A Large Reconstruction Model Driven Approach to Support Humans in Digitization of Dance Visual Material into 3D environments

Silvia Garzarella[1,**], Lorenzo Stacchio[2,**], Pasquale Cascarano[1], Allegra De Filippo[3], Elena Cervellati[1] and Gustavo Marfia[1]

[1]Department of the Arts, University of Bologna, Italy
[2]Department of Political Sciences, Communication and International Relations, University of Macerata, Italy
[3]Department of Computer Science and Engineering, University of Bologna, Italy

Abstract

Heritage in the domain of dance amounts to a vast set of multimodal information, representing both tangible and intangible materials. Modern systems leverage different Artificial Intelligence (AI)-driven paradigms to enhance the preservation, accessibility, quantitative data analysis, and valorization of dance heritage. One particular outcome of this application is the generation of linked semantic information among multimodal data regarding a particular dance entity, which is, however, hard to interpret and visualize. For this reason, Extended Reality and Immersive paradigms could be employed to visualize it through immersive approaches, also easing its manipulation. However, involving tangible material, there is still a gap on how to directly project objects, entities, and processes that were captured in flat 2D pictures into the 3D realm. Since manual 3D modeling are labor-intensive, we here introduce and discuss a Large Reconstruction Driven Framework for accelerating digitazion of visual material, also integrating discriminative AI approaches to generate 3D models starting from 2D pictures, through a human-in-the-loop (HITL) and controllable approach. To validate the approach, we applied it to a specific case study, linked to the artistic legacy of the dancer and choreographer Rudolf Nureyev, to digitize its multimodal materials. The implications of the proposed framework could impact various creative industries and cultural heritage preservation efforts.

Keywords

Artificial Intelligence, Extended Reality, Cultural Heritage, Computational Dance, Human Collaboration

1. Introduction

Dance cultural heritage (DCH) encompasses a wide array of both tangible and intangible elements, reflecting its inherently multimodal nature. These sources vary in type, geographical origin, and medium, resulting in a complex and diverse body of information that must be carefully considered during the digitization process [1, 2]. In the context of theatrical dance in particular, historiographical approaches often rely on comparisons between written documentation, choreographers' descriptions, and dancers' embodied memory. These are essential not only for the study of choreography but also for understanding the performance as a dramaturgical unit in its entirety, encompassing scenography, costumes, music, and other stage elements. Among the various sources that constitute dance cultural heritage, theatrical costumes occupy a particularly significant role. Their inherent fragility, resulting from their organic composition, combined with frequent use and a high risk of loss—whether due to company relocations, deterioration, or dispersal through auctions—renders them especially vulnerable. Nevertheless, costumes are far from marginal: they transmit key elements essential to the understanding

International Workshop on Human-AI Collaborative Systems (HAIC 2025), co-located with ECAI, October 25–30, 2025, Bologna, Italy
*Corresponding author.
**Equal Contribution.
silvia.garzarella2@unibo.it (S. Garzarella); lorenzo.stacchio@unimc.it (L. Stacchio); pasquale.cascarano2@unibo.it (P. Cascarano); allegra.defilippo@unibo.it (A. De Filippo); elena.cervellati@unibo.it (E. Cervellati); gustavo.marfia@unibo.it (G. Marfia)
0000-0001-8499-2542 (S. Garzarella); 0000-0002-9341-7651 (L. Stacchio); 0000-0002-1475-2751 (P. Cascarano); 0000-0002-1954-7271 (A. De Filippo); 0000-0002-8038-001X (E. Cervellati); 0000-0003-3058-8004 (G. Marfia)

of choreographic works, both in visual and performative terms [3, 4]. Beyond their function in defining the aesthetic and narrative dimensions of the choreography and the characters embodied, costumes also encapsulate the specific technical features of the performances for which they were created, the physicalities of the dancers for whom they were designed, and the traces left by those who have worn them over time. As such, the costume may be understood as a 'speaking' object, endowed with an agency capable of conveying crucial narratives—narratives that are instrumental in preserving the memory of the performative event [5]. Historically theorized since the 18th century, as in Noverre's *Lettres sur la danse et sur les Ballets* [6], dance costumes have been recognized not only for their visual function but also for their influence on expressivity and technique. Recent studies have further explored their impact on bodily movement and identity formation [7].

However, despite their significance, costumes are still predominantly archived and studied in two-dimensional formats, such as photographs, sketches, or videos, which limits the understanding of their volumetric, material, and functional properties. To overcome this constraint, we propose the projection of these 2D representations into 3D digital forms. Doing so would enable a more realistic and interactive approach to analysis, preservation, and application, particularly in Extended Reality (XR) environments, enhancing the accessibility and interpretability of dance heritage assets [8]. Indeed, XR environments allow users—researchers, curators, students, or performers—to engage with heritage materials in a spatial and embodied manner, thus bridging the gap between data representation and experiential understanding [9].

While 3D models offer the potential to overcome these limitations by enabling more realistic and immersive representations, the process of manually modeling and reconstructing them from the available set of 2D images remains a highly laborious task. It requires specialized skills from both developers and designers, involving complex workflows such as photogrammetry, mesh modeling, texture mapping, and rigging. These challenges often make large-scale digitization efforts impractical without substantial time and resources. This lays the opportunity for computer science, particularly in the domains of artificial intelligence (AI). Recent advances have indeed provided new tools for addressing these challenges, particularly to project 2D images in the 3D realm, by pipelining Discriminative and Generative Artificial Intelligence (GenAI) approaches, such as Diffusion Models and Neural Rendering [10, 11, 12, 13]. Modern Deep Learning (DL) models can indeed detect and isolate target visual object in archival material, enabling not only their classification across image and videos, but also to re-use this information to generate novel 3D data using GenAI, in particular Large Reconstruction models (LRM) [14, 13]. LRM is typically a transformer-based model designed to reconstruct 3D representations (e.g., Neural Radiance Fields, Gaussian splats, Triplanes) directly from a single input image or sparse view image [14, 13].

The combination of those two approaches could, in principle, accelerate 3D reconstruction of dance costumes. However, an easy and usable way for humans to analyze, correct, and modify such output is required in order to increase the artistic value of the generated object and correct possible errors.

To address this gap, we propose a novel AI pipeline that combines discriminative and generative paradigms within a human-in-the-loop (HITL) framework, enabling the automatic generation of 3D representations from 2D images. Unlike fully automated 3D reconstruction methods, which often struggle with artifacts, inconsistencies, or missing semantic details, our approach integrates expert feedback at critical stages of the pipeline. The involvement of the human not only guides the system toward higher-quality reconstructions but also ensures cultural and historical accuracy, which is particularly crucial in the context of heritage digitization.

This pipeline is implemented as an extension of the DanXe framework [15], an open and extensible platform designed to accelerate the labeling, digitization, and reconstruction of cultural heritage data. DanXe provides modular tools for integrating computer vision models, metadata enrichment, and human–AI collaboration, for building domain-specific extensions (i.e., dance domain) such as the one proposed in this work.

To validate our approach, we conducted a case study centered on the artistic legacy of Rudolf Nureyev, one of the most iconic figures in 20th-century dance. Nureyev's legacy, comprising thousands of multimodal materials, offers an ideal testbed: his international career and the extensive mediation of his public image make the study of his costumes particularly challenging. At the same time, however,

this research is particularly challenging: many of the costumes he wore have been sold at auction, are dispersed across multiple archives, and have been extensively used, rendering them extremely fragile and difficult to study [16]. In this context, the human-in-the-loop is crucial for resolving ambiguities and maintaining semantic fidelity, highlighting the added value of our approach compared to fully automated AI pipelines.

This work contributes to the emerging field of GenAI for digital humanities 3D preservation by demonstrating how hybrid AI workflows with a HITL approach can scale the preservation of tangible cultural heritage, promoting interpretability and supporting creative re-use in artistic and curatorial practices. Beyond dance, our methods hold promise for a broad range of cultural heritage applications and creative industries, from museum curation to fashion, theater, and performative arts.

2. Related Works

In this Section, we review recent works that allow projecting 2D images to 3D models, involving various methods, described below.

Among the most widespread approaches, there is classical Photogrammetry. We here mention Structure-from-Motion (SfM) [17], which reconstructs 3D geometry from a series of 2D images by analyzing their relative positions and orientations. Specifically, SfM methods detect and match key points between pairs of images, estimate camera poses, and triangulate 3D coordinates by intersecting the rays back-projected from the cameras. The process culminates in a sparse 3D model, which can be further densified and textured through additional steps. Furthermore, Multi-View Stereo (MVS) [18] methods have been developed. Unlike SfM, MVS estimates depth information by analyzing multiple images of the same scene, leading to dense 3D reconstructions. MVS offers improved accuracy, particularly in detailed environments, but it can be computationally intensive and sensitive to image quality and viewpoint variations. These techniques are typically organized into sequential pipelines comprising stages such as feature detection, camera calibration, depth estimation, mesh generation, and texture mapping. While effective, these pipelines are labor-intensive and demand significant expertise from developers and designers. Moreover, they require a large enough set of pictures depicting the scene/object we want to reconstruct and may not cope with illuminations variations and occlusions.

Recent advancements in GenAI, particularly diffusion models, enabled the creation of detailed 3D models from few or a single 2D image. These models leverage learned priors from extensive datasets to infer depth, geometry, and texture, effectively "hallucinating" unseen perspectives [19, 20, 21, 13]. The most commonly adopted approach involves two key stages: **(1)** Multi-View Image Synthesis: Starting from a single input image, a multi-view diffusion model generates a set of novel views from different angles. This process simulates capturing the object from multiple viewpoints, providing diverse perspectives necessary for 3D reconstruction. **(2)** 3D Reconstruction via Neural Rendering (NR): The synthesized multi-view images are then input into a neural rendering pipeline, such as Neural Radiance Fields (NeRF) or 3D Gaussian Splatting (3DGS) or Neural Implicit Surfaces (Neus) to reconstruct a coherent 3D model [22, 11].

For instance, Zero-1-to-3 [19] utilizes a conditional diffusion model trained on synthetic datasets to generate novel views of an object from a single image, facilitating 3D reconstruction without the need for multiple viewpoints. Similarly, Wonder3D [21] employs a cross-domain diffusion approach to produce consistent multi-view normal maps and color images, which are then fused to reconstruct high-quality 3D meshes. With a similar approach, [20] focuses on 3D point cloud reconstruction from a single image by developing a Consistency Diffusion Model. The model incorporates 3D structural priors and 2D image priors to guide the diffusion training process, enhancing the consistency and quality of the reconstructed 3D models.

While this GenAI approach has reached a level of maturity that enables their applications, they still face fundamental challenges. A key limitation is the fragmentation of 3D data formats—meshes, point clouds, Radiance Fields, and 3D Gaussians—each optimized for specific rendering or geometry tasks. Existing 3D models often perform well in one domain (e.g., appearance with NeRFs or geometry

with meshes), but struggle to generalize across formats or simultaneously model detailed shape and texture. A recent work introduced Trellis [13], a framework that employs a unified and versatile latent space for 3D generation, known as Structured LATents (SLAT). The main innovation lies in combining sparse voxel-based spatial structures with powerful pretrained vision features to enable high-quality, representation-agnostic 3D modeling. Once such a latent structure is built, dedicated decoders can translate it into different output formats, including Radiance Fields (NeRF-like rendering), 3DGS, and classical Triangle Meshes. For this reason, we employed such an approach in our Dance Costumes reconstruction pipeline.

However, despite the goodness of such models, images in the Dance Heritage domain always contain several elements, making hit hard to isolate a particular element to be fed into such models. For this reason, we also employed a state-of-the-art Segmentation model (i.e, Segment Anything [23]) to let humans easily isolate Dance Costumes within an image and then perform the 3D reconstruction [24].

3. Methodology

We here provide a detailed overview of the materials and methodologies employed in our study. We begin with the **Dance Costumes Dataset** subsection, which describes the collection and characteristics of the images used for analysis. Following this, the **AI Augmented Human 3D Generator** subsection outlines the HITL approach that leverages our DanXe extension to support humans in reconstructing annotation efficiency and accuracy.

3.1. Dance Costumes Dataset

As previously noted, the choice to focus on Rudolf Nureyev stemmed from the complexity and richness of the documentary traces surrounding his career. His international prominence and the highly mediated nature of his public image make the study of his costumes particularly meaningful, as they are deeply intertwined with the way he constructed his stage presence and shaped his persona in the collective imagination. At the same time, these costumes present significant challenges: many have been sold at auction, are dispersed across various archives, and have undergone extensive use, making them extremely fragile and difficult to access. [16].

The majority of Nureyev's costumes are preserved in the collection of the Rudolf Nureyev Foundation, the institution that safeguards and manages the rights to his legacy, and have been donated to the Centre national du costume et de la scène in Moulins, France. These collections are accessible in person by visiting the museum, online through digital images hosted on its website, or via printed volumes dedicated to the collection. In addition to this central repository, further documentation exists in the form of photographs, video recordings—both commercially distributed and informally shared online—as well as in the archives of theatres where Nureyev performed, which often still hold original costumes. His costumes are also frequently referenced in the extensive press coverage devoted to his career. To develop a Dance Costumes Dataset capable of capturing the complexity of such a case study, we sought to simulate a use context closely aligned with real-world scholarly research. To this end, we adopted the perspective of a researcher engaging with these materials, aiming not only to support academic inquiry, but also to inform the design of a digitization strategy that meets the needs of both researchers and cultural institutions. This approach was intended to enhance the research experience, facilitate access, and offer archives and collections a tool to valorize their holdings while preserving the integrity and material specificity of these fragile artifacts. In the case of costumes, this means preserving and reactivating their full agency, while also maintaining a strong awareness of their physical presence and tactile qualities.

To this end, we tested the potential of digital technologies to transform widely accessible two-dimensional resources into a three-dimensional digital object capable of approximating the experience of direct engagement with the original costume. The images used for this purpose—shown in Figure 2—were sourced online from the website of the Centre national du costume et de la scène in Moulins (Samples 1), from a press survey (Sample 2), through dialogue with the archivists of Teatro alla Scala

in Milan (Sample 3)and from a publicly available video (Sample 4). These materials, although varied in quality and format, reflect realistic conditions for digital reconstruction work, especially when direct access to physical collections is limited. The reconstruction process was implemented through a pipeline articulated in four distinct phases, each designed to recover and convey the visual, material, and performative dimensions of the costume in a coherent and meaningful way.

3.2. AI Augmented Human 3D Generator

We propose a HITL pipeline for digitizing and 3D reconstructing historical dance costumes from single-view archival images. The proposed pipeline is visually depicted in Figure 1. Such a pipeline was designed to address the digitization of dance costumes from archival images, aiming to balance automation with expert oversight.

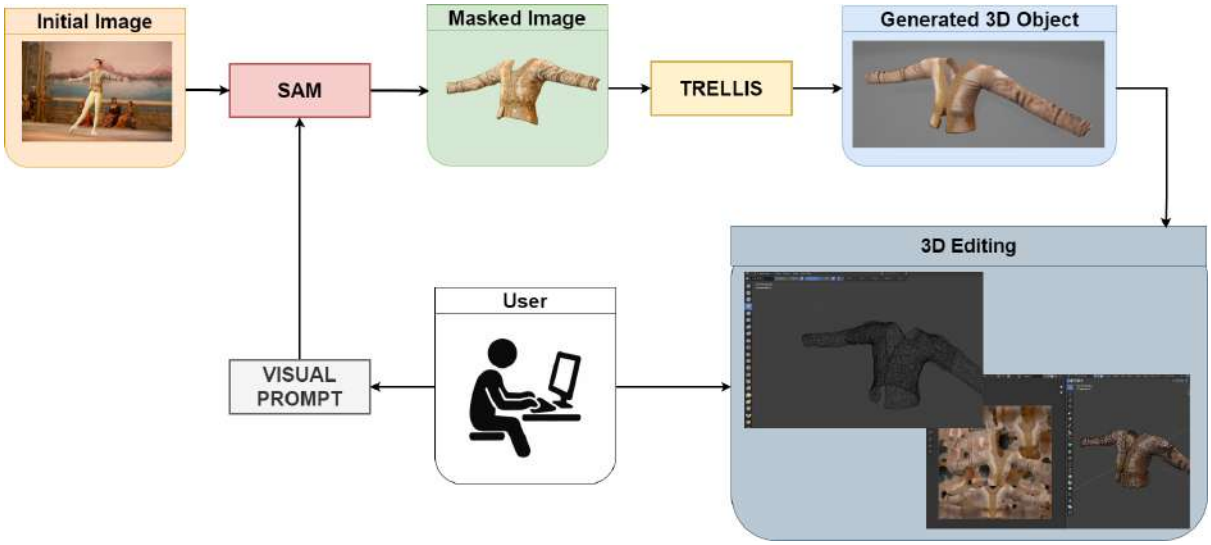

Figure 1: Overview of the proposed human-in-the-loop 3D digitization pipeline for costume heritage reconstruction. Starting from an initial image containing a dance costume, the SAM segmentation model is used to extract a masked image of the costume based on user-provided visual prompts. The isolated costume is then fed into TRELLIS, which synthesizes a 3D object from the isolated 2D image. The resulting 3D model is exported to a 3D editing environment (Blender in our case), where experts or designers can refine geometry and texture.

The pipeline proceeds through the following sequential steps:

- **Input Acquisition:** The process begins with a 2D archival image that typically portrays a dancer in full costume during a historical performance or staged photograph.
- **Dance Costume Isolation:** The image is processed using the SAM model, which enables the user to isolate the costume by providing visual prompts (e.g., points or bounding boxes). This generated a masked image of the costume, preserving only the visual information relevant to the garment.
- **3D Reconstruction:** The masked image is fed into the TRELLIS model, which infer novel views and internal 3D structure. The model then decodes the latent representation into a *triangle mesh*, suitable for further manipulation.
- **Export and Post-Processing:** The generated 3D mesh is then exported. The user can then import it into 3D editors (in our case, *Blender*), where they can refine, for example, the topology, UV mapping, and material properties.

It is worth mentioning that the process is iterative and interactive: the user can iteratively guide the segmentation or prompt generation for improved reconstructions. This system offers a scalable and interpretable workflow for reconstructing tangible dance cultural heritage artifacts (in our case, costumes), preserving not only their visual aesthetics but also their volumetric and functional traits for archival, educational, and curatorial purposes.

4. Results

We recruited a Dance Researcher and Expert to use the previously described extension of the DanXe framework on the Dance Costumes dataset.

To let our users easily employ those models, we developed custom user interfaces in Python (v3.10). In particular, the expert first segmented dance customers' pixels in images through a basic SAM interface, taken from the different samples included in our test dataset. Then, s/he employs a custom-implemented User Interface (via the gradio library [1]) to generate the final 3D model starting from the masked image, employing the TRELLIS architecture. We here state that all the considered models and user interfaces were deployed on a high-performance Linux server equipped with dual Intel® Xeon® Gold 6254 CPUs (72 threads total) and an NVIDIA RTX A6000 GPU (48 GB VRAM), running CUDA 12.2.

For each generated 3D model, we implemented a multi-angle rendering to generate standardized 2D projections, leveraging the trimesh and pyrender libraries. For each mesh file in the dataset, we extracted the primary geometry and constructed a scene with a white background and uniform ambient lighting. A perspective camera was positioned at a fixed distance along the z-axis, and a directional light source was co-located with the camera to ensure consistent illumination across renderings. We applied a series of rigid body rotations to the mesh around the canonical X and Y axes, rotating in 45-degree increments, resulting in a total of eight distinct viewpoints per model. The rendered RGB images were captured offscreen at a resolution of 1280×1280 pixels and saved to disk with systematic naming based on the axis and rotation degree.

Some qualitative results are reported in Figure 2, which compares 2D reference images and reconstructed 3D garment models across multiple views. Each row corresponds to a different sample, showcasing four views: the original 2D reference on the left, followed by three rendered views of the corresponding 3D reconstruction—front, left, and right.

To further confirm the perceptual quality of the 3D renderings shown in Figure 2, we conducted a quantitative image quality assessment across all generated views. Specifically, considering that we don't possess any 3D ground truth, we employed the No Reference metric CLIP-IQA model [25] with a ResNet-50 backbone and resolution of 512 pixels, as implemented in the PyIQA library, to compute no-reference perceptual quality scores. Our evaluation considers all the rendered images for each generated costume mesh, which is fed to the considered model, returning a scalar score that reflects the perceptual alignment of the content. These scores were collected and summarized per folder using statistical metrics for each object. The results are reported in Table 1.

Sample	Min	Max	Mean	Std Dev
Sample 1	0.3742	0.5116	0.4420	0.0331
Sample 2	0.3092	0.4312	0.3522	0.0316
Sample 3	0.3324	0.4339	0.3872	0.0312
Sample 4	0.3962	0.4752	0.4408	0.0199

Table 1
CLIP-IQA scores for rendered 3D views of each sample.

The results in Table 1 show how, across all the considered samples, the mean CLIP-IQA scores range from 0.3522 to 0.4420 (on average), indicating a generally good alignment with the perceptual priors encoded in CLIP embeddings. Sample 1 achieves the highest mean score (0.4420) and also the widest dynamic range between minimum and maximum values (0.1374), suggesting a mix of high-quality views and some lower-quality ones, possibly due to view-specific artifacts. Sample 4 closely follows in quality with a slightly lower mean (0.4408) but exhibits the smallest standard deviation (0.0199), indicating more consistent visual quality across views. In contrast, Sample 2 presents the lowest mean (0.3872), reflecting a perceptual quality that is less aligned with high-fidelity reference images, and may reflect limitations in texture reconstruction or shape fidelity in that sample. Finally, Sample 2 exhibited

[1] https://www.gradio.app/

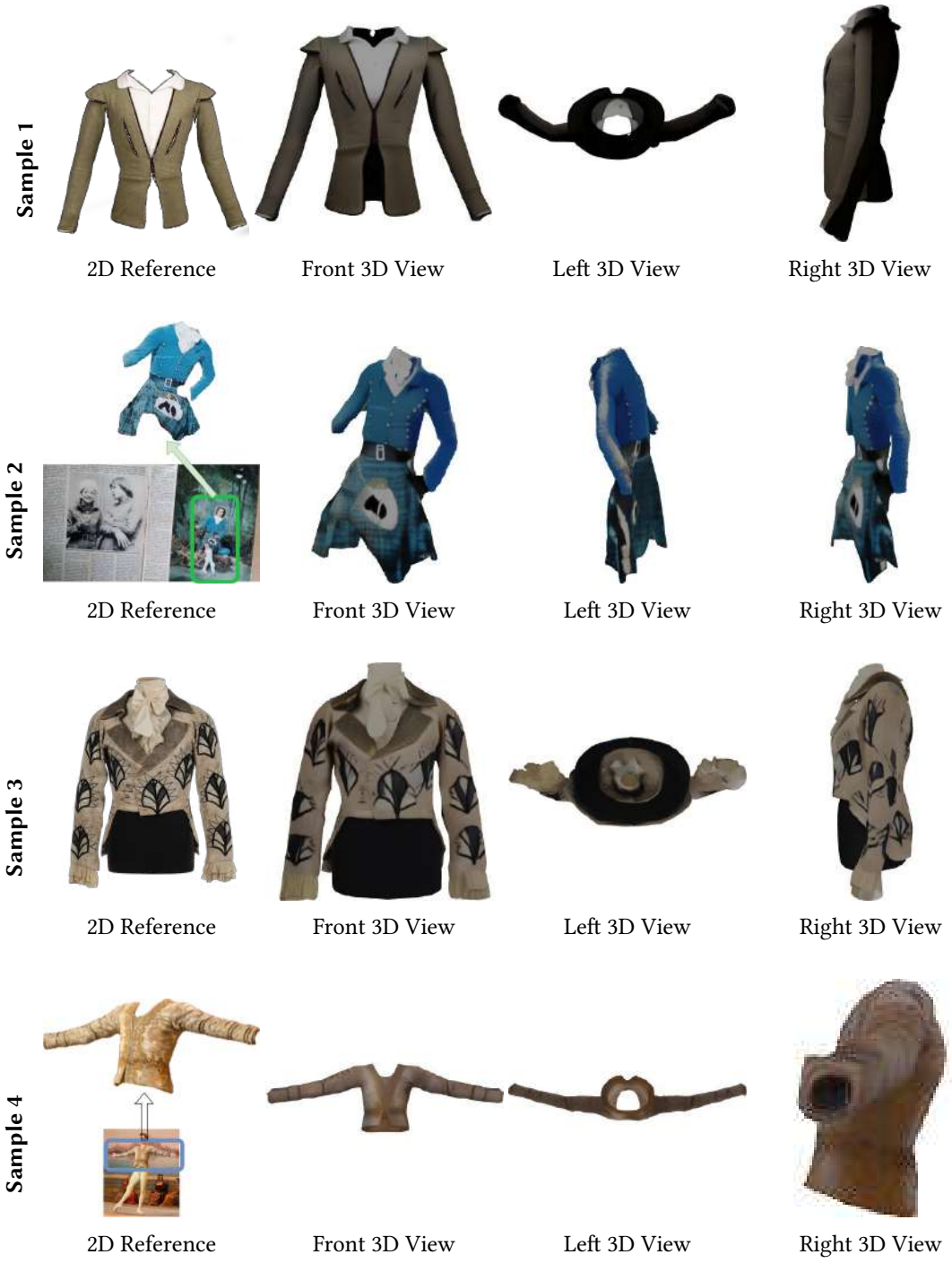

Figure 2: Comparison between 2D references and 3D model views for different samples. Sample 1 shows the costume for Giselle (1962), all rights reserved by the CNCS. Sample 2 features the costume for La Sylphide, taken from an article published in The Observer on July 2, 1972. Sample 3 presents the costume for The Nutcracker (1969), all rights reserved by the Historical Archives of Teatro alla Scala, Milan. Sample 4 is taken from the film The White Crow (2018), all rights reserved by Sony Pictures Classics.

the worst result, which could be attributed to the fact that the costume image is a low-resolution crop from a printed journal, exhibiting compression artifacts and lacking fine detail. Additionally, the figure is partially occluded and captured in a dynamic pose with complex clothing.

5. Discussions and Conclusions

In this work, we presented a HITL framework as an extension of the previously introduced DanXe toolbox[15], to support the 3D reconstruction of dance costumes from archival 2D imagery. Our proposed system integrates different AI models to map 2D visual material into editable 3D assets. The HITL approach aims to balance efficiency and accuracy with the domain expertise of cultural heritage professionals.

The adoption of state-of-the-art Generative AI models such as TRELLIS enabled the generation of high-quality 3D reconstructions from single masked costume images. From the results obtained, it is clear that the model is capable of reconstructing dance costumes with notable fidelity. These promising outcomes were further validated through expert-driven analysis, demonstrating that the pipeline can support culturally meaningful reconstructions. Additionally, the quality of the generated models was assessed using the CLIP-IQA metric, confirming alignment with visual expectations.

However, the specific context of use imposes important considerations that extend beyond technical performance. First, it must be acknowledged that these technologies rely on predictive models and therefore inherently involve a degree of approximation. While this may be acceptable for general visual representation, it becomes problematic in the case of historical dance costumes, where surface details—such as embroidery, ornamentation, and textile textures—are often essential. These elements are not only crucial to the historical and aesthetic integrity of the garments, but also play a significant role in conveying choreographic meaning, as illustrated in Sample 3. Inaccuracies at this level risk compromising both scholarly interpretation and the performative value of the original object.

The quality and source of input images also have a marked impact on reconstruction fidelity. High-resolution images, such as those provided by the Centre national du costume et de la scène (Samples 1 and 2) or archival photographs from institutions like Teatro alla Scala (Sample 3), yield particularly strong results. In more realistic and widespread scenarios, however, such as when working with press clippings or video stills, limitations quickly become apparent.

One notable issue is that models generated from performance photographs tend to replicate the dynamic pose of the costume as it appears in motion (Sample 4), rather than presenting it in a neutral, static posture suitable for exhibition or study. This necessitates a post-processing phase to reposition the model, adding to the complexity of the workflow. Nonetheless, these same dynamic representations offer a unique opportunity: they enable reconstructions that reflect the dancer's original body in motion, rather than the idealized or standardized form of a display mannequin. This opens the possibility for a more advanced and context-aware use of 3D modeling—not simply as a tool for static digitization, but as a medium to explore the costume's performative agency.

However, several technical limitations remain. The effectiveness of the reconstruction is highly dependent on the quality of the segmentation input, which can be compromised by cluttered backgrounds or poor-resolution imagery—common issues in archival video or press photography. Furthermore, while TRELLIS provides a flexible and efficient pipeline, the current implementation would benefit from domain-specific fine-tuning to enhance the fidelity of both geometry and textures. This is especially important in the case of dance costumes, where the rear of the garment often differs from the front in both design and material treatment, posing specific challenges for reconstruction algorithms and reinforcing the need for targeted calibration and expert oversight.

In our future works, we will first consider a more varied and curated dataset of historical costumes to evaluate the overall quality of the actual system. Then, we will explore the fine-tuning for the considered image-to-3D approach and also integrate additional deep learning models to restore damaged images (e.g., super-resolution methods). Finally, we planned to conduct structured evaluations with curators, archivists, and dance historians to assess the usability of the implemented pipeline, as well as the interpretability and perceived authenticity of the reconstructions. To conclude, this work contributes to the growing body of research at the intersection of GenAI, digital humanities, and HITL approaches, demonstrating how generative modeling and human-centered design can support the preservation and reinterpretation of dance cultural heritage material.

Acknowledgments

This work was partly funded by: (i) the PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 8 "Pervasive AI", funded by the European Commission under the NextGeneration EU program; (ii) MICS (Made in Italy – Circular and Sustainable) Extended Partnership and received funding from the European Union Next-Generation EU (PNRR – M4C2, Investimento 1.3 - D.D. 1551.11-10-2022, Partenariato Esteso PE00000004); (iii) a Ph.D. scholarship funded by the "PON Ricerca e Innovazione 2014-2020" project.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-5 to: Grammar and spelling check, and reword. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] J. Adshead-Lansdale, J. Layson, Dance history: An introduction, Routledge, 2006.
[2] M. De Marinis, Il corpo dello spettatore. performance studies e nuova teatrologia, Sezione di Lettere (2014) 188–201.
[3] E. Giannasca, Dance in the ontological perspective of a document theory of art, Danza e ricerca. laboratorio di studi, scritture, visioni 10 (2018) 325–346.
[4] E. Randi, Primi appunti per un progetto di edizione critica coreica, SigMa-Rivista di Letterature comparate, Teatro e Arti dello spettacolo 4 (2020) 755–771.
[5] V. Isaac, Donatella barbieri, costume in performance: Materiality, culture and the body, 2018.
[6] P. Dotlačilová, Costume in the Time of Reforms: Louis-René Boquet Designing Eighteenth-Century Ballet and Opera, Ph.D. thesis, Stiftelsen för utgivning av teatervetenskapliga studier (STUTS), 2020.
[7] E. Cervellati, et al., Storia della danza, Pearson Italia spa, 2020.
[8] I. Rakkolainen, A. Farooq, J. Kangas, J. Hakulinen, J. Rantala, M. Turunen, R. Raisamo, Technologies for multimodal interaction in extended reality—a scoping review, Multimodal Technologies and Interaction 5 (2021) 81.
[9] P. Kourtesis, A comprehensive review of multimodal xr applications, risks, and ethical challenges in the metaverse, Multimodal Technologies and Interaction 8 (2024) 98.
[10] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, et al., Sam 2: Segment anything in images and videos, arXiv preprint arXiv:2408.00714 (2024).
[11] A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk, W. Yifan, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi, et al., Advances in neural rendering, in: Computer Graphics Forum, volume 41, Wiley Online Library, 2022, pp. 703–735.
[12] R. Gao, A. Holynski, P. Henzler, A. Brussee, R. Martin-Brualla, P. Srinivasan, J. T. Barron, B. Poole, Cat3d: Create anything in 3d with multi-view diffusion models, arXiv preprint arXiv:2405.10314 (2024).
[13] J. Xiang, Z. Lv, S. Xu, Y. Deng, R. Wang, B. Zhang, D. Chen, X. Tong, J. Yang, Structured 3d latents for scalable and versatile 3d generation, in: Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 21469–21480.
[14] Y. Hong, K. Zhang, J. Gu, S. Bi, Y. Zhou, D. Liu, F. Liu, K. Sunkavalli, T. Bui, H. Tan, Lrm: Large reconstruction model for single image to 3d, arXiv preprint arXiv:2311.04400 (2023).
[15] L. Stacchio, S. Garzarella, P. Cascarano, A. De Filippo, E. Cervellati, G. Marfia, Danxe: an extended artificial intelligence framework to analyze and promote dance heritage, Digital Applications in Archaeology and Cultural Heritage (2024) e00343.
[16] C. Barnes, Nureyev, New York, NY: Helene Obolensky Enterprises, 1982.

[17] M. J. Westoby, J. Brasington, N. F. Glasser, M. J. Hambrey, J. M. Reynolds, 'structure-from-motion'photogrammetry: A low-cost, effective tool for geoscience applications, Geomorphology 179 (2012) 300–314.

[18] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, A comparison and evaluation of multi-view stereo reconstruction algorithms, in: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), volume 1, IEEE, 2006, pp. 519–528.

[19] R. Liu, R. Wu, B. Van Hoorick, P. Tokmakov, S. Zakharov, C. Vondrick, Zero-1-to-3: Zero-shot one image to 3d object, in: Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 9298–9309.

[20] G. Daras, Y. Dagan, A. Dimakis, C. Daskalakis, Consistent diffusion models: Mitigating sampling drift by learning to be consistent, Advances in Neural Information Processing Systems 36 (2023) 42038–42063.

[21] X. Long, Y.-C. Guo, C. Lin, Y. Liu, Z. Dou, L. Liu, Y. Ma, S.-H. Zhang, M. Habermann, C. Theobalt, et al., Wonder3d: Single image to 3d using cross-domain diffusion, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, pp. 9970–9980.

[22] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, W. Wang, Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction, arXiv preprint arXiv:2106.10689 (2021).

[23] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al., Segment anything, arXiv preprint arXiv:2304.02643 (2023).

[24] H. Qu, Y. Zhang, C.-H. Zhang, M. Huang, C. Hua, Sam-assisted 3d reconstruction: A novel framework for unseen object modeling in unknown environments, in: 2024 China Automation Congress (CAC), IEEE, 2024, pp. 6732–6737.

[25] J. Wang, K. C. Chan, C. C. Loy, Exploring clip for assessing the look and feel of images, in: Proceedings of the AAAI conference on artificial intelligence, volume 37, 2023, pp. 2555–2563.