



CECOR

Automation of Data Labeling

Mathew Trotter, Corithian Williams
North Carolina A&T State University

William Marchetto, 06773



Problem Statement

Data labeling is a very time-consuming and tedious task, but is a necessary aspect of developing, training and testing models used in various machine learning tasks. In fact, many more hours and resources are spent in ensuring quality and quantity of the data that is used for machine learning tasks.

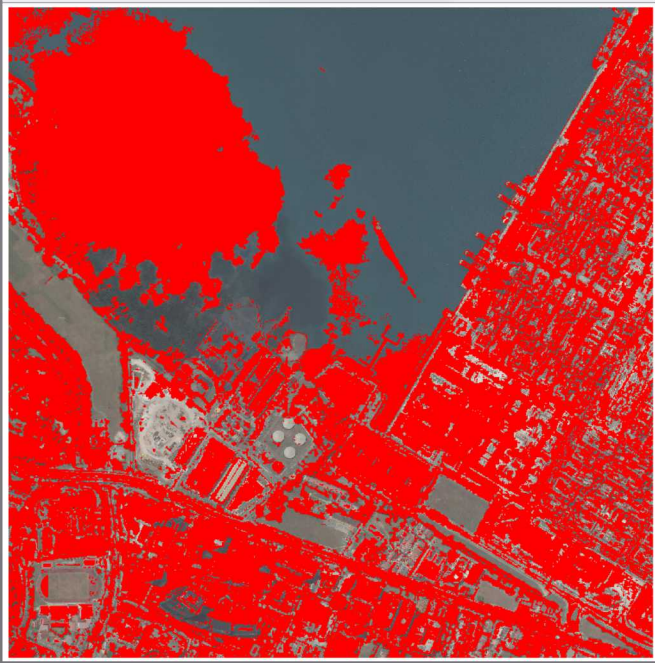
For this project, we were tasked with labeling a large amount of data to be utilized in testing various machine learning algorithms. We quickly realized that the quantity of data would needed to label was infeasible to accomplish manually. We then developed a method to accomplish this task based around the fact that all of our targets were various sized boats in the water. Using the static background of water in our images, we are able to develop a system that can isolate the targets and create a polygon around all desired targets in the image.

Technical Challenges

- The second mask may filter out parts of our targets, affecting the accuracy of the polygons we generate around them
- Automating classifications of ships has been difficult as slight differences in targets can change the classification
- Properly filtering our land and targets next to the land was tricky. We were able to apply a convex polygon solution to properly differentiate land and a target next to land.



Once we have the photo only displaying water, we found that all that remains in the image were our targets and piers extending into the water. We apply a second filtering mask as before to exclude the water and the piers. Finally our image should only display our targets. From this final image, we simply need to find the areas not covered by our two masks, then we can place a polygon around each target.



Our overall objective for this project is to develop a method to input an image containing desired label targets, being ships of various sizes, and obtain a polygon around all the targets along with a label for the type of ship. Our main assumption is that all of the targets reside on the water. Working from this assumption, we apply a mask to isolate the water and everything in the water.

Objectives & Approach

