Low Light Face Enhancement

Liming Gong
Instructor: Ha Le
Professor: Ioannis Kakadiaris

Background

Face detection and landmark detection have always suffered from low image qualities. One of these obstacles is poor lighting condition. The underexposure images will make face detection harder and significantly deteriorate face detection and landmark detection accuracy. One obvious solution is to do image enhancement before doing face detection and landmark detection. This project is trying to evaluate how the image enhancement process will affect the result of face and landmark detection.

There are lots of face enhancement methods, one of them is based on Retinex theory. It states the relationship between what human see when a certain light is casted on a certain surface. With some hypothesizes, Retinex theory can do this inverse process and decompose what human seen into the original illumination and the surface albedo. By adjusting the illuminance and enhance it, then cast this brighter lighting on the same surface, we can enhance the image.

However, how to decompose the received scene into the source lighting and the surface albedo is an ill-posed problem. Most existing Retinex-based methods have carefully designed hand-crafted constraints and parameters for this highly ill-posed decomposition, which may be limited by model capacity when applied in various scenes. This paper "Deep Retinex Decomposition for Low-Light Enhancement" gives a promising result by apply deep learning on this problem. Thus, we choose the method stated by this paper to do the evaluation.

Project Overview

This project contains multiple steps:

1, thoroughly read and understand the paper "Deep Retinex Decomposition for Low-Light Enhancement".

2, reimplement this paper and reproduce the result the author proclaimed.

3, acquire no less than 100 selfies by cellphone camera from 10 people in closed environment with different lighting conditions. The lighting should slightly adjust from darkest to normal with no less than 10 levels.

4, apply the Retinex-Net image enhancement method on the selfies.

5, find, download and install a pretrained face detection and landmark detection method.

6, test the face detection and landmark detection on relighted selfies and compare with the result from original images.

Approach and Result

Retinex theory

The classic Retinex theory models the human color perception. It assumes that the observed images can be decomposed into two components, reflectance and illumination. Let S represent the source image, then it can be denoted by: S = R*I, where R represents reflectance, I represents illumination and * represents element-wise multiplication. Reflectance describes the intrinsic property of captured objects, which is considered to be consistent under any lightness conditions. The illumination represents the various lightness on objects. On low-light images, it usually suffers from darkness and unbalanced illumination distributions.

More than this, to decompose this S into R and I, the Retinex theory also has below hypothesizes: the original image has variations, which can be clustered into 2 parts, the slow changes and sharp changes. The slow changes are caused by the illuminance, and the sharp changes are caused by the shape or surface discontinuity. This is a strong assumption which makes the original image easy to decompose. It also shows some limitation: generally, retinex theory misses one possibility, the illimitation change can also introduce sharp edges.

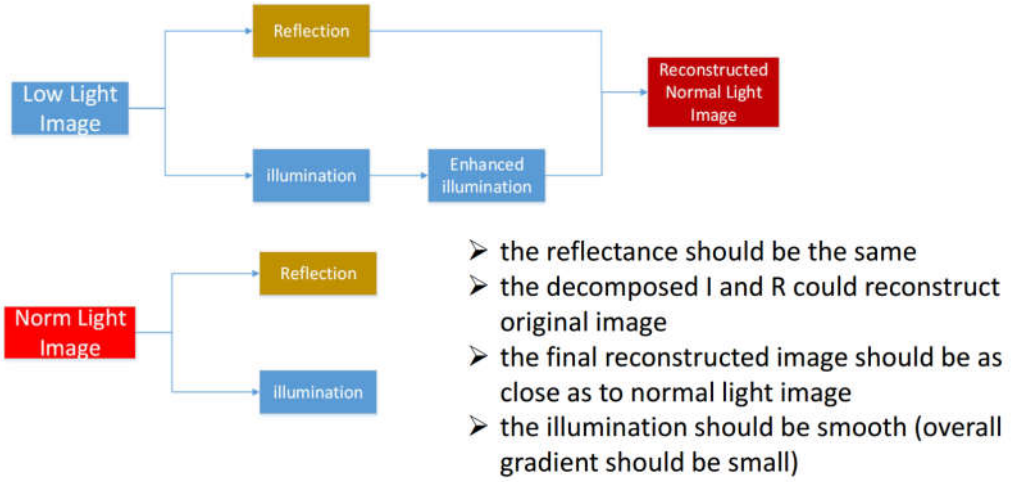
Retinex-Net

The overall idea of Retinex-Net is to decompose 1 image pair: a low light image and its corresponding normal light image. The normal light image is used as reference.

The dataset is acquired by the author themselves by cameras. They adjust the exposure time and ISO, and choose the most reliable pairs.

Since the image pair is the same scene with different exposure time, if decomposed into reflectance and illumination, there are several requirements should be satisfied:

1) The reflectance of low light image and normal light image should be the same.
2) The decomposed reflectance and illumination should be able to reconstruct the original image if multiply back. Both low and normal light image should satisfy this.
3) After do some illumination enhancement, if the result is multiplied with the reflectance, then generated image should be as close to normal light image as possible.
4) To diminish the limitation of Retinex theory, the decomposed illumination map should be smooth as the theory stated, it should also be able to keep some structure, which may be caused by illumination change.

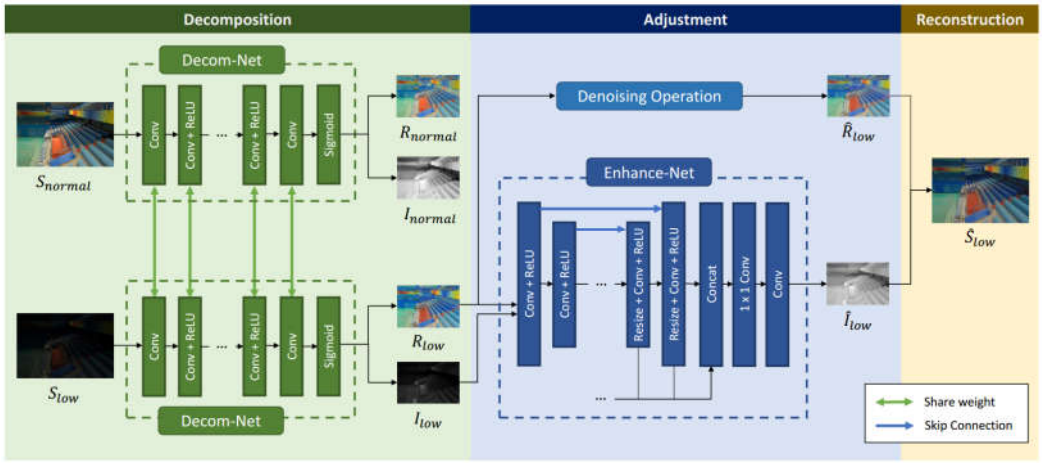


Below picture shows the overall architecture of the author proposed Retinex-net. It composes 3 parts: the decomposition, adjustment, and reconstruction.

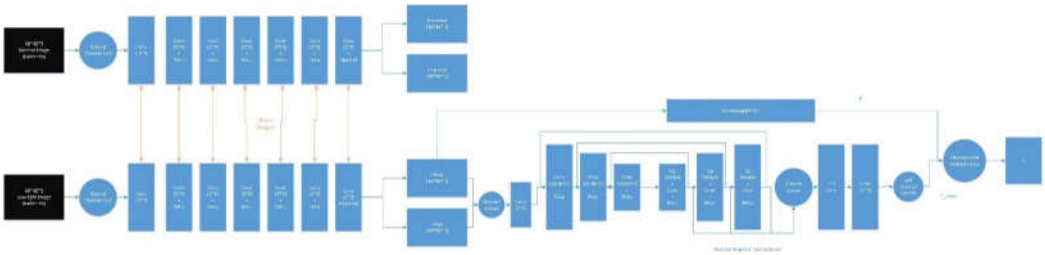
The input is image pair is decomposed respectively by the Decom-Net. The Decom-Net for low light image and normal light image share the same weight. Each Decom-Net is composed of several convolutional layers. The outpu R_{normal} and I_{normal}, is the decomposed reflectance and illumination for normal light image. R_{low} and I_{low} is the same for low light image.

The adjustment part is mainly composed of the Enhance-Net, which is a encoder-decoder structure with residual connection. It takes the decomposed R_{low} and I_{low} concatenated together as input, and generate the relighted illuminance. There is also a denoising branch for the reflectance map R_{low}, which is performed offline. The intuition to do denoising is that, the original low light image if decomposed into reflectance, certainly would not perform as good as normal light image, and it will have a lot of noise due to the information loss.

The reconstruction part is to reconstruct the normal light image by multiplying the denoising reflectance map and the enhanced illumination.



Below is a more detailed version, with the kernel size and activation function shown.



Loss function

Below is the loss function for the Decom-Net, it contains 3 parts:

$$\mathcal{L} = \mathcal{L}_{recon} + \lambda_{ir}\mathcal{L}_{ir} + \lambda_{is}\mathcal{L}_{is}$$

The reconstruction loss, which denote how difference between the original image and the reconstructed image from the decomposed R and I.

$$\mathcal{L}_{recon} = \sum_{i=low,normal} \sum_{j=low,normal} \lambda_{ij} ||R_i \circ I_j - S_j||_1.$$

The reflectance difference loss, which shows the reflectance difference between low light image and normal light image for the same scene.

$$\mathcal{L}_{ir} = ||R_{low} - R_{normal}||_1.$$

The illumination smoothness loss, which add over the gradient of the image, and loose the constrain where big variations occurs.

$$\mathcal{L}_{is} = \sum_{i=low,normal} ||\nabla I_i \circ exp(-\lambda_g \nabla R_i)||$$

Below is the loss function for Enhance-Net, which is composed by 2 parts: the reconstruction loss and the smooth loss. The reconstruction loss denotes the difference between the enhanced low light image with the original normal light image.

$$\mathcal{L} = \mathcal{L}_{recon} + \lambda_{is}\mathcal{L}_{is}$$

Training and testing

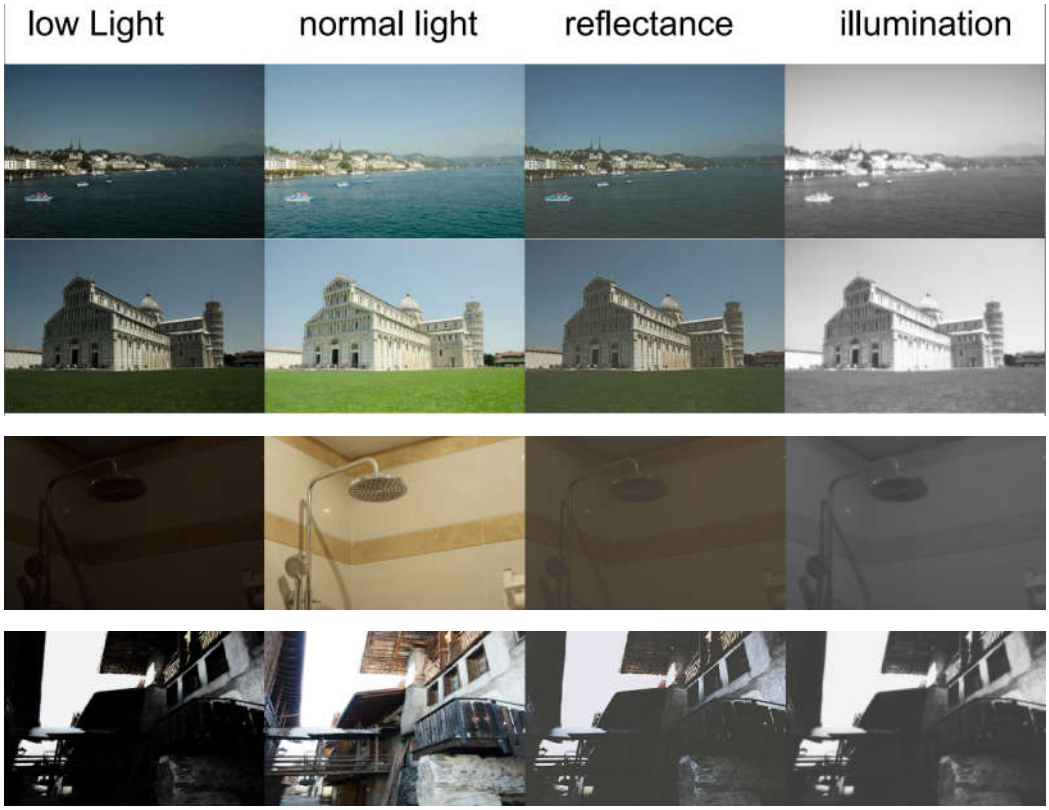
The general process is 2 phases with training the Decom-Net first, then Enhance-Net. The whole network is end to end trainable.

The detail steps are as flows:

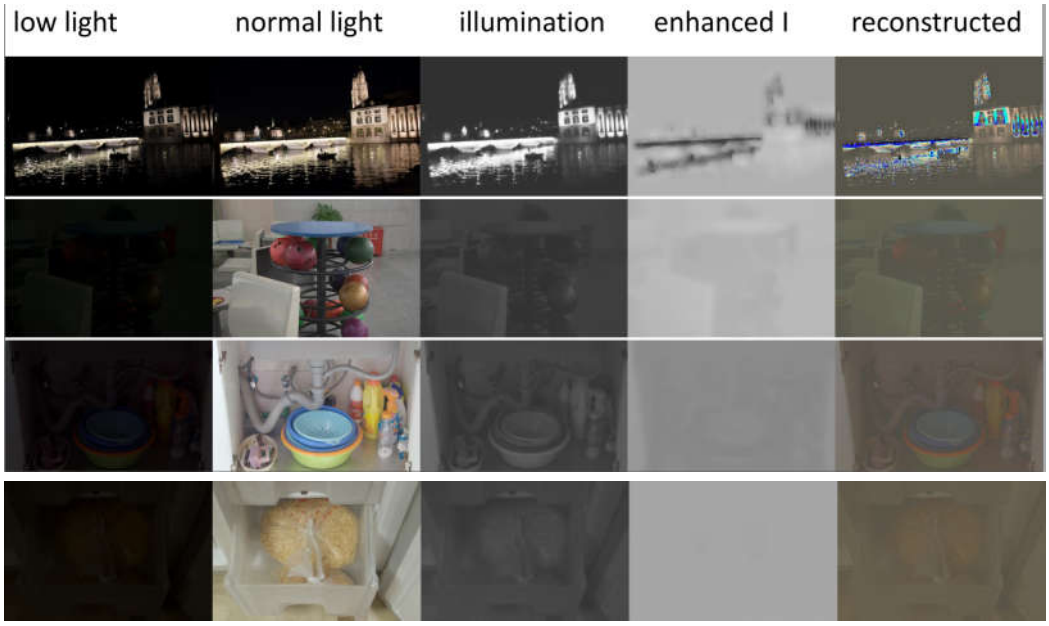
1) Random crop the input image pair into 48*48 patch pair. Make sure crop the same position for the low light and normal light image.

2) Send low light and normal light patch to their respective Decom-Net.
3) Train the network with only the loss function of the Decom-Net. Save the decomposition result and evaluate the decomposed I and R. Then, figure out the epochs needed for decomposition. Evaluating the result of decomposition is subjective, there is no ground truth of which decomposition is best.
4) When the Decom-Net training is done, fix its weight.
5) Use 1) to generate input patch pair, train the Enhance-Net only with its loss function. Check the reconstructed result and find the time to stop.
6) For testing, since the whole network is a fully convolutional network, there is no need to crop test image to fit the input size. Just feed in the whole image is fine. Feed 2 low light images for the 2 Decom-Net. No need to feed normal light image.
7) After get enhanced low light image, do face detection and landmark detection, which will be describe after.

Here is the result of doing decomposition, the low light and normal light image are the input, the reflectance and illuminance are the output.

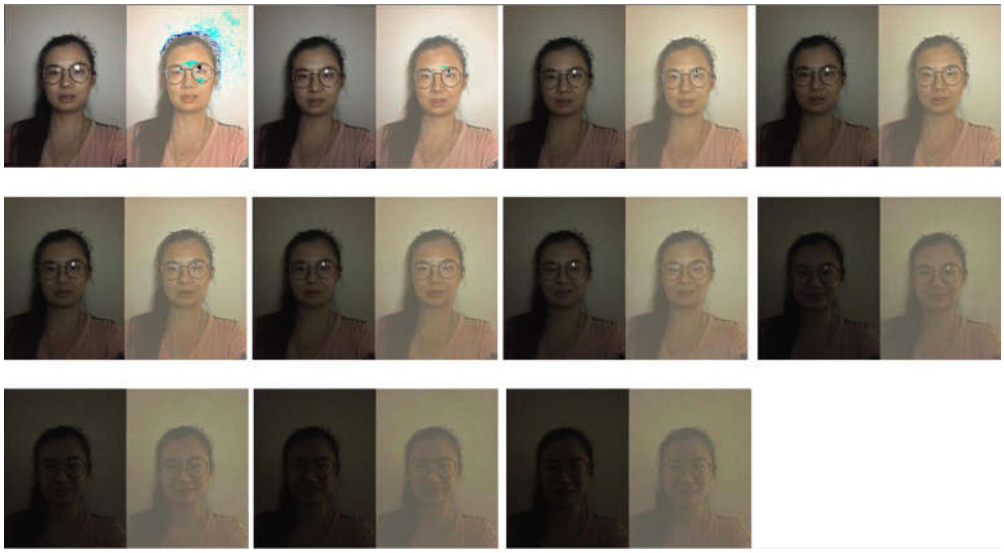


Here is the result of Enhance-Net: the left 2 columns are the input image pair. The illumination is the decomposed result from Decom-Net, the enhance I is the enhanced illumination of Enhance-Net. The last column shows the reconstructed result by multiplying the enhanced I and the decomposed low light reflectance.



Face enhancement

The above result shows this method is effective to generate reasonable enhanced image. Below results shows how it performs when applying on the acquired selfies:



This image contains 11 image pair with the original image on the left and the enhanced image on the right side. From above to bottom, left to right, the lighting condition is getting worse and the image is darker.

From the above result, for low light images, the enhancement result is fine, while for the top left normal light image, the enhanced image contains a lot of noise and green stains. The reason is maybe due to the network only learned to enhance low light image, rather than normal light image.

Face and landmark detection

After we get the relighted selfies, we apply one state-of-art face detection and landmark detection method. The network is well pretrained and easy to use. It comes from the paper *How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks),* which is accepted by ICCV2017. The face detection method is SFD, which comes from paper *S^3FD: Single Shot Scale-invariant Face Detector*.

The test set contains 124 selfies. After do face detection, the original images successfully detected 118 faces out of 124 images. The detection rate is 95.16%. Then I tested my relighted selfies, the result is 84.68%, which is much worse than the original images.

This result is beyond my expectation. After image enhancement, the image quality certainly is better, whereas the face detection rate is dropping. To examine this problem, I did more tests. For this project, I tested 3 versions of enhancement: my implementation of the Retinex-Net, the author's implementation of Retinex-Net, and Gamma enhancement.

As I was reimplementing the Retinex-Net, the author also published their code. So, I also tested the author's result. The Decom-Net and Enhance-Net result could be found on the paper, here I only give the enhancement result on selfies. I also tested Gamma enhancement to further check whether this result on happens on Retinex-Net, or it also happens on other kinds of image enhancement.

Below shows the result of face detection for all these 3 enhancements:

- The original face images: 118/124 = 95.16%
- My relighted faces: 105/124 = 84.68%
- Author's relighted faces: 79/124 = 63.71%
- Gamma enhancement: 116/124 = 93.54%

The best detection rate of enhancement method comes from Gamma enhancement, which is still worse than the original images. Actually, all 3 methods of enhancement show worse result than the original images.

I also tested the landmark detection on all selfies. Below picture shows some of the results. We can see the original low light images get a relatively good landmark detection, the enhanced image may look brighter and easier for human to detect face and landmarks, but it does not help the deep learning model to recognize landmark features. This picture just shows some of the result of landmark detection, some result also shows improvement on landmark detection than original image, which I didn't post here. Overall, the landmark detection improvement is undecided, image enhancement may help, also may not.

The possible reason may as below:

1) The face detection and landmark detection is trained on purely natural images, which are images captured by cameras and no more operation is added. By enhancing the image quality, it may change the input distribution and make the pretrained model perform worse since the input is not the same.

2) For any image, the original one contains the most information. If we build a deep learning model and believe the training is effective, we should count on the model itself to find the way to best fit the input data and let the model decide how to use the data. Thus, human intervene may just make the result worse and any operation may lose some information. Even we do image enhancement, we should still send the original images to the model, and let the model decide which one to use. As below picture, add the dash red line may make the result better.

Conclusion

After done all the tests, the conclusion is as below: Image enhancement is not helping face detection and landmark detection on pretrained deep learning model. For face detection and landmark detection, try to solve the task directly may lead to better performance.

Supplementary material

Paper website: https://daooshee.github.io/BMVC2018website/
My GitHub code repo: https://github.com/stephenkung/FaceEnhancement
Face alignment method: https://adrianbulat.com/face-alignment
BM3D denosing: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.300.5214&rep=rep1&type=pdf