WiDS
Assignment

1) This article discusses a technique for colourizing black and white photos using neural networks. The author Emil Wallner aims to explain how to build a colorization neural network step by step with increasing complexity.
The problem is broken down into three stages - an alpha, beta, and full version of the colorization model.

- In the alpha version, we create a 40-line neural network to colourize a single image as a proof of concept. This helps us get familiar with the core syntax and concepts.

- In the beta version, we develop a network that can generalize and colourize images which it has not seen earlier during training.

- In the full version, we combine the neural network with a pre-trained classifier to incorporate object recognition and produce more realistic colours. We train this final model on portrait images from Unsplash.

Author first explains some of the technical background of the situation **:** Black and white images are represented as grids of pixel values from 0-255 indicating brightness. Color images have three channels - red, green, and blue. The task is to link grayscale values to these color channels using a neural network. But mapping this grey scale image to 3 different colour channels can be a bit hard so a better way to implement this is to use another colour space called Lab. This also involves 3 layers, grey image, red-green and blue-yellow. Grey image is the representative of brightness which too plays a significant role in colourization.

For the alpha version code, the author loads a sample image, converts it to the 'Lab' colour space, and extracts the grayscale layer as input and the 'a' and 'b' colour layers as output. A simple convolutional neural network is designed that takes the input and predicts the outputs. After training on the single image, it can colourize that picture well.

In the beta version, the author takes multiple images as input instead of one unlike the previous model. He splits the data into training and testing sets. A deeper convolutional network is built to encode more complex patterns from the input images. It is trained on multiple pictures to generalize colouring rather than just memorizing/overfitting one image.

For the full version, he uses Keras' functional API since it combines multiple models. He downloads Inception ResNet V2, a powerful pre-trained classifier and extracts its final classification layer. In parallel, the input images are passed through this classifier and the colorization encoder.
Their outputs are fused together through a "fusion layer". By transferring learning from the classifier to the colouring network, it can match object representations to colouring schemes. The fused output passes through a decoder to make the final colour predictions. This architecture is trained on portrait images with data augmentation techniques.

Apart from all this, he also analyzes and shares some key lessons learned.
- Simplifying problems into smaller contexts helped address errors more systematically.
- A diverse dataset yields unsatisfactory brownish results, so he focused on very similar training samples instead.
- Maintaining proper image sizes and ratios is crucial, as networks rely on pixel locations rather than just information content.

- Publicly breaking down challenges motivated self-guided solutions.
- Asking questions to online communities proved less helpful than directly contacting researchers.

At the end, he provides recommendations for continuations or modifications of this colourization process ahead.

2) The neural network architecture used in the article in the final version of the colourization model involves a multi component design including encoder, decoder and a fusion layer. The encoder is responsible for extracting features from the input grayscale images, while the decoder generates the colourized output.
Along with the encoder, input images also run through a pre-trained classifier : Inception ResNet V2, which is a powerful neural network trained on 1.2 million images. The final classification layer from the classifier is merged with the output from the encoder using a fusion layer. This fusion layer combines the learned representations from the classifier with the colorization network, enabling the network to match object representations with a colouring scheme.

The article mentions the use of Mean Square Error (MSE) as the loss function, it was used for all the three versions of our model. MSE is a common choice for regression problems, including image colorization as it measures the average of the squares of the errors or differences between the predicted colour values and the actual colour values in the training dataset. Additionally, the article mentions the use of classification and weighted classification loss functions, but it specifically uses MSE in the context of the colorization neural network design.

3) The article discusses several challenges and limitations encountered during the development of the colorization neural network.

- Poor generalization of the network in the alpha version, where the network memorizes the information from the training image, colourizes the b/w version of the same image well but struggles to colourize images it hasn't seen before. This limitation is taken care of in the beta version by training the network on a larger and more diverse dataset to improve generalization.

- Unsatisfactory results obtained from training the beta and/or full version of the model on a diverse dataset leading to most images turning out brownish. This problem can be addressed by focusing on training the model on datasets which have very similar samples rather than the diverse one, which will result in more consistent and realistic colorization.

- Difficulty and importance of maintaining proper image sizes and ratios as the networks rely on pixel locations and not the information content. This problem/ challenge can be addressed by doing proper preprocessing as required for the dataset to ensure that the input images are standardized according to the design of the model. Doing so can improve the network's performance and accuracy significantly.