Unsupervised Low Light Image Enhancement Transformer Based on Dual Contrastive Learning

Fengji Ma[1], Jinping Sun[1]

[1] School of Electronic and Information Enginnering, Beihang University (BUAA)

BMVC 2022

Abstract

Low-light image enhancement aims to recover normal-light images from the images captured under very dim environments. While deep learning-based methods have achieved substantial success in this field, most of them require paired training data, which is difficult to be collected. We propose an Unsupervised Dual Contrastive Learning Transformer (UDCL-Transformer) where the unsupervised contrastive learning is for the first time introduced to the low light image enhancement task. From a different yet new perspective, we explore contrastive learning with an adversarial training effort to leverage unpaired low-light images and normal-light images. Our proposed method leveraged dual contrastive learning and generative adversarial networks to restore low light image. Patch-wise contrastive learning maximizes the mutual information between raw and restored images. Pixel-wise contrastive learning encourages the restored images to approach the positive samples and keep away from the negative samples in the embedding space. Generator based on Parallel Convolution Transformer (PC-Former) is proposed to capture the rich features of global and local context for better aggregate information. Extensive experiments with comparisons to recent approaches further demonstrate the superiority of our proposed method

Formulation

Patch-wise Contrastive Learning:

$$\ell(\mathbf{v}, \mathbf{v}^+, \mathbf{v}^-) = -log\left(\frac{\exp(\text{sim}(\mathbf{v}, \mathbf{v}^+)/\tau)}{\exp(\text{sim}(\mathbf{v}, \mathbf{v}^+)/\tau) + \sum_{i=1}^{N} \exp(\text{sim}(\mathbf{v}, \mathbf{v}_i^-)/\tau)}\right)$$

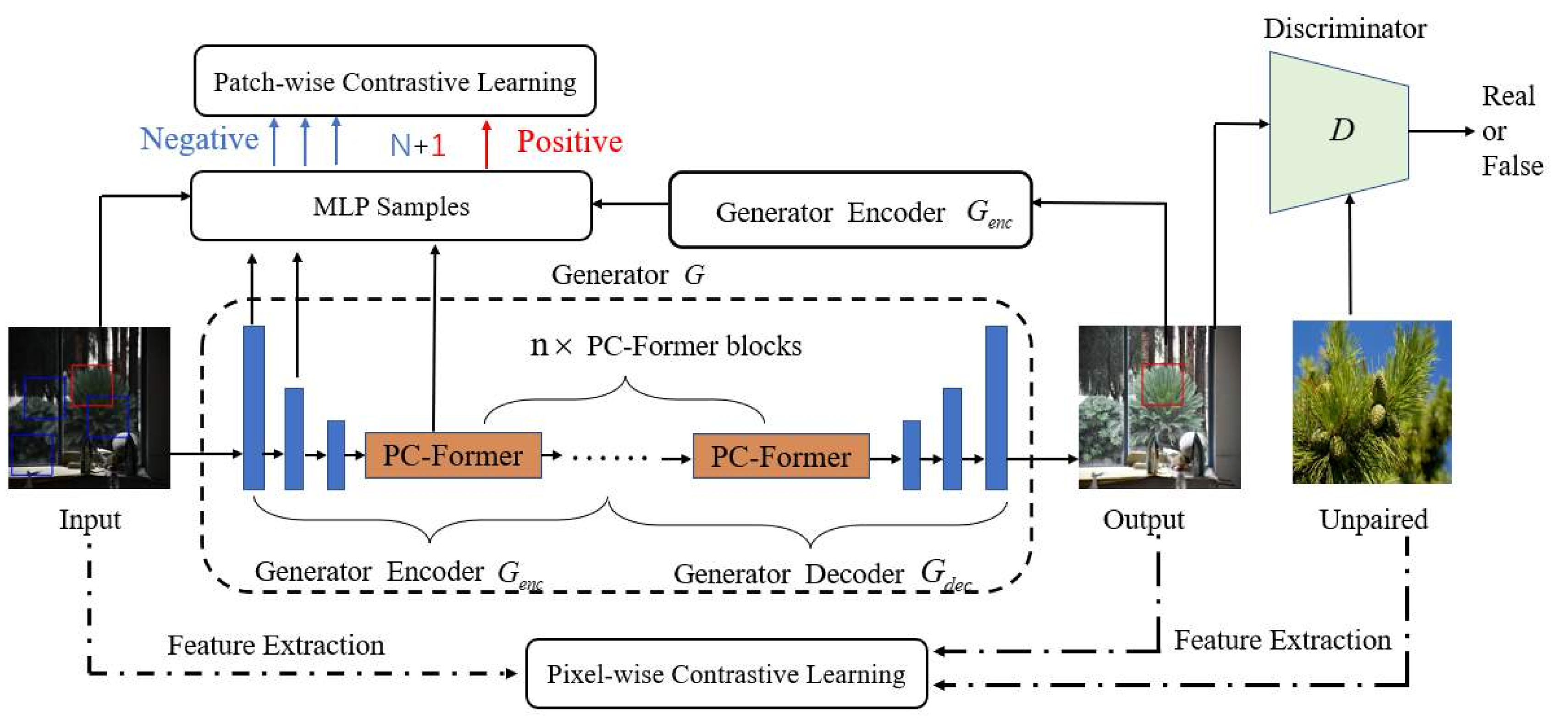
$$L_{PaC} = \mathrm{E}_x \sum_{l=1}^{L} \sum_{s=1}^{S_l} \ell(\hat{g}_l^s, g_l^s, g_l^{S \backslash s})$$

Pixel-wise Contrastive Learning:

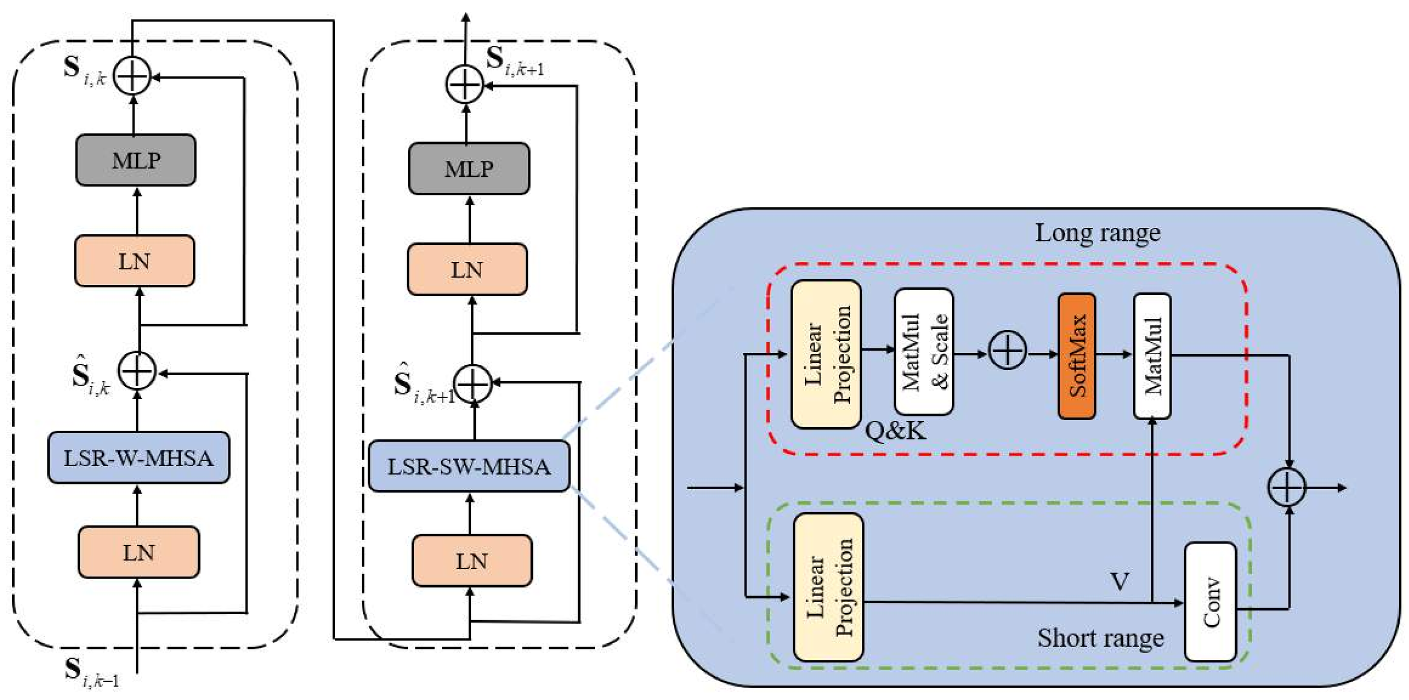
$$L_{PiC} = \sum_{i=1}^{n} \omega_i \frac{\|\psi_i(\tilde{r}) - \psi_i(G(x))\|_1}{\|\psi_i(\tilde{x}) - \psi_i(G(x))\|_1}$$

Method

Network Architecture:



PC-Former



Experiments & Results

Quantitative Results:

	Datasets	LOL[33]		MIT[1]		LSRW[9]	
	Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
T	LIME[7]	15.7586	0.4439	17.5976	0.8179	15.4775	0.4627
	JIEP[2]	16.7856	0.5664	19.5241	0.8690	14.9076	0.5039
	SRLL[17]	15.9872	0.5109	17.6464	0.7793	14.6694	0.5061
S	RetinexNet[33]	16.7741	0.4287	13.7474	0.7394	15.9062	0.4765
	KinD[39]	18.7913	0.7086	17.0935	0.8307	14.8176	0.5691
	STAR[40]	19.9301	0.7896	21.3597	0.8405	15.9629	0.5881
U	EnGAN[13]	15.6314	0.5781	16.4371	0.7966	16.0677	0.4755
	RUAS[18]	19.1076	0.7168	20.0945	0.8734	16.3186	0.6814
	RRD[41]	14.2261	0.5316	18.5372	0.8642	15.8906	0.5276
	Ours	19.6394	0.6901	20.8741	0.8721	16.5984	0.6903

Table 1: Quantitative results (PSNR and SSIM) of state-of-the-art methods and ours on the MIT-Adobe FiveK[1], LOL[33] and LSRW[9] datasets. The best results is in red whereas the second best one is in blue. T, S and U are traditional methods, supervised learning methods and unsupervised learning methods, respectively.

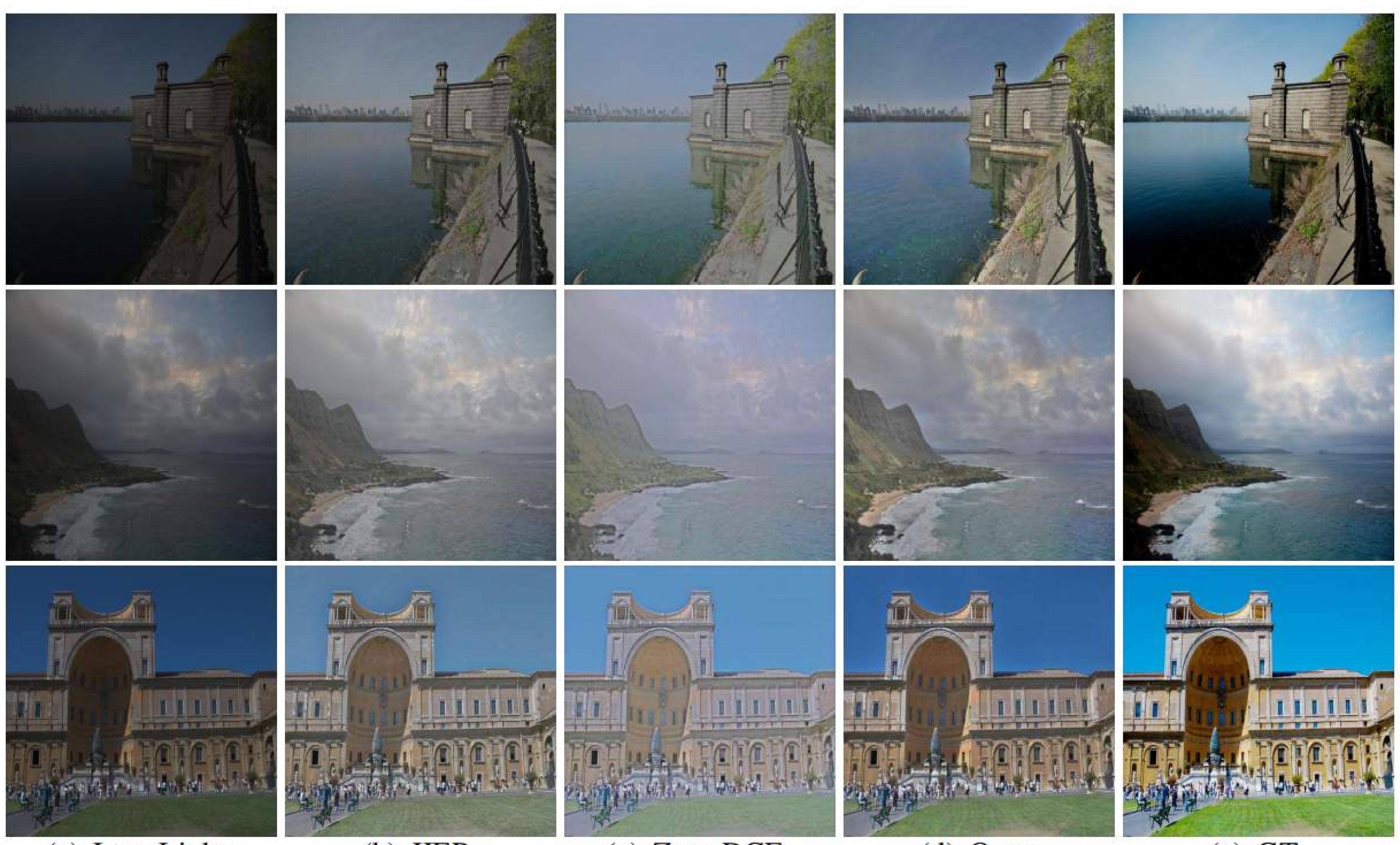
Ablation Results:

Generator (PC-Former)		Dual Contrastive Learning			Baseline(Ours)
U-shape skip connection	residual block	w/o L_{PiC}	w/o L_{PaC}	w/o $L_{PiC}+L_{PaC}$	**19.6394**
18.9043	19.3917	18.8561	18.7611	18.0816	
hyperparameters($\lambda_{adv}, \lambda_{PiC}, \lambda_{PaC}, \lambda_{idt}$)					
(0.1, 0.5, 1, 10)	(1, 0.05, 1, 10)	(1, 1, 1, 10)	(1, 0.5, 0.1, 1)	(1, 0.5, 1, 1)	
18.1762	18.5182	18.9091	18.7716	18.8162	

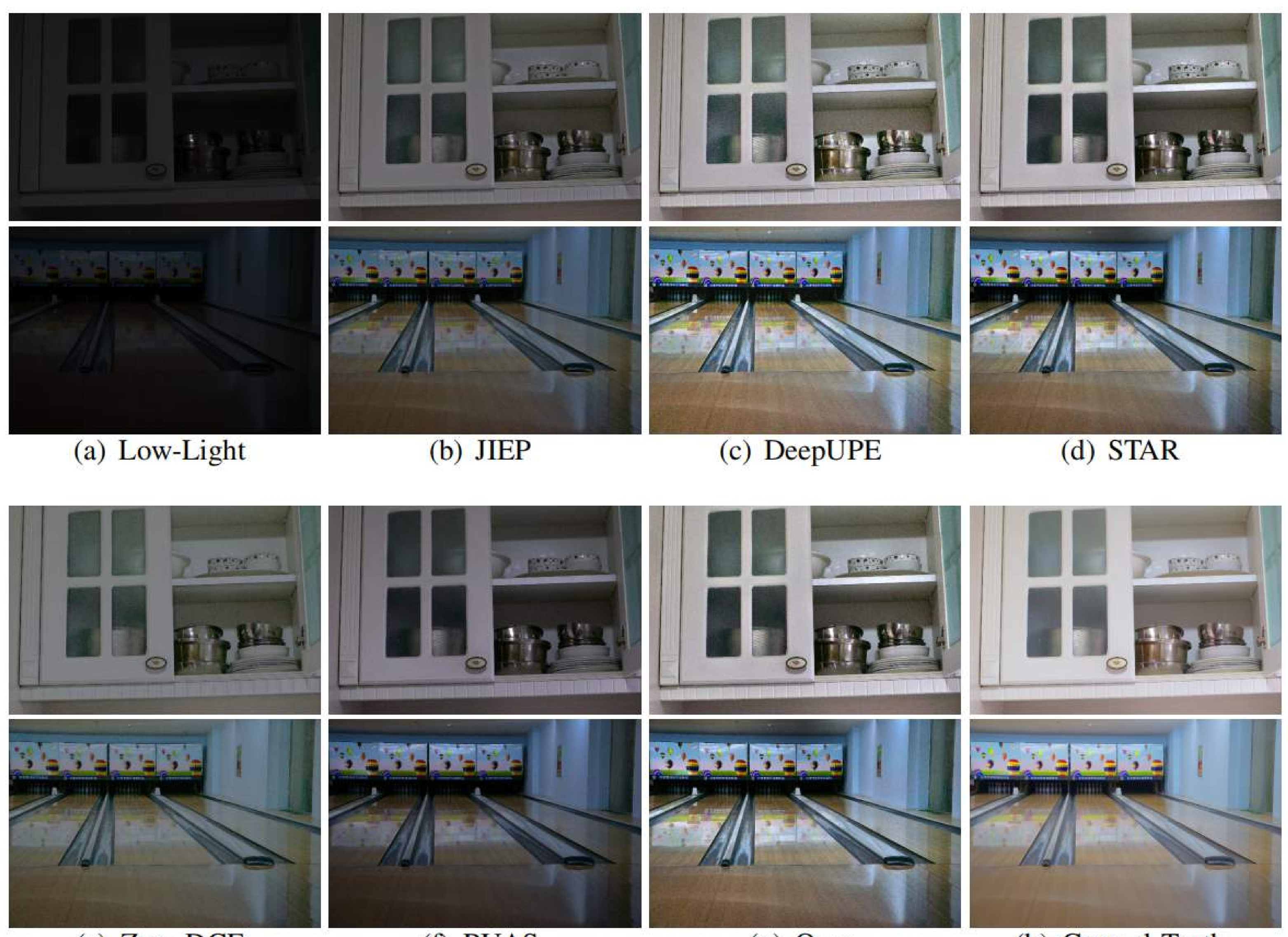
Table 2: Ablation Analysis (PSNR) on UDCL-Transformer, loss functions and hyperparameters on LOLdataset [33]. w/o represents without.

Dark Face Detection

Visual Compsarison:



(a) Low-Light (b) JIEP (c) Zero-DCE (d) Ours (e) GT



(a) Low-Light (b) JIEP (c) DeepUPE (d) STAR

(e) Zero-DCE (f) RUAS (g) Ours (h) Ground-Truth