

尊敬的各位老师

大家好。我是答辩学生唐麒。我的毕设题目是联合单目深度估计的深度图像超分辨率重建算法研究，指导老师是冯凤娟老师。我将从以下五个方面进行展示。

### 第一部分 项目简介

我的毕设项目来源于在北京交通大学信息科学研究所的实习，岗位是算法研发。下面我将对具体内容进行展示。

消费级深度相机的普及为获取深度图像提供了便利。但由于成像能力的限制，深度图像的分辨率通常较低。面对诸多应用领域对高质量深度图像的需求，深度图像超分辨率重建技术获得了广泛关注。深度图像超分辨率重建是在深度相机不变的前提下，通过算法重建出高质量、高分辨率的深度图像。随着深度学习的快速发展，该任务的性能得到了很大提升。

在获得高分辨率深度信息后，可以对许多任务起到促进作用，如虚拟现实、三维重建和自动驾驶。

### 第二部分 相关工作

在实际应用中，高分辨率彩色图像易于获得，且与深度图像具有结构相似性，因而可以为深度图像超分辨率重建提供先验信息。现有的颜色指导的深度图像超分辨率重建算法通常需要一个分支来从彩色图像获取信息，然后用于指导重建分支。但彩色图像与深度图像的边缘并非一一对应，这种不一致性可能会造成纹理复制等问题。为了寻找解决方案，我们将目光聚焦在单目深度估计。单目深度估计实现了彩色图像到深度图像的跨模态转换，因而面向单目深度估计学习到的彩色特征更适合指导深度图像超分辨率重建。在网络设计方面，为了替换现有方法中的颜色分支，拟采用多任务学习的方式进行不同任务间的信息传递。且单目深度估计和深度图像超分辨率重建联合学习无需引入额外的监督信息。

### 第三部分 研究内容

项目方案采用卷积神经网络，兼以多尺度、注意力机制等技术。

基于理论分析和实验验证，设计了联合学习网络 BridgeNet，该网络具有高度的可移植性，可以为关联深度图像超分辨率重建和单目深度估计提供范例。在探讨任务相关性后，提出了用于联合学习的高频注意力桥和内容引导桥。在不引入监督信息的前提下，我们的方法在多个基准数据集上达到了具有竞争力的性能。

网络由深度图像超分辨率重建子网络和单目深度估计子网络及两个桥接器组成，并以高分辨率彩色图像和插值到相同分辨率的深度图像作为输入，重建的深度图像为主要输出。

受编码器-解码器在单目深度估计任务取得成功的驱动，本文的单目深度估计子网络也遵循了这一结构，包括特征提取器，特征金字塔网络和深度预测模块三部分。特征提取器从彩色图像提取多级特征，并通过横向连接送到特征金字塔网络。特征金字塔网络将语义信息从高层特征传播到低层特征，进而生成优化的多级特征。在深度预测模块中，跨层恒等特征融合模块用于逐步融合优化的多层特征并完成深度估计。CLIFF 模块以插值的高层特征和低层特征作为输入，由于高层特征噪声较少，故通过相乘的方式来对低层特征进行强化。使得低层特征中的准确响应得到增强，而噪声受到抑制。为了得到高层特征、原始的和强化后的低层特征的最佳组合，CLIFF 模块通过两个卷积层来对三个特征进行选择。

深度图像超分辨率重建子网络同样采用了编码器-解码器结构。编码器由一个浅层特征提取器和三个特征编码模块及堆叠的残差块构成。特征编码模块中的最大池化层可以确保两个子网络编码器对应层特征的分辨率相同。较深的网络可以帮助深度图像恢复出更分明的边缘和形状，故将编码器提取的特征传递到由  $N$  个残差块组成的特征提取模块中。而为了恢复深度图像中的精细结构和微小物体，引入了多尺度策略，通过进一步融合编码器中间层的特征来优化高层特征。解码器使用了三个相同的上采样模块，特征图每经过一个上采样模块，分辨率都变为之前的两倍。编码器提取的浅层特征和深层特征通过一个跳连接进行融合。这样做是为了直接提供重建需要的低频信息，从而迫使网络更加专注学习重建需要的高频信息。此外，特征图在经过编码器中三个最大池化后，分辨率下降为输入的  $1/8$ 。浅层特征将被送到下采样块以保证跳连接两侧的特征拥有相同的分辨率。

在联合学习视角下，单目深度估计为解决 RGB-D 图像对结构不一致问题提供了新的思路。由单目深度

估计编码器提供的彩色特征更接近深度模态，可以避免明显的伪影。因此，提出在编码器使用单目深度估计代替现有方法的颜色分支。随着单目深度估计子网络的深入，特征的分辨率逐渐降低，其中高层特征具有丰富的语义信息，而低层特征则具有更多的结构信息。由于低分辨率图像缺少高频信息，因此超分辨率重建更需要彩色图像提供的是高频信息。基于此，设计了高频注意力桥。

具体来说，首先使用平均池化和反卷积对单目深度估计子网络的特征进行模糊，然后，将原始特征与模糊特征相减来获得高频信息，进而生成高频信息的注意力，接着用高频注意力对原始特征进行优化，通过残差连接最终得到引导特征。为了更好地利用来自单目深度估计子网络的指导特征，首先将它们与深度图像超分辨率重建子网络的特征级联，这种融合在空间和通道维度上会有很多冗余，因此引入了一个包括通道注意力和空间注意力的注意力块来增强特征融合能力。融合了高频信息的特征将作为深度图像超分辨率重建子网络编码器下一层的输入。

单目深度估计由于尺度模糊性而被认知为不适定的逆问题。尽管超分辨率重建也是一个不适定问题，但它仍在相同的域中学习映射关系，故相对单目深度估计而言更简单。由于两个任务性能之间存在较大的差距，单目深度估计不再适合为深度图像超分辨率重建提供指导信息。遵循简单任务指导困难任务的原则设计了内容引导桥。

具体来说，分别对两个子网络采用  $1 \times 1$  卷积以获得重建的深度图像和估计的深度图像，然后计算二者的差异图。差异图显示了估计的深度图像相对于超分辨重建的深度图像需要优化的位置。在此基础上，应用卷积和 softmax 来获得差异权重，从而为单目深度估计子网络提供内容引导，使得差异在学习过程中越来越小。最后，使用注意力块对特征进行优化，优化后的特征将作为单目深度估计子网络解码器下一层的高层特征输入。

考虑到两个任务的性能差距，很难设置损失函数的权重使得两个任务都达到最佳性能。因而提出分别对两个任务的相关部分进行优化的策略。

#### **第四部分，研究成果**

我们分别在 MIDDLEBURY 数据集和 NYU DEPTH 数据集上进行了训练和测试，并选择 MAD 和 RMSE 两个指标在不同的上采样因子与一些方法进行了比较。以\*16的上采样因子为例，对于 Books 图像，与排名第二的算法相比，MAD 从 0.67 提升至 0.51，百分比增益为 23.9%。此外，与其他方法相比，我们的方法可以恢复深度图像更多的细节。例如，Art 图像中棍子周围的伪影更少，Doll 图像中玩具头部轮廓更准确。

在 NYU 数据集上，我们的方法同样表现出具有竞争力的结果。

此外，我们还设计了一些消融实验，验证了设计模块对于提升模型性能的有效性。

作为第一作者，我们的研究已投稿至多媒体领域的顶会 2021 ACM MM，为 CCF A 类。

#### **第五部分 项目总结**

综上，项目探索了一种具有高度可移植性的联合学习框架，可以为关联深度图像超分辨率重建和单目深度估计提供范例。项目的核心思想是如何设计两个任务之间的交互，由此提出了高频注意力桥和内容引导桥。

在项目的研究过程中，整体执行情况良好，前期阅读了数十篇英文论文，并借助周报进行总结梳理，为模型设计奠定了基础。然后通过实验确定了模型方案，最终完成了论文的撰写。

以下为主要参考文献。

以上就是我展示的全部内容，感谢各位老师的聆听，敬请批评指正。