Final Report

**Title:** 3D Model Reconstruction from 2D Images  
**Authors:** Shubh Patel, Tejas Patel

**Abstract**

This deep learning project focuses on reconstructing 3D objects from 2D images captured from different angles. The goal is to take multiple 2D images of an object and reconstruct its 3D shape using advanced deep learning techniques. Bridging the gap between 2D visual data and 3D object reconstruction has significant applications in fields like Augmented Reality (AR), Virtual Reality (VR), gaming, robotics, and 3D printing. Utilizing convolutional neural networks (CNNs) and advanced reconstruction algorithms, this study achieves accurate and efficient transformations of 2D images into 3D voxel grids.

As input we provided some images with different angled photos as well as suitable masks for each photo and then feed it to the network and as a result we got the output as a 3D object constructed by very small dots. We have achieved a satisfactory result by using this model with pix3D dataset and handpicked input, validation and testing sets.

We have successfully achieved 70% of precision, accuracy and recall with this model.

**1. Introduction**

Reconstructing 3D models from 2D images is a critical task in computer vision with numerous applications, including AR/VR, robotics, gaming, and 3D printing. Most visual data, such as images and videos, are captured in 2D, while objects naturally exist in 3D. Bridging this gap enhances the realism and functionality of virtual environments and enables efficient 3D model generation for practical applications. This project explores deep learning techniques to create an end-to-end pipeline for transforming multiple 2D images into 3D models. Our approach emphasizes efficiency and accuracy in generating high-quality 3D reconstructions.

**2. Related Work**

Significant research has been conducted in the domain of 3D reconstruction.

1. Choy et al. proposed 3D-R2N2, which uses recurrent neural networks to predict volumetric shapes from 2D views

2. Wang et al. developed Pixel2Mesh, leveraging graph convolution networks for mesh-based 3D reconstruction.

3. Mescheder et al. introduced Occupancy Networks, which represent shapes as continuous fields, enabling high-resolution reconstruction.

Our work integrates aspects of these approaches, enhancing scalability and accuracy through preprocessing and tailored CNN architectures.

**3. Overview**

Here, we have mainly 3 components in our model,

1. Feature extractor – we are using CNN model to extract features from the images.
2. Feature Aggregator – Here, we are combining all collected features into a latent.
3. Decoder – It transforms the latent files into 3D voxel grid.

**4. Problem Statement**

Most visual data captured in daily life, such as images or videos, are 2D, while objects exist in 3D. Bridging this gap enables the creation of realistic models for AR/VR applications, gaming, and robotics. This technology can also be applied to 3D printing, where a 2D image generates a 3D model ready for printing, eliminating the need for manual creation in tools like Blender.

We are trying to make our lives and technology we use easy to use by integrating it into our life, 3D technology is very useful in that. Our project is makes 3D model from same object but in different scenarios, As we remove background with the help of masks it looses the focus on background and can know where the object is situated in the image.

There are so many specific challenges here, like too few images to train the model for one object and we use augmentation techniques like rotation and flipping to overcome that.

Another thing was we had a very big dataset of 13 GB and we downsized it to only 3 GB by manually filtering all the proper images, their masks as well as the voxel grid to train the model.

Our voxel grid size was also 80\*80\*80, which would take very long to be trained. So, we downsampled it to 50\*50\*50.

Every object has different threshold which has to manually adjusted to get proper output.

**5. Dataset**

The project employs the Pix3D dataset, which contains 2D images and corresponding 3D models in .obj format. This dataset spans nine object categories, including beds, bookcases, and chairs, among others. Each object has multiple 2D images captured from different angles, paired with masked images and ground-truth 3D models for reconstruction. Extensive preprocessing, including manual cleaning, alignment, manually created annotation files as well as handpicked images, was performed to create a structured and standardized dataset for training, validation and testing.

We has 70% training, 20% Validation and 10% of training data in dataset and all the data was fed into batches of 5 images along with their masks.

**6. Methods**

Our methodology includes:

1. **Feature Extraction:** CNNs extract spatial features from 2D images [5].
2. **Voxel Grid Generation:** Intermediate 3D representations are generated, allowing for structured learning.
3. **Model Architecture:** Advanced algorithms such as 3D-R2N2 and Pixel2Mesh are integrated to refine the reconstruction process.
4. **Data Handling:** Batch processing using a Dataloader framework ensures efficient training and scalability.

**7. Experiments**

Initial experiments validated the feasibility of the pipeline:

* **Data Preprocessing:** Successful standardization of the Pix3D dataset for model compatibility with manual filtering and augmentation techniques.
* **Training:** Iterative refinement of the model architecture improved early performance.

We have used adam and L2 regularization loss to make the model performance better.

* **Testing:** Validation against the structured dataset demonstrated robust 2D-to-3D mapping.

**8. Conclusion**

Our project has achieved significant milestones in dataset preparation, model design, and initial training. Future work includes further optimization, testing, and comparative analysis with existing approaches to measure the system’s effectiveness. This research aims to contribute to advancements in 3D reconstruction technologies, driving applications in multiple fields.

For future improvements we will refine the model, so it could handle more complex models and we are trying to enhance the resolution of the output 3D object.

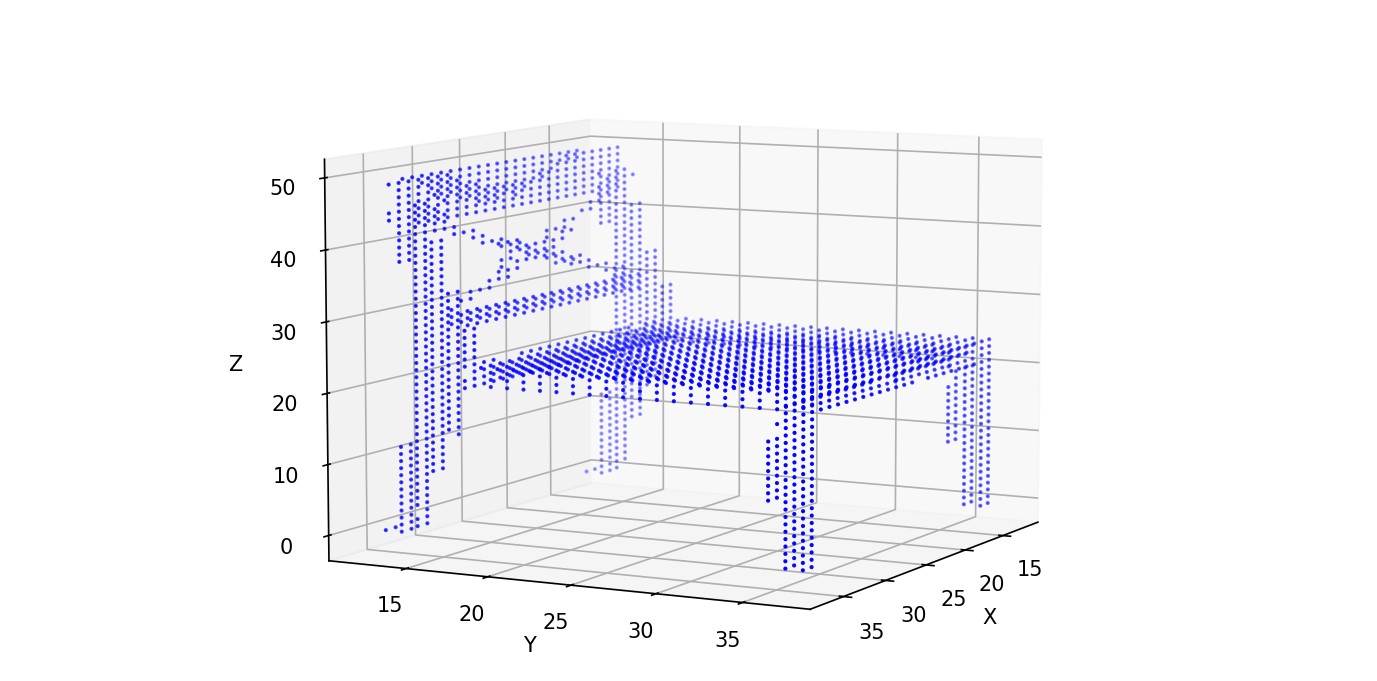
**Supplementary Material**

* **Source Code:** It is available on our Github Repository.
* **Dataset Samples:**

Input images –

* 1. Chair

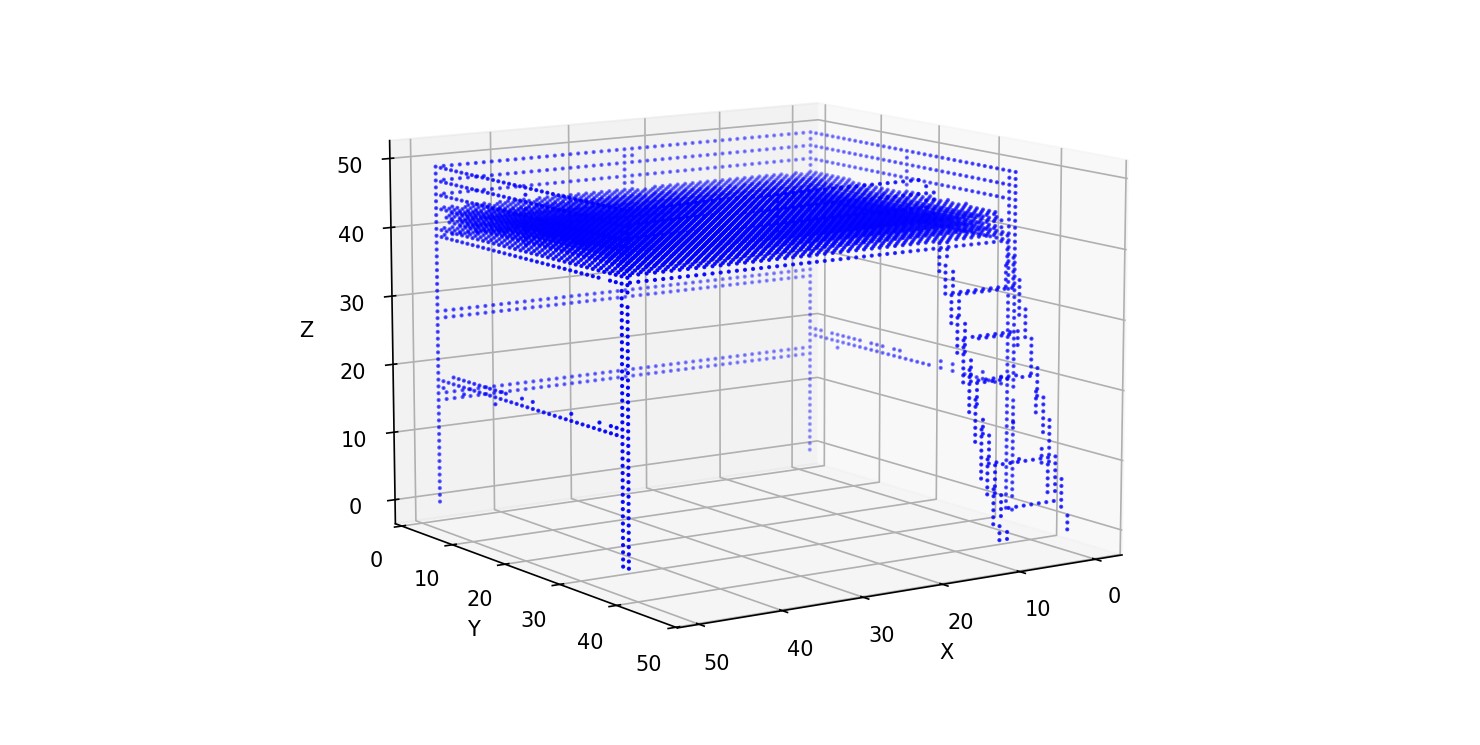


Output image –

* 1. Bed

Input image -



Output model -

**References**

1. C. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese, "3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction," *European Conference on Computer Vision (ECCV)*, 2016.
2. Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. Bronstein, and J. M. Solomon, "Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images," *European Conference on Computer Vision (ECCV)*, 2018.
3. L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy Networks: Learning 3D Reconstruction in Function Space," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
4. X. Sun, X. Wu, Z. Wang, X. Zhou, and S. Deng, "Pix3D: Dataset and Methods for Single-Image 3D Shape Modeling," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
5. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.