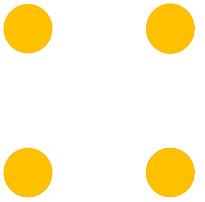


Super-Resolution

汇报人：谢孔伟

2024/7/19

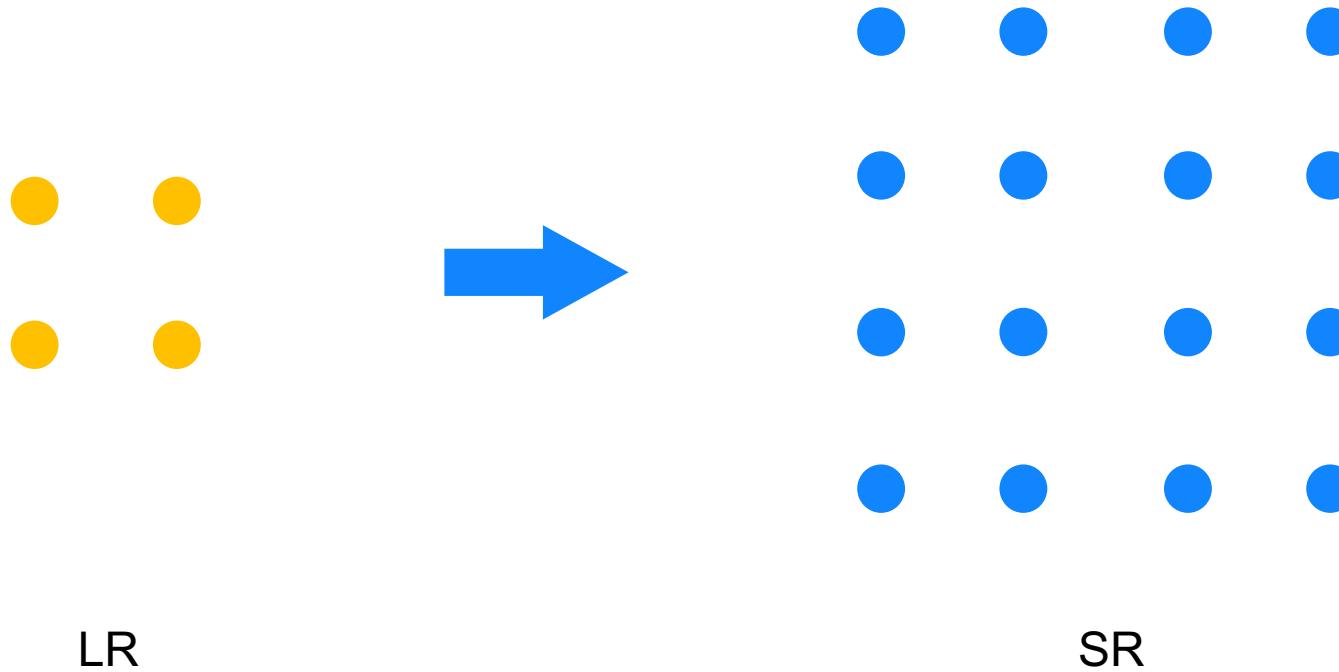
Preface



LR

Preface

Super-resolution



Preface

Super-resolution



LR

SR

简单来说就是把模糊的图片变清晰；有插值，边缘、学习等方法。图像超分辨率广泛应用于监控与安全、医学图像处理、媒体和艺术、航空航天和军事应用等领域

Paper report

CONTENT

- Feedback Network for Image Super-Resolution
- SwinIR: Image Restoration Using Swin Transformer
- Symmetric Uncertainty-Aware Feature Transmission for Depth Super-Resolution
- Recurrent Structure Attention Guidance for Depth Super-Resolution
- SGNet: Structure Guided Network via Gradient-Frequency Awareness for Depth Map Super-Resolution

Paper report

Feedback Network for Image Super-Resolution

Zhen Li¹ Jinglei Yang² Zheng Liu³ Xiaomin Yang^{1*} Gwanggil Jeon⁴ Wei Wu^{1*}
¹Sichuan University, ²University of California, Santa Barbara, ³University of British Columbia,
⁴Incheon National University

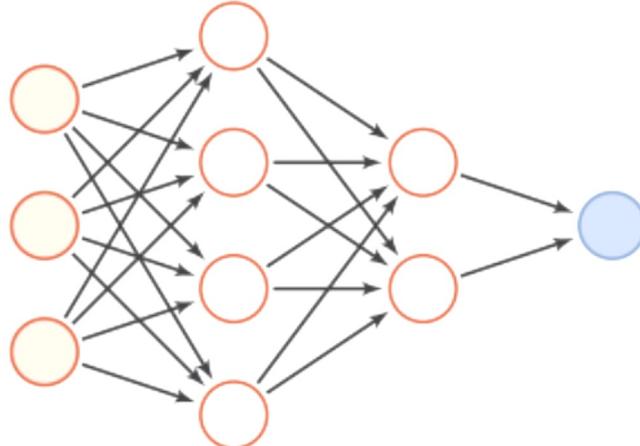
CVPR 2019

Paper report——SRFBN

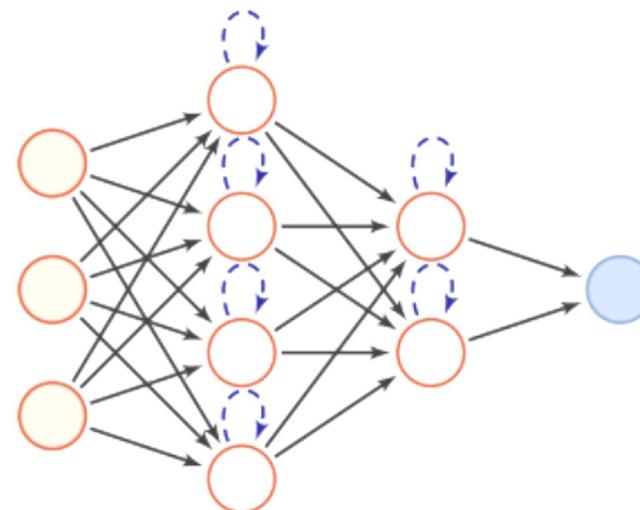
动机

- 如何将反馈机制引入深度学习中以改进图像超分辨率的性能？
- 如何利用高层信息来细化低层表示，从而提升图像重建质量？

同时，在人类视觉系统中存在反馈机制



(a) 前馈网络



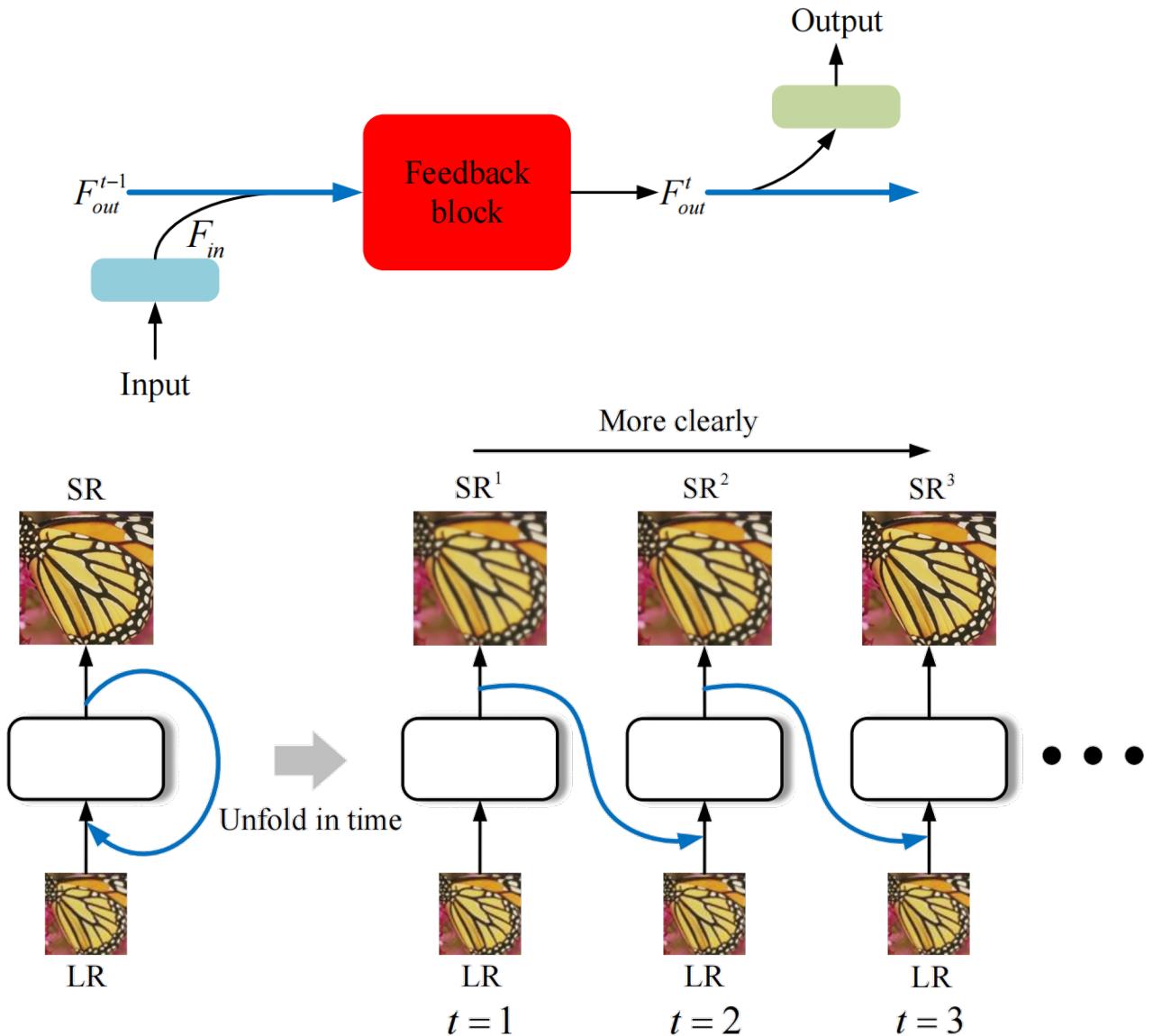
(b) 反馈网络

Paper report——SRFBN

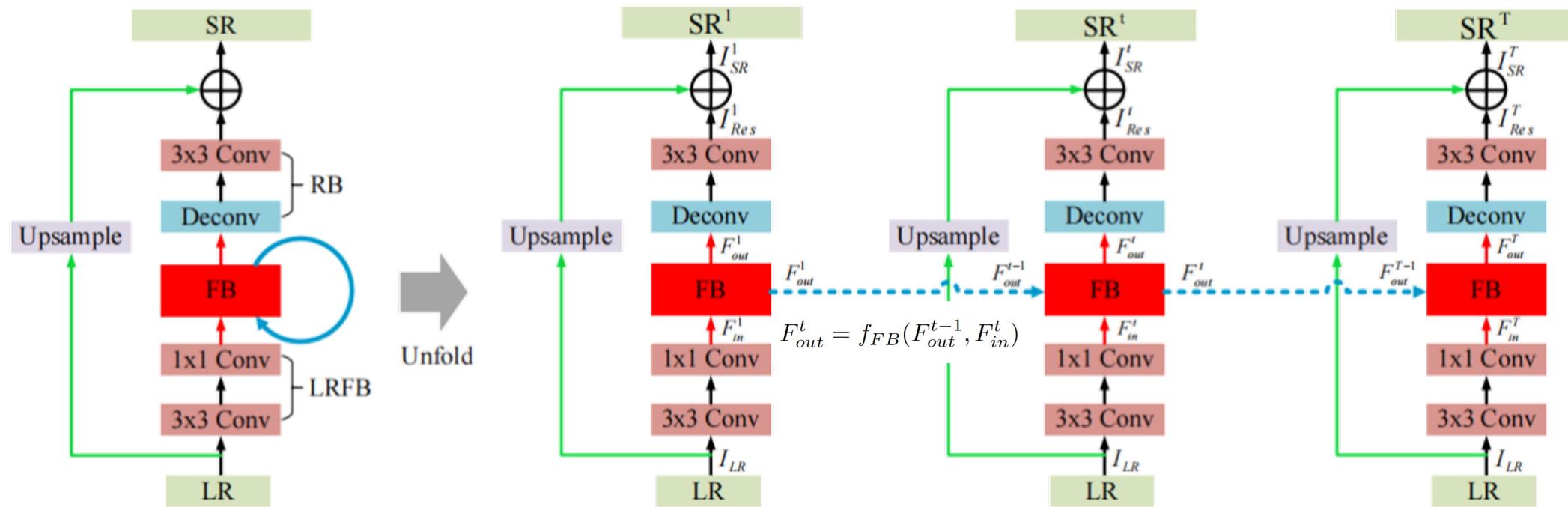
主要思路

提出了一种**回传机制**

优势：参数量较少，同时加深网络，
不断地的优化增强图像



Paper report——SRFBN



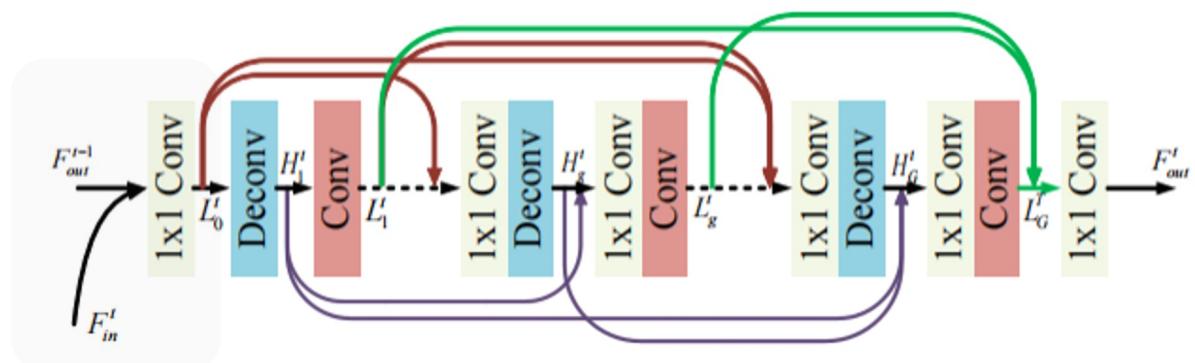
本质上是一个结合了 FB 结构的 RNN 网络，使用受约束的 RNN 的隐藏状态来实现反馈方式

Paper report——SRFBN

FB(反馈块)

但是加深网络也会带来梯度爆炸或消失等问题

这是 SRFBN 的核心部分，LR 的浅层特征和 $t-1$ 次的输出压缩融合作为第 t 次迭代的输入。
每次的迭代都会计算 loss，反向传播更新参数以减小损失函数的值



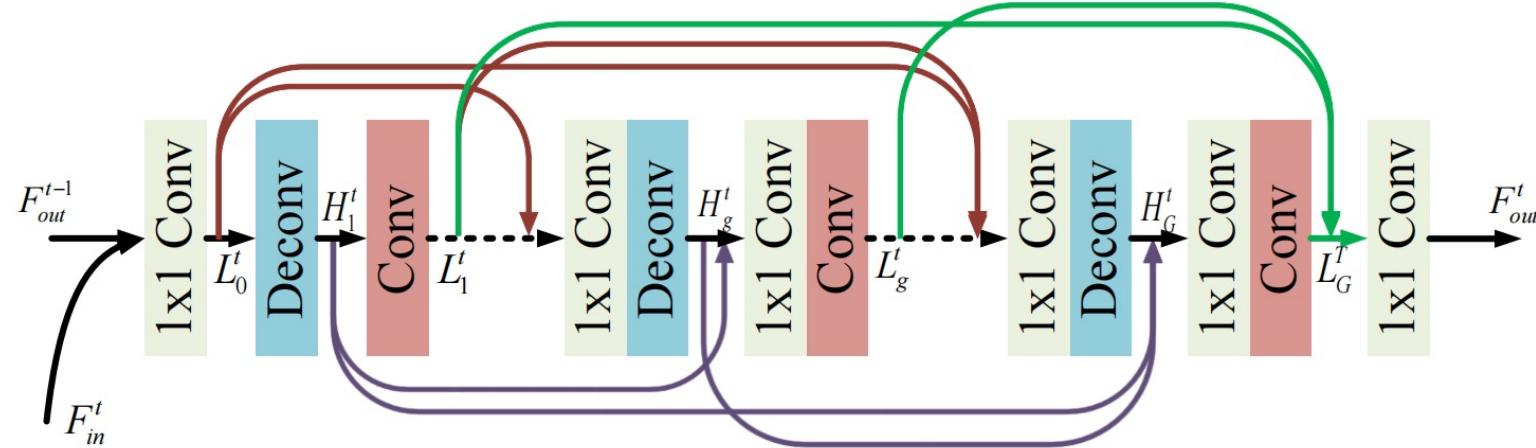
$$F_{in}^t = f_{LRFB}(I_{LR})$$

密集的残差跳跃

包含原始图像细节，具有准确性，多尺度特征

$$\begin{aligned} L_0^t &= C_0([F_{out}^{t-1}, F_{in}^t]) \\ H_g^t &= C_g^\uparrow([L_0^t, L_1^t, \dots, L_{g-1}^t]) \\ L_g^t &= C_g^\downarrow([H_1^t, H_2^t, \dots, H_g^t]) \\ F_{out}^t &= C_{FF}([L_1^t, L_2^t, \dots, L_G^t]) \end{aligned}$$

Paper report——SRFBN



激活函数

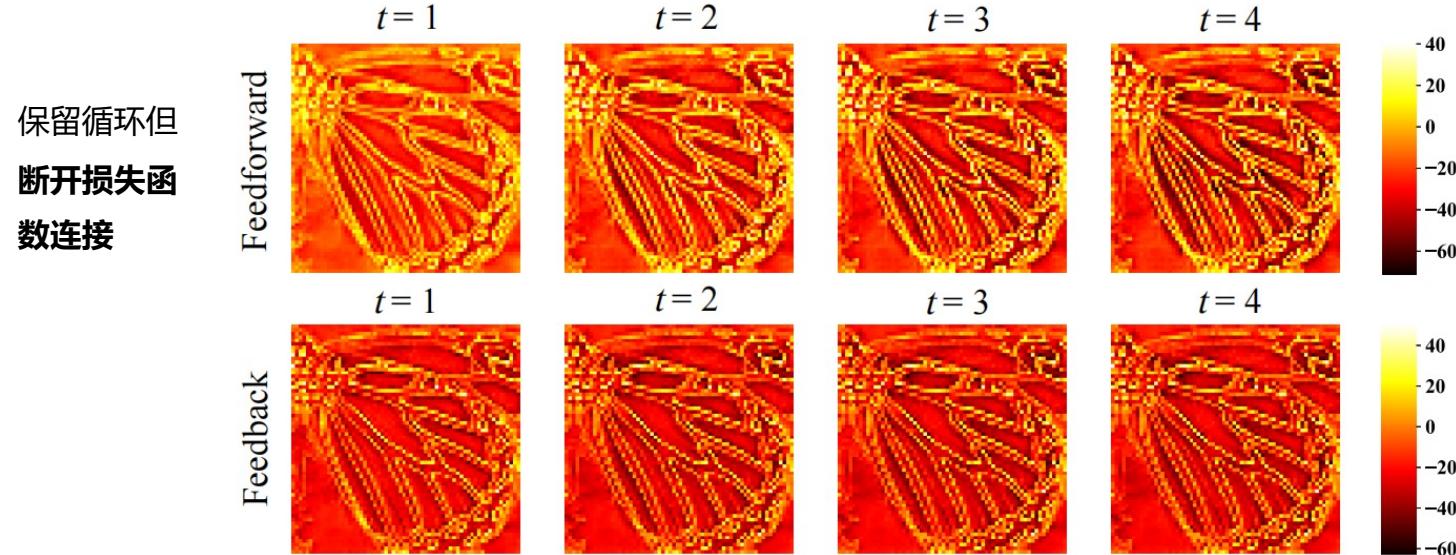
除最后一层外，其他卷积层操作后的都是 ReLU 作为激活函数

损失函数

使用 L1 损失，用来衡量生成图像与真实图像之间的差异 $L(\Theta) = \frac{1}{T} \sum_{t=1}^T W^t \|I_{HR}^t - I_{SR}^t\|_1$

对于复杂的退化模型， $(I_{HR}^1, I_{HR}^2, \dots, I_{HR}^T)$ 是根据 T 迭代任务的任务难度来执行课程的

Paper report——SRFBN



前馈与反馈的可视化特征图对比

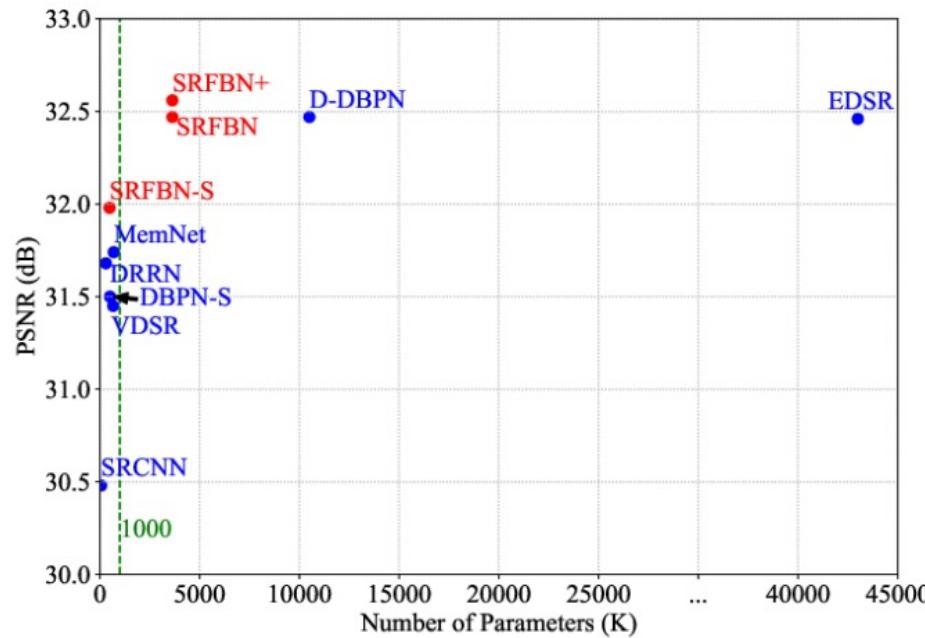
反馈网络获得的特征映射包含更多的**负值**。增强细节和纹理的捕捉能力，抑制平滑区域的效果

迭代次数

T 为迭代次数，G 为投影组；实验发现图像质量在 T=4，G=6 时基本达到收敛值

No. Prediction	1st	2nd	3rd	4th
SRFBN-L-FF	30.69	31.74	32.00	32.09
SRFBN-L	31.85	32.06	32.11	32.11

Paper report——SRFBN



优点和问题

优点：使用反馈机制，引入较少参数，效果较好
问题：迭代多次会增加计算量

那么有没有什么方法计算量较小还能达到更好的效果？

Paper report

SwinIR: Image Restoration Using Swin Transformer

Jingyun Liang¹ Jiezhang Cao¹ Guolei Sun¹ Kai Zhang^{1,*} Luc Van Gool^{1,2} Radu Timofte¹

¹Computer Vision Lab, ETH Zurich, Switzerland ²KU Leuven, Belgium

{jinliang, jiezcao, guosun, kai.zhang, vangool, timofter}@vision.ee.ethz.ch

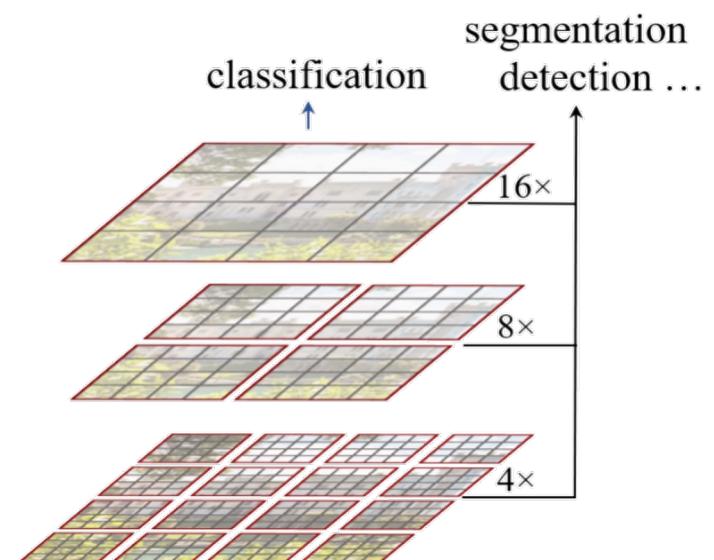
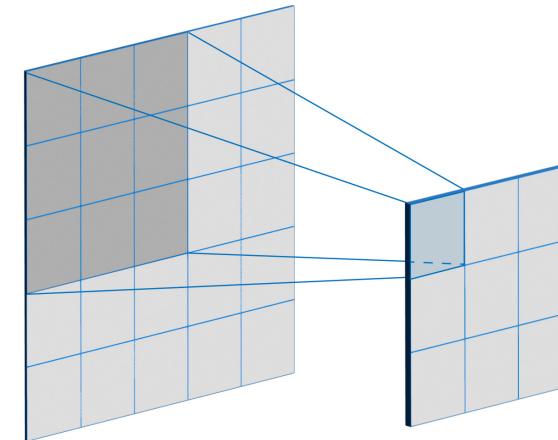
ICCV 2021

Paper report

SwinIR: Image Restoration Using Swin Transformer

SwinIR是一个基于 **Swin Transformer 架构**的图像恢复模型，利用了 Swin Transformer 的优势，**分层分块的自注意力和窗口滑动机制**，设计了一种高效且强大的图像恢复框架

性能优势：相比传统的卷积神经网络（CNN）架构，SwinIR 利用 Transformer 的能力更好地捕捉图像中的长距离依赖关系，并且能够有效地处理大规模的图像数据



Paper report——SwinIR

$$\begin{array}{c} Q \\ \begin{matrix} \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} \\ \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} \\ \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} \end{matrix} \end{array} = \begin{array}{c} I \\ \begin{matrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{matrix} \end{array} \times \begin{array}{c} W^Q \\ \begin{matrix} \textcolor{lightblue}{\square} & \square & \square \\ \square & \textcolor{lightblue}{\square} & \square \\ \square & \square & \textcolor{lightblue}{\square} \\ \square & \square & \square \end{matrix} \end{array}$$
$$\begin{array}{c} K \\ \begin{matrix} \textcolor{yellow}{\square} & \textcolor{yellow}{\square} & \textcolor{yellow}{\square} \\ \textcolor{yellow}{\square} & \textcolor{yellow}{\square} & \textcolor{yellow}{\square} \\ \textcolor{yellow}{\square} & \textcolor{yellow}{\square} & \textcolor{yellow}{\square} \end{matrix} \end{array} = \begin{array}{c} I \\ \begin{matrix} \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \end{matrix} \end{array} \times \begin{array}{c} W^K \\ \begin{matrix} \textcolor{yellow}{\square} & \square & \square \\ \square & \textcolor{yellow}{\square} & \square \\ \square & \square & \textcolor{yellow}{\square} \\ \square & \square & \square \end{matrix} \end{array}$$
$$\begin{array}{c} V \\ \begin{matrix} \textcolor{red}{\square} & \textcolor{red}{\square} & \textcolor{red}{\square} \\ \textcolor{red}{\square} & \textcolor{red}{\square} & \textcolor{red}{\square} \\ \textcolor{red}{\square} & \textcolor{red}{\square} & \textcolor{red}{\square} \end{matrix} \end{array} = \begin{array}{c} I \\ \begin{matrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{matrix} \end{array} \times \begin{array}{c} W^V \\ \begin{matrix} \textcolor{red}{\square} & \square & \square \\ \square & \textcolor{red}{\square} & \square \\ \square & \square & \textcolor{red}{\square} \\ \square & \square & \square \end{matrix} \end{array}$$

Self-Attention

输入数据内部进行关注不同位置的方法，而不依赖于输入的绝对位置或者顺序。对于图像，可以将自注意机制看作是通过计算每个像素点或者像素块之间的相似度来分配权重，以捕捉全局和局部的语义信息

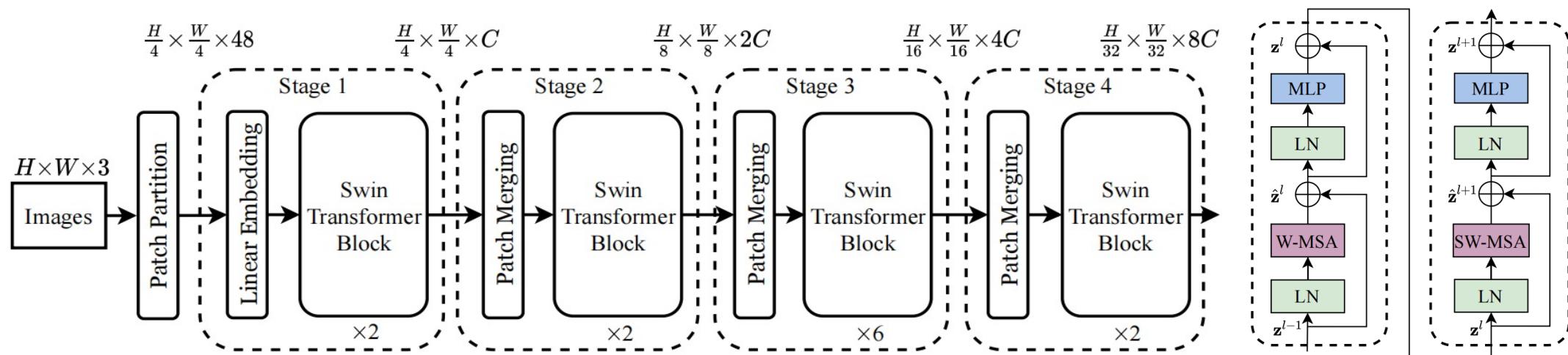
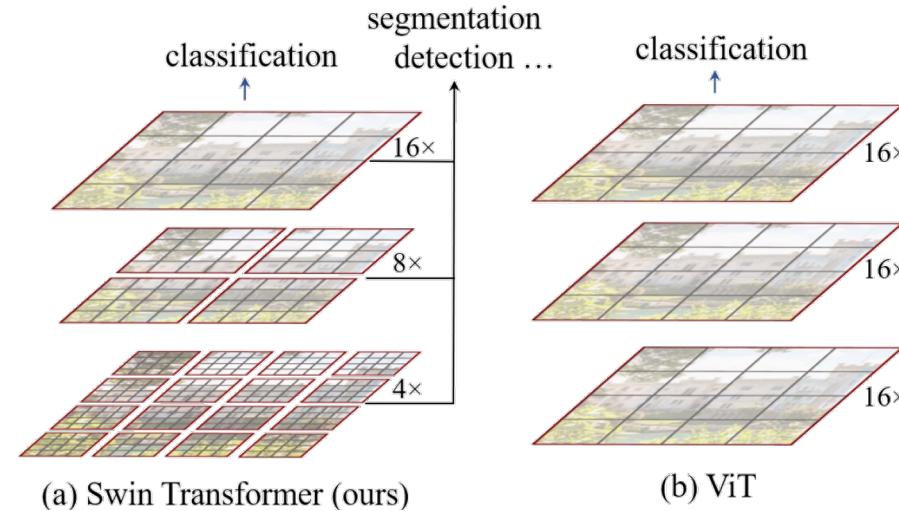
$$\text{Attention}(Q, K, V) = \text{SoftMax}(QK^T / \sqrt{d} + B)V$$

$$\text{softmax} \left(\frac{\begin{array}{c} Q \\ \begin{matrix} \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} \\ \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} \\ \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} & \textcolor{lightblue}{\square} \end{matrix} \end{array} \times \begin{array}{c} K^T \\ \begin{matrix} \textcolor{yellow}{\square} & \textcolor{yellow}{\square} \\ \textcolor{yellow}{\square} & \textcolor{yellow}{\square} \\ \textcolor{yellow}{\square} & \textcolor{yellow}{\square} \end{matrix} \end{array}}{\sqrt{d_k}} + \begin{array}{c} B \\ \begin{matrix} \square & \square \\ \square & \square \end{matrix} \end{array} \right) \times \begin{array}{c} V \\ \begin{matrix} \textcolor{red}{\square} & \textcolor{red}{\square} & \textcolor{red}{\square} \\ \textcolor{red}{\square} & \textcolor{red}{\square} & \textcolor{red}{\square} \\ \textcolor{red}{\square} & \textcolor{red}{\square} & \textcolor{red}{\square} \end{matrix} \end{array}$$

Paper report——SwinIR

Swin Transformer

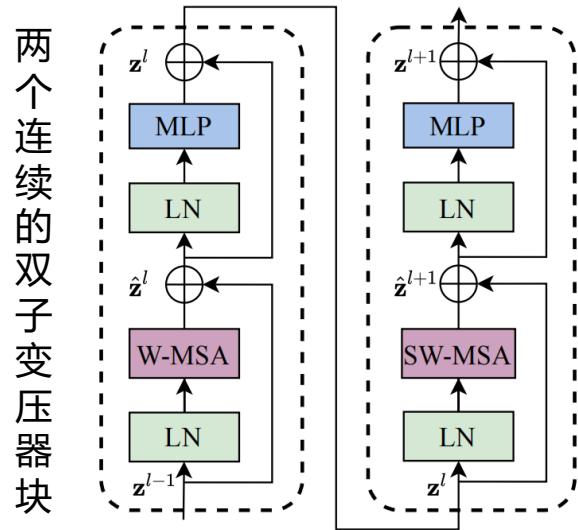
- 分层分块的自注意力



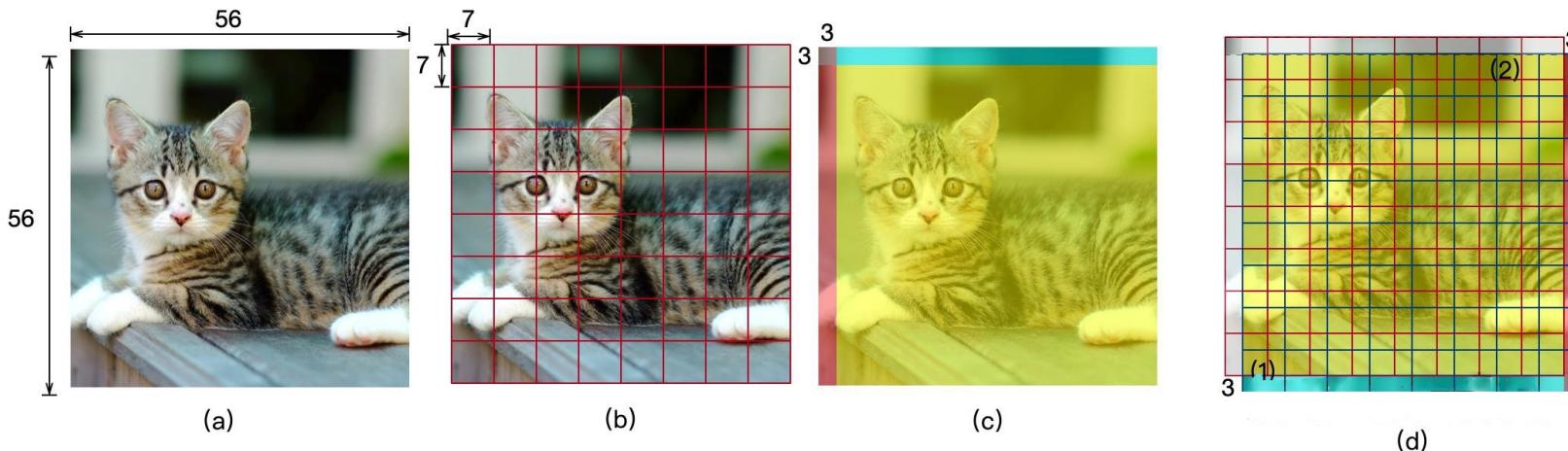
Swin Transform 分块只是单独计算各个块中的注意力，会造成弱交互能力

Paper report——SwinIR

动画模拟



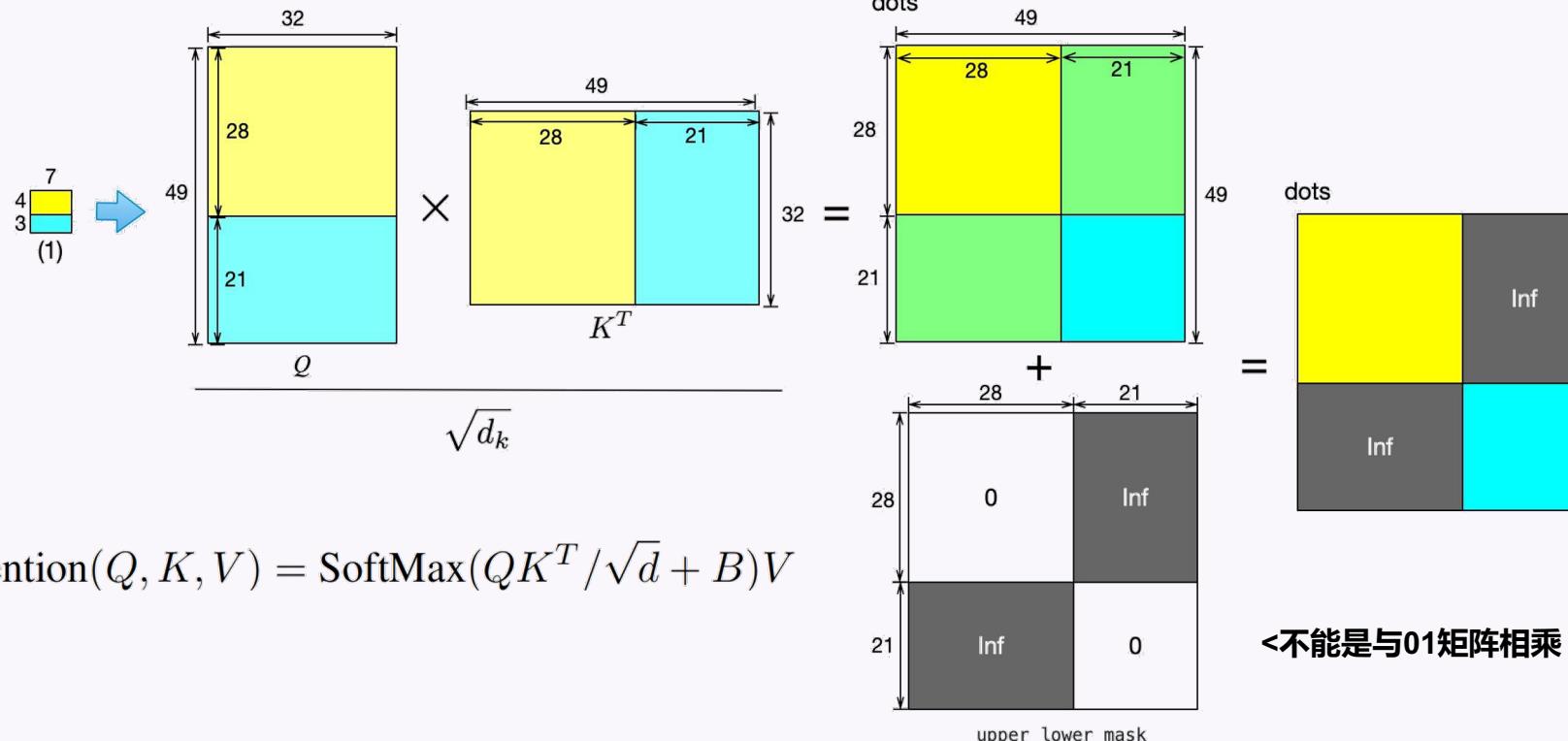
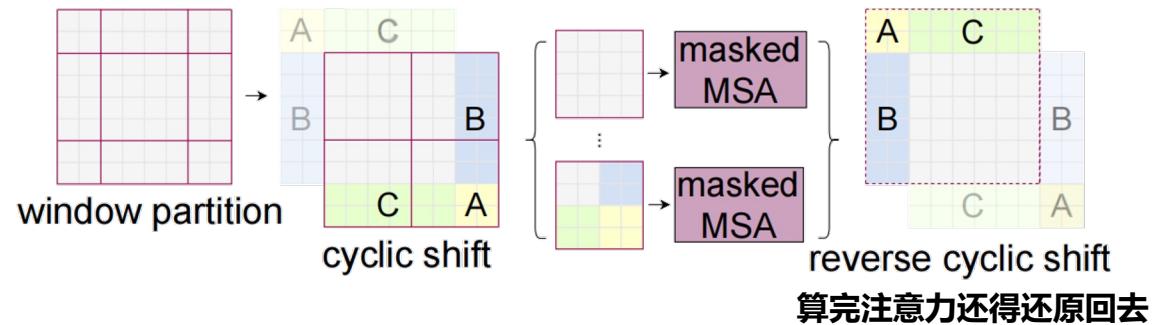
窗口移动机制
计算跨窗口注意力
增强交互能力



问题：
对于不连续块之间的
注意力计算问题

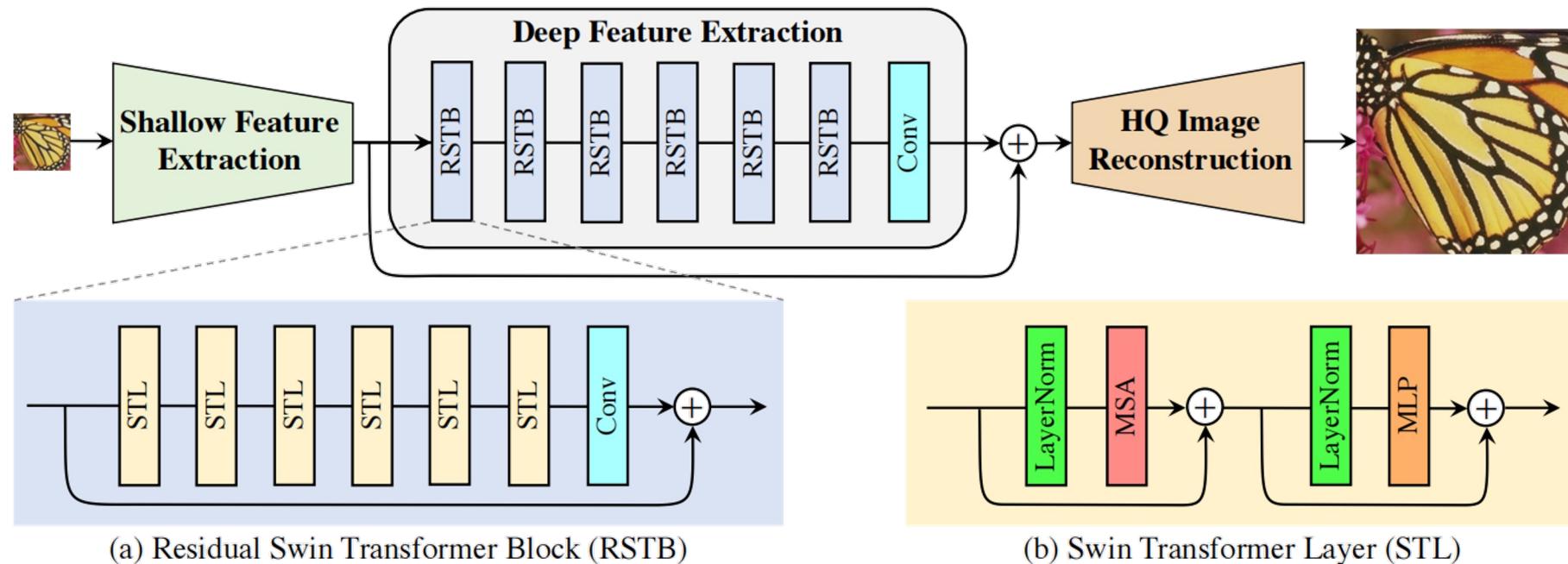
Paper report——SwinIR

为了过滤有害信息，需要进行
掩码操作(Mask)



Paper report——SwinIR

SwinIR: Image Restoration Using Swin Transformer的网络架构



采用密集的STL串联层提取深度特征，因为Swin Transformer架构具有高效的特征提取能力，
多尺度的特征融合，建立有效的长距离依赖，在图像处理上能达到很好的效果

Paper report——SwinIR

Method	Scale	Training Dataset	Set5 [3]		Set14 [87]		BSD100 [58]		Urban100 [34]		Manga109 [60]	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RCAN [95]	×2	DIV2K	38.27	0.9614	34.12	0.9216	32.41	0.9027	33.34	0.9384	39.44	0.9786
SAN [15]	×2	DIV2K	38.31	0.9620	34.07	0.9213	32.42	0.9028	33.10	0.9370	39.32	0.9792
IGNN [100]	×2	DIV2K	38.24	0.9613	34.07	0.9217	32.41	0.9025	33.23	0.9383	39.35	0.9786
HAN [63]	×2	DIV2K	38.27	0.9614	34.16	0.9217	32.41	0.9027	33.35	0.9385	39.46	0.9785
NLSA [61]	×2	DIV2K	38.34	0.9618	34.08	0.9231	32.43	0.9027	33.42	0.9394	39.59	0.9789
SwinIR (Ours)	×2	DIV2K	38.35	0.9620	34.14	0.9227	32.44	0.9030	33.40	0.9393	39.60	0.9792
SwinIR+ (Ours)	×2	DIV2K	38.38	0.9621	34.24	0.9233	32.47	0.9032	33.51	0.9401	39.70	0.9794
DBPN [31]	×2	DIV2K+Flickr2K	38.09	0.9600	33.85	0.9190	32.27	0.9000	32.55	0.9324	38.89	0.9775
IPT [9]	×2	ImageNet	38.37	-	34.43	-	32.48	-	33.76	-	-	-
SwinIR (Ours)	×2	DIV2K+Flickr2K	38.42	0.9623	34.46	0.9250	32.53	0.9041	33.81	0.9427	39.92	0.9797
SwinIR+ (Ours)	×2	DIV2K+Flickr2K	38.46	0.9624	34.61	0.9260	32.55	0.9043	33.95	0.9433	40.02	0.9800
RCAN [95]	×3	DIV2K	34.74	0.9299	30.65	0.8482	29.32	0.8111	29.09	0.8702	34.44	0.9499
SAN [15]	×3	DIV2K	34.75	0.9300	30.59	0.8476	29.33	0.8112	28.93	0.8671	34.30	0.9494
IGNN [100]	×3	DIV2K	34.72	0.9298	30.66	0.8484	29.31	0.8105	29.03	0.8696	34.39	0.9496
HAN [63]	×3	DIV2K	34.75	0.9299	30.67	0.8483	29.32	0.8110	29.10	0.8705	34.48	0.9500
NLSA [61]	×3	DIV2K	34.85	0.9306	30.70	0.8485	29.34	0.8117	29.25	0.8726	34.57	0.9508
SwinIR (Ours)	×3	DIV2K	34.89	0.9312	30.77	0.8503	29.37	0.8124	29.29	0.8744	34.74	0.9518
SwinIR+ (Ours)	×3	DIV2K	34.95	0.9316	30.83	0.8511	29.41	0.8130	29.42	0.8761	34.92	0.9526
IPT [9]	×3	ImageNet	34.81	-	30.85	-	29.38	-	29.49	-	-	-
SwinIR (Ours)	×3	DIV2K+Flickr2K	34.97	0.9318	30.93	0.8534	29.46	0.8145	29.75	0.8826	35.12	0.9537
SwinIR+ (Ours)	×3	DIV2K+Flickr2K	35.04	0.9322	31.00	0.8542	29.49	0.8150	29.90	0.8841	35.28	0.9543
RCAN [95]	×4	DIV2K	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087	31.22	0.9173
SAN [15]	×4	DIV2K	32.64	0.9003	28.92	0.7888	27.78	0.7436	26.79	0.8068	31.18	0.9169
IGNN [100]	×4	DIV2K	32.57	0.8998	28.85	0.7891	27.77	0.7434	26.84	0.8090	31.28	0.9182
HAN [63]	×4	DIV2K	32.64	0.9002	28.90	0.7890	27.80	0.7442	26.85	0.8094	31.42	0.9177
NLSA [61]	×4	DIV2K	32.59	0.9000	28.87	0.7891	27.78	0.7444	26.96	0.8109	31.27	0.9184
SwinIR (Ours)	×4	DIV2K	32.72	0.9021	28.94	0.7914	27.83	0.7459	27.07	0.8164	31.67	0.9226
SwinIR+ (Ours)	×4	DIV2K	32.81	0.9029	29.02	0.7928	27.87	0.7466	27.21	0.8187	31.88	0.9423
DBPN [31]	×4	DIV2K+Flickr2K	32.47	0.8980	28.82	0.7860	27.72	0.7400	26.38	0.7946	30.91	0.9137
IPT [9]	×4	ImageNet	32.64	-	29.01	-	27.82	-	27.26	-	-	-
RRDB [81]	×4	DIV2K+Flickr2K	32.73	0.9011	28.99	0.7917	27.85	0.7455	27.03	0.8153	31.66	0.9196
SwinIR (Ours)	×4	DIV2K+Flickr2K	32.92	0.9044	29.09	0.7950	27.92	0.7489	27.45	0.8254	32.03	0.9260
SwinIR+ (Ours)	×4	DIV2K+Flickr2K	32.93	0.9043	29.15	0.7958	27.95	0.7494	27.56	0.8273	32.22	0.9273

实验结果 (classical image)

红色为最佳，蓝色为其次

SwinIR模型的效果在不同放大倍数不同指标下均达到非常好的效果，尤其是在大scale下

Symmetric Uncertainty-Aware Feature Transmission for Depth Super-Resolution

Wuxuan Shi

wuxuansi@whu.edu.cn

School of Computer Science, Wuhan
University
Wuhan, China

Mang Ye*

yemang@whu.edu.cn

School of Computer Science, Wuhan
University
Hubei Luojia Laboratory
Wuhan, China

Bo Du

dubo@whu.edu.cn

School of Computer Science, Wuhan
University
Hubei Luojia Laboratory
Wuhan, China

AAAI 2023

Paper report

深度图(depth map):

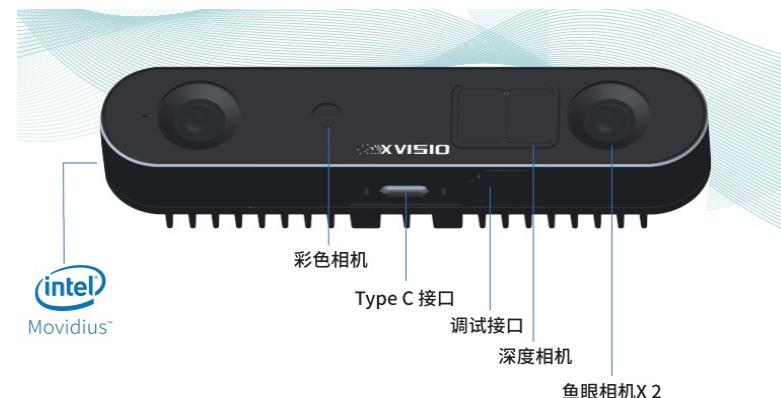
深度图是一种灰度图像，其中每个像素的值表示该像素对应场景中的距离或深度。深度图通常用于3D重建、虚拟现实、增强现实、自动驾驶等领域。

深度图超分:

指从低分辨率的深度图像中重建出更高分辨率、更精确的深度信息的技术。传统方法通常使用常规深度学习网络框架直接对深度图进行超分，但为了提升准确性，近年来**色彩引导**深度图超分作为一种新兴方法逐渐受到关注。



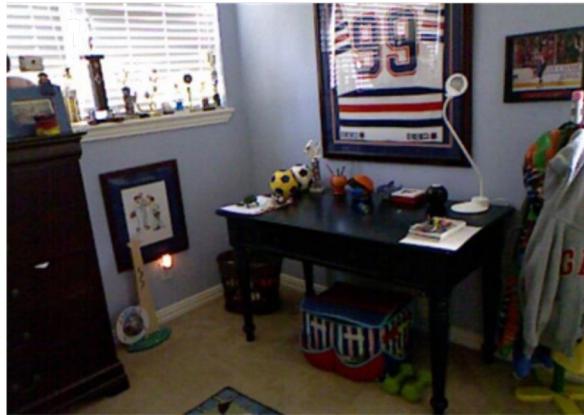
RGB图像与深度图



RGB-D相机

Paper report——SUFT

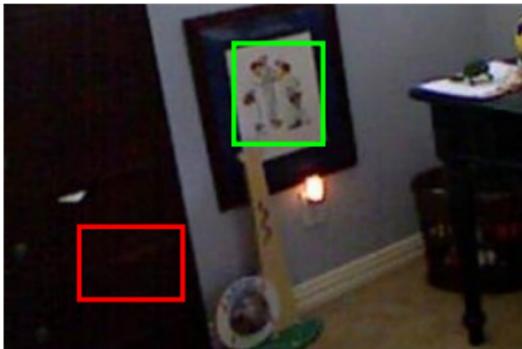
SUFT(Symmetric Uncertainty-Aware Feature Transmission for Depth Super-Resolution)



Resolution
Mismatch



(a) Resolution Gap



□ Texture Inconsistent Region



□ Texture Consistent Region

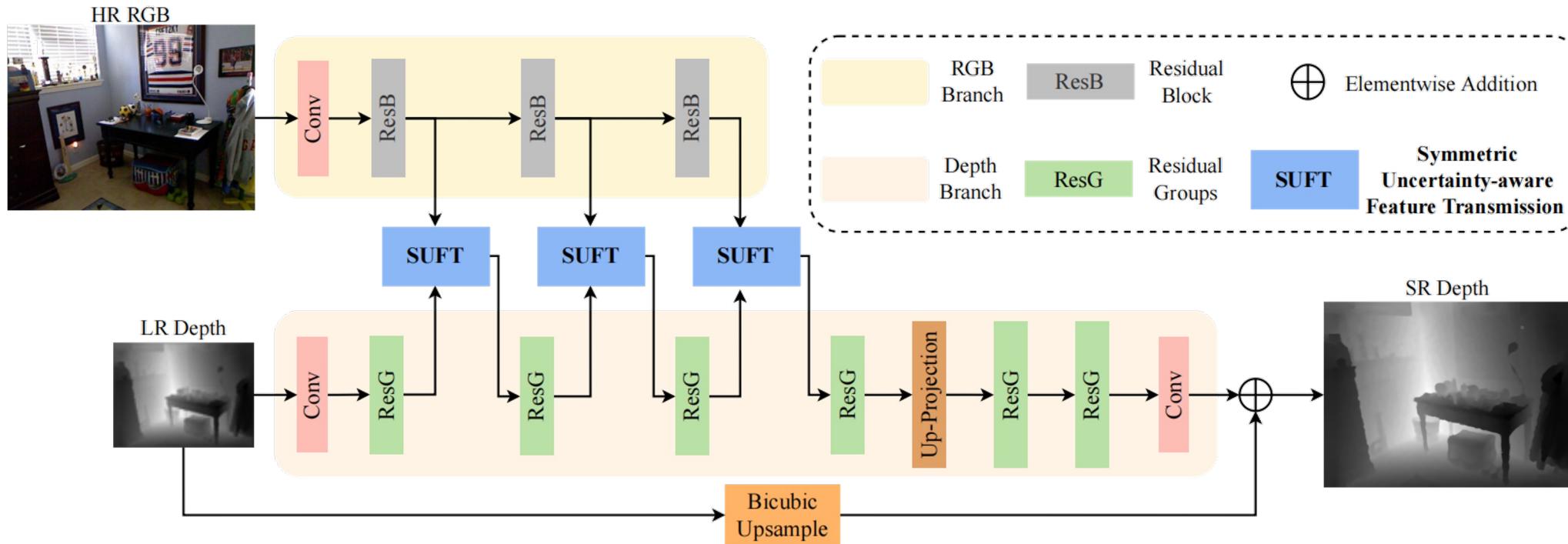
色彩引导深度超分的问题：

- 分辨率差异
- 跨模态差异

现有色彩引导方法通常通过插值将低分辨率深度图像放大后再输入网络，以从高分辨率RGB图像中提取的高频信息来指导深度图的重建。然而，这种插值操作会增加噪声并模糊图像

Paper report——SUFT

通过建立一个迭代的上采样和下采样管道，同时对RGB图像和深度图进行特征提取，从而解决分辨率差异问题，同时抑制噪声放大和模糊化，而不是简单地插值预上采样

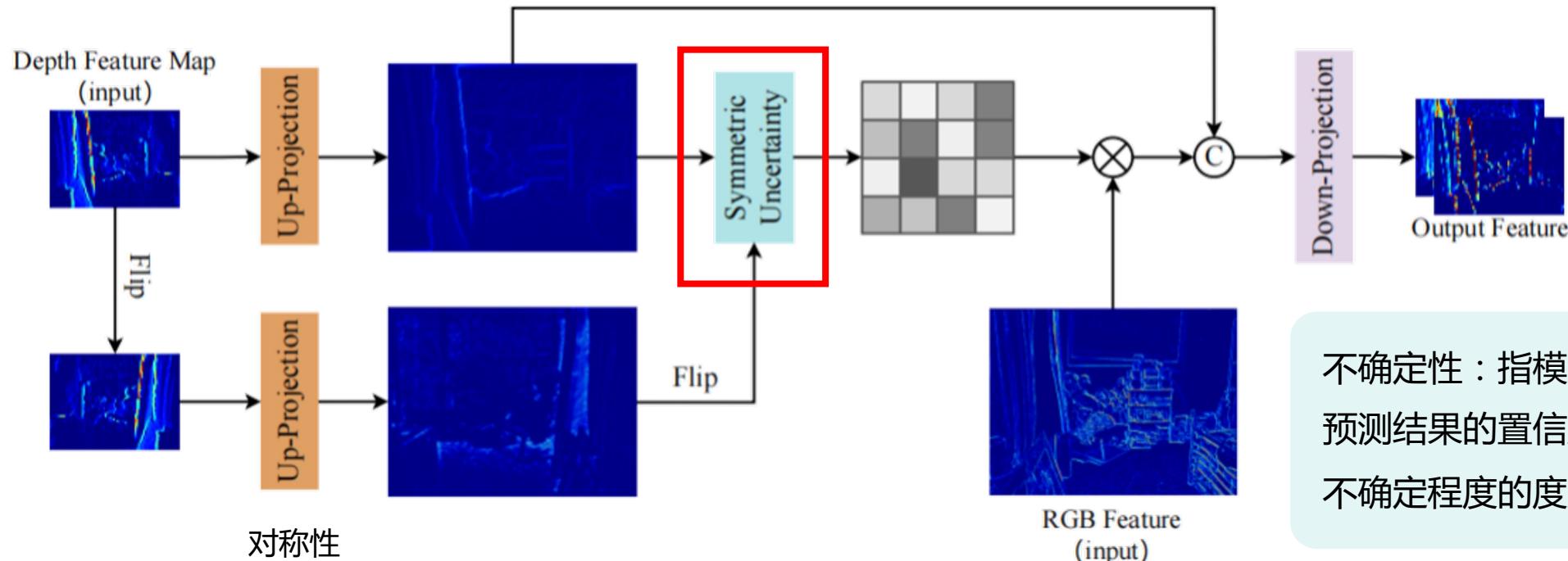


Paper report——SUFT

为了缩小交叉模态间隙，提出了SUFT结构

© Concatenation

\otimes Elementwise Multiplication

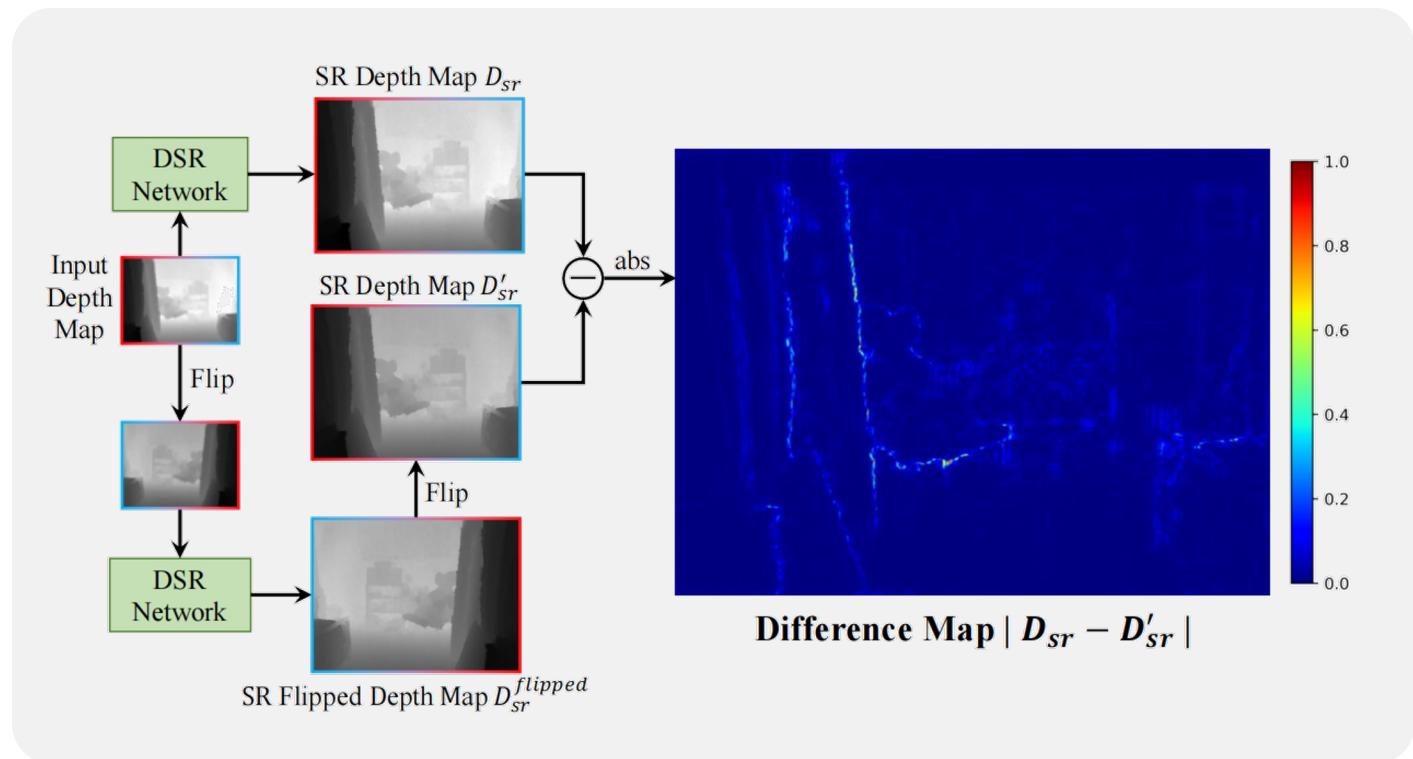
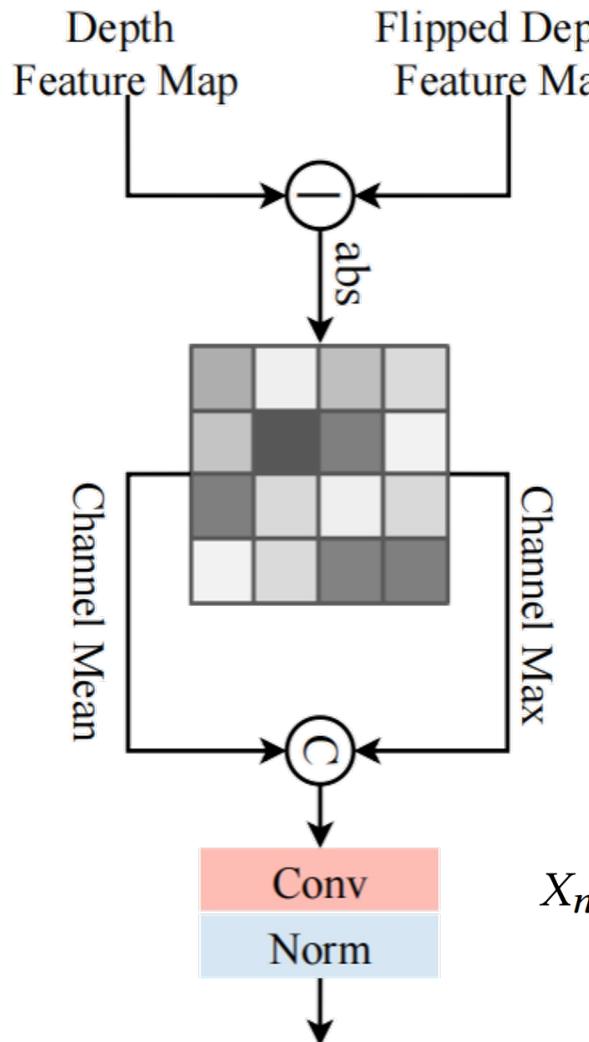


不确定性：指模型对预测结果的置信度或不确定程度的度量

卷积使 F_{hr}^{depth} 和 $F_{hr}^{flipped}$ 在相应位置上存在差异，并利用该差异来计算不确定度

Paper report——SUFT

对称不确定度

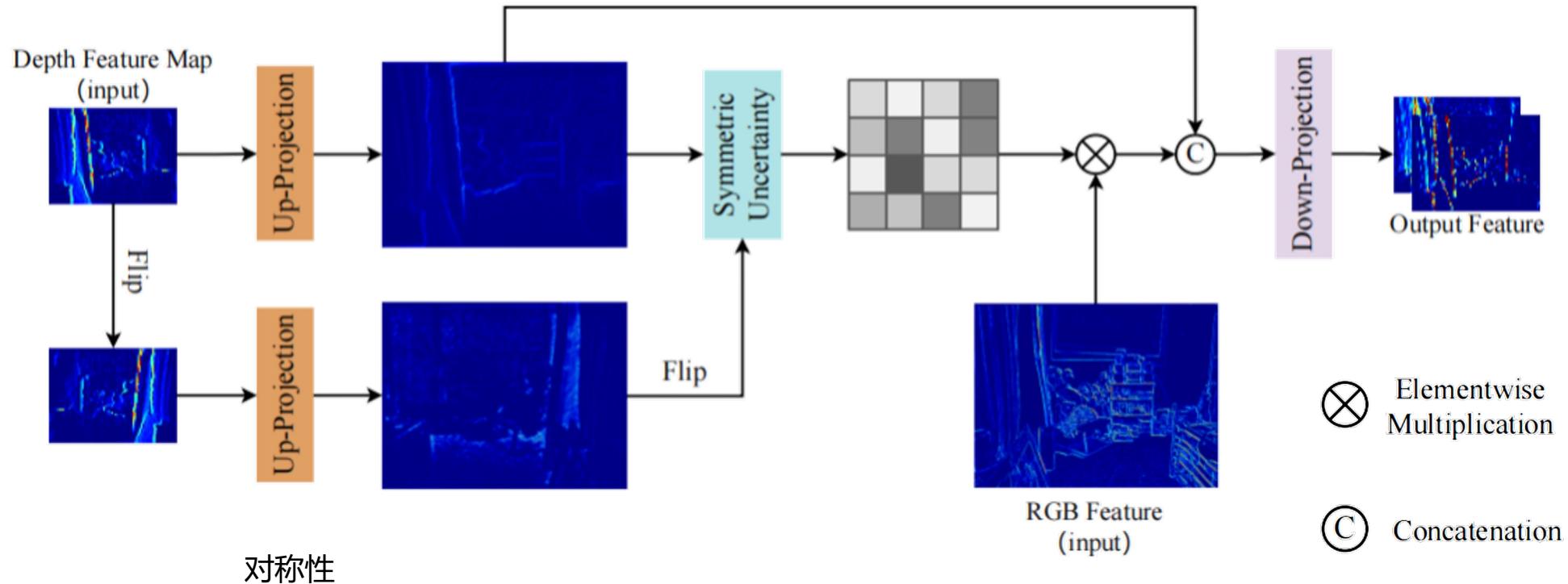


深度图中不确定性较高的区域（**纹理或边缘不清晰的地方**）在 RGB 特征传输过程中被赋予较高的权重。相反，深度图中不确定性较低的区域则接收较低的权重，实现有害信息的过滤。

$$X_{norm} = \frac{X - min}{(max - min) + \epsilon}$$

Paper report——SUFT

SUFT结构



用它与RGB特征进行乘法，掩盖有差异的纹理信息，实现有害信息的过滤

Paper report——SUFT

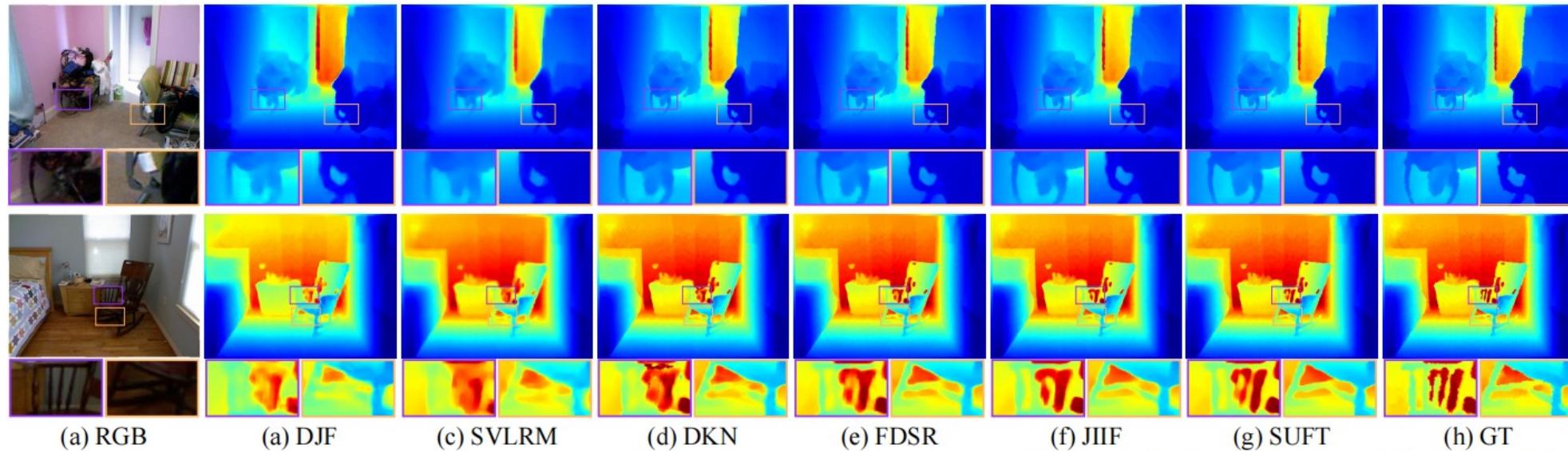


Figure 5: Visual comparison of $\times 8$ DSR results on NYU v2 dataset. (a) RGB images. (b) DJF [32]. (c) SVLRM [42]. (d) DKN [28]. (e) FDSR [19]. (f) JIIF [50]. (g) SUFT. (h) GT. Best viewed in color.

Table 1: Quantitative DSR results (in average RMSE) on NYU v2 dataset. The average RMSE is measured in centimeter. The best performance is displayed in bold, and the second best performance is underlined.

RMSE	Bicubic	MRF [58]	TGV [10]	SDF [15]	DG [12]	SVLRM [42]	DKN [28]	FDSR [19]	CTKT [49]	JIIF [50]	BridgeNet [51]	SUFT
$\times 4$	8.16	7.84	6.98	3.04	1.56	1.74	1.62	1.61	1.49	<u>1.37</u>	1.54	1.12
$\times 8$	14.22	13.98	11.23	5.67	2.99	5.59	3.26	3.18	2.73	<u>2.76</u>	<u>2.63</u>	2.51
$\times 16$	22.32	22.20	28.31	9.97	5.24	7.23	6.51	5.86	5.11	5.27	<u>4.98</u>	4.86

实验结果（指标为均方根误差）

Recurrent Structure Attention Guidance for Depth Super-Resolution

Jiayi Yuan*, Haobo Jiang*, Xiang Li, Jianjun Qian[†], Jun Li[†], Jian Yang

PCA Lab, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education

Jiangsu Key Lab of Image and Video Understanding for Social Security

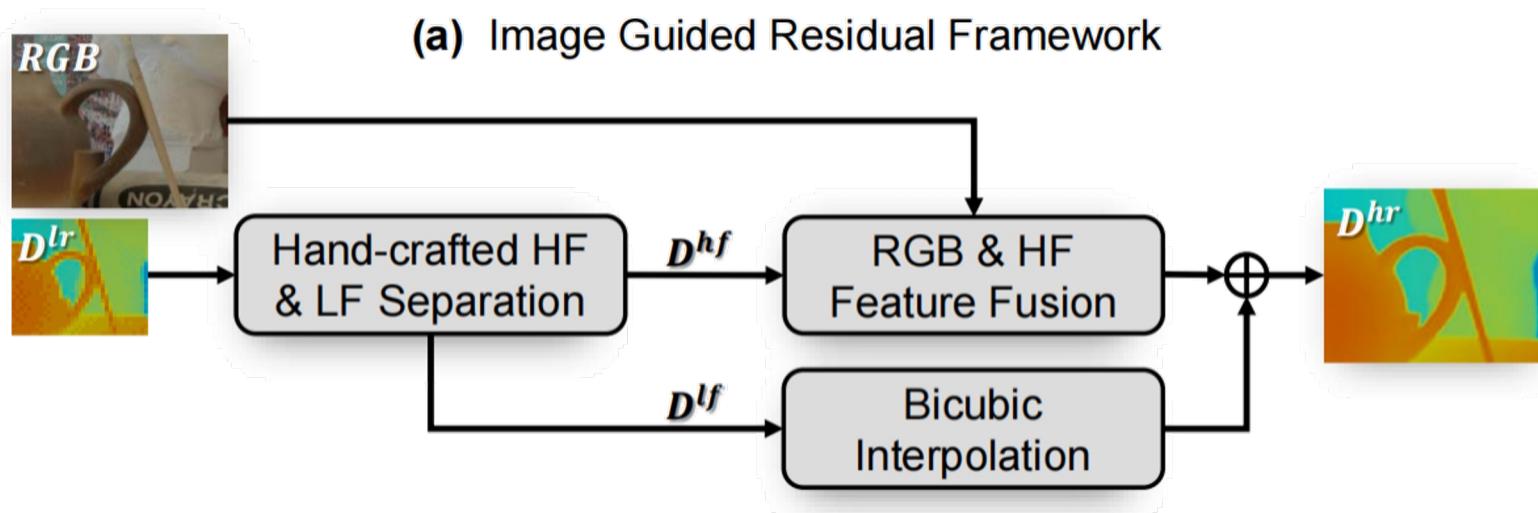
School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

{jiayiyuan, jiang.hao.bo, xiang.li.implus, csjqian, junli, csjyang}@njust.edu.cn

arXiv:2301.13419v1 [cs.CV] 31 Jan 2023

Paper report——RSAG

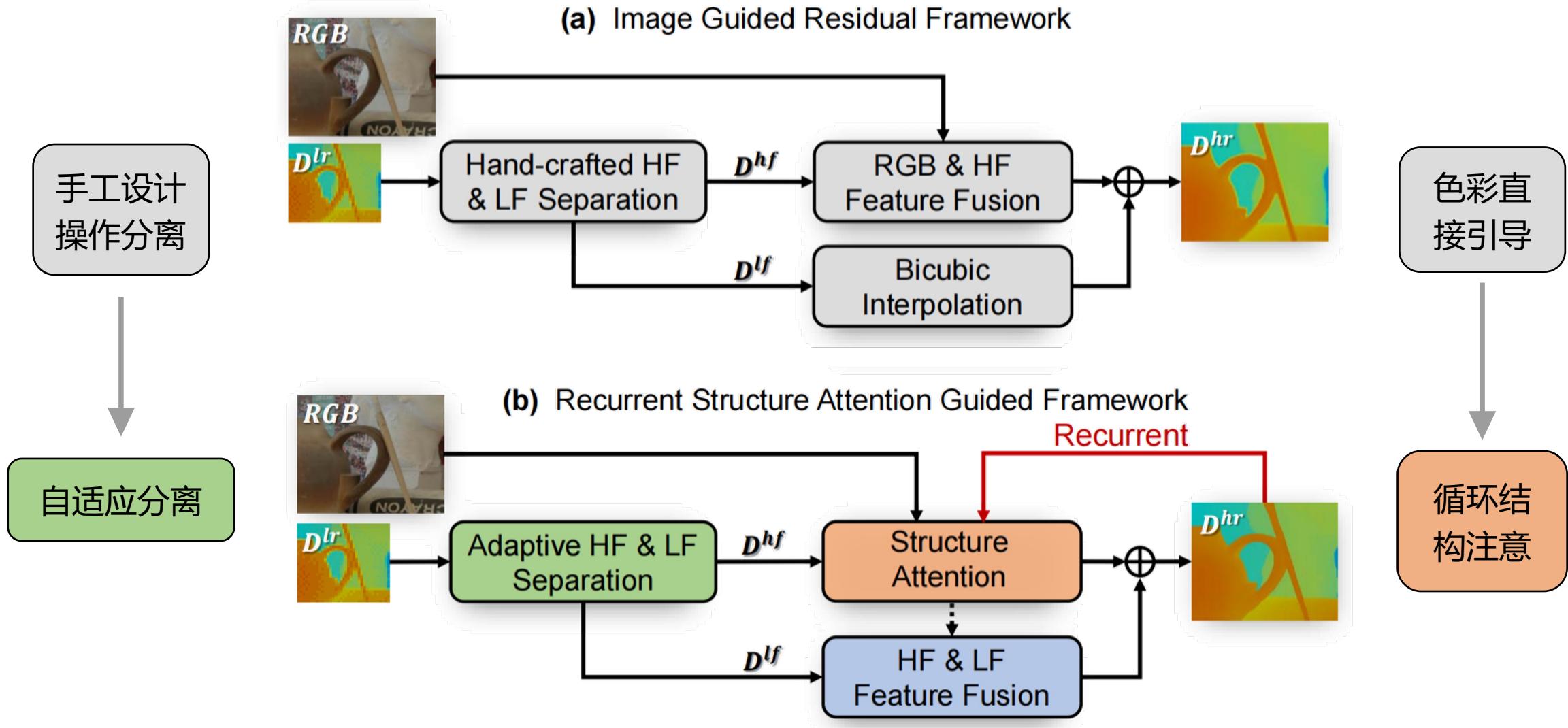
4. Recurrent Structure Attention Guidance for Depth Super-Resolution



动机：常规色彩引导深度超分方法的问题

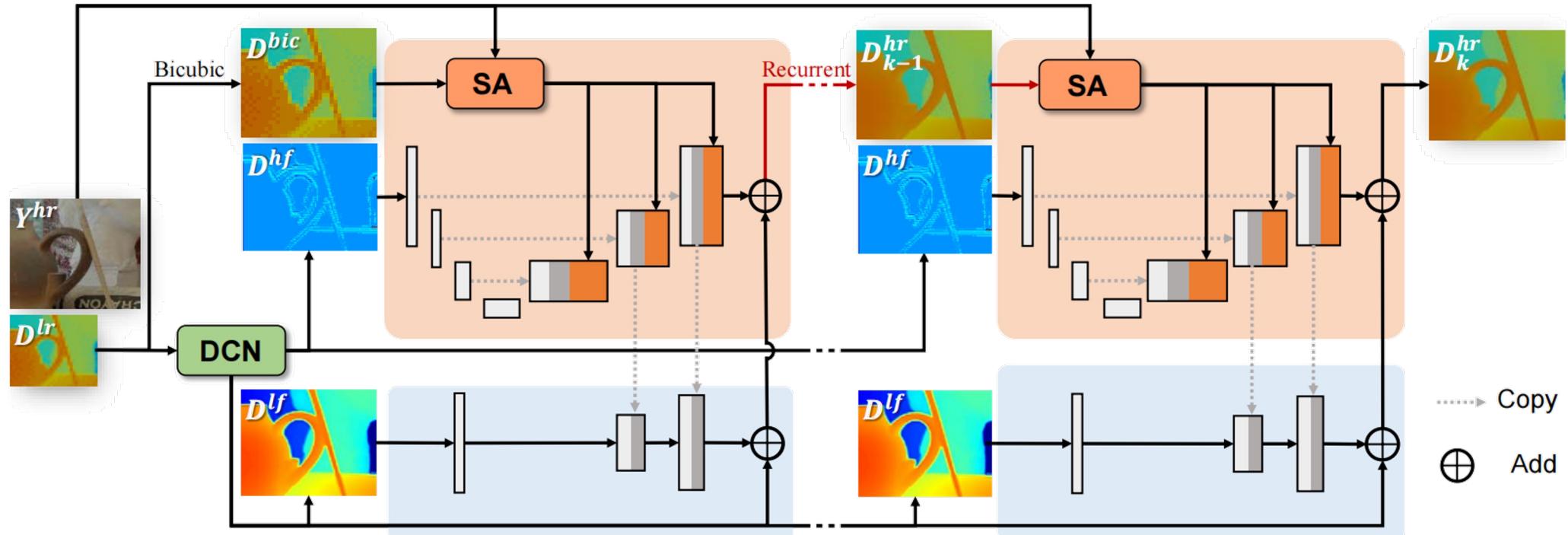
- 手工设计的操作符光谱分解较弱
- 特征直接连接导致图像引导性弱

Paper report——RSAG



Paper report——RSAG

- 展开的网络框架



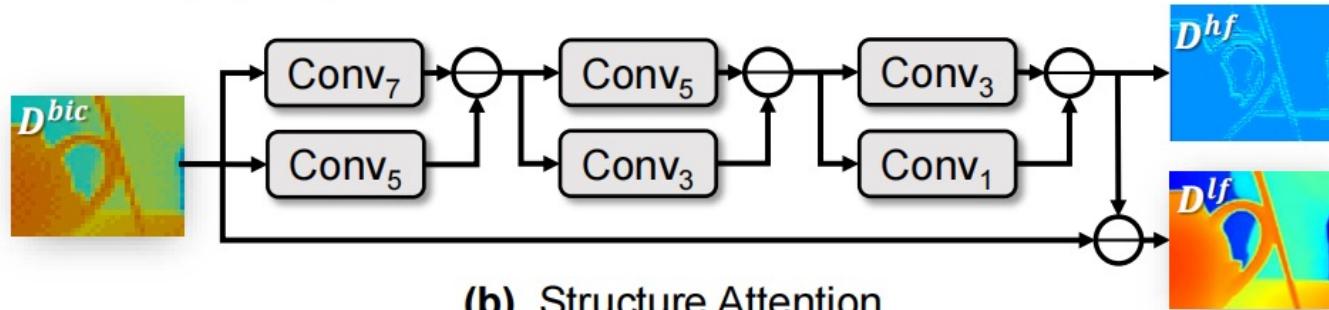
HF重建快：U-Net架构

通过卷积神经网络的各个层级和跳跃连接提取多尺度的特征

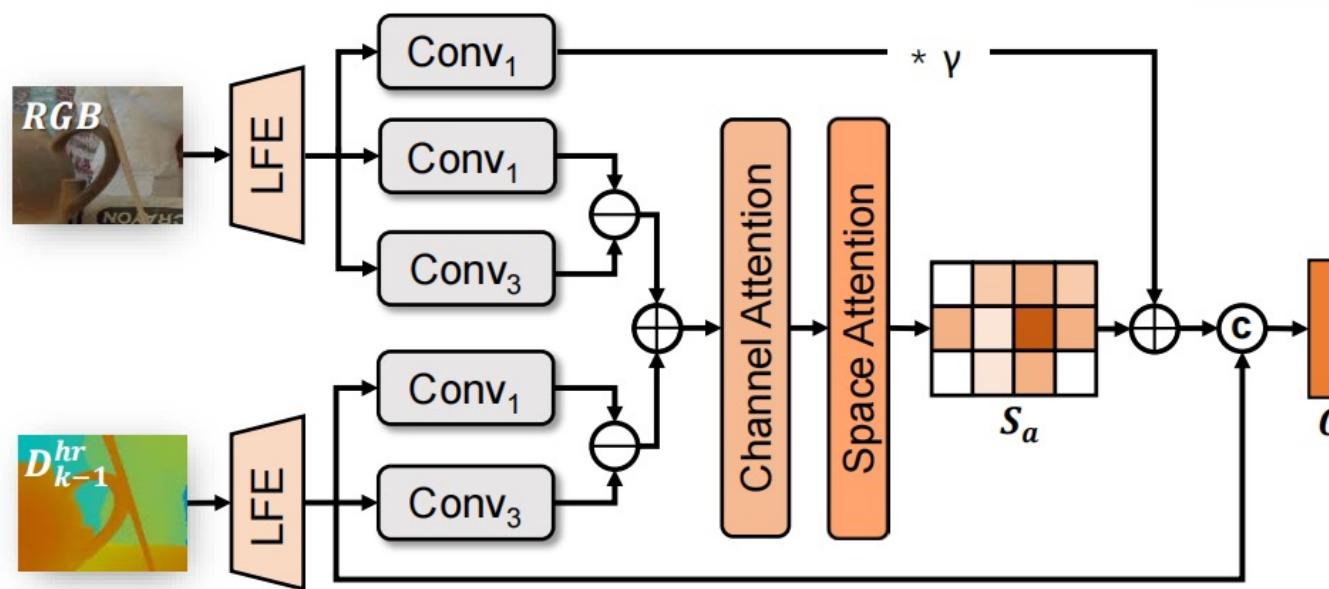
循环结构使将先前的深度预测
作为下一次递归的引导信息

Paper report——RSAG

(a) Deep Contrastive Network with Multi-scale Filters



(b) Structure Attention



DCN

采用大核到小核的像素对比叠
加来实现高频特征的提取，解
决手工设计的操作符光谱分
解较弱问题

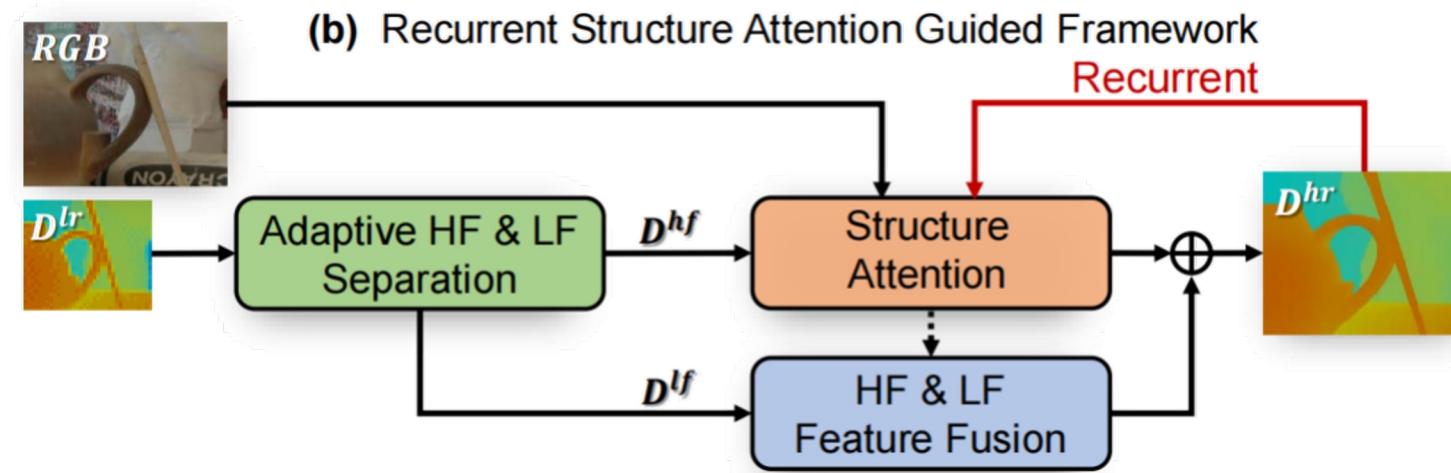
SA

提取有用特征，抑制有害信息
解决模态差异问题

$$J = |F_{y_1} - F_{y_2}| + |F_{d_1} - F_{d_2}|$$

Paper report——RSAG

损失函数：
smooth L1



$$\mathcal{L}_k(D_k^{hr}, D_i^{gt}) = \sum_{i=1}^N \text{smooth}_{L1}(D_{k,i}^{hr}, D_i^{gt}) \quad \text{smooth}_{L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases}$$

$$\mathcal{L}_s = \sum_{k=1}^K \lambda_k \mathcal{L}_k$$

<对损失进行求和

Paper report——RSAG

Model	<i>Art</i>			<i>Books</i>			<i>Dolls</i>			<i>Laundry</i>			<i>Mobius</i>			<i>Reindeer</i>		
	$\times 4$	$\times 8$	$\times 16$	$\times 4$	$\times 8$	$\times 16$	$\times 4$	$\times 8$	$\times 16$	$\times 4$	$\times 8$	$\times 16$	$\times 4$	$\times 8$	$\times 16$	$\times 4$	$\times 8$	$\times 16$
Bicbuic	1.15	2.15	4.04	0.41	0.72	1.32	0.44	0.76	1.31	0.65	1.17	2.17	0.41	0.76	1.37	0.66	1.16	2.26
DJF	0.40	1.07	2.78	0.16	0.45	1.00	0.20	0.49	0.99	0.28	0.71	1.67	0.18	0.46	1.02	0.23	0.60	1.36
DMSG	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	0.48	1.20	2.44	0.30	0.58	1.02	0.34	0.63	0.93	0.35	0.86	1.56	0.28	0.58	0.98	0.35	0.73	1.29
GSPRT	0.48	0.74	1.48	0.21	0.38	0.76	0.28	0.48	0.79	0.33	0.56	1.24	0.24	0.49	0.80	0.31	0.61	1.07
DJFR	0.33	0.71	1.72	0.19	0.38	0.78	0.25	0.44	0.79	0.22	0.50	1.12	0.20	0.38	0.76	0.24	0.45	0.96
PacNet	0.40	0.82	1.59	0.22	0.49	0.84	0.28	0.53	0.85	0.28	0.56	1.08	0.23	0.44	0.79	0.29	0.53	1.00
CUNet	0.47	1.06	2.34	0.33	0.63	1.41	0.40	0.67	1.27	0.41	0.80	1.88	0.29	0.65	1.12	0.35	0.69	1.14
PMBAN	0.28	0.55	1.11	0.19	0.30	0.53	0.23	0.37	0.64	0.21	0.36	0.74	0.18	0.31	0.57	0.22	0.39	0.75
DKN	0.25	0.51	1.22	0.16	0.30	0.52	0.21	0.35	0.61	0.17	0.34	0.81	0.16	0.28	0.54	0.20	0.38	0.70
AHMF	0.22	0.50	1.04	0.14	0.30	0.50	0.18	0.35	0.62	0.15	0.34	0.73	0.14	0.28	0.53	0.18	0.37	0.64
CTKT	0.25	0.53	1.44	0.11	0.26	0.67	0.16	0.36	0.65	0.16	0.36	0.76	0.13	0.27	0.69	0.17	0.35	0.77
RSAG	0.13	0.23	0.88	0.09	0.14	0.50	0.15	0.20	0.57	0.10	0.19	0.58	0.12	0.17	0.42	0.13	0.18	0.52

Table 1: Quantitative comparisons (in MAD) on Middlebury dataset.

Middlebury 数据集上平均绝对视差

SGNet: Structure Guided Network via Gradient-Frequency Awareness for Depth Map Super-Resolution

Zhengxue Wang, Zhiqiang Yan*, Jian Yang*

PCA Lab, Nanjing University of Science and Technology, China
`{zxwang,yanzq,csjyang}@njust.edu.cn`

AAAI 2024

Paper report——SGNet

5.SGNet: Structure Guided Network via Gradient-Frequency Awareness for Depth Map Super-Resolution

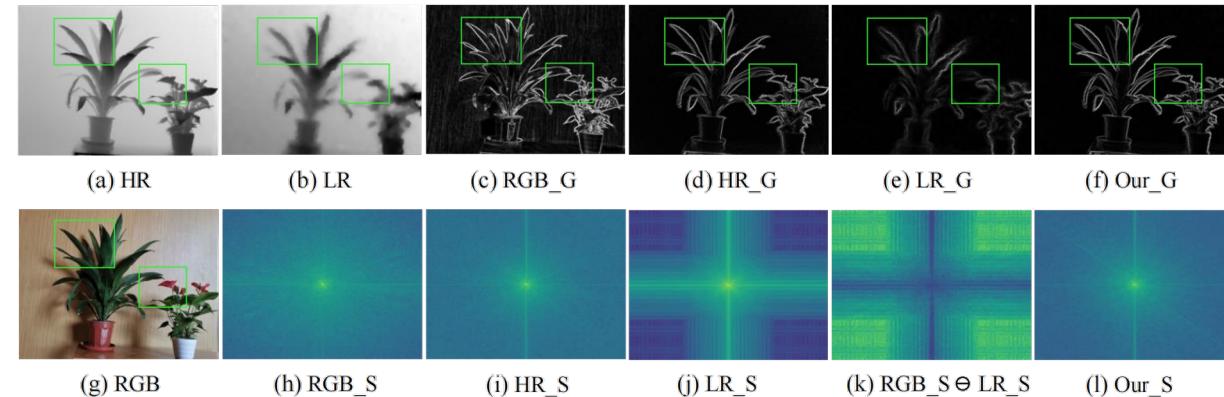
大多数 DSR 方法集中在空间域进行重建，空间域处理通常直接操作图像像素，适合进行大多数基本的图像操作和分析

那么如果摆脱这个限制呢？

Paper report——SGNet

5.SGNet: Structure Guided Network via Gradient-Frequency Awareness for Depth Map Super-Resolution

动机：大多数 SDR 方法仅关注于空间域，忽略了**梯度域**和**频域捕获**高频信息的特性



通过实验发现，梯度域和频域在表示高频结构信息方面有着天然的优势

梯度图的优势

边缘保持、减少噪声影响、信息增强

频域图的优势

频率特征分析、滤波和增强

SGNet 提出了梯度校准模块 (GCM) 和频率感知模块 (FAM)，在梯度域和频域解决 LR 深度图结构模糊问题

Paper report——SGNet

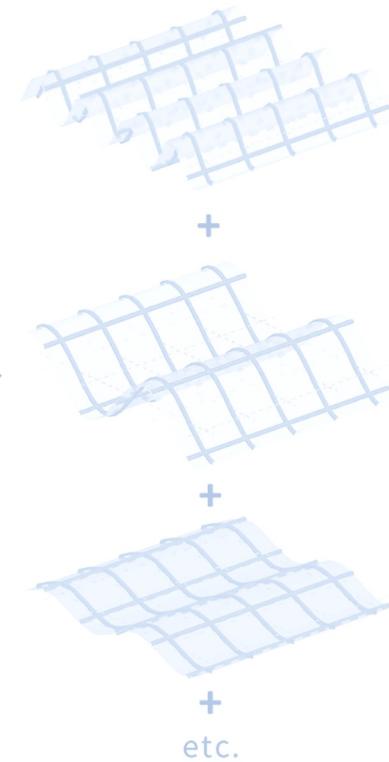
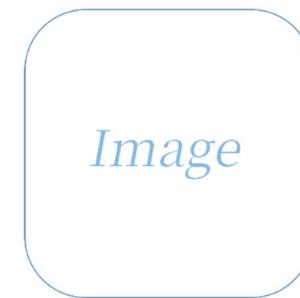
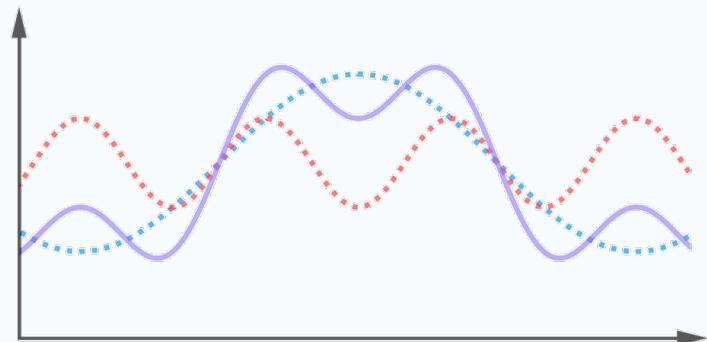
频域特征：

DFT (离散傅里叶变换)

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})}$$

$$f(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{j2\pi(\frac{ux}{M} + \frac{vy}{N})}$$

一维的傅里叶变换



Paper report——SGNet

频域特征：

每个 (u, v) 点对应的正弦波实际上是一个二维的“正弦函数面”

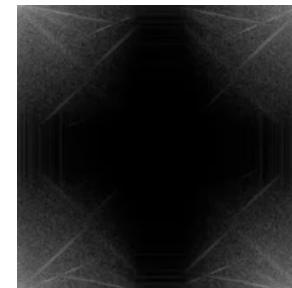
性质：中心化后的频谱图在靠近中心为体现低频特征，在四周体现为高频特征

应用：在频域中，高频分量通常对应图像中的**边缘、纹理等细节信息**，而低频分量则对应图像的整体轮廓和背景。SGNet通过提取和传播这些高频分量，来增强深度图的结构细节

Cameraman.tif courtesy of MIT and its Fourier spectrum.



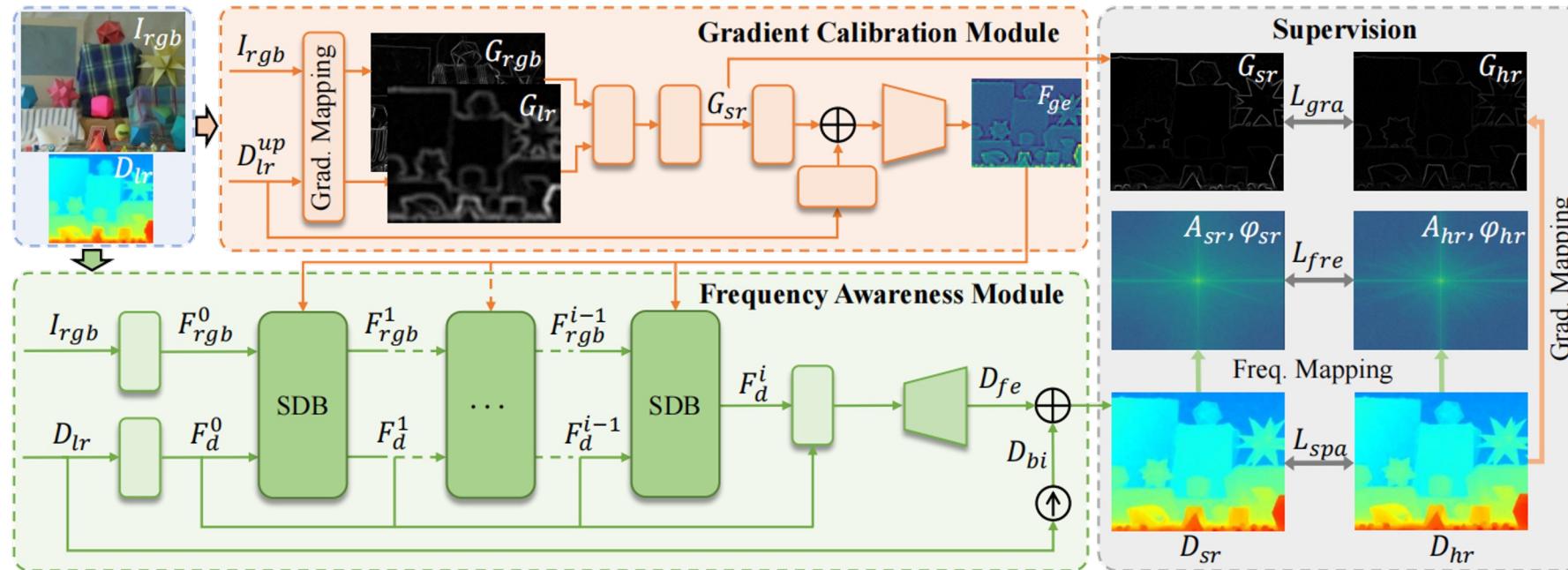
灰度图和中心化后的频谱图



中心化前的的频谱图

Paper report——SGNet

网络架构

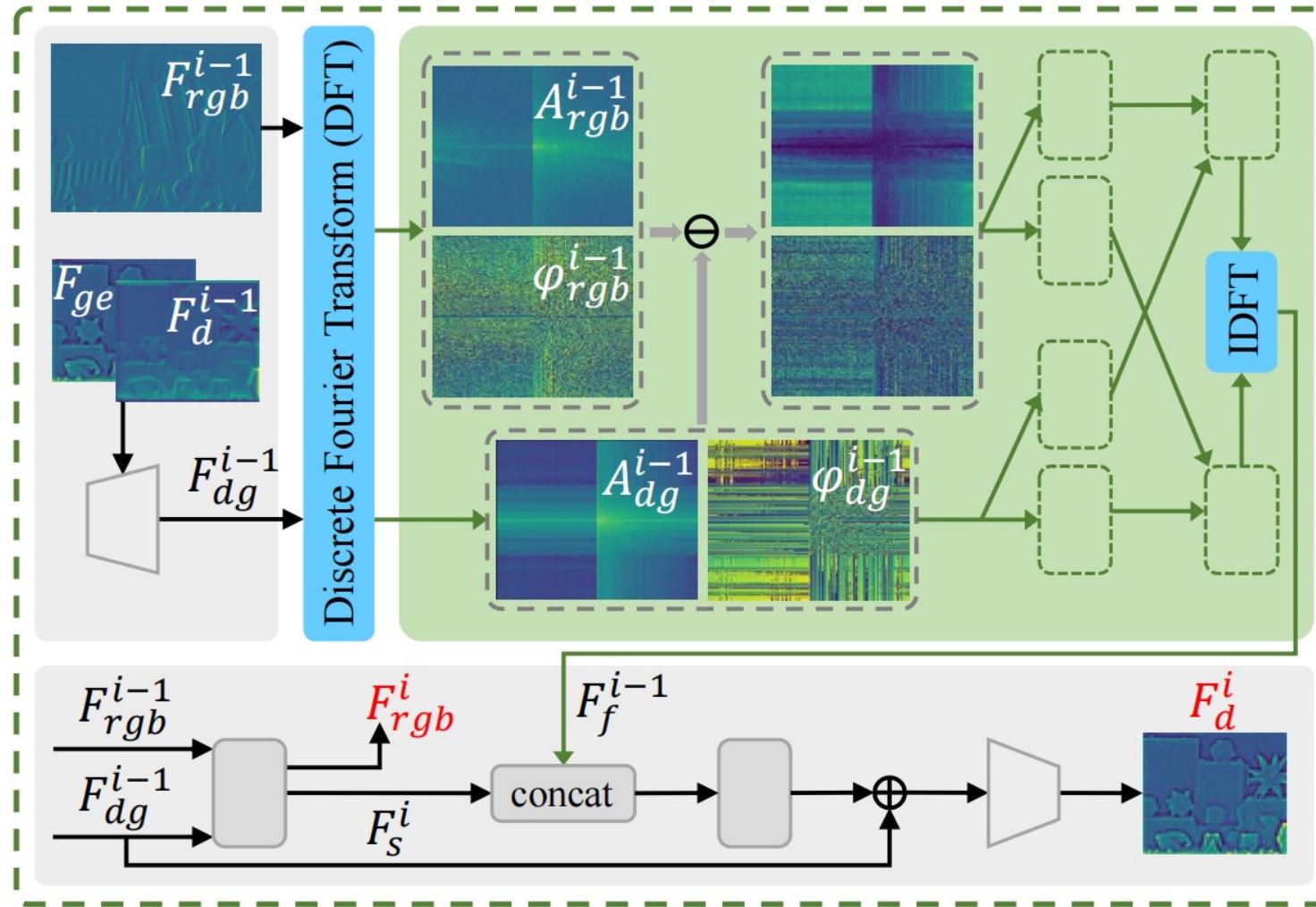


递归地进行多个频谱差分块

主要贡献：提出了梯度域和频域在深度图的一个指导作用
结构特点：分别在梯度域和频域恢复更精确的深度结构

$$f_{gm} = \|(Z_{x+1,y} - Z_{x-1,y}, Z_{x,y+1} - Z_{x,y-1})\|_2$$

Paper report——SGNet



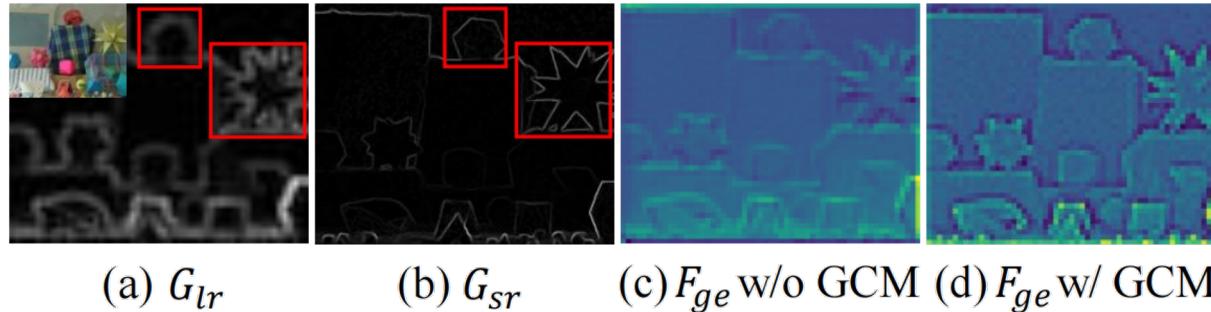
频谱分割块 (SDB) 用于增强深度特征

通过离散傅里叶变换 (DFT) 得到频域特征，使得特征在频域上进行计算，最后再还原回梯度域

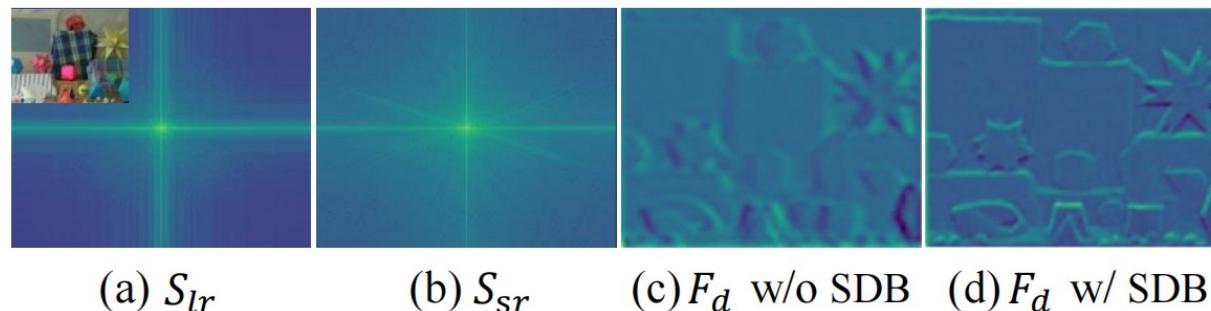
Paper report——SGNet

各个模块效果：

梯度校准
模块



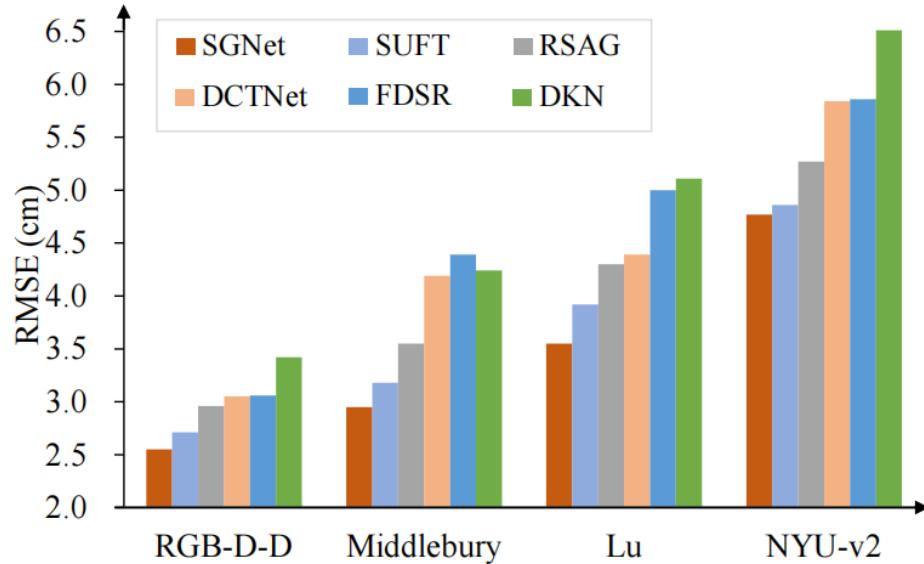
频域感知
模块



GCM成功地学习了清晰的梯度特征和深度特征

SDB成功恢复了LR频谱的高频分量和学习了清晰的结构特征

Paper report——SGNet



实验结果：

RGB-D-D 等四种指标以 RMSE (均方根误差) 为指标 , SGNet 均小于其他最先进的技术 , 重建效果最好

Scale	Bicubic	TGV	DJF	DMSG	GbFT	DKN	FDSR	CTKT	DCTNet	AHMF	RSAG	SUFT	SGNet
$\times 4$	8.16	4.98	3.54	3.02	3.35	1.62	1.61	1.49	1.59	1.40	1.23	<u>1.12</u>	1.10
$\times 8$	14.22	11.23	6.20	2.99	5.73	3.26	3.18	2.73	3.16	2.89	<u>2.51</u>	<u>2.51</u>	2.44
$\times 16$	22.32	28.13	10.21	9.17	9.01	6.51	5.86	5.11	5.84	5.64	5.27	<u>4.86</u>	4.77

Table 1: Quantitative comparison with existing state-of-the-art methods on NYU-v2 dataset.

效果比较好的超分模型的共性：

- 多尺度处理
- 引入残差网络
- 使用迭代，进行逐步恢复
- 对特定的任务执行特定的网络
- 利用了深度学习领域最新的研究进展

Thanks