

**TONGJI UNIVERSITY**

**专业型硕士研究生学位论文选题报告及工作计划**

**课 题 名 称 室内图像的家具多标签标注与箱体重建算法研究**

**学 号 1541447**

**研 究 生 马天瑶**

**专 业 软件工程**

**所 在 院 系 软件学院**

**导 师 贾金原**

**副 导 师 粱爽**

**选 题 时 间 2016.10**

**同济大学研究生院**

**2016年11月23日**

1. 研究问题

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1）在预研究的基础上提出应用研究（设计）中的科学问题  随着互联网的迅速发展，家居虚拟展示跨越了时空的局限，帮助用户直观、全面地了解到室内设计方案，因而被家居行业广泛接受。现有的家居虚拟展示技术大多采用纯3D的技术方式，其优势在于操作灵活、交互简单，具有良好的设计性。但是，该虚拟家居技术过分依赖于丰富的三维模型库，并且家居模型需要经过复杂的制作过程和大量繁琐的人工操作。  近年来，基于图像进行家居虚拟设计的方法不断涌现，在一定程度上克服了纯3D技术渲染速度慢、模型制作复杂等缺点，且素材丰富、获取容易。但是，图像往往缺少一些关键信息，如场景的深度信息、空间结构信息等。同时，室内图像中通常包含丰富的物品，物品之间的相互遮挡，这些都会给获取图像信息带来一定的挑战。  本课题在前人工作的基础上，结合图形、图像、自然语言处理等方面的内容，着力于获取丰富、准确的图像信息，设计出一个合理的、高效的解决方案来对室内图像中的家具进行多标签标注，以获取其属性信息，包括种类、位置、颜色以及材质等，并进行图像分割来实现家具的箱体重建，进而降低家居虚拟展示过程的复杂性，并提高虚拟展示的内容丰富性和精确性。  2）课题来源、选题依据和背景情况  **课题来源：**  本课题来源于国家自然科学基金面上项目(No.61272276)；国家自然科学基金青年项目(No.61305091)；国家十二五计划重大科技支撑项目(No. 2012BAC11B00-04-03)；高等学校博士学科点专项科研基金项目（博导类）(No.20130072110035)；中央高校基本科研业务费专项资金(No.2100219054)。  **选题依据：**  在过去的家居行业中，购买者主要通过卖方提供的家居效果图和家居实物作为参照在脑海中形成家居效果图，从而决定具体要购买的家具。这种传统的选择方式会经常出现因大小不合适，颜色不搭配等产生的实际效果与预想效果不符的问题，从而引发购买者的退货或是换货行为，造成不必要的经济损失。利用虚拟现实技术将三维虚拟对象与真实世界物体进行合成与交互，为家居行业中的应用问题提供了新的技术解决手段。借助于虚拟现实技术，可以为家居购买者快速生成所要购买的家具在房间之中的三维实景图，更为直观、全面的为购买者提供家居的效果展示。此项技术不仅可以方便用户更好的选择中意的家居，并减少因为实际效果与想象效果不符而引起的退换货问题，还可以扩展家居卖方的销售方式增加家居销量。因此，家居DIY算法拥有很广泛的应用需求和重要的研究价值。  在现今的家居DIY算法中，主要采用的技术方式有：（1）3D家居环境+3D家具模型（3D+3D）；（2）2D家居环境背景图像+2D家具图像（2D+2D）；（3）2D家居环境背景图像+3D家具模型（2D+3D）。  这三种技术方式在场景的真实感、过程的复杂性、操作的灵活性、交互的易用性方面各有特点,具体的对比见下表。  表1 家居DIY技术性能对比   |  |  |  |  |  | | --- | --- | --- | --- | --- | | 家居DIY采用的技术方式 | 场景的真实感 | 过程的复杂性 | 操作的灵活性 | 交互的易用性 | | 3D+3D | 低 | 高 | 高 | 高 | | 2D+2D | 高 | 低 | 低 | 低 | | 2D+3D | 高 | 低 | 高 | 高 |   **3D+3D**是现在受到各大家居企业青睐的虚拟家居展示方式，利用其操作的灵活性和交互的易用性，使客户得到身临其境的漫游体验和随心所欲的编辑体验。图形图像实验室自主开发的HouseDIY[10]，就是采用这种3D家居环境模型+3D家居模型的技术方式。此项目使用Flash 3D作为3D引擎，在轻量化、实时性、易用性方面都取得了良好的效果。但是，在场景的真实性和建模过程的复杂性方面，这种3D+3D的技术方式表现得差强人意。如何能让生成的虚拟场景更加自然，让家居模型更加真实的同时，简化建模和渲染的过程，一直是研究者们想要攻克的技术性难题。  hdiy  图1 HouseDIY效果图  **2D+2D**由于利用真实世界的图像，在场景的真实感和制作过程的复杂性方面，相比3D家居环境模型+3D家居模型的方式，有了明显的提高。但正是因为图像所包含的信息是部分的、缺失的，所以这种虚拟家居展示方式所能满足的用户交互操作非常有限。虽然，这种技术方式并不能带来更好的用户体验，却在虚拟家居研究中有着很好的启发作用。  575897802350563837  图2 2D家居DIY效果图  **2D+3D**是将2D和3D进行融合，利用两者的优点来弥补它们的不足并从中找到平衡。基于虚实融合的3D家居DIY算法就是对这种技术进行扩展研究。其中，一种基于图像三维重建的、可以对物体进行交互的图像交互算法有很好的启发作用和应用扩展价值。在该算法中，主要是利用家居是人造物品具有很好的规则性这一特点对家居做简单的建模，以达到能够探查到图像中物体之间的相互关系，能够对图像的需要编辑的物体进行简单的交互操作，并保证生成图像的真实性。  **背景情况：**  **1）箱体重建**  目前，实验室主要关注2D+3D的家居DIY设计，并取得了一定的成果。硕士生王海菲研究了基于灭点的单张图片的三维箱体重建，最终推测出场景布局。该过程主要分为三个具体阶段：首先，通过灭点生成数量固定的场景候选布局，并通过布局评价函数筛选出与图像内容最为相符的初级场景布局；之后，通过图像特征可以提取图像内显著物体的立方体模型，并利用这些物体模型对场景布局进行优化；最后，通过灭点进行相机参数的推导，从而将图像内的二维点还原到三维空间中。  其中，在场景内物品模型生成时，使用了两种不同的图像特征：由Lee[68]的线段推移算法得到的平面方向特征和通过Hoiem[69]的图像区域分类算法得到的几何内容特征。在得到两种图像特征后，通过对图像特征数据的推测可以自动检测出图像中较为重要的物品，如椅子，桌子，柜子等。首先，通过平面特征的分析找到图像中物品的立方体模型。对平面特征中属于两种不同方向的平面进行成对地检测，取两个平面上的三个角点，则可以判定是否这两个平面能确定一个立方体的平面投影。对于可以组成立方体的每组平面，找到其最为合适的三个角点，并生成相应的立方体模型。之后，将得到的立方体模型与图像的几何特征相结合，从而得到筛选后的物体模型。对于每个立方体模型，计算其图像投影内所包含的潜在物体像素与投影内像素的比率，并将比率较低的立方体模型移除。  这种生成立方体的方法存在以下缺点：  （1）生成的立方体模型比较杂乱，无法准确表示图像中的物品；  （2）只能对图像内比较显著的物品进行建模，无法满足用户建模的要求；  因此，这种建模方法并不能满足家居虚拟交互的需要。于是，如何准确地识别出图像中的家具需要进一步地研究。    图3 物体模型生成流程图  **2）特征提取**  在早期计算机视觉领域的应用中，对于图像特征的提取主要采用尺度不变特征变换[2]（Scale-invariant feature transform, SIFT）和方向梯度直方图[3]（Histogram of oriented gradients, HOG）等方法，然后将提取到的特征作为输入用于系统中。这些特征本质上是一种手工设计的特征，针对不同的问题，提取到的特征好坏对系统性能有着直接的影响。由于其泛化能力较差，因此难以在实际应用问题当中得到精准的效果。  2006年，机器学习领域的泰斗Geofery Hinton和他的学生Ruslan Salakhutdinov在国际顶尖学术刊物《Science》上发表了一篇文章[64]，第一次提出了深度学习的思想。与传统机器学习方法相比，深度学习具有更强大的特征学习和特征表达能力，成为了近些年来学术界的研究热点之一。2011年以来，研究人员首先在语音识别问题上应用深度学习技术，将准确率提高了20% ~ 30%，取得了十多年来最大的进展。仅一年之后，基于卷积神经网络的深度学习模型就在大规模图像分类任务上取得了非常大的性能提高[1]。可见，深度学习在语音识别、图像识别和自然语言处理等领域获得了突破性的进展。  目前，深度学习在人脸识别[65]、行人检测[66]、自动驾驶[67]等应用领域取得了显著的成果，但是在家居智能交互领域尚未得到应用。  综上所述，本课题将对深度学习在图像标注和图像分割方向的应用进行研究，并寻求一种解决方案能够结合图形学中的方法，将其运用至家居智能交互系统中，获取丰富、准确的图像信息，对室内图像中的家具进行多标签标注，并进行图像分割来实现家具的箱体重建，在一定程度上对家居DIY算法在家具识别上的精确度进行优化，为家居智能交互作了很好的铺垫。  3）课题的研究目标以及理论意义和实际应用价值  **研究目标：**  本课题以家居虚拟展示为实际应用背景展开研究，主要针对的是家居物品，家居物品属于人造物品的范围，因此具有结构上的规则性和非局部关系的丰富性的特点。利用该特点，通过对室内图像的特征进行分析，从而得到虚拟对象更为丰富的信息以及其在真实空间中的定位，进而重建出家具的三维箱体模型，以近似表示室内家居图像中的家具模型。从而为达到模仿真实世界中的操作，实现对家居物品的交互而奠定了基础，也为家居自动布局提供了丰富的案例。  **理论意义：**  在理论方面，本课题涉及到了图像标注、图像分割、物体的近似三维重建等方面，结合了计算机视觉、自然语言处理和计算机图形学等领域的研究内容，着力于获取丰富、准确的图像信息，使得家居虚拟展示过程的复杂性得到降低，并提高虚拟展示的内容丰富性和精确性，对2.5D的家居DIY算法进行一定的完善和优化，具有较高的科学价值。  **实际应用价值：**  本课题所提出的方法完全摆脱了通过用户手动标注物体去进行建模，而是通过对图像特征的学习能够自动获取家具的属性信息，并且获取的属性信息更为详细，生成的家具箱体模型更为精确、不杂乱，不仅为家居图像智能交互奠定了扎实的基础，也为家居自动布局提供了丰富的案例。因此，本课题在家居虚拟展示的研究问题上作出了较大的贡献。这样势必可以大力推动移动互联网在线可视化电子商务的发展，并给家居行业带来无限的商机。 |

2．文献综述（文献综述不得少于2000字）

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1）国内外在该研究方向的研究现状及发展动态  本课题主要从图像标注和图像分割两个方面展开研究，具体的研究现状和发展状态如下：  **1. 图像标注**  近几年来，鉴于图像标注技术的实用性和有效性等特点，该项技术己经渐渐发展成为多媒体技术研究领域的热点问题，以及图像检索领域研究的核心内容。在互联网图像数据的急剧增长下，对大规模和开放的互联网图像数据进行标注是主要研究的重点[24]。  图像的标注问题经历了早期的人工标注阶段，发展到现今的由计算机技术实现图像标注，不论是国内，或者是国外的研究人员，都在这方面花费了大量的时间和精力。  **1.1 国外研究现状**  图像标注技术能够大大改善人工标注带来的各种弊端，并提高了图像管理、检索的效率。国外对于图像标注技术的研究起步较早，其发展历程如表2所示。1999年，Mori Y等人提出了一种在图像与语义之间建立映射关系的方法[25]。  这种方法首先将图像按照区域关键词的不同划分为各个子块，然后提取各个子块的特征向量[26]。利用这种共生模型对图像进行标注，即对图像子块进行关键词提取的过程。随着共生模型的诞生，开启了图像标注技术这一新的研究领域。随后，K Bamard[27]等人将统计模型以及全局特征等方法运用于图像标注技术中，大大的改善了标注的性能，使图像标注技术得到进一步发展[28]。  近年来，鉴于自动图像标注技术所具有的优点及其商业价值，对这一技术的研究受到了越来越多学者和研究员的青睐，从而出现大量新的技术和研究方法，尤其是机器学习[29]的兴起，促进了自动图像标注技术的进一步发展。随之，Hinton[30]提出了深度信念网，它由一组受限玻尔兹曼机（Restricted Boltzmann Machines，RBM）组成，可以实现自主地进行特征学习，这一发现促使了对深度学习[31]领域的研究，并将这项技术应用于图像分类标注领域中。2009年，Kai Yu等人实现通过Sparse Coding[32]的方法进行特征学习，并用于实现图像的分类。2010年，Honglak Lee等人[33]使用无监督特征学习方法实现了分层的特征学习，完成对图像语义表征，并用于分类和标注中，这一研究成果，掀起了图像标注研究领域的又一新的浪潮。2011年，Marc’Aurelio Ranzato等人[34]利用深度学习的思想，设计实现了深度生成模型完成特征学习，并将该模型应用于图像识别和分类工作中。深度学习的兴起，促使大量的科研工作者涌向特征学习领域，并将该研究成果应用于其他领域，使得图像识别、分类、标注以及检索工作得到了进一步的完善。  表2 国外图像标注技术的发展历程   |  |  |  |  | | --- | --- | --- | --- | | 发展阶段 | 代表人物 | 研究内容 | 研究成果 | | 第一阶段 | Mori Y | 提出图像子块图关键字的映射方法 | Word co-occurrence model | | 第二阶段 | K Barnard, D A Forsyth | 提出图像集合语义信息的统计模型 | Hierarchical Aspect Cluster Model | | J Z Wang, J Li | 介绍一种统计建模方法，解 决图像语义自动索引问题 | Statistical models | | A Yavlinsky, E Schofield | 介绍了一种基于图像全局特征的图像自动标注框架 | 基于图像全局特征的图像内容标注 | | D Blei, A Ng, M Jordan | 提出结合LDA生成模型的图像内容标注方法 | Latent Dirichlet Allocation model | | 第三阶段 | J Jeon, R Manmatha | 提出使用最大熵方法进行图像内容标注 | 最大熵模型 | | C Cusano, GCiocca | 结合SVM对已标注图像进行分类，实现对未知图像的标注 | 基于SVM的图像标注方法 | | Honglak Lee, Yan Largman | 提出无监督特征学习并运用于图像分类与标注中 | Deep Learning |   **1.2 国内研究现状**  自动图像标注技术的研究在国内兴起较晚，从理论到广泛运用于实践中大约经历了近十年的时间。随着相关领域技术的快速发展，自动图像标注技术也在不断完善中。现阶段，按照类型的不同，标注模型主要可以分成以下两大类[35]：即基于全局特征的自动图像标注方法和基于区域划分的自动图像标注方法。其中后者又分为基于分类的自动图像标注方法、基于概率关联模型的自动图像标注方法和基于图学习的自动图像标注方法三类[36]。  **（1）基于全局特征的自动图像标注方法。**全局特征指的是图像的视觉特征，如颜色特征、纹理特征等底层特征。这种方法的核心思想就在于利用这些底层视觉特征，对图像内容进行识别，得到视觉特征与其高层语义的对应关系，利用某种模型或判别规则，实现图像内容的标注[37]。这种以图像特征为基础的方法，采用一个特征或是利用混合特征进行图像标注，实现对图像对象进行分析。概括来讲，这种方法将图像的标注问题转换成了图像对象的识别问题。  当人们对一幅图像进行观察时，只会关注自己感兴趣的部分，即视觉系统会将图像自动地划分为前景区域和背景区域[38]。因此，当使用基于全局特征的图像标注方法进行图像标注时，标注信息更多的是侧重于描述图像中的主要对象语义信息，而忽略了所谓的背景及其他对象信息。这种标注方法给出的图像信息相对简单，对于查询图像内容相对丰富时，可能会存在偏差。这种标注方法只适用于标注图像内容较为单一的图像，而当图像内容较为复杂时，只能实现图像语义内容的粗略描述，忽略了图像的其它语义部分。  **（2）基于分类的自动图像标注方法。**分类即按照图像内容、性质的不同，将其分为不同的种类。这种标注方法是相对较为直观的，它通过对图像语义进行分类处理实现标注问题。其核心思想是将图像的语义标注关键词按照其性质，表述内容的不同进行划分，分为若 干个类别，每个类别均用一个类别标记进行描述，则对图像进行标注的过程就转换成了对图像按照其语义不同进行分类的过程[39]。所以，可以从对图像进行分类的角度出发，来解决图像的标注问题。然而，图像标注问题并不是纯粹的分类问题，它具有一定的特殊性，因为一幅图像中可能包含一个或多个对象， 使其语义信息较为丰富，用于标注的关键词可能属于多个类别，则标注问题可以进一步通过多标记学习问题来描述。就图像而言，将整幅图像看作是由多个示例(即内容区域)组成的用于训练的包。如果对于概念集合中的每一个概念， 包中均可找到一个与之对应的示例，那么该包则可标记为正包。  这种标注方法的转换，需要某些先验知识以及训练样本集，来实现分类模型的建立。因此，对于训练样本的选择，训练样本集的大小，以及分类器的选择在一定程度上都会影响标注结果的效果。  **（3）基于概率的自动图像标注方法。**这种方法与基于分类的图像标注方法一样，都属于基于区域划分的图像标注方法。因此，其基本思想是类似的。基于分类的图像标注方法是通过训练分类器，对待标注图像进行学习，判断其类别属性。而这里，则是利用概率统计 模型，计算图像区域特征与其语义信息之间的联合概率关系[40]，实现对未知信息图像的标注。  **2. 图像分割**  图像分割的研究已经有四十余年的历史。在此期间，国内外学者们为之付出了不懈的努力和汗水。然而，到目前为止，令学者们感到非常遗憾的是一种分割算法至少对某一类图像的分割是失败的，同时一幅图像至少对某一种分割算法是不适用的。因此，为了解决某些具体的问题，各种具有针对性目的的分割算法层见叠出。虽然图像分割方法已有几十种，但是从基本上来讲图像分割主要分为[41]：基于阈值的图像分割、基于边缘的图像分割、基于区域的图像分割和结合一些特殊方法和工具的图像分割等。  **（1）基于阈值的图像分割。**是指在图像灰度范围内寻找一个灰度阈值，然后将图像中像素的灰度值与之比较，将图像分成目标和背景两类。此方法简单意义上讲具有计算简单、运算效率高的优点，除此之外它还可以极大的压缩数据量，大大简化了图像信息的分析和处理步骤。因此阈值分割方法在过去的几十年间备受重视和广泛的应用。在过去的研究中基于阈值的分割方法有:直方图与直方图变换法[42]、最大类间方差法[43][44][45]、最大熵自动阈值法[46]、最小误差的分割方法[47]以及其推广等。  **（2）基于边缘的图像分割。**是在边缘概念的基础上提出的，边缘[41]是指所要提取的目标和背景的分界线，是图像中从一个区域到另一个区域过度的表示，是图像分割所依据的重要特征。因此基于边缘的图像分割也是最早的图像分割技术。该方法的分割原理是构造对图像灰度阶跃变化敏感的差分算子，所应用的差分算子主要有：Robert算子、Sobel算子、Prewitt算子、Laplacian算子等[48][49]。  **（3）基于区域的图像分割。**是以区域为基础，把具有某种相似性质的像素连通起来， 从而构成最终分割区域的方法。区域分割的分类[41]主要有：以像元为出发点的区域归属，将属性接近的连通像元聚集为区域的区域增长分割方法；以全图出发，按区域属性特征一致的准则，决定每个像元的区域归属形成区域图的区域生长分割方法；以及将以上两种方法综合起来的分裂-合并方法。  除上述分割方法外，近年来对结合特殊方法和工具分割方法的研究也变得越来越普遍，当前关于特殊理论分割方法的研究主要有：基于数学形态学的分割技术[50]；基于人工神经网络技术的图像分割方法[51]；基于模糊技术的分割方法[52]；基于遗传算法的分割方法[53]；结合蚁群算法的分割方法[54]；基于能量思想的分割算法[55]；基于水平集的分割方法[56]；基于信息论的图像分割技术[57]；基于图论的分割方法[58][59][60]；使用小波变换运用到图像中的分割方法[61]等。  2015年，Long J等人[62]提出可以将一些深度神经网络改为全卷积网络来做图像分割。他们首先利用一些流行的分类网络（AlexNet，VGGNet，GoogleNet），在保留一些它们在图像分类方面训练所得参数基础上，进行“修剪”，转变为针对图像分割的模型。然后，他们将一些网络较深的层的所得特征和一些较浅的层所得特征结合起来，最后用一个反卷积层放大到原始图像大小来提供一个更为准确的分割结果，并称之为跳跃结构。随着深度学习的兴起，将推动图像分割方法向更快速、更精确的方向发展。  2）研究问题在本研究领域应用上的地位与价值  目前，基于图像的家居DIY算法[11]是根据传统的图像特征提取方法（线段推移算法和图像区域分类算法）来提取室内场景中显著物品的立方体模型。这样得到的家具模型杂乱，并且定位不精确，需要通过用户手动标注进行修正。  本课题着力于获取丰富、准确的图像信息，提出一个基于深度学习的方法，对室内图像中的家具进行多标签标注，并且进行图像分割来实现家具的箱体重建。该方法能获取到更为详细的属性信息，为家居自动布局提供了丰富的案例。同时，该方法完全摆脱了通过用户手动标注物体去进行建模，生成的家具箱体模型更为精确、不杂乱。进而降低家居虚拟展示过程的复杂性，并提高虚拟展示的内容丰富性和精确性，为家居智能交互作了很好的铺垫。  综上所述，本课题在虚拟家居展示系统的研究领域上具有一定的地位和价值。 |

附：参考文献

|  |  |
| --- | --- |
| 序号 | 文献目录（作者、题目、刊物名、出版时间、页次） |
| 1 | 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理, 2016, 31(1):1-17. |
| 2 | Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(60):91-110. |
| 3 | Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. IEEE Conference on Computer Vision & Pattern Recognition. 2005:886-893. |
| 4 | Lecun Y L, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. [J]. Proceedings of the IEEE, 1998, 86(11):2278-2324. |
| 5 | Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2):2012. |
| 6 | Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks[M]. Computer Vision – ECCV 2014. Springer International Publishing, 2013:818-833. |
| 7 | Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014. |
| 8 | Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[J]. 2015:1-9. |
| 9 | He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. Computer Science, 2015. |
| 10 | 陈超, 张博, 贾金原. 基于Web的在线家居设计系统[C]. 2010国际数字科技博物馆学术论坛暨数字科技馆技术与应用研讨会. 2010. |
| 11 | 王海菲, 贾金原, 谢宁. 复杂室内图像的灭点检测与箱体重建方法[J]. 计算机科学与探索, 2016, 10(5). |
| 12 | Chung J, Gulcehre C, Cho K H, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. Eprint Arxiv, 2014. |
| 13 | Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780. |
| 14 | Schuster M, Paliwal K K. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681. |
| 15 | Irsoy O, Cardie C. Opinion Mining with Deep Recurrent Neural Networks[C]. EMNLP. 2014: 720-728. |
| 16 | Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. Computer Science, 2014:580-587. |
| 17 | Girshick R. Fast R-CNN[J]. Computer Science, 2015. |
| 18 | Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016:1-1. |
| 19 | Zitnick C L, Dollár P. Edge Boxes: Locating Object Proposals from Edges[M]. Computer Vision – ECCV 2014. Springer International Publishing, 2014:391-405. |
| 20 | Karpathy A, Joulin A, Li F F. Deep Fragment Embeddings for Bidirectional Image Sentence Mapping[J]. Advances in Neural Information Processing Systems, 2014, 3:1889-1897. |
| 21 | Karpathy A, Li F F. Deep visual-semantic alignments for generating image descriptions[C]. Computer Vision and Pattern Recognition. IEEE, 2015:3128-3137. |
| 22 | Johnson J, Karpathy A, Fei-Fei L. Densecap: Fully convolutional localization networks for dense captioning[J]. arXiv preprint arXiv:1511.07571, 2015. |
| 23 | Dai J, He K, Sun J. Instance-aware Semantic Segmentation via Multi-task Network Cascades[J]. Computer Science, 2015. |
| 24 | Zhigang Ma, Feiping Nie, Yi Yang, Jasper R.R.Uijlings, Nicu Sebe. Web Image Annotation Via Subspace—Sparsity Collaborated Feature Selection[C]. IEEE Transactions on Multimedia, 2012, 14(4):1021-1030. |
| 25 | Ameesh Makadia, Vladimir Pavlovic, Sanjiv Kumar. Baselines for Image Annotation[J]. International Journal of Computer Vision, 2010, 90:88-105. |
| 26 | Feichao Wang. A Survey on Automatic Image Annotation and Trends of the New Age[J]. Procedia Engineering, 2011, 23:434-438. |
| 27 | Rui Jesus, Arnaldo J, Abrantes, Nuno Correia. Methods for automatic and assisted image annotation[J]. Multimed Tools Appl, 2011, 55:7-26. |
| 28 | Yong Wanga, TaoMei, Shaogang Gong, Xian-Sheng Hua. Combining global, regional and  contextual features for automatic image annotation[J]. Pattern Recognition, 2009, 42:259-266. |
| 29 | Nasullah Khalid Alham, Maozhen Li, Suhel Hammoud, Hao Qi. Evaluating Machine Learning Techniques for Automatic Image Annotations[C]. Fuzzy Systems and Knowledge Discovery, Sixth International Conference on, 2009, 7:245-249. |
| 30 | Geoffrey Hinton. A Practical Guide to Training Restricted Boltzmann Nachines[M]. Neural Networks: Tricks of the Trade. Springer Berlin Heidelberg, 2010:599-619. |
| 31 | Itamar Arel, Derek C.Rose, Thomas P.Karnowski. Deep Machine Learning-A New Frontier in Artificial Intelligence Research[C]. IEEE Computational Intelligence Magazine, 2009:13-18. |
| 32 | Jianchao Yang, Kai Yu, Yihong Gong, et a1. Linear spatial pyramid matching using sparse coding for image classification[C]. Computer Vision and Pattern Recognition(CVPR), IEEE, 2009:1794-1801. |
| 33 | Honglak Lee. Unsupervised Feature Learning via Sparse Hierarchical Representations[D]. America: Stanford University, 2010. |
| 34 | Marc’Aurelio Ranzato, Joshua Susskind, Volodymyr Mnih, et a1. On deep generative models with applications to recognition[C]. Computer Vision and Pattern Recognition(CVPR), 2011 IEEE Conference on. IEEE, 2011:2857-2864. |
| 35 | 许红涛, 周向东, 向宇, 施伯乐. 一种自适应的Web图像语义自动标注方法[J]. 软件学报, 2010, 21(9):2183-2195. |
| 36 | 王长虎. 互联网环境下大规模图像的内容分析、检索和自动标注的研究[D]. 合肥：中国科学技术大学博士学位论文, 2009. |
| 37 | 虎晓红. 用于图像检索的语义标注技术的研究[D]. 北京：中国矿业大学博士学位论文, 2010. |
| 38 | 张树臣. 融合颜色和形状特征的图像检索技术[D]. 吉林：吉林大学硕士学位论文, 2012. |
| 39 | 程环环. 基于贝叶斯网络的图像内容表述与分类[D]. 长沙：国防科技大学博士学位论 文, 2011. |
| 40 | 宋海玉. 自动图像标注及标注改善算法的研究[D]. 长春：吉林大学博士学位论文, 2012. |
| 41 | 章毓晋. 图像分割[M]. 北京：科学出版社, 2001. |
| 42 | 吴薇, 刘军, 李旭霞. 图像分割中的阈值选取方法[J]. 西安工业学院学报, 2002, 22(4):309-313. |
| 43 | Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems Man and Cybernetics, 1979, 9(1):62-66. |
| 44 | Liao P S, Chen T S, Chung P C. A fast algorithm for multilevel thresholding [J]. Journal of  Information Science and Engineering, 2001, 17(5):713-727. |
| 45 | 孔明, 孙希平, 王永骥. 一种改进的基于类间方差的阈值分割[J]. 华中科技大学学报：自然科学版, 2004, 32(7):46-47. |
| 46 | Pun.T. A new method for gray-level picture thresholding using the entropy of the histogram[J]. Signal Processing, 1980, (2):223-237. |
| 47 | Kittler J, Illing worth J. Minimum error thresholding[J]. Pattern Recognition, 1986, 19(1):  41-47. |
| 48 | Gonzalez R C, Woods R E. Digital Image Processing[M]. New York:Prentic Hall, 2002. |
| 49 | Yang Sheng-ke, Zhang Chao. Advanced Algorithm for Edge Detection Using SUSAN Operator[J]. Chinese:Tactical Missile Technology, 2007, (6):83-86. |
| 50 | 史健芳, 张富军. 基于小波变换和数学形态学的图像分割算法[J]. 太原理工大学学报, 2009, 40(5):490-493. |
| 51 | 杨治明, 王晓蓉, 彭军. BP人工神经网络在图像分割中的应用[J]. 计算机科学, 2007, 34(4):234-236. |
| 52 | YOU J, COHEN H A. Use of fuzzy set in textured image segmentation[C]. International Symposium on Signal Processing and Its Applications Gold Coast The International Society for Optical Engineering, 1996, (6):585-588. |
| 53 | Jiang Tian-zh, Yang Fa-guo. A parallel genetic algorithm for cell image segmentation [J]. Electronic Notes in Theoretical Computer Science, 2001, 46(8):1-11. |
| 54 | 韩彦芳, 施鹏飞. 基于蚁群算法的图像分割方法[J]. 计算机工程与应用, 2004, (18): 5-7. |
| 55 | 丁红, 张晓峰. 基于能量思想的图像分割算法[J]. 计算机应用与软件, 2011, 28(1): 259-262. |
| 56 | Chunming Li, Chenyang Xu, et al. Distance regularized level set evolution and its application to image segmentation [J]. IEEE transactions on image processing, 2010, 19(20):3243-3254. |
| 57 | Junmo Kim, John W.Fisher, et al. A Nonparametric Statistical Method for Image Segmentation Using Information Theory and Curve Evolution [J]. IEEE Transactions on Image Processing, 2005,14(10):1486-1502. |
| 58 | 刘松涛, 殷福亮. 基于图割的图像分割方法及其新进展[J]. 自动化学报, 2012, 38(6):911-922. |
| 59 | Tao wenbing. Interactively multiphase image segmentation based on variational formulation and graph cuts[J]. Pattern Recognition, 2010, (43):3208-3218. |
| 60 | 龚劬, 廖武忠, 卢力, 余维. 基于图论的快速FCM图像分割算法[J]. 计算机工程, 2012,38(8):192-197. |
| 61 | 李朝晖, 陈明. 基于小波和自学习神经网络的图像分割[J]. 计算机应用研究, 2006, 23(1):246-249. |
| 62 | Long J, Shelhamer E, Darel T. Fuly convolutional networks for semantic segmentation[C]. Computer Vision and Patern Recognition (CVPR), IEEE Conference on. Boston USA:IEEE, 2015:3431-3440. |
| 63 | Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context[M]. Computer Vision – ECCV 2014. Springer International Publishing, 2014:740-755. |
| 64 | Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507. |
| 65 | 林妙真. 基于深度学习的人脸识别研究[D]. 大连理工大学, 2013. |
| 66 | 肖同录. 基于深度学习的行人检测方法研究[D]. 山东科技大学, 2015. |
| 67 | Chen, Xiaozhi, et al. Monocular 3d object detection for autonomous driving. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. |
| 68 | Lee D C, Hebert M, Kanade T. Geometric reasoning for single image structure recovery[C]. 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009:2136-2143. |
| 69 | Hoiem D, Efros A A, Hebert M. Recovering Surface Layout from an Image[J]. International Journal of Computer Vision, 2007, 75(1):151-172. |

3．研究内容

|  |
| --- |
| 明确研究对象、研究内容及工作范围  **研究对象：**  本课题拟设计出一种基于深度学习的解决方案，通过分析室内图像中的特征，尽可能详细得获取家具的属性信息，然后重建出可近似表示家具模型的三维箱体模型。该方法主要利用图像标注、图像分割和空间坐标转换的手段对二维家居图片进行家具的多标签标注以及三维箱体重建，形成虚拟的家居环境，从而降低了家居虚拟展示过程的复杂性，并提高虚拟展示的内容丰富性和精确性。  **研究内容：**  本课题旨在研究利用用户输入的图像，自动地对图像特征进行分析，从而得到虚拟对象更为丰富的信息以及其在真实空间中的定位。该方法主要包含以下几个过程：首先，利用深度学习对带有分割和标注的图像数据集分别进行训练，得到已训练的图像标注网络和图像分割网络；其次，将室内图像输入至训练好的模型中，完成图像的多标签标注和分割，以获取家具的属性信息，包括种类、位置、颜色以及材质等；然后，根据已有的灭点检测算法，获取初始相机的位置；最后，结合相机的坐标信息和家具的像素级坐标信息，通过空间坐标转换重建出家具的三维箱体模型，以近似表示室内家居图像中的家具模型。  研究内容大致可以分为三个模块：训练深度学习网络、室内图像处理和家具箱体重建。其中，图像处理模块包含图像多标签标注、图像分割以及灭点检测三个部分。总体架构图如下所示：    图4 总体架构图  **工作范围：**  对于每个模块，所需完成的工作具体如下：  **1）深度学习神经网络的训练**  首先，选择合适的带有标注信息和分割信息的图像数据集。在选择时，需要考虑数据集中包含的图像数量、图像是否包含多个物体等方面。接着，选择合适的深度学习框架来搭建相应的网络模型。在选择时，需要考虑实现模型所使用的语言、适用的操作系统、适用的模型、是否支持GPU等因素。然后，在明确需要调试的超参数以及损失函数后，初始化权重、偏置常量等参数，选择合适的参数更新方法（例如SGD、Momentum等），将两个数据集分别输入至之前搭建好的图像标注和图像分割网络模型进行训练。在训练过程中，超参数取不同的值进行多次训练，选择其中表现最好的超参数。最后，将训练好的模型保存。  **2）图像的处理**  将室内图像输入至已训练的标注网络中，得到图像中包含物体的区域的坐标信息以及对应的物体的属性信息，完成家具的多标签标注。  将室内图像输入至已训练的分割网络中，得到图像中包含物体的区域的坐标信息以及对应的物体的像素级mask，进而可以计算出物体的像素级坐标信息。另外，将室内图像输入至已有的灭点检测算法中，获得该图像的相机内外参数信息。  **3）物体的箱体重建**  结合物体的像素级坐标信息和相机的内外参数信息，进行坐标的空间转换。由于家具的形状较为规则，所以只需要选取家具顶点的坐标，便可以构建立方体，完成物体的箱体重建。  于是，要完成这三个模块的工作，本课题主要需要实现以下两个方面的内容：  **1）训练基于Faster R-CNN的图像标注和图像分割网络**  在训练图像标注和图像分割网络时，本课题都将使用Faster R-CNN[18]来生成图像的region proposal。Faster R-CNN的损失函数可定义为：  （1）  其中，i表示第i个anchor，表示预测其是物体的概率，表示预测bounding box的4个参数化坐标。分类损失是一个二值分类器的softmax loss，回归损失。和是两个normalization参数，是平衡因子。  **a）图像标注网络：**用于对图像中的物体进行带有多标签的物体分类。  首先，将Faster R-CNN的最后一层与两个全连接层相连，每个全连接层都使用ReLU和Dropout。然后，把上述生成的图像特征和Word2vec生成的词向量特征作为输入，连接到RNN语言模型（GRU）中。  在训练RNN时，对于文本，其词向量标记为。其中，表示区域编码，对应特定符号<START>，对应特定符号< END>。在测试时，只需要输入，当输出的预测值为<END>，当前过程终止，完成该物体的多标签标注。  RNN的损失函数通常被定义为交叉熵（cross entropy）。t时刻的损失函数计算如下：  （2）  对于大小为T的文本，其损失函数定义如下：  （3）    图5 图像标注网络的架构  **b）图像分割网络：**用于对图像中的物体进行像素级的物体定位。  将Faster R-CNN的最后一层与两个全连接层相连，第一个全连接层使用ReLU，第二个全连接层进行pixel-wise mask的逻辑回归。最终的输出是对每一个proposal进行的像素级mask分割。  在回归训练时，其损失函数定义如下：  （4）  其中，表示整个网络需要优化的参数，表示一系列的bounding box，表示box的各种参数，表示一系列的mask，表示通过sigmoid激发后的输出。    图6 图像分割网络的架构  **2）基于灭点的物体三维箱体重建**  在进行基于灭点的物体箱体重建时，本课题的工作在于计算图像上的点对应的三维空间的坐标，因此需要定义图像坐标与三维空间坐标的转换关系，并且根据灭点信息重新定义相机的内外参数，包括标定矩阵K，世界坐标系到相机坐标系的旋转矩阵R和平移向量t。  假定空间中存在一点P，在世界坐标系下的齐次坐标记为，在图像平面上对应的投影点的齐次坐标记为。C为投影中心，投影中心到图像平面的垂线称为相机的主轴Z，其交点为主心，投影中心C到主心c的距离为焦距f。由针孔模型可知三维空间坐标到图像坐标的变换定义如下：  （5）  a）**标定矩阵K的定义：**假定摄像机和像素都是理想状态，那么标定矩阵K可以直接由代表三组正交方向的灭点计算得到。代表三个相互正交灭点，其在图像平面的坐标记为，在相机坐标系下的坐标记为，其中，。可知，主心p即为以灭点为顶点的三角形的垂心，焦距f满足关系：  （6）  b）**旋转矩阵R的定义：**已知灭点代表世界坐标系中三个方向的无穷远点，且这三个方向分别为世界坐标系中三个主轴的方向。记V为灭点方向在世界坐标系下的单位向量，为灭点方向在相机坐标系下的单位向量，则，其中V为单位矩阵。由此可得，其中，的计算公式为：  （7）  c）**平移向量t的定义：**，是相机中心在世界坐标系中的三维坐标。    图7 基于灭点的坐标转换示意图 |

4．拟解决的关键技术

|  |
| --- |
| 明确工作中的关键技术难点，提出解决的方法  基于上述的研究内容分析，拟解决的关键技术如下：  **1. 图像标注和图像分割网络的end-to-end学习**  为了给模型更多可以根据数据自动调节的空间，增加模型的整体契合度，尽可能使模型从原始输入到最终输出，实现end-to-end学习，本课题将在图像分割和图像标注网络中增加以下两个层：  **1）RoI pooling层**  由于RoI的尺寸不固定，为了保证每个RoI的后续工作能够正常进行，需要有一个RoI pooling层来提取出固定维度的特征表示。RoI pooling的具体实现可以看做是针对RoI区域的普通feature map的max pooling。  **2）RoI warping层**  为了使得梯度可以反向传播到坐标上，在RPN之后对region proposal的位置做二次调整，需要在RoI pooling层之前添加可以求导的RoI warping层来将残差传递给bounding box。对于每个RoI，其feature map都会被crop和warp到固定大小，转换方法如下：  （8）  其中，和分别表示warp之前和之后feature map的元素位置。表示长度和宽度两个方向上的双线性插值函数乘积。  于是，根据链式法则，bounding box的参数导数计算如下：  （9）  **2. 权重和偏置的初始化**  1）权重的初始化  由于权重初始值要非常接近0，但又不能等于0，所以需要将权重初始化为很小的数值，以此来打破对称性（如果神经元刚开始的时候是随机且不相等的，那么它们将计算出不同的更新，并将自身变成整个网络的不同部分）。如果选择小随机数权重初始化，随着输入数据量的增长，随机初始化的神经元的输出数据分布中的方差也在增大，所以用除以输入数据量的平方根的方式来调整其数值范围，这样神经元输出的方差就归一化到1了。  因此，本课题将采用随机数初始化结合使用校准方差的方式进行权重的初始化。其实现方法为。其中n是输入数据的数量，randn函数是基于零均值和标准差的一个高斯分布来生成随机数的。根据这个式子，每个神经元的权重向量都被初始化为一个随机向量，而这些随机向量又近似服从一个多变量高斯分布，于是在输入空间中，所有神经元的指向是随机的。  2）偏置的初始化  本课题选择将偏置初始化为0，因为随机小数值权重矩阵已经打破了对称性。  **3. 超参数的调优**  越大的神经网络需要越长的时间去训练，所以调参可能需要几天甚至几周。在调试超参数训练神经网络得到最优模型时，设计如下：用仆程序持续地随机设置参数并进行最优化。在训练过程中，仆程序会对每个周期后验证集的准确率进行监控，然后向文件系统写下一个模型的记录点（记录点中有各种各样的训练统计数据，比如不同时间的损失值变化等）。这个文件系统是可共享的，文件名中包含验证集的算法表现，以方便地查找和排序。另外还有一个主程序，它可以启动或者结束计算集群中的仆程序，也可以根据条件查看仆程序写下的记录点，输出它们的训练统计数据等。  对于超参数的选择范围，本课题将选择在对数尺度上进行超参数搜索。例如，超参数学习率被设置为，表示从标准分布中随机生成了一个数字，然后让它成为10的阶数。对于正则化强度，可以采用同样的策略，这是因为学习率和正则化强度都对于训练的动态进程有乘的效果。因此，比起加上或者减少某些值，思考学习率的范围是乘以或者除以某些值更加自然。但是有一些超参数还是需要在原始尺度上进行搜索。例如，超参数随机失活被设置为。  在进行超参数搜索时，采取从粗到细的分阶段搜索。首先进行初略范围搜索（比如10 \*\* [-6, 1]），然后根据出现较好结果时的位置，缩小范围进行搜索。进行粗搜索的时候，让模型训练一个周期就可以了，因为很多超参数的设定会让模型没法学习，或者突然就爆出很大的损失值。第二个阶段就是对一个更小的范围进行搜索，这时可以让模型运行5个周期。最后一个阶段就在最终的范围内进行仔细搜索，运行很多次周期。 |

5．工作方法

|  |
| --- |
| 1）选择科学的工作方法，制订完整的技术路线  **研究方法：**  1）文献研究法：阅读国内外先进的文献，包括CVPR、ICCV、NIPS等顶级期刊、会议上发表的具有创新意义的文章。了解深度学习技术前沿，对现有方法进行分析和归类。研究各类家居算法的优势、瓶颈和发展趋势，进行归纳与分析，针对实验室已有的虚实融合的家居DIY算法提出优化方案。  2）实验研究法：对提出的方案进行编程实现，验证所提方案并对实现结果进行归纳与分析。对没达到预期结果的部分进行改进，并再次通过编程实现验证方案的可行性。  **技术路线：**  本课题总体技术路线如下图所示：    图8 总体技术路线图  共分为以下两个阶段：  **1）图像标注和图像分割网络训练**  MS COCO是一种包含图像分割和标注的数据集。本课题将选择它作为数据集输入至网络模型中进行训练。对于该数据集，将随机选取70%的数据集用于训练，剩余的30%作为验证集。  由于本课题针对室内图像进行研究，因此，在把数据集输入至网络模型之前需要对其做预处理。首先，明确和家具相关的物体类别；然后，遍历MS COCO中的标注文件（格式为JSON），删除和家具无关的标注信息，即将室外图像和室内图像中和家具无关的物体当成负样本。另外，在之后的训练过程中，可以进行迭代训练以提高家具识别的准确率。  在网络模型训练过程中，主要有以下三个步骤：  **a）模型搭建**  首先，本课题将选择深度学习框架torch作为平台进行模型搭建，因为该框架同时支持对CNN和RNN的训练。然后，根据图像分割和图像标注的网络模型结构设置各层的排列方式，并且定义filter的数量和尺寸以及pooling的方式。最后进行参数初始化以及定义超参数、损失函数和优化方式。这里，超参数包括mini-batch的大小、学习率的大小、正则化的程度以及dropout的比例等。  **b）模型检查**  为了确保模型的正确合理，需要对搭建的模型进行一些检查，检查的内容包括：   * 梯度检查   把解析梯度和数值计算梯度进行比较。本课题将使用中心化公式近似表示数值计算梯度：  （10）   * 合理性检查   （1）寻找特定情况的正确损失值：在使用小参数进行初始化时，确保得到的损失值与期望一致，最好先单独检查数据损失（使正则化强度为0）。  （2）对小数据子集过拟合：在整个数据集进行训练之前，尝试在一个很小的数据集上进行训练（比如20个数据），然后确保能到达0的损失值。在这个过程中，最好让正则化强度为0，不然它会阻止得到0的损失。除非能通过这个正常性检查，不然进行整个数据集训练是没有意义的。  **c）模型跟踪**  在超参数调优时，需要跟踪一些重要的数值并绘制图表，从而知道应该如何修改超参数以获得更高效的学习过程。最后选取表现较好的超参数。这里需要跟踪的数值包括：   * 损失值   绘制损失值随时间变化的图表，曲线形状会给出关于学习率的设置情况。过低的学习率导致算法的改善是线性的；高一些的学习率会看起来呈几何指数下降；更高的学习率会让损失值很快下降，但是接着就停在一个不好的损失值上。   * 训练集和验证集的准确率   绘制训练集和验证集的准确率随时间变化的图表，训练集准确率和验证集准确率中间的空隙指明了模型过拟合的程度。一种可能是，相较于训练集，验证集的准确率低了很多，说明模型有很强的过拟合，遇到这种情况，就应该增大正则化强度（更强的L2权重惩罚，更多的随机失活等）或收集更多的数据；另一种可能是验证集曲线和训练集曲线如影随形，说明模型容量还不够大，应该增加参数数量。  将该超参数定义下的已经训练好的模型进行保存。  **2）家具多标签标注和箱体重建**  将室内图像分别输入至已训练好的模型和已有的灭点检测算法中，依次得到物体标注的结果、物体的像素级坐标信息以及相机的参数信息。结合物体的坐标信息和相机参数信息，通过坐标转换，得到物体在世界坐标系下的空间坐标。最终完成室内图像中家具的多标签标注和三维箱体重建。  2）工作方案的可行性分析，预设工作中可能遇到的难点，提出解决的方法  **人员的可行性：**  本课题有专门的研究团队进行相关研究，其中包括多名老师、1名博士研究生、3名硕士研究生。此外，实验室与微软亚洲研究院的专家有着密切的合作关系。  **方案的可行性：**  （1）课题组已经在前期进行了很多相关工作的研究，尤其是在图像灭点检测方面得到了很好的效果，为本课题的研究提供了较好的支持。  （2）本人前期研究了深度学习的基本内容，并在计算机视觉、自然语言处理等方面实现了一些应用示例，为本课题的研究打下了基础。  （3）目前，深度学习在图像标注、图像分割等领域有着很好的发展，已有较为成熟的算法。本课题将在此基础上进行研究并应用于家居虚拟展示系统中。 |

6．预期成果和结论

|  |
| --- |
| 对研究问题的解答进行科学预设，提出预期的工作成果  **预期的工作成果：**  （1）针对室内家居图像，实现家具的多标签标注；  WechatIMG3.png  图9 家具多标签标注效果图  （2）针对室内家居图像，进行图像分割；  （3）根据图像的相机信息和家具的位置信息，对家具进行箱体重建；    图10 家具箱体重建效果图  （4）整理研究成果，发表1-2篇高水平论文；  （5）申请一项专利。 |

7．工作基础

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1）科学评估研究条件和实验条件以及自己的专业基础和导师的专业把控能力  **研究和实验条件：**  （1）具有最新可编程显卡的通用PC一台  （2）高性能图形工作站：  名称:超微7046GT-TRF-TC4  CPU型号:Xeon E5620 2.4GHZ主频；  内存:16GB ECC DDR3；  标配硬盘容量:4TB；  网络控制器:Intel 82574L双千兆网卡。  （3）OpenCV、Torch开发文档和相关参考文献  （4）室内家居图片数据集  （5）以基于虚实融合的3D家居DIY算法为核心的技术研究  **自己的专业基础：**  在研究生一年里上过高级图形学、高级数字图像处理等相关课程，并对三维家居展示、机器学习和深度学习等内容进行过有针对性的研究。具有一定的专业基础和编程能力，适合进行图形图像相关工作。  **导师的专业把控能力：**  本人导师为贾金原教授，贾老师在虚拟现实、图形图像等研究领域钻研多年，有着深厚的研究基础，同时有多年的项目实战经验，目前已有多篇高水平论文发表，具备很好地专业把控能力。  2）所需经费，经费来源，开支预算（工程设备、材料须填写名称、规格、数量）  **本课题所需经费来源于以下项目支持：**  （1）国家自然科学基金面上项目(No.61272276)；  （2）国家自然科学基金青年项目(No.61305091)；  （3）国家十二五计划重大科技支撑项目(No. 2012BAC11B00-04-03)；  （4）高等学校博士学科点专项科研基金项目（博导类）(No.20130072110035)；  （5）吉林省重点科技攻关课题(No.20140204088GX)；  （6）上海市重点科技攻关课题(No.08511501000)；  （7）中央高校基本科研业务费专项资金(No.2100219054)。  **经费的开支预算情况如下表所示：**   |  |  |  | | --- | --- | --- | | 开支项目 | 内容 | 费用（元） | | 参考资料 | 相关书籍 | 500 | | 耗材 | A4纸张 | 100 | |

8．工作计划（含实验、实践、设计、写作）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **阶段及内容** | **工作量估计**  **（时数）** | | **起讫日期** | **阶段成果形式** |
| 1  2  3  4  5  6  7  8 | 文献研读  学习Lua和Python，熟悉深度学习框架，配置开发环境，处理数据集  搭建并训练深度学习神经网络，实现图像标注  搭建并训练深度学习神经网络，实现图像分割  实现物体箱体重建  整合前期工作，撰写小论文  后续工作，继续优化多标签标注和箱体重建的效果  硕士研究生论文撰写及修改 | 100小时  150小时  400小时  400小时  200小时  200小时  100小时  250小时 | | 16.10.15-16.11.30  16.12.01-16.12.30  17.01.01-17.02.15  17.02.16-17.03.31  17.04.01-17.04.30  17.05.01-17.05.31  17.06.01-17.06.30  17.07.01-17.09.30 | 研究报告  程序  算法设计及实现，程序  算法设计及实现，程序  算法设计及实现，程序  小论文  算法设计及实现，程序  论文 |
| **合计** | 1800小时 |

9. 同济大学硕士研究生学位论文选题报告评分表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 评审项目 | 权重 | 评 分 标 准 | | 得分（百分制） |
| A.研究问题 | 30% | 80~100分 | 研究问题明确，具有前沿性，反映出明确的应用研究中的科学内涵，具有较强的应用价值。 |  |
| 60~80分 | 研究问题明确，反映出应用研究中的科学内涵，具有应用价值。 |
| 60分以下 | 研究问题不明，缺乏理论意义和实用价值。 |
| B.理论基础和专门知识 | 20% | 80~100分 | 较好地掌握并能综合运用科学理论、方法和技术手段解决研究问题。 |  |
| 60~80分 | 基本掌握理论基础和系统专业知识。 |
| 60分以下 | 缺乏理论基础和系统知识。 |
| C.工作成果 | 20% | 80~100分 | 成果科学、完整、扎实，体现出明确的学术思想。 |  |
| 60~80分 | 研究成果完整，体现出一定的学术思想。 |
| 60分以下 | 研究成果不完整，未体现学术观点。 |
| D.工作方法 | 20% | 80~100分 | 工作方法科学合理，条理清楚，针对性强。 |  |
| 60~80分 | 制定了明确的工作方案，有针对性。 |
| 60分以下 | 研究方法不明确。 |
| E.报告报告 | 10% | 80~100分 | 报告严密、逻辑性强、文笔流畅，表达清楚。 |  |
| 60~80分 | 基本概念清晰、层次分明，表达较清楚。 |
| 60分以下 | 写作、表达较差。 |
| 总分 |  | 总分=0.3A+0.2B+0.2C+0.2D+0.1E | |  |

备注：评审专家只对五项指标每一项的最后一栏内打分（百分制），不必计算总分。

评审小组组成：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 组成 | 姓名 | 职称 | 单位 | 签字 |
| 导师 | 贾金原 | 教授 | 同济大学 |  |
| 成员 | 杜庆峰 | 教授 | 同济大学 |  |
| 张苗苗 | 教授 | 同济大学 |  |
| 张惠娟 | 副教授 | 同济大学 |  |
|  |  |  |  |

注：此评分表作为硕士研究生课程成绩单必备的材料之一

年 月 日

10.评审意见

|  |
| --- |
| 导师（或导师组）对本课题的评价  本课题将先进的深度学习技术首次应用于虚实融合的家居在线设计中，对室内图像中的家具进行多标签标注，并进行图像分割来实现家具的箱体重建，能够有效地提升家居照片中的家具识别上的精确度，为家居在线可视化电子商务提供关键技术支撑。该报告所拟定的技术路线基本可行，具有一定的创新性，研究方案合理，也具有潜在的实际商务价值。该课题理论联系实际，十分符合工程硕士论文的开题要求。同意开题。  导师签名  年 月 日 |
| 评审小组的审查结论  组长签名 组员签名  年 月 日 年 月 日 |
| 学科专业委员会意见  负责人签名  年 月 日 |