Combat Long-tails in Medical Classification with Relation-aware Consistency and Virtual Features Compensation

Li Pan[1*], Yupei Zhang[2*], Qiushi Yang[3], Tan Li[4], Zhen Chen[5✉]

[1] The Chinese University of Hong Kong
[2] The Centre for Intelligent Multidimensional Data Analysis (CIMDA)
[3] City University of Hong Kong
[4] Department of Computer Science, The Hang Seng University of Hong Kong
[5] Centre for Artificial Intelligence and Robotics (CAIR), HKISI-CAS
zhen.chen@cair-cas.org.hk

Abstract. Deep learning techniques have achieved promising performance for computer-aided diagnosis, which is beneficial to alleviate the workload of clinicians. However, due to the scarcity of diseased samples, medical image datasets suffer from an inherent imbalance, and lead diagnostic algorithms biased to majority categories. This degrades the diagnostic performance, especially in recognizing rare categories. Existing works formulate this challenge as long-tails and adopt decoupling strategies to mitigate the effect of the biased classifier. But these works only use the imbalanced dataset to train the encoder and resample data to re-train the classifier by discarding the samples of head categories, thereby restricting the diagnostic performance. To address these problems, we propose a Multi-view Relation-aware Consistency and Virtual Features Compensation (MRC-VFC) framework for long-tailed medical image classification in two stages. In the first stage, we devise a Multi-view Relation-aware Consistency (MRC) for representation learning, which provides the training of encoders with unbiased guidance in addition to the imbalanced supervision. In the second stage, to produce an impartial classifier, we propose the Virtual Features Compensation (VFC) to recalibrate the classifier by generating massive balanced virtual features. Compared with the resampling, VFC compensates the minority classes to optimize an unbiased classifier with preserving complete knowledge of the majority ones. Extensive experiments on two long-tailed public benchmarks confirm that our MRC-VFC framework remarkably outperforms state-of-the-art algorithms.

Keywords: Class Imbalance · Dermoscopy · Representation Learning.

1 Introduction

Recent years have witnessed the great success of deep learning techniques in various applications on computer-aided diagnosis [5, 6, 9, 23]. However, the chal-

∗ Equal contribution.

lenge of class imbalance inherently exists in medical datasets due to the scarcity of target diseases [7], where normal samples are significantly more than diseased samples. This challenge leads the model training biased to the majority categories [13] and severely impairs the performance of diagnostic models in real-world scenarios [8,18]. Therefore, it is urgent to improve the performance of diagnostic models in clinical applications, especially to achieve balanced recognition of minority categories.

Technically, the issue of class imbalance is formulated as a long-tailed problem in existing works [10, 12, 15], where a few head classes contain numerous samples while the tail classes comprise only a few instances [28]. To address this issue, most of the previous methods have typically attempted to rebalance the data distribution through under-sampling the head classes [2], over-sampling the tail classes [21], or reweighting the contribution of different classes during the optimization process [4, 8]. Nevertheless, these resampling methods can encounter a decrease in performance on certain datasets since the total information volume of the dataset is either unchanged or even reduced [29]. Recent advantages in long-tailed medical image classification have been achieved by two-stage methods, which first train the model on the entire dataset and then fine-tune the classifier in the second stage using rebalancing techniques to counteract the class imbalance [12, 14, 15, 19]. By decoupling the training of encoders and classifiers, the two-stage methods can recalibrate the biased classifiers and utilize all of the training samples to enhance representation learning for the encoder.

Although the aforementioned decoupling methods [14,15] have somewhat alleviated the long-tails, the classification performance degradation in the minority classes remains unsolved, which can be attributed to two challenges. First, in the first stage, the decoupling methods train the model on the imbalanced dataset, which is insufficient for representation learning in the rare classes due to the scarcity of samples [17]. To this end, improving the first-stage training strategy to render effective supervision on representation learning is in great demand. The second problem lies in the second stage, where decoupling methods freeze the pre-trained encoder and fine-tune the classifier [14, 15]. Traditional rebalancing techniques, such as resampling and reweighting, are used by the decoupling methods to eliminate the bias in the classifier. However, these rebalancing strategies have intrinsic drawbacks, e.g., resampling-based methods discard the samples of head classes, and reweighting cannot eliminate the imbalance with simple coefficients [26]. Thus, a novel approach that can perform balanced classifier training by generating abundant features is desired to recalibrate the classifier and preserve the representation quality of the encoder.

To address the above two challenges, we propose the MRC-VFC framework that adopts the decoupling strategy to enhance the first-stage representation learning with Multi-view Relation-aware Consistency (MRC) and recalibrate the classifier using Virtual Features Compensation (VFC). Specifically, in the first stage, to boost the representation learning under limited samples, we build a two-stream architecture to perform representation learning with the MRC module, which encourages the model to capture semantic information from images under

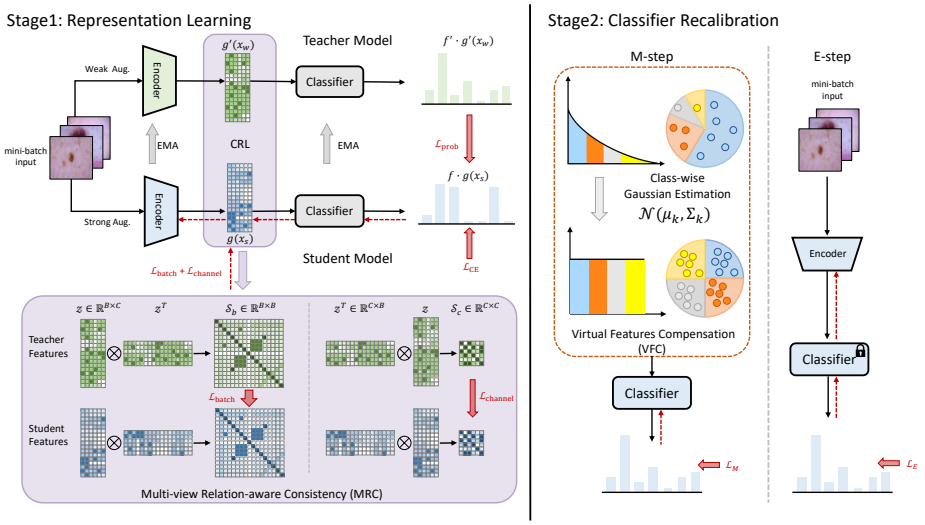

Fig. 1. The MRC-VFC framework. In stage 1, we perform the representation learning with the MRC module for the encoder on the imbalanced dataset. In stage 2, we recalibrate the classifier with VFC in two-step of the expectation and maximization.

different data perturbations. In the second stage, to recalibrate the classifier, we propose to generate virtual features from multivariate Gaussian distribution with the expectation-maximization algorithm, which can compensate for tail classes and preserves the correlations among features. In this way, the proposed MRC-VFC framework can rectify the biases in the encoder and classifier, and construct a balanced and representative feature space to improve the performance for rare diseases. Experiments on two public dermoscopic datasets prove that our MRC-VFC framework outperforms state-of-the-art methods for long-tailed diagnosis.

2 Methodology

As illustrated in Fig. 1, our MRC-VFC framework follows the decoupling strategy [14,31] to combat the long-tailed challenges in two stages. In the first stage, we introduce the Multi-view Relation-aware Consistency (MRC) to boost representation learning for the encoder g. In the second stage, the proposed Virtual Features Compensation (VFC) recalibrates the classifier f by generating massive balanced virtual features, which compensates the tails classes without dropping the samples of the head classes. By enhancing the encoder with MRC and recalibrating the classifier with VFC, our MRC-VFC framework can perform effective and balanced training on long-tailed medical datasets.

2.1 Multi-view Relation-aware Consistency

The representation learning towards the decoupling models is insufficient [28, 29]. To boost the representation learning, we propose the Multi-view Relation-aware Consistency to encourage the encoder to apprehend the inherent semantic features of the input images under different data augmentations. Specifically, we build a student neural network $f \cdot g$ for the strong augmented input $\boldsymbol{x}_s$ and duplicate a teacher model $f' \cdot g'$ for the weak augmented input $\boldsymbol{x}_w$. The two models are constrained by the MRC module to promote the consistency for different perturbations of the same input. The parameters of the teacher model are updated via an exponential moving average of the student parameters [24].

To motivate the student model to learn from the data representations but the ill distributions, we propose multi-view constraints on the consistency of two models at various phases. A straightforward solution is to encourage identical predictions for different augmentations of the same input image, as follows:

$$\mathcal{L}_{\text{prob}} \;=\; \frac{1}{B}\text{KL}(f \cdot g(\boldsymbol{x}_s)\;,\; f' \cdot g'(\boldsymbol{x}_w)), \tag{1}$$

where $\text{KL}(\cdot,\cdot)$ refers to the Kullback–Leibler divergence to measure the difference between two outputs. As this loss function calculates the variance of classifier output, the supervision for the encoders is less effective. To this end, the proposed MRC measures the sample-wise and channel-wise similarity between the feature maps of two encoders to regularize the consistency of the encoders. We first define the correlations of individuals and feature channels as $\mathcal{S}_b(\boldsymbol{z}) \;=\; \boldsymbol{z} \cdot \boldsymbol{z}^\intercal$ and $\mathcal{S}_c(\boldsymbol{z}) \;=\; \boldsymbol{z}^\intercal \cdot \boldsymbol{z}$, where $\boldsymbol{z} = g(\boldsymbol{x}_s) \in \mathbb{R}^{B\times C}$ is the output features of the encoder, and B and C are the batch size and channel number. $\mathcal{S}_b(\boldsymbol{z})$ denotes the Gram matrix of feature $\boldsymbol{z}$, representing the correlations among individuals, and $\mathcal{S}_c(\boldsymbol{z})$ indicates the similarities across feature channels. Thus, the consistency between the feature maps of two models can be defined as:

$$\mathcal{L}_{\text{batch}} \;=\; \frac{1}{B}||\mathcal{S}_b(g(\boldsymbol{x}_s)) \;-\; \mathcal{S}_b(g'(\boldsymbol{x}_w))||_2, \tag{2}$$

$$\mathcal{L}_{\text{channel}} \;=\; \frac{1}{C}||\mathcal{S}_c(g(\boldsymbol{x}_s)) \;-\; \mathcal{S}_c(g'(\boldsymbol{x}_w))||_2. \tag{3}$$

Furthermore, we also adopt the cross-entropy loss $\mathcal{L}_{\text{CE}} \;=\; \frac{1}{B}L(f \cdot g(\boldsymbol{x}_w),\; y)$, where y denotes the ground truth, between the predictions and ground truth to ensure that the optimization will not be misled to a trivial solution. The overall loss function is summarized as $\mathcal{L}_{\text{stage1}} \;=\; \mathcal{L}_{\text{CE}} + \lambda_1\mathcal{L}_{\text{batch}} + \lambda_2\mathcal{L}_{\text{channel}} + \lambda_3\mathcal{L}_{\text{prob}}$, where λ_1, λ_2 and λ_3 are coefficients to control the trade-off of each loss term. By introducing extra semantic constraints, the MRC can enhance the representation capacity of encoders. The feature space generated by the encoders is more balanced with abundant semantics, thereby facilitating the MRC-VFC framework to combat long-tails in medical diagnosis.

2.2 Virtual Features Compensation

Recalling the introduction of decoupling methods, the two-stage methods [14] decouple the training of the encoder and classifier to eliminate the bias in the classifier while retaining the representation learning of the encoder. However, most existing decoupling approaches [12, 15] employ the resampling strategy in the second stage to rebalance the data class distribution, causing the intrinsic drawbacks of the resampling of discarding the head class samples. To handle this issue, we propose Virtual Features Compensation, which generates virtual features $z_k \in \mathbb{R}^{N_k \times C}$ for each class k under multivariate Gaussian distribution [1] to combat the long-tailed problem. Different from existing resampling methods [2], the feature vectors produced by the VFC module preserve the correlations among classes and the semantic information from the encoder. Given the k-th class, we first calculate the class-wise Gaussian distribution with mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$, as follows:

$$\boldsymbol{\mu}_k \;=\; \frac{1}{N_k} \sum_{\boldsymbol{x} \in X_k} g^I(\boldsymbol{x}),\;\; \boldsymbol{\Sigma}_k \;=\; \frac{1}{N_k - 1} \sum_{\boldsymbol{x} \in X_k} (\boldsymbol{x} - \boldsymbol{\mu}_k)^\intercal (\boldsymbol{x} - \boldsymbol{\mu}_k), \tag{4}$$

where X_k denotes the set of all samples in the k-th class, and $g^I(\cdot)$ denotes the encoder trained in the first stage on the imbalanced dataset and N_k is the sample number of the k-th class. We then randomly sample R feature vectors for each category from the corresponding Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ to build the unbiased feature space, as $\{V_k \in \mathbb{R}^{R \times C}\}_{k=1}^{K}$. We re-initialize the classifier and then calibrated it under cross-entropy loss, as follows:

$$\mathcal{L}^M_{\text{stage2}} \;=\; \frac{1}{RK} \sum_{k=1}^{K} \sum_{\boldsymbol{v}_i \in V_k} L_{\text{CE}}(f(\boldsymbol{v}_i), y), \tag{5}$$

where K is the number of categories in the dataset. As the Gaussian distribution is calculated according to the statistics from the first-stage feature space, to further alleviate the potential bias, we employ the expectation-maximization algorithm [20] to iteratively fine-tune the classifier and encoder. At the expectation step, we freeze the classifier and supervise the encoder with extra balancing constraints to avoid being re-contaminated by the long-tailed label space. Thus, we adopt the generalized cross-entropy (GCE) loss [30] for the expectation step as follows:

$$\mathcal{L}^E_{\text{stage2}} \;=\; \frac{1}{N} \sum_{\boldsymbol{x} \in X} \frac{(1 - (f \cdot g^I(\boldsymbol{x}) y)^q)}{q}, \tag{6}$$

where q is a hyper-parameter to control the trade-off between the imbalance calibration and the classification task. At the maximization step, we freeze the encoder and train the classifier on the impartial feature space. By enriching the semantic features with balanced virtual features, our MRC-VFC framework can improve the classification performance in long-tailed datasets, especially the performance of minority categories.

Table 1. Comparison with state-of-the-art algorithms on the *ISIC-2019-LT* dataset.

ISIC-2019-LT			
Methods	Acc(%) @ Factor=100	Acc(%) @ Factor=200	Acc(%) @ Factor=500
CE	56.91	53.77	43.89
RS	61.41	55.12	47.76
MixUp [27]	59.85	54.23	43.11
GCE+SR [32]	64.57	58.28	54.36
Seesaw loss [26]	68.82	65.84	62.92
Focal loss [16]	67.54	65.93	61.66
CB loss [8]	67.54	66.70	61.89
FCD [15]	70.15	68.82	63.59
FS [12]	71.97	69.30	65.22
Ours *w/o* MRC	75.04	73.13	70.13
Ours *w/o* VFC	72.91	71.07	67.48
Ours	**77.41**	**75.98**	**74.62**

3 Experiments

3.1 Datasets

To evaluate the performance on long-tailed medical image classification, we construct two dermatology datasets from ISIC[1] [25] following [12]. In particular, we construct the *ISIC-2019-LT* dataset as the long-tailed version of ISIC 2019 challenge[2], which includes 8 diagnostic categories of dermoscopic images. We sample the subset from Pareto distribution [8] as $N_c = N_0(r^{-(k-1)})^c$, where the imbalance factor $r = N_0/N_{k-1}$ is defined as the sample number of the head class N_0 divided by the tail one N_{k-1}. We adopt three imbalance factors for *ISIC-2019-LT*, as $r = \{100, 200, 500\}$. Furthermore, the *ISIC-Archive-LT* dataset [12] is sampled from ISIC Archive with a larger imbalance factor $r \approx 1000$ and contains dermoscopic images of 14 classes. We randomly split these two datasets into train, validation and test sets as 7:1:2.

3.2 Implementation Details

We implement the proposed MRC-VFC framework with the PyTorch library [22], and employ ResNet-18 [11] as the encoder for both long-tailed datasets. All the experiments are done on four NVIDIA GeForce GTX 1080 Ti GPUs with a batch size of 128. All images are resized to 224×224 pixels. In the first stage of MRC-VFC, we train the model using Stochastic Gradient Descent (SGD) with a learning rate of 0.01. For the strong augmentation [3], we utilize the random flip, blur, rotate, distortion, color jitter, grid dropout, and normalization, and adopt the random flip and the same normalization for the weak augmentation. In the second stage, we use SGD with a learning rate of 1×10^{-5} for optimizing the classifier and 1×10^{-6} for optimizing the encoder, respectively. The loss

[1] `https://www.isic-archive.com/`

[2] `https://challenge.isic-archive.com/landing/2019/`

Table 2. Comparison with state-of-the-art algorithms on the *ISIC-Archive-LT* dataset.

ISIC-Archive-LT				
Methods	Head (Acc%)	Medium (Acc%)	Tail (Acc%)	All (Acc%)
CE	**71.31**	49.22	38.17	52.90
RS	70.17	55.29	34.29	53.25
GCE+SR [32]	64.93	57.26	38.22	53.47
Seesaw loss [26]	70.26	55.98	42.14	59.46
Focal loss [16]	69.57	56.21	39.65	57.81
CB loss [8]	64.98	57.01	61.61	61.20
FCD [15]	66.39	61.17	60.54	62.70
FS [12]	68.69	58.74	64.48	63.97
Ours *w/o* MRC	69.06	62.14	65.12	65.44
Ours *w/o* VFC	65.11	62.35	67.30	64.92
Ours	69.71	**63.47**	**70.34**	**67.84**

weights λ_1, λ_2 and λ_3 in the first stage are set as 10, 10 and 5, and the q in the second stage is set as 0.8. We set training epochs as 100 for the first stage and 500 for the second stage. The source code is available at `https://github.com/jhonP-Li/MRC_VFC`.

3.3 Comparison on ISIC-2019-LT Dataset

We evaluate the performance of our MRC-VFC framework with state-of-the-art methods for long-tailed medical image classification, including (i) baselines: fine-tuning classification models with cross-entropy loss (CE), random data resampling methods (RS), and MixUp [27]; (ii) recent loss reweighting methods: Generalized Cross-Entropy with Sparse Regularization (GCE+SR) [32], Seesaw loss [26], focal loss [16], and Class-Balancing (CB) loss [8]; (iii) recent works for long-tailed medical image classification: Flat-aware Cross-stage Distillation (FCD) [15], and Flexible Sampling (FS) [12].

As illustrated in Table 1, we compare our MRC-VFC framework with the aforementioned methods on the *ISIC-2019-LT* dataset under different imbalance factors. Among these methods, our MRC-VFC framework achieves the best performance with an accuracy of 77.41%, 75.98%, and 74.62% under the imbalance factor of 100, 200, and 500, respectively. Noticeably, compared with the state-of-the-art decoupling method FCD [15] on long-tailed medical image classification, our MRC-VFC framework surpasses it by a large margin of 11.03% accuracy when the imbalance factor is 500, demonstrating the effectiveness of representation learning and virtual features compensation in our framework. Furthermore, our MRC-VFC framework outperforms FS [12], which improves the resampling strategy and achieves the best performance on the *ISIC-2019-LT* dataset, with an accuracy increase of 9.4% under imbalance factor = 500. These experimental results demonstrate the superiority of our MRC-VFC framework over existing approaches in long-tailed medical image classification tasks.

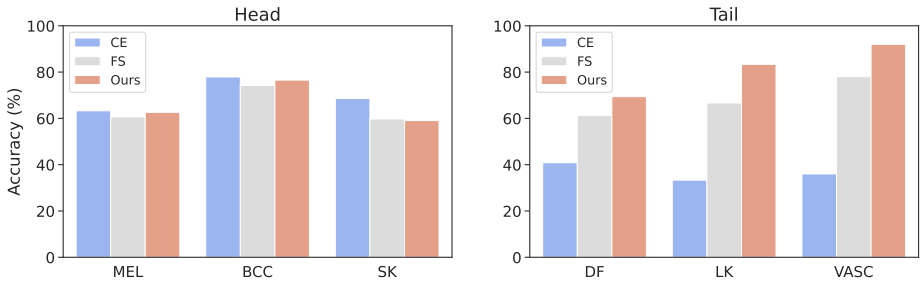

Fig. 2. The performance comparison of head/tail classes in *ISIC-Archive-LT* dataset.

Ablation Study. We perform the ablation study to validate the effectiveness of our proposed MRC and VFC modules on two long-tailed datasets. As shown in Table 1 and 2, both MRC and VFC modules remarkably improve the performance over the baselines. In particular, we apply two ablative baselines of the proposed MRC-VFC framework by disabling the MRC (denoted as Ours w/o MRC) and the VFC (denoted as Ours w/o VFC) individually. In detail, as shown in Table 1, when the imbalance factor is 500, the accuracy increases by 4.49% and 7.14% for MRC and VFC, respectively. In addition, as illustrated in Table 2, the mean accuracy of all classes in the *ISIC-Archive-LT* shows an improvement of 2.40% and 2.92% for MRC and VFC correspondingly. The ablation study verifies the effectiveness of our MRC and VFC modules.

3.4 Comparison on ISIC-Archive-LT Dataset

To comprehensively evaluate our MRC-VFC framework, we further perform the comparison with state-of-the-art algorithms on a more challenging *ISIC-Archive-LT* dataset for long-tailed diagnosis. As illustrated in Table 2, our MRC-VFC framework achieves the best overall performance with an accuracy of 67.84% among state-of-the-art algorithms, and results in a balanced performance over different classes, *i.e.*, 69.71% for head classes and 70.34% for tail classes. Compared with the advanced decoupling method [15] for medical image diagnosis, our MRC-VFC framework significantly improves the accuracy with 4.73% in medium classes and 8.87% in tail classes, respectively.

Performance Analysis on Head/Tail Classes. We further present the performance of several head and tail classes in Fig. 2. Our MRC-VFC framework outperforms FS [12] on both tail and head classes, and significantly promotes the performance of tail classes, thereby effectively alleviating the affect of long-tailed problems on medical image diagnosis. These comparisons confirm the advantage of our MRC-VFC framework in more challenging long-tailed scenarios.

4 Conclusion

To address the long-tails in computer-aided diagnosis, we propose the MRC-VFC framework to improve medical image classification with balanced perfor-

mance in two stages. In the first stage, we design the MRC to facilitate the representation learning of the encoder by introducing multi-view relation-aware consistency. In the second stage, to recalibrate the classifier, we propose the VFC to train an unbias classifier for the MRC-VFC framework by generating massive virtual features. Extensive experiments on the two long-tailed dermatology datasets demonstrate the effectiveness of the proposed MRC-VFC framework, which outperforms state-of-the-art algorithms remarkably.

Acknowledgments. This work was supported in part by the InnoHK program.

References

1. Ahrendt, P.: The multivariate gaussian probability distribution. Technical University of Denmark, Tech. Rep p. 203 (2005)
2. Buda, M., Maki, A., Mazurowski, M.A.: A systematic study of the class imbalance problem in convolutional neural networks. Neural networks **106**, 249–259 (2018)
3. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. Information **11**(2) (2020)
4. Cao, K., Wei, C., Gaidon, A., Arechiga, N., Ma, T.: Learning imbalanced datasets with label-distribution-aware margin loss. NeurIPS **32** (2019)
5. Chen, Z., Guo, X., Woo, P.Y., Yuan, Y.: Super-resolution enhanced medical image diagnosis with sample affinity interaction. IEEE Transactions on Medical Imaging **40**(5), 1377–1389 (2021)
6. Chen, Z., Guo, X., Yang, C., Ibragimov, B., Yuan, Y.: Joint spatial-wavelet dual-stream network for super-resolution. In: MICCAI. pp. 184–193. Springer (2020)
7. Chen, Z., Yang, C., Zhu, M., Peng, Z., Yuan, Y.: Personalized retrogress-resilient federated learning toward imbalanced medical data. IEEE Transactions on Medical Imaging **41**(12), 3663–3674 (2022)
8. Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S.: Class-balanced loss based on effective number of samples. In: CVPR. pp. 9268–9277 (2019)
9. De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., et al.: Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature medicine **24**(9), 1342–1350 (2018)
10. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
12. Ju, L., Wu, Y., Wang, L., Yu, Z., Zhao, X., Wang, X., Bonnington, P., Ge, Z.: Flexible sampling for long-tailed skin lesion classification. In: MICCAI. pp. 462–471. Springer (2022)
13. Kang, B., Li, Y., Xie, S., Yuan, Z., Feng, J.: Exploring balanced feature spaces for representation learning. In: ICLR (2021)
14. Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J., Kalantidis, Y.: Decoupling representation and classifier for long-tailed recognition. ICLR (2020)

15. Li, J., Chen, G., Mao, H., Deng, D., Li, D., Hao, J., Dou, Q., Heng, P.A.: Flat-aware cross-stage distilled framework for imbalanced medical image classification. In: MICCAI. pp. 217–226. Springer (2022)
16. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV. pp. 2980–2988 (2017)
17. Liu, J., Sun, Y., Han, C., Dou, Z., Li, W.: Deep representation learning on long-tailed data: A learnable embedding augmentation perspective. In: CVPR. pp. 2970–2979 (2020)
18. Liu, Z., Miao, Z., Zhan, X., Wang, J., Gong, B., Yu, S.X.: Large-scale long-tailed recognition in an open world. In: CVPR. pp. 2537–2546 (2019)
19. Luo, M., Chen, F., Hu, D., Zhang, Y., Liang, J., Feng, J.: No fear of heterogeneity: Classifier calibration for federated learning with non-iid data. NeurIPS **34**, 5972–5984 (2021)
20. Moon, T.K.: The expectation-maximization algorithm. IEEE Signal processing magazine **13**(6), 47–60 (1996)
21. More, A.: Survey of resampling techniques for improving classification performance in unbalanced datasets. arXiv preprint arXiv:1608.06048 (2016)
22. 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)
23. Srinidhi, C.L., Ciga, O., Martel, A.L.: Deep neural network models for computational histopathology: A survey. Medical Image Analysis **67**, 101813 (2021)
24. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
25. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**(1), 1–9 (2018)
26. Wang, J., Zhang, W., Zang, Y., Cao, Y., Pang, J., Gong, T., Chen, K., Liu, Z., Loy, C.C., Lin, D.: Seesaw loss for long-tailed instance segmentation. In: CVPR. pp. 9695–9704 (2021)
27. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
28. Zhang, Y., Kang, B., Hooi, B., Yan, S., Feng, J.: Deep long-tailed learning: A survey. arXiv preprint arXiv:2110.04596 (2021)
29. Zhang, Y., Wei, X.S., Zhou, B., Wu, J.: Bag of tricks for long-tailed visual recognition with deep convolutional neural networks. In: AAAI. vol. 35, pp. 3447–3455 (2021)
30. Zhang, Z., Sabuncu, M.: Generalized cross entropy loss for training deep neural networks with noisy labels. NeurIPS **31** (2018)
31. Zhou, B., Cui, Q., Wei, X.S., Chen, Z.M.: Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In: CVPR. pp. 9719–9728 (2020)
32. Zhou, X., Liu, X., Wang, C., Zhai, D., Jiang, J., Ji, X.: Learning with noisy labels via sparse regularization. In: ICCV. pp. 72–81 (2021)