层主要作用是融合全局信息, 以纠正 3D 模型中错误恢复的部分, 本文使用 PSA 模块替换全连接层, PSA 模块具有局部和全局信息融合的效果, 并且参数量会大大降低, 参数量对比见后文分析.

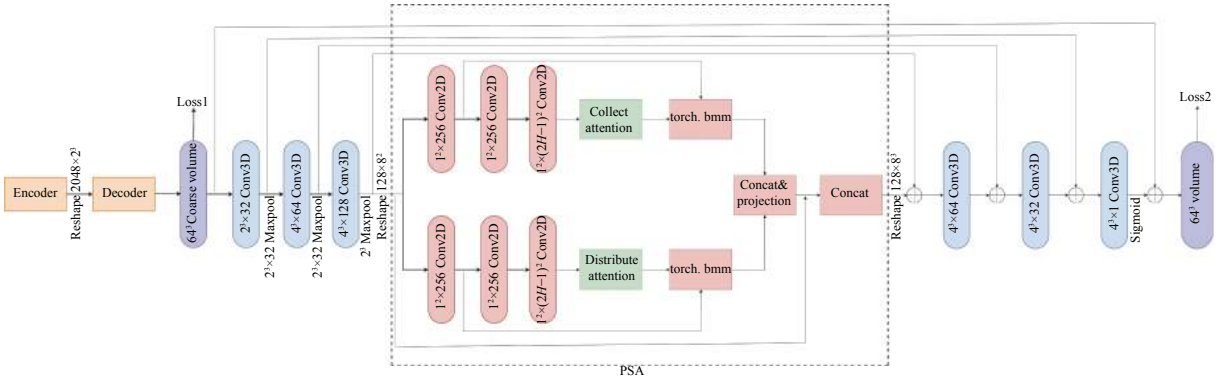

图 3 3D 重建网络结构图

表 1 编码器-解码器网络结构

编码器	解码器
VGG前8层	4^3×512 Deconv3D, BN, ReLU
3^2×512 Conv2D, BN, ELU 3^2×512 Conv2D, BN, ELU	4^3×128 Deconv3D, BN, ReLU
3^2 MaxPool	4^3×32 Deconv3D, BN, ReLU
1^2×256 Conv2D, BN, ELU	4^3×8 Deconv3D, BN, ReLU 4^3×8 Deconv3D, BN, ReLU*
Reshape 2^3×2048	1^3×1 Deconv3D, BN, Sigmoid

注: *为添加的反卷积层

PSA 包括上下并行的两个分支, 如图 3 中虚线部分, 在实现上两个分支是完全一样的. 在每个分支中, 首先应用具有 1×1 的卷积层减少输入特征图 X 的通道数以降低计算开销, 得到 $H\times W\times C$ 的特征图 X^r. 然后再应用一个 1×1 的卷积层得到 $H\times W\times[(2H-1)\times(2W-1)]$ 的 X^c, 最后经过收集和分散操作获得两个具有全局融合信息的特征图. 收集或分散操作如图 4 所示, 图 4 对应图 3 中的 CollectAttention 和 DistributeAttention 具体操作. 特征图 X^r 上每个位置 i, 对应特征图 X^c 的 i 位置 $1\times1\times[(2H-1)\times(2W-1)]$ 的特征图, 将其转换成 $(2H-1)\times(2W-1)\times1$ 的特征图 X^i, 最后以 i 作为特征图 X^i 中心点, 如图 4 中以虚线突出显示的区域是用于特征的收集和分散, 将该区域与 H^r 进行矩阵相乘得到具有全局融合信息的特征图. 上下两层各得到这样的一个分支特征图, 我们将该两个分支特征图进行通道数叠加并进行 1×1 的卷积操作得到具有融合全局信息的特征图, 最后与具有局部信息的特征图 X 进行通道数叠加, 完成全局信息的融合.

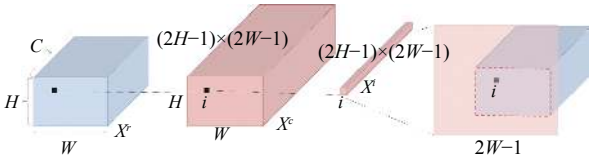

图 4 Collect / Distribute Attention

网络采用的损失函数是体素交叉熵的平均值, 公式如下:

$$L=\frac{1}{N}\sum_{i=1}^{N}[y_i\log(p_i)+(1-y_i)\log(1-p_i)] \qquad (2)$$

其中, N 为标签的数量, y_i 为真实概率, p_i 为预测概率, p_i 越接近 y_i, L 值越小.

2.3 后处理模块

3D 重建网络输出的 3D 模型是一个三维矩阵, 每个位置的值代表了是该点的置信度. Trimesh 是一个专门用来加载和使用三角网格的 Python 库, 可以直接得到包围该重建模型的最小外接立方体的归一化的长宽高, 但是因为在重建结果中会在边界生成一些错误的点, 如图 5 中红色圈标注所示, 如果用 Trimesh 库来计算会有较大的误差, 所以本文设计了算法 1 来剔除这些噪点.

算法 1. 噪点剔除算法

1) 将生成的 3D 模型中小于阈值 $T1$ 的点去除.
2) 如图 5 中坐标系所示, 假定符合 $y=0$, $z=0$ 所有点的集合为 $S1$, $S1$ 中所有点中 x 坐标最大值即为快递纸箱归一化后的长 l.
3) 在 $z=0$ 的 xy 这个面上, 统计 y 从 max 到 0 每行点的个数, 假定 $S1$ 中点的个数为 $N1$, 若某行中的点的个数大于$N1(1-T1+T2)$时, 则此时 y 值为快递纸箱归一化后的 h.

4) 假定符合 $x=0, z=0$ 所有点的集合为 $S2$, 假定 $S2$ 中点的个数为 $N2$, 在 $x=0$ 的 yz 这个面上, 统计 z 从 max 到 0 每列点的个数, 若某列中点的个数大于 $N2(1-T1+T2)$ 时, 则此时 z 值为快递纸箱归一化后的 w.

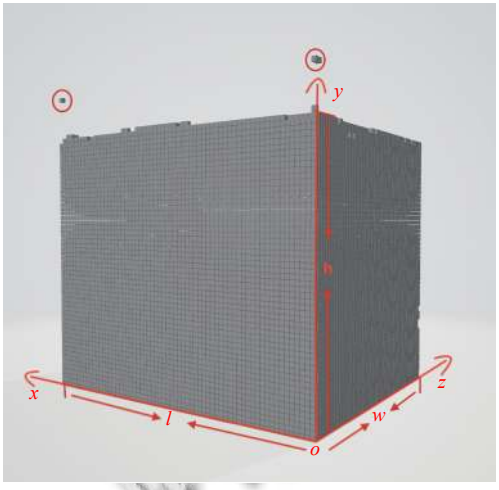

图 5 三维重建结果图

通过噪点剔除算法可以得到快递纸箱精确地归一化后的长宽高值, 由 2.1 节图像预处理操作可以得到快递纸箱实际的长度 L, 假定快递纸箱实际的宽度为 W, 实际的高度为 H, 存在如下的比例关系:

$$\frac{l}{L}=\frac{w}{W}=\frac{h}{H} \tag{3}$$

通过式 (3) 我们可以求出快递纸箱实际的长宽高, 即能算出快递纸箱的实际体积.

3 实验分析

3.1 数据集

按照文献 [9] 实验的数据集设置, 我们也使用了 ShapeNet[15] 的子集, 包含 13 个类别, 共 43 783 个 3D 模型组成. 由于目前的相关的 3D 重建的开源数据集中没有快递纸箱的模型, 所以为了验证本文算法的有效性, 我们使用 Cinema 4D 软件制作了与 ShapeNet 相同格式的快递纸箱数据集, 总共 150 个快递纸箱的 3D 模型, 3600 张 2 维图片, 这里简称该数据集为 box-150.

3.2 评价指标

首先为了评估改进后的 3D 重建网络的性能, 使用 3D 体素重建与真实体素标签之间的体素 IoU 作为相似性度量, 其公式如下:

$$IoU=\frac{\sum_{i,j,k}I\left(p_{i,j,k}>t\right)I\left(y_{i,j,k}\right)}{\sum_{i,j,k}I\left[I\left(p_{i,j,k}>t\right)+I\left(y_{i,j,k}\right)\right]} \tag{4}$$

其中, $p_{(i,j,k)}$ 和 $y_{(i,j,k)}$ 分别表示预测概率和真实标签. $I(\cdot)$ 是指示函数, t 表示体素化阈值. IoU 越高, 重建越好.

我们用体积的相对误差衡量一个快递纸箱体积测量结果的好坏, 具体公式如下:

$$\delta=\frac{\left|V_y-V_p\right|}{V_y} \tag{5}$$

其中, V_p 和 V_y 分别代表预测体积和真实体积, δ 越小表示体积测量的越准确.

3.3 实验结果

在实际训练中, 由于快递纸箱数据集的限制, 本文采用迁移学习的思想, 网络先用 ShapeNet 子集进行预训练 150 个周期, 接着在 box-150 数据集上训练 150 个周期. 网络采用 224×224 RGB 图像作为输入, 使用 Adam 优化器, β_1 为 0.9, β_2 为 0.999, 预训练时, 批处理大小为 64, 初始学习率都设置为 0.001, 在 box-150 训练时, 批处理大小为 10, 初始学习率都设置为 0.0005.

本文选择 3 种 3D 重建算法作为对比, 第 1 种是原始的 Pix2Vox, 第 2 种是 Pix2Vox 去除网络中的全连接层, 简称为 Pix2Vox-NF32, 第 3 种是 Pix2Vox 去除网络中全连接层并将网络输出分辨率提高至 64^3, 这里简称为 Pix2Vox-NF64. 表 2 为 4 个网络在不同阈值下重建性能的对比, 以 IoU 作为评价标准, 实验数据表明, 直接去除全连接层会降低网络的性能, 但去除全连接层后提高网络的输出分辨率能够提升网络的模型, 当同时采用提高分辨率与 PSA, 模型更优. 图 6 为 4 种模型的参数量比较, 通过比较可以发现我们的网络比 Pix2Vox 参数量更低.

表 2 不同阈值下模型性能对比

模型	t=0.20	t=0.30	t=0.40	t=0.50
Pix2Vox	0.8109	0.8218	0.8259	0.8238
Pix2Vox-NF32	0.7904	0.8032	0.8140	0.8128
Pix2Vox-NF64	0.8183	0.8249	0.8302	0.8350
Our	**0.8263**	**0.8390**	**0.8449**	**0.8445**

体积测量的误差主要来自两个部分, 第 1 个是在图像预处理阶段, 获取快递纸箱的长度时会存在误差; 第 2 个误差是 3D 重建得到的归一化长宽高的值, 这一误差可以归结为后处理阶段. 我们对 box-150 中 25 组

数据进行了第一个误差的统计分析, 具体如图 7 所示, 根据预处理获取快递纸箱的长度的误差平均仅为 0.6%, 80% 的数据的相对误差都小于 1%, 证明了图像预处理模块对于计算快递纸箱长度的有效性. 第 2 个误差可以通过统计快递纸箱的体积相对误差来分析, 并与用 trimesh 库来计算体积进行对比, 通过图 8 可以看出由后处理得出的平均体积相对误差比用 trimesh 库的低了很多, 证明了后处理模块的有效性.

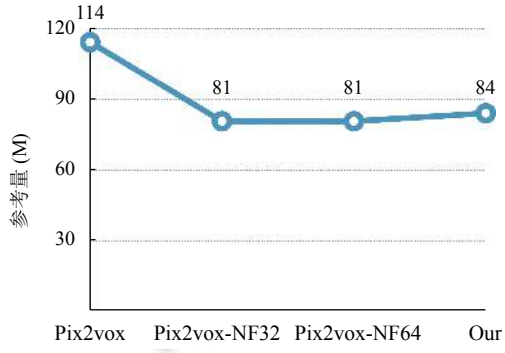

图 6 不同模型的参数量对比

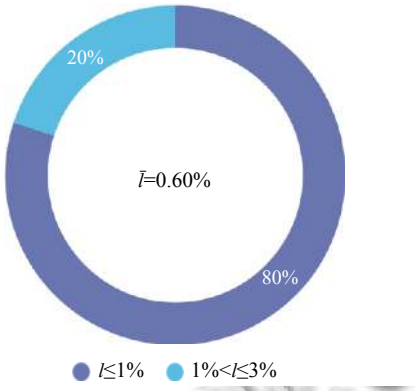

图 7 25 组快递纸箱长度相对误差分析

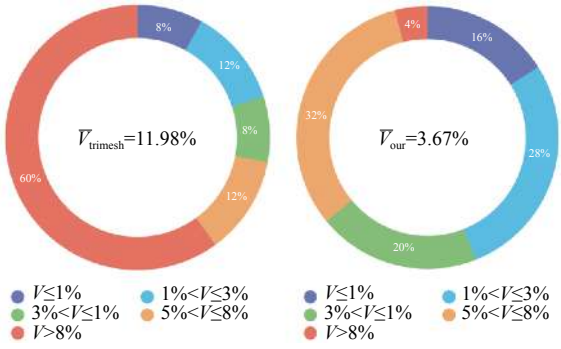

图 8 基于 trimesh 库与后处理程序的体积误差对比图

考虑算法的实用性, 本文还测试了 6 组真实数据, 具体的误差分析见表 3, 可以看出相对误差基本在 5% 以内, 满足实际的需求. 但相较于 box-150 数据集效果差了一点, 这是因为真实数据与模型训练使用的数据还是存在着一些差异.

表 3 手机拍照快递纸箱体积测量误差分析

纸箱编号	实际体积(cm^3)	测量体积(cm^3)	相对误差(%)
01	25 238.85	26 498.97	4.99
02	15 932.83	14 888.38	6.56
03	9962.66	9745.42	2.18
04	36 493.15	38 295.04	4.94
05	58 162.20	60 294.61	3.67
06	21 933.60	22 543.33	2.78

4 总结展望

本文基于 3D 重建网络, 设计了一个通过手机对快递纸箱拍照就能获得其体积的算法, 通过实验证明了图像预处理模块计算快递纸箱长度的有效性以及提高网络的重建分辨率并结合 PSA 模块有助于提高快递纸箱的重建性能, 此外后处理模块能更精确的计算快递纸箱的体积. 最后考虑算法的实用性, 本文还对真实数据做了测试.

下一步的研究工作在于如何让该算法对拍摄快递纸箱的角度更加鲁棒性, 保证在实际操作中更方便快捷地得到快递纸箱的体积数据.

参考文献

1 异方科技. 智能体积测量仪 快递物流行业的“加速器”. https://baijiahao.baidu.com/s?id=1622629186506567650&wfr=spider&for=pc. [2019-01-14].

2 刘士兴, 宓逸舟, 张阳阳, 等. 改进型测量光幕体积计量系统. 电子测量与仪器学报, 2016, 30(9): 1313–1319.

3 毛丹辉, 单彬, 王勇, 等. 激光技术在智慧物流中的应用. 物流科技, 2017, 40(2): 84–86, 95. [doi: 10.3969/j.issn.1002-3100.2017.02.026]

4 王玉伟, 尹颜朋. 基于 RCF 边缘检测和双目视觉的箱体体积测量算法. 现代计算机, 2017, (35): 71–74. [doi: 10.3969/j.issn.1007-1423.2017.35.015]

5 宓逸舟. 基于双目视觉的快递包裹体积计量系统 [硕士学位论文]. 合肥: 合肥工业大学, 2017.

6 Han XF, Laga H, Bennamoun M. Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019. [doi: 10.1109/TPAMI.2019.

2954885]

7 Wu JJ, Zhang CK, Xue TF, *et al*. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain. 2016. 82–90.

8 Wu JJ, Wang YF, Xue TF, *et al*. Marrnet: 3D shape reconstruction via 2.5D sketches. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA. 2017. 540–550.

9 Choy CB, Xu DF, Gwak JY, *et al*. 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 628–644.

10 Xie HZ, Yao HX, Sun XS, *et al*. Pix2Vox: Context-aware 3D reconstruction from single and multi-view images. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 2690–2698.

11 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.

12 Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv: 1804.02767, 2018.

13 Rother C, Kolmogorov V, Blake A. "GrabCut": Interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics, 2004, 23(3): 309–314. [doi: 10.1145/1015706.1015720]

14 Zhao HS, Zhang Y, Liu S, *et al*. PSANet: Point-wise spatial attention network for scene parsing. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany. 2018. 267–283.

15 Chang AX, Funkhouser T, Guibas L, *et al*. ShapeNet: An information-rich 3D model repository. arXiv: 1512.03012, 2015.