



# 计算机科学

COMPUTER SCIENCE

## 基于环境信息挖掘的体素形变网络

刘娜丽, 田彦, 宋亚东, 江腾飞, 王勋, 杨柏林

### 引用本文

刘娜丽, 田彦, 宋亚东, 江腾飞, 王勋, 杨柏林. [基于环境信息挖掘的体素形变网络](#)[J]. 计算机科学, 2022, 49(10): 207-213.

LIU Na-li, TIAN Yan, SONG Ya-dong, JIANG Teng-fei, WANG Xun, YANG Bai-lin. [Voxel Deformation Network Based on Environmental Information Mining](#)[J]. Computer Science, 2022, 49(10): 207-213.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

### Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于全局增强图神经网络的序列推荐](#)

Sequence Recommendation Based on Global Enhanced Graph Neural Network

计算机科学, 2022, 49(9): 55-63. <https://doi.org/10.11896/jsjcx.210700085>

#### [基于文本行匹配的跨图文本阅读方法](#)

Cross-image Text Reading Method Based on Text Line Matching

计算机科学, 2022, 49(9): 139-145. <https://doi.org/10.11896/jsjcx.220600032>

#### [多层注意力机制融合的序列到序列中国连续手语识别和翻译](#)

Sequence-to-Sequence Chinese Continuous Sign Language Recognition and Translation with Multi-layer Attention Mechanism Fusion

计算机科学, 2022, 49(9): 155-161. <https://doi.org/10.11896/jsjcx.210800026>

#### [基于值分解的多智能体深度强化学习综述](#)

Overview of Multi-agent Deep Reinforcement Learning Based on Value Factorization

计算机科学, 2022, 49(9): 172-182. <https://doi.org/10.11896/jsjcx.210800112>

#### [基于 Key-Value 关联记忆网络的知识图谱问答方法](#)

Key-Value Relational Memory Networks for Question Answering over Knowledge Graph

计算机科学, 2022, 49(9): 202-207. <https://doi.org/10.11896/jsjcx.220300277>

# 基于环境信息挖掘的体素形变网络

刘娜丽<sup>1,3</sup> 田彦<sup>1,3</sup> 宋亚东<sup>4</sup> 江腾飞<sup>3</sup> 王勋<sup>1,2</sup> 杨柏林<sup>1</sup>

1 浙江工商大学计算机与信息工程学院 杭州 310018

2 之江实验室 杭州 311121

3 先临三维科技股份有限公司研究院 杭州 310013

4 华东理工大学信息科学与工程学院 上海 200237

(2473447208@qq.com)

**摘要** 3D形变技术是计算机图形学领域的研究热点之一。当前的3D形变方法主要通过聚合局部相邻的体素特征来学习物体形变前后的变化,未充分挖掘非局部体素特征之间的相互关系,这种环境信息的缺失导致模型无法捕获更具辨识性的特征。针对上述问题,设计了一种基于环境信息挖掘的体素形变网络,该网络能够同时对局部和环境信息进行提取,从不同的空间域中挖掘环境信息以提升网络的表征性能,进而建模物体形变前后的变化关系。引入自注意力机制,通过学习特征空间中不同体素的非局部依赖性,以提升体素特征的辨别力;引入一种多尺度分析方法,使用不同扩张率的空洞卷积分别提取不同感知域中的环境信息,为模型提供了更丰富的上下文特征。此外,文中分析了特征融合对模型的影响,并设计了一种基于编码器-解码器特征融合方法,自适应地对编码器和解码器提取的特征进行融合,提高了模型的非线性映射能力。在自建的齿科数据集上进行了充分的对比实验,结果表明,与现有方法相比,所提方法在形变预测任务的准确率上有一定的提升。

**关键词:** 形状变形;体素;注意力机制;特征融合;多尺度分析

**中图法分类号** TP391

## Voxel Deformation Network Based on Environmental Information Mining

LIU Na-li<sup>1,3</sup>, TIAN Yan<sup>1,3</sup>, SONG Ya-dong<sup>4</sup>, JIANG Teng-fei<sup>3</sup>, WANG Xun<sup>1,2</sup> and YANG Bai-lin<sup>1</sup>

1 School of Computer Science & Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China

2 Zhejiang Lab, Hangzhou 311121, China

3 Shining 3D Research, Shining 3D Tech Co., Ltd, Hangzhou 310013, China

4 School of Information Science Engineering, East China University of Science and Technology, Shanghai 200237, China

**Abstract** The technique of 3D deformation is one of the hot topics in the field of computer graphics. Current 3D deformation methods mainly learn the changes before and after deformation by aggregating localized adjacent voxel features, and fail to exploit the interrelationship between non-local voxel features, and the absence of contextual information prevents the model from capturing more discriminative features. To address the above problems, this paper designs a voxel deformation network based on environmental information mining, which can extract local and environmental information simultaneously, and extract environmental information from different spatial domains to improve the representation performance of the network, further modeling the relationship before and after the deformation of the object. Firstly, a novel self-attention mechanism is introduced. Specifically, the learning of the non-local dependence of different voxels is proposed to improve the ability of voxel discrimination. Then, a multi-scale analysis method is introduced to extract environmental information in different perceptual fields via multiple dilated convolution with different dilation rates, which provides more informative contextual features for the subsequent models. In addition, this paper analyzes the impact of feature fusion on the model and designs a method based on encoder-decoder feature fusion, which adaptively fuses the features extracted from the encoder and decoder to improve the nonlinear mapping capability of the model.

到稿日期:2021-09-08 返修日期:2022-01-16

基金项目:国家重点研发计划(2018YFB1404102,2018YFB1403200);国家自然科学基金(61972351,61976188,61972353);浙江省自然科学基金(LY19F030005);北京航空航天大学开放课题基金(VRLAB2020B15);之江实验室资助项目(2019KD0AC02)

This work was supported by the National Key R&D Program of China(2018YFB1404102,2018YFB1403200), National Natural Science Foundation of China(61972351,61976188,61972353), National Natural Science Foundation of Zhejiang Province(LY19F030005), Opening Foundation of State Key Laboratory of Virtual Reality Technology and System of Beihang University, China(VRLAB2020B15) and Zhejiang Laboratory Funded Project(2019KD0AC02).

通信作者:田彦(tianyan@zjgsu.edu.cn)

Extensive experiments are conducted on our tooth dataset. The results show that the deformation prediction accuracy of the proposed method is improved compared to existing methods.

**Keywords** Shape deformation, Voxel, Attention mechanism, Feature fusion, Multiscale analysis

## 1 引言

3D形变技术是计算机图形学与几何计算等领域的研究热点之一,被广泛应用于互动娱乐、虚拟现实、影视特效、智慧医疗等领域中,具有重要的研究意义和应用价值。

近年来,深度学习在3D形变领域的发展日渐成熟,应用也越来越广泛。传统的3D形变方法使用优化函数优化源模型与目标模型上对应关键点的偏移量来得到3D形变的预测结果。然而,这些方法需要源模型和目标模型具有相同的拓扑,因此存在一定的局限性。基于数据驱动的3D形变卷积神经网络<sup>[1-7]</sup>自动地学习源模型和目标模型之间的形变映射。例如,Wang等<sup>[1]</sup>使用PointNet<sup>[8]</sup>网络将源模型和目标模型编码到隐空间,通过在隐空间计算源模型到目标模型中顶点的位移向量来实现形变预测。然而,基于PointNet网络的特征提取方式对局部特征和全局特征进行独立分析,忽略了特征之间的相关性。Yumer等<sup>[2]</sup>对形变前源模型和变形指示符(Deformation Indicator)进行编码,学习源模型到真实形变结果目标模型的形变特征,从而实现了对源模型的形变预测,但是该方法未考虑环境信息对形变预测的影响。对于形变任务而言,如果仅使用局部特征描述牙齿形变前后的变化关系,而忽略牙齿上的非局部特征之间的联系,比如,牙齿上远距离的对称点以及具有依赖关系的特征点等能提供牙齿关键信息的特征,如图1(a)中的关键特征点1和2所示,这种环境信息的缺失会导致模型无法捕获更具有辨识性的信息,从而影响形变的结果。如图1(b)所示,编码器-解码器体系结构提取的特征不能充分表达邻域几何信息,造成形变预测结果不理想。

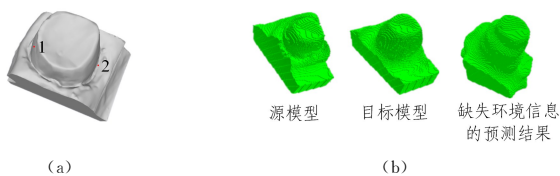


图1 形变问题图示

Fig.1 Illustrations of deformation problems

当前,在图像环境信息挖掘<sup>[9-10]</sup>方面,注意力机制(Attention Mechanism)<sup>[11]</sup>和多尺度分析<sup>[12]</sup>的应用越来越普遍。注意力机制能够提高深度学习模型对空间、特征通道等信息建模的能力,从而有效地提升卷积神经网络的表征性能;多尺度分析是研究图像不同尺度的特征以及不同尺度特征之间的联系,通过多尺度的分解,在不同分辨率上分析图像的信息,增加图像特征描述的丰富性。除此之外,深度卷积神经网络将不同阶段的特征融合,促进了不同层级特征信息的充分聚合,在保留物体位置细节的同时提升了网络的感知能力。

针对卷积神经网络仅提取局部特征,缺乏对非局部特征关系建模的问题,本文设计了一种基于环境信息挖掘的体素形变网络,该网络能够同时对局部和环境信息进行提取,通过引入环境上下文特征,增强特征的表达能力,从而提高了形变

预测结果的准确率;此外,本文设计了一种编码器-解码器特征融合方式,通过对编码器和解码器提取的特征进行融合,提高了模型的非线性映射能力,从而更好地学习形变。本文的创新性在于:

(1)提出一种基于环境信息挖掘的体素形变网络模型,该模型以V-Net<sup>[13]</sup>作为基础网络,通过引入自注意力机制,学习特征空间中不同体素的非局部依赖性,提升了体素特征的辨别力;

(2)引入空洞空间卷积池化金字塔(Atrous Spatial Pyramid Pooling, ASPP),使用不同扩张率空洞卷积提取不同感知域中的环境信息,为模型提供了丰富的上下文特征,提高了模型的预测能力;

(3)设计了一种编码器-解码器特征融合方式,自适应地对编码器和解码器提取的特征进行融合,有效地提高了模型的非线性映射学习能力,增强了模型的鲁棒性。

## 2 相关工作

为了解决3D形变问题,Noh等<sup>[14]</sup>提出在源模型与目标模型的顶点之间建立对应关系,通过网络学习顶点位置的变化关系,从而实现形变预测。Sumner等<sup>[3]</sup>提出学习源模型到目标模型的三角仿射变换矩阵,使用仿射矩阵对源模型进行预测。Chu等<sup>[4]</sup>在网络中引入一组3D实例模型,使用实例模型对形变网络预测的结果进行约束。但是,上述方法需要源模型和目标模型之间具有连通性和相同的拓扑结构。为了解决该问题,Wang等<sup>[1,15-16]</sup>对源模型和目标模型进行编码,在隐空间学习源模型与目标模型的映射关系。Wang等<sup>[1]</sup>使用PointNet网络作为自编码器,将源模型和目标模型编码到隐空间。Gao等<sup>[16]</sup>使用传统的卷积网络对源模型和目标模型进行编码,通过在隐空间学习源模型和目标模型的双向映射,建立3D模型之间的变换关系。Yumer等<sup>[2]</sup>使用传统的卷积网络对源模型和形变指示符进行编码,学习源模型到目标模型的形变特征,从而实现了对源模型的形变预测。上述方法主要使用自编码器对源模型和目标模型的特征进行编码,通过在隐空间学习特征之间的映射关系进行形变,该方法虽然解决了拓扑对应问题,但是存在数据计算量大,网络复杂度高等问题,并且传统的自编码器仅对3D模型的局部信息进行提取,缺乏对非局部特征之间相关性的分析,无法获得有辨识性的特征。

## 3 网络设计

给定3D齿科数据的源模型和目标模型,首先分别对源模型和目标模型进行网格体素化处理,得到对应的体素化数据。然后将源模型的体素化数据作为形变网络的输入,通过本文网络进行形变预测,使网络得到的体素预测结果与目标模型的体素化数据尽可能接近。

V-Net<sup>[13]</sup>是基于图像分割网络U-Net<sup>[17]</sup>从2D到3D空间的改进版本,可以在3D数据上直接实现端到端的分割。

V-Net 在 U-Net 上引入残差结构,增进了网络的收敛性,并使用卷积层代替采样操作。V-Net 网络结构如图 2 所示,其中

$E_i, Do_i, Up_i$  分别为网络中第  $i$  个残差块、下采样操作和上采样操作。

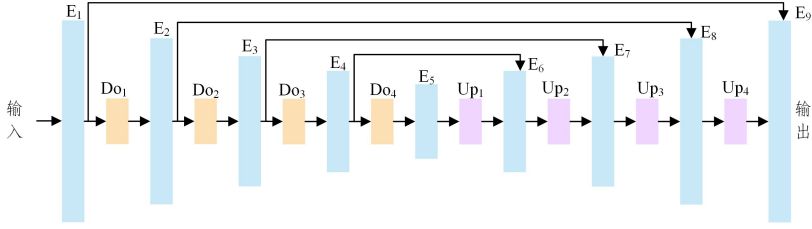


图 2 V-Net 网络结构

Fig. 2 V-Net network architecture

本文在 V-Net 网络的基础上设计了两种环境信息挖掘方法,通过挖掘特征空间中的环境信息丰富特征的表达,从而提高形变预测结果的准确性。此外,本文设计了一种基于编码器-解码器的特征融合方法,对形变任务中的非线性关系进行建模。

基于环境信息挖掘的体素形变网络结构如图 3 所示, SAM,ASPP,F<sub>i</sub>分别为本文提出的自注意力模块(Self-Attention Module,SAM)、空洞空间卷积池化金字塔(Atrous Spatial Pyramid Pooling,ASPP)和特征融合模块(Feature Fusion Module,FFM)。

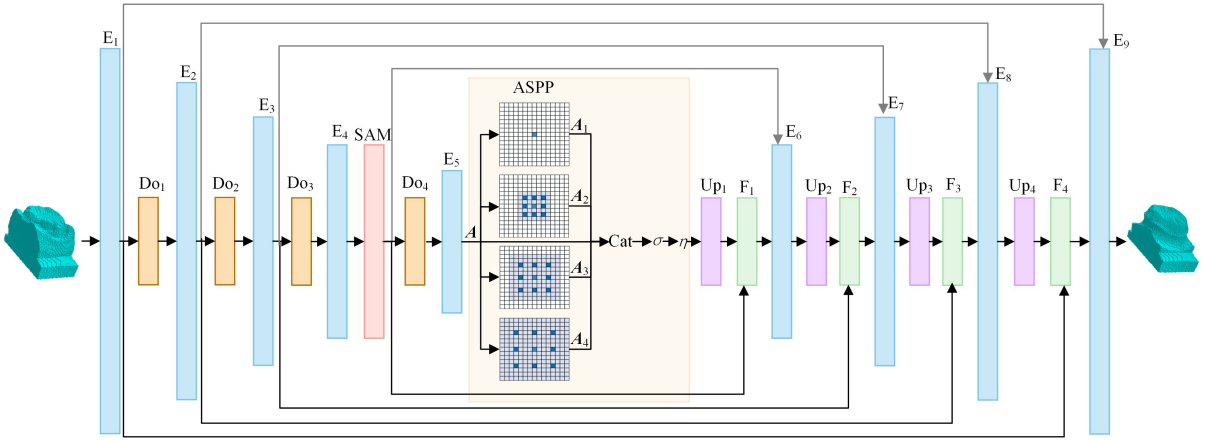


图 3 基于环境信息挖掘的体素形变网络

Fig. 3 Voxel deformation network based on environmental information mining

### 3.1 自注意力模块

传统上下文特征的提取方法仅对局部进行特征提取,忽略了非局部特征之间的依赖关系,比如远距离的对称点以及能提供关键信息的远距离特征点。这种环境信息的缺失导致模型无法捕获更具有辨识性的信息,从而影响形变预测的结果。

Wang 等<sup>[18]</sup>提出一种非局部神经网络,该网络使用自注意力机制,通过计算所有位置的特征加权和作为某个像素的响应值,在 2D 图像上建立远距离像素之间的联系,从而为网络提供有效和密集的上下文特征。

受 Wang 等<sup>[18]</sup>的启发,为了减少环境信息的缺失对形变预测网络的影响,本文引入自注意力机制对形变任务中的环境信息建模。首先,计算特征之间的相关程度,相关程度越高代表信息越重要;然后对特征的相关程度进行特征加权,并将结果作为某一特征的响应值,从而建立特征之间的关系。自注意力机制通过挖掘非局部特征之间的依赖关系,增强了特征的辨识能力,从而提高了形变预测网络的性能。

自注意力模块如图 4 所示。首先,对输入特征张量  $\mathbf{X} \in R^{C \times H \times W \times L}$  (其中  $C, H, W, L$  依次为通道数、长、宽、高)执行卷积操作进行降维,从而得到新的特征张量  $\mathbf{X}_0 \in R^{C_0 \times H \times W \times L}$  ( $C_0 = \frac{C}{8}$ )。然后,引入自注意力算法,通过 3 个并行的卷积

层分别对  $\mathbf{X}_0$  进行特征蒸馏得到相应的特征  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ 。最后,对  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$  执行点乘操作,得到非局部特征关系的环境特征。自注意力算法的实现如下:

$$S(x_i, x_j) = \frac{\exp(\mathbf{X}_1(x_i)^T \mathbf{X}_2(x_j))}{\sum_{j=1}^N \exp(\mathbf{X}_1(x_j)^T \mathbf{X}_2(x_j))} \quad (1)$$

$$Attention(i) = \sum_{j=1}^N S(x_i, x_j) \mathbf{X}_3 \quad (2)$$

其中,  $x_i$  和  $x_j$  分别表示  $\mathbf{X}_1$  中标号为  $i$  的体素和  $\mathbf{X}_2$  中标号为  $j$  的体素,对  $\mathbf{X}_1$  和  $\mathbf{X}_2$  进行加权求和得到  $S(x_i, x_j), S(x_i, x_j)$  表示  $\mathbf{X}_1$  和  $\mathbf{X}_2$  的注意力得分。最后将注意力得分  $S(x_i, x_j)$  和  $\mathbf{X}_3$  相乘,得到非局部依赖关系的环境特征。

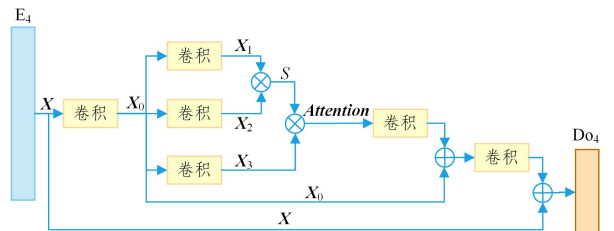


图 4 自注意力模块

Fig. 4 Self-Attention module

基于自注意力机制的环境信息挖掘方法通过建模非局部

特征之间的依赖关系,增强了特征的表达能力,从而更好地表达物体形变前和形变后的变化关系。

### 3.2 空洞空间卷积池化金字塔

多尺度分析方法的主要思想是研究不同分辨率下的对象特征及其特征之间的相互关系。通过多尺度分解对象,在不同层次结构上分析对象的特征信息,能够得到更丰富的特征描述。

针对传统上下文特征的提取方法仅关注局部特征,忽略了非局部特征的依赖关系以及特征之间的相关性,无法有效地对形变任务中的环境信息建模。本文引入空洞空间卷积池化金字塔,使用带有不同扩张率的空洞卷积在不同的感受野中建立特征之间的联系,通过融合多个尺度的特征来增强特征表达的丰富性,从而对形变任务中特征的环境信息进行建模。基于 ASPP 的环境信息挖掘模块的形变网络如图 3 所示。

ASPP 由 4 个并行的具有不同扩张率的卷积层组成,分别从不同的感受野分析输入的特征信息,然后将 4 个卷积层的输出结果  $A_1, A_2, A_3, A_4$  及其原始输入  $A$  堆叠并输入到一个串联的卷积网络中进行融合,得到更加丰富的多尺度的上下文信息。ASPP 实现的具体参数如表 1 所列。

表 1 扩张卷积的实现

Table 1 Implementation of dilated convolution

卷积核	扩张率	覆盖尺寸
$1 \times 1 \times 1$	1	$1 \times 1 \times 1$
$3 \times 3 \times 3$	2	$7 \times 7 \times 7$
$3 \times 3 \times 3$	3	$11 \times 11 \times 11$
$3 \times 3 \times 3$	4	$15 \times 15 \times 15$

### 3.3 基于编码器-解码器的特征融合模块

3D V-Net 网络使用线性连接进行学习,但是线性连接仅能表达线性映射,提取到的特征不够鲁棒,无法有效建模实际环境中非线性分布的特征,例如 3D 形变任务中物体形变前和形变后的变化关系。

假设高层语义信息与低层表征信息是相互关联的,可以采用特征融合的方式将高层语义信息与低层表征信息进行结合,从而学习形变前和形变后特征的非线性变换。由于编码器和解码器的特征具有一定的差异,直接融合特征容易产生语义鸿沟,因此,本文设计了一种特征融合方法,该方法自适应地对编码器和解码器提取的特征进行融合,从而增加了形变网络的非线性映射学习能力。

特征融合模块由一个小型的 3D 卷积神经网络实现,如图 5 所示。

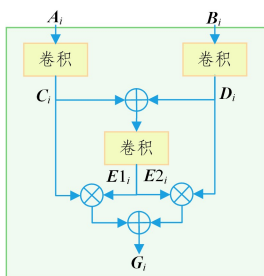


图 5 特征融合模块

Fig. 5 Feature fusion module

首先,对编码器和解码器所输入的特征张量  $A_i$  和  $B_i$  分别进行卷积操作,得到相应的映射结果  $C_i$  和  $D_i$ 。然后,对  $C_i$  和  $D_i$  进行逐元素相加并将结果输入到卷积层,得到根据特征自身属性学习的融合系数矩阵  $E_1$  和  $E_2$ 。最后,将系数矩阵  $E_1$  和  $E_2$  与映射结果  $C_i$  和  $D_i$  分别进行点乘,并对两个点乘结果进行求和,得到由编码器和解码器融合后的特征  $G_i$ 。

### 3.4 网络损失函数

本文构建的损失函数包括两项:前向传递得到的特征张量和中间层监督信号之间的相似性误差,网络预测的物体空间区域和真实形变的物体空间区域计算的相似性误差。

$$L_f(out_i, mask) = 1 - \text{mean} \left( \frac{2 \times out_i \times mask}{out_i + mask} \right) \quad (3)$$

$$L_p(pred, mask) = 1 - \text{mean} \left( \frac{2 \times pred \times mask}{pred + mask} \right) \quad (4)$$

其中,  $out_i$  是对前向传递第  $i$  层解码器的输出特征张量进行处理得到的数据(其中,  $i = [1, 2, 3, 4]$ ,  $out_i$  分别对应图 3 中的 ASPP,  $E_6, E_7$  和  $E_8$ ),  $pred$  表示网络预测的形变结果,  $mask$  表示真实的形变结果,  $\text{mean}$  表示均值操作。将误差项  $L_f$  和  $L_p$  分别对优化函数进行线性加权,得到网络的整体损失函数  $L_{total}$ :

$$L_{total} = \lambda \sum_{i=1}^4 L_f(out_i, mask) + L_p(pred, mask) \quad (5)$$

其中,  $\lambda$  是衡量  $L_f$  贡献量的超参数,  $\lambda = 0.25$ 。

## 4 实验结果以及分析

### 4.1 实验数据和数据的体素化

齿科数据集:本文所使用的齿科数据均来自去隐私化处理之后的口腔医院数据。首先,扫描人体口腔环境并重建,实例化原始数据的 3D 模型;然后,对实例化的原始 3D 模型进行手工处理,得到剔除牙龈后的牙齿模型,将剔除牙龈后的牙齿模型作为目标模型。本数据集包含 262 对齿科数据,训练集和验证集的选取比例为 9:1。同时,通过数据增强对训练集进行扩增。

网络的体素化处理步骤如下:

步骤 1 对 3D 牙齿模型所在空间划分网格。首先,建立涵盖物体的正方体包围盒;然后,根据空间分辨率对正方体包围盒进行  $n \times n \times n$  的划分网格,得到体素化网格,其中  $n = 128$ 。本文对空间分辨率  $n$  的不同取值构造了消融实验,实验结果以及分析见 4.4.1 节。

步骤 2 计算体素值。本文使用了二值化和截断符号距离函数(Truncated Signed Distance Function, TSDF)<sup>[19]</sup>两种方法对体素进行赋值,并通过实验比较了两种方法的性能,实验结果以及分析参考 4.4.2 节。

(1)二值化:遍历所有体素,判断每一个体素位置。当某一体素在物体内部或表面时,则该体素值为 1;否则,该体素值为 0。以在空间中  $(x, y, z)$  处的体素为例,若体素  $P_{(x, y, z)}$  在物体表面或物体内部,则该处的体素值  $V_{(x, y, z)} = 1$ ;否则,  $V_{(x, y, z)} = 0$ 。

(2)TSDF:遍历所有体素,计算每一个体素到物体表面的距离  $D$ ,根据截断距离值  $t$  对距离值  $D$  进行归一化,最后,通过归一化后的距离值确定该体素的 TSDF 值。以坐标  $(x, y, z)$  处



提高了特征的表达能力,从而对形变前和形变后物体的变化关系建模;FFM 能够有效地建模实际环境中非线性分布的特征,提高了模型的鲁棒性。

表 5 不同模块的消融实验对比

Table 5 Comparison of ablation experiments on different modules (单位: %)

模型	MIOU(齿科数据集)
V-Net <sup>[13]</sup>	95.28
V-Net+SAM	97.91
V-Net+FFM	97.97
V-Net+ASPP	98.88

#### 4.5 实验结果以及分析

为了进一步验证本文算法的有效性,本文选取 3DN<sup>[1]</sup>, V-Net<sup>[13]</sup>, U-Net<sup>[17]</sup>, LOGAN<sup>[20]</sup> 和 DispVoxNets<sup>[21]</sup> 这些相关的卷积网络来进行形变预测,通过计算网络的预测结果和真实形变结果之间的重合度来评估形变网络性能,结果如表 6 所列。

由表 6 可知,本文提出的形变网络模型在齿科数据集上的准确率均优于其他网络模型。LOGAN<sup>[20]</sup> 将源模型和目标模型编码到隐空间,通过在隐空间学习源模型到目标模型的转换进行形变,但是该方法使用无监督的方法学习形状之间的变形,缺少点与点的对应关系等强监督信息,从而导致形变结果不理想。3DN<sup>[1]</sup> 和 DispVoxNets<sup>[21]</sup> 通过学习原始数据的

3D 模型到真实形变结果中特征点的偏移量来实现形变预测,这种方法主要依赖于网络对形变前的特征和形变后特征点之间的变化关系建模,从而学习特征点的偏移量,但是传统的编码器-解码器结构的特征提取方法忽略了非局部特征之间的相关性对形变任务的影响。本文提出的基于自注意力机制的环境信息挖掘方法和基于空洞空间卷积池化金字塔的环境信息挖掘方法充分挖掘了非局部特征之间的关系,通过对环境信息建模,更准确地学习形变前和形变后特征的变化,从而提高了形变预测结果的准确性。本文算法和其他算法的形变预测结果可视化如图 7 所示,其中黑色虚线框突出显示了不同形变方法预测结果的主要差异。由图 7 可以看出,结合了环境信息的体素形变预测网络能够得到更准确的形变结果。

表 6 本文模型与其他形变网络的性能对比

Table 6 Performance comparison of the proposed model and other deformation networks (单位: %)

模型	MIOU(齿科数据集)
LOGAN <sup>[20]</sup>	29.67
3DN <sup>[1]</sup>	77.60
DispVoxNets <sup>[21]</sup>	89.00
U-Net <sup>[17]</sup>	93.50
V-Net <sup>[13]</sup>	95.28
本文模型	<b>98.88</b>

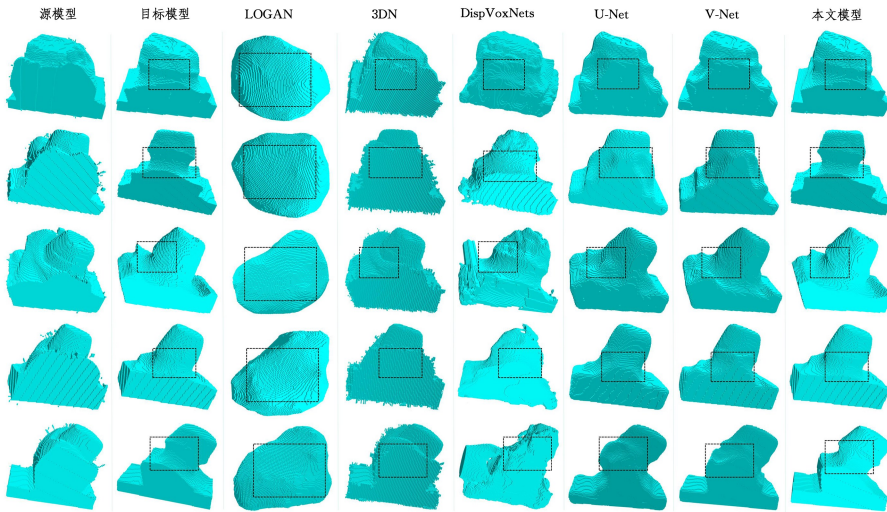


图 7 本文模型与其他形变网络的效果对比

Fig. 7 Comparison of effect of the proposed model and other deformation networks

**结束语** 针对当前的形变预测网络缺乏对环境信息的关注,本文提出在形变网络中挖掘环境信息,增强特征的辨识度,从而提高形变预测网络的性能。本文通过实验对形变网络的性能进行了验证,实验结果表明,基于环境信息挖掘的体素形变网络模型可有效提升形变预测网络的性能。鉴于 GAN 网络在对抗学习中的出色表现,在接下来的工作中将尝试使用 GAN 网络模型学习形变预测,进一步提升形变预测网络的性能。

#### 参考文献

[1] WANG W, CEYLAN D, MECH R, et al. 3DN: 3D Deformation Network[C]//Proceedings of the IEEE Conference on Compu-

ter Vision and Pattern Recognition. Long Beach: IEEE Press, 2019:1038-1046.

[2] YUMER M E, MITRA N J. Learning Semantic Deformation Flows with 3D Convolutional Networks[C]//Proceedings of the European Conference on Computer Vision. Amsterdam: Springer International Publishing Press, 2016:294-311.

[3] SUMNER R W, POPVIC J. Deformation Transfer for Triangle Meshes [J]. ACM Transactions on Graphics, 2004, 23(3): 399-405.

[4] CHU H K, LIN C H. Example-based Deformation Transfer for 3D Polygon Models[J]. Journal of Information Science and Engineering, 2010, 26(2): 379-391.

- [5] GROUEIX T, FISHER M, KIM V G, et al. 3D-CODED: 3D Correspondences by Deep Deformation[C]//Proceedings of the European Conference on Computer Vision. Munich; Springer International Publishing Press, 2018; 230-246.
- [6] YE Y T. Research of Deep Neural Network Based Face Guided Image Completion[D]. Harbin; Harbin Institute of Technology, 2018.
- [7] RU X Q, HUA G G, LI L H, et al. Handwritten Digital Recognition Based on Deformable Convolutional Neural Network[J]. *Microelectronics & Computer*, 2019, 36(4): 47-51.
- [8] QI C R, SU H, MO K, et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu; IEEE Computer Society Press, 2017; 77-85.
- [9] TIAN Y, CHEN T, CHENG G, et al. Global Context Assisted Structure-Aware Vehicle Retrieval[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 23(1): 165-174.
- [10] TIAN Y, CHENG G, GELERNTER J, et al. Joint Temporal Context exploitation and Active Learning for Video segmentation[J]. *Pattern Recognition*, 2020, 100: 107158.
- [11] TIAN Y, ZHANG Y, ZHOU D, et al. Triple Attention Network For video Segmentation[J]. *Neuro Computing*, 2020, 417: 202-211.
- [12] TIAN Y, GELERNTER J, WANG X, et al. Traffic Sign Detection Using a Multi-Scale Recurrent Attention Network[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 20(12): 565-571.
- [13] MILLETARI F, NAVAB N, AHMADI S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation[C]//Proceedings of Fourth International Conference on 3D Vision. Stanford; IEEE Computer Society, 2016; 565-571.
- [14] NOH J, NEUMANN U. Expression Cloning[C]//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York; Association for Computing Machinery Press, 2001; 277-288.
- [15] TAN Q, GAO L, LAI Y K, et al. Mesh-Based Autoencoders for Localized Deformation Component Analysis[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans; AAAI Press, 2018; 2452-2459.
- [16] GAO L, YANG J, QIAO Y L, et al. Automatic Unpaired Shape Deformation Transfer[J]. *ACM Transactions on Graphics*, 2018, 237: 1-15.
- [17] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich; Springer International Publishing Press, 2015; 234-241.
- [18] WANG X, GIRSHICK R, GUPTA A, et al. Non-local Neural Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake; IEEE Press, 2018; 7794-7803.
- [19] WERNER D, AL-HAMADI A, WERNER P. Truncated Signed Distance Function: Experiments on Voxel Size[C]//International Conference Image Analysis and Recognition. Algarve; Springer International Publishing Press, 2014; 357-364.
- [20] YIN K, CHEN Z, HUANG H, et al. LOGAN: Unpaired Shape Transform in Latent Overcomplete Space[J]. *ACM Transactions on Graphics*, 2019, 38(6): 1-13.
- [21] SHIMADA S, GOLYANIK V, TRETSCHEK E, et al. Dispvoxnets: Non-Rigid Point Set Alignment with Supervised Learning Proxies[C]//2019 International Conference on 3D Vision. Canada; IEEE Press, 2019; 27-36.



**LIU Na-li**, born in 1996, postgraduate, is a member of China Computer Federation. Her main research interest include computer vision and deep learning.



**TIAN Yan**, born in 1982, Ph.D, associate professor, master supervisor, is a member of China Computer Federation. His main research interests include machine learning and video analysis.

(责任编辑:何杨)