Week 3 Progress
Jordaine Wisdom
Overview of Progress

- Pointnet/pointnet++ → Pre-processing point cloud data to be fed into neural networks (extracting global and local features from PC’s)
- Transformers → attention head used for matching query and key to get attention matrix to generate loss based on how close predicted is to the ground truth
- 2DETR → extension of transformer but uses CNN to extract features from images to be fed into transformer and then perform the attention head
- 3DETR → extension of 2DETR with the main difference using methods from pointnet++ (sampling based on local PC density) to extract features and then feed into transformer network to generate predictions and perform matching
Pointnet/Pointnet++

- Handling point cloud data

Original 3d model

2048 sampled points
Point net classification

pred: sofa, label: sofa
pred: chair, label: chair
pred: toilet, label: toilet
pred: table, label: night_stand

pred: table, label: desk
pred: monitor, label: monitor
pred: chair, label: chair
pred: table, label: table
Transformers

Multi head attention $\rightarrow$ weighted attention matrix
DETR (Transformers for images)

- CNN extracts features
- Features positional encoded and flattened to be fed into transformer
- Predictions generated for labels and bounding box
- Predictions are compared with ground truth using 1 to 1 matching to generate loss
Main difference is the input method
- Down samples the point cloud using sampling techniques from Point net++ to capture both global and local features
- Transformer encoder decoder structure is the same as the 2DETR
- Similar 1 to 1 matching used to calculate loss from predictions and ground truth
Next steps

• Reading Track former → tracking objects across 2d frames
• Start working to extend 3DETR by implementing latest concepts to improve object detection