Virtual Content Generation via Deep Learning

Yinda Zhang

Research Scientist @ Google
yindaz@google.com, www.zhangyinda.com

GAMES Webinar

A bit about me...

Yinda Zhang

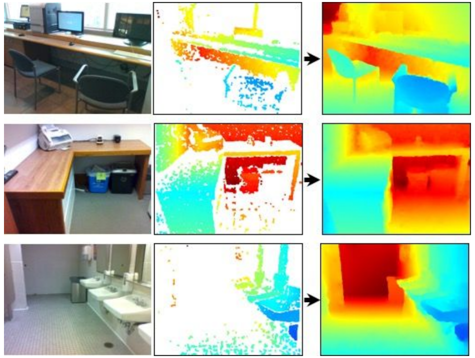
I am a Researsh Scientist at Google. My research interests lie at the intersection of computer vision, computer graphics, and machine learning. Recently, I focus on empowering 3D vision and percetion via machine learning, including dense depth estimation, 3D shape analysis, and 3D scene understanding. I received my Ph.D. in Computer Science from Princeton University, advised by Professor Thomas Funkhouser. Before that, I received a Bachelor degree from Dept. Automation in Tsinghua University, and a Master degree from Dept. ECE in National University of Singapore co-supervised by Prof. Ping Tan and Prof. Shuicheng Yan.

Email: yindaz (at) gmail (dot) com

Find me on:

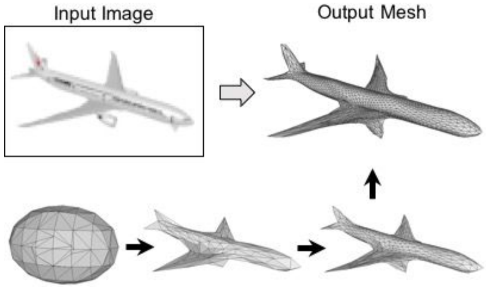
A bit about me...

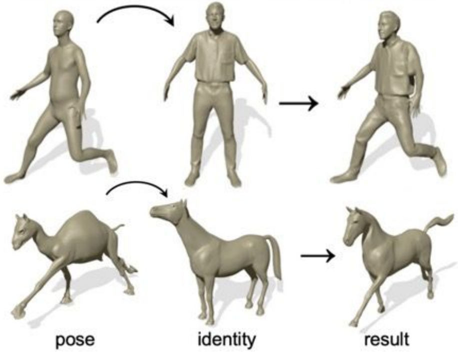
3D Scene Understanding



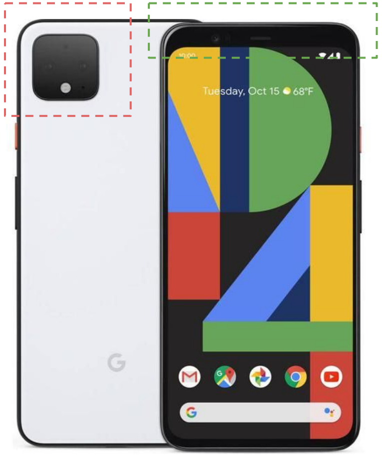
Data-Driven Approach

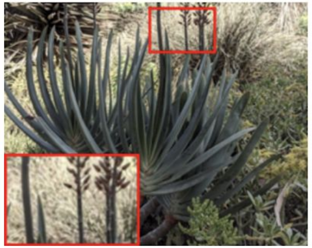
Shape Analysis



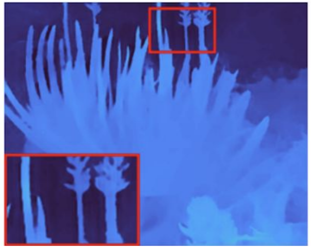


Depth Sensing on Device





Image



Depth

Image

Depth

Pixel4 Rear Facing Camera: https://ai.googleblog.com/2019/12/improvements-to-portrait-mode-on-google.html
Pixel4 Front Facing Camera: https://ai.googleblog.com/2020/04/udepth-real-time-3d-depth-sensing-on.html
Zhang et.al., Du2Net: Learning Depth Estimation from Dual-Cameras and Dual-Pixels, arXiv:2003.14299

Augmented Reality


ASE OPS
GREEN

A pipeline to generate virtual content...

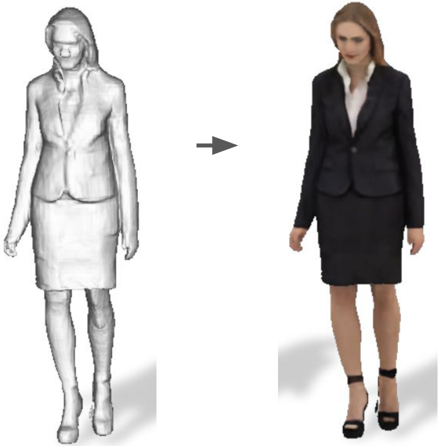
Generate 3D

Guo et.al., The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting, SIGGRAPH 2019.

Manipulate 3D

Wang et.al., Neural Pose Transfer by Spatially Adaptive Instance Normalization, CVPR 2020.

Render 3D



Saito et.al., PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization, ICCV 2019.

A pipeline to generate virtual content...

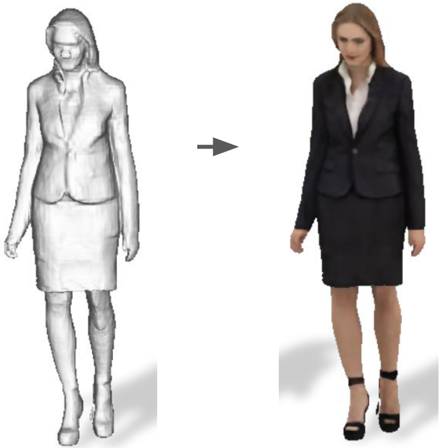
Generate 3D

Guo et.al., The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting, SIGGRAPH 2019.

Manipulate 3D

Wang et.al., Neural Pose Transfer by Spatially Adaptive Instance Normalization, CVPR 2020.

Render 3D



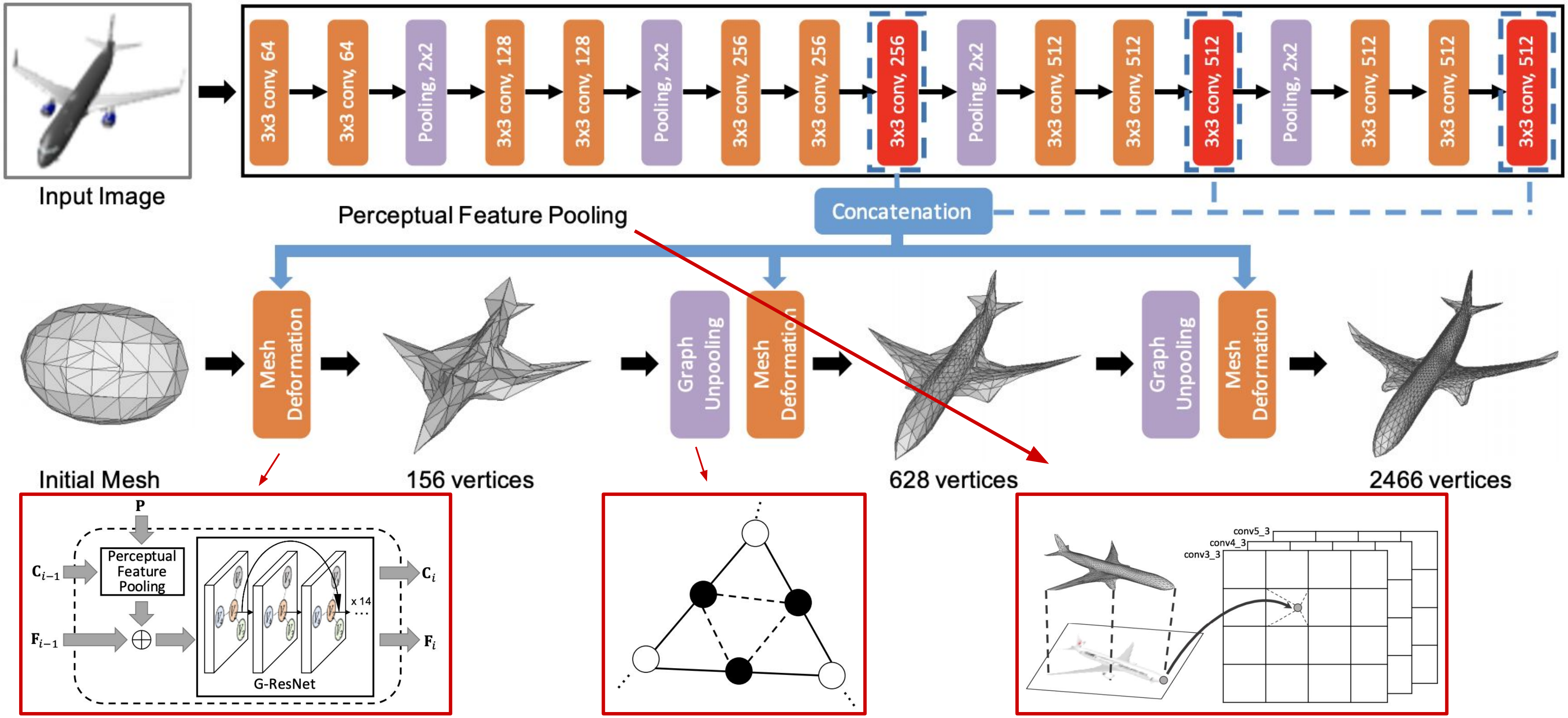
Saito et.al., PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization, ICCV 2019.

3D Shape Generation from a Single Image

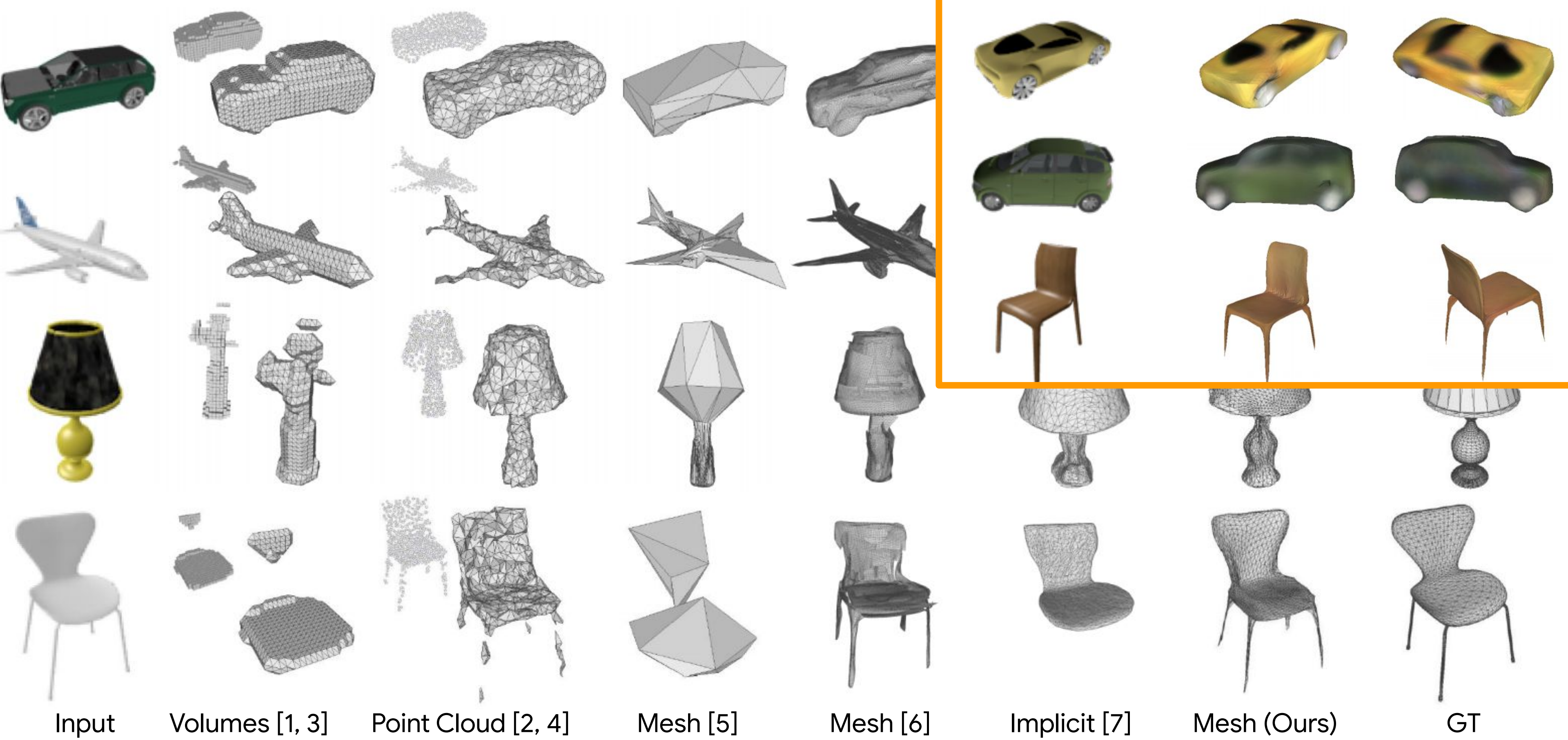
Input Volumes [1, 3] Point Cloud [2, 4] Mesh (Ours) Mesh + Texture (Ours)

[1] Choy et.al., 3dr2n2: A unified approach for single and multi-view 3d object reconstruction, ECCV 2016.
[2] Fan et.al., A point set generation network for 3d object reconstruction from a single image, CVPR 2017.
[3] Lorensen et.al., Marching cubes: A high resolution 3d surface construction algorithm, SIGGRAPH 1987.
[4] Bernardini et.al., The ball-pivoting algorithm for surface reconstruction, IEEE Trans. Vis. Comput. Graph 1999.

Pixel2Mesh

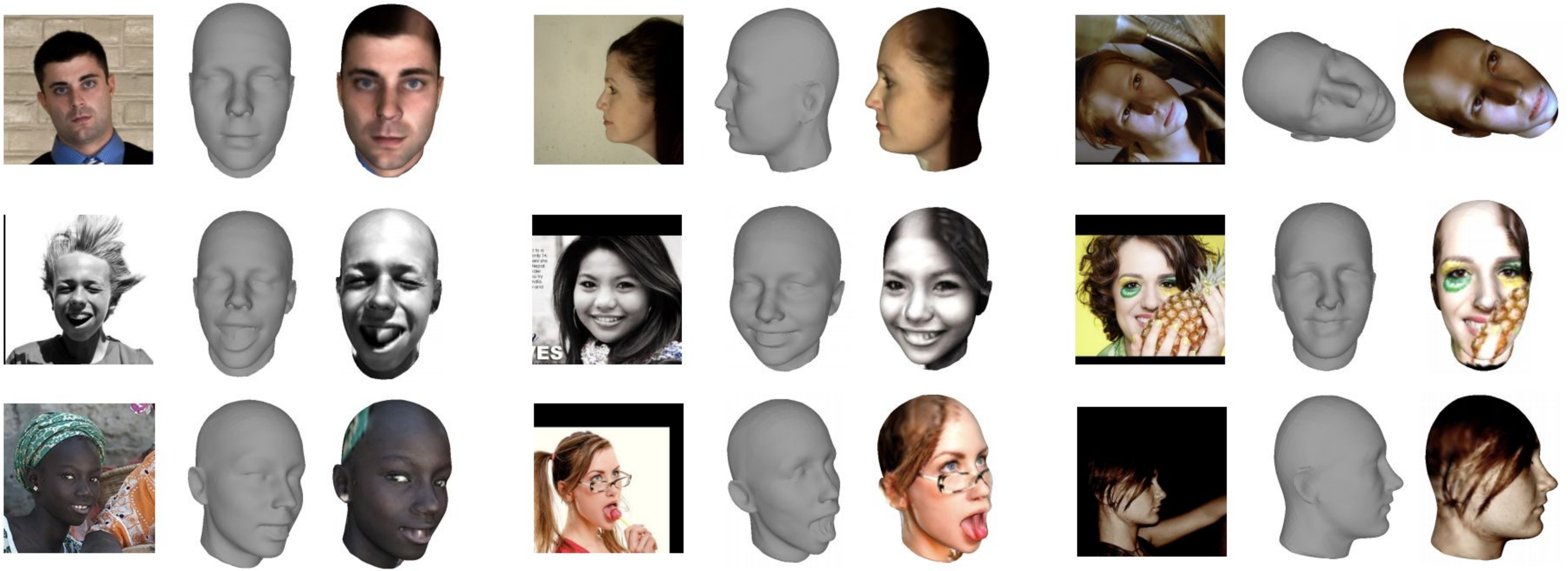


Input Image
3x3 conv, 64
3x3 conv, 64
Pooling, 2x2
3x3 conv, 128
3x3 conv, 128
Pooling, 2x2
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
Pooling, 2x2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
Pooling, 2x2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
Perceptual Feature Pooling
Concatenation
Mesh Deformation
Graph Unpooling
Mesh Deformation
Graph Unpooling
Mesh Deformation
Initial Mesh
156 vertices
628 vertices
2466 vertices
P
Perceptual Feature Pooling
C_{i-1}
F_{i-1}
C_i
F_i
x 14
G-ResNet
conv5_3
conv4_3
conv3_3



[1] Choy et.al., 3dr2n2: A unified approach for single and multi-view 3d object reconstruction, ECCV 2016.
[2] Fan et.al., A point set generation network for 3d object reconstruction from a single image, CVPR 2017.
[3] Lorensen et.al., Marching cubes: A high resolution 3d surface construction algorithm, SIGGRAPH 1987.
[4] Bernardini et.al., The ball-pivoting algorithm for surface reconstruction, IEEE Trans. Vis. Comput. Graph 1999.
[5] Kato et.al., Neural 3d mesh renderer, CVPR 2018.
[6] Groueix et.al., Atlasnet: A papier-maˆche´ approach to learning 3d surface generation, CVPR 2018.
[7] Mescheder et.al., Occupancy networks: Learning 3d reconstruction in function space, CVPR 2019.

Pixel2Mesh



Pixel2Mesh: 3D Mesh Model Generation via Image Guided Deformation

Nanyang Wang⋆, Yinda Zhang⋆, Zhuwen Li⋆, Yanwei Fu⋆, Hang Yu, Wei Liu, Xiangyang Xue and Yu-Gang Jiang†

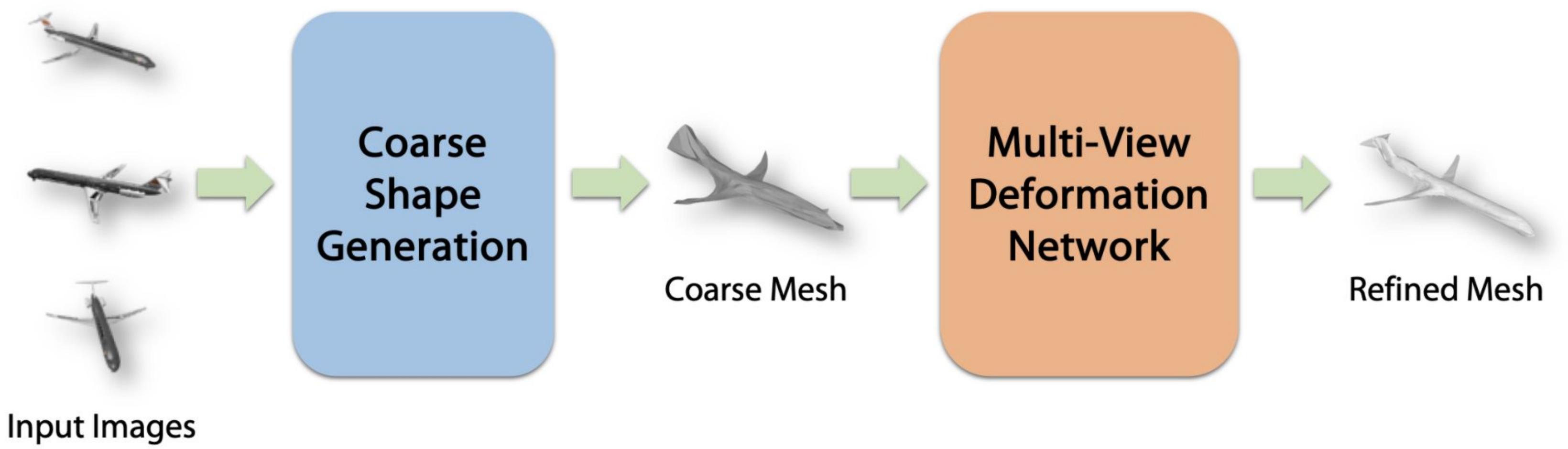
Abstract—In this paper, we propose an end-to-end deep learning architecture that generates 3D triangular meshes from single color images. Restricted by the nature of prevalent deep learning techniques, the majority of previous works represent 3D shapes in volumes or point clouds. However, it is non-trivial to convert these representations to compact and ready-to-use mesh models. Unlike the existing methods, our network represents 3D shapes in meshes, which are essentially graphs and well suited for graph-based convolutional neural networks. Leveraging perceptual features extracted from an input image, our network produces the correct geometry by progressively deforming an ellipsoid. To make the whole deformation procedure stable, we adopt a coarse-to-fine strategy, and define various mesh/surface related losses to capture properties of various aspects, which benefits producing the visually appealing and physically accurate 3D geometry. In addition, our model by nature can be adapted to objects in specific domains, e.g., human faces, and be easily extended to learn per-vertex properties, e.g., color. Extensive experiments show that our method not only qualitatively produces the mesh model with better details, but also achieves the higher 3D shape estimation accuracy compared against the state-of-the-arts.

Index Terms—3D shape generation, graph convolutional neural network, mesh reconstruction, coarse-to-fine, end-to-end framework.

Pixel2Mesh++: Multi-View 3D Mesh Generation via Deformation

Chao Wen Yinda Zhang Zhuwen Li Yanwei Fu

Fudan University Google LLC Nuro, Inc.



https://walsvid.github.io/Pixel2MeshPlusPlus/

A pipeline to generate virtual content...

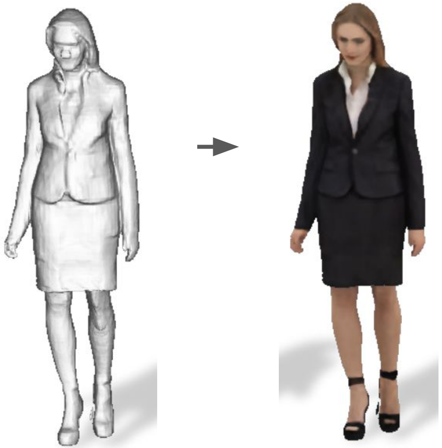
Generate 3D

Guo et.al., The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting, SIGGRAPH 2019.

Manipulate 3D

Wang et.al., Neural Pose Transfer by Spatially Adaptive Instance Normalization, CVPR 2020.

Render 3D



Saito et.al., PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization, ICCV 2019.

A pipeline to generate virtual content...

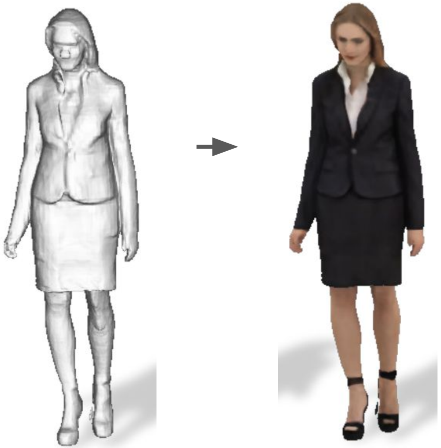
Generate 3D

Guo et.al., The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting, SIGGRAPH 2019.

Manipulate 3D

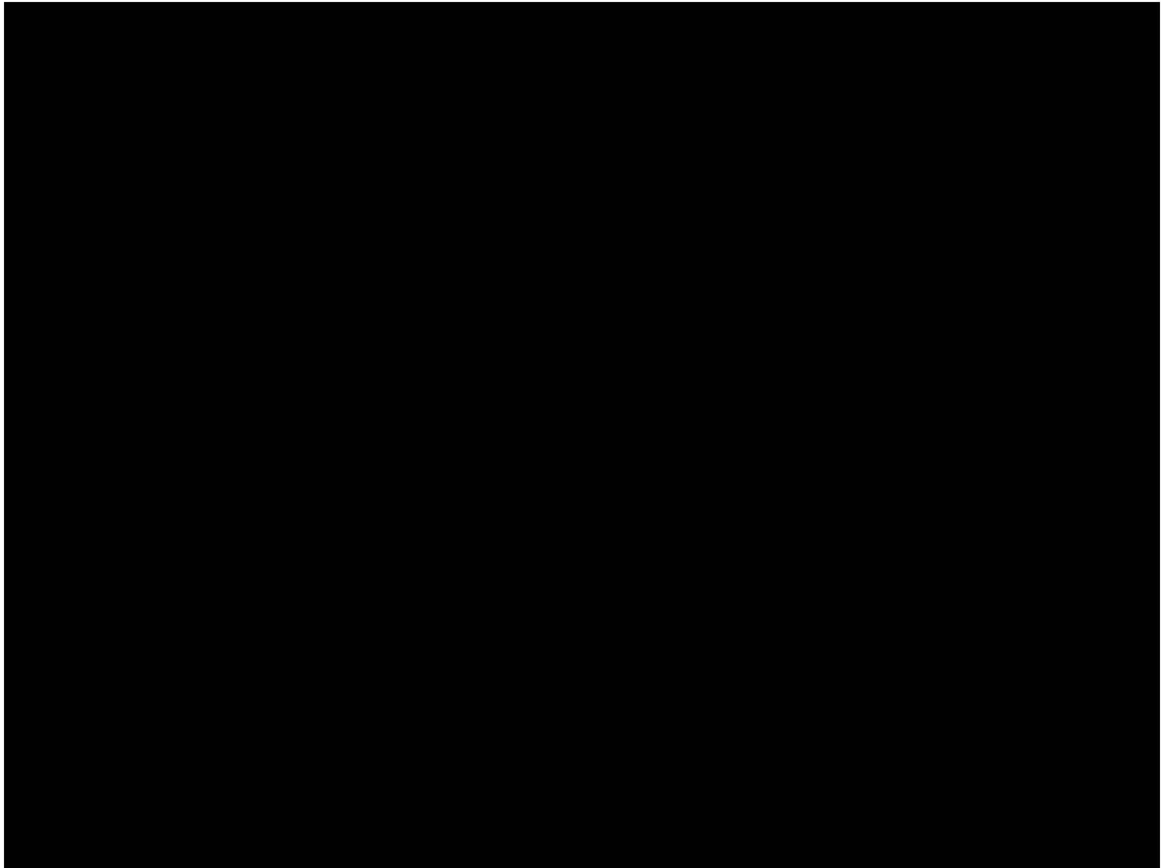
Wang et.al., Neural Pose Transfer by Spatially Adaptive Instance Normalization, CVPR 2020.

Render 3D



Saito et.al., PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization, ICCV 2019.

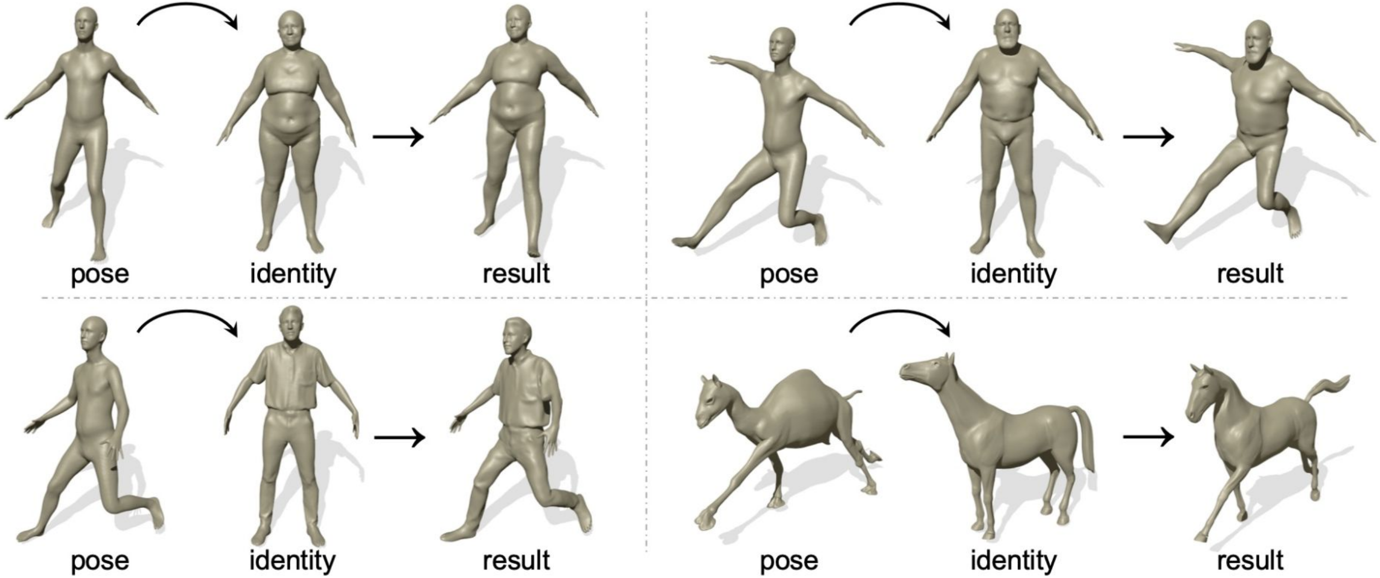
Shape Manipulation -- Pose Transfer



Neural Pose Transfer by Spatially Adaptive Instance Normalization

Jiashun Wang Chao Wen Yanwei Fu Haitao Lin Tianyun Zou Xiangyang Xue Yinda Zhang

Fudan University Google Research



https://jiashunwang.github.io/Neural-Pose-Transfer/

A pipeline to generate virtual content...

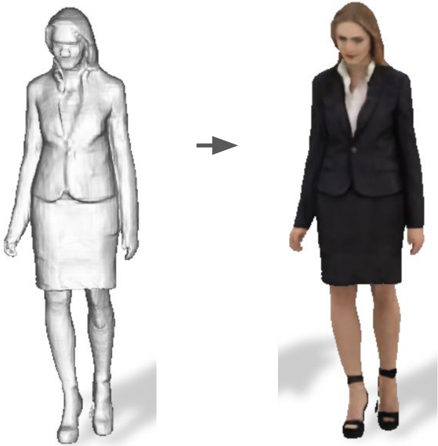
Generate 3D

Guo et.al., The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting, SIGGRAPH 2019.

Manipulate 3D

Wang et.al., Neural Pose Transfer by Spatially Adaptive Instance Normalization, CVPR 2020.

Render 3D



Saito et.al., PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization, ICCV 2019.

A pipeline to generate virtual content...

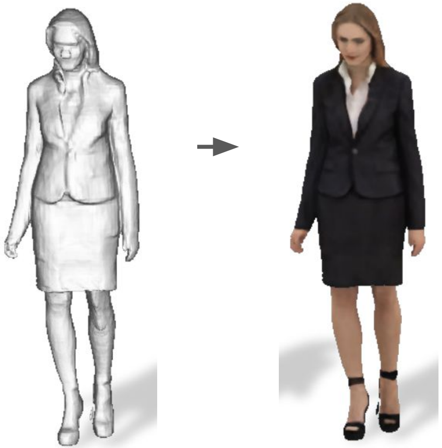
Generate 3D

Guo et.al., The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting, SIGGRAPH 2019.

Manipulate 3D

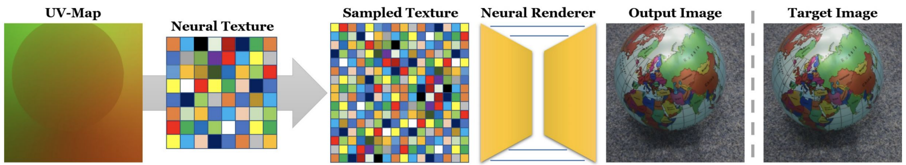
Wang et.al., Neural Pose Transfer by Spatially Adaptive Instance Normalization, CVPR 2020.

Render 3D

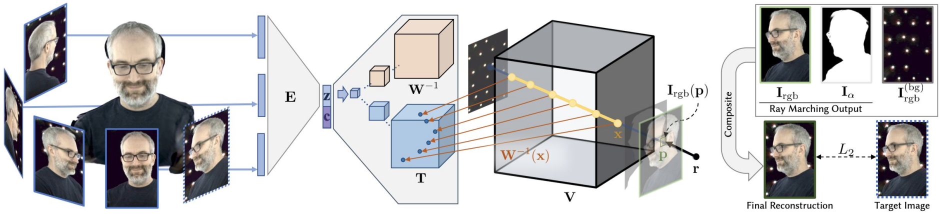


Saito et.al., PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization, ICCV 2019.

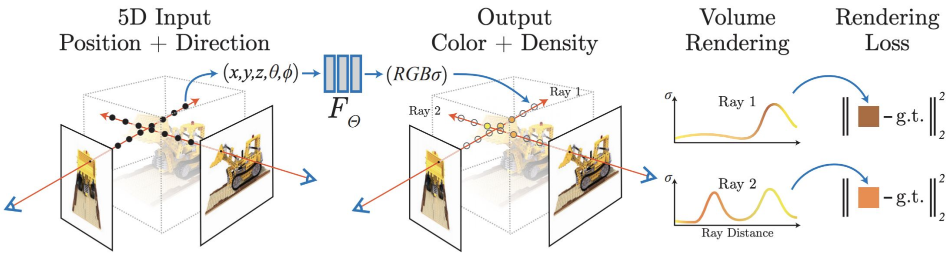
Neural Rendering



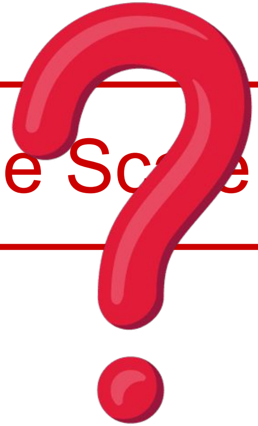
Rely on texture map.



Rely on volume.



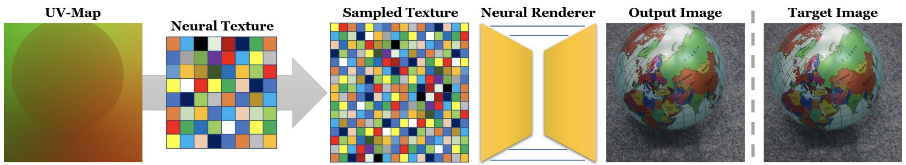
Rely on implicit representation.



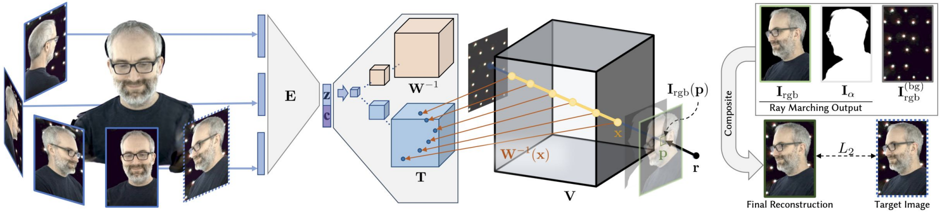
For Large Sc e Scene

Thies et.al., Deferred Neural Rendering: Image Synthesis using Neural Textures, SIGGRAPH 2019.
Lombardi et.al., Neural Volumes: Learning Dynamic Renderable Volumes from Images, SIGGRAPH 2019.
Mildenhall et.al., NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, arXiv:2003.08934

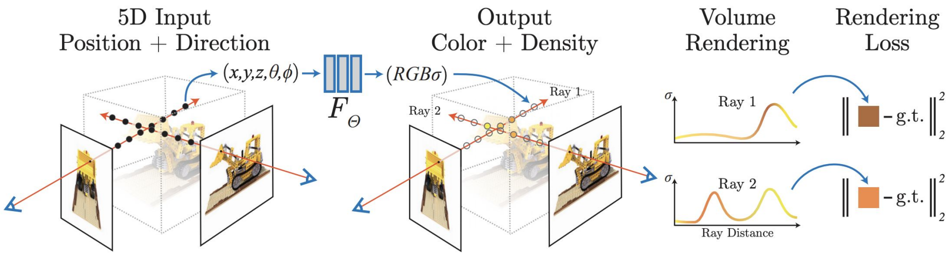
Neural Rendering



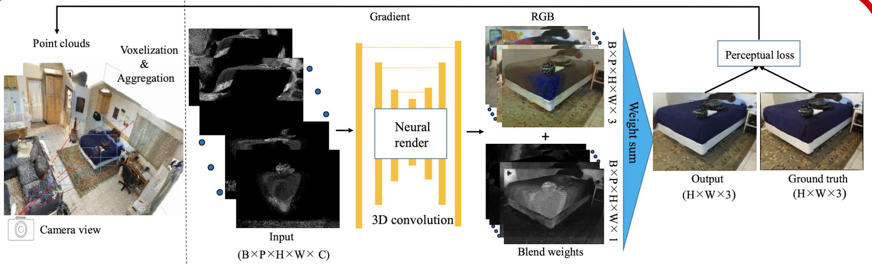
Rely on texture map.



Rely on volume.



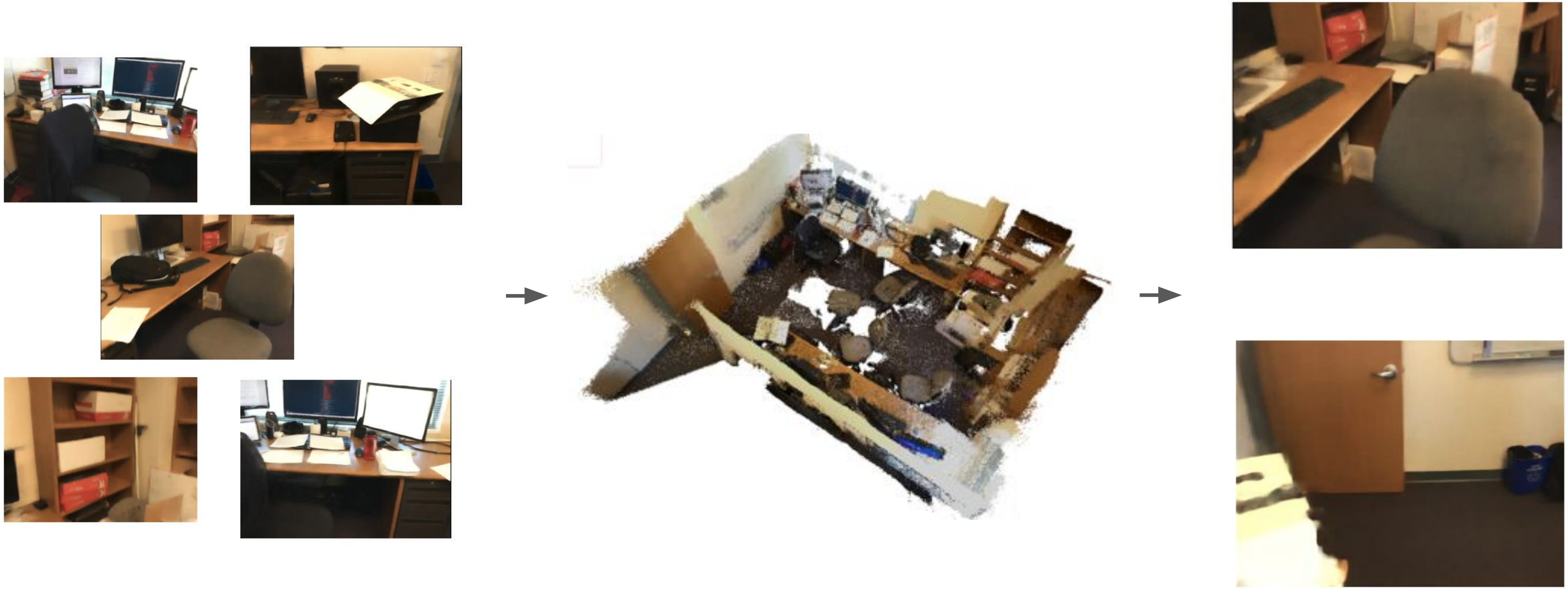
Rely on implicit representation.



Rely on point cloud.

Thies et.al., Deferred Neural Rendering: Image Synthesis using Neural Textures, SIGGRAPH 2019.
Lombardi et.al., Neural Volumes: Learning Dynamic Renderable Volumes from Images, SIGGRAPH 2019.
Mildenhall et.al., NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, arXiv:2003.08934

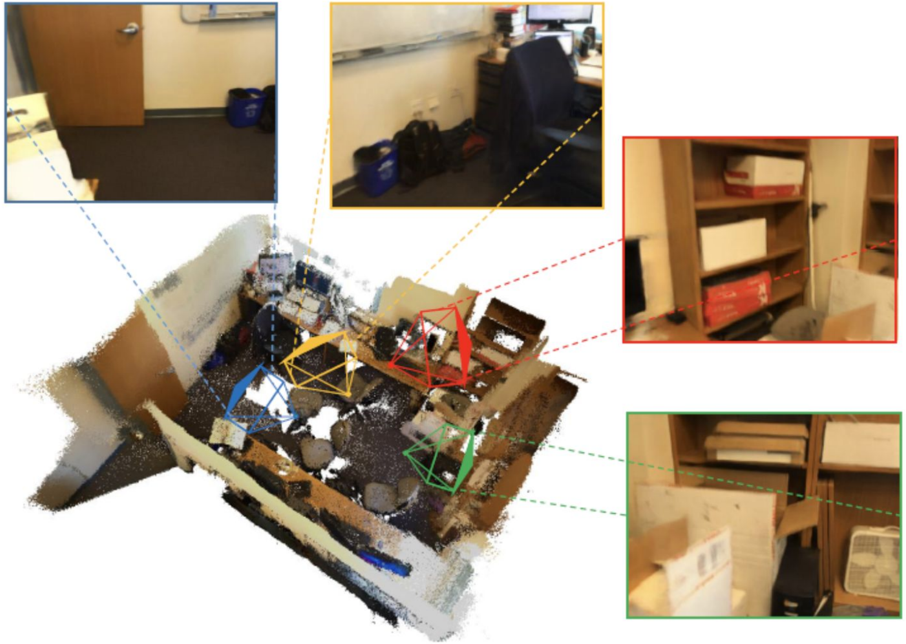
Neural Point Cloud Rendering



RGB-D Scans | Point Cloud | Rendering

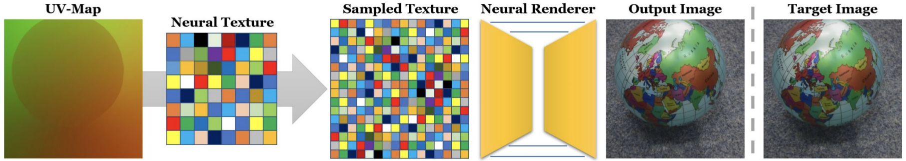
Neural Point Cloud Rendering via Multi-Plane Projection

Peng Dai* Yinda Zhang* Zhuwen Li* Shuaicheng Liu Bing Zeng
University of Electronic Science and Technology of China
Google Research
Nuro.Inc

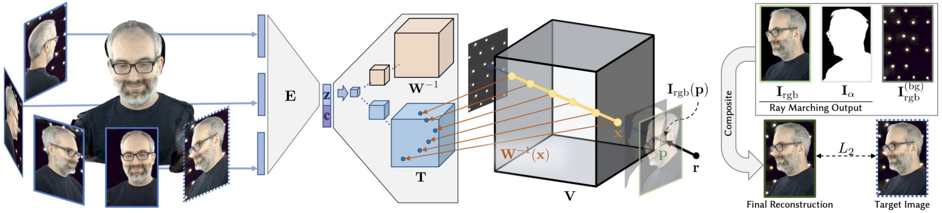


https://daipengwa.github.io/NeuralPointCloudRendering_ProjectPage/

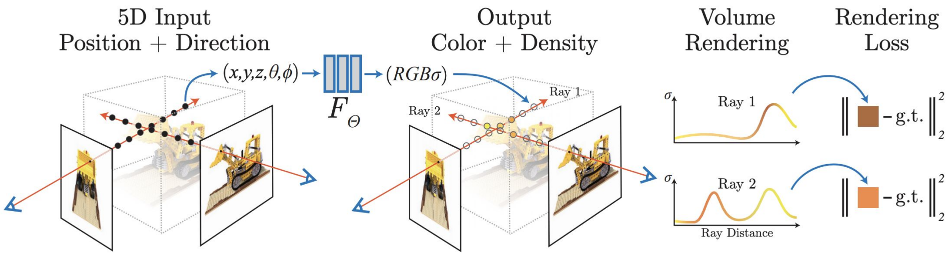
Neural Rendering



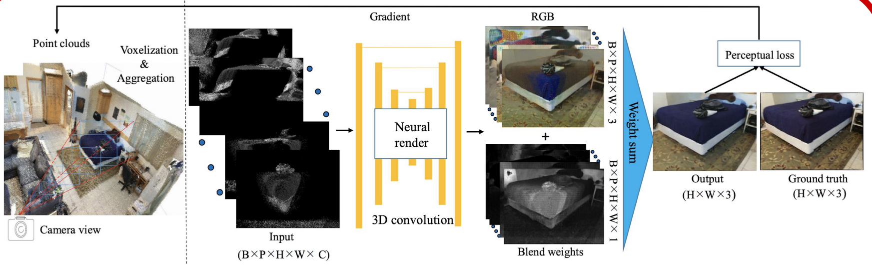
Rely on texture map.



Rely on volume.



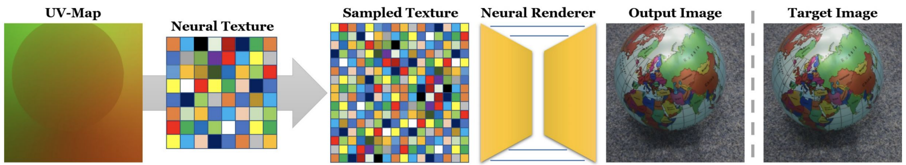
Rely on implicit representation.



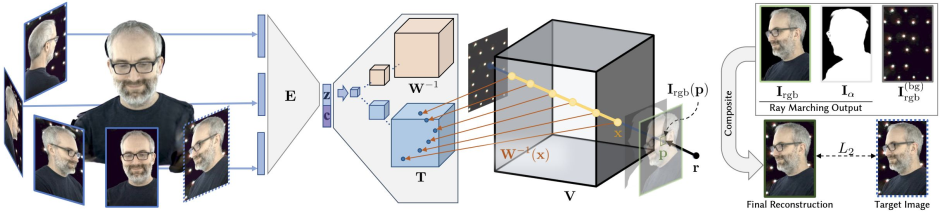
Rely on point cloud.

Thies et.al., Deferred Neural Rendering: Image Synthesis using Neural Textures, SIGGRAPH 2019.
Lombardi et.al., Neural Volumes: Learning Dynamic Renderable Volumes from Images, SIGGRAPH 2019.
Mildenhall et.al., NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, arXiv:2003.08934

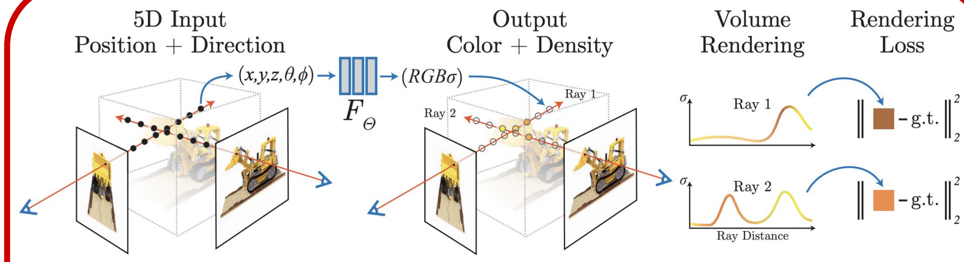
Neural Rendering



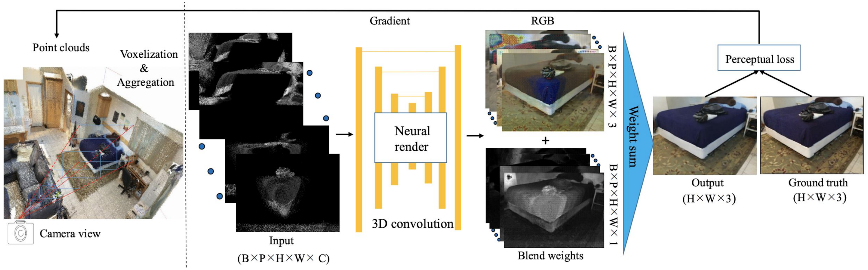
Rely on texture map.



Rely on volume.



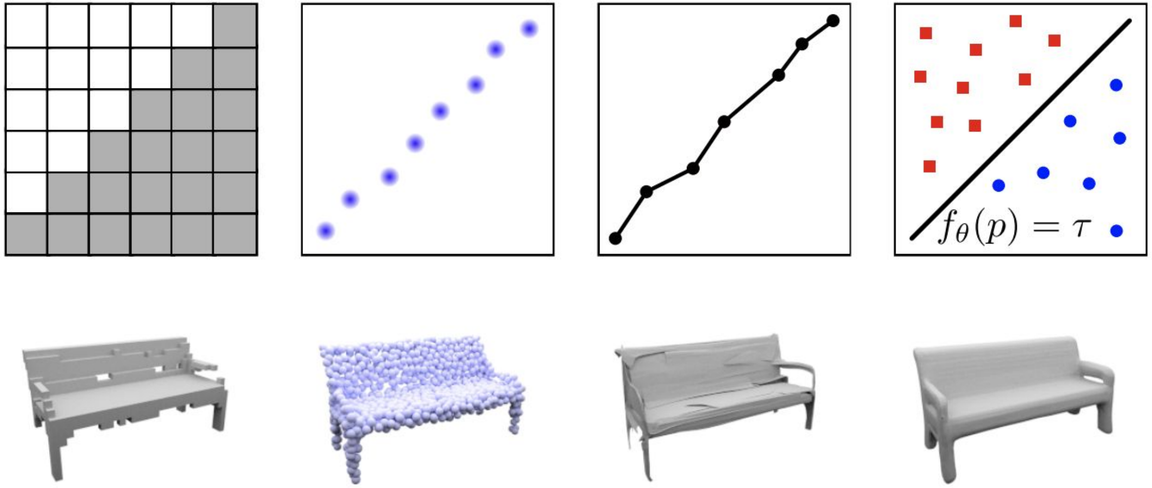
Rely on implicit representation.

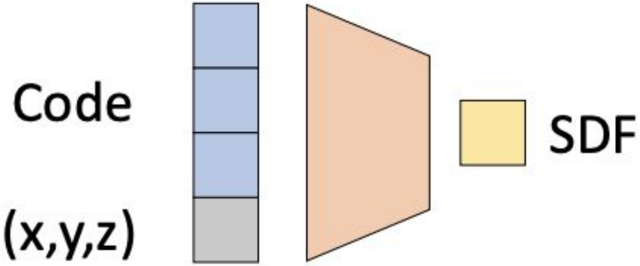


Rely on point cloud.

Thies et.al., Deferred Neural Rendering: Image Synthesis using Neural Textures, SIGGRAPH 2019.
Lombardi et.al., Neural Volumes: Learning Dynamic Renderable Volumes from Images, SIGGRAPH 2019.
Mildenhall et.al., NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, arXiv:2003.08934

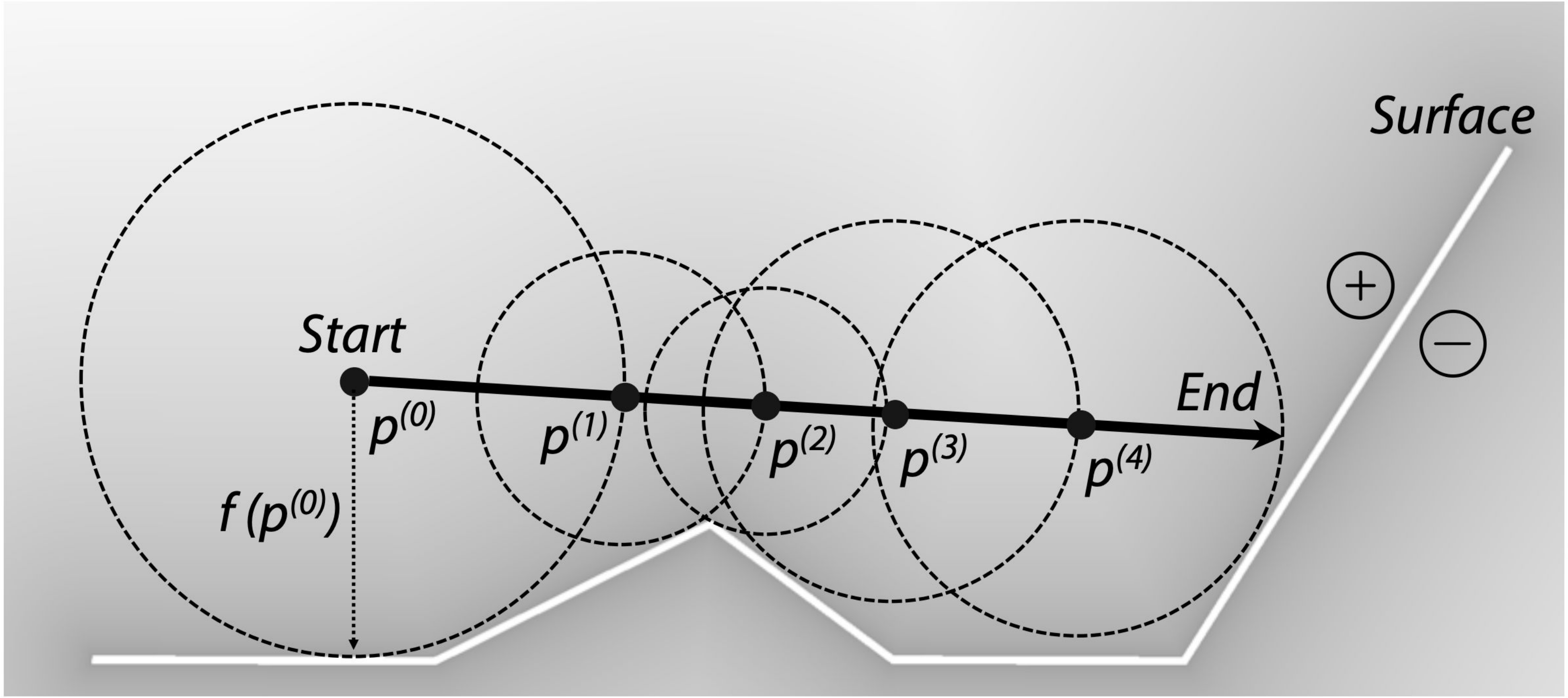
Implicit Function for 3D





Park et.al., DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR 2019.
Mescheder et.al., Occupancy Networks: Learning 3D Reconstruction in Function Space, CVPR 2019.

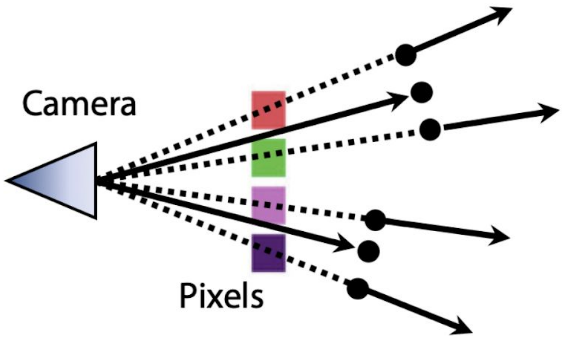
Render Implicit Function



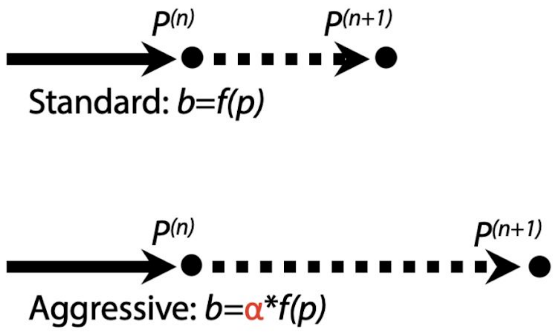
Hart et.al., Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces. The Visual Computer 1996.

Render Deep Implicit Function

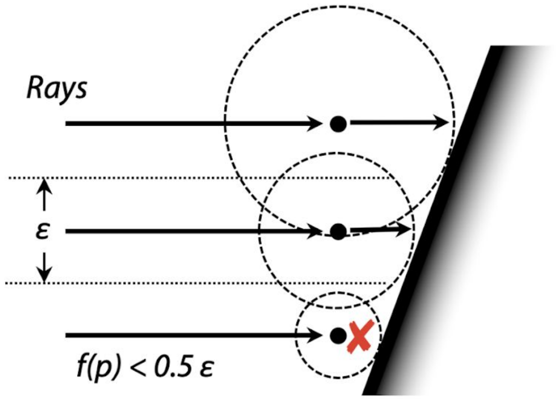
- Extremely time consuming
  - Too many queries for the rendering process.
  - Unroll multiple times for backpropagation.
  - Differentiable.



Coarse-to-fine

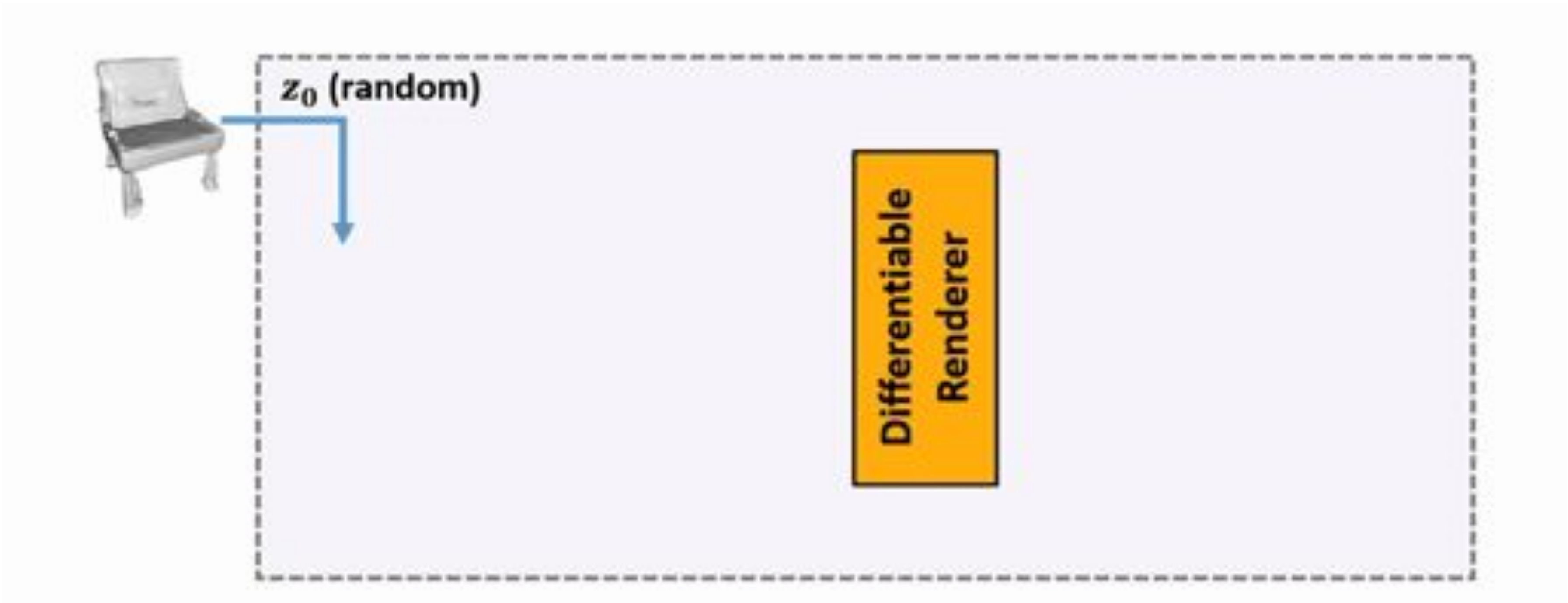


Aggressive Marching

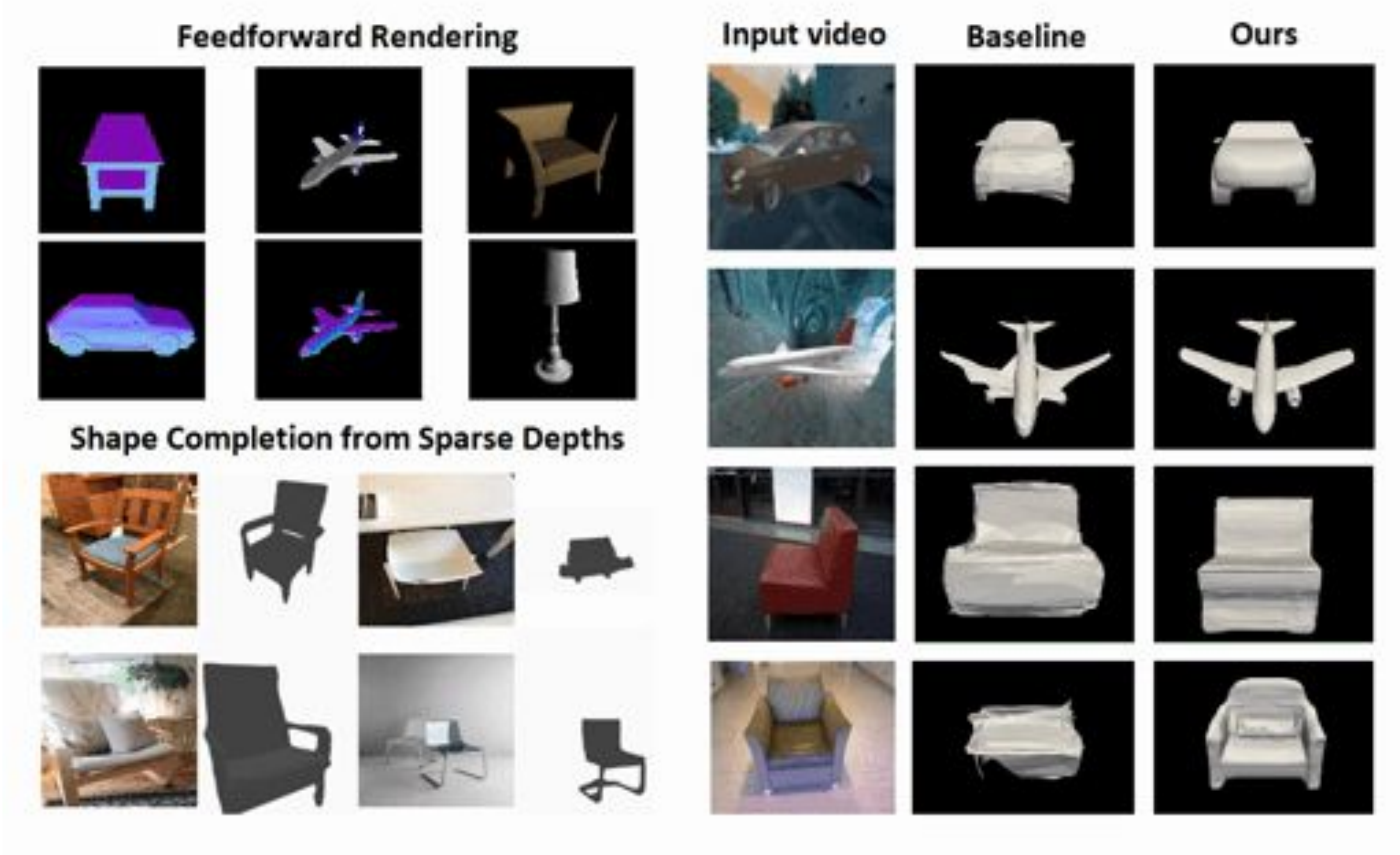


Converge Criteria

Render Deep Implicit Function



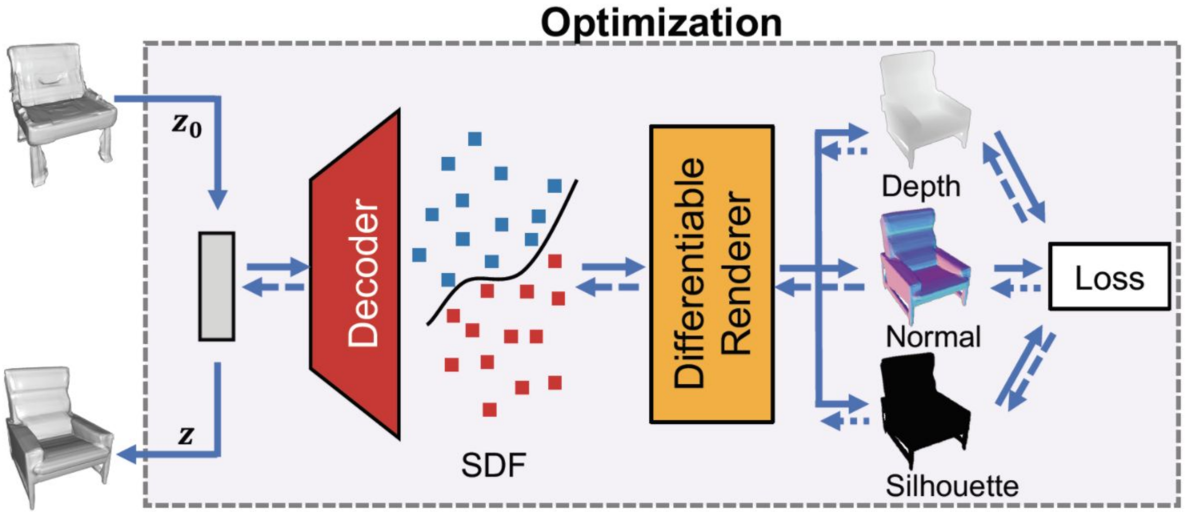
Render Deep Implicit Function



DIST: Rendering Deep Implicit Signed Distance Function with Differentiable Sphere Tracing

Shaohui Liu[1,3] Yinda Zhang[2] Songyou Peng[1] Boxin Shi[4] Marc Pollefeys[1,5] Zhaopeng Cui[1]

[1]ETH Zurich [2]Google [3]Tsinghua University [4]Peking University [5]Microsoft



http://b1ueber2y.me/projects/DIST-Renderer/

A pipeline to generate virtual content...

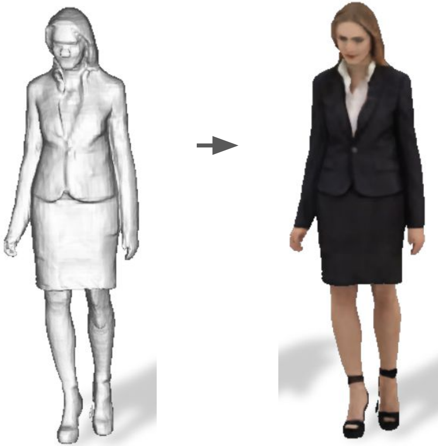
Generate 3D

Guo et.al., The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting, SIGGRAPH 2019.

Manipulate 3D

Wang et.al., Neural Pose Transfer by Spatially Adaptive Instance Normalization, CVPR 2020.

Render 3D



Saito et.al., PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization, ICCV 2019.

Thanks!