之江实验室 ZHEJIANG LAB HONOR SmartMore

Low-Light Image Enhancement via Structure Modeling and Guidance

Xiaogang Xu[1], Ruixing Wang[2], Jiangbo Lu[3]

[1] Zhejiang Lab [2] Honor Device Co., Ltd. [3] SmartMore Corporation

Paper Tag: WED-AM-159

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Preview of This Work

ZHEJIANG LAB HONOR SmartMore

Low-Light Image Enhancement via Structure Modeling and Guidance

Xiaogang Xu, Ruixing Wang, Jiangbo Lu

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Introduction

- **We propose a new framework by conducting structure modeling and guidance simultaneously.**
  - First, a novel structure modeling method with a GAN loss
  - Second, a novel structure-guided enhancement approach
- **Our framework**
  - Consistently achieves SOTA performance on different representative benchmarks with the same structure
  - Superior perceptual quality in a large-scale user study with 100 participants

Motivation of Our Framework

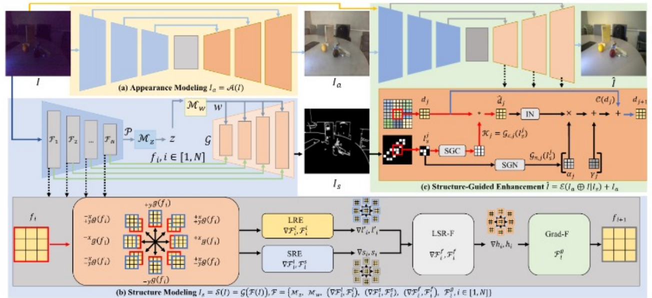
Structure modeling can be utilized to enhance the appearance predictions

- Structural information can enhance the image details
- Structural information help to distinguish different dark areas and build better relations among them

Structure of Our Framework

Appearance modeling $\mathcal{A}$, Structure Modeling $\mathcal{S}$

Structure-Guided Enhancement Module $\mathcal{E}$



Structure-Aware Feature Extractor (SAFE)

- Obtain the feature gradients

$$\{g_{+x}(f_i),\, g_{-x}(f_i),\, g_{+y}(f_i),\, g_{-y}(f_i),\, g_{+x,+y}(f_i),\, g_{+x,-y}(f_i),\, g_{-x,+y}(f_i),\, g_{-x,-y}(f_i)\}$$

- Spatial-varying feature extraction

$$l_i = \mathcal{F}_i^l(f_i), \qquad s_i = \mathcal{F}_i^s(f_i),$$
$$\bigtriangledown l_i = \bigtriangledown\mathcal{F}_i^l(\bigtriangledown g(f_i)), \quad \bigtriangledown s_i = \bigtriangledown\mathcal{F}_i^s(\bigtriangledown g(f_i)),$$
$$\bigtriangledown \in \{^{+x}, ^{-x}, _{+y}, _{-y}, ^{+x}_{+y}, ^{-x}_{+y}, ^{+x}_{-y}, ^{-x}_{-y}\}$$

- Long-short-range feature fusion

$$h_i = \mathcal{F}_i^f(l_i, s_i), \quad \bigtriangledown h_i = \bigtriangledown\mathcal{F}_i^f(\bigtriangledown l_i, \bigtriangledown s_i).$$

- Fusion from different directions

$$f_{i+1} = \mathcal{F}_i^g(h_i, {}^{+x}h_i, {}^{-x}h_i, {}_{-y}h_i, {}_{+y}h_i, {}^{+x}_{+y}h_i, {}^{-x}_{+y}h_i, {}^{+x}_{-y}h_i, {}^{-x}_{-y}h_i),$$

Structure-Guided Enhancement Module

- Overall Formulation $\hat{I} = \mathcal{E}(I_a \oplus I | I_s) + I_a$
- Structure-Guided Feature Synthesis

Loss Terms

- **Loss for appearance modeling**

$$\mathcal{L}_a = \|I_a - \bar{I}\| + \|\Phi(I_a) - \Phi(\bar{I})\|,$$

- **Loss for structure modeling**

$$\mathcal{L}_s = -[\bar{I}_s \log I_s + (1-\bar{I}_s)\log(1-I_s)],\ \bar{I}_s = C(\bar{I}),$$
$$\mathcal{L}_g = \mathbb{E}_I(\log(1+\exp(-\mathcal{D}(I_s)))),$$
$$\mathcal{L}_d = \mathbb{E}_I(\log(1+\exp(-\mathcal{D}(\bar{I}_s)))) + \mathbb{E}_I(\log(1+\exp(+\mathcal{D}(I_s)))),$$

- **Loss for SGEM** $\mathcal{L}_m = \|\hat{I} - \bar{I}\| + \|\Phi(\hat{I}) - \Phi(\bar{I})\|.$

Evaluation Results

- **Quantitative Evaluation**

Methods	SID [3]	DeepUPE [44]	KIND [69]	DeepLPF [30]	FIDE [55]	LPNet [22]	MIR-Net [67]	RF [19]	3DLUT [68]	UNIE [14]	LCDR [43]	LLFlow [49]
PSNR	13.24	13.27	14.74	14.10	16.85	17.80	20.02	14.05	17.59	20.85	18.57	19.36
SSIM	0.442	0.452	0.641	0.480	0.678	0.792	0.820	0.458	0.721	0.724	0.641	0.705

Methods	A3DLUT [46]	Band [61]	EG [13]	Retinex [24]	Sparse [62]	DSN [70]	RCTNet [17]	UTVNet [71]	SCI [28]	URetinex [54]	SNR [56]	Ours
PSNR	18.19	20.29	18.23	18.37	20.06	19.23	20.51	20.37	20.28	21.16	21.48	**24.62**
SSIM	0.745	0.831	0.617	0.723	0.815	0.736	0.831	0.834	0.752	0.840	0.849	**0.867**

Table 1. Quantitative comparison on the LOL-real dataset.

Methods	SID [3]	DeepUPE [44]	KIND [69]	DeepLPF [30]	FIDE [55]	LPNet [22]	MIR-Net [67]	RF [19]	3DLUT [68]	UNIE [14]	LCDR [43]	LLFlow [49]
PSNR	15.04	15.08	13.29	16.02	15.20	19.51	21.94	15.97	18.04	21.84	18.91	19.69
SSIM	0.610	0.623	0.578	0.587	0.612	0.846	0.876	0.632	0.800	0.884	0.825	0.871

Methods	A3DLUT [46]	Band [61]	EG [13]	Retinex [24]	Sparse [62]	DSN [70]	RCTNet [17]	UTVNet [71]	SCI [28]	URetinex [54]	SNR [56]	Ours
PSNR	18.92	23.22	16.57	16.55	22.05	21.22	22.44	21.62	22.20	22.89	24.14	**25.62**
SSIM	0.838	0.927	0.734	0.652	0.905	0.827	0.891	0.904	0.887	0.895	**0.928**	0.905

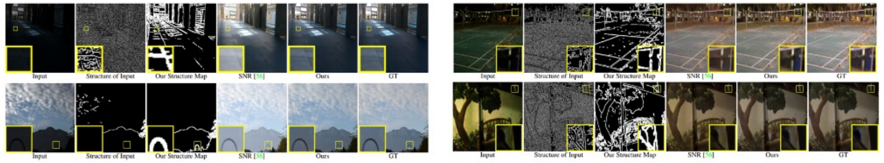
Table 2. Quantitative comparison on the LOL-synthetic dataset.

Methods	SID [3]	DeepUPE [44]	KIND [69]	DeepLPF [30]	FIDE [55]	LPNet [22]	MIR-Net [67]	RF [19]	3DLUT [68]	UNIE [14]	LCDR [43]	LLFlow [49]
PSNR	16.97	17.01	18.02	18.07	18.34	20.08	20.84	16.44	20.11	20.67	18.55	20.33
SSIM	0.591	0.604	0.583	0.600	0.578	0.598	0.605	0.596	0.592	0.602	0.587	0.611

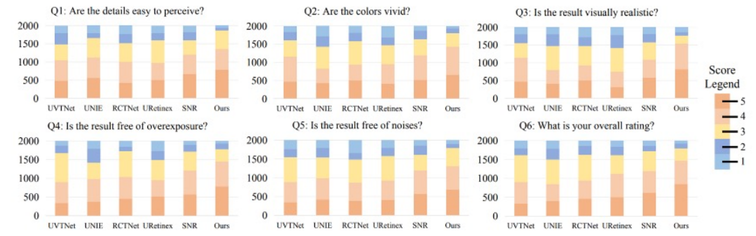
Methods	A3DLUT [46]	Band [61]	EG [13]	Retinex [24]	Sparse [62]	DSN [70]	RCTNet [17]	UTVNet [71]	SCI [28]	URetinex [54]	SNR [56]	Ours
PSNR	20.32	19.02	17.23	18.44	18.68	18.85	20.34	20.93	19.09	21.56	22.87	**23.18**
SSIM	0.595	0.577	0.543	0.581	0.606	0.617	0.601	0.614	0.585	0.619	0.625	**0.664**

Table 3. Quantitative comparison on the SID dataset (sRGB domain).

- **Qualitative Evaluation**



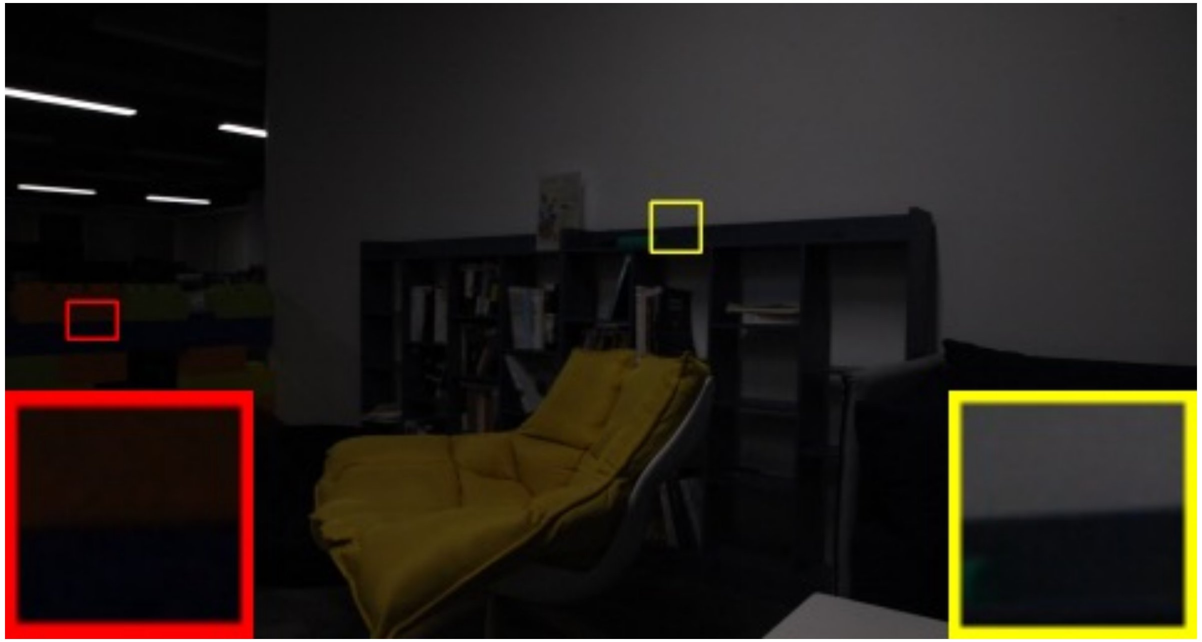
- **User Study**



Low-light Enhancement

Low-light enhancement:

- Enhance the illumination and suppress the noise
- Previous methods focus on appearance modeling



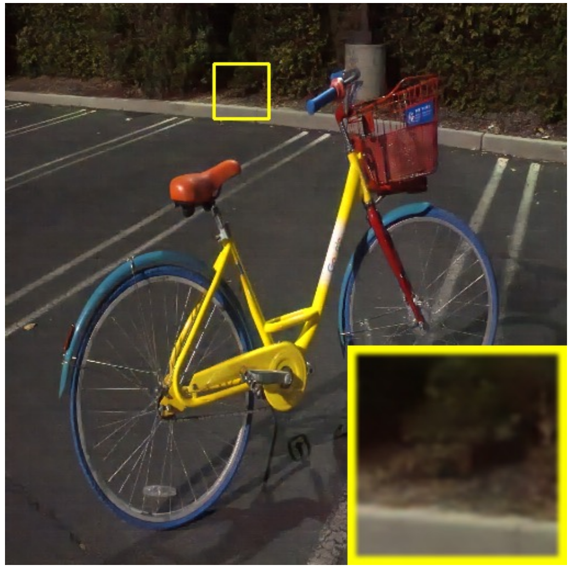
Low-light Enhancement

With only appearance modeling:

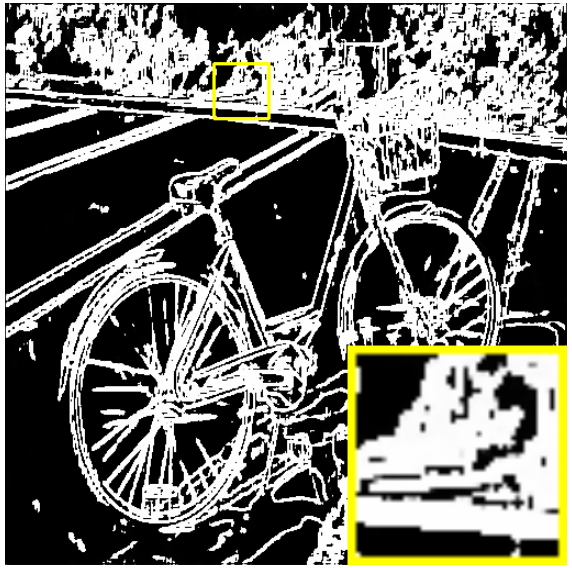
- Will result in blurry outcomes and low SSIM
- We need structure modeling

Input Image

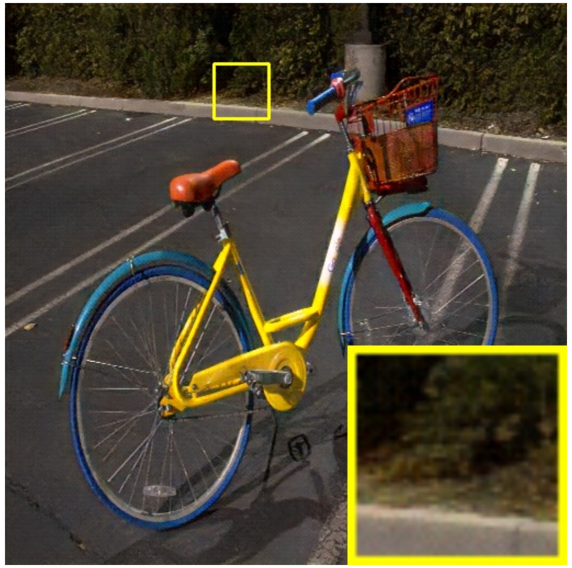
With Appearance Modeling



Structure Modeling



With Appearance & Structure Modeling

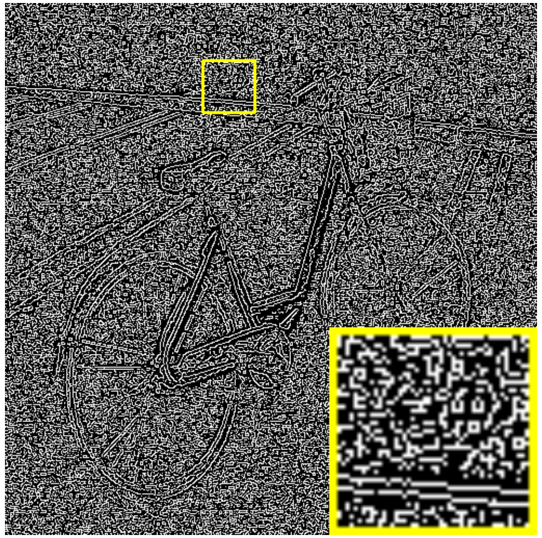


Low-light Enhancement

The challenges in structure modeling for low-light images:

- Highly ill-posed
- The influence of multiple degradations, e.g., noise

Edge detection approach



Noisy outcomes

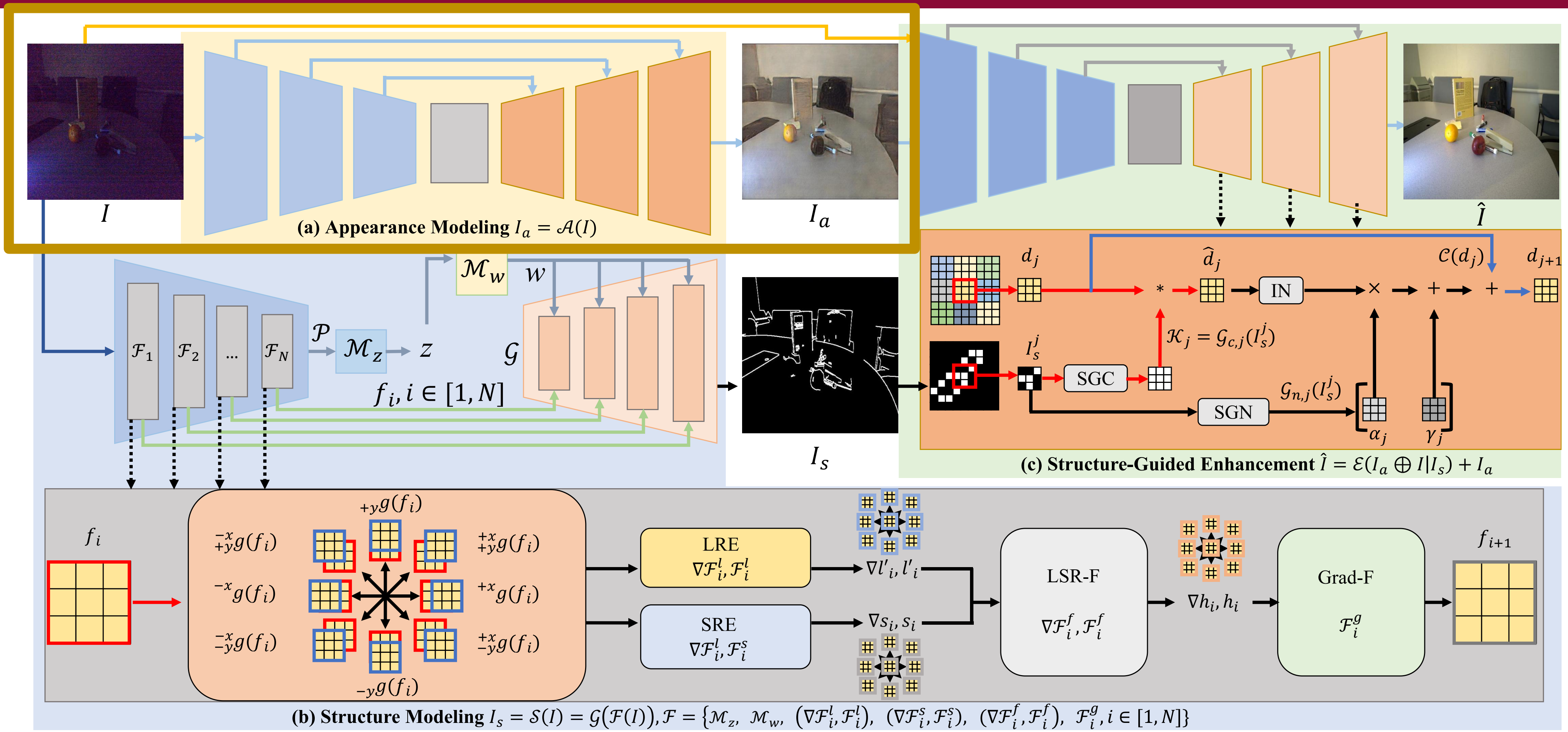
Not suitable for helping appearance modeling

Our Framework

In this paper, we:

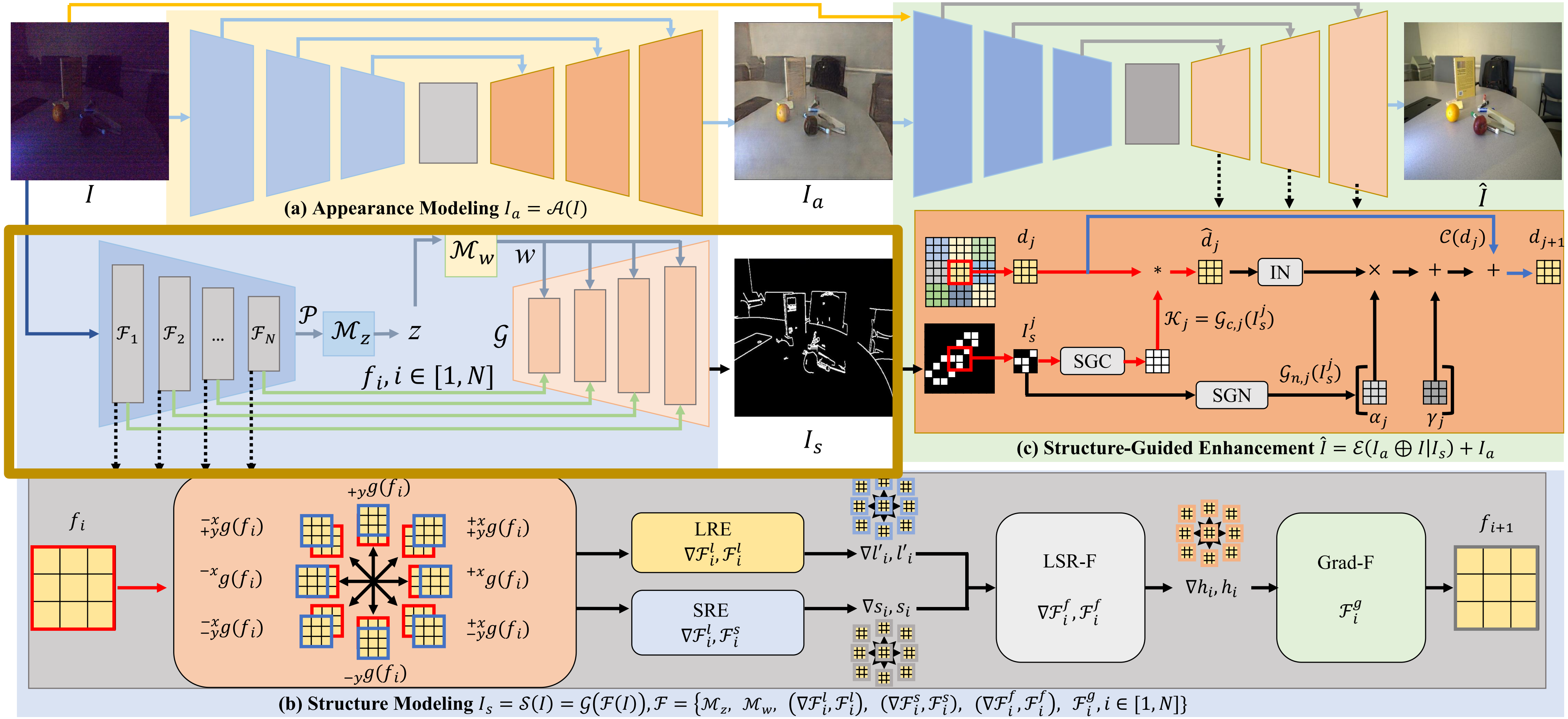
- propose a new framework for low-light enhancement by conducting **structure modeling** and **guidance** simultaneously.
- design a novel **structure modeling method**, where structure-aware features are formulated and trained with a **GAN loss**.
- formulate a novel **structure-guided enhancement approach**, for appearance improvement guided by the restored structure maps.

Our Framework



Appearance Modeling is a common U-Net

Our Framework

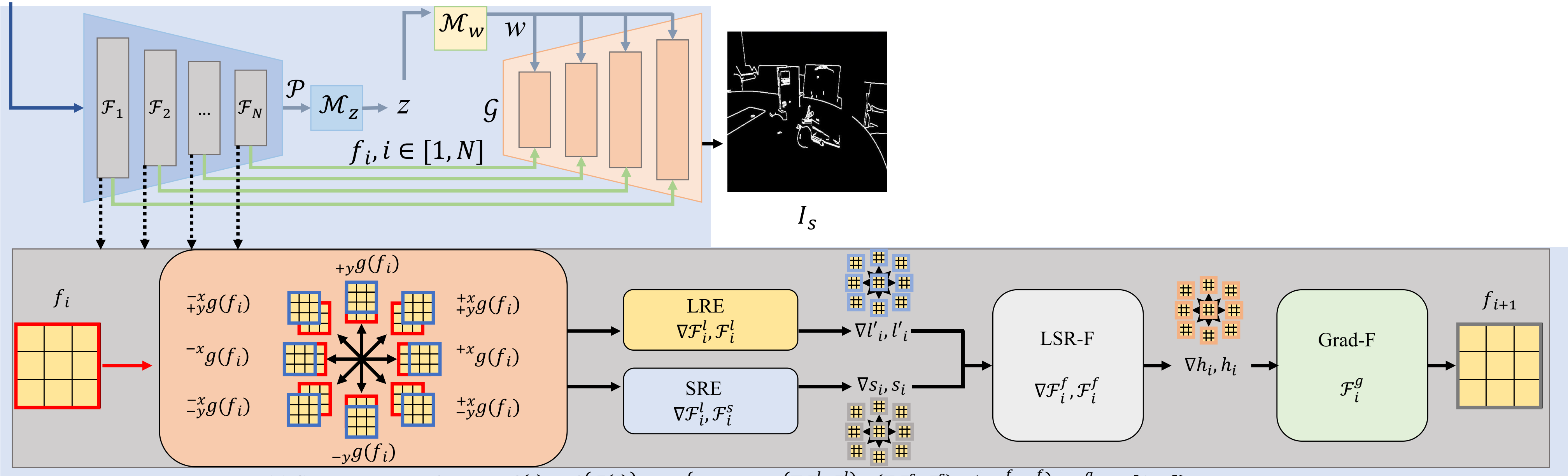


Structure Modeling is implemented with a StyleGAN backbone, $I_S = \mathcal{S}(I) = \mathcal{F}(\mathcal{G}(I))$

Structure-Aware Feature Extractor (SAFE)

Modify the encoder of StyleGAN for structure modeling:

- Compute gradient maps from features
- Spatially-varying feature extraction based on features and gradient maps



(b) Structure Modeling $I_s = \mathcal{S}(I) = \mathcal{G}(\mathcal{F}(I)), \mathcal{F} = \{\mathcal{M}_z,\ \mathcal{M}_w,\ (\nabla\mathcal{F}_i^l, \mathcal{F}_i^l),\ (\nabla\mathcal{F}_i^s, \mathcal{F}_i^s),\ (\nabla\mathcal{F}_i^f, \mathcal{F}_i^f),\ \mathcal{F}_i^g, i \in [1, N]\}$

Structure-Aware Feature Extractor (SAFE)

Modify the encoder of StyleGAN for structure modeling:

- Compute gradient maps from features
- Spatially-varying feature extraction based on features and gradient maps

(1) *Obtain the feature gradients*

$$\{g_{+x}(f_i),\, g_{-x}(f_i),\, g_{+y}(f_i),\, g_{-y}(f_i),\, g_{+x,+y}(f_i),\\ g_{+x,-y}(f_i),\, g_{-x,+y}(f_i),\, g_{-x,-y}(f_i)\}$$

(2) *Spatial-varying feature extraction*

$$\begin{aligned} l_i &= \mathcal{F}_i^l(f_i), & s_i &= \mathcal{F}_i^s(f_i), \\ \triangledown l_i &= \triangledown \mathcal{F}_i^l(\triangledown g(f_i)), & \triangledown s_i &= \triangledown \mathcal{F}_i^s(\triangledown g(f_i)), \end{aligned}$$
$$\triangledown \in \{{}^{+x}, {}^{-x}, {}_{+y}, {}_{-y}, {}^{+x}_{+y}, {}^{-x}_{+y}, {}^{+x}_{-y}, {}^{-x}_{-y}\}$$

(3) *Long-short-range feature fusion*

$$h_i = \mathcal{F}_i^f(l_i, s_i), \quad \triangledown h_i = \triangledown \mathcal{F}_i^f(\triangledown l_i, \triangledown s_i).$$

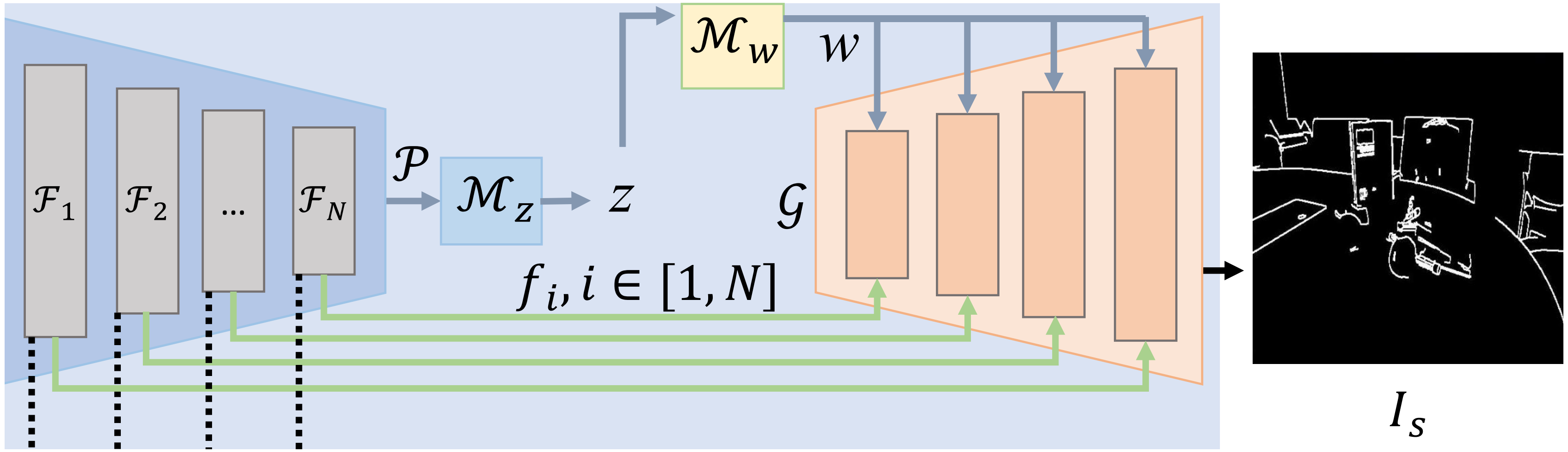
(4) *Fusion from different directions*

$$f_{i+1} = \mathcal{F}_i^g(h_i,\, {}^{+x}h_i,\, {}^{-x}h_i,\, {}_{-y}h_i,\, {}_{+y}h_i,\\ {}^{+x}_{+y}h_i,\, {}^{-x}_{+y}h_i,\, {}^{+x}_{-y}h_i,\, {}^{-x}_{-y}h_i),$$

Structure-Aware StyleGAN Generator (SAG)

Equipped with SAFE, we formulate SAG:

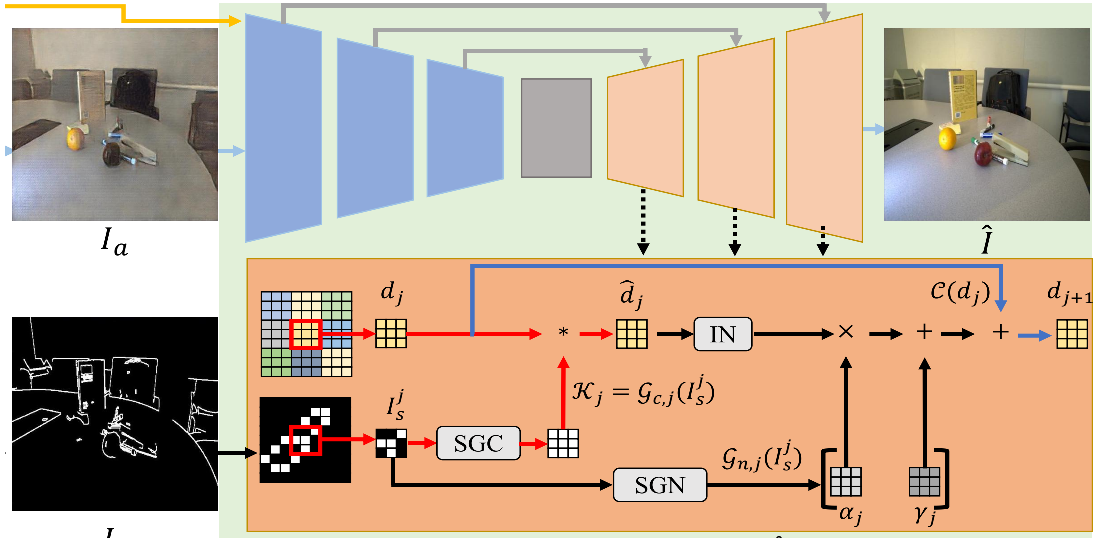
- The features from SAFE as $f_i, i \in [1, N]$
- Obtain w space of StyleGAN as $w = \mathcal{M}_w(z) = \mathcal{M}_w(\mathcal{M}_z(\mathcal{P}(f_N)))$
- Feed the structural information into the generator's different layers



Structure-Guided Enhancement Module (SGEM)

SGEM can also be implemented as a U-Net:

- We denote SGEM as $\mathcal{E}$
- The enhancement is denoted as $\hat{I} = \mathcal{E}(I_a \oplus I|I_s) + I_a$
- The structural information is inserted via Structure Guided Convolutions (SGC) and Structure Guided Normalizations (SGN)



(c) Structure-Guided Enhancement $\hat{I} = \mathcal{E}(I_a \oplus I|I_s) + I_a$

Loss Functions

Loss for appearance modeling:

- The loss is computed at both the pixel level and perceptual level

$$\mathcal{L}_a = \|I_a - \bar{I}\| + \|\Phi(I_a) - \Phi(\bar{I})\|,$$

Loss for structure modeling:

- Consists of regression loss and GAN loss
- The GT is obtained via edge detection in normal-light data

$$\mathcal{L}_s = -[\bar{I}_s \log I_s + (1 - \bar{I}_s) \log(1 - I_s)], \; \bar{I}_s = C(\bar{I}),$$

$$\begin{aligned}\mathcal{L}_g =& \mathbb{E}_I(\log(1 + \exp(-\mathcal{D}(I_s)))),\\ \mathcal{L}_d =& \mathbb{E}_I(\log(1 + \exp(-\mathcal{D}(\bar{I}_s))))+\\ & \mathbb{E}_I(\log(1 + \exp(+\mathcal{D}(I_s)))),\end{aligned}$$

Loss Functions

Loss for SGEM:

- The loss is computed at both the pixel level and perceptual level

$$\mathcal{L}_m = \|\widehat{I} - \bar{I}\| + \|\Phi(\widehat{I}) - \Phi(\bar{I})\|.$$

Overall Loss:

- The weighted sum of different loss functions

$$\mathcal{L} = \lambda_1 \mathcal{L}_a + \lambda_2 \mathcal{L}_s + \lambda_3 \mathcal{L}_g + \lambda_4 \mathcal{L}_m,$$

Experiments

Evaluation in sRGB Domain: Quantitative analysis

Methods	SID [3]	DeepUPE [44]	KIND [69]	DeepLPF [30]	FIDE [55]	LPNet [22]	MIR-Net [67]	RF [19]	3DLUT [68]	UNIE [14]	LCDR [43]	LLFlow [49]
PSNR	13.24	13.27	14.74	14.10	16.85	17.80	20.02	14.05	17.59	20.85	18.57	19.36
SSIM	0.442	0.452	0.641	0.480	0.678	0.792	0.820	0.458	0.721	0.724	0.641	0.705
Methods	A3DLUT [46]	Band [61]	EG [13]	Retinex [24]	Sparse [62]	DSN [70]	RCTNet [17]	UTVNet [71]	SCI [28]	URetinex [54]	SNR [56]	Ours
PSNR	18.19	20.29	18.23	18.37	20.06	19.23	20.51	20.37	20.28	21.16	21.48	**24.62**
SSIM	0.745	0.831	0.617	0.723	0.815	0.736	0.831	0.834	0.752	0.840	0.849	**0.867**

Table 1. Quantitative comparison on the LOL-real dataset.

Methods	SID [3]	DeepUPE [44]	KIND [69]	DeepLPF [30]	FIDE [55]	LPNet [22]	MIR-Net [67]	RF [19]	3DLUT [68]	UNIE [14]	LCDR [43]	LLFlow [49]
PSNR	15.04	15.08	13.29	16.02	15.20	19.51	21.94	15.97	18.04	21.84	18.91	19.69
SSIM	0.610	0.623	0.578	0.587	0.612	0.846	0.876	0.632	0.800	0.884	0.825	0.871
Methods	A3DLUT [46]	Band [61]	EG [13]	Retinex [24]	Sparse [62]	DSN [70]	RCTNet [17]	UTVNet [71]	SCI [28]	URetinex [54]	SNR [56]	Ours
PSNR	18.92	23.22	16.57	16.55	22.05	21.22	22.44	21.62	22.20	22.89	24.14	**25.62**
SSIM	0.838	0.927	0.734	0.652	0.905	0.827	0.891	0.904	0.887	0.895	**0.928**	0.905

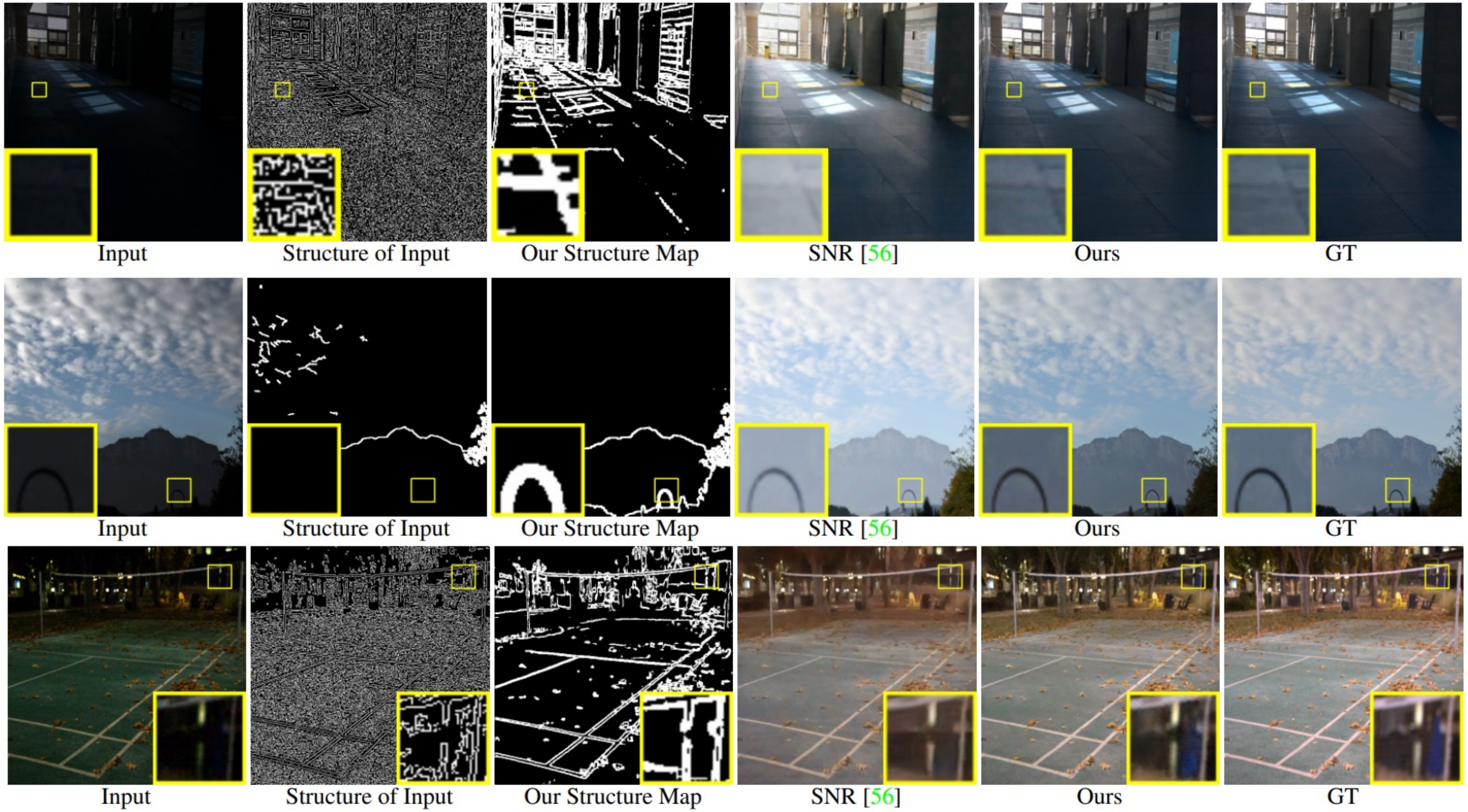
Table 2. Quantitative comparison on the LOL-synthetic dataset.

Methods	SID [3]	DeepUPE [44]	KIND [69]	DeepLPF [30]	FIDE [55]	LPNet [22]	MIR-Net [67]	RF [19]	3DLUT [68]	UNIE [14]	LCDR [43]	LLFlow [49]
PSNR	16.97	17.01	18.02	18.07	18.34	20.08	20.84	16.44	20.11	20.67	18.55	20.33
SSIM	0.591	0.604	0.583	0.600	0.578	0.598	0.605	0.596	0.592	0.602	0.587	0.611
Methods	A3DLUT [46]	Band [61]	EG [13]	Retinex [24]	Sparse [62]	DSN [70]	RCTNet [17]	UTVNet [71]	SCI [28]	URetinex [54]	SNR [56]	Ours
PSNR	20.32	19.02	17.23	18.44	18.68	18.85	20.34	20.93	19.09	21.56	22.87	**23.18**
SSIM	0.595	0.577	0.543	0.581	0.606	0.617	0.601	0.614	0.585	0.619	0.625	**0.664**

Table 3. Quantitative comparison on the SID dataset (sRGB domain).

Experiments

Evaluation in sRGB Domain: Qualitative analysis



Experiments

Evaluation in RAW Domain

Methods	DeepUPE [44]	SID [3]	EEMEFN [74]	DCE [9]
PSNR	29.13	28.88	29.60	26.53
SSIM	0.792	0.787	0.795	0.730
Methods	LLPackNet [20]	FIDE [55]	DID [29]	SGN [8]
PSNR	27.83	29.56	28.41	28.91
SSIM	0.750	0.799	0.780	0.789
Methods	RED [21]	ABF [6]	SNR [56]	Ours
PSNR	28.66	29.65	29.75	**30.17**
SSIM	0.790	0.797	0.812	**0.834**

Input Structure of Input Our Structure Map SNR [56] Ours GT

Experiments: Ablation Study

Ablation Settings

1. **“Ours w/o $\mathcal{A}$”: remove the module of $\mathcal{A}$, only input image and the structure map are set as the input of $\mathcal{E}$**
2. **“Ours w/o $\mathcal{S}$”: remove the module of $\mathcal{S}$, the structure of two concatenated networks for appearance modeling**
3. **“Ours w/o $\mathcal{F}$”: replace SAFE with traditional encoder for the StyleGAN**
4. **“Ours w/o $\mathcal{G}$”: remove the Structure-Guided Feature Synthesis in $\mathcal{E}$, set output of $\mathcal{S}$ as input of $\mathcal{E}$**
5. **“Ours w/o S.G.”: use other edge prediction network to implement $\mathcal{S}$**
6. **“Ours w/o GAN”: train $\mathcal{S}$ without GAN loss**

Experiments: Ablation Study

Results of Ablation Study

	LOL-real		LOL-synthetic		SID	
Methods	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Ours w/o $\mathcal{A}$	20.17	0.801	23.59	0.879	22.59	0.639
Ours w/o $\mathcal{S}$	18.14	0.773	21.20	0.881	20.47	0.623
Ours w/o $\mathcal{F}$	20.21	0.812	23.05	0.888	22.35	0.635
Ours w/o $\mathcal{G}$	19.39	0.784	21.71	0.868	21.15	0.629
Ours w/o S.G.	20.73	0.820	23.30	0.898	22.50	0.632
Ours w/o GAN	21.28	0.812	23.17	0.883	22.14	0.642
Results with $\mathcal{A}$	18.99	0.715	21.76	0.863	19.34	0.556
Ours with noise	24.15	0.832	24.07	0.880	22.86	0.648
Ours	**24.62**	**0.867**	**25.62**	**0.905**	**23.18**	**0.664**

Experiments: Evaluation for Structural Modeling

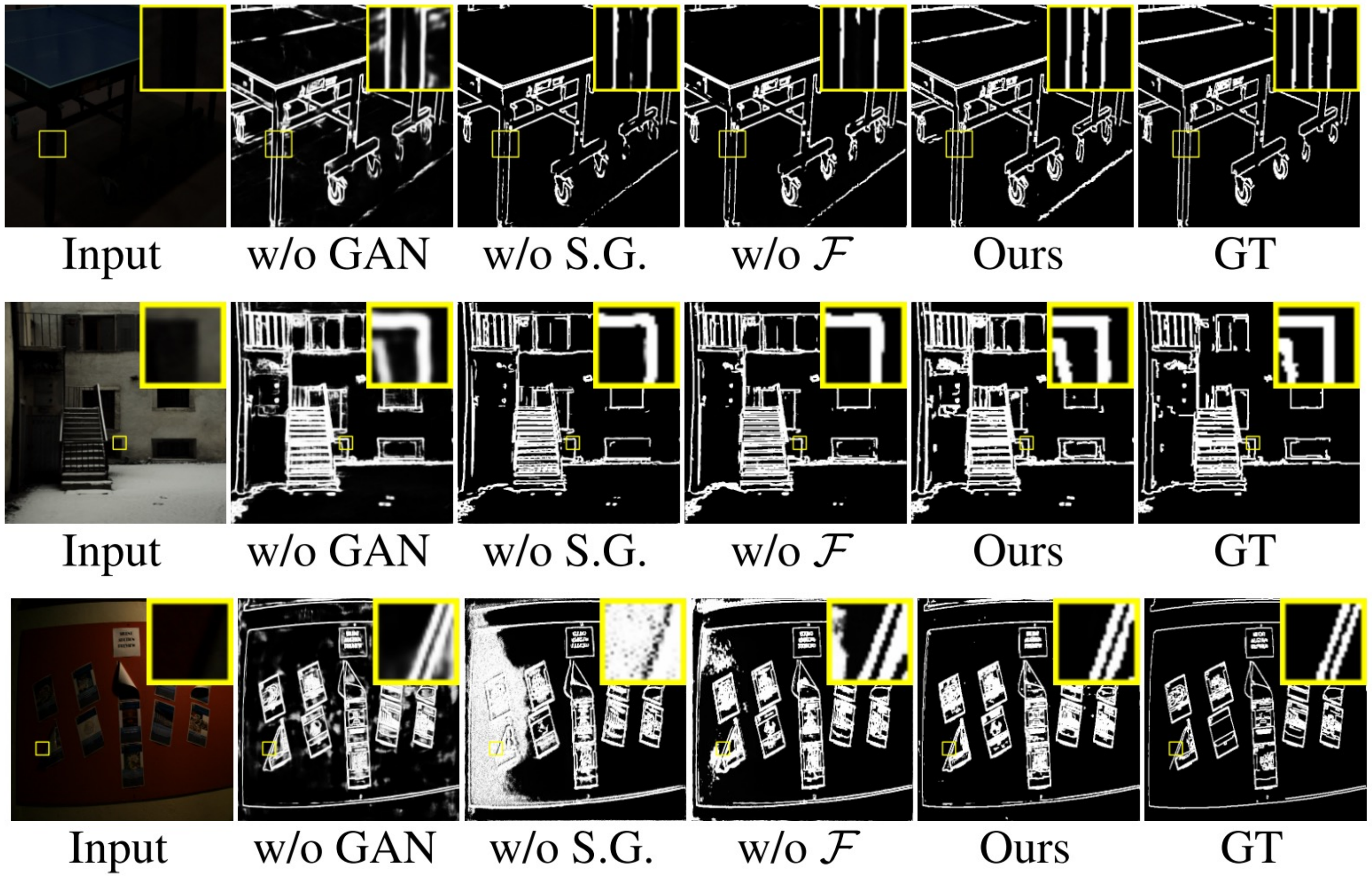
Metrics:

- The cross-entropy (CE) between the prediction and the ground truth
- The L_2 distance between the prediction and the ground truth

	LOL-real		LOL-synthetic		SID	
Methods	CE	L_2	CE	L_2	CE	L_2
Ours w/o GAN	0.2581	0.3650	0.2144	0.3936	0.5335	0.5034
Ours w/o S.G.	0.2923	0.3805	0.2133	0.3833	0.5035	0.5405
Ours w/o $\mathcal{F}$	0.3070	0.3553	0.2795	0.3675	0.5351	0.4905
Ours	**0.2130**	**0.3042**	**0.2072**	**0.3032**	**0.4352**	**0.4541**

Experiments: Evaluation for Structural Modeling

Qualitative analysis:

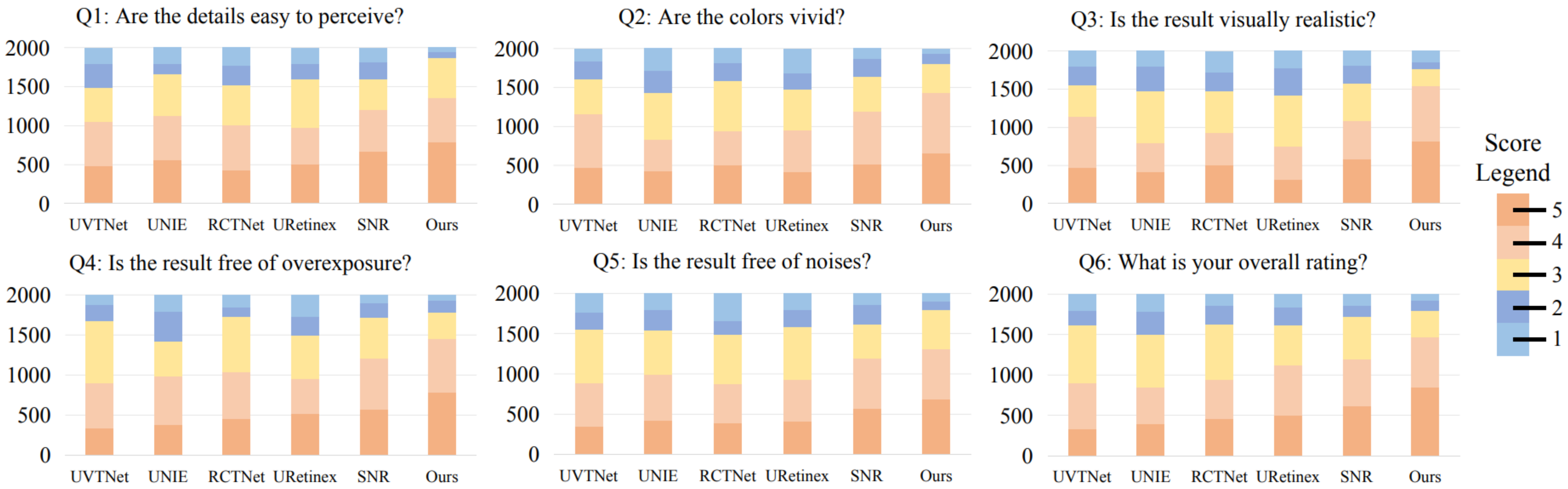


Experiments: User Study

User study from multiple dimensions with 100 participants

1. *Are the details easy to perceive?*
2. *Are the colors vivid?*
3. *Is the result visually realistic?*
4. *Is the result free of overexposure?*
5. *Is the result free of noises?*
6. *What is your overall rating?*

Experiments: User Study



Thanks