Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/20/20231070

An overview of algorithms for discovering new categories of problems based on unlabeled data

Chuanzhi Tong

The Dalian University of Technology, International School of Information Science&Engineering, Dalian, China, 116620

dlut.tongcz@gmail.com

Abstract. In traditional machine learning, supervised and semi-supervised learning is designed to be used in a closed-world setting where the training data is fixed and does not change over time. Unfortunately, these methods still require a large number of labels for the categories to be categorized, which is expensive and impractical. A new category discovery algorithm is designed so that it can discover new categories while classifying and recognizing labeled images. In this case, the machine can automatically identify new categories without manual marking of image feature categories, which can greatly reduce the cost of image classification. Kai Han et al. named this problem a new category discovery problem and proposed that deep clustering can be used to solve it well. This paper focuses on the comparison of two commonly used robust baselines in the new category discovery and proposes that adding a post-processing model can better improve the accuracy of the model result. This paper applied the relaxed contrast learning method to the Ranking Statistics, and the accuracy of CIFAR-100 is improved by 6%.

Keywords: NCD, GCD, representation learning, RankStats, UNO.

1. Introduction

Imagine a scenario where a baby goes to the zoo to observe animals, some of which their parents have told them in advance what category it is (this is a cat, this is a dog). Of course, there are still a lot of animals that babies don't know (cobras, geckos), and it can't be expected to tell babies about all the categories they haven't seen, because that's a lot of work and expensive to do, so it is expected that babies will see a lot of these new categories and their visual recognition system will automatically recognize them as new categories. That's the problem with this work: provide a data set, but only part of it is labeled with their category tags.

Because of the open-world setting, this has a high application prospect in the real world 0. Imagine in the supermarket, there will be many new products launched for sale every day. If these products have to be manually registered every day, it will be a waste of time and meaningless, because most of the new products are very similar and also some products may not be available the next day. So, an algorithm needs to be designed that can learn from the open data set to determine whether this is an already labeled or completely new category, and if it is a new category, put it into a new category with a simple classification, so that it is not a category that has never been seen before in later learning.

Kai Han et al. named this problem the new category discovery (NCD) problem, and they believe that this kind of problem is a deep transferring clustering problem, they suggest first training a model with

DOI: 10.54254/2755-2721/20/20231070

labeled images 0. Then, the model used the weight of labeled images to annotate the unlabeled images with pseudo-labels to realize the transfer of knowledge. By transferring the knowledge, the model can realize the functionality of simultaneously recognizing new categories and distinguishing old ones. The first step of training labeled data is mainly to use self-supervised learning skills. A common training model is to use a residual network for training because it is stable enough and through the design of residual blocks it can successfully increase the network depth. In the migration clustering in Step 2, this article will compare two strong baselines for solving the NCD problem: RankStats and Unified Objective function (UNO) [4]. RankStats is to find the old class most similar to the new class using the rank statistics method, while UNO finds the new class using the multi-view method.

And to avoid the restrictive assumption that all unlabeled classes come from new categories, Sagar Vaze et al. further proposed the Generalized Category Discovery (GCD) based on the NCD problem [3]. This paper mainly starts from these two questions to carry on the concrete discussion. Through a review of the new category discovery algorithms in recent years, this paper will also put forward some shortcomings and restrictive assumptions for NCD problems and explores the feasible research direction and content.

2. Related work

This paper involves semi-supervised learning and transfer learning. Both of these areas have been extensively studied, and a brief review of them follows.

2.1. Semi-supervision learning

Semi-supervised learning (SSL) assumes that both labeled and unlabeled categories come from the same data set [6]. Its purpose is to use labeled and unlabeled data in the training process to learn robust classification models. In recent years, many robust semi-supervised learning model algorithms have been proposed. And in existing models, more and more people think that consistence-based approaches are more effective and popular, such as Mean-teacher [7].

2.2. Transfer learning

Transfer learning begins by training the model on one tagged data set and then fine-tuning it with another tagged data set containing different categories [8]. Thus, NCD is more akin to transfer learning, because to address the new category of records, this project also needs to transfer tag knowledge from the source dataset to the target ones. It is worth mentioning that in NCD, the target dataset is unlabeled. As deep learning models continue to evolve, the most common form of migration learning today is using one pre-trained model on ImageNet for a specific task of tagging data [9]. In NCD, however, because there is no label available for new category identification, this project needs some robust baseline to achieve category differentiation.

3. Problem description

3.1. Category discovery algorithm

Compared with NCD, GCD is closer to real life, because it is no longer limited to all unlabeled images coming from new categories, so this paper mainly starts from GCD to summarize and expand relevant research. But no matter the GCD or NCD, their basic idea is the same, is to use the ResNet network to pre-train the labeled graph, and then get a pre-training model. The ResNet model’s main idea is to introduce the residual block to replace the original neural network and remove the same body part to highlight small changes through the residual block. Thus, increasing the depth of the network can avoid degeneration problems and ensure the training accuracy of the model.

The second step is to solve the essence of the GCD problem, design a baseline through the transfer clustering method, realize the discovery of new categories, gather similar categories together, and then use the link training method to train labeled pictures and unlabeled pictures at the same time, realize the

DOI: 10.54254/2755-2721/20/20231070

simulation of training in the real open environment. The two most useful baselines are RankStats and UNO.

3.2. RankStats

Because in the open world assumption, the unseen classes will have some degree of visual similarity to the seen classes. Thus, the equal application of learned representations to both old and new classes should be ensured. Based on this assumption, Kai Han et al. designed the RankStats algorithm using a more robust ranking statistical approach. To be Specific, they sort the values in the feature tensor vector of the picture by size. Then, if you get the same ranking for two unlabeled images A and B, these two images are probably in the same(new) category, so specify a rank statistic label S_{ij} , set to 1. Otherwise, set S_{ij} to 0.

Once the similarity labels S_{ij} of the pictures are obtained, they are used to train the comparison function for the unlabeled data as pseudo-labels. In addition, a new category header is designed to extract a new descriptor vector to optimize for the unlabeled data with the binary cross entropy (BCE) loss:

$$L_{BCE} = -\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left[S_{ij}\log\alpha^u\beta(x_i^u)^T\alpha^u\beta\left(x_j^u\right) + \left(1 - S_{ij}\right)\log\left(1 - \alpha^u\beta(x_i^u)^T\alpha^u\beta\left(x_j^u\right)\right)\right] \quad (1)$$

Where M indicates the number of unlabeled images, α^u is a new head to implement the migration to the image representation $\beta(x_i^u)$ and $\beta\left(x_j^u\right)$ to extract a new descriptor vector $\alpha^u\beta(x_i^u)$ and $\alpha^u\beta\left(x_j^u\right)$ to optimize for the unlabeled data. Therefore, the inner product $\alpha^u\beta(x_i^u)^T\alpha^u\beta\left(x_j^u\right)$ is used as a check point for whether x_i^u and x_j^u images belong to the same category or not.

3.3. UNO

Existing methods usually involve designing multiple objective functions to solve the GCD problem for special loss items of labeled and unlabeled samples respectively, and usually require auxiliary regularization items. Like RankStats, UNO moves away from this traditional restriction and introduces a unified objective function to discover new categories to enable synergy between unsupervised and supervised learning. By the multiple views self-labeling technology, UNO generates pseudo-labels that can be treated identically to the ground truth value labels. This results in a single classification target for operations on known and unknown classes.

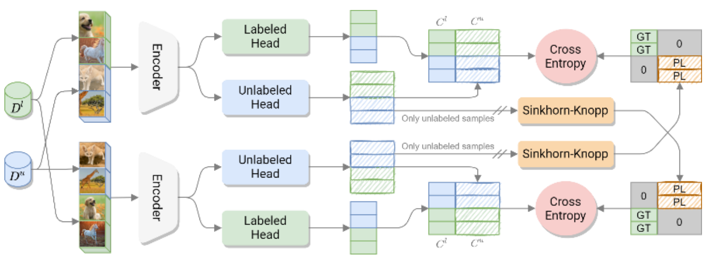

Figure 1. Overview of the UNO architecture [4].

As shown in Figure 1, in green are the "labeled components", in blue are their unlabeled counterparts, and in orange are the pseudo-labeling with its outputs. Sketchiness indicates that the unlabeled images and pseudo-labels are uncertain. The encoder E and the heads (h and g) share parameters for both two views.

4. Optimization algorithm

Based on the research method proposed by Kai Han et al in the auto-novel, they summarize the NCD problem into three steps, so it can better understand the problem and facilitate the search for the following research findings [3]. The first step of training is to learn unbiased image representation by

DOI: 10.54254/2755-2721/20/20231070

self-supervised learning with both labeled and unlabeled data. In this step, a learning representation model with a good shallow layer can be obtained. In the second step, the labelled data set is used to fine-tune the final layers of the resulting model through full-supervised learning; Finally, a robust baseline is used, using fine-tuning representations to induce clustering of unlabeled(unseen) data while still maintaining a good representation on the labelled (seen) set. Based on this, this paper decides to optimize the three steps respectively to improve the final accuracy and results of the algorithm.

4.1. Pretraining model

ResNet (Residual Network) is a popular deep-learning architecture that has been widely used for image classification tasks. Despite its success, there is always room for improvement in terms of the model's accuracy. In this section, this paper will discuss several techniques that can be used to improve the accuracy of ResNet models[10].

One approach to improving ResNet models' accuracy is adding another dense layer before the dense layer in the final. This can increase the depth and the capacity of the model to learn more complex representations of the data, which can help improve its ability to make precision predictions. However, it's important to note that it may also increase the risk of overfitting by adding more layers to a model, so it's important to carefully monitor the model's performance and parameters and use techniques such as regularization to prevent overfitting.

Another approach is to train the ResNet model from scratch. This can help model better capture the underlying patterns in the data and also improve its ability to make precision predictions. However, training a deep learning model from scratch can be time-consuming and computationally expensive. Using heavier data augmentation is another technique that can be used to improve the accuracy of ResNet models. Data augmentation is applying various transformations to the images, such as rotation, scaling, and flipping to artificially increase the size of training dataset. This can help the model better generalize to new data and improve its ability to make accurate predictions.

Experimenting with different learning rates can also help improve the accuracy of ResNet models. The learning rate controls how quickly the model updates its weights in response to the error it makes on the training data. Choosing an appropriate learning rate is important for ensuring that the model converges to a good solution.

4.2. Post-processing model

In addition to the techniques discussed earlier, another approach to improving the accuracy of ResNet models is to use relaxation contrast learning as a post-processing technique. Relaxation contrast learning is a method that involves adjusting the model's predictions to improve its accuracy.

The basic idea behind relaxation contrast learning is to use the model's predictions on a set of validation data to identify areas where the model is making incorrect predictions[11]. The model's predictions are then adjusted to improve its accuracy on these examples. This can be done by applying a relaxation operation to the model's output, which involves adjusting the predicted probabilities to make them more consistent with the true labels. The basic idea of relaxation contrast loss is based on semantic similarity captured as knowledge in the source embedding space, pulling or pushing a pair of samples in the target embedding space. Unlike the original contrast loss, it loosens the binary label indicating the equivalence relation of the class, by using the paired similarity given in the transferred knowledge.

One advantage of using relaxation contrast learning as a post-processing technique is that it can be applied to any pre-trained model without retraining. This makes it a fast and efficient way to improve the accuracy of ResNet models.

4.3. Experience

Through the above method, this paper successfully makes the code in the auto-novel paper get nearly 1% accuracy improvement on CIFAR-10 and nearly 6% accuracy improvement on CIFAR-100 [12].

DOI: 10.54254/2755-2721/20/20231070

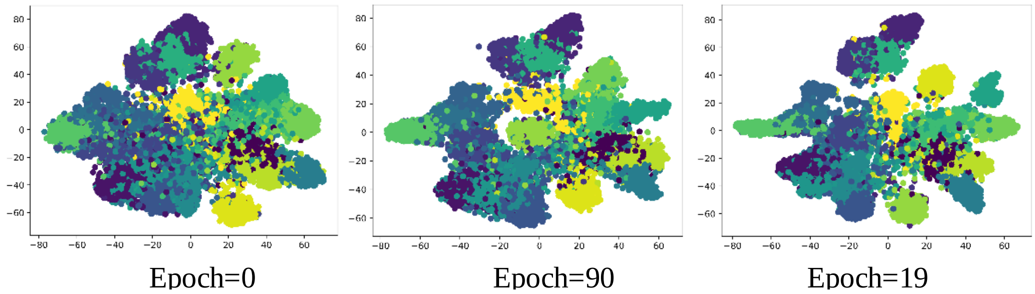

Figure 2. Cluster visualization by t-SNE dimensionality reduction during CIFAR-100 training about auto-novel.

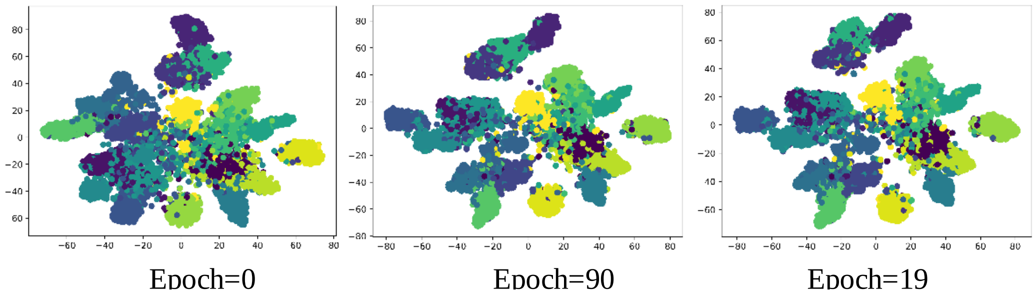

Figure 3. Cluster visualization by t-SNE dimensionality reduction during CIFAR-100 training using relaxation contrast learning to improve the result of auto-novel.

As shown in Figure 2 and Figure 3, through visualization, this paper demonstrates that the auto-novel can well separate 100 categories in cifar100, and the clustering effect is further improved significantly by adding the relaxation contrast learning as the post-processing algorithm. There is no doubt that the training accuracy of the model can be greatly increased by adding post-processing methods, and it also has a good application prospect for NCD problems.

Based on the above description, this paper believes that to solve the GCD problem, the best way is to add a post-processing model, through which the originally robust baseline accuracy can be further improved, which is very helpful for the algorithm research of non-annotated data. Furthermore, compared with designing a new baseline to greatly improve the clustering accuracy of unlabeled images, designing a post-processing model can improve the accuracy of the model better and more efficiently, while still maintaining the setting requirements of open data sets. After the tests, this paper found that relaxed comparative learning is a good post-processing method and can greatly improve the training effect.

5. Conclusion

In conclusion, this paper summarizes and sorts out the problem of finding new categories, and focuses on the comparison of the most applicable two robust baselines: RankStats and UNO. In addition, this paper also believes that compared with finding a new baseline to identify unlabeled new classes, finding a more suitable post-processing algorithm can more effectively improve the accuracy of the original NCD model. This paper also tests the relaxed comparative learning method and finds that it has a good effect.

However, this paper does not propose a new algorithm, and the application of relaxed contrast learning is limited to the auto-novel article, which has no universal applicability. In that case, this paper hopes that other researchers can find better post-processing methods to apply to NCD problems, to further increase the prediction accuracy of the model to the data.

DOI: 10.54254/2755-2721/20/20231070

Therefore, this paper believes that the most effective way to study the problems found in new categories at present is to find a post-processing method to refine the original model, which can not only maintain the setting of open world data set, realize the classification of known classes and learn new classes at the same time but also further improve the progress and achieve better accuracy, which is very valuable.

References

[1] Cao, Kaidi, Maria Brbic, and Jure Leskovec. "Open-world semi-supervised learning." arXiv preprint arXiv:2102.03526 (2021).

[2] Han, Kai, Andrea Vedaldi, and Andrew Zisserman. "Learning to discover novel visual categories via deep transfer clustering." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[3] Han, Kai, et al. "Automatically discovering and learning new visual categories with ranking statistics." arXiv preprint arXiv:2002.05714 (2020).

[4] Fini, Enrico, et al. "A unified objective for novel class discovery." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[5] Vaze, Sagar, et al. "Generalized category discovery." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[6] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. In NeurIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2011.

[7] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In NeurIPS, 2017.

[8] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2010.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In CVPR, 2009.

[10] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[11] Kim, Sungyeon, et al. "Embedding transfer with label relaxation for improved metric learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

[12] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.