Reciprocal Collaboration for Semi-supervised Medical Image Classification

Qingjie Zeng[1], Zilin Lu[1], Yutong Xie[2], Mengkang Lu[1], Xinke Ma[1], and Yong Xia[1,3,4†]

[1] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[2] Australian Institute for Machine Learning, The University of Adelaide, Australia
[3] Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, China
[4] Ningbo Institute of Northwestern Polytechnical University, Ningbo 315048, China
yxia@nwpu.edu.cn

Abstract. To acquire information from unlabeled data, current semi-supervised methods are mainly developed based on the mean-teacher or co-training paradigm, with non-controversial optimization objectives so as to regularize the discrepancy in learning towards consistency. However, these methods suffer from the consensus issue, where the learning process might devolve into vanilla self-training due to identical learning targets. To address this issue, we propose a novel **Re**ciprocal **Co**llaboration model (ReCo) for semi-supervised medical image classification. ReCo is composed of a main network and an auxiliary network, which are constrained by distinct while latently consistent objectives. On labeled data, the main network learns from the ground truth acquiescently, while simultaneously generating auxiliary labels utilized as the supervision for the auxiliary network. Specifically, given a labeled image, the auxiliary label is defined as the category with the second-highest classification score predicted by the main network, thus symbolizing the most likely mistaken classification. Hence, the auxiliary network is specifically designed to discern *which category the image should* ***NOT*** *belong to*. On unlabeled data, cross pseudo supervision is applied using reversed predictions. Furthermore, feature embeddings are purposefully regularized under the guidance of contrary predictions, with the aim of differentiating between categories susceptible to misclassification. We evaluate our approach on two public benchmarks. Our results demonstrate the superiority of ReCo, which consistently outperforms popular competitors and sets a new state of the art.

Keywords: Semi-supervised learning · Medical image classification · Contrary predictions.

1 Introduction

The development of data-driven deep learning models [27,8] has significantly advanced the performance of medical image classification [1,23]. However, their success can largely be attributed to the myriad number of paired image-label data. Given that data annotation within a clinical practice context is time-consuming and often demands expert knowledge, the application of deep learning models remains challenging. The ease of acquiring unlabeled data from clinical sites presents a potential solution. Semi-supervised learning (SSL) [20,2], which effectively utilizes abundant unlabeled data alongside limited labeled data, is increasingly gaining popularity as an alternative solution [22,9,24].

Existing SSL methods primarily fall into three categories. First, **pseudo-labeling** methods (see Fig. 1(a)) elaborately design pseudo-label selection standards to identify credible pseudo-labels from the perspective of probability-based threshold [16] or loss-base estimation [26]. Despite their ability to utilize correct pseudo-labels, these methods heavily rely on the warm-up on labeled data and, hence, are potentially vulnerable to erroneous pseudo-labels with high classification scores, a problem known as confirmation bias or overconfidence issue [10]. Second, **Co-training** methods (see Fig. 1(b)) usually employ weak-to-strong regularization [19] with varying network architectures. Although these methods can capture complementary knowledge by constraining discrepancy [25], they bear the risk of devolving into naive self-training, due to the same training objective [18]. Third, **Self-ensembling** methods (see Fig. 1(c)) predominantly operate within a teacher-student framework, where the parameters of the teacher are updated by the student via exponential moving average [21]. While this approach seems advantageous as it utilizes a stable teacher to supervise the student for unlabeled data learning, the teacher and the student will gradually converge to the same target [15], resulting in a decrease in information gain.

To address these issues, it is necessary to explore the SSL paradigm that can maintain network independence for differential information mining, while also minimizing the impact of erroneous predictions. As we delved into the co-training or self-ensembling framework, we found that there are two networks. Both networks are forced to learn an identical target, *i.e.*, *determining the category to which an image belong*. As a result, these networks tend to produce similar predictions when confronted with unlabeled data. However, as evidenced by [25,18,15], this is not a promising SSL method desires. Rather, disagreement between the networks is the crux of the matter. To this end, we advocate a viable principle, *i.e.*, learning two distinct but implicitly consistent networks (see Fig. 1(d)). One network is trained to recognize the category by default, while the other is regularized conversely by learning *which category the image should* ***NOT*** *belong to*. As a result, this principle preserves network discrepancy and mitigates the error rate from a contrary prediction perspective.

Accordingly, we propose a **Re**ciprocal **Co**llaboration framework (**ReCo**) for semi-supervised medical image classification. ReCo is composed of a main network and an auxiliary network. Given a labeled image, the main network is supervised by the ground truth, while the auxiliary network is constrained jointly

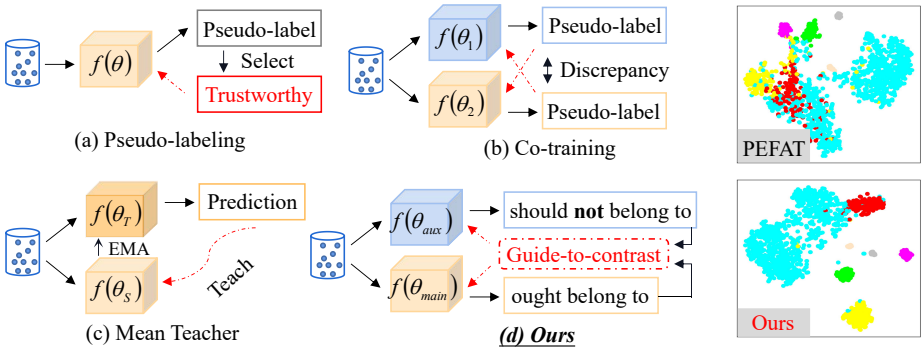

Fig. 1. A brief comparison between (a∼c) current popular SSL schemes and (d) our scheme. The right two t-SNE visualizations are produced by PEFAT [26] and our proposed ReCo, on the ISIC 2018 dataset.

by the ground truth and an auxiliary label. This auxiliary label is derived from the prediction of the main network. Concretely, the classification score corresponding to the auxiliary label is the second highest, thus the auxiliary label can represent the category most susceptible to misclassification. Thus, our goal to facilitate two differentiated but coherent networks is achieved. As for unlabeled data learning, reversed predictions are utilized to perform cross pseudo supervision. To make full use of the information related to which category is likely to be miscalculated, we conduct a feature-level contrast under the guidance of opposing predictions, aiming to intentionally segregate class representations prone to confusion. Based on the reciprocal collaboration, model accuracy is further enhanced due to improved formation of clusters and clearer boundary delineations, as illustrated in the t-SNE visualization on the right side of Fig. 1.

The main contributions are three-fold: (1) we propose a novel perspective of learning opposite targets, bringing the superiority of maintaining network diversity; (2) benefiting from the contrary predictions, categories that are prone to confusion can be distinguished more effectively, with the by-products of improved accuracy and visible decision boundary; and (3) extensive experiments validate the advantage of the proposed ReCo, which outperforms cutting-edge SSL methods on two public benchmarks, setting a new state of the art.

2 Method

2.1 Preliminaries

In semi-supervised learning (SSL), a labeled set $\mathcal{D}_l = \{(x_i^l, y_i^l)\}_{i=1}^{N_l}$ and an unlabeled set $\mathcal{D}_u = \{(x_i^u)\}_{i=1}^{N_u}$ are typically given, where N_l and N_u are the number of images with $N_l \ll N_u$. SSL aims at developing an algorithm that can effectively acquire knowledge from limited labeled data and sufficient unlabeled data.

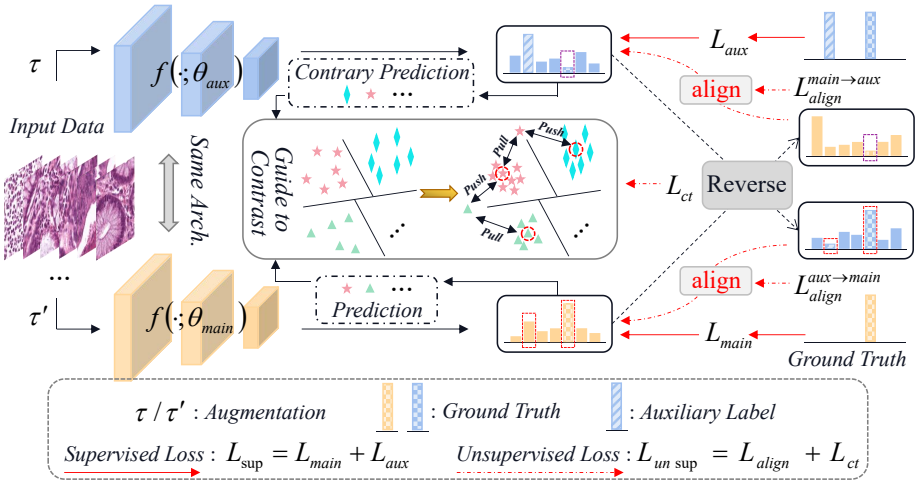

Fig. 2. Workflow of our proposed ReCo. ReCo is composed of two randomly initialized networks with same architecture, *i.e.*, DenseNet-121 [7]. The training objectives of $f(\cdot, \theta_{main})$ and $f(\cdot, \theta_{aux})$ are entirely different. Specifically, for an input image, $f(\cdot, \theta_{main})$ aims for finding out the ground truth, while $f(\cdot, \theta_{aux})$ focuses on which category this image shouldn't belong to. Despite the goals are different, their reversed predictions are implicitly consistent. Thus, unlabeled data can be mined grounded on the contrary predictions from both feature-level and logit-level perspectives.

Fig. 2 presents the diagram of our proposed ReCo. For labeled data learning, the main network $f(\cdot; \theta_{main})$ is supervised by the ground truth, with the aim of learning which category this image ought belong to. Whereas the auxiliary network $f(\cdot; \theta_{aux})$ is taught to predict which category this image shouldn't belong to, under the assistance of generated auxiliary label. As for unlabeled data learning, bi-level constraints are designed according to the contrary predictions. We now delve into details of each component.

2.2 Reciprocal Cooperation Enhances Labeled Data Learning

Current SSL methods are mainly designed based on the self-ensembling mean teacher or co-training paradigms. Despite appeared to be promising, models regularized under these schemes tend to produce similar predictions due to exponentially moving-averaged weights and identical training targets. To maintain the specificity of sub-networks, we propose to train two networks (denoted as $f(\cdot; \theta_{main})$ and $f(\cdot; \theta_{aux})$) with distinct while potentially consistent (if reversed) learning objectives. Specifically, for labeled data learning, the main network $f(\cdot; \theta_{main})$ follows a common supervised learning process, denoted as:

$$\mathcal{L}_{main} = -\frac{1}{N_l} \sum_{i=1}^{N_l} y_i^l \log(f(\hat{y}_{i,main}^l | x_i^l; \theta_{main})), \tag{1}$$

where $\hat{y}^l_{i,main}$ is the predicted label with the highest classification score. Also, we can obtain the category $\overline{y}^l_{i,main}$ that shares the second highest classification score. Here $\overline{y}^l_{i,main}$ can be calculated as:

$$\overline{y}^l_{i,main} = \arg\max(f(x^l_i;\theta_{main})\backslash\left\{\hat{y}^l_{i,main}\right\}), \tag{2}$$

where $\overline{y}^l_{i,main}$ (denoted as auxiliary label) represents the category that has the characteristic of being easily misclassified. To effectively leverage this information and potentially avoid misclassification, we introduce an auxiliary network $f(\cdot;\theta_{aux})$ to predict which category the image shouldn't belong to. To achieve this goal, we constrain $f(\cdot;\theta_{aux})$ using the following formula:

$$\mathcal{L}_{aux} = -\frac{1}{N_l}\sum_{i=1}^{N_l}\left[\underbrace{\overline{y}^l_{i,main}\log(f(\hat{y}^l_{i,aux}|x^l_i;\theta_{aux}))}_{term\ for\ auxiliary\ label} + \underbrace{y^l_i\log(1-f(\widetilde{y}^l_{i,aux}|x^l_i;\theta_{aux}))}_{term\ for\ ground\ truth}\right], \tag{3}$$

$$\hat{y}^l_{i,aux} = \arg\max(f(x^l_i;\theta_{aux})),\quad \widetilde{y}^l_{i,aux} = \arg\min(f(x^l_i;\theta_{aux})), \tag{4}$$

where $\hat{y}^l_{i,aux}$ and $\widetilde{y}^l_{i,aux}$ stand for the predicted classes produced by $f(\cdot;\theta_{aux})$, which have the highest and lowest classification scores, respectively. The first term of $\mathcal{L}_{aux}$ forces $f(\cdot;\theta_{aux})$ to predict the class that shouldn't belong to, and the second term makes model produce the lowest score to the ground truth. Although the training objectives of $f(\cdot;\theta_{main})$ and $f(\cdot;\theta_{aux})$ are completely different, the reversed predictions of these two networks are supposed to be identical. Under such a reciprocal effect, networks can not only alleviate the risk of model training regressing to vanilla self-training, but also make the relation of easy-to-confuse categories distinguishable.

2.3 Contrary Prediction Boosts Unlabeled Data Mining

Based on the learning from labeled data, the labor-division of $f(\cdot;\theta_{main})$ and $f(\cdot;\theta_{aux})$ are clear-cut. Concretely, the former focuses on finding out the ground truth, while the latter pays attention to the targets that tend to be misclassified. Benefiting from the ability of contrary predictions, we further devise bi-level constraints to advance the process of unlabeled data mining.

Logit-level Alignment. Given an unlabeled image x^u_i, we can acquire pseudo-label $\widetilde{y}^u_{i,aux}$ from $f(\cdot;\theta_{aux})$ in a reversed manner, and can also obtain default pseudo-label $\hat{y}^u_{i,main}$ from $f(\cdot;\theta_{main})$. Specifically, $\widetilde{y}^u_{i,aux}$ and $\hat{y}^u_{i,main}$ can be derived from:

$$\widetilde{y}^u_{i,aux} = \arg\max(1-f(x^u_i;\theta_{aux})),\quad \hat{y}^u_{i,main} = \arg\max(f(x^u_i;\theta_{main})). \tag{5}$$

So far, we can use the reversed pseudo-label of $f(\cdot;\theta_{aux})$ to supervised the prediction of $f(\cdot;\theta_{main})$. Conversely, the default pseudo-label produced by $f(\cdot;\theta_{main})$ can also be leveraged to regularize $f(\cdot;\theta_{aux})$, in the way of minimizing the class

probability whose reversed version indicates the potential pseudo-label. And this process can be written as:

$$\mathcal{L}_{align} = -\frac{1}{N_u}\sum_{i=1}^{N_u}\left[\underbrace{\widetilde{y}^u_{i,aux}\log(f(x^u_i;\theta_{main}))}_{\mathcal{L}_{align}^{aux\to main}} + \underbrace{\hat{y}^u_{i,main}\log(1-f(x^u_i;\theta_{aux}))}_{\mathcal{L}_{align}^{main\to aux}}\right], \tag{6}$$

where $\mathcal{L}_{align} = \mathcal{L}_{align}^{aux\to main} + \mathcal{L}_{align}^{main\to aux}$ is the loss calculated on unlabeled data from a predictive perspective.

Feature-level Contrast. In addition to the alignment of reciprocal predictions, the information of which category the image shouldn't belong to can be further employed for feature separation. Specifically, for any input image x^u_i, the predictions of $f(\cdot;\theta_{main})$ and $f(\cdot;\theta_{aux})$ are opposite. Therefore, we can maximize or minimize the feature distance of certain specified categories under the guidance of contrary predictions. This contrastive process can be formulated as:

$$\mathcal{L}_{ct} = \frac{1}{N_u}\sum_{i=1}^{N_u}\left[\underbrace{1-\frac{< z^u_{i,main}\cdot Pro^l_{\hat{y}^u_{i,main}} >}{\|z^u_{i,main}\|_2\cdot\|Pro^l_{\hat{y}^u_{i,main}}\|_2}}_{intra-class\ cluster} + \underbrace{\frac{\|z^u_{i,main}\cdot Pro^l_{\hat{y}^u_{i,aux}}\|_2}{\|z^u_{i,main}\|_2\cdot\|Pro^l_{\hat{y}^u_{i,aux}}\|_2}}_{contrary\ class\ separation}\right], \tag{7}$$

$$\hat{y}^u_{i,aux} = \arg\max(f(x^u_i;\theta_{aux})), \quad Pro^l_{\#} = \frac{1}{N^{\#}_l}\sum_{i=1}^{N^{\#}_l} Norm(f(x^l_{i,\#},\theta_{main})), \tag{8}$$

where $z^u_{i,main}$ is the normalized feature embeddings produced by $f(\cdot;\theta_{main})$ after global average pooling. $\hat{y}^u_{i,aux}$ is the predicted class that x^u_i shouldn't belong to. $Pro^l_{\#}$ is the feature prototype of class-# that calculated on labeled data. $x^l_{i,\#}$ is the i-th labeled data from class-#, and $N^{\#}_l$ is the number of labeled images in class-#. The contrastive loss $\mathcal{L}_{ct}$ is conductive to better clusters and more distinguishable decision boundary, thereby enhancing the accuracy.

2.4 Training and Inference

Training. The objective function of our method contains supervised loss from labeled data and unsupervised loss from unlabeled data, defined as:

$$\theta^t_{main}, \theta^t_{aux} \leftarrow \underset{\theta^{t-1}_{main},\theta^{t-1}_{aux}}{\arg\min} \underbrace{\mathcal{L}_{main}+\mathcal{L}_{aux}}_{\mathcal{L}_{sup}} + \underbrace{\mathcal{L}_{align}+\mathcal{L}_{ct}}_{\mathcal{L}_{unsup}}, \tag{9}$$

where θ^t_{main} and θ^t_{aux} are updated from the $t-1$ step by minimizing overall loss.

Inference. Given an unseen test image, the main network $f(\cdot,\theta_{main})$ is only used to predict the label with the highest classification score.

Table 1. 5-fold cross-validation results (mean±std) on the **NCT-CRC-HE** dataset, when leveraging 100 and 200 labeled data. The best and second best results are shown in **bold** and underline, respectively.

Method	NCT-CRC-HE (200 labeled data)				NCT-CRC-HE (100 labeled data)			
	ACC	SENS	PREC	F1	ACC	SENS	PREC	F1
Baseline	78.26±1.67	77.68±0.97	81.46±1.06	74.55±0.82	70.96±1.53	70.94±1.79	75.78±0.62	72.25±0.96
MT [21]	80.87±0.92	80.36±1.21	82.67±0.83	80.45±0.87	76.28±0.94	75.25±0.70	77.68±1.16	76.92±0.61
FixMatch [19]	82.62±0.57	83.96±0.66	83.18±0.71	83.28±0.97	78.29±1.23	78.75±0.94	80.68±0.46	79.80±0.42
SimPLE [6]	83.95±0.97	84.26±0.77	84.03±0.92	84.72±0.32	80.71±0.81	79.98±1.16	81.78±0.80	81.17±0.72
CoMatch [13]	85.72±0.56	86.06±0.73	87.73±0.25	85.58±0.53	82.86±0.79	83.27±0.81	83.47±0.43	83.92±0.55
RAC-MT [5]	86.06±0.55	86.27±0.51	88.11±0.54	86.29±0.30	82.17±0.98	82.91±0.74	82.50±0.62	83.27±0.81
SimMatch [28]	87.68±0.53	86.78±0.79	88.21±0.28	87.37±0.42	83.16±0.71	83.91±0.73	83.21±0.50	84.01±0.36
PEFAT [26]	89.27±0.58	88.92±0.37	89.76±0.89	89.53±0.66	85.77±0.82	84.98±0.79	85.76±0.47	85.15±0.68
ReCo (Ours)	**91.51±0.47**	**90.96±0.56**	**91.93±0.37**	**91.22±0.38**	**87.56±0.73**	**86.17±0.56**	**87.57±0.32**	**86.72±0.35**

Table 2. 5-fold cross-validation results (mean±std) on the **ISIC 2018** dataset, when leveraging 5% and 20% labeled data. The best and second best results are shown in **bold** and underline, respectively.

Method	ISIC 2018 (20% labeled data)				ISIC2018 (5% labeled data)			
	ACC	SENS	SREC	F1	ACC	SENS	SREC	F1
Baseline	88.36±0.97	67.03±0.87	89.10±0.79	47.87±1.62	83.41±0.89	53.27±0.87	84.41±0.75	40.77±1.98
DS^3L [4]	89.72±0.91	68.78±0.74	90.06±1.28	58.79±0.86	84.72±0.80	56.47±0.79	88.03±0.59	42.32±0.86
FixMatch [19]	90.14±0.55	69.79±0.68	90.21±0.73	57.83±0.47	85.72±0.46	57.75±0.64	88.38±0.86	44.81±0.87
SRC-MT [14]	90.31±0.73	70.36±0.91	90.39±0.87	57.39±0.72	86.72±0.77	60.15±0.97	88.58±0.74	45.15±0.63
CoMatch [13]	90.78±0.62	71.60±0.82	91.02±0.66	60.39±0.76	87.15±0.71	60.67±0.51	89.06±0.77	46.75±0.69
SimMatch [28]	91.16±0.56	72.77±0.63	91.65±0.71	61.80±0.58	88.30±0.82	61.03±0.74	89.52±0.91	47.18±0.80
RAC-MT [5]	91.37±0.71	73.57±0.90	91.55±0.40	62.10±0.73	88.96±0.50	61.92±0.86	89.71±0.82	47.90±0.57
PEFAT [26]	91.96±0.56	74.43±0.72	91.70±0.38	64.83±0.68	89.92±0.74	62.29±0.53	90.02±0.68	48.86±0.82
ReCo (Ours)	**93.25±0.61**	**76.55±0.89**	**93.13±0.19**	**66.07±0.52**	**91.10±0.60**	**64.21±0.74**	**91.31±0.51**	**50.73±0.38**

3 Experiments and Results

3.1 Datasets and Implementation Details

Datasets. We evaluate our method on two public medical image classification datasets, including NCT-CRC-HE dataset [11] and ISIC 2018 dataset [3]. In detail, NCT-CRC-HE contains 100,000 colorectal cancer histology patches with 9 categories. Following [26], 100 and 200 labeled data are respectively used to assess the model performance under an annotation-efficient scenario. ISIC 2018 provides 10,015 skin lesion dermoscopy images, which comprise 7 categories. Following [26,5], 5% and 20% label percentages are considered. For both datasets, we conduct 5-fold cross-validation and report the results using evaluation metrics of Accuracy (ACC), Sensitivity (SENS), Precision (PREC) and F1. The datasets are split into 70%/10%/20% for training/validation/test in each fold.

Implementation Details. For model training, we adopted DenseNet-121 [7] as backbone with resized input size of 224 × 224. This baseline setting was consistent with compared methods [14,5,26]. Our framework was implemented based on Pytorch [17], using four NVIDIA Gefore RTX 3080Ti GPUs. For a mini-batch, 16 labeled and 48 unlabeled images were included. Adam optimizer [12] was employed with an initialized learning rate of 0.001, and the learning rate would be decayed with a power of 0.9 after each epoch. We trained our model

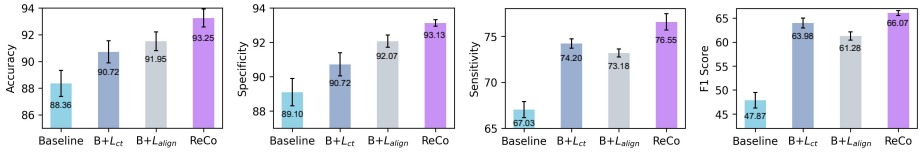

Fig. 3. Ablation study conducted on the ISIC 2018 dataset with 20% labels.

for 60 epochs on both classification tasks. Rotation, affine transformation, flip and cutout were utilized as data augmentation strategies.

3.2 Comparisons and Ablations

We compared our methods with three types of SSL methods, including (1) **Mean teacher-based self-ensembling**: relation-driven mean teacher SRC-MT [14] and reliability-aware contrastive mean teacher RAC-MT [5]. (2) **Consistency-based co-training**: weak-to-strong alignment FixMatch [19], instance-semantics contrast matching SimMatch [28] and memory-smoothed graph regularization CoMatch [13]. (3) **Pseudo-labeling-based**: pair-wise pseudo-label exploration SimPLE [6] and pseudo-loss distribution estimation PEFAT [26].

Results on the NCT-CRC-HE. Table 1 shows the performance comparison on the NCT-CRC-HE dataset. As indicated by the results, we can find: (1) our method consistently outperforms competitors on all metrics, in the scenario of merely providing 100 or 200 annotated data. For instance, compared to the second-best SSL method PEFAT, our proposed ReCo presents 1.79% and 1.57% gains in terms of accuracy and f1 scores, when leveraging 100 labeled data. (2) Compared to RAC-MT, a teacher-student-based cooperation framework with identical learning targets, the proposed strategy of learning distinct while implicitly compatible objectives is more beneficial. This is evidenced by higher accuracy (91.51% vs 86.06%, 5.45%↑) and f1 (91.22% vs 86.29%, 4.93%↑) scores, under the setting of utilizing 200 annotations. And (3) compared to contrast-based SimMatch, the improvements further demonstrate the success of our proposed contrary prediction-guided feature separation.

Results on the ISIC 2018. We also report the performance on the ISIC 2018 dataset. According to Table 2, similar findings can be observed, *e.g.*, ReCo again achieves the first place when compared to cutting-edge SSL methods. Specifically, without relying on carefully designed pseudo-label selection standard, ReCo surpasses PEFAT by 1.29% and 1.18% in term of accuracy, under 20% and 5% label percentages, respectively. This result is mainly attributed to the strategy of separating easy-to-confuse categories, showcasing ReCo's promising capability.

Ablation Study. We perform ablation studies to verify the effects of logit-level alignment $\mathcal{L}_{align}$ and feature-level contrast $\mathcal{L}_{ct}$ included in the ReCo. As Fig. 3 shows, both of $\mathcal{L}_{align}$ and $\mathcal{L}_{ct}$ have a positive impact on the model performance. This indicates the success of contrary predictions for unlabeled data mining. We also analysis the pseudo-label quality on labeled and unlabeled data (presented

in the supplementary), and we can find the performance gap produced by ReCo is much smaller than those produced by other SSL methods.

4 Conclusion

This paper introduces a novel method ReCo for semi-supervised medical image classification. ReCo differs from existing SSL methods as ReCo learns from two distinct while potentially consistent training objectives. Thanks to this operation, model discrepancy is retained and the relation of easy-to-confuse categories are legible, based on the guidance of contrary predictions. Also, sufficient experiments on two datasets validate the effectiveness of ReCo, which consistently achieves the first place on all evaluation metrics.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China under Grants 62171377, in part by Shenzhen Science and Technology Program under Grants JCYJ20220530161616036, and in part by the Ningbo Clinical Research Center for Medical Imaging under Grant 2021L003 (Open Project 2022LYKFZD06).

Disclosure of Interests. The authors have no competing interests.

References

1. Chen, X., Wang, X., Zhang, K., Fung, K.M., Thai, T.C., Moore, K., Mannel, R.S., Liu, H., Zheng, B., Qiu, Y.: Recent advances and clinical applications of deep learning in medical image analysis. Medical Image Analysis **79**, 102444 (2022) 2
2. Chen, Y., Mancini, M., Zhu, X., Akata, Z.: Semi-supervised and unsupervised deep visual learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022) 2
3. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368 (2019) 7
4. Guo, L.Z., Zhang, Z.Y., Jiang, Y., Li, Y.F., Zhou, Z.H.: Safe deep semi-supervised learning for unseen-class unlabeled data. In: ICML. pp. 3897–3906 (2020) 7
5. Hang, W., Huang, Y., Liang, S., Lei, B., Choi, K.S., Qin, J.: Reliability-aware contrastive self-ensembling for semi-supervised medical image classification. In: MICCAI. pp. 754–763. Springer (2022) 7, 8
6. Hu, Z., Yang, Z., Hu, X., Nevatia, R.: Simple: Similar pseudo label exploitation for semi-supervised classification. In: CVPR. pp. 15099–15108 (2021) 7, 8
7. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: CVPR. pp. 4700–4708 (2017) 4, 7
8. Jia, Z., Sun, S., Liu, G., Liu, B.: Mssd: multi-scale self-distillation for object detection. Visual Intelligence **2**(1), 8 (2024) 2
9. Jiao, R., Zhang, Y., Ding, L., Xue, B., Zhang, J., Cai, R., Jin, C.: Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. Computers in Biology and Medicine p. 107840 (2023) 2

10. Jin, Y., Wang, J., Lin, D.: Semi-supervised semantic segmentation via gentle teaching assistant. NeurIPS **35**, 2803–2816 (2022) 2
11. Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.A., Gaiser, T., Marx, A., Valous, N.A., Ferber, D., et al.: Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS Medicine **16**(1), e1002730 (2019) 7
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) 7
13. Li, J., Xiong, C., Hoi, S.C.: Comatch: Semi-supervised learning with contrastive graph regularization. In: ICCV. pp. 9475–9484 (2021) 7, 8
14. Liu, Q., Yu, L., Luo, L., Dou, Q., Heng, P.A.: Semi-supervised medical image classification with relation-driven self-ensembling model. IEEE Transactions on Medical Imaging **39**(11), 3429–3440 (2020) 7, 8
15. Miao, J., Chen, C., Liu, F., Wei, H., Heng, P.A.: Caussl: Causality-inspired semi-supervised learning for medical image segmentation. In: ICCV. pp. 21426–21437 (2023) 2
16. Nguyen, K.B., Yang, J.S.: Boosting semi-supervised learning by bridging high and low-confidence predictions. In: ICCV. pp. 1028–1038 (2023) 2
17. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. NeurIPS **32** (2019) 7
18. Shen, Z., Cao, P., Yang, H., Liu, X., Yang, J., Zaiane, O.R.: Co-training with high-confidence pseudo labels for semi-supervised medical image segmentation. In: IJCAI (2023) 2
19. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. NeurIPS **33**, 596–608 (2020) 2, 7, 8
20. Song, Z., Yang, X., Xu, Z., King, I.: Graph-based semi-supervised learning: A comprehensive review. IEEE Transactions on Neural Networks and Learning Systems (2022) 2
21. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. NeurIPS **30** (2017) 2, 7
22. Weng, Y., Zhang, Y., Wang, W., Dening, T.: Semi-supervised information fusion for medical image analysis: Recent progress and future perspectives. Information Fusion p. 102263 (2024) 2
23. Yuan, L., Liu, X., Yu, J., Li, Y.: A full-set tooth segmentation model based on improved pointnet++. Visual Intelligence **1**(1), 21 (2023) 2
24. Zeng, Q., Xie, Y., Lu, Z., Lu, M., Wu, Y., Xia, Y.: Segment together: A versatile paradigm for semi-supervised medical image segmentation. arXiv preprint arXiv:2311.11686 (2023) 2
25. Zeng, Q., Xie, Y., Lu, Z., Lu, M., Xia, Y.: Discrepancy matters: Learning from inconsistent decoder features for consistent semi-supervised medical image segmentation. arXiv preprint arXiv:2309.14819 (2023) 2
26. Zeng, Q., Xie, Y., Lu, Z., Xia, Y.: Pefat: Boosting semi-supervised medical image classification via pseudo-loss estimation and feature adversarial training. In: CVPR. pp. 15671–15680 (2023) 2, 3, 7, 8
27. Zhao, W., Xu, L.: Weakly supervised target detection based on spatial attention. Visual Intelligence **2**(1), 1–11 (2024) 2

28. Zheng, M., You, S., Huang, L., Wang, F., Qian, C., Xu, C.: Simmatch: Semi-supervised learning with similarity matching. In: CVPR. pp. 14471–14481 (2022) 7, 8