

REVIEW

Open Access



A survey on deep clustering: from the prior perspective

Yiding Lu^{1†}, Haobin Li^{1†}, Yunfan Li¹, Yijie Lin¹ and Xi Peng^{1*}

Abstract

Facilitated by the powerful feature extraction ability of neural networks, deep clustering has achieved great success in analyzing high-dimensional and complex real-world data. The performance of deep clustering methods is affected by various factors such as network structures and learning objectives. However, as pointed out in this survey, the essence of deep clustering lies in the incorporation and utilization of prior knowledge, which is largely ignored by existing works. From pioneering deep clustering methods based on data structure assumptions to recent contrastive clustering methods based on data augmentation invariances, the development of deep clustering intrinsically corresponds to the evolution of prior knowledge. In this survey, we provide a comprehensive review of deep clustering methods by categorizing them into six types of prior knowledge. We find that in general the prior innovation follows two trends, namely, i) from mining to constructing, and ii) from internal to external. Besides, we provide a benchmark on five widely-used datasets and analyze the performance of methods with diverse priors. By providing a novel prior knowledge perspective, we hope this survey could provide some novel insights and inspire future research in the deep clustering community.

Keywords Clustering, Deep clustering, Unsupervised learning

1 Introduction

As a fundamental problem in machine learning, clustering aims at grouping data instances into several clusters, where instances from the same cluster share similar semantics and instances from different clusters are dissimilar. Clustering could reveal the inherent semantic structure underlying the data, which benefits the downstream analysis such as anomaly detection [84], person re-identification [113], community detection [94], and domain adaption [109], etc.

In the early stage, various classic clustering methods are developed, such as centroid-based clustering [62], density-based clustering [19], hierarchical clustering [69],

and so on. These shallow methods are grounded in theory and enjoy high interpretability. Later on, some works extend shallow clustering methods to diverse data types, such as multi-view [73, 74, 96, 115] and graph data [71, 85]. Other efforts have been made to improve the scalability [116] of shallow clustering methods.

However, shallow clustering methods partition instances based on the similarity [62] or density [19] of the given raw or linear transformed data. Due to the limited feature extraction ability, shallow clustering methods would achieve sub-optimal results when confronted with complex, high-dimensional, and non-linear data in the real world. To tackle this challenge, deep clustering techniques are proposed to incorporate neural networks into clustering methods. In other words, deep clustering simultaneously learns discriminative representations and performs clustering on the learned features, progressively benefiting each other.

Over the past few years, many efforts have been devoted to improving the clustering performance from

[†]Yiding Lu and Haobin Li contributed equally to this work.

*Correspondence:

Xi Peng

pengx.gm@gmail.com

¹ College of Computer Science, Sichuan University, Chengdu, Sichuan, China

various aspects, such as network architectures [8, 72], training strategies [67], and loss functions [39, 122]. However, we would like to highlight that the fundamental challenge of deep clustering is the absence of data annotations. Consequently, the key to deep clustering lies in introducing proper *prior knowledge* to construct the supervision signals. From the early data structure assumption to the recent data augmentation invariance, the development of deep clustering methods intrinsically corresponds to the evolution of prior knowledge. In this survey, we provide a comprehensive review of deep clustering methods from the perspective of prior knowledge.

Inspired by traditional clustering and dimensionality reduction approaches [4, 83], the early deep clustering methods [32, 77, 89] build upon the *structure prior* of data. Based on the assumption that the inherent data structure could reflect the semantic relation, these methods incorporate classic manifold [83] or subspace learning [99] objectives to optimize the neural network for feature extraction and clustering. The second type of prior knowledge is the *distribution prior*, which assumes that instances from different clusters follow distinct distributions. Based on such a prior, several generative deep clustering methods [39, 67] propose to learn the latent distribution of samples for the data partition. In the past few years, the success of contrastive learning spawns a new category of prior knowledge, namely, *augmentations invariance*. Instead of mining data priors, researchers turn to constructing additional priors with the data augmentation technique. Leveraging the invariance across different augmented samples at the instance representation and clustering assignment levels, numerous contrastive clustering methods [38, 51] significantly improve the feature discriminability and clustering performance. Further, researchers find that instances of the same semantics are likely to be mapped into nearby points in the latent space, and accordingly propose the *neighborhood consistency* prior. Specifically, by encouraging neighboring samples to have similar cluster assignments, several works [95, 122] alleviate the false-negative problem in the contrastive clustering paradigm, thus advancing the clustering results. Another branch of progress is made based on the *pseudo label* prior, namely, cluster assignments with high confidence are likely to be correct. By selecting confident predictions as pseudo labels, several studies further boost the clustering performance through pseudo-labeling [52, 79] and semi-supervised learning [75]. Very recently, instead of pursuing internal priors from the data itself, some works [7, 53] attempt to introduce abundant *external knowledge* such as textual descriptions to guide clustering.

In summary, the essence of deep clustering lies in how to find and leverage effective prior knowledge, for both feature extraction and cluster assignment. To provide an overview of the development of deep clustering, in this paper, we categorize a series of state-of-the-art approaches according to the taxonomy of prior knowledge. We hope such a new perspective for deep clustering could inspire future research in the community. The rest of this paper is organized as follows: First, Section 2 introduces the preliminaries on deep clustering. Section 3 reviews the existing deep clustering methods from the prior knowledge perspective. Then, Section 4 provides experimental analyses of deep clustering methods. After that, Section 5 briefly introduces some applications of deep clustering in the vicinagearth security. Lastly, Section 6 summarizes some notable trends and challenges for deep clustering.

1.1 Related surveys

We notice that several surveys on deep clustering have been proposed in recent years. Briefly, Min et al. [64] categorizes deep clustering methods according to the network architecture. Dong et al. [18] focuses on applications of deep clustering. Ren et al. [82] summarizes existing methods from the view of data types, such as single- and multi-view. Zhou et al. [123] discusses various interactions between representation learning and clustering. Distinct from existing surveys, this work systematically provides a new perspective from the prior knowledge, which plays a more intrinsic and essential role in deep clustering.

2 Problem definition

In this section, we introduce the pipeline of deep clustering, including the notation and problem definition. Unless specially notified, in this paper, we use bold uppercase and lowercase to denote matrices and vectors, respectively. The commonly used notations are summarized in Table 1.

The deep clustering problem is formally defined as follows: given a set of instances $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N \in \mathcal{X}$ that belongs to C classes, deep clustering aims to learn discriminative features and group the instances into C clusters according to their semantics. Specifically, deep clustering methods first learn a deep neural network $f: \mathcal{X} \rightarrow \mathcal{Z}$ for feature extraction, i.e., $\mathbf{z}_i = f(\mathbf{x}_i)$. Given instance features in the latent space, clustering results could be obtained in two ways. The most straightforward way is to apply classic algorithms such as K-means [62] and DBSCAN [19] on the learned features. The other solution is to train an additional cluster head $h: \mathcal{Z} \rightarrow \mathbb{R}^C$ to produce soft cluster assignment $\mathbf{p}_i = \text{softmax}(h(\mathbf{z}_i))$ which satisfies $\sum_{j=0}^K \mathbf{p}_{ij} = 1$. The hard cluster assignment

Table 1 Commonly used mathematical notations

Notation	Explanation
N	Number of data instances
B	Size of a mini-batch
C	Number of clusters
$f(\cdot)$	Encoder network
$h(\cdot)$	Cluster head
\mathbf{x}_i	i -th data instance
\mathbf{z}_i	Feature of the i -th instance
\tilde{y}_i	Pseudo label of the i -th instance
$\ \cdot\ $	L2-norm of a vector
$\langle \cdot \rangle$	Dot product operator
$s(\mathbf{a}, \mathbf{b})$	Cosine similarity, i.e., $s(\mathbf{a}, \mathbf{b}) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\ \mathbf{a}\ \ \mathbf{b}\ }$
\mathbf{c}_i	Centroid of the i -th cluster
$H(\cdot)$	Entropy, i.e., $H(X) = -\sum_{x \in X} p(x) \log p(x)$
$H(\cdot \cdot)$	Conditional entropy, i.e., $H(Y X) = -\sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)}$
$I(\cdot; \cdot)$	Mutual Information, i.e., $I(X; Y) = H(X) - H(X, Y)$
τ	Temperature coefficient of contrastive loss

for the i -th instance could be computed by arg max operation, namely,

$$\tilde{y}_i = \arg \max_j \mathbf{p}_{ij}, 1 \leq j \leq C. \tag{1}$$

The cluster assignments provide the inherent semantic structure underlying the data, which could be utilized in various downstream analyses.

3 Priors for deep clustering

In this section, we review existing deep clustering methods from the perspective of prior knowledge. The priors are illustrated in Fig. 1 and the method categorization is summarized in Table 2.

3.1 Structure prior

Structure prior is mostly inspired by traditional clustering methods. Traditional cluster is mainly rooted in assumptions about the structural characteristics of clusters in data space. For example, K-means [62] aims to learn k cluster centroids, which assumes that instances in each cluster form a spherical structure around its centroid. DBSCAN [19] is based on the assumption that a cluster in data space is a contiguous region of high point density, separated from other such clusters by regions of low point density. Spectral clustering [4] assumes data lies on a locally linear manifold so that the local neighborhoods' relation should be preserved in latent space. Those methods partition instances according to the graph Laplacian. Agglomerative clustering [24] considers the hierarchical structure of data and performs clustering with merging and splitting. Motivated by the success of classic clustering methods, the early exploration of deep clustering mainly focuses on adapting mature structure priors as objective functions to optimize neural networks.

Given well-structured data in the latent space, ABDC [93] iteratively optimizes the data representation and clustering centers motivated by K-means. As the deep extension of classic spectral clustering,

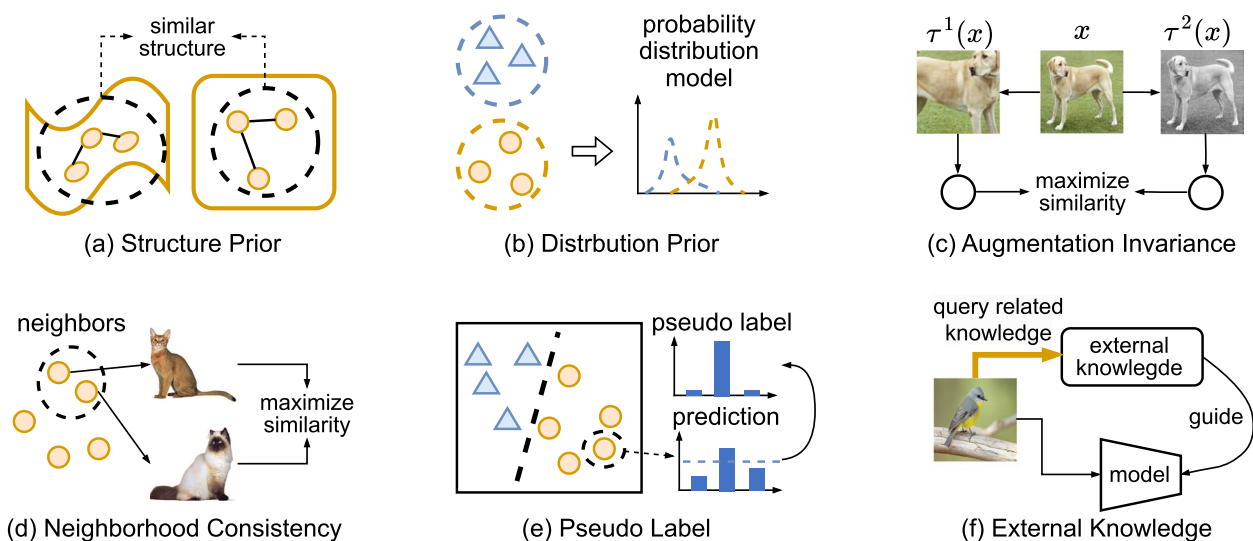


Fig. 1 Six categories of prior knowledge for deep clustering. **a** Structure Prior: data structure could reflect the semantic relation between instances. **b** Distribution Prior: instances from different clusters follow distinct data distributions. **c** Augmentation Invariance: samples augmented by the same instance have similar features. **d** Neighborhood Consistency: neighboring samples have consistent cluster assignments. **e** Pseudo Label: cluster assignments with high confidence are likely to be correct. **f** External Knowledge: abundant knowledge favorable to clustering exists in open-world data and models

DEN [32], SpectralNet [89], and MvLNet [33, 34] compute the graph Laplacian in the latent space learned by auto-encoder [5] and SiameseNets [27, 88], respectively. Likewise, DCC [87] extends the core idea of RCC [86] by performing a relation matching based on the similarity between latent features. The auto-encoder is then optimized by minimizing the distance of paired instances in the latent space. PARTY [77] is the first deep subspace clustering method, which introduces the sparsity prior and self-representation property in subspace learning to optimize neural networks. Motivated by the hierarchical structure of clusters, JULE [108] achieves agglomerative deep clustering by progressively merging clusters and optimizing the features.

3.2 Distribution prior

Distribution prior refers to instances of different semantics following distinct data distributions. Such a prior arouses the generative deep clustering paradigm, which employs variational autoencoder [42] (VAE) and generative adversarial network [23] (GAN) to learn the underlying distribution. Instances generated from similar distributions are then grouped together to achieve clustering.

VaDE [39] is the first deep generative clustering method, which computes different data distributions by fitting the Gaussian mixture model (GMM) in the latent space. To generate an instance, VaDE first samples a cluster distribution $p(c)$ to generate a latent vector $p(z | c)$, and then reconstructs the instance in the input space $p(x | z)$. The cluster assignment and neural network are jointly optimized by maximizing the log-likelihood of instance, i.e.,

$$\log p(x) = \log \int_z \sum_c p(x | z)p(z | c)p(c)dz. \quad (2)$$

Since directly computing Eq. 2 is intractable, the optimization is approximated by the evidence lower bound (ELBO) of variational inference objective, namely,

$$\mathcal{L} = \mathbb{E}_{q(z,c|x)} \left[\log \frac{p(x, z, c)}{q(z, c | x)} \right], \quad (3)$$

where $q(z, c | x)$ is variational posterior, which approximates the real posterior. The reparameterization trick introduced in VAE [42] is adopted to make the sampling process differentiable.

Though GMM could effectively distinguish distributions, Gaussian components are proved to be redundant, which harms the discriminability between different clusters [26]. As an improvement, ClusterGAN, DCGAN [67, 80] proposes to adopt GAN to implicitly learn the latent distributions. Specifically, as shown in Fig. 2, in addition

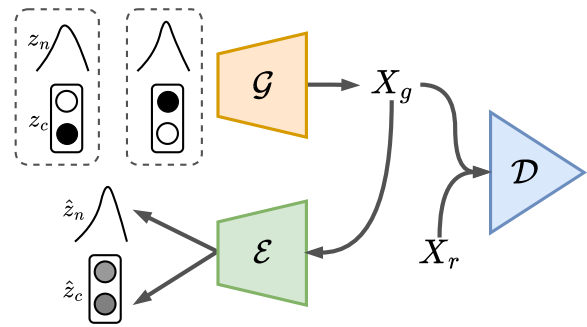


Fig. 2 The framework of distribution prior based methods. In addition to the standard continuous latent variable \mathbf{z}_n , generative deep clustering methods further introduce a discrete variable \mathbf{z}_c to capture the cluster information

to the continuous latent variable \mathbf{z}_n , it introduces a one-hot encoding \mathbf{z}_c to capture cluster distribution during the generation. The objective function of ClusterGAN is formulated as follows:

$$\begin{aligned} \mathcal{L} = & \mathbb{E}_{\mathbf{x} \sim p_X(\mathbf{x})} q(\mathcal{D}(\mathbf{x})) + \mathbb{E}_{\mathbf{z} \sim \mathbb{P}_z} q(1 - \mathcal{D}(\mathcal{G}(\mathbf{z}))) \\ & + \beta_n \mathbb{E}_{p_Z(\mathbf{z})} \|\mathbf{z}_n - \mathcal{E}(\mathcal{G}(\mathbf{z}_n))\|_2^2 \\ & + \beta_c \mathbb{E}_{p_Z(\mathbf{z})} \mathcal{H}(\mathbf{z}_c, \mathcal{E}(\mathcal{G}(\mathbf{z}_c))), \end{aligned} \quad (4)$$

where $\mathbf{z} = (\mathbf{z}_n, \mathbf{z}_c)$ is the mixed latent variable, \mathcal{E} is the inverse network which maps data from the raw to latent space, $\mathcal{H}(\cdot, \cdot)$ denotes the cross-entropy, and β_n, β_c are the weight parameters. The first two terms are consistent with standard GAN. The last two clustering-specific terms encourage a more distinct cluster distribution, as well as map inputs to the latent space to achieve clustering.

3.3 Augmentation invariance

In recent years, image augmentation methods [91] have gained widespread attention, grounded in the prior that augmentations of the same instance could preserve consistent semantic information. This augmentation-invariance character inspires exploration of how to leverage the positive pairs (i.e., different augmentations of the same image) with similar semantic information, as shown in Fig. 3. Notably, mutual-information-based methods and contrastive-learning-based methods have emerged as pioneers in the realm of deep clustering. In this section, we delve into the fundamental concepts and related works of both mutual-information-based and contrastive-learning-based methods.

Firstly, mutual information is a measure of dependence between two continuous random variables X and Y , formally,

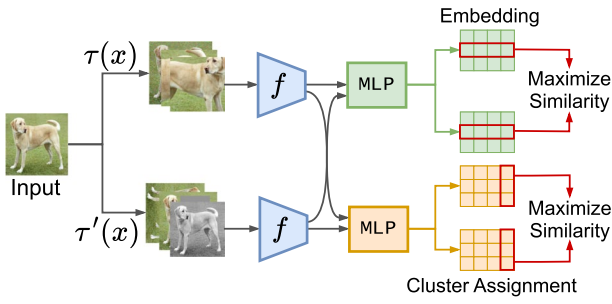


Fig. 3 The framework of augmentation invariance based methods. Diverse transformations are first applied to augment the input data x , after which the shared deep neural network is utilized to extract features. The augmented samples of the same instance are encouraged to have similar features and cluster assignments

$$I(X; Y) = \int_Y \int_X p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy, \quad (5)$$

where $p(x, y)$ is the joint probability mass function of X and Y , $p(x)$ and $p(y)$ are the marginal probability mass functions of X and Y respectively. In the context of information theory, leveraging the mutual information between variables of positive instances could enhance the optimization of clustering-related information.

IMSAT [30] stands as a typical information-theoretic approach to deep clustering. Its fundamental concept includes enforcing invariance on pair-wise augmented instances and achieving unambiguous and uniform cluster assignments. Specifically, IMSAT encourages the representations of augmented instances to closely match the representations of the original instances, i.e.,

$$\mathcal{L} = - \sum_{i,k} \mathbf{p}_{ik} \log \mathbf{p}'_{ik} \quad (6)$$

where \mathbf{p}' is the prediction representations of augmented instances. This aspect can be viewed as exploring the maximization of mutual information between data and its augmentations. Besides, IMSAT implements regularized information maximization for deep clustering inspired by RIM [43] to keep the cluster assignments unambiguous and uniform. Specifically, IMSAT seeks to maximize the mutual information between instances and their cluster assignments, expressed as:

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y | X) \\ &= - \sum_k \mathbf{p}_{\cdot k} \log \mathbf{p}_{\cdot k} + \frac{1}{N} \sum_{i,k} \mathbf{p}_{ik} \log \mathbf{p}_{ik}, \quad (7) \end{aligned}$$

where $H(\cdot)$ and $H(\cdot|\cdot)$ the entropy and conditional entropy, and $\mathbf{p}_{\cdot k} = \frac{1}{N} \sum_i \mathbf{p}_{ik}$. Increasing the first term (marginal entropy $H(Y)$) encourages uniform cluster assignments, i.e., the number of instances in each cluster tends to be the same. Conversely, decreasing the second term (conditional entropy $H(Y | X)$) encourages each instance to be unambiguously assigned to a certain cluster.

IIC [38] and Completer [56, 57] take a further step in exploring the mutual information between instances and their augmentations. The fundamental concept is to maximize the mutual information between the cluster assignments of pair-wise augmented instances. Specifically, IIC achieves semantically meaningful clustering and avoids trivial solutions by maximizing the mutual information between the cluster assignments,

$$\begin{aligned} \mathcal{L} &= I(Z, Z') = \sum_i I(\mathbf{z}_i, \mathbf{z}'_i) = I(\mathbf{P}), \\ &= \sum_{c=1}^C \sum_{c'=1}^C \mathbf{P}_{cc'} \cdot \ln \frac{\mathbf{P}_{cc'}}{\mathbf{P}_c \cdot \mathbf{P}_{c'}}, \quad (8) \end{aligned}$$

where \mathbf{z} and \mathbf{z}' are the representations of the original instance x and its augmentation x' , respectively. The conditional joint distribution of \mathbf{z} and \mathbf{z}' is given by the matrix $\mathbf{P} \in \mathbb{R}^{C \times C}$ which is constituted by,

$$\mathbf{P} = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \cdot (\mathbf{z}'_i)^\top, \quad (9)$$

where $\mathbf{P}_{cc'} = P(z = c, z' = c')$ denotes the element of c -th row and c' -th column. Additionally, the marginals $\mathbf{P}_c = P(z = c)$ and $\mathbf{P}_{c'} = P(z' = c')$ can be obtained by summing over the rows and columns of this matrix. Notably, IIC stands out as one of the earliest deep frameworks designed entirely under the framework of information theory, distinguishing itself from IMSAT.

Similar to mutual-information-based methods, contrastive-learning-based methods treat instances augmented from the same instance as positive samples and the rest as negative samples. Let \mathbf{z}_{2i} and \mathbf{z}_{2i-1} represent two augmented representation of the i -th instance, the contrastive loss is formulated as:

$$\begin{aligned} \mathcal{L} &= \sum_i^N (\ell(2i, 2i-1) + \ell(2i-1, 2i)), \\ \ell(i, j) &= - \log \frac{\exp(s(\mathbf{z}_i \cdot \mathbf{z}_j)/\tau)}{\sum_{j=1}^{2N} \mathbf{1}_{j \neq i} \exp(s(\mathbf{z}_i \cdot \mathbf{z}_j)/\tau)}, \quad (10) \end{aligned}$$

where $\ell(i, j)$ represents the pairwise contrastive loss and τ controls the temperature of the softmax. The function $s(\mathbf{z}_i, \mathbf{z}_j)$ denotes the similarity between representations \mathbf{z}_i and \mathbf{z}_j . This loss encourages representations of positive instances to be closer while being separated from negative instances, encouraging meaningful clustering patterns.

Notably, some theoretical works [58, 66, 76] have demonstrated that contrastive learning is equivalent to maximizing the mutual information from the instance level. Motivated by this observation, researchers have further explored the application of contrastive loss at the cluster level, proving beneficial for deep clustering. PICA [31] is one of the pioneer works of this domain. The fundamental concept behind it is to maximize the similarity between the cluster assignment of original and augmented data. This objective can be likened to conducting contrastive learning [59] at the cluster level. Motivated by PICA, CC [51] and DRC [121] conduct contrastive learning on both instance level and cluster level. Specifically, cluster-level contrastive loss helps learn discriminative cluster assignment, which is the key to the clustering task. Formally, the cluster-level contrastive loss is,

$$\mathcal{L} = \frac{1}{2C} \sum_{i=1}^C (\ell(2i-1, 2i) + \ell(2i, 2i-1)) - H(\mathbf{Y}),$$

$$\ell(i, j) = -\log \frac{\exp(s(\mathbf{y}_i, \mathbf{y}_i)/\tau)}{\sum_{j=1}^{2C} \mathbf{1}_{j \neq i} [\exp(s(\mathbf{y}_i, \mathbf{y}_j)/\tau)]}, \tag{11}$$

where $\mathbf{y}_i \in \mathbb{R}^{1 \times N}$ is the cluster-level assignment and τ is the cluster-level temperature parameter. $H(\mathbf{Y}) = H(\mathbf{Y}^1) + H(\mathbf{Y}^2)$ is the cluster assignment probabilities entropy of two augmentations. The inclusion of $H(\mathbf{Y})$ helps avoid the trivial solution where most instances are assigned to the same cluster. Notably, the utilization of contrastive learning at the cluster level in CC and DRC has inspired subsequent works in the field.

TCC [90] takes a step further in exploring the interaction between instance-level and cluster-level representations. The core idea is to leverage a unified representation combined of the cluster semantics and instances, enhancing the representation with cluster information to facilitate clustering tasks. Formally, for an instance representation \mathbf{z}_i , the enhanced representation is given by:

$$\hat{\mathbf{z}}_i = (\mathbf{z}_i + \text{NN}_\theta(\mathbf{c}_i)) / \|\mathbf{z}_i + \text{NN}_\theta(\mathbf{c}_i)\|_2, \tag{12}$$

where \mathbf{c}_i represents the cluster assignment of i -th instance after Gumbel Softmax. $\text{NN}_\theta(\cdot)$ denotes a single fully connected network, which is the learnable

cluster representation. Different from CC which performs contrastive loss on cluster assignment, TCC conducts contrastive loss on the unified representation to better capture cluster semantics. Inspired by TCC, some works [49, 106] explore the fusion of instance-level and cluster-level representation in various domains. and then conduct contrastive loss on the unified representation, which further explores its effectiveness.

3.4 Neighborhood consistency

Thanks to the advancements in self-supervised representation learning, the features acquired through discriminative pretext tasks can unveil high-level semantics in the latent space. This provides a crucial prior for clustering, as instances and their neighborhoods in the latent space are likely to belong to the same semantic cluster. Leveraging neighborhood-consistent semantics can further enhance clustering, as shown in Fig. 4.

SCAN [95] first observes that similar instances will be mapped closely in latent space through self-supervised pretext tasks. Motivated by this observation, SCAN trains a cluster head based on the cluster neighborhood consistency within neighbors. Specifically, SCAN first obtains an encoder f by a pretext task [22, 29, 102, 117]. It then optimizes a cluster head h by requiring it to make consistent predictions between instances and their nearest neighbors:

$$\mathcal{L} = -\frac{1}{B} \sum_{i=1} \sum_{j \in \mathcal{N}_i^k} \log(\mathbf{p}_i, \mathbf{p}_j) - \lambda H(Y). \tag{13}$$

Here \mathcal{N}_i^k denotes the k -nearest neighbors of the i -th instance. The second term in Eq. 13 prevents h from assigning all instances to a single cluster which is also used in Eq. 11.

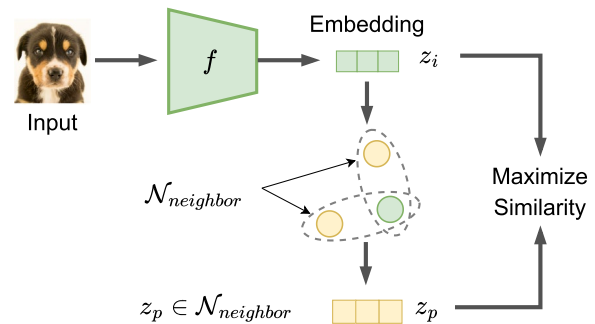


Fig. 4 The framework of neighborhood consistency-based methods. Such a paradigm encourages neighboring samples z_i and z_p in the latent space to have consistent features and cluster assignments, which improves the compactness of clusters

NNM [15] and GCC [122] incorporate neighbor information into the framework of contrastive learning to group instances within neighborhoods. In particular, NNM aligns the clustering assignment of an instance with its neighbors through cluster-level contrastive learning:

$$\mathcal{L} = -\frac{1}{C} \sum_{i=1}^C \log \frac{\exp(s(\mathbf{q}_i, \mathbf{q}_{N(i)}))}{\sum_{j=1}^C \exp(s(\mathbf{q}_i, \mathbf{q}_j))}, \quad (14)$$

where $\mathbf{q}, \mathbf{q}_N \in \mathbb{R}^{C \times B}$ represent the transpose matrix of \mathbf{p} and \mathbf{p}_N , respectively. In contrast, GCC introduces the graph structure of the latent space to modify the vanilla instance-level contrastive loss. It constructs a normalized symmetric graph Laplacian \mathbf{L} based on the K -nn graph:

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}, \quad (15)$$

with $\mathbf{A}_{ij} = \begin{cases} 1, & \text{if } j \in \mathcal{N}_i^k \text{ or } i \in \mathcal{N}_j^k \\ 0, & \text{otherwise} \end{cases}$.

Then, the loss function is given by the following form:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\sum_{\mathbf{L}_{ij} < 0} -\mathbf{L}_{ij} \exp(s(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{\mathbf{L}_{ij} = 0} \exp(s(\mathbf{z}_i, \mathbf{z}_j)/\tau)}, \quad (16)$$

where τ is the temperature. The Graph Laplacian guides the model to attract instances within neighborhoods rather than just augmentation of themselves so that the influence of potential false negative samples [110, 112] can be mitigated. As a result, GCC can better minimize the intra-cluster variance and maximize the inter-cluster variance. The success of this approach has inspired numerous contrastive learning methods [37, 61] in various domains to leverage neighbor relationships that effectively address the false negative challenge.

3.5 Pseudo-labeling

As a prevalent paradigm of semi-supervised classification [6, 47, 92], pseudo-labeling has been extended to deep clustering in recent years. The fundamental assumption of pseudo-labeling is that the predictions on unlabeled data, especially the confident ones, can provide reliable supervision and guide model training. Motivated by this, recent deep clustering works leverage confident predictions to boost clustering performance.

DEC [104] is a pioneering work that utilizes labels generated by itself to simultaneously enhance feature representations and optimize clustering assignments. DEC initializes with a pre-trained auto-encoder and C learnable cluster centroids. The soft assignment is calculated using the Student's t -distribution, based on the distance between the representation \mathbf{z}_i and centroid \mathbf{c}_j :

$$\mathbf{q}_{ij} = \frac{(1 + \|\mathbf{z}_i - \mathbf{c}_j\|^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_k (1 + \|\mathbf{z}_i - \mathbf{c}_k\|^2/\alpha)^{-\frac{\alpha+1}{2}}}, \quad (17)$$

where α is the hyper-parameter and \mathbf{q}_{ij} denotes the probability of assigning the instances i to the cluster j . DEC refines the clusters by emphasizing the high-confidence assignments and making predictions more confident. Specifically, DEC uses the second power of \mathbf{q}_i as a sharpened assignment to guide the training, i.e.,

$$\mathbf{p}_{ij} = \frac{\mathbf{q}_{ij}^2/\text{freq}_j}{\sum_k \mathbf{q}_{ik}^2/\text{freq}_k}, \quad (18)$$

where $\text{freq}_j = \sum_i \mathbf{q}_{ij}$ is the soft cluster frequency and the sharpened assignment is normalized by f_j to prevent feature collapse. Finally, a KL divergence loss between \mathbf{p} and \mathbf{q} minimizes the distances between the two distributions, i.e., $\mathcal{L} = \text{KL}(\mathbf{p}|\mathbf{q})$.

Another notable method of pseudo-labeling is DeepCluster [8]. As illustrate in Fig. 5, this approach employs K -means clustering on the learned representations to obtain cluster assignments as pseudo-labels. DeepCluster iteratively performs representation learning and clustering

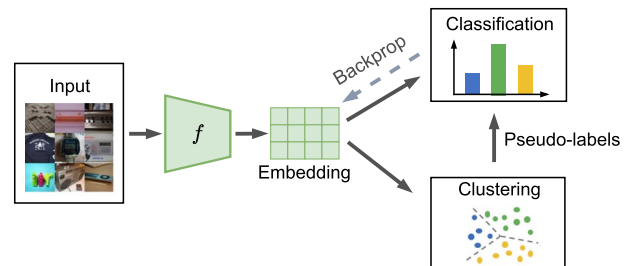


Fig. 5 The framework of pseudo-labeling based methods. Given features in the latent space, clustering algorithms such as K -means are performed to get pseudo labels. The pseudo labels, usually filtered by confidence, are then used as supervision signals to guide clustering

in a mutually beneficial manner to bootstrap each other. However, DeepCluster faces limitations in achieving outstanding performance, primarily due to the restricted semantics of the initial representation. Similar to DeepCluster, ProPos [35] proposes an EM framework of pseudo-labeling, iteratively performing K-means to obtain pseudo labels (E step) and the representation updating (M step). Notably, ProPos significantly outperforms DeepCluster and other methods because ProPos performs K-means on the learned feature of state-of-the-art self-supervised paradigm BYOL [25]. This observation has demonstrated that the semantics of the representation is vital to pseudo-label generation and clustering. Low-quality features would introduce potential noise in pseudo-labels, impact subsequent pseudo-label generation, and mislead representation learning, which accumulates the error in the process.

In addition to the progression of self-supervised paradigms, researchers are actively investigating strategies to alleviate the issue of error accumulation in pseudo-labeling. To be specific, the challenges in the realm of pseudo-labeling deep clustering remain two-fold: enhancing the accuracy of generating pseudo-labels and maximizing the utility of these pseudo-labels for effective clustering. On the one hand, inaccurate pseudo-labels pose a risk of degradation in clustering performance. On the other hand, determining how to effectively leverage these pseudo-labels for clustering is a critical consideration. These two challenges underscore the ongoing efforts in the pseudo-labeling learning of deep clustering.

The first challenge has been addressed by many works through carefully designing selection methods. For instance, SCAN [95] empirically observed that instances exhibiting highly confident predictions (i.e., $\max(\mathbf{p}_i) \approx 1$) tend to be correctly clustered by the cluster head. Building on this insight, SCAN opts to choose instances with the most confident predictions as labeled data to fine-tune the model using the cross-entropy loss,

$$\mathcal{L} = \frac{1}{|Y|} \sum_{i \in Y} -\tilde{y}_i \log(\mathbf{p}_i), \quad (19)$$

$$Y = \{i \mid \text{conf}_i \geq \eta\}, \text{ with } \text{conf}_i = \max(\mathbf{p}_i)$$

where η is the threshold hyper-parameter to filter the uncertain instances. TCL [52] and SPICE [75] have devised more effective selection strategies to enhance the accuracy of pseudo-labeling. Specifically, TCL selects the most confident predictions as pseudo labels from each cluster c :

$$Y^c = \{\text{topK}(\text{conf}_i) \mid \tilde{y}_i = c\}$$

$$Y = \bigcup_{c=1}^C Y^c \quad (20)$$

where $\text{topK}(\cdot)$ returns the indices of the top K confident instances and \bigcup denotes the union of the pseudo labels from all clusters. Here $K = \gamma N / C$ and γ is the selection ratio. The cluster-wise selection leads to more class-balanced pseudo labels compared to threshold-based criteria. It improves the clustering performance, especially for challenging classes.

SPICE introduces a prototype-based pseudo-labeling approach. Specifically, it first re-computes the centroids of each cluster only using the instances with confident predictions, then re-assign each instance with new pseudo labels according to the similarity to the new centroids, formally:

$$\mathbf{c}'_i = \frac{1}{|Y^c|} \sum_{i \in Y^c} \mathbf{z}_i, \quad (21)$$

$$\tilde{y}'_i = \arg \max_j s(\mathbf{z}_i, \mathbf{c}'_j).$$

This operation helps mitigate the influence of potentially incorrect pseudo labels used in calculating centroids, which might accumulate errors in the iterative self-training process.

To address the second challenge, i.e., better utilizing the confident labels, TCL removes negative pairs with the same label in contrastive loss, preventing intra-class instances from pushing apart, i.e., the false negative issue. Meanwhile, SPICE and TCL adopt some semi-supervised classification techniques like FixMatch [92] that impose the pseudo-label consistency for strong augmentations of the same instance. The marvelous results achieved by these works show the effectiveness of combining reliable pseudo-labeling methods and semi-supervised paradigms in clustering.

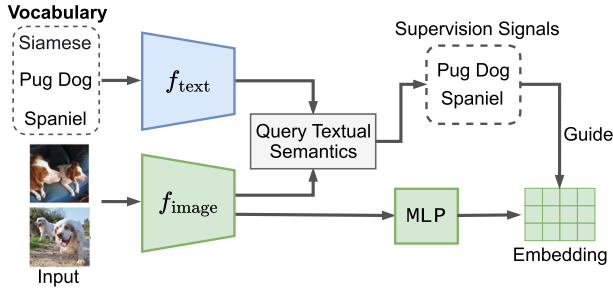


Fig. 6 The framework of external knowledge based methods. Instead of mining internal priors from the samples themselves, such a paradigm seeks external information like textual semantics to help distinguish the given samples

3.6 External knowledge

Most clustering approaches focus on grouping data based on inherent characteristics such as structural priors, distribution priors, and augmentation invariance priors. As shown in Fig. 6, instead of pursuing internal priors from the data itself, some recent works [7, 53] attempt to introduce abundant external knowledge such as textual descriptions to guide clustering. These methods prove effective because the utilization of semantic information from natural language offers valuable supervisory signals that enhance the quality of clustering.

SIC [7] is one of the first works in incorporating external knowledge guidance into clustering. The fundamental concept revolves around generating image pseudo-labels from a textual space pre-trained by CLIP [81]. The process involves three main steps: i) *Construction of Semantic Space*: SIC selects meaningful texts resembling category names to build a semantic space. ii) *Pseudo-labeling*: Pseudo-labels are generated using text semantic centers \mathbf{h} and image representations \mathbf{z}_i , formally,

$$\mathbf{q}_i = \text{one-hot} \left(c, \arg \max_l \frac{\exp(\mathbf{z}_i^T \mathbf{h}_l)}{\sum_{l'}^c \exp(\mathbf{z}_i^T \mathbf{h}_{l'})} \right), \quad (22)$$

where c is the number of semantic centers, \mathbf{h}_l is the l -th center of semantic centers, one-hot operator will generate a c -bit one-hot vector. The pseudo-labels is utilized to guide the clustering similar to SCAN [95],

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n CE(\mathbf{q}_i, \mathbf{p}_i), \quad (23)$$

where $CE(\cdot)$ is the cross entropy function. iii) *Consistency learning*: Enhancing clustering effect by enforcing the consistency between the images and their neighbors in the image space,

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^n \log \mathbf{p}_i^T \mathbf{p}_j, \quad (24)$$

where j is an instance index randomly selected from the nearest neighbors $\mathcal{N}_k(\mathbf{z}_i)$ of i -th instance. Note that, SIC essentially pulls image embeddings closer to embeddings in semantic space, while ignoring the improvement of text semantic embeddings.

TAC [53] focuses on leveraging textual semantics to enhance the feature discriminability. Specifically, it retrieves a text counterpart among representative nouns for each image, which improves K-means performance without any additional training. Besides, TAC proposes a mutual distillation paradigm to incorporate the image and text modalities, which further improves the clustering performance. The cross-modal mutual distillation strategy is formulated as follows:

$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^C \mathcal{L}_i^{v \rightarrow t} + \mathcal{L}_i^{t \rightarrow v}, \\ L_i^{v \rightarrow t} &= -\log \frac{\exp(\text{sim}(\hat{\mathbf{q}}_i, \hat{\mathbf{p}}_i^N)/\tau)}{\sum_{k=1}^K \exp(\text{sim}(\hat{\mathbf{q}}_i, \hat{\mathbf{p}}_k^N)/\tau)}, \\ L_i^{t \rightarrow v} &= -\log \frac{\exp(\text{sim}(\hat{\mathbf{p}}_i, \hat{\mathbf{q}}_i^N)/\hat{\tau})}{\sum_{k=1}^K \exp(\text{sim}(\hat{\mathbf{p}}_i, \hat{\mathbf{q}}_k^N)/\tau)}, \end{aligned} \quad (25)$$

where τ is the softmax temperature parameter, $\hat{\mathbf{p}}_i, \hat{\mathbf{q}}_i \in \mathbb{R}^{1 \times N}$ is the i -th column of image and text assignment matrix, $\hat{\mathbf{p}}_i^N, \hat{\mathbf{q}}_i^N \in \mathbb{R}^{1 \times N}$ is the i -th column of image and text random nearest neighbor matrix. The mutual distillation strategy has two advantages. On the one hand, it generates more discriminative clusters through cluster-level contrastive loss. On the other hand, it encourages consistent clustering assignments between each sample and its cross-modal neighbors, which bootstraps the clustering performance in both modalities.

Table 2 The summary of deep clustering methods from the perspective of prior knowledge

Prior Knowledge		Method	Major Contribution
Structure Prior	Inherent data structure reflect semantic relation	ABDC [93]	optimize features and clustering assignment in an EM manner
		DEN [32], SpectralNet [89]	extend spectral clustering from shallow to deep
		PARTY [77]	introduce the sparsity prior from subspace learning to deep clustering
		JULE [108]	extend agglomerative clustering from shallow to deep
		DCC [87]	propose relation matching to achieve non-parametric deep clustering
Distribution Prior	Instances of different semantics follow distinct data distribution	VaDE [39]	learn distinct cluster distributions by Gaussian mixture model
		ClusterGAN [67], DCGAN [80]	implicitly learn cluster distribution with GAN
Augmentation Invariance	Instance features are invariant to data augmentation	IMSAT [30]	propose the invariance between pair-wise augmented samples
	Cluster assignments are invariant to data augmentation	IIC [38], Completer [56]	propose the mutual information framework with respect to augmentation invariance
		PICA [31]	explore invariance between cluster assignments of augmented samples
		CC [51], DRC [121]	simultaneously explore augmentation invariance at instance and cluster level
		TCC [90]	leverage a unified representation combined of the cluster semantics and instances
Neighborhood Consistency	Neighboring instances have similar semantics	SCAN [95]	impose consistent cluster assignments between neighboring instances
		NNM [15]	perform cluster-level contrastive learning between neighbors
		GCC [122]	perform instance- and cluster-level contrastive learning between neighbors
Pseudo Label	Cluster assignments with high confidence are reliable	DEC [104], DeepCluster [8]	construct target cluster distribution via sharpening
		SCAN [95]	generate pseudo labels with K-means
			select high-confident predictions and finetune the model with strong augmented samples
		SPICE [75]	select pseudo labels with the help of prototypes and adopt semi-supervised learning to fine-tune the model
		TCL [52]	use pseudo labels to mitigate false negative pairs in contrastive learning
External Knowledge	Abundant clustering-favorable knowledge exists in open world	ProPos [35]	use pseudo label from K-means to increase cluster compactness
		SIC [7], TAC [53]	generate image pseudo labels from the textual space from pre-trained vision-language model construct more discriminative text counterparts and perform cross-modal distillation to improve clustering

4 Experiment

In this section, we introduce the evaluation of deep clustering. Briefly, we first present the evaluation metrics and common benchmarks. Then we analyze the results of the existing deep clustering methods.

4.1 Evaluation metrics

For clustering evaluation, three metrics are commonly used to measure how the predicted cluster assignments \tilde{y} match the ground truth labels y , including accuracy (ACC), normalized mutual information (NMI), and adjusted rand index (ARI). A higher value of the metrics corresponds to better clustering performance. The definitions of the three metrics are as follows:

- ACC [1] indicates the correct rate of clustering predictions:

$$\text{ACC} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{y_i = \tilde{y}_i\}, \quad (26)$$

where the Hungarian matching [45] is first applied to align the predictions and labels.

- NMI [63] quantifies the mutual information between the predicted labels $\tilde{\mathbf{Y}}$ and ground truth labels \mathbf{Y} :

$$\text{NMI} = \frac{I(\tilde{\mathbf{Y}}; \mathbf{Y})}{\frac{1}{2}[H(\tilde{\mathbf{Y}}) + H(\mathbf{Y})]}, \quad (27)$$

where $H(\mathbf{Y})$ denotes the entropy of Y and $I(\tilde{\mathbf{Y}}; \mathbf{Y})$ denotes the mutual information between $\tilde{\mathbf{Y}}$ and \mathbf{Y} .

- ARI [36] is the normalization of the rand index (RI), which counts the number of instances pairs in the same cluster and different clusters:

$$\text{RI} = \frac{\text{TP} + \text{TN}}{C_N^2}, \quad (28)$$

where TP and TN refer to the number of true positive pairs and true negative pairs, C_N^2 is the number of possible instance pairs. ARI is computed by adding the following normalization:

$$\text{ARI} = \frac{\text{RI} - \mathbb{E}(\text{RI})}{\max(\text{RI}) - \mathbb{E}(\text{RI})}, \quad (29)$$

where $\mathbb{E}(\text{RI})$ denotes the expectation of RI.

4.2 Datasets

In the early stage, deep clustering methods are evaluated on relatively small and low-dimensional datasets (e.g. COIL-20 [70], YaleB [21]). Recently, with the rapid development of deep clustering methods, it has become more popular to evaluate clustering performance on more complex and challenging datasets. There are five widely used benchmark datasets:

- CIFAR-10 [44] consists of 60,000 colored images from 10 different classes including airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck.
- CIFAR-100 [44] contains 100 classes grouped into 20 superclasses. Each image comes with a “fine” class label and a “coarse” superclass label.
- STL-10 [13] contains 13,000 labeled images from 10 object classes. Besides, it provides 100,000 unlabeled images for self-supervised learning to enhance the clustering performance.
- ImageNet-10 [9] is a subset of the ImageNet dataset [17]. It contains 10 classes, each with 1,300 high-resolution images.
- ImageNet-Dog [9] is another subset of ImageNet. It consists of images belonging to 15 dog breeds, which is suitable for fine-grained clustering tasks.

Apart from them, some recent works employ two more challenging large-scale datasets, Tiny-ImageNet [48] and ImageNet-1K [17], to evaluate the effectiveness and efficiency. A brief description of these datasets is summarized in Table 3.

Table 3 A summary of datasets commonly used for deep clustering

Dataset	Split	Samples	Classes	Image Size
CIFAR-10	Train+Test	60,000	10	32×32
CIFAR-100	Train+Test	60,000	20	32×32
STL-10	Train+Test	13,000	10	96×96
ImageNet-10	Train	13,000	10	96×96
ImageNet-Dogs	Train	19,500	15	96×96
Tiny-ImageNet	Train	100,000	200	64×64
ImageNet-1K	Train	1,281,167	1000	224×224

Table 4 Clustering performance on five widely-used image clustering datasets. SCAN* denotes the clustering results using only neighborhood consistency loss without the self-labeling step. † denotes using the train and test split for training and testing respectively, instead of using both splits for training and testing. Horizontal lines separate methods with different priors. From top to bottom are structure prior, distribution prior, augmentation invariance, neighborhood consistency, pseudo-labeling, and external knowledge

Method	CIFAR-10			CIFAR-100			STL-10			ImageNet-10			ImageNet-Dogs		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
K-means [62]	22.9	8.7	4.9	13.0	8.4	2.8	19.2	12.5	6.1	24.1	11.9	5.7	-	-	-
JULE [108]	27.2	19.2	13.8	13.7	10.3	3.3	27.7	18.2	16.4	30.0	17.5	13.8	13.8	5.4	2.8
DCGAN [80]	31.5	26.5	17.6	15.1	12.0	4.5	29.9	22.7	16.2	31.3	18.6	14.2	17.8	9.8	7.3
IIC [38]	61.7	51.3	41.1	25.7	22.5	11.7	59.6	49.6	39.7	-	-	-	-	-	-
PICA [31]	69.6	59.1	51.2	33.7	31.0	17.1	71.3	61.1	53.1	87.0	80.2	76.1	35.3	35.2	20.1
CC [51]	79.0	70.5	63.7	42.9	43.1	26.6	85.0	76.4	72.6	89.3	85.9	82.2	42.9	44.5	27.4
TCC [90]	90.6	79.0	73.3	49.1	47.9	31.2	81.4	73.2	68.9	89.7	84.8	82.5	59.5	55.4	41.7
SCAN* [95]	81.8	71.2	66.5	42.2	44.1	26.7	75.5	65.4	59.0	-	-	-	-	-	-
NNM† [15]	83.7	73.7	69.4	45.9	48.0	30.2	76.8	66.3	59.6	-	-	-	58.6	60.4	44.9
GCC [122]	85.6	76.4	72.8	47.2	47.2	30.5	78.8	68.4	63.1	90.1	84.2	82.2	52.6	49.0	36.2
DEC [104]	30.1	25.7	16.1	18.5	13.6	5.0	35.9	27.6	18.6	38.1	28.2	20.3	19.5	12.2	7.9
DeepCluster [8]	37.4	-	-	-	-	-	33.4	-	-	18.9	-	-	-	-	-
SCAN† [95]	87.6	78.7	75.8	48.3	48.5	31.4	81.8	70.3	66.1	-	-	-	59.3	61.2	45.7
SPICE [75]	83.8	73.4	70.5	46.8	44.8	29.4	90.8	81.7	81.2	92.1	82.8	83.6	64.6	57.2	47.9
TCL [52]	88.7	81.9	78.0	53.1	52.9	35.7	86.8	79.9	75.7	89.5	87.5	83.7	64.4	62.3	51.6
ProPos [35]	94.3	88.6	88.4	61.4	60.6	45.1	86.7	75.8	73.7	95.6	89.6	90.6	74.5	69.2	62.7
SIC† [7]	92.6	84.7	84.4	58.3	59.3	43.9	98.1	95.3	95.9	98.2	97.0	96.1	69.7	69.0	55.8
TAC [53]	92.3	84.1	83.9	60.7	61.1	44.8	98.2	95.5	96.1	99.2	98.5	98.3	83.0	80.6	72.2

4.3 Performance comparisons

The clustering performance on five widely used datasets is shown in Table 4. Thanks to the feature extraction ability of deep neural networks, early deep clustering methods based on structure and distribution priors achieve much better performance than the classic K-means. Then, a series of contrastive clustering methods significantly improve the performance by introducing additional priors through data augmentation. After that, more advanced methods boost the performance by further considering the neighborhood consistency (GCC compared with CC) and utilizing pseudo labels (SCAN compared with SCAN*). Notably, the performance gains of different priors are independent. For example, ProPos remarkably outperforms DEC and CC by additionally utilizing the augmentation invariance or pseudo-labeling priors, respectively. Very recently, external-knowledge-based methods achieved state-of-the-art performance, which proves the promising prospect of such a new deep clustering paradigm. In addition, clustering becomes more challenging when the category number grows (from CIFAR-10 to CIFAR-100) or the semantics becomes more complex (from CIFAR-10 to ImageNet-Dogs). Such results indicate that more challenging datasets such as full ImageNet-1K are expected to benchmark in future works.

5 Application in Vicinagearth

In this section, we explore some typical applications of deep clustering within the domain of Vicinagearth, a term crafted from the fusion of "Vicinage" and "Earth." Vicinagearth represents the critical spatial expanse ranging from 1,000 meters below sea level (the depth at which sunlight ceases to penetrate) to 10,000 meters above sea level (the typical cruising altitude of commercial aircraft). This zone is of great importance as it encompasses the core regions of human activity including areas of habitation and production. Recently, deep clustering has emerged as an indispensable analytical tool within Vicinagearth, instrumental in unveiling complex patterns and structures of data within the vicinal space. The diverse applications of deep clustering in this zone include anomaly detection, environmental monitoring, community detection, person re-identification, and more.

Anomaly Detection, also known as *Outlier Detection* [14] or *Novelty Detection* [19], attempts to identify abnormal instances or patterns. In the context of Vicinagearth, deep clustering proves valuable for analyzing sensor data obtained from diverse sources like underwater monitoring systems, aerial sensors, or ground-based sensors [10]. Through the analysis of the patterns and typical behaviors from the sensor data, the system

becomes adept at detecting anomalies, which may signal security threats or irregular activities.

Environmental Monitoring involves the analysis of data collected from environmental sensors [103], such as monitoring air quality, water conditions, and geological factors. The primary goal is to ensure the health of ecosystems [101] and detect potential environmental threats, such as pollution events or natural disasters. Deep clustering techniques play a crucial role in grouping similar environmental patterns, facilitating the identification of abnormalities. This application contributes to real-time environmental monitoring [46], enhancing the ability to respond promptly to environmental challenges.

Community Detection [20, 40] involves evaluating how groups of nodes are clustered or partitioned and their tendency to strengthen or break apart within a network. In the context of Vicinagearth, this technique is applied to identify groups of species [68] that interact closely or share similar ecological niches. Deep clustering plays a pivotal role in the analysis of complex ecological networks [65], contributing to a deeper understanding of ecological communities and their dynamics.

Person Re-identification [100, 113] is a crucial task that involves recognizing and matching individuals across different camera views [111]. This technology plays a significant role in public safety and law enforcement initiatives, as it helps to monitor densely populated areas for including potential threats or subjects on the watchlist. The integration of deep clustering algorithms has remarkably improved the scalability and efficiency [107] of person re-identification systems. Deep clustering effectively enables the management of the complexities presented by large and dynamically changing crowds. Furthermore, the adaptability of deep clustering techniques broadens their use to include the monitoring of natural habitats and the tracking of wildlife in diverse and uncontrolled settings.

6 Future challenges

Although existing works achieve remarkable performance, some practical challenges and emerging requirements have yet to be fully addressed. In this section, we delve into some future directions of modern deep clustering.

6.1 Fine-grained clustering

The objective of fine-grained clustering is to discern subtle and intricate variations within data, which is particularly advantageous in research like the identification of biological subspecies [54, 55]. The primary challenge is that fine-grained classes exhibit a high degree of similarity, where distinctions often lie in coloration, markings, shape, or other subtle characteristics. In such scenarios,

traditional coarse-grained clustering priors frequently prove inadequate. For instance, color and shape augmentations in augmentation invariance prior would become ineffective. Recently, C3-GAN [41] employs contrastive learning within adversarial training to generate lifelike images, enabling the nuanced capture of fine-grained details and ensuring the separability between clusters.

6.2 Non-parametric clustering

Many clustering methods typically require a predefined and fixed number of clusters. However, real-world datasets often present a challenge with an unknown number of clusters, reflecting situations closer to reality. Only a few works [11, 87, 98, 120] have been devoted to solving this problem. These methods often rely on calculating global similarity and introduce huge computational costs, especially in large-scale datasets. Therefore, efficiently determining the optimal value of cluster number C remains an open challenge, often involving the incorporation of human priors. Among existing works, DeepDPM introduces Dirichlet Process Gaussian Mixture Models (DPGMM) [3] that utilize the Dirichlet Process as the prior distribution over mixture components. DeepDPM dynamically adjusts the number of clusters C through split and merge operations guided by the Metropolis-Hastings framework [28].

6.3 Fair clustering

Collecting Real-world datasets from diverse sources with various acquisition methods can enhance the generalization of machine learning models. However, these datasets frequently manifest inherent biases, notably in sensitive attributes such as gender, race, and ethnicity. These biases would introduce disparities among individuals and minority groups, leading to cluster partitions that deviate from the underlying objective characteristics of the data. The pursuit of fairness is particularly pertinent in applications where unbiased and equitable analyses are crucial, such as employment, healthcare, and education. To tackle this challenge, fair clustering seeks to mitigate the influence of these biases given the biased attributes for each sample.

To address this daunting task, [12] first introduces a data pre-processing method known as fairlet decomposition. Recent advancements address this issue on large-scale data through adversarial training [50] and mutual information maximization [114]. Notably, [114] designs a novel metric that assesses both clustering quality and fairness from the perspective of information theory. Despite these developments, there is still room for improvement, and the establishment of better evaluation metrics is a continuing area of this research.

6.4 Multi-view clustering

Multi-view data [60, 105] is common in real-world situations where information is captured from a variety of sensors or observed from multiple angles. This data is inherently rich, offering diverse yet consistent information. For example, an RGB view would provide color details while the depth view reveals spatial information, which represents the complementary aspects of the views. Simultaneously, there exists a level of view consistency as the same object possesses common attributes across different views. To deal with multi-view data, multi-view clustering [16, 60] is proposed to exploit both the complementary and consistent characters. The goal is to integrate information from all views to produce a unified and insightful clustering result.

Over recent years, several deep-learning approaches [2, 78, 97, 119] have been developed to address this challenge. Binary multi-view clustering [118] simultaneously refines binary cluster structures alongside discrete data representations, ensuring cohesive clustering. In pursuit of view consistency, Lin et al. [56, 57] maximize mutual information across views, thus aligning common properties. SURE [112] aims to strengthen the consistency of shared features between views by utilizing robust contrastive loss. Recently, Li et al. [49] performs bound contrastive loss to preserve the view complementary at the cluster level. These innovative methodologies demonstrate the significant strides made in the field of multi-view analysis, where clustering continues to play a pivotal role in enhancing the synergistic exploitation of multi-view data.

7 Conclusion

The key to deep clustering or unsupervised learning is to seek effective supervision to guide representation learning. Different from traditional taxonomies from the network structure or data type, this survey offers a comprehensive review from the perspective of prior knowledge. With the evolution of clustering technologies, there is a discernible trend shifting from exploring priors within the data itself to external knowledge like natural language guiding. The exploration of external pre-trained models like ChatGPT or GPT-4V(ision) might emerge as a promising avenue. This survey potentially provides some valuable insight and inspires further exploration and advancements in deep clustering.

Authors' contributions

All authors contributed to the core insights presented in this paper. Xi Peng supervised this survey and provided valuable guidance throughout the process. Yiding Lu, Haobin Li, Yunfan Li, and Yijie Lin collaboratively wrote Priors for Deep Clustering. Yiding Lu took the lead in crafting Introduction, Application, and Future Challenges. Haobin Li was responsible for collecting and analyzing experimental results, creating figures, and summarizing tables. Yunfan Li and Yijie Lin designed the outline, wrote Abstract, and refined the manuscript.

Funding

This work was supported in part by NSFC under Grant 62176171 and in part by the Fundamental Research Funds for the Central Universities under Grant CJ202303.

Data availability

The datasets utilized in this survey are publicly available and can be accessed from the following sources:

- CIFAR-10 and CIFAR-100: <https://www.cs.toronto.edu/~kriz/cifar.html>.
- STL-10: <https://cs.stanford.edu/~acoates/stl10/>.
- ImageNet-10 and ImageNet-Dogs: [Google Drive](#) (Preprocessed versions)
- Tiny-ImageNet: <http://cs231n.stanford.edu/tiny-imagenet-200.zip>.
- ImageNet-1K: <https://www.image-net.org/>.

Code availability

Not applicable.

Declarations

Competing interests

The authors declare no competing interests.

Received: 27 February 2024 Revised: 1 March 2024 Accepted: 7 March 2024

Published online: 25 June 2024

References

1. E. Amigó, J. Gonzalo, J. Artilles et al., A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Inf. Retr.* **12**, 461–486 (2009)
2. G. Andrew, R. Arora, J. Bilmes et al., Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning*. PMLR, vol 28 (Atlanta, GA, USA, 17–19 June 2013), pp. 1247–1255
3. C.E. Antoniak, Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *Ann. Stat.* **2**(6), 1152–1174 (1974)
4. M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering. *Adv. Neural Inf. Process. Syst.* **14**, 585–591 (2001)
5. Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013)
6. D. Berthelot, N. Carlini, I. Goodfellow et al., Mixmatch: A holistic approach to semi-supervised learning. *Adv. Neural Inf. Process. Syst.* **32**, 5050–5060 (2019)
7. S. Cai, L. Qiu, X. Chen et al., Semantic-enhanced image clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol 37. (Washington, DC, USA, 7–14 February 2023), pp. 6869–6878
8. M. Caron, P. Bojanowski, A. Joulin et al., Deep clustering for unsupervised learning of visual features. In *Computer Vision – ECCV 2018*, ed. by V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss. Lecture Notes in Computer Science. vol. 11218 (Springer, Cham, 2018), pp. 139–156
9. J. Chang, L. Wang, G. Meng et al., Deep adaptive image clustering. In *2017 IEEE International Conference on Computer Vision (ICCV)*. (Venice, Italy, 22–29 October 2017), pp. 5880–5888
10. A. Chatterjee, B.S. Ahmed, IoT anomaly detection methods and applications: A survey. *Internet Things* **19**, 100568 (2022)
11. G. Chen, Deep learning with nonparametric clustering. arXiv preprint (2015) arXiv:150103084. <http://arxiv.org/abs/1501.03084>
12. F. Chierichetti, R. Kumar, S. Lattanzi et al., Fair clustering through fairlets. *Adv. Neural Inf. Process. Syst.* **30**, 5029–5037 (2017)
13. A. Coates, A. Ng, H. Lee, An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings*. JMLR, vol 15 (Fort Lauderdale, FL, USA, 11–13 April 2011), pp. 215–223
14. D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 603–619 (2002)

15. Z. Dang, C. Deng, X. Yang et al., Nearest neighbor matching for deep clustering. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Nashville, TN, USA, 20-25 June 2021), pp. 13688–13697
16. C. Deng, Z. Lv, W. Liu et al., Multi-view matrix decomposition: A new scheme for exploring discriminative information. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*. (Buenos Aires, Argentina, 25–31 July 2015), pp. 3438–3444
17. J. Deng, W. Dong, R. Socher et al., ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. (Miami, FL, USA, 20-25 June 2009), pp. 248–255
18. S. Dong, P. Wang, K. Abbas, A survey on deep learning and its applications. *Comput. Sci. Rev.* **40**, 100379 (2021)
19. M. Ester, H.P. Kriegel, J. Sander et al., A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD-96 Proceedings*. (AAAI Press, Portland, Oregon, USA 1996), pp. 226–231
20. S. Fortunato, Community detection in graphs. *Phys. Rep.* **486**(3–5), 75–174 (2010)
21. A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 643–660 (2001)
22. S. Gidaris, P. Singh, N. Komodakis, Unsupervised representation learning by predicting image rotations. In *6th International Conference on Learning Representations, ICLR 2018*, (Vancouver, BC, Canada, 30 April–3 May 2018) <https://openreview.net/forum?id=S1v4N2iO->
23. I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27**, 2672–2680 (2014)
24. K.C. Gowda, G. Krishna, Agglomerative clustering using the concept of mutual nearest neighbourhood. *Pattern Recogn.* **10**(2), 105–112 (1978)
25. J.B. Grill, F. Strub, F. Alché et al., Bootstrap your own latent—a new approach to self-supervised learning. *Adv. Neural Inf. Process. Syst.* **33**, 21271–21284 (2020)
26. S. Gurumurthy, R. Kiran Sarvadevabhatla, R. Venkatesh Babu, DeLiGAN: Generative adversarial networks for diverse and limited data. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI, USA, 21–26 July 2017), pp. 4941–4949
27. R. Hadsell, S. Chopra, Y. LeCun, Dimensionality reduction by learning an invariant mapping. In *Proceedings of the 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)* (New York, NY, USA, 17–22 June 2006), pp. 1735–1742
28. W.K. Hastings, Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57**(1), 97–109 (1970)
29. K. He, H. Fan, Y. Wu et al., Momentum contrast for unsupervised visual representation learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA, USA, 13–19 June 2020), pp. 9726–9735
30. W. Hu, T. Miyato, S. Tokui et al., in *ICML'17: Proceedings of the 34th International Conference on Machine Learning*. Learning discrete representations via information maximizing self-augmented training. *PMLR*, vol 70 (Sydney, NSW, Australia, 6–11 August 2017), pp. 1558–1567
31. J. Huang, S. Gong, X. Zhu, Deep semantic clustering by partition confidence maximisation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA, USA, 13–19 June 2020), pp. 8846–8855
32. P. Huang, Y. Huang, W. Wang et al., Deep embedding network for clustering. In *2014 22nd International conference on pattern recognition* (Stockholm, Sweden, 24–28 August 2014), pp. 1532–1537
33. Z. Huang, J.T. Zhou, X. Peng et al., Multi-view spectral clustering network. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19* (Macao, China, 10–16 August 2019), pp. 2563–2569. <https://doi.org/10.24963/ijcai.2019/356>
34. Z. Huang, J.T. Zhou, H. Zhu et al., Deep spectral representation learning from multi-view data. *IEEE Trans. Image Process.* **30**, 5352–5362 (2021)
35. Z. Huang, J. Chen, J. Zhang et al., Learning representation for clustering via prototype scattering and positive sampling. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(6), 7509–7524 (2022). <https://doi.org/10.1109/TPAMI.2022.3216454>
36. L. Hubert, P. Arabie, Comparing partitions. *J. Classif.* **2**, 193–218 (1985)
37. T. Huynh, S. Kornblith, M.R. Walter et al., Boosting contrastive self-supervised learning with false negative cancellation. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (Waikoloa, HI, USA, 3–8 January 2022), pp. 986–996
38. X. Ji, J.F. Henriques, A. Vedaldi, Invariant information clustering for unsupervised image classification and segmentation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul, Korea, 27 October–2 November 2019), pp. 9864–9873
39. Z. Jiang, Y. Zheng, H. Tan et al., Variational deep embedding: An unsupervised and generative approach to clustering. In *IJCAI'17: Proceedings of the 26th International Joint Conference on Artificial Intelligence* (Melbourne, Australia, 19–25 August 2017), pp. 1965–1972. <https://doi.org/10.24963/ijcai.2017/273>
40. D. Jin, Z. Yu, P. Jiao et al., A survey of community detection approaches: From statistical modeling to deep learning. *IEEE Trans. Knowl. Data Eng.* **35**(2), 1149–1170 (2021)
41. Y. Kim, J.W. Ha, Contrastive fine-grained class clustering via generative adversarial networks. In *the Tenth International Conference on Learning Representations, ICLR 2022* (Virtual Event, 25 April 2022), <https://openreview.net/forum?id=XW0De7ZLn8f>
42. D.P. Kingma, M. Welling, Auto-encoding variational bayes. In *International Conference on Learning Representations, ICLR 2014*. (Banff, AB, Canada, 14–16 April 2014), <https://openreview.net/forum?id=33X9fd2-9FyZd>
43. A. Krause, P. Perona, R. Gomes, Discriminative clustering by regularized information maximization. *Adv. Neural Inf. Process. Syst.* **23**, 775–783 (2010)
44. A. Krizhevsky, G. Hinton et al., Learning multiple layers of features from tiny images. Master's thesis, Department of Computer Science, University of Toronto, 2009
45. H.W. Kuhn, The hungarian method for the assignment problem. *Nav. Res. Logist. Q.* **2**(1–2), 83–97 (1955)
46. A. Kumar, H. Kim, G.P. Hancke, Environmental monitoring systems: A review. *IEEE Sensors J.* **13**(4), 1329–1339 (2012)
47. S. Laine, T. Aila, Temporal ensembling for semi-supervised learning. In *5th International Conference on Learning Representations, ICLR 2017*, (Toulon, France, 24–26 April 2017), <https://openreview.net/forum?id=BJ6oOfqge>
48. Y. Le, X. Yang, Tiny ImageNet visual recognition challenge. CS231n: Convolutional Neural Networks for Visual Recognition, Course Project Report, Stanford University (2015) http://vision.stanford.edu/teaching/cs231n/reports/2015/pdfs/yle_project.pdf
49. H. Li, Y. Li, M. Yang et al., Incomplete multi-view clustering via prototype-based imputation. In *IJCAI'23: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence* (Macao, China, 19–25 August 2023), pp. 3911–3919. <https://doi.org/10.24963/ijcai.2023/435>
50. P. Li, H. Zhao, H. Liu, Deep fair clustering for visual learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA, USA, 13–19 June 2020), pp. 9067–9076
51. Y. Li, P. Hu, Z. Liu et al., Contrastive clustering. In *Proceedings of the Thirty-Fifth AAAI conference on artificial intelligence* (Virtual Event, 2–9 February 2021), pp. 8547–8555
52. Y. Li, M. Yang, D. Peng et al., Twin contrastive learning for online clustering. *Int. J. Comput. Vis.* **130**(9), 2205–2221 (2022)
53. Y. Li, P. Hu, D. Peng et al., Image clustering with external guidance. *arXiv preprint* (2023) arXiv:231011989. <https://doi.org/10.48550/arXiv.2310.11989>
54. Y. Li, Y. Lin, P. Hu et al., Single-cell RNA-seq debiased clustering via batch effect disentanglement. *IEEE Trans. Neural Netw. Learn. Syst.* (2023). <https://doi.org/10.1109/TNNLS.2023.3260003>
55. Y. Li, D. Zhang, M. Yang et al., scBridge embraces cell heterogeneity in single-cell RNA-seq and ATAC-seq data integration. *Nat. Commun.* **14**, 6045 (2023)
56. Y. Lin, Y. Gou, Z. Liu et al., COMPLETER: Incomplete multi-view clustering via contrastive prediction. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Nashville, TN, USA, 20–25 June 2021), pp. 11169–11178
57. Y. Lin, Y. Gou, X. Liu et al., Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(4), 4447–4461 (2022). <https://doi.org/10.1109/TPAMI.2022.3197238>
58. Y. Lin, M. Yang, J. Yu et al., Graph matching with Bi-level noisy correspondence. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (Paris, France, 2–3 October 2023), pp. 23305–23314
59. J. Liu, Y. Lin, L. Jiang et al., Improve interpretability of neural networks via sparse contrastive coding. In *Findings of the Association for*

- Computational Linguistics: EMNLP 2022* (Abu Dhabi, United Arab Emirates, 7–11 December 2022), pp. 460–470
60. X. Liu, X. Zhu, M. Li et al., Multiple kernel k -means with incomplete kernels. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(5), 1191–1204 (2019)
 61. Y. Lu, Y. Lin, M. Yang et al., in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38 (AAAI Press, Washington, DC, 2024) pp. 14193–14201. <https://doi.org/10.1609/aaai.v38i13.29330>
 62. J. MacQueen et al., Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA* (University of California, Los Angeles, CA, USA, 1967), pp. 281–297
 63. A.F. McDaid, D. Greene, N. Hurley, Normalized mutual information to evaluate overlapping community finding algorithms. *arXiv preprint* (2011) arXiv:11102515. <http://arxiv.org/abs/1110.2515>
 64. E. Min, X. Guo, Q. Liu et al., A survey of clustering with deep learning: From the perspective of network architecture. *IEEE Access* **6**, 39501–39514 (2018)
 65. J.M. Montoya, S.L. Pimm, R.V. Solé, Ecological networks and their fragility. *Nature* **442**, 259–264 (2006)
 66. A. Moskalev, I. Sosnovik, V. Fischer et al., Contrasting quadratic assignments for set-based representation learning. In *European Conference on Computer Vision*, ed. by A. Moskalev, I. Sosnovik, V. Fischer, et al. Lecture Notes in Computer Science. vol. 13687 (Springer, Heidelberg, 2022), pp. 88–104
 67. S. Mukherjee, H. Asnani, E. Lin et al., ClusterGAN: Latent space clustering in generative adversarial networks. In *AAAI'19: AAAI Conference on Artificial Intelligence* (AAAI Press, Honolulu, HI, USA, 27 January–1 February 2019), pp. 4610–4617
 68. J. Murdock, L.S. Yaeger, Identifying species by genetic clustering. In *ECAL 2011: The 11th European Conference on Artificial Life* (Paris, France, 8–12 August 2011), <https://doi.org/10.7551/978-0-262-29714-1-ch087>
 69. F. Murtagh, P. Contreras, Algorithms for hierarchical clustering: an overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2**(1), 86–97 (2012)
 70. S.A. Nene, S.K. Nayar, H. Murase et al., Columbia object image library (COIL-20) (Department of Computer Science, Columbia University, 1996), <https://www.bibsonomy.org/bibtex/2e21afb22e024792723fc3b9f659c522e/jabrefest>
 71. M.E. Newman, M. Girvan, Finding and evaluating community structure in networks. *Phys. Rev. E* **69**(2), 026113 (2004)
 72. X.B. Nguyen, D.T. Bui, C.N. Duong et al., Clusformer: A transformer based clustering approach to unsupervised large-scale face and visual landmark recognition. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Nashville, TN, USA, 20–25 June 2021), pp. 10842–10851. <https://doi.org/10.1109/CVPR46437.2021.01070>
 73. F. Nie, J. Li, X. Li et al., Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification. In *IJCAI'16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, NY, USA, 9–15 July 2016), pp. 1881–1887. <https://dblp.org/rec/conf/ijcai/NieLL16.bib>
 74. F. Nie, J. Li, X. Li et al., Self-weighted multiview clustering with multiple graphs. In *IJCAI'17: Proceedings of the 26th International Joint Conference on Artificial Intelligence* (Melbourne, Australia, 19–25 August 2017), pp. 2564–2570. <https://doi.org/10.24963/ijcai.2017/357>
 75. C. Niu, H. Shan, G. Wang, SPICE: Semantic Pseudo-labeling for image clustering. *IEEE Trans. Image Process.* **31**, 7264–7278 (2022)
 76. A.V.D. Oord, Y. Li, O. Vinyals, Representation learning with contrastive predictive coding. *arXiv preprint* (2018) arXiv:180703748. <http://arxiv.org/abs/1807.03748>
 77. X. Peng, S. Xiao, J. Feng et al., Deep subspace clustering with sparsity prior. In *IJCAI'16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, NY, USA, 9–15 July 2016), pp. 1925–1931
 78. X. Peng, Z. Huang, J. Lv et al., COMIC: Multi-view clustering without parameter selection. In *Proceedings of the 36th International Conference on Machine Learning* PMLR, vol. 97 (Long Beach, California, USA, 9–15 June 2019), pp. 5092–5101
 79. Q. Qian, Stable cluster discrimination for deep clustering. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, Paris, France, 1–6 October 2023), pp. 16599–16608
 80. A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks. In *4th International Conference on Learning Representations, ICLR 2016, (San Juan, Puerto Rico, 2–4 May 2016)*. <http://arxiv.org/abs/1511.06434>
 81. A. Radford, J.W. Kim, C. Hallacy et al., Learning transferable visual models from natural language supervision. in *Proceedings of the 38th International Conference on Machine Learning*. PMLR, vol. 139 (Virtual, 18–24 July 2021), pp. 8748–8763
 82. Y. Ren, J. Pu, Z. Yang et al., Deep clustering: A comprehensive survey. *arXiv preprint* (2022) arXiv:221004142. <https://doi.org/10.48550/arXiv.2210.04142>
 83. S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**(5500), 2323–2326 (2000)
 84. H. Saeedi Emadi, S.M. Mazinani, A novel anomaly detection algorithm using DBSCAN and SVM in wireless sensor networks. *Wirel. Pers. Commun.* **98**, 2025–2035 (2018)
 85. S.E. Schaeffer, Graph clustering. *Comput. Sci. Rev.* **1**(1), 27–64 (2007)
 86. S.A. Shah, V. Koltun, Robust continuous clustering. *Proc. Natl. Acad. Sci.* **114**(37), 9814–9819 (2017)
 87. S.A. Shah, V. Koltun, Deep continuous clustering. *arXiv preprint* (2018) arXiv:180301449. <http://arxiv.org/abs/1803.01449>
 88. U. Shaham, R.R. Lederman, Learning by coincidence: Siamese networks and common variable learning. *Pattern Recogn.* **74**, 52–63 (2018)
 89. U. Shaham, K. Stanton, H. Li et al., Spectralnet: Spectral clustering using deep neural networks. In *6th International Conference on Learning Representations, ICLR 2018, (Vancouver, BC, Canada, 30 April–3 May 2018)*. https://openreview.net/forum?id=HJ_a0CyRZ
 90. Y. Shen, Z. Shen, M. Wang et al., You never cluster alone. *Adv. Neural Inf. Process. Syst.* **34**, 27734–27746 (2021)
 91. C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019)
 92. K. Sohn, D. Berthelot, C.L. Li et al., Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, ed. by H. Larochelle, M. Ranzato, R. Hadsell, et al. Neural Information Processing Systems Foundation, San Diego, CA, USA, 2020). <https://proceedings.neurips.cc/paper/2020/hash/06964dce9addd1c5cbs5d6e3d9838f733-Abstract.html>
 93. C. Song, F. Liu, Y. Huang et al., Auto-encoder based data clustering. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 18th Iberoamerican Congress, CIARP 2013, Havana, Cuba, November 20–23, 2013, Proceedings, Part I 18* (Springer, Berlin, Heidelberg, 2013), pp. 117–124
 94. X. Su, S. Xue, F. Liu et al., A comprehensive survey on community detection with deep learning. *IEEE Trans. Neural Netw. Learn. Syst.* **35**, 4682–4702 (2024) <https://doi.org/10.1109/TNNLS.2021.3137396>
 95. W. Van Gansbeke, S. Vandenhende, S. Georgoulis et al., SCAN: Learning to classify images without labels. In *Proceedings of 16th European conference on computer vision* (Virtual, 23–28 August 2020), pp. 268–285
 96. Q. Wang, M. Chen, F. Nie et al., Detecting coherent groups in crowd scenes by multiview clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(1), 46–58 (2018)
 97. W. Wang, X. Yan, H. Lee et al., Deep variational canonical correlation analysis. *arXiv preprint* (2016) arXiv:161003454. <http://arxiv.org/abs/1610.03454>
 98. Z. Wang, Y. Ni, B. Jing et al., DNB: A joint learning framework for deep bayesian nonparametric clustering. *IEEE Trans. Neural Netw. Learn. Syst.* **33**(12), 7610–7620 (2022)
 99. J. Wright, Y. Ma, J. Mairal et al., Sparse representation for computer vision and pattern recognition. *Proc. IEEE* **98**(6), 1031–1044 (2010)
 100. D. Wu, S.J. Zheng, X.P. Zhang et al., Deep learning-based methods for person re-identification: A comprehensive review. *Neurocomputing* **337**, 354–371 (2019)
 101. M. Wu, L. Tan, N. Xiong, Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications. *Inf. Sci.* **329**, 800–818 (2016)
 102. Z. Wu, Y. Xiong, S.X. Yu et al., Unsupervised feature learning via non-parametric instance discrimination. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Salt Lake City, UT, USA, 18–23 June 2018), pp. 3733–3742
 103. D. Xia, N. Vljajic, Near-optimal node clustering in wireless sensor networks for environment monitoring. In *21st International conference on advanced information networking and applications (AINA'07)* (Niagara Falls, ON, Canada, 21–23 May 2007), pp. 632–641

104. J. Xie, R. Girshick, A. Farhadi, Unsupervised deep embedding for clustering analysis. In *Proceedings of The 33rd International Conference on Machine Learning* PMLR, vol 48 (New York City, NY, USA, 20-22 June 2016) pp. 478–487
105. C. Xu, D. Tao, C. Xu, A survey on multi-view learning. arXiv preprint (2013) arXiv:13045634. <http://arxiv.org/abs/1304.5634>
106. J. Xu, S. De Mello, S. Liu et al., GroupViT: Semantic segmentation emerges from text supervision. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (New Orleans, LA, USA, 18-24 June), pp. 18113–18123. <https://doi.org/10.1109/CVPR52688.2022.01760>
107. Y. Yan, J. Li, J. Qin et al., Efficient person search: An anchor-free approach. *Int. J. Comput. Vis.* **131**(7), 1642–1661 (2023)
108. J. Yang, D. Parikh, D. Batra, Joint unsupervised learning of deep representations and image clusters. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV, USA, 27-30 June 2016), pp. 5147–5156
109. J. Yang, J. Liu, N. Xu et al., TVT: Transferable vision transformer for unsupervised domain adaptation. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (Waikoloa, HI, USA, 2-7 January 2023), pp. 520–530
110. M. Yang, Y. Li, Z. Huang et al., Partially view-aligned representation learning with noise-robust contrastive loss. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Nashville, TN, USA, 20-25 June 2021), pp. 1134–1143
111. M. Yang, Z. Huang, P. Hu et al., Learning with twin noisy labels for visible-infrared person re-identification. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (New Orleans, LA, USA, 18-24 June 2022), pp. 14288–14297
112. M. Yang, Y. Li, P. Hu et al., Robust multi-view clustering with incomplete information. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(1), 1055–1069 (2023). <https://doi.org/10.1109/TPAMI.2022.3155499>
113. M. Ye, J. Shen, G. Lin et al., Deep learning for person re-identification: A survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(6), 2872–2893 (2022)
114. P. Zeng, Y. Li, P. Hu et al., Deep fair clustering via maximizing and minimizing mutual information: Theory, algorithm and metric. in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Vancouver, BC, Canada, 17-24 June 2023), pp. 23986–23995
115. C. Zhang, H. Fu, S. Liu et al., Low-rank tensor constrained multiview subspace clustering. In *2015 IEEE International Conference on Computer Vision (ICCV)* (Santiago, Chile, 7-13 December 2015), pp. 1582–1590
116. H. Zhang, F. Nie, X. Li, Large-scale clustering with structured optimal bipartite graph. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(8), 9950–9963 (2023). <https://doi.org/10.1109/TPAMI.2023.3277532>
117. L. Zhang, G.J. Qi, L. Wang et al., AET vs. AED: Unsupervised representation learning by auto-encoding transformations rather than data. in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA, USA, 15-20 June 2019), pp. 2542–2550
118. Z. Zhang, L. Liu, F. Shen et al., Binary multi-view clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(7), 1774–1782 (2019)
119. H. Zhao, H. Liu, Y. Fu, Incomplete multi-modal visual data grouping. In *IJCAI'16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, NY, USA, 9-15 July 2016), pp. 2392–2398
120. T. Zhao, Z. Wang, A. Masoomi et al., Streaming adaptive nonparametric variational autoencoder. arXiv preprint (2019) arXiv:190603288. <http://arxiv.org/abs/1906.03288>
121. H. Zhong, C. Chen, Z. Jin et al., Deep robust clustering by contrastive learning. arXiv preprint (2020) arXiv:200803030. <https://arxiv.org/abs/2008.03030>
122. H. Zhong, J. Wu, C. Chen et al., Graph contrastive clustering. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. (Montreal, QC, Canada, 10-17 October 2021), pp. 9204–9213
123. S. Zhou, H. Xu, Z. Zheng et al., A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions. arXiv preprint (2022) arXiv:220607579. <https://doi.org/10.48550/arXiv.2206.07579>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.