Anthony J Vasques

Date: 10312023

Deep Learning Final Project Proposal

What:

Vision transformer for Object Detection

Why:

Vision Transformer (VT) is the current cutting edge in computer vision, and it has been used as a backbone for classification and object detection algorithms because of its computational complexity of over convolution neural networks (CNN).

Project:

Because of the time constraints of this project, I will need to procure a dataset that has already been curated, since non-curated image data can take a long time to process. For that reason, I will use Hand-Written Numbers MNIST dataset, which has 70k grey scaled images of dimensions 28x28 pixels over 10 classes of numbers (1-10). This dataset is well known and often used, so data curation is not a concern with this dataset. Noisy grayscale images will be created of dimension 128x128. These noisy images will be overlaid by the MNIST number at a random location. The random location will be documented and stored in a commonly used format. The data will undergo smoothing, and will split into train, test, and validation sets. An object detector with a vision transformer will be trained and tested. For comparison, an off-the-shelf CNN object detector will also be used to compare results.

References:

https://www.kaggle.com/datasets/oddrationale/mnist-in-csv