International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

3D Mesh Reconstruction from Single 2D Image Using DBSCAN and CNN Architecture.

[1]Akshay Marathe, Harshil Lakkad, Athava Undale, Dr. K. Nandhini, [2]Dr. Shilpa Gite, Dr. Smita Mahajan

Submitted: 27/10/2023 **Revised**: 16/12/2023 **Accepted**: 24/12/2023

Abstract: One of the most intriguing challenges in this domain is the reconstruction of 3-D objects or scenes from a single 2D image. 3D Reconstruction from a single-view image aims to reconstruct the 3D object from a single-colored image. Previously, the 3D Reconstruction was done using multiple images taken from different angles. But it has limitations in accessibility. Generating a 3D model from a 2D image has broad applications in various industries such as Augmented reality (AR), Virtual reality (VR), Robotics, Video games, and Medical Imaging. We have broken the process into two parts, in the first part we generate a 3D point cloud from a 2D image. For the next part, we then generate a 3D mesh from the point cloud map. A CNN is later used to further optimize the mesh. In the result and conclusion, we have generated an optimized mesh but the end output has not reached the desired accuracy. This accuracy can be increased by creating a Depth-map and view-point angle prior to calculating a point cloud.

Index Terms: *3D-Reconstruction, Computer Vision, 3D Mesh, 3D from 2D, Point Cloud.*

1. Introduction

In recent years, with the advancement in technology (both hardware and software), the use of 2D images for visual representation is becoming outdated. With the shift in trend from machine efficiency to a more human-centric experience, the need for 3D visual representation becomes more and more prominent. To solve this problem one of the most popular technique in Computer Vision is to reconstruct a 3D model from 2D image or images. By using multiple views of a single object to calculate the depth field a 3D model can be reconstructed. This process has its own limitations in accessibility as it requires multiple views and their viewing angle, which requires special equipment. A noveler approach to the process of 3D reconstruction is to use a single image to create a 3D mesh and then use the mesh to refine a 3D model.

Single-view 3D mesh reconstruction aims to reconstruct a 3D mesh using a single view of an object. The process faces some challenges, as for a single view of an object, there can be multiple 3D views that can correspond to that viewing angle. Recently, Deep learning methods have made major progress in this field, it can learn to extract crucial features from the images and use these features to predict 3D meshes. A common approach to Single-view mesh reconstruction is to use a two- stage pipeline. First, the model predicts the depth map from the image and then uses the depth map to reconstruct the 3D mesh.

Another approach to 3D mesh reconstruction uses a more direct approach as it directly predicts the 3D mesh from the input image without first predicting a depth map. For the 3D image reconstruction, there are various algorithms and some of them that are most widely used are multi-View stereo, Structure from motion(SFM), LiDAR, etc. Multi-view stereo algorithm requires various images of the same object with various angles so that we can see the object through various light effects. The field of 3D reconstruction is witnessing a huge transformation, driven by the influence of cutting-edge technologies and the ever-expanding range of its applications. The motivation behind this paper lies in the recognition of the pivotal role that 3D reconstruction plays in shaping the future of various fields and domains.

Despite the number of challenges that are present, the practical applications of this technique are too broad. It has implementations in various industries such as Augmented reality (AR), Virtual reality (VR), Robotics, Video games, and Medical Imaging. In robotics, the ability to perceive its surroundings from its cameras and interpret it into a scene which will help it to interact with its surroundings and plan future movements too. It can also have applications in Archaeology and can be used to recreate historical monuments. In the field of Medicine, it can be used to enhance medical diagnostics by providing medical experts better visual representation compared to traditional ones.

We further discuss the topic of single-view 3D mesh reconstruction in the following sections. Through this

[1]*Artificial Intelligence and Machine Learning Department, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India.*
[2]*Symbiosis Centre for Applied AI, Symbiosis International (Deemed University), Pune, India.*
Akkim17@gmail.com, Harshilpatel98981@gmail.com, Atharvaundale60@gmail.com, nandhini.k@sitpune.edu.in, shilpa.gite@sitpune.edu.in, smita.mahajan@sitpune.edu.in
Corresponding Author- Dr. Smita Mahajan

paper, we hope to shed some light on the potential and application into this field and provide some direction for future research.

2. Related Works

The objective of single view 3D reconstruction is to generate a 3D object from a single, colored, or RGB image. This is a challenging task in and of itself when a single image might be expressed with multiple 3D views or shapes.

The need for 3D model generation is an age-old challenging requirement for many fields. To accurately reconstruct it requires integration of strong geometric priors of our world [1]. Historically, this process was achieved using shape-from-shading [7], [8], [9]. Recently, learning-based approach have become a trend since they are more accessible and robust. The use of machine learning for 3D model reconstruction is achieved by various techniques. Such as voxel based [2], [3], [4], [5], [6], mesh

based [11], [12], [13], [14], [15], [16], [17], [18], [19], point cloud [23], [24], [25],

and implicit function [20], [21], [22] based techniques. Among those techniques, mesh based framework is the we have implemented in our work.

The model that is used by researchers is Structure from Motion (SfM) algorithms and Multi View Stereo (MVS) algorithms are faster than other toolkits that's why they propose effective methods for generating 3D models of objects, constructions, and environments.[30][32]. Neural style transfer and customized CNN are latest techniques in this domain [33][34].

The Long Short-Term Memory (LSTM) networks and single-view 3D reconstruction using CNNs model used by researcher for 3D-R2N2 model for Single and Multi- view 3D Object Reconstruction and paper provide comparison between 3D-R2N2 model and MVS(Multi View Stereo) algorithm and conclusion was 3D-R2N2 model is better for 3D reconstruction.[2]

3. Methodology

In figure 1, the input consists of a pulley image, which is loaded as a 2D image. Following this, an image resizing function is applied to adjust its dimensions. Subsequently, depth estimation is calculated on the resized image, leading to the generation of a Point Cloud representation of the image. This Point Cloud serves as the foundation for Mesh generation, resulting in the creation of a three-dimensional Mesh. For visualization of this mesh, MeshLab software is employed.

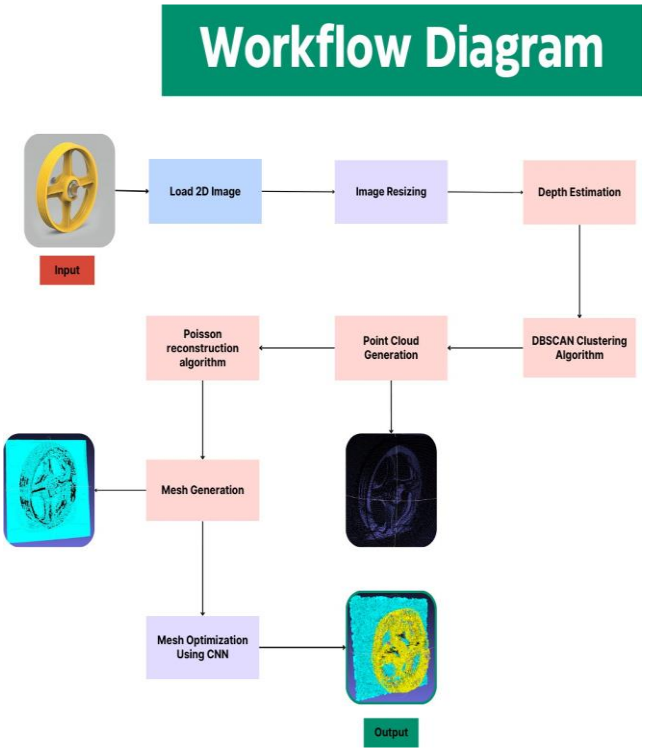

Fig 1 Workflow Diagram

3.0.1. 3D point cloud generation from 2D images

One of the most important steps in creating a 3D point cloud from a set of 2D images is to turn each image into a NumPy array and then store them in the empty list. This initial step lays the groundwork for further image data manipulation and analysis. Next, a crucial part of the algorithm is to create an empty Open3D Point Cloud object which will act as the canvas on which the 3D points that will be aggregated and arranged from the 2D images.

The second stage proceeds in a step-by-step iteration through every 2D NumPy array for the point cloud generation. In order to identify the spatial information contained in the images, the extensive extraction of 2D coordinates corresponding to non-black pixels during this iterative process is done. The extracted coordinates are subjected to a column swap operation in order to seamlessly integrate these coordinates into the Open3D framework. By aligning the coordinates in this manner, the Y and X axes are oriented in accordance with the Open3D format.

The measurement of depth estimation is required to build a complete three- dimensional model which is derived by specifically gathering values from an image's red channel. This assumption is based on the concept that the red channel provides as a substitute for depth mapping in this scenario. Further the organization of the combination of the extracted 2D. coordinates and depth values, resulting in an integrated 3D point cloud. Concentrating the points inside the chosen geographic coordinate system is a crucial refinement step. By using DBSCAN clustering to these point clouds, it becomes easier to identify physically homogeneous sets and helps create a more accurate 3D point cloud. Important parameters, like minimum points and epsilon can be adjusted according to the given image data.

In the last step of this complex process includes collection of finely tuned points and then sequentially merges them into the Open3D Point Cloud object. In addition to capturing the unique geometrical details of every image, this accumulative combination of points creates a comprehensive three-dimensional model that captures the spatial connections which are present in the original set of 2D images. The final result of this method is the all-inclusive Open3D Point Cloud, which is a concrete and comprehensible representation of the underlying three-dimensional framework contained in the 2-dimensional picture. In below figure 2 the point cloud is generated using this process and the generation of point cloud in 3D space can be seen. To view this 3D .ply file the software called MeshLab is used.

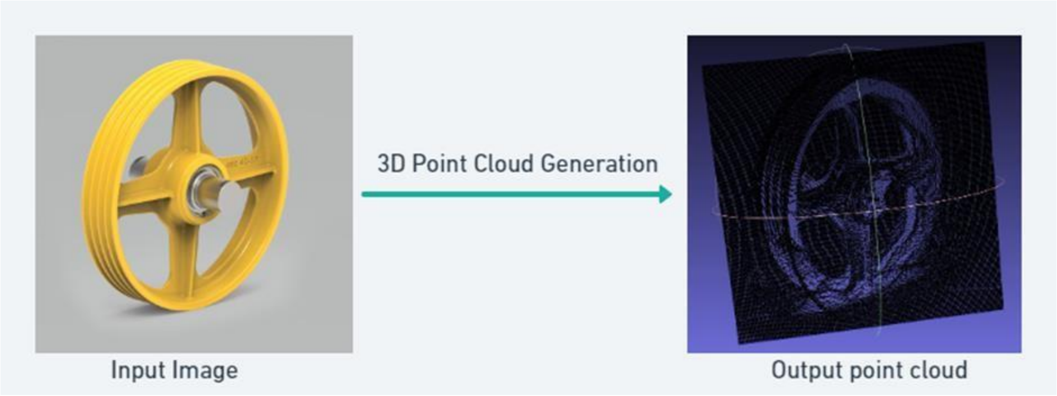

Fig 2 *3D Point cloud Generation*

3.0.2. 3D point cloud to 3D Mesh

An ordered set of computational phases is involved in creating a 3-dimensional (3D) mesh from a 3D point cloud. These procedures are intended to convert a collection of distinct points throughout space into a more organized and substantial mesh representation. Usually, the first step in the procedure is to acquire a 3D point cloud. Each point in the point cloud represents a 3 dimensional positional coordinate, resulting in an ordered collection of data points.

Upon obtaining the point cloud, the subsequent task involves computing normals for each individual point. Normals offer valuable information about surface orientations at each point within the point cloud, facilitating the reconstruction of a more coherent, meaningful, and visually appealing mesh. This process is pivotal for capturing the geometric characteristics of underlying surfaces, where the estimate_normals method employs a search approach dependent on a specified radius and maximum nearest neighbors to calculate these normals.

Following the computation of normals, the process of mesh reconstruction is initiated. The Poisson reconstruction algorithm is a frequently utilized method for this. Using the input point cloud as a starting point, this algorithm creates a triangle mesh and estimates the underlying surface using the Poisson equation. The function's depth parameter regulates the reconstruction's level of detail, enabling modifications in accordance with the particular needs of the application. The reconstruction process generates a three-dimensional mesh, which is seen as a group of interlocked triangles that roughly correspond to the surfaces that the original point cloud represents. Depending on the uses for this mesh, additional computation or analysis may be required. In below figure 3 the 3D Mesh is generated using this process and output can be seen very close to the original image.

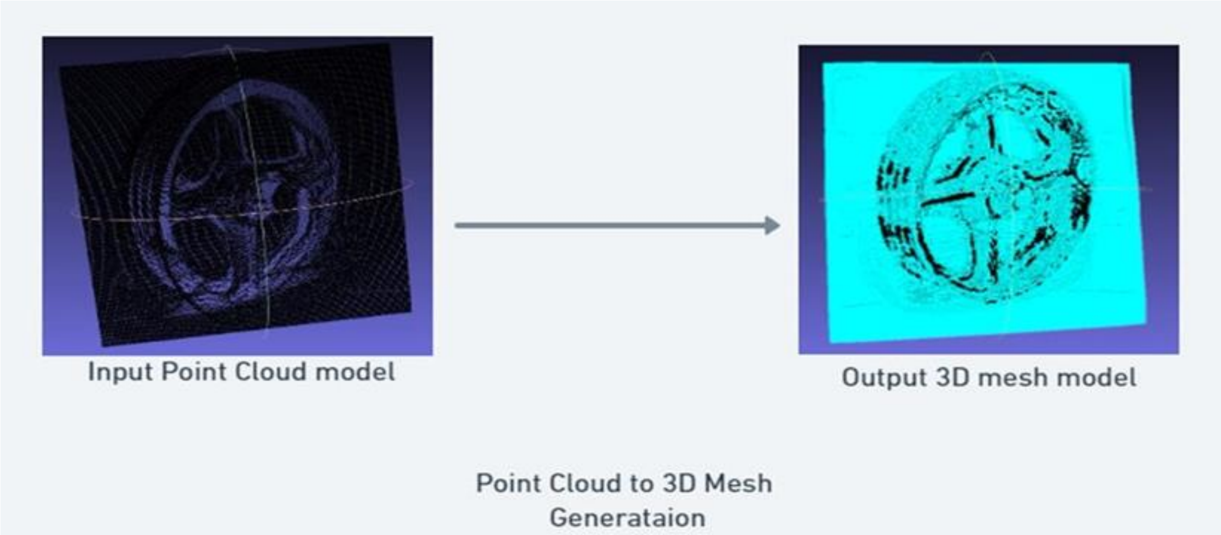

Fig 3 *3D Mesh Generation*

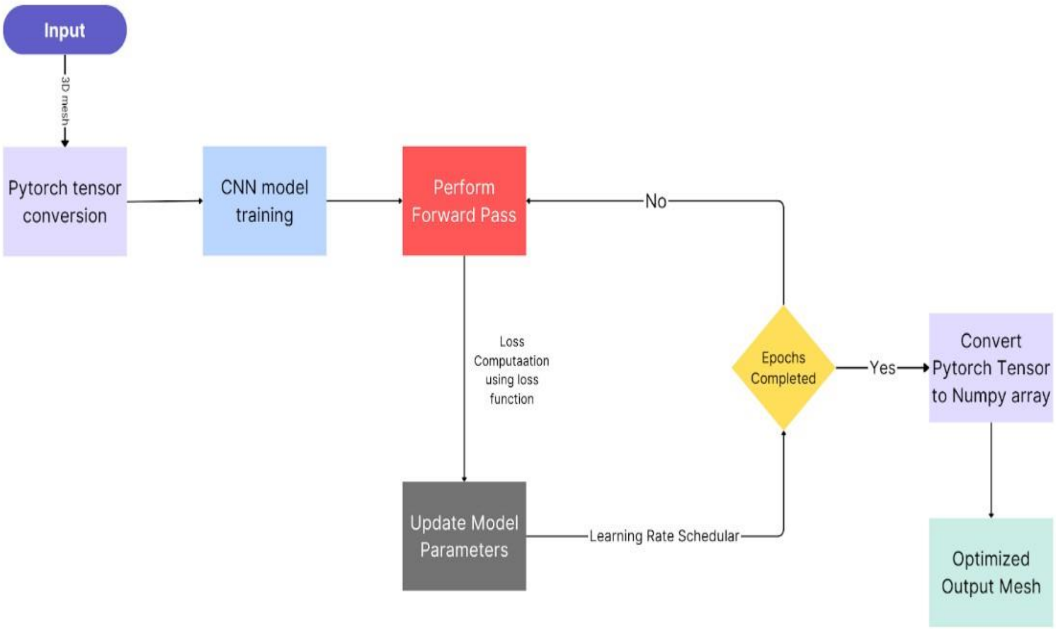

Fig 4 *3D Optimization*

For the optimization of 3D mesh as mentioned in figure 4 a novel convolutional neural network (CNN) architecture is used. The Improved Mesh Generator CNN is introduced, comprising three convolutional layers and two fully connected layers. The input for architecture is a 3D mesh which is converted to a PyTorch tensor, and the model undergoes training for 50 epochs. During each epoch, a forward pass is performed, computing the loss using the MSE loss function. The model is trained to learn the intricate patterns of 3D shapes, allowing it to generate detailed and realistic meshes. The optimization process involves the use of the Adam optimizer with a learning rate of 0.001 and a learning rate scheduler, which dynamically adjusts the learning rate during training. The backward pass and optimization steps follow, updating the model's parameters through gradient descent. The chosen loss function is Mean Squared Error (MSE), measuring the disparity between the generated output mesh and the input 3D mesh. The training progress, including the epoch and loss, is monitored, and the final output mesh is obtained by converting the PyTorch tensor to a NumPy array.

4. Result

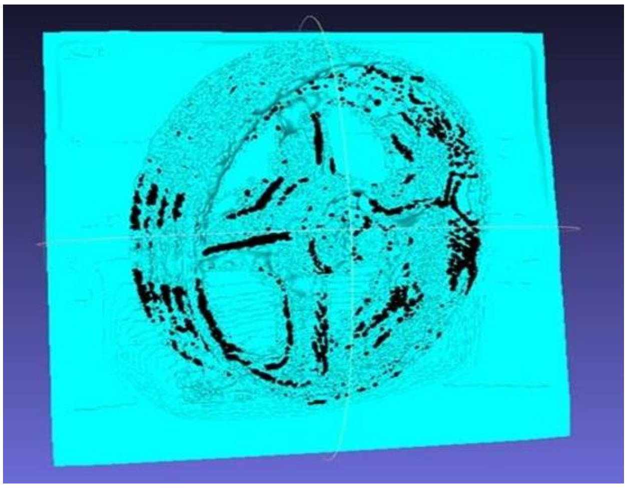

Fig 5 *3D Mesh before optimization*

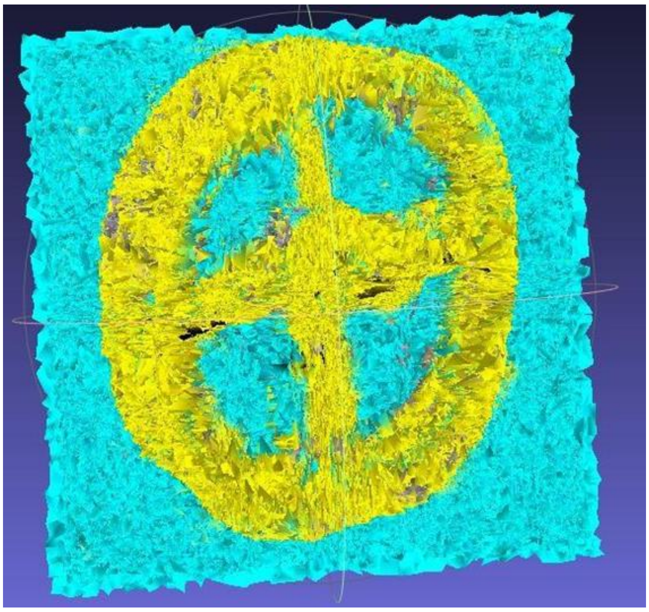

Fig 6 *3D Mesh after optimization*

As Figure 5 illustrate that the 3D mesh that is generated by using 3D point clouds and DBSCAN algorithm appears substantially similar to the 2D input image. But as can be seen in the previously described in figure 6, the 3D model created and refined solely using the CNN design has flaws in how well it represents the input 2D image. The reason for this variation could be that the differential rendering technique was not included in the mesh generation process, and the only source of information used was a single-view image; the generated mesh exhibits imperfections in accurately representing the input 2D image.

5. Conclusions

For this paper, we have introduced the framework for the reconstruction of single-view 3D mesh. Here, we have broken down the process into 3 parts, in the first phase generation of a 3D point cloud from a 2D image process is done. In the second Phase creation of a 3D mesh from the 3D point cloud process is done. And in the third and final phase, the optimization of 3D mesh for a better model is done. This limitation underscores the importance of considering multi-view approaches for more comprehensive 3D reconstructions along with the differential rendering in future iterations of the methodology. The findings of our study suggest that reduction of noise and the need for careful computation in order to produce an accurate depth map for the image has to be done perfectly to achieve the desired accuracy level. However, the obtained precision is great enough to make the process of creating a 3D model easier. Using a Deep Learning (DL) model, such as a Convolutional Neural Network (CNN), for depth map estimation and feature extraction is suggested as a potential direction for future research. Furthermore, there is potential to improve the 3D model creation process with the application of a

Generative Model, as demonstrated by Generative Adversarial Networks (GANs) for the single image dataset.

5. Future Scope

From the result and conclusion, we know that 3D Mesh can be further optimized. We can provide a calculated Depth-Map and Viewing Angle to the Point cloud generator as input. The mesh optimization can also be improved by further training it to handle noise better. A Generative Neural Network like GANs can be implemented to generate a 3D model from the optimized 3D mesh.

References

[1] Yang, Xianghui, Guosheng Lin, and Luping Zhou. "Single-view 3D Mesh Reconstruction for Seen and Unseen Categories." *IEEE Transactions on Image Processing* (2023).

[2] Choy, Christopher B., Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. "3d-r2n2: A unified approach for single and multi-view 3d object reconstruction." In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, pp. 628-644. Springer International Publishing, 2016.

[3] Mescheder, Lars, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. "Occupancy networks: Learning 3d reconstruction in function space." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4460-4470. 2019.

[4] Richter, Stephan R., and Stefan Roth. "Matryoshka networks: Predicting 3d geometry via nested shape layers." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1936-1944. 2018.

[5] Popov, Stefan, Pablo Bauszat, and Vittorio Ferrari. "Corenet: Coherent 3d scene reconstruction from a single rgb image." In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pp. 366-383. Springer International Publishing, 2020.

[6] Wu, Jiajun, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. "Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling." *Advances in neural information processing systems* 29 (2016).

[7] Durou, Jean-Denis, Maurizio Falcone, and Manuela Sagona. "Numerical methods for shape-from-shading: A new survey with benchmarks." *Computer Vision and Image Understanding* 109, no. 1 (2008): 22-43.

[8] Horn, Berthold KP. "Shape from shading: A method for obtaining the shape of a smooth opaque object from one view." (1970).

[9] Zhang, Ruo, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. "Shape-from-shading: a survey." *IEEE transactions on pattern analysis and machine intelligence* 21, no. 8 (1999): 690-706.

[10] M. Tatarchenko, S. R. Richter, R. Ranftl, Z. Li, V. Koltun and T. Brox, "What Do Single-View 3D Reconstruction Networks Learn?," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 3400-3409, doi: 10.1109/CVPR.2019.00352.

[11] Gkioxari, Georgia, Jitendra Malik, and Justin Johnson. "Mesh r-cnn." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9785-9795. 2019.

[12] Liu, Shichen, Tianye Li, Weikai Chen, and Hao Li. "Soft rasterizer: A differentiable renderer for image-based 3d reasoning." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7708-7717. 2019.

[13] Li, Xueting, Sifei Liu, Kihwan Kim, Shalini De Mello, Varun Jampani, Ming-Hsuan Yang, and Jan Kautz. "Self-supervised single-view 3d reconstruction via semantic consistency." In *Computer Vision– ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*, pp. 677-693. Springer International Publishing, 2020.

[14] Kanazawa, Angjoo, Shubham Tulsiani, Alexei A. Efros, and Jitendra Malik. "Learning category-specific mesh reconstruction from image collections." In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 371-386. 2018.

[15] Kato, Hiroharu, Yoshitaka Ushiku, and Tatsuya Harada. "Neural 3d mesh renderer." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3907-3916. 2018.

[16] Chen, Wenzheng, Huan Ling, Jun Gao, Edward Smith, Jaakko Lehtinen, Alec Jacobson, and Sanja Fidler. "Learning to predict 3d objects with an interpolation-based differentiable renderer." *Advances in neural information processing systems* 32 (2019).

[17] Pontes, Jhony K., Chen Kong, Sridha Sridharan, Simon Lucey, Anders Eriksson, and Clinton Fookes.

"Image2mesh: A learning framework for single image 3d reconstruction." In *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part I 14*, pp. 365-381. Springer International Publishing, 2019.

[18] Ye, Yufei, Shubham Tulsiani, and Abhinav Gupta. "Shelf-supervised mesh prediction in the wild." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8843-8852. 2021.

[19] Nie, Yinyu, Xiaoguang Han, Shihui Guo, Yujian Zheng, Jian Chang, and Jian Jun Zhang. "Total3dunderstanding: Joint layout, object pose and mesh reconstruction for indoor scenes from a single image." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 55-64. 2020.

[20] Chen, Zhiqin, and Hao Zhang. "Learning implicit fields for generative shape modeling." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5939-5948. 2019.

[21] Saito, Shunsuke, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. "Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2304-2314. 2019.

[22] Saito, Shunsuke, Tomas Simon, Jason Saragih, and Hanbyul Joo. "Pifuhd: Multi-level pixel- aligned implicit function for high-resolution 3d human digitization." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 84-93. 2020.

[23] Fan, Haoqiang, Hao Su, and Leonidas J. Guibas. "A point set generation network for 3d object reconstruction from a single image." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 605-613. 2017.

[24] Han, Zhizhong, Chao Chen, Yu-Shen Liu, and Matthias Zwicker. "DRWR: A differentiable renderer without rendering for unsupervised 3D structure learning from silhouette images." *arXiv preprint arXiv:2007.06127* (2020).

[25] Chao, Chen, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. "Unsupervised learning of fine structure generation for 3d point clouds by 2d projection matching." *arXiv preprint arXiv:2108.03746* (2021).

[26] Gwak, JunYoung, Christopher B. Choy, Manmohan Chandraker, Animesh Garg, and Silvio Savarese. "Weakly supervised 3d reconstruction with adversarial constraint." In 2017 International Conference on 3D Vision (3DV), pp. 263-272. IEEE, 2017.

[27] Afifi, Ahmed J., Jannes Magnusson, Toufique A. Soomro, and Olaf Hellwich. "Pixel2Point: 3D object reconstruction from a single image using CNN and initial sphere." IEEE Access 9 (2020): 110- 121.

[28] Ugrinovic, Nicolas, Albert Pumarola, Alberto Sanfeliu, and Francesc Moreno-Noguer. "Single-view 3D Body and Cloth Reconstruction under Complex Poses." arXiv preprint arXiv:2205.04087 (2022).

[29] Siddique, Ashraf, and Seungkyu Lee. "Sym3DNet: Symmetric 3D prior network for single-view 3D reconstruction." Sensors 22, no. 2 (2022): 518.

[30] Zhao, Meihua, Gang Xiong, MengChu Zhou, Zhen Shen, and Fei-Yue Wang. "3D-RVP: A method for 3D object reconstruction from a single depth view using voxel and point." *Neurocomputing* 430 (2021): 94-103.

[31] Tatarchenko, Maxim, Stephan R. Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. "What do single-view 3d reconstruction networks learn?." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3405-3414. 2019.

[32] Moons, Theo, Luc Van Gool, and Maarten Vergauwen. "3D reconstruction from multiple images part 1: Principles." *Foundations and Trends® in Computer Graphics and Vision* 4, no. 4 (2010): 287-404.

[33] A. Singh, V. Jaiswal, G. Joshi, A. Sanjeeve, S. Gite and K. Kotecha, "Neural Style Transfer: A Critical Review," in *IEEE Access*, vol. 9, pp. 131583-131613, 2021, doi: 10.1109/ACCESS.2021.3112996.

[34] Mishra, A.; Dharahas, G.; Gite, S.; Kotecha, K.; Koundal, D.; Zaguia, A.; Kaur, M.; Lee, H.-N. ECG Data Analysis with Denoising Approach and Customized CNNs. *Sensors* **2022**, *22*, 1928. https://doi.org/10.3390/s22051928