Parker Williamson

2/5/2018

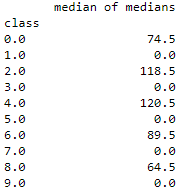
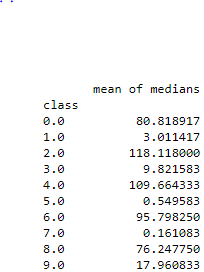
Springboard Data Science Career Track

Capstone Project 1 – Data Wrangling

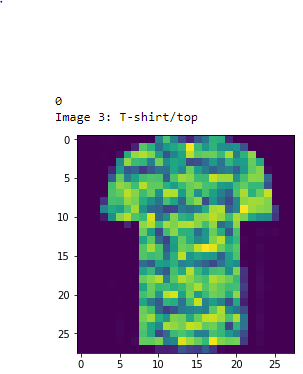
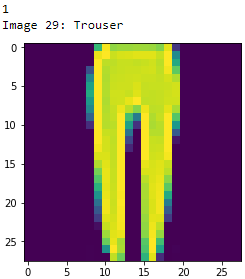
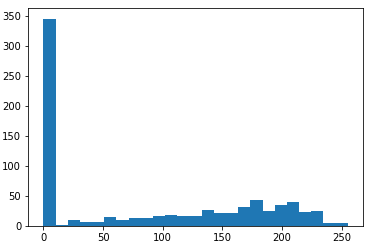
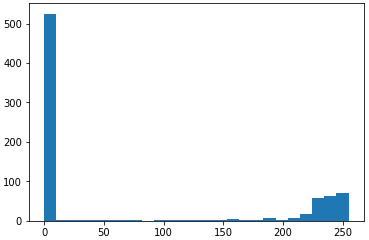
Image data needs to be broken down by image and category for meaningful analysis. Histograms can show the pixel range used and intermediate images of the CNN models can show important features. I will also look at the mean and median values of each image and how preprocessing changes the image of the biggest difference. Show an example of each category and figure out which categories are the most commonly confused.

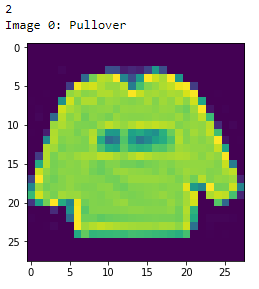
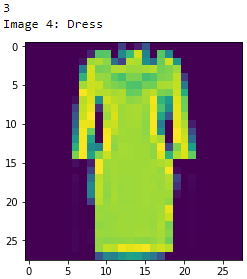
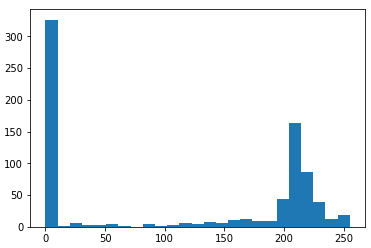
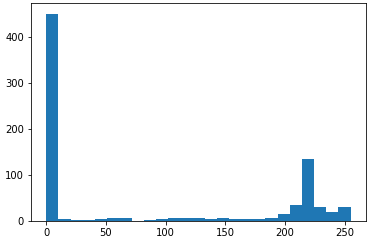
Mean and median values start to show some of the differences between categories. Mode wouldn’t make sense to examine because it would be zero for every class since the zero is the most common value, because of the background. Although groups definitely mix, it is a decent start to look at their mean and median because categories like 5:Sandels will clearly have a much lower average than others such as 2:Pullovers. Shown below are the median by class in order and two different analysis of medians.

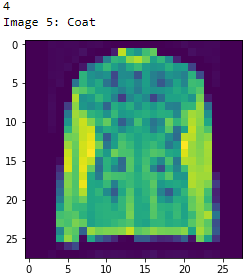
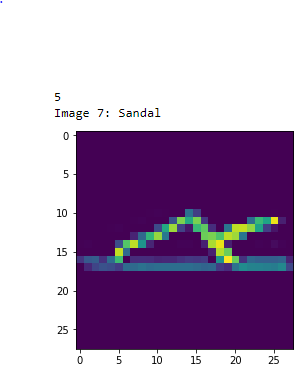
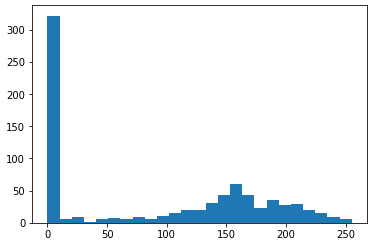
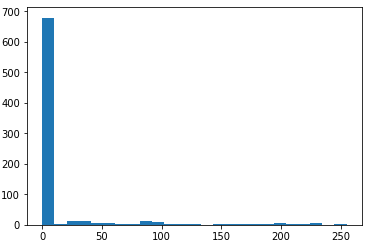


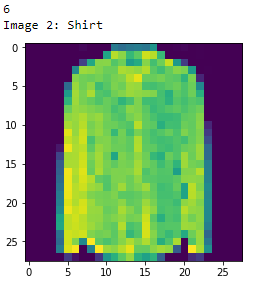
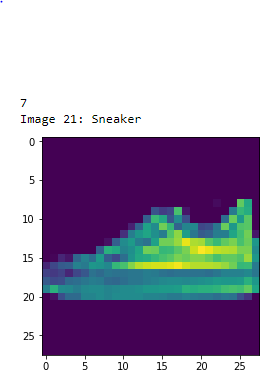
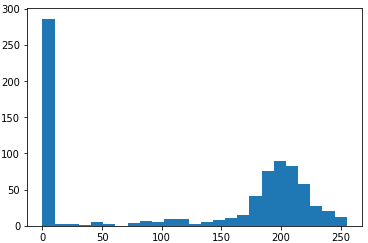
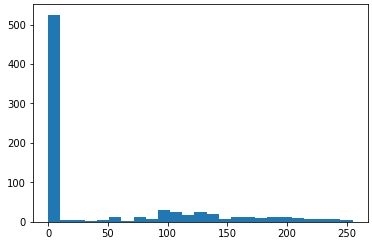
 

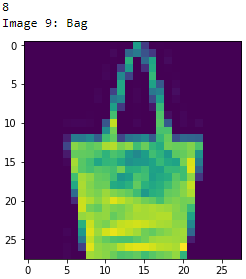
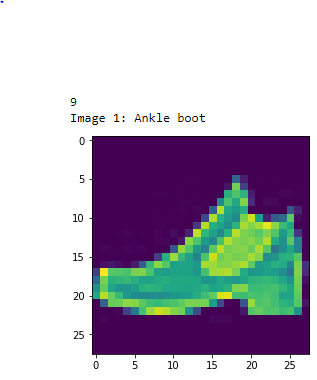
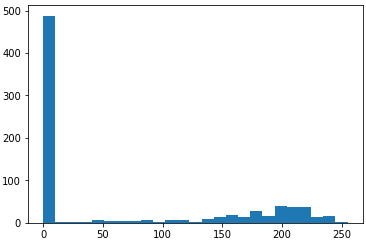
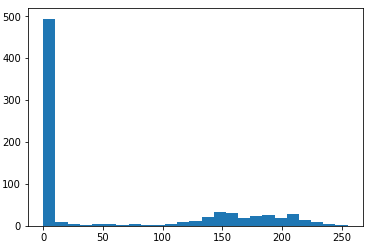
An example of all 10 classes of images are on the following pages, clearly shirt, t-shirt, pullover and coat categories are a lot more similar than some of the other classes, and their histograms are more similar to match. The histograms all largely have a spike at zero because of the background but then deviate, the shoes are all flat across the range but the rest build to the high value end. The tops seem to have more variation than others, because they have decals or color variation.

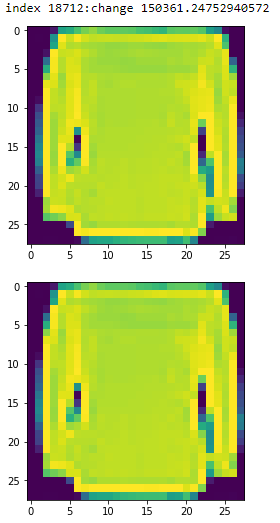
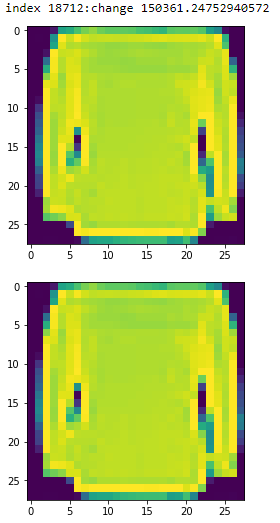
  

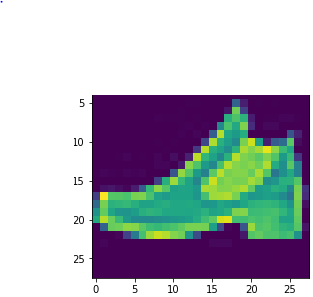
  

The effects of normalization on one of the images with the larges changes are below. The variation per pixel is very small so clearly the affect it minor. It seems to make the armpit areas a little bit brighter.

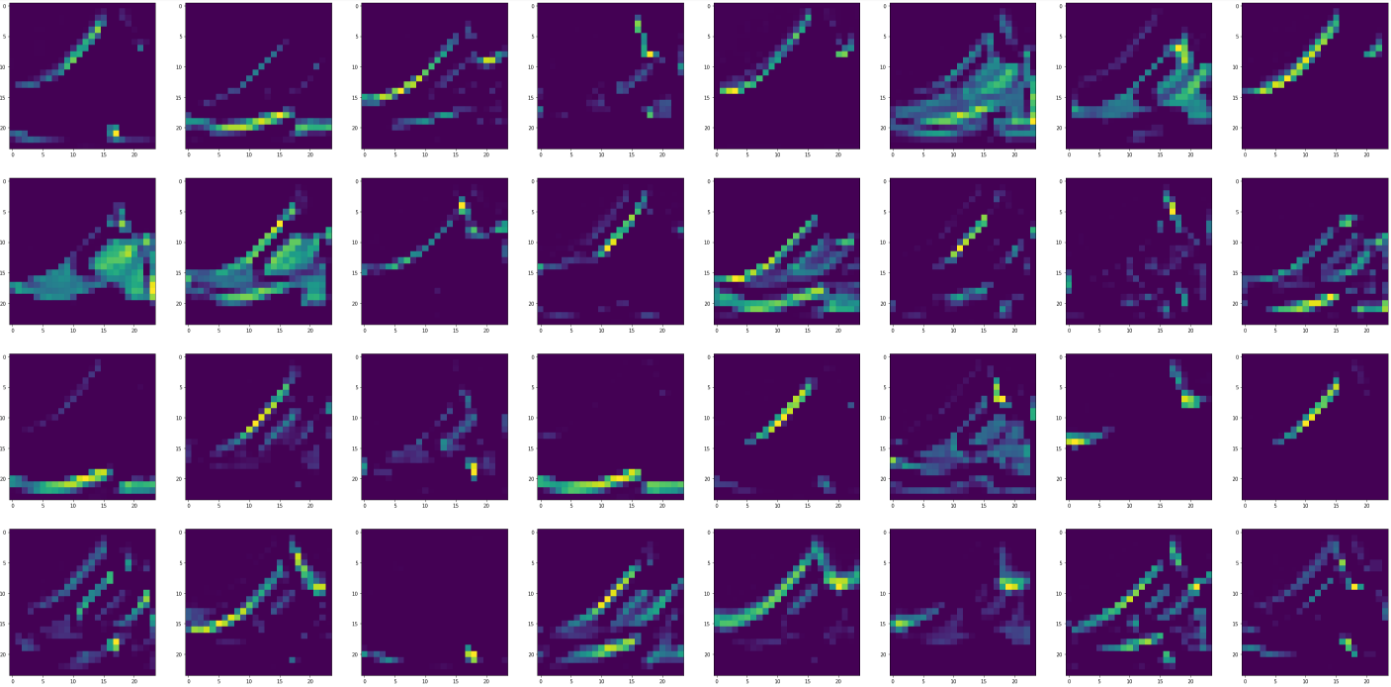


Finally there are visualizations of the convolutional neural network (CNN) intermediate layers. The first layer breaks the image into many different pieces and the second layer abstracts them to focus more on the general form. This is of the first CNN architecture and layout. Results may be improved by changing it, but these results are representative of the filtration and abstraction that the neural network does on every image.

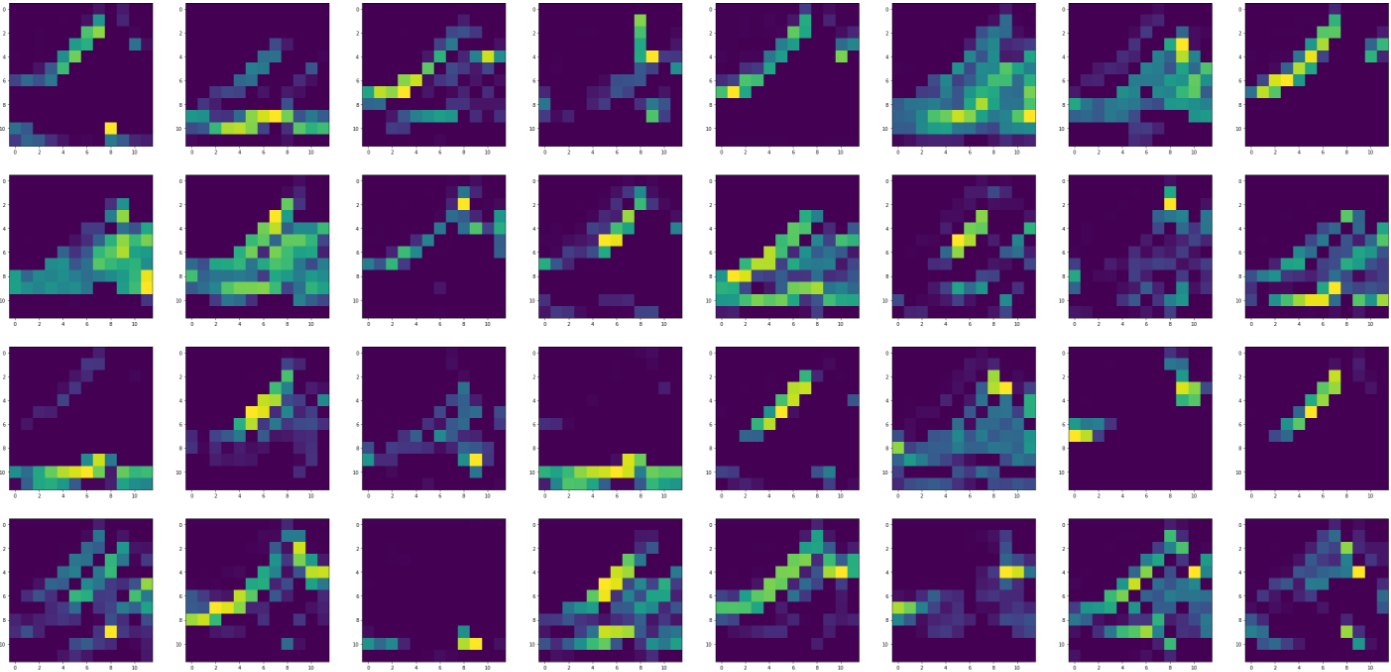
Original image

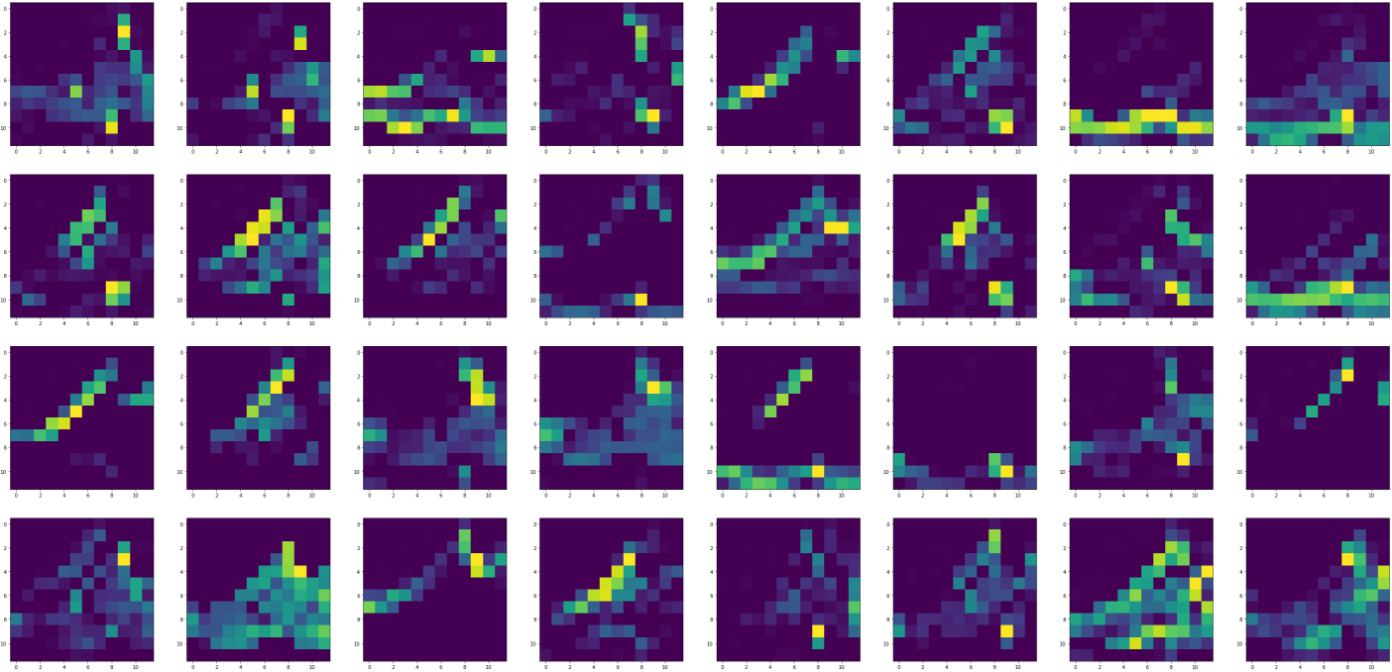


PHASE 1



PHASE 2





Further tweaking of the neural network is needed to improve results, but there are many ways that images can be separated and identified through the visual properties show in this document.