Pattern Recognition Letters 165 (2023) 33–38

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Weakly-supervised semantic segmentation via online pseudo-mask correcting

Jiapei Feng[a], Xinggang Wang[a], Te Li[b], Shanshan Ji[b], Wenyu Liu[b,*]

[a] School of EIC, Huazhong University of Science and Technology, China
[b] Zhejiang Lab, China

ARTICLE INFO

Article history:
Received 8 June 2022
Revised 20 October 2022
Accepted 26 November 2022
Available online 28 November 2022

Edited by: Jiwen Lu

Keywords:
Weakly-supervised learning
Semantic segmentation
Noisy label learning

ABSTRACT

Many existing weakly-supervised semantic segmentation methods focus on obtaining more accurate pseudo-masks with weak labels. So far pseudo-masks have come close to the ground truth. However, the potential of these high-quality pseudo-masks has not been fully explored. This is because pseudo-masks inevitably contain partial noisy labels. Deep segmentation networks tend to overfit noisy labels, which leads to poor generalization performance. In this work, we propose a new method to mitigate the damage caused by noisy labels. First, We use the exponential moving average (EMA) model of the online segmentation model as the teacher. Then, predictions from the teacher model are used to correct pseudo-masks online. Besides, learning with noisy labels has been extensively studied in classification tasks. We also introduce these anti-noise techniques and find them also effective for the segmentation task. Our proposed method can be easily embedded into existing weakly-supervised semantic segmentation algorithms and bring 2.3% IoU improvement without expensive computational cost. It also achieves the state-of-the-art performance on the PASCAL VOC 2012 benchmark dataset.

1. Introduction

Semantic segmentation is one of the fundamental computer vision tasks. It aims to assign semantic labels to all pixels of an image. Benefit from the rapid development of deep neural networks (DNNs), segmentation models [1] have made great progress in performance. However, these models rely on large amounts of pixel-level annotated data which calls for expensive cost. Weakly-supervised semantic segmentation (WSSS) aims to learn a segmentation model with weak labels. It relieves the dependence on pixel-level annotated data.

The commonly-used weak labels contain bounding box [2–4], scribble [5], point [6] and image-level labels [7–9]. In this paper, we focus on the weakest one, i.e., image-level label. This weak label does not provide any location or shape information of objects. Thus, most WSSS methods use the Class Activation Map (CAM) [10] from a trained classifier to generate initial pseudo-masks. Besides, many approaches obtain more accurate pseudo-masks through semantic expansion [7,11], erasure [12], adversarial attack [13] and saliency prior [14,15].

However, existing pseudo-masks can already assign accurate labels to most pixels of an image. It is the inevitable noisy labels that seriously damage the generalization performance of segmentation models and prevent the full potential of pseudo-masks from being exploited. Prior works on noisy labels have focused on classification task [16–19]. For segmentation, there are a few works [20,21] designed for medical images. Among them, ADELE [20] further concerns noisy labels in natural images and has provided an effective correction method for them. But it requires massive computing resources and memory footprint. Thus, we want to provide a new strategy that can correct noisy labels with a small computational cost.

ADELE has observed that segmentation models tend to overfit noisy labels in the latter phase of training, shown in Fig. 1. They emphasize the differences of this phenomenon between segmentation and classification: spatial dependence, asynchronous occurrence for all categories, and ubiquitous noise for each image. However, we pay more attention to their commonness. In many cases, segmentation can be regarded as a pixel-level classification task. Therefore, anti-noise techniques from classification methods can be used for segmentation. On the one hand, robust loss functions [17,19] re-weight gradients of all pixels and give lower weights for potential noisy pixels. On the other hand, regularization techniques [16,18] prevent segmentation models from overfitting. Furthermore, we propose a new method to correct noisy labels with

* Corresponding author.
E-mail address: liuwy@hust.edu.cn (W. Liu).

https://doi.org/10.1016/j.patrec.2022.11.024
0167-8655/

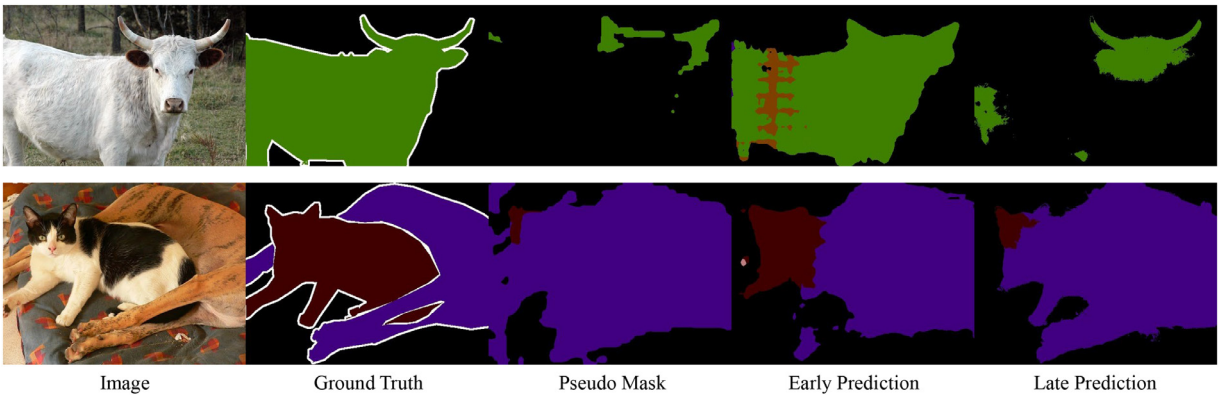

Fig. 1. Visual examples of the overfitting phenomenon. Predictions of the early model (at 3–5 epochs and there are 18 epochs in total in our experiments) are closer to the ground truth, while predictions of the late model are closer to the pseudo-masks.

a low computational cost. Firstly, we build the teacher model by means of the exponential moving average manner. Then, we explore three strategies (*Direct replacement, EMA update*, and *Non-noise sample selection*) to correct noisy and prove the effectiveness through experiments. During the whole online correcting process, our increased computational cost only comes from the generation and prediction of the teacher model.

In summary, the main contributions of this paper are as follows:

1. We introduce robust loss functions and regularization techniques against noisy labels in weakly-supervised semantic segmentation.
2. We explore and propose correcting strategies to reduce the harm of noisy labels, which correct them online without excessive computational costs.
3. Experimental results conducted on PASCAL VOC 2012 dataset proves that the proposed correcting method can be easily embedded into the existing WSSS algorithms and bring 2.3% performance improvement.

2. Related work

2.1. Weakly-supervised semantic segmentation

In this paper, we only consider weakly-supervised semantic segmentation (WSSS) methods with image-level annotations. This kind of WSSS method follows the paradigm that generates the pseudo-masks initially with the CAM [10], expands or updates the pseudo-masks, and then trains segmentation models with pseudo-masks. Since classification networks can only locate the most discriminative regions of objects, the initial pseudo-masks usually contain sparse semantic labels. Several methods use semantic expansion [7,11] and erasure technique [12] to expand initial pseudo-masks. AdvCAM [13] explores more regions of target objects in an anti-adversarial manner. SEAM [8] use the equivariance regularization to make sure that activation maps obtained from multiscale images are equivariant. Others [11] further extend labeled regions with contextual similarity. In addition, many methods take the saliency map as prior to improve the WSSS model. AuxSegNet [14] proposes a multi-task framework, in which classification, segmentation, and salient object segmentation are trained together. Different tasks interact cooperatively through attention maps. EPS [15] uses an off-the-shelf saliency detection model to generate saliency maps as pixel-level annotations. And then, the saliency loss enforces the WSSS model to learn the accurate segmentation knowledge. Different from the above methods, ADELE [20] focuses on improving segmentation models with existing pseudo-masks. After each training epoch, it adaptively corrects pseudo-masks for different categories. In contrast, our method updates pseudo-masks after each iteration, which is easier and more convenient.

2.2. Noisy label learning

Existing works on learning with noisy labels usually focus on classification tasks. And the above overfitting phenomenon is also discovered in this task [22]. In order to mitigate the problem, several methods [17,19] exploit more robust loss functions to improve the generalization performance of classification networks. Others [16,18] use regularization techniques to enforce the network overfitting less explicitly or implicitly. Besides, several methods utilize the multi-network learning [23,24] and the multi-round learning [25] techniques to identify clean samples from training. In this paper, we introduce the robust loss functions and regularization techniques against noise in pseudo-masks.

Label noises are inevitable in semantic segmentation due to subjective judgments from annotators. There are few works focusing on segmentation tasks. VPLR [26] proposes a novel boundary label relaxation technique and makes segmentation models robust for noisy labels of boundary pixels. Both URN [27] and LLU [28] determine the noisy labels through uncertainty estimation and then use the uncertainty map to weight the segmentation loss. BAP [29] uses the distance between features to determine the confidence of labels. Different from them, ADELE [20] and our proposed method try to use the generalization ability of the segmentation model for noisy label correction.

3. Method

3.1. Strategies for correcting labels

Segmentation models can predict accurate labels of pixels in the early stage of training. Our core idea is to correct noisy labels in pseudo-masks with predictions of neural networks. There are two key elements. The first one is when to correct. It remains to be studied when predictions are suitable for correcting. Second, how to correct the pseudo-masks with predictions from segmentation models also needs to be further explored. The framework of our proposed method is shown in Fig. 2.

3.1.1. When to correct

The performance of segmentation models increases rapidly in the first few epochs. We have observed that the performance reaches the inflection point of curves in various categories at the time of 3–5 epochs. This inflection point is the critical moment for correcting noisy labels. At this point, the segmentation model

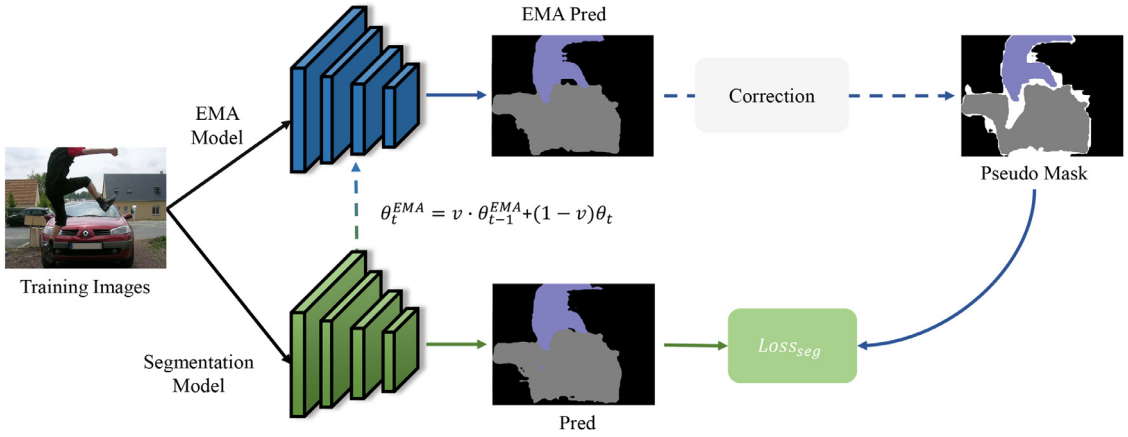

Fig. 2. The framework of our correcting method. Predictions from the EMA model are used to correct pseudo-masks online.

learns knowledge from samples with clean annotations. ADELE [20] adaptively determines the update timing for each category. But We take the unified inflection point as a hyperparameter and determine it through experiments.

3.1.2. How to correct

Inspired by the MixTraining [30] method, we build an exponential moving average (EMA) model of the online segmentation model to generate more robust predictions. The parameters of EMA model are updated by the following formula:

$$\theta_t^{EMA} = \upsilon \cdot \theta_{t-1}^{EMA} + (1-\upsilon) \cdot \theta_t, \tag{1}$$

where θ_t represents the parameters of the online segmentation model at time t and $\upsilon = 0.999$. Predictions from the EMA model are more stable than that from the online model. Thus, they are used as a reference for correcting pseudo-masks. As shown in Fig. 2, training images are fed into the online segmentation model and the EMA model. Before the inflection point, pseudo-masks are used as the supervision to update the online model and the EMA model is updated in the exponential moving average manner. After the inflection point, we use predictions from the EMA model to correct pseudo-masks and the revised pseudo-masks are used to update the online model. We next explore how to properly correct the pseudo-masks with predictions from the EMA model. For this, we designed three strategies as follows:

1. *Direct replacement* In this strategy, we replace pseudo-masks directly with predictions with high confidence.
2. *EMA update* We update pseudo-masks with predictions in the exponential moving average manner: $M_t = \upsilon_m \cdot M_{t-1} + (1-\upsilon_m) \cdot P_t$. M_t presents the pseudo-masks at time t, P_t presents predictions at time t, and υ_m presents the momentum of update.
3. *Non-noise sample selection* Samples whose labels in pseudo-masks are consistent with predictions are more likely to be non-noise samples. These potential non-noise samples are retained, while others are ignored during training.

We have tried both cumulative correction and online correction for the above strategies. The former means that the revised pseudo-masks will be passed to the next correction process, while the latter one is to correct the original pseudo-masks each time.

3.2. Techniques from classification methods

SCE loss Wang et al. [19] have found that the Cross Entropy Loss (CE Loss) is suboptimal loss fuction when labels contain noise. Its class-biased speciality make models overfit for easy classes and underfit for hard classes. To address this issue, they propose a Symmetric Cross Entropy Loss (SCE Loss). Inspired by the symmetric KL-divergence, the SCE Loss is composed of the Cross Entropy Loss and the Reverse Cross Entropy Loss (RCE Loss). The latter one makes models no longer blindly learn noisy labels but learn in a symmetric way. Thus, the SCE Loss is more robust than the CE Loss. And it is extremely simple to use in segmentation tasks. We simply replace the segmentation loss fuction with the following formula:

$$LOSS = -\alpha \cdot \frac{1}{N}\sum_{i}^{N}\sum_{c}^{C} y_i^c \log p_i^c - \beta \cdot \frac{1}{N}\sum_{i}^{N}\sum_{c}^{C} p_i^c \log y_i^c, \tag{2}$$

where N stands for the number of pixels and C stands for the number of classes. And Y and P refer to labels and predictions respectively. α and β are the weights of loss functions. In our experiments, $\alpha = 0.1$ and $\beta = 1.0$.

Noise pruned curriculum loss Lyu et al. [17] confirm the robustness of 0–1 loss to noisy labels. And they propose the Curriculum Loss (CL Loss) as an upper bound of the 0–1 loss. Compared with the 0–1 loss, the Curriculum Loss is differentiable and easy to optimize. Besides, they further propose the Noise Pruned Curriculum Loss to extend the CL Loss for more general scenes. Specifically, the Noise Pruned Curriculum Loss discards the samples with large loss values at a certain rate ϵ. This is based on the assumption that samples with large loss values are more likely to be noisy samples. We set $\epsilon = 0.25$ for better results.

Robust early-learning regularization Xia et al. [18] propose the Robust Early-learning regularization against noise. The method is based on two assumptions. The first one is that deep networks will preferentially learn training data with the clean labels, and then fit samples with noisy labels. (Fig. 1 shows the same phenomenon in segmentation). The second assumption is that only a few key parameters in the deep neural network are important for learning knowledge, while others tend to fit the noise. We believe that these assumptions are also valid for segmentation tasks. Therefore, we follow the Robust Early-learning method [18] and use the g_i as the criterion to determine whether a parameter is a key parameter.

$$g_i = |\nabla L(w_i; S) \times w_i|, i \in [m], w_i \in W, \tag{3}$$

where W represents the set of all parameters, m represents the number of parameters, and $\nabla L(w_i; S)$ represents the gradient of w_i. We sort all parameters according to the g_i and select the top K parameters as the key parameters. During the training stage, the key parameters are updated normally and the non-key parameters were suppressed to zero. We finally set $K = 0.4$ for superior performance.

Label smoothing Label Smoothing is a commonly-used method that can prevent models from overfitting in classification tasks. Thus, Lukasik et al. [16] have proved that Label Smoothing can also mitigate the damage of noisy labels. It works for two reasons. On the one hand, Label Smoothing reduces the confidence of noisy labels. On the other hand, the uniform noise from Label Smoothing competes with the noise in labels. The influence of noisy labels on the models is further weakened. Besides, Label Smoothing is easy to utilize in the segmentation task simply by modifying pseudo-masks. And we have tried several smoothing factors and finally set it to 0.2.

4. Experiments

4.1. Experimental setup

4.1.1. Datasets and metrics

We conduct extensive experiments on the PASCAL VOC 2012 dataset [31], which is the most commonly-used benchmark for weakly-supervised semantic segmentation. This dataset contains 1464 training images, 1449 validation images and 1456 testing images. Following the common practice in WSSS, we use the SBD dataset [32] to augment the PASCAL VOC 2012 dataset. Then the training data grows to 10,582 images.

For evaluation, we adopt the standard mean Intersection over Union (mIoU) as the metric.

4.1.2. Implementation details

Following previous work [11], we adopt ResNet-50 [33], ResNet-101 [33] and ResNet-38 [34] as backbones of segmentation models. For the convenience, we use ResNet-50 for all ablation studies. And we use the pseudo-masks generated by SEAM [8] and AdvCAM [13] for the superiority.

In the training phase, we crop the input image randomly to 448 $\times$ 448 and use the common random augmentation techniques. Then, we employ the stochastic gradient descent (SGD) optimizer with the initial learning rate of 0.001, a momentum of 0.9, and a weight decay of 0.0005. And we use the polynomial learning rate policy to adjust for the decay of learning rate. Specifically, the initial learning rate is multiplied by $(1-(\frac{iter}{iter_{\max}}))^{power}$ after each iteration with $power = 0.9$. The training batchsize is set to 10 and the number of iterations is set to 20K. The correct timing selects the 4 epoch, and the confidence threshold of the *Non-noise sample selection* strategy is set to 0.7. In the inference phase, we adopt multi-scale fusion and dense CRF (dense Conditional Random Field) [35] as post-processing techniques to further improve the performance of segmentation models.

4.2. Comparison with state-of-the-arts

We compare our approach with the state-of-the-art WSSS methods on the validation set and test set of PASCAL VOC 2012, shown in Table 1. For a fair comparison, we report the supervision of all methods. *P* stands for pixel-level labels, *B* stands for bounding box, *S* stands for saliency prior, and *I* stands for image-level labels.

With the pseudo-masks from SEAM [8], our proposed method achieves mIoU values of 66.8 and 67.2 for the PASCAL VOC 2012 validation images and test images. Compared to original SEAM [8], our method obtains 2.3% and 1.5% mIoU improvements respectively. It should be noted that we compare with ADELE [20] which uses the multiscale input augmentation rather than multiscale consistency, called ADELE*. This is to make a clearer comparison of the effectiveness of label correction strategies.

Table 2 sshows the computational costs of ADELE* and our approach (for ResNet-101). From the table, our additional EMA branch brings 1.1$\times$ GPU memory and 1.19$\times$ training time. It should be noted that we ensure training parameters and hardware environment are consistent across these experiments. And the above results also turn out that our approach achieves close performance with ADELE [20], but the manner of online updating requires lower computational costs.

Besides, with the pseudo-masks from AdvCAM [13], our proposed method achieves mIoU values of 70.5 and 71.8 for the PASCAL VOC 2012 validation images and test images. This result is even better than WSSS methods with saliency prior, and achieves the new state-of-the-art performance with image-level labels.

Table 1
Comparison of weakly-supervised semantic segmentation methods on PASCAL VOC 2012 val and test images.

Methods	Sup.	Backbone	Val	Test
DeepLab [1]	*P*	ResNet-101	76.8	76.2
SDI [2]	*B*	ResNet-101	69.4	–
Box2Seg [3]	*B*	ResNet-101	76.4	–
BBAM [4]	*B*	ResNet-101	73.7	73.7
NSRM [36]	*I, S*	ResNet-101	68.3	68.5
AuxSegNet [14]	*I, S*	ResNet-38	69.0	68.6
EPS [15]	*I, S*	ResNet-101	71.0	71.8
IRN [11]	*I*	ResNet-	63.5	64.8
SEAM [8]	*I*	ResNet-38	64.5	65.7
ADELE*$_{\text{SEAM}}$ [20]	*I*	ResNet-38	67.3	–
WOoD*$_{\text{AdvCAM}}$ [37]	*I*	ResNet-101	69.8	69.9
AdvCAM [13]	*I*	ResNet-101	68.1	68.0
PMM [38]	*I*	ResNet-38	68.5	69.0
URN [27]	*I*	ResNet-101	69.5	69.7
Our$_{\text{SEAM}}$	*I*	ResNet-38	66.8	67.2
Our$_{\text{AdvCAM}}$	*I*	ResNet-101	70.5	71.8

Table 2
Comparison with ADELE* on computational costs.

	GPU memory (GB)	Training time (h)
Baseline	20.55	5.25
ADELE*	22.63	11.37
Ours	22.52	6.27

Table 3
Performance improvements of different components on PASCAL VOC 2012 validation set (for ResNet-50 model).

	Cumulative	Online	mIoU
Baseline			62.4
Direct replacement	✓		57.1
Direct replacement		✓	64.4
EMA update		✓	63.4
Non-noise sample selection	✓		58.7
Non-noise sample selection		✓	**65.3**

4.3. Ablation study

4.3.1. Effectiveness of different correcting strategies

We present three strategies for correcting noisy labels and compare cumulative correction with online correction, and the results are shown in Table 3. First, it can be observed that online correction is far superior to cumulative correction. Compared with the baseline model, correcting strategies in a cumulative manner may lead to performance degradation. Through the visualization of the modified pseudo-masks, we find that it is due to the accumulation of incorrect modifications. Besides, the *Non-noise sample selection* strategy outperforms others and bring 2.9% mIoU improvement.

4.3.2. Effectiveness of robust loss functions and regularization techniques

Fig. 3 shows the effects of robust loss functions and regularization techniques on the baseline model. All of them can effectively

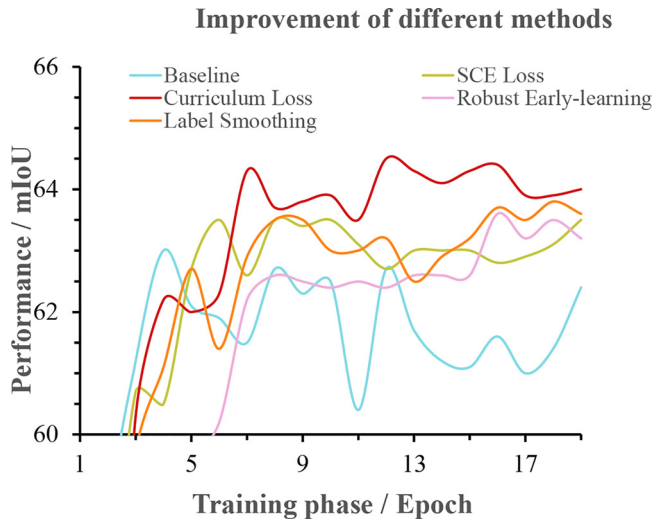

Fig. 3. Performance improvements from robust loss functions and regularization techniques.

improve segmentation models trained with pseudo-masks. Among them, Curriculum Loss is the most effective technique. In addition, it can be observed that the Robust Early-learning technique slows down the convergence speed. Fig. 3 demonstrates that anti-noise techniques from classification methods can also alleviate the overfitting problem in the segmentation task.

4.3.3. Effects of different components

Table 4 further shows the effectiveness of robust loss functions, regularization techniques and our *Non-noise sample selection* strategy (NNSS) respectively. However, these components do not complement each other. The *Non-noise sample selection* strategy can accurately select non-noise samples in pseudo-masks. When used with label smoothing, the latter technique is equivalent to adding smoothing noise. Similarly, Curriculum Loss forces the segmentation model to focus more on simple samples. And non-noise samples are not fully utilized. This also reflects the fact that the non-noise samples that we selected are clean enough.

Table 4
Performance improvements of different components on PASCAL VOC 2012 validation set (for ResNet-50 model).

Label smoothing	CL Loss	NNSS	mIoU
			62.4
✓			63.6
	✓		64.0
		✓	**65.3**
✓		✓	62.8
	✓	✓	63.0

4.3.4. Effectiveness on fully-supervised segmentation network

To our surprise, our method can also improve the performance of fully-supervised segmentation models. The baseline model is set to the fully-supervised DeepLab [1] (ResNet-50) model. With our *Non-noise sample selection* strategy, we improve the baseline from 74.7 mIoU to 75.8 mIoU. This is because the Ground Truth also contains noisy labels, as shown in Fig. 4. Besides, two categories with the most performance improvement are the dining table and sofa. Many dining tables in training images are incorrectly labeled as background. And in some cases, sofa and chair are easily confused. Our method can effectively mitigate these problems and help segmentation models learn better. Two samples of the PASCAL VOC 2012 test set are visualized in the Fig. 4.

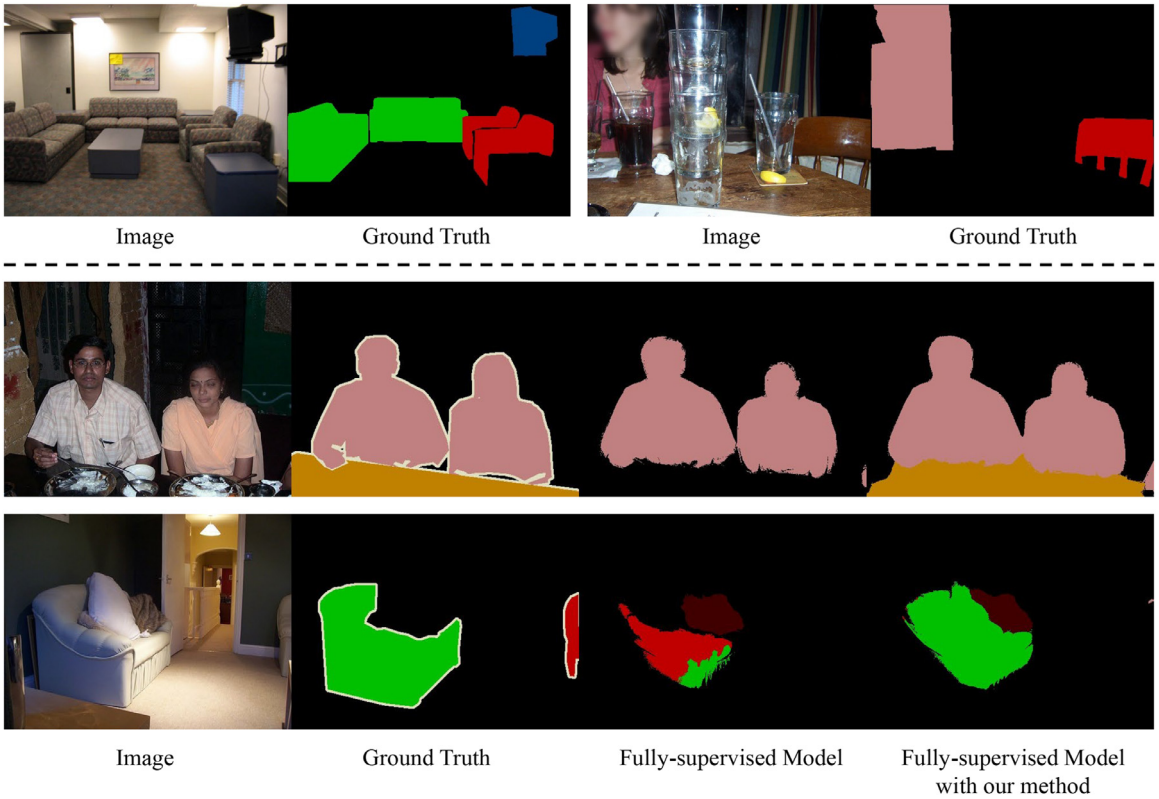

Fig. 4. The top row gives examples of class confusion and labels missing problems. The following two rows show that our approach can mitigate the damage of these problems to segmentation models.

5. Conclusion

In this paper, we propose a new method to correct noisy labels online for weakly-supervised semantic segmentation. The approach takes predictions from the teacher model as guidance and corrects pseudo-masks at each iteration. Besides, we introduce anti-noise techniques from classification methods to mitigate the same overfitting problem of segmentation models. The proposed method improve the performance of existing weakly-supervised semantic segmentation models by 2.3% IoU with less computational cost. Furthermore, we find it can also work for fully-supervised segmentation. In the future, we will explore more effective label correcting methods for fully-supervised segmentation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets used in this paper can be publicaly downloaded from the dataset websites.

Acknowledgment

This work was sponsored by Zhejiang Lab (No. 2019NBOABO2) and NSFC (No. 61876212).

References

[1] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2017) 834–848.
[2] A. Khoreva, R. Benenson, J. Hosang, M. Hein, B. Schiele, Simple does it: weakly supervised instance and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 876–885.
[3] V. Kulharia, S. Chandra, A. Agrawal, P. Torr, A. Tyagi, Box2Seg: attention weighted loss and discriminative feature learning for weakly supervised segmentation, in: European Conference on Computer Vision, Springer, 2020, pp. 290–308.
[4] J. Lee, J. Yi, C. Shin, S. Yoon, BBAM: bounding box attribution map for weakly supervised semantic and instance segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2643–2652.
[5] D. Lin, J. Dai, J. Jia, K. He, J. Sun, Scribblesup: scribble-supervised convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3159–3167.
[6] A. Bearman, O. Russakovsky, V. Ferrari, L. Fei-Fei, What's the point: semantic segmentation with point supervision, in: European Conference on Computer Vision, Springer, 2016, pp. 549–565.
[7] Z. Huang, X. Wang, J. Wang, W. Liu, J. Wang, Weakly-supervised semantic segmentation network with deep seeded region growing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7014–7023.
[8] Y. Wang, J. Zhang, M. Kan, S. Shan, X. Chen, Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12275–12284.
[9] D. Zhang, H. Zhang, J. Tang, X.-S. Hua, Q. Sun, Causal intervention for weakly-supervised semantic segmentation, Adv. Neural Inf. Process. Syst. 33 (2020) 655–666.
[10] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.
[11] J. Ahn, S. Cho, S. Kwak, Weakly supervised learning of instance segmentation with inter-pixel relations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2209–2218.
[12] Q. Hou, P. Jiang, Y. Wei, M.-M. Cheng, Self-erasing network for integral object attention, Adv. Neural Inf. Process. Syst. 31 (2018).
[13] J. Lee, E. Kim, S. Yoon, Anti-adversarially manipulated attributions for weakly and semi-supervised semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4071–4080.
[14] L. Xu, W. Ouyang, M. Bennamoun, F. Boussaid, F. Sohel, D. Xu, Leveraging auxiliary tasks with affinity learning for weakly supervised semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 6984–6993.
[15] S. Lee, M. Lee, J. Lee, H. Shim, Railroad is not a train: saliency as pseudo-pixel supervision for weakly supervised semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 5495–5505.
[16] M. Lukasik, S. Bhojanapalli, A. Menon, S. Kumar, Does label smoothing mitigate label noise? in: International Conference on Machine Learning, PMLR, 2020, pp. 6448–6458.
[17] Y. Lyu, I.W. Tsang, Curriculum loss: robust learning and generalization against label corruption, arXiv preprint arXiv:1905.10045 (2019).
[18] X. Xia, T. Liu, B. Han, C. Gong, N. Wang, Z. Ge, Y. Chang, Robust early-learning: hindering the memorization of noisy labels, in: International Conference on Learning Representations, 2020.
[19] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, J. Bailey, Symmetric cross entropy for robust learning with noisy labels, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 322–330.
[20] S. Liu, K. Liu, W. Zhu, Y. Shen, C. Fernandez-Granda, Adaptive early-learning correction for segmentation from noisy annotations, CVPR 2022, 2022.
[21] Y. Shu, X. Wu, W. Li, LVC-Net: medical image segmentation with noisy label based on local visual cues, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 558–566.
[22] S. Liu, J. Niles-Weed, N. Razavian, C. Fernandez-Granda, Early-learning regularization prevents memorization of noisy labels, Adv. Neural Inf. Process. Syst. 33 (2020) 20331–20342.
[23] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, M. Sugiyama, Co-teaching: robust training of deep neural networks with extremely noisy labels, Adv. Neural Inf. Process. Syst. 31 (2018) 8536–8546.
[24] X. Yu, B. Han, J. Yao, G. Niu, I. Tsang, M. Sugiyama, How does disagreement help generalization against label corruption? in: International Conference on Machine Learning, PMLR, 2019, pp. 7164–7173.
[25] J. Huang, L. Qu, R. Jia, B. Zhao, O2U-Net: a simple noisy label detection approach for deep neural networks, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3326–3334.
[26] Y. Zhu, K. Sapra, F.A. Reda, K.J. Shih, S. Newsam, A. Tao, B. Catanzaro, Improving semantic segmentation via video propagation and label relaxation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8856–8865.
[27] Y. Li, Y. Duan, Z. Kuang, Y. Chen, W. Zhang, X. Li, Uncertainty estimation via response scaling for pseudo-mask noise mitigation in weakly-supervised semantic segmentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, 2022, pp. 1447–1455.
[28] R. Neven, D. Neven, B. De Brabandere, M. Proesmans, T. Goedemé, Weakly-supervised semantic segmentation by learning label uncertainty, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1678–1686.
[29] Y. Oh, B. Kim, B. Ham, Background-aware pooling and noise-aware loss for weakly-supervised semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6913–6922.
[30] M. Xu, Z. Zhang, F. Wei, Y. Lin, Y. Cao, S. Lin, H. Hu, X. Bai, Bootstrap your object detector via mixed training, Adv. Neural Inf. Process. Syst. 34 (2021) 11315–11325.
[31] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (VOC) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[32] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, J. Malik, Semantic contours from inverse detectors, in: 2011 International Conference on Computer Vision, IEEE, 2011, pp. 991–998.
[33] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[34] Z. Wu, C. Shen, A. Van Den Hengel, Wider or deeper: revisiting the resnet model for visual recognition, Pattern Recognit. 90 (2019) 119–133.
[35] P. Krähenbühl, V. Koltun, Efficient inference in fully connected CRFs with Gaussian edge potentials, Adv. Neural Inf. Process. Syst. 24 (2011) 109–117.
[36] Y. Yao, T. Chen, G.-S. Xie, C. Zhang, F. Shen, Q. Wu, Z. Tang, J. Zhang, Non-salient region object mining for weakly supervised semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2623–2632.
[37] J. Lee, S.J. Oh, S. Yun, J. Choe, E. Kim, S. Yoon, Weakly supervised semantic segmentation using out-of-distribution data, arXiv preprint arXiv:2203.03860 (2022).
[38] Y. Li, Z. Kuang, L. Liu, Y. Chen, W. Zhang, Pseudo-mask matters in weakly-supervised semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 6964–6973.