

Automatizing 3D Reconstruction Pipelines for Speeding-Up Cultural Heritage Digitization

Gianluca Bison^{1†}, Luca Palmieri^{1†}, Sinem Aslan^{1,3},
Sebastiano Vascon^{1,2*}, Marcello Pelillo^{1,2}

¹*Department of Environmental Science, Informatics and Statistics, Ca' Foscari University of Venice, Via Torino 155, Mestre - Venice, 30172, VE, Italy.

²European Center for Living Technology, Ca' Foscari University of Venice, Dorsoduro 3911, Venice, 30123, VE, Italy.

³Department of Historical Studies, University of Milan, Via Festa del Perdono, 7, Milan, 20122, MI, Italy.

*Corresponding author(s). E-mail(s): sebastiano.vascon@unive.it;

Contributing authors: 870028@stud.unive.it; luca.palmieri@unive.it;
sinem.aslan@unimi.it; marcello.pelillo@unive.it;

[†]These authors contributed equally to this work.

Abstract

Reconstructing accurate 3D models of archaeological artifacts, particularly painted ones like frescoes, presents significant challenges due to their complex geometries and the need to capture detailed pictorial information. Traditional 3D scanning technologies excel in geometry but fall short in capturing texture quality, making RGB cameras a better choice for painted artifacts. This paper proposes a comprehensive pipeline that integrates automated segmentation with photogrammetric tools to address these issues. The semi-automatic segmentation phase minimizes manual effort while ensuring accurate mask generation for intricate artifacts, and the reconstruction process efficiently produces dense 3D models with high geometric and textural fidelity. The final registration of models ensures precise alignment, resulting in an accurate and detailed reconstruction. Tested on fragmented frescoes as part of the EU Horizon 2020 RePAIR project, our pipeline demonstrates a scalable and computationally efficient solution for large-scale CH digitization and restoration, aiding the reassembly of fragmented artifacts with minimal manual input.

Keywords: 3D Reconstruction, cultural heritage digitization, automated generation of segmentation masks

1 Introduction

Digitizing cultural heritage (CH) artifacts into *complete* 3D models is crucial for preserving, studying, and making these objects accessible to a wider audience (Quattrini et al (2020); Di Angelo et al (2022); Pollefeys et al (2003)). Beyond academia, 3D digitization plays a significant role in various industrial sectors, supporting applications like rapid prototyping, quality control, and digital inspection (Javaid et al (2021); Raytrix (2008); Rodríguez-González et al (2017)). The high demand for accurate and efficient 3D modeling underscores the importance of robust digitization techniques across both domains.

In CH, the quality requirements of 3D reconstructions differ based on their intended use. *Online rendering* for interactive VR/AR experiences often prioritizes real-time performance, necessitating compromises in geometric detail and texture resolution. In contrast, *offline rendering* for archival and research purposes requires high-resolution textures and fine geometric details to maintain historical authenticity (Rodríguez-García et al (2024)). This dichotomy emphasizes the need for adaptable 3D reconstruction processes that can cater to diverse application requirements.

Traditional 3D scanning methods, such as structured light scanners, are known for their geometric precision (Brown et al (2012); Theoharis et al (2013)). However, they often fail to capture the fine texture details necessary for preserving the visual and historical authenticity of CH artifacts. To overcome these limitations, RGB image-based methods (e.g., Javaid et al (2021); Rodríguez-González et al (2017); Kong and Hucks (2023); Schonberger and Frahm (2016); Schönberger et al (2016); Griwodz et al (2021); Carsten et al (2021)) have gained traction for their superior textural mapping capabilities. Despite their strengths, these methods face challenges in background noise elimination, integrating images from various angles, handling textureless areas, and managing complex object geometries. Additionally, preserving high-resolution textures is essential to maintaining an artifact’s historical significance, necessitating methods that balance geometric accuracy with visual fidelity.

To address these challenges, we propose a comprehensive pipeline¹ that integrates automated segmentation and photogrammetric tools, particularly Metashape (Agisoft (2006)). This pipeline covers the entire process from image acquisition to 3D model generation (Figure 1). Our segmentation phase leverages RITM (Reviving Iterative Training with Mask Guidance) (Sofiuk et al (2021)) and STCN (Space-Time Correspondence Network) (Cheng et al (2021b)), reducing manual input while ensuring precise mask generation, critical for artifacts with complex geometries and textures. Compared to COLMAP and Meshroom, Metashape was selected for its efficiency in generating dense reconstructions using fewer images, enhancing the computational feasibility of large-scale CH digitization (DJurić et al (2021)). The final stage employs the TEASER algorithm (Yang et al (2020)) for precise 3D model registration, using histogram-based feature descriptors to ensure robust alignment.

While recent learning-based 3D reconstruction methods have demonstrated success, particularly in full-scene reconstructions (Murez et al (2020)), they often rely on synthetic datasets or predefined shapes, limiting their adaptability to irregular, texture-rich CH artifacts (Chibane et al (2020); Genova et al (2020); Yavartanoo et al (2021)). Conversely, traditional geometry-based methods like Structure from Motion (SfM) (Schonberger and Frahm (2016))

¹The code related to the paper will be released upon publication.

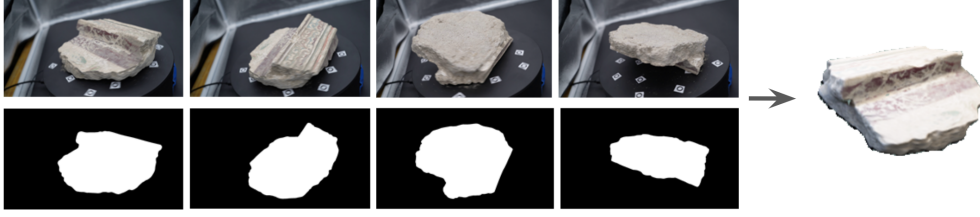


Fig. 1: A schematic representation of a result of the proposed approach. Starting from a sequence of images, the artifact is isolated from the background and reconstructed accurately both in terms of geometry and color information.

and Multi-View Stereo (MVS) (Luo et al (2024)) offer greater flexibility but are computationally intensive and prone to errors in low-texture regions. Our integrated approach seeks to balance the strengths of both learning-based and geometric methods, providing a scalable, computationally efficient solution adaptable to various CH artifacts.

This research is part of the EU Horizon 2020 project, RePAIR² (Tsesmelis et al (2024)), aimed at reassembling fragmented frescoes, with approximately *16,000 fragments* targeted for digitization. Our pipeline has proven exceptionally useful for efficiently scanning a large set of fragments within a reasonable timeframe while producing high-quality 3D models with precise shape details and rich texture. These models play a crucial role in subsequent steps of the automated fresco reassembly process, ensuring both accuracy and efficiency in reconstructing the original fresco. By integrating high-resolution digitization with automated segmentation and photogrammetric tools, our approach enhances the overall workflow, making it more scalable and effective for large-scale cultural heritage restoration projects.

To the best of our knowledge, our work is the first to comprehensively address the challenges of fragment digitization in CH by combining automated segmentation, precise 3D model registration, and efficient photogrammetry in a unified pipeline. This integrated solution specifically meets the unique requirements of CH applications in terms of scalability, adaptability, and precision for irregular and texture-rich artifacts.

2 Related Work

The process of 3D digitization for cultural heritage artifacts involves multiple stages, each contributing to the creation of a high-fidelity 3D model. Our work integrates state-of-the-art methods across these stages, which include 2D segmentation, 3D reconstruction, and 3D registration. In this section, we discuss the relevant research in these areas, positioning our approach within the broader context of 3D digitization for cultural heritage.

2.1 2D Segmentation for Object Isolation

Image segmentation is a foundational task in computer vision, with methods ranging from simple threshold-based approaches (Otsu (1979)) to more advanced techniques such as region-growing and graph-cut methods (Boykov Y. (2006)). The advent of deep learning has revolutionized segmentation tasks, with models like U-Net (Ronneberger et al (2015)), and

²<https://www.repairproject.eu/>

Mask R-CNN and its variations, i.e., fast R-CNN and faster R-CNN (He et al (2017); Ren et al (2015); Girshick et al (2014)) achieving state-of-the-art results. These deep learning-based approaches have proven to be highly effective for detecting and segmenting objects in complex images, including cultural heritage artifacts (Hou et al (2024); Hatir et al (2021)).

To address the challenges of labor-intensive full-mask annotations, recent advancements have shifted towards weakly-supervised techniques that reduce manual effort while maintaining high accuracy. These methods leverage simpler forms of annotation, such as bounding boxes (Cheng et al (2015); Zemene and Pelillo (2016)), scribbles (Aslan and Pelillo (2019); Bai and Wu (2014)), or clicks (Sofiuk et al (2020a, 2022)). Approaches like RITM (Reviving Iterative Training with Mask Guidance) (Sofiuk et al (2021, 2020b)) and STCN (Space-Time Correspondence Network) (Cheng et al (2021b,a)) have successfully utilized weak annotations and leverage temporal consistency, respectively, to enhance segmentation accuracy with minimal human input. In our pipeline, we explored these two approaches to automate segmentation by significantly reducing the labor-intensive task of manual annotation while fine-tuning a Mask R-CNN model for improved accuracy.

2.2 3D Reconstruction from Image sequences

3D reconstruction from image sequences has been widely studied, particularly in the fields of photogrammetry and structure-from-motion (SfM), with applications spanning both industrial contexts (Javaid et al (2021); Rodríguez-González et al (2017)) and cultural heritage studies (Kong and Hucks (2023)). While active scanning technologies, such as structured light and laser scanners (Brown et al (2012); Theoharis et al (2013)) offer high accuracy in geometry, they are less effective in capturing the fine textures needed for painted artifacts like frescoes. To overcome this limitation image-based methods have been commonly employed for capturing detailed surface textures (Javaid et al (2021); Rodríguez-González et al (2017)).

Tools such as Metashape (Agisoft (2006)), COLMAP (Schonberger and Frahm (2016); Schönberger et al (2016)) and Meshroom (Griwodz et al (2021); Carsten et al (2021)) are widely used in SfM and Multi-View Stereo (MVS)-based 3D reconstruction. These methods extract features from multiple images to estimate camera poses and generate sparse 3D models, which are subsequently refined through dense reconstruction techniques.

While these approaches are effective, they can face challenges when applied to noisy or incomplete data, such as irregular object geometries often encountered in cultural heritage artifacts. They are designed for full-scene reconstructions, which may not be suitable for fragmented artifacts like frescoes.

Moreover, recent research in 3D representation has introduced techniques that aim to facilitate efficient, detailed reconstructions. These include voxel-grid representations for regular volumetric structures and Signed Distance Functions (SDFs) (Park et al (2019)) to retain fine geometrical details. Additionally, implicit function-based methods (Chibane et al (2020); Genova et al (2020); Yavartanoo et al (2021)) and neural radiance fields (Mildenhall et al (2021)) allow for the representation of complex objects and high-quality rendering, with particular suitability for neural network integration. While these advanced techniques yield impressive results with synthetic data, they are often computationally demanding and rely on idealized input conditions, such as known camera poses and perfect segmentation masks. These requirements limit their applicability to real-world noisy data, where cumulative estimation errors can compromise accuracy.

In the field of 3D reconstruction for *cultural heritage*, many *photogrammetry-based studies* have made valuable contributions but often focus on specific stages of the reconstruction pipeline, lacking an integrated approach. For example, (Llull et al (2023)) excel in evaluating reconstruction methods but rely heavily on manual processes, which reduces adaptability to varying artifact conditions. Similarly, (Wang et al (2023)) prioritize registration but fail to robustly address segmentation or the integration of segmentation, reconstruction, and registration phases. In contrast, our work provides a unified solution, bridging these gaps by incorporating fine-tuned segmentation with minimal manual annotation, automated background removal, and automatic registration into a modular pipeline. Unlike prior works that often bypass detailed segmentation or rely on manual interventions, our pipeline ensures efficiency and precision tailored to the complex demands of fragmented fresco reconstruction, addressing both partial and complete 3D artifact assembly seamlessly.

2.3 3D Registration for Model Alignment

3D registration is the process of aligning multiple partial 3D models to create a complete representation of an object. Techniques like the iterative closest point (ICP) algorithm (Besl and McKay (1992)) have long been used for aligning 3D point clouds. However, ICP is highly sensitive to noise and requires precise initial alignment, making it less suitable for real-world scenarios involving fragmented cultural heritage artifacts with irregular shapes and limited surface overlap.

To address these limitations, various improvements have been introduced, including refined ICP variants (Rusinkiewicz and Levoy (2001)) and learning-based approaches (Wang and Solomon (2019)) that enhance robustness and accuracy. Additionally, advanced feature extraction methods (Choy et al (2019)) and robust global optimization frameworks (Huang et al (2021)) have been developed to improve performance in noisy or low-overlap conditions.

More recently, TEASER (Yang et al (2020)) has gained significant attention for its ability to handle outlier-contaminated data while guaranteeing globally optimal solutions. This robustness makes TEASER particularly effective for complex, fragmented datasets like cultural heritage artifacts with inconsistent surfaces and minimal overlap. According to Osipov et al (2023), feature correspondence-based methods, including TEASER, outperform ICP-based methods in challenging scenarios. The study also highlights that ICP can still be beneficial in refining feature-based methods, a role TEASER leverages by integrating feature-based alignment with ICP refinement, ensuring both robustness and precision in the final registration.

3 Proposed Approach

To reconstruct high-quality 3D models of cultural heritage artifacts from sequential RGB images, we propose a pipeline designed to enable precise artifact-specific segmentation and 3D reconstruction with *minimal manual intervention*. The pipeline begins by capturing a series of images, which can either be ordered or unordered, depicting a fresco fragment or other artifacts from multiple perspectives. Following this, the process includes 2D image segmentation, refinement through morphological erosion, camera pose estimation, 3D reconstruction, and final model registration, ensuring the extraction of detailed geometric and

textural information from standard RGB images. The overview of the pipeline is presented in Figure 2.

This approach is particularly effective for artifacts with complex geometries and textures. By integrating 2D segmentation, 3D reconstruction, and a registration phase to align different views, the pipeline ensures both geometric and textural accuracy. Importantly, the integration of automated segmentation reduces the need for manual annotation, significantly improving the runtime of the overall process. In the field of cultural heritage, time efficiency is critical, as the digitization of large volumes of artifacts or fragments within a limited timeframe is often necessary. Additionally, the reduced reliance on manual input allows the workflow to be scalable, enabling the processing of large datasets with consistency and minimal labor. This combination of high accuracy and fast runtime makes our pipeline particularly valuable for cultural heritage digitization and other fields requiring high-precision 3D modeling. The following subsections describe each stage of the pipeline in detail.

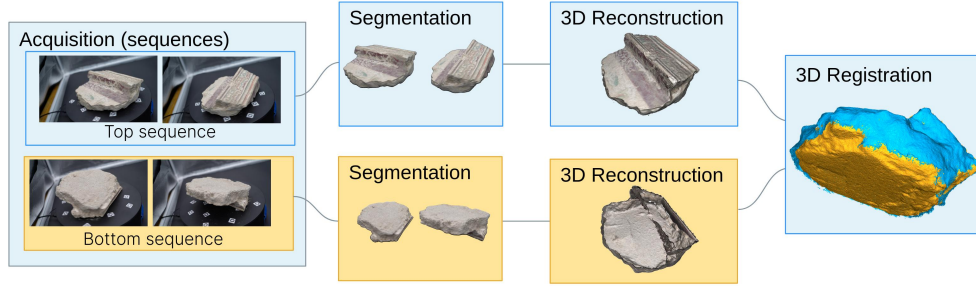


Fig. 2: Pipeline of the 3D reconstruction of a fresco fragment. After the acquisition of the images of both sides of the fragment, they are segmented to remove the background and used for the 3D reconstruction. Finally, the two partial reconstructions, colored in blue and yellow, are fused together via 3D registration.

3.1 Image segmentation of fresco fragments

Segmentation is a critical step in isolating the artifact from its background, ensuring that only the object of interest is used for accurate 3D reconstruction (See Fig 3). Conventional segmentation methods typically rely on fully supervised techniques, such as Mask RCNN [He et al (2017)], requiring full mask annotations for every image in the training dataset. While this approach is highly effective, its reliance on extensive manual annotations can be time-consuming and impractical in cultural heritage applications, especially when multiple views of the same object are needed. Additionally, fully annotating these masks for each image in a collection is not always feasible.

To mitigate the annotation burden, we propose an integrated strategy that combines Mask R-CNN with two automated annotation generation methods: RITM (Sofiuk et al (2021)) and STCN (Cheng et al (2021b)). RITM (Region-based Interactive Tracking Method) is an interactive segmentation technique that requires minimal user input, typically just a few strategically placed clicks to delineate object boundaries accurately. Selected for its effectiveness in

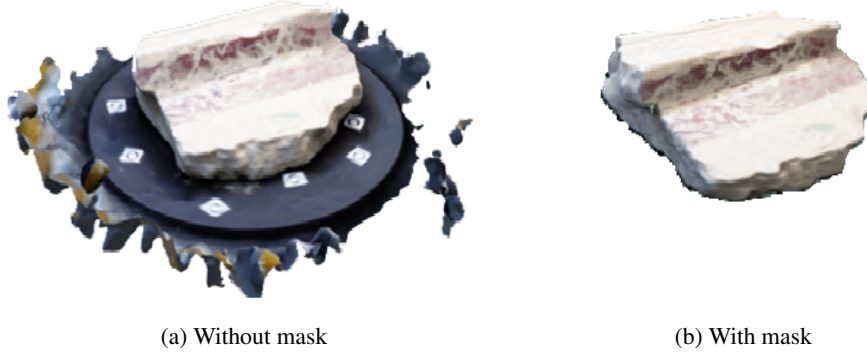


Fig. 3: A visual comparison of a complete reconstruction with and without the use of segmentation masks.

reducing manual effort while maintaining high accuracy, RITM³ remains highly competitive despite newer techniques and is well-suited for our application, particularly for generating high-precision initial masks when no prior annotations exist.

STCN (Space-Time Correspondence Network) complements RITM by enabling the propagation of initial segmentation masks across multiple frames from different perspectives. This is especially useful for reducing manual annotation efforts when dealing with large datasets consisting of multiple views of the same object. STCN ensures temporal consistency, allowing masks generated for one frame to be propagated to others with minimal human intervention. This method, particularly when combined with RITM, reduces the need for repeated annotations across sequential images, making the segmentation process more efficient. Moreover, STCN’s documented implementation within frameworks such as MiVOS (Modular Interactive Video Object Segmentation)⁴ enhances its applicability to custom datasets, aligning well with our need for automated yet accurate mask propagation in sequential image datasets.

By combining Mask R-CNN with RITM and STCN, we achieve a balance between automation and manual effort, allowing us to efficiently generate high-quality segmentation masks suitable for 3D reconstruction. Mask R-CNN is integrated into the pipeline for its strong generalization ability, crucial for segmenting fresco fragments captured from varying angles and conditions. Its capacity to recognize objects with different visual characteristics, even when not explicitly seen during training, ensures robust and consistent segmentation across diverse perspectives. This makes it an ideal tool for refining the initial masks generated by RITM and STCN, enhancing accuracy for the subsequent 3D reconstruction process. The segmentation process is structured into three strategies:

- **Strategy 1:** Initial segmentation is achieved by applying RITM for mask generation, followed by mask propagation using STCN across frames. Mask R-CNN is then applied to the resulting masks to refine the segmentation. After generating the initial masks using RITM

³<https://paperswithcode.com/paper/reviving-iterative-training-with-mask#code>

⁴<https://github.com/hkchengrex/MiVOS>



Fig. 4: An overview of different segmentation masks generated using the different approach described.

and propagating them across frames with STCN, we finetune a pre-trained Mask R-CNN model⁵ (He et al (2017)) using them.

- **Strategy 2:** Building on the predictions obtained by Mask R-CNN as a result of Strategy 1, RITM is then applied using *randomly* selected points to simulate user interaction.
- **Strategy 3:** After applying Strategy 1, the first frame’s mask is generated using the Mask R-CNN prediction, and subsequent frames are predicted using STCN propagation. This additional STCN round enhances segmentation consistency across frames, leveraging temporal coherence to maintain high precision with minimal manual intervention.

Each of these strategies optimizes the trade-off between manual effort and automation, enabling an evaluation of their impact on segmentation performance and the final 3D reconstruction quality, and emphasizing the importance of adapting workflows to efficiently process large datasets of fresco fragments captured from multiple angles. Table 1 summarizes these strategies and their descriptions.

Strategy	Method	Description
Strategy 1	RITM + STCN + Mask R-CNN	Initial RITM annotation followed by STCN mask propagation; Mask R-CNN is trained on resulting masks.
Strategy 2	Strategy 1 + RITM	Additional refinement of predicted masks with RITM by applying random user clicks for increased precision.
Strategy 3	Strategy 1 + STCN	Additional STCN propagation for refined mask consistency across frames.

Table 1: Summary of segmentation strategies with methods and descriptions

The integration of Mask R-CNN ensures high accuracy, while RITM and STCN contribute by automating initial mask generation and improving consistency across multiple frames. Figure 4 demonstrates the results of each strategy, showing how the segmentation quality improves from Strategy 1 to Strategy 3. Strategy 1 provides a reasonable approximation with slight inaccuracies, Strategy 2 refines the segmentation with RITM where simulated user clicks were randomly generated this time, and Strategy 3 minimizes the degree of human interaction required through additional STCN propagation.

⁵Used pre-trained Mask-RCNN model: https://github.com/facebookresearch/detectron2/blob/main/MODEL_ZOO.md

3.2 3D Reconstruction

This work focuses on reconstructing 3D models from only RGB images taken from multiple viewpoints, without requiring prior information such as camera poses. Therefore, we use methods that can estimate camera poses automatically. Three photogrammetry tools were evaluated for this purpose: COLMAP⁶ (Schonberger and Frahm (2016); Schönberger et al (2016)), Meshroom⁷ (Griwodz et al (2021)), and Metashape⁸ (Agisoft (2006)).

Colmap is widely used for pose recovery and sparse 3D reconstruction from sequential images. However, its dense reconstructions are often noisy and lack mesh generation capabilities (Rahaman and Champion (2019)). Metashape, on the other hand, has been shown to require fewer images for accurate reconstruction compared to Meshroom by (DJurić et al (2021)), leading us to select it for the experiments in this study.

All three tools follow a similar reconstruction workflow. Initially, the images are aligned to estimate the camera poses and generate a sparse point cloud, as shown in Figure 5. Next, depth maps are produced for each image using stereo matching, and these maps are merged to create a dense point cloud.

In our dataset, each fragment is captured from two perspectives: one focusing on the top and the other on the bottom. A 3D reconstruction is generated for each set using Metashape, resulting in two partial models. These models are then registered and merged to create a complete 3D representation of the fragment.

3.3 3D Registration

The registration of two partial 3D reconstructions—specifically, the upper and lower portions of each artifact fragment—is achieved through the state-of-the-art TEASER algorithm for 3D registration (Yang et al (2020)). TEASER finds the optimal transformation (scale, translation, and rotation) required to align two point clouds, effectively merging them into a cohesive model with the best possible fit.

To achieve this alignment, TEASER requires distinct feature descriptors from each point cloud. We extract these descriptors using a histogram-based approach from Rusu et al (2009),

⁶<https://github.com/colmap/colmap>

⁷<https://github.com/alicevision/meshroom>

⁸Agisoft Metashape: <https://www.agisoft.com/downloads/installer/>

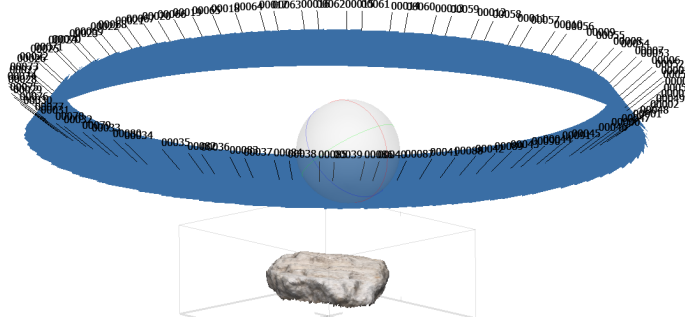


Fig. 5: 3D reconstruction with Metashape; camera poses are highlighted in blue.

where each descriptor vector comprises 33 values that capture intrinsic geometric properties of each point, such as distances, angles, and curvatures. These feature descriptors are invariant to transformations, ensuring that the characteristics of points are comparable across different perspectives.

Once the point descriptors are obtained, correspondences between points in the two partial reconstructions are established. TEASER then computes the optimal transformation based on these correspondences, aligning the two point clouds. Following this, we apply Iterative Closest Point (ICP) (Besl and McKay (1992)) for fine-tuning the alignment locally.

The results of the registration process are shown in Figure 6, where the upper (blue) and lower (yellow) sections of the fragment are seamlessly aligned, illustrating the effectiveness of the employed approach in achieving a precise 3D model.

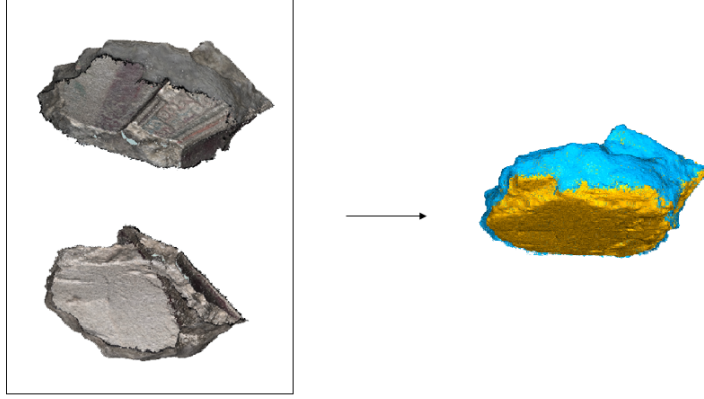


Fig. 6: Example of 3D registration for two partial models of a fresco fragment. The left side shows the separate upper and lower sections, while the right side illustrates the merged model, with blue and yellow shading to indicate their original alignment.

4 Experiments

To assess the effectiveness of our pipeline, we conducted comprehensive experiments focusing on both 2D segmentation accuracy and the quality of 3D reconstructions.

4.1 Dataset

Our dataset consists of 49 fresco fragments, separated into three sets: 10 for training Mask R-CNN, 14 for segmentation evaluation, and 25 for 3D reconstruction (see Table 2). We intentionally limited the training set to 10 fragments to minimize user intervention and demonstrate the pipeline’s efficiency with minimal data.

Each fragment was recorded with two 20-second videos (upper and lower parts) at 50 fps, yielding approximately 1000 frames per video. The fragments were placed on a rotating table under a fixed Sony a7c camera. To reduce the dataset size while preserving temporal

information, we sampled 10% of frames, averaging 100 frames per video at a resolution of 1920×1080 pixels. An example frame can be seen in Figure 7.

	Training Set	Segmentation Set	3D Reconstruction Set
# Fragments	10	14	25
Frames	2000	2800	5000

Table 2: The datasets used in this study: **# Fragments** is the total number of fresco’s fragment in each set, **Frames** is the total number of frames (images) within that set.



Fig. 7: An example of a frame of a fragment extracted from a video.

The ground truth segmentation masks were generated interactively using the RITM method (Sofiuk et al (2021)) with user verification. For the 3D models, ground truth data was obtained using a Polyga FlexScan3D scanner (Polyga (2022)), which employs structured light-based sensor technology to directly scan the fragments. The collected data was then processed with Polyga FlexScan3D software. The dataset will be made publicly available.

4.2 Performance metrics

To evaluate performance at each stage of the pipeline, we utilized different metrics tailored to the specific nature of the data, ranging from 2D segmentation to 3D reconstruction.

For the 2D segmentation task, the *Intersection over Union (IoU)* metric was used to assess the accuracy of the predicted segmentation masks. This metric compares the overlap between the estimated segmentation mask and the ground truth, measured at the pixel level.

Evaluating 3D reconstruction quality, however, is more complex due to the irregular and complex structure of the data. To address this, we used three point-based distance metrics to compare the reconstructed 3D model with the ground truth point cloud:

- *Mean Euclidean distance*: The average Euclidean distance between corresponding points in the ground truth and the reconstructed point cloud.
- *Chamfer distance*: A measure of the squared distances between points in the two point clouds, capturing overall discrepancies between the ground truth and reconstruction.
- *Hausdorff distance*: The maximum distance between a point in the ground truth and its nearest neighbor in the reconstructed model, highlighting the largest deviation in the geometry.

Finally, a *voxel-based IoU* metric is used to assess the volume. In this approach, the 3D data is re-sampled into a regular grid of fixed-size voxels. This allows us to apply standard techniques, such as the IoU used in 2D segmentation, to evaluate the overlap between the predicted and ground truth 3D volumes. The metric is calculated based on the average IoU across the voxel grid, with a voxel size of 2 used in the experiments.

4.3 Evaluation of 2D Segmentation Strategies

To evaluate the effectiveness of various 2D segmentation strategies for fresco fragments, three distinct approaches were tested, each designed to balance automation and manual effort while maintaining high segmentation accuracy across diverse perspectives. These strategies leverage Mask R-CNN, RITM, and STCN, with varying levels of refinement and automation, as detailed in Table 1. In the experiments, Mask R-CNN was pre-trained on the COCO dataset (Lin et al (2014)).

The performance of these strategies is summarized in Table 3, with key metrics being the mean Intersection over Union (IoU) and its standard deviation. All strategies demonstrated high accuracy, with mean IoU values exceeding 96%.

Strategy 2 achieved the highest mean IoU of 0.98544, reflecting its superior precision due to interactive refinement. However, this precision comes with a slight increase in variability (Std. Dev. 0.01222) compared to Strategy 1. Strategy 1 exhibited the lowest standard deviation (0.01117), indicating consistent segmentation across diverse fragment types, though its mean IoU of 0.96769 was slightly lower than the other strategies. Strategy 3 provided a balanced performance with a mean IoU of 0.97619. The additional STCN propagation improved temporal consistency but introduced a slightly higher standard deviation (0.02393), likely due to its reliance on initial Mask R-CNN predictions.

IoU	Strategy 1	Strategy 2	Strategy 3
Mean	0.96769	0.98544	0.97619
Std. Dev.	0.01117	0.01222	0.02393

Table 3: Comparative performance evaluation for 2D Segmentation Strategies

While accuracy is critical, execution time plays a significant role in the feasibility of large-scale applications (Table 4). Strategy 2, although the most precise, is considerably slower, taking 6 seconds per image and up to 10 minutes for two fragments. In contrast, Strategy 1 and Strategy 3 processed images in under 0.015 seconds and two fragments in less than 15 seconds, making them more suitable for large datasets where speed is essential.

IoU	Strategy 1	Strategy 2	Strategy 3
1 image	0.015 sec.	6 sec.	0.010 sec.
$\frac{1}{2}$ fragment	15 sec.	10 min.	10 sec.

Table 4: Execution Time Comparison for 2D Segmentation Strategies

Figure 8 presents the visual results of applying the segmentation masks to the input image, based on three different strategies. In (b) to (d), the segmentation masks obtained by Strategy 1, Strategy 2, and Strategy 3 are displayed, each corresponding to a different method for isolating the object of interest from the background. (e) to (g) show the results of applying these masks to the original image, effectively highlighting the segmented object. As seen in these visual outcomes, Strategy 2 yields the clearest object segmentation with minimal background interference, consistent with its superior performance in the quantitative metrics in Table 4. Strategy 1, while still effective, introduces some inaccuracies, leading to a less precise segmentation. Strategy 3, although refined through additional steps, still results in segmentation with slightly more background noise, failing to achieve the level of precision demonstrated by Strategy 2.

4.3.1 Erosion

The role of 2D segmentation goes beyond simply locating the fragment in the image as accurately as possible; it also provides crucial data for the subsequent 3D reconstruction step to maximize the accuracy of the final model. Therefore, we must not only focus on the precision of the segmentation in generating the masks but also consider how these masks impact the overall quality of the 3D reconstruction.

Through several experiments, it was observed that applying the erosion operation to the generated masks generally leads to a significant improvement in the 3D reconstruction, helping eliminate impurities such as outliers or small noise along the edges, as shown in Figure 9. Erosion is a morphological operator that reduces the mask’s area while maintaining its overall shape.

This improvement can be attributed to two main factors: first, the presence of frames from different viewpoints, where the loss of information from one frame is compensated by the data from another. Second, erosion reduces the likelihood of background pixels near the fragment’s edges being included in the 3D reconstruction, preventing them from introducing noise into the final model.

4.4 Partial 3D Reconstruction Analysis

In this phase, partial 3D reconstructions are generated for the bottom and top sides of each fragment. These partial models offer insights into the accuracy of the segmentation strategy before the registration step, which merges them into a complete 3D model. While the partial reconstructions provide valuable information about segmentation quality, they might exhibit minor inaccuracies along the borders due to the limitations in the camera and scanner’s field of view, resulting in slight discrepancies at the edges. It is important to note, however, that for a final, comprehensive analysis, the complete reconstruction—discussed in Section 4.5—should be considered.



(a) Input image



(b) Strategy 1 mask



(c) Strategy 2 mask



(d) Strategy 3 mask



(e) Strategy 1 result



(f) Strategy 2 result



(g) Strategy 3 result

Fig. 8: Visual segmentation results for the same input image using each strategy. (b) to (d) show the segmentation masks, and (e) to (g) show the masked images highlighting the object of interest. Strategy 2 provides the most accurate segmentation with minimal background interference.

Nevertheless, these partial reconstructions allow us to assess the segmentation's effectiveness independently of the complexities introduced by the registration process. Table 5 shows low error levels across the different strategies, confirming the reliability of the segmentation and 3D reconstruction algorithms. However, erosion of the segmentation masks did not significantly improve accuracy. The overall performance of the different strategies remains consistent, with no clear leader in terms of overall results.

Nonetheless, Mask R-CNN combined with STCN (Strategy 3) stands out as the most robust approach, particularly excelling in the Hausdorff distance, indicating better spatial alignment of the reconstructed surfaces. As shown in Table 4, this method is also the fastest solution among the tested strategies. Despite this, Strategy 2 offers higher IoU, suggesting that it captures a more accurate overall volume overlap with the ground truth. This difference can likely be attributed to Strategy 3's focus on preserving finer surface details, which contributes

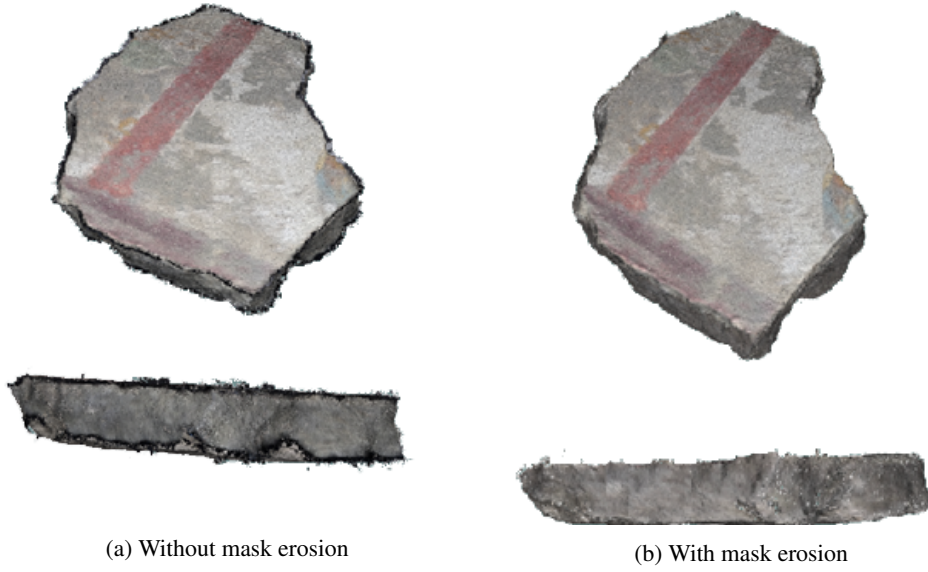


Fig. 9: Partial reconstruction with and without segmentation masks. The effect of erosion is particularly visible on the black points on the border of the artifact.

to improved spatial alignment but slightly reduces the volume captured, as indicated by the IoU.

Metrics	Strategy 1	Strategy 2	Strategy 3	Strategy 1 + 5% erosion	Strategy 2 + 5% erosion	Strategy 3 + 5% erosion
Mean ↓	0.442 ± 0.28	0.433 ± 0.30	0.433 ± 0.29	0.449 ± 0.28	0.438 ± 0.29	0.442 ± 0.26
Hausdorff ↓	2.022 ± 1.33	2.652 ± 1.58	1.852 ± 1.32	1.927 ± 1.40	2.220 ± 1.50	1.870 ± 1.34
Chamfer ↓	1.230 ± 1.69	2.552 ± 2.25	1.805 ± 1.67	1.192 ± 1.75	1.596 ± 2.23	1.116 ± 1.62
IoU ↑	0.731 ± 0.12	0.747 ± 0.12	0.734 ± 0.12	0.722 ± 0.12	0.742 ± 0.86	0.723 ± 0.19

Table 5: Results of partial 3D reconstruction using Metashape (Agisoft (2006)) with different 2D segmentation strategies.

4.5 Complete 3D Reconstruction Analysis

In this phase, individual parts of the fragment are merged into a complete 3D model, with the 3D registration process playing a key role in the final accuracy. While the partial reconstructions offer insight into the segmentation strategies, the complete model comparison takes into account all stages of the process: segmentation, reconstruction, and registration. This comprehensive analysis provides a clearer picture of each strategy’s overall performance.

The error values for the complete models, as shown in Table 6, are generally higher than those observed in the partial reconstructions. This is expected since the 3D registration process introduces additional complexity, amplifying the errors from the previous stages. However, the erosion operation significantly improves the accuracy across all segmentation approaches, highlighting its positive impact on the final model’s precision.

Metrics	Strategy 1	Strategy 2	Strategy 3	Strategy 1 + 5% erosion	Strategy 2 + 5% erosion	Strategy 3 + 5% erosion
Mean ↓	0.929 ± 0.35	1.051 ± 0.52	0.817 ± 0.35	0.772 ± 0.26	0.877 ± 0.30	0.756 ± 0.2
Hausdorff ↓	2.841 ± 1.24	4.187 ± 2.37	2.975 ± 1.34	2.378 ± 0.93	3.104 ± 1.30	2.389 ± 0.85
Chamfer ↓	2.138 ± 1.60	5.266 ± 5.84	2.356 ± 2.14	1.502 ± 0.94	2.640 ± 1.97	1.511 ± 0.88
IoU ↑	0.605 ± 0.12	0.552 ± 0.77	0.602 ± 0.11	0.619 ± 0.11	0.568 ± 0.12	0.604 ± 0.11

Table 6: Results on the complete 3D model after applying registration using Teaser [Yang et al \(2020\)](#).

The results in Table 6 demonstrate that Strategy 2 produces the worst results across all metrics. Both Strategy 1 and Strategy 3 achieve better performance, which aligns with the findings from the partial reconstructions. While the results from Strategy 1 and Strategy 3 are similar, Strategy 1 is more versatile since it does not rely on any assumptions about the temporal relationship between frames. In contrast, Strategy 3 leverages STCN, which requires temporal consistency between successive frames of the image sequence, making it more constrained in certain scenarios. Despite this, Strategy 3 ultimately delivers more refined results in terms of accuracy.

The presence of erosion enhances the accuracy of the 3D reconstruction, as demonstrated in the bottom row of Figure 10. Erosion helps to mitigate the issue of non-fragment regions (shown as black areas) interfering with the reconstruction, especially around the edges. This suggests that the erosion operation is particularly useful in refining the model by eliminating noise from peripheral areas, even though it does not entirely solve the issue. This reinforces the importance of using erosion as part of the 3D reconstruction pipeline, particularly for fragment-based objects where precise boundary delineation is critical.

In summary, while Strategy 1 performs well overall and Strategy 3 excels after erosion, Strategy 2 does not show strong results across all metrics. The application of erosion generally enhances reconstruction quality, but its effects are not uniform across all segmentation strategies.

In Figure 11, we present a series of 3D models reconstructed from various fragment types, showcasing the diversity of our dataset and the range of challenges these fragments present. (a) shows a flat fragment, while (b) highlights a textured fragment, demonstrating the complexity of capturing both geometric detail and intricate surface patterns. (c) displays an elongated fragment, and (d) focuses on a fragment with an irregular bottom, both of which introduce challenges in geometry alignment and texture mapping. The corresponding 3D reconstructions in (e) through (h) illustrate our pipeline’s ability to handle these diverse shapes, preserving both high geometric precision and texture fidelity. This variety highlights the versatility of our approach, which is capable of adapting to artifacts with differing shapes and surface characteristics, making it suitable for large-scale applications in cultural heritage digitization.

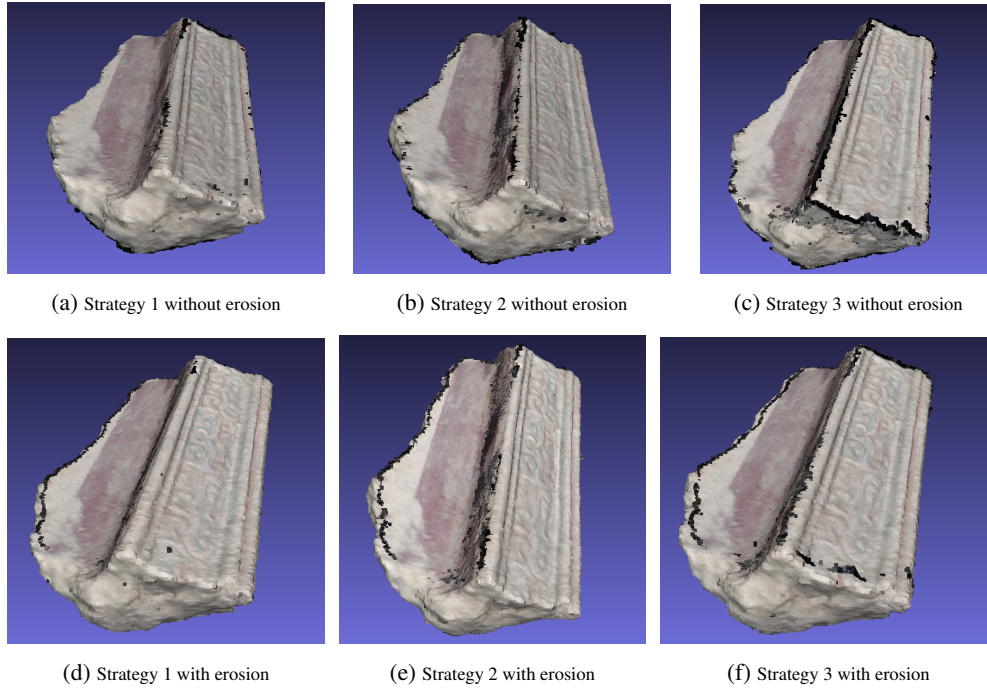


Fig. 10: 3d reconstructions with different strategies, with and without erosion.

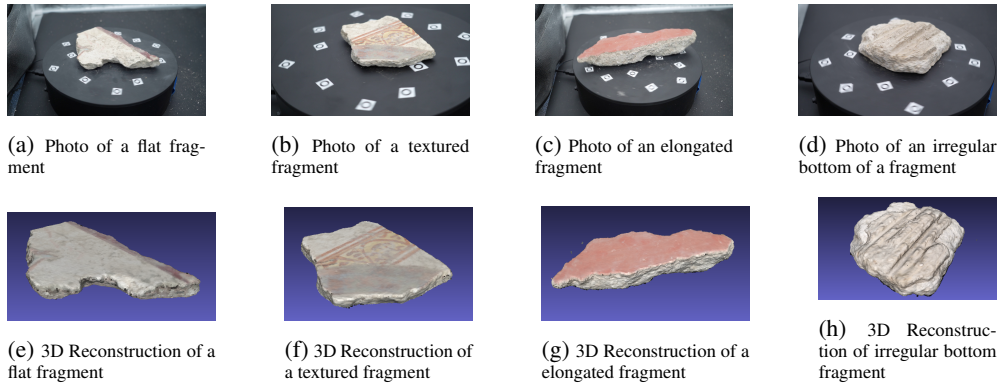


Fig. 11: 3D reconstructions of different artifact fragments.

4.6 Runtime Analysis

In this section, we present the time required to execute our pipeline, from the sequence of images to the final 3D reconstruction of fresco fragments.

Image segmentation of fresco fragments

Table 7 shows the time needed to perform the dataset annotation used for fine-tuning the Mask R-CNN in Strategy 1. The first column (RITM) represents the time taken to interactively annotate all frames of one side of the object using RITM. The second column outlines the case where only the first frame is annotated manually with RITM, while the remaining 99 frames are automatically annotated through mask propagation by the STCN.

As shown, combining the two semi-supervised approaches (RITM and STCN) significantly speeds up the annotation process, achieving a reduction of three orders of magnitude in time. Annotating the entire training dataset for Mask R-CNN, which consists of 10 fragments $\times$ 2 views $\times$ 100 images each = 2000 images, takes approximately 500 seconds. Once the training set is generated, fine-tuning the Mask RCNN model takes about 90 minutes to complete 3000 iterations on a T4 GPU.

Once the Mask R-CNN is fine-tuned, the time required to segment the fragments is relatively low, ~ 0.015 seconds per images. A single face of the fragments requires ~ 15 seconds. The Table 8 presents the time required to perform segmentation with RITM using simulated random user clicks and fragment segmentation with finetuned Mask R-CNN.

3D Reconstruction

The 3D reconstruction of each fragment’s face is performed with Metashape. The time required when considering the frames and their masks is ~ 173 seconds while around ~ 346 seconds when the masks are not available. The use of masks helps us in halving the time needed for the 3D reconstruction and at the same time avoid any post-processing on the 3D models.

3D Registration

Preprocessing the images is not affecting the timing of the 3D registration. The time required by TEASER is almost the same considering both the mask or not, and it is around ~ 13 seconds to merge two point clouds.

Considering the design choices of our pipeline, we need 1 annotated image for each partial reconstruction (2 in our case). Performing the 3D reconstruction of a single fragment using our proposed pipeline takes around 3 minutes: 15 + 15 seconds (Mask RCNN) + 173 seconds (MetaShape) + 13 seconds (Teaser) $\simeq$ 3 minutes. Performing the 3D reconstruction using only interactive segmentation with RITM would take around 23 minutes: 600 + 600 seconds (RITM) + 173 seconds (MetaShape) + 13 seconds (Teaser) = $\simeq$ 23 minutes.

Training Set Generation	RITM	RITM (1 frame) + STCN (99 frames)
100 Frames (seconds)	2000	50

Table 7: The time (in seconds) required to annotate the dataset for Mask R-CNN fine-tuning using different segmentation strategies.

Image Segmentation	RITM	Mask R-CNN
100 Frames (seconds)	~ 600	~ 15

Table 8: The time (seconds) required to perform mask generation using RITM and Mask R-CNN for 100 frames.

5 Conclusions

In this work, we presented a comprehensive pipeline for digitizing and reconstructing fragmented fresco artifacts, addressing the challenges of large-scale cultural heritage restoration. By integrating automated segmentation techniques and photogrammetric tools, our approach reduces the need for extensive manual intervention while ensuring high-quality 3D reconstructions with detailed geometry and textures.

Our pipeline demonstrates significant scalability and adaptability, particularly for handling extensive datasets of fragmented cultural heritage objects. The produced 3D models have been shown to meet the stringent quality requirements for both archival purposes and the reassembly of frescoes, making them indispensable for the broader objectives of the EU Horizon 2020 RePAIR project.

We emphasized that the incorporation of segmentation masks in the 3D reconstruction process enhances the quality of the models, minimizes the creation of unwanted scene elements, and accelerates the overall reconstruction workflow. Our modular pipeline accommodates various segmentation methods, which were analyzed to assess their respective strengths and limitations. Additionally, we quantitatively evaluated the effectiveness of applying erosion to segmentation masks and examined the errors introduced during the registration process when merging multiple partial reconstructions into a complete model.

Declarations

5.1 Funding

This work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 964854

5.2 Conflict of interest/Competing interests

Not applicable

5.3 Consent for publication

All authors given their consent for the publication

5.4 Data availability

No data will be made available through this work. The full dataset of the RePAIR project, on which this work is grounded, will be made available to the public once it will be concluded.

5.5 Ethics approval and consent to participate

Not applicable

5.6 Materials availability

Not applicable

5.7 Code availability

Not applicable

5.8 Author contribution

Conceptualization: Gianluca Bison, Luca Palmieri, Sinem Aslan, Sebastiano Vascon;
Methodology: Gianluca Bison, Luca Palmieri, Sinem Aslan, Sebastiano Vascon;
Formal analysis and investigation: Gianluca Bison, Luca Palmieri, Sebastiano Vascon;
Writing - original draft preparation: Gianluca Bison (initial submission), Luca Palmieri (initial submission), Sebastiano Vascon (initial submission), Sinem Aslan (full rewrite and restructuring during major revision, including response to reviewers);
Writing - review and editing: Gianluca Bison, Luca Palmieri, Sinem Aslan, Sebastiano Vascon, Marcello Pelillo;
Funding acquisition: Marcello Pelillo;
Supervision: Sebastiano Vascon, Marcello Pelillo;

References

- Agisoft (2006) Agisoft Metashape. URL <https://www.agisoft.com/>, accessed: Nov. 12, 2024
- Aslan S, Pelillo M (2019) Weakly supervised semantic segmentation using constrained dominant sets. In: International Conference on Image Analysis and Processing, Springer, pp 425–436
- Bai J, Wu X (2014) Error-tolerant scribbles based interactive image segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 392–399
- Besl P, McKay ND (1992) A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2):239–256. <https://doi.org/10.1109/34.121791>
- Boykov Y, FLG (2006) Graph cuts and efficient n-d image segmentation. International Journal of Computer Vision 70:109–131
- Brown BJ, Laken L, Dutré P, et al (2012) Tools for virtual reassembly of fresco fragments. International Journal of Heritage in the Digital Era 1(2)

- Carsten G, Simone G, Lilian C, et al (2021) Alicevision meshroom: An open-source 3d reconstruction pipeline. In: Proc. 12th ACM Multimed. Syst. Conf. - MMSys '21. ACM Press, <https://doi.org/10.1145/3458305.3478443>
- Cheng HK, Tai YW, Tang CK (2021a) Modular interactive video object segmentation: Interaction-to-mask, propagation and difference-aware fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 5559–5568
- Cheng HK, Tai YW, Tang CK (2021b) Rethinking space-time networks with improved memory coverage for efficient video object segmentation. *Advances in Neural Information Processing Systems* 34:11781–11794
- Cheng MM, Prisacariu VA, Zheng S, et al (2015) Densecut: Densely connected crfs for realtime grabcut. In: *Computer Graphics Forum*, Wiley Online Library, pp 193–201
- Chibane J, Alldieck T, Pons-Moll G (2020) Implicit functions in feature space for 3d shape reconstruction and completion. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 6968–6979, <https://doi.org/10.1109/CVPR42600.2020.00700>
- Choy C, Park J, Koltun V (2019) Fully convolutional geometric features. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 8958–8966
- Di Angelo L, Di Stefano P, Guardiani E (2022) A review of computer-based methods for classification and reconstruction of 3d high-density scanned archaeological pottery. *Journal of Cultural Heritage* 56:10–24. <https://doi.org/10.1016/j.culher.2022.05.001>
- DJurić I, Vasiljević I, Obradović M, et al (2021) Comparative analysis of open-source and commercial photogrammetry software for cultural heritage. In: eCAADe 2021 International Scientific Conference, pp 8–10
- Genova K, Cole F, Sud A, et al (2020) Local deep implicit functions for 3d shape. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 4856–4865, <https://doi.org/10.1109/CVPR42600.2020.00491>
- Girshick R, Donahue J, Darrell T, et al (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
- Griwodz C, Gasparini S, Calvet L, et al (2021) Alicevision Meshroom: An open-source 3D reconstruction pipeline. In: Proceedings of the 12th ACM Multimedia Systems Conference - MMSys '21. ACM Press, <https://doi.org/10.1145/3458305.3478443>
- Hatır ME, İnce İ, Korkanç M (2021) Intelligent detection of deterioration in cultural stone heritage. *Journal of Building Engineering* 44:102690

- He K, Gkioxari G, Dollár P, et al (2017) Mask r-cnn. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp 2980–2988, <https://doi.org/10.1109/ICCV.2017.322>
- Hou M, Huo D, Yang Y, et al (2024) Using mask r-cnn to rapidly detect the gold foil shedding of stone cultural heritage in images. *Heritage Science* 12(1):46
- Huang S, Gojcic Z, Usvyatsov M, et al (2021) Predator: Registration of 3d point clouds with low overlap. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 4267–4276
- Javaid M, Haleem A, Pratap Singh R, et al (2021) Industrial perspectives of 3d scanning: Features, roles and it's analytical applications. *Sensors International* 2:100114. <https://doi.org/10.1016/j.sintl.2021.100114>
- Kong X, Hucks RG (2023) Preserving our heritage: A photogrammetry-based digital twin framework for monitoring deteriorations of historic structures. *Automation in Construction* 152:104928
- Lin T, Maire M, Belongie SJ, et al (2014) Microsoft COCO: common objects in context. CoRR abs/1405.0312. URL <http://arxiv.org/abs/1405.0312>, [arXiv:1405.0312](https://arxiv.org/abs/1405.0312)
- Llull C, Baloian N, Bustos B, et al (2023) Evaluation of 3d reconstruction for cultural heritage applications. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 1642–1651
- Luo H, Zhang J, Liu X, et al (2024) Large-scale 3d reconstruction from multi-view imagery: A comprehensive review. *Remote Sensing* 16(5):773
- Mildenhall B, Srinivasan PP, Tancik M, et al (2021) Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* 65(1):99–106
- Murez Z, van As T, Bartolozzi J, et al (2020) Atlas: End-to-end 3d scene reconstruction from posed images. In: Vedaldi A, Bischof H, Brox T, et al (eds) European Conference on Computer Vision (ECCV), pp 414–431
- Osipov A, Ostanin M, Klimchik A (2023) Comparison of point cloud registration algorithms for mixed-reality cross-device global localization. *Information* 14(3):149
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9(1):62–66. <https://doi.org/10.1109/TSMC.1979.4310076>
- Park JJ, Florence P, Straub J, et al (2019) DeepSDF: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 165–174
- Pollefeys M, Van Gool L, Vergauwen M, et al (2003) 3d recording for archaeological fieldwork. *IEEE Computer Graphics and applications* 23(3):20–27

- Polyga (2022) Polyga Flexscan 3D Software. URL <https://www.polyga.com/flexscan3d-software/>, accessed: Nov. 12, 2024
- Quattrini R, Pierdicca R, Paolanti M, et al (2020) Digital interaction with 3d archaeological artefacts: evaluating user's behaviours at different representation scales. *Digital Applications in Archaeology and Cultural Heritage* 18:e00148. <https://doi.org/https://doi.org/10.1016/j.daach.2020.e00148>
- Rahaman H, Champion E (2019) To 3d or not 3d: Choosing a photogrammetry workflow for cultural heritage groups. *Heritage* 2(3):1835–1851. <https://doi.org/10.3390/heritage2030112>
- Raytrix (2008) Inspection. URL <https://raytrix.de/inspection/>, accessed: Nov. 12, 2024
- Ren S, He K, Girshick R, et al (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28
- Rodríguez-Gonzálvez P, Rodríguez-Martín M, Ramos LF, et al (2017) 3d reconstruction methods and quality assessment for visual inspection of welds. *Automation in Construction* 79:49–58. <https://doi.org/https://doi.org/10.1016/j.autcon.2017.03.002>, URL <https://www.sciencedirect.com/science/article/pii/S0926580517301863>
- Rodriguez-Garcia B, Guillen-Sanz H, Checa D, et al (2024) A systematic review of virtual 3d reconstructions of cultural heritage in immersive virtual reality. *Multimedia Tools and Applications* pp 1–51
- Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, et al (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, Cham, pp 234–241
- Rusinkiewicz S, Levoy M (2001) Efficient variants of the icp algorithm. In: *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, pp 145–152, <https://doi.org/10.1109/IM.2001.924423>
- Rusu RB, Blodow N, Beetz M (2009) Fast point feature histograms (fpfh) for 3d registration. In: *2009 IEEE International Conference on Robotics and Automation*, pp 3212–3217, <https://doi.org/10.1109/ROBOT.2009.5152473>
- Schonberger JL, Frahm JM (2016) Structure-from-motion revisited. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4104–4113
- Schönberger JL, Zheng E, Frahm JM, et al (2016) Pixelwise view selection for unstructured multi-view stereo. In: *14th European Conference on Computer Vision (ECCV 2016)*, Amsterdam, The Netherlands, October 11-14, 2016, *Proceedings, Part III* 14, Springer, pp 501–518

- Sofiuk K, Petrov I, Barinova O, et al (2020a) f-brs: Rethinking backpropagating refinement for interactive segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 8623–8632
- Sofiuk K, Petrov I, Barinova O, et al (2020b) f-brs: Rethinking backpropagating refinement for interactive segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 8623–8632
- Sofiuk K, Petrov I, Konushin A (2021) Reviving iterative training with mask guidance for interactive segmentation. arXiv preprint arXiv:210206583
- Sofiuk K, Petrov IA, Konushin A (2022) Reviving iterative training with mask guidance for interactive segmentation. In: 2022 IEEE International Conference on Image Processing (ICIP), IEEE, pp 3141–3145
- Theoharis T, Papaioannou G, Bjørlykke K, et al (2013) Multi-scale 3d digitization at nidaros cathedral: from archiving to large-scale visualization. In: 18th Conference on Cultural Heritage and New Technologies
- Tsesmelis T, Palmieri L, Khoroshiltseva M, et al (2024) Re-assembling the past: The repair dataset and benchmark for real world 2d and 3d puzzle solving. arXiv preprint arXiv:241024010 Project page: https://repairproject.github.io/RePAIR_dataset/
- Wang J, Zhang C, Wang P, et al (2023) Batch-based model registration for fast 3d sherd reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 14519–14529
- Wang Y, Solomon J (2019) Deep closest point: Learning representations for point cloud registration. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp 3522–3531, <https://doi.org/10.1109/ICCV.2019.00362>
- Yang H, Shi J, Carlone L (2020) Teaser: Fast and certifiable point cloud registration. IEEE Transactions on Robotics 37(2):314–333
- Yavartanoo M, Chung J, Neshatavar R, et al (2021) 3dias: 3d shape reconstruction with implicit algebraic surfaces. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp 12446–12455
- Zemene E, Pelillo M (2016) Interactive image segmentation using constrained dominant sets. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VIII 14, Springer, pp 278–294