electronics

MDPI

Article

IFSrNet: Multi-Scale IFS Feature-Guided Registration Network Using Multispectral Image-to-Image Translation

Bowei Chen [1], Li Chen [1,*], Umara Khalid [1,†] and Shuai Zhang [2,†]

[1] School of Information Science and Technology, Northwest University, Xi'an 710127, China; bwoweichen@stumail.nwu.edu.cn (B.C.); 2018880039@stumail.nwu.edu.cn (U.K.)
[2] Dongdian Testing Technology Xi'an Corporation, Xi'an 710075, China; zhang.shuai.eric@dgddt.com
* Correspondence: chenli@nwu.edu.cn
† These authors contributed equally to this work.

Abstract: Multispectral image registration is the process of aligning the spatial regions of two images with different distributions. One of the main challenges it faces is to resolve the severe inconsistencies between the reference and target images. This paper presents a novel multispectral image registration network, Multi-scale Intuitionistic Fuzzy Set Feature-guided Registration Network (IFSrNet), to address multispectral image registration. IFSrNet generates pseudo-infrared images from visible images using Cycle Generative Adversarial Network (CycleGAN), which is equipped with a multi-head attention module. An end-to-end registration network encodes the input multispectral images with intuitionistic fuzzification, which employs an improved feature descriptor—Intuitionistic Fuzzy Set–Scale-Invariant Feature Transform (IFS-SIFT)—to guide its operation. The results of the image registration will be presented in a direct output. For this task we have also designed specialised loss functions. The results of the experiment demonstrate that IFSrNet outperforms existing registration methods in the Visible–IR dataset. IFSrNet has the potential to be employed as a novel image-to-image translation paradigm.

Keywords: multispectral registration; intuitionistic fuzzy set; multi-scale features; multi-headed attention

Citation: Chen, B.; Chen, L.; Khalid, U.; Zhang, S. IFSrNet: Multi-Scale IFS Feature-Guided Registration Network Using Multispectral Image-to-Image Translation. *Electronics* **2024**, *13*, 2240. https://doi.org/10.3390/electronics13122240

Academic Editor: Silvia Liberata Ullo and Li Zhang

Received: 30 April 2024
Revised: 27 May 2024
Accepted: 30 May 2024
Published: 7 June 2024

1. Introduction

The information derived from infrared and visible imaging is complementary, as the former captures details on the intensity of temperature radiation emitted by the target, whereas the latter reflects information regarding the texture and contours of the target. Despite multispectral images being captured concurrently within the same environment, disparities persist among them due to the diverse intensities, gradients, and structures associated with different wavelengths of light. Multispectral image registration [1] aims to establish the mapping relationship between the reference image and the image to be registered, achieving geometric calibration through various methods. The primary process involves selecting the appropriate image registration technique, extracting feature points, performing feature matching, and evaluating the registration accuracy. In the field of medical image diagnosis [2], the registration captured at different times or in different modalities can assist physicians in formulating more accurate treatment plans. In the context of remote sensing [3], the registration of multispectral images can provide detailed information about the ground. Additionally, this technology can be employed in monitoring the distribution and dispersion of environmental pollutants [4] and in precision agricultural management [5]. It is evident that the research into multispectral image registration has a wide range of possible applications.

The visible image is typically a composite image of the entire visible wavelength band (380–780 nm), whereas the infrared image involves separate imaging of a single band to capture the spectral information. Figure 1 depicts multispectral images captured at

various wavelengths. Consequently, the radiometric difference between visible and infrared images is nonlinear. The proportion and repeatability of local feature points of the same objects occupied in images of different bands will decrease. This elevates the mismatching rate of local features, and compromises the quality and precision of image registration. The principal obstacle in multispectral image registration is the inconsistency in feature intensities between pairs of images, which precludes the direct registration of images from the same scene through the matching of feature descriptors. The method of leveraging Generative Adversarial Network (GAN)-based image-to-image translation [6] ingeniously circumvents the intricate spectral difference issue, simplifying the registration process.

(a) (b) (c) (d)

Figure 1. Multispectral images. The images are presented in the following order from left to right: (**a**) RGB, (**b**) panchromatic, (**c**) NIR, and (**d**) LWIR image. Different numbers of channels and wavelengths are the criteria for classifying these images.

However, owing to the instability in data quality, it is challenging for a GAN to discern the interest part from the vast array of features solely through the fusion of a reference images. In essence, this precludes the model from generating images with impeccable detail. Notably, the majority of current mainstream feature fusion techniques [7] rely on linear addition; their fitting capabilities require further enhancement. As shown in Figure 2. CycleGAN [8] employs a spatial loss of cyclic consistency to facilitate the transformation of images from infrared to visible and vice versa. This bidirectional mapping of image-to-image transformations also ensures the more accurate and complete preservation of the structural information of the object. Nevertheless, this bidirectional mapping is constrained in its ability to accommodate complex scenarios. This is because, regardless of whether the discriminator generates a score value or a score map for the input image, it is merely a method of providing an approximate score for the entire image. The judging process is rendered too coarse due to the failure to consider local features. Furthermore, the generation of a score map for each pixel in the image would be excessively detailed, thereby rendering it challenging to maintain consistency in the judgement of both local and global features. It is not appropriate to employ either the entire image or individual pixels as a reference for details in the modal transformation, as visible images are rich in background information.

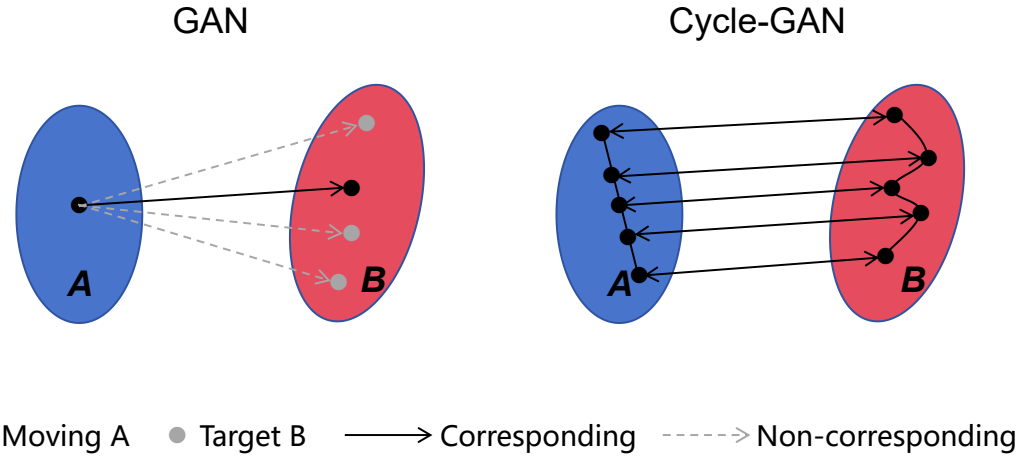
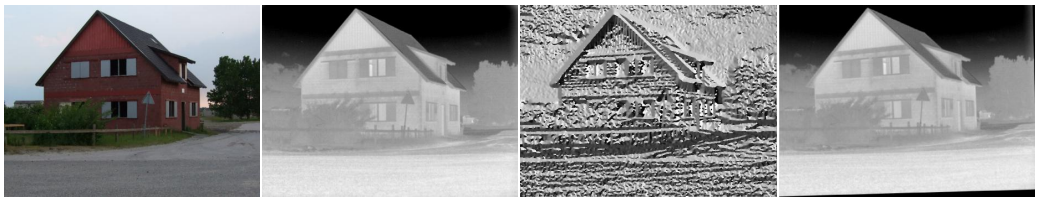


Figure 2. GAN-based methods can only ensure that the distribution of domain A corresponds to the distribution of domain B. It is desirable that the two domains can feed into each other, ideally.

Given that fuzzy theory offers a solution to uncertain data aggregation, this study endeavours to refine the feature additivity assumption and incorporate the concept of intuitionistic fuzzy sets [9]. The IFS-SIFT feature maps contain richer details as can be observed in Figure 3. This paper proposes a multi-scale [10] IFS feature-guided multispectral image registration method using image-to-image translation, which is termed IFSrNet. This involves a nonlinear fusion between extracted features on different scales and reference to obtain relatively correct feature detail from the pseudo-images generated by the CycleGAN with multi-head attention module.



(a) (b) (c) (d)

Figure 3. (**a**) Visible image as a reference. (**b**) Infrared images to be registered. (**c**) IFS-SIFT feature image. (**d**) Infrared image that has been registered.

This paper has constructed an end-to-end network [11], and the experiments demonstrate that the registration network exhibits comparable accuracy to other registration methods. The principal contributions of the proposed approach are asfollows:

1. The paper uses pseudo-infrared (IR) images created from reference images to overcome the discrepancy between multispectral paired images. The training data and multi-head attention module facilitates the learning of the generative model, thereby enabling the utilisation of convolutional neural networks (CNNs) for image registration.
2. This paper introduces the concept of intuitionistic fuzzy set features, which serve as an extension of gradient information. Furthermore, the two channels of the target and reference images are integrated by a multi-scale concatenation.
3. A novel loss function is designed for the registration network to increase the weight of matching unambiguous feature information. This approach not only preserves the structural integrity of the generated images, but also mitigates the inherent limitation of generative models, which cannot be aligned pixel by pixel.

2. Related Work

In recent years, a number of methodologies have been developed with the aim of extracting consistent features from disparate image modalities for image registration. Nunes et al. [12] developed a multispectral feature descriptor (MFD) to extract invariant gradient information in both the spatial and frequency domains via LogGabor filters. Gao et al. [13] constructed a partial principal orientation map to obtain robust orientation information, and simultaneously employed gradient location and orientation histogram (GLOH) descriptors to achieve intensity invariance. Furthermore, a number of methods for multimodal image registration based on deep features have been proposed. Xu et al. [14] adopted a coarse-to-fine approach to registration and employed image fusion techniques to facilitate multimodal image registration. Wei et al. [15] proposed a gradient-guided multispectral image registration method utilising a convolutional neural network, known as the gradient-guided registration network for multispectral images; RegiNet is an end-to-end network that takes the gradient maps of both the target image and the reference image as inputs, generating the registered image as its output. Zhang et al. [16] proposed the histogram of weighted phase direction (HOWP), which is employed to reduce the discrepancy between multimodal contrasts.

The optical, geometric, and spatial features expressed by infrared and visible images are significantly different. Establishing spatial relationships between two or more points using convolutional neural networks is a challenging task. Registration methods that rely

on image-to-image translation [17] can successfully map visible images to infrared images. Some attempts based on image-to-image translation [18,19] employed image generation techniques to transform visible images into infrared images. This approach provided training data and circumvented the issue of cross-modal matching. However, these methods still require the traditional utilisation of feature point extraction operators, such as Scale-Invariant Feature Transformation (SIFT) [20], Speeded Up Robust Features (SURF) [21], and Partial Intensity Invariant Feature Descriptor (PIIFD) [22]. Kumari et al. [23] achieved the registration of infrared and visible light images by utilising a generative adversarial network equipped with a spatial transformer module. The adversarial loss compelled the generator to produce a pseudo-infrared image, which was then compared to the original infrared image in a discriminator to assess the realism of the generated pseudo-infrared image. Mao et al. [24] leveraged the benefits of transfer learning to enhance feature matching. They used a parallel convolutional autoencoder to reconstruct images in both the visible and infrared branches before they are entered into the adversarial sub-network. Bingchao Yang et al. [25] chose to harness the modal shifting capabilities of GAN to generate pseudo-infrared images from infrared images. They combined the SURF algorithm with the PIIFD feature descriptor to extract and generate image feature points.

3. Methodology

3.1. IFS Feature Image

Infrared images are abundant in structural information. In the training of generative adversarial networks, it is imperative to segregate and eliminate distributional disparities while giving sole attention to spatial structural differences. This is crucial for simplifying the subsequent registration task. Traditionally, GAN-based methods strive to eradicate distributional differences by transforming images from the source domain into the target domain. Nevertheless, even though translated and target images may appear isospectral, residual distribution disparities can still be substantial. This renders the mean absolute error (MAE) [26] or mean squared error (MSE) [27] unsuitable for optimising the registration network.

To tackle this issue, this paper devised an intuitionistic fuzzy feature image grounded in gradient information. This preliminary design is intended to address the issue of non-consistency in the output image of the GAN model. It amplifies the impact of the large gradient direction and reduces the noise in the small gradient direction. Intuitionistic fuzzy sets represent the most significant expansion and development of fuzzy set theory. Fuzzy sets can be used to describe the concept of 'both positive and negative'. Intuitionistic fuzzy sets propose non-membership and hesitancy based on membership, which can be used to describe the neutral state of 'neither one nor the other'. In the application of IFS, the determination of three fuzzy description measures is crucial. This is a hot and difficult research topic nowadays and one of the main contributions of this paper. The feature vector of the ith feature of the image R is denoted by SIFT [28] as $R_i = \left(\varphi_1^i, \varphi_2^i, \varphi_3^i, \varphi_4^i, \varphi_5^i, \varphi_6^i, \varphi_7^i, \varphi_8^i\right)$, and the jth feature vector of image S is represented by IFS-SIFT as $\S_j = \left(\mu_1^j, \mu_2^j, \mu_3^j\right)$. The eight feature vectors φ_k^i presented here have been derived from the design in SIFT. As there are eight neighbouring pixels to the target pixel point, the gradient between each neighbouring pixel and the target is defined as a feature vector in the direction of that gradient. Although multi-dimensional descriptors may be more robust in labelling features, some directions with small gradient differences are not worthy of consideration in this task. Furthermore, this approach places a significant computational burden on the network. For the kth moment $\varphi_k (1 \leqslant k \leqslant 8)$ and $\mu_k (1 \leqslant k \leqslant 3)$ subset φ_k^i and μ_k^j, this article aims to find the direction with the highest magnitude in φ_k^i, designated as the direction of membership. To calculate the membership μ_1^j, add the φ_{k+1}^i and φ_{k-1}^i magnitudes of the two directions with the smallest adjacent angle. Then, determine the proportion of this sum to the total gradient magnitudes of all eight directions to obtain the membership of the main direction. The unaffiliated degree μ_2^j is the ratio of

the gradient magnitudes in the three opposite directions. The directions perpendicular to the main direction are the hesitant degree directions μ_3^j. Therefore, the membership $\S_i\left(\mu_1^j\right)$, non-membership $\S_j\left(\mu_2^j\right)$, and hesitancy $\S_j\left(\mu_3^j\right)$ of μ_k^j in $\S_j$ are defined as follows. The idea of computation is inspired by [29].

$$\S_j\left(\mu_1^j\right) = \frac{\varphi_k^i + \varphi_{k+1}^i + \varphi_{k-1}^i}{\sum_{k=1}^{k=8} \varphi_k^i} \tag{1}$$

$$\S_j\left(\mu_2^j\right) = \frac{\varphi_{k'}^i + \varphi_{k'+1}^i + \varphi_{k'-1}^i}{\sum_{k=1}^{k=8} \varphi_k^i}, k' = k \pm 4 \tag{2}$$

$$\S_j\left(\mu_3^j\right) = \frac{\varphi_{k''}^i + \varphi_{k''\pm4}^i}{\sum_{k=1}^{k=8} \varphi_k^i}, k'' = k \pm 2 \tag{3}$$

The feature descriptor based on the IFS-SIFT reduces computational complexity compared to the SIFT feature descriptor while maintaining high accuracy and accounting for the correlation of neighbouring pixels. When defining the direction of key points in SIFT, the gradient direction and magnitude of all pixels within a circle centred on the feature point and radiuses by 1.5 times the scale of the Gaussian image in which the feature point is located are counted. The 8-direction histograms of the key points are obtained by Gaussian filtering. When constructing the key point descriptor, the region is divided into 4× 4 sub-blocks. The gradient magnitude of each direction is obtained by performing histogram statistics of 8 directions for each sub-block, resulting in a total of 128 dimensional descriptor vectors. The IFS-SIFT feature descriptor has the advantage of reflecting correlation between adjacent pixels due to its 48-dimensional vectors. In Figure 4, although gradient maps are adept at communicating information regarding the edges of an image, the presence of disparities in residual distributions can increase the model's sensitivity to local features within the image task. Given the high performance of infrared images in the transmission of structural information, structural features should be accorded more attention in subsequent registration tasks. Intuitionistic fuzzy feature maps, which place more emphasis on global structural information, exhibit significant potential.

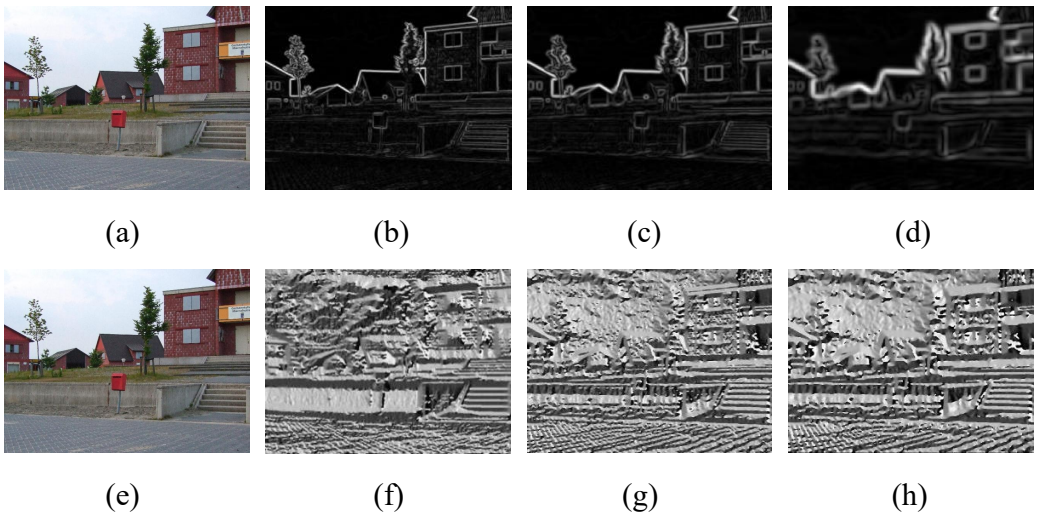


Figure 4. Results of feature maps in the registration network. The first row of (**a**–**d**) are the reference, and gradient maps of 60, 20, and 5 epochs, respectively. The second row of (**e**–**h**) shows the opposite reference, and IFS maps for the same epoch.

3.2. Model Frame

The paper has built an end-to-end network that connects the features of an IR image with a pseudo-IR image to perform a registration transform on the IR image. Figure 5

presents the comprehensive architectural framework of the network. Pseudo-infrared images were produced from visible images by CycleGAN [30]. Since the image-to-image translations involved in this paper are global cross-modal transformations, the sensitivity of the generative model to global information is crucial. The attention [31] mechanism is capable of mining remote dependencies in order to obtain global information through global interaction. This approach is more advantageous for high-level semantic feature extraction. Specifically, this paper adopts the multi-head attention [32] mechanism proposed in Transformer [33], whereby the three modular inputs are passed through a convolutional block to obtain Q, K, and V. Subsequently, the acquired Q, K, and V are employed to execute multi-head attention computation. This attention mechanism enables the generation of a more comprehensive feature representation by leveraging the multi-head attention.

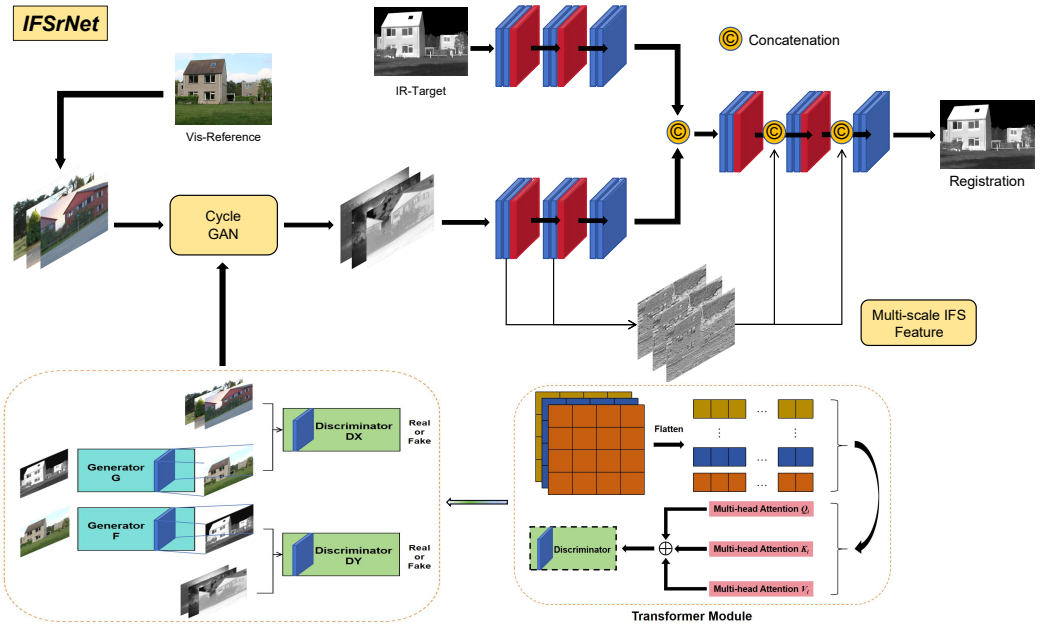


Figure 5. The network architecture of IFSrNet. IFSrNet is an end-to-end network that utilises pseudo-IR images created by generative model as the reference, which are fed into the registration network along with the target image registration.

The generator in the modified CycleGAN framework used in this paper employs a standard encoder–decoder network as the generator, which is composed primarily of a deep feature extraction module and an image reconstruction module. This generator is capable of producing high-quality images, including images of optimal resolution with rich detail. The U-shaped discriminator network, as depicted in Figure 6, incorporates a multi-head attention module that performs true–false judgements for different patch sizes. This differs from the conventional approach of improving the global judgement of the generated image over the whole image. This is accomplished by fusing the encoder and decoder output feature maps of varying spatial dimensions and subsequently passing them through distinct convolutional layers, thereby generating a single-channel feature map comprising three distinct resolution patches. The design facilitates the enhancement of accuracy in classification. The regular updating of a generator can better simulate the distribution of real data, thereby generating images with superior quality.

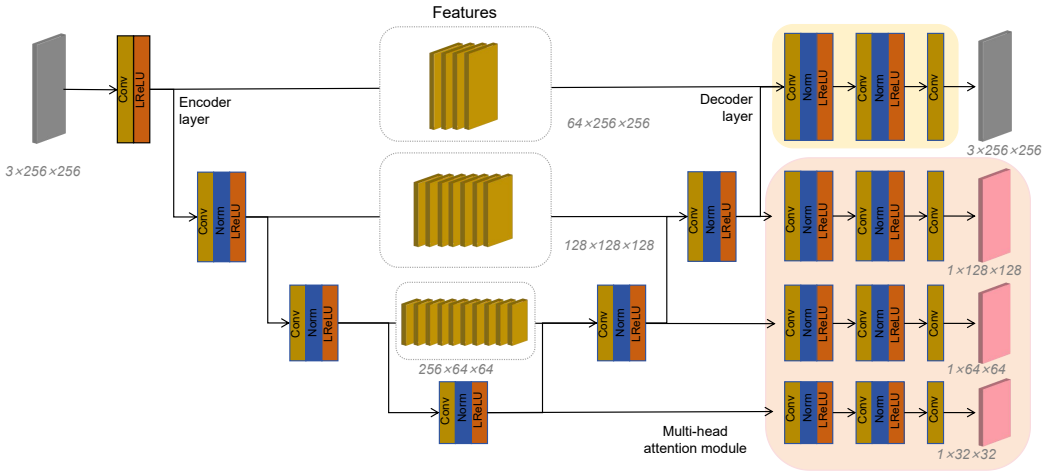


Figure 6. Architecture of the multi-headed discriminator of the proposed IFSrNet.

The input and output default image size is 512 × 512, but other sizes can also be supported. Rectified Linear Unit (ReLU) activation functions follow each convolutional layer except the last, which has an output. Skip connections are added to the network and combined by concatenation. A detailed enumeration of the specific parameters pertaining to the registration network is presented in Table 1. The IFSrNet input consists of two branches that extract the features of the target image and the intuitionistic fuzzy set feature map of the reference image. In Algorithm 1, the paper presents a pseudo-code flow for model training.

Table 1. Summary of the registration network.

Number	Layer Index	Branch	Depth	Stride	Output Size (Modifiable)
1	Convolution	2	64	1	512 × 512
2	Convolution	2	64	1	512 × 512
3	Conv+Dropout	2	64	0.5	512 × 512
4	Convolution	2	128	1	256 × 256
5	Convolution	2	128	1	256 × 256
6	Conv+Dropout	2	128	0.5	256 × 256
7	Convolution	2	256	1	128 × 128
8	Convolution	2	256	1	128 × 128
9	Convolution	1	256	1	128 × 128
10	Convolution	1	256	1	128 × 128
11	Conv+Dropout	1	256	2	128 × 128
12	Convolution	1	128	1	256 × 256
13	Convolution	1	128	1	256 × 256
14	Conv+Dropout	1	128	2	256 × 256
15	Convolution	1	64	1	512 × 512
16	Convolution	1	64	1	512 × 512

The potential issues of overfitting and weak generalisation in complex registration models have been addressed by the effectiveness of Dropout in numerous deep learning vision tasks. The Dropout mechanism is employed to randomly disable some units in a network and generate multiple sub-networks. This approach is intended to alleviate the overfitting problem and enhance the generalisation performance of the network. In this network, a dropout layer is incorporated subsequent to the activation layer, and distinct dropout parameters are set to perform the registration training. The quality of the registered images output from the network is evaluated by Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) in order to identify the optimal Dropout parameters. The initial Dropout module employs a probability of 0.1, 0.2, and 0.3.

Algorithm 1 Registration network training procedure

Initialise the input size i, the kernel size k, and the stride s, and the output size o. Initialise model parameters according to pre-training settings;
Input: training multispectral image dataset D;
Output: trained model N;
for epochs **do**
  for I_{Vis},I_{IR} in D **do**
    # Forward propagation
    $I_{Vis}{}'$ = $DeCNN$ (I_{Vis}) , $I_{IR}{}'$ = $DeCNN$ (I_{IR})
    # Calculated IFS-SIFT feature map
    $\S_j = \left(\mu_1^j, \mu_2^j, \mu_3^j\right)$
    # Multi-scale feature concatenation
    $O_{Reg}{}'$ = Concatenation ($I_{Vis}{}'$ + I_{IR} + $\S_j$)
    # Computation of loss
    L_{total}=($L_{membership}$, $L_{non-membership}$, $L_{hesitancy}$)
    # Backward propagation and update parameters
    $w = w - lr \times dl/dw$
  end for
  # Test model performance
end for

3.3. Design of Loss Function

This paper proposes a novel approach for loss function. During the training of the registration network, additional reference images are not used to calculate the loss of the model. The original multispectral and generated images are the main components that make up the loss. The similarity measure of the reference and the image to be registered is derived by calculating the inter-feature distance. The current research on distance is primarily concerned with the calculation of feature similarity like SSIM [34]. Furthermore, Intuitionistic Fuzzy Sets possess the capacity to describe dissimilarity. This paper employs loss functions to constrain the membership and non-membership between images, in accordance with the aforementioned concept. In calculating the distance between features, both similarity and dissimilarity are taken into account. The specifics of the loss design originate from [9].

Suppose the intuitionistic fuzzy set is $A = \{\langle x, \theta_A(x), \nu_A(x), \xi_A(x)\rangle \mid x \in X\}$, the membership $\theta_A(x)$, non-membership $\nu_A(x)$, and hesitancy $\xi_A(x)$ together form an ordered interval pair $(\theta_A(x), \nu_A(x), \xi_A(x))$ which is the IFS distance. $\theta_A(x) \in [0,1], \nu_A(x) \in [0,1]$, $\xi_A(x) \in [0,1]$, and $0 \leqslant \theta_A(x_j) + v_A(x_j) + \xi_A(x_j)) \leqslant 1$.

Intuitionistic fuzzy distance measures the distance between features, defined as follows: $I_{ij}(\delta(R_i, S_j), \bar{\delta}(R_i, S_j), \dot{\delta}(R_i, S_j))$ and $R_i \in R, S_j \in S$. $\delta(R_i, S_j)$ is the matching distance for R_i and S_j.

$$L_{membership} = \delta(R_i, S_j), \delta(R_i, S_j) = 1 - \left(\frac{1}{N}\sum_{1}^{k}\omega_k\left|\theta_{R_i} - \theta_{S_j}\right|\right) \tag{4}$$

where N is the dimension, ω is the normalised weighting factor, and $\sum_1^k \omega_k = 1$. Since the stability of images to be registered in the invariant moments is not uniform, it is possible to adjust the weighted coefficients for moments with large changes in amplitude. $\bar{\delta}(R_i, S_j)$ is the mismatch distance between feature R_i and S_j that is defined with respect to hesitancy as follows.

$$L_{non-membership} = \bar{\delta}(R_i, S_j), \bar{\delta}(R_i, S_j) = \max\{\min\{\nu_{R_i}, \nu_{S_i}\}\} \tag{5}$$

$$L_{hesitancy} = \dot{\delta}(R_i, S_j), \dot{\delta}(R_i, S_j) = \left(\frac{1}{N}\sum_{1}^{k}\omega_k\left|\theta_{R_i} - \theta_{S_j}\right|\right) - \max\{\min\{\nu_{R_i}, \nu_{S_i}\}\} \tag{6}$$

The total loss function is defined as follows:

$$L_{total} = \alpha L_{membership} + \beta L_{non-membership} + \gamma L_{hesitancy} \quad (7)$$

4. Experiments

4.1. Experiment Settings

The TNO [35] multispectral image dataset and an expanded version based on sample modification were utilised in this study. In order to expand the size of the training dataset, a manual method of overlapping cropping and multi-angle rotation was employed in the enhancement process. The final enhanced dataset for the experiments comprises a total of nearly 90,000 infrared and visible image pairs, with a batch size of 16. Figure 7 illustrates some of the enhanced data samples.The image pairs are randomly divided into three groups: the training set, the validation set, and the test set. The relevant data pertaining to the size and parameter information for each of the datasets can be found in Table 2.

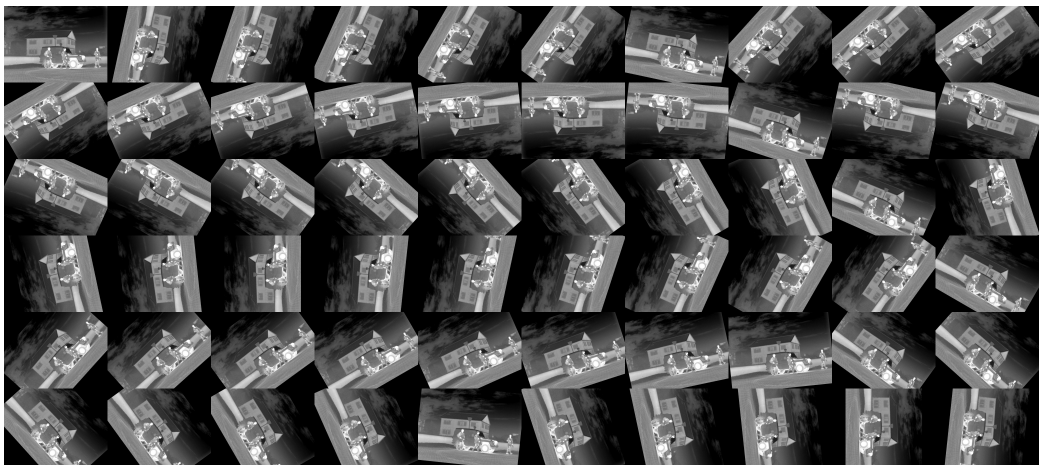


Figure 7. A preliminary presentation of the manually produced rotated dataset.

Table 2. A brief description of the datasets used in the study.

Source	Modality	Size	Train	Test	Resize
TNO Origins	Visible–IR	640 × 480	87	6	NO
TNO Enhancement	Visible–IR	512 × 512	89,646	24	YES

The pseudo-IR images input to the registration network are generated by the generative model in the initial stage. The comparison in Figure 8 shows that the trained generative model used in this paper can output high-quality pseudo-IR images, which provides for the later image registration.

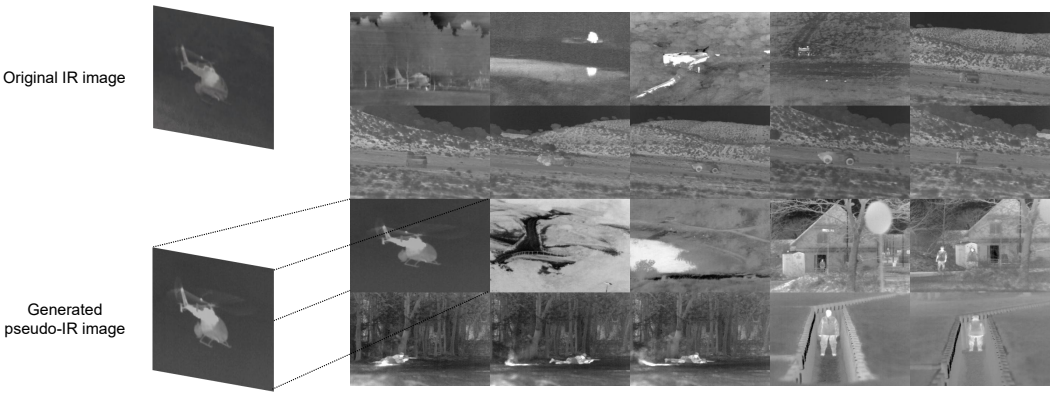


Figure 8. A portion of the pseudo-infrared image generated by the trained CycleGAN.

For the image-to-image translation network section, the batch size is 2. The entire CycleGAN model is optimised using the Adam optimiser [36], with the learning rates for the discriminator and generator set to 0.0004 and 0.0001, respectively. The model is trained for 200 epochs, with the learning rate remaining fixed for the first 100 epochs. Subsequently, from the 150th epoch onward, the learning rate gradually decays to 0. Additionally, for the registration network, the experiment trained the model for 100 epochs with a mini-batch size of 16. The study was conducted on a Windows 10 operating system, employing a desktop computer equipped with 32 GB memory, a Core i7-10700K CPU, and an NVIDIA RTX 3090 GPU. The experimental framework chosen wasTensorflow 2.12.0, Cuda 12.1, Python 3.11, and Cudnn 11.2.

To demonstrate the performance of the method proposed in this paper, the experiment compares its registration accuracy and the magnitude of network parameters against several existing registration methods, including DASC [37], DHN [38], MHN [39], NTG [40], SMILE [41], and MURF [42].

4.2. Evaluation Metrics

Since subjective evaluation information is highly influenced by individuals, quantitative metrics to evaluate the results of image registration are more objective and uniform. The following metrics mainly serve to evaluate the results of image registration: MAE, PSNR [43], Normalised Mutual Information (NMI) [44], SSIM, and Learned Perceptual Image Patch Similarity (LPIPS). MAE represents the mean of the absolute errors between the predicted and observed values. PSNR is an image quality reference value that measures the discrepancy between the maximum signal and the background noise. The greater the PSNR value, the less image distortion will be present. In this paper, NMI is employed as a metric to assess the accuracy of an image in comparison with the ground truth. The value range of NMI is between 0 and 1, with higher values denoting greater accuracy in image registration. SSIM is a quantitative measure used to assess the structural similarity between two images. SSIM values are expressed as a ratio between 0 and 1, with higher values indicating greater structural similarity. LPIPS, also known as perceptual loss, is used to measure the difference between two images. The metric facilitates the learning of the inverse mapping from a generated image to the ground truth, thereby compelling the generator to learn the inverse mapping from a reconstructed real image to a pseudo-image and to prioritise the perceived similarity between them. A lower value of LPIPS indicates that the two images are more similar to each other, and vice versa.

4.3. Experiment Results and Analysis

4.3.1. IFS-SIFT Validation

To compare the impact prior and subsequent to the integration of the multi-scale IFS-SIFT feature, the experiment employed an identical generator and discriminator with the same loss function as the optimisation target. All parameters and environmental settings were assigned identically. The results of the image registration are presented in Figure 9, where the disparities in registration effects are more intuitively displayed through image fusion.

(a) (b) (c) (d)

Figure 9. Fusion of the registration result with the reference. (**a**) shows the infrared image to be registered, (**b**) is the unguided registration result, (**c**) is the registration result achieved by multi-scale IFS-SIFT feature guidance, and (**d**) represents the reference.

The comparison shows that registration results guided by multi-scale IFS-SIFT features are closer to the ground truth. This is because multi-scale IFS-SIFT features enhance the neural network's feature representations and broaden its receptive field size to input. Fine-grained features at lower scales capture local details, while high-scale features grasp global semantic information. Additionally, the multi-scale IFS-SIFT features share weights among convolutional layers and reduce feature dimensions, resulting in a significant reduction in network parameters. This implies that the registration results remain unaffected even with reduced memory usage, thereby lowering the hardware requirements for the model.

4.3.2. Visual Comparison

The registration results of visible and IR images using DASC, DHN, MHN, NTG, SMILE, MURF, and the network proposed in this paper are presented in Figure 10. DASC employs DSC+GAN, which is the first successful use of GAN for unsupervised clustering. However, it achieves low accuracy in multispectral image registration due to the presence of projection residuals in the generative model subspace. DHN has a fast registration speed, which is achieved by obtaining an intermediate variable homography through the neural network.

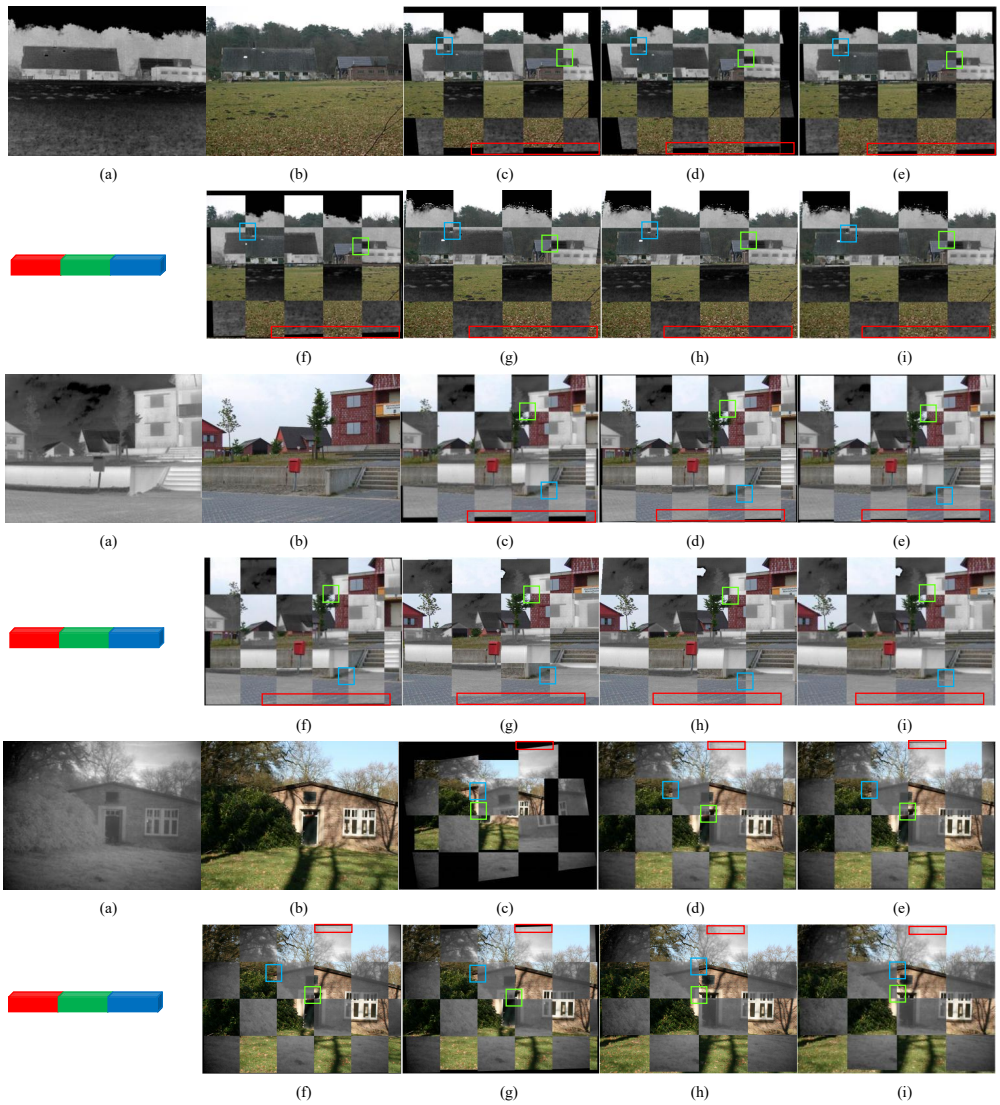


Figure 10. The figure shows from left to right (**a**) image to be registered; (**b**) reference; (**c**) DASC; (**d**) DHN; (**e**) MHN; (**f**) NTG; (**g**) SMILE; (**h**) MURF; (**i**) IFSrNet proposed in this paper. To concentrate on the region of interest, the lack of information on the edges of the image after registration is marked by the red boxes, and the detailed structural information is marked by the green and blue boxes.

This eliminates the need to separate feature point detection from transform estimation, as is carried out in traditional methods that use techniques such as ORB for corner detection and RANSAC [45] for matrix estimation. To preprocess the homography, denoising is added. However, this can result in the loss of feature information for noisy IR images, which can severely impact the registration accuracy. MHN is capable of processing large global motions and providing current single response matrix estimation results, making it more suitable for image registration tasks that involve dynamic scenes, blurred scenes, or lack of texture. The NTG method for multispectral image registration is based on the principle that the gradient of difference image is sparsest when two images are perfectly registered. However, it may not be stable enough when dealing with dynamic and non-rigid objects. The SMILE method typically necessitates an initial registration estimate as a point of departure. The initial estimate is inaccurate, which leads to subsequent processing steps that deviate in the correct direction to the extent that the results of each experiment differ significantly. The MURF algorithm employs a coarse-to-fine registration strategy, whereby both global rigid and local non-rigid transformations are considered. However, the registration network may become less accurate if the input image contains serious noise, artefacts, or distortion. To display the registration results, they are crossed in a checkerboard diagram. As shown in the figure, the method proposed in this paper achieves superior or comparable registration accuracy to other methods.

4.3.3. Quantitative Assessments

Table 3 presents the mean performance value of seven different algorithms on the same test set for each metric, allowing for quantitative analysis of registration. The best result for each evaluation metric is highlighted in red, while the second best is highlighted in blue.

Table 3. This study tested the average MAE, SSIM, PSNR, NMI, and LPIPS values obtained from 30 sets of multispectral images. The best performance is marked in red font, and blue is the second best.

Metric	DASC	DHN	MHN	NTG	SMILE	MURF	IFSrNet
MAE	77.6754	75.5950	71.1627	71.2080	73.2551	68.0920	67.4420
SSIM	0.5608	0.4071	0.5702	0.6826	0.6642	0.7093	0.7201
PSNR	56.9867	56.0058	57.0732	55.0581	56.0154	56.8961	57.5722
NMI	0.1708	0.2465	0.2717	0.2487	0.2612	0.3176	0.3092
LPIPS	0.3102	0.3294	0.3601	0.3768	0.3200	0.3749	0.3794

The study tested thirty sets of multispectral images using five registration algorithms and four evaluation metrics. Figure 11a,b show that DASC has the highest MAE, indicating that it is not the best fit for the dataset. On the other hand, IFSrNet demonstrates superior MAE and SSIM values compared to the other algorithms, indicating its superiority in terms of registration accuracy on this dataset. Figure 11c shows that the proposed algorithm records slightly lower than the others at individual points.

Noise interference in the image pairs, particularly those with intricate modal features, during the encoding process may account for this difference. However, our proposed algorithm, as shown in Figure 11d, successfully preserves the intricate details of the image by using multi-scale IFS feature guidance, which enhances the robustness of image registration. It is worth noting that this approach achieves the highest NMI. Furthermore, IFSrNet also demonstrates superior performance in the comparison of inter-image perceptual similarity, as illustrated in Figure 11e, which is comparable to MURF. In contrast to traditional methods, LPIPS is more aligned with the human perceptual situation.

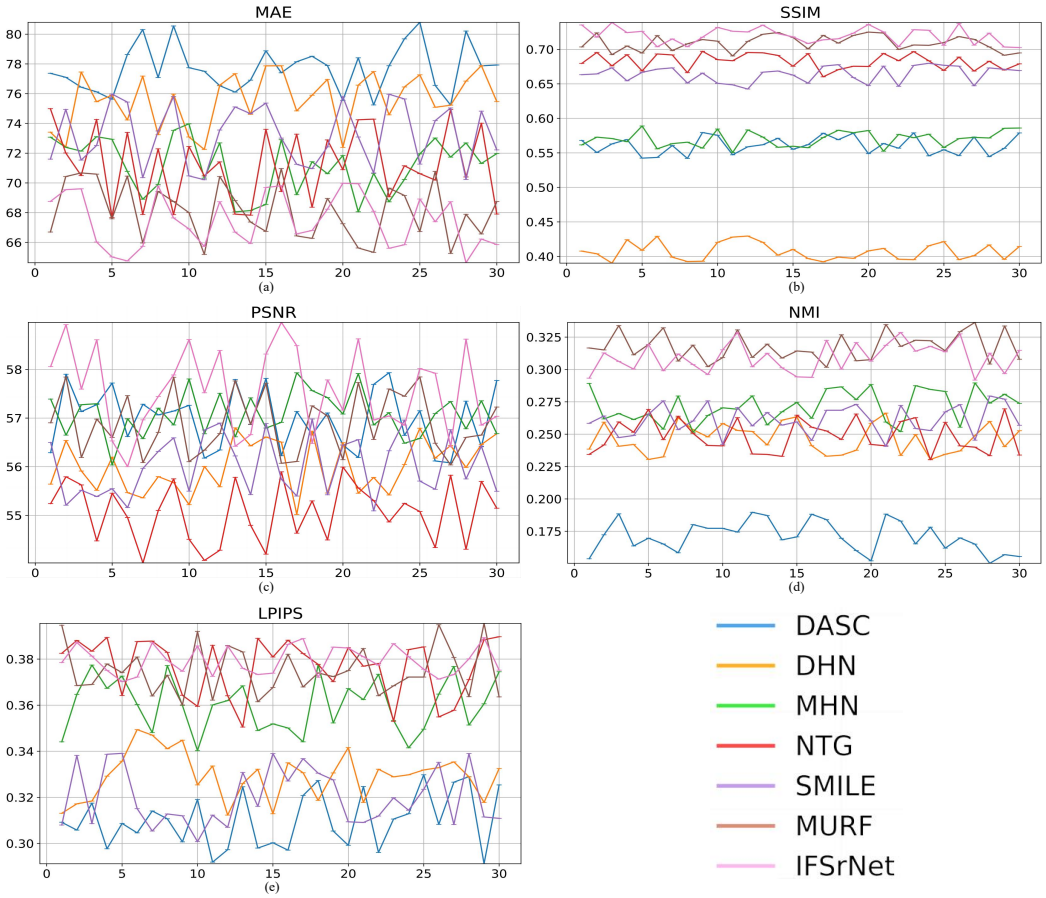


Figure 11. Experimental results on real datasets. Shown in order is the performance in each evaluation metric, (**a**) MAE, (**b**) SSIM, (**c**) PSNR, (**d**) NMI, and (**e**) LPIPS. The horizontal axis of the graph represents the number of each sample, while the vertical axis represents the value obtained. The various algorithms have been identified with different colours for differentiation purposes.

The efficiency of the models is evaluated, with particular attention paid to the parameter size and run speed. The experiments ensure that all models are run in the same environment and with the same equipment. Table 4 shows that IFSrNet has an advantage in network parameters but an average performance in inference time. If the network is equipped with GPU acceleration, there is a potential for much higher computational efficiency.

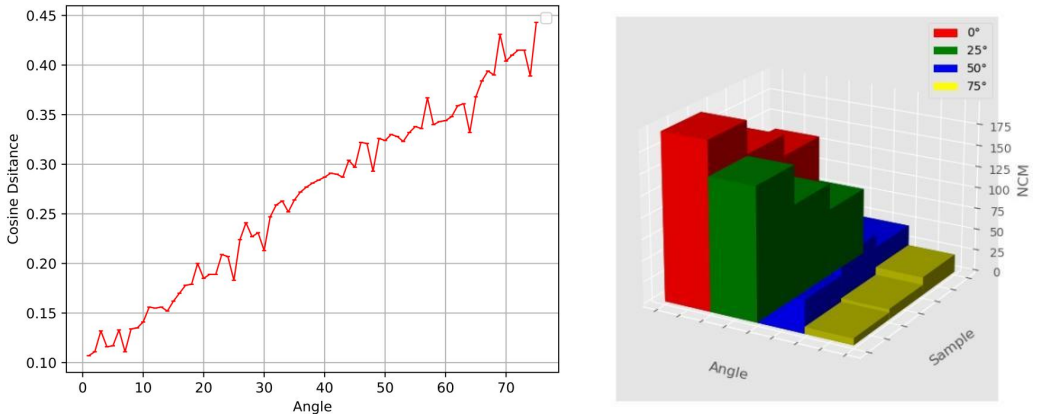
Table 4. Number of parameters and inference time for IFSrNet and other registration algorithms, where the parameters and times are in M and seconds, respectively. The red font parameter represents the optimal performance, while the blue font is just below it.

Efficiency	DASC	DHN	MHN	NTG	SMILE	MURF	IFSrNet
Parameters	17.35 M	35.26 M	3.18 M	20.38 M	13.61 M	25.21 M	12.16 M
Times	1.283 s	0.005 s	0.014 s	0.022s	2.196 s	0.429 s	0.585 s

4.3.4. Rotation Robust Experiment

To determine the effect of image training with different rotation angles on the registration network's performance, three sets of images were randomly selected from the dataset. Each set of images was rotated in 25° intervals from 0° to 75°, resulting in three rotated images per group. Nine new IFS feature matching maps were generated and used to evaluate the algorithm's robustness under rotated conditions. Simultaneously, the image to be registered was rotated in 1° steps until it was rotated to 75°. The similarity index was calculated as the cosine distance between the rotated image and the reference features. The

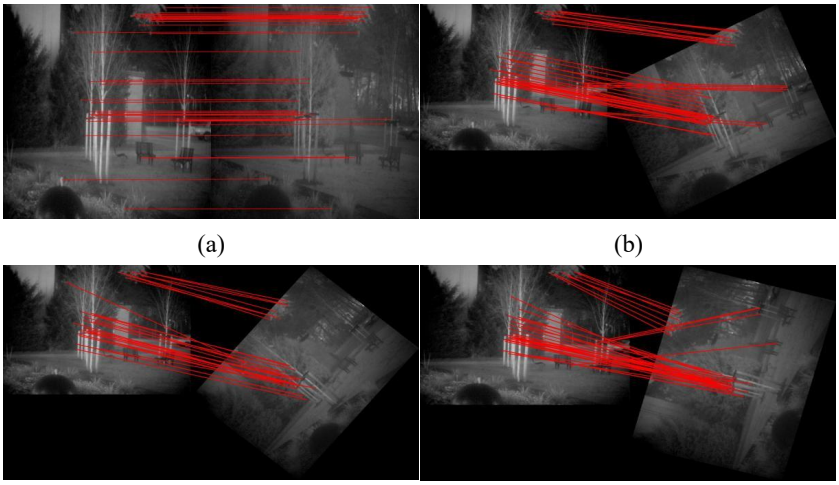
experimental results are presented in Figure 12a, while Figure 12b shows the number of positive matches (NPM) of the four groups of rotated images.



(a) (b)

Figure 12. Quantitative statistics of rotational invariance. (**a**) shows the average feature cosine distances for different rotation angles, and (**b**) demonstrates the number of correctly matched features for the four rotation angles in the three datasets.

The result of matching on instances as illustrated in Figure 13. The accuracy and NPM are contingent upon the specific distortion angle of the floating image. Therefore, if there are variations in the angle of distortion, the resulting accuracy and NPM will also vary. The analysis indicates that the algorithm maintains good feature similarity for a range of angles due to the network learning the invariance of small rotations during training, which is inseparable from the rotational invariance of SIFT feature descriptors. However, for rotation angles exceeding 25°, the NPM exhibits a substantial decrease and the cosine distance between features increases significantly.



(a) (b)

(c) (d)

Figure 13. Example of feature matching of two images to be processed, where the left is a reference and the right is a rotated image to be registered. The matching result is shown in (**a**) with 0° of rotation, (**b**) shows the result with 25° of rotation, and (**c**,**d**) show the results with 50° and 75° of rotation, respectively.

4.3.5. Ablation Study

To assess the impact of the loss function on the registration network's performance, this paper conducted an ablation study using the traditional similarity metrics SSIM, loss

of single membership, and loss of IFS. As a comparison, SSIM loss [46] $L_{spatial}$ is defined as follows:

$$\mathrm{SSIM}(h,p) = \left(\frac{2\mu_h\mu_p + C_1}{\mu_h^2 + \mu_p^2 + C_1} \cdot \frac{2\sigma_{hp} + C_2}{\sigma_h^2 + \sigma_p^2 + C_2}\right) \tag{8}$$

$$L \text{ spatial } (h,p) = \frac{1}{n}\sum_{i=0}^{n}\left(1 - \frac{2\mu_h\mu_p + C_1}{\mu_h^2 + \mu_p^2 + C_1} \cdot \frac{2\sigma_{hp} + C_2}{\sigma_h^2 + \sigma_p^2 + C_2}\right) \tag{9}$$

where h represents the high resolution multispectral image obtained by the network, p represents the original panchromatic image, n represents the number of pixels in the image block, μ_h represents the mean of h, μ_p represents the mean of p, σ_h^2 is the variance of h, σ_p^2 is the variance of p, and σ_{hp} is the covariance of h and p. $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are two constants to avoid dividing by 0, L is the range of pixel values, and $k_1 = 0.01$ and $k_2 = 0.03$ are default values. Since the SSIM value range is from -1 to 1, the closer the SSIM value of the two images is to 1, the more similar the images are. And the smaller the value of the loss function in the network the better; the optimisation of the network is also in the direction of smaller loss values, so the texture loss of the image is calculated using $1 - \mathrm{SSIM}(h,p)$.

Table 5 presents the performance of three loss functions: $L_{spatial}$, $L_{membership}$, and L_{Total}. The results indicate that, although the loss of membership outperforms SSIM in terms of similarity, there is potential for improvement. When only $L_{spatial}$ is used, the registration network's NMI falls below 0.2. This highlights the difficulty in capturing the correlation between multispectral paired images when relying solely on the structural similarity loss. The optimal solution to this problem is to implement a complete IFS loss. The data presented in red font indicate that L_{Total} performs the best on all evaluation measures.

Table 5. An ablation study on the IFSrNet loss function. Values marked in red represent the best performance.

Metric	$L_{spatial}$	$L_{membership}$	L_{Total}
MAE	78.6194	75.3547	67.4420
SSIM	0.6035	0.6349	0.7201
PSNR	49.9183	51.1351	57.5722
NMI	0.1961	0.2392	0.3092
LPIPS	0.2112	0.2830	0.3794

The transformation of multispectral image modalities is challenging in the absence of labelling. In practice, CycleGAN is a suitable approach for transformation tasks involving images with texture and colour variations. However, the estimated spatial error of GAN and CycleGAN cannot exclude the effect of residual distribution between the moving and target image. In contrast, CycleGAN embedded with a multi-head attention mechanism can circumvent the larger spatial error. In order to demonstrate the potential of the combination of CycleGAN and multi-head attention for the transformation of multispectral images, this paper explored the correlation between the model output and the Dice coefficient. The mean of the model output was computed based on the joint region of the moving and target masks. Subsequently, the mean absolute value was subtracted from 1 to obtain a normalised value indicative of the performance of the model generation. Similarly, the generative performance of both the GAN and the CycleGAN models were calculated. Figure 14 provides clear evidence of the positive correlation between each model and Dice, thereby confirming the potential of the combination of CycleGAN and multi-head attention to accurately transform the image modality. The GAN and CycleGAN have no significant positive correlation with Dice.

In order to ascertain the impact of the training set employed in the model on the registration network, three datasets were established for the purpose of separate experiments. As shown in Table 6, the mean number of matches for 286.8 is the highest of the

three strategies. The single TNO native image set, the single manufactured images after enhancement, and the hybrid dataset are presented. The hybrid training set not only solves the limitation of the insufficient size of the native dataset, but also ensures the number of features that can be matched.

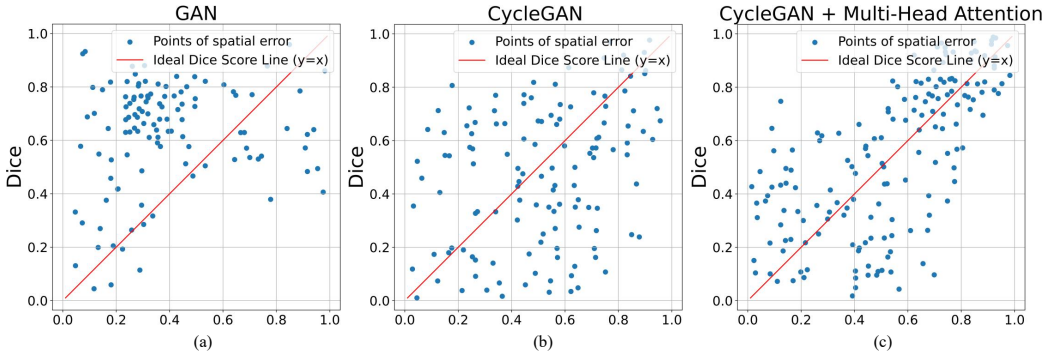


Figure 14. Correlation of estimated spatial errors with Dice in three generation strategies. The distribution of spatial error points in (**a**,**b**) does not exhibit a strong positive correlation with the Dice coefficient. In contrast, the generative model employed in this study, as illustrated in (**c**), exhibits a clear positive correlation.

Table 6. In the initial training phase of the model, ablation experiments were conducted with different training sets, with the average number of matched pairs serving as the standard. The letters ✓ and × are used to indicate whether the dataset was or was not included in the analysis.

Training Strategy	TNO	Artificial Enhancement	Number of Matches
1	✓	×	154.6
2	×	✓	81.9
3	✓	✓	286.8

In the field of image registration, research on multi-scale feature guidance has concentrated on gradient-based edge detection for features at each stage. In order to analyse the effectiveness of the multi-scale IFS feature map guidance, the experiment involved the alteration of the type of multi-scale features without modifying the skeleton of the registration network. The application of deep learning to the task of edge detection enables the DexNet and BDCN models to learn the singularity of edge information. Nevertheless, the failure to recognise weak edges may result in an uneven detection of features. The results of the ablation can be found in Table 7. Sobel edge mapping describes changes in gradient that retains more edge information than Canny and Laplace edge mapping. IFS feature descriptors benefit from advantages in dealing with biases caused by the generative model, and the efficiency gap with Sobel is closer.

Table 7. Registration results of ablation studies using different feature map extraction methods. The red font parameter represents the optimal performance, while the blue font is just below it.

Method	Precision	Recall	F1-score	Summation of TP
Canny	89.19	92.10	90.62	8051
Laplacian	88.93	87.58	88.25	7258
Sobel	88.08	96.11	91.93	18,521
BDCN	85.18	96.61	90.54	11,731
DexiNed	90.03	89.32	89.67	4883
IFS-SIFT	88.16	96.12	91.96	7202

The consistency between the matches identified by the registration method and the set of ground truth matches is verified, and pairs of true positive (TP), false positive (FP), and false negative (FN) matches are defined to calculate the precision and recall scores. The sum of TP and FN denotes all correct matches from putative feature points. Consequently, the values of precision and recall can be calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

The larger precision and recall are, the more accurate the feature matching and the stronger the adaptability of the registration method to distinguish feature points, respectively. F1-score is the harmonic mean and summary statistic of precision and recall, which can be calculated as follows:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

The total summation of TP points from all image pairs in the construed datasets is calculated as the number of correct matches, which is denoted as Summation of TP [34].

5. Applications

The feasibility of applying IFSrNet to other multimodal image registration tasks has been tried in medical images [47]. Chemical exchange saturation transfer (CEST) is a magnetic resonance imaging (MRI) technique for enhancing the contrast of images, which indirectly identifies the metabolites in tissue at millimolar concentrations through the water proton signal. Because the samples must have a sufficient saturation frequency, a large time span is usually required to acquire the spectrum. It is important to note that subject movement throughout the scan can lead to errors in CEST quantification. Even minor movements can have a significant impact on CEST analysis, resulting in the appearance of unusual peaks or dips in the spectrum and an uneven signal distribution on the image. To mitigate motion artefacts, image registration is a commonly employed method to ensure high-quality CEST-MRI images. Figure 15 shows the application of IFSrNet for registering T1 and Gd image pairs in CEST-MRI images of a rat brain. The lack of an objective gold standard for assessing the registration results of medical images means that the physician's judgement is crucial. After observation by senior experts, IFSrNet was able to accurately reconstruct the target images, which illustrates the potential of the network's application. All the CEST-MRI images in this experiment were obtained from Johns Hopkins University.

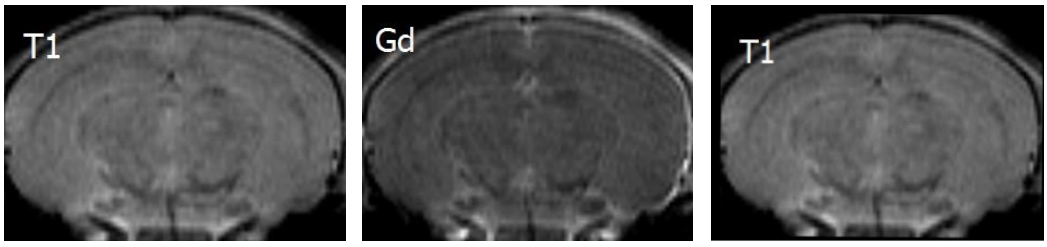


Figure 15. CEST-MRI images for a rat stroke lesion in which (**a**) is T1 as the image registration to be registered, (**b**) is Gd as the reference, and (**c**) is the T1 image that has been registered.

6. Discussion

The image-to-image translation registration strategy employs a novel approach which effectively circumvents the complex multimodal issue while simultaneously reducing the difficulty of registration. The approach proposed in this paper requires the preparation of multimodal data for the training of the generator. When encountering previously unknown data, the performance of the trained model will inevitably decline, although the present

network has been designed to mitigate this effect. Furthermore, the method in this paper introduces a small amount of bias and variance in the process of remapping the data distribution between different modalities, which requires subsequent work to optimise these negative effects. In future work, the authors will endeavour to extend this study to other modal image registration tasks beyond multispectral images. Additionally, the lightweight improvement of the network is a promising avenue for further investigation.

The precise registration of images constitutes a fundamental precondition for multispectral image processing. The related extended works encompass the fusion of image information, the localisation of targets, and the detection of changes, along with the reconstruction of high-resolution images. The procedure of multispectral image fusion is expedited by image registration, which allows for the organic combination of the advantages or complementarities of the information comprised in each image dataset, thereby giving rise to the production of a more comprehensive and accurate image. Furthermore, the image registration is also more conducive to change detection, which enhances the accuracy of target positioning and discovers alterations in features or the ground surface through comparing the image disparities at different times or under different circumstances in remote sensing operations. Within the domain of high-resolution image reconstruction, the aim is to coalesce multiple low-resolution images into a sole high-resolution image by means of registration, thereby enhancing the pellucidity and detailed manifestation of the image.

7. Conclusions

In this paper, a Multi-scale IFS Feature-guided Registration Network Using Multispectral Image-to-image Translation is proposed. To tackle the challenge of matching infrared images with visible images during image registration, a pseudo-infrared image is generated from visible images using CycleGAN equipped with a multi-head attention module. The generative models can be interchanged between the two modalities due to the bidirectional feedback mechanism of CycleGAN, which allows for the transfer of information between the two domains. This approach reduces the difficulty in modal feature extraction by addressing the large modal differences between the two types of images. To avoid the problem that the generated images cannot achieve pixel-by-pixel correspondence with the ground truth, the feature vectors of the reference are first extracted by the improved robust IFS-SIFT feature descriptor in the case of scale transformation. Secondly, this paper establishes an end-to-end registration network model and designs a loss function that incorporates multi-scale feature guidance. The experiments demonstrate that modal transformed infrared spectral information is effective in extracting both structural and textural features from the image. IFS-SIFT demonstrates superior performance in extracting multi-scale feature vectors, and the established registration model is robust in estimating the transformation despite the presence of discrete points in the reference. The proposed model primarily focuses on addressing the challenges posed by significant scale differences between infrared and visible images, as well as the intricacies involved in image registration, in comparison to other algorithms. In the future, this work can be further extended to encompass tasks such as image fusion and image enhancement with GAN.

Author Contributions: This study's topic and problem were identified by B.C. who also designed the network architecture and research methodology, collected and analysed the data, completed the experiments, and wrote and significantly revised the paper. L.C. provided financial support and significantly revised the academic content. U.K. made grammatical corrections to the academic content's English. S.Z. provided support for the collection and organisation of the datasets. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shaanxi Provincial Key Research and Development Programme (No. 2024GX-YBXM-555).

Data Availability Statement: The details and code related to the content discussed in this study can be obtained by contacting the first author.

Acknowledgments: We would like to express our gratitude to the authors of DASC, DHN, MHN, and NTG for sharing their code. This greatly assisted us in comparing our experiments. We also thank the anonymous reviewers for their insightful and valuable suggestions, which undoubtedly improved the quality of our paper.

Conflicts of Interest: Author Shuai Zhang was employed by the company Dongdian Testing Technology Xi'an Corporation. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Sun, W.; Gao, H.; Li, C. A Two-Stage Registration Strategy for Thermal–Visible Images in Substations. *Appl. Sci.* **2024**, *14*, 1158 [CrossRef]
2. Yang, A.; Yang, T.; Zhao, X.; Zhang, X.; Yan, Y.; Jiao, C. DTR-GAN: An Unsupervised Bidirectional Translation Generative Adversarial Network for MRI-CT Registration. *Appl. Sci.* **2024**, *14*, 95. [CrossRef]
3. Peng, B.; Ren, D.; Zheng, C.; Lu, A. TRDet: Two-Stage Rotated Detection of Rural Buildings in Remote Sensing Images. *Remote Sens.* **2022**, *14*, 522. [CrossRef]
4. Caseiro, A.; Rücker, G.; Tiemann, J.; Leimbach, D.; Lorenz, E.; Frauenberger, O.; Kaiser, J.W. Persistent Hot Spot Detection and Characterisation Using SLSTR. *Remote Sens.* **2018**, *10*, 1118. [CrossRef]
5. Stempliuk, S.; Menotti, D. Agriculture Multispectral Uav Image Registration Using Salient Features and Mutual Information. In Proceedings of the IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium,Waikoloa, HI, USA, 26 September–2 October 2020.
6. Yang, S.; Sun, M.; Lou, X.; Yang, H.; Liu, D. Nighttime Thermal Infrared Image Translation Integrating Visible Images. *Remote Sens.* **2024**, *16*, 666. [CrossRef]
7. Li, L.; Liu, L.; He, Y.; Zhong, Z. USES-Net: An Infrared Dim and Small Target Detection Network with Embedded Knowledge Priors. *Electronics* **2024**, *13*, 1400. [CrossRef]
8. Sun, Y.; Yan, K.; Li, W. CycleGAN-Based SAR-Optical Image Fusion for Target Recognition. *Remote Sens.* **2023**, *15*, 5569. [CrossRef]
9. Luo, M.; Li, W.; Shi, H. The Relationship between Fuzzy Reasoning Methods Based on Intuitionistic Fuzzy Sets and Interval-Valued Fuzzy Sets. *Axioms* **2022**, *11*, 419. [CrossRef]
10. Zhou, K.; Zhang, M.; Wang, H.; Tan, J. Ship Detection in SAR Images Based on Multi-Scale Feature Extraction and Adaptive Feature Fusion. *Remote Sens.* **2022**, *14*, 755. [CrossRef]
11. Shivajirao, S.; Hantach, R.; Abbes, S.B.; Calvez, P. Mask R-CNN End-to-End Text Detection and Recognition. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA),Boca Raton, FL, USA, 16–19 December 2019.
12. Nunes, C.F.G.; Pádua, F.L.C. A local feature descriptor based on log-gabor filters for keypoint matching in multispectral images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1850–1854. [CrossRef]
13. Gao, C.; Li, W.; Tao, R.; Du, Q. MS-HLMO: Multiscale histogram of local main orientation for remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*,5626714. [CrossRef]
14. Xu, H.; Ma, J.; Yuan, J.; Le, Z.; Liu, W. RFNet: Unsupervised network for mutually reinforcing multi-modal image registration and fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR,New Orleans, LA, USA, 18–24 June 2022; pp. 19679–19688.
15. Wei, Z.; Jung , C.; Su, C. RegiNet: Gradient guided multispectral image registration using convolutional neural networks. *Neurocomputing* **2020**, *415*, 193–200. [CrossRef]
16. Zhang, Y.; Yao, Y.; Wan, Y.; Liu, W.; Yang, W.; Zheng, Z.; Xiao, R. Histogram of the orientation of the weighted phase descriptor for multi-modal remote sensing image matching. *ISPRS J. Photogramm. Remote Sens.* **2023**, *196*, 1–15. [CrossRef]
17. Li, Y.; Tang, S.; Zhang, R.; Zhang, Y.; Li, J.; Yan, S. Asymmetric GAN for Unpaired Image-to-image Translation. *IEEE Trans. Image Process.* **2019**, *28*,5881–5891. [CrossRef]
18. Ye, Y.; Tang, T.; Zhu, B.; Yang, C.; Li, B.; Hao, S. A multiscale framework with unsupervised learning for remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*,5622215. [CrossRef]
19. Wang, D.; Liu, J.; Fan, X.; Liu, R. Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration.*arXiv* **2022**, arXiv:2205.11876.
20. Tsourounis, D.; Kastaniotis, D.; Theoharatos, C.; Kazantzidis, A.; Economou, G. SIFT-CNN: When Convolutional Neural Networks Meet Dense SIFT Descriptors for Image and Sequence Classification. *J. Imaging* **2022**, *8*, 256. [CrossRef]
21. Juan, L.; Gwun, O. A Comparison of SIFT, PCA-SIFT and SURF. *Int. J. Image Process.* **2009**,*3*, 143–152.
22. Gao, C.; Li, W. Multi-Scale HARRIS-PIIFD Features for Registration of Visible and Infrared Images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 5437–5440.
23. Kumari, V.; Saxena, M.; Gupta, M. Reliability Comparison of Conventional & Two Finger AlGaN/GaN HEMT. In Proceedings of the 2019 IEEE MTT-S International Microwave and RF Conference (IMARC),Mumbai, India, 13–15 December 2019.
24. Mao, Y.; Xiao, D.; Cheng, J.; Che, D.; Le B.; Song, L.; Liu, X. Multigrades Classification Model of Magnesite Ore Based on SAE and ELM. *J. Sens.* **2017**,*2017*, 9846181. [CrossRef]

25. Yang, B.; Wang, P.; Li, X.; Li, L.; Cao, X Infrared and visible image registration based on modal transformation combined with robust features. *Adv. Lasers Optoelectron.* **2022**, *59*, 180–189.
26. Wu, J.; Liu, N.; Li, X.; Fan, Q.; Li, Z.; Shang, J.; Wang, F.; Chen, B.; Shen, Y.; Cao, P. Convolutional neural network for detecting rib fractures on chest radiographs: A feasibility study. *BMC Med. Imaging* **2023**, *23*, 18. [CrossRef]
27. Zhao, P.; Chen, B.; Fan, X.; Chen, H.; Zhang, Y. Image Inpainting with Parallel Decoding Structure for Future Internet. *Electronics* **2023**, *12*, 1872. [CrossRef]
28. Battiato, S.; Gallo, G.; Puglisi, G.; Scellato, S. SIFT Features Tracking for Video Stabilization. In Proceedings of the 14th International Conference on Image Analysis and Processing (ICIAP 2007),Modena, Italy, 10–14 September 2007.
29. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV),Venice, Italy, 22–29 October 2017; Volume 244.
30. Shah, C.; Du, Q.; Xu, Y. Enhanced TabNet: Attentive Interpretable Tabular Learning for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 716. [CrossRef]
31. Qing, Y.; Huang, Q.; Feng, L.; Qi, Y.; Liu, W. Multiscale Feature Fusion Network Incorporating 3D Self-Attention for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 742. [CrossRef]
32. Chen, G.; Shang, Y. Transformer for Tree Counting in Aerial Images. *Remote Sens.* **2022**, *14*, 476. [CrossRef]
33. Yin, L.; Gao, W.; Liu, J. Deep Convolutional Dictionary Learning Denoising Method Based on Distributed Image Patches.*Electronics* **2024**, *13*, 1266. [CrossRef]
34. Fan, Z.; Sohail, S.; Sabrina, F.; Gu, X. Sampling-Based Machine Learning Models for Intrusion Detection in Imbalanced Dataset. *Electronics* **2024**, *13*, 1878. [CrossRef]
35. Available online: https://figshare.com/articles/dataset/TNOImageFusionDataset/1008029 (accessed on 18 August 2023)
36. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization.*arXiv* **2014**, arXiv:1412.6980.
37. Kim, S.; Min, D.; Ham, B.; Ryu, S.; Do, M.N.; Sohn, K. DASC: Dense adaptive self-correlation descriptor for multi-modal and multi-spectral correspondence. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
38. Zhu, H.; Long, M.; Wang, J.; Cao, Y. Deep Hashing Network for efficient similarity retrieval. InProceedings of the AAAI conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp.2415–2421.
39. Chen, B.; Deng, W.; Hu, J. Mixed High-Order Attention Network for Person Re-Identification. InProceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; Volume 46.
40. Chen, S.-J.; Shen, H.-L.; Li, C.; Xin, J.H. Normalized Total Gradient: A New Measure for Multispectral Image Registration. *IEEE Trans. Image Process.* **2018**, *27*, 1297–1310. [CrossRef]
41. Gao, X.; Zheng, G. SMILE: Siamese Multi-scale Interactive-representation LEarning for Hierarchical Diffeomorphic Deformable image registration. *Comput. Med. Imaging Graph.* **2024**, *111*,102322. [CrossRef]
42. Xu, H.; Yuan, J.; Ma, J. MURF: Mutually Reinforcing Multi-Modal Image Registration and Fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 12148–12166. [CrossRef]
43. Sun, Z.; Lei, Y.; Wu, X. Chinese Ancient Paintings Inpainting Based on Edge Guidance and Multi-Scale Residual Blocks. *Electronics* **2024**, *13*, 1212. [CrossRef]
44. Amelio, A.; Pizzuti, C. Correction for Closeness: Adjusting Normalized Mutual Information Measure for Clustering Comparison. *Comput. Intell.* **2016**, *33*, 579–601. [CrossRef]
45. He, Z.; Shen, C.; Wang, Q.; Zhao, X.; Jiang, H. Mismatching Removal for Feature-Point Matching Based on Triangular Topology Probability Sampling Consensus. *Remote Sens.* **2022**, *14*, 706. [CrossRef]
46. Zhang, L.; Wang, X.; Rawson, M.; Balan, R.; Herskovits, E.H.; Melhem, E.R.; Chang, L.; Wang, Z.; Ernst, T. Motion Correction for Brain MRI Using Deep Learning and a Novel Hybrid Loss Function. *Algorithms* **2024**, *17*, 215. [CrossRef]
47. Liang, Y.; Bie, C.; Chen, B.; Hou, Y.; Song, X. Motion correction in CEST MRI series exploiting Adaptive Stochastic Gradient Descent (ASGD)—Based optimization algorithm.In Proceedings of the 2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE), Shenzhen, China, 22–24 November 2019.