

## 摘要

图片聚类是计算机视觉领域的基本问题之一，即将相似的图片归为一簇，不同的图片归为不同簇，由于图像内容各异，使其成为了一项非常有难度的任务。近年来，随着深度学习的发展，深度神经网络在图像聚类研究中大放异彩，通过将大量的无标签数据输入网络进行训练，使得聚类算法效果更上了一个台阶。

目前深度聚类算法可分为三类：自编码器类、子空间类、图卷积神经网络类。在本研究中，我们结合了自编码器和图神经网络两种技术，搭建了一深度聚类网络，研究结果如下：

在 DEC(Deep Embedded Clustering)[2]算法的研究基础上，将普通自编码器改为变分自编码器，提取到了更具代表性的特征向量，并引入高斯混合模型，使提取出的特征符合多个高斯混合分布，提高了算法的准确率；在算法中引入图神经网络，提取样本整体的结构性信息，并在变分自编码器与图神经网络之间设计了信息传递环节，使图神经网络能够融合二者的信息，输出最终聚类结果，进一步提高了算法的准确率，所提出模型名为 VGDCN(Variational Graph convolution Deep Clustering Network)。

最后，为了验证所提出算法的普适性，自制了植物叶片数据集，包含来自十种不同植物的叶片，将其输入到 VGDCN 中进行聚类，得到了可信的实验结果。

**关键词：**图像聚类，深度神经网络，自编码器，图神经网络

## Abstract

Image clustering is one of the basic problems in the field of computer vision. The computer can classify the images by their similarities and differences with clustering algorithms. But it is a very difficult task because of the diversity of image content. In recent years, deep neural network has been widely used in image clustering research with the development of deep learning. The clustering algorithms have been improved to a higher lever by inputting a large number of unlabeled images into the neural network for training.

At present, deep clustering algorithms can be divided into three categories: Autoencoders based methods, subspace clustering and graph convolution neural network based methods. In this study, we build a deep clustering network by combining autoencoder and graph neural network. The details are as follows:

Based on the research of DEC (Deep Embedded Clustering) [2], we replaced the ordinary autoencoder by a variational autoencoder and got features which are more representative. The accuracy of the algorithm has been improved by making the features lie on a Gaussian mixture distribution; Then, we introduced GCN(Graph Convolutional Network) into the algorithm for the structural information of samples. We designed a delivery operator between the variational autoencoder and GCN to combine two kinds of information. The proposed model was named VGDCN (Variational Graph convolution Deep Clustering Network).

Finally, we applied VGDCN on our self-made dataset to verify the generality of the algorithm. The self-made dataset contains ten different kinds of plant leaves. We got reliable experimental results on the dataset.

**Key words:** image clustering, deep neural network, autoencoder, graph neural network

## 目录

1 前言 .....	1
2 相关工作 .....	3
2.1 自编码器 .....	3
2.2 图卷积神经网络 .....	4
3 建立深度聚类网络模型 .....	6
3.1 面向聚类任务的变分自编码器 .....	7
3.1.1 一般变分自编码器 .....	7
3.1.2 变分自编码器与高斯混合模型 .....	8
3.2 引入图卷积神经网络 .....	9
3.2.1 样本拓扑图与邻接矩阵 .....	10
3.2.2 VAE 与 GCN 之间的信息传递 .....	10
3.2.3 GCN 的自监督机制 .....	11
3.3 深度聚类网络模型 VGDCN .....	12
4 实验 .....	14
4.1 设置实验初始条件 .....	14
4.1.1 实验数据集 .....	14
4.1.2 网络超参数 .....	14
4.1.3 评价指标 .....	14
4.2 高斯混合模型的重要性 .....	15
4.3 用变分自编码器生成新图片 .....	15
4.4 图神经网络对聚类效果的提升 .....	16
4.5 关于平衡系数 $\epsilon$ 的探讨 .....	16
4.6 关于邻近样本数 $M$ 的探讨 .....	17
5 应用 .....	19
5.1 数据集与应用场景 .....	19
5.2 实验结果 .....	19
6 总结与展望 .....	21
参考文献 .....	22
致谢 .....	24



# 1 前言

随着信息技术的日新月异,图像这种信息介质被应用的越来越广泛,然而大多数图像都处于未标记状态。标记图片是非常耗费时间与精力的,为了更好地利用这些未标记图片,图像聚类成为了近年来颇有热度的研究课题。

所谓聚类,即“物以类聚”[1],通过聚类算法,计算机能够自动将相似的图片归为一类,同时将区别较大的图片归为不同类。在过去的几十年中,很多图像聚类算法都被成功的应用到了现实问题中,如数据挖掘、推荐引擎、生物种群统计等。近年来,深度学习所取得的巨大突破也使包括聚类在内的人工智能和机器学习算法大放异彩,因此,深度聚类受到了无监督学习领域的广泛关注。深度聚类的基本思想便是利用深度网络强大的隐层特征提取能力,因此学习到有效的隐层特征便是深度聚类的一项关键的先决条件。例如,利用预先训练的自编码器学习样本特征,然后对特征向量进行 K-means 聚类。直至目前,深度聚类网络在多个数据集上都实现了最好的聚类效果,并成为了聚类任务不可或缺的方法。

在本研究中,作者以文献[2]的深度聚类算法 DEC 为研究基础,将原聚类算法的传统自编码器替换为变分自编码器,并引入高斯混合模型,力求通过建立样本的统计学模型从而提取到更有效的样本特征,提升聚类准确率,同时也使深度网络具有生成新样本的功能。该深度网络首先预训练一个变分自编码器,然后将编码部分应用到聚类网络中,采用高斯混合模型(GMM)对特征进行预聚类,最后用 ELBO 损失函数对网络整体进行反向传播,优化网络参数,实验表明该网络模型相比 DEC 具有更好的聚类效果。

目前,大多数深度聚类算法都关注样本本身,而很少考虑样本的结构性信息。值得提出的是,数据特征学习领域已经验证了样本结构性信息的重要性,这些结构性信息揭示了样本之间的潜在相似程度,能为学习隐层特征提供正向的引导。在之前,一种被称为谱聚类的传统聚类方法,就是将图片样本视为一张拓扑图中的多个节点,并利用拓扑图中结构信息形成的特征向量进行聚类;近期出现的图卷积神经网络也会对图结构和节点属性信息同时进行编码进而进行特征学习。总的来讲,样本结构性信息是样本特征学习的重要角色,然而它并没有广泛应用在深度聚类方法中。

因此,为了补齐这一短板,作者在网络模型中引入图神经网络,并在变分自编码器与图卷积层之间设计了信息传递环节,使得图神经网络既能得到样本本身的信息,也能提取到样本之间的结构性信息,最后以图神经网络的输出作为聚类结果,为了使图神经网络得到正确的引导,为其设计了自监督机制,使其网络参数能够不断更新优化。

总的来讲,本研究的主要贡献如下:

(1)以 DEC 算法为基础建立了新的网络模型,将变分自编码器与高斯混合模型二

者结合在一起，相比于深度聚类算法 DEC 具有更好的聚类效果。

(2)在变分自编码器的基础上引入图神经网络，并在变分自编码器与图神经网络之间设计传递器，使图神经网络融合了样本本身信息和样本结构信息，经过在 USPS 数据集上进行对比验证，进一步提高了算法准确率，所提出模型名为 VGDCN。

(3)自制植物叶片数据集，并将 VGDCN 应用到该数据集上，验证了算法的普适性。

本研究的全部代码可在 <https://github.com/BJFUWDW/A-Deep-Clustering-Network-based-on-VaDE-and-GCN> 下载。

## 2 相关工作

这一节主要介绍了与本研究相关性最大的工作：自编码器与图卷积神经网络。

### 2.1 自编码器

自编码器 (autoencoder, AE) 是经常在无监督学习中使用的一类神经网络, 其原理是通过将输入的数据作为标签, 力求通过降低 MSE 损失使网络的输出结果尽可能等于输入, 从而提取输入信息的隐层特征(representation learning)。

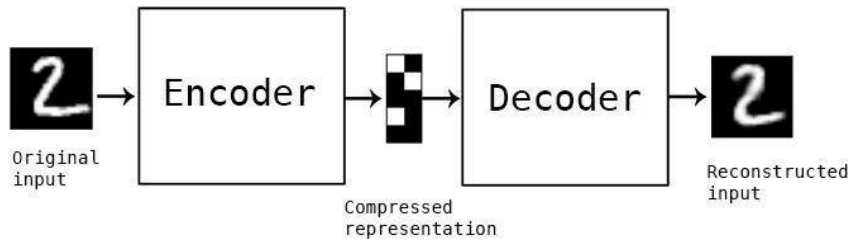


图 2.1 自编码器

Figure 2.1 AutoEncoder

1985 年, Hinton 与 Sejnowski 等发明了玻尔兹曼机, 是一个最早能够学习内部表达的神经网络模型, 但它只存在于理论中, 并没有解决机器学习中的实际问题; 为了解决玻尔兹曼机学习困难的问题, Hinton 与 Sejnowski 提出了自编码器的雏形, 在文献[3]中被称为编码问题(The encoder problem); 1986 年随着反向传播算法的提出 (Back-Propagation, BP), 自编码器得到了更深入的研究 [4]; 1987 年, LeCun 通过多层感知器构建了自编码器神经网络[5], 首次将自编码器作为一类正式的神经网络结构, 并将其定义为编码器(Encoder)和(Decoder)两部分; 2006 年, Hinton 和其学生 Salakhutdinov 将单层自编码器变成深度自编码器, 首次用自编码器做到了神经网络数据降维 [6]。

在图像聚类领域, 自编码器也受到了广泛应用。2016 年, Xie et al[2]预训练了一个深度自编码器提取输入图片的隐层特征, 首次提出了能够同时解决优化聚类效果与特征学习的数学模型。该模型首先通过深度自编码器初始化数据到特征的映射, 再使用 K-means 生成特征的聚类中心点。然后用 t-分布通过软分配方式生成每个样本的簇分配概率分布, 并以此为基础构造辅助目标分布, 最后使用 KL (Kullback-Leibler) 散度损失来衡量软分布与辅助目标分布之间的相似程度, 并优化聚类中心点, 直至网络收敛; 2017 年, Guo et al 发现了 DEC 中由于舍弃解码器部分而造成的特征空间损坏问题, 为解决此问题, 在原有研究的基础上引入了重构损失函数, 提出了 IDEC[7], 在优化 KL 散度的同时也优化重构过程, 在特征学习与聚类判别同时进行的情况下保护了原有数据结构, 实验证明其方法具有更好的聚类效果; 2018 年, 全卷积自编码器的判别增强图像聚类方法 (DBC) [8]将深度自编码器替换为深度卷积自编码器, 提取到了更具代表性的特征向量, 并以判别式增强分配的方式提高聚类分配的纯度。在

其他方面不变的情况下,得到了更优的结果;2020年,Tu等人在原有研究的基础上引入了图自编码器的概念,提出了DFCN[9],相比于原有方案中单一的使用自编码器提取图片结构,图自编码器能够提取图片中的节点属性与邻接矩阵,通过加入图自编码器能够使提取出的特征更好的代表原始图片。为了完成深度网络的反向传播,Tu等人提出了三通路式的优化方法,三条通路分别对自编码器重构损失、图自编码器重构损失、KL散度损失进行优化,在多个数据集上都显示出了迄今最好的聚类效果。

然而,上述基于自编码器的聚类方法都只专注于从样本本身学习到特征表达,忽略了样本之间的结构性信息;除此之外,其学习到的特征也只能应用于聚类任务,无法将其扩展到其他任务中,比如数据生成,因此,这些网络模型并没有得到样本数据真正的隐式表达。

## 2.2 图卷积神经网络

在过去的几年中,LeNet-5、AlexNet等经典卷积神经网络模型在多种深度学习任务上都得到了比传统方法更好的实验结果,因此,卷积神经网络受到了学术界的广泛关注。然而,传统的卷积神经网络最大的特点在于平移不变性,只能处理欧氏空间数据,不能处理非欧氏空间数据,所以会在现实中的多种应用场景中受限,如推荐系统、脑信号、计算几何等,图卷积神经网络便是为处理这种非欧氏空间数据而提出的。

图神经网络在很早便已出现,起初是为了网页排序而产生的,由于当时还没有机器学习、深度学习等概念的出现,因此,研究人员大多以统计的方法进行建模,如PageRank[10]、HITS[11]等;近年来,随着深度学习的兴起,科学家也开始思考如何利用深度学习的方式处理非欧氏空间数据,代表性的工作是2017年提出的网络表征学习方法——图嵌入(Network Embedding)。

近年来,研究人员倾向于关注如何借鉴传统卷积神经网络,利用深度学习工具对图卷积神经网络进行端到端的建模,图卷积神经网络便是其中最成功的一支。2013年,Bruna J等人建立了第一个图卷积神经网络,采用谱方法在非欧氏图上构建卷积算子,缺点是空间复杂度较高;ChebNet[13]和GCN[14]采用参数化卷积算子的方法降低了空间复杂度,虽然这两种方法在原理上也属于谱方法,但已经开始从空间角度定义图网络中节点的权重矩阵,对后续的空间方法起到了启发作用。

为了获取数据样本的结构性信息,一些基于图卷积神经网络的聚类方法也得到了广泛应用。2016年,Kipf[15]等人采用图卷积神经网络作为编码部分(Encoder),建立了图自编码器(GAE)和图变分自编码器(VGAE),从而将图结构信息和节点特征整合在一起,得到样本的隐式表达;Yu[16]等人提出了基于图表述的深度聚类算法,该算法首先用堆叠自编码器学习了一个对于原始图像的非线性嵌入表示,然后在嵌入的基础上运行一个it-means算法,得到最终聚类结果;Cai[17]等人提出了图卷积子空间聚类算法(GCSC),并应用在了高光谱图聚类任务上,该算法通过采用图神经网络将子



空间的自表达性质重建,使其成为了非欧氏域,并通过一个高鲁棒性的图嵌入字典来提高算法效果。

然而,上述所提到的基于图卷积神经网络的方法都依赖于重构邻接矩阵,并且只能够从图结构中学习数据样本的特征表达,忽略了数据样本本身的信息,可能会受限于图结构之间的交错重叠,产生过平滑现象[26]。

### 3 建立深度聚类网络模型

这一节主要介绍了本研究所提出的深度聚类网络模型，首先用变分自编码器替换了传统自编码器，提取到更具有表征力的特征向量，同时，为了提取到样本结构性信息，引入了图卷积神经网络，为了使图卷积神经网络能够同时结合样本本身信息与样本结构性信息，将变分自编码器的每一神经层都与图卷积网络中的对应层建立了连接，从而将变分自编码器提取到的特征信息传递到图卷积网络中。最后，为了使变分自编码器和图卷积网络能够在训练中进行网络参数的更新优化，分别为其设计了 ELBO 损失函数和自监督机制。下面将对所提出模型进行详细描述。

表 3.1 本文中所用符号及其意义

Table 3.1 The notations and their meaning

符号	意义
$X$	输入的图片样本
$Z$	变分自编码器提取的特征向量
$\hat{X}$	重构的图片
$S$	相似度矩阵
$A$	邻接矩阵
$Z_G$	图卷积神经网络提取的特征向量
$H^{(l)}$	变分自编码器第 $l$ 层学习到的特征向量
$W_e^{(l)}$	变分自编码器第 $l$ 层的权重
$b_e^{(l)}$	变分自编码器第 $l$ 层的偏差
$W^{(l)}$	图卷积神经网络第 $l$ 层的权重
$Q$	簇分配概率分布
$P$	辅助目标分布

### 3.1 面向聚类任务的变分自编码器

#### 3.1.1 一般变分自编码器

变分自编码器（Variational Auto-Encoders, VAE）是一种面向生成的深度网络模型，由 Kingma [18]等人于 2014 年提出。与传统的自编码器通过简单全连接线性层提取隐层特征不同，它以正态分布的形式描述隐层特征，因此能够生成输入样本中并不存在的数据，其生成数据的真实程度甚至可以和时下热门的生成对抗网络（Generative Adversarial Networks, GAN）相媲美，因此受到了无监督学习领域的广泛关注。

在[2]所建立的深度网络模型中，采用传统自编码器提取样本特征，并进行 Kmeans 聚类，本研究将传统自编码器改为变分自编码器，力求提取到更具有代表性的特征向量，下面对其数学模型进行简要描述。

给定数据样本  $X \in R^{N \times d}$ ，用  $X_i$  表示第  $i$  个样本， $N$  表示样本数量， $d$  表示样本维数，我们希望训练一深度网络，提取出样本的均值与方差，并用提取到的均值和方差计算一正态分布，从中随机采样得到其特征空间下的表示  $Z$ ，最终用  $Z$  生成新样本  $\hat{X}$  输出，整体过程可用下图表示：

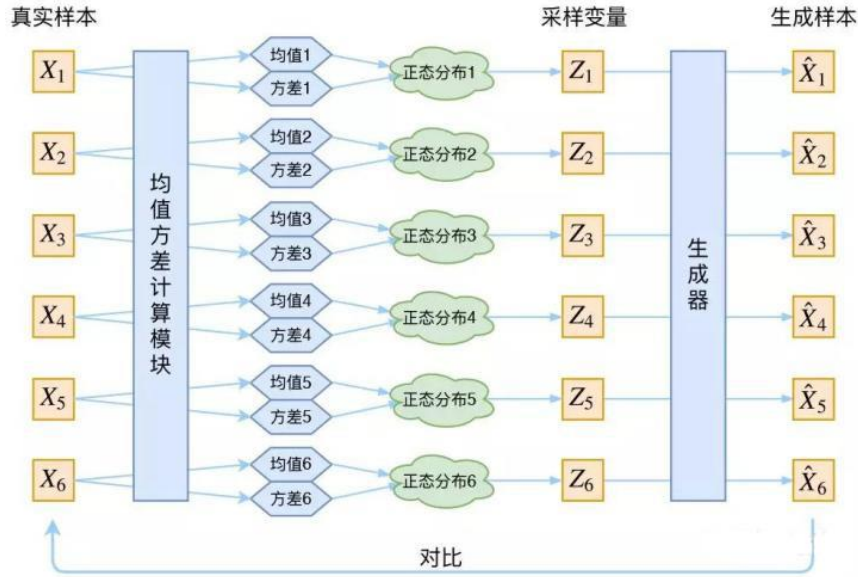


图 3.1 一般变分自编码器

Figure 3.1 General variational autoencoder

为了使生成样本尽可能近似于输入样本，定义重构损失函数作为网络的优化目标：

$$L_{res} = \frac{1}{N} (\hat{X} - X)^2 \quad (3.1)$$

随着网络训练次数的增加，该损失函数的值将会逐渐减小，即输出样本与输入样本的相似程度将逐步提高。但在这个重构过程中，特征向量  $Z$  是随机采样得来的，网络提取出的样本方差也就成了采样过程中的噪声，增加了重构难度，网络在一次又一次的训练迭代中，会倾向于让方差为 0，这样会使采样缺乏随机性，无论怎样采样， $Z$  都将是唯一结果。简而言之，变分自编码器(VAE)会逐渐退化成传统自编码器(AE)。

为了解决这一问题，变分自编码器力求使所有的后验分布(即前文提到的用均值和方差计算的正态分布) $p(Z|X)$ ，接近于一个标准正态分布 $N(0,1)$ ，既保证了方差不会为 0，也保证了模型具有生成能力，这是因为，若所有的 $p(Z|X)$ 都满足标准正态分布，那么采样变量 $Z$ 的分布 $p(Z)$ 将满足：

$$p(Z) = \sum_X p(Z|X)p(X) = \sum_X N(0,1)p(X) = N(0,1) \sum_X p(X) = N(0,1) \quad (3.2)$$

即采样变量 $Z$ 满足一标准正态分布，保证了模型的生成能力。

因此，为了使所有后验分布满足标准正态分布，在重构损失函数的基础上引入了新的损失函数：

$$L_{KL} = KL(N(\mu, \sigma^2) \| N(0,1)) = \frac{1}{2} \sum_{i=1}^d (\mu_{(i)}^2 + \sigma_{(i)}^2 - \log \sigma_{(i)}^2 - 1) \quad (3.3)$$

式(3.3)中，KL 表示数学中的 KL 散度，是用来衡量两个概率分布之间相似程度的函数，KL 散度越大，则两个概率分布之间的相似程度越小， $d$  表示样本维数， $\mu_{(i)}$ 、 $\sigma_{(i)}$  分别表示深度网络所提取的均值向量和方差向量中的第  $i$  个分量。

综上所述，一般变分自编码器深度网络的损失函数为：

$$L = L_{res} + L_{KL} \quad (3.4)$$

通过优化该损失函数，网络便可以生成近似于输入样本的图片数据。

### 3.1.2 变分自编码器与高斯混合模型

在上一小节中，我们对一般变分自编码器的数学模型进行了简要描述，一般变分自编码器为了使模型具有生成能力，采用了 KL 散度作为损失函数来使特征向量 $Z$ 满足标准正态分布，而这与聚类任务目标其实是不相符的，在这一小节中，我们将讨论如何将变分自编码器用于聚类任务。

给定图片数据样本 $X \in R^{N \times d}$ ，用 $x_i$ 表示第  $i$  个样本， $N$  表示样本数量， $d$  表示样本维数，假设我们要将其聚为  $K$  个簇，由于计算机无法直接对原数据进行聚类，因此需要将原数据映射到特征空间，找到合适的特征向量 $Z$ ，并用于聚类任务。在理想状况下，我们希望特征向量 $Z$ 来自于  $K$  个不同类别，并且类与类之间差异明显，这样计算机便能够进行识别，达到较好的聚类效果。而在一般变分自编码器中，为了保证其模型具有生成能力，使特征向量 $Z$ 满足标准正态分布，虽然方便了采样，但并没有保证特征向量 $Z$ 具有  $K$  个不同类别，因此，虽然 $Z$ 对原数据具有很好的表征力，但不能直接用于聚类任务，如何既保证其生成能力，又能让其能应用于聚类任务，是一个值得探讨的问题。高斯混合模型(Gaussian Mixture Model, GMM)[19]便为解决这一问题提供了方案。

高斯混合模型是一种流行的无监督学习算法，其原理类似于传统的 K-means 聚类方法[20]，都需要指定簇数  $K$ ，不同之处在于，K-means 只能将每个样本划分为一个类，而 GMM 可以给出一个样本对于所有类别的概率。

从数学上来讲, GMM 假设样本数据是由多个高斯分布(正态分布)混合生成的, 如式(3.5):

$$p(x) = \sum_{k=1}^K \pi_k N(x | \mu_k, \sigma_k) \quad (3.5)$$

式中,  $N(x | \mu_k, \sigma_k)$  称为混合模型中的第  $k$  个分量(component),  $\pi_k$  是每个分量对应的权重,  $K$  为分量总数。

GMM 实现聚类的过程如下: 若要将样本数据聚为  $K$  簇, 则首先假定样本数据由  $K$  个高斯分布混合生成, 然后计算出每个样本对应不同分量的权重  $\pi_k$ , 选取概率(权重)最大的作为其聚类结果并输出该簇对应的编号。

GMM 实现数据生成的过程如下: 根据权重随机选取一个高斯分布, 并从中采样得到  $Z$ , 利用  $Z$  生成新的样本。

因此, 假如我们能够设计损失函数来使特征向量  $Z$  满足 GMM 分布, 那么就既保证了其能够用于聚类, 也保证了生成新图片数据的能力, 其实现过程如算法 1 所示:

---

#### 算法 1: 面向聚类任务的变分自编码器

---

输入: 图片样本数据、聚类簇数  $K$

输出: 聚类结果, 生成的图片数据

- (1) 利用变分自编码器中的 Encoder 部分提取样本的均值与方差;
  - (2) 用该均值与方差计算一正态分布, 并从中随机采样得到特征向量  $Z$ ;
  - (3) 设置 GMM 的  $K$  值等于样本簇数, 对特征向量  $Z$  进行拟合, 得到  $K$  个不同的高斯分布, 用  $N(\mu_c, \sigma_c) (c = 1, 2, 3 \dots K)$  表示, 并根据不同高斯分布对应权重输出聚类结果;
  - (4) 将特征向量  $Z$  输入到解码器(Decoder), 生成新的图片数据并输出;
  - (5) 根据式(3.6)进行反向传播, 更新网络参数, 使得生成的图片数据近似于输入图片, 同时使  $Z$  的统计学分布  $p(Z)$  满足高斯混合分布, 不满足停止条件时回到(1)。
- 

根据算法 1, 我们完成了对一般变分自编码器的改进, 使其特征向量既能用于聚类, 也保留了生成数据的功能, (5)中所提到的损失函数如下:

$$L_{ELBO}(X) = E_{q(Z, c | X)} [\log p(X | Z)] - \text{KL}(q(Z, c | X) \| p(Z, c)) \quad (3.6)$$

式中的第一项可称为重构项, 它使网络生成的新图片数据更为近似于输入数据; 第二项为 KL 散度, 它使特征向量  $Z$  的统计学分布近似于 GMM 所拟合的高斯混合分布, 由于该损失函数涉及较为复杂的数学推导, 在此不详细列出, 只列出其简明意义, 具体可参考文献[21]。

## 3.2 引入图卷积神经网络

在上一节中, 我们对一般变分自编码器进行改进, 通过引入高斯混合模型, 使其能够应用于图片聚类任务, 提取到了相比于传统自编码器更能代表原数据的特征向量

$Z$ ，完成了对于样本本身信息的提取。这一节中，我们将针对于样本结构性信息的提取，在变分自编码器的基础之上引入图神卷积神经网络(Graph Convolutional Network, GCN)，完成最终深度聚类网络 VGDCN 的建模。

### 3.2.1 样本拓扑图与邻接矩阵

图卷积神经网络的输入不仅要有图片样本本身，同时还需要输入根据图片样本所建立的邻接矩阵(adjacency)。

若要建立图片样本邻接矩阵，则需要先建立样本拓扑图(graph)，在拓扑图中，图片样本将被视为一个个不同节点，节点之间为非欧氏距离关系，如图 3.2 所示：

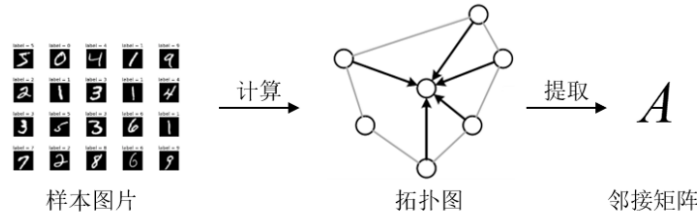


图 3.2 邻接矩阵计算过程

Figure 3.2 Calculation of adjacency matrix

假设有图片数据样本  $X \in R^{N \times d}$ ，用  $x_i$  表示第  $i$  个样本， $N$  表示样本数量， $d$  表示样本维数，对于每一个样本，我们首先要根据样本之间的相似程度找到其  $M$  个最相似的邻近图片，并把它和这些邻近图片连接起来，形成拓扑图，从而提取邻接矩阵。样本之间的相似程度可以由相似度矩阵  $S \in R^{N \times N}$  来评判，本研究中相似度矩阵采用的热核法进行计算[22]，第  $i$  个和第  $j$  个样本之间的相似度矩阵为：

$$S_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}} \quad (3.7)$$

式中， $t$  为热传导方程中的时间参数。

在计算出相似度矩阵  $S$  之后，便可以形成一张拓扑图，提取邻接矩阵  $A$ ，完成对样本结构性信息的提取，实验 4.4 中设置  $M = 3$ 。

### 3.2.2 VAE 与 GCN 之间的信息传递

在 3.1 中，我们建立了面向聚类任务的变分自编码器，得到了具有很好表征力的特征向量  $Z$ ，完成了样本本身信息的提取；在上一小节中，我们完成了邻接矩阵  $A$  的计算，提取了图片样本的结构性信息，那么，怎样同时利用两种信息实现聚类呢？下面将介绍如何利用图神经网络实现信息融合，并输出聚类结果。

首先，我们假设变分自编码器神经网络的编码部分(Encoder)共有  $L$  层，并用  $l$  表示层编号，因此，第  $l$  层编码器所学习到的特征向量  $H^{(l)}$  可以用下式得到：

$$H^{(l)} = \phi(W_e^{(l)} H^{(l-1)} + b_e^{(l)}) \quad (3.8)$$

式中， $\phi$  表示了全连接层的激活函数，如 Relu[23] 或 Sigmoid 函数， $W_e^{(l)}$  和  $b_e^{(l)}$  分别表示了第  $l$  层的权重和偏差， $H^{(0)}$  表示原样本数据。

为了使 GCN 能够得到 VAE 学习到的特征信息  $H^{(1)}, H^{(2)}, \dots, H^{(l)}$ , 设计平衡系数  $\varepsilon$ , 将 VAE 每一层学习到的特征向量乘以平衡系数加到 GCN 的对应卷积层作为该层输出, 如式(3.9)所示:

$$\tilde{Z}_G^{(l-1)} = (1 - \varepsilon)Z_G^{(l-1)} + \varepsilon H^{(l-1)} \quad (3.9)$$

式中,  $Z_G^{(l-1)}$  表示 GCN 第  $l-1$  层学习到的特征向量,  $\tilde{Z}_G^{(l-1)}$  表示结合了变分自编码器特征信息后的特征向量, 并作为该层的输出输入到下一层进行计算。

$$Z_G^{(l)} = \phi \left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \tilde{Z}_G^{(l-1)} W^{(l-1)} \right) \quad (3.10)$$

式中,  $A$  为邻接矩阵,  $\tilde{A} = A + I$ ,  $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ ,  $I$  为单位对角矩阵,  $\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$  为一正则化后的邻接矩阵,  $W$  为权重矩阵。

至此, 我们将 VAE 中编码器的每一层都与 GCN 中的对应卷积层连接了起来, 将 VAE 学习到的特征信息传递到了 GCN 中, 实现了二者的信息融合, 在第 4 章中将会对上述数学模型进行更深入的理论分析。

需要注意的是, GCN 的第一层输入为原始图片数据  $X$ , 即 GCN 第一层的输出为:

$$Z_G^{(1)} = \phi \left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X W^{(1)} \right) \quad (3.11)$$

GCN 的最后一层是一个多元分类层, 采用了 softmax 函数:

$$Z_G = \text{softmax} \left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} Z_G^{(L)} W^{(L)} \right) \quad (3.12)$$

最终的输出结果  $z_{ij} \in Z_G$  表示了第  $i$  个样本属于第  $j$  簇的概率, 因此我们可以把  $Z_G$  视作一簇分配概率分布。

### 3.2.3 GCN 的自监督机制

根据上一小节, 我们实现了 VAE 与 GCN 之间的信息传递, 并将 GCN 的最终输出作为聚类结果, 然而, 不同于 VAE 中采用了高斯混合模型, GCN 并不知道自己的特征向量要用于聚类任务, 其网络输出具有随机性, 因此, 需要设计目标损失函数, 对 GCN 进行约束, 从而实现网络参数的更新。

本研究的自监督机制设计参考了文献[2], 采用了目标辅助分布(target distribution)来对 GCN 的最终输出进行约束。

对于第  $i$  个样本与第  $j$  个簇, 我们采用学生  $t$  分布[24]作为内核来衡量特征向量  $h_i$  和簇中心向量  $\mu_j$  的相似程度:

$$q_{ij} = \frac{(1 + \|h_i - \mu_j\|^2 / \nu)^{-\frac{\nu+1}{2}}}{\sum_{j'} (1 + \|h_i - \mu_{j'}\|^2 / \nu)^{-\frac{\nu+1}{2}}} \quad (3.13)$$

式中,  $h_i$  表示  $H^{(L)}$  的第  $i$  个分量,  $\mu_j$  为在 VAE 学习到的特征向量  $Z$  上应用 GMM

后得到的结果， $v$ 表示学生  $t$  分布的自由度。

简单来讲， $q_{ij}$ 可理解为第  $i$  个样本属于第  $j$  簇的概率，并且能由 3.1.2 中的数学模型计算得到。在得到了所有样本的分布  $Q = [q_{ij}]$  之后，我们希望通过学习那些具有高自信度的簇来对特征向量提取进行优化，具体来说，我们想让特征向量更接近于簇中心，从而提高簇的凝聚力。因此，计算辅助目标分布  $P$  如下：

$$p_{ij} = \frac{q_{ij}^2 / f_j}{\sum_j q_{ij}^2 / f_j} \quad (3.14)$$

式中， $f_j = \sum_i q_{ij}$ ，为所有样本分配到簇  $j$  的频次。

直观来说，辅助目标分布  $P$  由  $Q$  计算得到，并能为簇分配提供正向指引，如果能设计目标函数来使  $Q$  更接近于  $P$ ，就能够使聚类效果更好。

3.2.2 中提到，GCN 最终的输出层为一多元分类层，输出的  $Z_G$  与  $Q$  形式相似，同为一簇分配概率分布，因此，如果能使  $Z_G$  接近于  $P$ ，那么便能够为 GCN 学习特征向量提供有利指引，利用 KL 散度设计 GCN 目标损失函数如下：

$$L_{GCN} = \text{KL}(P \parallel Z_G) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{z_{ij}} \quad (3.15)$$

由于辅助目标分布  $P$  是由深度聚类网络本身计算得到，所以称该机制为自监督机制。

### 3.3 深度聚类网络模型 VGDCN

至此，我们已经建立了面向聚类任务的变分自编码器、图神经网络的数学模型，并在两者之间设计了信息传递环节，使图神经网络既能够提取样本的结构性信息，也能提取到样本本身的信息，最后，利用自监督机制对图神经网络进行了约束，将其最后一层的输出作为聚类结果，上述整体过程可如图 3.3 和算法 2 所示。

---

#### 算法 2：深度聚类网络模型 VGDCN

---

输入：图片样本数据、邻接矩阵  $A$ 、聚类簇数  $K$

输出：聚类结果，生成的图片数据

(1)预训练一变分自编码器，初始化  $W_e^{(l)}$  和  $b_e^{(l)}$ ，得到特征向量  $Z$ ；

(3)设置 GMM 的  $K$  值等于聚类簇数，对特征向量  $Z$  进行拟合，得到  $K$  个不同的高斯分布，实现预聚类；

(4)初始化 GCN 的权重  $W^{(l)}$ ；

(5)将变分自编码器每一层学习到的特征向量  $H^{(1)}, H^{(2)}, \dots, H^{(L)}$  传递到对应的图卷积层；

(6)用(3)中的预聚类结果计算辅助目标分布  $P$ ；

(7)计算 GCN 最后一层输出的  $Z_G$  分布；

(8)用特征向量  $Z$  重构图片，输出重构的图片数据  $\hat{X}$ ；

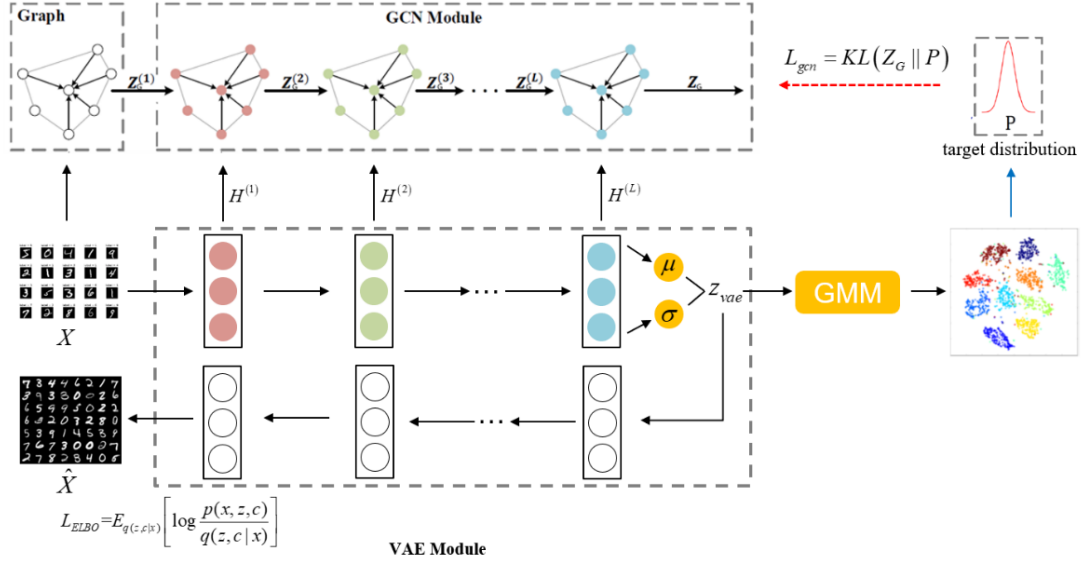
---



(9) 计算  $L_{ELBO}$ ,  $L_{GCN}$ ;

(10) 反向传播, 更新网络参数, 若未达到预先设置的迭代次数, 则回到(5);

(11) 将  $Z_G$  作为最终聚类结果输出。



本研究所提出的深度聚类网络模型 VGDCN 图中所示,  $X, \hat{X}$  分别为输入的图片样本数据和重构的图片数据;  $H^{(l)}$ ,  $Z_G^{(l)}$  分别为 VAE 和 GCN 第  $l$  层学习到的特征向量;  $Z_{vae}$  表示变分自编码器最终学习到的特征向量; 蓝色实线箭头表示辅助目标分布  $P$  是由在  $Z_{vae}$  上应用 GMM 之后计算得到的簇分配概率分布; 在训练过程中, 辅助目标分布  $P$  引导着 GCN 进行网络参数更新。

图 3.3 VGDCN 整体框架

Figure 3.3 The overall framework of VGDCN

## 4 实验

在这一部分中，我们主要设计了实验来验证 3 章中的深度聚类网络模型，并将实验结果和本研究的基准模型——DEC[2]进行了对比，实验结果表明通过将传统自编码器改为变分自编码器，成功提升了算法准确率；图卷积神经网络的引入也使算法的聚类效果更上了一个台阶，证明了提取样本结构性信息的重要性；另外，模型中一些重要的参数也会影响网络的聚类效果，本章中根据实验结果对其进行了理论分析。

### 4.1 设置实验初始条件

#### 4.1.1 实验数据集

实验中所用数据集为 USPS，USPS 为美国邮政服务手写数字识别库，包含 9298 个 16\*16 像素大小的手写数字灰度图片，图片共有十个类别(数字 0 到 9)，图片的特征为其每个像素点的灰度值，并且所有灰度值大小都处于[0,2]之间。

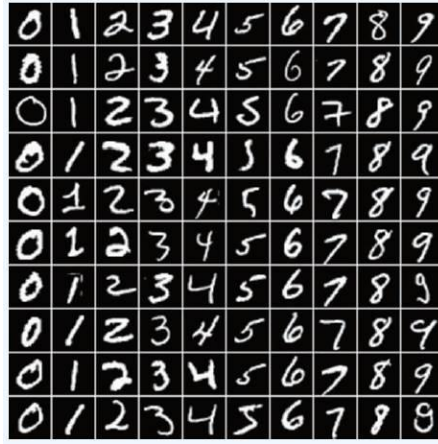


图 4.1 USPS 数据集

Figure 4.1 USPS dataset

#### 4.1.2 网络超参数

为了能够与 DEC 进行对比实验，将变分自编码器网络的结构设置为与 DEC 类似的形式，编码器(Encoder)部分为 D-500-500-2000-10，解码器(Decoder)部分为 10-2000-500-500-D，D 为样本维数；网络的所有神经层均为全连接层；优化器选为 Adam 优化器；学习率为 0.001，每 10 轮训练便会衰减 10%。

需要注意的是，为了使 GCN 能接收对应 VAE 全连接层输出的特征向量，GCN 的网络结构将与 VAE 的 Encoder 部分相同。

#### 4.1.3 评价指标

为了方便对比，本研究的聚类效果评价指标(ACC)与 DEC 中相同，将如下式定义：

$$ACC = \max_{m \in M} \frac{\sum_{i=1}^N \mathbb{1}\{l_i = m(c_i)\}}{n} \quad (4.1)$$

式中,  $l_i$  为样本的真实类别,  $c_i$  为算法计算得到的样本聚类结果,  $m$  为聚类结果与真实类别之间所有可能的一对一映射。

直观来说, 该指标表示了样本聚类结果与样本真实类别之间的最优匹配结果, 并可由文献[25]的算法计算得到。

## 4.2 高斯混合模型的重要性

在 3.1.2 中提到, 一般变分自编码器提取的特征向量不适用于聚类, 必须约束其特征向量处于高斯混合分布(GMM), 才能达到聚类的目的, 下面便通过实验来验证这一理论。

为了强调 GMM 的重要性, 我们先训练了一个一般变分自编码器, 并直接用其替换 DEC 中的传统自编码器, 由于 DEC 中采用了 Kmeans 方法对特征向量  $Z$  进行聚类, 我们把这一网络模型称为 VAE+Kmeans; 类似的, 我们把 3.1.2 中的网络模型称为 VAE+GMM, 二者采用同样的数据集、超参数进行训练, 实验结果如图 4.2 所示。

从图中可以看出, VAE+Kmeans 模型完全失效, 无法进行聚类; 而 VAE+GMM 模型的聚类准确率随着训练轮数的增加而增长, 最终稳定在 57.5%左右, 较 DEC 算法提升了 4.5%-5%, 证明了 GMM 的重要性。

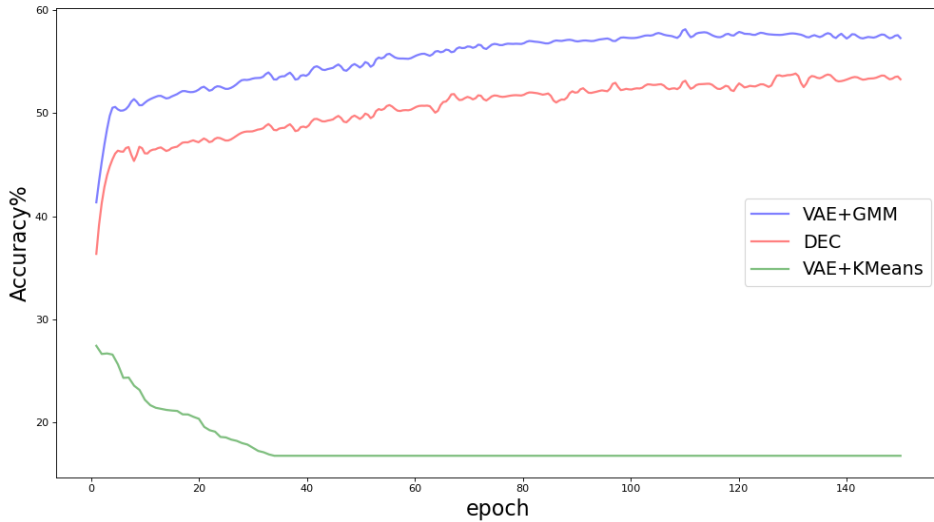


图 4.2 高斯混合模型的重要性  
Figure 4.2 The importance of GMM

## 4.3 用变分自编码器生成新图片

在 3.1.2 中提到, 通过引入高斯混合模型, 我们既让变分自编码器提取到的特征向量能够用于聚类, 也保留了其生成新图片样本的能力, 下图便是训练过程中变分自编码器生成的新图片, 从图中可以看出, 随着训练轮数的增加, 生成的数字图片可辨识度越来越高, 更加近似于输入样本。

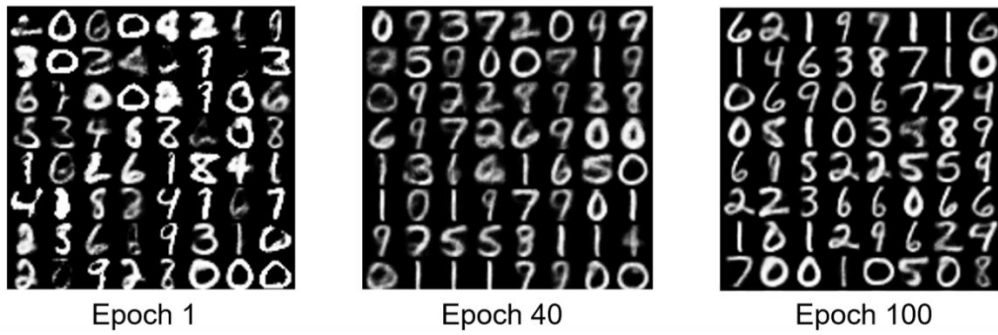


图 4.3 变分自编码器生成新图片

Figure 4.3 The images generated by VAE

#### 4.4 图神经网络对聚类效果的提升

3.2 中提到在变分自编码器的基础上引入图神经网络，从而提取样本的结构性信息，并在 3.3 中建立了深度聚类网络模型 VGDCN，实验结果如下。从图中可以看出，聚类准确率在 VAE+GMM 的基础上提升了 5-5.5%，稳定在 62.7% 左右。

值得注意的是，VGDCN 训练一开始的准确率要低于 VAE+GMM 很多，这是因为在训练轮数较低时，变分自编码器网络所计算的辅助目标分布  $P$  无法为 GCN 提供指引作用，因此 GCN 最后的输出聚类结果缺乏定向约束；随着训练轮数的增加，VAE 和 GCN 的网络参数不断更新，自监督机制开始生效，聚类准确率逐渐上升。

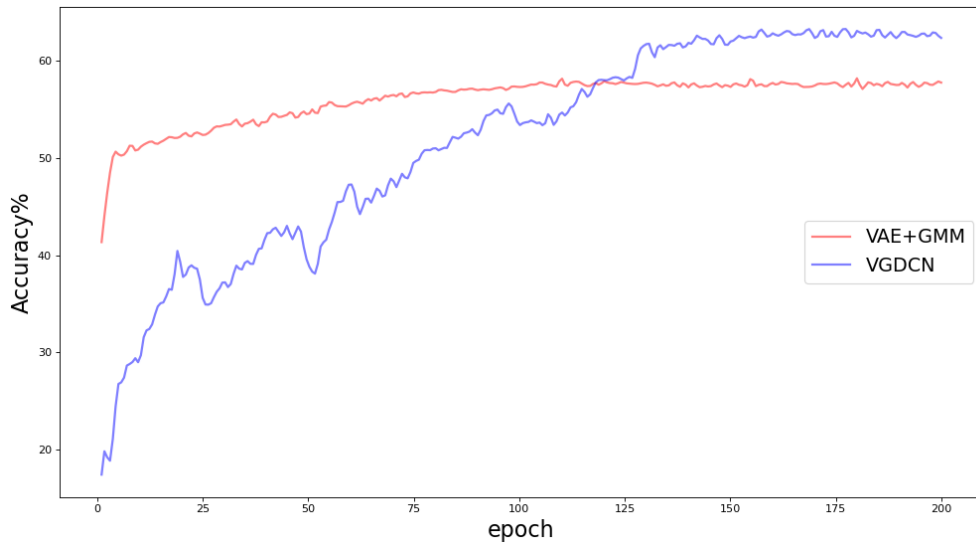


图 4.4 GCN 对聚类效果的提升

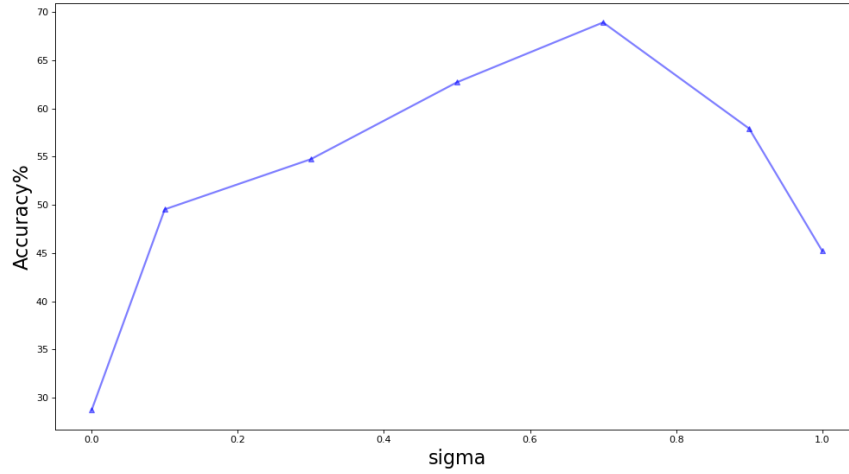
Figure 4.4 The promotion of accuracy by GCN

#### 4.5 关于平衡系数 $\varepsilon$ 的探讨

在 3.2.2 中，我们设置平衡系数  $\varepsilon$  将样本本身信息和样本结构性信息以权重的方式相加，并默认其值为 0.5。在本小节的实验中，我们将平衡系数  $\varepsilon$  设置为不同值，并探讨这些不同的  $\varepsilon$  是如何影响深度网络的聚类结果的。

实验中设置  $\varepsilon = \{0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0\}$ ，并记录了不同  $\varepsilon$  对应的网络聚类结果，

如图 4.4 所示:



注:  $\varepsilon = 0$  表示 GCN 没有输入 VAE 的特征信息;  $\varepsilon = 1.0$  表示 GCN 只采用了 VAE 的特征信息

图 4.4  $\varepsilon$  对聚类效果的影响

Figure 4.4 The influence of different  $\varepsilon$  on accuracy

从图 4.4 中可以看出, 当  $\varepsilon = 0.7$  时, VGDCN 的聚类效果最好, 这表明 USPS 数据集相比于样本结构性信息更侧重于样本本身的信息;  $\varepsilon = 0.0$  时, VGDCN 的聚类效果最差, 这是因为当  $\varepsilon = 0.0$  时, VGDCN 就等同于一多层的图卷积神经网络, 会有非常严重的过平滑问题[26], 使聚类准确率下降, 与之相比,  $\varepsilon = 0.1$  时的聚类效果便提升了很多, 这表明仅将一小部分变分自编码器学习到的特征信息注入到 GCN 中便能够缓解过平滑现象。

#### 4.6 关于邻近样本数 $M$ 的探讨

在 3.2.1 中提到, 为了提取样本的结构性信息, 针对于每一个样本, 我们计算了相似度矩阵  $S$ , 选取了  $M$  个最相似的邻近图片, 构建了拓扑图。在这个过程中,  $M$  为一重要参数, 因此设计了实验来探讨不同  $M$  对于聚类效果的影响。

实验中设置  $M = \{1, 3, 5, 10\}$ , 并记录了不同  $M$  对应的网络聚类结果, 如图 4.5 所示:

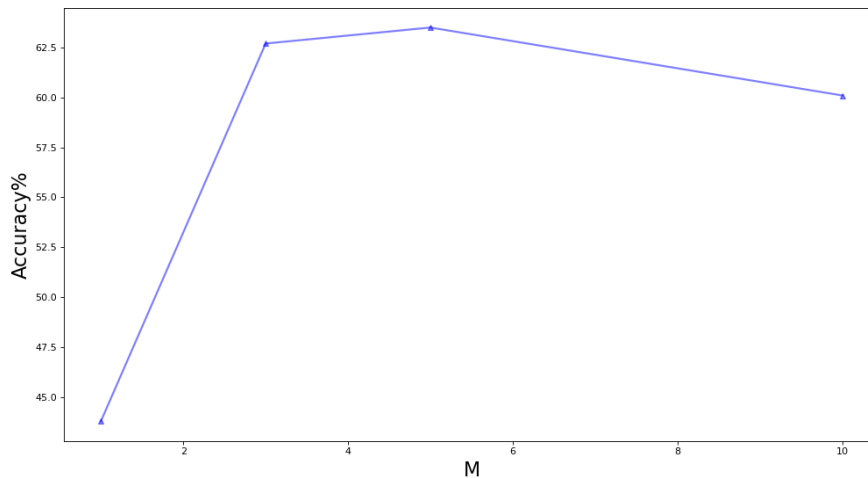


图 4.5  $M$  对聚类效果的影响

Figure 4.5 The influence of different  $M$  on accuracy

从图 4.5 中可以看出, 当 $M = 3,5$ 时, VGDCN 的聚类效果较好; 但当 $M = 1$ 时, 聚类效果会下降很多, 这是因为拓扑图中包含的样本结构性信息较少;  $M = 10$ 时, 聚类准确率略有下降, 这是因为拓扑图中的非欧式结构过于复杂, 导致了重叠现象。

## 5 应用

在 4 章中，我们利用标准数据集设计实验验证了所提出的深度聚类网络模型，在本章中，我们将 VGDCN 应用在了一自制数据集上，为其设计了具体应用场景，进一步验证了网络模型的普适性。

### 5.1 数据集与应用场景

自制数据集来自于大型植物叶片数据库 Flavia，为植物叶片识别领域广泛采用的数据库，作者在其中随机选取了十种不同植物叶片，并以 imagefolder 的方式将其整合为能够输入到网络中的数据格式。



图 5.1 自制植物叶片数据集

Figure 5.1 Self-made dataset of plant leaves

目前的植物叶片识别多为监督学习，即需要提前将植物叶片的标签输入到计算机中，这需要耗费大量的人力物力。设想有一未知的生态环境，植物学家需要利用相机采集植物样本图片，对该生态环境中的物种进行统计，若能够以无监督学习的方式让计算机对其进行识别，便能够大幅提升工作效率。

### 5.2 实验结果

将自制数据集分别输入到 VGDCN 与 DEC 中，得到结果如下：

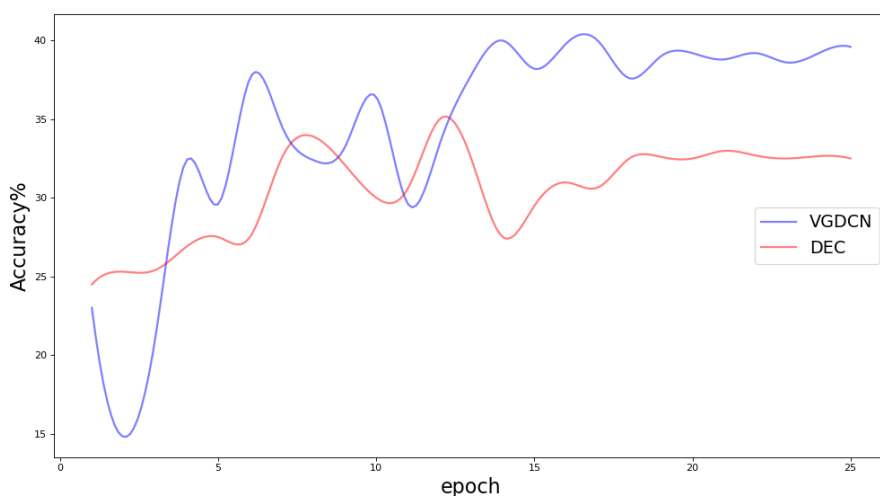
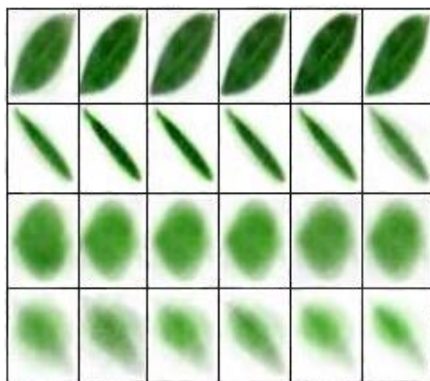


图 5.2 实验结果

Figure 5.2 Result of experiment

从图中可以看出，网络的聚类准确率在第 20 轮训练便已达到稳定态，约为 40%；

在整体的训练过程中，网络的聚类效果并不是稳定上升，发生了振荡。上述现象是因数据集中的样本数量过少导致的，该数据库针对于每一类叶片仅提供了 60-70 张样本图片，若能提高样本数量，该深度聚类网络完全可以和目前的有监督植物识别算法相媲美。



图为 VGDCN 聚类结果的可视化，每一行对应的簇分配结果分别为(由上到下)：珊瑚树、毛竹、紫荆、木莲

图 5.3 聚类结果可视化

Figure 5.3 The visualization of cluster assignment



## 6 总结与展望

本研究以 DEC 算法为基础模型,设计搭建了新的深度聚类网络模型 VGDCN。首先用变分自编码器替换了传统自编码器,提取到更具有表征力的特征向量,随后,为了提取到样本结构性信息,引入了图卷积神经网络,并在变分自编码器与图卷积神经网络之间设计了信息传递机制以实现信息融合。最后,为了使变分自编码器和图卷积神经网络能够在训练中进行网络参数的更新优化,分别为其设计了 ELBO 损失函数和自监督机制。实验证明所提出深度聚类网络模型 VGDCN 具有更好的聚类效果。

为了进一步验证算法的普适性,将其应用在自制植物叶片数据集上,得到了可信的实验结果。

目前在图像聚类领域,识别具有复杂成分的图像是较为困难的,如 imagenet 等数据集,作者的硕士阶段将对该问题进行研究,本论文为解决该问题提供了可能的思路,即:利用结合了样本本身信息与结构性信息的特征向量进行聚类。

## 参考文献

- [1]周涛, 陆慧玲. 数据挖掘中聚类算法研究进展[J]. 计算机工程与应用, 2014, 12:534-540.
- [2]Junyuan Xie. Unsupervised deep embedding for clustering analysis. In International Conference on Machine Learning(ICML), 2016.
- [3]Ackley, D.H., Hinton, G.E. and Sejnowski, T.J.. A learning algorithm for Boltzmann machines[J]. Cognitive science, 1985, 9(1):147-169.
- [4]Hintong. E . Connectionist learning procedures[J]. Artificial Intelligence, 1989.
- [5]Cun Y L . Modèles connexionnistes de l'apprentissage. 1987.
- [6]Hinton, G, E, et al. Reducing the Dimensionality of Data with Neural Networks.[J]. Science, 2006.
- [7]Guo X, Gao L, Liu X and Yin, J. Improved Deep Embedded Clustering with Local Structure Preservation. In IJCAI, 2017.
- [8]Li F F, Qiao H, Zhang B et al. Discriminatively boosted image clustering with fully convolutional auto - encoders[J]. Pattern Recognition, 2018, 83:161-173.
- [9]Wenxuan Tu. Deep Fusion Clustering Network. arXiv - CS - Machine Learning, 2020.
- [10]Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the Web. Stanford Digital Libraries Working Paper, 1998, 9(1): 1-14.
- [11]Kleinberg J M. Hubs, authorities, and communities, ACM Computing Surveys, 1999, 31(4es): 5.
- [12]齐金山, 梁循, 李志宇等. 大规模复杂信息网络表示学习: 概念、方法与挑战[J]. 计算机学报, 2018, 41(10): 2394-2419.
- [13]Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. Proceedings of the Advances in Neural Information Processing Systems, 2016, 1:3844-3852.
- [14]Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016).
- [15]Thomas N Kipf and Max Welling. Variational graph auto-encoders. arXiv preprint arXiv:1611.07308 (2016).
- [16]X. Yu, Z. Pan, Y. Gao. A graph-based image clustering method using mutual information maximization. In 2019 IEEE International Conference on Robotics and Biomimetics, 2019: 2322-2326.
- [17]Y. Cai, Z. Zhang, Z. Cai, X. Liu, X. Jiang, Q. Yan. Graph convolutional subspace clustering: A robust subspace clustering framework for hyperspectral image. IEEE Transactions on Geoscience and Remote Sensing, 2020, 8: 1-12.
- [18]Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. In ICLR, 2014.
- [19]Bishop C M. Pattern Recognition and Machine Learning (Information Science and Statistics)[M]. Springer-Verlag New York, 2006.
- [20]John A Hartigan and Manchek A Wong. Algorithm AS 136: A k-means clustering algorithm[J]. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979, 1:100-108.
- [21]Jiang Z, Zheng Y, Tan H, et al. Variational Deep Embedding: An Unsupervised and Generative Approach to Clustering. In IJCAI, 2017.
- [22]Bo D , Wang X , Shi C , et al. Structural Deep Clustering Network. In WWW, 2020.
- [23]Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In ICML, 2010: 807-814.
- [24]Laurens van der Maaten and Hinton G.E.. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 1:2579-2605.
- [25]Kuhn, HaroldW. The hungarian method for the assignment problem[J]. Naval research logistics

quarterly, 1955, 2(1-2):83–97.

[26]Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In AAAI, 2018.

## 致谢

时光荏苒，岁月如歌，四年的机械专业本科生涯还有一个多月的时间便要结束了。这四年时间，我从一个懵懂的高中生，不断地学习、成长和进步，直到今天完成了本科毕业论文。在这篇论文中，不仅仅凝聚了我四年的学习成果，也包括父母、老师和同学们对我无私的帮助。

感谢我的父母，是他们的默默支持，让我下定决心在大学里努力学习，并获得了学校的推免资格。也是他们无微不至的关怀，让我有信心去选择学习我最感兴趣的研究生专业，开始人工智能的学术生涯。

感谢我的本科毕设导师郑嫦娥副教授。郑老师大一的时候曾教授我们工程图学，可以说是我们的启蒙老师。在整个毕业设计期间，老师悉心 and 耐心的指导使我即使身处异地，也没有落下学校这边毕设的各项工作。

感谢我的研究生导师韩志研究员以及实验室的师兄师姐们。由于是机械专业出身，这次毕业设计是我第一次深度学习的实践，韩老师和师兄师姐们对我的关心照顾使我有勇气去探索、去学习。在算法改进创新的过程中遇到了很多问题，是韩老师和师兄师姐们的指导让我攻克了一个又一个难关，组内的学术交流会也使我增长了见识，更深入的理解了计算机视觉领域的前沿工作。期待未来三年与你们共同度过。

感谢我的舍友，在大学四年的学习生活中给了我多方面的关心照顾。

感谢马宇政同学等一同奋斗的挚友们，在遇到挫折时，能有人分享心中不快；收获胜利果实时，能有人分享心中喜悦。那些在实验室奋斗的一幕幕仿佛就发生在昨天，是你们的陪伴让我度过了丰富又精彩的本科四年。

最后，感谢本文的评审专家以及答辩评委对本研究提出的评审意见。