RGB and LUT based Cross Attention Network for Image Enhancement

Tengfei Shi[1,3,4], Chenglizhao Chen[2*], Yuanbo He[1,4], Aimin Hao[1]

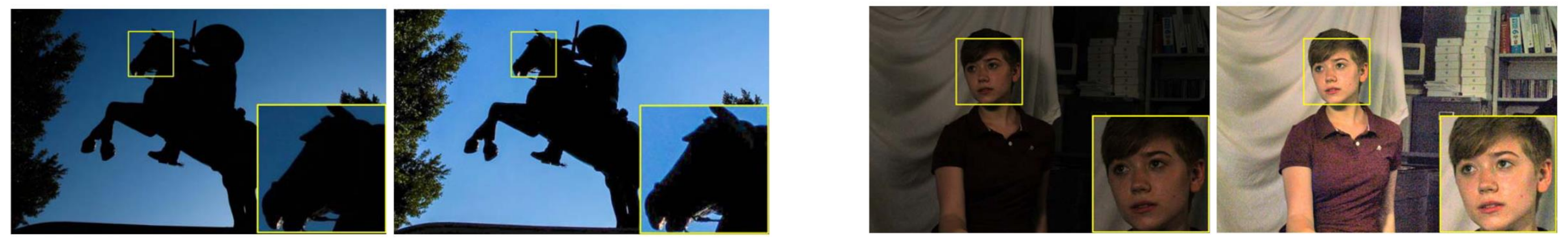
[1]Beihang University, [2]China University of Petroleum (East China),
[3]Qingdao Research Institute, [4]Peng Cheng Lab,

BMVC 2023

Problem Description

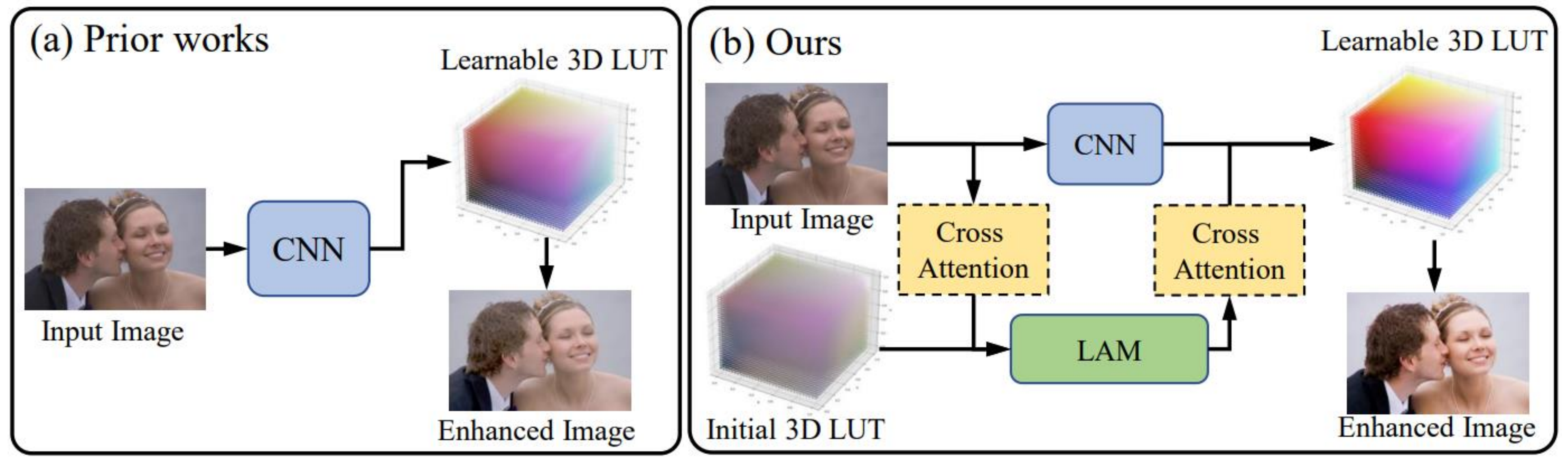
- Image enhancement aims at adapting low-light conditions and distorted colors.
- Single input might lead to image enhancement bias.

Challenge



Details Missing   Noise

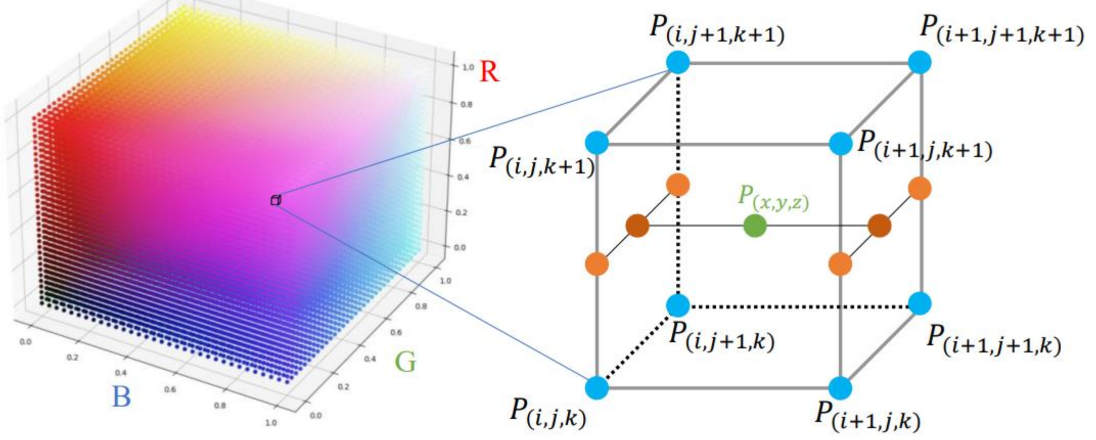
Motivation



- **(a)** 3D LUT learns the color transform through a serial schema from only single image, it is difficult to model the precise relationship between semantic and color transform
- **(b)** We take image and LUT features into consider, and adopt cross attention architecture and LUT-aware module to construct the fine-grained LUT

Method

3D LUT involves: look-up table and trilinear interpolation.



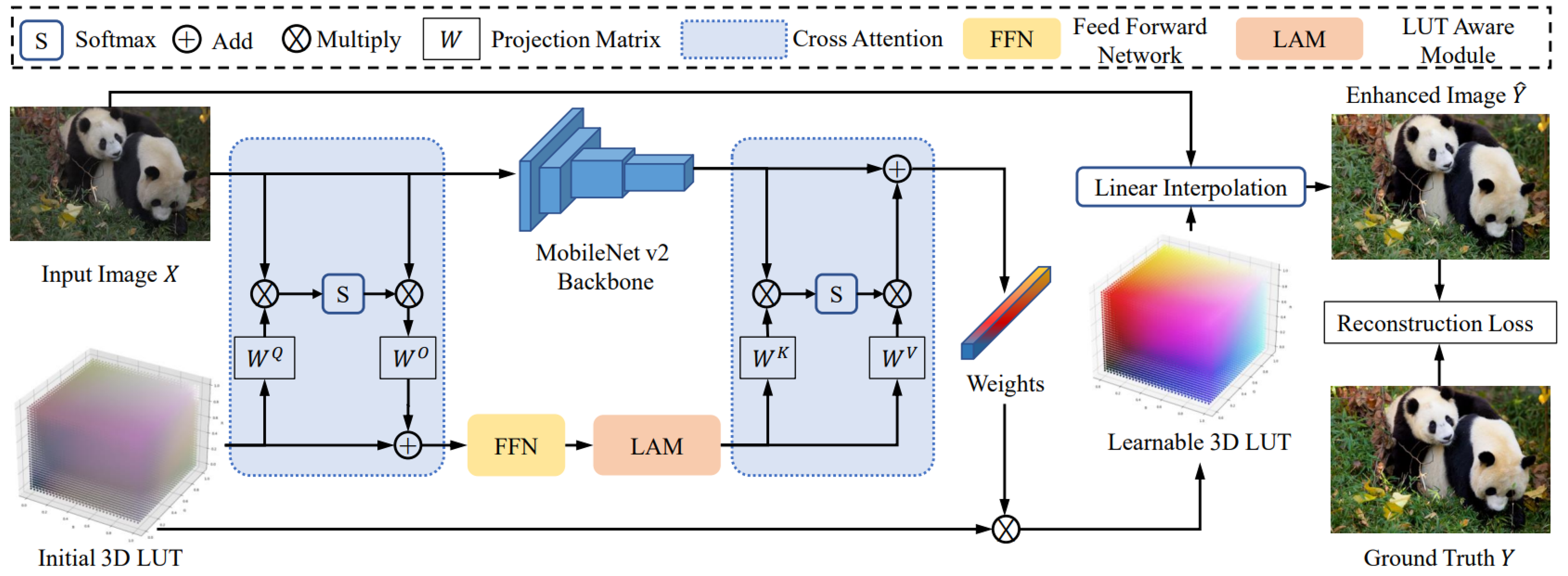
(a)   (b)

Input image I, the converted image O through the 3D LUT:

$$O_{(x,y,z)} = \psi(I^R_{(i,j,k)}, I^G_{(i,j,k)}, I^B_{(i,j,k)})$$

(i, j, k) are the pixel coordinates of the image in RGB space, and $\psi(.)$ is defined as the learnable LUT weight

Pipeline

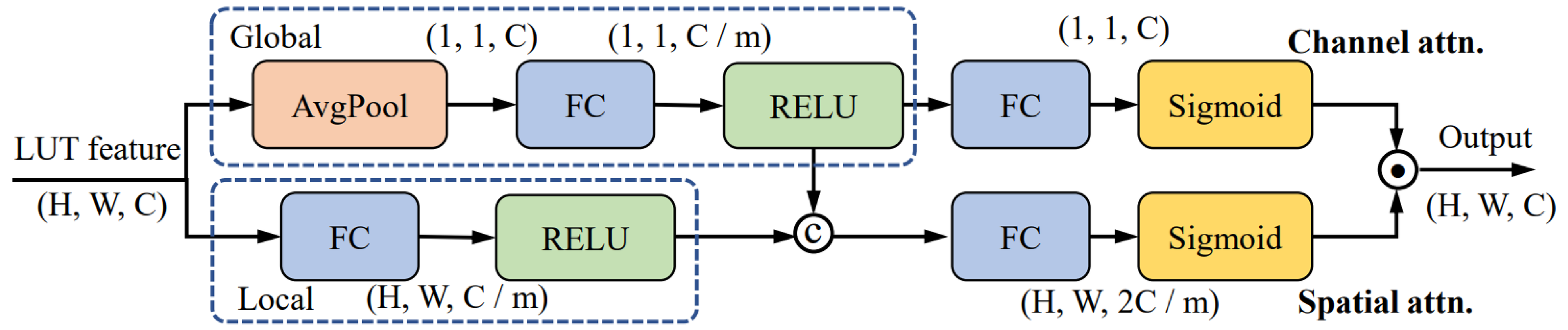


Two-way Interactive Bridge

$$image \leftarrow LUT: \quad \mathcal{A}_{X\to L} = [Attn(\tilde{l}_i W_i^Q, \tilde{x}_i, \tilde{x}_i)]_{i=1:h} W^O$$

$$image \rightarrow LUT: \quad \mathcal{A}_{L\to X} = [Attn(\tilde{x}_i, \tilde{l}_i W_i^K, \tilde{z}_i W^V)]_{i=1:h}$$

LUT Attention Module



Channel global attention and Spatial local attention

Experimental Results

Table 1: Quantitative comparisons state-of-the-art methods on the FiveK.

Method	$PSNR\uparrow$	$\Delta E_{ab}\downarrow$	$SSIM\uparrow$
Dis-Rec [16]	21.98	10.42	0.856
HDRNet [5]	24.32	8.49	0.912
DeepLPF [13]	24.73	7.99	0.916
CSRNet [7]	25.17	7.75	0.924
3D LUT [24]	25.21	7.61	0.922
STAR-DCE [26]	24.50	-	0.893
AdaInt [22]	25.28	7.48	**0.925**
CANet (Ours)	**25.49**	**7.25**	**0.925**

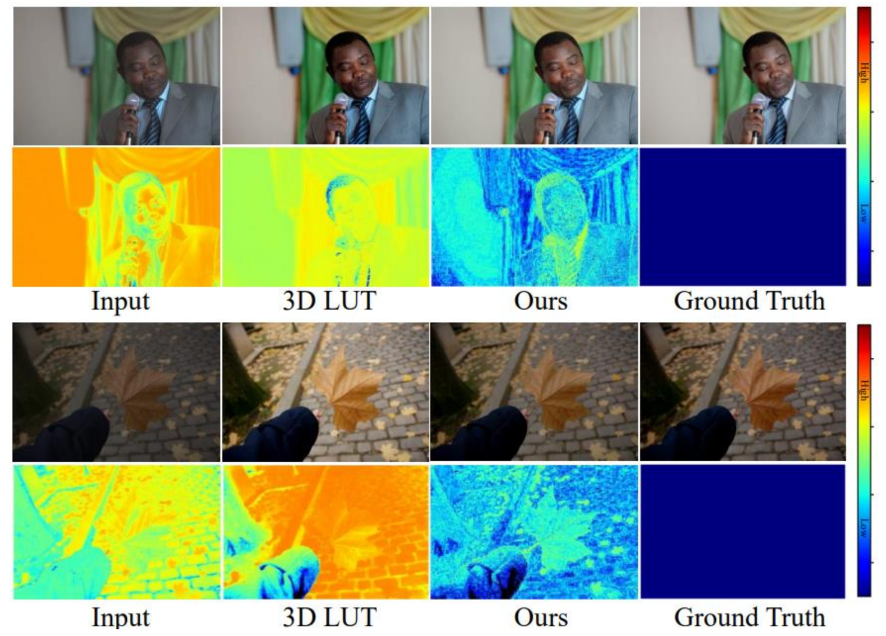
Table 2: Quantitative comparisons state-of-the-art methods on the HDR.

Method	$PSNR\uparrow$	$\Delta E_{ab}\downarrow$	$SSIM\uparrow$
Camera Raw	19.86	14.98	0.791
UPE [5]	21.21	13.05	0.816
DPE [13]	22.56	10.45	0.872
HDRNet [7]	23.04	8.97	0.879
3D LUT [24]	23.54	7.93	0.885
CANet (Ours)	**23.82**	**7.85**	**0.890**

Our proposed method outperforms related image enhancement methods on **FiveK** and **HDR** datasets

Table 3: Effectiveness comparisons of lastest methods on the MIT FiveK.

Method	Params↓	PSNR↑	SSIM↑	Runtime↓
3D LUT [24]	593.5K	25.21	0.922	**1.99ms**
AdaInt [22]	619.7K	**25.49**	**0.926**	2.56ms
SepLUT [23]	**119.8K**	25.47	0.921	2.25ms
DualNet [25]	11.25M	25.42	0.917	56.12ms
4D LUT [12]	924.4K	24.96	0.924	5.75ms
FlexiCurve [11]	130K	24.74	0.920	2.82ms
CANet (Ours)	1.52M	**25.49**	0.925	8.55ms



- Our method leverages a parallel learning process requires more model parameters but remains within the **real-time requirement**
- Visual results demonstrate our CANet produces the color transformation is **closest** to the ground-truth

Contribution

- A novel CANet adapts to uses the cross attention architecture to fuse image and LUT feature in a parallelize way
- A LUT-Aware Module fuse multi-channel and spatial attention features for enhancing the color transform
- We conduct comprehensive experiments on FiveK and HDR, the results show that our model outperforms state-of-the-art methods