MedGCD: Generalized Category Discovery in Medical Imaging

Ankit Das[⋆1] (✉), Chandan Gautam[⋆2,3](✉), Pritee Agrawal[4], Feng Yang[1], Yong Liu[1], and Ramasamy Savitha[2,3]

[1] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore
[2] Institute for Infocomm Research (I^{2}R), Agency for Science, Technology and Research (A*STAR), Singapore
[3] International Research Laboratory on Artificial Intelligence (IPAL), CNRS@CREATE, Singapore
[4] Singapore University of Technology and Design (SUTD)
{dasak,yangf,liuyong}@ihpc.a-star.edu.sg, {gautamc,ramasamysa}@i2r.a-star.edu.sg, pritee_agrawal@sutd.edu.sg

Abstract. Existing semi-supervised learning methods in medical imaging assume that unlabeled and labeled data share identical classes. However, in real-world medical scenarios, unlabeled datasets often contain novel categories not present in the labeled data. To address this problem, we propose MedGCD (Generalized Category Discovery for Medical Images), a method that identifies seen categories in labeled data and clusters novel categories in unlabeled data. Specifically, MedGCD introduces a dual stream of strong views in a weak-to-strong framework coupled with a confidence-aware pairwise objective for discovering novel categories. This dual view approach enhances feature extraction from unlabeled data, while the confidence-aware pairwise objective ensures the selection of reliable samples, enabling effective clustering of novel categories. Extensive experiments on benchmark datasets demonstrate the effectiveness of the proposed model in discovering novel categories while maintaining consistent performance on seen categories, with improvements in novel category ranging from 4% to 15%, leading to an overall accuracy improvement of 2% to 8%. [5]

Keywords: GCD · Medical imaging · Novel categories

1 Introduction

Deep learning has advanced medical imaging tasks [9], primarily in supervised settings using large labeled datasets [21]. However, acquiring labeled data in medical applications is often costly and challenging. Few-shot learning methods [26] address this issue by utilizing minimal labeled samples, yet they depend

⋆ Joint first and corresponding author.
[5] https://github.com/Chandan-IITI/MedGCD

entirely on labeled data and overlook the vast potential of unlabeled data. Semi-Supervised Learning (SSL) [18, 20] improves performance and generalization by utilizing the limited labeled data together with abundant unlabeled data. However, SSL operates under the assumption that labeled and unlabeled data share identical classes (closed world learning), a condition often unmet in medical applications. New patterns, conditions, or diseases may emerge over time, leading SSL to misclassify novel categories as known ones, resulting in poor performance. This underscores the need for methods that can handle both seen and novel categories effectively.

Existing methods have addressed novel categories in unlabeled data (open world learning) [16, 25, 4, 14] in diverse ways. Some reject novel categories to protect seen performance [29, 7], while open-set recognition treats them as outliers [12, 10]. Zero-shot learning classifies both seen and novel categories, but relies on semantic information [11]. Novel Category Discovery (NCD) methods focus solely on clustering novel categories in unlabeled data [16, 14]. This paper focuses on Generalized Category Discovery (GCD) [4, 25, 19], capable of handling both seen and novel categories in unlabeled data without requiring additional information. While GCD methods have predominantly focused on natural images (i.e, diverse in content), the main focus of our work lies in the intersection of GCD and medical domain that still remains unexplored. Medical images present unique challenges due to their focus on anatomical structures, physiological functions, and pathological conditions. Current GCD methods struggle to effectively identify novel categories in medical data, likely due to insufficient exploitation of unlabeled data. This highlights the need for specialized models tailored to medical images, capable of leveraging unlabeled data to improve novel category detection, aiding in diagnosing rare diseases and atypical symptoms while maintaining robust performance on seen categories.

To that end, we propose Medical Generalized Category Discovery (MedGCD), a novel framework designed to harness the potential of unlabeled data by learning accurate representations and effectively distinguishing between seen and novel categories, thereby enhancing representation learning and generalization. Overall, we make the following key contributions in this paper. First, MedGCD employs a weak-to-strong perturbation framework, inspired by SSL, where weak perturbations serve as ground truth for guiding strong perturbations. Second, to further enhance robustness and mitigation of confirmation bias [1], we introduce an additional strong view of the data, with both strong views sharing a common weak view as ground truth. This dual strong-view setup generates diverse yet consistent predictions from the same input, capturing complementary aspects of the data distribution while minimizing the distance between them, akin to contrastive learning known to enhance the learning of discriminative features. Third, we propose a confidence-aware pairwise objective that focuses on sample similarity to form cohesive clusters to improve the discovery of novel categories. Finally, in three benchmark medical classification datasets (i.e., PathMNIST, OrganAMNIST, and BloodMNIST), we show that MedGCD performance sig-

nificantly surpasses state-of-the-art methods, with improvements in the accuracy for novel categories ranging from 4% to 17% in different datasets.

2 Proposed Method: MedGCD

In the GCD setting, we have a labeled set $\mathcal{X}^l = \{(x_i^l, y_i^l)\}_{i=1}^n$ and an unlabeled set $\mathcal{X}^u = \{x_i^u\}_{i=1}^m$, with batch sizes B_l and B_u, respectively. Here, x_i^l denotes the i^{th} training labeled sample and y_i^l represents its corresponding ground truth label. Categories in the labeled set are seen categories (C_s), while those in the unlabeled set but not in the labeled data are novel categories ($C_n = C_u \backslash C_s$). The unlabeled data contains samples from both seen and novel categories. The goal is to build a model that classifies samples into seen categories and assigns novel categories to unseen samples, assuming the number of novel categories $|C_n|$ is known a priori.

2.1 Overview of MedGCD

Traditional semi-supervised medical image classification methods cannot identify novel categories in unlabeled data because they assume the presence of only seen (known) categories. To address this limitation, we adapt the **GCD** to introduce the capability of classifying novel categories from the unlabeled medical images, and it is termed as **MedGCD**. We present a simple yet effective way to utilize unlabeled data through a dual-view perturbation of the data space. This approach helps in learning relevant and distinct feature representation for both seen and novel categories. Upon learning the relevant feature representation, we employ a confidence-aware pairwise similarity objective that utilizes these features to discover novel clusters in the data. The MedGCD framework is detailed in Fig. 1, providing a comprehensive overview of its functionality and various components. The **overall objective** of MedGCD is structured into four key components: i) Margin-based cross-entropy loss ($\mathcal{L}_{mce}$); ii) Maximum entropy regularization loss ($\mathcal{L}_{mer}$); iii) Dual strong weak loss ($\mathcal{L}_{dsv}$); and iv) Confidence-aware pairwise objective loss ($\mathcal{L}_{capo}$) :

$$\mathcal{L} = \lambda_{mce}\mathcal{L}_{mce} + \lambda_{mer}\mathcal{L}_{mer} + \lambda_{dsv}\mathcal{L}_{dsv} + \lambda_{capo}\mathcal{L}_{capo}, \tag{1}$$

where λ_{mce}, λ_{mer}, λ_{dsv} and λ_{capo} represent the coefficients associated with different loss functions. First loss function in the above objective function is the margin-based cross-entropy loss, denoted as $\mathcal{L}_{mce}$, used to train the model using labeled data. The loss function is as follows.

$$\mathcal{L}_{\text{mce}} = \frac{1}{n}\sum_{i=1}^{n} -\log \frac{e^{(W_{y_i^l}^T \cdot f_i^l + m)/\mathcal{T}}}{e^{(W_{y_i^l}^T \cdot f_i^l + m)/\mathcal{T}} + \sum_{j \neq i} e^{W_{y_j^l}^T \cdot f_i^l/\mathcal{T}}}, \tag{2}$$

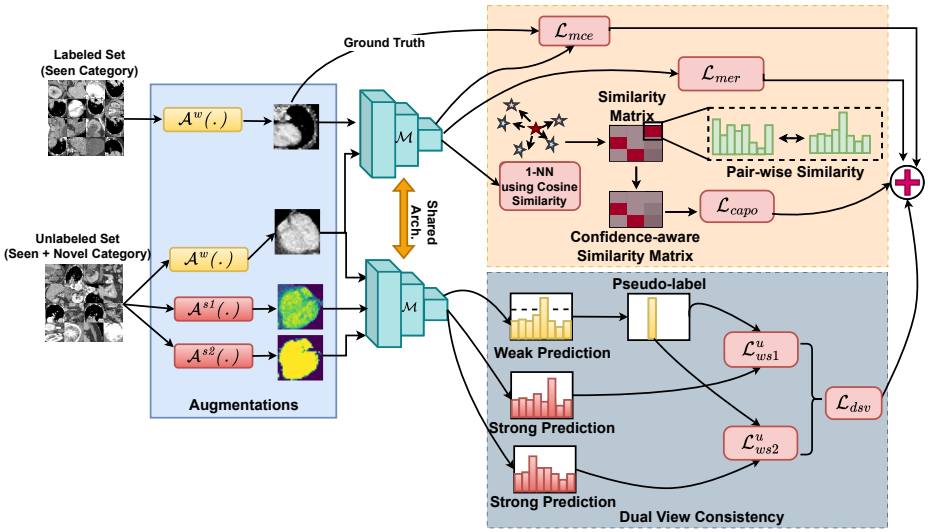

Fig. 1: Illustration of the proposed MedGCD. The process begins by inputting labeled and unlabeled datasets into model $\mathcal{M}$. Unlabeled samples undergo weak and dual strong augmentations, while labeled data only undergoes weak augmentations, serving as ground truth for the dual strong view loss (Eq. 6). Nearest neighbors are computed for each sample, and a confidence-aware pairwise objective (Eq. 7) is applied to confident pairs to discover novel categories. Labeled data is trained using margin-based cross-entropy loss (Eq. 2) and entropy regularization (Eq. 3) to avoid trivial solutions.

where $\mathcal{T}$ is the scaling parameter that controls the temperature of the loss, $W_{y_i^l}$ denotes linear layer weights, f_i^l denotes feature of i-th sample in labeled data, margin parameter m regulates the learning process of the model by reducing bias towards labeled data and m is computed similarly to [5]. The second loss in the objective is the maximum entropy regularization loss ($\mathcal{L}_{mer}$) [24, 2] that is integrated to prevent the occurrence of trivial solutions where all unlabeled data are assigned to a single class. It is represented as follows:

$$\mathcal{L}_{mer} = -\sum_{i=1}^{m+n} p_i \log_b(p_i), \tag{3}$$

where $p_i = \mathcal{M}(\mathcal{A}^w(x_i))$ represents the probability distribution and $x_i = x_i^l \cup x_i^u$. The third and fourth loss components are detailed in the following sections.

2.2 Representation Learning using Dual-View Augmentation

To fully exploit the unlabeled data for learning effective representations, we introduce a simple yet effective dual-view augmentation strategy. Inspired by

recent works in semi-supervised [3] and self-supervised [6] learning, which use multiple views to probe the perturbation space for discriminative representations, we investigate whether this strategy can benefit medical GCD tasks by learning diverse representations essential for effective performance. To this end, we apply two independent strong perturbations from a fixed pool [8] to the unlabeled sample x^u, generating two distinct views, $\mathcal{A}^{s1}(x^u)$ and $\mathcal{A}^{s2}(x^u)$. These views differ due to the non-deterministic nature of the perturbation set. We utilize the two strong views alongside the weak view $\mathcal{A}^w(x^u)$, which serves as the ground truth, as follows:

$$\mathcal{L}^u_{wsj} = \frac{1}{B_u}\sum_{i=1}^{B_u} \mathbb{1}(\max(p_i^w) \geq \eta)\mathbb{H}_c(argmax(\mathcal{M}(\mathcal{A}^w(x_i^u)), \mathcal{M}(\mathcal{A}^{sj}(x_i^u))), \forall j = 1, 2, \quad (4)$$

where H_c denotes cross-entropy, η represents a confidence threshold indicating the value above which a pseudo-label will be selected, and $p_i^w = \mathcal{M}(\mathcal{A}^w(x_i^u))$ denotes the class distribution of the pseudo-label. Using the weak view as ground truth for the strong views encourages consistent predictions across augmentations, helping the model focus on invariant representations and improve generalization. Our ablation studies show substantial gains from dual strong views compared to a single strong view. Even using one strong view with a weak view as the reference significantly boosts performance in medical GCD tasks. Enforcing consistency between strong views and a shared weak view also indirectly aligns the strong views. Intuitively, this aligns with contrastive learning principles, where the model learns to produce similar representations for different views of the same sample (positive pairs) and distinguish between views of different samples (negative pairs). This approach helps the model learn robust and distinctive representations, satisfying the InfoNCE objective [22] as follows:

$$\mathcal{L}_{\mathcal{A}^{s1}\leftrightarrow\mathcal{A}^{s2}} = -\log\frac{\exp(\mathcal{A}^{s1}(x^u)\cdot\mathcal{A}^{s2}(x^u))}{\sum_{k=1}^{m}\exp(\mathcal{A}^{si}(x^u)\cdot\mathcal{A}^{sj}(x^k)}, \forall i, j = 1, 2, \quad (5)$$

where m represents number of unlabeled samples. The final dual strong view loss $\mathcal{L}_{dsv}$ is as follows:

$$\mathcal{L}_{dsv} = \mathcal{L}^u_{ws1} + \mathcal{L}^u_{ws2} \quad (6)$$

2.3 Discover Novel Categories Using Pairwise Objective

We propose a confidence-aware pairwise objective to cluster samples into new categories by framing it as a pairwise similarity problem. This approach relies on the principle that any two images either belong to the same cluster or not. The pairwise objective aims to learn and predict the closest pairs among the unlabeled data, facilitating the identification of new clusters. While traditional BCE is commonly used for pairwise similarity [4], it struggles in GCD due to the mix of seen and novel categories in unlabeled data. Samples from novel categories may initially appear distant due to poor representations. To address this, we

focus on pulling closer samples rather than pushing dissimilar ones and restrict the objective to high-similarity pairs to avoid incorrect clustering. In a mini-batch with labeled samples $\mathcal{X}^l_{B_l}$ and unlabeled samples $\mathcal{X}^u_{B_u}$, we pair labeled samples with others from the same class. For each unlabeled sample $x^u_{B_u}$, we compute its cosine similarity with all samples in the batch and select the nearest neighbor (1-NN) $\hat{x}^u$. A sample $x^u_{B_u}$ is included in the pairwise objective if its similarity with $\hat{x}^u$ exceeds a confidence threshold θ. Based on this, we present a Confidence-Aware Pairwise Objective (CAPO) given by:

$$\begin{aligned}\mathcal{L}_{capo} = &-\frac{1}{B_l}\sum_{i=1}^{B_l} log(\mathcal{M}(\mathcal{A}^w(x^l_i))^T \mathcal{M}(\mathcal{A}^w(\hat{x}^l_i))) \\ &-\frac{1}{B_u}\sum_{i=1}^{B_u} \mathbb{I}(sim(f^u_i, \hat{f}^u_i) \geq \theta) log(\mathcal{M}(\mathcal{A}^w(x^u_i))^T \mathcal{M}(\mathcal{A}^w(\hat{x}^u_i)),\end{aligned} \tag{7}$$

where f^u and $\hat{f}^u$ represent features of the unlabeled sample and its nearest neighbor, respectively. θ is the confidence threshold, and sim denotes cosine similarity. This approach ensures high-quality pairwise relationships by focusing on reliable, high-similarity pairs.

3 Experiments and Results

Datasets. We evaluate the performance of our proposed MedGCD on three benchmark datasets: PathMNIST, OrganAMNIST and BloodMNIST [28] comprising 9, 11 and 8 categories, respectively. Each dataset is split into approximately 50% seen and 50% novel categories. Specifically, for PathMNIST and OrganAMNIST, 5 categories are considered seen, while in case of BloodMNIST, 4 categories are seen, and the remaining categories as novel. Further, we consider 50% of the samples in seen category as labeled and remaining 50% as unlabeled.

Implementation Details. The proposed method was implemented in PyTorch on an NVIDIA A40 GPU with 120 GB RAM. ResNet-18 was used as the backbone. Model convergence was achieved using the SGD optimizer over 200 epochs, with a batch size of 512, learning rate of 0.1, momentum of 0.9, and weight decay of $5e-4$. Uniform weights of 1 were assigned to λ_{mce}, λ_{mer}, λ_{dsv}, and λ_{capo}. Confidence thresholds θ and η were set to 0.95, and temperature $\mathcal{T}$ to 1.

Evaluation Metrics. For evaluation, we follow the protocol in [4]. We compute classification accuracy for seen and clustering accuracy for novel categories, along with Normalized Mutual Information (NMI). Clustering accuracy is determined using the Hungarian algorithm [17]. Joint accuracy across all categories is reported using clustering accuracy for both seen and novel categories.

Data Augmentation. In all experiments, MedGCD employed random horizontal flipping for weak augmentation and randaugment [8] for strong augmentation.

Method	PathMNIST				OrganAMNIST				BloodMNIST			
	All	**Seen**	**Novel**	**NMI**	**All**	**Seen**	**Novel**	**NMI**	**All**	**Seen**	**Novel**	**NMI**
Fixmatch	47.68	70.61	35.38	25.58	57.51	90.66	56.6	53.3	48.37	73.17	49.82	38.67
DS3L	47.86	67.4	36.91	23.94	59.19	90.6	67.45	52.82	48.84	65.99	50.77	40.70
DTC	43.41	72.32	33.33	15.21	54.3	92.3	50.67	47.20	43.8	59.51	46.96	33.27
Rankstats	51.18	88.21	39.55	25.35	60.5	94.51	58.63	53.41	51.09	52.66	58.66	57.65
Rankstats+	52.54	89.22	44.52	28.44	62.05	97.09	57.68	56.64	56.61	77.11	60.03	57.6
GCD	61.59	94.33	44.04	28.80	73.61	96.67	68.88	62.67	64.64	82.7	64.13	58.13
OpenNCD	46.71	67.34	35.17	11.82	36.11	69.45	31.94	30.12	48.7	34.37	59.31	49.05
ORCA	64.94	**95.56**	47.80	33.48	76.87	97.98	72.54	65.59	65.97	83.39	65.43	52.55
SimGCD	66.88	95.35	50.12	34.85	79.54	98.15	76.45	71.22	71.59	84.76	71.34	62.01
MedGCD	**67.42**	95.12	**51.92**	**35.22**	**81.92**	**98.98**	**81.3**	**83.04**	**73.83**	**87.63**	**80.83**	**69.88**

Table 1: Comparison of the proposed MedGCD with SOTA methods.

3.1 Comparison with State-of-the-Art Methods

We compare the performance of MedGCD with SOTA methods, including OpenNCD [19], ORCA [4], SimGCD [27], GCD [25], Rankstats [14], Rankstats+ [15], and DTC [16]. Rankstats, Rankstats+, and DTC belong to the NCD category, capable of identifying novel categories rather than seen ones. We extend Rankstats, Rankstats+ and DTC to be able to classify the seen categories as well, employing the Hungarian algorithm [17]. Additionally, we compare our results with the popular SSL methods Fixmatch [23] and DS3L [13], which only classify the seen categories. We extend these methods to classify the novel categories by estimating out-of-distribution samples and applying K-means clustering.

Table 1 shows the overall accuracy ('All'), seen accuracy, novel accuracy, and NMI results of MedGCD compared to baseline methods. We observe that MedGCD outperforms all the baseline methods. It should be noted that all the baseline methods struggle to improve the accuracy of novel categories while MedGCD achieves significant accuracy improvements, ranging from 4% to 15% compared to ORCA, the best performing baseline across all three benchmark datasets. Moreover, MedGCD provides upto 2% improvements on seen accuracy compared to ORCA. Additionally, it enhances NMI measures by at least 10% more than the best baseline in two out of the three datasets, indicating superior cluster quality of MedGCD. Finally, it achieves an overall accuracy improvement in the range of 2% to 8% compared to the best baseline, demonstrating the effectiveness of MedGCD in learning superior representations from unlabeled data to accurately classify both seen and novel categories.

3.2 Ablation

Analysis of losses. The proposed MedGCD algorithm incorporates four losses: $\mathcal{L}_{mer}$, $\mathcal{L}_{capo}$, $\mathcal{L}_{mce}$, and $\mathcal{L}_{dsv}$, which includes two views ($\mathcal{L}^u_{ws1}$ and $\mathcal{L}^u_{ws2}$). To assess the contribution of each loss, we systematically set each to 0 and train the model without it, evaluating its impact on performance. Table 2 shows the effect of each loss on the OrganAMNIST dataset. The first row highlights a performance drop when $\mathcal{L}_{mce}$ is removed, emphasizing its role in leveraging labeled

data. The second row shows a decline in novel accuracy without $\mathcal{L}_{mer}$, indicating a bias toward seen categories. The third row reveals a 28% performance degradation when $\mathcal{L}_{capo}$ is excluded, underscoring its importance in identifying pairs for labeled and unlabeled data. The fourth row demonstrates a 14% decrease without $\mathcal{L}_{dsv}$, highlighting its role in exploiting unlabeled data. These results confirm that all losses contribute significantly to MedGCD's performance.

Use of multiple views. We analyze the impact of number of views on model performance using OrganAMNIST (Table 3). Increasing augmentations does not consistently improve performance, reaching a saturation point.

Methods	All	Seen	Novel	NMI
w/o $\mathcal{L}_{mce}$	32.80	40.82	39.02	39.33
w/o $\mathcal{L}_{mer}$	44.54	**99.86**	41.95	49.61
w/o $\mathcal{L}_{capo}$	53.61	68.92	52.96	45.53
w/o $\mathcal{L}_{dsv}$	67.54	96.13	64.54	60.60
w/o $\mathcal{L}_{ws1}$	71.92	98.96	72.51	75.90
w/o $\mathcal{L}_{ws2}$	71.48	95.67	69.40	71.81
MedGCD	**81.92**	98.98	**81.30**	**83.04**

Table 2: Ablation of different losses on OrganAMNIST.

# Views	All	Seen	Novel	NMI
1	71.92	98.96	72.51	75.90
2	81.92	98.98	81.30	83.04
3	83.03	98.33	82.39	83.06
4	83.20	99.31	82.52	82.68

Table 3: Ablation of Views in MedGCD on OrganAMNIST.

Analysis of different labeled ratios. We evaluate MedGCD on OrganAMNIST by varying labeled data percentages, with results visualized in Fig. 2. MedGCD maintains stable performance, with slight improvements as the percentage of labeled classes increase.

Impact of confidence-aware threshold. We evaluated the impact of varying the confidence-aware threshold from 0.8 to 0.99 using OrganAMNIST data, as shown in Fig. 3. The results indicate optimal performance at a threshold of 0.95.

4 Conclusion

In this work, we proposed MedGCD to address the problem of generalized class discovery in the medical domain, aiming to improve diagnoses and treatments for rare, atypical, and emerging diseases, thereby enhancing patient care. MedGCD identifies novel categories in unlabeled data while accurately classifying seen categories by exploiting the unlabeled data via a dual strong view in a weak-to-strong framework and a confidence-aware pairwise objective. Our extensive experiments on benchmark medical imaging classification datasets demonstrate the superior performance of MedGCD in discovering novel categories while maintaining the performance on seen categories compared to state-of-the-art methods. Limitations: The proposed method, like most existing methods, requires knowing

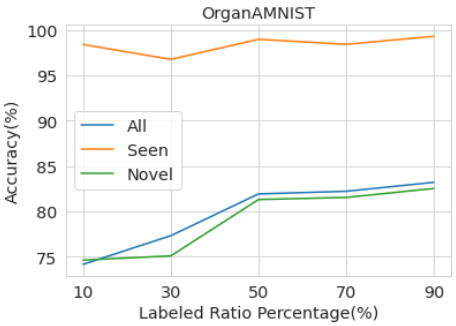

Fig. 2: Ablation of labeled ratio(%).

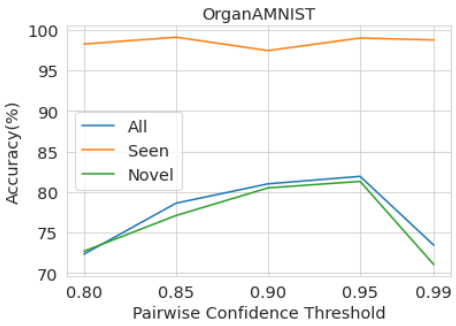

Fig. 3: Ablation by varying θ.

the number of unknown classes in advance and is not suitable for cross-domain generalization.

Acknowledgments. This work was supported by the Agency for Science, Technology and Research (A*STAR) through its DesCartes programme and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. This work was also supported by the Agency for Science, Technology and Research (A*STAR) through Singapore Enhanced Eye-care Referral System (SEERS): Questionnaire and Image-based Models project SC23/24-8504CI.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arazo, E., Ortego, D., Albert, P., O'Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International joint conference on neural networks (IJCNN). pp. 1–8. IEEE (2020)
2. Arazo, E., Ortego, D., Albert, P., O'Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2020)
3. Berthelot, D., Carlini, N., Cubuk, E.D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C.: Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. arXiv preprint arXiv:1911.09785 (2019)
4. Cao, K., Brbic, M., Leskovec, J.: Open-world semi-supervised learning. In: International Conference on Learning Representations (2022)
5. Cao, K., Chen, Y., Lu, J., Arechiga, N., Gaidon, A., Ma, T.: Heteroskedastic and imbalanced deep learning with adaptive regularization. In: International Conference on Learning Representations (2020)
6. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. Advances in neural information processing systems **33**, 9912–9924 (2020)

7. Cen, J., Zhang, S., Wang, X., Pei, Y., Qing, Z., Zhang, Y., Chen, Q.: Enlarging instance-specific and class-specific information for open-set action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15295–15304 (2023)
8. 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/CVF conference on computer vision and pattern recognition workshops. pp. 702–703 (2020)
9. Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., Socher, R.: Deep learning-enabled medical computer vision. NPJ digital medicine **4**(1), 5 (2021)
10. Gautam, C., Parameswaran, S., Kane, A., Fang, Y., Ramasamy, S., Sundaram, S., Sahu, S., Li, X.: Class name guided out-of-scope intent classification. In: Findings of the Association for Computational Linguistics: EMNLP 2024. pp. 9100–9112 (2024)
11. Gautam, C., Parameswaran, S., Verma, V., Sundaram, S., Ramasamy, S.: Refinement matters: Textual description needs to be refined for zero-shot learning. In: Findings of the Association for Computational Linguistics: EMNLP 2022. pp. 6127–6140 (2022)
12. Geng, C., Huang, S.j., Chen, S.: Recent advances in open set recognition: A survey. IEEE transactions on pattern analysis and machine intelligence **43**(10), 3614–3631 (2020)
13. Guo, L.Z., Zhang, Z.Y., Jiang, Y., Li, Y.F., Zhou, Z.H.: Safe deep semi-supervised learning for unseen-class unlabeled data. In: International Conference on Machine Learning. pp. 3897–3906. PMLR (2020)
14. Han, K., Rebuffi, S.A., Ehrhardt, S., Vedaldi, A., Zisserman, A.: Automatically discovering and learning new visual categories with ranking statistics. In: International Conference on Learning Representations (ICLR) (2020)
15. Han, K., Rebuffi, S.A., Ehrhardt, S., Vedaldi, A., Zisserman, A.: Autonovel: Automatically discovering and learning novel visual categories. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2021)
16. Han, K., Vedaldi, A., Zisserman, A.: Learning to discover novel visual categories via deep transfer clustering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8401–8409 (2019)
17. Kuhn, H.W., Yaw, B.: The hungarian method for the assignment problem. Naval Res. Logist. Quart pp. 83–97 (1955)
18. Kurakin, A., Li, C.L., Raffel, C., Berthelot, D., Cubuk, E.D., Zhang, H., Sohn, K., Carlini, N., Zhang, Z.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In: Proceedings of Advances in Neural Information Processing Systems (2020)
19. Liu, J., Wang, Y., Zhang, T., Fan, Y., Yang, Q., Shao, J.: Open-world semi-supervised novel class discovery. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23. pp. 4002–4010. International Joint Conferences on Artificial Intelligence Organization (8 2023)
20. Liu, Q., Yu, L., Luo, L., Dou, Q., Heng, P.A.: Semi-supervised medical image classification with relation-driven self-ensembling model. IEEE transactions on medical imaging **39**(11), 3429–3440 (2020)
21. Miranda, E., Aryuni, M., Irwansyah, E.: A survey of medical image classification techniques. In: 2016 international conference on information management and technology (ICIMTech). pp. 56–61. IEEE (2016)

22. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
23. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)
24. Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M., Van Gool, L.: Scan: Learning to classify images without labels. In: European conference on computer vision. pp. 268–285. Springer (2020)
25. Vaze, S., Han, K., Vedaldi, A., Zisserman, A.: Generalized category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7492–7501 (2022)
26. Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M.: Generalizing from a few examples: A survey on few-shot learning. ACM computing surveys (csur) **53**(3), 1–34 (2020)
27. Wen, X., Zhao, B., Qi, X.: Parametric classification for generalized category discovery: A baseline study. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 16590–16600 (2023)
28. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. Scientific Data **10**(1), 41 (2023)
29. Zhou, D.W., Ye, H.J., Zhan, D.C.: Learning placeholders for open-set recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4401–4410 (2021)