Journal of Artificial Intelligence and Technology, 2026, 6, 1-8
https://doi.org/10.37965/jait.2025.0589

ISTP
RESEARCH ARTICLE

Paper Drone Design Based on Pose Estimation Technology

Jiazheng Wang and Parvathy Rajendran
School of Aerospace Engineering, Universiti Sains Malaysia, Nibong Tebal, Pulau Pinang, Malaysia

(Received 12 July 2024; Revised 04 November 2024; Accepted 06 November 2024; Published online 30 January 2025)

Abstract: Paper drones, as a low-cost green new type of drone, have played a role in various fields. However, traditional pose estimation methods have high requirements for cost and environment, and they are not suitable for paper drones. Therefore, in the context of artificial intelligence, a data augmentation method and vision-based pose estimation technology are proposed using deep learning algorithms. This technology is used to achieve accurate pose estimation and effective motion control of paper drones, and the effectiveness of this technology is verified. The experimental results show that the highest accuracy achieved by using data augmentation methods is 96.31%, and it can enhance the generalization performance of the algorithm. When using the technology proposed in the study for pose estimation, the average errors of roll angle, pitch angle, and yaw angle are 0.07°, 0.18°, and 0.31°, respectively. When using the technology proposed in the study for motion control, the flight path is closest to the specified path. Research can effectively improve the pose estimation and motion control performance of paper drones, providing novel methods and ideas for the design of paper drones, which is of great significance for promoting the intelligent development of drone technology.

Keywords: artificial intelligence; data augmentation; motion control; paper drones; pose estimation

I. INTRODUCTION

The updates and iterations of technology are leading the development of the times, and the mature drone technology has brought many conveniences to people's lives and work [1]. Drones have the characteristics of time-saving, labor-saving, and easy deployment, playing an important role in military reconnaissance, environmental monitoring, logistics distribution, and other fields [2]. With the rapid development of artificial intelligence technology, intelligent drones have achieved breakthroughs in image recognition, data processing, and other areas with the help of deep learning algorithms [3]. Paper drones, as a new type of drone that uses biodegradable materials to reduce costs and environmental impact, provide a new direction for the development of drone technology, which is in line with the green development concept advocated by the country [4]. Compared with traditional drones, paper drones have the characteristics of low cost, high portability, high maneuverability, and high flexibility [5]. However, due to the limitations of physical properties of paper materials, the flight performance of paper drones may not be as good as traditional drones. Therefore, when designing paper drones, the requirements for their flight control algorithms are more stringent, requiring high-precision pose estimation technology and motion control technology to assist [6,7]. However, traditional drone flight control relies on inertial measurement units and Global Positioning System (GPS) for pose and position calculation, which contradicts the low-cost design goal of paper drones, and the performance of traditional methods is not high in environments with weak GPS signals [8,9]. In view of this, the study focuses on the design and motion control of paper drones and proposes a vision-based pose estimation technology (VbPET) using deep learning algorithms in artificial intelligence technology. The research aims to improve the accuracy and robustness of paper drone pose estimation, thereby enhancing its motion control performance in automated tasks and promoting the development of paper drone technology.

The research faces two major challenges: first, paper materials limit flight performance and require more accurate flight control algorithms. The second challenge is the need to achieve high-precision attitude estimation without relying on GPS. The innovation of this research is to propose a VbPET specially designed for paper UAV to improve the accuracy of attitude estimation and motion control performance of paper UAV. VbPET optimizes the flight control of paper drones by introducing data enhancement and deep learning algorithms, including self-attention mechanisms and spatial pyramid pool, and does not rely on GPS, solving the limitations of traditional methods in weak signal environments.

The research is divided into five parts. Section II introduces the current research on drone technology worldwide. Section III provides a detailed explanation of data augmentation algorithms and pose estimation techniques. Section IV focuses on conducting experiments on the innovative methods proposed in the study to verify the effectiveness of the research content. Section V summarizes the entire article, pointing out shortcomings and future research directions.

II. RELATED WORK

The rapid development of drone technology not only promotes technological innovation in related industries but also provides new ways for society to develop technological intelligence. In addition, the emergence of artificial intelligence has opened new doors for drone technology. Therefore, a large number of scholars around the world have conducted research on artificial intelligence-oriented drone technology in various fields. Mandloi D *et al.* proposed a path optimization algorithm based on an

Corresponding author: Parvathy Rajendran (e-mail: aeparvathy@163.com).

improved A * algorithm to address the problem of optimal path selection and obstacle avoidance for drones, thereby improving the efficiency of drone path planning and enhancing the performance of autonomous intelligent flight [10]. Wang X *et al.* proposed an intelligent drone-assisted fault diagnosis algorithm using drone technology combined with deep deterministic strategy gradients to address the issue of low reliability of intelligent devices in 5G space air ground ocean networks, thereby improving the performance of fault detection and multi-fault classification [11]. James K *et al.* proposed a detection algorithm that combines practical environmental changes for high-resolution imaging applications of drones in plant species mapping, thereby improving the accuracy of drone detection of plant species in the field and providing a new approach for drone operations in the field [12]. Ramadass L *et al.* proposed a drone automatic inspection system that combines the YOLOv3 algorithm to address the issues of maintaining social distance and wearing masks in public places. This system can automatically send alarm signals to nearby police stations and issue warnings to the public when regulations are not followed [13].

Pose estimation and motion control can improve the performance and application range of drones, which is a key technology in drone design and the foundation for achieving autonomous and intelligent drones. Kahouadji M *et al.* proposed a drone control technology based on an improved super torsion control algorithm to address the pose control problem of quadcopters in uncertain and interference filled environments, thereby eliminating vibration and effectively improving drone control accuracy [14]. Ulus Ş proposed a control algorithm that combines artificial neural fuzzy inference with Proportional-Integral-Derivative and other controllers to effectively improve the control effect of small drones used for pesticide spraying and weed control in the agricultural field under uncertain weather conditions and interference effects [15]. Wang B *et al.* proposed a composite adaptive fault-tolerant control strategy combining neural networks for the control problem of quadcopter drones, constructed a neural network adaptive control scheme, and incorporated it into baseline sliding mode control for processing, thereby improving the control performance of quadcopter drones [16]. Reinhardt *et al.* proposed an improved active disturbance suppression control scheme for the stability of a six-degree-of-freedom quadcopter system in the face of external disturbances and system uncertainties. Four active disturbance suppression control units were designed using a nonlinear model, thereby improving the altitude control and pose stability of unmanned aerial vehicles [17]. De La Rosa applied machine learning technology to address the multirotor activity recognition problem and analyze the flight data of drones to rebuild its trajectory [18].

In summary, the innovation of drone technology in various fields is closely related to artificial intelligence technology, and many scholars have provided new methods and ideas for drone pose estimation technology. However, from existing research results, it can be seen that there is almost no content related to the design of new paper drones, and the pose estimation and motion control problems for paper drone design urgently need to be solved. In this context, research focuses on paper drones and designs VbPET to fill the gap in this field. The innovation of the research lies in the data augmentation operation on the pose estimation data sequence of paper drones, and the introduction of self-attention mechanism and spatial pyramid pooling (SPP) to improve the performance of monocular vision estimation network, thereby improving the pose estimation and motion control performance of paper drones.

III. DESIGN OF PAPER DRONES BASED ON POSE ESTIMATION TECHNOLOGY

A. DATA AUGMENTATION METHODS IN POSE ESTIMATION AND MOTION CONTROL

In response to the limitations of traditional drone pose estimation and motion control, a study is conducted using deep learning algorithms in artificial intelligence to propose VbPET. When training a model using deep learning algorithms, it is necessary to use large-scale datasets for fast computation. However, paper drones have the characteristics of being lightweight and flexible. When operating paper drones, their motion is often not uniform, and the motion path is not fixed due to external factors such as wind speed [19,20]. Therefore, the workload of collecting all posture and motion data is extremely high. If the sample size in the dataset is small, it will limit the generalization ability of the algorithm, thereby affecting the accuracy of VbPET. Therefore, before introducing VbPET, the study first proposed a Mask-based Data Augmentation (MbDA) method for pose estimation. The core idea of the MbDA method is to randomly mask a portion of frames in the image sequence in the dataset. Specifically, while keeping the overall displacement and rotation of the paper drone basically unchanged, MbDA changes the average speed by changing the number of frames in the sequence, thereby introducing more samples with different speeds in the dataset and enhancing the generalization performance of VbPET.

Research suggests that the core objective of pose estimation tasks is to determine the motion changes of the camera in two consecutive frames of the image, namely the camera's translation and rotation. This process is divided into two stages, namely feature extraction and pose regression. Among them, feature extraction is the first step of pose estimation task, and features usually include key visual elements such as edges, corners, and the textures in the image. Assuming two input images are I_1 and I_2, respectively, the extracted features are shown in equation (1):

$$F = f(I_1, I_2) \tag{1}$$

In equation (1), F represents the feature and $f(\cdot)$ represents the operation related to feature extraction. Further process the extracted feature F and map it to an N dimensional space, with the aim of transforming the feature into points in a high-dimensional space for subsequent pose regression, as shown in equation (2):

$$a^F = E(f(I_1, I_2)) \tag{2}$$

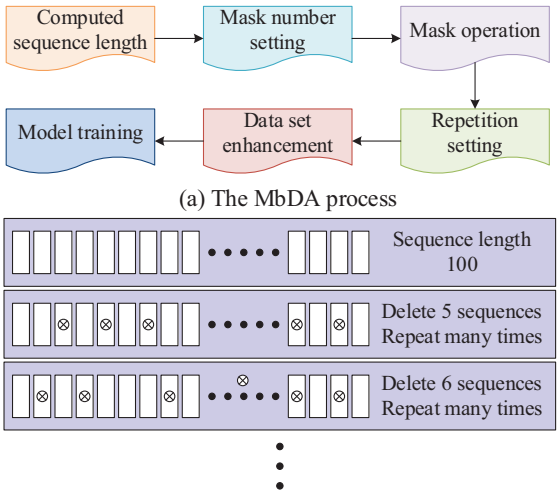
In equation (2), a^F represents a vector composed of features, where each feature is an element in F and $E(\cdot)$ represents the mapping function. The a^F expression is shown in equation (3):

$$a^F = (a_1^F, a_2^F, a_3^F, \ldots, a_N^F)^T \tag{3}$$

Finally, the embedded feature a^F is mapped onto the pose transformation relationship using the pose estimation function, as shown in equation (4):

$$\begin{aligned} G(I_1, I_2) &= b + W \cdot E(f(I_1, I_2)) \\ &= b + \textstyle\sum_{i=1}^{N} a_i^F W_i = \begin{pmatrix} T \\ R \end{pmatrix} \end{aligned} \tag{4}$$

In equation (4), $G(\cdot)$ represents the pose estimation function, b represents the bias term, W represents the weight matrix used to map features to the pose space, T represents the displacement of two input images, and R represents the rotation of two input images. For a fixed frame rate camera, the time difference between



(a) The MbDA process

(b) MbDA operation diagram with sequence length of 100

Fig. 1. The specific process of the Mask-based Data Augmentation method.

two input images is also fixed. Therefore, T and R can be used to obtain the average linear velocity and average angular velocity of the paper drone in the two input images, as shown in equation (5):

$$\begin{pmatrix} T \\ R \end{pmatrix} = \begin{pmatrix} \overline{v} \cdot \Delta t \\ \overline{w} \cdot \Delta t \end{pmatrix} \tag{5}$$

In equation (5), $\overline{v}$ represents the average linear velocity, $\overline{w}$ represents the average angular velocity, and Δt represents the time difference, which is inversely proportional to the camera frame rate. It can be inferred that when MbDA changes the number of frames in the input image sequence, the camera frame rate will change, which in turn affects the $\overline{v}$ and $\overline{w}$ of the paper drone, thus completing the data augmentation work on the dataset. Therefore, the specific process of constructing the MbDA is shown in Fig. 1.

As shown in Fig. 1(a), MbDA is mainly divided into six steps. First, it calculates the sequence length of paper drone flight data. For example, in the 7Sense dataset containing image sequences for drone pose estimation and motion control research, a single sequence may contain 1000 consecutive flight state data. Then it determines the number of Masks, which is the number of frames to be masked in the sequence. When the sequence length is greater than 100, research suggests that the number of masks should be between 5 and 50. After determining the quantity, it is necessary to perform a Mask operation and randomly select frame sequences and delete them. To ensure the diversity of data augmentation operations, after the first deletion, it is necessary to perform duplicate deletion and set the number of masks before repeating deletion, and each deletion operation randomly selects a frame sequence. After generating multiple subsequences with different velocity characteristics, data augmentation of the dataset is completed, and finally VbPET can be used for training to improve its generalization performance.

When using pose estimation for motion control of paper drones, it is also necessary to perform data augmentation on the motion control dataset. Due to the influence of external environmental factors and internal sensors on the path generated by paper drones from the starting point to the end point, there is diversity in the actual path, and collecting all path samples requires a huge amount of calculation [21]. In fact, to save time, the motion path of

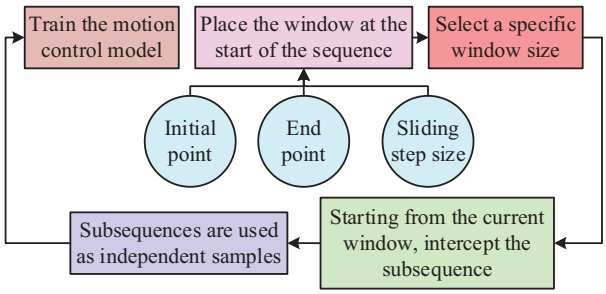


Fig. 2. Data augmentation operations based on sliding windows.

paper drones is usually designed based on the shortest path, so there must be repeated common parts in all path samples [22]. In response to this phenomenon, research proposes using sliding windows to perform data augmentation operations, as shown in Fig. 2.

As shown in Fig. 2, when applying the sliding window method, the first step is to set the sequence start point, end point, window size, and sliding step size. Among them, the starting and ending points determine the range of the original sequence, the window size determines the length of the subsequence, and the sliding step determines the number of subsequences. Subsequently, the subsequence is intercepted, a window is placed at the starting point of the original sequence, and a specific window size is selected from the set window size range. Then, starting from the current window, the sliding window is used to sequentially capture subsequences until the end point. Finally, each captured subsequence is treated as an independent sample and used to train the paper drone motion control model, thereby completing the data augmentation of the motion control training set.

B. POSE ESTIMATION AND MOTION CONTROL BASED ON VbPET

On the basis of data augmentation, a detailed study to VbPET is conducted, and the pose estimation and motion control in paper drone design are completed based on VbPET. The VbPET proposed in the study selected a monocular depth estimation network based on Laplacian and residual structures as the backbone network. On this basis, ACMix self-attention mechanism and SPP are added to the feature extraction network to enhance the scale information of features and enhance feature diversity. The application of ACMix self-attention mechanism in VbPET feature extraction network is shown in Fig. 3.

From Fig. 3, the first step is to use a monocular depth estimation network to process the RGB images captured by the drone camera. It obtains an RGBD image that combines the RGB image and depth estimation results. “D” represents depth information, which can provide three-dimensional position information of objects in the scene. Subsequently, residual networks are used to extract deep features of the image. ACMix combines convolutional networks and self-attention mechanisms to study the use of ACMix to capture long-distance dependencies between different regions in the initially obtained feature images, resulting in new multidimensional feature images. However, if the new multidimensional feature map is directly flattened into a one-dimensional vector input into the regression network, some important information in the spatial structure that affects the pose estimation of paper drones will be ignored due to the inability of one-dimensional vectors to

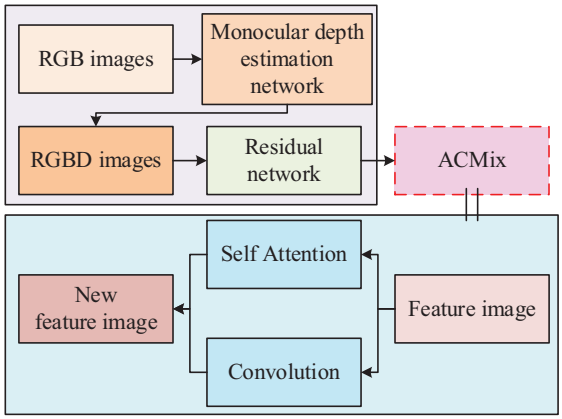


Fig. 3. ACMix self-attention mechanism in feature extraction network.

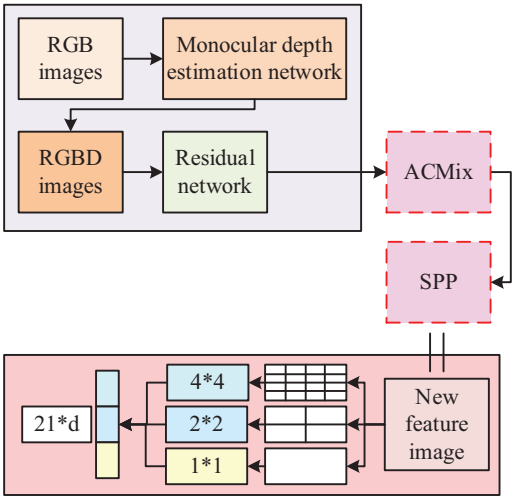


Fig. 4. Feature extraction network after SPP optimization.

preserve the spatial relationships and scale information in the original feature map [23,24]. Therefore, the study combines SPP to optimize the feature extraction network, as shown in Fig. 4.

From Fig. 4, SPP performs pooling operations on the new multidimensional feature maps obtained using ACMix. Specifically, the study sets SPP to use pooling kernels of 4 * 4, 2 * 2, and 1 * 1 to capture spatial feature information at different scales, divide the feature maps into multiple subregions, and perform average pooling operations on each subregion. Finally, the pooling results of three different scales are concatenated to form a multi-scale feature vector of (16+4+1) * Channel, achieving feature diversity.

According to equation (4), when estimating the pose of a paper drone based on image features, the final target to be obtained is the displacement T of the image and the selection R of the image. If this is considered as a classification task, that is, to preset the drone T and R based on the maximum flight speed and maximum rotation angular velocity of the paper drone, and discretize the preset space to form a finite number of possible states, learning the relationship between input image frames and discretized states will cause the loss of continuous pose and position information of the paper drone, reducing estimation accuracy [25]. Therefore, the research regards pose estimation task as a regression task, which utilizes the features extracted from the optimized VbPET feature extraction network and uses a regressor to predict the T and R of the drone. The paper drone pose estimation network based on VbPET studied and constructed is shown in Fig. 5.

From Fig. 5, it can be seen that for each pose estimation, two frames of RGBD images at time t and time t+1 are selected, and the two RGBD images are stacked to form new image data, which is then processed through a feature extraction network. It is worth noting that the study incorporated Long Short-Term Memory (LSTM) into the pose network, which serves as a recurrent neural network for processing sequential data and can effectively learn the temporal information of input data. The output of LSTM is processed through a fully connected layer to obtain pose information. After calculating the poses at time t and time t+1, the network will receive the images at time t+1 and time t+2 as new inputs and repeat the above operation to achieve continuous pose estimation of the drone. In the process of pose estimation, traditional methods use Euler angles for estimation, but due to the fact that the universal joint lock can result in multiple different Euler angle representations of the same pose, it does not have uniqueness [26]. Therefore, the study uses quaternions to replace Euler angles, as shown in equation (6):

$$\begin{cases} q_x = \cos(\frac{\alpha}{2}) + 0 \cdot i + 0 \cdot j + (\sin\frac{\alpha}{2}) \cdot k \\ q_y = \cos(\frac{\beta}{2}) + (\sin\frac{\beta}{2}) \cdot i + 0 \cdot j + 0 \cdot k \\ q_z = \cos(\frac{\gamma}{2}) + 0 \cdot i + (\sin\frac{\gamma}{2}) \cdot j + 0 \cdot k \end{cases} \tag{6}$$

In equation (6), q_x, q_y, and q_z represent the quaternions corresponding to rotation around the X, Y, and Z axes, respectively. α, β, and γ are Euler angles, representing the roll angle for rotation around the X axis, the pitch angle for rotation around the Y axis, and the yaw angle for rotation around the Z axis. i, j, and k are imaginary units, respectively. Finally, q_x, q_y, and q_z are multiplied in order of rotation to obtain quaternions. Finally, the

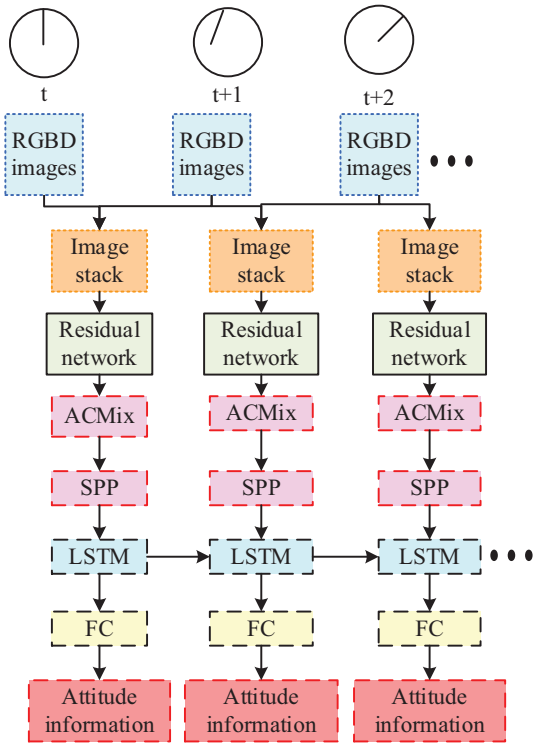


Fig. 5. Pose estimation network based on VbPET.

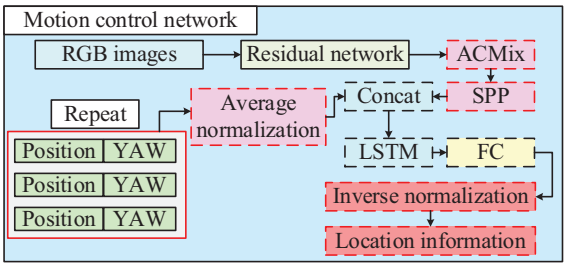


Fig. 6. Motion control network based on VbPET.

core idea of constructing a paper drone motion control network based on VbPET is to combine feature data and position data to predict the future flight status of paper drones, as shown in Fig. 6.

As shown in Fig. 6, in the motion control network, the direct combination of position information data and image feature data may cause problems due to their different magnitudes and distributions. Therefore, the study repeats the feature vectors of position information to ensure dimensional compatibility when merging position information and image features. In addition, to maintain the stability of the motion control network, the study also performs average normalization on the input position feature data. Finally, during output, the position and yaw angle data of the paper drone are restored through inverse normalization to obtain real physical quantities, completing prediction and motion control.

IV. VERIFICATION OF POSE ESTIMATION TECHNOLOGY FOR PAPER DRONES

A. VALIDATION AND ANALYSIS OF DATA AUGMENTATION METHODS

In order to verify the effectiveness of the proposed MbDA method and the sliding window method, the 7Sense dataset was selected to test the two methods, which included a total of 7 application scenarios. 7Sense, a publicly available annotated dataset widely used in drone research, is shared by the drone researcher and developer community and contains flight image sequences for a variety of environments and conditions, each with a corresponding attitude label. The 7Sense dataset provides sequences of RGBD camera frames from multiple scenes for drone research, each of which is compressed into a zip file containing 500 to 1,000 frames. Each frame consists of an RGB color map, a depth map (in millimeters, invalid depth value is 65535), and a 4×4 homogeneous coordinate matrix for recording camera poses. In addition, each scene is provided with a split file for evaluation and a Truncated

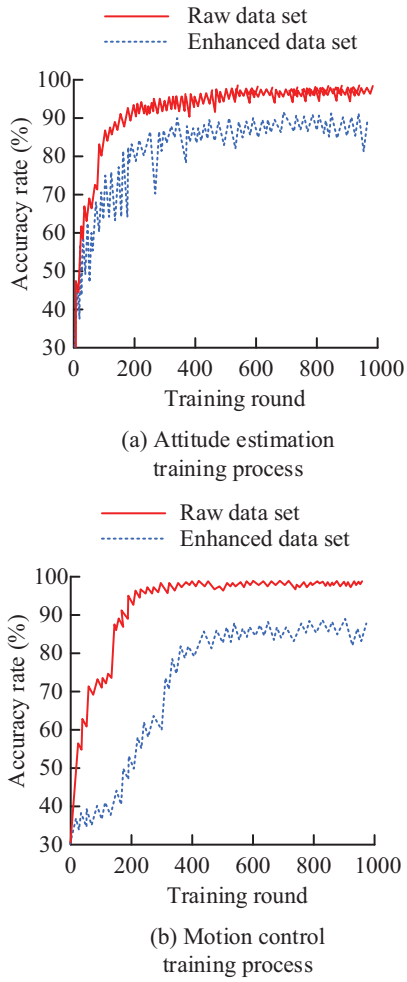


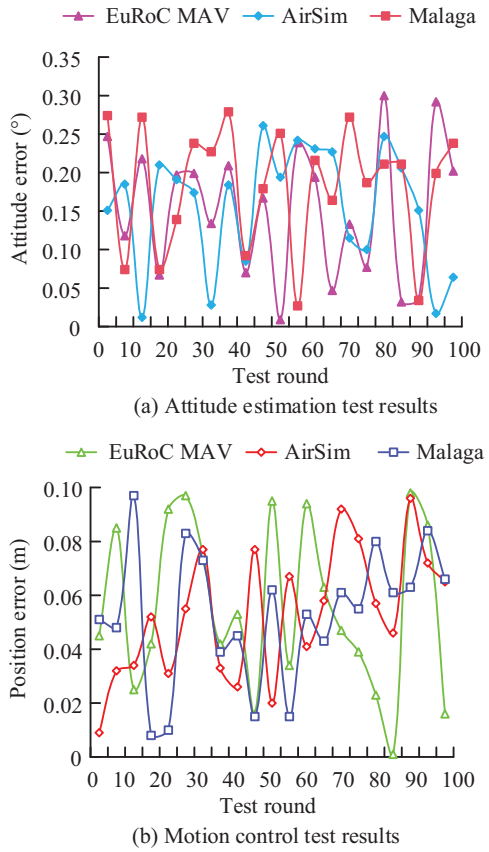
Fig. 7. Experimental results of data enhancement methods.

Signed Distance Function volume file for frame-to-model alignment, the latter stored in MetaImage format with 512x512x512 voxel resolution. Rafique A. *et al.* proposed a new framework for robot reality segmentation and conducted experiments on 7Sense, achieving an accuracy of 74.85%, which is better than the comparison method [27]. The experimental software and hardware configuration and parameter settings are shown in Table I.

In Table I, the initial learning rate of the research set algorithm was 0.0001, which increased linearly to 0.003 in the first 6 rounds of training. Then, for every 200 rounds of training, the learning rate was halved. During the training process, the study used the original

Table I. Experimental hardware and software configuration

Operating environment		Parameter configuration	
CPU	Intel Core I7 13700KF	Momentum	0.90
GPU	NVIDIA RTX 3090	Attenuation	0.0004
Internal memory	64G	Training round	1000
Operating system	Ubuntu22.04	Initial learning rate	0.0001
Programming language	Python 3.9	Attenuation cycle	200
Deep learning framework	Pytorch 1.16	Attenuation rate	1/2
Dataset	7Sense	/	/



(a) Attitude estimation test results

(b) Motion control test results

Fig. 8. Test results from different datasets.

7Sense dataset for training as a comparative method to explore the differences after training using data augmentation methods on the 7Sense dataset. The result is shown in Fig. 7.

From Fig. 7(a), for pose estimation, the estimation accuracy for the original dataset was around 90.00% after 1000 training sessions, while the highest estimation accuracy reached 96.31% after using data augmentation methods. As shown in Fig. 7(b), when performing motion control, the accuracy of motion control trained on the dataset using data augmentation methods was also better than the original dataset, with the former reaching 97.81% and the latter only 88.34%. From this, the MbDA method and sliding window method proposed in the study could effectively improve accuracy during the training process. After training, to verify the generalization performance improvement effect of the data augmentation method, the EuRoC MAV dataset, AirSim dataset, and Malaga dataset were selected as the test sets for a total of 100 tests. The results are shown in Fig. 8.

According to Fig. 8(a), in the three test sets, the average pose errors of the X, Y, and Z axes did not exceed 0.30 °. Among them, the EuRoC MAV dataset even had an error close to 0 ° in the 53rd test, and the AirSim dataset also had multiple tests approaching 0 °. As shown in Fig. 8(b), in the motion control test, the maximum error occurred during the 12th test on the Malaga dataset, with an average position error of 0.096 m. However, the error in all test results did not exceed 0.100 m. From this, after data augmentation dataset testing, pose estimation and motion control showed good performance in different test sets, with outstanding generalization ability.

B. POSE ESTIMATION AND MOTION CONTROL VERIFICATION ANALYSIS BASED ON VbPET

To verify the effectiveness of the VbPET proposed in the study, ablation experiments were conducted first, with a total of four experimental methods setup. Error rate (E), Precision rate (P), and F1 score (F1) were selected as validation indicators, and the data enhanced 7Sense dataset was used as the test and validation sets. The specific methods and results are shown in Fig. 2.

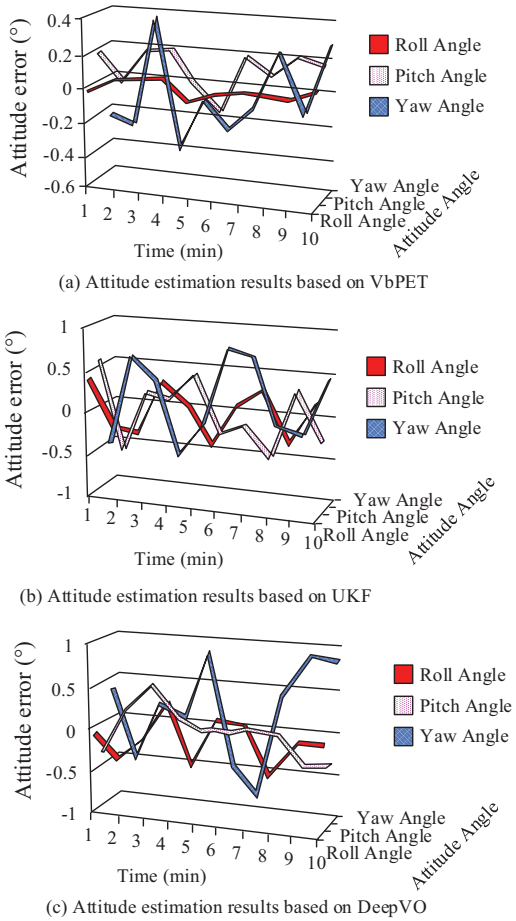
According to Table II, when using only the backbone network for monocular depth estimation, the E value was 9.46%, the P value was 83.48%, and the F1 was 84.37%. When adding ACMix on the basis of the backbone network without using SPP or adding SPP without using ACMix, the E value decreased and the P value and F1 increased. For method 4, which simultaneously incorporated ACMix and SPP, that is, the VbPET proposed in the study, the E value was reduced to 3.46%, the P value was increased to 97.46%, and the F1 value was increased to 97.53%. This indicated that the addition of ACMix and SPP could improve the performance of the network, and when combined with ACMix and SPP, the performance improvement was the most significant, verifying the effectiveness and superiority of VbPET. On this basis, the effectiveness and superiority of pose estimation based on VbPET were verified using paper drones in real scenarios. Unscented Kalman Filter (UKF) and Deep Visual Odometry (DeepVO) were selected as comparison algorithms, and the results are shown in Fig. 9.

As shown in Fig. 9(a), for VbPET, the average error of roll angle, pitch angle, and yaw angle during a 10-minute flight of a paper drone were 0.07°, 0.18°, and 0.31°, respectively. As shown in Fig. 9(b), for UKF, the average errors in pose estimation for roll angle, pitch angle, and yaw angle were 0.19°, 0.34°, and 0.47°, respectively. As shown in Fig. 9(c), for DeepVO, the average errors in pose estimation for roll angle, pitch angle, and yaw angle were 0.20°, 0.31°, and 0.54°, respectively. From this, pose estimation based on VbPET had the smallest error and the highest accuracy. Finally, the motion control based on VbPET was validated, and UKF and DeepVO were also selected as comparison algorithms. The results are shown in Fig. 10.

As shown in Fig. 10, under the motion control based on VbPET, the flight path of the paper drone was almost in line with the specified flight path, with only a significant deviation at the end of the path. Both UKF-based motion control and DeepVO-based motion control caused significant deviation in the flight path

Table II. Results of ablation experiment

Method number	Network structure			Validation index		
	Backbone network	AC Mix	SPP	E	P	F1
1	√	×	×	9.46%	83.48%	84.37%
2	√	√	×	6.81%	89.71%	88.64%
3	√	×	√	6.27%	91.16%	90.33%
4	√	√	√	3.46%	97.46%	97.53%



(a) Attitude estimation results based on VbPET

(b) Attitude estimation results based on UKF

(c) Attitude estimation results based on DeepVO

Fig. 9. Comparison of experimental results of pose estimation algorithms.

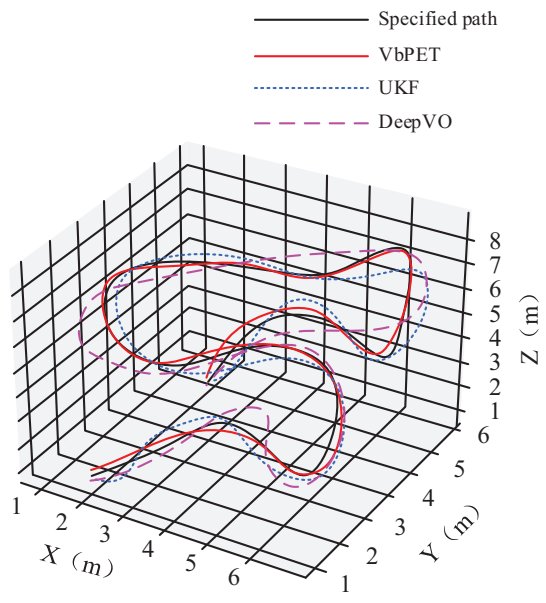


Fig. 10. Results of motion control comparison experiment.

of paper drones, with DeepVO even showing a deviation of nearly 50 cm. From this, the motion control performance based on VbPET was good, effective, and superior.

V. CONCLUSIONS

In response to the challenges faced by paper drones in pose estimation and motion control, the research first proposed MbDA method and sliding window method for data augmentation to address the problem of insufficient flight data. Second, ACMix self-attention mechanism and SPP were added to the monocular depth estimation network to construct a VbPET network. Finally, a method for pose estimation and motion control was proposed based on VbPET, and the effectiveness and superiority of the research technology were verified. The experimental results showed that the accuracy of training completion did not exceed 90.00% without the use of data augmentation methods, while the accuracy exceeded 96.00% with the use of data augmentation methods. In addition, after data augmentation, the pose estimation error did not exceed 0.30° in all three test sets, and the motion control error did not exceed 0.100 m. The results of ablation experiments showed that the combination of ACMix and SPP could significantly improve the performance of monocular depth estimation networks, reducing the E value to 3.46%, increasing the P value to 97.46%, and increasing the F1 value to 97.53%. Meanwhile, the pose estimation based on VbPET had an average roll angle error of 0.07°, pitch angle error of 0.18°, and yaw angle error of 0.31° in the data enhanced 7Sense dataset, which was significantly lower than the comparison algorithms. The motion control performance based on VbPET was significantly better than the comparison algorithm, and the actual flight path was closest to the specified path. Overall, the method proposed in the study was effective and had practical application potential for the design of paper drones. However, further validation is needed for the robustness of research methods in complex environments. Future research can explore the application of VbPET to more types of drones and more complex tasks.

CONFLICT OF INTEREST STATEMENT

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] L. Wang and A. A. Siddique, "Facial recognition system using LBPH face recognizer for anti-theft and surveillance application based on drone technology," *Meas. Control*, vol. 53, nos. 7–8, pp. 1070–1077, 2020.

[2] A. Upadhyaya, P. Jeet, P. K. Sundaram, A. K. Singh, K. Saurabh, and M. Deo, "Efficacy of drone technology in agriculture: a review: drone technology in agriculture," *J. AgriSearch*, vol. 9, no. 3, pp. 189–195, 2022.

[3] M. Mohammed, N. A. Hazairin, S. Al-Zubaidi, A. K. Sairah, S Mustapha, and E. Yusuf, "Toward a novel design for coronavirus detection and diagnosis system using IoT based drone technology," *Int. J. Psychosoc. Rehabil.*, vol. 24, no. 7, pp. 2287–2295, 2020.

[4] J. Demuyakor, "Ghana go digital Agenda: 'The impact of zipline drone technology on digital emergency health delivery in Ghana'," *Humanities*, vol. 8, no. 1, pp. 242–253, 2020.

[5] Y. J. Liang and Z. X. Luo, "A survey of truck–drone routing problem: literature review and research prospects," *J. Oper. Res. Soc. China*, vol. 10, no. 2, pp. 343–377, 2022.

[6] S. Meng, Z. Gao, Y. Zhou, B. He, and A. Djerrad, "Real-time automatic crack detection method based on drone," *Comput.-Aid. Civil Infrastruct. Eng.*, vol. 38, no. 7, pp. 849–872, 2023.

[7] N. Lin, L. Fu, L. Zhao, G. Min, A. Al-Dubai, and H. Gacanin, "A novel multimodal collaborative drone-assisted VANET networking model," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4919–4933, 2020.

[8] A. M. Alanen, A. M. Räisänen, L. C. Benson, and K. Pasanen, "The use of inertial measurement units for analyzing change of direction movement in sports: a scoping review," *Int. J. Sports Sci. Coach.*, vol. 16, no. 6, pp. 1332–1353, 2021.

[9] G. Coviello and G. Avitabile, "Multiple synchronized inertial measurement unit sensor boards platform for activity monitoring," *IEEE Sens. J.*, vol. 20, no. 15, pp. 8771–8777, 2020.

[10] D. Mandloi, R. Arya, and A. K. Verma, "Unmanned aerial vehicle path planning based on A* algorithm and its variants in 3d environment," *Int. J. Syst. Assur. Eng. Manag.*, vol. 12, no. 5, pp. 990–1000, 2021.

[11] X. Wang, H. Lin, H. Zhang, D. Miao, Q. Miao, and W. Liu, "Intelligent drone-assisted fault diagnosis for B5G-enabled space-air-ground-space networks," *IEEE Trans. Netwk. Sci. Eng.*, vol. 8, no. 4, pp. 2849–2860, 2020.

[12] K. James and K. Bradshaw, "Detecting plant species in the field with deep learning and drone technology," *Methods Ecol. Evol.*, vol. 11, no. 11, pp. 1509–1519, 2020.

[13] L. Ramadass and S. Arunachalam, "Applying deep learning algorithm to maintain social distance in public place through drone technology," *Int. J. Pervasive Comput. Commun.*, vol. 16, no. 3, pp. 223–234, 2020.

[14] M. Kahouadji, M. R. Mokhtari, A. Choukchou-Braham, and B. Cherki, "Real-time attitude control of 3 DOF quadrotor UAV using modified super twisting algorithm," *J. Franklin Inst.*, vol. 357, no. 5, pp. 2681–2695, 2020.

[15] Ş. Ulus and I. Eski, "Neural network and fuzzy logic-based hybrid attitude controller designs of a fixed-wing UAV," *Neural Comput. Appl.*, vol. 33, 14, pp. 8821–8843, 2021.

[16] B. Wang, Y. Zhang, and W. Zhang, "A composite adaptive fault-tolerant attitude control for a quadrotor UAV with multiple uncertainties," *J. Syst. Sci. Complex.*, vol. 35, 1, pp. 81–104, 2022.

[17] A. A. Najm and I. K. Ibraheem, "Altitude and attitude stabilization of UAV quadrotor system using improved active disturbance rejection control," *Arab. J. Sci. Eng.*, vol. 45, no. 3, pp. 1985–1999, 2020.

[18] M. De La Rosa and Y. Chen, "A machine learning platform for multirotor activity training and recognition," *Proceedings of the 2019 IEEE 14th International Symposium on Autonomous Decentralized Systems (ISADS)*, Utrecht, Netherlands, 2019, pp. 1–8.

[19] M. A. Siddiqi, C. Iwendi, K. Jaroslava, and N. Anumbe, "Analysis on security-related concerns of unmanned aerial vehicle: attacks, limitations, and recommendations," *Math. Biosci. Eng.*, vol. 19, no. 3, pp. 2641–2670, 2022.

[20] F. Ahmed, J. C. Mohanta, A. Keshari, and P. S. Yadav, "Recent advances in unmanned aerial vehicles: a review," *Arab. J. Sci. Eng.*, vol. 47, 7, pp. 7963–7984, 2022.

[21] A. Ait Saadi, A. Soukane, Y. Meraihi, A. Benmessaoud Gabis, S. Mirjalili, and A. Ramdane-Cherif, "UAV path planning using optimization approaches: a survey," *Arch. Comput. Methods Eng.*, vol. 29, no. 6, pp. 4233–4284, 2022.

[22] S. Zhang and R. Zhang, "Radio map-based 3D path planning for cellular-connected UAV," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1975–1989, 2020.

[23] H. Stuckey, A. Al-Radaideh, L. Sun, and W. Tang, "A spatial localization and attitude estimation system for unmanned aerial vehicles using a single dynamic vision sensor," *IEEE Sens. J.*, vol. 22, no. 15, pp. 15497–15507, 2022.

[24] M. Tao, Q. Chen, X. He, and S. Xie, "Fixed-time filtered adaptive parameter estimation and attitude control for quadrotor UAVs," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 5, pp. 4135–4146, 2022.

[25] S. Kim, V. Tadiparthi, and R. Bhattacharya, "Computationally efficient attitude estimation with extended h 2 filtering," *J. Guid. Control Dyn.*, vol. 44, no. 2, pp. 418–427, 2021.

[26] A. Kaba, "Unscented Kalman filter based attitude estimation of a quadrotor," *J. Aeronaut. Space Technol.*, vol. 14, no. 1, pp. 79–88, 2021.

[27] A. A. Rafique, A. Jalal, and K. Kim, "Statistical multi-objects segmentation for indoor/outdoor scene detection and classification via depth images," *Proceedings of the 2020 17th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, Islamabad, Pakistan, IEEE, Piscataway, NJ, USA, 2020, pp. 271–276.