第 58 卷 第 14 期/2021 年 7 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于多任务监督学习的实时室内布局估计方法

黄荣泽，孟庆浩，刘胤伯*

天津大学电气自动化与信息工程学院，天津大学机器人与自主系统研究所，天津市过程检测与控制重点实验室，天津 300072

摘要 室内布局估计是计算机视觉领域的研究热点之一，在三维重建、机器人导航和虚拟现实等方面具有广泛的应用。目前室内布局估计的解决方案存在实时性较差、计算量较大等问题。针对这些问题，本文提出了一种基于多任务监督的轻量卷积网络，该网络模型基于编码器-解码器结构，使用室内边缘热图与平面语义分割实现多任务监督学习。此外本文对卷积模块进行了改进，使用 1×1 卷积替换了 1×3、3×1 卷积，在保证模型精度的情况下提升了网络的实时性。在公共数据集 LSUN 上进行实验，结果表明，本文方法具有良好的实时性和准确性。

关键词 图像处理；卷积神经网络；室内布局估计；多任务监督；语义分割

中图分类号 TP391.4 **文献标志码** A **doi**：10.3788/LOP202158.1410023

Real-Time Indoor Layout Estimation Method Based on Multi-Task Supervised Learning

Huang Rongze, Meng Qinghao, Liu Yinbo*

School of Electrical and Information Engineering, Institute of Robotics and Autonomous Systems, Tianjin Key Laboratory of Process Detection and Control, Tianjin University, Tianjin 300072, China

Abstract Indoor layout estimation is one of the important research topics in the field of computer vision, and it is widely used in three-dimensional reconstruction, robot navigation, and virtual reality. The current indoor layout estimation solutions have problems such as poor real-time performance and large calculations. To deal with these problems, this paper proposes a lightweight convolutional network based on multi-task supervision. The network model is based on the encoder-decoder structure and uses indoor edge heatmaps and planar semantic segmentation to achieve multi-task supervised learning. In addition, this paper modifies the convolution module, replaced 1×3 and 3×1 convolution with 1×1 convolution, which improves the real-time performance of the network while ensuring the accuracy of the model. The experimental results conducted on the public dataset LSUN show that the proposed method has good real-time performance and accuracy.

Key words image processing; convolutional neural network; indoor layout estimation; multi-task supervised learning; semantic segmentation

OCIS codes 100.4996; 100.2000; 100.5010

1 引 言

近年来，室内布局估计已成为计算机视觉领域热点问题之一。其目的是通过一张室内的二维图像估计出室内的三维空间布局信息，即直接获取室内地面、墙面和天花板平面的位置。这一任务可为室内场景的三维重建[1]、虚拟现实、增强现实[2]、机器人导航与定位等应用提供有效的信息[3]。室内布局估计这一任务的难点在于室内场景总是存在大量的物体，这些物体会对室内的地面、墙面或者天花板产

收稿日期：2020-08-28；**修回日期**：2020-09-21；**录用日期**：2020-09-30

基金项目：国家自然科学基金(61573252)、国家重点研发计划项目(2017YFC0306200)

通信作者：*liuyinbo@tju.edu.cn

生遮挡；另外，室内环境还会受到光照、物体表面纹理等多种因素的影响。

已有的室内布局估计方法主要分为两类：一类是非深度学习的传统方法，另一类是基于深度学习的方法。传统方法通常使用消失点等信息生成室内候选布局，并对候选布局进行排序获得最终结果。Hedau 等[4]最先提出室内布局估计的问题并给出了解决方案。他们先使用传统图像处理方法获取消失点，进而通过消失点在图像上的均匀采样获取候选布局，最后利用低层次图像特征筛选候选布局并获得结果。Wang 等[5]引入了多种图像特征，并使用复合特征对候选布局进行评价。近年来，随着深度学习的飞速发展，研究人员开始使用深度学习解决室内布局估计这一问题。Mallya 等[6]搭建了一个全卷积神经网络(FCN)[7]，FCN 是将卷积神经网络(CNN)最后的全连接层替换成卷积层，实现像素级别的分类[8]，他们使用 FCN 提取室内的平面边缘，再进一步生成室内候选布局方案。Lee 等[9]基于编码器-解码器架构搭建了一个网络，用来预测室内的关键角点热图和室内布局分类，接着使用室内布局分类的结果对关键角点热图进行筛选，进而生成最终的室内布局估计。Dasgupta 等[10]使用以 VGG[11]为骨干的 FCN 生成室内平面的语义分割图，并提出了一种基于室内布局合理性的后处理框架生成室内布局估计。相比于基于深度学习的方法，传统方法往往基于局部特征而缺失语义上下文信息，其室内布局估计的精度较低；而深度学习模型更加强调从庞大的数据中通过多层神经元组织自动学习特征[12]，基于深度学习的方法通过深度卷积神经网络提取全局语义特征，大大提升了室内布局估计的精度。上述基于深度学习的方法大多基于一种特征，或者是在后处理阶段融合多个特征，前者准确性有所欠缺，而后者在后处理阶段需要花费额外的时间。

针对以上不足，本文提出了一种基于多任务监督的轻量卷积神经网络。本文网络基于编码器-解码器结构，其骨干网络为高效残差分解卷积网络(ERFNet)[13]。本文对网络的子模块进行了改进优化，将 3×1 与 1×3 的连续卷积层替换为 1×1 的卷积层，在保证运算精度的前提下，减少了网络的参数量，提升了网络的计算速度。此外本文网络使用了平面边缘热图与室内平面语义分割图共同监督学习，提升了室内布局估计结果的准确性。在大规模场景理解挑战(LSUN)[14]数据集上进行实验，结果表明，本文方法具备良好的实时性，且室内布局估计的准确性优于现有研究成果。

2 基于多任务监督的轻量卷积神经网络

2.1 多任务监督的轻量卷积神经网络总体结构

本文提出的网络结构如图 1 所示，本文网络模型基于编码器-解码器架构。编码器部分通过不断对原始图像进行下采样和卷积操作来提取图像全局语义特征；而解码器部分用于对编码器提取的特征张量进行解码生成不同的特征图，本文网络具有两个独立的解码器，其中边缘解码器用于提取室内的平面边缘热图，分割解码器用于提取室内布局的平面语义分割图，两解码器共享同一个特征张量，互相分享，互相补充其学习到的信息，并解码为不同的全局特征；网络结构的最后将这提取的特征图进行叠加，并通过融合模块获得室内布局估计的最终语义分割结果。

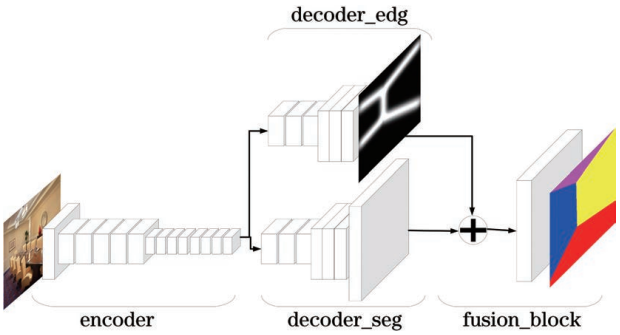

图 1 多任务监督的轻量卷积神经网络总体结构图

Fig. 1 General structure of multi-task supervised lightweight convolutional neural network

编码器部分由下采样模块(DSBlock)和改进的轻量分解卷积模块(LFBlock)组成；两个解码器除了输出层的通道数分别为 1 和 15 以外，其余结构完全一致，这两个分支由上采样模块(USBlock)、LFBlock 和转置卷积层组成。Fusion_block 由一个 1×1 卷积层和一个 SoftMax 分类器组成，其中 1×1 卷积用于将不同通道的特征图进行融合，而 SoftMax 分类器用于对每个像素点进行语义分类。

编码器与解码器的具体参数见表 1 和表 2。表中 Layer ID 为不同层的编号；Block type 是使用的模块类型；Dilation 是使用的空洞卷积[15]的空洞数，在不增加参数量的基础上扩大卷积层的感受野[16]；Dropout 是每一层随机丢弃的参数的比例[17]，用于缓解网络训练中的过拟合情况；Output channels 是该层输出的特征张量的通道数；Output resolution

表 1　编码器参数表

Table 1　Parameters of the encoder

Layer ID	Block type	Dilation	Dropout	Output channels	Output resolution
1	DSBlock	–	–	16	128×128
2	DSBlock	–	–	64	64×64
3-7	LFBlock	1	0.3	64	64×64
8	DSBlock	–	–	128	32×32
9	LFBlock	2	0.3	128	32×32
10	LFBlock	4	0.3	128	32×32
11	LFBlock	8	0.3	128	32×32
12	LFBlock	16	0.3	128	32×32
13	LFBlock	2	0.3	128	32×32
14	LFBlock	4	0.3	128	32×32
15	LFBlock	8	0.3	128	32×32
16	LFBlock	16	0.3	128	32×32

表 2　解码器参数表

Table 2　Parameters of the decoder

Layer ID	Block type	Dilation	Dropout	Output channels	Output resolution
1	USBlock	–	–	64	64×64
2-3	LFBlock	1	0.3	64	64×64
4	USBlock	–	–	16	128×128
5-6	LFBlock	1	0.3	16	128×128
7	Deconvolution	–	–	1/15	256×256

是该层输出的特征张量的分辨率。

一维非瓶颈层(Non-bottleneck-1D)、轻量分解卷积模块(LFBlock)、下采样模块(DSBlock)和上采样模块(USBlock)模块的结构如图 2 所示。本文中的 LFBlock 是基于 ERFNet[13] 中的 Non-bottleneck-1D 模块改进的。Non-bottleneck-1D 模块相比于 Non-bottleneck 模块[13]和 Bottleneck 模块[18]，其参数量更少且精度接近，同时非线性更强。而在室内

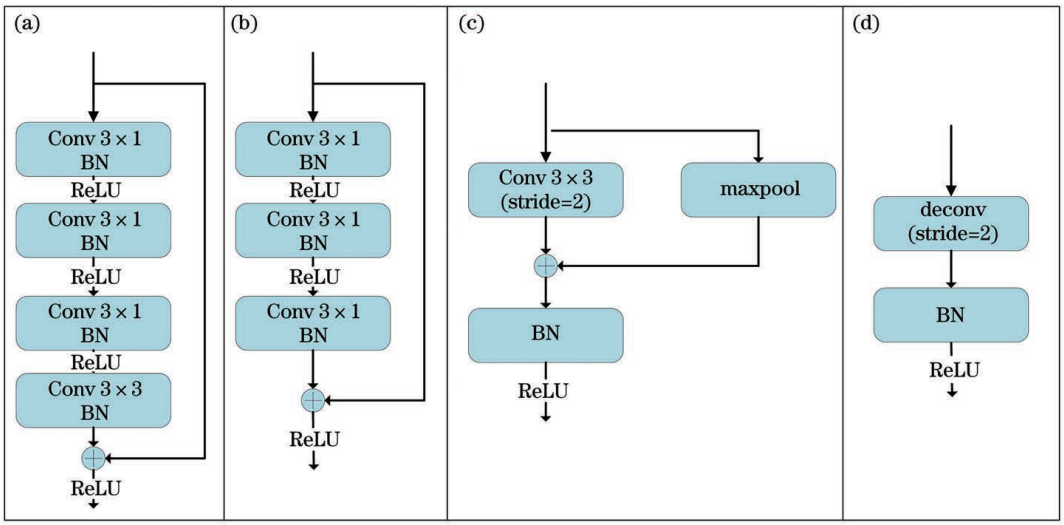

图 2　多种卷积模块结构图。(a)一维非瓶颈层；(b)轻量分解卷积模块；(c)下采样模块；(d)上采样模块

Fig. 2　Structures of various convolution modules. (a) Non-bottleneck-1D; (b) LFBlock; (c) DSBlock; (d) USBlock

布局估计问题上，使用本文提出的 LFBlock 模块的网络实时性更好。LFBlock 将 Non-bottleneck-1D 中连续的 3×1 卷积和 1×3 卷积替换为一个 1×1 卷积，进一步减少了参数量。有研究者[19]提出单独使用最大池化层或者单独使用大步长卷积都会丢失一部分特征信息，因此 DSBlock 模块中采用了 Max Pooling 与步长为 2 的卷积混合的方式。USBlock 采用了步长为 2 的 Deconvolution 操作进行尺度的扩增。上述的所有模块的卷积层后都采用了批标准化操作（BN）[20]，这一操作对网络每一层的输出进行归一化处理，降低了网络在反向传播中对大尺度参数变化的敏感性，加快了网络收敛速度。

2.2　多任务监督网络设计

多任务监督学习是指多个相关的任务同时进行学习，梯度同时反向传播。不同任务在学习各自的领域信息的同时，通过浅层网络的共享表示互相分享、互相补充学习到的信息，提升网络模型的泛化能力。已有的研究[21]证明，引入边缘特征可提高预测语义分割图的准确性。因此，本文网络在进行室内语义分割图预测的同时增加了预测室内边缘热图的任务。本文网络的边缘解码器与分割解码器共享一个编码器，实现了浅层网络的参数共享，提升了网络的泛化能力和语义分割图预测的准确性。

边缘热图与语义分割图数据的标注示例如图 3 所示。本文使用热图的形式对平面边缘进行标注，每个点的像素值反映了当前位置为边缘的概率，像素的位置越接近平面边缘，其概率值越接近 1，远离边缘的概率值为 0。室内平面的语义被分为 5 类，其语义编号为 0～4，分别代表地面、中间墙壁、右侧墙壁、左侧墙壁和天花板，当室内只有两面墙壁时，将其标注为中间墙壁与右侧墙壁。在网络训练的时候采用平面边缘热图和平面语义分割图作为共同监督数据，而在网络预测时直接生成室内平面的语义分割图。

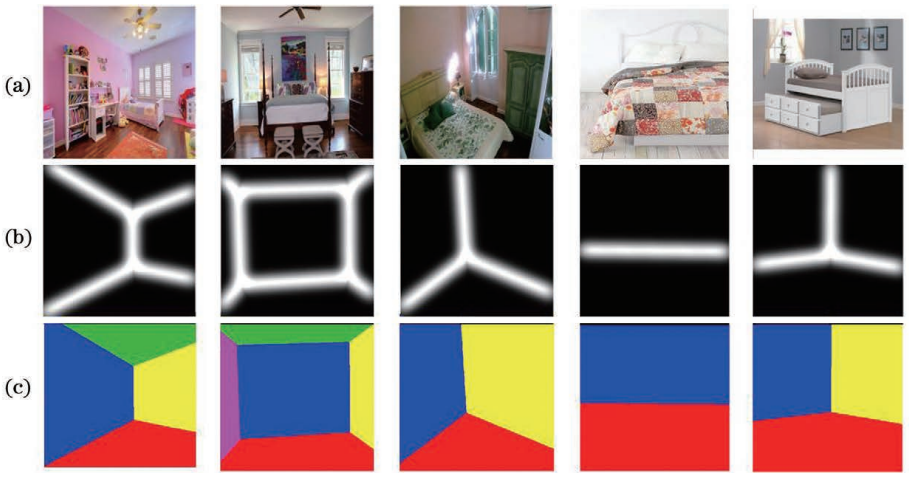

图 3　数据标注示例。(a)原始图像；(b)平面边缘标注热图；(c)语义分割图标注可视化结果

Fig. 3　Examples of labels. (a) Original images; (b) edge annotation heat maps; (c) visualization result of semantic segmentation labels

2.3　损失函数设计

训练过程中的整体损失函数公式为

$$l_{\text{total}}=l_{\text{seg}}+\lambda\cdot l_{\text{edge}},\tag{1}$$

式中：l_{total} 代表整体的损坏函数；l_{edge} 和 l_{seg} 分别代表边缘热图预测损失值和语义分割图预测损失值；λ 为边缘热图预测的损失值所占权重，文中该权重取值为 0.5。边缘热图和语义分割图训练时使用的损失函数分别是二分类交叉熵函数和多分类交叉熵函数，其定义为

$$l_{\text{edge}}=-[y_i\cdot\log_2(p_i)+(1-y_i)\cdot\log_2(1-p_i)],\tag{2}$$

$$l_{\text{seg}}=-\log_2\left[\frac{\exp(p_{\text{class}})}{\sum_j\exp(p_j)}\right],\tag{3}$$

式中：y_i 代表样本 i 的标注值；p_i 表示样本 i 的预测值；p_{class} 为该样本通过网络后其真值节点输出的概率值；p_j 是其他类节点输出的概率值。

3　实验结果与分析

3.1　数据集与训练细节

目前大多数研究者均采用公共数据集 LSUN 进行室内布局估计算法的训练和验证，该数据集包含 4000 张训练集图片，394 张验证集图片，还有

1000 张测试集图片。由于 LSUN 数据集仅含有室内的关键角点热图和平面边缘热图，因此需要生成室内布局的语义分割图。本文首先通过室内关键角点热图获取每个关键角点的坐标；其次，按照室内的拓扑结构对关键角点进行连接，获取室内布局线框图；然后，利用线框图对图像进行划分，并分别为不同区域进行了标注；最后，生成室内布局的语义分割图数据集。

使用 PyTorch1.1 并基于 Python 语言实现本文所提方法，并采用 Intel Xeon E5430 和 NVIDIA GTX 980Ti GPU 对其进行了训练。本文网络的原始输入图像、边缘热图和室内语义分割图的分辨率均为 256×256。由于 LSUN 数据集规模相对较小，因此在训练时对原始训练数据进行了数据扩增，采用的数据扩增方法包括图像水平镜像翻转和图像随机裁剪操作，随机裁剪的范围为 0～64。整体训练样本的迭代次数为 200，每次迭代所用的样本量为 28，模型在训练中采用了随机梯度下降优化器(SGD)，初始学习率设置为 0.001，每 20 个 epoch 学习率变为原来的 0.1。

3.2 实验结果及分析

针对室内布局估计这一问题，研究人员[4,6,9-10,22-24]一般采用两个标准的评价指标，即角点误差(CE)和像素误差(PE)。CE 反映的是每个关键角点的预测坐标与真实坐标之间的欧氏距离与图像的对角线长度的比值；PE 表示预测的语义分割结果与真实的语义分割结果之间的像素误差。CE 和 PE 的定义为

$$E_{Ci}=\frac{\sqrt{(x_{\text{pred}}-x_{\text{gt}})^2+(y_{\text{pred}}-y_{\text{gt}})^2}}{\sqrt{H^2+W^2}},\tag{4}$$

$$E_{\text{C}}=\frac{\sum E_{Ci}}{n}\times 100\%,\tag{5}$$

$$E_{\text{P}}=1-\frac{\sum_{i=0}^{k}p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}},\tag{6}$$

式中：E_{C} 代表角点误差；E_{P} 代表像素误差；$(x_{\text{pred}},y_{\text{pred}})$和$(x_{\text{gt}},y_{\text{gt}})$分别代表室内关键角点预测坐标值与真实坐标值；$H$ 和 W 表示图像的高和宽；p_{ij} 是真实类别为 i 类却被预测为 j 类的像素点个数；p_{ii} 代表预测类别与真实类别均为 i 的像素点个数；k 为语义分割的类别数，本文中为 5。

此外，为评估多任务监督与改进 LFBlock 对室内平面分割任务的精度和速度的影响，本文还引入了语义分割的评价指标，包括类别平均像素准确率(MPA)、平均交互比(MIoU)、频权交互比(FWIoU)、模型大小和执行时间。MPA、MIoU 和 FWIoU 的表示为

$$M_{\text{MPA}}=\frac{1}{k}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}},\tag{7}$$

$$M_{\text{MIoU}}=\frac{1}{k}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}(p_{ji}-p_{ii})},\tag{8}$$

$$F_{\text{FWIoU}}=\frac{1}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}}\sum_{i=0}^{k}\frac{p_{ii}\sum_{j=0}^{k}p_{ij}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}(p_{ji}-p_{ii})}。\tag{9}$$

为了获得室内关键角点的坐标，需要对室内平面语义分割图进行简单的后处理。首先，使用尺寸为 3×3 的十字形结构元对语义分割图的不同区域进行膨胀操作，使相邻的语义区域相交；然后，计算不同语义区域的交集，其中图像内部的交集通过计算三个相邻区域相交像素点获取，而图像边缘的交集通过计算两个相邻区域及图像边缘的相交像素点获取；最后，通过计算每个区域的质心坐标获取室内关键角点的坐标。

多任务监督和改进的 LFBlock 模块对网络性能的影响如表 3 所示。加入多任务监督机制后，使用 Non-bottleneck-1D 模块网络的 MPA、MIOU、FWIOU 分别提升了 3.27%、1.39%、1.84%，CE 和 PE 分别提升了 0.37% 和 0.22%；而使用 LFBlock 模块的网络的 MPA、MIOU 与 FWIOU 分别提升了 4.52%、2.49%、2.05%，CE 和 PE 分别提升了 0.78% 和 0.81%。证明了多任务监督机制对室内平面语义分割的准确度和室内布局估计的准确度均有所提升。此外，从表 3 可以看出，使用 LFBlock 对最终结果的准确度影响不大，但是可以使模型大小减少 30%，执行时间减少 10%，提升了模型的实时性。

下面将对本文方法在 CE 和 PE 两个指标上进行量化分析。本文方法和 7 种主流方法在 LSUN 数据集上的测试结果对比如表 4 所示。从表中可以看出，本文方法的 CE 和 PE 指标分别比最好的文献[9]结果提升了 0.04% 和 0.81%。基于本文方法的室内布局可视化结果如图 4 所示。从图中可以看

表 3　模型性能评估表

Table 3　Model performance evaluation

Add multi-task supervised?	Use LFBLock?	MPA /%	MIOU /%	FWIOU /%	CE /%	PE /%	Size /MB	Time /ms
No	Yes	77.44	66.42	71.97	7.04	9.86	6.2	43.26
No	No	77.76	66.65	72.01	6.92	9.69	8.8	48.02
Yes	No	81.03	68.04	73.85	6.55	9.47	8.8	47.92
Yes	Yes	81.96	68.91	74.02	6.26	9.05	6.2	43.13

表 4　不同方法在 LSUN 数据集上的性能对比

Table 4　Performance comparison of different methods on LSUN dataset

Method	CE /%	PE/%
Ref. [4]	15.48	24.23
Ref. [6]	11.02	16.71
Ref. [22]	10.13	14.82
Ref. [23]	8.70	12.49
Ref. [10]	8.20	10.63
Ref. [24]	7.95	9.31
Ref. [9]	6.30	9.86
Proposed	**6.26**	**9.05**

出，本文方法可以很好地区分出室内的不同平面，且提取的室内边缘比较整齐，对于被物体遮挡的平面边缘也能很好地区分。

4　结　　论

本文提出了一种基于多任务监督的轻量卷积神经网络，采用多种特征对网络进行共同监督训练，提升了室内布局估计结果的准确性。此外，本文网络修改了卷积模块结构，减少了网络卷积模块的参数量，在不损失模型精度的前提下提升了网络的实时性。本文方法在公开数据集 LSUN 上取得了良好的效果，CE 为 6.26%，PE 为 9.05%，其预测精度优于其他室内布局估计方法，同时具有良好的实时性。本文方法依赖于大量数据集及数据标注的准确性，

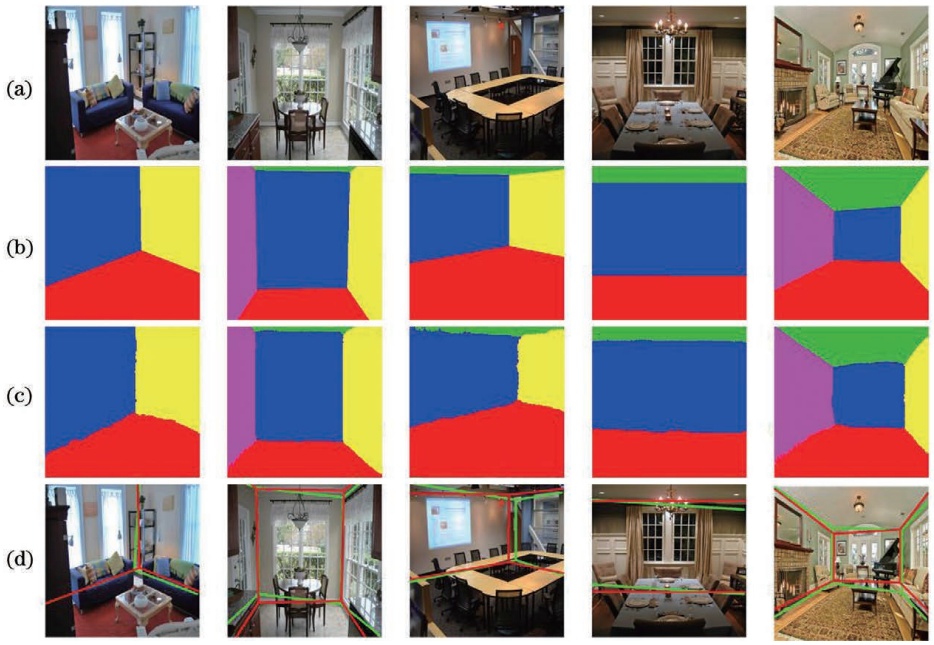

图 4　本文网络模型的可视化结果。(a)原始图像；(b)语义分割真值图；(c)本文方法语义分割预测图；(d)本文方法估计布局与真实布局对比图(绿色为估计布局，红色为真实布局)

Fig. 4　Visualization results of the proposed network model. (a) Original images; (b) semantic segmentation ground truth maps; (c) semantic segmentation prediction maps of the proposed method; (d) comparison maps between the estimated layouts of the proposed method and the real layouts (green is the estimated layout, red is the real layout)

当数据量不足或标注有误时会影响本文的实验结果，在未来的工作中考虑利用虚拟数据集等方法扩增室内布局估计的样本量。

参 考 文 献

[1] Izadinia H, Shan Q, Seitz S M. IM2CAD[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2422-2431.

[2] Liu C X, Schwing A G, Kundu K, et al. Rent3D: floor-plan priors for monocular layout estimation [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3413-3421.

[3] Hedau V, Hoiem D, Forsyth D. Thinking inside the box: using appearance models and context based on room geometry [M] // Daniilidis K, Maragos P, Paragios N. Computer vision-ECCV 2010 Lecture notes in computer science. Heidelberg: Springer, 2010, 6316: 224-237.

[4] Hedau V, Hoiem D, Forsyth D. Recovering the spatial layout of cluttered rooms[C]//2009 IEEE 12th International Conference on Computer Vision, September 29-October 2, 2009, Kyoto, Japan. New York: IEEE Press, 2009: 1849-1856.

[5] Wang H Y, Gould S, Koller D. Discriminative learning with latent variables for cluttered indoor scene understanding[M]//Daniilidis K, Maragos P, Paragios N. Computer vision-ECCV 2010 Lecture notes in computer science. Heidelberg: Springer, 2010, 6314: 497-510.

[6] Mallya A, Lazebnik S. Learning informative edge maps for indoor scene layout prediction [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 936-944.

[7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.

[8] Dai P Q, Ding L X, Liu L J, et al. Tree species identification based on FCN using the visible images obtained from an unmanned aerial vehicle[J]. Laser & Optoelectronics Progress, 2020, 57(10): 101001.
戴鹏钦, 丁丽霞, 刘丽娟, 等. 基于 FCN 的无人机可见光影像树种分类[J]. 激光与光电子学进展, 2020, 57(10): 101001.

[9] Lee C Y, Badrinarayanan V, Malisiewicz T, et al. RoomNet: end-to-end room layout estimation[C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4875-4884.

[10] Dasgupta S, Fang K, Chen K, et al. DeLay: robust spatial layout estimation for cluttered indoor scenes [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 616-624.

[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2020-08-28]. https://arxiv.org/abs/1409.1556.

[12] Wang E D, Qi K, Li X P, et al. Semantic segmentation of remote sensing image based on neural network[J]. Acta Optica Sinica, 2019, 39(12): 1210001.
王恩德, 齐凯, 李学鹏, 等. 基于神经网络的遥感图像语义分割方法[J]. 光学学报, 2019, 39(12): 1210001.

[13] Romera E, Álvarez J M, Bergasa L M, et al. ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(1): 263-272.

[14] Yu F, Seff A, Zhang Y, et al. Lsun: construction of a large-scale image dataset using deep learning with humans in the loop[EB/OL]. (2016-06-04)[2020-08-28]. https://arxiv.org/abs/1506.03365.

[15] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [EB/OL]. (2016-04-30)[2020-08-28]. https://arxiv.org/abs/1511.07122.

[16] Cai Y, Huang X G, Zhang Z A, et al. Real-time semantic segmentation algorithm based on feature fusion technology [J]. Laser & Optoelectronics Progress, 2020, 57(2): 021011.
蔡雨, 黄学功, 张志安, 等. 基于特征融合的实时语义分割算法[J]. 激光与光电子学进展, 2020, 57(2): 021011.

[17] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

[18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[19] Gao Z T, Wang L M, Wu G S. LIP: local

importance-based pooling[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3354-3363.

[20] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-03-02)[2020-08-28]. https://arxiv.org/abs/1502.03167.

[21] Chen L C, Barron J T, Papandreou G, et al. Semantic image segmentation with task-specific edge detection using CNNs and a discriminatively trained domain transform [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4545-4554.

[22] Liu T L, Gu Y Q, Cao D D, et al. A coarse-to-fine estimation method for spatial layout of indoor scenes [J]. Robot, 2019, 41(1): 58-64.
刘天亮, 顾雁秋, 曹旦旦, 等. 一种由粗至精的室内场景的空间布局估计方法[J]. 机器人, 2019, 41(1): 58-64.

[23] Zhang W D, Zhang W, Liu K, et al. Learning to predict high-quality edge maps for room layout estimation[J]. IEEE Transactions on Multimedia, 2017, 19(5): 935-943.

[24] Ren Y Z, Li S W, Chen C, et al. A coarse-to-fine indoor layout estimation (CFILE) method[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ACCV 2016 Lecture notes in computer science. Cham: Springer, 2017, 10115: 36-51.