

三 维 动 画 和 交 互 设 计 课 程 读 书 报 告



作者姓名 殷曜东

作者学号 21851020

年级专业 2018级软件工程

所在学院 软件学院

完成日期 2018.12.24

Research on Computer Animation and Interaction Design

A Report Submitted to

Zhejiang University

Major Subject: Software Engineering

Advisor: Li Qilei

By

Yin Yaodong

College of Software Technology, Zhejiang University

2018

**摘要**

室内设计和平面制图以往需要耗费大量的人力和时间，随着深度学习理论的发展，卷积神经网络在图像识别处理等领域的卓越表现获得了更多的关注，将该理论用于室内设计或者平面制图等方面具有充足的现实意义。

本文基于深度卷积神经网络和生成模型理论，将试图搭建一个能够生成高质量室内平面图的模型，在经过训练之后，模型将能够生成这种图片，而且和人手工绘制的图片难以区分。这种技术如果能够发展成熟，将有效地节省在该方面所投入的各种人力物力，人类社会走向人工智能时代的步伐也会更进一步。

**关键词**：室内场景；卷积神经网络；生成模型；

**Abstract**

Interior design and graphic design used to require a lot of manpower and time, with the development of deep learning theory, convolutional neural networks have gained more attention in the field of image recognition and processing.The application of this theory to interior design or plane drawing has sufficient practical significance.   
 In this paper, based on deep convolutional neural network and generating model theory, we will try to build a model that can generate high quality interior plans. After training, the model will be able to generate such pictures, and it is indistinguishable from the pictures drawn by people.If this kind of technology can be developed and matured, it will effectively save all kinds of manpower and material resources invested in this respect, and the pace of human society moving towards the era of artificial intelligence will also be further.

**Keywords:** indoor scene; convolutional neural network; generating model;

1. **论文简介**

论文题目：Deep Convolutional Priors for Indoor Scene Synthesis，用于室内场景合成的深度卷积先验模型

作者：KAI WANG, Brown University

MANOLIS SAVVA, Princeton University

ANGEL X. CHANG, Princeton University

DANIEL RITCHIE, Brown University

发表于ACM Transactions on Graphics,August 2018.

**二、研究背景与问题描述**

2.1研究背景

人类的生活活动很大一部分在室内度过，如办公室，客厅，卧室和其他室内空间。 许多行业使用这种室内场景的虚拟复制品，如室内设计，建筑，游戏和虚拟现实等等。能够理解这些场景结构的计算机模型可以产生新的场景，通过实现全自动或半自动的形式生成虚拟室内环境来服务这些行业。

2.2问题描述

卷积神经网络（CNN）对于解决场景合成问题具有一定的潜力，在其他图形域中，它们已经被证明能够可靠地学习识别并生成视觉模式和关系。由于深度网络需要大量的训练数据，所以没有将其用于仅拥有小型场景数据集的早期场景合成工作。然而，近年来3D场景的大型数据集已经可用。在本文中，作者们使用这样的数据来训练基于CNN的模型，在仅给定所需房间类型和其室内结构作为输入的情况下，训练模型选择和放置房间中的对象。

为了将卷积网络应用于场景合成问题，我们使用了这样一套理论：当室内房间是三维空间时，房间的大多数特征对象都排列在其二维地面平面上。因此，我们通过正交的自上而下的视图来表示房间，这样就可以使用卷积网络进行操作。这里定义的自上而下的视图表示是一种语义丰富的、多通道的图像，它捕获每个像素的多个场景特征，例如对象的类型和方向。将多层可学习卷积应用于该表示，使我们的模型能够处理房间的整个状态，作为其采样决策、捕获模式和对象之间关系的基础。我们提出的模型通过一串连续地添加一个对象来生成一个房间图，考虑步骤如下：首先决定是否添加一个对象，然后再决定放置哪种类型的对象以及放置在何处。这些决策都是由一个卷积神经网络来控制的。如图2-1所示。模型通过连续迭代地插入一个对象来生成场景。这里是一个客厅场景的对象插入序列。图像顺序分别为中间场景的渲染和模型的预测空间概率分布（用蓝色表示）。

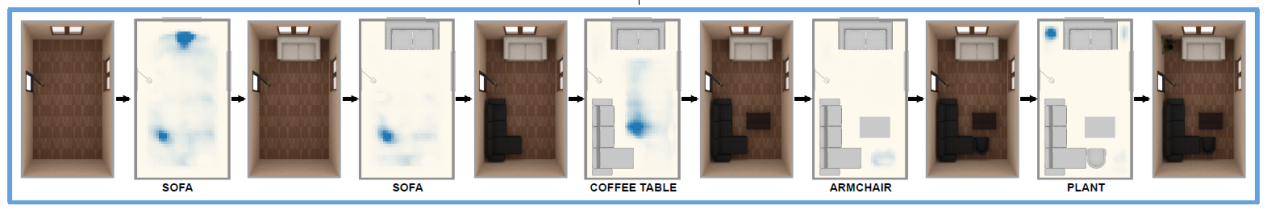


图2-1 客厅场景的对象插入序列

先前有其他研究者已经研究了室内场景合成的这个问题：考虑到房间的建筑几何形状（即墙壁，地板和天花板），选择一组在房间内放置和排列的物体。已经提出了各种方法，包括约束 满意度，手工制作的室内设计原则的优化，对象之间成对关系的统计先验，以及以人为中心的关系先验等。然而，在室内规模的无约束合成仍然是一个具有挑战性的问题。先前的工作要么专注于在较大的房间（即工作台）内建模较小的功能区域，要么引入额外的输入来约束问题：固定的对象和手动注释重要的关系，草图，自然语言描述或房间的3D扫描等。

2.3 论文贡献

1）提出了第一个基于卷积网络的室内场景合成系统：该系统只需要接受待合成场景的房间的几何图形作为输入。

2）引入了一种全新的、基于正交自上而下视图的、语义丰富的的场景表示方法。

3）通过一系列消元，分析了所学到的卷积先验，并展示了它们根据房间的状态捕捉对象排列的常识模式。

4）进行强制选择感知研究，以比较使用该方法生成的房间的合理性，与人类设计师或其他角色创作的场景相比较。我们的方法生成的场景表现高于基准线，可以说足够好，并且在某些情况下与人工创建的场景表现同样好。

1. **模型架构实现**

3.1设计思路

本文主要研究室内场景的生成模型，特别是包含一个或多个对象的房间的生成模型。我们方法的输入是一个房间，即描述地板、墙和天花板的几何图形。在这项工作中，我们的重点是生成和安排主要的功能对象，如放置在地板上的家具。我们不处理挂在墙上或第二层的对象（比如桌子上的物品）。

我们的目标是建立一个场景的生成模型，该模型可以编码涉及对象和房间几何图形的关系。我们使用深度网络（DNN）作为我们模型的主要构建方式，DNN已经被证明能够可靠地识别和生成其他视觉领域中的复杂视觉模式。为了做到这一点，我们利用了这样一个事实，即当室内场景存在于三维空间中时，重力决定大多数物体排列在二维地板平面上，因此，我们可以将2D卷积网络应用于场景S的自上而下的视图表示V(S)，其中网络学习识别对象的存在和对象之间的关系。

DNN中的第一个预测为模型是否应该继续向场景添加对象。如果此组件返回true，则下一个模型组件将选择要添加的对象类别以及添加它的位置。该组件的工作原理是在许多可能的场景位置上构建可能的对象类别的概率分布，然后从该分布中进行采样。给定对象类别和位置，模型必须在该位置实例化属于该类别的特定对象，在该过程中适当地定向对象。这三个模型组件都使用深度卷积网络，将场景视图作为输入。这些网络是在从3D场景的大型数据库中采样的场景上训练的。

综上所述，模型的实现包括三个主要部分，分别为判断什么时候停止添加对象，判断应当在哪添加什么类型的对象，添加对象。

3.2模型设计

模型生成场景S的方式为迭代地向场景中添加对象，对象的存在与否和位置都取决于场景当前的状态，每一次迭代，模型的动作分为三步：决定是否添加另一个对象（Continue？），决定添加对象的类型和位置（CategoryLocation），插入对象实体（InstanceOrientation）。

3.2.1 是否添加另一个对象

决定是否添加另一个对象的是一个函数，它的输入为当前场景S，输出为一个伯努利分布Pcontinue(T|S)，计算当前场景分为两步，1）计算当前已存在与场景的对象数量，将结果表示为counts（S），作为非常重要的global information，2）用一个深度卷积网络从俯视图中抽取高级别的特征，这些特征提供一个信息：当前的场景是否能够表达一个完整的房间。

用图3-1举例来说，左上角的图，这个房间里没有床所以Pcontinue 接近1，说明这种图很可能应该继续添加对象；右上角虽然有床，但是还是比较空，所以Pcontinue为0.64，模型认为它还是有较大的可能应继续添加对象；右下角的空间被利用得相当充分，所以模型认为应该继续添加对象的可能性很低，只有0.03.

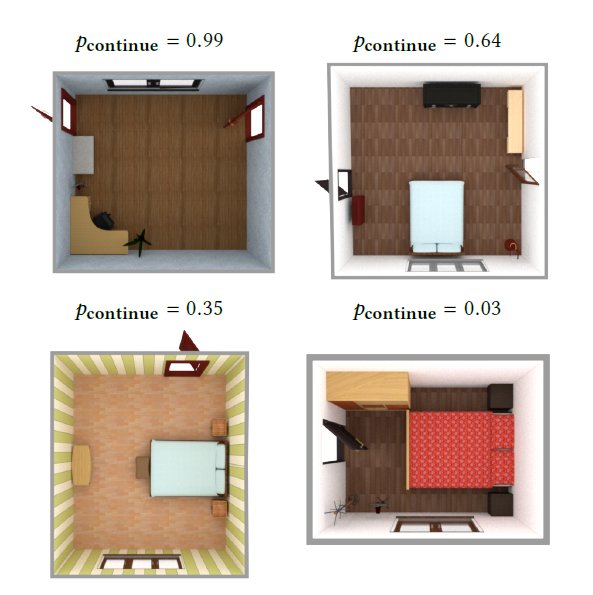


图3-1

这个Continue?模块使用标准的交叉熵损失函数进行训练，在实际进行合成时，当Pcontinue的值小于0.5时，就应该停止合成。

3.2.2 决定添加对象的类型和位置

如果Continue？决定添加另一个对象时，模型必须同时决定要添加哪一类对象以及在房间中的什么位置添加该对象。这两个决定是密切相关的：有些位置只对某些类型的对象有意义，反之亦然。理想情况下，我们希望模型了解场景中所有类别和所有可能位置的联合分布P(c，x，y|S)（c表示种类，xy表示位置）。为了使这个问题适用于卷积网络的使用，我们学习了一个条件分布，具体地说，我们的下一个模型组件类别位置计算PCAT(c|S，x，y)，即c是在位置(x，y)处添加到场景S中的类别的概率。

要计算这些概率，CategoryLocation使用与Continue？类似的结构：卷积网络从房间的当前状态中提取空间特征，以及房间的当前不同类型的对象计数。由于CategoryLocation计算特定位置(x，y)的概率，因此它必须包含其他信息，以指示CNN关注该位置。为此，我们在自上而下的视图图像V(S)中添加了一个额外的attention mask通道。这是一个ImageMattn(x，y)包含一个以(x，y)为中心的小(9×9)mask。图3-2显示了一个mask运用的例子。

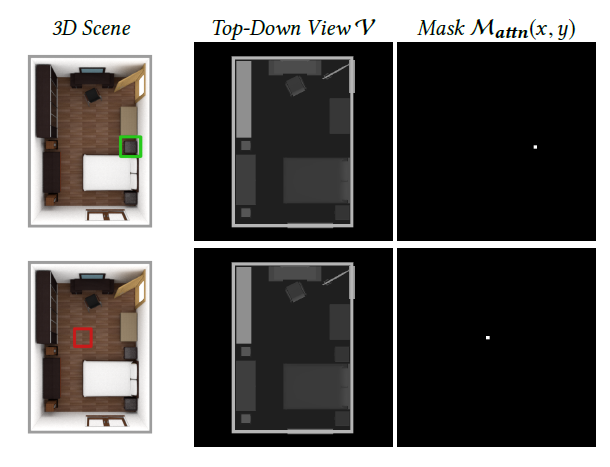


图3-2

Pcat(c |S, x,y)给出了一个概率分布，表示（x，y）位置的类型概率。既然它是一个概率分布意味着它的和为1，表示每个位置上都会有某个对象，这显然是不合理的，一个房间很大一部分区域都是空闲的。所以在对象分类中加入一个辅助类型表示“没有任何对象质心”同时加入另外两个辅助类型分别表示“某个位置已被占用“和“房间外的区域”，如图3-3所示。

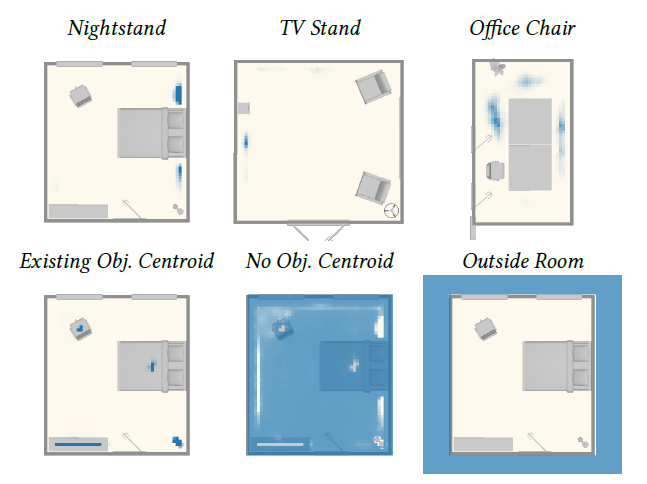


图3-3

1）训练（Training）

为了生成CategoryLocation的训练数据，我们从数据集中提取房间，并从其中随机删除对象。为了生成对象类别的示例，选择一个随机删除的对象，并生成关于该对象的质心(x，y)的attention MaskMattn(x，y)。为了生成辅助类别的示例，我们选择房间外部、房间内的空白空间或现有对象中的一个随机点，并围绕该点生成attention mask。。我们生成的训练集包含95%的辅助类别示例和5%的对象类别示例。考虑到场景中对象质心位置的稀疏性，这种平衡具有直观的意义。对象类别实例的比例较低会导致网络训练速度慢，为了加快训练速度，我们只对对象类别实例进行训练，并逐步将辅助类别实例的比例提高到95%。

给定大量可能的条件输入（S，x，y），模型可能看不到足够的数据来学习稀有规则，例如： 不要在已经有两张床的卧室里放第三张床。 为了帮助解决这种情况，我们使用标准的分类交叉熵损失以及额外的损失Lglobal来训练CategoryLocation。 这种损失通过惩罚模型来提供额外的全局信息，以便将概率分配给不在Cremoved中的类别，即从当前场景中移除的对象类型。

2）消融（Ablation）

模型训练了CategoryLocation的不同版本，这些版本从场景视图图像V（S）中省略了不同特征。除了完整的图像特征集（All）之外，我们还考虑了删除类别掩码通道（NoCategories），以及除去深度以外的所有通道（DepthOnly）的场景。我们还在用于训练CategoryLocation网络的数据集中试验了辅助类别示例（NoObject，ExistingObject，OutsideRoom）的百分比。图3-3显示了每种设置的典型预测分布。在第一行，展示了如何需要所有功能来防止床放置在阻挡门的位置。在最下面的一行中，展示了如何增加训练数据中辅助类别示例的百分比，从衣柜位置的分布中去除虚假概率质量，从而导致可靠地将其质量集中在可用墙壁上的分布。

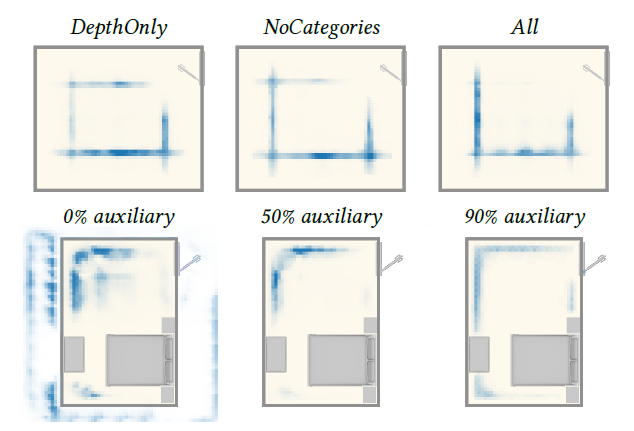


图3-3

3）采样（Sampling）

为了在合成时采样新的类别和位置，需要使用Pcat（c | S，x，y）的值构造联合分布P（c，x，y | S）。为了使这个问题易于处理，我们考虑在场景上定义的N×N规则网格上的候选位置。然后，我们需要在网格位置上定义先验分布p（x，y | S），此外我们已经考虑了物体在特定位置存在或不存在的可能性。

3.2.3插入对象实例

给定位置（x，y）和对象类别c，模型的最后一步是在该位置插入该类别的实例。这里，'实例'指的是特定的3D模型。本文中，从SUNCG数据集中使用的模型集中绘制这些3D模型。

对于给定类别可能存在许多可能的实例（例如，SUNCG中有近300个椅子模型），并非所有实例都在视觉上彼此兼容。因此，我们首先将所考虑的实例集限制为适合在房间中存在的实例集。SUNCG源于室内设计工具，包含许多从风格一致的家具系列中汲取的3D模型。我们用他们的系列注释了所有SUNCG 3D模型。在合成时，我们允许插入来自已经在房间中的一个对象，它的那个实例集中的另一个对象。我们稍微放宽了这个约束。

为了将实例插入房间，我们的模型还必须为该实例选择方向，其中方向被定义为关于重力矢量的角度θ。为此，我们使用第三个也是最后一个卷积网络组件InstanceOrientation.这个组件将候选实例i插入到候选方向θ的场景中，然后评估返回pinst的网络（T| S，i，x，y ，θ），在（x，y，θ）处插入实例i的概率.InstanceOrientation用插入实例的几何图形的maskMgeo（i，x，y，θ）增强俯视图V（S）。计算插入的概率所得为Pinst(⊤|S, i, x,y, θ)。如图3-4所示，当床头方向靠墙时和不靠墙时，Pinst的值分别接近1和0.

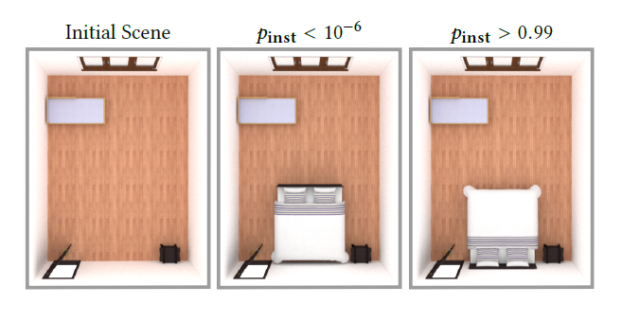


图3-4

通过从数据集中取出房间来训练InstanceOrientation，再次从每个房间中删除一组随机对象，然后选择一个要重新插入的已删除对象。我们通过将对象插入其原始方向生成50％的正面训练示例，并通过将对象插入不同方向来生成50％的负面训练示例。对于这些负示例方向，将[0,2π]量化为16个离散方向，对象的原始方向为0，我们从其他15个方向中进行选择。训练使用标准二元交叉熵损失。

在合成时，模型尝试在16个离散方向中的每一个插入所有允许的实例，并选择具有最高概率的插入方向。如果没有无冲突插入或没有插入概率高于50％，模型会从CategoryLocation重新采样不同的（x，y，c）元组。模型会自动拒绝任何使用超过20个此类重采样步骤的合成房间。实验中未被拒绝的房间平均使用少于三个重采样步骤。

1. **实验结果与分析**

本文中所有模型都是使用PyTorch进行实现和训练的，实验所用GPU为NVIDIA GeForce GTX 1080 Ti ，最终在验证集上的分类准确率为：Continue? 85%，InstanceOrientation 94%。

为了评估模型生成具有类似于SUNCG培训场景的特征的合理房间的能力。，通过对Amazon Mechanical Turk的感知研究，将模型合成的场景与几条基线进行比较。所有这些比较都使用了相同的研究设计，扩展了以往对自动图像着色方法的评价工作。研究参与者要求进行一些必须进行选择的图像比较任务。每个任务都显示了一对图像，一幅来自模型合成，另一幅来自基准源。两个场景使用相同的房间几何图形。图像在每对内的顺序是随机的。每个场景图像中的对象根据其类别进行着色，以便参与者根据类别进行判断。这项任务要求参与者从这对图片中选择一张他们认为更好的。每个参与者完成55个比较任务。每11个任务中就有一个是“警惕性测试”：与明显错误的答案进行比较(特别当一幅图像描绘了随机、混乱的图像)。对于每种房间类型的每一项研究，收集10名参与者的回复，并丢弃了在警戒测试中未达到100%准确的员工的所有回复。结果如图4-1所示，图4-2为部分图像对比。

在所有房间类型中，此方法生成的场景明显优于那些，通过先采样后排列而生成的场景(Occurrence Baseline)。在排列固定的对象时，我们的模型也优于基于成对对象关系统计(Arrangement Baseline)的模型。与测试数据集中的人造场景相比，我们生成的办公场景同样好，卧室和客厅场景则稍弱一点。

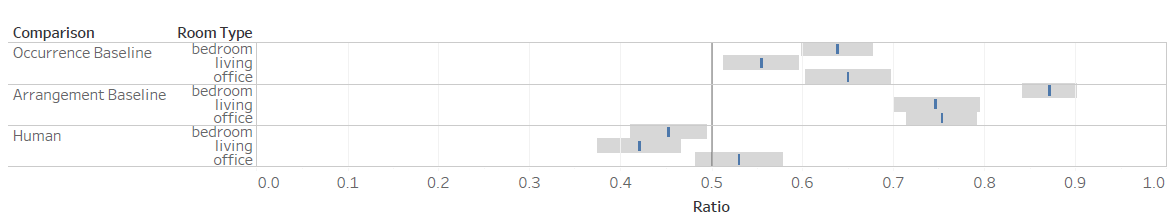


图4-1

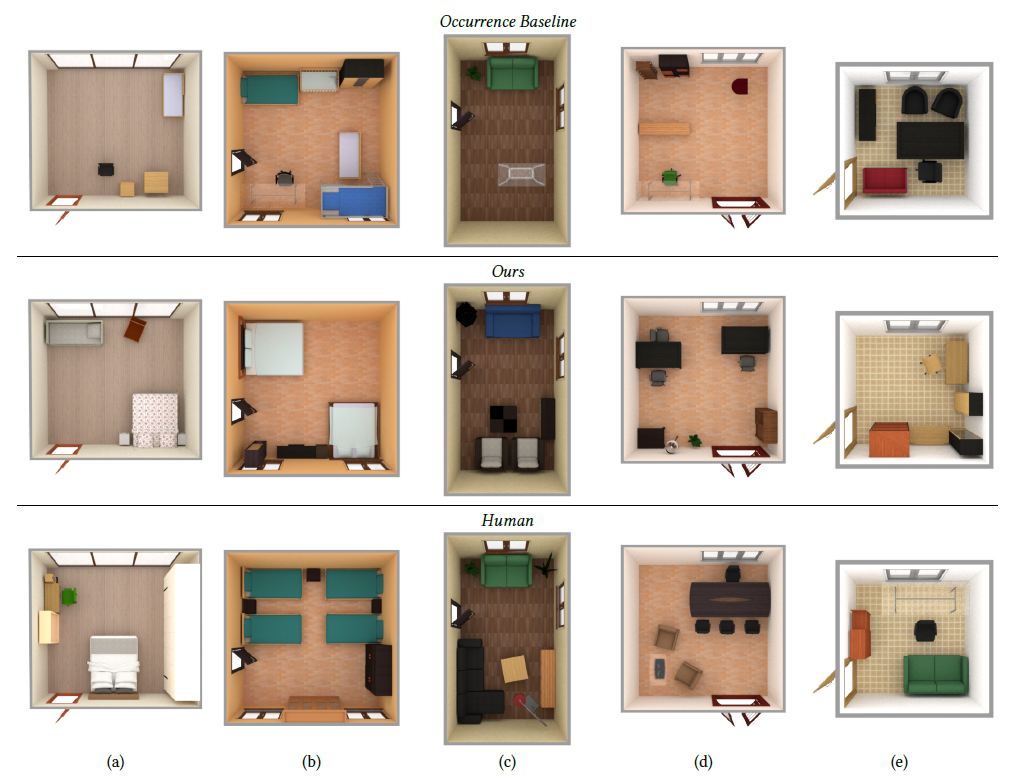


图4-2

1. **思考与扩展建议**

1）这篇文章中，生成图像的对象之间只有平面关系，这显然是和显示情况不符的，现实中，各种物件间的关系十分复杂，可能有层叠，可能有重合，为了考虑到这一点，我认为应当考虑到物件之间的关系，如床经常和床头柜同时出现，餐桌旁的椅子分布往往是对称的等等，应当把这些物件之间的关系当做特征去考虑。

2）模型的时间效率堪忧，文中提到CategoryLocation部分对于一个场景的评估可能要花几分钟时间，这种时间效率在使用中显然是不能接受的，原因是存在大量的重复计算，当图片的大小和物件的种类数量较大时，这些重复计算带来大量无意义的开销，应该寻找办法共享这些计算结果，可以参考Faster-RCNN中的RPN所使用的的方法，在评估候选对象区域时，对所有图像使用通用特征的map。

3）因为概率计算方式的限制，模型倾向于将物件的分布稀疏化，因此对于墙面上的物件（窗帘等）几乎未做考虑，如果想要加入未考虑部分，如墙面上、物件平面上等区域，当前的计算方式是需要改变的。

参考文献

[1]柴梦婷,朱远平.生成式对抗网络研究与应用进展[J/OL].计算机工程:1-16[2018-12-27].https://doi.org/10.19678/j.issn.1000-3428.0051964

[2]Chuan Zhang. Neural Networks:Efficient Implementations and Applications[A]. IEEE Beijing Section、Fudan University.Proceedings of 2017 IEEE 12th International Conference on ASIC[C].IEEE Beijing Section、Fudan University:IEEE BEIJING SECTION,2017:4.

[3]Shuang Liu. Image Captioning Based on Deep Neural Networks[A]. 上海工程技术大学.Proceedings of 2018 2nd International Conference on Electronic Information Technology and Computer Engineering (EITCE 2018)[C].上海工程技术大学:广东省人力资源研究会国际学术交流研究院,2018:7.

.