Dimensionally Reduced Open-World Clustering: DROWCULA

Erencem Ozbey[1][0009−0009−9281−4856] and Dimitrios I. Diochnos[2][0000−0002−2934−606X]

[1] Bogazici University, Istanbul, Turkey
[2] University of Oklahoma, Norman, OK, USA

Abstract. Working with annotated data is the cornerstone of supervised learning. Nevertheless, providing labels to instances is a task that requires significant human effort. Several critical real-world applications make things more complicated because no matter how many labels may have been identified in a task of interest, it could be the case that examples corresponding to novel classes may appear in the future. Not unsurprisingly, prior work in this so-called 'open-world' context has focused a lot on semi-supervised approaches.
Focusing on image classification, somehow paradoxically, we propose a fully unsupervised approach to the problem of determining the novel categories in a particular dataset. Our approach relies on estimating the number of clusters using Vision Transformers, which utilize attention mechanisms to generate vector embeddings. Furthermore, we incorporate manifold learning techniques to refine these embeddings by exploiting the intrinsic geometry of the data, thereby enhancing the overall image clustering performance. Overall, we establish new State-of-the-Art results on single-modal clustering and Novel Class Discovery on CIFAR-10, CIFAR-100, ImageNet-100, and Tiny ImageNet. We do so, both when the number of clusters is known or unknown ahead of time.
Details omitted due to space limitations are available in [40]. The code is available at: https://github.com/DROWCULA/DROWCULA.

Keywords: Clustering · Novel Class Discovery · Dimension Reduction

1 Introduction

Labeled data are the foundation of supervised learning. In the case of classification, such information captures phenomena as observations are *labeled* by human experts in various domains. This labeling process requires significant human effort and perhaps also to monetary cost. Unsurprisingly, a big portion of machine learning research focuses on developing the most performant models for certain datasets, or, to put this differently, on trying to utilize to the largest extent possible the information available to a learning algorithm in a specific context.

Sometimes, however, this is not the whole story. Many critical applications, e.g., identifying - potentially new - viruses, requires the development of mechanisms that can categorize instances to *novel* classes, leading to the branch of

machine learning algorithms that operate in an *open-world* setting [4,28]. *Semi-supervised learning (SSL)* methods [10,57,17] develop a preliminary model using a small labeled dataset and with the use of unlabeled data a better model is learned. Thus, SSL reduced the human effort for data labeling and, in certain cases, allows the development of algorithms that operate in an open-world scenario where not all labels are known ahead of time.

Such *novel class discovery (NCD) algorithms* utilize the existing labeled data, and their performance is tested by how accurately they can predict novel classes. Even though NCD has been investigated in the area of SSL, any clustering algorithm can also detect novel classes since no labels are needed for clustering. In this sense NCD can be viewed as a more informed problem than clustering, where some labels are available and some classes are known a priori.

NCD in Open-World SSL setting is useful in some scenarios, nevertheless, we find three drawbacks in these lines of work: *(i)* Real-world scenarios ordinarily deal with the emergence of an unknown number of novel classes, *(ii)* Larger datasets capture a bigger portion of the underlying distribution that governs a particular phenomenon, and typically this also translates to a need for a larger portion of the dataset to be labeled (by humans), *(iii)* Despite their unfair advantage, NCD methods fall behind recent clustering algorithms.

Along these lines, we propose a methodology to completely remove the labeling process for NCD and approach this problem as an area of *Open-World Unsupervised Learning (OWUL)*. Our proposed methodology *Dimensionally Reduced Open-World Clustering (DROWCULA)* generalizes to large datasets without any prior information and outperforms previous clustering and NCD methods. DROWCULA leverages the power of Vision Transformers (ViT) and dimension reduction techniques to create powerful vector embeddings and estimate the number of clusters in the meantime. Figure 1 provides a schematic overview of our method.

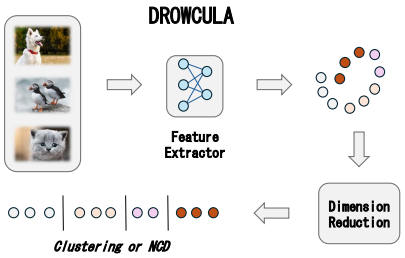

Fig. 1: Overview of DROWCULA (our method). An image dataset is fed into a feature extraction mechanism and in sequence the dimension of the resulting embeddings is reduced, yielding the low-dimensional instances that we ultimately cluster.

Contributions. Our contributions are summarized as follows. First, we propose a *fully unsupervised* setting for NCD in Open-World scenarios. Second, we establish state-of-the-art (SoTA) accuracies for single-modal clustering and NCD by increasing the efficiency of vector embeddings in categorization tasks. Third, for clustering in high-dimensional vector spaces, we investigate non-Euclidean metrics and manifold learning methods, while providing a comparison of vision models and clustering algorithms. Omitted details due to space limitations are available in [40]. Our code is available at `https://github.com/DROWCULA/DROWCULA`.

2 Related Work

Image categorization has been an important problem in the field of Computer Vision. Although supervised and unsupervised machine learning algorithms are broadly investigated, open-world settings have received less attention than the development of closed-set machine learning. In an open-world setting, the number of categories and labels are unknown, hence some of the well-accepted classification and clustering methods fall short in these conditions.

Unsupervised learning has been identified as a primary branch of statistical learning for discovering hidden patterns without needing labeled examples. Traditional clustering algorithms include K-means [2], K-medoids [29], Gaussian Mixture Models (GMM) [43], density-based methods like DBSCAN [48], and more recently deep clustering approaches [9]. Advancement of pretrained models took the image clustering problem one step further, creating powerful methods such as TURTLE, PRCut, PRO-DSC, and TEMI. [20,21,38,1]

The main approaches for closed-world SSL are *pseudo-labeling* [34,41,44], *consistency regularization* [46,51], as well as hybrid approaches [6,5,50]; see [27] for a survey and taxonomy of these and related methods. There is evidence that working in an open-world setting is harder than in a closed-world setting, since novel classes among the unlabeled data result in performance loss; e.g., [11]. Various approaches have tried to deal with this issue; e.g., [22,16,55,8,45].

While SSL makes a lot of sense in closed-world settings, NCD has been conceived as a subfield of machine learning with an open-world setting in mind, right from the start [8]. NCD relies on the use of a labeled set that bootstraps an initial model that can identify classes, but nevertheless, that learned model has the capability of detecting novel classes [24,25,18,56]. It is therefore not surprising that a large body of work in NCD has focused on open-world SSL approaches, such as [45]. Acknowledging these SSL methods, our goal is to discover novel categories in a completely unsupervised manner and, thus, we call this new setting *Open-World Unsupervised Learning (OWUL)*.

Feature extraction and vector embeddings are essential for various computer vision tasks, as they capture the important characteristics of images. Deep learning models with Convolutional Neural Network (CNN) architectures had been the standard way of feature extraction; e.g., [31,26,49]. However, for open-world image clustering or NCD, CNNs become limited due to their dependent nature to the dataset they are trained on. On the other hand, Vision Transformers (ViT) have emerged as a promising alternative for these tasks [15,12]. ViTs leverage attention mechanisms [54], and their ability to handle large datasets makes them suitable for open-world [23]. Utilizing ViTs as foundation models we obtain flexible vector representations and improved performance on various tasks. Hence we propose using ViTs for estimating the number of clusters and performing clustering on entirely unknown datasets.

In what follows, we propose a generalizable OWUL-based methodology leveraging the power of ViTs. Our proposed method *Dimensionally Reduced Open-World Clustering (DROWCULA)* not only eliminates the need for labeled data but also surpasses previously proposed Clustering and NCD methods.

3 Preliminaries

Clustering. We want to cluster images in similar groups. However, instead of clustering images directly, we cluster images based on the *embeddings* that we obtain using some vectorization operation on these images.

K-means is one of these clustering algorithms. K-means relies on initializing random centroids, assigning data points to the closest centroid, and updating these centroids by calculating the mean of the assigned data points in that cluster. Different distance metrics result in different K-means clustering performance. Moreover some metrics may be incompatible with the representation of the instances; which is true also in some of our scenarios. Instances may be lying on a lower-dimensional manifold, which can deceive Euclidean distances between points and adversely affect the clustering performance.

We investigate clustering algorithms beyond K-means; such as K-medoids. Instead of using the mean of the data points to determine centroids, it chooses a real data point for this purpose. These center data points are called *medoids*, and they are updated with different heuristic algorithms; Partitioning Around Medoids (PAM) [42] is one such common implementation. The main problem with PAM is the run-time complexity, and for this reason, in practice, one works with optimized algorithms like FastPAM and FasterPAM [47].

Applying the K-medoids algorithm and then evaluating the Silhouette score is a reasonable methodology; however it is also possible to directly optimize the Silhouette score in the K-medoids algorithm. This technique is known as PAM-SIL [32]; with run-time optimized versions being FastMSC and FasterMSC [35].

In this work, we make our preliminary investigation of distance measures using FasterMSC. Additional details are available in [40].

Performance Measures. In a totally unlabeled scenario, the performance evaluation of the clustering is not as easy as it is in an image classification problem. To overcome this issue, we use internal performance measures for clustering and external measures for evaluation. External measures do not determine the clustering process, but they are still essential to see the real categorization performance in the end. We use Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and Clustering Accuracy (ACC) as the external measures. As internal measures, we use *Cluster Validity Indices (CVI)* [3]. Among many options, we chose to use *Silhouette Score (Sil)* for the main clustering evaluation. Details and experiments with *Calinski-Harabasz (CHI)* [7] - *Davies-Bouldin (DBI)* [13] indices, and an investigation on the Average Clustering Coefficient (C_{avg}) are included in [40].

Dimension Reduction. In Figure 2 we provide a toy illustration on how treating a lower dimensional manifold in a higher space can cause data points that were originally distant to appear closer to each other. Case in point are the blue and red points in Figure 2. This compression may lead to catastrophic effects for clustering. Hence, we investigate different methods for mapping the data into a

meaningful lower-dimensional space. Operating in this space not only improves the clustering quality but also improves the time and space efficiency.

When it comes to dimension reduction, one of the most common approaches is Principal Component Analysis (PCA). PCA finds vectors maximizing the variance of the data. As a linear dimension reduction method, PCA struggles with more complex structures as they cannot be captured linearly. To address these limitations, we propose apply Manifold Learning methods to preserve the underlying geometry of the data.

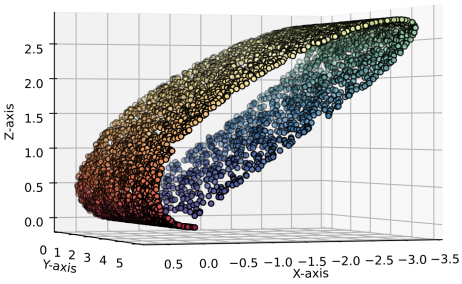

Fig. 2: A sample 2D manifold in 3D space. Points that were originally distant in the 2D manifold (e.g., red and blue) end up being close to each other in the 3D space. Illustration created with artificial data.

Multidimensional Scaling (MDS) is one of these powerful projection methods that preserves the pairwise distance between data points in a lower-dimensional space by applying an eigen decomposition. By operating on Geodesic distance local structure of the manifold can be retained. This technique is called Isomap [52].

Another powerful non-linear method is t-distributed Stochastic Neighbor Embedding (t-SNE) [36]. This method preserves local similarities by modeling them with a Student-t distribution and minimizing KL divergence between high and low-dimensional representations. While t-SNE is mostly used in data visualization, we believe it is a powerful technique to enhance the performance of clustering. One issue with t-SNE is its computationally expensive nature. As an alternative to t-SNE, Uniform Manifold Approximation and Projection (UMAP) [37] can be leveraged. UMAP is a non-linear manifold learning technique, which constructs a KNN graph and optimizes a similar low-dimensional representation. All of these properties of UMAP make it a strong candidate.

4 Preliminary Investigation on Clustering

In this section we investigate key components of the methodology that we propose in Section 5. Thus, the current section provides a justification of our design choices and can also serve as an ablation study among competing metrics or methods that can be utilized for clustering.

In order to overcome the problematic effect of high-dimensional ViT embeddings, we investigate different distance measures in clustering. Since the K-Means algorithm relies on continuous measures, we preferred to use the FasterMSC algorithm in this experiment.

Distances. In Table 1 we present external performance measurements for the FasterMSC clustering results using different distance measures. We use DINOv2 [39] embeddings for the evaluation. Our experiments show that Geodesic distance

performs better than other measures by overcoming the curse of dimensionality. Moreover, we observe that FasterMSC is highly sensitive to normalization.

Since we are expected to put similar points in the same cluster, neighboring points in the space should also be clustered into the same category. The superiority of geodesic distance in Table 1 is related to the prominence of local structure.

Table 1: Impact of Distance Measures on Clustering Performance using FasterMSC at CIFAR-10 Dataset.

Measure	**Norm.**	**ACC**	**NMI**	**ARI**
Euclidean	✗	26.3	38.8	6.3
Euclidean	✓	73.1	72.5	62.4
Manhattan	✗	26.3	38.4	6.2
Manhattan	✓	72.7	71.7	61.4
Chebyshev	✗	41.1	31.6	16.9
Chebyshev	✓	56.6	39.5	34.3
Geodesic	✗	**83.8**	**88.9**	**81.9**

Dimension Reduction. Since we need to overcome the curse of dimensionality while conserving the local structure of the data manifold, we have evaluated several dimension reduction techniques on the same DINOv2 embeddings. After the dimension reduction we obtain an Euclidean space, thus the K-means algorithm is viable in this case.

Table 2: Impact of Dimension Reduction Techniques on Clustering Performance using K-means at CIFAR-10 Dataset.

Method	**Norm.**	**ACC**	**NMI**	**ARI**
None	✗	91.0	90.8	84.3
PCA (30D)	✗	90.8	90.3	83.8
MDS (24D)	✗	90.7	90.1	83.5
Isomap (32D)	✗	88.4	92.5	86.8
AE (16D)	✗	84.6	90.6	83.0
t-SNE (2D)	✗	98.0	95.6	95.7
UMAP (3D)	✗	95.2	93.7	95.3
MDS	✓^{post}	**98.2**	**95.3**	**96.0**
Isomap	✓^{post}	**97.9**	**94.6**	**95.4**
AE	✓^{post}	**99.2**	**97.6**	**98.1**
UMAP	✓^{pre}	**99.1**	**97.6**	**98.1**
t-SNE	✓^{pre}	**98.6**	**96.6**	**96.9**

In Table 2 we present detailed results of dimension reduction techniques. When compared to Table 1, we observe that the K-means algorithm on dimensionally reduced space gives much more robust results than FasterMSC on the initial space. Since we want to obtain a transformed space where the local structure is conserved, non-linear dimension reduction methods improve the external clustering measures significantly. Isomap, UMAP, and t-SNE all somehow use neighboring points for dimension reduction, which is essential for clustering tasks. Unlike these non-linear dimension reduction methods, Autoencoders (AEs) don't directly utilize the intrinsic geometry of the manifold. Hence, vector spaces reduced by AEs are not necessarily well correlated with our evaluation measures; please see [40] for more details.

Feature Extractors. In this part, we make a comparison between feature extractors and show the generalizability of the dimension reduction on different models. Our experiments on various pre-trained feature extractors demonstrate that applying dimension reduction can drastically improve the clustering.

In Figure 3, we observe that t-SNE (2D) enhances the performance of the K-means on all feature extractors. Considering DINOv2-giant's 1536D vectors, our

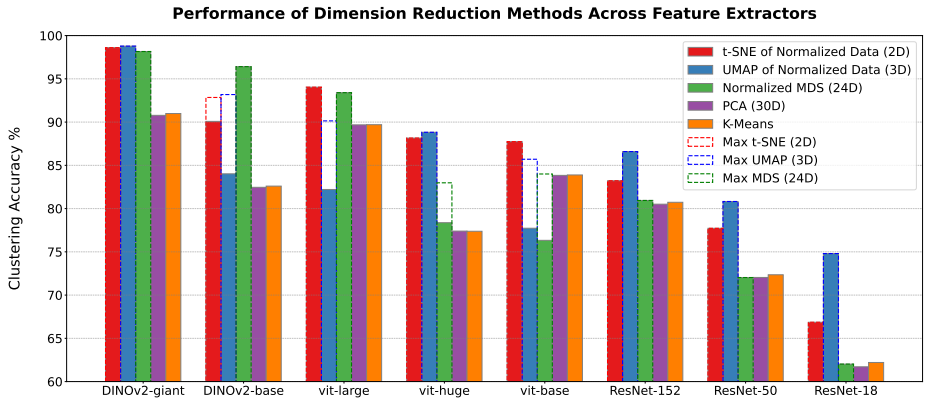

Fig. 3: Performance of Feature Extractors on CIFAR-10 Dataset.

new 2D vectors are not only better in clustering but also substantially better in memory efficiency. Our experiments further reveal that 30D PCA-reduced features are comparable to the original higher-dimensional data in clustering. Although the linear nature of PCA couldn't improve the performance, reducing the number of dimensions saves a significant amount of memory.

Since DINOv2-giant performs better than the other feature extractors among all dimension reduction techniques, we choose to present our proposed methodology using this model. We are aware that most of the pretrained backbones are trained in a supervised/self-supervised way. This latent supervision can also be eliminated by using unsupervised pretrained models.

5 Methodology

In Open-World scenarios we investigate three main tasks: clustering, determining the number of clusters, and NCD.

Clustering. As we discussed in the previous chapters, it is possible to obtain better performing embeddings with the help of non-linear dimension reduction techniques. In light of our preliminary investigations, we construct our base methodology DROWCULA in the following way.

- **Obtain embeddings:** Use a pretrained Vision Transformer (ViT) to extract feature embeddings from input images.
- **L2 normalization:** Apply L2 normalization to each embedding vector.
- **Dimensionality reduction:** Reduce the dimensionality of the normalized embeddings using UMAP or t-SNE.
- **Clustering:** Apply the K-means algorithm on the reduced embeddings to group them into clusters.

To the best of our knowledge, this approach obtains SoTA results in image clustering where vision language models (VLM) are not used adjunctively.

Number of Clusters Estimation. The totally unknown nature of OWUL makes estimating the number of clusters an essential issue. In this part we investigate the usage of DROWCULA in the absence of any prior information.

In Table 2 we showed that directly using DINOv2 extracted features with the K-Means algorithm can lead to an ACC of 91.0. Although high accuracies can be obtained in the original space, the high dimensional nature of the vector embeddings has a catastrophic effect on internal evaluation measures. Applying dimension reduction techniques, results in a significant improvement for these measures; please see [40].

Although we expect the best performance at the original number of clusters of the dataset, in some scenarios, overestimation of the number of clusters creates a better clustering performance. Hence, in Figure 3 we also indicate the best clustering performance of our clustering methods with dashed lines to present their real potential in some cases.

Algorithm 1 explains a generic way of estimating the number of clusters by applying DROWCULA. For simplicity, we define the DINOv2(img) function as a feature extractor and introduce DimReduce(X, method) to apply the chosen dimension reduction method over extracted features.

Algorithm 1: Number of Cluster Estimation

```
Data: Image set Images; min. clusters k_min = 2; max. clusters k_max;
      reduction method method ∈ {UMAP, t-SNE}
Result: Best label assignment BestLabels and silhouette score BestSil
F ← [];
for img ∈ Images do
    F ← F ∪ DINOv2(img);
X ← Normalize(F);
Y ← DimReduce(X, method);
BestSil ← −∞;
BestLabels ← ∅;
for k ← k_min to k_max do
    labels ← KMeans(Y, k);
    sil ← Silhouette(Y, labels);
    if sil > BestSil then
        BestSil ← sil;
        BestLabels ← labels;
return BestLabels, BestSil;
```

This process gives competitive clustering performances compared to knowing the real number of clusters, but computing silhouette scores can be computationally expensive. To maximize Sil in a faster and more robust way, we propose performing *Bayesian Optimization* (BO) [19] on the number of clusters.

Table 4: Accuracy with **Known** Number of Categories. Novel% column indicates the percentage of novel categories in the dataset, and sub-columns Known, Novel, and All indicate the clustering accuracy for the corresponding subset of the testing portion of the data. Most of the results from [45] and [8].

CIFAR-10 [30]				
Method	Novel%	Known	Novel	All
FixMatch [50]	50%	71.5	50.4	49.5
DTC [25]	50%	53.9	39.5	38.3
RankStats [24]	50%	86.6	81.0	82.9
UNO [18]	50%	91.6	69.3	80.5
ORCA [8]	50%	88.2	90.4	89.7
OpenLDN-MixMatch [45]	50%	95.2	92.7	94.0
OpenLDN-UDA [45]	50%	95.7	95.1	95.4
DROWCULA t-SNE	100%	–	98.6	98.6
DROWCULA UMAP	100%	–	**99.1**	**99.1**

CIFAR-100 [30]				
Method	Novel%	Known	Novel	All
FixMatch [50]	50%	39.6	23.5	20.3
DTC [25]	50%	31.3	22.9	18.3
RankStats [24]	50%	36.4	28.4	23.1
UNO [18]	50%	68.3	36.5	51.5
ORCA [8]	50%	66.9	43.0	48.1
OpenLDN-MixMatch [45]	50%	73.5	46.8	60.1
OpenLDN-UDA [45]	50%	74.1	44.5	59.3
DROWCULA t-SNE	100%	–	**81.8**	**81.8**
DROWCULA UMAP	100%	–	80.4	80.4

ImageNet-100 [53,14]				
Method	Novel%	Known	Novel	All
FixMatch [50]	50%	65.8	36.7	34.9
DTC [25]	50%	25.6	20.8	21.3
RankStats [24]	50%	47.3	28.7	40.3
ORCA [8]	50%	89.1	72.1	77.8
OpenLDN-UDA [45]	50%	89.6	68.6	79.1
DROWCULA t-SNE	100%	–	**88.6**	**88.6**
DROWCULA UMAP	100%	–	83.5	83.5

Tiny ImageNet (200) [33]				
Method	Novel%	Known	Novel	All
DTC [25]	50%	28.8	16.3	19.9
RankStats [24]	50%	5.7	5.4	3.4
UNO [18]	50%	46.5	15.7	30.3
OpenLDN-MixMatch [45]	50%	52.3	19.5	36.0
OpenLDN-UDA [45]	50%	58.3	25.5	41.9
DROWCULA t-SNE	100%	–	**78.4**	**78.4**
DROWCULA UMAP	100%	–	77.5	77.5

6 Performance Evaluation

We evaluate DROWCULA on clustering - NCD tasks and compare its performance with previous methods.

To the best of our knowledge, Table 3 includes the best-performing methods in image clustering. Among these methods, TEMI, PRO-DSC, and TURTLE leverage VLM embeddings (CLIP) while PRCut and DROWCULA only use vision-domain features (DINOv2). In this sense, DROWCULA gives SoTA clustering performance in single-modal vision settings without relying on any language supervision.

Table 3: Accuracies of recent clustering algorithms on CIFAR-10 and CIFAR-100 datasets

Method	**CIFAR-10**	**CIFAR-100**
TEMI [1]	96.9	73.7
PRO-DSC [38]	97.2	77.3
PRCut-D [21]	79.1	80.6
TURTLE 2-spaces [20]	**99.5**	**89.9**
DROWCULA t-SNE	**98.6**	81.8
DROWCULA UMAP	**99.1**	80.4

In Table 4, we included previous Semi-Supervised NCD and Unsupervised Learning approaches to compare with our proposed methods DROWCULA t-SNE and DROWCULA UMAP. In these experiments, both of our proposed methods outperformed prior methods in all datasets. Although the difference in ACC between our method and the previous best-performing NCD method (OpenLDN-UDA) is less than 4% on CIFAR-10, our methods outperform OpenLDN-

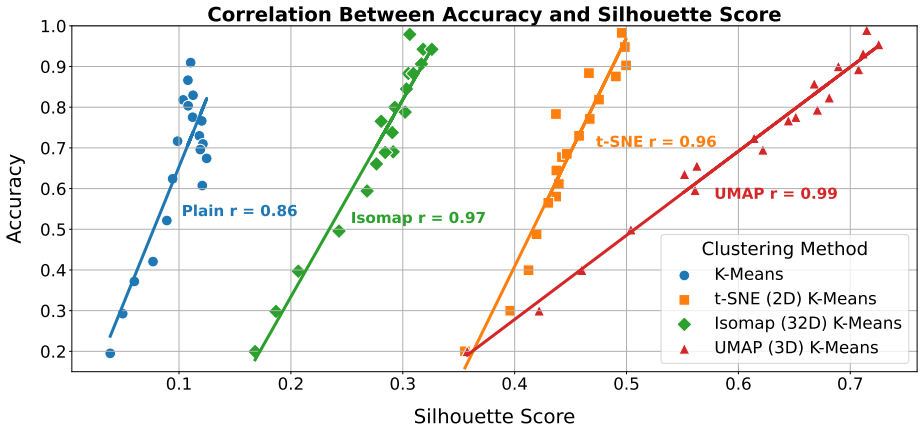

Fig. 4: Effect of Dimension Reduction on the Correlation Coefficients between Silhouette Score and Clustering Accuracy on CIFAR-10 Dataset.

UDA by 74% on CIFAR-100, 30% on ImageNet-100, and by 207% on Tiny ImageNet in terms of Novel Category accuracy.

When the number of categories is unknown, the correlation between these CVIs and our performance evaluation metrics is crucial since we are using CVIs to determine the number of clusters. In Figure 4, we observed that dimension reduction techniques increase the correlation coefficient between Sil and Acc by more than 10% in the CIFAR-10 dataset.

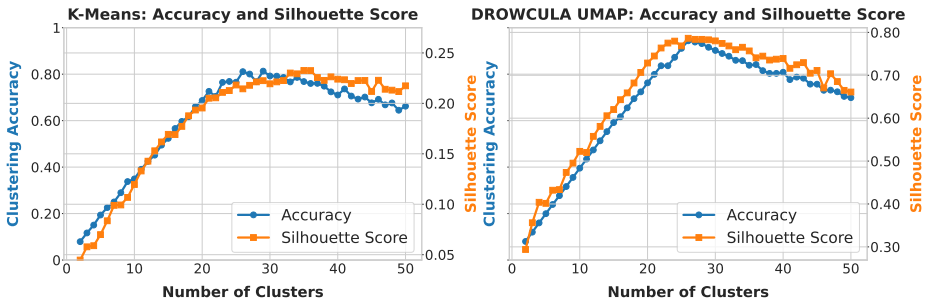

Fig. 5: Effect of the DROWCULA UMAP on clustering accuracy and silhouette score on a random CIFAR-25 dataset. Acc is marked at the left, and the Sil is marked at the right axis, respectively. Both plots are aligned according to the Acc values, but the Sil is scaled to match with Acc for demonstration purposes.

Figure 5 shows the performance of DROWCULA UMAP with the number of cluster estimation method on a 25-cluster subset of the CIFAR-100 dataset.

DROWCULA UMAP obtains its highest Sil value with 25 clusters (94.4% ACC), whereas K-Means takes it with 35 clusters (76.8% ACC).

Table 5 provides a direct comparison between our method and other NCD approaches capable of clustering without prior knowledge of the number of categories. On the CIFAR-100 dataset, our method achieves twice the ACC of ORCA, 2.6 times higher than RankStats, and 4.2 times higher than DTC [8,24,25].

Table 5: Accuracy with **Unknown** Number of Categories. The columns have similar interpretation as in Table 4.

Method	Novel%	**CIFAR-100**		
		Known	Novel	All
DTC [25]	50%	30.7	15.4	14.5
RankStats [24]	50%	33.7	22.1	20.3
ORCA [8]	50%	66.3	40.0	46.4
DROWCULA t-SNE	100%	-	79.0	79.0
DROWCULA UMAP	100%	-	**80.0**	**80.0**

Table 6 shows that BO can speed up the searching process with a relatively small compromise in ACC as we proposed in Section 5. Since the number of iterations for BO is a hyperparameter, the trade-off between ACC and the number of iterations can be externally determined.

7 Conclusion

Image categorization is an important area of machine learning, and we believe that removing human labeling from image categorization can speed up the development of various related tasks, including the development of foundation models.

Table 6: Performance of BO across large datasets with DROWCULA UMAP.

	BO		**Manual**	
Dataset	**ACC**	**#Iter**	**ACC**	**#Iter**
CIFAR-100	75.9	20	**80.0**	200
ImageNet-100	**84.0**	20	83.8	200
Tiny ImageNet	76.5	40	**78.4**	400

In this work we have presented a fully unsupervised methodology for clustering and NCD that leverages ViTs for creating vector embeddings that effectively capture complex features in the data. To address challenges caused by the high-dimensional nature of image embeddings, we employed various dimension reduction methods that not only enhance clustering performance but also improve computational efficiency.

Our comparisons across different datasets demonstrate the superior performance of our approach without having any prior knowledge of the number of clusters. Notably, we establish new SoTA accuracies for single-modal clustering and Novel Class Discovery. We believe that the implications of our work extend beyond clustering, offering a pathway to utilize vast amounts of unlabeled data effectively. By eliminating the human effort, our unsupervised framework may accelerate the development of foundation models and other data greedy tasks in computer vision.

Looking ahead, we intend to further develop our methodology by utilizing other unsupervised learning techniques. Leveraging the features of known datasets, we aim to enhance the quality of our feature representations. We also

believe new NCD algorithms can be directly implemented on this setup, thus expanding the literature in a completely new direction in this field.

Acknowledgments. This work was partially supported by the National Science Foundation under Grant No CISE-25-31450. This support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

This work was also partially supported by a seed fund sponsored by the Institute for Community and Society Transformation (ICAST) at the University of Oklahoma.

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

References

1. Adaloglou, N., Michels, F., Kalisch, H., Kollmann, M.: Exploring the limits of deep image clustering using pretrained models. arXiv preprint arXiv:2303.17896 (2023)
2. Ahmed, M., Seraj, R., Islam, S.M.S.: The k-means algorithm: A comprehensive survey and performance evaluation. Electronics **9**(8), 1295 (2020)
3. Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I.: An extensive comparative study of cluster validity indices. Pattern recognition **46**(1), 243–256 (2013)
4. Bendale, A., Boult, T.E.: Towards Open World Recognition. In: CVPR. pp. 1893–1902. IEEE Computer Society (2015)
5. Berthelot, D., Carlini, N., Cubuk, E.D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C.: ReMixMatch: Semi-Supervised Learning with Distribution Matching and Augmentation Anchoring. In: ICLR (2020)
6. Berthelot, D., Carlini, N., Goodfellow, I.J., Papernot, N., Oliver, A., Raffel, C.: MixMatch: A Holistic Approach to Semi-Supervised Learning. In: NeurIPS. pp. 5050–5060 (2019)
7. Caliński, T., Harabasz, J.: A dendrite method for cluster analysis. Communications in Statistics-theory and Methods **3**(1), 1–27 (1974)
8. Cao, K., Brbic, M., Leskovec, J.: Open-world supervised learning. In: ICLR (2022)
9. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: ECCV (2018)
10. Chapelle, O., Schölkopf, B., Zien, A. (eds.): Semi-Supervised Learning. The MIT Press (2006)
11. Chen, Y., Zhu, X., Li, W., Gong, S.: Semi-Supervised Learning under Class Distribution Mismatch. In: AAAI (2020)
12. Darcet, T., Oquab, M., Mairal, J., Bojanowski, P.: Vision Transformers Need Registers. In: ICLR (2024)
13. Davies, D.L., Bouldin, D.W.: A Cluster Separation Measure. IEEE Trans. Pattern Anal. Mach. Intell. **1**(2), 224–227 (1979)
14. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR (2009)
15. Dosovitskiy, A., et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: ICLR (2021)
16. Du, P., Zhao, S., Sheng, Z., Li, C., Chen, H.: Semi-Supervised Learning via Weight-aware Distillation under Class Distribution Mismatch. In: ICCV (2023)

17. van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Machine Learning **109**(2), 373–440 (2020)
18. Fini, E., Sangineto, E., Lathuilière, S., Zhong, Z., Nabi, M., Ricci, E.: A Unified Objective for Novel Class Discovery. In: ICCV (2021)
19. Frazier, P.I.: A tutorial on bayesian optimization. arXiv preprint arXiv:1807.02811 (2018)
20. Gadetsky, A., Jiang, Y., Brbic, M.: Let go of your labels with unsupervised transfer. arXiv preprint arXiv:2406.07236 (2024)
21. Ghriss, A., Monteleoni, C.: Deep clustering via probabilistic ratio-cut optimization. arXiv preprint arXiv:2502.03405 (2025)
22. Guo, L., Zhang, Z., Jiang, Y., Li, Y., Zhou, Z.: Safe Deep Semi-Supervised Learning for Unseen-Class Unlabeled Data. In: ICML (2020)
23. Gupta, A., Narayan, S., Joseph, K.J., Khan, S., Khan, F.S., Shah, M.: OW-DETR: Open-world Detection Transformer. In: CVPR (2022)
24. Han, K., Rebuffi, S., Ehrhardt, S., Vedaldi, A., Zisserman, A.: Automatically Discovering and Learning New Visual Categories with Ranking Statistics. In: ICLR (2020)
25. Han, K., Vedaldi, A., Zisserman, A.: Learning to Discover Novel Visual Categories via Deep Transfer Clustering. In: ICCV (2019)
26. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
27. Kage, P., Rothenberger, J.C., Andreadis, P., Diochnos, D.I.: A Review of Pseudo-Labeling for Computer Vision. arXiv preprint arXiv:2408.07221v3 (2025)
28. Kardan, N., Stanley, K.O.: Mitigating fooling with competitive overcomplete output layer neural networks. In: IJCNN. pp. 518–525 (2017)
29. Kaufman, L., Rousseeuw, P.J.: Finding groups in data: an introduction to cluster analysis. John Wiley & Sons (2009)
30. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009), toronto, ON, Canada
31. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. NeurIPS (2012)
32. Van der Laan, M., Pollard, K., Bryan, J.: A new partitioning around medoids algorithm. Journal of Statistical Computation and Simulation **73**(8), 575–584 (2003)
33. Le, Y., Yang, X.: Tiny imagenet visual recognition challenge. CS 231N **7**(7), 3 (2015)
34. Lee, D.H.: Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. Workshop on Challenges in Representation Learning, ICML **3**(2), 896 (2013)
35. Lenssen, L., Schubert, E.: Medoid silhouette clustering with automatic cluster number selection. Information Systems **120**, 102290 (2024)
36. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
37. McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
38. Meng, X., Huang, Z., He, W., Qi, X., Xiao, R., Li, C.G.: Exploring a principled framework for deep subspace clustering. arXiv preprint arXiv:2503.17288 (2025)
39. Oquab, M., et al.: DINOv2: Learning Robust Visual Features without Supervision. Trans. Mach. Learn. Res. **2024** (2024)
40. Ozbey, E., Diochnos, D.I.: Dimensionally Reduced Open-World Clustering: DROWCULA. arXiv preprint arXiv:2509.07184 (2025)

41. Pham, H., Dai, Z., Xie, Q., Le, Q.V.: Meta Pseudo Labels. In: CVPR. pp. 11557–11568 (2021)
42. Reynolds, A.P., Richards, G., Rayward-Smith, V.J.: The application of k-medoids and pam to the clustering of rules. In: IDEAL (2004)
43. Reynolds, D.A., et al.: Gaussian mixture models. Encyclopedia of biometrics **741**(659-663), 3 (2009)
44. Rizve, M.N., Duarte, K., Rawat, Y.S., Shah, M.: In Defense of Pseudo-Labeling: An Uncertainty-Aware Pseudo-label Selection Framework for Semi-Supervised Learning. In: ICLR (2021)
45. Rizve, M.N., Kardan, N., Khan, S., Khan, F.S., Shah, M.: OpenLDN: Learning to Discover Novel Classes for Open-World Semi-Supervised Learning. In: ECCV. pp. 382–401 (2022)
46. Sajjadi, M., Javanmardi, M., Tasdizen, T.: Regularization With Stochastic Transformations and Perturbations for Deep Semi-Supervised Learning. In: NeurIPS (2016)
47. Schubert, E., Rousseeuw, P.J.: Fast and eager k-medoids clustering: O (k) runtime improvement of the pam, clara, and clarans algorithms. Information Systems **101**, 101804 (2021)
48. Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X.: Dbscan revisited, revisited: why and how you should (still) use dbscan. ACM Transactions on Database Systems (TODS) **42**(3), 1–21 (2017)
49. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
50. Sohn, K., et al.: FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. In: NeurIPS (2020)
51. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: NeurIPS (2017)
52. Tenenbaum, J.B., Silva, V.d., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)
53. Tian, Y., Krishnan, D., Isola, P.: Contrastive multiview coding. In: ECCV (2020)
54. Vaswani, A., et al.: Attention is All you Need. In: NeurIPS (2017)
55. Yang, Y., Jiang, N., Xu, Y., Zhan, D.: Robust Semi-Supervised Learning by Wisely Leveraging Open-Set Data. IEEE Trans. Pattern Anal. Mach. Intell. **46**(12), 8334–8347 (2024)
56. Zhao, B., Han, K.: Novel Visual Category Discovery with Dual Ranking Statistics and Mutual Knowledge Distillation. In: NeurIPS (2021)
57. Zhu, X., Goldberg, A.B.: Introduction to Semi-Supervised Learning. Morgan & Claypool (2009)