

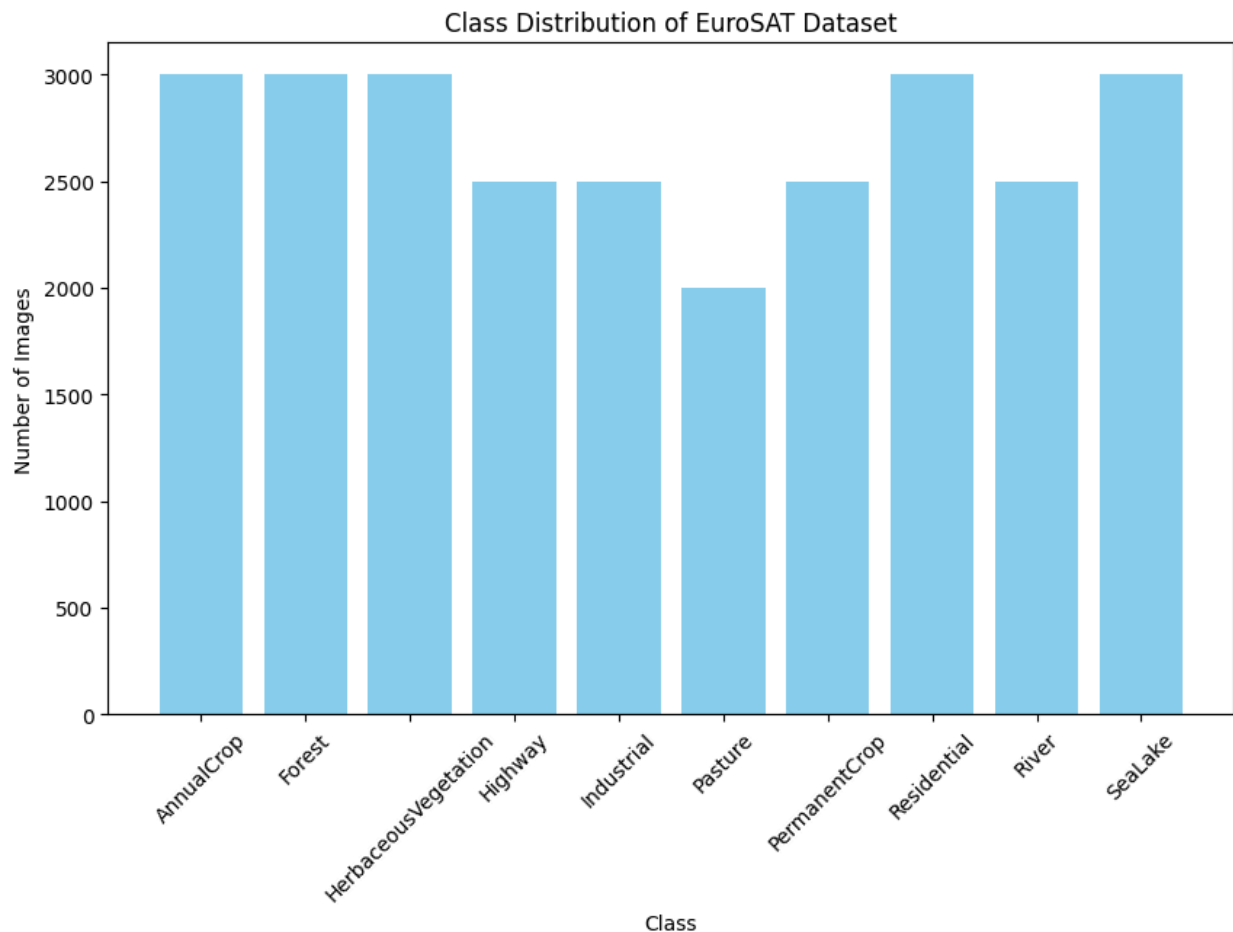
DATASET LINK BEFORE PREPROCESS:

[dataset link](#)

PRE EXPLORATORY ANALYSIS:

Data Distribution in classes:

The dataset has 27000 images which have a spectrum band of 13 and 10 classes. We first analyzed which class had more images. We found out that AnnualCrop, Forest, Herbaceous Vegetation, Residential and SeaLake had the highest number of images with 3000 images each. The Lowest was the Pasture class which had 2000 images only.



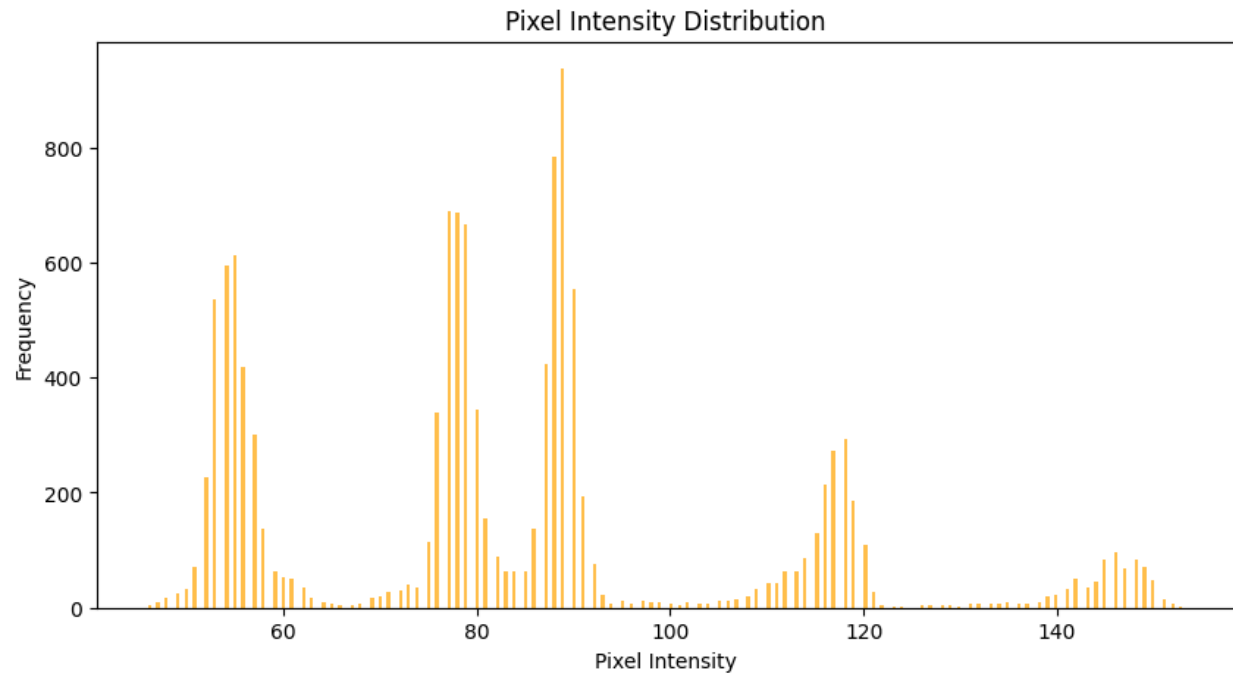
Visualize Sample Images:

After checking the class distribution of EUROSAT. We visualize images from each class to view if all classes and their data are loaded correctly. We generated the image randomly so that if there was any missing value the function would catch it.



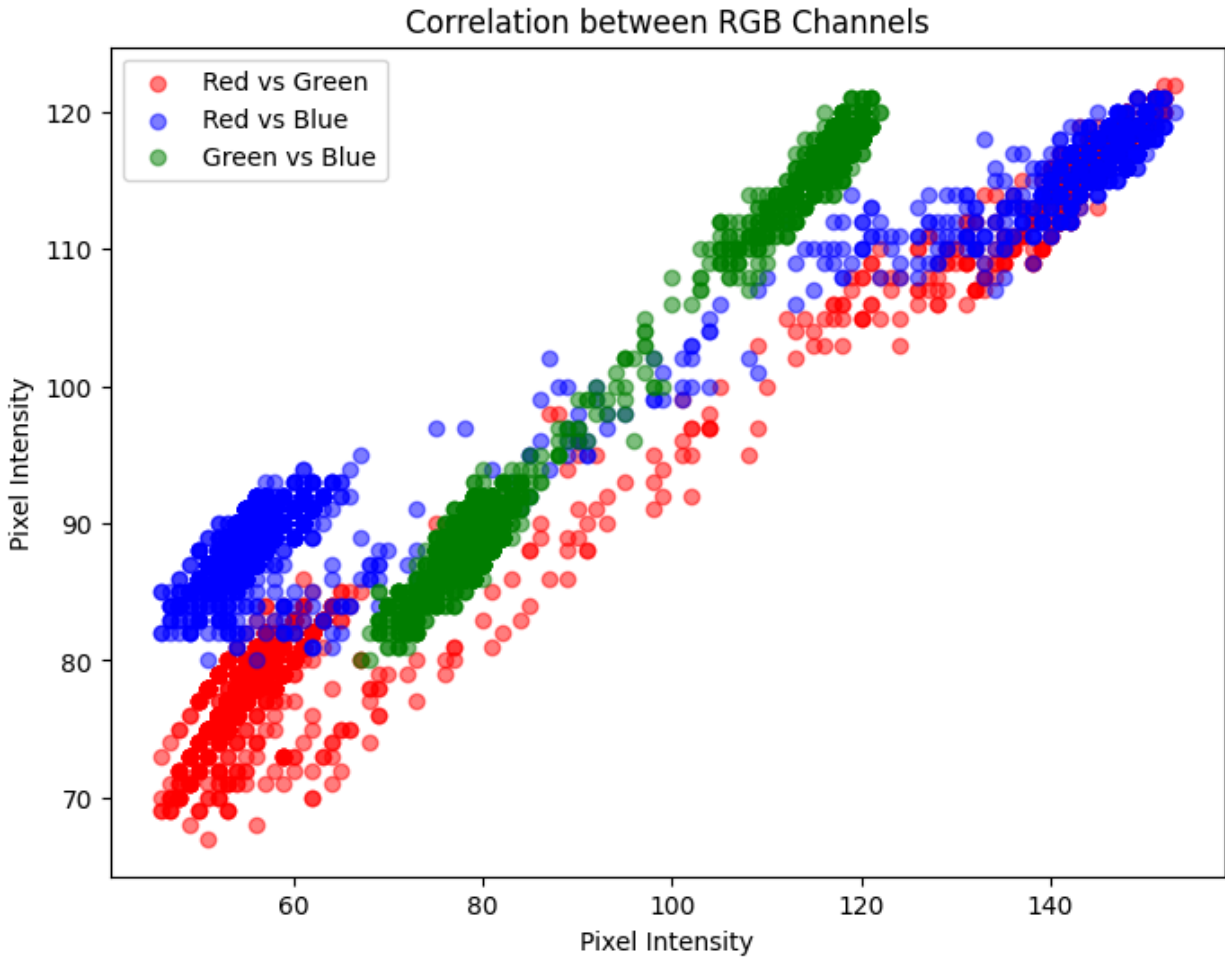
Pixel Intensity:

After viewing the images from classes. We took one sample image to check the pixel intensity. Pixel intensity means the sum, mean or median intensity pixels within a range. The pixel intensity can be 0 to 255 based on the sample image. In our sample image the intensity didn't get too high.



Correlation Between Channels:

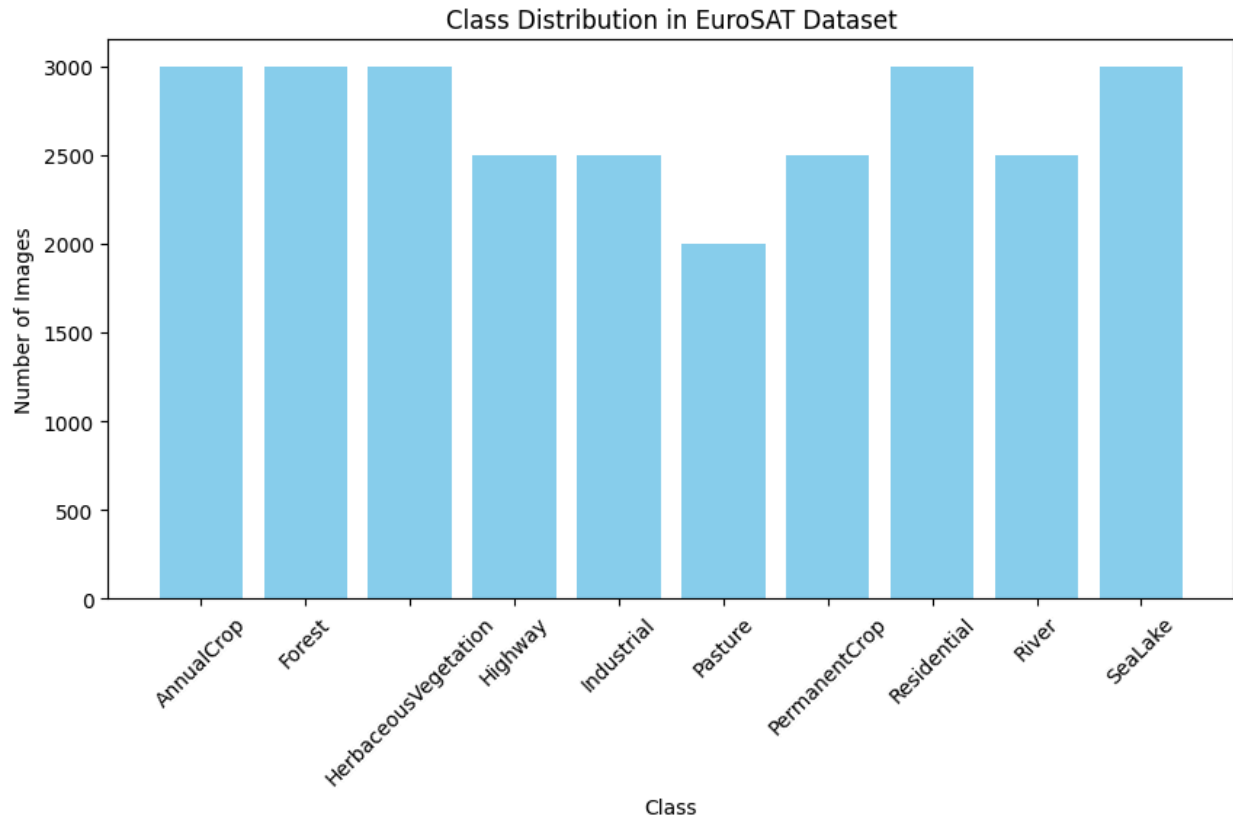
Correlation means a connection between two things which are dependent on each other. We have visualized the correlation of RGB channels in our sample image and we saw that in 100 pixel regions there are more green and blue. In 80-90 there are more red and blue. From 75 to 120 we can see more red and green. Other parts of the images contain the actual red, green and blue.



POST EXPLORATORY ANALYSIS:

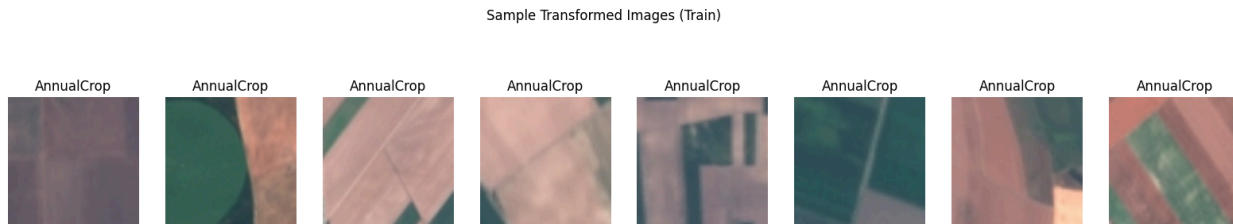
Data Distribution in classes:

We checked after preprocessing the data if there were any missing values but there were none.. All the data in 10 classes were successfully loaded. The histogram shows us that the highest and lowest values of class distribution remain the same.



Normalizing and Transforming images:

We convert the pixel values of the images from 0-255 to 0-1 which makes it easier to work on. We Randomly Resized, flipped horizontally, vertically. We used the mean value as a threshold. We viewed images from a particular class to understand if the transformation and normalization worked.



RGB DISTRIBUTION:

In Pre EDA we check rgb distribution for a particular image. Here we have also visualized the rgb distribution for the same image. From the below histogram we can see that for every pixel value how many times red,green,blue appeared. Now it has changed. Here the pixel intensity is till 255 because we visualized on transformed image only.

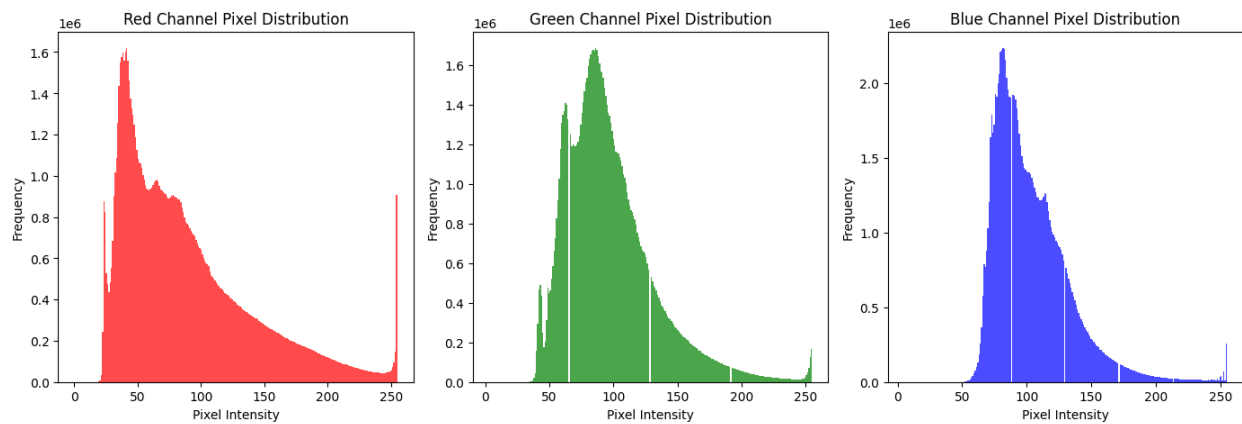
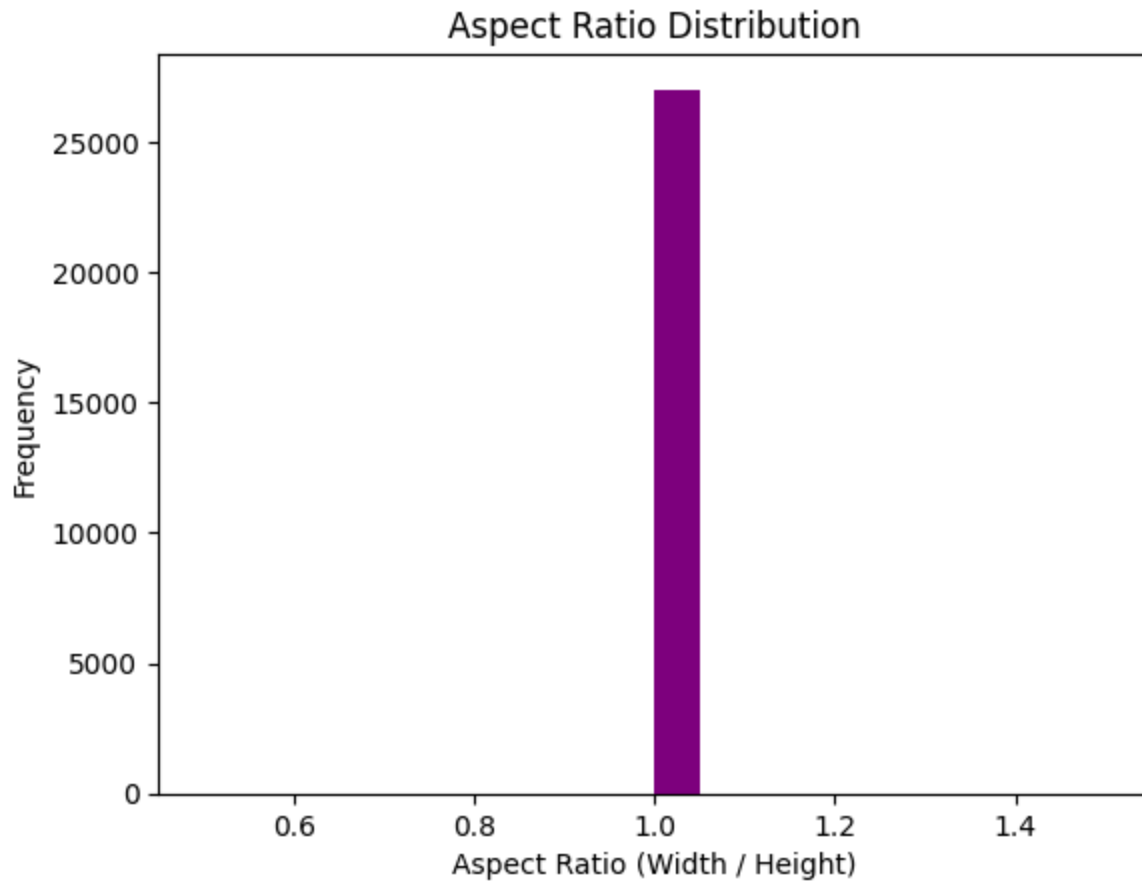


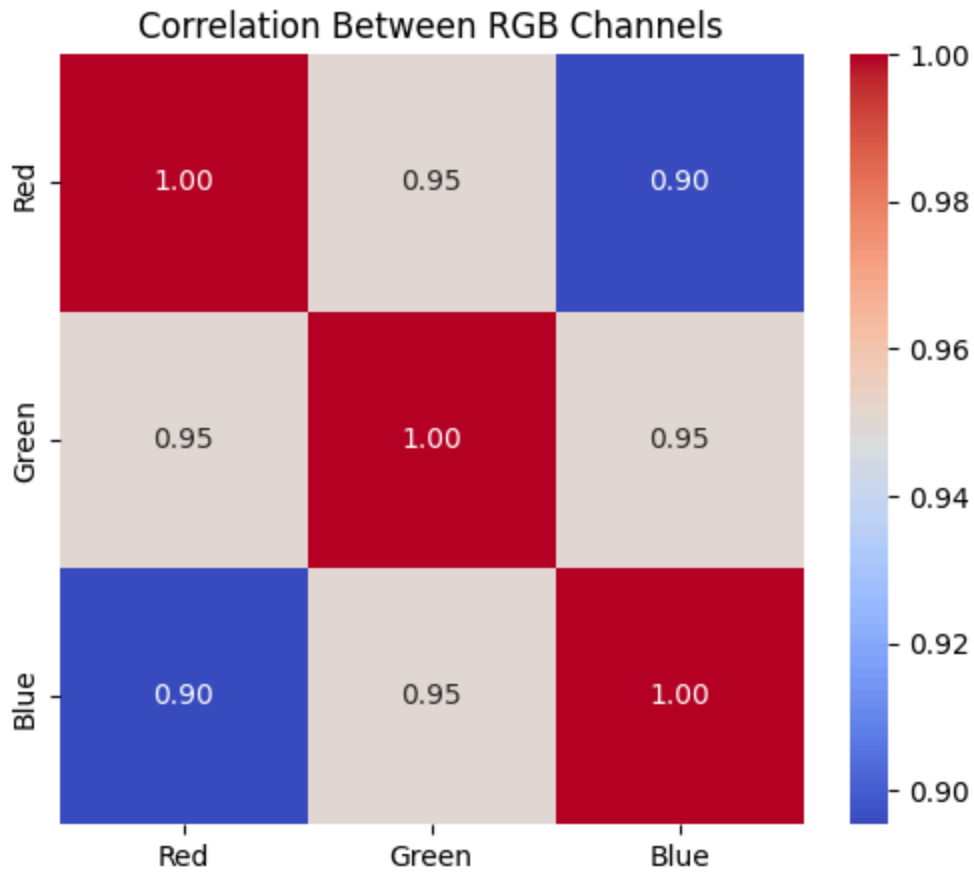
Image Ratio Distribution:

We have to check if the image size is still constant because the pixel size will change because of normalization but the height and width should remain the same. From the below distribution we can see that it has remained constant.



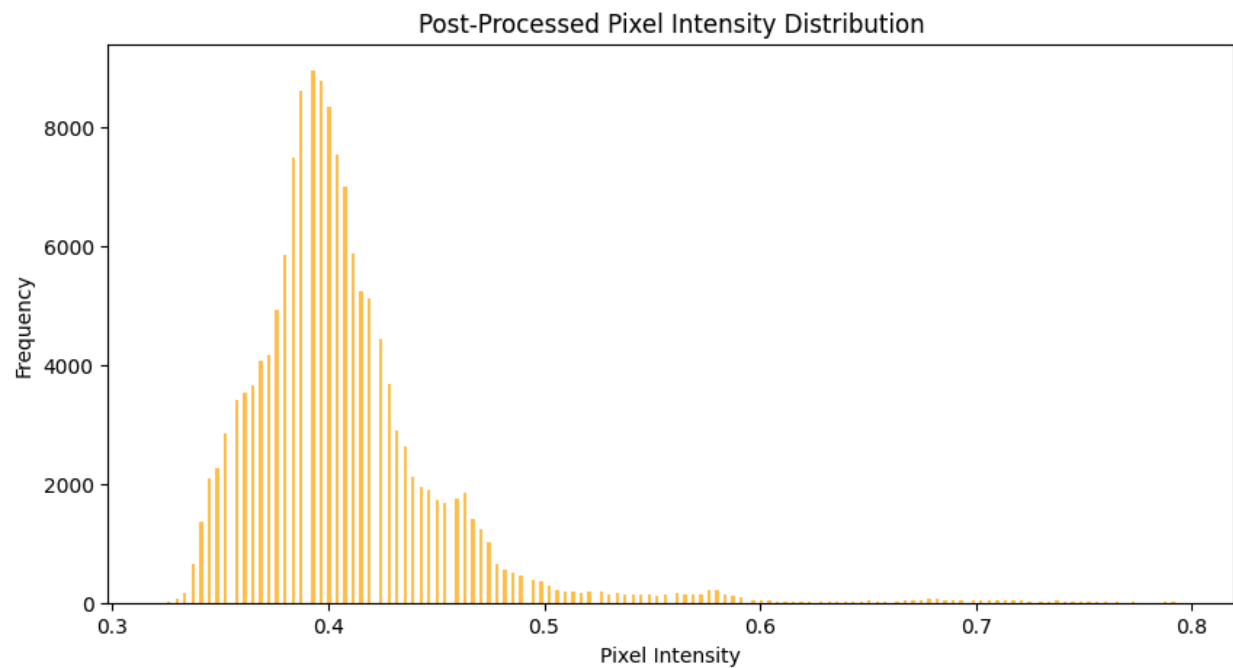
Correlation between Channels:

After Normalizing the images the correlation has improved significantly . The matrix shows us that the three variables are now more dependent on each other.



Post process pixel intensity distribution:

After processing the image the images pixel values are now in 0 to 1 so the pixel intensity has changed. In 0.4 we see that the pixels are more. In the pre process we saw that pixels were scattered but here they are more correlated to each other.



Data Augmentation:

We have applied the transform function which we had built previously. But here we have also used `permute` from `pytorch`. `permute()` rearranges the original tensor according to the desired ordering and returns a new multidimensional rotated tensor. It permutes images randomly to do augmentation.

Sample Augmented Images (Train)

