

## 基于变分自编码器与流形特征的聚类算法

陈俊芬, 韩金池, 谢博竣\*, 谢政豪

(河北大学 数学与信息科学学院 河北省机器学习与计算智能重点实验室, 河北 保定 071002)

**摘要:** 深度神经网络因具有优良的非线性映射能力和在不同场景下的灵活性而成为一种很有前景的聚类方法。为了得到易于聚类的特征,许多深度聚类方法从原始数据中提取特征是通过将原始数据映射到一个更低维的空间方式,而聚类指派依然假设是在欧式空间进行。为了探究特征提取和流形空间对聚类性能的影响,本文提出了一种基于变分自编码器与流形特征的聚类算法——MFVC(Clustering Algorithm Based on Variational Autoencoder and Manifold Features)。该方法通过具有残差连接层及无参数注意力机制 SimAM(A Simple, Parameter-Free Attention Module for Convolutional Neural Networks)组成的 $\beta$ -VAE(Learning Basic Visual Concepts with a Constrained Variational Framework)做特征提取器提取图像特征,采用流形 UMAP(Uniform Manifold Approximation and Projection for Dimension Reduction)方法提高特征的可分离性,然后用 K-Means 方法进行聚类学习。在 6 个基准数据集的实验结果表明该方法能够提供较好的性能, MFVC 在 MNIST(Mixed National Institute of Standards and Technology database)数据集上能够达到 0.981 的精度,在 Fashion-MNIST 数据集上能够达到 0.681 的精度。

**关键词:** 变分自编码器; 残差连接; UMAP; K-Means; 流形学习

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 0253-2395(2024)01-0069-12

## Clustering Algorithm Based on Variational Autoencoder and Manifold Features

CHEN Junfen, HAN Jinchi, XIE Bojun\*, XIE Zhenghao

(College of Mathematics and Information Sciences, Hebei Key Laboratory of Machine Learning and Computational Intelligence, Hebei University, Baoding 071002, China)

**Abstract:** Deep neural network has become a promising clustering method due to its excellent nonlinear mapping ability and flexibility in different scenarios. In order to map the original high-dimensional data to a feature space in where the clustering is easy to be done, feature extraction or feature transformation are done by many deep clustering methods, and then the extracted features are grouped into different clusters in the lower-dimensional space, which still are assumed in Euclidean space. In order to explore the impact of feature extraction and manifold space on clustering performance, in this paper, we propose a clustering algorithm based on variational autoencoder and manifold learning—MFVC (Clustering Algorithm Based on Variational Autoencoder and Manifold Features). In this method, the  $\beta$ -VAE (Learning Basic Visual Concepts with a Constrained Variational Framework) with residual connection layer is used as a feature extractor to extract image features, and the non-parameter attention mechanism SimAM (A Simple, Parameter-Free Attention Module for Convolutional Neural Networks) is added to improve the expressive ability of the convolutional network. For more favorable features, the Manifold UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) method is used to improve the separability of the features, and then the K-Means method is used for clustering learning. Experi-

**收稿日期:** 2023-06-06; **接受日期:** 2023-08-15

**基金项目:** 河北省引进留学人员资助项目(C20200302); 河北省教育教学改革研究与实践项目(2020GJJG007)

**作者简介:** 陈俊芬(1976-), 女, 河北阜城人, 博士, 副教授, 研究方向为机器学习、图像聚类。E-mail: chenjunfen2010@126.com

\* **通信作者:** 谢博竣(XIE Bojun), E-mail: xiebojun@126.com

**引文格式:** 陈俊芬, 韩金池, 谢博竣, 等. 基于变分自编码器与流形特征的聚类算法[J]. 山西大学学报(自然科学版), 2024, 47(1): 69-80. DOI: 10.13451/j.sxu.ns.2023139

mental results on six benchmark datasets show that this method can provide better performance. MFVC achieves with accuracy of 0.981 on the MNIST (Mixed National Institute of Standards and Technology database) dataset, and 0.681 on the Fashion-MNIST dataset.

**Key words:** variational autoencoder; residual connection; UMAP; K-Means; manifold learning

### 0 引言

随着网络媒体以及科学技术的发展,收集数据越来越简单数据量也越发大且复杂,并且相应的簇结构往往是没有规则形状的。无论在科研领域,医学领域还是商业领域里,研究者们都希望有相似特征的个体能够聚集在一起,为进一步展开研究做出贡献。近年来,利用深度神经网络对聚类的友好表示得到了广泛的应用。核心思想是通过网络学到的高质量特征能够提高聚类的精度,而聚类的结果又反作用于网络,使得网络学习到更好的特征。

过去的几十年,人们提出了大量的基于浅层模型的聚类方法,包括基于质心的聚类<sup>[1]</sup>,基于密度的聚类<sup>[2]</sup>、层次聚类<sup>[3]</sup>、多视图聚类<sup>[4]</sup>等。但是这些浅层模型在复杂数据上的表现比较差,原因是特征学习能力不够优秀,所以提升对图像特征的理解对于提升聚类性能很有意义。

自动编码器(AutoEncoder, AE)<sup>[5]</sup>是一种无监督的机器学习模型,通过对称的网络学习数据的分布规律,通过瓶颈层诱导得到数据的压缩表示,并重构原始数据(例如正则化数据)学习到数据的内在本质特征,结构如图1(a)所示。在自动编码器中,输入数据 $X$ 被转化为编码向量,随后解码器网络接收编码向量并尝试重构输入。自动编码器可以解释为非线性的降维模型,在数据压缩方面提供了良好的性能,

但是它们不能生成新实例,因为它们不会对潜在空间施加任何结构方面的约束。此外,最小化重构损失而不进行正则化容易导致过拟合。变分自编码器(Variational AutoEncoder, VAE)可以被看作是一种改进潜空间结构的方法,与自动编码器由编码器与解码器两部分构成相似,VAE利用两个神经网络建立两个概率密度分布模型:一个称为推断网络,用来进行原始输入数据的变分推断,生成隐变量的变分概率分布;另一个称为生成网络,根据隐变量变分概率分布还原生成原始数据的近似概率分布。其隐空间提取的是对重建图像更有帮助的特征,结构如图1(b)所示。我们需要做的是让其提取到的特征对于聚类更有益处。

流形是在局部与欧式空间同胚的空间,也就是它在局部也具有欧式空间的性质,可以使用欧式距离进行距离计算。Goodfellow 等于2016年提出了流形假设<sup>[6]</sup>,即该假设认为,现实世界的的数据形态位于高维空间中的低维流形上。流形学习(Manifold Learning)是从高维采样数据中恢复低维流形结构,即找到高维空间中的低维流形,并求出相应的嵌入映射,以实现维数约简或者数据可视化。

强大的推理网络是VAE的一个显著特征,其中数据点被投射到语义有意义的潜在空间。潜在空间内的特征点可看作一单位球面空间上的点,因此考虑流形的球面距离和流形学习,

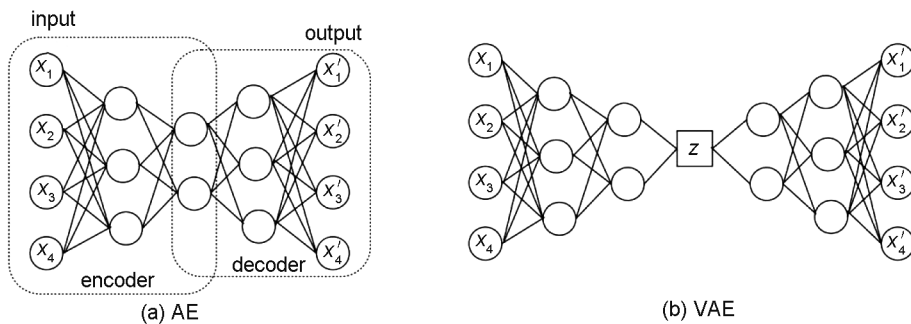


图1 AE和VAE结构示意图  
Fig. 1 Schematic structure diagram of the AE and VAE

替代了传统欧式空间内度量两点之间直线距离。使用流形学习学习更低维数的特征表示,降低数据分布的锐度和冗余关系,让网络专注于本质的特征表示,从而益于下游的聚类任务。针对以上讨论,本文提出了基于变分自编码器与流形特征的聚类算法(Clustering Algorithm Based on Variational Autoencoder and Manifold Features, MFVC),主要内容包括以下4个方面:

I. 提出了一种基于变分自编码器与流形特征的聚类算法——MFVC,在特征提取部分加入残差连接,解决网络退化的问题和防止信息丢失;

II. 进行特征早融合操作,最大可能提取局部特征和全局特征,使特征包含更多信息,有利于提升聚类性能;

III. 在编码器部分加入无参数注意力机制 SimAM (A Simple, Parameter-Free Attention Module for Convolutional Neural Networks),评估每个神经元的重要性,有效地改善了卷积神经网络的表征能力,获得对提升聚类效果有益的特征;

IV. 从聚类角度的理论分析发现流形学习算法 UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction)有明确的拓扑基础,增加了特征的可分离性。

本文组织如下:第1节回顾了相关工作。第2节详细介绍了本文所提方法。第3节描述了实验的数据集和实验设置,并给出具体的实验结果分析。第4节列出了结论。

## 1 相关工作

聚类分析在机器学习和数据挖掘中发挥着不可或缺的作用。学习一个好的数据表示是聚类算法的关键。利用深度神经网络进行聚类任务也有着广泛的应用和重要的研究意义。关于深度聚类方法的研究有很多,但大多数都是从代数角度对特征进行的研究。

### 1.1 深度聚类

深度聚类是指利用深度神经网络进行聚类的过程,其中常使用 Deep AutoEncoder<sup>[7]</sup>方法。该方法的损失函数通常由重构损失和面向聚类的损失组成,通过最小化网络的重构损失来学

习出一个低维嵌入特征空间。DEC (Unsupervised Deep Embedding for Clustering Analysis)<sup>[8]</sup>方法具有联合学习框架和较低的计算复杂度,是深度聚类中的一种代表性方法。DEC 首先通过重构损失,学习从数据空间到低维特征空间的映射,然后迭代优化聚类损失即 KL 散度(软分配-调整后的目标分布)。然而,聚类损失并不能保证局部结构的保护。特征转换可能会被误导,导致嵌入式空间的破坏。因此在 DEC 的基础上,IDECL (Improved Deep Embedded Clustering with Local Structure Preservation)<sup>[9]</sup>通过联合优化加权聚类损失和自动编码器的重构损失来保证微调阶段的局部结构。

在基于概率图模型的深度聚类研究中,更多的是将变分推理和深度自动编码器结合在一起。VAE<sup>[10]</sup>是重要的生成模型之一。

GMVAE (高斯混合变分自动编码器)<sup>[11]</sup>将 VAE 与高斯混合模型 (Gaussian Mixture Model, GMM) 相结合,根据平均场理论用神经网络对变化因素进行参数化,并对证据下界 (Evidence Lower Bound, ELBO) 损失进行优化。VAEIC (Variational Clustering: Leveraging Variational Autoencoders for Image Clustering)<sup>[12]</sup>的先验是确定性的,先验和后验参数是联合学习的,不需要预训练过程。VAEIC 采用了更直接的推理和更有原则的潜在空间先验知识,产生了更简单的推理模型以及更简单的近似后验。

Chen 等<sup>[13]</sup>以变分自编码器做基础,用混合高斯分布作为隐变量的先验分布,特征提取后再进行聚类。其中先验和后验中的高斯分量个数设置为相同,但在现实生活中不是如此,而且处理复杂纹理特征时效果也不理想。Guo 等<sup>[14]</sup>通过优化高斯混合模型先验来优化 VAE。Zhang 等<sup>[15]</sup>提出了一种基于变分自编码器的人脸修复方法,利用了变分自编码器。生成对抗网络实现特征解耦操作,最后获得良好的结果。

### 1.2 与流形学习结合的深度聚类

无论形状或大小,数据之间的位置关系不变谓之数据的拓扑结构,拓扑结构信息为聚类任务提供了很好地估计。图像包含着丰富的局部结构,而局部结构构成整体结构。在图像聚

类任务中考虑局部结构有助于描述图像数据的全局结构。因此,学习特征的拓扑结构是很有必要的,这样可以提升聚类的性能。

一个重要的问题是采用哪种流形方法来学习拓扑结构。有许多方法,比如主成分分析法(PCA<sup>[16]</sup>)、Isomap<sup>[17]</sup>、局部线性嵌入(Local Linear Embedding, LLE)<sup>[18]</sup>等。PCA是一种线性方法,目的是寻求一个低维空间,使得数据的重要结构得以保留,而LLE和Isomap等则倾向于保留数据的几何和邻居属性。

Allaoui等<sup>[19]</sup>只是应用UMAP<sup>[20]</sup>技术,发现应用UMAP可以显著提高聚类算法的性能,但并未深入探究深层次原因。Herrmann等<sup>[21]</sup>提出UMAP与DBSCAN(Density-Based Spatial Clustering of Applications with Noise)算法相结合,探究推断数据集的拓扑结构对聚类性能的影响,表明聚类嵌入向量代表数据流形的结构而不是观察到的特征向量本身是非常有效的,但未重视特征提取环节。

本文提出的基于变分自编码器与流形特征的聚类算法,在 $\beta$ -VAE(Learning Basic Visual Concepts with a Constrained Variational Framework)基础上加入残差结构和无参数注意力SimAM<sup>[22]</sup>构成一个特征提取器,即通过编码器-解码器网络架构来提取特征,之后隐空间的特征经过UMAP算法处理后得到低维特征,最后将低维特征提供给聚类算法K-means<sup>[23]</sup>。我们的方法不仅实现了数据的特征提取,还很大程度上保留了数据的局部拓扑结构,提高了聚类的性能。

## 2 方法

在本节中,我们将描述MFVC的架构,并简要回顾一些我们在工作中使用的常见神经网络架构。

表征学习理论<sup>[24]</sup>表明了特征学习(或表征学习)在机器学习任务中的重要性。我们提出一种新的无监督方法来表征输入图像以提升聚类效果,MFVC的网络架构如图2所示。在 $\beta$ -VAE<sup>[25]</sup>中加入残差连接和无参数注意力机制SimAM来构建一个特征提取器,编码器的每层输出后接池化以获得更好的特征表示。首先,

将大小统一成 $32 \times 32$ 的图像,输入进特征提取器中,并经过指定轮次的训练,这一步的输出是每个输入在隐空间中的高维特征向量 $Z = \{z_1, z_2, \dots, z_n\}$ ;其次,使用流形学习方法处理高维特征向量 $Z$ ,这一步的输出是高维特征向量映射的低维嵌入 $U = \{u_1, u_2, \dots, u_n\}$ ;最后,可使用经典的K-means<sup>[23]</sup>聚类算法或DBSCAN算法对低维嵌入 $U$ 进行聚类分析,输出是每个输入的所属类簇,完成聚类指派。我们将MFVC简单地表示为 $C = M_C(M_M(M_V(X)))$ ,其中 $C$ 为最终聚类, $M_C$ 为聚类算法, $M_M$ 为流形表示学习, $M_V$ 为特征提取器部分, $X$ 为原始数据。

### 2.1 与流形学习结合的深度聚类

本文以 $\beta$ -VAE为基础构建特征提取器, $\beta$ -VAE是一种变体VAE。 $\beta$ -VAE的损失函数如式(1)所示,即在VAE损失的第二项KL散度上添加了超参数 $\beta$ ,约束了潜在信息通道 $z$ 的容量。

$$L(x, x') = \|x - x'\|^2 + \beta D_{KL}(G(\mu(x), \sigma(x)), N(0, 1)). \quad (1)$$

如图3所示,我们使用两组卷积层和归一化层堆叠后添加无参数注意力机制SimAM块组成直接映射部分,使用一组卷积层和归一化层堆叠组成残差部分,直接映射部分和残差部分组成Res-VAE-Block模块。添加残差连接后,信息呈现逐层递增趋势,不用再担心信息的丢失。并且残差连接可以防止过拟合,因为残差连接降低了模型的复杂度,还可以解决网络退化问题,使模型更好地收敛。

通道注意力和空间注意力模块产生一维或二维权值,并平均处理每个通道或空间位置的神经元,这可能会限制它们学习更多辨别线索的能力。依据Yang等<sup>[22]</sup>SimAM的核心思想描述如下,无参数注意力机制SimAM产生三维权值,模块结构如图4所示。

SimAM不在原始网络中添加参数,而是通过优化一个能量函数来发现每个神经元的重要度,在一层中推断特征图的三维注意权重。为了寻找重要神经元即度量神经元之间的线性可分性,最小能量可以计算如下:

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda}, \quad (2)$$

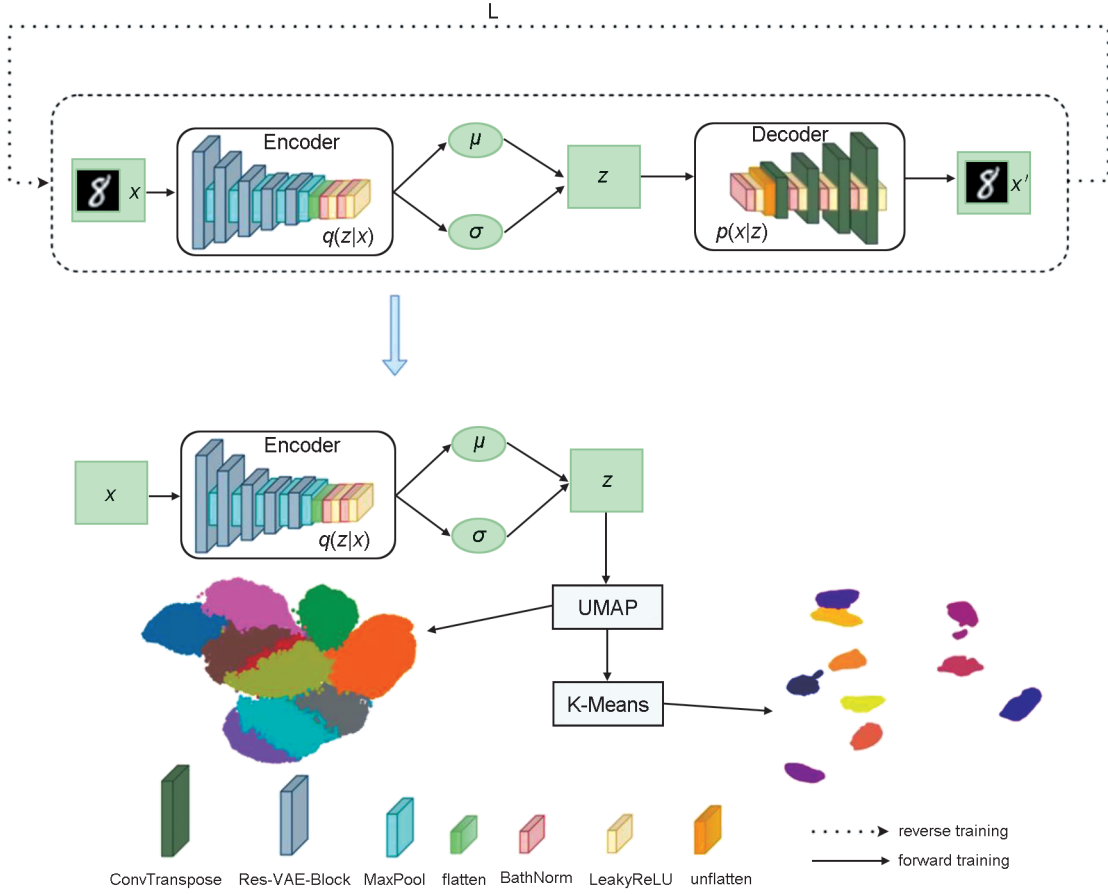


图2 MFVC网络结构示意图,由特征提取、流形学习以及聚类方法三部分组成

Fig. 2 Illustration of MFVC network structure, MFVC consists of three parts: feature extraction, manifold learning and clustering methods

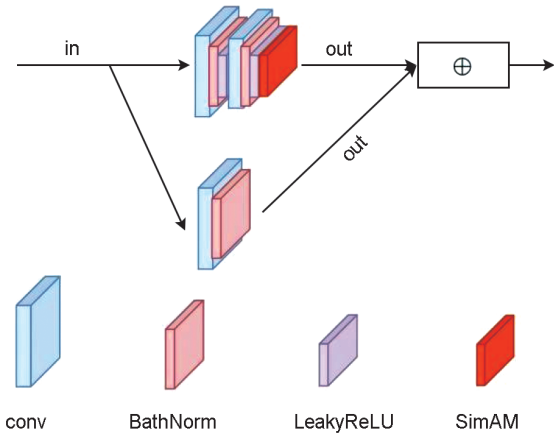


图3 Res-VAE-Block结构示意图

(此模块由卷积层、归一化层和无参数注意力机制SimAM块组成)

Fig. 3 Structure illustration of the SimAM

(This module consists of convolutional layers, normalization layers and a parameterless attention mechanism SimAM module)

其中  $\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i$ ,  $\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$ ,  $t$  和  $x_i$  是输入特征  $X \in R^{C \times H \times W}$  的单个通道中的目标神经元和其他神经元,  $i$  是空间维度上的索

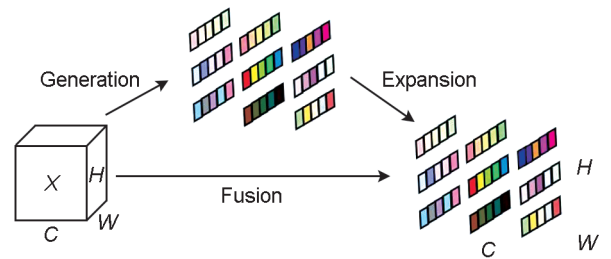


图4 SimAM结构示意图

Fig. 4 Structure illustration of the Res-VAE-Block

引,每个通道有  $M$  个能量函数,公式(2)表明能量越低,神经元  $t$  与周围神经元更不同,对视觉处理更重要。因此,每个神经元的重要性可以通过  $\frac{1}{e_i^*}$  得到。

我们在编码器阶段的每层 Res-VAE-Block 模块之后都添加了池化层,将特征图做全局平均池化(GAP),然后使用连接(concat)进行特征早融合。即将特征进行连接融合再训练。全局平均池化由 Lin 等<sup>[26]</sup>提出,GAP与全连接相比更加自然地加强了类别和特征图之间的联

系。GAP 不同于将特征图直接拉成一维向量的方法,它将每个通道的二维图像做平均,也就是每个通道对应一个均值。相比于全连接将特征图拉平,GAP 参数量显然更少,同时能够抑制有效过拟合。并且特征图经过 GAP 之后,神经网络的参数与输入图像尺寸的大小不再有关,即输入图像的长宽可以不固定。

## 2.2 流形映射

本节介绍 UMAP 算法如何对高维特征向量  $z$  得到低维嵌入  $U = \{u_1, u_2, \dots, u_n\}$ , 其中  $u_i = \text{UMAP}(z_i), u_i \in R^{du}, du < dz$ 。UMAP 基于以下三个假设:1)数据均匀分布在黎曼流形上,2)黎曼度量是局部常数,3)流形是局部连通的。依据 McInnes 等<sup>[20]</sup>,UMAP 的核心思想描述如下。

首先用指数概率分布计算高维数据点之间的相似度:

$$p_{ij} = \exp\left(-\frac{d(z_i, z_j) - \rho_i}{\sigma_i}\right), \quad (3)$$

其中,  $\rho_i$  为高维空间中观测点  $z_i$  到其最近邻  $z_j$  的距离,  $\sigma_i$  是为每个  $z_i$  设置的(光滑)归一化因子,  $\sum_{j=1}^k \exp\left(\frac{-\max(0, d(z_i, z_j) - \rho_i)}{\sigma_i}\right) = \log_2(k)$ , 这定义了  $z_i$  点的局部(黎曼)度量,  $d(\cdot, \cdot)$  表示欧氏距离。

$i$  和  $j$  结点之间的图的权重是不等于  $j$  和  $i$  结点之间的权重,使用高维概率的对称化来解决此问题:

$$p_{ij} = p_{ij} + p_{ji} - p_{ij}p_{ji} \quad (4)$$

接着使用概率度量  $q_{ij}$ , 构建并优化低维模拟图的布局,使其与高维图尽可能相似:

$$q_{ij} = (1 + a(u_i - u_j)^b)^{-1}, \quad (5)$$

其中默认 UMAP 中  $a \approx 1.93$  和  $b \approx 0.79$ 。

最后采用交叉熵  $L$  (公式 6) 作为代价函数,利用交叉熵的导数对低维数据点进行微调更新,优化投影空间直至收敛。

$$L(P, Q) = \sum_i \sum_j [p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right) + (1 - p_{ij}) \log\left(\frac{1 - p_{ij}}{1 - q_{ij}}\right)] \quad (6)$$

在 UMAP 算法中有许多影响性能的重要超参数。第一个是邻居的数量,这表示保留多少局部结构和捕获多少全局结构的粒度之间的权衡。二是目标嵌入的维数。还有嵌入空间中点

之间允许的最小间距。最小间距越小,越能准确地捕捉到真正的流形结构,但投影后相似的点越聚集,使可视化变得困难。

我们方法的伪代码在算法 1 中描述。

### 算法 1 MFVC 方法

```

输入:数据集 X, UMAP 算法参数近邻点数量 n_neighbors, 目标嵌入的维度 T, 嵌入空间中点之间允许的最小距离 d, K-Means 算法参数 n_clusters 和 random state;
输出:聚类结果 C = {c_1, c_2, ..., c_k}。
① X -> β-VAE; /*即 X -> encoder -> feature -> decoder, 训练指定轮次*/
② UMAP(feature, n_neighbors, d, T)
③ 计算高维空间点之间的距离, 计算每个样本点的参数 ρ_i 和 σ_i, 构造稀疏矩阵返回联合概率
④ 低维数据初始化, 利用交叉熵的导数对低维数据点进行协调更新, 返回降维结果 new_features
⑤ end UMAP
⑥ K-Means(n_clusters, random state) /*对经过 UMAP 算法降维后的特征进行聚类*/
⑦ 得到聚类结果 C

```

对算法 1 的时间复杂度分析如下: 设输入大小为  $H \times W$ , 通道数为  $C$ , Res-VAE-Block 深度为  $D$ 。Res-VAE-Block 时间复杂度为  $O(DCH^2W^2)$ 。卷积层的理论时间复杂度为  $O(P_i^2 \cdot R_i^2 \cdot C_i^{\text{in}} \cdot C_i^{\text{out}})$ , 其中  $P_i$  是每个卷积核输出特征图的边长,  $R_i$  是每个卷积核的边长,  $C_i^{\text{in}}$  即输入通道数, 也就是上一层的输出通道数,  $C_i^{\text{out}}$  本卷积层具有的卷积核个数, 也即输出通道数。UMAP 算法的经验复杂度为  $O(N^{1.14})$ <sup>[20]</sup>, K-Means 算法的时间复杂度为  $O(knm)$ , 其中  $k$  为簇数目,  $n$  为样本点数,  $m$  为样本点维度。

所以算法 1 时间复杂度为  $O(T \sum_{i=1}^Q P_i^2 \cdot R_i^2 \cdot C_i^{\text{in}} \cdot C_i^{\text{out}}) + O(N^{1.14}) + O(knm)$ , 其中  $Q$  为网络的深度,  $T$  为训练迭代次数。

## 3 实验

本节描述实验中所用到的数据集, 评价指标, 介绍实验设置并讨论实验结果。

### 3.1 数据集

(1) MNIST (Mixed National Institute of Standards and Technology database) 是一个手写数字数据集, 包含 10 个类别共 70 000 张图像, 图像尺寸为  $28 \times 28$ 。

(2) Fashion-MNIST 数据集: 该数据集包含

10类共70 000个正面图片,图像尺寸为 $28 \times 28$ 。

(3) USPS (United States Postal Service) 手写数字数据集,包括9 298张10类手写数字,图像尺寸为 $16 \times 16$ 。

(4) Emnist-digits 数据集:EMNIST (Extended MNIST) 在手写字体分类任务中更有挑战,共10类,每一类包含相同数量数据,每一类训练集24 000张,测试集4 000张,图像尺寸为 $28 \times 28$ 。

(5) Emnist-balanced,共47类,每一类训练集2 400张,测试集400张,图像尺寸为 $28 \times 28$ 。

### 3.2 评价指标

我们将使用以下三个标准评估指标来验证无监督聚类算法的性能。这三个评价指标的取值范围都在 $[0, 1]$ ,值越高,则表明聚类性能越好。

$A_{ACC}$  (Accuracy):是在所有预测中预测正确的比例。

$$A_{ACC} = \max_m \frac{\sum_{i=1}^n 1\{t_i = m(c_i)\}}{n}, \quad (7)$$

其中 $t$ 是真实标签, $c$ 是集群标签, $m$ 是集群与标签之间的映射关系。

$N_{NMI}$  (Normalized Mutual Information): NMI是一种聚类有效性外部评价指标,并且是从信息论的角度评估两个聚类结果的相似性。

$$N_{NMI} = \frac{2I(t, c)}{[H(t) + H(c)]}, \quad (8)$$

其中 $t$ 是真实标签, $c$ 是集群标签, $H$ 度量熵, $I(t, c)$ 表示真实标签与集群标签之间的互信息, $H(t)$ 和 $H(c)$ 表示 $t$ 和 $c$ 的熵。

**Purity**:把每个簇中最多的类作为这个簇所代表的类,然后计算正确分配的类的数量,然后除以总数。

$$P(\Omega, C) = \frac{1}{N} \sum_k \max_j |w_k \cap c_j|, \quad (9)$$

其中 $N$ 表示总的样本数, $\Omega = \{w_1, w_2, \dots, w_k\}$ 是聚类后簇的集合,有 $k$ 个簇, $C = \{c_1, c_2, \dots, c_j\}$ 是正确类别的集合,有 $j$ 个类别。类 $k$ 与出现次数最多的集群 $j$ 相关联。

### 3.3 实验设置

整个实验都是在Pytorch中进行,特征提取部分选择 $\beta$ -VAE网络作为基础,每层Res-

VAE-Block模块都有池化层,激活函数使用LeakyRelu<sup>[27]</sup>函数,通过最小化损失函数(1)式训练 $\beta$ -VAE,训练轮次为200轮。优化器是AdamW,学习率设为0.005并采用余弦退火调整学习率,批处理尺寸为64。我们在流形表示环节即提取特征后使用UMAP算法,其中邻居数设置为 $n\_neighbors = 27$ ,维数为数据集的类别数,流形中每个点之间的最小距离 $d = 0.00$ 。最后的聚类算法采用K-Means算法。编码器具体配置如表1所示,特征提取器的编码器部分由残差块部分堆叠得到,前四层由Res-VAE-Block模块和池化两部分堆叠得到,第五层由Res-VAE-Block模块、池化、flatten操作、Batch-Norm操作以及激活函数LeakyReLU组成。

表1 编码器的网络配置

Table 1 Configuration of encoder network

Layers	Configurations
layer1	Res_Block_VAE(in=1,out=16),Pool
layer2	Res_Block_VAE(in=16,out=32),Pool
layer3	Res_Block_VAE(in=32,out=64),Pool
layer4	Res_Block_VAE(in=64,out=128),Pool
layer5	Res_Block_VAE(in=128,out=256),Pool

### 3.4 实验结果与分析

在本小节中,首先介绍MFVC方法在MNIST、Fashion-MNIST、USPS、Emnist-digits等六个数据集上的聚类性能,在3.4.1小节中我们选择与一些非深度聚类方法和深度聚类方法进行对比,在下面的3.4.2小节中,我们主要探究MFVC各组件的作用。



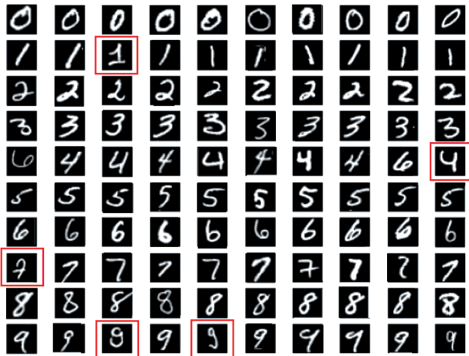
图5 MNIST数据集上的MFVC聚类结果可视化

Fig. 5 Visualization of the MFVC clustering results on the MNIST dataset

图5所示为在MNIST数据集上经过我们的方法所产生的类簇可视化结果。可以看到在我们的结果中,能够较为清晰地分辨出各类簇。

图6随机显示了MNIST数据集中每类手写数字各10个的可视化图像,其中红框标注的数字在特征识别能力不太强的情况下很容易出现混淆的情况,例如红框标注的数字7与最后一行的数字9就很相似,难以辨别。

我们的方法将MNIST、Fashion-MNIST、USPS、Emnist-digits等六个数据集上的聚类性能



注:红框标注的数字易与其他数字混淆。

图6 MNIST数据集的原始数据展示

Fig. 6 Several examples from the MNIST dataset

列于表2,包括 $A_{ACC}$ 、 $N_{NMI}$ 和 $P$ 的平均值和标准偏差(调整参数至最佳后的5次微调实验结果),在MNIST-Full数据集上准确度可达0.981,在Fashion-MNIST数据集上准确度可达0.681。

### 3.4.1 对比实验

对比实验所选择的聚类算法包括非深度聚类方法,如GMM<sup>[28]</sup>。也包括一些深度聚类的方法,如VaDE (Variational Deep Embedding)<sup>[29]</sup>、ClusterGAN<sup>[30]</sup>、DEC<sup>[8]</sup>、IDEC<sup>[9]</sup>、DSC<sup>[31]</sup> (Deep Spectral Clustering)、DeepDPM<sup>[32]</sup>和DC-VAE<sup>[33]</sup>。VaDE与DC-VAE都是以变分自编码器为基本框架,ClusterGAN是训练一个GAN和一个具有特定簇损失的逆映射网络来实现分簇,GAN与VAE都是生成式模型。DEC与IDEC使用卷积自编码器进行联合训练,虽然MFVC使用分步训练,性能并未逊色于这些经典聚类方法。DSC设计了一个双自动编码器将数据嵌入到潜在空间中,并且应用深度谱聚类来获得标签分配。还与基于流形学习方法UMAP和聚类算法结合的简单聚类算法(AI-laoui等<sup>[19]</sup>)做对比,虽然它与MFVC一样利用了UMAP与聚类算法,但并未重视特征提取部分。

表2 我们方法在6个数据集上的性能

Table 2 The performances of our method on the six datasets

	MNIST-Full	MNIST-test	Fashion-MNIST	USPS	Emnist-digits	Emnist-balanced
$A_{ACC}$	0.981±0.000 1	0.969±0.002	0.681±0.01	0.961±0.002	0.873±0.02	0.713±0.001
$N_{NMI}$	0.951±0.002	0.926±0.003	0.655±0.002	0.903±0.001	0.907±0.001	0.793±0.002
$P$	0.981±0.001	0.963±0.002	0.681±0.01	0.951±0.002	0.875±0.02	0.718±0.001

表3 我们方法在6个数据集上的性能

Table 3 Comparison of our method and other methods in  $A_{ACC}$  and  $N_{NMI}$  performance of each dataset

	Reference	MNIST		Fashion-MNIST		USPS	
		$A_{ACC}$	$N_{NMI}$	$A_{ACC}$	$N_{NMI}$	$A_{ACC}$	$N_{NMI}$
GMM <sup>[28]</sup>	Gaussian Mixture Models 2009	0.389	0.333	0.463	0.514	0.562	0.540
DEC	PMLR2016	0.863	0.834	0.518	0.546	0.762	0.767
IDEC	IJCAI2017	0.881	0.867	0.529	0.546	0.761	0.785
VaDE <sup>[29]</sup>	IJCAI2017	0.945	0.876	0.578	0.630	0.566	0.512
UMAP+GMM <sup>[19]</sup>	ICISP2020	0.916	0.866	0.589	0.651	0.811	0.900
ClusterGAN <sup>[30]</sup>	AAAI-19	0.890	0.900	0.500	0.640		
DSC <sup>[31]</sup>	IEEE/CVF 2019	0.978	0.941	0.662	0.645	0.869	0.857
DeepDPM <sup>[32]</sup>	IEEE/CVF 2022	0.980	0.940	0.620	0.680	0.890	0.880
DC-VAE <sup>[33]</sup>	Knowledge-Based Systems 2020	0.973	0.941	0.597	0.633	0.778	0.697
N2D <sup>[34]</sup>	ICPR 2020	0.979	0.942	0.672	0.684	0.958	0.901
MFVC		0.981	0.951	0.681	0.655	0.961	0.903

表3展示了MFVC方法与其他方法在MNIST、Fashion-MNIST、USPS三个数据集上 $A_{ACC}$ 与 $N_{NMI}$ 表现的比较情况,通过比对数据我们发现,在给定数据集上,MFVC能够达到更优的性能。

图7是经过UMAP简单调参降维后可视化特征图得到的结果,可以看出,有的类与类之间已经有了界限,但是有的类与类之间界限还是不甚明朗,数字3、5、8类簇之间的界限尤其模糊。

#### 3.4.2 消融实验

消融实验通过使用CBAM(Convolutional Block Attention Module)<sup>[35]</sup>代替SimAM,使用Isomap代替流形学习UMAP方法,使用普通 $\beta$

-VAE代替具有残差连接和无参数注意力机制的 $\beta$ -VAE以及使用DBSCAN<sup>[36]</sup>聚类算法代替K-Means聚类方法来探究各组件的作用。

表4显示的是消融方法与我们的方法在MNIST、Fashion-MNIST、USPS这3个数据集上的 $A_{ACC}$ 、 $N_{NMI}$ 的平均值和标准偏差(调整参数至最佳后的5次微调实验结果)。

我们在6个原始数据集中的三个数据集MNIST、Fashion-MNIST、USPS上做了实验,消融实验结果显示在这3个数据集上我们的方法MFVC都是更好的方法,具有UMAP的MFVC在所有数据集上都优于具有Isomap的方法,并且具有VAE特征提取器部分的性能也会相对



图7 经过UMAP处理后的MNIST数据集特征图

Fig. 7 MNIST data feature map after UMAP processing

表4 MFVC方法在3个数据集上的消融实验结果

Table 4 Ablation experiment results of MFVC method on 3 datasets

	MNIST		Fashion-MNIST		USPS	
	$A_{ACC}$	$N_{NMI}$	$A_{ACC}$	$N_{NMI}$	$A_{ACC}$	$N_{NMI}$
K-Means	0.532 0±0.000 0	0.500 0±0.000 0	0.474 0±0.000 0	0.512 0±0.000 0	0.668 0±0.000 0	0.626 0±0.000 0
MFVC(Isomap)	0.121 0±0.002 0	0.009 0±0.003 0	0.100 0±0.001 0	0.001 0±0.001 0	0.171 0±0.001 0	0.104 0±0.001 0
MFVC(DBSCAN)	0.881 0±0.001 0	0.909 0±0.001 0	0.644 0±0.001 0	0.683 0±0.001 0	0.863 0±0.010 0	0.833 0±0.002 0
MFVC(without SimAM)	0.857 0±0.010 0	0.831 0±0.005 0	0.584 0±0.002 0	0.607 0±0.030 0	0.536 0±0.020 0	0.668 0±0.010 0
$\beta$ -VAE+K-Means	0.782 0±0.001 0	0.719 0±0.001 0	0.491 0±0.002 0	0.524 0±0.010 0	0.683 0±0.001 0	0.601 0±0.002 0
UMAP+K-Means	0.869 0±0.002 0	0.850 0±0.001 0	0.410 0±0.003 0	0.590 0±0.001 0	0.517 0±0.002 0	0.619 0±0.004 0
MFVC(without Res lock)	0.882 0±0.003 0	0.878 0±0.003 0	0.604 0±0.000 1	0.673 0±0.000 2	0.468 0±0.040 0	0.657 0±0.001 0
MFVC(with CBAM)	0.977 0±0.000 2	0.941 0±0.000 1	0.662 0±0.001 0	0.657 0±0.003 0	0.953 0±0.000 2	0.889 0±0.000 4
MFVC	0.981 0±0.000 1	0.951 0±0.002 0	0.681 0±0.010 0	0.655 0±0.002 0	0.961 0±0.002 0	0.903 0±0.001 0

较好。

表4显示了我们各组件作用的对比,与 $\beta$ -VAE+K-Means相比,MFVC最大的收益体现在USPS上,性能提升了27个百分点。而在MNIST和Fashion-MNIST上应用我们的方法时,性能大约提升了20个百分点和19个百分点。与UMAP+K-Means相比,我们的方法最大的收益体现在USPS上,性能提升了44个百分点。而在MNIST和Fashion-MNIST上应用我们的方法时,性能大约提升了11个百分点和27个百分点。这说明提取好的特征对于聚类是非常重要的。

图8显示了在控制其他参数一定的条件下,UMAP算法参数 $n\_neighbors$ (近邻点)和 $min\_dist$ (低维下允许的行间最小距离)对准确

率 $A_{ACC}$ 的影响。图8(a)中显示在调整 $n\_neighbors$ 的过程中,我们发现结果都稳定在0.95左右。而图8(b)中显示, $min\_dist$ 在0~0.1范围内可以保持一个性能较好的结果,之后随着 $min\_dist$ 的增大,性能在逐渐下降,当 $min\_dist = 0.8$ 时的性能与 $min\_dist = 0$ 时的性能低26个百分点。就 $n\_neighbors$ 和 $min\_dist$ 这两个参数来说,参数 $eps$ 对于结果的影响较大。

在使用DBSCAN聚类算法代替K-Means聚类方法的消融实验中,如图9所示,经过在MNIST数据集和Fashion-MNIST数据集上的实验,结果显示通过UMAP算法对提取到的特征进行处理后,对于DBSCAN临近点阈值 $min\_points$ 有一个较高的容忍度,并且同时能够得到一个较好的性能效果。

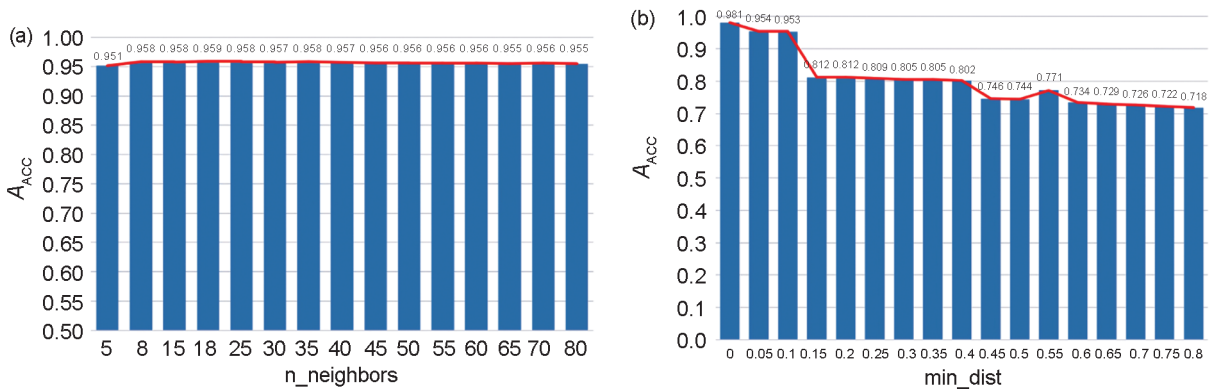


图8 UMAP的参数对准确率 $A_{ACC}$ 的影响

(a)参数 $n\_neighbors$ 准确度的影响;(b)参数 $min\_dist$ 对准确度的影响

Fig. 8 The influence of UMAP parameters on the accuracy rate  $A_{ACC}$

(a) Effect of parameter  $n\_neighbors$  on accuracy; (b) Effect of parameter  $min\_dist$  on accuracy

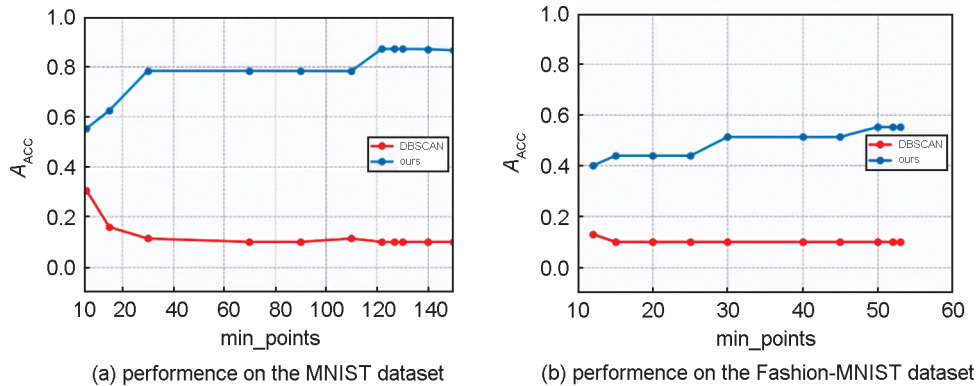


图9 在MNIST和Fashion-MNIST数据集上DBSCAN和有DBSCAN的MFVC的 $A_{ACC}$ 比较

(a)在MNIST数据集上的表现;(b)在Fashion-MNIST数据集上的表现

Fig. 9 Comparison of DBSCAN and MFVC with DBSCAN  $A_{ACC}$  on the MNIST and Fashion-MNIST dataset

(a) performance on the MNIST dataset; (b) performance on the Fashion-MNIST dataset

## 4 结论

本文提出了一种基于变分自编码器与流形特征的聚类算法,由特征提取,流形嵌入和聚类步骤组成。使用具有残差连接的 $\beta$ -VAE进行特征提取,可以保留了大部分信息;使用UMAP流形算法,其表示针对可分离性进行优化,提高了特征的可分离性,模糊拓扑表示放大了集群结构,提高了聚类性能;使用K-Means作为最后步骤的聚类算法,容易实现且结果好,我们的结果表明,结合 $\beta$ -VAE、UMAP和K-Means的方法甚至可以在Fashion-MNIST等复杂图像数据上优于基于深度学习的ClusterGAN方法。但是K-Means方法本身也有一定的缺点,比如初始点选取不好确定,受离群值影响较大等,并且经过调参后的DBSCAN算法性能也可能达到K-Means算法的性能,以及基于局部拓扑嵌入的深度聚类具有一定的研究价值,这都会是我们新的研究方向。

## 参考文献:

- [1] REN Y Z, KAMATH U, DOMENICONI C, *et al.* Parallel Boosted Clustering[J]. *Neurocomputing*, 2019, **351**: 87-100. DOI: 10.1016/j.neucom.2019.04.003.
- [2] REN Y Z, DOMENICONI C, ZHANG G J, *et al.* A Weighted Adaptive Mean Shift Clustering Algorithm[C]// Proceedings of the 2014 SIAM International Conference on Data Mining. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2014: 10.1137/1.9781611973440.91. DOI: 10.1137/1.9781611973440.91.
- [3] JAIN A K, MURTY M N, FLYNN P J. Data Clustering: a Review[J]. *ACM Comput Surv*, 1999, **31**(3): 264-323. DOI: 10.1145/331499.331504.
- [4] HUANG Z M, REN Y Z, PU X R, *et al.* Dual Self-paced Multi-view Clustering[J]. *Neural Netw*, 2021, **140**: 184-192. DOI: 10.1016/j.neunet.2021.02.022.
- [5] CUN Y L. Modeles connexionnistes de L'apprentissage [D]. Paris: Universite Paris, 1987, **2-3**: 114-143. DOI: 10.3406/intel.1987.1804.
- [6] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. Cambridge, MA, USA. MIT Press, 2016.
- [7] HINTON G E, SALAKHUTDINOV R R. Reducing the Dimensionality of Data with Neural Networks[J]. *Science*, 2006, **313**(5786): 504-507. DOI: 10.1126/science.1127647.
- [8] XIE J Y, GIRSHICK R, FARHADI A. Unsupervised Deep Embedding for Clustering Analysis[C]// Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48. New York: ACM, 2016: 478-487. DOI: 10.5555/3045390.3045442.
- [9] GUO X F, GAO L, LIU X W, *et al.* Improved Deep Embedded Clustering with Local Structure Preservation[C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. New York: ACM, 2017: 1753-1759. DOI: 10.5555/3172077.3172131.
- [10] KINGMA D P, WELLING M. Auto-encoding Variational Bayes[EB/OL]. arXiv Preprint: 1312.6114, 2013.
- [11] DILOKTHANAKUL N, MEDIANO P, GARNELO M, *et al.* Deep Unsupervised Clustering with Gaussian Mixture Variational Autoencoders[J]. arXiv Preprint: 1611.02648, 2016.
- [12] PRASAD V, DAS D, BHOWMICK B. Variational Clustering: Leveraging Variational Autoencoders for Image Clustering[C]// 2020 International Joint Conference on Neural Networks (IJCNN). 2020: 1-10. DOI: 10.1109/IJCNN48605.2020.9207523.
- [13] 陈华华, 陈哲, 郭春生, 等. 混合高斯变分自编码器的聚类网络[J]. 中国图象图形学报, 2022, **27**(7): 2148-2156. DOI: 10.11834/jig.200467.
- [14] CHEN H H, CHEN Z, GUO C S, *et al.* A Gaussian Mixture Variational Autoencoder Based Clustering Network [J]. *J Image Graph*, 2022, **27**(7): 2148-2156. DOI: 10.3772/j.issn.1000-0135.2018.09.006. DOI: 10.11834/jig.200467.
- [15] GUO C S, ZHOU J L, CHEN H H, *et al.* Variational Autoencoder with Optimizing Gaussian Mixture Model Priors[J]. *IEEE Access*, 2020, **8**: 43992-44005. DOI: 10.1109/ACCESS.2020.2977671.
- [16] 张雪菲, 程乐超, 白升利, 等. 基于变分自编码器的人脸图像修复[J]. 计算机辅助设计与图形学学报, 2020, **32**(3): 401-409. DOI: 10.3724/SP.J.1089.2020.17938.
- [17] ZHANG X F, CHENG L C, BAI S L, *et al.* Face Image Inpainting Variational Autoencoder[J]. *J Comput Aided Des Comput Graph*, 2020, **32**(3): 401-409. DOI: 10.3724/SP.J.1089.2020.17938.
- [18] ABDI H, WILLIAMS L J. Principal Component Analysis[J]. *Wiley Interdiscip Rev Comput Stat*, 2010, **2**(4): 433-459. DOI: 10.1002/wics.101.
- [19] TENENBAUM J B, DE SILVA V, LANGFORD J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. *Science*, 2000, **290**(5500): 2319-2323. DOI: 10.1126/science.290.5500.2319.
- [20] ROWEIS S T, SAUL L K. Nonlinear Dimensionality

- Reduction by Locally Linear Embedding[J]. *Science*, 2000, **290**(5500): 2323–2326. DOI: 10.1126/science.290.5500.2323.
- [19] ALLAOUI M, LAMINE KHERFI M, CHERIET A. Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 317–325. DOI: 10.1007/978-3-030-51935-3\_34.
- [20] MCINNES L, HEALY J, SAUL N, *et al.* UMAP: Uniform Manifold Approximation and Projection[J]. *J Open Source Softw*, 2018, **3**(29): 861. DOI: 10.21105/joss.00861.
- [21] HERRMANN M, KAZEMPOUR D, SCHEIPL F, *et al.* Enhancing Cluster Analysis via Topological Manifold Learning[J]. arXiv Preprint:2207.00510, 2022.
- [22] YANG L, ZHANG R Y, LI L, *et al.* Simam: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks[C]//Proceedings of the International conference on machine learning. PMLR, 2021, **139**: 11863–11874.
- [23] MACQUEEN J. Some Methods for Classification and Analysis of Multivariate Observations[C]//Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1967. **1**(14): 281–297.
- [24] BENGIO Y, COURVILLE A, VINCENT P. Representation Learning: A Review and New Perspectives[J]. *IEEE Trans Pattern Anal Mach Intell*, 2013, **35**(8): 1798–1828. DOI: 10.1109/TPAMI.2013.50.
- [25] HIGGINS I, MATTHEY L, PAL A, *et al.* BETA-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework[C]//Proceedings of the International conference on learning representations, 2017.
- [26] LIN M, CHEN Q, YAN S C. Network in network[J]. arXiv Preprint :1312.4400, 2013.
- [27] MAAS A L, HANNUN A Y, NG A Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models[C]//Proceedings of the International Conference on Machine Learning. ICML, 2013, **30**(1): 3.
- [28] REYNOLDS D. Gaussian Mixture Models[M]//Encyclopedia of Biometrics. Boston, MA: Springer US, 2009: 659–663. DOI: 10.1007/978-0-387-73003-5\_196.
- [29] JIANG Z X, ZHENG Y, TAN H C, *et al.* Variational Deep Embedding: An Unsupervised and Generative Approach to Clustering[J]. arXiv Preprint:1611.05148, 2016.
- [30] MUKHERJEE S, ASNANI H, LIN E, *et al.* ClusterGAN: Latent Space Clustering in Generative Adversarial Networks[J]. *Proc AAAI Conf Artif Intell*, 2019, **33**(1): 4610–4617. DOI: 10.1609/aaai.v33i01.33014610.
- [31] YANG X, DENG C, ZHENG F, *et al.* Deep Spectral Clustering Using Dual Autoencoder Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 4066–4075. DOI: 10.1109/CVPR.2019.00419.
- [32] RONEN M, FINDER S E, FREIFELD O. DeepDPM: Deep Clustering with an Unknown Number of Clusters [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 9851–9860. DOI: 10.1109/CVPR52688.2022.00963.
- [33] XU C Y, DAI Y F, LIN R J, *et al.* Deep Clustering by Maximizing Mutual Information in Variational Autoencoder[J]. *Knowl Based Syst*, 2020, **205**: 106260. DOI: 10.1016/j.knosys.2020.106260.
- [34] MCCONVILLE R, SANTOS-RODRÍGUEZ R, PIECHOCKI R J, *et al.* N2D: (not too) Deep Clustering via Clustering the Local Manifold of an Autoencoded Embedding[C]//2020 25th International Conference on Pattern Recognition (ICPR). 2021: 5145–5152. DOI: 10.1109/ICPR48806.2021.9413131.
- [35] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional Block Attention Module[M]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 3–19. DOI: 10.1007/978-3-030-01234-2\_1.
- [36] ESTER M, KRIEGEL H P, SANDER J, *et al.* A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise[C]//Knowledge Discovery and Data Mining, 1996, **96**(34): 226–231. DOI: 10.5555/3001460.3001507.