3D model reconstruction from 2D images

- Authors –

Shubh Patel, Tejas Patel

- Project Overview -

Our project is centered around developing a deep learning-based system capable of generating 3D models from 5 2D images which are taken from various angles of a single object. This research aims to push the boundaries of current computer vision applications, enabling more seamless transitions between two-dimensional image data and three-dimensional object reconstructions from different 2 dimensional images . By leveraging advanced deep learning techniques, we are focusing on creating a robust pipeline that ensures accurate and efficient 3D model creation as well It will the label using cnn for the images provided.

- Technical Approach –

We are using CNN to extract the features from different images and then we will use Fully connected layers to transform the features into 3d model and output will be a 3d tensor and for 3d model reconstruction we are going to use many algorithms such as 3D-R2N2, Pixel2Mesh, Occupancy Network and many such.

We are using "Pix3d" dataset which was unarranged for our use but we have arranged it and prepared our training data and we have done manual cleaning as well the we have taken the help of python to arrange our data.

The main objectives of this project include:

1. Data Extraction and Arrangement: Efficiently sourcing and organizing data for optimal input into the model such as (batch size, number of images, channels, image height, image width).
2. Voxel Grid Creation: Generating 3D voxel representations as intermediary structures.
3. Model Training and Testing: Constructing a training model that incorporates batch processing and preparing comprehensive testing data to validate the results.

The scope of this project includes detailed coding, data handling, model construction, and iterative improvements to achieve reliable 3D reconstructions from 2D images.

- Challenges Encountered -

Throughout the development process, we faced several challenges, including:

1. Data Arrangement Complexity: Ensuring that the images were correctly aligned and labeled required meticulous attention to detail.

2. Memory Constraints: Managing large datasets and voxel grids proved to be a resource-intensive task, which we mitigated by implementing batch processing.

3. Model Performance: Achieving optimal performance in training required iterative adjustments and debugging to refine the learning process.

- Work Completed -

1. Data Extraction and Preprocessing

One of the initial phases of the project involved data extraction from an unorganized dataset containing multiple 2D images of various objects. We implemented a series of preprocessing steps, which included image cleaning. The images were systematically arranged and labeled to create a structured dataset suitable for deep learning tasks.

Following the extraction process, we pass the mesh 3d model that act as intermediary 3D representations of the 2D images. These mesh grids provide a foundational structure that allows the model to better learn and interpret spatial relationships inherent in the data.

2. Dataset File Creation and DataLoader Integration to streamline the data handling process, we created a comprehensive dataset file. This file serves as a directory of image paths, ensuring seamless data integration into the deep learning pipeline. The DataLoader class was employed to handle data loading and batching efficiently. This method allows the dataset to be divided into manageable batches, defined by parameters such as:

- Batch Size: The number of images processed simultaneously.

- Number of Channels: The depth of the input (e.g., grayscale or RGB channels).

- Image Dimensions (Height and Width): Standardized image size to maintain consistency during model training.

This structured approach not only enhances the model’s learning capacity but also reduces memory consumption, facilitating smoother training and validation processes.

3. Initial Training Model Development significant progress has been made in constructing the training model. The architecture of this model is designed to take in 2D image batches and their corresponding voxel grids to learn the mapping between 2D views and 3D structures. Key components of the model include convolutional layers for feature extraction, followed by fully connected layers that compile and interpret learned features to predict the 3D shape.

We have also ensured that the testing dataset is fully prepared. This dataset comprises five 2D images of each object captured from different angles, paired with the corresponding 3D model and voxel grid representation. The structured nature of this testing data is critical for evaluating the model's accuracy and robustness.

- Current Status -

Although we have completed the preliminary phases of data extraction, arrangement, and initial coding, the training model is still undergoing development. Our efforts are currently focused on optimizing the training process to achieve convergence and improve model performance. Finetuning the hyperparameters and adjusting the architecture are ongoing tasks aimed at enhancing the accuracy and efficiency of the system.

The next step involves completing the training phase and beginning rigorous testing and validation. Once the training is complete, we will conduct a series of performance evaluations, including comparisons with existing 3D reconstruction models, to measure the effectiveness of our approach.

- Future Work and Goals -

- Completing Model Training: Ensuring that the model achieves satisfactory performance metrics.
- Validation and Testing: Implementing robust testing protocols using the prepared dataset.
- Model Optimization: Applying post-training improvements, such as fine-tuning and regularization, to enhance the output quality.
- Result Analysis and Reporting: Conducting a comprehensive analysis of the results and compiling a detailed report to document findings and observations.

- Conclusion -

This project has reached several important milestones, including successful data extraction, voxel grid creation, and the development of the initial training model. The training phase is still in progress. By continuing our focused efforts, we aim to complete the model training and move on to validation and refinement. Ultimately, our goal is to produce a reliable system capable of constructing accurate 3D models from 2D image inputs, contributing to advancements in the field of deep learning and 3D modeling.