International Journal for Housing Science and Its Applications
Publish August 5, 2025. Volume 46, Issue 3 Pages 4154-4163

https://doi.org/10.70517/ijhsa463355

Image pyramid-based multi-scale image alignment technique for computer vision applications

Huaijiang Teng[1] and Zhenbo Zhang[1,*]
[1] Heilongjiang Open University, Harbin, Heilongjiang, 150080, China
Corresponding authors: (e-mail: 15546202925@163.com).

Abstract Aiming at the problem of limited generalization ability of U-Net network to images of different scales and resolutions in image segmentation tasks. In this paper, a multi-scale feature extraction convolutional block is integrated into U-Net to enable the model to simultaneously consider image information at different scales, thus obtaining a more comprehensive and rich feature representation. In order to alleviate the limitation of the influence of the convolution kernel size in the convolution operation, a self-attention mechanism is embedded on the basis of the multi-scale feature learning image alignment network to cross-fertilize and match the feature representations of different images to form a new UNet backbone network, i.e., the LK-CAUNet model. Segmentation effects of different modal training on image alignment techniques are analyzed using T1, T2, T1ce, Flair single modality and a combination of all four modalities used simultaneously. The segmentation performance of LK-CAUNet model is analyzed on the dataset. The DSC metrics of LK-CAUNet model under the combined training of T1+T2+T1ce+Flair modalities are WT=93.56%, TC=89.42%, and ET=83.22% respectively.

Index Terms u-net network, multi-scale features, image segmentation, image alignment technique

I. Introduction

Image as a very important medium and means has received more and more attention in today's time. When researching and analyzing in the field of image processing, it is usually necessary to read the images by traditional means, such as box mounting devices or printing images on film [1], [2]. However, due to various differences or variations, it is difficult for researchers to synthesize the acquired images completely and accurately [3], [4]. Image alignment, on the other hand, can take advantage of computer image processing techniques so that these images can be unified into a common coordinate system, aided by computer visibility [5]-[7].

Image alignment aims to establish the spatial transformation relationship between two images, and to realize the alignment of two image regions by determining the parameters of the geometric transformation model and converting the coordinates of the image to be aligned into the coordinates of the reference image [8]. Using image alignment techniques, one can combine images collected from different sensors, or one can obtain variations between images under different conditions and at different times, or one can collect information in three dimensions and template-based patterns from images of moving objects [9]-[12]. However, digital image alignment can be limited by hardware devices such as the acquisition of digital images, so very often only one sensor or several sensors can be used to obtain information or a part of the picture of the same object or a part of the same scene [13]-[15]. Since there is a certain degree of connectivity between this data information, multi-scale image alignment techniques can be designed to obtain richer and more comprehensive information containing more detailed information [16]-[18]. Nowadays, image alignment techniques are also used in computer vision and other fields, and also play an important role in practical applications such as image stitching, image analysis, data fusion, target change recognition and detection [19], [20].

This paper analyzes the proposal and development of image pyramid, including subsampling pyramid, Gaussian pyramid, and the up-sampling and down-sampling process of image pyramid. Propose U-Net network, for the shortcomings of U-Net network in medical image segmentation task which is easy to be interfered by noise and has limited generalization ability, propose a new type of U-Net backbone network, i.e., LK-CAUNet.The LK-CAUNet model adds cross-attention module and loss function part based on deformable image alignment network based on multiscale feature learning. Preprocess the image data and analyze the effect of single modality and combined modality on the segmentation performance of LK-CAUNet model. Test the performance of LK-CAUNet model using Dice coefficient, sensitivity and positive predictive value indexes, and analyze the DICE score curve and MSE curve.

II. Image alignment techniques proposed

II. A.Image Pyramid

Image pyramid is a model for multi-scale feature representation of digital images. The simplest kind of image pyramid is the subsampling pyramid, which can be obtained by sampling each layer of image to the previous layer at equal intervals along both row and column directions [21].

The subsampling pyramid, although fast to construct, will reduce the accuracy due to too much variation in scale. In order to solve this problem it is necessary to introduce a filter, the filter can not be satisfied in the filtering process to add new extreme points, Gaussian filter can satisfy this condition, Gaussian filter is a low-pass filter, can remove the low-frequency components, and play the role of image smoothing.

Gaussian pyramid is used to smooth the image using a proper smoothingfilter, followed by downsampling the image after smoothing. Continue to do the same for the resulting resultant image, repeating this step several times, with each cycle resulting in a smaller, smoother, lower resolution image. If the original image is placed at the bottom and each subsequent resulting image is stacked on top of the previous result, the shape of a pyramid is formed. The Laplace pyramid is obtained by differencing two adjacent layers of images in the pyramid. Since the size of the two neighboring layers is different, some image interpolation operations are needed to calculate the difference between the pixels. Each layer in the Laplace pyramid is obtained by interpolating the corresponding layer of the Gaussian pyramid as well as the higher layer to enlarge the difference.

The image pyramid upsampling and downsampling process is shown in Fig. 1, the image pyramid contains 4 layers of images, and the collection of these 4 layers of images is compared to a pyramid. It can be acquired through gradient down acquisition until some termination requirements are met before stopping the acquisition, but in the down acquisition, the higher the level, the smaller the pixels and the lower the clarity.

There are two main ways to generate an image pyramid: down sampling and up sampling. Down sampling: the process of converting an image from G0 to G1, G2, G3, with further degradation of image clarity. Upward sampling: the image is transformed from G3 to G2, G1, G0, the process of improving the clarity of the image.

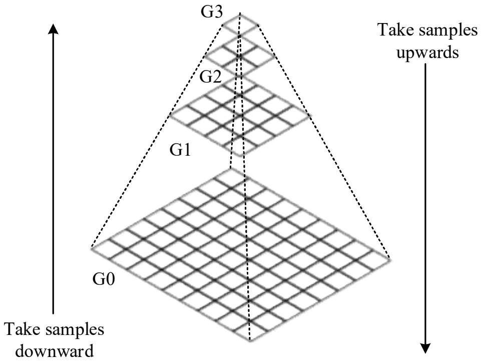


Figure 1: Image pyramid sampling and sampling process

II. A. 1) Downward sampling

In image down sampling, there are generally two steps:

The first step performs a Gaussian convolution kernel (Gaussian filtering) on the image.

The second step removes all even rows and columns.

Here, the Gaussian kernel convolution algorithm (Gaussian filtering) is simply a step of weighted averaging of the entire image, where each pixel value is derived from a weighted average of itself and the rest of the neighboring pixel values in the neighborhood (with proportional differences in weights). Typical 5*5 Gaussian kernels include:

$$K(5,5)=\frac{1}{273}\times\begin{bmatrix}1 & 4 & 7 & 4 & 1\\ 4 & 16 & 26 & 16 & 4\\ 7 & 26 & 41 & 26 & 7\\ 4 & 16 & 26 & 16 & 4\\ 1 & 4 & 7 & 4 & 1\end{bmatrix} \tag{1}$$

II. A. 2) Upward sampling

Sampling up an image is the process of enlarging an image from a small image. It enlarges the image in each direction to twice the original image, with new rows and columns filled with zeros, and uses the same convolution kernel as “down sampling” multiplied by 4, and then convolves with the enlarged image to obtain the new value of “new pixels”. This is shown in equation (2):

$$\begin{vmatrix} 45 & 123 \\ 89 & 149 \end{vmatrix} \rightarrow \begin{vmatrix} 45 & 0 & 123 & 0 \\ 0 & 0 & 0 & 0 \\ 89 & 0 & 149 & 0 \\ 0 & 0 & 0 & 0 \end{vmatrix} \tag{2}$$

It adds a new row and column of pixels with a value of 0 in each direction between the original pixels 45, 123, 89, and 149. In each direction it is expanded to twice the original size, and the new rows and columns are filled with zeros. Use the same convolution kernel as for “downsampling” and multiply by 4 to get the new values of the “new pixels”.

II. B.Image alignment method for multi-scale feature learning

Image alignment refers to the use of different imaging modalities formed by the meaning of different two images, to find out a transformation to achieve the purpose of the alignment of the corresponding points, the corresponding points include the location of anatomical structure points, angle, size [22], [23]. Image alignment is based on the need for two images with similar or identical parts, the process is a coordinate mapping between one image and the other, first spatial transformation, followed by interpolation, calculating the similarity function for comparison, searching for optimization, and iterating until the similarity function reaches the optimum, the essence of which is the spatial and grayscale transformations between two images. For image $A(a)$ and image $B(b)$:

$$M(T) = M(A(a), B(T(b))) \tag{3}$$

In Eq. M refers to the similarity function and T refers to the spatial transformation, the purpose of image alignment is to find the best T transformation that makes M optimal.

In this paper, the spatial position relationship of the corresponding points between two images is described as a rigid-body transformation, which can deal with the problem of simple differences in targets due to changes in position. Before and after the transformation is performed, the distances between all points in the image correspond to equal. In this paper, the rigid-body transformation has six parameters in the three-dimensional coordinate system of the human body, which are three parameters of rotation around the coordinate axis and three parameters of translation transformation along the coordinate axis.

II. B. 1) U-Net network

U-Net network is a U-shaped network with an encoder-decoder structure, which is mainly used to solve medical image segmentation problems [24], [25]. The encoder part of U-Net consists of multiple convolutional blocks, each of which includes convolutional layers, batch normalization, and activation functions for learning low-level features of the image. With the encoder part, the network gradually extracts the features of the input image and reduces the spatial resolution. The decoder part consists of multiple inverse convolution blocks, each containing transposed convolution, batch normalization and activation function, which helps to learn the high-level semantic information of the image. With the decoder part, the network restores the feature map to the size of the original input image and generates the segmentation result. The jump connection structure between the two enables the image feature information in the encoder to be passed to the decoder, which in turn utilizes these feature maps in the decoding process to restore the image details, realizing the effective fusion of semantic abstract information and positional information between different network layers.

The symmetric structure of U-Net enables the network to learn both local features and global context information of an image, which makes U-Net perform well in medical image segmentation tasks. However, it also has some shortcomings, including difficulty in balancing global and local information, susceptibility to noise interference, and limited generalization ability.

II. B. 2) LK-CAUNet method proposed

Ⅰ Network structure of Broad-UNet-diff

Aiming at improving and optimizing the shortcomings of U-Net network in the field of image alignment, Broad-UNet-diff, a deformable image alignment network based on multi-scale feature learning, is proposed.

Given source and target images in the spatial domain, the goal of deformable image alignment is to find an optimal nonlinear dense transformation or deformation field $\phi : \Omega \times R \rightarrow \Omega$ to minimize the energy:

$$\min_{\varphi} E_D(\varphi; f \circ I_m, f \circ I_f) + E_R(R) \tag{4}$$

where E_D is a data matching term used to evaluate the similarity between the alignment feature $f \circ I_m$ of the source image and the alignment feature $f \circ I_f$ of the target image. The E_R is a regularization term.

A model for medical image alignment is proposed by integrating a multi-scale feature extraction convolutional block into a common U-Net. In the process of image alignment, the encoder extracts features from the input image pairs, while the decoder is responsible for classifying each pixel to generate the segmentation. By jumping connection between the encoder and decoder, accurate localization can be achieved in the output image.

The proposed Broad-UNet-diff model has two inputs: the target image and the source image. It generates a nonlinear distortion function, which is then applied to the source image via a spatial transformation function. The bottom of the figure shows the structure of the alignment network. A multi-scale feature extraction convolutional block is embedded into the U-Net network, which ultimately generates the deformation field.

In the model, based on the stabilized displacement field in the final output of the Broad-UNet, seven scaling and squaring layers are used to compute integrals to induce anisotropy so that the final deformation field can be represented as shown as $\phi = exp\left(v\right)$.

II Generating an image pyramid

The original image is scaled into different scales to generate an image pyramid and the scaled image is fed into different sub-networks (P-Net, R-Net, O-Net, U-Net) for training with the aim of being able to detect faces of different sizes, thus enabling multi-scale target detection.P-Net is a candidate network for image regions.R-Net is used to veto P-Net to generate most of the wrong detection frames and the method applies detection frame regression and NMS merged detection frames. The P-Net candidate frames correspond to the original image and are intercepted, at the same time, the intercepted image is scaled to 24*24*3 as the input of R-Net, and the output of the network is the same as that of P-Net.O-Net screens the detection frames even further, which is the same as the processing of R-Net, and scales the face region of the output candidate frames of the previous layer to 48*48*3 as the input of O -Net's input, the output of the network is the same, including the coordinate information of the N bounding boxes, the score and the key point location. As can be seen from the network structure, this layer has one more convolutional layer than the R-Net layer, so it can get more fine processing results.

III Multi-scale feature extraction module

The model contains a 3D feature extraction convolution block with parallel arms, and the multi-scale feature extraction convolution block is shown in Figure 2. In this multiscale feature extraction convolution block, the data is forked into parallel convolution branches with different kernel sizes after initial convolution. A 3*3*3 kernel size convolution follows a set of parallel convolutions with 1*1*1, 3*3*3 and 5*5*5 kernel sizes. In order to reduce the large number of feature parameters generated by these parallel convolutions, a kernel factorization is applied in the convolution operation, which means that the N*N*N convolution is decomposed into three consecutive 1*1*N, 1*N*1 and N*1*1 convolutions. The outputs of different branches are then connected. In addition, residual connections are added alongside the parallel convolution in order to retain the complete feature information. Finally, the output of the convolution block is extracted using the ReLU activation function corrected for multi-scale features.

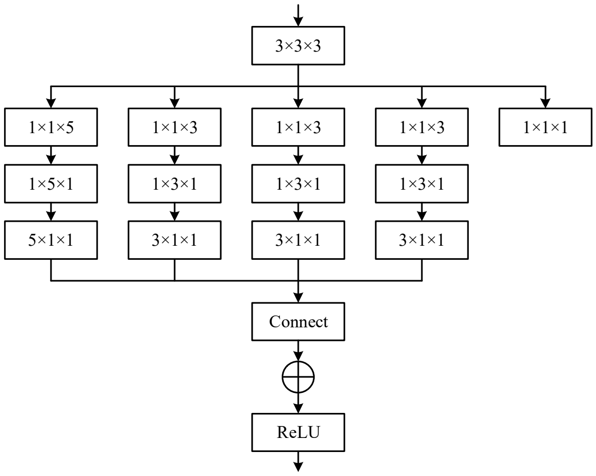


Figure 2: Multiscale features extract the convolution block

IV Overview of the LK-CAUNet algorithm

However, challenging applications and recent advances in computer vision have shown that the effective receptive field of convolutional operations is limited by the size of the convolutional kernel, and convolutional neural networks usually have limitations in modeling explicit remote spatial relationships present in images. To address this problem, on top of the network structure of Broad-UNet-diff, this paper optimizes and improves it, and proposes a novel UNet backbone network LK-CAUNet.

Ⅴ Network structure of LK-CAUNet

The proposed LK-CAUNet model framework of motion images and fixed images are input into their respective feature extraction sub-networks with large kernel multi-scale feature extraction convolutional blocks. Their features are fused and matched in the form of cross-attention and the fused features are up-sampled. After up-sampling, the full resolution displacement fields of the moving and fixed images are obtained. Next, a squaring and scaling layer is used to ensure the isoembryonicity of the final deformation. Finally, the exact amount of deformation can be estimated by minimizing the loss function, which deforms the motion image to resemble the fixed image.

Unlike the nonlocal blocks that compute self-attention on a single image, the cross-attention module proposed in this paper aims to establish the spatial correspondence between the features of two different images. The two input feature mappings of the module are denoted as the original input $P \in R^{LWH \times 64}$ and the crossover input $C \in R^{LWH \times 64}$. The L, W and H denote the dimensions of each 3-dimensional feature after flattening. The module computes the crossover features which can be expressed as:

$$y_i = \frac{\sum_{\forall j} f(\theta(C_i)^T \phi(P_j)) s(P_j)}{\sum_{\forall j} f(\theta(C_i)^T \phi(P_j))} \tag{5}$$

where C_i and P_i are the characteristics of the cross input and the original input at positions i and j, $\theta(\square), \phi(\square)$ and $s(\cdot)$ are linear embeddings, and $f(\cdot) = \exp(\cdot)$.

In Equation, $f(\cdot)$ computes a scalar representing the correlation between the features at the two locations C_i and P_i. The result y_i is a normalized summary of the features at all positions of the original input. These features are weighted with the correlation of the features on the location i of the cross input. Thus, the matrix Y consisting of y_i integrates the non-local information from the original input to each position in the cross input. Finally, the output of the module is the sum of Y and the original input to allow for efficient backpropagation. Thus, the feature at a location k in the output of the module summarizes the nonlocal correlation between the entire original input feature map and the location k in the cross-input feature map, as well as the information at k in the original input feature map.

II. B. 3) Loss function

The loss functions both consist of a mean square error (MSE) data term and a L_2 regularization term for the gradient of the deformation field, balanced by the hyperparameter λ. The mean square error data term can be expressed as:

$$loss_{sim} = \frac{1}{N} \sum_{i=1}^{N} (I_m^i \circ \phi_i(\Theta) - I_f^i)^2 \tag{6}$$

The L_2 regularization term can be expressed as:

$$loss_{smooth} = \frac{1}{N} \| \nabla v_i(\Theta) \|_2^2 \tag{7}$$

where N denotes the number of input image pairs, I_f denotes the target image, I_m denotes the source image, and Θ denotes the model parameters to be learned. The ∇ denotes the first gradient realized by finite difference. Thus, the total loss function can be expressed as:

$$loss_{total} = loss_{sim} + \lambda loss_{smooth} \tag{8}$$

where the value of λ is set to 0.5.

III. Evaluation of medical image segmentation models with improved UNet structure

III. A. Data pre-processing

In order to validate the effectiveness of the proposed model, experiments are conducted on four datasets in this paper. They are the training data provided by the Brain Tumor Segmentation Challenge organized by MICCAI in 2018 as well as 2020, and the MRI images of glioma as well as meningioma provided by a hospital in Beijing. For the sake of narrative simplicity, the datasets provided by the MICCAI challenge are noted as BraTS 2018 and BraTS 2020 in this paper, and the two types of tumor datasets provided by this hospital are named GLIA and MENIB.

In order to make the network training easier to converge, it is generally necessary to perform pixel normalization operations on the images. In this paper, zscoring is used to normalize the images. As shown in equation (9). Namely:

$$\frac{x-\mu}{\delta} \tag{9}$$

For each voxel x, first subtract the mean μ and divide by the standard deviation δ.

III. B. Experimental environment and parameter settings

The experimental models are run on a deep learning server with the relevant configurations shown in Table 1.

Table 1: Experimental environment configuration

Name	Type
Hard disk	6TB
CPU	Intel(R)Core(TM)i7-10750H
Memory	32GB
Graphics model	Geforce GTX 2080ti
Depth learning framework	pyTorch
Programming language	Python
Operating system	Linux

The specific model hyperparameter settings in the experiment are shown in Table 2. The deformation field regularization parameter λ is taken as 0.5 to ensure that the image will not be too smooth and lose details, and the learning rate is adopted as $1e^{-4}$ to reduce the overfitting phenomenon that occurs during network training.

Table 2: The concrete model superparameter Settings in the experiment

Parameter name	Parameter value
Learning rate	$1e^{-4}$
Regularization parameter λ	0.5
Loss	MSE
Epoch	500
Batch size	5
Transformer batch size	12
Optimizer	Adam

III. C. Evaluation indicators

Medical image segmentation does not reflect the advantages and disadvantages of different algorithms well using a single evaluation metric due to its complexity. Because of the different emphasis of different evaluation indexes, an algorithm may work well in one evaluation index, but perform poorly in another evaluation index.

In order to reflect the performance of different algorithms as objectively and fairly as possible, this paper adopts a variety of evaluation indexes to test the segmentation effect of algorithms. They include Dice coefficient (DSC), sensitivity (SEN), and positive predictive value (PPV).The formulas for Dice coefficient, sensitivity, and positive predictive value are shown in Eqs. (10) to (12):

$$Dice(P,T)=\frac{|P_f \wedge T_f|}{(|P_f|+|T_f|)/2} \tag{10}$$

$$SEN(P,T)=\frac{|P_f \wedge T_f|}{|T_f|} \tag{11}$$

$$PPV(P,T)=\frac{|P_f \wedge T_f|}{|P_f|} \tag{12}$$

where P , T 's are the prediction result and the true label, and the subscript f denotes the image foreground region.

III. D. Analysis of experimental results

III. D. 1) Analysis of the results of the effect of different modes on the segmentation effect

In order to verify the importance of different information for the segmentation task, the image alignment model for multi-scale feature learning is trained in this paper using data from four modalities of BraTS 2020 respectively.

BraTS 2020 segmentation results using different modalities are shown in Figure 3. The average Dice coefficients, sensitivities and positive predictive values of the trained models for different modalities are demonstrated. Where the whole tumor region (WT), tumor core region (TC), and tumor active region (ET).

It can be found that the model trained with a single modality cannot achieve good results on all three types of segmentation regions at the same time, such as Flair and T2 have very good prediction effect on WT, which reaches more than 80% respectively, but performs poorly on the segmentation of the two regions of TC and ET. T1 and T1ce also have this kind of problem, so it can be seen that the information provided by the different modalities for the training of the model is differentiated, and only using a single modality to train a model is not sufficient to predict the positive results. This shows that the information provided by different modes for model training is different, and using only a single mode to train the model often fails to get the optimal results.

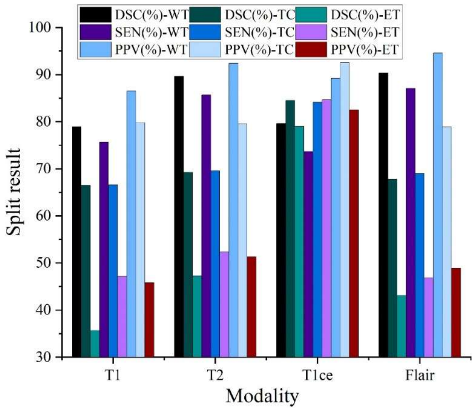


Figure 3: Brats 2020 USES different modal segmentation results

In addition, this paper also verifies the effect of modality combinations on the segmentation performance of the algorithm, including the use of T1 and T2, T1ce and Flair, and the simultaneous use of four modalities. For the two-modal combination, only two forms are used in this paper, the main reasons are: T1 and T2 is a more conventional combination, because these two modes can provide two types of information about the tumor anatomy and lesions respectively; T1ce and Flair is a combination because T1ce can show the details of the tumor interior, which makes it easy to distinguish between the tumor and the non-tumor lesion area. Flair, on the other hand, suppresses cerebrospinal fluid information, thereby highlighting the entire tumor profile. Thus, the pairing of these two modalities reflects information from both the inside and outside of the tumor.

The use of multimodal combination to train the model can provide more information for training, which can alleviate the above problems to a certain extent. The segmentation results of the algorithm by modality combination are shown in Figure 4.

The results of the model trained with T1+T2 modalities are better than the model trained with T1 or T2 alone in the segmentation of three types of regions under the Dice evaluation index.Under the Dice evaluation index, the results of the model trained with all the modalities inputted into the network at the same time are DSC(%)-WT=93.56%, DSC(%)-TC=89.42%, and DSC(%)-ET=83.22%. 83.22%.

Finally, the models trained using data from all modalities achieved very good results for the segmentation of all types of regions, and although they are not the optimal results in some of the item metrics, they are not far from the optimal results. Therefore, in this paper, for both the BraTS 2018 dataset and the BraTS 2020 dataset, all modalities are input into the network at the same time for model training, so as to maximize the use of the information contained in the data and improve the segmentation effect of the model.

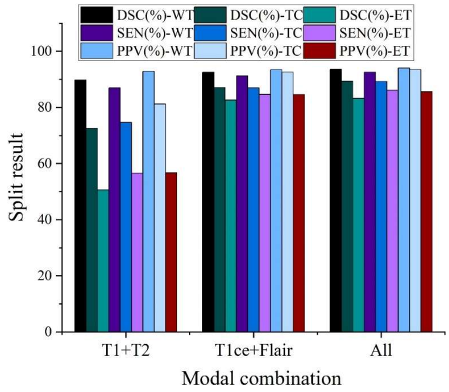


Figure 4: The split result of the modal combination on the algorithm

III. D. 2) Analysis of data set results

In this paper, we first validate the segmentation performance of LK-CAUNet and STUNet on the publicly available dataset BraTS 2020, and compare the segmentation results of the UNet network as well as the three UNet-based improved networks.

The segmentation results of this paper's algorithms on BraTS 2020 are shown in Figure 5. The average Dice coefficients, sensitivities and positive predictive values of various algorithms on the official evaluation criteria of BraTS 2020 are shown.

It can be seen that the segmentation method proposed in this paper outperforms other improved methods in the metrics of Dice coefficient (DSC) and positive predictive value (PPV).The LK-CAUNet model has 91.12%, 88.79%, and 79.93% for WT, TC, and ET on DSC, respectively. It is 12.5%, 4.19% and 9.71% higher than UNet network. The performance is obvious on the segmentation tasks of WT and ET. However, the performance in sensitivity index is inferior to other improved models.The DSC(%)-TC index of DeepResUNet model is 92.15%, which is higher than the LK-CAUNet model in this paper. Combining all the indexes, the LK-CAUNet model proposed in this paper is overall better than the other improved models.

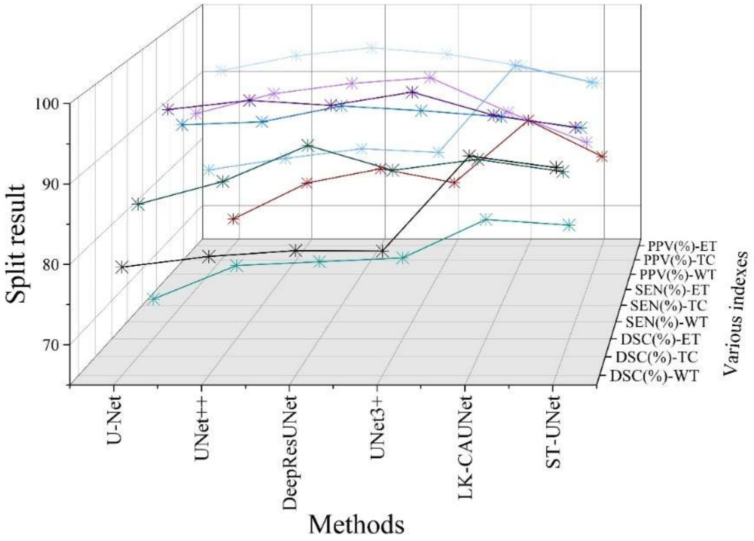
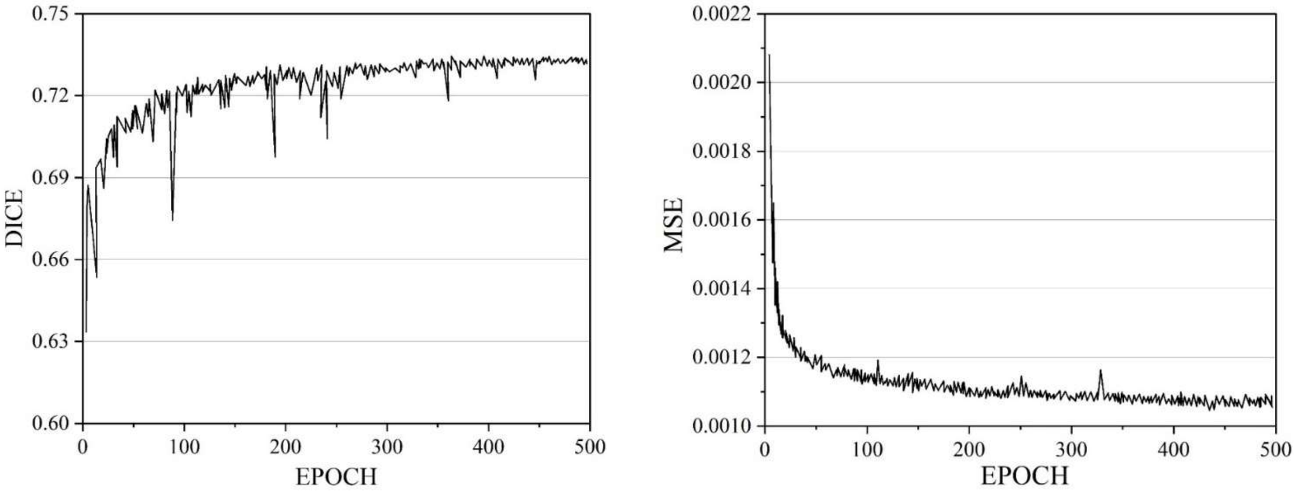


Figure 5: The algorithm split the results at brats 2020

III. D. 3) Dice Score Curve vs. MSE Curve

In order to verify the effectiveness of the image alignment model proposed in this paper to improve the multi-scale feature learning of U-net network, two types of tumor datasets GLIA and MENIB provided by the hospital are used, and the Dice scores and MSE curves obtained by applying the model training to the experimental alignment results of medical images of the proposed alignment method are shown in Fig. 6.The Dice scores curves and MSE curves are shown in Fig. 6.

It can be seen that the Dice score increases steadily with the increase of the number of iterations and tends to stabilize. And the MSE loss curve also decreases gradually with the increase of the number of iterations, and the final loss value curve is close to smooth and tends to zero.



(a) DICE score curve

(b) MSE loss curve

Figure 6: DICE score curve and MSE curve

IV. Conclusion

In this paper, we propose a deformable image alignment network based on multi-scale feature learning by improving the U-Net network. In order to alleviate the impact of the convolution operation limited by the size of the convolution kernel, a novel UNet backbone network, LK-CAUNet, is proposed on the basis of multiscale feature learning image alignment network.Preprocessing of the performance test data set of the image alignment technique, combined with Dice coefficients, sensitivity, and positive predictive value detection of the novel UNet backbone network proposed in this paper.T1, T2, T1ce, Flair single modality to train the LK-CAUNet model, the LK-CAUNet model can efficiently predict the WT region. However, the segmentation performance of both TC and ET regions needs to be optimized.When the LK-CAUNet model is trained with a combination of modalities, the combination of T1+T2+T1ce+Flair can train the segmentation performance of the LK-CAUNet model more efficiently. On the BraTS 2020 dataset, the LK-CAUNet model is able to outperform other improved methods. It is proved that modal combination training is effective for LK-CAUNet model. Meanwhile, the DICE score curves and MSE curves of the LK-CAUNet model on the GLIA and MENIB datasets indicate the stability of the LK-CAUNet model operation.

References

[1] Lo, S. J., Kuan, C. M., Hung, M. W., Fu, Y., Yeh, J. A., Yao, D. J., & Cheng, C. M. (2018). A simple imaging device for fluorescence-relevant applications. Micromachines, 9(8), 418.

[2] Debevec, P. E., & Malik, J. (2023). Recovering high dynamic range radiance maps from photographs. In Seminal Graphics Papers: Pushing the Boundaries, Volume 2 (pp. 643-652).

[3] Rao, Y., Zhao, W., Zhu, Z., Lu, J., & Zhou, J. (2021). Global filter networks for image classification. Advances in neural information processing systems, 34, 980-993.

[4] Bosse, S., Maniry, D., Müller, K. R., Wiegand, T., & Samek, W. (2017). Deep neural networks for no-reference and full-reference image quality assessment. IEEE Transactions on image processing, 27(1), 206-219.

[5] Tondewad, M. P. S., & Dale, M. M. P. (2020). Remote sensing image registration methodology: Review and discussion. Procedia Computer Science, 171, 2390-2399.

[6] Arar, M., Ginger, Y., Danon, D., Bermano, A. H., & Cohen-Or, D. (2020). Unsupervised multi-modal image registration via geometry preserving image-to-image translation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13410-13419).

[7] De Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., & Išgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis, 52, 128-143.

[8] Patel, A., Israni, D., Kumar, N. A. M., & Bhatt, C. (2019). An adaptive image registration technique to remove atmospheric turbulence. Statistics, Optimization & Information Computing, 7(2), 439-446.

[9] Polo, A. L., Nix, M., Thompson, C., O'Hara, C., Entwisle, J., Murray, L., ... & Svensson, S. (2024). Improving hybrid image and structure-based deformable image registration for large internal deformations. Physics in Medicine & Biology, 69(9), 095011.

[10] Abdel-Basset, M., Fakhry, A. E., El-Henawy, I., Qiu, T., & Sangaiah, A. K. (2017). Feature and intensity based medical image registration using particle swarm optimization. Journal of medical systems, 41, 1-15.

[11] Borovec, J., Munoz-Barrutia, A., & Kybic, J. (2018, October). Benchmarking of image registration methods for differently stained histological slides. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 3368-3372). IEEE.

[12] Liu, C., Xu, J., & Wang, F. (2021). A review of keypoints' detection and feature description in image registration. Scientific programming, 2021(1), 8509164.

[13] Velesaca, H. O., Bastidas, G., Rouhani, M., & Sappa, A. D. (2024). Multimodal image registration techniques: a comprehensive survey. Multimedia Tools and Applications, 83(23), 63919-63947.

[14] Lv, G. (2019). Self-similarity and symmetry with SIFT for multi-modal image registration. IEEE Access, 7, 52202-52213.

[15] Ye, Y., Tang, T., Zhu, B., Yang, C., Li, B., & Hao, S. (2022). A multiscale framework with unsupervised learning for remote sensing image registration. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-15.

[16] Lei, Y., Fu, Y., Wang, T., Liu, Y., Patel, P., Curran, W. J., ... & Yang, X. (2020). 4D-CT deformable image registration using multiscale unsupervised deep learning. Physics in Medicine & Biology, 65(8), 085003.

[17] Dogra, A., Goyal, B., & Agrawal, S. (2017). From multi-scale decomposition to non-multi-scale decomposition methods: a comprehensive survey of image fusion techniques and its applications. IEEE access, 5, 16040-16067.

[18] Yang, Z., Dan, T., & Yang, Y. (2018). Multi-temporal remote sensing image registration using deep convolutional features. Ieee Access, 6, 38544-38555.

[19] Baygin, M., Karakose, M., Sarimaden, A., & Akin, E. (2017, September). Machine vision based defect detection approach using image processing. In 2017 international artificial intelligence and data processing symposium (IDAP) (pp. 1-5). Ieee.

[20] Yue, Z., Huang, L., Lin, Y., & Lei, M. (2024). Research on image deformation monitoring algorithm based on binocular vision. Measurement, 228, 114394.

[21] Jintao Tan,Longyang Huang,Zhonghui Chen,Ruokun Qu & Chenglong Li. (2025). DarkSegNet: Low-light semantic segmentation network based on image pyramid. Signal Processing: Image Communication,135,117265-117265.

[22] Yi Zhu,Qinghua Wang,Xinyun Xie & Xiaojun Yan. (2025). Image registration method for full-field deformation measurement. Optics and Laser Technology,184,112427-112427.

[23] ZhongyuYang,MohsenMohammadi,HaolinWang & Yi Chang (James)Tsai. (2024). A feature-based pavement image registration method for precise pavement deterioration monitoring. Computer-Aided Civil and Infrastructure Engineering,40(10),1276-1294.

[24] Yukun Wang,Fan Ye,Xiaobo Chen & Juntong Xi. (2025). Image restoration of underwater fuel assembly thermal turbulence based on improved U-Net network. Nuclear Engineering and Technology,57(9),103615-103615.

[25] Maria Chiara Brunese,Aldo Rocca,Antonella Santone,Mario Cesarelli,Luca Brunese & Francesco Mercaldo. (2025). Explainable and Robust Deep Learning for Liver Segmentation Through U-Net Network. Diagnostics,15(7),878-878.