

尊敬的各位老师：

大家下午好。

我是答辩学生唐麒。我的毕设题目是联合深度估计的深度图像超分辨率重建算法研究。我的指导教师是冯凤娟老师。我的答辩将从以下四个部分展开。

第一部分 项目简介

我的毕设来源于目前在北京交通大学信息科学研究所的实习。下面我将对毕设内容进行介绍。

便携式消费级深度相机的普及为主动获取场景深度提供了便利，但由于成像技术的限制，拍摄获得的深度图像分辨率普遍较低。面对自动驾驶等领域对高质量深度图像的需求，提高深度图像分辨率成为一个亟待解决的问题。深度图像超分辨率重建技术是在硬件条件不变的前提下，利用算法恢复图像的高频信息，并有效抑制随机噪声等，重建出高质量的高分辨率深度图像。近年来，学习式的超分辨率重建技术取得了巨大进展。

深度图像超分辨率重建的研究不仅具有重要的理论意义，在无人驾驶、三维重建等实际应用中也有着迫切的需求。

第二部分 研究现状

高分辨率彩色图像包含低分辨率深度图像缺失的高频信息，因此可以为深度图像的超分辨率重建提供先验信息。但彩色图像与深度图像的边缘并非完全对应，如图中绿色矩形，彩色图像具有较多的纹理变化，而深度图像并没有对应的结构，在重建中会造成纹理复制。单目深度估计就是要找到一个场景从彩色图像到深度信息的映射模型。近年来，研究者通过设计良好的网络从数据中自动地学习恢复深度图像并取得了很大进展。同时，深度学习在深度估计上也取得了良好的性能。本项目旨在建立教师-学生模型，通过教师网络在深度估计任务上学习彩色图像与深度图像的映射关系，然后利用教师模型学习到的“知识”帮助学生网络进行深度图像的超分辨率重建，从而达到更好的重建效果。

受上述分析的驱动，我主要对学习式超分辨率重建算法、深度学习下的单目深度估计和面向视觉智能的教师-学生网络的研究现状进行了调研。

彩色图像引导的深度图像超分辨率重建在获取丰富的层次特征后，利用对应尺度下彩色图像的信息，引导深度图像重建。现有研究的网络可分为金字塔结构和 U-Net 结构。两种结构串联地从低尺度到高尺度重建高分辨率深度图像，均可提取丰富的层次特征。

Eigen 团队利用 CNN 对单目图像进行深度估计，网络分为全局深度的粗估计和局部优化的精估计两步。这一方法开创了深度学习在单目深度估计领域的先河。此后，研究者通过设计不同的网络结构、损失函数等对单目深度估计进行优化。

知识蒸馏被用来将模型学习到的知识迁移到另一个模型，该过程也被描述为教师-学生框架。较小的学生模型被训练来模仿预训练的较大的模型。Hong 等在图像重建和去雾两个任务探究了教师通过不同任务帮助学生网络学习的可行性及方法。

批注 [唐1]: (1) 在游戏领域，通过体感周边设备来获取玩家的姿态动作，进行人机交互操作，如果用超分辨率重建技术增强深度数据的分辨率和质量，有利于提高姿态识别率，进而大大提升玩家的游戏体验。

(2) 在三维重建方面，对于深度相机获取的点云数据，利用超分辨率重建技术提高它的密集度和精度，从而建模出更具真实感的三维物体表面模型，促进三维重建在生物医学、视频监控、刑事案件分析、增强现实等领域的应用。

(3) 在无人驾驶方面，利用深度信息来确定车辆的位置，深度图像超分辨率重建能帮助车辆获得更高的定位精度，同时进行更准确的环境描述和避障操作。

批注 [唐2]: Hui 等构建了金字塔结构，输入的低分辨率深度图像经逐级采样后得到多尺度的特征表示，再融合对应尺度的纹理特征；Guo 等采用 U-Net 结构对插值后的深度图像进行编码，在解码过程中融合对应尺度的纹理特征。

批注 [唐3]: Eigen 将深度估计、表面法线预测和语义标注三个任务统一在一个三级的神经网络中，并将结果的分辨率提升至输入图像分辨率的一半。Grigorev 等将长短期记忆 (LSTM) 用于循环网络以获取图像全局信息，将其与一般的卷积神经网络混合使用实现了端到端的单目图像深度估计。Liu 等把 CNN 与 CRF 统一于一个框架内，将两个卷积神经网络分别对应能量函数中包含超像素内深度信息的项和关于相邻超像素关系的项，计算其最大化后验概率。Laina 等采用了更深的残差网络并使用小卷积代替大卷积来实现上采样，使得深度估计更为高效，并且提出了新颖的损失函数，从而可以得到更好的结果。Cao 等将原本连续的图像深度离散化为一定深度范围的类别，并将深度估计问题转化为分类问题，使用全卷积的深度残差网络实现分类，最后使用条件随机场优化结果得到最终的深度估计值。

第三部分 研究内容

课题的研究基于多尺度特征、注意力机制等理论基础。**（知识蒸馏）**
低分辨率图像的捕获可视为图像的降质模型，其对场景中物体的影响是不同的，尤其是结构微小的物体。这就需要多尺度信息来增强特征表示，以精确恢复每个深度区域。
注意力机制可以帮助模型对输入的每个部分赋予权重，抽取出更加重要的信息，使模型做出更加准确的判断。

在研究中还存在如下问题：
当彩色图像引导深度图像重建时，违背共生关系的区域会出现纹理转移或深度流失两种失真。设计特定的网络结构来解决这些问题，是当前需要解决的关键问题。
其次，我们希望将深度估计学习到的“知识”用于低分辨率深度图像的超分辨率重建。目前上没有针对这一问题的研究，探究和明确两个任务的关联性是项目研究的**关键任务**。

综上，本项目以深度图像超分辨率重建为基本任务，联合深度估计，以期两个任务互相促进，从而获得更加优异的重建精度。主要包括：
模型设计：设计约束关系，如构建教师-学生模型等将单目深度估计与深度图像超分辨率重建联合起来。
数据准备：深度图像超分辨率重建的 RGB-D 数据集有 MIDDLEBURY 等，单目深度估计有 KITTI、NYU Depth 等。
实验评估：与其他深度图像超分辨率重建模型的性能进行对比。
分析总结：通过消融研究验证不同模块对建精度、抑制噪声等不同性能的影响，进行分析和总结。

项目预期成果为设计和编码神经网络对深度图像超分辨率重建,通过深度估计提高重建精度。

第四部分 研究计划

至开题答辩**已完成**文献调研和论文复现，**当前**工作为通过由弱到强的约束将两个任务联合，从而完成模型设计，并计划于 3 月 31 日前完成模型的编码、训练和调优。然后进行性能对比和论文**撰写**。
目前共计阅读相关论文 40 余篇，以下为前期研究的主要**参考文献**。
我的答辩到此结束，感谢各位的聆听，希望老师们批评指正。

批注 [唐4]: 截止到 2020 年 12 月 13 日，Hinton 等人的 KD 已经被引用 4955 次。在视觉领域，不少研究人员在传统 KD 的基础上进行了研究。Tian 等指出原始 KD 不适合跨模态蒸馏的情况，如将图像处理网络(处理 RGB 信息)的表示/特征迁移到深度处理网络(处理 depth 信息)。Mirzadeh S-I 等指出当教师与学生的模型大小差距太大时，KD 会失效，学生网络的性能会下降，因而在教师网络和学生网络之间引入了助教网络。Jin 等受课程学习启发，并且知道学生和老师之间的差距很大导致蒸馏失败，提出路由约束提示学习。Liu 等人的工作整合了暗知识的逐像素蒸馏、马尔科夫随机场的逐配对/特征块蒸馏和有条件生成对抗网络的整体蒸馏，用于密集预测任务：语义分割、深度估计和物体检测。在模型压缩中，由于模型蒸馏可以用于任意网络而被广泛研究，并应用于具体的视觉计算任务之中。

批注 [唐5]: 文献调研：阅读国内外相关任务主流算法和算法性能较高的论文，总结当前研究的进展、技术路线和发展趋势。**论文复现：**通过 Pytorch 实现论文中重要算法，直观理解其作用机理和实验效果。**模型设计：**根据理论分析和实验验证，设计和编码实现完整的网络架构进行深度图像的超分辨率重建，并对模型进行训练和调试。**对比分析：**将实验获得的模型指标与其他算法相比较，验证模型在性能提升上的有效性和显著性。**项目的验收以及论文的撰写等。**