ABC2: Journal of Architecture, Building, Construction, and Cities

Empowering research for Sustainable Development Goals, ABC2: Architecture, Building, Construction, and Cities is a fundamental manifesto to address these pressing issues, fostering dialogue and knowledge exchange among researchers, practitioners, and policymakers. Exploring sustainable design, resilient infrastructure, advanced construction methods, and equitable urban development, ABC2 aims to empower the global community to create adaptive, inclusive, and sustainable environments. The ABC2 focus on cutting-edge research, technological advancements, and transformative strategies is essential for navigating the future of our cities and communities.

Research Article

A Photo to 3D Workflow for Generation of LOD3 Digital Building Representations

Handan Aş Çemrek[1*], Ümit Işıkdağ[2]
Gebrail Bekdaş[3], Sinan Melih Niğdeli[4]

[1] Istanbul Health and Technology University, Istanbul, 34275, Turkey
[2] Mimar Sinan Fine Arts University, Istanbul, 34427, Turkey
[3,4] Istanbul University, Cerrahpasa, Istanbul, 34098, Turkey

DOI: https://doi.org/10.66408/abc2.2026.20

* Correspondence: handan.as@istun.edu.tr

Received: 01/01/2026
Revised: 09/05/2026
Accepted: 20/05/2026
Published: 25/06/2026

Volume: 2006
Issue: 04
Pages: 26-42

Abstract

This study proposes a sequential AI workflow for generating LOD3-oriented digital building representations from a single architectural photograph. The novelty of the approach lies in the use of Gemini Flash 2.5 for preprocessing and Hunyuan3D 2.1 for image-to-3D generation within a preprocessing–reconstruction decoupled pipeline. Rather than directly producing a 3D model from an uncontrolled photograph, the workflow first normalizes the image by isolating the building, converting the scene to daylight, and approximating an isometric view, and then converts this image into a textured 3D mesh. In this way, the method offers a rapid, low-data, and reproducible alternative for cultural heritage visualization, comparative analysis, and digital archiving. Experiments were conducted on 16 photographs of architectural structures representing typologies from the early Republican and Ottoman periods. The findings demonstrate that the method offers a tool for the digital documentation of cultural heritage, archiving and academic research. The resulting 3D outputs are suitable for urban silhouette analysis at the LOD3, volumetric comparison, testing urban design scenarios, and the creation of digital heritage archives. The method works with a single photograph and cloud-based AI models, which is advantageous when on-site scanning is difficult, risky, or costly. Furthermore, the workflow can be enhanced with multiple images, text input, and other sensor data to produce digital twins with geometric accuracy. In this respect, the study demonstrates that by combining cutting-edge technologies in 3D modeling from a single image, 3D twins of architectural objects can be produced, and proposes an AI-powered method for creating digital architectural representations.

Keywords: Single-image 3D reconstruction; Gemini Flash 2.5; Hunyuan3D 2.1; AI-based modeling; LOD3 digital building representations; Cultural heritage documentation; Image-to-3D generation; Diffusion model

Highlights

- An AI workflow based on Gemini Flash 2.5 and Hunyuan 3D 2.1 is proposed to generate a 3D building model from a single architectural photograph.
- The method's suitability for rapid, low-cost heritage documentation is demonstrated on 16 early Republican and Ottoman building photos.

It is shown that the resulting 3D models can directly provide input to architecture-focused applications such as urban silhouette analysis and digital heritage archives.

1 Introduction

The rapid development of digital technologies is fundamentally transforming approaches to documenting and preserving architectural heritage. In particular, the three-dimensional (3D) recording of historical structures within the scope of cultural heritage has become an increasingly indispensable tool for conducting academic research and creating virtual experiences and educational content (Münster et al., 2024). Against this background, this study proposes a sequential AI workflow that combines Gemini Flash 2.5 and Hunyuan3D 2.1 to generate LOD3-oriented digital building representations from a single architectural photograph. The main contribution of the study lies in its preprocessing–reconstruction decoupled design: instead of relying on a raw input image, the workflow first standardizes the visual input through background removal, daylight normalization, and near-isometric reformulation, and then converts the processed image into a textured 3D mesh. This design is particularly relevant in cultural heritage contexts where on-site capture is difficult and only a single archival or opportunistic photograph is available.

For many years, such 3D documents have been produced using traditional methods such as photogrammetry and laser scanning. These techniques allow for the creation of three-dimensional models using numerous photographs of a structure taken from different angles or various sensor data. Photogrammetry can produce highly accurate 3D reconstructions by matching common feature points in hundreds of images; however, it also has limitations such as the multi-stage processing, the need for intensive data collection, and the cost of time (Münster et al., 2024).

In many cultural heritage contexts, comprehensive on-site capture is constrained by access restrictions, safety concerns, limited time windows, and budget limitations. Consequently, researchers and practitioners often rely on existing photographs that were not acquired for metric surveying but still contain valuable architectural cues. A workflow that operates on a single photograph is valuable for documentation and for producing preliminary 3D assets that can be refined when higher-quality data become available.

Against this background, recent AI advances have made it possible to generate 3D models from a single photograph. To provide a clear theoretical basis for this shift, the following section reviews key single-image 3D reconstruction approaches, current AI tools, and the conceptual framing that informs the proposed workflow. Accordingly, the study frames the proposed pipeline as a lightweight alternative that prioritizes automation and reproducibility while remaining compatible with downstream analyses. The objective is not to claim survey-grade accuracy from one image, but to provide a practical route to LOD3-oriented representations that support visualization, comparison, and archiving.

While the workflow prioritizes speed and accessibility, it may not recover geometry, absolute scale, or ornamentation from a single view. Accordingly, outputs should be interpreted as plausible exterior representations and refined using additional photographs, text constraints, or points when available.

From a heritage informatics perspective, a key contribution of the proposed workflow is its explicit fitness-for-purpose framing for single-image 3D generation. Rather than claiming survey-grade accuracy, the study targets operational usability: producing a consistent textured mesh that imports reliably into common AEC and GIS environments and supports comparative and scenario-based analyses, and positions results for downstream archival reuse. This is particularly relevant for heritage collections where photographs may be the only record and rapid 3D surrogates can aid curation, teaching, and research. Reporting the steps, visual normalization choices, and output formats also improves transparency and repeatability in cloud-based generative pipelines.

The aim of this research is to propose a method for generating digital 3D models of selected examples of Turkish architectural heritage using Gemini Flash 2.5 and Hunyuan 3D 2.1 models in a sequential workflow. The method was tested on 16 photographs of buildings from the early Republican and late Ottoman periods of civil architecture; thus, its performance under different architectural styles and visual conditions was observed. The initial findings indicate that the method can offer a practical

solution for the digital documentation of cultural heritage and for academic studies in the field of architecture. The method steps followed and the results obtained from the example applications are presented below, followed by a general evaluation by discussing the gains and limitations obtained.

2 Theoretical Background and Conceptual Framework

This section establishes the theoretical basis for the study by situating the proposed workflow within the evolving field of single-image-to-3D reconstruction and digital cultural heritage documentation. It first clarifies the key concepts used throughout the paper (e.g., digital building representation, LOD3, digital heritage archives, and reproducibility) and then reviews the dominant AI paradigms and cloud-based toolchains that enable 3D asset generation from limited visual input for architectural objects. Building on this synthesis, the section identifies the main gaps that emerge when detailed on-site capture is infeasible and only a single photograph is available. Finally, it introduces the conceptual model that structures the Gemini Flash 2.5–Hunyuan3D 2.1 pipeline and its intended applications.

2.1 Defining Key Concepts

Recently, advancements in artificial intelligence have led to the emergence of new approaches that make it possible to generate 3D models from a single photograph. Generative AI techniques such as Generative Adversarial Networks (GANs) and diffusion-based models have made significant progress in deriving 3D shapes from incomplete, limited, or singular visual data. For example, various GAN-based methods have been developed to digitize objects from a single photograph or to complete missing 3D geometries. Similarly, approaches such as Neural Radiance Fields (NeRFs) can predict different views of a scene and the corresponding 3D geometry from a small number of images or even a single image (Münster et al., 2024). These developments make 3D modelling methods from a single image particularly attractive in cultural heritage sites where taking multiple shots is difficult, costly, or physically impossible.

In this paper, a digital building representation refers to a textured surface mesh that captures the perceptual identity of an architectural object for visualization, comparative analysis, and archiving, rather than a fully parametric BIM model. LOD3 is used to denote an exterior representation with architecturally meaningful articulation that supports skyline and silhouette studies at an urban scale. In established BIM and HBIM frameworks, LOD-based representations are typically associated with progressively richer geometric articulation and, in many cases, increasing semantic structure. In this study, however, LOD3 is used in a more bounded and operational sense: not to claim a fully parametric BIM or HBIM model, but to describe an exterior building representation with sufficient formal articulation to support visualization, comparative analysis, and archival use. Accordingly, the proposed workflow should be understood as producing a geometry-oriented, LOD3-like digital building representation that may complement BIM- and HBIM-related processes, rather than replacing them. Reproducibility denotes that the workflow, given the same input image and prompts, can be repeated with consistent steps and outputs that can be inspected and reported.

2.2 Existing Theories and Frameworks

Large technology companies have also accelerated the transformation in this field by making successive moves around 3D modelling from single images. Google DeepMind introduced Gemini Flash 2.5 as a lightweight and fast image model, and its relevance to the present study lies not in direct 3D generation, but in its ability to support image editing and normalization tasks such as background removal, lighting correction, and perspective reformulation from a single architectural photograph (Saadat et al., 2025; AINews, 2025). In this respect, Gemini Flash 2.5 functions in this study as a preprocessing tool that reduces visual ambiguity before 3D generation.

In parallel, the Hunyuan3D family represents the recent development of open-source image-to-3D systems. Hunyuan3D 2.0 introduced a scalable diffusion-based framework for textured 3D asset generation, while Hunyuan3D 2.1 further advanced this direction by focusing on the generation of high-

fidelity textured meshes from a single image (Tencent Hunyuan3D Team, 2025; Hunyuan3D et al., 2025). For the purposes of this study, the key point is that Hunyuan3D 2.1 enables the production of reusable textured 3D assets that can be inspected and transferred to common visualization and modelling environments. The public circulation of the release through online discussion communities also contributed to its broader visibility among potential users (Reddit, 2025).

Hunyuan3D 2.1 is an end-to-end 3D content creation system capable of generating a textured 3D mesh from a single image input using physically based materials. Architectural, design, and everyday objects can be modelled with PBR (Physically Based Rendering)-compatible materials based on a single photograph. The system integrates two core components: the Hunyuan3D-DiT module for shape generation and the Hunyuan3D-Paint module for texture generation (Hunyuan3D et al., 2025). Hunyuan3D-DiT generates the overall 3D geometry of the object, while Hunyuan3D-Paint performs multi-view conditional texture synthesis around the generated shape, producing PBR texture sets including albedo, metallic, roughness, and normal maps. This two-stage yet integrated approach enables the consistent combination of colour and geometry, resulting in high-quality 3D outputs (Hunyuan3D et al., 2025). Such systems contribute to advancing single-image 3D modelling in terms of both accuracy and visual quality.

The Hunyuan3D family has been compared with contemporary large-scale 3D generative models such as Michelangelo (Zhao et al., 2023), Craftsman 1.5 (Li et al., 2024), and Trellis (Xiang et al., 2024). The technical report indicates that Hunyuan3D 2.1 performs strongly in preserving detailed geometry, maintaining multi-view consistency, and generalizing to open-world objects. Likewise, when evaluated alongside methods such as TripoSG (Li et al., 2025) and Direct3D-S2 (Wu et al., 2025), Hunyuan3D 2.1 stands out for its ability to preserve complex surface detail and maintain texture-image consistency (Hunyuan3D et al., 2025).

In studies focusing on cultural heritage and architecture, there is a growing interest in image-to-3D production approaches. Montas-Laracuente et al. (2025), in their work on the digital reconstruction of historical Romanesque–Mudéjar churches, proposed a photo-based workflow instead of traditional LiDAR point cloud-based methods and designed a “photo-to-BIM” (HBIM) conversion process supported by Gaussian Splatting from a single facade photograph. In this approach, the radiance field obtained from the input images is approximately represented by surface-aligned Gaussian particles; then, this representation is transformed into a triangular mesh surface using the SuGaR (Surface-aligned Gaussian Reconstruction) algorithm. The study’s findings revealed that the 3D model generation process was significantly faster compared to classical point cloud-based approaches, and the resulting networks were consistent enough to be integrated into an HBIM (Historic Building Information Modelling) environment.

Methodologically, single-image 3D reconstruction is ill-posed because depth, scale, and occluded geometry cannot be uniquely determined from one view. Contemporary AI systems address this by learning strong priors from large datasets and by using generative modelling to propose plausible 3D hypotheses consistent with the input image. This contrasts with earlier geometry-driven pipelines that typically required multi-view imagery or calibrated capture, and it explains both the rapid progress and the persistent uncertainty in façade-level detail.

2.3 Knowledge Gaps and Research Opportunities

These examples demonstrate that artificial intelligence and advanced 3D generative methods are positioned as innovative tools in the digital representation of cultural heritage and architectural modelling processes. Consequently, diffusion-based models, NeRF-based representation techniques, GAN derivatives, and large-scale open-source systems are simultaneously developing within the single-image 3D modelling ecosystem, creating a significant area of knowledge accumulation and experience for both academia and industry, particularly in architectural and cultural heritage applications. However, despite the rapid proliferation of AI models for single-image 3D generation, their practical use for architecture-specific heritage documentation remains underexplored, particularly in workflows that

explicitly combine 2D image editing and 3D mesh generation for rapid, low-cost LOD3-oriented representations.

A further gap concerns evaluation and fitness-for-purpose. Heritage documentation often requires clear statements about what generated models can support (e.g., interpretive visualization) and what they should not be used for (e.g., precise measurement). In addition, outputs are rarely assessed in terms of practical transfer to AEC and GIS environments, where scale assumptions, mesh cleanliness, and metadata affect usability.

2.4 Proposed Conceptual Model

This study examines the process of obtaining a 3D model from a single architectural photograph using two of the most current tools in the field of artificial intelligence. The first is the Gemini 2.5 Flash Image model (commonly known as "Nano Banana") developed by Google DeepMind. Gemini Flash 2.5 is an advanced image production and editing model that can make targeted changes to images using natural language commands. For example, this model can perform complex edits such as removing unwanted objects from a photograph, recolouring a specific area, or changing the orientation of an object with a single prompt (Google Developers Blog, 2025). In our scenario, Gemini Flash 2.5 was used to convert a building photograph to an isometric perspective and isolate the structure by removing it from its background. Indeed, it was observed that the model, with a command such as "Make image daytime and isometric (building only)" could easily select the building in the photograph, separate it from its background, and transform the scene as if it were retaken from an isometric perspective during daylight hours. As highlighted in Google's own statements, Gemini Flash 2.5 can even take a night shot and convert it to daylight, realistically adding environmental details such as wall details or cables that are not visible in the original image (weixin_47221050, 2025).

The second key tool used in this study is the Hunyuan 3D 2.1 model. This open-source system, offered by Tencent, is designed as an artificial intelligence infrastructure capable of generating 3D models from text or image inputs (Echo3D, 2025). Thanks to its diffusion-based architecture, Hunyuan 3D can create a highly detailed 3D entity consisting of a triangular mesh and accompanying physically based (PBR) material textures, taking a single image as input (Hunyuan3D, 2025). One of the model's distinctive aspects is that it treats shape (geometry) and texture generation as a two-stage sequential process: in the first stage, the geometric mesh of the object is generated by the Hunyuan3D-DiT component, and in the second stage, the Hunyuan3D-Paint module synthesizes material and texture maps for this mesh through multi-view conditional diffusion (Team Hunyuan3D et al., 2025; Tencent Hunyuan3D Team, 2025). Thus, it becomes possible to obtain a highly realistic, three-dimensional, textured digital asset based on a single photograph (Echo3D, 2025).

Operationally, the conceptual model assumes a single input photograph and produces a textured mesh that can be reused across common software ecosystems. The preprocessing stage reduces visual ambiguity by isolating the building from cluttered backgrounds and by generating a controlled daylight, near-isometric view that standardizes perspective cues. The reconstruction stage converts this normalized image into 3D geometry and materials, after which the output can be inspected and exported for LOD3-oriented analyses, archiving, and scenario testing. The model is extensible to multi-image and text-guided inputs for improved geometric accuracy

3 Methodology

The workflow proposed in this study is structured in four main stages.

3.1 Selection of Input Image

First, a suitable entrance image is selected for the building to be converted into a 3D model. Ideally, the photograph should be taken during the day, from a diagonal perspective, at an angle close to an isometric viewpoint. Sufficient natural lighting and clear shadow information in exterior shots contribute

to a more reliable and legible 3D reconstruction. If possible, shots with minimal obscuring elements such as trees, vehicles, or pedestrian traffic that obscure a large portion of the building are preferred. However, if the photograph is taken at night or has a perspective significantly different from an isometric viewpoint, these deficiencies and distortions can be partially compensated for in the next step using the editing capabilities of the Gemini Flash 2.5 Image model.

The study used sixteen architectural photographs selected from Turkey, representing both the early Republican period and the more modern era, as input data. This dataset consists of iconic buildings such as Ankara Palace, the Turkish Grand National Assembly Museum, the Ankara Republic Museum, Gazi Education Institute, Aynalıkavak Palace, Baruthane, Baruthane Tower, Bruno Taut House, Doğan Apartment, Florya Atatürk Sea Pavilion, Frej Apartment, Arif Pasha Apartment, Narmanlı Han, Pera Palace, Safranbolu House, and İkinci Evkaf Apartment. Facade photographs of each building were provided and processed as separate inputs to the Gemini model. All of these buildings constitute the input dataset shown in Figure 1.

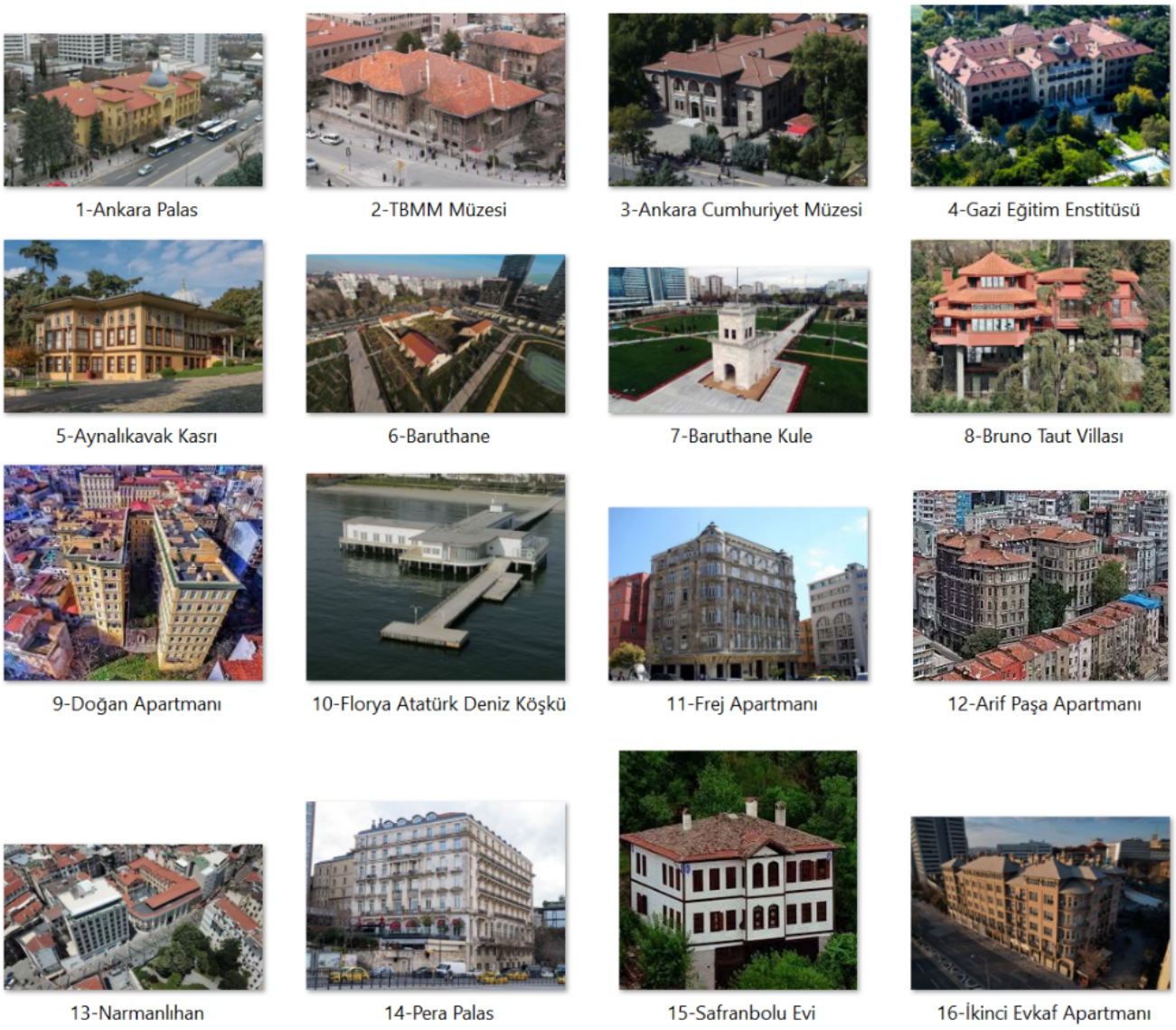

Figure 1. Input dataset of selected Turkish architectural landmarks provided to Gemini Flash 2.5. See Appendix A for the detailed case list.

3.2 Image Preprocessing Using the Gemini Flash 2.5 Model

In the second step, the selected building photograph is given as input to the Google Gemini 2.5 Flash Image model. After the photograph is loaded into the model interface, the text prompt “make image daytime and isometric (building only)” is applied. This prompt allows the model to focus only on the building mass in the scene, transforming the image from night conditions to daylight and making the perspective isometric. As a result of the process, the model separates the building from its background and reconstructs it on a neutral background. By applying slight perspective corrections to the facade, the building gains an appearance as if it were taken from a bird's-eye/isometric angle. For example, when this process is applied to a building photograph taken at night, the output produces an isolated and perspective-enhanced image of the building under daylight; however, it is observed that some details that were not visible in the original scene due to the dark shooting conditions cannot be fully restored by the model (weixin_47221050, 2025).

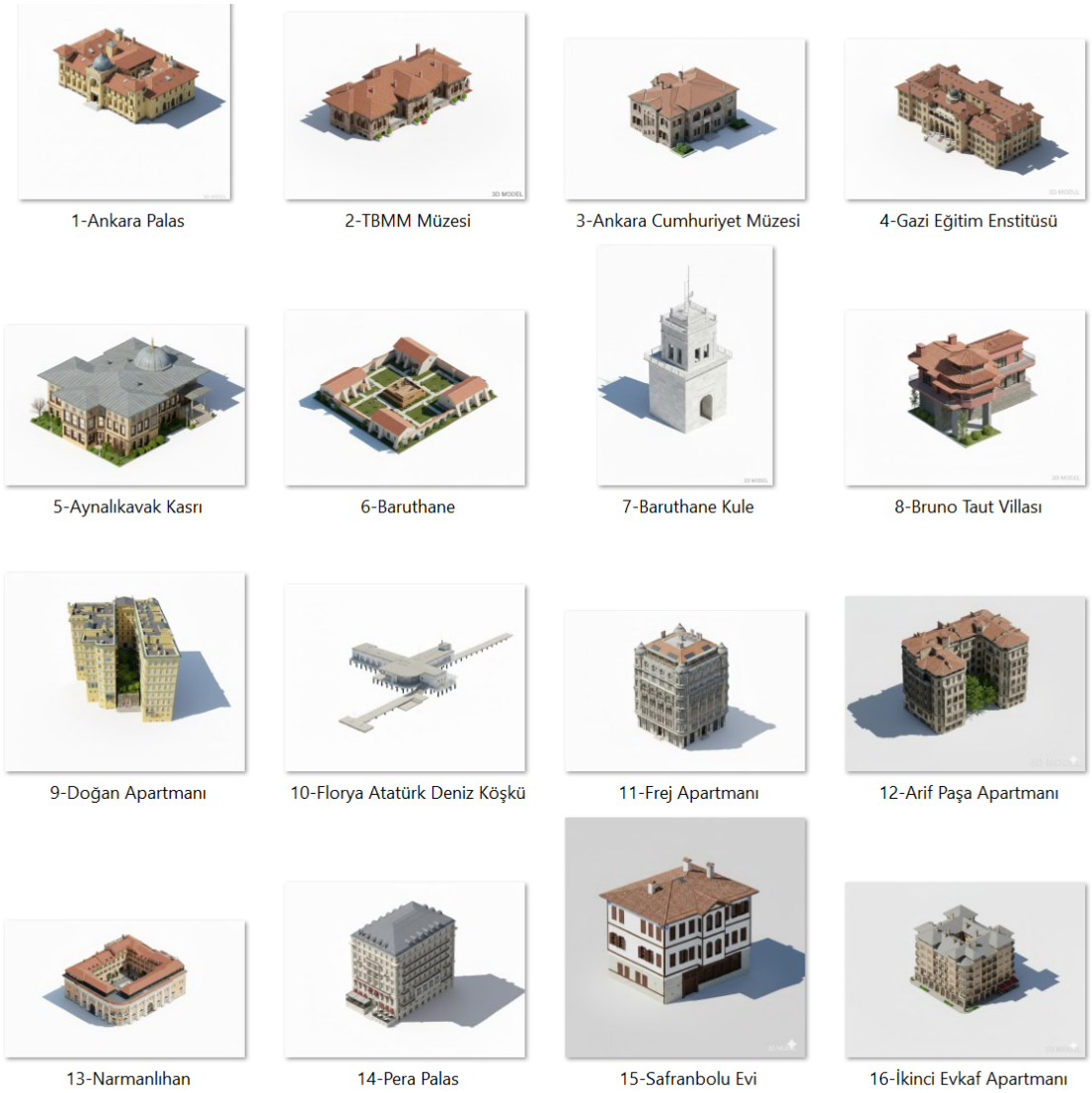

Figure 2. Output images generated by Gemini Flash 2.5 based on the real building photographs provided as input (see Figure 1). See Appendix A for the detailed case list.

The outputs obtained after preprocessing with the Gemini Flash 2.5 Image model demonstrate that the structures are successfully isolated from their original backgrounds, lighting conditions are converted to daytime views, and an approximate isometric perspective is achieved. Furthermore, in some examples, the model is observed to consistently complete details that are not clearly visible in the

original image, such as side facades or background architectural elements. This step plays a critical role in providing clean, legible, and geometrically more suitable input images for the 3D reconstruction process to be carried out with Hunyuan3D 2.1.

3.3 3D Reconstruction Using Hunyuan3D 2.1

In the third step, the isometric view of the building obtained in the second step, converted and isolated from its background, is transferred to the Hunyuan3D 2.1 model. In Hunyuan3D's image-to-3D working mode, when a single facade image is given as input, the system produces a holistic 3D object from this two-dimensional representation (Hunyuan3D, 2025). In this study, the outputs obtained with Gemini Flash 2.5 were loaded directly into the Hunyuan3D 2.1 interface without any additional editing. The model's generation process is architecturally structured as a pipeline consisting of two sub-stages: in the first stage, the three-dimensional form of the structure, i.e., the polygonal mesh topology, is estimated and created; in the second stage, realistic material and texture maps are applied to this geometric surface. Hunyuan3D 2.1's use of a physically based (PBR) material model for shading ensures that the generated 3D entity offers a realistic appearance compatible with contemporary game engines and visualization software in terms of features such as metallic surfaces, reflections, and roughness (Tencent-Hunyuan, 2025).

After the model is run, the user is offered the option to download the resulting 3D output in **.**glb (GLTF Binary) or **.**obj format. Hunyuan3D's documentation states that the calculated result is represented as a trimesh object with a triangular mesh structure, and this data can be easily converted to both file formats (Tencent-Hunyuan, 2025). In this study, the .glb format was preferred for the vast majority of examples; because this GLTF-based format can combine geometry, material definitions, and texture images in a single file and is naturally supported by browser-based viewers, game engines, and many 3D modelling and BIM software programs (Echo3D, 2025).

Figure 3. 3D building models reconstructed by Hunyuan3D 2.1 from the Gemini-processed façade images. The textured meshes correspond to the sixteen Turkish architectural landmarks in the input dataset (see Figure 1).

In this study, the image completion steps performed with Gemini Flash 2.5 were structured to precede and follow the 3D reconstruction process with Hunyuan3D 2.1. In other words, first, visually enhanced facade images were produced by completing the missing parts using Gemini Flash 2.5; then, these enhanced images were fed as input to the Hunyuan3D 2.1 model to obtain the final 3D models. Throughout all experiments, deterministic settings were preferred where possible to keep randomness under control (e.g., the seed was fixed in the Hunyuan3D code, and no separate random seed value was specified in Gemini Flash API calls, since the same image and the same request generally yield similar results). Therefore, if the experiments reported here were to be run again, it is expected that 3D models consistent in terms of general form and material properties, except for minor random differences, would be obtained. The general appearance of the sixteen 3D building models obtained because of this workflow is presented in Figure 3.

Figure 4. Three-step workflow of the proposed method illustrated using the Bruno Taut Villa: (left) original building photograph, (centre) isometric daylight image generated by Gemini Flash 2.5, (right) textured 3D building model reconstructed by Hunyuan3D 2.1.

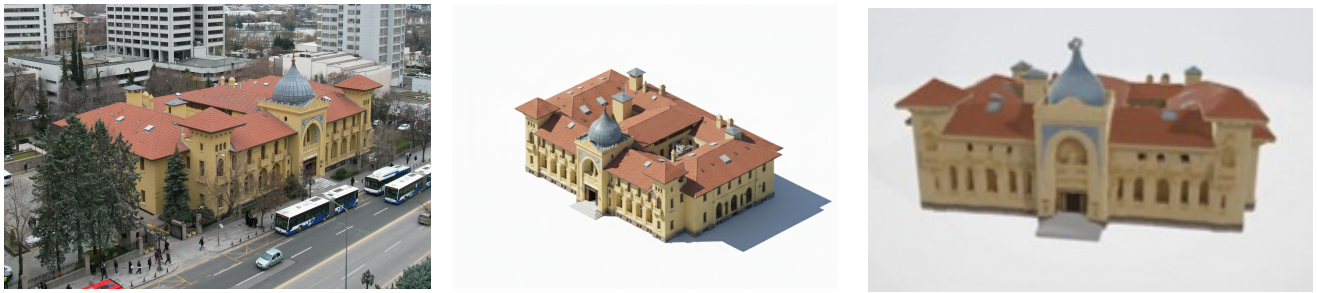

Figure 5. Three-step workflow of the proposed method illustrated using the Ankara Palace: (left) original building photograph, (centre) isometric daylight image generated by Gemini Flash 2.5, (right) textured 3D building model reconstructed by Hunyuan3D 2.1.

Figure 6. Three-step workflow of the proposed method illustrated using the Safranbolu House: (left) original building photograph, (centre) isometric daylight image generated by Gemini Flash 2.5, (right) textured 3D building model reconstructed by Hunyuan3D 2.1.

Figure 7. Three-step workflow of the proposed method illustrated using the İkinci Evkaf Apartment: (left) original building photograph, (centre) isometric daylight image generated by Gemini Flash 2.5, (right) textured 3D building model reconstructed by Hunyuan3D 2.1.

3.4 Evaluation of Reconstructed 3D Models

In the final step, the resulting 3D model files were examined and evaluated qualitatively and technically. Each model produced in this study was opened and viewed in software such as Autodesk Revit, ArcGIS Pro, and CloudCompare. Revit, a BIM-based tool, can be used to check the architectural scale and dimension compatibility of the models. ArcGIS Pro, on the other hand, is suitable for testing models in a geographic environment, performing experiments such as georeferencing and placement within an urban model.

To make this evaluation more systematic, the generated models were examined through a structured qualitative framework consisting of five criteria: (1) visual fidelity to the input image, (2) preservation of overall massing and proportions, (3) legibility of distinctive architectural features such as roofs, arches, projections, or façade rhythm, (4) interoperability and usability in common AEC/GIS environments, and (5) the extent of simplification or AI-based prediction of non-visible surfaces such as rear façades, inner courtyards, or roof areas. This framework was preferred because the study focuses on rapid single-image-based digital representation rather than survey-grade reconstruction, and because consistent ground-truth geometric data were not available for all sixteen cases. The case-based observations derived from this framework are summarized in Table 1.

The Hunyuan3D 2.1 model generated 3D entity outputs for each of the 16 structures in the study. The outputs were created in the default glTF (.glb) format and included both triangular mesh geometry and PBR-based texture maps. File sizes varied depending on the geometry of the structures, remaining approximately 4–5 MB. The generated .glb files were opened and examined in the open-source CloudCompare software, which allows for detailed analysis of point cloud and mesh datasets. In this environment, technical characteristics such as the triangular mesh structure, number of vertices, and accuracy of texture mapping of the models were evaluated, and it was observed that both the geometry and texture maps were processed as expected; an example of this process is shown in Figure 8.

Furthermore, thanks to the. glb format being an industry standard, the files can be opened seamlessly in commonly used software such as Blender, Autodesk 3ds Max, and Unity; this demonstrates that the models can be directly integrated into different platforms and workflows. If needed, it is also possible to convert the models to other formats such as OBJ, PLY, STL, or FBX. This conversion can be done via Hunyuan3D's output parameters or later through software such as Blender/MeshLab; however, care must be taken to preserve scale units during conversion, and additional scaling steps may be applied if necessary to ensure compatibility with real-world scales.

The general overview of the dataset of 16 structures used in the study and the 3D models generated from these structures is presented in the relevant figures (see Figures 3-7). Furthermore, the three stages of the workflow are visualized using four selected example structures: the original building photograph, the isometric daytime image obtained with Gemini Flash 2.5, and the textured 3D model

generated by Hunyuan3D 2.1. These visual comparisons show that basic architectural elements such as arches, columns, and timber joinery are largely accurately reproduced; however, because parts not directly visible in the photographs (rear facade, inner courtyard, etc.) are predicted and modelled by artificial intelligence, the generated 3D models are not exact copies of the original buildings, but rather partially predictive digital representations. For some selected structures, the generated models were also checked for scale and proportion by comparing them with existing drawing or measurement data. Applying this workflow to all selected buildings resulted in a series of case studies demonstrating the use of the proposed method.

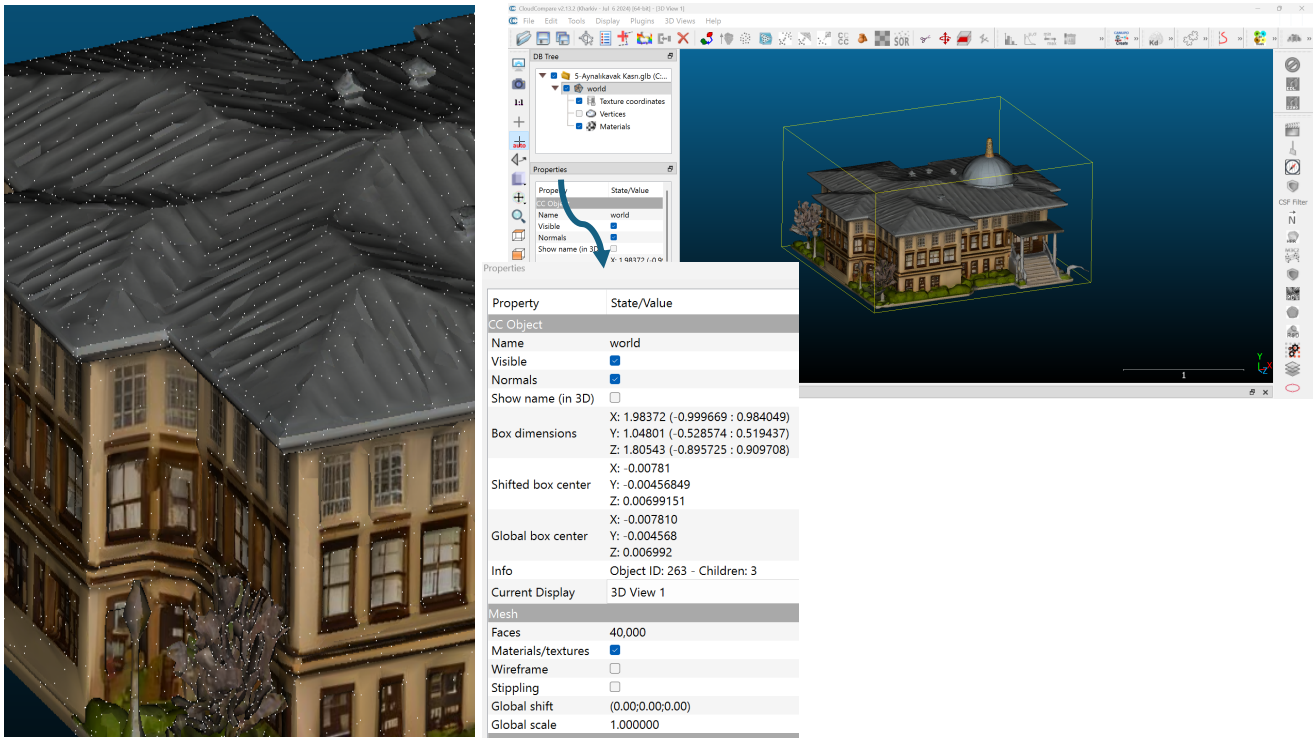

Figure 8. Evaluation of the reconstructed 3D model of Aynalıkavak Pavilion in CloudCompare.

4 Results

A summary of the sixteen case studies is presented in Table 1 in order to provide a more systematic overview of the input image conditions, processing workflow, output characteristics, and the main geometric limitations observed in the generated models.

Table 1. *Qualitative summary of the sixteen case studies, including input image characteristics and limitations, and the main geometric limitations observed in the generated outputs*

Case	Building	Input image characteristics and limitations	Main observed geometric limitations in the generated output
1	Ankara Palas (Ankara Palace)	Daylight, elevated oblique urban view; moderate resolution; dense urban background and limited façade detail visibility.	Fine façade ornament, roof junctions, and courtyard-facing surfaces appear simplified or partly inferred.
2	TBMM Müzesi (Turkish Grand National Assembly Museum)	Daylight, elevated oblique view; moderate resolution; relatively clear massing but limited close-up façade detail.	Entrance detailing, repetitive window articulation, and less visible roof areas are generalized.

3	Ankara Cumhuriyet Müzesi (Ankara Republic Museum)	Daylight, elevated oblique view; moderate resolution; partial shadowing and some visual interference from surrounding vegetation.	Arch profiles, façade ornament, and side/rear surfaces are simplified or inferred.
4	Gazi Eğitim Enstitüsü (Gazi Education Institute)	Daylight, high oblique aerial view; moderate resolution; distant capture reduces visibility of finer façade features.	Repetitive façade modules, roof intersections, and secondary masses are generalized.
5	Aynalıkavak Kasrı (Aynalıkavak Palace)	Daylight, ground-level oblique view; relatively good visibility and resolution; limited background interference.	Fine timber detailing, roof–dome transitions, and secondary side surfaces are simplified.
6	Baruthane	Daylight, high aerial wide shot; distant capture; complex surrounding context and reduced detail at building scale.	Inner-courtyard geometry, wall-depth perception, and peripheral landscape elements are partly inferred.
7	Baruthane Kulesi (Baruthane Tower)	Daylight, frontal-oblique view; the object occupies a relatively small part of the frame; limited material detail visibility.	Railings, antenna geometry, and surface detailing are simplified.
8	Bruno Taut Evi (Bruno Taut House)	Daylight, oblique view; partial occlusion by trees; moderate resolution and limited visibility of lower-level details.	Terrace railings, undercroft geometry, and fine window-frame articulation are generalized.
9	Doğan Apartmanı (Doğan Apartment)	Daylight, high aerial urban view; dense surrounding fabric; complex courtyard geometry; moderate resolution.	Courtyard depth, repetitive façade modules, and roof/parapet details are simplified or inferred.
10	Florya Atatürk Deniz Köşkü (Florya Atatürk Sea Pavilion)	Daylight, aerial oblique view over water; distant capture; slender structural elements and transparent components are difficult to resolve.	Slender supports, railings, glazed details, and underside/sea-facing elements are simplified.
11	Frej Apartmanı (Frej Apartment)	Daylight, street-level corner view; relatively good resolution; some perspective distortion due to upward viewing angle.	Fine façade ornament, fenestration profiles, and rooftop details are generalized.
12	Arif Paşa Apartmanı (Arif Pasha Apartment)	Daylight, aerial oblique urban view; dense context; complex massing and partial visual interference from adjacent buildings.	Courtyard-facing elevations, repetitive windows, and roof intersections are simplified; courtyard vegetation appears inferred.
13	Narmanlı Han (Narmanlı Han)	Daylight, aerial oblique view; dense urban background; interior court not fully visible in the selected image.	Inner-courtyard façades, rear-side details, and some roof articulations are predicted rather than directly observed.
14	Pera Palas (Pera Palace)	Daylight/overcast, street-level frontal-oblique view; relatively	Cornices, mansard detailing, awnings, and side/rear elevations are simplified.

		good resolution; vertical perspective distortion present.	
15	Safranbolu Evi (Safranbolu House)	Daylight, elevated oblique view; moderate-to-good visibility; some background vegetation and limited close-detail resolution.	Timber joinery, window lattice/detailing, and lower-base articulation are generalized.
16	İkinci Evkaf Apartmanı (İkinci Evkaf Apartment)	Daylight, aerial oblique urban view; moderate resolution; dense surrounding context and limited visibility of inner façades.	Courtyard-facing façades, balcony/railing details, and roofscape elements are simplified or inferred.

As a summarized in Table 1, across the sixteen cases, the workflow consistently preserved overall massing, roof morphology, courtyard configuration, and major stylistic cues, while fine façade ornament, repetitive window detailing, railings, and non-visible rear or inner-court surfaces were more frequently simplified or inferred. The quality of the generated outputs was also influenced by the conditions of the selected input photographs, particularly image resolution, viewing distance, occlusion, lighting, and surrounding visual complexity. In operational terms, Gemini Flash 2.5 typically processed each input image within a few seconds, while Hunyuan3D 2.1 generated the corresponding 3D model in approximately 1–2 minutes; the outputs were generally exported in .glb format and typically ranged around 4–5 MB, depending on model geometry. These workflow-level values are reported as approximate ranges rather than exact per-case measurements.

The outputs of the applied method show that the model largely successfully captures the general form and characteristic architectural features of each historical building examined. In early Republican-era examples (e.g., a government building from the 1920s), the building's facade composition, window arrangements, and the general geometry of the roof are distinctly preserved in the 3D models. In Ottoman-era civil architecture examples (e.g., a mansion from the 19th century), it was observed that period-specific facade elements such as wooden projections and bay windows could be read on the model. These results are particularly noteworthy for a system working with a single image; as they indicate that the artificial intelligence model, trained with deep learning on a large data pool, has partially ‘learned’ and become capable of reproducing the formal and stylistic features of different architectural typologies. Nevertheless, it should be considered that parts not directly visible in the photographs, such as rear facades or the hidden surfaces of the roof, are modelled by artificial intelligence based on predictions, and therefore the generated 3D models may not perfectly match the original buildings in these areas.

When evaluated in terms of architectural accuracy, the method was found to successfully capture the main mass and proportions of the structure, although discrepancies in fine details were observed in places. For example, in the model of a building from the Early Republican period, window frames were represented as simpler geometries instead of original profiles; in the example of an Ottoman-era mansion, wooden ornaments and fine joinery details were expressed in more general outlines. This is partly an expected result, as the AI model complements details that are not clearly discernible in the input image with general patterns "learned" from the large dataset on which it was trained. Nevertheless, the general observation is that both models (Gemini Flash and Hunyuan3D) produce consistent outputs in terms of architectural style: when modelling a modernist reinforced concrete structure, the resulting 3D object retained its modernist character; when modelling a traditional Ottoman house, distinctive elements such as the sloping roof and latticed windows were legibly included in the generated 3D model.

When evaluated in terms of performance and speed, the proposed approach is seen to provide a significant advantage compared to traditional methods. The Gemini Flash 2.5 model, which operates as a cloud-based service, can process a single image in just a few seconds; the Hunyuan3D 2.1 model, on

the other hand, completes the production of a 3D model from a single image via its online interface in an average of 1–2 minutes. In multi-image-based methods such as photogrammetry, achieving a similar level of detail requires taking dozens of photographs, aligning them, and performing steps such as generating a dense point cloud and mesh, so the process can often take hours. In contrast, the workflow proposed in this study can produce a usable 3D asset from a single photograph in minutes; this makes the method extremely efficient, especially in situations requiring rapid prototyping or scenarios where many structures need to be modelled sequentially.

5 Discussion

The results obtained demonstrate that the proposed method is a promising tool for producing digital architectural representations. Its most significant advantage is the ability to create a three-dimensional model with extremely limited data – in most cases, a single photograph. This is particularly important when it comes to historical structures for which only a few photographs exist in visual archives or which have completely disappeared today. For example, generating an approximate 3D reconstruction from a single available facade photograph of a demolished building becomes possible thanks to such AI-based approaches. Thus, new possibilities emerge for the digital revival of structures that have physically disappeared and for the transmission of cultural heritage to future generations. As emphasized in the literature, such digital representations offer an important starting point for the planning, preservation, exhibition, and educational applications of cultural heritage (Münster et al., 2024).

However, the method has some limitations and aspects that require attention. First, since the generated model is entirely "estimated" by artificial intelligence, it should not be considered an exact replica of the real structure. The model adapts, especially the unseen parts and fine details, based on its own learning data. Therefore, the resulting 3D model should be considered not a linear digital document of cultural heritage, but a kind of AI-assisted visualization. In academic research or restoration projects, the outputs of this method should be used for preliminary visualization and providing ideas, rather than for precise measurements or plans. For example, a hypothetical superstructure proposal can be shown on an archaeological ruin with the help of these models, but such a model should not be presented as a one-to-one representation of scientific reality.

Another aspect of the method is the optimization and integration of the outputs. When the mesh models generated by Hunyuan 3D are directly imported into game engines or real-time 3D applications, their high surface counts can lead to performance issues. However, because the model is open-source and allows for the editing of the outputs, lighter versions can be obtained through polygon reduction (decimation) or retopology operations if necessary. Furthermore, PBR texture maps can be user-edited or replaced with different, higher-resolution textures. This flexibility expands the practical application areas of the method.

The fundamental differences between classical photogrammetry and the method used in this study are also worth discussing. Photogrammetry is superior in terms of geometric accuracy and scale because it works based on real-world data; however, it requires many photographs and intensive labour. The approach of generating a model from a single photograph offers an alternative to photogrammetry in cases of data constraints. Especially in historical photographic archives, if only one image of a structure from a single angle is available, a three-dimensional representation can be obtained using this method. However, it is not feasible for photogrammetry to be completely replaced in engineering or conservation studies requiring high accuracy. Future evaluation of hybrid approaches would be appropriate: for example, if multiple historical photographs are available, the Hunyuan 3D model may yield more consistent results with multiple image inputs, or the AI-generated model could be combined with a laser-scanning point cloud to fill in missing parts.

From a cultural and educational perspective, this method also has the potential to present architectural heritage to a wider audience. The 3D models produced can be used in web-based 3D viewers or virtual/augmented reality applications. Thus, for example, a digital reconstruction of an Ottoman-era

building can be shown interactively to museum visitors or students. These models can also be used as a visual tool in urban planning or tourism promotion. It is important to explain both the capabilities and limitations of the artificial intelligence tools used in creating this content to the target audience; that is, it should be clearly stated that the model may not perfectly reflect reality but can provide a general idea.

6 Conclusions

This study presents a method for generating 3D digital models from architectural building photographs using two AI-based models. The Gemini Flash 2.5 model was used to isolate building photographs and adapt them to an isometric view thanks to its advanced editing capabilities on 2D images; then, the Hunyuan 3D 2.1 model automatically generated the three-dimensional shape and surface textures of the structure based on this processed image. Experiments conducted on selected examples from Turkish architectural heritage showed that the method generally yielded consistent results across different styles and periods. The ability to obtain a 3D model even from a single photograph offers new opportunities in the field of digital cultural heritage. These opportunities include the reconstruction of historical artifacts with limited visual documentation, the re-evaluation of archival photographs, and their use as a visualization tool in architectural history research.

Unlike traditional single-image 3D reconstruction approaches that rely on a single model to infer geometry directly from the original input, the proposed workflow separates visual normalization from 3D generation. In this respect, it also differs from multi-view-based AI generation, which typically depends on multiple photographs or synthetically expanded views to improve geometric consistency. The main advantage of the present approach lies in its preprocessing–reconstruction decoupled design: by first isolating the building, normalizing lighting conditions, and approximating a near-isometric view, the workflow reduces visual ambiguity before mesh generation. This sequential structure improves input consistency, supports more controllable and repeatable outputs, and offers a practical solution for heritage cases where only a single archival or opportunistic photograph is available.

However, it should be emphasized that the accuracy of the method's outputs must be carefully considered. AI-generated models are generalized products of the datasets on which they are trained and do not completely replace the real object. Therefore, in matters requiring scientific or technical precision, it is appropriate to use these models as a supporting visual tool, not as a definitive reference. In the future, the accuracy and scope of the method can be increased with improvements such as integrating multiple images and training AI models specifically for architectural details (e.g., better recognition of elements of a specific historical period's architecture). Furthermore, similar approaches can be extended to 3D city models or interior architecture at the urban scale.

In conclusion, the method developed through the sequential use of Gemini Flash 2.5 and Hunyuan 3D 2.1 models offer an innovative approach to digital architectural representation production, both for academic research and for the digital preservation of cultural heritage.

Ethical Approval Declaration
Not Applicable

Informed Consent Statement
Not Applicable

Acknowledgements
The authors would like to thank all individuals and institutions who supported this study.

Funding
This research received no external funding.

Data Availability Statement
The data are not publicly available due to copyright and usage restrictions.

Conflicts of Interest
The authors declare no conflict of interest.

Appendix A. Detailed case list for the input dataset and Gemini outputs
The case numbering used in Figures 1 and 2 is presented below for ease of reference.

Case 1: Ankara Palas (Ankara Palace)
Case 2: TBMM Müzesi (Turkish Grand National Assembly Museum)
Case 3: Ankara Cumhuriyet Müzesi (Ankara Republic Museum)
Case 4: Gazi Eğitim Enstitüsü (Gazi Education Institute)
Case 5: Aynalıkavak Kasrı (Aynalıkavak Palace)
Case 6: Baruthane (Baruthane)
Case 7: Baruthane Kulesi (Baruthane Tower)
Case 8: Bruno Taut Evi (Bruno Taut House)
Case 9: Doğan Apartmanı (Doğan Apartment)
Case 10: Florya Atatürk Deniz Köşkü (Florya Atatürk Sea Pavilion)
Case 11: Frej Apartmanı (Frej Apartment)
Case 12: Arif Paşa Apartmanı (Arif Pasha Apartment)
Case 13: Narmanlı Han (Narmanlı Han)
Case 14: Pera Palas (Pera Palace)
Case 15: Safranbolu Evi (Safranbolu House)
Case 16: İkinci Evkaf Apartmanı (İkinci Evkaf Apartment)

References

AINews. (2025, May 20). Google I/O: New Gemini native voice, Flash, DeepThink, AI Mode (DeepSearch+Mariner+Astra). *News.Smol.ai*. Retrieved from https://news.smol.ai/issues/25-05-20-google-io/, Last Access: December 6, 2025.

Echo3D. (2025, August 14). A guide to creating and sharing 3D models with Hunyuan-3D-2.1 and echo3D [Blog post]. *Medium*. Retrieved from https://medium.com/echo3d/a-guide-to-creating-and-sharing-3d-models-with-hunyuan-3d-2-1-and-echo3d-da99f677b740, Last Access: December 12, 2025.

Fortin, A., Vernade, G., Kampf, K., & Reshi, A. (2025, August 26). Introducing Gemini 2.5 Flash Image, our state-of-the-art image model. *Google Developers Blog*. Retrieved from https://developers.googleblog.com/en/introducing-gemini-2-5-flash-image/, Last Access: December 9, 2025.

Hunyuan3D. (2025). *Advanced 3D asset generation with AI*. Retrieved from https://hunyuan-3d.com/, Last Access: December 14, 2025.

Li, W., Liu, J., Yan, H., Chen, R., Liang, Y., Chen, X., Tan, P., & Long, X. (2024). CraftsMan3D: High-fidelity mesh generation with 3D native generation and interactive geometry refiner. *arXiv preprint arXiv:2405.14979*. https://doi.org/10.48550/arXiv.2405.14979

Li, Y., Zou, Z. X., Liu, Z., Wang, D., Liang, Y., Yu, Z., Liu, X., Guo, Y. C., Liang, D., Ouyang, W., & Cao, Y. P. (2025). TripoSG: High-fidelity 3D shape synthesis using large-scale rectified flow models. *arXiv preprint arXiv:2502.06608*. https://doi.org/10.48550/arXiv.2502.06608

Montas-Laracuente, N., Delgado Martos, E., Pesqueira-Calvo, C., Intra Sidola, G., Maitín, A., Nogales, A., & García-Tejedor, Á. J. (2025). Automatic 3D reconstruction: Mesh extraction based on Gaussian splatting from Romanesque–Mudéjar churches. *Applied Sciences*, *15*(15), 8379. https://doi.org/10.3390/app15158379

Münster, S., Maiwald, F., di Lenardo, I., Henriksson, J., Isaac, A., Graf, M. M., Beck, C., & Oomen, J. (2024). Artificial intelligence for digital heritage innovation: Setting up a R&D agenda for Europe. *Heritage*, *7*(2), 794–816. https://doi.org/10.3390/heritage7020038

Saadat, A., Aziz, S., Mahmud, S., Mahi, A. I. M., & Ahmed, S. (2025). VisionTrap: Unanswerable questions on visual data. *arXiv preprint arXiv:2507.17262*. https://doi.org/10.48550/arXiv.2507.17262

SysPsych. (2025). Hunyuan 3D 2.1 released today – Model, HF Demo, GitHub links on X. *r/StableDiffusion*. Retrieved from https://www.reddit.com/r/StableDiffusion/comments/1laq8he/hunyuan_3d_21_released_today_model_hf_demo_github/, Last Access: December 18, 2025.

Team Hunyuan3D, Yang, S., Yang, M., Feng, Y., Huang, X., Zhang, S., He, Z., Luo, D., Liu, H., Zhao, Y., Lin, Q., Lai, Z., Yang, X., Shi, H., Zhao, Z., Zhang, B., Yan, H., Wang, L., Liu, S., Zhang, J., Chen, M., Dong, L., Jia, Y., Cai, Y., Yu, J., Tang, Y., Guo, D., Yu, J., Zhang, H., Ye, Z., He, P., Wu, R., Wei, S., Zhang, C., Tan, Y., Sun, Y., Niu, L., Huang, S., Zheng, B., Liu, S., Chen, S., Yuan, X., Yang, X., Liu, K., Zhu, J., Chen, P., Liu, T., Wang, D., Liu, Y., Linus, Jiang, J., Huang, J., & Guo, C. (2025, June 18). Hunyuan3D 2.1: From images to high-fidelity 3D assets with production-ready PBR material. *arXiv*. https://doi.org/10.48550/arXiv.2506.15442

Tencent-Hunyuan. (2025). *Hunyuan3D-2: High-resolution 3D assets generation with large-scale Hunyuan3D diffusion models*. GitHub. Retrieved from https://github.com/Tencent-Hunyuan/Hunyuan3D-2, Last Access: December 21, 2025.

Tencent Hunyuan3D Team. (2025). *Hunyuan3D-2.1: From images to high-fidelity 3D assets with production-ready PBR material*. GitHub. Retrieved from https://github.com/Tencent-Hunyuan/Hunyuan3D-2.1, Last Access: December 25, 2025.

Tencent Hunyuan3D Team. (2025). Hunyuan3D 2.0: Scaling diffusion models for high resolution textured 3D assets generation. *arXiv preprint arXiv:2501.12202*. https://doi.org/10.48550/arXiv.2501.12202

Wu, S., Lin, Y., Zhang, F., Zeng, Y., Yang, Y., Bao, Y., Qian, J., Zhu, S., Cao, X., Torr, P., & Yao, Y. (2025). Direct3D-S2: Gigascale 3D generation made easy with spatial sparse attention. *arXiv preprint arXiv:2505.17412*. https://doi.org/10.48550/arXiv.2505.17412

Xiang, J., Lv, Z., Xu, S., Deng, Y., Wang, R., Zhang, B., Chen, D., Tong, X., & Yang, J. (2025). Structured 3D latents for scalable and versatile 3D generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 21469–21480). https://doi.org/10.48550/arXiv.2412.01506

Zhao, Z., Liu, W., Chen, X., Zeng, X., Wang, R., Cheng, P., Fu, B., Chen, T., Yu, G., & Gao, S. (2023). Michelangelo: Conditional 3D shape generation based on shape-image-text aligned latent representation. *arXiv preprint arXiv:2306.17115*. https://doi.org/10.48550/arXiv.2306.17115