MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 1111-1116

Graph Convolutional Networks for 3D Scene Reconstruction

Naveen Kumar N , GiribabuKande $, and B Prabhakar Rao #
1 Research Scholar, Department of ECE, JNTUK, Kakinada, 533003, Andhra Pradesh, India, mailnaveenn@gmail.com
$ Professor, ECE Department, VVIT, Nambur, 522509, Andhra Pradesh, India, kgiribabu@vvit.net
 # Professor, ECE Department, JNTUK, Kakinada, 533003, Andhra Pradesh, India, drbprjntu@gmail.com.

Abstract:
We introduce a novel framework for holistic 3D scene Reconstruction from a single RGB image, aiming to jointly infer object geometry, object poses, and global scene layout. Due to the severe ambiguity of monocular 3D perception, prior approaches often struggle to accurately recover object shapes and spatial layouts, particularly in cluttered environments with heavy inter-object occlusions. To address these challenges, we leverage recent advances in deep learning representations. Specifically, we design an image-driven, locally structured graphical network to enhance object-level shape reconstruction, and further improve 3D object pose estimation and scene layout reasoning through a new implicit scene graph neural network that effectively aggregates local object features. Comprehensive experiments on different datasets such as SUN RGB-D and Pix3D dataset demonstrate that our approach consistently surpasses state-of-the-art methods in object shape reconstruction, scene layout estimation, and 3D object detection.

Keywords: *Deep Learning, 3D Reconstruction, Computer Vision, 3D-Model, Multi-View Images*

1. Introduction

Researchers are increasingly interested in reconstructing 3D sceneries from 2D photographs. This is due to increased opportunities for 3D reasoning from photos made possible by the recent availability of extensive catalogues of 3D models. Three-dimensional (3D) reconstruction has become a fundamental research topic in computer vision, playing an essential role across numerous application domains such as autonomous driving [1,2], underwater exploration [3,4], medical diagnostics [5,6], precision agriculture [7,8], civil infrastructure monitoring [9,10], and robotic systems [11,12]. These diverse application settings range from well-controlled indoor scenes to highly complex environments like underwater imagery, which present specific challenges including light refraction, water turbidity, and low-texture regions. Such difficulties have motivated the development of advanced preprocessing strategies, including color normalization [13] and image denoising techniques [14], to improve reconstruction accuracy. Better decision-making, spatial comprehension, and interaction with real-world situations are made possible by the ability to rebuild precise three-dimensional representations from images or video sequences. In comparison to conventional approaches, deep learning techniques have recently produced significant breakthroughs in this field [15,16], allowing for more reliable, accurate, and computationally efficient solutions. However, the quick development and variety of deep learning-based techniques have produced inconsistent performance and applicability outcomes, igniting continuous discussions over the optimal methodology, measurements, and real-world applications. While techniques utilizing transformers [18], hybrid architectures [19], and convolutional neural networks (CNNs) [17] show great potential, there are differences in their effectiveness, computational cost, and context-specific flexibility. These differing viewpoints highlight the necessity of methodical assessments and comparative studies of current approaches. So, in this paper we propose a novel 3D reconstruction technique using deep learning techniques. Our proposed pipeline takes a single image as input, estimates layout and object poses, then reconstructs the scene from SUN RGB-D dataset as shown in figure 1.

Fig 1. 3D Reconstructed Image

2. Related work:

Several reconstruction algorithms are developed for scenes which are broadly classifies as follows:

(i) Single Image Scene Reconstruction: As a highly ill posed problem, single image scene reconstruction sets a high bar for learning-based algorithms, especially in a cluttered scene with heavy occlusion. The problem can be divided into layout estimation, object detection and pose estimation, and 3D object reconstruction. A simple version of the first problem is to simplify the room layout as a bounding box [22]. To detect objects and estimate poses in 3D space, recent works [23] try to

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 1111-1116

infer 3D bounding boxes from 2D detection by exploiting relationships among objects with a graph or physical simulation. At the same time, other works further extend the idea to align a CAD model with similar style to each detected object. Total3D [21] is proposed as an end-to-end solution to jointly estimate the layout box and object poses while reconstructing each object from the detection and utilizing the reconstruction to supervise the pose estimation learning.

(ii) Shape Representation: In the field of computer graphics, traditional shape representation methods include mesh, voxel, and point cloud. Some of the learning-based works try to encode the shape prior into a feature vector but stick to the traditional representations by decoding the vector into mesh [24], voxel [25] or point cloud [26]. Others try to learn structured representations which decompose the shape into simple shapes. Recently, implicit surface function [21] has been widely used as a new representation method to overcome the disadvantages of traditional methods (i.e. unfriendly data structure to neural network of mesh and point cloud, low resolution and large memory consumption of voxel). Most recent works try to combine the structured and implicit representation which provides a physically meaningful feature vector while introducing significant improvement on the details of the decoded shape.

(iii) Graph Convolutional Networks: Graph Convolutional Networks (GCN) have been widely used to learn from graph-structured data. Inspired by convolutional neural networks, convolutional operation has been introduced to graph either on spectral domain [27] or non-spectral domain which performs convolution with a message passing neural network to gather information from the neighboring nodes. Attention mechanism has also been introduced to GCN and has been proved to be efficient on tasks like node classification, scene graph generation and feature matching. Recently, GCN has been even used on super-resolution which is usually the territory of CNN. In the 3D world which interests us most, GCN has been used on classification and segmentation on point cloud, which is usually an enemy representation to traditional neural networks. The most related application scenario of GCN with us is 3D object detection on points cloud. Recent work shows the ability of GCN to predict relationship or 3D object detections [28] from point cloud data.

3. Proposed Method:

The proposed system consists of two stages, i.e., the initial estimation stage, and the refinement stage. In the initial estimation stage, a 2D detector is first adopted to extract the 2D bounding box from the input image, followed by an Object Detection Network (ODN) to recover the object poses as 3D bounding boxes and a new Local Implicit Embedding Network (LIEN) to extract the implicit local shape information from the image directly, which can further be decoded to infer 3D geometry. The input image is also fed into a Layout

Estimation Network (LEN) to produce a 3D layout bounding box and relative camera pose. In the refinement stage, a novel Scene Graph Convolutional Network (SGCN) is designed to refine the initial predictions via the scene context information. The proposed network is as shown in figure 2.

3.1 Implementation:

We use the outputs of the 2D detector from Total3D [21] as the input of our model. We also adopted the same structure of ODN and LEN from Total3D. LIEN is trained with LDIF decoder on Pix3D with watertight mesh, using Adam optimizer with a batch size of 24 and learning rate decaying from 2e-4 (scaled by 0.5 if the test loss stops decreasing for 50 epochs, 400 epochs in total) and evaluated on the original non-watertight mesh. SGCN is trained on SUNRGB-D, using Adam optimizer with a batch size of 2 and learning rate decaying from 1e-4 (scaled by 0.5 every 5 epochs after epoch 18, 30 epochs in total). When training SGCN individually, we use $_{Lj}$ without L_{phy}, and put it into the full model with pre-trained weights of other modules. In joint training, we adopt the observation from that objects reconstruction depends on clean mesh for supervision, to fix the weights of LIEN and LDIF decoder.

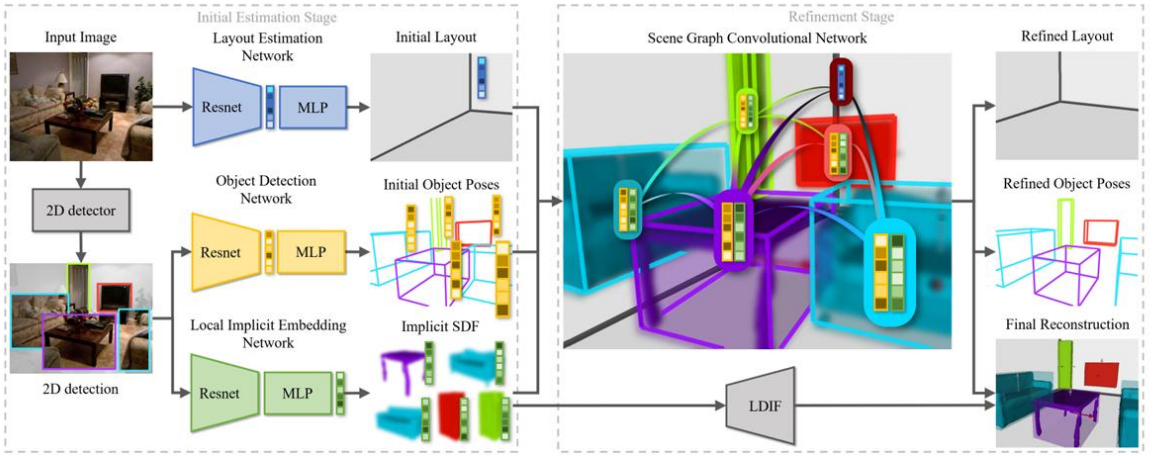

Fig 2. Proposed SGCN

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 1111-1116

3.2 Dataset:

We use two datasets to train each module individually and jointly. We use two datasets for training and evaluation. 1) Pix3D dataset [20] is presented as a benchmark for shape-related tasks including reconstruction, providing 9 categories of 395 furniture models and 10,069 images with precise alignment. We use the mesh fusion pipeline from Occupancy Network to get watertight meshes for LIEN training and evaluate LIEN on original meshes. 2) SUN RGB-D dataset [19] contains 10K RGB-D indoor images captured by four different sensors and is densely annotated with 2D segmentation, semantic labels, 3D room layout, and 3D bounding boxes with object orientations. Follow Total3D [21], we use the train/test split from [14] on the Pix3D dataset and the official train/test split on the SUN RGB-D dataset. The object labels are mapped from NYU-37 to Pix3D as presented by [21].

The SUN RGB-D dataset contains 10,335 indoor scene images, 3D camera poses annotations, 3D layout bounding boxes, semantic segmentation, and labels. From this dataset, 5050 images were used for testing and 5280 for training. The Pix3D dataset comprises 10,069 real-world images and 395 indoor object models under 9categories. These images and shapes are annotated using pixel-level 2D–3D alignment. From this dataset, 7556 images were used for training and 2513 for testing. To ensure a fair comparison, we used the same train/test splits as [18]. Figure 3 shows Sample images and shapes in Pix3D dataset.

3.3 Metrics:

We adopt the same evaluation metrics with [21], including average 3D Intersection over Union (IoU) for layout estimation, mean absolute error for camera pose, average precision (AP) for object detection, and chamfer distance for single-object mesh generation from single image.

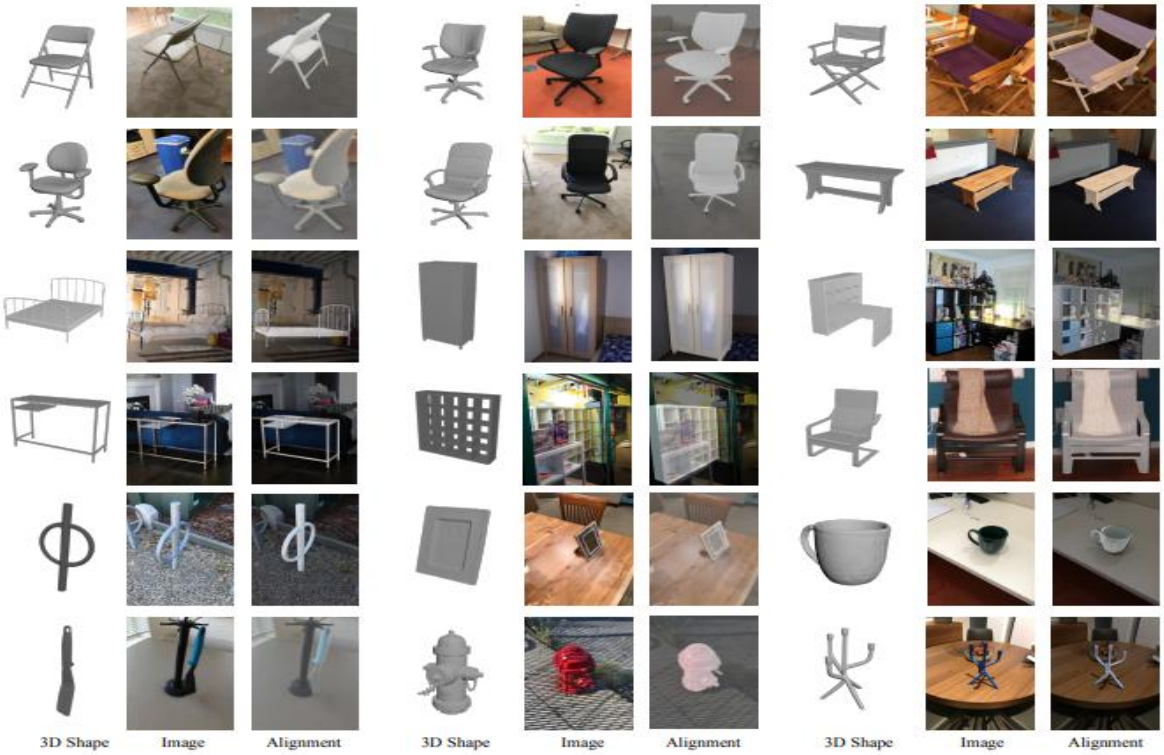

Fig 3. Sample images and shapes in Pix3D dataset

4. Experimental Results:

The proposed SGCN is evaluated using the above datasets. The mean value is compared with latest methods as given in table 1. Qualitative comparison on object detection and scene reconstruction. We compare object detection results with Total3D [21] and ground truth in both oblique view and camera view. The results show that our method gives more accurate bounding box estimation and with less intersection. We compare scene reconstruction results with Total3D in camera view and observe more reasonable object poses as shown in figure 4.

Table1. 3D Object Detection Comparison

Method	Bed	Chair	Sofa	Table	Desk	mAP
3DGP[29]	5.62	2.31	3.24	1.23	-	-
HoPR[30]	58.29	13.6	28.3	12.1	4.79	14.47
Coop[31]	57.51	15.21	36.7	31.6	19.9	21.7
Total[21]	60.65	17.55	44.9	36.48	27.9	26.38
Proposed method	88.3	34.8	68.6	57.8	52.1	44.3

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 1111-1116

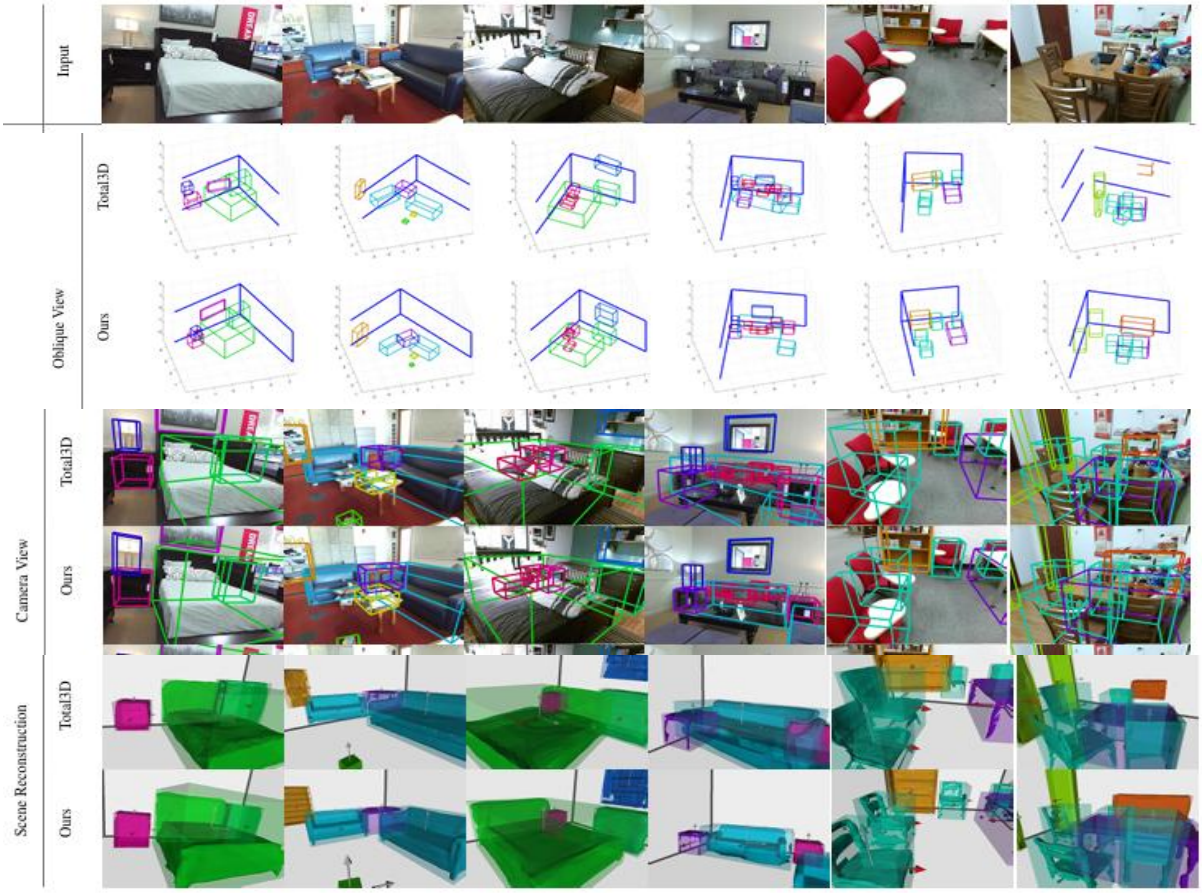

Fig 4. 3D Scene Reconstruction with the proposed SGCN and comparison with state of art.

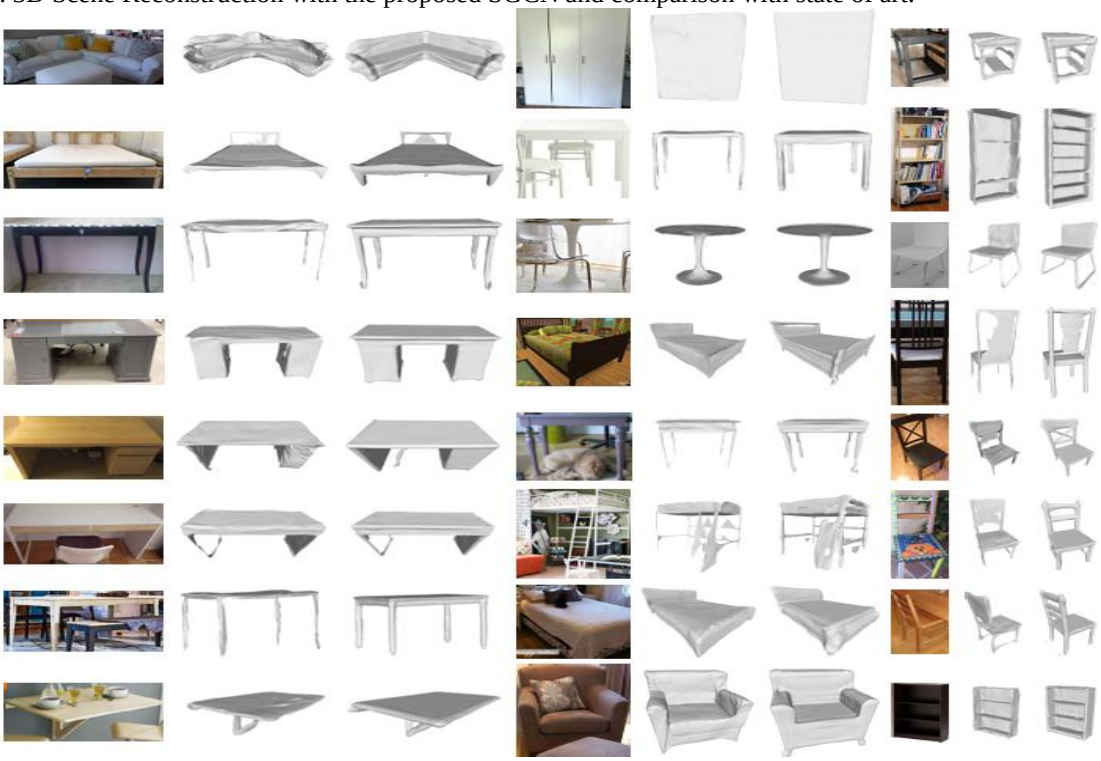

Fig 5. More qualitative comparisons on object reconstruction. We compare with Total 3D[21]. First column is input, second is reconstructed image from Total 3D[21], and third is the proposed SGCN.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 1111-1116

5. Conclusion:

We presented a holistic scene understanding framework based on deep implicit representations. The proposed method jointly reconstructs accurate 3D object geometry and models scene-level context through a graph convolutional network and a physics-aware violation loss. By integrating object-level and relational constraints, our approach produces more reliable scene layouts and object configurations. Extensive experimental results demonstrate consistent improvements over existing methods across multiple scene understanding tasks. Future work will explore incorporating object functionality and affordance cues to further enhance 3D scene reasoning.

References

1. Kompis, Y.; Bartolomei, L.; Mascaro, R.; Teixeira, L.; Chli, M. Informed sampling exploration path planner for 3d reconstruction of large scenes. IEEE Robot. Autom. Lett. 2021, 6, 7893–7900.
2. Ren, R.; Fu, H.; Xue, H.; Sun, Z.; Ding, K.; Wang, P. Towards a fully automated 3d reconstruction system based on lidar and gnss in challenging scenarios. Remote Sens. 2021, 13, 1981.
3. Hu, S.; Liu, Q. Fast underwater scene reconstruction using multi-view stereo and physical imaging. Neural Netw. 2025, 189, 107568.
4. Yu, Z.; Shen, Y.; Zhang, Y.; Xiang, Y. Automatic crack detection and 3D reconstruction of structural appearance using underwater wall-climbing robot. Autom. Constr. 2024, 160, 105322.
5. Maken, P.; Gupta, A. 2D-to-3D: A review for computational 3D image reconstruction from X-ray images. Arch. Comput. Methods Eng. 2023, 30, 85–114.
6. Zi, Y.; Wang, Q.; Gao, Z.; Cheng, X.; Mei, T. Research on the application of deep learning in medical image segmentation and 3d reconstruction. Acad. J. Sci. Technol. 2024, 10, 8–12.
7. Yu, S.; Liu, X.; Tan, Q.; Wang, Z.; Zhang, B. Sensors, systems and algorithms of 3D reconstruction for smart agriculture and precision farming: A review. Comput. Electron. Agric. 2024, 224, 109229.
8. Gu, W.; Wen, W.; Wu, S.; Zheng, C.; Lu, X.; Chang, W.; Xiao, P.; Guo, X. 3D reconstruction of wheat plants by integrating point cloud data and virtual design optimization. Agriculture 2024, 14, 391.
9. Lu, Y.; Wang, S.; Fan, S.; Lu, J.; Li, P.; Tang, P. Image-based 3D reconstruction for Multi-Scale civil and infrastructure Projects: Areview from 2012 to 2022 with new perspective from deep learning methods. Adv. Eng. Inform. 2024, 59, 102268.
10. Muhammad, I.B.; Omoniyi, T.M.; Omoebamije, O.; Mohammed, A.G.; Samson, D. 3D Reconstruction of a Precast Concrete Bridge for Damage Inspection Using Images from Low-Cost Unmanned Aerial Vehicle. Disaster Civ. Eng. Archit. 2025, 2, 46–62.
11. Banerjee, D.; Yu, K.; Aggarwal, G. Robotic arm based 3D reconstruction test automation. IEEE Access 2018, 6, 7206–7213.
12. Sumetheeprasit, B.; Rosales Martinez, R.; Paul, H.; Shimonomura, K. Long-range 3Dreconstruction based on flexible configuration stereo vision using multiple aerial robots. Remote Sens. 2024, 16, 234.
13. Wang,H.;Sun,S.; Ren, P. Underwatercolor disparities: Cues for enhancing underwater images toward natural color consistencies. IEEE Trans. Circuits Syst. Video Technol. 2023, 34, 738–753.
14. Wang,H.; Sun, S.; Chang, L.; Li, H.; Zhang, W.; Frery, A.C.; Ren, P. INSPIRATION: A reinforcement learning-based human visual perception-driven image enhancement paradigm for underwater scenes. Eng. Appl. Artif. Intell. 2024, 133, 108411.
15. Samavati, T.; Soryani, M. Deep learning-based 3D reconstruction: A survey. Artif. Intell. Rev. 2023, 56, 9175–9219.
16. Vinodkumar, P.K.; Karabulut, D.; Avots, E.; Ozcinar, C.; Anbarjafari, G. Deep learning for 3d reconstruction, augmentation, and registration: A review paper. Entropy 2024, 26, 235.
17. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. Proc. IEEE 2002, 86, 2278–2324.
18. Zhang, C.; Cui, Z.; Zhang, Y.; Zeng, B.; Pollefeys, M.; Liu, S. Holistic 3d scene understanding from a single image with implicit representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 8833–8842.
19. Song, S.; Lichtenberg, S.P.; Xiao, J. Sun rgb-d: A rgb-d scene understanding benchmark suite. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 567–576.
20. Sun, X.; Wu, J.; Zhang, X.; Zhang, Z.; Zhang, C.; Xue, T.; Tenenbaum, J.B.; Freeman, W.T. Pix3d: Dataset and methods for single image 3d shape modeling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2974–2983.
21. Yinyu Nie, 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 IEEE Conf. Comput. Vis. Pattern Recog., 2020.
22. Varsha Hedau, Derek Hoiem, and David Forsyth. Recover ing the spatial layout of cluttered rooms. In Int. Conf. Com put. Vis., 2009.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 1111-1116

23. Siyuan Huang, Siyuan Qi, Yinxue Xiao, Yixin Zhu, Ying Nian Wu, and Song-Chun Zhu. Cooperative holistic scene understanding: Unifying 3d object, layout, and cam era pose estimation. In Adv. Neural Inform. Process. Syst., 2018.
24. Junyi Pan, Xiaoguang Han, Weikai Chen, Jiapeng Tang, and Kui Jia. Deep mesh reconstruction from single rgb images via topology modification networks. In Int. Conf. Comput. Vis., pages 9964–9973, 2019.
25. Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In Adv. Neural Inform. Process. Syst., pages 82–90, 2016.
26. GuandaoYang,XunHuang,ZekunHao,Ming-YuLiu,Serge Belongie, and Bharath Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In Int. Conf. Comput. Vis., pages 4541–4550, 2019.
27. Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.
28. Mahyar Najibi, Guangda Lai, Abhijit Kundu, Zhichao Lu, Vivek Rathod, Thomas Funkhouser, Caroline Pantofaru, David Ross, Larry S Davis, and Alireza Fathi. Dops: Learn ing to detect 3d objects and predict their 3d shapes. In IEEE Conf. Comput. Vis. Pattern Recog., pages 11913–11922, 2020.
29. Wongun Choi, Yu-Wei Chao, Caroline Pantofaru, and Silvio Savarese. Understanding indoor scenes using 3d geometric phrases. In IEEE Conf. Comput. Vis. Pattern Recog., 2013.
30. Siyuan Huang, Siyuan Qi, Yixin Zhu, Yinxue Xiao, Yuanlu Xu, and Song-Chun Zhu. Holistic 3d scene parsing and reconstruction from a single rgb image. In Eur. Conf. Comput.Vis., 2018.
31. Siyuan Huang, Siyuan Qi, Yinxue Xiao, Yixin Zhu, Ying Nian Wu, and Song-Chun Zhu. Cooperative holistic scene understanding: Unifying 3d object, layout, and camera pose estimation. In Adv. Neural Inform. Process. Syst., 2018.

Author Bibliography

Mr. Naveen Kumar Nekkalapu received the AMIETE degree in Electronics and Communication Engineering from IETE, New Delhi, in 2004, and the M.Tech degree in Electronics and Communication Engineering from Nagarjuna University, Guntur, in 2006. He is currently pursuing his Ph.D. in Electronics and Communication Engineering from Jawaharlal Nehru Technological University, Kakinada. Since 2006, he has been working as an Assistant Professor in the Department of Electronics and Communication Engineering at various institutions, including St. Mary's College of Engineering and Technology, SLC's Institute of Science and Technology, and Made Easy Institute Private Limited. His research interests include Image Processing and Computer Vision.

Dr Giri Babu Kande (Senior Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Nagarjuna University, Guntur, in 1996, the M.E. degree in electronics and communication engineering from Andhra University, Visakhapatnam, in 2000, and the Ph.D. degree in electronics and communication engineering from Jawaharlal Nehru Technological University, Hyderabad, in 2010. From 2008 to 2019, he was a Professor and the Head of the Department of Electronics and Communication Engineering Department. Since 2019, he has been a Professor and the Dean of Academics with the Vasireddy Venkatadri Institute of Technology, Nambur. He is the author of one book and more than 75 publications. His research interests include computer vision, statistical modeling and learning, and medical image analysis.

Dr B PrabhakarRao has obtained his master's from SVU, Tirupati in 1982, Ph.D in Sonar Signal processing from Indian institute of Science, Bangalore in 1995. He is a Professor & Dean in the Department of Electronics and Communication Engineering at Godavari Global University Rajamahendravaram EGDistrict from March 2025. Prior to this he is having 43 years of Experience in teaching and administrative and Research experience. Dr Rao has contributed significantly in the areas of Optical Networks, Image Processing, Microwave Engineering, and emerging technologies such as Deep & Machine learning. He published 370 international And National journals and presented research papers in international And national conferences in India and Abroad like Singapore, Malaysia and USA. He served in different capacities staring from Hod,Vice Principal, Director of Evaluation, Rector and Vice Chancellor i/c.