EAI Endorsed Transactions
on e-Learning

Review Article EAI.EU

A review of research and development of semi-supervised learning strategies for medical image processing

Shengke Yang[1,*]

[1] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, P R China

Abstract

Accurate and robust segmentation of organs or lesions from medical images plays a vital role in many clinical applications such as diagnosis and treatment planning. With the massive increase in labeled data, deep learning has achieved great success in image segmentation. However, for medical images, the acquisition of labeled data is usually expensive because generating accurate annotations requires expertise and time, especially in 3D images. To reduce the cost of labeling, many approaches have been proposed in recent years to develop a high-performance medical image segmentation model to reduce the labeling data. For example, combining user interaction with deep neural networks to interactively perform image segmentation can reduce the labeling effort. Self-supervised learning methods utilize unlabeled data to train the model in a supervised manner, learn the basics and then perform knowledge transfer. Semi-supervised learning frameworks learn directly from a limited amount of labeled data and a large amount of unlabeled data to get high quality segmentation results. Weakly supervised learning approaches learn image segmentation from borders, graffiti, or image-level labels instead of using pixel-level labeling, which reduces the burden of labeling. However, the performance of weakly supervised learning and self-supervised learning is still limited on medical image segmentation tasks, especially on 3D medical images. In addition to this, a small amount of labeled data and a large amount of unlabeled data are more in line with actual clinical scenarios. Therefore, semi-supervised learning strategies become very important in the field of medical image processing.

Keywords: Deep Learning; semi-supervised learning; Medical Image Processing; neural network

Received on 01 January 2024, accepted on 15 January 2024, published on 16 January 2024

doi: 10.4108/eai.4822

*Corresponding author. Email: shengkeyang@home.hpu.edu.cn

1. Introduction

In recent years, various algorithms and methods of machine learning, especially deep learning, have been widely applied in various fields such as computer vision (including object detection, image classification, semantic segmentation, etc.), natural language processing (including text recognition, speech recognition, etc.). Deep learning methods have achieved many milestones in different fields, and these results have improved the state of the art in each field[1]. However, these milestones and technical improvements often require large amounts of labeled data for training. Moreover, most of these methods require that the training data and the test data have the same distribution. In reality, in many application scenarios, large amounts of labeled data are not easily accessible, and it is very expensive and time-consuming to annotate different data for different tasks. This fatal drawback has become a bottleneck limiting the development of supervised learning methods, and makes many models have low generalization ability on different tasks.

Deep learning techniques and deep convolutional networks (CNNs) have been achieving great success for some time in computer-aided therapy including analysis of medical images, medical image segmentation,

detection and diagnosis of medical images, and various other tasks that are included in the broader category of computer vision field[2]. In traditional computer vision tasks, i.e., natural image analysis tasks, the cost of labeling the data is low, so there is no need to worry about the difficulty of obtaining labeled data for such tasks. However, in the medical field, a large amount of labeled data is needed, but these labeled data need to be manually labeled by clinicians with rich expertise and also need the cooperation of radiologists and other experts, which is a time-consuming, labor-intensive, and costly process[3]. The difficulty of obtaining labeled data constrains the establishment of a reliable and robust machine learning model in the field of medical image analysis. On the other hand, there is a large amount of unlabeled medical image data that is not effectively utilized. Therefore, how to make full use of these large amounts of unlabeled medical image data has become a key issue in applying machine learning and deep learning algorithms to the field of medical image processing.

To address the above problems, there are many different training strategies such as weakly supervised learning, unsupervised learning, transfer learning, and semi-supervised learning that will be highlighted in this paper. The bottleneck problem in the medical image field is that labeled data is not easily accessible. Semi-supervised learning[4] and active learning can alleviate this problem to a large extent, so that a large amount of unlabeled data can be fully utilized, which in turn can improve the performance of machine learning deep learning algorithms in the field of medical image processing tasks. Semi-supervised learning is a branch of deep learning, and this approach mainly consider how to utilize a small amount of labeled data and a large amount of unlabeled data to obtain good performance on different tasks[4].

Semi-supervised learning learning strategies as well as active learning strategies are one of the most promising approaches in the field of computer vision for solving specific tasks with small amount of annotated data as well as difficult acquisition[5]. It is expected to alleviate the bottleneck problem of difficulty in acquiring annotated pathology images in the field of medical image processing. The robustness and generalization ability of the model is improved by this method, and a large number of unlabeled pathology images are effectively utilized[6]. It is widely recognized in the academic community that their research on semi-supervised learning began in 1994 and was first proposed by Shahshahani and Landgrebe[7]. Semi-supervised learning strategies have been rapidly developed because of the emergence of domains that train on numerous unlabeled data or have difficulty in acquiring labeled data. In semi-supervised learning includes semi-supervised support vector machines, graph semi-supervised learning methods, and divergence-based methods[8]. And in this paper only the method of semi-supervised learning based on deep neural network is introduced. And this deep neural network based method can be subdivided into three kinds after continuous development in different fields, firstly the first one is the end-to-end semi-supervised learning method; the second one is to train the network model in labeled data, and then train the network by using the deep features obtained from the network; and the last one is to train the network by using unlabeled data, and then fine-tune it by using the labeled data[9]. Among these three methods only the first one is semi-supervised learning of the network itself.

This paper summarizes the application of semi-supervised learning methods in the field of medical image processing[10], and introduces the relevant definitions, research directions, and classifications of semi-supervised learning methods, as well as representative cases of these classifications. In addition, it introduces the loss function used in semi-supervised learning and some methods to prevent overfitting. In addition, this paper also introduces the definition of active learning and its methods. Finally, the challenges of semi-supervised learning methods in the field of medical image processing and the future development direction are introduced.

2. Semi-supervised learning methods

This subsection categorizes semi-supervised learning methods into 6 classes[11]. These 6 classes of methods are consistency constraints; self-training methods; multi-view training; generative adversarial learning; multiple hybrid methods; and graph-based semi-supervised learning methods. In this section, the definitions, concepts, schematic diagrams, as well as the advantages and disadvantages of these 5 classes of methods are presented in detail in turn.

2.1 Consistency constraints

The main idea of the consistency constraint method is to have the same predicted output for similar inputs, which means that for an input[12], the predicted output results obtained should be consistent even if a noise disturbance is added to it. This method improves the generalization ability of the model by adding an unsupervised regularized loss term between the predicted results after the perturbation and the normal predicted results on unlabeled data.

And in the specific practice of consistency constraints for semi-supervised learning, in general the total model loss will consist of a loss function with annotated data and a loss-weighted summation with unannotated data. And for the unlabeled sample data it is necessary to use the consistency constraint method, and the loss of this kind of data is generally measured by using MSE (Mean Square Error) or KL [13]scatter. Some semi-supervised learning methods based on consistency constraints are described next.

2.1.1 The π-model and the temporal combinatorial model

These two types of methods were proposed by S Laine and T Aila in 2016[14], and a schematic of these methods is shown in Figure 1

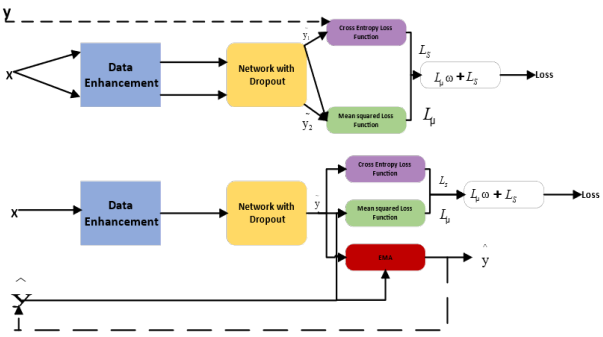


Figure 1. Methodological flow of the π-model and the timing combination model

Both the π-model and the time-series combination model[15] approaches use the weighted sum of the cross-entropy loss of the annotated data and the mean square error loss of the unannotated data samples as the overall loss function. For both methods, the first input to consistency constrained loss for the data with annotation samples is the predicted output obtained fro the random enhancement of the image (random enhancement includes random flipping, ad f noise, random cropping, etc.) fed into the ne ork[16], whereas the second input is different, being a xponenti sliding average of the previous predictio in the combination model[17], and the ne in ge in the π-model which is randomly au ented agai

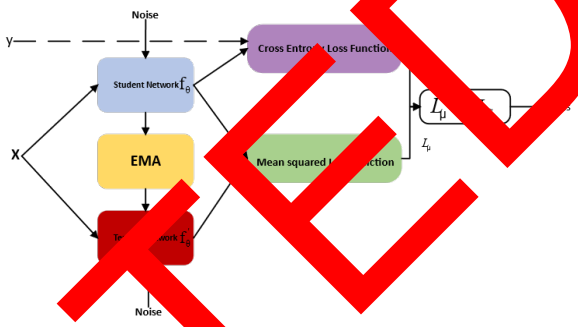
Comparing the tw he π model uires the network to predict th nput age twice, w eas the temporal combination model req s only one prediction in each round of t ning, and it is a more robust.

2.1.2 Mea eacher odule

The me-order combinatorial model also has its inherent dis ge, that is hen it performs sliding ave g, the ning para ers are updated only once i each und, a the work converges slowly. To solve thi problem, inen and Valpola improved on the π-m el and the time-sequence combination model by pro g the Teacher model[18].

In this Mean Teacher semi-supervised learning ap ch there is a student model and a teacher model. We t data into the student model to produce a corresponding output prediction, and then input data into the teacher model to produce an output prediction. The parameters of the student model are used to update the parameters of the teacher model[19]. The results of the student model and the results of the teacher model are then tested for consistency.

Both labeled and unlabeled data are fed into the student model and the teacher model, the student model calculates the cross-entropy loss function of the predictions of the labeled data and the true labels, and the predictions of the teacher model and the student model are used as the mean-square error to calculate the similarity, i.e., consistency detection[20].The flowchart of the methodology of the Mean Teacher mod s sho in Figure 2.



gure 2. Mean acher Model Algorithm Flowchart

1.3 Data Enh cement for Unannotated Data

In semi-s pervised learning methods with co tency co raints, the data enhancement used for image simple random enhancement (e.g., random ipping, random cropping, etc., which are simple random Xie[21] proposed a special stochastic enhancement method called unsupervised data augmentation (UDA), which encompasses both supervised and unsupervised loss components, and uses the cross-entropy loss function to compute the supervised loss component, whereas the unsupervised loss component is computed by using the KL dispersion between the predicted outputs of the annotated samples and the unannotated samples. Numerous experiments have demonstrated that the networks trained using this special unsupervised data augmentation method perform better and are more robust than the ordinary random data augmentation[22]. The algorithmic flow of UDA is shown in Fig. 3

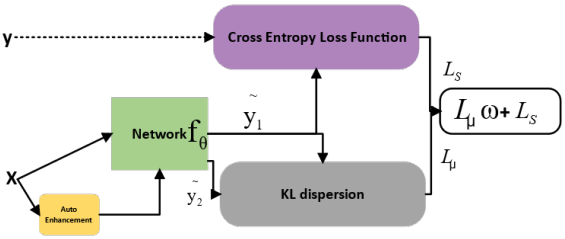


Figure 3. Flowchart of UDA algorithm

2.2 Self-Training

The self-training method is the most basic method in semi-supervised learning. In the self-training approach the model generates pseudo-labels for unlabeled data,

which are then retained with the updated labeled dataset to train the model. Since the predicted pseudo-labels are usually less reliable in the early training stages, their importance in the loss function gradually increases with training[23].

The specific steps are: first divide the annotated sample data into a training set and a test set, and train an initial model using the annotated sample data, second, input the unannotated sample data into the initial model to obtain the prediction results, and select the label with the highest probability among the predicted class labels as the pseudo-label of the unannotated sample data. In the third step, the pseudo-labeled data and labeled data are mixed together, and then the network is retrained using the sample data after mixing[24]. In the fourth step, this trained model is then used to predict the labeled data, using different metrics to evaluate the performance of the classifier. The specific steps are shown in Fig. 4.

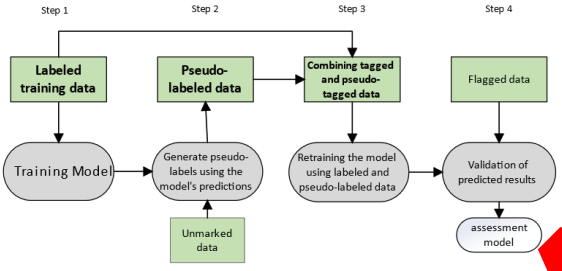


Figure 4. Self-training flowchart

2.3 Multi-view training

Multi-view training This approach assumes that each sample data can be processed from several different perspectives, and then the model trained from the different perspectives is used to process the unannotated sample data, and then the samples with higher confidence and their pseudo-labels are selected to be added to the training set[25]. The purpose of multi-view training is to learn different prediction functions in each perspective to target different specific tasks, and the different perspectives produce complementary predictions, thus improving the generalization ability of the whole model.

2.3.1 Co-Training

Co-training is a semi-supervised learning method based on "divergence"[26], which is designed for multi-view data. Co-training exploits the compatibility and complementarity of multiple views by assuming that the two sample data have two adequate and conditionally independent views, where "adequate" means that each view has enough feature information to generate an optimal method, and "conditionally independent" means that the different views are independent of each other in a given category[27]. The different views are independent of each other.

In the co-training approach there are two models which are trained on the dataset. The specific process is: first in each view with annotated sample data are used to train their corresponding different models[28], and then let each model pick the highest confidence unannotated sample data to give its pseudo-label, and this pseudo-labeled samples are provided to the other model as the new annotated sample data for training and parameter updating, this kind of " This process of "mutual learning and joint promotion" continues in a cycle until the parameters of the two models reach an optimal solution. This process of "mutual learning and joint promotion" continues until the parameters of the two models reach the optimal solution and are no longer updated, or until a preset number of rounds has been reached.

2.3.2 Tri-Training

Tri-training is an improvement of the co-training approach, proposed by Chen et al.[29], which is also based on divergence. The algorithm process can be summarized as follows, first three sub-datasets are sampled from the annotated sample dataset using bootstrapping, and these three sub-datasets are used to train three different basic classifier models. Second, for classifier I, all the unannotated datasets are predicted using the other two classifiers, and the data samples with the same prediction results are selected as the new annotated sample data, which are then added to the training set of classifier model I. Third, perform the second step for all three classifier models and update the parameters for all three classifier models using the new data[30]. Fourth, keep performing the third and fourth steps until the models converge to the optimal solution.The flowchart of the Tri-training algorithm is shown in Fig. 5

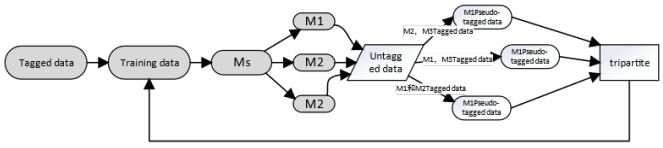


Figure 5. Tri-training's training flowchart

Multi-view training also has its drawbacks, that is, it requires multiple models and the training process is relatively complex with high computational overhead. In particular, the pseudo-labeling in the Tri-training method requires multiple predictions, which undoubtedly incurs greater computational Overhead.

2.4 Generative Adversarial Learning

This approach is mainly inspired by generative adversarial network, which is a generative model proposed by Goodfellow in 2014[31], and this model can be applied to some domains with scarce annotated datasets, which can generate realistic and high resolution images, and according to this approach, image repair, style migration, etc. can be realized. According to its

ability to generate high-resolution images, it can be combined with semi-supervised learning in domains where annotated data is scarce[32]. Thus, the problem of insufficient data with annotation can be alleviated. The following is a brief introduction of the structure of GAN first.

The core of GAN generative adversarial network is that there is a generator and a discriminator, and then the two discriminators play with each other, and finally achieve the data generated by the generator model so that the discriminator cannot be recognized, in order to achieve the effect of the false, at this time, the "fake data" also becomes the real data, which increases the amount of data, but also This increases the amount of data and alleviates the overfitting of the model[33]. The structure is shown in Figure 6 below.

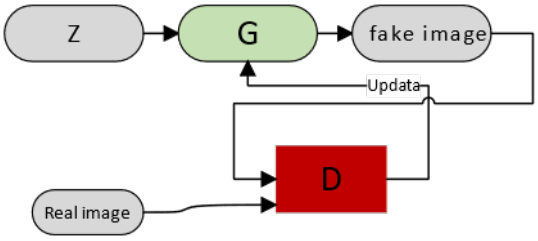


Figure 6. Schematic diagram of GAN

In the figure, z is a random noise, and the real an fake data are fed into the discriminator D a binary neural network training, and the generate can bricate a "fake image" based on a string of ra om numb s. The generated fake images are u to dec e the discriminator D, which is respo sibl for di nguish between real and fake ima and gi a score. For example, if G generates a ure that score ell with D, it proves that G is ver uc sful; if D can ffectively distinguish between r l and fa pictures, then G is not very effective and eeds to adjust parameters. GAN is such a game pr ss.

2.4.1 Se supervi d learning GAN

In s rvise learning methods, the model is constantly ad g the par eters by minimizing the v the c s-entro loss function. Until the parameter are fou h that the value of the cross-entropy unction is inimized. In 2016 Salimans combi d semi-supervised learning[34], and h faced with the K classification problem, the di iminator of the GAN was changed to a classifier of K+1 sses, and the last of the many classes was the generat generated pseudo-data, which was identified as an anomalous class, and the structure of such a model is illustrated in Fig. 7

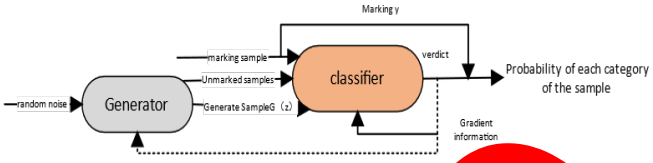


Figure 7. Semi-supervised learni GAN flowchar

The content received by e classi consists o three parts, including label data (x, y unlabel samples x and fake data G generate by the g or. For labeled data classi perform normal sup vised learning. While for u labe dat nd generat d data, the classifier then ne to categ e the sam s judged to be fake data int y last class w h is K class, and the samples jud t e real data i e of the first K classes, b there no need to judge whether this prediction is correct o ot. The loss function for this met sists of a we ted sum of three parts i.e. s ervised loss, loss functio for data judged to be false, d loss function for data judged to be true. The ultimate eal situation is t the generator fools the discriminator much as possib and generates data that is judged to be tr y the disc inator, which also expands the amount of ann a.

Semi-supervised GAN

The semi-supervised GAN proposed by Odena[35] further simplifies the semi-supervised learning GAN proposed by Salimans, and the network structure is shown in Fig. 7. Unlike the network proposed by Salimans, in this network the classifiers of GAN only receive labeled sample data and generated sample data, while the other structures are basically similar. The structure is shown in Figure 8

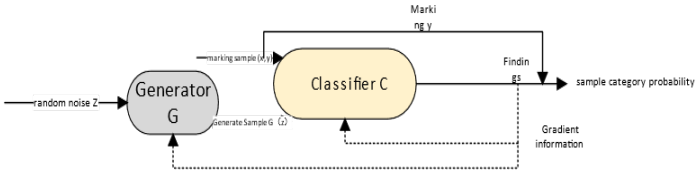


Figure 8. Flowchart of semi-supervised GAN model

However, the disadvantage of this method is also obvious, that is, the generator of the GAN generates random samples, and such random samples often do not have annotation information. In the subsequent use of the process also found that this method is prone to class imbalance problems.

2.5 Multiple mixing methods

This approach attempts to combine several semi-supervised learning methods and ideas, with the aim of drawing on the strengths of the various methods to continuously improve the performance of the modeling approach.

2.5.1 FixMatch

FixMatch is a diverse hybrid semi-supervised method proposed by Google Brain[36], which combines the ideas of consistency constraints and pseudo-labeling correlation in self-training, but simplifies them. For each unannotated sample data, two new sample data were obtained using both weak and strong enhancement. Predictions were obtained by modeling using these two sample data. The consistency regularization is then trained as a one-hot label between the weakly enhanced image and the prediction result of the strongly enhanced image using cross-entropy loss. The FixMatch structure is shown in Figure 9

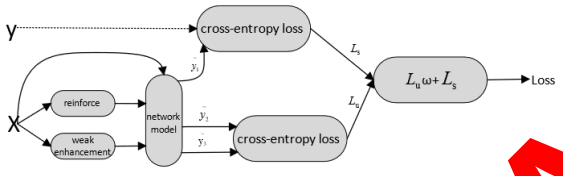


Figure 9. FixMatch model flowchart

2.5.2 MixMatch

MixMatch This approach weighs the sum of supervised and unsupervised losses as an overall loss function. Specifically, Berthelot[37] combines consistency constraints and entropy minimization into an overall loss function, introducing data augmentation for both annotated sample data and unannotated sample data. Each unannotated sample is augmented K times and then the predictions are averaged across the different data augmentations. [38] To reduce the value of entropy, the predicted labels are sharpened and then MixUp[39] regularization is applied between annotated and unannotated samples. The structure of MixMatch is shown in Figure 10

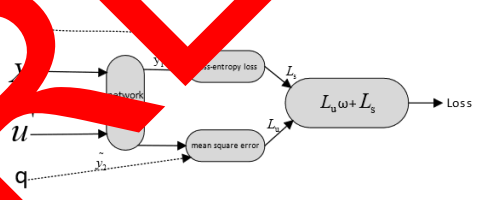


Figure 10. MixMatch Model Flowchart

2.6 Graph-based semi-supervised learning methods

The main idea of graph-based semi-supervised learning [40]is to extract a graph from the original data, where each node represents a training sample and the edges represent the similarity measures of sample pairs. This graph contains annotated samples and unannotated samples with the aim of passing labeled data from labeled nodes to unlabeled nodes.

2.6.1 SDNE

SDNE is based on an autoencoder[41] and the method consists of two parts: an unsupervised part and a supervised part. The first one uses an autoencoder for generating the embedding feature of each node to reconstruct the domain. The second part uses Laplacian for feature mapping to penalize the model when related vertices are far away.

2.6.2 Basic GNN Methods

GNN is a basic classifier that is first trained to obtain the class labels to which the labeled nodes belong, and then the hidden states based on this graph neural network model are applied to the unlabeled nodes. The relevant information is updated by transforming between each pair of nodes using graph neural network.

3. Semi-supervised learning related knowledge skills

In this section, we will introduce some common conceptual knowledge in semi-supervised learning and how it can improve the performance of semi-supervised learning in the field of medical image processing. This includes common loss functions, data enhancement methods, and reasons why combining semi-supervised learning with active learning can improve performance.

3.1 Loss function

3.1.1 Shannon entropy

Shannon entropy represents the uncertainty of the occurrence of an event, the greater the uncertainty of the occurrence of an event represents the greater the Shannon entropy, and the smaller the uncertainty, the smaller the Shannon entropy[42]. Specifically, let X be a discrete random variable with a finite number of values and a probability distribution of $P(X = x_i) = p_{i,i=1,2,\ldots n}$, For this discrete random variable, the total amount of uncertainty in this probability distribution is $H(X) = -\sum_{i=1}^{n} p_i \log p_i$,where H(X) is the Shannon entropy of the random variable X, which depends only on the distribution of X and has no relationship with the specific value of X. Therefore, sometimes the Shannon entropy can be also written as H (P).

3.1.2 Relative Entropy (KL Divergence)

Relative entropy, also known as KL scatter, is used to measure the distance between two separate probability distributions[43]. Let the random variable X have two separate probability distributions P(X) and Q(X), then the KL scatter between the two probability distributions is defined as:

$$D_{KL}\ (P||Q) = \sum_{i=1}^{n} P(x_i) \log \frac{P(x_i)}{Q(x_i)} \quad (1)$$

3.1.3 Cross-entropy loss function

Cross-entropy is a concept in information theory that measures the difference between two probability distributions and is often used in image classification problems. The cross-entropy loss function can usually be minimized to obtain an approximate distribution of target probabilities. It is used extensively in supervised learning. Specifically, the cross-entropy loss function is used to measure the difference between the predicted outcome and the true label.The final cross-entropy loss function is defined as follows:

$$H(P, Q) = -\sum_{i=1}^{n} P(x_i) \log Q(x_i) \quad (2)$$

3.1.4 Dice Loss function

Dice Loss is often used in medical im segmentation in the field of medical image process and in general, Dice Loss is a set similarity met function, which is usually used to calculate the similarit of two sample data (the range of values is [0,1]), and the calculation formula is as follows:

$$s = \frac{2|X \cap Y|}{|X|+|Y|} \quad (3)$$

3.1.5 Group Normalization (GN)

In the field of sem supervised dical image processing, the differenc ween different samples are often large, wh can be achieved using normalization to standardize the d thereby speeding up network conver ce.

At mos batch normalization is d to standardize data, but b n normali ion is affected by the size of the BatchSize, the B nSize is too small, the calculated mean and will b naccurate, and if the B e is too rge, the mputer's video memory will ot be ough. co parison, group normalization calculate the mean variance of each group in the chann tion and has nothing to do with the Bat Size size, so the performance of the model is not a ted by the BatchSize size.

3.2 Data sample enhancement methods

3.2.1 Pseudo-labeling

Also known as pseudo-labeling, pseudo-labeling is commonly used in semi-supervised learning[44]. The main idea of pseudo-labeling is to use annotated data samples to first train the model, and then use the trained model to predict the labels of the unlabeled data samples, and select the label types with the highest confidence as pseudo-labels, and after that the annotated sample data and the newly produced pseudo-labeled data are fed into the network model together to train the network. The specific process is shown in Figure 11

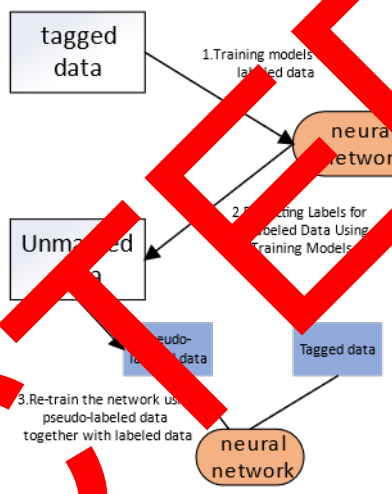


Figure 11. Flowchart of a semi-supervised network for pseudo-label training

3.2 ingle-sample data augmentation

In the field of deep learning, the final performance of a model is highly dependent on the amount of data. Often, the larger the amount of data, the better the model can achieve, but if a model is trained with little data, the model usually does not achieve the desired results. The purpose of data augmentation is to solve the problem that the quantity of annotated data is insufficient and the annotation quality is not high enough, and the main thing that needs to be discussed in this article is data augmentation for annotated data.

The so-called single-sample data augmentation means that when a sample is enhanced, it all revolves around the sample itself. Geometric transformations in single-sample data augmentation include flipping, random rotation; Shift; Random cropping, random scaling; deformation, scaling, etc. Flip and rotate operations are common for tasks that are not directionally sensitive, such as image classification.

The above geometric transformation data enhancement methods do not change the content of the image itself, but only redistribute the pixels. The enhancement method of color transformation in single-sample data augmentation is to change the content of the image itself, and the common methods include adding noise; Obscure; color transformation; Erasure; In general, adding noise to an image is Gaussian noise, and learning ability can be enhanced by adding an appropriate amount of noise. In addition, an important transformation of color transformation is color perturbation, which is to increase

or decrease certain color components in a color space, or to change the order of color channels.

3.2.3 Multi-sample data augmentation

The multi-sample data augmentation method uses multiple sample data to generate new samples, and there are three main methods: SMOTE, SamplePairing, and Mixup.

4. Semi-supervised Learning Methods in Medical Image Processing

Currently there are many tasks related to the field of medical image processing using semi-supervised learning, and there are many sites targeted including heart, brain tumors, and retina segmentation.

4.1 Example of the consistency regularization method to a single task

W Hang[45] proposed a structure-aware average teacher model, where we introduced the entropy minimization principle to the student network so as to adjust itself to produce high-confidence predictions for unlabeled images. Local structural consistency is formulated by encouraging pairwise pixel similarity consistency of the same local predictions between the teacher and student networks. Global structural consistency between the two networks can be promoted by aligning the weighted self-information map. In this way, our model captures structural information from local regions to global regions and minimizes prediction uncertainty for unlabeled images.

While X Li[46] proposed a Transform Consistent Self-Integration Model (TCSM) for semi-supervised medical image segmentation, there are two versions of this network structure. For TCSM-v1, the network optimizes the network by using a weighted combination of supervised loss from annotated data and unsupervised loss with unannotated data. To take advantage of the unannotated data, the network encourages the same inputs to end up with the same predictions in different noise regularization. The method enhances the predictive effect of regularization on the pixel level by introducing a consistency scheme including rotation and flipping in the self-integrating model. Unsupervised loss is designed by minimizing the difference between the network predictions under different transformations of the same input.

4.2 Example of Consistent Regularization Methods for Multitasking

This transform consistency self-integration model proposed by X Li[46] is the same as many previous semi-supervised medical image segmentation tasks that utilize consistency regularization tend to add different noises to the input sample data in the same task for perturbation in the expectation of achieving the same predicted output results thus utilizing a large amount of unannotated data[47]. Consistency based on multi-task level is another approach.

Y Zhang[48] proposed a network architecture for semi-supervised medical image segmentation with two tasks. This network architecture is an integration of two separate segmentation networks based on two tasks which learns region-based shape constraints and boundary-based surface adaptation. The two networks in the overall architecture can learn collaboratively and by collaboratively learning the target probability map and a signed distance map the network can enforce geometric shape constraints to learn more reliable information. The overall architecture of the network consists of two main tasks i.e. segmentation task and also regression task. The segmentation task aims at generating segmentation probability graphs while the regression task aims at generating signed distance graphs. This dual-task network can learn different representations of segmentation goals from different perspectives. The network for each task is mainly guided by the traditional supervised learning loss used for training.

4.3 Examples of multi-view trainingmethods

In addition to methods based on consistent regularization for similar inputs to have the same output, there are examples of applications of self-training methods based on pseudo-labeling in the field of semi-supervised medical image processing.

R Li and D Auer[49] et al. proposed a deep convolutional neural network (DCNN) based on generalized integration for semi-supervised medical image segmentation. The architecture of this network is based on an encoder and decoder structure. The network is initially trained using some annotated data. Immediately after this initial model is copied into sub-models and iteratively improved using a random subset of unlabeled data which contains pseudo-labels generated from the model trained in previous iterations.

4.4 Examples of multi-view trainingmethods

Semi-supervised methods for multi-view learning utilize multiple different views of the data and benefit from the resulting relationships. By minimizing disagreement on multiple different hypotheses. The error in each hypothesis will be minimized.

H Peiris as well as Z Chen[50] et al. proposed the Duo-SegNet net structure.Duo-SegNet proposes a two-view UNet model and equips it with a CriticNet.Each UNet provides a view of the data distribution. And it is trained by minimizing the supervised loss of labeled data and the consistent unsupervised loss of unlabeled data. In

addition CriticNet is mainly used to promote two goals, firstly to ensure that the output of the UNet resembles the true labeling and secondly to identify the plausible part of the predicted results, thus enforcing consistency between the views. In short CriticNet is utilized to normalize the training and identify the parts of the prediction results with high confidence in order to learn from unlabeled data. Namely the network consists of two modules, a dual view segmentation network and a CriticNet.

4.5 Recent advances and methodologies

Recently W. Zhang and L Zhu et al. proposed a recent approach to semi-supervised learning in the field of medical images, which they called BoostMIS[51]. This approach adaptively utilizes the cluster assumption and consistency regularization to exploit the unannotated sample data in semi-supervised learning based on the current learning state, and this strategy can adaptively generate pseudo-labels obtained from the prediction of the task model for better task training. And for the low-confidence unannotated sample data, active learning is introduced to find labeling candidates by utilizing the virtual adversarial perturbation and the density-aware entropy of the model as a query strategy for active learning, and eventually these information-rich labeling candidates are sent to the next training cycle. Adapti pseudo-labeling and information active learn annotations form a closed loop and both collaborate w each other thus enhancing semi-supervised medica images. The network framework contains following four main components: medical image sk; c sistent adaptive label propagator; adversarial stability lector; and balanced uncertainty selector.

In addition, X Zhao and C [52] oposed enhance local feature repres tation in i-supervised medical image segmentati by introducing cross-level comparison learning che into semi pervised learning. Enabling t netwo to explore relational features between lobal and lo representations. In addition to fully utilize the c -level semantic information new consistency constraint is designed to compare prediction of patches with the predictions of the full ima The mantic consistency between the two prediction lev elps to imp e the performance of the

5. E hallenges and future earch directions

e strategy of semi-supervised learning learning can al iate this bottleneck problem in the field of medical image processing with a small number of labeled datasets and high cost of labeling. It is also able to utilize a large amount of unlabeled data to avoid data waste. It should be said that the semi-supervised learning strategy is a good way to alleviate data scarcity, and this method has also made a lot of progress in the field of medical image processing. However, at this stage semi-supervised learning strategies still face the following challenges[53]:

(1) The problem of poor performance using only a single semi-supervised learning method: in current semi-supervised learning methods, most authors seek to use as little annotated data as possible to obtain the best model performance, and thus tend to use one semi supervised learning method such as consistency con ints ir models. Some basic methods are not p erly combi with each other.

(2) Scarcity of labeled data: though semi-supervised learning has reduc the dema for labele data, there is still a certain emand for the ount labeled data, but in some ecific do ns (e.g., ical image processing), the of lab d data is very high. So this becomes a h chal ge limit the formance of semi-supervis earning.

(3) Effe ness in fea ection: Semi-supervised l rning rategies have y similarities with unsupervised learni Semi-supervised algorithms are more sensitive to the se ion of features, if the selected fe es have a lot of ineffe e and bad features, then the t ning results may become very poor.

Future dire ons in the field of semi-supervised rning are predi ted on the consideration of solving the a e three pro ms. There are several trends in the fut develop t of semi-supervised learning.

For enge (1) the future direction is to consider mbining some methods such as entropy minimization, pse abeling, consistency regularization and so on may achieve good experimental results. In addition how to combine many semi-supervised learning methods reasonably rather than just a stack of methods is also a promising research direction in the future.

For challenge (2) consider using an active learning[53] method to label the data that is most meaningful for network performance improvement during training before training using a semi-supervised learning strategy. (Active learning is a method of labeling by actively selecting one of the most valuable samples, with the aim of achieving the best performance of the model using the fewest possible, high quality samples.) In this way the quality of the data is improved while the amount of labeling is reduced. Methods such as these that reduce the need for labeled data by improving the quality of the labeled data also need to be investigated in the future.

For the challenge (3) correct feature selection can reduce the number of features, reduce the dimension and increase the understanding between features and feature values[54]. Currently more commonly used feature selection methods include: discarding features with small changes in value; mutual information and maximum information coefficients; feature selection based on the distance correlation coefficient and other methods. How to reasonably choose the combination of feature selection methods, so as to effectively distinguish the quality of unlabeled samples, enhance the robustness of the method to unlabeled noise samples, and realize the automatic

selection of feature subsets is what semi-supervised learning needs to address in the future stage.

6. Conclusion

This paper introduces the main concepts and methods of semi-supervised learning, as well as the differences between various semi-supervised learning and their respective advantages and disadvantages; various concepts that may be used in semi-supervised learning. It also introduces the application of semi-supervised learning in the field of medical image segmentation. Finally, the challenges faced by semi-supervised learning strategies in the current stage and the future directions are clarified. We hope that through this review, readers can better understand the main methods and classifications of semi-supervised learning, as well as the current status of applications of semi-supervised learning in the field of medical image processing and future research directions.

References

[1] Xu Y, Wang Y. Intelligent recognition technology of heavy metal pollution based on deep learning and fuzzy clustering and its application. First International Meeting for Applied Geoscience & Energy; 20 2021.

[2] Razzak MI, Naz S, Zaib A. Deep Learning Medical Image Processing: Overview, Challenges an Future. 2018.

[3] Hesamian MH, Jia W, He X, K Deep Learning Techniques for Medical Image Segmentation: Achievements a Challenge Journal of Digital Imaging 2019, **32**(8

[4] Van Engelen JE, Hoos HH. sur on semi supervised learning. M ine Lear 2020, **109**(2): 373-440.

[5] Hudelot C, Tam li Y. An Over of Deep Semi-Supervis Learni 2020.

[6] Ren Z, K ng X, Zhang Wang S. UKSSL: Underly Knowledge b Semi-Supervised Learn for Medical Image ssification. IEEE O Journal of gineering in Medicine and Biology 3: 1-8.

[7] Yu C, Yang Liu M, Xia J, Liu S. A survey visual ana s techniques for machine arning. mputati l Visual Media 2021, **7**(1): 3-

[8] se NN, Kul i P. A Survey of Semi-Supervised Methods. IEEE 2008.

[9] Jian-We Yuan L, Xiong-Lin L. Semi-Supervised Learning Methods. Chinese Journal of Computers 2015.

[10] Wang S, Zhang Y. Grad-CAM:Understanding AI Models. □ □ □ □ □ □ □ □ □ □ (□ □) 2023, **76**(8): 1321-1324.

[11] Kulis B, Basu S, Dhillon I, Mooney R. Semi-supervised graph clustering: a kernel approach. Machine Learning 2009, **74**(1): 1-22.

[12] Cirean DC, Meier U, Gambardella LM, Schmidhuber J. Deep, big, simple neural nets for handwritten digit recognition. Neural computation 2010(12): 22.

[13] Pham DL, Xu C, Prince JL. Current methods in medical image segmentation. Annual Review of Biomedical Engineering 2000, **2**(2): 315-337.

[14] Laine S, Aila T. Temporal Ensembling for Semi-Supervised Learning. 2016.

[15] Hendrycks D, Lee K, Mazeika M. U g Pre-T g Can Improve Model Robustnes nd Uncertai 2019.

[16] Mei-Qin H, Jian-Hua S. π-Modu Algebra and π-Module Ideal. Mathemat in Prac and Theor 2015.

[17] Klinker F. Expo ial movi g avera us movingexponenti verage. Mathe sche Semesterberichte 2 **58**(1 7-107.

[18] Tarvainen A valpola an teachers better role models: ight-average nsistency rgets improve semi- ised deep learni resu 2017.

[19] Ma arin Chan LV. Analysis on the Competence Secondary School Mathematics Teachers: Spring rd for a Teacher Training Module. ernational Journ f Research and Innovation in Social Science 2021(6).

[20] O'Connor Michael, Mcgraw, Robert, Killen, Larry, Reich, D is. A computer-based training module for suturing 1] self-directed basic. Medical Teacher 1998.

[21] Xie Q Z, Hovy EH, Luong T, Le Q. Unsupervised Augmentation for Consistency Training. Neural Information Processing Systems; 2020; 2020.

[22] Wang J, Khan MA, Wang S, Zhang Y. SNSVM:SqueezeNet-Guided SVM for Breast Cancer Diagnosis. □ □ □ □ □ □ □ □ □ □ (□ □) 2023, **76**(8): 2201-2216.

[23] Tanha J, Someren MV, De Bakker M, Bouteny W, Shamounbaranesy J, Afsarmanesh H. Multiclass semi-supervised learning for animal behavior recognition from accelerometer data. IEEE International Conference on Tools with Artificial Intelligence; 2013; 2013.

[24] Chen M, Tan X, Zhang L. An iterative self-training support vector machine algorithm in brain-computer interfaces. Intelligent Data Analysis 2016, **20**(1): 67-82.

[25] Tanha J, Someren MV, Afsarmanesh H. Boosting for multiclass semi-supervised learning. Elsevier Science Inc 2014.

[26] Ferguson MBAL, emailprotected, Emailprotected E, Ferguson AL, Andrew L. Ferguson * Andrew L. FergusonPritzker School of Molecular Engineering UoC, Chicago, Illinois , United States*Email: emailprotectedMore by Andrew L. Fergusonhttps://orcid.org/---, a, Lee MBJ, Lee J, Junhee Lee Junhee LeePritzker School of Molecular Engineering UoC, Chicago, Illinois , United StatesMore by Junhee Lee. Permutationally Invariant Networks for Enhanced Sampling (PINES): Discovery of Multimolecular and Solvent-Inclusive Collective Variables.

[27] Tanha J, Van Someren M, Afsarmanesh H. [IEEE 2012 IEEE 12th International Conference on Data Mining (ICDM) - Brussels, Belgium (2012.12.10-2012.12.13)] 2012 IEEE 12th International

Conference on Data Mining - An AdaBoost Algorithm for Multiclass Semi-supervised Learning. 2012**:** 1116-1121.

[28] Dong CDC, Yin YYY, Guo XGX, Yang GYG, Zhou GZG. On Co-Training Style Algorithms. Fourth International Conference on Natural Computation; 2008; 2008.

[29] Chen DD, Wang W, Gao W, Zhou ZH. Tri-net for Semi-Supervised Deep Learning. Twenty-Seventh International Joint Conference on Artificial Intelligence IJCAI-18; 2018; 2018.

[30] Zhou ZH, Li M. Tri-training: exploiting unlabeled data using three classifiers. IEEE Transactions on Knowledge and Data Engineering 2005, **17**(11)**:** 1529-1541.

[31] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative Adversarial Networks. 2014.

[32] Duan X, Lieber CM. Laser-Assisted Catalytic Growth of Single Crystal GaN Nanowires. Journal of the American Chemical Society 2000, **122**(1)**:** 188--189.

[33] Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative Adversarial Networks: An Overview. IEEE Signal Processing Magazine 2017, **35**(1)**:** 53-65.

[34] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved Techniques for Training GANs. 2016.

[35] Odena A. Semi-Supervised Learning with Generative Adversarial Networks. arXiv; 2016.

[36] Sohn K, Berthelot D, Li CL, Zhang Z, Carlini N, Cubuk ED, Kurakin A, Zhang H, Raffel C. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. 2020.

[37] Berthelot D, Carlini N, Goodfellow I, Papernot N, Oliver A, Raffel C. MixMatch: A Holistic Approach to Semi-Supervised Learning. 2019.

[38] Shi Z, Liu L, Liu R. Ho and Pod Hybrid Supervised Sound Event Dete n wi Multi MixMatch and Comp ion Co ce Training. European Signal Pro sing Confere 2021; 2021.

[39] Zhang H, Cisse , Dauphin YN, z-Paz D. mixup: Beyon mpiric Risk Minimizatio 2017.

[40] Wang ZF. Graph-based mi-supervised learning. Artificial e&Robotics 200

[41] Wang , Cui P, Zhu W. Stru al Deep Network Em ding. Acm Sigkdd International Conference on wledge Di very & Data Mining; 2016; 2016.

[42] M atsu Miyake S. Construction of Codes for the Wire annel and th ecret Key Agreement From Correlat Source utputs Based on the Hash operty. I E Transactions on Information Theory 12, **58**(2)**:** 992.

[43] IM. Kullback-Leibler Divergence. Springer Berli lberg 2011.

Lee DH. Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. 2013.

[45] haaban A, Hilton B, Clements K, Dodwell D, Sharma N, Kirwan C, Sawyer E, Maxwell A, Wallis M, Stobart H. The presentation, management and outcome of patients with ductal carcinoma in situ (DCIS) with microinvasion (invasion ≤ 1mm in size)—results from the UK Sloane Project. British Journal of Cancer 2022, **127:** 2125 - 2132.

[46] Xiaomeng_Li, Yu L, Chen H, Heng PA. Semi-supervised Skin Lesion Segmentation via Transformation Consistent Self-ensembling Model. arXiv e-prints 2018.

[47] Li X, Yu L, Chen H, Fu CW, Xing L, Heng PA. Transformation-Consistent Self-Ensembling Model for Semisupervised Medical Image Segmentation. 2021.

[48] B SZA, B JZA, B BTA, C TL, Envel KABP. Multi-modal contrastive mutual lear g and p o-label re-learning for semi-supery d medical im segmentation - ScienceDire Medical Imag Analysis 2022.

[49] Li R, Wagner C, Che X, Auer A Generi Ensemble Based Deep nvolutional N l Netw for Semi-Supervise Medical I age Se n on. 2020 IEEE 1 Internati l Sympos on Biomedical Imagi ISBI): 20; 2020.

[50] Peiris H, C Z, Eg Harandi M do-SegNet: Adversa Dual-Views Semi-Su vised Medical Image entation. 2021.

[51] Zh W, L, Hallinan J, ur A, Zhang S, Cai Q, Ooi BC. E tMIS: Boosting Medical Image Semi-supervised Lea with Adaptive Pseudo Labeling Informative ve Annotation. arXiv e-prints 2022.

[52] Zhao X, Fang C, Fan DJ, Lin X, Gao F, Li G. Cross-level Co stive Learning and Consistency Constraint for Semi pervised Medical Image Segmentation. 2022.

[53] Su JC aji S. The Semi-Supervised iNaturalist-Aves nge at FGVC7 Workshop. 2021.

[54] Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, Liu H. Feature Selection: A Data Perspective. Association for Computing Machinery (ACM) 2017(6).