A Point Set Generation Network for 3D Object Reconstruction from a Single Image

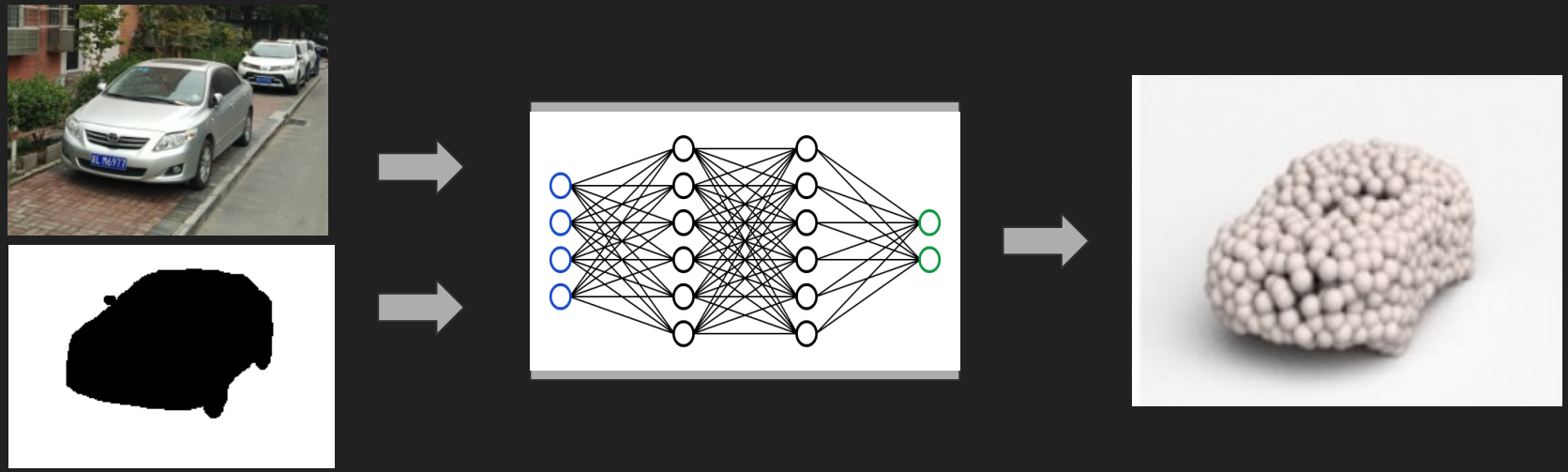
Paper by
- Haoqiang Fan*
- Hao Su*
- Leonidas Guibas

Presentation by
- Apoorva S(2019702014)
- Anurag S(2018121004)
- Kajal S(2019801006)

CVPR December 2016

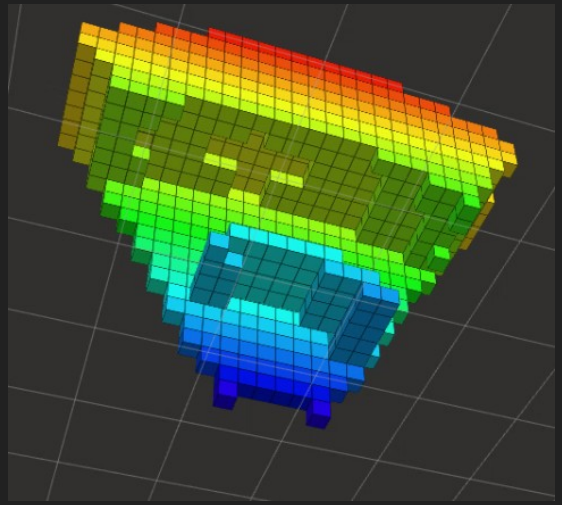
What is the project All About?

Addresses the problem of 3D reconstruction from a single image by giving 3D point cloud coordinates.



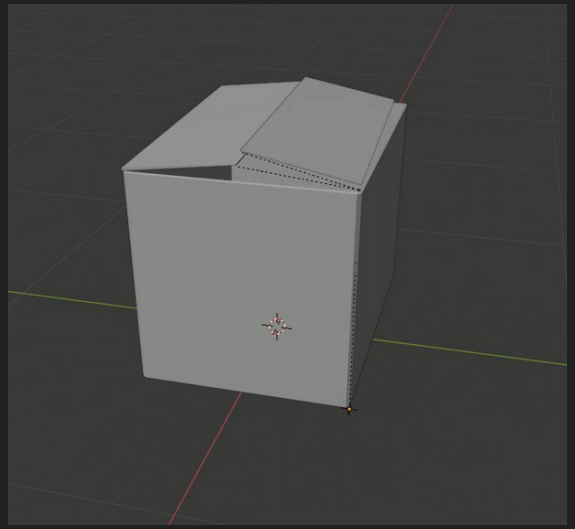
What's New in it?

- Till date, 3D representation is in the form of volumetric grids or collection of images, They suffered invariance during 3D geometric transformations.
- Point cloud representation can be easily transformed geometrically as associations between the components is not required to be updated during transformations.



Challenges for Point Cloud Generation using Neural Networks

- Neural network has to hallucinate complete 3D structure from a single view
- Training problem inherently different due to ground truth ambiguity
- Proper choice of loss function is crucial for meaningful results
- Design architecture has to encompass the possibility of multiple outputs



Proposed solutions to the Challenges

- The network has been designed in two ways to deal with different appearances : Vanilla Version and Two Branch Prediction
- The loss function has been chosen to account for the ground truth ambiguity and give the best output from the multiple possible results obtained.
- Loss function : Chamfer Distance for giving multiple possible results

$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2$$

- Additional Loss Function: MoN Loss for choosing best out of multiple choice

$$\underset{\Theta}{\text{minimize}} \quad \sum_{k} \min_{\substack{r_j \sim \mathbb{N}(\mathbf{0}, \mathbf{I}) \\ 1 \leq j \leq n}} \{d(\mathbb{G}(I_k, r_j; \Theta), S_k^{gt})\}$$

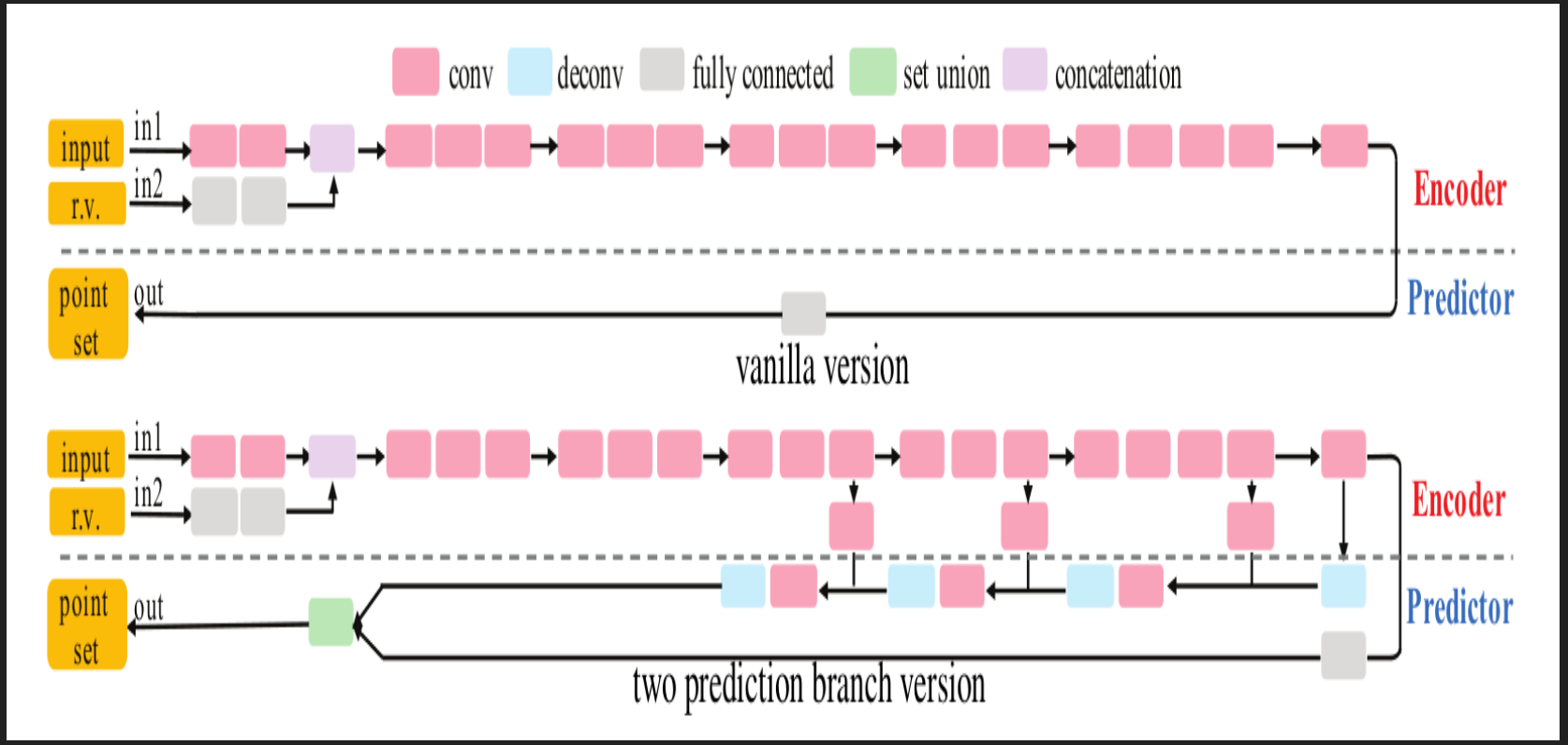
Conditional Generative Architecture Formulation

- **Goal -** Single 2D image (RGB or RGBD) to complete 3D shape

  $S = \{(x_i, y_i, z_i)\}^{N}_{i=1}$, N is predefined constant, N = 1024 seems sufficient

- The ground truth is a probability distribution, P(./I) over the shapes possible for input, I i.e. for given image the output 3D points come with a probability
- So, we train the Neural Network as a conditional sampler from P(./I)
  - $S = G (I, r, \theta)$
  - θ - n/w parameter
  - $r \sim N(O, I)$ - It is a random variable to perturb the input

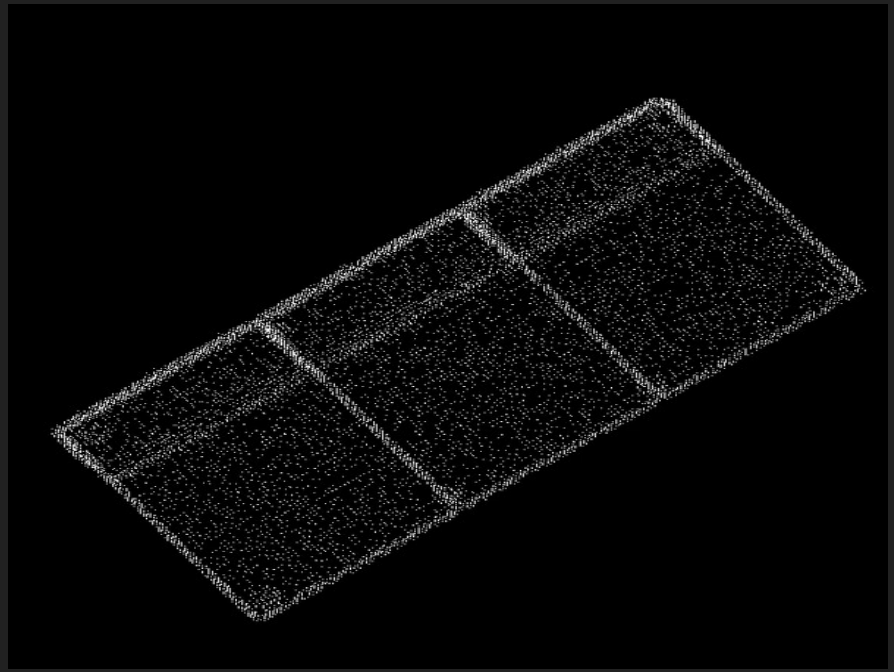
Conditional Generative Architecture



Experimental Details

- Input Images dimension 192x256
- The deconv branch produces 768 points which correspond to a 32x24 three-channel image.
- The fully connected branch produces 256 points.
- The training program is implemented in TensorFlow.
- Adam is used as the optimizer.
- Activation Function:ReLU

Input and Ground Truth

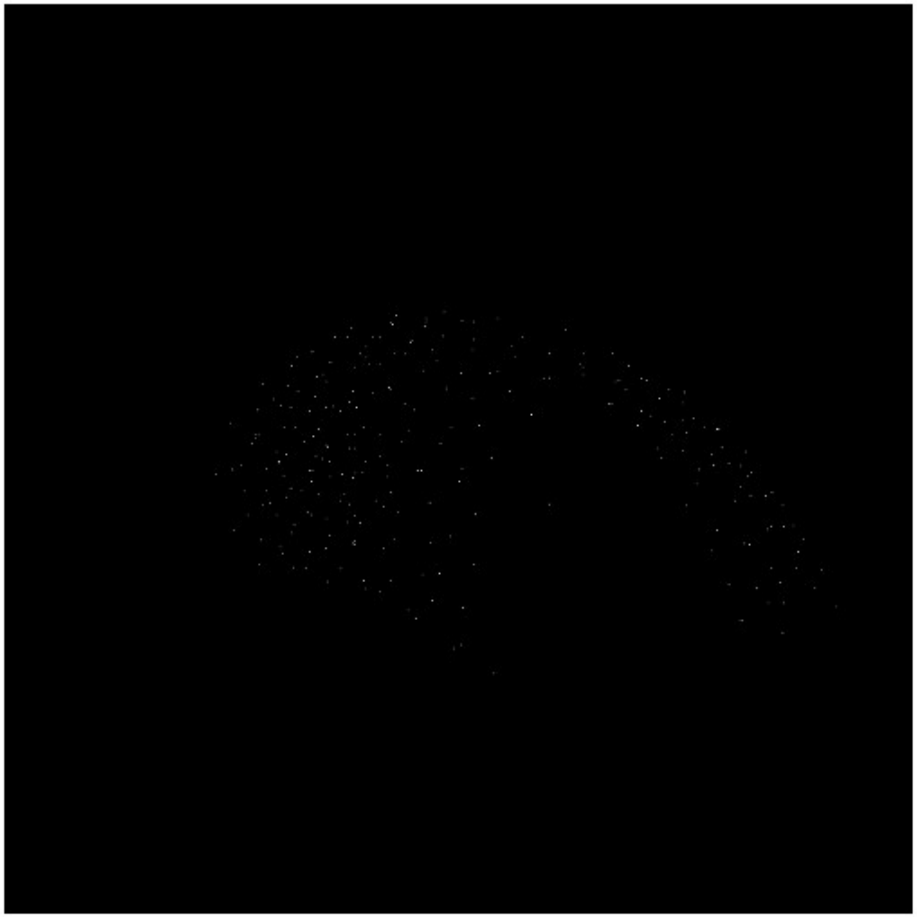


Experimental Results

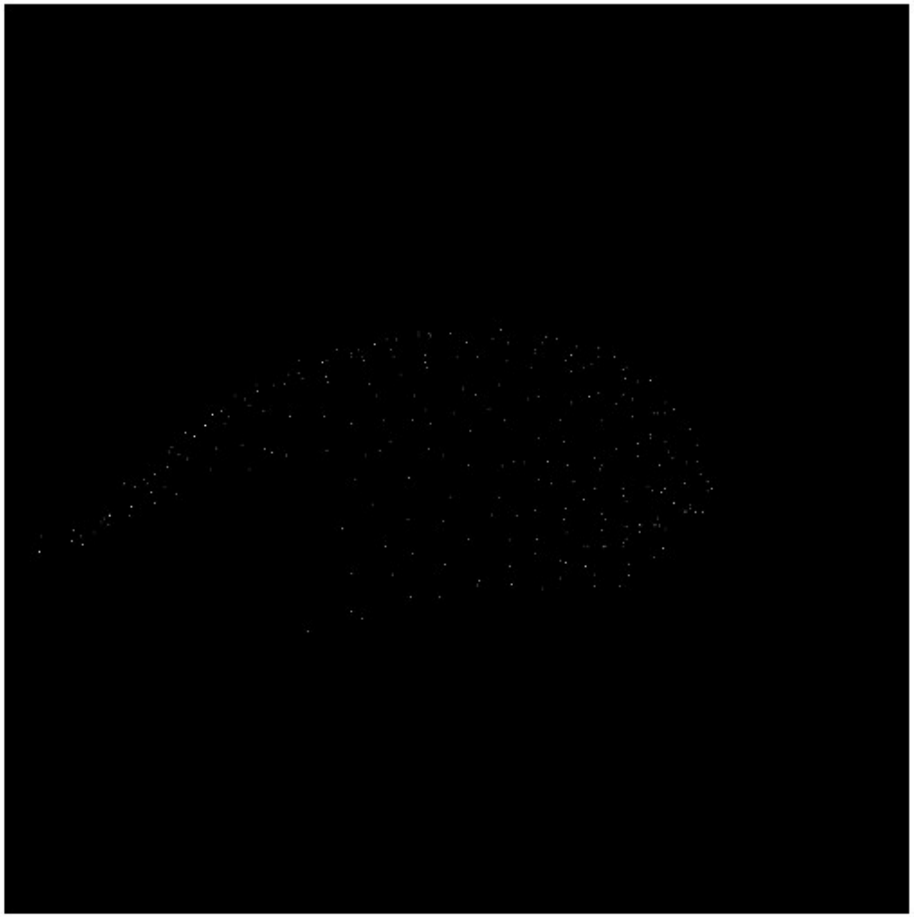
Input

Input_Mask

Output View 1



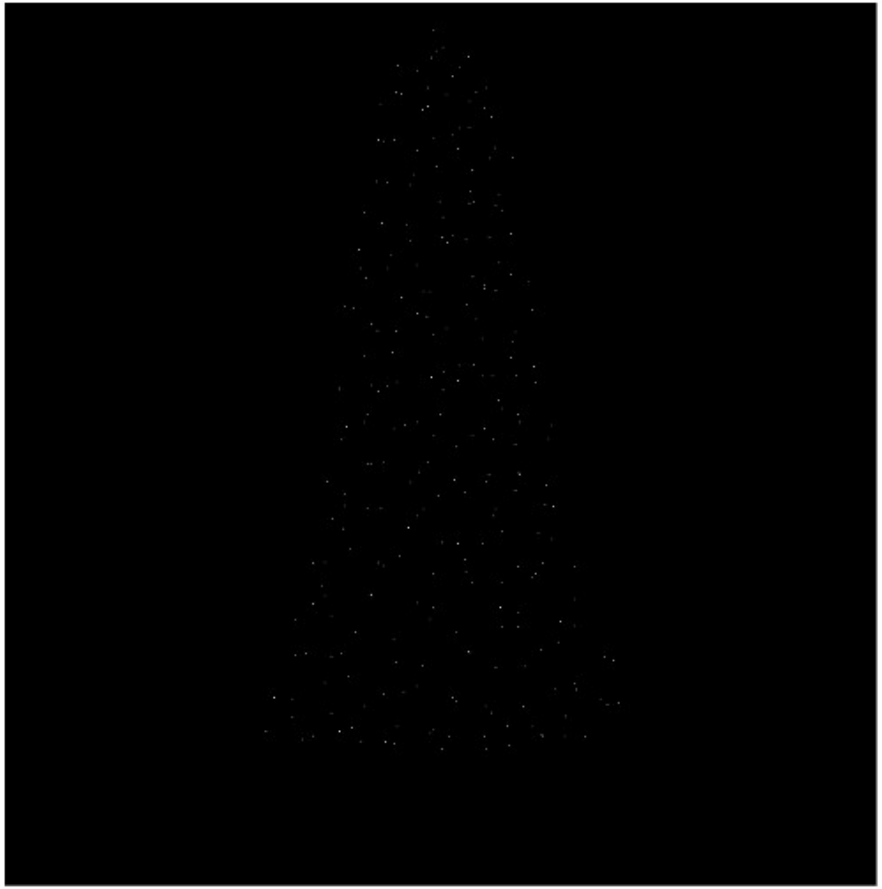
Output_View 2



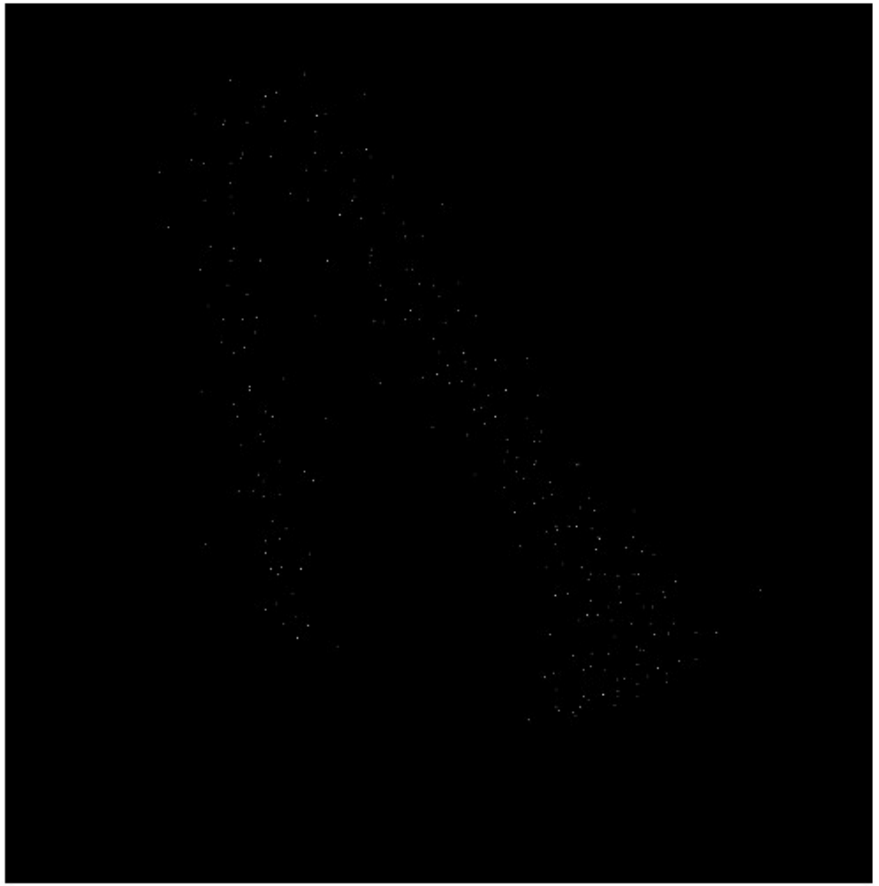
Input

Input_Mask

Output View 1



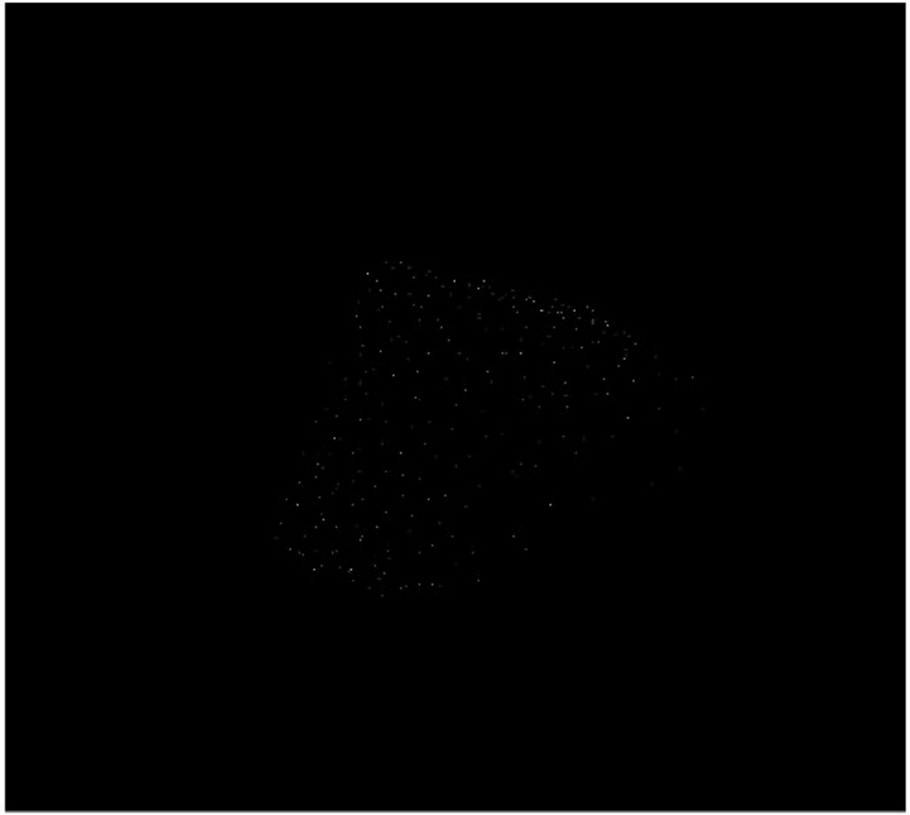
Output_View 2



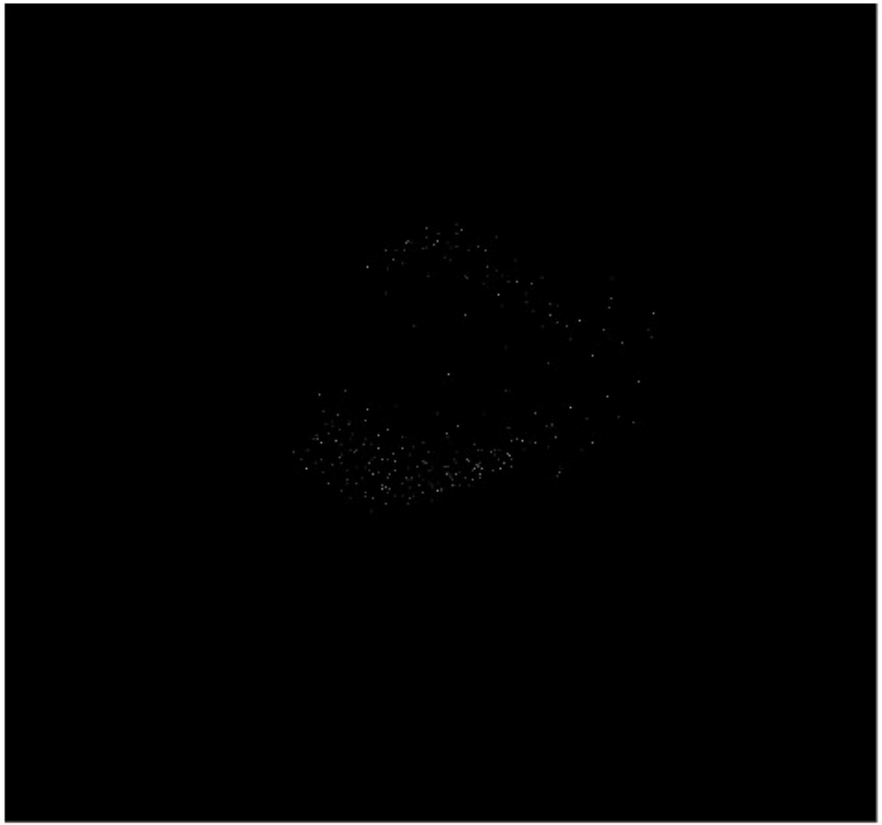
Input

Input_Mask

Output View 1



Output_View 2



Expected Learnings

- Learn and implement Conditional Generative Network
- Understanding 3D Convolution Network
- Implementation and Deployment of Neural Network
- Designing and Implementation of Custom Loss functions
- Working with TensorFlow and Cuda
- NN’s for 3D points reconstruction from single image

To be continued...

- Improvement of current results
- Implementation of Two Branch Prediction version
- Comparison of results
- Testing the model on real world data

Important Links!

- Link to Paper
- State of the Art for 3D reconstruction from single Image as of 2019 Mesh R-CNN by FAIR ICCV 2019.