

Learning to Reconstruct Temperature Field from Sparse Observations with Implicit Physics Priors

Shihang Li^{a,b,c}, Zhiqiang Gong^{b,c}, Weien Zhou^{b,c}, Yue Gao^a, Wen Yao^{b,c}

^a*MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai 200240, China*

^b*Defense Innovation Institute, Academy of Military Science, Beijing 10071, China*

^c*Intelligent Game and Decision Laboratory, Beijing 10071, China*

Abstract

Accurate reconstruction of temperature field of heat-source systems (TFR-HSS) is crucial for thermal monitoring and reliability assessment in engineering applications such as electronic devices and aerospace structures. However, the high cost of measurement acquisition and the substantial distributional shifts in temperature field across varying conditions present significant challenges for developing reconstruction models with robust generalization capabilities. Existing DNNs-based methods typically formulate TFR-HSS as a one-to-one regression problem based solely on target sparse measurements, without effectively leveraging reference simulation data that implicitly encode thermal knowledge. To address this limitation, we propose IPTR, an implicit physics-guided temperature field reconstruction framework that introduces sparse monitoring–temperature field pair from reference simulations as priors to enrich physical understanding. To integrate both reference and target information, we design a dual physics embedding module consisting of two complementary branches: an implicit physics-guided branch employing cross-attention to distill latent physics from the reference data, and an auxiliary encoding branch based on Fourier layers to capture the spatial characteristics of the target observation. The fused representation is then decoded to reconstruct the full temperature field. Extensive experiments under single-condition, multi-condition, and few-shot settings demonstrate that

Email addresses: lishh_07@sjtu.edu.cn (Shihang Li),
gongzhiqiang13@nudt.edu.cn (Zhiqiang Gong), weienzhou@outlook.com (Weien Zhou), yuegao@sjtu.edu.cn (Yue Gao), wendy0782@126.com (Wen Yao)

IPTR consistently outperforms existing methods, achieving state-of-the-art reconstruction accuracy and strong generalization capability.

Keywords: Temperature field reconstruction of heat-source systems, Implicit physics-guided, Cross-attention, Few-shot learning

1. Introduction

Accurate perception of temperature distribution plays a pivotal role in ensuring the operational safety, efficiency, and long-term reliability of modern engineered systems [1, 2, 3, 4, 5], particularly in electronic and aerospace where internal heat generation is substantial. With the ongoing trend toward higher integration and increased power density in electronic components [6], the resulting thermal behavior becomes more intricate and difficult to predict, presenting significant challenges for thermal monitoring and control. In many practical scenarios-such as printed circuit boards [7], honeycomb panels in spacecraft [8], satellite-mounted thermal components [9, 2], and launch vehicle [10]-only a limited number of temperature sensors can be deployed due to space, cost, or structural constraints. This sensor sparsity leads to a fundamental challenge: reconstructing the high fidelity temperature field of heat source systems from sparse point-wise measurements. This task, known as Temperature Field Reconstruction of Heat Source Systems(TFR-HSS), is inherently ill-posed [11], as the underdetermined nature of the observation makes it difficult to recover the full field through direct physical inversion.

Early efforts in TFR-HSS primarily adopted classical interpolation methods and traditional surrogate modeling techniques. Interpolation-based methods [12], such as linear, polynomial, and spline interpolation, estimate the temperature values in unsensed regions by fitting continuous functions over a discrete set of known sensor measurements. While these methods are computationally efficient and easy to implement, they neglect the intrinsic physical laws governing heat conduction, leading to poor accuracy and limited physical consistency. To address these limitations, traditional regression-based surrogate models have been introduced, including Kriging method [13], support vector regression [14], radial basis function networks [15]. These models can capture certain nonlinear relationships but still lack the representational capacity required for high-resolution or large-scale tasks. Shallow neural architectures, such as random vector functional-link (RVFL) networks [16] and GMDH networks [17, 18], have also been explored to enhance flexibility, yet

their limited depth and parameterization constrain their ability to serve as end-to-end approximators for complex thermal patterns.

In recent years, deep neural networks (DNNs) have emerged as powerful tools for physical field reconstruction [19, 20, 21, 22, 23, 24, 25], benefiting from their strong expressive capacity and end-to-end learning capabilities. By learning complex mappings from sparse monitoring measurements to global temperature field, DNNs have shown superior performance in handling non-linear dependencies and spatial heterogeneity. VoronoiCNN [22] incorporate nearest-neighbor interpolation to reorganize irregularly distributed monitoring data into structured grids, thereby making it compatible with convolutional neural network models for spatial field recovery. VoronoiUnet [26] develops a Unet-based [27] framework that incorporates Voronoi-based encoding and integrates uncertainty estimation, thereby improving the model’s reliability in handling sparse and irregular measurements. In parallel, physics-informed neural networks (PINNs) [28, 29, 30] incorporate governing physics by embedding partial differential equation (PDE) residuals and boundary conditions into the loss function, thus enforcing physical consistency and improving generalization under data-scarce conditions. Recently, operator learning [31, 32, 33, 34, 35] has emerged as a powerful paradigm for modeling mappings between infinite-dimensional function spaces. RecFNO [35] introduces three encoding schemes for sparse sensors and employs FNO [32] to reconstruct the corresponding flow and heat fields from sparse observations. By learning the underlying functional relationships—often through frequency-domain representations—these approaches enable resolution-independent modeling of systems governed by PDEs.

Nonetheless, despite these advances, most existing DNNs-based methods formulate temperature field reconstruction as a one-to-one regression problem: given sparse monitoring data, the model directly maps the input to the corresponding full field. Such formulation tends to overlook a potentially rich source of information—reference simulation data collected under similar thermal conditions—which may implicitly encode the spatial and physical regularities of the system. Effectively leveraging these priors can provide additional guidance and improve reconstruction accuracy, particularly under sparse observation scenarios where the inverse problem is inherently ill-posed.

To this end, we propose a novel implicit physics-guided temperature field reconstruction (IPTR) framework that conditions the reconstruction process on a reference pair comprising sparse measurements and its corresponding full-field distribution. By leveraging this prior information, the model is

encouraged to capture transferable physical patterns from thermally similar states, enriching the reconstruction process with inductive physical priors. This approach effectively bridges the gap between purely data-driven and physics-informed paradigms by embedding learned physical consistency through structural similarity. Specifically, our framework consists of three key components. First, the Sparse Field Encoding module transforms discrete sensor measurements into a continuous spatial representation through Voronoi-based interpolation, preserving the geometric structure of the sensing layout and ensuring spatial alignment. Next, the Dual Physics Embedding Module, which forms the core of our architecture, processes both the reference and the target inputs through two parallel branches: an Implicit Physics-Guided Branch that utilizes cross-attention to extract implicit physical knowledge from the reference pair, and an Auxiliary Encoding Branch that employs a Fourier-based encoder to capture the spatial characteristics of the target input. This dual-branch design enables the model to integrate global physical priors with local observation patterns into a unified representation. Finally, the fused features are passed to the Field Decoding Module, which reconstructs the full-resolution temperature field over the 2D spatial domain. This modular architecture enables IPTR to incorporate implicit physical priors from reference data, leading to improved reconstruction accuracy and enhanced generalization in multi-condition and few-shot transfer scenarios. Our main contributions can be summarized as follows:

- We propose IPTR, a novel framework that leverages reference simulation pairs—comprising sparse measurements and their corresponding full field—to embed implicit physical priors into the reconstruction process. By conditioning on such reference information, the model captures transferable thermal patterns, which significantly improve reconstruction fidelity in extreme regions and enhance generalization under limited data scenarios.
- To simultaneously leverage reference data and target monitoring data, we design a Dual Physics Embedding Module which consists of two branches: an implicit physics-guided branch using cross-attention to extract implicit physical priors from reference data, and an auxiliary encoding branch based on Fourier layers to preserve target field characteristics.
- Extensive experiments across single-condition, multi-condition, and few-

shot settings demonstrate that IPTR achieves state-of-the-art accuracy and strong generalization capability.

This paper is structured as follows: Section 2 introduces the temperature field reconstruction of heat-source systems and our motivation. Section 3 presents the proposed IPTR framework in detail, including the sparse field encoding, dual physics embedding module and temperature field decoding. Section 4 demonstrates the effectiveness of our method through experiments under single-condition, multi-condition, and few-shot settings. Finally, Section 5 concludes the paper with a summary of key findings.

2. Temperature Field Reconstruction of Heat-Source Systems(TFR-HSS)

2.1. Mathematical Definition

This work focuses on the heat-source systems comprising multiple internal heat transfer predominantly occurs via conduction within a two-dimensional spatial domain. To facilitate analysis while preserving generality, the system is modeled as a planar region containing localized heat source, each corresponding to a discrete component, such as an electronic component [19, 28]. Within this framework, the problem of Temperature Field Reconstruction for Heat Source Systems (TFR-HSS) involves inferring the full temperature distribution across the domain based solely on a limited number of temperature measurements obtained from strategically deployed sensors. Let $\Omega \subset \mathbb{R}^2$ denote a bounded spatial domain representing the layout of the HSS. Within this domain, there exists Λ distinct heat sources, each denoted by an intensity function $\phi_i(x, y)$, which characterizes the internal heat generation of the i -th source, as shown in Figure 1. The temperature field is represented by a scalar function $T(x, y)$, and the spatially varying thermal conductivity is denoted by λ . Under steady-state conditions, the temperature distribution satisfies the stationary heat conduction equation with internal heat sources, which takes the form of Laplace equation:

$$\frac{\partial}{\partial x} \left(\lambda \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(\lambda \frac{\partial T}{\partial y} \right) + \sum_{i=1}^{\Lambda} \phi_i(x, y) = 0, \quad (x, y) \in \Omega \quad (1)$$

subject to appropriate boundary conditions, which include the Dirichlet boundary condition (Dirichlet B.C.), the Neumann boundary condition (Neumann B.C.), and the Robin boundary condition (Robin B.C.). These can be

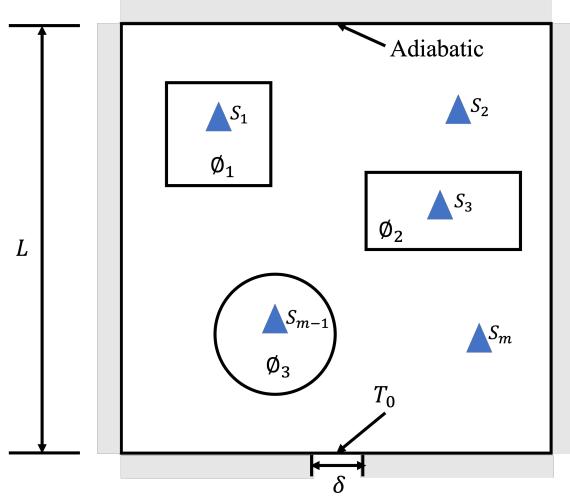


Figure 1: An illustration of the region Ω of the HSS subject to Dirichlet boundary condition.

expressed as:

$$T = T_0 \text{ (Dirichlet), or } \lambda \frac{\partial T}{\partial n} = 0 \text{ (Neumann), or } \lambda \frac{\partial T}{\partial n} = h(T - T_0) \text{ (Robin),} \quad (2)$$

where n is the outward normal on the boundary, and h is the convective heat transfer coefficient. The objective of TFR-HSS task is to estimate the full temperature distribution $T(x, y)$ over Ω using only sparse monitoring measurements collected from m sensors deployed at locations $\{(x_{s_i}, y_{s_i})\}_{i=1}^m$. The corresponding observed temperature values are denoted as $\{f(x_{s_i}, y_{s_i})\}_{i=1}^m$. This leads to the following inverse problem formulation:

$$T^* = \operatorname{argmin}_T \sum_{i=1}^m |T(x_{s_i}, y_{s_i} | \phi_1, \dots, \phi_\Lambda) - f(x_{s_i}, y_{s_i})| \quad (3)$$

subject to the Equation (1) and (2). This problem is inherently ill-posed, owing to the limited number of observations and the high dimensionality of the target temperature field. To tackle this challenge, we aim to construct a data-driven regression model that learns the underlying mapping from sparse monitoring measurements to the full temperature field. In the following, we provide a formal problem formulation and outline the motivation of our approach.

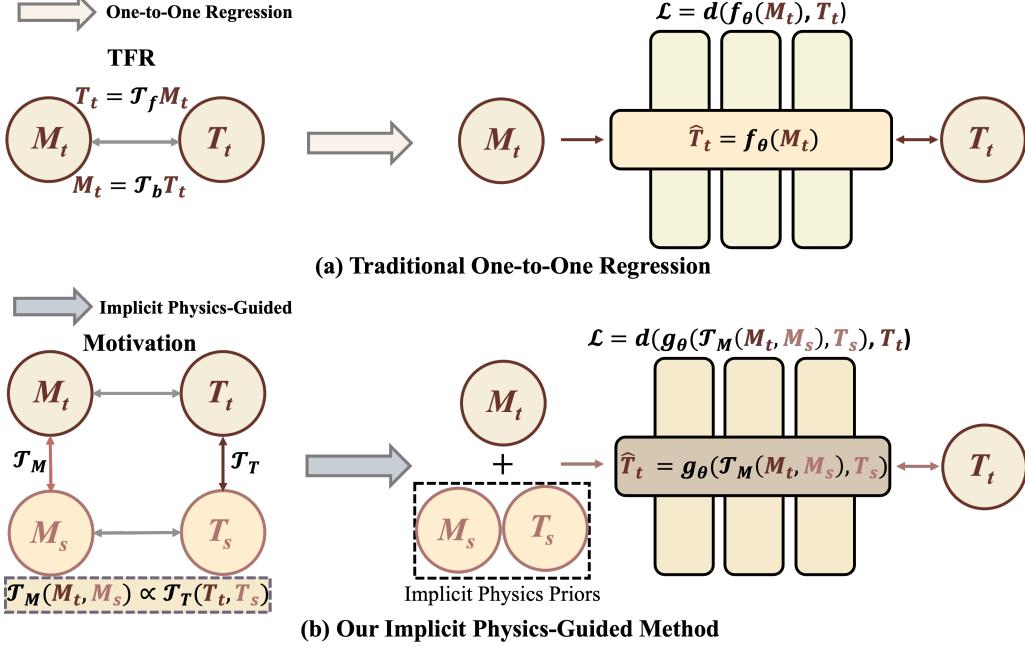


Figure 2: Overview of our reconstruction paradigm. (a) Existing methods directly regress the temperature field T_t from sparse measurements M_t , ignoring auxiliary information. (b) Our approach leverages a reference pair (M_s, T_s) from similar thermal conditions to provide implicit physics guidance. By modeling the relation between (M_t, M_s) and (T_t, T_s) , the network learns to reconstruct T_t via the mapping $g_\theta(M_t, M_s, T_s) \rightarrow \hat{T}_t$.

2.2. Problem Formulation and Motivation

The TFR-HSS involves predicting the full-field temperature distribution T_t from sparse monitoring measurements M_t . Notably, M_t can also be regarded as a sampled subset of T_t , i.e., $M_t = \mathcal{T}_b T_t$. Existing approaches typically treat this process as a one-to-one regression, learning a direct mapping from M_t to T_t via a surrogate model f_θ , as illustrated in Figure 2(a). While effective in controlled settings, such formulations ignore valuable prior knowledge available in practical scenarios, such as prior simulations data and their associated thermal conditions. To address this limitation, we propose a new paradigm that introduces an additional reference pair—sparse monitoring measurements M_s and its corresponding full-field temperature T_s under a similar thermal condition—to implicitly embed physical guidance into the reconstruction process. This strategy allows the model to exploit thermally relevant reference information under similar condition, providing a form of

implicit physics guidance that supplements the sparse input M_t .

Our motivation, shown in Figure 2(b), is based on the assumption that a transformation between M_t and M_s , denoted as $\mathcal{T}_M(M_t, M_s)$, should correspond to a similar transformation $\mathcal{T}_T(T_t, T_s)$ between the associated temperature fields. **This implicit proportional relationship enables us to reformulate the reconstruction task as a function of both the target and the reference data, rather than treating them in isolation.**

To this end, we define a neural network g_θ that models the mapping:

$$g_\theta : (M_t, M_s, T_s) \rightarrow \hat{T}_t, \quad (4)$$

where g_θ is parameterized by θ , and \hat{T}_t is the predicted temperature field. The model is trained to minimize the discrepancy between \hat{T}_t and the ground truth T_t .

3. Methodology

In this section, we begin by presenting an overview of the proposed framework, which takes as input the target sparse monitoring data together with a reference pair of sparse monitoring–temperature field as priors. These reference data serve as implicit physical priors that guide the reconstruction process and enhance the model’s generalization capability (Section 3.1). Section 3.2 introduces the Sparse Field Encoding module, where the sparse monitoring measurements are transformed into structured spatial representations using Voronoi-based interpolation. The core component of our approach is the Dual Physics Embedding Module (Section 3.3), which effectively integrates the implicit physical knowledge encoded in the reference data with the target observations, through two complementary branches: the Implicit Physics-Guided Branch and the Auxiliary Encoding Branch. Finally, the fused representation is passed to the Temperature Field Decoder, which reconstructs the full temperature field (Section 3.4).

3.1. Overview of Our Framework

As illustrated in Figure 3, the proposed IPTR framework aims to reconstruct the full temperature field from sparse monitoring measurements by leveraging reference data from a similar thermal condition. The overall architecture is composed of three key components: the Sparse Field Encoding module, the Dual Physics Embedding module, and the Temperature

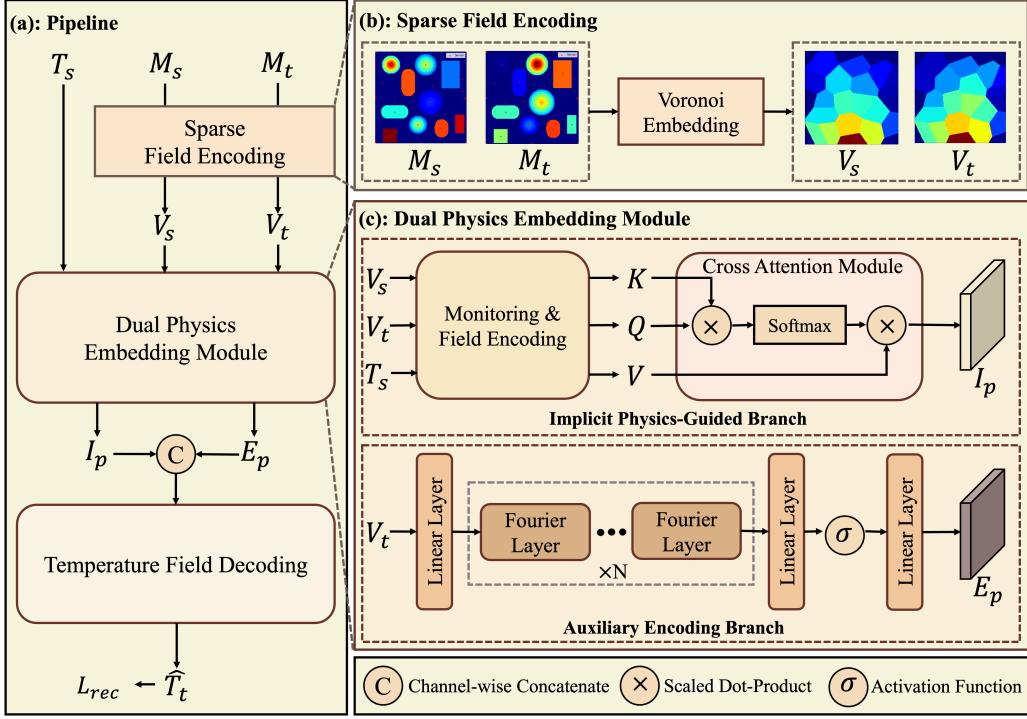


Figure 3: Architecture of the proposed IPTR framework. (a) shows the overall pipeline for temperature field reconstruction. Given a reference pair consisting of sparse monitoring data M_s , and corresponding temperature field T_s , along with target sparse monitoring data M_t , we first obtain interpolated field V_s , V_t via Voronoi-based encoding (b). These are passed to the Dual Physics Embedding Module (c), where the Implicit Physics-Guided Branch performs cross-attention among V_s , T_s and V_t to extract I_p , and the Auxiliary Encoding Branch encodes V_t into E_p . The fused representation $I_p \odot E_p$ is decoded to reconstruct the full temperature field \hat{T}_t , supervised by the ground-truth T_t .

Field Decoder. Given the target sparse monitoring temperature value and a selected pair of reference sparse monitoring-temperature field data, the framework first performs Voronoi-based interpolation to transform both the target and reference monitoring data into spatially structured fields. These interpolated fields are the passed into the Dual Physics Embedding module, which constitutes the core of our design. It consists of two parallel branches: the Implicit Physics-Guided Branch, which encodes reference information via a cross-attention mechanism to provide implicit physical priors; and the Auxiliary Encoding Branch, which encodes the target monitoring data using a Fourier-based encoder to enrich the learned representation. The features

from both branches are fused and subsequently fed into the Temperature Field Decoder, which maps the unified representation to a high fidelity temperature field reconstruction. By jointly utilizing reference physical patterns and target observations in a unified learning framework, IPTR enhances reconstruction accuracy and spatial coherence by implicitly leveraging physical priors, and enables effective adaptation to unseen scenarios through few-shot fine-tuning.

3.2. Sparse Field Encoding

The sparse and irregular distribution of sensors across the spatial domain introduces challenges for structured modeling of physical fields. Directly using raw monitoring data as model input often leads to poor spatial correlation modeling, as the lack of consistent spatial structure hinders effective information propagation and attention computation.

To address this, we employ a Voronoi-based interpolation strategy [22, 35] to transform the sparse monitoring measurements into structured grid-aligned representations. Let the target sparse monitoring measurements set be denoted as $M_t = \{(p_i^t, m_i^t)\}_{i=1}^{N_t}$, where $p_i^t \in \Omega \subset \mathbb{R}^2$ is the spatial location of the i -th sensor and $m_i^t \in \mathbb{R}$ denotes its corresponding temperature value. Here, Ω is discretized into a uniform 2D grid $\mathcal{G} = \{(x_i, y_i) | i = 1, \dots, H; j = 1, \dots, W\}$, which defines the spatial resolution of the reconstructed field.

To propagate sparse monitoring measurements to the full grid, we partition Ω using a Voronoi tessellation induced by the sensor locations $\{(p_i^t)\}_{i=1}^{N_t}$. Each grid point $(x_i, y_i) \in \mathcal{G}$ is assigned the temperature value of its nearest sensor in Euclidean distance, i.e.,

$$V_t(i, j) = m_k \quad \text{where } k = \arg \min_{1 \leq \ell \leq N_t} \|(x_i, y_j) - \mathbf{p}_\ell^t\|_2 \quad (5)$$

This process yields a fully populated 2D pseudo-field $V_t \in \mathbb{R}^{H \times W}$, referred to as the Voronoi-interpolated temperature map. A similar procedure is applied to the reference measurements M_s , resulting in $V_s \in \mathbb{R}^{H \times W}$. Figure 4 illustrates this process, where sparse and irregular measurements are first assigned to a uniform spatial grid via Voronoi tessellation, producing a structured pseudo-field that approximates the underlying temperature distribution. This transformation improves spatial continuity while preserving the sensor layout, enabling grid-aligned model inputs. The structured representations (V_t, V_s) facilitate subsequent cross-attention computation by aligning both the target and reference monitoring measurements to a common spatial

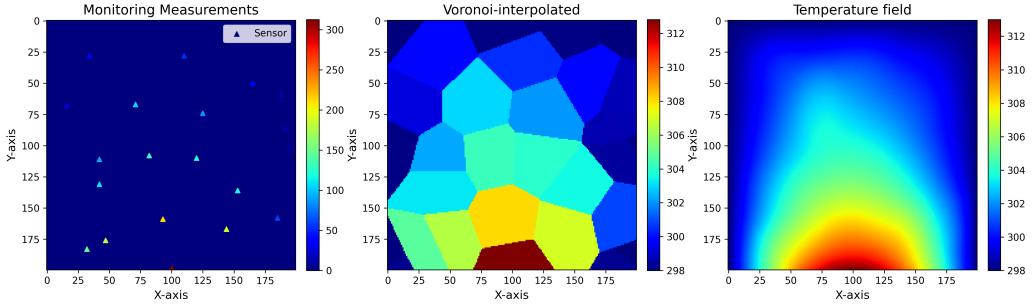


Figure 4: Voronoi-based pseudo-field generation.

grid. Compared to unordered sensor inputs, Voronoi interpolation enhances spatial consistency and interpretability, ultimately improving the model’s ability to infer underlying physical patterns under sparse sensing conditions, as discussed in Section 3.3.

3.3. Dual Physics Embedding Module

While the Voronoi-based interpolation provides a structured representation of the sparse monitoring data, it alone lacks the ability of capture deeper physical dependencies and correlations necessary for accurate reconstruction. To address this, we propose the Dual Physics Embedding Module, which forms the core of our framework. Its key motivation is to leverage reference sparse monitoring-temperature field as implicit physical priors, enabling the model to reason beyond the target sparse monitoring measurements and generalize better in regions with complex thermal variations.

As illustrated in Figure 3(c), this module consists of two complementary branches designed to encode both the target interpolated field and reference information. The resulting embeddings are fused to form a unified representation for temperature field reconstruction. In the following, we detail the designs of the Implicit Physics-Guided Branch and the Auxiliary Encoding Branch.

3.3.1. Implicit Physics-Guided Branch

The goal of this branch is to inject implicit physical priors derived from a reference pair into the reconstruction process. Rather than explicitly modeling physical laws, we encode these priors by utilizing a reference sparse

monitoring-tempearature field pair (M_s, T_s) , which captures thermal behaviors under similar thermal conditions.

To model the correlation between the target sparse measurements M_t and the reference measurements M_s —as conceptualized in Figure 2(b)—we employ a cross-attention mechanism. Specifically, we assume that the similarity between M_t and M_s can guide the transfer of information from T_s to assist in reconstructing the target temperature field T_t . In practice, this is implemented via a single scaled dot-product attention [36] operating on the target interpolated field V_t as the query, and the reference interpolated field V_s and full field T_s as the key and value, respectively. The attention operation is defined as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{Q \times K^T}{\sqrt{D}}\right) \times V \quad (6)$$

where Q , K and V denote the queries, keys and values, and D is a scalar scaling factor equal to the feature dimension. By computing attention in this way, the model learns to dynamically associate target observations with relevant reference patterns in a data-driven yet physically grounded manner, effectively enhancing its reconstruction capability, particularly in spatial regions with complex thermal variations.

Considering computational resources and efficiency, we apply attention in the latent space, rather than the pixel space. To this end, we employ a monitoring encoder \mathcal{E}_m to encode $V_s \in \mathbb{R}^{H \times W}$ and $V_t \in \mathbb{R}^{H \times W}$ into $\tilde{V}_s \in \mathbb{R}^{C \times \frac{H}{4} \times \frac{W}{4}}$ and $\tilde{V}_t \in \mathbb{R}^{C \times \frac{H}{4} \times \frac{W}{4}}$, and a temperature field encoder \mathcal{E}_{tf} to encode $T_s \in \mathbb{R}^{H \times W}$ into $\tilde{T}_s \in \mathbb{R}^{C \times \frac{H}{4} \times \frac{W}{4}}$. Note that these encoders are all UNet-based [27] encoders.

We assign the latent of \tilde{V}_t as the query because it represents the target to be reconstructed—it is where the model should attend to relevant information. The latent of reference monitoring data \tilde{V}_s is used as the key, providing a spatial anchor from reference measurements. Finally, the latent of reference temperature field \tilde{T}_s serves as the value, containing rich physical structures that can be transferred to guide the reconstruction. Note that Q , K and V are all reshaped from $(C, \frac{H}{4}, \frac{W}{4})$ to $(\frac{H}{4} \times \frac{W}{4}, C)$, we also add positional embeddings to Q and K . Finally, the implicit-physics guided representation is formulated as:

$$I_p = \text{Attention}(Q, K, V) \quad (7)$$

As illustrated in Figure 3(c), this branch provides the core physical guidance

for our model. The resulting embedding I_p will be later fused with auxiliary features for downstream reconstruction.

We next introduce the Auxiliary Encoding Branch, which complements this component by providing structural enhancement.

3.3.2. Auxiliary Encoding Branch

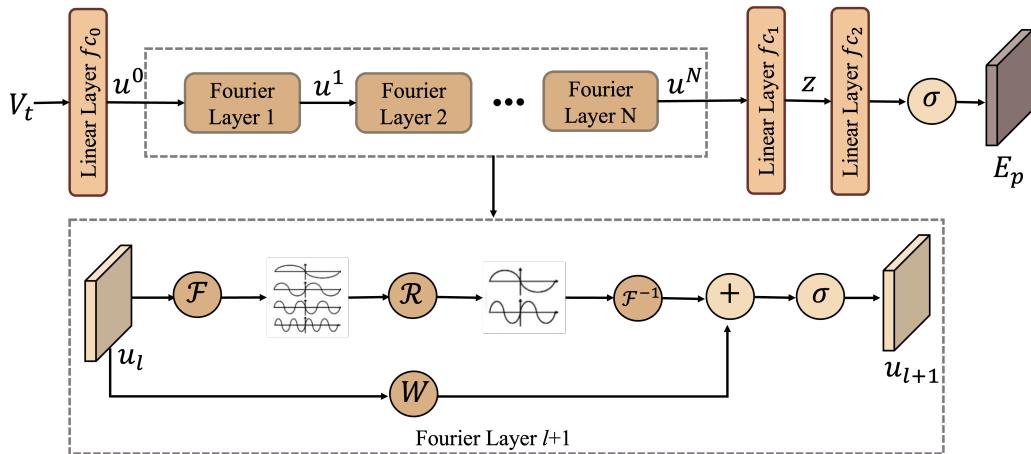


Figure 5: Architecture of the Auxiliary Encoding Branch. The target sparse interpolated field V_t undergoes a dimensional lifting, followed by spectral transformation through Fourier layers, and is subsequently projected to obtain auxiliary embeddings E_p for enhancing the reconstruction process.

To complement the implicit physics-guided reconstruction, we design an Auxiliary Encoding Branch that enhances the structural representation of the target input. Although not directly involved in physics prior extraction, this branch encodes the interpolated sparse field $V_t \in \mathbb{R}^{H \times W}$ into an auxiliary representation $E_p \in \mathbb{R}^{1 \times H \times W}$, which serves as a complementary source of spatial information. By capturing non-local correlations in the spectral domain, it enriches the feature space and promotes more robust and spatially consistent reconstruction.

As illustrated in Figure 5, this branch is based on a residual Fourier Neural Operator (FNO) [32] structure, implemented using a stack of Fourier-based convolutional layers interleaved with residual connections and non-linear activations. The processing begins with a point-wise linear transformation which lift the input V_t from a single-channel field into a high-dimensional

latent space:

$$u^{(0)} = fc_0(V_t), \quad u^{(0)} \in \mathbb{R}^{C' \times H \times W} \quad (8)$$

where C' denotes the channel dimension of the lifted space. This lifting step enables the subsequent spectral operations to operate in a richer representational space.

Following the lifting, the latent feature map is successively updated through N Fourier layers. In each layer, the update rule is defined as:

$$u^{(l+1)}(x) = \sigma(Wu^{(l)}(x) + \mathcal{K}(u^{(l)})(x)) \quad (9)$$

where $\sigma(\cdot)$ denotes the non-linear GELU activation function, W is a learnable 1×1 convolution representing a local linear transformation, and \mathcal{K} is a parameterized global operator defined in the frequency domain.

Specifically, the global operator \mathcal{K} is formulated as a kernel integral operator:

$$\mathcal{K}(u)(x) = \mathcal{F}^{-1}(\mathcal{R} \cdot \mathcal{F}(u))(x), \quad (10)$$

where \mathcal{F} and \mathcal{F}^{-1} denote the 2D Fourier transform and its inverse, and $\mathcal{R} \in \mathbb{C}^{F_1 \times F_2 \times C' \times C'}$ is a complex weight that operates as a spectral filter in the frequency domain, modulating selected low-frequency components (F_1, F_2). For implementation efficiency, we truncate the spectrum and retain only the lowest F_1 and F_2 modes (i.e., low-pass filtering). The forward Fourier transform is performed as:

$$\hat{u}^{(l)}(k_1, k_2) = \sum_{x_1=0}^{H-1} \sum_{x_2=0}^{W-1} u^{(l)}(x_1, x_2) \cdot e^{-2\pi i (\frac{x_1 k_1}{H} + \frac{x_2 k_2}{W})} \quad (11)$$

while the inverse transform is:

$$u'^{(l)}(x_1, x_2) = \sum_{k_1=0}^{F_1-1} \sum_{k_2=0}^{F_2-1} (\mathcal{R} \cdot \hat{u}^{(l)})(k_1, k_2) \cdot e^{2\pi i (\frac{x_1 k_1}{H} + \frac{x_2 k_2}{W})} \quad (12)$$

The multiplication in the frequency domain is defined element-wise for each mode (k_1, k_2) and each output channel $i \in \{1, \dots, C'\}$:

$$(\mathcal{R} \cdot \hat{u}^{(l)})_{k,i} = \sum_{j=1}^{C'} R_{k,i,j} \cdot \hat{u}_{k,j}^{(l)} \quad (13)$$

where $k = (k_1, k_2)$. After N such Fourier layers ($N=4$), the resulting representation $u^{(N)} \in \mathbb{R}^{C' \times H \times W}$ is first mapped back to a lower dimensional channel space through two fully connected layers:

$$z = fc_1(u^{(N)}), \quad \tilde{E}_p = fc_2(\sigma(z)) \quad (14)$$

To ensure compatibility with the latent resolution of the implicit physics-guided branch (e.g., due to spatial downsampling by the encoder), we apply a spatial downsampling via max pooling:

$$E_p = \text{MaxPool}_{4 \times 4}(\tilde{E}_p), \quad \text{where } \tilde{E}_p \in \mathbb{R}^{1 \times H \times W}, E_p \in \mathbb{R}^{1 \times \frac{H}{4} \times \frac{W}{4}} \quad (15)$$

This projection ensures the dimensional alignment with downstream modules and allows the auxiliary output E_p to be effectively fused with the implicit physics-guided features I_p during decoding. The entire architecture thus learns to represent spatial patterns in both physical and frequency domains, offering complementary guidance to the reconstruction task.

3.4. Temperature Field Decoding

The final stage of our framework aims to reconstruct the full temperature field from the learned physics-aware representations. To achieve this, we employ a decoding module that integrates both branches of the Dual Physics Embedding Module, effectively combining implicit physics prior knowledge and auxiliary Fourier-based features for high-fidelity reconstruction.

Specifically, we concatenate the implicit physics-guided features $I_p \in \mathbb{R}^{C \times \frac{H}{4} \times \frac{W}{4}}$ and the auxiliary encoded features $E_p \in \mathbb{R}^{1 \times \frac{H}{4} \times \frac{W}{4}}$ along the channel dimension, resulting a unified representation:

$$F_p = \text{Concat}(I_p, E_p) \in \mathbb{R}^{(C'' \times \frac{H}{4} \times \frac{W}{4})} \quad (16)$$

To decode this fused representation into the final temperature reconstruction, we adopt a SPADE decoder [37], which has shown strong performance in structured generation tasks due to its ability to incorporate spatial conditioning in a learnable, dynamic manner.

Formally, let $\mathcal{D}_{\text{SPADE}}$ denote the SPADE decoder. The final reconstructed temperature field \hat{T}_t is given by:

$$\hat{T}_t = \mathcal{D}_{\text{SPADE}}(F_p) \quad (17)$$

where $\hat{T}_t \in \mathbb{R}^{1 \times H \times W}$ corresponds to the model’s prediction of the dense thermal field under current observation.

By integrating both I_p and E_p into a unified decoding process, the framework benefits from complementary information sources. The representation I_p encodes implicit physical priors derived from reference pair through cross-attention, capturing spatial correlations that reflect underlying physical behaviors. In contrast, E_p provides auxiliary spatial structure information extracted from the target interpolated sparse observations via a Fourier-based encoder, serving as a complementary cue to enhance local consistency and improve robustness under measurement sparsity. This design enables the decoding module to leverage both implicit physics knowledge and target structural features in a unified reconstruction pipeline.

4. Experiments

4.1. Experiment setups

4.1.1. Experimental Objectives

To comprehensively evaluate the performance and generalization capability of the proposed IPTR framework, we conduct experiments under two categories of settings:

Benchmark Evaluation: We first evaluate IPTR on the TFRD benchmark dataset [19], which consists of various heat source configurations. Under this setting, we conduct two experiments:

- Single-condition reconstruction, where both training and testing are performed on the same heat configuration;
- Multi-condition reconstruction, where the model is trained on multiple heat configurations and tested on held-out ones within the same dataset.

Generalization to New Scenario: To evaluate the model’s generalization ability to previously unseen configurations, we introduce a new heat configuration as a target domain. In this setting: we perform:

- Sample-efficiency comparison, where we compare IPTR’s reconstruction performance with other models trained from scratch using varying numbers of samples;

- Few-shot transfer learning, where IPTR, pre-trained on the benchmark dataset, is fine-tuned using a limited number of labeled samples from the new scenario;
- Sensor sparsity analysis, where we evaluate the reconstruction performance of IPTR trained from scratch under different number of sensors in the new scenario.

4.1.2. Datasets

In this work, we employ two categories of datasets to evaluate the proposed method: the TFRD benchmark dataset and a newly constructed scenario dataset. Both datasets are generated based on the governing equation and boundary conditions defined in Equation (1) and (2), simulating steady-state heat conduction in a $0.1\text{m} \times 0.1\text{m}$ satellite panel. The temperature field is discretized into a 2D grid of resolution 200×200 , representing fine-grained thermal distributions across the panel.

To characterize typical thermal behaviors in HSS, two representative power distribution settings are widely used in engineering scenarios: uniformly distributed and ununiformly distributed heat sources. In the uniform setting, each heat-generating component maintains a constant power density across its surface, resulting in a spatially uniform thermal load. In contrast, the ununiform setting captures spatially varying heat generation, often modeled using a Gaussian distribution, where the power density peaks at the component center and decays towards the boundary. This can be formally expressed as:

$$\phi_i(x, y) = \begin{cases} Q_i \exp\left(-\sigma \frac{(x - x_0)^2 + (y - y_0)^2}{r^2}\right), & (x, y) \in \Omega_i, \\ 0, & \text{otherwise.} \end{cases} \quad (18)$$

where Q_i is the peak power, (x_0, y_0) is the center of the i -th component, r denotes the radius, and σ is the deviation coefficient.

The TFRD benchmark consists of three representative sub-tasks: HSink, ADlet and DSine, each defined by specific boundary conditions and power configurations: HSink applies a Dirichlet boundary with a fixed temperature of 298K on one side (heat sink) and Neumann (adiabatic) conditions on the remaining sides, with uniform power sources; ADlet imposes Dirichlet boundaries on all sides, where one side follows a sine-wave temperature distribution and others remain constant, using Gaussian-distributed heat sources; DSine

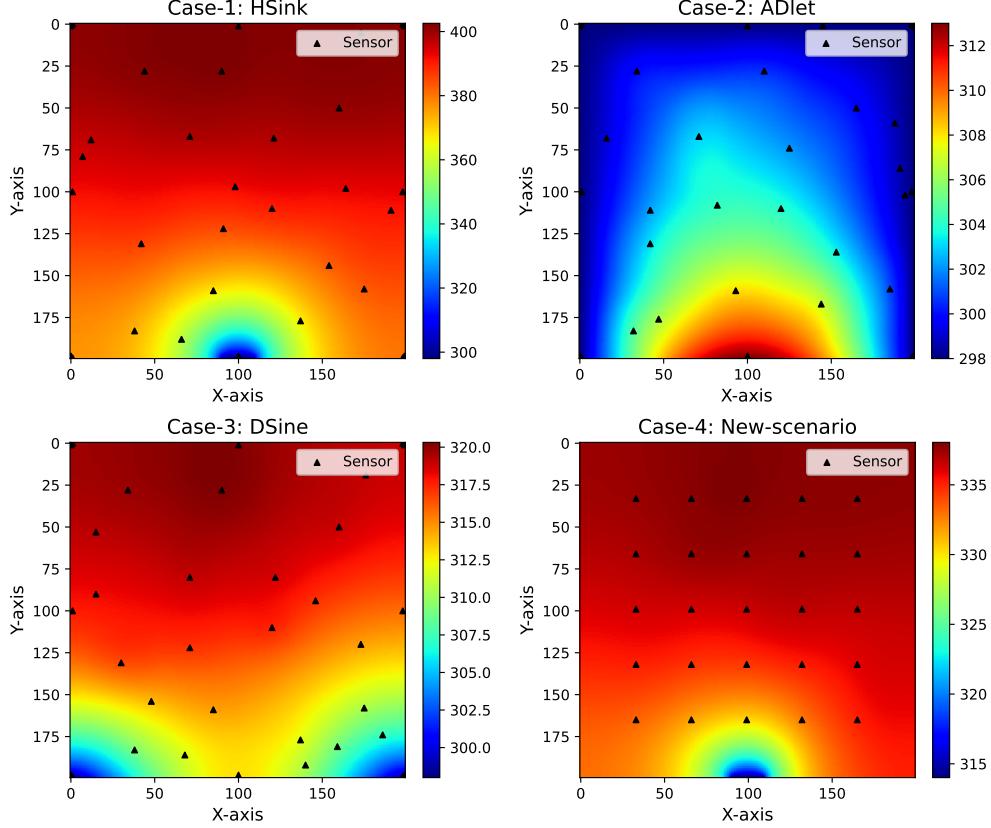


Figure 6: Visualization of four datasets used in this paper

adopts a single sine-wave Dirichlet boundary with the other three sides being adiabatic, also with Gaussian power sources. In addition, we construct a new scenario dataset with Gaussian heat sources and a sink-type Dirichlet boundary, the thermal conductivity is modeled as temperature-dependent: $\lambda = 1 + 0.05 \times (T - 298)$. Furthermore, 25 temperature sensors are uniformly placed across the spatial domain to reflect sparse real-world monitoring. Visualizations of the four datasets are shown in Figure 6.

4.1.3. Implementation and Evaluation Metrics

Implementation Details: All models are implemented using the PyTorch framework and trained on a single NVIDIA 3090 GPU. In the benchmark evaluation experiments, the model is trained from scratch for 150 epochs with a batch size of 16. For few-shot adaptation in the new scenario, fine-tuning

is conducted for 100 epochs. The batch size is set to 1 in few-shot fine-tuning experiments and remains 16 in all other settings. The Adam optimizer [38] is employed with an initial learning rate of 1.5e-4. A multi-step learning rate scheduler (MultiStepLR) [39] is used, where the learning rate is decayed by a factor of 0.1 at predefined epoch milestones. Specifically, the milestone is set to 100 for training from scratch, and 70 for fine-tuning.

Evaluation Metrics: To quantitatively evaluate the performance of our model in temperature field reconstruction, we adopt the following two commonly used metrics [9]:

- Mean Absolute Error (MAE). The MAE measures the average absolute deviation between the predicted temperature field \hat{T}_t and the ground truth field T_t , and is defined as

$$\text{MAE} = \frac{\|T_t - \hat{T}_t\|}{|T_t|} \quad (19)$$

where $|T_t|$ denotes the total number of points in the field.

- Maximum Absolute error (Max-AE). The Max-AE reflects the worst-case deviation and is especially relevant for evaluating safety-critical systems involving extreme temperatures.

$$\text{MAE-AE} = \|T_t - \hat{T}_t\|_\infty \quad (20)$$

4.2. Comparison with Other Methods

To comprehensively evaluate the proposed IPTR framework, we conduct a series of experiments on both benchmark and newly constructed datasets. We compare our model with four representative baselines: Vor-UNet and Vor-FNO, which take Voronoi-interpolated fields as input; and Mask-UNet and Mask-FNO, which directly process raw sparse monitoring maps. UNet and FNO serve as the backbone architectures to isolate the impact of model architecture, while the two input settings distinguish the role of spatial interpolation. These baselines provide a fair and comprehensive comparison to our IPTR.

4.2.1. Benchmark Evaluation

In the benchmark evaluation, we perform experiments under two settings: single-condition and multi-condition. We use the test set provided in

the benchmark to evaluate model performance. For the single-condition evaluation, we randomly sample 1,000 instances from each of the three TFRD sub-datasets (HSink, ADlet, and DSine). During training, we randomly form 1,000 reference–target pairs under the same thermal condition to supervise the reconstruction task. The specific sampling strategy is further discussed in Section 4.3.2. For the multi-condition setting, we construct a combined training set by mixing the sampled instances from all three sub-datasets, allowing the model to learn under diverse boundary and source configurations. During testing, we randomly select one sample from the training set under the same thermal condition as the reference pair, and use all corresponding samples in the test set as target inputs for evaluation.

Single-Condition Reconstruction: As shown in Table 1, our method demonstrates consistently strong performance across all three cases based on quantitative metrics. In terms of MAE, the results of our method are 0.1154, 0.0149, and 0.0347 respectively—either achieving or closely approaching the best-performing models—while maintaining minimal variation across tasks, indicating robust generalization. More notably, our method achieves the lowest Max-AE values in all cases (3.2731, 0.1236, and 0.3207), outperforming the second-best methods by an average of 76.5%, highlighting its superior ability to control local reconstruction errors. As the temperature span increases from 15.5K (ADlet) to 130.6K (HSink), most baseline methods show a significant rise in reconstruction error. For instance, Vor-UNet, as a representative baseline, sees its Max-AE grow from 1.1551 to 14.6286. In comparison, our method shows only a moderate increase—from 0.1236 to 3.2731—while still achieving the lowest error in both scenarios. This demonstrates that our model better preserves accuracy under more challenging temperature differences and maintains stable performance across varying conditions.

Method	HSink		ADlet		DSine	
	MAE↓	Max-AE↓	MAE↓	Max-AE↓	MAE↓	Max-AE↓
Mask-UNet	0.2715	<u>11.3475</u>	0.0342	<u>0.6265</u>	0.1148	2.7575
Mask-FNO	1.4606	11.4818	0.4393	1.6913	2.1168	5.9253
Vor-UNet	7.7331	14.6286	0.0115	1.1551	<u>0.0350</u>	<u>1.4806</u>
Vor-FNO	<u>0.2439</u>	12.6760	0.0444	5.7005	0.1168	4.6461
Ours	0.1154	3.2731	<u>0.0149</u>	0.1236	0.0347	0.3207

Table 1: Quantitative comparisons under single-condition reconstruction.

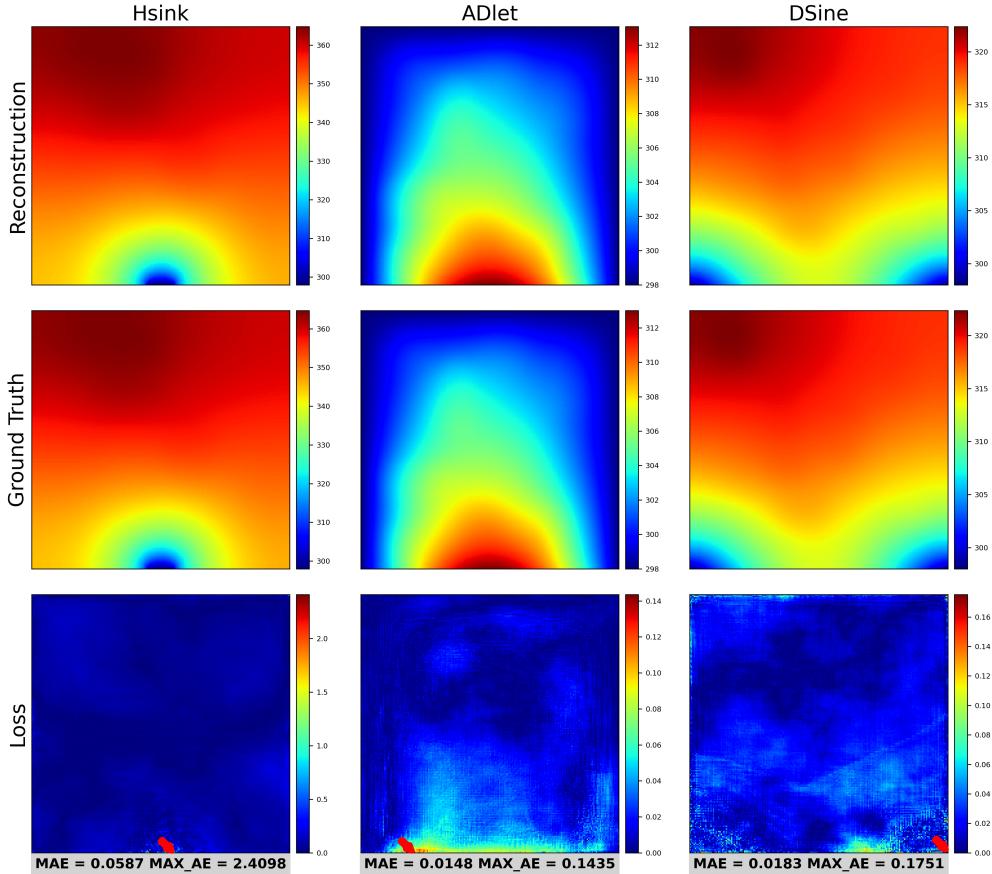


Figure 7: Visualization of our result under singe-condition reconstruction.

Qualitative results (Figure 7 and Figure 8) further support these findings. Across all datasets, our reconstructions closely align with the ground truth, with smoother error distributions and fewer high-error zones. This is especially evident in the HSink case, where baseline models such as Mask-FNO and Vor-UNet exhibit visible local artifacts. In ADlet and DSine, our model also preserves finer structural details and maintains uniform error fields. These visual patterns are consistent with the numerical metrics and underscore the reliability of our approach for single-condition reconstruction. **Multi-Condition Reconstruction:** Table 2 and Figure 9 report the results under the multi-condition setting, where the model is trained on a combined dataset containing samples from all three sub-datasets. Across all test cases,

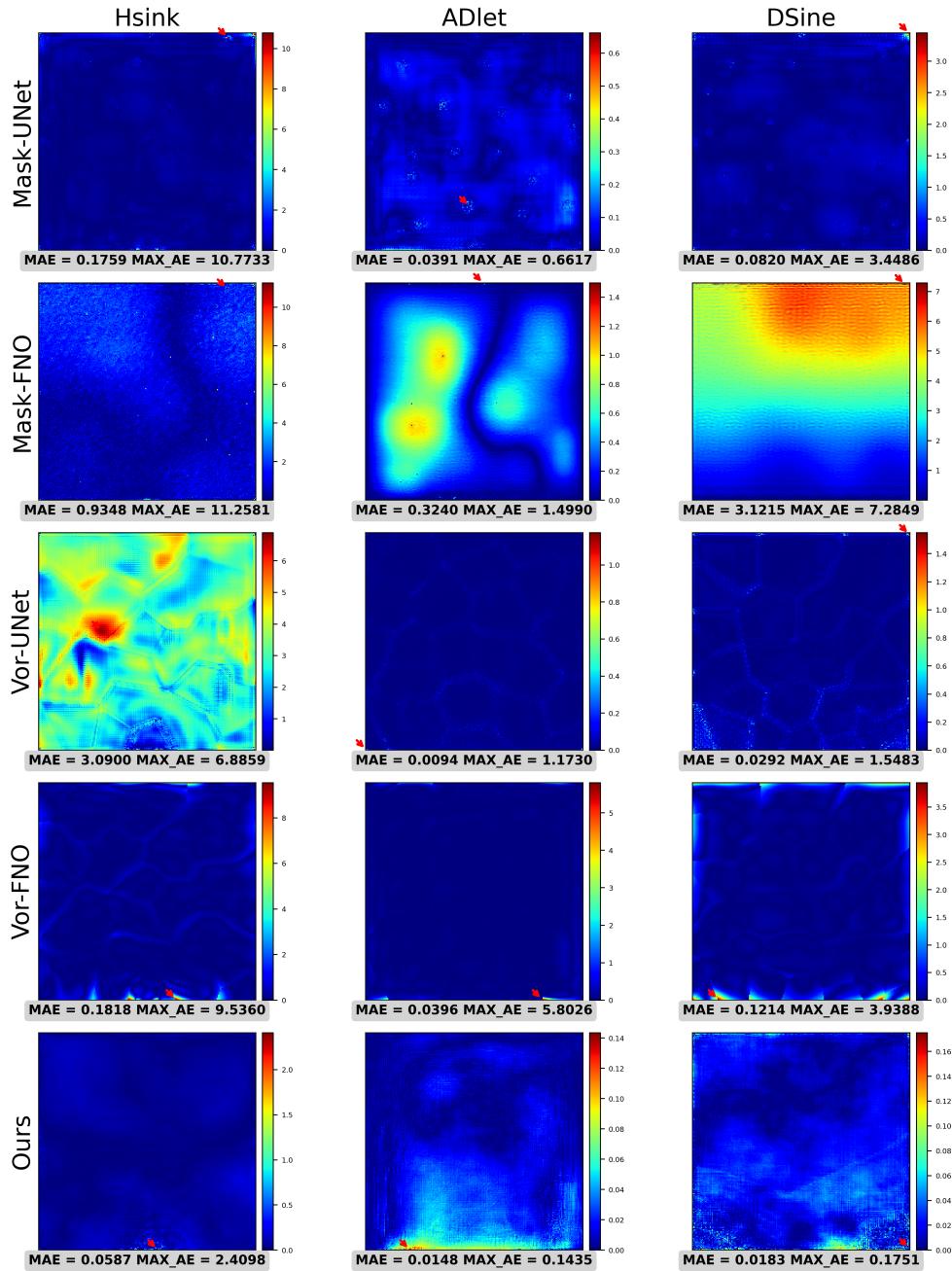


Figure 8: The qualitative comparison of our IPTR with other representative methods under single-condition reconstruction.

our method consistently outperforms existing approaches in both MAE and Max-AE, achieving the best performance on HSink (0.1375 / 1.7476), ADlet (0.0189 / 1.0168), and DSine (0.0391 / 1.1947). These results highlight the strong generalization ability of our model when exposed to diverse boundary conditions and source patterns during training. The qualitative comparisons in Figure 9 further support these findings. Our model produces spatially smooth and accurate reconstructions, with notably fewer high-error regions. In contrast, other methods—such as Mask-FNO and Vor-UNet—tend to exhibit large localized errors or structural distortions, particularly in the HSink and DSine cases. For instance, Vor-UNet suffers from banding artifacts in HSink and Mask-FNO shows noisy residuals across all scenarios, indicating difficulty in generalizing under mixed-condition inputs. Our results remain stable and visually consistent, demonstrating both high accuracy and reliability in challenging, heterogeneous settings.

Method	HSink		ADlet		DSine	
	MAE \downarrow	Max-AE \downarrow	MAE \downarrow	Max-AE \downarrow	MAE \downarrow	Max-AE \downarrow
Mask-UNet	<u>0.1559</u>	8.4566	0.1026	1.5651	<u>0.1188</u>	9.0847
Mask-FNO	0.8158	<u>4.6842</u>	0.4454	3.1466	0.4025	8.8669
Vor-UNet	6.2315	12.9236	<u>0.0632</u>	<u>1.3899</u>	0.1515	8.5528
Vor-FNO	0.1860	9.5279	0.2312	6.4576	0.1426	<u>4.9227</u>
Ours	0.1375	1.7476	0.0189	1.0168	0.0391	1.1947

Table 2: Quantitative comparisons under multi-condition reconstruction.

4.2.2. Generalization to New Scenario

To assess the generalization capability and robustness of IPTR, we construct a new scenario comprising 6000 previously unseen samples that are not included in the TFRD benchmark. Among them, 1000 samples are reserved for validation and another 1000 for testing, while the remaining 4000 samples are used for training with varying data sizes in different experiments. This section presents a series of experiments focusing on sample efficiency and few-shot learning, aiming to simulate real-world deployment conditions where labeled data are limited and sensor layouts are constrained, as exemplified by Case-4 in Figure 6.

Sample-Efficiency Comparison: To investigate the sample efficiency of different models under the new scenario, we train each method using vary-

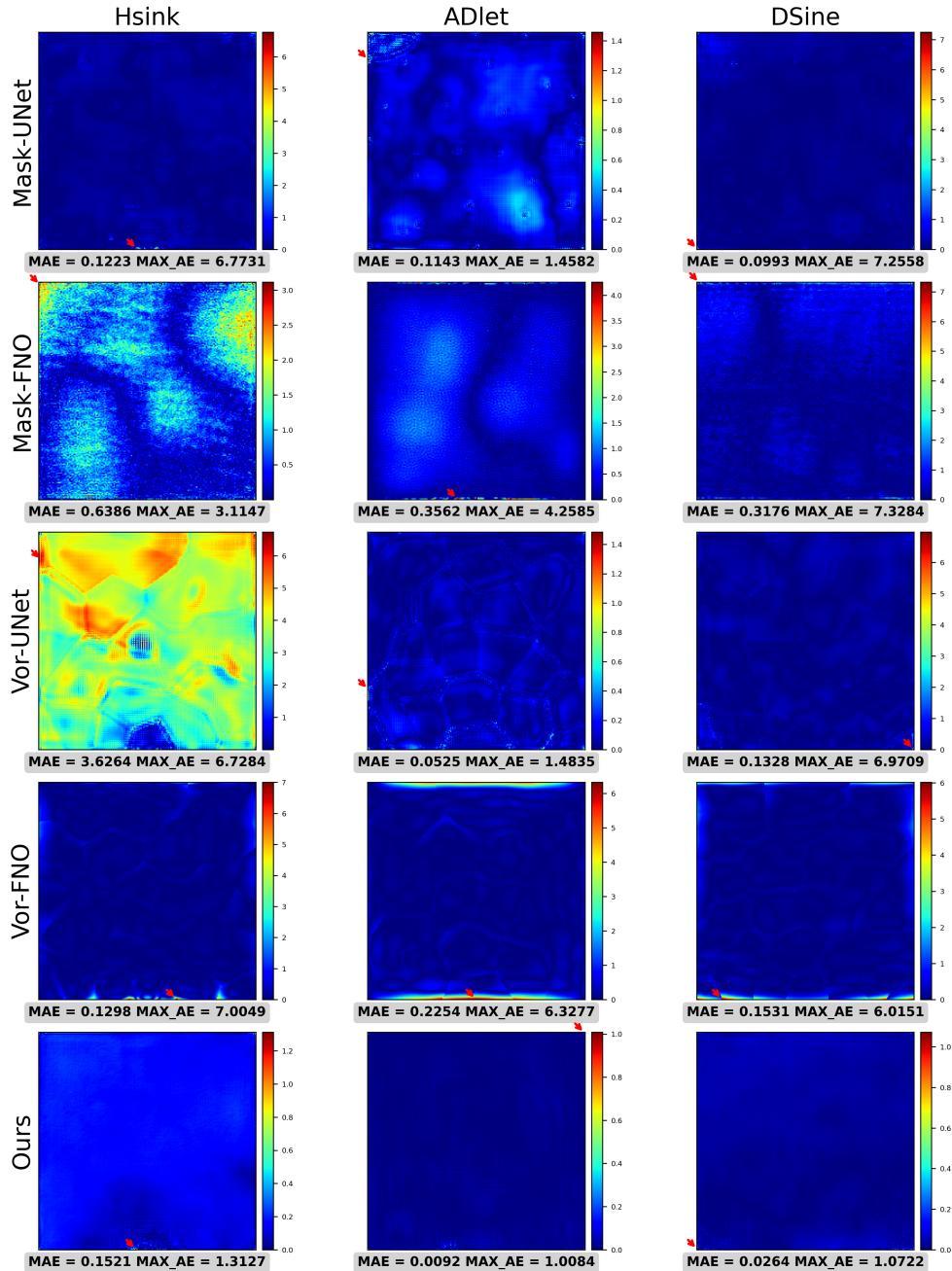


Figure 9: The qualitative comparison of our IPTR with other representative methods under multi-condition reconstruction.

ing amounts of labeled data: 1000, 1500, 2000, and 4000 samples. For each training configuration, a corresponding validation set of 100, 200, 500, and 1000 samples is used to monitor performance during training. The quantitative results, visualized in Figure 10, highlight a clear trend: our method consistently achieves the lowest MAE and Max-AE across all sample sizes, demonstrating strong performance across different amounts of samples.

In terms of MAE, our model reaches 0.1247 with 1000 samples and improves to 0.0297 at 4000 samples—surpassing all baselines at every point. Notably, the performance gap widens as the number of training samples increases. For example, when using 4000 samples, the closest competitor (Vor-UNet) reports an MAE of 0.0370, which is still 24.6% higher than ours. Similarly, Mask-UNet exhibits high variability in performance, starting with poor accuracy at low data volumes and only matching Vor-UNet’s performance at 2000 samples, where our model already achieves superior results.

When considering Max-AE, our method again leads by a significant margin. With only 1000 samples, our Max-AE is 2.7415—better than almost all other methods even when they are trained with four times more data. At 4000 samples, we achieve a Max-AE of 1.0279, which is less than half that of the best-performing baseline (Vor-UNet: 2.3932), indicating our model’s exceptional ability to constrain local errors.

These results collectively suggest that our method not only maintains high accuracy across different training scales, but also scales more effectively with increasing supervision—demonstrating strong generalization under varying amounts of training data. Rather than simulating limited-data conditions, this set of experiments reflects model performance under sufficiently sampled scenarios, allowing us to assess how well different methods benefit from additional supervision. In the next section, we further evaluate the model’s adaptability in few-shot settings by conducting few-shot transfer experiments, where only a limited number of labeled samples are available in the target domain.

Few-Shot Transfer Learning: Compared to the previous setting with sufficient training samples, this section evaluates model performance under few-shot condition using transfer learning, which is a practical strategy to leverage prior knowledge when training data in the target scenario are insufficient. All models are initialized from their counterparts pre-trained on the TFRD multi-condition benchmark. This choice reflects a practical consideration: in real-world deployments, the relevance between pre-training and downstream tasks is often unknown or difficult to quantify. Therefore, a

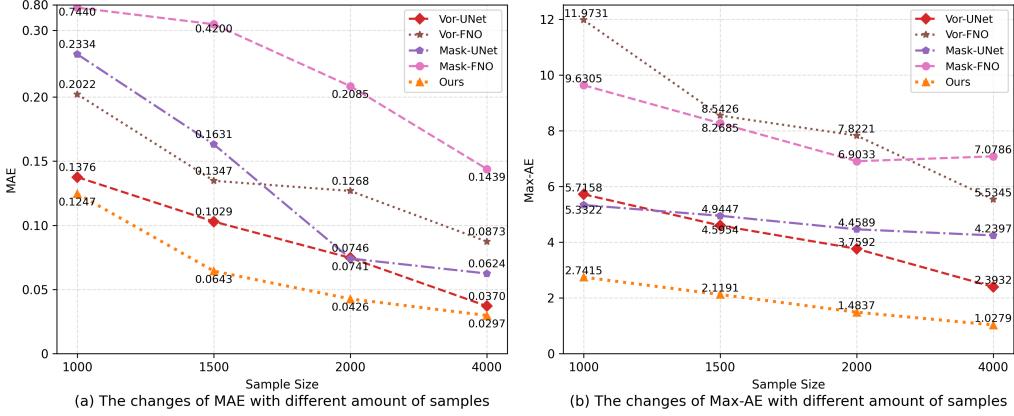


Figure 10: Sample-Efficiency Comparison.

model trained across diverse conditions serves as a more general and transferable backbone for downstream adaptation. Fine-tuning is performed with 2, 5, 10, 20, and 50 training samples, each paired with a validation set of 1, 2, 4, 8, and 16 samples, respectively.

Number of Samples	2		5		10		20		50	
	MAE↓	Max-AE↓	MAE↓	Max-AE↓	MAE↓	Max-AE↓	MAE↓	Max-AE↓	MAE↓	Max-AE↓
Mask-UNet	3.5720	14.6502	1.3315	<u>13.1989</u>	1.2247	14.2104	0.4999	10.4581	0.3058	<u>7.7115</u>
Mask-FNO	6.5030	101.0619	2.4550	46.4677	1.2713	12.4416	0.7302	9.7054	0.4021	8.8690
Vor-UNet	4.7273	<u>13.6441</u>	3.8922	14.0768	3.6637	<u>10.1455</u>	3.9723	<u>8.6465</u>	3.3265	8.1748
Vor-FNO	0.3892	18.4292	0.3019	16.3998	0.2507	15.5292	<u>0.2012</u>	13.0728	0.1471	9.9537
Ours	0.2349	7.1077	0.1918	7.0079	0.1291	5.7714	0.1049	4.4456	0.0614	3.8818

Table 3: Few-shot Transfer Learning.

As summarized in Table 3, our model consistently outperforms all baselines across various few-shot scenarios, achieving both lowest reconstruction error and best stability.

Convolution-based architectures such as Vor-UNet and Mask-UNet exhibit limited transferability in this setting. Vor-UNet, despite integrating geometric priors through Voronoi tessellation, consistently shows high errors across all sample sizes. For instance, with 50 training samples, it still yields a MAE of 3.3265 and a Max-AE of 8.1748. This indicates that its spatial encoding strategy fails to adapt effectively to the new domain. This indicates that its spatial encoding strategy fails to adapt effectively to the new domain. Mask-UNet performs somewhat better as the sample size increases, achiev-

ing an MAE of 0.3058 with 50 samples, but it still struggles in the few-shot regime, suggesting that the raw monitoring measurements are insufficient for robust knowledge transfer.

Among operator-based models, Vor-FNO demonstrates relatively stable performance under few-shot transfer. Even with only 10 training samples, it achieves reasonable accuracy, and its errors steadily decrease as the sample size increases. This suggests that its spectral-based encoding retains a certain level of generalization capability. In contrast, Mask-FNO suffers from pronounced instability across all data regimes. Particularly in extremely few-shot settings, such as with only 2 samples, it produces highly unreliable results with a Max-AE exceeding 100, and remains significantly worse than other baselines even at 50 samples. These results indicate that using the original monitoring measurements in Mask-FNO weakens its ability to transfer meaningful representations, likely by disrupting the operator network’s inherent inductive bias.

In contrast, our proposed IPTR consistently outperforms all baseline methods across all few-shot scenarios. With only 2 training samples, IPTR achieves an MAE of 0.2349 and Max-AE of 7.1077. When provided with 50 samples, the performance further improves to an MAE of 0.0614 and Max-AE of 3.8818. More importantly, this performance surpasses many existing methods trained from scratch using substantially larger datasets. For example, Mask-Unet requires 4000 training samples to reach comparable accuracy (MAE of 0.0624), while Vor-FNO with 4000 samples still results in a higher MAE of 0.0873 and Max-AE of 5.5345. To complement the quantitative results, Figure 11 presents the visualization of reconstructed temperature fields and corresponding MAE for all methods when fine-tuned with 50 samples. The visual comparison clearly highlights the superiority of our IPTR framework in capturing detailed temperature variations and preserving structural consistency. In contrast, other baselines show noticeable artifacts, particularly in regions with sharp gradients. This demonstrates that our method significantly enhances data efficiency and transferability.

The superior performance of IPTR is primarily attributed to its architectural design. By incorporating an Implicit Physics-guided branch, IPTR leverages historical simulation data not as direct training input, but as reference to encode implicit physical priors. The dual-branch physics embedding module enables the model to jointly exploit these implicit priors through a cross-attention mechanism, while preserving the target task’s characteristics via a Fourier-based auxiliary encoder. This synergy allows IPTR to main-

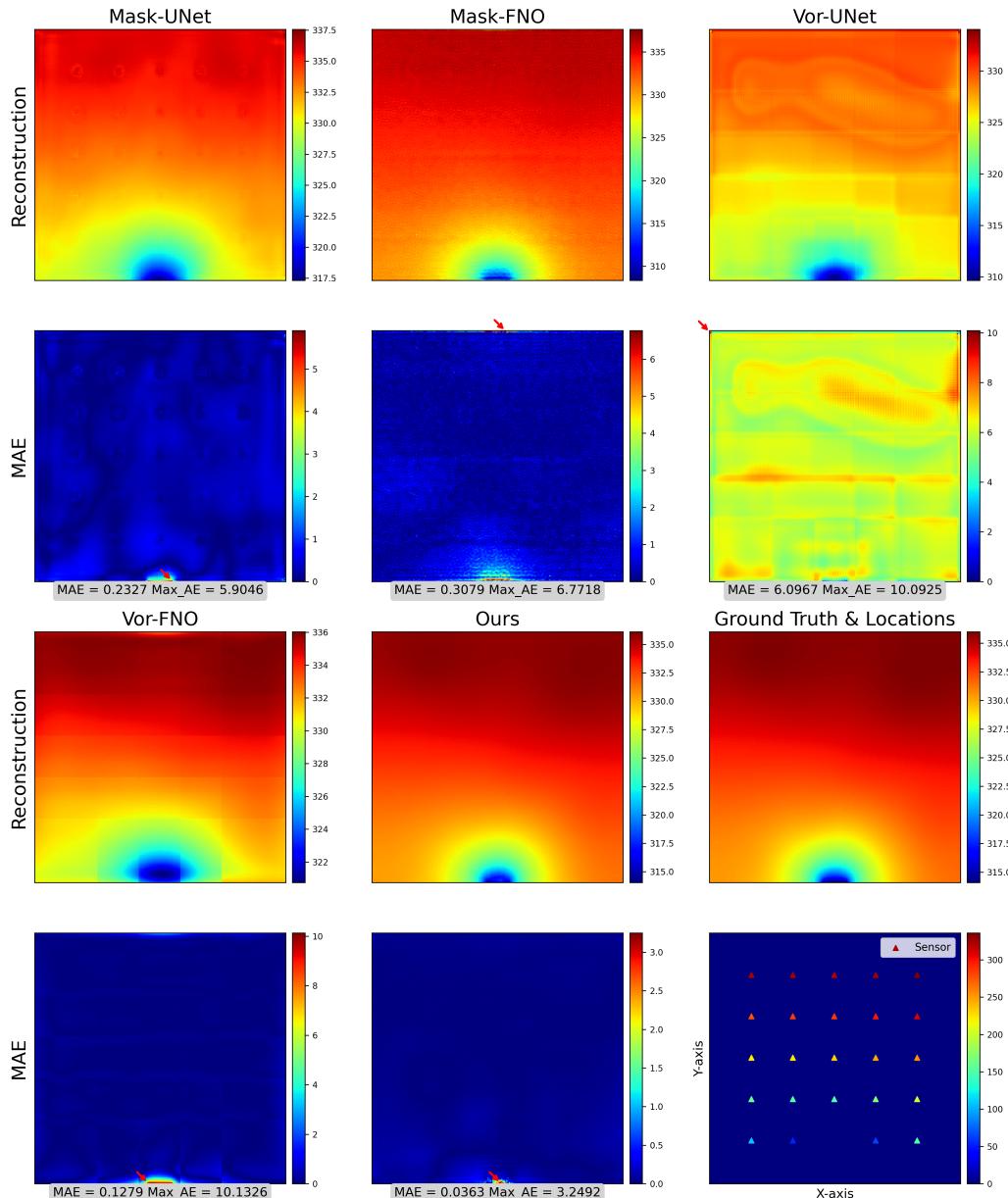


Figure 11: Few-shot transfer learning with 50 samples.

tain physical consistency and stable reconstruction, even when the available training data is extremely limited.

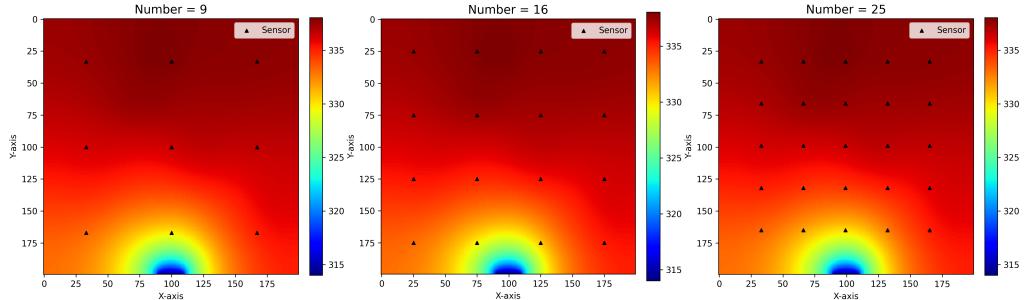


Figure 12: Visualizations of different number of sensors.

4.3. Further Discussion

4.3.1. Influence of the Number of Sensors

The number of deployed sensors plays a critical role in determining the reconstruction quality of temperature fields, as it directly affects the spatial coverage and informativeness of the input observations. To systematically investigate this factor, we conduct controlled experiments to evaluate the impact of varying sensor quantities on the performance of our proposed method. All experiments are performed on the newly generated dataset containing 2,000 training samples as illustrated in Section 4.2.2. To ensure consistency and isolate the effect of sensor quantity, we adopt a uniform placement strategy across all settings. Specifically, we evaluate three sensor configurations (Figure 12) with 9, 16, and 25 sensors, respectively. The results are summarized in Figure 13. Across all metrics and configurations, our proposed IPTR method consistently outperforms other baselines, achieving the lowest MAE and Max-AE under all sensor counts. Notably, IPTR maintains high reconstruction accuracy even with only 9 sensors, demonstrating its robustness under extremely sparse observation. Furthermore, the reconstruction performance peaks when using 16 sensors, slightly surpassing the 25-sensor case. This counter-intuitive observation may be attributed to the potential spatial redundancy introduced by a denser sensor layout. While additional sensors generally provide more information, they may not contribute equally due to overlapping receptive fields or limited additional value in relatively smooth regions of the field. In contrast, the 16-sensor configuration may

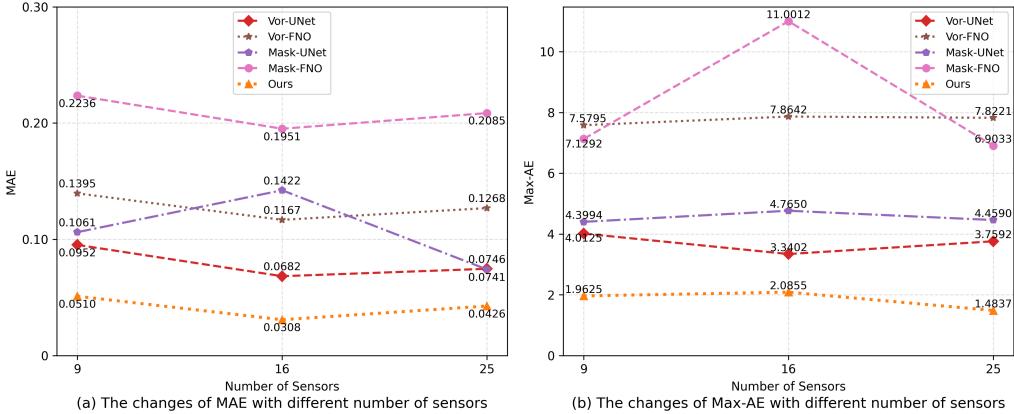


Figure 13: Influence of the Number of Sensors.

strike a better balance between spatial coverage and input sparsity, allowing the model to focus on more informative regions without redundancy, thus yielding better generalization.

4.3.2. Ablation Study

To evaluate the effectiveness of our design choices in IPTR, we perform ablation studies on the TFRD multi-condition dataset. Specifically, we assess (1) the impact of different reference sampling strategies for training data construction, and (2) the contribution of the Dual Physics Embedding Module.

The impact of different reference sampling strategy: Each training pair in IPTR requires a target sample and a corresponding reference sample. We compare two schemes: *Fixed-reference*, where all pairs share a common simulation instance as reference; *Sliding-reference*, where training pairs are formed by sequentially pairing samples with their immediate successors based on dataset index.

Quantitative results as shown in Table 4 demonstrate that the sliding-reference sampling strategy significantly outperforms the fixed-reference setting across all sub-tasks. It notably reduces both MAE and Max-AE in HSink and DSine, indicating stronger reconstruction accuracy and robustness. In ADlet, although the Max-AE slightly increases (from 0.2341 to 1.0168), the MAE is substantially reduced (from 0.0457 to 0.0189), suggesting that overall accuracy improves despite rare local deviations. These results confirm that introducing dynamic and diverse reference frames facilitates more effective

implicit-physics prior extraction. Accordingly, the sliding-reference strategy is adopted in all subsequent experiments.

Sampling Strategy	HSink		ADlet		DSine	
	MAE↓	Max-AE↓	MAE↓	Max-AE↓	MAE↓	Max-AE↓
Fixed-reference	0.9820	5.4610	0.0457	0.2341	0.5792	1.5822
Sliding-reference	0.1375	1.7476	0.0189	1.0168	0.0391	1.1947

Table 4: Ablation study on sampling strategy.

Effectiveness of Dual Physics Embedding: To investigate the contribution of the proposed Dual Physics Embedding Module, we conducted an ablation study summarized in Table 5. Specifically, we first validate the effect of the implicit physics-guided branch by removing the auxiliary branch (w/o Auxiliary Encoding Branch). This configuration still achieves strong performance across all three sub-tasks, particularly in terms of Max-AE, where the model consistently outperforms all baseline methods in Table 2. This confirms that the implicit physics priors extracted from reference data provide robust guidance for temperature field reconstruction, even without target spatial encoding. To further assess the role of the auxiliary encoder, we compare the original FNO-based variant with a UNet-based alternative. The model with UNet auxiliary encoder shows degraded performance, especially in HSink and DSine, with significantly increased Max-AE (e.g., 14.44 in HSink), indicating inferior generalization. By contrast, our full model—incorporating the FNO-based auxiliary encoder—achieves consistently strong performance across all tasks and metrics, highlighting the advantage of using Fourier-based representations to encode current field characteristics.

It is worth noting that removing the implicit physics-guided branch while keeping the auxiliary encoder and identical reference inputs (i.e., w/o Implicit Physics-Guided Branch in Table 5) does not yield comparable performance. This highlights that the superiority of our method arises not from the use of FNO itself or the naive concatenation of reference data, but from the synergistic design of the dual physics embedding. Among the two branches, the implicit physics-guided branch plays the dominant role by embedding physically consistent priors from reference data, ensuring robust reconstruction even under complex or extreme temperature variations. The auxiliary branch, based on a Fourier backbone, complements this process by refining

local details and preserving field-specific characteristics. The dual-branch structure thus provides a balanced yet physics-guided representation that enhances both stability and reconstruction fidelity.

Method	HSink		ADlet		DSine	
	MAE↓	Max-AE↓	MAE↓	Max-AE↓	MAE↓	Max-AE↓
w/o Auxiliary Encoding Branch	<u>0.3542</u>	<u>2.9001</u>	0.0547	0.4401	0.3475	1.4865
w/o Implicit Physics-Guided Branch	0.6641	4.2551	0.0833	<u>0.5611</u>	0.2193	1.1333
w/ UNet Auxiliary Encoder	0.5637	14.4374	<u>0.0523</u>	2.6832	<u>0.1228</u>	1.7970
Ours	0.1375	1.7476	0.0189	1.0168	0.0391	1.1947

Table 5: Ablation study on the role of dual physics embedding module.

5. Conclusion

In this paper, we present IPTR, a novel implicit physics-guided framework for temperature field reconstruction of heat-source systems (TFR-HSS) under sparse sensor observations. Unlike conventional approaches that treat TFR-HSS as a one-to-one regression from sparse monitoring measurements, IPTR introduces a reference monitoring–field pair (under thermal conditions similar to the target) to implicitly encode physical priors. This design enhances the model’s ability to recover extreme regions and improves generalization in data-scarce scenarios. To effectively integrate target monitoring data and reference information, we design a Dual Physics Embedding Module, consisting of an implicit physics-guided branch and an auxiliary encoder based on Fourier layers. The former extracts latent physical knowledge from reference data via cross-attention, while the latter preserves the spatial structure of target observations. These two branches work in synergy to form a robust representation for accurate field reconstruction.

Extensive experiments are conducted under various conditions—including single-condition, multi-condition, few-shot generalization settings, and different number of sensor placement—demonstrating that IPTR consistently achieves superior reconstruction performance and strong generalization capability in terms of both MAE and Max-AE metrics, validating the effectiveness of implicit physics guidance for solving ill-posed reconstruction problems. In future work, we plan to extend IPTR toward higher-dimensional physical field reconstruction problems, aiming to tackle more complex engineering scenarios with sparse observations. In addition, we will explore more effective

strategies for selecting reference pair, beyond the current random sampling scheme, in order to better capture relevant physical correlations and further enhance reconstruction accuracy and robustness.

Acknowledgments

This work was partly supported by the Young Elite Scientist Sponsorship Program By CAST (Grant No. YESS20240697) and the National Natural Science Foundation of China (Grant No. 92371206)

References

- [1] M Grujicic, CL Zhao, and EC Dusel. The effect of thermal contact resistance on heat management in the electronic packaging. *Applied surface science*, 246(1-3):290–302, 2005.
- [2] Xiaoyu Zhao, Zhiqiang Gong, Yunyang Zhang, Wen Yao, and Xiaoqian Chen. Physics-informed convolutional neural networks for temperature field prediction of heat source layout without labeled data. *Engineering Applications of Artificial Intelligence*, 117:105516, 2023.
- [3] Wen Yao, Xiaoqian Chen, Wencai Luo, Michel Van Tooren, and Jian Guo. Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. *Progress in Aerospace Sciences*, 47(6):450–479, 2011.
- [4] Ziqiang He, Yunfei Yan, and Zhien Zhang. Thermal management and temperature uniformity enhancement of electronic devices by micro heat sinks: A review. *Energy*, 216:119223, 2021.
- [5] Yunyang Zhang, Zhiqiang Gong, Weien Zhou, Xiaoyu Zhao, Xiaohu Zheng, and Wen Yao. Multi-fidelity surrogate modeling for temperature field prediction using deep convolution neural network. *Engineering Applications of Artificial Intelligence*, 123:106354, 2023.
- [6] Kai Chen, Shuangfeng Wang, and Mengxuan Song. Optimization of heat source distribution for two-dimensional heat conduction using bionic method. *International Journal of Heat and Mass Transfer*, 93:108–117, 2016.
- [7] Paisarn Naphon, Setha Klangchart, and Somchai Wongwises. Numerical investigation on the heat transfer and flow in the mini-fin heat sink for cpu. *International Communications in Heat and Mass Transfer*, 36(8):834–840, 2009.

- [8] Lu Wang, Jianyin Miao, Mingming Gong, Qiang Zhou, Chang Liu, Hongxing Zhang, and Hanlin Fan. Research on the heat transfer characteristics of a loop heat pipe used as mainline heat transfer mode for spacecraft. *Journal of Thermal Science*, 28(4):736–744, 2019.
- [9] Xianqi Chen, Xiaoyu Zhao, Zhiqiang Gong, Jun Zhang, Weien Zhou, Xiaoqian Chen, and Wen Yao. A deep neural network surrogate modeling benchmark for temperature field prediction of heat source layout. *Science China Physics, Mechanics & Astronomy*, 64(11):1, 2021.
- [10] Boris Benedikter, Alessandro Zavoli, Guido Colasurdo, Simone Pizzurro, and Enrico Cavallini. Convex optimization of launch vehicle ascent trajectory with heat-flux and splash-down constraints. *Journal of Spacecraft and Rockets*, 59(3):900–915, 2022.
- [11] Hanwei Zhao, Youliang Ding, Aiqun Li, Bin Chen, and Xiaonan Zhang. State-monitoring for abnormal vibration of bridge cables focusing on non-stationary responses: From knowledge in phenomena to digital indicators. *Measurement*, 205:112148, 2022.
- [12] Amedeo Capozzoli, Claudio Curcio, Angelo Liseno, and Pietro Vinetti. Field sampling and field reconstruction: A new perspective. *Radio Science*, 45(06):1–31, 2010.
- [13] Xuejun Zhou, Fangyuan Zhou, and Maryam Naseri. An insight into the estimation of frost thermal conductivity on parallel surface channels using kernel based gpr strategy. *Scientific reports*, 11(1):7203, 2021.
- [14] Razieh Razavi, Aida Sabaghmoghadam, Amin Bemani, Alireza Baghban, Kwok-wing Chau, and Ely Salwana. Application of anfis and lssvm strategies for estimating thermal conductivity enhancement of metal and metal oxide based nanofluids. *Engineering Applications of Computational Fluid Mechanics*, 13(1):560–578, 2019.
- [15] Tong Ma, Yuqian Liu, and Chengyu Cao. Neural networks for 3d temperature field reconstruction via acoustic signals. *Mechanical systems and signal processing*, 126:392–406, 2019.
- [16] Carlos Augusto Richter Do Nascimento, Viviana Cocco Mariani, and Leandro dos Santos Coelho. Integrative numerical modeling and thermodynamic optimal design of counter-flow plate-fin heat exchanger applying neural networks. *International Journal of Heat and Mass Transfer*, 159:120097, 2020.

- [17] Nait Amar Menad, Zeraibi Noureddine, Abdolhossein Hemmati-Sarapardeh, Shahaboddin Shamshirband, Amir Mosavi, and Kwok-wing Chau. Modeling temperature dependency of oil-water relative permeability in thermal enhanced oil recovery processes using group method of data handling and gene expression programming. *Engineering Applications of Computational Fluid Mechanics*, 13(1):724–743, 2019.
- [18] Mohammad Hossein Ahmadi, Milad Sadeghzadeh, Amir Hossein Raffiee, and Kwok-wing Chau. Applying gmdh neural network to estimate the thermal resistance and thermal conductivity of pulsating heat pipes. *Engineering Applications of Computational Fluid Mechanics*, 13(1):327–336, 2019.
- [19] Xiaoqian Chen, Zhiqiang Gong, Xiaoyu Zhao, Weien Zhou, and Wen Yao. A machine learning surrogate modeling benchmark for temperature field reconstruction of heat source systems. *Science China Information Sciences*, 66(5):152203, 2023.
- [20] Xingwen Peng, Xingchen Li, Zhiqiang Gong, Xiaoyu Zhao, and Wen Yao. A deep learning method based on partition modeling for reconstructing temperature field. *International Journal of Thermal Sciences*, 182:107802, 2022.
- [21] Xiaoyu Zhao, Xiaoqian Chen, Zhiqiang Gong, Wen Yao, and Yunyang Zhang. A hybrid method based on proper orthogonal decomposition and deep neural networks for flow and heat field reconstruction. *Expert Systems with Applications*, 247:123137, 2024.
- [22] Kai Fukami, Romit Maulik, Nesar Ramachandra, Koji Fukagata, and Kunihiko Taira. Global field reconstruction from sparse sensors with voronoi tessellation-assisted deep learning. *Nature Machine Intelligence*, 3(11):945–951, 2021.
- [23] Javier E Santos, Zachary R Fox, Arvind Mohan, Daniel O’Malley, Hari Viswanathan, and Nicholas Lubbers. Development of the senseiver for efficient field reconstruction from sparse observations. *Nature Machine Intelligence*, 5(11):1317–1325, 2023.
- [24] Wenzhe Zhang, Zixue Xiang, Ming Sun, and Wen Yao. In-context learning for temperature field reconstruction under multiple layouts. In *International Conference on Neural Information Processing*, pages 211–225. Springer, 2024.
- [25] Jiahuan Long, Wenzhe Zhang, Ning Wang, Tingsong Jiang, and Wen Yao. Fr-mamba: Time-series physical field reconstruction based on state space model. *arXiv preprint arXiv:2505.16083*, 2025.

- [26] Yufeng Xia, Wen Yao, Xiaohu Zheng, and Zhiqiang Gong. Reliability analysis of heat source layout temperature field prediction considering uncertainty in deep neural network surrogate models. *Quality and Reliability Engineering International*, 39(5):1775–1795, 2023.
- [27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
- [28] Zhiqiang Gong, Weien Zhou, Jun Zhang, Wei Peng, and Wen Yao. Joint deep reversible regression model and physics-informed unsupervised learning for temperature field reconstruction. *Engineering Applications of Artificial Intelligence*, 118:105686, 2023.
- [29] Xu Liu, Wei Peng, Zhiqiang Gong, Weien Zhou, and Wen Yao. Temperature field inversion of heat-source systems via physics-informed neural networks. *Engineering Applications of Artificial Intelligence*, 113:104902, 2022.
- [30] Yimeng Chi, Mingliang Li, Rui Long, Zhichun Liu, and Wei Liu. Heat source field inversion and detection based on physics-informed deep learning. *International Communications in Heat and Mass Transfer*, 164:108824, 2025.
- [31] Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.
- [32] 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.
- [33] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Andrew Stuart, Kaushik Bhattacharya, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. *Advances in Neural Information Processing Systems*, 33:6755–6766, 2020.
- [34] Nicholas H Nelsen and Andrew M Stuart. The random feature model for input-output maps between banach spaces. *SIAM Journal on Scientific Computing*, 43(5):A3212–A3243, 2021.

- [35] Xiaoyu Zhao, Xiaoqian Chen, Zhiqiang Gong, Weien Zhou, Wen Yao, and Yunyang Zhang. Recfno: A resolution-invariant flow and heat field reconstruction method from sparse observations via fourier neural operator. *International Journal of Thermal Sciences*, 195:108619, 2024.
- [36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [37] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2337–2346, 2019.
- [38] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [39] A Paszke. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.