激光与光电子学进展 先进成像

基于改进轻量网络的实时室内场景布局估计

岳有军[1]，张杰[1*]，赵辉[1,2]，王红君[1]

[1]天津理工大学电气电子工程学院天津市复杂系统控制理论与应用重点实验室，天津 300384；

[2]天津农学院工程技术学院，天津 300392

摘要 为简化布局估计网络结构，提高输出特征利用率，提出一种基于改进轻量网络的实时布局估计方法。利用轻量级的编码解码网络，端对端直接获得室内场景的主要平面分割图，实现实时的布局估计。针对以往联合学习方法特征利用率不高的问题，引入简化的联合学习模块，使用输出分割图的梯度作为输出边缘，将边缘的损失直接整合到整个网络输出损失中，提高特征利用率并精简联合学习网络。针对数据集正负标签不平衡和布局类型分布不平衡问题，使用分割型语义迁移，使用在LSUN数据集上训练得到的语义分割网络参数初始化所提网络参数，提高网络训练的稳定性。在两个基准数据集上对所提方法的性能进行评估。实验结果表明，在LSUN数据集上所提方法的平均像素误差为7.35%，在Hedau上为8.32%。通过消融实验证明了分层监督、简易学习联合和语义迁移机制对提高准确率的有效性。最终实验表明，所提方法能够实时获得准确的室内场景布局估计。

关键词 编码解码网络；室内场景；布局估计；端对端；语义分割

中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP202259.1810007

Real-Time Indoor Scene Layout Estimation Based on Improved Lightweight Network

Yue Youjun[1], Zhang Jie[1*], Zhao Hui[1,2], Wang Hongjun[1]

[1]*Tianjin Key Laboratory of Control Theory & Applications in Complicated Systems, School of Electrical and Electronic Engineering, Tianjin University of Technology, Tianjin* 300384, *China*;

[2]*College of Engineering and Technology, Tianjin Agricultural University, Tianjin* 300392, *China*

Abstract This study proposes a real-time layout estimation method based on an improved lightweight network to simplify the network structure of layout estimation and improve the use of output features. A lightweight coding and decoding network was used to obtain the main plane segmentation images of indoor scenes directly end-to-end and realize real-time layout estimation. Aiming at the problem of low feature usage in previous joint learning methods, a simplified joint learning module was introduced and the gradient of the output segmentation graph was used as the output edge. Additionally, the loss of the edge was directly integrated into the output loss of the entire network to improve feature utilization and simplify the joint learning network. Aiming at the imbalance of positive and negative labels of dataset and the imbalance of layout type distribution, to improve the stability of network training, segmentation semantic transfer was used to initialize the network parameters in this paper using the semantic segmentation network parameters trained on the LSUN dataset. The performance of the proposed method was evaluated using two benchmark datasets. The results show that the average pixel error of the proposed method is 7.35% and 8.32% on the LSUN and Hedau datasets, respectively. Ablation experiments prove the effectiveness of hierarchical supervision, simplified joint learning, and semantic transfer mechanism for improving accuracy. Finally, the experimental results show that the proposed method can estimate accurate indoor scene layout in real-time.

Key words encoding and decoding network; indoor scene; layout estimation; end-to-end; semantic segmentation

收稿日期：2021-07-08；**修回日期**：2021-07-21；**录用日期**：2021-07-28

基金项目：天津市科技计划(15ZXZNGX00290,19YFZCSN00360)

通信作者：*2580690058@qq.com

1 引 言

室内场景布局估计，旨在使用室内场景的图像，估计天花板、地板和左中右墙之间的边界，并将这五个主要平面区分开来，获得房间的整体框架。它在机器人导航[1]、目标检测[2]、深度预测[3]和增强现实[4]等领域发挥着重要的作用，是目前国内外研究的热点。

Hedau等[5]提出布局估计任务，并构建Hedau数据集，基于曼哈顿假设[6]，使用消失点和采样线获得大量候选布局，并使用结构化推理模型得到最优布局；Wang等[7]提出了一种判别学习方法，使用潜在变量来推断布局；Schwing等[8]应用积分图像的概念，优化布局推理框架。然而，上述提取图像底层特征的布局估计方法准确率较低，由于室内场景的复杂性，在布局估计任务中简单地提取线段、颜色和纹理等特征时极易受遮挡和误识别影响。因此，后来的研究者开始使用卷积神经网络提取更深层次的图像特征。

Mallya等[9]利用全卷积神经网络(FCN)[10]，联合学习边缘特征和语义特征，使用图像边缘特征来产生候选布局，并构建结构化学习模型来排序候选布局和择优；Dasgupta等[11]使用FCN获得5张主要平面语义标签图，然后使用标签图获得一张语义分割图作为粗略的分割型布局；Ren等[12]联合边缘和语义特征训练网络，得到粗略的边缘型布局，然后使用边缘直线度、表面平滑度和几何约束来生成更好的候选布局；Zhang等[13]使用VGG-16网络[14]为主干，通过构建编码解码网络来获取图像的边缘特征图，相较于先前FCN方法，提高了特征图的分辨率；Zhao等[15]使用语义迁移的方法，获得3种边缘线和背景的语义标签图。然而，上述方法使用的卷积神经网络在施加几何约束上的能力较弱，网络得到的粗略布局往往存在残影[11]和边缘线扭曲等不符合现实几何的情况，因此都附加后续优化过程以改进粗略布局。但优化方法也有不足：一是时效性较差，优化过程中往往会生成大量的布局候选项，影响效率；二是曼哈顿假设要求房间由严格的长方体盒子构成，由于室内设计变化多样，候选布局和非严格长方体房间的真实布局存在误差；三是精细化过程需要候选布局内的采样点或者采样线逐步移动和优化，计算量大。

Lee等[16]使用网络获得有序的关键点，相比于先提取特征再生成候选项最后精细化的方法，端对端方法更加省时；Lin等[17]基于ResNet101网络[18]，提出一种获取平面语义标签的方法，在网络里添加边缘直线度和表面光滑度约束，并提出布局结构退化方法来增加数据集；Hirzer等[19]使用3个网络，每个网络包含2个SegNet结构[20]，分别训练三类数据集，来缓解语义标签中墙歧义问题；黄荣泽等[21]使用编码解码网络，通过联合学习图像边缘和语义特征，直接生成布局分割图。然而large-scale scene understanding(LSUN)数据集[22]边缘标签和关键点标签正负项不平衡，一般的端对端的方法又缺少几何约束的后续处理，使得准确率进一步降低。再者，现有的端对端方法网络主干较深，结构较为复杂，常用的ResNet网络含多达101层，网络结构依然存在简化的空间。

为了提高布局估计准确率和时效性，简化网络结构，本文提出一种基于改进轻量网络的实时布局估计方法。相较以往的方法，所提方法使用更加轻量级的分割网络，实现端对端分割型布局估计，提高布局估计时效性；使用分层监督方式，层层推进，优化网络，提高布局估计精度；使用简化的联合学习，在联合语义和边缘特征训练网络的同时，提高对边缘特征的利用率，并简化网络结构；使用语义迁移的训练方法，缓解训练数据不平衡的问题，提高训练稳定性。

2 布局估计网络模型

2.1 轻量级分割网络

受语义分割网络[21,23]在现有布局估计工作应用的启发，本文选择在scene understanding RGB depth(SUNRGBD)数据集[24]上用于语义分割的efficient RGB-D semantic segmentation for indoor scene analysis(ESA)网络[25]，并针对布局估计任务进行优化。

ESA的实验表示，ResNet51语义分割准确率高，但是实时性较差，ResNet18反之。所以折中选择ResNet34作为编码器主要框架，用于高效且准确地提取图像高层特征。网络的输入图片不包含深度信息，所以删除了深度信息提取编码器和深度-RGB信息相加融合模块，设置中间上下文模块(CM)为两部分，分别提取全局和局部的上下文信息，并拼接融合。网络中存在的3个跳跃层(skip)用来弥补编码器下采样造成的损失。

采用5个主要平面的分割作为布局表示方式。ESA网络原本用于SUNRGBD数据集上37个目标的语义分割，所以在原网络的37分类层之后，添加输入通道数为37、输出通道数为5的分割型语义迁移(ST)层，实现37类目标到5个主要平面的迁移。在三个解码模块的分层输出后，也添加ST，实现分层输出的语义迁移，用以分层监督学习。除网络末端的最终输出外，三个解码模块也有不同尺寸大小的分割型布局输出。对于四个不同尺寸的输出，使用相应尺寸的标签来分层监督学习，层层递进，不断优化训练。修改后的网络结构如图1所示。

网络的语义分割损失包含4部分，所有分割损失为输出与标签的多分类交叉熵，其中最终输出的分割损失L_{out}为

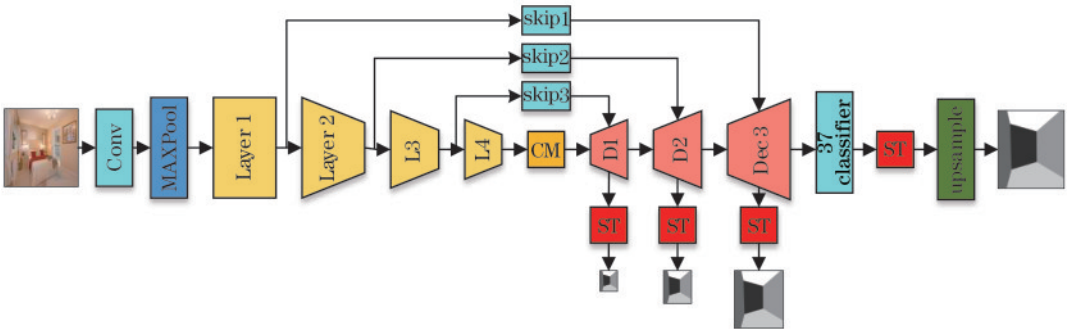

图 1　所提网络的结构

Fig. 1　Structure of the proposed network

$$L_{\text{out}}=-\log\left[\text{Softmax}\left(x_{\text{label}}\right)\right]=-\log\left[\frac{\exp\left(x_{\text{label}}\right)}{\sum_{j}\exp\left(x_{j}\right)}\right],\tag{1}$$

式中：x_{label} 表示真实标签对应的网络节点输出值；x_j 表示各个类别节点输出值，其他分割损失类似。整个网络的语义分割损失为

$$L_{\text{seg}}=a\cdot L_{\text{out}}+d_1\cdot L_{\text{dec1-32}}+d_2\cdot L_{\text{dec2-16}}+d_3\cdot L_{\text{dec3-8}},\tag{2}$$

式中：$L_{\text{dec1-32}}$ 为第一个解码模块的分割损失，其输出尺寸是原输入尺寸的 32 倍下采样；a 为最终输出损失的权重；d_1 为第一个解码模块输出损失的权重。

2.2　简化的联合学习

一般联合学习方法只取网络输出中的边缘特征或语义特征之一作为粗略布局，然后再精细化处理；或者只使用边缘特征作为粗略布局，语义特征作为后续评价的基准，输出特征的利用率较低。再者，对于多种特征的输出，往往需要分别构建语义特征输出通道和边缘特征输出通道，使得网络结构复杂。另外，多数联合学习方法都需要使用输出特征生成布局候选项，再经过精细化布局、排序、择优等后续处理，需要消耗大量的时间。

针对以上问题，本文引入边缘几何约束作为一种简化的联合学习方式，将边缘的损失直接整合到整个网络输出损失中，提高了边缘特征信息的利用率。由于语义分割和边缘布局实际上可以转换，边缘图实际为语义分割图的梯度，这为简化的联合学习提供了可行性依据。本文使用的边缘是前向传递得到的分割图梯度。所以网络后向优化时，直接优化分割图本身，避免了一般联合学习方法对某一部分特征利用率不足的问题。相比于具有双输出结构的联合学习网络，所提网络只采用单输出通道，能在使用边缘信息改善分割精度的同时，精简网络结构。并且所提精细化布局的方法直接包含在网络训练中，免去了后续处理，改善了时效性。引入的边缘直线度和表面光滑约束也使得预测结果的边缘看上去更平滑和更直，更加具有几何合理性。

表面光滑约束 L_{smooth} 表现为网络损失中真实标签和输出布局之间的均方差损失，它在增强每个平面内一致性的同时，对边缘施加平滑约束，具体公式为

$$L_{\text{smooth}}=\frac{1}{N}\sum\left(M-M_{\text{gt}}\right)^2,\tag{3}$$

式中：N 为样本个数；M 为预测分割图；M_{gt} 为真实的分割图。边缘直线度约束 L_{edge} 表现为输出布局的梯度边缘和真实边缘的二分类交叉熵损失，可使得边缘更直，更具有几何合理性，具体公式为

$$L_{\text{edge}}=-\left[y_i\cdot\log\left(p_i\right)+\left(1-y_i\right)\cdot\log\left(1-p_i\right)\right],\tag{4}$$

式中：y_i 代表样本 i 真实边缘的值；p_i 代表网络输出布局梯度的值。联合学习的损失函数为

$$L_{\text{joint}}=\lambda_{\text{s}}\cdot L_{\text{smooth}}+\lambda_{\text{e}}\cdot L_{\text{edge}},\tag{5}$$

式中：λ_{s} 和 λ_{e} 均为权重。联合式(2)和式(5)，最终得到整个网络的损失函数 L，具体公式为

$$L=L_{\text{seg}}+L_{\text{joint}}。\tag{6}$$

2.3　分割型语义迁移训练

由于 LSUN 数据集中不同布局类型的图片数量分布不平衡，有的类型图片不足 10 张。边缘标签和关键点标签正负项不平衡，在边缘标签上，负项标签背景占主要部分。因此，不同于以往的迁移学习和边缘型语义迁移，本文使用分割型语义迁移的训练方法，将 37 种目标分类迁移到天花板、地板和左中右墙 5 个主要平面上。因为整个图片的像素都可以被分类为若干个主要平面，所以避免了数据集正负项不均衡问题。

相对应地，使用在 SUNRGBD 数据集上预训练得到的 ESA 网络的部分参数，作为所提网络的初始化权重，在 LSUN 数据集上训练所提布局估计网络。SUNRGBD 数据集比 LSUN 大很多，常用于室内场景语义分割，和布局估计任务类似，所以能很好地解决不同布局类型的图片数量分布不平衡问题。

在网络训练时，将 ESA 网络的 RGB 编码器、跳跃层、上下文模块和解码器等模块的参数用于所提网络的初始化；深度编码器、深度-RGB 图像信息融合层和 37 类上采样层参数由于不适合本文的 5 类平面分割任务被删除；重新训练所提网络特有的 4 个语义迁移和 5 类上采样层。不同参数的应用如图 2 所示，蓝色表示

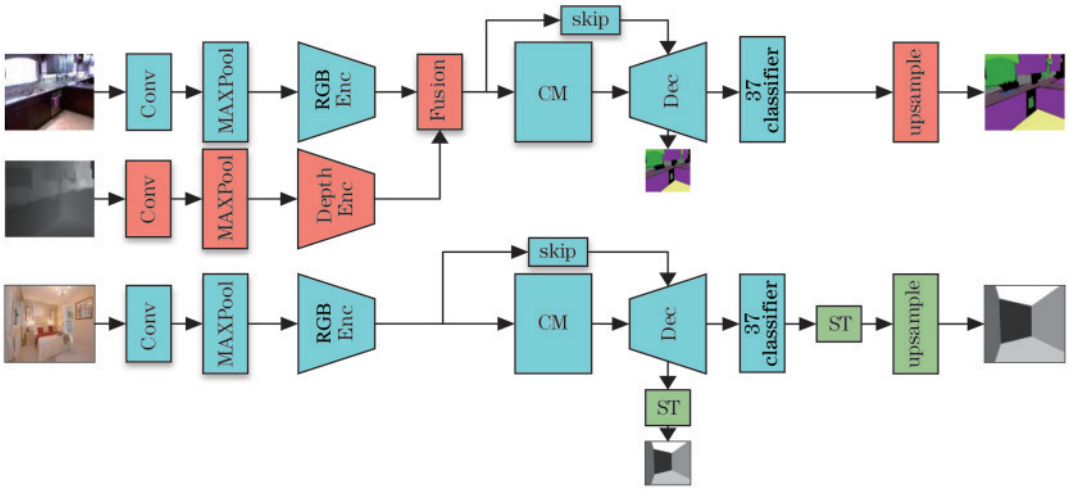

图 2　ESA 网络中不同参数的应用

Fig. 2　Applications of different parameters in ESA network

用于所提网络初始化的参数，橙色表示被删除的参数，绿色表示需要重新训练的参数。上部分为 ESA 文献中修改后的网络结构简化图，下部分为布局估计网络简化图。

3　实验与分析

在 Pycharm 平台上使用 Pytorch 搭建网络。利用布局结构退化方法来扩充数据集，在 LSUN 数据集上迭代 150 次，学习率为 0.0001。在 LSUN 和 Hedau 数据集上测试训练得到的网络，并与其他方法进行比较。通过消融实验，验证多层监督、简化联合学习和语义迁移训练方法的作用。最后通过大量实验，估计出用于训练的 4 个分割误差系数的最优值。由于所提网络生成分割型布局，所以使用像素误差（PE）作为定量评价标准，PE 计算代码由 LSUN 官方提供。

3.1　LUSN 数据集上实验

在 LSUN 验证集上测试所提方法，并和以往的方法进行比较。定量比较 PE，定性比较是否有端对端时效性和网络主干精简性，结果如表 1 所示。

表 1　LSUN 数据集上不同方法的性能评估

Table 1　Performance evaluation of different methods on the LSUN dataset

Method	PE /%	End-to-end	Network backbone
Method in Ref. [5]	24.23	No	No
Method in Ref. [9]	16.71	No	FCN(VGG-16)
Method in Ref. [11]	10.63	No	FCN(VGG-16)
Method in Ref. [12]	9.31	No	FCN(VGG-16)
Method in Ref. [16]	9.86	Yes	Encode and decode
Method in Ref. [13]	12.49	No	Encode and decode
Method in Ref. [19]	7.79	Yes	SegNet×6
Method in Ref. [21]	9.05	Yes	Encode and decode
Proposed method	7.35	Yes	ESA(ResNet-34)

与其他方法相比，所提方法得到最小的像素误差。所提方法使用精简的 ResNet-34 网络作为编码器主干，与文献[12]和文献[21]以往的联合学习网络相比，使用了简易的联合学习方式，不需要构建两个解码器，在提高布局估计精度的同时，简化了网络结构，并提高特征利用率；与文献[19]使用 3 个网络，分别训练只含一、二、三面墙数据集的方法相比，所提训练方法和网络结构较简单，并且像素误差更小；与文献[11]、文献[12]和文献[13]需要网络得到粗糙布局再进行后续迭代优化的多步走方法相比，所提方法使用端对端的布局估计方法，使用训练好的网络，输入室内场景原图直接得到布局估计，每张图片的处理时间约为 0.072 s，可以完成实时的布局估计。

图 3 展示了在 LSUN 验证集上视觉效果较好的 4 个结果（场景 1 到场景 4）和较差的 2 个结果（场景 5 和场景 6）。对于场景 1 和场景 2，所提方法很好地表现了在床、柜子和沙发等大型家具遮挡下的鲁棒性；在场景 3 光照条件差和场景 4 干扰直线较多情况下，所提方法也能有很好的准确率；场景 5 中由于房间右顶部和椅子右侧有明显的直线干扰，所提方法输出左中右墙，但实际只有两面墙，可能与网络的过拟合有关；场景 6 中真实标签不含中墙，但在仔细观察原图后，发现原图左侧确实存在被柜子遮挡的一个主要平面，实际原图有左中右三面墙，所提方法在遮挡严重下依旧有较好的识别效果。

3.2　Hedau 数据集上实验

在不用 Hedau 数据集训练的前提下，直接在 Hedau 数据集上测试在 LSUN 数据集上训练得到的网络，并与其他方法进行比较，得到的结果如表 2 所示。可以看出，所提方法在 Hedau 数据集上的测试结果比在 LSUN 数据集上的结果略差，这可能与 Hedau 数据集的标注方法比 LSUN 数据集更加严格有关，而有的方法使用了 Hedau 数据集进行训练。即便如此，所提

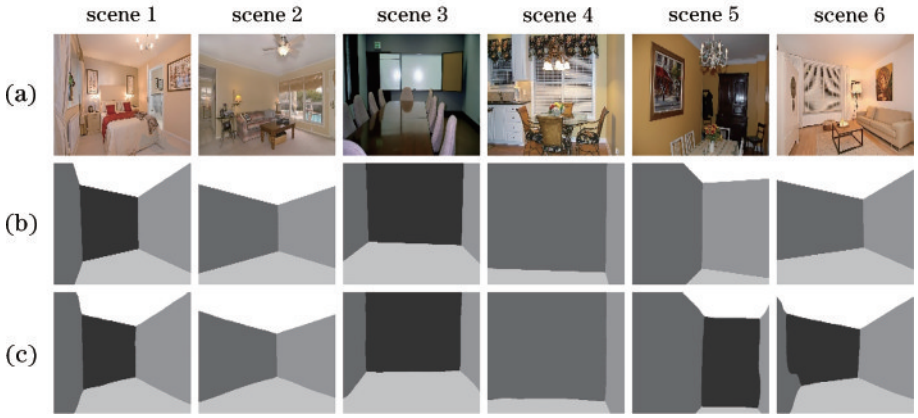

图 3　LSUN 数据集上的实验结果。(a)原图;(b)真实布局;(c)所提方法得到的布局

Fig. 3　Experimental results on LSUN dataset. (a) Original picture; (b) real layout; (c) layout obtained by proposed method

表 2　Hedau 数据集上不同方法的性能评估

Table 2　Performance evaluation of the different methods on the Hedau dataset

Method	PE /%
Method in Ref. [5]	21.20
Method in Ref. [7]	20.10
Method in Ref. [8]	12.80
Method in Ref. [9]	12.83
Method in Ref. [11]	9.73
Method in Ref. [12]	8.67
Method in Ref. [16]	8.36
Method in Ref. [13]	12.70
Method in Ref. [19]	7.44
Proposed method	8.32

网络也能在 Hedau 数据集上表现出较好的泛化能力。

图 4 展示了在 Hedau 数据集上视觉效果较好的 4 个结果(场景 7 到场景 10)和较差的 2 个结果(场景 11 和场景 12)。在场景 7 和场景 8 大型沙发遮挡、场景 9 物件众多且灯光较暗的条件下,所提方法依然有良好的效果,在 Hedau 数据集上有较好的泛化能力;在场景 11 中所提方法的效果较差,这与原图不严格遵循曼哈顿假设有关,原图中显然存在多个既不平行也不垂直的墙面。从场景 12 真实布局可以看出,Hedau 数据集的标注方法比 LSUN 数据集更为严格,这也是所提方法在 Hedau 数据集上的 PE 略高的原因。

3.3　消融实验

在 LSUN 数据集上进行消融实验,以验证本文的分层监督、简易联合和分割型语义迁移在提高布局估计精确上的有效性。添加三次实验,控制其他变量一定,分别消融这三种机制,重新进行三次网络训练,使用 PE 作为评价标准,并与同时使用三种机制训练网络的结果进行比较。实验结果如表 3 所示,4 种方法得到的输出布局如图 5 所示。

从图 5 可以看出:消融分层监督对输出准确率影响很大,场景 13 中出现明显的残影现象;消融简易联合的输出中,线条几乎都存在扭曲;消融分割型语义迁移产生的误差最大,在场景 14 中甚至未能识别出两面墙;而三种机制的结合不仅能准确地识别出墙的数量,还能减少残影的产生,并且能够合理地优化边缘线,使得各个主要平面之间的边界更直,更加符合几何实际,布局估计准确率更高。

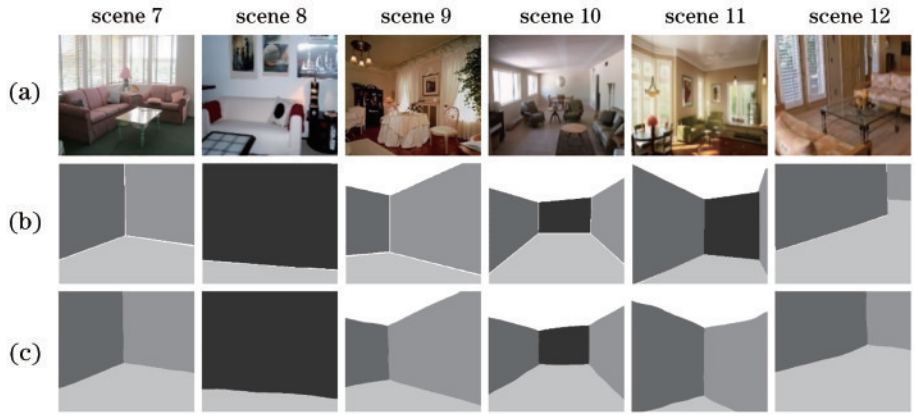

图 4　Hedau 数据集上的实验结果。(a)原图;(b)真实布局;(c)所提方法得到的布局

Fig. 4　Experimental results on Hedau dataset. (a) Original picture; (b) real layout; (c) layout obtained by proposed method

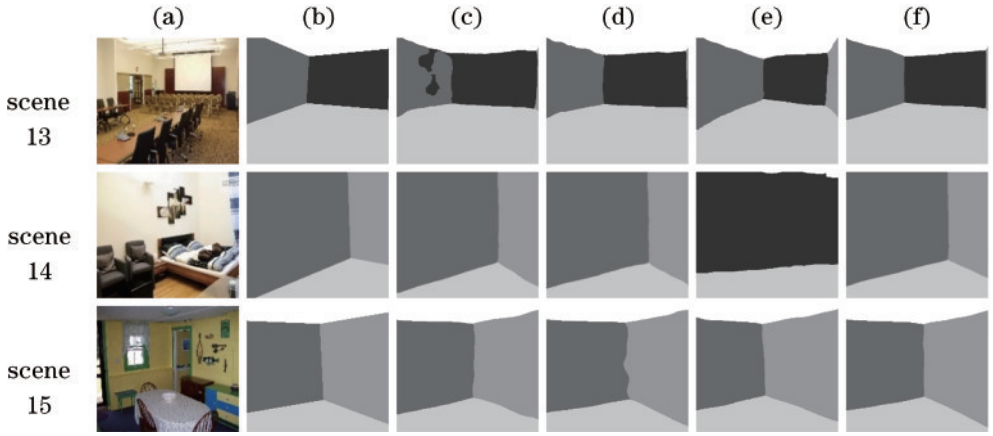

图 5　消融实验结果。(a)原图;(b)真实布局;(c)消融分层监督;(d)消融简易联合;(e)消融语义迁移;(f)结合三种机制

Fig. 5　Results of ablation experiment. (a) Original image; (b) real layout; (c) ablating layered monitoring (ALM); (d) ablating simple combination (ASC); (e) ablating semantic transfer (AST); (f) combination of three mechanisms (CTM)

与ESA网络不同,在式(2)中所提网络最终输出的损失和3个解码器输出的损失在损失函数中最优权重并不相同,这可能与ESA的任务不同有关。再者ESA不存在语义迁移层,而所提方法在3个解码器模块输出和最终输出上都添加语义迁移层,最终输出和3个解码模块输出的差异较大。本文通过大量实验,估计出最优训练权重,实验结果如表4所示。当a为1,d_1、d_2、d_3为0.2时,达到最小像素误差。表1和表2中所用的和其他文献对比的所提网络,就是表3中同时使用三种机制和表4中使用最优参数的网络,所提网络在LSUN数据集上的像素误差都是7.35%。

表3　消融实验结果

Table 3　Results of the ablation study

Method	PE /%
ALM	7.84
ASC	7.67
AST	14.30
CTM(ours)	7.35

表4　不同损失权重的结果

Table 4　Results of different loss weights

a	d_1	d_2	d_3	PE /%
0.25	0.25	0.25	0.25	8.81
0.5	0.5	0.5	0.5	7.99
0.7	0.7	0.7	0.7	8.07
1	1	1	1	7.72
1.25	1.25	1.25	1.25	8.07
1	0.5	0.5	0.5	7.67
1	0.2	0.2	0.2	7.35
1	0	0	0	7.84

4　结　　论

提出一种改进的轻量级网络,实现端对端的实时室内场景布局估计。使用多层监督优化损失函数,提高准确率;针对卷积神经网络在几何约束上的欠缺,使用简化的联合学习方法,在网络中精细化布局,简化网络结构并提高时效性;使用分割型语义迁移的训练方式,缓解训练数据集不平衡问题。最后在LSUN和Hedau数据集上,比较了所提方法和几种基准方法的性能。通过大量实验,证明了多层次监督、简易联合学习和分割型语义迁移对提高布局估计准确率的有效性,并估计出最优参数权重。最终表明,所提网络能实时且准确地完成布局估计任务,很好地解决遮挡和误识别问题。但此方法的泛化能力还有提升的空间,往后的工作将会考虑使用其他数据增强方法或者扩大数据集来训练网络。

参　考　文　献

[1] Zhu F D, Zhu L C, Yang Y. Sim-real joint reinforcement transfer for 3D indoor navigation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 11380-11389.

[2] 冀中, 孔乾坤, 王建. 一种双注意力模型引导的目标检测算法[J]. 激光与光电子学进展, 2020, 57(6): 061008.
Ji Z, Kong Q K, Wang J. Object detection algorithm guided by dual attention models[J]. Laser & Optoelectronics Progress, 2020, 57(6): 061008.

[3] Wong A, Cicek S, Soatto S. Learning topology from synthetic data for unsupervised depth completion[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 1495-1502.

[4] 李雪婷, 党建武, 王阳萍, 等. 基于文字特征的增强现实识别注册方法[J]. 激光与光电子学进展, 2020, 57(2): 021502.
Li X T, Dang J W, Wang Y P, et al. Augmented reality recognition registration method based on text features[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021502.

[5] 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. New York: IEEE Press,

2009: 1849-1856.

[6] Coughlan J M, Yuille A L. The manhattan world assumption: regularities in scene statistics which enable Bayesian inference[C]//Annual Neural Information Processing Systems Conference, November 27-December 2, 2000, Cambridge. New York: Curran Associates, 2000: 845-851.

[7] 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, 6312: 435-449.

[8] Schwing A G, Hazan T, Pollefeys M, et al. Efficient structured prediction for 3D indoor scene understanding [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI. New York: IEEE Press, 2012: 2815-2822.

[9] 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.

[10] 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: 431-440.

[11] 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.

[12] Ren Y Z, Chen C, Li S W, et al. A coarse-to-fine indoor layout estimation (CFILE) method[EB/OL]. (2016-07-03)[2021-05-08]. https://arxiv.org/abs/1607.00598.

[13] 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.

[14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2021-05-06]. https://arxiv.org/abs/1409.1556.

[15] Zhao H, Lu M, Yao A B, et al. Physics inspired optimization on semantic transfer features: an alternative method for room layout estimation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 870-878.

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

[17] Lin H J, Huang S W, Lai S H, et al. Indoor scene layout estimation from a single image[C]//2018 24th International Conference on Pattern Recognition (ICPR), August 20-24, 2018, Beijing, China. New York: IEEE Press, 2018: 842-847.

[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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[19] Hirzer M, Roth P M, Lepetit V. Smart hypothesis generation for efficient and robust room layout estimation [C]//2020 IEEE Winter Conference on Applications of Computer Vision, March 1-5, 2020, Snowmass, CO, USA. New York: IEEE Press, 2020: 2901-2909.

[20] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[21] 黄荣泽, 孟庆浩, 刘胤伯. 基于多任务监督学习的实时室内布局估计方法[J]. 激光与光电子学进展, 2021, 58(14): 1410023.
Huang R Z, Meng Q H, Liu Y B. Real-time indoor layout estimation method based on multi-task supervised learning[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410023.

[22] Yu F, Seff A, Zhang Y D, et al. LSUN: construction of a large-scale image dataset using deep learning with humans in the loop[EB/OL]. (2015-06-10)[2021-05-08]. https://arxiv.org/abs/1506.03365.

[23] 王旭初, 刘辉煌, 牛彦敏. 基于双流加权Gabor卷积网络融合的室内RGB-D图像语义分割[J]. 光学学报, 2020, 40(19): 1910001.
Wang X C, Liu H H, Niu Y M. Indoor RGB-D image semantic segmentation based on dual-stream weighted Gabor convolutional network fusion[J]. Acta Optica Sinica, 2020, 40(19): 1910001.

[24] Song S R, Lichtenberg S P, Xiao J X. SUN RGB-D: a RGB-D scene understanding benchmark suite[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 567-576.

[25] Seichter D, Kohler M, Lewandowski B, et al. Efficient RGB-D semantic segmentation for indoor scene analysis [C]//2021 IEEE International Conference on Robotics and Automation (ICRA), May 30-June 5, 2021, Xi'an, China. New York: IEEE Press, 2021.