# TOPOAUDIT3D: GEOMETRY IS NOT TOPOLOGY IN IMPLICIT SURFACE RECONSTRUCTION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Implicit surface reconstruction is usually ranked by overlap or surface distance, although a geometrically accurate level set can contain the wrong number of components, tunnels, or cavities. We present TopoAudit3D, a source-audited framework with two explicitly separated components: a model-agnostic Betti-stable output audit and TopoAudit3D-SDF, a controlled localized-residual SDF reconstruction probe. The audit maps heterogeneous outputs to one physical coordinate system, evaluates topology across reference-stable offsets, and binds aggregates to case identities, evaluator replay, and provenance. Cardiac surfaces are used only as a demanding topology stress test for closed implicit surfaces. The frozen archive contains 2,793 instances, five reconstruction tracks, and 13,965 method–instance records. We reserve the 1,920-instance, repair-free M&Ms track as primary evidence. NKSR obtains lower ASSD than TopoAudit3D-SDF (0.4479 vs. 0.5200 mm) but only 7.76% topology success versus 68.70%. Among 1,203 cases where TopoAudit3D-SDF passes the complete stable-offset criterion and NKSR fails, NKSR still has lower ASSD in 784. A controlled 1,080-run study shows why evaluation and causal attribution must remain separate: residual gating and hard-example sampling help within TopoAudit3D-SDF, yet a capacity-matched flat MLP is stronger on every reported endpoint of the 60-case subset, while removing gradient detachment slightly improves geometry. We therefore make neither an identical-input claim nor an architecture-superiority claim. Instead, we establish that geometric leadership does not certify topological reliability and derive sufficient conditions under which a bounded residual deformation preserves the isotopy class of a zero level set.

## 1 INTRODUCTION

Continuous implicit fields represent a surface as a level set rather than a fixed mesh. This flexibility underlies DeepSDF, Occupancy Networks, neural point reconstruction, learned-kernel reconstruction, and modern PDE-inspired solvers (Park et al., 2019; Mescheder et al., 2019; Gropp et al., 2020; Ma et al., 2021; Peng et al., 2021; Huang et al., 2023; Wang et al., 2025). Yet the community still compares reconstructions mainly through overlap and surface distance. Those metrics are indispensable, but they answer a local geometric question. A pinhole, thin bridge, spurious component, or enclosed cavity can change a Betti number while moving only a small fraction of the surface.

Three scientific confounders follow. First, geometry and topology are mathematically distinct: small average distance does not certify a correct level-set connectivity. Second, reconstruction papers often compare methods with different observations, such as a dense signed field, oriented points, or unoriented points; a common table can then blur method effects with information effects. Third, topology is sensitive to coordinate conversion, offset choice, and case inclusion, so an untracked evaluator change can alter the conclusion even when predictions are frozen. The central question is therefore not simply which method has the smallest error, but which claims remain valid after geometry, topology, supervision, and provenance are separated.

Existing work improves field fitting through sign-agnostic objectives, local patches, query pulling, Poisson solvers, learned kernels, and analytic or Green-identity formulations (Atzmon & Lipman, 2020; Erler et al., 2020; Ma et al., 2021; Peng et al., 2021; Huang et al., 2023; Feng et al., 2026; Wu et al., 2026). A complementary line uses persistent homology, Euler characteristic, skeletons,



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

or Betti matching to impose or evaluate structure (Hu et al., 2019; Clough et al., 2022; Shit et al., 2021; Stucki et al., 2023; Dong et al., 2022; Jignasu et al., 2024; Tang et al., 2019). These methods show that topology can be optimized. They do not eliminate the need for an output-level protocol that remains meaningful when observations and solvers differ.

We answer this need with TopoAudit3D. The protocol maps each reconstruction to a common physical field, measures Betti agreement over a frozen set of reference-stable offsets, reports graded and strict topology endpoints, and adds paired "geometry-better/topology-worse" contradictions. We instantiate it on three public cardiac datasets because closed, thin, multi-compartment anatomy is a demanding topology stress test, not because the protocol assumes a heart-specific model. To study representation mechanisms without conflating them with the cross-paradigm comparison, we use TopoAudit3D-SDF as the controlled dense-field reconstruction probe. We separate its representation from its optimization, derive a sufficient zero-set stability condition, and preserve negative controls. Throughout the paper, TopoAudit3D denotes the overall framework, while TopoAudit3D-SDF denotes only the reconstruction instantiation used in comparison and ablation tables. To preserve the frozen visual artifacts exactly, all archived plots retain the historical legend label "HSDF"; in every figure, "HSDF" denotes TopoAudit3D-SDF.

Our contributions are:

1. **A general topology audit.** TopoAudit3D combines canonical physical conversion, reference-stable Betti curves, BME/StableD/StableA/TSR, paired rescue–regression analysis, and source closure over 13,965 frozen records.
2. **A repair-free empirical finding.** On 1,920 M&Ms instances, geometry and topology disagree both in aggregate ranking and in hundreds of paired cases, directly establishing that lower geometric error is not a topology certificate.
3. **Claim-scoped representation evidence.** A controlled TopoAudit3D-SDF study identifies supported internal components, falsifies a broad architecture-superiority claim through a stronger capacity-matched MLP control, and links bounded residuals to topology through an explicit sufficient-condition theorem rather than an informal guarantee.

## 2 Related Work

**Implicit surface reconstruction.** DeepSDF and Occupancy Networks established continuous neural shape representations (Park et al., 2019; Mescheder et al., 2019). IGR, SAL, Points2Surf, and Neural-Pull address raw, unsigned, or sparse observations through gradient regularization, sign-agnostic objectives, local features, or query pulling (Gropp et al., 2020; Atzmon & Lipman, 2020; Erler et al., 2020; Ma et al., 2021). Classical Poisson reconstruction solves a global indicator-field problem, while Shape as Points couples oriented samples to a differentiable Poisson solver (Kazhdan et al., 2006; Peng et al., 2021); NKSR learns a sparse kernel hierarchy (Huang et al., 2023); HotSpot derives a screened-Poisson-inspired condition for SDF optimization (Wang et al., 2025); PAT and G3R use local tori and Green's third identity (Feng et al., 2026; Wu et al., 2026). Their different observations and inductive biases make a common output audit informative, but not an identical-input competition.

**Topology-aware learning and reconstruction.** Persistent-homology losses, clDice, and induced Betti matching reveal structural errors missed by local losses (Hu et al., 2019; Clough et al., 2022; Shit et al., 2021; Stucki et al., 2023). Topology-controllable B-spline reconstruction optimizes persistence-based targets (Dong et al., 2022), while STITCH adds persistent-homology constraints to neural implicit reconstruction from sparse point clouds (Jignasu et al., 2024). Skeleton-bridged reconstruction uses a topology-preserving intermediate representation for single-view mesh generation (Tang et al., 2019). These methods primarily ask how to impose a topology prior. TopoAudit3D instead asks whether topology is correct and stable after reconstruction, including when no explicit topology objective is present.

**Application stress tests and provenance.** Cardiac reconstruction motivates valid surfaces for measurement and simulation through template deformation, learned SDFs, and simulation-oriented meshing (Kong et al., 2021; Verhülsdonk et al., 2024; Yang et al., 2024; Narayanan et al., 2024). We use this domain because topology failures are consequential and readily exposed, while our definitions

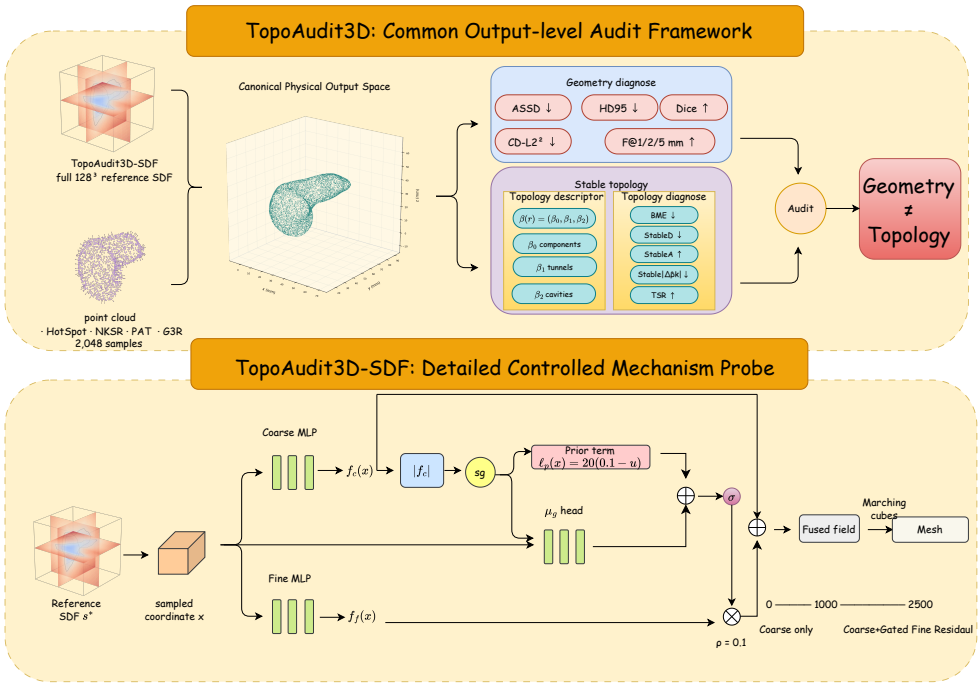


Figure 1: Overview of TopoAudit3D.Top: heterogeneous reconstructions are mapped to a canonical physical output space and audited separately for geometry and stable topology. Bottom: TopoAudit3D-SDF reconstructs the reference SDF with coarse and gated fine-residual MLP branches, followed by field fusion and Marching Cubes extraction; training switches from coarse-only optimization to coarse + gated fine residual at step 1,000.

require only a reference closed surface and physical coordinate map. Case closure, evaluator replay, supervision disclosure, and artifact hashes are treated as part of the method rather than supplementary bookkeeping.

# 3 METHODOLOGY

## 3.1 OVERVIEW AND PROBLEM FORMULATION

TopoAudit3D is the proposed framework: it combines a common output-level topology audit with a controlled dense-field instantiation, TopoAudit3D-SDF, used to study representation mechanisms under identical dense-SDF supervision. Figure 1 separates these two roles explicitly. The audit answers whether reconstructed zero level sets are geometrically accurate and topologically stable; the controlled probe asks which representation and optimization choices are supported when supervision is held fixed.

## 3.2 TOPOAUDIT3D: CANONICAL OUTPUT SPACE

Let method $m$ produce a mesh or field for instance $i$. We convert every output to a canonical physical-coordinate SDF $F_i^m : \Omega_i \to \mathbb{R}$ and evaluate it against reference field $G_i$ on the same domain. Mesh inside–outside queries use generalized winding numbers (Jacobson et al., 2013). This equalizes the *output and evaluator space*; it deliberately does not claim that methods receive identical information.

### 3.3 REFERENCE-STABLE BETTI CURVES

For offset $r$, define the sublevel set $V_{i,r}^m = \{x : F_i^m(x) \leq r\}$ and Betti tuple

$$\boldsymbol{\beta}_{i,r}^m = (\beta_0, \beta_1, \beta_2), \tag{1}$$

counting connected components, tunnels, and enclosed cavities. We use the frozen symmetric physical grid $\mathcal{R} = \{-10, -5, -2.5, 0, 2.5, 5, 10\}$ mm. These offsets are not presented as clinical thresholds; they are a fixed local-to-coarse perturbation grid frozen before comparative analysis. For each instance, $\mathcal{R}_i \subseteq \mathcal{R}$ retains offsets at which the reference topology agrees with its zero-offset tuple, preventing a method from being penalized for reference topology changes. With prediction $P$ and reference $G$,

$$\mathrm{StableD}_i = \frac{1}{|\mathcal{R}_i|} \sum_{r \in \mathcal{R}_i} \|\boldsymbol{\beta}_{i,r}^P - \boldsymbol{\beta}_{i,r}^G\|_1, \tag{2}$$

$$\mathrm{StableA}_i = \frac{1}{|\mathcal{R}_i|} \sum_{r \in \mathcal{R}_i} \mathbf{1}[\boldsymbol{\beta}_{i,r}^P = \boldsymbol{\beta}_{i,r}^G], \tag{3}$$

$$\mathrm{TSR}_i = \mathbf{1}[\forall r \in \mathcal{R}_i : \boldsymbol{\beta}_{i,r}^P = \boldsymbol{\beta}_{i,r}^G]. \tag{4}$$

StableD and StableA are cardinality-normalized graded endpoints. TSR is intentionally stricter: one mismatch makes the instance fail. We therefore never interpret TSR alone. BME measures unmatched spatially corresponding persistent features (Stucki et al., 2023); stable $|\Delta\beta_k|$ decomposes StableD by homology dimension.

### 3.4 PAIRED CONTRADICTIONS AND SOURCE CLOSURE

Dataset means can hide the direction of instance-level changes. Let $p$ denote the TopoAudit3D-SDF reconstruction track and $b$ a baseline. We define a topology rescue

$$R_i(b) = \mathbf{1}[\mathrm{TSR}_i^p = 1 \wedge \mathrm{TSR}_i^b = 0] \tag{5}$$

and a regression by reversing the two success indicators. Within rescued cases we also count

$$C_i^{\mathrm{ASSD}}(b) = R_i(b)\mathbf{1}[\mathrm{ASSD}_i^b < \mathrm{ASSD}_i^p], \tag{6}$$

with an analogous Dice contradiction. A large count is direct evidence that better geometry and worse topology coexist on the *same* object.

Every aggregate is tied to case identifiers, raw evaluator records, and SHA-256 manifests. We require exact case-set closure within a dataset, recompute all table values from case records, and replay each dataset–method block through the frozen evaluator. This separates numerical authenticity from causal interpretation: the audit can prove that a result is reproducible without pretending that heterogeneous observations establish architecture superiority.

### 3.5 TOPOAUDIT3D-SDF: LOCALIZED RESIDUAL RECONSTRUCTION PROBE

#### 3.5.1 REPRESENTATION AND OPTIMIZATION

Let $s^* : \Omega \to \mathbb{R}$ be a reference SDF on $\Omega = [-1, 1]^3$. Two independent three-layer ReLU MLPs of width 128 predict a base field $f_c$ and residual $f_f$. For a generic input $z$,

$$h_1 = \phi(W_1 z + b_1), \quad h_2 = \phi(W_2 h_1 + b_2), \quad m(z) = W_3 h_2 + b_3. \tag{7}$$

The deterministic gate uses

$$u(x) = \mathrm{sg}(|f_c(x)|), \qquad \ell_p(x) = \alpha(\tau - u(x)), \tag{8}$$

$$g(x) = \sigma\left(\frac{\ell_p(x) + \mu_g(x, u(x))}{T}\right), \tag{9}$$

with $\tau = 0.1$, $\alpha = 20$, $T = 1$, and final field

$$f_\theta(x) = f_c(x) + \rho g(x) f_f(x), \qquad \rho = 0.1. \tag{10}$$

The frozen code retains a log-variance head for backward compatibility, but stochastic sampling is disabled ($\varepsilon = 0$); it has no effect on the reported output and is not claimed as a contribution. Since $0 \le g \le 1$,

$$|f_\theta(x) - f_c(x)| \le \rho |f_f(x)|. \tag{11}$$

Stop-gradient removes the coarse-to-gate chain-rule term, so it controls an optimization path rather than guaranteeing better accuracy. The controlled experiment tests this distinction explicitly.

Training is also part of the probe. For steps $t < 1000$, only $f_c$ forms the output; afterward Equation 10 is activated. The frozen objective is L1 field reconstruction plus fine-branch and gate regularization; no Eikonal, Euler, persistent-homology, or multiscale topology loss is active. Adam runs for 2,500 steps with learning rate $10^{-4}$ and batch size 4,096. The base sampler mixes near-surface, uniform, and axial-stratified pools in proportions $(0.6, 0.2, 0.2)$. Every 1,000 steps, $15\%$ of subsequent samples are reassigned to high-error and boundary-neighborhood locations.

**Complexity.** For $N$ query points and hidden width $H$, the coordinate branches and deterministic gate require $O(NH^2)$ arithmetic and $O(NH)$ activation memory up to constant factors; the frozen instantiation uses $H = 128$ and 68,868 parameters. Hard-pool refresh evaluates a $128^3$ field every 1,000 steps, so its cost is periodic rather than incurred at every update. Algorithm 1 and the exact parameter count are provided in Appendix B.

#### 3.5.2 When can a bounded residual preserve topology?

The pointwise bound in Equation 11 alone is insufficient: an arbitrarily small perturbation can change topology near a critical zero. The missing link is a margin and regularity condition.

**Theorem 1** (Sufficient zero-set isotopy condition)**.** *Let $\Omega$ be compact, $f_c, h \in C^1(\Omega)$, and $f_t = f_c + th$ for $t \in [0,1]$. Suppose an open $U \Subset \operatorname{int}\Omega$ contains $f_c^{-1}(0)$ and*

$$\inf_{x \in \Omega \setminus U} |f_c(x)| > \|h\|_\infty, \tag{12}$$

$$\inf_{x \in U} \|\nabla f_c(x)\|_2 > \sup_{x \in U} \|\nabla h(x)\|_2. \tag{13}$$

*Then $0$ is a regular value of every $f_t$, all zero sets $f_t^{-1}(0)$ lie in $U$, and $f_0^{-1}(0)$ and $f_1^{-1}(0)$ are ambiently isotopic. Consequently, their Betti numbers agree.*

For TopoAudit3D-SDF, $h = \rho g f_f$, so

$$\|h\|_\infty \le \rho \|f_f\|_\infty, \tag{14}$$

$$\|\nabla h\|_\infty \le \rho\big(\|\nabla g\|_\infty \|f_f\|_\infty + \|g\|_\infty \|\nabla f_f\|_\infty\big). \tag{15}$$

A small residual scale and localized gate therefore make the sufficient inequalities easier to satisfy, but the frozen training does not enforce them. The theorem supplies a mechanism, not an unconditional guarantee. Appendix A gives the proper-submersion/isotopy proof (Hirsch, 1976).

## 4 Experiments

### 4.1 Testbed and data protocol

An instance is one labeled structure at one phase for one subject. The frozen archive contains 600 ACDC instances (Bernard et al., 2018), 273 MM-WHS instances (Zhuang et al., 2019), and 1,920 M&Ms instances (Campello et al., 2021). These datasets instantiate a demanding family of closed surfaces; they are not used to define a heart-specific algorithm. The five frozen tracks are TopoAudit3D-SDF, HotSpot, NKSR, PAT, and G3R. TopoAudit3D-SDF fits the complete normalized $128^3$ SDF per instance. The four comparison tracks reconstruct from 2,048 surface samples and differ in normal assumptions. Thus Table 1 is a cross-paradigm behavior study under one output evaluator, not a same-input leaderboard.

The historical ACDC/MM-WHS correction stage used each test instance's reference SDF to localize and accept compact changes: 272 attempts yielded 187 successes. These results are not deployable inference and do not support the main method claim. Main-text comparisons therefore center M&Ms; complete upper-bound tables are in Appendix D.

Table 1: Evidence roles are fixed before interpreting results. ACDC and MM-WHS preserve historical same-case GT-assisted corrections and are diagnostic upper bounds; repair-free M&Ms is the primary comparative evidence.

| Dataset | Inst. | Eval. | Role | TopoAudit3D-SDF correction status |
|---|---|---|---|---|
| M&Ms | 1,920 | 9,600 | Primary repair-free stress test | None; frozen backbone only |
| ACDC | 600 | 3,000 | Diagnostic upper bound | 123 GT-assisted successes |
| MM-WHS | 273 | 1,365 | Diagnostic upper bound | 64 GT-assisted successes |
| Total | 2,793 | 13,965 | Exact method-wise case closure in each dataset. | |

Table 2: Frozen settings used by the proposed framework and controlled reconstruction probe.

| Block | Setting | Value |
|---|---|---|
| TopoAudit3D | Physical offsets (mm) | $\{-10, -5, -2.5, 0, 2.5, 5, 10\}$ |
| TopoAudit3D-SDF | Input field | normalized $128^3$ SDF |
| TopoAudit3D-SDF | MLP / gate | 3 layers, width 128; $\tau = 0.1$, $\alpha = 20$, $T = 1$ |
| TopoAudit3D-SDF | Residual / detach | $\rho = 0.1$; stop-gradient enabled |
| Optimization | Steps / stage switch | 2,500 / 1,000 |
| Optimization | Adam / batch | $10^{-4}$ / 4,096 |
| Sampling | Base / hard pool | $(0.6, 0.2, 0.2)$ / 15%, refresh 1,000 |
| Controlled study | Repeats | seeds 42, 123, 456; 60 cases; 6 variants |

For controlled evidence, six variants are trained on the same deterministic 60-instance ACDC subset (20 RV, 20 MYO, 20 LV) with seeds 42, 123, and 456: full TopoAudit3D-SDF, a 69,001-parameter flat MLP matched to TopoAudit3D-SDF's 68,868 parameters, coarse only, no detach, no hard pool, and an ungated residual. This yields 1,080 raw evaluations. Statistical inference uses the anatomical case after averaging its three seeds, paired Wilcoxon tests with Pratt zero handling, and Holm correction over five variants per endpoint.

### 4.2 Settings and configurations

Table 2 records the frozen settings that directly affect the reported analysis. The comparative archive was generated on heterogeneous historical hardware, so hardware is treated as provenance rather than a ranking variable; we do not report a synthetic runtime leaderboard from incomparable logs.

### 4.3 Evaluation metrics and source closure

Geometry uses Dice↑, ASSD↓, HD95↓, squared-$L_2$ Chamfer↓, and F-score at 1/2/5 mm↑. Topology uses BME↓, stable $|\Delta\beta_k|$ ↓, StableD↓, StableA↑, and TSR↑. Full formulas are in Appendix C. All 13,965 records pass schema and finiteness checks; 210 aggregate values recompute with zero mismatch; and 15/15 dataset–method blocks replay with maximum absolute difference zero. Runtime is not ranked because archived predictions were generated on heterogeneous hardware; the paper does not convert incomparable historical timings into a false efficiency claim.

### 4.4 Comparative experiments: repair-free geometry–topology discordance

Table 3 reports the repair-free M&Ms track. No method is a universal geometric winner. NKSR has the best ASSD and Chamfer, TopoAudit3D-SDF the best Dice and HD95, and HotSpot the second-best ASSD. The topology ordering is much clearer: TopoAudit3D-SDF has the lowest BME and StableD and highest StableA and TSR. This does not identify the cause as architecture, because observation regimes differ. It does establish that ranking by one geometric column cannot certify topology.

Figure 2 strengthens this from an aggregate observation to a paired empirical fact. Relative to NKSR, TopoAudit3D-SDF rescues 1,203 topology failures and regresses on 33. In 784 of the rescued cases, NKSR still has lower ASSD. The same pattern is not unique to NKSR: within TopoAudit3D-SDF rescues, HotSpot has lower ASSD in 379 cases, PAT in 230, and G3R in 34. Baselines also have

Table 3: Repair-free M&Ms comparison. All methods share 1,920 instance IDs, physical conversion, and evaluator; observations differ. Arrows indicate direction. Best and second-best values are bold and underlined.

| Method | ASSD ↓ | HD95 ↓ | CD-$L_2^2$ ↓ | Dice ↑ | BME ↓ | StableD ↓ | StableA ↑ | TSR ↑ |
|---|---|---|---|---|---|---|---|---|
| TopoAudit3D-SDF | 0.5200 | **1.6030** | <u>0.8442</u> | **0.9571** | **2.316** | **0.1289** | **90.50%** | **68.70%** |
| HotSpot | <u>0.4706</u> | 1.9870 | 1.3092 | <u>0.8877</u> | <u>46.436</u> | 5.2040 | <u>69.61%</u> | <u>38.80%</u> |
| NKSR | **0.4479** | <u>1.6898</u> | **0.4789** | 0.8336 | 189.510 | 26.8396 | 48.81% | 7.76% |
| PAT | 1.0548 | 5.9529 | 8.4341 | 0.8007 | 74.284 | 5.7630 | 51.85% | 16.61% |
| G3R | 1.1578 | 5.2620 | 6.3690 | 0.8123 | 10.905 | <u>1.5979</u> | 62.74% | 28.28% |

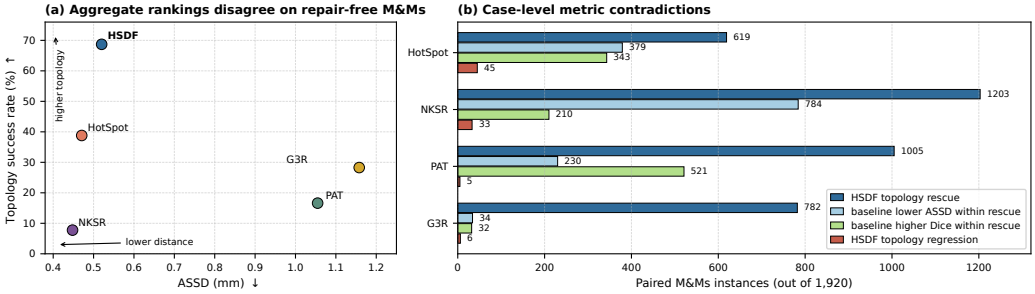


Figure 2: Geometry–topology discordance on repair-free M&Ms. (a) Aggregate ASSD and TSR rankings disagree. (b) A rescue means TopoAudit3D-SDF passes the complete stable-offset criterion while the baseline fails; the lighter bars count rescued cases where the failing baseline nevertheless has lower ASSD or higher Dice. Regressions are shown separately. The frozen legend label "HSDF" denotes TopoAudit3D-SDF.

higher Dice in 343, 210, 521, and 32 rescued cases, respectively. Geometry-better/topology-worse is therefore not a rhetorical possibility or a few outliers; it occurs repeatedly on the same objects.

Graded metrics reach the same conclusion without TSR's all-or-nothing threshold. TopoAudit3D-SDF StableA is 90.50%, versus 69.61% for the next highest method, and StableD is 0.1289 versus 1.5979 for the next lowest. Thus the finding is not created by one strict binary endpoint.

**Diagnostic upper bounds.** ACDC and MM-WHS show the same geometric–topological ranking mismatch, but their TopoAudit3D-SDF rows include 187 successful same-case GT-assisted corrections and are therefore treated only as diagnostic upper bounds. They are absent from the main comparison table and from the abstract's method-performance claim. Appendix D reports them with the disclosure adjacent to the numbers. Removing them would reduce study breadth but would not change the primary M&Ms conclusion.

### 4.5 Ablation experiments: controlled representation study

Table 4 answers a narrower causal question under identical dense-SDF supervision. Removing the fine residual (A2), hard pool (A4), or gate (A5) worsens the principal geometric endpoints. This supports their roles *inside the frozen system*. It does not support overall hierarchy superiority: the capacity-matched flat MLP is best on every reported endpoint, including descriptive topology values. No-detach also slightly improves geometry, so stop-gradient is retained as interpretable gradient control rather than an indispensable performance module.

This negative control sharpens the paper's novelty. The contribution is not "TopoAudit3D-SDF beats a plain MLP." Rather, the archive exposes why a method paper can appear compelling under cross-paradigm topology numbers while failing a same-supervision architecture test. TopoAudit3D makes that distinction explicit and reusable.

Table 4: Three-seed controlled study on a fixed 60-instance subset. Values are mean ± standard deviation across seed-level means. † denotes Holm-adjusted $p < 0.05$ versus A0 after case-level seed averaging; it marks a difference, not necessarily an improvement. Best values are bold.

| Variant | Dice ↑ | ASSD (mm) ↓ | HD95 (mm) ↓ | StableD ↓ | TSR ↑ |
|---|---|---|---|---|---|
| A0 Full TopoAudit3D-SDF | 0.9633±0.0013 | 0.5181±0.0117 | 1.4395±0.0433 | 0.1491±0.0049 | 71.11±1.92% |
| A1 Flat MLP (matched) | **0.9709±0.0003**† | **0.4345±0.0031**† | **1.0860±0.0236**† | **0.1389±0.0127** | **75.00±1.67%** |
| A2 Coarse only | 0.9590±0.0016† | 0.5692±0.0158† | 1.6586±0.0807† | 0.1638±0.0050 | 68.33±0.00% |
| A3 No detach | 0.9646±0.0008† | 0.5061±0.0077† | 1.3913±0.0194† | 0.1493±0.0061 | 69.44±2.55% |
| A4 No hard pool | 0.9621±0.0012† | 0.5356±0.0112† | 1.5567±0.0543† | 0.1726±0.0049† | 66.11±0.96% |
| A5 Ungated residual | 0.9611±0.0013† | 0.5441±0.0122† | 1.5684±0.0667† | 0.1551±0.0010 | 70.56±0.96% |

## 5 Discussion

**What is established.** The repair-free evidence supports a general empirical statement: surface-distance leadership and stable Betti reliability can disagree at scale, including on the same instance. Source replay establishes that the finding is not an aggregation or case-selection artifact. The controlled study supports the fine residual, gate, and hard-error sampling as internal mechanisms, while falsifying a broader hierarchy-superiority claim.

**What is not established.** The cross-paradigm comparison cannot attribute TopoAudit3D-SDF's topology values to architecture because the dense SDF contains more information than 2,048 surface samples. ACDC and MM-WHS are upper-bound diagnostics, not deployable tests. Detachment changes gradient routing but is not empirically necessary here. The theorem gives sufficient conditions but the training objective does not enforce them. These boundaries are part of the result, not admissions hidden after the conclusion.

**Scope.** Cardiac surfaces are one high-stress application family; extending the audit to synthetic controlled-genus objects, CAD surfaces, and non-anatomical benchmarks would test external generality. The seven offset values form a frozen symmetric perturbation grid rather than clinically privileged thresholds. StableD and StableA normalize by the retained offset count, while strict TSR may still be harder for cases with more valid offsets; future releases should include per-case $|\mathcal{R}_i|$ and offset-grid sensitivity.

## 6 Conclusion

We introduced TopoAudit3D, a source-audited protocol for separating geometric accuracy, topological reliability, observation regime, and provenance in implicit surface reconstruction. On 1,920 repair-free topology-stress-test instances, methods with lower surface error frequently fail topology where the dense-field probe succeeds, and hundreds of these contradictions occur on the same objects. A controlled study then prevents the wrong causal conclusion: a matched flat MLP outperforms TopoAudit3D-SDF on the available same-supervision subset. The resulting message is stronger than a conventional leaderboard claim. Geometry is necessary, topology must be measured directly, and architecture claims must be scoped to experiments that actually isolate architecture.

The main limitation is external identification rather than numerical reliability: the cross-paradigm archive does not hold input information constant, and the present stress tests are all cardiac surfaces. The highest-value next tests are therefore (i) an identical-supervision comparison between TopoAudit3D-SDF and simpler capacity-matched implicit networks on a full repair-free set, and (ii) validation on synthetic controlled-genus and non-anatomical closed surfaces together with offset-grid sensitivity. These extensions would test causal attribution and external generality without changing the central audit principle.

## Reproducibility Statement

The source bundle includes the 13,965-row case-level table, frozen aggregate tables, derived M&Ms contradiction counts, figure-generation code, evaluator replay summaries, and SHA-256 manifests. No prediction was regenerated or altered for this rewrite.

## Ethics Statement

The work analyzes public challenge data and does not make clinical claims. Topology metrics are research endpoints and must not be interpreted as evidence that a reconstruction is safe for diagnosis or intervention.

## Statement on the Use of Generative AI

Generative AI tools were used for language editing, code assistance, and figure planning. All numerical values, equations, citations, and claims were checked against the frozen experiment archive and cited primary records by the authors.

## References

Matan Atzmon and Yaron Lipman. SAL: Sign agnostic learning of shapes from raw data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2565–2574, 2020.

Olivier Bernard et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11): 2514–2525, 2018.

Víctor M. Campello et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms challenge. *IEEE Transactions on Medical Imaging*, 40(12):3543–3554, 2021.

James R. Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A. Zimmer, Julia A. Schnabel, and Andrew P. King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 8766–8778, 2022.

Zhetong Dong, Jinhao Chen, and Hongwei Lin. Topology-controllable implicit surface reconstruction based on persistent homology. *Computer-Aided Design*, 150:103308, 2022.

Philipp Erler, Paul Guerrero, Stefan Ohrhallinger, Niloy J. Mitra, and Michael Wimmer. Points2Surf: Learning implicit surfaces from point clouds. In *Computer Vision – ECCV 2020*, volume 12350 of *Lecture Notes in Computer Science*, pp. 108–124, 2020.

Nicole Feng, Ioannis Gkioulekas, and Keenan Crane. Points as tori: Fast pointwise signed distance for point clouds. *ACM Transactions on Graphics*, 45(4):53:1–53:24, 2026.

Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3789–3799, 2020.

Morris W. Hirsch. *Differential Topology*, volume 33 of *Graduate Texts in Mathematics*. Springer, New York, 1976.

Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. In *Advances in Neural Information Processing Systems*, volume 32, pp. 5658–5669, 2019.

Jiahui Huang, Zan Gojcic, Matan Atzmon, Or Litany, Sanja Fidler, and Francis Williams. Neural kernel surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4369–4379, 2023.

Alec Jacobson, Ladislav Kavan, and Olga Sorkine-Hornung. Robust inside-outside segmentation using generalized winding numbers. *ACM Transactions on Graphics*, 32(4):33:1–33:12, 2013.

Anushrut Jignasu, Ethan Herron, Zhanhong Jiang, Soumik Sarkar, Chinmay Hegde, Baskar Ganapathysubramanian, Aditya Balu, and Adarsh Krishnamurthy. STITCH: Surface reconstrucTion using Implicit neural representations with Topology Constraints and persistent Homology. *arXiv preprint arXiv:2412.18696*, 2024.

Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, pp. 61–70, 2006.

Fanwei Kong, Nathan Wilson, and Shawn C. Shadden. A deep-learning approach for direct whole-heart mesh reconstruction. *Medical Image Analysis*, 74:102222, 2021.

William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4):163–169, 1987.

Baorui Ma, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. Neural-Pull: Learning signed distance function from point clouds by learning to pull space onto surface. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7246–7257, 2021.

Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3D reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4460–4470, 2019.

Arjun Narayanan, Fanwei Kong, and Shawn C. Shadden. LinFlo-Net: A two-stage deep learning method to generate simulation ready meshes of the heart. *Journal of Biomechanical Engineering*, 146(7):071005, 2024.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 165–174, 2019.

Songyou Peng, Chiyu Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: A differentiable Poisson solver. In *Advances in Neural Information Processing Systems*, volume 34, pp. 13032–13044, 2021.

Suprosanna Shit, Johannes C. Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien P. W. Pluim, Ulrich Bauer, and Bjoern H. Menze. clDice – a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16560–16569, 2021.

Nico Stucki, Johannes C. Paetzold, Suprosanna Shit, Bjoern Menze, and Ulrich Bauer. Topologically faithful image segmentation via induced matching of persistence barcodes. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32698–32727, 2023.

Jiapeng Tang, Xiaoguang Han, Junyi Pan, Kui Jia, and Xin Tong. A skeleton-bridged deep learning approach for generating meshes of complex topologies from single RGB images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4541–4550, 2019.

Jan Verhülsdonk, Thomas Grandits, Francisco Sahli Costabal, Thomas Pinetz, Rolf Krause, Angelo Auricchio, Gundolf Haase, Simone Pezzuto, and Alexander Effland. Shape of my heart: Cardiac models through learned signed distance functions. In *Proceedings of the 7th International Conference on Medical Imaging with Deep Learning*, volume 250 of *Proceedings of Machine Learning Research*, pp. 1584–1605, 2024.

Zimo Wang, Cheng Wang, Taiki Yoshino, Sirui Tao, Ziyang Fu, and Tzu-Mao Li. HotSpot: Signed distance function optimization with an asymptotically sufficient condition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1276–1286, 2025.

Zhonghao Wu, Dong Xiao, and Renjie Chen. Surface reconstruction from unoriented points via Green's 3rd identity. In *Proceedings of the ACM SIGGRAPH 2026 Conference Papers*, pp. 110:1–110:10, 2026.

Jiancheng Yang, Ekaterina Sedykh, Jason Ken Adhinarta, Hieu Le, and Pascal Fua. Generating anatomically accurate heart structures via neural implicit fields. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume 15001 of *Lecture Notes in Computer Science*, pp. 264–274, 2024.

Xiahai Zhuang et al. Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. *Medical Image Analysis*, 58:101537, 2019.