IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 4, August, 2024, pp. 2115-2122
Journal homepage: http://iieta.org/journals/ts

Application of Deep Learning-Based Image Registration Techniques in Autonomous Robot Navigation

Xuhan Chen[1], Tianze Wang[2*], Chye En Un[3], Hongwu Qin[4]

[1] School of Information and Control, Ustinov Baltic State Technical University, St. Petersburg 190005, Russia
[2] School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China
[3] Higher School of Cybernetics and Digital Technologies, Pacific National University, Khabarovsk 680035, Russia
[4] College of Electronic and Information Engineering, Changchun University, Changchun 130022, China

Corresponding Author Email: wangtianze@mails.cust.edu.cn

https://doi.org/10.18280/ts.410439

Received: 3 April 2024
Revised: 10 July 2024
Accepted: 26 July 2024
Available online: 31 August 2024

Keywords:
autonomous robot navigation, image registration, deep learning, network-in-network, dual-attention mechanisms, feature matching, parameter regression network

ABSTRACT

Autonomous robot navigation is widely applied across various domains, with one of the core challenges being accurate image registration under varying time frames, perspectives, and complex environmental changes. Existing image registration methods address some of these challenges but still face significant limitations, such as insufficient model generalization and low computational efficiency when dealing with highly dynamic and irregular environmental changes. To enhance the accuracy, robustness, and real-time performance of image registration, this paper proposes a deep learning-based image registration technique. The approach comprises three key components: model hypothesis, dataset generation, and overall network design. Through innovative designs such as network-in-network, dual-attention mechanisms, a bidirectional correlation operation in the feature matching layer, and a parameter regression network, this study aims to provide more reliable visual support for autonomous robot navigation.

1. INTRODUCTION

Autonomous robot navigation has made significant progress in recent years, with widespread applications in industrial, agricultural, medical, and service fields [1-3]. As an important automation technology, autonomous robots can perform complex tasks in unknown or dynamic environments, achieving efficient and accurate autonomous movement [4, 5]. However, one of the core challenges in realizing autonomous navigation is how to accurately perform image registration under different times, different perspectives, and complex environmental changes, thus providing reliable visual information for the robot's path planning and decision-making [6]. The rapid development of deep learning technology provides new ideas and methods to solve this problem.

As a key aspect of autonomous robot navigation, image registration technology directly affects the robot's ability to perceive and understand the environment [7-9]. Using deep learning for image registration can not only significantly improve registration accuracy and robustness but also maintain high real-time performance in complex and dynamic environments [10, 11]. Therefore, research on deep learning-based image registration technology is not only of great significance for enhancing the performance of autonomous robot navigation but also will promote technological advancements in related fields and the application and popularization of intelligent robots in more scenarios.

Although existing image registration methods have addressed some issues in autonomous robot navigation to some extent, there are still many shortcomings. For example, traditional feature point-based methods are prone to registration failure or accuracy degradation when faced with complex backgrounds and large-scale perspective changes [12-15]. On the other hand, although image registration methods based on classical deep learning have improved performance to some extent, they still suffer from insufficient model generalization ability and low computational efficiency when dealing with highly dynamic and irregular environmental changes [16-21]. Therefore, it is urgent to develop more effective deep learning image registration techniques to overcome the limitations of the above methods.

This paper aims to study the application of deep learning-based image registration technology in autonomous robot navigation, focusing on three main parts. The first part is the model hypothesis, proposing and verifying deep learning model hypotheses suitable for image registration. The second part is dataset generation, constructing an image registration dataset suitable for autonomous robot navigation to provide data support for model training and evaluation. The last part is the overall network design, including the design and implementation of network-in-network, dual-attention mechanisms, feature matching layers using bidirectional correlation operations, and parameter regression networks. Through these innovative designs, this study aims to improve the accuracy, robustness, and real-time performance of image registration, thereby providing more reliable visual support for autonomous robot navigation, which has important theoretical significance and application value.

2. MODEL HYPOTHESIS

In the multi-view image registration task for autonomous robot navigation, this paper proposes a series of key hypotheses to effectively address diverse and complex environmental challenges. First, it is assumed that in most cases, the distance between the imaging device and the environment is sufficiently large so that the imaging area can be approximated as a plane. Second, it is assumed that the motion of the imaging device relative to the target area can be simplified to pure rotational motion around the target area. This means that during the imaging process, the translational motion of the device relative to the target is not considered, and the focus is on the impact of rotational motion on the change in perspective. This pure rotational motion assumption simplifies the multi-view image registration problem in autonomous robot navigation, allowing it to be described as a perspective transformation problem and solved using a homography matrix. The expression of the homography matrix is given by the following formula, where the parameters $g_{u,k}$ map the pixel coordinates (a,b) in the image to be registered during autonomous robot navigation to (a',b').

$$\begin{pmatrix} a' \\ b' \\ 1 \end{pmatrix} = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \begin{pmatrix} a \\ b \\ 1 \end{pmatrix} \tag{1}$$

In this paper, the 4-point parameter model G_{4PO} is used to describe the above formula, thereby enabling the neural network to optimize the output parameters. Let the coordinates of the corresponding four points of the two images be represented by (a_u,b_u) and (a'_u,b'_u), i.e., $\Delta a_u=a'_u-a_u$, $\Delta b_u=b'_u-b_u$. The model expression is shown in the following formula:

$$G_{4PO} = \begin{pmatrix} \Delta a_1 & \Delta b_1 \\ \Delta a_2 & \Delta b_2 \\ \Delta a_3 & \Delta b_3 \\ \Delta a_4 & \Delta b_4 \end{pmatrix} \tag{2}$$

3. DATASET GENERATION

In the multi-view image registration task for autonomous robot navigation, key steps in generating the training dataset include performing a series of random transformations on the original images to simulate image variations under different perspectives and generate image pairs required for the registration task. Specifically, the dataset generation principle is as follows: 1) A fixed-size square area is randomly selected from the original image U, denoted as O_{X1}, with its center coordinate as o. To simulate perspective changes, the four vertices of O_{X1} are randomly perturbed by a magnitude of $[-a, a]$, resulting in a new square area O_Y. By calculating the homography transformation matrix G_{XY} from the four vertices of O_{X1} to the four vertices of O_Y, the parameters describing this transformation can be obtained. Next, the inverse matrix of GXY is applied to the original image U. In the newly obtained image U', a square area of the same size is selected again, centered at point o, denoted as O_{X2}. Thus, O_{X1} and O_{X2} serve as the reference image and the image to be registered in the multi-view image dataset for autonomous robot navigation, with G_{XY} considered as the transformation parameter between the two images. 2) To further enrich the multi-view image dataset for autonomous robot navigation, several groups of sampling point coordinates are randomly and uniformly selected in the image to be registered, O_{X2}. The transformation matrix G_{XY} is applied to these sampling points to obtain the corresponding new sampling point coordinates. Therefore, the new and old sampling point coordinates respectively represent the labels of the two images in the dataset. Furthermore, based on the squared distance of the sampling points, a loss function for the model is constructed. Let V be the number of sampling points, the function representing the distance between the two groups of sampling points is denoted as f, the sampling points are denoted as T, the parameters predicted by the model are denoted as ϕ', the real parameters in the registration dataset are denoted as ϕ_{HS}, the corresponding real transformation in the registration dataset is denoted as $S_{\phi HS}$, and the transformation predicted by the model is denoted as S_{ϕ}, then the function expression is:

$$M(\varphi', \varphi_{HS}) = \frac{1}{V} \sum_{u=1}^{V} f\left(S_{\varphi}(T_u), S_{\varphi_{HS}}(T_u)\right)^2 \tag{3}$$

4. OVERALL NETWORK DESIGN

To construct a multi-view image registration model for autonomous robot navigation, this paper designs a deep learning network architecture consisting of three parts: feature extraction network, feature matching network, and parameter regression network. The structure diagram is shown in Figure 1. Each part is designed to address specific registration task requirements, ensuring that the model can accurately perform image registration under different perspectives. The feature extraction network uses a weight-sharing dual-channel network-in-network to extract features from the reference image and the image to be registered, respectively. This network leverages the mechanism of weight sharing to ensure consistency and efficiency in the features extracted from different channels, thereby enhancing the ability to extract high-level semantic features from the images. To further enhance the effectiveness of feature extraction, a dual-attention module combining channel and spatial attention is integrated into the network-in-network. This module improves the network's ability to distinguish and locate features by focusing on important feature channels and spatial positions, making the extracted features more refined and accurate. Next is the feature matching network part. By using bidirectional cross-correlation layers, this paper matches the feature maps between the reference image and the image to be registered. The bidirectional cross-correlation layers can capture similar features between images and match them accurately, thereby improving the performance of feature matching. This part of the design ensures the robustness and accuracy of the feature matching process under multi-view conditions. Finally, the matched images are input into the parameter regression network, which is responsible for predicting the final transformation parameters. Through the feature extraction and matching in the first two parts, the parameter regression network can learn the transformation relationship between the images from the matched feature maps and output eight homography transformation parameters. These parameters are used to describe the perspective transformation relationship between the reference image and the image to be registered, and the final image registration is performed using these

parameters, resulting in the registered image. Let the output of each network-in-network layer be denoted as L, the pixel coordinate index in the input feature map be denoted as i and n, the feature map value centered at point (i,n) be denoted as $a_{i,n}$, the channel index of the input feature map be denoted as j, the network index be denoted as v, the weight of the neural network be denoted as QZ, and the bias of the neural network be denoted as y. Then, the computation method of the network-in-network is as follows:

$$\begin{aligned} L_{i,n,j_1}^{1} &= \mathrm{ReLU}\left(BN\left(QZ_{j_1}^{1\,S} a_{i,n} + y_{j_1}\right)\right), \\ &\vdots \\ L_{i,n,jv}^{v} &= \mathrm{ReLU}\left(BN\left(QZ_{j_v}^{v\,S} d_{i,n}^{v-1} + y_{j_v}\right)\right) \end{aligned} \tag{4}$$

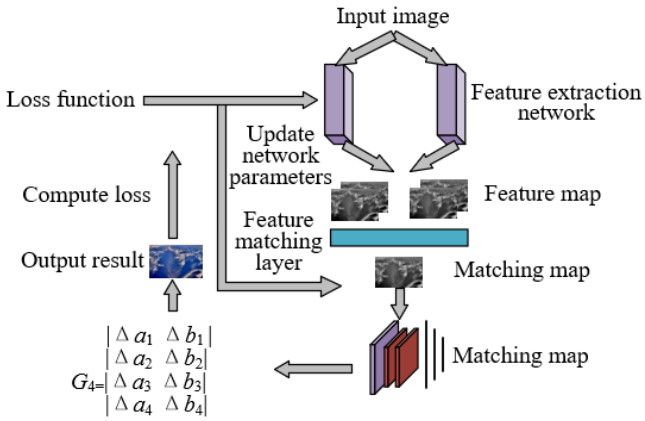

Figure 1. Overall structure diagram of multi-view image registration model for autonomous robot navigation

In the multi-view image registration process for autonomous robot navigation, only important features in the image can provide accurate registration references, while secondary features or noise points may interfere with the registration results. The attention mechanism, by simulating the selective attention of human vision, can effectively enhance the neural network's ability to extract key features from the image. This is particularly important for autonomous robot navigation, as the robot needs to navigate precisely in various complex environments, where accurate recognition of critical features such as roads and obstacles is crucial. Meanwhile, traditional image registration methods reduce the impact of mismatched features on the results by eliminating noise points. In deep learning models, the attention mechanism can dynamically adjust the focus, achieving a similar effect. In this way, the model can more robustly perform multi-view image registration, effectively improving registration performance. Figure 2 shows the operating principle of the feature extraction network.

This paper introduces a dual attention mechanism in the multi-view image registration task for autonomous robot navigation. During the navigation process, the autonomous robot needs to recognize and match images from different perspectives in complex environments, which may contain various rotated, deformed, and scaled features. The spatial attention mechanism in the dual attention mechanism can automatically adjust and optimize these features, thereby improving the accuracy of image registration. The robot needs to accurately recognize and match key features, such as roads, obstacles, and signs, in complex environments. Through the channel attention mechanism, the model can assign higher weights to critical channels, thereby highlighting important features and suppressing irrelevant background information.

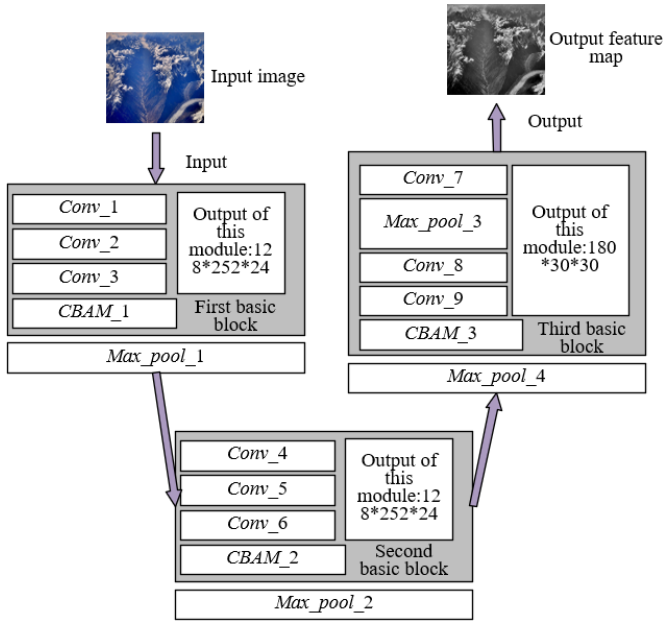

Figure 2. Operating principle of the feature extraction network

Specifically, the feature maps generated by each layer of the network can be regarded as detectors for a certain type of feature. The channel attention mechanism, through the structure of a multi-layer perceptron, evaluates the importance of each channel globally. To improve computational efficiency, max pooling and average pooling are used to compress the spatial dimensions, generating pooled feature descriptors. These are then further processed by the multi-layer perceptron to finally obtain the channel attention map, which adjusts the weight of each channel. Suppose the number of channels is denoted by Z, the input feature map by d, the sigmoid activation function by δ, the average pooling and max pooling operations by AP and MP respectively, the average and max pooling feature maps across channels by d^{Z}_{AVG} and d^{Z}_{MAX}, and the shared multi-layer perceptron weights for channel attention and spatial attention by Q_0 and Q_1. The generated max pooling feature descriptor is denoted by d^{Z}_{MAX}, the average pooling descriptor by d^{Z}_{AVG}, and the final obtained channel attention map by X_C. The calculation process of the attention map is expressed as follows:

$$\begin{aligned} X_Z(d) &= \delta\left(MLP\left(AP(d)\right), MLP\left(MP(d)\right)\right) \\ &= \delta\left(Q_1\left(Q_0\left(d_{AVG}^{Z}\right)\right) + Q_1\left(Q_0\left(d_{MAX}^{Z}\right)\right)\right) \end{aligned} \tag{5}$$

The spatial attention mechanism generates feature descriptors by performing average pooling and max pooling across the channel dimension, and then concatenates these two descriptors to capture key spatial features in the image. Next, through standard convolution operations and activation functions, the final spatial attention map $X_T(d)$ is generated. This process ensures that the model can identify and focus on important regions in the image, thereby enhancing the effectiveness of image registration. Assuming the convolution operation with a kernel size of 7 is denoted by $COV^{7\times7}$, and the feature maps obtained using average pooling and max pooling

are denoted by d^{T}_{AVG} and d^{T}_{MAX}, the calculation process of $X_T(d)$ is as follows:

$$\begin{aligned} X_T(d) &= \delta\left(COV^{7\times7}\left(\left[AP(d);MP(d)\right]\right)\right) \\ &= \delta\left(d^{7\times7}\left(\left[d^{T}_{AVG};d^{T}_{MAX}\right]\right)\right) \end{aligned} \tag{6}$$

The input to the dual attention module is the feature map d output by the convolutional network. The one-dimensional channel attention map and two-dimensional spatial attention map generated by the module are denoted by X_Z and X_T, respectively. Assuming element-wise multiplication is denoted by the symbol $\otimes$, the intermediate result of the attention module by d', and the final result of the attention module by d'', the process is expressed by the following equations:

$$\begin{aligned} d' &= X_Z(d)\otimes d \\ d'' &= X_T(d)\otimes d' \end{aligned} \tag{7}$$

In autonomous robot navigation tasks, the robot needs to recognize and match key features such as roads, obstacles, and signs from different perspectives in various complex environments. By normalizing the image size at the image input stage, generating high-dimensional feature descriptors during the feature extraction stage, and performing feature matching during network operation, the model can efficiently and accurately complete the image registration task.

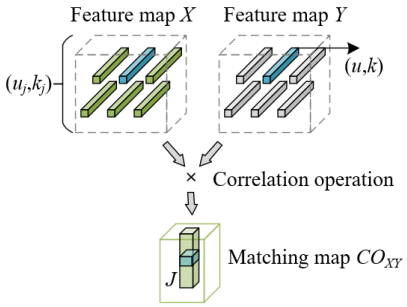

Figure 3. Operating principle of the feature matching layer

In the multi-view image registration model for autonomous robot navigation, the feature matching layer is also a critical component. Traditional feature matching methods often rely on operations such as element-wise subtraction or channel-wise concatenation. However, in this model, the feature extraction network employs a bidirectional correlation operation to enhance the accuracy of feature matching. Figure 3 shows the operating principle of the feature matching layer. It not only extracts rich features from the image but also retains the spatial location information of these features. The design of the bidirectional correlation operation is inspired by the mutual nearest neighbor algorithm, whose core advantage is the ability to effectively avoid mismatches. Suppose the input feature maps are denoted by X and Y, the bidirectional correlation operation achieves feature matching by calculating the similarity CO_{XY} between a feature (u, k) in Y and all features (u_j,k_j) in X. Suppose the feature maps of the reference image and the image to be registered are denoted by d_X and d_Y, the sizes of the feature maps by g and q, and the number of channels by f. The indices on the channel slices of the two feature maps are denoted by u and k, and the auxiliary index is denoted by $J=g(k_j\text{-}1)+u_j$. The calculation formula is as follows:

$$CO_{XY}(u,k,j) = d_Y(u,k)^{S}\, d_X(u_j,k_j) \tag{8}$$

The matched map is further normalized to eliminate mismatched features, as shown in the following formula. Suppose the matched map output by the correlation operation is denoted by CO, the first dimension size of the matched map by v, and the normalized matched map by d_v.

$$d_v = \frac{CO}{\sqrt{\sum_o^v CO^2 + \zeta}} \tag{9}$$

In the model, the parameter regression network module is used to predict the homography transformation parameters between images, thereby achieving accurate image registration. This module consists of two basic network-in-network blocks and three fully connected layers, each followed by a batch normalization layer and a ReLU activation function. Its input is a 256-dimensional feature map of size 16*16. First, the feature map passes through the first convolutional layer, with a convolutional kernel size of 5, an input channel of 256 dimensions, and an output channel of 128 dimensions. The main purpose of this layer is to extract high-level features within a relatively large receptive field. Next, the feature map sequentially passes through two convolutional layers with a kernel size of 1, where the number of channels is reduced to 64 and 320, respectively. These convolutional layers are primarily used to further compress and extract feature information in preparation for the subsequent fully connected layers. Afterward, the processed feature map is flattened into a 1152-dimensional vector and input into the fully connected layers. The structure of the fully connected layers includes a hidden layer, which reduces the dimensionality of the input feature vector from 1152 to 8, and ultimately outputs the homography transformation parameters. Through this process, the model can extract crucial parameter information for image registration from the high-dimensional features. Figure 4 shows the architecture of the parameter regression network.

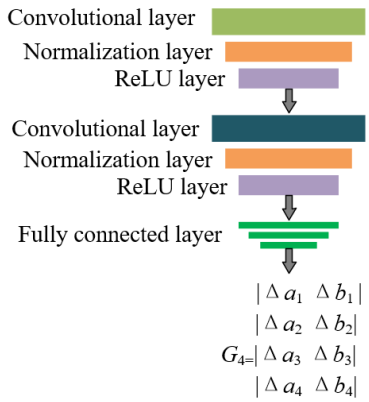

Figure 4. Parameter regression network architecture

The following are the specific steps for registration:

Step 1: Image preprocessing

First, preprocess the multi-view images during the autonomous robot navigation process, including resolution adjustment and enhancement processing. The purpose of this step is to standardize the size of the input images and improve

the quality of the images through enhancement processing, thereby providing a better foundation for subsequent feature extraction.

Step 2: Feature extraction

In the preprocessed images, use the trained network-in-network model with integrated attention mechanisms to perform feature extraction on the reference image and the image to be registered. This model, through the integrated attention mechanism, can better capture important features and details in the image, providing rich information for feature matching.

Step 3: Feature matching

Use the bidirectional correlation layer to process the two feature maps. The bidirectional correlation layer can calculate the similarity of all feature points in the two feature maps and output detailed matching information. This process draws on the mutual nearest neighbor algorithm to ensure the accuracy of the matching and reduce the possibility of mismatches.

Step 4: Parameter prediction

Based on the matching map output by the feature matching layer, use the parameter regression network to predict the homography transformation parameters from the image to be registered to the reference image. The parameter regression network, through multiple convolutional and fully connected layers, extracts key transformation information from the matched features, providing accurate transformation parameters for image registration.

Step 5: Solve the transformation model

Finally, compute the image transformation matrix to complete the multi-view image registration. By solving the transformation model, align the image to be registered with the reference image, achieving precise image registration. This step ensures that the robot can recognize the same scene features from different viewpoints for effective navigation and path planning.

5. EXPERIMENTAL RESULTS AND ANALYSIS

As seen in Table 1, different feature extraction networks show significant differences in terms of registration accuracy, memory usage, model size, and average prediction speed. Although LeNet-5 is an early feature extraction network, it achieved a registration accuracy of 88.5%, with memory usage of 689MB, a relatively small model size of 17.5MB, and an average prediction speed of 0.93s. ResNet-50 had a slightly lower registration accuracy of 87.6%, but its memory usage and model size were larger, at 700MB and 536MB respectively, with an average prediction speed of 1.24s. DenseNet-169 and DenseNet-201 achieved registration accuracies of 89.4% and 92.3%, with memory usage of 791MB and 834MB, model sizes of 584MB and 106.2MB, and average prediction speeds of 1.22s and 0.98s respectively. The proposed method achieved the highest registration accuracy of 93.4%, with memory usage of 832MB, a model size of 98.5MB, and an average prediction speed of 0.92s, demonstrating excellent performance. The comprehensive experimental results indicate that the proposed method exhibits a good balance and optimization in registration accuracy, memory usage, model size, and average prediction speed. Compared to DenseNet-201 and DenseNet-169, the proposed method shows a significant improvement in registration accuracy, with an increase of 1.1% and 4.0% respectively, while also demonstrating better optimization in memory usage and model size. Although the memory usage is slightly higher than DenseNet-169, it is lower than DenseNet-201. Compared to ResNet-50 and LeNet-5, the proposed method not only shows a substantial improvement in registration accuracy, with an increase of 5.8% and 4.9% respectively, but also exhibits clear advantages in model size and average prediction speed.

Table 1. Performance comparison of different feature extraction network structures

Model	Registration Accuracy	Memory Usage	Model Size	Average Prediction Speed
LeNet-5	88.5%	689MB	17.5MB	0.93s
ResNet-50	87.6%	700MB	536MB	1.24s
DenseNet-169	89.4%	791MB	584MB	1.22s
DenseNet-201	92.3%	834MB	106.2MB	0.98s
The Proposed Method	93.4%	832MB	98.5MB	0.92s

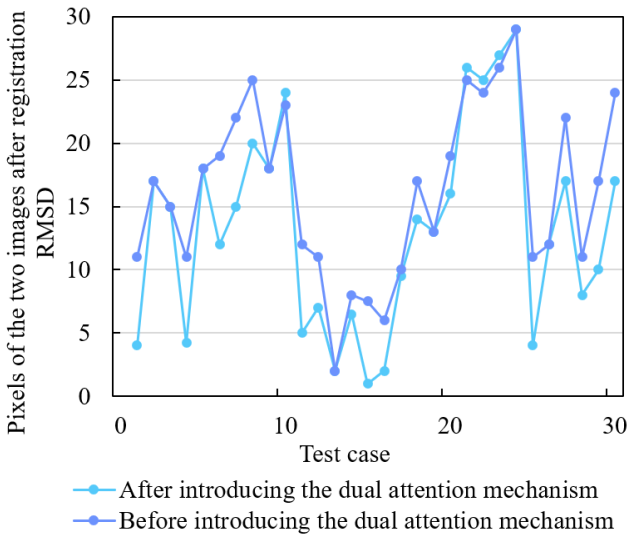

Figure 5. Registration error comparison with dual attention mechanism

Based on the data comparison in Figure 5, the image registration model with the dual attention mechanism exhibits lower registration errors across multiple test cases. Specifically, in subtests of test case numbers 0, 10, 20, and 30, the registration errors are significantly reduced after introducing the dual attention mechanism. For example, in test case 0, the registration error decreased from 11, 17, 15, 11 to 4, 17, 15, 4.2; in test case 10, the error decreased from 18, 19, 22, 25 to 18, 12, 15, 20; in test case 20, the error decreased from 12, 11, 2, 8 to 5, 7, 2, 6.5; and in test case 30, the error decreased from 10, 17, 13, 19 to 9.5, 14, 13, 16. Overall, the model with the dual attention mechanism shows a more concentrated error distribution across different test cases, with smaller error fluctuations. The comparison indicates that introducing the dual attention mechanism significantly improves the accuracy of the image registration model. Specifically, in multiple test cases, the model with the dual attention mechanism generally shows reduced errors, particularly at some high-error test points where the error reduction is especially noticeable. This indicates that the dual

attention mechanism effectively enhances the accuracy of feature matching, reduces error fluctuations, and improves the robustness of the model in different scenarios.

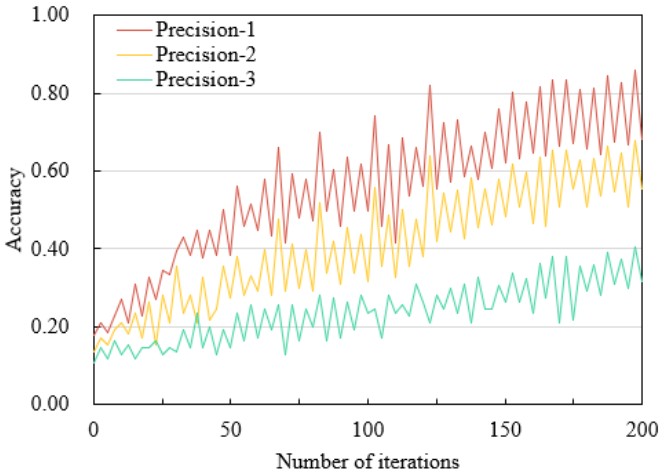

Figure 6. Model accuracy on training set for errors of 1, 3, and 5

From the data in Figure 6, it can be seen that the accuracy of the image registration model improves significantly at different iteration counts when errors are set to 1, 3, and 5. When the error is set to 1, the accuracy gradually increases from the initial 0.18 to 0.68 after 200 iterations. When the error is set to 3, the initial accuracy is 0.11, and after 200 iterations, the final accuracy reaches 0.32. When the error is set to 5, the accuracy increases from 0.14 at 0 iterations to 0.55 after 200 iterations. Overall, the accuracy of the model shows a steady growth trend with increasing iterations, with particularly significant growth after 50 iterations. Through comparative analysis, the model's accuracy with an error of 1 exhibits the fastest improvement and the highest final accuracy throughout the training process, indicating that the model is easier to optimize and improve accuracy under low-error conditions. The accuracy with an error of 3 is second, showing that the model's training effect under moderate error conditions is also quite good. The model's accuracy with an error of 5 increases more slowly, and the final accuracy is the lowest, possibly because the model is more challenging to optimize under high-error conditions, limiting the training effect. In summary, the model training effect is best under low-error conditions, moderate under moderate-error conditions, and more difficult to optimize under high-error conditions. However, overall, the model accuracy can significantly improve with increasing iterations, validating the effectiveness and feasibility of the proposed image registration model in autonomous robot navigation applications. Figure 7 provides a more intuitive display of the registration results on real multi-view images of autonomous robot navigation when errors are set to 1, 3, and 5.

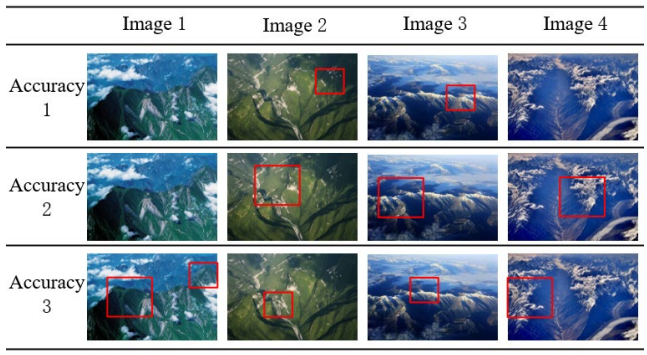

Figure 7. Comparison of registration results on real multi-view images in autonomous robot navigation

Table 2. Quantitative analysis of comparison experiments average data results

	RMSD		MSD		STD		MD	
The proposed method	5.62	Min:3.46 Max:17.58	6.35	Min:2.89 Max:25.36	4.56	Min:1.78 Max:15.58	5	Min:3 Max:25.36
Reference method 1	15.68	Min:3.36 Max:75.48	16.58	Min:4.57 Max:62.35	14.26	Min:4.05 Max:43.48	12	Min:2.14 Max:65.48
Reference method 2	7.2	Min:2.4 Max:33.65	8.69	Min:1.75 Max:39.36	6.35	Min:1.85 Max:18.36	10.6	Min:1.4 Max:32
Reference method 3	35.6	Min:18.9 Max:62.58	25	Min:21.72 Max:52.34	23.58	Min:16.35 Max:36.55	22	Min:3.3 Max:51.2
Reference method 4	37.54	Min:16.58 Max:151.8	42.6	Min:12.36 Max:123	47.59	Min:16.25 Max:88.54	40.26	Min:4.3 Max:132.85
Reference method 5	30	Min:14.5 Max:168.5	25.69	Min:12.15 Max:153.68	32.65	Min:8.36 Max:1.804	36.56	Min:9.5 Max:136.5

According to the data shown in Table 2, the proposed method outperforms Reference Methods 1-5 in RMSD, MSD, STD, and MD indicators. Specifically, the proposed method has an RMSD mean of 5.62, which is significantly lower than Reference Method 1's 15.68, Reference Method 3's 35.6, and Reference Method 4's 37.54. For MSD, the proposed method has a mean of 6.35, also significantly lower than other methods such as Reference Method 4's 42.6 and Reference Method 5's 25.69. In terms of STD, the proposed method has a mean of 4.56, showing lower error variability, whereas Reference Methods 4 and 5 have STD means of 47.59 and 32.65, respectively, indicating higher variability. For MD, the proposed method has a mean of 5, while Reference Methods 1 and 4 have means of 12 and 40.26, respectively, demonstrating the proposed method's clear advantage in maximum deviation. The comparative analysis shows that the proposed method significantly surpasses traditional methods in image registration accuracy. This is mainly due to the introduction of the dual attention mechanism and bidirectional correlation operations in the feature matching layer, which greatly enhances the model's ability to capture and match image features. The proposed method not only performs excellently in terms of mean values but also shows higher stability and robustness within the range of minimum and maximum values, especially in RMSD and STD indicators, indicating that the model can maintain high registration accuracy and stability in

various scenarios.

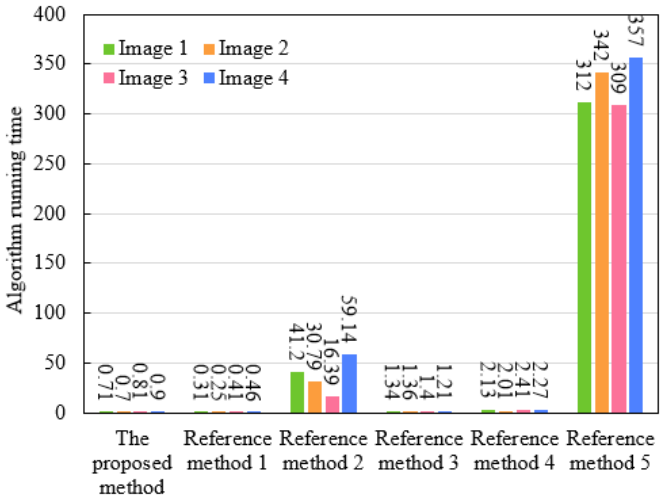

Figure 8. Time performance of each algorithm on test data

The data in Figure 8 shows that the proposed method exhibits significant advantages in time performance for image registration. For the four test images, the processing times of the proposed method are 0.71, 0.7, 0.81, and 0.9 seconds, all significantly lower than those of other reference methods. For example, Reference Method 1's processing time ranges from 0.25 to 0.46 seconds, although slightly faster than the proposed method, it performs worse in registration accuracy. Reference Methods 2, 3, 4, and 5 have processing times of up to 16.39 to 59.14 seconds, 1.21 to 1.4 seconds, 2.01 to 2.41 seconds, and 309 to 357 seconds, respectively, all far exceeding the proposed method, especially Reference Method 5, with the longest processing time of up to 357 seconds. Overall, the proposed method significantly reduces processing time while maintaining high registration accuracy. The comparative analysis reveals that the proposed method has a notable advantage in time performance for image registration, completing high-precision registration in a shorter time. This advantage is primarily due to the introduction of the dual attention mechanism and bidirectional correlation operations in the feature matching layer, which optimize the feature extraction and matching process, improving processing efficiency. In contrast, Reference Method 1, while having a shorter processing time, does not achieve the same level of registration accuracy as the proposed method. Meanwhile, Reference Methods 2, 3, 4, and 5 perform significantly worse in registration time compared to the proposed method, with Reference Method 5 in particular showing extremely long processing times, making it less suitable for practical applications.

6. CONCLUSION

This paper studies the application of deep learning-based image registration techniques in autonomous robot navigation, proposes an innovative image registration model, and conducts systematic experimental validation. The research covers three main aspects: the proposal and validation of the model hypothesis, dataset generation, and overall network design. Specifically, this paper proposes a deep learning model hypothesis suitable for image registration and provides data support for the model's training and evaluation by constructing an autonomous robot navigation image registration dataset. The network structure designed in this paper includes network-in-network, dual attention mechanisms, bidirectional correlation operations in the feature matching layer, and parameter regression networks, optimizing the image registration effect from multiple perspectives. Experimental results show that different feature extraction network structures have varying performance, and the introduction of dual attention mechanisms significantly reduces the registration error of the image registration model. Especially, when the error is 1, 3, or 5, the image registration model performs excellently on the training set, far surpassing traditional methods. The registration results on real multi-view images in autonomous robot navigation also demonstrate the practicality and reliability of the proposed method. Through quantitative analysis and time performance comparison, the proposed method significantly shortens processing time while maintaining high-precision registration, showing its efficiency in practical applications.

The research value of this paper lies mainly in the following aspects: first, it proposes a deep learning model suitable for image registration, effectively improving the accuracy and efficiency of autonomous robot navigation; second, it proves the advantages of dual attention mechanisms and bidirectional correlation operations in image registration through systematic experimental validation. The limitations of this study include the limited diversity and scale of the dataset, which may affect the model's generalization capability. Future research directions include expanding the dataset scale, improving the model's generalization ability, further optimizing the network structure, and exploring other deep learning technologies in image registration. Through these efforts, it is expected to further enhance the intelligence level and practical application effectiveness of autonomous robot navigation.

FUNDING

This paper was supported by the project of Jilin Provincial Science and Technology Department (Grant No.: 20210402081GH), the Innovation and Entrepreneurship Talent Funding project of Jilin Province (Grant No.: 2023RY17) and the project of Jilin Provincial Development and Reform Commission (Grant No.: 2023C042-4).

REFERENCES

[1] Bose, D., Mohan, K., CS, M., Yadav, M., Saini, D.K. (2023). Review of autonomous campus and tour guiding robots with navigation techniques. Australian Journal of Mechanical Engineering, 21(5): 1580-1590. https://doi.org/10.1080/14484846.2021.2023266

[2] Wijayathunga, L., Rassau, A., Chai, D. (2023). Challenges and solutions for autonomous ground robot scene understanding and navigation in unstructured outdoor environments: A review. Applied Sciences, 13(17): 9877. https://doi.org/10.3390/app13179877

[3] Ravankar, A., Ravankar, A.A., Rawankar, A., Hoshino, Y. (2021). Autonomous and safe navigation of mobile robots in vineyard with smooth collision avoidance. Agriculture, 11(10): 954. https://doi.org/10.3390/agriculture11100954

[4] De Luca, A., Muratore, L., Tsagarakis, N.G. (2023).

Autonomous navigation with online replanning and recovery behaviors for wheeled-legged robots using behavior trees. IEEE Robotics and Automation Letters, 8(10): 6803-6810. https://doi.org/10.1109/LRA.2023.3313052

[5] Jiménez, D.D.J.G., Olvera, T., Orozco-Rosas, U., Picos, K. (2021). Autonomous object manipulation and transportation using a mobile service robot equipped with an RGB-D and LiDAR sensor. Optics and Photonics for Information Processing XV, 11841: 92-111. https://doi.org/10.1117/12.2594025

[6] de Sousa Bezerra, C.D., Teles Vieira, F.H., Queiroz Carneiro, D.P. (2023). Autonomous robotic navigation approach using deep q-network late fusion and people detection-based collision avoidance. Applied Sciences, 13(22): 12350. https://doi.org/10.3390/app132212350

[7] Frisk, H., Burström, G., Persson, O., El-Hajj, V.G., Coronado, L., Hager, S., Edström, E., Elmi-Terander, A. (2024). Automatic image registration on intraoperative CBCT compared to surface matching registration on preoperative CT for spinal navigation: Accuracy and workflow. International Journal of Computer Assisted Radiology and Surgery, 19(4): 665-675. https://doi.org/10.1007/s11548-024-03076-4

[8] de Geer, A.F., de Koning, S.B., van Alphen, M.J.A., Van der Mierden, S., Zuur, C.L., Van Leeuwen, F.W.B., Loeve, A.J., van Veen, R.L.P., Karakullukcu, M.B. (2022). Registration methods for surgical navigation of the mandible: A systematic review. International Journal of Oral and Maxillofacial Surgery, 51(10): 1318-1329. https://doi.org/10.1016/j.ijom.2022.01.017

[9] Smit, J.N., Kuhlmann, K.F., Ivashchenko, O.V., Thomson, B.R., Langø, T., Kok, N.F., Fusaglia, M., Ruers, T.J. (2022). Ultrasound-based navigation for open liver surgery using active liver tracking. International Journal of Computer Assisted Radiology and Surgery, 17(10): 1765-1773. https://doi.org/10.1007/s11548-022-02659-3

[10] Schreurs, R., Baan, F., Klop, C., Dubois, L., Beenen, L.F.M., Habets, P.E.M.H., Becking, A.G., Maal, T.J.J. (2021). Virtual splint registration for electromagnetic and optical navigation in orbital and craniofacial surgery. Scientific Reports, 11(1): 10406. https://doi.org/10.1038/s41598-021-89897-8

[11] Chiurillo, I., Sha, R.M., Robertson, F.C., Liu, J., Li, J., Le Mau, H., Amich, J.M., Gormley, W.B., Stolyarov, R. (2023). High-accuracy neuro-navigation with computer vision for frameless registration and real-time tracking. Bioengineering, 10(12): 1401. https://doi.org/10.3390/bioengineering10121401

[12] Taleb, A., Guigou, C., Leclerc, S., Lalande, A., Bozorg Grayeli, A. (2023). Image-to-patient registration in computer-assisted surgery of head and neck: State-of-the-art, perspectives, and challenges. Journal of Clinical Medicine, 12(16): 5398. https://doi.org/10.3390/jcm12165398

[13] Zhang, F., Zhang, S., Sun, L., Zhan, W., Sun, L. (2022). Research on registration and navigation technology of augmented reality for ex-vivo hepatectomy. International Journal of Computer Assisted Radiology and Surgery, 17: 147-155. https://doi.org/10.1007/s11548-021-02531-w

[14] Alvarez-Breckenridge, C., Muir, M., Rhines, L.D., Tatsui, C.E. (2021). The use of skin staples as fiducial markers to confirm intraoperative spinal navigation registration and accuracy. Operative Neurosurgery, 21(3): E193-E198. https://doi.org/10.1093/ons/opab132

[15] Hiep, M.A., Heerink, W.J., Groen, H.C., Ruers, T.J.M. (2023). Feasibility of tracked ultrasound registration for pelvic–abdominal tumor navigation: A patient study. International Journal of Computer Assisted Radiology and Surgery, 18(9): 1725-1734. https://doi.org/10.1007/s11548-023-02937-8

[16] Hayashi, Y., Misawa, K., Mori, K. (2023). Database-driven patient-specific registration error compensation method for image-guided laparoscopic surgery. International Journal of Computer Assisted Radiology and Surgery, 18(1): 63-69. https://doi.org/10.1007/s11548-022-02804-y

[17] Loerch, A.C., Stow, D.A., Coulter, L.L., Nara, A., Frew, J. (2022). Comparing the accuracy of sUAS navigation, image co-registration and CNN-based damage detection between traditional and repeat station imaging. Geosciences, 12(11): 401. https://doi.org/10.3390/geosciences12110401

[18] Gambrych, J., Gromek, D., Abratkiewicz, K., Wielgo, M., Gromek, A., Samczyński, P. (2023). SAR and orthophoto image registration with simultaneous SAR-based altitude measurement for airborne navigation systems. IEEE Transactions on Geoscience and Remote Sensing, 61: 5219714. https://doi.org/10.1109/TGRS.2023.3327090

[19] Taeger, J., Müller-Graff, F.T., Neun, T., Köping, M., Schendzielorz, P., Hagen, R., Rak, K. (2021). Highly precise navigation at the lateral skull base by the combination of flat-panel volume CT and electromagnetic navigation. Science Progress, 104(3): 00368504211032090. https://doi.org/10.1177/00368504211032090

[20] Sommer, F., Goldberg, J.L., McGrath, L., Kirnaz, S., Medary, B., Härtl, R. (2021). Image guidance in spinal surgery: A critical appraisal and future directions. International Journal of Spine Surgery, 15(s2): S74-S86. https://doi.org/10.14444/8142

[21] Dhoju, R., Alsadoon, A., Prasad, P.W.C., Al-Saiyd, N.A., Alrubaie, A. (2021). Augmented reality navigation for liver surgery: An enhanced coherent point drift algorithm based hybrid optimization scheme. Multimedia Tools and Applications, 80(18): 28179-28200. https://doi.org/10.1007/s11042-021-11070-0