Prof. Wan-Chi Siu
(Speaker)

Deep Lightening Network for Low-Light Image Enhancement

Authors: Li-Wen Wang, Zhi-Song Liu, **Wan-Chi Siu**, and Daniel P. K. Lun

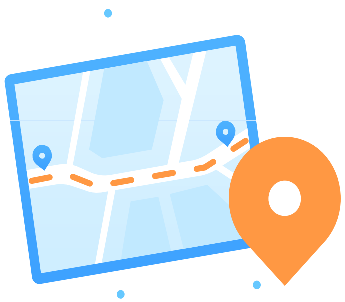
Department of Electronic and Information Engineering

The Hong Kong Polytechnic University

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Opening Minds • Shaping the Future • 啟迪思維 • 成就未來

❖ Outline



1. Background
2. The Proposed Method
3. Experiments
4. Conclusion

Background

❖ Background

- Images taken in low-light conditions are usually very **dim**, which is difficult to recognize the scene or object.
- **Darkness** brings us uncertainty, worry and low-confidence, especially when we are walking or driving. We may use **visual aid** (including small devices such as **mobile phones**) which is more convenient compared with dedicated approaches such as **infrared detection**.

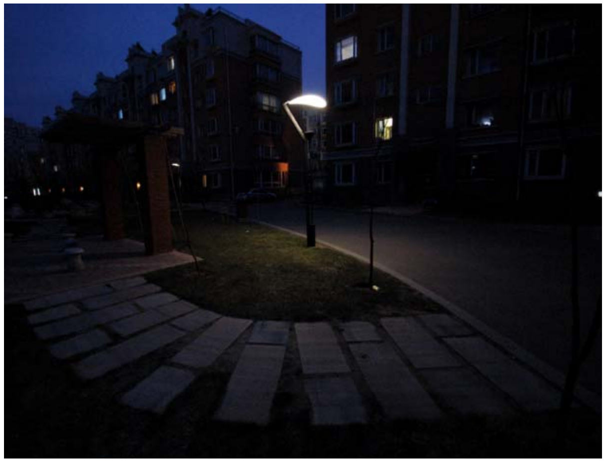


Image taken in low-light condition

Our proposed method

❖ Background (cont'd)

To take good visual-quality images in dark environment, there are some possible solutions:

- **Flash:**

  It may **bother** others and is **not allowed** in some places (museum).

- **High ISO sensitivity:**

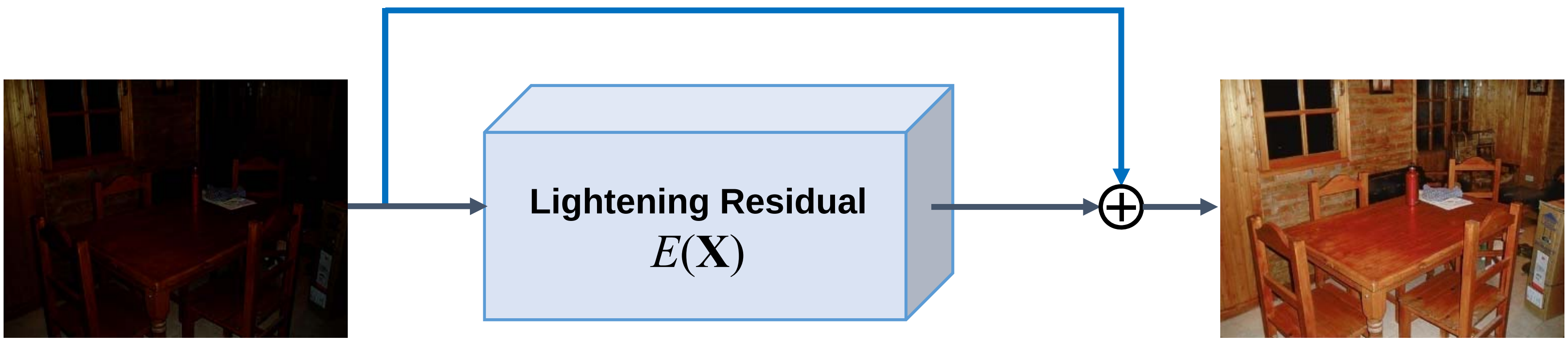
  It brings more **noise** and **overexposes** to the normal-light areas.

- **Longer exposure time:**

  It may suffer from **blur** problems and is not suitable for shooting **videos** (the interval of video frames may be too short).

The Proposed Method

❖ **Assumption:** Residual Learning



Low-Light image
($\mathbf{X}$)

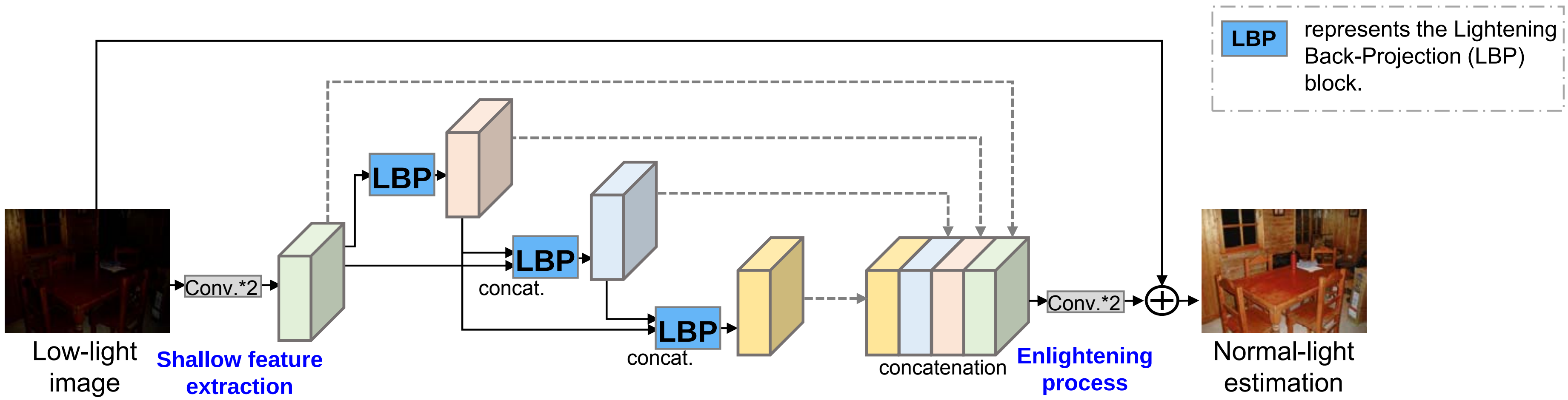
Normal-Light image
($\mathbf{Y}$)

Problem definition:

$$\mathbf{Y} = \mathbf{X} + E(\mathbf{X})$$

where $E(\cdot)$ denotes the enhancing operator that estimates the residual (LL(Residual) between low-light (LL) and normal-light (NL) images.

❖ The Proposed Method – Deep Lightening Network (DLN)



The proposed **Deep Lightening Network** (**DLN**) consists of **three parts**:

(i) shallow feature extraction,
(ii) Lightening Back-Projection (LBP) blocks, and
(iii) enlightening process.

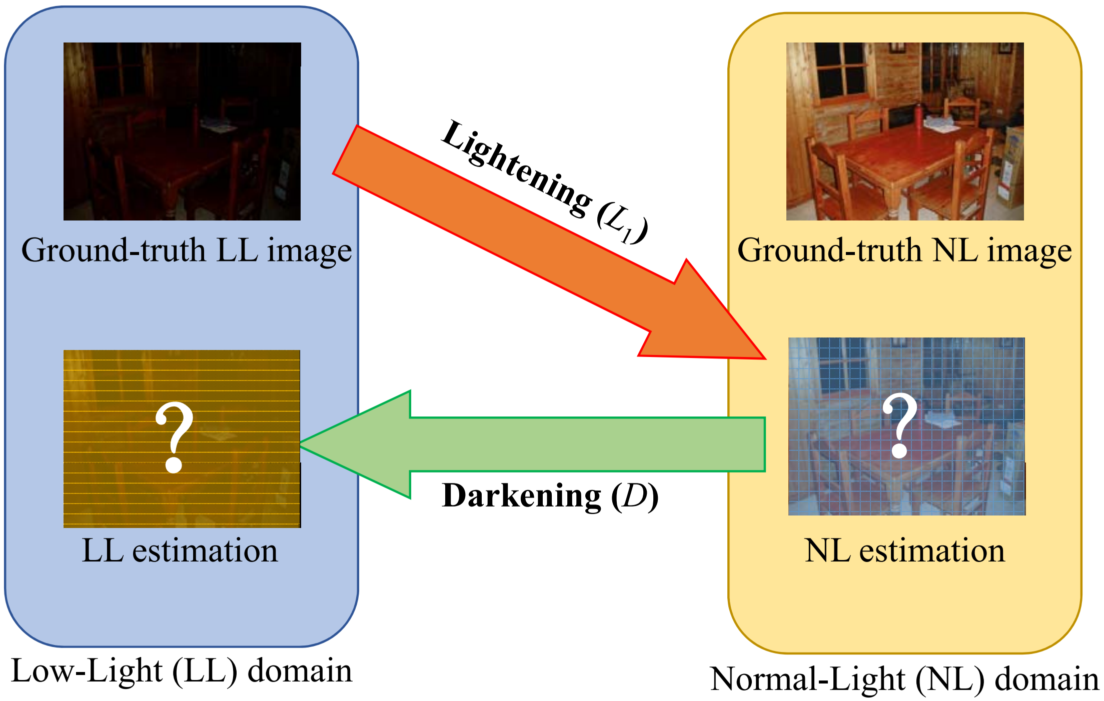
- The DLN takes the LL image as the input. It firstly enters into the **feature extraction** part that maps the image from the RGB space to feature space.
- Then, the **Lightening Back-Projection (LBP) blocks** scheme starts to enhance the LL image accumulatively.
- Finally, the **enlightening** process receives the results from LBPs and **estimates the NL image.**

❖ The Proposed Method – domain transfer learning

Mapping between low- and normal-light domain:

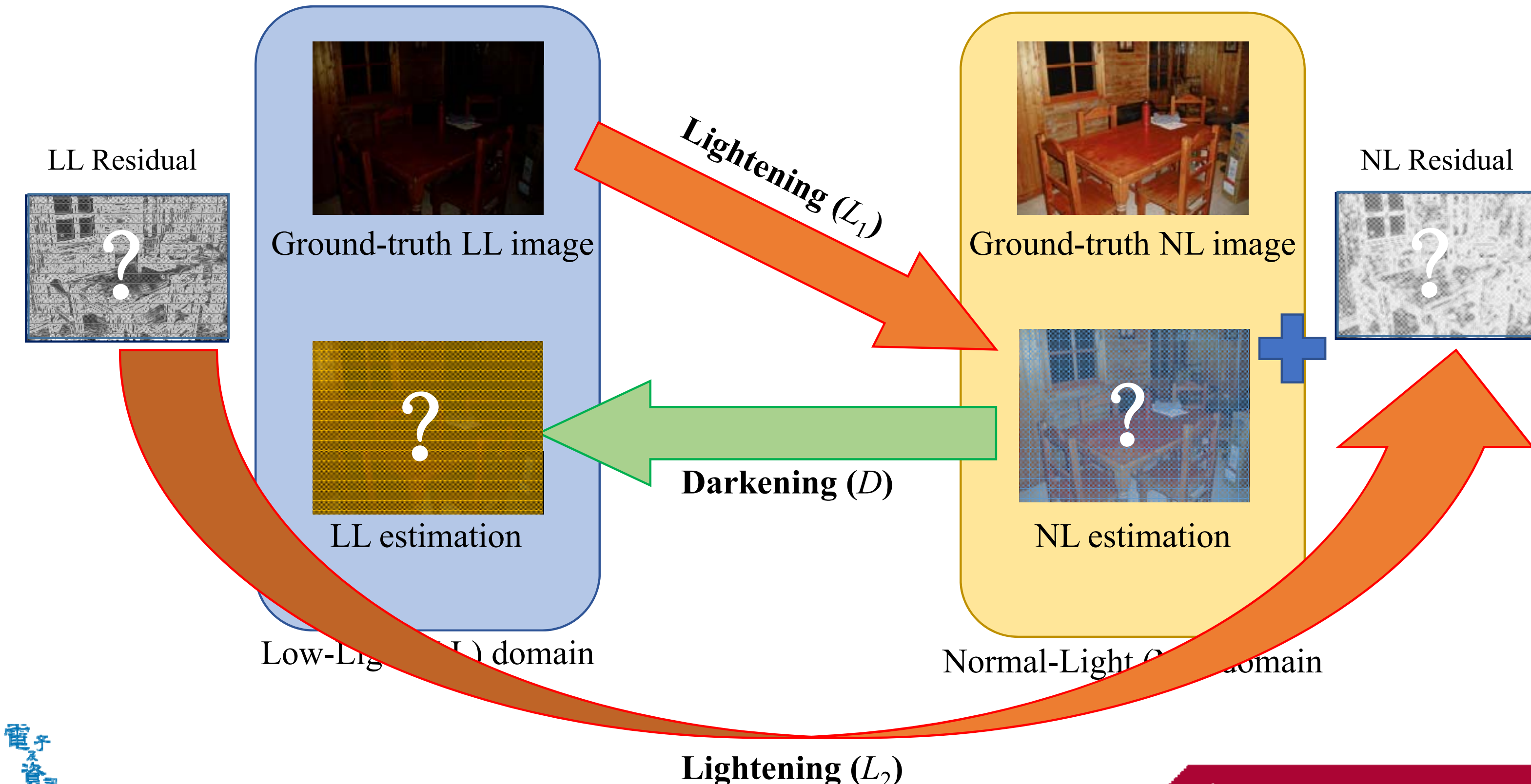
Our approach is to **make "Lightening" process and "Darkening" process iteratively,** which means to **learn the lightening mechanism gradually**, and finally obtain the normal-light image.

This process can also be roughly described as **domain transfer learning**.

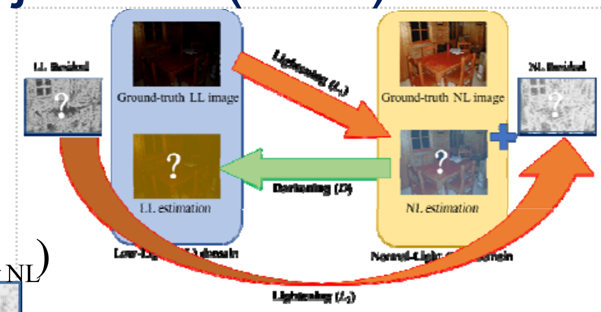


❖ **The Proposed Method** – domain transfer learning

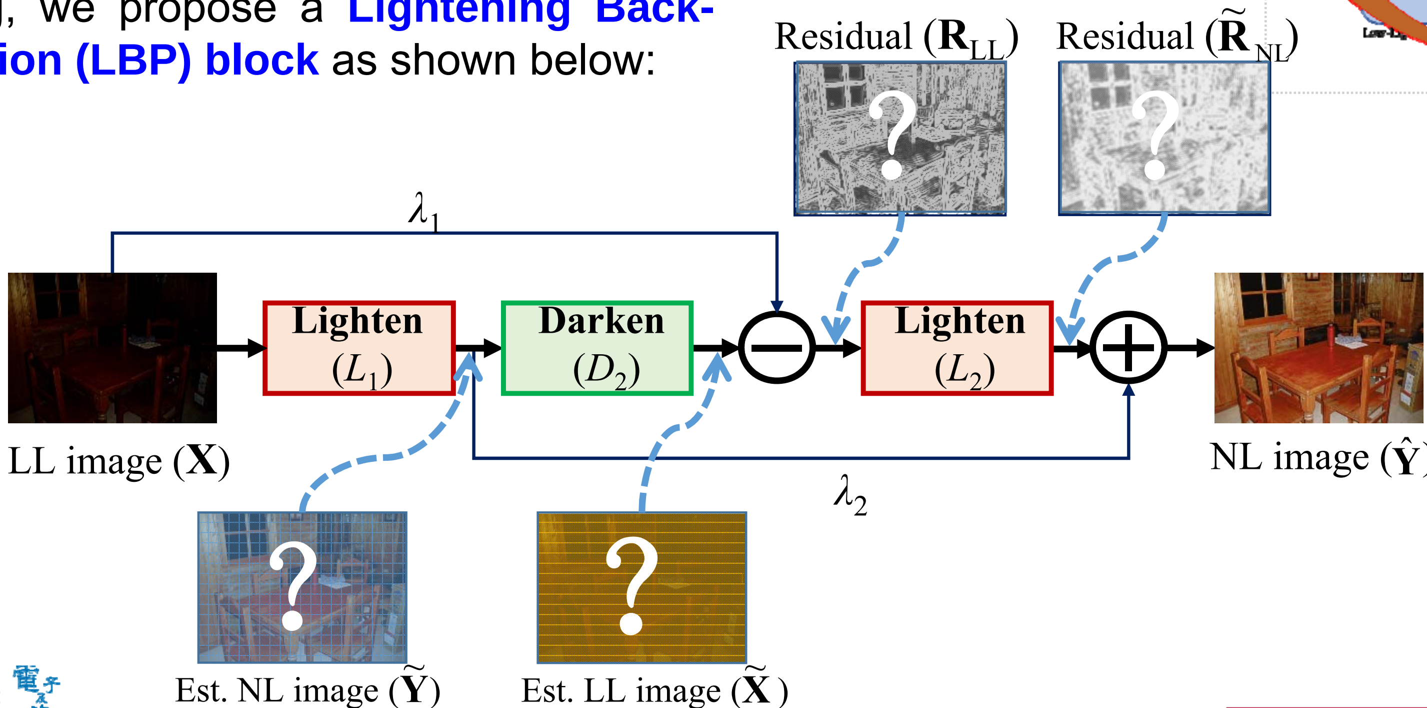
Mapping between low- and normal-light domain:



❖ The Proposed Method – Lightening Back-Projection (LBP)



Based on the ideas of the domain transfer learning, we propose a **Lightening Back-Projection (LBP) block** as shown below:



❖ The Proposed Method – Lightening Back-Projection (LBP)

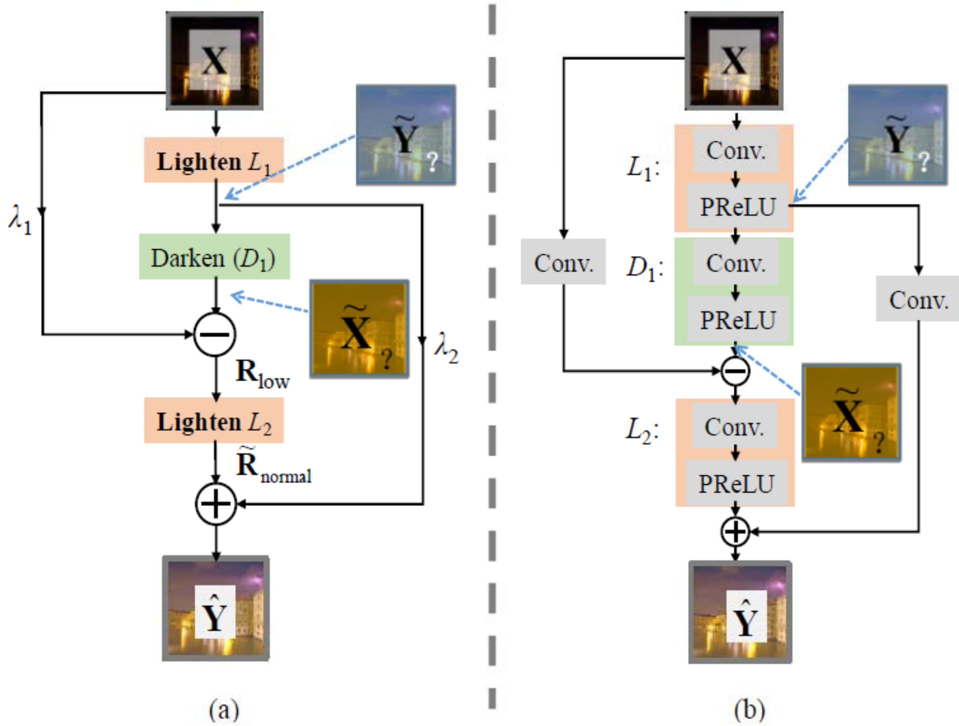


Figure 4. Lightening Back-Projection (LBP) Block: (a) framework, (b) CNN structure

An advantage of CNN is that the convolutional layers contain trainable parameters (weight w and bias b), which makes CNN be adaptive to different tasks. It is interesting that we can use the **same CNN structure** for **both Lightening and Darkening processes**.

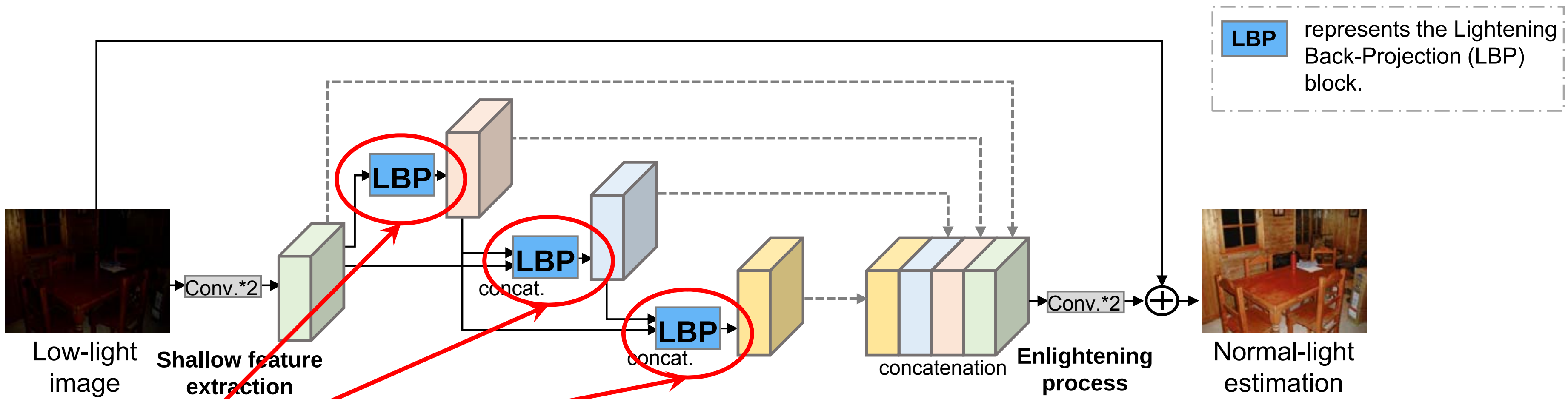
Explanation:

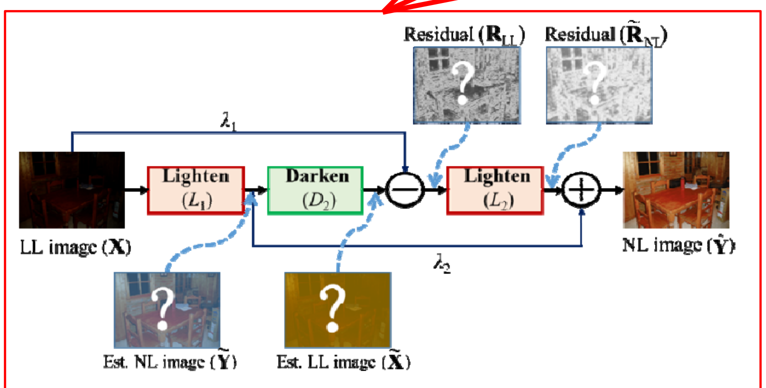
***L_2*-lightening -** The input of L_2 is the residual information of the low-light image, which makes the output of L_2 also be residual. Considering the LBP predicts normal-light image at the output, the output of L_2 should be the residual $\tilde{\mathbf{R}}_{normal}$ of normal-light condition, so ***L_2* lightens** the residual from the low- to normal-light domain.

***L_1*-Lightening -** To obtain the normal-light image after an addition process, it needs another information $\tilde{\mathbf{Y}}$ that is from the output of L_1. The input of L_1 is low-light image $\mathbf{X}$ and output is the normal-light image $\tilde{\mathbf{Y}}$, hence L_1 achieves the **Lightening** process.

***D_1*-Darkening -** Because the input of L_2 is the residual $\mathbf{R}_{low}$, D_1 should act **darken** process that transfer the normal-light estimation $\tilde{\mathbf{Y}}$ to low-light's $\tilde{\mathbf{X}}$.

❖ **The Proposed Method** – multi-level LBPs





•For low-light image enhancement task, both **global** and **local** information are important, as we need **global information to evaluation the light condition** of the whole image, and **local information to enrich the details**. Hence, lightening image at multi-level features can benefit the performance.

•Instead of stacking blocks end-to-end, we add extra **short connections** among the **LBPs** such that the latter block can obtain results from all previous stages. Then, a convolutional block receives results from **different levels** and **fuses the information**.

❖ The Proposed Method – loss function

- **For Training**, we regard the low-light video enhancement as **a supervised learning task**. For each low-light input, there is a normal-light target that is known to us.

- We need to measure the difference between the estimation and the ground-truth target. The loss is defined as two parts:

  - **Structure Similarity:**. **SSIM** measure the structure similarity. For the low- and normal-light images, structure is the most important element. The range of SSIM is [0, 1], and a larger SSIM value means more similarity.

$$SSIM(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \cdot \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}$$

  - **Constraint of Smoothness:** The estimated image may have inconstant illumination and noises. To handle the problem, we can use a regularization term that is **Total Variation (TV)**

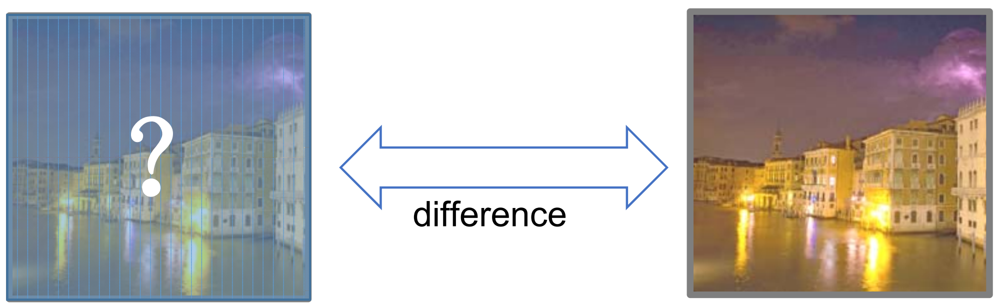
$$Loss_{TV}(\mathbf{P}) = \sum_{i,j} \sqrt{(p_{i,j} - p_{i+1,j}) \cdot (p_{i,j} - p_{i,j+1})}$$

❖ The Proposed Method – loss function (cont'd)

- the differences (or loss) between the estimation and ground-truth images can be defined as:

$$Loss(\hat{\mathbf{Q}}, \mathbf{Q}) = 1 - SSIM(\hat{\mathbf{Q}}, \mathbf{Q}) + \lambda * Loss_{TV}(\hat{\mathbf{Q}}, \mathbf{Q})$$

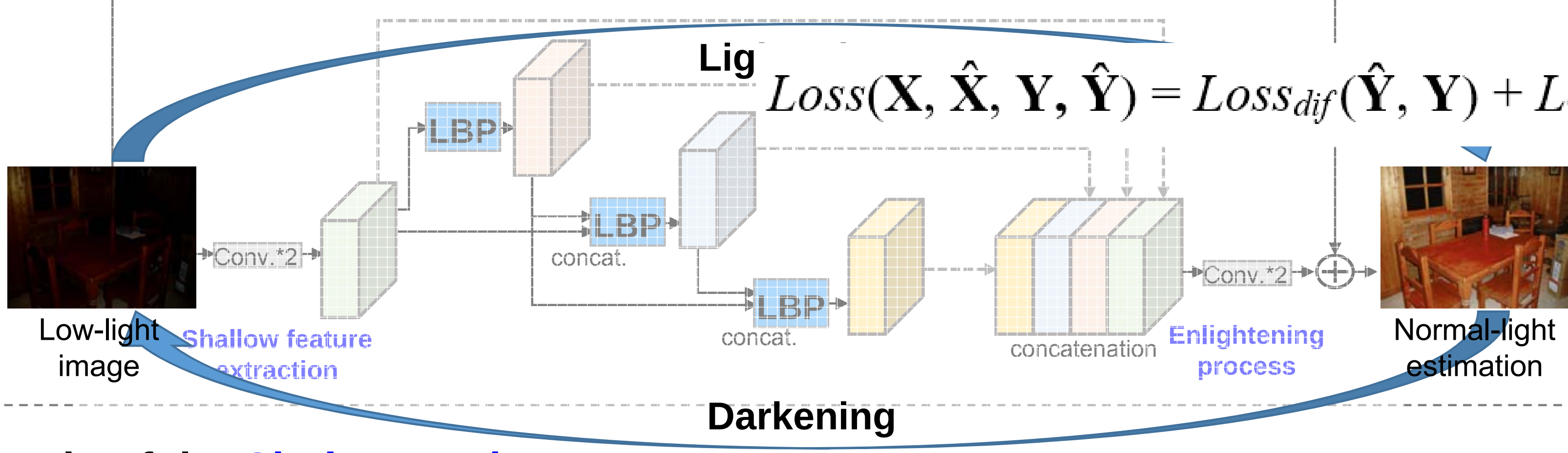
where $\hat{\mathbf{Q}}$ and $\mathbf{Q}$ denote the estimated and ground-truth images separately. λ is the balance coefficient (we use 0.001 in our experiments).



Normal-light image (estimation $\hat{\mathbf{Q}}$)

Normal-light image (ground truth $\mathbf{Q}$)

❖ **The Proposed Method** – loss function (cont'd)



$$Loss(\mathbf{X}, \hat{\mathbf{X}}, \mathbf{Y}, \hat{\mathbf{Y}}) = Loss_{dif}(\hat{\mathbf{Y}}, \mathbf{Y}) + Loss_{dif}(\hat{\mathbf{X}}, \mathbf{X})$$

Constrain of the Circle Learning:

The estimated normal-light image should contain enough information, which can darken back as a low-light image. **We design our training structure that contains two DLN models**: one is for Lightening, the other is for Darkening. We train the two DLN models together by a common loss function which combines the two losses:

$$Loss(\mathbf{X}, \hat{\mathbf{X}}, \mathbf{Y}, \hat{\mathbf{Y}}) = Loss_{dif}(\hat{\mathbf{Y}}, \mathbf{Y}) + Loss_{dif}(\hat{\mathbf{X}}, \mathbf{X})$$

where $\mathbf{X}$ and $\mathbf{Y}$ denote the low- and normal light images, $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ are the estimations from the DLNs

It is interesting to note the the same network structure is used for both Lightening and Darkening. Why?

Experiments

❖ Experiments

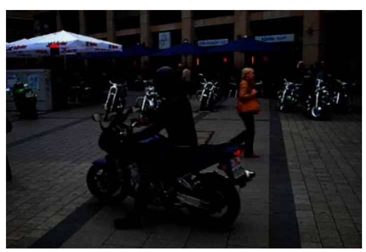
- It is **difficult** to capture two images of **different illumination** at the same time.
- The normal-light images usually contain more information and less noise than the low-light ones. It is **feasible** to synthesize the low-light images from the normal-light ones.
- Following the analysis, **a LL image can be simulated from an NL image** through the following simulation equation [18]:

$$\overline{\mathbf{X}}^{(i)} = \beta(\alpha\mathbf{Y}^{(i)})^{\gamma}$$

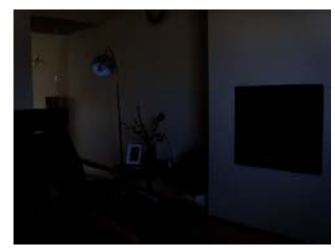
Where $\overline{\mathbf{X}}^{(i)}$ represents a simulated LL image. The piel value of the NL image $\mathbf{Y}$ is compressed to [0, 1]. i denotes the R, G, or B channel of the image. $\alpha \sim U(0.9, 1)$, $\beta \sim U(0.5, 1)$ and $\gamma \sim U(1.5, 5)$ which control the effect of low-light simulation.

Examples of simulated LL-NL image pair:


NL image


Simulated LL image


NL image

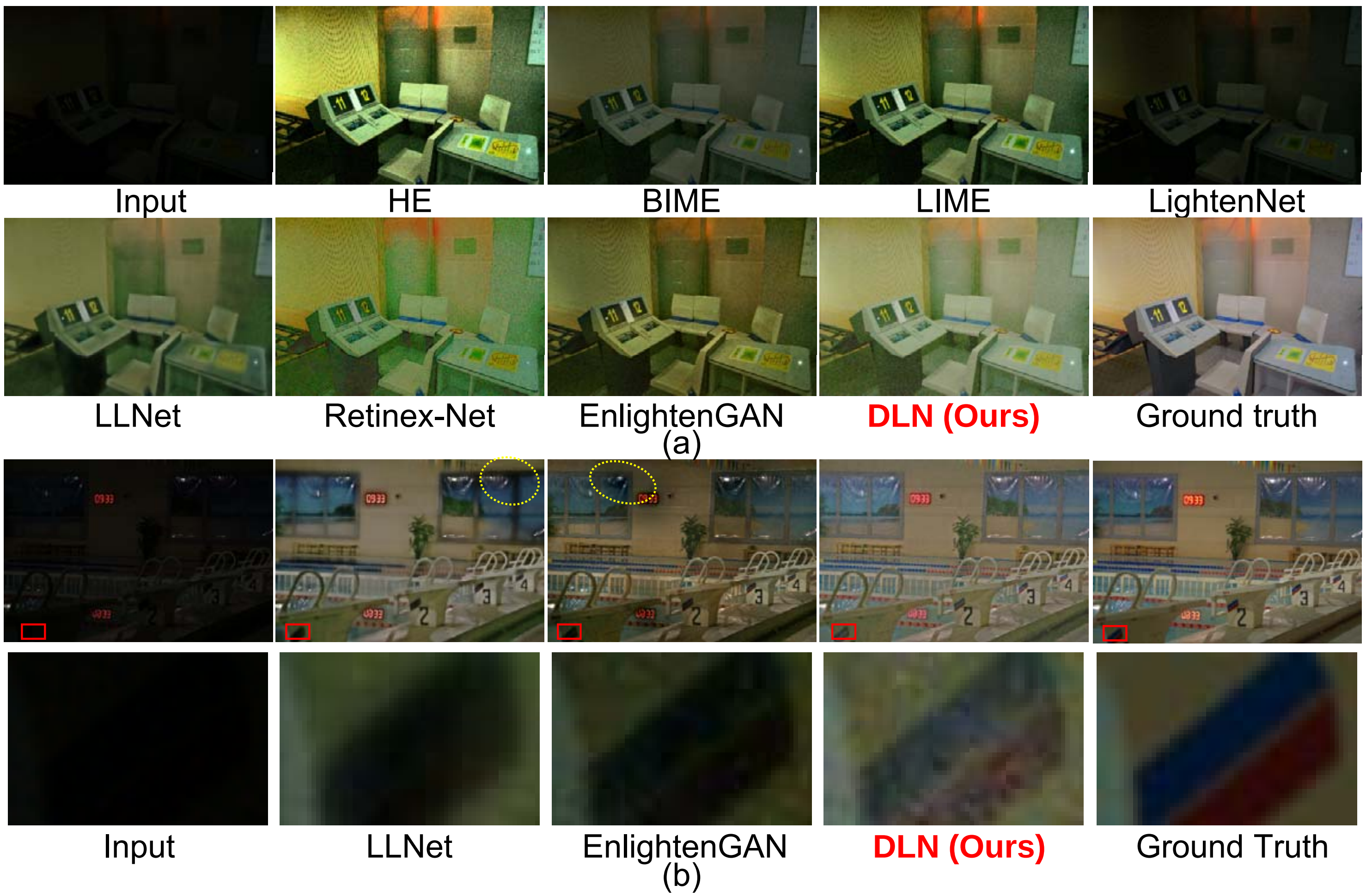

Simulated LL image

[18] Feifan Lv and Feng Lu, “Attention-guided Low-light Image Enhancement,” *arXiv preprint arXiv:1908.00682*, 2019.

❖ Experiments

- **Training data**: **Images in PASCAL VOC 2007 dataset** have good visual quality that were initially used **for image object detection**. We used all 9,963 images in PASCAL VOC 2007 dataset (train + validation + test) as the ground-truth NL images to simulate the low-light images.
- **Testing data**: **LOw-Light (LOL) dataset contains 500** (485 for training (not used here) and 15 for testing) paired LL-NL images that were captured by a camera with two different **exposure** times and **ISO** for each **still** scene. Note that the same training and testing data were used by all approaches.

❖ Experimental result - visualization



❖ **Experimental result –** performance at LOL testing dataset

We believe that the enhanced LL image should be close to the ground-truth NL image. Therefore, we adopt Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity (SSIM), which are widely used in image restoration field to measure the quality of the estimation.

Method	**PSNR**	**SSIM**
HE	15.467	0.504
BIMEF	13.875	0.577
LIME	16.920	0.599
LightenNet	10.301	0.361
LLNet	**17.953**	**0.704**
Retinex-Net	16.774	0.559
EnlightenGAN	**17.483**	**0.658**
DLN (proposed)	**19.256**	**0.724**

❖ Conclusion

- We propose a **Deep Lightening Network (DLN)** for low-light image enhancement. Experimental results show that the DLN **outperforms** other **state-of-the –art methods in all objective and subjective measures**.
- We regards the low-light enhancement as a **domain transfer learning** that learns the mapping function between low- and normal-light domain. We **propose** a **Lighting Back-Projection (LBP) block** that iteratively learns the residual between low-light and normal-light images.
- To aggregate the global and local information, we propose a **fusion block** that receives the lightening results from **the LBPs of different scales**.

More work and Future Work

Since the submission of this paper at the end of last year, there are some good development of the topic. Say

(i) we have modified the **lightening and darkening blocks** making them more distinct for lightening and darkening objectives,

(ii) we have also ported architectural structure matching with some **existing deep learning networks**, such as the U-net successfully, etc.

The results of which give **comparable or even better performance**. These are certainly **directions** for future research.

The end,
thank you!