

# Unsupervised Deep Learning: Taxonomy and Algorithms

Aida Chefrour<sup>1,2,\*</sup> and Labiba Souici-Meslati<sup>2</sup>

E-mail: aida.chefrour@univ-soukahras.dz, souici\_labiba@yahoo.fr

<sup>1</sup> Computer Science Department, Mohamed Cherif Messaadia University, Souk Ahras, Algeria

<sup>2</sup> LISCO Laboratory, Computer Science Department, Badji Mokhtar University, B.P-12 Annaba, 23000, Algeria

## Overview paper

**Keywords:** clustering, deep learning, autoencoder, taxonomy

**Received:** November 10, 2021

*Clustering is a fundamental challenge in many data-driven application fields and machine learning techniques. The data distribution determines the quality of the outcomes, which has a significant impact on clustering performance. As a result, deep neural networks can be used to learn more accurate data representations for clustering. Many recent studies have focused on employing deep neural networks to develop a clustering-friendly representation, which has resulted in a significant improvement in clustering performance. We present a systematic survey of clustering with deep learning in this study. Then, a taxonomy of deep clustering is proposed, as well as some sample algorithms for our overview. Finally, we discuss some exciting future possibilities for clustering using deep learning and offer some remarks.*

*Povzetek: Ta članek opisuje metode globokega združevanja v skupine in predlaga taksonomijo globokega združevanja v skupine.*

## 1 Introduction

Clustering is one of the most important aspects of unsupervised machine learning. Its main goal is to separate a data set into subsets or clusters so that data values in the same cluster have some common characteristics or attributes. It aims to divide the data into groups (clusters) of similar objects. The objects in the same cluster are more identical to each other than to those in other clusters. Clustering is widely used in Artificial Intelligence, pattern recognition, statistics, and other information processing fields. The input of a cluster analysis system is a set of samples and a measure of similarity (or dissimilarity) between two samples. The output is a set of clusters that form a partition, or a structure of partitions of the data set. Generally, finding clusters is not a simple task and the current clustering algorithms take a long time when they are applied to large databases [1].

In addition, the transformation of input data into a feature space where separation is easier concerning the problem's context, dimensionality reduction, and representation learning has been widely applied to clustering, because the similarity measurements utilized in these procedures are ineffective.

Existing data transformation methods generally include linear transformations such as Principal Component Analysis (PCA) and non-linear transformations such as kernel approaches and spectral methods [2].

To solve this problem, Deep Neural Networks (DNNs) are used to train non-linear mappings that allow the data to be transformed into clustering-friendly representations because they have a significant non-linear transformation feature. In this paper, we refer to clustering approaches involving deep learning as deep clustering for simplicity.

In our research, we focus on Deep Clustering, which represents a family of clustering algorithms that adopt deep neural networks to learn cluster-oriented features [3]. Deep clustering has recently become popular as a method for data classification and feature representation discovery, a solution for large-scale and high-dimensional learning problems [4,5]

We were particularly interested in the studies conducted in deep clustering for image recognition. We give an overview of deep clustering to review most methods and implementations in this field.

The main contributions treated in this paper are:

- use of a deep autoencoder for embedding the data into a lower-dimensional space;
- integrate the extracting intermediate features phase and the performing phase of the traditional clustering algorithm;
- employ the similarity of the representation features if they are assigned to the same cluster;
- add dimensionality reduction and temporal clustering into a single unsupervised learning framework;

---

\* Corresponding author

- apply the impressive ability to deal with unsupervised learning for structure analysis of high-dimensional visual data;
- find a solution to the problem of subspace clustering by partitioning data drawn from a union of multiple subspaces.

The contribution of this study is (1) to provide an overview of various deep learning-based clustering algorithms. It includes an explanation of the most recent improvements in unsupervised clustering; (2) propose a taxonomy of methods that use deep learning for clustering.

We chose to synthesize studies published in the previous 3-4 years since they used deep learning to increase unsupervised clustering performance. On the MNIST dataset, several algorithms achieve more than 96% accuracy without using a single labeled datapoint. However, for more difficult datasets like CIFAR-10 and ImageNet, they are still a long way from achieving good accuracy.

We'll go over all of the most recent deep learning-based clustering approaches in this article. The aim of most of these strategies is to discover feature representation and solve the problem of large-scale, high-dimensional learning, as well as to respond to the contributions mentioned above.

The rest of the paper is organized as follows. In the next section, we survey in brief the literature on deep clustering overviews. We present the most recent works using unsupervised deep learning in section 3, with a synthesis of all of this work in section 4. In section 5, we describe the proposed taxonomy of clustering with deep learning algorithms and we introduce some representative methods. Section 6 includes a conclusion and proposals for further research.

## 2 Related work

Several custom taxonomies for clustering with deep learning have been proposed in the literature. In this section, we outline the best known and most recent ones:

[6] focus on a review of deep learning for multimodal data fusion, which provides readers with the fundamentals of the multimodal deep learning fusion methods. This study summarizes the representative architectures—DBN, SAE, CNN, and RNN—which are fundamental to understanding multimodal deep learning fusion models. This work summarizes the pioneering multimodal deep learning fusion models from the task, model framework, and data set perspectives, and groups them by the deep learning architecture used.

[2] divide deep clustering algorithms into four categories: AE-based (Autoencoder), CDNN-based (Clustering DNN), VAE-based (Variational encoder), and GAN-based deep clustering (Generative Adversarial Network). Each category has some representative methods as well.

- (a) AE-based has a (1) Deep Clustering Network (DCN), which combines an autoencoder with the k-means algorithm; (2) Deep Embedding Network (DEN) to extract effective representations for clustering, which utilizes a

deep autoencoder; (3) Deep Subspace Clustering Networks (DSC-Nets) which introduces a novel autoencoder architecture; (4) Deep Multi-Manifold Clustering (DMC); (5) Deep Embedded Regularized Clustering (DEPICT); and (6) Deep Continuous Clustering (DCC);

- (b) CDNN-based deep clustering algorithms can be divided into three categories according to the way of network initialization, i.e., unsupervised pre-trained (Deep Nonparametric Clustering (DNC), Deep Embedded Clustering (DEC), Discriminatively Boosted Clustering (DBC)), supervised pre-trained (Clustering Convolutional Neural Network (CCNN)), randomly initialized (non-pre-trained) (Information Maximizing Self-Augmented Training (IMSAT), Joint Unsupervised Learning (JULE) and Deep Adaptive Image Clustering (DAC));
- (c) VAE-based deep clustering, which can be considered as a generative variant of AE. It presents two algorithms: (1) Variational Deep Embedding (VaDE) and (2) Gaussian Mixture VAE (GMVAE);
- (d) GAN-based deep clustering contains a (1) Deep Adversarial Clustering (DAC), (2) Categorical Generative Adversarial Network (CatGAN), and (3) Information Maximizing Generative Adversarial Network (InfoGAN).

[7] propose a taxonomy of clustering algorithms that employ deep learning. Their taxonomy helps the user to see what methods are available and to create new ones by combining the best characteristics of existing methods in a simple context. This taxonomy's main principle is representation learning with DNNs and using these representations as input to a specific clustering approach. Every method is divided into the following parts, each of which has a variety of options: (1) Architecture of the main neural network branch (Multilayer perceptron (MLP), Convolutional neural network (CNN) and Deep Belief Network (DBN)); (2) Set of deep features used for clustering (one layer, several layers); (3) Non-clustering loss (No non-clustering loss, Autoencoder reconstruction loss); (4) Clustering loss (No clustering loss, k-Means loss, Cluster assignment hardening, Balanced assignments loss, Locality-preserving loss, Group sparsity loss, Cluster classification loss, and Agglomerative clustering loss); (5) Method to combine the losses (Pre-training, fine-tuning, Joint training and Variable schedule); (6) Cluster updates (Jointly updated with the network model, and Alternatingly updated with the network model); (7) After network training (Clustering a similar dataset and Obtaining better results). The methods which use this taxonomy are Deep Embedded Clustering (DEC), Deep Clustering Network (DCN), Discriminatively Boosted Clustering (DBC), Joint Unsupervised Learning of Deep Representations and Image Clusters (JULE), and Clustering CNN (CCNN).

[8] propose a simplified taxonomy based on deep clustering algorithms' overall procedural structure or design. Deep Clustering techniques can be classified into three broad families according to this taxonomy: (a) Sequential multistep Deep Clustering approaches: these approaches have two basic steps. The first stage involves learning richer deep (also known as latent) representation of the input data, followed by clustering on this deep or latent representation in the second step; (b) Joint Deep Clustering approaches: Instead of two independent processes for representation learning and clustering, this family of approaches includes a step where the representation learning is intimately associated with the clustering. Tight coupling is usually achieved by optimizing a combined or joint loss function that promotes good reconstruction while accounting for some sort of data grouping, clustering, or codebook representation; (c) Closed-loop multistep Deep Clustering approaches: Similar to the first family (sequential multistep Deep Clustering), this family of algorithms has two key phases that alternate in an iterative loop rather than being conducted in a single feedforward linear approach.

### 3 Contributions of deep clustering

In recent years, we have noticed that there are many applications in the field of deep learning using unsupervised learning algorithms for image recognition. We now discuss some of the most common deep clustering approaches.

[9] find that existing deep clustering algorithms either do not take advantage of convolutional neural networks well enough or do not preserve the local structure of data-generating distribution in the learned feature space sufficiently. In this research, they suggest a deep convolutional embedded clustering method as a solution to this problem. They create a convolutional autoencoder structure to learn embedded features from start to finish. Then, using embedded features, a clustering-oriented loss is created to accomplish feature refinement and cluster assignment simultaneously. They keep the decoder, which can preserve the local structure of data in feature space, to avoid feature space being affected by clustering loss. In summary, they minimize both the reconstruction and clustering losses of convolutional autoencoders. Mini-batch stochastic gradient descent with back-propagation can effectively solve the resulting optimization issue. Experiments on benchmark datasets (MNIST-full, MNIST-test, and USPS) empirically verify the usefulness of local structure preservation and the power of

convolutional autoencoders for feature learning in terms of accuracy (acc) and the normalized mutual information (NMI).

DeepCluster [10] is a clustering algorithm developed by the authors that learn both the parameters of a neural network and the cluster assignments of the generated features. DeepCluster uses a typical clustering technique, k-means, to iteratively group the features and uses the following assignments as supervision to update the network's weights. They use DeepCluster to train convolutional neural networks unsupervised on big datasets like ImageNet and YFCC100M, using accuracy criteria evaluation. On all typical benchmarks, the generated model exceeds the present state of the art by a significant margin.

This study's [11] concern is that data representation affects the performance of subspace clustering. Subspace clustering data representation translates data from one space to another with higher separability. In recent years, a slew of new data visualization techniques has emerged. Low-rank representation (LRR) and an autoencoder are two examples. LRR is a low-rank constraint linear representation method that captures the global structure of data. An autoencoder, on the other hand, uses a neural network to nonlinearly map data into a latent space by minimizing the difference between the reconstruction and the output. The authors of this work suggest a unique data representation approach for subspace clustering that combines the benefits of an LRR (globality) and an autoencoder (self-supervision-based locality). The low-rank constrained autoencoder (LRAE) method introduced in this research causes the neural network's latent representation to be of low rank, and the low-rank constraint is derived as a prior from the input space. One of the most significant advantages of the LRAE is that the learned data representation not only preserves the data's local properties but also serves as a precursor to the underlying low-rank global structure. Extensive subspace clustering tests were carried out on a variety of datasets (MNIST, COIL-100, and ORL), using ACC, NMI, and the adjusted rand index (ARI). They showed that the suggested LRAE outperformed state-of-the-art subspace clustering approaches significantly.

The researchers in this paper [12] created a hybrid autoencoder (BAE) model for image clustering by combining three AE-based models: the convolutional autoencoder (CAE), adversarial autoencoder (AAE), and stacking autoencoder (SAE). The MNIST and CIFAR-10 datasets are used to test the suggested models' results and compare them to those of other researchers. The proposed models outperform others in the numerical experiment, according to the clustering criteria: ACC, NMI, and ARI.

GANs have demonstrated great performance in a variety of unsupervised learning problems, and clustering is unquestionably an important unsupervised learning challenge. While the latent-space back-projection in GANs could be used to cluster, they show that the cluster structure is not preserved in the GAN latent space. ClusterGAN is a new mechanism for clustering using GANs proposed by the authors in this study [13]. They achieve clustering in the latent space by sampling latent

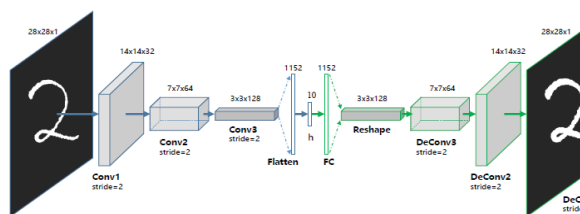


Figure 1: The structure of proposed Convolutional AutoEncoders (CAE) for MNIST [9].

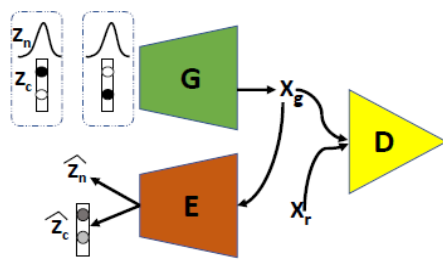


Figure 2: ClusterGAN Architecture [13].

variables from a mixture of one-hot encoded variables and continuous latent variables, together with an inverse network (which projects the data to the latent space) trained jointly with a clustering specific loss. GANs can maintain latent space interpolation across categories, even though the discriminator is never exposed to such vectors, according to their findings. They compared their results to a variety of clustering benchmarks (MNIST, Synthetic, Fashion-10,6 Fashion-5, 10x\_73k, and Pendigits) and showed that they outperformed them on both synthetic and real-world datasets according to the following evaluation criteria: ACC, NMI, and ARI.

This work [14] proposes a new approach to this study, in which the embedding is performed using a differentiable model such as a deep neural network. They create a fully differentiable loss function that can be minimized concerning both the embedding parameters and the cluster parameters via stochastic gradient descent by rewriting the k-means clustering method as an optimal transport problem and adding an entropic regularization. They show that by including limits on cluster sizes, this new formulation generalizes a previously suggested state-of-the-art soft-k-means technique. According to empirical evaluations of image classification benchmarks (MNIST, CIFAR-10), their optimum transport-based technique provides greater unsupervised accuracy and does not require a pre-training step when compared to state-of-the-art methods.

The researchers of this work [15] present a deep Generative Adversarial Clustering Network (ClusterGAN) in this publication, which addresses the challenges of unsupervised deep clustering model training. ClusterGAN is made up of three networks that include a discriminator, a generator, and a clustered (i.e. a clustering network). They use an adversarial game between these three players to use the generator to synthesize actual samples given discriminative latent variables, and the clustered to learn the inverse mapping of the real samples to the discriminative embedding space. Furthermore, they use a conditional entropy minimization loss to increase/decrease Intra/inter-cluster sample similarity. Because the ground-truth similarities in the clustering task are unknown, they offer a new balanced self-paced learning algorithm for gradually incorporating data into training from simple to tough while taking into account the diversity of selected samples from all clusters. Their unsupervised learning approach allows them to train clusters with a lot of depth quickly. On numerous datasets (MNIST, USPS, FRGC, CIFAR-10, and STL-10),

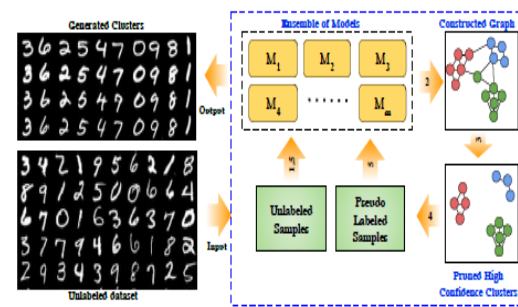


Figure 3: Kingdra overview . They train all the models using the unlabeled samples, in step 1. In step 2, they construct a graph modeling the pairwise agreement of the models. In step 3, they get k high confidence clusters by pruning out data points for which the models do not agree. In step 4 they take the high confidence clusters and generate pseudo labels. In step 5 they train the models using both unlabeled samples and pseudo labeled samples. They iterate from step 2 to step 5 and final clusters are generated [16].

ClusterGAN produces competitive outcomes when compared to state-of-the-art models, according to experimental results, using accuracy criteria evaluation Acc and NMI.

The main topic of this work [3] is that deep clustering outperforms conventional clustering by combining feature learning and cluster assignment. Although several deep clustering algorithms have been developed for various purposes, the majority of them fail to learn robust cluster-oriented features, resulting in poor final clustering performance. The authors suggest a two-stage deep clustering technique (ASPC-DA) that incorporates data augmentation and self-paced learning to overcome this challenge. They discover robust features in the first stage by training an autoencoder with examples that have been enhanced by random shifting and rotating the clean instances. Then, in the second stage, they alternate between finetuning the encoder with augmented examples and modifying the cluster assignments of the clean examples to encourage the learned features to be cluster-oriented. The center of the cluster to which the clean example is assigned is the target of each augmented example in the loss function during finetuning of the encoder. The targets could be computed improperly, and the encoder network could be misled by instances of inaccurate targets. They use adaptive self-paced learning to select the most confident instances in each iteration to stabilize the network training. Extensive testing shows that their algorithm outperforms the competition on four image datasets (MNIST-full, MNIST-test, USPS, and Fashion) in terms of ACC and NMI.

The authors of this study [16] present a system for improving unsupervised clustering performance using semi-supervised models called Kingdra. To use semi-supervised models, they must first create pseudo-labels, which are automatically generated labels. Prior approaches to creating pseudo-labels have been found to degrade clustering performance due to their low precision.

Instead, they generate a similarity graph using an ensemble of deep networks, from which they extract high-accuracy pseudo labels. The method of employing ensembles to find high-quality pseudo-labels and training the semi-supervised model is iterated, resulting in continual improvement. For numerous image and text datasets, they show that their approach beats state-of-the-art clustering findings. To evaluate their method, they used the accuracy evaluation criteria and five datasets (MNIST, STL, CIFAR10, Reuters, and 20news). They reached 54.6 % accuracy for CIFAR-10 and 43.9 % for 20 news.

In [17], discriminative models are the most common in the literature, and they produce the best results. These algorithms learn a deep discriminative neural network classifier with latent labels. As is common in supervised learning, they typically use multinomial logistic regression posteriors and parameter regularization. Discriminative objective functions (e.g., those based on mutual information or KL divergence) are generally thought to be more flexible than generative approaches (e.g., K-means) in that they make fewer assumptions about data distributions and, as a result, produce much better unsupervised deep learning results. Several contemporary discriminative models may appear to be unrelated to K-means at first glance. Under mild conditions, these models are similar to K-means, common posterior models, and parameter regularization, as demonstrated in this paper. The authors show that maximizing the L2 regularized mutual information via an approximate alternating direction method (MI-ADM) for commonly used logistic regression posteriors is comparable to minimizing a soft and regularized K-means loss. Their theoretical study not only ties numerous recent state-of-the-art discriminative models directly to K-means but also leads to a novel soft and regularized deep K-means algorithm that performs well on a variety of image clustering benchmarks. They used the accuracy and normalized mutual information criteria for the evaluation of five datasets: USPS, MNIST, YTF, CMU-PIE, and FRGC.

The researchers [18] introduced a new clustering objective that develops a neural network classifier from the start using only unlabeled input samples. In eight unsupervised clustering benchmarks spanning image classification and segmentation, the model discovers clusters that accurately match semantic classes, delivering state-of-the-art performance. These include STL10, an unsupervised ImageNet variation, and CIFAR10, which outperformed their closest competitors by 6.6 and 9.5 absolute percentage points, respectively. The strategy isn't limited to computer vision and can be applied to any paired dataset sample; in their studies, they used random transforms to generate a pair from each image. Instead of high-dimensional representations that require further processing to be useable for semantic clustering, the trained network outputs semantic labels directly. The goal is simple: to maximize the mutual information between each pair's class assignments. It's simple to use and is firmly rooted in information theory, so it easily avoids the degenerate solutions that other clustering algorithms are prone to. The experiments used four datasets: STL10,

CIFAR10, CIFAR 100-20, and MNIST. They examine two semi-supervised settings in addition to the unsupervised mode. The first achieves a global state-of-the-art of 88.8% accuracy in STL10 classification, surpassing all current approaches (whether supervised, semi-supervised or unsupervised). The second reveals that it can withstand 90 percent reductions in label coverage, which is useful for applications that just need a few labels.

In [19], the authors of this paper discuss a variant of variationally-oriented autoencoders where the superstructure of latent variables is on top of the features of the autoencoders. Their model is based on a tree structure that consists of multiple super latent variables. When there is only one active variable in the superstructure, it generates a model that assumes the latent features of that variable are generated by the Gaussian mixture model. The model, known as the Latent Tree Variational AutomaticEncoder (LTVAE) learns by creating multiple partitions of data, each containing a super latent variable. It is a type of deep learning method that produces multiple partitions of data. This method allows us to partition high-dimensional data into multiple ways. To evaluate this model, they used four datasets: MNIST, STL-10, Reuters, and HHAR, the criteria for clustering accuracy.

In [20], to resolve the problem of high-dimensional dataset clustering difficulties, the authors of this paper describe a clustering approach that simultaneously conducts nonlinear dimensionality reduction and clustering. A deep autoencoder embeds the data in a lower-dimensional space. As part of the clustering process, the autoencoder is optimized. The resulting network generates data that is clustered. The proposed method, Deep Continuous Clustering (DCC) does not rely on knowing the number of ground-truth clusters in advance. The optimization of a global continuous objective is used to combine nonlinear dimensionality reduction and clustering. As a result, they avoid the discrete reconfigurations of the objective that previous clustering algorithms are known for. Experiments on six datasets (MNIST, Coil100, YTF, YaleB, Reuters, and RCV1) using the accuracy evaluation criteria (AMI) show that the proposed approach outperforms current clustering approaches, including deep network-based approaches like k-means, DBSCAN, AC-W, SEC, LDMGI, GDL, and RCC.

Deep clustering through a Gaussian-mixture variational autoencoder (VAE) with Graph embedding is proposed by the authors in [21]. They use the Gaussian mixture model (GMM) as the prior in VAE to make clustering easier. They use graph embedding to handle data with a complicated spread. Their hypothesis is that graph data, which captures local data structures, is a great complement to deep GMM. When they're combined, the network can develop more powerful representations that adhere to global models and local structural restrictions. As a result, their method unites model-based and similarity-based clustering methodologies. They propose a novel stochastic extension of graph embedding to combine graph embedding with probabilistic deep GMM: they consider samples as nodes on a graph and minimize

the weighted distance between their posterior distributions. The distance is calculated using the Jensen-Shannon divergence. They integrate the deep GMM's divergence minimization and log-likelihood maximization. They came up with formulations to achieve a unified goal that allows deep representation learning and clustering to happen at the same time. Their findings on four datasets (MNIST, STL-10, Reuters, and HHAR) in terms of accuracy reveal that their suggested DGG outperforms recent deep Gaussian mixture approaches (model-based) and deep spectral clustering techniques (similarity-based). The benefits of integrating model-based and similarity-based clustering, as advocated in this paper, are highlighted by their findings.

The authors [22] present a shared learning paradigm for discriminative embedding and spectral clustering in this research. To embed the inputs into a latent space for clustering, they first build a dual autoencoder network that enforces the reconstruction requirement for the latent representations and their noisy variants. As a result, the learned latent representations may be more noise-resistant. Then, to give more discriminative information from the inputs, mutual information estimation is used. Furthermore, a deep spectral clustering method is used to embed the latent representations in the eigenspace and then cluster them, allowing for optimal clustering outcomes by fully exploiting the link between inputs. Experiments on benchmark datasets (MNIST-full, MNIST-test, USPS, Fashion-10, and YTF) reveal that their strategy outperforms state-of-the-art clustering algorithms significantly (k-means, NMF,...) in terms of ACC and NMI.

The researchers [23] offer a unique clustering framework called deep comprehensive correlation mining (DCCM) in this paper for analyzing and exploiting various types of correlations behind unlabeled data from three perspectives: 1) Pseudo-label supervision is presented as an alternative to employing only pair-wise information to examine category information and develop discriminative features. 2) The resilience of the features to picture alteration in the input space is completely studied, which aids network learning and boosts performance greatly. 3) For the clustering problem, triplet mutual information among features is introduced to lift the recently discovered instance-level deep mutual information to a triplet-level formation, which aids in the learning of more discriminative features. Extensive tests on a variety of tough datasets (CIFAR-10, CIFAR-100, STL-10, ImageNet-10, Imagenet-dog-15, and Tiny-ImageNet) in terms of ACC, NMI, and adjusted rand index (ARI) reveal that their method works well, with 62.3 % clustering accuracy on CIFAR-10, which is 10.1 % better than the state-of-the-art results (k-means, AE,...).

By jointly maximizing a clustering loss and a non-clustering loss, deep clustering algorithms combine representation learning with clustering. In such systems, a deep neural network is combined with a clustering network to learn representations. Rather than using this framework to increase clustering performance, the researchers [24] offer a simpler method of maximizing the entanglement of an autoencoder's learned latent code

representation. They define entanglement as the distance between pairs of points belonging to the same class or structure and pairs of points belonging to different classes or structures. They employ the soft closest neighbor loss and expand it by adding an annealing temperature factor to assess the entanglement of data points. The test clustering accuracy was 96.2% on the MNIST dataset, 85.6% on the Fashion-MNIST dataset, and 79.2% on the EMNIST Balanced dataset when they used their proposed approach, beating their baseline models.

The Matching Priors and Conditionals for Clustering (MPCC) is a GAN-based model featuring an encoder for inferring latent variables and cluster categories from data and a flexible decoder for generating samples from a conditional latent space, according to the researchers of [25]. They show via MPCC that a deep generative model may compete/outperform discriminative approaches in clustering tasks, outperforming the state of the art across a variety of benchmark datasets (MNIST, CIFAR10). In CIFAR10, their tests show that adding a learnable prior and increasing the number of encoder updates improves the quality of the generated samples, resulting in an inception score of  $9.49 \pm 0.15$  and a 46.9% improvement in the Fréchet inception distance above the state of the art.

The researchers of [26] show that greedy or local methods of maximizing mutual information (such as stochastic gradient optimization) identify local optimal for the mutual information criterion; as a result, the resulting representations are less-than-ideal for complex downstream tasks. This problem has not been identified or addressed in previous research. They introduced deep hierarchical object grouping (DHOG), which generates representations that better optimize the mutual information objective by computing many separate discrete representations of pictures in a hierarchical sequence. They also discovered that these representations are more suited to the task of grouping objects into underlying object classes. They put DHOG to the test on unsupervised clustering, which is a natural downstream test given that the target representation is discrete data labeling. They produced new state-of-the-art scores on the three key benchmarks (CIFAR-100-20, STL-10, and SVHN) without any of the pre-filtering or Sobel-edge detection that many earlier approaches needed to work. They obtained accuracy improvements of 4.3% on CIFAR-10, 1.5% on CIFAR-100-20, and 7.2% on SVHN.

The researchers in this work [27] tackle the problem of Federated Learning (FL), where users are spread and partitioned into clusters. This configuration represents scenarios in which separate groups of users have their own goals (learning tasks), but by aggregating their data with those of others in the same cluster (same learning task), they can take advantage of the power of numbers to execute more efficient Federated Learning. They present the Iterative Federated Clustering Algorithm (IFCA), a new framework that uses gradient descent to estimate user cluster identities and improve model parameters for user clusters. They investigated the algorithm's convergence rate in a linear model with squared loss, as well as for generic strongly convex and smooth loss functions. They demonstrate that IFCA converges at an exponential rate in



both scenarios with good initialization, and they explain the statistical error rate's optimality. They propose training the models by combining IFCA with the weight sharing strategy in multi-task learning when the clustering structure is uncertain. They show that our technique can succeed even if we reduce the initialization criteria by using random initialization and repeated restarts in the tests. They also offer practical data demonstrating the efficiency of our technique in non-convex problems like neural networks. On numerous clustered FL benchmarks (Rotated MNIST, Rotated CIFAR), they show how IFCA outperforms the baselines in terms of precision.

The problem with this work [28] is that unsupervised image classification is a difficult computer vision task. Deep learning-based algorithms have produced excellent results, with the most recent technique using uniform embedding and class assignment losses. Because these processes have distinct goals fundamentally, improving them together may result in a suboptimal solution. To overcome this problem, the researchers suggest the IIC model (Invariant Information Clustering), a novel two-stage approach in which a pretraining embedding module is followed by a refining module that does both embedding and class assignment simultaneously. When evaluated with different datasets (CIFAR-10, CIFAR-100-20, and STL-10), their model outperforms SOTA in unsupervised tasks, with an accuracy of 81.0% for the CIFAR-10 dataset (an increase of 19.3% points), 35.3 % for CIFAR-100-20 (9.6 pp), and 66.5 % for STL-10 (6.9 pp).

Deep clustering has demonstrated an excellent ability to deal with unsupervised learning for structure analysis of high-dimensional visual data by learning visual features and data grouping at the same time. Local learning constraints based on inter-sample relations and/or self-estimated pseudo labels are commonly used in existing deep clustering algorithms. This is vulnerable to unavoidable errors that spread throughout the neighborhood, as well as to error propagation during training. Based on the observation that assigning samples from the same semantic categories into different clusters

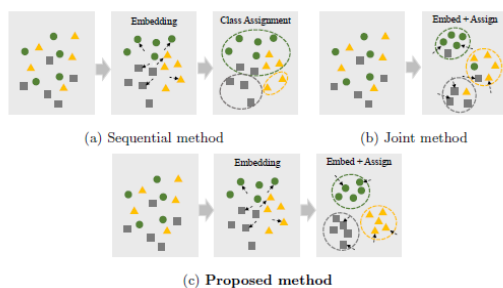


Figure 4: Methods for unsupervised image classification. (a) The sequential method embeds and assigns data points to classes one by one, whereas (b) the joint technique embeds and organizes data points into classes all at once. (c) The proposed technique performs embedding learning as a pretraining step to determine suitable initialization, then optimizes the embedding and class assignment processes simultaneously. During the pretraining stage of their two-stage design, they experience distinctive losses [28].

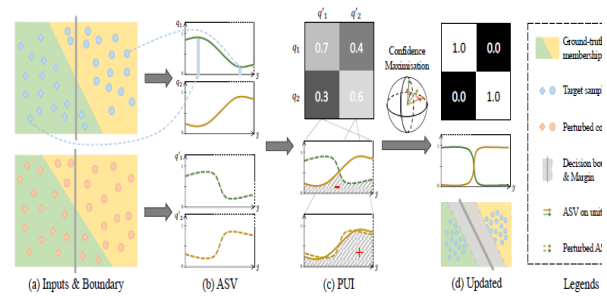


Figure 5: Unsupervised deep clustering using the proposed PartItion Confidence mAXimisation (PICA) approach. (a) Given the input data as well as the CNN model's decision bounds, (b) Using a mini-batch of data and its randomly perturbed copy, PICA computes the cluster-wise Assignment Statistics Vector (ASV) in the forward pass. (c) To reduce the partition uncertainty index as much as possible (PUI), (d) PICA is taught to use a specific objective loss function to distinguish the ASV of all clusters on the hypersphere to discover the most confident and potentially promising clustering solution [29].

reduces both intra-cluster compactness and inter-cluster diversity, i.e. lower partition confidence, the authors of [29] propose to solve this problem by learning the most confident clustering solution from all possible separations. In particular, they present PartItion Confidence MAXimisation, a unique deep clustering method (PICA). It is based on the principle of learning the most semantically plausible data separation, in which all clusters may be mapped one-to-one to the ground-truth classes, by increasing the "global" partition confidence of the clustering solution. This is accomplished by introducing a differentiable partition uncertainty index and its stochastic approximation, as well as a principled objective loss function that minimizes such an index, all of which, when combined, allow for direct application of traditional deep networks and mini-batch based model training. Extensive testing on six frequently used clustering benchmarks (CIFAR-10, CIFAR-100, STL-10, imageNet-10, ImageNet-dogs, and Tiny-ImageNet) demonstrates that their model outperforms a wide range of state-of-the-art techniques in terms of ACC, NMI, and ARI.

The challenge with this study [30] is that there is no obvious easy-cost function that can capture the major variables of differences and similarities in unsupervised learning. Because natural systems feature smooth dynamics, if an unsupervised objective function remains static during the training process, an opportunity is missed. Smooth dynamics should be introduced in the absence of concrete monitoring. Dynamic goal functions, as opposed to static cost functions, enable greater use of the progressive and unpredictable knowledge gained through pseudo supervision. In this study, they present Dynamic Autoencoder (DynAE), a new deep clustering model that eliminates the clustering reconstruction trade-off by gradually and seamlessly removing the reconstruction objective function in favor of a

construction one. In comparison to the most relevant deep clustering algorithms, experimental evaluations on benchmark datasets (MNIST-full, MNIST-test, USPS, and Fashion-MNIST) reveal that our methodology achieves state-of-the-art outcomes in terms of ACC and NMI.

The problem addressed in this paper [31] is: Clustering with deep autoencoders has received a lot of attention in recent years. Current methods rely on learning embedded features and clustering data points in the latent space at the same time. Although many deep clustering algorithms beat shallow models in achieving good findings on a variety of high-semantic datasets, a major flaw in such models has gone unnoticed. The embedded clustering objective function may distort the latent space by learning from faulty pseudo-labels in the absence of concrete supervisory signals. As a result, the network can learn non-representative features, lowering its discriminative ability and resulting in inferior pseudo-labels. Modern autoencoder-based clustering articles advocate using the reconstruction loss for pretraining and as a regularizer during the clustering phase to mitigate the effect of random discriminative features. Feature Drift can, however, be caused by a clustering reconstruction trade-off. The authors suggest ADEC (Adversarial Deep Embedded Clustering), a novel autoencoder-based clustering model that uses adversarial training to handle a dual problem, namely, Feature Randomness and Feature Drift. They use benchmark real datasets (MNIST-full, MNIST-test, USPS, Fashion-MNIST, Reuters-10K, and Mice Protein) to empirically illustrate the applicability of their model for dealing with these difficulties. The researchers' model outperforms state-of-the-art autoencoder-based clustering approaches in terms of ACC and NMI.

For image clustering, the authors of [32] suggest a self-supervised Gaussian Attention network (GATCluster). GATCluster delivers semantic cluster labels without further post-processing, rather than extracting intermediate features first and then conducting the standard clustering technique. The Label Feature Theorem is used to ensure that the learned features are one-hot encoded vectors and that trivial solution are avoided. They created four self-learning tasks with the restrictions of transformation invariance, separability maximization, entropy analysis, and attention mapping to train the GATCluster unsupervised. The transformation invariance and separability maximization tasks, in particular, are used to understand the relationships between sample pairs. The goal of the entropy analysis task is to avoid finding simple solutions. They created a self-supervised attention method that incorporates a parameterized attention module and a soft attention loss to capture object-oriented semantics. During the training process, all of the clustering guiding signals are self-generated. Furthermore, they create a memory-efficient two-step learning approach for grouping large-size images. Extensive trials show that their suggested method outperforms the current state-of-the-art image clustering benchmarks (CIFAR-10, CIFAR-100, STL-10, imageNet-

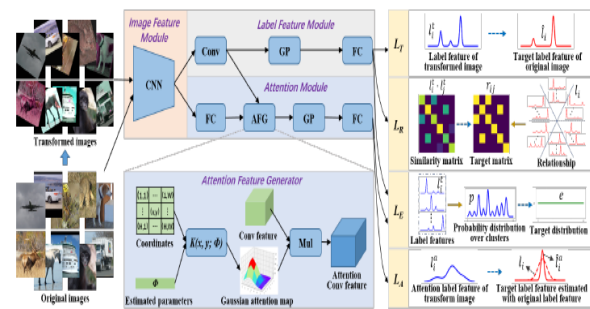


Figure 6: GATCluster framework. CNN is a convolutional neural network, GP means global pooling, Mul represents channel-independent multiplication, Conv is a convolution layer, FC is a fully connected layer, and AFG represents an attention feature generator [32].

10, ImageNet-dogs, and Tiny-ImageNet) in terms of ACC, NMI, and ARI.

Deep learning has recently demonstrated its ability to learn strong feature representations for images. The work of image clustering necessitates appropriate feature representations to capture the data distribution and, as a result, distinguish data points from one another. Often, these two aspects are dealt with independently, and thus, traditional feature learning alone does not suffice in partitioning the data meaningfully. Variational Autoencoders (VAEs) naturally lend themselves to learning data distributions in a latent space. The authors [33] suggest a method based on VAEs that uses a Gaussian Mixture before helping cluster the images appropriately since they seek to efficiently differentiate between distinct clusters in the data. They learn the parameters of both the prior and posterior distributions at the same time. Their method represents a true Gaussian Mixture VAE. In this way, their system learns a prior that captures the latent distribution of the images as well as a posterior that aids in data point discrimination. They also suggest a new reparametrization of the latent space that includes both discrete and continuous variables. One important takeaway is that, unlike existing methods, their method generalizes well across diverse datasets without the use of pre-training or learned models, allowing it to be trained from scratch in an end-to-end manner. They demonstrate our efficacy and generalizability in the lab by achieving state-of-the-art outcomes on a variety of datasets using unsupervised approaches. To the best of their knowledge, they are the first to use VAEs for image clustering on real image datasets (MNIST, Fashion-MNIST, STL-10, CIFAR10, CIFAR100, and FRGCv2) in an unsupervised manner and the accuracy evaluation criteria.

The authors of this research [34] deviate from current work by advocating the SCAN method (Semantic Clustering by Adopting Nearest neighbors), a two-step strategy in which feature learning and clustering are separated. To obtain semantically relevant features, a self-supervised task from representation learning is used first. Second, in a learnable clustering strategy, they employ the collected features as a prior. They accomplish so by removing cluster learning's capacity to rely on low-level



features, which are present in existing end-to-end learning systems. In terms of classification accuracy, they surpass state-of-the-art approaches by substantial margins, with +26,6 % on CIFAR10, +25,0 % on CIFAR100-20, and +21,3 % on STL10, respectively. Furthermore, their technology is the first to successfully classify images on a large-scale dataset.

In this paper [35], the authors offer a new deep image clustering framework for learning a category-style latent representation (Deep Clustering with Category-Style representation (DCCS) for unsupervised image clustering), in which the category information is decoupled from the image style and may be used directly for cluster assignment. Mutual information maximization is used to embed relevant information in the latent representation to achieve this goal. Furthermore, the augmentation-invariant loss is used to separate the representation into two parts: category and style. Last but not least, the latent representation is given a prior distribution to ensure that the elements of the category vector can be used as probabilities over clusters. Extensive tests show that the suggested method significantly outperforms state-of-the-art approaches on a variety of public datasets (MNIST and Fashion-MNIST) in terms of ACC, NMI, and ARI.

The study's authors [36] proposed Deep Robust Clustering (DRC). Unlike existing methods, DRC approaches deep clustering from two perspectives: semantic clustering assignment and representation features, which can simultaneously improve inter-class and intra-class diversities. Furthermore, by examining the internal relationship between mutual information and contrastive learning, they established a generic framework that may change maximizing mutual information into minimizing contrastive loss. They used it to learn invariant features and robust clusters in DRC with great success. Extensive tests on six widely used deep clustering benchmarks (CIFAR-10, CIFAR-100, STL-10, imageNet-10, ImageNet-dogs, and Tiny-ImageNet) show that DRC outperforms them in terms of both stability and accuracy. For example, on CIFAR-10, they achieved a mean accuracy of 71.6%, which is 7.1% higher than current values.

In this research [37], they introduced Contrastive Clustering (CC), a one-stage online clustering algorithm that performs explicit instance- and cluster-level contrastive learning. To be more exact, the positive and negative instance pairs for a given dataset are created using data augmentation and then projected into a feature space. In this case, instance- and cluster-level contrastive learning are carried out in the row and column space, respectively, by maximizing positive pair similarities while minimizing negative pair similarities. Their main finding is that the feature matrix's rows can be thought of as soft labels, for instance, and the columns can be thought of as cluster representations. The model learns row representations and cluster assignments in an end-to-end way by maximizing the instance- and cluster-level contrastive loss at the same time. On six challenging image benchmarks (CIFAR-10, CIFAR-100, STL-10, imageNet-10, ImageNet-dogs, and Tiny-ImageNet),

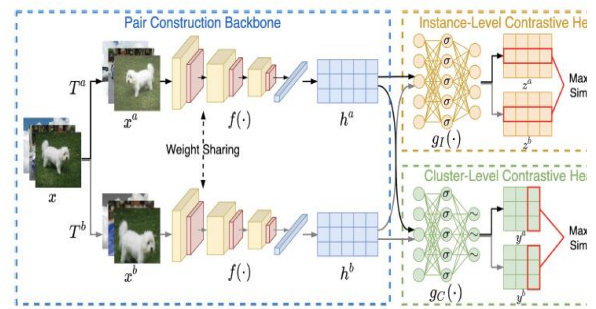


Figure 7: Contrastive Clustering framework. Two data augmentations are used to create data pairs. One shared deep neural network is utilized to extract features from distinct augmentations given data pairs. To project the features into the row and column space, two distinct MLPs (denotes the ReLU activation and denotes the Softmax operation to produce soft labels) are utilized to undertake instance- and cluster-level contrastive learning, respectively [37].

extensive experimental data shows that CC beats 17 competitive clustering approaches. On the CIFAR-10 (CIFAR-100) dataset, in particular, CC obtains an NMI of 0.705 (0.431), which is a performance gain of up to 19% (39%) above the best baseline.

The authors of this paper [38] propose learning an autoencoder embedding and then searching for the underlying manifold using it. They then cluster this using a shallow clustering technique rather than a deeper network for simplicity. They investigated a variety of local and global manifold learning methods on both raw data and autoencoder embeddings, concluding that UMAP in their framework is capable of determining the optimal clusterable manifold of the embedding. This shows that using local manifold learning on an autoencoder embedding to find higher-quality clusters is a good idea. They show numerically that their method outperforms the existing state-of-the-art on a variety of image and time-series datasets (MNIST, MNIST-test, USPS, Fashion, Pendigits, and HAR) including outperforming the current state-of-the-art on numerous in terms of ACC and NMI. They believe these findings point to a viable research direction in deep clustering.

SPICE, a Semantic Pseudo-labeling framework for Image Clustering, is presented in this work [39]. SPICE generates pseudo-labels by self-learning and directly employs the pseudo-label-based classification loss to train a deep clustering network, rather than requiring indirect loss functions as required by the recently proposed approaches. The core idea behind SPICE is to use a semantically-driven paradigm to improve the clustering network by combining the discrepancy between semantic clusters, similarity across instance samples, and semantic consistency of local samples in an embedding space. To train a clustering network by unsupervised representation learning, a semantic-similarity-based pseudo-labeling approach was presented initially. A local semantic consistency principle is employed to pick a set of consistently labeled samples based on the initial clustering results, and a semi-pseudo-labeling technique (SPICE-

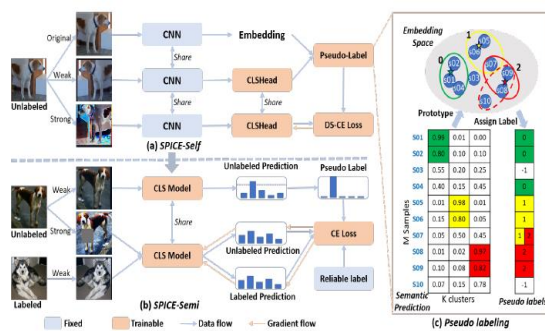


Figure 8: Representation of the SPICE framework. (a) SPICE-Self uses pseudo labeling to train a classification model, with CNN fixed after pretraining using representation learning. (b) SPICE-Semi retraining the classification model by semi-pseudo-labeling, in which reliable labels are chosen from the SPICE-Self findings based on the local consistency of nearby samples. (c) A simple example of pseudo labeling, with red, green, and blue indicating different clusters [39].

Semi) is adopted for performance boosting. On six typical benchmark datasets, including STL10, Cifar10, Cifar100-20, ImageNet-10, ImageNet-Dog, and Tiny ImageNet, extensive studies show that SPICE outperforms existing approaches. In terms of adjusted rand index, normalized mutual information, and clustering accuracy, the proposed SPICE technique improves the existing best results by roughly 10% on average.

Unsupervised image clustering approaches are prone to incorrect predictions and overconfident outcomes since they use alternate objectives to indirectly train the model. To address these issues, the current study [40] provides a new RUC model that is based on resilient learning. RUC is unique in that it uses the pseudo-labels of existing picture clustering algorithms as a noisy dataset with potentially misclassified samples. Its retraining method can correct mismatched knowledge and reduce the problem of overconfidence in forecasts. The model's flexible structure allows it to be used as an add-on module to existing clustering algorithms, allowing them to perform better on a variety of datasets (CIFAR-10, CIFAR-20, STL-10). Extensive studies show that the suggested approach can improve model confidence and gain additional robustness against adversarial noise by properly calibrating it. RUC is a module that may be added to any off-the-shelf unsupervised learning method to improve its performance. RUC is motivated by a desire to learn more. It separates clustered data points into clean and noisy sets before fine-tuning the clustering results. SCAN and TSUC, two state-of-the-art unsupervised clustering algorithms, exhibited considerable performance increases with RUC. (STL-10 : 86.7 %, CIFAR-10 : 90.3 %, CIFAR-20 : 54.3 %).

In the research [41], the authors use instance discrimination and feature decorrelation to propose a clustering-friendly representation learning approach. The principles of classical spectral clustering inspired their deep-learning-based representation learning method. Instance discrimination discovers data commonalities,

whereas feature decorrelation eliminates redundant correlation between features. They employ a method of instance discrimination in which knowing individual instance classes leads to learning similarities between examples. They show that the methodology may be extended to learning a latent space for clustering through comprehensive experimentation and examination of the benchmark datasets (CIFAR-10, CIFAR-100, STL-10, ImageNet-10, and ImageNet-Dog). For learning, they create new softmax-formulated decorrelation constraints. Their method achieves an accuracy of 81,5% and 95,4% in image clustering tests using CIFAR-10 and ImageNet-10, respectively. They also demonstrate that the softmax-formulated constraints work with a variety of neural networks.

The authors of this study [42] introduced Mixture of Contrastive Experts (MiCE), a unified probabilistic clustering approach that concurrently uses contrastive learning's discriminative representations and a latent mixture model's semantic structures. MiCE uses a gating function to partition an unlabeled dataset into subsets according to latent semantics and numerous experts to differentiate separate subsets of instances allotted to them in a contrastive learning method, which is motivated by the mixing of experts. They designed a scalable form of the Expectation-Maximization (EM) algorithm for MiCE and showed proof of convergence to overcome the nontrivial inference and learning challenges caused by latent variables. They tested MiCE's clustering performance empirically on four frequently used natural image datasets (CIFAR-10, CIFAR-100, STL10, and ImageNet-Dog). MiCE outperforms a variety of earlier approaches and provides a strong contrastive learning baseline using the criteria ACC, NMI, and ARI.

The problem with this study [43] is that, as measured by curated class-balanced datasets, unsupervised feature learning has made significant progress with contrastive learning based on instance discrimination and invariant mapping. Natural data, on the other hand, maybe highly linked and skewed. The supposed instance distinction clashes with natural between-instance similarity, resulting in inconsistency in training and poor performance. The goal is to identify and integrate between-instance similarity into contrastive learning via cross-level discrimination (CLD) between instances and local instance groups rather than instance grouping directly. While attraction inside each instance's augmented perspectives forces invariant mapping, between-instance similarity comes via common repulsion against instance groupings. The batch-wise and cross-view comparisons also help to increase contrastive learning's positive/negative sample ratio and produce improved invariant mapping. We impose both grouping and discrimination objectives on characteristics obtained separately from a shared representation to achieve both goals. For the first time, they also present normalized projection heads and unsupervised hyper-parameter adjustment. CLD is a lean and powerful add-on to existing methods (e.g., NPID, MoCo, InfoMin, BYOL) on highly correlated, long-tail, or balanced datasets, as demonstrated by considerable experimentation. It not only sets new

benchmarks (CIFAR-10, CIFAR-100, and ImageNet) for self-supervision, semi-supervision, and transfer learning, but it also outperforms MoCo v2 and SimCLR on every reported performance achieved with a far larger compute in terms of accuracy. Unsupervised learning is effectively extended to natural data with CLD, bringing it closer to real-world applications.

## 4 Discussion

Based on this short and selective survey of deep clustering algorithms, we make the following observations:

- most deep clustering techniques have been tested in the area of image recognition;
- performances of these techniques are great in terms of recognition accuracy, as the study of [35], where obtained recognition accuracy achieves 98.9 %;
- most studies enhance the embedding of the data into a lower-dimensional space;
- several researchers use the MNIST database for experimentation and k-means algorithm for results comparison;
- we remark that the appearance of the hybrid version of Autoencoder gives satisfactory results too ;
- deep learning is a technology that continues to mature and has been applied to pattern recognition to great effect;
- we have identified the name of the proposed method, the category to which it belongs, a dataset of each approach with the methods of comparison, these are seen in table 1;
- Table 1 summarizes the sorted works in chronological order. We observed in Table 1, that the MNIST dataset provides good results compared to other databases like USPS; CIFAR-10; CIFAR-100;

Table 1: General comparison of various deep clustering algorithms for image recognition.

References	Method	Category	Dataset	Compared results with	Obtained results
[24]	SNNL Soft Nearest Neighbor Loss	AE	MNIST; Fashion-MNIST; and EMNIST Balanced.	SNNL-2; SNNL-4; Baseline AE; DEC; VaDE; N2D; and ClusterGAN;....	1. The best accuracy (acc)=96.2% with MNIST; 2. The best NMI=90.3% with MNIST; 3. The best ARI=91.8% with MNIST;
[25]	MPCC Matching Priors and Conditionals for Clustering	AE	MNIST; Onmigtot; FMNIST; CIFAR-10; and CIFAR-20.	DEC; VADE; InfoGAN; ClusterGAN; DAC; IMSAT (VAT); ADC; SCAE; and IIC.	The best accuracy (acc)= 98.76 $\pm$ 0.03% with MNIST;
[10]	DeepCluster is a new clustering strategy for large-scale end-to-end convent training.	AE	ImageNet; Places.	The methods have a standard AlexNet architecture.	The best is 73.7% on classification with deepCluster
[11]	Low-rank Constrained Deep Autoencoder for Subspace Clustering (LRAE)	AE	MNIST; COIL-100, and ORL	SSC; LRR; LRSC; LSR; AESC, and PARTY.	1. The best accuracy (acc)= 81.49 $\pm$ 2.19 with ORL; 2. The best NMI= 90.77 $\pm$ 2.01 with ORL; 3. The best ARI= 73.92 $\pm$ 2.11 with ORL;
[12]	Hybrid Autoencoder (BAE), the combination of three AE-based models—the convolutional autoencoder (CAE), adversarial autoencoder (AAE), and stacked autoencoder (SAE)	AE	MNIST and CIFAR-10.	Fuzzy objective function algorithm (FCM), Spectral clustering algorithm (SC), Low-rank representation algorithm (LRR), LSR1 and LSR2 are the variants of the least-squares regression (LSR), SLRR is the scalable LRR, LSC-R and LSC-K are the variants of the large-scale spectral clustering (LSC) algorithms, NMF is the non-negative matrix factorization algorithm, ZAC is the Zeta function based agglomerative clustering algorithm, and DEC is the deep embedding clustering algorithm.	1. The best accuracy (acc)= 83.67% with MNIST; 2. The best NMI= 80.85% with MNIST;

References	Method	Category	Dataset	Compared results with	Obtained results
[14]	Clustering with Optimal Clustering (OT) is a new approach where the embedding is performed by a differentiable model such as a deep neural network	GAN	MNIST and CIFAR10	k-means; AE + k-means; soft k-means and soft k-means (p)	The best NMI= 85.10% with MNIST;
[15]	ClusterGAN is a deep Generative Adversarial Clustering Network.	GAN	MNIST; USPS; FRGC; CIFAR-10 and STL-10.	Kmeans; N-Cuts; SC-LS; AC-PIC; SEC and LDMGI.	1. The best accuracy (acc)= 97% with USPS; 2. The best NMI= 93.10% with USPS; 3. The best accuracy (acc)= 96.4% with MNIST; 4. The best NMI= 92.10% with MNIST;
[27]	IFCA a new framework dubbed the Iterative Federated Clustering Algorithm	AE	Rotated MNIST; and Rotated CIFAR	The global model for IFCA; and local model	The best accuracy (acc)= 95.25 ± 0.40% with Rotated MNIST;
[9]	Deep Convolutional Embedded Clustering (DCEC)	AE	MNIST -full; MNIST-test; USPS	1. Deep Embedded Clustering (DEC); 2. K-means; 3. Stacked AutoEncoders (SAE).	1. The best accuracy (acc)=88.97% with MNIST-full; 2. The best NMI=88.49% with MNIST-full.
[3]	ASPC-DA is an Adaptive Self-Paced deep Clustering with Data Augmentation	AE	MNIST-full; MNIST-test; USPS and Fashion		1. The best accuracy (acc)=98.8% with MNIST-full; 2. The best NMI=96.6% with MNIST-full.
[16]	Kingdra is a framework that leverages semi-supervised models	AE	MNIST; STL; CIFAR10; Reuters and 20news.	k-means; AC; DEC; Deep RIM; and IMSAT...	The best accuracy (acc)=98.5% with MNIST.
[28]	A novel two-stage algorithm in which an embedding module for pretraining precedes a refining module that concurrently performs embedding and class assignment	AE	CIFAR-10; CIFAR-20; and STL-10	Random network; k-means; Autoencoder (AE); SWWAE; GAN; JULE; DEC; DAC; DeepCluster; ADC; and IIC	The best accuracy (acc)= 81% with CIFAR-10;
[29]	PICA a novel deep clustering method named PartItion Confidence mAximisation	AE	CIFAR-10; CIFAR-100; STL-10; ImageNet-10; ImageNet-Dogs and Tiny-ImageNet	K-means; SC; AC; NMF; AE; DAE; and IIC;...	1. The best accuracy (acc)= 87% with ImageNet-10; 2. The best NMI= 80.2% with ImageNet-10; 3. The best ARI= 76.1% with ImageNet-10;
[17]	MIADM is an approximate alternating direction method.	AE	USPS; MNIST-test; MNIST-full; YTF; CMU-PIE and FRGC	SR-K-means; DEPICT; DCN (K-means based) and DEC (KL based).	1. The best accuracy (acc)=97.9% with USPS; 2. The best NMI=94.8% with USPS.
[18]	IIC Invariant Information Clustering	AE	STL10; CIFAR10; CFR100-20 and MNIST.	Random network; Kmeans; Spectral clustering; Triplets; Variational Bayes AE and DeepCluster 2018,.....	The best accuracy (acc)=99.2% with MNIST;

[19]	LTVAE latent tree variational autoencoder.	VAE	MNIST; STL-10; Reuters and HHAR	AE+GMM; VAE+GMM; DEC and DCN.	The best accuracy (acc)=90% with STL-10;
[37]	CC Contrastive Clustering is an online clustering method	AE	CIFAR-10; CIFAR-100; STL-10; ImageNet-10; ImageNet-Dogs; and Tiny-ImageNet	k-means; SC; AC; NMF; DEC; JULE; VaE; DCGAN; DeCNN; DCCM; IIC; and PICA;...	1. The best accuracy (acc)= 89.3% with ImageNet-10; 2. The best NMI= 85.9% with ImageNet-10; 3. The best ARI= 82.2% with ImageNet-10;
[38]	N2D: (Not Too) Deep Clustering via Clustering the Local Manifold of an Autoencoded Embedding	AE	MNIST; MNIST-test; USPS; Fashion; pendigits; and HAR	k-means; SC; GMM; DEC; DCN; JULE; VaDE; DEPICT; DBC; and ASPC-DA;...	1. The best accuracy (acc)= 97.9% with MNIST; 2. The best NMI= 94.2% with MNIST;
[30]	DynAE Dynamic Autoencoder, a novel model for deep clustering that addresses a clustering–reconstruction trade-off.	AE	MNIST-full; MNIST-test; USPS; and Fashion-MNIST	K-Means; GMM; LSNMF; AC; SSC-OMP; EnSC; LMVSC; RBF K-Means –; DEC; JULE; and DEPICT;....	1. The best accuracy (acc)= 98.7% with MNIST-full; 2. The best NMI= 96.4% with MNIST-full;
[31]	ADEC (Adversarial Deep Embedded Clustering) is a novel autoencoder-based clustering model	AE	MNIST-full; MNIST-test; USPS; Fashion-MNIST; REUTERS-10K; and Mice Protein	DEC*; IDEC*; k-means; GMM; LSNMF; AC; RBF k-means; .....	1. The best accuracy (acc)= 98.6% with MNIST-full; 2. The best NMI= 96.1% with MNIST-full;
[13]	ClusterGAN method is a new mechanism for clustering using GANs (Generative Adversarial Networks )	GAN	Synthetic data; MNIST; Fashion-MNIST; 10x_73k and Pendigits.	WGAN (normal); WGAN (One-Hot) and Info GAN.	The best accuracy (acc)= 95% with MNIST;
[32]	SPICE, a Semantic Pseudo-labeling framework for Image Clustering	AE	STL10; ImageNet-10; ImageNet-Dog-15; Cifar10; Cifar100-20; and Tiny-ImageNet-200	JULE; DEC; DAC; DeepCluster; DDC; IIC; DCCM; GATCluster; PIC; and CC	1. The best accuracy (acc)= 93.8% with STL10; 2. The best NMI= 87.2% with STL10; 3. The best ARI= 87% with STL10;
[40]	RUC is inspired by robust learning. RUC's novelty is at utilizing pseudo-labels of existing image clustering models as a noisy dataset.	AE	CIFAR-10; CIFAR-20; and STL-10	k-means; SC; Triplets; AE; GAN; JULE; DAC; DEC; DeepCluster; IIC; TSUC and SCAN;...	The best accuracy (acc)= 90.1% with CIFAR-10;
[33]	A method based on VAEs where we use a Gaussian Mixture before help cluster the images accurately	VAE	STL-10; CIFAR10; MNIST; and Fashion-MNIST	k-means; AE+k-means; and DEC	The best accuracy (acc)= 98.4% with MNIST;
[20]	DCC Deep Continuous Clustering	AE	MNIST; Coil100; YTF; YaleB; Reuters and RCV1	k-means++; AC-W; DBSCAN; SEC and LDMGI;.....	1. The best accuracy (acc)=91.3% with MNIST; 2. The best accuracy (acc)=98.5% with YaleB;



[41]	IDFD a clustering-friendly representation learning method using instance discrimination and feature decorrelation.	AE	CIFAR-10; CIFAR-100; STL-10; Imagenet-10; and Imagenet-Dog	AE; DEC; DAC; DCCM; ID; IIC; IDFO; and SCAN	1. The best accuracy (acc)= 95.4% with ImageNet-10; 2. The best NMI= 89.8% with ImageNet-10; 3. The best ARI= 90.1% with ImageNet-10;
[42]	MiCE Mixture of Contrastive Experts, a unified probabilistic clustering framework	AE	CIFAR-10; CIFAR-100; STL-10; and Imagenet-Dog	K-means; AE; DHOG; DAC; DCCM; MMDC; IIC; IDFO; and MoCo	1. The best accuracy (acc)= 83.5% with CIFAR-10; 2. The best NMI= 73.7% with CIFAR-10; 3. The best ARI= 69.8% with CIFAR-10;
[34]	SCAN Semantic Clustering by Adopting Nearest neighbors	AE	CIFAR10; CIFAR100- 20; STL10; and ImageNet	k-means; SC; Triplets; JULE; AEVB; SAE; DAE; GAN; DAC; and IIC	1. The best accuracy (acc)= 88.3% with CIFAR10; 2. The best NMI= 79.7% with CIFAR10; 3. The best ARI= 77.2% with CIFAR10;
[43]	CLD cross-level discrimination	AE	STL10; CIFAR10; CIFAR100; and ImageNet100	DeepCluster; MoCo; Exemplar; Inv. Spread; NPID; and BYOL;....	1. The best retrieval= 78.6% with CIFAR-10; 2. The best NMI= 69% with CIFAR-10; 3. The best kNN= 86.7% with CIFAR-10;
[23]	DCCM is a deep comprehensive correlation mining	AE	CIFAR-10; CIFAR-100; STL-10; ImageNet-10; Imagenet-dog-15; and Tiny-ImageNet.	K-means; SC; AC; NMF; AE; and DAE;.....	1. The best accuracy (acc)=60.8% with ImageNet-10; 2. The best NMI=71% with ImageNet-10; 3. The best ARI=55.5% with ImageNet-10;
[21]	DGG: Deep clustering via a Gaussian mixture variational autoencoder (VAE) with Graph embedding	VAE	MNIST; STL-10; Reuters and HHAR.	AE+GMM; DEC; IMSAT; VaDE; SpectralNet; and LTVAE.	The best accuracy (acc)=97.58±0.1% with MNIST;
[22]	A joint learning framework for discriminative embedding and spectral clustering	AE	MNIST-full; MNIST-test; USPS; Fashion-10; and YTF.	K-means; SC-Ncut; SC-LS; NMF; AC-GDL; and DASC;.....	1. The best accuracy (acc)=98% with MNIST-test; 2. The best NMI=94.6% with MNIST-test;
[35]	DCCS a novel deep image clustering framework to learn a category-style latent representation	AE	MNIST; and Fashion-MNIST	k-means; SC; AC; NMF; DEC; JULE; VaDE; DEPICT; IMSAT; ClusterGan; IIC; and DLS-clustering;...	1. The best accuracy (acc)= 98.9% with MNIST; 2. The best NMI= 97% with MNIST; 3. The best ARI= 97.6% with MNIST;
[36]	DRC Deep Robust Clustering	AE	CIFAR-10; CIFAR-100; STL-10; ImageNet-10; Imagenet-dog-15; and Tiny-ImageNet	k-means; SC; AC; NMF; DEC; JULE; VaDE; DEPICT; IMSAT; DCCM; IIC; and PICA;...	1. The best accuracy (acc)= 88.4% with ImageNet-10; 2. The best NMI= 83% with ImageNet-10; 3. The best ARI= 79.8% with ImageNet-10;

## 5 Proposed taxonomy of deep clustering

Figure 9 illustrates the taxonomy of Deep Clustering techniques that we describe, which in turn indicates the study's structure. The basic algorithmic structure, network

architecture, loss functions, and training optimization methodologies for deep clustering systems vary (or learning the parameters).

We focus on deep learning for clustering approaches in this paper, where those approaches either use deep learning for grouping (or partitioning) the data and/or creating low-rank deep representations or embeddings of

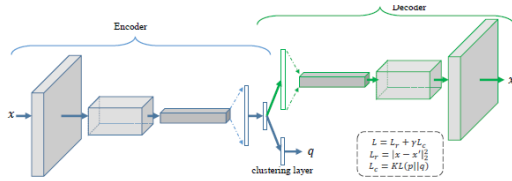


Figure 10: The structure of deep convolutional embedded clustering (DCEC). It is composed of convolutional autoencoders and a clustering layer connected to the embedded layer of autoencoders [9].

the data, which could play a significant supporting role as a building block of supervised learning, among other goals. There are numerous approaches to developing a taxonomy of deep clustering algorithms; in this study, we took the approach of seeing the methods as a process. As a result, we provide a simplified taxonomy based on deep clustering algorithms' overall procedural structure or architecture. Beginners and experienced readers will benefit from the simplified classification.

We have chosen to propose to divide deep learning into three categories:

#### AE-based deep clustering:

Artificial neural networks (ANNs) are a type of machine learning model made up of numerous nodes grouped in layers that compute an output depending on node activation mediated by weights in the connections between them. ANNs are capable of solving a variety of machine learning tasks, including classification, regression, and dimensionality reduction [44].

A neural network that has been trained to duplicate its input to its output is called an autoencoder. It has a hidden layer  $h$  on the inside that defines the code used to represent the input. The network is made up of two parts: an encoder function  $h=f(x)$  and a decoder function  $r=g(h)$  that provides a reconstruction. Figure 10 illustrates this architecture. If an autoencoder only succeeds in learning to set  $g(f(x)) = x$  everywhere, it isn't particularly useful. Autoencoders, on the other hand, are meant to be incapable of flawless copying. They are usually limited in some way, allowing them to copy only roughly and only input that closely mimics the training data. Because the model must prioritize which features of the input should be duplicated, it frequently discovers interesting data attributes. The following is an overview of representative methods of Autoencoder:

1. Deep Convolutional Embedded Clustering (DCEC): the DCEC system is composed of Convolutional Clustering (CAE) and a clustering layer that is connected to the embedded layer of CAE [9]. Each embedded point  $z_i$  of the input image  $x_i$  is mapped into a soft label by the clustering layer. The Kullback-Leibler divergence (KL divergence) between the distribution of soft labels and the precisely defined distribution is then defined as the clustering loss  $L_c$ . The clustering loss leads the embedded features to be resistant to forming clusters, and CAE is used to learn embedded features.

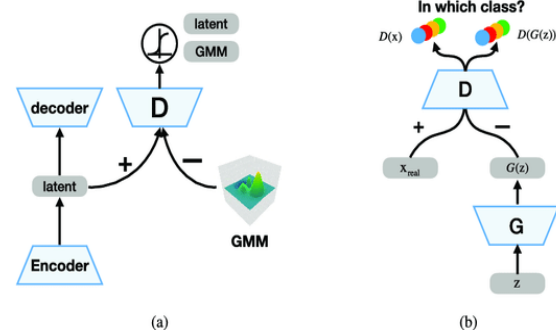


Figure 11: GAN-based deep clustering [47].

The objective of DCEC is:

$$L = L_r + \gamma L_c \quad (1)$$

where  $L_r$  and  $L_c$  are reconstruction loss and clustering loss respectively, and  $\gamma > 0$  is a coefficient that controls the degree of distorting embedded space. When  $\gamma = 1$  and  $L_r \equiv 0$ , (1) reduces to the objective of DEC.

2. Deep Clustering Network (DCN): this method [45] which combines the autoencoder and the k-means algorithm, is one of the most remarkable in the field. It pre-trains an autoencoder in the first stage. The reconstruction loss and the k-means loss are then optimized together. Because k-means relies on discrete cluster assignments, it necessitates the employment of a different optimization procedure. When compared to other methods, DCN's goal is simple, and the computing complexity is modest;
3. Hybrid Autoencoder (HAE): [46] CAE (convolutional autoencoder), VAE (adversarial autoencoder), and SAE (stacked autoencoder) combine the advantages of three autoencoders to learn low and high-level feature representation.

#### GAN-based deep clustering:

In recent years, the Generative Adversarial Network (GAN) has become a popular deep generative model. A min-max adversarial game is established between two neural networks in the (GAN) [47]: a generating network,  $G$ , and a discriminative network,  $D$ . The generative network attempts to map a sample  $z$  from a prior distribution  $p(z)$  to the data space, whereas the discriminative network attempts to compute the probability that an input is a real sample from the data distribution rather than one created by the generative network. GAN is an exciting idea since it offers an adversarial approach to matching the distribution of data or its representations to an arbitrary prior distribution.

#### VAE-based deep clustering:

[48] VAE is a generative variant of AE since it causes AE's latent code to follow a predetermined distribution. VAE blends variational Bayesian approaches with neural network flexibility and scalability. It applies neural networks to the conditional posterior and uses stochastic gradient descent and standard backpropagation to optimize the variational inference objective. It employs the reparameterization of the variational lower bound to produce a simple, differentiable, unbiased lower bound

estimator. In nearly every model with continuous latent variables, this estimator can be utilized for efficient approximate posterior inference:

1. Deep clustering via a Gaussian mixture VAE with Graph embedding (DGG): [21] a new VAE-based model that assumes the latent variables have a tree structure;
2. Variational Deep Embedding (VaDE): introduces a VAE-based generative model that assumes the latent variables are a mixture of Gaussians with trainable means and variances [49];
3. Latent Tree Variational Autoencoder (LTVAE): a VAE-based model that assumes the latent variables have a tree structure [19].

## 6 Conclusion and perspectives

Deep learning is made up of a number of well-known and effective models that are used to solve a variety of problems [50].

In the context of deep clustering, we have presented, in this article, an introductory study of the main deep unsupervised learning algorithms that have been found in the last 3-4 years in the literature.

We have presented an overview of clustering methods and algorithms for deep learning. We noticed the multitude of contributions developed in the area of image recognition and we studied and synthesized different recent works in this context.

We have proposed a taxonomy of clustering with deep learning algorithms based on previous studies and some treated representative methods in the survey.

This study is the first step of our research for which we can consider several future extensions, such as exploring the possibilities of hybridization between different deep clustering approaches and their application in evolving patterns. We will be able to make a comparative study of the performance of deep learning approaches based on the autoencoder, such as the work of [51]. We will be able to apply the deep clustering method in fields such as face recognition, etc [52].

## Acknowledgment

The authors would like to thank the DGRSDT (General Directorate of Scientific Research and Technological Development) - MESRS (Ministry of Higher Education and Scientific Research), ALGERIA, for the financial support of LISCO Laboratory.

## References

- [1] A. Chefrour, and L. Souici-Meslati (2019). AMF-IDBSCAN: Incremental Density Based Clustering Algorithm Using Adaptive Median Filtering Technique. *Informatica*, vol. 43(4). <https://doi.org/10.31449/inf.v43i4.2629>
- [2] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long (2018). A survey of clustering with deep learning: From the perspective of network architecture. *IEEE Access*, vol. 6, pp. 39501-39514. <https://doi.org/10.1109/access.2018.2855437>
- [3] X. Guo, X. Liu, E. Zhu, X. Zhu, M. Li, X. Xu, and J. Yin (2019). Adaptive self-paced deep clustering with data augmentation. *IEEE Transactions on Knowledge and Data Engineering*, vol. 32(9), pp. 1680-1693. <https://doi.org/10.1109/tkde.2019.2911833>
- [4] C. C. Wang, K. L. Tan, C. T. Chen, Y. H. Lin, S. S. Keerthi, D. Mahajan, and C. J. Lin (2018). Distributed newton methods for deep neural networks. *Neural computation*, vol. 30(6), pp. 1673-1724. [https://doi.org/10.1162/neco\\_a\\_01088](https://doi.org/10.1162/neco_a_01088)
- [5] Z. Shen, H. Yang, and S. Zhang, S (2021). Deep network with approximation error being reciprocal of width to power of square root of depth. *Neural Computation*, vol. 33(4), pp. 1005-1036. [https://doi.org/10.1162/neco\\_a\\_01364](https://doi.org/10.1162/neco_a_01364)
- [6] J. Gao, P. Li, Z. Chen, and J. Zhang (2020). A survey on deep learning for multimodal data fusion. *Neural Computation*, vol. 32(5), pp.829-864. [https://doi.org/10.1162/neco\\_a\\_01273](https://doi.org/10.1162/neco_a_01273)
- [7] E. Aljalbout, V. Golkov, Y. Siddiqui, M., Strobel, and D. Cremers (2018). Clustering with deep learning: Taxonomy and new methods. *arXiv preprint arXiv:1801.07648*.
- [8] G. C. Nutakki, B. Abdollahi, W. Sun, and O. Nasraoui (2019). An introduction to deep clustering. *Clustering Methods for Big Data Analytics*, pp. 73-89. Springer, Cham. [https://doi.org/10.1007/978-3-319-97864-2\\_4](https://doi.org/10.1007/978-3-319-97864-2_4)
- [9] X. Guo, X. Liu, E. Zhu, and J. Yin (2017, November). Deep clustering with convolutional autoencoders. In *International conference on neural information processing*, pp. 373-382. Springer, Cham. [https://doi.org/10.1007/978-3-319-70096-0\\_39](https://doi.org/10.1007/978-3-319-70096-0_39)
- [10] M. Caron, P. Bojanowski, A. Joulin, and M. Douze (2018). Deep clustering for unsupervised learning of visual features. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 132-149. [https://doi.org/10.1007/978-3-030-01264-9\\_9](https://doi.org/10.1007/978-3-030-01264-9_9)
- [11] Y. Chen, L. Zhang, and Z. Yi (2018). Subspace clustering using a low-rank constrained autoencoder. *Information Sciences*, vol. 424, pp. 27-38. <https://doi.org/10.1016/j.ins.2017.09.047>
- [12] P. Y. Chen, and J. J. Huang (2019). A hybrid autoencoder network for unsupervised image clustering. *Algorithms*, vol. 12(6), pp.122. <https://doi.org/10.3390/a12060122>
- [13] S. Mukherjee, H. Asnani, E. Lin, and S. Kannan (2019, July). ClusterGAN: Latent space clustering in generative adversarial networks. In *Proceedings of*

- the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 01, pp. 4610–4617.
- [14] A. Genevay, G. Dulac-Arnold, and J. P. Vert (2019). Differentiable deep clustering with cluster size constraints.
  - [15] K. Ghasedi, X. Wang, C. Deng, and H. Huang (2019). Balanced self-paced learning for generative adversarial clustering network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4391–4400. <https://doi.org/10.1109/CVPR.2019.00452>
  - [16] Gupta, D., Ramjee, R., Kwatra, N., & Sivathanu, M. (2019, September). Unsupervised clustering using pseudo-semi-supervised learning. In *International Conference on Learning Representations*.
  - [17] M. Jabi, M. Pedersoli, A. Mitiche, and I. B. Ayed, (2019). Deep clustering: On the link between discriminative models and k-means. *IEEE transactions on pattern analysis and machine intelligence*. <https://doi.org/10.1109/TPAMI.2019.2962683>
  - [18] X. Ji, J. F. Henriques, and A. Vedaldi (2019). Invariant information clustering for unsupervised image classification and segmentation. <https://doi.org/10.1109/ICCV.2019.00996>
  - [19] X. Li, Z. Chen, L. K. Poon, and N. L. Zhang (2018). Learning latent superstructures in variational autoencoders for deep multidimensional clustering. *arXiv preprint arXiv:1803.05206*. iclr 2019.
  - [20] S. A. Shah, and V. Koltun (2018). Deep continuous clustering. *arXiv preprint arXiv:1803.01449*.
  - [21] L. Yang, N. M. Cheung, J. Li, and J. Fang (2019). Deep clustering by gaussian mixture variational autoencoders with graph embedding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6440–6449. <https://doi.org/10.1109/ICCV.2019.00654>
  - [22] X. Yang, C. Deng, F. Zheng, J. Yan, and W. Liu (2019). Deep spectral clustering using dual autoencoder network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4066–4075
  - [23] J. Wu, K. Long, F. Wang, C. Qian, C. Li, Z. Lin, and H. Zha (2019). Deep comprehensive correlation mining for image clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8150–8159. <https://doi.org/10.1109/ICCV.2019.00824>
  - [24] A. F. Agarap, and A. P. Azcarraga (2020, July). Improving k-Means Clustering Performance with Disentangled Internal Representations. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE. <https://doi.org/10.1109/IJCNN48605.2020.9207192>
  - [25] N. Astorga, P. Huijse, P. Protopapas, and P. Estévez (2020, August). MPCC: Matching Priors and Conditionals for Clustering. In *European Conference on Computer Vision*, pp. 658–677. Springer, Cham. [https://doi.org/10.1007/978-3-030-58592-1\\_39](https://doi.org/10.1007/978-3-030-58592-1_39)
  - [26] L. N. Darlow, and A. Storkey (2020). DHOG: Deep Hierarchical Object Grouping. *arXiv preprint arXiv:2003.08821*.
  - [27] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran (2020). An efficient framework for clustered federated learning. In *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, Canada.
  - [28] S. Han, S. Park, S. Park, S. Kim, and M. Cha (2020, August). Mitigating Embedding and Class Assignment Mismatch in Unsupervised Image Classification. In *16th European Conference on Computer Vision, ECCV 2020*. Springer Science and Business Media Deutschland GmbH. [https://doi.org/10.1007/978-3-030-58586-0\\_45](https://doi.org/10.1007/978-3-030-58586-0_45)
  - [29] J. Huang, S. Gong, and X. Zhu (2020). Deep semantic clustering by partition confidence maximisation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8849–8858. <https://doi.org/10.1109/CVPR42600.2020.00887>
  - [30] N. Mrabah, N. M. Khan, R. Ksantini, and Z. Lachiri (2020). Deep clustering with a Dynamic Autoencoder: From reconstruction towards centroids construction. *Neural Networks*, vol. 130, pp. 206–228. <https://doi.org/10.1016/j.neunet.2020.07.005>
  - [31] N. Mrabah, M. Bouguessa, and R. Ksantini (2020). Adversarial deep embedded clustering: on a better trade-off between feature randomness and feature drift. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/TKDE.2020.2997772>
  - [32] C. Niu, J. Zhang, G. Wang, and J. Liang (2020, August). Gatcluster: Self-supervised gaussian-attention network for image clustering. In *European Conference on Computer Vision*, pp. 735–751. Springer, Cham. [https://doi.org/10.1007/978-3-030-58595-2\\_44](https://doi.org/10.1007/978-3-030-58595-2_44)
  - [33] V. Prasad, D. Das, and B. Bhowmick (2020, July). Variational Clustering: Leveraging Variational Autoencoders for Image Clustering. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–10. IEEE. <https://doi.org/10.1109/IJCNN48605.2020.9207523>
  - [34] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, M. Proesmans, and L. Van Gool (2020, August).

- Scan: Learning to classify images without labels. In *European Conference on Computer Vision*, pp. 268-285. Springer, Cham.  
[https://doi.org/10.1007/978-3-030-58607-2\\_16](https://doi.org/10.1007/978-3-030-58607-2_16)
- [35] J. Zhao, D. Lu, K. Ma, Y. Zhang, and Y. Zheng (2020, August). Deep Image Clustering with Category-Style Representation. In *European Conference on Computer Vision*, pp. 54-70. Springer, Cham.  
[https://doi.org/10.1007/978-3-030-58568-6\\_4](https://doi.org/10.1007/978-3-030-58568-6_4)
- [36] H. Zhong, C. Chen, Z. Jin, and X. S. Hua (2020). Deep robust clustering by contrastive learning. *arXiv preprint arXiv:2008.03030*.
- [37] Y. Li, P. Hu, Z. Liu, D. Peng, J. T. Zhou, and X. Peng (2021). Contrastive Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35(10), pp. 8547-8555.
- [38] R. McConville, R. Santos-Rodriguez, R. J., Piechocki, and I. Craddock (2021, January). N2d:(not too) deep clustering via clustering the local manifold of an autoencoded embedding. In *25th International Conference on Pattern Recognition (ICPR)*, pp. 5145-5152. IEEE.  
<https://doi.org/10.1109/ICPR48806.2021.9413131>
- [39] C. Niu, and G. Wang (2021). SPICE: Semantic Pseudo-labeling for Image Clustering. *arXiv preprint arXiv:2103.09382*.
- [40] S. Park, S. Han, S. Kim, S. Kim, D. Park, S. Hong, and M. Cha (2021). Improving Unsupervised Image Clustering With Robust Learning. *Accepted at Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence (cs.AI); Machine Learning (cs.LG)*.
- [41] Y. Tao, K. Takagi, and K. Nakata (2021). Clustering-friendly Representation Learning via Instance Discrimination and Feature Decorrelation. *arXiv preprint arXiv:2106.00131. ICLR 2021 Workshop on Embodied Multimodal Learning (EML)*.
- [42] T. W. Tsai, C. Li, and J. Zhu (2021). MiCE: Mixture of Contrastive Experts for Unsupervised Image Clustering. *ICLR 2021*.
- [43] X. Wang, Z. Liu, and S. X. Yu (2021). Unsupervised Feature Learning by Cross-Level Instance-Group Discrimination. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12586-12595.
- [44] L. Amado, and F. Meneguzzi (2018). Q-Table compression for reinforcement learning. *The Knowledge Engineering Review*, vol. 33.  
<https://doi.org/10.1017/S0269888918000280>
- [45] B. Yang, X. Fu, N. D. Sidiropoulos, and M. Hong (2017, July). Towards k-means-friendly spaces: Simultaneous deep learning and clustering. In *International conference on machine learning*, pp. 3861-3870.
- [46] K. Gupta, M. Y. Raghuprasad, and P. Kumar (2018). A hybrid variational autoencoder for collaborative filtering. *arXiv preprint arXiv:1808.01006*.
- [47] F. Shoeleh, N. M. Yadollahi, and M. Asadpour. (2020). Domain adaptation-based transfer learning using adversarial networks. *The Knowledge Engineering Review*, vol. 35.  
<https://doi.org/10.1017/S0269888920000107>
- [48] K. L. Lim, X. Jiang, and C. Yi (2020). Deep clustering with variational autoencoder. *IEEE Signal Processing Letters*, vol. 27, pp. 231-235.  
<https://doi.org/10.1109/LSP.2020.2965328>
- [49] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou (2016). Variational deep embedding: An unsupervised and generative approach to clustering. *arXiv preprint arXiv:1611.05148*
- [50] W. Etaiwi, D. Suleiman, and A. Awajan (2021). Deep Learning Based Techniques for Sentiment Analysis: A Survey. *Informatica*, vol. 45(7).  
<https://doi.org/10.31449/inf.v45i7.3674>
- [51] A. S. Gaafar, J. M. Dahr, and A. K. Hamoud (2022). Comparative Analysis of Performance of Deep Learning Classification Approach based on LSTM-RNN for Textual and Image Datasets. *Informatica*, vol. 46(5).  
<https://doi.org/10.31449/inf.v46i5.3872>
- [52] H. Ni. (2020). Face recognition based on deep learning under the background of big data. *Informatica*, vol. 44(4).  
<https://doi.org/10.31449/inf.v44i4.3390>