水生生物学报
ACTA HYDROBIOLOGICA SINICA

一种单视图江豚三维模型重建方法

黄志勇 杨晨龙 石小涛 华喜锋 涂法宪 丁妥君 佘雅丽 向梦丽

A SINGLE-VIEW 3D MODEL RECONSTRUCTION METHOD FOR YANGTZE FINLESS PORPOISE

HUANG Zhi-Yong, YANG Chen-Long, SHI Xiao-Tao, HUA Xi-Feng, TU Fa-Xian, DING Tuo-Jun, SHE Ya-Li, XIANG Meng-Li

在线阅读 View online: https://doi.org/10.7541/2025.2024.0183

doi: 10.7541/2025.2024.0183
CSTR: 32229.14.SSSWXB.2024.0183

一种单视图江豚三维模型重建方法

黄志勇[1, 2, 3] 杨晨龙[1, 3] 石小涛[2, 4] 华喜锋[1, 3] 涂法宪[2, 4] 丁妥君[1, 3]
佘雅丽[1, 3] 向梦丽[1, 3]

(1. 三峡大学湖北省水电工程智能视觉监测重点实验室, 宜昌 443002; 2. 三峡大学湖北省鱼类过坝技术国际科技合作基地, 宜昌 443002; 3. 三峡大学计算机与信息学院, 宜昌 443002; 4. 三峡大学水利与环境学院, 宜昌 443002)

摘要: 在江豚三维重建领域, 存在水下图像色偏失真、江豚数据集不足、获取江豚多视角图像困难等问题, 而新兴方法尚未出现针对江豚的应用研究。为了解决这些难题, 文章提出了一种结合扩散模型和神经辐射场的单视图江豚三维模型重建方法。首先, 改进水下图像增强方法, 有效地解决水下图像色偏失真的问题。其次, 自制江豚多视角图像数据集, 微调视角条件扩散模型, 实现由单视图合成多视角图像, 为单张图像重建江豚提供了新思路。最后, 由神经辐射场进行重建, 得到江豚三维模型。对江豚三维重建的结果使用平均倒角距离和法向量一致性进行了对比评估, 平均倒角距离低于现有方法, 法向量一致性高于现有方法, 表明文章方法能够有效重建出符合江豚体色及形态的三维模型, 合成新视角图像PSNR、SSIM、LPIPS值分别为38.968、0.972和0.294, 效果优于现有方法, 经过水下图像增强的重建结果的平均倒角距离值最低为0.428, 法向量一致性最高达到0.882。

关键词: 扩散模型; 新视角合成; 神经辐射场; 三维重建; 长江江豚
中图分类号: Q-334 **文献标识码:** A **文章编号:** 1000-3207(2025)04-042510-11

长江江豚(Yangtze finless porpoise)是国家一级保护动物, 因江豚的生存环境受到威胁, 以及人类的过度捕捞, 导致其数量锐减, 所以对江豚的保护变得迫在眉睫[1]。随着近几年对江豚保护力度的提升, 有关部门建立江豚繁育基地[2], 江豚的种群数量逐渐回升[3], 但对其个体的监测和江豚繁育基地的高效管理成为了一个新的问题。三维重建是一类根据现实场景或二维图像建立其三维模型的技术。对长江江豚进行三维重建, 旨在以一种非入侵的方式捕捉其三维身体形态及行为动作, 对它们的行为习性、健康状况和社交行为等方面进行研究。

针对动物的单视图三维重建已有大量研究。其中, Zuffi等[4]提出用于重建四足哺乳动物的蒙皮多动物模型(Skinned Multi-Animal Linear Model, SMAL), 该模型将动物模板网格模型分割为带有混合权重的许多部分, 再通过线性混合蒙皮进行姿势变形, 能够重建出动物的不同身体姿势。Rueegg等[5]修改SMAL的形状空间, 使其更适合表示宠物犬的身体形状, 从而提出基于品种信息的回归增强分类模型(Breed-Augmented Regression using Classification, BARC), 采用直接从图像像素回归参数化的三维形状模型的方法, 能够有效重建出120个品种宠物犬的模型。然而, 这2种方法也存在可扩展性差的缺点, 难以扩展到其他形态的动物, 对于江豚来说存在着巨大的挑战性。

近年来, 随着扩散模型[6]的飞速发展, 以及涵盖物体多视图的图像数据集规模逐渐增大, 我们可以通过微调扩散模型以从物体多视图图像中学习控制拍摄视角及相机外参的能力, 从而实现新视角图像合成。微调扩散模型后预测出来的新视图具有较高的空间一致性, 可以为单视图重建提供一定的几何先验, 为解决单目重建自遮挡问题提供新的有效途径。Chan等[7]利用现有的2D扩散骨干网络, 引入以3D特征体形式的几何先验, 使得即使存在遮挡

收稿日期: 2024-05-02; **修订日期:** 2024-09-18
基金项目: 国家自然科学基金 (52279069) 资助 [Supported by the National Natural Science Foundation of China (52279069)]
作者简介: 黄志勇(1979—), 男, 博士, 副教授; 主要从事计算机视觉方面研究。E-mail: hzy@hzy.org.cn
通信作者: 石小涛(1981—), 男, 博士, 教授; 主要从事生态水利方面研究。E-mail: fishlab@163.com

或阴影, 也能够生成多样化且可信的新视图。Watson等[8]提出3DiM, 使用一种称为“随机条件”的新技术, 在每个去噪步骤中从可用视图集中随机选择一个条件视图, 这种随机条件的引入显著提高了生成视图的3D一致性。Melas-Kyriazi 等[9]提出RealFusion, 通过引入一种新的单图像文本反演变体, 能够360度合成多角度视图, 而无需对图像的对象类型或任何形式的三维监督做出假设。Liu等[10]提出Zero-1-to-3, 在大规模多视角图像数据集Objaverse[11]中学习对相机视角的控制, 实现了在指定相机变换下生成相应视角的新视图。然而, 上述几种单目重建方法都是通用的, 他们的目标是重建任意的对象, 并没有针对江豚进行特别的优化。

此外, 上述几个基于扩散模型的方法虽然取得了出色的效果, 但是仍然存在以下不足: 在实际运用中, 由于水对不同波长光线的吸收及散射, 导致在室内拍摄的水下江豚图像往往带有严重的偏色, 以及较低的对比度和亮度, 这极大影响了最后重建模型的颜色以及质量。综上, 本文从实际出发, 引入了一种水下图像增强方法, 实现对水下江豚图像的增强处理, 并针对Objaverse数据集中不包含江豚的问题, 制作了江豚多视角图像数据集, 用于训练微调的视角条件扩散模型, 以合成多视角图像, 最后利用改进的神经辐射场对其进行重建。

1 数据采集与制作

本文的目标是获取处于不同身体姿势的江豚三维模型。由于难以让江豚保持一个特定的姿势同时配合实验人员进行扫描, 故我们采用手持3D激光扫描仪扫描的江豚雕塑获得江豚点云, 再对点云进行去除孤立点、平滑等处理操作, 最后构建网格模型(图 1)。再将此模型经过骨骼绑定、蒙皮、涂抹权重等步骤制作出了江豚运动的三维动画(图 2), 以尽可能多地包含江豚在运动中的各种姿势, 将其导出为不同姿势的模型, 从而扩充江豚模型数据集。

将得到不同姿势的江豚模型分别通过Blender软件进行随机渲染, 得到12幅不同相机位姿的视图, 和12个与视图对应的相机位姿, 以及1个包含三维模型关键属性统计信息的json文件, 其内容如表 1所示, 数据集如图 3所示。本文一共制作了76组这样的数据, 构成了江豚多视图数据集。

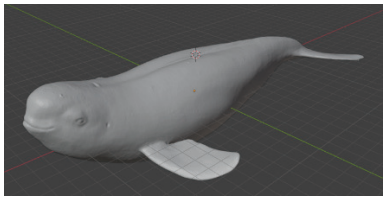

图 1 扫描获得的江豚基础模型

Fig. 1 Basic model of finless porpoise obtained by scanning

2 方法

2.1 方法框架

本文设计了一种基于视角条件扩散模型、神经辐射场重建江豚的方法, 方法框架如图 4所示。对于水下摄像机拍摄的江豚图像, 为了消除严重的蓝色偏色及去除背景对江豚重建的干扰, 首先对其进行预处理工作, 包括水下图像增强, 以及对江豚进行图像前景分割和深度估计[12]。然后, 再使用视

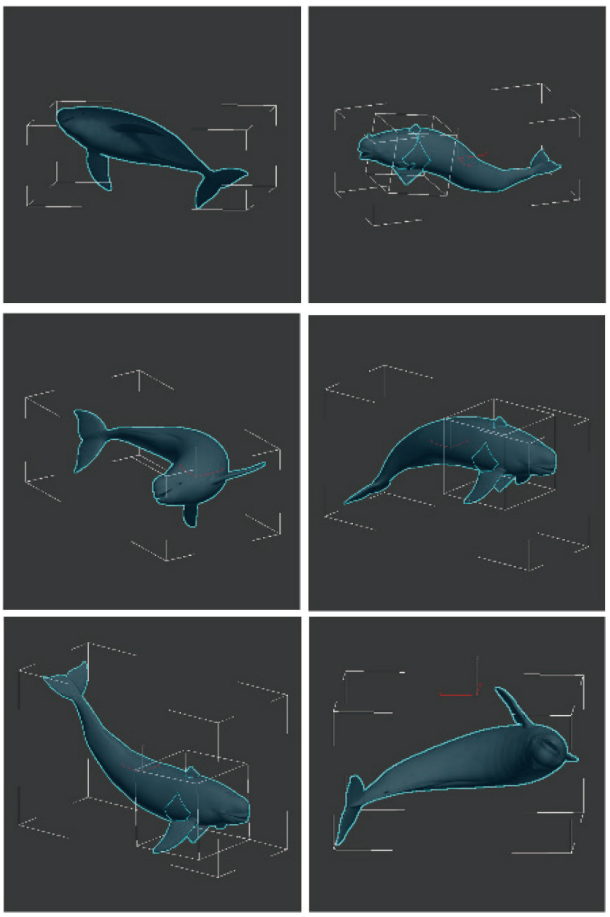

图 2 经过骨骼绑定、蒙皮后的江豚动作模型

Fig. 2 A model of a finless porpoise after bone rigging and skinning

表 1 关键属性统计信息

Tab. 1 Key attribute statistics

属性Property	值Value
边的数量	2187315
面的数量	1458210
点的数量	729107
随机渲染颜色	(0.94, 0.27, 0.73)

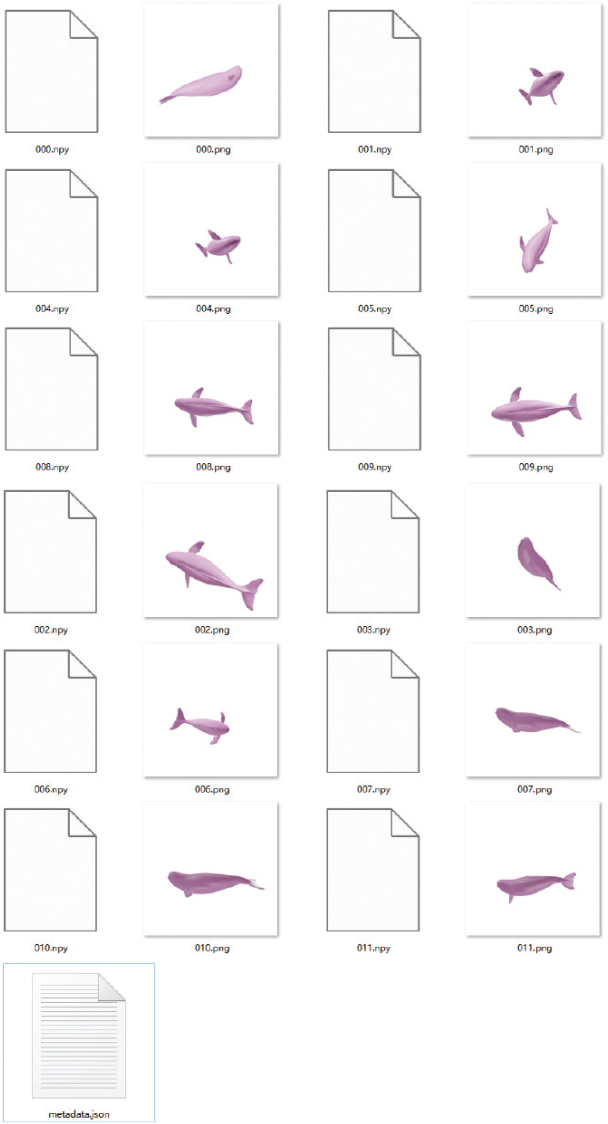

图 3　一组江豚视图和相机位姿数据集(包括江豚模型的12个不同角度的视图, 以及其对应的相机位姿)

Fig. 3　A set of views and camera pose datasets including 12 different angle views of the finless porpoise model and their corresponding camera pose

角条件扩散模型根据输入图像合成江豚多视角图像。最后, 利用神经辐射场根据输入的江豚多视角图像进行场景优化, 最终从场景中提取出三维模型。

2.2　预处理

自然光线在水下传播时, 由于不同波长的光线受到水的吸收程度不同, 其中红光衰减最快, 蓝绿光衰减最慢, 导致水下拍摄的江豚图像呈现出蓝色的偏色[13—16], 这使得江豚的细节信息出现大量的丢失, 严重影响最终的重建结果。

因此, 为对拍摄的水下江豚图像纠正色偏、提高对比度, 本文使用改进的水下图像增强模型PUIE-Net[17]对初始图像进行校正。PUIE-Net将条件变分自动编码器与自适应实例标准化相结合, 从而构建增强分布, 最后通过共识过程在多个不同的预测结果中计算出最可靠的一个。由于不同波长光线保留的程度不同, 使得水下图像的不同通道所需要处理程度也是不同的。因此, 通过在PUIE-Net中引入通道注意力机制ECA-Net[18], 根据不同通道的重要性对其赋予权重, 从而让神经网络重点关注需要增强的红通道。图像增强结果如图 4所示。

同时, 为了消除背景信息对重建主体的影响, 采用了Dense Prediction Transformer[19]对处理好的图像进行江豚前景分割, 并进行深度估计, 获得深度图。

2.3　江豚视角条件扩散模型

传统的江豚多视图合成需要采集静止状态下的江豚多视角图像, 然而, 自然状态下的江豚难以配合研究人员长时间保持静止或特定姿势。扩散模型的出现则使得获得物体多视角图像成为可能。首先, 目前的扩散模型[20, 21]在包含数十亿文本–图像对的大规模数据集[22]上进行了训练, 学习到了强大的先验知识, 其中涵盖许多对象的多角度视图先验, 有利于微调扩散模型以学习对相机位姿的控

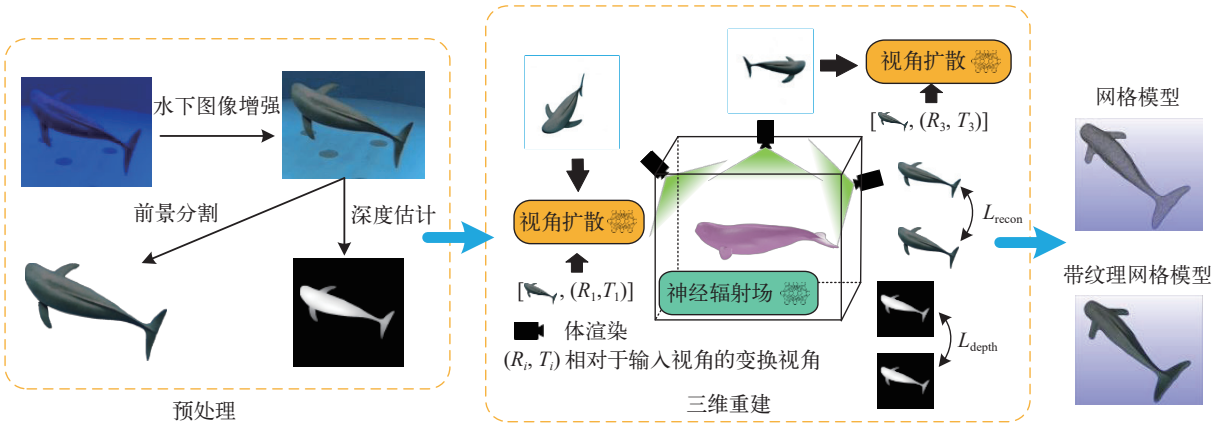

图 4　新视图合成及三维重建示意图

Fig. 4　Schematic diagram of new view synthesis and 3D reconstruction

制。其次, 从头开始训练1个具有强大泛化能力的扩散模型需要耗费大量的资源。因此, 本研究通过对大型扩散模型Stable Diffusion[21]进行微调来执行江豚新视角图像合成任务(图 5和图 6)。

在预训练的扩散模型中, 为了能够获得控制生成江豚新视角图像的能力, 需要添加控制视角的条件机制。给定一张江豚的输入图像$x \in \mathbb{R}^{H\times W\times 3}$, 令$R \in \mathbb{R}^{3\times 3}$和$T \in \mathbb{R}^{3\times 3}$分别表示目标视角相对输入视角的相机旋转和平移矩阵, 我们的目标是训练一个模型f, 该模型根据相机旋转平移的变换合成一幅新视角图像:

$$\hat{x}_{R,T} = f(x, R, T) \tag{1}$$

式中, $\hat{x}_{R,T}$表示合成的新视角图像。

然而, 要构建模型f, 仍需要面对两个挑战。其一, 尽管大型生成式扩散模型在不同角度上对数据集中涉及的各种物体进行了训练, 但是它们并没有直接编码各个视角之间的关系。其二, 扩散模型受到了互联网数据中存在的视角偏见的影响, 例如, Stable Diffusion倾向于生成处于正面视角的物体图像。这两个问题严重制约了从大型扩散模型中提取三维信息的能力。

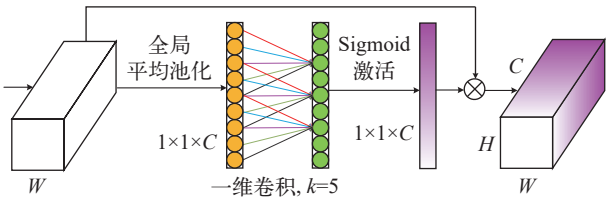

图 5 通道注意力ECA-Net模块

Fig. 5 Channel Attention ECA-Net Module

为了解决这两个问题, 江豚视角条件扩散模型将江豚多视图的相机位姿作为条件机制[21] (Conditioning mechanisms)引入扩散模型, 以实现通过相机视角变换引导扩散模型生成目标视角的新视图。为了使图像和相机位姿这两种不同的数据串联在一起, 视角条件扩散模型将输入图像使用CLIP[23](Contrastive Language-Image Pre-training, 对比文本–图像对预训练模型)进行编码, 再与相机外参(R,T)矩阵拼接起来, 形成一个联合的相机位姿CLIP嵌入, 表示为$c(x,R,T)$, 作为条件机制引入扩散模型, 引导去噪自编码器的去噪方向朝目标视角靠近, 网络结构如图 7所示。此时损失函数可以写成:

$$\boldsymbol{L}_{\mathrm{VCDM}} = \mathbb{E}_{z\sim\mathcal{E}(x),t,\epsilon\sim\mathcal{N}(0,1)}\|\epsilon - \epsilon_\theta(z_t, t, c(x,R,T))\|_2^2 \tag{2}$$

式中, $\mathcal{E}$表示一个编码器, 它将RGB图像$x \in \mathbb{R}^{H\times W\times 3}$编码至低维的隐空间$z = \mathcal{E}(x)$。隐空间相比原图像具有更小的尺寸, 能够以更高的计算效率对其进行训练。

为训练江豚视角条件扩散模型, 我们制作了江豚图像–相机外参对$\{(x, x_{(R,T)}, R, T)\}$数据集。其中, 视图图像和它的相机外参是一一对应的, 以此来训练微调的视角条件扩散模型来学习对视角的控制。本文在1.1节中制作的江豚多视图数据集就是这样一个图像–相机外参对数据集。

2.4 神经辐射场

在传统三维重建流程复杂的背景下, 神经辐射

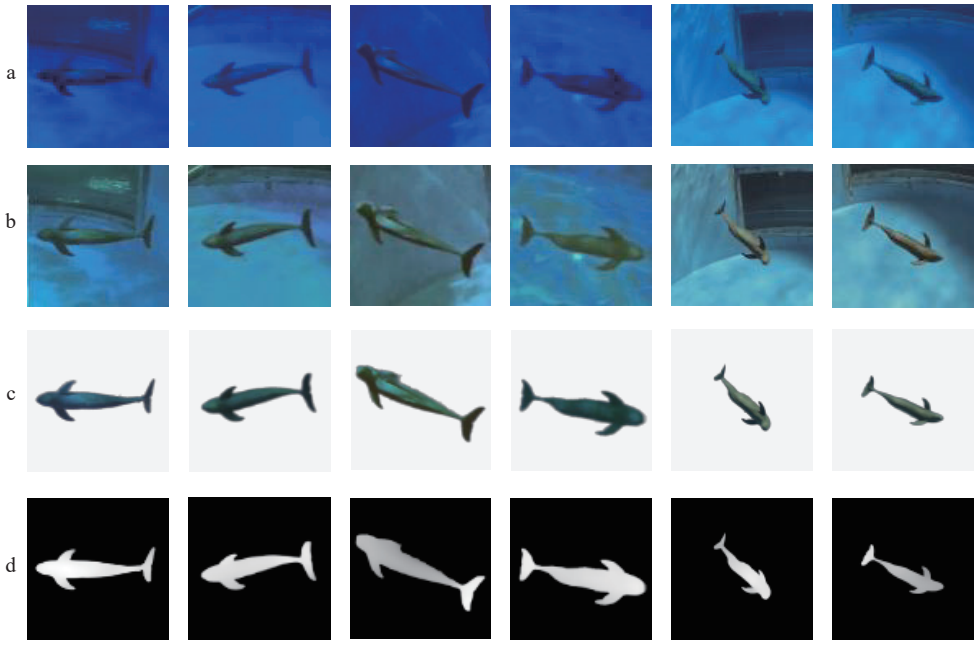

图 6 原始图像(a)、增强处理后的图像(b)、分割后的图像(c)和深度图(d)

Fig. 6 Original image (a), enhanced image (b), segmented image (c), and depth map (d)

场(NeRF, Neural radiance fields)[24]为江豚的三维重建提供了创新且高效的方法。神经辐射场表示为一个映射函数F_θ, 该函数的输入为一组江豚图像的相机位姿, 这是一个5D向量(x,y,z,θ,φ), 包括采样点坐标$p=(x,y,z)$和相机视角方向$d=(\theta,\varphi)$, 输出为颜色$c=(r,g,b)$和体素密度σ, 实现了从3D点坐标及相机视角方向到颜色和体素密度的映射, 即$F_\theta:(p,d)\rightarrow(c,\sigma)$。从以相机为起点, 方向为$(\theta,\varphi)$的射线上获取采样点, 其坐标$p$可由射线方程计算得出。将该映射函数用一个全连接神经网络来近似表示, 其网络架构如图 8所示。

由于全连接神经网络倾向于学习江豚图像中灰度值较平滑的低频细节, 如果将5D向量(x,y,z,θ,φ)直接输入多层感知机, 会导致渲染出的新视图缺乏灰度值快速变化的高频细节, 而高频细节包含较多的纹理细节信息——例如物体边缘等, 具有丰富的图像特征。因此, 为了更多地学习江豚输入图像中的高频细节, 神经辐射场引入了位置编码:

$$\gamma(\pi p)=\left(\sin\left(2^0\pi p\right),\cos\left(2^0 p\right),\cdots,\sin\left(2^{L-1}\pi p\right),\cos\left(2^{L-1}\pi p\right)\right) \tag{3}$$

式中, 函数应用于$p=(x,y,z)$的三个坐标值时令$L=10$, 得到$\gamma(p)$为60维; 应用于相机视角方向的单位向量的3个分量时$L=8$, 得到$\gamma(d)$为24维。通过位置编码, 将连续输入的低维坐标映射到更高维的空间, 使得全连接神经网络更容易逼近高频函数, 学习到江豚图像更多的高频细节。

经过全连接神经网络得到的神经辐射场还需要经过体渲染(Volume rendering)才能得到辐射场中各个点的颜色等信息, 方便后续提取江豚网格模型。从一个新视角处的相机发射的一条穿过江豚的辐射场的射线$r(t)=o+td$(其中o是射线起点, t是射线参数, d是射线方向, t_n和t_f分别是射线的近端和远端边界)的期望颜色$C(r)$可以表示为:

$$C(r)=\int_{t_n}^{t_f}T(t)\sigma(\boldsymbol{r}(t))\boldsymbol{c}(\boldsymbol{r}(t),\boldsymbol{d})dt \tag{4}$$

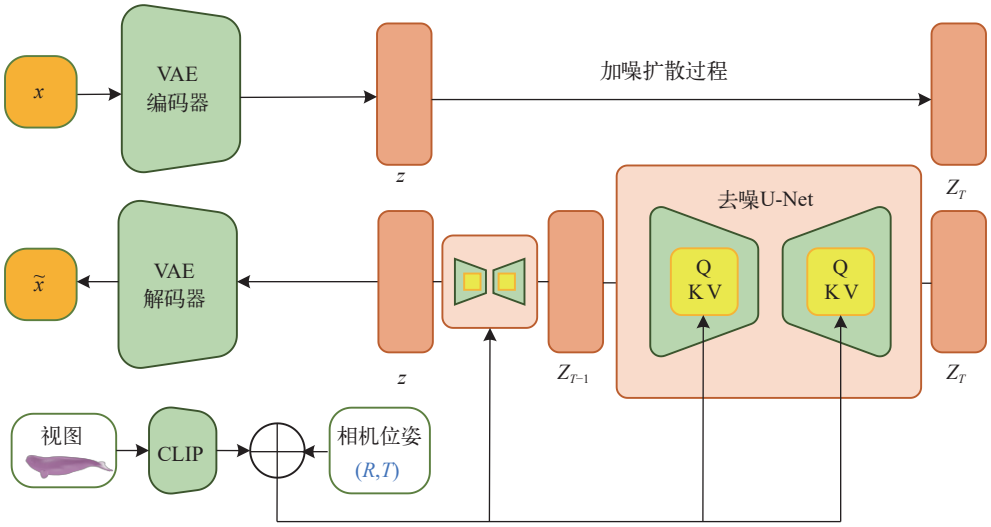

图 7　视角条件扩散模型原理示意图

Fig. 7　Schematic diagram of the principle of view-condition diffusion model

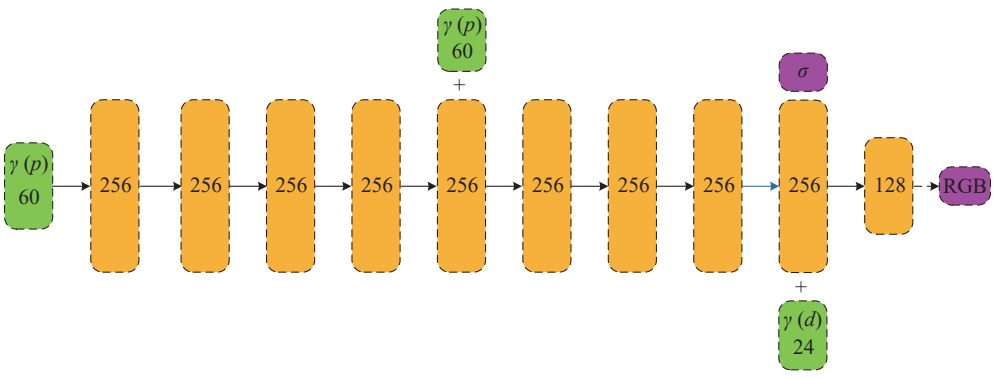

图 8　神经辐射场使用的全连接神经网络架构

Fig. 8　Fully connected neural network architecture used by neural radiance fields

式中, $\sigma(r(t))$表示射线某一采样点处的体素密度, 反映的是该点的不透明度, $c(r(t),d)$表示射线上某一采样点处的发射颜色, 且颜色只与位置和射线方向有关, $T(t)$表示射线从t_n到t的累计不透光率, 其表达式为:

$$T(t)=exp\left(-\int_{t_n}^{t}\sigma(r(s))ds\right) \tag{5}$$

为了从扩散模型生成的多视角图像中重建江豚, NeRF的优化还需要计算参考视图重建损失:

$$\begin{aligned}L_{\text{recon}}=&\lambda_{\text{rgb}}\|M\odot(I^r-G_\theta(v^r))\|^2+\\&\lambda_{\text{mask}}\|M-M(G_\theta(v^r))\|^2\end{aligned} \tag{6}$$

式中, θ是待优化的NeRF参数, ⊙表示Hadamard积, M表示神经辐射场中沿每个像素的射线积分获得的前景掩膜, 由于前景对象已经分割出来, 因此不对背景进行渲染, λ_{rgb}和λ_{mask}是前景RGB图和掩膜的权重。

为了避免重建出的模型过于平坦或过于凹陷, 呈现合理的凹凸, 参考单目深度估计[25]的损失函数, 使用皮尔逊相关系数作为深度的损失函数:

$$L_{\text{depth}}=\frac{1}{2}\left[1-\frac{cov(M\odot d^r,M\odot d)}{\sigma(M\odot d^r)\sigma(M\odot d)}\right] \tag{7}$$

式中, d^r表示利用预训练的单目深度估计器获取的参考视图的深度, d表示NeRF模型输出的深度, $cov(\cdot)$表示协方差, $\sigma(\cdot)$表示标准差。

总而言之, 重建的总体损失函数为:

$$L_{\text{total}}=\lambda_{\text{VCDM}}L_{\text{VCDM}}+L_{\text{recon}}+\lambda_{\text{depth}}L_{\text{depth}} \tag{8}$$

式中, λ_{VCDM}为视角条件扩散模型损失的权重。

将重建对象表示为神经辐射场的隐式表示形式之后, 利用行进立方体算法(MC, Marching Cubes)从其中提取出粗糙的江豚网格模型, 再通过深度移动四面体算法[26](DMTet, Deep Marching Tetrahedra)将其优化为更为平滑的江豚网格模型。

3 结果

3.1 实验平台及评价指标

江豚三维重建模型基于Pytorch深度学习框架搭建, 实验平台服务器使用的处理器型号为Intel (R) Xeon (R) Platinum 8358P CPU @2.6GHz, 显卡型号为Nvidia A40 GPU (显存为48GB)。

本研究模型使用PSNR、LPIPS[27]、SSIM[28]作为实验的评价指标, 用以评估合成江豚新视角图像的质量。

PSNR (Peak Signal-to-Noise Ratio, 峰值信噪比)基于MSE(Mean Square Error, 均方误差)而定义, 用于衡量经过处理后的图像品质, 特别是经过图像压缩之后, 输出图像与原始图像之间的差异程度。

SSIM (Structural Similarity Index, 结构相似性指标)基于人眼对图像的感知, 通过比较图像的亮度、对比度和结构等方面的相似性来评估图像质量。其值越大, 表明感知相似度越高。

LPIPS (Learned Perceptual Image Patch Similarity, 学习感知图块相似度)用于衡量两张图像之间的感知相似度, 也被称为“感知损失”, 与PSNR、SSIM等指标相比, LPIPS更符合人类的感知情况。其值越低, 表示两张图像在感知上越相似; 反之, 则表明两张图像的差异越大。

此外, 为了评估江豚三维模型的重建效果, 采用平均倒角距离(Average Chamfer Distance, ACD)和法向量连续性(Normal Consistency, NC)作为评价指标。

平均倒角距离用于衡量两个点云或网格模型点集之间的平均欧氏距离。假设有两个点云$P=\{p_1,p_2,\cdots,p_N\}$和$Q=\{q_1,q_2,\cdots,q_M\}$, 点云P中所有的点都在Q中寻找对应的距离最近的N个点, 计算欧氏距离的平均值; 点云Q中所有的点都在P中寻找对应的距离最近的M个点, 计算欧氏距离的平均值; 最后再将两个平均值相加。ACD越小说明重建效果越好, 其公式如下:

$$d_{\text{ACD}}(P,Q)=\frac{1}{N}\sum_{p\in P}\min_{q\in Q}\|p-q\|_2^2+\frac{1}{M}\sum_{q\in Q}\min_{p\in P}\|q-p\|_2^2 \tag{9}$$

法向量一致性为两个网格模型的每个面片的法向量点积的L_1范数, 可以衡量三角形面片的方向一致性及模型的几何保真度, 其范围为[0, 1], NC值越大, 表明重建效果越好。

3.2 实验结果

图 9展示了根据输入江豚视图合成新视图的效果, 可以看到, 合成视图基本反映了江豚各部位的形态。此外, 我们将新视图合成的质量使用PSNR、LPIPS和SSIM三个指标进行了定量评估, 并将3种方法进行了对比, 如表 2所示, 表明本文方法在合成江豚新视角图像方面优于另外两种方法。

图 10展示了本文方法与RealFusion、One-2-3-45[29]重建结果的各视角截图对比, 输入图像均经过水下图像增强处理。RealFusion和One-2-3-45都是一种通用型的重建算法, 但是它们没有针对江豚进行专门的训练优化, 因此其重建质量明显低于本文方法。

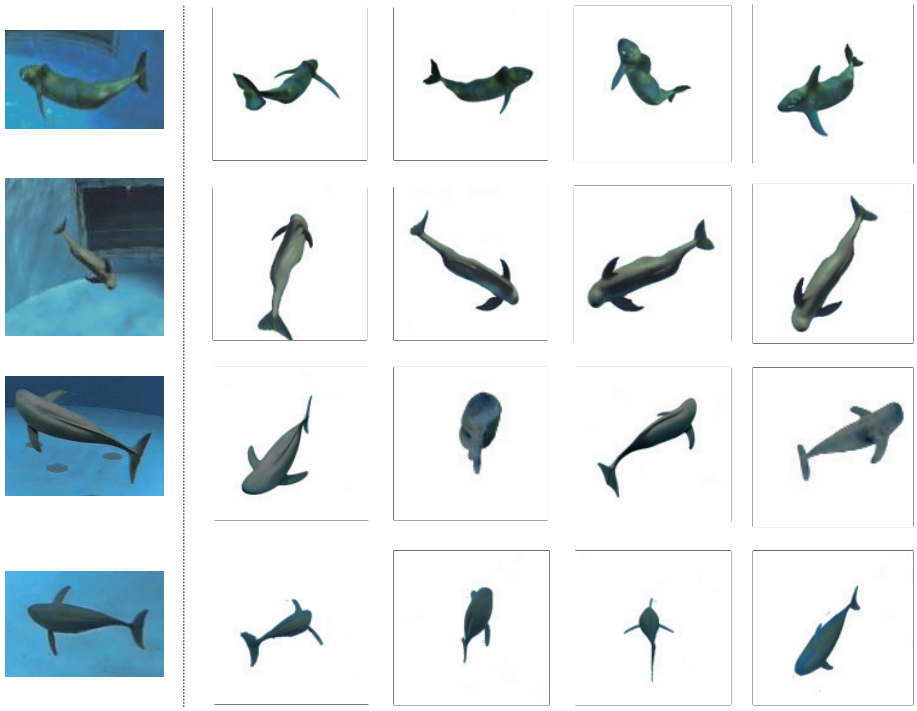

图 9　利用视角条件扩散模型进行新视图合成的效果

Fig. 9　The effect of new view synthesis using the perspective conditional diffusion model

表 2　新视图合成指标对比

Tab. 2　Comparison of composite indicators for new views

指标Index	RealFusion	One-2-3-45	Ours
PSNR↑	37.784	38.132	**38.968**
SSIM↑	0.943	0.955	**0.972**
LPIPS↓	0.305	0.298	**0.294**

图 11展示了输入图像未经水下增强处理和经过水下增强处理的重建结果。从室内获取的水下环境的江豚图像往往存在蓝色色偏、低亮度、低对比度, 这使得图像的纹理不清晰, NeRF渲染过程中的深度估计产生了偏差, 这最终导致提取的网格模型出现异常的凹凸部分。例如, 江豚1出现的头部凸起, 江豚2、3、4、6的腹部也有异常凸起, 江豚4和5出现三角形尾部。

表 3和表 4通过网格模型倒角距离和法向量一致性对重建结果进行了对比评估。从图 10可以看出, RealFusion方法重建出许多漂浮的杂点, One-2-3-45方法重建出异常的块状物, 以及图 11中未增强一栏模型的异常凸起, 这使得它们的平均倒角距离偏大, 法向量一致性偏小。

此外, 本文的方法也存在一定的局限性, 江豚视角条件扩散模型倾向于将输入视角作为正面视角, 并在此基础上推断其他视角, 导致部分视角重建质量不佳。如图 12所示, 江豚7由于角度的原因, 错误地认为头部和尾部在同一深度上, 导致尾部出现在头部正上方。而江豚8的右胸鳍呈现为微小的凸起, 以及存在输入视角和与之相对的背面视角颜色不一致的Janus问题。这主要是由于江豚视角条件扩散模型提供的先验知识比较有限造成的, 而这也是许多基于扩散模型的单视图重建方法存在的问题, 需要我们进一步改进江豚视角条件扩散模型的泛化能力。

4　结论

本文针对当前江豚三维重建应用领域存在的着水下图像色偏失真、江豚数据集不足、获取江豚多视角图像困难的问题, 提出一种基于视角条件扩散模型和神经辐射场的江豚单视图三维重建方法, 训练出了自己的江豚视角条件扩散模型, 经过评估, 效果好于目前的两种重建方法, 并取得了较好的重建结果。然而, 本方法仍然存在一些限制。例如, 江豚视角条件扩散模型倾向于将输入视角作为正面视角, 导致对于部分视角的图像重建质量不佳, 以及存在一定的Janus问题。这主要与江豚视角条件扩散模型提供的先验知识有限有关。未来, 我们计划从扩充数据集、改进深度估计等角度对模型进行优化, 以提高对各个角度江豚图像重建的泛化能力。

(作者声明本文符合出版伦理要求)


江豚1


江豚2


江豚3


江豚4


江豚5


江豚6

输入图像 a. RealFusion b. One-2-3-45 c. Ours

图 10 本文方法与RealFusion、One-2-3-45的重建结果对比

Fig. 10 Comparison of the reconstruction results of the proposed method with RealFusion and One-2-3-45

a. 未经水下图像增强　　b. 经过水下图像增强

图 11　未经水下图像增强的重建结果和经过水下图像增强的重建结果对比

Fig. 11　Comparison of the reconstruction results without and with after underwater image enhancement

表 3　网格模型平均倒角距离评估

Tab. 3　Evaluation of average chamfer distance of mesh model

江豚Finless porpoise	RealFusion	One-2-3-45	Ours (未增强)	Ours (增强)
江豚1	1.146	2.352	0.689	0.503
江豚2	0.871	2.341	0.649	0.554
江豚3	1.217	1.761	0.733	0.535
江豚4	1.473	3.576	0.675	0.583
江豚5	1.591	2.287	0.874	0.759
江豚6	1.748	1.237	0.462	0.428

表 4　法向量一致性评估

Tab. 4　Evaluation of normal vector consistency evaluation

江豚Finless porpoise	RealFusion	One-2-3-45	Ours (未增强)	Ours (增强)
江豚1	0.624	0.602	0.853	0.866
江豚2	0.705	0.483	0.824	0.841
江豚3	0.718	0.472	0.819	0.837
江豚4	0.531	0.415	0.807	0.849
江豚5	0.683	0.585	0.763	0.774
江豚6	0.524	0.673	0.876	0.882

图 12　重建质量不佳的模型示意图

Fig. 12　Schematic diagram of a poorly reconstructed model

致谢:

本文江豚图像数据由中国科学院水生生物研究所王克雄老师团队提供, 感谢王克雄团队对本研究的支持。

参考文献:

[1] Cheng Z L, Li Y T, Zuo T, *et al*. Threats and conservation strategies of the East Asian finless porpoises in China [J]. *Journal of Applied Oceanography*, 2024, **43**(3): 597-606. [程兆龙, 李永涛, 左涛, 等. 我国东亚江豚的研究现状、面临的威胁与保护建议 [J]. 应用海洋学学报, 2024, **43**(3): 597-606.]

[2] Wang K W, Zhou K Y, Chen M M, *et al*. Beware of several problems in ex-situ protection of Yangtze finless porpoise [J]. *Journal of Nanjing Normal University* (Natural Science Edition), 2024, **47**(2): 91-98. [王康伟, 周开亚, 陈敏敏, 等. 长江江豚迁地保护需要注意的几个问题 [J]. 南京师大学报(自然科学版), 2024, **47**(2): 91-98.]

[3] Hao Y J, Tang B, Mei Z G, *et al*. Further suggestions on conservation of the Yangtze finless porpoise based on retrospective analysis of the current progress [J]. *Acta Hydrobiologica Sinica*, 2024, **48**(6): 1065-1072. [郝玉江, 唐斌, 梅志刚, 等. 长江江豚保护进展的回顾性分析及进一步保护建议 [J]. 水生生物学报, 2024, **48**(6): 1065-1072.]

[4] Zuffi S, Kanazawa A, Jacobs D W, *et al*. 3D Menagerie: Modeling the 3D Shape and Pose of Animals [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 5524-5532.

[5] Rüegg N, Zuffi S, Schindler K, *et al*. BARC: Learning to Regress 3D Dog Shape from Images by Exploiting Breed Information [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 3866-3874.

[6] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models [J]. *Advances in Neural Information Processing Systems*, 2020(33): 6840-6851.

[7] Chan E R, Nagano K, Chan M A, *et al*. Generative Novel View Synthesis with 3d-aware Diffusion Models [C]. 2023 IEEE/CVF International Conference on Computer Vision (ICCV). October 1-6, 2023, Paris, France. IEEE, 2023: 4217-4229.

[8] Watson D, Chan W, Martin-Brualla R, *et al*. Novel view synthesis with diffusion models [EB/OL]. 2022: 2210.04628. https://arxiv.org/abs/2210.04628v1.

[9] Melas-Kyriazi L, Laina I, Rupprecht C, *et al*. RealFusion 360° Reconstruction of Any Object from a Single Image [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17-24, 2023. Vancouver, BC, Canada. IEEE, 2023: 8446-8455.

[10] Liu R, Wu R, Van Hoorick B, *et al*. Zero-1-to-3: Zero-shot One Image to 3D Object [C]. 2023 IEEE/CVF International Conference on Computer Vision (ICCV). October 1-6, 2023, Paris, France. IEEE, 2023: 9264-9275.

[11] Deitke M, Schwenk D, Salvador J, *et al*. Objaverse: A Universe of Annotated 3D Objects [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 13142-13153.

[12] Arampatzakis V, Pavlidis G, Mitianoudis N, *et al*. Monocular depth estimation: a thorough review [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, **46**(4): 2396-2414.

[13] Hu K, Weng C, Zhang Y, *et al*. An overview of underwater vision enhancement: from traditional methods to recent deep learning [J]. *Journal of Marine Science and Engineering*, 2022, **10**(2): 241.

[14] Jaffe J S. Computer modeling and the design of optimal underwater imaging systems [J]. *IEEE Journal of Oceanic Engineering*, 1990, **15**(2): 101-111.

[15] Mobley C D. Light and Water: Radiative Transfer in Natural Waters [M]. Academic Press, 1994.

[16] Anwar S, Li C, Porikli F. Deep underwater image enhancement [EB/OL]. 2018: 1807.03528. https://arxiv.org/abs/1807.03528.

[17] Fu Z, Wang W, Huang Y, *et al*. Uncertainty Inspired Underwater Image Enhancement [C]. 2022 European Conference on Computer Vision (ECCV). Cham: Springer Nature Switzerland, 2022: 465-482.

[18] Wang Q, Wu B, Zhu P, *et al*. ECA-net: Efficient Channel Attention for Deep Convolutional Neural Networks [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 11531-11539.

[19] Ranftl R, Bochkovskiy A, Koltun V. Vision Transformers for Dense Prediction [C]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). October 10-17, 2021, Montreal, QC, Canada. IEEE, 2021: 12159-12168.

[20] Saharia C, Chan W, Saxena S, *et al*. Photorealistic text-to-image diffusion models with deep language understanding [J]. *Advances in neural information processing systems*, 2022(35): 36479-36494.

[21] Rombach R, Blattmann A, Lorenz D, *et al*. High-Resolution Image Synthesis with Latent Diffusion Models [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 10674-10685.

[22] Schuhmann C, Beaumont R, Vencu R, *et al*. Laion-5b: An open large-scale dataset for training next generation image-text models [J]. *Advances in Neural Information Processing Systems*, 2022(35): 25278-25294.

[23] Radford A, Kim J W, Hallacy C, *et al*. Learning Transfe-

rable Visual Models from Natural Language Supervision [C]. Proceedings of the 38th International Conference on Machine Learning (ICML). July 18-24, 2021, New York, NY, USA. PMLR 139: 8748-8763.

[24] Mildenhall B, Srinivasan P P, Tancik M, *et al*. Nerf: Representing scenes as neural radiance fields for view synthesis [J]. *Communications of the ACM*, 2021, **65**(1): 99-106.

[25] Ranftl R, Lasinger K, Hafner D, *et al*. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **44**(3): 1623-1637.

[26] Shen T, Gao J, Yin K, *et al*. Deep marching tetrahedra: a hybrid representation for high-resolution 3d shape synthesis [J]. *Advances in Neural Information Processing Systems*, 2021(34): 6087-6101.

[27] Zhang R, Isola P, Efros A A, *et al*. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 586-595.

[28] Wang Z, Bovik A C, Sheikh H R, *et al*. Image quality assessment: from error visibility to structural similarity [J]. *IEEE Transactions on Image Processing*, 2004, **13**(4): 600-612.

[29] Liu M, Xu C, Jin H, *et al*. One-2-3-45: Any single image to 3d mesh in 45 seconds without per-shape optimization [J]. *Advances in Neural Information Processing Systems*, 2024: 36.

A SINGLE-VIEW 3D MODEL RECONSTRUCTION METHOD FOR YANGTZE FINLESS PORPOISE

HUANG Zhi-Yong[1, 2, 3], YANG Chen-Long[1, 3], SHI Xiao-Tao[2, 4], HUA Xi-Feng[1, 3], TU Fa-Xian[2, 4], DING Tuo-Jun[1, 3], SHE Ya-Li[1, 3] and XIANG Meng-Li[1, 3]

(1. *Key Laboratory of Intelligent Vision Monitoring for Hydroelectric Engineering*, *Hubei Provincial University*, *China Three Gorges University*, *Yichang* 443002, *China*; 2. *International Science and Technology Cooperation Base for Fish Passage Technology in Hubei Province*, *China Three Gorges University*, *Yichang* 443002, *China*; 3. *School of Computer and Information Science*, *China Three Gorges University*, *Yichang* 443002, *China*; 4. *College of Hydraulic & Environmental Engineering*, *China Three Gorges University*, *Yichang* 443002, *China*)

Abstract: In the field of 3D reconstruction of Yangtze finless porpoises, challenges such as underwater image color distortion, limited datasets, and difficulty in capturing multi-view images of Yangtze porpoises remain significant. Emerging methods have yet to address these issues specifically for Yangtze finless porpoises. To tackle these challenges, this paper proposes a novel single-view 3D reconstruction method for Yangtze finless porpoises, combining diffusion models and neural radiance fields. First, an improved underwater image enhancement technique is developed to effectively address the issue of underwater color distortion. Second, a custom multi-view image dataset of Yangtze finless porpoises is created to fine-tune a view-conditioned diffusion model, enabling the synthesis of multi-view images from a single view. This provides a new approach for reconstructing Yangtze finless porpoises from a single image. Finally, a neural radiance field is employed to reconstruct the 3D model of the porpoise. The reconstruction results were evaluated using the average chamfer distance (ACD) and normal consistency (NC). The proposed method achieved lower ACD and higher NC compared to existing methods, demonstrating its effectiveness in reconstructing 3D models that accurately capture the coloration and morphology of Yangtze finless porpoises. The synthesized multi-view images achieved PSNR, SSIM, and LPIPS values of 38.968, 0.972, and 0.294, respectively, surpassing the performance of existing methods. Additionally, the reconstruction results after underwater image enhancement yielded the lowest ACD of 0.428 and the highest NC of 0.882, further highlighting the superiority of the proposed approach.

Key words: Diffusion models; New view synthesis; Neural radiance fields; 3D reconstruction; Yangtze finless porpoise