International Journal of Education and Social Development
ISSN: 3078-9931 (Print), ISSN: 3078-2287 (Online) | Vol. 9, No. 1, 2026

Diffusion-Based Generative Image Technologies for Urban Cultural Heritage: New Shifts in Public Perception of Authenticity and Cultural Cognition.

Mingyu Zheng

Bartlett School of Environment, Energy and Resources, University College London, WC1E 6BT, United Kingdom

Abstract: This paper studies how diffusion-based generative image technology may change public perception of authenticity and cultural cognition in urban cultural heritage interpretation. To avoid viewing generative digital technology as limitless, this paper will focus on text-to-image and image-to-image diffusion systems for generating, modifying and diversifying visual representations of historical streets, buildings and urban areas. This paper presents concepts and practices; it does not conduct original research through data collection of visitors or other empirical evidence to support the conclusions. According to the doctrine of conservation, the theory of authenticity, media theory, digital heritage guidance and technical literature, it is proposed that generative images shift the concept of authenticity from an apparently self-evident visual attribute to a judgment about evidence, provenance, uncertainty and institutional trust. They provide comparisons and questions, but photorealism may also conceal speculation and reduces controversial urban history. Therefore, the responsible workflow should separate records from hypotheses, disclose the generation process, keep the source link, mention communities, and not present them as fact by virtue of visual plausibility.

Keywords: Diffusion models; urban cultural heritage; perceived authenticity; cultural cognition; digital interpretation; provenance.

1. Introduction

Urban cultural heritage can be experienced through old buildings and streets, as well as maps, pictures, shows, exhibitions, cell phones, virtual reconstructions, etc. Generative Visual Systems are another kind of mediated environment. The user can enter a text description, add an old photo, or mark a damaged area; then, the system will generate a realistic scene of the restored facade, remove later additions, or simulate a former time. The output is attractive and convenient to share, but due to its fluidity, there is an interpretation problem: although an image may be historically accurate, it may contain unsupported details, and these details could be composed of different periods or have been added by the model.

This paper narrows the general term "generative digital technologies" to diffusion-based text-to-image and image-to-image systems applied in the public interpretation of urban heritage. Traditional photogrammetry, laser scanning and manually created 3D models are not included unless they provide support or constraints for generation. Not the problem of physical protection or automated repair. It is the public-facing production of hypothetical still images and visual variations of historical streets, buildings, districts, and daily urban life.

Diffusion-based heritage imagery does not just expand access to the past, according to the above analysis. It modifies the conditions under which people perceive the authenticity of culture. Visual Realism is attention-grabbing and prompts people to think, but it can also be mistaken for the truth. Therefore, the actual problem is to build an interpretation system that uses the generated images as disclosed hypotheses rather than silent substitutions for records. Based on the analysis of concepts, this study proposes a governance-oriented workflow. There are no original datasets, surveys, experiments, visitor numbers or statistical tests in this paper to show that generative images have produced any measured changes in a particular population.

2. Conceptual and Doctrinal Foundations

The first type of modern conservation was based on material continuity and the readability of historical evidence. According to the Venice Charter, conservation and restoration should protect monuments as works of art and historical evidence, and restoration should cease where conjecture begins [1]. The rule does not forbid imaginative visualisation of the generated urban images; however, there should be an obvious division between the documented fabric and the proposed completion. A seamless facade generated by a short prompt cannot automatically obtain the evidential status of measured drawings just because it is in the same frame.

In addition to form and material, authenticity at the level of value and credibility in the Nara Document is also determined by factors such as use, tradition, environment, spirit, and others [2]. In cities, the broader concept of heritage may also be embodied in practices, routes, memories and social relations, etc., in addition to masonry. Authenticity does not need to take the form of pixel-level similarity. Although a new market street has been built in the same style, living vendors, minority history and the changing demands of the community have not been incorporated.

Public authenticity is also formed by the organisation of experience. MacCannell's idea of staged authenticity is that, by arranging the environment, one can make it seem as though one has entered an authentic old area [3]. Generative reconstruction may achieve a similar effect by focusing more visual attention on "the city as it was originally", and later Wang distinguished between objective, constructive and existential authenticity to show that meaningful experience

cannot always be explained by the verified status of an object alone [4]. Therefore, emotional engagement in the interpretation of heritage can be genuine without requiring historical verification of the represented details. In short, they are not the same: experience is one's own actual experience, and documentary truth is based on the verification and traceability of historical facts.

Digital images also remedy older problems. Bolter and Grusin have presented new media as forms of photography, painting, film, etc., that compete for the desire of transparent immediacy and visible mediation [5]. Diffusion-generated heritage scenes increase this tension. Photorealism can allow us to forget about the medium, and prompts, masks, confidence labels, and source panels display the constructed interface. Responsible heritage communication should not be achieved through technical obscurity. The Medium can be used to add comments and context for the picture, etc.

3. Definition of Technology

One particular kind has been chosen intentionally. Diffusion models learn to reverse the addition of noise and generate samples from the training data. The first Denoising Diffusion Model is a parameterised process that learns to reverse gradual noise addition [6]. Later, Latent Diffusion introduced a compressed form of these operations and added conditioning mechanisms for text-guided synthesis, inpainting, and super-resolution efficiently [7]. In a heritage environment, the above functions can generate a scene according to a verbal description, modify a selected part of a photograph, extend an incomplete image, or produce multiple visual options from the same input.

Generation is not equivalent to retrieval. The output is not a photograph taken in an archive, and it is also not an automatic geometric reconstruction of surveyed ruins. It is a new image obtained by the model under prompt and interface conditions. Although historical photographs and drawings provide conditions for the process, the system may generate reasonable textures and objects that do not exist in the sources. Therefore, "reconstruction" should be used only for the results of explicit evidence and reconstruction rules. Looser-constrained results are better described as generated hypotheses, interpretations or visual scenarios.

It is not the same as the one in the paper. There is no particular model, comparison of output accuracy, or display of user responses. No public databases have been used, and no generated corpora are provided as evidence. Technical papers are used only to show what the diffusion system will achieve. Claims of possible heritage uses are reasoned applications, not findings that a specific technology improves learning, trust, visitation or conservation outcomes.

4. Authenticity as Provenance

The first wave of public authenticity is now about origins rather than appearance. Most traditional exhibitions present the evidence to the public in familiar ways, such as archive photos, measured drawings, contemporary artists' works or physical remains of old buildings. Diffusion systems can integrate the above differences to produce a single style of photorealistic images. Visual Consistency is not a good index of historical accuracy. A reasonable explanation should state the source of the information, which parts were recorded directly, which parts were deduced, and what was added to fill in the missing links.

Established Documentation Rules Support This Way. ICOMOS guidance on recording monuments, groups of buildings and sites requires records to identify their creators, dates, cross-references, locations and information sources, and it regards recording as part of understanding and conservation [8]. UNESCO's Charter on the Protection of Digital Heritage also demands a reliable system, continuous access, and technical and legal guarantees for the authenticity of records and prevention of unauthorised alteration [9]. Therefore, it is proposed to save the prompts, model versions, input files, masks, editing steps, approvals and output dates as interpretive records rather than discarding them after a visually satisfactory image has been exported.

Display the origin for non-experts as well. The ICOMOS Charter for the Interpretation and Presentation of Cultural Heritage Sites lists sources of information, context, authenticity, inclusivity, sustainability and continuous evaluation [10]. A public interface can generate the necessary disclosures at various levels. A small label can be added to show that the image has been created or processed by AI; an expandable window can display the archive origin and disputed parts, and an overlay can distinguish between recorded, inferred, and speculative areas. Thus, it will not be too technical and can still be enjoyed by all people.

The London Charter for computer-based visualisation calls for intellectual transparency, documentation, sustainability, access and communication of evidential status [11]. In recent years, in addition to its application in three-dimensional data display, people have also begun to explore various uses for visualisation in learning and understanding. For an observer to recognise a diffusion-based image as such, the image needs to possess intelligence. Not all the proprietary parameters need to be published; however, sufficient information must be provided so that the conjectural image does not appear to be an archival record.

5. Cultural Cognition and the City

The second is cultural cognition, that is, the frames in which the public organises knowledge about the urban past. Generated Images can show alternatives more quickly. A curator can introduce several representative historical roof styles, show how the neighbourhood has changed over time, and have some older residents share their memories of the past. Thus, the generation function is to pose a problem. It offers a first-order organisation of the uncertainty and can identify defects in the evidence. If only one polished version is shown and not others, it will block the opportunity for exploration and present a single interpretation of the entire visual memory.

Urban heritage is relatively simple to organise because a city is a multi-level space rather than an organised arrangement of buildings. UNESCO's Historic Urban Landscape recommendation defines urban heritage as the interaction of physical form, setting, infrastructure, social and cultural practices, economic processes, perception, and intangible attributes [12]. A generated image of only the picturesque facades will thus restrict cultural cognition. Informal trade, labour and migration; religious differences, political conflicts and modern additions may be excluded, and a stylistically unified past that never existed in that form is created.

Training data and prompting further shape what is shown. A model will favour frequently repeated monuments, well-known architectural styles and famous scenes from films.

Prompt writers may use the general tag "ancient", "traditional" or "authentic" to refer to many eras and cultures. Not all of the outputs are therefore biased or incorrect due to these risks. Therefore, during the process of curatorial evaluation, one should also point out any omissions, anachronisms, stereotypes or unequal distribution of visibility, and not just determine whether a building appears authentic.

Changes in Interface Sequence also affect Cognition. If the generated image is presented first and then the evidence follows, the initial feeling of reality may be false. An evidence-first sequence can start with the surviving traces and records, introduce the historical problem, and then show one or more generated hypotheses. Comparison of views is more suitable for learning what is lacking than revelation. Next to the original visuals are the corresponding side-by-side comparisons and brief explanations of the modifications. The desired outcome is not universal distrust of all digital pictures, but rather graduated trust according to the presented evidence.

6. A Responsible Practice Framework

A feasible project should start with a narrow interpretive question, for example, how might a recorded street front have looked in a particular decade? Terms, place, period, audience and purpose of use should be specified before generation. Then, the project team will collect and organise materials such as surveys, photos, maps, planning documents, oral histories, material culture, relevant academic research, etc. Document the conflicts among the sources and do not choose the option that creates a more attractive image.

Next is the map of evidence. The attributes of the proposed visual elements listed above can be one of the following: directly observed, strongly inferred, weakly inferred or illustrative. Generation should only occur after this map exists. Input images should not be changed in the archive, and derivative files need unique identifiers. Log prompts and negative prompts, model and interface versions, random seeds (if applicable), masks, control images and human modifications. Such a record does not change the model output into historical facts; instead, it can enable the reproduction of interpretation reviews under practical conditions.

The human reviewers should have technical, historical, ethical and community knowledge. Conservationists can access the physical evidence, historians may find a time anomaly, curators will assess the continuity of the narrative, community members report any missing or incorrect parts, and technical staff check the artifacts and workflow integrity. Participation needs to be meaningful, not perfunctory. If the image is connected to living traditions, displacement, sacred sites or trauma and identifiable individuals, the relevant group shall have the right to object to the use of their likeness and to request that it not be circulated.

Add a notice at the contact point for publication. Use a simple expression for the label, such as "AI-generated interpretive hypothesis", instead of "digital restoration". Link to the sources in the display, point out the main uncertainties, mention who ordered and approved the output, and clearly differentiate generated content from archival materials. Watermarks are not very effective because they can be cropped, and metadata may be removed on social media. Redundant disclosure in the image, caption, metadata and project page is more robust.

Finally, based on the results of the project, conduct an evaluation. The group may ask the visitors whether they believe the picture was created by artificial intelligence, if there are any supporting materials or no evidence, why the other results are different, and whose story is being told. If these problems are not solved, the report will only list the authors' wishes for interpretation and fail to show an improvement in cultural cognition or trust. Therefore, the present system is a proposal for application rather than an empirically verified intervention model. Its utility will be introduced in the following visitor studies, comparative research and long-term observations of digital preservation.

7. Risks and Institutional Responsibilities

The most immediate risk is epistemic overreach: the transformation of possibility into perceived fact. Photorealistic manipulation can hide the absence of evidence, and after multiple circulations, the original disclaimer may be lost. The institution can add an "Afterlife" section, preserve the original source, and post an updated version of the data after detecting an error. Avoid marketing words that promise to bring the past "back to life" or show a lost city "exactly as it was", as these are outside the scope of probabilistic generation and lack full historical data.

Another is the risk of falsification. Generative systems can reproduce stereotypes, exclude marginalised groups, normalise idealised cleanliness, and add culturally inappropriate symbols. The problem is not addressed by increased visual quality. Determine what is and is not shown in the judgment, whether the sources are authentic, whether the image contradicts the main narrative of the heritage and community consensus, etc. If they still cannot agree, inform both parties of what happened and do not fabricate a story.

The third reason is sustainability. Generated assets are subject to changes in software, models, formats, licenses and computing environments. Institutions should store the source record and interpretive document even if the generation service cannot be retained. Determine whether repeated generation at high resolution is necessary for the educational goals and feasible given the resources. A small number of well-documented pictures can help the public learn better than a large number of unverified modifications.

Legal and organisational Review is also required. Historical pictures, building drawings, contemporary photographs and community contributions may have rights or conditions of use. Restrict access to personal data and sensitive locations. The model will not be responsible; a particular institution and its responsible staff will be designated to manage the interpretation of problems and changes. Procurement contracts should maintain access to project records and should not restrict the institution's disclosure of the system's role.

8. Conclusion

Diffusion-based text-to-image and image-to-image systems offer a focused way to study generative digital technology in urban cultural heritage. Their value is less in their ability to recreate the past and more in their capacity for rapid production of impressive visual hypotheses. With the expansion of scale, the focus of public authentication has gradually shifted from physical characteristics to other factors such as the origin and legitimacy of evidence, uncertainty, environment, institutional responsibility, etc. Although it provides other situations for cultural cognition, new problems

of simplification, exclusion and false closure also arise at the same time.

Therefore, responsible use should keep the difference between recording and creation, material existence and visual presentation, experiential meaning and documentary evidence, interpretation and verified facts. Evidence mapping, process documentation, layered disclosure, plural review, community participation and durable preservation can all be made visible. The generated pictures should prompt people to consider the problems in the city themselves.

A concept system and a workflow have been introduced in this paper; however, empirical evidence of the public effect is lacking. There is no original visitor information, and it does not mention that the generated images help with learning, authenticity, trust or participation. In the future, various labels, uncertainty displays and comparison interfaces can be explored to enhance comprehension in different cultures. In the absence of such evidence, the institution should present the expected benefits as possibilities, clearly indicate the limitations, and never present visual plausibility as historical evidence.

References

[1] ICOMOS. (1965). International charter for the conservation and restoration of monuments and sites (The Venice Charter 1964).

[2] ICOMOS. (1994). The Nara document on authenticity.

[3] MacCannell, D. (1973). Staged authenticity: Arrangements of social space in tourist settings. American Journal of Sociology, 79(3), 589–603. https://doi.org/10.1086/225585

[4] Wang, N. (1999). Rethinking authenticity in tourism experience. Annals of Tourism Research, 26(2), 349–370. https://doi.org/10.1016/S0160-7383(98)00103-0

[5] Bolter, J. D., & Grusin, R. (1999). Remediation: Understanding new media. MIT Press.

[6] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840–6851.

[7] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042

[8] ICOMOS. (1996). Principles for the recording of monuments, groups of buildings and sites.

[9] UNESCO. (2003). Charter on the preservation of digital heritage.

[10] ICOMOS. (2008). Charter for the interpretation and presentation of cultural heritage sites.

[11] Denard, H. (Ed.). (2009). The London charter for the computer-based visualisation of cultural heritage (Version 2.1).

[12] UNESCO. (2011). Recommendation on the historic urban landscape, including a glossary of definitions.