

Introduction

Goal: Restoring extremely low-light images in hope that the texts can be detected as if they are under normal light.

Integral factors for better restoring extremely low-light images :

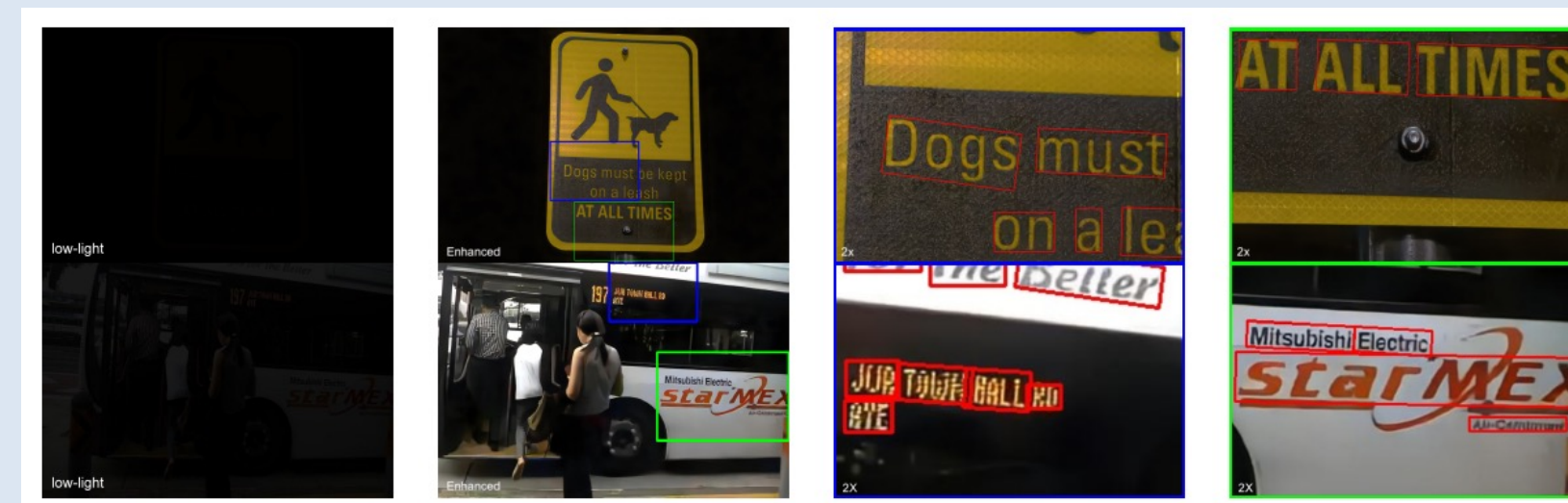
- Overall image quality
- Edge sharpness
- Text detectability

Evaluation tool:

- Image quality: PSNR & SSIM
- Text detection models: CRAFT & PAN
- Text recognition models: TRBA & ASTER

Hindrance:

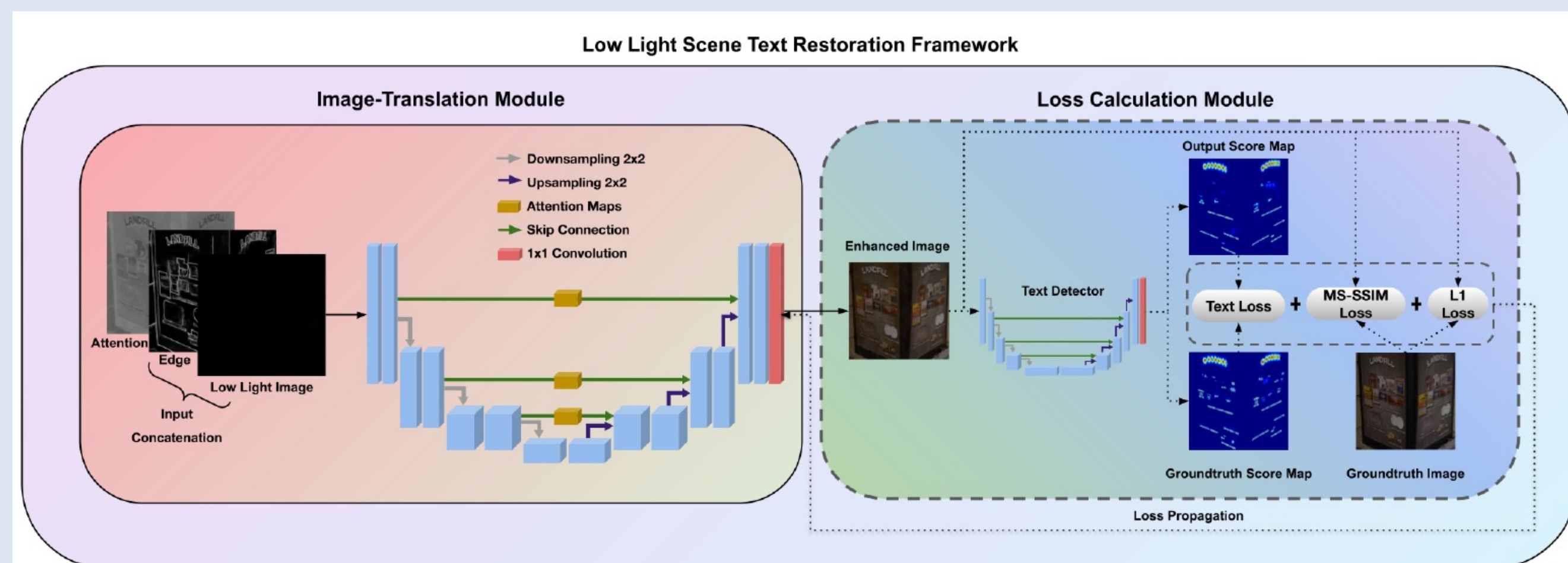
- After 24-bit quantizing, the intensity is mostly below 10
- There were no text-related datasets collected in extremely low-light
- The number of low-light images is very limited.



Extremely low-light
images

Restored images

Methods



Objective Loss Terms

- L1 loss
- SSIM Loss
- Text detection loss

SSIM loss

$$SSIM_{MS}(I', I^{GT}) = [l_M(I', I^{GT})]^\alpha \prod_{j=1}^M [c_j(I', I^{GT})]^\beta [s_j(I', I^{GT})]^\gamma,$$

Total loss

$$\mathcal{L}_{Total} = \omega_1 \mathcal{L}_{\ell_1} + \omega_2 \mathcal{L}_{SSIM_{MS}} + \omega_3 \mathcal{L}_{text}.$$

Text detection loss

$$\mathcal{L}_{text} = \|R(I') - R(I^{GT})\|_1,$$

L1 loss

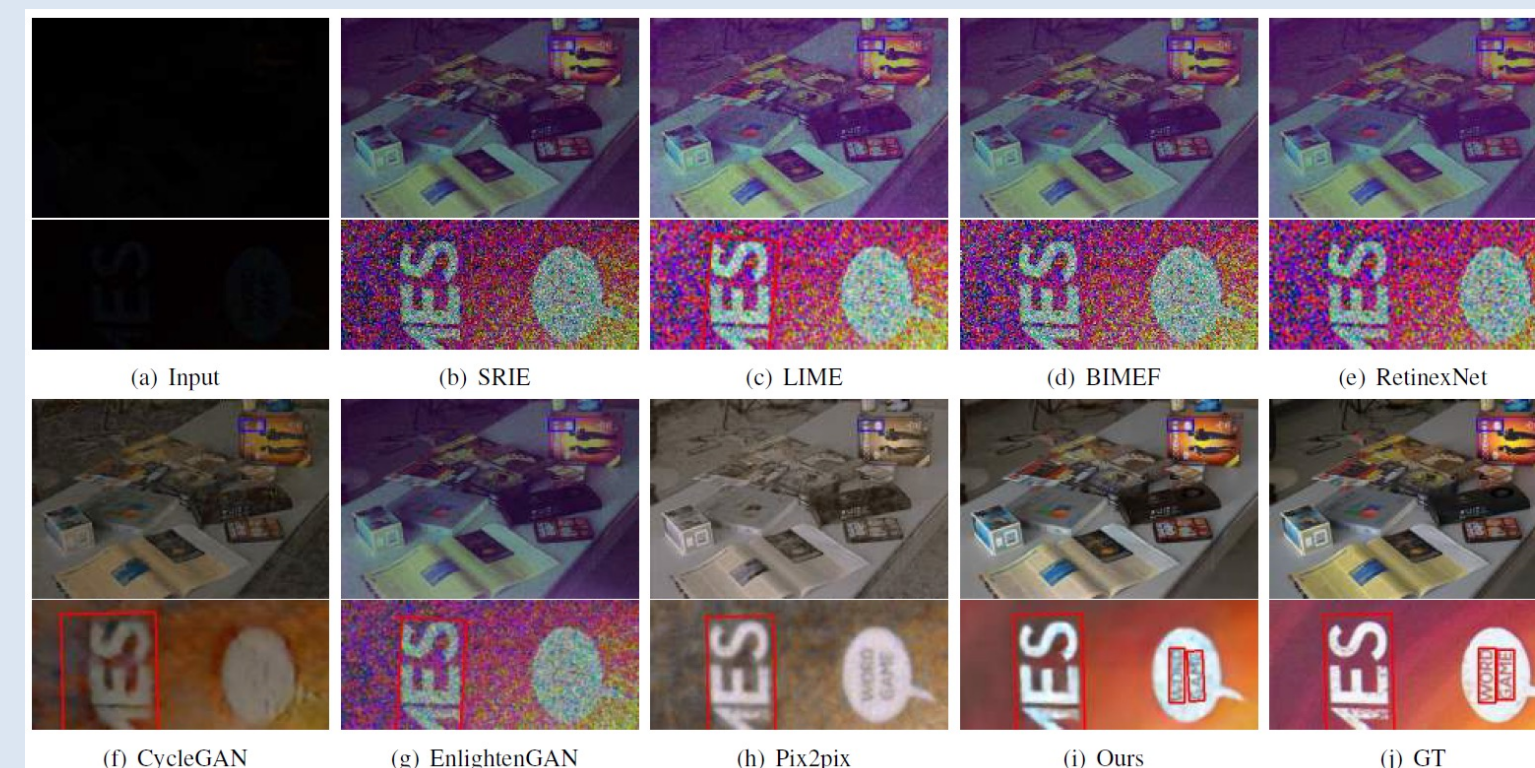
$$\mathcal{L}_{\ell_1} = \|I' - I^{GT}\|_1.$$

Results

Datasets

- SID Sony: 2697 short-exposure images and 231 long-exposure images at the resolution of 4240×2832.
- ICDAR dataset: 1500 images at the resolution of 1280×720.

Qualitative results on SID-Sony



Qualitative results on ICDAR15



Quantitative results (PSNR/SSIM)

	SRIE [22]	LIME [23]	BIMEF [24]	RetinexNet [10]	CycleGAN [11]	EnlightenGAN [12]	Pix2pix [13]	Ours
Sony	12.59/0.104	13.87/0.135	12.87/0.110	15.49/0.368	15.34/0.453	14.59/0.426	21.07/0.662	25.51/0.716
Syn. ICDAR15	10.42/0.409	12.04/0.522	16.00/0.581	15.55/0.637	23.42/0.720	21.03/0.661	24.87/0.727	28.41/0.840

Quantitative results (H-mean/Accuracy)

Model	Text Detection H-Mean				Two-Stage Text Spotting Case Insensitive Accuracy			
	Sony		Syn. ICDAR 15		Syn. ICDAR 15			
	CRAFT	PAN	CRAFT	PAN	CRAFT + TRBA	CRAFT + ASTER	PAN + TRBA	PAN + ASTER
Input	0.057	0.026	0.355	0.192	0.114	0.118	0.062	0.067
SRIE [22]	0.133	0.076	0.465	0.423	0.127	0.144	0.117	0.122
LIME [23]	0.127	0.057	0.454	0.428	0.129	0.146	0.120	0.128
BIMEF [24]	0.136	0.079	0.450	0.411	0.124	0.138	0.123	0.122
RetinexNet [10]	0.115	0.040	0.374	0.325	0.090	0.096	0.069	0.076
CycleGAN [11]	0.090	0.053	0.428	0.458	0.119	0.144	0.122	0.138
EnlightenGAN [12]	0.146	0.075	0.458	0.461	0.140	0.157	0.123	0.139
Pix2pix [13]	0.266	0.190	0.559	0.542	0.183	0.207	0.182	0.195
Ours	0.324	0.266	0.623	0.631	0.193	0.219	0.197	0.221
GT	0.842	0.661	0.800	0.830	0.526	0.584	0.555	0.591

Ablation study on SID in terms of H-mean

A: Attention map

E: Edge map

M: MS-SSIM

T: Text detection loss

baseline	w/ A	w/ A+E	w/ A+E+M	w/ A+E+M+T (full model)
0.235	0.242	0.269	0.291	0.324

baseline	w/o A	w/o E	w/o M	w/o T	full model
0.235	0.308	0.302	0.304	0.307	0.324

Quantitative Results on SID in mixed data setting:

Training data: SID-Sony-train + synthetic

ICDAR-train

Testing data: SID-Sony-test

Model	CRAFT	PAN
EnlightenGAN [12]	0.205	0.146
Pix2pix [13]	0.281	0.224
Ours	0.348	0.278