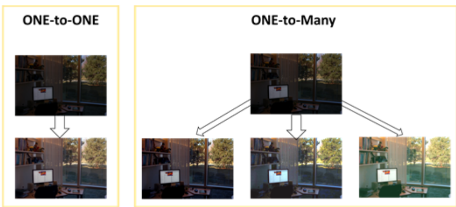
Enhance Images as You Like with Unpaired Learning

Xiaopeng Sun, Muxingzi Li*, Tianyu He, Lubin Fan*

Xidian University
Alibaba Group

Introduction

Image Enhancement



Contributions

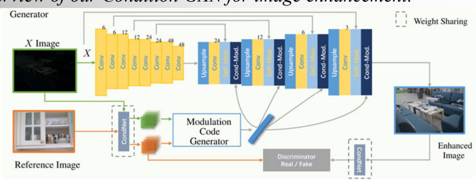
(1) cGAN-inspired framework for one-to-many image enhancement
(2) Carefully designed losses that enable unpaired learning:

- an idempotence loss that assumes a normal-light image should be mapped to itself when conditioned on itself
- a spatial consistency loss that facilitates the generated image to have more spatial coherence with the input
- a global color consistency loss that makes the overall color coherent with the input
- a GAN loss that tries to make the outputs more realistic

Our Method

General Framework

Overview of our Condition GAN for image enhancement.

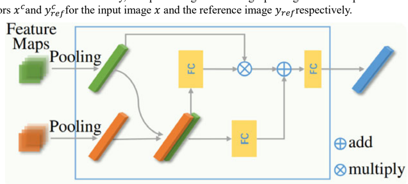


We employ a U-Net Translator that performs conditional translation on the input low-light image. Our U-Net Translator consists of two complementary modules: the Pixel-wise Self-Modulation (PSM) and the Channel-wise Conditional Modulation (CCM), where both of them are designed to adjust the feature distribution of the low-light input image but from different aspects. Specifically, the PSM is designed to learn modulation parameters from previously upsampled features, while the CCM is designed to learn from the features of both the low-light and the reference image. In particular, the features fed into CCM are generated by our Condition Net (CondNet), which consists of three convolutional layers.

Details on Network Architecture

- *Modulation Code Generator*

We combine the information from both the input image and the reference one to facilitate the learning process of Modulation Code Generator. Formally, we perform global average pooling on the outputs of CondNet, forming two feature vectors x^c and y^c_{ref} for the input image x and the reference image y_{ref} respectively.



Our Modulation Code Generator can be formulated as:

$$c_{ref} = fc_{\text{out}}\left(fc_{\text{in}}(c'_{ref}) \odot y^c_{ref} \oplus fc_y(c'_{ref})\right),$$

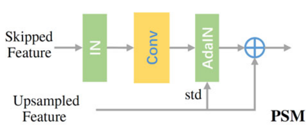
where $c'_{ref} = \text{concat}(x^c, y^c_{ref})$. The concat and fc indicate concatenation operation and fully-connected layer respectively, $\odot$ and $\oplus$ denote element-wise multiplication and addition respectively.

- *U-Net Translator*

We tailor the standard U-Net architecture for our task from three aspects: 1) we propose a Pixel-wise Self-Modulation to match the statistics of the upsampled feature with the skipped feature. 2) we propose a Channel-wise Conditional-Modulation to polish the feature with the generated modulation code, which is vital for conditional image enhancement. 3) we remove the original batch normalization layers which destroy the relative distribution across channels, as our Channel-wise Conditional-Modulation is learned more efficiently by seeing the original distributions.

- *Pixel-wise Self-Modulation (PSM)*

we propose a Self-modulation Block to adjust the statistics of skipped feature by the upsampled feature from the previous layer. The skipped feature is first processed by an instance normalization layer and two 3×3 convolution layers with leaky ReLU, then fed into AdaIN, whose mean and variance are calculated from the upsampled feature. The goal of our PSM block is to enhance the skipped feature, which consists of multi-scale texture information, in an adaptive way. Intuitively, it can be viewed as modulating the lower-level representation using the higher-level one.

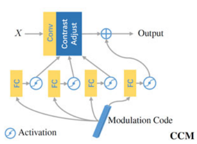


PSM

- *Channel-wise Conditional-Modulation (CCM)*

Our CCM block plays two important roles: 1) it transfers the style of the reference image, which encoded in a modulation code, to the input low-light image. 2) it performs learnable non-linear transformation on the learned feature. Formally, let c_{ref} be the modulation code inferred from y_{ref}, our CCM block first generates four coefficient vectors α^1_{ref}, α^2_{ref}, $\bar{\alpha}_{ref}$ and β_{ref} based on the modulation code c_{ref} through two fully-connected layers. Then the retouch operation $m(\cdot)$ can therefore be formulated as:

$$m(x) = \begin{cases} \alpha^1_{ref} \odot (x \ominus \bar{\alpha}_{ref}) \oplus \beta_{ref}, if\ x > \bar{\alpha}_{ref} \\ \alpha^2_{ref} \odot (x \ominus \bar{\alpha}_{ref}) \oplus \beta_{ref}, if\ x \le \bar{\alpha}_{ref} \end{cases}$$



CCM

Ablation Study

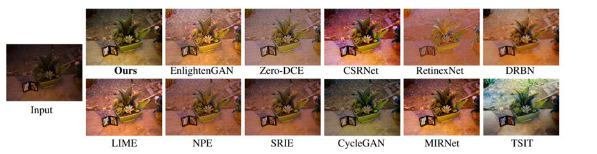
- *Ablation study on losses and the CCM block*

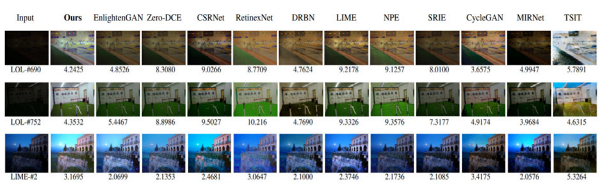
Experiment Results

Quantitative Results

Method \ Metric		Unpaired	Conditional	LOL-690	FiveK
SRIE				15.35 \ 0.559 \ 7.4022	16.90 \ 0.750 \ 4.1352
LIME				17.97 \ 0.512 \ 8.2972	16.67 \ 0.772 \ 3.7043
NPE				17.62 \ 0.481 \ 8.5105	15.60 \ 0.736 \ 3.6475
RetinexNet				16.17 \ 0.420 \ 9.2652	11.89 \ 0.644 \ 4.4298
DRBN				18.71 \ 0.784 \ 4.5612	15.07 \ 0.562 \ 7.1623
CSRNet				15.69 \ 0.408 \ 8.1343	23.68 \ 0.896 \ 3.7492
EnlightenGAN		✓		18.89 \ 0.692 \ 5.0857	15.47 \ 0.734 \ 3.7616
Zero-DCE		✓		18.47 \ 0.598 \ 7.8224	13.01 \ 0.557 \ 7.3117
CycleGAN		✓		17.42 \ 0.576 \ 4.0663	17.04 \ 0.681 \ 4.8327
TSIT			✓	13.14 \ 0.533 \ 5.5965	14.35 \ 0.638 \ 5.3926
MIRNet		✓		12.90 \ 0.432 \ 4.2501	19.36 \ 0.806 \ 3.9225
Ours	Min.			12.24 \ 0.609 \ –	11.97 \ 0.655 \ –
	Avg.	✓	✓	17.00 \ 0.671 \ –	17.37 \ 0.750 \ –
	Max.			22.45 \ 0.732 \ 4.0733	20.87 \ 0.797 \ 4.0305

Qualitative Results





Acknowledgements

We thank Jing Ren for useful suggestions on the manuscript and all the anonymous reviewers for their valuable comments.

Contact Information

Email : xpsun@stu.xidian.edu.cn; muxingzi.lmxz@alibaba-inc.com
Paper:
Github : https://github.com/sxpro/ImageEnhance_cGAN