**Mathematical Review Notes:**

Understanding “Colorful image Coloriazation” – paper by Richard Zhang

**The General Gist:**

* This paper is all about trying to fool someone into thinking that the output achieved from their colourisation of greyscale image method is in fact the real image.
* It is NOT trying to recreate the images colour precisely, rather give a realistic interpretation of what the colours could be.
* It is doing this, put simply, by predicting the a and b colour values (in Lab colour space) based on the given Lightness channel (the L of Lab)

**The Objective Function:**

* (found in section 2.1)
* Works with the Lab colour space
  + a 3-axis **color** system with dimension L for lightness and a and b for the **color** dimensions.
* Their objective is to “learn a mapping” (where is a prediction, and is some function carried out on X – the lightness channel, where X is a member of all real numbers and has image dimensions H, W?)
* As the Lab colour space is used, there is perceptual distance – therefore a “natural objective function” is the Euclidean Loss between predicted and ground truth colors:
  + Meaning: Measuring the distance between two points to check how far away the predicted and true values are, and therefore how well performing the model is doing.
  + The output of this is the amount of error.
* However: the loss isn’t robust in the case of this paper
  + Why?
    - There’s ambiguity as the aim Is NOT to recreate the actual colours, but rather suggest a colourisation that is realistic and believable.
    - “if an object can take on a set of distinct ab values, the optimal solution to the Euclidean loss will be the mean of the set”. Meaning the following: As there are, as it were, multiple solutions, this kind of loss function will say, let’s just take the average of these possibilities! BUT this results in the desaturated results as described!
    - Also mentions plausible colourisations being in a non-convex set, meaning there are multiple feasible regions where the colours could be, but finding an optimal solution that is in fact the global optimal solution is in fact incredible difficult to know?
* So instead, this problem is now treated ass multinomial classification:
  + ab output space is quantized (so can take only a specific set of values) into bins with grid size 10 and keep the Q = 313 values which are in-gamut (i.e. they are within the ranges desired)
  + is prediction Z based off some function G on X, which a mapping is learnt to a probability distribution over possible colours
    - This is where things get a little fuzzy for me!!!
    - Possible colours is a member of some vector? Which has height width and quantized ab values?
* Now there is a need to compare the predicted value against ground truth, where the paper defines their own function for this which will convert the ground truth colour into a vector using some soft encoding scheme? (not sure what that is), and then use multinomial cross entropy loss (which I’ve tried looking for online and am struggling) to figure out the error?
* The Weighting and rebalancing maths I am also having difficulty with, although I do understand the need for this step:
  + Looking at the distribution of the ab values in figure 3b does show the strong bias towards low ab values as indicated by the red, orange, and yellow accumulating around 0.
  + I also understand that this is because of the regular occurrences of backgrounds with these values like clouds and pavements etc.
  + I also think I understand that without taking this into account then the loss function (however it is working) will then be dominated by desaturated ab values and skew the model’s predictions?