AIClub Project Member Application

Pranav S Ramanujam

EP23B038

1 Section A: Managerial Questionnaire

1.1 Motivation

I wish to join the AI Club as I have an interest in AI and would love to work in this field. I picked up an interest in AI in my DC tenure at the AI Club. I enjoyed the post session tasks and reading / understanding whatever I could about things related to what was taken in the sessions. I feel I am a decent fit for this position as I am not starting AI from scratch and was a DC at the club as I mentioned before. I am also hard working and I want to do something substantial in AI, which I can achieve by being a part of the club.

1.2 Commitments / PORs

I feel I can devote 25-30 hrs a week for AI club work. I feel so as I do not plan on taking up any other POR. I will have only academic committments other than this, which I can manage.

2 Section B: Common Technical Questionnaire:

2.1 Quantile Regression:

The notebook which contains the code has been hyperlinked below.
Link
In the notebook, I have implemented a simple Neural Network, without any

activation functions. The input layer contains 4 nodes, but in the sample template, the input size is 2*4. Hence I have assumed that we are giving two inputs to the network simultaneously, and have added the Quantile losses corresponding to each of the inputs to find total Quantile loss.

I feel the Quantile for fastest convergence depends on the input data X and Y and have included some code to find the quantile for fastest convergence for a given X and Y.

3 Section C: Project Specific Questionnaire

3.1 Retinex

3.1.1 Advantages of Retinex

The main goal of Image Enhancement Algorithms is to modify the image into such a form that resembles the way the scene contained in the image would appear to a human. Keeping this in mind, I think Retinex is used extensively in deep learning for the following reasons:

- Retinex is based upon the Retinex theory that proposes the following: The perception of brightness and colour in a scene is influenced by the product of the reflectance of the objects and the Illumination falling on them. This theory also plays a role in the explaining human vision. Hence, it is widely used in many deep learning models.
- The main idea of Retinex is the separation of the reflectance and illutmination components of the image and enhancement of the reflectance component. This makes it a relatively simpler model.
- It has applications in various areas such as low light image enhancement, enhancing colour constancy, improvement of brightness and contrast etc.

3.1.2 Reflectance, Illumination Map and Corruption

The Retinex theory proposes that any image is the product of its reflectance component and Illumination component. Reflectance, Illumination and Corruption are given as follows:

- Reflectance: This is assumed to be the component of the image which contains the intrinsic and salient features of the image.This component encapsulates the inherent attributes of objects, remaining uninfluenced by external illumination.
- Illumination: This is the component of the image dealing with the background and setting of the image. It is often considered to be the low frequency component of the image and is sometimes approximated to be the image obtained by performing Gaussian Convolution over the image in simpler algorithms. The illumination represents the dark background in case of low light images.
- Corruption: In Retinex, the term corruption refers to the artifacts that can be introduced in the enhanced image while using Retinex algorithms. The separation of the reflectance and the illumination is not always perfect, especially in the case of complex lighting. This may represent in the formation of halos around object edges while separating reflectance from illumination.

3.1.3 Deep Learning Models with Retinex

Given below is a deep learning architecture which employs Retinex in the enhancement of low light images:
Overall framework:

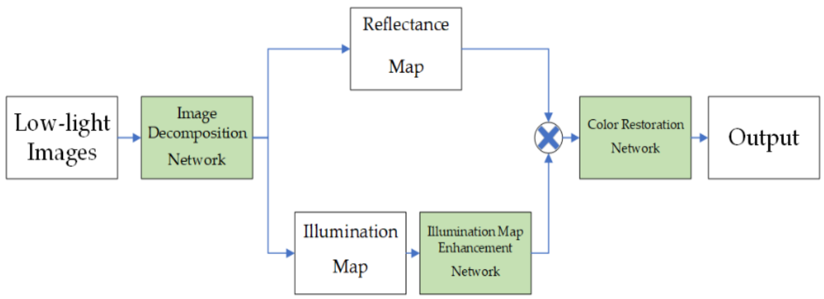
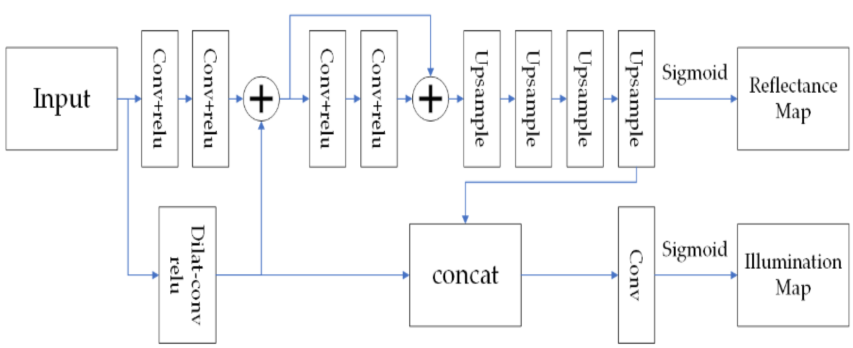


Figure 1. Algorithm framework diagram.

It consists of:

- Separation of the image into its Reflectance and illumination component.
- Enhancement of the Illumination component which is what we need in low light image enhancement.
- Recombine the improved illumination and original reflectance maps.
- Pass it through a Colour restoration network to avoid the introduction of artifacts and colour distortions (halo effects etc).

Structure of the Decomposition Network:



- Each of the Upscale corresponds to upscaling or increasing the resolution of its input by using interpolation techniques.

- Sigmoid activations are provided at the end to compress all the values to the range (0, 1)

The structure of the illumination enhancer is as follows:

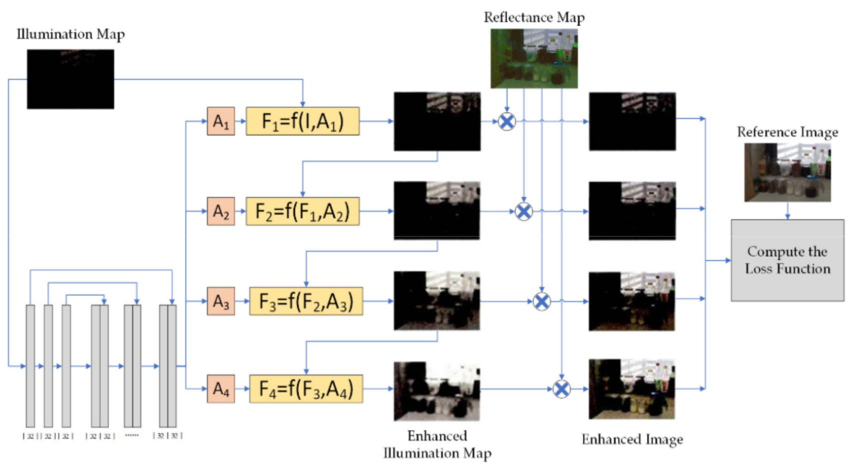


Figure 3. Structure of illumination map enhancement network.

where:

- Each A_n corresponds to the parameters of the mapping functions F_n.
- $F_n(I(x, y))$ for $n > 1$ is given by:

$$F[I(x,y)]_n = F[I(x,y)]_{n-1} + A_n \cdot F[I(x,y)]_{n-1} \cdot \{1 - F[I(x,y)]_{n-1}\}$$

For $n = 1$ we use:

$$F[I(x,y)] = I(x,y) + \alpha I(x,y)[1 - I(x,y)]$$

where $I(x, y)$ is the illumination map of the image

- The A_n are probably obtained by applying convolution filters over the illumination map to obtain an output with the same size as the illumination map. (In the research paper it is not clearly specified as to how the A_n values are calculated.)

The structure of the Colour restoration network is as follows:

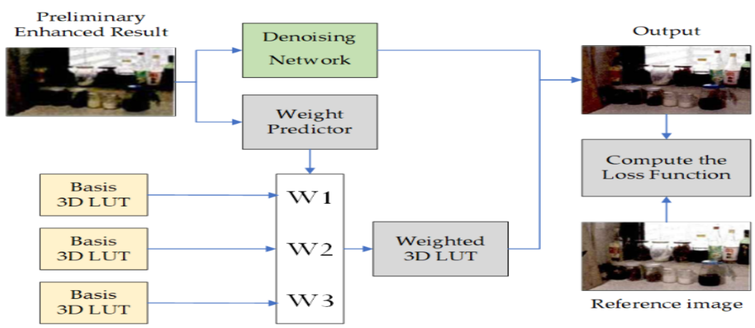


Figure 4. Structure of color restoration network.

The Denoising network is made of:

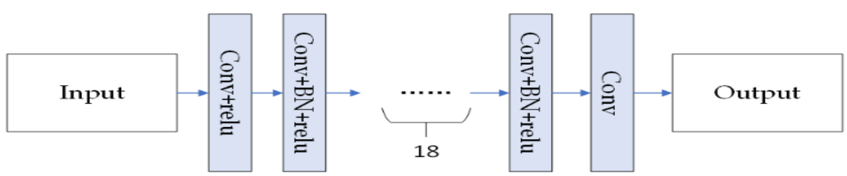


Figure 5. Structure of denoising network.

- The 3D LUTs are used to bring transformations to the original image.
- The output from the 2 channels are used to get the final image output.

Other simpler algorithms which emply Retinex to extract the reflectance include SSR (SIngle Scale Retinex), MSR (Multi scale retinex) , MSRCR (multi scale retinex with colour restoration).

Research paper explaining this deep learning model: "Low-Light Image Enhancement Algorithm Based on Deep Learning and Retinex Theory"

3.2 U - Net

3.2.1 Up Scaling Of the Image

U - Nets are mainly used for semantic segmentation.
The Encoder part is responsible for the feature extraction, whereas the decoder does the up - scaling of the Encoder output feature map so that the final output of the U - Net is a segmentation mask of the same size as the original image.
A segmentation mask is the output of an image segmentation algorithm, in which each pixel of the image is assigned to a certain class. The Encoder part outputs a segmentation map of smaller sizeas it extracts features by Convolution and Pooling. Important Spatial information is lost in this. Hence the up - scaling in the Decoder part is important, so that our final output is a segmentation mask of the same size as the image, which will ensure that all the objects are in the correct position in the image.
The scaling up is achieved by using various techniques such as transpose Convolution and skip - connections.

3.2.2 Skipped Connections in U - Nets

In the feature extraction stage of the U - Net, it is possible that to capture all the features thoroughly, the original image must be passed through numerous Convolution and Pooling layers. This results in a very small feature map outputted by the Encoder, which contains information about the various features present in our image.
However in such cases where the original image is passed through many layers to get encoded, there is a possibility of loss of spatial information.
Hence skipped connections were introduced. These introduce a direct link between stages of the decoder and the corresponding stage in the encoder. By means of skipped connections, the input to the next stage is fed as the concatenated matrix, which is made up of the previous output in the decoder stage and the output in the corresponding stage in the Encoder. Since the Encoder output is from one of the earlier stages of the feature extraction, it will provide spatial information about the segments of the image.

3.2.3 Skipped Connections in ResNets

The use of Skipped Connections in U-nets is explained above.
Skipped connections howeversolve a different problem in ResNets. ResNets are usually very deep models, which are deeper than U-Nets. Hence there is a high chance of encountering the vanishing Gradients problem (The activation they use is ReLU, hence the exploding gradients problem is avoided). The Resnet is made of numerous residual blocks which consist of convolution layers, activation functions etc. The output of the previous block serves as input to the current block. Now by means of the skipped connections, the final output of the current residual block is the sum of the output of this block and the output of the previous block. This way, the weights and biases of the earlier layers have a more direct influence on the final output and hence the vanishing gradient problem is mitigated, as the gradient of the loss with respect to earlier weights will not consist only of long chain rule differentiation.

3.2.4 U - Nets in Low light Media enhancement

In my view, U - Nets are used in low light media enhancement as they can perform semantic segmentation of an image. Hence they can be used to extract the useful features which get hidden in low light images.

3.2.5 Attention Is All You Need

Humans process information which they pay attention to. The Triesmann theory, which explains the human attention mechanism, states that Attention in humans is an attenuating filter. Out of all the stimuli in the sensory buffer, only the stimuli left after filtering are processed further.
Attention in Computer Vision also works on the same lines. By means of Attention, the model will know which parts of the image or input it needs to give importance to. That way, many computer vision tasks like image classification, image segmentation, face detection etc will be simplified.
The basic types of Attention mechanisms in Computer Vision are:

- Spatial Attention:
  In this form of Attention, the model identifies the relevant spatial regions in an image.

- Channel Attention:
  In this form of Attention, the model identifies the important channels or feature maps present in the image.

  The above two forms of Attention can be implemented together also as in BAM and CBAM.

3.2.6 Challenges

Although Attention Mechanisms are very useful, they are very complex. Some of the disadvantages of attention in my view are:

  – Attention Mechanisms are complex to implement
  – They also introduce many new weights and biases, and hence training time is increased and the computation power requirement is high.

3.2.7 Applications

Attention mechanisms will be of great use in this field as in Low light images, only some portion of the image contains important features whose visibility needs to be improved, and attention is all about focusing on selective regions of the input.
Attention Mechanisms are also of great use in NLP in the Seq2Seq models such as translators etc. Here, when the input sequence gets very long, the context vector fails to capture all the details of the input. Hence a weighted sum of all the outputs at the Encoder stage (this is implemented using attention) along with the current decoder output are used as input to the next stage of the decoder in translator models.

3.3 Section C: Coding questionnaire

3.3.1 VGG19

Details:

- Notebook: **notebook**
- I have made a scaled down version for 32*32 images.
- To make it similar to VGG19, I make the image pass through 2 convolutions at a time, with number of output channels being multiplied by 2 in the second layer.
- I used ReLU activation like in VGG19. Following this, I have 3 Linear layers.

3.3.2 LeNet

Details:

- Notebook: **notebook**
- As LeNet is for 32*32 images, I could replicate it. I have used the same number of Convolutional and linear layers as in the actual architecture and used tanh activation.

3.3.3 AlexNet

Details:

- Notebook: **notebook**
- I sclaed up images to 64*64 to incorporate the use of larger kernels in initial stages.
- For the convolutional layers, I decreased the kernel size for the layers in further stages, as in AlexNet. I have also used padding, to prevent changes in size of the image.
- I used ReLU activation like in AlexNet. I have also incorporated 2 Dropout layers (For regularization) and 3 linear layers.

3.4 Section D: Approach

3.4.1 Thoughts and Solutions Regarding the Problem

My approach to the problem would be:

- We should begin by choosing one of the well known methods for low light image enhancements such as Retinex.
- We should then implement simple algorithms related to the chosen method, for eg if it is Retinex, then implement things like SSR, MSR, MSRCR, IRIE etc and see results.
- Then we can move on to implementing some known deep learning approach related to this method.
- Then we can also try out other methods like Histogram Equalisation similar to how we analysed the first method (start small and expand).
- Using all the knowledge gained, we can then come up with something new.

3.4.2 Domains

Machine Learning domains this includes are:

- Computer vision
- Deep Learning

3.4.3 Challenges

The problem we are taking up is definitely a challenging one. Hence we will face problems in the implementation of Deep Learning architectures for low light image enhancement. I think the only solution to this would be to read research papers about existing architectures thoroughly and try implementing them first, make variations in them and analyse results, and then move on to implement something of our own.
Collaborating with the club's faculty advisor would also be helpful.

3.4.4 Applications

Low light image enhancement has many real world applications such as:

- in Surveillance (CCTVs)
- Celestial Body imaging
- Underwater imageing etc