Article

Semi-Supervised Learning for Medical Image Classification Based on Anti-Curriculum Learning

Hao Wu [1], Jun Sun [2,*] and Qi You [1]

1 School of IOT, Jiangnan University, Wuxi 214122, China
2 School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China
* Correspondence: junsun@jiangnan.edu.cn

Abstract: Although deep learning has achieved great success in image classification, large amounts of labelled data are needed to make full use of the advantages of deep learning. However, annotating a large number of images is expensive and time-consuming, especially annotating medical images, which requires professional knowledge. Therefore, semi-supervised learning has shown its potential for medical image classification. In this paper, we propose a novel pseudo-labelling semi-supervised learning method for medical image classification. Firstly, we utilize the anti-curriculum strategy for model training to prevent the model from producing predictions with a high value from the samples which are similar with existed labeled data. Secondly, to produce more stable and accurate pseudo labels for unlabeled data, we generate the pseudo labels with ensemble predictions provided by the model with samples augmented by different augmentations. In addition, we refine the generated pseudo labels using the prediction of the model at the current epoch in order to make the model learn from itself and improve the model performance. Comparative experiments on the Chest X-ray14 dataset for a multi-label classification task and the ISIC 2018 dataset for a multi-class classification task are performed, and the experimental results show the effectiveness of our method.

Keywords: medical image classification; semi-supervised learning; anti-curriculum learning

MSC: 68T07

1. Introduction

Image classification, which is a basic task in computer vision, has achieved fantastic improvements due to the developments in deep learning. In addition, deep learning is also beneficial to cancer classification [1–7]. The key point of the success of deep learning for image classification is the use of a huge amount of high-quality labelled data. However, it is expensive and time-consuming to collect such a hand-crafted labelled dataset. In addition, annotating the medical images requires professional knowledge. Thus, it is difficult to provide a large labelled medical image dataset for training deep neural networks. The limitations on the availability of a large quantity of labelled data promote the development of semi-supervised learning. Recently, semi-supervised learning has shown its potential for medical image classification [8–10].

Currently, semi-supervised learning methods can be approximately classified into two categories. The first category contains the consistency-based methods, which train the model by minimizing the model predictions under small perturbations. Image perturbations, such as adversarial [11] and data augmentations [12,13], are widely used in these methods. In addition, the MeanTeacher method [14] constructs a teacher model in which the parameters are updated by an exponential moving average (EMA) with the training model. Using such a teacher model can be regarded as a kind of model perturbation. As for the consistency-based methods, the design of perturbation functions plays an important role in the effectiveness of these methods.

The other kind of semi-supervised learning methods, i.e., the pseudo-labelling methods, estimate the pseudo labels for unlabeled data and train the model with these data and the existing labelled data. There are two challenges for pseudo-labelling methods. The first one is that the accuracy of the pseudo labels cannot be guaranteed. Therefore, some researchers have attempted to solve this problem. For instance, Lee et al. [15] annotated unlabeled data by using the prediction of the model. Mamshad et al. [16] considered the uncertainty and calibration of the model prediction and proposed a novel framework to improve the accuracy of pseudo labels. Liu et al. [17] combined the model prediction and cluster prediction to produce more reliable pseudo labels.

The second challenge is how to avoid confirmation bias. Since the accuracy of the pseudo labels cannot be guaranteed, training the model with these pseudo-labelled data could make the model accumulate errors and lead to confirmation bias, and thus degrade the model performance. Addressing these issues, some researchers have designed the multi-network architecture [18–21]. However, constructing and training multi-networks need a large memory cost. Hence, there are also some researchers who have employed other strategies to avoid confirmation bias. For example, Sohn et al. [13] just generated pseudo labels for high-confidence samples. Cascante-Bonilla et al. [22] trained the model in several stages and reinitialized the model at each stage.

In this paper, we propose a simple and effective framework for semi-supervised medical image classification. We first employ the anti-curriculum strategy to train the model. Specifically, the model is firstly trained by the existing labelled data and then produces the prediction for each piece of unlabeled data by using unlabeled data as the input of the model. Afterwards, we select the unlabeled data to be pseudo labelled by the prediction. Since the anti-curriculum strategy makes the model learn from hard samples to easy samples, we select the samples with a low prediction value to be pseudo labelled. We divide the training process into several stages. At each stage, we select samples for training by percentage k, which is set as a hyper-parameter.

In addition, to improve the accuracy of pseudo labels, we generate the pseudo labels by averaging the predictions for the data with different data augmentation. As the predictions are produced by the model for the same data, we assume that the predictions for the data with different data augmentation should be as similar as possible. Averaging the predictions can make the pseudo labels more stable and accurate according to the majority rule [12].

According to the anti-curriculum strategy, the pseudo labels are generated by the model of the last training stage. However, in general, the model performance becomes better after training for more epochs unless overfitting is reached. A better model can produce more accurate predictions. Thus, we propose the temporal refinement by linearly combining the pseudo labels with the current model prediction to make the pseudo labels more accurate and further improve the performance of the model.

Finally, we test our method on two different medical image datasets and also conduct ablation studies to investigate the impact of each component. The experimental results show the effectiveness of our method on both multi-label and multi-class datasets.

The paper is structured as follows. Section 2 presents the related work about semi-supervised learning and curriculum learning. In Section 3, the details of our proposed method are described. The experimental results are given and analyzed in Section 4. Finally, the conclusion is drawn in the last section.

2. Related Work

2.1. Semi-Supervised Learning Methods for Image Classification

As the amount of data continues to increase, annotating data becomes an expensive and time-consuming task. Therefore, in real life, the amount of unlabeled data is much greater than that of the labeled data. An effective way to solve this problem is semi-supervised learning. In the past few years, scholars have proposed many semi-supervised algorithms, including generative methods [23–26], graph neural networks based meth-

ods [27–29], consistency-based methods [11–14] and pseudo-labeling methods [15–17]. The current widely used semi-supervised classification methods are consistency-based methods and pseudo-labeling methods. Pseudo-labeling methods utilize the model to generate pseudo labels for unlabeled data and then train the model with these data. Hence, Lee et al. [15] utilized the model prediction to generate the pseudo labels and then trained the model with these pseudo labels. Mamshad et al. [16] generated the pseudo labels for the samples selected by the uncertainty estimation method to improve the accuracy of the pseudo labels, and reinitialized the model at the beginning of each training stage to reduce the accumulation of errors. Consistency-based methods train the model to produce the same predictions by the same samples. For example, Laine et al. [30] perturbed the input with different regularizations and trained the model to produce the same prediction with the same input. Furthermore, the MeanTeacher method [14] improves the model performance with the employment of the EMA model. MixMatch [12] mixes the labeled data and unlabeled data and augments data several times to produce more stable predictions and also trains the model to produce the same prediction with the same data. FixMatch [13] introduces more powerful data augmentation, such as CTAugment [31] and RandAugment [32], to train the model to produce the same predictions with weakly augmented data and strongly augmented data. Recently, several methods [33–35] based on the architecture of FixMatch were proposed to further improve the model performance.

2.2. Semi-Supervised Learning Methods for Medical Image Classification

Since annotating the medical images needs professional knowledge, it is difficult to obtain enough labeled data for deep learning. Thus, semi-supervised learning has become popular for medical image classification. Diaz-Pinto et al. [8] proposed a generative adversarial network (GAN)-based architecture, named Deep Convolutional Generative Adversarial Network (DCGAN), for retinal image synthesis, and trained a semi-supervised learning method based on the architecture of DCGAN. Liu et al. [10] improved the Mean-Teacher method [14] by modeling the feature-level correlation. NoTeacher [36] constructs two student models and trains them with a novel loss based on a probabilistic graphical model. Graph XNet [37] designs a novel graph-based framework for X-ray semi-supervised classification. Liu et al. [38] combined self-supervised pre-training and semi-supervised fine-tuning and achieved great success. ACPL [17] selects the informative unlabeled data to be pseudo labeled and trains the model with pseudo labels.

2.3. Curriculum and Anti-Curriculum Learning

Curriculum learning is first proposed by Bengio et al. [39], the basic idea of which is to make the model learn from easy samples to hard samples. For instance, in image classification, the model is trained by the images which are easy to first be recognized. On the contrary, anti-curriculum learning focuses on hard samples first and then easy samples. Most previous works based on curriculum learning are designed for supervised learning [40,41]. Recently, Cascante-Bonilla et al. [22] proposed a novel pseudo-labeling method based on a curriculum learning strategy and proved the effectiveness of curriculum learning. Liu et al. [17] utilized the anti-curriculum learning strategy for semi-supervised medical image classification and achieved a state-of-the-art result.

3. The Proposed Method

This section shows the details of our proposed method. We firstly trained the model with the existing labeled data. Then, we selected samples to be pseudo labeled and added them into the labeled dataset according to the anti-curriculum learning strategy. Next, we reinitialized the model and trained it with the labeled dataset expanded by the selected samples. Afterwards, we repeated these actions until all the data were added into the labeled dataset. Finally, we trained the model with all the data with pseudo labels. The procedure of our method is described in Algorithm 1. The rest of this section first describes the anti-curriculum learning strategy, then presents the details of the genera-

tion of pseudo labels, and finally introduces the temporal refinement of pseudo labels.


Algorithm 1 Anti-curriculum learning based semi-supervised learning method

---

1   **Input** training model θ, EMA model θ_{ema}, labeled dataset D_L, unlabeled dataset D_U, selected dataset D_S, percentage k, training stage t

2   **Initialize** $D_S = D_L$, $t = 1$, θ, θ_{ema}

3    $\theta \leftarrow$ train with D_S

4    $\theta_{ema} \leftarrow$ update with θ

5   **do**

6       $D_S \leftarrow$ sort $(\theta_{ema}(D_U))$ and select top-k samples to be pseudo labeled by anti-curriculum learning strategy

7       $D_S = D_S \cup D_L$

8   reinitialize θ, θ_{ema}

9       $\theta \leftarrow$ train with D_S

10       $\theta_{ema} \leftarrow$ update with θ

11       $t = t + 1$

12       $k = k * t$

13   **While** $D_S \neq D_L \cup D_U$

14   **end**

3.1. Anti-Curriculum Learning Strategy

Curriculum learning has shown its effectiveness in supervised learning [42]. Cascante-Bonilla et al. [22] further verified its effectiveness in semi-supervised learning. However, according to the idea of curriculum learning, we trained the model by the labeled data first and then chose the samples which were easily recognized to further train the model. Since the difficulty level of samples was determined by the model which was trained by the labeled data, the samples which were similar to the labeled data may be selected. As a result, the model was trained by similar samples and was easy to be overfitted. Thus, we utilized the anti-curriculum learning strategy to train the model and improve the model generalization.

Inspired by the method in [22], we selected samples by percentage, that is, at each training stage, we chose $k * t$ samples to train the model. Here, k was a percentage, which was set to be 20%, and t was the current training stage. Let D_L and D_U mean the labeled dataset and unlabeled dataset, respectively. We firstly trained the model on D_L and then put the data in D_U to the model to obtain the predictions of each unlabeled sample. Next, we sorted the samples by the maximum probability predictions produced by the model from small to large. In addition, the EMA model was employed to produce the sample predictions, for the purpose of improving the quality of the prediction. Afterwards, we annotated the top-k samples and added them to the labeled data to generate the selected dataset D_S. Then, we reinitialized the model and trained it with the data in D_S. Finally, at the last training stage, all the data were added to the labeled data and then the model was retrained by these data. Therefore, at the last training stage, $D_S = D_L \cup D_U$. The anti-curriculum learning strategy is illustrated in Figure 1.
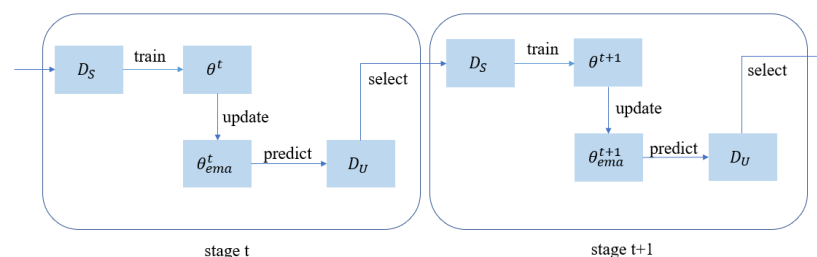
Figure 1. Anti-curriculum learning strategy. The training process is divided into several stages according to the proportion of selected samples. At each training stage, we initialized the model and trained it with D_S. It should be noticed that $D_S = D_L$ at the first stage.

3.2. The Generation of Pseudo Labels

The accuracy of pseudo labels plays an important role in model training. The inaccurate labels make the model learn negative information and accumulate errors, which may greatly degrade the model performance. Thus, it is meaningful to improve the accurate of pseudo labels.

It was proved that the EMA model can effectively improve the performance of the model [10,14,17]. Therefore, we also used the EMA model to produce more accurate predictions. The parameters of the EMA model are updated by the training model:

$$\theta_{ema}^t = \alpha\theta_{ema}^t + (1 - \alpha)\theta^t \tag{1}$$

where θ_{ema}^t and θ^t are the parameters of the EMA model and the training model, respectively, at training iteration t. α is a momentum coefficient. In addition, inspired by the effectiveness of data augmentations in [12], we averaged the sample predictions with different data augmentations. Specifically, we augmented the data for T times and generated the pseudo labels by the average predictions. Here, T is a hyper-parameter. Thus, the pseudo labels $\overline{y}$ can be generated by:

$$\overline{y} = \frac{1}{T}\sum_{j=1}^{T}\theta_{ema}^t(x)_j \tag{2}$$

Figure 2 illustrates the generation of pseudo labels in our method.
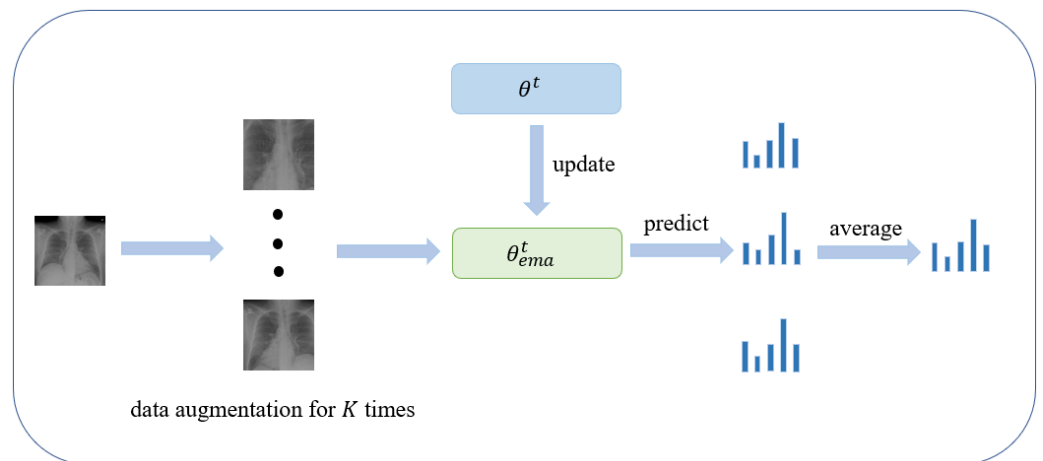
Figure 2. The generation of pseudo labels. Each piece of unlabeled data is augmented for several times and the average predictions produced by the EMA model are used as pseudo labels.

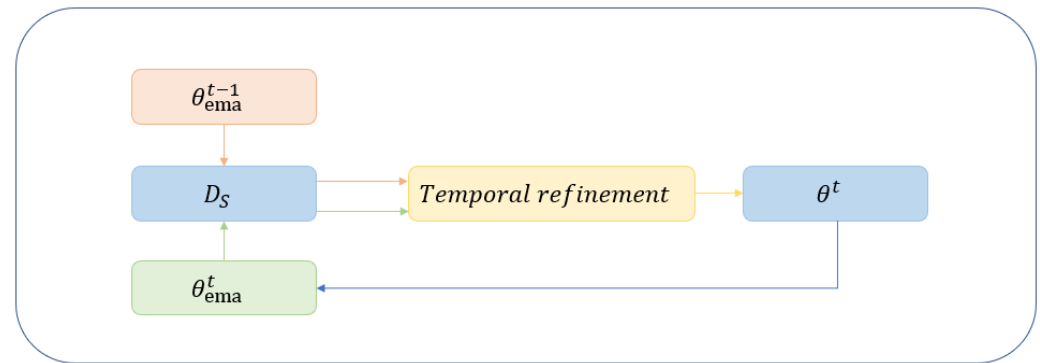### 3.3. Temporal Refinement of Pseudo Labels

According to the anti-curriculum learning strategy, we generated the pseudo labels by the last EMA model. It is known that the model performance can be improved with the training of the model, especially at the early training stage. Therefore, to enhance the model performance, we proposed the temporal refinement to combine the pseudo labels with the prediction of current EMA model. The refined labels can be written as:

$$y' = (1 - \beta)\overline{y} + \beta\theta_{ema}^t(x) \tag{3}$$

where $\overline{y}$ is the pseudo labels and $\theta_{ema}^t(x)$ represents the predictions of the current EMA model, and β is the weight used to control the balance of the pseudo labels and the current predictions. Since the model performance is improved fast at the early training stage, the parameter β is large at the beginning and decreases with the training stage:

$$\beta = \max(\beta_{min}, \beta_{max} - \gamma * s) \tag{4}$$

where β_{min} and β_{max} are the minimum value and maximum value of the parameter β, respectively. γ represents the range of decline and s is the number of training stages. The temporal refinement procedure can be seen in Figure 3.

Figure 3. The temporal refinement of pseudo labels. The pseudo labels generated by the EMA model at the last training stage and the predictions produced by the EMA model at current training stage are linearly combined to generate the current labels to train the model.

4. Experiments

This section describes the datasets, the experiments and implementation details, and also presents the results with some analysis.

4.1. Datasets

The proposed method was evaluated on two different datasets, including Chest X-ray14 [43] and ISIC2018 [44,45]. Chest X-ray14 contains 112,120 images from 30,805 patients with 14 categories. As one patient can have different diseases, each image may also have multiple labels. Thus, the Chest X-ray14 dataset was designed for multi-label classification. We utilized the official training and test dataset, which includes 86,524 and 25,596 images, respectively. In the same way as the previous related work [17], we evaluated our method on the dataset with five different rates of labeled data.

ISIC2018 includes 10,015 images with 7 categories. Each image only has one possible label, so it is designed for multi-class classification. We followed the dataset partition setting in [17] and used 20% of samples in training set as labeled data and the remaining 80% of samples as unlabeled data.

4.2. Implementation Details

We trained the DenseNet121 [46] with ImageNet pretrained parameters for both datasets. For the Chest X-ray14 dataset, the Adam algorithm [47] was employed to train our model. The batch size was set to be 8 and the learning rate was set to be 0.03. For the first training stage, we trained the model with the original labeled data for 20 epochs in all the cases. Afterwards, we trained the model for 10 epochs at each stage. The images were randomly cropped and resized to 512×512 and the random horizontal flip was used as data augmentation. For the ISIC2018 dataset, we also used the Adam algorithm [47] for training. The batch size and the learning rate were set to be 32 and 0.01, respectively. The model was trained for 40 epochs at the first training stage and 20 epochs at other stages. The images were cropped and resized to 224×224 and the data augmentations were the same as that for Chest X-ray14. For both datasets, the parameter T was set as 3, β_{min} and β_{max} were 0.1 and 0.4, respectively. The hyper-parameter γ was set to be 0.1.

4.3. Comparison with Other Semi-Supervised Classification Methods

We compared our method with several other semi-supervised classification methods on the Chest X-ray14 dataset, including Graph XNet [37], UPS [16], SRC-MT [10], No Teacher [36], S^2MTS^2 [38] and ACPL [17]. The experimental results on Chest X-ray14 are

presented in Table 1. Following the settings in [17], we evaluated our method with the rate of number of labeled data in {2%, 5%, 10%, 15%, 20%} and scored the model performance by the area under the ROC curve (AUC). As can be seen in Table 1, our method performed well in all the rates, especially with few labeled data. When using 2% of the labeled data, our method achieved an AUC of 0.7539, which is higher than that of the state-of-the-art method ACPL [17] by 0.0057. The results listed in Table 1 verify the effectiveness of our method for multi-label classification task.

Table 1. Comparison results between our method and other compared methods in AUC on Chest X-ray14 dataset. The bold represents the best result.

Method	Label Percentage				
	2%	5%	10%	15%	20%
Graph XNet [37]	0.5300	0.5800	0.6300	0.6800	0.7800
UPS [16]	0.6551	0.7318	0.7684	0.7890	0.7992
SRC-MT [10]	0.6695	0.7229	0.7528	0.7776	0.7923
NoTeacher [36]	0.7260	0.7704	0.7761	N/A	0.7949
S^2MTS^2 [38]	0.7469	0.7896	0.7990	0.8031	0.8106
ACPL [17]	0.7482	0.7920	0.8040	0.8106	0.8177
Ours	**0.7539**	**0.7952**	**0.8068**	**0.8123**	**0.8197**

Table 2 shows the performance comparison of the ISIC2018 dataset. Since the ISIC2018 dataset contains a small number of images, we only conducted the experiments by using 20% labeled data, in the same way as the method in [17]. We compared our method with the self-training-based method [48], GAN-based method [8], TemporalEnsemble [30], MeanTeacher [14], TCSE [49], SRC-MT [10], ACPL [17] and S^2MTS^2 [38]. We also scored the model performance by AUC. As shown in Table 2, our method achieved an AUC of 0.9612, which is higher than those of other compared methods. The result demonstrates that our method is not only suitable for multi-label classification task, but also for multi-class classification task.

Table 2. Performance of different methods on ISIC2018 dataset with 20% labeled data for training. The bold represents the best result.

| Method	AUC
Self-training [48]	0.9058
SS-DCGAN [8]	0.9128
TCSE [49]	0.9224
TemporalEnsemble [30]	0.9270
MeanTeacher [14]	0.9296
SRC-MT [10]	0.9358
ACPL [17]	0.9436
S^2MTS^2 [38]	0.9471
Ours	**0.9612**

4.4. Ablation Study

In this subsection, we conducted ablation studies to investigate the influence of different components in our method on Chest X-ray14 dataset with 2% labeled data and ISIC2018 with 20% labeled data. First, we compared the performance of model with curriculum learning strategy and anti-curriculum learning strategy. As displayed in Table 3, by using the curriculum learning strategy on the Chest X-ray14 dataset, our method achieved an AUC of 0.7509, which is 0.003 lower than that of our method with the anti-curriculum learning strategy. For the ISIC2018 dataset, the result increased from 0.9553 to 0.9612 by using the anti-curriculum learning strategy instead of curriculum learning strategy. The experimental results on both datasets show the effectiveness of the anti-curriculum learning strategy. The model was trained by the original labeled data first, and thus it may select the

samples which are similar to the labeled data as easy samples. The curriculum learning strategy makes the model learn from easy samples to hard samples. However, training the model with similar samples cannot significantly improve the model performance. Therefore, at the early training stage, the performance of the model with the curriculum learning strategy was lower than that of model with anti-curriculum learning strategy, which can be seen in Figures 4 and 5. It may be the reason why the anti-curriculum learning strategy is more useful.

Table 3. Performance of different learning strategies on Chest X-ray14 dataset with 2% labeled data and ISIC2018 with 20% labeled data. CL: Curriculum learning; ACL: Anti-curriculum learning. The bold represents the best result.

| Method	Chest X-ray14	ISIC2018
Ours with CL strategy	0.7509	0.9553
Ours with ACL strategy	**0.7539**	**0.9612**
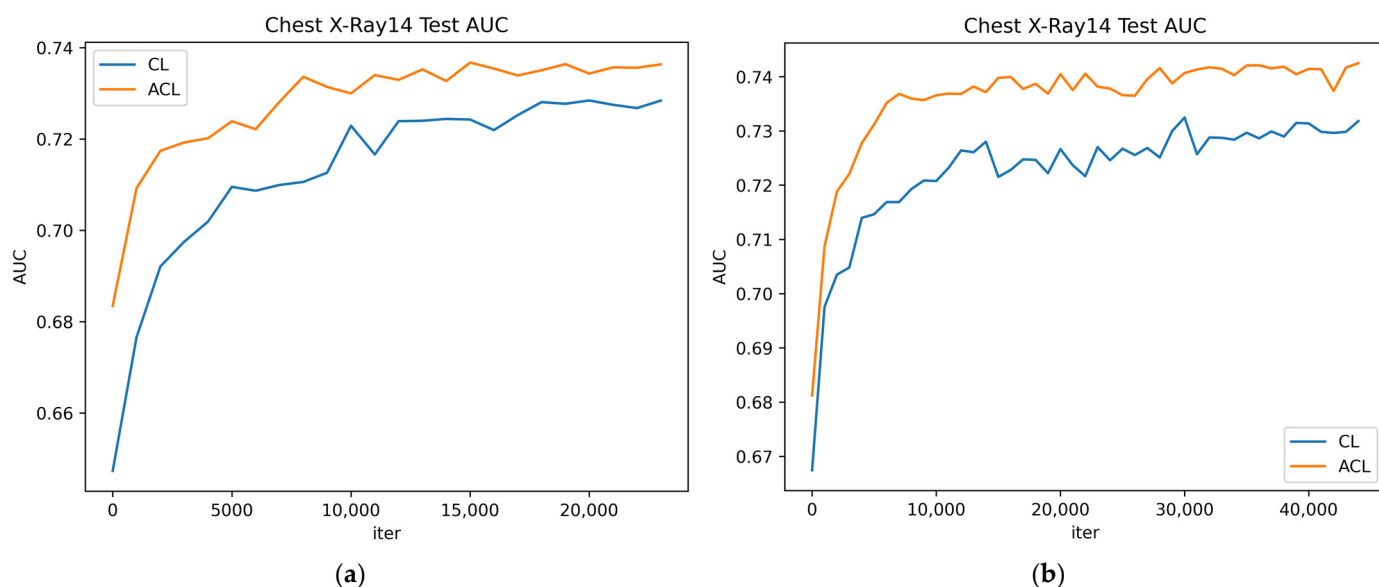
Figure 4. The test AUC on the Chest X-ray14 dataset at the second and third training stage. (**a**) The result at the second training stage; (**b**) The result at the third training stage.

Since the label distribution in a medical dataset is often imbalanced, Figure 6 illustrated the pseudo label distributions of the data selected by the models trained with the curriculum learning strategy and anti-curriculum learning strategy after the second training stage on the ISIC2018 dataset, respectively. It can be seen in Figure 6 that the model trained with the anti-curriculum learning strategy selects fewer data of the majority class and more data of the minority class, which means the selection of the model trained with anti-curriculum learning strategy is more balanced.

To increase the accuracy of pseudo labels, we augmented data for several times to generate the pseudo labels and employed temporal refinement to train the model. We carried out the experiments to inspect the effect of each component. Table 4 reports the results of our method with and without each component. Our method obtained the AUC of 0.7495 and 0.9588 without average prediction and temporal refinement on the Chest X-ray14 dataset and ISIC2018 dataset, respectively. By employing the temporal refinement in our method, the result increased to 0.7503 and 0.9592, which shows the effectiveness of temporal refinement. It can also observe an increase by using the average prediction to generate pseudo labels, with the AUC increasing to 0.7528 and 0.9597. By employing both components in our method, we achieved the AUC of 0.7539 and 0.9612, which strongly show the effectiveness of our method.

(**a**)   (**b**)
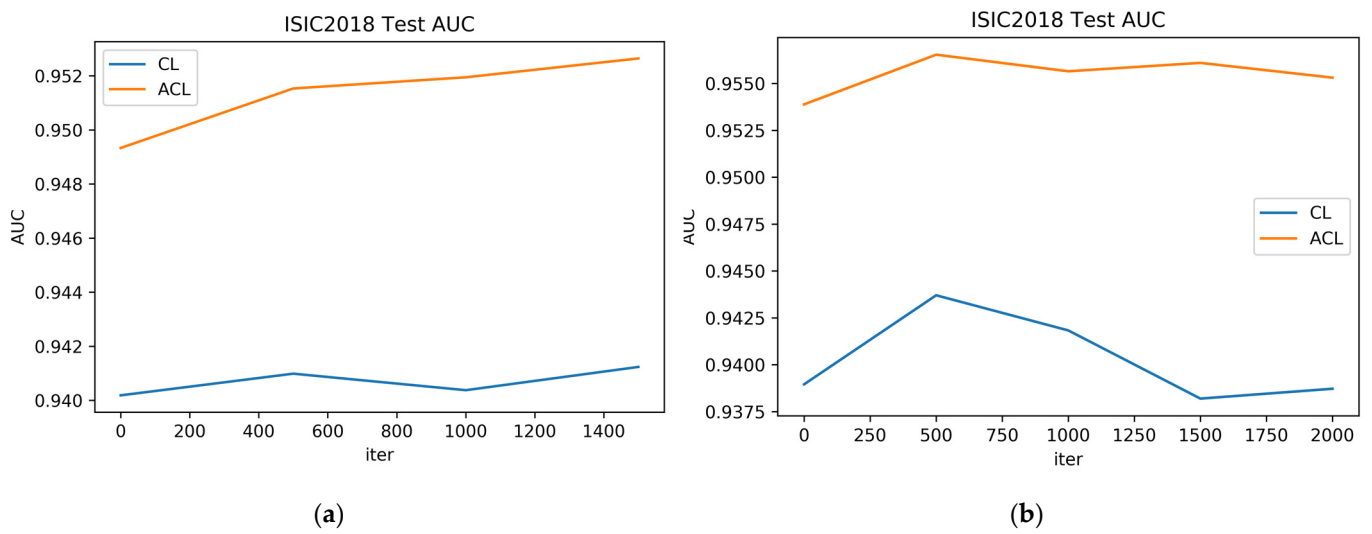
Figure 5. The test AUC on ISIC2018 dataset at the second and third training stage. (**a**) The result at the second training stage; (**b**) The result at the third training stage.
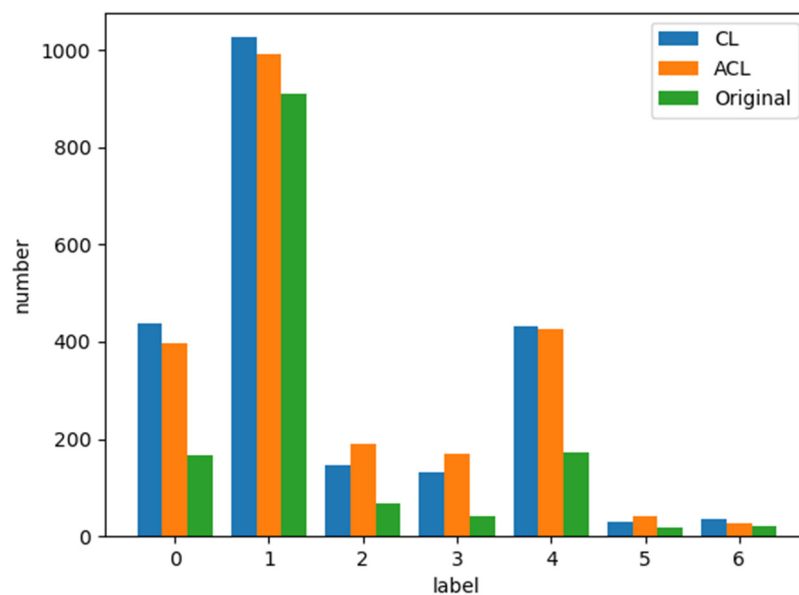

Figure 6. The pseudo label distributions. The original label distributions are the labels of samples selected as the labeled data.

Table 4. The effectiveness of different components on Chest X-ray14 dataset with 2% labeled data and ISIC2018 with 20% labeled data. The bold represents the best result.

Method		Chest X-ray14	ISIC2018
Average Prediction	**Temporal Refinement**		
		0.7495	0.9588
	√	0.7503	0.9592
√		0.7528	0.9597
√	√	**0.7539**	**0.9612**

We can visualize some images with their generated pseudo labels in Figure 7, where we can see the predictions of our method for thorax disease and skin cancers.
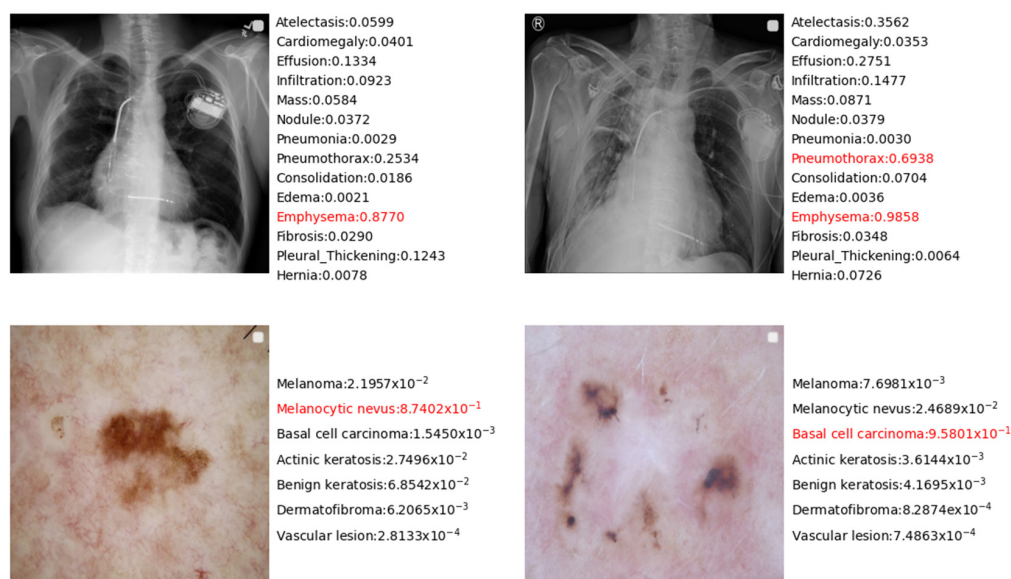
Figure 7. Some images and their corresponding predictions. Top: images from Chest X-ray14 dataset. Bottom: images from ISIC2018 dataset. The red classes mean the ground-truth.

5. Conclusions

In this paper, we proposed a simple and effective method for medical image classification. To avoid confirmation bias and promote the model training, we trained the model with the anti-curriculum learning strategy. We also averaged the model predictions of samples with different data augmentations to improve the accuracy of pseudo labels. Since the model performance improves over time, we designed the temporal refinement to linearly combine the pseudo labels generated from the model at the last stage and the predictions of the model at the current stage, in order to further improve the model performance and the accuracy of pseudo labels. Our method was evaluated on two different medical image datasets for multi-label and multi-class classification task. The experimental results verify the effectiveness of our method. In the future, we would like to evaluate our method in general computer vision tasks and other medical image processing such as image segmentation and tumor detection.

Author Contributions: Methodology, software, validation, writing—original draft preparation, H.W.; writing—review and editing, Q.Y.; supervision, J.S. All authors have read and agreed to the published version of the manuscript.

References

1. Liu, Z.; Su, W.; Ao, J.; Wang, M.; Jiang, Q.; He, J.; Gao, H.; Lei, S.; Nie, J.; Yan, X.; et al. Instant diagnosis of gastroscopic biopsy via deep-learned single-shot femtosecond stimulated Raman histology. *Nat. Commun.* **2022**, *13*, 4050. [CrossRef] [PubMed]
2. Nguyen, Q.H.; Nguyen, B.P.; Nguyen, M.T.; Chua, M.C.; Do, T.T.; Nghiem, N. Bone age assessment and sex determination using transfer learning. *Expert Syst. Appl.* **2022**, *200*, 116926. [CrossRef]
3. Zhang, X.; Huang, D.; Li, H.; Zhang, Y.; Xia, Y.; Liu, J. Self-training maximum classifier discrepancy for EEG emotion recognition. *CAAI Trans. Intell. Technol.* **2023**. [CrossRef]
4. Liu, H.; Liu, M.; Li, D.; Zheng, W.; Yin, L.; Wang, R. Recent advances in pulse-coupled neural networks with applications in image processing. *Electronics* **2022**, *11*, 3264. [CrossRef]

5. Nguyen, Q.H.; Muthuraman, R.; Singh, L.; Sen, G.; Tran, A.C.; Nguyen, B.P.; Chua, M. Diabetic retinopathy detection using deep learning. In Proceedings of the 4th International Conference on Machine Learning and Soft Computing, Haiphong City, Vietnam, 17–19 January 2020; pp. 103–107.

6. Nguyen, Q.H.; Nguyen, B.P.; Dao, S.D.; Unnikrishnan, B.; Dhingra, R.; Ravichandran, S.R.; Satpathy, S.; Raja, P.N.; Chua, M.C. Deep learning models for tuberculosis detection from chest X-ray images. In Proceedings of the 26th International Conference on Telecommunications (ICT), Hanoi, Vietnam, 8–10 April 2019; IEEE: New York, NY, USA, 2019; pp. 381–385.

7. Jin, K.; Huang, X.; Zhou, J.; Li, Y.; Yan, Y.; Sun, Y.; Zhang, Q.; Wang, Y.; Ye, J. Fives: A fundus image dataset for artificial Intelligence based vessel segmentation. *Sci. Data* **2022**, *9*, 475. [CrossRef]

8. Diaz-Pinto, A.; Colomer, A.; Naranjo, V.; Morales, S.; Xu, Y.; Frangi, A.F. Retinal Image Synthesis and Semi-Supervised Learning for Glaucoma Assessment. *IEEE Trans. Med. Imaging* **2019**, *38*, 2211–2218. [CrossRef] [PubMed]

9. Gyawali, P.K.; Ghimire, S.; Bajracharya, P.; Li, Z.; Wang, L. Semi-supervised medical image classification with global latent mixing. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin, Germany, 2020; Volume 12261 LNCS, pp. 604–613. [CrossRef]

10. Liu, Q.; Yu, L.; Luo, L.; Dou, Q.; Heng, P.A. Semi-Supervised Medical Image Classification with Relation-Driven Self-Ensembling Model. *IEEE Trans. Med. Imaging* **2020**, *39*, 3429–3440. [CrossRef] [PubMed]

11. Miyato, T.; Maeda, S.I.; Koyama, M.; Ishii, S. Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1979–1993. [CrossRef] [PubMed]

12. Berthelot, D.; Carlini, N.; Goodfellow, I.; Oliver, A.; Papernot, N.; Raffel, C. MixMatch: A holistic approach to semi-supervised learning. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation: La Jolla, CA, USA, 2019; Volume 32.

13. Sohn, K.; Berthelot, D.; Li, C.L.; Zhang, Z.; Carlini, N.; Cubuk, E.D.; Raffel, C. FixMatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation: La Jolla, CA, USA, 2020.

14. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation: La Jolla, CA, USA, 2017; pp. 1196–1205.

15. Lee, D.-H. Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. ICML 2013 Workshop: Challenges in Representation Learning. pp. 1–6. Available online: https://www.kaggle.com/blobs/download/forum-message-attachment-files/746/pseudo_label_final.pdf (accessed on 16 December 2022).

16. Rizve, M.N.; Duarte, K.; Rawat, Y.S.; Shah, M. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. *arXiv* **2021**. [CrossRef]

17. Liu, F.; Tian, Y.; Chen, Y.; Liu, Y.; Belagiannis, V.; Carneiro, G. ACPL: Anti-curriculum Pseudo-labelling for Semi-supervised Medical Image Classification. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 20665–20674. [CrossRef]

18. Xie, Q.; Luong, M.T.; Hovy, E.; Le, Q.V. Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10684–10695. [CrossRef]

19. Qiao, S.; Shen, W.; Zhang, Z.; Wang, B.; Yuille, A. Deep co-training for semi-supervised image recognition. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11219 LNCS, pp. 142–159. [CrossRef]

20. Chen, D.D.; Wang, W.; Gao, W.; Zhou, Z.H. Tri-net for semi-supervised deep learning. In Proceedings of the IJCAI International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 2014–2020. [CrossRef]

21. Ke, Z.; Wang, D.; Yan, Q.; Ren, J.; Lau, R. Dual student: Breaking the limits of the teacher in semi-supervised learning. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2019; pp. 6727–6735. [CrossRef]

22. Cascante-Bonilla, P.; Tan, F.; Qi, Y.; Ordonez, V. Curriculum Labeling: Revisiting Pseudo-Labeling for Semi-Supervised Learning. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*. [CrossRef]

23. Dai, Z.; Yang, Z.; Yang, F.; Cohen, W.W.; Salakhutdinov, R. Good semi-supervised learning that requires a bad GAN. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation: La Jolla, CA, USA, 2017; pp. 6511–6521.

24. Kumar, A.; Sattigeri, P.; Fletcher, P.T. Semi-supervised learning with GANs: Manifold invariance with improved inference. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation: La Jolla, CA, USA, 2017; pp. 5535–5545.

25. Wu, S.; Deng, G.; Li, J.; Li, R.; Yu, Z.; Wong, H.S. Enhancing triplegan for semi-supervised conditional instance synthesis and classification. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10083–10092. [CrossRef]

26. Liu, Y.; Deng, G.; Zeng, X.; Wu, S.; Yu, Z.; Wong, H.S. Regularizing discriminative capability of cGANs for semi-supervised generative learning. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5719–5728. [CrossRef]

27. Luo, Y.; Zhu, J.; Li, M.; Ren, Y.; Zhang, B. Smooth Neighbors on Teacher Graphs for Semi-Supervised Learning. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8896–8905. [CrossRef]

28. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017-Conference Track Proceedings, Toulon, France, 24–26 April 2017.

29. Iscen, A.; Tolias, G.; Avrithis, Y.; Chum, O. Label propagation for deep semi-supervised learning. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5065–5074. [CrossRef]

30. Laine, S.; Aila, T. Temporal ensembling for semi-supervised learning. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017-Conference Track Proceedings, Toulon, France, 24–26 April 2017.

31. Berthelot, D.; Carlini, N.; Cubuk, E.D.; Kurakin, A.; Raffel, C. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv* **2019**, arXiv:1911.09785.

32. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. Randaugment: Practical automated data augmentation with a reduced search space. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Washington, DC, USA, 16–18 June 2020; pp. 3008–3017. [CrossRef]

33. Zhang, B.; Wang, Y.; Hou, W.; Wu, H.; Wang, J.; Okumura, M.; Shinozaki, T. FlexMatch: Boosting Semi-Supervised Learning with Curriculum Pseudo Labeling. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 18408–18419.

34. Li, J.; Xiong, C.; Hoi, S.C.H. CoMatch: Semi-supervised Learning with Contrastive Graph Regularization. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 9455–9464. [CrossRef]

35. Zheng, M.; You, S.; Huang, L.; Wang, F.; Qian, C.; Xu, C. SimMatch: Semi-supervised Learning with Similarity Matching. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 14451–14461. [CrossRef]

36. Unnikrishnan, B.; Nguyen, C.; Balaram, S.; Li, C.; Foo, C.S.; Krishnaswamy, P. Semi-supervised classification of diagnostic radiographs with noteacher: A teacher that is not mean. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; Springer: Cham, Switzerland, 2020; pp. 624–634.

37. Aviles-Rivero, A.I.; Papadakis, N.; Li, R.; Sellars, P.; Fan, Q.; Tan, R.T.; Schönlieb, C.B. GraphX NET- Chest X-ray Classification Under Extreme Minimal Supervision. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; Springer: Cham, Switzerland, 2019; pp. 504–512.

38. Liu, F.; Tian, Y.; Cordeiro, F.R.; Belagiannis, V.; Reid, I.; Carneiro, G. Self-supervised Mean Teacher for Semi-supervised Chest X-ray Classification. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin, Germany, 2021; Volume 12966 LNCS, pp. 426–436. [CrossRef]

39. Bengio, Y.; Louradour, J.; Collobert, R.; Weston, J. Curriculum learning. In Proceedings of the ACM International Conference Proceeding Series, Athens, Greece, 29–31 October 2009; Association for Computing Machinery: New York, NY, USA, 2019; Volume 382. [CrossRef]

40. Jiang, L.; Meng, D.; Zhao, Q.; Shan, S.; Hauptmann, A.G. Self-paced curriculum learning. In Proceedings of the National Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; Volume 4, pp. 2694–2700. [CrossRef]

41. Jiang, A.H.; Wong, L.K.; Zhou, G.; Andersen, D.G.; Dean, J.; Ganger, G.R.; Joshi, G.; Kaminksy, M.; Kozuch, M.; Lipton, Z.C.; et al. Accelerating Deep Learning by Focusing on the Biggest Losers. *arXiv* **2019**, arXiv:1910.00762. [CrossRef]

42. Hacohen, G.; Weinshall, D. On the power of curriculum learning in training deep networks. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, CA, USA, 10–15 October 2019; pp. 4483–4496.

43. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 3462–3471. [CrossRef]

44. 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* **2018**, arXiv:1902.03368.

45. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 180161. [CrossRef] [PubMed]

46. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]

47. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.

48. Bai, W.; Oktay, O.; Sinclair, M.; Suzuki, H.; Rajchl, M.; Tarroni, G.; Rueckert, D. Semi-supervised learning for network-based cardiac MR image segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10434 LNCS, pp. 253–260. [CrossRef]

49. Li, X.; Yu, L.; Chen, H.; Fu, C.W.; Heng, P.A. Semi-supervised Skin Lesion Segmentation via Transformation Consistent Self-ensembling Model. In Proceedings of the British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, 3–6 September 2018.