Characterizing Label Errors: Confident Learning for Noisy-Labeled Image Segmentation

Minqing Zhang[1], Jiantao Gao[1,3], Zhen Lyu[2], Weibing Zhao[1], Qin Wang[1], Weizhen Ding[1], Sheng Wang[4], Zhen Li[1(✉)], and Shuguang Cui[1]

[1] Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, Guangdong, China
lizhen@cuhk.edu.cn
[2] Warshel Institute for Computational Biology, The Chinese University of Hong Kong, Shenzhen, Guangdong, China
[3] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China
[4] Tencent AI Lab, Bellevue, USA

Abstract. Convolutional neural networks (CNNs) have achieved remarkable performance in image processing for its mighty capability to fit huge amount of data. However, if the training data are corrupted by noisy labels, the resulting performance might be deteriorated. In the domain of medical image analysis, this dilemma becomes extremely severe. This is because the medical image annotation always requires medical expertise and clinical experience, which would inevitably introduce subjectivity. In this paper, we design a novel algorithm based on the teacher-student architecture for noisy-labeled medical image segmentation. Creatively, We introduce confident learning (CL) method to identify the corrupted labels and endow CNN an anti-interference ability to the noises. Specifically, the CL technique is introduced to the teacher model to characterize the suspected wrong-labeled pixels. Since the noise identification maps are a little away from sufficient precision, the spatial label smoothing regularization technique is utilized to generate soft-corrected masks for training the student model. Since our method identifies and revises the noisy labels of the training data in a pixel-level rather than simply assigns lower weights to the noisy masks, it outperforms the state-of-the-art method in the noisy-labeled image segmentation task on the JSRT dataset, especially when the training data are severely corrupted by noises.

1 Introduction

Convolutional neural network (CNN)-based methods currently dominate various computer vision tasks, such as face recognition [9] and object detection [10], attesting their capabilities to tackle challenging problems. In the domain of medical image analysis, CNN-based methods also achieved great successes

A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12261, pp. 721–730, 2020.
https://doi.org/10.1007/978-3-030-59710-8_70

in numerous clinical practice, including lung nodule detection [6] and pediatric bone age assessment [15]. These applications brought an intelligent and efficient diagnosis process to reality and the success is usually based on well-performed models.

To obtain a well-trained CNN model, clear-labeled data are crucial. If the labels in the training set are corrupted by noises, the resulting performance might become moderate. To avoid these potential noises, a third party is usually consulted to provide a review for the preliminary annotation and uncover the disputed-labeled data. This review was found to be tedious and inefficient. Still, noise-free labels could not be guaranteed. Besides, unlike the natural image labeling where noises mainly come from random errors, noises of medical image labeling have broad sources. These sources can be separated into two parts. First, high-resolution medical images often require pixel-level manual annotations for segmentation tasks. This onerous labeling process would unavoidably bring in random noises such as missed, wrong or inaccurate annotations. Second, to accelerate the labeling process, medical experts usually cooperate. This cooperation would induce individual subjectivity based on different clinical experience and personal opinions. This subjectivity would also introduce noises.

CNN models are struggling with the noises, meanwhile it is usually time-consuming to obtain the noise-free labels. Researchers have proposed various methods to alleviate the negative effects brought by the noises so that the model performance will not be severely deteriorated even with their presence. Ren et al. [11] uncovered the noisy labels by the gradient and assigned lower weights to the samples with these noisy labels. Goldberger et al. [3] designed an adaptation layer to model the process where the latent true labels were corrupted into the noisy ones. Jiang et al. introduced MentorNet [7] to discover the 'correct-ish' samples and concentrate more on them. For the medical image applications, Xue et al. [14] designed an online uncertainty sample mining method and a re-weighting strategy to eliminate the disturbance from the noisy labels. These researches provide practical solutions to the noisy label problem. However, most studies focus on the classification task because it is the most basic problem in natural image processing domain. However, the medical image domain also requires segmentation tasks to provide the radiologists with pixel-level analysis results. Yet, this task is not comprehensively studied considering the existing difficulties. In the literature, Zhu et al. [16] proposed an end-to-end trainable architecture to automatically evaluate the label quality and, in an image-level, assign lower weights to the noisy labels. This approach did improve the performance of the model trained with noisy labels. Nevertheless, the corrupted areas of each label mask remain unknown. Besides, the image-level re-weighting strategy targets the whole image, while the noises might only corrupt the pixels in a local region. Consequently, the ability of noise identification in this method might be limited.

In this paper, a novel method is proposed to endow the model interpretability and anti-interference ability to the noises existing in the medical image segmentation tasks. Specifically, our algorithm is based on the teacher-student architecture [5]. First, we creatively applied confident learning (CL) technique [8] to

the teacher model to characterize the label noises from the training data. As the pixel-level noise identification maps could only roughly distinguish the corrupted area, the spatial label smoothing regularization (SLSR) technique [1] is utilized as a soft label correction method to train the student model.

To summarize, our research has a dual contribution. First, CL is applied to the segmentation task. The noises identified by CL can help visualize the specific area corrupted by the noisy labels. Second, a soft pixel-level label correction module based on SLSR is designed to train the student model. The algorithm was evaluated on the public JSRT dataset [13]. The result outperforms the *state-of-the-art* method in the noisy-labeled image segmentation tasks. Such anti-interference ability becomes more obvious as the noises increase.

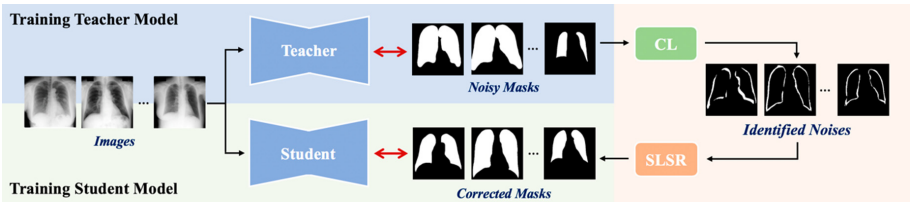

Fig. 1. The pipeline of the teacher-student architecture method. CL module identifies the label noises based on the teacher model. SLSR revises the noises. The student model trained with the soft-corrected masks provides the segmentation results.

2 Method

Figure 1 illustrates the pipeline of our method. The method consists of three parts: a teacher-student architecture containing two segmentation models, CL and SLSR. First, a teacher model is trained with the noisy-labeled data. Then, the CL module will identify the label noises at a pixel-level from the training set. Afterwards, SLSR will correct the noisy labels smoothly. Finally, a student model can be further supervised by the soft-corrected masks.

2.1 Segmentation Network

We adopt a U-Net [12] network with residual blocks [4] as the segmentation network for our framework. The modified U-Net consists of a top-down contracting path and a bottom-up expansive path. The contracting path follows the repeated application of downsampling blocks and residual blocks, while the expansive path consists of repeated upsampling blocks and residual blocks. The downsampling block is a 2×2 convolution with stride 2 and doubles the output channels, and the upsampling block is a 2×2 transposed convolution with stride 2 and reduces the output channels by half. The residual blocks consist of multiple 3×3 convolutions with identity shortcut connections [4]. Each convolution layer in the network is followed by batch normalization and rectified linear unit.

2.2 The Confident Learning Module

Based on the assumption of Angluin [2], CL [8] can identify the label errors in the datasets and improve the training with noisy labels by estimating the joint distribution between the noisy (observed) labels $\tilde{y}$ and the true (latent) labels y^*. Remarkably, no hyper-parameters and few extra computations are required.

Specifically, given a training set $\mathbf{X} = (\mathbf{x}, \tilde{y})^n$ containing n samples $\mathbf{x}$ with noisy label $\tilde{y}$, the predicted probabilities $\hat{p}$ of m classes can be obtained through the teacher model. Assuming that a sample $\mathbf{x}$ labeled $\tilde{y}{=}\, i$ has large enough predicted probabilities $\hat{p}_j(\mathbf{x}) \geq t_j$, it can be suspected that the annotation is wrong and $\mathbf{x}$ might belong to the true latent label $y^* = j$. Here, we select the threshold t_j as the average predicted probabilities $\hat{p}_j(\mathbf{x})$ of all samples labeled $\tilde{y}{=}\, j$:

$$t_j := \frac{1}{|\mathbf{X}_{\tilde{y}=j}|} \sum_{\mathbf{x} \in \mathbf{X}_{\tilde{y}=j}} \hat{p}_j(\mathbf{x}). \tag{1}$$

Based on this assumption, we can construct the confusion matrix $\mathbf{C}_{\tilde{y},y^*}$ by counting the number $\mathbf{C}_{\tilde{y},y^*}[i][j]$ of the samples $\mathbf{x}$ (labeled $\tilde{y}{=}\, i$) which, yet, may belong to the true latent label $y^* = j$:

$$\begin{aligned} \mathbf{C}_{\tilde{y},y^*}[i][j] &:= \left|\hat{\mathbf{X}}_{\tilde{y}=i,y^*=j}\right| \ , where \\ \hat{\mathbf{X}}_{\tilde{y}=i,y^*=j} &:= \{\mathbf{x} \in \mathbf{X}_{\tilde{y}=i} : \hat{p}_j(\mathbf{x}) \geq t_j,\ j = \underset{k \in M: \hat{p}_k(\mathbf{x}) \geq t_k}{\arg\min}\ \hat{p}_k(\mathbf{x})\}. \end{aligned} \tag{2}$$

The required confusion matrix $\mathbf{C}_{\tilde{y},y^*}$ is then normalized, and the joint distribution $\mathbf{Q}_{\tilde{y},y^*}$ between the noisy labels and the true labels could be obtained:

$$\mathbf{Q}_{\tilde{y},y^*}[i][j] = \frac{\frac{\mathbf{C}_{\tilde{y},y^*}[i][j]}{\sum_{b=1}^{m} \mathbf{C}_{\tilde{y},y^*}[i][b]} \cdot |\mathbf{X}_{\tilde{y}=i}|}{\sum_{a,b=1}^{m} \left(\frac{\mathbf{C}_{\tilde{y},y^*}[a][b]}{\sum_{b=1}^{m} \mathbf{C}_{\tilde{y},y^*}[a][b]} \cdot |\mathbf{X}_{\tilde{y}=a}|\right)}. \tag{3}$$

Finally following the confusion matrix $\mathbf{C}_{\tilde{y},y^*}$ or the joint distribution $\mathbf{Q}_{\tilde{y},y^*}$, the wrong-labeled sample set $\tilde{\mathbf{X}}$ can be discovered by the following four options:

Option 1: Confusion Matrix $\mathbf{C}_{\tilde{y},y^*}$. The wrong-labeled samples are selected from the off-diagonals of confusion matrix $\mathbf{C}_{\tilde{y},y^*}$.

Option 2: Joint Distribution $\mathbf{Q}_{\tilde{y},y^*}$. For each class $i \in 1, 2, \ldots m$, the $n \cdot \sum_{j=1, j\neq i}^{m} (\mathbf{Q}_{\tilde{y},y^*}[i][j])$ samples with the lowest self-confidence $\hat{p}_i(\mathbf{x} \in \mathbf{X}_{\tilde{y}=i})$ are selected as the wrong-labeled samples.

Option 3: $\mathbf{C}_{\tilde{y},y^*} \bigcap \mathbf{Q}_{\tilde{y},y^*}$. The element-wise set intersection of the option 1 and 2.

Option 4: $\mathbf{C}_{\tilde{y},y^*} \bigcup \mathbf{Q}_{\tilde{y},y^*}$. The element-wise set union of the option 1 and 2.

2.3 Spatial Label Smoothing Regularization

After identifying the wrong-labeled sample set $\tilde{\mathbf{X}}$ from the training set, traditional methods based on the CL technique clean the data by pruning. However,

the samples of the medical images often undergo the onerous labeling process. Therefore, in our binary segmentation task, rather than directly removing the noisy-labeled data, the SLSR technique is introduced to correct the noisy labels $\tilde{y}$. Specifically, based on the wrong-labeled pixels $\tilde{\mathbf{X}}$ identified by the CL module, SLSR revises the corresponding noisy labels $\tilde{y}$. The corrected labels $\dot{y}$ of noisy samples are defined as:

$$\dot{y}(\mathbf{x}) = \tilde{y}(\mathbf{x}) + \mathbb{1}(\mathbf{x} \in \tilde{\mathbf{X}}) \cdot (-1)^{\tilde{y}} \cdot \epsilon, \tag{4}$$

where $\epsilon \in [0, 1]$ is the hyper-parameter. If $\epsilon = 0$, there will be no modification to the noisy labels, while $\epsilon = 1$ indicates that the noisy labels $\tilde{y}$ will be revised by the output of the CL module directly. Since CL may have uncertainties, ϵ is chosen as 0.8 empirically in our experiments. Based on the training data with the soft-corrected labels, the segmentation network can be further trained by the cross-entropy loss:

$$\mathbf{L} = \sum_{\mathbf{x} \in \mathbf{X}} \log(\hat{p}(\mathbf{x})) \cdot \dot{y}(\mathbf{x}) = \sum_{\mathbf{x} \in \mathbf{X}} \log(\hat{p}(\mathbf{x})) \cdot (\tilde{y}(\mathbf{x}) + \mathbb{1}(\mathbf{x} \in \tilde{\mathbf{X}}) \cdot (-1)^{\tilde{y}} \cdot \epsilon). \tag{5}$$

Fig. 2. An example radiograph with clean mask and three types of corrupted masks.

3 Experiments and Results

3.1 Dataset and Noisy Labels Synthesis

Our method was evaluated on the JSRT [13] dataset containing 247 chest radiographs. Each radiograph has a $2,048 \times 2,048$ resolution and was annotated with three anatomical structures: clavicle, heart, and lung. The data were resized into 256×256 pixels and randomly split into training (197) and validation (50) set.

To better simulate the label noises brought by the manual annotating process, a variety of digital image processing techniques, including dilating, eroding, and edge-distorting, were applied to corrupt the original (clean) masks, which is shown in Fig. 2. To comprehensively investigate our algorithm, we generate several noisy-labeled training sets. For each training set, the proportion and extent of the synthesized noisy-labeled data were various, with their emergence being controlled by two variables, α and β. Specifically, given a noisy-labeled training set with α and β, the data were corrupted by α proportion,

with each mask being dilated, eroded, or edge-distorted by $\beta \pm 3$ pixels. It is noted that edge-distorted was implemented by randomly dilating or eroding the pixels on the boundary with a circle ($radius = \lceil \beta/5 \pm 2 \rceil$ pixel). To generate training sets with several noisy levels, we set α as 0.3, 0.5, and 0.7 respectively, representing three different proportions of noisy-labeled data. Simultaneously, two sets of β, termed A and B (A: $\beta_{clavicle} = 5, \beta_{heart} = 10, \beta_{lung} = 15$; B: $\beta_{clavicle} = 10, \beta_{heart} = 15, \beta_{lung} = 20$), were used to distinguish the size difference for each anatomical structure in the following experiments.

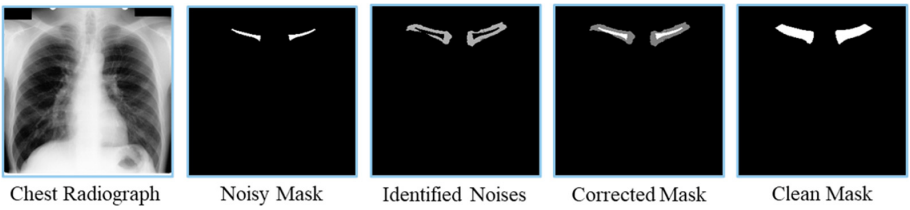

Fig. 3. The noises in the clavicle mask identified by the CL module.

Table 1. Quantitative comparisons among the four CL options on the training set. The best performance is marked in bold.

Noise levels		Recall				Precision				F1-Score			
		C	Q	C∩Q	C∪Q	C	Q	C∩Q	C∪Q	C	Q	C∩Q	C∪Q
$\alpha = 0.3$	$\beta = A$	0.64	**0.83**	0.64	**0.83**	**0.71**	0.64	**0.71**	0.62	0.67	**0.72**	0.67	0.71
	$\beta = B$	0.69	**0.77**	0.68	**0.77**	0.76	0.74	**0.77**	0.73	0.72	**0.75**	0.72	**0.75**
$\alpha = 0.5$	$\beta = A$	0.55	0.70	0.52	**0.73**	0.69	**0.74**	**0.74**	0.70	0.60	**0.72**	0.59	0.71
	$\beta = B$	0.53	**0.64**	0.53	**0.64**	0.67	**0.72**	0.67	0.71	0.59	**0.68**	0.59	0.67
$\alpha = 0.7$	$\beta = A$	0.49	**0.66**	0.49	**0.66**	0.67	**0.70**	0.67	**0.70**	0.56	**0.68**	0.56	0.67
	$\beta = B$	0.52	**0.57**	0.52	**0.57**	0.68	**0.69**	0.68	0.66	0.59	**0.62**	0.59	0.61

3.2 Experimental Settings

The experiments were implemented in PyTorch. Horizontal flipping was applied for data augmentation. All models were trained for 1,000 epochs with Adam optimizer (learning rate = 0.001 and decayed by 0.95 every 50 epochs). The same data split and training settings were applied to all the following experiments.

3.3 Identifying Error from Noisy Labels

To comprehensively investigate the CL module in our algorithm, we test all four CL implementing options on the datasets corrupted by different noise levels. Each option was evaluated by ***Recall***, ***Precision***, and ***F1-Score*** to estimate their abilities to identify the noise. The quantitative comparisons of these

options are shown in Table 1. It can be observed that compared with $\mathbf{C}_{\tilde{y},y^*}$, $\mathbf{Q}_{\tilde{y},y^*}$ achieved higher resulting ***Recall*** and ***Precision*** over almost all noise levels. This means that the option with normalized joint distribution, $\mathbf{Q}_{\tilde{y},y^*}$ can characterize the noises more thoroughly with fewer false positives. By comparison, the noise identification result of $\mathbf{C} \cap \mathbf{Q}$ excluded some true positive noises from $\mathbf{Q}_{\tilde{y},y^*}$, and $\mathbf{C} \cup \mathbf{Q}$ introduced some false positive noises from $\mathbf{C}_{\tilde{y},y^*}$. Therefore, in Sect. 3.4, we chose $\mathbf{Q}_{\tilde{y},y^*}$ to implement the CL module in the experiments for its superior ***F1-Score*** under all noise settings. Figure 3 exhibits the efficacy of noise identification from the CL module.

Table 2. Quantitative comparisons of our methods and the state-of-the-art method pick-and-learn [16] under all noise settings. The best performance is marked in bold.

Noise levels		Methods	Clavicle	Heart	Lung	Average
No noise		Baseline(CE)	0.910	0.939	0.977	0.942
		Baseline + CL	0.920	0.945	0.978	0.948
		Baseline + CL + SLSR	**0.935**	**0.949**	**0.979**	**0.954**
		Pick-and-Learn [16]	**0.935**	0.945	0.978	0.953
$\alpha = 0.3$	$\beta = A$	Baseline(CE)	0.875	0.908	0.948	0.910
		Baseline + CL	0.880	0.925	0.957	0.921
		Baseline + CL + SLSR	0.893	**0.938**	**0.972**	**0.934**
		Pick-and-Learn [16]	**0.894**	0.931	**0.972**	0.932
	$\beta = B$	Baseline(CE)	0.857	0.892	0.938	0.896
		Baseline + CL	0.860	0.921	0.948	0.910
		Baseline + CL + SLSR	**0.890**	**0.932**	**0.968**	**0.930**
		Pick-and-Learn [16]	0.876	0.925	0.967	0.923
$\alpha = 0.5$	$\beta = A$	Baseline(CE)	0.806	0.891	0.898	0.865
		Baseline + CL	0.825	0.908	0.928	0.887
		Baseline + CL + SLSR	**0.856**	**0.927**	**0.966**	**0.916**
		Pick-and-Learn [16]	0.844	0.925	0.965	0.911
	$\beta = B$	Baseline(CE)	0.718	0.861	0.868	0.816
		Baseline + CL	0.730	0.878	0.903	0.837
		Baseline + CL + SLSR	**0.786**	**0.913**	**0.958**	**0.886**
		Pick-and-Learn [16]	0.764	0.911	0.956	0.877
$\alpha = 0.7$	$\beta = A$	Baseline(CE)	0.762	0.826	0.806	0.798
		Baseline + CL	0.774	0.849	0.895	0.839
		Baseline + CL + SLSR	**0.812**	**0.903**	**0.957**	**0.891**
		Pick-and-Learn [16]	0.769	0.896	0.948	0.871
	$\beta = B$	Baseline(CE)	0.614	0.785	0.718	0.706
		Baseline + CL	0.641	0.805	0.831	0.759
		Baseline + CL + SLSR	**0.745**	**0.885**	**0.948**	**0.859**
		Pick-and-Learn [16]	0.630	0.874	0.940	0.815

3.4 Robust Training with Noisy-Labeled Data

In this section, we conducted experiments to perform ablation analysis of our method and compared it with others. All methods were trained on the training set with different noise levels and tested on the clean validation set. In the experiments, ***Dice*** coefficient is introduced to quantify the matching score between predicted and clean masks. Table 2 illustrates the experimental results of the baseline U-Net model without teacher-student architecture trained with the cross-entropy loss, our method (with or without CL module and SLSR), and the leading **Pick-and-Learn** method in noisy-label alleviation. Although the baseline method achieves high ***Dice*** in all three segmentation tasks without noise, as the noisy levels increase, the segmentation performance decreases sharply, especially for a small anatomical structure like clavicle. By contrast, with the aid of the CL module and SLSR, our method can retain high ***Dice*** even with the interference of strong noises. Surprisingly, our method also outperforms the baseline method on the"original" dataset. To explain this, we visualized the label noise identification maps generated by the CL module. As shown in Fig. 4, it can be observed that the "original" dataset is not completely "clean" due to human errors. In our method, these errors can be corrected and thus elevating the ***Dice***. On the whole, our method can help fix the label errors existing in the dataset and thus significantly reduce the labeling requirements. Finally, our method was also compared with the leading **Pick-and-Learn** method in noisy-label alleviation. As our method could ultimately correct the noisy labels rather than simply reducing their weights, it can achieve higher ***Dice*** compared with **Pick-and-Learn**. The qualitative comparisons of the validation set of each method with intensive noises ($\alpha = 0.7, \beta = B$) are shown in Fig. 5.

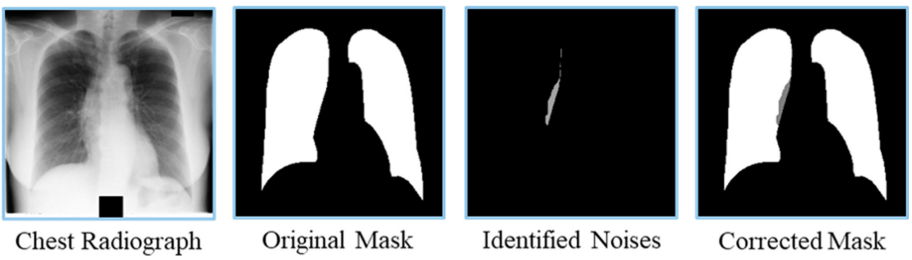

Fig. 4. An example of label error existing in the JSRT dataset. The error is identified by the CL module.

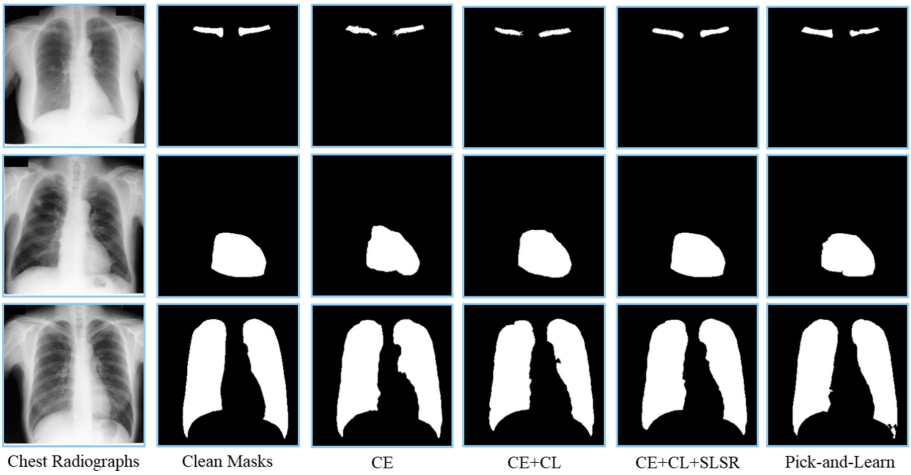

Fig. 5. Qualitative comparisons on the validation set with noise level, $\alpha = 0.7, \beta = B$.

4 Conclusion

In this paper, we propose a two-stage method to address the noisy label issue in the medical image segmentation. The method consists of a teacher-student architecture, a confident learning (CL) module, and a spatial label smoothing regularization (SLSR) technique. This is the first time that CL is involved in the segmentation tasks. This statistic-based technique can identify the label errors in the training set by estimating the joint distribution between the noisy (observed) labels and the true (latent) labels. After recognition of the noises, SLSR will correct the noisy labels smoothly instead of directly pruning these significant data away. The efficacy and necessity of the CL module and SLSR were supported by extensive experiments. Based on the experimental results, our model keeps positive segmentation performance against the increase of the noise levels. The method identifies and corrects the noisy labels in a pixel level instead of directly deleting it. Therefore it outperforms the *state-of-the-art* method in segmentation tasks and could be employed to correct the datasets with noisy labels, especially when treating the intensive noises.

Acknowledgement. The work was supported in part by the Key Area R&D Program of Guangdong Province with grant No. 2018B030338001, by the National Key R&D Program of China with grant No. 2018YFB1800800, by Natural Science Foundation of China with grant NSFC-61629101, by Guangdong Zhujiang Project No. 2017ZT07X152, by Shenzhen Key Lab Fund No. ZDSYS201707251409055, by NSFC-Youth 61902335, by Guangdong Province Basic and Applied Basic Research Fund Project Regional Joint Fund-Key Project 2019B1515120039 and CCF-Tencent Open Fund.

References

1. Ainam, J.P., Qin, K., Liu, G., Luo, G.: Sparse label smoothing regularization for person re-identification. IEEE Access **7**, 27899–27910 (2019)
2. Angluin, D., Laird, P.: Learning from noisy examples. Mach. Learn. **2**(4), 343–370 (1988)
3. Goldberger, J., Ben-Reuven, E.: Training deep neural-networks using a noise adaptation layer (2016)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
5. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
6. Huang, X., Shan, J., Vaidya, V.: Lung nodule detection in CT using 3D convolutional neural networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 379–383. IEEE (2017)
7. Jiang, L., Zhou, Z., Leung, T., Li, L.J., Fei-Fei, L.: Mentornet: learning data-driven curriculum for very deep neural networks on corrupted labels. arXiv preprint arXiv:1712.05055 (2017)
8. Northcutt, C.G., Jiang, L., Chuang, I.L.: Confident learning: estimating uncertainty in dataset labels. arXiv preprint arXiv:1911.00068 (2019)
9. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition (2015)
10. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
11. Ren, M., Zeng, W., Yang, B., Urtasun, R.: Learning to reweight examples for robust deep learning. arXiv preprint arXiv:1803.09050 (2018)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Shiraishi, J., et al.: Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. Am. J. Roentgenol. **174**(1), 71–74 (2000)
14. Xue, C., Dou, Q., Shi, X., Chen, H., Heng, P.A.: Robust learning at noisy labeled medical images: applied to skin lesion classification. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1280–1283. IEEE (2019)
15. Zhang, M., Wu, D., Liu, Q., Li, Q., Zhan, Y., Zhou, X.S.: Multi-Task convolutional neural network for joint bone age assessment and ossification center detection from hand radiograph. In: Suk, H.-I., Liu, M., Yan, P., Lian, C. (eds.) MLMI 2019. LNCS, vol. 11861, pp. 681–689. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32692-0_78
16. Zhu, H., Shi, J., Wu, J.: Pick-and-learn: automatic quality evaluation for noisy-labeled image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 576–584. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_64