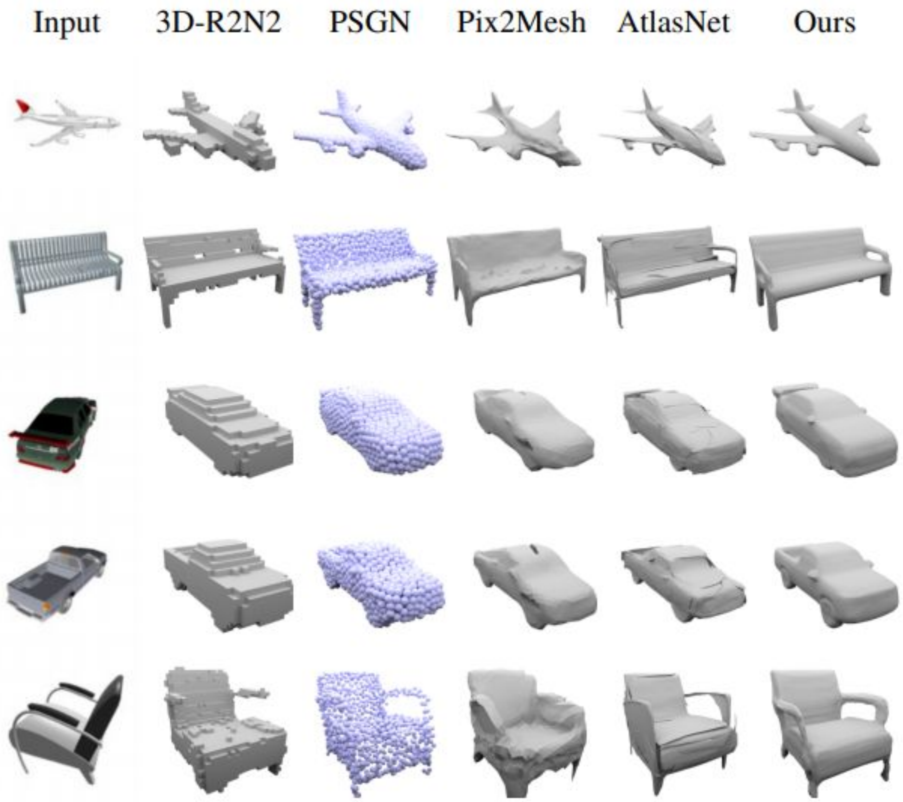
3D Reconstruction Occupancy Networks

Team Noisy Pixels:
Shubham Dokania (2020701016)
Shanthika Naik (2020701013)
Sai Amrit Patnaik (2020701026)
Madhvi Panchal (2019201061)

3D Reconstruction



Problem Statement

With the considerable pace of advancement in the field of deep learning and its applications, notable developments have been seen in the domain of 3D vision.

Unlike Images, there is no well defined, structured representation method in 3D which is both computationally and memory efficient, but allows arbitrary topology to represent high-resolution geometry.

However, considering functional space, we can map the 3D space to an implicit function, for which a non-linear classifier may exist.

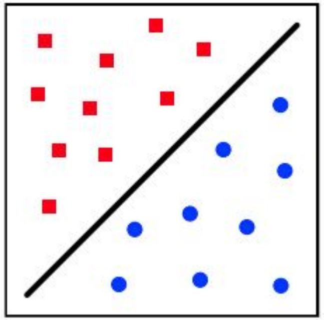
The task we wish to solve here is that of 3D reconstruction from a single view image, i.e. given an image of an object (without camera pose), we want to retrieve the 3D geometry of the object and reconstruct it in 3D space.

Goals

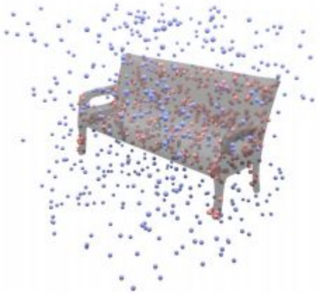
- In this project the 3D surface is implicitly learnt as a continuous decision boundary of a non-linear classifier instead of learning to predicting an explicit voxelized representation at a fixed resolution.
- A neural network learns the entire occupancy function that can be evaluated at any arbitrary resolution.
- At inference time, the mesh is extracted from the learnt occupancy function by using a Multi-resolution IsoSurface Extraction(MISE) technique and Marching cubes.
- The final mesh is refined using a 2 step process using Fast-Quadric-Mesh-Simplification algorithm and first and second order (i.e., gradient) information.

Occupancy Networks

- Model 3D surface as a decision boundary of a non-linear classifier.
- Returns an Occupancy probability of a 3D point conditioned on an input image.

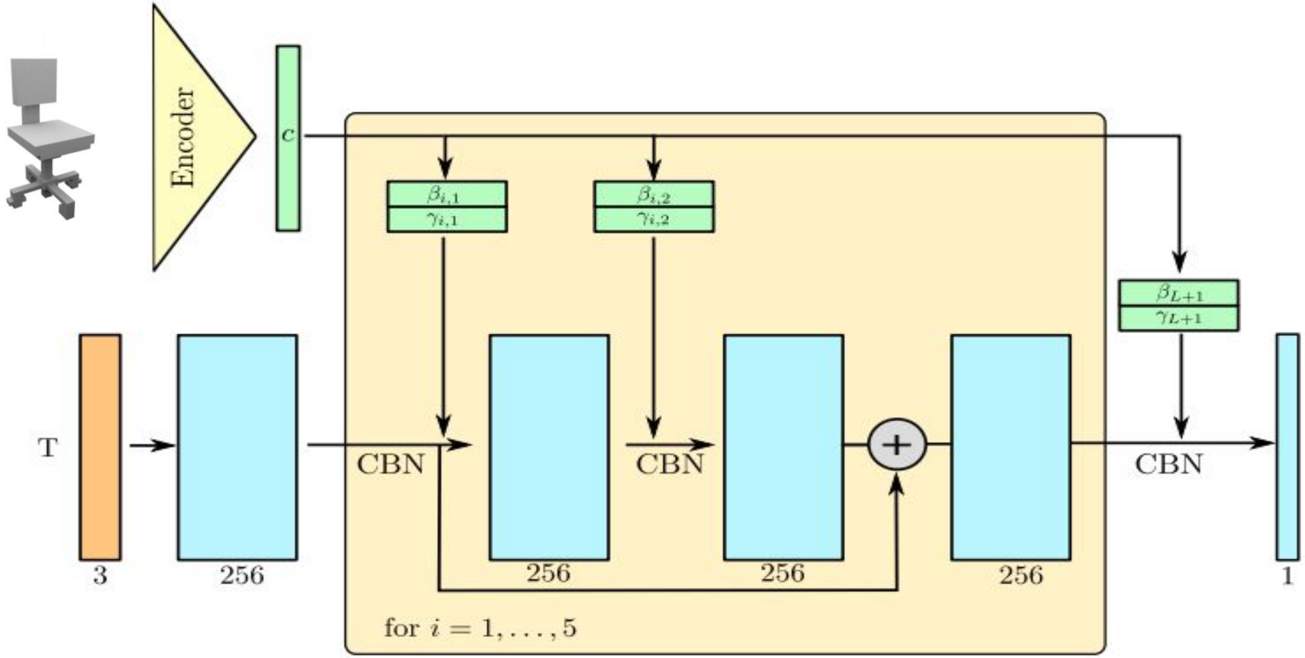


$$\underbrace{f_\theta : \mathbb{R}^3}_{\text{3D Location}} \times \underbrace{\mathcal{X}}_{\text{Condition (eg, Image)}} \rightarrow \underbrace{[0, 1]}_{\text{Occupancy Probability}}$$



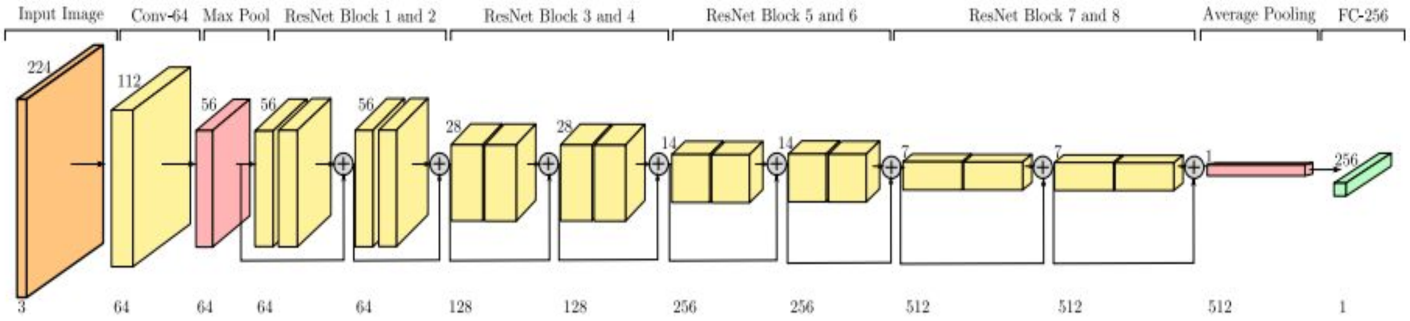
- Model 3D surface as a decision boundary of a non-linear classifier.
- Returns an Occupancy probability of a 3D point conditioned on an input image.

Network Architecture



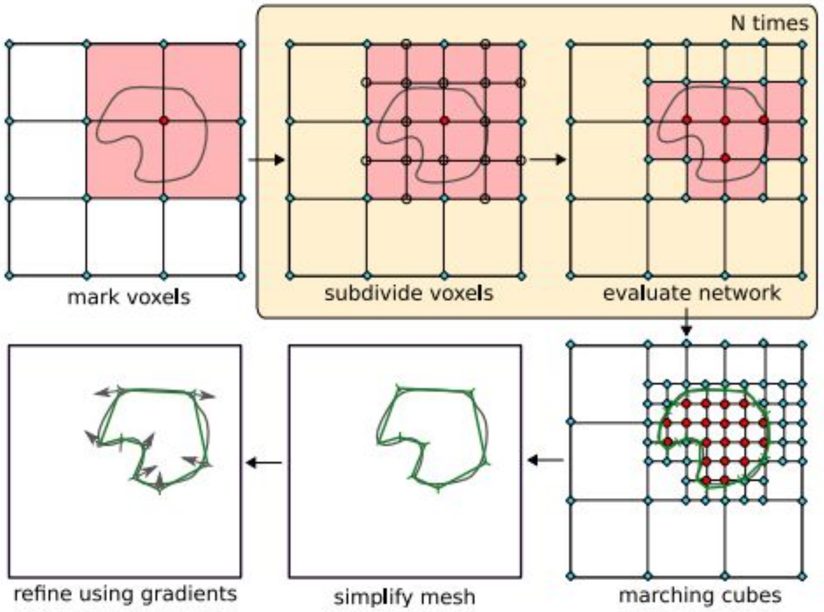
Occupancy Networks

- Occupancy network consists of a fully connected neural network with 5 ResNet blocks with the input conditioned it using Conditional Batch Normalization.
- The network takes as input a set of uniformly sampled 3D points.
- Image encoder encodes the input image to a feature vector.
- The extracted feature vector for the image is fed to the Occupancy Net using a Conditional Batch Normalization operation.
- Image Encoder is a ResNet 18 network with the final layer returning a 256D feature vector of the input image.



Iso-surface Extraction

We use the MISE (Multi-resolution Iso-Surface Extraction) method proposed in the paper; this is a hierarchical method which incrementally builds an octree.



We first mark all points at a given resolution which have already been evaluated as either occupied (red circles) or unoccupied (cyan diamonds).

We then determine all voxels that have both occupied and unoccupied corners and mark them as active (light red) and subdivide them into 8 sub-voxels each.

Next, we evaluate all new grid points (empty circles) that have been introduced by the subdivision.

The previous two steps are repeated until the desired output resolution is reached. Finally we extract the mesh and refine, as explained further.

Mesh extraction

- Once the desired resolution is reached, we use the Marching Cube algorithm[1] to extract an approximate isosurface :

$$\{p \in \mathbb{R}^3 \mid f_\theta(p, x) = \tau\}.$$

- The mesh obtained is further simplified using Fast-Quadric-Mesh-Simplification[2] algorithm.
- For this purpose, random points are sampled from each face of output and minimizing the loss :

$$\sum_{k=1}^{K} (f_\theta(p_k, x) - \tau)^2 + \lambda \left\| \frac{\nabla_p f_\theta(p_k, x)}{\|\nabla_p f_\theta(p_k, x)\|} - n(p_k) \right\|^2$$

Expected Results

Input image

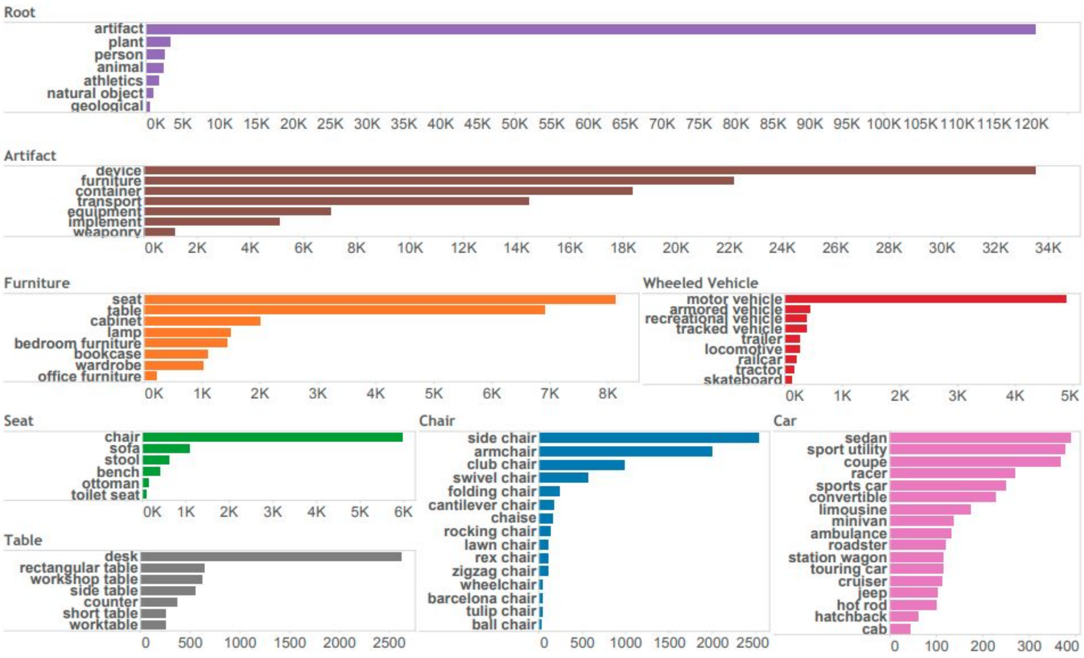
Output mesh

Expected Results

Input image

Output mesh

Dataset



- The dataset used in this paper is the *ShapeNet* dataset.
  - ShapeNet is a richly-annotated, large-scale repository of shapes represented by 3D CAD models of objects. It is the ImageNet equivalent for 3D data.
  - ShapeNet contains 3D models from a multitude of semantic categories and organizes them under the WordNet taxonomy.
  - ShapeNet has indexed more than 3,000,000 models, 220,000 models out of which are classified into 3,135 categories (WordNet synsets).

Expected Deliverables

The overall expectation from this project is an implementation of the occupancy network approach for 3D reconstruction. To outline the tasks in milestones, we shall follow the below deliverables:

1. Analysis of the ShapeNet Dataset relevant to 3D reconstruction.
2. PyTorch implementation of Occupancy Network for Single View Reconstruction.
3. Visualization of point clouds as inside/outside the surface as output of network.
4. Prepare demo for 3D mesh reconstruction using mesh simplification as available in open-source mesh-fusion library.
5. Compare performance metrics such as Chamfer distance, IoU, and Normal consistency with similar methods (available implementations).
6. Release all code as open-source and with proper instructions.

References

1. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S. and Geiger, A., 2019. **Occupancy networks: Learning 3d reconstruction in function space**. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4460-4470).
2. Vakalopoulou, M., Chassagnon, G., Bus, N., Marini, R., Zacharaki, E.I., Revel, M.P. and Paragios, N., 2018, September. **AtlasNet: multi-atlas non-linear deep networks for medical image segmentation**. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 658-666). Springer, Cham.
3. Bartsch, M., Weiland, T. and Witting, M., 1996. **Generation of 3D isosurfaces by means of the marching cube algorithm.** *IEEE transactions on magnetics*, *32*(3), pp.1469-1472.
4. M. Garland and P. S. Heckbert. **Simplifying surfaces with color and texture using quadric error metrics**. In Visualization'98. Proceedings, pages 263–269. IEEE, 1998.
5. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H. and Xiao, J., 2015. **Shapenet: An information-rich 3d model repository**. *arXiv preprint arXiv:1512.03012*.
6. H. de Vries, F. Strub, J. Mary, H. Larochelle, O. Pietquin, and A. C. Courville. **Modulating early visual processing by language**. In Advances in Neural Information Processing Systems (NIPS), 2017.

Thanks!

...