Indoor 3D Reconstruction from a Single Image

Chirag Garg

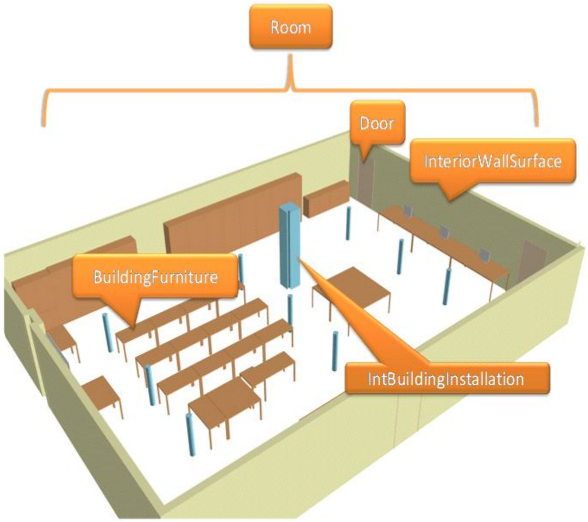
Mentor #1: Liangliang Nan
Mentor #2: Jan van Gemert
Mentor #3: Seyran Khademi

TU Delft

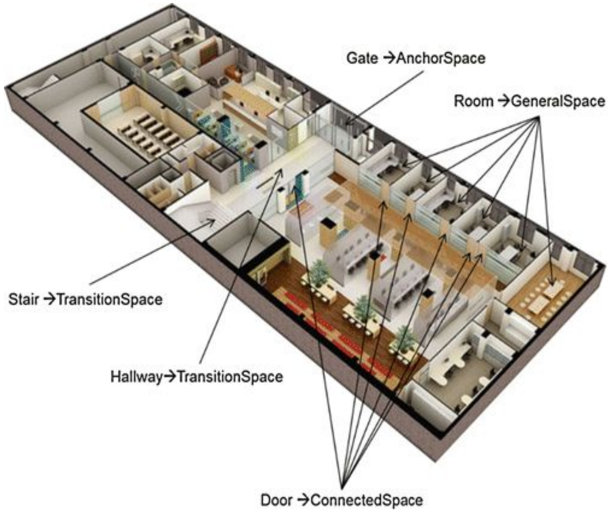
- Motivation
- Related Work
- Research Questions
- Methodology
- Results
- Conclusion

Motivation

- **Applications of 3D indoor reconstruction**



Infrastructure Management

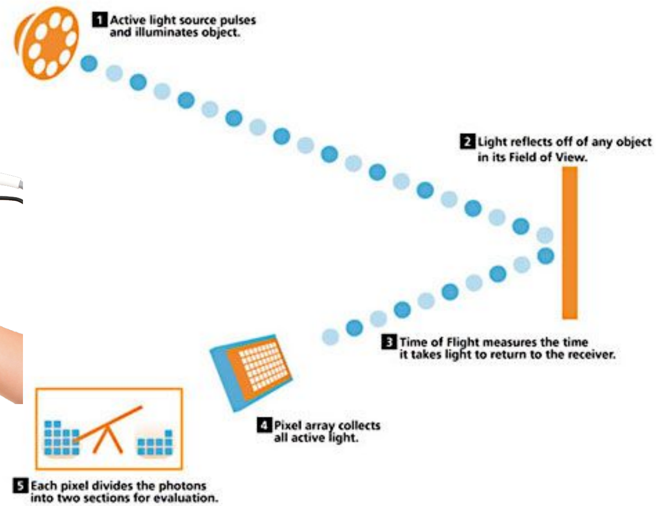


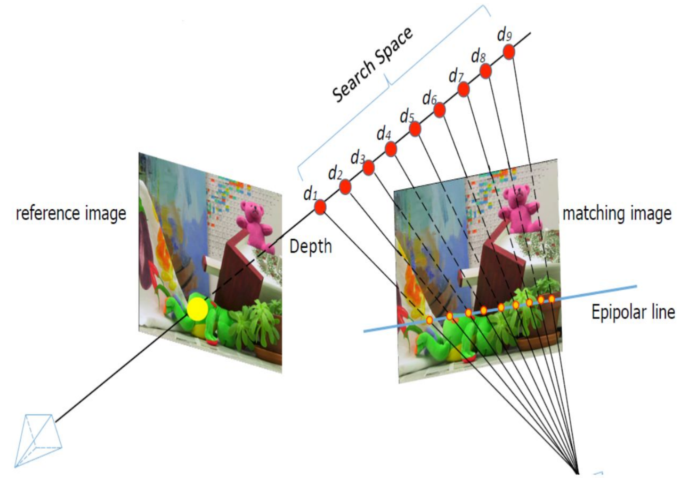
Indoor Navigation, Simulation, Virtual Reality

[Donaubauer et al., 2010], [Zlatanova and Isikdag, 2017]

TU Delft

Conventional Approaches for 3D Reconstruction

- 




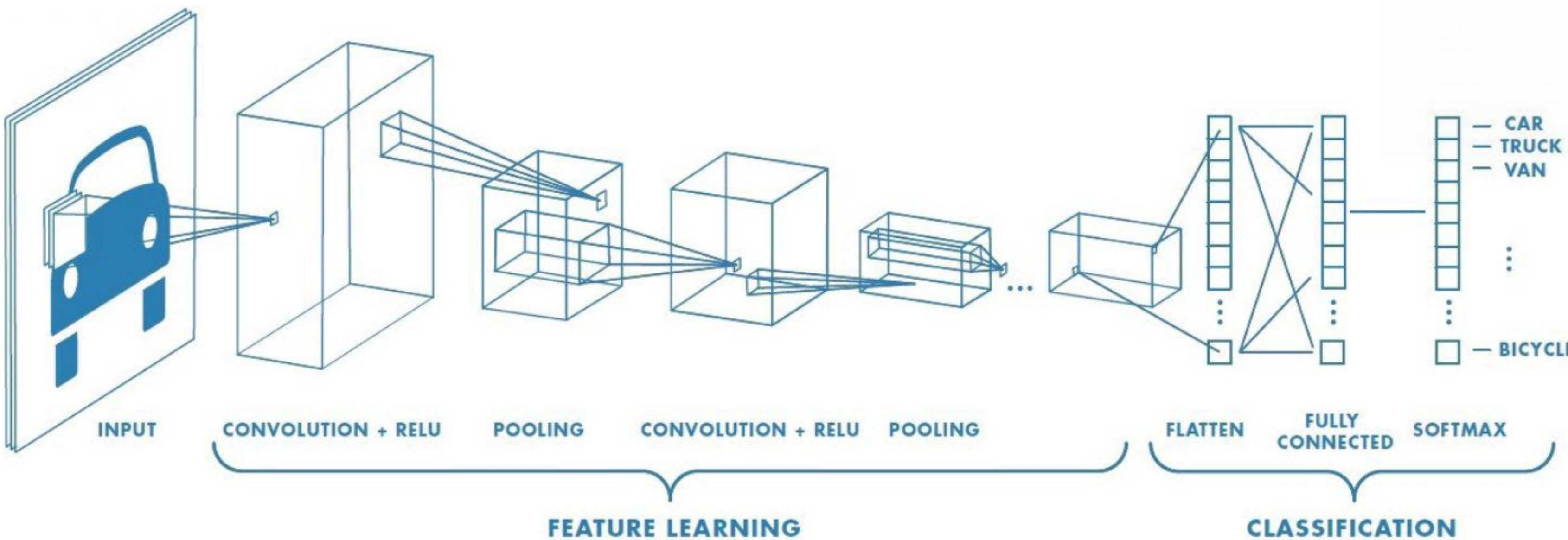
  - Using sensors (laser scanner, IMU, GPU devices) - requires manpower & equipments
  - Using multiple images (SFM/MVS) - needs considerable processing

- **Motivation :**
  - Minimize user effort for data acquisition
    - Use Single Image for understanding an indoor scene
    - Explore possibilities to extract 3D information

TU Delft

Deep learning Approach

- **Convolutional Neural Networks (CNN)**
  - Feature Learning using deep neural networks
  - Task Specific network



[Saha, 2018]

Object Level 3D Reconstruction

- Deformation based method for mesh of single objects

Input Image **GEOMetrics** **Pixel2Mesh**

[Smith et al., 2019]

- Mesh R-CNN : Multiple objects using real world dataset

Input Image

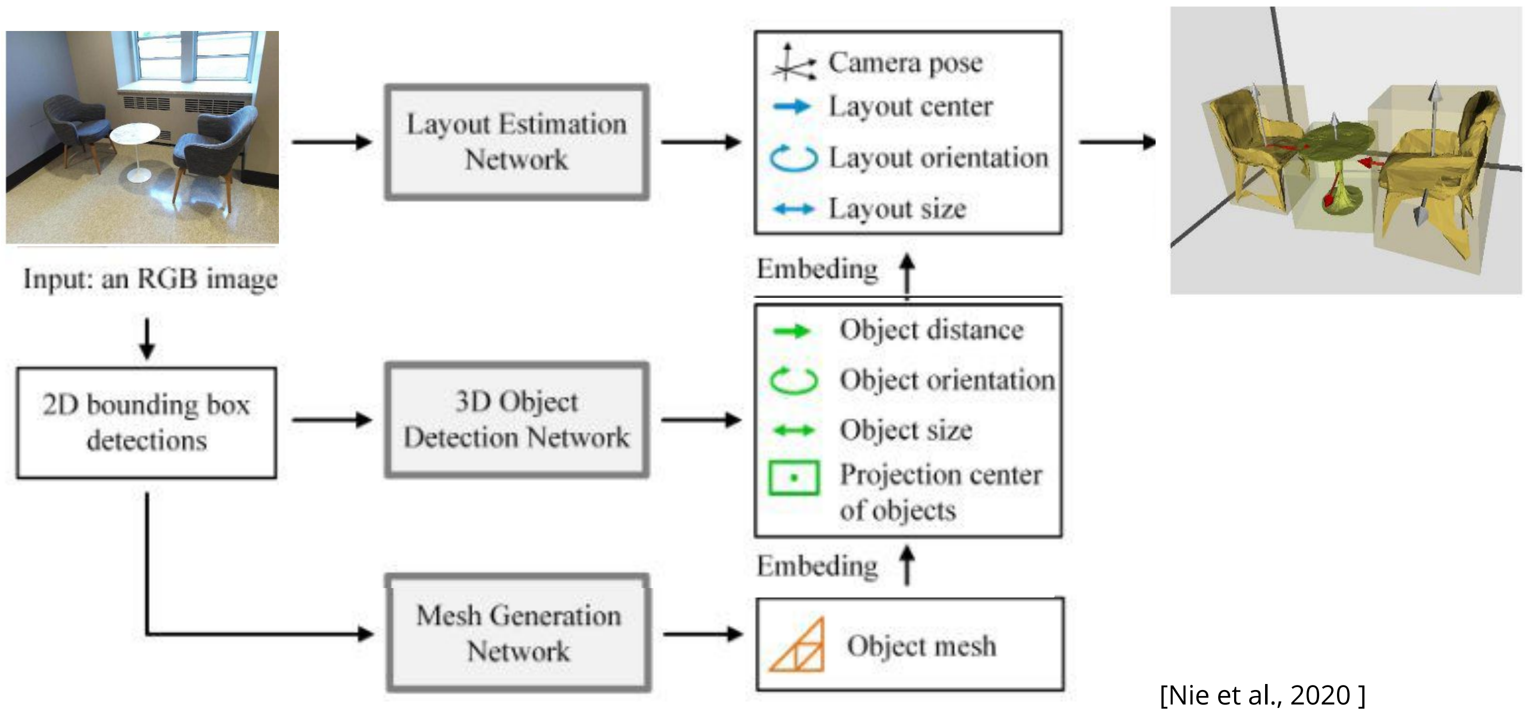
3D Meshes

[Gkioxari et al., 2019]

TUDelft

Scene Level 3D Reconstruction

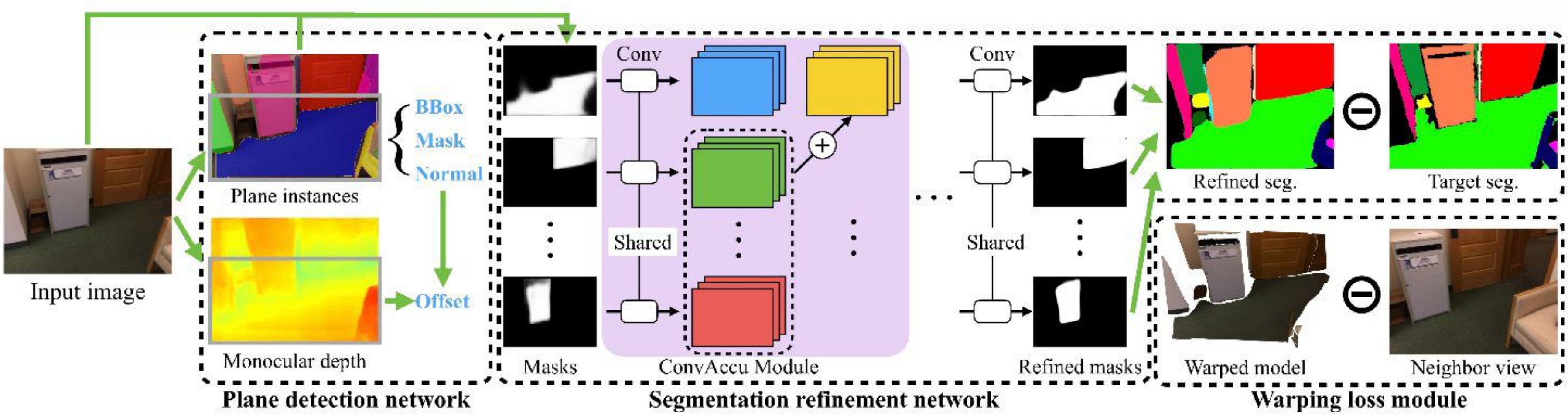
- **Mesh Based Approach**
  - Total3DUnderstanding : Combined scene understanding and mesh reconstruction



[Nie et al., 2020]

Scene Level 3D Reconstruction

- **Piecewise Planar Approach**
  - PlaneRCNN : Plane Detection and 3D Reconstruction using single image



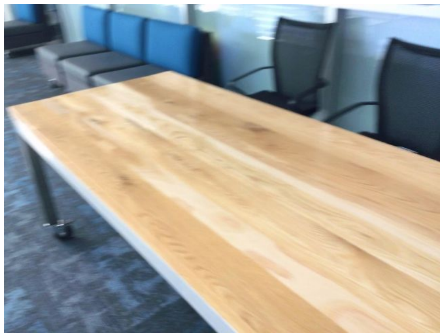
[Liu et al., 2019]

  - Jointly refines all the segmentation masks with a novel loss enforcing the consistency with a nearby view during training.

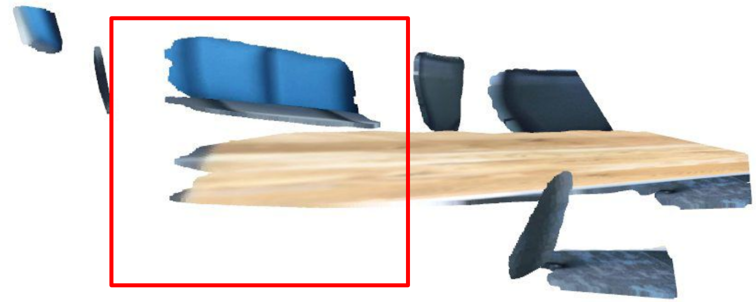
Investigation of the basic model of Planercnn

Input Image | Piecewise Planar Model

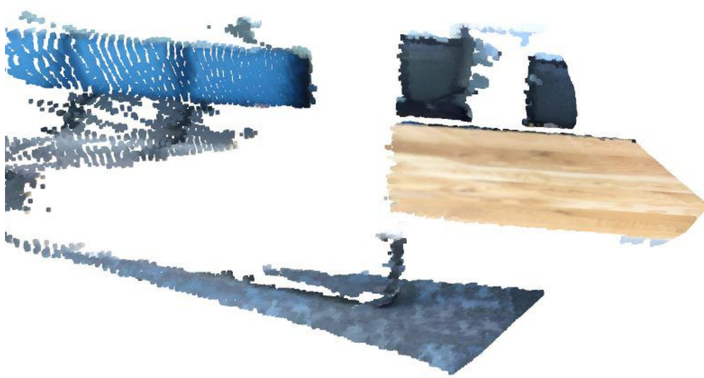
Investigation of the basic model of Planercnn



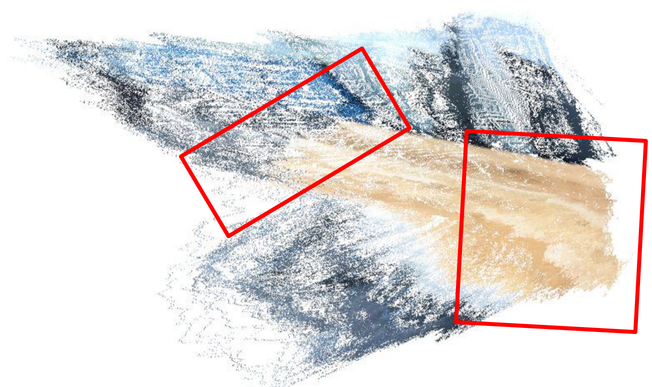
Input Image



Piecewise Planar Model



Ground Truth



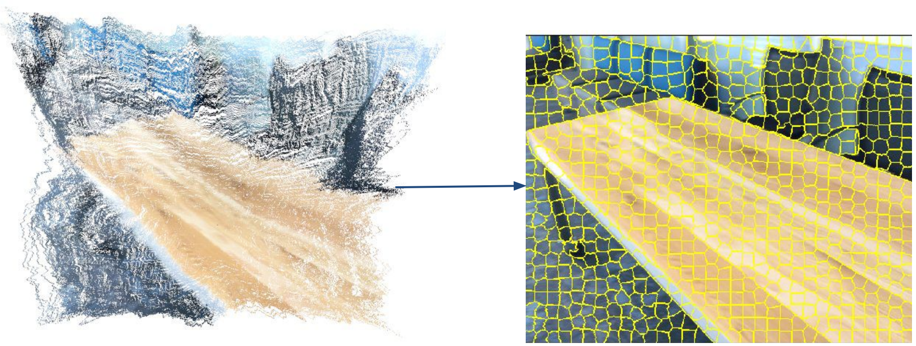
Point Cloud

TUDelft

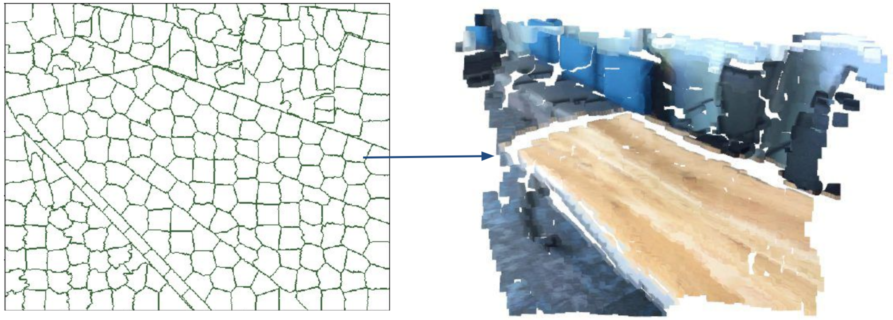
Motivation

- Spatial compatibility within neighbourhood is not maintained
- Inconsistent boundaries and extent of surfaces in reconstructed scene

Potential in using color information for guiding depth consistency at local level during supervision and 3D reconstruction



Segmentation based on spatial and color compatibility



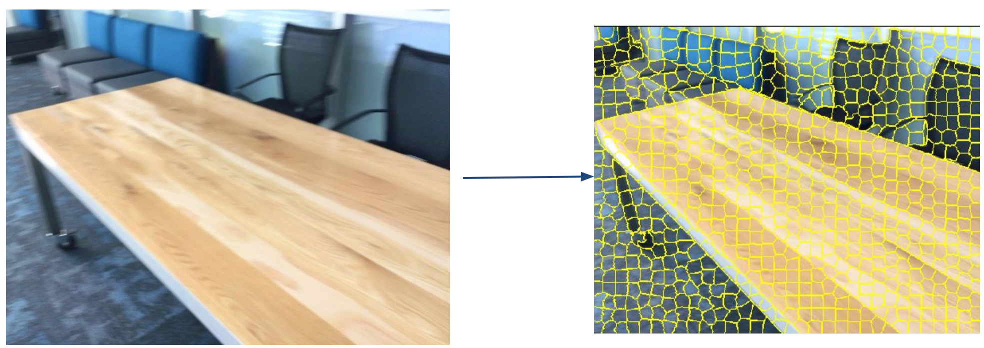
Design learning algorithm using segmented mask

TU Delft

Research questions

Can optimization based on the spatial and color compatibility of pixels within image, help in the improvement of 3D reconstruction from a single image ?

- How does the optimization approach influence the process of 3D Reconstruction in an indoor environment ?

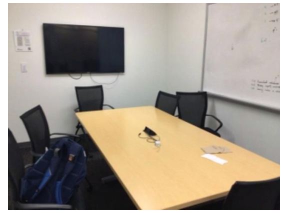


Segmentation based on spatial and color compatibility

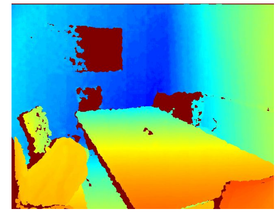
TU Delft

Methodology

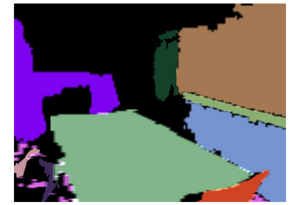
Data Collection



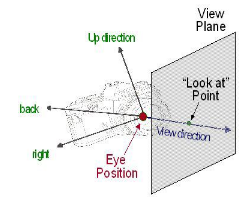
RGB Image



Ground truth Depthmap



Plane annotations



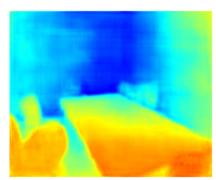
Camera Parameters

CNN Model Setup and Training ← Hyperparameter Optimization using validation dataset

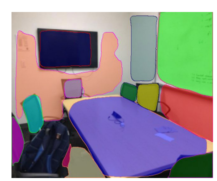
Learning Algorithm ← Use the proposed loss function for *depth optimization*

Output

3D Reconstruction



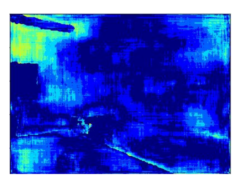
Global Depth

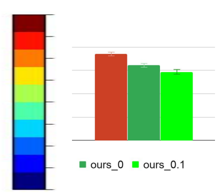


Plane Instances

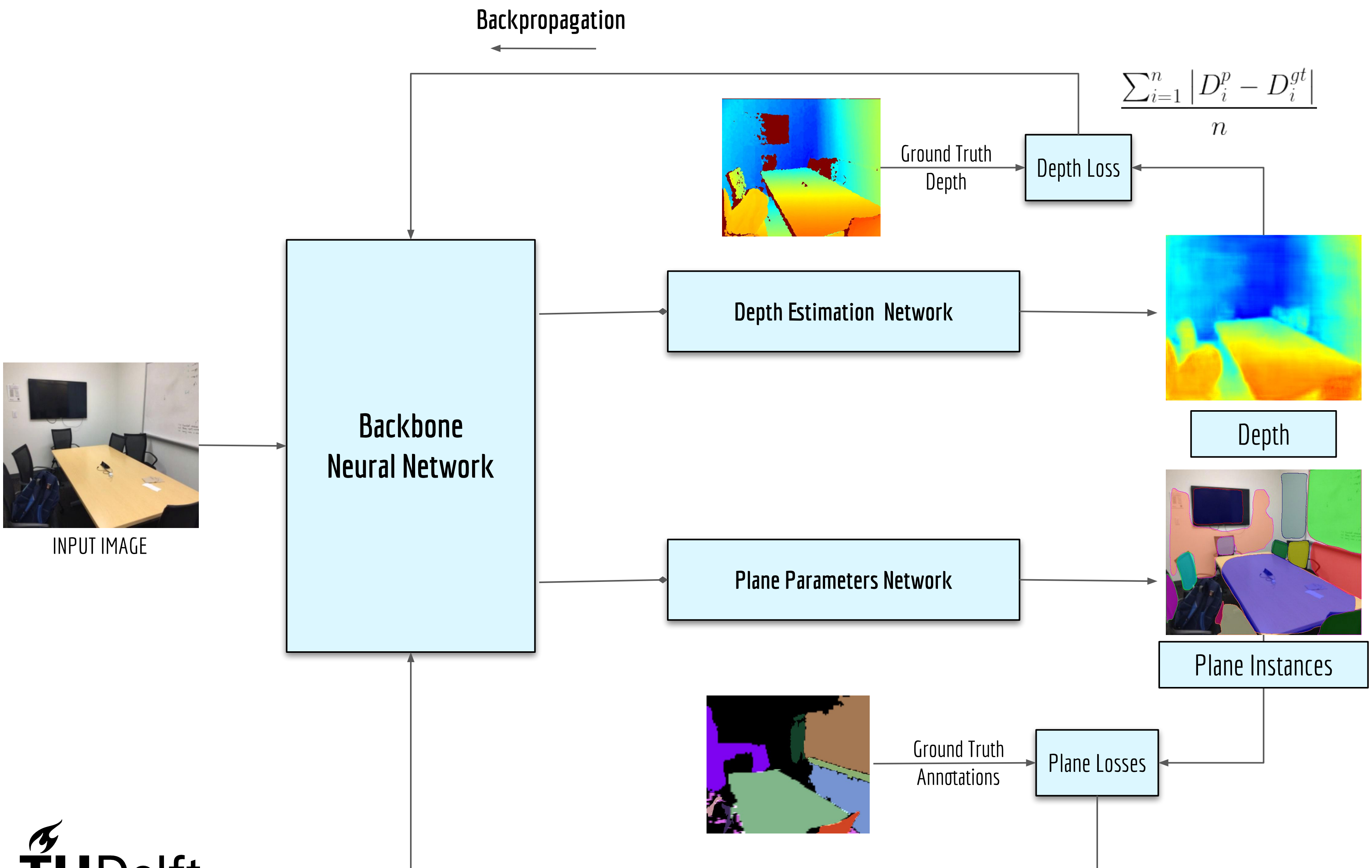
3D Model

Evaluation

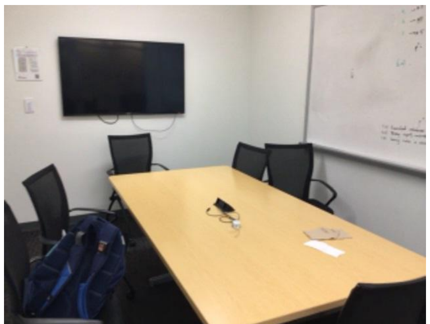




Neural Network Architecture

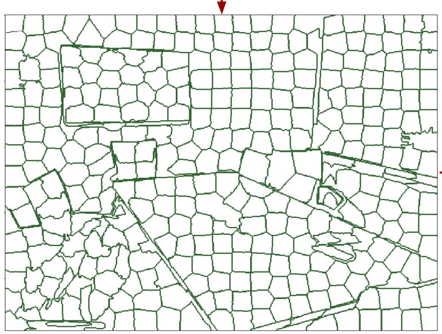


Geometry Aware Depth Loss



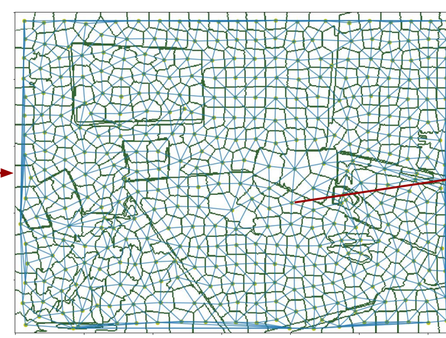
RGB Image

Oversegmentation

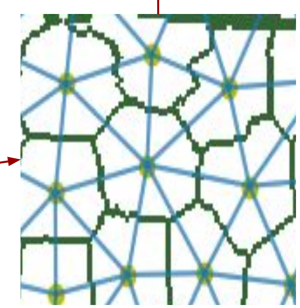


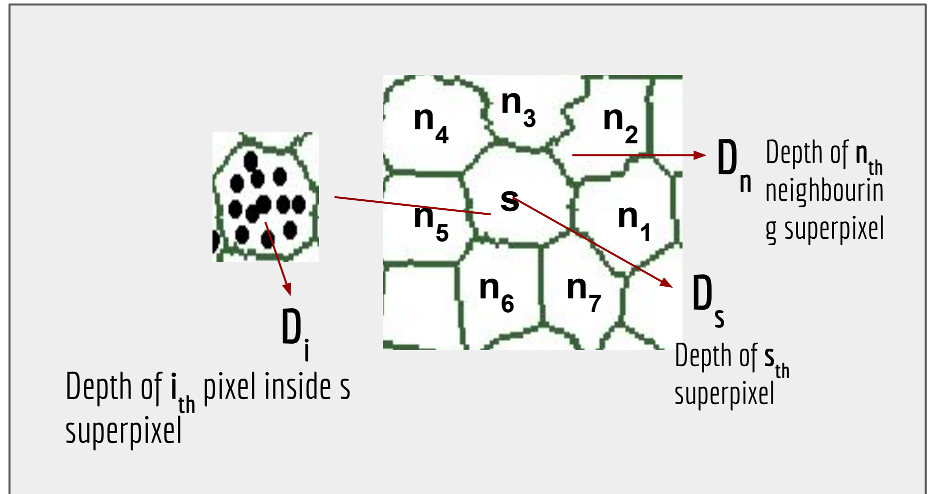
Boundaries of Superpixels

Delaunay Triangulation on geometric centers of superpixels



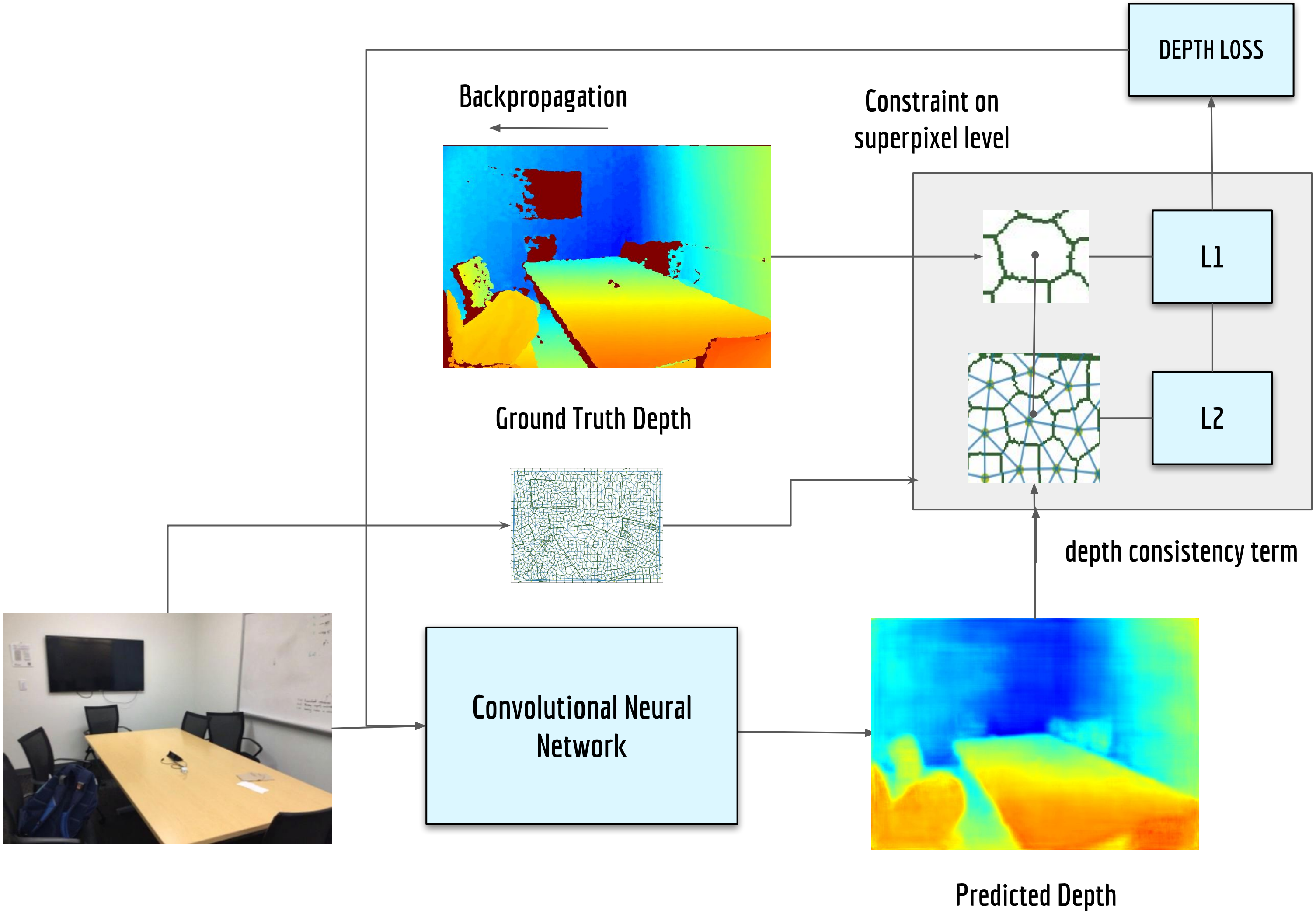
Superpixel Patch with connected neighbors



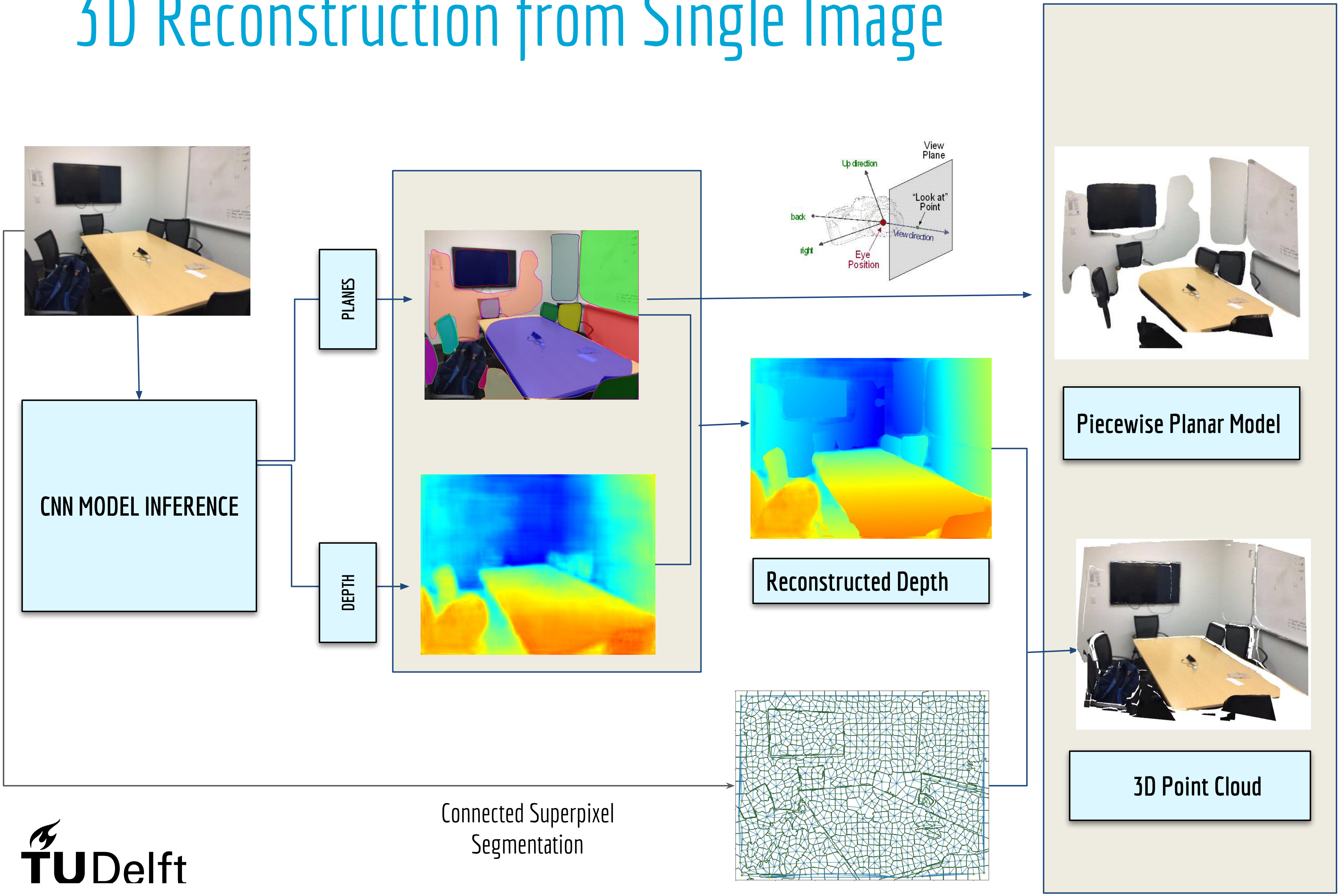


Geometry Aware Depth Loss

$$L = (1 - w)L_1 + wL_2$$



3D Reconstruction from Single Image



TU Delft

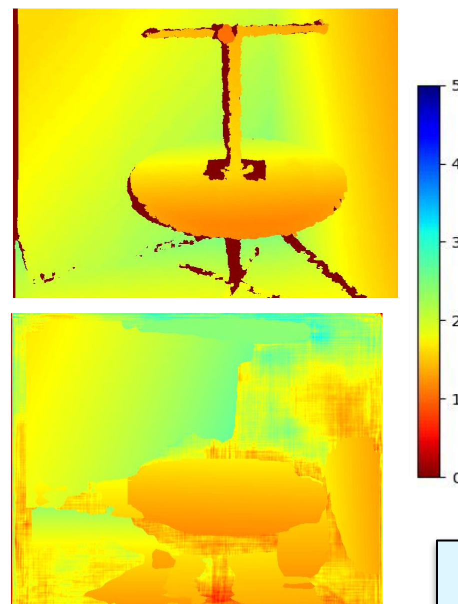
Evaluation

- **Depth Estimation :**
  - Mean Relative Error
  - Root Mean Square Error
  - Accuracy with respect to depth error threshold

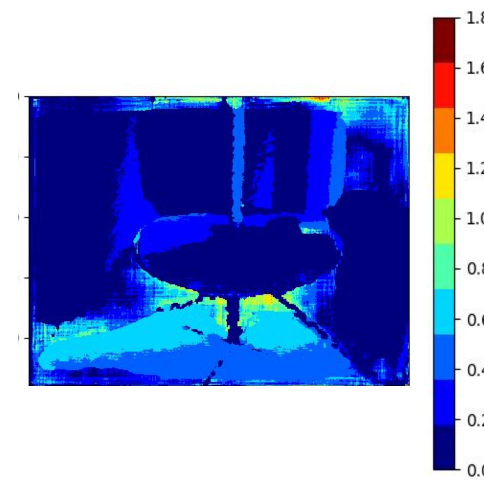
- **Plane Detection :**
  - Average Precision
  - Segmentation Cover
  - Variation of Information
  - Rand Index

Curved Dataset

Planar Dataset



Reconstructed Depth



Error :
| Reconstructed - Ground Truth |

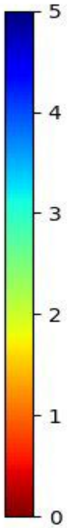
Experiments Setup

- **Training :**
  - Load the weights of pre-trained MaskR-CNN (coco dataset)
  - All layers using randomly sampled images (minibatch : 15)
  - Optimizer : Stochastic Gradient Descent
  - LR =0.00001, momentum =0.9, weight decay = 0.0001

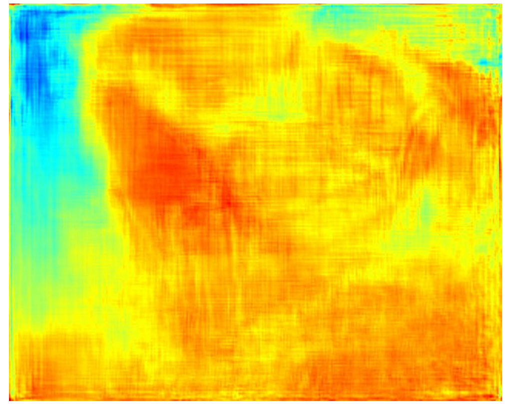
- **Data :**
  - ScanNet : 7000, 1000 and 800 : Training, validation, testing
  - NYU-Depth v2: 645 test images
  - Plane Annotations using benchmark from PlaneR-CNN

- **Tools :**
  - Ubuntu 18.04 + 4GB on-board memory
  - HPC cluster , TU Delft server
  - Deep Learning Ecosystem: Pytorch, skit-learn, numpy, opencv, python, scikit-image
  - Open3D : visualization, rendering 3D models

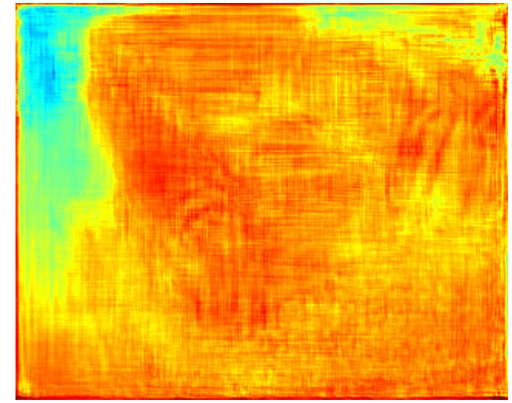
Effect of Superpixel Representation



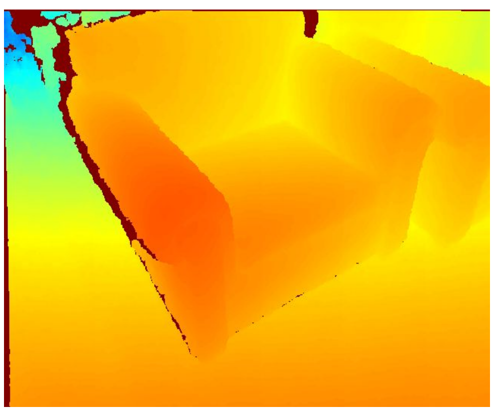
Input



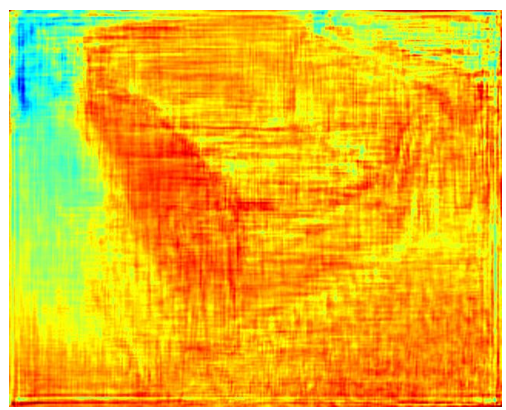
Baseline



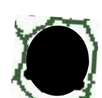
Ours (mean_gt)

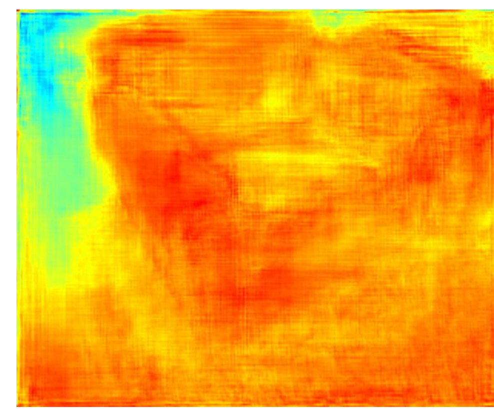


Ground Truth

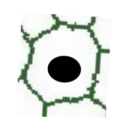


Ours (mean)

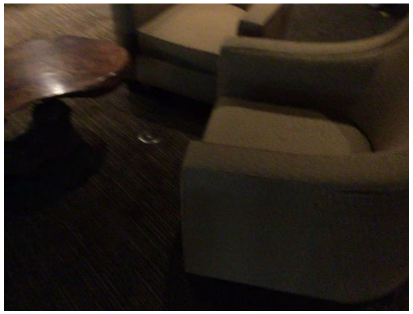




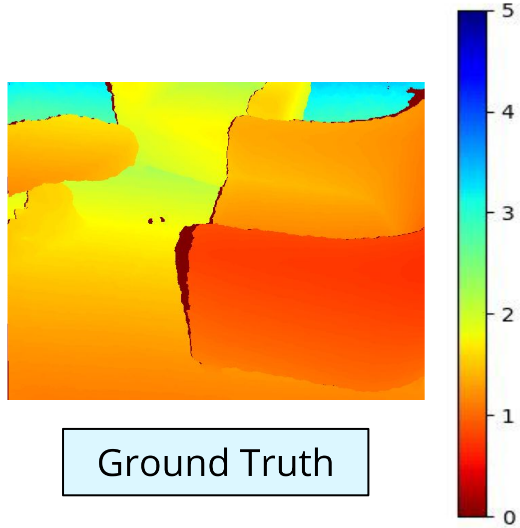
Ours (center)



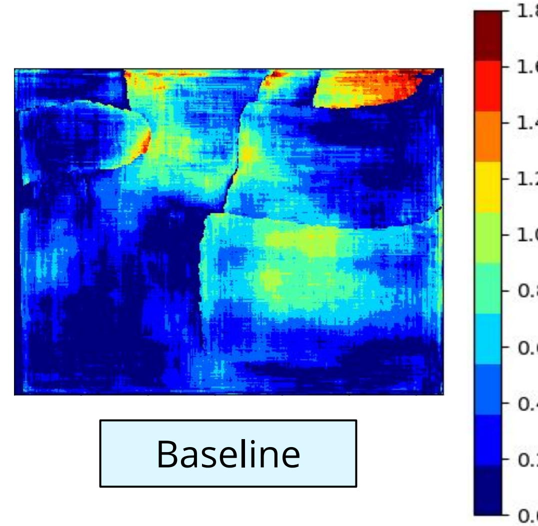
Effect of weight of depth consistency term



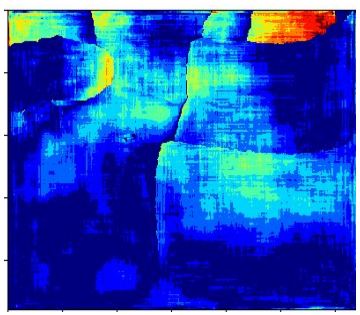
Input



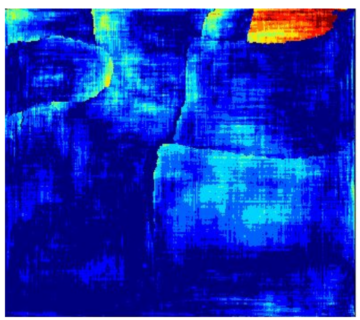
Ground Truth



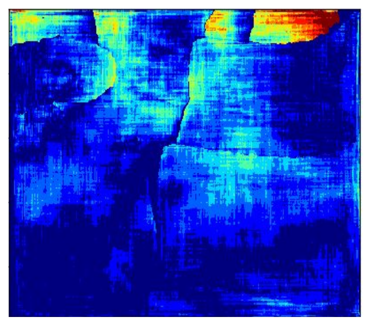
Baseline



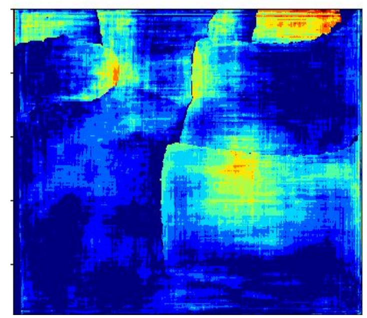
w = 0



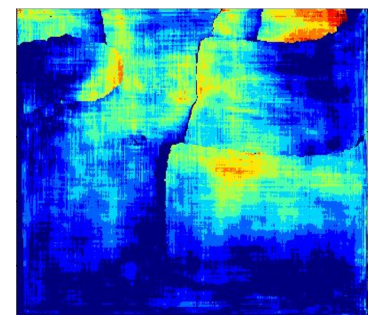
w = 0.1



w = 0.2



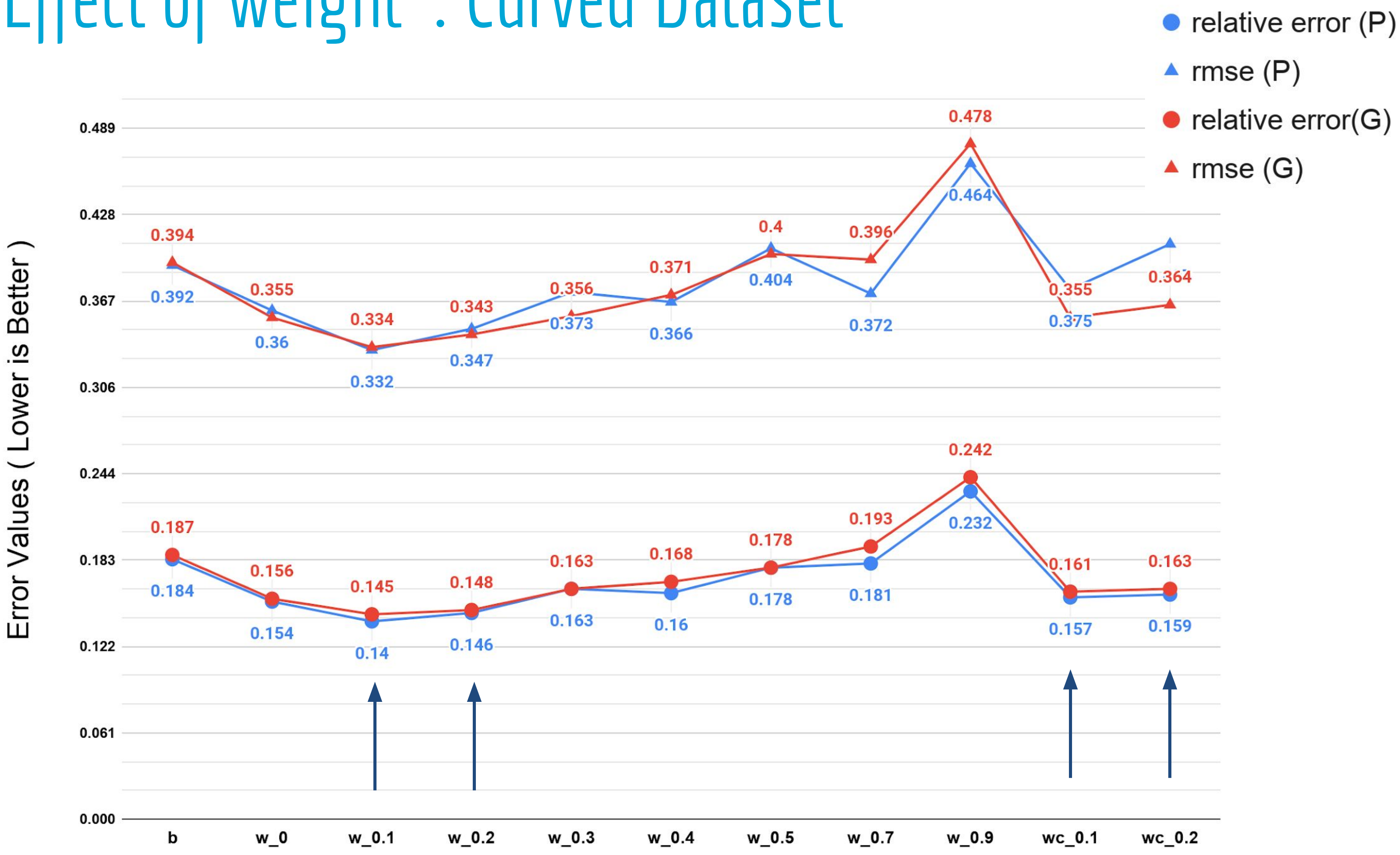
w = 0.4



w = 0.7

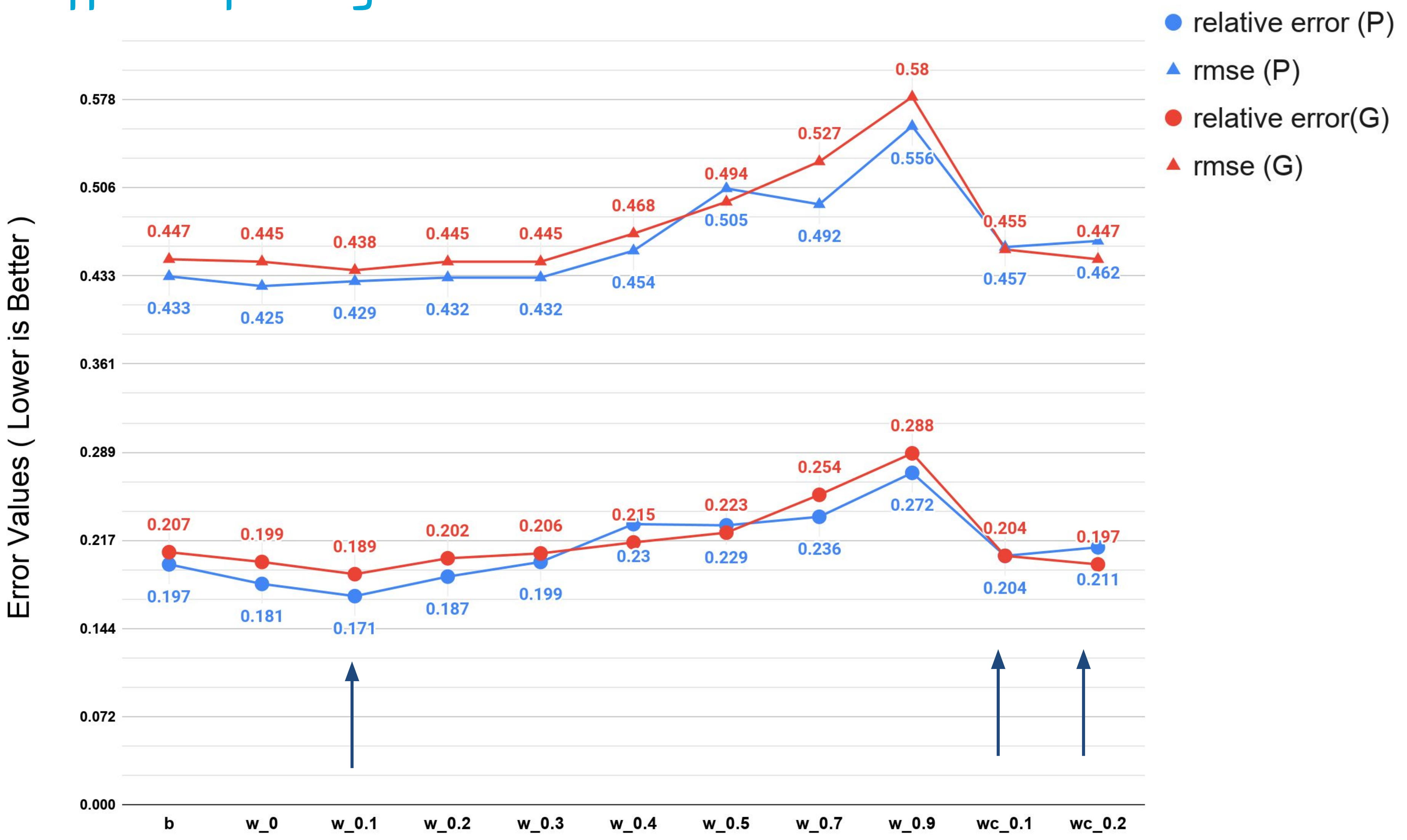
TU Delft

Effect of weight : Curved Dataset



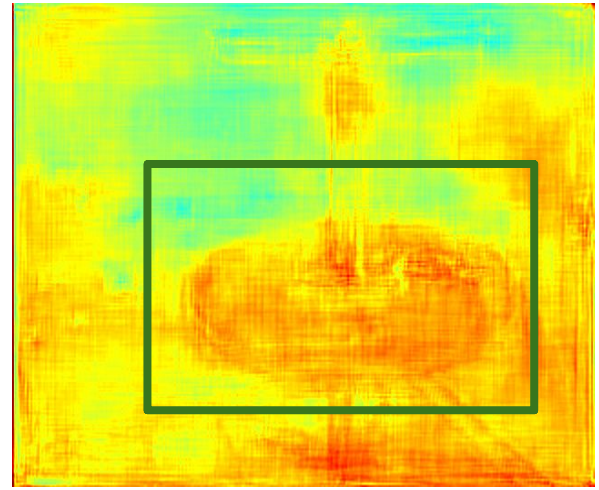
TUDelft

Effect of weight : Planar Dataset

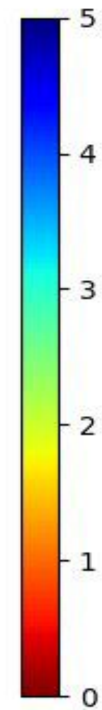


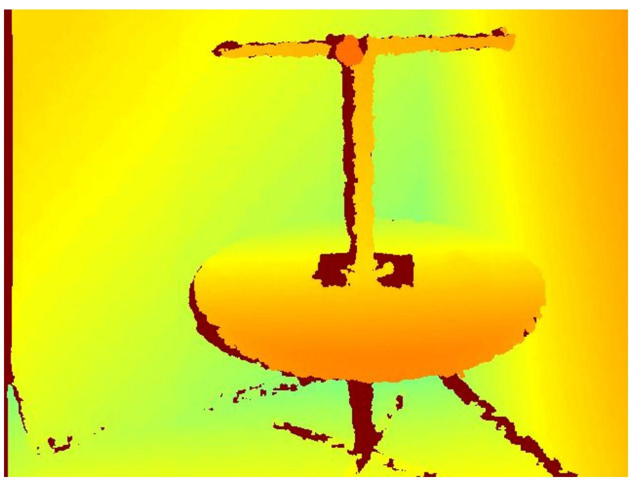
Depth Estimation Results

Input

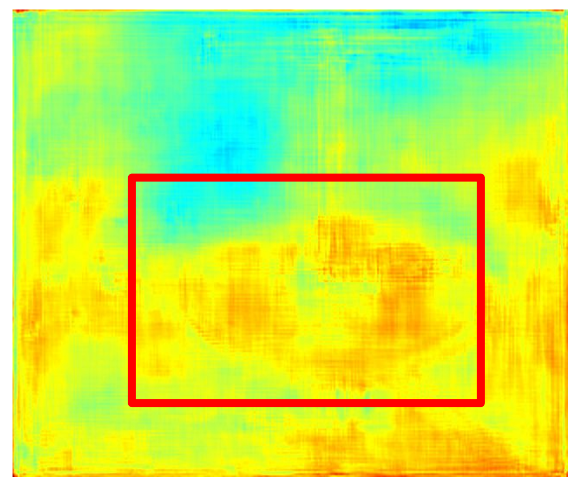


Ours



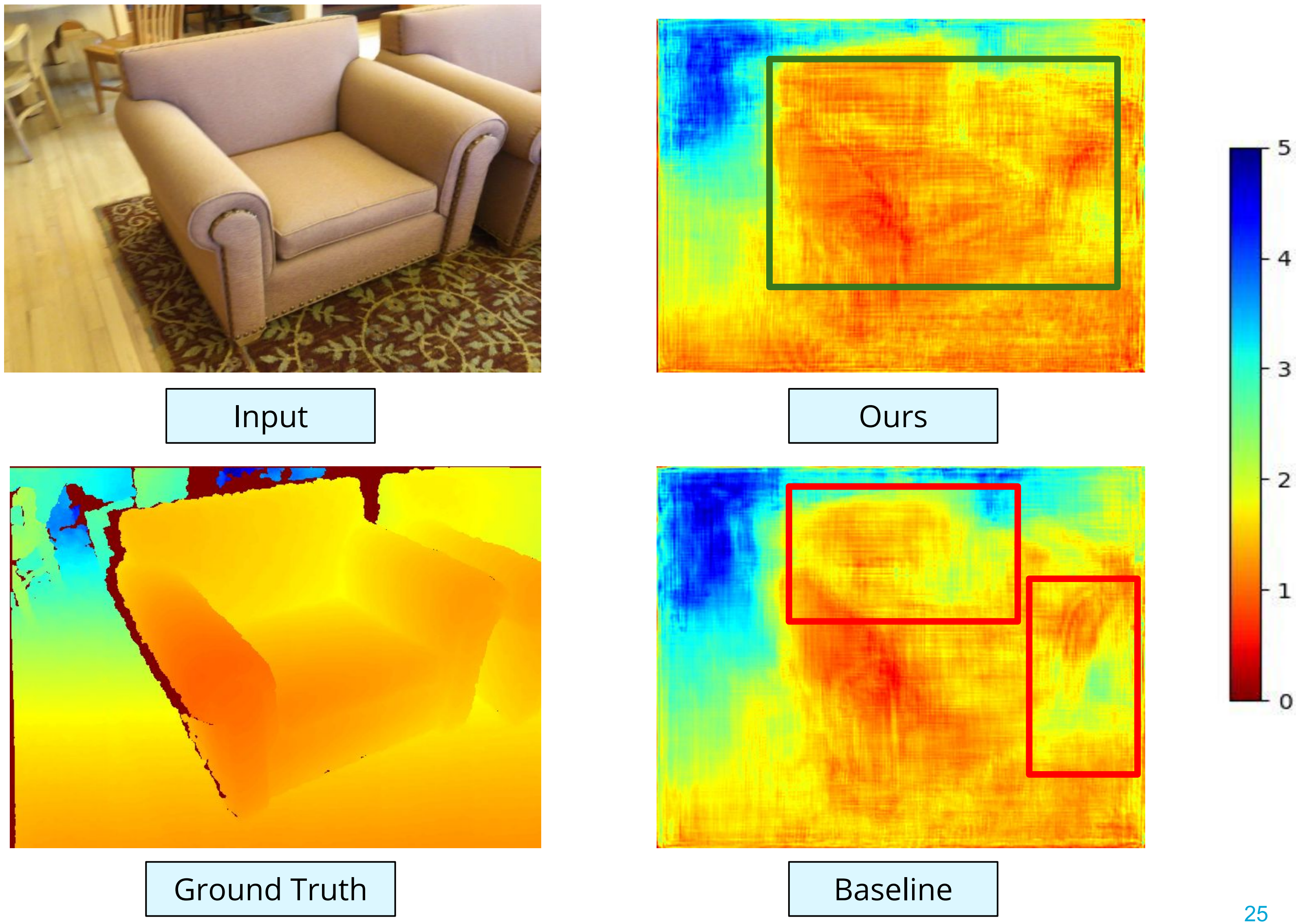


Ground Truth

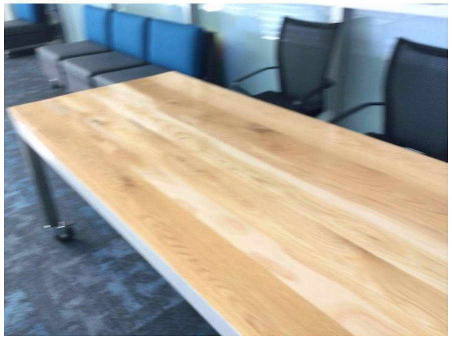


Baseline

Depth Estimation Results



Piecewise Planar Reconstruction Results

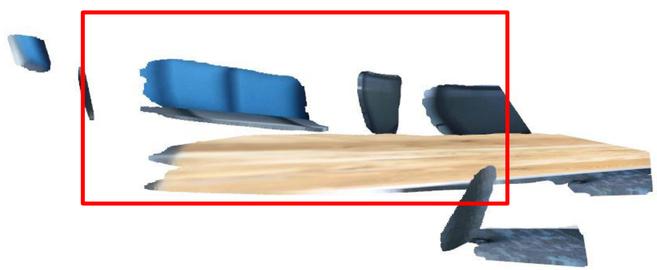


Input

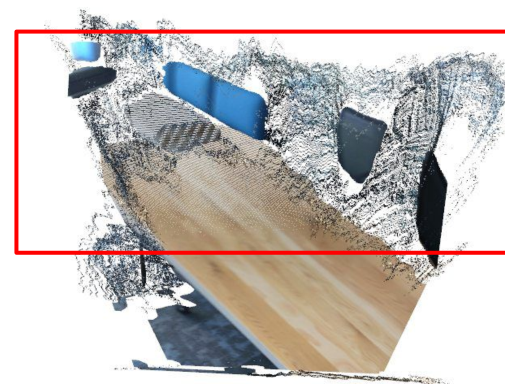
Ours

Ours

Ground Truth



Baseline



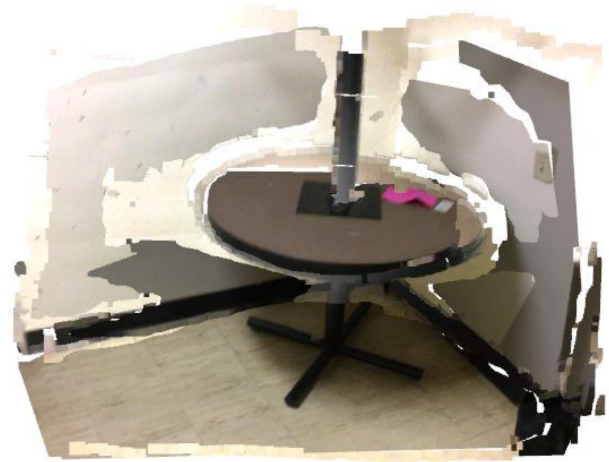
Baseline

TUDelft

Piecewise Planar Reconstruction Results

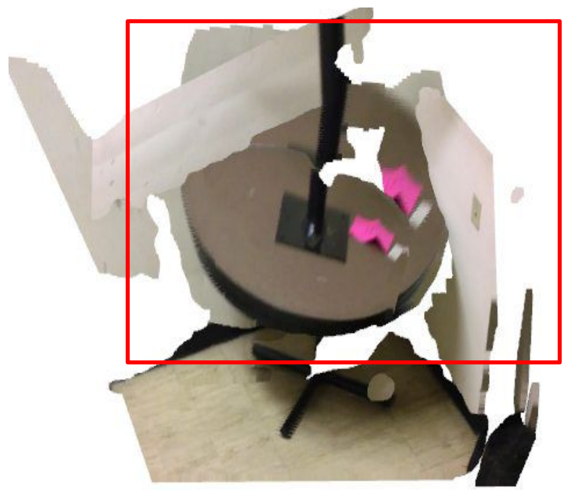
Input

Ours

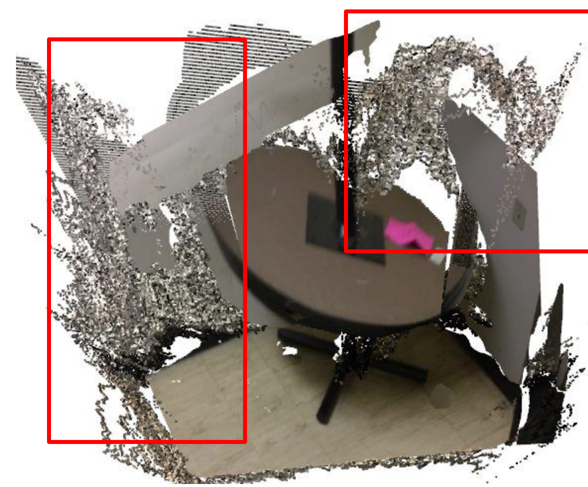


Ours

Ground Truth

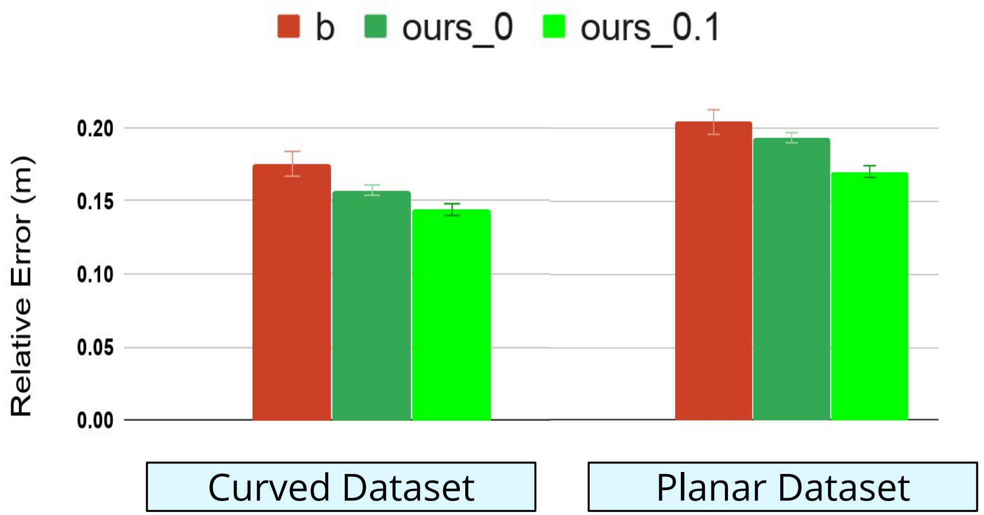


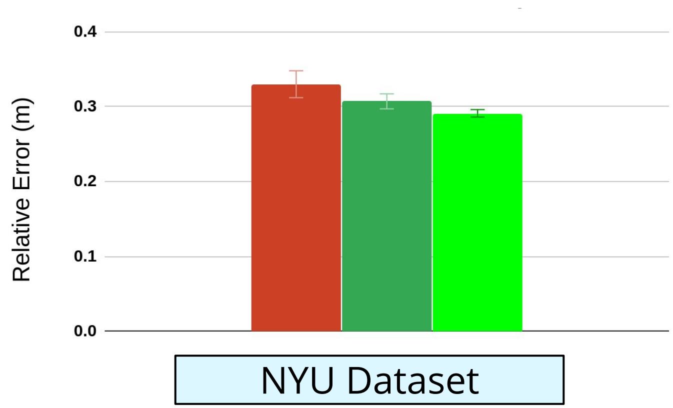
Baseline



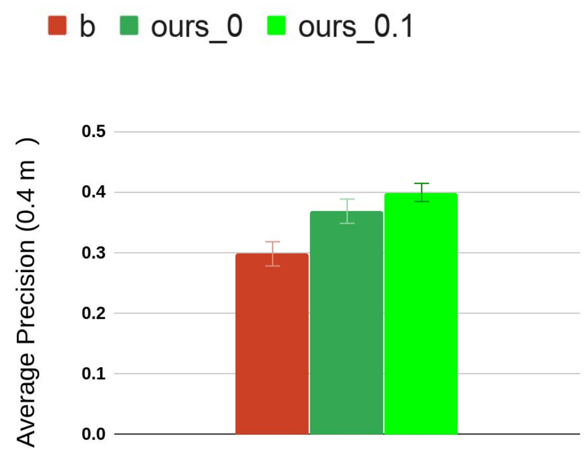
Baseline

Quantitative Evaluation: Piecewise Planar Depth

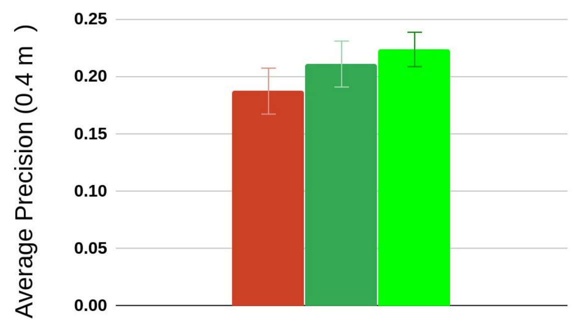




Quantitative Evaluation : Planar Reconstruction

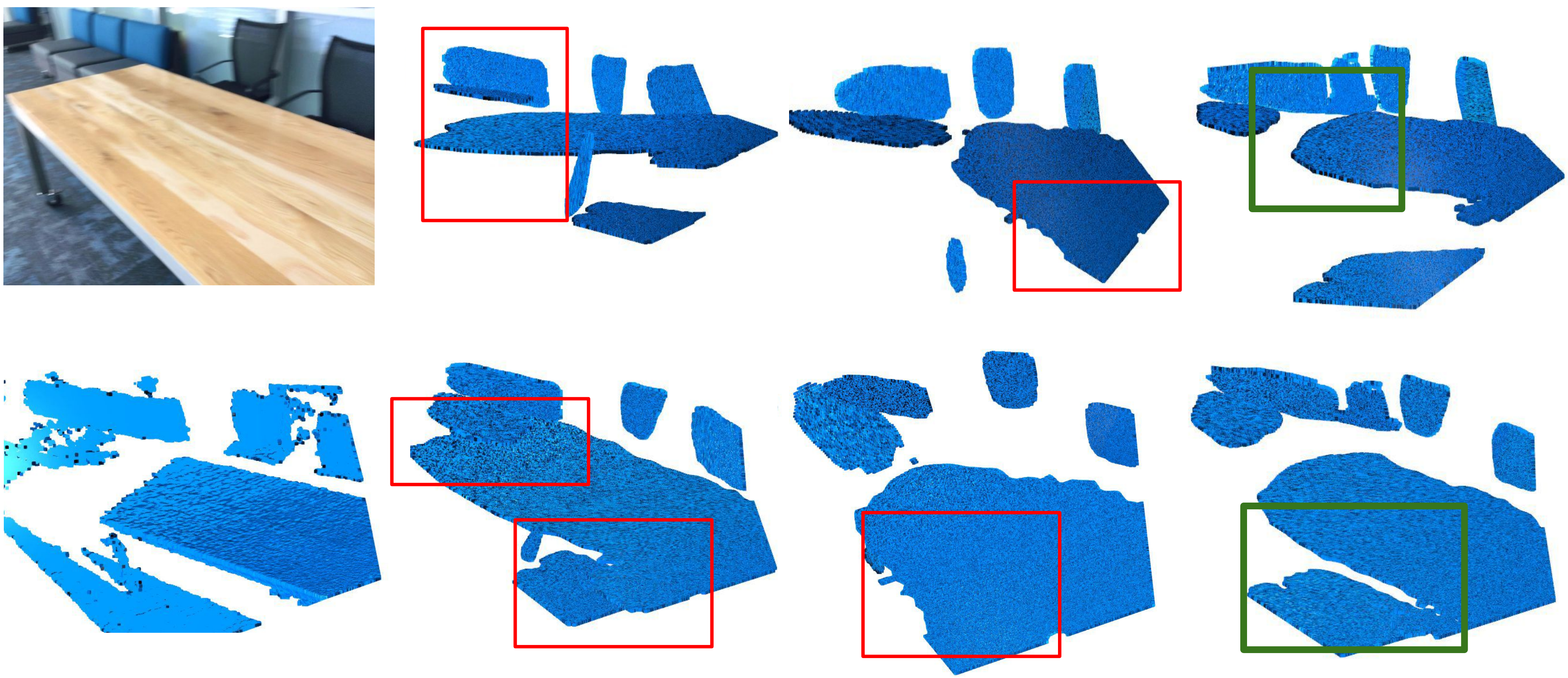


Curved Dataset



Planar Dataset

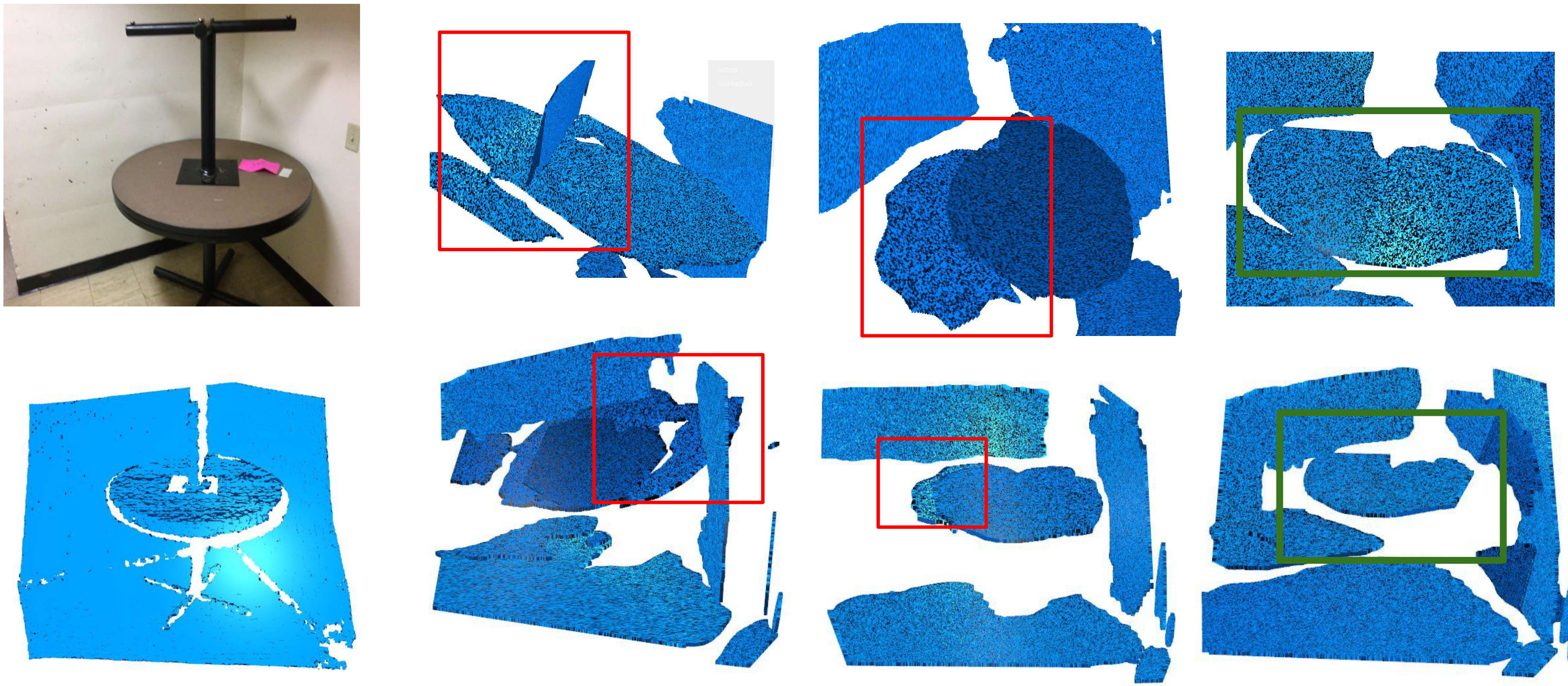
Evaluation : Piecewise Planar Reconstruction



Input | Baseline | Ours (w:0) | Ours (w:0.1)

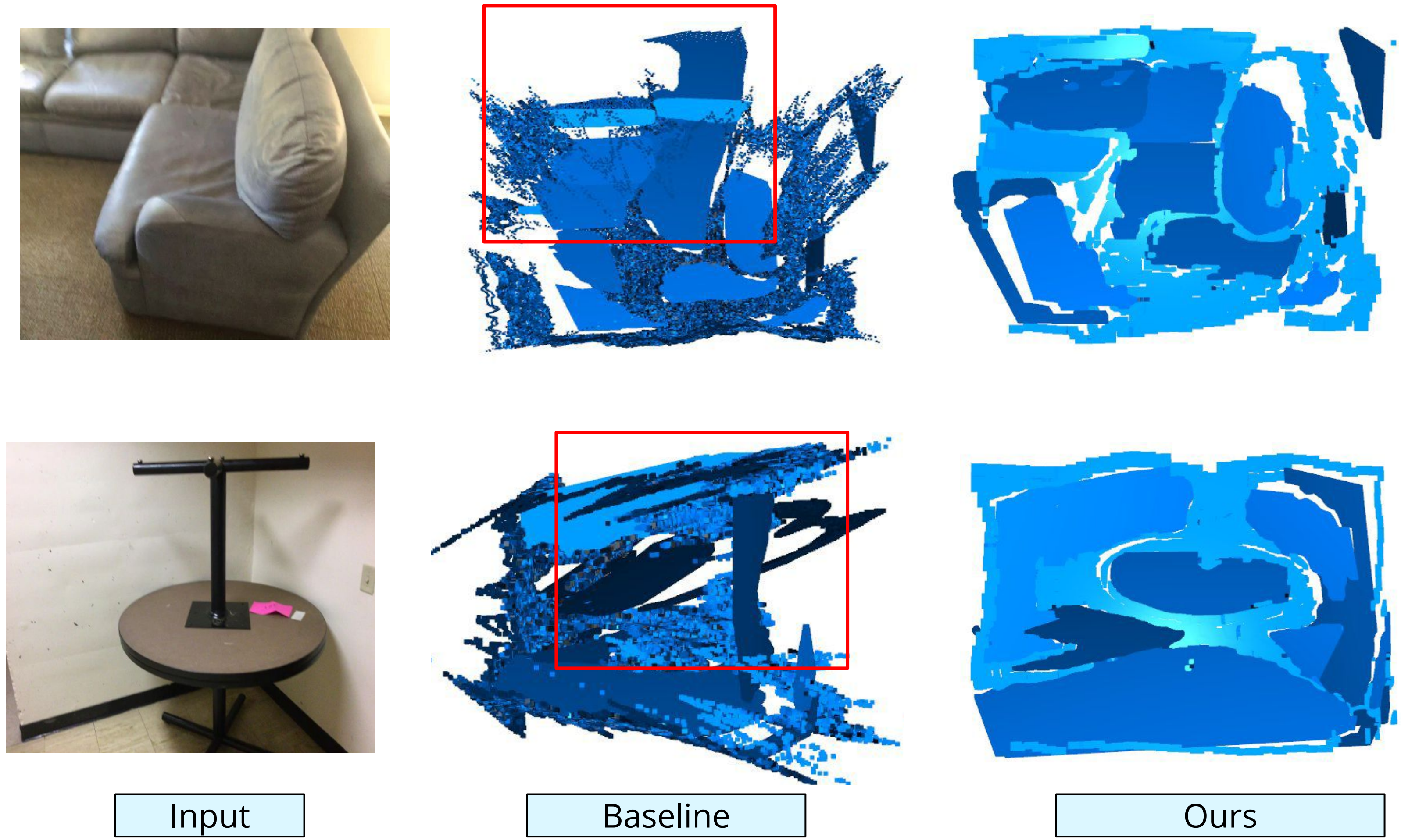
TU Delft

Evaluation : Piecewise Planar Reconstruction

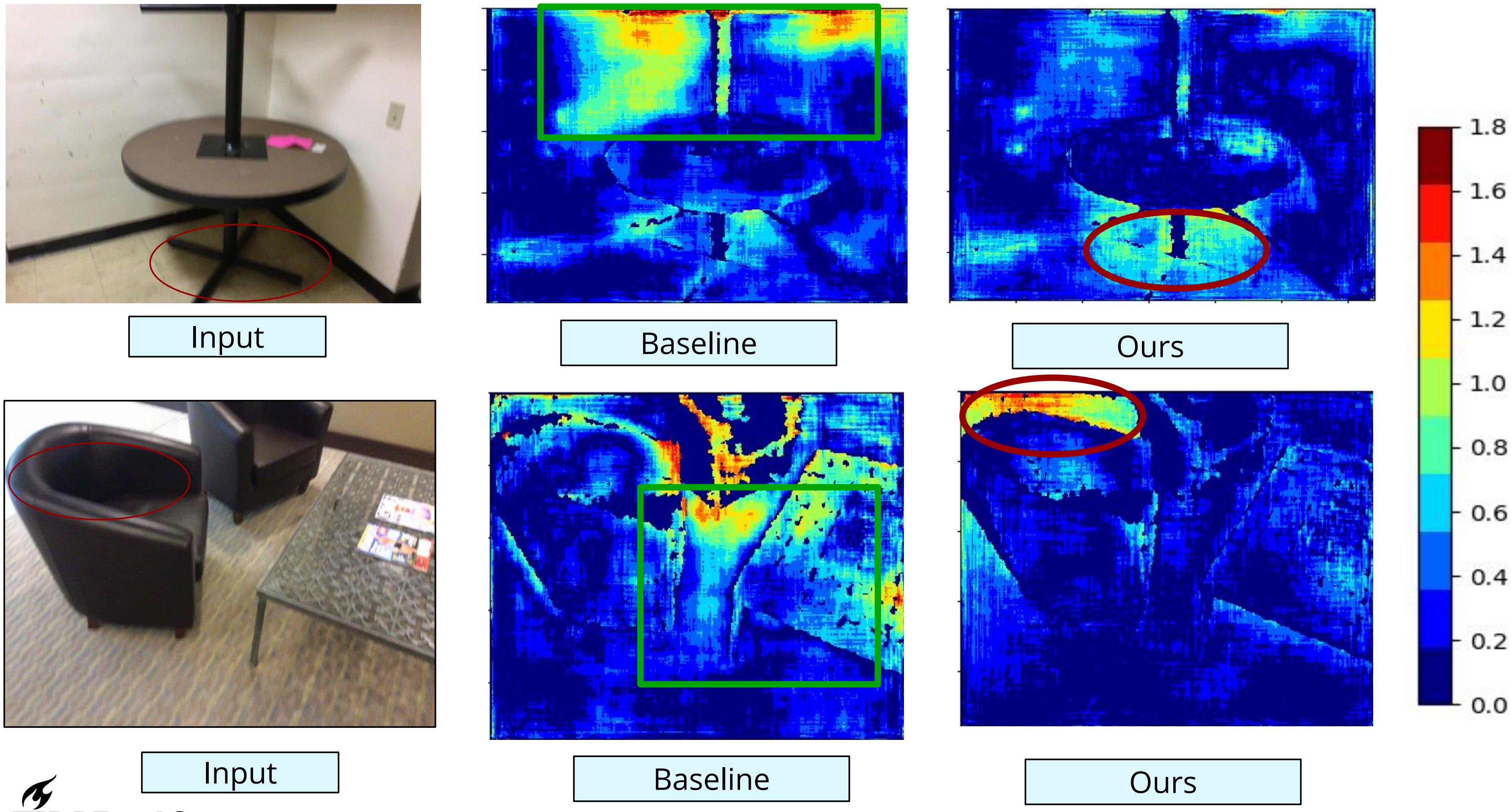


Input | Baseline | Ours (w:0) | Ours (w:0.1)

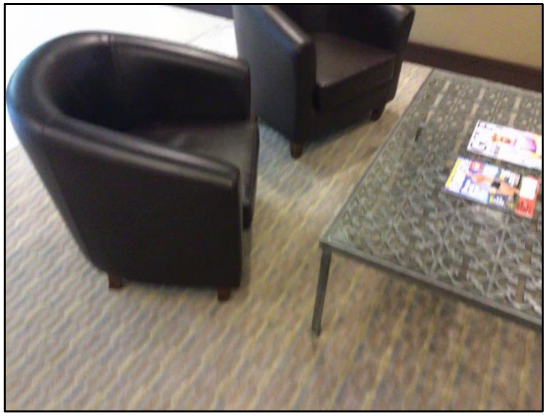
Evaluation : 3D Reconstructed Point Cloud



Limitations and Challenges

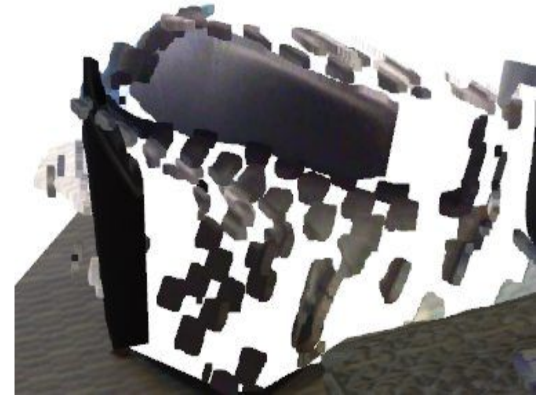


Limitations and Challenges

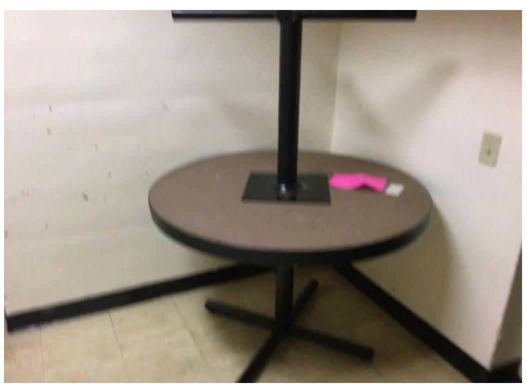


Input

Image View



Side View



Input

Image View

Side View

Limitations and Challenges

- No error reduction or over smoothing in non-differentiable color regions
- Training Time (7000 images)
  - Base: 6 hrs
  - Ours: 1st term : 9 hrs; both terms : 18 hrs
    - Create metadata beforehand
    - Using CUDA compatible preprocessing
- Scale of research framework the experiment
  - Limited computation power
  - Generalization using more datasets
- Limitations of superpixel segmentation and histograms

Conclusion

- The proposed optimization approach helps in improving the 3D reconstruction in indoor environment
- Depth consistency term refines the reconstructed depth within local neighborhood based on spatial and color compatibility
- Second term affects both curved and planar surfaces while first term based on superpixel has more effect on curved surfaces in depth estimation
- In piecewise planar models, the surface extent and orientation improves for detected planar regions
- Consistency during 3D reconstruction step helps in better understanding of non-planar regions in the scene and has further potential

Future Work

- Using other superpixel segmentation and color comparison methods
- Testing with other datasets and neural networks for more insights
  - improved real world dataset
  - Synthetic Dataset
  - Different depth of network
- Exploration in Applications :
  - Using multiple images for full 3D reconstruction
  - Using semantic labels for direct analysis and further processing
  - Indoor Navigation and localisation using signature of 3D model
  - Using old historic images for virtual models in culture and heritage
- Explore using normal orientation term during supervision and 3D reconstruction

Contribution

- Introducing new learning approach for neural networks in the context of 3D reconstruction

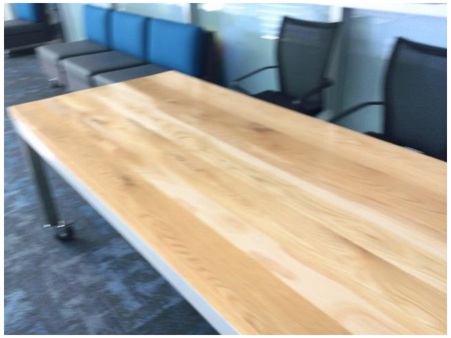
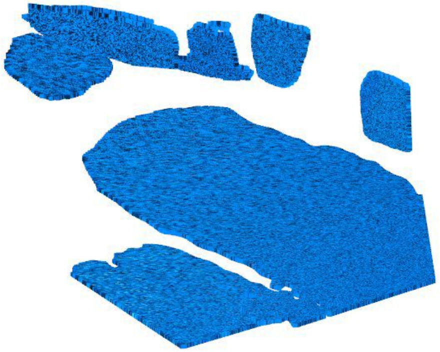
- Open source code for research community : https://github.com/cgarg-tud/GeomAwareLoss

- Working on paper :

Indoor 3D Reconstruction using Single Image

Abstract : 3D indoor reconstruction has been an important research area in the field of computer vision and photogrammetry. While the initial techniques developed for this purpose use sensor devices and multiple images for data acquisition and extracting 3D information and representation of the scene, with the advent of deep learning techniques, there has been a good progress in extracting 3D information of an indoor scene reconstruction using a single image. This has potential in minimizing user efforts and cost for data acquisition. The current state of the art method involves two main components, the global depth map and plane instances. After investigating the current state of the art methods, it is observed that there is inconsistency in reconstructed surface

Thank You !!

TU Delft

References

A. Donaubauer, T. K. Kohoutek, and R. Mautz. CityGML als Grundlage für die Indoor Positionierung mittels Range Imaging. abc-Verl., Heidelberg, 2010. ISBN 978-3-938833-42-1.

Canvas. Canvas: Create a 3d model of your home in minutes, 2016.https://www.youtube.com/watch?v=XA7FMoNAK9M.

S. Zlatanova and U. Isikdag.3d indoor models and their applications. Springer, 2017.

S.-H. Tsang. Review:Deepmask, an instance segment proposal method driven by convolution neural networks.2018. https://towardsdatascience.com/review-deepmask-instance-segmentation-30327a072339

E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation.2016

H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. 2016.

K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, and J. Wang. High-resolution representations for labeling pixels and regions. arXiv preprint arXiv:1904.04514, 2019.

Kim. Deep object detectors. techreport, Slideshare. 2017 .https://www.slideshare.net/IldooKim/deep-object-detectors-1-20166

B. D. Brabandere, D. Neven, and L. V. Gool. Semantic instance segmentation with a discriminative loss function. 2017. https://arxiv.org/pdf/1708.02551.pdf

D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In Advances in neural information processing systems, pages 2366–2374, 2014.

C. Liu, J. Yang, D. Ceylan, E. Yumer, and Y. Furukawa. Planenet: Piece-wise planar reconstruction from a single rgb image. In Computer Vision and Pattern Recognition(CVPR), 2018. https://arxiv.org/abs/1804.06278.pdf .