Generalized Categories Discovery for Long-tailed Recognition

Ziyun Li
Hasso Plattner Institut
University of Potsdam
ziyun.li@hpi.de

Christoph Meinel
Hasso Plattner Institut
University of Potsdam
christoph.meinel@hpi.de

Haojin Yang
Hasso Plattner Institut
University of Potsdam
haojin.yang@hpi.de

Abstract

Generalized Class Discovery (GCD) plays a pivotal role in discerning both known and unknown categories from unlabeled datasets by harnessing the insights derived from a labeled set comprising recognized classes. A significant limitation in prevailing GCD methods is their presumption of an equitably distributed category occurrence in unlabeled data. Contrary to this assumption, visual classes in natural environments typically exhibit a long-tailed distribution, with known or prevalent categories surfacing more frequently than their rarer counterparts. Our research endeavors to bridge this disconnect by focusing on the long-tailed Generalized Category Discovery (Long-tailed GCD) paradigm, which echoes the innate imbalances of real-world unlabeled datasets. In response to the unique challenges posed by Long-tailed GCD, we present a robust methodology anchored in two strategic regularizations: (i) a reweighting mechanism that bolsters the prominence of less-represented, tail-end categories, and (ii) a class prior constraint that aligns with the anticipated class distribution. Comprehensive experiments reveal that our proposed method surpasses previous state-of-the-art GCD methods by achieving an improvement of approximately 6 - 9% on ImageNet100 and competitive performance on CIFAR100.

1. Introduction

In the context of modern machine learning approaches, the effectiveness of models is heavily dependent on their training data, especially when trained on large datasets that have been annotated by humans. This limitation restricts their effectiveness in real-world situations where encountering unannotated data from previously unseen categories is common. To fill this research gap, scholars have explored several avenues. One approach is open-set recognition [16], which aims to determine if an unlabeled piece of data belongs to a class we've seen before. Another is the Novel Category Discovery (NCD) [7, 6, 4, 12]. How-

ever, early NCD methods were flawed, assuming all unlabelled images represented entirely new categories. This isn't typically the case in real-world scenarios. Recognizing this limitation, the Generalized Category Discovery (GCD) [18] framework was developed. GCD understands that unlabeled data can be from both previously known or entirely new categories, making it a more realistic approach for diverse data settings.

A pivotal challenge in GCD is identified: the *lack of information about occurrence frequency* of categories, as mention in [11], whether known or unknown, forming the crux of our research. Notably, much of the existing research, as seen in [18, 17, 1, 15], neglects this frequency issue. Long-tailed GCD aligns more accurately with the real-world, emphasizing the long-tailed distribution of visual classes with known classes like dominate the distribution's "head", whereas unknown classes being rarer, populate the sparse "tail". We present a novel method to address the challenges inherent to long-tailed GCD. One predominant challenge in long-tailed GCD is the tendency of the system to exhibit bias towards head classes, often at the expense of tail/unknown classes. This is manifested when tail or unknown samples are frequently misclassified as head or known classes. The primary reason for this is the model's inclination to learn predominantly from the patterns of the more prevalent classes, which in turn compromises the accuracy for the minority classes.

To rectify this, our approach diverges from the ImbaGCD [11] of utilizing optimal transport alignment in accordance with the estimated prior distribution. Instead, we introduce two strategic constraints: (i) Reweighting of tail/unknown classes: We amplify the weight attributed to the tail distribution. (ii) Alignment of the estimated prior distribution: We impose constraints on the estimated class distribution. Compared to ImbaGCD [11], our method substantially reduces computational complexity by eliminating the EM and Sinkhorn iterations, which are used to solve the optimal transport function. Concurrently, we observe a performance improvement of 6 - 9% on ImageNet100 and achieve competitive results on CIFAR100.

2. Problem Setup

Definition 1 (Long-tailed Generalized Category Discovery). *Let $\mathbb{P}_{\mathbf{X}_l, Y_l}$ and $\mathbb{Q}_{\mathbf{X}_u, Y_u}$ represent labeled and unlabeled probability measures on $\mathcal{X}_l \times \mathcal{Y}_l$ and $\mathcal{X}_u \times \mathcal{Y}_u$ respectively, with $\mathcal{Y}_l \subset \mathcal{Y}_u$ ($\mathcal{Y}_u$ represents all classes in the unlabeled set, comprising known classes $\mathcal{Y}_l$ and unknown classes $\mathcal{Y}_n = \mathcal{Y}_u \backslash \mathcal{Y}_l$) and $\mathcal{Y}_n = \mathcal{Y}_u \backslash \mathcal{Y}_l$. Given datasets $\mathcal{L}_n$ and $\mathcal{U}_m$ sampled from these measures, the goal of GCD is to predict labels for $\mathcal{U}_m$, which can belong to known or unknown classes. In long-tailed GCD, the class distributions in the unlabeled set are skewed such that $\mathbb{P}_{Y_l}(i) > \mathbb{P}_{Y_n}(j)$ for every $i \in \mathcal{Y}_l$ and $j \in \mathcal{Y}_n$.*

3. Method

3.1. Class Prior Estimation

Moving-average distribution update We suggest employing model predictions for class prior estimation following [5, 11]. However, due to potential inaccuracies and biases in early training stages, we propose a moving-average update mechanism to enhance reliability. Starting with a uniform class prior $\mathbf{r} = [1/C, \ldots, 1/C]$, we iteratively refine the distribution per epoch.

$$\mathbf{r} := \mu \mathbf{r} + (1-\mu)\mathbf{z},$$
$$\text{where } \mathbf{z}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(j = \arg\max_{j'} f_{j'}(\mathbf{x}_i)\right), \mu \in [0,1]$$

The class prior is continuously updated through a linear function, resulting in more stable training dynamics. As the training progresses, the model's accuracy improves, making the estimated distribution increasingly dependable.

3.2. Representation Improvement

To enhance model representation, we adopt unsupervised contrastive learning [3, 8] for unlabeled data and supervised contrastive learning [9] for labeled data, as in [17, 1]. Specifically, let $\mathbf{v}_i$ and $\mathbf{v}'_i$ represent features from two views (random augmentations) of the same image within a mini-batch B. The instance-level unsupervised contrastive loss and the supervised contrastive loss are defined as follows:

$$\mathcal{L}_{ins}^{(i)} = \frac{1}{|B_u|}\sum_{i \in B_u} -\log\frac{\exp(\mathbf{v}_i \cdot \mathbf{v}'_i/\tau)}{\sum_{i}^{i \neq j}\exp(\mathbf{v}_i \cdot \mathbf{v}_j/\tau)}, \quad (1)$$

$$\mathcal{L}_{sup}^{(i)} = \frac{1}{|B_l|}\sum_{i \in B_l}\frac{1}{|\mathcal{N}^{(i)}|}\sum_{q \in \mathcal{N}^{(i)}} -\log\frac{\exp(\mathbf{v}_i \cdot \mathbf{v}_q/\tau)}{\sum_{i}^{i \neq j}\exp(\mathbf{v}_i \cdot \mathbf{v}_j/\tau)}, \quad (2)$$

where B_u and B_l represent the unlabeled and labeled subsets of mini-batch B, respectively. Furthermore, $\mathcal{N}^{(i)}$ refers to the indices of other images in the batch sharing the same label, and τ signifies a temperature parameter.

3.3. Constrain on Class Distribution

Unlike ImbaGCD [11] involve sophisticated techniques such as optimal transport for aligning class prior distribution, the EM algorithm for generating pseudo labels, and updating prototypes. In contrast, our approach is geared towards simplicity and directness, focusing on addressing the inherent challenges of class imbalances, especially in the context of tail classes.

We propose two constraints: (i) *Reweighting tail samples*: A prevalent challenge in dealing with imbalanced datasets is the difficulty of learning patterns from tail classes, given their notably fewer samples compared to the head classes. One direct and effective way to mitigate this is through sample reweighting. Previous works have demonstrated the efficacy of reweighting schemes in various contexts [2, 13, 10]. To effectively balance the information derived from tail classes with the rest, we incorporate a constraint between model predictions $\mathbf{q}$ and the uniform distribution $\mathbf{u}$:

$$\mathrm{H}(\mathbf{q}, \mathbf{u}) = \mathbf{q}\log\mathbf{u}, \quad \text{where } \mathbf{u} = [1/C, ..., 1/C] \quad (3)$$

(ii) *Aligning Class Prior:* While optimal transport methods like Sinkhorn have their merits, they can also be reduced to simpler forms such as KL divergence [14]. For our work, in pursuit of simplicity without significant trade-off in performance, we employ entropic regularization between the estimated distribution and the output. It is represented as:

$$\mathrm{H}(\mathbf{q}, \mathbf{r}) = \mathbf{q}\log\mathbf{r} \quad (4)$$

3.4. Overall Loss Objective

The overall loss objective is a weighted sum of the unsupervised and supervised contrastive losses and two regularizations:

$$\mathcal{L}_{overall} = \mathcal{L}_{ins} + \lambda\mathcal{L}_{sup} + \alpha\mathrm{H}(\mathbf{q}, \mathbf{r}) + \beta\mathrm{H}(\mathbf{q}, \mathbf{u}) \quad (5)$$

4. Experiments

In this section, we conduct an experimental analysis of our proposed method under both original and varying imbalanced scenarios.

4.1. Setup

Datasets Experiments were conducted on CIFAR10, CIFAR100, and ImageNet100. Classes were split evenly into known and unknown. From the known classes, 50% were randomly selected as a balanced labeled dataset. In the unlabeled set, the class sample sizes were imbalanced, adjusted by factor ρ, defined as the ratio $\rho = \frac{n_k}{n_u}$ where n_k and n_u are the sample sizes of known and unknown classes. We considered $\rho \in \{0.5, 1, 5, 10\}$. Previous studies typically had a fixed imbalance of $\rho = 0.5$ by using the remaining 50% of known class samples as the unlabeled set.

Evaluation metrics   We use the evaluation method from [1, 17, 18], focusing on (1) overall accuracy, (2) accuracy for known classes, and, as in ImbaGCD [11], on novel data under settings: (3) **unknown-aware** and (4) **unknown-agnostic**. In the **unknown-aware** evaluation, we separate unlabeled samples from unknown classes and cluster them. However, this might not mirror real-life situations where telling apart known and unknown classes isn't straightforward. Hence, we introduce the **unknown-agnostic** metric that evaluates without pre-classifying samples, offering a more genuine and neutral assessment.

Implementation details   We use ResNet18 for CIFAR10/100 and ResNet50 for ImageNet100. Training lasts 200 epochs for CIFAR10/100 and 120 for ImageNet100, with batch sizes of 512. We apply stochastic gradient descent (momentum 0.9, weight decay $10e - 4$), starting with a learning rate of 0.02 that decays 10x at 50% and 75% milestones. The momentum, μ, for class prior remains at 0.99.

4.2. Compared with SOTA

Baselines   We benchmark our method against four leading GCD techniques: (1) GCD [18] - a foundational GCD framework that clusters unknown categories without exhaustive labeling. (2) ORCA [1] - melds supervised and unsupervised learning, addressing open-world class imbalances. (3) OpenCon [17] - uses contrastive learning to adapt to new categories with minimal supervision. (4) ImbaGCD [11] - employs the EM algorithm, combining Sinkhorn-based pseudo-labeling with contrastive updates. The class prior and prototype are refreshed every epoch. Our evaluation spans original GCD, balanced, and various imbalanced scenarios for a thorough comparison.

Comparisons with the state-of-the-art   In Table 1, our method stands out on ImageNet datasets, especially amidst greater imbalance. We best the leading baseline by approximately 6 - 7% and 7 - 9% in the unknown-aware and unknown-agnostic scenarios for $\rho = 5$ and $\rho = 10$, respectively. For CIFAR100, it also achieves competitive results across a range of imbalanced settings. Notably, the unknown-agnostic assessment is tougher than its unknown-aware counterpart, especially in highly skewed scenarios. In these demanding conditions, our method shines brighter, with its gains in unknown-agnostic results being especially pronounced compared to the unknown-aware ones.

4.3. Hyper-parameters Analysis

Analysis of β   We carried out experiments on the CIFAR100 dataset with an imbalance factor of $\rho = 5$. By fixing $\alpha = 0$, we analyzed the impact of varying β on different accuracy metrics as shown in Figure 1. Our results show
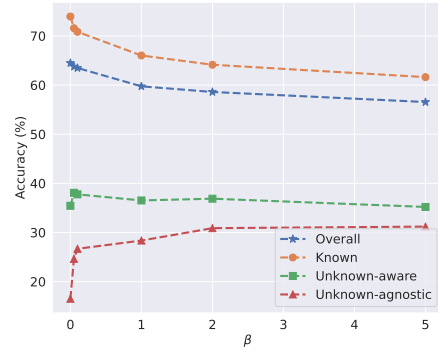

Figure 1. Performance trends of Overall, Known, Unknown-Aware, and Unknown-Agnostic accuracies across different β values with $\alpha = 0$ constant on CIFAR100 with $rho = 5$.
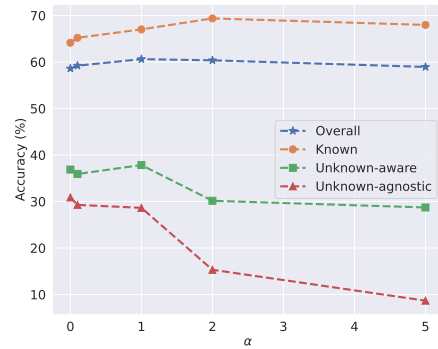

Figure 2. Performance trends of Overall, Known, Unknown-Aware, and Unknown-Agnostic accuracies across different α values with $\beta = 2$ constant on CIFAR100 with $\rho = 5$.
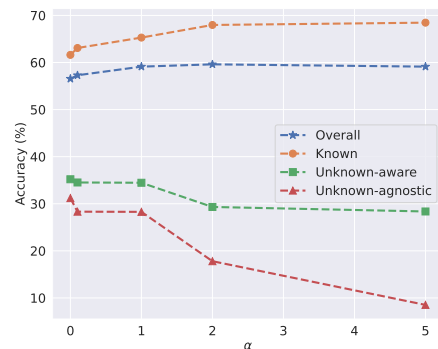

Figure 3. Performance trends of Overall, Known, Unknown-Aware, and Unknown-Agnostic accuracies across different α values with $\beta = 5$ constant on CIFAR100 with $\rho = 5$.

that a higher β enhances the unknown-aware and unknown-agnostic accuracies. However, this increase comes at the expense of known accuracy.

Table 1. Performance comparison on CIFAR100 and ImageNet100 with varying imbalance factors ρ ($\rho = 0.5$ for original GCD setting). Un1 represents unknown-aware accuracies, while Un2 refers to unknown-agnostic accuracies. Our method consistently excels in novel class performance under imbalanced settings and remains competitive in balanced settings. Mean and standard deviation of clustering accuracy are reported over 3 runs for multiple methods. Higher means are bolded, while results within one standard deviation are not.

		Methods									
IMF	Metrics	CIFAR100					ImageNet100				
		GCD	ORCA	OpenCon	ImbaGCD	Ours	GCD	ORCA	OpenCon	ImbaGCD	Ours
$\rho = 0.5$	All	$45.41_{\pm 0.13}$	$55.68_{\pm 0.33}$	$51.85_{\pm 0.63}$	$53.51_{\pm 0.26}$	$52.20_{\pm 0.30}$	$77.12_{\pm 0.56}$	$74.93_{\pm 0.34}$	$82.22_{\pm 0.56}$	$81.90_{\pm 0.67}$	$81.06_{\pm 0.39}$
	Known	$67.61_{\pm 0.12}$	$66.41_{\pm 0.31}$	$69.07_{\pm 0.29}$	$68.09_{\pm 0.13}$	$70.45_{\pm 0.31}$	$87.02_{\pm 0.36}$	$89.21_{\pm 0.06}$	$90.65_{\pm 0.04}$	$91.19_{\pm 0.15}$	$91.03_{\pm 0.03}$
	Un1	$34.31_{\pm 0.22}$	$42.63_{\pm 0.67}$	$45.76_{\pm 0.32}$	$\mathbf{47.92}_{\pm 0.33}$	$46.96_{\pm 0.35}$	$57.48_{\pm 0.76}$	$67.18_{\pm 0.27}$	$78.12_{\pm 0.80}$	$77.89_{\pm 0.88}$	$77.22_{\pm 0.25}$
	Un2	$18.12_{\pm 0.34}$	$38.95_{\pm 0.76}$	$42.11_{\pm 0.54}$	$46.22_{\pm 0.33}$	$46.25_{\pm 0.25}$	$42.68_{\pm 0.77}$	$65.43_{\pm 0.37}$	$78.01_{\pm 0.83}$	$77.79_{\pm 0.97}$	$77.68_{\pm 0.30}$
$\rho = 1$	All	$48.36_{\pm 0.08}$	$47.37_{\pm 0.52}$	$53.20_{\pm 0.33}$	$54.06_{\pm 0.45}$	$55.82_{\pm 0.40}$	$63.06_{\pm 0.66}$	$68.01_{\pm 0.31}$	$82.44_{\pm 0.24}$	$82.34_{\pm 0.39}$	$82.28_{\pm 0.68}$
	Known	$71.48_{\pm 0.24}$	$65.14_{\pm 0.17}$	$68.00_{\pm 0.06}$	$67.98_{\pm 0.37}$	$68.00_{\pm 0.35}$	$88.80_{\pm 0.43}$	$88.99_{\pm 0.05}$	$90.62_{\pm 0.12}$	$90.56_{\pm 0.18}$	$\mathbf{91.30}_{\pm 0.10}$
	Un1	$25.24_{\pm 0.07}$	$34.93_{\pm 1.04}$	$43.58_{\pm 0.34}$	$43.39_{\pm 0.59}$	$\mathbf{44.54}_{\pm 0.40}$	$37.29_{\pm 1.11}$	$47.68_{\pm 0.70}$	$74.45_{\pm 0.36}$	$74.56_{\pm 0.79}$	$75.06_{\pm 1.00}$
	Un2	$12.02_{\pm 0.46}$	$29.61_{\pm 1.02}$	$38.40_{\pm 0.65}$	$38.76_{\pm 0.9}$	$\mathbf{42.65}_{\pm 0.89}$	$31.16_{\pm 1.48}$	$47.28_{\pm 0.66}$	$74.35_{\pm 0.38}$	$74.24_{\pm 0.89}$	$73.39_{\pm 1.38}$
$\rho = 5$	All	$63.13_{\pm 0.12}$	$56.37_{\pm 0.12}$	$61.84_{\pm 0.29}$	$59.48_{\pm 0.48}$	$\mathbf{62.85}_{\pm 0.53}$	$77.29_{\pm 0.17}$	$76.79_{\pm 0.13}$	$81.87_{\pm 0.23}$	$83.01_{\pm 0.11}$	$\mathbf{86.14}_{\pm 0.08}$
	Known	$70.96_{\pm 0.13}$	$64.40_{\pm 0.11}$	$69.37_{\pm 0.23}$	$67.82_{\pm 0.06}$	$\mathbf{70.15}_{\pm 0.05}$	$87.51_{\pm 0.21}$	$89.02_{\pm 0.12}$	$90.77_{\pm 0.10}$	$88.89_{\pm 0.06}$	$\mathbf{91.69}_{\pm 0.31}$
	Un1	$24.04_{\pm 0.12}$	$25.36_{\pm 0.40}$	$35.06_{\pm 0.34}$	$37.87_{\pm 1.59}$	$36.72_{\pm 1.23}$	$23.83_{\pm 0.24}$	$20.19_{\pm 0.55}$	$39.76_{\pm 1.4}$	$54.35_{\pm 0.38}$	$\mathbf{61.17}_{\pm 0.91}$
	Un2	$7.18_{\pm 0.34}$	$16.18_{\pm 0.23}$	$24.21_{\pm 0.67}$	$\mathbf{27.64}_{\pm 2.26}$	$26.36_{\pm 1.98}$	$14.80_{\pm 0.23}$	$16.42_{\pm 0.73}$	$37.93_{\pm 1.3}$	$53.97_{\pm 0.38}$	$\mathbf{59.10}_{\pm 0.59}$
$\rho = 10$	All	$66.21_{\pm 0.23}$	$59.61_{\pm 0.27}$	$65.05_{\pm 0.29}$	$63.21_{\pm 0.05}$	$62.84_{\pm 0.06}$	$83.32_{\pm 0.15}$	$81.98_{\pm 0.03}$	$84.60_{\pm 0.10}$	$83.23_{\pm 0.17}$	$\mathbf{87.44}_{\pm 0.31}$
	Known	$70.36_{\pm 0.31}$	$64.27_{\pm 0.32}$	$69.80_{\pm 0.23}$	$67.82_{\pm 0.07}$	$66.78_{\pm 0.04}$	$89.71_{\pm 0.03}$	$89.24_{\pm 0.02}$	$90.85_{\pm 0.12}$	$87.48_{\pm 0.23}$	$\mathbf{91.51}_{\pm 0.11}$
	Un1	$24.74_{\pm 0.45}$	$26.21_{\pm 0.69}$	$33.01_{\pm 0.34}$	$34.87_{\pm 0.70}$	$34.08_{\pm 0.54}$	$22.17_{\pm 0.18}$	$16.94_{\pm 0.07}$	$26.70_{\pm 0.11}$	$42.62_{\pm 0.50}$	$\mathbf{51.18}_{\pm 2.53}$
	Un2	$7.28_{\pm 0.21}$	$13.01_{\pm 0.34}$	$17.57_{\pm 0.67}$	$21.68_{\pm 0.29}$	$23.52_{\pm 0.32}$	$11.47_{\pm 0.28}$	$10.31_{\pm 0.22}$	$22.89_{\pm 0.27}$	$41.23_{\pm 0.58}$	$\mathbf{47.61}_{\pm 2.59}$

Analysis of α To understand the influence of α, we set $\beta = 2$ (Figure 2) and $\beta = 5$ (Figure 3), respectively, and observed how changes in α affected the accuracy metrics. In contrast to the behavior of β, higher values of α led to an improvement in known accuracy but resulted in reductions in both unknown-aware and unknown-agnostic accuracies.

Significantly, our findings highlight an apparent trade-off between known and unknown accuracies. By optimally tuning these parameters, especially with an $\alpha \approx 1$ when $\beta = 2$ and $\beta = 5$, we can enhance known accuracy without adversely impacting the unknown accuracies.

5. Conclusion

In this work, we turn our attention to the Long-tailed Generalized Category Discovery (Long-tailed GCD), where the unlabeled data predominantly consists of known classes, leading to an inherent imbalance. To counteract this challenge, we proposed a simple and effective method fortified by two regularizations: tail reweighting and a class prior constraint. Through comprehensive experimentation covering both balanced and multiple imbalanced contexts, our method emerges as a formidable contender. It consistently matches or exceeds the performance of leading-edge solutions across a variety of imbalanced settings, underscoring its adaptability and paramount effectiveness in real-world applications.

References

[1] Kaidi Cao, Maria Brbic, and Jure Leskovec. Open-world semi-supervised learning. In *ICLR*, 2020. 1, 2, 3

[2] Kaihua Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in Neural Information Processing Systems*, 32:1565–1576, 2019. 2

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020. 2

[4] Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. A unified objective for novel class discovery. In *ICCV*, 2021. 1

[5] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017. 2

[6] Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Autonovel: Automatically discovering and learning novel visual categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1

[7] Kai Han, Andrea Vedaldi, and Andrew Zisserman. Learning to discover novel visual categories via deep transfer clustering. In *CVPR*, 2019. 1

[8] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual rep-

resentation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020. 2

[9] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020. 2

[10] Mingchen Li, Xuechen Zhang, Christos Thrampoulidis, Jiasi Chen, and Samet Oymak. Autobalance: Optimized loss functions for imbalanced data. *Advances in Neural Information Processing Systems*, 34:3163–3177, 2021. 2

[11] Ziyun Li, Ben Dai, Furkan Simsek, Christoph Meinel, and Haojin Yang. Imbagcd: Imbalanced generalized category discovery. 1, 2, 3

[12] Ziyun Li, Jona Otholt, Ben Dai, Christoph Meinel, Haojin Yang, et al. A closer look at novel class discovery from the labeled set. *arXiv preprint arXiv:2209.09120*, 2022. 1

[13] Seulki Park, Jongin Lim, Younghan Jeon, and Jin Young Choi. Influence-balanced loss for imbalanced visual classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 735–744, 2021. 2

[14] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport. *Center for Research in Economics and Statistics Working Papers*, (2017-86), 2017. 2

[15] Mamshad Nayeem Rizve, Navid Kardan, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. Openldn: Learning to discover novel classes for open-world semi-supervised learning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXI*, pages 382–401. Springer, 2022. 1

[16] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1757–1772, 2012. 1

[17] Yiyou Sun and Yixuan Li. Opencon: Open-world contrastive learning. *Transactions of Machine Learning Research*. 1, 2, 3

[18] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 1, 3