Fighting Class Imbalance with Contrastive Learning

Yassine Marrakchi[1,2(✉)], Osama Makansi[1], and Thomas Brox[1,2]

[1] Department of Computer Science, University of Freiburg, Freiburg im Breisgau, Germany
{marrakch,makansio,brox}@cs.uni-freiburg.de
[2] CIBSS – Centre for Integrative Biological Signalling Studies, University of Freiburg, Freiburg im Breisgau, Germany

Abstract. Medical image datasets are hard to collect, expensive to label, and often highly imbalanced. The last issue is underestimated, as typical average metrics hardly reveal that the often very important minority classes have a very low accuracy. In this paper, we address this problem by a feature embedding that balances the classes using contrastive learning as an alternative to the common cross-entropy loss. The approach is largely orthogonal to existing sampling methods and can be easily combined with those. We show on the challenging ISIC2018 and APTOS2019 datasets that the approach improves especially the accuracy of minority classes without negatively affecting the majority ones.

Keywords: Imbalance classification · Medical imaging · Contrastive learning

1 Introduction

Convolutional networks (CNN) have much to offer for computer-aided diagnostics (CAD) as they can potentially bring cheap pre-screening to people, who do not have regular access to medical experts, or they can decrease the screening intervals even for those who have this access. Several domain specific issues in medical image processing, like data scarcity, noisy labels, and low image quality, have been addressed. Another issue, yet with less attention for far, is the often imbalanced distribution of classes in medical datasets. Some classes are much more common than others, hence it is difficult to collect a dataset where all classes are represented equally. For example, among retinal diseases, diabetic retinopathy is more common than fundus pulverulentus and fundus albipunctatus [41]. In Fig. 1, we show the histogram of the classes in the ISIC dataset, a dermatoscopic dataset of common pigmented skin lesions, where few major classes have orders of magnitude larger frequency than others. Neural networks

This study was supported by the Excellence Strategy of the German Federal and State Governments, (CIBSS - EXC 2189).

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12903, pp. 466–476, 2021.
https://doi.org/10.1007/978-3-030-87199-4_44

trained on imbalanced datasets tend to perform worse on the minority classes – a problem that is well-known in machine learning and has been studied thoroughly in the last few years.

Few recent works in the medical domain apply the common technique of resampling the training data [44] or reweighting the loss function to give more attention to minority classes [2,44], hence improving their accuracy. Despite the marginal success of these methods, a thorough analysis of the problem and an effective approach towards a solution is still missing. In addition to the imbalance problem, medical image classification typically focuses on subtle, fine-grained differences. In Fig. 1, we show three samples from different classes which are hard to distinguish due to their subtle differences.

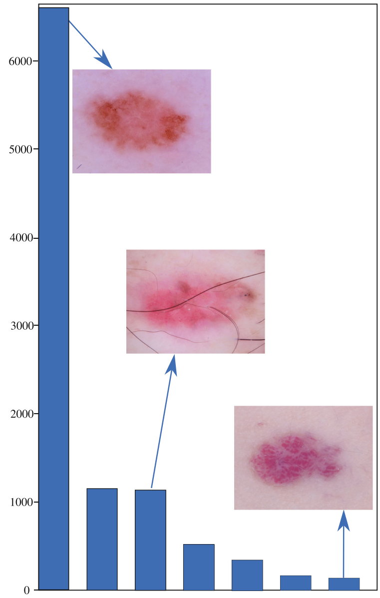

Fig. 1. Histogram of the ISIC dataset showing the majority (left) and minority classes (right). We also show a sample from three classes. These look very similar, yet should be classified differently.

In this paper, we address the above-mentioned problems by explicitly separating the feature space into different clusters by minimizing the distance between samples from the same class (intra-class) and maximizing the distance between samples from different classes (inter-class). We achieve this using a contrastive loss on the learned feature space. With this approach, the minority samples receive enough attention without negatively affecting performance on majority classes.

In summary, we (1) emphasize the issue of imbalanced datasets in the medical domain, (2) propose a framework based on contrastive learning to better arrange the feature space for minority and majority classes, (3) show quantitatively on the challenging ISIC2018 [7] and APTOS2019 [1] datasets that our approach outperforms or performs on parwith existing techniques on all metrics, and (4) discuss the complementary of our method to existing techniques (resampling).

2 Related Work

Classification in Medical Imaging. Disease diagnosis and disease grading are major applications of computer-aided diagnosis that benefited from the flourishing era of medical image analysis [9]. The recent improvement is mainly due to the emergence of deep learning techniques at large scale as recent CAD methods are driven by learning-based classification methods employing neural networks [4]. Although adopting standard methods, such as finetuning pretrained networks, using ensembles and heavy data augmentation, led to decent improvement, more pronounced and domain specific issues have been tackled to gain further quantitative improvement. For instance, fighting against hard examples [14], inter-class correlation [42], and small inter-class differences [45], domain shift [13], catastrophic forgetting [29] and annotation scarcity [17,28] have been deeply studied in the last year and several new methods based on semi-supervised learning and meta-learning have been introduced. Unlike data for common diseases, which is available in large amounts with typically clean labels, it is hard to collect as much annotated data for rare diseases. As data for majority classes is not supposed to be thrown away (a.k.a. undersampling), this leads to a systematic data imbalance. Hence, training a neural network to classify all diseases at comparable accuracy is challenging, and so far, has seen very little attention in the medical imaging community. For instance, the winners of the ISIC-2018 challenge handled the data imbalance with a simple loss reweighting scheme [2,44]. The problem has been also implicitly tackled by splitting the data into two subsets consisting of majority classes and minority classes [28]. After training the network on the majority classes, it is finetuned to recognize the minority classes as well.

Learning from Imbalanced Datasets. There is a long line of works addressing the task of learning from datasets with class-imbalance. The most common technique is to manipulate the training data distribution by oversampling the data from minority classes [36,38] or undersampling the majority classes [12,20]. To avoid potential overfitting on the minority, recent methods proposed to generate new data from minority classes by simulation [5,18,26,33]. Instead of changing the data distribution, other works proposed to introduce a weight in the loss of samples based on their class frequency [8,22,31,39] or optimize a loss that is not sensitive to the class imbalance [15,35]. Recently, Kang et al. [23] proposed the two-stage approach of imbalanced feature learning and balanced classifier learning. In the latter, they use the common oversampling technique [38] for training the classifier. Others proposed to shape the feature space by explicitly designing a loss function to increase the inter-class distance while decreasing the intra-class one [22,43]. Recent works follow up on this idea by explicitly enlarging the margin for the minority to boost their performance [3,10,19,24]. We compare to the most relevant of these recent methods in Sect. 4.4.

Contrastive Learning. The concept of contrastive learning was first proposed to learn better feature representation in a self-supervised manner and was referred to as noise-contrastive learning [6,11,16,34]. In particular, for any input sample, a set of positives and negatives are constructed and the loss function is optimized to separate them. The positives are noisy versions of the input sample (e.x, augmentation) and the negatives are the other samples in the batch. Recently, [25] extended the positives to also include images from the same class yielding impressive results in image classification on ImageNet [37]. Despite the success of contrastive learning in both self-supervised and supervised settings, we are the first to adapt this concept in the task of learning from imbalance datasets.

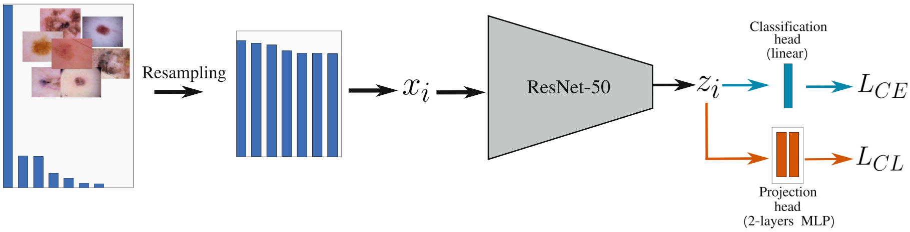

Fig. 2. Two-stages Framework. In the first stage, we learn the feature space via the contrastive loss (orange) by projecting the feature vector using 2-layers MLP. In the second stage, we through the projection head, freeze the backbone and only learn the classification head (blue). (Color figure online)

3 Supervised Representation Disentanglement via Contrastive Learning

Representation learning aims to disentangle the feature space, such that the resulting embedding can be mapped well to classes. However, when datasets are highly imbalanced, networks focus on the majority classes, as they dominate the loss. In order to overcome this limitation, we propose to separate the minority from the majority classes in feature space with contrastive learning. In particular, for every sample x_i in the batch, its positive samples, i.e., samples from the same class in the batch, are pushed closer while the negatives ones, i.e. all samples from different classes in the batch, are pushed away. Formally:

$$L_i^{\mathrm{CL}} = -\frac{1}{N_{\mathbf{c}_i} - 1} \sum_{j \in \mathbf{c}_i} \mathbb{1}_{i \neq j} \cdot \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_j / \tau)}{\sum_{k=1}^{N} \mathbb{1}_{i \neq k} \cdot \exp(\mathbf{z}_i \cdot \mathbf{z}_k / \tau)}, \tag{1}$$

where $\mathbf{z}_i$ is the normalized feature vector of an image x_i, $\mathbf{c}_i$ is the positive set for sample i and corresponds to all samples in the batch with the same class label as i. $\mathbb{1}_{cond}$ is the indicator function that returns 1 if the condition *cond* returns true and 0 otherwise. N is the total number of samples in the batch, and $N_{\mathbf{c}_i}$ the size of the positive set of sample x_i. $\tau > 0$ is the temperature parameter, which controls the balance between positives and negatives. In the standard setting, cross-entropy is a function of the pseudo-probabilities generated by the Softmax function and the labels for each input image separately. Thus the resulting gradients for each single input depend solely on the considered image. Given the data unbalance, the network will prioritize learning major classes. Using contrastive loss mitigates this effect by computing the pseudo-probabilities based on the cosine similarity between feature embeddings of the full batch. Consequently, gradients depend on the feature embedding of all batch elements and minimizing the loss **explicitly** moves feature embeddings to form clusters.

After learning the feature space using the contrastive loss, a second stage is needed to learn to map from the feature space to the target class labels. To this end, we use the common cross-entropy loss:

Table 1. Quantitative evaluation on the ISIC2018 dataset. Our method using contrastive learning yields the best performance. Results denoted with † are taken from the respective papers.

	Accuracy	F-score
CE	$0.850 \pm 0.7e-3$	$0.716 \pm 6e-3$
Focal Loss [31]	$0.849 \pm 4.5e-3$	$0.728 \pm 5e-3$
LDAM [3]	$0.857 \pm 9.0e-3$	$0.734 \pm 24e-3$
CL (ours)	$\mathbf{0.865} \pm 4.0e-3$	$\mathbf{0.739} \pm 13e-3$
CE + resample	$0.861 \pm 0.9e-3$	$0.735 \pm 6e-3$
CL (ours) + resample	$\mathbf{0.868} \pm 7.0e-3$	$\mathbf{0.751} \pm 16e-3$
OHEM [40]	$0.818^\dagger$	$0.660^\dagger$
MTL [30]	$0.811^\dagger$	$0.667^\dagger$
DANIL [14]	$0.825^\dagger$	$0.674^\dagger$

$$L_i^{\mathrm{CE}} = -\sum_{j=1}^{M} y_{i,j} \log p_{i,j} \, , \tag{2}$$

where M is the number of classes, $y_{i,j}$ is an indicator if j is the class label for the sample i, and $p_{i,j}$ is the predicted probability that the sample i is classified as class j. When training the second stage, we freeze the learned backbone and only learn the classification head as illustrated in Fig. 2. As the cross-entropy loss is not sensitive to the class distribution, it is important to sample the classes uniformly across batches as suggested by [23].

In Sect. 4.4, we compare and combine our approach with an oversampling strategy during the backbone training. We follow a simple oversampling strategy by extending the original data set with copies of samples from minority classes so that the artificially balanced dataset has the exact same number of items per class. For the second stage we always use artificially balanced data by following this simple scheme.

4 Experiments

4.1 Experimental Setup

We evaluate the proposed method on the ISIC2018 lesion diagnosis dataset [7] which consists of 10015 skin lesion images and 7 predefined categories and APTOS2019 [1] for diabetic retinopathy which has 5 classes and 3662 images. We split images randomly to a train and test set with a ratio of 7:3 as in [14]. Beside the average accuracy which is very sensitive to data imbalance, we report the F-score (also known as Dice similarity coefficient) which is the average of the classwise harmonic mean of precision and recall. Since the classwise F-score is normalized, this metric is particularly sensitive to the performance on minority classes. To evaluate the stability of our method, we report the means and standard deviations over 3 independent runs.

Table 2. Quantitative evaluation on the APTOS2019 dataset. Contrastive learning clearly outperforms the CE baseline and is on-par with the state of the art on this dataset. Results denoted with † are taken from the respective papers.

	Accuracy	F-score
CE	$0.812 \pm 7e-3$	$0.608 \pm 18e-3$
Focal Loss [31]	$0.815 \pm 1e-3$	$0.629 \pm 8e-3$
LDAM [3]	$0.813 \pm 3e-3$	$0.620 \pm 5e-3$
CL (ours)	$\mathbf{0.825} \pm 1e-3$	$\mathbf{0.652} \pm 3e-3$
CE + resample	$0.802 \pm 23e-3$	$0.583 \pm 55e-3$
CL (ours) + resample	$0.816 \pm 1e-3$	$0.608 \pm 4e-3$
CANet [27]	0.813 †	0.631 †
OHEM [40]	0.813 †	0.632 †
DANIL [14]	**0.825** †	**0.660** †

4.2 Baselines

We compare the proposed method to previous works that use the same experimental setup: **OHEM** [40]: a hard example mining method that samples training images according to a non-uniform distribution based on their current loss. **MTL** [30]: a deep multi-task learning framework that optimizes body location classification along with the skin lesion classification. **CANet** [27]: a CNN with an attention module to learn disease-specific features and a disease-dependent attention module to learn internal relationships between pairs of diseases. **DANIL** [14]: a method that synthesizes distractors in the feature space to learn stronger features. We also compare to popular model-based methods to learn from unbalanced data. Beside **focal loss** [31], a sophisticated way to balance the loss depending on the class frequency, we consider **LDAM** [3], a label-distribution-aware margin loss that applies stronger regularization to minority classes to improve the generalization of the model. We further investigate the effect of **Resampling**, which ensures having the same number of images per class per epoch by sampling images from the minority classes multiple times within the same epoch. We apply this method in combination with both the standard network training and with contrastive learning.

4.3 Implementation Details

Following the default data augmentation policy for training ResNet [21] on Imagenet [37], we use scaling, random flipping, color distortion, normalization and random cropping in training. We use Resnet50 as backbone and append a 2-layer perceptron of size 2048 and 128 respectively to apply the contrastive loss. The network is trained with SGD optimizer where the learning rate is initialized to 10^{-1} and decayed to 10^{-4} using the cosine schedule without restart [32]. The weight decay is set to 10^{-4}. For networks without data resampling and ISIC2018

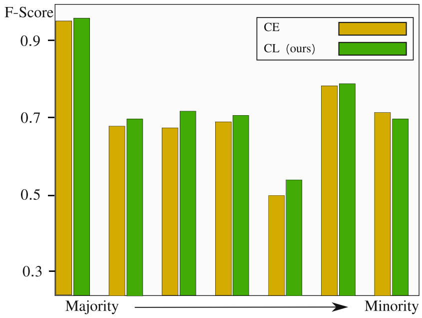

Fig. 3. F-score evaluation of our approach and the common cross-entropy baseline across different classes. This shows that our contrastive-based approach yields larger improvements on the minority (right) while the performance on the major class is not affected.

datasets, we train the backbone for 1000 epochs. For sake of fairness, we train networks with data resampling for 250 epochs only so that in both cases we run roughly 40k updates. For APTOS2019, we double the number of epochs. It takes about 8 h to train the backbone on 4 NVIDIA TITAN RTX GPUs with a batchsize of 192. Training the classification head follows the same settings except the batch size which is set to 512 and the learning rate which decays from 1 to 10^{-4}. This step takes less than an hour to train on a single GPU.

For testing, we resize the images to 256×256 and then process the normalized image. The code is implemented using PyTorch v1.5.1 and will be publicly available upon acceptance.

4.4 Results and Discussion

Tables 1 and 2 summarize the quantitative evaluation of our approach and several baselines. First we show a comparison of different model-based approaches for addressing the imbalance classification (top part). Here we clearly show that our approach based on contrastive learning is consistently superior to other techniques due to the explicit disentanglement of the feature space on both datasets. Next, we study the effect of the common data-based method (resampling) in conjunction with contrastive learning (middle part) and find that when resampling improves over standard cross-entropy, it is complementary to our approach and further improves its performance as in Table 1. In case resampling harms the performance, as in Table 2, contrastive learning helps stabilize the learning process and leads to clearly lower standard deviation. We think the drop in performance on APTOS19 is mainly due to the small size of the dataset. Thus, making copies from the very few samples results in overfitting to those. Finally, we compare to the best existing methods (lower part) and show that contrastive learning performance is on-par or even significantly better than those works without using their special modifications to standard classification. The gap in performance on ISIC shows that pretraining on ImageNet [37], used as

Table 3. Effect of contrastive loss temperature on the f-score of different classes without/with resampling

Temperature	Major	Middle	Minor
0.01	0.941/0.939	0.702/0.712	0.692/0.710
0.05	0.945/0.938	0.717/0.711	0.713/0.721
0.1	0.942/0.937	0.716/0.711	0.696/0.709
0.5	0.942/0.937	0.696/0.698	0.575/0.720
1.0	0.941/0.934	0.669/0.623	0.531/0.651

a common practice in existing methods, does not necessarily yield improvement due to the large domain shift. Figure 3 shows a break-down of the performance of our method and the cross-entropy baseline across all classes. This shows that the main source of improvement is due to better handling the minority while not affecting the performance of the major classes.

4.5 Ablation Study

We ran experiments on ISIC data to study the effect of the temperature used in contrastive loss without/with resampling. We let the temperature take values in $\{0.01, 0.05, 0.1, 0.5, 1.0\}$ for our experiments and report the averaged F-scores for 3 groups of classes: the major class ($\geq$6000 images), the 2 medium size classes ($\sim$1100 images) and the 4 minor classes ($\leq$550 images). Recall that the temperature controls the balance between positives and negatives and higher temperatures put more attention to pushing positives close to each other while lower temperatures focus more on pushing negatives away from each other which should theoretically lead to better performance on minority classes for lower temperatures and more stability using the resampling than without resampling. Experimentally, we show in Table 3 that the performance on the major class stays almost constant for different temperatures. We also notice a significant drop in performance for both middle and minor classes for increasing temperatures, which matches the theoretical explanation. We find out that the choice of the temperature for the contrastive loss with resampling is less critical as we recorded non-significant fluctuations in all classes for all temperatures in $\{0.01, 0.05, 0.1, 0.5\}$ while setting the temperature to 0.05 gives a performance boost for the plain contrastive learning. This fact aligns well with our expectation on the stability of both versions. Finally, we record a drop in performance for contrastive learning in conjunction with the lowest temperature. This drop can be explained by an excessive focus on pushing negatives away from each other, which slows down the learning process.

5 Conclusion

In this paper we highlighted the importance of learning from imbalanced datasets in medical image analysis. We proposed a new approach based on contrastive learning to better separate minority from majority classes in feature space. The approach consistently improved over the cross-entropy and oversampling baselines in our evaluation. Moreover, we showed that it is complementary to oversampling and sets a new state of the art for imbalanced datasets.

References

1. Aptos 2019 blindness detection (2019). https://www.kaggle.com/c/aptos2019-blindness-detection/data
2. Aleksey Nozdryn-Plotnicki, J.Y., Yolland, W.: Ensembling convolutional neural networks for skin cancer classification. ArXiv (2018)
3. Cao, K., Wei, C., Gaidon, A., Aréchiga, N., Ma, T.: Learning imbalanced datasets with label-distribution-aware margin loss. In: NeurIPS (2019)
4. Chan, H.P., Samala, R.K., Hadjiiski, L.M., Zhou, C.: Deep learning in medical image analysis (2020)
5. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. JAIR **16**, 321–357 (2002)
6. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.E.: A simple framework for contrastive learning of visual representations. In: ICML (2020)
7. Codella, N.C.F., et al.: Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC). CoRR abs/1902.03368 (2019)
8. Cui, Y., Jia, M., Lin, T., Song, Y., Belongie, S.J.: Class-balanced loss based on effective number of samples. In: CVPR (2019)
9. Doi, K.: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput. Med. Imaging Graph. **31**, 198–211 (2007)
10. Dong, Q., Gong, S., Zhu, X.: Imbalanced deep learning by minority class incremental rectification. IEEE TPAMI **41**, 1367–1381 (2019)
11. Dosovitskiy, A., Fischer, P., Springenberg, J.T., Riedmiller, M.A., Brox, T.: Discriminative unsupervised feature learning with exemplar convolutional neural networks. IEEE TPAMI, 1734–1747 (2016)
12. Drummond, C., Holte, R.: C4.5, class imbalance, and cost sensitivity: why undersampling beats oversampling. In: ICML Workshop (2003)
13. Fotedar, G., Tajbakhsh, N., Ananth, S., Ding, X.: Extreme consistency: overcoming annotation scarcity and domain shifts. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 699–709. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_68
14. Gong, L., Ma, K., Zheng, Y.: Distractor-aware neuron intrinsic learning for generic 2D medical image classifications. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 591–601. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_57
15. Guanjin Wang, K.W.W., Lu, J.: AUC-based extreme learning machines for supervised and semi-supervised imbalanced classification. IEEE Trans. Syst. Man Cybern.: Syst., 1–12 (2020)
16. Gutmann, M., Hyvärinen, A.: Noise-contrastive estimation: a new estimation principle for unnormalized statistical models. In: AISTATS (2010)
17. Gyawali, P.K., Ghimire, S., Bajracharya, P., Li, Z., Wang, L.: Semi-supervised medical image classification with global latent mixing. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 604–613. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_59

18. Han, H., Wang, W.-Y., Mao, B.-H.: Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 878–887. Springer, Heidelberg (2005). https://doi.org/10.1007/11538059_91
19. Hayat, M., Khan, S., Zamir, S.W., Shen, J., Shao, L.: Gaussian affinity for max-margin class imbalanced learning. In: ICCV (2019)
20. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE TKDE **21**, 1263–1284 (2009)
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
22. Huang, C., Li, Y., Loy, C.C., Tang, X.: Learning deep representation for imbalanced classification. In: CVPR (2016)
23. Kang, B., et al.: Decoupling representation and classifier for long-tailed recognition. In: ICLR (2020)
24. Khan, S.H., Hayat, M., Zamir, S.W., Shen, J., Shao, L.: Striking the right balance with uncertainty. In: CVPR (2019)
25. Khosla, P., et al.: Supervised contrastive learning. In: NeurIPS (2020)
26. Kim, J., Jeong, J., Shin, J.: M2m: imbalanced classification via major-to-minor translation. In: CVPR (2020)
27. Li, X., Hu, X., Yu, L., Zhu, L., Fu, C.W., Heng, P.: CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. IEEE Trans. Med. Imaging, 1483–1493 (2020)
28. Li, X., Yu, L., Jin, Y., Fu, C.-W., Xing, L., Heng, P.-A.: Difficulty-aware meta-learning for rare disease diagnosis. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 357–366. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_35
29. Li, Z., Zhong, C., Wang, R., Zheng, W.-S.: Continual learning of new diseases with dual distillation and ensemble strategy. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 169–178. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_17
30. Liao, H., Luo, J.: A deep multi-task learning approach to skin lesion classification. In: AAAI workshop (2017)
31. Lin, T., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV (2017)
32. Loshchilov, I., Hutter, F.: SGDR: stochastic gradient descent with warm restarts. In: ICLR (2017)
33. Mullick, S.S., Datta, S., Das, S.: Generative adversarial minority oversampling. In: ICCV (2019)
34. van den Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. CoRR abs/1807.03748 (2018)
35. Peilin Zhao, Steven C. H. Hoi, R.J., Yang, T.: Online AUC maximization. In: ICML (2011)
36. Peng, J., Bu, X., Sun, M., Zhang, Z., Tan, T., Yan, J.: Large-scale object detection in the wild from imbalanced multi-labels. In: CVPR (2020)
37. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**, 211–252 (2015)
38. Shen, L., Lin, Z., Huang, Q.: Relay backpropagation for effective learning of deep convolutional neural networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 467–482. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_29

39. Shrivastava, A., Gupta, A., Girshick, R.B.: Training region-based object detectors with online hard example mining. In: CVPR (2016)
40. Shrivastava, A., Gupta, A., Girshick, R: Training region-based object detectors with online hard example mining. In: CVPR (2016)
41. Skorczyk-Werner, A., et al.: Fundus albipunctatus: review of the literature and report of a novel RDH5 gene mutation affecting the invariant tyrosine (p. Tyr175Phe). J. Appl. Genet. **56**, 317–327 (2015)
42. Wei, D., Cao, S., Ma, K., Zheng, Y.: Learning and exploiting interclass visual correlations for medical image classification. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 106–115. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_11
43. Zhang, X., Fang, Z., Wen, Y., Li, Z., Qiao, Y.: Range loss for deep face recognition with long-tailed training data. In: ICCV (2017)
44. Zhuang, J.X., et al.: Skin lesion analysis towards melanoma detection using deep neural network ensemble (2018)
45. Zhuang, J., Cai, J., Wang, R., Zhang, J., Zheng, W.-S.: Deep kNN for medical image classification. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 127–136. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_13