**SS23 3D ML Project Proposal**

Team Members:

1. Biray Sutcuoglu – XXXXX
2. Youssef Youssef - 03766163

**Project 1:** Text2Scene Synthesis

**Summary**: Creates a 3D scene from a given text input query using a diffusion model.

**Paper**:

* DiffuScene: Scene Graph Denoising Diffusion Probabilistic Model for Generative Indoor Scene Synthesis
* Link: <https://arxiv.org/abs/2303.14207>

**Dataset**: 3D-FRONT dining and living rooms

**Modifications**:

* Change the diffusion model to an autoregressive **random** transformer with attention. The random transformer is to preserve the main idea behind the paper of order-invariance of the furniture in the scene.

**Notes**:

* Not very sure the transformer architecture will yield better results

**Project 2:** Reconstruction from single 2D Images

**Summary:** Creates a 3D shape from a single 2D image

**Paper**:

* 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction
* Link: <https://arxiv.org/abs/1604.00449>

**Dataset**: Shapenet

**Modifications:**

* Change encoder to use a transformer with attention
* Change decoder to use a transformer with attention
* If there is enough time, try to reverse the process.
  + Given a 3D shape, output its corresponding 2D Image

**Notes**:  
 Looks straight forward

**Project 3**: Dense Captioning from 3D scenes

**Summary**: Given a 3D input scene, perform object segmentation and caption each identified object

**Paper**:

* Scan2Cap: Context-aware Dense Captioning in RGB-D Scans
* Link: <https://openaccess.thecvf.com/content/CVPR2021/html/Chen_Scan2Cap_Context-Aware_Dense_Captioning_in_RGB-D_Scans_CVPR_2021_paper.html?ref=https://githubhelp.com>

**Dataset:** ScanRefer (Requires author permission)

**Modifications:**

* Change Votenet module to use softgroup instead

**Notes:**

* Very hard and complicated

**Project 4**: 3D shape reconstruction & completion from deficient 3D representations

**Summary**: Reconstruct high quality shapes from deficient 3D representations such as incomplete point clouds or low-resolution voxel grids. We will probably focus on one of the deficient representations

**Paper**:

* Implicit Functions in Feature Space for 3D
* Link: https://openaccess.thecvf.com/content\_CVPR\_2020/html/Chibane\_Implicit\_Functions\_in\_Feature\_Space\_for\_3D\_Shape\_Reconstruction\_and\_CVPR\_2020\_paper.html

**Dataset:** 3D human scans (requires purchase) – ShapeNet – Text2Shape Dataset

**Modifications:**

* Apply text guided reconstruction
* Train on shape net
* Finetune on Text2Shape to learn text guided reconstruction

**Notes:**

* Super nice paper but not sure about the modifications or ideas

**Project 5**: Instance Segmentation

**Summary**: Given a 3D scene in the form of a point cloud, perform instance segmentation

**Paper**:

* PointGroup: Dual-Set Point Grouping for 3D Instance Segmentation
* Link: <https://arxiv.org/abs/2004.01658>

**Dataset:** Scannet

**Modifications:**

* Not quite sure here – maybe changing the backbone network to something else?