problem and contributes a valuable dataset to the community, which is fundamental to future research on aircraft and airport semantic perception.

Julien Thomasa and Boyu Kuang have contributed equally to this work.

✉ Boyu Kuang
neil.kuang@cranfield.ac.uk

Julien Thomas
thomasjulien92140@gmail.com

Yizhong Wang
yizhong.wang@cranfield.ac.uk

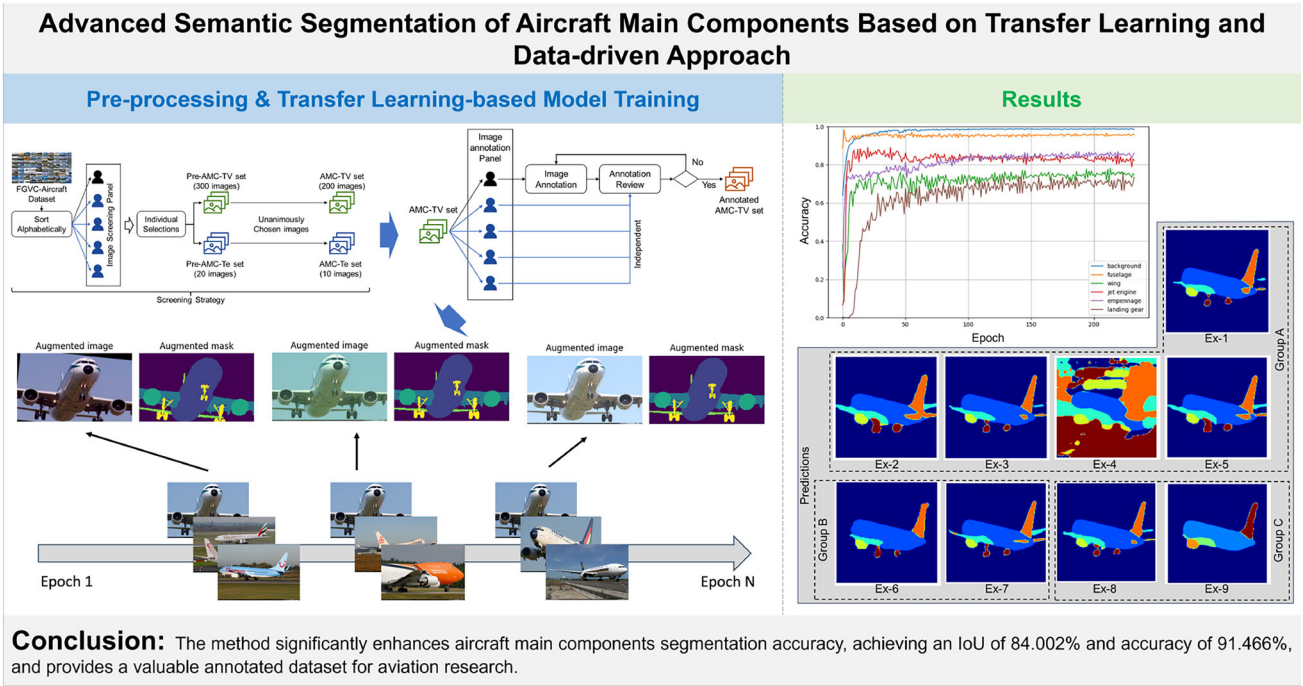
Stuart Barnes
stuart.barnes@cranfield.ac.uk

Karl Jenkins
k.w.jenkins@cranfield.ac.uk

[1] School of Aerospace, Transport, and Manufacturing, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK

Published online: 05 November 2024

Graphical abstract



Keywords Aircraft main component segmentation · Transfer learning · DeepLabV3 · Focal dice loss · Data augmentation · Aviation semantic perception

1 Introduction

Smart Airport and Airport 4.0 visions spearhead an unprecedented revolution in the modern aviation industry [1]. These emerging trends signify a shift toward automation and intelligence in airport operations and are also increasingly recognized globally as critical factors in enhancing passenger experiences and operational efficiency. With technological advancements and the progression of innovative projects, there is a proactive exploration and development of more automated solutions for smart airports, such as the *UK Research and Innovation* (UKRI)-funded *Future Flight Challenge Phase Three* and *Out of Cycle NExt generation highly efficient air transport* (ONEHeart) projects [2, 3]. These projects propel the advancement of intelligent technologies and provide novel automated solutions for maintenance, repair, and operations (MRO), heralding significant transformations in future airport and aviation services.

Modern airports and aircraft are equipped with high-resolution cameras and 3D laser scanners, capturing detailed structural information to support vision-based maintenance tasks [4, 5]. Combined with the growing computational power of AI accelerators (GPUs and TPUs), this enables real-time diagnostics and predictive maintenance [6, 7]. However, varying lighting and complex visual conditions in airport environments pose challenges to the accurate segmentation of aircraft components. To address these issues, this study employs DeepLabV3 [8], a model integrating atrous convolution and spatial pyramid pooling, which effectively segments aircraft main components (AMC) under diverse conditions. Using pretrained ResNet101 weights, this model achieves high accuracy even with limited training data.

In the process of implementing automated aviation MRO, precise semantic segmentation is essential for accurately identifying and localizing key aircraft components, such as wings, fuselage, and engines. This accuracy is critical for early detection of defects, enabling targeted maintenance and reducing inspection time. Unlike traditional manual inspection, which is time-consuming and prone to human error, semantic segmentation offers a systematic approach to handling complex backgrounds and varying lighting conditions, ensuring consistency and reliability. However, high-quality and task-specific data is often severely limited in the aviation sector, a field with stringent demands for confidentiality and safety, which complicates the development of robust segmentation models [9]. Furthermore, the limited availability of labeled data and the common issue of class imbalance pose significant challenges, as they can lead to biased model predictions and hinder the recognition of critical minority classes [10]. This limitation makes it challenging for models

to learn and generalize effectively, thereby hindering their performance in real-world applications. Therefore, addressing data imbalance and label scarcity is crucial to achieving precise segmentation of AMC, which is not only essential for ensuring high safety standards but also for enhancing operational efficiency and supporting the implementation of Smart Airport and Airport 4.0 visions.

This study focuses on the semantic segmentation for commercial aircraft equipped with either twin or quad-jet engines, aiming to achieve precise semantic segmentation of AMC through advanced computer vision technologies. This study has adopted an architecture based on DeepLabV3 [8], which integrates deep convolutional networks, atrous convolution, and spatial pyramid pooling. This model is particularly suited for processing objects of varying sizes and shapes in images. By initializing the model with pretrained ResNet101 weights from ImageNet, this study leverages transfer learning strategies to quickly adapt to segmenting AMC, ensuring efficient learning and generalization capabilities even with limited training samples.

This study introduces a comprehensive approach for AMC segmentation through a task-specific DeepLabV3-based model and the creation of a novel AMC-Tr dataset with detailed pixel-level annotations. The dataset was developed using a rigorous image screening and annotation protocol that accounts for human bias, aircraft types, annotation error, and component representation, ensuring high-quality annotations for training, validation, and testing. Additionally, to address the categorical imbalance, a custom Focal Dice Loss (FDL) function was designed, combining the benefits of focal loss and dice loss to enhance sensitivity to minority categories and improve boundary recognition. This model leverages transfer learning with pretrained ResNet101 weights, enabling efficient learning even under limited data conditions. Together, these contributions significantly advance semantic segmentation performance for automated aircraft maintenance, supporting precise and robust visual recognition in challenging environments.

The contributions of this study are summarized as follows.

- This study employs a transfer learning strategy coupled with a DeepLabV3-based neural network to provide a novel solution for the automated semantic segmentation of aircraft main components (AMC). Utilizing pretrained ResNet101 weights from ImageNet, our model achieves rapid and efficient learning on a limited training dataset, significantly enhancing performance in aircraft component recognition tasks. As far as the authors are aware, this is the first successful solution for AMC semantic segmentation.
- Facing the challenge of task-orientated data availability and annotation scarcity, this study has designed and implemented a systematic protocol for image screening and pixel-level annotation. This protocol not only enhances the specificity of the data and consistency of the annotations but also reduces human errors through an independent review mechanism involving multiple researchers. The pixel-level annotated dataset has been made available to the academic community via CORD, providing valuable resources for subsequent related research, which can be accessed via: https://cord.cranfield.ac.uk/
- To address the challenge of categorical imbalance, this study successfully introduces the Focal Dice Loss (FDL) into the AMC semantic segmentation task. FDL effectively suppresses categorical imbalance and improves model performance when dealing with minority categories and complex visual content. This innovation enables the model to maintain high accuracy and robustness under challenging conditions, such as adverse weather and complex lighting scenarios.

The structure of this article is divided into six sections to present systematically an in-depth analysis of this study. Sect. 2: Related study reviews prior work related to the AMC semantic segmentation, providing a historical and theoretical foundation for the innovations and needs of this study. Section 3 depicts the dataset configuration. Section 4: Method details the data selection processes, image annotation protocols, and the design and implementation of the model architecture, clarifying the leading techniques and strategies employed in the experiments. In Sect. 5: Experimental Design, the article meticulously introduces the experimental setups, specific usage of datasets, and evaluation criteria to ensure the reliability and validity of the experimental results. This is followed by Sect. 6: Results and Discussion, where this study analyzes the experimental outcomes in detail, discusses the model performance and the effects of different loss functions, and delves into the reasons and potential impacts behind the results by comparing performance across different configurations. The article concludes with Sect. 7: Conclusion, summarizing the findings, discussing the limitations of the study, and suggesting possible directions for future work.

2 Related study

2.1 Aircraft main component

There is no unified definition for AMC, which varies depending on specific tasks. For example, Park et al. [11] identified the main components of a typical commercial aircraft as the fuselage, empennage, powerplant, wings, and landing gears, focusing primarily on common commercial jets. Figure 1 illustrates the main components of a typical commercial aircraft, highlighting the fuselage, empennage, powerplant,

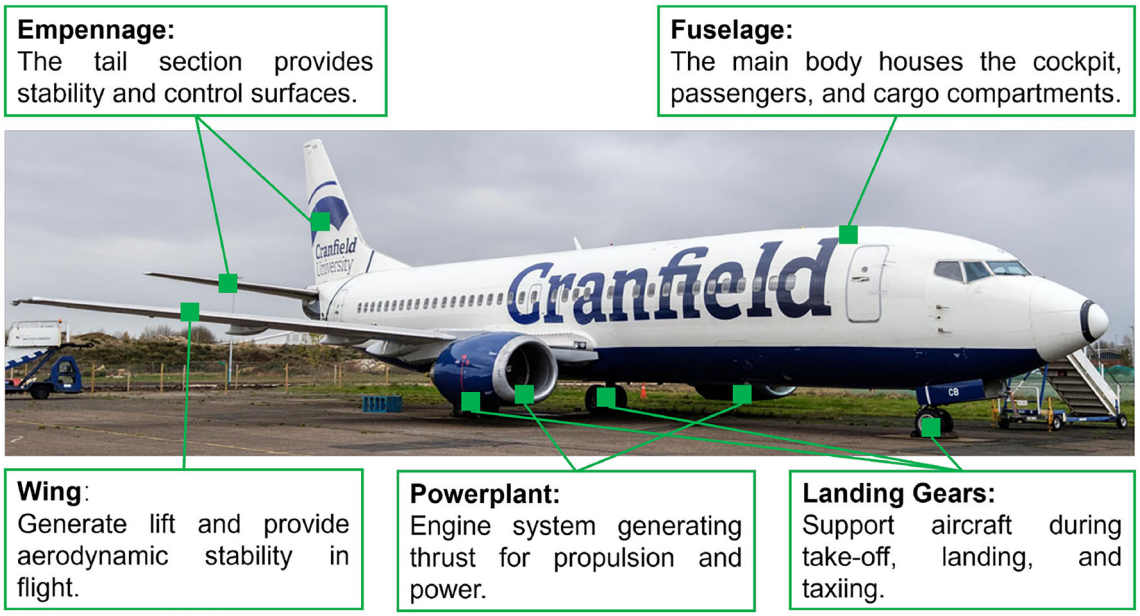


Fig. 1 The main aircraft components (AMC) and their functions

wings, and landing gears, along with their respective functions. Meanwhile, Khalid et al. [12] center on military or specialized aircraft, including the empennage, wings, cockpit, and fuselage as main structural components, excluding landing gears. This categorization reflects the specific focus on noncommercial aircraft, where landing gears are not a primary focus. Furthermore, Faisal et al. [13] categorize fighter jets into 12 distinct parts, and Hefazi et al. [14] divide a commercial plane into approximately 20 separate components, demonstrating the detailed segmentation in the highly complex aircraft systems.

This study primarily works on providing visual semantic information necessary for the autonomous operation of robots and other intelligent devices within intelligent and autonomous aircraft and airport operations. Clear and practical visual identification of aircraft components is essential to ensure these devices can accurately navigate to the correct aircraft positions. Therefore, this study adopts the AMC definition in Park et al. [11], renaming *powerplant* to engines to enhance its semantic intuitiveness. This decision is grounded in several considerations. First, the objectives of Park et al. [11] align with this study, both concentrating on commercial aviation. Second, the definition of aircraft types is suitable for commercial jets with twin and quad-jet engines used in this study. Lastly, selecting these components as the focus of the study ensures accuracy in annotation while reducing the complexity and labor costs associated with pixel-level labeling. This approach, driven by practicality and study objectives, not only provides the necessary theoretical and practical foundation for intelligent and autonomous aircraft and airport operations but also ensures high data quality and repeatability of the research through a consistent annotation protocol. These decisions reflect the complexity and precision required in applying machine learning and computer vision technologies within the aerospace field.

2.2 Semantic segmentation

Semantic segmentation has become a core technology in computer vision, enabling precise identification and categorization of objects in images. It builds on the fundamental concept of image segmentation but extends beyond mere region separation by associating each pixel with a meaningful class label, such as roads, vehicles, or buildings [15, 16]. The development of deep learning, particularly convolutional neural networks (CNNs), has revolutionized semantic segmentation by allowing models to learn complex feature hierarchies directly from raw images, significantly enhancing performance. Traditional image segmentation techniques relied heavily on handcrafted features and were limited in capturing high-level semantic information. In contrast, deep learning-based semantic segmentation techniques have demonstrated remarkable success in various visual recognition tasks, from autonomous driving to medical imaging, due to their ability to combine low-level features with high-level contextual information.

The evolution of deep neural networks has led to the emergence of several pioneering architectures specifically designed for semantic segmentation. The Fully Convolutional Network (FCN) [17] was the first architecture to repurpose CNNs for pixel-wise classification, replacing fully connected layers with convolutional layers to output dense

predictions. Building on FCN, more sophisticated models like U-Net [18] and SegNet [19] were introduced, incorporating encoder–decoder structures that allow for better preservation and refinement of spatial information. DeepLab [8], one of the most influential series, introduced atrous convolution and spatial pyramid pooling to capture multi-scale contextual features, achieving state-of-the-art results in several benchmark datasets. Recently, attention mechanisms and transformer-based architectures, such as the Vision Transformer (ViT) [20] and SegFormer [21], have further improved semantic segmentation by enabling long-range dependency modeling, thereby overcoming some of the limitations of convolutional layers in capturing global context.

Despite these advancements, semantic segmentation models face several critical challenges in real-world applications. One of the most prevalent issues is class imbalance, where certain categories, such as background regions, have far more samples than others (e.g., small objects). This imbalance can lead to biased predictions favoring majority classes and reduced accuracy for minority categories [22]. To address this, researchers have developed various loss functions, such as focal loss and weighted cross-entropy loss, which dynamically adjust the contribution of each class during training [23]. Another challenge is the precise segmentation of small or thin structures, which are often overlooked by standard CNN-based models. Techniques like boundary-aware refinement, multi-scale feature fusion, and context aggregation strategies have been proposed to improve the representation of these objects and enhance boundary delineation [24]. These methods ensure that models not only focus on large, prominent features but also maintain high precision on fine details.

A newer trend in semantic segmentation research is the integration of self-supervised learning and weakly supervised techniques to overcome data limitations. High-quality labeled data is often scarce and expensive to obtain, especially for specialized tasks like aircraft component segmentation. Methods leveraging self-supervised pretraining, such as contrastive learning [25], and weakly supervised learning using image-level labels or bounding boxes [26], provide effective solutions to this problem. By learning from large-scale unlabeled datasets or weak annotations, these techniques significantly reduce the reliance on dense pixel-level annotations while maintaining competitive segmentation performance. This has enabled the deployment of semantic segmentation models in domains where obtaining precise labels is challenging, thereby expanding the applicability of these models.

2.3 Transfer learning

Transfer learning has become a cornerstone in machine learning, enabling models to transfer knowledge from one domain to another and address the challenges of data scarcity. This approach involves various strategies depending on the relationship between the source and target tasks. A common technique is to share pretrained weights from models trained on large-scale datasets, such as ImageNet [27]. This method helps capture general visual features and allows models to adapt to new tasks with limited data. However, the use of pretrained weights is only one aspect of transfer learning. More advanced strategies include fine-tuning, where certain layers of the model are frozen while others are retrained, allowing for a balance between general feature retention and task-specific adaptation [28]. Researchers have also explored domain-specific transfer learning, where models are fine-tuned on a small set of target domain samples to improve performance in visually distinct domains, such as medical imaging and autonomous driving [29, 30].

The choice of transfer learning strategy significantly impacts model performance, particularly in visual recognition tasks. Sharing pretrained weights is effective for general feature extraction, but it may not capture fine-grained details in tasks such as semantic segmentation [31]. Therefore, a common practice is to use a hybrid approach: some layers, typically the initial layers, are frozen to retain generic features, while deeper layers are fine-tuned to adapt to the specific characteristics of the target task. This selective fine-tuning helps preserve the robustness of the pretrained model while enhancing its ability to distinguish subtle differences in the new task [32]. Another method is to employ a two-phase training strategy, where initial training is conducted with frozen layers, followed by end-to-end fine-tuning [33]. These techniques are particularly useful in applications where training data is imbalanced, or annotations are limited, as they reduce the risk of overfitting and ensure stable learning dynamics.

Recent studies have also explored more sophisticated forms of transfer learning, such as domain adaptation and multi-task learning. Domain adaptation methods aim to minimize the discrepancy between source and target domains, using techniques like adversarial training to align feature distributions across domains [34]. This approach is effective when there is a substantial difference between the visual characteristics of the source and target datasets. Multi-task learning, on the other hand, leverages shared representations across related tasks, improving generalization and robustness [35]. These techniques expand the applicability of transfer learning, making it a powerful tool not only for traditional vision tasks but also for specialized applications like aircraft component segmentation, where labeled data is scarce and expensive to obtain [36].

In this study, we adopt a transfer learning strategy using the DeepLabV3 model pretrained on the ImageNet dataset. By leveraging pretrained ResNet101 weights, the model captures high-level visual features that generalize well across

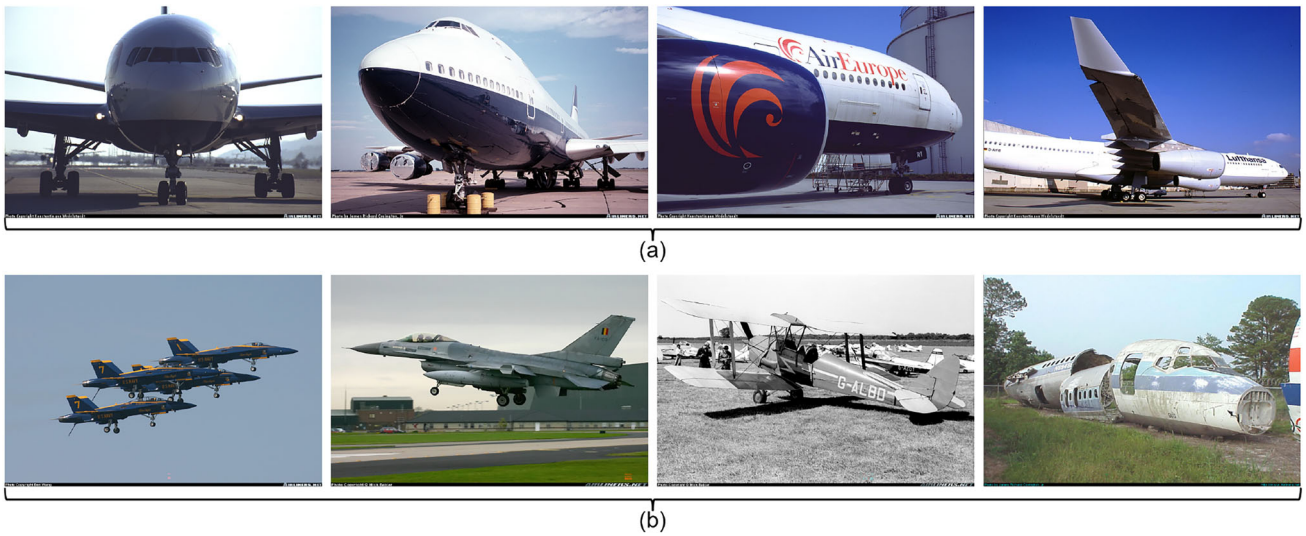


Fig. 2 Example images in FGVC-Aircraft dataset. **a** and **b** refer to various angles/distances images and irrelevant aircraft types, respectively

different contexts [8]. To adapt these features to the AMC segmentation task, the model is partially fine-tuned, with lower layers frozen to retain general patterns, while upper layers are retrained to identify aircraft-specific components. This strategy not only saves on computational costs and reduces the amount of required labeled data but also enhances the model's robustness and accuracy in identifying complex aircraft structures. Furthermore, by adjusting the learning rates of different layers during fine-tuning, we ensure a smooth adaptation process and prevent catastrophic forgetting, a common issue in transfer learning scenarios [37].

3 Dateset configuration

The primary challenge faced in this study is the scarcity of data available for the AMC segmentation task, which includes scene images targeted at AMC tasks and the corresponding pixel-level segmentation masks for AMC. Generating pixel-level masks is a labor-intensive and time-consuming process highly susceptible to human error [38]. Human annotators may overlook small or inconspicuous target objects, especially when images contain numerous objects or when some objects are obscured. The proximity of object categories or unclear annotation guidelines may lead to misclassification, causing category confusion. The accuracy of annotations is also constrained by the image resolution, particularly in marking highly detailed or small-sized objects, where insufficient resolution can prevent the precise delineation of complete boundaries of an object. Operational errors, such as accidentally deleting annotations or improper use of annotation tools, further introduce human errors [38]. It is essential to establish an annotation protocol to address these issues and enhance the precision and efficiency of the annotation process. This protocol should have consistent annotation standards, high repeatability, and be accurate and effective. This section introduces the datasets used as the data source and the established criteria for the segmentation mask annotations.

3.1 FGVC-aircraft dataset

The training and validation datasets utilized in this study are sourced from the Fine-Grained Visual Categorisation Aircraft (FGVC-Aircraft) dataset [39], first introduced in 2013 by Yang et al. titled *Fine-grained visual classification of aircraft* without AMC annotations. This dataset is commonly employed in aircraft-related computer vision tasks, as detailed in references [40, 41]. As depicted in Fig. 2a, the FGVC-Aircraft benchmark comprises images of aircraft captured from various angles and distances, many of which prominently display the main components of the aircraft. Thus, the FGVC-Aircraft dataset provides a rich repository of images tailored to the AMC-focused scenarios that this study addresses. However, the dataset includes over 100 types of aircraft, many of which do not fit the scope of this study, including images of fighter jets, wreckage, or small single-engine planes (as shown in Fig. 2b). The screening strategy outlined herein details the process for selecting targeted images from the FGVC-Aircraft dataset to align with the specific requirements of the segmentation task.

3.2 Semantic screening strategy

The objective of the image screening strategy is to select imagery from the FGVC-Aircraft dataset that is relevant

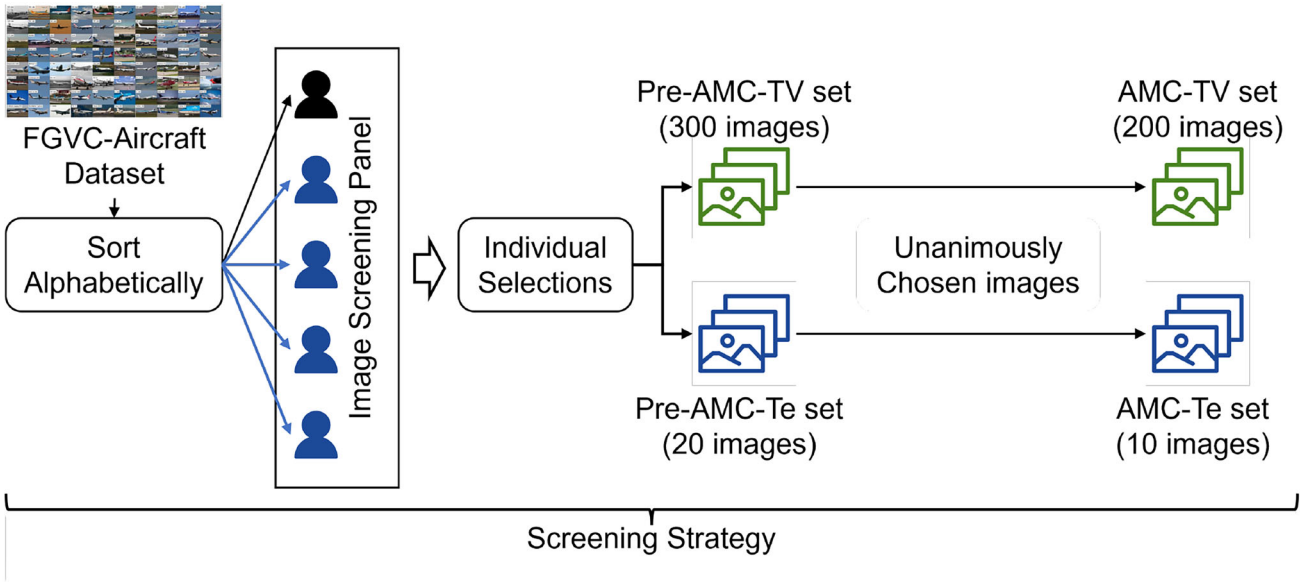


Fig. 3 The image screening strategy

to the AMC segmentation task. This task provides visual semantic understanding for automated and robotic systems operating within Smart Airport and Airport 4.0 visions. The image selection process considers four main factors: the visibility of AMC, aircraft type, the volume of images, and the risk of human bias. First, a more extensive image set (AMC-TV set) is chosen for model training, which includes images that clearly and comprehensively display AMC. One smaller dataset is selected to test the generalizability performance of the model under difficult weather and lighting conditions, which is named AMC-Te, while *Te* stands for various weather and lighting conditions for testing. Second, this study focuses on modern and new models of commercial jets. There are no restrictions on the manufacturer or model, and only the overall distribution of components is essential. An exception is considered in the selection of datasets, where an aircraft appears as a typical twin-engine model but has an engine mounted on the vertical stabilizer—a design combining rear-mounted and wing-mounted engines. [42] This study considers this aircraft type no longer suitable for modern aviation. Thus, images of such aircraft designs are excluded. Thirdly, the AMC-TV set comprises 200 images, while the AMC-Te set contains ten images. Fourthly, to minimize human bias in image screening, a selection process involving five independent researchers as panel members with sufficient background in aviation and image processing is implemented, where images selected by all five panel members are included in the final image set.

The steps in the image screening strategy are as follows, illustrated in Fig. 3:

step 1. Sort the FGVC-Aircraft dataset alphabetically.

step 2. Begin selection based on the image screening factors mentioned above. Each panel member selects 300 candidate images for the pre-AMC-TV set in the order of the images.

step 3. Compare the images selected by the five panel members, and those unanimously chosen are added to a final AMC-TV set.

step 4. Further, for the AMC-Te set, which is designed for testing in different visual conditions, including challenging conditions, select 20 images with different weather and illumination conditions for the pre-AMC-Te set, skipping any images already included in the AMC-TV set.

step 5. Compare the pre-AMC-Te set images selected by the researchers; those unanimously chosen are included in the final AMC-Te set, and the inclusion will stop when the AMC-Te set contains 10 images.

This meticulous strategy ensures the relevancy and quality of the images used and supports the accuracy and efficacy of the segmentation tasks in automated airport systems.

3.3 The pixel-level segmentation mask annotation protocol

The annotation protocol delineated herein is essential for the precise semantic annotation of AMC masks and ensures rigorous standardization across all actions and rules within the dataset (as shown in Fig. 4). The protocol consists of four rules, and the structural approach prevents human error and ensures the uniformity of annotations, overcoming variations that may arise from ambiguous junctures or aircraft model-

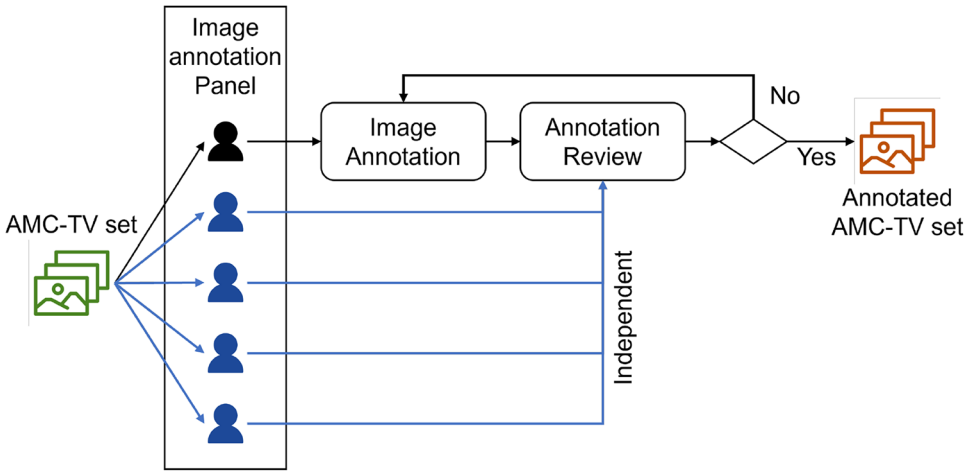


Fig. 4 The annotation protocol

specific nuances. These guidelines are integral to maintaining the high quality and replicability required for deploying machine learning and computer vision technologies in aviation applications.

Rule 1. A dual review pipeline is stipulated, where the initial annotation is followed by a critical review process by four independent assessors. The central question posed during the review phase is whether the pixel-level annotation is precise enough. A majority consensus either confirms the annotation or mandates re-annotation, with a maximum allowance of three attempts. Failure to achieve consensus after these attempts results in the exclusion of the image, which indicates that the image is too complex for human expert discernment. Notably, while machine learning models that outperform human expertise represent an engaging avenue for future research, they fall outside the ambit of this particular study.

Rule 2. The selections and annotations for the AMC-TV set cease when their image counts reach 200 images, respectively. The AMC-Te set is not annotated because of the difficult images included, which can cause significant human error in the annotated masks.

Rule 3. This rule is motivated by the operational context within intelligent airports, where proximity between cameras and aircraft is typical, thus rendering distant aircraft segmentation less significant. When aircraft in the background is remote and occupy small pixel areas, they are uniformly categorized as 'background', thereby simplifying the segmentation process. However, in cases where multiple aircraft are in proximity and constitute substantial pixel regions, all aircraft components must be annotated to preserve the semantic clarity of the visual data.

Rule 4. The AMC in this study are categorized into four categories—engines, wings, fuselage, landing gear, and empennage— and provide a precise mask annotation protocol for each:

(a) Engines, typically jet engines or turbines, are affixed to the wings via pylons. In the annotation protocol, the pylon is not isolated but considered an extension of the wing, recognizing the absence of clear demarcation and the shared attributes such as paint and materials.

(b) The interface between wings and fuselage is invariably labeled as part of the fuselage to promote consistency across the dataset. This policy effectively counters the potential variability introduced by diverse perspectives and illumination conditions, thereby safeguarding the accuracy of data analysis.

(c) The empennage requires meticulous differentiation, particularly of the junctions of the vertical and horizontal stabilizers with the fuselage. The annotation protocol mandates that these boundaries be well defined for the detailed identification and segmentation of these structures.

(d) The landing gear, typically obscured by the shadow of the aircraft, is annotated only for its visible segments. In cases where ground obstructions partially conceal the gear, the annotations are curtailed accordingly to exclude nonrelevant elements.

(e) Lastly, as previously discussed, the fuselage annotation must outline all visible fuselage sections as accurately as possible.

These carefully formulated guidelines not only equip the autonomous systems of smart airports with a robust dataset but also underpin the potential advancements in automated

aviation MRO processes. By adopting such stringent protocols, this study endeavors to set a precedent for precision and reliability in applying advanced computational techniques to aerospace semantic segmentation.

4 Method

4.1 Preprocessing

Preprocessing is a fundamental step in machine learning solutions. It transforms raw data into a format that is more amenable to learning, thereby enhancing the quality, efficiency, and performance of the model. The preprocessing primarily involves one-hot mask encoding, dataset partitioning, image enhancement, and image normalization.

The masks in this study encompass six categories, including the background, engine, wing, empennage, landing gear, and fuselage, distinguished by integers during the annotation process. Since AMC semantic segmentation is a pixel-level multi-class task, this study has employed one-hot encoding for different category masks. The one-hot encoding provides clear categorical separation in multi-class mask segmentation tasks, encoding each category into separate channels to reduce confusion between categories and enhance the classification accuracy of the model. Moreover, one-hot encoding allows the model to learn and distinguish the specific features of each category more effectively, thus improving the accuracy of the segmentation.

For dataset partitioning, the AMC-TV set and AMC-Te set are subdivided and reorganized for training, validation, and testing purposes. After shuffling, the AMC-T set is divided into 160 images for training (AMC-Tr set), and 40 for validation (AMC-Va set). The AMC-Tr set is used for model training, the AMC-Va set is used for validation during training, and the AMC-Te set is used for independent testing. Furthermore, the AMC-Te includes both normal scenarios similar to AMC-Tr/Va sets, while some difficult scenarios are also included to test the robustness under difficult visual conditions.

Data augmentation is essential in machine learning and computer vision, significantly enhancing model performance and generalization capability. Data augmentation involves strategically manipulating existing training data to generate variants without introducing new samples. By applying carefully designed transformations, such as rotation, flipping, cropping, and modifications to color and lighting, to the original datasets, data augmentation aims to expand the effective size of the training dataset. Traditional data augmentation involves pre-generating enhanced versions of the dataset and saving them to disk, significantly limiting the number of possible enhancements to the available disk space. Moreover, each enhanced instance must be loaded during training, which can slow down the training process. This work dynamically applies data augmentation, with transformations randomly applied to each training batch. Each transformation combination is different, ensuring the model "sees" different versions of each instance in every training set (as shown in Fig. 5).

Enhancements are equally applied to the corresponding masks. This study has developed a data augmentation pipeline using the Albumentations package, passed as parameters to the data loader handling dynamic image transformations. The data augmentation transformation pipeline consists of three geometric transformations and four photometric transformations:

- (1a) Random horizontal and/or vertical shifts, up to 5% of the image size.
- (1b) Random scaling of images, with a maximum scale factor of 10%.
- (1c) Random rotation of images within a range of -10 to 10 degrees.
- (2a) Random RGB shifts for each channel, up to 25.
- (2b) Random brightness and contrast enhancement of 30%.
- (2c) Random brightness and contrast reduction of 30%.
- (2d) Random Gaussian blur.

4.2 AMC segmentation model design and training strategy

This study adopts the DeepLabV3 architecture as the backbone model for precise semantic segmentation of aircraft main components (AMC) due to its proven effectiveness in complex segmentation tasks. DeepLabV3 has been widely used in various domains because of its ability to handle intricate object shapes and boundaries while maintaining high segmentation accuracy [8]. The key strength of DeepLabV3 lies in its use of atrous (dilated) convolutions, which enable a larger receptive field without losing resolution, allowing it to capture multi-scale contextual information. This capability is particularly important for AMC segmentation, where components vary significantly in size and spatial arrangement. The ability of DeepLabV3 to retain both local and global features makes it a robust choice for segmenting aircraft components that exhibit complex geometric characteristics, such as curved wing edges and detailed fuselage features.

Furthermore, the Atrous Spatial Pyramid Pooling (ASPP) module in DeepLabV3 enhances its ability to capture objects at varying scales, which is essential for accurately segmenting aircraft components that may appear differently due to changes in perspective or distance within the image. By employing multiple dilation rates, the ASPP module aggregates contextual information at different resolutions, enabling the model to segment large structures like wings

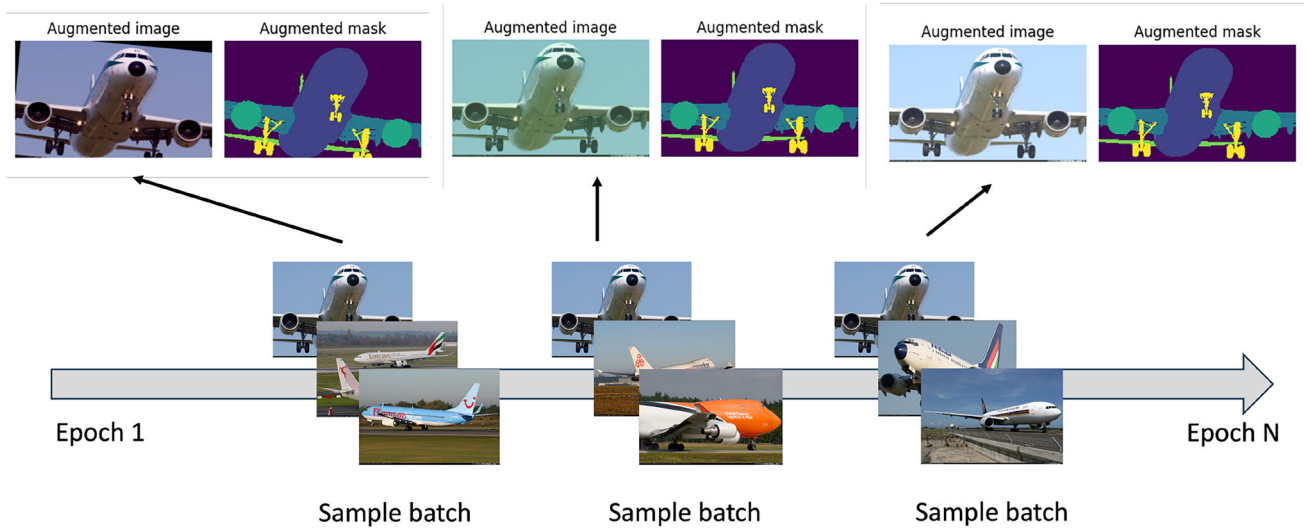


Fig. 5 The dynamic data augmentation process

and smaller components like landing gears simultaneously. Given the variability in the scale and shape of AMC, the architecture of DeepLabV3 is highly suited to maintaining boundary precision while preserving semantic information. Besides, the flexibility to adapt through minor adjustments, such as customizing the output layer for specific class labels, ensures its compatibility with the proposed AMC-Tr dataset. These considerations make DeepLabV3 an ideal foundation for developing an effective and specialized AMC segmentation model.

This study adopts a transfer learning strategy to enhance the model segmentation performance on the AMC semantic segmentation task. The model backbone network is initialized using pretrained ResNet101 weights from the ImageNet dataset, leveraging the low-level image features learned from large-scale image classification tasks. This approach allows the model to effectively learn on the AMC-Tr dataset, which consists of only 160 images—even after data augmentation, the dataset only expands to 1280 images, which is insufficient for deep learning training from scratch. Through transfer learning, the model quickly adapts to the new task, accelerates the training process, and swiftly converges on the specific semantic segmentation task. This training strategy not only saves computational resources but also enhances the generalization ability of the model across new datasets, enabling it to extract underlying patterns and representations beyond a single dataset more effectively. Besides, to prevent overtraining, this study has implemented an early stopping strategy based on the performance of the validation dataset, and it avoids the risk of overfitting. This implementation of transfer learning and early stopping strategy optimizes the semantic segmentation of aircraft components, which is essential to achieving efficient and precise segmentation results in our research.

4.3 Loss function design

A loss function called Focal Dice Loss (FDL) was specifically designed to address the category imbalance in the AMC semantic segmentation, combining the focal loss mechanism and dice loss. In the AMC segmentation, the number of pixels belonging to the background mask significantly outweighs those of other categories. This may cause the model to predominantly predict most pixels as background, neglecting other crucial components at the training stage. Traditional cross-entropy loss, commonly used in semantic segmentation tasks, can penalize misclassified pixels but is limited under severe category imbalance scenarios (see Eq. 1). To overcome this limitation, dice loss was introduced, which focuses on the similarity between the predictions and actual labels, especially in terms of overlap (see Eqs. 2 and 3). The advantage of dice loss is that it does not directly rely on the number of pixels per category; thus, it is more suitable for handling minority categories. However, dice loss may under-penalize the misclassification of small objects, which can be problematic when segmenting smaller components like landing gears. To further optimize the loss function, the weight design of the focal loss (see Eq. 4) was integrated into the FDL. Focal Dice Loss adjusts the focus parameter (γ) and category weights (w_i, see Eq. 5) to enhance penalties on difficult-to-classify or frequently misclassified samples (for example, samples with a lower Dice Similarity Coefficient, DSC) and reduce penalties on well-classified samples (see Eq. 6). This design helps manage category imbalance and focuses on difficult-to-segment areas, including the boundaries between wings and the background. FDL provides a robust tool to tackle specific challenges in the semantic segmentation of AMC, optimize the training process, and improve the accuracy and

generalization of segmentation.

$$L(y, \hat{y}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i), \tag{1}$$

$$\mathrm{DSC} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{DL} = 1 - \mathrm{DSC} = 1 - \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{3}$$

$$\mathrm{FL} = -\alpha(1 - p_t)^{\gamma} \log(p_t) \tag{4}$$

$$w_{ic} = \frac{1}{N_{ic}}, \quad \text{where } N_{ic} \text{ is the number of pixels of category} c \text{ in image } i. \tag{5}$$

$$\mathrm{FDL} = \sum_{i} w_i (1 - \mathrm{DSC}^{1/\gamma}), \quad \text{where } w_i \text{ is accumulation of category } c. \tag{6}$$

5 Experimental design

This study establishes a uniform experimental environment to evaluate the effectiveness of various loss functions, input image sizes, and models for semantic AMC segmentation. The experimental setup included a computer running Ubuntu 20.04, equipped with an Intel Core i7-13700H processor, an Nvidia RTX 4090 GPU with 24GB of VRAM, and 64GB of system memory. To ensure the accuracy and repeatability of the experiments, the specifications of the software are PyTorch (version 2.1.0 + cu118), OpenCV (version 4.8.1.78), pandas (version 2.1.1), Albumentations (version 1.4.3), and Matplotlib (version 3.8.2). During the experiments, the models were trained using the Adam optimizer from Pytorch with a learning rate set to 0.0001 and a batch size of 32 images per batch. Considering efficiency in training and preventing overfitting, an early stopping mechanism ceased training when the validation loss continued to increase over 20 consecutive epochs, with a maximum limit of 1000 epochs set for the training period. This study systematically explores various experimental settings, including different loss functions, input image sizes, and models, to assess the impact of these variables on the performance of the AMC semantic segmentation task (as detailed in Table 1). By comparing the experimental results, this study aims to identify the optimal model configurations and parameters to enhance the performance of ACM segmentation and the practicality of the model.

Table 1 Experiment design for AMC semantic segmentation

Index	Group	Loss	Image size (pixels)	Model
Ex-1	A/B/C	FDL	256 × 256	DeeplabV3
Ex-2	A	CE	256 × 256	DeeplabV3
Ex-3	A	DL	256 × 256	DeeplabV3
Ex-4	A	FL	256 × 256	DeeplabV3
Ex-5	A	CE-DL	256 × 256	DeeplabV3
Ex-6	B	FDL	128 × 128	DeeplabV3
Ex-7	B	FDL	512 × 512	DeeplabV3
Ex-8	C	FDL	256 × 256	FCN
Ex-9	C	FDL	256 × 256	LRASPP

Table 1 presents nine experimental setups conducted in this study, further divided into three groups—A, B, and C—for straightforward comparative analysis. Group A consists of Ex-1, Ex-2, Ex-3, Ex-4, and Ex-5, designed to assess the impact of various loss functions on AMC segmentation. The loss functions experimented with include the proposed *Focal Dice Loss* (FDL) as well as *Cross-Entropy* (CE), *Dice Loss* (DL), *Focal Loss* (FL), and a Combined Loss between *Cross-Entropy* and *Dice Loss* (CE-DL). Group B comprises Ex-1, Ex-6, and Ex-7, which explore the effect of different input image sizes on AMC segmentation. Ex-1, Ex-6, and Ex-7 utilize medium, small, and large image sizes, respectively, with specific dimensions of 256 × 256, 128 × 128, and 512 × 512 pixels. Group C includes Ex-1, Ex-8, and Ex-9, focusing on examining the influence of different segmentation models. The models evaluated are the *Fully Convolutional Network* (FCN) and *Lite Reduced Atrous Spatial Pyramid Pooling* (LRASPP). This structured experimental design facilitates a thorough evaluation of the impact of each variable and enhances the precision and applicability of the segmentation models in discerning AMC accurately.

6 Results and discussion

6.1 Aircraft main components segmentation dataset

This section extensively analyzes the structure and challenges of the proposed AMC segmentation dataset, delving into the characteristics of the dataset and the issues based on sample distribution and statistical analysis. First, the sample image and annotation results shown in Fig. 6 clearly illustrate the condition of occlusions within the AMC structure, which affects the complete display of components and increases the complexity of the segmentation task. The partial or complete occlusion of structures, particularly where the engine obscures the wing, poses a significant challenge to the precision of segmentation algorithms. Second, the chal-

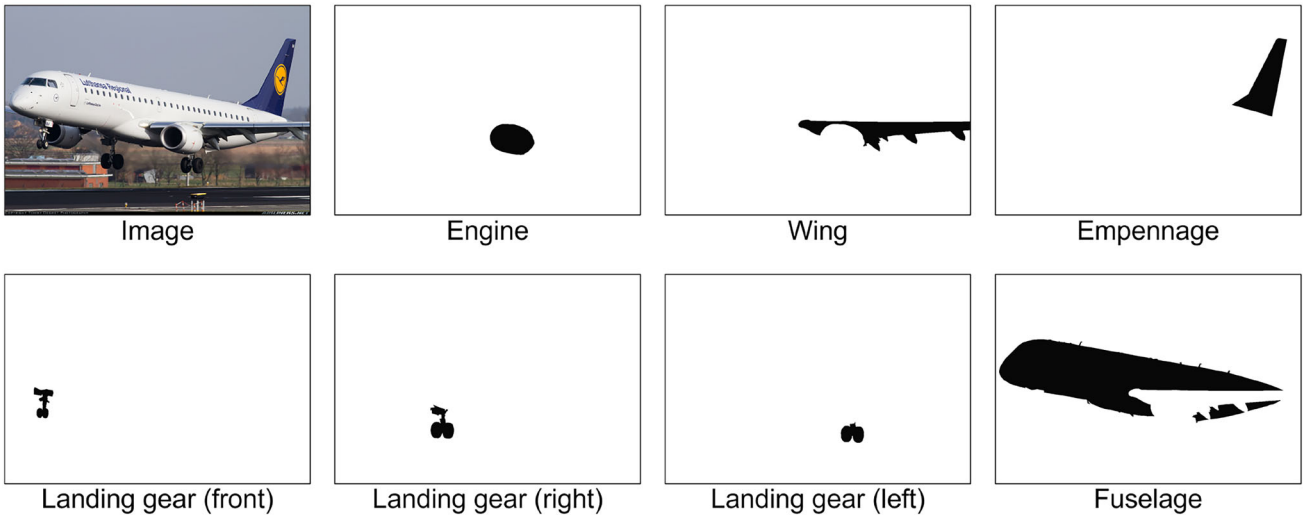


Fig. 6 The annotations of the sample image from AMC-Tr set

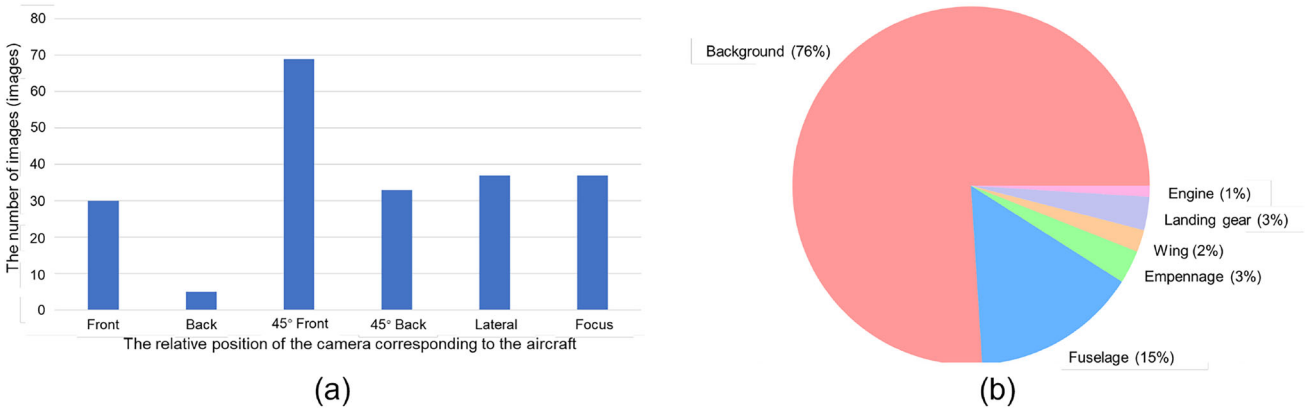


Fig. 7 The statistic analysis of the AMC dataset. **a** and **b** refer to the relative position of the camera corresponding to the aircraft and pixel ratio distributed by categories, respectively

lenge of category imbalance within the dataset is further confirmed in Fig. 7b, where the background mask accounts for 76% of the pixels, whereas the engine mask constitutes only 1%. This pronounced imbalance highlights the need for focused strategies on handling minority categories to prevent the model from overfitting to the majority category while neglecting minority-category recognition. In practical applications, such as automated equipment in smart airports, accurate identification of minority-category AMC is crucial. Third, the distribution of viewing angles shown in Fig. 7a also reveals strategic considerations in data collection, with the 45-degree front view having the highest sample volume, aligning with the typical observational needs of automated devices in airports to ensure precise capture of AMC from multiple directions. This strategy reflects the operational demands and enhances the practicality and effectiveness of the AMC segmentation dataset.

6.2 Image segmentation

This section delves into the AMC segmentation results through both quantitative and qualitative experimental results. First, data from Table 2 reveals significant variations in segmentation outcomes across different experimental setups. Ex-1, employing FDL, demonstrates the highest IoU values on the validation set, indicating superior performance over other loss functions in AMC segmentation. Although Ex-5 achieved the highest validation accuracy, the results depicted in Fig. 8 suggest this high accuracy came from an increased prediction boundary of the AMC, a phenomenon also observed in Ex-2, hinting at potential misclassification issues due to the cross-entropy component in the combined loss function. Moreover, Ex-4, while showing the lowest values in training and validation losses, does not perform well qualitatively, suggesting that a low numerical value of the loss does not directly correlate with high segmentation results. In

Table 2 Experiment results for AMC semantic segmentation

Index	Group	$\text{Loss}_{\text{train}}$	$\text{Loss}_{\text{valid}}$	$\text{Accuracy}_{\text{valid}}$	$\text{IoU}_{\text{valid}}$	Epoch_{es}
Ex-1	A/B/C	5.046×10^{-3}	8.327×10^{-3}	83.420%	75.268%	263 epochs
Ex-2	A	105.832×10^{-3}	317.089×10^{-3}	89.831%	66.798%	68 epochs
Ex-3	A	21.983×10^{-3}	38.429×10^{-3}	83.642%	74.985%	261 epochs
Ex-4	A	0.018×10^{-3}	0.071×10^{-3}	74.854%	22.952%	33 epochs
Ex-5	A	61.273×10^{-3}	198.208×10^{-3}	90.113%	70.622%	119 epochs
Ex-6	B	8.294×10^{-3}	12.981×10^{-3}	74.538%	65.992%	477 epochs
Ex-7	B	2.389×10^{-3}	5.233×10^{-3}	91.466%	84.002%	106 epochs
Ex-8	C	3.773×10^{-3}	7.427×10^{-3}	84.919%	75.656%	191 epochs
Ex-9	C	9.662×10^{-3}	13.696×10^{-3}	63.662%	56.237%	429 epochs

Group B, Ex-7, which increased image resolution, displays better quantitative and qualitative outcomes than Ex-1 and Ex-6, underscoring the significant impact of image resolution on segmentation results. However, higher resolutions imply more significant computational costs and longer training times, necessitating a balance in practical applications. For Group C, although the quantitative results of Ex-8 are close to those of Ex-1, the mask edges in Ex-8 are more irregular. A slightly higher accuracy indicates a better background pixel prediction in Ex-8, which does not entirely align with the primary goals of AMC semantic segmentation. Thus, it can be concluded that FDL effectively mitigates the category imbalance caused by a predominance of background masks, and where computational power and time permit, increasing image resolution substantially enhances the performance of AMC semantic segmentation. This thorough analysis advances the understanding of various experimental factors on AMC segmentation and provides valuable guidance for future algorithm improvements and experimental designs.

The curves depicted in Fig. 9a represent the training trajectory of Ex-1, showcasing a rapid initial descent in loss values indicative of the capability of the model to learn and adapt its weights to the dataset swiftly. As epochs progress, the decline in loss decelerates, ultimately stabilizing, which suggests the approach of the model to convergence and robust understanding of the data. The modest gap between the training and validation loss curves denotes the proper generalization of the model, a crucial marker for circumventing overfitting. Concurrently, qualitative predictive results infer that the model demonstrates substantial segmentation performance against normal and clear backgrounds, accurately delineating each AMC as demonstrated in the first row of Fig. 9b. However, precision falters as backgrounds complexify, particularly in scenarios with shadow occlusions, as observed in the second row of Fig. 9b. This could stem from the limited adaptability of the model to environmental changes, especially sensitivity to shadow and illumination conditions variations. Furthermore, despite the satisfied predictive performance on frontal components, room for refinement persists in the detailed treatment of regions, notably at the fringes of the wings (as shown in the third row in Fig. 9b).

Figure 10a and b presents the accuracy and IoU curves for individual sub-categories, respectively. Analysis of the curves displayed in these figures reveals distinct characteristics in recognizing the model of various AMC throughout the training process. First, a rapid improvement in both accuracy and IoU for all categories indicates the swift learning of key features distinguishing each AMC. As the training progresses, the rate of increase for these metrics decelerates and ultimately plateaus, suggesting that the model is approaching its performance limit. The background category consistently shows the highest accuracy and IoU, likely reflecting its predominant presence and more straightforward features within images. In contrast, the accuracy and IoU for engines, empennages, and landing gears are relatively lower, indicating challenges encountered when processing these smaller, more complex parts and pointing to potential areas where model optimization could enhance performance. Notably, improving the recognition accuracy of these components, particularly under complex backgrounds or occlusions, is of significant importance. These observations underscore the necessity of devising customized strategies tailored to different aircraft parts to ensure precise segmentation across all conditions.

Table 3 provides an insightful evaluation of the impact of varying the γ parameter in the FDL on the segmentation task, and Figure describes the predictions of different γ values. A clear trend is observed: as γ increases from 1 to 4, there is a notable decrease in training loss, indicating a more focused training process on challenging samples. Validation loss follows a similar trend, albeit with a slight uptick for $\gamma = 3$, which suggests an optimal range for γ may exist that balances attention to complex samples without overfitting. Interestingly, the validation accuracy and IoU gradually increase with higher γ values, with the highest IoU of 75.992% at $\gamma = 4$, despite more epochs needed. This could imply that an increased γ helps the model better delin-

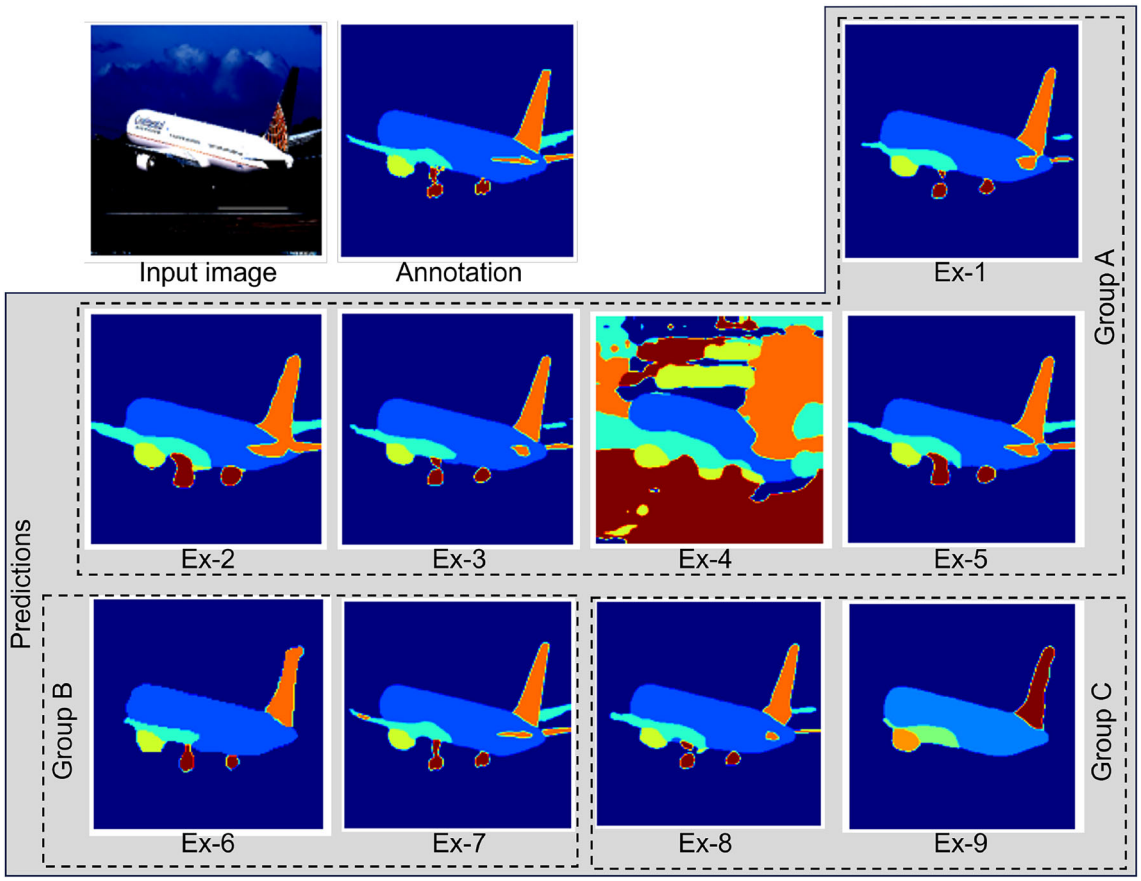


Fig. 8 The segmentation predictions of all experiments in Table 1

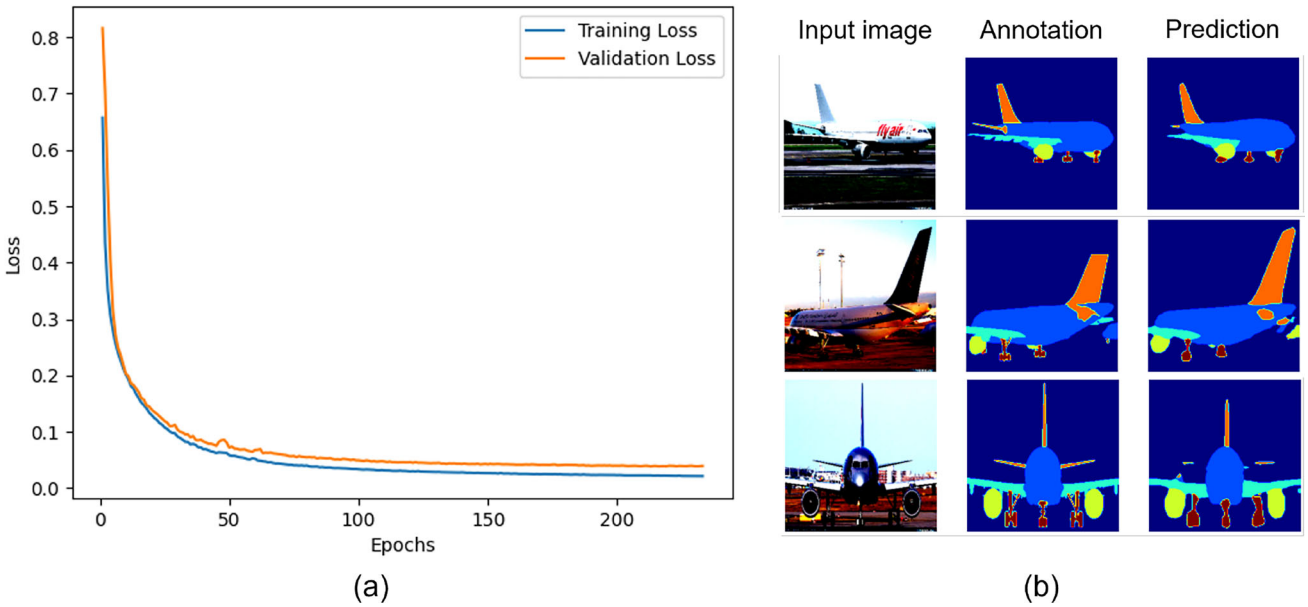


Fig. 9 The training and validation loss curves in the AMC-Tr set and the AMC-Va set

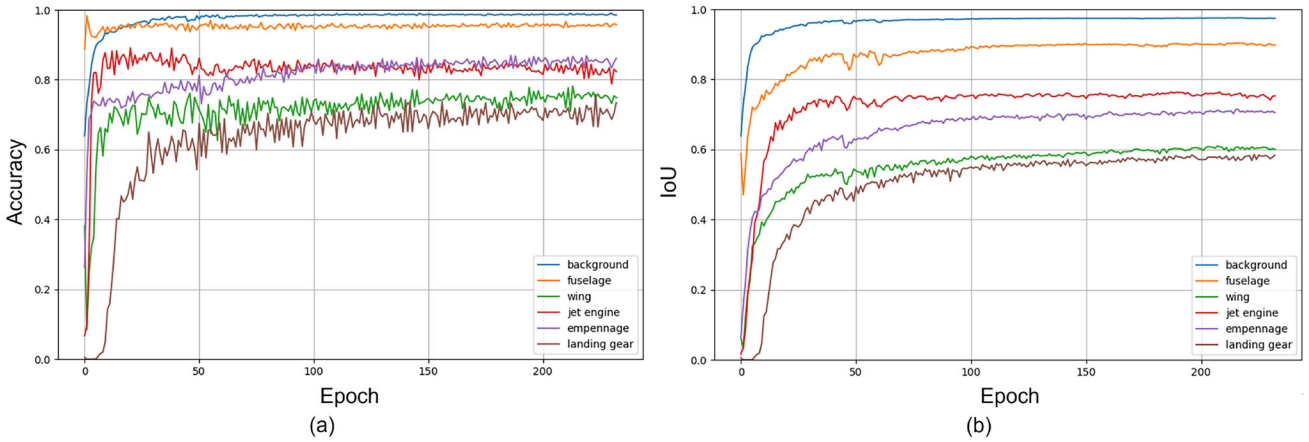


Fig. 10 The validation accuracy and IoU curves in the AMC-Va set. **a** and **b** refer to the training curves and predictions in the AMC-Va set, respectively

eate between the segments, improving the overall precision of segmentation, albeit at the cost of extended training time. The trend highlights the delicate balance between focusing on hard-to-classify pixels and achieving general segmentation accuracy, showcasing the importance of fine-tuning the γ parameter within the loss function for optimal model performance.

6.3 Further tests for difficult visual conditions

This section deliberates on the performance of the proposed AMC segmentation dataset across varying visual conditions. Distribution samples from the AMC-Te set, as depicted in Fig. 12, span a spectrum of challenging visual scenarios. Conversely, the AMC-Te set underscores exceedingly difficult visual settings, which pose significant challenges even for human observers, thus precluding manual annotation to avoid high-error labeling due to extreme conditions.

Predictive results from the AMC-Va dataset demonstrate proficiency in accurately recognizing and segmenting AMC (as shown in Fig. 10b). This corroborates the efficacy of the advanced AMC segmentation approach under optimal lighting without complex interferences. However, arduous conditions reflected in the AMC-Te dataset, such as lens blurring from rain, impose considerable impediments to the segmentation task, often precluding any reliable structure recognition (as shown in Fig. 13). Such instances illustrate the performance limitations of even sophisticated segmentation models under extreme conditions. Nonetheless, it is noteworthy that the model exhibited robust performance under suboptimal lighting, glare, dense fog, and low visibility, albeit with less detailed precision than in normal lighting scenarios. These observations affirm the efficacy of the transfer learning strategy. When confronted with challenging visual environments, the pretrained weights endow the model with the essential knowledge to tackle segmentation tasks, ensuring effectiveness and adaptability across variable conditions.

7 Conclusion

This study advances the field of automated AMC semantic segmentation through a tailored DeepLabV3-based model, rigorous annotation protocols, and innovative loss function design. The introduction of the novel AMC-Tr dataset, created using a meticulous image screening and annotation strategy, addresses data limitations and ensures high-quality pixel-level annotations. A key contribution is the custom Focal Dice Loss (FDL) function, which effectively handles category imbalance, improving the model's sensitivity to minority components like engines and landing gears. Leveraging transfer learning with pretrained ResNet101 weights, the model achieves robust performance under constrained data conditions, enabling precise segmentation of complex aircraft structures and advancing automated MRO capabilities.

The empirical results of this study demonstrate a notable high segmentation IoU of 84.002% and an accuracy of 91.466%, validating the efficacy of the proposed segmentation approach. These outcomes signify a leap in performance and cement the practical utility of the model in intelligent and autonomous aircraft and airport operations. The impact of these findings sets a precedent for future research in the domain and offers a rich repository for continued exploration and development.

However, while the achievements are significant, the research does not escape limitations. For instance, under extreme visual conditions like lens blurring from rain, the segmentation task faces substantial hurdles, revealing performance bottlenecks of the current model. In particular, small

Table 3 Experiment results for Ex-1 setting with different γ values in FDL

Index	γ	$\text{Loss}_{\text{train}}$	$\text{Loss}_{\text{valid}}$	$\text{Accuracy}_{\text{valid}}$	$\text{IoU}_{\text{valid}}$	Epoch_{es}
Ex-1a	1	22.400×10^{-3}	38.200×10^{-3}	83.524%	75.172%	213 epochs
Ex-1	2	5.046×10^{-3}	8.327×10^{-3}	83.420%	75.268%	263 epochs
Ex-1b	3	7.112×10^{-3}	12.732×10^{-3}	84.309%	75.413%	283 epochs
Ex-1c	4	5.066×10^{-3}	9.277×10^{-3}	84.864%	75.992%	338 epochs

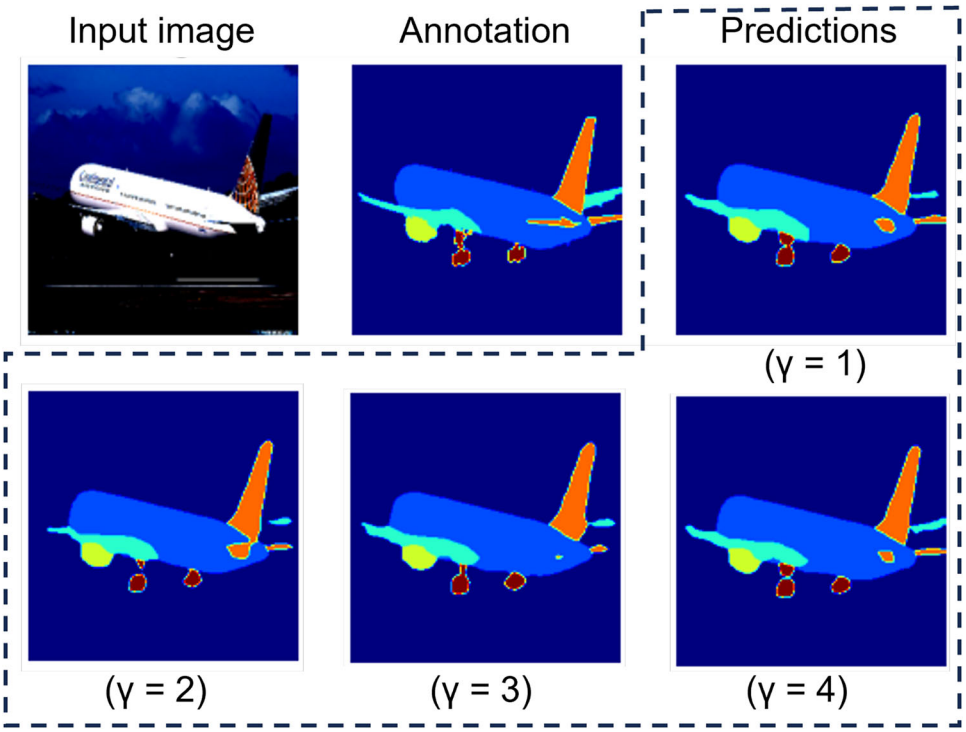


Fig. 11 The predictions of different γ values

Fig. 12 The image distribution by weather and illumination conditions in the AMC-Te-D set

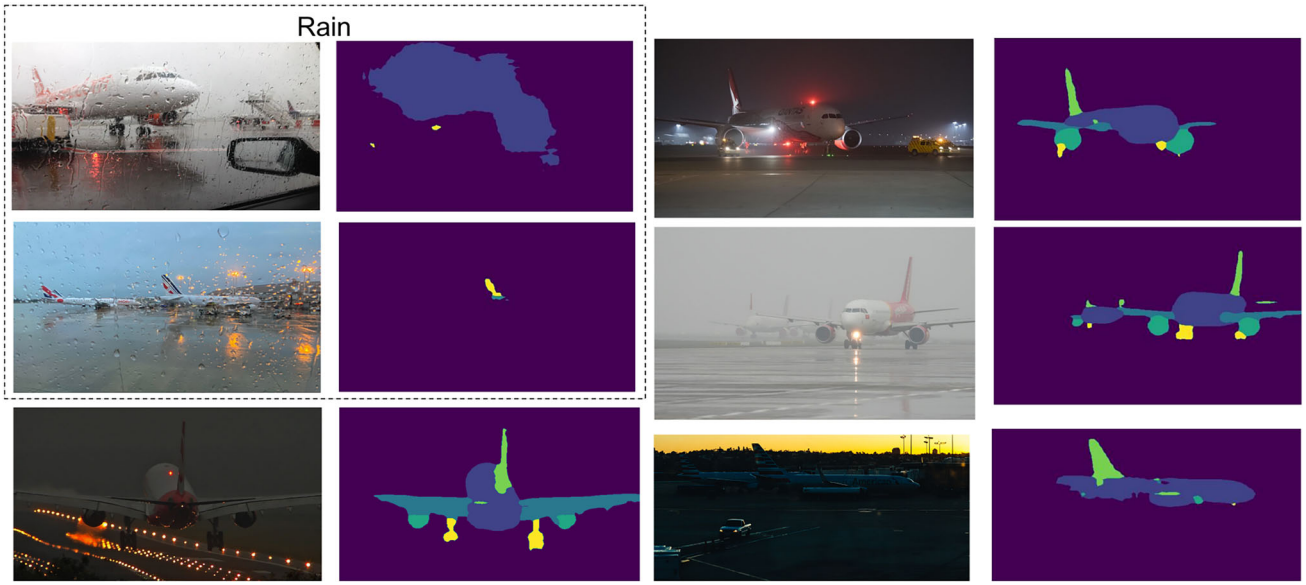


Fig. 13 The predictions in the AMC-Te set

objects, such as landing gears and certain aircraft components, remain challenging to segment accurately due to their limited pixel size and complex backgrounds. As discussed in the literature, small objects pose unique difficulties in image recognition, primarily because they occupy a relatively small area within the image and are often overshadowed by larger objects, resulting in lower detection and segmentation accuracy [43]. This issue is further compounded by the imbalance between small and large objects, as noted in maritime and aerial applications [44]. Future work will focus on addressing these limitations by incorporating advanced small object detection strategies and integrating novel techniques to enhance robustness and adaptability in challenging environments.

One potential direction is to incorporate uncertainty modeling to improve the model's reliability in uncertain scenarios, as explored in recent studies using deep learning-based uncertainty-aware methods [45]. This approach can provide better confidence estimates for the predictions and help in identifying areas where the model is prone to errors, especially under visual disturbances. Moreover, combining the current framework with 3D modeling can enhance spatial understanding, allowing the model to better segment components with complex geometric structures. Expanding the dataset to include more diverse weather and lighting conditions, as well as incorporating real-time learning capabilities, will be crucial for creating a more dynamic and responsive system. These advancements will contribute toward achieving the broader Smart Airport and Airport 4.0 visions by providing a more comprehensive solution capable of handling complex scenarios with greater accuracy and robustness.

Author Contributions Julien Thomas and Boyu Kuang were involved in conceptualization, investigation, methodology, resources, validation, visualization, writing—original draft; Julien Thomas, Boyu Kuang, Stuart Barnes, Yizhong Wang, and Karl Jenkins helped in data curation, writing—review and editing; Julien Thomas, Boyu Kuang, Stuart Barnes, and Yizhong Wang contributed to formal analysis; Karl Jenkins was involved in funding acquisition; Boyu Kuang, Stuart Barnes, and Karl Jenkins helped in project administration, supervision; software was done by Julien Thomas.

Funding The authors did not receive support from any organization for the submitted work.

Data Availability The data that support the findings of this study will be available at the Cranfield Online Research Data (CORD) platform after acceptance (https://cord.cranfield.ac.uk/).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Financial interests The authors declare that they have no financial interests.

Ethical and informed consent for data used All data have been used with ethical and informed consent.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material

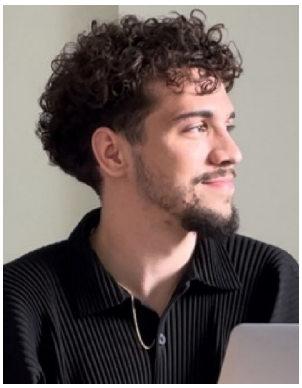
in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

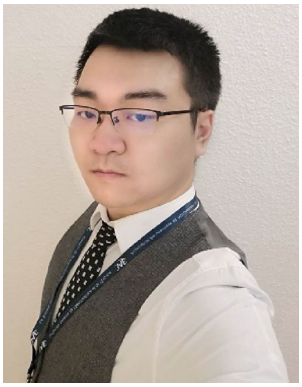
1. Rajapaksha, A., Jayasuriya, N.: Smart airport: a review on future of the airport operation. Global J. Manag. Bus. Res. **20**(3), 25–34 (2020)
2. UKRI, £65 million future flight challenge phase 3 competition launches-ukri. https://www.ukri.org/news/65-million-future-flight-challenge-phase-3-competition-launches/ (2021)
3. UKRI, Out of cycle next generation highly efficient air transport (oneheart). https://gtr.ukri.org/projects?ref=10003388 (2023)
4. Zhao, J., Conrad, C., Delezenne, Q., Xu, Y., Tsourdos, A.: A digital twin mixed-reality system for testing future advanced air mobility concepts: a prototype. In: Integrated Communication. Navigation and Surveillance Conference (ICNS) 2023, pp. 1–10 (2023). https://doi.org/10.1109/ICNS58246.2023.10124310
5. Zhao, J., Li, Y.-G., Sampath, S.: Convolutional neural network denoising auto-encoders for intelligent aircraft engine gas path health signal noise filtering. J. Eng. Gas Turbin. Power **145**(6) (2023). https://doi.org/10.1115/1.4056128
6. Misra, J., Saha, I.: Artificial neural networks in hardware: a survey of two decades of progress. Neurocomputing **74**(1–3), 239–255 (2010). https://doi.org/10.1016/j.neucom.2010.03.021
7. Menghani, G.: Efficient deep learning: a survey on making deep learning models smaller, faster, and better. ACM Comput. Surv. **55**(12), 1–37 (2023). https://doi.org/10.1145/3578938
8. Chen, L.-C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587 (2017)
9. Kuang, B., Barnes, S., Tang, G., Jenkins, K.: A dataset for autonomous aircraft refueling on the ground (agr). In: 2023 28th International Conference on Automation and Computing (ICAC), pp. 55–60. IEEE. (2023) https://doi.org/10.1109/icac57885.2023.10275212
10. Kim, G., Choi, J.G., Ku, M., Lim, S.: Developing a semi-supervised learning and ordinal classification framework for quality level prediction in manufacturing. Comput. Ind. Eng. **181**, 109286 (2023). https://doi.org/10.1016/j.cie.2023.109286
11. Park, P., Di Marco, P., Nah, J., Fischione, C.: Wireless avionics intracommunications: a survey of benefits, challenges, and solutions. IEEE Internet Things J. **8**(10), 7745–7767 (2021). https://doi.org/10.1109/JIOT.2020.3038848
12. Khalid, S., Song, J., Azad, M.M., Elahi, M.U., Lee, J., Jo, S.-H., Kim, H.S.: A comprehensive review of emerging trends in aircraft structural prognostics and health management. Mathematics **11**(18), 3837 (2023). https://doi.org/10.3390/math11183837
13. Faisal, N., Cora, O.N., Bekci, M.L., Åšliwa, R.E., Sternberg, Y., Pant, S., Degenhardt, R., Prathuru, A.: Defect Types, pp. 15–72 Springer (2021). https://doi.org/10.1007/978-3-030-72192-3_3
14. Hefazi, H.: Aerospace Engineering, pp. 1085–1137. Springer, Berlin (2021). https://doi.org/10.1007/978-3-030-47035-7_24
15. Yu, H., Yang, Z., Tan, L., Wang, Y., Sun, W., Sun, M., Tang, Y.: Methods and datasets on semantic segmentation: a review. Neurocomputing **304**, 82–103 (2018)
16. Sehar, U., Naseem, M.L.: How deep learning is empowering semantic segmentation: traditional and deep learning techniques for semantic segmentation: A comparison. Multimedia Tools Appl. **81**(21), 30519–30544 (2022)
17. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
18. Ronneberger, O., Fischer, P., Brox, T.: 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, (2015)
19. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
20. Alexey, D.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv: 2010.11929 (2020)
21. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: simple and efficient design for semantic segmentation with transformers. Adv. Neural. Inf. Process. Syst. **34**, 12077–12090 (2021)
22. Hossain, M.S., Betts, J.M., Paplinski, A.P.: Dual focal loss to address class imbalance in semantic segmentation. Neurocomputing **462**, 69–87 (2021)
23. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA (2017), and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3, pp. 240–248. Springer (2017)
24. Bertasius, G., Shi, J., Torresani, L.: Semantic segmentation with boundary neural fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3602–3610 (2016)
25. Hu, H., Cui, J., Wang, L.: Region-aware contrastive learning for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 16291–16301 (2021)
26. Chibane, J., Engelmann, F., Anh Tran, T., Pons-Moll, G.: Box2mask: Weakly supervised 3d semantic instance segmentation using bounding boxes. In: European Conference on Computer Vision, pp. 681–699. Springer (2022)
27. Huang, S.-C., Le, T.-H.: Transfer Learning, pp. 219–233. Elsevier, New York (2021). https://doi.org/10.1016/b978-0-323-90198-7.00008-2
28. Vrbančič, G., Podgorelec, V.: Transfer learning with adaptive fine-tuning. IEEE Access **8**, 196197–196211 (2020)
29. Kora, P., Ooi, C.P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W.Y., Meenakshi, K., Swaraja, K., Plawiak, P., Acharya, U.R.: Transfer learning techniques for medical image analysis: a review. Biocybern. Biomed. Eng. **42**(1), 79–107 (2022)
30. Liu, X., Li, J., Ma, J., Sun, H., Xu, Z., Zhang, T., Yu, H.: Deep transfer learning for intelligent vehicle perception: a survey. Green Energy Intell Transp 100125 (2023)
31. Wang, R., Yao, X., Yang, J., Xue, L., Hu, M.: Hierarchical deep transfer learning for fine-grained categorization on micro datasets. J. Vis. Commun. Image Represent. **62**, 129–139 (2019)
32. Yuan, X., Shi, J., Gu, L.: A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Syst. Appl. **169**, 114417 (2021)
33. Arshad, M., Qureshi, M., Inam, O., Omer, H.: Transfer learning in deep neural network based under-sampled mr image reconstruction. Magn. Reson. Imaging **76**, 96–107 (2021)
34. Wang, S., Zhang, L., Fu, J.: Adversarial transfer learning for cross-domain visual recognition. Knowl.-Based Syst. **204**, 106258 (2020)
35. Thung, K.-H., Wee, C.-Y.: A brief review on multi-task learning. Multimedia Tools Appl. **77**(22), 29705–29725 (2018)

36. Kuang, B., Nnabuife, S.G., Whidborne, J.F., Sun, S., Zhao, J., Jenkins, K.: Self-supervised learning-based two-phase flow regime identification using ultrasonic sensors in an s-shape riser. Expert Syst. Appl. **236**, 121414 (2024). https://doi.org/10.1016/j.eswa.2023.121414.
37. Boschini, M., Bonicelli, L., Porrello, A., Bellitto, G., Pennisi, M., Palazzo, S., Spampinato, C., Calderara, S.: Transfer without forgetting. In: European Conference on Computer Vision, pp. 692–709. Springer (2022)
38. Liu, M., Yan, X., Wang, C., Wang, K.: Segmentation mask-guided person image generation. Appl. Intell. **51**(2), 1161–1176 (2020). https://doi.org/10.1007/s10489-020-01907-w
39. Maji, S., Rahtu, E., Kannala, J., Blaschko, M., Vedaldi, A.: Fine-grained visual classification of aircraft. arXiv:1306.5151 (2013)
40. Ma, Z., Wu, X., Chu, A., Huang, L., Wei, Z.: Swinfg: A fine-grained recognition scheme based on swin transformer. Expert Syst. Appl. 123021 (2023)
41. Du, R., Xie, J., Ma, Z., Chang, D., Song, Y.-Z., Guo, J.: Progressive learning of category-consistent multi-granularity features for fine-grained visual classification. IEEE Trans. Pattern Anal. Mach. Intell. **44**(12), 9521–9535 (2021)
42. Nicolosi, F., Corcione, S., Trifari, V., De Marco, A.: Design and optimization of a large turboprop aircraft. Aerospace **8**(5), 132 (2021)
43. Cheng, G., Yuan, X., Yao, X., Yan, K., Zeng, Q., Xie, X., Han, J.: Towards large-scale small object detection: survey and benchmarks. IEEE Trans. Pattern Anal. Mach. Intell. (2023)
44. Shao, Z., Yin, Y., Lyu, H., Soares, C.G., Cheng, T., Jing, Q., Yang, Z.: An efficient model for small object detection in the maritime environment. Appl. Ocean Res. **152**, 104194 (2024)
45. Kim, G., Yang, S.M., Kim, D.M., Choi, J.G., Lim, S., Park, H.W.: Developing a deep learning-based uncertainty-aware tool wear prediction method using smartphone sensors for the turning process of ti-6al-4v. J. Manuf. Syst. **76**, 133–157 (2024)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Julien Thomas is a Junior Flight Simulation Software Engineer at Airbus Defence and Space, where he has been contributing to innovative simulation solutions since March 2024. He holds an M.Sc. in Computational and Software Techniques in Engineering from Cranfield University, where his thesis focused on aircraft components' semantic segmentation for future civil aviation autonomy. Prior to Airbus, Julien completed internships at Cranfield University and Safran Data Systems, where he worked on deep learning for civil aviation and static code analysis, respectively. He is proficient in C++, Python, and PyTorch, with experience in hybridizing loss functions and code standards.



Boyu Kuang is a research fellow in Computer Vision and AI at the Centre for Computational Engineering Sciences, Cranfield University. He holds a Ph.D. in Computer Vision, with his research focusing on deep learning, machine vision, and multiple domain adaptation. Dr. Kuang has published extensively in top journals and has worked on prestigious projects, including the Airbus-funded Autonomous Aircraft Ground Refueling system. He is also an active IEEE member and professional reviewer for leading journals such as IEEE Transactions on Industrial Electronics and Elsevier Expert Systems with Applications. His expertise spans aerospace, robotics, and digital twin applications.

Yizhong Wang is a research student at Cranfield University's Centre for Life-cycle Engineering and Management. He earned dual bachelor's degrees in industrial engineering from Sichuan University and the University of Pittsburgh in 2020. Afterward, he pursued aerospace manufacturing at Cranfield University, graduating in 2021. His current research focuses on non-destructive testing and artificial intelligence. Yizhong has co-authored several publications, including work on laser scanning thermography for composite inspection and 3D rail track reconstruction. He is supervised by Professor Yifan Zhao and Dr. Isidro Durazo Cardenas.

Stuart Barnes is a lecturer in Computational Intelligence and Data Analytics at Cranfield University and serves as the Course Director for the MSc in Computational and Software Techniques in Engineering. He holds a PhD and MSc in Computer Vision from Cranfield, followed by his BSc and MSc in Physics from the University of Kent. Dr. Barnes has worked across academia and industry, contributing to commercial scientific software development. His research interests focus on Computer Vision, Artificial Intelligence, and Software Engineering, with projects in areas such as gesture recognition, security, surveillance, and autonomous vehicle operations. He has collaborated with major clients such as Jaguar Land Rover, Airbus, and Saab UK.

Karl Jenkins is the Head of the Centre for Computational Engineering Sciences and Professor of Computational Engineering at Cranfield University. He holds a PhD from the University of Manchester, where he specialized in computational and experimental water wave interactions with coastal structures. His expertise includes Computational Fluid Dynamics (CFD), High-Performance Computing (HPC), and turbulent combustion, which he expanded during his postdoctoral research at Cambridge University. Professor Jenkins has over 100 publications, and his research spans reacting flows, turbulence, and large-scale data processing. He has collaborated with industry leaders such as Airbus, Rolls-Royce, and Siemens.

https://dspace.org/

Advanced semantic segmentation of aircraft main components based on transfer learning and data-driven approach

Thomas, Julien; Kuang, Boyu; Wang, Yizhong; Barnes, Stuart; Jenkins, Karl

2025-05
Springer
https://doi.org/10.1007/s00371-024-03686-8
Thomas J, Kuang B, Wang Y, et al., (2025) Advanced semantic segmentation of aircraft main components based on transfer learning and data-driven approach. The Visual Computer, Volume 41, Issue 7, May 2025, pp. 4703-4722

Downloaded from a DSpace repository. https://dspace.org/