Problem Solving Question

Question 1 – Imaging Science

Please take single grayscale photograph of a textured surface (such as paper, fabric, or a wall) that appears non-uniformly lit due to uneven illumination — some areas are brighter and some are shadowed.
The challenge is to reconstruct the true texture pattern of the surface as if it were uniformly illuminated, without any prior knowledge of the lighting source or direction.

1. Model Formulation

Assume the observed image I(x, y) is formed as:

$$I(x, y) = R(x, y) \times L(x, y)$$

where

1. R(x, y) = reflectance (the actual texture pattern you want),
2. L(x, y) = illumination (a smooth, slowly varying function).

   Explain why simple histogram equalization cannot recover R(x, y) and propose a mathematical or algorithmic strategy to estimate both R and L from a single image (without knowing either).

2. Write a programming code in Python or Matlab that converts the image in to log domain and separates high and low frequency component manually without any inbuilt function and recover an estimate of the reflectance image $\hat{R}(x, y)$.

3. Now suppose the image is color and illumination varies spectrally in red, green and blue colour space. Convert your algorithm to jointly correct illumination across channels which preserving true color ratios.

Question 2 – Computer Vision

Take a single photograph of a planar rectangular grid (like a tiled floor or printed checkerboard) taken with an unknown camera and an unknown radial distortion. The grid may be partially occluded, captured at an oblique perspective, and the image may contain moderate noise or lighting variation. Use whichever distortion model you prefer, but state it precisely

Your task is to estimate the camera radial distortion parameters

1. Formulate a robust cost function to estimate the distortion parameters as well as camera intrinsics, extrinsics and principal point if needed.
2. Implement a robust optimation pipeline to refine distortion parameters.
3. Use RANSAC to remove outliers
4. Using the estimated parameters, undistort the image and compute an undistorted grid (image of grid corners on an undistorted plane).
5. Reproject the undistorted grid back into the distorted image and compute residuals between detected features and model-predicted positions and compute the reprojection error.

Technical Aptitude

Question – Deep Learning

You must use Python for your implementation.

You can run the code in a jupyter notebook on Colab/Kaggle by enabling GPUs.
You must also provide a link to your GitHub code.

Follow good software engineering practices The commits in GitHub should reflect how the code has evolved during the course of the assignment.

In this assignment, you will experiment with a sample of the Aksharantar dataset released by AI4Bharat.

This dataset contains pairs of the following form:
x,y
ajanabee,अजनबी
i.e., a word in the native script and its corresponding transliteration in the Latin script (how we type while chatting with our friends on WhatsApp etc). Given many such $(x_i, y_i)_{i=1}^{n}$pairs your goal is to train a model $y = \hat{f}(x)$which takes as input a romanized string (ghar) and produces the corresponding word in Devanagari (घर).
As you would realize, this is the problem of mapping a sequence of characters in one language to a sequence of characters in another. Notice that this is a scaled-down version of the problem of translation where the goal is to translate a sequence of words in one language to a sequence of words in another language (as opposed to a sequence of characters here).

Read this blog to understand how to build neural sequence-to-sequence models.

Build a RNN based seq2seq model which contains the following layers: (i) input layer for character embeddings (ii) one encoder RNN which sequentially encodes the input character sequence (Latin) (iii) one decoder RNN which takes the last state of the encoder as input and produces one output character at a time (Devanagari).

The code should be flexible such that the dimension of the input character embeddings, the hidden states of the encoders and decoders, the cell (RNN, LSTM, GRU), and the number of layers in the encoder and decoder can be changed.

What is the total number of computations done by your network? (assume that the input embedding size is m, the encoder and decoder have 1layer each, the hidden cell state is $k$for both the encoder and decoder, and the length of the input and output sequence is the same, i.e., T, the size of the vocabulary is the same for the source and target language, i.e., V)

What is the total number of parameters in your network? (assume that the input embedding size is m, the encoder and decoder have 1layer each, the hidden cell state is $k$for both the encoder and decoder, and the length of the input and output sequence is the same, i.e., T, the size of the vocabulary is the same for the source and target language, i.e., V)

----------x----------