Between the two models, both are unable to properly differentiate between the boat types. The second model has a much greater ability to classify the boats it was trained on and has a slightly better ability for classifying the test data. I would not use either model in its current trained state for the application.

Over the training epochs, the first model had a smaller amount of change between the initial validation loss and training loss. Same goes for the accuracy which started at a minimum of 0.32 and increased to a max of 0.41. On the other hand, the second model had a better performance in decreasing the loss and increasing accuracy. The test data for both however showed that both models were equally unable to differentiate untested data. The first model showed a loss of 1.68 and an accuracy of 0.37. The second model showed a loss of 1.63 and an accuracy of 0.44. Both these values means that on untested data, the models cannot classify between the images.

The heatmaps based on the two models which predictions are true and which are not. Both models only predict classifications for the gondola, kayak, and sailboat. I believe that with a larger dataset, the second model would have surpassed the first model in terms of classification. There is either an issue of overfitting to the training data, which showed a greater ability to classify the images on the transfer learning model, or an issue with the complexity of the neural network, one with less layers may be enough to classify the images. In the future, I would like to add object detection, then classification to allow the model to isolate and focus solely on areas where the boat is located.