AI CLUB PM APPLICATION : NIGHT VISION

Sreeram R

April 2024

Managerial Questionnaire

1 Essentials:

AI/ML is something which has fascinated me. My deep interest in the domain of AI/ML is what has motivated me to apply to the role of a project member for the project Nightvision. I felt extremely interested and enthusiastic when I got to learn about the project, its applications and wide scope of research that it has. I believe the interest in the domain of AI/ML and the enthusiasm within me to learn new things, to learn from my peers, the desire within me to have a high learning curve makes me fit for this role. I believe good communication skills, leadership qualities. motivating my peers to work on a common goal are some of the qualities which I have. With my motivation and my work ethic,I believe that I would be able to contribute to project as well as the club to a great extent.

2 Commitments / POR's:

Considering the fact that I will not be having any other POR's in my second year and my motivation to apply for this role, practically speaking I would say 7-8 hrs weekly. It might change a little here and there due to various reasons but I feel on an average 1 hr a day is something which I would certainly be able to commit to AI club. I am also applying for the role of a coordinator. In case I need to spend more hours since I will have added responsibilities as a coordinator if I get selected, The fact that I am extremely motivated and self driven, I believe that I will be able to spend even more hours if required.

Common Technical Questionnaire

3 Quantile Regression

3.1

A neural network with the given specification has been trained using quantile loss using pytorch and the link to the google collab notebook is here. I played around with the τ values and found $\tau = 0.05$ to be doing good. Values between 0.1 and 0.2 also seem to do fairly well.

Project Specific Questionnaire

4 Retinex

4.1

- Retinex algorithms draw their inspiration from the functioning of the human visual system. In the human visual system the image is formed with the help of human eye (retina) and human brain (cortex), hence comes the name Retinex. These algorithms seperate the image into two components, Illumination

and Reflectance. Retinex algorithms have a wide scale of adaptability and has applications in a wide variety of domains like surveillance, astronomy, video surveillance, autonomous driving etc.

- The ability of the Retinex based algorithms to decompose the image into two components, Illumination and Reflectance is a great feature extraction mechanism and this can help the deep learning models in detecting complex patterns, being more efficient in the desired task of image enhancement. Also the adaptability of the deep learning models based on Retinex algorithms and its wide scale applications in various domains as said before makes it a powerful tool in enhancing the visibility of low light images.

4.2

- By Retinex theory, An image can be seperated into two components, Illumination and Reflectance. Illumination depends on the source of illumination whereas Reflectance depends on the nature of the object in the image. Since having information about illumination and reflection is not feasible in practice, We come up with the idea of convolving the image with an appropriate smoothening filter (for example Gaussian filter) and using the smoothened image as illumination. The equations given below are for the Single Scale Retinex (SSR).

$$I(m,n) = L(m,n) \times R(m,n) \tag{1}$$

$$R(m,n) \approx log(I(m,n)) - log(L(m,n)) \tag{2}$$

$$R(m,n) \approx log(I(m,n)) - log(F(m,n) * I(m,n)) \tag{3}$$

  - where $I(m,n)$ represents the distribution of image.
  - $L(m,n)$ refers the the illumination component.
  - $R(m,n)$ refers to the reflectance component.
  - $F(m,n)$ represents the surround function of the filter.

- In the third equation the illumination component is rewritten as the convolution of the surrounding function of the filter and the input image. Its called single scale because of the fact that here we are using only one surround function. In the case of Multi Scale Retinex (MSR) and other better models we use more than one surround function and we do weighted summation of the results of the reflectance for different surround functions.

- Illumination map and reflectance image are kind of complementary components. Illumination map represents the spatial distribution and intensity of the light which gives a fair idea about the brightness of the scene. The variations in the brightness in the illumination map gives idea about lighting conditions in the scene. Illumination maps are useful for tasks like Dynamic range compression, Shadow removal etc. As briefly mentioned before, Illumination maps are obtained by performing convolution operation of the image with an appropriate smoothening filter (like the Gaussian filter) and this removes the reflectance component (colours, textures and other fine details) of the image.

- Reflectance image represents the properties of the objects and the surfaces in the scene like colour, texture etc. A reflectance image is obtained by removing the illumination map from a given image and this is mathematically depicted in the third equation. Reflectance image preserves the fine details of the objects and this really helps in analysing and understanding the events in the image/ footage. Both reflectance image and illumination map together help us in image enhancement.

- In the context of Retinex, Corruption refers to the image degradation caused by noise, shadow, uneven illumination and various other disturbances. Noise can corrupt the appearence of the image by bringing in some random variations in the pixel values of the image. Uneven illumination is yet another problem since it can cause problems with brightness and contrast of the image. The aim of image enhancement algorithms like Retinex is to do away with corruption (image degradation).

4.3

There are several deep learning models which use Retinex based techniques. One of them is

- U-Net based architectures using Retinex algrithms:
  - In U-Net based architectures we have a contraction path (encoder network) and an expansion path (decoder network) and skip connections. I have elaborated on the architecture and the functioning of various parts of the U-Net in question 5. The encoder path is responsible for feature extraction and it reduces the spatial dimensions through various maxpooling layers. The decoder path reconstructs the illumination and reflectance components from the encoded features, with the help of skip connections which help in regaining the information which might have been lost due to maxpooling. During the upsampling process, the decoder network focuses on reconstructing the low-frequency components of the input image and during this the decoder network effectively estimates the overall lighting conditions present in the image. This way we get the illumination map and once we get the illumination map we can get the reflectance map by removing the illumination map from the image since by Retinex theory an image has two components Reflectance and Illumination. Attention mechanisms along with Unet architecture based on Retinex algorithms help in focusing on the important parts and enhancing the contrast of the low lighted image which helps in getting a greater understanding of the image.
  - I found two interesting research papers on Low light image enhancement using Retinex based network with attention mechanism. The first research paper I referred to is here . I found it to be very relevant to the project and extremely interesting. Another research paper I found is here.

5 U-Net

U-Net is a Convolutional Neural Network Architecture which is popularly used for the image segmentation task. It is U shaped encoder decoder network consisting of 4 encoder blocks, 4 decoder blocks, connected by a bridge.

- **Encoder network** consists of a series of convolutional layers followed by max pooling layers. Each convolutional layer is followed by a ReLU (Rectified Linear Unit) layer which is responsible for bringing in non linearity into the network helping it to learn complex patterns. Then it consists of max pooling layer which extracts feature and also reduces the spatial dimensions. Since the Encoder network captures the features of the input image at multiple scales and reduces its dimensions, its also called Contraction path. The encoding network is responsible for extracting all the important
- The **bridge** connects the the input f encoder network and the decoder network. It consists of a series convolutional layers followed by ReLU activation function. At this stage we get a feature map which has reduced spatial dimensions but consisting of a greater number of channels. The feature map contains a lot of features which are extracted throughout the contraction path.
- The **decoder network** also called the expansion path. It consists of upsampling layers followed by convolutional layers. Since the contracting path involves several steps of downsampling there is a chance of loss of information. So to account for this connections are established between the contraction path and the corresponding layers of the expansion path. The feature maps captured in the contraction path contains a lot of features and valuable information and these are cropped and concatenated with the upconvolved feature map. The connections between the expansion path and the corresponding contraction path layer is known as **skip connection**. These skip connections help in recovering the spatial information which might have been lost due to down sampling in the contraction path. Then the final layer consists of a convolutional layer with a softmax function which generates the segmentation mask.

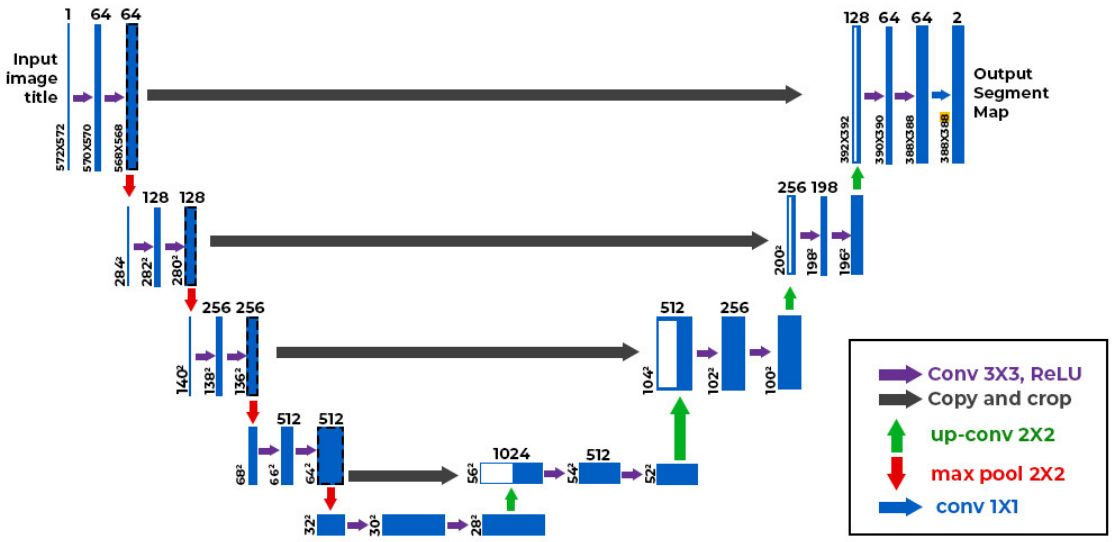

Figure 1: U-Net architecture

5.1

Scaling up of the image takes place in the expansion path (decoder network). In the contraction path, the input image undergoes several max pooling operations which might lead to loss of important information. Scaling in the expansion path helps in regaining this lost information due to downsampling through skip connections where the feature map of the corresponding layer in the contraction path is cropped and concatenated with the upconvolved feature map. This increases the efficiency of the architecture in tasks like image segmentation. Also the fact that additional convolutional layers present in the expansion path helps in enhancing the features of the image, increasing the efficiency in the task of image segmentation. Scaling the image also helps in localizing objects more accurately and produce segmentation masks with clear boundaries.

5.2

In the context of U-net Skip connections are the connections between the contraction path (encoder network) and the expansion path (decoder network) which pass a feature map of the contraction path to the expansion path where the passed feature map is cropped and concatenated with the upconvolved feature map helping in regaining the lost information due to maxpooling. Due to downsampling in the contraction path, there is a chance of loss of information. Skip connections help in recovering those lost information.

5.3

- To come up with a solution to complex tasks like image recognition, most of the times of we stack up some additional layers into our deep neural network with the expectation that it helps to learn complex features of the image. But it was observed that there is a maximum threshold of depth (number of layers stacked) and adding layers further degrades the performance and accuracy of our model due to the vanishing gradient problem. Hence Microsoft research experts came up with a new architecture called ResNet. Resnet is made up of residual blocks where each residual block contains some convolutional layers. Like Unet, there are skip connections in Resnet as well, though their objective and purposes are different in these two architectures.

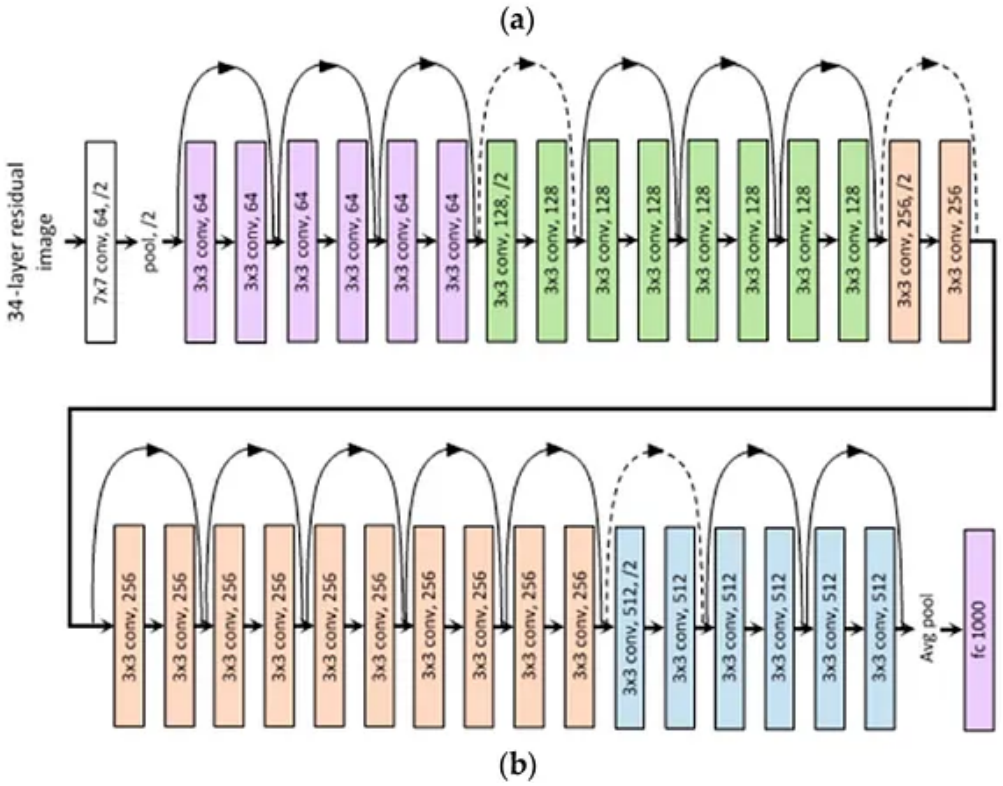

Figure 2: ResNet architecture

- Skip connections in ResNet bypass the convolutional layers in the residual block and connect the input to the output of the residual block. Adding the input to the output of the residual block helps in solving the vanishing gradient problem and also ensures that the higher layer will perform atleast as good as the input layer an not worse as in the case of deep neural networks.
- In the case of Unet, skip connections help in regaining the information which might have lost due to maxpooling (downsampling) in the contraction path, by cropping and concatenating the feature map of the corresponding layer in contraction map with the upconvolved feature map of the expansion path. Skip connections in ResNet are designed to facilitate residual learning by allowing the network to learn the difference between the input and output of a block, thereby making it easier to train deeper networks. There is a difference between the adding and the concatenating operation. In adding operation, the feature maps are added together ie. the corresponding element values in the tensors are summed up. Whereas in concatenation operation , they are combied along a new dimension, basically the feature maps are stacked together icreasing the channel depth.For example if you concatenate a feature maps with depths 64 and 32 then you get a resultant feature map of depth 96.
- Also the these two architectures U-Net and ResNet are used for different purposes. U-Net architecture is image segmentation focused whereas ResNet architecture is focused on taining deep neural networks and solving the vanishing gradient problem.

5.4

- U-Nets play a very important role when it comes to the task of low light video enhancement. U-Net helps in feature extraction which helps in getting a greater understanding of the scene. The segmentation masks that we get out of U-Net helps in seperating the objects from the background. The skip connections in the U-Net architecture helps in complex feature extraction. Also the adaptability of the U-Net architecture helps in generalizing it to various low light scenarios and lighting conditions.

- Segmenting out the objects of interest from the footage by U-Net helps in applying various image enhancement techniques like for example Retinex algorithms with attention modules (which focuses on objects of interest), along with U-Net can be used to increase the clarity, contrast, visibility and overall appearence of the image. One of the research papers attached in the question 4.3 is about the application of retinex algorithms with attention modules along with U-Net architeture and how it is used in low light video enhancement. I found this research paper very relevant to our project and extremly interesting.

6 Attention Is All You Need

6.1

- Attention mechanisms are crucial components in the design of neural networks especially in the field of Natural Language Processing and Computer Vision. It mimics the human brain's ability to pay attention to more important things than the less important ones. For example let us consider a scene of a social gathering, and you meet your friend and you talk to him/her. A lot of people are talking at the same time but still you will be able to understand what your friend wants to convey, this is because you are paying more attention to the conversation than the background noise.

- This is exactly what attention mechanisms do when it comes to the field of computer vision as well. Attention mechanisms allow our model to selectively focus on the parts of the input that are most important for making a prediction, and ignores the less important parts.This is done by focusing on specific important parts/areas of the input image by assigning weights to different elements in the input indicating their relative importance.This helps in discriminating various regions of the image based on the weights and help in complex feature extraction, therefore plays a very important role in image segmentation, object detection, image classification and many other tasks in the domain of computer vision. For example in the image classification task , attention mechanism helps in focusing more on the object and ignoring the rest.

6.2

Though Attention mechanisms have many advantages, there are also some challenges when it comes to implementation.

- These increase the complexity of the model which can make the training and optimization of model more difficult.

- The flexibility and adaptability of the model is severely affected since the weights or the priorities might change with different inputs.

- Overfitting is yet another problem when it comes to attention mechanisms. The model might perform well on the training data but might struggle when it comes to generalizing it.

- Also, if inappropriate weights are assigned, It can lead to ambiguous outputs when in comes to image classification, image segmentation and various other tasks.

6.3

- When it comes to low light video enhancement, Attention mechanisms can help in enhancing contrast in the image by focusing on the more important regions within each frame. They help in dynamic range compression which helps in increasing the visibility of a low light image or video. In some cases the noise levels in the image can be very high which might be responsible for image degradation. Attention mechanisms can help in applying dynamic noise reduction techniques only to those regions of the image with noises/ disturbances. They can also help in identifying the important regions of the image and increasing brightness of these regions by choosing appropriate weights might help in getting a greater understanding of the scene.

- Attention mechanisms also have great applications in various other domains of AI/ML especially in Natural Language Processing (NLP) tasks like Machine Translation, Speech recognition and Computer vision tasks like image enhancement, object detection and also for other tasks like Music generation and speech recognition.
- In NLP tasks, attention mechanisms help in tasks like translation by selectively focusing and giving more importance to certain word or phrases in a sentence. This increases the model's efficiency in getting an in depth understanding and clarity about the context.
- It also plays an important role in various computer vision tasks like object detection. Google streetview's identification of house numbers is a great example of application of attention mechanism in the field of computer vision.
- It also has application is Music generation tasks like creation of melodies. The attention mechanism helps in focusing on more relevant musical sounds and in order to create pleasent and coherent music.

7 Coding Questionnaire

I implemented the VGG19 architecture from scratch using pytorch to classify images of the CIFAR10 dataset. The google collab notebook link is attached here

8 Section D: Approach

8.1

I believe application of Retinex based algorithms along with attention modules and using the U-Net architecture and training the model and testing the generalizability of our model using various datasets should be our approach towards this project. This is just a brief explanation of an idea which has immense scope of innovation and research. A pipeline solution is suggested below

- First step should be to collect data and have a datset with wide variety of low light images along with their high quality versions. Data preprocessing and splitting dataset into training dataset and testing dataset should be done.
- Then we can implement the retinex based algorithms on our dataset. A Wide variety of algortihms like SSR (single scale retinex), MSR (multi scale retinex) , MSRCR (multi scale retinex with colour restoration) appropirately. This will help to get a greater understanding of the image / scene.
- We can design a U-Net architecture for enhancing the image. The contraction path (encoder network), the expansion path (decoder network) along with the skip connections helps in image segmentation which can help us in distinguishing the object from the background and this along with the attention mechanisms when implemented should increase the visibility, contrast and the appearence of the image.
- After this we need to train our model with the training dataset, compute the loss and minimize it. We might to need to use techniques like dropout to prevent overfitting which might lead to the model performing very well with the training data but poorly with the testing data.
- Finally we need to check how good our model is by implementing it on the testing dataset and make further improvements in the architecture or algorithms as per the need.

8.2

These are the domains of AI/ML which we will be exploring while solving this problem.

- Deep learning
- Computer vision (Image processing, Image segmentation)

- Attention mechanisms
- Retinex algorithms

8.3

These can be the challenges that we might face while we work on this project

- Coming up with large data of low light video images with high quality images of the same scene might be difficult. Sometimes we might have complex lighting conditions in our image which might make it difficult for our model to train for ie. the adaptability of our model is something which we should be trying to increase. This can be yet another challenge.In applications like surveillance, real time analysis of the scene requires efficient algorithms with less computational complexity and our architecture must be robust to handle that.
- Some of the solutions which I can think of are coming up with better architecture and innovative methods which might address the issue of real time analysis. We need to focus on optimizing our architecture to decrease the computational complexity for this. Implementing domain adaptation techniques and self-supervised learning methods can increase the efficiency and adaptability of our model.

8.4

The project has a wide variety of applications in Surveillance, Cinematography, Astrophysics, Medical imaging etc. This makes this project extremely interesting. I believe the best resources to start with are the research papers in the domain of Low light video/image enhancement. Well I believe studying these research papers would for sure give us insights on how to take this project forward and make it a success. Collecting a wide variety of low lighted footage with the high quality image is something which we should focus on which would help us build robust architectures and implement algorithms which can increase the adaptability of our model to a great extent. I also believe that collaborating with companies who are working on the same domain in IITM research park (if possible) might help us come up with great ideas and take this project to greater heights.