Step 1: Data  
Before starting to train a model, please get familiar with the dataset. When you look at the  
dataset, please answer the following questions:

1. How many data samples are included in the dataset?

There are 80 training images, paired with their masks. There are 20 test images, paired with their masks.

1. Which problem will this dataset try to address?

The problem is a binary segmentation, predict a mask. Which means each pixel 0 or 1 if it’s a blood vessel or not. For application, this dataset is to help blood vessels be extracted to an easier to see image for diagnosis.

1. What is the dimension ranging in the dataset?

All images are 512 by 512 resolution with three 0-255 RGB values for each pixel.

1. Does the dataset have any missing information? E.g., missing features.

Nope, all images are solid, none are corrupted or empty. It might have a little noise, but it’s clean for the most part.

1. What is the label of this dataset?

A binary mask, a 512 by 512 binary pixel image.

1. How many percent of data will you use for training, validation and testing?

I will use all 80 images available for training, and then 10 of the test images for validation and the rest for testing. Thus it will be a 80/10/10 split.

1. What kind of data pre-processing will you use for your training dataset?

I will need to normalize the mask images to binary, because they start in grayscale.

I will need to normalize the images as well using global mean and std.

I will not be rotating, scaling, or cropping any images as it does not seem like the given masks are doing any of that.