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

Presentation Overview

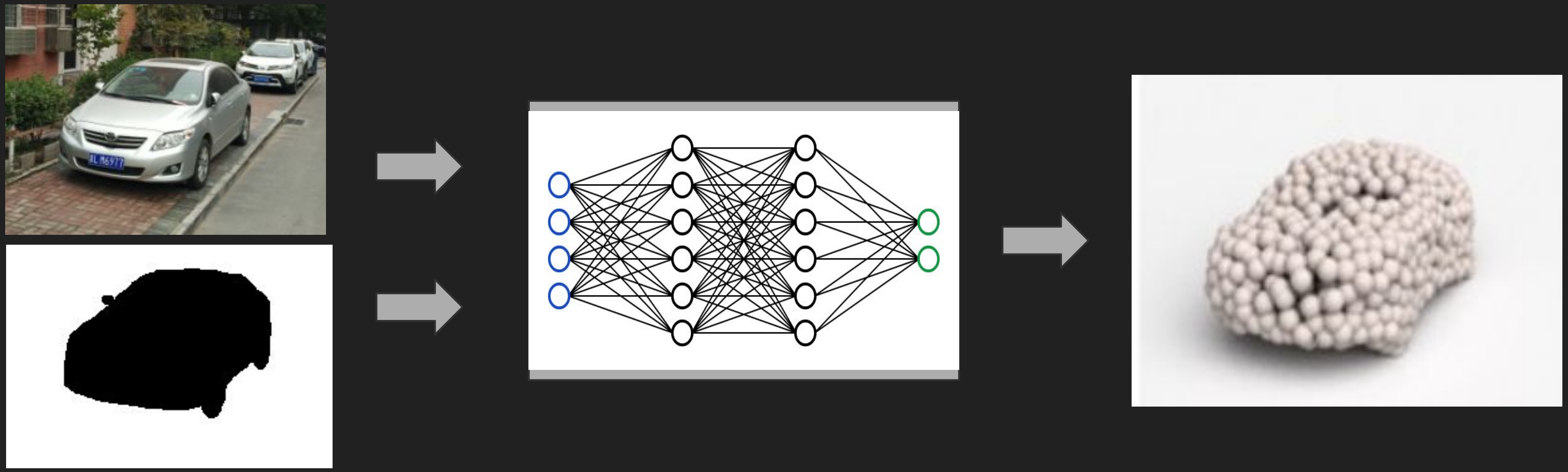
- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- How Loss function helps to overcome the challenge
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

What is the Project All About?

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

How to obtain 3D Point Cloud Coordinates from a Single View!!!

Simply by training a Deep Neural Network to do so!!!!

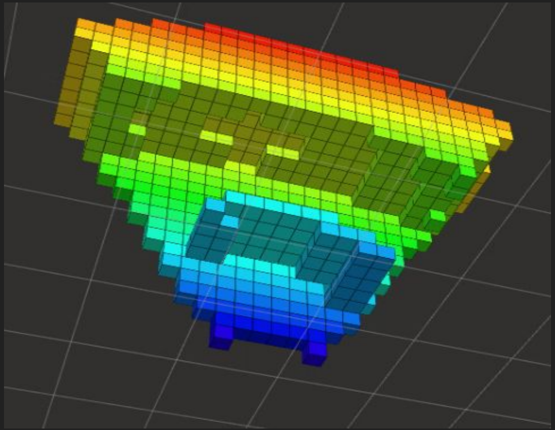


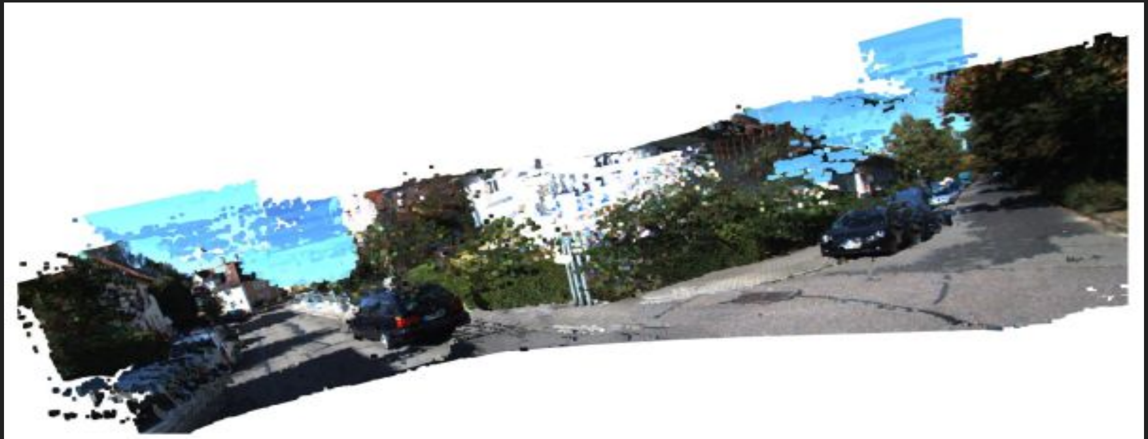
Presentation Overview

- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- How Loss function helps to overcome the challenge
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

Why Point Clouds instead of Volumetric Representation or Meshes??

- Volumetric Grids and Meshes suffered invariance during 3D geometric transformations.
- In Point cloud representation associations between the components is not required to be updated during transformations hence it is not stored.





Presentation Overview

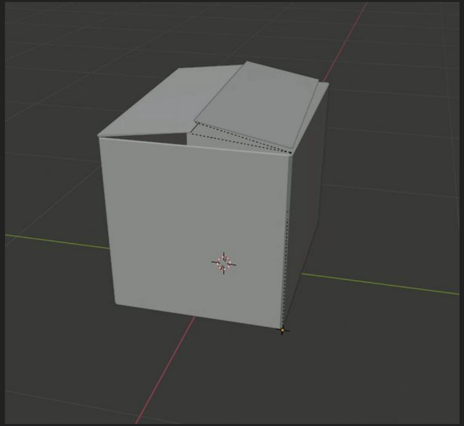
- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- How Loss function helps to overcome the challenge
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

Challenges for Point Cloud Generation using Neural Networks

- Problem is under-determined
- Hence,no fixed annotation, Multiple outputs valid for same input
- Conclusion:

Network Architecture | Loss Function | Learning Paradigm

has to be chosen is such a way that ground truth ambiguity can be resolved

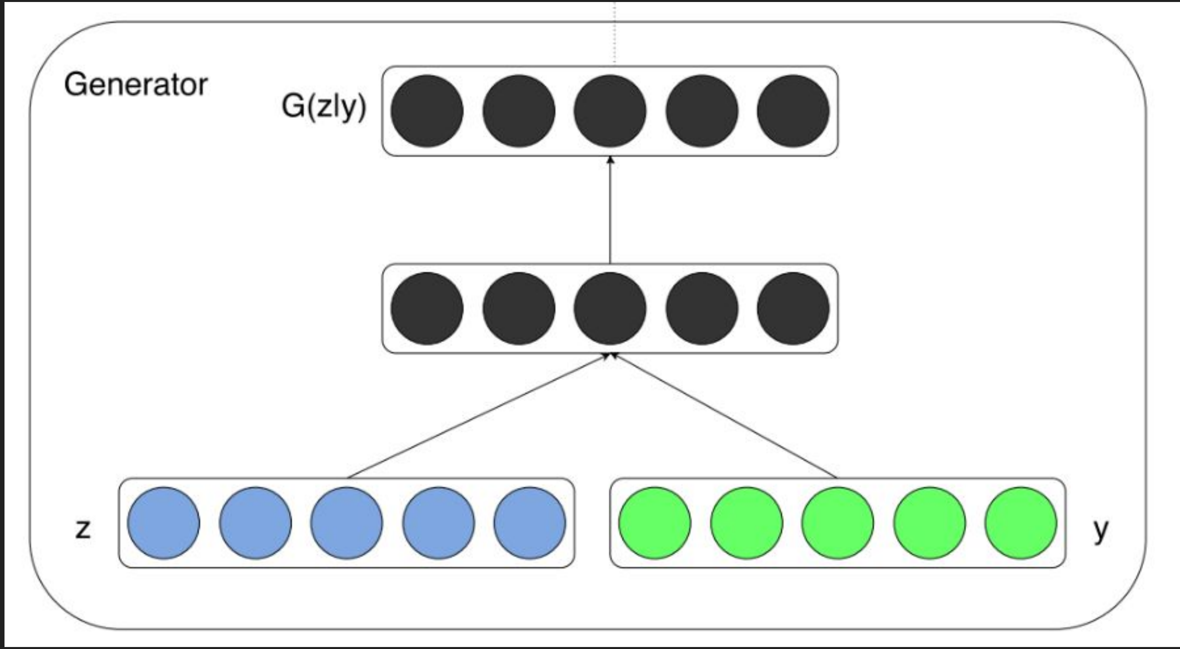


Presentation Overview

- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- <u>How Network Architecture helps to overcome the challenge</u>
- How Loss function helps to overcome the challenge
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

Choice of Network??

Conditional Generative Architecture



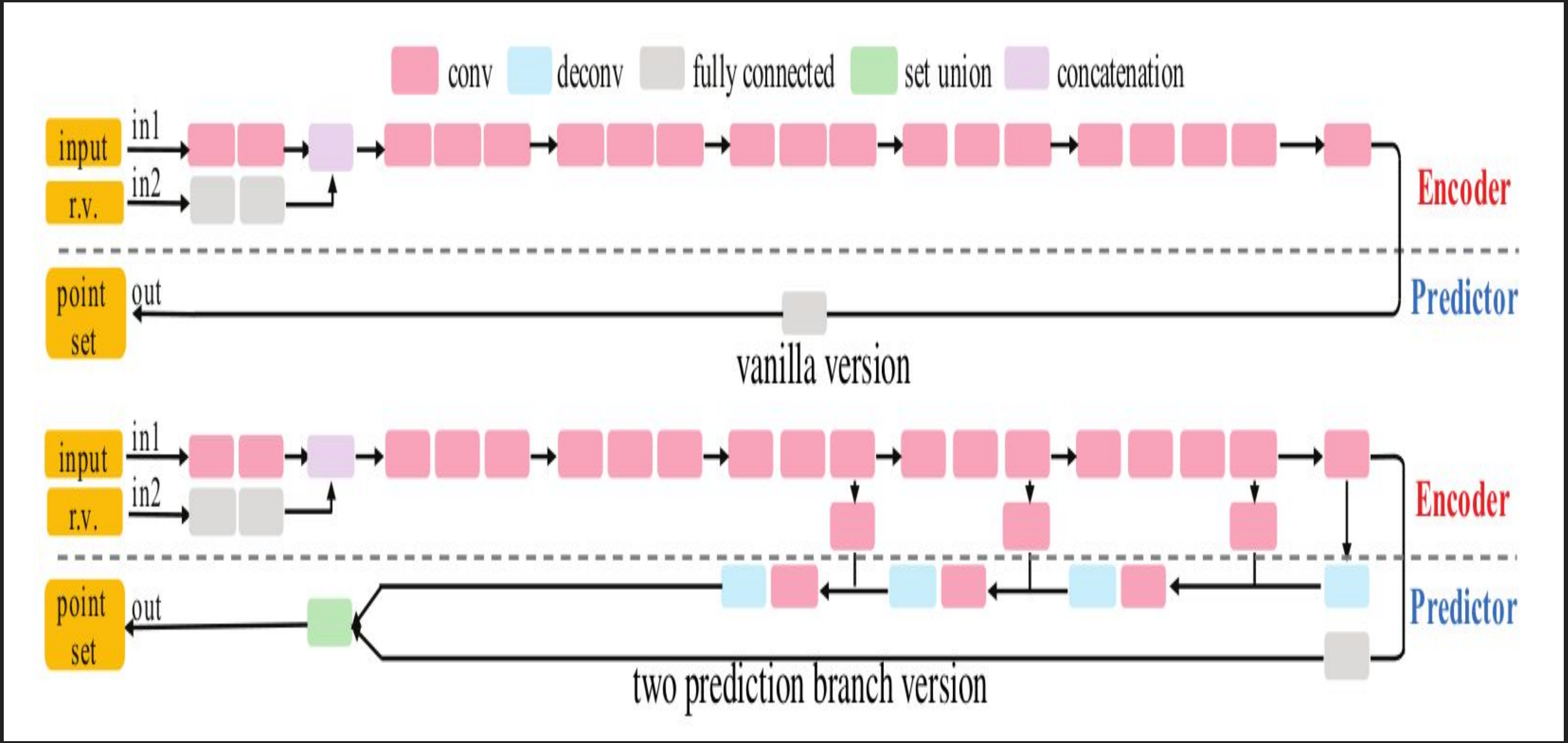
Formulation of Conditional Generative Architecture

- **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



Advantages of 2 Branch Prediction Version over Vanilla Version

Two Branch Version	Vanila Version
Gives Output in 1024 points	Gives Output in 256 points
Encoder and Predictor branches are connected via multiple Skip Connections	Encoder and Predictor branches are not connected with any Skip Connections
Predictor Branch has both Convolutional Layer and Fully Connected Layer	Predictor Branch has only Fully Connected Layer with no skip connections from the Encoder
Enjoys high flexibility in capturing Complicated Structures and exploits Geometric Continuity	Can exploit Complicated Structures but lacks in providing Geometrical Continuity

Presentation Overview

- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- <u>How Loss function helps to overcome the challenge</u>
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

Loss Function

Role of Loss Function:

$$L(\{S_i^{pred}\}, \{S_i^{gt}\}) = \sum d(S_i^{pred}, S_i^{gt}),$$

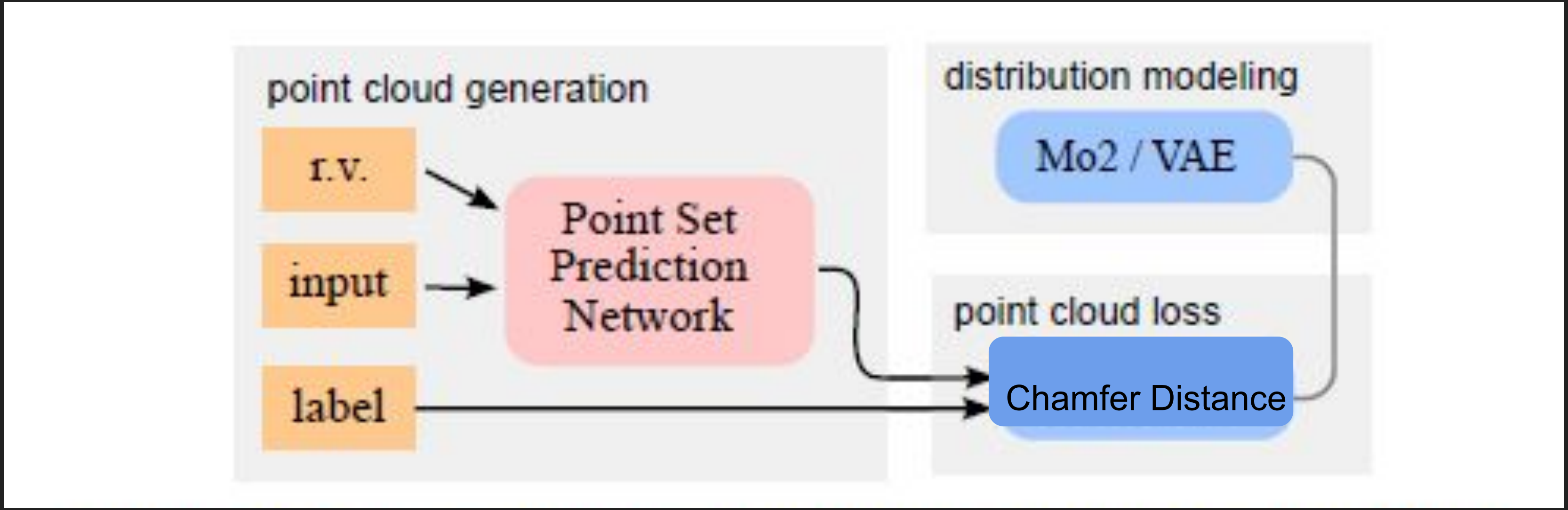
Chamfer Distance:

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

Presentation Overview

- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- How Loss function helps to overcome the challenge
- <u>How Learning Paradigm helps to overcome the challenge</u>
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

Learning Paradigm



Need of Additional Loss Function

- The ambiguity of groundtruth shape may significantly affect the trained predictor, as the loss function induces our model to predict the mean of the possible shapes.
- Mean of the shapes depends more on the Loss Function than on the input, which is highly undesirable.
- Additional Loss Function: MoN Loss for choosing best out of multiple choice

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

Random Vector

- What is the use of Random Vector?

Random Vector perturbs the input in multiple possible ways still preserving the original Image distribution leading the network to predict multiple mean outputs hence exploring the ambiguity space of the 3D shape and choosing the one closest to the Ground Truth.

- How many random vectors are being used and where they are coming from?

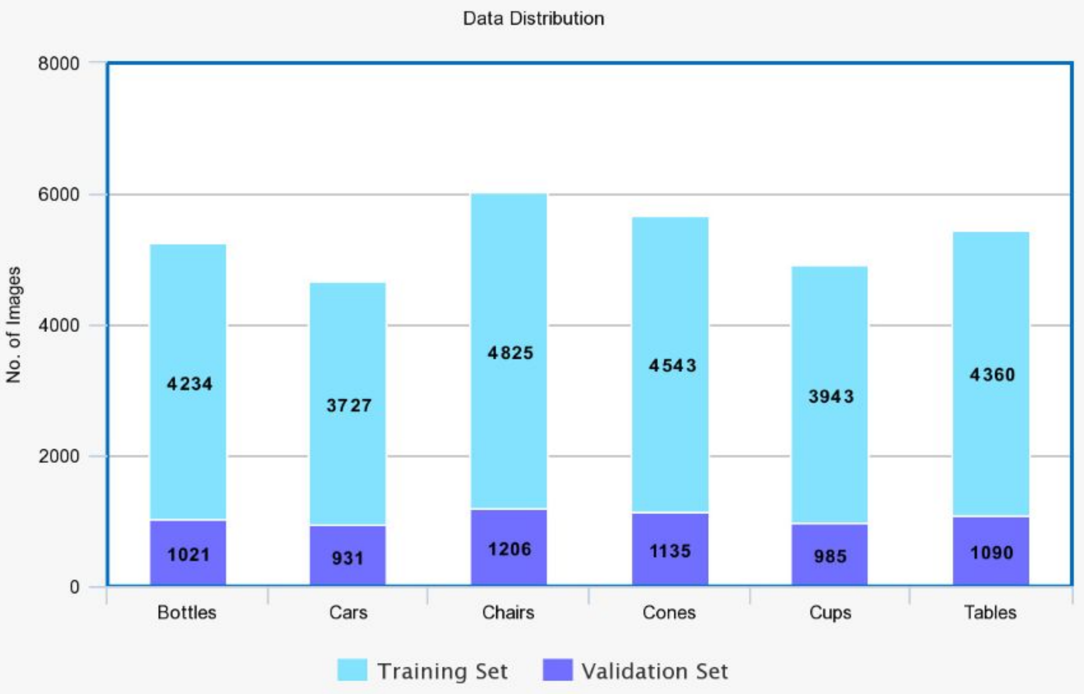
Two random vectors are being used.The depth map of the image is being provided as the random vector along with the input to do the perturbation.

Presentation Overview

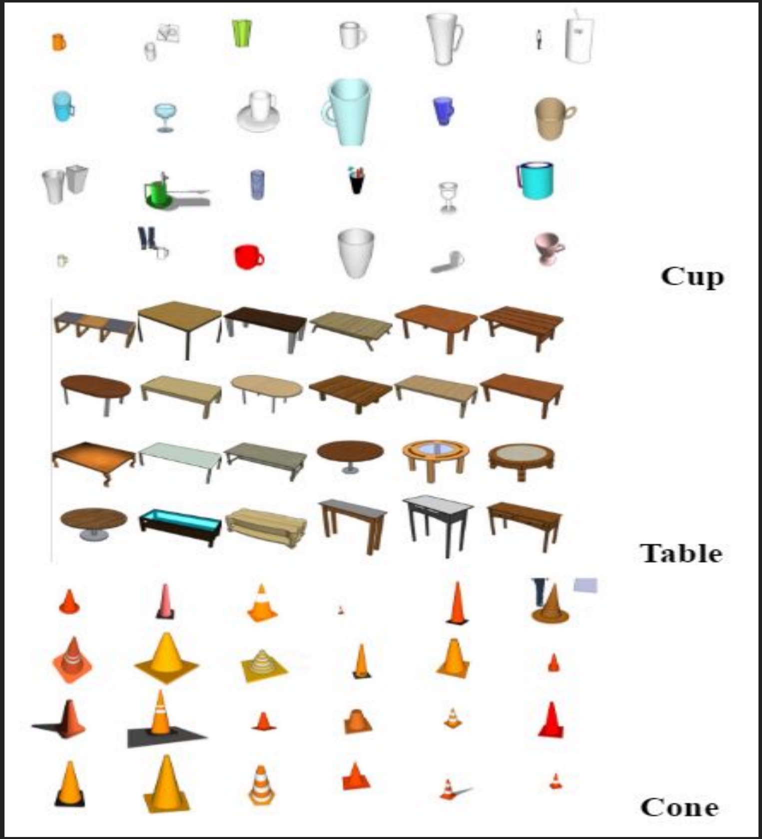
- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- How Loss function helps to overcome the challenge
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

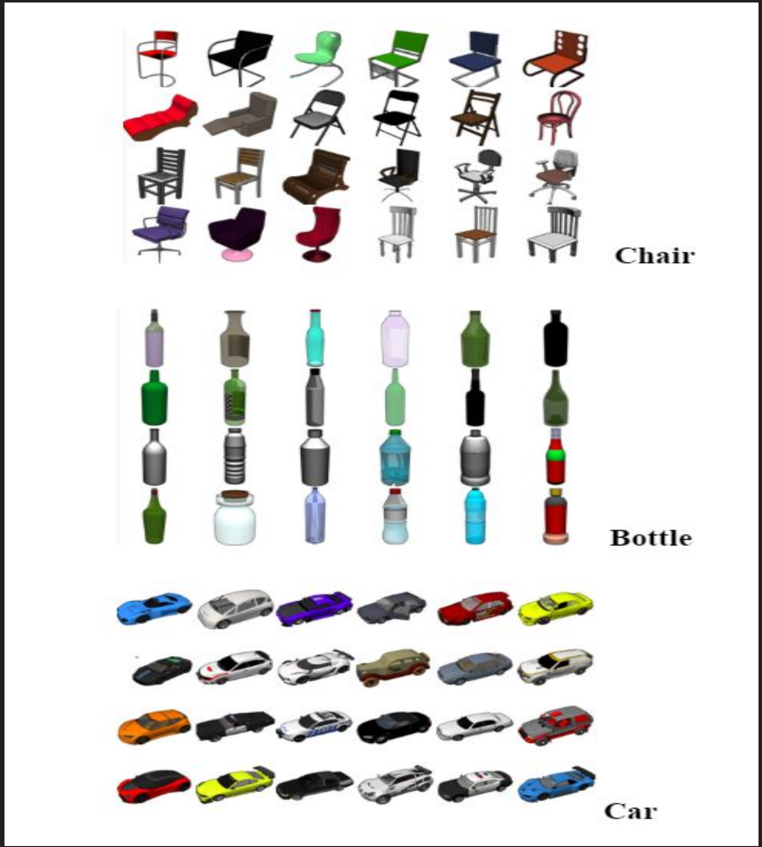
Data Distribution

Data has been divided in 6 classes and 20% data of each class is in validation set

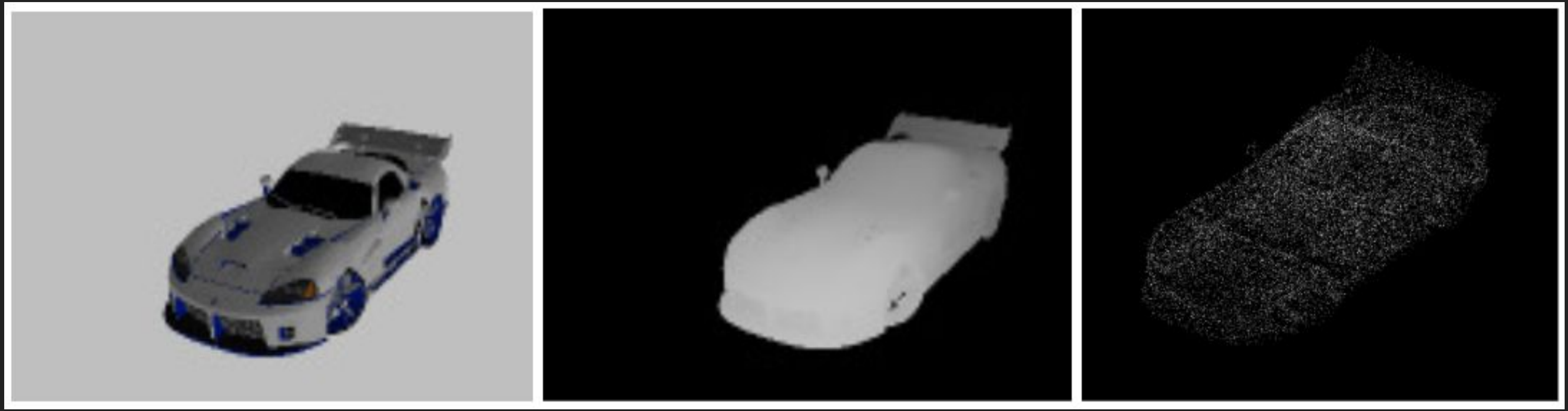


Dataset and its Preparation





Input Sample,Random Vector and Ground Truth



Presentation Overview

- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- How Loss function helps to overcome the challenge
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

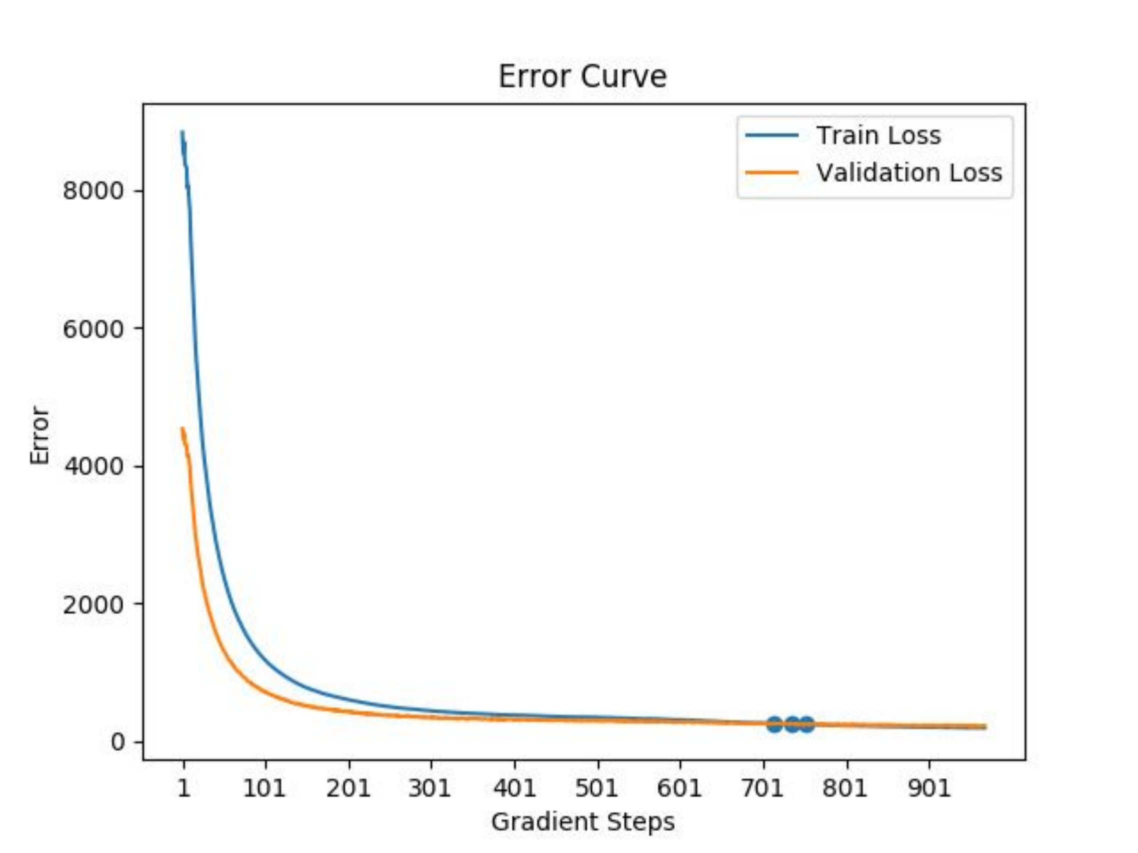
Implementation Details

- Our network works on input images of 192x256.
- The deconv branch produces 768 points, which correspond to a 32x24 three channel image. The fully connected branch produces 256 points. So the final image is 32x32x3,where each location has 3 values depicting x,y,z.
- The training program is implemented in TensorFlow.
  1000 gradient steps are taken, each computed from a minibatch of 32.
- Adam is used as the optimizer.
- All activation functions are ReLU.
- Regularization: L2 Regularizer
- Number of points in Ground Truth: 16384
- Loss Value: 191.39 for last gradient step possible for Vanilla Version
- Loss Value: 175.38 for last gradient step possible for Two Branch Version

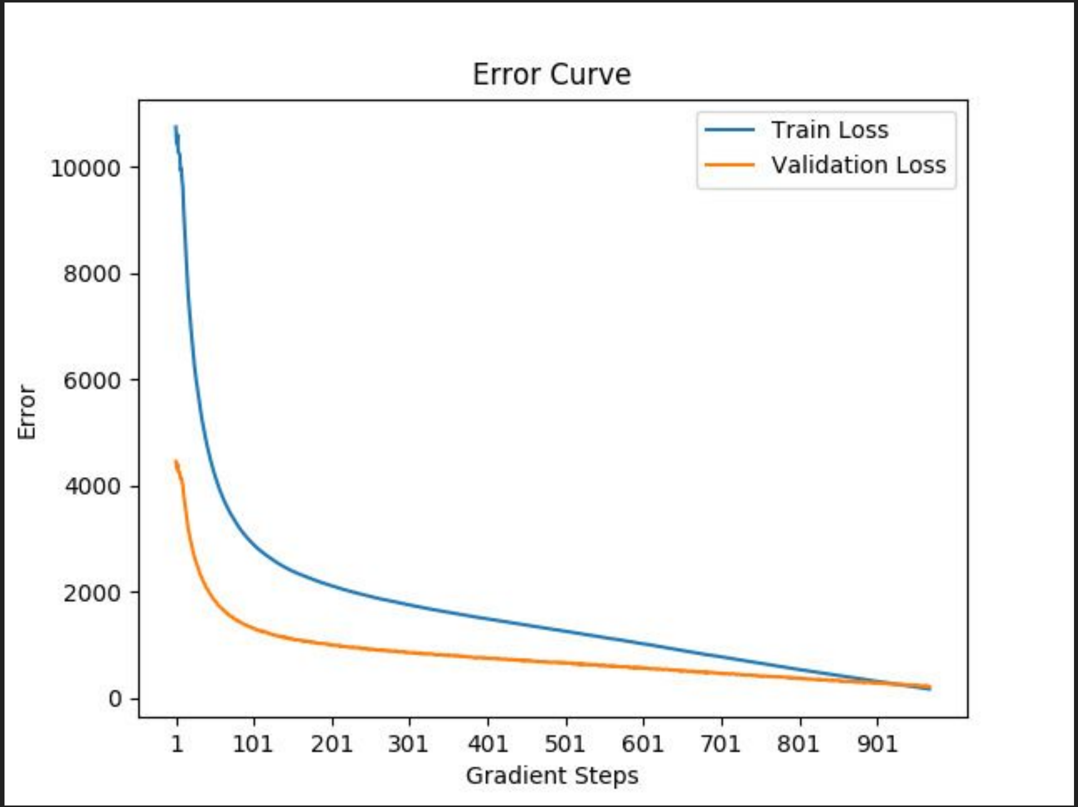
Presentation Overview

- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- How Loss function helps to overcome the challenge
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

Loss Curve for Vanilla Version

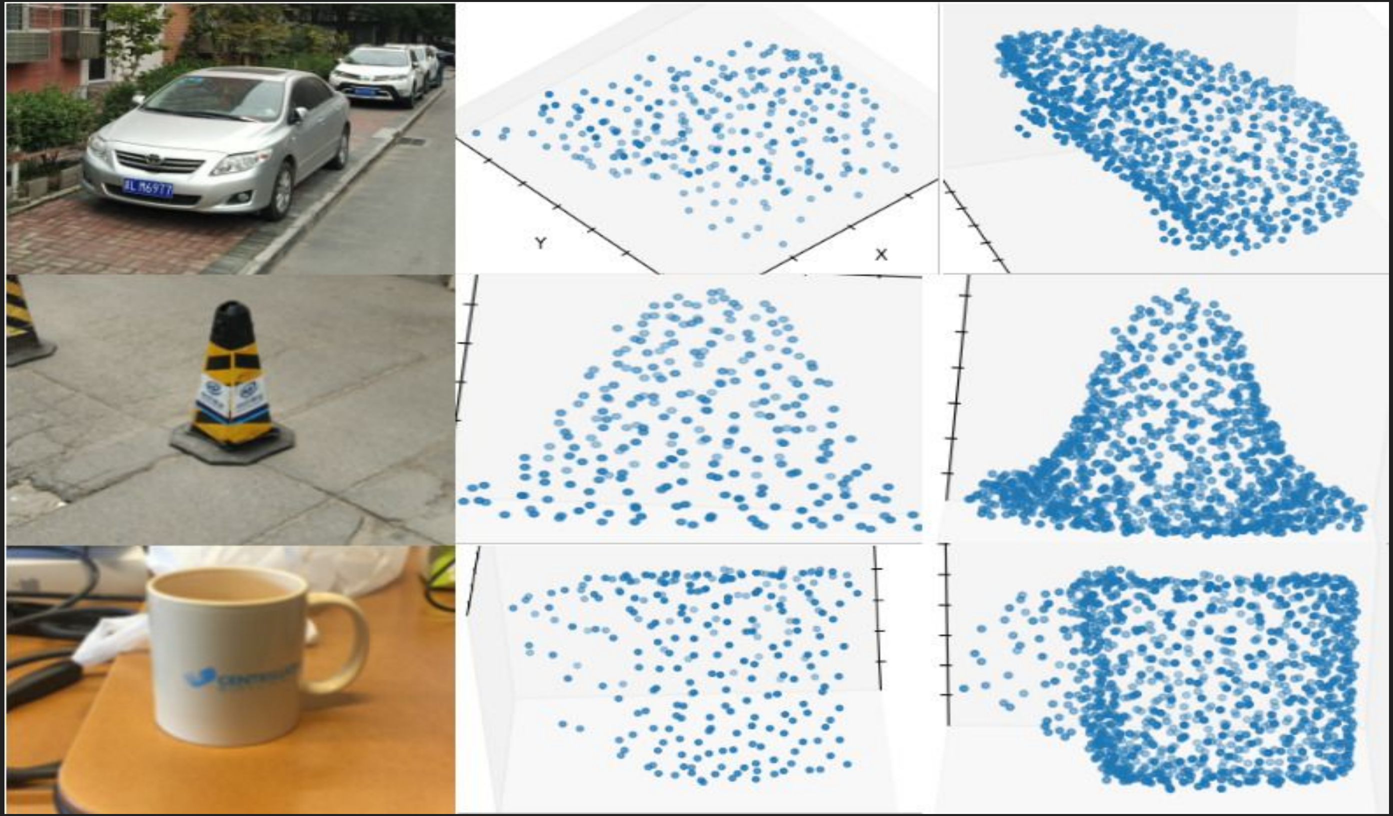


Loss Curve for Two Branch Prediction Version



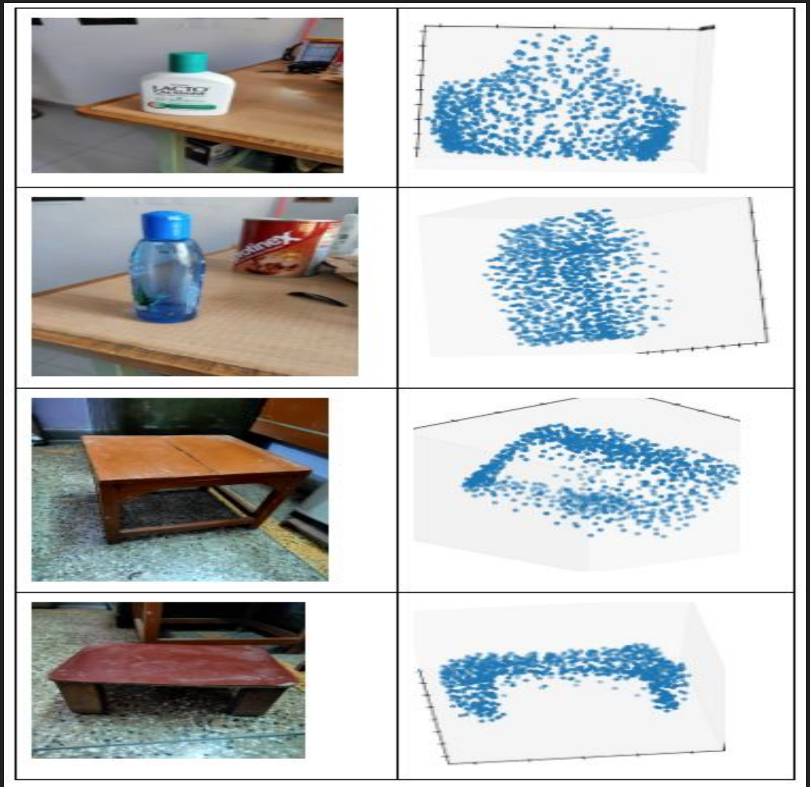
Results on Their Data

Input Output_Vanilla Output_Two_Branch



Results on Real-World Data

Input | Output_Two_Branch



Presentation Overview

- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- How Loss function helps to overcome the challenge
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

Conclusion

- Implementing the project gave us a good exposure to Deep Networks.
- It helped in Introduction to the Research Area of 3D Reconstruction from monocular image using Deep Neural Networks.
- It helped in understanding the steps involved in finding a solution from scratch.
- However,Still much scope remains in terms of improvement of results and understanding the concepts deeper.

Presentation Overview

- Problem Statement addressed by the Presentation
- Solution proposed and its Merits
- Challenges in the path
- How Network Architecture helps to overcome the challenge
- How Loss function helps to overcome the challenge
- How Learning Paradigm helps to overcome the challenge
- Dataset and its Preparation
- Our Implementation Details
- Results
- Conclusion
- References

Important Links and References

- Link to Paper
- Link to Detailed Report
- Link to our Implementation on GitHub

- M. Mirza and S. Osindero. Conditional generative adversarial nets.arXiv preprint arXiv:1411.1784,2014.
- ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], 2015
- A unified approach for single and multiview 3d object reconstruction. arXiv preprint arXiv:1604.00449, 2016.
- D. F. Fouhey, A. Gupta, and M. Hebert. Data-driven 3D primitives for single image understanding. In ICCV, 2013.