1. Problem Statement

本文试图解决的研究问题是什么?这个问题重要吗?为什么?

当前的技术状态(即其他竞争方法/解决方案)是什么?

这篇论文的动机是什么?这篇论文的定位是什么?

如果有必要，你可以用图片来更好地说明研究问题。

本文的研究问题是3D indoor scene understanding，使用一张彩色图像重建房间布局以及每个单独的物体，并估计其在三维空间中的语义。该问题是一个长期存在的计算机视觉问题，对机器人、虚拟现实等应用有重要影响。

当前的方法主要聚焦在两个焦点。一个是保留细粒度集合细节的三维形状表示。另一种是利用情境信息对场景中的多个物体和布局进行联合推断。然而杂乱的场景增加了理解的复杂性，表现不能让人满意。在Single Image Scene Reconstruction方面，最近的方法通过利用物体之间的关系从2D检测中推断3D包围盒，拓展方法用CAD模型校准检测对象。然而该方法受到CAD模型数据库大小的限制，结果不准确。一些基于学习的方法先将形状编码为特征向量并用传统方式表示，或将形状分解为简单形状的结构化表示方法。近年来隐式曲面函数被广泛应用，也有研究将结构化和隐式表示结合起来。图卷积网络已经被广泛应用，并能够从点云数据预测关系或检测3d物体。

本文提出了一个基于深度隐式表示的全面的三维场景理解的深度学习系统，提出了一种基于图像的隐式形状的场景上下文网络，利用局部对象中的信息来细化初始三维姿态和场景布局。他们还提出了一种物理侵犯损失以防止交叉对象和产生合理的对象布局。

2. Summarise the paper’s main contributions

这篇论文声称有哪些贡献?

这篇论文有什么新内容吗?这篇论文有什么创新之处?

作者是否夸大了他们的贡献，如果是，又是如何夸大的?

作者设计了一个基于单图像的使用局部结构隐式网络的两阶段整体三维场景理解系统。其次，他们提出了一种新的基于图像的隐式形状嵌入网络来提取潜在形状信息。他们还提出了一种基于GCN的场景上下文网络来细化对象的排列。最后他们还提出了一种物理侵犯损失以防止交叉对象和产生合理的对象布局。

他们给出了实验方法和数据。他们还与最先进的方法在3D Object Reconstruction、3D Object Detection、Layout Estimation、Camera Pose Estimation和Holistic Scene Reconstruction方面做了对比，结果显示他们的方法表现更好。

3. Method and Experiment

他们提出了什么方法/想法/见解? 解释他们方法的关键(关键思想/洞察力/聪明才智)。

主要观点/论点是什么?

作者如何证实他们的主张?进行了哪些实验?这些实验的主要结果或发现是什么?

如图一所示，系统分为两个阶段，the initial estimation stage with Local Implicit Embedding Network (LIEN), and the refinement stage with Scene Graph Convolutional Network (SGCN)。

在第一个阶段，首先从图像中提取二维边界盒，然后利用目标检测网络将目标恢复为三维包围盒，并利用LIEN直接从图像中学习隐式局部形状信息并推断三维形状。他们的LIEN包括一个图像编码器Resnet-18以及一个三层的MLP来获得分析和潜在代码，并将类别代码与编辑器的图像特征连接起来，在LIEN中引入先验以提升性能。输入的图像也被输入到局部估计网络（LEN）以产生一个3D布局包围框和相对的相机姿态。

在细化阶段中，他们设计了一种新的SGCN，通过场景上下文细化初始预测。他们基于Graph R-CNN把整个3D场景建模为一个图G，其中的节点表示对象、场景布局及他们之间的关系。对于不同类型的节点他们设计了不同的特性。对于布局节点，他们使用LEN图像编码器来编码布局的外观以及布局边界框和摄像机姿态的参数化输出以作为布局节点特征，并使用相机固有参数作为相机先验。对于对象节点，他们从ODN收集外观关系特征和对象边界框的参数化输出，并从LIEN收集世界坐标中的元素中心和分析代码。他们还使用2D检测器中的一类热点标签将语义信息引入到SGCN中。至于关系节点，对于连接两个不同物体的节点，使用2D物体包围盒的几何特征和两个被连接物体的盒角坐标作为特征并区分了方向。对于连接对象和布局的节点，他们使用常量初始化表示，并使用SGCN推断。他们为每个源-目的地类型定义了独立的消息传递权值。对于一个N个对象和1个布局的图，分别用Z0、Zr两个矩阵表示对象-布局节点和关系节点。With adjacent matrix αsr = αdr = 1 - IN+1, Z0和Zr的更新可以表示为：

文本

描述已自动生成

图片包含 图示

描述已自动生成

训练LIEN的损失函数为训练LEN和ODN的损失函数为文本, 信件

描述已自动生成。他们还提出了一种新的物理冲突损耗以保证目标之间不会相交。他们在每个对象内部取样点，将高斯元素的中心点作为候选点放入LDIF解码器中滤除目标表面外的点，最后将Si queue到k-nearest object的LDIF中以验证是否与其他对象相交。损失函数可以表示为

图片包含 文本

描述已自动生成

他们使用Pix3D和SUN RGB-D数据集，先分别训练每个模块，然后共同训练。

他们的Ablation Study显示GCN、deep implicit feature、physical violation loss and some other factors will influence the performance. 他们的方法在ObjectNet3D dataset上也展现出了较好的泛化能力。

5. Critical Analysis

**5.1. Are the paper’s contributions** **significant?**

贡献/改进是微不足道的、增量的吗?

为什么之前的努力失败了?

论文的贡献是显著的。相比已经存在的方法，他们给出了新的网络模型和损失函数，使用深度隐式表达提高了建模的准确度和精细度，并展现出了较好的泛化性能。

**5.2. Are the authors’ main claims valid?**

他们是否令人信服地证实了自己的主要观点?

他们的论点，推导，实验有漏洞吗?

作者的main claims是有效的。他们给出了方法的数学表示和网络模型，并对比了不同方法之间的差异，结果显示他们的方法表现更好。他们还做了烧蚀对比。

**5.3. Limitation and weaknesses**

他们的方法有什么限制/弱点吗?可以做些什么来改进这项工作?

你会如何解决/克服他们的弱点?

给出的结果图片显示，在建模时有一些细节仍然与图片存在差异。在估计场景布局时有些细节与grand truth存在差异，似乎更倾向于规整的布局。这可能是受限于数据集的丰富程度。增加数据集的数量和丰富程度可能会有帮助。

**5.4. Extension and future work**

你会建议作者做哪些额外的实验来强化结果?

你能想到论文中提出的方法/想法(假设有效)的其他可能应用吗?

未来可能的作品是什么?

作者可以尝试使用更多的数据集训练网络以提高3d建模和布局估计在细节上的准确度。作者也可以尝试合成数据集。

未来该技术可能应用于VR、机器人、设计等领域。

**5.5. Is the paper stimulating or inspiring ?**

许多论文(甚至那些已经发表的论文)都是枯燥乏味的，而有些则是令人兴奋的。你对这篇论文有什么看法?为什么?

这篇论文是令人兴奋的。他们提出的新的网络结构和损失函数使得3D建模和场景布局估计的准确性和精细度大大提高。

**5.6. Conclusion and personal reflection**

首先，对本文进行总结。

那么，如果让你来解决这个研究问题，你会有什么不同的做法?还有别的解决方案吗?

最后，用一句话总结你从阅读本文中学到的东西。

In conclusion，这篇论文提出了一个基于学习和深度隐式表达的方法来端到端地解决使用单个图像进行室内3D建模和场景布局估计问题，给出了新的网络结构和损失函数，提高了算法的性能。

如果是我，我可能会尝试在此基础上结合光流法来辅助建模。

这篇文章告诉我深度隐式表达可以更紧凑、信息丰富地表达特征测量，更有效地从上下文中学习。