



Università
Ca'Foscari
Venezia

Corso di Laurea magistrale
in Digital and Public Humanities

Tesi di Laurea

**3D Computer Vision and Structure from Motion
to analyse Udine's ancient wells: a comparison
between COLMAP and TRELLIS**

Relatore

Ch. Prof. Filippo Bergamasco

Correlatore

Ch. Prof. Fabio Pittarello

Laureanda

Matilde Moretti
907785

Anno Accademico

2025/2026

Ai miei genitori

*Che hanno continuato a crederci
fino a mettere le mani in pasta*

Index

1. Introduction	2
1.1. Udine's ancient wells.....	2
1.2. Computer Vision and Structure from Motion (SfM)	5
2. Comparing COLMAP and TRELLIS: a description of the methodologies.....	11
2.1. The wells.....	11
2.2. COLMAP and TRELLIS.2 v2 techniques.....	13
2.3. Creating the dataset of images.....	21
2.4. 3D imaging and rendering	24
2.5. Comparing the results.....	38
3. Online visualizer for the 3D models.....	45
References	47

1. Introduction

1.1. Udine's ancient wells

The founding of the city of Udine is closely linked to water, or rather to its absence, as the city is located in an area of “barren plain”¹ in the Friuli-Venezia Giulia region; precisely for this reason, the ways in which the “canals reached Udine and who dug the five ancient wells”² have been widely discussed. Specifically, their purpose was to ensure a constant supply of drinking water that could be used for both public and private purposes.

The locations of the five ancient wells of Udine are well documented:

1. in the courtyard of Palazzo Filitini, Caimo, and Dragoni on Via Manin: in 1493, the municipality granted the noble Filitini family permission to construct buildings around the well, on the condition that they install a “gate opening onto the square to allow free public access”³. The well disappears from the municipal records, only to reappear after 1590 “in which it is mentioned as a means of supplying water to the public fountains.”⁴ ;
2. in Piazza Giacomo Matteotti (also known as Piazza Mercato Nuovo), near the corner formed by Via delle Erbe and Via Mercerie: no records have survived indicating when the well ceased to be used or when it was actually covered, however, Federico Braidotti, author and public official from Udine who lived between 1843 and 1922, hypothesised that this may have occurred at the same time as the construction of the cistern located beside the Church of San Giacomo;
3. on Via Savorgnana, in the small square known as Piazza dei Teatri: it was called the Well of San Giovanni, after the ancient name of that district, and the humanist Gregorio Amaseo “while recounting the massacres that took place in Udine on

¹ F. BRAIDOTTI, “I primi pozzi di Udine”, in *Pagine Friulane* n.2, 1891, *Patria del Friuli*, p.17, [“arida pianura”]

² A. FRANGIPANE “La costruzione di Udine: tracce di lettura lungo le acque, le mura, le case”, *Cultura in Friuli III* 2017, Società Filologica Friulana, p.1, [“rogge siano arrivate a Udine e su chi abbia scavato gli antichi cinque pozzi”]

³ F. BRAIDOTTI, “L'acqua potabile a Udine : lettura”, in *Atti della Accademia di Udine*, serie 2, v. 9 (1890/1893), p. 5, [“porta verso la piazza per libero accesso del pubblico”]

⁴ *Ibidem*, [“in cui se ne parla come mezzo per fornir d'acqua le pubbliche fontane”]

Shrove Thursday in 1511”⁵ recounts the grim memory associated with it. Alvise da Porto, a member of the Savorgnan faction, “wishing to rid himself of the assassins Vergon and Bernardino de Narni, accomplices in his many acts of treachery, and also out of greed for the spoils they had taken, [...] he strangled them while they slept and then had them thrown into the nearby, exceedingly deep Well of San Giovanni, together with a maidservant of Pietro Urbano, who was cast in alive because she had witnessed them being thrown in, lest she reveal what had happened”⁶. The bodies were discovered on 28 March 1511, after the well was opened by order of Bernardino Manin, the municipal treasurer: it was believed, in fact, that releasing the fumes from the wells could help alleviate the problem of the earthquakes that were devastating the region in those years. It can therefore be inferred that, as early as 1511, that well “was not used as a source of drinking water”⁷;

4. in Piazza S. Cristoforo: this well was covered in 1601 by order of the Greater Council, however, in 1806, Prefect Somenzari decreed that water could once again be drawn from it (an event that was commemorated with an inscription carved inside the structure). Today, the well is no longer located in the centre of the square, in fact, the original structure is kept in the municipal storages.
5. on via Cisis: in 1535, the Convocation granted the neighbouring resident, Giuseppe Ettoreo, permission to build a wall near the well, while requiring him to construct a gate through which citizens could access it.

To these, we can add the well belonging to the ancient complex of the Hospital of Santa Maria della Misericordia in Udine, located in Piazza XX Settembre (also known as Piazza dei Grani).

⁵ *Ibidem*, p.6, [“narrando le stragi udinesi del Giovedì grasso 1511]

⁶ *Ibidem*, [“volendo liberarsi dei sicari Vergon e Bernardino de Narni, complici dei molti suoi tradimenti, ed anche per cupidigia della preda fatta [...] li strangolò dormienti e poi li fè gettare nel vicino profundissimo pozzo di S. Giovanni et con loro una fantesca, viva, di Pietro Urbano, per aver vistoli precipitar, accio la non discoprisse tal cosa”]

⁷ *Ibidem*, [“non serviva per gli usi potabili”]

However, it was precisely the great depth of the wells (approximately 65 metres) that made them difficult to use, and with the demographic increase of the population “especially after the transfer of the patriarchal residence to Udine”⁸, it became necessary to devise more effective methods of supplying the city with drinking water. As a result, the construction and use of cisterns gradually became the preferred solution.

⁸ *Ibidem*, p. 7, [“specialmente dopo il trasferimento della residenza patriarcale in Udine”]

1.2. Computer Vision and Structure from Motion (SfM)

Computer Vision is the discipline that “studies how to enable computers to understand and interpret the visual information contained in images and videos”⁹. “Computer vision trains machines to perform object identification, distance, and movement using cameras, data, and algorithms instead of retinas, optic nerves, and the visual cortex”¹⁰, and it’s a process that can analyse thousands of products per minute, surpassing human capabilities by far. The first studies regarding Computer vision started in the late 1960s with the aim to mimic the human visual system in order to “enable robots to behave intelligently”¹¹. In the 1990s, there was a further advancement in the field thanks to research focusing on projective and sparse 3D reconstruction, camera calibration, dense stereo correspondence, and image segmentation.

Today, computer vision is a crucial technique in numerous fields, such as medicine, industry, defence, and autonomous vehicles. In medicine, for instance, “it is used for diagnosing patients, enhancing images interpreted by humans, and supporting research”¹²; in industry, instead, it is used for quality control and inspection of the final products; meanwhile, in defence applications, it aims to detect enemy forces and missile guidance.

The way computer vision works is by preprocessing an image and analysing it for details like edges, shapes, or colours; after doing that, several things happen:

1. the system “trains the model using thousands of positive and negative [...] pictures to learn unique patterns”¹³;
2. the object of a new image is predicted and classified thanks to the learned patterns;

⁹ V. BARRILE, G. BILOTTA, D. LAMARI, G. MEDURI, U. TRUNGADI, A. RICCIARDI, “Computer vision/structure from motion per la diffusione dei beni culturali”, in Atti della XIX Conferenza Nazionale ASITA, 2015, p. 52, [“studia come abilitare i computer alla comprensione e alla interpretazione delle informazioni visuali presenti in immagini o video”]

¹⁰ M.M. HASSAN, S. KARUNGARU, R. BASHAR, “A Short Review on Computer Vision: Visualizing the World Through Machine”, 2025, 10.20944/preprints202506.0630.v1, p. 2

¹¹ *Ibidem*.

¹² *Ibidem*.

¹³ *Ibidem*.

3. the result is outputted by labelling correctly the subject of the image analysed.

Computer vision processes images using various algorithms and techniques, some of which have important applications also in the field of Machine Learning:

- Convolutional Neural Networks (CNNs): they are “deep learning models designed to process and analyse data with a grid-like topology, such as images”¹⁴, which allows them to be particularly effective for classification, object detection, and segmentation thanks to their ability to “learn spatial hierarchies of features”¹⁵;
- Object Detection: this technique both identifies objects within an image and determines their position “through bounding boxes”¹⁶. There are two popular methods for object detection: R-CNN and YOLO. The first one, *in primis*, generates possible areas where an object might be located and then classifies these regions using convolutional neural networks; meanwhile, the second one processes the entire image in a single instance, it divides it into a grid and “predicts the bounding boxes and class probabilities for each cell”¹⁷;
- Semantic Segmentation: this task classifies each pixel of an image into a predefined category, therefore, all pixels belonging to the same object or section are assigned to the same class. Unlike object detection, which creates bounding boxes around objects, semantic segmentation assigns a label to every pixel, which allows us to understand the scene with more depth;
- Generative Models: they are machine learning models whose aim is to “learn the underlying distribution of data and generate new data samples that resemble the original dataset”¹⁸. In comparison with discriminative models, generative models apply a different approach, in fact, while the first ones focus on predicting labels for given input, generative models attempt to generate new data instances similar to the training set by modelling the joint probability distribution;

¹⁴ *Ibidem*, p. 10

¹⁵ *Ibidem*.

¹⁶ *Ibidem*, p. 12

¹⁷ *Ibidem*.

¹⁸ *Ibidem*, p. 13

- Vision Transformers (ViTs): with this deep learning model, transformer architectures are applied to computer vision tasks. Unlike CNNs that rely on pooling to process pictures, ViTs “breaks an image into a series of fixed-size non-overlapping patches and treat each patch as a token”¹⁹. In the following instance, the patches are embedded into a sequence and conveyed to a transformer model to capture the spatial relationships and the dependencies between them.

Other techniques, like Structure from Motion, for example, are methodologies purely algorithmic, ultimately based on signal analysis:

- Edge Detection: “edges are the lines where colours and brightness change sharply in an image”²⁰, therefore, detecting these elements allows the computer to identify the shape of objects portrayed in it;
- Histogram of Oriented Gradients (HOG): this technique grants the computer the ability to detect objects in an image, meaning that if in a picture there is a face or a human figure, this feature “captures the shape of objects based on pixel intensities, focusing on edges and contours”²¹;
- Feature Extraction: features are unique patterns in an image, such as lines, corners, and textures, and computers pay attention to these elements to recognise objects. This type of association is critical for object recognition, because “instead of comparing every pixel in an image, a computer looks for key features that stand out and give clues about the object”²²;

It becomes clear that one of the fundamental goals of computer vision is to recover the 3D structure of a scene from a series of images. In this context, a particularly effective approach to 3D reconstruction involves the use of many images of a stationary scene and the application of digital photogrammetry with Structure from Motion (SfM) algorithms.

¹⁹ *Ibidem* p. 14

²⁰ *Ibidem*, p. 3

²¹ *Ibidem*, p. 4

²² *Ibidem*, p. 5

SfM allows “to form 3D point clouds by taking numerous multiview images with relatively inexpensive non-metric cameras taken by the ground or the air”²³. In addition, this technique “is based on the use of passive sensors and generates accurate metric and geometric information by taking into account camera calibration and image orientation”²⁴.

Many characteristics that define Structure from Motion (highly automated low-cost software, the use of digital cameras easily available to the consumer, handling generic image network configuration, and an easy implementation that doesn’t require specialist knowledge) stand in contrast to traditional photogrammetry, where expensive equipment and trained operators are required, in addition to costly specialist software. A downside of SfM is that the automation of the software (the reconstruction of the scene with the positioning and the orientation of the camera are automatically resolved by the software itself) “often depends on great overlap between adjacent images to perform image matching”²⁵, and changes in lighting conditions, shadows, and lack of texture can lead to issues.

Ultimately, the purpose of using Structure from Motion is to get a “cloud of 3D points in the scene, which in the process of feature matching and meshing, give the final 3D model”²⁶ (Fig. 1-2); therefore, given the position of an object in one picture, we need to find the corresponding position of the same object in the successive ones. This is called the Correspondence Problem, which is based on principles of multiple view geometry

²³ S. GRANSHAW, “Structure from motion: origins and originality”, in *The Photogrammetric Record* 33, 2018, p. 1

²⁴ A. BOSCO, E. MINUCCI, “Rendering rti ed editing d'immagine per elaborazioni SfM: confronto tra tecniche e strumenti di visualizzazione per la documentazione di graffiti in contesto archeologico” in *Newsletter di Archeologia CISA* 11, 2020, Centro Interdipartimentale di Servizi di Archeologia, p. 45, [“si basa sull’utilizzo di sensori passivi e genera informazioni accurate, metriche e geometriche, tenendo conto della calibrazione della fotocamera e dell’orientamento delle immagini”]

²⁵ S. GRANSHAW, “Structure from motion: origins and originality”, in *The Photogrammetric Record* 33, 2018, p. 1

²⁶ S. MIJAKOVSKA, “Overview of Structure from Motion” in *Advanced Computing: An International Journal* 15, 2024, Academy and Industry Research Collaboration Center (AIRCC), p. 26

“from its projected position in an [...] image [...], a scene point can only be recovered up to one-parameter ambiguity corresponding to its distance from the camera”²⁷.

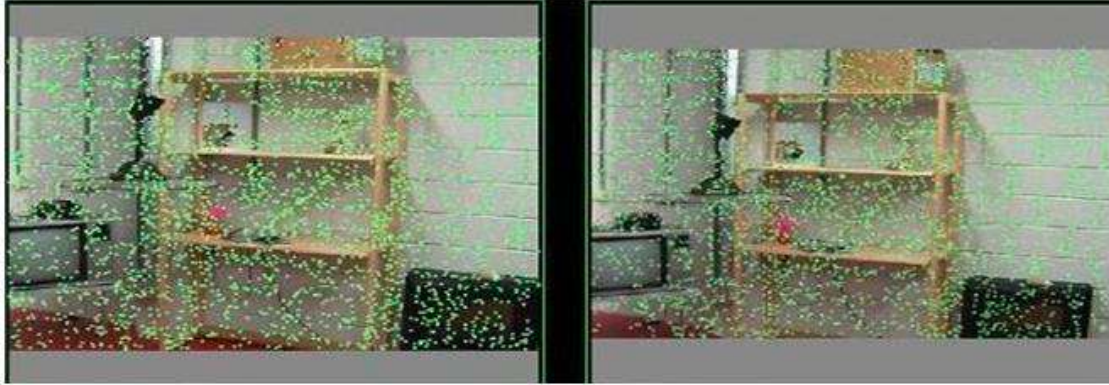


Fig. 1 Example of points of interest detected by the Scale Invariant Feature Transform (SIFT) algorithm implemented by D. Lowe in 2004



Fig. 2 Example of corresponding position of the same object in successive images

Moreover, in any image representing a real-world scene, there are numerous pixels that are surrounded by pixels of very similar intensity and colour (regions of homogeneous texture), which makes it virtually impossible to distinguish between the correct corresponding pixel and other nearby ones. Issues can also be caused by occlusion, where objects can be obstructed in following views by other parts that compose the scene, rendering them invisible.

Fortunately, the additional information we need to resolve the Correspondence Problem can be acquired in two different ways:

²⁷ *Ibidem*.

- use prior knowledge of the scene to reduce the number of degrees of freedom and reconstruct simple geometric shapes (line segments and polygons) from their projected position;
- use corresponding image points in multiple views, in fact, given an image in two or more views, the 3D point associated with it can be reconstructed by triangulation.

2. Comparing COLMAP and TRELLIS: a description of the methodologies

2.1. The wells

For this analysis, two of the six ancient wells in the city of Udine were taken into consideration. The first is the well of the former Hospital of Santa Maria della Misericordia in Piazza XX Settembre (Fig. 3), whose primary function was to ensure an independent supply of clean drinking water for the healthcare facilities in the area.



Fig. 3 Well in Piazza XX Settembre

The second is a considerably more unusual example: it is an octagonal stone well built near the Church of San Giacomo in 1487 (Fig. 4). It was commissioned by Nicolò Lipomanno, Lieutenant of the Patriarchate of Friuli, from “Lombard craftsmen who were in the city at the time due to the saturation of the Venetian market”²⁸. This curious structure was renamed ‘Diogenes’ Lantern’ by art historian Giovanni Battista Cavalcaselle, as its form resembles the famous monument in Athens: on the octagonal cistern, “four columns rise up to support the cover, above which a single isolated column stands”²⁹. On the base, decorated with coats of arms, the acronym C.VT.F.F. (*Civitas Utinensis Fieri Fecit*) can also be read.



Fig. 4 Well in Piazza San Giacomo

²⁸ L. MARIONI BROS, D. NOBILE, “La chiesa di San Giacomo in Mercatonuovo”, Deputazione di storia patria per il Friuli, 2008, p. 6 [“maestranze lombarde in quei tempi presenti in città a causa della saturazione del mercato veneziano”]

²⁹ *Ibidem*, [“s’impongono [...] quattro colonne che sostengono il coperchio sopra cui s’innalza una colonna isolata”]

2.2. COLMAP and TRELIS.2 v2 techniques

COLMAP is a free, open-source multi-view 3D reconstruction pipeline based on Structure from Motion (to generate sparse 3D point clouds alongside estimated camera poses) and Multi-View Stereo (to synthesise dense depth maps, dense 3D point clouds, and surface meshes) algorithms. It provides both graphical and command-line interfaces and “can reconstruct geometry from ordered or unordered multi-view image datasets”³⁰ by analysing spatial relationships across a collection of overlapping 2D photographs.

At its core, the pipeline (Fig. 5) enforces a clear distinction between physical optical sensors and captured image frames by defining a dedicated camera model abstraction. When processing datasets captured by a consistent device and focal length, COLMAP allows intrinsic parameters to be shared across multiple images, constraining the degrees of freedom during bundle adjustment and improving stability. In fact, each image entity in the project maps to a bitmap file and contains extracted keypoints alongside high-dimensional visual feature descriptors. When this step is performed, the descriptors are “stored as an array to describe a local feature of an image”³¹.

Then, since there is a need to convert these feature pairs into reliable spatial inputs, COLMAP subject them to geometric verification: only correspondences that pass a strict geometric threshold are classified as verified and forwarded to the reconstruction graph.

Furthermore, to handle complex acquisition scenarios like multi-camera setups, stereo rigs, or synchronised capture systems, COLMAP integrates explicit rig support. In fact, a camera rig defines a rigid structural assembly where relative spatial transformations among attached sensors remain strictly fixed over time. Therefore, instead of optimising extrinsic parameters for every individual image independently, COLMAP uses a frame:

³⁰ X. LI, S. YAN, Z. ZHU, C. ZHANG, Y. WU, F. MENG, L. YUN, "Relative Agreement and Feasibility of Low-Cost Passive Stereophotogrammetry for Nasal Three-Dimensional Surface Imaging: A Comparison of COLMAP and Agisoft Metashape with a Handheld Structured-Light Scanner" *Bioengineering*, 2026, 13, no. 8: 899. <https://doi.org/10.3390/bioengineering13080899>, p. 2

³¹ Team, TELKOMNIKA. “A Practical Approach to Candi Siwa 3D Reconstruction with COLMAP and Nerfstudio.” *TELKOMNIKA Telecommunication Computing Electronics and Control*, 2026, p. 250

a single temporal snapshot that groups all synchronised captures from the rig, allowing the optimisation algorithms to adjust a single rigid-body pose per frame.

The next step in the COLMAP pipeline is the use of the Structure from Motion algorithm, which is designed to reconstruct camera poses and sparse 3D geometry from verified image correspondences. The process begins with the construction of a scene graph, where the nodes and edges represent, respectively, images and verified feature matches. From the graph created, the algorithm selects an optimal initial two-view pair to seed the reconstruction (the pair needs to have a high density of overlapping matches combined with a large baseline to ensure a sufficient triangulation angle). After the first set of 3D points is triangulated successfully, new ones are continuously generated from overlapping registered views using a multi-view triangulation method.

Interestingly, COLMAP continuously performs aggressive filtering cycles throughout the incremental process to remove degenerate 3D points displaying errors or unrealistic triangulation angles: by doing that, structural accuracy and mathematical stability are maintained. Then, the result of the reconstruction is a “mesh file that is formed by point clouds”³² calculated incrementally through image registration, triangulation, and bundle adjustments.

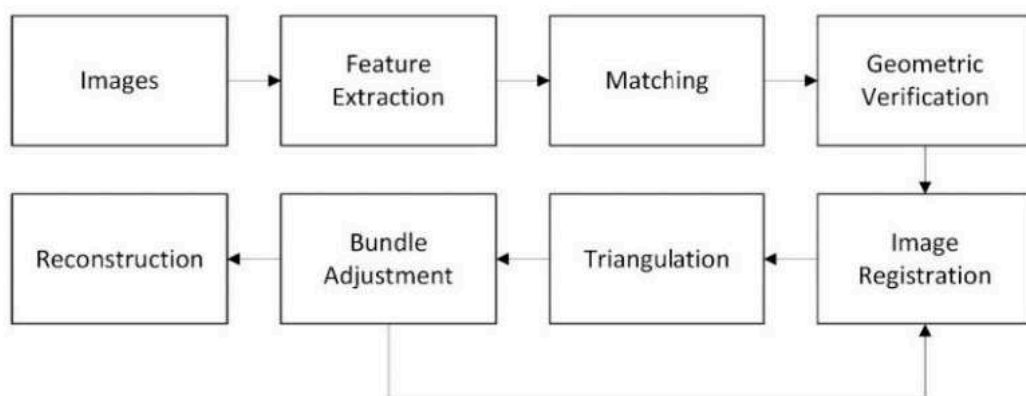


Fig. 5 COLMAP pipeline

³² *Ibidem.*

Ultimately, the primary strength of COLMAP resides in the ground-truth metric accuracy, since it reproduces real-world objects and physical dimensions with precision, but, on the other hand, surfaces that are occluded or unphotographed cannot be reconstructed, leading to holes or missing geometry.

TRELLIS.2 v2, instead, represents a modern paradigm “designed to generate 3D objects from diverse inputs, such as text prompts, multiple 2D views of an object, or another 3D object directly”³³ (Fig. 6). It uses deep learning architectures both to synthesise high-quality 3D assets from a single 2D image and to infer shapes not visible in the input image. While this is one of its primary strengths, this technique can hallucinate over exact measurements: in fact, generated unseen areas are guessed based on dataset priors rather than physical measurements of the original object.



Fig. 6 Example of a 3D asset generated with TRELLIS.2

³³ C. HERVÉ, P. GARNIER, J. VIQUERAT, E. HACHEM, “TRELLIS -enhanced surface features for comprehensive intracranial aneurysm analysis”, in Neuroscience Informatics, 2026, vol.6 issue 1, pp. 2-3

Unlike traditional pipeline-based 3D generators, which in most cases focus on 3D shape generation while “neglecting the appearance and material information inherent in 3D assets that are fundamentally correlated with shape”³⁴, TRELIS uses a structured, multi-stage generative approach led by O-Voxel (Fig. 7), which operates on a unified representation.

In fact, O-Voxel is an omni-voxel representation that encodes both geometry and appearance, creating a bridge between mesh assets and neural networks. Moreover, in order to handle diverse 3D formats seamlessly, TRELIS avoids directly generating unstructured geometry like raw point clouds (as COLMAP does) or implicit fields.

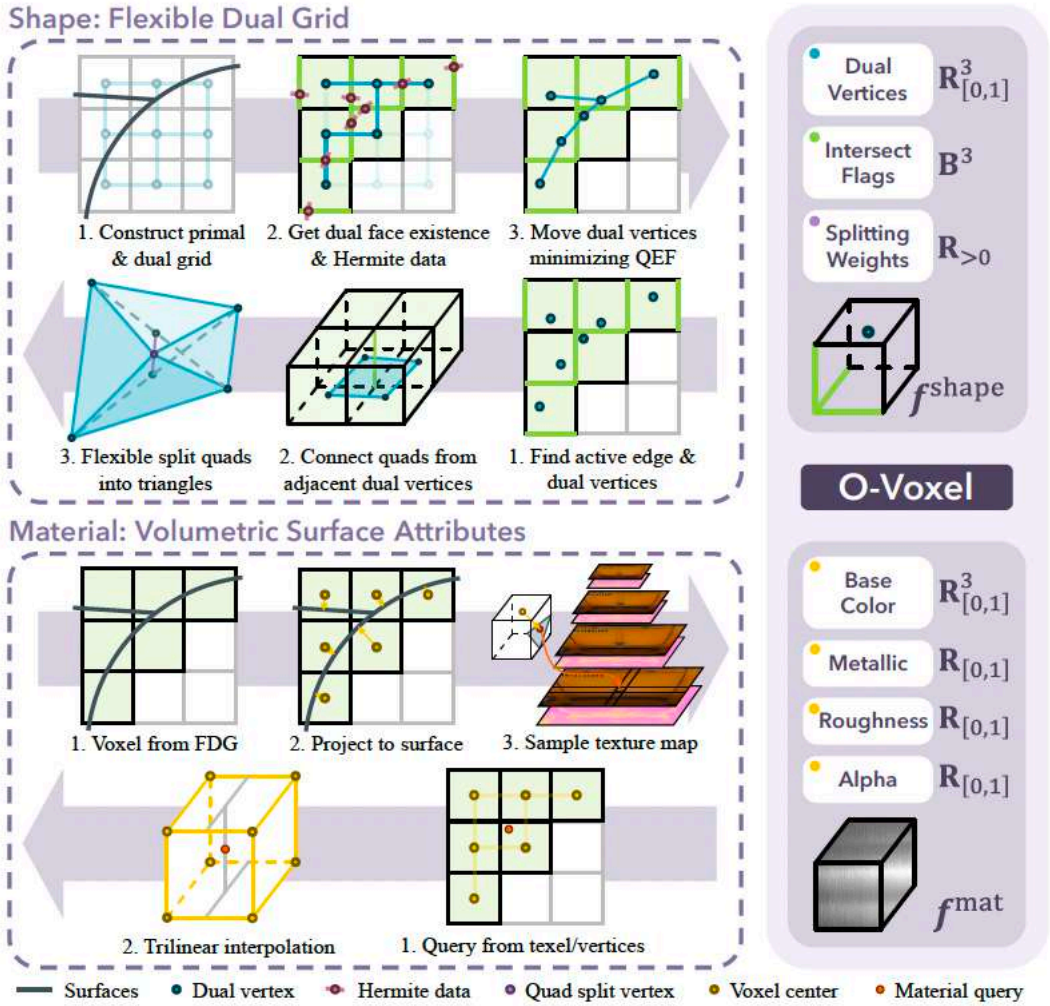


Fig. 7 Illustration of O-Voxel and the bidirectional conversion

³⁴ X. JIAFENG, X. CHEN, S. XU, R. WANG, Z. LU, Y. DENG, H. ZHU et al., “Native and compact structured latents for 3d generation”, in Proceedings in the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2026, p. 2

Instead, it uses a flexible dual grid structure which corresponds to the primal sparse voxel, and “it can accommodate complex topology without lossy data preprocessing and preserve sharp edge features and normal discontinuities for accurate geometry reproduction”³⁵. Inside the dual grid, in fact, one vortex is defined “per primal cell and one quadrilateral face per primal edge, connecting dual vertices in adjacent primal cells”³⁶. By constructing this grid, the algorithm can adjust the position of the dual vertices with a high level of flexibility.

Another interesting feature is that it can capture arbitrary surface attributes beyond just colour: in fact, physically-based rendering (PBR) parameters aligned with surface geometry were implemented in order to enable re-lighting capabilities (particularly useful to handle translucent surfaces). Moreover, O-Voxel features a custom sparse 3D variational autoencoder (Sparse Compression VAE, or SC-VAE - Fig. 8) by employing a

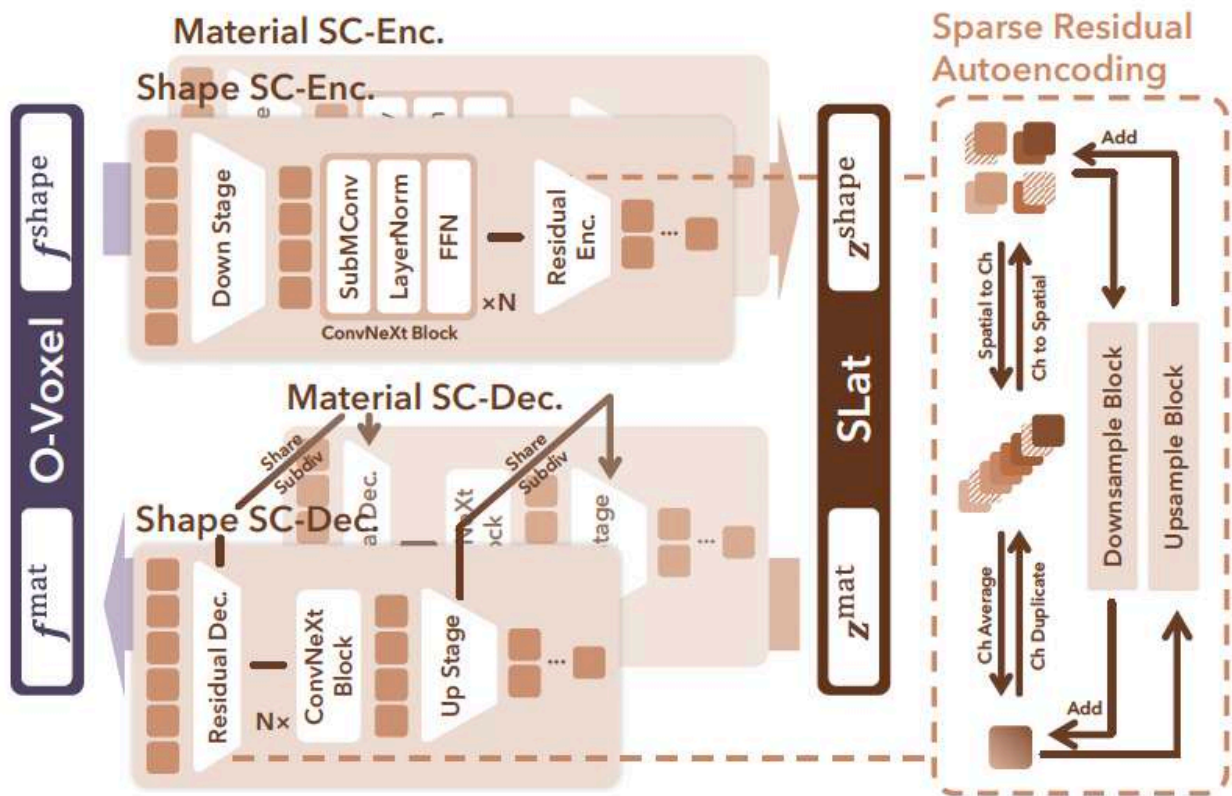


Fig. 8 The network structure of SC-VAE

³⁵ *Ibidem*.

³⁶ *Ibidem*, p. 3

“fully sparse-convolutional network”³⁷ to learn the proper latent space and facilitate efficient, high-resolution 3D generation.

The generation process is based on three different stages characterised by three different models:

1. Sparse structure generation: it predicts the occupancy layout of the sparse voxel grid;
2. Geometry generation: it produces geometry latents within active voxels;
3. Material generation: it synthesises material latents aligned to the geometry structure, and it allows the modelling of PBR materials directly in the native 3D space.

In addition, the third stage in the pipeline can be “independently used as a 3D PBR texture synthesis model given 3D mesh and reference image”³⁸ (Fig. 9). It is widely known that multi-view approaches often suffer from inconsistencies both regarding the shape and the different views, resulting in ghosting or blurred textures. TRELIS represents an advancement in this aspect since the framework “performs appearance

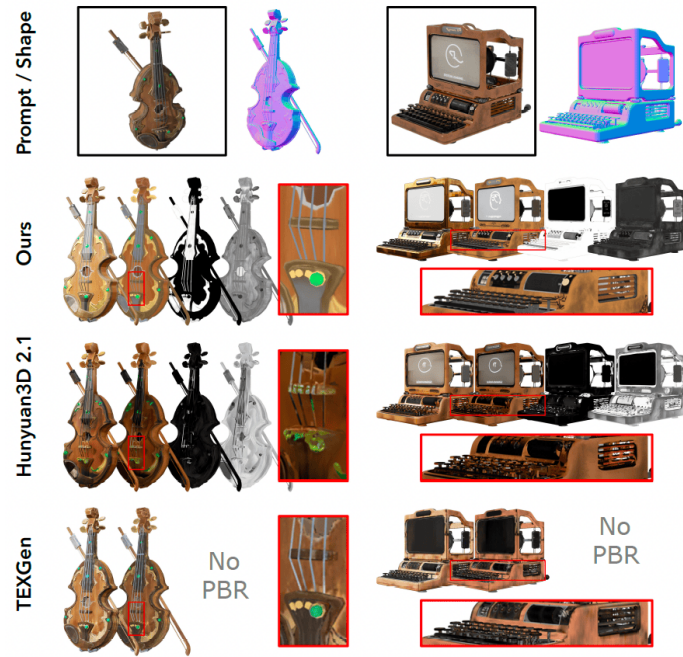


Fig. 9 Comparison of PBR texture generation

³⁷ *Ibidem*, p. 5

³⁸ *Ibidem*, p. 7

reasoning natively in 3D”³⁹, enabling sharper textures, consistent shape-material alignment, and synthesis of texture for internal surfaces.

Comparing the methodologies

Feature / Aspects	COLMAP (Photogrammetry)	TRELLIS (Generative AI)
Input data	Tens to hundreds of overlapping photos from multiple angles	A single 2D image (or text prompt)
Reconstruction Engine	Deterministic mathematics (triangulation, bundle adjustment)	Machine learning inference (Neural networks / Latent transformers)
Handling Unseen Geometry	Fails / Leaves gaps: Requires line-of-sight exposure for every polygon	Predicts / Hallucinates: Fills in occluded geometry using learned dataset priors
Accuracy and Realism	High metric fidelity to the physical object	High visual plausibility; approximate geometric scale
Processing Time	Minutes to hours depending on photo count and resolution	Seconds
Material / Texture Mapping	Projects original camera pixels onto mesh surfaces	Synthesizes PBR materials (Base color, roughness, normals)

Lastly, I based the main comparison analysis of the wells on the models reconstructed with COLMAP and TRELLIS.2 v2 (this latest version of the technique represents a major architectural upgrade from the previous one because it focuses on extreme geometric detail, complex topology handling, and a full PBR material pipeline), but I also weighted the different versions of TRELLIS to highlight how this methodology has evolved over time, as well as its strengths and weaknesses.

In general:

- TRELLIS v1 (standard): it represents the foundation of the framework and provides superior visual presentation without altering the underlying polygon count or geometric topology of v1. The model code remains open-source under the MIT License.

³⁹ *Ibidem*.

- TRELLIS v2: it transitions to a sparse voxel representation (O-Voxel), and it is capable of handling non-manifold geometry, open surfaces, thin structures (like foliage or cloth), and complex cutouts without topological artefacts. Significantly reduces generative "hallucinations" on occluded scenes. This version operates on a freemium model: it provides some free initial trial credits, followed by a paid credit pack or monthly subscription tier.

Comparing the two TRELLIS.2 versions

Feature/Metric	TRELLIS v1	TRELLIS v2
Latent/Voxel Grid	SLAT Sparse Grid	O-Voxel Grid (up to 1536^3)
Texture Map Resolution	512×512 to 1024×1024	Up to 4096×4096 (4K)
Material Map Support	Base Color (Diffuse)	Full PBR (Base Color, Roughness, Metallic, Normal, Opacity)
Topology and Open Surfaces	Limited to closed / manifold meshes	Native support for open surfaces and non-manifold meshes
Generation Speed	10–30 seconds	1-3 minutes
Cost	Free	Freemium and subscription

2.3. Creating the dataset of images

The images comprising the dataset used to create the 3D models were captured using different equipment and camera settings, in order to adapt to the lighting conditions encountered during the shooting days.

Specifically, for the well in Piazza XX Settembre, a Canon R5 camera was used with a Canon RF 100mm Macro f/2.8 lens and a tripod to prevent image blur. The camera was set to evaluative metering, with an aperture of f/11, an ISO value of 200, and a focal length of 100 mm, while manual exposure was used. The images were captured in JPEG format at a resolution of 8192×5464 px. The photographs were taken between late afternoon and evening to ensure that each shot had soft and uniform lighting, avoiding strong direct light that could produce harsh shadows and reflections.

As for the “Diogenes’ Lantern,” a Canon R5 was also used, this time with a Canon RF 24–105mm f/4–7.1 lens and a tripod. The camera was again set to evaluative metering, with an aperture of f/8, an ISO value of 200, and a focal length of 70 mm, while manual exposure was used. In this case, the photographs were taken in the early morning, in JPEG format with a resolution of 8192×5464 px.

The way in which the individual images were captured varied depending on the software used during the 3D processing phase and the size of the well, in fact:

- For COLMAP, the camera was mounted on a tripod at a distance of 2.7 m from the monument. The image acquisition started from the bottom and proceeded upward in vertical overlapping rings, ensuring that each image overlapped with the subsequent one by at least 70–80% (Fig. 10). This resulted in a dataset of approximately 960 photographs: 800 for the well in Piazza XX Settembre and 162 for the “Diogenes’ Lantern” (this difference stems from the diverse sizes between the two wells: while the diameter of the basin is similar, the stepped base platform of the well in Piazza XX Settembre is much wider than “Diogenes’ Lantern”’s, therefore, in order to maintain the correct overlap, the number of images had to increase). Moreover, since “Diogenes’ Lanterns” nears 3 m in height, it was impossible to capture it in its entirety without using a drone, therefore, the decision was made to focus on the central part (the basin with the bas-reliefs) and the base of the columns;

- For TRELLIS, on the other hand, the camera was positioned 5 m away from the target, and a single image of each well was captured.



Fig. 10 Example of overlapping

To ensure that the wells were correctly identified by the TRELLIS algorithm, the images were edited using Photoshop and Camera Raw. The background was selected, then the exposure, highlights, whites, and blacks cursors were set to 0, and the contrast was increased to 100% (Fig. 11).



Fig. 11 Example of editing for TRELLIS.2

2.4. 3D imaging and rendering

I began the 3D modelling process using TRELIS.2, selecting the first version (v1) from the ones available. During the initial processing stage, I left the default settings unchanged (1024 px resolution, a guidance strength of 7.5, and 12 sampling steps), resulting in a model being generated in just 17 seconds (Fig. 12-15).

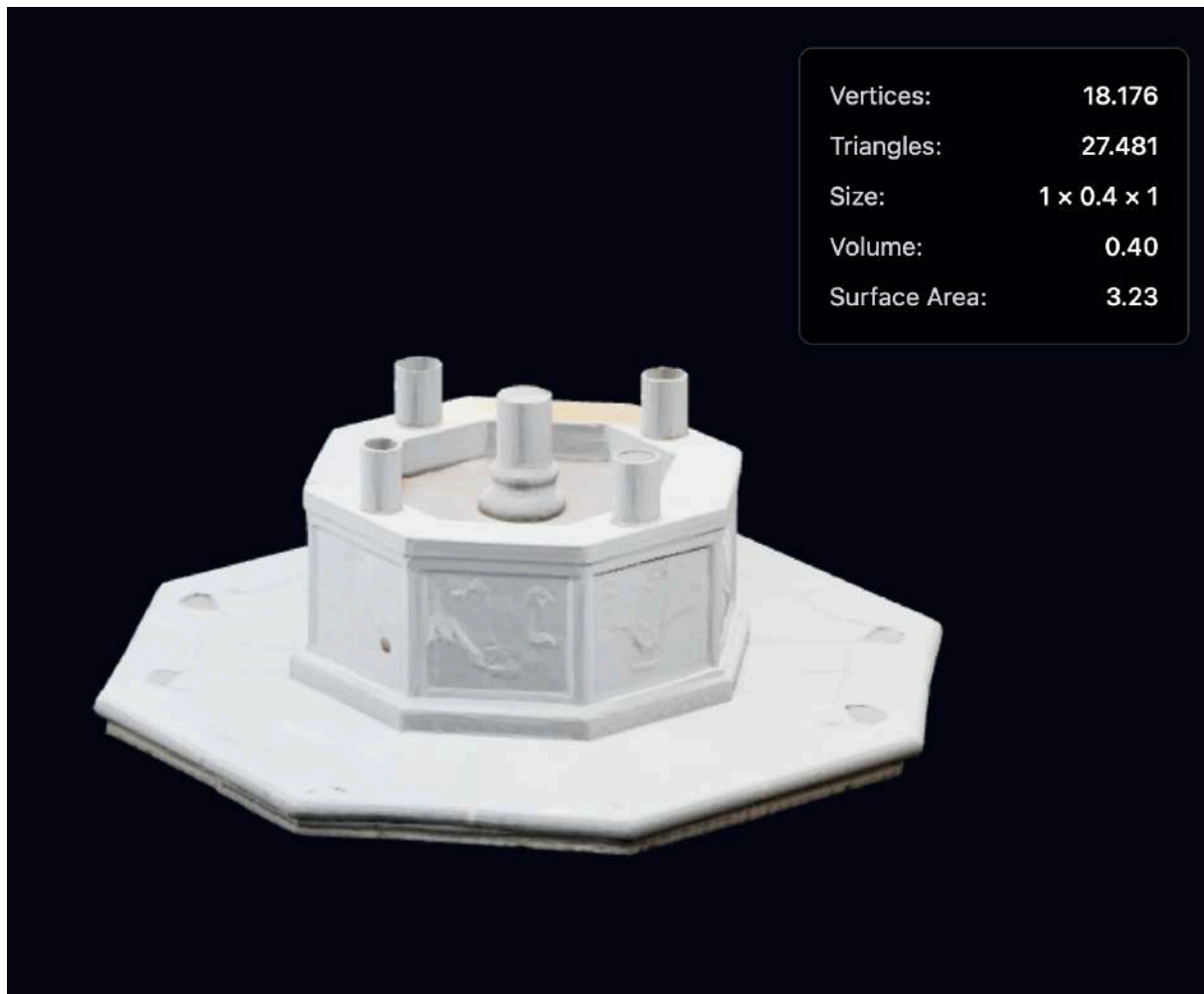


Fig. 12 3D reconstruction of the well in Piazza San Giacomo with a non-existent column at the centre of it (TRELIS.2 v1)



Fig. 13 Detail of the coats of arms (TRELLIS.2 v1)



Fig. 14 3D reconstruction of the well in Piazza XX Settembre (TRELLIS.2 v1)



Fig. 15 Detail of the texture (TRELLIS.2 v1)

I subsequently performed the modelling process again, this time modifying some of the settings: I increased the resolution to 2048 px, the guidance strength to 10, and the sampling steps to 29. In this case as well, the models were generated in approximately 17 seconds (Fig. 16-19).



Fig. 16 Second attempt with increased resolution: the coats of arms are much more visible, but the number of columns increased (TRELLIS.2 v1)



Fig. 17 Detail portraying a hole in the reconstruction (TRELLIS.2 v1)

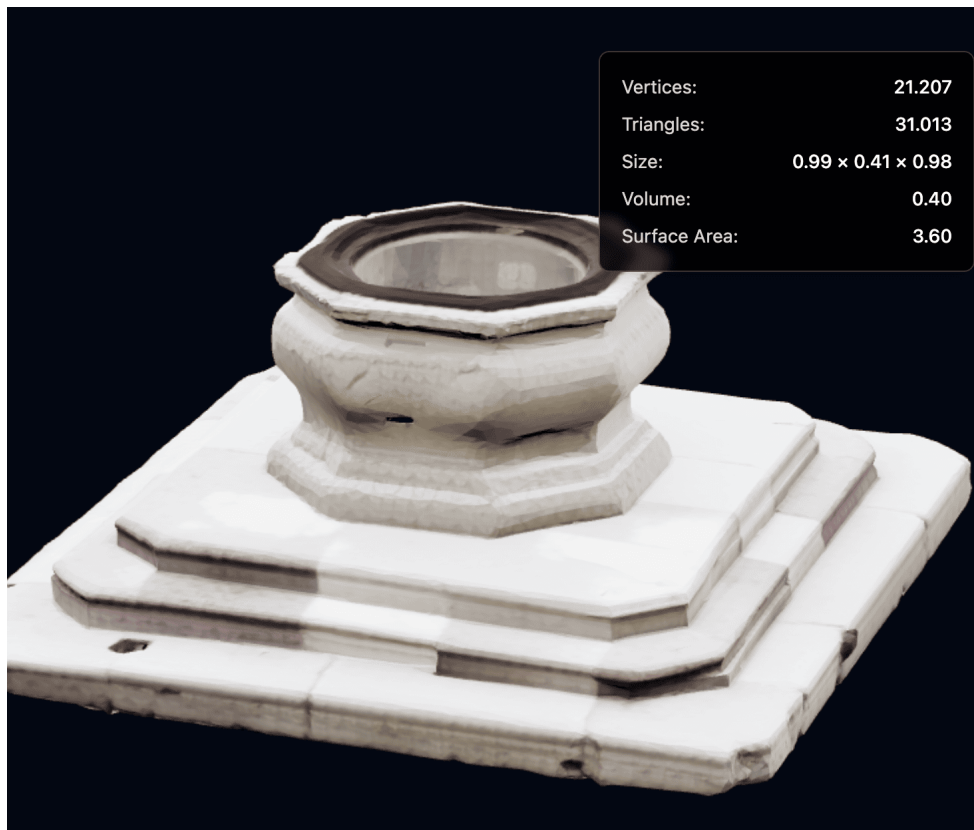


Fig. 18 Second attempt with higher resolution (TRELLIS.2 v1)



Fig. 19 Detail of the improved texture (TRELLIS.2 v1)

Then I switched to v2, which gave me the opportunity to manage not only the resolution of the model (1024px) but also the resolution of the texture (2048px). The guidance strength was set to 7.5 and the sampling steps to 12; in this case, the processing time was longer, taking between 60 and 120 seconds (Fig. 20-23).



Fig. 20 3D model created using v2: the reconstruction improved significantly (TRELLIS.2 v2)



Fig. 21 Detail of the coat of arms (TRELLIS.2 v2)

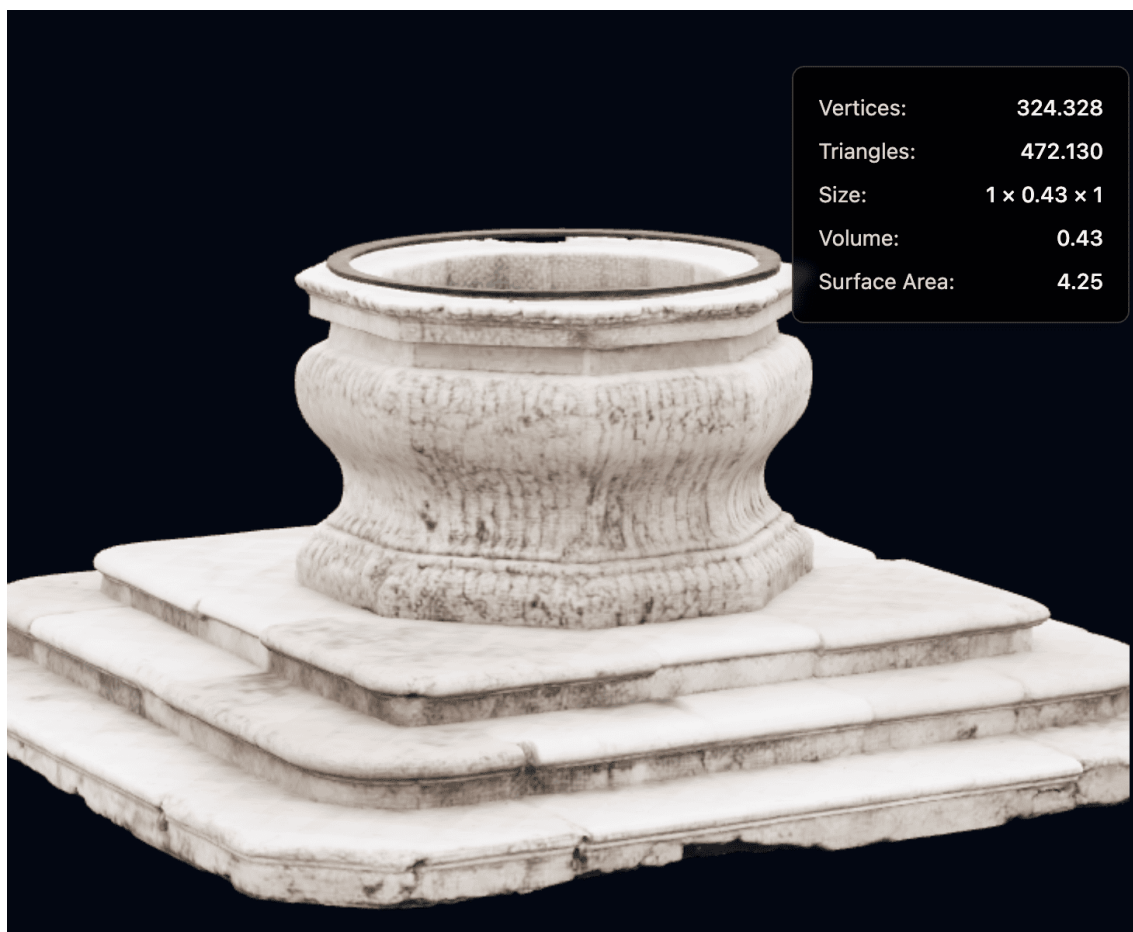


Fig 22. 3D model using v2, displaying a high level of fidelity to the original picture (TRELLIS.2 v2)



Fig. 23 Detail of the texture (TRELLIS.2 v2)

I continued the analysis using COLMAP: after importing the folder containing the 164 photographs of the “Diogenes’ Lantern” into the software, the processing phase began immediately, resulting in the creation of the point cloud. This initial phase took 26 minutes (Fig. 24-25).

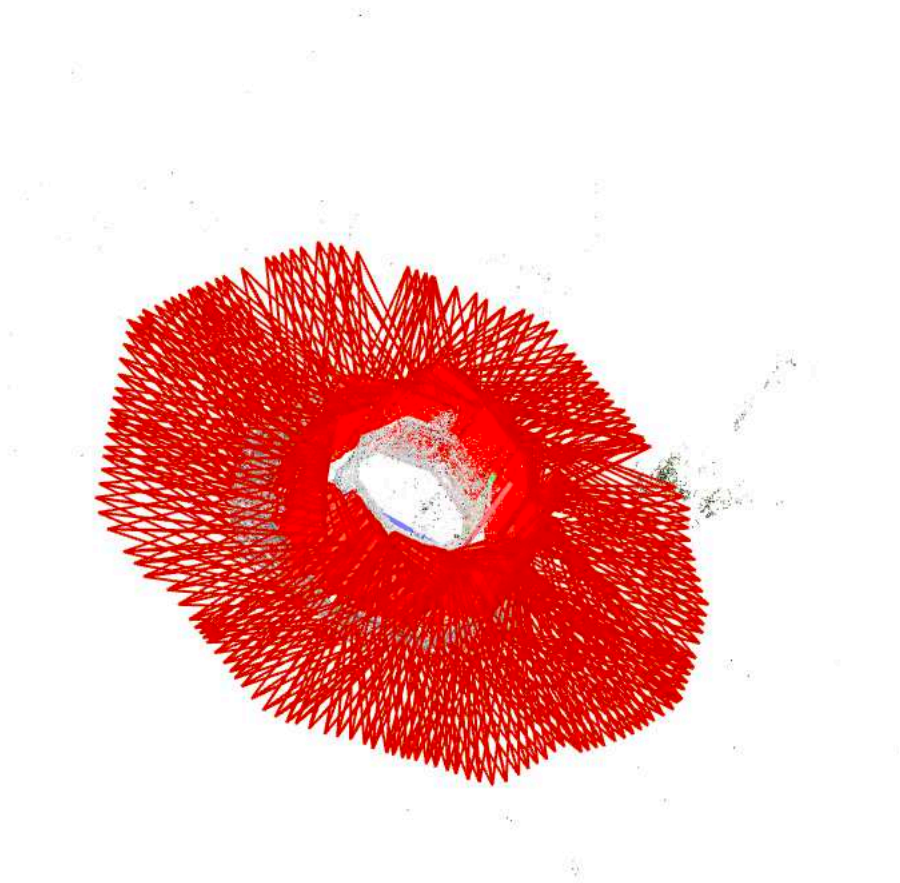


Fig. 24 The result of COLMAP’s image mapping

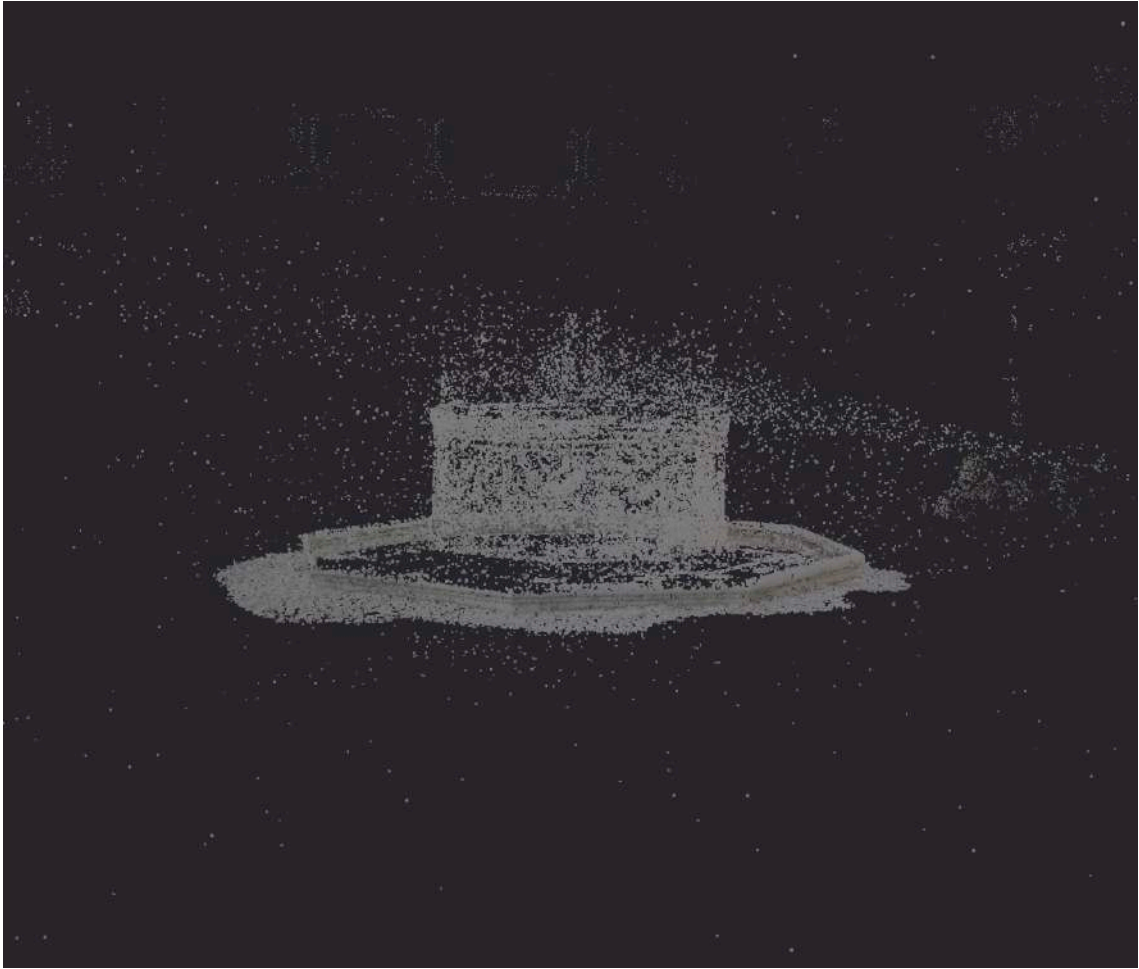


Fig. 25 The cloud of points created by COLMAP

Subsequently, I carried out the same process for the well in Piazza XX Settembre. Given the large number of photographs contained in the folder (approximately 800), the process took 8 hours to complete (Fig. 26-27).

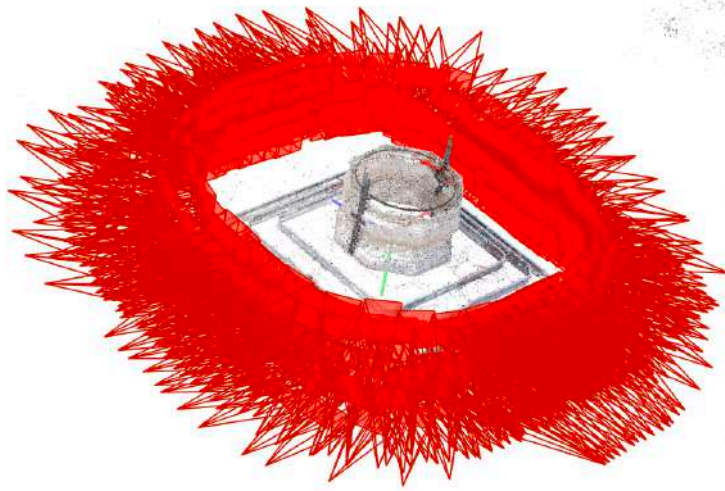


Fig. 26 The result of COLMAP's image mapping

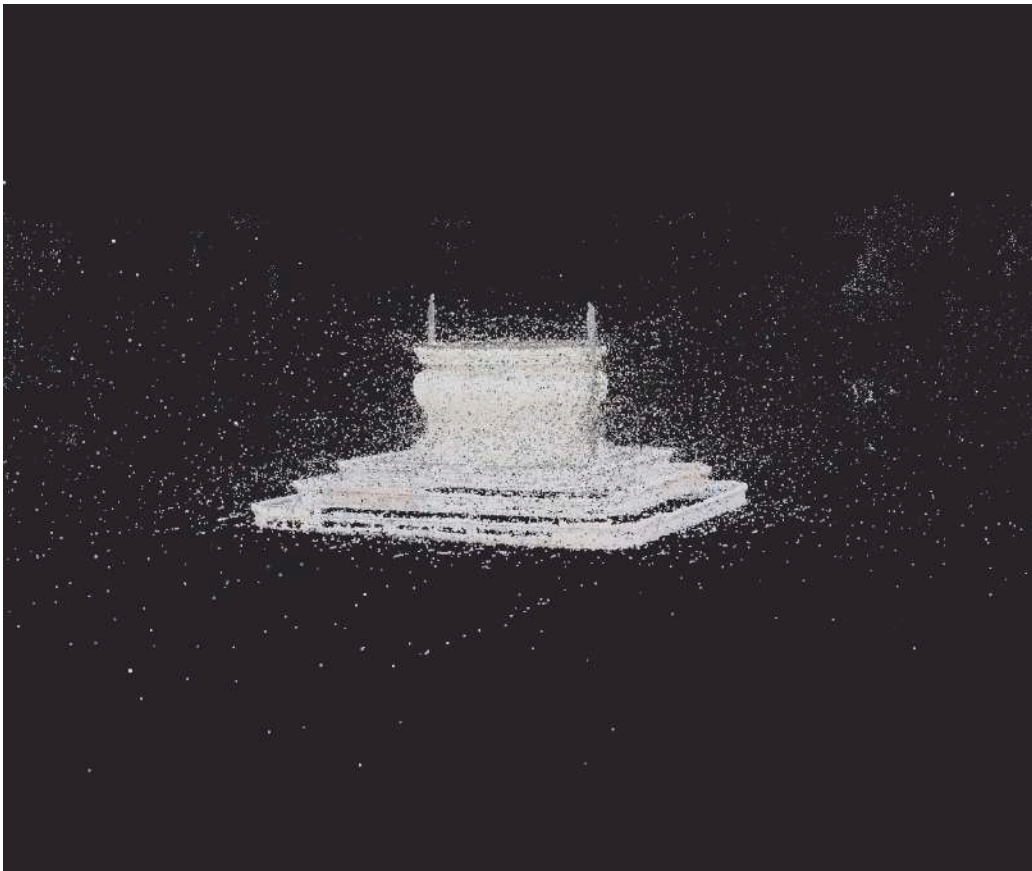


Fig. 27 The cloud of points created by COLMAP

The files containing the point clouds in .ply format were subsequently opened in CloudCompare and finally exported in the .obj format for the final rendering stage (the models created with TRELIS.2 were instead exported in .glb). (Fig. 28-31)



Fig. 28 The 3D model of the well in Piazza San Giacomo (COLMAP)



Fig. 29 Detail of the coat of arms (COLMAP)



Fig. 30 3D model of the well in Piazza XX Settembre (COLMAP)



Fig. 31 Detail of the well's body (COLMAP)

For the rendering of the wells, I used the open-source software Blender, which offers interesting features for both 3D modelling and final export. For each well, I positioned the camera and added a light source: I chose natural sunlight, setting its strength between 2.0 and 3.0 W/m² to enhance the bas-reliefs of the “Diogenes’ Lantern” and the marble details of the well from the ancient Hospital of Udine. Finally, I set the angle to 0.5° to achieve soft lighting without harsh transitions between light and shadow (Fig. 32-33).

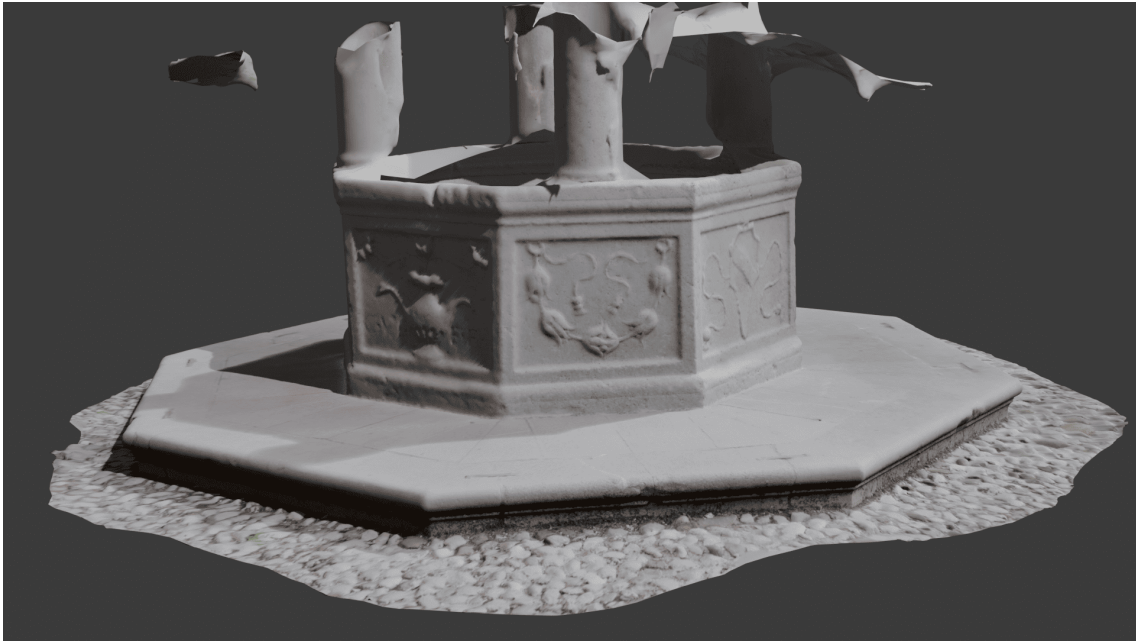
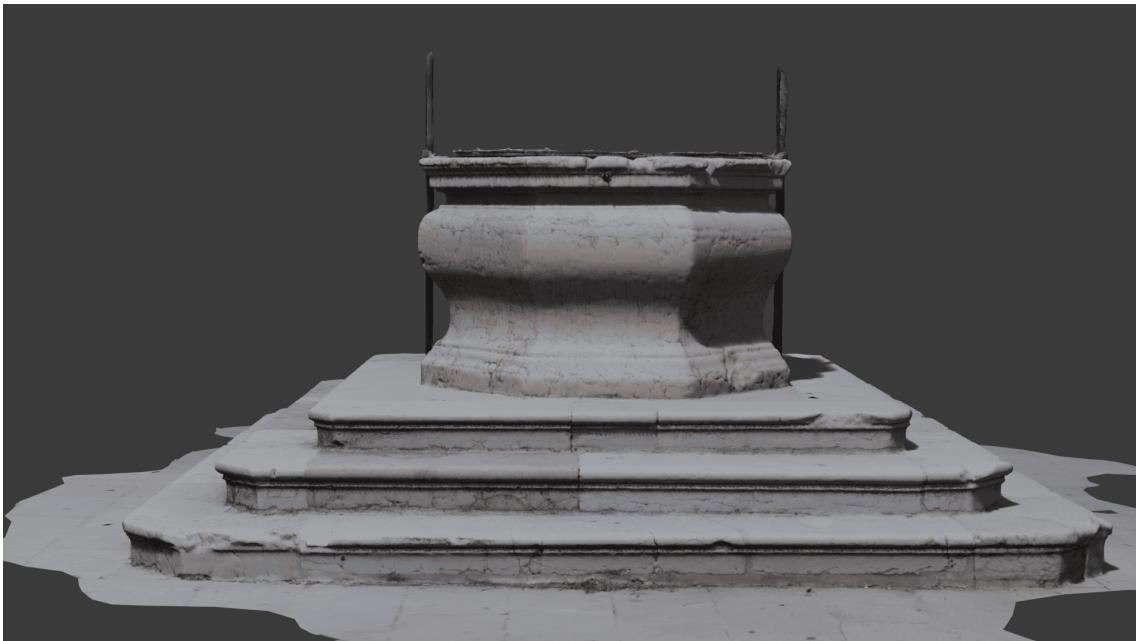


Fig. 32 Rendering of the 3D models created with COLMAP

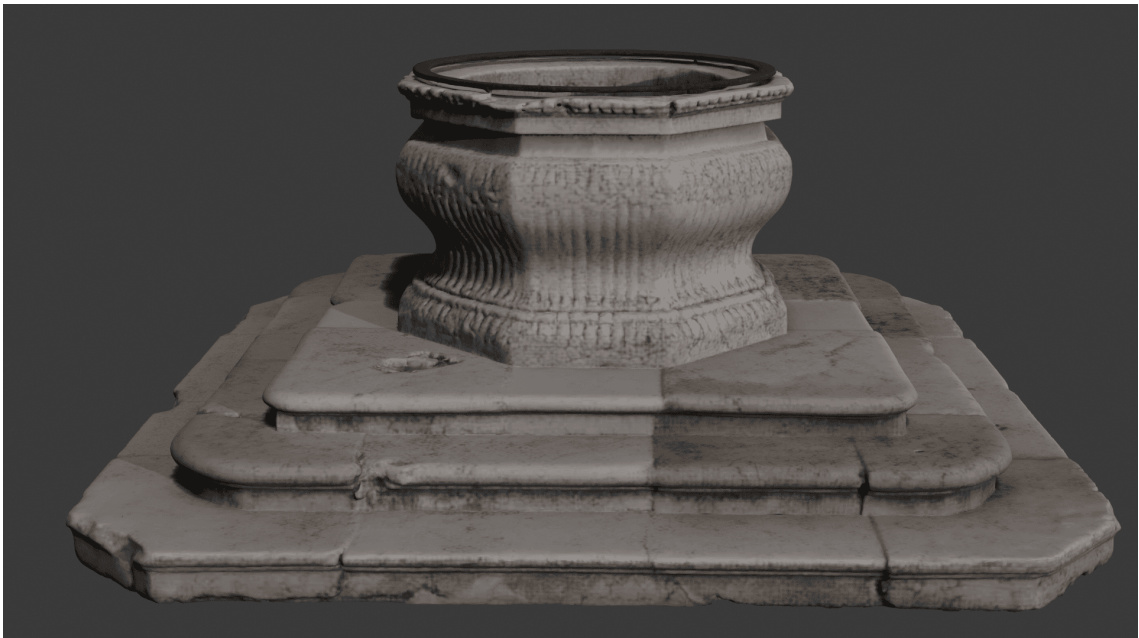


Fig. 33 Rendering of the 3D models created with TRELLIS.2 v2

2.5. Comparing the results

All three 3D models created with TRELLIS.2 (v1, v1 with higher resolution, and v2) share the same geometric structure: every model has successfully identified and reconstructed the general octagonal structure of the “Diogenes’ Lantern”, characterised by a wide stepped octagonal base platform, an elevated octagonal main wall enclosure and vertical cylindrical columns rising from the upper rim. Moreover, all three models display a uniform off-white to light grey colour scheme, designed to imitate light-colored limestone, and each one attempts to interpret the central decorative bas-relief (the coats of arms) as a raised feature on the front of the octagonal pane.

The primary differences, instead, among the TRELLIS 3D models (Fig. 34) lie in mesh resolution, hallucination of structural elements, and surface texture quality: in fact, v1 variants suffer from severe spatial hallucinations, generating nonexistent columns in the upper ring in addition to the central one inside the basin. We can further notice that the high-resolution v1 model further damages the surface quality by introducing holes and deep indentations on the front relief panel. V2, instead, eliminates these incorrect elements, introducing correctly scaled, thick cylindrical columns. Furthermore, v2 reproduces the central bas-relief vase-like design with sharp volumetric depth and smooth shadowing, whereas the v1 models present flattened details.

Another aspect of discordance stems from the target image displaying fine details (like stone porosity, surface weathering, subtle cracks and carved inscriptions) that are reduced in the v1 model to plastic-like surfaces. Even the highly detailed v2 produces a synthetic, modern marble finish that lacks the authentic texture, subtle ageing marks, and natural stone grain.

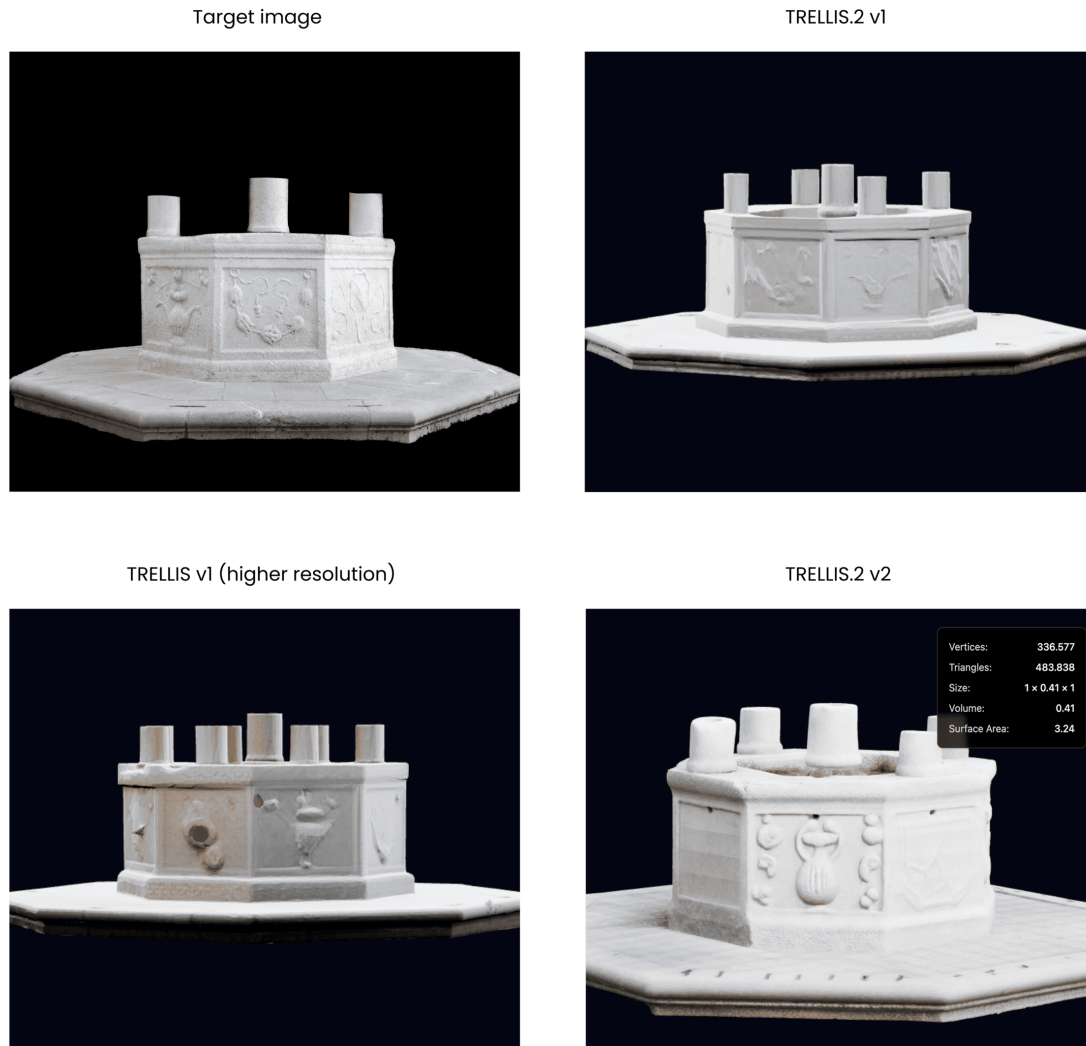


Fig. 34 Comparison between target image, v1, v1 with higher resolution and v2 (TRELLIS.2)

For what pertains to the well in Piazza XX Settembre, we can notice that the models again share a consistent architectural schema (Fig. 35): each model generates a stepped base platform, an elevated central well body, and an inner dark rim placed in the opening. All three digital assets portray the material colouring of the real-world structure, mapping distinct pinkish/reddish stone blocks onto the front edges of the base, while preserving the off-white for the basin and upper steps.

The core differences among the three 3D outputs, instead, manifest in the surface texture, resolution, and geometric fidelity: both v1s render the basin as a faceted shape with blurry texture. Instead, v2 was able to achieve tactile realism by reconstructing

vertical lines, grooves, stone weathering, and elements of micro-erosion. In addition, the high-resolution v1 model adds dark indentations and glitchy holes near the lower rim of the well's body, whereas v2 created a clean, high-fidelity mesh with realistic marble veining along the stepped base.

Another big difference concerns the two prominent vertical iron bars that are completely absent in the reconstructed assets.



Fig. 35 Comparison between target image, v1, v1 with higher resolution and v2 (TRELLIS.2)

Then, if we compare the v2 reconstruction of the “Diogenes’ Lantern” made with TRELIS.2 with the one produced with COLMAP, we’re able to take into consideration several characteristics that they have in common (Fig. 36). At the geometric level, again both digital assets successfully identify and reproduce the primary architectural aspects of the well: a wide octagonal base supporting an elevated, multi-sided body that culminates in an upper ring. Moreover, both models attempt to capture the columns situated at the top edge of the well. Regarding the details of the surface, instead, both reconstructions preserve the placement and spatial distribution of the relief carvings on the front octagonal panels, and portray an off-white/cream colored stone, capturing the general characteristic of the original stone.

In contrast, while TRELIS v2 generates a solid mesh with smooth surfaces, clean edges, and fully enclosed geometry (the columns), the COLMAP model exhibits geometric tearing, especially in the upper structure, due to insufficient feature matching. COLMAP, in fact, attempted to capture the full height of the original pillars, but that unfortunately resulted in hollowed-out, vertical structures. Despite COLMAP showing clear limitations when dealing with partially visible elements, the robustness of the tool is undeniable: the use of a large number of images allows it to achieve a level of accuracy that TRELIS can’t match. This accuracy could have also been further improved by capturing the well in its entirety and from different heights.

Another interesting point of comparison regards the texture of the surface: COLMAP, relying on photographic projection, was able to preserve the authentic light exposure from the source photos, whereas TRELIS instead synthesises an artificially shaded texture map.

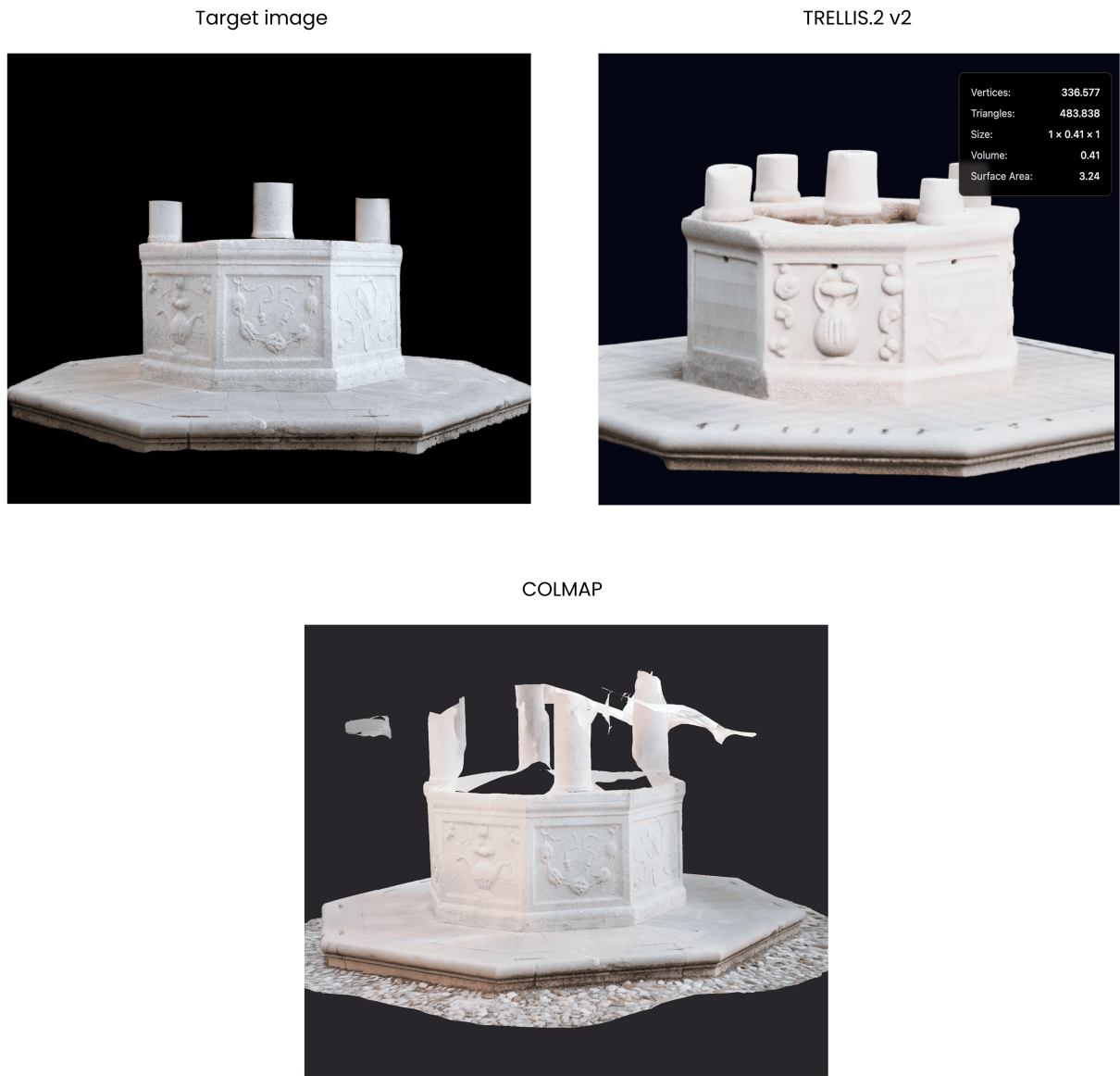


Fig. 36 Comparison between target image, TRELLIS.2 v2, and COLMAP

Lastly, for the well in Piazza XX Settembre (Fig. 37), the two models find themselves again in architectural alignment: both successfully reconstruct the primary structure consisting of a three-stepped base, a centrally elevated stone basin, and a circular top opening fitted with an inner metal ring. As for the decorative elements, both models accurately map the distinct dual-tone stonework, the outward convex curvature, and the vertical fluting along the main body of the well. At first glance, we can also notice that the most striking difference lies in the upper metallic elements: COLMAP successfully

reconstructs the two vertical black iron poles rising from the upper rim; in contrast, TRELLIS completely omits them.

Target image



TRELLIS.2 v2



COLMAP



Fig. 37 Comparison between target image, TRELLIS.2 v2 and COLMAP

To conclude, the whole process of obtaining the 3D reconstruction first with TRELLIS and then with COLMAP went entirely differently: while the acquisition of the models produced by TRELLIS was almost instantaneous, COLMAP required from half an hour to almost a whole day to obtain the point clouds. Moreover, although TRELLIS.2 v2 is available only by purchasing a monthly subscription (COLMAP is instead completely open-source), its use is extremely intuitive, with only one target image required to produce the 3d model and a settings panel really easy to understand. On the other hand, it clearly lacks the geometric and textural precision that characterises COLMAP and its Structure from Motion algorithm.

3. Online visualizer for the 3D models

The tool is a single-page 3D viewer, built with the focus of letting the user manipulate and switch between the four reconstructed 3D models of the two wells analysed, each reconstructed with two different photogrammetry pipelines (COLMAP and TRELLIS) (Fig. 38).

The SPA has five conceptually distinct layers that interact through an explicit data flow:

- A data layer that describes what models exist and where their assets live;
- A state layer that tracks which model is currently being viewed;
- A 3D rendering layer that is responsible for tuning a glb URI into pixels on screen;
- A UI layer that lets the user change which model is active;
- And a loading layer that gives the user feedback while a model asset is being fetched and parsed.

Ultimately, the libraries chosen for this tool are mainly three:

- React, which allowed the application's central interaction (swapping the displayed model while keeping the rest of the UI intact) to be expressed as a function of state: in fact, the “activateId” state variable represents the entire routing mechanism of the app;
- Three.js, a tool well known for being able to manage scene graphs, cameras, lights, loaders, and materials with ease;
- Lastly, React Three Fiber, the reconciler that lets Three.js objects be described as JSX and keeps them synchronised with React's state and props system.

Udine ancient wells: a comparison between COLMAP and TRELLIS

Diogenes' Lantern - COLMAP



Diogenes' Lantern - COLMAP



Well of Piazza XX Settembre - COLMAP



Diogenes' Lantern - TRELLIS



Well of Piazza XX Settembre - TRELLIS

Open-source code created by [Matilde Moretti](#) with React

Fig. 38 The online visualizer coded with React, Three.js, and React Three Fiber

References

1. A. FRANGIPANE, “La costruzione di Udine: tracce di lettura lungo le acque, le mura, le case”, in *Cultura in Friuli III*, 2017, Società Filologica Friulana, pp. 689-698
2. F. BRAIDOTTI, “I primi pozzi di Udine”, in *Pagine Friulane* n.2, 1891, Patria del Friuli, pp.17-18
3. F. BRAIDOTTI, “L'acqua potabile a Udine : lettura”, in *Atti della Accademia di Udine*, serie 2, v. 9 (1890/1893), pp. 3-86
4. M.M. HASSAN, S. KARUNGARU, R. BASHAR, “A Short Review on Computer Vision: Visualizing the World Through Machine”, 2025, 10.20944/preprints202506.0630.v1, pp. 1-21
5. V. BARRILE, G. BILOTTA, D. LAMARI, G. MEDURI, U. TRUNGADI, A. RICCIARDI, “Computer vision/structure from motion per la diffusione dei beni culturali”, in *Atti della XIX Conferenza Nazionale ASITA*, 2015, pp. 51-60
6. S. GRANSHAW, “Structure from motion: origins and originality”, in *The Photogrammetric Record* 33, 2018, pp. 1-5
7. A. BOSCO, E. MINUCCI, “Rendering rti ed editing d'immagine per elaborazioni SfM: confronto tra tecniche e strumenti di visualizzazione per la documentazione di graffiti in contesto archeologico” in *Newsletter di Archeologia CISA* 11, 2020, Centro Interdipartimentale di Servizi di Archeologia, pp. 43-66
8. S. MIJAKOVSKA, “Overview of Structure from Motion” in *Advanced Computing: An International Journal* 15, 2024, Academy and Industry Research Collaboration Center (AIRCC), pp. 25-32
9. L. MARIONI BROS, D. NOBILE, “La chiesa di San Giacomo in Mercatonuovo”, *Deputazione di storia patria per il Friuli*, 2008, pp. 1-61
10. X. LI, S. YAN, Z. ZHU, C. ZHANG, Y. WU, F. MENG, L. YUN, "Relative Agreement and Feasibility of Low-Cost Passive Stereophotogrammetry for Nasal Three-Dimensional Surface Imaging: A Comparison of COLMAP and Agisoft Metashape with a Handheld Structured-Light Scanner" *Bioengineering*, 2026, 13, no. 8: 899, pp. 1-12

11. Team, TELKOMNIKA. "A Practical Approach to Candi Siwa 3D Reconstruction with COLMAP and Nerfstudio." TELKOMNIKA Telecommunication Computing Electronics and Control, 2026, pp. 249-257. doi:10.12928/TELKOMNIKA.V24I1.27212
12. C. HERVÉ, P. GARNIER, J. VIQUERAT, E. HACHEM, "TRELLIS -enhanced surface features for comprehensive intracranial aneurysm analysis", in Neuroscience Informatics, 2026, vol.6 issue 1, pp. 1-14
13. X. JIAFENG, X. CHEN, S. XU, R. WANG, Z. LU, Y. DENG, H. ZHU et al., "Native and compact structured latents for 3d generation", in Proceedings in the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2026, pp. 1-24