Transformers for Point-cloud Data

Week 8

Prject lead: Aakash Kumar
By: Jordaine Wisdom
Overview:

- 3DETR
- Loss convergence on KITTI dataset
- Multimodal 3D object detection
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 Dataset (1.3 TB)

- RGB-D video dataset
- more than 1500 scans
KITTI dataset
Updated training on KITTI

**Total Loss**

- **Last week:**

- **This week:**
Total loss

- Summation of all the individual losses

```python
self.loss_functions = {
    "loss_sem_cls": self.loss_sem_cls,
    "loss_angle": self.loss_angle,
    "loss_center": self.loss_center,
    "loss_size": self.loss_size,
    "loss_giou": self.loss_giou,
    # this isn't used during training and is logged for debugging.
    # thus, this loss does not have a loss_weight associated with it.
    "loss_cardinality": self.loss_cardinality,
}
```
loss_center

tag: Train_details/loss_center
Next Steps

- Fix camera to Velodyne transformation
- Retrain and get better results