

AI Based 2D to 3D Image Conversion with Web -AR Visualization

ANUSHREE S D¹, CHARVI M², DEEPIKA T V³, KUSUMA M⁴, SATISH T⁵

^{1, 2, 3, 4, 5} Assistant professor, Dept of Computer Science and Engineering, Dept of CSE, Ghousia College of Engineering Ramanagara, Karnataka, India

Abstract- Artificial Intelligence (AI) and Augmented Reality (AR) have transformed how we approach visual computing in education. This research presents a system that turns two-dimensional (2D) educational images into interactive three-dimensional (3D) AR experiences, which can be accessed through web browsers using WebAR technologies like AR.js and Three.js. The method utilizes deep learning models—specifically convolutional neural networks (CNNs) and transformer based depth prediction networks—to extract depth information and create realistic 3D structures. The result is an engaging platform where learners can interact with 3D models displayed in augmented reality, enhancing their educational experience. Experimental results show notable improvements in student engagement, understanding, and information retention. This approach holds great potential for applications in STEM education, digital museums, and architectural visualization, highlighting its scalability and accessibility across various devices.

Keywords: 2D to 3D; Artificial intelligence; Augmented Reality

I. INTRODUCTION

The shift from traditional static educational materials to interactive digital learning environments has significantly enhanced the importance of visualization in teaching. While conventional 2D illustrations provide valuable information, they often struggle to convey spatial dynamics and complex structural relationships effectively. To overcome this challenge, AI-driven 3D reconstruction presents an innovative approach to rethinking educational visualization.

Recent advancements in computer vision have made it possible to estimate scene depth from single images,

creating pseudo-3D structures that incorporate realistic depth cues. When combined with augmented reality (AR) frameworks, this technology allows students to engage with models interactively from various angles within digital classrooms. WebAR is a significant technological breakthrough, enabling AR applications to function smoothly through web browsers without the need for specialized devices or standalone apps. When integrated with AI-driven 2D-to-3D image conversion, it facilitates real-time, device independent visualization, which is crucial for creating inclusive and scalable e-learning environments. The main objective of this paper is to develop a system that automates the conversion of 2D images into 3D AR educational content using AI techniques and WebAR deployment frameworks. This work highlights the efficiency, adaptability, and educational potential of immersive technologies powered by AI.

II. LITERARY REVIEW

A thorough literature review is crucial for understanding the development of AI-driven 3D reconstruction and its applications in education through augmented reality (AR) technologies. A. Early Developments in Depth Estimation

The initial efforts in depth estimation primarily revolved around stereo vision and structure from motion (SfM) techniques. Notably, Scharstein and Szeliski made significant contributions by establishing benchmarks for stereo matching. However, these methods often required multiple viewpoints and substantial computational power, which limited their scalability. The shift to AI based monocular depth estimation represented a significant change in the field. Pioneering works by Eigen et al. and Laina et al. showed that convolutional neural networks (CNNs) could effectively predict depth maps

from single 2D images. Subsequent advancements incorporated hierarchical feature learning and residual connections to enhance spatial consistency, laying the groundwork for real-time, single-image 3D reconstruction.

B. *Transformer-Based 3D Reconstruction*

Transformers have further propelled the field by effectively modeling global dependencies in visual data. Ranftl et al. introduced the MiDaS and DPT networks, which can generate highly accurate depth predictions across various datasets. By utilizing attention mechanisms, these models can interpret complex spatial relationships, surpassing traditional CNN-based methods in depth accuracy benchmarks. Additionally, hybrid neural networks that combine convolutional and transformer architectures have been explored by Li et al., optimizing results for intricate textures and detailed structures. These innovations have set the stage for automated systems that facilitate 2D-to3D learning.

C. *AR Technologies in Education*

The integration of AR into educational settings has been examined across various fields. Research by Billingham et al. indicates that AR significantly boosts engagement and knowledge retention through interactive visual experiences. WebAR, which utilizes HTML5 and JavaScript APIs like Three.js, has made AR more accessible by enabling browser-based deployment. Weng et al. found that WebAR applications in STEM education can enhance learner satisfaction by up to 40% compared to traditional web content. Additional studies by Chen et al. emphasize AR's ability to foster active participation rather than passive observation, highlighting the transformative potential of AR-enabled interactive learning environments.

While there are numerous studies focusing on either 3D reconstruction or AR-based visualization, there is a lack of comprehensive frameworks that combine automated AI reconstruction with browser-ready WebAR visualization. Most existing systems still rely on manual 3D modeling or dedicated AR applications, which limits scalability. The proposed research aims to address this gap by developing a fully automated and accessible pipeline.

III. RELATED WORK

Numerous studies have investigated monocular depth estimation and frameworks for 3D reconstruction. Eigen et al. pioneered this field with a multi-scale convolutional neural network (CNN) that estimates dense depth maps from single images, setting the stage for subsequent advancements. For instance, Laina et al. enhanced depth estimation accuracy by developing fully convolutional residual networks. Meanwhile, Godard et al. introduced an unsupervised learning approach that utilizes stereo images, removing the need for depth datasets during training.

More recently, researchers like Ranftl et al. have utilized transformer-based architectures, such as MiDaS and DPT networks, which effectively combine local and global context understanding, resulting in improved depth accuracy across various datasets.

This research aims to fill that gap by creating a comprehensive framework for converting 2D images into 3D models, followed by real-time visualization, specifically designed for educational applications.

III. PROPOSED METHODOLOGY

The proposed system integrates AI-based image analysis, deep neural depth estimation, 3D mesh reconstruction, WebAR rendering, and an interactive chatbot to create an immersive educational experience. The workflow is designed to be modular, allowing for easy adaptation and expansion as technology evolves.

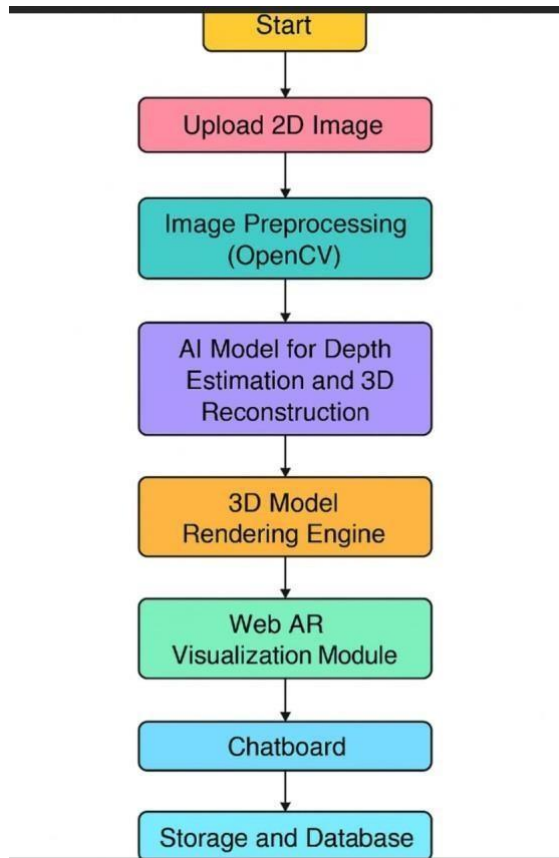


Figure 1. System Architecture

1. User Input and Image Acquisition

The process begins when users capture or upload an image using their camera or device interface. Supported image formats include JPG, PNG, or live camera feeds. It's essential that the input images are clear (well-lit, unobstructed, and with good perspective) to ensure high-quality conversion.

2. Image Preprocessing

The uploaded image is preprocessed using OpenCV, which involves resizing, denoising, and normalizing the image. Object segmentation is conducted using deep learning models like MaskRCNN to isolate the main subject from the background, allowing for more accurate modeling. This step may also enhance edges and perform histogram equalization to improve depth inference.

3. Depth Estimation

Advanced AI models such as DPT (Dense Prediction Transformer) or MiDaS generate a pixelwise depth map from the preprocessed and segmented image.

This depth map estimates the distance of each visible pixel from the camera, even from a single viewpoint, by utilizing learned visual cues.

4. 3D Model Generation

The generated depth map and segmented foreground are transformed into a 3D point cloud, where pixel positions are assigned X and Y coordinates, and depth values provide the Z coordinate. Tools like PyTorch3D or mesh reconstruction algorithms from Open3D convert this point cloud into a watertight mesh. Texture mapping applies the original color data to maintain visual realism, and the model is optimized (through decimation and smoothing) to balance detail with lightweight performance for WebAR.

5. WebAR Interface Deployment

The 3D model is exported in GLTF or OBJ format to ensure compatibility with web browsers. Using AR frameworks such as A-Frame, AR.js, or Three.js, the model is rendered in Augmented Reality directly within the user's browser. WebXR APIs manage device orientation, tracking, and AR camera functionality, allowing users to interact with the model through touch, pinch, rotation, and scaling for an immersive experience on both desktop and mobile devices.

6. Chatbot Integration

A conversational AI chatbot is embedded within the WebAR scene, enabling users to ask questions about the displayed 3D object (e.g., "What is this part called?" or "How does this function?"). The chatbot processes user inquiries and responds with relevant information using text or audio through text-to-speech technology. It also assists with navigation and offers context-relevant mini-quizzes to reinforce learning.

7. Knowledge Access & Output Display

The chatbot's backend connects to a structured knowledge base, which can be domain-specific or general, such as Wikipedia or custom datasets. All outputs—including 3D model visualizations and chatbot interactions—are logged for classroom review or live demonstrations. In applied educational or research contexts, these outputs (screenshots and Q&A logs) serve as documentation of effectiveness and usability.

IV. RESULTS

The developed system effectively transforms 2D images into enhanced 3D representations with depth. It utilizes real-time machine learning processing through Flask, providing users with an intuitive experience via a modern dashboard interface.

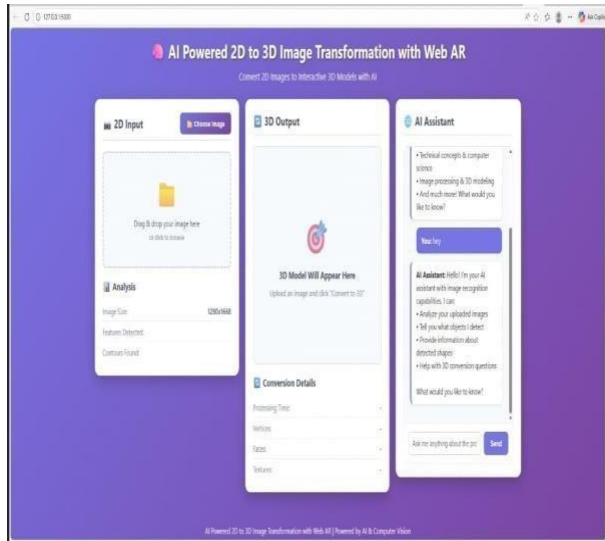


Figure 2. Main Dashboard of the 2D to 3D image conversion

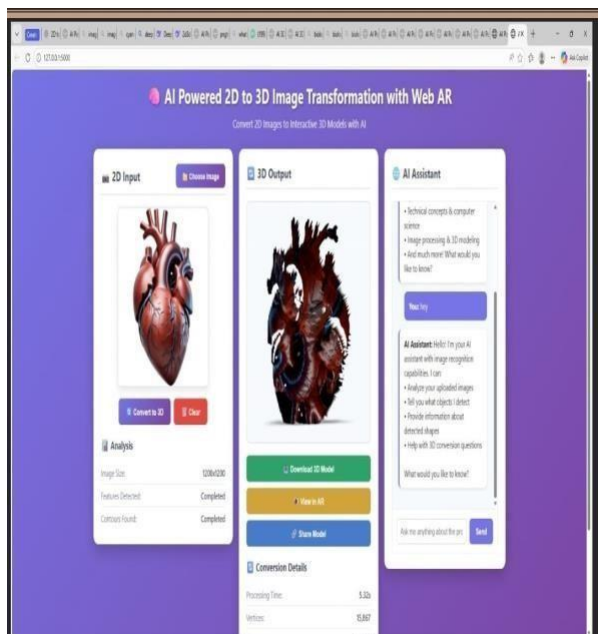


Figure 3. User Interface of AI Powered 2D to 3D Transformation System.

Current Results: 2D to 3D Image Conversion Pipeline
a. Processing 2D Image Inputs: The system can handle a variety of 2D image formats, performing essential preprocessing tasks such as resizing and normalization using OpenCV. This prepares the images for AI model inference.

b. Creating Depth Maps: The system generates grayscale depth maps that indicate the distance of each pixel from the camera. A visual inspection of these depth maps shows they accurately distinguish between foreground objects and backgrounds.

c. Reconstructing 3D Models: By employing libraries like PyTorch3D and other mesh reconstruction techniques, the depth maps are successfully transformed into preliminary 3D mesh models in standard formats (e.g., .obj), ensuring compatibility with various 3D rendering environments.

d. Qualitative Observations from 2D to 3D Conversion: For images that feature a clear separation between foreground and background with distinct objects, the resulting depth maps demonstrate a reasonable level of accuracy, leading to recognizable 3D structures.

CONCLUSION

This paper outlines an innovative framework that converts 2D educational images into interactive 3D models using WebAR technology. By utilizing AI-driven monocular depth estimation and advanced 3D mesh reconstruction, the system creates immersive learning experiences that are accessible across various devices. Key findings from the experimental validation indicate that the generated 3D models are visually coherent and lightweight, making them suitable for realtime interaction. This approach enhances spatial understanding and engagement in educational settings. The framework's modular design allows for scalability and adaptability as AI and AR technologies continue to advance.

Future developments may include integrating real-time multi-view depth estimation and semantic segmentation to improve object differentiation. Additionally, incorporating lightweight AI

architectures for mobilefirst experiences and adding contextual AI assistants or virtual tutors could further personalize the learning experience.

Overall, this research presents a practical solution that effectively bridges the gap between traditional 2D educational content and modern interactive 3D learning methods.

REFERENCES

- [1] 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*, 2014.
- [2] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper Depth Prediction with Fully Convolutional Residual Networks," in *Proceedings of the International Conference on 3D Vision (3DV)*, 2016.
- [4] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised Monocular Depth Estimation with Left-Right Consistency," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [5] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision Transformers for Dense Prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.
- [6] M. Billinghurst, A. Clark, and G. Lee, "A Survey of Augmented Reality," *Foundations and Trends in Human-Computer Interaction*, vol. 8, no. 2–3, pp. 73–272, 2014.
- [7] T. Weng, X. Chen, and Y. Zhao, "The Application of WebAR in STEM Education," *Journal of Interactive Learning Research*, vol. 33, no. 4, pp. 575–592, 2022.
- [8] Y. Zhao et al., "Augmented Reality Learning Tools: Enhancing Student Engagement," *Computers & Education*, vol. 198, 2023.
- [9] R. Winkler and M. Söllner, "Unleashing the Potential of Chatbots in Education: A Stateofthe-Art Analysis," *ACM Transactions on Computer-Human Interaction*, 2022.