CENG 796 - Peer-review form

Reviewed project ID: Group 02

Reviewed project's title (title of the paper): Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model

Reviewer name(s): Devrim Çavuşoğlu, Ahmet Burak Yıldırım

Instructions:

- Answer = *Yes*, *No* or *Partial*.
- You may expand sections as necessary.
- For most questions, you do not need to add comments, unless the instructions tell you otherwise.
- "Notebook" refers to "Jupyter Notebook" file that is expected to be named as main.ipynb

Question	Answer	Comments
Contains a jupyter notebook file	Yes	
Notebook is located at <project_root>/main.ipynb	Yes	
Notebook's first section contains paper information (paper title, paper authors, and project group members' name & contact information) Some good examples: see group03, group10, group11 (and a couple of other groups).	Partial	Few things need to be addressed: ☑ ~~Paper title.~~ ☐ Paper authors. (I think the provided URL is sufficient, but pointing it out as the guide requires.) ☑ ~~Link to the paper.~~ ☐ Group name. ☑ ~~Group members.~~ ☐ Contact information.
Notebook contains a section for hyper-parameters of the model.	Yes	
Notebook contains a section for training & saving the model.	Yes	
Notebook contains a section (or a few sections) for loading a pre-trained model & computing qualitative samples/outputs.	Partial	There is a shared pretrained model in the codebase (“TDN.pth”), but there are no “shared” pretrained diffusion models.
Notebook contains reproduced plots and/or tables, as declared.	No	
Notebook contains pre-computed outputs.	Partials	There are pre-computed outputs for most of the sections. As far as we observed, the TDN training part has not precomputed outputs.

Data is included and/or a proper download script is provided.	Yes	
Notebook contains a section describing the difficulties encountered.	No	
The paper has achieved its goals and/or explained what is missing.	No	The goals are stated as achieving the qualitative and quantitative (FID, PSNR, SSIM) similarity with the metric results reported on the paper. However, there is no metric calculation in the codebase and the final light-enhanced outputs could not obtained because of the lack of a main model component (incomplete implementation). Only TDN outputs are provided but it is an intermediate network in the framework therefore we couldn't evaluate its performance. The authors did not provide future work or the current status for the version 1 submission (e.g. no `-- version 1 submission –` and no follow up)
The notebook contains a section that reproduces the figure(s) and table(s) declared in the goals.	No	As we mentioned above, the implementation is incomplete. This is why there is no evaluatable outputs in the notebook both qualitatively and quantitatively. Only reflection and illumination maps obtained from the TDN network are provided in the notebook. There is no final image reported.
The notebook also reports the original values of the targeted quantitative results, for comparison.	Yes	
MIT License is included.	Yes	
As the reviewer(s), you have read the paper & understood it.	Partial	● *"In MDLA, for a feature $X \in R^{h \times w \times c}$ obtained from Layer-Norm …"* In Eq. 9, X is defined as the LayerNorm output but in Figure 3, the skip connection is shown through the pre-layernorm features. This is an ambiguity in the paper, in the implementations code authors may have made assumptions on this. However, it would be better to note that in the implementation with comments. ● In the paper, it's not explicitly stated how the reduction/concat/FFN kind of components are carried out, especially in the MDLA part. ● Eq. 10 is ambiguous in terms of weights it looks like (from the eq.) they create one depth-wise conv. weights and use them twice in a single pass (e.g. W_dc * act(W_pc * W_dc * X) + X) ● "*In the illumination decomposition branch, it makes up of several convolutional layers to reduce the amount of calculation on the premise of ensuring the decomposition effect"* In the paper, this seems like the only information about the illumination maps (convnet), and no additional information is given aside from Figure 2.

Implementation of the model seems correct.	Partial	● Eq. 8 is not present in the TDN implementation (FFN description is not explicitly given in the paper, but can assume 1x1 conv., maybe this is the reason they did not implement, but no information is given.) ● “We first aggregate the information of its channel directions with a 1×1 convolution. Subsequently, 3×3, 5×5 and 7×7 convolutions aggregate the information.” is written in the paper. But in the implementation only 1x1 and 3x3 convolutions are used. ● Skip connection at Eq. 9 seems missing. ● Eq. 10 seems to be not implemented correctly, pointwise separable conv. is not present in the attention module (Eq. 10), also skip connection seems to be missing. ● In TDN network, there should be 6 convolution (including the projection) in total but there are 7 convolution layers implemented including the layers named feature extractor and deconvoluzier in the implementation. ● We think that (from 3.1.2.) F_init should be shared in the reflectance and illumination branches, however, in the implementation they are separate. It might be assumed this way by the code authors as there is no explicit mention for that. ● There should be 2 different diffusion models for reflection and illumination branches. In the implementation, only the reflection model is trained. Illumination branch should be added and trained to be able to obtain the final image outputs. ● In the notebook, they set `shuffle=False` for TDN training, can be noted as an undesired setting for the actual training (it might’ve been set and forgotten from experimentation/trials). ● The notebook outputs and training of the diffusion models seems OK (in terms of loss and qualitative outputs), however there is no way currently to make assessment on the complete framework (the illumination diffusion model is not implemented yet). ● Although it is ambiguous in the paper, the paper mentions a “small auxiliary function” (cross multiplication of R & L) in the Reconstruction Loss which is not present in the implementation/notebook. The paper does not explicitly mention this and the statement is a little vague.
Notebook looks professional (in terms of notation, readability, etc.)	Partial	● The images of the “Inference TDN” section can be labeled for understandibility. ● There are unrelated comments under the “Future work for diffusion” section. Also, the section is empty
Source code looks professional (in terms of coding style, comments, etc.)	Yes	As a suggestion, model.train() and model.eval() is used in the diffusion training process, which is confusing because the main purpose is to change the “diffusion” model state which has the ”model” variable inside. Using diffusion.train() and diffusion.eval() could be more appropriate. Both methods are functioning the same but this is a suggestion for readability.

Additional comments:

Overall, the naming convention can be improved, and in the network components naming variables closer/similar to the paper notation/wording might be better and can improve readability a lot.

In the `IlluminationSmoothnessLoss` the code authors applied sigmoid to images after mean operation, we just wondered why not it's a simple normalizaiton to range (0,1). We are not claiming sigmoid is incorrect.