Received August 12, 2020, accepted August 21, 2020, date of publication August 31, 2020, date of current version September 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3020322

Image Recognition and Analysis of Intrauterine Residues Based on Deep Learning and Semi-Supervised Learning

TAO TAO, KAN LIU, LI WANG, AND HAIYING WU
Department of Gynecology and Obstetrics, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University, Zhengzhou 450003, China
Corresponding author: Haiying Wu (whysunnyzg@163.com)

ABSTRACT Residual placenta is one of the common types of postpartum complications in clinical practice. Residual placenta is also the main and most common cause of late postpartum hemorrhage. This article proposes a spatial pyramid loop module, which solves the problem that the existing network structure cannot effectively extract the semantic information and category information in the image at the same time. The spatial pyramid structure is used to effectively extract the semantic information and category information. In addition, this article proposes to use cyclic convolutional network to realize the transfer function of information at different scales, and build it in the spatial pyramid structure to further strengthen the ability to extract semantic information and category information. This article proposes a feature fusion module to solve the impact of image classification network used in the base network in the existing network structure. The attention mechanism is used to achieve the effective fusion of high-dimensional features and low-dimensional features in the base network to reduce the influence of the base network, so as to better recover the recognition and prediction results. A semantic category loss function is proposed to supervise the categories of objects in images. This article builds it on the feature layer with the smallest scale, which not only increases the intermediate supervision to make the network fully converge, but also reduces the difficulty of extracting category information, and makes full use of the information transfer function of the cyclic convolutional network. This article introduces uncertainty information into the field of image segmentation to provide the accuracy of segmentation. For the purpose of uncertainty information, this article improves the network structure. At the same time for the image segmentation task, this article improves the Bayesian cross entropy loss function. The experiment verifies the necessity of improving the Bayesian crossover function in this article and the effectiveness of the conditional random field used in this article, and also verifies the effectiveness of the proposed semi-supervised learning method.

INDEX TERMS Deep learning, semi-supervised learning, image recognition, intrauterine residue.

I. INTRODUCTION

The placenta will be delivered from the birth canal 5-15 minutes after the fetus is delivered in a natural childbirth. If the placenta is not completely delivered out of the body, and there are still some tissues remaining inside the mother's uterus, it is called postpartum remnant placenta [1]. There is also a view that residual placenta refers to the phenomenon that a part of the placenta remains in the uterus after 0.5h after delivery of the fetus during the third stage of delivery [2]. Residual placenta is a common complication of natural childbirth after artificial abortion. The placental tissue remaining in the uterus is in most cases relatively small fragments, which is difficult to be found. Residual placenta leads to postpartum hemorrhage, uterine cavity infection, endometritis, and abdominal pain. If it is not treated in a timely, effective and thorough manner, adhesions and organization will occur, which will bring great pain to the patient [3]. At the same time, it seriously threatens the life safety and quality of life of women of childbearing age, and even leads to secondary amenorrhea or intrauterine adhesions, which affects reproductive health [4]. It also increases the medical burden and affects the postpartum rehabilitation process.

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

Diagnosis of postpartum residual placenta is not complicated, and the medical history has an important reference role, which can indicate whether the placenta delivered after delivery is intact [5]. The clinical manifestations of residual placenta are different. In mild cases, it can be manifested as intermittent vaginal bleeding after delivery or sudden small amount of bleeding and secondary anemia [6]. Physical examination can find that the cervix becomes soft and loose, the enlarged uterus becomes soft, the uterus is poorly restored, and the uterine cavity patients with internal infections have obvious tenderness in the lower abdomen, and those with severe conditions can threaten their lives due to massive bleeding [7]. When the clinical symptoms of residual placenta are not typical, it is easy to be missed. B-ultrasonic examination of residual placenta has the advantages of simple operation, high diagnostic accuracy, convenient and quick, etc. It can clearly scan the tissue in the uterus and is the preferred imaging method for the diagnosis of placental residual [8]. B-ultrasonography shows the structure of placenta lobules, which is ring-shaped with strong echo [9]. The light mass seen by B-ultrasound is mostly located in the uterine cavity and does not invade the myometrium, unless part of the placenta is implanted. Most patients with short course of disease have irregular medium and low echoes. With the course of disease, patients with longer course of disease have tissue degeneration, necrosis, organization, echo enhancement, and rough spots of light [10]. B-ultrasound can also know whether there is blood accumulation, endometritis and uterine involution in the uterine cavity [11].

Machine learning is mainly divided into three categories: supervised learning, unsupervised learning and semi-supervised learning [12], [13]. Among them, supervised learning and unsupervised learning are relatively frequently encountered, and semi-supervised learning is less exposed than the first two methods [14]. Semi-supervised learning is a combination of supervised and unsupervised learning. Its main idea is to use a small amount of labeled data and a large amount of unlabeled data to train a learner with better results [15]–[17]. With the gradual improvement of the research system of semi-supervised learning methods, and the advantages of using unlabeled data for better learning performance gradually appear, the related theoretical results of semi-supervised learning are also applied in practical problems [18], [19]. One of the typical fields and the more widely used is the related applications of natural language processing. Related scholars have proposed an active semi-supervised syntactic analysis method based on collaborative training, which can effectively reduce the workload of manual marking by about half [20]–[22]. Another very important application field of semi-supervised learning is the application of content-based image retrieval [23], [24]. This method can effectively improve the performance of image retrieval. The generative model is a semi-supervised learning method that appeared early. It assumes that all complete data comes from a limited mixed model, and treats all labeled samples as complete data, treats all unlabeled samples as incomplete data. The number is the total number of categories [25]. The goal of the model is the maximum value of the likelihood function. Usually maximum likelihood estimation is performed, and then each component parameter of the mixed model is obtained [26]–[28]. Each category in the generative model algorithm only needs one labeled sample to determine the mixed model [29], [30]. A single-layer feedforward neural network can represent any function, but the actual number of hidden units required increases exponentially and may be too large to be realized. Using deeper neural networks can reduce the number of hidden units required, but deep neural networks have problems such as obvious gradient disappearance and are difficult to train [31], [32]. There is evidence that, although the stochastic gradient descent algorithm with adaptive learning rate can show fast convergence speed in the early stage of optimization, its performance in the test quickly stagnates and is finally used by the basic or momentum-added stochastic gradient descent algorithm [33], [34]. In order to achieve semi-supervised image semantic segmentation, many researches use these images by giving unlabeled images a pseudo-label, and then add labeled data sets for training, so as to improve network performance. In order to obtain pseudo-labeled images, weak labeling information is often added to such methods. In image segmentation, weak annotations can be divided into image-level annotations, object-level annotations, and stroke-level annotations. Among them, image-level annotation has the least cost, so most of the methods for semi-supervised learning based on weak annotations use image-level label.

This article proposes a multi-scale cyclic network, and designs a spatial pyramid cyclic module and a feature fusion module to improve the accuracy of the network. In addition, the benefits brought by the information transmission process at different scales are explored. Next, this article studies the loss function and proposes a semantic category loss function. This loss function can be used as an intermediate supervision and solves the problem that the existing convolutional neural network cannot effectively process large and small objects at the same time. Specifically, the technical contributions of this article can be summarized as follows:

First: A semi-supervised spectral clustering algorithm based on deep learning is proposed. Aiming at the problem of how to introduce prior information in the clustering algorithm, this algorithm proposes to construct the similarity matrix of spectral clustering by using the distance learned by the paired constraint information, and uses the constrained K-means clustering algorithm to map the spectrum. The feature vector is clustered. The algorithm makes full use of the prior information of the data and improves the stability of the algorithm.

Second: This article sorts out the data set of fully-supervised semantic recognition, and establishes a data set of semi-supervised semantic recognition. In the experiment of fully supervised semantic recognition, multiple sets of comparative experiments were carried out, including the comparative experiment of using different times of downsampling in the spatial pyramid cycle module, and the improved modules proposed in this article were added to the multi-scale cyclic convolution network and Deeplab v3+. These experimental results fully illustrate the effectiveness and transferability of the spatial pyramid loop module, loop neural network, feature fusion module and semantic category loss function proposed in this article.

Third: In the semi-supervised semantic recognition experiment, multiple sets of comparative experiments were also carried out, including comparative experiments with different amounts of data, comparative experiments with different trade-off coefficients in the Bayesian cross-entropy function, and Bayesian crossover. These experiments fully illustrate the necessity of improving the Bayesian crossover function and the effectiveness of semi-supervised learning in this article.

The rest of this article is organized as follows. Section 2 discusses related theories and technologies. Section 3 builds an image recognition model based on semi-supervised feature extraction of deep learning. Section 4 presents the experimental results and experimental analysis. Section 5 summarizes the full text.

II. RELATED THEORIES AND TECHNOLOGIES

A. IMAGE FEATURE DESCRIPTION

Feature extraction is a step in pattern recognition, which refers to the use of algorithms to transform the measured value of a pattern. The extracted features are usually representative, repeatable and distinguishable. Commonly used image features include color features, shape features, texture features, and spatial relationship features.

Color feature refers to the color model established based on all pixels of the picture, which describes the overall nature of the image area. Color features are intuitive and efficient, and can distinguish images based on colors; however, changes in the angle and scale of the image are ignored, and the local features of the target cannot be described well. In addition, tracking using only color features is susceptible to interference from objects with similar colors in the background. Color histogram is the most common method of describing color features, which is invariant to image rotation and translation changes, but the disadvantage is that it cannot describe the spatial distribution of colors.

The shape feature describes the boundary information of the object in the image, and the result is intuitive and easy to understand. This feature describes the boundary of the target through a mathematical model, thereby distinguishing the target and the background; when the shape of the target changes, the results often have various degrees of deviation, and many shape features are only a description of the target part, because the overall shape of the target is often described.

Texture feature describes the surface characteristics of a certain area of an image, is a statistical feature based on pixels in a certain area, and is a regional feature. As a statistical feature, texture features are rotation invariant and have a certain anti-interference ability.

If there are multiple targets in an image, the mutual spatial position or relative direction relationship between the targets can be extracted as the spatial relationship feature. The spatial position relationship is divided into two categories: relative position relationship and absolute position relationship. The spatial relationship feature is easily disturbed by image rotation and target reversal. Therefore, the spatial relationship feature needs to be used in combination with other features in practical applications.

Local features refer to a class of algorithms that describe the local features of a target. Local features solve the problem that global features cannot reflect the local information of objects. Local features usually focus on the moving object itself, regardless of the characteristics of the background. The local features of an object refer to the corner points of the object area that appear locally and stably. These corner points often have good distinguishability. When the target is occluded, because some local features are still stable, they can still be distinguished by local features aims.

B. FEATURE DESCRIPTION BASED ON DEEP LEARNING

The essence of the machine learning model is to map data to a feature space that is more conducive to classification. However, with limited samples and computational costs, the above-mentioned shallow structure cannot accurately simulate a complex nonlinear mapping. There are certain restrictions when the two-classification problem occurs, and it is prone to overfitting. Deep learning uses multiple linear functions to iteratively approximate complex nonlinear functions by training deep network structures, showing a powerful ability to learn the essential characteristics of data sets. The image detection and recognition algorithm process of intrauterine residues based on the deep multi-core convolutional neural network model is shown in Figure 1.

There are n layers in deep learning, the input of each layer is the output result of the previous layer, and the layers are connected by a weight matrix. In fact, each layer of calculation can be seen as a layer of data mapping, and the output of each layer is another expression of the input data. By adjusting the parameters, the output of the network result is closest to the input data, and then a series of hierarchical features of the input data are obtained.

The large number of unlabeled samples collected are divided into many sub-sample sets containing a certain number of samples according to the size of the sample set, and the size of each sub-sample set can be adjusted according to the total number of samples. The input of the first layer is V, and the output is H. Back-propagating H is to obtain the observed value V'of H. The error between the actual value and the measured value e=V-V' can be obtained. After each layer of training, the output of the previous layer is used as input, and the parameters are adjusted by repeated training. Finally, you stack the trained shallow models to form the initialized deep model.

Based on the Wake-Sleep algorithm, you further fine-tune the weight coefficients of each depth model. The sleep phase

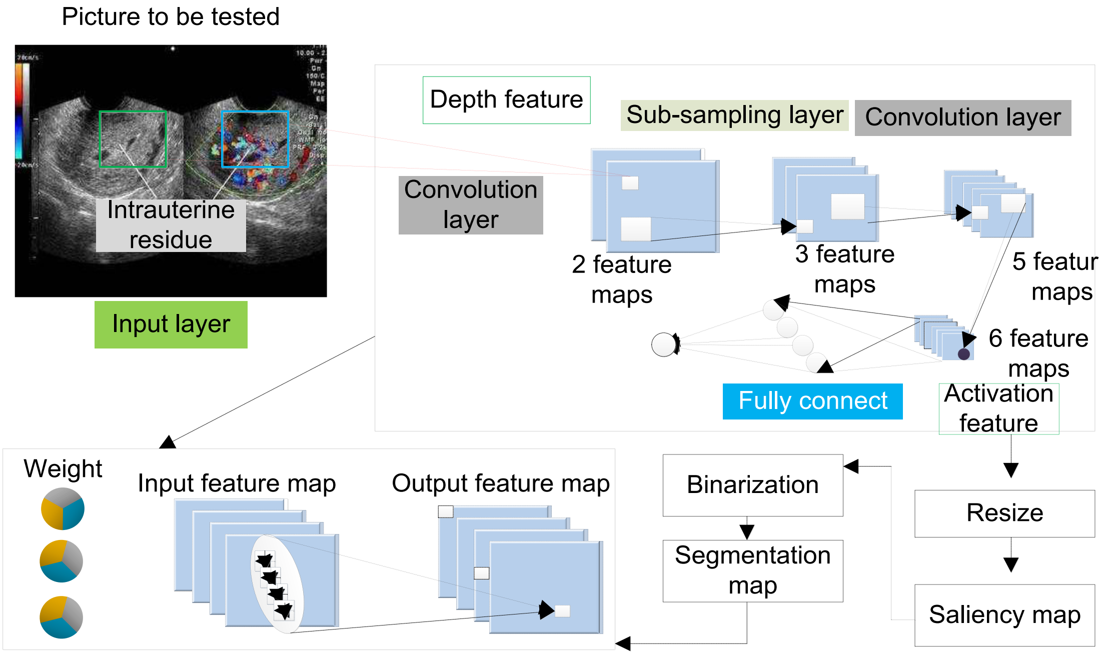


FIGURE 1. Intrauterine residue image detection and recognition algorithm process based on deep multi-core convolutional neural network model.

is the process of bottom-up propagation of excitation. The output of each layer is calculated by input and connection weights. It is a process from concrete to abstract; the wake phase is the process of error transmission from top to bottom. The weight matrix between layers is a process from abstract to concrete. The main purpose of Wake-Sleep is to make the abstract output generated upward through the sleep phase to restore the input through the wake phase.

C. GENERATIVE MODEL ANALYSIS

The main advantage of self-learning is its simplicity and effectiveness. Self-learning is actually a "wrapper", which means that any applicable model can be fitted with self-learning algorithms, such as simple nearest neighbor algorithms or more complex deep networks. Self-learning implies a very important assumption: the high-confidence label predicted by the model is likely to be correct. This is very likely to be the case when samples of the same class tend to cluster, so the self-learning model implies a smooth hypothesis.

However, even under this premise, the self-learning model often does not have good generalization performance in the initial training model with a small amount of data, so high-confidence samples are also easy to make mistakes. The samples are added to the training. Without additional supervision, the model will "reinforce" its own mistakes, so the final model usually makes mistakes on fuzzy samples.

In the classification problem, the machine learning model predicts the label of the sample. From the perspective of probability, the goal of learning is to estimate the conditional distribution. Most methods directly use parametric models for modeling, and use labeled data for training. Such models are called discriminant models. Another method is to infer the conditional probability according to the Bayesian formula. Such a model is called a generative model. Many generative models are only an approximation of the real data distribution, and the optimization and reasoning complexity are high, and the discriminant model can be relatively simple to directly learn the classification boundary, so when the amount of data is sufficient, the discriminant model usually performs in the prediction accuracy. However, when the amount of data is insufficient, directly training the discriminant model with a small number of labeled samples can easily lead to overfitting; and the generative model will make specific assumptions about the data generation process during the modeling process. This assumption is used in model learning. At the same time, the generative model also has the ability to model the complex relationship between the observed variables and the hidden variables. The priori of the problem structure or data generation can be naturally introduced into the modeling process, which makes the generative model suitable for semi-supervised learning. The schematic diagram of the tool chain of deep learning combined with semi-supervised learning algorithm is shown in Figure 2.

1) HYBRID MODEL AND EXPECTATION MAXIMIZATION

The Gaussian mixture model assumes that the data is a mixture of several Gaussian distributions. Each Gaussian distribution represents a category, and the probability that each sample belongs to each Gaussian analysis is used as a priori. This prior can be used for prediction by a classifier trained

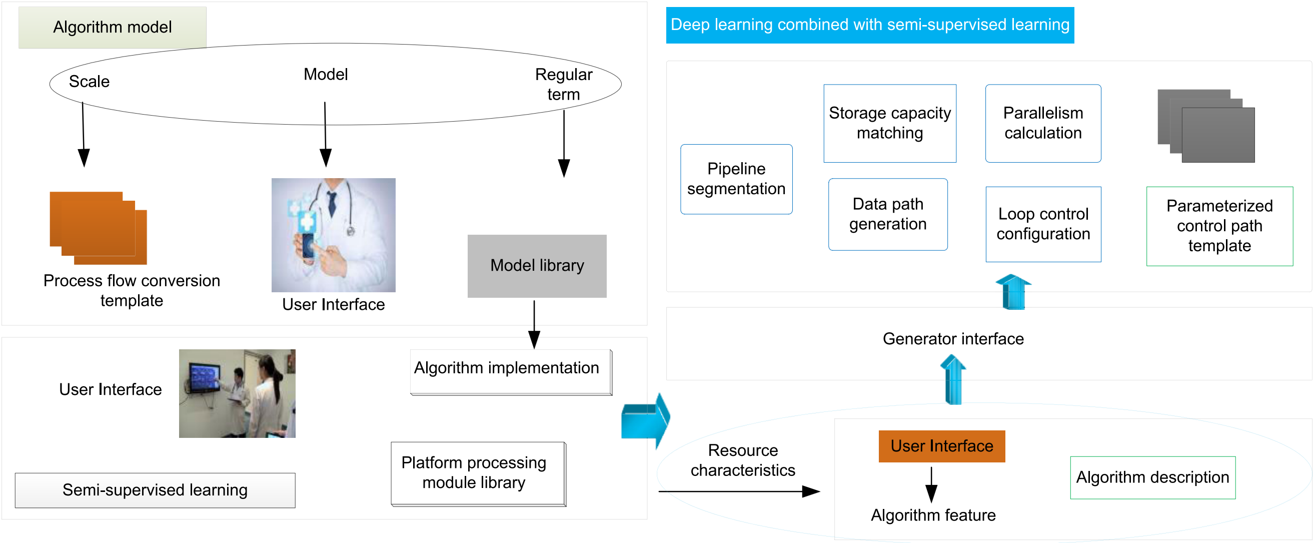


FIGURE 2. Schematic diagram of the tool chain of deep learning combined with semi-supervised learning algorithm.

on the labeled data, or directly initialized with a clustering algorithm.

The expectation-maximization algorithm can also be regarded as a special case of self-learning. Unlike self-learning that gives each sample a fixed pseudo label, the expectation-maximization algorithm estimates the probability that each sample belongs to each component in the mixed model.

The hybrid model needs to be identifiable, which means that only one model parameter can explain the current observable data. When the unlabeled data gradually increases, the optimized model can restore the accurate hybrid model. For example, the Gaussian mixture model is identifiable, but the homogeneous mixture model is not.

2) CLUSTER AND LABEL

A more direct method is to cluster the data and then label them based on the labeled samples. First, you cluster all the data. For each cluster obtained by clustering, you train a classifier based on the labeled samples in the cluster, and classify the unlabeled samples in the cluster according to this classifier; if there is no labeling samples, you use all labeling samples to train a classifier, and use this classifier for classification. Such an algorithm can exert better performance only when the adopted clustering algorithm just matches the true distribution of the data.

D. SEMI-SUPERVISED METHOD BASED ON LOW-DENSITY SEPARATION

1) DIRECT PUSH SVM

Although the direct push SVM implements the idea of direct push learning, the learned classifier can actually be used outside the training sample. Direct push SVM is an extension of standard SVM. The goal of standard SVM is to find a maximum separation hyperplane of labeled data in the regenerated kernel Hilbert space. In the direct push SVM, in addition to labeled data, there is also unlabeled data. The goal is to find a separating hyperplane so that there is a maximum gap between labeled data and pseudo-labeled unlabeled data. Intuitively, it is necessary to make all data fall outside the separation hyperplane as much as possible, so that the low-density area is within the boundary. Such a separating hyperplane has the smallest generalization error bound on unlabeled data.

Specifically, the objective function of direct SVM optimization can be seen as adding a regular term for unlabeled data to the standard SVM, taking linear SVM as an example:

$$SVM = \min_{f} \sum_{i=1}^{l} \max(1 - f(x_i), y_i) + \max(1 - |f(x_i)| , 0) \tag{1}$$

The last item is the regularization item for unlabeled data. This regularization term is non-convex.

2) GAUSSIAN PROCESS

In the discriminant model, according to the d-separation theory, when the label is unknown, the data and the model parameters are independent of each other. Therefore, the usual way to use unlabeled data in the discriminant model is to add regular terms. Another way is to expand the probability graph model and add a permanently observable leaf node to the label.

E. GRAPH-BASED METHODS

The graph-based method builds a graph based on the similarity between samples on the labeled and unlabeled data, and assumes that the labels on the graph are smooth. Its optimization function usually contains two items: one is the loss function on the labeled data, and the other is regularization, usually a smoothness constraint. Most graph-based methods differ only in the two choices. In addition, graph construction and computational complexity are also two key issues in practical applications.

1) MINIMUM CUT

This method regards semi-supervised classification as a network flow segmentation problem and uses the minimum cut algorithm to solve it. Specifically, only the two-class classification problem is considered. The positive samples are regarded as the source vertices and the negative samples are regarded as the convergent vertices. The goal is to find a set of edges with the smallest sum of weights, so that these edges can be removed from the source vertex to the convergence vertex. According to the segmentation result, the vertices connected to the source vertex are marked as positive samples, and the vertices connected to the convergent vertices are marked as negative samples. The minimum cut algorithm can be seen as optimizing the following objective function:

$$y = \sum_{i,j=1}^{l+u} w_{ij}(f(x_i) - \left|f(x_j)\right|) + \sum_{i=1}^{L} f(x_i) \tag{2}$$

The first term is the loss function, that is, the prediction on the labeled data is required to be consistent with the real label, otherwise there will be an infinite loss; the second term is a regular term, which is actually a minimum cut.

2) GAUSSIAN RANDOM FIELD AND HARMONIC FUNCTION

This method is actually an extension of the minimum cut algorithm. Although the objective form of optimization is the same, in the minimum cut algorithm, f(x) can only take 1, and the Gaussian random field is defined in a continuous state space, which allows to take real value. Such relaxation makes the optimal solution of the graph the only closed-form solution, which can be described by the harmonic function. Specifically, you let W denote the edge weight matrix and D be the diagonal matrix, then the unnormalized graph Laplacian matrix is L=D-W.

Note that there is also a constraint that the predicted label on the labeled data should be equal to the true label. You divide the Laplacian matrix into blocks with labeled data and unlabeled data:

$$L = \begin{bmatrix} L_{ll} & L_{lu} \\ L_{ul} & L_{uu} \end{bmatrix} \tag{3}$$

3) MANIFOLD REGULARIZATION

Both the minimum cut algorithm and the harmonic function belong to direct learning, and the labels on the labeled data are limited to the given real labels, which cannot handle the case of real labels with noise. The method based on manifold regularization solves these two problems. The goal is to make the solution smooth with respect to the surrounding space and edge distribution.

III. IMAGE RECOGNITION MODEL CONSTRUCTION BASED ON SEMI-SUPERVISED FEATURE EXTRACTION OF DEEP LEARNING

A. IMAGE FEATURE EXTRACTION

1) IMAGE FEATURE EXTRACTION BASED ON GRAY LEVEL CO-OCCURRENCE MATRIX

The Gray Level Co-occurrence Probabilities (GLCP) method of extracting texture features is currently recognized as an important texture analysis method. It is defined by the joint probability density of pixels in two positions, which reflects the comprehensive information of the image gray level on the direction, the adjacent interval, and the amplitude of change. It can be used as the basis for analyzing the local pattern structure and arrangement rules of the image.

Given a two-dimensional image with a size of and a gray level of G, the gray level co-occurrence matrix that satisfies a certain spatial relationship is:

$$P(i,j) = \#\left[(x_1, y_1), (x_2, y_2) \in M * N\right] \tag{4}$$

where #{x} represents the number of elements in the set x, and P is a G∗G matrix. If the distance between pixels (x_1, y_1) and (x_2, y_2) is d, the angle between the two and the horizontal axis of the coordinate is θ, the gray level co-occurrence matrix of various pitches and angles can be obtained.

As the feature vector of texture analysis, the calculated gray-level co-occurrence matrix is often not directly used, but the texture feature is extracted on the basis of the gray-level co-occurrence matrix, which is called secondary statistics. Generally, the following secondary statistics can be extracted from the gray level co-occurrence matrix of the image as the feature coefficients for classification and recognition.

The second-order moment of angle reflects the uniformity of image gray distribution and the thickness of texture. Looking at the image as a whole, when the image texture is thicker, the ASM value is larger, and vice versa, the ASM value is smaller. Since the angular second moment is the sum of squares of the co-occurrence matrix elements, it is also called energy. The coarse texture ASM value is larger, which can be understood as the coarse texture contains more energy, and the fine texture ASM value is smaller, that is, it contains less energy.

Entropy is a measure of the amount of information an image has, and texture information also belongs to the information of an image, which is a measure of randomness. If the image has no texture, the gray-level co-occurrence matrix is almost zero, and the ENT value is almost zero. If the image is full of fine texture, the values of the co-occurrence matrix are approximately equal, and the ENT value of the image is the largest.

2) IMAGE ENERGY FEATURE EXTRACTION BASED ON WAVELET TRANSFORM

The decomposition of the two-dimensional image can be obtained by one-dimensional filtering along the x direction and the y direction respectively. Therefore, the orthogonal wavelet decomposition of an image can be understood as a signal decomposition on a set of independent spatially directed frequency channels. Each scale is decomposed into four sub-bands LL, HL, LH and HH, which represent the low-frequency information of the image and the details in the horizontal, vertical and oblique directions respectively.

The energy information of each subband decomposed by wavelet can describe the texture feature of the image well, and it is more stable than other subband coefficient statistical features. For an image with a size of N*M, the formula used to calculate the energy is as follows:

$$E = \frac{1}{M * N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \tag{5}$$

where f(x,y) is the wavelet coefficient of the xth row and yth column in the subband. In this chapter, we perform a three-layer wavelet transform on the image, extract the energy features of the 10 subbands, and obtain a 10-dimensional feature vector.

B. MULTI-SCALE CYCLIC NETWORK DESIGN

In order to solve the above-mentioned problems in the existing image recognition network structure, this article proposes the Multi-Scale Recurrent Network (MSR-net). The multi-scale cyclic convolutional network is an end-to-end image recognition network. Its input is an image of any size, and the output prediction result uses a codec structure. In the encoding process, the residual 101 network (Resnet101) is used as the base network, and the 16 times downsampling multiple is used to complete the encoding process; in the decoding process, the spatial pyramid loop module is proposed in the multi-scale loop network to extract the categories in the image Information and location information, as well as the feature fusion module to recover the prediction result image with the same size as the original image from the feature to complete the decoding process.

In order to extract the category information and location information in the image at the same time, this article designs the Spatial Pyramid Recurrent module (SPR). The spatial pyramid loop module adopts a spatial pyramid structure, that is, the image is transformed to features of different sizes, and the convolution operation is used to extract category information and location information. When the feature size is smaller, the convolutional layer is easier to extract category information; conversely, when the feature size is smaller, the convolutional layer is easier to extract location information. In addition, a circular convolutional network is added to the spatial pyramid loop module to increase the difference between dimensional features. In order to alleviate the impact of using the network structure of image classification as a base network, the Feature Fusion (FF) module is designed in the multi-scale recurrent network. In the feature fusion module, a combination of shallow features and high-dimensional features is added to the network structure, but instead of simply connecting directly in the channel dimension, it uses the attention mechanism to perform weighted fusion in the channel dimension. The image recognition flow chart of intrauterine residue based on semi-supervised deep learning is shown in Figure 3.

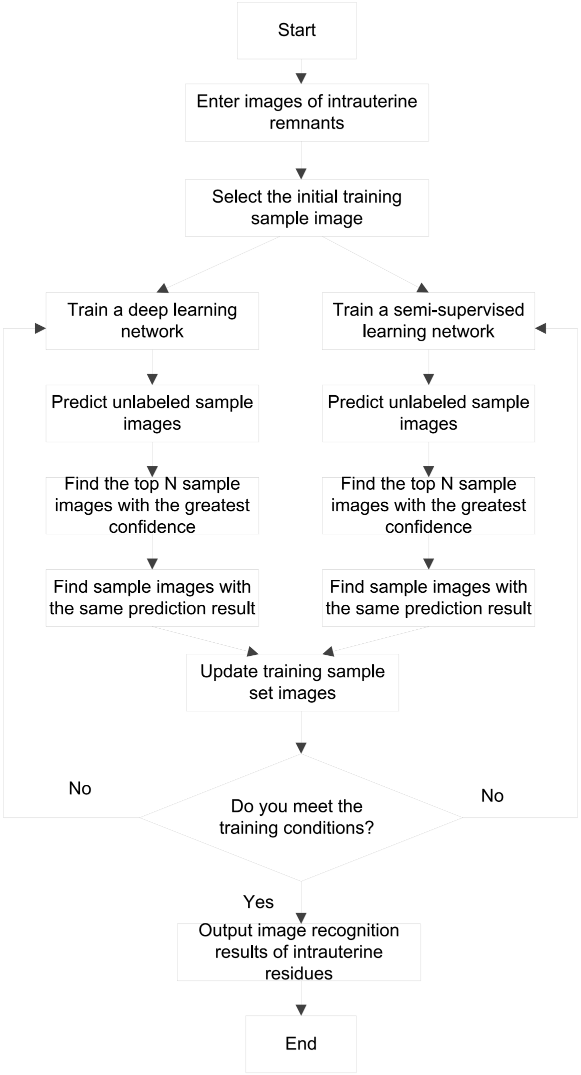

Start
Enter images of intrauterine remnants
Select the initial training sample image
Train a deep learning network
Train a semi-supervised learning network
Predict unlabeled sample images
Predict unlabeled sample images
Find the top N sample images with the greatest confidence
Find the top N sample images with the greatest confidence
Find sample images with the same prediction result
Find sample images with the same prediction result
Update training sample set images
No
Do you meet the training conditions?
No
Yes
Output image recognition results of intrauterine residues
End

FIGURE 3. Image recognition flow chart of intrauterine residue based on semi-supervised deep learning.

1) BASE NETWORK

The multi-scale recurrent network proposed in this article uses the residual 101 network as the base network. In the residual 101 network, first downsampling is implemented twice, where the first downsampling is achieved through a convolutional layer with a 7*7 step size of 2, and the second downsampling is implemented through a 5*5 maximum pool. Then through 4 blocks, including 3, 4, and 23 residual modules, each block is followed by a pooling operation until the network reaches the preset downsampling multiple. Finally, the global average pooling is used to unify the features to a scale of 1*1, and then through a fully connected layer and the Softmax function, the probabilities of 1000 categories are obtained. Different from the purpose of the standard residual 101 network, since this article does not require image classification, this article only takes the part of the standard residual 101 network before the global average pooling as the base network in the multi-scale recurrent network.

2) SPACE PYRAMID CYCLE MODULE

The Spatial Pyramid Recurrent (SPR) module is a module connected to the base network. It takes the features obtained from the base network as input and outputs a feature with both category information and location information to better solve the problem of image recognition.

In order to obtain several features of different scales, this article first adjusts the size of the input features. Specifically, this article is implemented by bilinear interpolation. The convolution kernels for all convolution operations are 3∗3, and the number of output channels is 256. According to the characteristics of the spatial pyramid structure, when the feature size is different and the convolution kernel size is the same, the smaller the feature scale, the larger the receptive field of the convolution operation relative to the original image size, the larger the feature scale is, the larger the convolution operation. The smaller the receptive field of the original image size, and the larger the receptive field in the convolution operation, the easier it is to extract category information. The larger the receptive field in the convolution operation, the easier it is to extract position information. Thus, the category information and location information in the image are extracted at the same time.

A cyclic convolutional neural network is added to the spatial pyramid cycle module to realize the transmission of information at different scales. Specifically, this article designs two forms. After the features obtained at each scale in the spatial pyramid structure, Uni-Direction RNN (UDR) and Bi-direction RNN (BDR) are added to obtain the spatial pyramid unidirectional loop module.

3) FEATURE FUSION MODULE

The feature fusion module is a module of the network decoding part. Its input consists of two parts and outputs the predicted segmentation result. In the feature fusion module, the features obtained by the spatial pyramid cycle module are first upsampled by 4 times according to the interpolation method, and then input to the feature fusion module together with the features obtained from the second block in the residual network, and the output is the same as the original image size 1/4 segmentation prediction result graph.

In order to recover the segmented prediction structure with better high-dimensional features, it is often necessary to add shallow low-dimensional features. In this process, the most common method is to directly connect high-dimensional features and low-dimensional features with the same feature size in the channel dimension, and then increase the complexity of the network through the convolutional layer. However, experiments show that this simple method cannot make full use of neural networks, resulting in the ineffective combination of low-dimensional features and high-dimensional features. Therefore, in order to better integrate low-dimensional features and high-dimensional features, and add low-dimensional detail information while ensuring high-dimensional semantic information, this experiment designed a feature fusion module based on the attention mechanism.

In the feature fusion module, first you connect the high-dimensional features and low-dimensional features in the channel dimension, then pass through several convolutional layers to obtain the feature F1, and then add an attention mechanism to the channel dimension of the feature F1. The attention mechanism is implemented in three parts. First, the weight W_1 of each channel to all channels is obtained, and then normalized by the Softmax function, and finally the weight is applied to the feature F_1 in the form of residual, as follows:

$$F_2 = F_1 * [1 + \sigma(W_1)] \tag{6}$$

Since the network will learn this weight autonomously during the training process, it can better realize the fusion of high-dimensional features and low-dimensional features. Finally, through a convolutional layer, a segmentation prediction result image of 1/4 of the original image size is obtained.

C. SEMI-SUPERVISED SPECTRAL CLUSTERING ALGORITHM BASED ON DEEP LEARNING

In semi-supervised clustering, the data information we get is generally data with class labels or paired constraint information. In many practical problems, it is easier to obtain class label information than to obtain paired constraint information, and the class label information can be transformed into paired constraint information, but not vice versa. Here, we use class-labeled data, and the obtained class-labeled data can be converted into pairwise constraints for calculation when learning distance.

The core step of the semi-supervised spectral clustering algorithm based on deep learning is the adjustment of the similarity matrix. The adjustment method is to learn the distance using paired constraint information, and then use the distance to measure the similarity between two samples. The deep learning neural network architecture is shown in Figure 4.

However, for a given data set containing n data samples, the size of the similarity matrix is n∗n, then the complexity of calculating the matrix eigenvector is $O(n^3)$. This makes it difficult to apply spectral clustering algorithms to large-scale data clustering problems like image recognition.

D. LOSS FUNCTION DESIGN

Loss function, also called cost function, refers to a function that maps an event to a real number that expresses the cost associated with the event. In deep learning, the loss function is usually associated with optimization problems as a learning criterion, and network parameters are estimated by minimizing the loss function.

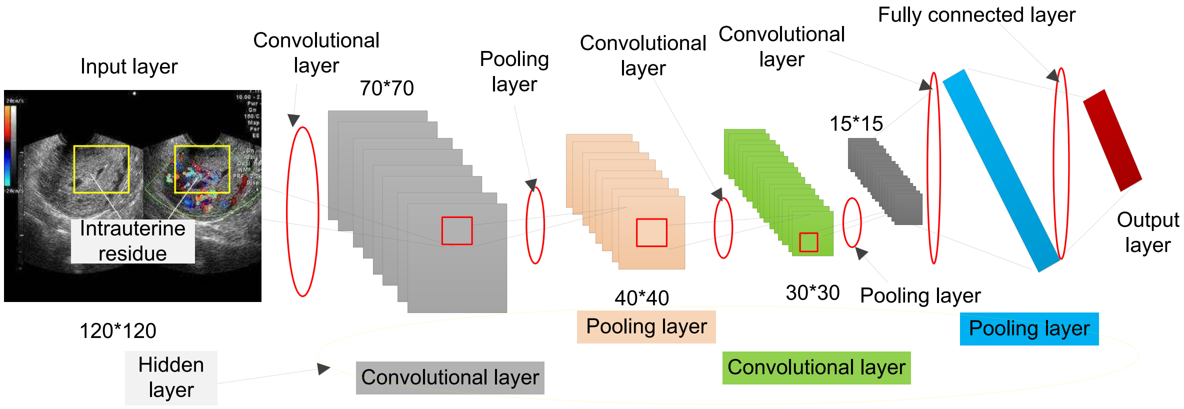


FIGURE 4. Deep learning neural network architecture.

In the training process of image recognition, the prediction result is often first normalized by the Softmax function for probability, and then the cross entropy function is used as the loss function to train the network parameters. This article also uses the above process in the training process, that is, first passes through the Softmax function, and then uses the multi-class cross entropy function as the loss function of the prediction result. However, in this article, a semantic category loss function is proposed to supervise the categories contained in the image.

In the image recognition task, a prediction result map is required, that is, a unique category is assigned to each pixel, and the probability of multiple categories can be obtained on each pixel after the network, which requires finite item discretization for multiple types of objects. The probability is normalized, and then the category with the highest probability is taken as the prediction result of the pixel. The normalization of the discrete probability of the finite item is realized with the help of the Softmax function, which is as follows:

$$P_j = e^{z_j} / \sum_{i=1}^{C} e^{z_i} \tag{7}$$

Among them, j is the current category and C is the total number of categories (C-1 foreground and background).

Cross entropy was first proposed in information theory. It means that if there are true probability distribution p and observation probability distribution q based on the same event measurement, then the cross entropy between them refers to the average code length when encoding the observation probability distribution q.

In image recognition tasks, each pixel needs to be assigned a unique category, then the truth map can be regarded as a discrete distribution, the probability value of the correct category corresponding to each pixel is 1, and the probability value of the misaligned category is 0. At the same time, the prediction result map is regarded as a discrete distribution, and the probabilities of all categories are given in each pixel. Then you use the expression of the discrete distribution of the cross entropy function as the loss function to obtain multiple types of cross entropy loss functions, and supervise the final prediction result graph to train the network parameters. Multi-type cross entropy functions are shown in the following formula:

$$Loss = \sum_{c=0}^{C} \sum_{i} \log(p_i) * q_i - \log(1 - p_i) * (1 - q_i) \tag{8}$$

where q is the true probability and p is the predicted probability.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL RESULTS OF FULLY SUPERVISED SEMANTIC RECOGNITION

1) TRAINING METHOD

In the training process, this article found that it is affected by batch processing operations. When the number of batches is larger, the training effect of the network is better, and when the input image size of the network is larger, the training effect of the network is better. However, whether increasing the number of batches or expanding the input image size of the network will occupy more video memory. Due to the limited video memory of the experiment, this article adopts a three-stage training strategy in the training process. In the first stage of training, the number of batches is set to 20, and the input image size of the network is set to 256. Its main purpose is to train the parameters in the batch normalization operation; in the second stage of training, the number of batches is set to 7, and the image size is set to 512, and the parameters in the batch normalization operation are frozen. The main purpose is to train parameters other than the batch operation; in the third stage of training, the batch number is set to 8, and the input image size of the network is set to 512, which also freezes the parameters in the batch normalization operation. Its main purpose is to prevent the network from falling into a local minimum and make the network further converge.

In each stage of training, a total of 40 rounds of training are carried out. This article adopts the momentum gradient

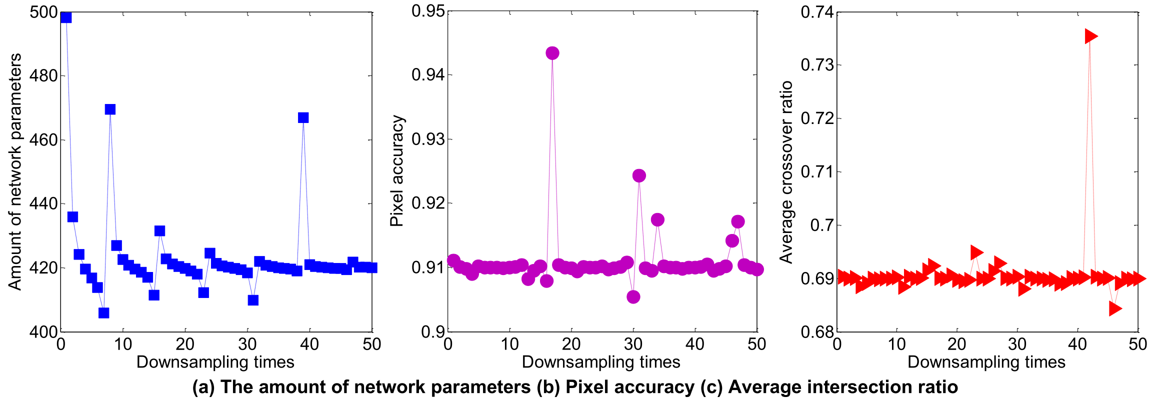


FIGURE 5. Comparison results of spatial pyramid unidirectional loop modules.

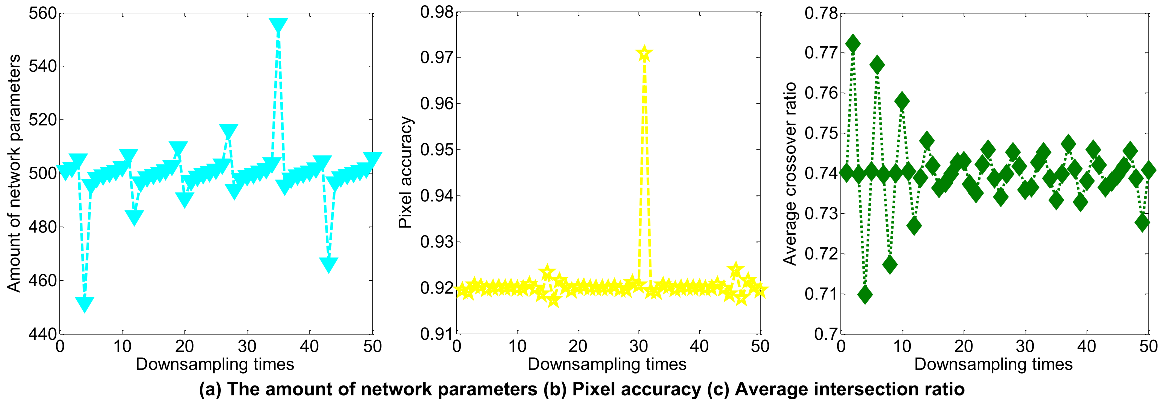


FIGURE 6. Comparison results of the two-way circulation module of the spatial pyramid.

descent method, in which the momentum factor is set to 0.9. At the same time, the learning rate adjustment strategy of polynomial decay is adopted, that is, the learning rate will decrease with the increase of the number of training iteration steps. The initial learning rate is set to 1e-3, and the minimum learning rate setting is reached when training for 30 rounds, and its power exponent is 0.88. In addition, in the training process, this article also adds regularization constraints. Regularization constraints refer to regularization of each parameter, which increases the sparsity of network parameters, which can effectively prevent network overfitting and improve the generalization ability of the network. The L2 norm regularization constraint adopted in this article is to find the square mean of all parameters, and finally set the weight of the regularization loss to 2e-4.

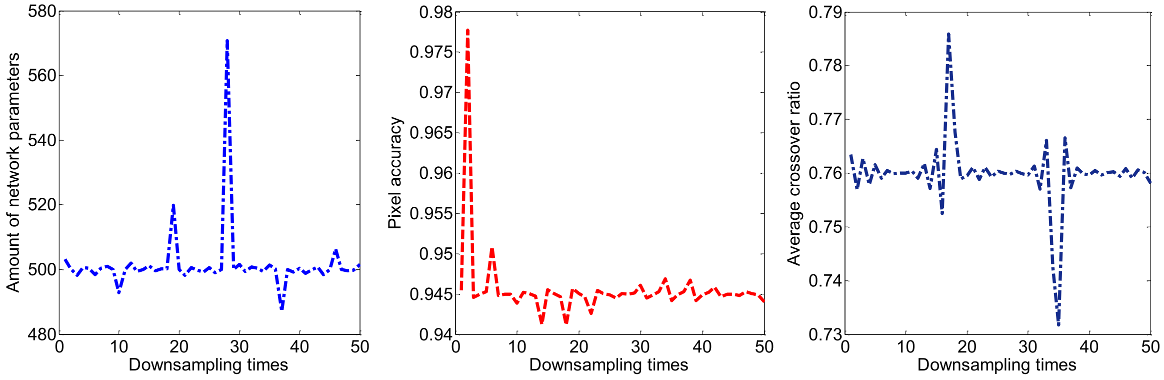
2) EXPERIMENTAL RESULTS AND ANALYSIS

This article first compares the performance of the network when using different times of downsampling in the spatial pyramid loop module module, and then compares the improvements brought by the improved module proposed in this article on the multi-scale loop network. In order to verify these comparative experiments, this article selects the medical data set as the verification data set, and finally this article gives the recognition effect in the real scene.

In order to compare only the performance of the network when different times of downsampling are used in the spatial pyramid recurring module, the multi-scale recurrent network built in this article uses a common feature fusion module, and only cross-entropy loss is used in the training process. The comparison of this part is divided into two parts: one-way cyclic network (SPUDR) and two-way cyclic network (SPBDR). The specific results are shown in Figure 5 and Figure 6.

Through comparison, it can be found that the more downsampling is used, the better the performance of the network. When 5 downsampling is used, the network has good performance. In addition, it can be found that the performance of the two-way loop network is generally better than that of the one-way loop network. Through image comparison and analysis, it can be found that the more down-sampling is used, the classification accuracy of its recognition fluctuates, which also conforms to the nature of the spatial pyramid structure.

The improved modules proposed in this article can be divided into three: spatial pyramid loop module, feature fusion module, and semantic category loss function. In order



(a) The amount of network parameters (b) Pixel accuracy (c) Average intersection ratio

FIGURE 7. Comparison results of improved modules.

to illustrate the role of these three improved modules, this article first builds a basic network that does not include these three modules. The spatial pyramid loop module removes the circular convolutional network and uses 50 downsampling operations, using ordinary feature fusion. The module removes the attention mechanism, and only uses the cross-entropy loss during the training process. Then we add these improvement modules in turn for comparison experiments. After adding the semantic category loss function, this article sets $\alpha 1$ and $\alpha 2$ in the overall loss function to 1.2 and 0.7, respectively. The specific results of the final comparison are shown in Figure 7.

Through comparison, it can be found that when the three improvement modules proposed in this article are added, the network performance has been improved to different degrees, and the improvement effects of these modules can be superimposed, especially when the feature fusion module and the semantic category loss function are added, the network parameters do not increase. With just a little increase, the network performance will be significantly improved. Through image comparison and analysis, it can be found that after the semantic category loss function is added, the network's ability to predict categories is significantly improved. After the feature fusion module is added, the network's ability to predict details has been improved.

This article adds the improved module to Deeplab v3+, which is currently recognized as the best-performing network structure, and explains the improvements brought by the improved module proposed in this article. This article first builds Deeplab v3+, and then adds these improvement modules in turn for comparison experiments, and adds semantic category loss functions. The comparison result of the improved module and Deeplab v3+ network is shown in Figure 8.

Through comparison, it can be found that when the three improved modules proposed in this article are added to Deeplab v3+, the network performance has been improved to different degrees, and the improvement effects of these modules can be superimposed. The characteristics of these three improved modules in Deeplab v3+ are consistent with the multi-scale recurrent network proposed in this article, indicating that the improved modules proposed in this article have good mobility.

B. SEMI-SUPERVISED SEMANTIC RECOGNITION EXPERIMENT RESULTS

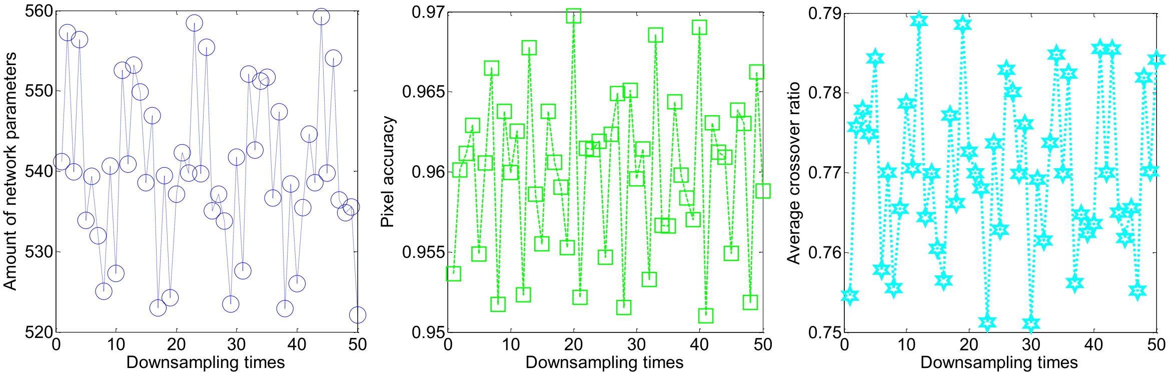
1) TRAINING METHOD

In the training process, this article found that the larger the input image size of the network, the better the training effect of the network. However, expanding the input image size of the network will take up more video memory. Due to the limited video memory of the experiment, this article adopts the strategy of two graphics cards training together in the training process. The batch number is set to 20, and each graphics card is 10.

In the training, a total of 40 rounds of training are carried out. This article adopts the Adam method, in which the momentum factor is set to 0.9. At the same time, the learning rate adjustment strategy of polynomial decay is adopted, that is, the learning rate will decrease with the increase of the number of training iteration steps. The initial learning rate is set to 4e-5, and the minimum learning rate setting is reached when training for 30 rounds is 5e-6, and its power exponent is 0.91.

2) ANALYSIS OF EXPERIMENTAL RESULTS

This article first compares the performance of the network when the amount of data is used, then compares the performance of the network when using different trade-off coefficients in the Bayesian cross entropy function, and then compares the performance of the network when the Bayesian cross entropy function is used with the ordinary cross entropy function. The results of network performance improvement after adopting the semi-supervised learning method proposed in this article are given, and finally the recognition effect of the face recognition network obtained in this article in real scenes is given. In order to verify these experiments, this



(a) The amount of network parameters (b) Pixel accuracy (c) Average intersection ratio

FIGURE 8. Comparison result of improved module and Deeplab v3+ network.

article uses the face recognition data set established in this article to verify the experiments.

The semi-supervised portrait recognition data set compiled in this article contains 4 data sets with different amounts of data: overall training set, 1/2 training set, 1/4 training set, and 1/8 training set. This part of the comparison uses different amounts of data for training, and the specific results are shown in Figure 9.

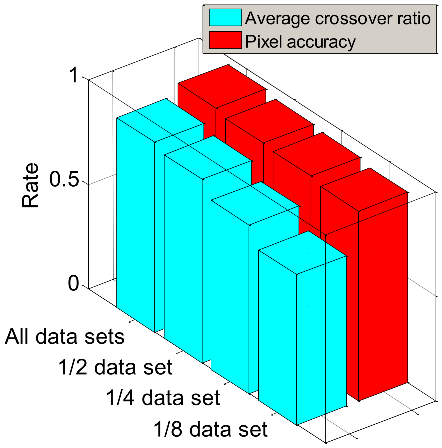


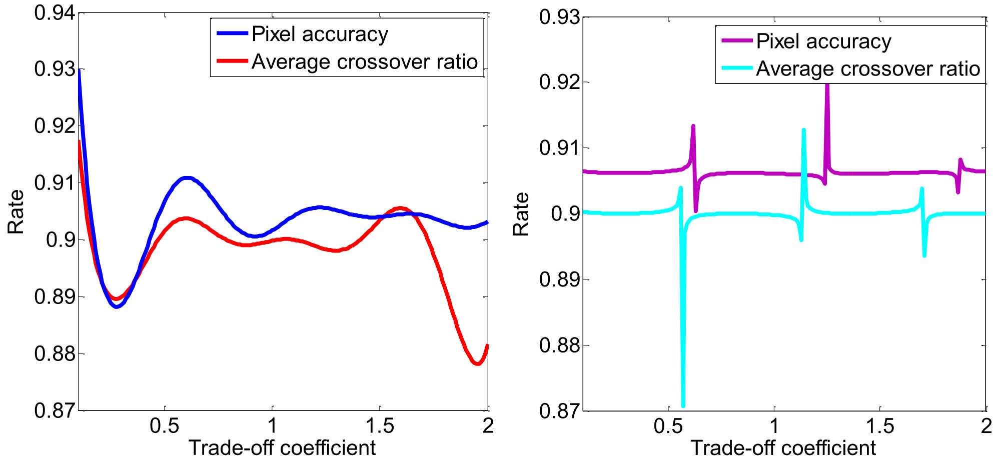
FIGURE 9. Comparison results under different data volumes.

Through comparison, it can be found that when the amount of training data is larger, the performance of the network is better. The network performance when using the full data volume is 9.5% higher than the network performance using 1/8 the data volume. In the case of a small amount of data, the performance of the network can be significantly improved by increasing the amount of data, but when the amount of data is large, the effect of increasing the amount of data is not obvious. The 1/4 data volume has doubled the data volume of 1/8 data volume, and the network performance has increased by 4.6%, while the data volume has also doubled. The total data volume is only increased compared to the 1/2 data volume 2.8%. Through image comparison and analysis, it can be found that, compared with the small amount of data, the recognition accuracy of the network under the large amount of data is mainly improved by reducing the background misjudgment rate and improving the accuracy of the edge of the portrait.

This article selects 1/8 training set as the comparison data set. Since the trade-off coefficient in the Bayesian cross-entropy function can be a continuous value, in order to better express the comparison results and save the experimental cost, this article only selects a few representative trade-off coefficients for experiments, and selects 0.1∼2 as the trade-off coefficient, At the same time, this article has been tested on the test set and the remaining 7/8 training set. The specific results are shown in Figure 10.

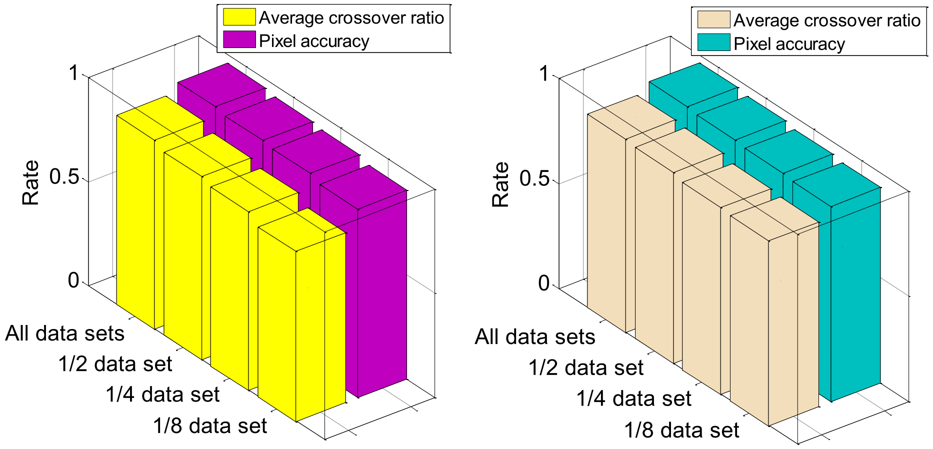
Through comparison, it can be found that when the value of the trade-off coefficient in the Bayesian cross entropy function is different, there is no obvious difference between the performance of the network in the test set and the remaining 7/8 training set, but through image comparison analysis, it can be found that the determination graph is obviously different. The larger the trade-off coefficient, the greater the uncertainty in the uncertainty graph. The uncertainty is mainly concentrated in the edge area of the image or the area that is difficult to identify, which is consistent with the meaning of the variance in the Bayesian cross entropy function in this article. At the same time, it also fully shows that the Bayesian cross entropy function can be improved. It is required to obtain uncertain graphs of different magnitudes. If the Bayesian cross entropy function is not improved, that is, the trade-off coefficient is kept at 1.0, the uncertainty graph obtained is almost 0.

In order to compare the Bayesian cross-entropy function with the ordinary cross-entropy function, this article first sets the trade-off coefficient in the Bayesian cross-entropy function to 0.7, and then sets the overall training set, 1/2 training set, and 1/4 training set. A comparison experiment was carried out on the 1/8 training set, and the specific results are shown in Figure 11.



(a) Test set **(b) Remaining training set**

FIGURE 10. Comparison results of different trade-off coefficients in Bayesian cross entropy function.



(a) Ordinary cross entropy function (b) Bayesian cross entropy function

FIGURE 11. Comparison results under different data volumes.

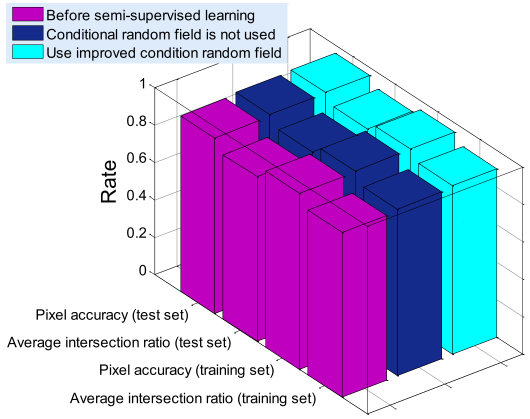


FIGURE 12. Comparison results of semi-supervised learning.

Through the comparison, it can be found that the network performance will be improved when the Bayesian cross entropy function is used compared with the ordinary cross entropy function. When the amount of data is small, the performance of the network is improved after switching to the Bayesian cross entropy function. When the amount of data is large, the performance of the network is improved less after switching to the Bayesian cross entropy function. When using all data sets, the network performance is improved by 1.35% after switching to the Bayesian cross entropy function, and when using the 1/8 data set, the network performance is improved by 6.67% after switching to the Bayesian cross entropy function.

In order to perform semi-supervised learning, this article uses 1/8 of the data set as the labeled data set, and the remaining 7/8 data set as the unlabeled data.

This article compares the three cases of not using the conditional random field, using the ordinary conditional random field, and using the improved conditional random field. Specifically, in the semi-supervised learning process, this article first uses the network trained on the labeled data set to predict in the unlabeled data to obtain the recognition result and uncertainty of the unlabeled data. When the

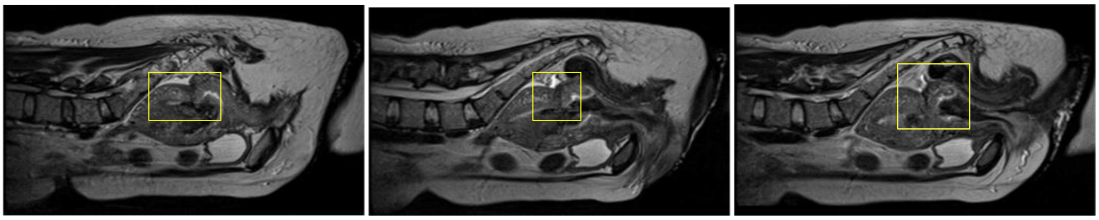


FIGURE 13. Recognition result of intrauterine residue image based on deep learning combined with semi-supervised learning.

conditional random field is not used, the prediction result is directly used as a pseudo-labeled map.

In order to distinguish the deterministic area from the uncertainty area in the uncertainty graph, this article uses a certainty value of 0.8 as the threshold. The specific semi-supervised learning for comparison is shown in Figure 12.

Through comparison, it can be found that after semi-supervised learning, the network performance on the test set has been significantly improved, which is better than the network performance obtained by using 1/4 of the data volume as the training data, even if the conditional random field is not used. After semi-supervised learning, the network performance is also higher than the network performance obtained by using 1/4 data volume as training data. This fully illustrates the effectiveness of the semi-supervised method proposed in this article, and also fully illustrates the effectiveness of the improvement of the conditional random field in this article. Through the performance comparison on the remaining 7/8 training set, it can be found that the network performance has been significantly improved after semi-supervised learning, which fully illustrates the effectiveness of the uncertainty loss function for unlabeled images.

Through comparative analysis on unlabeled images, it can be found that after semi-supervised learning, the original prediction errors will be automatically corrected at the cost of greater uncertainty. After semi-supervised learning, the original prediction errors are not corrected, but the uncertainty often indicates that the predictions of these error regions are not accurate. These images illustrate the role of uncertainty in semi-supervised learning. Figure 13 shows the image recognition results of intrauterine residues by deep learning combined with semi-supervised learning.

V. CONCLUSION

From the perspective of network structure and loss function, this article proposes a method based on multi-scale cyclic convolutional network. Aiming at the problem of not being able to effectively extract category information and location information in an image at the same time, this article proposes a spatial pyramid loop module, which uses a spatial pyramid structure to extract semantic information and category information, and creatively adds circular convolution to this structure. The network to realize the function of information transmission at different scales further strengthens the ability to extract semantic information and category information. In the research of semi-supervised image semantic recognition, this article proposes a method based on uncertainty and conditional random field. This is a method based on pseudo-labeled images. Aiming at the problem that false labeling in pseudo-labeled images will affect training, this article adds uncertainty information for image recognition to provide the accuracy of recognition. In order to obtain the uncertainty information of the recognition image, this article adopts Bayesian cross entropy loss, and on the basis of it, an improvement is made for the image recognition task. In order to obtain pseudo-labeled images with higher accuracy, this article adopts the method based on conditional random field. In view of uncertainty, this article improves the conditional random field to eliminate the influence of uncertainty information and make it more reasonable. The uncertainty loss function is used to complete semi-supervised learning of image semantic recognition on unlabeled data. Combined with the fully-supervised and semi-supervised image semantic recognition method based on deep neural network proposed in this article, this article carries out experimental verification and comparative analysis. The effectiveness of the method proposed in this article is verified by experiments, and the results of other methods in the current related fields prove that the method proposed in this article has outstanding performance.

REFERENCES

[1] Q. Zhu, B. Du, and P. Yan, "Boundary-weighted domain adaptive neural network for prostate MR image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 753–763, Mar. 2020.

[2] M. Mehdipour Ghazi, B. Yanikoglu, and E. Aptoula, "Plant identification using deep neural networks via optimization of transfer learning parameters," *Neurocomputing*, vol. 235, pp. 228–235, Apr. 2017.

[3] J. Bi, H. Yuan, and M. Zhou, "Temporal prediction of multiapplication consolidated workloads in distributed clouds," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1763–1773, Oct. 2019.

[4] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers, Z. Xu, and D. J. Mollura, "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Comput. Methods Biomech. Biomed. Eng., Imag. Vis.*, vol. 6, no. 1, pp. 1–6, Jan. 2018.

[5] A. Karargyris, J. Siegelman, D. Tzortzis, S. Jaeger, S. Candemir, Z. Xue, K. C. Santosh, S. Vajda, S. Antani, L. Folio, and G. R. Thoma, "Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 1, pp. 99–106, Jan. 2016.

[6] A. Nibali, Z. He, and D. Wollersheim, "Pulmonary nodule classification with deep residual networks," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 10, pp. 1799–1808, Oct. 2017.

[7] C. Qin, D. Yao, Y. Shi, and Z. Song, "Computer-aided detection in chest radiography based on artificial intelligence: A survey," *Biomed. Eng. OnLine*, vol. 17, no. 1, pp. 1–23, Dec. 2018.

[8] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1207–1216, May 2016.

[9] E. Elyan and M. M. Gaber, "A fine-grained random forests using class decomposition: An application to medical diagnosis," *Neural Comput. Appl.*, vol. 27, no. 8, pp. 2279–2288, Nov. 2016.

[10] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, and W. Hsu, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Syst. Appl.*, vol. 128, pp. 84–95, Aug. 2019.

[11] Y. Li, C. Huang, L. Ding, Z. Li, Y. Pan, and X. Gao, "Deep learning in bioinformatics: Introduction, application, and perspective in the big data era," *Methods*, vol. 166, pp. 4–21, Aug. 2019.

[12] X. Gao, S. Lin, and T. Y. Wong, "Automatic feature learning to grade nuclear cataracts based on deep learning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2693–2701, Nov. 2015.

[13] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Isgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, May 2016.

[14] R. Ashraf, M. Ahmed, S. Jabbar, S. Khalid, A. Ahmad, S. Din, and G. Jeon, "Content based image retrieval by using color descriptor and discrete wavelet transform," *J. Med. Syst.*, vol. 42, no. 3, p. 44, Mar. 2018.

[15] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognit.*, vol. 85, pp. 109–119, Jan. 2019.

[16] Y. Li, S. Wang, R. Umarov, B. Xie, M. Fan, L. Li, and X. Gao, "DEEPre: Sequence-based enzyme EC number prediction by deep learning," *Bioinformatics*, vol. 34, no. 5, pp. 760–769, Mar. 2018.

[17] A. S. Qureshi, A. Khan, A. Zameer, and A. Usman, "Wind power prediction using deep neural network based meta regression and transfer learning," *Appl. Soft Comput.*, vol. 58, pp. 742–755, Sep. 2017.

[18] E. Elyan and M. M. Gaber, "A genetic algorithm approach to optimising random forests applied to class engineered data," *Inf. Sci.*, vol. 384, pp. 220–234, Apr. 2017.

[19] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017.

[20] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[21] S. Zhou, Q. Chen, and X. Wang, "Active deep learning method for semi-supervised sentiment classification," *Neurocomputing*, vol. 120, pp. 536–546, Nov. 2013.

[22] Z. Zou, S. Tian, X. Gao, and Y. Li, "MlDEEPre: Multi-functional enzyme function prediction with hierarchical multi-label deep learning," *Frontiers Genet.*, vol. 9, p. 714, Jan. 2019.

[23] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[24] G. Wu, W. Lu, G. Gao, C. Zhao, and J. Liu, "Regional deep learning model for visual tracking," *Neurocomputing*, vol. 175, pp. 310–323, Jan. 2016.

[25] R. Ashraf, K. Bashir, A. Irtaza, and M. Mahmood, "Content based image retrieval using embedded neural networks with bandletized regions," *Entropy*, vol. 17, no. 6, pp. 3552–3580, May 2015.

[26] Q. Zhu, B. Du, B. Turkbey, P. Choyke, and P. Yan, "Exploiting interslice correlation for MRI prostate image segmentation, from recursive neural networks aspect," *Complexity*, vol. 2018, pp. 1–10, Feb. 2018.

[27] R. Ashraf, M. Ahmed, U. Ahmad, M. A. Habib, S. Jabbar, and K. Naseer, "MDCBIR-MF: Multimedia data for content-based image retrieval by using multiple features," *Multimedia Tools Appl.*, vol. 79, pp. 1–27, Jun. 2018.

[28] O. Yilmaz, "Machine learning using cellular automata based feature expansion and reservoir computing," *J. Cellular Automata*, vol. 10, nos. 5–6, pp. 435–472, Sep. 2015.

[29] M. Reza Zare, W. Chaw Seng, and A. Mueen, "Automatic classification of medical X-ray images using a bag of visual words," *IET Comput. Vis.*, vol. 7, no. 2, pp. 105–114, Apr. 2013.

[30] Q. Zhu, B. Du, P. Yan, H. Lu, and L. Zhang, "Shape prior constrained PSO model for bladder wall MRI segmentation," *Neurocomputing*, vol. 294, pp. 19–28, Jun. 2018.

[31] S. A. El-Regaily, M. A. Salem, M. H. Abdel Aziz, and M. I. Roushdy, "Survey of computer aided detection systems for lung cancer in computed tomography," *Current Med. Imag. Rev.*, vol. 14, no. 1, pp. 3–18, Dec. 2017.

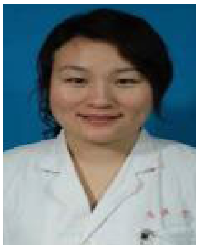
[32] J. Bi, T. Feng, and H. Yuan, "Real-time and short-term anomaly detection for GWAC light curves," *Comput. Ind.*, vol. 97, pp. 76–84, May 2018.

[33] D. Cheng, G. Meng, G. Cheng, and C. Pan, "SeNet: Structured edge network for sea–land segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 247–251, Feb. 2017.

[34] J. Bi, H. Yuan, L. Zhang, and J. Zhang, "SGW-SCN: An integrated machine learning approach for workload forecasting in geo-distributed cloud data centers," *Inf. Sci.*, vol. 481, pp. 57–68, May 2019.

TAO TAO graduated from the Tongji Medical College, Huazhong University of Science and Technology, in 2011. He worked with the Department of Gynecology and Obstetrics, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University. His research interest includes fetal medicine.

KAN LIU graduated from Tianjin Medical University, in 2011. She worked with the Henan Provincial People's Hospital. Her research interest includes maternal-fetal medicine.



LI WANG graduated from Xi'an Jiaotong University, in 2008. She is currently working with the Henan Provincial People's Hospital. Her research interest includes obstetric emergencies.

HAIYING WU graduated from Henan Medical University, majoring in clinical medicine, in 1991. She received the master's degree from Henan Medical University, in 1999, and the Ph.D. degree from Zhengzhou University, in 2012. She is currently working with the Henan Provincial People's Hospital. Her research interests include critical obstetrics and fetal medicine.