Project Proposal: 3D Vertebra Generation with Diffusion Models

Ege Ozgul
ozgul.e@northeastern.edu
Northeastern University

Andrew Sayegh
sayegh.a@northeastern.edu
Northeastern University

Abstract

This project develops a 3D diffusion model for generating anatomically accurate thoracic vertebrae (T1–T12). Using high-quality STL scans from the MedShapeNet dataset, we will train a generative model that produces novel vertebral structures while preserving key anatomical features. Our approach converts mesh data to voxel or point cloud representations, applies diffusion-based generation, and evaluates outputs using FID and Precision/Recall metrics. Primary challenges include limited training samples, computational constraints, and ensuring 3D spatial invariance.

1 Introduction

The thoracic spine stands out as a complex and essential region of the human skeleton. It is composed of twelve distinct vertebrae labeled T-1 through T-12. These vertebrae connect the ribcage to the spinal column, protect the spinal cord, and support upper posture. Modeling these bones is crucial for medical visualization, education and surgical planning. Our project focuses on generating high quality 3D representations of the thoracic vertebrae using diffusion. Diffusion models have proved effective within recent advances of generative artificial intelligence, and we strive to utilize this technology to diffuse together voxels which represent the stated bones. We then would have a pipeline which would input the models output into a convex hull or marching cube algorithm to generate the meshed structure.

From a personal standpoint, both of us are new to 3D medical data and diffusion-based generation, and thus we will be learning the concepts as we work.

2 Objectives

2.1 Main Goal

Build a 3D diffusion model that generates unique human spine T1-T12 sections.

2.2 Stretch Goals

- Evaluate generation quality using both visual inspection and quantitative metrics
- Explore bone repair and completion from partial scans
- Train the model to generate other parts of the human skeleton

3 Technical Details

3.1 Training Dataset

We plan to use the MedShapeNet dataset, which contains tens of thousands of modeled bones. However, we will focus on training the diffusion model on thoracic vertebrae (T1–T12) datasets. There are approximately 700 3D object files per thoracic vertebra.

MedShapeNet Dataset Resources:

- **Main landing page:** `https://medshapenet-ikim.streamlit.app/`
- **vertebrae_T12 Dataset:** 766 STL files available (search "vertebrae_T12" at the main page)

3.2 Input Data Format

The training dataset consists of STL mesh files, which define single objects using vertices and triangles. Each STL object will be converted to either a 3D point cloud or 3D voxel representation for model training.

3.3 Output Data Format

The model will generate either voxel or 3D point cloud representations, which can be converted to proper mesh objects using Convex Hull or Marching Cubes algorithms.

3.4 Diffusion Model Architecture

The architecture is currently under investigation as we progress through the project. We are considering the approach used in this paper as a potential starting point.

4 Project Schedule

4.1 Milestone 1 (October 23): Input Data Preprocessing

- Convert STL files to voxel or point cloud format
- Store data in a compact file format

4.2 Milestone 2 (November 6): Draft Diffusion Model

- Train a preliminary model using a small subset of input models
- Implement 3D output visualization
- Validate model architecture using low voxel resolution for rapid iteration

4.3 Milestone 3 (November 17): Solidified Diffusion Model

- Improve and optimize the model architecture
- Train the model with all T1–T12 sections
- Implement text-conditioned generation (input: T1–T12 label, output: corresponding vertebra)
- Increase training resolution

4.4 Milestone 4 (November 28): Advanced Optimizations

Depending on progress, pursue one of the following directions:

- Further refinement of T1–T12 generation.
- Expansion to complete spine sections (T1–T12 represent approximately 1/4 of the spine)
- Generate the other parts of the human skeleton.

4.5 Milestone 5 (December 5): Output Processing and Documentation

- Generate STL files from 3D point clouds/voxels
- Complete project paper

5 Main Challenges

- Limited training data ($<$1,000 models per bone type)
- Constrained project timeline
- Limited computational resources (GPU availability)
- Achieving translation and orientation invariance

5.1 Rationale for Vertebrae Selection

We selected human vertebrae as our target for several strategic reasons. First, the MedShapeNet dataset provides freely available, high-quality 3D models in sufficient quantity for training generative models. We considered alternative datasets such as automotive models from GrabCAD, but these exhibit excessive geometric variation, making it difficult to obtain consistent, high-quality training data at scale.

The thoracic spine (T1–T12) is particularly well-suited for this project due to the availability of high-quality scans and anatomical consistency within vertebral structures. Our primary objective is to successfully train a diffusion model on these vertebrae and generate anatomically plausible results. As a stretch goal, if we achieve strong performance with the spine, we may expand the approach to other skeletal components where sufficient training data is available.