Handwritten Equation Solver Using CNN

Chase Clark

In this project, I created a CNN that can recognize math symbols in images. My project breaks down into three main phases or sections. Data preprocessing, model creation, and testing of the model. In this document I will explain how I approached each section and what challenge I faced along the way.

I found my dataset on a popular data science site called Kaggle. There, I found a dataset that includes handwritten numbers and symbols in image format. After extracting each symbol into their own subfolder I was ready to start processing them into format that is easier to work. Next, I created a method that loops through every image in the subfolder and extracts the important features. Before working with the images, I inverted the colors so it is a black background with a white foreground. This supposedly increases the accuracy of the opencv library I was using. Then for each image, I used opencv to identify contours and create bounding rectangles around them. Next, I used the dimensions of the bounding rectangle to crop the image down to a normalized size of 28 x 28. Finally, I flatten the image and store the data inside a csv file and append the label in the final column. Now the data is ready to be used for training.

Creating the model was the hardest part of the project for me. I had to research and find examples on the web before I was able to put together an effective model.( Reference: Vipul Gupta <https://medium.com/@vipul.gupta73921/handwritten-equation-solver-using-convolutional-neural-network-a44acc0bd9f8> ) Using keras, I added the first layer to the model which is Conv2D. This layer