CALIBRATED HEAT-LOSS RECOVERY FROM BUILDING SCANS WITH NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

We introduce a calibrated thermal twin that localizes thermal bridges and quantifies section-level heat loss from as-built geometry, material information, and exterior infrared measurements. The central challenge is that distinct interior conductivity fields can produce the same surface temperature, making full-field reconstruction an ambiguous target for learned inversion. Our approach instead targets quantities supported by the boundary observation. The twin combines a geometry-conditioned neural surrogate for surface-temperature prediction with a diagnostic stage that recovers surface transmittance (U-value) through a physical surface-flux identity and localizes the lateral footprint of thermal bridges without reconstructing the interior conductivity field. Split-conformal calibration provides U-value intervals with distribution-free finite-sample marginal coverage under exchangeability. On the reported synthetic junction evaluation split generated with a solver validated against ISO 10211 reference cases, direct surface readout achieves a U-value mean absolute error of $0.039\,\mathrm{W\,m^{-2}\,K^{-1}}$, compared with $0.046\,\mathrm{W\,m^{-2}\,K^{-1}}$ when the same quantity is read from an optimized conductivity field. The surface-anomaly localizer achieves a true bridge-footprint IoU of 0.35, compared with 0.15 for optimization and at most 0.08 for learned field inverses. Across increasing measurement noise, direct U-value recovery remains more stable than optimization-based field recovery, while split-conformal intervals attain the nominal 90% marginal coverage target. Experiments in heat conduction, Darcy flow, and electrical-impedance tomography further show that low-dimensional boundary functionals can remain accurately recoverable as field reconstruction deteriorates. Experiments on real building geometry with synthetic bridges and on measured infrared data provide additional localization plausibility checks, while differentiable retrofit optimization illustrates a downstream use of the thermal twin. These results support designing inverse methods around quantities that the available observations actually constrain.

1 INTRODUCTION

Buildings lose a substantial fraction of their energy through the envelope, the outer shell of walls, roof, and floor that separates the heated interior from the outdoors (Pérez-Lombard et al., 2008). A disproportionate share of that loss escapes through a small number of localized defects called thermal bridges, where a balcony slab or a steel stud penetrates the insulation and opens a concentrated path for heat loss. Standard engineering models approximate the heat loss of an envelope by a one-dimensional flow perpendicular to each surface and so ignore thermal bridges altogether (ISo, 2017; ISO, 2017), and in practice the bridges are still identified manually. This challenge can be addressed by a thermal twin, a virtual model of a building that combines its geometry with its thermal properties. We propose a thermal twin that localizes thermal bridges and quantifies the heat loss of the wall section that carries them, which asks whether a building's thermal properties can be learned from a scan of its envelope together with a measurement taken only at its exterior surface.

Learned inverse operators and differentiable-solver inversions approach this by recovering the full interior conductivity field and scoring it against the truth (Molinaro et al., 2023; Raissi et al., 2019; Zhong et al., 2026), which assumes, rarely explicitly, that the field is determined by the observation once data and model capacity suffice. Under the surface-only observation available to a thermal camera the assumption fails, in principle and in practice. Heat spreads sideways on its way out, so

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.

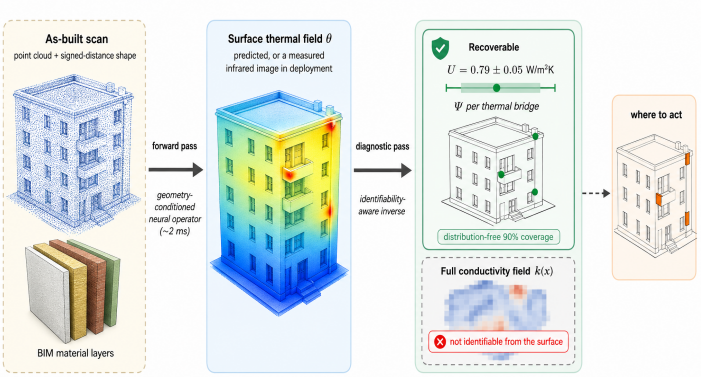


Figure 1: The thermal twin in two passes. A *synthesis* pass maps an as-built scan and its material layers to the surface thermal field, and a *diagnostic* pass reads the surface transmittance off that field with a calibrated interval and locates the bridge footprint.

the fine detail of where a defect sits is blurred away before it reaches the surface and many different interior fields produce the same surface temperature, exactly so for a layered wall and to within the measurement noise once a bridge is present, a fact classical in the theory of elliptic inverse problems (Calderón, 2006; Sylvester & Uhlmann, 1987; Alessandrini, 1988) whose consequence for learned inversion is what we draw out. On our ISO 10211-validated junction corpus of 120 validation details observed through a slab at the exterior face, optimization through the forward model returns a field with relative error 0.94 and three amortized inverses 0.68 to 0.80, where the same networks reached 0.26 to 0.46 when handed the interior, the optimized field fitting the observed slab to within about ten percent, and none reads the U-value better on that corpus than a direct calibrated surface readout, 0.042 to 0.206 against 0.036 W/m^{2}K. *The surface determines the U-value but not the depth at which the insulation fails.*

Our thermal twin therefore recovers the surface constant that the measurement does determine and returns it with a calibrated interval, rather than the interior conductivity field that it does not. That constant is the surface transmittance, or U-value, the heat lost through a square metre of wall for each degree of indoor-to-outdoor temperature difference, in W/m^{2}K, the quantity building codes are written against (ISo, 2017). The standards also define a linear thermal-bridge transmittance per metre of bridge (ISO, 2017), which needs the excess flow attributed among neighbouring bridges, so we locate the bridges here and leave that attribution to future work. In machine-learning terms the U-value is a low-dimensional quantity of interest, a scalar functional of the observed surface field rather than of the interior coefficient, which is why a boundary measurement determines it while it leaves the coefficient free.

The twin applies learning at three points, one in its synthesis pass and two in its diagnostic pass. In the *synthesis* pass a geometry-conditioned neural operator, trained on solver-generated fields, predicts the surface temperature of an envelope geometry it has not seen, a differentiable forward model in milliseconds where the solver takes seconds. Next, in the *diagnostic* pass, the departure of the observed face from the analytic clear-wall field locates the lateral footprint of the bridges, and an amortized inverse trained across many envelopes offers a learned alternative that localizes fins where its prior applies. Lastly, the U-value is read from a conservation identity with a two-parameter calibration fit on training envelopes rather than from any recovered field (Tejedor et al., 2017), so it does not inherit the surrogate's error, and split-conformal calibration wraps it in an interval whose coverage holds without any assumption on the error distribution (Vovk et al., 2005; Angelopoulos & Bates, 2023). Figure 1 sets out both passes.

Our central result is that as we raise the measurement noise, which is what makes the inverse problem more ill-posed, field recovery by optimization through the forward model degrades past the error of a homogeneous guess while the surface constant read directly off the boundary stays accurate. The pattern holds on heat conduction, Darcy flow, and electrical-impedance tomography, the archetype of the identifiability literature, on every one of 144 independent samples at every noise level. On the envelope corpora we compare the direct readout against the U-value read off every recovered field under one readout and one calibration, so the comparison is on the same target, and no route through a recovered field reads the U-value better at zero noise on two of the three corpora, on which neither

swept field route does at any noise level. On the third a global-attention inverse whose training prior recovers the field, which it does on fin corpora and not on junction details, reads it as well as or better than the direct readout at every noise level.

No prior method predicts a building's surface thermal field from an as-built point cloud with its material layers, and none reads a calibrated surface transmittance off that field while declining to reconstruct the interior. Geometry-conditioned neural operators already map point clouds to physical surface fields for aerodynamics and stress (Li et al., 2023; Wu et al., 2024), while learned building-thermal surrogates work on cross-sections that cannot represent a thermal bridge. On the diagnostic side, quantitative thermography reads a U-value from a single wall (Tejedor et al., 2017), deep networks detect bridges without quantifying them, and Bayesian inversion reconstructs the interior (De Simon et al., 2018). Targeting a functional rather than a field, and calibrating it without distributional assumptions, are established moves in inverse problems (Lieberman & Willcox, 2013; Wen et al., 2024), and our contribution is to make that choice for a stated identifiability reason, to ground the functional in a physical flux identity, and to show the whole system working on as-built envelopes.

Contributions. We make the following four contributions.

1. **Identifiability characterization.** Under single-excitation boundary observation, the interior conductivity field is not uniquely determined, exactly for layered walls and increasingly so with bridge depth. In contrast, surface transmittance remains directly recoverable and lateral bridge information remains observable. To our knowledge, this is the first identifiability treatment of envelope inverse thermography.
2. **Calibrated readout and footprint localization.** A two-parameter calibrated surface-flux readout of the U-value with a split-conformal interval of marginal coverage under exchangeability, and a footprint localizer that reads the lateral extent of a bridge off the surface anomaly and on real junction geometry locates the true bridge footprint where learned field inverses do not.
3. **Same-target evidence.** On every corpus the U-value read off a recovered field is compared with the direct readout under one readout and one calibration with geometry-level statistics over five seeds. No field route reads it better at zero noise on two of the three corpora, on which neither swept field route does at any noise level, the readout's margin over optimization widens with noise on all three, and the one exception, a global-attention inverse on a fin corpus, is exactly the case where its prior recovers the field.
4. **Generality and a real-geometry instance.** The characterization holds across heat conduction, Darcy flow, and electrical-impedance tomography, and the twin runs on as-built city-model geometry with ISO 10211-validated synthetic ground truth.

2 Related Work

Neural operators as forward surrogates. A neural operator learns a fast, discretization-agnostic surrogate for the forward map of a partial differential equation, on regular grids (Li et al., 2020), between function spaces (Lu et al., 2021), and on irregular as-built shapes (Li et al., 2023; Wu et al., 2024; Hao et al., 2023; Pfaff et al., 2020; Qi et al., 2017), and we use one as the differentiable engine of our synthesis pass.

Learned and optimization-based inverses. Learned operators and differentiable-solver optimization recover a spatially varying coefficient from observations. Neural inverse operators amortize the inversion across instances (Molinaro et al., 2023), physics-informed and differentiable-solver methods optimize a coefficient field to fit the data (Raissi et al., 2019; Zhong et al., 2026), and amortized inference places calibrated posteriors over the recovered parameters (Tong et al., 2024; Cranmer et al., 2020), all reporting field-recovery error as their measure of success. We instead target the surface constant a measurement determines, the goal-oriented view of inference (Lieberman & Willcox, 2013), for a different reason, that boundary observation leaves the field to the prior rather than merely making it high-dimensional.

Inverse thermography for buildings. Quantitative thermography estimates a building's thermal properties from a thermal image, per pixel under one-dimensional assumptions (Tejedor et al., 2017) or by Bayesian inference on idealized cross-sections (De Simon et al., 2018), and standards define

both constants (ISo, 2017; ISO, 2017). All operate on flat sections rather than three-dimensional geometry and pair no identifiability analysis with a calibrated recovery, whereas we operate on as-built envelopes and return a calibrated surface transmittance.

Identifiability and distribution-free calibration. A classical theory governs when a coefficient is recoverable from boundary data. The full Dirichlet-to-Neumann map determines the coefficient (Calderón, 2006; Sylvester & Uhlmann, 1987) with only logarithmic stability (Alessandrini, 1988; Mandache, 2001), and a single boundary excitation is less informative still. Split-conformal prediction gives finite-sample coverage for a scalar estimate without distributional assumptions (Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023) and has been used to calibrate a functional of an ill-posed reconstruction (Wen et al., 2024). We apply it to an identifiability-selected physical functional rather than to a generic task output.

3 Problem Setup

The twin has two jobs, predicting a building's surface temperature from what a scan provides and then reading the surface transmittance and the lateral footprint of the bridges back off that surface. The synthesis pass does the first, mapping an envelope geometry and its material layers to a temperature at every surface point. The diagnostic pass of Section 4 does the second, and because a thermal camera sees only the outer face of the envelope, it must recover both from that face alone. This section makes each object precise, the heat conduction the synthesis pass approximates, its differentiable surrogate, the surface observation, and the two surface constants, with notation in Appendix A.11.

Heat conduction in the envelope. An as-built envelope occupies a Lipschitz domain $\Omega \subset \mathbb{R}^3$ with boundary $\Gamma = \Gamma_{\mathrm{int}} \cup \Gamma_{\mathrm{ext}}$. Given a conductivity field $k(x) > 0$ and fixed indoor and outdoor temperatures $T_{\mathrm{in}}, T_{\mathrm{out}}$, the steady temperature T solves the elliptic boundary-value problem

$$\nabla \cdot (k \nabla T) = 0 \text{ in } \Omega, \quad -k\,\partial_n T = \tfrac{T - T_{\mathrm{in}}}{r_{si}} \text{ on } \Gamma_{\mathrm{int}}, \quad -k\,\partial_n T = \tfrac{T - T_{\mathrm{out}}}{r_{se}} \text{ on } \Gamma_{\mathrm{ext}}. \tag{1}$$

where r_{si}, r_{se} are the interior and exterior surface film resistances and ∂_n is the outward normal derivative, so that $-k\,\partial_n T$ is the heat flux leaving Ω across either face, negative on Γ_{int} where heat enters. We rescale to the dimensionless temperature

$$\theta \;=\; \frac{T - T_{\mathrm{out}}}{T_{\mathrm{in}} - T_{\mathrm{out}}} \;\in\; [0, 1], \tag{2}$$

which is invariant to the boundary temperatures and depends only on the geometry and k, so a result transfers across weather. The surface conditions in equation 1 are of Robin type, fixing the flux across the film, and the problem is uniformly elliptic and well posed, defining the forward solution operator

$$\theta \;=\; F(\Omega, k). \tag{3}$$

For a plain layered wall, with layers of thickness d_i and conductivity λ_i, equation 1 reduces to the one-dimensional profile

$$\theta_{\mathrm{prior}}(x) \;=\; \frac{R_{>}(x)}{R_{\mathrm{tot}}}, \qquad R_{\mathrm{tot}} \;=\; r_{si} + \sum_i \frac{d_i}{\lambda_i} + r_{se}, \tag{4}$$

with $R_{>}(x)$ the resistance between x and the exterior. We use θ_{prior} as an analytic reference throughout, exact on a clear wall and departing from the true field only near a bridge, where the flow stops being one-dimensional (ISo, 2017).

Surface observation. A thermal camera does not observe θ throughout Ω but only its values at the exterior surface, under a single fixed excitation. Writing $\mathcal{O}$ for the operator that restricts a field to that surface, the measurement is

$$y \;=\; \mathcal{O}[\theta] + \eta, \qquad \mathcal{O}[\theta] \;=\; \theta|_{\Gamma_{\mathrm{ext}}}, \tag{5}$$

with η a bounded sensor perturbation of scale σ_{obs}. In our experiments the envelope is a point cloud sampled through the wall, and the observation is the set of points within a slab of normalised through-wall width 0.08 at the exterior face, about 330 of the 4096 points, so no observed point lies

exactly on the face. Every inverse below, the optimization, the amortized networks, and the readout, sees only these points, and an interior-face variant covers internal thermography.

Assumption 1 (Single-excitation boundary observation)**.** The data y is the noisy exterior trace equation 5 under one fixed indoor-to-outdoor excitation, not the full Dirichlet-to-Neumann map. The noise η has zero mean, is independent of k and across observed points, and has bounded scale σ_{obs}. A perturbation shared by all points, such as an emissivity offset, lies outside this assumption and is not reduced by averaging.

Composing equation 3 with equation 5 gives the map from k to y, $G(\Omega, \cdot) = \mathcal{O} \circ F(\Omega, \cdot)$. Because y is a single surface trace rather than the full boundary map, the central question is whether k is determined by y at all, which Section 4 answers.

Target surface constants. The engineering value of an envelope is captured by two surface constants. The first, the surface transmittance, is a scalar functional of the observed surface field rather than a field over the interior, which is exactly the property that makes it recoverable.

Definition 1 (Surface constants)**.** The surface transmittance U, known in building physics as the U-value, and the linear thermal-bridge transmittance Ψ are

$$U \;=\; \frac{1}{|\Gamma_{\mathrm{ext}}|} \int_{\Gamma_{\mathrm{ext}}} \frac{\theta}{r_{se}}\, dS, \qquad \Psi \;=\; L^{2\mathrm{D}} - \sum_j U_j\, \ell_j, \tag{6}$$

where $L^{2\mathrm{D}}$ is the two-dimensional thermal coupling coefficient across a junction and $U_j \ell_j$ are the clear-wall contributions of the adjoining walls (ISo, 2017; ISO, 2017), U being the constant this paper recovers and Ψ defined for reference.

The surface transmittance U is much better posed than Ψ. From equation 6, U is a bounded linear functional of the surface temperature, a plain average of the observed trace, whereas Ψ is a difference of coupled multidimensional heat flows that also needs the excess flow attributed to one bridge among its neighbours. We recover U and the lateral footprint of the bridges, and we do not estimate Ψ.

Differentiable surrogate. The diagnostic stage of Section 4 differentiates through the forward map and runs across many envelopes, so it needs a forward model that is fast and differentiable. We replace the elliptic solve equation 1 with a surrogate that keeps the analytic prior and learns only a geometry-conditioned residual,

$$\hat{\theta} \;=\; \theta_{\mathrm{prior}} + \Delta_\phi(\Omega), \tag{7}$$

where Δ_ϕ maps the per-point geometry and material description of the envelope, its position, signed distance to the interfaces, film resistances, layer thicknesses and conductivities, and the analytic prior, to a scalar temperature correction on Γ_{ext}. We instantiate it as a geometry-conditioned neural operator following Li et al. (2023), which encodes the point cloud onto a latent grid with a graph neural operator, mixes globally with a Fourier neural operator (Li et al., 2020), and decodes back to the points, with transformer (Wu et al., 2024) and DeepONet (Lu et al., 2021) variants, trained by supervised regression against the validated finite-volume field. Predicting a residual on the analytic prior confines the learning problem to the departures from one-dimensional flow, which is where a thermal bridge lives. We use the surrogate primarily for its speed and differentiability. The proposed U-value readout is evaluated directly from the observation and therefore does not depend on surrogate accuracy.

4 METHOD

The diagnostic stage takes the surface temperature field, produced by the synthesis stage of Section 3 or measured by an infrared camera in deployment, and returns the surface transmittance together with the lateral footprint of the bridges. Figure 2 shows the full pipeline. We recover the U-value by integrating over the observed face and reading a physical identity rather than by reconstructing the interior, we locate the bridge footprint from the surface anomaly, with a learned amortized alternative, and we return the transmittance with an interval whose marginal coverage is guaranteed under exchangeability. The subsections below make each step precise, starting with why the interior field cannot be recovered while the surface transmittance can.

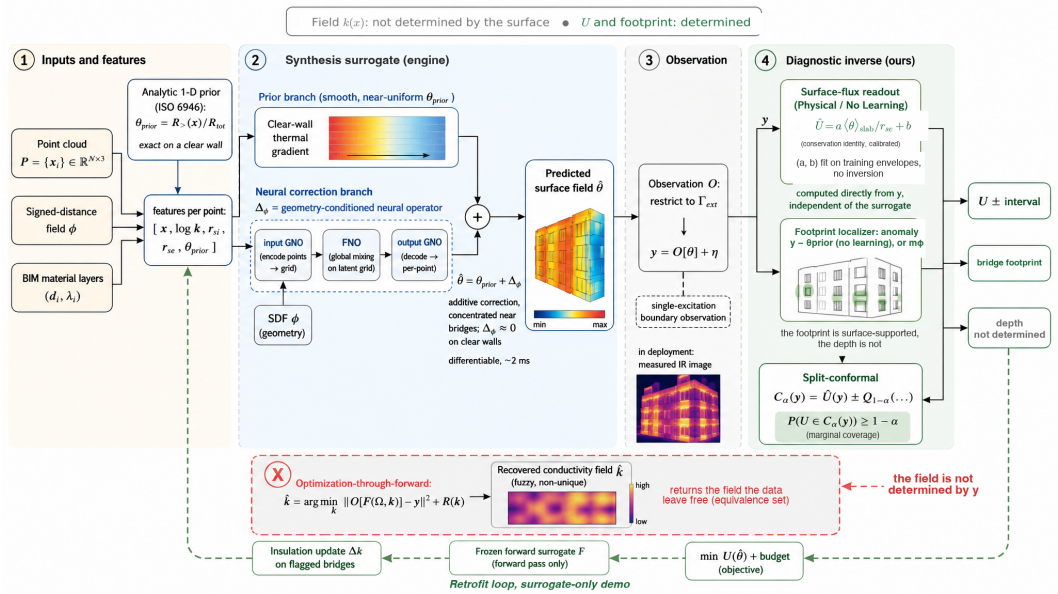


Figure 2: Method architecture. A per-point featurization of the point cloud, its signed-distance field, and its material layers feeds an additive forward surrogate for the surface temperature field. The diagnostic stage reads the U-value by a calibrated surface-flux identity, locates the bridge footprint from the surface anomaly, and returns a split-conformal interval. Optimization-through-forward (red) targets the field the data leave free, and the surrogate supports a retrofit demonstration (green).

Non-identifiability under boundary observation. Minimizing $\|\mathcal{O}[F(\Omega, k)] - y\|^2$, the natural way to recover k from the measurement $y = \mathcal{O}[F(\Omega, k)]$, cannot identify k from a single trace, because for a layered wall many different conductivity fields drive that objective to the same minimum and for a bridged wall a whole band of bridge depths drives it to within the noise. The coefficient is recoverable from a full characterization of the boundary but not from a single image of it, which we now make precise.

Proposition 1 (Non-identifiability under boundary observation)**.** *Under Assumption 1, for a layered wall the map $G(\Omega, \cdot)$ is not injective, because the exterior trace depends on k only through the total resistance $\int_0^L dx/k(x)$ and any rearrangement of layers with the same total leaves y unchanged. For a wall with a thermal bridge the trace depends on the bridge's depth only through a signal that diffusion damps, so the depth is determined only up to a band that widens as the bridge sits deeper. Appendix A.1 gives the explicit pair for the layered case, the measured depth bands for the bridged case, and the classical stability results both rest on.*

Diffusion is what places the field beyond reach, because it smooths interior structure before that structure reaches the surface, so a whole family of coefficient fields produce the same trace to within the noise and G acquires a large observational equivalence set $\mathcal{N}(k) = \{k' \mid \|G(\Omega, k') - G(\Omega, k)\| \leq \sigma_{\text{obs}}\}$ within which any inverse returns what its prior prefers. The surface constant survives this set. From equation 6, U is a linear functional of the observed trace, so it is constant on $\mathcal{N}(k)$ up to the noise, a perturbation η moving it by exactly its surface mean $\langle\eta\rangle/r_{se}$, which contracts as $\sigma_{\text{obs}}/\sqrt{N}$ over N independent observed points, against the logarithmic stability of field recovery in equation 11. *Average what the diffusion leaves behind and calibrate it, and do not score on what it erased.*

Surface-flux readout of the U-value. Energy conservation turns the definition of U into a direct readout, because every watt entering at the interior face must leave across the exterior face, so the mean exterior film flux equals the U-value. Integrating equation 1 over Ω and applying the film conditions gives

$$\int_{\Gamma_{\text{ext}}} \frac{\theta}{r_{se}}\, dS \;=\; \int_{\Gamma_{\text{int}}} \frac{1-\theta}{r_{si}}\, dS \;=\; L, \tag{8}$$

the thermal coupling coefficient of the section in W/K, the heat flow per kelvin of indoor-to-outdoor difference, so the surface average in equation 6 equals $L/|\Gamma_{\text{ext}}|$ with no inversion. On a sam-

pled point cloud no point lies exactly on the face, so the slab reads θ a little inside the wall where it exceeds its face value $\theta_{\mathrm{ext}} \approx r_{se}U$, and we calibrate the readout with two parameters, $\hat{U} = a\,\bar{\theta}_{\mathrm{slab}}/r_{se} + b$, fit by least squares on the training envelopes and applied unchanged to every U-value in this paper, read off the observation or off a recovered field, so every comparison is on one formula. We evaluate $\hat{U}$ from y directly, so it does not inherit the surrogate's error, and as an ablation we also train a network that regresses U from the observed slab.

Amortized bridge localization. Where a bridge sits laterally is surface-supported, the footprint on Γ_{ext} over which the anomaly $y - \theta_{\mathrm{prior}}|_{\Gamma_{\mathrm{ext}}}$ concentrates, and it is determined by the data even though the bridge's depth is not. We therefore score localization by the footprint, the projection of the bridge onto the observed face, recovered either by thresholding the anomaly itself with no learning or by a network m_ϕ that maps the observed slab to the conductivity at every point of the cloud, a point being called a bridge where the prediction exceeds the clear-wall value by a fixed margin. Trained once and applied in one forward pass, m_ϕ is far cheaper at deployment than a per-scene optimization, and what it recovers of the field is set by the prior it encodes. Optimization through the forward model, by contrast, solves

$$\hat{k} \;=\; \arg\min_{k} \;\big\|\mathcal{O}[F(\Omega,k)] - y\big\|^2 + R(k), \tag{9}$$

which fits the observed slab yet returns a k displaced within $\mathcal{N}(k)$, because it chases the field the data do not determine (Molinaro et al., 2023).

Distribution-free calibration. We equip the surface constant with a split-conformal interval. Given a calibration set $\{(y_i, U_i)\}_{i=1}^n$ and the readout $\hat{U}$, the interval and its guarantee are

$$C_\alpha(y) \;=\; \hat{U}(y) \;\pm\; Q_{1-\alpha}\big(\{\,|\hat{U}(y_i) - U_i|\,\}_{i=1}^n\big), \qquad \mathbb{P}\big(U \in C_\alpha(y)\big) \;\geq\; 1-\alpha, \tag{10}$$

where $Q_{1-\alpha}$ is the conformal quantile and the coverage is marginal, holding on average over exchangeable calibration and test data with no distributional assumption on the errors (Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023). The guarantee holds for any estimator, including a U-value read off a recovered field, so identifiability does not make the interval valid, and what a well-conditioned estimator buys is width, which we report for every route. Exchangeability holds by construction on our synthetic corpora and is stated plainly as an assumption on measured buildings.

5 Experiments

Datasets, metrics, and ground truth. Ground truth on the synthetic corpora is the temperature field throughout the envelope and the transmittance computed from it, produced by an in-house steady finite-volume conduction solver validated against the two two-dimensional reference cases of ISO 10211 (ISO, 2017) (Appendix A.2). Synthetic evaluation uses three bridge corpora, junction, hard, and box, each observed through the exterior slab of Section 3. For real geometry we use the TUM2TWIN city model (Wysocki et al., 2026), whose envelopes carry synthetic bridges with solver ground truth, and we corroborate against measured infrared from ThermoScenes (Hassan et al., 2024) and the TBBR rooftop benchmark (Mayer et al., 2023) (Appendix A.9). Metrics are a relative L_2 field error, the U-value mean absolute error in W/m^{2}K, footprint intersection-over-union, and conformal coverage and width. Every test statistic is computed over validation geometries, 40 on junction and 32 on hard and box, with errors averaged over five training seeds per geometry before a paired Wilcoxon test and a bootstrap interval.

Robustness under increasing ill-posedness. On a layered wall and a two-dimensional wall with a movable bridge, Tikhonov, total-variation, and Bayesian inversions (Tikhonov & Arsenin, 1978; Rudin et al., 1992; De Simon et al., 2018) recover the U-value to the noise floor yet return flaw depths spread across most of the wall (Appendix A.1). On the three corpora we vary the measurement noise, added independently at every observed point as a fraction of the indoor-to-outdoor contrast, and compare the direct readout against the U-value read, with the same calibrated readout, off the field recovered by optimization and off the field returned by the Transolver-architecture amortized inverse (Table 1, full sweeps in Appendix A.5). The readout has a lower mean error than optimization at

Table 1: Noise sweep under exterior-slab observation, averaged over five seeds per geometry. U-value errors compare the direct readout with the same calibrated readout applied to optimized and Transolver-recovered fields. Win rates are reported against optimization, Transolver, and the clear-wall baseline. Full field errors, p-values, and interval widths are in Appendix A.5.

Measurement noise	0	0.01	0.02	0.04
junction, U error (clear 0.299)	**0.039**/0.047/0.164	**0.043**/0.053/0.176	**0.053**/0.071/0.176	**0.081**/0.110/0.175
win-rate (%)	57/90/92	78/92/98	80/92/95	78/85/80
hard, U error (clear 0.098)	**0.029**/0.042/0.047	**0.031**/0.046/0.054	**0.036**/0.058/0.067	**0.049**/0.085/0.086
win-rate (%)	69/78/97	66/84/97	69/81/91	81/88/78
box, U error (clear 0.078)	**0.042**/0.054/0.038	**0.045**/0.067/0.037	**0.053**/0.087/0.042	**0.076**/0.149/0.052
win-rate (%)	66/62/81	75/34/78	84/38/66	91/19/31

every level on every corpus, significantly so at every level on the 120-geometry sweep and from the smallest nonzero noise on the 40-geometry sweep, where the zero-noise level gives $p = 0.07$, and its margin widens with noise as the optimized field's surface residual grows. On 120 junction geometries the residual grows from 0.11 to 0.48, the readout's error from 0.036 to 0.086 against 0.042 to 0.111 off the optimized field, the win-rate rises from 65% to 72% with p falling from 2.8×10^{-4} to 8.2×10^{-8}, while Table 1 reports the 40-geometry set. Against the Transolver inverse the readout wins at every level on junction and hard, but on box, where that inverse recovers the field (relative error 0.19 to 0.24), the inverse reads U better, 0.052 against 0.076 at the highest level ($p = 6.0 \times 10^{-5}$), where the readout has fallen to the clear-wall level of 0.078.

Generality across elliptic inverse problems. The field collapse is a property of boundary-observed inverse problems in general. We reproduce the optimization-versus-constant comparison on heat conduction, Darcy flow, and electrical-impedance tomography, each solved with an exact finite-difference operator and observed only at the boundary, over six independent draws of 24 random coefficient fields. On all three the field error rises past the norm of the field as the noise grows while the surface-constant error is at machine precision without noise and stays small with it, and the surface constant beats the field on every sample in every draw, a paired Wilcoxon p of 1.2×10^{-7} per draw (Appendix A.3).

Surface constants versus field recovery. On the three synthetic corpora every route is compared on the same target under the exterior-slab observation, on the enlarged validation sets of 120, 96, and 96 geometries with five training seeds, and Table 2 reports the original 40, 32, and 32 geometries with the same seeds, which agree with them to within 0.02 W/m^{2}K in every column and within 0.01 on the readout and the head. The U-value is recovered from the exterior face on every corpus. The calibrated readout reaches 0.036, 0.024, and 0.033 W/m^{2}K on junction, hard, and box against clear-wall guesses of 0.283, 0.079, and 0.059 that use no observation, at p below 10^{-7}, and it matches or beats the learned regression head trained on U labels, our ablation, at 0.044, 0.025, and 0.026. It beats the optimized field on every corpus, at $p \leq 2 \times 10^{-3}$, and every field route on junction and hard, while on box the two global-attention inverses that recover the field there read U better, 0.029 and 0.028 against 0.033 at $p \leq 0.01$. No route recovers the field from the exterior face of the junction corpus, all four landing between 0.68 and 0.94 where the same networks reached 0.26 to 0.46 under full-volume observation, while on the fin corpora the two global-attention inverses recover it, to 0.23 and 0.19 for the Transolver-architecture inverse and 0.26 and 0.23 for the DeepONet-architecture inverse, where the local PointNet2 inverse and optimization do not (Appendix A.4). What a learned inverse recovers of the conductivity is set by the prior its architecture and corpus encode, not by the observation, and the interior face constrains more of it (Appendix A.8).

Footprint, not depth. A single exterior observation retains useful information about a bridge's lateral footprint while providing substantially less information about its depth. Table 9 scores that footprint, the projection of the true bridge onto the observed face, against the footprint recovered by thresholding the surface anomaly with no learning and by each inverse. Volumetric overlap with the true bridge collapses from 0.74 under full-volume observation to 0.17 and 0.25 for our and the Transolver inverse on 120 junction geometries, so depth is not recovered. Against the true prism footprint on junction the surface anomaly itself scores 0.36 while the learned field inverses score at most 0.09 ($p = 1.6 \times 10^{-5}$ against the Transolver inverse), because they reconstruct the conductive structure their labels encode, a whole structural layer on this corpus, rather than the bridge, while on the fin corpora, whose labels are the fins, they localize the fins (Table 9). On the TUM2TWIN city-

Table 2: Same-target U-value comparison under exterior-slab observation, averaged over five seeds per geometry. All recovered-field estimates use the same calibrated readout as the direct method. Raw slab, learned-head, and clear-wall baselines are also shown. Field-recovery errors are reported in Table 4.

	U-MAE (W/m^{2}K), exterior slab				U-MAE off a recovered field		
Corpus	raw slab	**direct (ours)**	head	clear wall	opt.	Transolver	DeepONet
junction	0.098	**0.039**	0.045	0.299	0.046	0.166	0.217
hard	0.173	**0.029**	0.028	0.098	0.039	0.047	0.053
box	0.200	**0.042**	0.028	0.078	0.048	0.039	0.033

model shells (Wysocki et al., 2026), whose bridges are synthetic prisms inserted into real envelopes, the footprint is recovered by the surface anomaly (0.27 and 0.38 at LoD2 and LoD3) and better still by optimization through the forward model (0.38 and 0.48), which places its conductivity excess in the right columns while missing the depth (field error 0.99 and 1.03), and not by the amortized inverses (0.04 to 0.22), with volumetric overlap below 0.23 for every amortized route.

Calibrated uncertainty. We compare split-conformal intervals on every route at the same calibrated readout against the ensemble error bar of the optimization inverse (Appendix A.7). The ensemble bar covers the truth at one standard deviation about half the time, whereas every conformal interval attains its marginal coverage at the 0.90 target, so validity is not what the surface constant buys. At 90% the interval on the direct readout is 0.25, 0.24, and 0.38 W/m^{2}K wide on junction, hard, and box. On junction and hard it matches the optimized field that fits the slab, 0.26 and 0.24, as the residual bound predicts, it is three times narrower than off our amortized field on junction, 0.85, that does not, and the learned head is tightest there, at 0.20 and 0.13. On box the optimized-field route is wider, 0.47, while the amortized and Transolver routes that recover the field there are tighter than the readout, at 0.20 and 0.28, with the head at 0.36.

6 Discussion and Limitations

The result states what a single exterior observation of a wall determines. It determines the transmittance of the section, which the calibrated readout recovers on every corpus at 1.8 to 7.9 times the accuracy of the layer table alone on the enlarged validation sets and which no route through a recovered field reads better at zero noise on two of the three corpora, on which neither swept field route does at any noise level. It determines the lateral footprint of a bridge, which the surface anomaly locates on real junction geometry at 0.36 overlap on 120 geometries where learned field inverses reach 0.09. It does not determine the conductivity field or the depth of a bridge, and what an inverse returns for them is the prior it encodes, so scoring an inverse by field recovery under boundary sensing scores its prior, as does the U-value it reads.

Several boundaries mark the scope of this claim. The observation is a thin slab of a sampled point cloud at the face, not a dense thermal image. We recover the transmittance of a section and the footprint of its bridges, not the per-bridge transmittance Ψ. Absolute U-value recovery is established on the ISO 10211-validated corpora and not on measured buildings, the conformal guarantee is marginal under exchangeability, noise shared across pixels is not averaged away, the analysis is steady-state, the retrofit is a surrogate-only demonstration (Appendix A.6), and the identifiability we build on is classical, our contribution being its consequence for learned inversion.

7 Conclusion

We asked whether the coefficient field, the quantity learned inverses recover, is the right target under a single boundary measurement, and it is not. A single exterior observation determines the transmittance of the section and the lateral footprint of its bridges, and it does not determine the conductivity field or the depth of a bridge, which no inverse we tested recovers from the exterior face of the junction corpus, so scoring an inverse by field recovery scores its prior. The calibrated readout recovers the U-value at 1.8 to 7.9 times the accuracy of the layer table alone on the enlarged validation sets, and no route through a recovered field reads it better at zero noise on two of the three corpora, on which neither swept field route does at any noise level. Under partial boundary observation the surface constant is the right target.

AI USE STATEMENT

Generative AI tools were used for proofreading and language refinement, assistance with schematic figure creation, and limited support in organizing the paper structure. All scientific content, methodology, experiments, results, mathematical claims, and conclusions were developed and verified by the authors. The authors reviewed all AI-assisted content and take full responsibility for the final manuscript.

ETHICS STATEMENT

This work uses synthetic corpora generated by our own solver and publicly released building datasets, involves no human subjects and no personal data, and its intended use is the diagnosis of building heat loss, whose benefit is reduced energy use and emissions. We foresee no misuse of a method that estimates the thermal transmittance of a wall from its exterior temperature.

REPRODUCIBILITY STATEMENT

The synthetic corpora are generated by the scripts from fixed seeds (1337 for training, 99 for validation), with $150/40$ (junction), $96/32$ (hard), and $96/32$ (box) envelopes sampled at 4096 (2048 for box) points, solved by the finite-volume reference solver validated in Appendix A.2. The observation is the slab of points with normalised through-wall coordinate above 0.92 (exterior) or below 0.08 (interior). The forward surrogate is a residual PointNet++ operator trained for 200 epochs, the amortized inverses map the masked slab to the conductivity and are trained for 200 epochs with a mean-squared loss, the optimization inverse runs 400 steps in the inverse benchmark and armor runs and 300 steps in the noise sweeps, at learning rate 0.05 with sparsity, proximity, and smoothness weights 5×10^{-3}, 10^{-3}, and 5×10^{-3}, and the ensemble has five members. The U-value readout averages θ over the slab points within 0.02 of the face, divides by the film resistance, and applies the affine calibration fit by least squares on the training envelopes, whose coefficients are stored with every result. Noise is independent Gaussian per observed point with the stated fraction of the dimensionless range. Conformal intervals use repeated $50/50$ calibration-test splits of the validation geometries with the finite-sample quantile at $\alpha = 0.1$. Statistics are paired Wilcoxon tests and percentile bootstrap intervals over validation geometries, with errors averaged over the five training seeds $\{1337, 1, 2, 3, 4\}$ per geometry. The cross-problem study uses six independent random-field draws of 24 samples each. Five-seed runs on enlarged validation sets of 120, 96, and 96 geometries give calibrated readout errors of 0.036, 0.024, and 0.033 W/m^{2}K against clear-wall guesses of 0.283, 0.079, and 0.059. The July full-volume armor summary for the junction corpus was reconstructed from its job log after a file-name collision, and a full-volume control run with the current code reproduces its field errors to within 0.006. Code, configurations, the job manifest, and every result file will be released upon acceptance.

ACKNOWLEDGMENTS

REFERENCES

Giovanni Alessandrini. Stable determination of conductivity by boundary measurements. *Applicable Analysis*, 27(1-3):153–172, 1988.

Anastasios N Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023.

Alberto P Calderón. On an inverse boundary value problem. *Computational & Applied Mathematics*, 25(2-3):133–138, 2006.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

Lia De Simon, Marco Iglesias, Benjamin Jones, and Christopher Wood. Quantifying uncertainty in thermophysical properties of walls by means of bayesian inversion. *Energy and Buildings*, 177: 220–245, 2018.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. Gnot: A general neural operator transformer for operator learning. In *International conference on machine learning*, pp. 12556–12569. PMLR, 2023.

Mariam Hassan, Florent Forest, Olga Fink, and Malcolm Mielle. Thermonerf: Joint rgb and thermal novel view synthesis for building facades using multimodal neural radiance fields. *arXiv preprint arXiv:2403.12154*, 2024.

EN ISO. 10211; thermal bridges in building construction—heat flows and surface temperatures—detailed calculations. *International Organization for Standardization (ISO): Geneva, Switzerland*, 2017.

En ISo. 6946: 2017; building components and building elements—thermal resistance and thermal transmittance—calculation methods. *International Organization for Standardization: Geneva, Switzerland*, 2017.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113 (523):1094–1111, 2018.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, et al. Geometry-informed neural operator for large-scale 3d pdes. *Advances in Neural Information Processing Systems*, 36:35836–35854, 2023.

Chad Lieberman and Karen Willcox. Goal-oriented inference: Approach, linear theory, and application to advection diffusion. *siam REVIEW*, 55(3):493–519, 2013.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Niculae Mandache. Exponential instability in an inverse problem for the schrödinger equation. *Inverse Problems*, 17(5):1435–1444, 2001.

Zoe Mayer, James Kahn, Markus Götz, Yu Hou, Tobias Beiersdörfer, Nicolas Blumenröhr, Rebekka Volk, Achim Streit, and Frank Schultmann. Thermal bridges on building rooftops. *Scientific Data*, 10(1):268, 2023.

Roberto Molinaro, Yunan Yang, Björn Engquist, and Siddhartha Mishra. Neural inverse operators for solving pde inverse problems. *arXiv preprint arXiv:2301.11167*, 2023.

Luis Pérez-Lombard, José Ortiz, and Christine Pout. A review on buildings energy consumption information. *Energy and buildings*, 40(3):394–398, 2008.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W Battaglia. Learning mesh-based simulation with graph networks. *arXiv preprint arXiv:2010.03409*, 2020.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

John Sylvester and Gunther Uhlmann. A global uniqueness theorem for an inverse boundary value problem. *Annals of mathematics*, pp. 153–169, 1987.

Blanca Tejedor, Miquel Casals, Marta Gangolells, and Xavier Roca. Quantitative internal infrared thermography for determining in-situ thermal behaviour of façades. *Energy and Buildings*, 151: 187–197, 2017.

Andrej Nikolaevich Tikhonov and Vasiliy Yakovlevich Arsenin. Solutions of ill-posed problems. *mathematics of computation*, 32(144):1320, 1978.

Guoxiang Grayson Tong, Carlos A Sing Long, and Daniele E Schiavazzi. Invaert networks: A data-driven framework for model synthesis and identifiability analysis. *Computer Methods in Applied Mechanics and Engineering*, 423:116846, 2024.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer, 2005.

Jeffrey Wen, Rizwan Ahmad, and Philip Schniter. Task-driven uncertainty quantification in inverse problems via conformal prediction. In *European Conference on Computer Vision*, pp. 182–199. Springer, 2024.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. *arXiv preprint arXiv:2402.02366*, 2024.

Olaf Wysocki, Benedikt Schwab, Manoj Kumar Biswanath, Michael Greza, Qilin Zhang, Jingwei Zhu, Thomas Froech, Medhini Heeramaglore, Ihab Hijazi, Khaoula Kanna, et al. Tum2twin: Introducing the large-scale multimodal urban digital twin benchmark dataset. *ISPRS Journal of Photogrammetry and Remote Sensing*, 232:810–830, 2026.

Tao Zhong, Yixun Hu, Dongzhe Zheng, Aditya Sood, and Christine Allen-Blanchette. Neural field thermal tomography: A differentiable physics framework for non-destructive evaluation. *arXiv preprint arXiv:2603.11045*, 2026.

A Appendix

A.1 Non-identifiability, classical results and an explicit construction

Proposition 1 rests on classical results for elliptic inverse-coefficient problems. The full Dirichlet-to-Neumann map determines the coefficient (Calderón, 2006; Sylvester & Uhlmann, 1987), recovery from it is only logarithmically stable (Alessandrini, 1988; Mandache, 2001), so any reconstruction $\hat{k}$ obeys a worst-case bound

$$\sup_{k} \left\|\hat{k} - k\right\| \;\geq\; \frac{c}{\left(\log(1/\sigma_{\text{obs}})\right)^{s}} \tag{11}$$

for constants $c, s > 0$, and a single excitation is strictly less informative than that map. Neither statement by itself proves non-injectivity for a single trace, so we give the two cases separately. Our contribution is not the impossibility itself but its consequence, that field recovery scores a prior, and the method that acts on it.

Layered wall, exact. For steady one-dimensional conduction through layers the exterior surface temperature is $\theta_{\text{ext}} = r_{se}/R_{\text{tot}}$ with $R_{\text{tot}} = r_{si} + \int_0^L dx/k(x) + r_{se}$, so every conductivity profile with the same integral produces the identical trace. The controlled study of Section 5 realizes this on a two-layer wall of thickness 0.30 m. The true field and a twin with the insulation flaw moved from depth 0.18 m to 0.03 m differ by a relative L_2 distance of 1.36 and produce exterior traces whose maximum difference is exactly 0, with the same U-value of 0.471 W/m^{2}K. Tikhonov inversions under three priors and a total-variation inversion all recover U to the noise floor and return confidently wrong depths spread over 0.24 m of the wall, and only a Bayesian posterior reports the depth as unresolved, with a standard deviation of 0.088 m. This is non-injectivity of $G(\Omega, \cdot)$ exhibited, not assumed.

Bridged wall, effective. With a laterally localized bridge the trace does depend on the bridge's depth, but the depth signal is damped by diffusion. On a two-dimensional wall with a movable bridge, the set of depths whose trace matches a noisy observation to within twice the noise floor spans 0.04 m for a shallow bridge and 0.15 m for a deep one in a 0.30 m wall, while the ground-truth U and Ψ of the forward model vary by less than 8% across the whole depth range, a property of the solutions and not an estimate. This is the effective non-identifiability that the three-dimensional corpora inherit, and it is a property of the forward map's fibres under noise, so it binds every inverse regardless of its regularization.

A.2 Ground-truth solver validation

The synthetic corpora rely on an independent conduction solver, which we validate against the two two-dimensional reference cases of the ISO 10211 thermal-bridge standard (ISO, 2017) before using it as ground truth. The three-dimensional cases of the standard are not run. The first case has prescribed temperatures on every boundary and no surface films, and the solver matches the reference temperatures to within 0.010 K away from the singular corner, where the standard allows 0.1 K. The second case exercises the surface-film conditions the corpora use, and the solver matches the nine reference temperatures to within 0.039 K and returns a thermal coupling coefficient of 9.491 W/mK against the reference 9.5 W/mK. Against our production solver it reproduces the temperature field to a relative error near 10^{-15}. We use it as the reference for all synthetic corpora.

A.3 Cross-problem identifiability, full breakdown

Table 3 gives the per-noise-level detail behind the generality experiment, for each of the three problems over six independent draws of 24 random coefficient fields, with the mean and the standard deviation over draws. The surface constant beats the field on every sample of every draw at every level, so the paired Wilcoxon test returns its smallest attainable value at 24 samples, 1.2×10^{-7}, in every draw.

Table 3: Cross-problem identifiability, full per-level breakdown over six independent draws of 24 samples. Field error is a relative L_2 distance, the mean over draws with the standard deviation of the six draw means in parentheses, and surface-constant error a relative error, mean over draws. The win-rate is the fraction of samples on which the surface-constant error is below the field error, and the p-value is the per-draw paired Wilcoxon value, identical in every draw.

Problem	Noise	Field error	Surface-constant error	Win-rate	Wilcoxon p
Heat conduction	0	0.994 (0.003)	3.7×10^{-17}	100%	1.2×10^{-7}
	0.01	1.078 (0.004)	0.0304	100%	1.2×10^{-7}
	0.03	1.454 (0.015)	0.0911	100%	1.2×10^{-7}
	0.06	1.817 (0.018)	0.1823	100%	1.2×10^{-7}
Darcy flow	0	0.932 (0.007)	7.2×10^{-17}	100%	1.2×10^{-7}
	0.01	0.950 (0.006)	0.0163	100%	1.2×10^{-7}
	0.03	1.045 (0.011)	0.0489	100%	1.2×10^{-7}
	0.06	1.209 (0.016)	0.0977	100%	1.2×10^{-7}
Electrical-impedance tomography	0	0.748 (0.026)	0	100%	1.2×10^{-7}
	0.01	0.836 (0.026)	6.3×10^{-4}	100%	1.2×10^{-7}
	0.03	1.097 (0.025)	0.0019	100%	1.2×10^{-7}
	0.06	1.430 (0.032)	0.0038	100%	1.2×10^{-7}

A.4 Field recovery per route

Table 4 gives the relative L_2 error of the recovered conductivity for every route under the exterior-slab observation, beside the same routes with the full volume observed, so the effect of the observation model is explicit. The junction full-volume row is a control run with the current code and five seeds, which reproduces the July run to within 0.006, and the hard and box rows are the July five-seed runs. The same control gives the paired localization statement. With the interior visible the learned inverses recover the prism footprint at 0.32 to 0.34 where from the exterior slab they reach 0.06 to 0.08, while the anomaly's footprint is unchanged, 0.31 against 0.35.

Table 4: Field rel-L2 per route, geometry-level means over five seeds, exterior slab against full volume (the junction full-volume row is the geometry-level five-seed control with the current code, and the hard and box full-volume rows are the July pooled five-seed summaries over $n = 160$ geometry-seed rows).

	Field rel-L2, exterior slab				Field rel-L2, full volume, five seeds			
Corpus	opt.	PointNet2	Transolver	DeepONet	opt.	PointNet2	Transolver	DeepONet
junction	0.93	0.76	0.67	0.80	0.58	0.26	0.31	0.46
hard	1.00	0.68	0.23	0.26	0.77	0.20	0.18	0.24
box	0.57	0.71	0.19	0.23	0.50	0.20	0.21	0.20

A.5 Noise sweeps, full tables

Tables 5, 6, and 7 give the full sweeps behind Table 1, with the field error of each route, the surface residual of the optimized field, paired Wilcoxon p-values over geometries, and the width of the readout's 90% interval.

Table 5: Noise sweep on the junction corpus, exterior-slab observation, 40 validation geometries, errors averaged over five training seeds. Clear-wall error 0.299 W/m^{2}K.

Measurement noise	0	0.0025	0.005	0.01	0.02	0.04
Optimization, field rel-L2	0.923	0.932	0.946	0.971	1.021	1.095
Transolver inverse, field rel-L2	0.671	0.670	0.674	0.710	0.791	0.871
Optimization, surface residual	0.101	0.104	0.115	0.155	0.263	0.430
U off the optimized field	0.047	0.046	0.045	0.053	0.071	0.110
U off the Transolver-inverse field	0.164	0.162	0.162	0.176	0.176	0.175
U **direct calibrated readout (ours)**	**0.039**	**0.039**	**0.040**	**0.043**	**0.053**	**0.081**
Win-rate, ours over opt./Transolver/clear (%)	57/90/92	72/88/95	72/92/98	78/92/98	80/92/95	78/85/80
Wilcoxon p, ours vs optimization	6.6×10^{-2}	9.9×10^{-4}	2.2×10^{-2}	5.4×10^{-5}	2.4×10^{-5}	6.3×10^{-5}
Wilcoxon p, ours vs Transolver inverse	1.4×10^{-9}	1.9×10^{-9}	2.5×10^{-10}	4.6×10^{-10}	1.3×10^{-10}	1.3×10^{-7}
Conformal 90% width, ours	0.222	0.219	0.215	0.214	0.240	0.361

Table 6: Noise sweep on the hard corpus, exterior-slab observation, 32 validation geometries, errors averaged over five training seeds. Clear-wall error 0.098 W/m^{2}K.

Measurement noise	0	0.0025	0.005	0.01	0.02	0.04
Optimization, field rel-L2	0.979	0.987	0.993	1.002	1.015	1.059
Transolver inverse, field rel-L2	0.227	0.231	0.239	0.244	0.258	0.302
Optimization, surface residual	0.158	0.162	0.174	0.201	0.266	0.405
U off the optimized field	0.042	0.045	0.046	0.046	0.058	0.085
U off the Transolver-inverse field	0.047	0.048	0.046	0.054	0.067	0.086
U **direct calibrated readout (ours)**	**0.029**	**0.029**	**0.030**	**0.031**	**0.036**	**0.049**
Win-rate, ours over opt./Transolver/clear (%)	69/78/97	69/81/97	66/75/97	66/84/97	69/81/91	81/88/78
Wilcoxon p, ours vs optimization	3.1×10^{-2}	8.8×10^{-3}	6.5×10^{-3}	1.9×10^{-2}	1.7×10^{-3}	6.0×10^{-5}
Wilcoxon p, ours vs Transolver inverse	9.1×10^{-4}	4.3×10^{-4}	1.8×10^{-4}	9.5×10^{-7}	1.3×10^{-6}	3.1×10^{-6}
Conformal 90% width, ours	0.128	0.127	0.126	0.125	0.144	0.196

A.6 Differentiable retrofit

Because the twin is differentiable, we optimize a retrofit through it under a budget, acting on the true conductivity field and scoring on the surrogate, and a sparse targeted fix captures about 93% of the heat-loss reduction of the unconstrained optimum at about 28% of its intervention. This demonstrates the differentiable pipeline. A validated retrofit decision needs the recovered rather than the true field and a reference-solver evaluation, which we leave to future work.

A.7 Calibrated uncertainty, full comparison

Table 8 gives the full calibration comparison summarized in Section 5, under exterior-slab observation with one training seed. Coverage is the mean empirical coverage over repeated $50/50$ calibration-test splits of the validation geometries at $\alpha = 0.1$ with the finite-sample quantile, and width is the full interval width in W/m^{2}K. Every interval off a recovered field uses the same calibrated readout as the direct readout.

Table 7: Noise sweep on the box corpus, exterior-slab observation, 32 validation geometries, errors averaged over five training seeds. Clear-wall error 0.078 W/m^{2}K.

Measurement noise	0	0.0025	0.005	0.01	0.02	0.04
Optimization, field rel-L2	0.594	0.607	0.619	0.653	0.713	0.759
Transolver inverse, field rel-L2	0.191	0.192	0.193	0.196	0.207	0.239
Optimization, surface residual	0.093	0.117	0.148	0.220	0.327	0.439
U off the optimized field	0.054	0.058	0.056	0.067	0.087	0.149
U off the Transolver-inverse field	0.038	0.037	0.038	0.037	0.042	0.052
U **direct calibrated readout (ours)**	**0.042**	**0.042**	**0.043**	**0.045**	**0.053**	**0.076**
Win-rate, ours over opt./Transolver/clear (%)	66/62/81	69/66/81	66/59/81	75/34/78	84/38/66	91/19/31
Wilcoxon p, ours vs optimization	5.4×10^{-3}	1.8×10^{-3}	2.4×10^{-3}	7.7×10^{-4}	3.9×10^{-5}	1.1×10^{-6}
Wilcoxon p, ours vs Transolver inverse	3.0×10^{-1}	2.8×10^{-1}	4.4×10^{-1}	2.6×10^{-1}	2.4×10^{-1}	6.0×10^{-5}
Conformal 90% width, ours	0.275	0.273	0.272	0.269	0.262	0.336

Table 8: Calibrated uncertainty under exterior-slab observation, 90% target. Every split-conformal interval attains its marginal coverage. On junction and hard the direct readout and the optimized field that fits the observed slab have the same width and the learned head is tightest, while on box the optimized-field route is wider than the readout and the amortized and Transolver field routes, which recover the field there, are tighter. The ensemble spread of the optimization inverse has no guarantee and covers the truth at one standard deviation about half the time. Full width in W/m^{2}K.

	junction ($n = 40$)		hard ($n = 32$)		box ($n = 32$)	
Estimator	Cov.	Width	Cov.	Width	Cov.	Width
Ensemble spread, optimization inverse (1σ)	0.52	–	0.53	–	0.63	–
Conformal, raw slab readout	0.90	0.45	0.94	1.05	0.94	1.62
Conformal, calibrated direct readout (ours)	**0.90**	**0.25**	**0.94**	**0.24**	**0.94**	**0.38**
Conformal, off the optimized field	0.90	0.26	0.95	0.24	0.94	0.47
Conformal, off our amortized field	0.91	0.85	0.94	0.22	0.94	0.20
Conformal, off the Transolver-inverse field	0.91	0.83	0.94	0.31	0.94	0.28
Conformal, learned head (ablation)	0.90	0.20	0.95	0.13	0.94	0.36

A.8 Interior-face observation

The exterior face is the deployment case, a camera outside the building, and it is the primary observation of this paper, but the corpus bridges start at or near the interior face, so an interior-face observation carries more field information and we report it for internal thermography, geometry-level over five seeds. On junction, field recovery improves from the interior face, optimization from 0.93 to 0.61 and the Transolver inverse from 0.67 to 0.54, and the learned inverses begin to locate the true prism footprint, our PointNet2 inverse at 0.21 and the Transolver inverse at 0.16 against the anomaly detector's 0.30. The U-value is recovered from either face. The calibrated readout reaches 0.054 W/m^{2}K on junction, 0.014 on hard, and 0.011 on box from the interior face against 0.039, 0.029, and 0.042 from the exterior, all far inside the clear-wall guesses, the optimized field gives 0.060, 0.017, and 0.012, and on junction the 90% interval on the readout is 0.27 wide against 0.55 off the optimized field, the latter at its raw rather than calibrated readout. Across the interior-face noise sweep on junction the readout's error grows from 0.054 to 0.061 and the optimized-field route's from 0.060 to 0.093, the win-rate rises from 65% to 75% with p from 0.19 to 4.4×10^{-3}, significant from $\sigma = 0.02$ ($p = 0.023$). The face changes how much of the field the data constrain, not whether the transmittance is recoverable.

A.9 Measured-data validation

Measured infrared tests plausibility, not accuracy, because no dataset available to us pairs a thermal image with a volumetric or transmittance ground truth. On ThermoScenes (Hassan et al., 2024), calibrated facade infrared in degrees Celsius, a local smooth model of the measured field explains about 63% of its variance on one scene and the heat-loss sources recovered by our graph inverse overlap a threshold detector on the same residual at about six times the overlap expected of a random mask over seven scenes, which corroborates the inverse's localization against a simple rule and no more. On airborne infrared over the TUM2TWIN city model (Wysocki et al., 2026) the heat-loss anomalies land about 2.1 times more often on the buildings we model than off them. On the TBBR rooftop bridge benchmark (Mayer et al., 2023) our saliency rule enriches at 0.77 times chance, below random, a negative result we report as such. The TUM2TWIN localization corpora of Section 5 use the real building shells with synthetic bridges and solver ground truth.

A.10 FOOTPRINT LOCALIZATION

Table 9: Footprint intersection-over-union under exterior-slab observation, geometry-level mean over five seeds. Threshold labels mark every point whose conductivity exceeds the clear wall by a fixed margin, which on junction covers whole structural layers. The true prism footprint is the projection of the inserted bridges, stored for the junction corpus, and on the TUM2TWIN shells (20 geometries each) the threshold labels are exactly the inserted prisms.

Corpus, label	Anomaly only	PointNet2 (ours)	Transolver	DeepONet	Optimization
junction, threshold labels	0.29	0.54	0.52	0.51	0.36
junction, true prism footprint	0.35	0.08	0.06	0.06	0.15
hard, threshold labels	0.16	0.49	0.52	0.45	0.27
box, threshold labels	0.16	0.55	0.64	0.60	0.22
TUM2TWIN LoD2, prism labels	0.27	0.05	0.13	0.04	0.38
TUM2TWIN LoD3, prism labels	0.38	0.12	0.15	0.22	0.48

A.11 NOTATION

Table 10 collects the notation used throughout.

Table 10: Notation.

Symbol	Meaning
$\Omega,\ \Gamma = \Gamma_{\mathrm{int}} \cup \Gamma_{\mathrm{ext}}$	envelope domain and its interior and exterior boundary
$k(x)$	spatially varying thermal conductivity field
θ	dimensionless temperature equation 2, 0 outside and 1 inside
$r_{si},\ r_{se}$	interior and exterior surface film resistances
θ_{prior}	analytic clear-wall temperature equation 4
$F,\ \mathcal{O},\ G = \mathcal{O} \circ F$	forward map, surface observation, and their composition $k \mapsto y$
$\eta,\ y$	sensor noise and measurement equation 5
$\mathcal{N}(k)$	observational equivalence set of G within which k is unconstrained
$U,\ \Psi$	transmittance and linear thermal-bridge transmittance equation 6
L	thermal coupling coefficient of the section equation 8, in W/K
$\hat{U},\ (a, b)$	calibrated slab readout of U and its two calibration coefficients