RECONCILING DEEP LEARNING WITH OCEANOGRAPHY: SEAMLESS FIELD REPRESENTATION VIA HEAT-CONDUCTION INFILLING

Zhaolin Sun

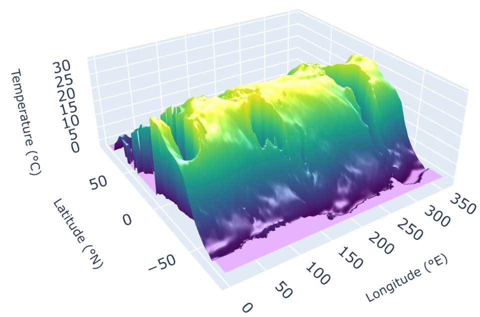
ABSTRACT

Applying deep learning to model high-fidelity ocean fields is often hindered by a fundamental inductive bias mismatch. Standard Convolutional Neural Networks (CNNs), designed for natural images, struggle with the irregular, sharp land-sea boundaries and non-local teleconnections inherent in oceanographic data. To address this, we propose Heat-Conduction Infilling (HCI), a novel, physics-driven preprocessing technique that eliminates spurious boundary artifacts. HCI models the land infilling problem as a steady-state heat conduction process, solving the Laplace equation to generate a globally continuous and seamless field. This fundamentally aligns the data representation with the intrinsic preferences of CNNs, making the learning task significantly more tractable. We further introduce a Climate-State Conditioning mechanism that embeds the Niño index to capture global climate patterns like ENSO. Our framework demonstrates transformative performance improvements across multiple challenging tasks. In multi-layer sea temperature reconstruction, we reduce the state-of-the-art Mean Squared Error (MSE) from 0.648(Song et al., 2025) down to an unprecedented 0.20. Furthermore, by enabling a pre-trained VAE from Stable Diffusion(Rombach et al., 2022) to learn a superior latent space, we achieve a reconstruction MSE of just 0.04 on complex surface layers. The comprehensive validation suggests that HCI is a universally applicable and highly effective technique. We propose that it should become a standard preprocessing paradigm for applying deep learning models to gridded oceanographic data, such as sea temperature and salinity fields.Our open-source code is available at ecnu-Sun/Heat-Conduction-Infilling

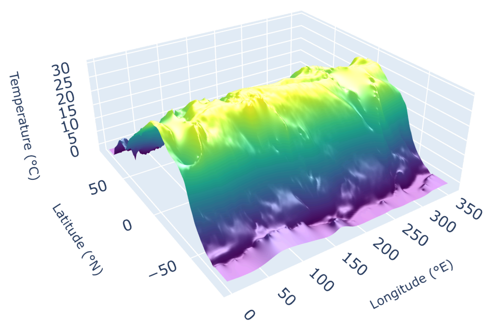
1 INTRODUCTION

High-fidelity, high-resolution ocean temperature field data are crucial for modern climate science, ocean forecasting, and ecosystem research. To obtain high-quality ocean temperature data from sparse, low-resolution observations, deep learning has emerged as a powerful tool, particularly in two key applications: reconstruction from incomplete data and super-resolution for recovering fine-scale structures from low-resolution observations(Reichstein et al., 2019). Furthermore, with the advancement of generative models such as diffusion models (Ho et al., 2020), learning latent space representations of data promises to better address core problems like reconstruction and super-resolution with greater computational efficiency in the future(Rombach et al., 2022).

However, a fundamental inductive bias mismatch arises when directly applying standard deep neural networks designed for natural images to oceanographic physical fields. The success of Convolutional Neural Networks (CNNs) stems from their powerful inductive biases of spatial invariance and locality(Wang & Wu, 2023). In sea temperature data, however, irregular land-sea boundaries are treated as sharp numerical cliffs, introducing severe boundary artifacts. Conventional methods circumvent land regions using a land-sea mask, allowing for the direct application of standard CNNs.(Song et al., 2025; Izumi et al., 2022; Hirahara et al., 2022; Barth et al., 2020) Yet, due to the spatial invariance bias, CNN models erroneously interpret these cliffs, which contain no physical information, as strong features that demand attention at any location. This boundary artifact issue is exacerbated in practical applications by inconsistencies across data sources. Different datasets, such as outputs from various models or reanalysis data, often employ distinct native resolutions, statistical conventions, and grid types (e.g., tri-polar grids), resulting in disparate land-sea masks. Particularly at coarse resolutions, numerous grid cells contain both land and ocean, leading to ambiguity in mask definition(Shah et al., 2023). When multiple datasets are used for joint training, these inconsistent boundaries introduce a de facto domain mismatch between samples, posing a significant challenge to training a robust model capable of handling diverse data sources(Ben-David et al., 2010). Moreover, geophysical systems like the ocean possess intrinsic properties that are fundamentally different from natural images; they are governed by physical laws and replete with non-local teleconnections—for instance, the El Niño–Southern Oscillation (ENSO) can exert profound influences(McPhaden et al., 2006) globally—which directly conflicts with the locality assumption of CNNs.



(a) original data



(b) processed data

Figure 1: Data Comparison(Before and After HCI)

To resolve the mismatch between oceanographic datasets and the spatial invariance inductive bias of CNNs, we propose that instead of forcing the model to learn these complex and physically meaningless boundaries, a more effective approach is to eliminate them at the data level. To this end, we draw inspiration from classical PDE-based interpolation(Bae & Weickert, 2008) and devise a novel, physics-driven infilling method for the data preprocessing stage, which we term Heat-Conduction Infilling (HCI). This method is inspired by thermodynamics and models the land infilling problem as a steady-state heat conduction process(Kraemer & Chen, 2014). By solving the Laplace equation, it generates a smooth, continuous, and physically consistent transition zone over land areas, creating a seamless field that connects with the ocean data. This operation elegantly resolves the fundamental conflict between the data representation and the model architecture: it transforms the original data, defined on an irregular ocean domain, into a globally continuous and seamless field representation devoid of sharp boundaries, thereby better aligning with the intrinsic preferences of convolutional neural networks. By eliminating boundary artifacts, HCI significantly reduces the data's Total Variation (TV) and noise (as will be demonstrated via frequency-domain analysis), thus creating a data representation that is more compatible with the inductive biases of CNNs and fundamentally easier to learn.

We further address the conflict between the locality assumption of CNNs and the non-local nature of ocean systems. Standard CNNs struggle to capture teleconnection signals(Wang & Wu, 2023). We therefore propose to condition the generative process on the macro-climate state. Specifically, we embed the Niño index—a quantitative metric for the ENSO phenomenon—into the generative model. This design provides the model with critical climate context, enabling it to learn a conditional probability distribution. Consequently, the model can intelligently generate more physically plausible SST patterns corresponding to El Niño, La Niña, or neutral states, thereby better capturing teleconnection effects.

In summary, this paper introduces a novel framework for high-fidelity ocean temperature modeling with deep learning, which consists of two independently applicable, pluggable modules:

- A physics-driven infilling technique, Heat-Conduction Infilling (HCI), that resolves the fundamental mismatch between sea temperature datasets and the spatial invariance inductive bias of CNNs. By solving the heat conduction equation to seamlessly fill land areas, HCI eliminates physically meaningless boundary artifacts and creates a data representation that is fundamentally easier for neural networks to learn.
- A climate-state conditioning mechanism that mitigates the inconsistency between oceanographic tasks and the locality inductive bias of CNNs. By introducing the Niño index as a condition, our generative model can learn the vacillations between different climate states. This enables it to capture non-local teleconnections and generate physically plausible oceanic patterns corresponding to anomalous states like El Niño, its atmospheric counterpart La Niña, or neutral conditions.

Our comprehensive experiments validate the superior performance of this framework, demonstrating fundamental performance enhancements in reconstruction and VAE learning tasks and elevating the accuracy of global-scale super-resolution to a new level. Crucially, we observe a universal and significant acceleration in model convergence across all tasks that employed HCI, providing strong evidence for its efficacy.

2 Related Work

2.1 Deep Generative Models for Ocean Field Reconstruction

In recent years, the reconstruction of geophysical fields has undergone a paradigm shift from traditional statistical methods to data-driven deep learning approaches. For instance, Su et al. (2022) successfully inverted the three-dimensional temperature field in the offshore China and Northwest Pacific regions using a U-Net-based deep learning model. Similarly, Wu et al. (2024) utilized a Convolutional Long Short-Term Memory (ConvLSTM) network to reconstruct a long-term dataset of subsurface temperatures covering the upper 2000 meters of the global ocean.

Of particular note, deep generative models, exemplified by Denoising Diffusion Probabilistic Models (DDPMs), have emerged as one of the frontier technologies in this domain. A state-of-the-art model, ReconMOST, has been successfully applied to the challenging task of reconstructing global multi-layer ocean temperature data from extremely sparse observations (e.g., with over 92.5% of data missing) (Song et al., 2025). The core of its framework is a two-stage process: first, an unconditional diffusion model is pre-trained on large-scale historical numerical simulation data to learn a physically consistent spatial prior; subsequently, during the generation phase, sparse real-world observations are used as guidance to constrain and direct the reverse diffusion process.

Although these methods have achieved significant progress on their respective problems, their underlying model architectures (e.g., U-Net and other CNN-based models) are still susceptible to the inductive bias mismatch issue we identified in the introduction. This is particularly evident with the sharp numerical cliffs introduced by land-sea boundary artifacts, which directly conflict with the spatial invariance inductive bias of CNNs. Furthermore, while many efforts in ocean field reconstruction have incorporated auxiliary information such as longitude, latitude, buoy data, and temporal information into generative models, there remains a lack of effective mechanisms to embed the NIÑO index as a climatic background condition to counteract the locality inductive bias of CNNs.

2.2 Super-resolution of Ocean Temperature Fields

Concurrently, Single-Image Super-Resolution (SISR) techniques, originally developed in the computer vision domain, have been adapted for applications in oceanography. Advanced CNN architectures, such as RCAN and RRDBNet, have demonstrated proficiency in recovering high-frequency details . Izumi et al. (2022) systematically explored the feasibility of applying visual deep learning models to super-resolve Sea Surface Temperature (SST) data.

Notably, Izumi et al. (2022) pointed out that the performance of CNN models deteriorates in regions containing land, an issue especially pronounced in simpler architectures like SRCNN. We argue that this finding strongly illustrates the inductive bias mismatch problem triggered by boundary artifacts. Specifically, the model's process of learning fine ocean textures is dominated and corrupted by the physically meaningless gradients at the land-sea boundary, which possess an intensity far exceeding that of the oceanic features themselves.

2.3 Representation Learning for Oceanographic Data

Learning low-dimensional, structured representations from high-dimensional data has been a central theme in modern deep learning. The Variational Autoencoder (VAE) is a powerful tool for this purpose, adept at capturing the principal modes of variation in data by learning a continuous and regularized latent space. The success of large-scale models like Stable Diffusion has further underscored the power of a high-quality latent space representation (Rombach et al., 2022).

Undoubtedly, applying VAEs to oceanographic datasets is a promising direction. Compressing ocean data into a well-structured latent space can facilitate the use of more advanced models for faster and improved sea temperature reconstruction and super-resolution tasks. Several excellent studies have already designed advanced VAEs for ocean temperature, such as the Spatio-Temporal Attention VAE (STAVAE) proposed by Zheng et al. (2024). However, we note that VAEs, as CNN-based models, are also susceptible to the sharp numerical cliffs introduced by land-sea boundary artifacts, which conflict with the spatial invariance inductive bias inherent to CNNs.

3 Method

3.1 Heat-Conduction Infilling

To fundamentally resolve the conflict between the boundary artifacts of the data representation and the spatial invariance inductive bias of CNN models, we propose a physics-informed infilling method applied during the data preprocessing stage, which we term Heat-Conduction Infilling (HCI). This method aims to reshape the original data, defined on an irregular domain, into a globally continuous and seamless field representation without sharp boundaries, thereby better aligning the data representation with the intrinsic preferences of convolutional neural networks.

3.1.1 General Problem Formulation

We first consider a general N-dimensional data field infilling problem. Given an N-dimensional data field $I(\mathbf{x})$ defined on an N-dimensional hyper-rectangle $\Omega \subset \mathbb{R}^N$, where $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ is a coordinate vector in the N-dimensional space. The domain Ω is partitioned into two disjoint subsets: an observed domain Ω_{obs}, where the field values I_{obs} are known, and a missing domain to be infilled, Ω_{missing}. Our objective is to find an infilled field I_{filled} that satisfies two conditions:

1. **On the observed domain (Ω_{obs})**, the field must remain identical to the original data: $I_{\text{filled}}(\mathbf{x}) = I_{\text{obs}}(\mathbf{x})$.
2. **On the missing domain (Ω_{missing})**, the field must be continuous and smoothly transition from the values at the boundary $\partial\Omega_{\text{obs}}$.

To find a physically plausible solution, we model this infilling task as a steady-state heat conduction process. In physics, heat flux—the flow of heat from warmer to cooler regions—is described by the gradient of the temperature field, ∇I. A system reaches a steady state when the temperature at any point ceases to change over time, which requires the net heat flux into any infinitesimal volume to be zero. This physical condition is mathematically expressed by stating that the divergence of the gradient is zero:

$$\nabla \cdot (\nabla I) = 0, \quad \forall \mathbf{x} \in \Omega_{\text{missing}}$$

The divergence of the gradient is precisely the definition of the Laplacian operator, ∇^2. This leads to our governing equation—the N-dimensional Laplace's Equation:

$$\nabla^2 I(\mathbf{x}) = \sum_{i=1}^{N} \frac{\partial^2 I}{\partial x_i^2} = 0, \quad \forall \mathbf{x} \in \Omega_{\text{missing}}$$

where the equation is subject to a Dirichlet boundary condition, with the values on the boundary of the missing domain, Ω_{missing}, being set by the known values from the surrounding observed domain, Ω_{obs}. The properties of the solution to this partial differential equation are governed by well-established mathematical principles. Specifically, the Maximum Principle guarantees that the solution's maximum and minimum values must lie on the boundary of the domain, while the Mean Value Theorem dictates that the value at any point is the average of its surrounding values (Medková, 2018). These two principles collectively ensure that our infilling result is continuous and does not introduce any local extrema within the missing domain. In fact, solutions to the Laplace equation are infinitely differentiable (i.e., smooth) within the interior of the domain (Medková, 2018).

In this work, we apply this general framework to the 2D spatial domain of ocean temperature fields (i.e., N=2), but the principle is equally applicable to higher-dimensional data. For instance, it can be extended to fill missing volumes in 3D ocean data by solving the 3D Laplace equation, or to obtain a maximally smooth solution in both space and time by solving the 4D Laplace equation for spatio-temporal data.

3.1.2 Application to Ocean Temperature Datasets

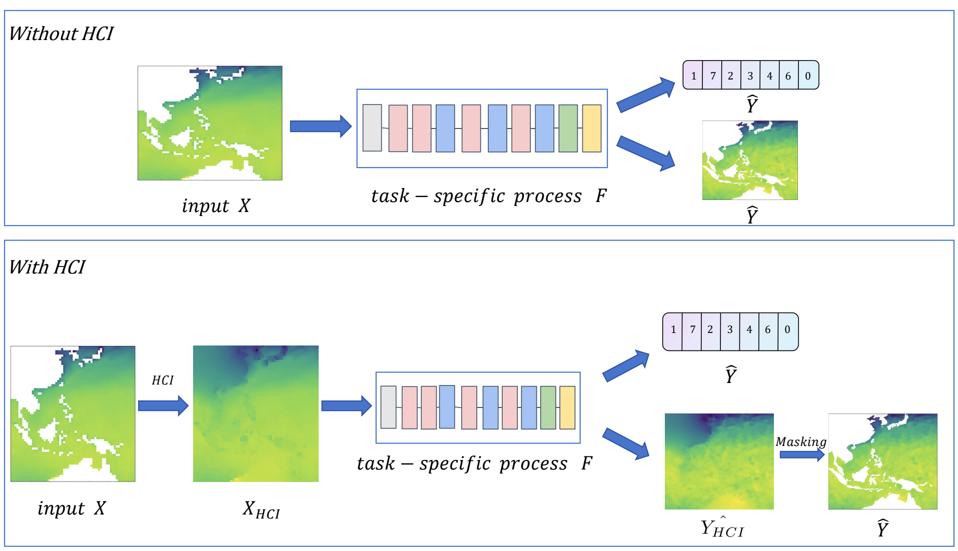


Figure 2: HCI pipeline

We now consider how to apply the aforementioned method to ocean temperature datasets by formulating a unified preprocessing and training pipeline. This pipeline is designed to be model-agnostic and broadly applicable to various downstream tasks. The workflow, as illustrated in Figure 2, consists of two distinct phases:

Training Phase. For any given data sample, including the input features X and the ground truth label Y, we first apply the HCI method to create complete, seamless fields, denoted as X_{HCI} and Y_{HCI}. Given that oceanographic data are typically represented on a latitude-longitude grid, we have developed an efficient implementation of HCI using the Successive Over-Relaxation (SOR) method with Red-Black ordering (Medková, 2018)to solve the underlying system. The processed data is then fed into a **task-specific process**, which refers to a computational model or algorithmic procedure designed for a specific scientific objective (provided it contains CNNs). The output of this computational process can be diverse: for instance, in a field reconstruction or super-resolution task, the output is a 2D or 3D field; whereas in a climate index prediction task, the output might be several numerical values. A key aspect of our framework is that when the process yields a field, the model's loss between $\hat{Y_{HCI}}$ and Y_{HCI} is computed over the entire seamless domain I_{filled}, rather than exclusively on the observed domain Ω_{obs} (the ocean region). This strategy ensures that the model's convolutional filters operate on a continuous and well-behaved data domain at all spatial locations, thereby preventing the model from learning spurious features from the artificial land-sea boundary artifacts and guiding it to focus on the physically meaningful spatial patterns within the actual ocean data.

Application and Evaluation Phase. During the application and evaluation phase, a test input X^{test} is first preprocessed by HCI to generate $X^{\text{test}}_{\text{HCI}}$, which is then fed into the trained task-specific process to obtain an output. When the output is a physical field that requires evaluation, the portion corresponding to Ω_{missing} (the original land area) is discarded. Specifically, the original land-sea mask is used to extract the predicted values only on the observed domain Ω_{obs}. The final performance metrics (e.g., MSE) are then computed between this masked prediction and the ground truth, which is defined on the ocean domain. Thus, the infilled values in Ω_{missing}, having fulfilled their role as a bridge between the ocean dataset and the CNN model, are removed from the final product.

3.2 CLIMATE-STATE CONDITIONING

To address the inherent conflict between the locality inductive bias of CNNs and the non-local teleconnections prevalent in ocean systems, we introduce a climate-state conditioning mechanism. Standard convolutional operations excel at capturing local spatial patterns but struggle to model long-range dependencies, such as the global impacts of the El Niño–Southern Oscillation (ENSO). An unconditional model, blind to the overarching climate state, would erroneously learn an average representation, failing to capture the distinct oceanic patterns characteristic of anomalous climate phases like El Niño or La Niña.

Our solution is to explicitly provide this global context to the model by conditioning the generation process on a macroscopic climate indicator. We select the Niño 3.4 index, a widely recognized metric that quantifies the state of ENSO. By incorporating this index, we transform the learning objective from modeling an unconditional probability distribution $p(\text{ocean field})$ to a conditional one, $p(\text{ocean field}|\text{Niño index})$. This enables the model to learn the specific spatial signatures associated with different climate regimes.

The core idea of our conditioning approach is to inject the global climate information by modulating the intermediate feature maps within the neural network. This is achieved through a mechanism of **feature-wise modulation**(Perez et al., 2018). Specifically, the scalar Niño index is first projected by a small multi-layer perceptron (MLP) to generate a set of channel-wise scale (γ) and shift (β) parameters. These parameters are then applied to the feature maps h at various layers within the main network's convolutional blocks via an affine transformation:

$$h' = \gamma \odot h + \beta$$

where $\odot$ denotes element-wise multiplication. This operation allows the single global climate value to dynamically recalibrate the activations across all spatial locations for each feature channel.

By integrating this mechanism, the model's convolutional filters, which are inherently local operators, become sensitive to the global climate context. This allows the network to learn the distinct spatial characteristics of different ENSO phases. For instance, when presented with a strong positive Niño index, the model is guided to generate patterns consistent with an El Niño event, such as anomalously warm surface waters in the central and eastern tropical Pacific. This approach effectively bridges the gap between the local receptive fields of CNNs and the global-scale influence of climate teleconnections, empowering the model to generate more physically consistent and accurate oceanic fields.

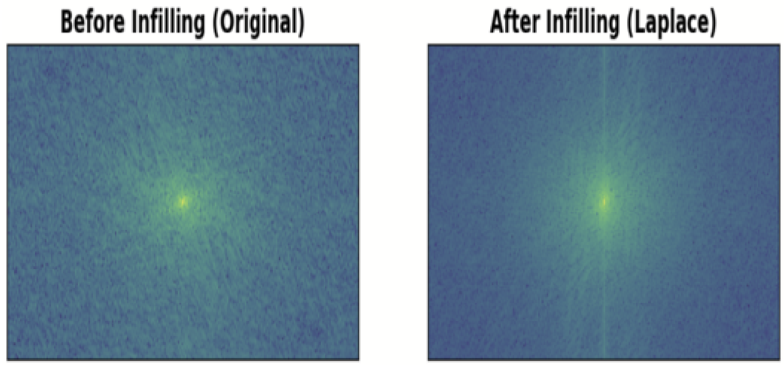


Figure 3: fft comparison

4 EXPERIMENTS

4.1 DATA

Our experiments leverage a combination of numerical simulation outputs and observation-based analysis datasets to train and evaluate our models across three distinct tasks.

For the sea temperature reconstruction and the VAE-based representation learning experiments, the training data is sourced from historical simulations of the Coupled Model Intercomparison Project Phase 6 (CMIP6)(Eyring et al., 2016). Specifically, we utilize monthly sea temperature fields from three Earth System Models: FIO-ESM-2-0(Bao et al., 2020), BCC-CSM2-MR(Wu et al., 2019), and MRI-ESM2-0(Yukimoto et al., 2019), covering the period from 1850 to 2010. To thoroughly assess model robustness and generalization, different training runs were conducted using data from a single model, as well as combinations of two or all three models. For the evaluation and testing phase of these two tasks, we use the EN4 analysis dataset(Good et al., 2013) from the UK Met Office. EN4 is a high-quality, global dataset derived from quality-controlled subsurface ocean profile observations, and it serves as the ground truth for quantifying model performance.

For the sea temperature super-resolution task, our experimental design is aligned with the methodology presented by Izumi et al. (2022). The low-resolution input data for model training is sourced from the NOAA/CIRES/DOE 20th Century Reanalysis (V3) dataset(Slivinski et al., 2019). The high-resolution target data, which serves as the ground truth for both training and evaluation, is the Optimum Interpolation Sea Surface Temperature (OISST) dataset(Reynolds et al., 2007). OISST provides globally gridded, high-resolution (0.25-degree) sea surface temperature fields, making it a standard benchmark for evaluating super-resolution performance in oceanography.

4.2 ANALYSIS OF HCI'S IMPACT ON DATA REPRESENTATION

To further investigate the fundamental impact of Heat-Conduction Infilling (HCI) on the data representation, we conducted a frequency-domain analysis using the 2D Fast Fourier Transform (FFT). Figure3 illustrates the spectral distribution of a representative sea temperature field before and after applying HCI.

Before the application of HCI, the spectrum exhibits significant energy dispersion across a wide range of frequencies. Notably, substantial high-amplitude components are scattered throughout the high-frequency regions (far from the center of the spectrum). These high-frequency signals do not correspond to physical oceanic phenomena; instead, they are spurious artifacts generated by the sharp numerical discontinuities at the land-sea boundaries. These artificial cliffs introduce strong, physically meaningless gradients that dominate the spectral landscape.

In stark contrast, the spectrum of the HCI-processed field shows a dramatic redistribution of energy. The spectral power becomes highly concentrated within the low-frequency domain near the origin. Specifically, the energy is primarily aligned along the central vertical axis, which corresponds to the large-scale, latitudinal temperature gradients (e.g., from the equator to the poles) that are physically characteristic of global ocean temperature distributions. This demonstrates that HCI effectively acts as a low-pass filter, selectively suppressing the high-frequency artifacts associated with land boundaries while preserving the essential, large-scale features of the ocean field.

This qualitative observation from the frequency domain is quantitatively corroborated by a significant reduction in the data's Total Variation (TV). We observed that applying HCI reduces the average TV of the dataset by approximately 60%. This substantial decrease confirms that HCI generates a much smoother and more continuous field by eliminating the most extreme gradients. By removing these high-frequency, non-physical artifacts, HCI simplifies the learning task for Convolutional Neural Networks. It allows the model to focus its capacity on learning physically meaningful oceanic patterns rather than being confounded by spurious boundary signals, thus leading to more efficient training and improved generalization.

4.3 Sea temperature reconstruction

To validate the efficacy of our proposed methods, we conducted experiments on the challenging task of multi-layer sea temperature reconstruction. Our experimental setup, including the U-Net architecture, datasets, and evaluation protocol, strictly follows the state-of-the-art framework, ReconMOST Song et al. (2025). The baseline model in our experiments is a faithful re-implementation of their proposed method. To ensure a fair comparison, the baseline models without our proposed Heat-Conduction Infilling (HCI) were trained for the full 200,000 steps, consistent with the original work. In contrast, due to the substantially accelerated model convergence observed when applying HCI, all HCI-equipped models were trained for just 60,000 steps. This significant reduction in training time provides strong evidence for our method's efficiency.

The quantitative results of our ablation study are presented in Table 1. The analysis clearly demonstrates the profound impact of our contributions. The introduction of Heat-Conduction Infilling (HCI) alone yields a remarkable performance leap across all training configurations. For instance, when trained on the fused dataset, HCI reduces the baseline MSE from 1.31 to 0.21. Our full framework, integrating both HCI and the Climate-State Conditioning Mechanism (CCM), achieves the best overall performance with an MAE of **0.26** and an MSE of **0.20**. This result represents a substantial improvement, drastically lowering the state-of-the-art MSE of 0.648 reported by ReconMOST.

Table 1: Ablation study of sea temperature reconstruction performance on the EN4 test set. The baseline model is our re-implementation of ReconMOST. All metrics are lower-is-better. The best performance is highlighted in bold.

Method	**FIO Dataset**		**BCC Dataset**		**Fused Dataset**	
	MAE	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**
baseline	0.57	1.29	0.59	1.61	0.57	1.31
baseline+CCM	0.46	1.14	0.47	1.30	0.47	1.17
baseline+HCI	0.31	0.33	0.35	0.29	0.28	0.21
baseline+HCI+CCM	0.29	0.33	0.33	0.28	**0.26**	**0.20**

Notably, our framework fundamentally alters the effect of training on diverse datasets. In the original ReconMOST study, fusing data from multiple CMIP6 models did not consistently improve and could even degrade performance, a phenomenon attributed to increased distributional variance (Ben-David et al., 2010). In contrast, our results show that with HCI preprocessing, training on the fused dataset (FIO+BCC+MRI) consistently yields the best performance, surpassing results from any single dataset. This suggests that HCI, by creating a seamless and unified data representation, effectively mitigates the domain mismatch caused by differing land-sea masks and enables the model to learn a more robust and generalized representation from varied sources.

A qualitative analysis of the error distribution reveals that the dramatic reduction in MSE, especially from HCI, is not merely the result of correcting a few extreme outliers. Rather, the improvement is observed to be widespread and relatively uniform across the entire globe and throughout all vertical sea temperature layers. This indicates a fundamental enhancement in the model's ability to capture the underlying physical dynamics of the ocean field, rather than just mitigating isolated anomalies.

Furthermore, we observe a nuanced but important role for the CCM module. While its incremental improvement on the overall performance metrics appears modest after the massive gains from HCI, its benefits are more pronounced under specific climatic conditions. In our detailed analysis of the test set, we found that for months corresponding to strong El Niño events, the addition of CCM led to a more significant reduction in reconstruction error. This highlights its value in enabling the model to accurately capture the distinct spatial signatures associated with anomalous climate states, a capability that global average metrics may not fully reflect.

4.4 Sea temperature super-resolution

In this section, we evaluate the effectiveness of our framework on the task of Single-Image Super-Resolution (SISR) for Sea Surface Temperature (SST) fields. Our experimental methodology is designed to be consistent with the work of Izumi et al. (2022) , employing several established deep learning models for comparison. While adhering to their evaluation protocol, we use the NOAA/CIRES/DOE 20th Century Reanalysis (V3) dataset as the source for our low-resolution inputs, with the high-resolution Optimum Interpolation Sea Surface Temperature (OISST) dataset serving as the ground truth . To rigorously test the models' capabilities on complex coastal geographies, we selected a large evaluation area in the Southeast Asia (SEA) region, performing a 4$\times$ super-resolution from 64$\times$64 pixels to 256$\times$256 pixels.

The performance of various super-resolution models, with and without our proposed framework, is summarized in Table 2. The results indicate that applying Heat-Conduction Infilling (HCI) and Climate-State Conditioning provides a consistent performance improvement across all tested architectures, including sophisticated models like RCAN, RRDBNet, and SANSISR.

The most striking result, however, is observed with the simpler SRCNN model. Izumi et al. (2022) specifically noted that the performance of SRCNN is significantly lower in areas where the data contains land . They attributed this to the model's shallow, three-layer architecture being insufficient to handle the sharp, artificial gradients introduced by missing values over landmasses . Our baseline experiment confirms this finding emphatically; the baseline SRCNN yields a very high MSE of **3.81**, performing substantially worse than the other, deeper models.

Upon applying our HCI and climate conditioning framework, the MSE for SRCNN plummets to **0.917**, a dramatic reduction of over 75%. This improvement is transformative, effectively closing the vast performance gap between SRCNN and much more complex architectures. This strongly validates our central hypothesis: the primary bottleneck for simpler CNNs in this task was not their limited capacity to learn oceanic textures, but their inability to cope with the physically meaningless boundary artifacts. By providing a smooth, continuous, and physically consistent data representation, HCI fundamentally resolves this issue. It makes the learning task more tractable for all models, and proves to be especially revolutionary for simpler architectures that are highly sensitive to data quality.

Table 2: Comparison of MSE for various super-resolution models on the SEA region test set. The baseline models are trained on original data, while the (HCI+CCM) models are trained with our proposed framework. The results for SRCNN are highlighted.

Model	MSE (baseline)	MSE (HCI+CCM)
SRCNN	**3.81**	**0.917**
RCAN	0.803	0.792
RRDBNET	0.800	0.791
SANSISR	0.855	0.801

4.5 Latent Space Representation of Ocean Fields

As stated in our introduction, learning a compact and structured latent space is a crucial step toward enabling more efficient and powerful generative models for complex scientific data. A high-quality latent representation should capture the principal modes of physical variation, not spurious artifacts from the data format. In this section, we investigate whether our proposed HCI method can fundamentally improve the quality of a learned latent space, thereby validating our claim that it creates a data representation more aligned with the inductive biases of deep neural networks.

To test this hypothesis, we leverage the powerful pre-trained Variational Autoencoder (VAE) from the Stable Diffusion 2.1 model, a cornerstone of modern large-scale generative modeling(Rombach et al., 2022). We designed an experiment to fine-tune this VAE on multi-layer ocean temperature data. To adapt our data to the VAE's 3-channel input architecture, we partitioned the vertical ocean layers into groups of three, treating each group as an image-like tensor. The VAE was then fine-tuned on the FIO dataset. A parallel experiment was conducted where the VAE was fine-tuned on the same FIO data preprocessed with HCI. For a rigorous evaluation, we assessed the VAE's reconstruction performance on the most challenging and dynamically complex ocean surface layers from the EN4 analysis dataset. A lower reconstruction Mean Squared Error (MSE) serves as a direct indicator of a more effective and informative latent space.

The results provide compelling evidence for our central thesis. The baseline VAE, fine-tuned on the original data containing sharp land-sea boundaries, achieved a reconstruction MSE of approximately **0.15**. In stark contrast, the VAE fine-tuned on the HCI-processed data achieved a significantly lower MSE of approximately **0.04**. This dramatic reduction in error demonstrates that by eliminating the high-frequency, physically meaningless boundary artifacts, HCI enables the VAE's encoder to learn a much cleaner, more coherent, and physically representative latent space. The model no longer needs to waste its capacity on encoding spurious "numerical cliffs," and can instead focus entirely on the true oceanic patterns. This state-of-the-art reconstruction accuracy underscores the critical importance of our physics-driven preprocessing, proving that it is an essential step for successfully applying large-scale, pre-trained generative models to scientific domains.

5 Conclusion

In this work, we addressed the fundamental inductive bias mismatch between standard deep learning models and oceanographic physical fields. We introduced two synergistic mechanisms: Heat-Conduction Infilling (HCI), a physics-driven preprocessing method to eliminate spurious land-sea boundary artifacts, and a Climate-State Conditioning mechanism to account for non-local teleconnections. By transforming irregularly defined ocean data into a seamless and continuous representation, our framework makes the data inherently more compatible with the architectures of modern neural networks.

Our comprehensive experiments validate the profound impact of this approach. We achieved state-of-the-art performance in multi-layer sea temperature reconstruction, drastically reducing the error of existing models. We also demonstrated that HCI enables simpler super-resolution models to rival complex architectures and empowers large-scale VAEs to learn significantly more accurate and physically coherent latent representations. These results collectively confirm that our approach makes the learning task fundamentally easier for a wide range of deep learning models.

Despite the exceptional effectiveness of HCI as a preprocessing step, it remains an explicit and separate procedure. Future work could explore more integrated, end-to-end solutions. One promising direction is to train a specialized neural network that learns an optimal infilling strategy, potentially capturing more complex data-driven priors than the smooth solution of the Laplace equation. An even more advanced avenue would be to design novel algorithms or model architectures that can directly learn a CNN-friendly data representation from the raw, masked fields without explicit infilling. Such an approach could learn a continuous mapping from the irregular ocean domain to a regular latent grid, offering greater flexibility. Ultimately, our work provides a robust and universally applicable framework that significantly closes the gap between deep learning and oceanography, paving the way for more accurate, efficient, and physically consistent modeling of the Earth's climate system.

References

Egil Bae and Joachim Weickert. Partial differential equations for interpolation and compression of surfaces. In *International Conference on Mathematical Methods for Curves and Surfaces*, pp. 1–14. Springer, 2008.

Ying Bao, Zhenya Song, and Fangli Qiao. Fio-esm version 2.0: Model description and evaluation. *Journal of Geophysical Research: Oceans*, 125(6):e2019JC016036, 2020.

Alexander Barth, Aida Alvera-Azcárate, Matjaz Licer, and Jean-Marie Beckers. Dincae 1.0: A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. *Geoscientific Model Development*, 13(3):1609–1622, 2020.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1):151–175, 2010.

Veronika Eyring, Sandrine Bony, Gerald A Meehl, Catherine A Senior, Bjorn Stevens, Ronald J Stouffer, and Karl E Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016.

Simon A Good, Matthew J Martin, and Nick A Rayner. En4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12):6704–6716, 2013.

Nobuyuki Hirahara, Motoharu Sonogashira, and Masaaki Iiyama. Cloud-free sea-surface-temperature image reconstruction from anomaly inpainting network. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022. doi: 10.1109/TGRS.2021.3111649.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Tomoki Izumi, Motoki Amagasaki, Kei Ishida, and Masato Kiyama. Super-resolution of sea surface temperature with convolutional neural network- and generative adversarial network-based methods. *Journal of Water and Climate Change*, 13(4):1673–1683, 2022. doi: 10.2166/wcc.2022.291.

D Kraemer and G Chen. A simple differential steady-state method to measure the thermal conductivity of solid bulk materials with high accuracy. *Review of Scientific Instruments*, 85(2), 2014.

Michael J McPhaden, Stephen E Zebiak, and Michael H Glantz. Enso as an integrating concept in earth science. *science*, 314(5806):1740–1745, 2006.

Dagmar Medková. The laplace equation. *Boundary value problems on bounded and unbounded Lipschitz domains. Springer, Cham*, 2018.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and F Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.

Richard W Reynolds, Thomas M Smith, Chunying Liu, Dudley B Chelton, Kenneth S Casey, and Michael G Schlax. Daily high-resolution-blended analyses for sea surface temperature. *Journal of climate*, 20(22):5473–5496, 2007.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Meghal Shah, Amit Thakkar, and Hiteshri Shastri. A comparative study of spatial interpolation methods for cmip6 monthly historical and future hydro-climatic datasets for indian region. In *2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)*, volume 1, pp. 1–4. IEEE, 2023.

Laura C Slivinski, Gilbert P Compo, Jeffrey S Whitaker, Prashant D Sardeshmukh, Benjamin S Giese, Chesley McColl, Rob Allan, Xungang Yin, Russell Vose, Holly Titchner, et al. Towards a more reliable historical reanalysis: Improvements for version 3 of the twentieth century reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, 145(724):2876–2908, 2019.

Yuanyi Song, Pumeng Lyu, Ben Fei, Fenghua Ling, Wanli Ouyang, and Lei Bai. Reconmost: Multi-layer sea temperature reconstruction with observations-guided diffusion. *arXiv preprint arXiv:2506.10391*, 2025. doi: 10.48550/arXiv.2506.10391.

Hua Su, Jinwen Jiang, An Wang, Wei Zhuang, and Xiao-Hai Yan. Subsurface temperature reconstruction for the global ocean from 1993 to 2020 using satellite observations and deep learning. *Remote Sensing*, 14(13):3198, 2022. doi: 10.3390/rs14133198.

Zihao Wang and Lei Wu. Theoretical analysis of the inductive biases in deep convolutional networks. *Advances in Neural Information Processing Systems*, 36:74289–74338, 2023.

Tongwen Wu, Yixiong Lu, Yongjie Fang, Xiaoge Xin, Laurent Li, Weiping Li, Weihua Jie, Jie Zhang, Yiming Liu, Li Zhang, et al. The beijing climate center climate system model (bcc-csm): The main progress from cmip5 to cmip6. *Geoscientific Model Development*, 12(4):1573–1600, 2019.

Xiang Wu, Mingxing Zhang, Qiliang Wang, Xiaofeng Wang, Jinnian Chen, and Yan Qin. A deep learning method for inversing 3d temperature fields using sea surface data in offshore china and the northwest pacific ocean. *Journal of Marine Science and Engineering*, 12(12):2337, 2024. doi: 10.3390/jmse12122337.

Seiji Yukimoto, Hideaki Kawai, Tsuyoshi Koshiro, Naga Oshima, Kohei Yoshida, Shogo Urakawa, Hiroyuki Tsujino, Makoto Deushi, Taichu Tanaka, Masahiro Hosaka, et al. The meteorological research institute earth system model version 2.0, mri-esm2. 0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan. Ser. II*, 97(5):931–965, 2019.

Qingyu Zheng, Guijun Han, Wei Li, Lige Cao, Gongfu Zhou, Haowen Wu, Qi Shao, Ru Wang, Xiaobo Wu, Xudong Cui, and Xiaomin Wang. Generating unseen nonlinear evolution in sea surface temperature using a deep learning-based latent space data assimilation framework. *arXiv preprint arXiv:2412.13477*, 2024.