

基于Point Transformer 方法的鱼类三维点云模型分类

胡少秋 段瑞 张东旭 鲍江辉 吕华飞 段明

CLASSIFICATION OF 3D POINT CLOUD MODELS OF FISH BASED ON POINT TRANSFORMER APPROACH

HU Shao-Qiu, DUAN Rui, ZHANG Dong-Xu, BAO Jiang-Hui, L Hua-Fei, DUAN Ming

在线阅读 View online: <https://doi.org/10.7541/2024.2024.0053>

您可能感兴趣的其他文章

Articles you may be interested in

基于帧间光流特征和改进RNN的草鱼摄食状态分类

FEEDING STATE CLASSIFICATION OF GRASS CARP BASED ON OPTICAL FLOW AND IMPROVED RNN

水生生物学报. 2022, 46(6): 914–921 <https://doi.org/10.7541/2022.2021.0358>

基于生物能量学模型的尖吻鲈精准投喂管理辅助决策系统构建

ESTABLISHMENT OF PRECISE FEEDING MANAGEMENT ASSISTANT SYSTEM OF *LATES CALCARIFER* BASED ON BIOENERGY MODEL

水生生物学报. 2021, 45(2): 237–249 <https://doi.org/10.7541/2021.2019.258>

基于RS与水动力模型的鄱阳湖草洲产卵型鱼类潜在产卵场识别

IDENTIFICATION OF POTENTIAL SPAWNING GROUNDS OF FISH SPAWNING ON GRASSLANDS IN POYANG LAKE BASED ON RS AND HYDRODYNAMIC MODEL

水生生物学报. 2023, 47(3): 389–398 <https://doi.org/10.7541/2022.2022.0398>

基于机器视觉和改进YOLOv5s的鲫病害轻量级无损检测模型

LIGHTWEIGHT NONDESTRUCTIVE DETECTION MODEL OF CRUCIAN CARP DISEASE BASED ON MACHINE VISION AND IMPROVED YOLOv5s

水生生物学报. 2024, 48(7): 1141–1148 <https://doi.org/10.7541/2024.2023.0226>

基于Ecopath模型的鄱阳湖生态系统“十年禁渔”效果评估

EVALUATION OF THE EFFECTS ON “10-YEAR FISHING BAN” IN POYANG LAKE ECOSYSTEM BASED ON ECOPATH MODEL

水生生物学报. 2024, 48(8): 1402–1413 <https://doi.org/10.7541/2024.2023.0429>

中国布氏属鱼类的分类探讨

TAXONOMIC DISCUSSION OF GENUS *EUBLEEKERIA* FROM THE COASTAL WATERS OF CHINA

水生生物学报. 2021, 45(1): 69–75 <https://doi.org/10.7541/2021.2019.156>



关注微信公众号，获得更多资讯信息

doi: 10.7541/2024.2024.0053

CSTR: 32229.14.SSSWXB.2024.0053

基于Point Transformer方法的鱼类三维点云模型分类

胡少秋^{1,2} 段瑞³ 张东旭^{2,4} 鲍江辉² 吕华飞² 段明^{1,2}

(1. 华中农业大学, 武汉 430070; 2. 中国科学院水生生物研究所, 武汉 430072; 3. 武汉中科慧景
科技有限公司, 武汉 430074; 4. 中国科学院大学, 北京 100049)

摘要: 为实现对不同鱼类的精准分类, 研究共采集110尾真实鱼类的三维模型, 对获取的3D模型进行基于预处理、旋转增强和下采样等操作后, 获取了1650尾实验样本。然后基于Point Transformer网络和2个三维分类的对比网络进行数据集的分类训练和验证。结果表明, 利用本实验的目标方法Point Transformer获得了比2个对比网络更好的分类结果, 整体的分类准确率能够达到91.9%。同时对所使用的三维分类网络进行有效性评估, 3个模型对于5种真实鱼类模型分类是有意义的, 其中Point Transformer的模型ROC曲线准确率最高, AUC面积最大, 对于三维鱼类数据集的分类最为有效。研究提供了一种可以实现对鱼类三维模型进行精准分类的方法, 为以后的智能化渔业资源监测提供一种新的技术手段。

关键词: 点云处理; Point Transformer; 三维模型; 鱼类分类

中图分类号: Q-31; S932.4 **文献标识码:** A **文章编号:** 1000-3207(2025)02-022515-10



鱼类是我国重要的蛋白质来源之一, 在世界各地的粮食安全和营养战略中发挥着关键作用^[1]。因此作为重要渔业资源的水生态系统受到了高度关注, 对渔业资源进行监测的重要性不仅在于了解鱼类的种群动态、空间分布格局, 还会影响水生态系统的调控策略, 例如提供需要进行保护及恢复的渔业区域信息, 保证渔业资源的健康发展, 以供人们进行后续的渔业捕捞及保护^[2,3]。

目前水下监测鱼类的主要手段是利用水下摄像机来获取鱼类的视频数据, 并对监测到的鱼类进行分类识别, 基于视频的方法不仅能够克服人工调查的局限性(深度、时间)还能够提供一个永久存在且能够反复分析的数据^[4]。但是该方法与传统的鱼类分类方法一样, 也是通过对鱼体的大小、形状和颜色等视觉特征进行区分^[5], 分类结果的准确性主要依靠鱼类学专家及该专业的受培训人员的经验, 这仍是一项要求很高且耗时费力的工作^[4], 因此, 快

速、精准的对获取的鱼类数据进行分类具有重要意义。

深度学习模型具有很强的学习能力, 能够从环境变化和变化不敏感的物种独特的视觉特征中获取更深层次的特征信息, 因此已广泛应用于水生动物的物种分类。目前多个网络模型已经在公共数据集 LifeCLEF14 和 LifeCLEF15^[6-8]得到验证, 分类准确率达到90%以上^[9]。大多数学者倾向于研究二维图像的鱼类分类, 主要是因为公开的鱼类图像数据集丰富^[9-12], 且二维图像数据更易获取, 数据量足够庞大^[13]。但是在遇到种类特征相似及背景信息复杂时, 效果并不理想^[1,9]。而且在不受限制的自然栖息地中对活鱼的识别分类也存在根本性的挑战。与多数的水下监测环境一样, 水下监测时难以避免由于光线的快速衰减、鱼体的非横向视图或弯曲的身体形状, 导致了許多数据的不确定性, 在这种情况下, 用于对象识别的常见图像特征通常

收稿日期: 2024-02-05; **修订日期:** 2024-04-23

基金项目: 国家重点研发计划(2022YFB3206900和2023YFD2400600); 中国科学院“中央级科学事业单位改善科研条件专项资金”科研装备项目(GSZXKYZB2023019); 中国科学院科研仪器设备研制项目(YJKYYQ20190055)资助 [Supported by the National Key R & D Programme (2022YFB3206900 and 2023YFD2400600); Scientific Research Equipment Project of “Special Fund for Improving Scientific Research Conditions of Central Scientific Institutions” of the CAS (GSZXKYZB2023019); Chinese Academy of Sciences Research Instruments and Equipment Development Project (YJKYYQ20190055)]

作者简介: 胡少秋(1998—), 男, 硕士研究生; 主要从事渔业智能技术与装备研究。E-mail: shaoqiuhu@ihb.ac.cn

通信作者: 段明, 研究员; 主要从事渔业智能技术与装备研究。E-mail: duanming@ihb.ac.cn

会出现一定程度的缺失或者遮挡,这不可避免地降低了对于鱼类的识别性能^[14, 15]。因此,迫切需要开发新监测模式来对鱼类实现更精准的识别,分类以提供准确、高效、可重复和具有成本效益的水生态系统监测结果。

随着三维模型不断应用与发展,给现有的监测方法提供了一个新的途径。三维(3D)模型作为计算机视觉和计算机图形学领域的重要发展方向之一,其应用前景正在不断增加^[16]。在现有的研究中,由于三维数据较为缺乏以及应用场景较少等限制,对于三维数据处理及应用主要基于手工制作的特征。随着三维数据的采集设备不断更新,各种LiDAR类型、3D扫描仪及RGB-D相机等设备变得更易获取,弥补了3D模型在数据量上不足的缺点^[17]。

目前有关3D模型分类的数据类型主要分为基于点云、多视图^[18]及体积^[19]等方法,其中点云数据因其可在3D空间中保留原始几何的信息,受到了多种应用场景的青睐^[20]。并且现在已经有很多三维模型应用方面的尝试集中在3D形状表示上,所提出来的方法也是成功用于多种物体的分类和检索等领域^[3, 21]。Feng等^[16]提出了一种名为MeshNet的网格神经网络,能够从网格数据中学习各个3D模型的表示,在ModelNet40数据集上^[21],获得了91.9%的准确率。Qi等^[22]提出的pointnet网络直接在点云上进行学习,并且利用多层感知器(MLP)和对称函数解决了点云数据无序性的问题,在ModelNet40公共数据集上进行了验证并取得了89.2%分类准确率^[23]。Maturana等^[19]提出了一个体积占用网络(VoxNet)以获得稳健的3D形状分类。虽然这些分类算法在合成数据集上取得了可喜的结果,但是由于合成数据集上训练的分类模型无法很好地应用到真实世界的数据集上,所以对真实世界的3D模型进行分类仍具有挑战性。

目前对真实世界生物个体的3D模型分类应用研究有很多,但是大部分只针对人体,例如人体3D模型的形态分类^[24]、3D人脸识别^[25]及人体姿势估计与动作分类^[26]等。因为在获取3D模型的难易程度上人与动物是完全不同的,首先是不同物种之间的形状差异远大于人类,例如鱼类身体表面覆盖了鳞片,兽类全身具毛发等特征。其次在认知程度上,人类是特别容易相处和合作的主体,让野生动物配合实验人员进行模型扫描是难以进行的^[27]。鱼类也是一样,目前针对真实动物个体的3D模型研究还停留在重建良好且接近真实个体3D模型层面。

目前基于除人类以外的动物三维模型的重建与应用技术已经在多个种类上取得了成功^[27-33],但

是从三维角度将深度学习应用于分类方面的研究还很少,而且目前也没有与鱼类相关的真实三维模型数据集。因为真实的三维模型数据在获取的过程中还易受到背景噪声污染和不同程度的缺失甚至遮挡^[20]。所以目前在分类方面所使用的三维数据集,其模型大多都是合成的,这也导致了众多的三维模型分类网络在真实的三维模型数据集下的分类效果也有待重新评估。

在此背景下,本研究提出了一种基于多头注意力机制的真实三维鱼类模型分类方法,基于Point Transformer算法对鱼类的三维模型进行深度学习和训练,通过对自建的鱼类三维模型数据进行预处理与验证,能够对5种鱼类的三维模型实现精准分类,为未来的渔业资源的监测与保护开拓了一种新的途径。

1 材料与方法

1.1 实验对象

本研究使用的鱼类三维模型均采自中国安徽省巢湖市巢湖流域的鱼类,模型采集的时间为2021年2月14日—2023年6月10日。从所采样本中筛选了5种共110尾鱼类三维模型用于分类研究。

1.2 鱼类三维点云数据的采集

本研究所采用的鱼体扫描仪器是武汉中观科技公司的智能式蓝色激光三维扫描仪,型号为RigelScan。为了方便本研究更好的进行,我们自行设计了一个便捷的鱼类三维模型采集装置(图1),该装置主要由载鱼设施及操作台两个部分组成,其中平台长1.1 m,宽90 cm,载鱼设施靠滑轮来移动,高1.5 cm。模型获取的操作流程:首先是选择体表完好,无损伤的鱼固定在载鱼设施上,并在鱼体表现均匀贴上扫描贴片,然后使用三维激光扫描仪对扫描鱼体进行自上而下360度的全方位扫描,同时在扫描期间要控制扫描仪器鱼体之间的距离保持在30 cm左右,在扫描完毕后,利用扫描仪自带的ZGscan 2.0软件,将获取的三维鱼类点云进行保存并导出。利用自研的采集平台能够有效减少鱼体受到干扰,提高了模型获取的效率。

1.3 鱼类三维模型的预处理及数据集的构建

在本研究中,所有的模型在使用前会人为进行标注。因为目前得到的模型是一种无序的、离散的三维数据类型,通常包括三维坐标(X, Y, Z)及其他属性(如颜色、法向量等),并且各个种鱼类的三维模型因为体长、体重以及体宽不同,所以组成模型的点和面在数量上也存在差异。因此,在进行鱼类三维模型分类实验之前,需要对点云数据进行预

处理,主要包括数据降噪、降采样和数据增强等步骤。对模型进行处理之后,按种类汇总成一个真实的鱼类三维模型数据集(图2)。

数据降噪 在三维鱼类模型的获取过程中,会导致扫描仪器将不属于鱼体部分的点云数据也扫入三维模型中,所以获取的鱼类点云数据常常存在噪点、离群点和重复点等问题。因此利用软件对数据清洗等步骤可以去除这些干扰因素,提高分类算法的鲁棒性和准确性,本实验对于模型的预处理主要包括两个步骤,首先通过利用扫描仪器自带的ZGscan软件能够对获取的模型进行检查,并去除扫描过程中与鱼体无关的杂点;其次在制作鱼类三维模型数据集的过程中,我们将所有模型依次导入MeshLab软件,通过软件的模型可视化功能对鱼类点云数据进行二次杂点去除,以消除因其他区域的点导致实验模型结果不精确的问题(图3)。

降采样 在获取的鱼类三维模型中,由于点

云数量庞大,模型复杂度高,导致在训练过程中消耗的计算资源较多,速度慢。虽然高度详细的模型能够保持令人信服的真实感,但是并不总是需要此类模型的全部复杂性,并且使用模型的计算成本与其复杂性直接相关^[34]。因此,本研究希望通过降采样技术来减少降低模型表面的点的数量来简化模型,这能够降低模型的存储空间、减少计算资源的消耗及提高三维鱼类模型的训练效率。

本实验我们基于二次误差度量(Quadric Error Metric)的增量式方法来简化鱼类三维模型的网格结构,利用Garland和Heckberts提出的边缘折叠原理,将一个顶点、两个面及一侧的线被迭代删除^[34](图4)。由于该简化算法是基于顶点对的迭代收缩,本文定义: $(v_i, v_j) \rightarrow \bar{v}$ 边缘折叠的顺序取决于量化点到面的距离误差。如果将一组平面与每个顶点相关联,则每个顶点的误差定义为到所有平面的距离平方和。因此该顶点所产生的误差定义为:

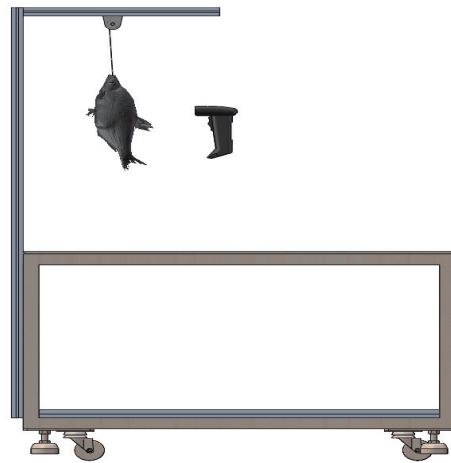
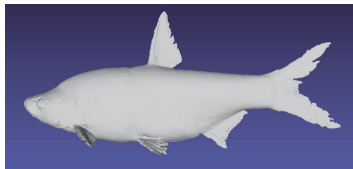


图1 鱼类三维点云数据采集装置

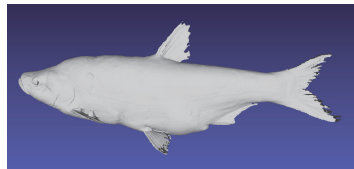
Fig. 1 Fish 3D point cloud data acquisition device

左. 现实点云数据采集装置; 右. 点云数据采集装置CAD平面图

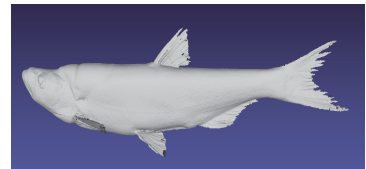
Left. realistic point cloud data acquisition device; Right. CAD plan of point cloud data acquisition device



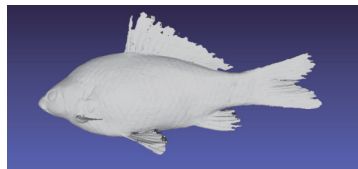
达氏鲃 *Culter dabryi*



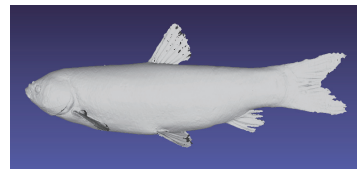
蒙古鲃 *Culter mongolicus*



翘嘴鲃 *Culter alburnus*



鲫 *Carassius auratus*



草鱼 *Ctenopharyngodon idella*

图2 鱼类三维点云数据集的制作

Fig. 2 Fish 3D point cloud dataset production

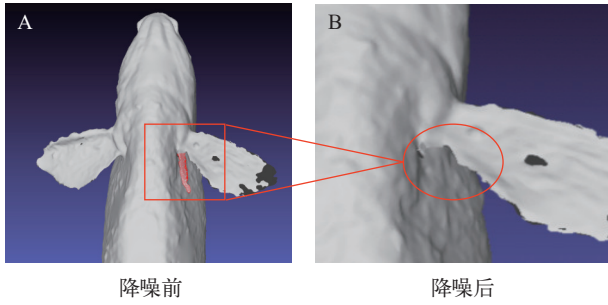


图3 鱼类三维数据清洗示例

Fig. 3 Example of fish 3D data cleaning

A图代表鱼类数据清洗之前; B图代表清洗后的模型

A represents the fish data before cleaning; B represents the model after cleaning

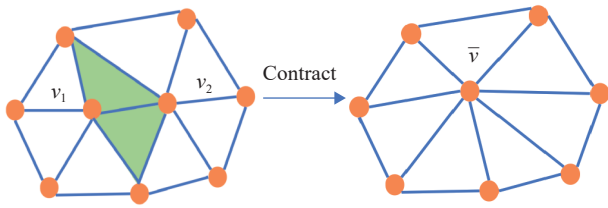


图4 边缘收缩: 突出显示的边缘收缩为单个点。阴影三角形将在收缩中并被移除

Fig. 4 Edge contraction: the highlighted edge is contracted into a single point. The shaded triangles are removed during the contraction

$$\Delta(v) = \sum_{p \in \text{planes}(v)} D^2(v) = \sum_{p \in \text{planes}(v)} v^T (K_p) v$$

$$= v^T \left(\sum_{p \in \text{planes}(v)} K_p \right) v \quad (1)$$

式中, 平面 (V) 表示在该顶点相交的所有三角形。并且在收缩过程中, 所产生的收缩误差我们定义为: $\Delta(v_1) = \Delta(v_1) + \Delta(v_1)$, 其中 $\Delta(\bar{v})$ 是收缩后产生的顶点误差。

数据增强 由于本实验所使用的数据集为自建数据集, 其自身的数据量对于深度学习模型的泛化性有一定的限制。所以为了保证实验效果, 我们对已有三维模型采用旋转的增强方式进行扩增, 如图5所示本文定义了3个旋转轴axis_X、axis_Y、axis_Z, 并且设置旋转轴序列旋转顺序。然后, 对于每个旋转轴, 旋转角度为 θ , 步长为20, 其中 $\beta \in (0, 360)$ 。本文将绕每个轴的三维点旋转表示为矩阵形式, 其公式如下:

(1) 绕Z轴旋转

$$R_Z = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

(2) 绕Y轴旋转

$$R_Y = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

(3) 绕X轴旋转

$$R_X = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

模型经过增强后, 我们将旋转后的顶点和面信息导出为新的模型文件, 以此实现模型数量上的扩增。扩增后的鱼类模型数量为1650个, 按照9:1的比例划分训练集和测试集。

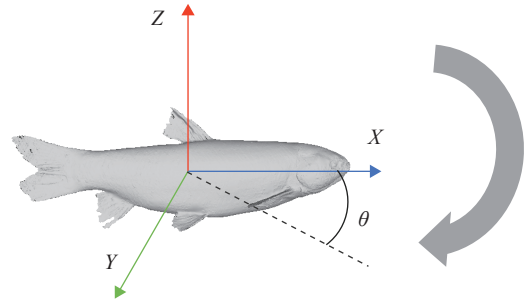


图5 鱼类三维模型数据增强过程示例

Fig. 5 Example of data enhancement process for a fish 3D model

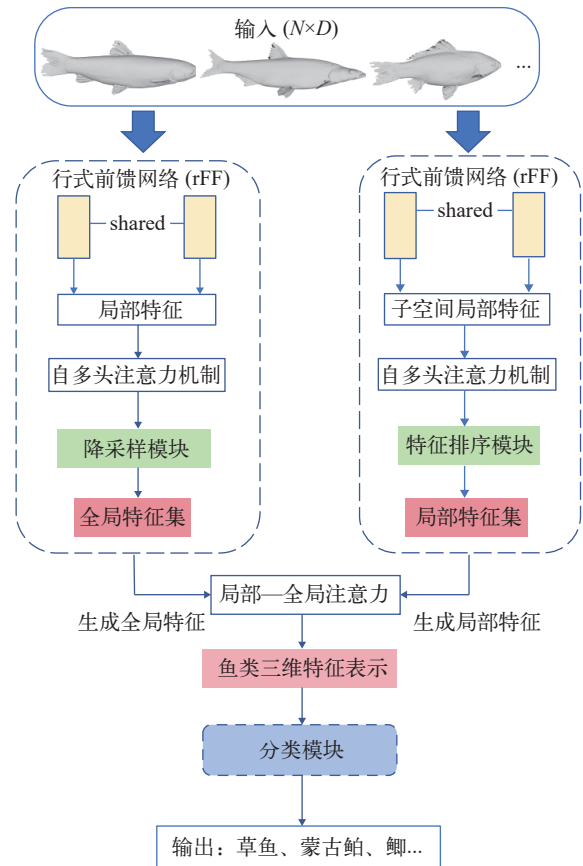


图6 基于Point Transformer的三维鱼类分类模型

Fig. 6 A 3D fish classification model based on Point Transformer

1.4 鱼类三维点云分类模型

本文主要探究Point Transformer算法对鱼类三维点云模型进行分类应用的效果。该算法利用基于多头注意力机制的特征排序模块(SortNet), 能够将目标鱼类输入的局部和全局特征联系起来, 并显示鱼类点云模型的形状信息^[35]。网络模型如图6所示, 主要分为3个部分: (1)提取不同子空间的局部特征, (2)全局特征的生成, (3)局部-全局注意力。

提取不同子空间的局部特征: SortNet作为该模型的主要分支, 通过学习不同局部特征的分数的对其进行排序, 从而获得在不同子空间内排列不变的局部特征。如图4所示, SortNet通过行式前馈网络(rFF)接收原始点云 $P \in P^{N \times D}$ 和潜在特征表示为 $P = [p_1^{\text{latent}}, \dots, p_N^{\text{latent}}] \in R^{N \times d_m}$ 。为获取每个 $p_i \in P$ 之间的空间关系, 我们利用额外的自多头注意力机制进行操作。随后又通过行式前馈网络将输入的鱼类点云特征转化为一维数据, 给输入的 P_i 创建一个可学习的标量分数, 其中 $s_i \in R$, 其中由于自多头注意层而结合了空间关系。我们现在定义一个对, 将相应的分数分配给每个输入点 $\langle p_i, s_i \rangle_{i=1}^N$ 。并且令 Q 为降序排列的一个集合 $\{S_1 > S_2 > S_3 \dots S_N\}$ 。通过在输入的鱼类最初的原始点云数据中选择 $K \leq N$, 且分值较高的点进行排序, 获得的结果我们定义为:

$$Q = \{q_j, j = 1, \dots, K\} \quad (2)$$

式中, $q_j = \langle q_j, s_i^j p_i^j \rangle_{j=1}^K$, 因为 $p_i^j \in P$ 所以 $s_i^1 \geq \dots \geq s_i^K$ 。该操作利用top-k操作来搜索 K 个最高分数 S_i 并选择相关的输入点 P_i 。在如何获取鱼类三维模型的局域性特征上, 是利用 Q_i 等^[36]所提出的球查询搜索方法, 对局部特征进行编码, 这里我们定义:

$$g^j \in R^{d_m - 1 - D}, j = 1, \dots, K \quad (3)$$

在得到编码结果后, 我们利用分组点 g^j 的特征维度将局部特征向量的结果与模型纬度 d_m 一一对应。分数 s_i^j 及来自分组层的局部特征 g^j 被连接到相应的输入点 p_i^j , 将分数的计算纳入优化问题中, 并将局部特征编码到所选点。因此, 我们将获得了局部特征向量定义:

$$f_i^j = p_i^j \oplus s_i^j \oplus g^j, f_i^j \in R^{d_m} \quad (4)$$

此外, 由SortNet 网络网络结构输出的局部特征数据则被定义为:

$$F_m^L = \{f_i^j, j = 1, \dots, K\} \quad (5)$$

此时 F_m^L 是有序的, 但是为了获取更多的子空间的局部特征, 该模型采用 M 个独立的SortNet结构,

获得由 M 个特征集组成的局部特征集, 定义为:

$$F^L = F_1^L \cup \dots \cup F_m^L, F^L \in R^{K \cdot M \cdot d_m} \quad (6)$$

全局特征的生成: 利用集合抽象多尺度分组(MSG)层的方法从鱼类三维点云数据中提取鱼体的全局特征, 首先通过最远点采样(Farthest Point Sampling, FPS)来选取整个鱼体的中心点云, 此时点云的数量为 $N' < N$ 点^[36]。接着利用相邻点获得维度为 $N' \times d_m$ 的全局特征。但是, 此时的全局特征是无序排列的, 还需要进行排序和池化操作。

局部-全局注意力: Point Transformer将局部特征集(F^L)和全局特征集(F^G)联系起来获取鱼类点云数据整体形状信息。在获取局部和全局特征集之后, 自多头注意力机制(A^{self})输出一个维度为 $N \times d_m$ 的矩阵, 但是此时的全局特征还是无序排列的, 而利用交叉多头注意力层通过对局部特征进行评分, 这样全局特征就具有了一定的顺序。鱼体三维模型的整体特征与局部特征也联系起来, 称为局部全局注意力(A^{LG})。此外, 本方法通过将 A^{LG} 多头注意力机制中的最后一个行前馈层亏的特征纬度降低至 $d'_m < d_m$, 以此降低模型鱼类计算过程的复杂程度。因此, 本文对 A^{LG} 定义为:

$$A^{LG} : R^{K \cdot M \cdot d_m}, R^{N' \times d_m} \rightarrow R^{K \cdot M \cdot d'_m} \quad (7)$$

此外, 在评估模型在训练过程中受到的损失时, 我们选择使用交叉熵损失函数作为分类任务的损失函数。计算公式定义为:

$$L_{CE} = - \sum_{i=1}^m \sum_{j=1}^n P(y_{ij}) \log_2 [Q(y_{ij})] \quad (8)$$

式中, L_{CE} 为交叉熵损失函数, n 为分类的类别数。 $P(y_{ij})$ 代表 y_i 类对应的真实标签, 也就是 n 个类别的预测概率, $Q(y_{ij})$ 代表 y_i 类的预测值。

1.5 评价指标

将划分为训练集和测试集的鱼类三维点云模型输入训练完成的分类模型后, 并输出所有种类鱼类的分类结果。实验选择总体样本分类准确率(Over accuracy)、平均准确率(Average accuracy)、单个类别的准确率(Accuracy)和F1-score对模型的分类效果进行评估, 同时利用模型训练过程中的ROC曲线来评估本实验分类模型性能, 计算公式如下:

$$\text{Acc} (\%) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (9)$$

$$\text{F1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (10)$$

式中, TP表示将正类预测为正类数; TN表示将负类

预测为负类数; FP将负类预测为正类数误报 (Type I error); FN将正类预测为负类数。

2 结果

2.1 实验环境

本实验采用 Windows 10作为实验环境配置的操作系统, 其中CPU为Intel (R) Xeon (R) Gold 6248 RI7-7700@3.00GHz; GPU为NVIDIA GeForce RTX 3090Ti。分类网络框架是Python语言编写并基于Pytorch建立三维模型分类网络框架。对三维模型进行处理的平台使用MeshLab2022。

2.2 实验参数设置

实验中采用“RangerVA”作为模型训练的优化器, Batchsize设置为11个。网络的初始学习率设置为0.001, Dropout丢弃率设置为0.5, 最后设置训练的总轮数为500 epoch。

2.3 分类实验

本文采用总体准确率(Overall accuracy, OA)、平均准确率(Average accuracy), F1分数作为各个分类模型的性能评价指标。通过准确性指标计算模型在整个数据集中做出正确预测的次数, 来确定不同模型对三维鱼类模型的分类效果。

(1) 鱼类三维模型分类结果

为了验证Point Transformer在处理真实复杂世界数据的表现能力, 本文在自建的鱼类数据集上进行实验, 实验中输入模型的点云个数为2048。对比实验结果见表1。目标模型Point Transformer对5种鱼类三维模型取得了91.9%的总体准确率、91.7%的平均分类精度, 整体准确率比Meshnet网络高了10.5%, 五种鱼类的平均准确率高了11.7%, 同时比Pointnet网络的整体准确率高了16.9%, 平均准确率

高了18.6%, 与其他模型进行对比中 Point Transformer模型的效果更优。

此外在利用F1-score对模型的整体性能进行评估时, Point Transformer算法的F1分数也是3种方法中最高的, 为91.9%, 表明了Point Transformer算法的优越性。

对单个类别的分类精度对比上, 如表2所示, Point Transformer模型有3种鱼的分类精度都超过了90%, 分别是草鱼、蒙古鲌和鲫。Meshnet只有1种鱼的分类准确率达到90%以上, 分别是草鱼和鲫草鱼识别率最高为87.9%, 鲫为88.7%, 最差的为蒙古鲌, 总体的识别准确率为68.7%。此外, 准确率最低的3种鱼分别为达氏鲌、蒙古鲌和翘嘴鲌, 尤其是蒙古鲌和达氏鲌, 因为都属于鲌类且表型性状相近, 所以易产生次优的结果, 最容易被混淆, 识别误差率高。

(2) ROC曲线比较与训练过程的可视化评估

本实验对鱼类3D模型进行分类的模型性能评估, 采用了正确识别概率(P_{TP})和错误识别概率(P_{FP})来衡量, 其中绘制的Roc曲线在虚线的左上方, 说明3个模型对于5种真实鱼类模型的分类是有意义的, 并且越靠近左上角的ROC曲线, 其分类的准确率就越高, 并且ROC曲线的AUC值也能够描述不同分类模型在鱼类3D模型上进行分类的能力, AUC值越大的分类器, 正确率越高。

通过对3种算法分类结果的ROC曲线及AUC值对比分析(图7), Point Transformer模型的ROC曲线最靠近左上角, 且AUC值最高为0.96, 所以具有更好的分类性能和鲁棒性。Meshnet和Pointnet的ROC曲线虽然均靠近左上角并且高度接近, 但是在AUC值上 Meshnet为0.89>Pointnet为0.84, 所以 Meshnet

表 1 在鱼类数据集上的实验结果

Tab. 1 Experimental results on fish dataset (%)

算法 Algorithm	输入 Import	模型点数 Model points	模型个数 Number of models	总体分类准确率 OA (%)	平均准确率 m-Acc (%)	F1分数 F1 score
Meshnet	面片	2048	550	81.4	73.1	82.7
Pointnet	点坐标	2048	550	75	80	76.4
Point transformer	点坐标	2048	550	91.9	91.7	91.9

表 2 在五种鱼类三维模型上的实验结果

Tab. 2 Classification results on a three-dimensional model of five fish species (%)

算法 Algorithm	草鱼 <i>Ctenopharyngodon idella</i>	达氏鲌 <i>Culter dabryi</i>	蒙古鲌 <i>Culter mongolicus</i>	翘嘴鲌 <i>Culter alburnus</i>	鲫 <i>Carassius auratus</i>
Meshnet	95	74.4	85.4	60	85
Pointnet	71.1	89.8	27.8	88.9	87.9
Point transformer	97.6	88.8	92.9	85.7	93.3
over-accuracy	87.9	84.3	68.7	78.2	88.7

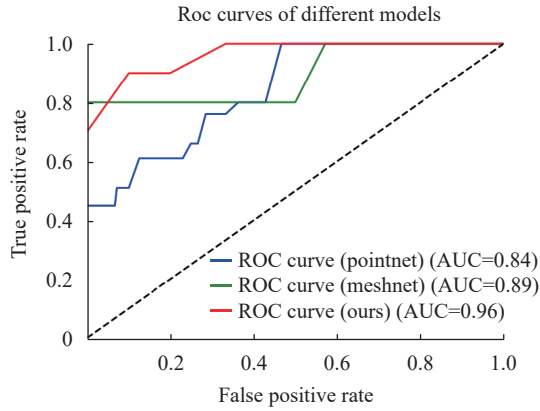


图7 利用ROC曲线和AUC (曲线下面积)评估分类模型性能

Fig. 7 Assessing classification model performance using ROC curves and AUC (area under the curve)

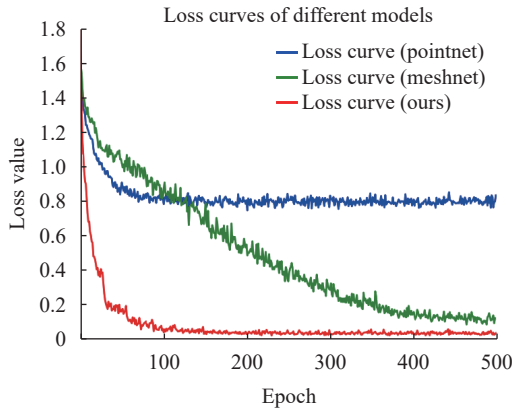


图8 利用loss曲线可视化不同分类模型的训练过程

Fig. 8 Visualisation of the training process of different classification models using loss curves

模型的分类效果次于Point Transformer, 且Pointnet模型性能最差。

图8可视化了不同分类模型对3维鱼类数据集进行分类的准确率损失情况, 可以看出随着迭代次数的增加, 模型分类准确率的loss值呈现: Point Transformer<Meshnet<Pointnet。其中Point Transformer模型在0—100次的迭代训练次数内就能够有效的降低分类准确率的loss值, 获得更好且更快的分类效果。Meshnet和Pointnet在迭代训练过程也是随着迭代次数的增加, 分类准确率的loss曲线逐渐降低, 其中meshnet所需要的训练次数明显高于Point Transformer和Pointnet模型, 且loss值高于最低的Point Transformer, 说明在模型的分类效率上Point Transformer是优于Meshnet。虽然Pointnet模型在0—100次的迭代训练次数内能够获得比较稳定的loss值, 但是其准确率的loss值要大于Meshnet和Point Transformer模型, 所以其分类效率与Point Transformer相近, 但是分类效果要比其他两种方法

要差。综上, 可视化3种方法的训练过程, 可知Point Transformer模型具有更好的分类效率和应用潜力。

3 讨论

在本实验中基于不同网络对鱼类3D模型呈现了不同程度的分类效果, 以及实验中采用的处理方法, 主要从以下几种原因进行讨论。

从鱼类3D模型识别特征上来看, 收集的鱼类模型样本中草鱼和鲫的表型特征是比较鲜明的, 且易于辨认的, 而3种鲃类在形态特征相似度较高, 这可能是3种鲃亚科鱼类分类准确率偏低的原因之一^[37]。此外, Pointnet网络模型对于翘嘴鲃的分类准确度高于蒙古鲃, 对蒙古鲃的分类准确率甚至低于30%, 这可能是由于翘嘴鲃与蒙古鲃在形态特征上可能更为接近, 虽然在摄食器官的形态上翘嘴鲃与达氏鲃也是比较相近的但是身体的其他特征上尤其是头部至背鳍区是存在差异^[37, 38]。这说明基于普通的点云特征提取的Pointnet网络难以获取相似物体的准确特征, 导致对分类个体的误判, 准确率降低。而Point Transformer基于多头注意力机制, 注重局部特征的获取, 对于相似个体具有更好的分类结果。因此后续应重点针对相似鱼类模型及其细微结构的表型数据获取和分类研究, 这提高了三维模型的获取的标准与要求。

本实验在进行模型分类之前使用了数据降采样技术, 虽然基于二次误差度量(Quadric Error Metric)的增量式方法, 能够有效解决因点云数量过多导致分类过程的分类效率过低的问题, 模型的总体占用空间过高的问题, 但是减少点云数量导致的整个鱼体表面的微结构出现的变化的程度大小, 以及对后续的总体分类准确率的影响, 我们并未探究。因此, 在后续鱼类三维模型的细粒度分类方向, 我们希望能够获得一个准确的结果, 并提出一个标准来对其进行度量, 为后续鱼类三维模型的细粒度方向分类, 提供一定的理论基础。

目前基于三维鱼类模型数据集构建的研究相对较少, 真实鱼类的3D模型获取困难且速度较慢。本实验鱼类模型样本虽然利用旋转增强进行了扩充, 但是相对于二维图像分类所要求的数据量, 目前的三维数据集规模仍较小。这在一定程度上也是制约现有鱼类3D模型分类效果的提升, 后续还需要扩大鱼类三维模型的种类及单个类别的数据量。

目前对于鱼类三维模型的获取手段, 具有速度慢、模型有噪声、耗费大量人力、自动化水平低等特点。伴随传统的各项渔业技术的革新, 后续的鱼

类似现有的智能投喂技术^[39]、养殖监测模型^[40,41]、鱼类疾病预测系统等^[42,43]。因此,将人工智能技术应用鱼类三维方面,并搭载智能仪器进行应用,不仅可以有效缩短鱼类三维模型获取的时间,还能推动鱼类三维模型数据库的建设与应用。

鱼类三维模型数据集的构建需要丰富且准确的表型特征信息,虽然目前有许多数据库例如Google等可以很好地支持和显示常见三维模型的特征信息,但是针对真实鱼类三维模型,则需要构建与鱼类相适应的特征展示模型及将鱼类三维模型数据放入存储和调用的专用库,这两个原因相互影响,对真实鱼类三维模型数据集构建和获取带来了较大的挑战。

4 结论

本文提出一种基于Point Transformer的鱼类三维点云分类模型,用于对真实的鱼类三维点云进行分类识别。该方法通过将point transformer用于真实的鱼类三维点云模型,并且通过调整其在初始阶段的预处理策略,来提高其分类性能,在自建的鱼类三维模型数据集上的实验结果表明,该方法能够获得较高的准确率,与对比网络Pointnet和Meshnet相比本方法的整体性能占优。此外我们还在数据预处理阶段,利用点云降采样的方法有效降低了鱼类三维模型的点云数量,提高了鱼类三维模型的分类的效率,并且在鱼类数据样本较少的情况下,通过利用三维模型旋转的数据增强技术来扩充本研究实验数据集,进一步提高模型分类的鲁棒性。(作者声明本文符合出版伦理要求)

参考文献:

- [1] Yang X, Zhang S, Liu J, *et al.* Deep learning for smart fish farming: applications, opportunities and challenges [J]. *Reviews in Aquaculture*, 2021, **13**(1): 66-90.
- [2] Longépé N, Hajduch G, Ardianto R, *et al.* Completing fishing monitoring with spaceborne Vessel Detection System (VDS) and Automatic Identification System (AIS) to assess illegal fishing in Indonesia [J]. *Marine Pollution Bulletin*, 2018(131): 33-39.
- [3] Chang A X, Funkhouser T, Guibas L, *et al.* Shapenet: An information-rich 3d model repository [J]. *arXiv preprint arXiv*, 2015(3012): 1512.
- [4] Villon S, Mouillot D, Chaumont M, *et al.* A deep learning method for accurate and fast identification of coral reef fishes in underwater images [J]. *Ecological Informatics*, 2018(48): 238-244.
- [5] Hu J, Li D, Duan Q, *et al.* Fish species classification by color, texture and multi-class support vector machine using computer vision [J]. *Computers and Electronics in Agriculture*, 2012(88): 133-140.
- [6] Banan A, Nasiri A, Taheri-Garavand A. Deep learning-based appearance features extraction for automated carp species identification [J]. *Aquacultural Engineering*, 2020(89): 102053.
- [7] Mathur M, Vasudev D, Sahoo S, *et al.* Crosspooled FishNet: transfer learning based fish species classification model [J]. *Multimedia Tools and Applications*, 2020(79): 31625-31643.
- [8] Ishaq O, Sadanandan S K, Wählby C. Deep fish: deep learning-based classification of zebrafish deformation for high-throughput screening [J]. *SLAS Discovery: Advancing Life Sciences R & D*, 2017, **22**(1): 102-107.
- [9] Salman A, Jalal A, Shafait F, *et al.* Fish species classification in unconstrained underwater environments based on deep learning [J]. *Limnology and Oceanography: Methods*, 2016, **14**(9): 570-585.
- [10] Xu W, Zhu Z, Ge F, *et al.* Analysis of behavior trajectory based on deep learning in ammonia environment for fish [J]. *Sensors*, 2020, **20**(16): 4425.
- [11] Li X, Shang M, Hao J, *et al.* Accelerating Fish Detection and Recognition by Sharing CNNs with Objectness Learning [C]. OCEANS 2016-Shanghai. IEEE, 2016: 1-5.
- [12] Shah S Z H, Rauf H T, IkramUllah M, *et al.* Fish-Pak: Fish species dataset from Pakistan for visual features based classification [J]. *Data in Brief*, 2019(27): 104565.
- [13] Duan Y E, Li D L, Li Z B, *et al.* Review on visual characteristic measurement research of aquatic animals based on computer vision [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, **31**(15): 1-11. [段延娥, 李道亮, 李振波, 等. 基于计算机视觉的水产动物视觉特征测量研究综述 [J]. *农业工程学报*, 2015, **31**(15): 1-11.]
- [14] Chuang M C, Hwang J N, Williams K. A feature learning and object recognition framework for underwater fish images [J]. *IEEE Transactions on Image Processing*, 2016, **25**(4): 1862-1872.
- [15] Ali-Gombe A, Elyan E, Jayne C. Fish Classification in Context of Noisy Images [C]. Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25-27, 2017, Proceedings. Springer International Publishing, 2017: 216-226.
- [16] Feng Y, Feng Y, You H, *et al.* Meshnet: Mesh Neural Network for 3d Shape Representation [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, **33**(1): 8279-8286.
- [17] Liang Z, Guo Y, Feng Y, *et al.* Stereo matching using multi-level cost volume and multi-scale feature constancy [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **43**(1): 300-315.
- [18] Su H, Maji S, Kalogerakis E, *et al.* Multi-View Convolutional Neural Network for 3D Object Classification

- tional neural Networks for 3d Shape Recognition [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 945-953.
- [19] Maturana D, Scherer S. Voxnet: A 3d Convolutional Neural Network for Real-time Object Recognition [C]. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2015: 922-928.
- [20] Guo Y, Wang H, Hu Q, *et al.* Deep learning for 3d point clouds: A survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **43**(12): 4338-4364.
- [21] Wu Z, Song S, Khosla A, *et al.* 3d Shapenets: A Deep Representation for Volumetric Shapes [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1912-1920.
- [22] Charles R Q, Hao S, 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, 2017: 652-660.
- [23] Uy M A, Pham Q H, Hua B S, *et al.* Revisiting Point Cloud Classification: A New Benchmark Dataset and Classification Model on Real-world Data [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1588-1597.
- [24] Wang J, Li X, Pan L, *et al.* Parametric 3D modeling of young women's lower bodies based on shape classification [J]. *International Journal of Industrial Ergonomics*, 2021(84): 103142.
- [25] Zhou S, Xiao S. 3D face recognition: a survey [J]. *Human-centric Computing and Information Sciences*, 2018, **8**(1): 1-27.
- [26] Presti L L, La Cascia M. 3D skeleton-based human action classification: A survey [J]. *Pattern Recognition*, 2016(53): 130-147.
- [27] Zuffi S, Kanazawa A, Jacobs D W, *et al.* 3D Menagerie: Modeling the 3D Shape and Pose of Animals [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6365-6373.
- [28] Yang G, Vo M, Neverova N, *et al.* Banmo: Building Animatable 3d Neural Models from Many Casual Videos [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 2863-2873.
- [29] Rüegg N, Zuffi S, Schindler K, *et al.* Barc: Learning to Regress 3d Dog Shape from Images by Exploiting Breed Information [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 3876-3884.
- [30] Mathis M W, Mathis A. Deep learning tools for the measurement of animal behavior in neuroscience [J]. *Current Opinion in Neurobiology*, 2020(60): 1-11.
- [31] Li C, Ghorbani N, Broomé S, *et al.* hSMAL: Detailed horse shape and pose reconstruction for motion pattern recognition [EB/OL]. arXiv preprint arXiv: 2106.10102, 2021.
- [32] Zuffi S, Kanazawa A, Black M J. Lions and Tigers and Bears: Capturing Non-rigid, 3d, Articulated Shape from Images [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3955-3963.
- [33] Labuguen R, Matsumoto J, Negrete S B, *et al.* Macaque-Pose: a novel "in the wild" macaque monkey pose dataset for markerless motion capture [J]. *Frontiers in Behavioral Neuroscience*, 2021(14): 581154.
- [34] Garland M, Heckbert P S. Surface Simplification Using Quadric Error Metrics [C]. Proceedings of the 24th annual conference on Computer graphics and interactive techniques, 1997: 209-216.
- [35] Engel N, Belagiannis V, Dietmayer K. Point transformer [J]. *IEEE Access*, 2021, **9**: 134826-134840.
- [36] 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.
- [37] Hu H, Di Y, Zhao Y, *et al.* Comparative study on the morphological characteristics of four species of culter and culterichthys in Lihu Lake [J]. *Journal of Yunnan Agricultural University (Natural Science)*, 2011, **26**(4): 488-494. [胡海彦, 狄瑜, 赵永锋, 等. 蠡湖 4 种鲌鱼形态特征的比较研究 [J]. 云南农业大学学报: 自然科学版, 2011, **26**(4): 488-494.]
- [38] Wang Y L, Li H C, He Y F, *et al.* Morphological variations of feeding organs of five species of culter from Changhu Lake [J]. *Freshwater Fisheries*, 2016, **46**(6): 26-32. [王亚龙, 李昊成, 何勇凤, 等. 长湖 5 种鲌摄食器官形态学的比较 [J]. 淡水渔业, 2016, **46**(6): 26-32.]
- [39] Chao Z, Daming X, Kai L, *et al.* Evaluation of fish feeding intensity in aquaculture based on near-infrared machine vision [J]. *Smart Agriculture*, 2019, **1**(1): 76.
- [40] Fernandes A F A, Turra E M, de Alvarenga E R, *et al.* Deep Learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia [J]. *Computers and Electronics in Agriculture*, 2020(170): 105274.
- [41] Chen L, Yang X, Sun C, *et al.* Feed intake prediction model for group fish using the MEA-BP neural network in intensive aquaculture [J]. *Information Processing in Agriculture*, 2020, **7**(2): 261-271.
- [42] Xu M, Wu Y. Fish diseases diagnosis based on rough set and neural network [J]. *Computer Engineering and Design*, 2009, **30**(7): 1738-1741.
- [43] Darapaneni N, Sreekanth S, Paduri A R, *et al.* AI Based Farm Fish Disease Detection System to Help Micro and Small Fish Farmers [C]. 2022 Interdisciplinary Research in Technology and Management (IRTM). IEEE, 2022: 1-5.

CLASSIFICATION OF 3D POINT CLOUD MODELS OF FISH BASED ON POINT TRANSFORMER APPROACH

HU Shao-Qiu^{1,2}, DUAN Rui³, ZHANG Dong-Xu^{2,4}, BAO Jiang-Hui², LÜ Hua-Fei² and DUAN Ming^{1,2}

(1. *Huazhong Agricultural University, Wuhan 430070, China*; 2. *Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China*; 3. *Huijing Technology, Co., Ltd., Wuhan 430074, China*; 4. *University of Chinese Academy of Sciences, Beijing 100049, China*)

Abstract: Phenotypic data serve as the foundation for effective monitoring of fish species. Currently, fish classification heavily relies on expertise from relevant professionals, leading to issues such as low efficiency, high errors, potential damage to fish bodies, and susceptibility to subjective factors affecting data quality. In this study, we developed a simplified device for acquiring three-dimensional models of fish, pioneering the creation of a dataset comprising authentic three-dimensional fish models. By leveraging the Point Transformer algorithm, we can rapidly, efficiently, and accurately extract phenotypic features from three-dimensional fish bodies, enabling precise classification of different fish species. A total of 110 authentic fish three-dimensional models were collected in this research, resulting in 1650 experimental samples after preprocessing, rotation enhancement, and downsampling operations on the acquired 3D models. Subsequently, through classification training and validation using the Point Transformer network and two comparative networks for three-dimensional classification, the results indicate that the proposed Point Transformer method outperforms the two comparative networks, achieving an overall classification accuracy of 91.9%. Simultaneously, an effective evaluation of the utilized three-dimensional classification networks was conducted, demonstrating the meaningful classification of the three models for five authentic fish species models. The Point Transformer model exhibited the highest ROC curve accuracy and the largest AUC area, proving its effectiveness in classifying three-dimensional fish datasets. This study presents a method for accurately classifying three-dimensional fish models, offering a new technological approach for intelligent monitoring of fisheries resources in the future.

Key words: Point Cloud Processing; Point Transformer; Three-dimensional Model; Fish Classification