Semi-Supervised Breast Histological Image Classification by Node-Attention Graph Transfer Network

LIHENG GONG, JINGJING YANG, AND XIAO ZHANG
School of Information Science and Engineering, Hebei North University, Zhangjiakou 075000, China
Corresponding author: Jingjing Yang (r78z@foxmail.com)

This work was supported in part by the Population Health Informatization in Hebei Province Engineering Technology Research Center, Hebei North University under Grant JYT2019016, and in part by the Zhangjiakou Municipal Science and Technology and Seismological Bureau under Grant 1911016C-9.

ABSTRACT
As a major cause of leading female death, breast cancer is often diagnosed by histological images which has been resolved by many deep learning methods with the assistance of large amounts of annotated data. However, their performances are severely limited by the lack of sufficient labeled data in clinical practice. This paper aims to relieve the annotating workload by a semi-supervised transfer learning algorithm to conduct knowledge distillation from a completely labeled source domain. To achieve this goal, we propose a node-attention graph transfer network to exploit the inherent correlation between individual samples by graph convolutional network, along with a cross-domain graph learning module to stimulate the graph construction in target domain. In the meanwhile, we design a node-attention mechanism to learn the individual contribution of each source image for target domain, which can further leverage the domain-gap by our node-attention transfer learning. Results of semi-supervised breast histological image classification with various scales of annotated training images are performable and further experiments demonstrate the significant contributions of each component we proposed.

INDEX TERMS Semi-supervised, histological image classification, graph convolutional network, node-attention, transfer learning.

I. INTRODUCTION
Breast cancer has severe health threats to women due to its considerable patients, who suffer from a rigorous mortality in the whole world. From the report of the World Heath Organization, breast cancer is the second threatening cancer disease for women patients after skin cancer. Though breast cancer has a high mortality, its deterioration can be prevented by the early diagnosis and appropriate treatment, which demonstrates the significance of responsive detection. As an important measurement, the histology examination can obtain sufficient images within a limited period to provide reliable diagnostic evidence, where costs large amounts of manpower to conclude the clinical manifestation even with misdiagnosis accidentally. To make this laborious work easier to execute, a number of automatic breast cancer diagnosis models have been proposed to improve the examination effectiveness and reduce the work intensity of pathologists.

Among existing breast histological image classification methods, they are mainly composed by hand-crafted feature and deep learning based methods on histological image analysis, which predicts the different type of the breast lesions. The first category of hand-crafted feature based methods employs prior knowledge to design robust feature representations, and utilize independent classifier on them [5], [25], [28]. Another solution for breast histological image classification is based on deep learning technology, with the developments of computing power (GPUs) in recent years, which combines feature learning and classification into a unified framework [17], [20], [40]. However, they have an important weakness that requires a large number of annotated samples to provide training direction, named as supervised learning. In contrast to general image annotation, the symbol in breast histological image is more complicated to discover...
than the general examine disease (e.g., skin cancer). Thus, annotating large amounts of the breast histological images is more challenging in practical and is sure to cost enormous time and need more well-trained domain professionals. In this paper, we propose a semi-supervised learning framework for solving this challenge, rather than unsupervised framework with poor performance. The semi-supervised learning only requires a small number of annotated samples to exploit the interactive influence from unlabeled data, which performance is between supervised and unsupervised approaches. There is a major concern that existing semi-supervised medical image classification methods ignore the interactive influence between image samples, which is illustrated by the graph convolutional network (GCN) with a graph learning module [12]. There exists a major challenge for the automatic histological image classification that is the limited amount of data available under supervised learning and temporally annotating a large amount of breast histological images is unsustainable in clinical applications as demonstrated in [1], [18], [30]. Nevertheless, transfer learning provides a novel ideology that has been proved to be effective for solving the challenge of limited labeled histology data [27], [29], [33], and it needs a completely labeled dataset as the source domain with a target domain of partially labeled data. When we exploit the interactive influence between samples in transfer learning, there exists a major problem that the topology graph learning cannot be well transferred because the semi-supervised GCN [12] only builds graph learning in a single domain, which will lose efficacy moving to another.

To achieve both of the transfer learning and interactive correlation exploration for solving histological images classification, a novel Semi-Supervised Transfer Learning (SSTL) algorithm is designed to integrate semi-supervised GCN and transfer learning into an unified framework. As the important component in GCN, graph learning should be transferable between different domains in our SSTL algorithm. In order to leverage the source samples with different contributions to target domain, it is necessary to estimate the importance of each source image (illustrated in Figure 1), which can be solved by the attention mechanism.

According to the analysis above, this paper proposes a novel Semi-Supervised Graph Transfer Network (NaGTN) for SSTL problem, which is composed of three crucial modules, including graph representation, cross-domain graph learning, and node-attention transfer learning. This network jointly integrates the graph convolutional network into transfer learning architecture by the cross-domain graph learning, along with a node-attention mechanism. Firstly, NaGTN uses Convolutional Neural Network (CNN) to learn sample-level CNN feature representation and then builds the cross-domain graph structure by the cross-domain graph learning module. After that, the learned graph and the sample-level CNN features are together fed into a graph convolutional network to train the semi-supervised classifier, which leverages the domain gap by the node-attention transfer learning, and learns the contributing degree of source image to target domain, optimized by a domain critic constraint.

Our contributions are summarized below. (1) A novel semi-supervised transfer learning framework of Node-attention Graph Transfer Network (NaGTN) is proposed for breast histological image classification; (2) A transferable graph learning mechanism is designed to assist the graph convolution; (3) A node-attention transfer learning approach for source images is employed to leverage the domain gap between source and target domains; (4) Extensive experiments of breast histological image classification are implemented to evaluate the effectiveness of NaGTN.

II. RELATED WORK

Up to now, many feature learning researches have been designed for histological image classification. Those methods can be roughly categorized into hand-crafted and learning based methods.

From the successful application in natural image classification, several works [5], [25], [36] adopted conventional feature descriptors, e.g. histogram, SIFT [21], HOG [7], and LBP [26] to represent the histological images, and achieved effective performance in histological image classification. In the meanwhile, many researches [16], [22] attempted to learn representations according to the pathologist’s point of view. Kowal et al. [16] utilized the statistics in terms of the morphology of cells in histological images, where the statistic can represent the general characteristics of the cell nuclei, yet may lose some important information. Furthermore, the graph-based representations [3], [9] and mixed features [23], [39] were integrated to histological image classification for extracting meaningful semantic information from the images, which have achieved satisfactory performance.

Due to the development of Convolutional Neural Networks (CNN) in recent years, more and more research works...
introduce CNN into histological image classification [34], [38], [40], which integrates feature learning and classification model into an unified framework without any hand-designed work. For example, Wang et al. [34] devised a classification framework based on histology images by combining deep learning with machine learning methodologies. It proposed a multi-network feature extraction model by using pre-trained deep convolution neural network, and developed an effective feature dimension reduction method with an ensemble support vector machine. Yu et al. [38] proposed an automatic breast cancer detection method based on hybrid features, where utilizes a 3-output convolutional neural network to segment the nuclei, and extract the weak correlation between the hematoxylin and eosin channels, along with the texture feature.

Besides, a popular strategy of transfer learning aims to relax the limitation of annotations, and it has been adopted for diagnosing breast cancers. For example, De et al. [8] utilized the transfer learning to extract features from Histological Images (HI) using the Inception-V3 CNN pre-trained with ImageNet with a support vector machine classifier on tissue labeled colored cancer dataset, which aims to filter the patches from a breast cancer HI and remove the irrelevant ones; Alinaif et al. [1] compared two common techniques to deal with limited domain data by using deep features and fine-tuning convolutional neural networks, which demonstrate that using feature vectors with a classical support vector machine for training and testing can lead to higher accuracy on publicly available datasets; Chougrad et al. [6] proposed a multi-label transfer learning with end-to-end image representation learning and a novel customized label decision scheme, which can estimate the optimal confidence for each visual concept.

Inspired by these successful applications of CNN and transfer learning, this paper employs the cross-domain graph learning and node-attention graph feature learning to implement a joint semi-supervised transfer learning algorithm for automatic histological image classification.

III. OUR APPROACH

Aiming at solving semi-supervised histological image classification problem, this paper is inclined to utilize a completely annotated source domain to accelerate the training efficiency of target data, which is often achieved by transfer learning. Concretely, there are two crucial obstacles to be solved. Firstly, the inherent correlations between the labeled and unlabeled samples should be exploited both for source and target domains. Secondly, several source samples that have less contribution to target domain should weaken their influence over the transfer learning process.

In order to settle the first obstacle, we employ the Graph Convolutional Network (GCN) to synthesize the topology correlation and image features into an unified graph feature representation framework under transfer learning manner, which can explore the topological influence between unlabeled and labeled samples. To tackle another problem, we design a node-attention mechanism to measure the contributing degrees of each source samples, where the sample with higher degree could provide more discriminative information for the classification task in target domain. Specifically, this paper proposes a Node-attention Graph Transfer Network (NaGTN), by designing two modules of cross-domain graph learning and node-attention transfer learning for the semi-supervised feature representation of histological images. As illustrated in Figure 2, NaGTN is comprised by three major components, including modules
of graph feature representation, cross-domain graph learning, and node-attention transfer learning. By the integration of them, NaGTN can exploit the cross-domain discriminative correlations both in source and target domains by GCN, and learn a node-attention graph feature representation for semi-supervised histological image classification by the knowledge distillation from source samples with higher contributing degrees. By the proposed NaGTN approach, this paper enables to train an efficient classification model from limited labeled data and abundant unlabeled data from the target domain.

A. GRAPH FEATURE REPRESENTATION

We conduct node-attention transfer network between a labeled source domain \( X^s = \{x^s_i\}_{i=1}^{N_s} \), with the correlated label set \( Y^s = \{y^s_i\}_{i=1}^{N_s} \), and a partially labeled target domain \( X^t = \{x^t_i\}_{i=1}^{N_t} \), where \( x^s_i/x^t_i \) is a source/target sample and \( y^s_i \) is the category annotation of \( x^s_i \). \( N_s/N_t \) is the image number of source/target images. Note that, there are several target samples containing labels, which is represented by \( Y^t = \{y^t_j\}_{j=1}^{N_t^l} \) and \( y^t_j \) is the label of \( j \)-th target sample. The number of labeled and unlabeled target images are \( N_t^l \) and \( N_t^u \) (\( N_t^l + N_t^u = N_t \)), respectively.

For the graph feature learning, we firstly extract independent representations for each image in source and target domains by a backbone CNN (Convolutional neural Network) network, and secondly conduct graph convolutional network on the learned CNN features.

Formally, the backbone CNN for the independent feature representation is defined by \( r(\cdot; \theta_r) \), where \( \theta_r \) is the trainable parameters. The CNN feature collection \( H_c \) of source and target domains is defined by,

\[
H_c = \{h^c_i|_{i=1}^{N_s}, h^c_j|_{j=1}^{N_t}\} = \{r(x^s_i, \theta_r)|_{i=1}^{N_s}; r(x^t_j, \theta_r)|_{j=1}^{N_t}\} \quad (1)
\]

where \( h^c_i/h^c_j \) is the learned \( i/j \)-th CNN feature for source/target sample \( x^s_i/x^t_j \). Importantly, the parameters \( \theta_r \) is shared across domains, which can be trained by the transfer learning with excellent feature learning capability both for source and target domains [11], [19].

Aiming to learn graph representation in the second feature learning stage, we introduce graph convolutional network to exploit the inherent correlations between the samples both in source and target domains, where GCN has been proved its high-efficiency in several semi-supervised learning methods [12], [15]. In particular, the shared GCN is defined by \( G(A, H_c) \), where \( A \in \mathbb{R}^{(N_s+N_t) \times (N_s+N_t)} \) is the pairwise relationship adjacent matrix and \( A_{ij} \) denotes the relationship (such as distance, similarity) between \( i \)-th and \( j \)-th samples in source and target domains.

GCN offers an essential operation to jointly integrate topology correlations and node features for extracting valuable graph representations. For the given topology structure relationships \( A \) and CNN feature set \( H_c \), the graph convolution is conducted following a layer-wise propagation in hidden layers, where the graph representation \( H_g \) is learned by,

\[
H_g^{(k)} = \sigma \left( A H_g^{(k-1)} W_g^{(k)} \right) \quad (2)
\]

where \( H_g^{(k-1)} \) denotes the output of the previous GCN layer, and the first input of \( H_g^{(0)} = H_c \). \( W_g^{(k)} \) is the trainable parameters in \( k \)-th hidden layer, and \( \sigma(\cdot) \) represents the activation function. As demonstrated in Figure 2, NaGTN adopts graph convolutional layers to explore the inherent discriminative correlations across source and target domains, where the \( W_g^{(k)} \) is shared in them so that conducts the cross-domain knowledge distillation in transfer learning procedure. To simplify the formula, we re-define the GCN as \( g(A, H_c; \theta_g) \), which employs the topology adjacent matrix \( A \) and CNN feature set \( H_c \) as inputs, and \( \theta_g \) is the optimizable convolutional parameters. In this feature learning stage, the graph representation of images from source and target can be mathematically represented by,

\[
H_g = \{h^g_i|_{i=1}^{N_s}, h^g_j|_{j=1}^{N_t}\} = \left\{ g(A, h^c_i; \theta_g)|_{i=1}^{N_s}; g(A, h^c_j; \theta_g)|_{j=1}^{N_t} \right\} \quad (3)
\]

B. CROSS-DOMAIN GRAPH LEARNING

Though GCN can exploit correlations between image samples as illustrated above, there still remains a major problem of GCN to be solved. That is the graph learning (topology structure building) of \( A \) in Eq. 3. The popular solution for this problem is to employ \( k \)-nearest neighbor rule to establish graph, which is too inflexible to conquer the transfer learning in our network. Therefore, we propose a cross-domain graph learning method to establish the graph structure \( A \) in Eq.3 with sufficient scalability across different domains, which is demonstrated below.

To achieve cross-domain graph learning, we build three different graph building layers by a nonnegative function \( A_{ij} = b(h^c_i, h^c_j; \theta^d) \) that transforms the pairwise correlation between any features \( h^c_i \) and \( h^c_j \) in \( H_c \) into an element of adjacent matrix, and \( \theta^d \) denotes the parameters for different domains, where \( d = \{s, t, st\} \) that \( \theta^d_s, \theta^d_t \) denotes the graph building parameters for source/target domain, and \( \theta^d_{st} \) is the correlation between pairwise samples from different domains. Mathematically, we implement the cross graph learning functions following,

\[
A_{st,ij} = \frac{\exp(\text{ReLU}(\theta^d_{st}^T [h^c_i - h^c_j]))}{\sum_{l=1}^{(N_s+N_t)} \exp(\text{ReLU}(\theta^d_{st}^T [h^c_i - h^c_l]))} \quad (4)
\]

\[
A_{si,ij} = \frac{\exp(\text{ReLU}(\theta^d_{si}^T [h^c_i - h^c_j]))}{\sum_{l=1}^{(N_s+N_t)} \exp(\text{ReLU}(\theta^d_{si}^T [h^c_i - h^c_l]))} \quad (5)
\]

\[
A_{ti,ij} = \frac{\exp(\text{ReLU}(\theta^d_{ti}^T [h^c_i - h^c_j]))}{\sum_{l=1}^{(N_s+N_t)} \exp(\text{ReLU}(\theta^d_{ti}^T [h^c_i - h^c_l]))} \quad (6)
\]

where the characteristic of \( \sum_{l=1}^{N_s+N_t} A_{ij} = 1, A_{ij} \geq 0 \) is constrained by \( \text{ReLU}(\cdot) = \max(0, \cdot) \).
In the real cross-domain graph topological correlations, the intra-domain correlations are stronger and more intensive than the cross-domain connections. That is, \( \sum_{v_i \in V} A_{s,i,j} + \sum_{v_i \in V} A_{t,i,j} > 2 \sum_{v_i \in V} A_{s,i,j} \) because the relationships between samples in different domains are scarce. In addition, we expect that smaller distance between samples has a larger value \( A_{i,j} \). To achieve these goals, we design the Cross-domain Graph Learning (CGL) loss function to learn the optimal parameters \( \theta^g \),

\[
L_{CGL} = \sum_{v_i \in H} \| h_i^g - h_i^t \|_2 \| A_{i,j} \|
- \lambda_1 (\sum_{v_i \in V} A_{s,i,j} + \sum_{v_i \in V} A_{t,i,j} - 2 \sum_{v_i \in V} A_{s,i,j} - \alpha)
+ \lambda_2 \| A \|_F^2
\]

where the term of \( \| A \|_F^2 \) encourages the scarcity of \( A \), and \( \alpha \) is a margin parameter.

As discussed before, the cross-domain graph learning module proposed ensures the reasonable graph topology structure by three graph building layers, constrained by the CGL loss. That provides the cross-domain graph structure for the graph transfer network, which will be illustrated in next subsection.

**C. NODE-ATTENTION TRANSFER LEARNING**

Though the cross-domain graph can be built to serve the graph convolutional network, the transfer learning for semi-supervised histological image classification is another major task in this paper. Aiming at learning the independent contributing degrees of source samples to target domain, we propose a node-attention transfer learning module in our graph transfer network, which can integrate the contributing degree into the graph feature representation when conduct knowledge distillation between source and target domains.

Based on different source nodes have individual importances to target domain, we firstly estimate the contribution of each node in the cross-domain graph by a fully connected layer on the learned GCN features, which is named as node-attention mechanism. Then, we integrate the importance into the calculation of final feature representation for the semi-supervised histological image classification.

In particular, given the learned graph feature set \( H_g = \{h_i^{s,v}_i; h_j^{t,v}_j\} \) after GCN, the node-attention mechanism attaches a fully connected layer and a sigmoid function on each feature to calculate the attention weight of its contribution to target domain. Mathematically, the contribution attention score \( \mu^i \) of \( i \)-th source image is represented by,

\[
\mu^i = a(h_i^{s,v}; \theta_a)
\]

where \( a(\cdot) \) is the contribution estimator composed by a fully connected layer and a sigmoid function, and \( \theta_a \) is the trainable parameters.

Otherwise, the self-contribution importance \( \mu^j \) of \( j \)-th target image to the target domain can be calculated by,

\[
\mu^j = a(h_j^{t,v}; \theta_a)
\]

where the parameters \( \theta_a \) are shared between the calculations of \( \mu^i \) and \( \mu^j \).

Through the obtained contributions of each image for target domain, we introduce them into the final feature representations of source images, which are calculated by,

\[
h^i_t = \mu^i h^i_g \frac{1}{1/N_t \sum_{j=1}^{N_t} \mu^j_t}
\]

and the final feature of target images are obtained by,

\[
h^j_t = \mu^j h^j_g \frac{1}{1/N_t \sum_{j=1}^{N_t} \mu^j_t}
\]

where we can obtain all the final image features \( H = \{h^i_t; h^j_t\} \) in source and target domains.

To guarantee the reflection of image contributions to target domain, we design a Node-Attention (NA) loss to enforce the maximum estimated contribution of source images is smaller than the average contribution of target images, because the target images are the main contributors to the target distribution. Mathematically, the NA loss is presented by,

\[
L_{NA} = \max\{0, \beta - (\mu^s - \mu^s_{max})\}
\]

where \( \mu^s = \frac{1}{N_s} \sum_{j=1}^{N_s} \mu^s_j \), \( \mu^s_{max} \) is the maximum value in \( \mu^s_{i=0} \), and \( \beta \) is a hyper parameter as a margin.

After node-attention feature representation, we further alleviate the domain gap between source and target domains by the empirical Domain Critic (DC) loss by,

\[
L_{DC} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mu^s_i h^i_g - \frac{1}{N_t} \sum_{j=1}^{N_t} \mu^j h^j_g
\]

With respect to the losses of node-attention, and domain critic, NaGTN can achieve the node-attention transfer learning in the knowledge distillation between source and target domains and obtain the final features \( H = \{h^i_t; h^j_t\} \). Note that, our node-attention transfer learning is flexible enough to be integrated into the semi-supervised learning, which is described in Section III-D.

**D. SEMI-SUPERVISED LEARNING**

In NaGTN approach, the node-attention transfer learning based graph feature representations are fed into a classifier with annotations, following the semi-supervised learning configuration in [15]. The classifier is achieved by a single layer logistic regression classifier, which is represented by \( f(H; \theta_f) \), where \( \theta_f \) denotes the trainable parameters in the classifier. Then, the classifier predicts the category probabilities of the annotated samples with their labels \( Y = \{Y^s; Y^t\} \) in the source and target domains, by,

\[
\hat{Y} = f(H; \theta_f)
\]

where \( \hat{Y} \) is the predicted probability matrix of the annotated samples both in the source and target domains.
Then, we attach a sigmoid activation function on the prediction matrix, and utilize the Cross-Entropy (CE) loss on the all labeled data as,

$$L_{CE} = -\frac{1}{N_s + N_i} \sum_{i=1}^{N_s} \sum_{k=1}^{C} Y_{ik} \log(\hat{Y}_{ik})$$  \hspace{1cm} (15)$$

where $C$ is the category number.

Based on the graph feature representation, cross-domain graph learning and node-attention transfer learning modules, our proposed NaGTN model can be optimized by the overall loss function,

$$L = L_{CE} + \gamma_1 L_{DC} + \gamma_2 L_{NA} + \gamma_3 L_{CGL}$$  \hspace{1cm} (16)$$

where $\gamma_1$, $\gamma_2$, and $\gamma_3$ are balance parameters to these four terms. The summary of our NaGTN model is concluded in Algorithm 1.

**Algorithm 1 Node-Attention Graph Transfer Network (NaGTN)**

**Input:** Source images $X^s = \{x^s_i\}_{i=1}^{N_s}$ and their labels $Y^s = \{y^s_i\}_{i=1}^{N_s}$; Target images $X^t = \{x^t_i\}_{i=1}^{N_i}$ and their partial labels $Y^t = \{y^t_i\}_{i=1}^{N_i}$; Coefficients of $\lambda_1$, $\lambda_2$, $\alpha$, $\beta$, $\gamma_1$ and $\gamma_3$; learning rates $lr$.

**Initialization:** The parameters $\theta_s$ for CNN feature extractor, $\theta_g$ for GCN module, and $\theta_b^{\beta}, \theta_b^{\alpha}, \theta_b^{\gamma}$ for graph building layers, $\theta_a$ for node-attention learner, and $\theta_f$ for semi-supervised classifier.

**While not converge do**  ///Forward propagation

Obtain CNN feature representation $H_c$ by Eq.1 with $\theta_s$;

Construct graph structure $A$ by Eq.4, 5, and 6 with $\theta_b^{\beta}, \theta_b^{\alpha}, \theta_b^{\gamma}$;

Obtain GCN feature representation $H_g$ by Eq.3 with $\theta_g$;

Learn node-attention features $H$ by Eq.10 and 11 with $\theta_a$;

Compute the gradients for each parameter ($\theta_s, \theta_g, \theta_b^{\beta}, \theta_b^{\alpha}, \theta_b^{\gamma}, \theta_a, \theta_f$) according to back propagation of Eq.16.

Update the parameters $\theta_s, \theta_g, \theta_b^{\beta}, \theta_b^{\alpha}, \theta_b^{\gamma}, \theta_a, \theta_f$ by Eq.16 with learning rate $lr$ and their gradients.

**End**

**Return** the parameters of NaGTN.

**IV. EXPERIMENTS**

**A. DATABASE AND CONFIGURATION**

In our Node-attention Graph Transfer Network (NaGTN), two public databases of the ICIAR grand challenge [2] (BACH) and BreakHis [31] are employed as source and target domains, separately. Both of them are captured by microscope of the whole slide which is stained by hematoxylin and eosin (H&E). BACH dataset contains 400 histological images of breast cell in H&E stained which is captured by microscope as 2048 × 1536 pixels. All of them are captured in magnification of 200× and pixel size of 0.42μm × 0.42μm, and are annotated by four categories of benign, in situ carcinoma, invasive carcinoma, and normal, where each category includes 100 images. We also divide them into benign (benign and normal) and malignant (in situ carcinoma, and invasive carcinoma). As for the BreakHis dataset, it consists 7909 images and 8 sub-categories of breast cancer, which is a challenging database in large scale. To keep consistence with our binary classification task, we summarize the multiple sub-categories into benign and malignant, which follows the binary classification in BACH dataset. In our experiments, we set the BACH as source domain to conduct classification on BreakHis dataset, which is the evaluated database in charge of demonstrating the effectiveness of our NaGTN.

As for the network configuration, we implement the whole network by PyTorch on two NVIDIA GeForce 2080Ti GPUs, and employ the ImageNet pre-trained ResNet [10] model as the initialization of our CNN feature extractor. In addition, we set two graph convolutional layers to learn graph representation, which follows the research [12]. Furthermore, Adam optimizer [14] is utilized to update the weight parameters of each layer with an initial learning rate of 0.01 which will be ×0.1 after 50 epochs. For the coefficients in each losses, parameters $\lambda_1 = 0.8$, $\lambda_2 = 0.2$, $\alpha = 0.02$, $\beta = 0.03$, $\gamma_1 = 0.5$, $\gamma_2 = 0.5$ and $\gamma_3 = 0.3$ are chosen to achieve the best performance of NaGTN, and the maximum number of training epochs is set as 100. For the setting of semi-supervised learning, we randomly split the BreakHis into testing (20%) and training (80%) set following research [37], and define the Annotated Percentage (AP) of target images (BreakHis) to represent the annotated images in network training, while the source images (BACH) are completely labeled. To better demonstrate the effectiveness, we randomly conduct 10 times for each experiments and report the average value of them.

**B. EVALUATION METRICS**

In the experiments of NaGTN, we validate the performance of the network on BreakHis dataset, with several measurements, including accuracy, precision, Recall and F1-score. Their detailed definitions are illustrated below.

- Accuracy = \( \frac{TP + TN}{TP + TN + FP + FN} \)
- Precision = \( \frac{TP}{TP + FP} \)
- Recall = \( \frac{TP}{TP + FN} \)
- F1-score = \( \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \)

In these definitions, $TP$, $TN$, $FP$ and $FN$ denotes the number of True Positive, True Negative, False Positive, and False Negative, individually. Moreover, we adopt Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) to evaluate the performance of classifier. Note that, we report the average results for the various images under 40×, 100×, 200×, and 400× magnifications in the NaGTN and compared methods.

**C. EXPERIMENTAL RESULTS**

We report the evaluated performance on target dataset BreakHis when the BACH dataset is employed as the source domain. Table 1 reports the evaluation results (Accuracy,
TABLE 1. The results of NaGTN with different annotated percentage (AP) of training images.

| AP  | Accuracy | Precision | Recall | F1-score |
|-----|----------|-----------|--------|----------|
| 20% | 0.865    | 0.979     | 0.746  | 0.847    |
| 40% | 0.894    | 0.980     | 0.804  | 0.883    |
| 60% | 0.928    | 0.892     | 0.974  | 0.931    |
| 80% | 0.964    | 0.977     | 0.950  | 0.963    |
| 100%| 0.961    | 0.987     | 0.934  | 0.960    |

Precision, Recall, and F1-score) of our NaGTN, given different percentage AP (20%, 40%, 60%, 80%, 100%) of annotated training images. From the results, given 80% labeled training images, our NaGTN approach achieves the best results of 0.964, 0.977, 0.950, and 0.963 for accuracy, precision, recall and F1-score, individually. Along with the other annotated percentage, NaGTN obtains weaker results than AP of 80% but satisfactory results, e.g. 0.865 accuracy with 20% annotated training images. From the 20% annotated training images, it can be observed that the model reaches more than 80% accuracy but save 80% annotating workload. Moreover, the NaGTN model achieves more than 90% of accuracy at 60% labeled training images, combined with unannotated samples. Thus, our network significantly relieves the annotating workload with a competitive performance.

To explicitly demonstrate the effectiveness of our NaGTN model, we draw the ROC curve with AP of 80%, combined with its AUC value, as shown in Figure 3. Through the analysis of the ROC curve, it is also obvious that our NaGTN obtains performable classification results with partially labeled training images (semi-supervised learning), and the AP of 80% annotated images facilitates the performance (AUC=0.93), which illustrates that the semi-supervised framework generates positive effect to the semi-supervised breast histological image classification task.

Furthermore, to evaluate the separability of the learned graph features before feeding into the classifier, we visualize the final features (AP of 80%) from target domain \( \{h_j^t\}_{j=1}^N \) into 2-d domain by t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization [24]. t-SNE converts similarities between data points to the joint probabilities and tries to minimize the Kull-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data, thus it is a powerful and popular tool to visualize the representation capability of the learned features. As shown in Figure 4 (a), NaGTN performs the separable feature points, where the red dots associate the benign sample features and green points represent the malignant features. From the t-SNE performance along with 80% AP, it is proved that the graph feature representations have better separability than other compared methods Figure 4 (b-d), and keep consistence with the reported experimental results in Table 1 and ROC curves 3.

To visualize the convergence of our NaGTN method, the variation curves of training and testing accuracy with AP of 80% per epoch have plotted in Figure 5. The variation curves clearly show that NaGTN can successfully solve the overfitting problem, which is usually caused by limited labeled data. By introducing the semi-supervised learning, our NaGTN is rapidly stabilized in the training process. The testing accuracy curve also conforms to the training curve. This visualization of our NaGTN makes the semi-supervised learning of breast histological image classification overcome the overfitting problem and appears a robust classification trend.

D. COMPARISON WITH RECENT SEMI-SUPERVISED GCN METHODS

To reveal the superiority of our proposed NaGTN, several GCN based semi-supervised classification methods are
compared to conduct on the BreaKHis dataset. The compared methods are recently proposed, consisting of GCN (ICLR17) [15], GAT (ICLR18) [32], and GLCN (CVPR19) [12] with the annotated percentage of 80% and their settings follows the original papers. GCN [15] developed the CNN architecture into a scalable semi-supervised learning approach on graph-structured data, which is the first application of graph convolutional network. GAT [32] is a graph attention network that leverages masked self-attention layers to address the shortcomings of prior methods based on graph convolutions or their approximations. GLCN [12] proposed a semi-supervised graph learning-convolutional network for graph data representation and it can learn an optimal graph structure that best serves graph CNN by integrating graph learning and graph convolution into an unified network architecture.

TABLE 2. Comparison with semi-supervised GCN methods (AP=80%), and transfer learning models.

| Method  | Accuracy | Precision | Recall | F1-score |
|---------|----------|-----------|--------|----------|
| GCN [15] | 0.826    | 0.837     | 0.816  | 0.826    |
| GAT [32] | 0.858    | 0.864     | 0.852  | 0.858    |
| GLCN [12] | 0.916    | 0.920     | 0.919  | 0.919    |
| DTL [8]  | 0.893    | -         | -      | -        |
| DFTL [1] | 0.980    | -         | -      | -        |
| ECNN [13] | 0.981    | -         | -      | -        |
| NaGTN   | 0.964    | 0.977     | 0.950  | 0.963    |

Their results are summarized in Table 2, and our NaGTN achieves the best result, compared to those semi-supervised GCN models. Our approach outperform them at least 4.8% of accuracy (GLCN), and it is also superior in other evaluation metrics (Precision, Recall, and F1-score). Besides, their t-SNE and feature map visualizations are drawn in Figure 4, and 6. Compared to them, the NaGTN achieves the best separated feature points in t-SNE, and its feature map is more clears than others. Different from them, the most contributions of our NaGTN are introducing transfer learning and the node-attention mechanism into GCN framework by knowledge distillation from a completely labeled source domain. That guarantees the NaGTN can learn a robust graph structure in the target domain and produce a better attention on the important samples, while the compared method only utilizes hard or nontransferable graph building strategy in the single domain. The compared experiments further prove that our NaGTN can solve the semi-supervised histological image classification effectively by the assistance of source domain, and it has preferable results than the recently proposed GCN methods.

E. COMPARISON WITH RECENT BREAST CANCER CLASSIFICATION METHODS BASED ON TRANSFER LEARNING

To present the overall superiority more intuitively, three recent state-of-the-art transfer learning methods for automatic breast cancer diagnosis are employed as baselines to compare, including DTL [8], DFTL [1], and ECNN [13]. DTL [8] transfers knowledge from double source domains with two separate steps from a different and similar datasets, successively. It firstly transfers feature representations from the ImageNet. And then utilize CRC to provide the structural information about the tissue types of histological images with a SVM classifier. Finally, the target domain is to classify histopathologic images from the BreaKHis dataset into malign and benign. DFTL [1] introduces infinite latent feature selection for the pre-trained CNN model, and employs the fine-tuning transfer learning into the target domain,
which utilizes the target labels into the training procedure. ECNN [13] applies various pre-processing and CNN tuning techniques such as stain-normalization, data augmentation, hyper-parameter tuning, and fine-tuning on the ensemble breast cancer classification models of VGG19, MobileNet, and DenseNet, which is for feature representation and extraction with multi-layer perceptron classifier. Note that, both of DTL and ECNN models transfer knowledge from source domain but fine-tune the parameters in the target domain assisted by the target labels. Thus, both of them are intrinsically supervised methods.

The results of DTL, DFTL, and ECNN are also reported in Table 2. It is obvious that NaGTN outperforms DTL with a margin of 7.1%, while DTL is an unsupervised transfer learning method. Besides, our method presents a weakness no more than 1.7% compared to DFTL and ECNN. In practice, this comparison is conducted between our SSTL method and state-of-the-art transfer learning methods under supervision (DTL) or unsupervised frameworks (DFTL and ECNN). The results in Table 2 present that our method has a superior progress than unsupervised method, but a negligible distance to the supervised methods. The proposed SSTL method (NaGTN) has significant value in practical applications with limited labeled target data (80%).

F. DISCUSSIONS ON MAJOR COMPONENTS

Through the experimental analysis above, it can be observed that our node-attention graph transfer network achieves satisfactory results in semi-supervised breast histological image classification, while the effects of major components haven’t discussed. Thus, this subsection evaluates the influence of each important module in NaGTN, including node-attention mechanism, graph convolutional layers, and the cross-domain graph learning modules.

As comparison, we modify the original NaGTN into three baselines. Specifically, the node-attention mechanism (node-attention loss) is removed, which is denoted by GTN. As for the GCN, we remove graph convolutional layers (directly feeding CNN features into semi-supervised learning, NaTN). As for the other major contribution of NaGTN, the cross-domain graph learning module is replaced by KNN graph structure (Node-attention Transfer Graph Convolutional Network, NTGCN). We compare the results of these modified method and our original NaGTN in Table 3, with an unified annotated percentage of 80%. Furthermore, to demonstrate the overall influence of each module, we conduct the parameters analysis for $\gamma_1$, $\gamma_2$, and $\gamma_3$ with AP of 80%.

1) NODE-ATTENTION MECHANISM

Node-attention mechanism provides the importance of each source image to target domain, which can leverage the contributed knowledge in source domain to stimulate the network training for semi-supervised classification in target domain. From the results in Table 3, GTN achieves 0.924, 0.955, 0.890, and 0.921 of accuracy, precision, recall and F1-score, respectively. The results are weaker than NaGTN, which denotes the node-attention mechanism contributes an accuracy increasing of 4% with AP of 80%. The reason why is that the node-attention is practical for evaluating the different importance for each node and produces a reasonable inference for the messaging in GCN. So as to prove that the node-attention transfer learning in our graph convolutional network architecture is a major contribution to the semi-supervised histological image classification, which is indispensable in our NaGTN.

2) GRAPH CONVOLUTIONAL NETWORK

To exploit the inherent correlation between image samples in feature extraction, this paper introduces graph convolutional network to assist the semi-supervised learning. This modified method without graph convolution layers (NaTN) only achieves 0.830 of accuracy, which is lower than NaGTN of 13.4% (Table 3). This comparison demonstrates the importance of the correlation between individual samples when we conduct semi-supervised learning on histological image classification, because GCN module can learn the inherent correlations and alleviate the distribution gap between labeled and unlabeled images according to the graph topology structure in them.

3) CROSS-DOMAIN GRAPH LEARNING

As for the cross-domain graph learning module, it transfers the topology constructing ability to target domain, which enhances the graph learning ability in target domain without needing of sufficient annotated data. The validated method NTGCN obtains 0.917 of accuracy, which is less than NaGTN with a 4.7% distance. That illustrates the cross-domain graph learning module plays an important role in the graph transfer network. The reason of this result is that the graph topology construction of source domain is a significant stage in GCN, which transfers the correlation knowledge to target domain.

4) PARAMETER ANALYSIS

Besides the impacts of the major modules, the crucial components ($L_{DC}$, $L_{NA}$, and $L_{CGL}$) in the overall loss $L$ (Eq.16) contribute considerable progress for the semi-supervised method. To evaluate their effects on the network, we set the hyper-parameters of $\gamma_1$, $\gamma_2$, and $\gamma_3$ by the values in [0:0.1:1] to report their accuracy performance, as shown in Figure 7.

The parameter $\gamma_1$ aims to balance the term $L_{DC}$, and it achieves 93.2% when $\gamma_1 = 0$, which produces the best results.
performance at 0.5. That demonstrates domain critic loss contribute a improvement of 3.2% in accuracy, and expresses the stable progress up to 0.5. At the mean while, the parameter γ₃ for Lₕₐₜ₅ presents the same trend line with a contribution of 4% accuracy. As for the hyper-parameter of γ₃, it makes a progress of 6.5% and achieves the best performance when γ₃ = 0.3. The overall evaluation of the hyper-parameters demonstrates that the optimization of NaGTN network has close relation to the hyper-parameters in Eq. 16, and the losses of Lₚₐₜ₈, Lₙₐₐ₅ and Lₕₐ₅ₖ are essential for the semi-supervised learning.

V. CONCLUSION

Semi-supervised learning of breast histological image classification provides an effective solution to the lacking of sufficient annotated data in practice. This paper designs a semi-supervised transfer learning algorithm to strengthen the performance by a proposed Node-attention Graph Transfer Network (NaGTN). This approach aims to exploit the inherent correlation between labeled and unlabeled samples by Graph Convolutional Network (GCN), and utilizes a completely labeled source domain to conduct knowledge distillation for the target domain, with a cross-domain graph learning module proposed to reinforce the graph learning ability in target domain. Moreover, NaGTN learns the individual importance of each source image related to target domain, because the contribution degrees of source samples are various for the target model. The proposed NaGTN is implemented on the BreaKHis dataset to show its effectiveness on semi-supervised breast histological image classification, which achieves 89.4% with only 40% annotated training images, and reaches 86.5% when employs less labeled data of 20%. Further evaluated experiments also demonstrate the significant contributions of node-attention learning, cross-domain graph learning and the graph convolutional network on the semi-supervised breast histological image classification task.

REFERENCES

[1] S. Alinsaif and J. Lang, “Histological image classification using deep features and transfer learning,” in Proc. 17th Conf. Comput. Robot Vis. (CRV), May 2020, pp. 101–108.
[2] G. Aresta et al., “BACH: Grand challenge on breast cancer histology images,” Med. Image Anal., vol. 56, pp. 122–139, Aug. 2019.
[3] A. N. Basavanahally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhattacharjee, and A. Madabhushi, “Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology,” IEEE Trans. Biomed. Eng., vol. 57, no. 3, pp. 642–653, Mar. 2010.
[4] C. Baur, S. Albarquoni, and N. Navab, “Semi-supervised deep learning for fully convolutional networks,” in Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. Quebec City, QC, Canada: Springer, 2017, pp. 311–319.
[5] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, “Histopathology image classification using bag of features and kernel functions,” in Proc. Conf. Artif. Intell. Med. Eur. Verona, Italy: Springer, 2009, pp. 126–135.
[6] H. Chougad, H. Zouaki, and O. Alheyane, “Multi-label transfer learning for the early diagnosis of breast cancer,” Neurocomputing, vol. 392, pp. 168–180, Jun. 2020.
[7] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Jun. 2005, pp. 886–893.
[8] J. D. Matos, A. D. S. Britto, L. E. S. Oliveira, and A. L. Koerich, “Double transfer learning for breast cancer histopathologic image classification,” in Proc. Int. Joint Conf. Neural Net. (IJCNN), Jul. 2019, pp. 1–8.
[9] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, “Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features,” in Proc. 5th IEEE Int. Symp. Biomed. Imag., From Nano Macro, May 2008, pp. 496–499.
[10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.
[11] J. Hu, J. Lu, and Y.-P. Tan, “Deep transfer metric learning,” in Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 325–332.
[12] B. Jiang, Z. Zhang, D. Lin, J. Tang, and B. Luo, “Semi-supervised learning with graph learning-convolutional networks,” in Proc. IEEE/ICVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 11313–11320.
[13] S. Hosseinzadeh Kassani, P. Hosseinzadeh Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, “Classification of histopathological biopsy images using ensemble of deep learning networks,” 2019, arXiv:1909.11870. [Online]. Available: http://arxiv.org/abs/1909.11870
[14] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980
[15] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in Proc. 5th Int. Conf. Learn. Representations (ICLR), Toulon, France, Apr. 2017, pp. 1–14.
[16] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, “Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images,” Comput. Biol. Med., vol. 43, no. 10, pp. 1563–1572, Oct. 2013.
[17] A. Kumar, S. K. Singh, S. Saxena, K. Lakshmanan, A. K. Sangai, H. Chauhan, S. Shrivastava, and R. K. Singh, “Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer,” Inf. Sci., vol. 508, pp. 405–421, Jan. 2020.
[18] N. Kumar and S. P. Awate, “Semi-supervised robust mixture models in RKHS for abnormality detection in medical images,” IEEE Trans. Image Process., vol. 29, pp. 4772–4787, 2020.

[19] S. Li, Q. Yang, H. Jiang, J. A. Cortés-Vecino, and Y. Zhang, “Parasitologist-level classification of apicomplexan parasites and host cell with deep cycle transfer learning (DCTL),” Bioinformatics, to be published, doi: 10.1093/bioinformatics/btaa513.

[20] Y. Li, J. Wu, and Q. Wu, “Classification of breast cancer histology images using multi-scale and discriminative patches based on deep learning,” IEEE Access, vol. 7, pp. 21400–21408, 2019.

[21] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.

[22] C. Lu and M. Mandal, “Automated analysis and diagnosis of skin melanoma on whole slide histopathological images,” Pattern Recognit., vol. 48, no. 8, pp. 2738–2750, Aug. 2015.

[23] Y. Ma, Z. Jiang, H. Zhang, F. Xie, Y. Zheng, H. Shi, and Y. Zhao, “Breast histopathological image retrieval based on latent Dirichlet allocation,” IEEE J. Biomed. Health Inform., vol. 21, no. 4, pp. 1114–1123, Jul. 2017.

[24] L. V. D. Maaten and G. Hinton, “Visualizing data using t-SNE,” J. Mach. Learn. Res., vol. 9, pp. 2579–2605, Nov. 2008.

[25] E. Mercan, S. Aksoy, L. G. Shapiro, D. L. Weaver, T. Brunye, and J. G. Elmore, “Localization of diagnostically relevant regions of interest in whole slide images,” in Proc. 22nd Int. Conf. Pattern Recognit., Aug. 2014, pp. 1179–1184.

[26] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” Pattern Recognit., vol. 29, no. 1, pp. 51–59, Jan. 1996.

[27] H. Ravishankar, P. Sudhakar, R. Venkataratnam, S. Thiruvenkadam, P. Anang, N. Babu, and V. Vaidya, “Understanding the mechanisms of deep transfer learning for medical images,” in Deep Learning and Data Labeling for Medical Applications. Athens, Greece: Springer, 2016, pp. 188–196.

[28] D. Sanchez-Morillo, J. González, M. García-Rojo, and J. Ortega, “Classification of breast cancer histopathological images using KAZE features,” in Proc. Int. Conf. Bioinf. Biomed. Eng. Granada, Spain: Springer, 2018, pp. 276–286.

[29] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mohrara, and R. M. Summers, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1285–1298, May 2016.

[30] S. Y. Shin, S. Lee, I. D. Yun, S. M. Kim, and K. M. Lee, “Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images,” IEEE Trans. Med. Imag., vol. 38, no. 3, pp. 762–774, Mar. 2019.

[31] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A dataset for breast cancer histopathological image classification,” IEEE Trans. Biomed. Eng., vol. 63, no. 7, pp. 1455–1462, Jul. 2016.

[32] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” in Proc. 6th Int. Conf. Learn. Representations (ICLR), Vancouver, BC, Canada, Apte/May 2018, pp. 1–12.

[33] P. Wang, Q. Song, Y. Li, S. Lv, J. Wang, L. Li, and H. Zhang, “Cross-task extreme learning machine for breast cancer image classification with deep convolutional features,” Biomed. Signal Process. Control, vol. 57, Mar. 2020, Art. no. 101789.

[34] Y. Wang, B. Lei, A. Elazab, E.-L. Tan, W. Wang, F. Huang, X. Gong, and T. Wang, “Breast cancer image classification via multi-network features and dual-network orthogonal low-rank learning,” IEEE Access, vol. 8, pp. 27779–27792, 2020.

[35] Y. Xie, J. Zhang, and Y. Xia, “Semi-supervised adversarial model for benign-malignant lung nodule classification on chest CT,” Med. Image Anal., vol. 57, pp. 237–248, Oct. 2019.

[36] Y. Xu, J.-Y. Zhu, E. I.-C. Chang, M. Lai, and Z. Tu, “Weakly supervised histopathology cancer image segmentation and classification,” Med. Image Anal., vol. 18, no. 3, pp. 591–604, Apr. 2014.

[37] Z. Yang, L. Ran, S. Zhang, Y. Xia, and Y. Zhang, “EMS-net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images,” Neurocomputing, vol. 366, pp. 46–53, Nov. 2019.

[38] C. Yu, H. Chen, Y. Li, Y. Peng, J. Li, and F. Yang, “Breast cancer classification in pathological images based on hybrid features,” Multimedia Tools Appl., vol. 78, no. 15, pp. 21325–21345, Aug. 2019.

[39] X. Zhang, H. Dou, T. Ju, J. Xu, and S. Zhang, “Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis,” IEEE J. Biomed. Health Inform., vol. 20, no. 5, pp. 1377–1383, Sep. 2016.

[40] Y. Zheng, Z. Jiang, F. Xie, H. Zhang, Y. Ma, H. Shi, and Y. Zhao, “Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification,” Pattern Recognit., vol. 71, pp. 14–25, Nov. 2017.

[41] Y. Zhou, X. He, L. Huang, L. Liu, F. Zhu, S. Cui, and L. Shao, “Collaborative learning of semi-supervised segmentation and classification for medical images,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 2079–2088.