ICJK

Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2024.344084

RESEARCH ARTICLE

Quantitative Evaluation Method for Anomaly Levels of Complex Flight Maneuver Based on Multi-sensor Data

Liqiang Ren[1], **Haipeng Wang**[1,*], **Xinlong Pan**[1,*], **Shuyi Jia**[1] and **Bing Wan**[1]

[1] Naval Aviation University, Yantai 264001, China

Abstract

The methods that identify complex flight maneuvers from multi-sensor flight parameter data fusion and conduct automated quantitative evaluations of anomaly levels could play an important role in enhancing flight safety and pilot training. However, existing methods focus on anomaly detection at individual flight parameter data points, making it challenging to accurately quantify the overall abnormality of a flight maneuver. To address this issue, this paper proposes a novel method for the quantitative evaluation of anomaly levels in complex flight maneuvers using multi-sensor data. The proposed method comprises two stages: complex flight maneuver recognition and anomaly level quantification. In the complex flight maneuver recognition stage, a one-dimensional dual attention mechanism (1D-DAM) is introduced to capture discriminative fusion features in both the temporal and variable dimensions. Based on this mechanism, we develop a one-dimensional dual attention mechanism ResNet (1D-DAMResNet) model to achieve the recognition of complex flight maneuvers. Subsequently, in the anomaly level quantification stage, we employ a clustering technique to establish a standard maneuver benchmark library, which serves as a reference for the flight maneuver evaluations of different categories. According to the results of flight maneuver recognition, the corresponding category of standard maneuver from the library is automatically selected, and the dynamic time warping algorithm is then utilized to compute the anomaly quantification score of the test maneuver, thereby determining its anomaly level. Compared to contrastive methods, the proposed complex flight maneuver recognition model demonstrates significant advantages in both accuracy and stability, with an average precision, recall, and F1 scores of 99.75%. Additionally, the proposed anomaly level quantification method provides an automatic quantification of the overall anomaly level of maneuvers, and the results are highly interpretable. Overall, this paper introduces a novel approach for the quantitative evaluation of anomaly levels in maneuvers, which not only contributes to improving the accuracy of flight training evaluation but also significantly enhances the efficiency and quality of flight training.

Academic Editor:
Tiancheng Li

Submitted: 10 October 2024
Accepted: 12 March 2025
Published: 17 March 2025

Vol. 2, **No.** 1, 2025.
10.62762/CJIF.2024.344084

***Corresponding authors:**
✉ Haipeng Wang
whp5691@163.com
✉ Xinlong Pan
airadar@126.com

Citation

Ren, L., Wang, H., Pan, X., Jia, S., & Wan, B. (2025). Quantitative Evaluation Method for Anomaly Levels of Complex Flight Maneuver Based on Multi-sensor Data. *Chinese Journal of Information Fusion*, 2(1), 14–26.

© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Keywords: complex flight maneuver recognition, anomaly level quantification, dual attention mechanism, dynamic time warping, flight training evaluation, data fusion.

1 Introduction

In the daily training of military fighter pilots, complex flight maneuvers training is a crucial and indispensable subject. Its purpose is to enable pilots to rapidly adjust the aircraft's flight status within a short period, thereby enhancing their ability to respond to various challenges in complex environments [1]. Quantifying the anomaly levels of these flight maneuvers is vital for evaluating their quality. It plays a key role in ensuring flight safety, optimizing training programs, and improving pilots' operational skills and combat capabilities [2].Traditional evaluation methods for anomaly levels depend on subjective observations and manual records, which are inefficient and lack precision. With the increase of novice pilots, these limitations become more apparent in processing multi-sensor flight data, failing to meet modern flight training requirements. Effective information fusion techniques are crucial for integrating heterogeneous flight parameters from multiple sensors to achieve comprehensive maneuver assessment. Therefore, developing maneuver anomaly level quantification evaluation methods to achieve automated evaluation of flight training quality holds significant military value and practical significance.

The quantification of flight maneuver anomaly levels involves two primary aspects: the recognition of complex flight maneuver and the quantification of anomaly level based on the recognition results. Currently, both domestic and international researchers have conducted some research in these two areas. In terms of complex flight maneuver recognition, existing methods can be categorized into three types: pattern matching-based methods, expert system-based methods, and machine learning-based methods. Traditional pattern matching-based methods typically rely on predefined similarity metrics, such as Dynamic Time Warping [3] (DTW) and its improved algorithms [4–6]. Then, subsequences with similarity exceeding a specific threshold with template sequences are labeled as the same flight maneuver category. This method solves the classification problem of non-equally long time series data and does not rely on domain knowledge. However, it requires calculating the similarity with all known flight maneuvers, resulting in low algorithm efficiency. Moreover, this method is difficult to distinguish between complex flight maneuver s with high similarity and requires manual adjustment of the threshold. Expert system-based methods depend on the prior knowledge of domain experts to construct an artificial rule knowledge base and achieve flight maneuver recognition through pattern matching techniques [7, 8]. These methods demonstrate high efficiency and accuracy in recognizing basic flight maneuvers. However, the construction and maintenance of the knowledge base require high manual costs, and each model is usually only suitable for specific types of aircraft or flight tasks, limiting its generalizability in practical applications. To address these challenges, machine learning-based methods have emerged and shown significant advantages, such as decision trees, support vector machines [9], hidden Markov models, and naive Bayes models [10]. These shallow learning methods are suitable for handling linearly separable or simply nonlinear problems and have achieved satisfactory results in inferring the details of flight maneuvers. Furthermore, due to the high dimensionality, strong coupling, and non-linear characteristics of multi-dimensional flight maneuver sequence data, shallow learning methods struggle to capture the significant distinguishing features of complex flight maneuvers, thereby further limiting their effectiveness in recognizing complex flight maneuvers. Therefore, exploring more efficient and generalizable recognition methods is an important direction for current research.

In the aspect of quantifying flight maneuver anomaly levels, current research primarily focuses on the anomaly detection of flight parameter data, covering the identification of point anomalies, sequence anomalies, and pattern anomalies [11–14]. For the detection of anomalies in overall flight trajectories, the main methods include rule-based, model-based, and deep learning-based anomaly level evaluation methods. Rule-based methods quantify the anomaly levels of detected flight maneuvers through predefined rules and standards [15]. While this approach is straightforward and intuitive, its effectiveness is heavily reliant on the formulation of the rules, making it difficult to adapt to complex and variable flight environments. In contrast, model-based methods involve the establishment of mathematical models representing aircraft and pilot operations, utilizing simulation techniques to evaluate abnormal situations in flight maneuvers [16]. Although this approach can accurately simulate the flight process, the complexity involved in model construction and simulation

calculations limits its generalizability and applicability across different scenarios. Deep learning-based methods, on the other hand, leverage deep neural network models trained on large datasets to learn and extract anomaly patterns, achieving automated anomaly detection and evaluation [17–19]. Despite the promising application potential of this method, the acquisition of labeled data in supervised learning and the setting of anomaly thresholds in unsupervised learning pose certain limitations to its practical application.

To address the aforementioned challenges, this paper proposes a novel method for the automatic quantification of anomaly levels in complex flight maneuvers based on multi-sensor data. This method consists of two primary stages: complex flight maneuver recognition and anomaly level quantification. In the first stage, we tackle the issue of insufficient accuracy in recognizing complex flight maneuvers by employing a one-dimensional ResNet18 [20] (1D-ResNet) neural network enhanced with a dual attention module (DAM). In the second stage, the flight parameter sequence samples of flight maneuvers are normalized in time to ensure consistent sequence lengths. Then, by clustering similar flight maneuvers, a standard flight maneuver benchmark library is constructed for different categories. Next, the dynamic time warping algorithm is used to calculate the similarity between the test flight maneuver samples and the reference set. Finally, the similarity scores are normalized to abnormality quantification scores, which is categorized into anomaly levels based on predetermined score intervals. The proposed method for quantifying anomaly levels in complex flight maneuvers effectively supports automated evaluation of flight training quality. It enhances the efficiency and accuracy of flight training quality evaluation, aligning with the requirements of modern flight training programs.

Notably, fighter aircraft can perform numerous complex flight maneuvers. For the sake of simplicity, in this paper, we selected eight typical complex flight maneuvers as the research objects, which include the following: loop, cloverleaf loop, aileron roll, immelmann turn, split-S, pull-up, push-over and circulating.

2 Problem Formulation

Definition 1. (Flight Maneuver Recognition, FMR). Given a flight maneuver sequence $\mathbf{X} = \{x_1, \ldots, x_T\} \in \mathbb{R}^{T \times d}$, where T denotes the number of time steps and d denotes the number of parameters closely related to the flight state. The FMR can be regarded as given input maneuver sequence $\mathbf{X}$, we want to accurately predict the class label $y \in \{1, \ldots, C\}$ from C predefined classes.

We aim to train a deep neural network classifier, which is a mapping function $T \to L$, where T is the space of all maneuver samples and L is the set of all class labels, and the class label is a discrete variable. The deep neural network classifier is trained on a training dataset D_{Train} and subsequently evaluated on an independent test dataset D_{Test}.

Definition 2. (Flight Maneuver Abnormality Quantification). Defined as a metric function that measures the similarity $Sim(\mathbf{X}_{test}, \mathbf{X}_{std})$ between the test maneuver sequence $\mathbf{X}_{test}$ and its corresponding standard flight maneuver $\mathbf{X}_{std}$ from the predefined categories. The anomaly quantification score (AQS) can be expressed as:

$$AQS(\mathbf{X}_{test}) = f\left(Sim(\mathbf{X}_{test}, \mathbf{X}_{std})\right) \tag{1}$$

where f maps the similarity score to a value between 0 and 1, indicating the degree of abnormality. A score of 0 indicates completely normal performance, while a score of 1 signifies a high degree of abnormality.

Based on the AQS, we define the anomaly level as follows:

$$AL = \begin{cases} \text{Level1}, & \text{if } 0 \le AQS < \theta_1 \\ \text{Level2}, & \text{if } \theta_1 \le AQS < \theta_2 \\ \vdots & \\ \text{LevelK}, & \text{if } \theta_{k-1} \le AQS \le 1 \end{cases} \tag{2}$$

where $\theta_1, \theta_2, \ldots, \theta_{K-1}$ are predefined thresholds that divide the AQS into K abnormality levels, each corresponding to a different severity of abnormality.

3 Methodology

Accurately identifying complex flight maneuvers is crucial for quantifying flight maneuver anomaly levels. During flight, parameters such as position, velocity, acceleration, heading angle, pitch angle, and roll angle undergo dynamic changes [21]. Since different maneuvers exhibit distinct patterns in these parameter changes, effective analysis of these parameters is essential for precise recognition of complex maneuvers. Given the multi-sensor nature of flight data acquisition, our methodology focuses on discriminative feature fusion from heterogeneous parameters including

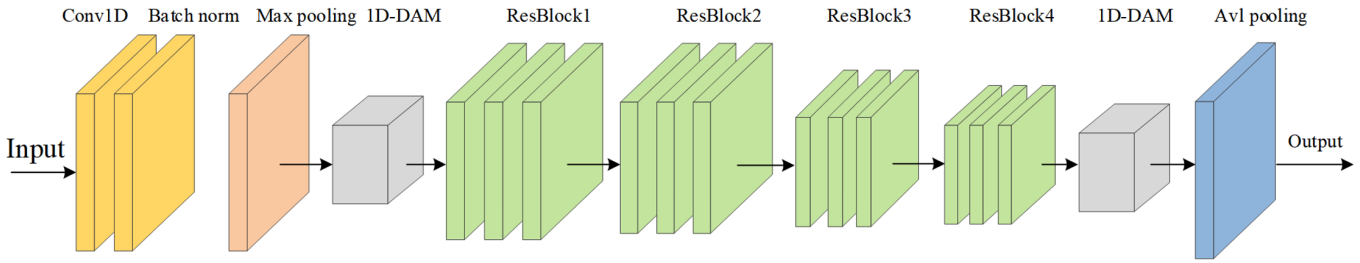


Figure 1. The overall architecture of the proposed 1D-DAMResNet.

position, velocity, acceleration, and attitude angles. The proposed 1D-DAMResNet model specifically addresses the challenge of fusing temporal and cross-sensor correlations to extract comprehensive maneuver signatures. The ResNet18 [22] network, with its residual block design, effectively mitigates the issue of vanishing gradients in deep neural networks through skip connections, making it well-suited for handling the nonlinear and complex characteristics of flight data.

To further improve the model's feature extraction capability, we optimized the ResNet18 architecture by proposing a novel one-dimensional dual attention ResNet18 (1D-DAMResNet). Specifically, a one-dimensional dual attention module is incorporated before the first residual block and after the last residual block. This design enhancement enables the model to better capture critical features from flight parameters. Figure 1 illustrates the architecture of the optimized network. The following sections provides a detailed explanation of the design of the one-dimensional dual attention module, along with the model's structural parameters and optimization details.

Figure 2, the 1D-DAM consists of two sub-modules: variable attention (VA) submodule and temporal attention (TA) submodule, which interact to capture long-range dependencies and variable interactions.

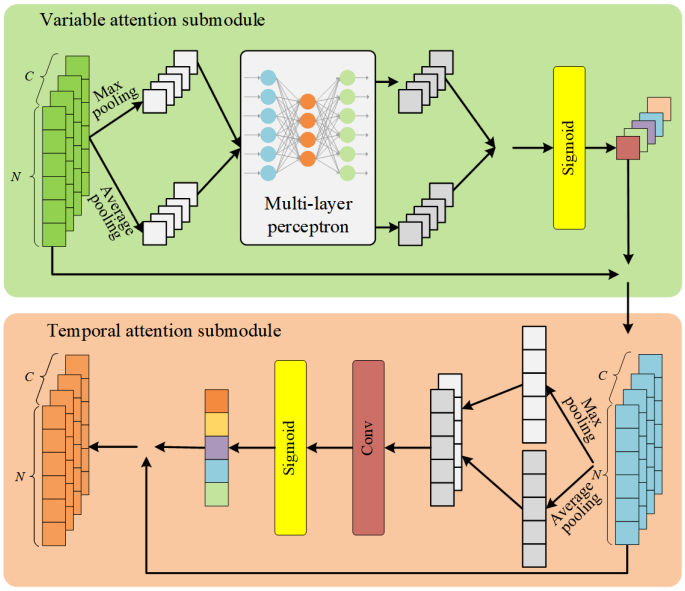


Figure 2. The architecture of the proposed 1D-DAM.

3.1 Proposed 1D-DAMResNet

3.1.1 One-Dimensional Dual Attention Module

In the task of flight maneuver recognition, the multi-dimensional flight parameter sequence exhibits dependencies over time and correlations among different features. Depending on the situation, features such as normal load factor, heading angle, or altitude become critical. To address this, we propose a strategy that incorporates an attention mechanism within the feature representation, it becomes feasible to focus on important features while suppressing irrelevant ones. Inspired by the Convolution Block Attention Module [23] (CBAM) in computer vision, we introduce a one-dimensional dual attention module (1D-DAM) that highlights meaningful features in both the channel and temporal dimensions. As shown in

The Variable Attention (VA) submodule implements sensor-level information fusion by automatically learning optimal weighting across different flight parameters. This enables adaptive fusion of multi-sensor data based on their relative importance for specific maneuver recognition tasks. Assuming the input to the VA submodule corresponds to a feature map $y \in \mathbb{R}^{C\times N}$, where C and N represent the variable and temporal dimensions, respectively. To efficiently compute variable attention, global max pooling and global average pooling are applied across the temporal dimension for each feature map $y^i \in \mathbb{R}^N, i = 1, 2, \ldots, C$, generating two feature vectors: max pooling feature $M^V = [M_1^V, M_2^V, \ldots, M_C^V]$ and average pooling feature $A^V = [A_1^V, A_2^V, \ldots, A_C^V]$. The i-th element of

these two vectors is calculated as follows:

$$\begin{aligned} M_i^V &= \max(y^i) \\ A_i^V &= \text{Average}(y^i) \end{aligned} \tag{3}$$

where $\max(\cdot)$ denotes the maximum element in the vector, and average$(\cdot)$ denotes the average of all elements in the vector.

These two vectors M^V and A^V are then fed into a multi-layer perceptron (MLP) with two fully connected layers: a C dimensional input layer, a C/r dimensional hidden layer, and a C dimensional output layer. The goal of this MLP is to *adaptively* adjust the weights for each channel by first reducing the feature map channels using a reduction factor r and then re-scaling them to the original size. The MLP produces two new feature vectors $\hat{M}^V$ and $\hat{A}^V$, which are merged by element-wise summation and then normalized using the sigmoid function to produce the variable attention weight vector $W_V \in \mathbb{R}^C$, as follows:

$$W_V = \sigma\left(F_2\left(\delta\left(F_1(M^V)\right)\right) + F_2\left(\delta\left(F_1(A^V)\right)\right)\right) \tag{4}$$

where σ and δ denote the GELU and sigmoid activation functions, respectively, F_1 and F_2 are convolutional mapping functions with kernel size 1.

Finally, the variable attention weight vector W_V is element-wise multiplied with the corresponding feature maps to generate a new feature representation $f \in \mathbb{R}^{C\times N}$. The calculation formula for the $i\text{-}th$ feature map is as follow:

$$f^i = W_V(i) \times y^i, \quad i = 1, 2, \ldots, C \tag{5}$$

The TA submodule focuses on the crucial positions in the flight maneuver sequences along the temporal dimension. The output of the VA submodule f serves as the input to this submodule. Similarly, we employ global max pooling and global average pooling operations to compute the maximum and average values across the variable dimension, resulting in two feature vectors: the max-pooled feature $M^T = [M_1^T, M_2^T, \ldots, M_N^T]$ and the average-pooled feature $A^T = [A_1^T, A_2^T, \ldots, A_N^T]$. The calculation formulas for the $i\text{-}th$ element of these two feature vectors are as follows:

$$\begin{aligned} M_j^T &= \max_{i=1,2,\ldots,C}(f(i,j)) \\ A_j^T &= \frac{1}{C}\sum_{i=1}^{C} f(i,j) \end{aligned} \tag{6}$$

Unlike the additive strategy used in the VA submodule, these two pooled vectors are concatenated to aggregate variable dimension information, forming a feature tensor $S^V \in \mathbb{R}^{2\times N}$. A standard one-dimensional convolution with a kernel size of k, followed by a sigmoid activation function, generates the temporal attention weight vector, as follow:

$$W_T = \sigma\left(F_3\left([M^T; A^T]\right)\right) \tag{7}$$

where σ denotes the sigmoid activation function, F_3 is a convolutional mapping function with kernel size k.

Finally, the output feature map $\hat{y} \in \mathbb{R}^{C\times N}$ is computed by element-wise multiplication of f each channel dimension feature map with the corresponding element of W_T. This feature map simultaneously captures more critical information in both the temporal and variable domains. The calculation formula for the $i\text{-}th$ feature map of $\hat{y}$ is as follows:

$$\hat{y}^i = W_T \odot f^i, \quad i = 1, 2, \ldots, C \tag{8}$$

3.1.2 *Improved Classification Layer*

The original classification layer in the ResNet18 network consists of a global average pooling layer followed by a fully connected layer. In this paper, we introduce an improved classification layer while retaining the network structure before the global average pooling layer. The modified classification layer includes a 1×1 convolutional layer, a global average pooling layer, another 1×1 convolutional layer, and a final linear transformation layer. The linear transformation layer is designed to map the extracted high-dimensional features to the flight maneuver classification categories. To enhance the generalization capability of the flight maneuver recognition model, a Dropout layer is inserted between the 1×1 convolutional layer and the linear transformation layer following the global average pooling layer. By integrating the attention mechanism and a modified classification layer within the ResNet18 framework, the proposed 1D-DAMResNet structure is shown in Table 1.

Remarks: n is the number of flight parameter features; k represents the kernel size; C denotes the number of flight action categories.

4 Quantification of Anomaly Levels

During flight training, pilots may not always perform maneuvers according to the ideal standards, leading to significant deviations from the standard values specified in the training outline. To address this issue, this section proposes a method for quantifying

Table 1. The proposed 1D-DAMResNet structure.

Layers	Parameters	Input Size	Output Size
Conv1D	k=7	512	64256
MaxPool	k=3	64256	64128
1D-DAM	/	64128	64128
ResBlock1	k=3	64128	64128
ResBlock2	k=3	64128	12864
ResBlock3	k=3	12864	25632
ResBlock4	k=3	25632	51216
1D-DAM	/	51216	51216
Conv1D	k=1	51216	51216
AvgPool	k=16	51216	512
Dropout	/	512	512
Conv1D	k=1	512	128
Dropout	/	128	128
Linear	/	128	C

anomaly levels in flight maneuvers, building upon the prior recognition of maneuver categories. The approach first applies time normalization to the flight parameter samples, ensuring consistency in data length across different categories and samples within the same category. Next, clustering is performed on maneuvers within the same category to establish the standard maneuvers sets for each class. Then, the DTW algorithm is used to calculate the similarity between the test flight maneuvers and the standard maneuver sets. and the similarity scores are normalized to obtain abnormality quantification scores. Finally, the similarity scores are normalized into abnormality quantification scores, which are categorized into different anomaly levels based on predefined score ranges. Notably, this method provides a visual representation of the deviations between the test flight maneuvers and the standard maneuvers at critical points and features. This visualization serves as a valuable tool for instructor pilots to evaluate trainees' performance, as well as for trainees to engage in self-directed learning and training.

4.1 Standard Maneuver Benchmark Library Construction

To quantify the anomaly levels for different categories of flight maneuvers, it is necessary to establish a standard maneuver benchmark library. This library serves as a reference standard and guides the threshold setting for subsequent scoring rules. The establishment process of the standard maneuver benchmark library is as follows.

Due to differences in pilot habits and proficiency levels, variations arise in the execution of the same flight maneuvers, particularly in terms of duration and roll direction. These differences are even more pronounced across different maneuver categories. To facilitate effective anomaly analysis, it is essential to standardize flight maneuvers. The first step in this process is to standardize the roll direction. For example, in roll maneuvers, if both left and right rolls are present, right roll angles should be converted to left roll angles, ensuring consistency in roll direction. Next, the duration of each maneuver must be normalized to a uniform length using linear interpolation or down-sampling. After duration normalization, amplitude normalization is performed to eliminate discrepancies in the scale of different feature parameters. This step is crucial for ensuring that the analysis is not affected by differences in scale. Finally, maneuvers of the same category are grouped using a clustering algorithm. Taking into account the computational complexity, time-series data characteristics, and sensitivity of the parameters, the K-Means clustering algorithm is employed. The number of clusters is set to three, and the flight maneuvers closest to the center of each cluster are selected as the standard maneuver set for that category. These standardized maneuver sets form the standard maneuver benchmark library, which serves as a reference for subsequent anomaly analysis and evaluation.

4.1.1 *Quantification Method for Anomaly Levels*

To quantify the abnormality of flight maneuvers, we first calculate the similarity between a test maneuver and all corresponding standard maneuvers using the DTW algorithm. The lowest similarity value among these comparisons is taken as the similarity score for the test sample. This process can be represented as follows:

$$\mathrm{DTW}^{i}_{\mathrm{distance}} = \frac{\mathrm{DTW}(X_{\mathrm{test}}, X^{i}_{\mathrm{std}})}{\mathrm{len}(\mathrm{path}_{i})} \tag{9}$$

$$Sim(X_{\mathrm{test}}, X_{\mathrm{std}}) = \min(\mathrm{DTW}^{i}_{\mathrm{distance}}) \tag{10}$$

where X_{test} is the test flight maneuver sample, X^{i}_{std} denotes the $i\text{-}th$ sample from the standard maneuver set, $\mathrm{DTW}(X_{\mathrm{test}}, X^{i}_{\mathrm{std}})$ is the similarity between X_{test} and X^{i}_{std}, and $\mathrm{len}(\mathrm{path}_{i})$ represents the shortest path distance in the DTW algorithm, while $\min(\mathrm{DTW}^{i}_{\mathrm{distance}})$ is the minimum similarity value between the test sample and the standard maneuvers, effectively serving as the similarity score for the test sample.

Next, this similarity score is normalized to quantify the anomaly score, as expressed by:

$$\text{Score} = \frac{1}{Sim(X_{\text{test}}, X_{\text{std}}) + 1} \tag{11}$$

$$AQS(X_{\text{test}}) = 1 - \text{Score} \tag{12}$$

where Score is the normalized similarity score, with a range of $(0, 1]$, and $AQS(X_{\text{test}})$ is abnormality quantification score for the test maneuver sample, with a range of $[0, 1)$. A higher value indicates a greater level of anomaly.

Subsequently, a mapping is established between the abnormality quantification score and anomaly levels. Specifically, according to the flight training guidelines, maneuver quality is categorized as "Poor" for scores in the range $[0, 2.5)$, "Good" for $[2.5, 4)$, and "Excellent" for $[4, 5]$. From each maneuver category, 100 samples are randomly selected along with their instructor-assigned scores. For samples with scores closest to 2.5 and 4, the abnormality quantification scores are calculated. By averaging the abnormality quantification scores of the samples nearest to 2.5 and 4, the threshold values θ_1 and θ_2 are determined. Based on these thresholds, the mapping between abnormality quantification scores and anomaly levels is defined as follows: a score in $[0, \theta_2)$ is classified as "Excellent," a score in $[\theta_2, \theta_1]$ is classified as "Good," and a score in $(0.79, 1]$ is classified as "Poor." Additionally, these mappings can also be defined manually by domain experts.

To facilitate anomaly analysis, we identify the key dimensional features that flight instructors focus on for each type of maneuver. By visualizing the curves of the test maneuver and the corresponding standard maneuver for these key features, we can highlight deviations at critical points. This visualization enables instructors to diagnose the causes of anomalies more effectively.

5 Experiments

5.1 Dataset Construction

We used Prepar3D to generate data for eight flight maneuvers, simulating each 250 times for a total of 2,000 samples. Multi-sensor data fusion was implemented by synchronously recording seven key flight parameters from integrated avionics systems: roll angle (AHRS), yaw angle (GPS/INS), pitch angle (AHRS), relative altitude (radar altimeter), vertical speed (air data computer), velocity (GPS), and normal load factor (accelerometers).

Table 2. The details of the complex flight maneuver dataset.

Flight maneuver	Training set size	Validation set size	Test set size	Maneuver type	Label
Loop	150	50	50	M0	0
Cloverleaf	150	50	50	M1	1
Aileron Roll	150	50	50	M2	2
Immelmann turn	150	50	50	M3	3
Split-S	150	50	50	M4	4
Pull-up	150	50	50	M5	5
Push-over	150	50	50	M6	6
Circulating	150	50	50	M7	7

Given that the original flight data was sampled at 30Hz, each maneuver generated a large number of data points. To reduce computational complexity, we down-sampled each sequence, dividing it into 512 intervals and randomly selecting one point from each interval. This ensured that all maneuver sequences had a uniform length of 512 data points. The final dataset consists of eight categories of complex maneuvers, with 250 samples per category. The dataset was split into training, validation, and test sets in a 6:2:2 ratio. The dataset details are presented in Table 2.

5.2 Result Analysis of Flight Maneuver Recognition

5.2.1 Experimental Setting

In this experiment, We used Python 3.9 and PyTorch 2.0.1 to train the model on an i9-13900HX CPU and RTX 4090 GPU. Training lasted 50 epochs with a batch size of 64, using Adam [24] optimizer and Label Smoothing Regularization [25] (LSR) loss function.

5.2.2 Model Training

To evaluate the effectiveness and accuracy of the proposed flight maneuver recognition model, we conducted a comparative study using 1D ResNet18 [22], 1D VGG11 [26], 1D MobileNet [27], and the multi-scale feature extraction deep residual network (MSDRN) proposed in [1]. To accommodate the time-series characteristics of flight parameter data, the ResNet18, VGG11, and MobileNet models were modified by replacing their original 2D convolutional layers with 1D convolutional layers, while keeping the rest of the network architecture unchanged. For a fair comparison, all models were trained on the same dataset from scratch, without using any pre-trained weights, to eliminate the potential influence of prior knowledge on the results. Each experiment was repeated five times to mitigate the effect of randomness, and the loss and accuracy were recorded

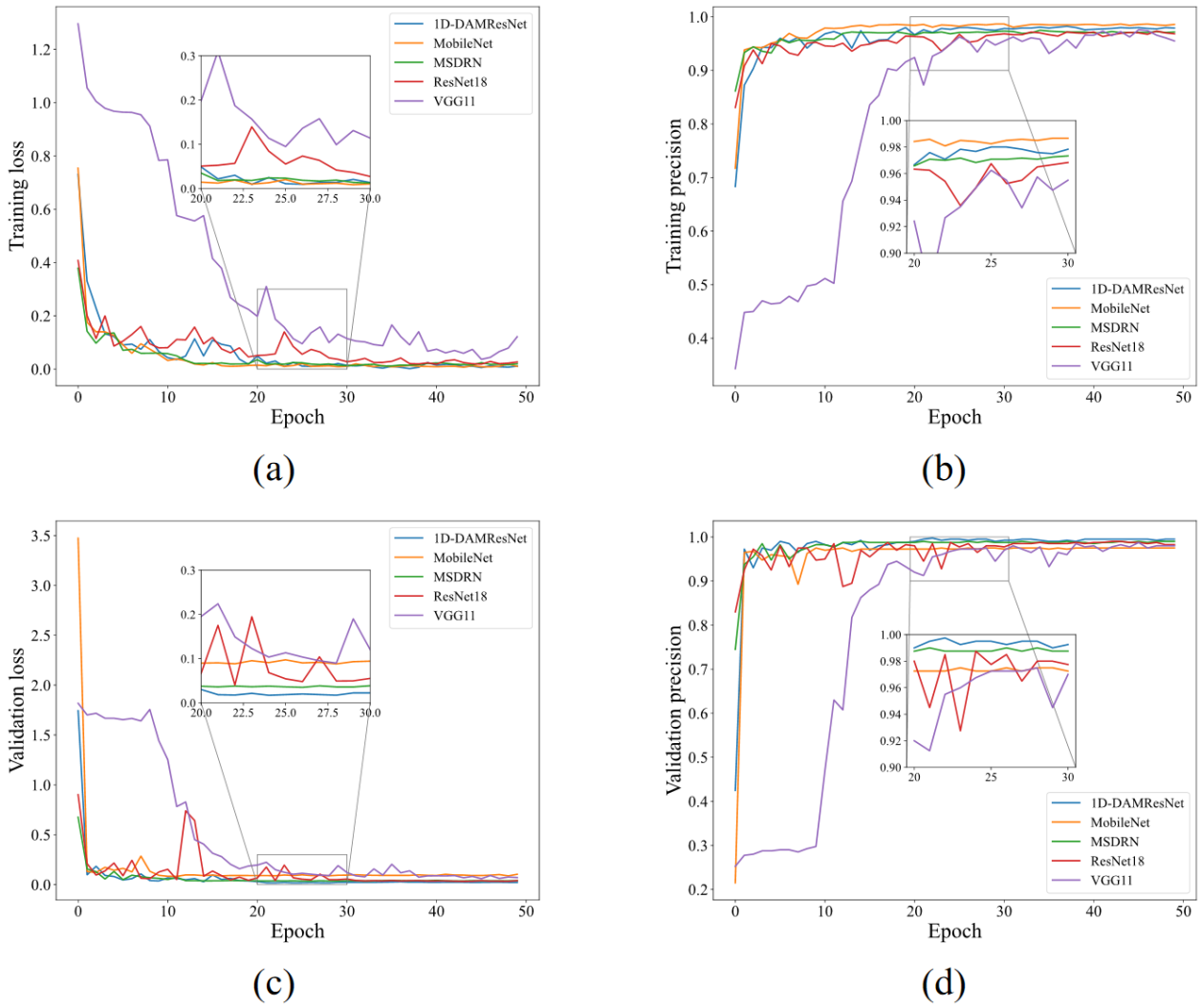


Figure 3. The average loss and precision of each method: (a) Training loss; (b) Training precision; (c) Validation loss; (d) Validation precision.

for both the training and validation sets. The average loss and accuracy from these five experiments are showed in Figure 3.

As shown in Figure 3, during the early training stages, all methods exhibited significant fluctuations in loss and precision due to the random initialization of model parameters. However, after approximately 20 epochs, all models converged, achieving low loss and high precision on both the training and validation sets, demonstrating good stability. Notably, VGG11 showed a slower convergence rate and lower precision on the validation set, both in terms of loss and precision. This is possibly due to the larger number of parameters in VGG11, which may have led to underfitting on small samples. On the other hand, MobileNet showed lower loss and higher precision on the training set but underperformed on the validation set, suggesting overfitting and reduced generalization ability. In contrast, the proposed 1D-DAMResNet model consistently achieved lower average loss and higher average precision on the validation set, with both metrics exhibiting high stability. This excellent performance can be attributed to the model's ability to extract robust feature representations. Overall, the experimental results validated the accuracy and stability of the proposed method.

5.2.3 *Model Test*

Table 3. Recognition performance of each method.

Method	Average Precision(%)	Average Recall(%)	Average F1 Score(%)
1D-DAMResNet	**99.75**	**99.75**	**99.75**
MSDRN [1]	99.50	99.50	99.50
ResNet18 [22]	99.25	99.25	99.25
MobileNet [27]	97.50	97.50	97.50
VGG11 [26]	98.75	98.75	98.75

Table 3 shows the test set recognition performance of various methods, with the 1D-DAMResNet18 model

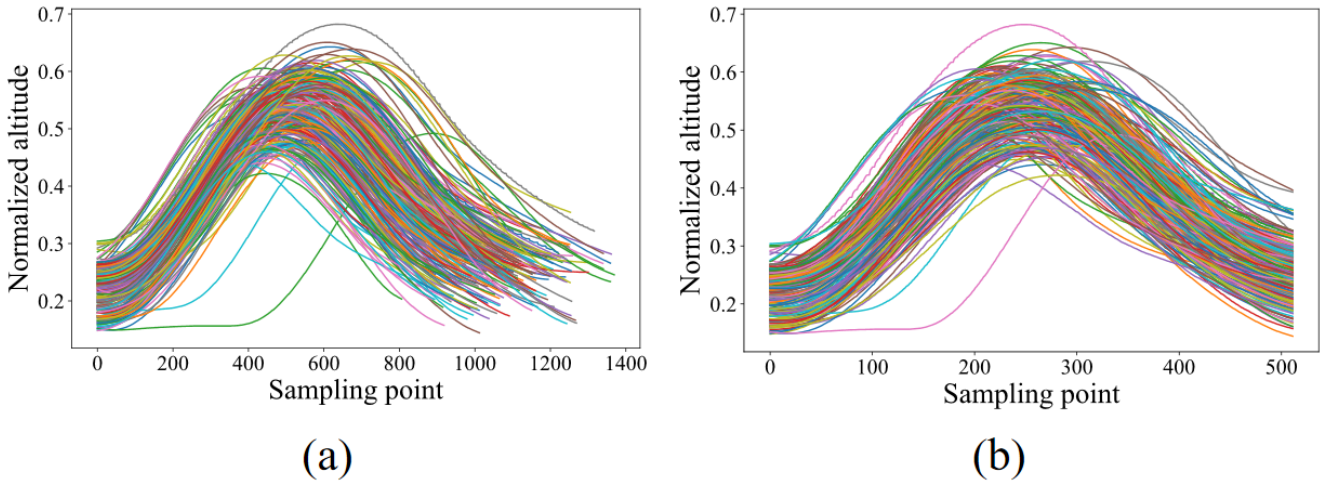


Figure 4. Time Normalization Process: (a) the original altitude data; (b) the altitude data after time normalization.

achieving the highest average precision, recall, and F1 score of 99.75%. It outperforms the second-best MSDRN model by 0.25% in these metrics. The excellent performance of 1D-DAMResNet18 is due to its dual attention mechanism (DAM), which enhances feature extraction. ResNet18 and VGG11 also show high accuracy but are limited by traditional convolutional layers in capturing feature correlations. MobileNet has the lowest scores (97.50%) due to its focus on computational efficiency over accuracy. Overall, the 1D-DAMResNet model excels in recognition performance, making it a strong foundation for anomaly level quantification.

5.3 Quantitative Evaluation Results of Anomaly Levels

5.3.1 Generation of Standard Flight Maneuver Set

First, following the method described in Section 4.1, the duration of different flight maneuvers was normalized. Taking the loop maneuver as an example, the altitude changes before and after time normalization are shown in Figure 4. It is evident from Figure 4 that before normalization, the durations of the flight maneuvers vary significantly, while after normalization, the durations become uniform. This consistency facilitates the subsequent quantification of anomalies and comparative visual analysis. This consistency facilitates subsequent quantitative analysis and visualization of anomaly levels. Next, the K-Means clustering algorithm was applied to the training and validation datasets for each type of flight maneuver, with the number of clusters set to three. The results are shown in Figure 5, where each point represents a flight maneuver sample, different colors indicate different clusters, and the red stars mark the cluster centers. Finally, the sample closest to the center of each cluster was selected to represent the standard flight maneuver for that category. As a result, each flight maneuver category includes three representative samples. This process ensures that the standard flight maneuver set is both representative and accurate, providing a reliable reference for further anomaly quantification and analysis.

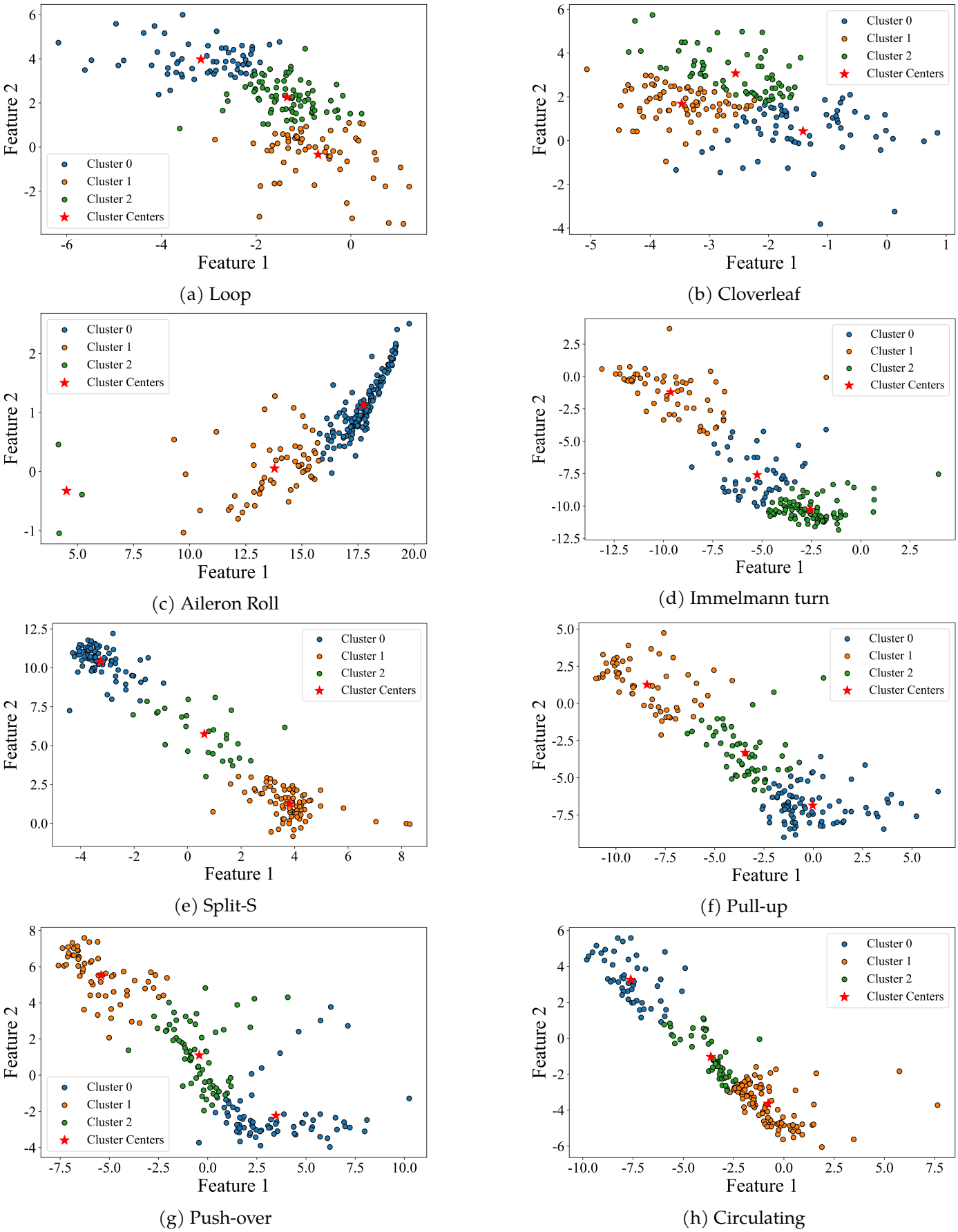
5.3.2 Quantitative Evaluation Results of Anomaly Levels

In this section, we randomly selected one sample from the test set of each flight maneuver category. Using the corresponding standard flight maneuver set for each category, we calculated the similarity score based on Equations (9) and (10). The results are shown in Table 4, where the minimum similarity score between the test flight maneuver and the standard maneuver is highlighted in bold. The corresponding standard flight action is used as the reference for that test sample.

Next, based on the similarity scores in Table 4, we applied Equations (11) and (12) to compute the anomaly quantification scores for each category's test flight maneuver, as shown in Table 6. To establish

Table 4. Similarity Scores of Test Flight Maneuvers for Different Categories.

Test flight maneuver	Standard flight maneuver 1	Standard flight maneuver 2	Standard flight maneuver 3
Loop	0.169	**0.147**	0.423
Cloverleaf	0.353	**0.3097**	0.504
Aileron Roll	0.965	0.922	**0.857**
Immelmann turn	**0.356**	0.387	0.592
Split-S	0.553	**0.531**	0.609
Pull-up	0.628	**0.563**	0.734
Push-over	0.672	0.730	**0.646**
Circulating	0.637	**0.609**	0.760



(a) Loop (b) Cloverleaf

(c) Aileron Roll (d) Immelmann turn

(e) Split-S (f) Pull-up

(g) Push-over (h) Circulating

Figure 5. Clustering Results of Flight Maneuver Samples.

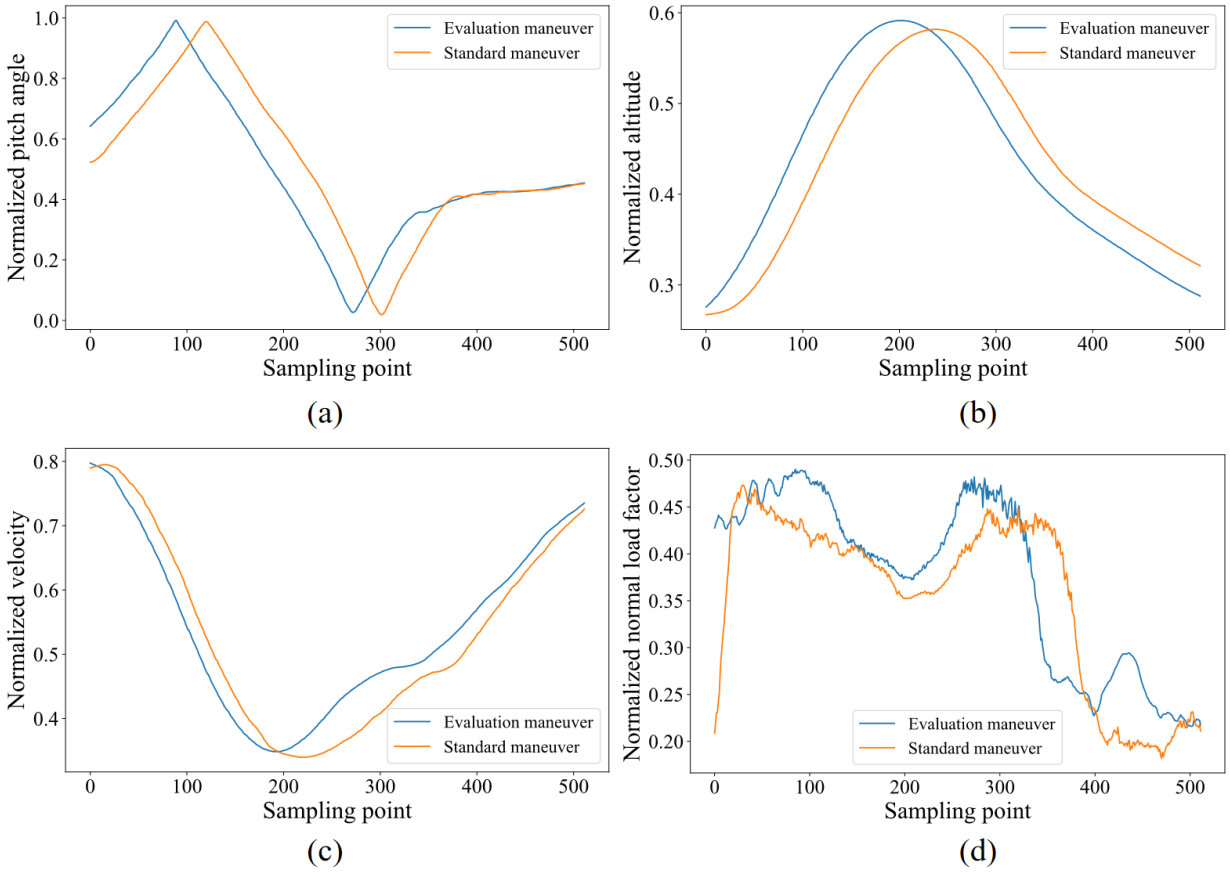


Figure 6. Comparison Curves of Select Parameters Between the Test Sample and Standard Maneuver for the Loop Maneuver: (a) Pitch Angle; (b) Altitude; (c) Velocity; (d) Normal Load Factor.

threshold values for defining different anomaly levels, we selected 10 samples from the training and validation sets with expert ratings closest to 2.5 and 4 points, then calculated their abnormality quantification scores. The mean values of these scores were used to establish the threshold values for each category, denoted as θ_1 and θ_2, as shown in Table 5.

Table 5. Threshold Values for Anomaly Levels of Flight Maneuvers.

Flight maneuver	θ_1	θ_2
Loop	0.653	0.264
Cloverleaf	0.579	0.256
Aileron Roll	0.765	0.342
Immelmann turn	0.834	0.401
Split-S	0.612	0.314
Pull-up	0.654	0.273
Push-over	0.721	0.354
Circulating	0.752	0.342

Based on these threshold values, we define the mapping between abnormality quantification scores and performance grades as follows: scores in the range $[0, \theta_2)$ were classified as "Excellent," scores in the range $[\theta_2, \theta_1]$ as "Good," and scores in the range $(\theta_1, 1]$ as "Poor". Consequently, the anomaly levels for each category's test flight maneuver are shown in Table 6. These results indicate that the proposed quantitative evaluation method for anomaly levels of complex flight maneuver is effective in evaluating the anomaly levels of different flight maneuvers.

Table 6. Anomaly Quantification Scores and Grading Results of Test Flight Maneuvers for Different Categories.

Test flight maneuver	Anomaly Quantification Score	anomaly level
Loop	0.128	Excellent
Cloverleaf	0.236	Excellent
Aileron Roll	0.461	Good
Immelmann turn	0.263	Excellent
Split-S	0.347	Good
Pull-up	0.360	Good
Push-over	0.392	Good
Circulating	0.378	Good

To provide a more intuitive understanding of the

abnormality quantification results, we used the "loop" maneuver as an example. Figure 6 shows the comparison of the test flight maneuver sample and the corresponding standard maneuver sample across four dimensions: pitch angle, altitude, velocity, and normal load factor. It is clear that deviations between the test maneuver and the standard maneuver in these flight parameters can be easily observed, demonstrating the strong interpretability of the proposed method.

6 Conclusion

This paper presents a novel method to quantify anomaly levels in flight maneuvers, overcoming the limitations of current flight data analysis, which primarily detects anomalies at the parameter level. We developed the 1D-DAMResNet neural network, incorporating a one-dimensional dual attention module, achieving 99.75% accuracy in recognizing complex maneuvers. For anomaly quantification, we used clustering to build a standard maneuver database, setting anomaly thresholds based on training standards and instructor scores. This similarity-based approach ensures interpretable results and supports automated flight training evaluation, promising enhanced safety and reduced costs. Future work will explore feature importance across maneuvers to further improve performance.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by the National Natural Science Foundation of China under Grant 62076249.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Tian, W., Zhang, H., Li, H., & Xiong, Y. (2022). Flight maneuver intelligent recognition based on deep variational autoencoder network. *EURASIP Journal on Advances in Signal Processing*, 2022(1), 21. [CrossRef]

[2] Lu, J., Pan, L., Deng, J., Chai, H., Ren, Z., & Shi, Y. (2023). Deep learning for Flight Maneuver Recognition: A survey. *Electronic Research Archive*, 31(1).

[3] Soheily-Khah S, Marteau P F. Sparsification of the alignment path search space in dynamic time warping. *Applied Soft Computing*, 2019, 78: 630-640. [CrossRef]

[4] Vasimalla, K., Challa, N., & Naik, S. M. (2016). Efficient dynamic time warping for time series classification. *Indian Journal of Science and Technology*, 9(21), 1-7.

[5] Trigeorgis, G., Nicolaou, M. A., Schuller, B. W., & Zafeiriou, S. (2017). Deep canonical time warping for simultaneous alignment and representation learning of sequences. *IEEE transactions on pattern analysis and machine intelligence*, 40(5), 1128-1138. [CrossRef]

[6] Wei, Z., Ding, D., Zhou, H., Zhang, Z., Xie, L., & Wang, L. (2020). A flight maneuver recognition method based on multi-strategy affine canonical time warping. *Applied Soft Computing*, 95, 106527. [CrossRef]

[7] Lu, J., Chai, H., & Jia, R. (2022). A general framework for flight maneuvers automatic recognition. *Mathematics*, 10(7), 1196. [CrossRef]

[8] Samuel, K., Gadepally, V., Jacobs, D., Jones, M., McAlpin, K., Palko, K., ... & Kepner, J. (2021, September). Maneuver identification challenge. In *2021 IEEE High Performance Extreme Computing Conference (HPEC)* (pp. 1-7). IEEE. [CrossRef]

[9] Yang, J., & Xie, S. (2005). Aircraft flight action recognition based on fuzzy support vector machine. *Acta Aeronautica Sinica*, 26(6), 738-742.

[10] Shen, Y. C., Ni, S. H., & Zhang, P. (2017). Flight action recognition method based on Bayesian network. *Computer Engineering and Applications*, 53(24), 161-167.

[11] Li, L., Hansman, R. J., Palacios, R., & Welsch, R. (2016). Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring. *Transportation Research Part C: Emerging Technologies*, 64, 45-57. [CrossRef]

[12] Schumann, J., Rozier, K. Y., Reinbacher, T., Mengshoel, O. J., Mbaya, T., & Ippolito, C. (2015). Towards real-time, on-board, hardware-supported sensor and software health management for unmanned aerial systems. *International Journal of Prognostics and Health Management*, 6(1). [CrossRef]

[13] Gupta, M., Gao, J., Aggarwal, C. C., & Han, J. (2013). Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and data Engineering*, 26(9), 2250-2267. [CrossRef]

[14] Puranik, T. G., & Mavris, D. N. (2017). Identifying instantaneous anomalies in general aviation operations. In *17th AIAA Aviation Technology, Integration, and Operations Conference* (p. 3779). [CrossRef]

[15] Bu, J., Sun, R., Bai, H., Xu, R., Xie, F., Zhang, Y., & Ochieng, W. Y. (2017). Integrated method for the UAV navigation sensor anomaly detection. *IET Radar, Sonar & Navigation*, 11(5), 847-853. [CrossRef]

[16] López-Estrada, F. R., Ponsart, J. C., Theilliol, D., Zhang, Y., & Astorga-Zaragoza, C. M. (2016). LPV

model-based tracking control and robust sensor fault diagnosis for a quadrotor UAV. *Journal of Intelligent & Robotic Systems*, 84, 163-177. [CrossRef]

[17] e Silva, L. C., & Murça, M. C. R. (2023). A data analytics framework for anomaly detection in flight operations. *Journal of Air Transport Management*, 110, 102409. [CrossRef]

[18] Yang, L., Li, S., Li, C., Zhu, C., Zhang, A., & Liang, G. (2023). Data-driven unsupervised anomaly detection and recovery of unmanned aerial vehicle flight data based on spatiotemporal correlation. *Science China Technological Sciences*, 66(5), 1304-1316. [CrossRef]

[19] Li, H., Sang, Y., Ge, H., Yan, J., & Li, S. (2024). Anomaly detection of aviation data bus based on SAE and IMD. *Computers & Security*, 137, 103619. [CrossRef]

[20] REN, L., JIA, S., WANG, H., & WANG, Z. (2024). A Review of Research on Time Series Classification Based on Deep Learning. *Journal of Electronics & Information Technology*, 46(8), 3094-3116. [CrossRef]

[21] FANG W, WANG Y, YAN W J. (2022). Symbolized flight action recognition based on neural network. *Systems Engineering and Electronics*, 44(3), 737-745.

[22] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[23] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision* (*ECCV*) (pp. 3-19).

[24] Loshchilov, I., & Hutter, F. (2017). Fixing weight decay regularization in adam. *arXiv preprint arXiv:1711.05101*.

[25] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826). [CrossRef]

[26] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. [CrossRef]

[27] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

Liqiang Ren Ph.D. candidate, research interests include information fusion, intelligent analysis and engineering applications of flight data. (Email: 1194153993@qq.com)

Haipeng Wang Male, Professor, research interests include information fusion. (Email: whp5691@163.com)

Xinlong Pan Associate Professor, research interests include information fusion. (Email: airadar@126.com)

Shuyi Jia Female, Associate Professor, research interests include information fusion.

Bing Wan Ph.D., research interests include flight control theory and engineering applications. (Email: 2412696166@qq.com)