CS 7643
Final Project Proposal

Team Name: Dimensional Dominators

Project Title: 3D image classification

Project summary (4-5+ sentences). *Fill in your problem and background/motivation (why do you want to solve it? Why is it interesting?). This should provide some detail (don't just say "I'll be working on object detection")*

3D data is important in building world models. It can unlock and bridge the interface between machine learning and the physical world. Unlike traditional 2D image classification, 3D object classification presents unique challenges that translate to physical perception problems. The goal for doing the project is to explore methodologies pertaining to 3D point cloud and how it can be used to train deep learning networks. These techniques can eventually be applied to more complex problems for spatial systems and 3D reasoning.

What you will do *(Approach, 4-5+ sentences) - Be specific about what you will implement and what existing code you will use. Describe what you actually plan to implement or the experiments you might try, etc. Again, provide sufficient information describing exactly what you'll do. One of the key things to note is that just downloading code and running it on a dataset is not sufficient for a description or a project! Some thorough implementation, analysis, theory, etc. have to be done for the project.*

For this project we plan on manually reimplementing the PointPillars and PointNet models. We will train our models on the ModelNet dataset. We will be using the ModelNet dataset because it is readily obtained and also used in multiple benchmarking, making our results comparable to other models. We plan to tune the hyperparameters of our model to find the best hyperparameter configuration and explain theoretically why these values are the way they are. We also plan to create validation curves for our hyperparameters such as number of epochs, batch size, etc. To determine how well our models perform, we will look at the accuracy and loss curves of our model and compare it to the state of the art model performance (SPNet).

Resources / Related Work & Papers (4-5+ sentences). *What is the state of art for this problem? Note that it is perfectly fine for this project to implement approaches that already exist. This part should show you've done some research about what approaches exist.*

Point clouds are usually transformed to 3D voxels or collection of images, which causes issues due to data volume. PointNet uses a novel type of neural network that directly learns from point clouds in a unified architecture for object classification, segmentation, etc. PointPillars builds on PointNet with a lean downstream network for increased speed and accuracy.

The state-of-the-art model on this dataset is SPNet. It uses stereographic projection to transform 3D inputs into 2D images, which are then learned by a shallow 2D CNN for estimation of the object category and feeding into further classification layers.

SPNet: Deep 3D Object Classification and Retrieval using Stereographic Projection, ACCV2018.
https://arxiv.org/abs/1811.01571
PointPillars: Fast Encoders for Object Detection from Point Clouds
https://arxiv.org/abs/1812.05784
PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation
https://arxiv.org/abs/1612.00593
PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space
https://arxiv.org/abs/1706.02413
PointCNN: Convolution On X-Transformed Points
https://arxiv.org/abs/1801.07791
Dynamic Graph CNN for Learning on Point Clouds
https://arxiv.org/abs/1801.07829

Datasets: https://modelnet.cs.princeton.edu/#
We will initially use ModelNet10 to achieve best performance. We will then attempt ModelNet40 and see how well our models generalize to more output classes.

List your Group members
Mitchell Gray (mgray88)
Timo (htsui6)
Shaheer Amjad (samjad7)
Yonghao Li (yli3850)