

Improved DETR-Based Detection of Icing and Snow on Aircraft Surfaces

Shi Yongsheng and Xiang Yuhang

College of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China

Keywords: DETR, Object Detection, RefineBox, De-Icing detection, Transformer.

Abstract: Utilizing neural network models to detect icing and snow accumulation on aircraft surfaces can significantly reduce the workload of maintenance personnel, enhance operational efficiency, and lower aircraft operating costs. This proposal marks the first application of the transformer-based object detection model DETR to the detection of icing and snow on aircraft surfaces. To address the issue of significant boundary box prediction deviations in DETR, the RefineBox localization optimization network was employed for improvements. Performance was compared and analyzed on a custom dataset, revealing a 1.8% increase in the mAP metric for the enhanced model. Ground trials were conducted to validate the accuracy and feasibility of the improved model in detecting aircraft surface icing and snow. The results demonstrate that the enhanced model performs well, exhibits strong environmental adaptability, and can operate stably on mainstream devices.

1 INTRODUCTION

Snow, frost, and ice accumulation on the surfaces of aircraft can compromise the clean aerodynamic state critical for flight, posing a threat to flight safety. Not only in snowy weather, but whenever meteorological conditions reach icing thresholds, aircraft must undergo rigorous de-icing procedures before takeoff. Therefore, checking for surface icing during the pre-flight walk-around is essential.

Currently, aircraft surface icing checks are primarily conducted through manual inspections and contact-type icing sensors. Manual inspections depend on the visual acuity and judgment of maintenance personnel, making this method subjective. It's impractical to comprehensively inspect an entire aircraft's exterior solely by human height, requiring tools and presenting significant limitations; moreover, inspecting an aircraft, especially large ones, is time-consuming, involves multiple personnel, is costly, inefficient, and prone to issues during handover. Contact-type icing sensors can only detect icing at a single point or small area. For detecting widespread surface icing, these methods often require the installation of numerous sensors (Zhou et al, 2021).

In contrast, computer vision-based detection methods are unrestricted, relying on camera-captured

images to perform inspections without directly contacting the aircraft surface and capable of wide-area detection. This approach offers good protection for the aircraft, allowing for quantitative representation of detection results. Before 2014, traditional algorithms dominated computer vision-based object detection. Subsequently, deep learning-based object detection algorithms rapidly evolved. Convolutional Neural Networks (CNNs) utilize fixed weights in convolutional layers to extract features from specific parts of an image, then apply these invariant weights across the entire image through convolution. This approach has two main benefits: First, the invariance of weights ensures that the same features are extracted from any sub-image within the same image; second, it significantly reduces the amount of input data for the image. These advantages are crucial for the training speed and robustness of the neural network and have been proven to far surpass traditional feature-based image recognition methods in accuracy (Wei, 2022).

In 2020, Glenn Jocher (Carion et al, 2020) released the YOLOv5 model, and the Facebook AI Research team proposed the DETR neural network model. Li Gang (Li et al, 2023) from North China Electric Power University and others integrated DETR with prior knowledge to address the issue of sample imbalance in bolt defect detection. In 2023, Zhou Jing and Li Xin (Zhou and Li, 2023) used the e-

efficientNet network to extract image features, which were then merged via the BiFPN network and analyzed using DETR, enhancing the detection efficiency in tasks involving the inspection of anti-vibration hammers on power transmission lines. Representative models such as DETR and YOLOv5 have proven their effectiveness in a broad range of object detection tasks.

DETR's self-attention mechanism endows it with strong global contextual awareness, beneficial for fully considering the relationships between ice/snow targets and the aircraft, and operates without the need for preset anchor boxes, thus flexibly adapting to ice formations of varying sizes and shapes. Given these advantages, applying the DETR model to the task of detecting ice on aircraft surfaces holds great potential. However, during its application in detecting ice and snow accumulation on aircraft surfaces, issues such as positioning deviations, especially for small targets like icicles and clear ice, were noted, indicating that the model's localization performance needs enhancement. In 2023, Chen Y (Chen, 2023), from the Institute of Artificial Intelligence, Chinese Academy of Sciences and the University of Chinese Academy of Sciences proposed a localization optimization network tailored to the DETR model and its derivatives. This network extracts multi-scale features from DETR's Resnet backbone using a Feature Pyramid Network (FPN) and uses these features alongside Ground Truth to correct predicted bounding boxes, thereby improving the localization accuracy of the DETR model. This development is significant for addressing the aforementioned application issues. Employing advanced deep learning techniques for detecting ice on aircraft surfaces to enhance flight safety and efficiency provides maintenance personnel with a precise, efficient, and automated icing detection method, offering reliable decision support and further elevating flight safety standards.

2 IMPROVED DETR MODEL BASED ON REFINEBOX

2.1 DETR Model

DETR (Detection Transformer) is an object detection network based on the Transformer architecture. Unlike traditional object detection methods, DETR adopts an end-to-end approach, outputting the classes and positions of objects directly through the Transformer network, thus accomplishing the task of object detection.

DETR provides a novel approach to end-to-end object detection algorithms by combining CNNs and the Transformer model to predict the class information of N objects, including both targets and background, in parallel. Leveraging the Transformer's focus on global features, the DETR model possesses powerful global feature learning capabilities (Zhang et al, 2022). Specifically, DETR first encodes the input image into feature vectors via a CNN, which are then combined with positional encodings. The computation of positional encodings is as follows (Chen et al, 2023):

$$\begin{aligned} PE_{(pos,2i)} &= \sin(pos / 10000^{2i/d}) \\ PE_{(pos,2i+1)} &= \cos(pos / 10000^{2i/d}) \end{aligned} \quad (1)$$

In the formula, "pos" represents the position of the image block; "d" represents the dimension of the vector; and "2i" and "2i+1" represent the even and odd dimensions within "d", respectively. After the positional encoding, the feature vector is processed by the encoder which modifies the feature map. Through a linear layer and a multi-head self-attention mechanism, DETR generates a set of encoded vectors of specific sizes, representing the objects present in the image. These encoded vectors are matched with known category vectors, thus determining the probability distribution of classes for each object.

The Encoder in DETR receives feature vectors and processes them through a series of self-attention layers and feedforward neural networks, encoding them to extract high-level feature representations. These representations are then passed to the Decoder, serving as inputs for subsequent processes. The Decoder receives these feature representations from the Encoder and generates the object detection results. Typically, the Decoder is composed of a series of self-attention layers and feedforward neural networks, which allow it to merge and process features at different levels. In each Decoder layer, the model generates new predictions based on the current feature representations and previous prediction outcomes. These predictions include information about the object's class and location.

The interaction between the Encoder and Decoder is facilitated by a cross-layer multi-head self-attention mechanism. This mechanism allows the model to exchange information across different levels, thereby better capturing the global context and the relationships between objects (Fan and Ma, 2023; Vaswani, 2017; Chen et al, 2018). The design of the Encoder and Decoder in DETR aims to utilize the Transformer's self-attention mechanism and feedforward neural networks to accomplish end-to-end object detection tasks. This architecture not only

improves the efficiency of object detection by reducing the reliance on traditional detection steps such as region proposal generation but also enhances the model's ability to understand and interpret complex scenes where multiple objects interact or overlap.

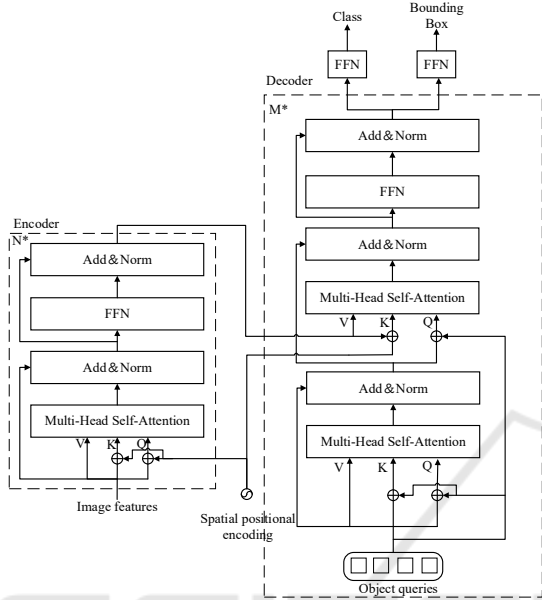


Figure 1: The structure of the transformer in DETR.

In order to improve the performance of the DETR model during training, it is necessary to use a loss function to measure the difference between the predictions and the actual targets. DETR employs a method called the Hungarian algorithm to match predicted bounding boxes with their corresponding ground truth values, then calculates the cross-entropy loss for the classes and the Smooth L1 loss for the bounding boxes to achieve the minimal loss. The formula for the cross-entropy loss is (Huang et al, 2019):

$$L_{class} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C \hat{y}_{ij} \log(y_{ij}) \quad (2)$$

N is the number of actual targets in the image, C represents the number of categories (including the background), y_{ij} is the predicted value for the j -th category in the i -th prediction box, and \hat{y}_{ij} is the actual value for the j -th category in the i -th prediction box.

Smooth L1 loss function:

$$L_{box} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^4 \hat{m}_i \text{SmoothL1}(b_{ij} - \hat{b}_{ij}) \quad (3)$$

b_{ij} represents the coordinates of the i -th predicted box, \hat{b}_{ij} represents the coordinates of the i -th actual box, and m_i represents the existence indicator of the i -th actual box (1 for present, 0 for absent).

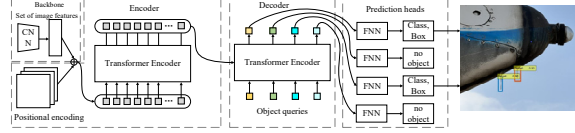


Figure 2: The structure of DETR.

2.2 RefineBox Location Optimization Network

RefineBox Localization Optimization Network is a method used for localization enhancement in object detection. It effectively improves the accuracy of object localization in the DETR model by optimizing bounding boxes without affecting the classification results. By utilizing multi-scale features and a series of Refiner modules, this network efficiently refines the bounding boxes predicted by the detector, further enhancing model performance. This approach adopts a two-stage detection philosophy. However, unlike typical two-stage methods, the RefineBox Localization Optimization Network is built upon a well-trained detection model with frozen parameters. It can be directly applied to pre-trained models without the need for retraining, significantly reducing time and computational costs.

First, the Feature Pyramid Network (FPN) is used to extract multi-scale features from the DETR's Backbone, reducing the channel count to a specific number C , considered as the model dimension. These features, along with the bounding boxes predicted by the Detector, serve as inputs. This enables the RefineBox to perform effectively in detecting both large and small objects. Subsequently, through a series of Refiner modules, the extracted multi-scale features are fully utilized to improve the bounding boxes predicted by the Detector, as shown in Figure 3. The weights of the Refiner modules are shared.

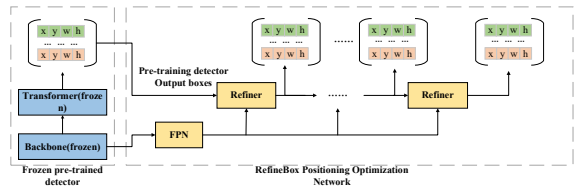


Figure 3: Structure of the RefineBox Localization Optimization Network.

The Refiner component consists of an ROI Align layer, a residual block, and a Multi-Layer Perceptron (MLP). Its internal structure is shown in Figure 4. The FPN extracts multi-scale features from the Backbone and inputs them into the ROI Align for feature alignment across different scales. Subsequently, the bounding boxes are optimized through the residual block and the MLP.

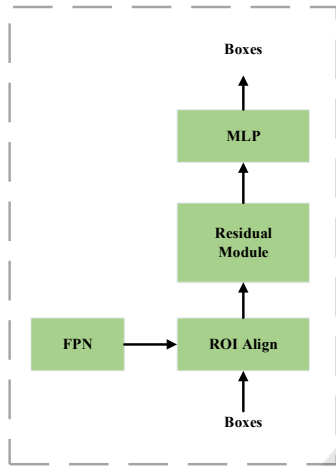


Figure 4: Structure of the Refiner.

3 EXPERIMENTS AND ANALYSIS

3.1 Dataset and Experimental Setup

In order to compare the performance of the models before and after the improvement, a custom dataset of icing or snow on the surface of an airplane was used to train and test the models before and after the improvement. A high-quality dataset is required to contain a large variety and sufficient number of images covering a rich range of application scenarios, such as variations in different angles, lighting conditions, and so on. 493 image data of icing snow on the surface of the airplane were obtained by collecting online and taking field photos on the ramp, and data augmentation methods such as flipping and adding noise were used to balance the categories and reduce overfitting. The total amount of data after augmentation was 800. LabelMe was chosen as the annotation software, the decision to use LabelMe was driven by the need for a reliable tool that can handle complex annotations with ease, facilitating the preparation of high-quality training data., and a COCO-format dataset was created for training the model and testing the model's performance metrics.

Neural network training environment and configuration:

Table 1: Environment and configuration.

Learning Rate	1e-4
Encoder+Decoder	6+6
Epochs	250
Device	NVIDIA Quadro P6000
Operating System	Windows10
Framework	Pytorch2.0.0+cuda11.8
Programming Language	Python

3.2 Experimental Results and Analysis

In the field of object detection, evaluating the performance of models is crucial, directly impacting the effectiveness and feasibility of models in practical applications. This experiment aims to conduct a detailed data analysis and comparison to comprehensively assess the performance of several mainstream object detection models and the RefineBox DETR model on a dataset of aircraft surfaces covered with ice and snow. A horizontal comparison not only showcases the unique advantages and potential limitations of each model but also highlights the effectiveness of the improvement methods and the advantages of the improved model.

The design of this experiment strictly adheres to scientific evaluation standards and principles of fair comparison. Tests are conducted under identical experimental conditions, including the same dataset, hardware, and software configurations, to ensure the objectivity of the results. This approach not only helps to verify the improvements made to the RefineBox DETR model but also provides valuable data support and theoretical guidance for future technological innovations in this field.

Table 2: Results.

Method	F1-Score	Recall	mAP(%)	Precision
Faster-RCNN	0.872	0.861	82.5	0.815
YOLOv5	0.859	0.778	79.6	0.863
DETR	0.863	0.932	81.9	0.802
ours	0.880	0.953	83.7	0.837

From the experimental data in Tables 2, it is evident that under the same environmental configurations and using the same training dataset, the two-stage object detection algorithm Faster-RCNN shows moderate recall and precision, with higher F1-Score and mAP, indicating its good performance and balanced recall and precision. However, due to the more complex structure of the two-stage detection algorithm compared to single-stage algorithms, it requires more computation time. The YOLOv5 object detection model has the lowest recall and highest precision, indicating that this model is less likely to produce false positives in the context of detecting icing and snow accumulation on aircraft surfaces, but it has a higher rate of false negatives, suggesting that it is prone to under-detecting the actual number of targets, leading to missed detections. The performance of the DETR model is more balanced compared to YOLOv5. Contrary to YOLOv5, it has higher recall but lower precision, indicating that it is less likely to miss detections, but more prone to false positives, and its higher mAP score also demonstrates its superior overall performance.

The improved RefineBox DETR model achieves the best results in all three metrics, proving the stability and effectiveness of its enhancements. The high recall ensures the reliability of the detection results, and the highest mAP score demonstrates its superior overall performance.

3.3 Application Testing


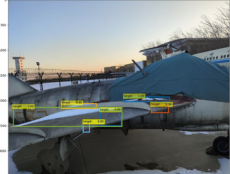


To validate the practicality of the model, it was chosen to conduct field tests on the apron after snowfall. The test environment included two lighting conditions: morning and evening. Factors such as the shooting angle, location of ice and snow on the aircraft, and the conditions of ice and snow accumulation were considered. The test subjects were two types of aircraft: a silver-grey skinned J-5 and a milky-white skinned An-24. The purpose of introducing these variables was to fully verify the impact of environmental factors on the accuracy of detection results.

The results presented in Table 3 show the detection performance of RefineBox DETR in real-world applications. The first group of photos and the second group were taken under conditions of shadow and direct sunlight, respectively; from angles of top-down and bottom-up views; with different skin colours; ice and snow located on the upper surface of the wing and the leading edge of the wing; under conditions of snow + transparent ice and snow +

icicles. These varying conditions were used to test the performance of the RefineBox DETR Detector in different real-world scenarios.

The test results indicate that RefineBox DETR achieves good detection results in various real-world scenarios under different conditions, reaching a high level of accuracy. This demonstrates the model's robustness and effectiveness in practical applications, particularly in challenging environmental conditions encountered in aircraft operations on snow-covered aprons.

Table 3: Application test results.

Capture images	Results
	
	

The results in Table 3 demonstrate the detection effectiveness of the RefineBox DETR in practical applications. From the results, it is evident that the improved RefineBox DETR object detection model exhibits strong performance in the application scenario of detecting ice and snow on aircraft surfaces, achieving high accuracy and recall rates. This indicates that the model is capable of reliably identifying areas of icing and snow accumulation, essential for maintaining aircraft safety and operational efficiency.

4 CONCLUSIONS

This article aims to enhance the accuracy of the DETR model in the application scenario of detecting ice and snow on aircraft surfaces by integrating the localization optimization network RefineBox with DETR. This integration improves the model's prediction of bounding boxes, reduces loss values, and increases the accuracy of the results.

A custom dataset of images showing ice and snow on aircraft surfaces was created to train and test the

metrics of RefineBox DETR. After selecting suitable hyperparameters for training, the final results demonstrated that RefineBox DETR exhibits higher robustness and accuracy compared to the original DETR model, proving the effectiveness of the improvements.

To verify the real-world performance of RefineBox DETR, it was tested in an engineering application on the apron. The testing process involved different lighting conditions, shooting angles, aircraft skin colours, positions of ice and snow, and ice and snow conditions, to highlight RefineBox DETR's inclusivity to different environments during practical applications. Field test results indicate that RefineBox DETR has good environmental inclusivity, high detection accuracy, and minimal errors, accurately detecting snow, transparent ice, and icicles.

REFERENCES

- Zhou, S., Zhao, M., Hu, X., Yu, Q., Xu, C., 2021. Research on Quantitative Detection Method of Icing Based on PCA of Guided Wave Energy Features. *Measurement & Control Technology*.
- Wei, P., 2022. Research on Wing Ice Shape Recognition Technology Based on Improved U-Net. Civil Aviation Flight University of China.
- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S., 2020. End-to-end object detection with transformers. *In European conference on computer vision*. Cham: Springer International Publishing.
- Li, Gang., Zhang, Y., Wang W., et al., 2023. Transmission Line Bolt Defect Detection Method Using DETR and Prior Knowledge Integration. *Journal of Graphics*.
- Zhou, J., Li, X., 2023. Transmission Line Anti-vibration Hammer Detection Based on Improved DETR. *Computer Simulation*.
- Chen, Y., Chen, Q., Sun, et al., 2023. Enhancing Your Trained DETRs with Box Refinement. arXiv preprint arXiv.
- Zhang, N., Zhong, Y., Zhao, Tao., Dian S., 2022. Small-Size Defect Detection Algorithm for Product Surfaces Based on Smooth-DETR. *Computer Applications Research*.
- Chen, L., Lin, C., Zheng, Z., Mo, Z., Huang, X., Zhao, G., 2023. *Review of Transformer in Computer Vision Scenes*. Computer Science.
- Fan, R., Ma, X., 2023 Improved DETR Algorithm for Crowded Pedestrian Detection. *Computer Engineering and Applications*.
- Vaswani, A., Shazeer, N., Parmar N., et al., 2017. Attention is all you need. *In Proc of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc.
- Chen, L., Zhu, Y., Papandreou, G., et al., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. *In Proc of the 15th European Conference on Computer Vision*. Cham: Springer.
- Huang, Q., Song, K., Lu, J., 2019. Loss Balance Function Applied to Imbalanced Multiclass Problems. *Journal of Intelligent Systems*.