Lightweight and Fast Real-time Image Enhancement via Decomposition of the Spatial-aware Lookup Tables

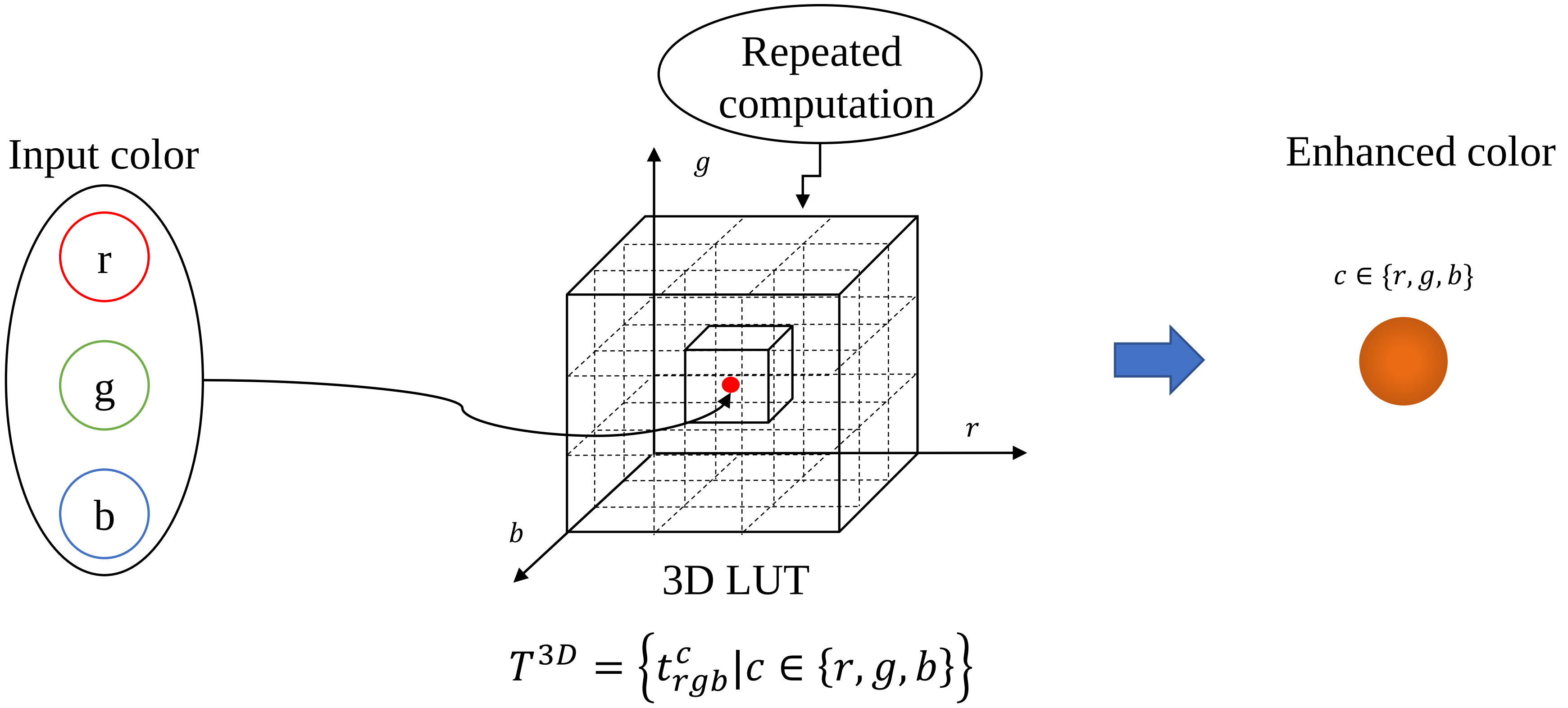
Wontae Kim[1,3], Keuntek Lee[2] and Nam Ik Cho[1,2]

[1] IPAI, Seoul National University, Seoul, Korea

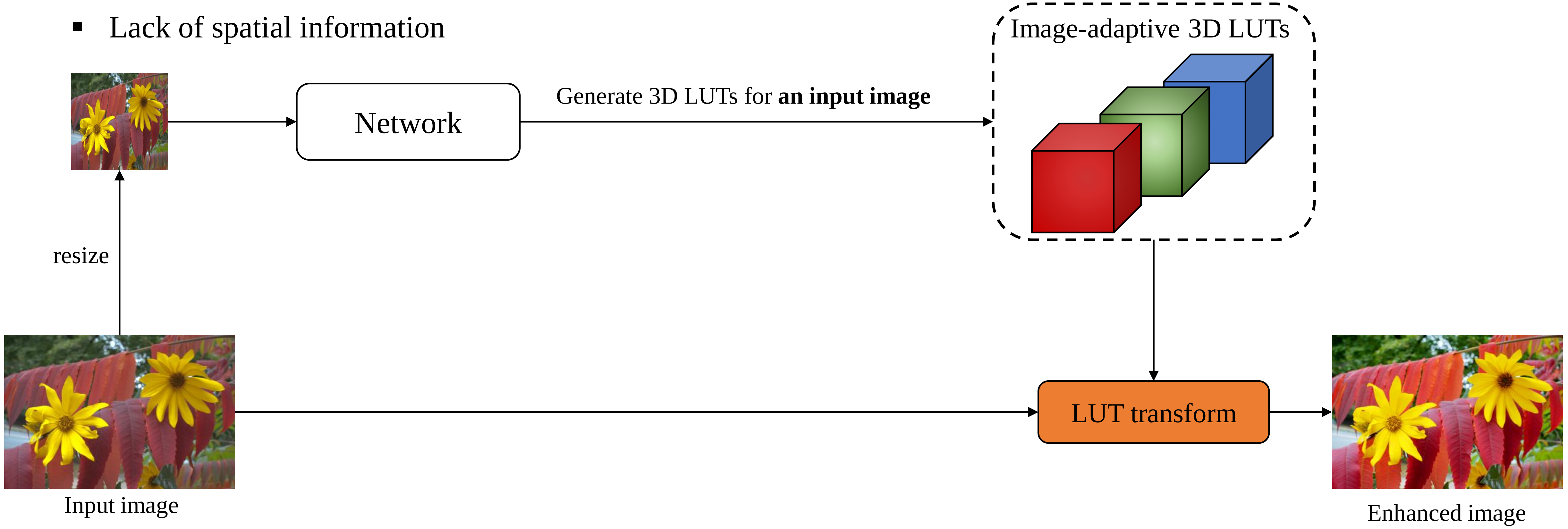
[2] Department of ECE, INMC, Seoul National University, Seoul, Korea

[3] LG Electronics

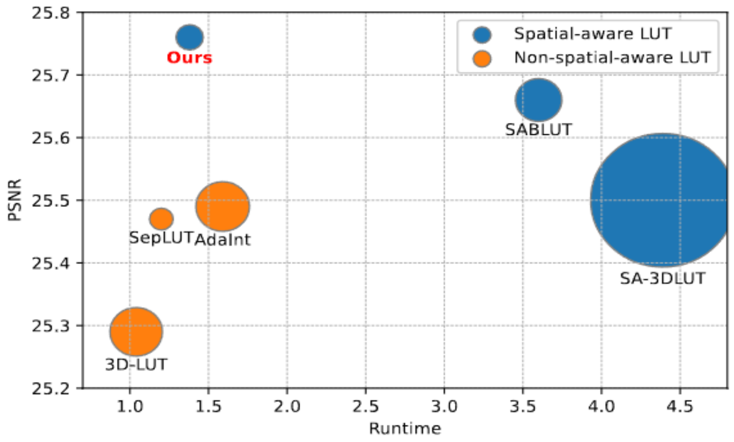
- What is 3D Lookup Table (3D LUT)?
  - A 3D LUT ($T^{3D} = \left\{t_{rgb}^{c} | c \in \{r, g, b\}\right\}$) comprises sparsely sampled input values and corresponding output values on a 3D lattice.
  - The 3D LUT transform can save computational costs and inference time by interpolating pre-calculated values.

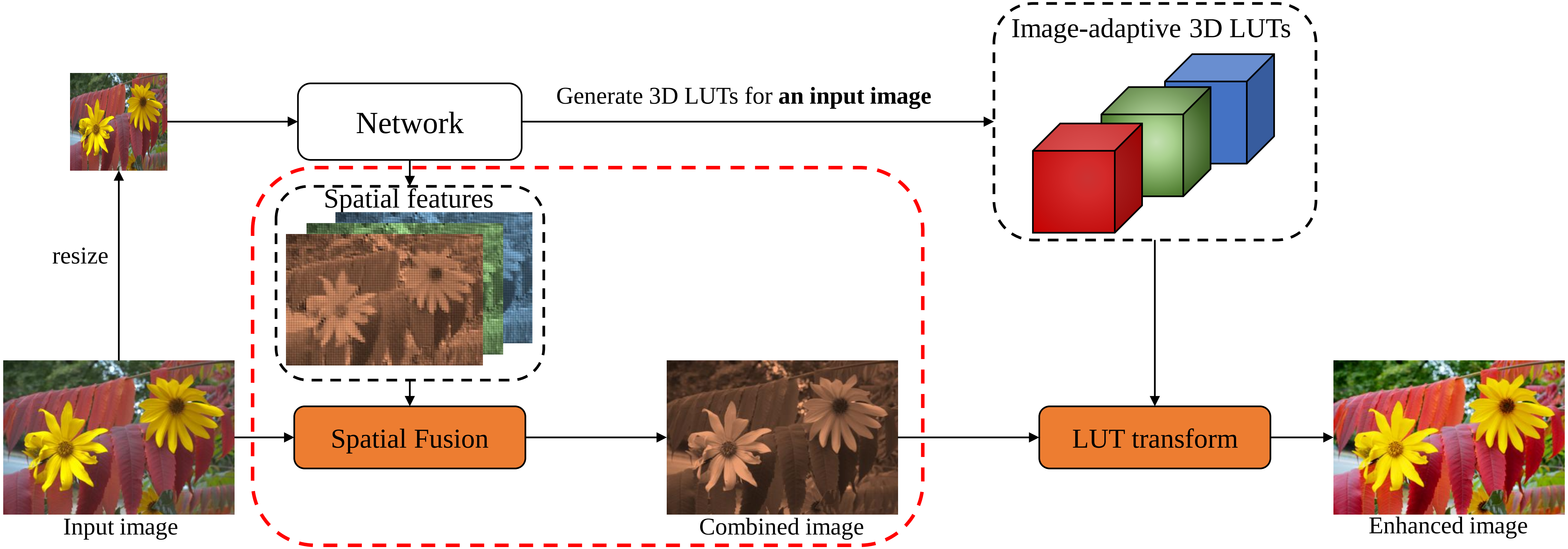


- Image-adaptive 3D LUT image enhancement methods
  - The image-adaptive 3D LUT methods achieved better performance than the fixed one by **generating 3D LUTs for each image**.
  - Benefit
    - Low computational costs
    - Short inference time
  - Problem
    - Lack of spatial information



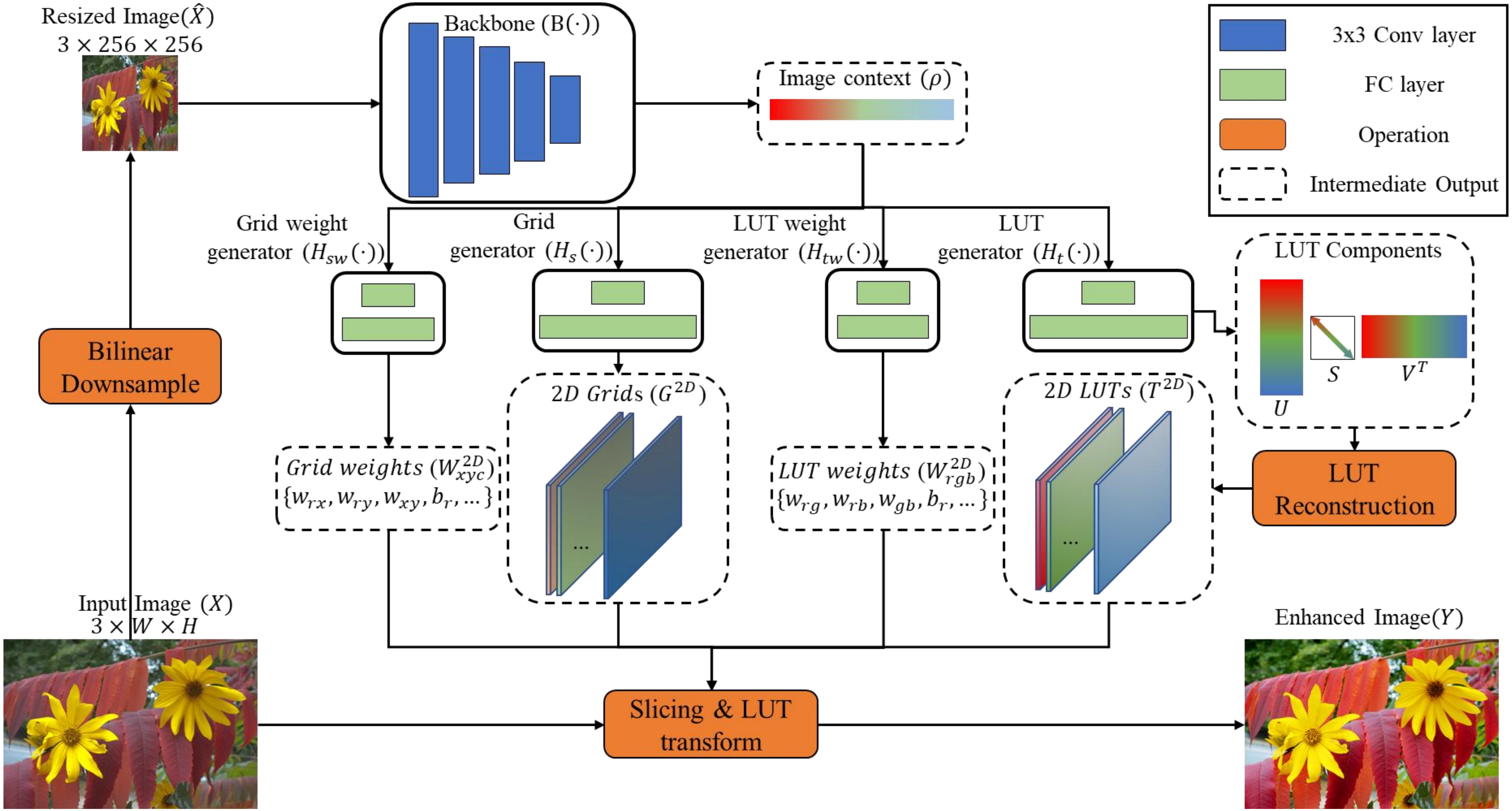
- Spatial aware image-adaptive 3D LUT image enhancement methods
  - Some spatial aware image-adaptive 3D LUT methods overcome the limitation with spatial feature fusion
  - However, additional modules introduce **a substantial number of parameters** and **long inference time**.



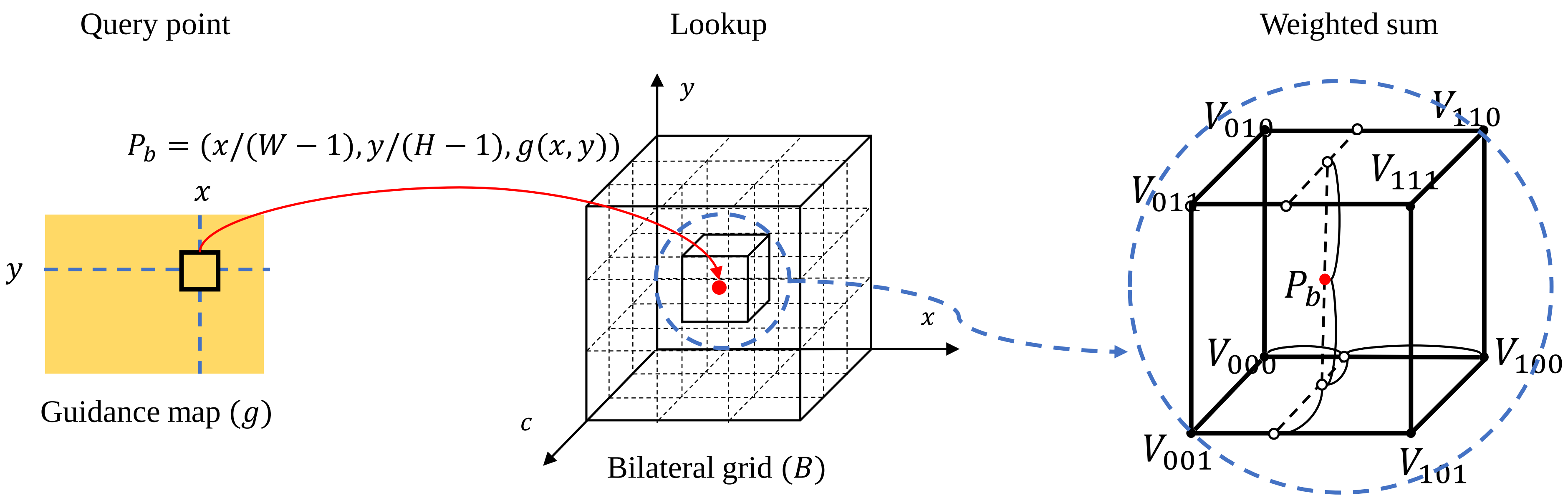


- Key idea
  - The 3D LUTs generated for each image can often be redundant through the analysis
    → We propose the decomposition techniques by **low dimensional LUTs** and **singular value decomposition(SVD)**.
  - The previous spatial-aware methods are not cache-efficient when incorporating spatial information.
    → We propose **cache-efficient spatial fusion** structure to deliver quick inference.

- Overall architecture



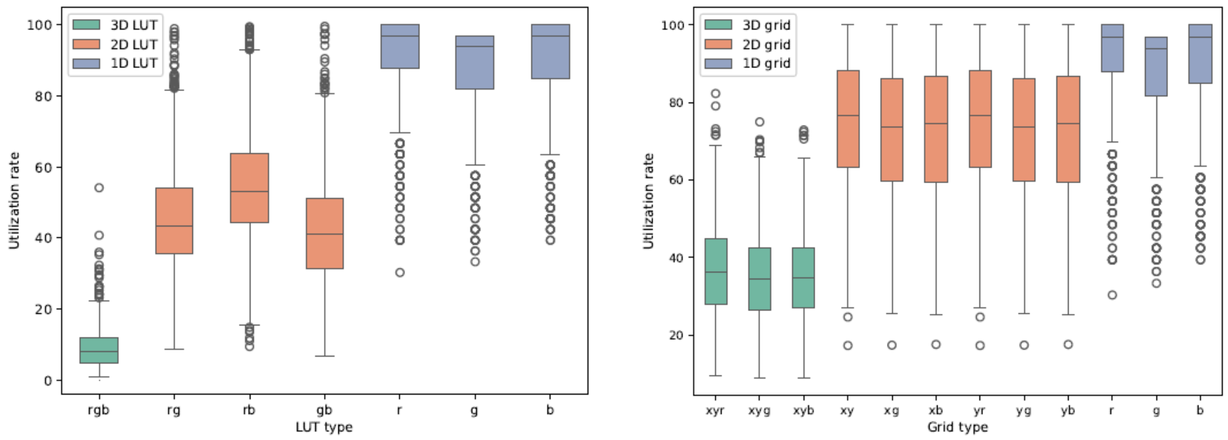
- Spatial feature fusion using bilateral grids
  - The **bilateral grid** is a data structure capable of **effectively providing spatial features.**[1]
  - We adopt the bilateral grid to incorporate spatial features, owing to its structural similarity to 3D LUTs, which **enables the application of LUT decomposition techniques**.



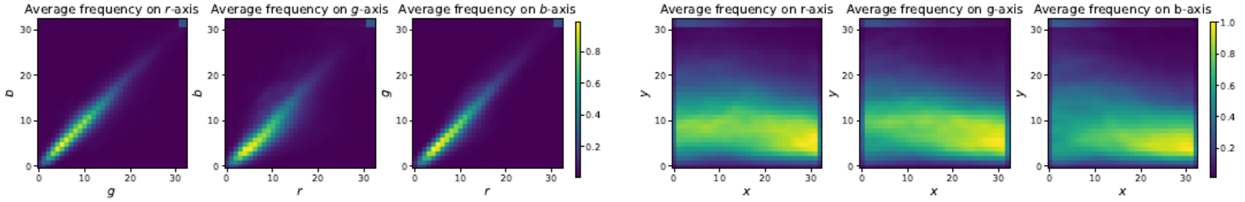
$$G^{3D} = \left\{ g^c_{xyc} | c \in \{r, g, b\} \right\}$$

[1] Kim et al. "Image-adaptive 3d lookup tables for real-time image enhancement with bilateral grids." ECCV2024

- Decomposition into lower dimension
  - The utilization rate ($\frac{\text{\#referenced vertices}}{\text{\#generated vertices}} \times 100$) show that **3D are redundant** and **1D are saturated**.
  - Most of the higher occurrence frequencies are distributed in a specific region.



(a) LUT utilization rate (b) Grid utilization rate



(c) LUT occurrence statistics (d) Grid occurrence statistics

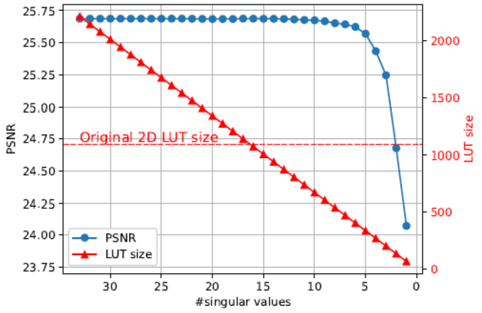
- Hypothesis
  - The combination of 2D LUTs and bilateral grids can serve as an alternative to 3D LUTs and bilateral grids.

$Slicing_{2D}^{c_k'}$	$Transform_{2D}^{c}$
$\omega_{xy}^{c_k'} \cdot I\left(X, g_{xy}^{c_k'}\right) + \omega_{xc_k'}^{c_k'} \cdot I\left(X, g_{xc_k'}^{c_k'}\right) + \omega_{xy}^{c_k'} \cdot I\left(X, g_{yc_k'}^{c_k'}\right)$	$\omega_{rg}^{c} \cdot I\left(\bar{X}, t_{rg}^{c}\right) + \omega_{rb}^{c} \cdot I\left(\bar{X}, t_{rb}^{c}\right) + \omega_{gb}^{c} \cdot I\left(\bar{X}, t_{gb}^{c}\right)$

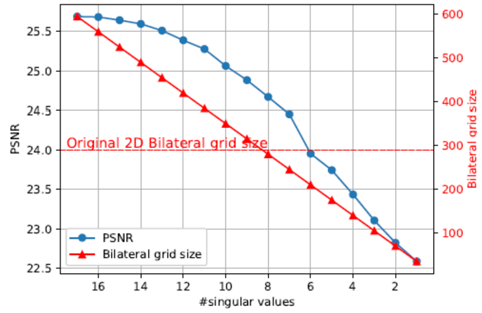
- Verification
  - We assess the PSNR(model size) under different dimensions

		Bilateral grid		
		3D	2D	1D
LUT	3D	25.68 (1.3M)	25.67 (1.1M)	25.54 (1.0M)
	2D	25.67 (421.5K)	25.68 (205.3K)	25.53 (161.2k)
	1D	25.37 (335.9K)	25.53 (119.8K)	25.22 (75.7K)

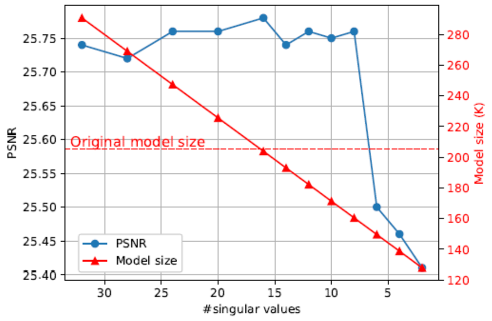
- Example of spatial feature fusion
  - The LUT maintains its performance with up to eight singular values.
  - The bilateral grid experiences a performance drop from the beginning.
  - We decide to **apply SVD to the LUT** but **not to the bilateral grid**.



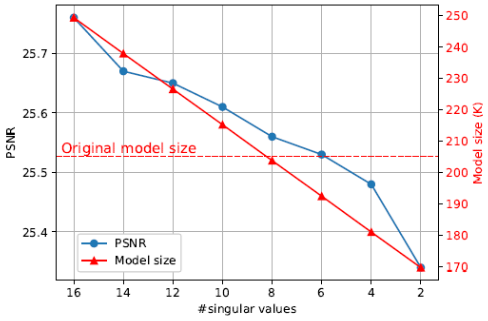
(a) Toy experiment for LUT



(b) Toy experiment for Grid

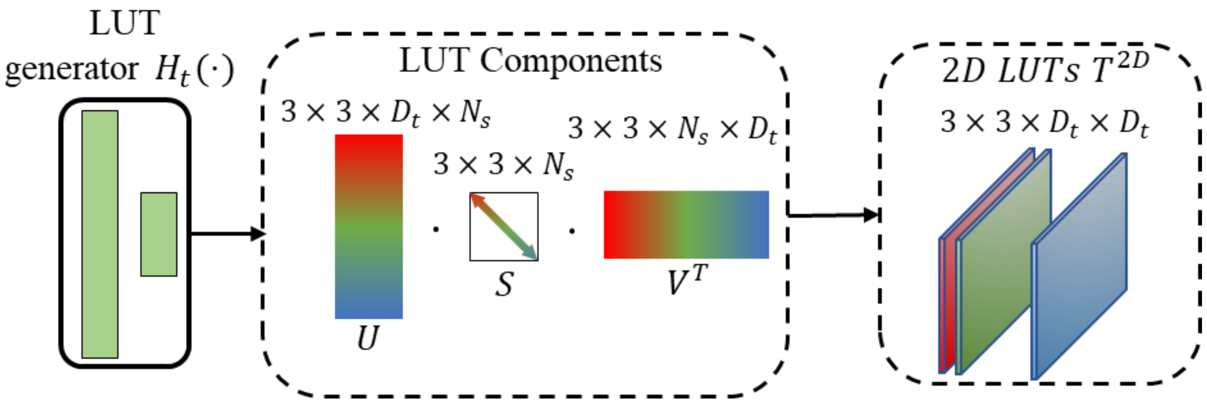


(c) Trained model for LUT

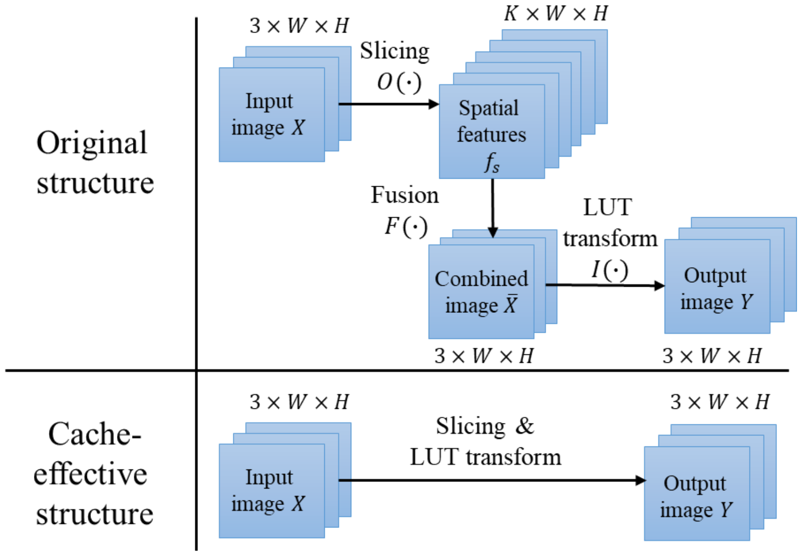


(d) Trained model for Grid

$$U \cdot S \cdot V^T = H_t(\rho) \qquad T^{2D} = U \cdot S \cdot V^T$$



- Cache-Effective Spatial Feature Fusion
  - Previous spatial feature fusion read high-resolution input and write high-resolution intermediate outputs.
  - We conduct slicing and LUT transformation at the same time.



$$Y_{(c,x,y)} = Transform_{2D}^{c}(X_{(c,x,y)}, T^{2D}) + \sum_{K=0}^{\frac{k}{3}-1} Slicing_{2D}^{c'_k}(X_{(c,x,y)}, G^{2D})$$

- # Datasets

  - FiveK (expert C)
    - 4,500 training set / 500 testing set
    - Photo retouch task (8-bit sRGB → 8-bit sRGB) / Tone mapping task (16-bit XYZ → 8-bit sRGB)
  - PPR10K (expert a/b/c)
    - 8,875 training set / 2,286 testing set
    - Photo retouch task (16-bit sRGB → 8-bit sRGB)

- # Loss Function

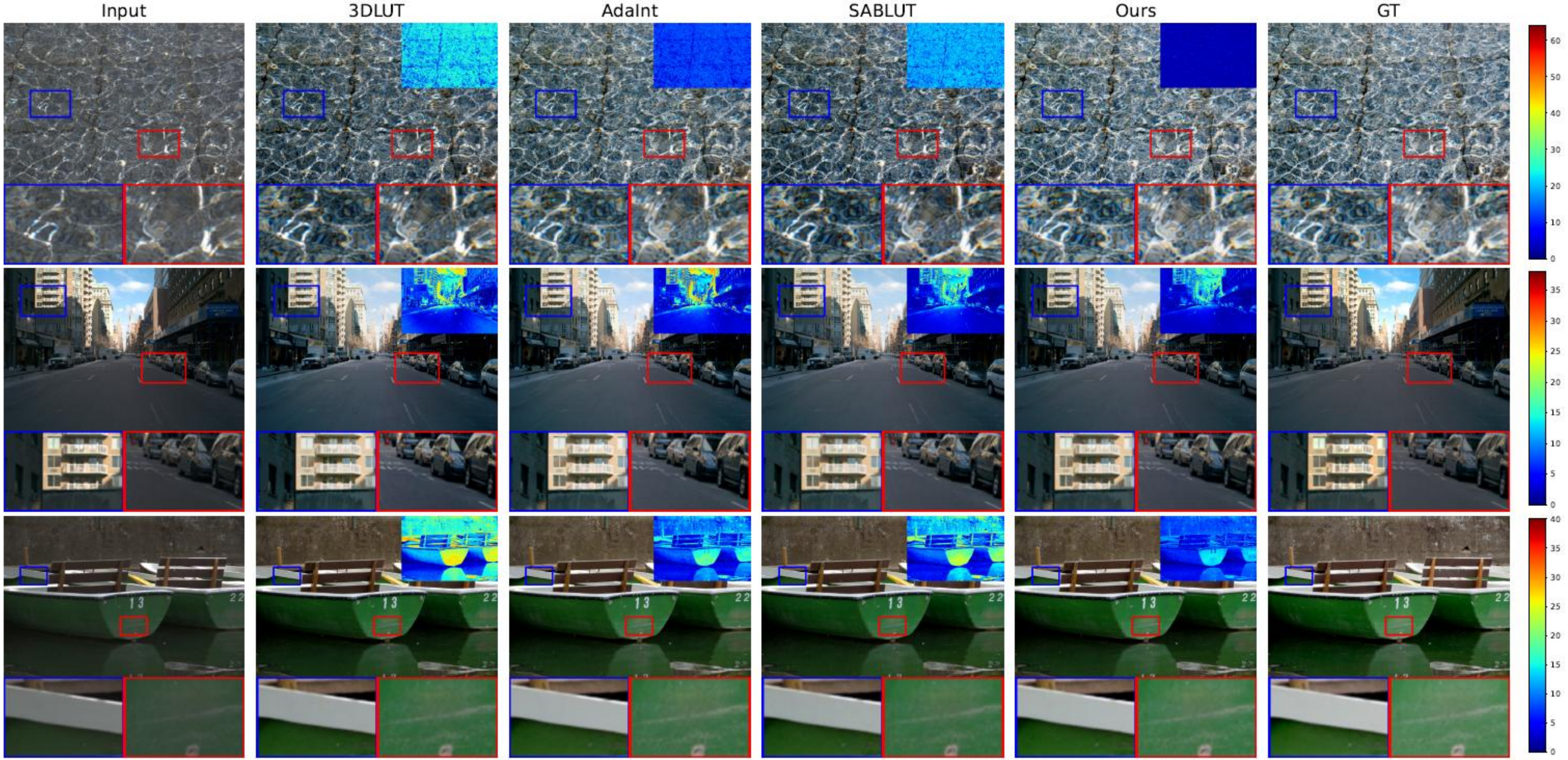
$$\mathcal{L} = \mathcal{L}_{mse} + 0.005 * \mathcal{L}_p + 0.05 * \mathcal{L}_c$$

- $\mathcal{L}_{mse}$: MSE loss
- $\mathcal{L}_p$: perceptual loss
- $\mathcal{L}_c$: color difference loss

- Quantitative comparison results of **photo retouch** on the **FiveK**
  - This results shows that our method alleviate the intrinsic problems of the spatial-aware LUT method, the tremendous model size, and the long inference time.

Method	#param	480p				Full Resolution (4K)			
		PSNR	SSIM	ΔE_{ab}	Runtime (ms)	PSNR	SSIM	ΔE_{ab}	Runtime (ms)
UPE [37]	927.1K	21.88	0.853	10.80	4.27	21.65	0.859	11.09	56.88
DPE [9]	3.4M	23.75	0.908	9.34	7.21	-	-	-	-
HDRNet [12]	483.1K	24.66	0.915	8.06	3.49	24.52	0.921	8.20	56.07
DeepLPF [33]	1.7M	24.73	0.916	7.99	32.12	-	-	-	-
CSRNet [13]	36.4K	25.19	0.925	7.76	3.09	24.82	0.924	7.94	77.10
3D LUT [46]	593.5K	25.29	0.923	7.55	1.02	25.25	0.932	7.59	1.04
SA-3DLUT* [38]	4.5M	25.50	/	/	2.27	/	/	/	4.39
SepLUT* [44]	119.8K	25.47	0.921	7.54	1.10	25.43	0.932	7.56	1.20
AdaInt [43]	619.7K	25.49	0.926	7.47	1.29	25.48	0.934	7.45	1.59
SABLUT [22]	463.7K	25.66	0.930	7.29	1.20	25.66	0.937	7.27	3.64
Ours	160.5K	25.76	0.931	7.26	1.37	25.69	0.938	7.27	1.38

- Qualitative comparisons on the **FiveK** dataset



Thank you

Github: https://github.com/WontaeaeKim/SVDLUT.git