**文章编号**　1004-924X(2022)18-0001-12

# 局部-整体双向推理的文物无监督表征学习

刘　杰[1,2], 耿国华[1,2*], 田　煜[1,2], 王　毅[1,2], 刘阳洋[1,2], 周明全[1,2]

(1. 西北大学 文化遗产数字化国家地方联合工程研究中心,陕西 西安 710127;
2. 西北大学 信息科学与技术学院,陕西 西安 710127)

**摘要**:针对现有陶制文物表征学习方法是基于大量带标签数据的有监督学习方法,人工标记费时耗力且不能有效地学习到点云内在结构信息等问题,本文提出一种基于局部-整体双向推理的无监督表征学习方法。首先,提出多尺度壳卷积层级结构编码器提取不同尺度的文物碎片局部特征。其次,利用局部到整体推理模块将提取的局部特征映射得到全局特征,通过度量学习衡量两者之间差异,进行反复学习。然后,利用整体到局部推理模块以确保获取到的全局特征的质量。最后,在不同层次的局部结构和整体形状之间通过双向推理来学习文物点云表征,并将学习到的点云表征应用于分类下游任务。该网络模型在兵马俑数据集和ModelNet40公开数据集上的分类精度分别达到了93.33%和92.02%,分别高于PointNet 4.4%和2.82%。同时缩小了下游分类任务中无监督和有监督学习方法之间的差距。

**关　键　词**:无监督表征学习;多尺度;深度学习;点云分类;文物虚拟修复

**中图分类号**:TP391　　**文献标识码**:A　　**doi**:10.37188/OPE.20223018.0001

# Unsupervised representation learning for cultural relics based on local-global bidirectional reasoning

LIU Jie[1,2], GENG Guohua[1,2*], TIAN Yu[1,2], WANG Yi[1,2], LIU Yangyang[1,2], ZHOU Mingquan[1,2]

(1. *National and Local Joint Engineering Research Center for Cultural Heritage Digitization, Northwest University, Xi'an* 710127, *China*;
2. *College of Information Science and Technology, Northwest University, Xi'an* 710127, *China*)
* *Corresponding author, E-mail: ghgeng@nwu.edu.cn*

**Abstract**: Existing representation learning methods of cultural relics require numerous labels. Manual labeling is time-consuming and labor-intensive. Furthermore, supervised learning methods cannot effectively learn the internal structure information of point clouds. We propose an unsupervised representation learning network to extract the deep features of ceramic cultural relics. The approach is based on local-global bidirectional reasoning. First, we propose a multi-scale shell convolution-based hierarchical encoder to extract local features at different scales. Second, the local-to-global reasoning module is used to map the extracted local features to the global features. The differences between the two types of features are measured using metric learning for iterative learning. Third, a fold-based decoder is used to obtain better

**收稿日期**:2022-03-27;**修订日期**:2022-04-27.

**基金项目**:国家重点研发计划项目(No. 2019YFC1521103);国家自然科学基金重点项目(No. 61731015);陕西省重点产业链项目(No. 2019ZDLSF07-02, No. 2019ZDLGY10-01);陕西省重点研发项目(No. 2022GY-331);陕西省教育厅专项项目(No. 19JK0842);青海省重点研发与转化计划资助项目(No. 2020-SF-140)

reconstruction effects from the acquired global features in a coarse-to-fine manner. A local-to-global reasoning module supervises only the local representation to be near the global one. We propose using a low-level generation task as a self-supervision signal. The global feature can capture more basic structural information about point clouds, and the bidirectional inference between local structures and global shapes at different levels was used to learn point cloud representations. Finally, the learned representations are applied in the downstream task of point cloud classification. Experiments on the Terracotta Warriors and ModelNet40 datasets show that the proposed model significantly improves in terms of classification accuracy. The classification accuracies were 93.33% and 92.02%, respectively. The algorithm improved by approximately 4.4% and 2.82% compared with the supervised algorithm PointNet. The results demonstrate that our model achieves a comparable performance and narrows the gap between unsupervised and supervised learning approaches in downstream object classification tasks.

**Key words**: unsupervised representation learning; multi-scale strategy; deep learning; point cloud classification; virtual restoration of cultural relics

## 1 引　言

随着传统博物馆向数字博物馆的转变,文物的展示方式突破了藏品展陈的时空限制,变得丰富多样,同时稀缺文物也可以得到很好地交流与共享。作为考古史上的重大发现之一,由于自然环境和人为因素的影响,兵马俑大多以碎片的形式出土。因此,兵马俑的修复工作亟待解决。传统的人工修复方法不仅费时耗力,同时会对文物造成二次伤害。随着激光扫描仪技术的飞速发展,文物虚拟修复技术成为研究热点。作为文物修复的关键步骤,强大的碎片特征表示可大大提高碎片分类、匹配和拼接等工作的效率,对文物保护起着不可或缺的作用。

文物表征学习方法通常分为传统机器学习方法和基于深度学习的算法。传统方法一般采用专家设计的特征描述算子。Rasheed等[1]提出了依赖于RGB颜色特征和纹理特征、文物碎片之间以及基于灰度共生矩阵(Gray Level Cooccurrence Matrix,GLCM)从碎片中提取纹理特征的算法。路正杰等[2]提出了一种点云局部信息与显著性多特征描述子,并结合旋转投影特征,解决文物破损严重时分类效果差的问题。上述算法依赖于专家的先验知识,需花费大量时间,且手工设计提取特征的方法表达能力较弱,使得分类模型的泛化能力不强。近年来,随着深度学习的快速发展,一些学者将深度学习理论应用于自动驾驶[3-4]、遥感[5]、文化遗产保护[6-10]等领域,其中文物修复包括文物模型简化[6],文物碎片分类、分割[7-9]和文物碎片拼接[10]等。随着激光雷达传感器和深度传感器等三维数据采集设备的普及,点云已成为一种常见的三维模型数据,其结构简单且描述的形状信息丰富。如何将点云数据应用于上述领域,已经成为深度学习计算机视觉领域新的热点。由于点云是无序的且分布稀疏,使得将二维卷积直接应用在点云上是困难的。Charles等[11]提出的PointNet是利用深度网络直接处理点云的开创性工作,并取得了较好的效果,但该模型没有考虑局部特征。许多后续工作通过设计能够更好地获取点云的局部特征的卷积来扩展这个方向[12-17]。PointNet++[12]通过将点云划分为小的子集,构建了通过层级结构学习局部区域特征的网络。Yang等[7]提出一种基于双模态神经网络,能够较好地同时提取兵马俑碎片的空间特征和图像纹理特征。Liu等[17]提出一种基于多尺度和自注意力机制的深度神经网络,可以较好地提取兵马俑碎片点云的局部特征和全局特征。

上述有监督学习方法虽然取得一定的成绩,但往往需要大量的人工标记数据做训练,人工标注工作费时耗力[18],尤其是对于真实场景数据集。因此,无监督特征表示方法的研究具有一定

的现实意义。从无标签的数据中学习有用的表征是点云分析中一个具有挑战性的问题。现有方法大多数主要是基于生成或重建任务提供的自我监督信号，包括自我重建[19-21]和局部到整体重建[22]。Achlioptas等[19]通过多层感知器（Multi-Layer Perceptrons，MLPs）从特征表示中生成与原模型相似的结构，即 Auto-Encoder（AE）结构。但该AE模型生成的点云模型较稀疏、粗糙。Yang等[20]提出一个通过深度网格变形的点云自编码器（FoldingNet），以此来得到可以表示高维嵌入点云的特征表示，并以基于折叠的解码器替换原有的全连接解码器。Li等[21]将已有的生成对抗网络框架（Generative Adversarial Networks，GAN）扩展到处理三维点云数据，提出一个用于点云分层采样和推理网络的深度生成对抗网络（Point Cloud-GAN，PC-GAN）。该网络通过使用层次贝叶斯建模和隐性生成模型的思想来学习生成点云。Liu等[22]通过局部到整体重建方法（L2G Auto-Encoder），同时学习点云的局部和全局结构。Hassani等[23]利用聚类、自动编码和自监督分类这三个无监督任务来学习点云上的点和形状特征并取得较好结果，但网络架构复杂。上述无监督学习方法在提取点云低层次结构信息时是有效的，但通常不能或难以从点云中学习高层次的语义信息。

针对上述问题，本文通过不同抽象层次的局部表征与文物碎片对象的全局表征之间的双向推理，提出一种能够同时学习点云的结构和语义信息的无监督点云表征学习。该局部-整体双向推理模型由两部分组成：（1）局部到整体推理：首先定义预测网络，将局部特征和全局特征映射到一个共享的特征空间内，然后衡量由局部特征得到的全局特征与直接提取的全局特征之间差异，进行反复学习，使其局部特征向全局特征靠拢；（2）整体到局部推理：为保证全局特征的质量，利用自重建任务学习文物碎片对象必要结构信息的全局特征。本文算法在兵马俑数据集和Modelnet40公开数据集上的准确率分别达到93.33%和92.02%。其中在兵马俑数据集上，该算法优于除有监督模型AMS-Net（较AMS-Netoge准确率低2.35%）之外的所有方法。实验结果表明，本文无监督方法的点云表征在下游分类任务中比部分有监督表征更有鉴别力，同时缩小了无监督和监督学习方法之间的差距。作为将无监督表示学习应用于3D兵马俑数据集的一次新的尝试，该方法减少了文物数据集在收集和注释过程中所花费的大量成本，对文物虚拟修复提供一种可行的方法，具有一定意义。

## 2　兵马俑数据集制作

本实验所使用的兵马俑碎片数据集是由西北大学文化遗产数字化国家地方联合工程研究中心可视化研究所的学生使用Creaform VIU718手持式3D扫描仪采集得到的，共来自3 250块兵马俑碎片，并根据每个碎片所属部位的不同，将其分为手臂（800块）、身体（810块）、头部（810块）和腿（830块）四类。兵马俑碎片数据集示例如图1所示。对扫描得到的点云进行去噪声和下采样等一系列预处理操作。首先，使用Geomagic软件获得如图2（a）所示的原始稠密点云。其次，为减少网络的过拟合，同时提高网络预测的鲁棒性，需对稠密点云按一定比例进行下采样。使用随机采样方法，将一个碎片点云重新采样为四个不完全重叠的点云，且保证每个点云包含固定的点个数（1 024点），如图2（b）所示。图2（c）表明上述四个点云集合不完全重合。通过上述一些列操作，最终得到包含11 996块碎片的扩增数据集，将其划分为训练集和测试集两部分。其中，训练集包含10 144块碎片（Arm：2 656，Body：2 720，Head：2 272，Leg：2 496），测试集包含1 852块碎片（testArm：476，testBody：504，tes-


图1　兵马俑碎片数据集示例

Fig. 1　Examples of the Terracotta Warriors fragments datasets

tHead:428,testLeg:444)。

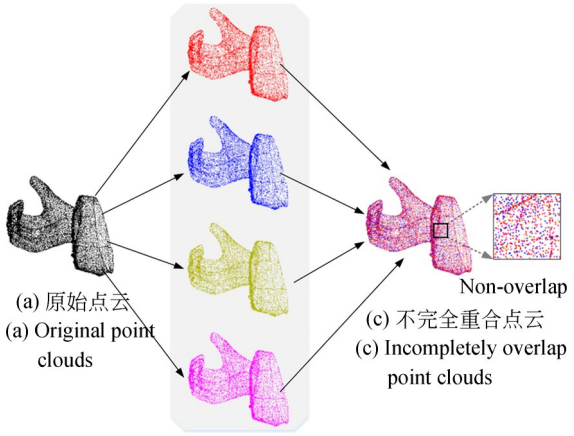



图 2　扩展数据集制作流程

Fig. 2　Approach of the extend dataset

## 3　基于局部-整体双向推理无监督表征学习方法

利用三维模型的底层语义信息和结构信息融合在对象的整体结构中这一独特属性，使得三维模型可以由单个部分推理整个对象。如图 3 所示，通过给定兔子耳部的点云，可推断出相应整体对象；整个兔子的表示也包含所有必要的细节来推断该兔子的局部结构。受上述启发，本文通过局部-整体之间的双向推理问题来实现点云表征学习。

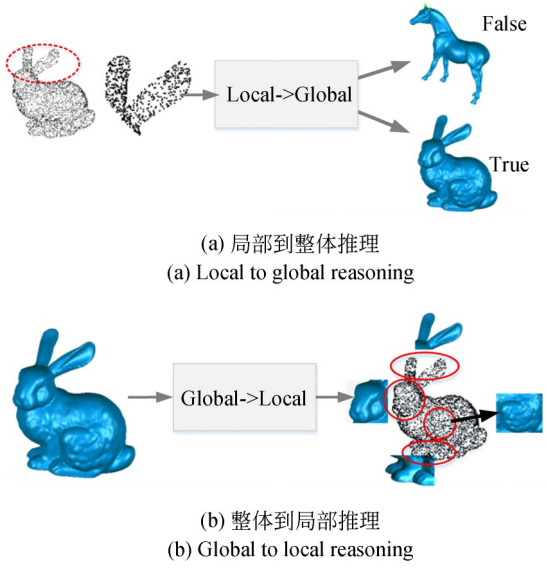



图 3　局部-整体双向推理

Fig. 3　Local-global bidirectional reasoning

本文提出的局部-整体双向推理无监督表征学习整体网络框架图如图 4 所示。首先，将点云 $P$ 输入至由两个多尺度壳模块组成的分层结构。编码器通过构建分层结构的点分组，逐步扩大感受野，最终得到第 $l$ 层局部特征 $F^l$ 和维度为 512 的全局特征 $G$。为保证全局特征 $G$ 能够获得更丰富的点云表示，本文建立由局部到整体推理模块，从而使得全局特征包含更多的结构信息和语义信息。为保证全局表征不偏离原始点云，再次提出从整体到局部的推理模块，利用折叠解码器(Folding-based Decoder)将学习到的全局特征 $G$ 重新解码为 3D 坐标，使得全局特征 $G$ 能够捕捉到更多的点云结构信息。接下来将详细介绍每个模块的组成部分。

### 3.1　多尺度壳卷积层级结构特征提取

多尺度壳卷积模块如图 5 所示。本节将从多尺度局部区域的构造、单尺度特征提取和多尺度特征融合三部分分别予以介绍。

#### 3.1.1　多尺度局部区域的构造

与 PointNet++[12] 和 ShellNet[13] 类似，该模型的多尺度局部区域构造分别由采样层、搜索层和分组层组成。首先，采样层利用最远点采样算法(Farthest Point Sampling，FPS)从输入点云 $P$ 中选择 $M$ 个点作为局部区域的质心，即 $P_c^M=\left\{p_c^m|p_c^m\in\mathbf{R}^3, m=1,2,\cdots,M\right\}$。在每个局部区域中，搜索层通过最近邻搜索算法(K-Nearest Neighbors algorithm，KNN)寻找质心点的 $k_i(i=1,2,3)$ 个最近邻点，并返回对应的点索引。其中，$k_i$ 为各个局部区域每个尺度包含点云数目，本文使用三个不同大小的尺度，故 $i$ 取值为 3。最后分组层将点云集分成多组具有不同尺度大小的局部点集，其大小为 $M\times k_i\times 3$。

#### 3.1.2　单尺度特征提取

为便于理解，接下来先介绍单尺度模块结构及特征提取，如图 5(b)所示。

本文提出一种邻域的划分方法。对于每个采样点，将其邻域划分为一组同心球体，两个同心球体之间的空隙定义为同心球壳。对于任意采样点，壳卷积算子计算局部特征的算法如下：

输入：采样点 $p_i\in P, i=1,2\cdots N$，邻点

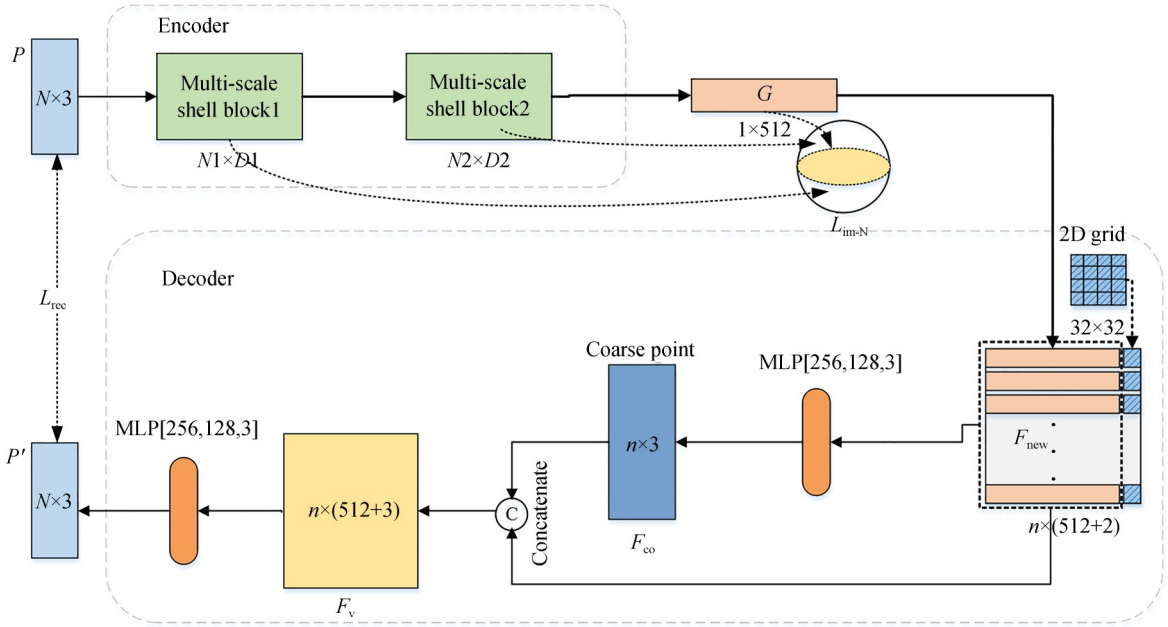



图 4　整体网络框架

Fig. 4　Overall network framework

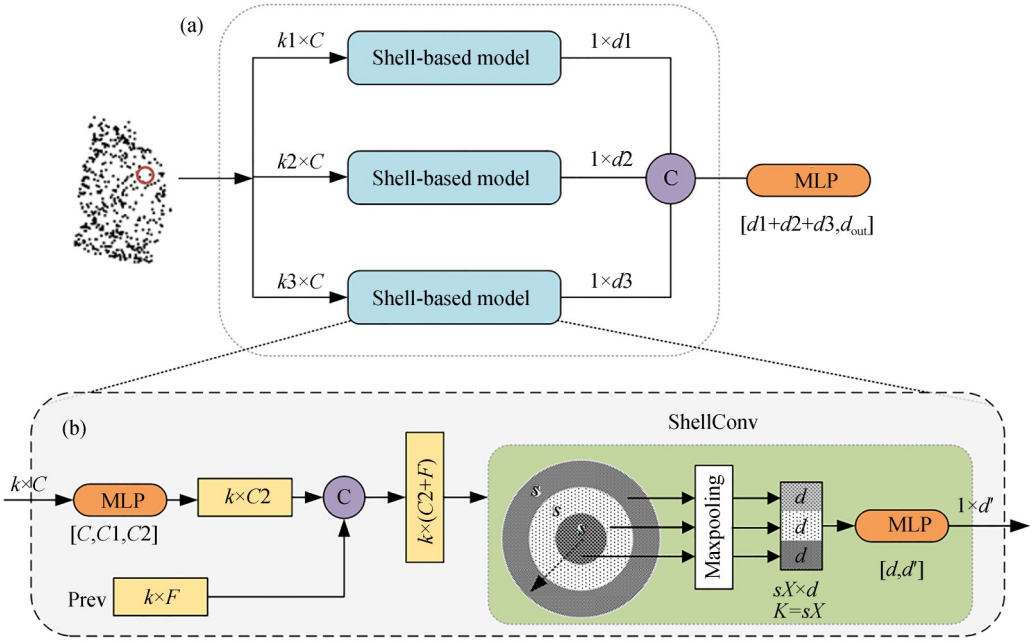



图 5　多尺度壳卷积模块

Fig. 5　Multi-scale shell convolution block

$p_j, \forall p_j \in B_p$，邻域 $B_p$，以及前一层特征 $F^{(l-1)}(p_j)$（上标 $l$ 表示第 $l$ 层）。

输出：采样点 $p_i$ 经过壳卷积的输出特征 $F_p$。

Step1：通过 KNN 搜索采样点 $p_i$ 的邻点 $p_j$，并将邻点特征维度提高到 $F^{(l)}(p_j)$。

Step2：若已存在前层特征 $F^{(l-1)}(p_j)$，则将 $F^{(l-1)}(p_j)$ 与 $F^{(l)}(p_j)$ 拼接在一起作为该点提取得到的特征 $F^{(l-1)}(p_j)$。

Step3：根据邻点 $p_j$ 到采样点 $p_i$ 的距离确定 $p_j$ 属于哪个壳，同一个壳内的点用 $X$ 表示。

Step4：通过最大池化操作获取每个壳的局部特征。

Step5：按照从内到外的顺序对所有壳的局部特征执行一维卷积，得到输出特征 $F_p$。

对于任意采样点 $p_i$ 来说，$p_i$ 点上传统的卷积为：

$$F(p_i)^{(l)}=\sum_{x\in B_p^{(l)}}W(x)^{(l)}F(x)^{(l-1)},\tag{1}$$

其中：$F(\cdot)$表示输入点的特征，$W$表示卷积的权重。每个点均需一个与之对应的权重，但点云是无序的，为每个邻点 $p_j$ 都分配卷积权重 $W(p_j)$ 是不切实际的。为解决该问题，本文将划分为同一壳内的点的特征分配相同的卷积权重。如图 5 所示。规定每个壳中的点个数是固定的，确保每个壳中包含的点数达到阈值 $s$ 后，向外扩展，直到下一个壳内点个数也达到 $s$，依此类推。假设该采样点的邻域由 $X$ 个壳组成，则邻域数 $k$ 为 $sX$，那么采样点的邻域可表示为 $M\times k\times 3$。改进后的卷积可定义为：

$$F(p_i)^{(l)}=\sum_{X\in B_p^{(l)}}W_X{}^{(l)}F(X)^{(l-1)}.\tag{2}$$

按照由内到外的顺序对壳进行排序，每个壳的权重为 $W_X$。但每个壳中的点仍然是无序的，为了产生对输入顺序不敏感的输出，通过最大池化来聚合同一个壳内的点，并使用一维卷积整合所有壳的特征以获得采样点的融合特征 $F_p$：

$$F_p=\xi\left(C\left(F(X_i)\right)\right),i=1,2,\cdots,X,\tag{3}$$

$$F(X)=\max_{p_j\in B_X}F(p_j),\tag{4}$$

其中：$C(\cdot)$为拼接操作，$\xi(\cdot)$为一维卷积操作。

3.1.3　多尺度特征融合

受文献[13]启发，本文提出一种基于多尺度壳卷积点云特征提取方法，其结构如图 5 所示。

每个单尺度模块输出的特征向量在进入下一个多尺度壳模块之前还将进行特征聚合，以形成包含不同尺度局部信息的全局特征向量。其中，三个不同尺度的特征分别记为 $F_p^{k_1}$，$F_p^{k_2}$，$F_p^{k_3}$，融合后的多尺度特征记为 $F_{\text{multi}}$：

$$F_{\text{multi}}=\zeta\left(C\left(F_p^{k_1},F_p^{k_2},F_p^{k_3}\right)\right).\tag{5}$$

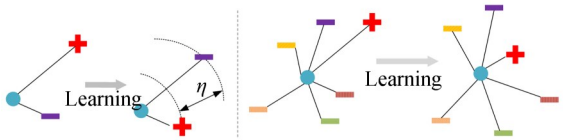
### 3.2　局部到整体推理模块

本文利用局部和整体形状之间的关系作为监督信号，用于训练丰富的表征进一步理解点云。由于全局表征通常比局部表征能更好地捕捉对象的语义信息，所以局部到目标的推理是通过预测局部的全局表征来进行的。为了评估预测结果，本文将预测视为自监督度量学习问题，并使用改进多类 $N$ 对损失（$N$-pair loss）来监督预测任务。

首先，简单回顾三元组损失函数[24]（Triplet loss）。三元组 $\{p_a,p_p,p_n\}$ 由锚点 $p_a$，正样本 $p_p$ 和负样本 $p_n$ 组成。映射函数 $f_\omega(\cdot)$ 可以将点 $p_i$ 映射到嵌入特征向量 $f_\omega(p_i)\in\mathbf{R}^d$。为方便表示，令 $\chi_a=f_\omega(p_a)$，$\chi_p=f_\omega(p_p)$，$\chi_n=f_\omega(p_n)$。Triplet loss 的目标是使同类样本的特征在空间位置上尽量靠近，不同类样本位置尽量远离，并要求 $p_a$ 到负样本 $p_n$ 的距离与 $p_a$ 到正样本 $p_p$ 的距离之差至少大于阈值 $\eta$，图 6 为 Triplet loss 与 N-pair loss 示意图，其中，“＋”代表正样本，“－”代表负样本。Triplet loss 表示形式如式 6 所示：

$$L_{\text{tri}}=\max\left(0,\left\|\chi_a-\chi_p\right\|_2^2-\left\|\chi_a-\chi_n\right\|_2^2+\eta\right),\tag{6}$$

其中，$\|\cdot\|_2^2$ 为欧式距离。




(a) 三元组损失　(b) 多类N对损失
(a) Triplet loss　(b) *N*-pair loss

图 6　Triplet loss 与 N-pair loss 示意图

Fig. 6　Illustration of Triplet loss and N-pair loss

Triplet loss 在学习参数的更新过程中，只比较了一个负样本，而忽略了其他类的负样本，故 Triplet loss 收敛速度慢。针对上述情况，$N$-pair loss[25]被提出，它是由多个负样本对组成，即一对正样本对，选取其他所有不同类别的样本作为负样本与其组合得到负样本对。如果数据集中有 C 类别，则每个正样本都对应了 $N-1$ 个负样本对。通过图 6 中两者比较，Triplet loss 可看作是 $N$-pair loss 的一个特例（$N=2$）。$N$-pair loss 的形式如式（7）所示：

$$L_{N\text{-pair}}=\frac{1}{Q}\sum_{i=1}^{Q}\left\{\log\left[1+\sum_{i\neq j}\exp\left(\chi_i^{\mathrm{T}}\chi_j'-\chi_i^{\mathrm{T}}\chi_i'\right)\right]\right\},\tag{7}$$

其中，$\chi_i=f_\omega(p_i)$和$\{p_i,p_1',p_2',\cdots,p_Q'\}$为来自$Q$个不同类别的样本对。为学习每个对象的不同语义信息，将当前对象的全局表示作为正样本，将其他类对象的全局表示作为负样本。由于局部特征$F_{pi}^{(l)}$和全局特征$G$的维度不同，无法直接衡量它们之间的相似度。故先使用$\Gamma^{(l)}(\cdot)$和$B(\cdot)$将它们分别嵌入到一个共享的特征空间中。

$$L_{\text{im-}N}=\frac{1}{Q}\sum_{i,l}\left\{\log\left[1+\sum_{G_j\neq G}\exp\left(\Gamma^{(l)}\left(F_{pi}^{(l)}\right)^{\mathrm{T}}B(G_j)-\Gamma^{(l)}\left(F_{pi}^{(l)}\right)^{\mathrm{T}}B(G)\right)\right]\right\},\tag{8}$$

其中：$\{G_j\}_{j=1}^{q}$是批量大小为$q$的批量中不同点集的全局表示，$M$是局部特征的数量。

### 3.3　整体到局部推理模块

局部到整体的推理只能监督局部表征接近全局表征，全局表征的好坏至关重要。若全局表征效果好，则会对局部表征提供较好的监督，从而为局部和全局特征的学习创造一个良性循环。反之，会导致网络学习到不可预知的结果。为避免该问题的产生，本节提出自重建辅助任务来监督网络共同学习有用的表征。自重建模块采用3.1节提出的基于多尺度壳卷积层级编码器，解码器结构图如图4中Decoder所示。首先，将编码器所得到的全局特征$G\in\mathbf{R}^{1\times512}$复制$m$次，可得$F_G\in\mathbf{R}^{m\times512}$。其次，将大小为$m\times512$的特征矩阵$F_G$与包含$m\times2$网格矩阵拼接得到$F_{\text{new}}\in\mathbf{R}^{m\times514}$，其中$m\times2$矩阵包含以原点为中心的正方形上的$m$个网格点。然后，将$F_{\text{new}}$送入MLPs得到粗输出$F_{\text{coarse}}$，其特征矩阵大小为$m\times3$。为得到更好的重建点云，最后将$m\times512$矩阵$F_G$与粗输出$F_G$拼接，再经过MLPs，得到最终重建点云。其中，网格大小$m^2$是根据输入点云大小设置的，且需满足$m^2\geqslant N$。本实验$m=32$，$N=1\,024$。

在给定输入点云$P$时，重建点云是由基于折叠解码器$D$将规范的2D网格变形到以全局表示$G$为条件的点云的3D坐标上，记为$J(G)$，其中$J(G)\in\mathbf{R}^{N\times3}$。重建误差$L_{\text{rec}}$定义为$P$和$J(G)$之间的倒角距离，故其损失函数定义为：

优化预测的一种直接方法是最小化$\Gamma^{(l)}\left(F_{pi}^{(l)}\right)$，$B(G)$之间的整体差异，即最小化$\sum_{i,l}\left\|\Gamma^{(l)}\left(F_{pi}^{(l)}\right)-B(G)\right\|_2^2$。然而，该目标可能会导致将所有输入映射到一个常数。本文使用自监督度量学习任务来监督预测的好坏。对于每个嵌入的局部表示$F_{pi}^{(l)}$，强制其特征比任何其他类别对象更接近同一对象的全局表示，故本文使用的改进$N$-pair loss表示为：

$$L_{\text{rec}}=\sum_{p\in P}\min_{p'\in J(G)}\|p-p'\|_2^2+\sum_{p'\in J(G)}\min_{p\in P}\|p'-p\|_2^2,\tag{9}$$

其中：$P$为输入点云，$p$为输入点云中的点，$J(G)$代表重建点云，$p'$是重建点云中的点。

## 4　实验结果及分析

### 4.1　实验数据集

本文分别选用兵马俑数据集和公开数据集ModelNet40作为基准数据集并完成对该模型的测试，将其实验结果与其他方法进行比较。其中，兵马俑数据集已在第2节介绍。ModelNet40包含来自40个类别的12 311个CAD模型，其中包含9 843个训练样本和2 468个测试样本。本实验中所有数据只使用点云的坐标特征$(x,y,z)$。

### 4.2　实验设置

编码器网络结构包含两层多尺度壳卷积模块，每个模块均包含三个不同尺度的壳卷积算子。其中，参数$M_j$、$s_j$和$X_i(i=0,1;j=0,1,2)$分别表示每层的采样点的个数、每层壳的大小和不同尺度中壳的数量。因此，每个采样点的邻域个数为$s_j\times X_i$。具体参数取值如表1所示。对于第一层模块来说，三个尺度邻点个数分别等于32，64，128. 同理可得，第二层模块中不同尺度的邻点个数为16，32，64。

本文使用$L_{\text{im-}N}$和$L_{\text{rec}}$联合损失为损失函数，并使用Adam优化网络，初始学习率为0.001，动

量为0.9，批量大小为22。同时使用Lambda学习率调度器，每20轮将学习率衰减0.5。

实验硬件环境为Intel Core i7处理器、16 G内存、2 T硬盘、显卡NVIDIA GTX 1080Ti；软件环境为Ubuntu16.04 x64＋CUDA9.0＋cuDNN 7.0＋PyTorch1.2＋Python3.7.

**表1　编码器网络参数**

Tab. 1　Encoder network parameters

| Layer | $M_j$ | $s_j$ | $X_i$ | $k_i$ | MLP1 | $F_{prev}$ | MLP2 out size | Shellconv out size | MLP3 Outsize |
|---|---|---|---|---|---|---|---|---|---|
| 1st | 512 | 32 | 1 | 32 | [32, 64] | — | (512,32,64) | (512,1,64) | (512,1,256) |
| | | | 2 | 64 | | | (512,64,64) | (512,1,64) | |
| | | | 4 | 128 | | | (512,128,64) | (512,1,128) | |
| 2nd | 128 | 16 | 1 | 16 | [32, 64] | 256 | (128,16,320) | (128,1,128) | (128,1,896) |
| | | | 2 | 32 | | | (128,32,320) | (128,1,256) | |
| | | | 4 | 64 | | | (128,64,320) | (128,1,512) | |

### 4.3　不同方法实验对比分析

为验证本文方法的有效性，选取一种传统点云分类方法和五种深度学习点云分类方法与之比较。目前基于兵马俑点云集的无监督表征学习方法研究甚少，上述对比方法只有Foldingnet[20]为无监督方法，其余方法均为有监督方法。

从表2可以看出，除AMS-Net方法之外（本文方法准确率低2.35%），本文方法在兵马俑数据集上取得了93.33%的最高分类准确率。表中输入数据类型字段中“P”代表点云，“G”代表图像。基于模板的兵马俑碎片传统分类算法[26]将分类问题转化为形状匹配问题，分类准确率为表2所列出方法中最低。PointNet[11]缺乏局部特征的提取，易造成细节特征的丢失，故分类精度仅为88.93%。文献[7]在PointNet基础上结合图像轮廓特征，提出一种融合点云和图像轮廓的双模态网络，相比于单独使用PointNet网络，正确率提升了2.48%。然而本文提出的无监督方法比文献[7]提高了1.92%，并远超于无监督方法Foldingnet（提高了11.42%）。另外，本文所提出模型的编码部分是在ShellNet的基础上结合了多尺度策略，最终分类准确率高于ShellNet方法0.89%。但在ModelNet40数据集上，ShellNet高于该无监督方法1.08%（见表10），由此可以表明，该方法能够针对兵马俑碎块数据集进行较好的特征表示，为兵马俑碎块分类任务提供了更加可靠的信息。

**表2　兵马俑数据集上不同方法的分类准确率**

Tab. 2　Classification of different methods on Terracotta Warrior fragments dataset.

| Method | Input data type | Deep model | Supervised | Overall accuracy |
|---|---|---|---|---|
| Template-based method[26] | P | F | T | 87.64% |
| PointNet[11] | P | T | T | 88.93% |
| Dual-modal[7] | P, G | T | T | 91.41% |
| ShellNet[13] | P | T | T | 92.44% |
| AMS-Net[17] | P | T | T | 95.68% |
| Foldingnet[20] | P | T | F | 81.91% |
| Proposed method | P | T | F | 93.33% |

为了更科学地评估本文算法的分类效果，引入精确率（$I_{Precision}$）、召回率（$I_{Recall}$）和F1值（$I_{F1}$）作为评价指标。由于数据集包含多个类别，故文本采用平均精确率（$P$）、平均召回率（$R$）、平均F1值（$F$）作为整体的评价指标，三者分别为所有类对应值的算术平均值，其计算公式如式（10）～（12）。其中，$N_{TP}$为正确判断的正样本数，$N_{FN}$为误判的正样本数，$N_{FP}$为误判的负样本数，类别个

数为 $T$。本组实验仅对表 2 代码公开的方法进行对比。从表 3 可以看出，本文算法在上述三个评价指标的结果与准确率保持一致，进一步说明该无监督方法在文物分类任务上具有一定的稳定性。

$$I_{\text{Precision}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FP}}},\ P=\frac{1}{T}\sum_{t}^{T}I_{\text{Precision}_t},\tag{10}$$

$$I_{\text{Recall}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FN}}},\ R=\frac{1}{T}\sum_{t}^{T}I_{\text{Recall}_t},\tag{11}$$

$$I_{\text{F1}}=\frac{2\times I_{\text{Precision}}\times I_{\text{Recall}}}{I_{\text{Precision}}+I_{\text{Recall}}},\ F=\frac{1}{T}\sum_{t}^{T}I_{\text{F1}_t}.\tag{12}$$

**表 3　不同模型的分类性能**

Tab. 3　Classification performance of different models

| Method | $P$ | $R$ | $F$ |
|---|---|---|---|
| ShellNet[13] | 92.55% | 92.61% | 92.49% |
| AMS-Net[17] | 95.68% | 95.71% | 95.65% |
| Foldingnet[20] | 82.66% | 81.84% | 82.29% |
| Proposed method | 93.33% | 93.34% | 93.25% |

从表 4 可以看出，表中的所有方法身体和头部两个类别的准确率要高于胳膊类和腿类。身体类的准确率最高，胳膊类的准确率最低。主要原因是身体大部分部位是衣服或盔甲，褶皱较多，特征较为明显（见图 1）。部分胳膊类的特性与腿较为相似，容易被误分类。整体上本文方法相较于文献[7]都有很大的提升。

**表 4　4 个类别中的分类精确度**

Tab. 4　Classification accuracies of the four classes

| Method | Arm | Body | Head | Leg |
|---|---|---|---|---|
| Dual-modal[7](G) | 77.75% | 92.75% | 91.50% | 76.25% |
| Dual-modal[7](P) | 82.51% | 96.45% | 92.36% | 84.41% |
| Dual-modal[7](G+P) | 87.55% | 87.55% | 94.37% | 88.41% |
| AMS-Net[17] | 92.40% | 98.10% | 98.00% | 94.20% |
| Proposed method | 84.88% | 97.08% | 94.97% | 91.76% |

### 4.4　消融实验分析

为进一步评估该网络模型中每个组成模块对网络性能的影响，在兵马俑数据集分别进行三组对比实验。

（1）编码器结构：单尺度与多尺度对比。从表 5 可以看出，多尺度模型比单尺度模型准确率高 2.40%，这表明该模型编码器可以有效地提取多个尺度的细节特征。

**表 5　编码器对分类精度的影响**

Tab. 5　Effect of the encoder on classification accuracies

| Encoder | Accuracy |
|---|---|
| Single-scale | 90.93% |
| Multi-scale | 93.33% |

（2）解码器结构：MLPs 与 Folding-based 进行对比，其中，MLPs 输出通道数为[512，256，3]。从表 6 可以看出，本文采用的折叠解码器比直接经过简单的 MLPs 准确率提升 1.88%。为进一步证明该模型的有效性，将两个网络的收敛性进行了比较，结果如图 7 所示。该网络在训练 25 轮后逐渐变得平稳，且损失值低于 MLPs 解码器。损失函数值越小，表明解码生成的点云与输入点云越相近。

**表 6　解码器对分类精度的影响**

Tab. 6　Effect of the decoder on classification accuracies

| Decoder | Accuracy |
|---|---|
| MLPs | 91.45% |
| Folding-based | 93.33% |

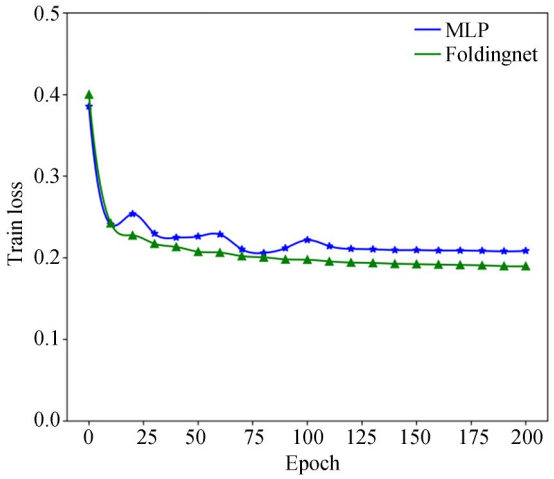



图 7　不同解码器训练损失曲线

Fig. 7　Training loss curves for different decoders

（3）损失函数：$L_{\text{rec}}$ 与 $L_{\text{im-N}}+L_{\text{rec}}$ 进行对比。从表 7 可以看出，仅通过自重建损失进行训练，得到 89.47% 的低分类准确率。当使用本文提出的

联合损失时，准确率可提升3.86%。结果表明通过度量每层局部特征与编码获取的全局特征，可以使网络进一步捕获兵马俑碎片的潜在语义特征，从而提高分类准确率。

**表7　损失函数对分类精度的影响**

Tab. 7　Effect of the loss function on classification accuracies

| loss | Accuracy |
|---|---|
| $L_{rec}$ | 89.47% |
| $L_{im-N}+L_{rec}$ | 93.33% |

### 4.5　鲁棒性

为进一步验证该模型的鲁棒性，将大小在[−1.0,1.0]的随机噪声添加到输入点云中，当一个点替换为噪声点($nn=1$)时，该模型的准确率为92.13%，其中$nn$为噪声点数量。从表8可以看出，当噪声数量高达100时，该模型仍然有很好的分类结果。表明该方法更适合真实场景数据集。

**表8　不同噪声点的结果**

Tab. 8　Results for different noise points

| Noise Number | Accuracy |
|---|---|
| 1 | 92.13% |
| 10 | 88.93% |
| 50 | 75.62% |
| 100 | 72.77% |

### 4.6　在Modelnet40公共数据集的结果分析

为评估该方法在点云表征学习方面的性能，接下来在Modelnet40公共数据集上对其进行实验测试。首先从上述预训练模型中提取ModelNet40数据集的全局特征；然后将其放入线性分类器进行训练，无需任何微调，得出分类结果。将本文方法与已有无监督表征学习方法进行比较，实验结果如表9所示，其中，$l$-GAN的实验结果为文献[22]的复现结果。本文方法只使用ModelNet40作为训练数据，而$l$-GAN[19]和FoldingNet[20]的训练数据是包含超过57 000个3D对象的ShapeNet数据集。尽管如此，本文方法仍比它们分别提高4.75%和3.62%。从定量结果可以看出，本文方法取得较好的分类性能，准确率为92.02%，分别比L2G和Multi-Task方法高1.38%和2.92%。

**表9　与已有无监督方法在ModelNet40上的对比结果**

Tab. 9　Comparisons of the classification accuracy of our method against the unsupervised learning methods on ModelNet40

| Unsupervised Method | Accuracy |
|---|---|
| FoldingNet[20] | 88.40% |
| $l$-GAN (M40)[19] | 87.27% |
| $l$-GAN[19] | 85.70% |
| Multi-Task[23] | 89.10% |
| L2G Auto-encoder[22] | 90.64% |
| Proposed method | 92.02% |

从表10中可以看出，本文方法分类准确率分别比PointNet和PointNet++提高2.82%和1.32%。即使SO-Net输入点云个数为2 048，本文方法仍相比提高1.12%。结果表明本文方法获取的特征表示比一些监督表示更具辨别力。

**表10　与已有监督方法在ModelNet40上的对比结果**

Tab. 10　Comparisons of the classification accuracy of our method against the supervised learning methods on ModelNet40

| Supervised method | Accuracy |
|---|---|
| PointNet[11] | 89.20% |
| PointNet++[12] | 90.70% |
| ShellNet[13] | 93.10% |
| SO-Net[14] | 90.90% |
| Proposed method | 92.02% |

## 5　结　论

本文提出一种无监督表征学习网络，应用于3D文物碎片分类的下游任务。首先通过层级结构的多尺度壳卷积编码器来学习点云模型不同区域之间的相关性。其次通过各层局部特征和全局特征之间的相似性度量以捕获模型的结构和语义信息。最后将学习到的点云表征应用于点云分类下游任务。实验结果表明，本文算法在兵马俑数据集和ModelNet40数据集的分类准确率分别为93.33%和92.02%。在ModelNet40

数据集上,ShellNet比该无监督方法高1.08%;而在兵马俑碎块数据集上,该方法比ShellNet高0.89%。结果表明该方法更适合于兵马俑碎块分类,同时缩小了下游分类任务中无监督和有监督学习方法之间的差距。本文尝试将点云无监督表征学习网络应用于兵马俑碎片数据集,结果也验证了该方法的有效性。接下来将继续该方法扩展到更多的文物点云分析场景中。

**参考文献:**


[1] RASHEED N A, NORDIN M J. Classification and reconstruction algorithms for the archaeological fragments[J]. *Journal of King Saud University-Computer and Information Sciences*, 2020, 32(8): 883-894.

[2] 陆正杰, 李纯辉, 耿国华, 等. 基于多特征描述子自适应权重的文物碎片分类[J]. 激光与光电子学进展, 2020, 57(4): 321-329.
LU Z J, LI C H, GENG G H, *et al*. Classification of cultural fragments based on adaptive weights of multi-feature descriptions[J]. *Laser & Optoelectronics Progress*, 2020, 57(4): 321-329. (in Chinese)

[3] CHEN X Z, MA H M, WAN J, *et al*. Multi-view 3D object detection network for autonomous driving[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA*. IEEE, 2017: 6526-6534.

[4] 陈苑锋. 视觉深度估计与点云建图研究进展[J]. 液晶与显示, 2021, 36(6): 896-911.
CHEN Y F. Progress of visual depth estimation and point cloud mapping[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(6): 896-911. (in Chinese)

[5] WANG Q, LIU S T, CHANUSSOT J, *et al*. Scene classification with recurrent attention of VHR remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(2): 1155-1167.

[6] GENG G H, LIU J, CAO X, *et al*. Simplification method for 3D Terracotta Warrior fragments based on local structure and deep neural networks[J]. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 2020, 37(11): 1711-1720.

[7] YANG K, CAO X, GENG G H, *et al*. Classification of 3D terracotta warriors fragments based on geospatial and texture information[J]. *Journal of Visualization*, 2021, 24(2): 251-259.

[8] 鱼跃华, 张海波, 李昕, 等. 基于数据增强的秦俑碎片深度分类模型[J/OL]. 激光与光电子学进展, 2021, doi:10.3788/ lop59.1810010.
YU Y H, ZHANG H B, LI X, *et al*. Data Enhanced Depth Classification Model for Terra-Cotta Warriors Fragments[J/OL]. *Laser & Optoelectronics Progress*, 2021, doi: 10.3788/ lop59.1810010. (in Chinese)

[9] GAO H J, GENG G H, ZENG S. Approach for 3D cultural relic classification based on a low-dimensional descriptor and unsupervised learning[J]. *Entropy* (*Basel, Switzerland*), 2020, 22(11): 1290.

[10] YAO W M, CHU T, TANG W L, *et al*. SPPD: a novel reassembly method for 3D terracotta warrior fragments based on fracture surface information[J]. *ISPRS International Journal of Geo-Information*, 2021, 10(8): 525.

[11] CHARLES R Q, HAO S, MO K C, *et al*. PointNet: deep learning on point sets for 3D classification and segmentation[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA*. IEEE, 2017: 77-85.

[12] QI C R, YI L, SU H, *et al*. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. *Advances in Neural Information Processing Systems*, 2017, 30.

[13] ZHANG Z Y, HUA B S, YEUNG S K. ShellNet: efficient point cloud convolutional neural networks using concentric shells statistics[C]. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *Seoul, Korea* (*South*). IEEE, 2019: 1607-1616.

[14] LI J X, CHEN B M, LEE G H. SO-net: self-organizing network for point cloud analysis[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA*. IEEE, 2018: 9397-9406.

[15] 杨军, 党吉圣. 采用深度级联卷积神经网络的三维点云识别与分割[J]. 光学 精密工程, 2020, 28(5): 1187-1199.
YANG J, DANG J S. Recognition and segmenta-

tion of three-dimensional point cloud based on deep cascade convolutional neural network [J]. *Opt. Precision Eng.*, 2020, 28(5): 1187-1199. (in Chinese)

[16] 伍锡如，薛其威. 基于激光雷达的无人驾驶系统三维车辆检测[J]. 光学 精密工程，2022，30(4)：489-497.
WU X R, XUE Q W. 3D vehicle detection for unmanned driving systerm based on lidar [J]. *Opt. Precision Eng.*, 2022, 30(4): 489-497. (in Chinese)

[17] LIU J, CAO X, ZHANG P C, *et al*. AMS-net: an attention-based multi-scale network for classification of 3D terracotta warrior fragments [J]. *Remote Sensing*, 2021, 13(18): 3713.

[18] 徐哲，耿杰，蒋雯，等. 联合训练生成对抗网络的半监督分类方法[J]. 光学 精密工程，2021，29(5)：1127-1135.
XU Z, GENG J, JIANG W, *et al*. Co-training generative adversarial networks for semi-supervised classification method [J]. *Opt. Precision Eng.*, 2021, 29(5): 1127-1135. (in Chinese)

[19] ACHLIOPTAS P, DIAMANTI O, MITLIAGKAS I, *et al*. Learning representations and generative models for 3d point clouds[C]. *Conference on Computer Vision Theory and Applications* (*VISAPP*), 27-29, 2020, *Valletta*, *MALTA*, *USA*. IEEE, 2018:421-428.

[20] YANG Y Q, FENG C, SHEN Y R, *et al*. FoldingNet: point cloud auto-encoder via deep grid deformation [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 206-215.

[21] LI C L, ZAHEER M, ZHANG Y, *et al*. Point cloud GAN [EB/OL]. *2018*: *arXiv*: 1810.05795. https://arxiv.org/abs/1810.05795

[22] LIU X H, HAN Z Z, WEN X, *et al*. L2G auto-encoder: understanding point clouds by local-to-global reconstruction with hierarchical self-attention [C]. *Proceedings of the 27th ACM International Conference on Multimedia*. *Nice France*. *New York*, *NY*, *USA*: *ACM*, 2019: 989-997.

[23] HASSANI K, HALEY M. Unsupervised multi-task feature learning on point clouds [C]. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *Seoul*, *Korea* (*South*). IEEE, 2019: 8159-8170.

[24] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition*. *Boston*, *MA*, *USA*. IEEE, 2015: 815-823.

[25] KIHYUK Sohn. Improved deep metric learning with multi-class n-pair loss objective [C]. 2016 *Conference on Neural Information Processing Systems* (*NIPS*), 5-10, 2016, *Barcelona*, *SPAIN*, 2016: 1857-1865.

[26] DU G G, ZHOU M Q, YIN C L, *et al*. Classifying fragments of terracotta warriors using template-based partial matching [J]. *Multimedia Tools and Applications*, 2018, 77(15): 19171-19191.

**作者简介：**

**刘　杰**(1989—)，女，河南新乡人，博士研究生，主要从事三维重建和智能信息处理方面的研究。E-mail: jieliu2017@126.com

**耿国华**(1955—)，女，山东莱西人，教授，博士生导师，主要从事智能信息处理、数据库与知识库、图像处理方面的研究。E-mail:ghgeng@nwu.edu.cn