A Hybrid Ensemble Approach For 3D Object Reconstruction from Multi-View Monocular RGB images

Ajith K B
Intern CTO
Wipro Turbo

Guided by : Jijith Nadumuri Ravi
Principal Consultant
Wipro CTO Office

July 4, 2019

Introduction

- Computer vision and augmented reality industry focus on reconstructing 3D view of an object for the purpose of improvising the visual effect.
- The 3D reconstruction of objects is a generally scientific problem and core technology of a wide variety of fields, robotics, entertainment, AR, HMI and animation.
- Traditional pipelines such SFM and V-SLAM are fails if the multiple viewpoints are separated by large baseline or self occlusions.
- Another method Space carving, assumes that the objects are accurately segmented from the background or that the cameras are calibrated.

Introduction(contd...)

WHY CNN?

- The main advantages of CNN in 3D model reconstruction are
  - Quality, output resolution.
  - Speed, generation time.
  - Simplicity, easy to use.
- For the product level approach CNN are more easier to integrate with servers
- Creation of 3D objects can be a complex and time-consuming task for those who have little to none prior experience in 3D modeling.

Literature review

Title of the paper	Methodology	Merits	Demerits
Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling,29th NIPS 2016, Barcelona	*VAE+GAN used for generation *Loss function-Cross entropy loss kl-divergence and optimizer-ADAM	*Higher output size(64*3) *High accuracy for less class	*Multi-class 3D reconstruction is not good * Less IOU value for multi class
3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction ,ECCV 2016	*Resnet based VAE network is used for extraction and generation *RNN based LSTM used for feature selection * Cross entropy loss and ADAM	*Better IOU Values *Single and multi-view 3D Reconstruction	*Time consuming due ti LSTM network. *Permutation variant

Literature review (contd...)

Title of the paper	Methodology	Merits	Demerits
Attentional Aggregation of Deep Feature Sets for Multi-view 3D Reconstruction ,IEEE 2018	*residual net based VAE + FC based network Architecture. *FC based Attsets- for feature extraction *cross entropy loss *JTSO and ADAM	*Outperforms other methods * Computationally efficient Permutation invariant	*small output voxel size(32*3)
Pix2Vox: Context-aware 3D Reconstruction from Single and Multi-view Images,2019	*VGG16 based VAE for feature extraction. *U-net based architecture for optimization * cross entropy loss and ADAM	*Unet provides better optimized models	*small output voxel size(32*3) *Compared to Resnet, VGG provides less features

Objectives

- The idea is to develop a unique plugin for Unity 3D which helps to create 3d models automatically
- To develop a better accurate 3d models using economically feasible system compatible with Unity 3D.
- Develop an network which can integrate with AttSets and optimizes the network output
- To interface a suitable algorithm for surface reconstruction from the network predicted voxalized shape.
- Develop an improved algorithm for optimization
- To Optimize the network for custom automobile data-set

Methodology

Network Architecture Comparison

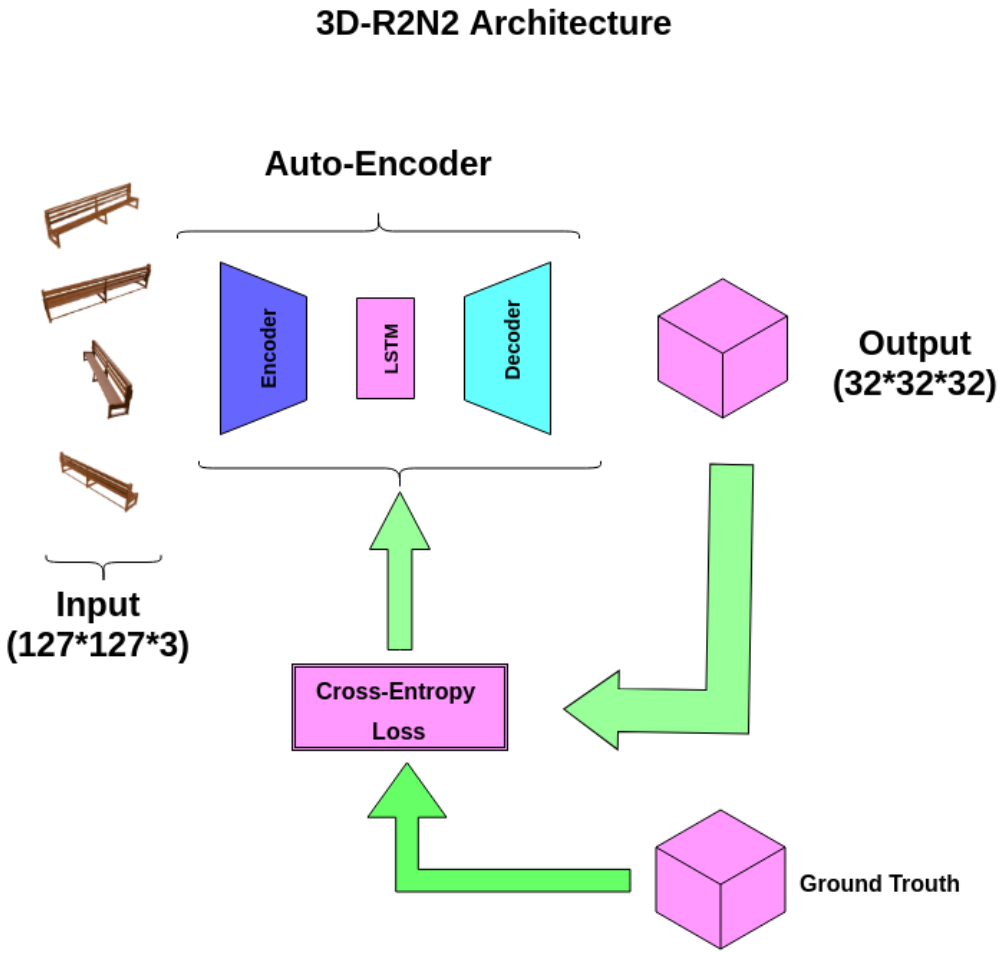

Figure: Block diagram of 3D-R2N2 Architecture

Methodology(Contd...)

Network Architecture Comparison

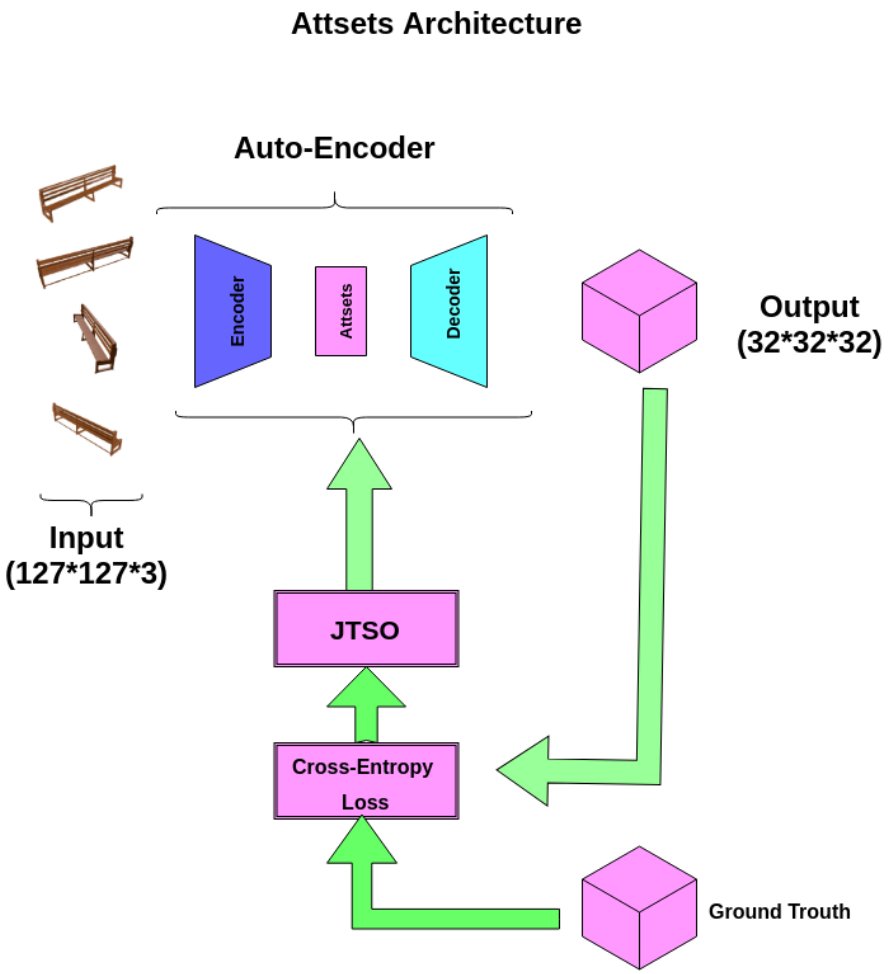

Figure: Block diagram of Attsets Architecture

Methodology(contd...)

Network Architecture Comparison

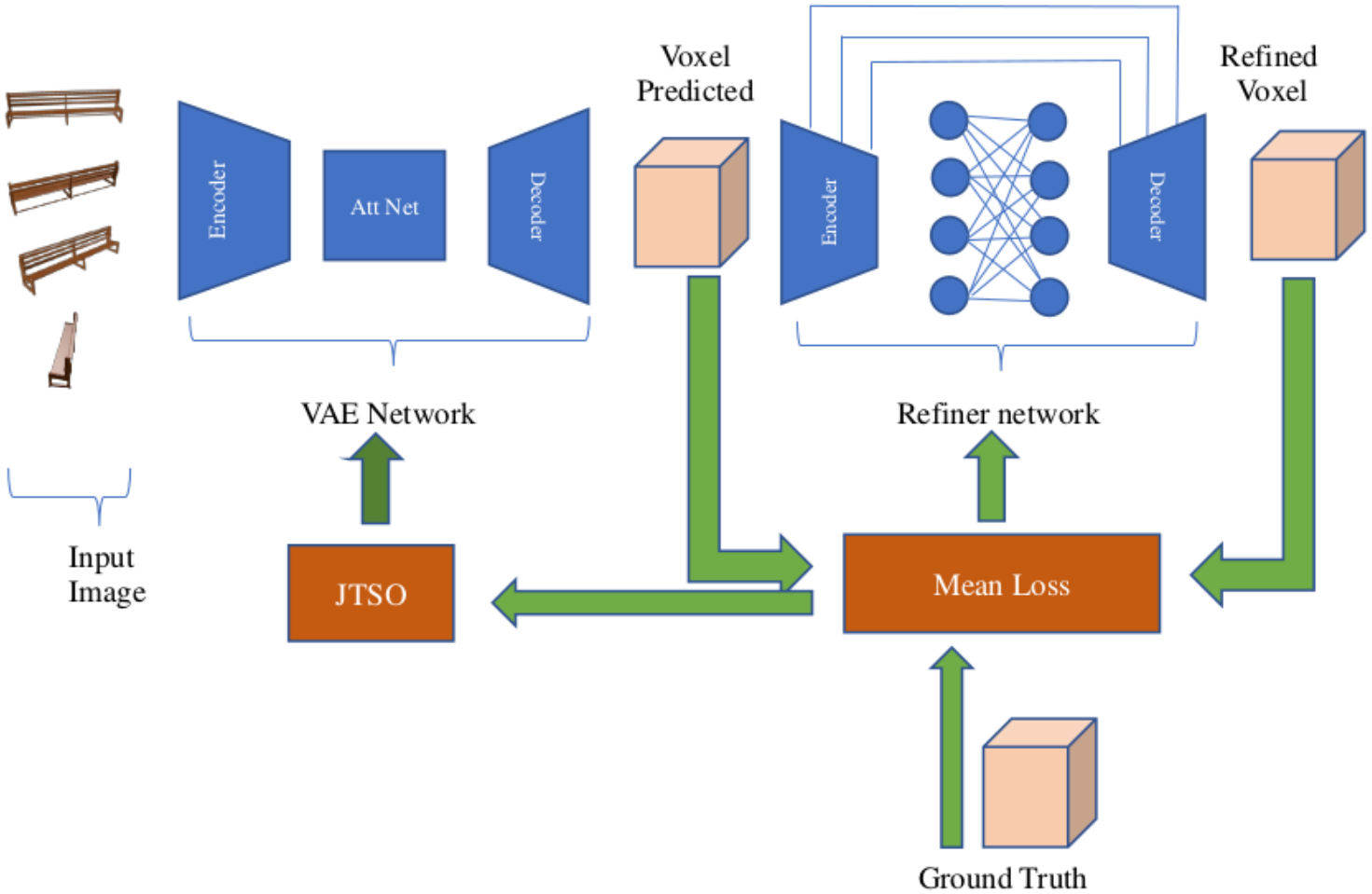

Figure: Block diagram of the New Architecture

Methodology(contd...)

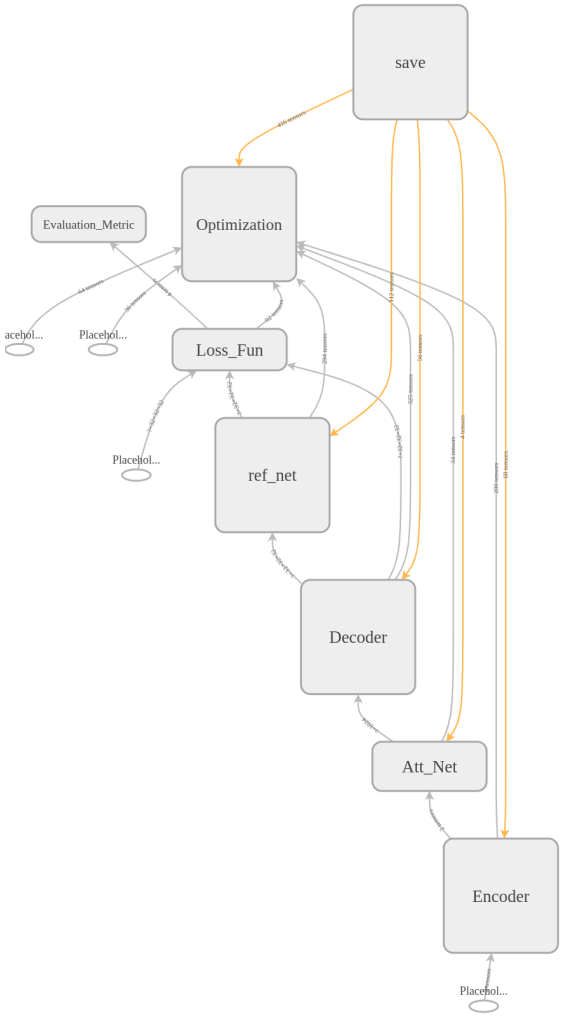

Figure: TensorBoard Graph Visualization of the proposed system

Architecture of Network

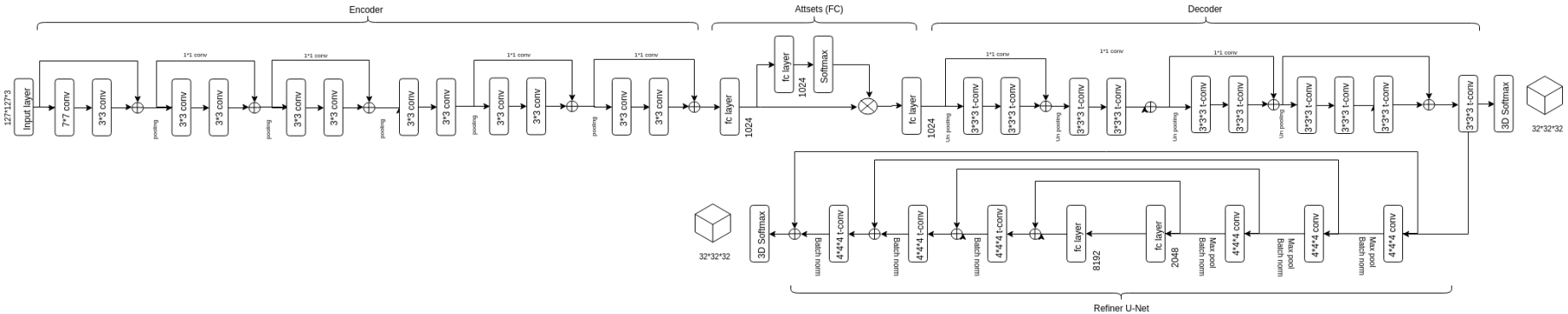

Figure: Layer Architecture of the network

Hardware and Software used for training

- Nvidia Quadro P4000, 32GB RAM, 16GB GPU Ubuntu 16.04 LTS.
- Tensorflow 1.11.0 and tensorflow-gpu 1.11.0
- Cuda (version 9.0) is a parallel computing platform and API developed by Nvidia for its GPUs
- Cudnn 7.4.2 library which is used to accelerate performance of Dense Neural Network.

Selected data-set

Shapenet Data-set

- Selected subset of Shapenet data-set consists of 5 categories of 16,896 common objects with synthesized RGB images
- For each 3D object, 24 images are rendered from different viewing angles circling around.
- For better network optimization dataset divided in to 4 and each part contains train and test set- Cross validation technique.
-

Progress of Work

- Attsets network is integrated with **U-Net** based optimizer architecture.
- New architecture trained with Shapenet Data-set using Tensorflow
- New improved JTSO algorithm for optimization is implemented
- To improve optimization, a custom feedback line is used to generate mean loss to optimize VAE network
- Integrated surface reconstruction feature which is based on **Marching cubes** algorithm to the network
- Generated 3D models are converted to .obj and .dae (collada) format which can be directly open in Unity

Results and Discussion

Loss functions

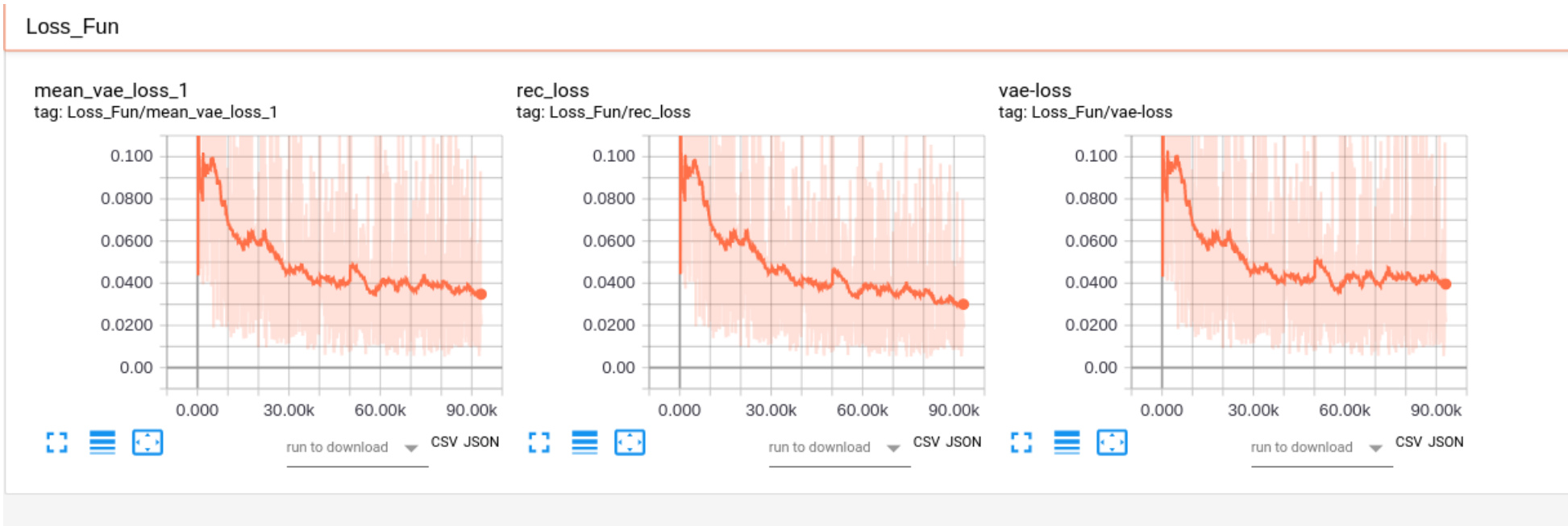

Figure: Multi-view 3D Reconstruction loss

Results and Discussion(contd...)

Evaluation Matrix

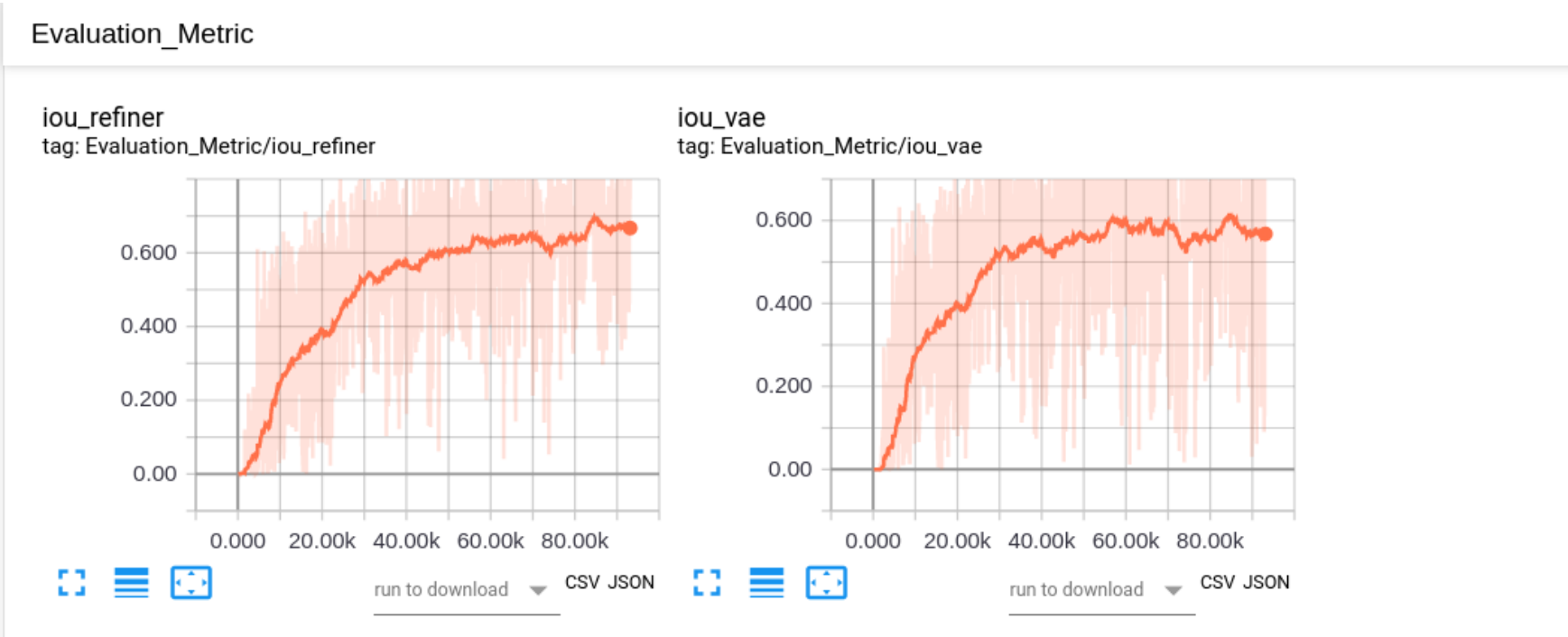

Figure: Accuracy graph - Mean IOU

Results and Discussion(contd...)

loss Distribution

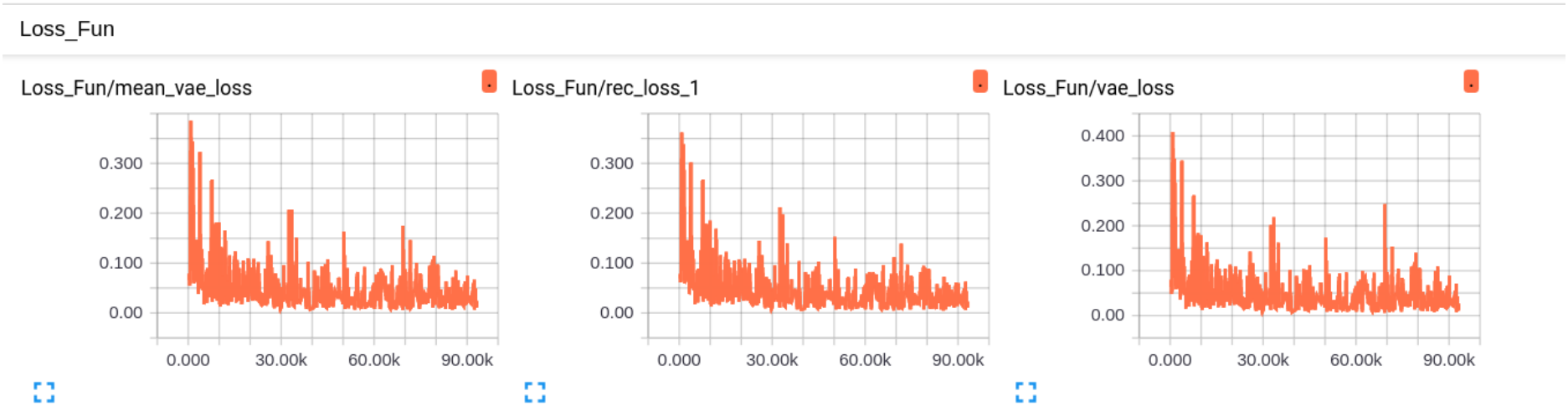

Figure: Multi-view 3D Reconstruction loss

Results and Discussion(contd...)

Histogram

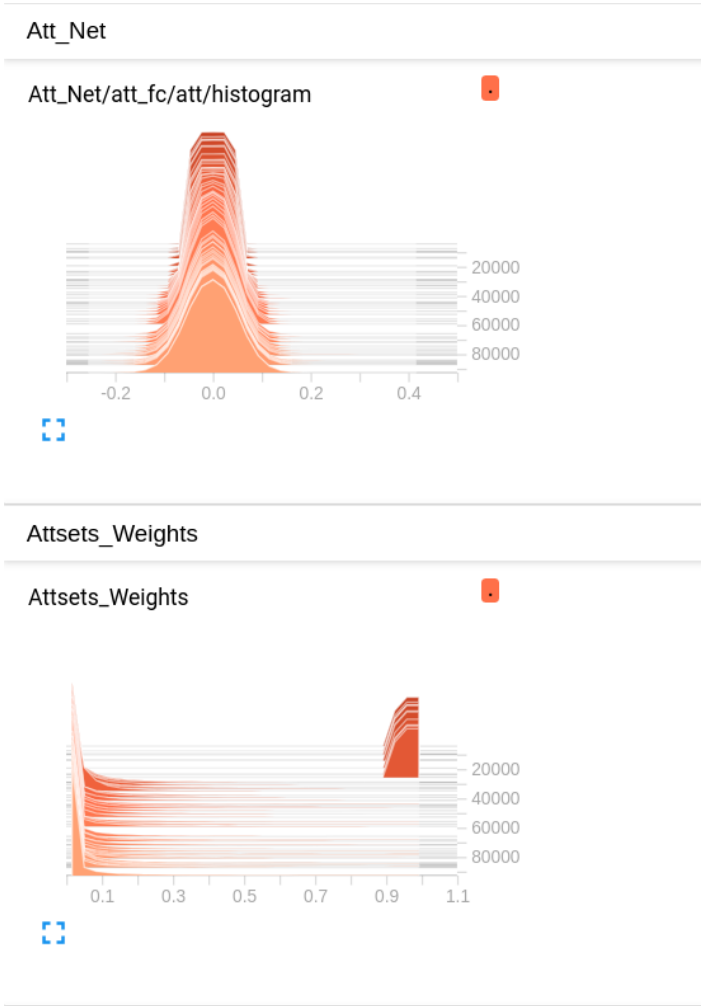

Figure: Histogram of Attsets Fully connected leyers

Results and Discussion(contd...)

Scalar-weights

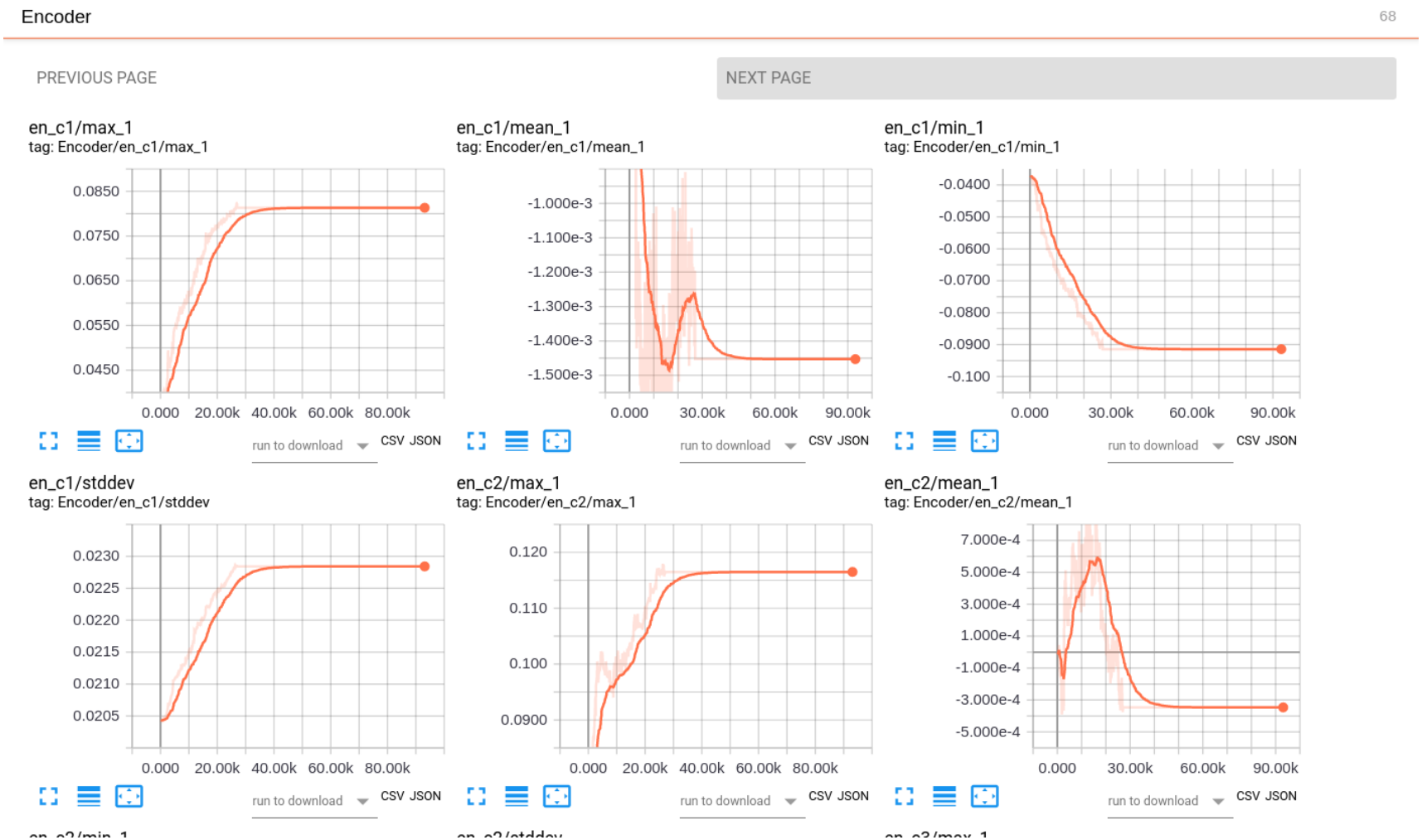

Figure: Scalar graph of Attsets Encoder weight updation

Results and Discussion(contd...)

Scalar-weights

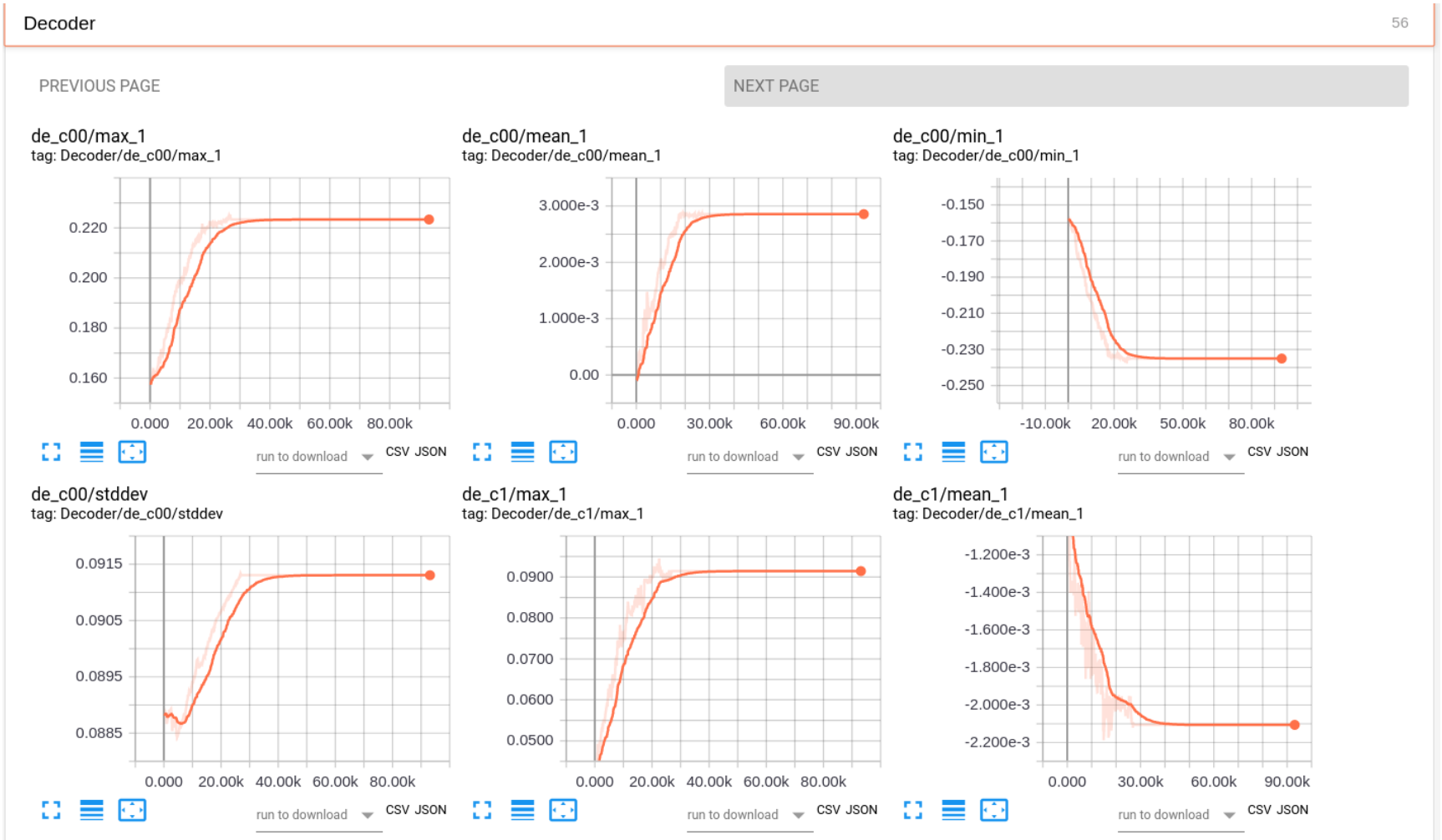

Figure: Scalar graph of Attsets Decoder weight updation

Results and Discussion(contd...)

Unity Screen-shot

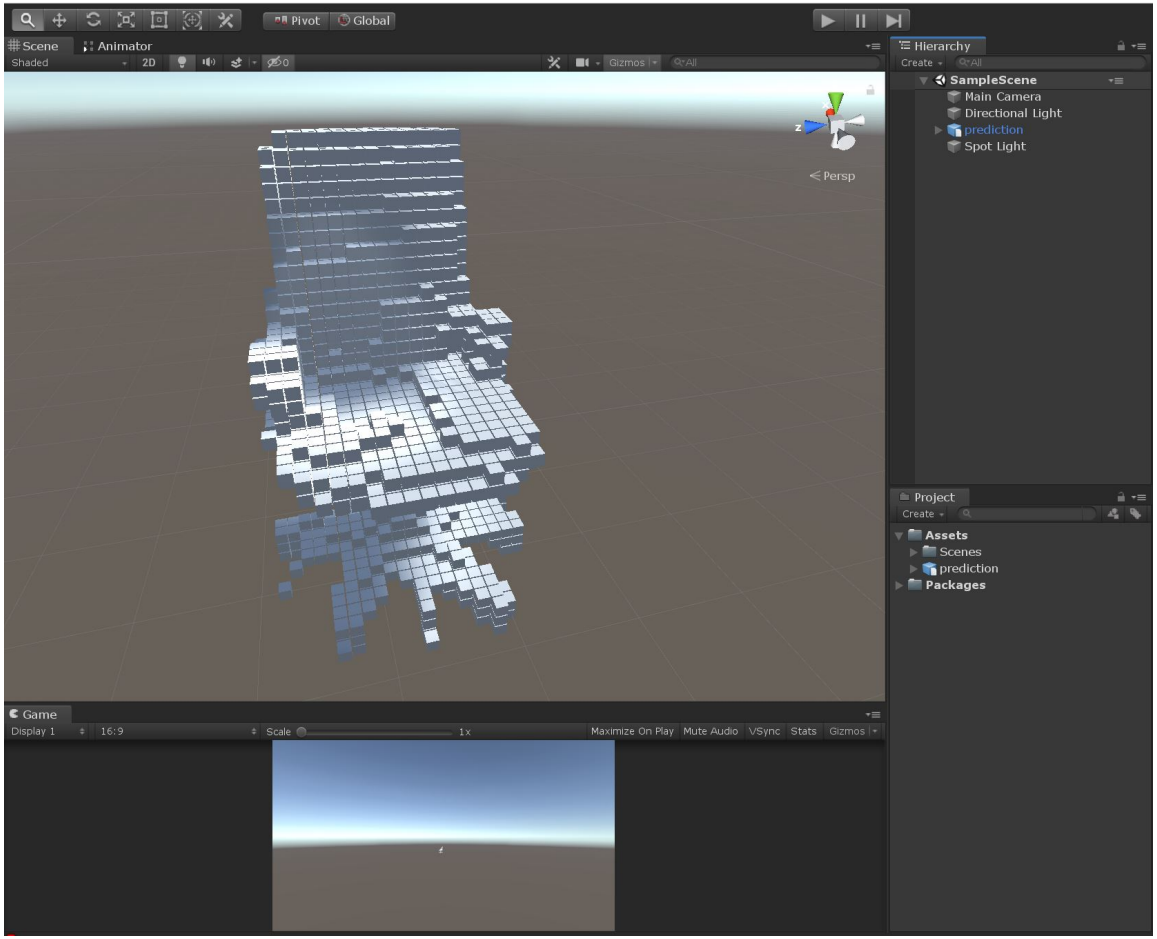

Figure: Screen-shot of Generated voxel model rendered in Unity

Test Results

Generated 3D Models for class-desk after 45K iterations

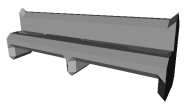

Figure: Ground Truth

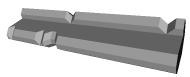

Figure: Optimizer Output

Figure: VAE Output

Test Results(contd...)

Generated 3D Models for class-desk after 90K iterations

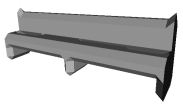

Figure: Ground Truth

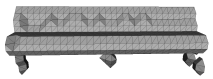

Figure: Optimizer Output

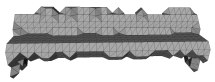

Figure: VAE Output

Test Results(contd...)

Generated 3D Models for class-desk after 90K iterations

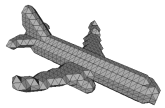

Figure: Ground Truth

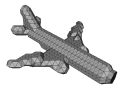

Figure: Optimizer Output

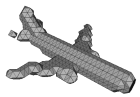

Figure: VAE Output

Test Results(contd...)

IOU Readings for class-desk

Figure: IOU readings after 90K iterations

Test Results(contd...)

Generated 3D Models for class-boat after 90K iterations

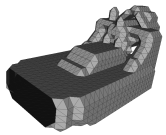

Figure: Ground Truth

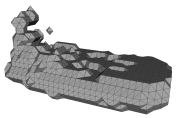

Figure: Optimizer Output

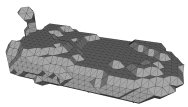

Figure: VAE Output

Test Results(contd...)

IOU Readings for Class-Boat

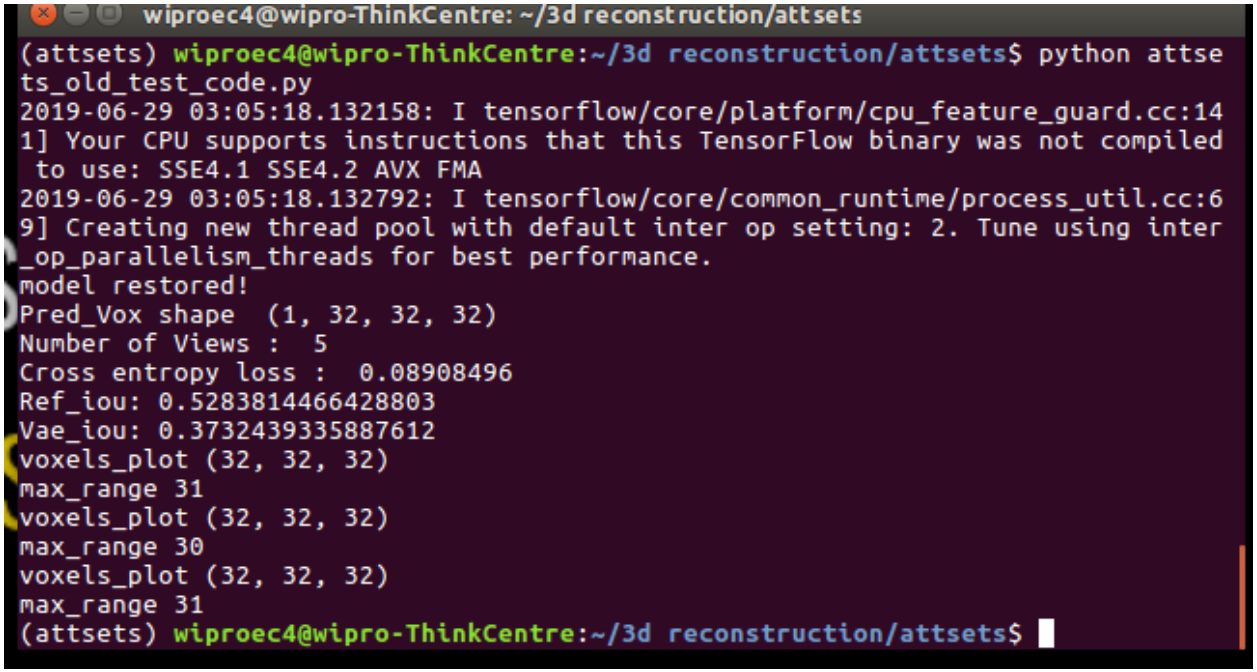

Figure: IOU readings after 90K iterations

Summary

- 30 % of the training of new architecture is finished and it shows encouraging results
- surface reconstruction from voxels is implemented using marching cubes algorithm.
- Implemented one feedback network to improve network optimizaion
- To extract more features need to try with other feature extraction methods like Resnet V2,Inception Net etc.

Difficulties in the Project

- Network is not trained for automobile parts due to the lack of availability of enough data-set
- Need powerful GPU to train - New network takes 3hr for 1 epoch in Quadro series GPU
- But same network takes only 110 minutes in our Alienware Machine which loaded with Nvidia 1080GPU which is twice faster than our desktop GPU.

References

- Bo Yang , Sen Wang , Andrew Markham , Niki Trigoni *Attentional Aggregation of Deep Feature Sets for Multi-view 3D Reconstruction*. , IEEE 2018
- Christopher B. Choy, Danfei Xu, JunYoung Gwak Kevin Chen, Silvio Savarese *3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction*. ECCV 2016
- jiajun Wu, Chengkai Zhang, Tianfan Xue, William T Freeman, Joshua B Tenenbaum *Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling*,29th NIPS 2016, Barcelona
- Haozhe Xie, Hongxun Yao, Xiaoshuai Sun, Shangchen Zhou, Shengping Zhang, Xiaojun Tong *Pix2Vox: Context-aware 3D Reconstruction from Single and Multi-view Images* jan - 2019

References

- Olaf Ronneberger, Philipp Fischer, and Thomas Brox *U-Net: Convolutional Networks for Biomedical Image Segmentation*. , IEEE 2015
- Li Yi , Lin Shao , Manolis Savva , Haibin Huang , Yang Zhou , Qirui Wang , Benjamin Graham , Martin Engelcke , Roman Klokov , Victor Lempitsky , Yuan Gan *Large-Scale 3D Shape Reconstruction and Segmentation from ShapeNet Core55*. ECCV 2017