Crowdsourced Neural Rendering for Participatory Cultural Digital Twins in Smart Cities

Leonardo Bordoni*, Riccardo Rosati*, Lorenzo Stacchio†, Juri Vitali*, Primo Zingaretti*
*Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy
†Department of Political Sciences, Communication and International Relations, University of Macerata, Macerata, Italy

***Abstract*—Neural rendering can support cultural heritage, entertainment, tourism, smart-city services, and real-time digital twinning by transforming visual observations into navigable 3D representations. However, its practical adoption in urban and cultural scenarios is still constrained by acquisition requirements that often depend on professional operators, drones, controlled capture protocols, and expertise in photogrammetry or AI-based 3D reconstruction. This paper presents a novel participatory pipeline developed for the generation of cultural digital twins from crowdsourced images and videos. The key idea is to exploit data collected by citizens and tourists while visiting cultural places, turning discontinuous individual recordings into a more continuous and updatable visual memory of monuments and urban assets. We describe a workflow that combines mobile acquisition, rewarding, administrative validation, frame selection, pose estimation, neural rendering, and Web and XR visualization. Preliminary experiments on the Palla di Pomodoro and Palazzo della Prefettura use cases show the potential of smartphone videos and drone imagery, while highlighting problems of data scarcity, insufficient view overlap, device variability, reflective surfaces, and unstable convergence. The paper discusses how crowdsensing can complement professional 3D digitization and enable scalable cultural digital twins for smart cities.**

***Index Terms*—Neural Rendering, 3D Gaussian Splatting, NeRF, Cultural Heritage, Digital Twins, Crowdsourcing, Smart Cities**

I. Introduction

Neural rendering methods, including Neural Radiance Fields (NeRF) [1], instant neural graphics primitives [2], and 3D Gaussian Splatting (3DGS) [3], have changed the way image collections can be transformed into photorealistic 3D scenes. These representations are particularly relevant for cultural heritage and smart-city applications, where the objective is not only to reconstruct isolated objects, but also to support continuous visualization, immersive storytelling, remote access, and public engagement. In this sense, a cultural digital twin can become a reusable interface for tourism, education, entertainment, urban monitoring, and XR-based civic services. Despite this potential, neural reconstruction remains difficult to scale in real city contexts. High-quality capture is usually performed by trained professionals with drones, calibrated cameras, or dense acquisition trajectories. This creates a gap between the ambition of continuously updated city-scale twins and the practical availability of images, camera poses, and metadata. The problem becomes more evident when cultural assets cannot be captured with drones, when access is limited, or when public institutions do not have continuous access to AI and 3D digitization expertise.

We introduce **CRONR** (*Crowdsourced Reconstruction through Open Neural Rendering*), a participatory pipeline that transforms images and videos collected by citizens and tourists into reusable neural-rendered representations and digital-twin assets. Instead of relying only on professional campaigns, the system exploits crowd data collection, allowing visitors to contribute short videos or image sequences while exploring a place. This approach is consistent with Smart Cities ecosystems, whose verticals include culture, tourism, and engagement, and whose research lines include proof of attendance, real-time digital twins, and interactive multimedia environments. The contribution of this work is a compact description of the CRONR pipeline and of the practical issues observed during preliminary neural rendering experiments on cultural assets in Pesaro.

II. Participatory Pipeline

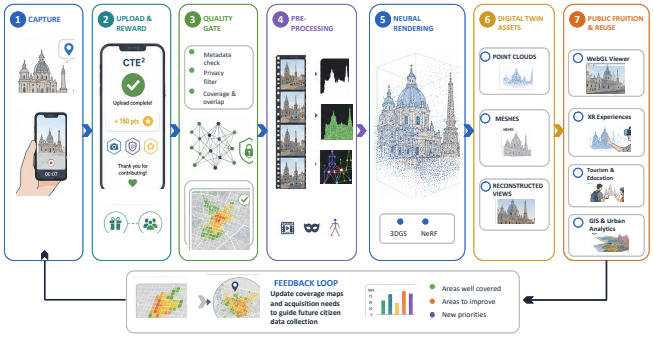
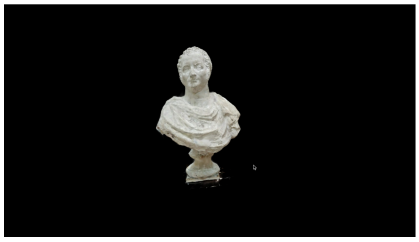


Fig. 1: CTE^2 workflow: crowdsourced capture is filtered, curated, reconstructed through neural rendering, and published as reusable digital twin content for Web, VR/AR, tourism, and smart-city services.

Fig. 1 summarizes the proposed workflow. The first layer concerns data capture. Citizens and tourists acquire images or videos using common smartphones; professional or institutional acquisitions, such as drone imagery, can be added when available to improve coverage. The second layer concerns participation. A mobile application allows users to upload and visualize data, while rewarding mechanisms make the contributor an active protagonist in territorial promotion. This layer is central because it transforms data acquisition from an expert-only task into a distributed cultural activity.

(a) Palla di Pomodoro

(b) Prefettura



(c) Cultural object

Fig. 2: Qualitative neural-rendering results used as visual evidence for the CRONR pipeline.

The third layer is technical validation. Crowdsourced data are heterogeneous and must be filtered before reconstruction. The system therefore needs quality gates for blurred frames, redundant views, missing metadata, privacy-sensitive content, and insufficient overlap. Selected frames are then aligned through structure-from-motion or neural pose estimation, optionally supported by masks, segmentation, inpainting, or manual curation. The fourth layer is neural rendering, where the processed data can be reconstructed using NeRF-based models, 3DGS, or commercial/cloud-based pipelines. Finally, the resulting assets can be exported as novel views, point clouds, meshes, textured models, or WebGL/XR visualizations. This architecture is designed to support two complementary goals. On the one hand, it produces digital twin assets for cultural fruition, such as virtual tours, interactive museum content, or city-scale visualizations. On the other hand, it produces a feedback loop for acquisition: missing views, weakly reconstructed regions, and poor-quality areas can be translated into new capture needs shown to future visitors or administrators.

III. Preliminary Experiments

The initial testbed focused on two urban-cultural use cases: the Palla di Pomodoro and Palazzo della Prefettura in Pesaro [1]. Experiments were conducted with Instant-NGP using smartphone videos, smartphone photographs from multiple devices, combinations of photographs and video frames, and drone photographs. Qualitatively, smartphone videos produced usable reconstructions in several cases, especially with mobile devices. By contrast, sparse photographs from two devices often failed to converge, probably because of the limited number of samples and insufficient overlap. Combining sparse photographs with video frames did not automatically improve the results and sometimes worsened convergence. Drone photographs produced the most stable qualitative outputs, but they reintroduce the acquisition constraints that the participatory scenario aims to reduce.

The experiments expose three main limitations. First, data scarcity is not only a matter of the number of images, but also of trajectory quality: neural rendering requires viewpoint diversity, stable exposure, and sufficient parallax. Second, public cultural scenes contain reflective surfaces, transparent materials, moving people, cars, vegetation, and variable illumination; these elements can produce floaters, blurred geometry, or inconsistent appearance. Third, crowdsourced acquisition increases device variability and makes automatic reconstruction more fragile, because image quality, focal length, compression, rolling shutter, and capture behavior are not controlled. The results indicate that participatory reconstruction requires selective data validation. The pipeline should guide users toward useful captures, assess coverage and quality, and support curators in selecting reliable sequences before publishing validated digital twin assets.

IV. Discussion and Conclusion

CTE^2 proposes a scalable alternative to expert-only cultural digitization. The goal is not to replace professional surveys, but to complement them with a continuous, low-cost, and socially engaging acquisition layer. In smart-city contexts, such a layer can help keep cultural digital twins updated, extend coverage to places that are not frequently surveyed, and connect digital reconstruction with tourism, proof-of-attendance, gamification, and immersive fruition. Future work will compare 3DGS, NeRFs, SDF-based methods, LUMA-like pipelines, and photogrammetry baselines under controlled and crowdsourced acquisition conditions, also including quantitative novel-view synthesis metrics such as PSNR, SSIM, and LPIPS on held-out views. However, these metrics will need to be interpreted carefully in crowdsourced settings, where viewpoint distribution, illumination changes, device variability, and pose-estimation errors may affect direct comparisons. Further developments will address automatic quality assessment, 360° imagery, reflection handling, inpainting-assisted reconstruction, mesh extraction, and WebXR deployment. These steps are necessary to transform isolated neural rendering experiments into robust participatory digital twins for cultural heritage and smart cities.

[1]This work was partially funded by MIMIT, under FSC project "Pesaro CTE SQUARE", CUP D74J22000930008.

References

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[2] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics*, vol. 41, no. 4, pp. 1–15, 2022.

[3] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 1–14, 2023.

[4] M. Tancik et al., "Nerfstudio: A modular framework for neural radiance field development," in *Proc. ACM SIGGRAPH*, 2023.