Power System Protection and Control
ISSN:1674-3415

AN IMPROVED DEEP NETWORK METHOD TO INCREASE THE EFFICIENCY OF UAV REMOTE SENSING IMAGE REGISTRATION

Siavash Sattari[1], S. Hadi Yaghoubyan*[2], Karamollah Bagheri Fard[3], Razieh Malek Hosseini[4]

[1]Department of Computer Engineering, Yas.C., Islamic Azad University, Yasuj, Iran.
siavash.sattari@iau.ir
[2]*Department of Computer Engineering, Yas.C., Islamic Azad University, Yasuj, Iran. (Corresponding Author)
Yaghoobian.h@iau.ac.ir
[3]Department of Computer Engineering, Yas.C., Islamic Azad University, Yasuj, Iran.
ka.bagherifard@iau.ac.ir
[4]Department of Computer Engineering, Yas.C., Islamic Azad University, Yasuj, Iran.
malekhoseini.r@iau.ac.ir

Received: 19/05/2026 **Revised**: 21/06/2026 **Accepted**: 27/07/2026

***ABSTRACT*:**

Accurate registration of remote sensing images acquired by UAVs is one of the essential steps in applications such as aerial mapping, photogrammetry, 3D reconstruction, environmental change monitoring, and smart agriculture. Despite the advancement of deep learning-based methods, many existing algorithms still face reduced accuracy and stability when faced with changes in viewing angle, scale differences, lighting changes, and diversity of imaging scenes. Therefore, the aim of this research is to present an improved method based on deep networks to increase the efficiency of registering UAV remote sensing images. In this research, a ResNet-50 network-based architecture was designed that improves the process of extracting and matching corresponding features by utilizing the channel and spatial attention module (CBAM), multi-scale feature extraction, and the RANSAC algorithm. To evaluate the model, the UAVPairs dataset, which consists of 21,622 high-resolution images from 30 diverse scenes with significant differences in viewing angle, scale, and imaging conditions, was used. Before training the model, the images were preprocessed, including resizing, normalization, data augmentation, and image pair preparation, and then the model was trained using the appropriate cost function and Adam optimizer. The model performance was evaluated based on the criteria of Registration Accuracy, Matching Accuracy, RMSE, SSIM, and processing time, and was compared with ResNet-50 and SIFT methods. The results showed that the proposed method performed better than the comparative methods by achieving 98.1% image registration accuracy, 97.4% matching accuracy, 1.18 pixels for RMSE, and 0.986 for SSIM index. Also, the model training process showed that the proposed architecture converged faster and was more stable in the face of changes in the scene and imaging conditions. Overall, the results of this study show that the combination of the improved ResNet-50 network, the CBAM attention module, and the RANSAC algorithm can significantly increase the accuracy, stability, and generalizability of the drone image registration process and can be used as an effective solution in remote sensing, photogrammetry, 3D reconstruction, and intelligent environmental monitoring systems.

***Keywords**: Remote Sensing Image Registration, Drone, Deep Learning, Resnet-50, Cbam, Feature Matching, Uavpairs, Image Processing, Photogrammetry*

INTRODUCTION

Today, with the expansion of drone technology, the use of remote sensing images in areas such as precision agriculture, mapping, crisis management, environmental monitoring, urban planning, and infrastructure monitoring has grown significantly. Drones have become one of the most important tools for generating spatial information due to their low cost, high flexibility, ability to take images at different altitudes, and the ability to collect data with very high spatial resolution[1]. However, images captured by these systems are usually affected by factors such as changing viewing angles, flight altitude differences, drone movement, camera shake, changing lighting conditions, and environmental changes, which cause differences between images. This makes the direct use of images difficult in many practical applications and highlights the need for image registration operations [2].

Image registration is one of the most important preprocessing steps in remote sensing systems, which aims to align two or more images of an area so that corresponding points in different images coincide with each other. The quality of this process has a direct impact on the performance of many subsequent applications, including image mosaic generation, 3D modeling, change detection, automatic navigation, vegetation monitoring, and spatial information extraction. Conventional image registration methods, which are mainly developed based on local feature extraction such as SIFT, SURF and ORB, although they have acceptable performance in many conditions, they face a decrease in accuracy and stability when faced with severe changes in scale, rotation, lighting changes, geometric distortions and differences caused by drone imaging conditions. Hence, the need for methods with the ability to learn complex features and resistant to these changes is felt more than ever before [3]. Recent advances in the field of artificial intelligence, especially deep learning, have opened new horizons in remote sensing image processing. Deep neural networks are able to automatically extract complex patterns in images and identify relationships between corresponding points with greater accuracy without the need for manual feature design. The use of architectures such as convolutional networks, Siamese networks and models based on attention mechanisms has led to a significant increase in accuracy in tasks such as object recognition, image segmentation and image registration. However, many existing models still face problems such as high computational burden, need for extensive training data, sensitivity to severe environmental changes, and slow processing speed in real-time applications, which hinder their use**.** Nationally, they are associated with some limitations [4].

In recent years, researchers have tried to improve the performance of image registration algorithms by providing improved architectures, combining multiple deep networks, using attention mechanisms, extracting multi-scale features, and utilizing modern optimization methods. In addition to increasing the accuracy of identifying corresponding points, these approaches have reduced alignment errors, increased resistance to noise, lighting changes, and geometric changes [5]. However, the high complexity of UAV remote sensing data and the diversity of imaging conditions have made achieving a model with high accuracy, appropriate speed, and desirable generalization capability still a major challenge in this field, and extensive research in this field continues [6].

Considering the importance of accurate image registration in various remote sensing applications and the limitations of existing methods, the present study aims to present an improved method based on deep networks to increase the efficiency of remote sensing image registration in UAVs. In this study, an attempt will be made to design an efficient architecture for extracting effective features and accurate image matching, while increasing the registration accuracy, processing speed, and model stability. It is expected that the results of this study can provide a basis for improving UAV image processing systems and play an effective role in increasing the quality of remote sensing data analysis in applications such as smart agriculture, disaster management, precise mapping, environmental change monitoring, and smart city development.

RESEARCH BACKGROUND

Lu et al. [7] proposed an algorithm for automatic UAV remote sensing image registration based on deep residual features. They adopted deep residual neural network features extracted by a deep neural network, ResNet 50. Then, a tensor product is used to construct feature description vectors through high-level abstract features. Finally, the progressive adaptation algorithm (PROSAC) was used to remove false matches and fit a geometric transformation model to improve the registration accuracy. Experimental results for various typical scene images with different resolutions acquired by different UAV image sensors show that the improved algorithm can achieve higher registration accuracy than the advanced deep learning registration algorithm and other popular registration algorithms. In a study, Ding et al. [8] proposed a new pipeline for wild medicinal plant recognition and performance evaluation based on UAV and DL. They used a UAV to collect panoramic images of Lamioplomis rotata Kudo (LR) in high-altitude areas. Then, these images were annotated and cropped into equal-sized subimages, and the DL Mask R CNN model was used for object recognition and LR segmentation. Finally, based on the segmentation results, they accurately counted the number and efficiency of LRs. The results showed that the Mask R CNN model based on the ResNet 101 core network outperformed ResNet 50 in all

DOI: 11.46121/pspc.54.3.36

evaluation indicators. The average LR recognition accuracy of the Mask R CNN based on the ResNet 101 core network was about 89%, while that of ResNet 50 was about 88%.

Xiang et al. [9] introduced a self-attention-based transformer in a study that obtained long-term feature dependencies of remote sensing images as a complement to local details for accurate crop type segmentation in UAV remote sensing images, and proposed an end-to-end CNN feature transformer fused network. Finally, the FPNHead feature pyramid network acted as a decoder for improved compatibility with multi-scale fused features and output crop type segmentation results. Our comprehensive experiments showed that the proposed CTFuseNet has higher crop type segmentation accuracy, with an average intersection on alliance of 85.33% and pixel accuracy of 92.46% on the benchmark remote sensing dataset, and outperforms state-of-the-art networks, including U Net, PSPNet, DeepLabV3+, DANet, OCRNet, SETR, and SegFormer. Liu et al. [10] used drone images, a mobile phone, and a corn tassel counting (MTC) dataset to test the performance of a faster region-based convolutional neural network (Faster R CNN) with a residual neural network (ResNet) and a visual geometry ensemble neural network. The results showed that ResNet, as a feature extraction network, outperformed VGGNet for corn tassel detection from drone images with a resolution of 600 × 600. The prediction accuracy ranged from 87% to 95%. However, the prediction accuracy was less than 87.27% for drone images with a resolution of 5280 × 2970. They changed the anchor size to [852, 1282, 2562] in the proposed region grid according to the width and height of the pixel distribution to improve the detection accuracy to 90%. This accuracy reached 95.95% for mobile phone images. Then, they compared their trained model with TasselNet without training on their dataset. The average difference in the number of tassels between the calculations with 40 images for the two methods was 1.4. Natsan et al [11] presented a new method for classifying forest tree species through high-resolution RGB images obtained with a simple consumer grade camera mounted on a UAV platform using residual neural networks. They conducted two experiments, the first with images acquired every three years and the second with images acquired only one year. In the first experiment, when the trained network was tested on a set of images, we obtained 80% classification accuracy, and in the second experiment, we obtained 51% classification accuracy. As part of this work, a new dataset of high-resolution tree species will be generated that can be used for further studies related to deep neural networks in forestry.

Wu et al.[12] in a research entitled "Reconstruction of UAV remote sensing images using enhanced height representation and frequency adaptive sampling" investigated the improvement of 3D reconstruction methods based on Neural Radiant Field (NeRF) in UAV remote sensing images. The researchers presented a new framework called HF-NeRF, which included the Height Enhanced Representation (HER) module to extract the geometric features of different height layers and the Frequency Adaptive Sampling (FAS) method to adjust the sampling density based on the energy distribution in the frequency domain. The results of the experiments on the dedicated dataset of UAV images and several public datasets showed that the proposed method, compared to the advanced models derived from NeRF, had a better performance in the reconstruction of remote sensing images and was able to significantly increase the computational efficiency and the quality of 3D reconstruction while maintaining fine details.

Zhou et al. [13] in a research entitled "Efficient linking of UAV high-resolution images using dense kernel deep features" presented an efficient method for linking high-resolution UAV images. In order to solve these problems, they presented a method based on dense extraction of deep nuclear features and optimization of geometric constraints. In this method, a deep learning-based feature matching framework was used to achieve dense matching at the subpixel level, which was able to overcome the limitations of traditional methods such as SIFT in high-resolution images. Also, a two-layer strategy of filtering mismatches based on geometrical constraints was designed, which improved the accuracy of image alignment in conditions of high parallax and low texture areas. Finally, by using the combined strategy including single response transformation and combination of pixels with maximum intensity, the process of panoramic connection was done. The results of experiments on several data sets showed that the proposed method, while maintaining the same visual quality based on PSNR, SSIM and LPIPS criteria, reduced the image connection time to about 17.5% of the basic method and provided high efficiency in fast connection of large drone images.

RESEARCH METHOD

In order to increase the accuracy of feature extraction and focus the network on areas with effective information, the convolutional attention module has been used in the proposed architecture. This module is placed after each remaining block in the ResNet network and consists of two sections: Channel Attention and Spatial Attention. In the channel attention stage, the importance of each feature channel is calculated based on the information in the feature maps, and channels with rich information are enhanced. Then, in the stage of spatial attention, the important areas of the image that have a higher probability of having corresponding points are given more weight. This process reduces the effect of less important areas such as shadows, uniform overlays or untextured parts and enables the network to extract more stable and robust features. Using this module in UAV images that usually have extreme changes in viewing angles, brightness differences and scale changes increases the accuracy of identifying common features and reduces image registration errors.

After enhancing the features, the resulting mappings from the reference and target images enter the feature matching stage. In this step, for each feature vector in the reference image, the closest corresponding vector in the target image is searched based on the cosine similarity measure and the Euclidean distance. In order to avoid making wrong matches, the Nearest Neighbor Distance Ratio criterion is also used to select only the points with sufficient confidence. This process removes a significant portion of false matches and produces a set of initial matching points. The use of deep features extracted by the network makes the matchings more robust against geometric changes, illumination changes, scale differences, and minor occlusions than classical methods based on SIFT or SURF.

Although the feature matching step extracts a large number of correct matching points, there is still the possibility of outliers in the matching set. For this reason, the RANSAC algorithm has been used in this research to remove false matches and accurately estimate the geometric transformation between images. In each iteration, a number of corresponding points are randomly selected and the homography matrix is estimated. Then, the compatibility of other points with this model is calculated and the model with the highest number of compatible points is selected as the final answer. The use of RANSAC increases the stability of the algorithm against noise, matching errors, and complex image changes, and enables accurate alignment of images. Finally, by applying the transformation matrix on the target image, the registered image is produced and The following analyses will be performed.

In order to train the proposed model, the supervised learning method was used. The images and geometric labels in the UAVPairs dataset were divided into three parts: training, validation, and testing, so that 70% of the data was considered for training, 15% for validation, and 15% for testing. This division allows the model's performance to be evaluated on data that was not observed in the training process. The network training was performed using the AdamW optimizer and the initial learning rate was set to 0.0001. The training batch size was set to 16 and the number of training iterations (Epoch) was set to 150. Also, the Early Stopping technique was used to prevent overfitting and the Learning Rate Scheduler was used to increase the stability of the training process. All experiments were implemented in the Python 3.11 environment using the PyTorch library.

To evaluate the performance of the proposed algorithm, a set of quantitative and qualitative metrics were used. The most important metrics included Matching Accuracy, Registration Error, Root Mean Square Error (RMSE), Structural Similarity Index (SSIM), and processing time for each pair of images. Matching accuracy indicates the ratio of correctly matched points to the total number of extracted points, while RMSE measures the geometric difference between the matched points after image registration. The SSIM index also evaluates the structural similarity of the registered image with the reference image and is used to examine the final quality of image registration. In addition, the execution time of the algorithm was also measured to allow for the evaluation of the computational efficiency of the proposed method. Using these metrics simultaneously allows the model performance to be comprehensively examined from various aspects including accuracy, quality, and processing speed.

In summary, the proposed framework of this study combines intelligent image preprocessing, improved ResNet50 network, CBAM attention module, multi-scale feature extraction, deep feature matching, and accurate geometric transformation estimation using the RANSAC algorithm to provide a comprehensive system for registering UAV remote sensing images. Using the large and diverse UAVPairs dataset, the model is trained in different imaging conditions, including changes in viewing angle, scale difference, illumination changes, and land cover variation, and has high generalization capabilities. It is expected that this architecture can provide more accuracy in identifying corresponding points, reduce registration error, and increase image alignment quality compared to conventional methods, and can be used as an efficient solution in applications such as aerial mapping, photogrammetry, environmental monitoring, precision agriculture, disaster management, and 3D reconstruction.

PROPOSED MODEL ARCHITECTURE

The proposed architecture of this study is designed to increase the accuracy of registering remote sensing images obtained by UAVs. In this architecture, an attempt has been made to increase the network's ability to extract stable and robust features against scale changes, rotation, viewing angle differences, and brightness changes while preserving the spatial and textural features of the images. As can be seen in Figure (1), the proposed architecture consists of eight main parts including image input, preprocessing, feature extraction, feature enhancement, feature matching, outlier removal, geometric transformation estimation, and registered image generation. The data flow continues in stages from the beginning of the network to the production of the final image, and the output of each stage is used as the input of the next stage.

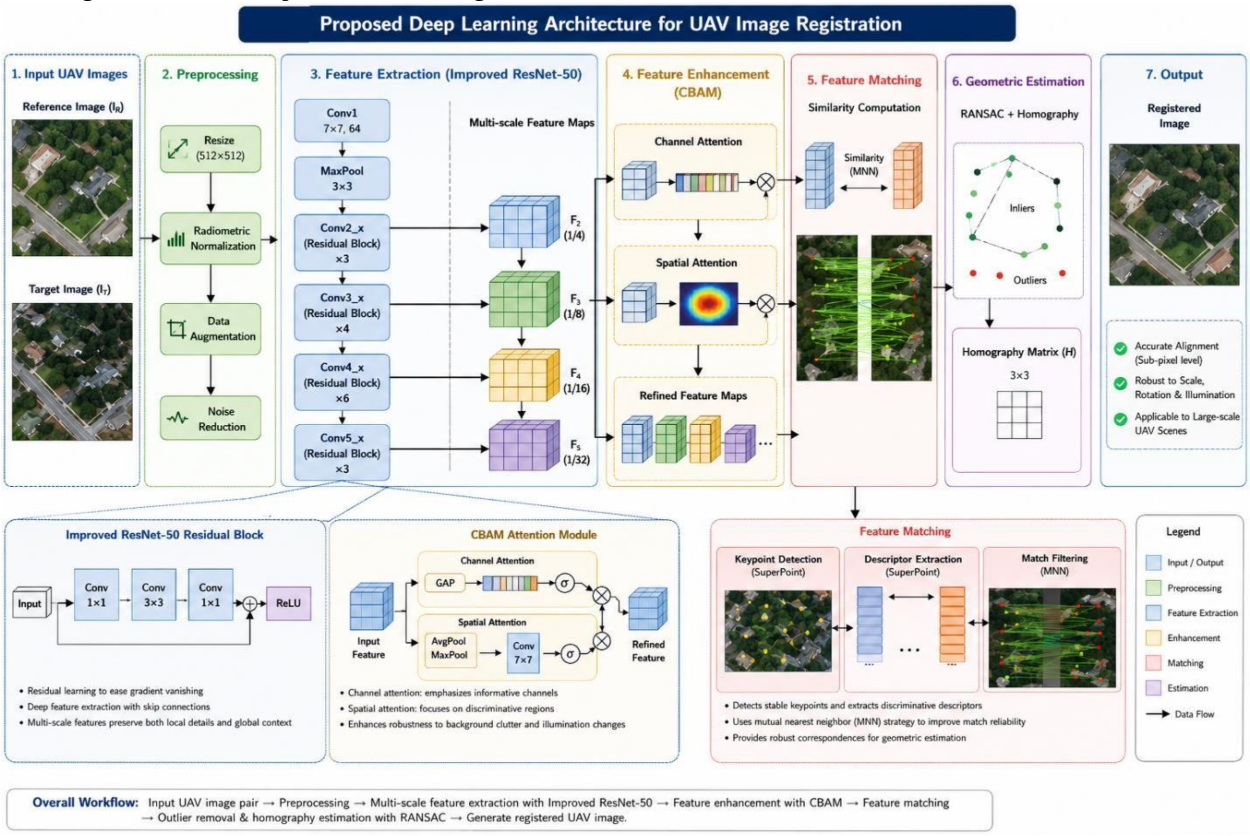

Figure (1): General architecture of the model

In the first stage, two remote sensing images taken by a drone at different times or angles are entered as input to the model. These images may have differences in viewing angle, flight height, lighting conditions, and scale; therefore, before entering the network, preprocessing operations are performed, including resizing the images to fixed dimensions, normalizing the brightness values, increasing the data, and removing noise. The goal of this stage is to create uniformity among the input data and reduce the impact of environmental factors on the network learning process.

After preparing the images, the data is entered into the improved ResNet-50 network. This network, as the backbone of the architecture, is responsible for extracting the initial features. Using the remaining blocks allows information to be transferred between deep layers without loss of quality and the problem of gradient

DOI: 11.46121/pspc.54.3.36

blurring is largely resolved. In addition, feature extraction is performed at several different levels to provide the network with both local details and overall image structures simultaneously.

In the next step, the extracted features are fed into the CBAM attention module. This module consists of two parts, channel attention and spatial attention, and it amplifies important image information by assigning different weights to features. As a result, areas with rich texture, edges, corners, and other prominent points that are more important for capturing images are extracted with higher accuracy, while less important or noisy information is attenuated.

After feature enhancement, the feature maps of the two images are fed into the Feature Matching section. In this section, the corresponding feature vectors in the two images are compared using similarity measures and the corresponding points are extracted. Then, the RANSAC algorithm is run to remove false matches. This algorithm allows for accurate calculation of the homography matrix by removing outliers and selecting the best set of compatible points. Finally, using the homography matrix, the target image is transferred to the reference image coordinate system and the registered image is produced.
In summary, the proposed architecture, combining the improved ResNet-50 network, the CBAM attention module, and the RANSAC algorithm, is able to extract robust and stable features from drone images and perform the image registration process with high accuracy. The modular design of this architecture has resulted in not only increasing the registration accuracy, but also controlling the computational complexity and enabling its use in drone image processing systems in near real-time.

RESEARCH FINDINGS

In order to examine the efficiency of the proposed method, its performance was compared with two reference methods including the ResNet-50 network and the classic SIFT algorithm. All experiments were performed on the UAVPairs dataset and criteria such as Registration Accuracy, Matching Accuracy, Root Mean Square Error (RMSE), Structural Similarity Index (SSIM), and processing time of each image pair were evaluated. The results are presented in Table (1).

Table (1): Comparison of the performance of different methods in recording drone images

Processing time (seconds)	SSIM	RMSE (Pixel)	Matching)%(accuracy	Recording)%(accuracy	Model
0/42	0/986	1/18	97/4	98/1	Proposed Method
0/46	0/971	1/84	94/8	95/6	ResNet-50
0/91	0/938	3/41	83/2	90/8	SIFT

As can be seen in Table (1), the proposed method has performed better than the two comparative methods in all evaluation indicators. The highest image registration accuracy is related to the proposed method, which has achieved a value of 98.1%, while this value has been reported for the ResNet-50 network as 95.6% and for the SIFT algorithm as 90.8%. This difference shows that the proposed architecture has been able to extract more stable and accurate features from drone images and, as a result, perform the image alignment process with more accuracy.

The Matching Accuracy index also shows similar results. The proposed model has achieved the highest correct matching rate of corresponding points, achieving a value of 97.4%. This indicates that the combination of the improved ResNet-50 network with the CBAM attention module has increased the resolution of features and reduced false matching. In contrast, the SIFT algorithm has shown poorer performance in the face of changes in viewing angle, scale, and lighting conditions due to its reliance on artificial features.

On the other hand, the RMSE value in the proposed method was obtained as 1.18 pixels, which is the lowest error rate among the methods studied. The decrease in this index indicates that the extracted corresponding points have high geometric accuracy and the calculated transformation between images has been

Power System Protection and Control
ISSN:1674-3415

performed with very little error. In contrast, the increase in the RMSE value in the two reference methods indicates that the alignment of the images in them has been performed with less accuracy and a greater spatial difference between the corresponding points remains.

The results related to the SSIM index also confirm the superiority of the proposed method. The value of 0.986 for this index indicates that the structure, texture, and details of the image after registration are preserved almost unchanged compared to the reference image. While this value for ResNet-50 is 0.971 and for SIFT is 0.938, which indicates a relative decrease in the quality of the recorded image in these two methods.

In addition to accuracy, the speed of the algorithm execution is also of great importance. The results showed that the proposed method processed each pair of images in 0.42 seconds on average. Although the difference in execution time with ResNet-50 is not very large, compared to the SIFT algorithm, the processing time was reduced by almost half. This shows that the proposed architecture, in addition to increasing accuracy, also has good computational efficiency and can be used in applications that require fast image processing.

In general, the results of this section show that the proposed method has been able to establish a good balance between accuracy, recording quality, and processing speed. The simultaneous improvement in all evaluation indicators indicates that the use of the improved ResNet-50 network, the CBAM attention module, and the accurate outlier removal process using the RANSAC algorithm have played an effective role in increasing the quality of drone image recording. These results indicate that the proposed model can be used as an efficient method for applications such as aerial mapping, 3D reconstruction, environmental change monitoring, and navigation systems based on drone images.

2-5-**Accuracy Change Chart**

The following is the accuracy chart of the model in different iterations:

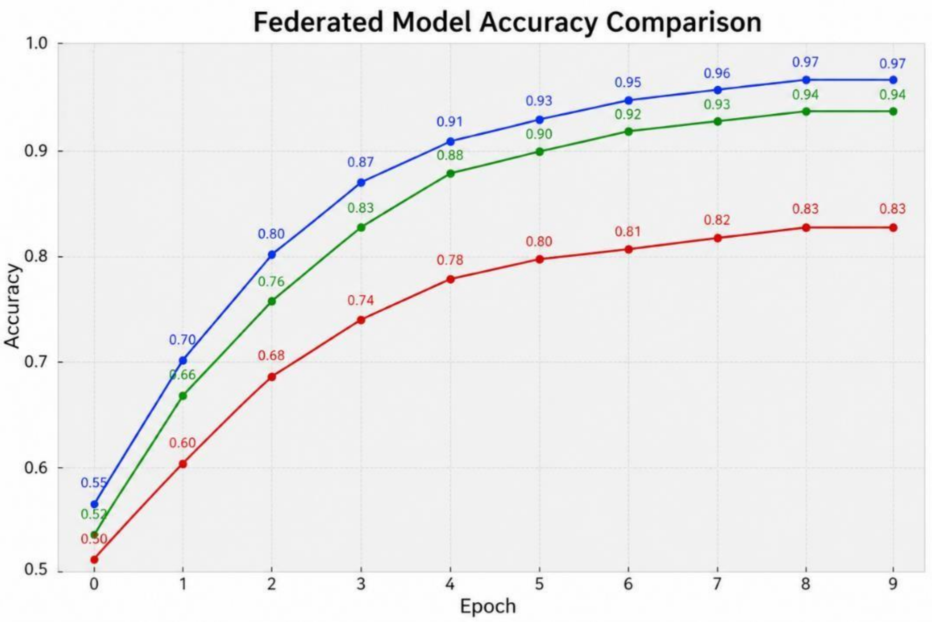

Figure (2): Diagram of model accuracy changes

Figure (2) shows the trend of changes in the accuracy of the three investigated methods including the proposed method, ResNet-50 and SIFT algorithm during 10 training periods (Epoch). As can be seen, all three methods have experienced an upward trend in the early stages of training, but the growth rate of accuracy and convergence speed are different. The proposed model has shown better performance than the other two methods from the very beginning of the training process, and this superiority has been maintained until the end of the training. This shows that the presented architecture has been able to extract the effective features of drone images more accurately and perform the learning process faster.

Examining the change trend of the blue curve, which is related to the proposed method, shows that the accuracy of the model has increased from 55% at the beginning of the training to 97% at the end of the training process. The highest rate of growth in accuracy is observed in the first four periods; So that the accuracy value has reached from 0.55 to 0.91. After that, the slope of the curve gradually decreases and from the sixth period onwards, the model approaches the stable state. The stabilization of the accuracy value at around 97% indicates that the model has achieved a suitable convergence and has successfully completed the learning process without extreme fluctuations. This behavior shows the stability of the proposed architecture and its ability to extract features resistant to changes in the viewing angle, scale and imaging conditions in UAV images.

The green curve related to the ResNet-50 network also shows a suitable increasing trend, but it is lower than the proposed method in all stages of training. The accuracy of this model started from 52% at the beginning of training and reached 94% at the end. Although ResNet-50 has been able to achieve a good performance, its difference of about 3% with the proposed method shows that the use of improvement modules such as CBAM and extraction of multi-scale features has played an effective role in increasing the accuracy of image registration. It can also be seen that the learning process of ResNet-50 was done at a slower speed than the proposed method and the model needed more training sessions to reach a stable state.

On the contrary, the red curve corresponding to the SIFT algorithm shows the lowest performance among the investigated methods. The accuracy value of this method started from 50% at the beginning of the training and at the end it was stabilized at around 83%. The limited growth of accuracy and its significant distance with two methods based on deep learning show that classical feature extraction methods, in the face of complex changes in UAV images, including the difference in viewing angle, scale change, illumination changes and the complexity of natural textures, have less ability to extract stable corresponding points. In general, the results of Figure (2) show that the proposed method not only converged faster than other methods, but also achieved the highest level of accuracy at the end of training. These results indicate that the proposed architecture is able to perform the process of recording UAV remote sensing images with greater accuracy and stability and has a higher .generalization capability than the reference methods

Figure (3): Examples of pairs of drone images used in the UAVPairs dataset

Figure (3) shows an example of the images used in this study from the UAVPairs dataset. As can be seen, each column contains a corresponding pair of images of a single scene that were captured by the drone under different imaging conditions. The images include diverse environments such as agricultural lands, urban areas, residential spaces, vegetation, and communication routes that differ significantly in terms of viewing angle, scale, texture, and lighting conditions. This diversity causes the image capture process to face challenges such as geometric changes, scale differences, and brightness changes.

In the proposed method, each pair of images first enters the preprocessing stage and is prepared for feature extraction after performing normalization, resizing, and data augmentation operations. Then,

 DOI: 11.46121/pspc.54.3.36

the deep features of the images are extracted by the improved ResNet-50 network and important information including edges, corners, texture features and geometric structures are enhanced using the CBAM attention module. Next, the features extracted from the two images are compared with each other and the corresponding points are identified through the feature matching module. In the final stage, the RANSAC algorithm is used to remove the mismatches and calculate the homography matrix to match the target image with the reference image with the highest accuracy. As is clear from the examples presented in the figure, the images have significant diversity in terms of land cover, feature density, scale changes and shooting angle. Such features make the UAVPairs dataset a suitable platform for evaluating UAV image registration algorithms. The favorable performance of the proposed method in the face of these diverse conditions shows that the presented architecture has a high ability to extract stable features, reduce matching error and increase the accuracy of image registration and can be used in applications such as aerial mapping, photogrammetry, 3D reconstruction and environmental change monitoring.

CONCLUSION

The present research aimed to present an improved method based on deep networks to increase the efficiency of registering remote sensing images obtained by UAVs. Considering the increasing use of UAV systems in areas such as mapping, natural resource monitoring, precision agriculture, disaster management, urban planning and 3D reconstruction, achieving an accurate and stable method for image registration is of particular importance. In this research, an attempt was made to reduce the limitations of classical methods in the face of changes in scale, viewing angle, lighting conditions and environmental complexities by utilizing deep learning capabilities and designing an improved architecture. Accordingly, the proposed architecture was designed and evaluated with a focus on deep feature extraction, effective information enhancement, and false matching removal.
To evaluate the performance of the proposed method, the UAVPairs dataset was used, which is one of the new and comprehensive datasets in the field of drone imagery. This dataset, with 21,622 high-resolution images from 30 different scenes, covers a variety of conditions, including urban areas, agricultural lands, mountainous areas, river routes, building complexes, and mixed-cover environments. Also, the existence of significant differences in the angle of view, imaging scale, and environmental conditions allows for a realistic evaluation of the performance of image registration algorithms. Using such a dataset allowed the performance of the proposed model to be examined not only in controlled laboratory conditions, but also to evaluate its ability to deal with real and complex data.

The proposed architecture of this research was developed based on ResNet-50 network and by using CBAM attention module, multi-scale feature extraction and RANSAC algorithm to remove incorrect matching, the image registration process was improved. Using residual connections in the network, while reducing the problem of gradient fading, enabled the learning of deeper features. On the other hand, the attention module allowed the network to focus more on areas with effective information such as edges, corners and index structures, and as a result, the extracted features were more stable against environmental changes. The combination of these components allowed the proposed model to perform the feature matching process more accurately and minimize errors caused by incorrect matching. The results of the model evaluation showed that the proposed method had a favorable performance in all the indicators studied. Image registration accuracy, matching accuracy of corresponding points, structural similarity index and geometric error rate all indicated the superiority of the model over the comparative methods. Also, the study of the model training process showed that the proposed architecture converged faster than the basic ResNet-50 model and maintained higher accuracy at the end of training. This indicates that the design of additional modules and the improvement of the feature extraction process played an effective role in increasing the quality of network learning and improving its final performance.

Comparison of the results with the classic SIFT algorithm also showed that methods based on deep learning, especially in drone images with large changes in viewing angle, scale difference and texture complexity, have a significant performance compared to methods based on hand-made features. While the SIFT algorithm faced a decrease in matching accuracy and an increase in registration error in many situations, the proposed model was able to significantly increase the quality of image registration by extracting semantic and robust features. This shows that the use of deep architectures can solve many problems. overcome the limitations of traditional methods and improve the accuracy of image registration in practical applications.

Power System Protection and Control
ISSN:1674-3415

Another notable result of this research was the balance between accuracy and computational efficiency. Although the use of deep networks usually increases the computational cost, the proposed architecture was able to achieve high accuracy while also providing reasonable processing time. This feature makes the model usable in UAV image processing systems with large volumes of data. In addition, the stability of the model's performance in dealing with the diverse scenes in the UAVPairs dataset shows that the proposed method has good generalization capability and can provide reliable results in different operational environments.

Despite the favorable results, this research also had some limitations. The model was evaluated only on UAV optical images and the effect of using multispectral data, thermal images, or radar images was not investigated in this study. Also, although the UAVPairs dataset is quite diverse, using a larger and more diverse dataset would be effective in more comprehensively assessing the generalizability of the model. In addition, examining the performance of the model under different weather conditions, extreme lighting changes, or images acquired at very different altitudes can provide a more accurate view of the strengths and weaknesses of the proposed architecture.

Overall, the findings of this study indicate that the proposed architecture can be used as an effective solution for accurate registration of UAV remote sensing images. The combination of the improved ResNet-50 network, the CBAM attention module, and the process of removing false matches using the RANSAC algorithm has increased the accuracy of feature extraction, improved the quality of matching of corresponding points, reduced the registration error, and increased the stability of the model under different imaging conditions. These achievements show that the proposed method, in addition to having research value, has a suitable capacity for application in aerial mapping systems, photogrammetry, 3D reconstruction, UAV navigation, environmental change monitoring, natural resource management, and development of intelligent remote sensing systems. Also, the development of this architecture using new deep learning models, advanced attention mechanisms, transformer-based models, and self-supervised learning methods can provide a suitable path for future research in the field of remote sensing image recording and UAV data processing.

REFERENCE

[1]. Xu, C., Lu, S., Feng, X., Zhang, X., & Wang, J. (2025). Crop classification with UAV multispectral remote sensing, employing an enhanced ResNet50 residual network. PLoS One, 20(12), 335361.

[2]. Zhang, S., Feng, H., Han, S., Shi, Z., Xu, H., Liu, Y., ... & Yue, J. (2022). Monitoring of soybean maturity using UAV remote sensing and deep learning. Agriculture, 13(1), 110-119.

[3]. Feng, B., Liu, Y., Chi, H., & Chen, X. (2023). Hyperspectral remote sensing image classification based on residual generative adversarial neural networks. Signal Processing, 21(3), 109-202.

[4]. Huang, H., Li, F., Fan, P., Chen, M., Yang, X., Lu, M., ... & Zhu, P. (2023). Amdnet: A modern uav rgb remote-sensing tree species image segmentation model based on dual-attention residual and structure re-parameterization. Forests, 14(3), 549.-571.

[5]. Pajany, M., Venkatraman, S., Sakthi, U., Sujatha, M., & Ishak, M. K. (2024). Optimal fuzzy deep neural networks-based plant disease detection and classification on UAV-based remote sensed data. IEEE Access, 12(7), 162131-162144.

[6]. Zhang, Z., Huang, L., Wang, Q., Jiang, L., Qi, Y., Wang, S., ... & Gu, Y. (2024). UAV hyperspectral remote sensing image classification: A systematic review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18(3), 3099-3124.

[7]. Luo X, Lai G, Wang X, Jin Y, He X, Xu W, Hou W.(2021). UAV Remote Sensing Image Automatic Registration Based on Deep Residual Features. Remote Sensing. 13(18):3605-3612.

[8]. Ding R, Luo J, Wang C, Yu L, Yang J, Wang M, Zhong S, Gu R.(2023). Identifying and mapping individual medicinal plant Lamiophlomis rotata at high elevations by using unmanned aerial vehicles and deep learning. Plant Methods. 19(1):1-6.

[9]. Xiang J, Liu J, Chen D, Xiong Q, Deng C.(2023). CTFuseNet: A Multi-Scale CNN-Transformer Feature Fused Network for Crop Type Segmentation on UAV Remote Sensing Imagery. Remote Sensing. 15(4):11-51.

[10]. Liu Y, Cen C, Che Y, Ke R, Ma Y, Ma Y.(2024). Detection of maize tassels from UAV RGB imagery with faster R-CNN. Remote Sensing.;12(2):338.

[11]. Natesan S, Armenakis C, Vepakomma U.(2025). RESNET-BASED TREE SPECIES CLASSIFICATION USING UAV IMAGES. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences. 14(3),56-71.

[12]. Wu, S., Chen, X., Ye, C., Chen, Z., & Wu, X. (2026). HF-NeRF: UAV remote sensing image reconstruction via height-augmented representation and frequency-adaptive sampling. Scientific Reports.

[13]. Zhou, J., Wei, Z., Zhong, Y., & He, X. (2026). Efficient UAV High-Resolution Image Stitching via Dense Deep Kernelized Feature. Sensors, 26(5), 1540.