Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 218–228
https://doi.org/10.54941/ahfe1008256

Qualitative Study Updated With Recent 3D Object Generation Models for Immersive Technologies

Pio Henrique Ordozgoith Da Frota Sobrinho, Gilberto Oliveira Neto, Agustin Alejandro Ortiz Diaz, Sergio Cleger Tamayo, Heider Betcel Lima De Souza, Geovana Amorim Abensur, Delrick Nunes de Oliveira, Gustavo De Freitas Martins, and Osvaldo Vitalino Dos Santos Junior

Sidia Institute of Science and Technology. Manaus, Amazonas, 69055-035, Brazil

ABSTRACT

As a meeting point between AI-generated content and extended reality (XR), 3D generative content has rapidly evolved from text and image synthesis to the creation of 3D objects. This article investigates user perceptions of two 3D generative models—Trellis and TripoSR—which operate from a single image within an XR environment using a head-mounted display. Employing a qualitative analysis methodology with 21 evaluators, the study provides valuable insights into overall 3D object quality, image-to-object correspondence, and user satisfaction regarding generation time, highlighting the models' potential for immersive applications. The two models were tested using a minimal image dataset categorized into four groups: humanoids, animals, places, and objects. A direct comparison of the two models yielded clear results: users noted that the Trellis model provided a significantly more satisfying experience than the TripoSR model.

Keywords: Artificial intelligence generated content, 3D Generative content, Extended reality, Qualitative methods

INTRODUCTION

The increasing development of large-scale pre-trained models has notably strengthened AI-generated products. These products have been transformed into generative tools with unlimited potential that have directly driven innovation in multiple application areas (Wu, 2023). These models and products have become prominent in the generation of different types of content, such as text, images, audio, video, and, currently experiencing rapid growth, 3D content. It is important to highlight 3D content due to its close relationship with emerging markets such as immersive technologies, especially Extended Reality (XR). These technologies provide users with realistic experience, featuring several innovative elements (Fu, 2021; Li, 2023).

The generation of 3D objects has significantly transformed how users interact with immersive environments. In particular, the use of automatically

Received October 7, 2026; Revised November 7, 2026; Accepted November 15, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

generated 3D objects from limited data, such as two-dimensional images or textual descriptions, has expanded the possibilities for content creation in XR applications. Within this context, head-mounted displays (HMDs) play a fundamental role in providing immersive experiences that enhance the user's sensory and cognitive perception of these objects (Cleger, 2025-a).

User perception when interacting with AI-generated 3D objects in immersive environments is an emerging area of research involving multiple dimensions, including visual quality, geometric consistency, real-time responsiveness, and the sense of presence. These variables directly influence the acceptance, trust, and usability of the systems, especially in critical applications such as simulation, education, and training. Furthermore, factors such as latency, rendering fidelity, and adaptation to the user's context shape the overall experience, creating new challenges for the evaluation and validation of these systems (Cleger, 2025-b).

A recent study presented a qualitative evaluation based on three works that introduced models for generating 3D meshes from a single image and their application in immersive environments for HMDs (Cleger, 2025-b): Unique3D (Wu, 2024), CRM (Wang, 2024), and InstantMesh (Xu, 2024). These three models follow the same strategy for generating 3D meshes, this strategy is structured in two stages. First, the input image is fed into a diffusion model capable of generating a set of multi-view images. Second, these multi-view images are fed into another model to finally reconstruct the 3D mesh. However, all three works highlight the existence of potential limitations in their first stage, which could result in low-quality meshes (Cleger, 2025-a).

However, during this period (2024-2025), two other interesting generative models (2D to 3D) were published with divergent positioning in the quadrant defined by the structural quality of the generated 3D asset and its generation time.

- **TripoSR** (Tochilkin, 2024) focuses on real-time reconstruction from a single image using a TriPlane-NeRF architecture based on transformers; generates textures meshes in less than 0.5 seconds.
- **Trellis** (Xiang, 2025) is a large-scale 3D asset generation model that stands out for its flexible output format selection and advanced local 3D editing capabilities.

The objective of this work is to explore the intersection of AI-based content generation and XR technology, developing a qualitative study with users to evaluate the perceived quality of 3D object generation visualized on devices such as HMDs. For this reason, we propose conducting a subjective evaluation of the quality of 3D resources automatically generated by specialized models.

RELATED WORK

The development of single-view 3D reconstruction has evolved considerably, driven by advances in Artificial Intelligence-Generated Content (AIGC), which enables more efficient and scalable alternatives to traditional 3D modeling

techniques (Wu, 2023). Early approaches relied on photogrammetry methods such as Structure from Motion (Schönberger, 2016-a) and Multi-View Stereo (Schönberger, 2016-b), while the introduction of Neural Radiance Fields (NeRF) marked a breakthrough by representing scenes as neural radiance fields for high-quality 3D reconstruction.

More recently, large-scale datasets including Objaverse (Deitke, 2023), ShapeNet (Chang, 2015), and Google Scanned Objects (Downs, 2022) have enabled the development of data-driven reconstruction models. Among them, Large Reconstruction Model (Hong, 2023) and InstantMesh (Xu, 2024) significantly improved scalability and inference speed by combining diffusion models with sparse-view reconstruction, although challenges remain in preserving fine textures and geometric consistency. CRM further addressed these limitations by directly mapping image features into 3D representations through fine-tuned multi-view diffusion models trained on Objaverse (Deitke, 2023) and ShapeNet (Chang, 2015), improving both reconstruction quality and efficiency (Wang, 2024).

Building on these advances, Unique3D (Wu, 2024) introduced multi-level upscaling and the Instant and Consistent Mesh Reconstruction algorithm, producing high-fidelity mesh within seconds and outperforming previous feed-forward methods such as CRM (Wang, 2024) and InstantMesh (Xu, 2024). However, recent progress in AI-based 3D generation has significantly improved the reconstruction of textured 3D objects from a single-view images. Some of the models that today are part of the state of the art were published just two years ago.

- A little less recent (2024) and focusing on computational efficiency, TripoSR (Tochilkin, 2024) employs a transformer-based TriPlane-NeRF architecture that reconstructs textured meshes from a single image in less than 0.5 seconds, achieving state-of-the-art performance among feed-forward methods.
- The Trellis model (Xiang, 2025) introduces a unified latent representation capable of generating multiple 3D formats, including meshes, radiance fields, and Gaussian splats, demonstrating excellent geometric accuracy and perceptual quality through Rectified Flow Transformers.
- Hunyuan3D model (Zhao, 2025) extends this direction by combining multi-view diffusion with transformer-based feed-forward reconstruction, improving view consistency and geometric completeness while maintaining relatively fast inference.

Although these approaches have substantially advanced AI-driven 3D reconstruction, they reflect different design priorities, ranging from maximum visual fidelity to real-time efficiency. Consequently, selecting the most appropriate model for immersive XR applications requires balancing reconstruction quality with inference speed and hardware constraints, motivating comprehensive comparative evaluations under practical XR deployment scenarios.

PROPOSED METHODOLOGY AND RESULTS

In this work, the research methodological guidelines defined in (Cleger, 2025-b) were followed. This methodology is divided into five well-defined stages, which are represented in a general way in Figure 1.

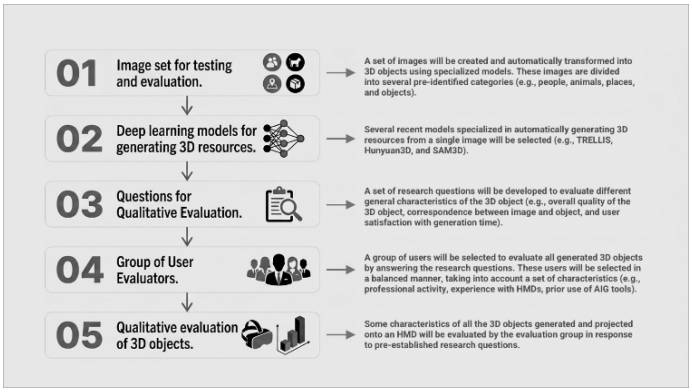


Figure 1: Methodological strategy used for the comparative analysis of generative 3D asset models.

Image Set for Testing and Evaluation

The 3D objects analyzed in this work were generated from four images shown in Figure 2. These images represent four main categories of interest: humanoids, animals, places, and objects, identified by users in previous research.



Figure 2: Set of images used for the evaluation of generative models.

Deep Learning Models for 3D Resource Generation

To generate 3D objects from a single image, two generative models with different characteristics were selected: TripoSR (Tochilkin, 2024) and Trellis (Xiang, 2025). Four objects were generated by each model, totalling eight 3D objects for the study. The selection of these models was based on previous research and their availability. Figure 3 shows the reference images and 3D assets generated by each model.

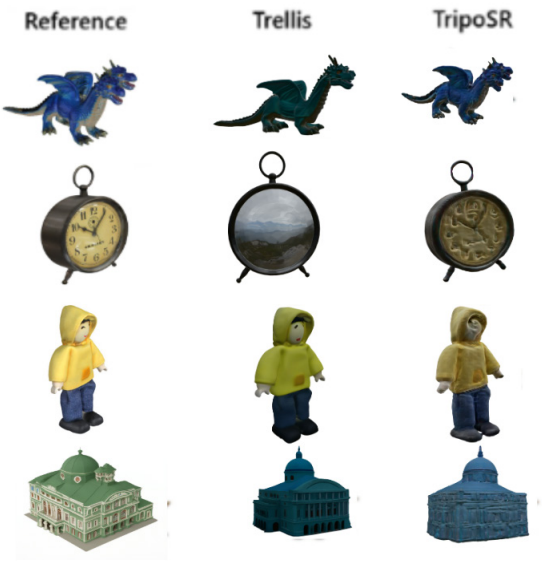


Figure. 3: Reference images and 3D objects generated by each model studied.

To carry out this study, an immersive application was developed in Unity (Figure 4). This immersive application simulates the process of generating a 3D object from a single image for the evaluators. Since the 3D objects were generated beforehand, outside of the Unity application, a delay was applied to the visualization of the 3D object to simulate the waiting time for its generation. The delay time was equivalent to the real time spent pre-generating the object.

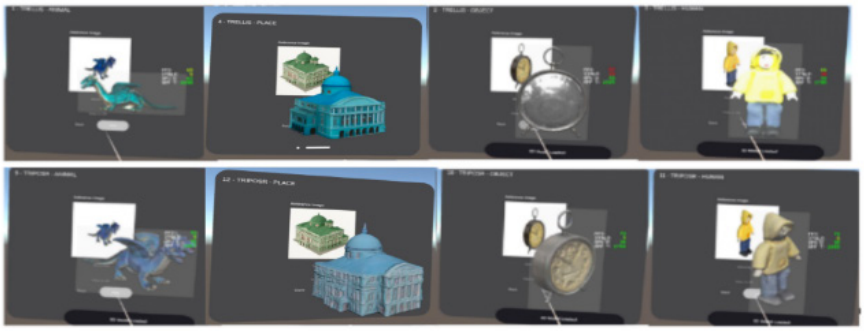


Figure 4: The immersive application with interactable 3D objects.

Description of the Evaluator Group

To evaluate the 3D resources resulting from the generative process of the studied models and to gather information about the interaction and performance of the 3D object generation process in immersive technologies, a qualitative survey was conducted with a group of 21 evaluator users. The characteristics of this group are described in Table 1.

Table 1: User demographics.

Professional Activity		Level of Experience with HMD Devices	
Artist.	3	**No experience.**	0
Designer.	6	**Know that exist.**	0
Product Owner.	1	**Little experience.**	1
Project Manager.	2	**Moderate experience.**	9
Software developer.	7	**Advanced experience.**	9
Tester.	1	**Extensive experience.**	2
Coordinator.	1		21

Questions and Results

The tests were divided in two rounds, one for TrispoSR and one for Trellis. Each round would display the same four different image categories of interest: humanoids, animal character, place, and objects. The participants would generate a 3D object for each image in each model. In total, eight 3D images were evaluated. The goal was to evaluate each model's loading time, visual quality and the relation between these two aspects. For each model presented, the users would click "generate" for generating a 3D object from each image, a loading time would be applied and an interactable virtual object would be available user analysis. Users would be able to freely move, scale, and rotate the objects (Figure 4).

The following four tables (2, 3, 4, and 5) summarize the users' responses to the key questions for analysing the model outputs according to different aspects.

Table 2: Questions and answers about the visual quality of the 3D objects.

1.	What called your attention the most in the object's visual quality?
2.	What could be improved in the object's visual or structure?

Trellis model. Evaluation of visual quality in the Trellis architecture reveals a clear division in participant perceptions between non-organic geometric objects and complex organic entities. For simple blocky shapes, silhouettes, and basic structures (such as an object or place figures), the model was commended for effectively capturing outer outlines and overall proportions. Notably, dynamic lighting rendering was highlighted as an advantageous feature for artistic workflows. In contrast, evaluations regarding organic representations—specifically animal and humanoid figures—were significantly more critical, with outputs described as blurry, lacking definition, and exhibiting poor texture quality alongside a severe loss of fine detail. A recurring pattern indicates that detail fidelity varies directly according to object complexity: non-organic items yield cleaner geometries, whereas organic subjects account for most perceived flaws regarding texture, depth resolution, and anatomy. Facial synthesis and anatomical representation represent the primary failure points of the architecture, with participants reporting that the model fails to properly identify facial features and hallucinates details absent from source images. Texture execution falls short across multiple cases, failing to render stripes, patterns, and surface finishes. Because non-organic assets perform significantly better, user recommendations center on post-processing mechanisms, chromatic treatment for unfinished assets, and the integration of multi-angle inspection tools prior to accepting generated outputs.

(Continued)

Table 2: Continued.

TripoSR model. User assessment of the TripoSR model produced a virtually unanimous and strongly negative consensus. Synthesized visual quality was consistently categorized as unsatisfactory across all evaluation metrics. Reconstructed geometries exhibit low polygon density, severe surface distortion, and structural collapse (metaphorically described as objects appearing to "melt"). Architecture frequently introduces hallucinatory geometric features while compromising color fidelity in favor of rudimentary structural approximations. Consequently, dissatisfaction with TripoSR was expressed globally across entire assets rather than localized features. When specific failure modes were enumerated, participants highlighted general mesh distortion, porous textures, washed-out color reproduction, unfinished surface finishes, and a complete absence of geometric coherence. In contrast to Trellis—where criticisms were constrained to specific object categories—visual quality degradation in TripoSR occurs universally across all evaluated object types.

Table 3: Questions and answers about the 3D objects resemblance to the original single image, it was based.

3.	How does the 3D object faithfully resembles the original single image?
4.	What visual aspects could contribute for better visual resemblance with the original image?

Trellis model. Evaluation reference fidelity metrics for Trellis mirror the trends observed in visual quality: high fidelity is achieved for static geometric assets, whereas humanoid and organic entities exhibit severe degradation. Indeed, humanoid figures suffered from extreme divergence, rendering generated subjects virtually unrecognizable. The model consistently preserves outer silhouettes and dominant color distributions but fails to maintain depth resolution and internal geometric symmetry. Overall, participants regard Trellis outputs as adequate for basic structural approximation or illustrative visualization, but unreliable for precise reproduction. While the model captures geometric fundamentals (silhouettes, primary colors, basic proportions), it fails on fine refinement (textures, fine symmetry, anatomical details), reinforcing its utility as a structural approximation framework rather than a faithful replication system.
TripoSR model. Participants reporting near-zero correspondence to reference inputs. The model retains a generic notion of the initial silhouette, while entirely losing proportions, fine detail, and chromatic accuracy. Even benchmark test cases (such as place figure) exhibited severe color and structural divergence, indicating that the model's best-case outputs fail to reach the quality of worst-case outputs in competing architecture. Without explicit knowledge of the source image, most generated models are rendered unrecognizable. While shape and color are the only attributes partially preserved, they are characterized as weak attempts rather than functional strengths. Texture, depth, and fine detail were universally cited as absent or detrimental to reference similarity.

Table 4: Questions and answers about the loading time for 3D object generation.

5.	How did the loading time affect your perception of overall 3D object quality?
6.	How adequate do you consider the loading time?

Trellis model. User perception regarding generation latency in Trellis (~30 seconds) reveals three distinct user stances: (1) **Acceptable Latency:** A majority considers processing duration consistent with output quality for simple geometric forms. (2) **Frustrated Expectations:** Extended processing time increases expectations of high detail that are unmet in complex models, undermining perceived quality. (3) **Preference for Extended Processing:** Users express a positivity to tolerate longer generation times if additional processing yields tangible detail improvements.

(Continued)

Table 4: Continued.

Satisfaction with generation latency correlates directly with asset complexity. While processing speed is deemed adequate for static items, latency is perceived as unacceptable for humanoid figures due to severe hallucination and quality defects, prompting requests for extended compute times in exchange for geometric precision.
TripoSR model. TripoSR achieves rapid generation times (~3 seconds), minimal latency fails to drive user satisfaction and instead reinforces disappointment. The prevailing user perception is that minimal computational effort directly results in poor volumetric depth and degraded mesh quality. A significant majority of participants explicitly prefer longer generation latencies in exchange for heightened structural accuracy and detailed fidelity. Ambiguity arises as direct evaluations often acknowledge that processing time is consistent with the meager output delivered. However, this acknowledgment reflects awareness of low processing effort rather than satisfaction with system performance. User demand centers entirely on increasing process duration to enhance geometric precision, with no participants requesting further reductions in generation time.

Table 5: Percentage of responses from evaluator users to the corresponding statement. The meaning of the abbreviated column headings is as follows: **S.D.** = Strongly disagree; **P.D.** = Partially disagree; **N.** = Neutral; **P.A.** = Partially agree; and **S.A.** = Strongly agree.

Statement	Model	S. D.	P. D.	N.	P. A.	S. A.
"The 3D object presents a satisfactory level of visual quality and details."	**Trellis**	28%	17%	17%	19%	19%
	TripoSR	88%	12%	0%	0%	0%
"The 3D object corresponds faithfully to the image from which it was generated, in terms of shape and overall appearance."	**Trellis**	35%	12%	21%	19%	13%
	TripoSR	79%	20%	1%	0%	0%
The 3D object's loading time was adequate and did not negatively impact my experience."	**Trellis**	18%	18%	24%	21%	19%
	TripoSR	20%	12%	%7	%19	40%

SUMMARY AND ANALYSIS OF THE RESULTS

These two models occupy clearly different positions on a spectrum of quality vs. speed. However, an important aspect that must be highlighted is that neither model resolves the humanoid figure well. Tables 6 and 7 summarize the general characteristics of the analysed models and summarize the advantages and disadvantages of each of them according to the aspects studied.

Table 6: Synthesis of the characteristics of the two models analysed.

Model	Perceived Quality	Generation Time	General Loyalty
Trellis	Inconsistent median.	Moderate (~30s).	Average (good at geometric shapes, poor at organic ones).
TripoSR	Low, negative consensus.	Very low (~3s).	Very low in almost all objects.

Table 7: Summary with the advantages and disadvantages of the analysed models.

Trellis model	**Pros.**	Good balance between time and quality for geometric/static objects. Praised dynamic lighting; well-captured silhouette.
	Cons.	Generally weak texture. Significant hallucination on faces/humanoid figures. Inconsistent quality across object categories.
TripoSR model	**Pros.**	Very high speed (almost instantaneous). The only positive aspect consistently highlighted by participants.
	Cons.	Quality and fidelity considered insufficient even for a low-demand illustrative use case. Structural flaws (not merely issues of refinement). Even the attributes that are most "resilient" in other models (shape, color) appear degraded here.

Trellis Model:

- Has a solid silhouette and geometric structure, but it's described as 'blurry' and lacking textural definition, especially with organic objects.
- The criticisms are primarily focused on the face or humanoid figure (hallucinations) and insufficient texture. Tone of responses: "Good but needs improvement in X aspect."
- Regarding conditional fidelity, it is high for static objects (clock) and low for organic objects (animal, humanoid). It is good for basic visualization but unreliable for faithful reproduction.

TripoSR Model:

- Receives unanimously negative feedback on this. Terms such as 'melting,' 'looks like a cake,' and 'low polygon count' appear with unusual frequency, indicating a near-total lack of structural quality. The evaluators consider the problems with this model to be fundamental structural flaws, rather than matters of refinement.
- A significant portion of the responses literally boils down to 'improve everything' or 'redo the AI'; there is no specific improvement to suggest, as the foundation itself is considered flawed. Tone: 'everything is wrong'.
- Fidelity is virtually non-existent for most; even the "best-case" example displays colours completely different from the reference.

CONCLUSION

In this work, we conduct an exploratory study on the intersection of two research areas that have seen significant advancement in recent years: AI-generated content (AIGC) and extended reality (XR). Our proposal aims to qualitatively evaluate how users in immersive environments perceive the quality of 3D objects generated by deep learning models from a single image. To structure our work, we followed a methodology proposed in 2025 for this type of study (Cleger, 2025-b). The results reveal complex, interconnected assessments within the immersive environment.

For this study, we selected two well-known generative models with distinct characteristics—Trellis (Xiang, 2025) and TripoSR (Tochilkin, 2024)—and tested them using a minimal image dataset divided into four categories: humanoids, animals, places, and objects. A group of 21 evaluators were assembled to analyze the 3D objects generated by these models within a unified immersive environment. Direct comparison of the two models yielded clear results; generally, users highlighted that the Trellis model provides a more promising experience. Key conclusions included the following:

- If perceived quality is the priority, even at the cost of longer wait times, Trellis is the highest-rated model across all questions, if expectation management is addressed through some form of feedback mechanism.
- If delivery speed is the priority, Trellis appears to be a safer choice than TripoSR, offering reasonable speed without triggering the near-unanimous rejection observed with TripoSR. Despite being the fastest, TripoSR carries the risk of negatively affecting the perception of overall product quality. The repeated assertion—"yes, but nothing was delivered"—indicates that speed is insufficient to compensate for a delivery falling below the minimum acceptable standard.
- In both cases, the humanoid figure is the critical cross-cutting element; any chosen model must be treated with caution regarding this specific type of reference image, perhaps by using a product rule that assigns lower reliability to objects containing humanoid figures, or by applying additional post-processing in such instances.

ACKNOWLEDGMENT

This paper was presented as part of the results of the Project "SIDIA-M_VST_PLATFORM_AND_APPLICATIONS", carried out by the Institute of Science and Technology – SIDIA, in partnership with Samsung Electrônica da Amazônia LTDA, in accordance with the Information Technology Law n.8387/91 and article at the 39 of Decree 10,521/2020.

REFERENCES

Chang, A., Funkhouser, T., Guibas, L., et al. (2015). "ShapeNet: An Information-Rich 3D Model Repository". arXiv preprint arXiv:1512.03012.

Cleger, S., Abensur, G., Ortiz, A., et al. (2025). "Comparative Analysis Between Different 3D Object Generation Models. Adaptation to Immersive Technologies". In International Conference on HCI, pp. 18–35. Springer Nature Switzerland.

Cleger, S., Vitalino, O., Ortiz, A., et al. (2025). "Qualitative Study of 3D Object Generation in Immersive Technologies". In HCI International 2025. Lecture Notes Computer Science, Springer. https://doi.org/10.1007/978-3-032-12657-3_11.

Deitke, M., Schwenk, D., Salvador, J. et al. (2023). "Objaverse: A universe of annotated 3D objects". In IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, pages 13142–13153. IEEE, 2023.

Downs, L., Francis, A., Koenig, N. et al. (2022). “Google Scanned Objects. A High-Quality Dataset of 3D Scanned Household Items”, https://arxiv.org/pdf/2204.11918.

Fu, K., Peng, J., He, Q. et al. (2021). Single image 3D object reconstruction based on deep learning. A review. Multimed Tools, https:/doi.org/10.1007/s11042-020-09722-8.

Hong, Y., Zhang, K., Gu, J. et al. (2023). “LRM. Large Reconstruction Model for Single Image to 3D”. CoRR, abs/2311.04400.

Li, C., Zhang, C., Waghwase, A., et al. (2023). “Generative ai meets 3d: A survey on text-to-3d in aigc era”, arXiv preprint arXiv:2305.06131.

Schönberger, J., Frahm, J. (2016). “Structure-from-motion revisited”. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4104–4113.

Schönberger, J., Zheng, E., Frahm, J. et al. (2016). “Pixelwise view selection for unstructured multi-view stereo”. In Computer Vision–ECCV 2016.

Tochilkin, D., Pankratz, D., Liu, Z. et al. (2024) “TripoSR: Fast 3D Object Reconstruction from a Single Image”, https://arxiv.org/abs/2403.02151.

Wang, Z., Wang, Y., Chen, Y., et al. (2024). “CRM: Single Image to 3D Textured Mesh with Convolutional Reconstruction Model”. In: European Conference on Computer Vision (ECCV), pp. 57–74. Springer.

Wu, J., Gan, W., Chen, Z., et al. (2023). “AI-Generated Content (AIGC): A Survey”. arXiv preprint arXiv:2304.06632.

Wu, K., Liu, F., Cai, Z., et al. (2024). “Unique3D. High-Quality and Efficient 3D Mesh Generation from a Single Image”. https://arxiv.org/abs/2405.20343.

Xiang, J., Lv, Z., Xu, S. et al. (2025). “Structured 3D Latents for Scalable and Versatile 3D Generation”, https://arxiv.org/abs/2412.01506.

Xu, J., Cheng, W., Gao, Y., et al. (2024). “InstantMesh: Efficient 3D Mesh Generation from a Single Image with Sparse-view Large Reconstruction Models”. arXiv preprint arXiv:2404.07191.

Zhao, Z., Lai, Z., Lin, Q. et al. (2025). “Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation”, https://arxiv.org/abs/2501.12202.