

# 工作总结与计划

## 一、工作总结（按时间划分）

日期	工作内容			完成情况及主要问题
	上午	下午	晚上	
星期一	处理杂事	回顾论文	读论文	回顾之前阅读的论文部分 继续阅读未完成的部分
星期二	读论文	开会	读论文	对论文中不明白的地方反复阅读 大体了解了论文的思想
星期三	读论文	读论文	形教课	基本完成论文的阅读，同时对两篇论文的相似之处有了新的理解
星期四	处理杂事	回顾上一篇论文	休息	通过再次阅读两篇论文，对阅读中似是而非的部分有了新的理解
星期五	处理杂事	读论文	回顾论文	对之前的阅读和思考进行思考和总结 思考周报内容
星期六	写周报	写周报	写周报	完成周报

## 二、下一步计划（按任务划分）

编号	工作内容	目标	相关配合
1	在毕设系统中完成毕设申报	11 月 16 日完成	无
2	继续读有关论文	11 月 20 日完成	获取新的论文
3			
4			
5			
主要风险	无		

## 三、个人分析与总结

内容提要	
1	进度方面：本周整体进度按照预期计划进行，在此基础上可适当提升效率
2	课题方面：通过阅读论文，对上一篇论文有了新的认识，也对一些方法有了了解
3	其他思考：相较于阅读第一篇论文时存在的阅读方法问题，通过对论文的回顾和反复阅读，有了改善
4	

四、论文总结

论文标题	Depth Super-Resolution via Deep Controllable Slicing Network
作者及单位	Xinchen Ye, Baoli Sun, Zhihui Wang, Jingyu Yang, Rui Xu, Haojie Li, Baopu Li
论文出处	2020-ACM
创新点提炼	<p>本文主要贡献可总结如下：</p> <p>1） 一种端到端的深度可控切片网络，可实现区域级深度恢复并具有较高的 DSR 任务泛化能力；</p> <p>2） 尺度可控模块（scale-controllable module, SCM），在一个统一的模型中以任意放大倍数实现深度恢复的细粒度控制；</p> <p>3） 深度切片模块（depth slicing module, DSM），区别对待地使用具有不同深度范围的深度图特征以分而治之的方式超分辨率深度图；</p> <p>本文的方法基于单一深度图 SR，无需颜色信息，但在合成数据集和实际数据集上，其性能优于其他颜色指导的 DSR 方法。此外，在所有尺度上一起训练的模型也优于在每个尺度上独立训练的最新的模型。</p>
个人想法	<p>本文以集成策略为主体思路，通过多个网络分支对深度图像的不同深度区域特征进行提取和超分辨率重建，仅通过单一深度图就可完成较高性能的 DSR 任务，同时其提出的其他模块和训练策略也对实际应用中的 DSR 任务有一定的指导意义。</p>

论文方法及结论：

1 提出的问题

先前的方法均等地对待所有深度区域，而不考虑区域级别的不同退化程度，并且将不同尺度下的 DSR 视为独立任务，而没有考虑不同尺度的建模，这阻碍了 DSR 的进一步性能改进和实际使用。

- 场景深度恢复取决于场景特征，即，近景和微小物体的深度区域倾向于比远景物体和背景更严重地受到下采样降级的破坏；
- 大多数现有的 DSR 方法在一个深度图中通过相同模型会有不平衡的估计，可能会导致不同深度区域的深度恢复不准确。此外，由深度传感器捕获的深度图通常会受到距离相关的高斯噪声的污染，但以前的研究主要集中在处理深度图的所有区域的单个模型，而没有考虑上述在区域级别的复杂退化；
- 大多数现有算法将不同比例因子（×2，×4，×8，×16）的 DSR 任务视为独立问题，并且需要许多明确尺度的网络，这些网络需要进行独立训练才能处理各种规模。但在实际应用中，给定场景真正想要的上采样因子不是整数甚至是未知的；
- 此外，还原后的深度图的感知质量是相对主观的，因此模型必须根据图像特征细粒度地控制深度还原等。

2 解决的方法

为了缓解这些问题，本文提出了一种深度可控的切片网络（deep controllable slicing network），该模型将以分而治之的方式学习一组切片分支，并通过距离感知加权模式（distance-aware weighting scheme）进行参数化，以自适应地集成不同深度。指定深度切片（某个深度范围内的区域）的每个分支往往会准确地恢复深度。同时，提出了一种尺度可控模块（scale-controllable module），该模块可提取不同比例下的深度特征，并将其插入到切片网络前，从而可以对带有尺度超参数的切片网络的深度恢复结果进行精细控制。

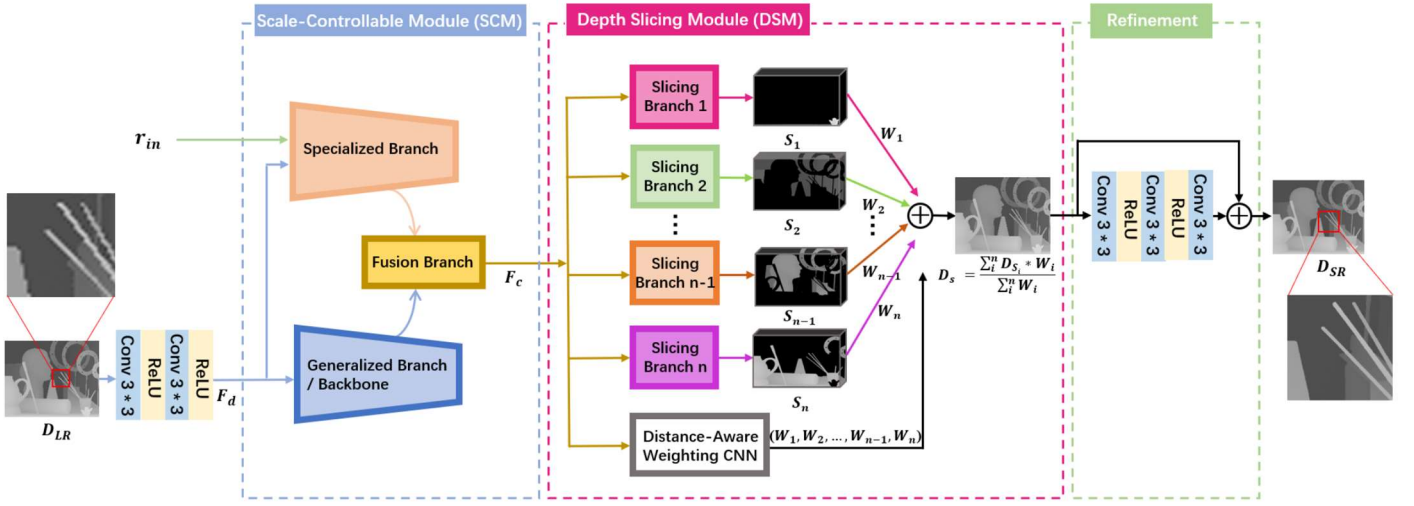


图 1 网络结构

令  $D_{LR}$  和  $\gamma_{in}$  分别为 LR 深度图（内插到所需的输出大小）和放大系数作为输入。

该网络可以分为三个部分：

- 1) **尺度可控制模块 (scale-controllable module, SCM)** : SCM 由三个分支组成, 即 generalized branch (GB, our backbone), specialized branch (SB), and fusion branch (FB)。从  $D_{LR}$  提取的浅层特征  $F_d$  被送入 GB 和 SB 中,  $\gamma_{in}$  被另外喂入 SB 以表示所需的尺度。FB 以多尺度方式聚合 SB 和 GB 的特征, 然后输出中间  $F_c$ 。
- 2) **深度切片模块 (depth slicing module, DSM)** : 在 DSM 中, 根据不同的深度范围将  $F_c$  在空间上划分为  $n$  个区域, 并将其送入相应的切片分支中。然后, 通过对切片分支的输出进行加权平均来获得中间深度图  $D_s$ 。
- 3) **最终优化 (final refinement)** : 通过简单的残差结构精炼  $D_s$ , 以输出最终的超分辨率深度图  $D_{SR}$ 。

## 2.1 Scale-Controllable Module

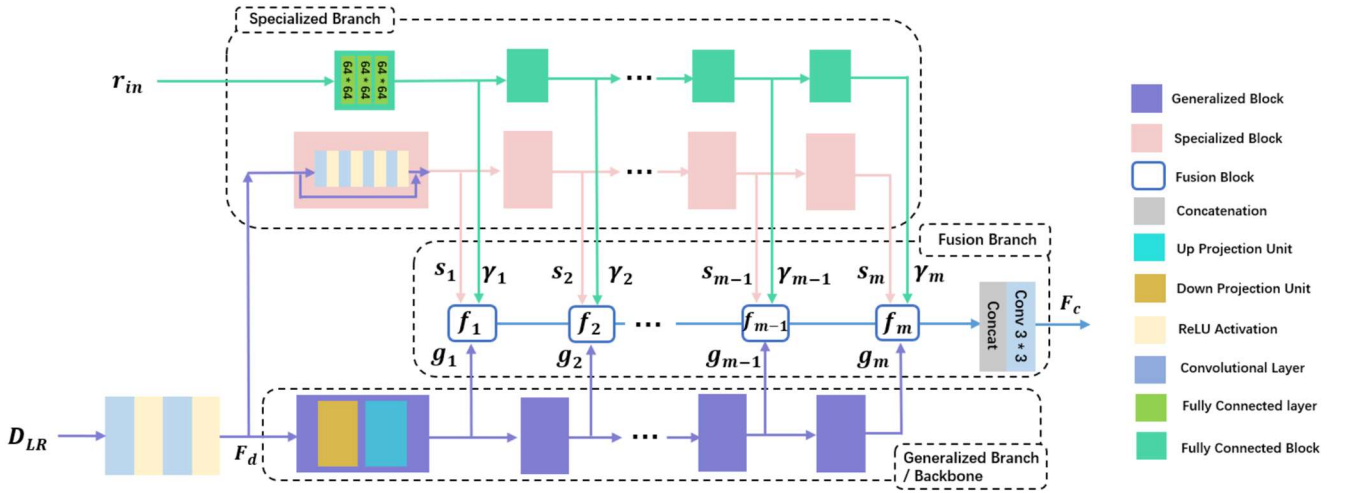


图 2 Scale-Controllable Module

GB 包含  $M$  个堆叠的 generalized blocks, 旨在提取共同特征 (common features); 而 SB 包含  $M$  个堆叠的 specialized blocks 以及  $M$  个连续的完全块连接块, 将给定的尺度参数及其相应的 LR 深度图作为输入, 然后将生成的包含不同尺度信息的特定特征与 GB 的特征通过 FB (包含  $M$  融合块) 以多尺度的方式汇总来自不同阶段的 generalized block, specialized block 和 fully connected block 的输出特征, 从而实现更丰富的深度表示。

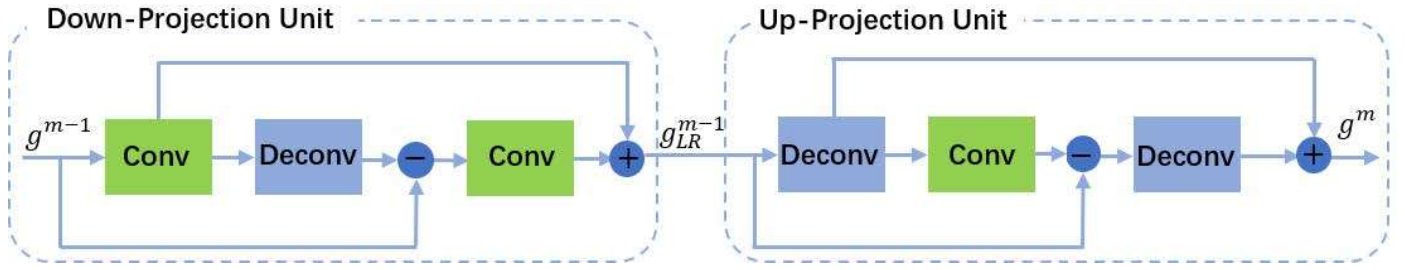


图 3 Down- and up-projection units

generalized block 基于 down-projection unit 和 an up-projection unit，通过迭代地将 HR 表示投影到 LR 空间域，然后将重构误差映射回 HR 域，可以有效地改善深度边界处的特征表示。specialized block 是一个简单的残差结构，由四个卷积层组成，每个卷积层均跟随一个 ReLU。

引入  $\gamma_{in}$  来明确控制深度特征的尺度，它可能进一步引导 generalized block 和 specialized block 的输出的自适应融合。具体来说，引入完全连接块以将输入尺度超参数  $\gamma_{in}$  映射到不同的向量中。对于 generalized block 和 specialized block 的多通道特征，每个向量具有不同的融合系数。

线性融合操作以在每个特征提取阶段选择性地处理输出特征

$$f^m = g^m + r^m \cdot s^m$$

最终，融合块的所有输出通过  $3 \times 3$  卷积进行连接和滤波，以输出特征图  $F_c$ 。

**训练策略：**为分别从 GB 和 SB 中获取通用特征和特定特征，通过基于不同优化目标的两阶段训练策略来学习整个网络。在第一阶段，将所有向量  $\{\gamma^m, m \in [1, M]\}$  固定为 1 全向量，并从 SB 中删除完全连接块，然后使用所有降采样输入将其余模块一起训练。该阶段旨在从各种降级的深度图中学习通用特征。在第二阶段，修复 GB 的参数，并将完全连接块添加回 SB，将尺度因子  $\gamma_{in}$  设置为与输入深度图的类型动态匹配的指定值，即  $\gamma_{in}$  可以设置为 1、2、3、4，对应于  $\times 2$ 、 $\times 4$ 、 $\times 8$  和  $\times 16$  LR 深度图。网络通过更新 SB 和 DSM 的参数来学习特定功能。在测试阶段，可以将尺度因子设置为任何值（从 1 到 4）以调整恢复性能，而无需事先知道放大因子的实际值，这对于在实际输入上测试的 DSR 任务非常实用。

## 2.2 Depth Slicing Module

DSM 由一组切片分支和一个距离感知加权 CNN 组成，其中每个分支都具有与 specialized block 相同的网络，即简单的残差块。每一个切片分支  $S_n$  都倾向于从共享的深度特征  $F_c$  的不同深度区域生成准确地深度估计  $D_{S_n}$ 。

为了实现准确的回归，本文提出了一种距离感知加权方案以学习融合权重，从而自适应地聚合集合中的所有切片分支。加权 CNN 的输出是一组加权掩码， $W_n$ （由 0 和 1 构成，其中 1 表示此像素属于相应的深度范围）。距离感知加权方案会根据不同的深度范围（ $D_{S_n}$  内的距离）自动识别多个深度区域，并协助每个切片分支恢复专注于其相应深度区域的深度值。

当融合切片分支时，中间深度图可能会在不同深度区域的交界处表现出一些误差。因此，在 DSM 模块的末尾添加了一个细化模块，以进一步增强中间深度图。

## 3 实验结果结论

- (1) **Depth SR under Noiseless Cases:** Our network almost obtains the best objective scores in all cases, especially for the  $\times 8$  and  $\times 16$  cases which are more difficult to restore. Our method can recover more pleasing structures and depth details.
- (2) **Depth SR under Noisy Cases:** Our method achieves the best objective performance. The results of other methods present excessive cotton-like blurring, while our method can remove the noise and keep the sharpest depth boundaries.

(3) **Evaluation on Generalization:** We obtain more accurate and clear depth details in the recovered results. Our model also achieves high perceptual quality which validates our high generalization ability due to the proposed scale-controllable module.

本文的方法基于单一深度图 SR，无需颜色信息，但在合成数据集和实际数据集上，其性能优于其他颜色指导的 DSR 方法。通过将 DSR 任务划分为应用于每个特定深度范围的多个子问题，本文的方法与单一 DSR 网络相比，它具有集成策略的精髓，并且得出的估计更可靠。

Table 1: Quantitative depth upsampling results on Middlebury dataset. (lower MADs, better performance)

	<i>Art</i>				<i>Books</i>				<i>Dolls</i>				<i>Laundry</i>				<i>Moebius</i>				<i>Reindeer</i>			
	×2	×4	×8	×16	×2	×4	×8	×16	×2	×4	×8	×16	×2	×4	×8	×16	×2	×4	×8	×16	×2	×4	×8	×16
Bicubic	0.48	0.97	1.85	3.59	0.13	0.29	0.59	1.15	0.20	0.36	0.66	1.18	0.28	0.54	1.04	1.95	0.13	0.30	0.59	1.13	0.30	0.55	0.99	1.88
FGI [26]	0.70	1.29	2.41	4.51	0.43	0.74	1.16	1.91	0.54	0.93	1.44	2.12	0.51	0.91	1.59	2.68	0.42	0.72	1.13	1.81	0.50	0.87	1.58	2.72
TGV [6]	0.45	0.65	1.17	2.30	0.18	0.27	0.42	0.82	0.21	0.33	0.70	2.20	0.31	0.55	1.22	3.37	0.18	0.29	0.49	0.90	0.32	0.49	1.03	3.05
DJF [25]	0.12	0.40	1.07	2.78	0.05	0.16	0.45	1.00	0.06	0.20	0.49	0.99	0.07	0.28	0.71	1.67	0.06	0.18	0.46	1.02	0.07	0.23	0.60	1.36
MSG [17]	-	0.46	0.76	1.53	-	0.15	0.41	0.76	-	0.25	0.51	0.87	-	0.30	0.46	1.12	-	0.21	0.43	0.76	-	0.31	0.52	0.99
DGDIE [12]	0.20	0.48	1.20	2.44	0.14	0.30	0.58	1.02	0.16	0.34	0.63	0.93	0.15	0.35	0.86	1.56	0.14	0.28	0.58	0.98	0.16	0.35	0.73	1.29
DEIN [41]	0.23	0.40	0.64	1.34	0.12	0.22	0.37	0.78	0.12	0.22	0.38	0.73	0.13	0.23	0.36	0.81	0.11	0.20	0.35	0.73	0.15	0.26	0.40	0.80
CCFN [40]	-	0.43	0.72	1.50	-	0.17	0.36	0.69	-	0.25	0.46	0.75	-	0.24	0.41	0.71	-	0.23	0.39	0.73	-	0.29	0.46	0.95
GSRT [3]	0.22	0.48	0.74	1.48	0.11	0.21	0.38	0.76	0.13	0.28	0.48	0.79	0.12	0.33	0.56	1.24	0.12	0.24	0.49	0.80	0.14	0.31	0.61	1.07
DSR_N [39]	0.12	0.25	0.61	1.80	0.04	0.11	0.28	0.69	0.06	0.14	0.33	0.73	0.06	0.15	0.43	1.24	0.05	0.13	0.29	0.67	0.07	0.15	0.35	0.92
Ours	0.10	0.23	0.58	1.30	0.06	0.09	0.26	0.63	0.07	0.11	0.31	0.69	0.06	0.14	0.34	0.77	0.06	0.12	0.27	0.64	0.07	0.14	0.33	0.79

Table 2: Quantitative depth upsampling results on real NYU v2 dataset.

Method	JBU [22]	DJF [25]	DGDIE [12]	GbFT [1]	PAC [37]	SVLRM [28]	DKN [20]	Ours
×4	4.07	3.54	1.56	3.35	2.39	1.74	1.62	<b>1.33</b>
×8	8.29	6.20	2.99	5.73	4.59	5.59	3.26	<b>2.87</b>
×16	13.35	10.21	5.24	9.01	8.09	7.23	6.51	<b>5.12</b>

Table 3: Quantitative depth upsampling results noisy Middlebury dataset.(lower MADs, better performance)

	<i>Art</i>		<i>Books</i>		<i>Dolls</i>		<i>Laundry</i>		<i>Moebius</i>		<i>Reindeer</i>	
	×8	×16	×8	×16	×8	×16	×8	×16	×8	×16	×8	×16
TGV [6]	2.76	6.87	1.49	2.74	1.75	3.71	1.89	4.16	1.72	3.99	1.75	4.40
MSG [17]	1.57	2.98	1.18	1.48	1.12	1.78	<b>1.03</b>	1.89	1.13	1.76	1.12	1.87
DGDIE [12]	1.84	3.34	1.29	2.04	1.39	2.05	1.73	2.67	1.37	2.16	1.33	2.19
DEIN [41]	2.44	4.24	1.44	2.38	1.55	2.45	1.77	3.20	1.64	3.29	1.46	3.87
GSRT [3]	1.33	2.47	<b>0.87</b>	1.37	1.26	<b>2.03</b>	1.24	1.86	<b>1.03</b>	1.68	1.04	<b>1.70</b>
DSR_N [39]	1.60	3.25	1.21	1.98	1.33	2.16	1.44	2.64	1.24	2.16	1.29	2.35
Ours	<b>1.20</b>	<b>2.13</b>	0.96	<b>1.33</b>	<b>1.11</b>	2.06	1.19	<b>1.82</b>	<b>1.03</b>	<b>1.61</b>	<b>0.91</b>	1.82

Table 4: Generalization on MPI Sintel datasets.

	<i>Ambush_2-15</i>		<i>Ambush_4-12</i>		<i>Ambush_5-41</i>		<i>Tumple_3-23</i>	
	×8	×16	×8	×16	×8	×16	×8	×16
MSG [17]	0.51	1.12	1.10	1.82	1.36	2.01	0.82	1.78
DGDIE [12]	0.65	1.24	1.26	2.23	1.79	3.10	1.01	1.90
DEIN [41]	0.47	1.08	1.12	1.76	1.69	2.32	0.89	1.82
GSRT [3]	0.62	1.44	1.32	2.45	1.98	3.46	1.19	2.07
DSR_N [39]	<b>0.30</b>	0.85	0.88	2.21	0.98	2.07	<b>0.59</b>	1.35
Ours	0.42	<b>0.64</b>	<b>0.82</b>	<b>1.62</b>	<b>0.77</b>	<b>1.88</b>	0.81	<b>1.30</b>

Table 5: Generalization on real ToFMark dataset.

Method	MSG [17]	DEIN[41]	DGDIE [12]	GSRT [3]	DSR_N [39]	Ours
<i>Books</i>	12.26	12.78	12.31	13.21	11.15	<b>11.03</b>
<i>Shark</i>	14.11	15.11	14.06	15.03	13.26	<b>12.08</b>
<i>Devil</i>	12.45	14.25	9.66	12.27	9.54	<b>9.33</b>

## 4 存在的问题

暂无