

ITC 2/53 Information Technology and Control Vol. 53 / No. 2 / 2024 pp. 429-441 DOI 10.5755/j01.itc.53.2.35021	Application Simulation Research Based on Visual Image Capture Technology in Sports Injury Rehabilitation	
	Received 2023/09/04	Accepted after revision 2023/11/19
	HOW TO CITE: Tang, X. (2024). Application Simulation Research Based on Visual Image Capture Technology in Sports Injury Rehabilitation. <i>Information Technology and Control</i> , 53(2), 429-441. https://doi.org/10.5755/j01.itc.53.2.35021	

Application Simulation Research Based on Visual Image Capture Technology in Sports Injury Rehabilitation

Xun Tang

College of Physical Education, Chengdu University, Chengdu, 610106, China

Corresponding author: tangxun2023@126.com

Abstract: To capture and analyze the motion state of patients in real time and improve the evaluation effect of sports injury, the research is based on image recognition in visual image capture technology. Firstly, multi-scale attention mechanism was introduced into U-Net image segmentation model to improve the pre-processing of image recognition. Then, the image recognition model of convolutional neural network is optimized by gradient class weighted activation mapping. The combination of the two is applied to the sports injury image processing to verify the effect. The results show that the F1 score and Precision values of the improved segmentation model in the database reach 98.85% and 98.74%, respectively. The segmentation accuracy is obviously improved. The accuracy of the optimized image recognition method in the training set and the test set is about 96% and 98%, respectively. After the combination of the two methods, the processing accuracy of sports injury medical images is 97%, and the running time is within 4s. It has high accuracy and processing efficiency, providing a technical and methodological basis for sports injury rehabilitation training.

KEYWORDS: Image segmentation; Classification; Motion impairment; Visualization; CNN; U-Net.

1. Introduction

Exercise is an important activity to promote physical health. It is a necessary content for the people to achieve a healthy life [15]. With the extensive promotion of national fitness activities, sports have gradually become the only way for most people to pursue a

better life. In some sports, it is inevitable that tissue and organ injuries closely related to sports events, intensity and personal quality, that is, sports injuries, which causes serious threat to human health. With the development of digital technology, image capture

technology is gradually becoming mature. Applying it to the medical image recognition can provide doctors with the organ and tissue information needed for disease diagnosis, providing technical reference for subsequent rehabilitation training [1]. As a pre-processing step of medical image recognition, image segmentation (IS) technology can provide reference information for quantitative evaluation such as brain tumor segmentation and injury IS. It acts a crucial role in the automatic identification of lesion information. The current IS is mainly based on the supervised segmentation algorithm. Although the segmentation effect is good, it lacks accurate image data annotation. Meanwhile, the edge part effect is still poor [27]. At the same time, Convolutional Neural Network (CNN) uses artificial neurons to respond to surrounding units to achieve large-scale image processing, which is superior in the image recognition. However, the CNN also needs to be further improved to enhance the recognition accuracy and efficiency [25]. Therefore, the research improves the U-Net segmentation model. It is combined with the optimized CNN to get better application in sports injury rehabilitation. The research is mainly divided into four parts. The first part is a literature review on IS and recognition algorithms in visual image capture technology. The second part is about the research methods used in this paper. The first section is about IS through multi-scale attention mechanism. The second section is about sports injury image recognition through improved CNN. The third part validates the proposed image capture technology, including performance testing of the proposed method and actual application effects analysis. The last part summarizes the application of the proposed visual image capture technology in sports injury.

2. Related Works

IS in visual image capture technology is a crucial component influencing how well image recognition works. Many academics, both domestically and internationally, have become interested in it recently. Cui X's team [5] developed a recognition method based on dual-channel deep learning (DL) for the limitations faced by the information-diverse image recognition from local or global features. The results showed that this method obtained higher recognition accuracy. To

improve the application of IS platform in medicine, Müller and Kramer [6] proposed an open source Python library MIS CNN. This database contained DL models, training models and smart technologies. The data showed that the model framework could quickly produce the required medical images [17]. Saood and Hatem [21] proposed to use DL model and U-NET for image classification and segmentation to better diagnose patients with new crowns. Combined with computer technology, the patient's lungs were scanned. The results showed that the method had high usability in medical IS. To enhance the accuracy of medical IS, You C's team [26] proposed to introduce the contrastive distillation framework Sim CVD into the IS. The method was based on an unsupervised training strategy with a large amount of data. The results showed that the average scores of the framework were 90.85% and 89.03%, which were 0.91% and 2.22% higher than the traditional advanced methods. Scholars such as Feng et al. [7] proposed an interactive few-shot learning (IFSL) method to solve the problem that the existing DL model needed a lot of data annotation during training. It achieved IS by introducing interactive learning into a few shot learning strategy. When there were four image lenses, the performance of this method was significantly improved by more than 20%.

Image recognition has also achieved a series of breakthrough research results. Raza and Singh [20] analyzed the application of IS in the medical field. The application of common models in medicine was compared. The advantages and disadvantages of supervised learning and unsupervised learning were discussed. The final results showed that both learning methods could be applied in the medical field, but the recognition effect of unsupervised learning technology was better [20]. To better grasp the application principles of DL models in the medical area, Guo et al. [10] combined machine learning (CML) with DL as a hybrid intelligence-driven medical image recognition method. The data showed that this method could improve the image recognition accuracy by 2% to 3%. Sarvamangala and Kulkarni [22] conducted research on the application of CNN in the medical field. The research discussed the various frameworks included in CNN and introduced the applications in medical IS, recognition. It was found that CNN had better performance and feasibility in medical image recognition. Li et al. [23] proposed several different super-resolution

methods to solve the limitations of long medical image acquisition time and high radiation dose. The DL and different frameworks were described and compared. The findings demonstrated that DL technology could be widely used in the medical field, which was conducive to intuitively discovering the location of patients' lesions. Alalwan et al. [2] proposed a semantic segmentation DL model for human liver and tumor segmentation to improve the accuracy of medical IS and recognition. Compared with traditional models, this model effectively utilized laterally separable convolutional images. The final results showed that the model had high performance. It could be effectively applied to relevant medical IS and recognition.

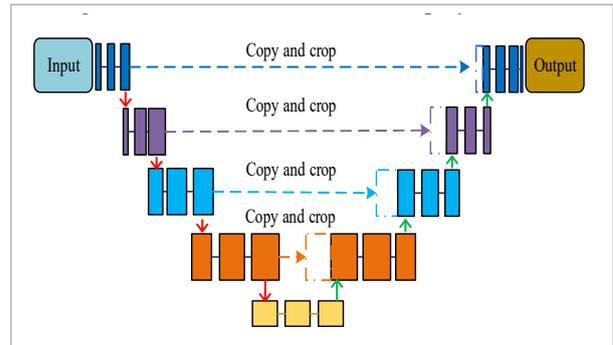
In summary, for the application of IS and image recognition technology in the medical field, most researchers have combined and improved the DL model with other algorithms. Therefore, it can be better applied to medical image tasks. However, there are few applications for the fusion of the two methods. The accuracy needs to be enhanced. Therefore, IS and recognition technology will be improved separately. It is applied to image recognition of sports injury to provide more reliable information support for rehabilitation training.

3. Sports Injury Rehabilitation Training Based on Image Capture Technology

3.1. IS Based on Multi-scale AM

In this study, the U-Net framework is used to build a segmentation network model for sports injury images. U-Net is developed from a fully CNN. The output resolution is improved through up sampling [16]. When segmenting boundaries, image data of the outer circle is needed for specific areas. When the image itself is the boundary, mirror extrapolation is used to supplement the missing data to achieve segmentation training [4]. The U-Net structure is U-shaped on the whole. Down sampling is maximum pooling and convolution. The up sampling is transposed convolution. Finally, a 1×1 convolution operation with dimensionality reduction as the main function is performed. In Figure 1, the U-Net structure is displayed.

Figure 1
U-Net general structure diagram



In Figure 1, the U-NET network structure is symmetrical, forming a structure similar to the letter “U”. The left part is a down sampling encoding network. In the coding structure, convolution, pooling, and activation function are used to extract image features. The right part of the U-Net network structure is the up sampling decoding network. The decoding network restores images through repeated up sampling, convolution, and activation function. The difference between U-Net network and FCN network is that U-Net network uses the same convolutional operation in the encoding and decoding stage. Between layer encoding and decoding, skip connections are used to fuse more features with low-level image details and high-level semantic information. U-Net adds the boundary weight to the Loss function, as shown in Formula (1).

$$E = \sum_{x \in \Omega} w(x) \log(p_{l(x)}(x)). \quad (1)$$

In Formula (1), Ω represents the label value. $p_{l(x)}(x)$ is the softmax. w represents the weight value. x represents the pixel point. The calculation method of $w(x)$ is shown in Formula (2).

$$w(x) = w_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right) + w_c(x). \quad (2)$$

In Formula (2), d_1 represents the distance from the pixel point to the nearest cell. d_2 is the distance from the second nearest cell. $w_c(x)$ represents the weight that balance the foreground-background ratio. σ and w_0 represent constant values. Although U-Net can cope with the fewer training samples, it causes overlapping and redundant calculations, leading in the

loss of background information. Therefore, the study introduces a multi-scale attention network for IS, and establishes an image network segmentation model. This model includes 5 parts, convolution layer, de-convolution layer, pooling layer, spatial attention module and IR-Block module. The input is the sports injury image to be segmented. The IR-Block block, the 2×2 convolution kernel and the pooling layer with a stride of 2 are used as the structure of the encoder layer. The primary purposes are to lower resolution and extract features. After adding the pooling layer, the size of the feature map (FM) varies. The image FM is reduced to half of the upper image feature. After the first pooling operation, the obtained image features are reduced to $1/2$ of the original image. After the second layer pooling operation, it is reduced to $1/4$. After the third layer, it is $1/8$. In the decoder layer of the model, the image resolution and feature size are reduced due to the pooling operation of the encoder layer. To restore it to the original size, the de-convolution operation is used to realize the mapping from small resolution to large resolution. The first de-convolution operation expands the size to 2 times, the second to 4 times, and the third to 8 times. The last layer of the model is the convolution operation of the 1×1 convolution kernel. The resolution and size of the obtained image are consistent with the size of the input image. They are mapped through the Sigmoid activation function. Then the type of pixel is determined to realize segmentation prediction. The encoder and decoder layers of the model are connected through a skip connection layer. An attention mechanism (AM) is added simultaneously. The skip connection can provide high-resolution image features to the decoder layer, which combines the bottom and high-level features of the image, improving the model performance. Figure 2 shows the framework of the IRS-Block model.

In the sports injuries IS task, based on AM, all attention is focused on key information, which mainly includes two main steps. The first one is to use the global scan to extract local useful information. The other is to suppress redundancy information and enhance useful information. Since the channel and spatial information of the FM are fused in the local receptive field when the convolutional layer extracts features, the information contained in the spatial and channel positions is quite different. Therefore, the research introduces a spatial attention module to preserve the global information in the image. Figure 3 is the structure of the spatial attention module.

Figure 2

The proposed IRS-Block model framework

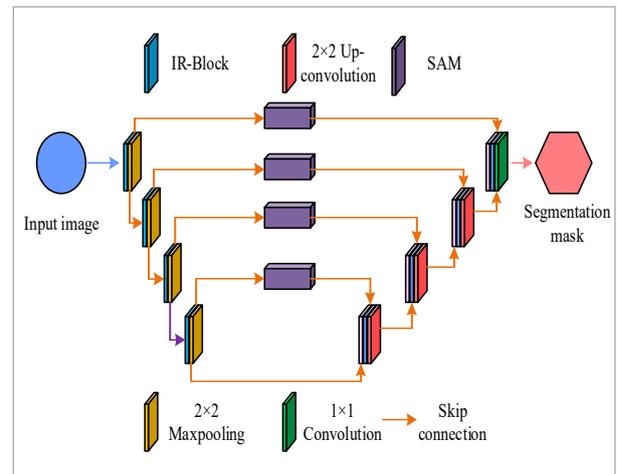
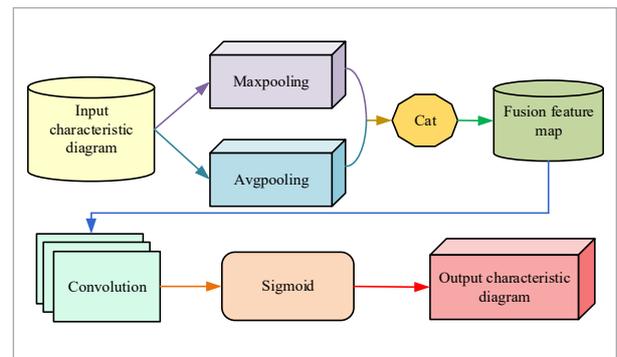


Figure 3

Frame structure of spatial attention module



The primary role of the attention module is to convert the key information contained in the image into the spatial domain, because the semantic consistency of the image is obtained according to the weighting process of the channel AM [6]. The input of the spatial attention module is a FM. Two branches are generated to perform operations. One mainly performs a global maximum pooling operation on the feature information, so as to reduce the dimensions and parameters in the network and avoid over fitting to the greatest extent. Another branch performs the global average pooling operation, so that all positions in the FM can get a pixel-level global attention map. Then the FMs processed by the two branches are connected. The dimensionality reduction and feature fusion are realized through the convolutional layer with a convo-

lution kernel size of 7×7 . Finally, the spatial attention FM is obtained by the Sigmoid activation function, as shown in Formula (3).

$$M_s(F) = \text{Sigmoid} \left(f^{7 \times 7} (\text{AvgPool}(F); \text{MaxPool}(F)) \right) \\ = \text{Sigmoid} \left(f^{7 \times 7} (F_{\text{avg}}^S; F_{\text{max}}^S) \right). \quad (3)$$

In Formula (3), $f^{7 \times 7}$ is a convolution operation with a convolution kernel size of 7×7 . F_{avg}^S is the FM obtained by global average pooling. F_{max}^S represents the FM obtained by global maximum pooling. Spatial attention has strong portability, which can be embedded in the U-Net network structure. It is conducive to the extraction of image detail features. To address the gradient decrease during training and enhance training stability, the optimization and update iteration of the network are realized through the binary cross-entropy loss function, which is defined as shown in Formula (4).

$$\text{LOSS}_{\text{BCE}} = - \sum_{i=1}^N \sum_{j=1}^L \left((1 - \log p_{ij})(1 - g_{ij}) + g_{ij} \log p_{ij} \right). \quad (4)$$

In Formula (4), p_{ij} is the predicted result value obtained by network segmentation. g_{ij} is the real category.

3.2. Image Recognition of Sports Injuries Based on Improved CNN

After IS, a CNN combined Gradient-weighted Class Activation Mapping (Grad-CAM) is proposed for image classification, which includes AlexNet network and Gradient-weighted Class Activation Mapping (Grad-CAM) layer. The AlexNet network consists of 3 fully connected layers and 5 convolutional layers. The output of the last connected layer is classified through the input of the Softmax function. The diagnostic classification process of medical images using AlexNet network is shown in Figure 4.

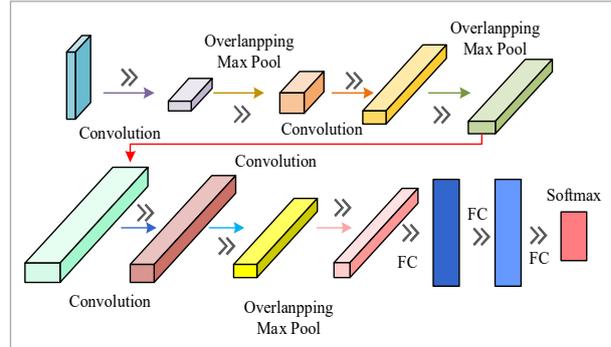
The convolution kernel contained in GCCV-CNN uses convolution operation to obtain the medical image feature, that is, FeatureMap. The convolution operation is shown in Formula (5).

$$x'_{i,j} = \sum_{c=0}^{C-1} \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} \left(u_{c,(n+1),(m+1)} \cdot x_{c,(i-1+1),(j-1+1+n)} \right). \quad (5)$$

In Formula (5), c indicates the quantity of channels. $x'_{i,j}$ indicates the element whose coordinate is (i, j) . The pooling kernel's size is represented by k . The values of m and n are $0, 1, \dots, k-1$. After the convolu-

Figure 4

AlexNet network for diagnosis and classification of medical images



tional layer operation is completed, the pooling layer is utilized to lower the network's computational loss. It is also to extract relevant features in the image. The pooling function selected in the research is the maximum value. It can overcome the blurring defect caused by the average pooling, thereby retaining the most significant features. The maximum pooling calculation is shown in Formula (6).

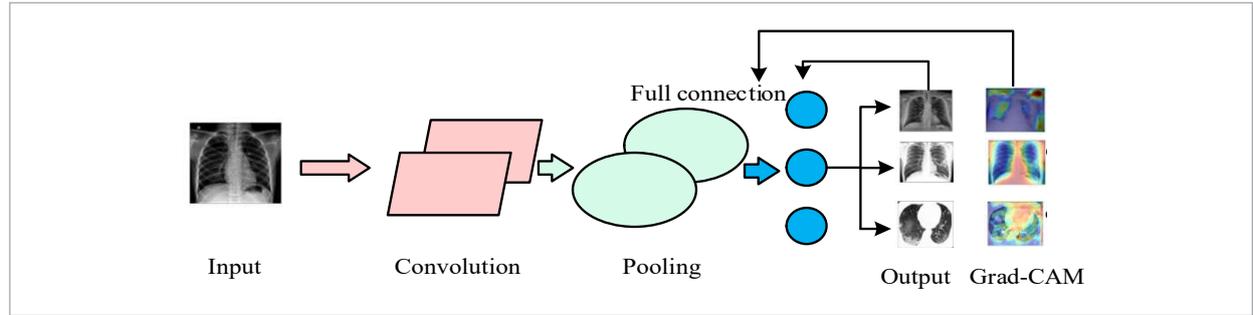
$$x_{l,i,j} = \max_{m \leq k-1, 0 \leq n} \left(x_{l-1,(j-1)stride+n+1,c,(i-1)stride+m+1} \right). \quad (6)$$

In Formula (6), $x_{l,i,j}$ represents the element at the coordinate (i, j) of the layer l in the FeatureMap after the pooling operation. When the pooling kernel is larger than $stride$, the output of the pooling layer overlaps each other, the feature richness is improved, and the information loss is less [8]. After the pooling layer in the fully connected layer, all neurons are fully connected to the neurons in the upper layer to achieve information separation in the integrated area. The ReLU function serves as the fully linked layer's activation function. The proposed GCCV-CNN classification process is shown in Figure 5.

When using CNN to classify images, the last convolutional layer can obtain the FeatureMap containing the key information of the original image, that is $R^{u \times v}$. u and v represent the breadth and height, respectively. The CAM technology removes the fully connected layer of the last convolutional layer and adds a new global average pooling (GAP) layer [14]. Each FeatureMap in the GAP layer calculates the mean value to obtain a new FeatureMap. Then it is directly input into the Softmax layer. Then, the weights from the GAP layer to the Softmax layer ω_k^c are multiplied

Figure 5

The Proposed GCCV-CNN classification process



and added to each FeatureMap, thereby obtaining a marked activation map (CAM) with the same size as the original image. The FeatureMap obtained by the last convolutional layer is k , denoted as A^k . Through the GAP linear transformation, the FeatureMap generates c categories. The score of the specific category is S^c , as shown in Formula (7).

$$S^d = \sum_k \omega_k^d \frac{1}{Z} \sum_i \sum_j A_{ij}^k. \quad (7)$$

In Formula (7), d represents the serial number of the target category. ω_k^d represents the weight from the GAP layer to the Softmax layer. j and i represents the high and wide dimension serial numbers, respectively. A represents the final output FeatureMap. Since the CNN structure does not have weight ω_k^d , the network must eventually be retrained. Therefore, the research introduces Grad-CAM to obtain the Grad-CAM diagram of the image, as shown in Formula (8).

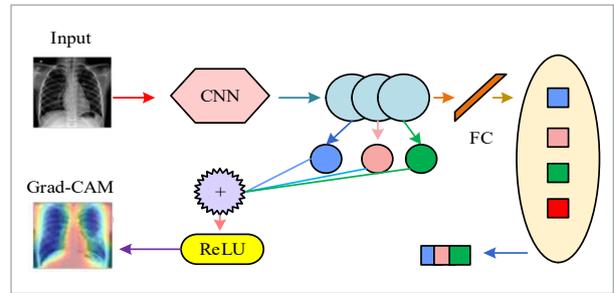
$$S^d = \frac{1}{Z} \sum_i \sum_j \sum_k A_{ij}^k \omega_k^d. \quad (8)$$

Adding Grad-CAM technology to the CNN structure can provide a better visual interpretation and interpretability view for the neural network. Therefore, more information that can explain CNN in classification tasks is obtained. The study uses Grad-CAM technology to obtain a gradient-weighted category activation map, highlighting important areas of the image and improving classification visualization through marking. The deep-level features in CNN can well represent the visual structure of the image. The final output FeatureMap retains the spatial information lost by the fully connected layer. This creates

necessary conditions for using FeatureMap to save the original image spatial information. The Grad-CAM technology learns the weight of all neurons to the final decision through the gradient information of the last convolutional layer, which is in line with CNN's features [24]. The CNN image classification based on Grad-CAM technology is shown in Figure 6.

Figure 6

CNN image classification process based on Grad-CAM technology



y^g is target class probability output by the Softmax layer in CNN. A_{ij} is the final output FeatureMap. The calculation method of partial derivative y^d of A_{ij} is shown in Formula (9).

$$a_{ij} = \frac{y^d \partial}{A_{ij} \partial}. \quad (9)$$

The partial derivative for all pixels of y^d on the FeatureMap and the global average are calculated.

$$a_k^d = \frac{1}{Z} \sum_i \sum_j \frac{y^d \partial}{A_{ij} \partial}. \quad (10)$$

According to Formula (10), y^d is the interest degree that the d class for the final FeatureMap of the k channel. a_k^d is weight. The final FeatureMap is linearly weighted and combined. The result is shown in Formula (11).

$$L_d = \sum_k a_k^d A^k . \quad (11)$$

The two-dimensional FeatureMap is passed to the ReLU activation function. The output is shown in Formula (12).

$$L_{Grad-CAM}^d = ReLU\left(\sum_k a_k^d A^k\right) . \quad (12)$$

Finally, the output result of the two-dimensional map is obtained, which is the Grad-CAM map obtained from the input image. Therefore, the final output FeatureMap can be retained through Grad-CAM technology. Then it is reversed to obtain the interest degree of the target category in the final output FeatureMap. The features are marked on the original image to realize the color visualization of image classification. For segmentation and recognition accuracy, the calculation of the evaluation index F1 score is shown in Formula (13).

$$F1 = \frac{2TP}{FN + FP + 2TP} . \quad (13)$$

In Formula (13), TP means the correct detection. The result of the positive class detection is the positive class. FN indicates false detections. The positive detection results are negative. FP indicates the false detections. The negative class results are positive class. The Accuracy and Precision are shown in Formula (14).

$$\begin{cases} Accuracy = \frac{TN + TP}{TN + FN + FP + TP} \times 100\% \\ Precision = \frac{TP}{FP + TP} \times 100\% \end{cases} . \quad (14)$$

In Formula (14), TN is the wrong detection. The negative class detection result is the negative class. The Sensitivity and Specificity are shown in Formula (15).

$$\begin{cases} Sensitivity = \frac{TP}{FN + TP} \times 100\% \\ Specificity = \frac{TN}{FP + TN} \times 100\% \end{cases} . \quad (15)$$

4. Analysis of the Application Effect of Image Recognition in Sports Injury Rehabilitation

Firstly, the IS effect of the proposed IRS-Net network is verified. To improve classification accuracy, the IRS-Net network model is pre-trained on the open Data set Lung Kaggle in the Lung Nodule Analysis competition in the 2017 Kaggle Data Science Bowl. During the training and testing, the data in nii format is converted to numpy format, which is more convenient for data loading. The data set consists of 3D CT images, including lung CT images and corresponding mask images, respectively. The experimental platform is Ubuntu operating system. The hardware configuration of the experiment is Tesla P4. The memory is 16 G. The software is CUDA10.1 and Python 3.6. The deep learning framework is Keras. The proposed segmentation algorithm uses Adam optimizer for network training. To improve the efficiency of the network, batch input of the network is adopted. The batch size is 8. A total of 50 epochs are trained to achieve the convergence of the model. After fine-tuning the model, to verify the IS effect of IRS-Net network, the method proposed in the research is compared with classic medical IS networks, including CA-Net, DO-UNet, and U-Net. Table 1 displays the fundamental environment.

Table 1

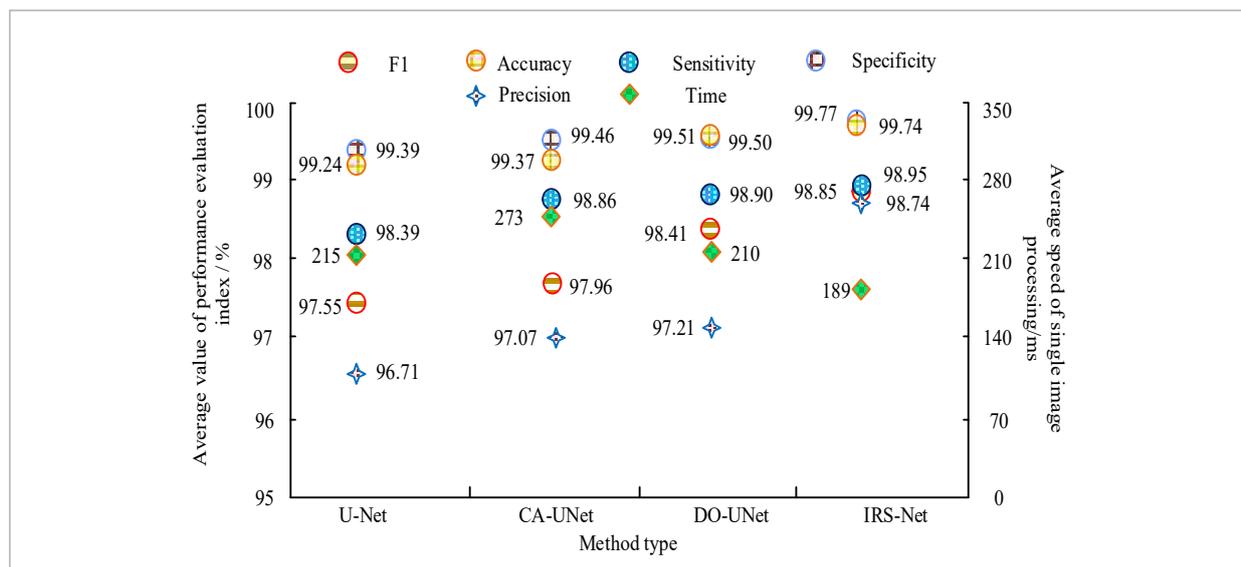
Basic environment of experiment

Project _	Configuration conditions
Experimental platform	Ubuntu operating system
GPU	Tesla P4
Internal memory	16G
Software configuration	CUDA10.1, Python3.6
DL framework	Keras

To verify the IS effect of IRS-Net network, it is compared with the U-Net Medical IS network and the CA-Network medical IS network in literature [11], and the DO-UNet medical IS network in literature [9, 11, 12]. The data set used in the test is the Musculoskeletal Radiographs Abnormalities Database

Figure 7

Performance evaluation index mean of four methods in MURA database comparison results of Average Speed of Single Image Processing



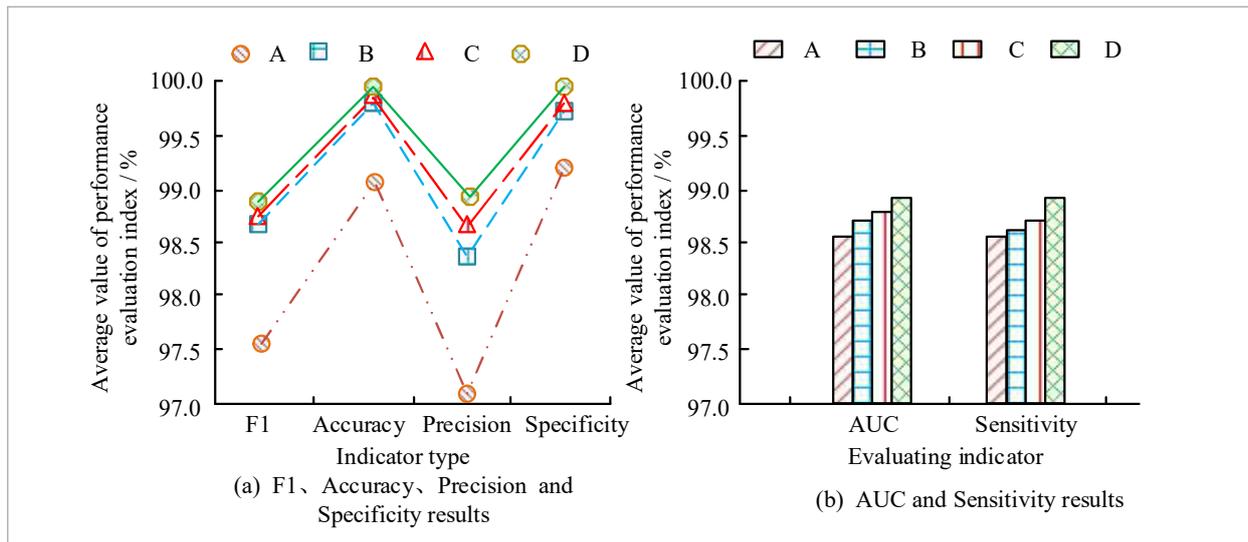
(MURA) database, which is a data set of musculoskeletal radiographs. This dataset includes 12,173 patients, a total of 14,863 researches, and 40,561 multi-view radiology images. The performance evaluation results of the four methods in this dataset are shown in Figure 7. From Figure 7, the F1 score of the proposed IRS-Net network is 98.85%, which is 1.30%, 0.89% and 0.44% higher compared with U-Net, CA-Net and DO-UNet, respectively. The IRS-Net network accuracy is 98.74%, which is 2.03%, 1.30% and 1.25% higher than U-Net, CA-Net and DO-UNet, respectively. It shows that the improved network enhances the accuracy index, because of the proposed method in this study significantly improves the segmentation accuracy of motor damage images. The IRS-Net network accuracy is 99.74%, 0.50%, 0.37%, and 0.23% higher compared to U-Net, CA-Net and DO-UNet, respectively. The proposed IRS-Net network has better F1 score, accuracy and accuracy evaluation metrics than the other three methods. Compared with other algorithms, the improved IRS-Net algorithm has lower sensitivity, because the edge features of the medical images are not significant. Finally, in the running time, the IRS-Net model is 189 ms, which is 26 ms, 84 ms, and 21 ms faster compared to U-Net, CA-Net, and DO-UNet, respectively. It has less runtime. Combined with the objective evaluation index, the proposed model is still superior. The reason

may be that the multi-scale AM in the IRS-Net network and embedded Grad-CAM technology provide high-resolution image features for the decoder layer. The bottom and advanced features of the image are combined to improve the performance of the model.

To further verify the influence of the AM and the multi-scale feature extraction module on the IS performance, different modules are added to the U-Net network. The optimal effect of IRS-Net is proved by ablation experiments. The test of the basic U-Net in the MURA database is recorded as Experiment A. The combination of AM and U-Net network is recorded as Experiment B. The combination of multi-scale feature extraction and U-Net network is recorded as Experiment C. The proposed IRS-Net model is recorded as experiment D. Figure 8 displays the comparison findings. Figure 8(a) displays the F1 score, Accuracy, Precision and Specificity results of the four experiments. Figure 8(b) shows the AUC and Sensitivity results of the four experiments. From Figure 8(a), the F1 score, Accuracy, Precision and Specificity values of Experiment D are 98.85%, 99.74%, 98.74% and 99.77%, respectively. The finding indicates that the IRS-Net model is better than the other three models. From Figure 8 (b), the AUC and sensitivity indexes of experiment D are better than the other three experiments. The AUC is the area under the ROC curve, be-

Figure 8

Mean value of performance evaluation indexes of four experiments



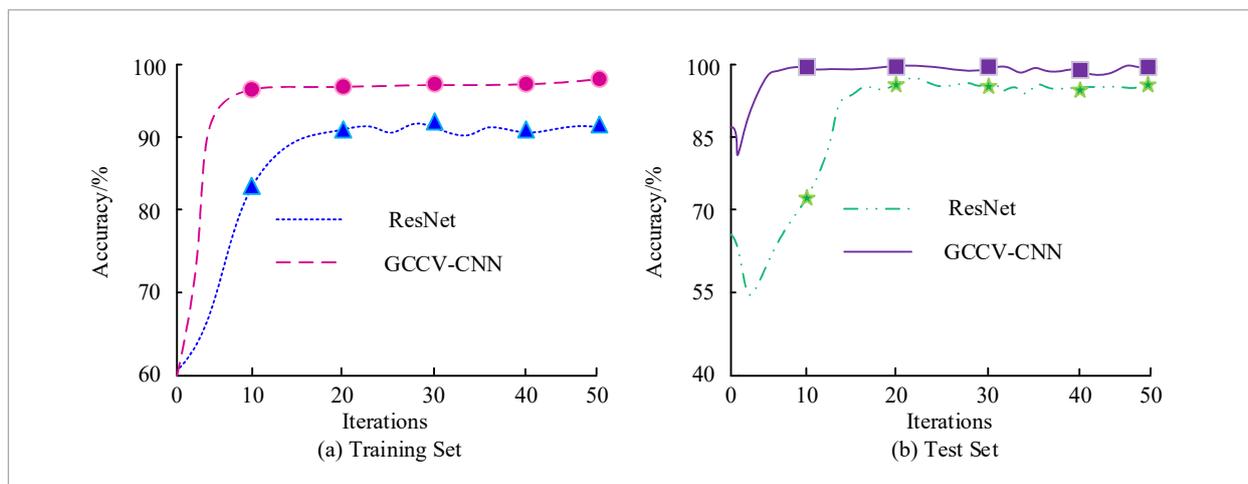
tween 0.1 and 1. The larger AUC in the horizontal axis shows better model performance. This indicates that the IS performance of the IRS-Net model is better. The AM and multi-scale feature extraction modules have the potential to significantly improve the IS performance.

The GCCV-CNN algorithm proposed in the study is verified for image classification. The ResNet model is commonly utilized in medical image classification

in the CNN model. It is compared with the proposed GCCV-CNN mode. A test set and a training set are separated in the MURA database, respectively, which are randomly selected according to the ratio of 25% and 75% respectively. Considering the CPU, GPU and memory performance, both models are run 50 times. Figure 9 displays the obtained accuracy curves for the training set and test set. Figure 9(a) and (b) respectively show the variation of the accuracy rate in

Figure 9

The accuracy curve of the training set and test set obtained by running the two models for 50 times



the training set and test set with the number of iterations. The accuracy rate is plotted on the vertical axis, with the number of iterations on the horizontal axis. From Figure 9(a), the ResNet model converges at 15 iterations, and then the accuracy rate stabilizes at about 90%. The GCCV-CNN model converges at 5 iterations, and the accuracy rate stabilizes at 96%, with an increase of 6%. From Figure 9(b), the GCCV-CNN model is still better than the ResNet model, converging only in 8 iterations. The accuracy rate can be maintained at about 98%. The ResNet model is only stable around 93%. Overall, the GCCV-CNN model achieves a superior classification effect and a better training effect.

Figure 10 shows the changes in the loss curves of the two models. The loss coefficient is represented by the vertical axis, while the number of iterations is represented by the horizontal axis. From Figure 10(a), the loss coefficient of the ResNet model fluctuates greatly with the iterations. The highest iteration is 0.7, and the lowest is about 0.1, which is very unstable. However, after the 10th iteration of the GCCV-CNN model, the loss coefficient gradually fluctuates around 0. The maximum value is only 0.4, and the waveform does not appear to vibrate significantly. From Figure 10(b), the fluctuation of the loss curve for the ResNet decreases, the highest is close to 1.2, and the loss coefficient approaches 0 only between 20 and 35 iterations. However, the fluctuation of the loss curve in the GCCV-CNN model basically disappears. The loss coefficient approaches 0 after 5 iterations. It shows that

the error of the GCCV-CNN model is smaller, which has stronger classification ability.

Finally, The IRS-Net IS algorithm is used as a preprocessing step. It is combined with GCCV-CNN algorithm in the recognition and classification of sports injury images. Through the investigation of 10 well-known rehabilitation institutions and 5 provincial hospitals, skeletal medical images, muscles and skin images in sports injuries are collected. The new data are formed after screening by doctors and experts in the field set to verify the actual application effect of the combined method. The data set contains a total of 1086 images of bone, muscle and skin injuries. The combined method is compared with GCCV-CNN, CNN in reference [12], and the M-C model in reference [27]. Each model is run 3 times. A total of 100 images are processed randomly. The outcomes are displayed in Figure 11. Figure 11(a) shows the image recognition accuracy of the 4 methods. Figure 11(a) displays the time required for 3 processing. According to Figure 11(a), the highest accuracy rates of CNN, M-C and GCCV-CNN are 90%, 94%, and 95%, respectively. The accuracy rate of CNN fluctuates relatively greatly. The lowest value is close to 85%. The accuracy of GCCV-CNN method is basically kept above 95%, with a maximum of 97%, because the improved method model is able to extract more abundant and more accurate features from medical images. From Figure 11(b), among the four methods, CNN consumes the shortest time, all within the range of 2s. The time difference between M-C and GCCV-CNN is small, all within the range of 3-4s. The combined combination is as low

Figure 10

Test set and training set loss curve results of the two models

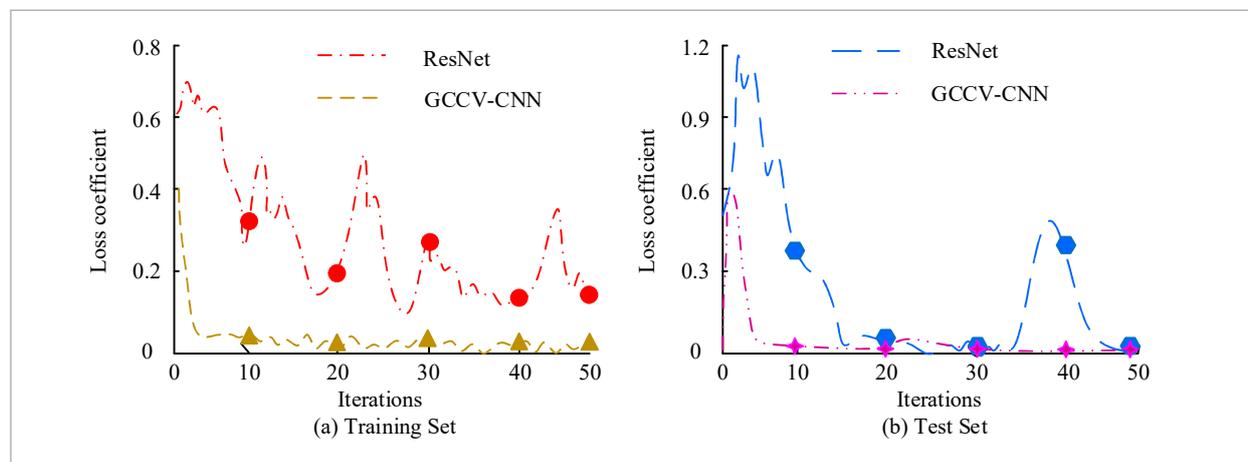
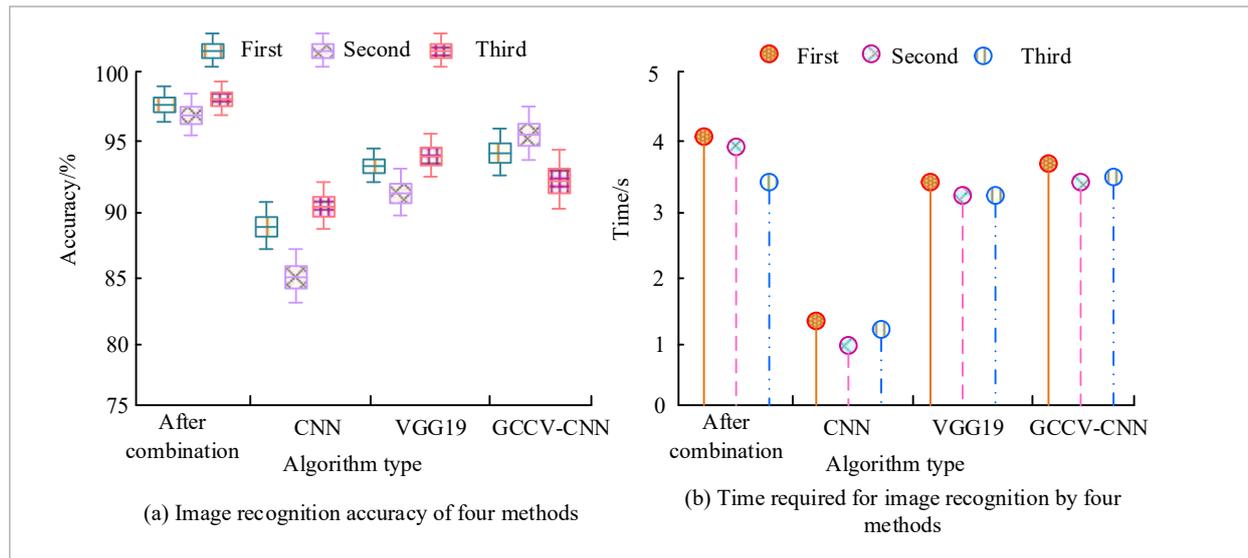


Figure 11

The results of four models in the recognition of actual sports injury images



as 3s and as high as 4s. The required time is gradually shortened. From the perspective of comprehensive time and accuracy, the running time of the combined method has increased, which may be because of the more detailed features, so it takes a long time. Compared with other methods, the time difference is small and the accuracy is high. Therefore, after comprehensive consideration, the improved combined method has better performance and has more advantages.

To further validate the superior performance of the proposed method, it is compared with the latest IS and recognition techniques. Five methods are independently run 10 times on a self-built dataset. The accuracy and average recognition time obtained are shown in Table 2.

Table 2

Comparison results between research methods and the latest IS techniques

Method Type	Run time/s	Accuracy/%
Medical IS technology [18]	6.37	92.65
Optimized multi-kernel FCM method [13]	5.92	91.07
Differential evolution [3]	7.83	93.55
Image processing-based system [19]	7.06	93.01
GCCV-CNN	3.44	98.96

The latest research methods in Table 2 are the optimized multi-kernel FCM method proposed in 2023 in literature [13], the Differential proposed in literature [3] evolution, the image processing based system proposed in 2023 in literature [19], and the research method proposed in this paper. From Table 2, the accuracy of the four methods in the reference is more than 90%, the maximum accuracy is 93.55%, and the running time is about 6s. The running time of the GCCV-CNN method proposed in this study is 3.44s, and the accuracy rate is 98.96%. This shows that GCCV-CNN has obviously better actual performance.

5. Conclusion

The promotion of national fitness has led to the rapid development of sports types. The subsequent sports injuries have gradually increased. Using modern technology to analyze sports injury has become a new mainstream trend. Visual image capture technology is used as a starting point to analyze the image recognition technologies it encompasses. Firstly, a multi-scale AM is introduced into the U-Net framework. An IS model is established. Then the Grad-CAM technology is embedded to improve the CNN medical image classification algorithm. Finally, the segmentation model and classification algorithm are combined to

recognize the sports injury image. The results show that the proposed IRS-Net network segmentation model is 0.50%, 0.37% and 0.23% higher than U-Net, CA-Net and DO-UNet in accuracy indicators. The highest AUC is 98.93%. The loss coefficient of the improved GCCV-CNN algorithm gradually fluctuates around 0 after the 10th iteration in the training set, the maximum value is only 0.4, and the waveform does not oscillate. The loss coefficient approaches 0 after 5 iterations in the test set. After taking the IRS-Net segmentation model as the preprocessing step of GCCV-CNN algorithm, the accuracy rate basically remains above 95%, the maximum can reach 98%, and the minimum running time is close to 3s. It shows that the IS and recognition method proposed in the research has higher accuracy and better comprehensive performance. The

proposed IS and recognition method has high accuracy and excellent overall performance. At the same time, it also shows that the application potential of this technology is of great significance for improving the rehabilitation effect of Sports injury. In the future, personalized and customized virtual rehabilitation applications can be studied according to the individual characteristics of patients, such as the type of injury, severity and specific rehabilitation needs. Visual image capture technology is used to promote the simulation research of sports injury rehabilitation. It improves the effectiveness, efficiency and accessibility of rehabilitation practice. However, the research does not pay enough attention to the speed of image recognition. The efficiency of image processing needs to be continuously improved in subsequent research.

References

1. Abumalloh, R. A., Nilashi, M., Ismail, M. Y., Alhargan, A., Alghamdi, A., Alzahrani, A. O., Asadi, S. Medical Image Processing and COVID-19: A Literature Review and Bibliometric Analysis. *Journal of Infection and Public Health*, 2022, 15(1), 75-93. <https://doi.org/10.1016/j.jiph.2021.11.013>
2. Alalwan, N., Abozeid, A., ElHabshy, A. A. A., Alzahrani, A. Efficient 3D Deep Learning Model for Medical Image Semantic Segmentation. *Alexandria Engineering Journal*, 2021, 60(1), 1231-1239. <https://doi.org/10.1016/j.aej.2020.10.046>
3. Chakraborty, S., Saha, A. K., Ezugwu, A. E., Agushaka, J. O., Zitar, R. A., Abualigah, L. Differential Evolution and Its Applications in Image Processing Problems: A Comprehensive Review. *Archives of Computational Methods in Engineering*, 2023, 30(2), 985-1040. <https://doi.org/10.1007/s11831-022-09825-5>
4. Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., Haworth, A. A Review of Medical Image Data Augmentation Techniques for Deep Learning Applications. *Journal of Medical Imaging and Radiation Oncology*, 2021, 65(5), 545-563. <https://doi.org/10.1111/1754-9485.13261>
5. Cui, X., Zou, C., Wang, Z. Remote Sensing Image Recognition Based on Dual-Channel Deep Learning Network. *Multimedia Tools and Applications*, 2021, 80(18), 27683-27699. <https://doi.org/10.1007/s11042-021-11079-5>
6. Fazilov, S. K., Yusupov, O. R., Abdiyeva, K. S. Mammography Image Segmentation in Breast Cancer Identification Using the Otsu Method. *Web of Scientist: International Scientific Research Journal*, 2022, 3(8), 196-205.
7. Feng, R., Zheng, X., Gao, T., Chen, J., Wang, W., Chen, D. Z., Wu, J. Interactive Few-Shot Learning: Limited Supervision, Better Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 2021, 40(10), 2575-2588. <https://doi.org/10.1109/TMI.2021.3060551>
8. Girum, K. B., Créhange, G., Lalande, A. Learning with Context Feedback Loop for Robust Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 2021, 40(6), 1542-1554. <https://doi.org/10.1109/TMI.2021.3060497>
9. Gu R, Wang G, Song T. CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation. *IEEE*, 2021, 40(2), 699-711. <https://doi.org/10.1109/TMI.2020.3035253>
10. Guo, Z., Shen, Y., Wan, S., Shang, W. L., Yu, K. Hybrid Intelligence-Driven Medical Image Recognition for Remote Patient Diagnosis in Internet of Medical Things. *IEEE Journal of Biomedical and Health Informatics*, 2021, 26(12), 5817-5828. <https://doi.org/10.1109/JBHI.2021.3139541>
11. Huang T, Liu Y. Research on the Magnetic Resonance Imaging Brain Tumor Segmentation Algorithm Based on DO-UNet. *International Journal of Imaging Systems and Technology*, 2023, 33(1), 143-157. <https://doi.org/10.1002/ima.22783>
12. Jiang H, Jiang A, Xu M. Research on Algorithm and Coupling Damage Model of Rock Under High Temperature.

- re and Loading Based on Mohr-Coulomb Criterion. *International Journal of Computational Methods*, 2022, 19(4), 2250001-2250023. <https://doi.org/10.1142/S0219876222500013>
13. Kollem, S., Prasad, C. R., Ajayan, J., Malathy, V., Subbarao, A. Brain Tumor MRI Image Segmentation Using an Optimized Multi-Kernel FCM Method with a Pre-processing Stage. *Multimedia Tools and Applications*, 2023, 82(14), 20741-20770. <https://doi.org/10.1007/s11042-022-14045-x>
 14. Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., ... & Acharya, U. R. Transfer Learning Techniques for Medical Image Analysis: A Review. *Bio-cybernetics and Biomedical Engineering*, 2022, 42(1), 79-107. <https://doi.org/10.1016/j.bbe.2021.11.004>
 15. Luo, X., Chen, J., Song, T., Wang, G. Semi-Supervised Medical Image Segmentation Through Dual-Task Consistency. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021, 35(10), 8801-8809. <https://doi.org/10.1609/aaai.v35i10.17066>
 16. Monga, V., Li, Y., Eldar, Y. C. Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing. *IEEE Signal Processing Magazine*, 2021, 38(2), 18-44. <https://doi.org/10.1109/MSP.2020.3016905>
 17. Müller, D., Kramer, F. MIScnn: A Framework for Medical Image Segmentation with Convolutional Neural Networks and Deep Learning. *BMC Medical Imaging*, 2021, 21(1), 1-11. <https://doi.org/10.1186/s12880-020-00543-7>
 18. Müller, D., Soto-Rey, I., Kramer, F. Towards a Guideline for Evaluation Metrics in Medical Image Segmentation. *BMC Research Notes*, 2022, 15(1), 1-8. <https://doi.org/10.1186/s13104-022-06096-y>
 19. Rahman, S. U., Alam, F., Ahmad, N., Arshad, S. Image Processing Based System for the Detection, Identification and Treatment of Tomato Leaf Diseases. *Multimedia Tools and Applications*, 2023, 82(6), 9431-9445. <https://doi.org/10.1007/s11042-022-13715-0>
 20. Raza, K., Singh, N. K. A Tour of Unsupervised Deep Learning for Medical Image Analysis. *Current Medical Imaging*, 2021, 17(9), 1059-1077. <https://doi.org/10.2174/1573405617666210127154257>
 21. Saood, A., Hatem, I. COVID-19 Lung CT Image Segmentation Using Deep Learning Methods: U-Net Versus SegNet. *BMC Medical Imaging*, 2021, 21(1), 1-10. <https://doi.org/10.1186/s12880-020-00529-5>
 22. Sarvamangala, D. R., Kulkarni, R. V. Convolutional Neural Networks in Medical Image Understanding: A Survey. *Evolutionary Intelligence*, 2022, 15(1), 1-22. <https://doi.org/10.1007/s12065-020-00540-3>
 23. Li, Y., Sixou, B., Peyrin, F. A Review of the Deep Learning Methods for Medical Images Super Resolution Problems. *Irbm*, 2021, 42(2), 120-133. <https://doi.org/10.1016/j.irbm.2020.08.004>
 24. Suganyadevi, S., Seethalakshmi, V., Balasamy, K. A Review on Deep Learning in Medical Image Analysis. *International Journal of Multimedia Information Retrieval*, 2022, 11(1), 19-38. <https://doi.org/10.1007/s13735-021-00218-1>
 25. Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., & Nandi, A. K. Medical Image Segmentation Using Deep Learning: A Survey. *IET Image Processing*, 2022, 16(5), 1243-1267. https://doi.org/10.1007/978-3-030-32606-7_2
 26. You, C., Zhou, Y., Zhao, R., Staib, L., Duncan, J. S. Simcvd: Simple Contrastive Voxel-Wise Representation Distillation for Semi-Supervised Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 2022, 41(9), 2228-2237. <https://doi.org/10.1109/TMI.2022.3161829>
 27. Ziller, A., Usynin, D., Braren, R., Makowski, M., Rueckert, D., Kaissis, G. Medical Imaging Deep Learning with Differential Privacy. *Scientific Reports*, 2021, 11(1), 1-8. <https://doi.org/10.1038/s41598-021-93030-0>

