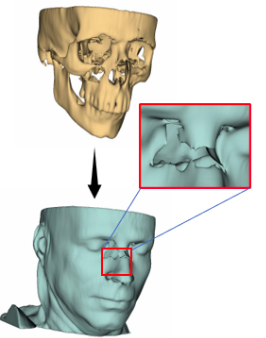
**3D Anatomy Completion (P-10)**

**Mentor:** Munish Daroch (D22001)

**Team Members:**

* Arnav Thakur (B22003 DSE)
* Shreyansh Gupta (B22131 CSE)
* Krish Mittal (B22214 CSE)
* Siddharth Amlavad (B22177 CSE)
* Yash Chaudhary (B22146 CSE)



(FIGURE-1.1)

Broken skull reconstruction with generative deep learning

model

**About the problem statement :**

In recent years, remarkable advancements have been achieved in computer vision, particularly in the reconstruction of two-dimensional (2D) images depicting various scenes, objects, and environments. However, transitioning from 2D image reconstruction to three-dimensional (3D) image reconstruction poses a significant challenge.

Our focus is on reconstructing missing anatomical structures within a human body based on a given 3D anatomy. This concept finds its inspiration in various biological conditions or traumatic incidents that can result in underdeveloped or absent organs in the human body. Whether due to mutations during birth, inadequate nutrition and diet, accidents, or traumatic injuries, all these factors may contribute to an underdeveloped human body.

The objective of our project is to develop a deep learning model capable of taking 3D scanned images of cross-sections of the human body in medical image formats as input and producing a reconstructed version of the image as output.

The inherent challenge lies in the fact that research in this specific field is still ongoing, with only a few selective deep learning models and papers focusing on 3D images. However, the ultimate aim is to create a model that can accurately reconstruct human anatomy and potentially facilitate biological 3D printing.

**Biological terminology and segmentation**

**Image format**

* **MRI (Magnetic Resonance Images):**

Magnetic Resonance Imaging (MRI) is a medical imaging technique used to visualize detailed internal structures of the body. It utilizes a powerful magnetic field, radio waves, and computer technology to generate highly detailed images. MRI is particularly valuable for imaging soft tissues such as the brain, spinal cord, muscles, joints, and organs like the heart, liver, and kidneys.

* **CT SCAN (COMPUTED TOMOGRAPHY):**

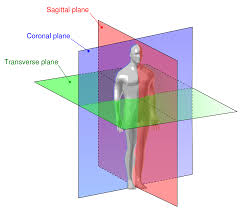
Computed Tomography (CT) scan, also known as CAT (Computerized Axial Tomography), is a medical imaging technique used to create detailed cross-sectional images of the body. CT scans utilize X-rays and computer processing to generate high-resolution images of internal structures.

****

(FIGURE-2.2) MRI VS CT SCAN Machine

**Anatomical Planes and Views:**

* **Sagittal Plane (Lateral View)**: This plane divides the body into right and left halves. When you view the body from the side, you're looking at it as if it were divided along the sagittal plane. A specific sagittal plane, called the mid-sagittal or median plane, divides the body into equal right and left halves.
* **Coronal Plane (Frontal View):** The coronal plane divides the body into front (anterior) and back (posterior) portions. Viewing the body from the front or back, as if you were looking at the face or behind someone, aligns with the coronal plane.
* **Transverse Plane (Axial or Horizontal View):** This plane divides the body into upper (superior) and lower (inferior) parts. When viewing the body as if you were looking down from the head (top view) or up from the feet (bottom view), you're looking at cross-sections taken along the transverse plane. This is common in imaging techniques like CT scans and MRIs, where cross-sectional images are produced.
* **Oblique Plane:** This is less commonly referenced but involves cuts (sections) made at an angle. This isn't one of the primary planes but is mentioned for completeness.



(FIGURE-2.2) Multiple cross sectional planes of human body

**BODY CAVITIES :**

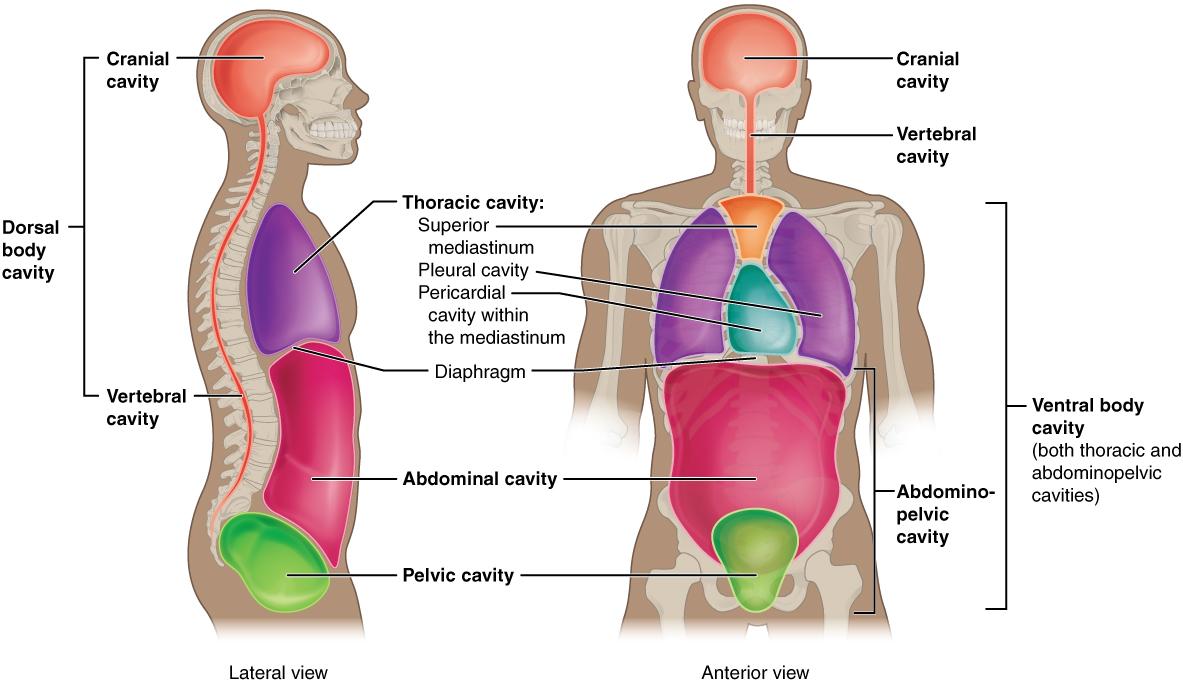
**Cranial Cavity:** Located within the skull, the cranial cavity houses the brain, encompassing structures such as the cerebrum, cerebellum, and brainstem. In cross-sections, the cranial cavity appears as a rounded or oval-shaped space surrounded by bone tissue, with intricate neural structures visible within.

**Thoracic Cavity:** Situated within the chest, the thoracic cavity contains vital organs such as the heart, lungs, and major blood vessels. Cross-sectional imaging reveals the heart centrally positioned, with the lungs bilaterally surrounding it. The mediastinum, a central compartment containing structures like the trachea and esophagus, is also discernible.

**Abdominal Cavity:** Positioned below the thoracic cavity, the abdominal cavity accommodates organs such as the stomach, liver, intestines, kidneys, and spleen. Cross-sectional views showcase a range of organs with varying densities and shapes, delineated by visceral and peritoneal layers.

**Pelvic Cavity:** Situated inferiorly to the abdominal cavity, the pelvic cavity contains reproductive organs, bladder, and portions of the digestive system. Cross-sectional imaging reveals structures such as the uterus, ovaries, prostate gland, bladder, and rectum, each exhibiting characteristic shapes and relationships within the pelvic space.

**Spinal Canal:** Extending longitudinally within the vertebral column, the spinal canal encases the spinal cord and nerve roots. Cross-sectional views depict the spinal cord centrally positioned within the vertebral canal, surrounded by cerebrospinal fluid and protected by vertebral bones and spinal ligaments.

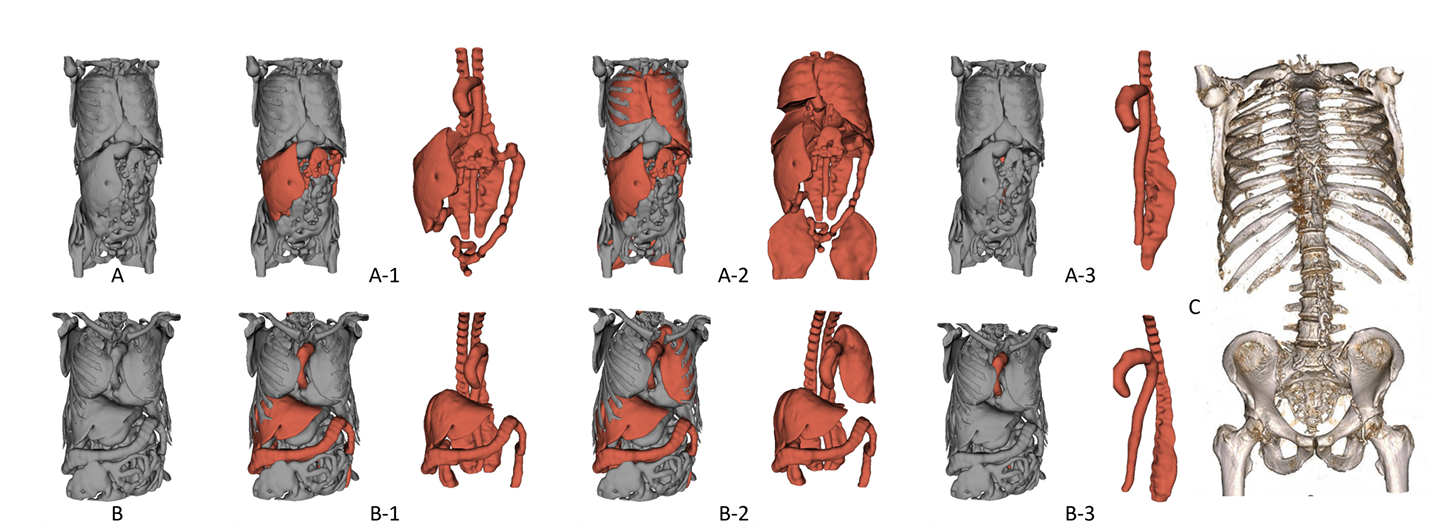
****

(FIGURE-2.3) Different cavities present in human body

**Available Dataset :**

MedShapeNet is a large collection of 3D medical shapes, including bones, organs, and surgical instruments. It contains over 100,000 shapes paired with annotations, making it valuable for computer vision research in the medical field.

For each set of segmentations, we randomly remove anatomies accounting for at least 10%, 20% and 40% of the entire segmentation’s volume to create the incomplete instances X. The original segmentations serve as the ground truth Y. Considering that anatomy ratios are subject-specific, different types and/or number of anatomies could have been removed for different subjects given the same threshold, as can be seen from Figure (3.1).



(Figure -3.1)

In general, using a 10% threshold removes more anatomies than using higher thresholds, and using a threshold of 40% removes only large anatomies, such as the aorta and the autochthonous back muscles. The small bones such as the individual ribs and vertebrae that form the skeleton enclosing the internal anatomies are generally not removed, providing a natural constraint for anatomy reconstruction.

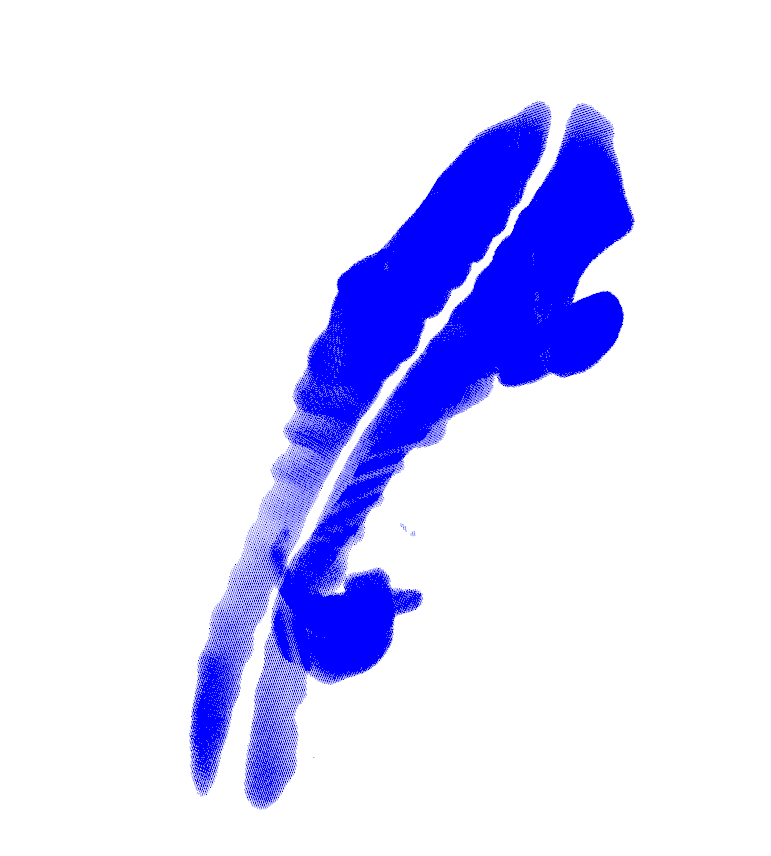
LINK TO AVAILABLE DATASET : [Download](https://files.icg.tugraz.at/f/b0623306eb9246be8c3c/?dl=1)

LINK TO VISUALIZATION OF DATASET :  [Visualize](https://drive.google.com/file/d/16TZf-CxGAcEgVSatO81urqseOyNz7KxC/view?usp=drive_link)

**Nifti image conversion to point cloud and visualization:**

A point cloud is a collection of data points defined by a given coordinate system. Each point represents a specific position in space and can have associated attributes such as color, intensity, or other information depending on the data acquisition method.

We will be working with point cloud representation, other possible formats to work with this is [Mesh](https://www.danthree.studio/en/blog-cgi/what-is-a-3d-mesh-model-definition-examples), [Voxel](https://en.wikipedia.org/wiki/Voxel).



(FIGURE 3.2) Point cloud image of Human spine

LINK TO VISUALIZATION IN POINT CLOUD - [Visualize](https://drive.google.com/file/d/12OdVpGB22diCA-2_Ak_CW9H9OecLOIXc/view?usp=drive_link)

**Data Augmentation Techniques:**

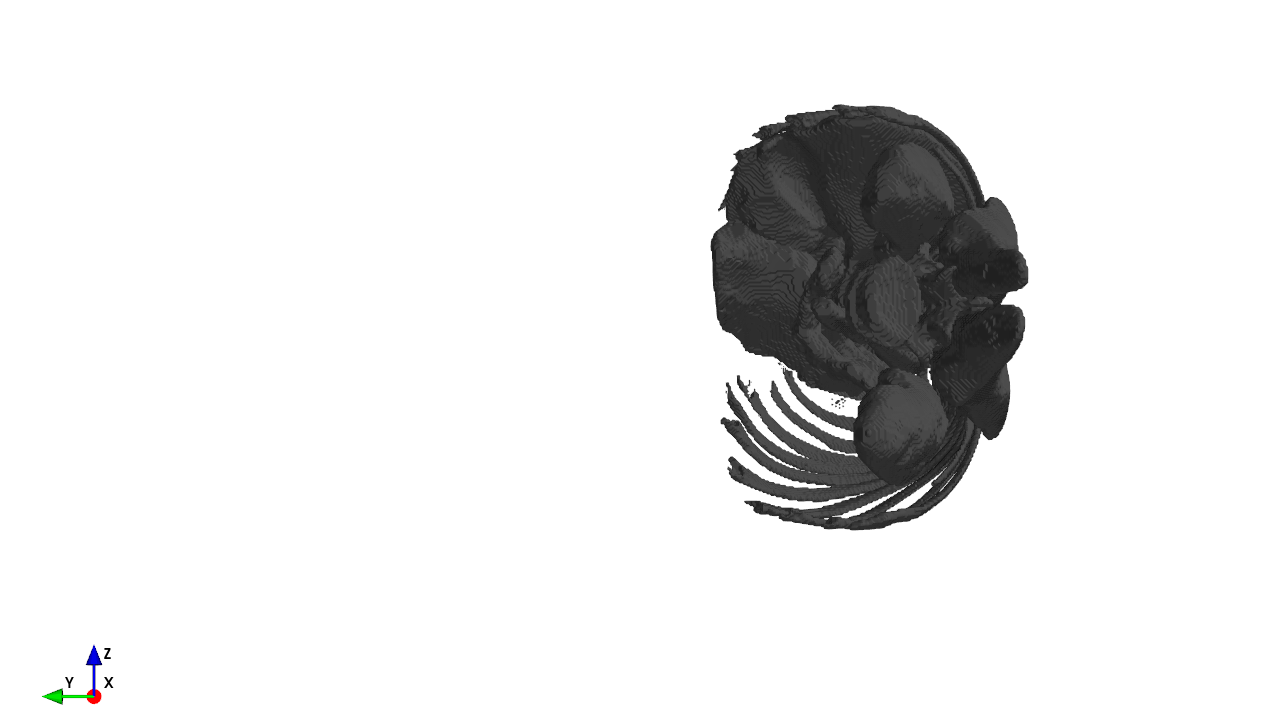
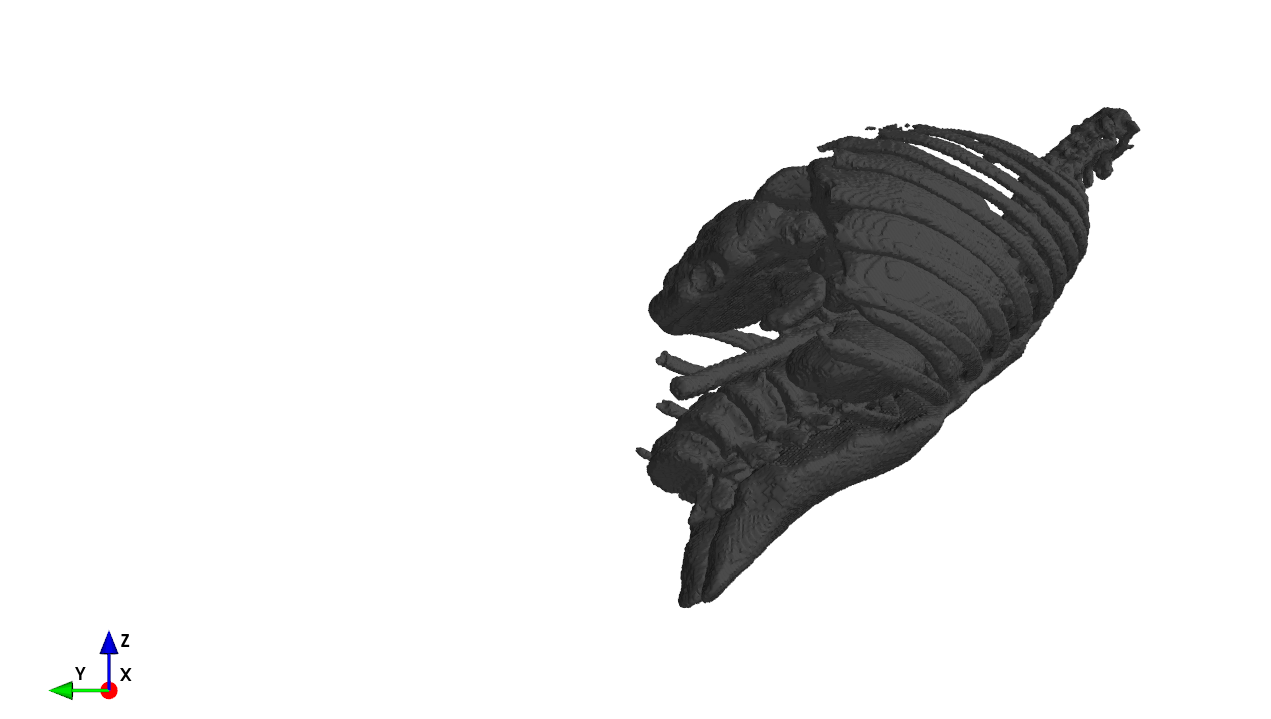
Our dataset leverages various data augmentation techniques to improve model robustness and generalization.The data is in two sets: a training set of **270** images and a testing set of **180** images.

Here's an overview of the augmentation techniques employed:

* **Baseline (No Augmentation):** This baseline serves as a reference point for evaluating the impact of other augmentation techniques.
* **Additive Gaussian Noise:** This technique introduces random noise to the images, mimicking real-world noise conditions and enhancing model resilience to noisy inputs.
* **Recursive Gaussian Filtering:** This method applies multiple Gaussian filters with varying parameters, simulating different levels of blurring and sharpening effects that can occur in real-world images.
* **Random Rotations (X, Y, Z Axes):** Images are randomly rotated around the X, Y, and Z axes, increasing the model's ability to recognize objects regardless of their orientation.
* **Random Flips (Horizontal/Vertical):** Images are randomly flipped horizontally or vertically, expanding the variety of object appearances and improving model performance under these transformations.
* **Brightness Adjustment:** Brightness levels are randomly augmented within a defined range, simulating different lighting conditions and enhancing model performance under diverse illumination settings.
* **Contrast Adjustment:** Contrast levels are randomly adjusted within a defined range, mimicking variations in object visibility and enhancing model robustness to image contrast variations.
* **Translation (X, Y):** Images are randomly translated in the X and Y directions, simulating slight shifts in object position and improving model ability to recognize objects regardless of their location within the image.

This comprehensive set of data augmentation techniques helps our model learn from a wider range of image variations, leading to improved generalization and robustness to real-world data.

Sample of rotation of nifti dataset .Figure 3.2 is before rotation and figure 3.3 is after rotation along X Axis.



(Figure-3.2) (Figure-3.3)

**Related Works:**

Since research in the field of 3D anatomical reconstruction is going, there have been a couple of Deep Learning models and papers published over medShapeNet dataset and have shown significant results.

* **3D anatomies reconstruction using Denoising Auto-encoders**

The medShapeNet dataset classified into complete and incomplete testing/training images is being used.

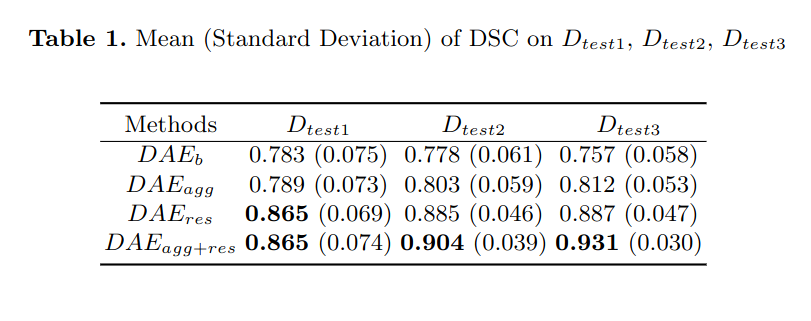
The main focus of the model and the paper is to denoise the 3D images and reconstruct them, since there have been a few missing organs by (10%, 20% and 40%) in an incomplete 3D image. These missing organs are treated as noise introduced in the images (we are working on the same dataset).

The images are passed to an encoder after certain preprocessing pipeline where encoder consist of various convolution layers followed by an activation function, this leads to reduced representation of the 3D images (encoding) which is then passed to a decoder having multiple deconvolution layers followed by activation function.

The final output image is then used with the complete corresponding 3D image of the input and dice loss is calculated and optimized. There have been certain variations in the architecture where instead of complete reconstruction the residual part is constructed.

There have been a total of 4 different methods used:

* Baseline model
* Aggregation
* Residual
* Aggregation+Residual



(FIGURE-4.1) Quantitative results of DAE based models

LINK TO THE PAPER - [View](https://arxiv.org/pdf/2309.04956.pdf)

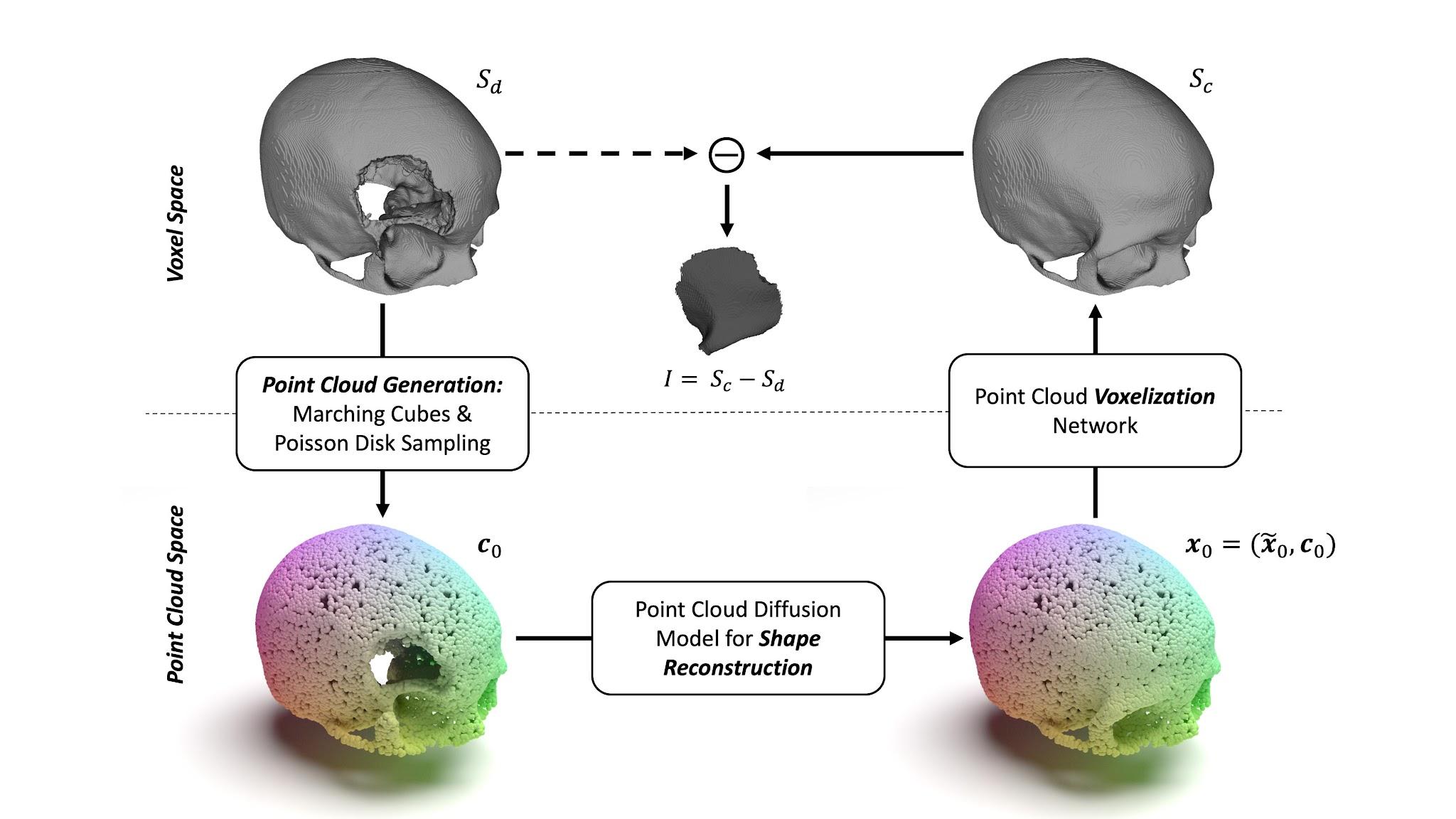
LINK TO THE CODE - [Github](https://github.com/shrey7ansh07/medshapenet-feedback/tree/main/anatomy-completor)

* **Point cloud diffusion model for broken skull reconstruction**

The paper describes the use of a diffusion model for reconstructing a broken skull. The image dataset is a baseline broken skull complete-incomplete dataset that works by adding noise to the missing segment of the skull.

It then uses the diffusion model for learning the noise and again introducing it in various time steps as per the traditional diffusion model. For the reconstruction loss it uses normal estimation and offset estimation on the complete image of the skull and uses a mean squared error for the optimization task.

Here is the view of the architecture of the model:



(FIGURE-4.2) Architecture of the point cloud diffusion model for broken skull reconstruction.

LINK TO THE PAPER - [View](https://pfriedri.github.io/pcdiff-implant-io/)

LINK TO THE CODE - [Github](https://github.com/pfriedri/pcdiff-implant)

**Our method:**

***The main outline of our method comes from the*** [***paper***](https://bmvc2022.mpi-inf.mpg.de/0569.pdf)***, where the very researched topic about converting a 2D image to its corresponding 3D version is discussed. The same have been incorporated in our paper with multiple variations and optimized blocks.***

We will train our code on the DAE model dataset consisting of 270 training 3D nifti images for complete/incomplete anatomies i.e a total of 27 complete anatomy images with their respective noised anatomies by 10, 20 and 40 percent of volume.

The overall architecture of our deep learning model will be inspired mainly by the paper above with some tweaking from earlier models.

Below is the architecture of our proposed model which is divided into 8 tasks:

1 - Image preprocessing

2 - Image conversion

3 - Missing organs image reconstruction

4 - Insightful information embedding (features concatenated)

5 - Image generation

6 - Discrimination (Fake/Real)

7 - Optimization (Loss function)

8 - Metrics and Visual representation.

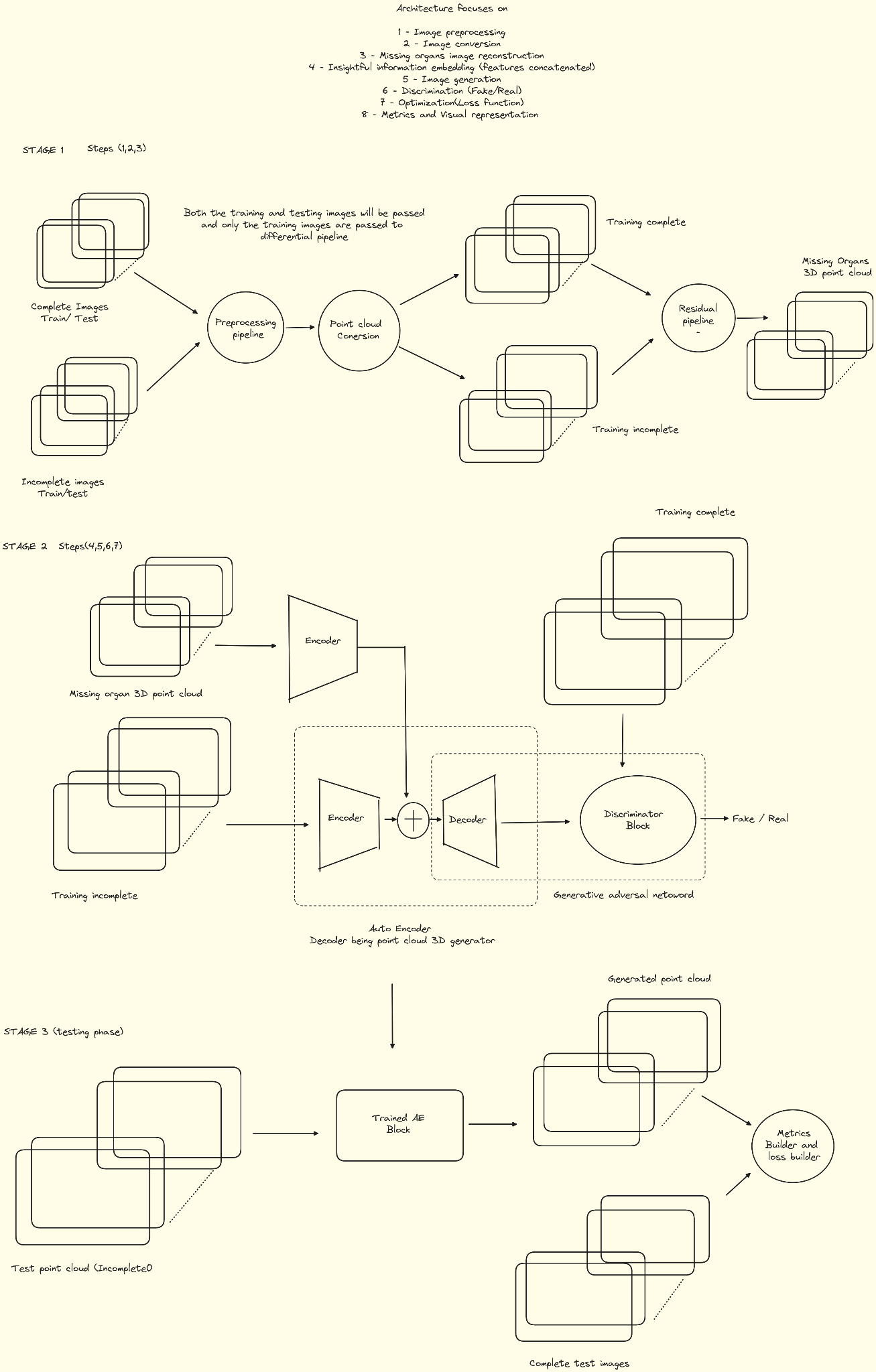
The training and testing have been divided into 3 stages:

1 - Data augmentation + image conversion + absent anatomy extraction.

2 - Model training (Auto-encoder+GAN).

3 - Optimization + metrics building + visualization (testing phase).

Here is the architecture of our proposed model:



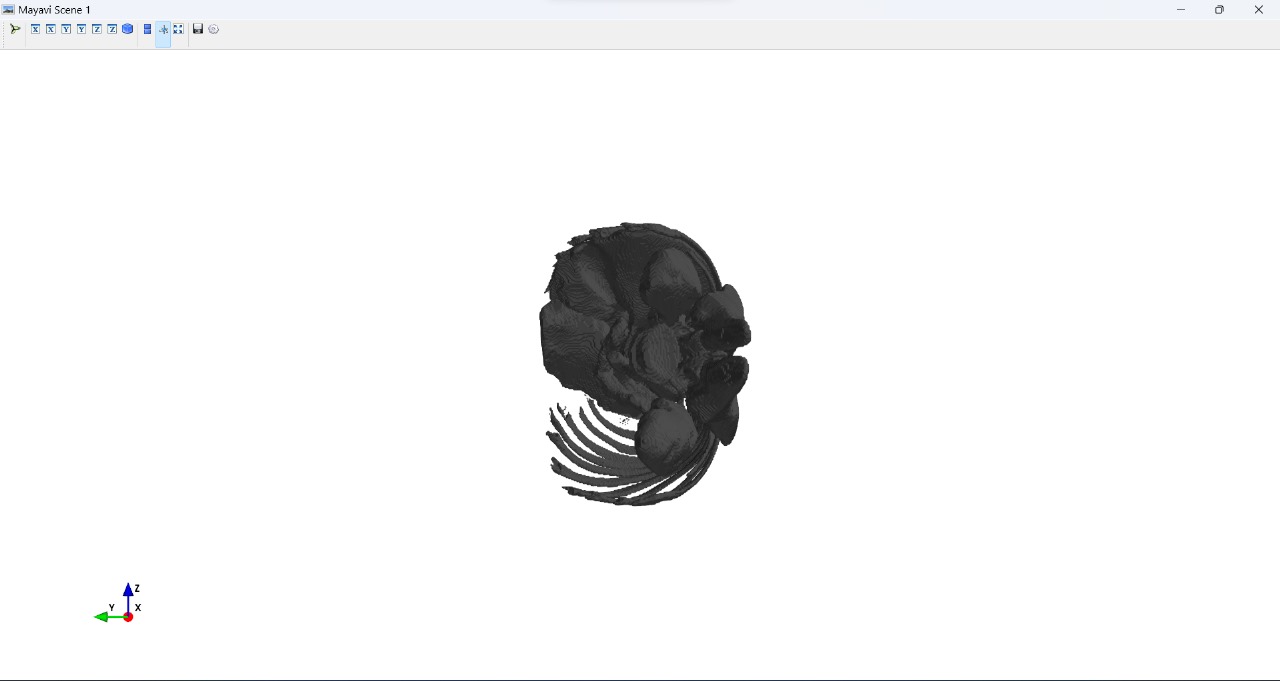
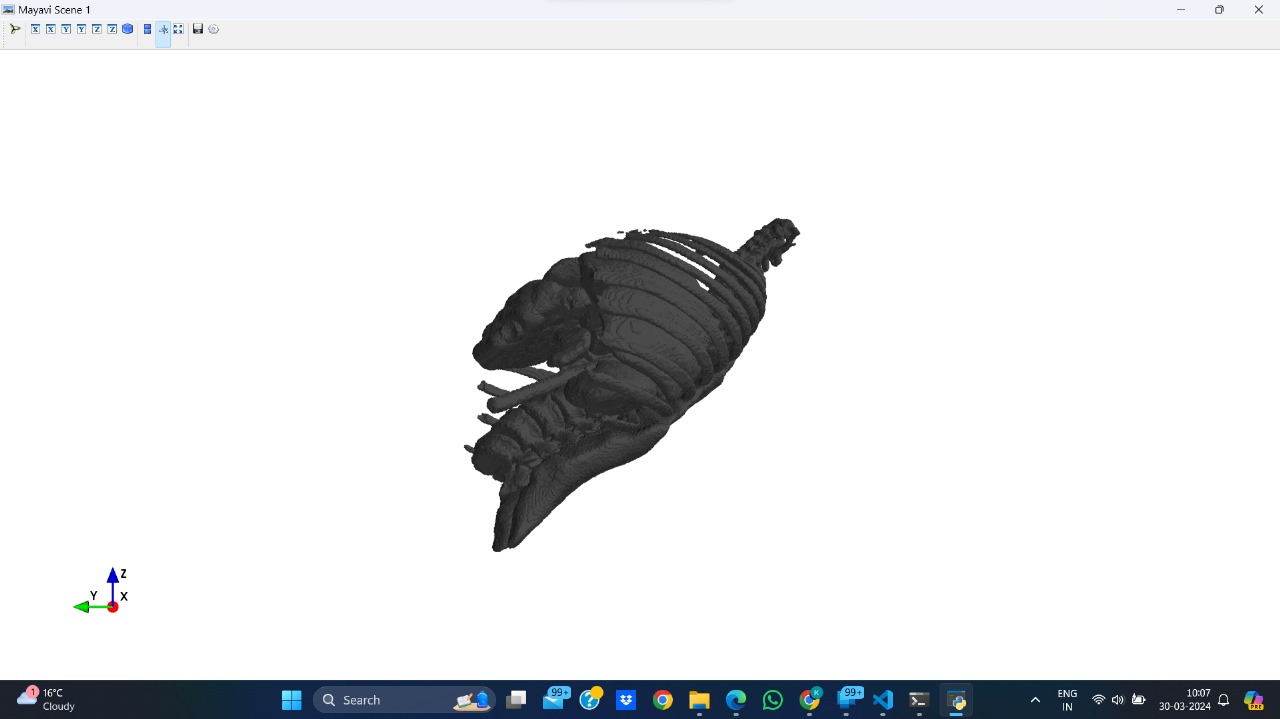
(FIGURE 5.1) Complete architecture of our missing anatomy completion

**STAGE 1 (Step 1 2 and 3) :**

* The first step is to perform data augmentation since we want variation in the dataset and also since the dataset is not very large as deemed for a precise training, augmentation techniques will help. Above were certain methods for the data augmentation, the code will randomly pick only 2-3 methods from the given set and will apply to the image input.

This is implemented only on the training data consisting of complete and incomplete anatomies.

One of the given implementation of the augmentation is given below where the image is rotated by 40 degree along all the combined axis.

Before rotation (real image) vs 40 degree rotation (augmented)

(FIGURE 5.2)

* The second step is to convert the nifti dataset images to corresponding point cloud image, this is done as we will be implementing point cloud auto-encoder+GAN, since working directly with 3D nifti images does not have many libraries, mesh images have their own complexity to work with, voxel images are make the model computationally intractable hence point cloud will be our go to solution.

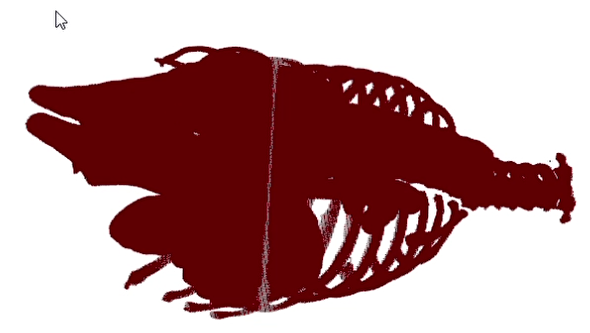
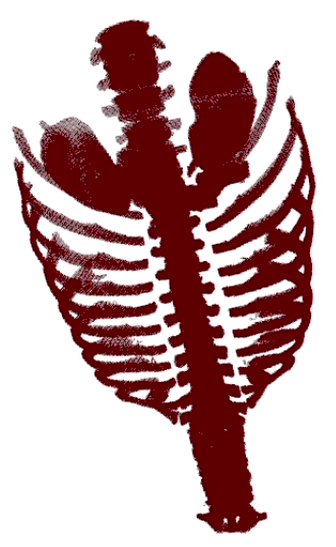
 

Nifti image vs point cloud

(FIGURE-5.3)

* The third step is extracting the missing anatomies for well supervised encoding, the point cloud complete and incomplete 3D images where a many-to-one incomplete to complete mapping, so once the difference is taken between the images we will get the corresponding missing anatomy.

This will also help to learn the reconstruction of the residual part as referenced from the DAE paper.

Complete Incomplete Missing

(FIGURE-5.4)

**STAGE 2 (Step 4 5 6 ) :**

Stage 2 will be the most essential part of the entire architecture having the entire training model. The idea behind the model is derived from the 2D -> 3D VAE+GAN paper + certain variations have been introduced from the diffusion point cloud and DAE paper, these have been a source of discussion so let’s break down what is happening in the model.

* An Auto-encoder block is used to encode the incomplete anatomy, for the encoder block different models have been viewed like 3D U-NET, Convolution on voxel space, PointNet encoder etc however the PointNet encoder will be most likely be finalized because of certain reasons. The pointNet encoder is specially meant for the 3D point cloud encoding with various optimization techniques, since point cloud is a dense representation, a traditional encoder block will have hard time encoding a nifti image or a point cloud image.
* A decoder block that will act as the generator as well for the GAN segment of the network, the decoder will be held responsible for generating image from the encoded vector representation of the 3D incomplete image, this will then be taken to the discriminator block which will have 0 and 1’s as their corresponding label for the image being generated and real one, various decoder can be used some of them comes from 3D U-NET, traditional decoder (deconvolution on voxel images), flow based. We will stick to the flow based decoder since it has a better ability to learn complex distributions of the latent space and can learn multi-modal data pretty well.
* Discriminator block will act as the judge of the generated image, being the part of the GAN that will take images in batches and label them as 0’s and 1’s and then it will help to discriminate between the real and generated images.
* Along with the Auto-encoder + gan segment we are also having another encoder that will take the residual or missing anatomies extracted from the STAGE 1 of the architecture, we will use this to shape our encoding more precisely since we know very well about the final outline of the image. The encoding vector from this encoder will be concatenated to the encoding of the pointNet encoder and then it will be passed to decoder, this will help in accurate learning of the model to lean towards precise reconstruction of small missing anatomies.

**STAGE 3 (Step 7 8) :**

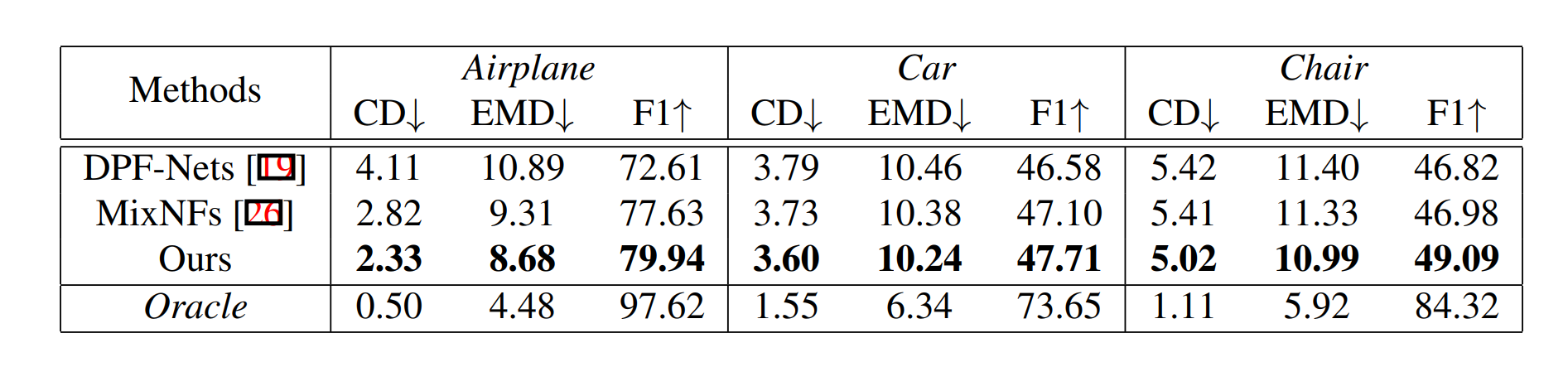
For optimization we are using three loss functions 👉

1. Chamfer Distance(CD) : It computes the squared

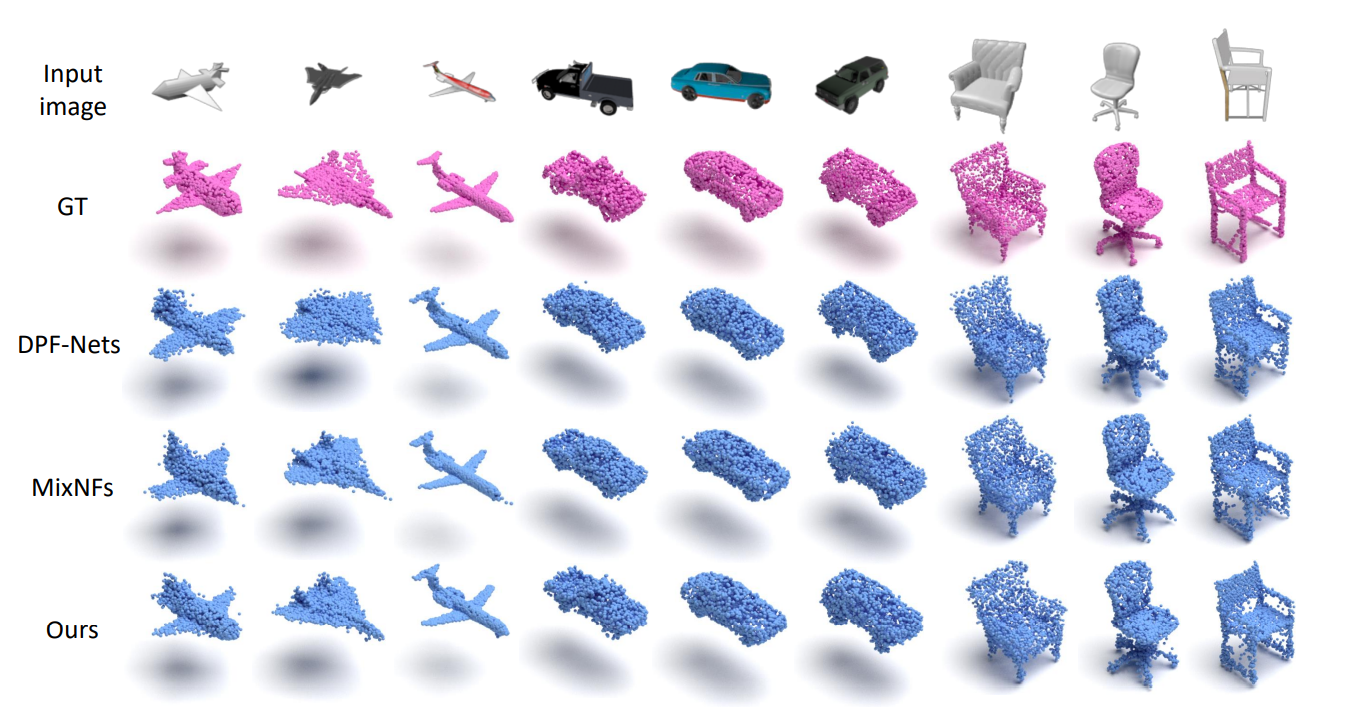
distance between each point in one set and its nearest neighbor in another set.

1. Earth Mover’s distance (EMD) : It measures the distance between two sets by attempting to transform one set into the other.
2. F1-Score (F1) : It is used to measure the similarity between the prediction and the corresponding ground-truth.

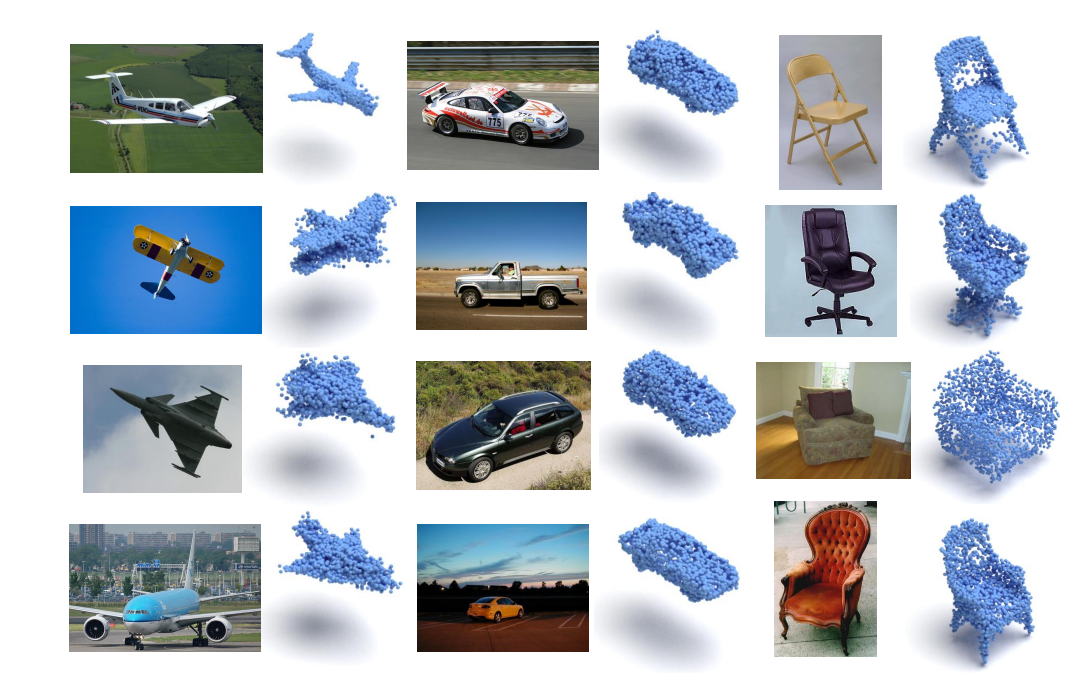
For CD and EMD, a lower score signifies a better model. F1 indicates the percentage of points that are correctly reconstructed, i.e., Euclidean distance between each prediction and ground-truth under a certain threshold **𝝉.**

(FIGURE 5.5)The best results are shown in **bold**

**Here is the View of output of the dataset from the PASCAL 3D+ dataset which is used in the research paper.**

****

(FIGURE 5.6)



(FIGURE 5.7) Qualitative Results of the PASCAL 3D+ dataset.

For our data visualization needs, we leverage several powerful libraries:

1. [Mayavi](https://docs.enthought.com/mayavi/mayavi/): A versatile scientific data visualization tool, Mayavi allows us to render complex 3D visualizations with ease.
2. [Pyvista](https://tutorial.pyvista.org/tutorial/00_intro/index.html): Another excellent library for 3D visualization, Pyvista enhances our ability to create stunning visual representations of our data.
3. [PyOpenGL](https://pypi.org/project/PyOpenGL/): PyOpenGL provides us with additional capabilities for rendering and interacting with 3D graphics in our visualizations.

In our visualization pipeline, we rely on Mayavi and SimpleITK to convert 3D images into point clouds and extract arrays for further analysis and visualization.

Explore our [GitHub repository](https://github.com/shrey7ansh07/3D-missing-anatomy-reconstruction) for a comprehensive look at our related works and projects.