Transformers for Point-cloud Data
Week 5 Progress

Project lead: Aakash Kumar
By: Jordaine Wisdom
Main Tasks

- Solving errors
- Training and testing 3DETR model
- Gaining an intuitive understanding of the code
3DETR

**Figure 2: Approach.** *(Left)* 3DETR is an end-to-end trainable Transformer that takes a set of 3D points (point cloud) as input and outputs a set of 3D bounding boxes. The Transformer encoder produces a set of per-point features using multiple layers of self-attention. The point features and a set of ‘query’ embeddings are input to the Transformer decoder that produces a set of boxes. We match the predicted boxes to the ground truth and optimize a set loss. Our model does not use color information (used for visualization only). *(Right)* We randomly sample a set of ‘query’ points that are embedded and then converted into bounding box predictions by the decoder.
ScanNet Data set

Input Point Cloud

Decoder Attention

Detections
Model outputs on ScanNet
Understanding the code base
Next

- Complete going through the code base
- Train and test on KITTI data set