Towards Novel Class Discovery: A Study in Novel Skin Lesions Clustering

Wei Feng[1,2,4,6], Lie Ju[1,2,4,6], Lin Wang[1,2,4,6], Kaimin Song[6], and Zongyuan Ge[1,2,3,4,5](✉)

[1] Faculty of Engineering, Monash University, Melbourne, Australia
zongyuan.ge@monash.edu
[2] Monash Medical AI Group, Monash University, Melbourne, Australia
[3] AIM for Health Lab, Monash University, Melbourne, VIC, Australia
[4] Airdoc-Monash Research Lab, Monash University, Suzhou, China
[5] Faculty of IT, Monash University, Melbourne, VIC, Australia
[6] Airdoc LLC, Beijing, China
https://www.monash.edu/mmai-group

Abstract. Existing deep learning models have achieved promising performance in recognizing skin diseases from dermoscopic images. However, these models can only recognize samples from predefined categories, when they are deployed in the clinic, data from new unknown categories are constantly emerging. Therefore, it is crucial to automatically discover and identify new semantic categories from new data. In this paper, we propose a new novel class discovery framework for automatically discovering new semantic classes from dermoscopy image datasets based on the knowledge of known classes. Specifically, we first use contrastive learning to learn a robust and unbiased feature representation based on all data from known and unknown categories. We then propose an uncertainty-aware multi-view cross pseudo-supervision strategy, which is trained jointly on all categories of data using pseudo labels generated by a self-labeling strategy. Finally, we further refine the pseudo label by aggregating neighborhood information through local sample similarity to improve the clustering performance of the model for unknown categories. We conducted extensive experiments on the dermatology dataset ISIC 2019, and the experimental results show that our approach can effectively leverage knowledge from known categories to discover new semantic categories. We also further validated the effectiveness of the different modules through extensive ablation experiments. Our code will be released soon.

Keywords: Novel Class Discovery · Skin Lesion Recognition · Deep Learning

1 Introduction

Automatic identification of lesions from dermoscopic images is of great importance for the diagnosis of skin cancer [16,22]. Currently, deep learning mod-

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43987-2_3.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 24–33, 2023.
https://doi.org/10.1007/978-3-031-43987-2_3

els, especially those based on deep convolution neural networks, have achieved remarkable success in this task [17,18,22]. However, this comes at the cost of a large amount of labeled data that needs to be collected for each class. To alleviate the labeling burden, semi-supervised learning has been proposed to exploit a large amount of unlabeled data to improve performance in the case of limited labeled data [10,15,19]. However, it still requires a small amount of labeled data for each class, which is often impossible in real practice. For example, there are roughly more than 2000 named dermatological diseases today, of which more than 200 are common, and new dermatological diseases are still emerging, making it impractical to annotate data from scratch for each new disease category [20]. However, since there is a correlation between new and known diseases, a priori knowledge from known diseases is expected to help automatically identify new diseases [9].

One approach to address the above problem is novel class discovery (NCD) [7,9,24], which aims to transfer knowledge from known classes to discover new semantic classes. Most NCD methods follow a two-stage scheme: 1) a stage of fully supervised training on known category data and 2) a stage of clustering on unknown categories [7,9,24]. For example, Han et al. [9] further introduced self-supervised learning in the first stage to learn general feature representations. They also used ranking statistics to compute pairwise similarity for clustering. Zhong et al. [24] proposed OpenMix based on the mixup strategy [21] to further exploit the information from known classes to improve the performance of unsupervised clustering. Fini et al. [7] proposed UNO, which unifies multiple objective functions into a holistic framework to achieve better interaction of information between known and unknown classes. Zhong et al. [23] used neighborhood information in the embedding space to learn more discriminative representations. However, most of these methods require the construction of a pairwise similarity prediction task to perform clustering based on pairwise similarity pseudo labels between samples. In this process, the generated pseudo labels are usually noisy, which may affect the clustering process and cause error accumulation. In addition, they only consider the global alignment of samples to the category center, ignoring the local inter-sample alignment thus leading to poor clustering performance.

In this paper, we propose a new novel class discovery framework to automatically discover novel disease categories. Specifically, we first use contrastive learning to pretrain the model based on all data from known and unknown categories to learn a robust and general semantic feature representation. Then, we propose an uncertainty-aware multi-view cross-pseudo-supervision strategy to perform clustering. It first uses a self-labeling strategy to generate pseudo-labels for unknown categories, which can be treated homogeneously with ground truth labels. The cross-pseudo-supervision strategy is then used to force the model to maintain consistent prediction outputs for different views of unlabeled images. In addition, we propose to use prediction uncertainty to adaptively adjust the contribution of the pseudo labels to mitigate the effects of noisy pseudo labels. Finally, to encourage local neighborhood alignment and further refine the pseudo

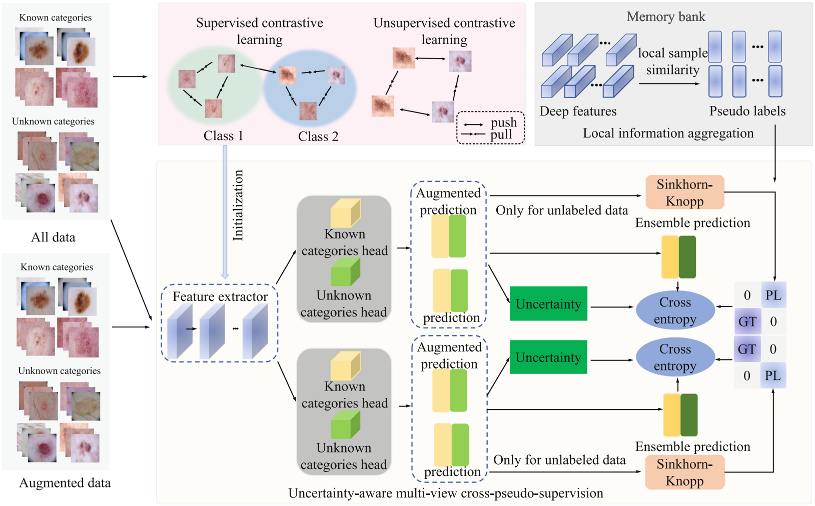

Fig. 1. The overall framework of our proposed novel class discovery algorithm.

labels, we propose a local information aggregation module to aggregate the information of the neighborhood samples to boost the clustering performance. We conducted extensive experiments on the dermoscopy dataset ISIC 2019, and the experimental results show that our method outperforms other state-of-the-art comparison algorithms by a large margin. In addition, we also validated the effectiveness of different components through extensive ablation experiments.

2 Methodology

Given an unlabeled dataset $\{x_i^u\}_{i=1}^{N^u}$ with N^u images, where x_i^u is the ith unlabeled image. Our goal is to automatically cluster the unlabeled data into C^u clusters. In addition, we also have access to a labeled dataset $\{x_i^l, y_i^l\}_{i=1}^{N^l}$ with N^l images, where x_i^l is the ith labeled image and $y_i^l \in \mathcal{Y} = \left\{1, \ldots, C^l\right\}$ is its corresponding label. In the novel class discovery task, the known and unknown classes are disjoint, i.e., $C^l \cap C^u = \varnothing$. However, the known and unknown classes are similar, and we aim to use the knowledge of the known classes to help the clustering of the unknown classes. The overall framework of our proposed novel class discovery algorithm is shown in Fig. 1. Specifically, we first learn general and robust feature representations through contrastive learning. Then, the uncertainty-aware multi-view cross-pseudo-supervision strategy is used for joint training on all category data. Finally, the local information aggregation module benefits the NCD by aggregating the useful information of the neighborhood samples.

Contrastive Learning. To achieve a robust feature representation for the NCD task, we first use noise contrastive learning [8] to pretrain the feature extractor network, which effectively avoids model over-fitting to known categories. Specifically, we use x_i and x_i' to represent different augmented versions of the same image in a mini-batch. The unsupervised contrastive loss can be formulated as:

$$L_i^{ucl} = -\log \frac{\exp\left(z_i \cdot z_i'/\tau\right)}{\sum_n \mathbb{1}_{[n \neq i]} \exp\left(z_i \cdot z_n/\tau\right)} \tag{1}$$

where $z_i = E(x_i)$ is the deep feature representation of the image x_i, E is the feature extractor network, and τ is the temperature value. $\mathbb{1}$ is the indicator function.

In addition, to help the feature extractor learn semantically meaningful feature representations, we introduce supervised contrastive learning [12] for labeled known category data, which can be denoted as:

$$L_i^{scl} = -\frac{1}{|N(i)|} \sum_{q \in N(i)} \log \frac{\exp\left(z_i \cdot z_q/\tau\right)}{\sum_n \mathbb{1}_{[n \neq i]} \exp\left(z_i \cdot z_n/\tau\right)} \tag{2}$$

where $N(i)$ represents the sample set with the same label as x_i in a mini-batch data. $|N(i)|$ represents the number of samples.

The overall contrastive loss can be expressed as: $L_{cl} = (1-\mu) \sum_{i \in B} L_i^{ucl} + \mu \sum_{i \in B_l} L_i^{scl}$, where μ denotes the balance coefficient. B_l is the labeled subset of mini-batch data.

Uncertainty-Aware Multi-view Cross-Pseudo-Supervision. We now describe how to train uniformly on known and unknown categories using the uncertainty-aware multi-view cross-pseudo-supervision strategy. Specifically, we construct two parallel classification models M_1 and M_2, both of them composed of a feature extractor and two category classification heads, using different initialization parameters. For an original image x_i, we generate two augmented versions of x_i, x_i^{v1} and x_i^{v2}. We then feed these two augmented images into M_1 and M_2 to obtain the predictions for x_i^{v1} and x_i^{v2}:

$$p_{i,1}^{v1} = M_1(x_i^{v1}), p_{i,1}^{v2} = M_1(x_i^{v2}), p_{i,2}^{v1} = M_2(x_i^{v1}), p_{i,2}^{v2} = M_2(x_i^{v2}). \tag{3}$$

The prediction outputs are obtained by concatenating the outputs of the two classification heads and then passing a softmax layer [7]. Then, we can compute the ensemble predicted output of M_1 and M_2: $p_i^{M_1} = \left(p_{i,1}^{v1} + p_{i,1}^{v2}\right)/2$, $p_i^{M_2} = \left(p_{i,2}^{v1} + p_{i,2}^{v2}\right)/2$.

Next, we need to obtain training targets for all data. For an input image x_i, if x_i is from the known category, we construct the training target as one hot vector, where the first C^l elements are ground truth labels and the last C^u elements are 0. If x_i is from the unknown category, we set the first C^l elements to 0 and use pseudo labels for the remaining C^u elements.

We follow the self-labeling method in [1,3] to generate pseudo labels. Specifically, the parameters in the unknown category classification head can be viewed as prototypes of each category, and our training goal is to distribute a set of samples uniformly to each prototype while maximizing the similarity between samples and prototypes [1]. Let $\mathbf{P} = \left[p_1^u; \ldots; p_{B_u}^u\right] \in \mathbb{R}^{B_u \times C^u}$ denotes the ensemble prediction of data of unknown categories in a mini-batch, where B_u represents the number of samples. Here we only consider the output of the unknown categories head due to the samples coming from unknown categories [7]. We obtain the pseudo label by optimizing the following objective:

$$\max_{\mathbf{Y} \in \mathcal{S}} \operatorname{tr}\left(\mathbf{Y}\mathbf{P}^\top\right) + \delta H(\mathbf{Y}) \tag{4}$$

where $\mathbf{Y} = \left[y_1^u; \ldots; y_{B_u}^u\right] \in \mathbb{R}^{B_u \times C^u}$ will assign B_u unknown category samples to C^u category prototypes uniformly, i.e., each category prototype will be selected B_u/C^u times on average. $\mathcal{S}$ is the search space. H is the entropy function used to control the smoothness of $\mathbf{Y}$. δ is the hyperparameter. The solution to this objective can be calculated by the Sinkhorn-Knopp algorithm [6]. After generating the pseudo-labels, we can combine them with the ground truth labels of known categories as training targets for uniform training.

To mitigate the effect of noisy pseudo labels, we propose to use prediction uncertainty [14] to adaptively adjust the weights of pseudo labels. Specifically, we first compute the variance of the predicted outputs of the models for the different augmented images via KL-divergence:

$$V_1 = E\left[p_{i,1}^{v1} \log\left(\frac{p_{i,1}^{v1}}{p_{i,1}^{v2}}\right)\right], V_2 = E\left[p_{i,2}^{v1} \log\left(\frac{p_{i,2}^{v1}}{p_{i,2}^{v2}}\right)\right], \tag{5}$$

where E represents the expected value. If the variance of the model's predictions for different augmented images is large, the pseudo label may be of low quality, and vice versa. Then, based on the prediction variance of the two models, the multi-view cross-pseudo supervision loss can be formulated as:

$$L_{cps} = E\left[e^{-V_1} L_{ce}\left(p^{M_2}, y^{v1}\right) + V_1\right] + E\left[e^{-V_2} L_{ce}\left(p^{M_1}, y^{v2}\right) + V_2\right] \tag{6}$$

where L_{ce} denotes the cross-entropy loss. y^{v1} and y^{v2} are the training targets.

Local Information Aggregation. After the cross-pseudo-supervision training described above, we are able to assign the instances to their corresponding clustering centers. However, it ignores the alignment between local neighborhood samples, i.e., the samples are susceptible to interference from some irrelevant semantic factors such as background and color. Here, we propose a local information aggregation to enhance the alignment of local samples. Specifically, as shown in Fig. 1, we maintain a first-in-first-out memory bank $\mathcal{M} = \{z_k^m, y_k^m\}_{k=1}^{N^m}$ during the training process, which contains the features of N^m most recent samples and their pseudo labels. For each sample in the current batch, we compute the similarity between its features and the features of each sample in the memory bank:

$$d_k = \frac{\exp\left(z \cdot z_k^m\right)}{\sum_{k=1}^{N^m} \exp\left(z \cdot z_k^m\right)}. \quad (7)$$

Then based on this feature similarity, we obtain the final pseudo labels as: $y^u = \rho y^u + (1-\rho) \sum_{k=1}^{N^m} d_k y_k^m$, where ρ is the balance coefficient. By aggregating the information of the neighborhood samples, we are able to ensure consistency between local samples, which further improves the clustering performance.

3 Experiments

Dataset. To validate the effectiveness of the proposed algorithm, we conduct experiments on the widely used public dermoscopy challenge dataset ISIC 2019 [4,5]. The dataset contains a total of 25,331 dermoscopic images from eight categories: Melanoma (MEL), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Actinic Keratosis (AK), Benign Keratosis (BKL), Dermatofibroma (DF), Vascular Lesion (VASC), and Squamous Cell Carcinoma (SCC). Since the dataset suffers from severe category imbalance, we randomly sampled 500 samples from those major categories (MEL, NV, BCC, BKL) to maintain category balance. Then, we construct the NCD task where we treat 50% of the categories (AK, MEL, NV, BCC) as known categories and the remaining 50% of the categories (BKL, SCC, DF, VASC) as unknown categories. We also swap the known and unknown categories to form a second NCD task. For task 1 and task 2, we report the average performance of 5 runs.

Implementation Details. We used ResNet-18 [11] as the backbone of the classification model. The known category classification head is an *l2*-normalized linear classifier with C^l output units. The unknown category classification head consists of a projection layer with 128 output units, followed by an *l2*-normalized linear classifier with C^u output units. In the first contrastive learning pre-training step, we used SGD optimizer to train the model for 200 epochs and gradually decay the learning rate starting from 0.1 and dividing it by 5 at the epochs 60, 120, and 180. μ is set to 0.5, τ is set to 0.5. In the joint training phase, we fix the parameters of the previous feature extractor and only fine-tune the parameters of the classification head. We use the SGD optimizer to train the model for 200 epochs with linear warm-up and cosine annealing ($lr_{\text{base}} = 0.1$, $lr_{\text{min}} = 0.001$), and the weight decay is set to 1.5×10^{-4}. For data augmentation, we use random horizontal/vertical flipping, color jitter, and Gaussian blurring following [7]. For pseudo label, we use the Sinkhorn-Knopp algorithm with hyperparameters inherited from [7]: $\delta = 0.05$ and the number of iterations is 3. We use a memory bank $\mathcal{M}$ of size 100 and the hyperparameter ρ is set to 0.6. The batch size in all experiments is 512. In the inference phase, we only use the output of the unknown category classification head of M_1 [9]. Following [9,23,24], we report the clustering performance on the unlabeled unknown category dataset. We assume that the number of unknown categories is known and it can also be obtained by the category number estimation method proposed in [9].

Table 1. Clustering performance of different comparison algorithms on different tasks.

Method	Task1			Task2		
	ACC	NMI	ARI	ACC	NMI	ARI
Baseline	0.4685	0.2107	0.1457	0.3899	0.0851	0.0522
RankStats [9]	0.5652	0.2571	0.2203	0.4284	0.1164	0.1023
RankStats+ [9]	0.5845	0.2633	0.2374	0.4362	0.1382	0.1184
OpenMix [24]	0.6083	0.2863	0.2512	0.4684	0.1519	0.1488
NCL [23]	0.5941	0.2802	0.2475	0.4762	0.1635	0.1573
UNO [7]	0.6131	0.3016	0.2763	0.4947	0.1692	0.1796
Ours	**0.6654**	**0.3372**	**0.3018**	**0.5271**	**0.1826**	**0.2033**

Following [2,9], we use the average clustering accuracy (ACC), normalized mutual information (NMI) and adjusted rand index (ARI) to evaluate the clustering performance of different algorithms. Specifically, we first match the clustering assignment and ground truth labels by the Hungarian algorithm [13]. After the optimal assignment is determined, we then compute each metric. We implement all algorithms based on the PyTorch framework and conduct experiments on 8 RTX 3090 GPUs.

Comparison with State-of-the-Art Methods. We compare our algorithms with some state-of-the-art NCD methods, including RankStats [9], RankStats+ (RankStats with incremental learning) [9], OpenMix [24], NCL [23], UNO [7]. we also compare with the benchmark method (Baseline), which first trains a model using known category data and then performs clustering on unknown category data. Table 1 shows the clustering performance of each comparison algorithm on different NCD tasks. It can be seen that the clustering performance of the benchmark method is poor, which indicates that the model pre-trained using only the known category data does not provide a good clustering of the unknown category. Moreover, the state-of-the-art NCD methods can improve the clustering performance, which demonstrates the effectiveness of the currently popular two-stage solution. However, our method outperforms them, mainly due to the fact that they need to generate pairwise similarity pseudo labels through features obtained based on self-supervised learning, while ignoring the effect of noisy pseudo labels. Compared with the best comparison algorithm UNO, our method yields 5.23% ACC improvement, 3.56% NMI improvement, and 2.55% ARI improvement on Task1, and 3.24% ACC improvement, 1.34% NMI improvement, and 2.37% ARI improvement on Task2, which shows that our method is able to provide more reliable pseudo labels for NCD.

Ablation Study of Each Key Component. We performed ablation experiments to verify the effectiveness of each component. As shown in Table 2, CL is contrastive learning, UMCPS is uncertainty-aware multi-view cross-pseudo-supervision, and LIA is the local information aggregation module. It can be observed that CL brings a significant performance gain, which indicates that

Table 2. Ablation study of each key component.

Method			Task1			Task2		
CL	UMCPS	LIA	ACC	NMI	ARI	ACC	NMI	ARI
✗	✗	✗	0.4685	0.2107	0.1457	0.3899	0.0851	0.0522
✓			0.5898	0.2701	0.2375	0.4402	0.1465	0.1322
✓	✓		0.6471	0.3183	0.2821	0.5012	0.1732	0.1851
✓		✓	0.6255	0.3122	0.2799	0.4893	0.1688	0.1781
✓	✓	✓	**0.6654**	**0.3372**	**0.3018**	**0.5271**	**0.1826**	**0.2033**

Table 3. Ablation study of contrastive learning and uncertainty-aware multi-view cross-pseudo-supervision.

Method	Task1			Task2		
	ACC	NMI	ARI	ACC	NMI	ARI
Baseline	0.4685	0.2107	0.1457	0.3899	0.0851	0.0522
SCL	0.5381	0.2362	0.1988	0.4092	0.1121	0.1003
UCL	0.5492	0.2482	0.2151	0.4291	0.1173	0.1174
SCL+UCL	**0.5898**	**0.2701**	**0.2375**	**0.4402**	**0.1465**	**0.1322**
w/o CPS	0.6021	0.2877	0.2688	0.4828	0.1672	0.1629
CPS	0.6426	0.3201	0.2917	0.5082	0.1703	0.1902
UMCPS	**0.6654**	**0.3372**	**0.3018**	**0.5271**	**0.1826**	**0.2033**

contrastive learning helps to learn a general and robust feature representation for NCD. In addition, UMCPS also improves the clustering performance of the model, which indicates that unified training helps to the category information interaction. LIA further improves the clustering performance, which indicates that local information aggregation helps to provide better pseudo labels. Finally, our algorithm incorporates each component to achieve the best performance.

Ablation Study of Contrastive Learning. We further examined the effectiveness of each component in contrastive learning. Recall that the contrastive learning strategy includes supervised contrastive learning for the labeled known category data and unsupervised contrastive learning for all data. As shown in Table 3, it can be observed that both components improve the clustering performance of the model, which indicates that SCL helps the model to learn semantically meaningful feature representations, while UCL makes the model learn robust unbiased feature representations and avoid its overfitting to known categories.

Uncertainty-Aware Multi-view Cross-Pseudo-Supervision. We also examine the effectiveness of uncertainty-aware multi-view cross-pseudo-supervision. We compare it with 1) w/o CPS, which does not use cross-pseudo-supervision, and 2) CPS, which uses cross-pseudo-supervision but not the uncertainty to control the contribution of the pseudo label. As shown in Table 3, it can be seen that CPS outperforms w/o CPS, which indicates that CPS encourages the model to maintain consistent predictions for different augmented versions of the input images, and enhances the generalization performance of the model. UMCPS achieves the best clustering performance, which shows its ability to use uncertainty to alleviate the effect of noisy pseudo labels and avoid causing error accumulation.

4 Conclusion

In this paper, we propose a novel class discovery framework for discovering new dermatological classes. Our approach consists of three key designs. First, contrastive learning is used to learn a robust feature representation. Second, uncertainty-aware multi-view cross-pseudo-supervision strategy is trained uniformly on data from all categories, while prediction uncertainty is used to alleviate the effect of noisy pseudo labels. Finally, the local information aggregation module further refines the pseudo label by aggregating the neighborhood information to improve the clustering performance. Extensive experimental results validate the effectiveness of our approach. Future work will be to apply this framework to other medical image analysis tasks.

References

1. Asano, Y.M., Rupprecht, C., Vedaldi, A.: Self-labelling via simultaneous clustering and representation learning. arXiv preprint arXiv:1911.05371 (2019)
2. Cao, K., Brbic, M., Leskovec, J.: Open-world semi-supervised learning. arXiv preprint arXiv:2102.03526 (2021)
3. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018. LNCS, vol. 11218, pp. 139–156. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01264-9_9
4. Codella, N.C., et al.: Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 168–172. IEEE (2018)
5. Combalia, M., et al.: BCN20000: dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288 (2019)
6. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: Advances in Neural Information Processing Systems, vol. 26 (2013)
7. Fini, E., Sangineto, E., Lathuilière, S., Zhong, Z., Nabi, M., Ricci, E.: A unified objective for novel class discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9284–9292 (2021)
8. Gutmann, M., Hyvärinen, A.: Noise-contrastive estimation: a new estimation principle for unnormalized statistical models. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, pp. 297–304 (2010)
9. Han, K., Rebuffi, S.A., Ehrhardt, S., Vedaldi, A., Zisserman, A.: AutoNovel: automatically discovering and learning novel visual categories. IEEE Trans. Pattern Anal. Mach. Intell. **44**(10), 6767–6781 (2021)

10. Hang, W., et al.: Local and global structure-aware entropy regularized mean teacher model for 3D left atrium segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020, Part I. LNCS, vol. 12261, pp. 562–571. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_55
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
12. Khosla, P., et al.: Supervised contrastive learning. In: Advances in Neural Information Processing Systems, vol. 33, pp. 18661–18673 (2020)
13. Kuhn, H.W.: The Hungarian method for the assignment problem. Naval Res. Logist. Q. **2**(1–2), 83–97 (1955)
14. Li, Z., Togo, R., Ogawa, T., Haseyama, M.: Learning intra-domain style-invariant representation for unsupervised domain adaptation of semantic segmentation. Pattern Recogn. **132**, 108911 (2022)
15. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20697–20706 (2022)
16. Mahbod, A., Schaefer, G., Ellinger, I., Ecker, R., Pitiot, A., Wang, C.: Fusing fine-tuned deep features for skin lesion classification. Comput. Med. Imaging Graph. **71**, 19–29 (2019)
17. Tang, P., Liang, Q., Yan, X., Xiang, S., Zhang, D.: GP-CNN-DTEL: global-part CNN model with data-transformed ensemble learning for skin lesion classification. IEEE J. Biomed. Health Inform. **24**(10), 2870–2882 (2020)
18. Yao, P., et al.: Single model deep learning on imbalanced small datasets for skin lesion classification. IEEE Trans. Med. Imaging **41**(5), 1242–1254 (2021)
19. You, C., Zhou, Y., Zhao, R., Staib, L., Duncan, J.S.: SimCVD: simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. IEEE Trans. Med. Imaging **41**(9), 2228–2237 (2022)
20. Zhang, B., et al.: Opportunities and challenges: classification of skin disease based on deep learning. Chin. J. Mech. Eng. **34**(1), 1–14 (2021)
21. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
22. Zhang, J., Xie, Y., Xia, Y., Shen, C.: Attention residual learning for skin lesion classification. IEEE Trans. Med. Imaging **38**(9), 2092–2103 (2019)
23. Zhong, Z., Fini, E., Roy, S., Luo, Z., Ricci, E., Sebe, N.: Neighborhood contrastive learning for novel class discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10867–10875 (2021)
24. Zhong, Z., Zhu, L., Luo, Z., Li, S., Yang, Y., Sebe, N.: OpenMix: reviving known knowledge for discovering novel visual categories in an open world. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9462–9470 (2021)