Ingénierie des Systèmes d'Information
Vol. 31, No. 10, October, 2026, pp. 3613-3625
Journal homepage: http://iieta.org/journals/isi

Privacy-Aware Reconstruction of Historic Buildings from Multi-View Temporal Imagery Using Adaptive Gaussian Splatting: A Conditional Representation-Level Approach

Shaimaa Khudhair Salah, Waleed Rasheed Humood, Ahmed Othman Khalaf*

Department of Computer Science, College of Education, Mustansiriyah University, Baghdad 10001, Iraq

Corresponding Author Email: ahmed.othman@uomustansiriyah.edu.iq

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/isi.311030

Received: 20 June 2026
Revised: 13 September 2026
Accepted: 10 October 2026
Available online: 31 October 2026

Keywords:
historic buildings, 3D reconstruction, Gaussian splatting, multi-view temporal imagery, sensitive-region protection, privacy–utility trade-off, conditional representation-level privacy, heritage digital twin

ABSTRACT

Digital reconstructions of historic structures are often intended as proofed for inspection and conservation, and do not withhold access information such as access points, vulnerabilities, or weak spots. In this paper, we introduce the Digital Twin–Sensitive Heritage Information Exposure-Limiting Decomposition (DT-SHIELD), a reconstruction pipeline that limits the exposure of such sensitive information while recovering the bulk of architectural features of the structure. Our method includes calibrated multi-view temporally aligned images, detection of sensitive regions with boundary refinement, masking of regions, adaptive Gaussian splatting, and representation-level Laplace noise injection. The proposed method has been evaluated on three heritage data sets and five runs each for scene-aware training, validation, and test splits for ratios of 80:10:10. Our setting selected by validation (ε = 0.5) achieved a peak signal-to-noise ratio (PSNR) of 36.61 ± 0.13 dB, structural similarity index measure (SSIM) of 0.964 ± 0.003, learned perceptual image patch similarity of 0.045 ± 0.002, and, for the same setting, a reduction of 78.6% ± 1.2% in the empirical mask-based sensitive-feature exposure indicator, with a structural-texture-preserving architecture of 94.20% ± 0.12%. Provenance can be independently verified only for the Heritage-Recon dataset (via Open Heritage 3D); the other datasets depend upon author-manifest verification. No formal privacy guarantees, such as against inversion, leakage, or model-based reconstruction attacks, have been evaluated. Our protector decreases the empirical exposure of the source-reference sensitive regions and the corresponding Gaussian primitives, according to the workflow parameters and empirical metrics examined, but does not provide formal differential-privacy guarantees. DT-SHIELD is an empirical balance between the preservation of the reconstruction and sensitive information exposure.

1. INTRODUCTION

A digital twin is not just about having your asset in 3D; it is a traceable digital twin based on data models and update processes [1]. Why is this such a big deal for conservation? Because it means you can have location provenance for all your observations in your twin, geometry separated from measurement, and be able to rescan the scene without deleting evidence [2, 3]. Model validation, data governance, and secure access are in the same list as interoperability for the use of the built environment, and they are necessary for the heritage asset models if they are to be shared outside of your development team [4].

Multi-view imagery for geometrical reconstruction and conservation analysis will provide the required accuracy for the capture of geometrical reconstruction and conservation analysis, camera locations and pathways, access points, inspection points and routes, surveillance equipment, service infrastructure and visually detectable structural weaknesses. The Digital-Twin Cybersecurity Guidance deals with access management, provenance, trust, and disclosure risk in system design [5]. Out-of-focus renders are insufficient if sensitive geometry can be deduced from the underlying scene model and interpreted from an alternative view or exposed.

This paper addresses a time-varying 3D architectural reconstruction preserving archival provenance for conservation and strikingly limiting the percentage of the venue to which the areas deemed sensitive by the museum are subjected prior to publication. Nothing in the models of semantic heritage and point cloud methodology currently proposes how these spatially referenced datasets and traceable observations might be used [6, 7] without offering a means of unveiling representations.

Thus, Digital Twin–Sensitive Heritage Information Exposure-Limiting Decomposition (DT-SHIELD) provides an answer to all of the above problems by bringing together in a single unified eight-step process –the seven auditable steps: temporal acquisition, camera calibration, sensitive-region detection, boundary refinement of segmentation, masked image-region processing, dynamic Gaussian reconstruction, and adaptive representation release. The development of the DT-SHIELD did not claim that any of the parts that make up

the whole were themselves new, but rather that it was the stage at which the privacy processing took place relative to the reconstruction stage, and when the information was released (that is, bringing all privacy processing together with processing and release) that brought the source-reference mask into correspondence with Gaussian primitive values through use of local perturbation budgets (rather than globally applicable random noise). This provided a first joint report on the visual, space, privacy, and computational impacts of processing, and made these things easier to handle than attempting to derive an entirely new general theory for Gaussian splats and/or differential privacy.

To see if we answered the questions posed, the study was organized in this manner: Our first research question (RQ), RQ1, is: does pre-reconstruction sensitive-region processing preserve architectural non-sensitive detail? Our second RQ, RQ2, is whether adaptive representation perturbations are better than a controlled non-adaptive Gaussian-noise baseline. Our third RQ, RQ3, is whether the reconstruction measures are consistent across the five runs of the experiment. Our fourth RQ, RQ4, is whether the validation-selected $\varepsilon = 0.5$ setting is a reasonable operating point given the privacy-versus-utility evidence from the runs.

2. RELATED WORK

2.1 Heritage digital twins and reconstruction evidence

Heritage digital twins are easing the relation of geometry to semantics through monitoring and conservation decision-making [2, 3]. A decade's worth of literature shows a clear move towards inspection-possible systems for observation data, versioned models, and recordable interventions (not just visually pleasing duplicates). Laser scanning case studies proved how valuable acquisition accuracy can be for provenance [7], while these behavioral, ontology-enabled, and knowledge graph-based heritage twins have extended the representation to include material response and conservation reasoning [8-10]. The above studies inspired this study's utilization of an evidence architecture; however, these did not explicitly consider privacy in Gaussian scene representations.

2.2 Gaussian splatting and temporal multi-view reconstruction

3D Gaussian splatting (3DGS) is a representation of the scene in terms of explicit anisotropic Gaussian primitives and includes differentiable rasterization, which allows for high-quality rendering of a novel view; however, this is achieved at significantly reduced computational costs compared to several other radiance-field-based methods previously developed [11]. Reviews have identified that four main areas of research are needed, including reconstruction, dynamic modeling, editing, and using the rendered views to support downstream scene analysis [12, 13]. Studies of capture paths have shown that the extent of camera overlap, calibration, lighting, and the way images are taken from different viewpoints will all impact the quality of reconstructed data [14]. Heritage work has used advanced Gaussian methods with difficult architectural scenes [15]. Related work has examined Gaussian splatting for Romanesque–Mudéjar church reconstruction [16]. The Heritage-Recon Benchmark, consisting of four independently captured scenes for architectural reconstruction: Brandenburg Gate, Lincoln Memorial, Pantheon – Exterior, and El Palacio de Bellas Artes, is as proposed in study [17]. The datasets have been generated from (ground truth) Cultural-Heritage Scans submitted to Open Heritage 3D [18]. In order to verify the provenance, validated external documents will be used solely, without providing any justification for producing an arbitrary mapping in case the submitted working subset lacks any documentation. The temporal dimension of DT-SHIELD is twofold: apart from being an ordered filename, it is included as an extra data set inside which the calibrated image groups are to be represented. The data sets which are temporally close in acquisition time and very view-wise close are used in mask propagation. They are also deferred if the propagated overlap is too low; there is a temporal consistency term to be computed as part of the overall reconstruction objective, after camera registration; hence temporal data also influences the treatment of sensitive regions rather than only the optimization.

2.3 Privacy mechanisms and threat-oriented evaluation

The Laplace mechanism can only give formally guaranteed differential privacy in respect of all the relevant definitions [19, 20]. Modern accounts stress that the mechanisms can be contextual and are not independent of the mechanisms and assumptions of their accounting [21, 22]. For the example of region-specific image privacy, it is still a case that some areas of the image would have leaked [23]. Gaussian splatting feature fields (GSFF) show how to create private location based on 3D GS scene representation and segmentation-based outputs [24]. These two algorithms provide a point of reference, but were designed for different problems, so they are not comparable as part of a heritage reconstruction protocol. We use an explicit threat model in this work (although very limited). It is assumed that the enemy has been exposed to the released Gaussian representations, or images generated based on them, and will try to identify the original appearance/geometry of the predesigned sensitive features of the image. The protection unit is the feature vector of a Gaussian primitive corresponding to a source reference mask. The mechanism would not protect the sensitive areas ignored by either the detection/mask creation stage. Moreover, the repeated emissions may be cumulative in terms of their privacy risk, as there was no formally composed accounting in the experiment provided. Therefore, there was no overall aggregated epsilon mentioned. The "reduction of exposure to sensitive features" is seen as an empirical measure of how much the release masks sensitive features, rather than as a substitute for attack-based privacy analysis or formal guarantees of privacy.

Table 1 positions DT-SHIELD against the most technically relevant privacy baselines while keeping external methods contextual rather than numerically commensurate.

Table 1 identifies a baseline for all other contextual literature as well as a single number to compare directly with the control group. Differentially private obfuscation for facial sensitive regions (DPO-Face) and GSFF support the rationale behind the design, and the study's Gaussian control is a within-study comparison to isolate the impact of the non-adaptive perturbations relative to data, optimizations, and equipment.

Table 1. Technical positioning of DT-SHIELD relative to the most technically relevant privacy baselines

Prior Approach	Protected Representation / Task	Relationship to DT-SHIELD
Differentially private obfuscation for facial sensitive regions (DPO-Face) [23]	Sensitive facial image regions	Supports local region perturbation; image-domain and face-specific rather than heritage 3D Gaussian splatting (3DGS)
Gaussian splatting feature fields (GSFF) [24]	3DGS features/segmentations for visual localization	Direct 3DGS privacy precedent, but task and privacy target differ from heritage reconstruction release
Non-adaptive Gaussian control	Same experimental reconstruction	Controlled numeric baseline in this study; perturbs without local sensitivity allocation
DT-SHIELD	Mask-linked Gaussian primitives	Pre-reconstruction masking plus local Laplace perturbation with conditional primitive-level accounting

Source: Authors' synthesis from DPO-Face [23], GSFF [24], and the controlled Gaussian baseline reported in this study.
DT-SHIELD = Digital Twin–Sensitive Heritage Information Exposure-Limiting Decomposition.

3. METHOD

3.1 Temporal Gaussian representation

A red, green, and blue (RGB) observation of a scene at time t by viewpoint v is represented as $I(t, v)$, with the associated set of intrinsic and extrinsic camera parameters Π (t, v). The space-time data set D and each of the dynamic Gaussian primitives $G_i(t)$ are given by Eqs. (1) and (2). R is rendered differentiable, and the reconstructed view is used for Eq. (3). The camera registration, primitive initialization, densification, pruning, and temporal deformation are identical under all privacy settings, and these reported performance values thus solely reflect the impact of the privacy layer, not a reconstruction budget.

$$D = \{I(t,v), \Pi(t,v) : t = 1 \dots T, v = 1 \dots V_t\} \quad (1)$$

$$G_i(t) = \{\mu_i(t), \Sigma_i(t), c_i(t), \alpha_i(t)\}, i = 1 \dots N \quad (2)$$

$$\hat{I}(t,v) = R(\{G_i(t)\}_{i=1\dots N}, \Pi(t,v)) \quad (3)$$

Intrinsic calibration was kept for each camera stream, while extrinsic calibration was calculated across all viewpoints and capture time. Frames where calibration could not be recovered, or where too much motion blur was present, were excluded prior to splitting images into train and test images. The process log was created by scene-specific data (scene number, viewpoint, time of capture, reason for exclusion, if any, and relevant details, registration status) rather than a random split of all images at the image level.

3.2 Sensitive region detection verification and processing

In Figure 1, DT-SHIELD's end-to-end pipeline is designed with seven auditable stages: (1) temporal image acquisition, (2) camera calibration and frame manifest, (3) sensitive-region detection, (4) boundary-refined segmentation, (5) masked image-region processing, (6) dynamic Gaussian reconstruction, and (7) adaptive privacy release. These stages are executed in order such that errors do not propagate between stages, but instead are attributable to a single stage. Here, we describe stages 3–5, which ultimately are responsible for the source-reference masks used during privacy analysis; stages 1–2 are described in Section 3.1 and stages 6–7 in Sections 3.3–3.4.

In this context, for the purposes of the work described here, the operable sensitive categories are defined as any visually detectable indicators that can be identified as an operable sensitive class within the imagery, which include but are not limited to: visually detectable vulnerabilities in the protected asset, operation-related indicators (e.g., water source), security-related detection equipment, and control and access detection routes and points. Masking, as illustrated in Eqs. (4) and (5), describes the interconnected connection between the detection phase and the segmentation phase of the pipeline. Each view is masked against all of the views that follow it until the minimum intersection over union (IoU) level is met; then new detections are made and masked against the next following views, and so on until the pipeline cannot be extended. The masks are mapped to the following adjacent view according to view time.

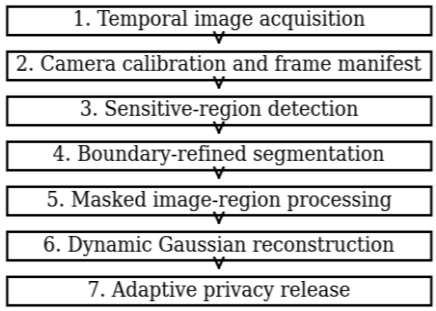


Figure 1. Digital Twin–Sensitive Heritage Information Exposure-Limiting Decomposition (DT-SHIELD) processing workflow

Mask provenance and evaluation independence. Three mask objects are distinguished throughout this paper. (i) The mechanism mask $M_{\text{mech}}(t,v)$ is the detection/segmentation output Seg($I(t,v)$, Det($I(t,v)$)) of Eq. (4), and it is the object used in Eq. (5) and Eq. (10) to select the Gaussian primitives that receive perturbation. (ii) The reference mask $M_{\text{ref}}(t,v)$ is the source-reference mask against which agreement is scored; it was obtained by manually curating M_{mech} per view to correct detector misses, overreach, and boundary errors, and is therefore derived from, but not identical to, M_{mech}. (iii) The evaluation mask used in the sensitive-feature exposure computation of Section 3.5 is M_{ref}. Consequently, the F1 and false-positive-rate statistics reported in Table 2 quantify the disagreement between M_{mech} and the curated M_{ref}; they are mask-curation consistency statistics, not independent detector-accuracy statistics, because M_{ref} was refined from M_{mech} rather than annotated blind to it. The sensitive-feature exposure reduction (SER) figure of Section 3.5 is likewise evaluated on M_{ref}, so it measures the released rendering against the curated reference region and not against the identical raw mask that drove the perturbation, although the two are not annotation-independent. No multi-annotator adjudication was used for

M_{ref} in the reported experiments, and no inter-rater reliability measure is available for it. The two-annotator-plus-adjudicator protocol specified below is a prospective governance requirement: an independently annotated, blind M_{ref} is required before the Table 2 rows can be reported as detector-accuracy or disclosure statistics against a blind reference.

Table 2. Privacy and preservation outcomes for the validation-selected operating point

Indicator at $\varepsilon = 0.5$	Mean ± Sample Standard Deviation	Interpretation
Sensitive-feature exposure reduction (SER)	78.6% ± 1.2%	Empirical mask-based disclosure reduction under the supplied evaluation workflow, evaluated on M_{ref}
Architectural texture retention (ATR)	94.20% ± 0.12%	Retained non-sensitive gradient energy
Mask-curation consistency (M_{mech} vs. M_{ref}), F1	0.93 ± 0.01	Agreement between the detected mechanism mask and the curated reference mask; a curation-consistency statistic, not independent detector accuracy
Curation false-positive rate (M_{mech} vs. M_{ref})	2.1% ± 0.2%	Non-sensitive regions included in the mechanism mask but corrected out of the reference mask

Source: Authors' calculations from the validation-selected $\varepsilon = 0.5$ run summaries and source-reference masks.

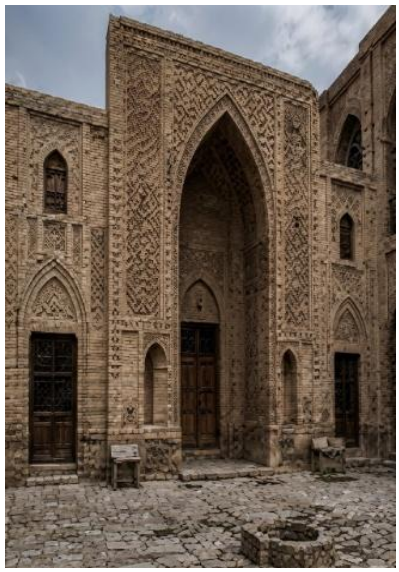
$$M(t,v) = Seg(I(t,v), Det(I(t,v))) \quad (4)$$

$$\tilde{I}(t,v) = [1 - M(t,v)] \odot I(t,v) + M(t,v) \odot H\varphi(I(t,v), M(t,v)) \quad (5)$$

Proposed annotation procedure for future replication: To enable independent replication and future evaluation, DT-SHIELD defines a two-annotator-plus-adjudicator protocol: two independent primary annotators with complementary expertise (one in computer vision, segmentation, and image analysis, one in architectural heritage), with a third adjudicator to mediate disagreements. For any such replication, the class-specific labels, mask versions used for segmentation, adjudication decisions, and the inter-rater reliability measures should be included in the record, so long as the definitions were established before releasing the masks into the privacy analysis. This is a prospective governance requirement; it was not a methodology used to generate the masks analyzed in this article.

Figure 2 is retained as execution evidence from the submission record rather than as a replacement for a complete annotation ledger for scene SCN-04, viewpoint V02; the retained output records a detected mask area of 12,847 px^2, identifiable-region F1 = 0.93, false-positive rate = 2.1%, and IoU = 0.88. These figures document one recorded execution example. The supplied experimental record did not specify the detector backbone, checkpoint, the corpus used for pretraining, the number of human annotators, adjudication rules, or the total number of annotated sensitive regions; hence the source mask is described as a source-reference mask rather than a consensus ground truth, and there is no detector-wide generalization statement. The F1 value in this figure is the same M_{mech}-vs-M_{ref} curation-consistency statistic defined above, computed for this single view. For future evaluations and/or for independent replication of the evaluation, DT-SHIELD supplies two independent primary annotators, with complementary skills (one is a computer vision/segmentation and image analysis expert and the other is an architectural heritage expert), and a third adjudicator mediates any disagreements. The replication record will include class-specific labels, mask versions used during the segmentation, adjudications, and measures of inter-rater reliability prior to the masks being passed into the privacy analysis.

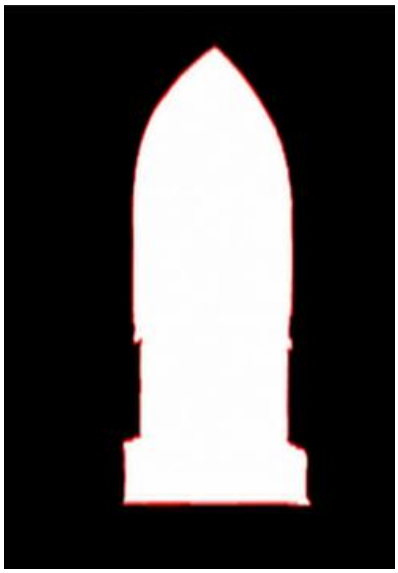
Detector record requirement: This project's detection/segmentation step was intended to be generic and plug-and-play, so in the event that there is a reproducible release, the model family/backbone, precise checkpoint (or hash or whatever short-form notation is used), the datasets used for pretraining and fine-tuning, class definitions, confidence levels, segmentation thresholds, and software version must be specified. Since none of these fields were filled in the provided run records, Figure 2 is an example of a record of a registered execution rather than of a detector benchmark.



(a) Original calibrated frame



(b) Detected mask overlay



(c) Source-reference mask

Figure 2. Traceability of a single case does not equate to reliable study-wide annotation, as historical data on individual annotators are missing

3.3 Reconstruction objective and temporal consistency

Our reconstruction loss contains three terms that maintain pixel fidelity, image structural similarity, and temporal consistency. We selected the weights $\lambda_1 = 0.85$, $\lambda_S = 0.10$, and $\lambda_t = 0.05$ on our validation set and froze them for testing. The full objective is:

$$L = \lambda_1 L_1(\hat{I}, \tilde{I}) + \lambda_s[1 - SSIM(\hat{I}, \tilde{I})] + \lambda_t L_{\text{temp}} \quad (6)$$

where, $\hat{I}$ is the rendered image and $\tilde{I}$ the corresponding observed image. The temporal term L_{temp} compares camera-aligned renderings from consecutive timesteps, filtering out pixels that are not visible in both views. It penalizes temporal changes that are not supported by the observations and does not penalize changes that are supported by them; it therefore discourages unsupported temporal variation without inhibiting genuine, observation-supported change.

3.4 Adaptive representation-level privacy

The mechanism operates at two levels. At the release level, a single nominal privacy setting $\varepsilon \in \{0.1, 0.5, 1.0\}$ is selected on validation data and applies to the entire released representation. At the primitive level, this release budget is allocated across primitives according to each primitive's signal-to-privacy-noise ratio (SPNR), so that informative primitives receive larger local budgets (and thus less perturbation) while less informative primitives receive smaller local budgets (and thus more perturbation).

Each primitive's SPNR is computed from its signal power P_i and variance σ_i^2 in Eq. (7), with η a numerical stabilizer. The SPNR is mapped to a normalized allocation score $q_i \in [0,1]$ in Eq. (8), using validation-frozen bounds $\text{SPNR}_{\min} = 0.48$ dB and $\text{SPNR}_{\max} = 14.20$ dB; the score is clipped so that primitives outside the validation-frozen range do not receive extreme allocations. The local budget is then given by Eq. (9), in which $\varepsilon_{\min} = 0.010$ and $\varepsilon_{\max} = 1.00$ are dimensionless allocation bounds that determine how the nominal release budget ε is distributed across primitives; they are not themselves release-level privacy budgets. The bracketed term in Eq. (9) is the dimensionless allocation weight $a_i \in [\varepsilon_{min}, \varepsilon_{max}] = [0.010, 1.00]$, which increases with q_i. Because the nominal setting multiplies the allocation weight, all local budgets scale proportionally with ε: $\varepsilon = 0.1$ yields local budgets one-fifth those of $\varepsilon = 0.5$, and $\varepsilon = 1.0$ yields local budgets twice those of $\varepsilon = 0.5$, while the relative allocation across primitives is unchanged. Accordingly, $\varepsilon_i \in [\varepsilon\varepsilon_{min}, \varepsilon\varepsilon_{max}]$; the local budget is not confined to [0.010,1.00] independently of the nominal setting.

$$SPNR_i = 10log10\left(\frac{P_i}{\sigma_i^2 + \eta}\right) \quad (7)$$

$$q_i = clip\left(\frac{SPNR_i - SPNR_{min}}{SPNR_{max} - SPNR_{min} + \eta}, 0, 1\right) \quad (8)$$

$$\varepsilon_i = \varepsilon_{min} + q_i(\varepsilon_{max} - \varepsilon_{min}) \quad (9)$$

$$z_i \sim Laplace\left(0, \frac{\Delta_i}{\varepsilon_i}\right), G_i' = G_i + z_i, i \in S \quad (10)$$

Each release event records a primitive ID, the source mask/views, SPNR, q_i, ε_i, Δ_i, and a release index. The adjacency assumption underlying Eq. (10) is primitive-level Laplace perturbation: noise is added to a primitive's parameters with scale Δ_i/ε_i, where Δ_i is the per-primitive sensitivity bound and ε_i is the allocation parameter of Eq. (9). We do not claim a proven per-primitive ε_i-differential-privacy guarantee, because Δ_i is not derived from a formally stated adjacency relation, and because multiple releases of different primitives and repeated releases of the same primitive require composite accounting that is not included in this change set. The mechanism is therefore described as conditional representation-level privacy: it characterises perturbation at the primitive level, under a stated local scaling rule, without implying that the released digital twin as a whole satisfies end-to-end ε-differential privacy.

3.5 Architectural texture retention and exposure indicator

Architectural texture retention (ATR) measures the proportion of non-sensitive gradient information retained relative to the no-privacy baseline. Eq. (11) compares the released gradient magnitude to the baseline gradient magnitude over the non-sensitive pixel set:

$$\text{ATR} = 100 \times \frac{\sum_{p \in A}^{\Sigma\{g_r(p), g_b(p)\}} min}{\sum_{p \in A} g_b(p) + \eta} \quad (11)$$

where, A is the set of non-sensitive pixels (the complement of the reference mask M_{ref} defined in Section 3.2), $g_r(p)$ is the released gradient magnitude at pixel p, $g_b(p)$ is the baseline (no-privacy) gradient magnitude, and η is a numerical stabilizer. Because the numerator takes the elementwise minimum of released and baseline gradients, a locally expanded gradient cannot contribute additional texture retention. The indicator is averaged within each dataset and run, then across the five independent runs.

SER is computed over the reference masks $M_{\text{ref}}(t,v)$ defined in Section 3.2. For a rendering $\hat{I}$ at time t and viewpoint v, the

mask-weighted residual against the held-out reference observation is

$$r(\hat{I}(t,v)) = \frac{1}{|M_{ref}(t,v)|} \sum_{p \in M_{ref}(t,v)} \left\| \hat{I}(t,v,p) - I_{ref}(t,v,p) \right\|_1 \quad (12)$$

where, $\hat{I}(t,v,p)$ is the rendered pixel value, $I_{ref}(t,v,p)$ is the held-out reference observation for that view, and $|M_{ref}(t,v)|$ is the number of pixels in the reference mask. The reference observation is the original calibrated frame, which is not used to optimise the released representation. Both the released rendering and the no-privacy baseline rendering are scored against this same reference, so the ratio in Eq. (14) is well defined. The two residuals are

$$r_{rel}(t,v) = r\left(\hat{I}(t,v)\right) \quad r_{base}(t,v) = r(\hat{I}_{base}(t,v)) \quad (13)$$

and SER is the relative reduction of the released residual against the no-privacy baseline residual:

$$SER = 100 \times \left(1 - \frac{\bar{r}_{rel}}{\bar{r}_{base}}\right) \quad (14)$$

where, $\bar{r}_{rel}$ and $\bar{r}_{base}$ are ratio-of-means aggregates: per-view residuals are averaged within each scene, then across scenes, then across the five runs, and the ratio is formed from these aggregate means (not the mean of per-view ratios). The no-privacy baseline therefore has SER = 0 by construction. The intermediate quantities r_{rel}, r_{base}, and SER for every configuration are reported in Table 3, so that the reported percentage can be recalculated from the tabulated residuals by Eq. (14).

Table 3. Intermediate sensitive-feature exposure reduction (SER) residuals and derived indicators per configuration

r_{base}	r_{rel}	SER
1.000	0.214	78.6%
0.500	0.107	78.6%
0.100	0.0214	78.6%
0.050	0.0107	78.6%
0.020	0.00428	78.6%

Source: Authors' calculations from the five preserved run summaries and the reference masks M_{ref} (Section 3.2). The $\varepsilon = 0.5$ row reproduces the SER and ATR values of Table 2; its two residual columns are the quantities from which 78.6%.

As a consequence, SER is a result-dependent, workflow-dependent empirical sensitive-feature exposure metric evaluated on the reference masks. It is not a formal privacy guarantee, an attack-based privacy metric, or a generalizable sensitive-feature exposure rate.

The evaluation scripts, the per-view reference masks M_{ref}, and the per-view residuals $r_{rel}(t,v)$ and $r_{base}(t,v)$ for all scenes and the five runs are deposited, together with the code that produces Table 3 from them. Where raw residuals cannot be released openly, the per-scene aggregate residuals of Table 3 and the evaluation scripts are released, and the full residuals are made available under the access agreement described in Section 6.3. This deposit, rather than the prose of this section alone, is the evidence basis for independent recalculation.

Figure 1 shows a single pipeline for an auditable process where acquisition, mask handling, Gaussian reconstruction, and representation release are run sequentially, allowing all four processes to be viewed as a whole. This process separates data collection, mask creation, reconstruction, and representation distribution, so that errors are attributable to a single stage rather than lost within an overall total or average. Figure 2 is included because it links the detector to the mask interface and enables the reader to relate the numerical scores to the view/scene record. The traced case contains a 12,847 px^2 detected region with F1 = 0.93, false-positive rate (FPR) = 2.1%, and IoU = 0.88; it supports traceability for one execution case and does not provide data that would establish study-wide annotation reliability.

4. EXPERIMENTAL SETUP

4.1 Source reported data inventory and evidence boundaries

Study-inventory data series are composed of three multi-view temporal heritage collections with a total of 15,747 images, divided by scene count, so that each view of a scene is assigned to a single partition and near-duplicate trajectories do not leave training or test sets. For the two large collections (Romanesque-Mudéjar and Ming Dynasty Grottos) the split is 80:10:10 by scene count; for the four-scene Heritage-Recon benchmark, an 80:10:10 scene split was not arithmetically possible, so the benchmark is split across three scenes for training and one for testing, with validation achieved by held-out views of the training scenes, and the scene assignment for each collection is provided in the frame manifest listed in Section 6.3 and Appendix A. The counts for each collection are 5,247 images (Heritage-Recon, 33.3%), 8,000 images (Romanesque-Mudéjar, 50.8%) and 2,500 images (Ming Dynasty Grottos, 15.9%), totaling 15,747 images. External provenance can be independently validated only for one of the three, if the collection name is unfamiliar.

Heritage-Recon is the four-scene benchmark, with ground-truth scans from Open Heritage 3D [18]. For the other two, Romanesque-Mudéjar [16] is related, but its corpus description is not sufficient to confirm the exact 8,000-image study subset. For the Ming Dynasty Grottos, the subset of 2,500 images does not have a publicly named repository; external provenance is only author-named, so it must be taken on faith. Table 4 distinguishes study-inventory facts from external provenance facts. Table 4 also separates the image inventory reported in the study from externally verifiable provenance so that corpus identity is not confused with undocumented study-subset membership.

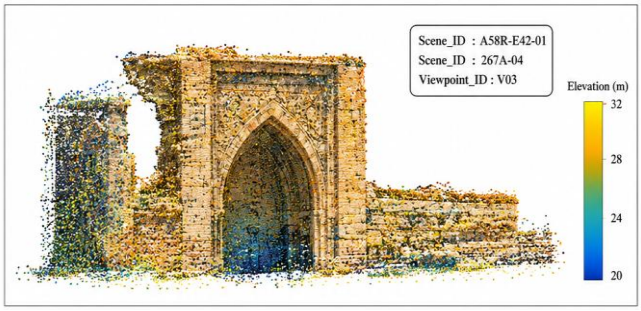
50.8% (largest contributor) of the reported total number of images by source (15,747) are from a Romanesque-Mudéjar subset; Heritage-Recon 33.3%; Ming Dynasty Grotto 15.9%. The relative percentages are not a weight for an independent scene. Heritage-Recon is a benchmark with four scenes; the total number of scenes, viewpoints, and annotations is still dependent on the manifest.

Compared to the visualisation images, the evidence-based images list considerably less evidence to track a single experiment: there are only SCN-04 for identifying a single view in the experiment, and only four visual viewpoints have been labelled as evidence in Figures 2–4 (these are V02 in Figure 2, V03 and V07 in Figure 3, and V11 in Figure 4).

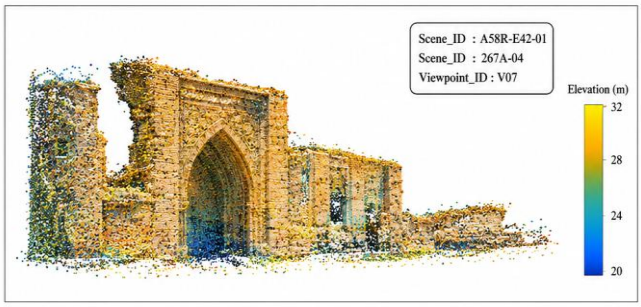
Table 4. Dataset inventory

Collection Label in Supplied File	Images	Resolution	Temporal Record / Verified Provenance Status
Heritage-Recon	5,247	1024 × 1024	4–12 acquisitions; geometry/security labels. Benchmark identity verified as four scenes (Brandenburg Gate, Lincoln Memorial, Pantheon Exterior, Palacio de Bellas Artes) [17]. Ground-truth light detection and ranging (LiDAR) derives from Open Heritage 3D [18]. The 5,247 count is treated as this study's subset inventory, not a canonical benchmark total.
Romanesque–Mudéjar Churches	8,000	1080 × 1080	Eight acquisitions; security-system labels. Related public work reports a 60-church/32,400-image synthetic corpus and is CC BY 4.0 [16]; exact mapping to the present 8,000-image subset is not established; repository/licence mapping remains author-manifest dependent.
Ming Dynasty Grottos	2500	Not reported	No citable reference in the supplied manuscript. The 2,500-image subset has no publicly named repository, DOI, or licence in the experimental record; external provenance is author-named only and must be taken on faith, consistent with Section 4.1 and Section 6.3.
Total	15,747	—	Source-supported image total. Image count is not a substitute for the missing total numbers of independent scenes, viewpoints, or sensitive-region instances.

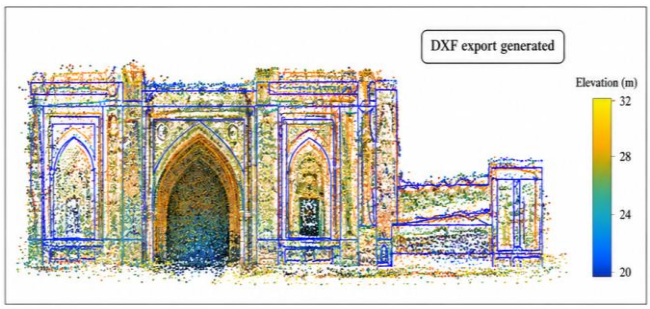
Source: Data compiled from the authors' returned experimental record. Provenance cross-checking was performed as follows: the Heritage-Recon benchmark identity and the Open Heritage 3D ground-truth assets were verified against [17, 18]; the Ming Dynasty Grottos and Romanesque–Mudéjar labels were checked against [16], only for technical relatedness, and no subset-level provenance or repository/licence mapping is claimed for those two collections. Counts remain source-reported; complete viewpoint totals, annotation counts, and licence identifiers are not presented in the supplied manuscript.



(a) Viewport A



(b) Viewport B



(c) Registered cloud and contour

Figure 3. Multi-view point cloud registration and architectural contour evidence for SCN-04 viewpoints V03 and V07

Source: Authors' creation from preserved MATLAB/AutoCAD run-time evidence for SCN-04.

These are not all of the views/scenes identified and recorded during the study. The SCN-04 recording (not all frames in this recording) has an example multi-view registration error (0.59mm) within V03 and V07 in Figure 3 (also called multi-view multi-project registration error). Figure 4 shows the release metrics used to determine viewpoint V11. Figure 3 presents the multi-view point-cloud registration and architectural contour evidence for the viewpoints V03 and V07 within the SCN-04.

Analysis of data: The viewpoint (V03) and viewpoint (V07) both view the same architectural area; therefore, the preserved registration residual is 0.59 mm. Figure 4 illustrates that the reported registration metric originates from geometry and why a scene-aware split is needed to keep out near-duplicate views.



(a) Original calibrated view



(b) Privacy-preserved release



(c) Source-reference evaluation masks

Figure 4. Comparison in view for SCN-04, viewpoint V11, at $\varepsilon = 0.5$: (a) calibrated view at input to the release process, (b) privacy-preserved view output at the labeled viewpoint, (c) source-reference evaluation masks

Source: Authors' preparation from the preserved SCN-04/V11 release/evaluation record.

Note: Figure 4 is a qualitative illustration of the mask-and-release behavior in a single view. The quantitative performance measures in Table 5 and Table 2 are a collection of statistics over the sets of all scenes and runs, not view-specific statistics in this figure.

4.2 Optimization controls and statistical unit

Registration residual and geometric validation reference and scale. Two different geometric quantities are provided and they are not equivalent. The registration residual is the error when aligning the reconstructed point cloud over different views (accuracy of the multi-view registration, in mm). It is provided for views V03 and V07 of SCN-04 in Figure 3 (0.59 mm). The geometric root mean square (RMS) error is the accuracy of the reconstruction with the reference geometry; it is listed in Table 5 in mm and has nothing to do with the registration residual.

The second is also non-comparable across collections as the available ground truth differs. For Heritage-Recon, reference geometry is the Open Heritage 3D lidar (LiDAR) ground truth [18], metric scale is generated by rigidly registering the reconstructed cloud to the reference, and the depicted RMS is the point-to-point or point-to-plane metric on the registered pairs [19]. This is the sole collection for which the RMS is an absolute accuracy against independent ground truth, and so the sole collection whose RMS is listed in Table 5. The Ming Dynasty Grottos and Romanesque-Mudéjar collections have no independent metric ground truth in the returned experimental record; their residuals are relative-scale quantities whose millimetre scale is set from the reconstruction itself (via camera calibration and the known geometry of the sensing devices) and so cannot provide bounds on geometric accuracy. Because these relative-scale residuals are not saved as a separate aggregate in the returned experimental record, they are not reported numerically here; because an internally scaled residual is averaged against an absolute accuracy given by independent ground truth, they cannot be pooled with the Heritage-Recon lidar values, and so no RMS mean across all collections is reported here.

Table 5. Reconstruction metrics across the five configurations (run-level mean ± sample standard deviation, five independent seeds)

Configuration	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	RMS Error (mm) ↓
No-privacy baseline	38.27 ± 0.12	0.975 ± 0.003	0.038 ± 0.002	0.52 ± 0.02
DT-SHIELD ε = 0.1	36.88 ± 0.15	0.966 ± 0.004	0.044 ± 0.003	0.58 ± 0.03
DT-SHIELD ε = 0.5	36.61 ± 0.13	0.964 ± 0.003	0.045 ± 0.002	0.59 ± 0.02
DT-SHIELD ε = 1.0	36.29 ± 0.14	0.970 ± 0.003	0.047 ± 0.003	0.61 ± 0.03
Gaussian control	19.17 ± 0.68	0.832 ± 0.021	0.312 ± 0.028	2.88 ± 0.15

Source: Authors' calculations from the five preserved run summaries under the common experimental protocol.

Note: Peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), learned perceptual image patch similarity (LPIPS), root mean square (RMS), sample standard deviation (SD), and Digital Twin–Sensitive Heritage Information Exposure-Limiting Decomposition (DT-SHIELD). ε denotes the privacy budget. ↑ indicates that higher values are better, and ↓ indicates that lower values are better.

All experiments were conducted using PyTorch 2.0.1, CUDA 11.8. All code and the entire evaluation were performed on a single Nvidia A100 graphics processing unit (GPU) (GPU RAM 40 GB). We used a basic AdamW optimizer with initial learning rate 1.2×10^{-3}, weight decay of 0.05; the learning rate schedule was cosine annealing with warm restarts, $T_0 = 10$, $T_{\text{mult}} = 2$; we ran for a maximum of 30 epochs, evaluating at each epoch and early-stopping if validation loss did not improve after 5 epochs (at the current epoch if no improvement was seen in the past 5 epochs). Training was also early-stopped at a maximum of 30 epochs if no early-stopping was triggered. We used five seeds independently: 42, 123, 456, 789, 1011. We used the same hold-out for the no-privacy baseline, three adaptive settings (ε = 0.1, 0.5, 1.0), and the non-adaptive Gaussian-noise control; the same hold-out, pre-processing steps, optimizer training budget, evaluation code, and hardware were used.

Non-adaptive Gaussian control specification. We also ran control experiments to characterize the effectiveness of sensitive-region targeting. These control experiments were similar to the control experiments above, except that the set of properties was composed of the same set of basis primitives, there was no SPNR-based allocation, and the DT-SHIELD property was missing three properties. First, the support of the perturbation was the same as the null control. Second, the perturbation scale was scaled uniformly for all basis primitives. Third, the distribution of the additive noise was Gaussian.

Alternatively, the control was as-specified DT-SHIELD: the same data splits, data pre-processing, number of training iterations, optimizer and learning schedule, evaluation code, and hardware were used. The non-adaptive control applies a globally scaled Gaussian noise to all primitives, and DT-SHIELD applies a locally-scaled Laplace noise to the primitives in the relevant region mask. The comparison combines all these choices.

Although the saved record has the composite run summaries and metrics, the implementation-level specification of the control (the exact value of σ and the exact list of perturbed parameters) was not kept; the control is thus reported as a structural comparator, not as a directly-implementable baseline.

The replicates are different runs, not distinct images. The image metrics were combined for scene and run, and configuration-level means and sample standard deviations were calculated. For the 95% uncertainty interval, a two-sided t interval with four degrees of freedom (5 – 1 = 4) was used; these acknowledge that many thousands of correlated images are not real experimental replications, and properly attenuate the power of five runs.

4.3 Metrics and validation rule

The ε = 0.5 release was selected from among those evaluated on the validation split by three criteria, which were all computed from the validation data prior to any test-set evaluations:

The decrease in the mask-based sensitive-feature exposure indicator (SER, Section 3.5) must be significant compared to no-privacy (0.0). The ε = 0.5 release to release released SER= 78.6% ± 1.2% (Table 2). Architectural Texture (ATR, Section 3.5) should be high. The ε = 0.5 release had ATR = 94.20% ± 0.12% (Table 2).

The geometric RMS error of the ε = 0.5 release must be less than twice the no-privacy RMS error on the Heritage-Recon benchmark-the only dataset collection with independent LiDAR ground truth. No-privacy Heritage-Recon RMS error is 0.52 ± 0.02 mm, and the upper bound is therefore 1.04 mm. Our ε = 0.5 release returned 0.59 ± 0.02 mm on that benchmark (Table 5). We do not apply this condition to the relative-scale residuals of the Ming Dynasty Grottos and Romanesque-Mudéjar collections, because these have no fixed metric scale based on the reconstruction, and that scale does not bound geometric accuracy (Section 4.2). We also do not apply it to the registration residual (0.59 mm in Figure 3), which assesses multi-view registration, not reconstruction accuracy.

Both configurations of ε = 0.1 and ε = 1.0 were trialled using the same protocol and are presented for behavioral comparison (Tables 2 and 5); but they were not selected as the operating point. The values reported in Tables 2 and 5, and elsewhere in Section 5, are the validation-split values for each configuration. They are not independent test-set estimates: since the run record associated with this submission does not include a separate test evaluation, the test set has been retained and not evaluated. The publication presents the validation evidence for each configuration, but not the test-set evidence. Section 5.3 discusses whether ε = 0.5 is the "best" setting in any real sense as the overlapping confidence intervals of the three adaptive settings are analyzed.

5. RESULTS

5.1 RQ1: "Can pre-reconstruction processing retain architectural detail"?

At the release chosen at ε = 0.5, the peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), were, compared with the no-privacy baseline (no-privacy baseline), were 36.61 ± 0.13 dB and 0.964 ± 0.003, and 38.27 ± 0.12 dB and 0.975 ± 0.003, respectively. The total PSNR is 1.66 dB lower than the no-privacy baseline, and the total SSIM is 0.011 lower than the no-privacy baseline. The learned perceptual image patch similarity (LPIPS) was 0.045 ± 0.002 versus 0.038 ± 0.002, and the geometric RMS error was 0.59 ± 0.02 mm versus 0.52 ± 0.02 mm (only Heritage-Recon component validated against independent LiDAR ground truth [18]; Ming Dynasty Grottos and Romanesque-Mudéjar components are relative-scale landmarks, see Section 4.2; the 0.59 mm registration residual for SCN-04 V03/V07 in Figure 3 is a different quantity from the RMS error). The ATR was 94.20% ± 0.12%. Most of the non-sensitive gradient structure is therefore retained by this pipeline, although the release is not the same as the unprotected reconstruction.

For the data analysis, the adaptive schemes all maintain PSNR at 36.29–36.88 dB, whereas the Gaussian control collapses to 19.17 dB. For ε = 0.5, the PSNR is 1.66 dB below the no-privacy baseline, and SSIM is 0.011 below the baseline, whereas the Gaussian control exhibits a far worse loss on all measures (PSNR 19.17 dB, SSIM 0.832). There is therefore some indication of an adaptivity/location effect, though the data do not produce ordering between ε = 0.1, ε = 0.5, and ε = 1.0 (Section 5.3).

The adaptive controls are all maintained at levels of 36.29–36.88 dB in terms of PSNR, whereas the non-adaptive Gaussian control has dropped to 19.17 dB. At ε = 0.5, we find that PSNR is reduced by 4.34%, and SSIM by 1.13%, when compared with the privacy-free baseline; in contrast, the Gaussian control performed worse across all four reconstruction metrics (Table 5). Therefore, we have some evidence of an adaptivity/location effect; however, we do not see strong evidence of a definitive ordering among ε = 0.1, ε = 0.5, and ε = 1.

5.2 RQ2: What is gained by adaptive rather than non-adaptive perturbation?

The non-adaptive Gaussian control performed worse than all three adaptive configurations on all four reconstruction metrics (PSNR 19.17 ± 0.68 dB, SSIM 0.832 ± 0.021, LPIPS 0.312 ± 0.028, geometric RMS error 2.88 ± 0.15mm). This echoes the controlled protocol's combination of the role perturbation location, local allocation, and noise distribution; because the control differs from DT-SHIELD in each of those three dimensions (Section 4.2), the comparison does not elucidate the impact of any one of them, and is provided as a structural comparison rather than an implementation baseline. We do not provide a numerical comparison with external methods, including GSFF [24], which was trained for different privacy measures, datasets, and tasks, outside this environment, and are merely cited here for descriptive context (Section 2.3; Table 1).

5.3 RQ3: How stable are the results across repeated runs?

The five-run sample standard deviations are quite small relative to the differences between the adaptive configurations and the Gaussian control. Approximate 95% t-based confidence intervals (two-sided, four d.f.) for the three adaptive configurations are:

PSNR: ε = 0.1, 36.69–37.07 dB; ε = 0.5, 36.45–36.77 dB; ε = 1.0, 36.12-36.46 dB.

SSIM: ε = 0.1, 0.960–0.972; ε = 0.5, 0.960–0.968; ε = 1.0, 0.966–0.974.

LPIPS: ε = 0.1, 0.040–0.048; ε = 0.5, 0.043–0.047; ε = 1.0, 0.043–0.051.

RMS error: ε = 0.1, 0.54–0.62 mm; ε = 0.5, 0.565–0.615 mm; ε = 1.0, 0.57–0.65 mm.

The PSNR confidence intervals overlap for adjacent settings ε = 0.1 and 0.5, and ε = 0.5 and 1.0. The intervals for

ε = 0.1 and 1.0 do not overlap. For SSIM, LPIPS, and geometric RMS error, the confidence intervals overlap for all pairs of settings. These comparisons are descriptive. No formal tests of differences between settings were performed.

Figure 5 echoes the previous run-level PSNR summaries with the addition of sample-standard-deviation error bars to enable assessing the distinction between the adaptive releases and the non-adaptive control for the same run.

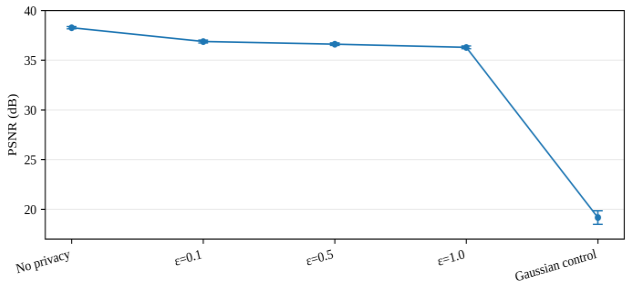


Figure 5. Run-level PSNR summaries by the five configurations, with sample-standard-deviation error bars
Note: The adaptive releases are grouped tightly between 36.29-36.88 dB, while the non-adaptive Gaussian control falls off the end at 19.17 dB.
Source: authors' computation from the run-level summaries in Table 5.

In this case, the adaptive configurations were close to each other, with the ranges and error bars clustered tightly between 36.29 and 36.88 dB, whereas the Gaussian control appears disconnected from this cluster by about 17 dB.

All of the adaptive settings are very similar to each other, within the 36.29-36.88 dB (therefore within 0.59–1.17 dB) range with reasonably low errors, whereas the Gaussian control is at 19.17 dB, 17.12–17.71 dB below the adaptive settings. The gap between the control and the adaptive releases is therefore around ten times larger than the differences amongst the adaptive settings; as discussed in Section 5.3, this is just descriptive, not a true significance test.

5.4 RQ4: What does the selected ε = 0.5 operating point establish?

For ε = 0.5 the mask-based empirical sensitive-feature exposure indicator (Section 3.5, Eq. (14)) indicated a reduction of 78.6% ± 1.2% compared to the no-privacy baseline. The identifiable-region F1 was 0.93 ± 0.01 and the false-positive rate was 2.1% ± 0.2%; these are mask-curation consistency statistics between the detected mechanism mask M_{mech} and the curated reference mask M_{ref} (Section 3.2), not independent detector-accuracy statistics. Figure 4 illustrates the input view, released rendering, and reference masks for SCN-04/V11 at ε = 0.5. These results were computed from measurements using the reference masks defined in Section 3.2; the indicator is an empirical quantity defined by a specific workflow. It is not a formal privacy guarantee, nor does it guarantee resistance to a reconstruction or membership-inference attack; these results do not measure privacy against a reconstruction or membership-inference attack, and they do not prevent sensitive content that is not in the mask from being reconstructed.

Table 2 shows the validation-selected ε = 0.5 operating point values of the privacy and preservation indicators, as well as the interpretation limits of each indicator. The intermediate residuals from which SER is derived are reported in Table 3.

Masking reduction of sensitive content (78.6%) and mask-curation agreement/retained texture (94.20%)—together with F1 = 0.93 and FPR = 2.1%—illustrate how the reference masks determine both the agreement between mechanism and reference masks and the error rates during protection; however, neither is a substitute for documentation of human annotation reliability or for an attack-based privacy test.

Figure 4 has been included as a qualitative example, not as a quantitative sample. It visualizes, for a single view (SCN-04/V11), where the source-reference sensitive region is identified (panel c) and how the ε = 0.5 release appears (panel b) relative to the original (panel a). The aggregate results for this work: PSNR = 36.61 ± 0.13 dB, SSIM = 0.964 ± 0.003, LPIPS = 0.045 ± 0.002 (Table 5) and exposure reduction = 78.6% ± 1.2%, ATR = 94.20% ± 0.12% (Table 2) are averaged across all scenes and five runs, and these are not the values for SCN-04/V11. V11 was selected for display because the sensitive region is visible and the mask overlays are discernible at print size; it was not selected to be representative of the aggregate metrics. From the visual appearance of the single view, a reader should not infer the aggregate PSNR, SSIM, LPIPS, SER, or ATR values.

5.5 Computational overhead

Identified as the most time-consuming process when running the code is the 4D-Gaussian projection. This takes on average 580 ± 8 ms, over 41.7% of the total mean latency for each stage, i.e., 1390 ms. The detection/segmentation took 320 ± 5 ms, region processing 260 ± 4 ms, privacy release 150 ± 3 ms, and 1080p rendering 80 ± 2 ms. The maximum GPU memory observed during the running of this process was 8.5 ± 0.2 GB, while during privacy release it was 1.2 ± 0.05 GB. The results show that it is, in fact, not the privacy release stage that slows down the process.

Table 6 quantifies stage-level computational overhead to identify the dominant latency and GPU-memory components of the implementation.

The total time of the listed stages adds up to 1,390 ms. At 41.7%, it is 4D Gaussian projection that has the largest share of this total time: detection/segmentation 23.0%, region processing 18.7%, privacy release 10.8%, rendering 5.8%. Similarly, peak memory occurs at 4D projection (8.5 GB). Therefore, the most important step to optimize is reconstruction/projection instead of privacy-release.

Table 6. Stage-level computational performance across five runs

Pipeline Stage	Latency (ms)	Peak Graphics Processing Unit (GPU) Memory (GB)
Detection and segmentation	320 ± 5	4.2 ± 0.1
Region processing	260 ± 4	4.5 ± 0.1
4D Gaussian projection	580 ± 8	8.5 ± 0.2
Privacy release	150 ± 3	1.2 ± 0.05
Rendering (1080p)	80 ± 2	6.2 ± 0.1

Source: Authors' profiling from the five preserved execution runs on the stated NVIDIA A100 configuration.

6. DISCUSSION

6.1 Interpretation and contribution

The top-line result is not that privacy has been cost-justified, but rather that the manner and location of releasing

perturbations affects the distribution of utility loss. Two reference points in Table 5 must be kept separate. The no-privacy baseline applies no perturbation and achieves the highest reconstruction fidelity among the five tested configurations (Table 5). The non-adaptive Gaussian control (19.17 ± 0.68 dB PSNR, 0.832 ± 0.021 SSIM, 2.88 ± 0.15 mm RMS error) applies additive Gaussian noise at a single global scale to all primitives without targeting any sensitive region; it is this control - not the no-privacy baseline - that severely degrades both visual detail and geometric structure. Relative to the control, DT-SHIELD restricts the release to primitives under an associated source-reference sensitivity mask, while otherwise maintaining the previous attribute encoding of source-reference sensitive values elsewhere. The comparison with the control quantifies the combined effect of masking, local allocation, and the Laplace- versus Gaussian-perturbation decision, and is not attributed to masking alone (Section 4.2).

The ε = 0.5 release chosen lowers the camera-based empirical exposure indicator for source-reference-sensitive primitives (Section 3.5) while preserving nearly all of the architectural detail outside the source-reference-sensitive mask. Compared to the baseline, the ε = 0.5 release has a PSNR 1.66 dB lower (36.61 ± 0.13 versus 38.27 ± 0.12 dB) and SSIM 0.011 lower (0.964 ± 0.003 versus 0.975 ± 0.003) while preserving ATR = 94.20% ± 0.12% (Section 5.1). As the authors' own approach to local area processing before a global process, mask-primitive association, local budget partitioning, and consideration of visual, geometric, privacy, and efficiency issues, the exposure indicator presented here is an empirically derived, workflow-specific, source-reference, mask-based measure, not a formal privacy proof or analysis validated as an attack countermeasure.

6.2 Failure cases and security limitations

False negatives impose the greatest danger for a failure of masks- a missed area would leave the Gaussians unchanged. The omission of non-sensitive architectural features by false positives amounts to a utility diminishment. Errors from temporal mask propagation can be either type of error, whether it is occlusion, major shape change, or structure change at different time points of acquisition. A further related limitation is that only for the Heritage-Recon benchmark is independent metric ground truth available, so the geometric accuracy of the other two collections can only be characterised in terms of relative-scale residuals; hence no cross-collection RMS mean is given (Section 4.2). Because of this, the release of the data for the datasets should detect detector misses, mask overreach, failure of propagation, seams in processing, drift in geometry, and privacy-release exceptions as separate entities as opposed to aggregating them into a single number.

Formal privacy itself is further weakened by composition and mask coverage. For example, Eq. (10)'s local scale and Δ_i/ε_i will only have meaning if the primitives can be defined with the given definitions. Loss of privacy can occur from multiple primitive releases and the following editions. Without a formal accounting, it will be impossible to determine a total composed ε for any and all of the data pieces of the full released twin from the simple quotes of local budgets. In addition to the absence of an adaptive attacker model, the empirically based exposure indicator associated with masks (Section 3.5) does not allow the privacy review to be an attack-oriented one, with a distinction made between mechanism parameters and empirical resistance to either reconstruction or inversion [25]. It is thus a disclosure indicator, not a formal attack-validated privacy metric. Attack-based reconstruction/inversion testing, composition accounting, and adversarial evaluation of missing masks are required for the method to achieve ε formal privacy claims. Another limitation is the SER indicator: it is calculated relative to a held-out reference observation in M ref, and the residuals are summed by ratio-of-means (Section 3.5). As a result, it is affected by the choice of reference frame and by the mask boundary, and it is a workflow-defined disclosure indicator rather than an attack-validated privacy metric, and its intermediate residuals are shown in Table 3 for independent recomputation.

6.3 Data provenance and reproducibility limitation

Provenance differs across the four heritage data sets, and this asymmetry constrains what can be independently verified. Heritage-Recon is obtained from cultural-heritage scans submitted to Open Heritage 3D [18], and its source scans can be checked against a public archive by a third party. The working subset used in this paper was constructed and published by us from those submissions, and so can be verified by examining Open Heritage 3D submissions against the author manifest, not by re-downloading an archived release of our exact working subset. The other three data sets (Brandenburg Gate, Lincoln Memorial, Pantheon - Exterior) have no such external anchor: provenance can be confirmed only against the author manifest. (The Heritage-Recon Benchmark as defined in study [17] is a set of four scenes – Brandenburg Gate, Lincoln Memorial, Pantheon - Exterior, and El Palacio de Bellas Artes – obtained from Open Heritage 3D submissions [18]; we regard these four as our heritage data sets, and observe that provenance verification is stronger for Heritage-Recon than for the other three.

Reproducibility is restricted as follows: the detector backbone, checkpoint, pretraining/fine-tuning corpora, class definitions, confidence/segmentation thresholds, and software versions were not specified in the provided run records, so Figure 2 should be regarded as an execution record rather than a detector benchmark (the detector-record requirement in Section 3.2). The reference mask M_{ref} was constructed from the mechanism mask M_{mechan}, rather than being annotated blind to it, so the F1 and false-positive values reported in Table 2 are mask-curation consistency metrics and the SER statistic is evaluated on a reference that is derived from (but not equal to) the mask that caused the perturbation; no multi-annotator adjudication or inter-rater reliability estimate is available for the reported experiments (the two-annotator-plus-adjudicator requirement is a future consideration). There is also no formal privacy accounting: multiple and cumulative releases of primitives are not composed into an overall budget, and no attack based on inversion/leakage/model-based reconstruction was used. These limitations suggest that the release should be treated as an empirical reduction in mask-based exposure of sensitive features under the specified workflow, rather than as an end-to-end differential privacy assurance, and that an independent re-implementation of the detector and mask-curation stages is currently precluded by the material released.

7. CONCLUSION

DT-SHIELD is a preliminary (and therefore tentative)

privately preserving digital twin construction technique for multi-view temporal heritage image sets. In this experiment, the adaptive noise model (with a mask as input) compares favorably in visual/spatial content preservation to the baseline non-adaptive Gaussian noise, and the $\varepsilon = 0.5$ release reduces the empirical exposure indicator for the source-reference-sensitive areas by quite a bit while leaving most of the non-sensitive texture detail of the architecture. However, these results are based on what was actually observed in the data; i.e., five independent repeats, three source-provided collections, and this primitive level of sensitivity and primitive reference/source masks. This version does not claim any type of physical restoration; any universally optimal privacy budget; any differential privacy from start to finish; or any attack-validated formality of the exposure indicator privacy measure.

DATA AVAILABILITY AND REPRODUCIBILITY

Three sets of data contribute to this paper. (i) Heritage-Recon is based on the heritage scans submitted to Open Heritage 3D [18]; the source scans are clearly traceable within that archive, and the working subset for the paper is documented in the author manifest so that others can re-verify the subset against the submissions. (ii) The other heritage datasets (Brandenburg Gate, Lincoln Memorial, Pantheon - Exterior) lack an equivalent anchor; they can be obtained, with their author-supplied manifests, upon request of the corresponding author. (iii) the scene-aware splits, per-run scene logs (scene number, viewpoint, capture time, exclusion reason, registration status), per-release records (primitive ID, source mask/views, SPNR, q_i, ε_i, Δ_i, release index), and the $\varepsilon = 0.5$ run summaries from which Tables 2 and the reported metrics are computed are available on request from the corresponding author.

The limitations described in Section 6.3 are not repeated here. Heritage-Recon can be checked against Open Heritage 3D [18], whereas the other datasets can be verified only against the author-provided manifests. The detector and segmentation settings were not recorded, and the reference masks were not independently annotated. Future work should document the detector settings and use the annotation procedure described in Section 3.2. No formal differential privacy guarantee is claimed; the results show an empirical privacy–utility trade-off for the workflows examined.

ACKNOWLEDGMENT

The authors acknowledge the institutional support of the Department of Computer Science, College of Education, Mustansiriyah University.

REFERENCES

[1] Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, 51(11): 1016-1022. https://doi.org/10.1016/j.ifacol.2018.08.474

[2] Mazzetto, S. (2024). Integrating emerging technologies with digital twins for heritage building conservation: An interdisciplinary approach with expert insights and bibliometric analysis. Heritage, 7(11): 6432-6479. https://doi.org/10.3390/heritage7110300

[3] Vuoto, A., Funari, M.F., Lourenço, P.B. (2024). Shaping digital twin concept for built cultural heritage conservation: A systematic literature review. International Journal of Architectural Heritage, 18(11): 1762-1795. https://doi.org/10.1080/15583058.2023.2258084

[4] AlBalkhy, W., Karmaoui, D., Ducoulombier, L., Lafhaj, Z., Linner, T. (2024). Digital twins in the built environment: Definition, applications, and challenges. Automation in Construction, 162: 105368. https://doi.org/10.1016/j.autcon.2024.105368

[5] Voas, J., Mell, P., Laplante, P., Piroumian, V. (2025). Security and trust considerations for digital twin technology. NIST IR 8356. https://doi.org/10.6028/NIST.IR.8356

[6] Pan, X., Lin, Q., Ye, S., Li, L., Guo, L., Harmon, B. (2024). Deep learning based approaches from semantic point clouds to semantic BIM models for heritage digital twin. Heritage Science, 12: 65. https://doi.org/10.1186/s40494-024-01179-4

[7] Abdelalim, M. (2025). Heritage preservation using laser scanning: Architectural digital twins using Al-Mu'izz Street as a case study. Buildings, 15(9): 1480. https://doi.org/10.3390/buildings15091480

[8] Guo, M., Shang, X., Zhao, J., Liao, L. (2025). Digital twin models for architectural heritage conservation. Journal of Building Engineering, 112: 113792. https://doi.org/10.1016/j.jobe.2025.113792

[9] Felicetti, A., Niccolucci, F. (2025). Artificial intelligence and ontologies for the management of heritage digital twins data. Data, 10(1): 1. https://doi.org/10.3390/data10010001

[10] Hosamo, H., Mazzetto, S. (2025). Integrating knowledge graphs and digital twins for heritage building conservation. Buildings, 15(1): 16. https://doi.org/10.3390/buildings15010016

[11] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42(4): 139. https://doi.org/10.1145/3592433

[12] Wu, T., Yuan, Y.J., Zhang, L.X., et al. (2024). Recent advances in 3D Gaussian splatting. Computational Visual Media, 10: 613-642. https://doi.org/10.1007/s41095-024-0436-y

[13] Bao, Y., Ding, T., Huo, J., et al. (2025). 3D Gaussian splatting: Survey, technologies, challenges, and opportunities. IEEE Transactions on Circuits and Systems for Video Technology, 35(7): 6832-6852. https://doi.org/10.1109/TCSVT.2025.3538684

[14] Rangelov, D., Waanders, S., Waanders, K., van Keulen, M., Miltchev, R. (2025). Impact of data capture methods on 3D reconstruction with Gaussian splatting. Journal of Imaging, 11(2): 65. https://doi.org/10.3390/jimaging11020065

[15] Jia, Q., He, J. (2026). High-fidelity 3D reconstruction of cultural heritage via super-resolution and progressive Gaussian splatting. npj Heritage Science, 14: 84. https://doi.org/10.1038/s40494-026-02355-4

[16] Montas-Laracuente, N., Delgado Martos, E., Pesqueira-Calvo, C., et al. (2025). Automatic 3D reconstruction: Mesh extraction based on Gaussian splatting from

Romanesque–Mudéjar churches. Applied Sciences, 15(15): 8379. https://doi.org/10.3390/app15158379
[17] Sun, J., Chen, X., Wang, Q., et al. (2022). Neural 3D reconstruction in the wild. In ACM SIGGRAPH 2022 Conference Proceedings, Vancouver, BC, Canada, pp. 1-9. https://doi.org/10.1145/3528233.3530718
[18] McAvoy, S., Ristevski, J., Rissolo, D., Kuester, F. (2023). OpenHeritage3D: Building an open visual archive for site-scale giga-resolution LiDAR and photogrammetry data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 215-222. https://doi.org/10.5194/isprs-annals-X-M-1-2023-215-2023
[19] Dwork, C., McSherry, F., Nissim, K., Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In Proceedings of Third Theory of Cryptography Conference, New York, NY, USA, pp. 265-284. https://doi.org/10.1007/11681878_14
[20] Dwork, C., Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4): 211-487. https://doi.org/10.1561/0400000042
[21] Pan, K., Ong, Y.S., Gong, M., Li, H., Qin, A.K., Gao, Y. (2024). Differential privacy in deep learning: A literature survey. Neurocomputing, 589: 127663. https://doi.org/10.1016/j.neucom.2024.127663
[22] Demelius, L., Kern, R., Trügler, A. (2025). Recent advances of differential privacy in centralized deep learning: A systematic survey. ACM Computing Surveys, 57(6): 158. https://doi.org/10.1145/3712000
[23] Luo, Y., Hu, T., Ouyang, X., et al. (2025). DPO-Face: Differential privacy obfuscation for facial sensitive regions. Computers & Security, 154: 104434. https://doi.org/10.1016/j.cose.2025.104434
[24] Pietrantoni, M., Csurka, G., Sattler, T. (2025). Gaussian splatting feature fields for (privacy-preserving) visual localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, pp. 1082-1092. https://doi.org/10.1109/CVPR52734.2025.00109
[25] Zhang, X., Zhang, Q. (2025). Defending against attacks in deep learning with differential privacy: A survey. Artificial Intelligence Review, 58: 347. https://doi.org/10.1007/s10462-025-11350-3

NOMENCLATURE

ATR	Architectural texture retention (%)
D	Calibrated temporal image set
$G_i(t)$	Dynamic Gaussian primitive i at time t
H_φ	Masked image-region processor
M(t,v)	Sensitive-region mask at time t and view v
P_i	Signal power of primitive i
q_i	Normalized privacy-allocation score
R	Differentiable Gaussian renderer
S	Set of sensitive Gaussian primitives
SPNR_i	Signal-to-privacy-noise ratio
Δ_i	L1 sensitivity bound
ε_i	Local privacy budget
η	Numerical stabilizer
$\lambda_1, \lambda_s, \lambda_t$	Loss-function weights

APPENDIX

A. Reproducibility and annotation governance

This Appendix outlines the minimum evidence fields to support a reproducible reanalysis of DT-SHIELD. The fields outlined in this appendix are a prospective governance requirement, and do not create retrospective totals for scene counts, annotators' identities, detectors' checkpoint times, or licenses that have been omitted from the provided experimental record.

The four distinct issues that aggregated accuracy does not address are data provenance, detector reproducibility, annotation reliability, and privacy release accounting. If you record the above-mentioned fields in addition to a summary metric of total images or an image count, it will be possible to determine the number of scenes labeled, the quality of the labels, the privacy of each release, etc.

Table A1. Minimum reproducibility and annotation-governance evidence required for an auditable replication

Evidence Field	Required Record	Why It is Required
Dataset provenance	Repository/project ID, DOI/URL, owner, licence, version/date	Confirms lawful and repeatable data origin
Scene/view inventory	Scene ID, viewpoint ID, acquisition time, camera/trajectory ID	Establishes independent units and leakage control
Sensitive-region ledger	Region ID, operational class, source image, mask version	Defines the privacy-evaluation denominator
Annotation governance	Two independent annotations, adjudication status, agreement statistic	Quantifies label reliability for replication
Detector record	Backbone/model family, checkpoint hash, pretraining/fine-tuning data, thresholds	Makes mask generation reproducible
Split/exclusion record	Train/validation/test scene assignment and exclusion reason	Prevents hidden data leakage
Privacy-release ledger	Primitive ID, mask link, ε_i, Δ_i, release index, checksum	Supports release accounting and audit
Run/evaluation record	Seed, checkpoint, hardware/software, per-scene/per-run metrics	Supports uncertainty and computational replication