# Self-Paced Learning With Diversity for Medical Image Segmentation by Using the Query-by-Committee and Dynamic Clustering Techniques

**XIAOYING PAN**[1,2], **DE WEI**[1], **YIZHE ZHAO**[3], **MINJIE MA**[1], **AND HAO WANG**[4]

[1]School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
[2]Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an 710121, China
[3]School of Information Science and Technology, Northwest University, Xi'an 710119, China
[4]School of Software, Northwestern Polytechnical University, Xi'an 710129, China

Corresponding author: De Wei (weide@stu.xupt.edu.cn)

**ABSTRACT** Convolutional neural networks (CNNs), as a typical deep learning technique, have been widely used in image segmentation, but they often require a large amount of annotated data. However, the number of available pixel-wise labeled medical images is extremely small, and this prevents the application of CNNs in many medical image segmentation tasks. We proposed a self-paced learning with diversity SPLD) framework to boost the performances of medical image segmentation models with a limited amount of annotated data. Self-paced learning (SPL) is a learning regime that selects training samples in order from the easiest to the most difficult for model training. In addition, we took the diversity of the data into consideration. The proposed self-paced learning with diversity by query-by-committee (SPLD-QBC) algorithm dynamically and diversely selects the appropriate training data to boost the performance of an image segmentation model. SPLD-QBC incorporates the query-by-committee (QBC) technique for data selection and affinity propagation for optimizing the data diversity. By dynamically selecting the optimal sequence of training samples from different probability distributions, the segmentation models achieved improved performances. To verify the effectiveness of the proposed SPLD-QBC framework, we conducted experiments on three medical image segmentation tasks with five different datasets. The experimental results indicated that the proposed SPLD-QBC algorithm significantly improved upon the segmentation performances of the baseline models and resulted in a higher Dice score, surface distance and mean intersection over union (mIoU). The proposed SPLD significantly boosts the segmentation performances of models and is easily embedded into CNN-based image segmentation models.

**INDEX TERMS** Self-paced learning, image segmentation, deep learning.

## I. INTRODUCTION

The precision of medical image segmentation is crucial for obtaining accurate diagnoses. In recent years, deep learning-based techniques have become the most popular approaches in the various subfields of medical image semantic segmentation, such as retinal vessel segmentation [1], organ segmentation [2], and cell segmentation [3]. To the best of our

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda .

knowledge, training convolutional neural networks (CNNs) requires a large amount of labeled training data. Furthermore, deep learning often requires strenuous annotation by clinical experts since a large amount of image data are required for training and validating deep neural networks. Under such circumstances, it is necessary to fully use the information behind the medical data to generate reliable results for clinical usage.

Most of the optimization functions used in deep learning are nonconvex functions [4], which increase the difficulty
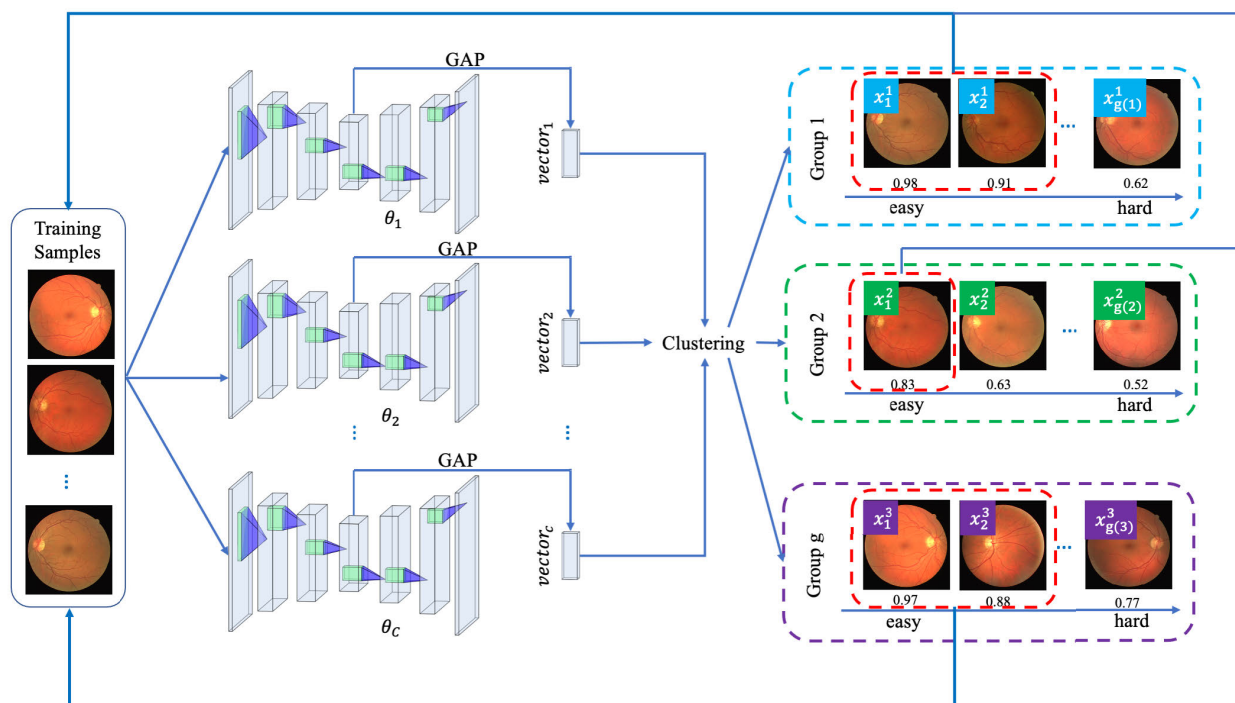
**FIGURE 1.** Overview of proposed SPLD approach for medical image segmentation with an application on retina vessel segmentation.

of finding global minima. Self-paced learning (SPL) [5] is a popular training strategy for enhancing model performance by first feeding easy samples to the model to further exploit the information behind the training data. SPL imitates the learning process of human beings, which also starts with easy samples and then gradually uses some complex data for model training. In SPL, an additional regularization term is added to the optimization function so that the loss and curriculum are trained jointly. However, the ease of a training sample is hard to determine, and what is intuitively "easy" for a human may not be as easy for the computer to comprehend.

According to [6], the probability distribution behind a dataset is difficult to capture. In addition, the distribution of the training data is independently and identically distributed. When directly applying the SPL learning framework to image segmentation tasks, the SPL model selects images that reside in the same cluster of the probability distribution so that the weights of the model are optimized according to a local minimum [7]. Hence, employing data from diverse distributions is important.

Self-paced learning with diversity (SPLD) was proposed in [7] by incorporating data diversity into the SPL learning strategy. The original SPLD algorithm demonstrated that the intuitive approach for SPLD is to select samples from different groups, i.e., data from different classes. However, in semantic image segmentation tasks, categorical labels are often missing, thereby limiting the application of SPL with diversity.

In this paper, a versatile self-paced learning with diversity by query-by-committee (SPLD-QBC) framework for medical image segmentation is presented. The proposed SPLD-QBC segmentation framework integrates a regularization term into the SPL loss function. Instead of using image-level accuracy, i.e., the total Dice score and pixel-level metrics, we incorporate the query-by-committee (QBC) [9] technique to determine the simplicity of the data in SPL regimes. A member in the committee is a deep CNN-based segmentation network, and each member is trained with the same dataset. Under the SPLD-QBC learning framework, the data simplicity and diversity are determined by the latent features of each member. The next query is chosen according to the principle of maximal similarity, which is calculated by the cosine distances between the extracted features among all members. An overview of the proposed SPLD-QBC framework is shown in Figure 1.

As depicted in Figure 1, the SPLD-QBC framework contains several members in the committee. Each member is a deep CNN for medical image segmentation. The feature maps with the smallest sizes, located in the deepest layer of the image encoder, are extracted to represent the feature of an image patch. All of the members work cooperatively to determine the learning pace: easy samples are utilized to train the model first, and then hard samples are iteratively added to the training set for model fine-tuning. Furthermore, we use the extracted feature vectors to group the training data and thus guarantee the diversity of the data.

The main contribution of this paper is that we propose a versatile medical image semantic segmentation framework that combines SPLD and the QBC technique. To the best of our knowledge, this paper is the first work that uses self-paced learning to perform medical image semantic segmentation. The proposed SPLD-QBC framework is a supervised learning regime that aims at improving the performance of an image segmentation model. Instead of determining the difficulty level of the training samples, the features extracted from the deep neural network are used as surrogates to measure the simplicity of the data and thus determine the training sequence. More impressively, quantitative and theoretical proofs of the effectiveness of the proposed SPLD-QBC framework are illustrated.

## II. RELATED WORK

### A. SELF-PACED LEARNING

Curriculum learning (CL) indicates that learning first from the easiest aspects of a task and gradually increasing the difficulty level can benefit model training. CL helps the model find improved local minima and speeds up the convergence of the training process towards the global minimum. SPL and CL share a similar learning regime concept but differ in their designs of the learning pace. In comparison to human education, CL is guided by instructors, while SPL is advised by students [8]. In CL, the syllabus or the data learning sequence is determined by prior knowledge; however, the learning sequence is dynamically updated during successive training iterations in SPL. An application using a neural network to determine an appropriate syllabus was presented in [9], and the learning progress was determined using reinforcement learning rewards to maximize learning efficiency.

In detail, SPL embeds curriculum learning and assigns a regularization term into the learning objective function [8]. During the training periods, the loss function and the SPL regularization term are optimized jointly. The goal of a typical SPL model is to jointly learn the model parameter w and the data weight variable $v = [v_1, v_2, \ldots, v_n]^T$ by minimizing the objective function in Eq. 1:

$$E(w, v; \lambda)$$
$$= \sum_{i=1}^{n} v_i L(y_i, f(x_i, w)) - \lambda \sum_{i=1}^{n} v_i, \quad s.t. \ v \in [0, 1]^n \quad (1)$$

where $\lambda$ is the hyperparameter for controlling the learning pace, $f$ is the machine learning model, $L$ is the loss function, and $x_i$ and $y_i$ are the training data and corresponding label, respectively.

From an extendibility perspective, SPL has been extended into self-paced curriculum learning [8], self-paced learning with reinforcement learning [10], and self-paced boosted learning [11]. In the extensions of SPL, several regularization terms are added to Eq. 1 for the optimization of specific tasks. From an applicability perspective, SPL has been applied to classification [11], matrix factorization [12], and mixtures of regressions [13]. Zhang *et al.* [14] proposed a self-paced

fine-tuning network for localizing and segmenting objects in weakly-labeled videos. A joint SPL regularizer was employed to compute the data priority values according to the localization task and segmentation task. Moreover, SPL is utilized in multi-instance learning for object detection [15]. In addition, CL and SPL are jointly utilized to perform weak object detection tasks, where the prior knowledge of the learning sequence is integrated by the corresponding regularization term [16]. For image segmentation tasks, Tong *et al.* [17] proposed a self-paced DenseNet that dynamically adjusts the weights of each target class to train the easiest classes first and then train the hard object classes.

However, most SPL applications are implemented on sparsely labeled classification data and for object detection tasks, where the label often exists as a scalar or a bounding box. However, studies related to dense label classification, such as image segmentation using SPL, are limited.

### B. SELF-PACED LEARNING WITH DIVERSITY

SPLD was first proposed by Jiang *et al.* [7], who embedded a regularization term independent of the specific model objectives to determine the diversity of the training data. SPLD integrates both diversity and the ease of generating a curriculum to sequentially feed reasonable data into the model for training and aims to guarantee improved model performance. The modification of SPLD involves adding a regularization term into Eq. 1 to form the diversity constraint. The proposed objective function is shown in Eq. 2.

$$\min_{w, v \in [0,1]^n} E(w, v; \lambda) = \sum_{i=1}^{n} v_i L(y_i, f(x_i, w))$$
$$- \lambda \sum_{i=1}^{n} v_i - \gamma \|v\|_{2,1} \quad (2)$$

where $\gamma$ is the newly added hyperparameter for the diversification pace. SPLD was first performed on an action recognition task and a multimedia event detection task with a random forest [18] and support vector machine (SVM). The datasets used in both tasks were sparsely labeled datasets in which the class of each data point was known by the SPLD model. However, in medical image semantic segmentation tasks, the image-level labels, i.e., the categorical labels, are unknown. Therefore, controlling data diversity by using previously obtained class labels is impossible.

### C. QUERY-BY-COMMITTEE

QBC is an active learning algorithm. The key technique of QBC is that it actively decides which data are critical for the model training process. In QBC, a set of active learners is trained, and the learners vote on the data to decide which data points need to be added to the training set. To some extent, the active learning of QBC can effectively evaluate the informativeness of the training data. In QBC, the cross-entropy method is often used to calculate the data similarity [19]. However, the entropy is always based on sparse label prediction, so it is not suitable for dense label classification tasks.

## D. MEDICAL IMAGE SEGMENTATION

Deep neural networks have been widely used in medical image segmentation tasks. For a long time, various algorithms in the research field of computer vision have been proposed and developed for the automatic segmentation of retinal vessels in fundus images [20]. Jung *et al.* proposed an iterative deep learning model for medical image segmentation by iteratively inputting the network output as the shape prior for model fine-tuning [21]. A curvilinear structure extraction algorithm was presented in [22], and it trained the model from a single patch to the entire image sequentially. Different from natural images with RGB channels, medical image data are always grayscale and contain multimodal data. A cross-modality feature representation method was proposed in [23]; the features from multimodal MR images were extracted, and a precise brain tumor segmentation model was trained. In medical image segmentation, incorporating a shape prior and clinical knowledge is important. Zhang *et al.* [24] proposed a task-structured brain tumor segmentation network that models the task-modality structure as a weighted combination structure and mimics clinical practices to extract brain tumors. For cell segmentation, splitting clumps of convex objects has many practical applications in biomedical and industrial fields. The splitting framework proposed in [25] successfully solved the overlapping and touching problem in cell segmentation and improved the robustness of the cell-splitting model.

## III. APPROACH

In this paper, we propose a new framework for medical image semantic segmentation using the SPLD strategy, which dynamically selects the samples in order from the easiest to the hardest ones while guaranteeing the diversity of the data to facilitate model convergence and to improve the performance of the image segmentation model.

## A. QBC BASED SPL REGULARIZATION TERM

Before illustrating the algorithm, we firstly give the definition of the variables. Let $L = Labeled\ Data = \{\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_n, y_n\}\}$ be a collection of labeled image patches cropped from medical images with ground truth. Let $C = \{\theta^1, \theta^2, \ldots, \theta^C\}$ be a committee with |C| members, each member $\theta^i$ is a deep CNN for medical image semantic segmentation.

Our approach utilizes the QBC framework. In each iteration, each committee predicts the features of all data points. A CNN-based image segmentation network contains an encoder and a decoder. The convolutional layer in the deep network provides the same receptive field as those of the large filters in the initial layers. When the network goes deeper, most of the existing CNNs employ the pooling layer to increase the receptive field. By deepening the network with small filters and pooling layers, the performance is enhanced [26]. In the encoder part of most image segmentation networks, the number of feature maps increases with

decreasing resolution. The shallowest layers of the CNN learn low-level image features, such as shapes and edges, while the highest layers learn high-level features, which are most important to the specific application [27]. In our approach, we extract the feature maps from the last layer of the encoder and use the global average pooling (GAP) layer to convert the feature maps into feature vectors. Each member in the committee generates a feature vector representing the features of the input image. The feature vector generated from $\theta^i$ is defined as vector$_i$. The similarity between two members is measured by the cosine distance between them, as illustrated in Eq. 3:

$$\text{sim}_{ij} = \frac{\text{vector}_i \cdot \text{vector}_j}{|\text{vector}_i| \times |\text{vector}_j|} \quad (3)$$

where $|\text{vector}_i|$ represents the length of the high-level feature vector extracted by member $\theta^i$ in committee C, '$\cdot$' represents the dot product operation and '$\times$' indicates the cross-product operation. All the members in the committee have the same CNN architecture, indicating that the meaning of the elements in the feature vector is the same. According to Eq. 3, the cosine function is used to calculate the global similarity using the two extracted feature vectors. Each pair of elements from the two vectors contributes to the cosine similarity. The similarity between each member can be formulated as a matrix:

$$\text{sim}_M(x_k)$$
$$= \begin{bmatrix} 0 & \text{sim}_{12}(x_k) & \cdots & \text{sim}_{1C}(x_k) \\ \text{sim}_{21}(x_k) & 0 & & \text{sim}_{2C}(x_k) \\ & \vdots & \ddots & \vdots \\ \text{sim}_{C1}(x_k) & \text{sim}_{C2}(x_k) & \cdots & 0 \end{bmatrix} \quad (4)$$

In our approach, we define the similarity of an input data $x_k$ as:

$$\text{sim}(x_k) = \sum_{i>j>0}^{n} \text{sim}_{ij}(x_k) \quad (5)$$

The hypothesis behind our framework is that a set of extracted SPLD-QBC feature vectors with a high similarity indicates that all of the members in the committee have effectively learned the input datum, $x_k$. Thus, a higher similarity indicates that the data are an easier sample and are thus first selected in the SPL training process. In contrast, a sample with a small similarity means that the features generated from each member possess large discrepancies between them, indicating a difficult sample.

After obtaining the similarities of all the training samples, we normalize them into [0, 1] to form the SPL regularization term in Eq. 1 by Eq. 6.

$$v_k = \frac{sim(x_k) - \min_{i \in [1,n]} sim(x_i)}{\max_{i \in [1,n]} sim(x_i) - \min_{i \in [1,n]} sim(x_i)} \quad (6)$$

To successfully apply the SPL strategy during the training process, we need to tackle the following three problems:

(1) How many data should be selected during the initial training epochs? The SPL strategy has to first exploit the information provided with a small number of data and then add more labeled data to the training set to control the learning pace. However, during the first several training epochs, the parameters in the CNNs have not yet been well optimized. If the data selection rule is directly applied to the SPL strategy, the model may collapse, and the SPL framework will not work. To handle this problem, we randomly select a small number of training samples to train the model. For example, we can train the committee with 10% randomly selected samples. The initial ratio of training data is defined as a hyperparameter named *ITDR* in our algorithm. If the model converges, we use Eq. 6 to control the pace of learning:

(2) Another problem for QBC-based learning is that all the members have the same CNN architecture, indicating that they would generate the same sample feature vector for an input image if their parameters are the same. To overcome this shortcoming, we randomly initialize the model variables in each CNN with different seeds. With different seeds, the parameters and layers in each member possess different meanings, and they could effectively determine the data similarity. In addition, during the training process, the parameters are optimized according to different solutions. The experimental evaluation section shows that our method is effective.

(3) How can we control the learning pace? In practice, we need to prevent SPL learners from greedily adding data. If the SPL learner adds a fixed number of training data after each training epoch, all the data will be added to the training set, and the proposed method will become a fully supervised one. This is not our original intention with regard to controlling the learning pace. To tackle this problem, we impose a restriction on the query rule by setting an uncertainty threshold, which is inspired by the uncertainty sampling strategy [28]. The threshold is named *UT* in our algorithm. If and only if the $v_k$ of the data evaluated by all the members in the committee is larger than the threshold, the SPL learner defines the sample as an ''easy'' sample and adds the data to the training set.

In addition, the *UT* is dynamically changed during the training process. At the first training epoch, *UT* equal to 0.9 times of the largest $v_k$ in $v$. With the increment of training iterations, UT dynamically changes according to the Eq. 7.

$$UT = \frac{0.9e}{m}(\max_{i\in[1,n]} v_i) \qquad (7)$$

where $m$ is the number of training epochs and $e$ is the current epoch during the training process.

## B. CLUSTERING BASED DATA DIVERSITY
Intuitively, selecting data from different classes is an easy method to maintain data diversity during model training. However, as demonstrated in section 1, the medical images used in segmentation often lack categorical labels. To address this circumstance, a cluster algorithm is employed to deter-

mine the groupings of the data. In section 3.A, the extracted features from each member determine the ease of the training data because the features contain the representations of the data. In the grouping stage, the extracted features are reutilized for clustering.

In a committee, for each training data, the model will generate |C| feature vectors. We use the average of the |C| features as the feature vector of the sample, as shown in Eq. 8.

$$u(x_k) = \frac{1}{|C|}\sum_{i=1}^{|C|} vector_i \qquad (8)$$

where $vector_i$ is pre-defined in Eq. 3.

However, even if we can use a clustering algorithm to classify all the data into several groups, the optimal number of groups is still unknown. If the number of groups is too large, all of the data will be separately grouped into cluster centers with small numbers of samples; in contrast, if the number of groups is too small, the diversity of the data is not effectively represented. To handle this problem, affinity propagation [29] is employed to cluster the generated feature vectors. In affinity propagation, three matrixes are calculated: (1) a similarity matrix that indicates the data similarity between the training samples; (2) a responsibility matrix that measures how well the sample serves as the exemplar relative to other candidates; and (3) an availability matrix that represents the appropriateness of a sample picked from the exemplar. During the training iterations, the three matrixes are updated until the cluster boundaries remain unchanged. The exemplars are extracted as the cluster centers for classification. The affinity propagation algorithm is not needed to determine or estimate the number of clusters before clustering because the number of groups in the diversity process is automatically calculated according to the extracted feature vectors.

During each training epoch, the affinity propagation algorithm is applied to extract features and cluster data into different groups. Then, the SPL approach demonstrated in section 3.A is employed to select several easy samples from each clustered group. Formally, by applying the cluster algorithm, the dataset is grouped into $g$ groups as $L_g = \{\{X^1, Y^1\}, \{X^2, Y^2\}, \ldots, \{X^g, Y^g\}\}$, where $X^i = \{x_1^i, x_2^i, \ldots, x_{g(i)}^i\}$, $Y^i = \{y_1^i, y_2^i, \ldots, y_{g(i)}^i\}$ s.t. $i \in [1, g]$, and $g(i)$ represents the number of samples in the corresponding group. Obviously, $\sum_{j=1}^{g} g(j) = n$. To guarantee the diversity of the training sample, the SPL learner selects the easiest samples from each group, and the corresponding data weight variables become $v = [v^1, v^2, \ldots, v^g]$, where $v^i = \left[v_1^i, v_2^i, \ldots, v_{g(i)}^i\right] s.t. i\epsilon[1, g]$.

By applying the grouping strategy, the loss function for SPLD changes from Eq. 2 into Eq. 9, as shown below.

$$E(w, v, \lambda) = \sum_{j=1}^{g}\sum_{i=1}^{g(j)} v_i^j L\left(y_i^j, f\left(x_i^j, w\right)\right)$$
$$- \lambda \sum_{j=1}^{g}\sum_{i=1}^{g(j)} v_i^j - \gamma \|v\|_{2,1} \qquad (9)$$

## C. ALOGIRITHM

In summary, the proposed SPLD-QBC algorithm for medical image segmentation contains two parts: one is SPL data selection, and another is data diversity calculation by using affinity propagation. Combining both parts, the algorithm can be summarized in Algorithm 1.

---

**Algorithm 1** Proposed SPLD-QBC Algorithm

**Input**:

$L = Labeled\ Data = \{\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_n, y_n\}\}$: training data, composed of n samples;

$C = \{\theta^1, \theta^2, \ldots, \theta^C\}$: $|C|$ members initialized by different seeds;

$m$: number of training epochs;

$ITDR$: the ratio of the number of data randomly chosen for training to the number at initial epochs;

**Output:**

$C = \{\theta^1, \theta^2, \ldots, \theta^C\}$: trained committee of medical image semantic segmentation networks;

1. randomly select $ITDR$ of the training data.

**While** model not converges:

   2. Train $C = \{\theta^1, \theta^2, \ldots, \theta^C\}$ with Loss function defined in Eq. 2 on randomly selected data;

**For** $e$ in 1 to $m$:

   3. Extract features $u$ defined in Eq. 8 for all the data in $L$;

   4. Group data with the extracted features by affine propagation into $g$ groups;

   **For** j in 1 to $g$:

     5. Calculate the $v^j$ for each data in the j-th group by Eq. 6;

     6. Select the easy samples in j-th group whose $v > UT$;

   7. Train $C = \{\theta^1, \theta^2, \ldots, \theta^C\}$ with Loss function defined in Eq. 9 on selected data from each group;

   8. Update $UT$ by Eq. 7

---

According to the description of Algorithm 1, there are four inputs, including the label dataset, the initialized committee with $|C|$ members, and two hyperparameters. The outputs of our algorithm are the trained members. Line 1 represents the random data selection process. The while loop and line 2 indicate the first initial training epochs. The first 'for' loop demonstrates the process of the SPLD algorithm. Lines 4 to 6 represent the cluster-based data diversity algorithm in section 3.B. Without the group operation listed in lines 4 to 6, the proposed SPLD algorithm would degrade into the SPL algorithm. In our experiments, the SPL strategy is used as a baseline.

We summarize our algorithm as below:

(1) During the initial training epochs, the proposed algorithm is an SPL-based algorithm.

(2) The presented algorithm uses $\lambda$, $\gamma$ and a dynamically changed $UT$ to control the learning pace and to prevent the algorithm from degrading into a fully supervised learning (FSL) algorithm.

(3) The segmentation results are generated by the member with the best performance. Like students in a class, the best student can produce the best results for a given task.

## D. PROOF OF CONVERGENCE

To guarantee the feasibility of the proposed SPLD-QBC algorithm, it is necessary to prove the convergence of the SPLD approach. The proposed algorithm aims at finding a global optimum to minimize $E(w, v; \lambda)$, as defined in Eq. 9 for any given $w$ and $v$.

*Optimize w with fixed v.* Firstly, if we fix v, then the second and the third terms in Eq. 9 are constants and our algorithm degrades into a SPL algorithm. It is proved that the SPL will converge after several training epochs [34]. And the author mentioned that the learning process of traditional SPL regime can be guaranteed to converge to rational critical points.

*Optimize v with fixed w.* Then, we need to prove v computed in algorithm 1 attains the global optimization with a fixed w. The loss function defined in Eq. 9 can be rewritten as Eq. 10.

$$E(w, v, \lambda)$$

$$= \sum_{j=1}^{g} \sum_{i=1}^{g(j)} v_i^j L\left(y_i^j, f\left(x_i^j, w\right)\right) - \lambda \sum_{j=1}^{g} \sum_{i=1}^{g(j)} v_i^j - \gamma \|v\|_{2,1}$$

$$= \sum_{j=1}^{g} \left( \sum_{i=1}^{g(j)} \left( v_i^j L\left(y_i^j, f\left(x_i^j, w\right)\right) - \lambda v_i^j \right) - \frac{\gamma g(j)}{n} \|v\|_{2,1} \right)$$

$$(10)$$

Eq. 10 indicates that the loss function in the proposed SPLD-QBC algorithm is decomposed into g groups of traditional SPLD loss functions, as defined in Eq. 2. Our proposed SPLD-QBC algorithm is viewed as a linear combination of several traditional SPLD algorithms. Since the traditional SPLD algorithm can converge [7], the proposed SPLD-QBC converges after several training iterations.

By the two proofs, we can conclude that our algorithm could converge. Through minimizing the loss function in Eq. 9, the optimal parameters are obtained.

## IV. RESULTS AND DISCUSSION

In this section, we introduce medical image segmentation experiments. Three types of medical image semantic segmentation tasks, including retina vessel segmentation, lung organ semantic segmentation and nuclear image segmentation, were performed on five different datasets to demonstrate the effectiveness of the proposed SPLD-QBC learning framework.

### A. IMPLEMENTATION DETAILS

Our implementation was based on the TensorFlow deep learning library. To apply the SPL framework, we used image patches cropped from the original medical images, each with a resolution of $128 \times 128$, for model training. During the testing stage, the CNN model of each member inferred the

whole segmentation results rather than the patched images. Due to hardware limitations, the number of members was set to 2, 3 and 4 for the SPLD-QBC environment. Inspired by [31], we used a poly learning strategy, and the learning rate equaled the base learning rate multiplied by $(1 - \frac{e}{m})^{0.9}$, where e indicates the current training epoch and m represents the total number of training iterations. Typically, we set $m = 1000$ for retinal vessel segmentation, $m = 500$ for lung organ segmentation and $m = 300$ for nuclear cell segmentation. The ITDR was set to 10 %. In addition, the hyperparameters $\gamma$ and $\lambda$ were both set as 1 in our experiments.

Furthermore, the architectural design of each member in the committee was different so that the proposed model could be capable of fitting the specific tasks. In the following sections, we discuss the features of these tasks and then illustrate the network for each member.

For each experiment, we compared the proposed SPLD-QBC algorithm with the SPL strategy and fully supervised learning (FSL) strategy. For the SPL baseline, the diversity module implemented by affinity propagation was removed, and the number of committee members was set as 4 ($|C| = 4$) and was fixed. For the FSL, all training patches were selected from the model during the entire training period. In the SPL and FSL experiments, we trained the model with the same number of epochs as that of SPLD-QBC. For SPL, the hyperparameter ITDR was the same as the setting in SPLD-QBC.

## B. RETINA VESSEL SEGMENTATION

Retinal vessel segmentation plays an important role in the automatic detection of retinal diseases with funduscopic images. Retinal blood vessel image analyses provide important information for the detection of several diseases, such as diabetes, retinopathy and glaucoma. The segmentation and localization of retinal blood vessels serve as important cues for the diagnosis of ophthalmological diseases, such as hypertension, microaneurysms and arteriosclerosis [32].

The DRIVE dataset is a public dataset consisting of 40 fundus images of size $565 \times 584$ [33]. The images were manually divided into a training set and a test set, both containing 20 images. The STARE dataset is another public dataset consisting of 20 fundus images of size $605 \times 700$ [34]. The images were manually divided into a training set and a test set, both containing 10 images.

Because the vessels in fundus images vary in shape, thickness and contrast, the model for retina vessel segmentation was designed accordingly. For the vessel segmentation tasks, we employed a full-resolution residual network (FRRN) [35] as the architecture for each member. The FRRN incorporates inputs of different sizes to successfully capture the segment boundaries and enhance the multiscale ability of the model. The last full-resolution residual unit defined in the FRRN was extracted as the latent feature of each member, and we used the feature to compute $v_k$ for SPLD-QBC training.

The retina vessel segmentation task is a binary classification task because each pixel belongs to the vessel or the

**TABLE 1.** Evaluation results on STARE and DRIVE datasets.

| DATASET | DRIVE | | STARE | |
|---|---|---|---|---|
| STRATEGY | DSC | HD | DSC | HD |
| FSL | 0.7700 | 43.1731 | 0.7966 | 60.5468 |
| SPL | 0.7988 | 36.6205 | 0.8118 | 57.5684 |
| SPLD-QBC, $|C|$=2 | 0.8097 | 39.1853 | 0.8446 | 58.7824 |
| SPLD-QBC, $|C|$=3 | 0.8155 | 34.6138 | 0.8359 | 50.6632 |
| SPLD-QBC, $|C|$=4 | **0.8263** | **31.4576** | **0.8520** | **44.9290** |

background. To evaluate the model performances, we calculated the Dice score (DSC) and Hausodorff distance (HD) between the segmentation results and the ground truths. The DSC considers the overlap between the segmented results and ground truths. A higher DSC indicates that the model performance is more powerful. The HD measures how far two subsets of a metric space are from each other. In addition, the lower the HD is, the better the model. The comparison results are as shown in Table 1.

Several examples are shown in Figure 2. We juxtaposed the original vessel image, the segmentation results obtained by the FSL, SPL, and SPLD approaches and the corresponding ground truths. It is observed that the proposed SPLD-SPLD-QBC algorithm extracted more capillaries than the SPL and FSL algorithms, thereby demonstrating that the model trained by the SPLD-QBC strategy has the strongest ability to capture both small and large objects because of diversity learning. The results also indicate that by using the SPLD training algorithm developed in Algorithm 1, the feature representation ability of the model was improved.

## C. ORGAN SEMANTIC SEGMENTATION IN CHEST X-RAYS

Scanning patient organs using chest X-rays is one of the most important procedures, and it creates significant diagnostic workloads [2]. Accurate segmentation of the lung and heart boundaries provides valuable information for developing a computer-aided diagnosis (CAD) system for inspecting lung and heart function. However, due to the high variations in the shapes, sizes and contrasts of chest X-rays, organ segmentation in the lungs remains a challenging task.

Different from retina vessel segmentation tasks, the lung organ segmentation task is semantic because the model predicts each pixel as belonging to the left lung, right lung, heart or background. In this application, the experiments were performed on the JSRT dataset [36] and Montgomery dataset [37]. The JSRT dataset contains 247 chest X-rays, and the annotations include left lung, right lung and heart. The number of total pixel classes is 4. The Montgomery dataset contains 138 chest X-rays, and the total number of classes is three, including left lung, right lung and background. We scaled all images to 256*256 pixels while guaranteeing that the details of organ structures remained clear. In the experiments, 80 % of the samples were randomly selected for training, and the remaining 20 % were selected for testing.

The organ segmentation task is a semantic segmentation task because the number of pixel categories is more than
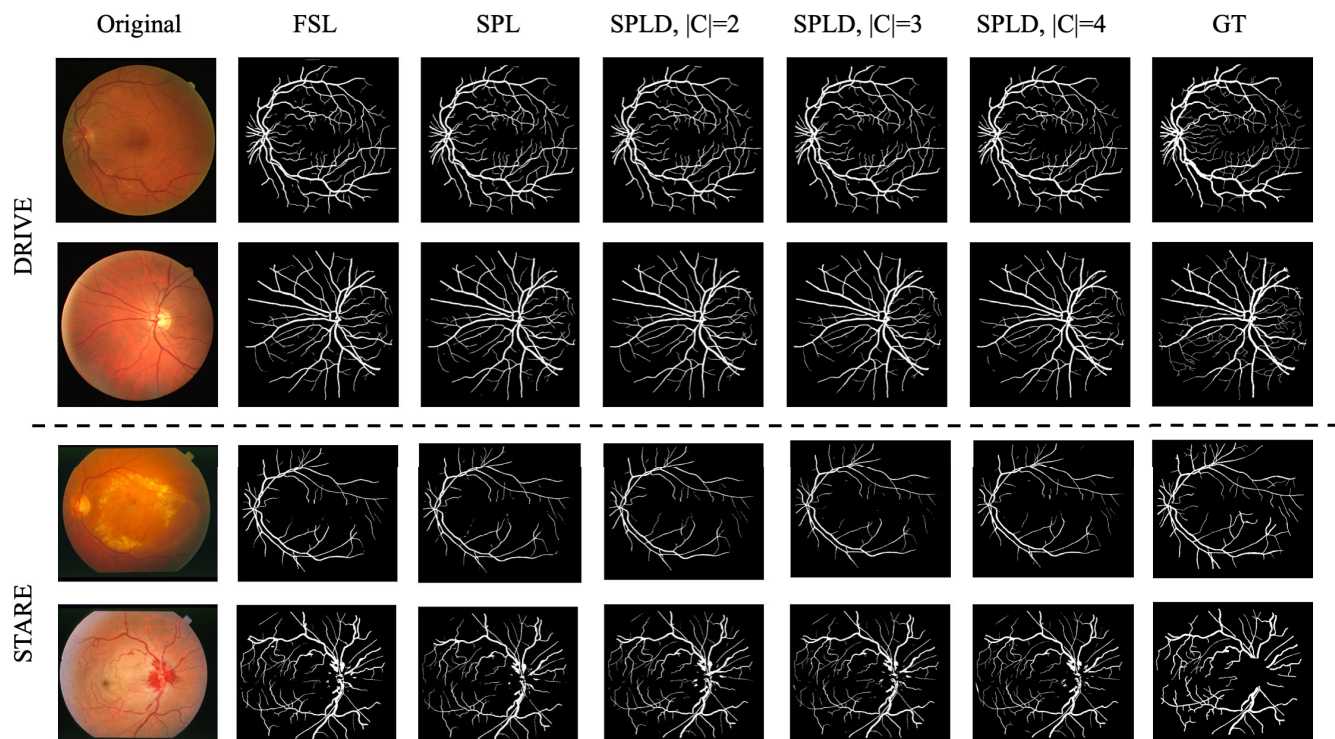
**FIGURE 2.** Segmentation results on DRIVE (Top) and STARE (Bottom) dataset.

**TABLE 2.** Evaluation results on Montgomery and JSRT datasets.

| DATASET | MONTGOMERY | | JSRT | |
|---|---|---|---|---|
| STRATEGY | MIOU | HD | MIOU | HD |
| FSL | 0.9256 | 40.5771 | 0.9254 | 44.2064 |
| SPL | 0.9421 | 30.6431 | 0.9399 | 42.8732 |
| SPLD-QBC, \|C\|=2 | 0.9583 | 23.7990 | 0.9299 | 47.9757 |
| SPLD-QBC, \|C\|=3 | **0.9599** | 31.1726 | 0.9342 | 41.3829 |
| SPLD-QBC, \|C\|=4 | 0.9584 | **22.7487** | **0.9403** | **37.3257** |

2. For this task, we employed AdapNet, which combines the convolutional mixture of deep experts (CMoDE) fusion scheme for learning robust kernels and enlarges the receptive field as the model of each member [38]. We extracted the feature maps of the layer directly before the first deconvolution layer as the latent features of each member. The mean intersection over union (mIoU) and HD between the segmentation results and the ground truths were used to measure the model performances. A higher mIoU indicates that the model performance is more powerful. The comparison results are as shown in Table 2.

As shown in Table 2, the SPLD training strategy outperformed the SPL and FSL methods. A further comparison showed that even though the mIoU was enhanced by limited amount, the HD measures were significantly improved. The proposed method achieved the best performance in terms of the HD metric, at approximately half the value achieved by the FSL strategy on the Montgomery dataset (from 40.5771 to 22.7487). For the JSRT dataset, the HD was significantly

reduced by a 5-pixel length using the proposed SPLD-QBC compared to the FSL baseline. Segmentation examples are shown in Figure 3. For the Montgomery dataset, as the number of members in the committee increased, the amount of noise in the segmentation maps was reduced significantly. For the JSRT dataset, our proposed SPLD model generated impressive results on heart RoIs, even though they are difficult for humans to determine. In addition, with the use of SPLD, the shapes of the organs in segmented images were more complete than those obtained with the use of FSL for both datasets.

## D. NUCLEAR IMAGE SEGMENTATION

Nuclei segmentation is one of the most important tasks for whole slide image analysis in digital pathology [39]. One of the most important steps in clinical practices is to extract information components from the whole slide images (WSIs) [40]. However, thousands of cells exist in a single WSI, and thus manual segmentation is tedious and time-consuming. To automatically segment the cells, we applied our model to H&E-stained multiorgan nuclei segmentation datasets [40] with 30 WSIs and 22,000 corresponding cell boundaries. Each pixel in the WSIs was classified into two categories: one is the cell, and the other is the background. We set each member as a fully convolutional DenseNet (FC-DenseNet) [41]. In this task, we randomly selected 20 samples for training and the remaining 10 for testing. Each image was resized to a resolution of $1024 \times 1024$; thus, we cropped
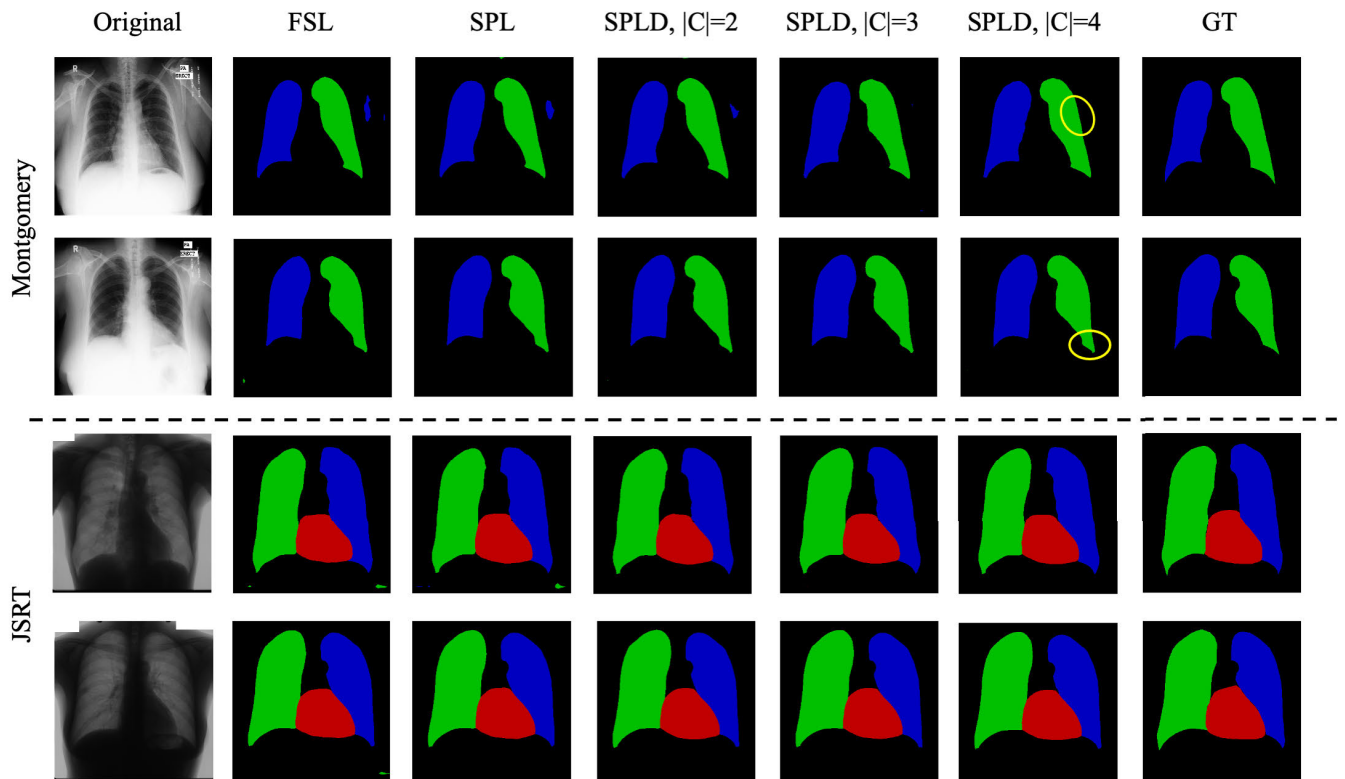
**FIGURE 3.** Segmentation results on Montgomery (Top) and JSRT (Bottom) datasets.
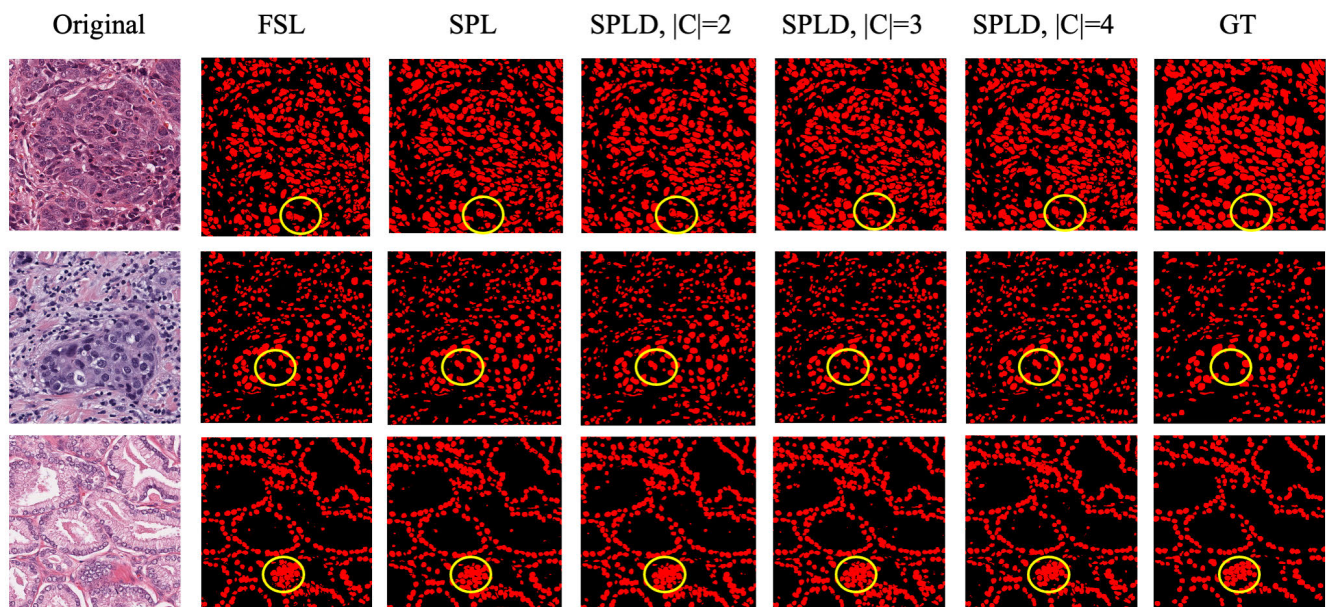


**FIGURE 4.** Segmentation results on WSI dataset.

64 patch images with a resolution of 128 × 128 from one WSI.

Similar to the retina vessel segmentation task, we used the DSC and HD to evaluate the model performances, as shown in Table 3.

As shown in Table 3, our model achieved the best mIoU and HD performance. Figure 4 shows several examples of the model outputs. The improved areas are annotated by yellow circles. Detailed improvements are circled in yellow circles. For example, in the first row of Figure 4, the cells segmented

**TABLE 3.** Evaluation results on WSI dataset.

| DATASET | WSI | |
|---|---|---|
| STRATEGY | MIOU | HD |
| FSL | 0.7463 | 78.9950 |
| SPL | 0.7878 | 75.0170 |
| SPLD-QBC, \|C\|=2 | 0.7850 | 74.7056 |
| SPLD-QBC, \|C\|=3 | 0.7758 | 74.7963 |
| SPLD-QBC, \|C\|=4 | **0.8090** | **65.3745** |

by SPLD with 4 members are fully connected regions without holes. However, holes exist in the segmented cells generated by the other four models. In addition, the number of false positive pixels is smaller in results obtained by the SPLD method than in the SPL- and FSL-generated results. This phenomenon indicates that learning in order from the easiest scenario to the hardest scenario facilitated model training, and the segmentation performance of the model was improved by adding a large number of members to the committee.

## V. CONCLUSION

In this paper, we proposed a versatile medical image semantic segmentation framework for medical image segmentation. The proposed approach can enhance the performances of medical image semantic segmentation models. By applying the query-by-committee (QBC) technique, we dynamically selected the optimal sequence of training samples from different probability distributions so that the models could achieve improved performances. To prevent the model from reaching local minima, we employed a clustering algorithm to guarantee data diversity and thus force the model to pursue global minima. The experimental results indicated that the proposed SPLD could significantly boost the segmentation performance of the model and that it is easy to embed into a CNN-based segmentation model.

## REFERENCES

[1] K. Hu, Z. Zhang, X. Niu, Y. Zhang, C. Cao, F. Xiao, and X. Gao, "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179–191, Oct. 2018.

[2] W. Dai, N. Dong, Z. Wang, X. Liang, H. Zhang, and E. P. Xing, "SCAN: Structure correcting adversarial network for organ segmentation in chest X-Rays," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 263–273.

[3] A. Arbelle and T. R. Raviv, "Microscopy cell segmentation via adversarial neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 645–648.

[4] Y. Bengio, "Curriculum learning," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 41–48.

[5] M. P. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, vol. 23, pp. 1189–1197.

[6] D. P Kingma and M. Welling, "Auto-encoding variational bayes," 2013, *arXiv:1312.6114*. [Online]. Available: http://arxiv.org/abs/1312.6114

[7] L. Jiang, "Self-paced learning with diversity," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2078–2086.

[8] L. Jiang, "Self-paced curriculum learning," in *Proc. AAAI*, 2015, p. 6.

[9] A. Graves, M. G. Bellemare, J. Menick, R. Munos, and K. Kavukcuoglu, "Automated curriculum learning for neural networks," 2017, *arXiv:1704.03003*. [Online]. Available: http://arxiv.org/abs/1704.03003

[10] T. Matiisen, A. Oliver, T. Cohen, and J. Schulman, "Teacher–Student curriculum learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3732–3740, Sep. 2020.

[11] T. Pi, X. Li, Z. Zhang, D. Meng, F. Wu, J. Xiao, and Y. Zhuang, "Self-paced boost learning for classification," in *Proc. IJCAI*, 2016, pp. 1932–1938.

[12] Q. Zhao, "Self-paced learning for matrix factorization," in *Proc. AAAI*, 2015, p. 4.

[13] L. Han, "Self-paced mixture of regressions," in *Proc. IJCAI*, 2017, pp. 1816–1822.

[14] D. Zhang, J. Han, L. Yang, and D. Xu, "SPFTN: A joint learning framework for localizing and segmenting objects in weakly labeled videos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 475–489, Feb. 2020.

[15] D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 865–878, May 2017.

[16] D. Zhang, J. Han, L. Zhao, and D. Meng, "Leveraging prior-knowledge for weakly supervised object detection under a collaborative self-paced curriculum learning framework," *Int. J. Comput. Vis.*, vol. 127, no. 4, pp. 363–380, Apr. 2019.

[17] N. Tong, S. Gou, T. Niu, S. Yang, and K. Sheng, "Self-paced DenseNet with boundary constraint for automated multi-organ segmentation on abdominal CT images," *Phys. Med. Biol.*, vol. 65, no. 13, Jul. 2020, Art. no. 135011.

[18] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[19] B. Settles, "Active learning literature survey," Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WA, USA, Tech. Rep., 2009.

[20] H. Fu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 132–139.

[21] J. U. Kim, H. G. Kim, and Y. M. Ro, "Iterative deep convolutional encoder-decoder network for medical image segmentation," in *Proc. 39th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 685–688.

[22] F. Wang, "Context-aware Spatio-recurrent curvilinear structure segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12648–12657.

[23] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, and Y. Yu, "Cross-modality deep feature learning for brain tumor segmentation," *Pattern Recognit.*, vol. 110, Feb. 2021, Art. no. 107562.

[24] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, Y. Wang, and Y. Yu, "Exploring task structure for brain tumor segmentation from multi-modality MR images," *IEEE Trans. Image Process.*, vol. 29, pp. 9032–9043, 2020.

[25] Q. Zhang, J. Wang, Z. Liu, and D. Zhang, "A structure-aware splitting framework for separating cell clumps in biomedical images," *Signal Process.*, vol. 168, Mar. 2020, Art. no. 107331.

[26] U. Nazir, "Tiny-inception-ResNet-v2: Using deep learning for eliminating bonded labors of brick kilns in South Asia," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 39–43.

[27] M. Raghu, "On the expressive power of deep neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2847–2854.

[28] D. D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *Machine Learning Proccessing*. Amsterdam, The Netherlands: Elsevier, 1994, pp. 148–156.

[29] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

[30] Z. Ma, "On convergence properties of implicit self-paced objective," *Inf. Sci.*, vol. 462, pp. 132–140, Dec. 2018.

[31] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*. [Online]. Available: http://arxiv.org/abs/1412.7062

[32] J. J. Kanski, *Bowling, Clinical Ophthalmology: A Systematic Approach*. Amsterdam, The Netherlands: Elsevier, 2011.

[33] J. Staal, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.

[34] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.

[35] T. Pohlen, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4151–4160.

[36] B. van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database," *Med. Image Anal.*, vol. 10, no. 1, pp. 19–40, Feb. 2006.

[37] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 577–590, Feb. 2014.

[38] A. Valada, "Adapnet: Adaptive semantic segmentation in adverse environmental conditions," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2017, pp. 4644–4651.

[39] I. Yoo, D. Yoo, and K. Paeng, "Pseudoedgenet: Nuclei segmentation only with point annotations," in *Medical Image Computing and Computer Assisted Intervention*. Springer, 2019.

[40] E. Arbel and I. A. Remer Ben-Dor, "Deep learning based partial annotation framework for instance segmentation in histopathology images," in *Proc. Int. Conf. Med. Imag. Deep Learn.–Extended Abstract Track*, 2019, pp. 1–4.

[41] S. Jégou, "The one hundred layers Tiramisu: Fully convolutional densenets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 11–19.

**XIAOYING PAN** received the B.S. degree in computer communications from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2002, and the M.S. degree in computer application technology and the Ph.D. degree in circuits and systems from Xidian University, Xi'an, in 2009. Since 2019, she has been working as a Professor with the School of Computer Science and Technology, Xi'an University of Posts and Telecommunications. She is the author of two books and more than 70 articles. Her research interests include medical image processing, artificial intelligence, and data mining.
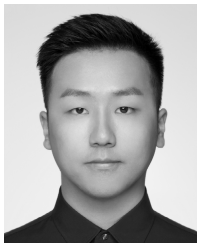
**DE WEI** received the B.S. degree in financial management from the College of Science and Technology, Xi'an University of Technology, Xi'an, China, in 2018. He is currently pursuing the M.S. degree in computer science and technology with the Xi'an University of Posts and Telecommunications. His research interests include medical image processing and deep learning.

**YIZHE ZHAO** received the B.S. and M.S. degrees in communication engineering from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2013 and 2017, respectively. She is currently pursuing the Ph.D. degree in computer science and technology with the School of Information Science and Technology, Northwest University, China. Her research interests include medical image processing and deep learning.

**MINJIE MA** received the B.S. degree in computer science and technology from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2020, where she is currently pursuing the M.S. degree in computer science and technology. Her research interests include medical image processing and deep learning.
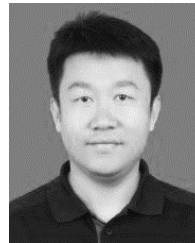
**HAO WANG** received the B.S. degree in computer communications from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2002, and the M.S. degree in computer application technology from Xidian University, Xi'an, in 2006. He is currently pursuing the Ph.D. degree in software engineering with Northwestern Polytechnical University. His research interests include deep learning and image processing.

• • •