

融合 PointNet 和 3D-LMNet 的单幅图像 三维重建及语义分割

陈辉^{1*}, 童勇¹, 朱莉¹, 梁维斌²

¹上海电力大学自动化工程学院, 上海 200090;

²开放智能机器(上海)有限公司, 上海 200233

摘要 由单幅图像重建三维结构并感知三维对象的语义理解极具挑战性。针对单幅图像难以直接生成三维重建点云问题,提出一种融合 PointNet 与 3D-LMNet 的联合优化网络模型进行三维重建并完成语义分割。基于 3D-LMNet 网络进行训练生成三维点云,并完成局部分割,同时,对网络损失函数进行联合优化来预测分割点云。通过分割点云的语义信息改善重建效果,生成带有语义分割信息的三维点云重建模型。针对联合训练中真值点云和预测点云类别标签无点对点的对应关系问题,引入联合优化损失函数来提高重建和分割效果,生成最终三维重建模型。通过在 ShapeNet 数据集上实验验证,并与 PointNet 和 3D-LMNet 单独训练相比,所提模型在平均交并比(mIoU)上提高了 4.23%,在倒角距离(CD)和 EMD(earth mover's distance)上分别降低了 7.97% 和 6.04%,联合优化网络明显改善了重建和分割的点云模型。

关键词 深度学习; 单幅图像; 联合优化; 三维重建; 语义分割

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP202259.1815010

3D Reconstruction and Semantic Segmentation Method Combining PointNet and 3D-LMNet from Single Image

Chen Hui^{1*}, Tong Yong¹, Zhu Li¹, Liang Weibin²

¹School of Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China;

²Open AI Lab (Shanghai) Co., Ltd., Shanghai 200233, China

Abstract It is very challenging to reconstruct the 3D structure from a single image and perceive the semantic information of 3D objects. Aiming at the problem that it is difficult to directly generate a 3D reconstruction model from a single image input, a joint optimization network model combining PointNet and 3D-LMNet is proposed for single image 3D reconstruction and semantic segmentation. First, a 3D point cloud is generated by training based on the 3D-LMNet network, and then local segmentation is performed. Meanwhile, the network loss function is jointly optimized to predict the segmented 3D point cloud. Then, the reconstruction effect is improved through the semantics information of segmented point cloud, and a 3D point cloud reconstruction model is generated with semantic segmentation information. Finally, in view of the problem that there is no point-to-point correspondence between the true value point cloud and the predicted point cloud category label during the joint training, the joint optimization loss function is introduced into the joint optimization network to improve the reconstruction and segmentation effect, and the 3D reconstructed model is made. Through verification on the ShapeNet dataset, and comparison with PointNet and 3D-LMNet training, the model in this paper improves mean intersection over union (mIoU) by 4.23%, and reduces chamfer distance (CD) and earth mover's distance (EMD) by 7.97% and 6.04%, respectively. The joint optimization network significantly improves the reconstruction and segmented point cloud model.

Key words deep learning; single image; joint optimization; 3D reconstruction; semantic segmentation

收稿日期: 2021-06-15; 修回日期: 2021-07-26; 录用日期: 2021-08-10

基金项目: 国家自然科学基金(51705304)、上海市自然科学基金面上项目(20ZR1421300)、上海市浦江人才计划项目(21PJJD025)

通信作者: *chenhui@shiep.edu.cn

1 引言

计算机视觉中,直接由单幅图像推断三维结构并感知三维语义理解极具困难。随着机器人技术和三维感知技术的广泛应用^[1],以及深度学习方法的快速发展^[2],通过神经网络模型^[3]仅依靠单幅图像实现三维世界感知,给三维重建及模式识别提供了有利支撑^[4-5]。Choy等^[6]提出基于体素表示的3D-R2N2完成由任意视点的单个或多个视图生成三维结构;Wu等^[7]提出MarrNet模型在真实图像上进行端到端的训练,为单幅图像的三维形状重建2.5D模型;Girdhar等^[8]将3D模型和相应的2D表示映射到公共嵌入空间,获得既可从2D图像预测又能够生成3D对象的表示。由于体素表示容易造成信息丢失,Fan等^[9]提出基于点云表达的PSGN网络,通过条件形状采样器,从输入图像中预测多个可信的3D点云;Mandikal等^[10]在点云自动编码器的潜在空间中训练了图像编码器;Charles等^[11]提出PointNet通过空间变换网络将无序的点云与点云特征对齐,提取点云的全局特征,提高点云数据处理效率;PointNet++^[12]用分层的学习结构提取了全局特征和局部特征,增强点云信息的提取;朱育正等^[13]引入基于注意力机制的图像编码器网络,获得较好的重建效果;张爱武等^[14]提出基于卷积神经网络(CNN)的多特征三维点云语义分割方法,实现高精度的三维点云语义分割;胡涛等^[15]提出基于多层深度特征融合的PolSAR图像语义分割方法,具有较高的分割精度;徐

聪等^[16]提出改进DeepLabV3+网络的图像语义分割方法,可预测出精细的结果;鲍海龙等^[17]提出基于区域自我注意力的实时分割网络,分割精度高、速度快。Song等^[18]以场景的深度图作为输入,并成功预测出基于体素表达包含语义信息的重建模型。上述方法专注于单独的点云重建和语义分割任务,无法完成端到端带有语义标签的三维点云模型。

针对三维模型缺乏语义信息的问题,本文结合PointNet分割网络和3D-LMNet重建网络,将分割网络和重建网络相结合并引入联合优化损失函数,提出一种基于单幅图像的联合优化网络模型以实现三维点云语义重建,与单独训练相比,可改善重建和分割效果。

2 基本原理

网络结构如图1所示,主要包含三维重建、局部分割和联合优化3个部分。首先,将单幅图像输入到重建网络,预测三维重建点云模型,以训练得到重建损失;在局部分割中,将输入图像真值点云输入到分割网络进行语义分割,预测出局部分割模型,以训练得到分割损失;在联合优化中,引入联合优化损失函数,对重建分割网络进行联合优化。联合优化即对重建与分割共同预测,利用联合优化损失函数来改善重建效果和分割效果。图1中 L_{rec} 为3D-LMNet独立重建网络的损失函数, L_{ce} 为PointNet独立分割网络的损失函数, L_{seg} 为提出的联合优化损失函数, a 为重建影响因子, b 为分割影响因子。

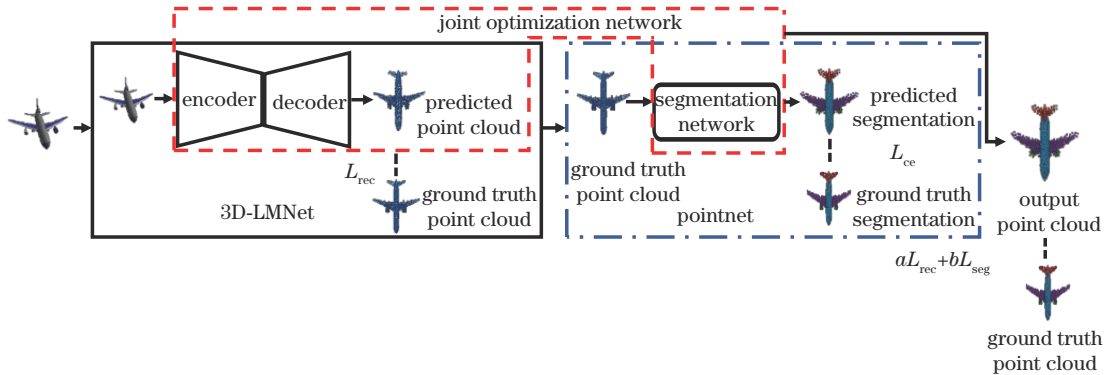


图1 网络结构流程图

Fig. 1 Flow diagram of network structure

2.1 3D-LMNet独立重建网络

该网络为编码-解码网络,在给定输入图像 I 时,训练得到重建点云 \hat{X}_p ,其中 $\hat{X}_p \in R^{N_p \times 3}$, R 代表几何空间, N_p 为点云中包含的点数。仅考虑重建过程中,对于重建网络部分,由于优化的损失函数需要阶保持不变,选择真值点云 X_p 和 \hat{X}_p 之间的倒角距离(CD)^[9]作为重建损失函数 L_{rec} 。CD用于计算预测点云 \hat{X}_p 与 X_p 的最短距离,CD值越小,预测点云与真值点云之间距离越小,重建效果越好。因此 L_{rec} 损失函数定义为

$$L_{rec} = d_{CD}(X_p, \hat{X}_p) = \sum_{x \in X_p} \min_{\hat{x} \in \hat{X}_p} \|x - \hat{x}\|_2^2 + \sum_{\hat{x} \in \hat{X}_p} \min_{x \in X_p} \|x - \hat{x}\|_2^2, \quad (1)$$

式中: x 为具体点; \hat{x} 为预测点云中的点。

2.2 PointNet独立分割网络

为训练获得分割损失 L_{ce} ,在独立分割部分,将输入图像的真值点云输入重建网络,训练得到带有类别标签的点云 \hat{X}_c 。在此采用PointNet^[11]网络作为独立分割网络模型。PointNet将无序的、不变的点云作为输

入,每个点包含基本的坐标 (x,y,z) 及颜色值 (R,G,B) 、法线值和点之间的距离等信息,以及每个点的分类标签。分割网络结构包含变换网络(T-Net)、多层感知机(MLP)以及作用于聚合所有点特征信息的对称函数(Max pooling)。

PointNet 使用 T-Net 来预测变换矩阵,并应用到输入点云和点云特征空间,以对齐不同变换后获得的点云。由于特征空间中的变换矩阵具有较高维度,为了约束输出的特征变换矩阵为正交矩阵,降低优化的难度,网络在 softmax 训练损失的基础上增加了正则化项

$$L_{\text{reg}} = \|I - \mathbf{A}\mathbf{A}^T\|_F^2, \quad (2)$$

式中: \mathbf{A} 是 T-Net 预测输出的特征对齐矩阵; F 是网络在 softmax 训练损失的基础上增加的正则化操作。

MLP 最底层是输入层,中间是隐藏层,最上层是输出层。如果输入一个 n 维向量 \mathbf{X} ,输入层有 n 个神经元,每一个神经元由一个线性拟合和一个非线性激活函数组成,隐藏层的输出为 $f(W_1\mathbf{X} + b_1)$, W_1 是权重, b_1 是偏置,函数 f 为 sigmoid 函数或 tanh 函数。隐藏层到输出层为 softmax 回归,输出层输出为 $\text{softmax}(W_2X_1 + b_2)$, X_1 表示隐藏层的输出 $f(W_1\mathbf{X} + b_1)$ 。MLP 所有的参数即各个层之间的连接权重及偏置,包括 W_1, b_1, W_2, b_2 。分割网络中的 MLP 通过权重共享卷积实现。

为处理点云模型无序性,保证模型对输入不同排序具有不变性,PointNet 通过应用对称函数来逼近定义上的一般函数,可表示为

$$f(\{x_1, x_2, \dots, x_n\}) = g(h(x_1), h(x_2), \dots, h(x_n)), \quad (3)$$

式中: h 通过 MLP 训练获得; f 的范围为 $2^{R^N} \rightarrow R$; h 的范围为 $R^N \rightarrow R^K$, N 为点云数量, K 为 Max pooling 层维度; g 为一个单变量函数和 Max pooling 的组合,表示为

$$g: \underbrace{R^K \times \dots \times R^K}_n \rightarrow R. \quad (4)$$

由于独立分割网络最上层为 softmax,而损失函数

对于最上层权重的梯度不再跟激活函数的导数相关,只跟输出值和真实值的差值成正比,此时收敛较快。另外,多分类交叉熵损失求导更简单,损失仅与正确类别的概率有关,而且损失对于 softmax 激活层的输入求导很简单。由于在分割独立训练阶段,真值点云分类标签 X_c 和预测点云类别标签 \hat{X}_c 之间存在直接的点对点对应关系,即 softmax 交叉熵,因此直接将 softmax 交叉熵函数 L_{ce} 作为分割网络的分割损失函数,可表示为

$$L_{ce}(X_c, \hat{X}_c) = \sum_{\substack{x \in X_c \\ \hat{x} \in \hat{X}_c}} [x \log(\hat{x}) + (1-x) \log(1-\hat{x})]. \quad (5)$$

2.3 联合优化损失函数设计

在联合优化训练期间,真值点云分类标签和预测点云的类别标签之间不存在点对点的对应关系。为了在训练中最小化模型误差,在分割损失函数的基础上引入优化分割损失,以便在匹配的点对之间传播语义信息。损失函数被用来估量模型的预测值与真实值之间不一致的程度,是模型对数据拟合程度的反映,它是一个非负实值函数,损失函数越小,代表模型的鲁棒性越好。3D-LMNet+PointNet 网络首先通过 3D-LMNet 重建网络,以图像为输入,生成预测点云模型,然后通过 PointNet 分割网络,以真值点云为输入,生成带有语义信息的预测点云模型。算法核心为联合优化损失函数,在联合优化阶段,将给定图像 I 作为输入,训练重建网络得到重建点云 \hat{X}_p ; 将 \hat{X}_p 作为输入,经过分割网络,训练得到带有类别标签的点云 \hat{X}_c , 同时在联合优化损失函数的作用下,改善重建和分割效果。

优化分割损失主要包括两个方面:前向分割损失和反向分割损失,如图 2 所示。错误的重建和错误的分割都会使网络受到相应的惩罚,因此将分割阶段的前向分割损失与反向分割损失的和作为分割阶段的损失函数。

1) 前向分割损失 ($L_{\text{seg-fwd}}$)

对于 X_p 中每一个点 i , 找到与之最邻近的点 $i' \in \hat{X}_p$, 并对应 L_{ce} 的类别标签, 将真值点云到预测点云

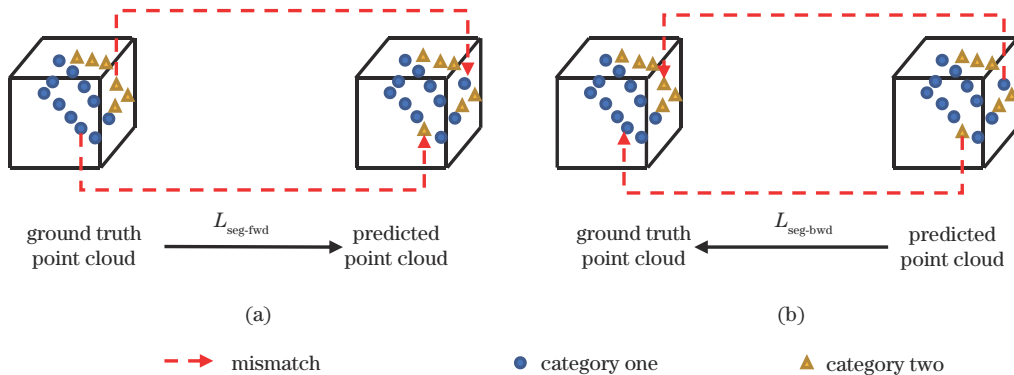


图 2 分割损失示意图。(a)前向分割损失;(b)反向分割损失

Fig. 2 Schematic diagram of segmentation loss. (a) Forward loss; (b) backward loss

所有点的交叉熵的平均值作为前向分割损失,定义为

$$L_{\text{seg-fwd}} = \frac{1}{N_p} \sum_{i \in X_p} L_{\text{ce}}(X_{ci}, \hat{X}_{ci}) \quad (6)$$

2) 反向分割损失 ($L_{\text{seg-bwd}}$)

对于 \hat{X}_p 中每一个点 i , 找到与之最邻近的点 $i' \in X_p$, 并对应 L_{ce} 的类别标签, 将预测点云到真值点云所有点的交叉熵的平均值作为反向分割损失, 定义为

$$L_{\text{seg-bwd}} = \frac{1}{N_p} \sum_{i \in \hat{X}_p} L_{\text{ce}}(X_{ci'}, \hat{X}_{ci}) \quad (7)$$

3) 优化分割损失 (L_{seg})

为在点对之间更好地传播语义信息, 将前向分割损失和反向分割损失的总和作为联合优化阶段最终的优化分割损失, 表示为

$$L_{\text{seg}} = L_{\text{seg-fwd}} + L_{\text{seg-bwd}} \quad (8)$$

4) 联合优化损失 (L)

为实现联合重建和分割两个任务来预测局部分割点云模型, 同时又从语义标签中传播信息来改善重建模型, 引入两个参数即重建影响因子 a 和分割影响因子 b , 构建联合优化的总损失, 表示为

$$L = aL_{\text{rec}} + bL_{\text{seg}} \quad (9)$$

3 分析与讨论

3.1 评价指标

为了检测评估结果, 将独立重建分割方法与联合优化方法进行比较和分析。

1) 重建部分

针对重建模型的评估, 将 CD 与 EMD (earth mover's distance) 作为重建评估标准。EMD^[19] 用来表征两个分布之间的距离, EMD 函数可以计算预测点云与真值点云之间的距离, 比较预测点云和真值点云的相似度。EMD 值越小, 预测点云与真值点云之间的相似度越高, 重建效果越好, 可表示为

$$d_{\text{EMD}}(X_p, \hat{X}_p) = \min_{\phi: X_p \rightarrow \hat{X}_p} \sum_{x \in X_p} \|x - \phi(x)\|_2 \quad (10)$$

式中: $\phi(x)$ 是 $X_p \rightarrow \hat{X}_p$ 的映射。

2) 分割部分

分割训练过程中将分割公式转换为按点分类的问题。评估指标按点计算, 对于类别 C 的每个形状 S , 均按以下方式来计算平均交并比 (mIoU): 对于类别 C 中的每种零件类型, 计算真值和预测点之间的交并比 (IoU)。如果真值和预测点的并集为空, 则将零件 IoU 计为 1。然后, 对 C 类中所有零件类型的 IoU 取平均值, 作为该形状的 mIoU。若要计算类别的 mIoU, 则对类别中所有形状的 mIoU 取平均值。由于真值和预测点之间没有对应关系, 因此使用类似于 2.3 节中描述的机制来计算前向和反向的 mIoU, 然后对它们进行平均, 最终得出 mIoU 的表达式为

$$M_{\text{IoU}}(X_c, \hat{X}_c) = \frac{1}{2N_c} \sum_i \frac{N_{ii}}{\sum_j N_{ij} + \sum_j N_{ji} - N_{ii}} + \frac{1}{2N_c} \sum_i \frac{\hat{N}_{ii}}{\sum_j \hat{N}_{ij} + \sum_j \hat{N}_{ji} - \hat{N}_{ii}} \quad (11)$$

式中: N_{ij} 是 X_c 中类别 i 和 \hat{X}_c 中正向点类别 j 的点数对应; 同理 \hat{N}_{ij} 是反向点类别 j 的点数对应; N_c 为类别的总数。

3.2 数据集

采用的数据集来自于 ShapeNet 数据集^[20], 训练测试样本来自于数据集中的 3 个类别, 即椅子、汽车和飞机, 共包含 7346 个模型, 并且每个模型都具有部分注释的真值。

3.3 参数指标选取

参数设置对神经网络模型的训练尤其重要, 合理的参数设置可以加快收敛, 提高模型的性能。对联合优化损失函数中的参数重建影响因子 a 和分割影响因子 b 做相关实验探究, 将分割影响因子 b 设置为 1, 重建影响因子 a 从 10 变化到 10^5 的过程中对评估指标 CD、EMD 和 mIoU 的实验结果如图 3 所示。

根据评估指标准则, CD、EMD 越小, 重建精度越高; mIoU 越大, 分割效果越佳。从图 3(a) 可看出, 在参数 a 从 10 变为 10^5 的过程中, CD 在 a 取 5×10^3 、 10^4 和 5×10^4 处取得最小值时, 重建点云与真值点云的点对之间距离最小, 重建精度比较高。从图 3(b) 可看出 EMD 在 a 取 10^4 处为最小值, 说明重建点云与真值点云相似度最高, 重建效果最好。从图 3(c) 可以观察到 a 在 10 和 10^4 处 mIoU 取得最大值, 分割效果最好。因此, 可以看出 a 为 10^4 时各个指标均可以达到最佳, 因此实验中 a 值取 10^4 , b 值取 1。

3.4 实验结果

实验应用 ShapeNet 数据集分别训练了 3 种目标类别: 飞机、汽车和椅子。针对 3D-LMNet+PointNet 网络和联合优化网络进行训练, 每个类别上的定量结果及重建指标值均缩放 100, 各个指标如表 1 所示。

由表 1 可知联合优化网络在重建和分割指标方面, 在飞机、汽车和椅子 3 个类别中的每个指标均优于独立训练的基准指标。整体上, 网络重建指标 CD 降低 0.38, EMD 降低 0.34, 分割指标提升了 4.23, 说明联合优化网络同时改善了重建和分割效果。由飞机、汽车和椅子的重建指标可知, 飞机和汽车的 CD 值相对椅子较小, 说明在神经网络训练中, 这两种类别的重建模型与真值的点对之间的距离相对较小, 错误的点相对椅子较少, 原因是椅子存在孔洞较多, 椅背易产生空洞纹理, 且椅子腿形状多样, 使神经网络学习训练的困难增加, 因此在重建过程中就会出现错误的特征重建, 存在错误的点云分布, 从而影响 CD 值。而飞机与汽车相对椅子具有更多的整体轮廓信息, 空洞较少, 出

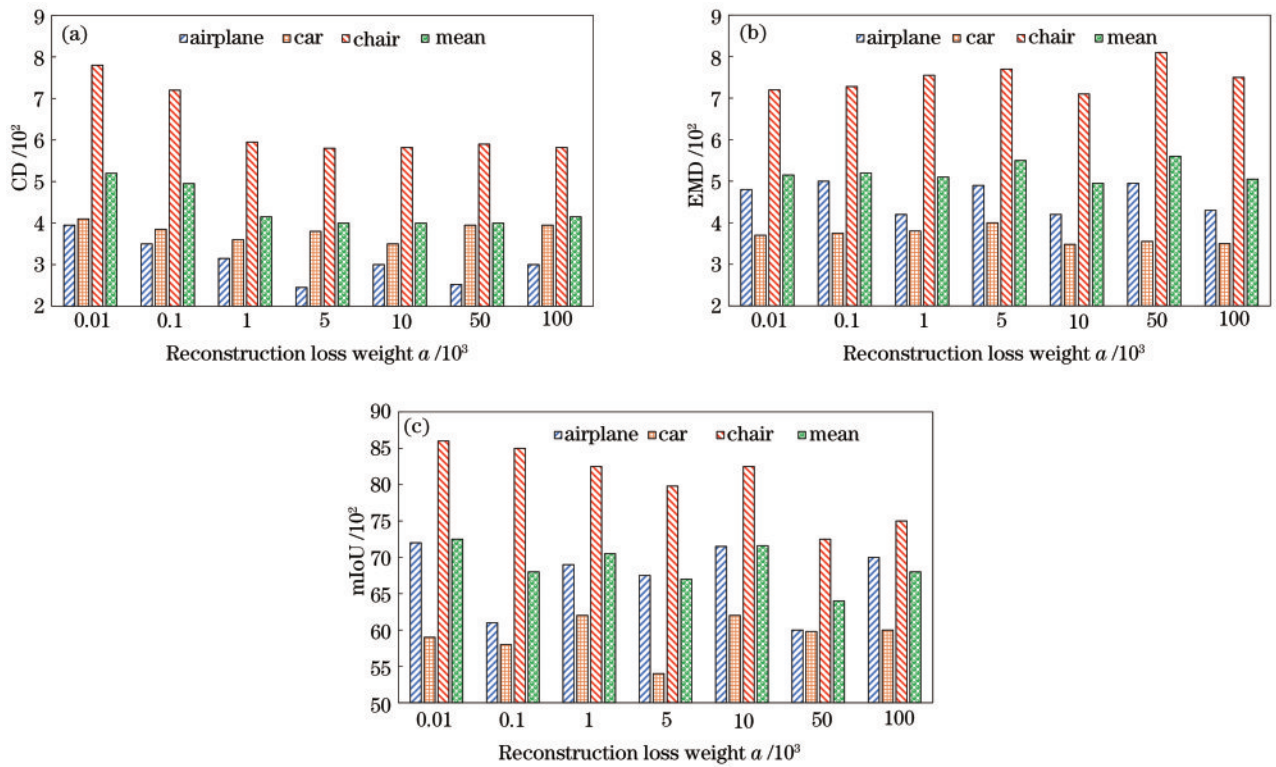


图3 重建影响因子 a 从 10^{-2} 变化到 10^2 时对CD、EMD和mIoU的影响。(a) CD; (b) EMD; (c) mIoU

Fig. 3 Effect of different a value on CD, EMD and mIoU. (a) CD; (b) EMD; (c) mIoU

表1 ShapeNet实验结果

Table 1 Experimental results on ShapeNet

Category	Metric	3D-LMNet+PointNet	Our model
Airplane	CD	3.34	3.01
	EMD	4.77	4.18
	mIoU	62.98	71.12
Car	CD	4.55	4.26
	EMD	4.10	3.85
	mIoU	60.01	61.68
Chair	CD	6.41	5.91
	EMD	8.02	7.83
	mIoU	79.15	82.03
Mean	CD	4.77	4.39
	EMD	5.63	5.29
	mIoU	67.38	71.61

现散点较少,所以飞机和汽车的CD值较椅子更低。在表1中发现汽车的EMD值比较低,原因是汽车结构集中,重建的点几乎分布在车身与车轮上,因此重建模型与真值模型最为相似。飞机和椅子结构比较分散,在重建过程中会出现不同程度的失真,导致点云分布不均,因此飞机与椅子的重建模型与真值点云模型差异较大,尤其是椅子类别。但在这3个类别中椅子的分割指标比较高,原因是椅子的各个结构大小比较均匀,从真值点云中也可以看出点云分布比较均匀,而飞机中涡轮相对机身比较小,汽车车轮相对车身较小,因此椅子分割效果更好。

为进一步验证联合优化网络模型效果,使用3D-LMNet+PointNet网络和联合优化后两个网络对ShapeNet数据集进行实验,图4~图6为实验结果。

图4中,6种飞机可清晰重建出机身、机翼、机尾和涡轮四部分,并以不同颜色呈现。由于机身比例较大,点云分布较多,因此两种网络在机身重建分割效果相近,但针对(1)至(6)飞机的机尾部分和机翼部分,联合优化网络能够将细节重建得更加清楚,更接近于真值点云,说明联合优化过程中,分割网络传播的语义信息改善了重建效果。(1)和(4)中飞机的涡轮均位于机翼下方,各有2个。由于涡轮比较小,均出现不同程度的失真,但是在联合优化网络损失函数的作用下,能更好地重构出涡轮部件。在(6)中,由于机尾和机翼比较靠近,导致重建点云分布不均,3D-LMNet+PointNet网络没有很好地分割出这两部分,但是在联合优化网络的重建效果中得到改善。

在图5所示的汽车类别中,由于汽车自身属性,车身、车轮和车盖部件比较集中,所以(1)至(6)中汽车几乎都能够重建出这三个部分,并得到很好的分割效果。但(1)汽车车身轮廓本为方形,3D-LMNet+PointNet网络重建的模型丢失了这一重要特征,而联合优化网络重建分割的模型更接近于真值点云的形状,整体来看汽车的重建结果也优于3D-LMNet+PointNet网络的重建结果。同样地,(6)中汽车的联合优化网络的重建结果相对于前一个网络更接近于真值点云的车身形状。(2)和(3)两个汽车的形状比较规则,所以两个网络

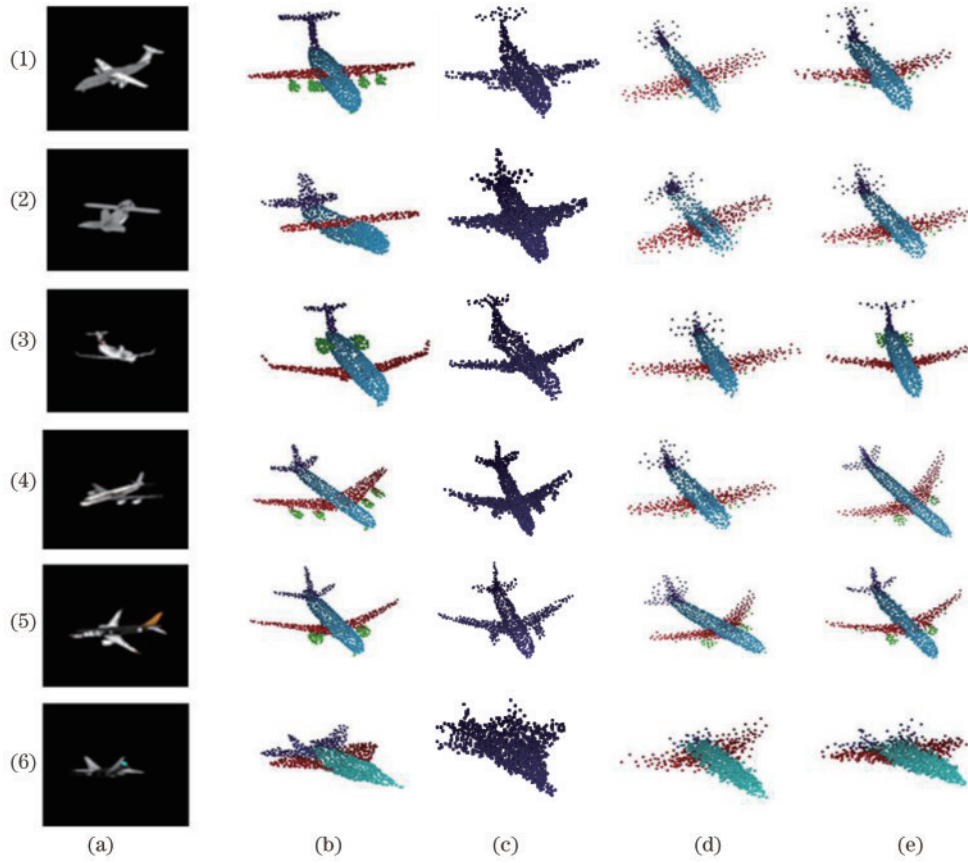


图 4 ShapeNet 中飞机的定性结果。(a)输入图像;(b)目视真值;(c) 3D-LMNet;(d) 3D-LMNet+PointNet;(e)所提网络
 Fig. 4 Qualitative results on airplanes from ShapeNet. (a) Input image; (b) ground truth; (c) 3D-LMNet; (d) 3D-LMNet+PointNet; (e) our model

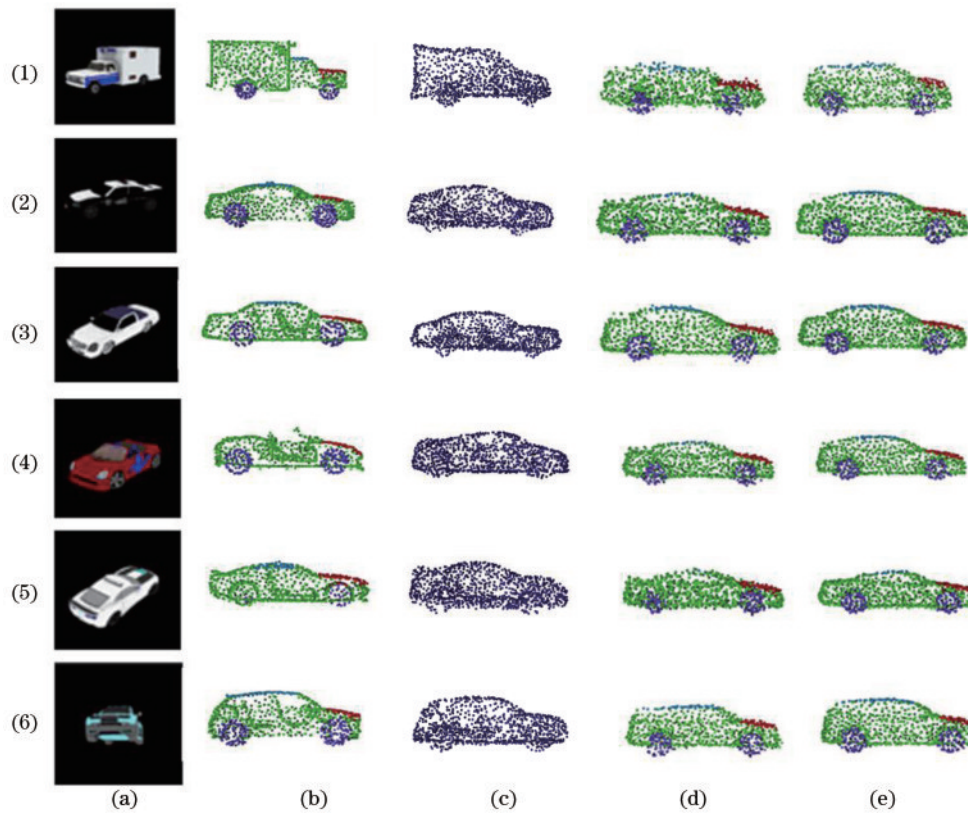


图 5 ShapeNet 中汽车的定性结果。(a)输入图像;(b)目视真值;(c) 3D-LMNet;(d) 3D-LMNet+PointNet;(e)所提网络
 Fig. 5 Qualitative results on cars from ShapeNet. (a) Input image; (b) ground truth; (c) 3D-LMNet; (d) 3D-LMNet+PointNet; (e) our model

的重建分割效果相差无几。但是(4)中跑车无论是在独立训练中,还是在联合优化网络中,均错误地重建出车顶和车身形状。

图 6 中,(1)和(3)两把椅子在联合优化网络中重建的椅子腿部分更加清晰,但是仍存在异常点,未能识

别出转椅的具体轮子数量,(3)中腿部的架子点云也比较模糊。但是(5)和(6)中联合优化网络成功重建出椅子把手细节。从这 3 个类别整体来看,联合优化网络更好地捕获了输入图像中的细节,重建出更多的部位,并且点云分布更为均匀。

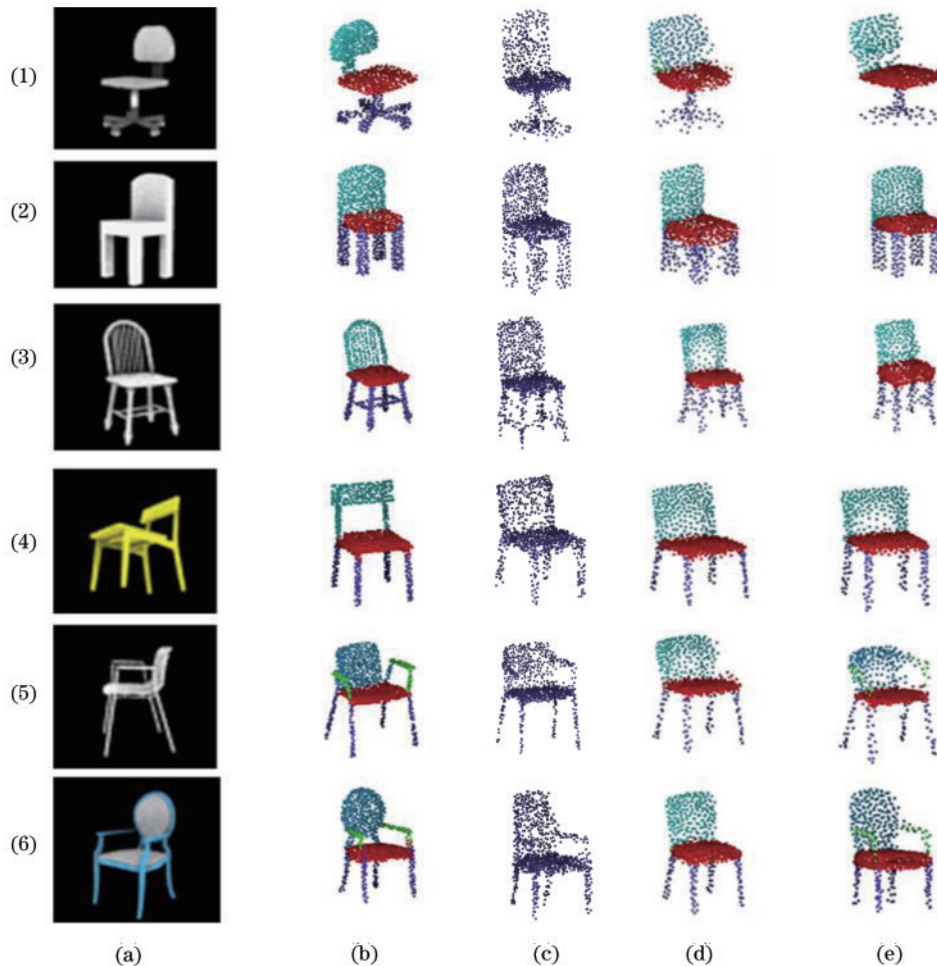


图 6 ShapeNet 中椅子的定性结果。(a)输入图像;(b)目视真值;(c) 3D-LMNet;(d) 3D-LMNet+PointNet;(e)所提网络
Fig. 6 Qualitative results on chairs from ShapeNet. (a) Input image; (b) ground truth; (c) 3D-LMNet; (d) 3D-LMNet+PointNet; (e) our model

4 结 论

提出 PointNet 与 3D-LMNet 结合的单幅图像三维点云语义重建方法,在直接结合重建和分割网络的基础上,引入联合优化损失函数,使信息能够在两个任务之间传播,使得网络预测出局部分割点云,并根据局部分割点云的标签信息对重建结果进行优化,从而生成更高精度的零件重构,同时还提高了分割精度。实验中对 ShapeNet 数据集的定量和定性评估证明,与当前最先进的重构和分割网络相比,此方法不仅能预测出模型的详细结构部件信息,还可以根据分割点云的语义信息来改善重建的点云模型,拥有更好的重建效果。

参 考 文 献

- [1] 佟帅, 徐晓刚, 易成涛, 等. 基于视觉的三维重建技术综述[J]. 计算机应用研究, 2011, 28(7): 2411-2417.
Tong S, Xu X G, Yi C T, et al. Overview on vision-based 3D reconstruction[J]. Application Research of Computers, 2011, 28(7): 2411-2417.
- [2] 郑远攀, 李广阳, 李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用, 2019, 55(12): 20-36.
Zheng Y P, Li G Y, Li Y. Survey of application of deep learning in image recognition[J]. Computer Engineering and Applications, 2019, 55(12): 20-36.
- [3] 杨斌, 钟金英. 卷积神经网络的研究进展综述[J]. 南华大学学报(自然科学版), 2016, 30(3): 66-72.
Yang B, Zhong J Y. Review of convolution neural network[J]. Journal of University of South China (Science

- and Technology), 2016, 30(3): 66-72.
- [4] 陈加, 张玉麒, 宋鹏, 等. 深度学习在基于单幅图像的物体三维重建中的应用[J]. 自动化学报, 2019, 45(4): 657-668.
Chen J, Zhang Y Q, Song P, et al. Application of deep learning to 3D object reconstruction from a single image [J]. Acta Automatica Sinica, 2019, 45(4): 657-668.
- [5] 董鹏辉, 柯良军. 基于图像的三维重建技术综述[J]. 无线电通信技术, 2019, 45(2): 115-119.
Dong P H, Ke L J. Overview of 3D reconstruction techniques based on images[J]. Radio Communications Technology, 2019, 45(2): 115-119.
- [6] Choy C B, Xu D F, Gwak J, et al. 3D-R2N2: a unified approach for single and multi-view 3D object reconstruction[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9912: 628-644.
- [7] Wu J J, Wang Y F, Xue T F, et al. Marmnet: 3D shape reconstruction via 2.5 D sketches[EB/OL]. (2017-11-08) [2021-05-07]. <https://arxiv.org/abs/1711.03129>.
- [8] Girdhar R, Fouhey D F, Rodriguez M, et al. Learning a predictable and generative vector representation for objects[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9910: 484-499.
- [9] Fan H Q, Su H, Guibas L. A point set generation network for 3D object reconstruction from a single image [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2463-2471.
- [10] Mandikal P, Navaneet K L, Agarwal M, et al. 3D-LMNet: latent embedding matching for accurate and diverse 3D point cloud reconstruction from a single image [EB/OL]. (2018-07-20)[2021-05-03]. <https://arxiv.org/abs/1807.07796>.
- [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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [12] Charles R Q, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [EB/OL]. (2017-06-07)[2021-05-03]. <https://arxiv.org/abs/1706.02413>.
- [13] 朱育正, 张亚萍, 冯乔生. 基于深度学习的单视图彩色三维重建[J]. 激光与光电子学进展, 2021, 58(14): 1410010.
Zhu Y Z, Zhang Y P, Feng Q S. Colorful 3D reconstruction from single image based on deep learning [J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410010.
- [14] 张爱武, 刘路路, 张希珍. 道路三维点多特征卷积神经网络语义分割方法[J]. 中国激光, 2020, 47(4): 0410001.
Zhang A W, Liu L L, Zhang X Z. Multi-feature 3D road point cloud semantic segmentation method based on convolutional neural network[J]. Chinese Journal of Lasers, 2020, 47(4): 0410001.
- [15] 胡涛, 李卫华, 秦先祥. 基于多层深度特征融合的极化合成孔径雷达图像语义分割[J]. 中国激光, 2019, 46(2): 0210001.
Hu T, Li W H, Qin X X. Semantic segmentation of polarimetric synthetic aperture radar images based on multi-layer deep feature fusion[J]. Chinese Journal of Lasers, 2019, 46(2): 0210001.
- [16] 徐聪, 王丽. 基于改进DeepLabv3+网络的图像语义分割方法[J]. 激光与光电子学进展, 2021(16): 1610008.
Xu C, Wang L. Image semantic segmentation method based on improved DeepLabv3+ network[J]. Laser & Optoelectronics Progress, 2021(16): 1610008.
- [17] 鲍海龙, 万敏, 刘忠祥, 等. 基于区域自我注意力的实时语义分割网络[J]. 激光与光电子学进展, 2021, 58(8): 0810018.
Bao H L, Wan M, Liu Z X, et al. Real-time semantic segmentation network based on regional self-attention[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810018.
- [18] Song S R, Yu F, Zeng A, et al. Semantic scene completion from a single depth image[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 190-198.
- [19] Rubner Y, Tomasi C, Guibas L J. A metric for distributions with applications to image databases[C]//Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271), January 7, 1998, Bombay, India. New York: IEEE Press, 1998: 59-66.
- [20] Chang A X, Funkhouser T, Guibas L, et al. ShapeNet: an information-rich 3D model repository[EB/OL]. (2015-12-09)[2021-05-07]. <https://arxiv.org/abs/1512.03012>.