TopoMamba: A Load-Support Relation-Guided Multi-Directional State-Space Model for Topology Optimization

Anonymous authors
Paper under double-blind review

Abstract

Deep learning has emerged as an efficient alternative for predicting high-performance material distributions in topology optimization. Existing methods struggle to accurately capture load-transfer information, limiting out-of-distribution generalization, while their model architectures often incur high computational costs. To address these challenges, this paper proposes TopoMamba, a topology prediction framework incorporating a load-support relation-guided multi-directional state-space model. Coupling physical fields with load-support relations enables more effective modeling of mechanical dependencies. A load-support relation-guided spatially adaptive fusion mechanism dynamically adjusts multi-directional scan features according to spatial conditions. Mamba is coupled with the solid isotropic material with penalty method to enhance structural mechanical performance while maintaining computational efficiency. Results on two-dimensional topology optimization benchmarks demonstrate that TopoMamba achieves superior topology prediction accuracy, out-of-distribution generalization, and computational efficiency over state-of-the-art models. The proposed load-support physics-guided framework enables efficient optimization of more complex structural systems. Codes are available at https://anonymous.4open.science/r/anonymous-ICLR-TM.

1 Introduction

As an advanced design method, topology optimization (TO) aims to obtain high-performance structures by optimizing the distribution of materials under the constraints of load, boundary conditions and material volume(Bendsøe, 1989; Bendsøe & Kikuchi, 1988). Gradient-based TO repeats the finite element analysis and material distribution update during the iterative process, leading to high computational cost. With the increase of grid resolution, the computational cost of TO rises sharply, which limits its widespread application (Träff et al., 2023).

Deep learning has been widely employed to reduce the computational cost of TO by learning an end-to-end mapping from design conditions to optimal material distributions(Woldseth et al., 2022; Shin et al., 2023; Sosnovik & Oseledets, 2019). The existing end-to-end methods include generative methods and deterministic methods. Generative methods offer strong distribution modeling capabilities but often incur high computational costs due to multi-step sampling or iterative refinement(Ho et al., 2020; Song et al., 2021). Deterministic methods enable efficient prediction through a single forward pass, but their generalization to unseen conditions is often limited by the training distribution. Therefore, improving generalization while retaining computational efficiency remains an important challenge.

Achieving this balance requires addressing several limitations in mechanical condition representation, long-range dependency modeling, and computational efficiency. Although the existing methods enhance the characterization of conditions by introducing physical fields(Nie et al., 2021; Mazé & Ahmed, 2023), it is still difficult to fully reflect the mechanical information contained in loads, boundary conditions and their relations. Moreover, the material distribution in TO has a significant spatial dependence. The local material distribution is not only affected by the neighborhood response, but also related to the long-distance load,

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

support and overall force-transfer process. Meanwhile, global self-attention captures long-range dependencies but incurs quadratic computational and memory costs with respect to the number of tokens (Vaswani et al., 2017).

To solve these problems, this paper proposes TopoMamba, a multi-directional state-space topology prediction framework guided by load-support relationship. We combine the physical field and the explicitly constructed load-support relationship to accurately characterize the information of conditions. We further adaptively reweight directional state-space features based on spatial position and load-support configuration, enabling long-range information propagation to better follow force-transfer patterns. Finally, the efficient deterministic prediction and the solid isotropic material with penalty (SIMP) method are combined to improve the mechanical properties of the generated structure while maintaining a low computational cost.

Our main contributions are summarized as follows:(1) an explicit load-support relation representation to enhance the modeling of global mechanical dependencies in topology prediction. (2) a load-support relation-guided spatially adaptive multi-directional state-space model, enabling long-range feature modeling to adapt to different mechanical conditions. (3) an efficient prediction-optimization collaborative topology design framework that further improves structural mechanical performance while achieving a favorable balance between structural quality and computational efficiency.

2 Related Work

2.1 Classical Topology Optimization

As shown in Figure 1, classical topology optimization aims to obtain high-performance structures by optimizing material distributions under prescribed design, loading, boundary, and material constraints, with objectives such as minimizing structural compliance, thermal compliance, or pressure loss.(Bendsøe, 1989; Sigmund & Maute, 2013). Among them, SIMP is one of the most widely used density methods in structural TO (Sigmund, 2001; Andreassen et al., 2011). SIMP relaxes the existence of discrete materials into continuous element density variables, and establishes the relationship between element density and material stiffness through the material interpolation model with penalty. The structural response and the sensitivity of the objective function are obtained according to the finite element analysis, and the gradient-based optimization algorithms such as the optimality criteria or the method of moving asymptotes (Svanberg, 1987) are used to iteratively update the material distribution. Although the classical TO can obtain the material distribution with excellent mechanical properties, the solution process usually needs to re perform the finite element analysis and sensitivity calculation after each design variable update. Therefore, with the increase of design freedom and problem complexity, a large number of repeated numerical solutions will bring higher computational costs.

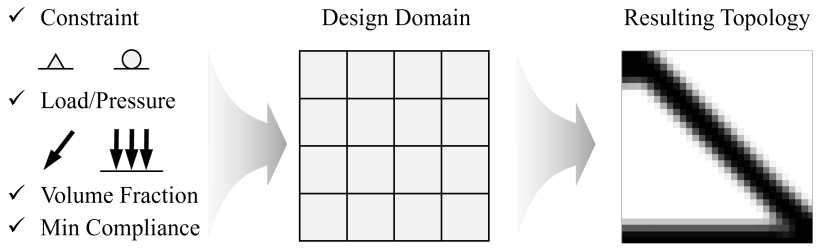


Figure 1: TO. TO introduces the design domain, load and support conditions, and volume fraction as input conditions. The optimization process iteratively updates the material distribution to minimize compliance while satisfying the volume fraction constraint.

2.2 Deep Learning in TO

The application of deep learning in TO is mainly carried out along two paths: one is to use neural network to approximate finite element analysis, so as to reduce the repeated calculation cost (Woldseth et al., 2022); The other class directly learns the mapping between the design conditions and the optimal material distribution to achieve fast end-to-end topology prediction. In recent years, end-to-end methods have gradually become an important direction of learning TO, and have developed different prediction paradigms based on Gan, diffusion model and neural implicit representation.(Nie et al., 2021; Mazé & Ahmed, 2023; Bastek et al., 2025; Nobari et al., 2025; Yang et al., 2026; Zheng et al., 2021).

Learning-based TO is gradually changing from only using design conditions such as load, boundary conditions and volume fraction to further using physical fields. Topologygan (Nie et al., 2021) takes the physical field in the initial design domain as conditional information to enhance topology prediction; Topodiff (Mazé & Ahmed, 2023) further combines conditional diffusion and physical performance to guide the generation of topology that meets the design requirements; Recent topotransformer (Rashed et al., 2026) further shows that the physical fields related to sensitivity plays an important role in generalization without working conditions. However, physical fields can not fully capture the mechanical interactions among loads and support conditions.

2.3 State-space Models

Transformer relies on self attention mechanism to achieve global feature interaction and can effectively model long-range dependencies (Dosovitskiy et al., 2021; Vaswani et al., 2017). However, its computational and storage overhead usually increases with the square of the sequence length, which will bring a large computational burden in high-resolution topology prediction. The state-space model provides another efficient paradigm for long-range dependency modeling (Gu et al., 2022). Mamba introduces a selective mechanism based on the structured state-space model, which enables the model to dynamically control the retention and propagation of information according to the input content, and reduces the complexity of sequence modeling to a linear scale with respect to sequence length (Yang et al., 2024; Li et al., 2024). Subsequently, vision Mamba and others further extended the mechanism to two-dimensional visual tasks, enabling the state-space model to realize large-scale spatial feature interaction at a low computational cost (Zhu et al., 2024). Vmamba further propagates and aggregates features along multiple spatial directions through two-dimensional selective scanning, enabling different locations to obtain more sufficient long-range context information (Liu et al., 2024).In this study, mamba is applied to the field of TO, and modified based on the TO task, without attention mechanism.

3 Method

3.1 Preliminaries

3.1.1 Topology Optimization

The TO problem can be written as follows:

$$\begin{aligned} \min_{\boldsymbol{\rho}}: \quad & c(\boldsymbol{\rho}) = \mathbf{U}^T \mathbf{K}(\boldsymbol{\rho}) \mathbf{U} \\ \text{subject to:} \quad & \frac{V(\boldsymbol{\rho})}{V_0} \leq v, \\ & \mathbf{K}(\boldsymbol{\rho})\mathbf{U} = \mathbf{F}, \\ & \rho_{\min} \leq \rho_e \leq 1, \qquad e = 1, \ldots, N. \end{aligned} \tag{1}$$

where c denotes the structural compliance, $\mathbf{U}$ is the global displacement vector, $\mathbf{K}$ is the global stiffness matrix, $\mathbf{F}$ is the global force vector, and ρ is the element density vector. The volume constraint requires the ratio of the material volume $V(\boldsymbol{\rho})$ to the design domain volume V_0 to be no greater than the target volume fraction v. The density variable ρ_e for

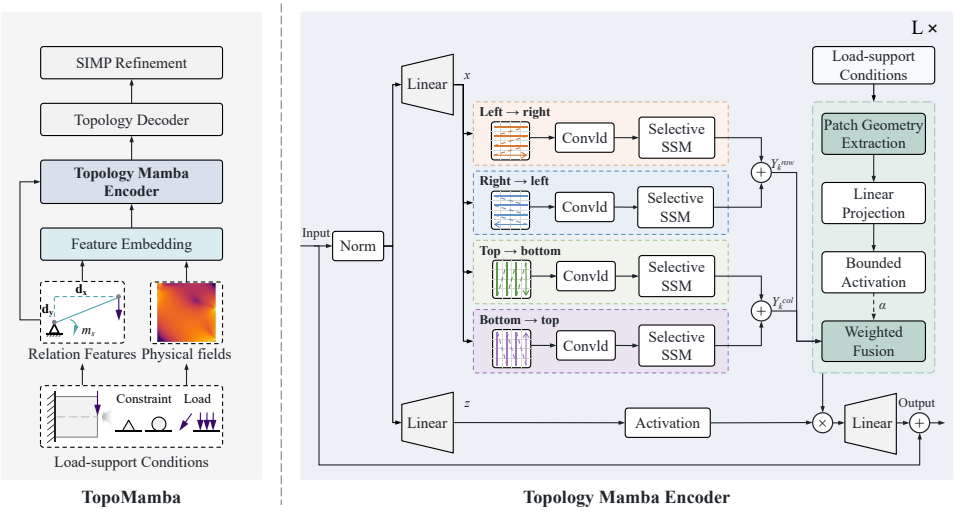


Figure 2: Overview of the proposed TopoMamba framework. The framework illustrates the complete workflow from prescribed loading and boundary conditions to SIMP refinement. The encoder models features along four spatial directions using selective state-space modules and adaptively fuses the directional features with spatial weights conditioned on the load-support relations.

each element e is bounded between $\rho_{\min}$ and 1, and N is the total number of elements in the design domain.

Based on SIMP method, the Young's modulus of element e is interpolated as

$$E_e(\rho_e) = E_{\min} + \rho_e^p \left(E_0 - E_{\min}\right), \tag{2}$$

where E_0 denotes the Young's modulus of the solid material, $E_{\min}$ is a small stiffness introduced to avoid singularity of the global stiffness matrix, and p is the penalization factor to discourage intermediate densities.

3.1.2 State-Space Models and Mamba

A continuous linear state-space model maps an input $x(t)$ to an output $y(t)$ through a hidden state $\mathbf{h}(t)$:

$$\dot{\mathbf{h}}(t) = \mathbf{A}\mathbf{h}(t) + \mathbf{B}x(t), \qquad y(t) = \mathbf{C}\mathbf{h}(t) + \mathbf{D}x(t). \tag{3}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are the state-space parameters. Mamba introduces a selective mechanism with input-dependent parameters, enabling adaptive information propagation along a sequence. For two-dimensional visual data, image features can be organized into sequences along spatial scanning paths, allowing Mamba to capture long-range spatial dependencies.

3.2 Architecture

The overall network architecture of TopoMamba is illustrated in Figure 2. It adopts an encoder-decoder architecture that embeds the input conditions into spatial features, processes them with stacked state-space blocks, and decodes them into a continuous material-density field, which is subsequently refined by SIMP.

Given the design conditions $\mathbf{c}$, we first construct two complementary condition representations: a physics-conditioned representation $\mathbf{X}$ and an explicit load–support relation representation $\mathbf{R}$. These representations are subsequently encoded and fused into an element-wise

feature map:

$$\mathbf{S} = \mathcal{E}_\theta(\mathbf{X}, \mathbf{R}), \tag{4}$$

where $\mathbf{S} \in \mathbb{R}^{H\times W\times C}$ denotes the resulting spatial feature representation.

To construct the input sequence for the state-space backbone, $\mathbf{s}$ is partitioned into non-overlapping $P \times P$ patches. Each patch is flattened and linearly projected into a D-dimensional token. The resulting token sequence is

$$\mathbf{T}_0 = \left[\mathbf{s}_p^1\mathbf{W}_p; \mathbf{s}_p^2\mathbf{W}_p; \ldots; \mathbf{s}_p^J\mathbf{W}_p\right] + \mathbf{E}_{\text{pos}}, \tag{5}$$

where $\mathbf{s}_p^j \in \mathbb{R}^{P^2C}$ denotes the j-th flattened patch, $\mathbf{W}_p \in \mathbb{R}^{(P^2C)\times D}$ is the learnable projection matrix, $J = HW/P^2$ is the number of patches, and $\mathbf{E}_{\text{pos}} \in \mathbb{R}^{J\times D}$ is the positional embedding. The token sequence is then processed by the state-space backbone:

$$\begin{aligned} \mathbf{T}_l &= \text{Mamba}\left(\mathbf{T}_{l-1}; \boldsymbol{\alpha}(\mathbf{R})\right) + \mathbf{T}_{l-1}, \\ \widehat{\boldsymbol{\rho}} &= \text{sigmoid}\left[-\text{Unpatchify}\left(\text{Linear}\left(\text{LN}(\mathbf{T}_l)\right)\right)\right], \end{aligned} \tag{6}$$

where $\boldsymbol{\alpha}(\mathbf{R})$ contains the fusion weights conditioned on the design problem, L is the number of mamba blocks, and $\mathbf{T}_l$ is the output of the l-th layer of the mamba block. The network is trained with a composite loss function that combines a binary cross entropy loss $\mathcal{L}_{\text{BCE}}$ between the predicted and reference topologies, volume fraction loss $\mathcal{L}_{\text{vol}}$, and the floating material loss $\mathcal{L}_{\text{floating}}$:

$$\mathcal{L} = \mathcal{L}_{\text{BCE}} + \lambda_v\mathcal{L}_{\text{vol}} + \lambda_f\mathcal{L}_{\text{floating}}. \tag{7}$$

Here, λ_v and λ_f denote the weighting coefficients for the volume fraction and floating material losses, respectively.

3.3 Load-Support Relation Embedding

Following prior learning-based methods (Mazé & Ahmed, 2023; Nie et al., 2021), we adopt strain energy density and von Mises stress as physical-field inputs obtained from finite-element analysis. While these fields characterize the structural response under the prescribed conditions, they do not explicitly represent the spatial relation between the load and supports. We therefore construct the explicit load-support relation representation.

With coordinates normalized to the unit domain, let $\boldsymbol{\ell}$, $\mathbf{f}$, and $\mathbf{s}$ denote the load position, applied force, and support position, respectively. We define

$$\begin{aligned} \mathbf{d}_s &= \boldsymbol{\ell} - \mathbf{s}, & \widehat{\mathbf{f}} &= \frac{\mathbf{f}}{\|\mathbf{f}\|_2}, \\ m_s &= d_{s,x}\widehat{f}_y - d_{s,y}\widehat{f}_x, & r_s^2 &= \|\mathbf{d}_s\|_2^2. \end{aligned} \tag{8}$$

For each support node $\mathbf{s}$, let $b_x(\mathbf{s})$ and $b_y(\mathbf{s})$ indicate whether the node is constrained along the x- and y-directions, respectively. The load–support relation feature is defined as

$$\mathbf{R}(\mathbf{s}) = \left[b_x(\mathbf{s})\,\mathbf{r}(\mathbf{s}); b_y(\mathbf{s})\,\mathbf{r}(\mathbf{s})\right], \tag{9}$$

where

$$\mathbf{r}(\mathbf{s}) = \left[1,\, d_{s,x},\, d_{s,y},\, \widehat{f}_x,\, \widehat{f}_y,\, m_s,\, r_s^2\right]. \tag{10}$$

3.4 Load-Support Relation-Guided Scan Fusion

The material distribution in TO is strongly influenced by the relative positions of loads and supports. Since their effects can vary across spatial directions under different design conditions, uniformly combining directional features may be suboptimal. We therefore use the load–support relations introduced in Sec. 3.3 to adaptively fuse multi-directional state-space features.

Let $\mathbf{T}_{k-1}$ denote the input feature grid of the k-th TopoMamba block. We apply selective state-space scans along four spatial directions, $\{\rightarrow, \leftarrow, \downarrow, \uparrow\}$. Opposite directions are aggregated into row-wise and column-wise representations:

$$\mathbf{Y}_k^{\text{row}} = \mathcal{S}_k^{\rightarrow}(\mathbf{T}_{k-1}) + \mathcal{S}_k^{\leftarrow}(\mathbf{T}_{k-1}), \qquad \mathbf{Y}_k^{\text{col}} = \mathcal{S}_k^{\downarrow}(\mathbf{T}_{k-1}) + \mathcal{S}_k^{\uparrow}(\mathbf{T}_{k-1}), \tag{11}$$

where $\mathcal{S}_k^d$ denotes the selective state-space scan along direction d, with all outputs aligned to their original patch locations.

For each patch position $\mathbf{q}$, we extract local relation features $\mathbf{g}(\mathbf{q};\mathbf{R})$ from the load–support representation $\mathbf{R}$ introduced in Sec. 3.3. A linear projection followed by a bounded activation produces a spatially varying fusion weight $\alpha_k(q) \in [0,1]$. The final representations are then combined as

$$\mathbf{Y}_k(\mathbf{q}) = \alpha_k(\mathbf{q};\mathbf{R})\mathbf{Y}_k^{\text{row}}(\mathbf{q}) + [1-\alpha_k(\mathbf{q};\mathbf{R})]\,\mathbf{Y}_k^{\text{col}}(\mathbf{q}). \tag{12}$$

Thus, the state-space scan paths remain fixed, while their relative contributions vary across spatial locations according to the load–support configuration.

3.5 SIMP Refinement

TopoMamba predicts a continuous material-density field, while the final topology is binary. Since structural compliance is not directly optimized by the loss, the prediction may still contain suboptimal regions; moreover, direct thresholding introduces a mismatch between the continuous representation used during training and the discrete topology required at inference. To address both issues, we apply a small number of SIMP iterations before binarization, using compliance sensitivities and the prescribed volume constraint to refine the predicted density field in the continuous design space. This refinement provides local corrections while preserving the efficiency of the learned prediction.

$$\boldsymbol{\rho}^{(0)} = \widehat{\boldsymbol{\rho}}, \qquad \boldsymbol{\rho}^{(j+1)} = \mathcal{U}_{\mathbf{c}}\left(\boldsymbol{\rho}^{(j)}\right), \quad j = 0,\ldots,K-1, \tag{13}$$

each update uses the current structural response and compliance sensitivities to correct mechanically important regions under the prescribed volume constraint. Binarization is applied only after refinement, so the network prediction and the physics-based updates remain in the same continuous design space.

4 Experiments

In this section, we evaluate the performance of TopoMamba on two-dimensional structural TO benchmarks. We compare TopoMamba with state-of-the-art deterministic and generative models in terms of compliance error, volume fraction error, and end-to-end runtime. We also conduct ablation studies to assess the contributions of key components.

4.1 Experimental Setup

4.1.1 Dataset and Baselines

We use the two-dimensional structural TO benchmarks from TopoDiff and NITO. The dataset includes 30,000 training samples and 1,800 in-distribution (ID) conditions samples for testing our models. Additionally, we include a dataset of 1,000 out-of-distribution (OOD) samples for generalization testing. We compare TopoMamba with the baseline of deterministic models and generative models.

4.1.2 Evaluation Metrics

In line with NITO and TopoDiff, we evaluate the models using three metrics, as follows:

Compliance Error. Structural performance is evaluated using the compliance error (CE), which compares the compliance of the predicted topology with that of a reference design obtained by SIMP under the same loading and boundary conditions. For the i-th sample, CE is defined as

$$\text{CE}_i = \frac{C(\hat{y}_i) - C(y_i)}{C(y_i)}, \tag{14}$$

where $C(\hat{y}_i)$ denotes the compliance of the predicted topology and C_i^{ref} denotes the compliance of the corresponding SIMP-optimized reference design. We report both the mean and median CE to evaluate the compliance performance of each model.

Volume Fraction Error. We further evaluate constraint satisfaction using the volume fraction error(VFE), which quantifies the relative deviation between the actual volume fraction of the generated topology and the prescribed target volume fraction:

$$\mathrm{VFE}_i = \frac{|VF(\hat{y}_i) - VF(y_i)|}{VF(y_i)}, \tag{15}$$

where, $VF(\hat{y}_i)$ denotes the actual volume fraction of the predicted topology, while $VF(y_i)$ denotes the target volume fraction.

End-to-End Runtime. We measure the end-to-end runtime required by each method to generate the final topology from a given problem specification. For comparison, we also report the runtime of the standard SIMP optimizer under the same problem setting. All experiments are conducted on an NVIDIA RTX 5880 Ada 48 GB GPU and an AMD EPYC 9654 CPU.

We exclude samples with compliance errors exceeding 1000%, consistent with a common practice in prior SOTA models (Mazé and Ahmed, 2023; Giannone et al., 2023; Nie et al., 2021b) for comparsion purposes. Such cases typically correspond to failed predictions in which the model does not produce a meaningful material distribution or satisfy the prescribed constraints.

4.2 Performance

Table 1: Performance comparison on ID tests. FS: the number of SIMP refinement steps. w/ G: using a classifier and regression guidance. HPG-Diff* and TopoViT* results from Yang et al. (2026) and Lutheran et al. (2026)

Method	FS	Mean CE (%)	Median CE (%)	VFE (%)
TopologyGAN	–	48.51	2.06	11.87
TopoDiff	–	3.23	0.45	1.14
TopoDiff w/ G	–	2.59	0.49	1.18
HPG-Diff*	–	0.87	0.16	1.38
NITO	5	0.81	0.098	0.5
NITO	10	0.48	0.067	0.41
TopoViT*	–	3.67	0.91	4.03
TopoMamba	–	2.91	0.27	0.96
TopoMamba	5	0.43	0.13	0.50
TopoMamba	10	**0.17**	0.074	**0.29**

Table 2: Performance comparison on OOD tests.

Method	FS	Parameters (M)	Mean CE (%)	Median CE (%)
TopoDiff	–	121	8.57	1.14
TopoDiff w/ G	–	239	7.79	1.26
TopoTransformer	–	34	5.73	0.53
NITO	5	22	9.33	2.37
TopoMamba	5	**7**	1.79	0.28
TopoMamba	10	**7**	**1.07**	**0.14**
HPG-Diff	–	177	5.29	0.61

Table 1 presents the performance of different methods on ID tests . TopoMamba achieves low errors in terms of mean compliance error, median compliance error, and volume fraction error, demonstrating favorable mechanical performance and constraint satisfaction. Its best results reach 0.17% Mean CE, 0.074% Median CE, and 0.29% VFE, respectively. Notably, even without SIMP refinement, TopoMamba already achieves performance comparable to several existing state-of-the-art methods. Further refinement leads to additional improvements across the evaluation metrics, indicating both the high quality of the initial designs generated by TopoMamba and their potential for further optimization.

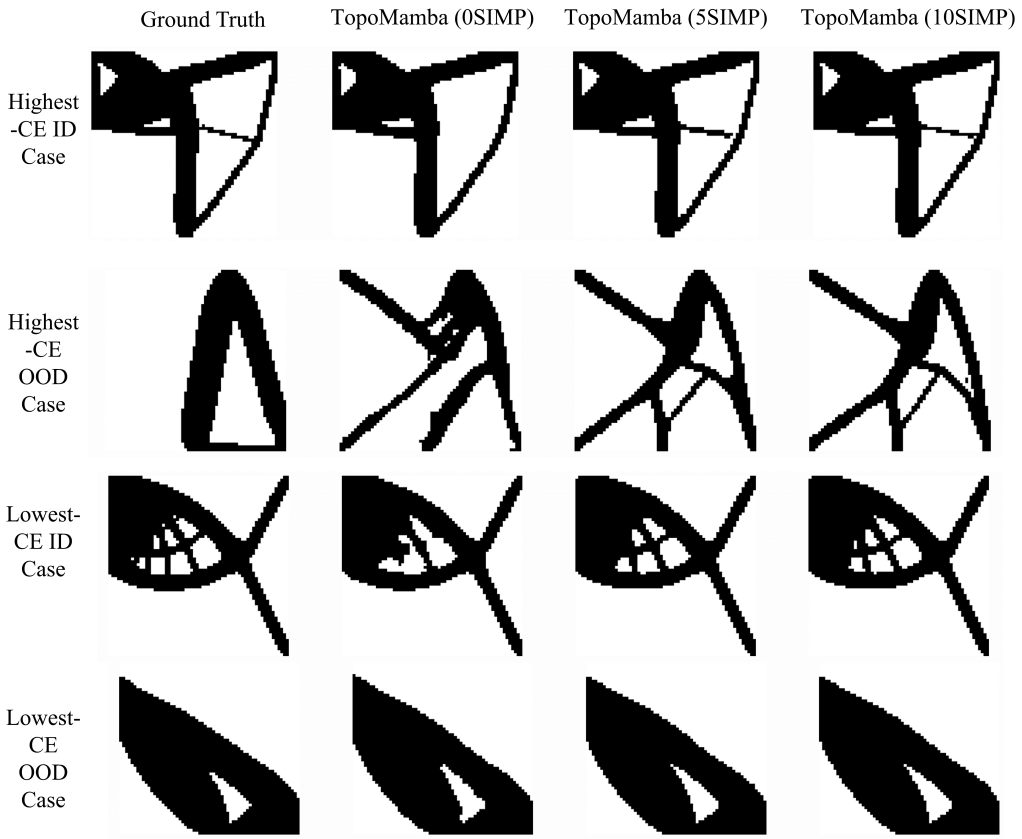


Figure 3: Qualitative comparison of TopoMamba before and after SIMP refinement on representative ID and OOD cases.

Table 2 shows the performance of different methods on OOD tests. TopoMamba achieves state-of-the-art performance while using only 7M parameters, demonstrating strong generalization capability and high parameter efficiency. Its best results reach 1.07% Mean CE and 0.14% Median CE, outperforming the compared methods on both metrics. These results highlight the effectiveness of TopoMamba in maintaining accurate and robust performance under distribution shifts, while requiring substantially fewer model parameters.

Figure 3 presents qualitative comparisons of TopoMamba before and after SIMP refinement on representative ID and OOD cases. The results demonstrate that TopoMamba can generate high-quality initial topologies that closely resemble the reference designs, even under challenging OOD conditions. The subsequent SIMP refinement further enhances the structural quality, yielding final designs with improved compliance and adherence to volume constraints. These visualizations underscore the effectiveness of TopoMamba in producing mechanically sound and visually coherent topologies across diverse scenarios.

End-to-End Runtime. Table 3 compares the end-to-end runtime for obtaining the final topology from a given problem specification, including physical field preparation, model inference, and numerical refinement. TopoMamba achieves an average end-to-end runtime of 0.48 s, substantially faster than TopoTransformer and HPG-Diff, while remaining close to the fastest learning-based method, NITO. This demonstrates that TopoMamba provides high computational efficiency in addition to its strong predictive performance.

Table 3: Averaged End-to-end times of tests.

Method	Time (s)
Standard SIMP	10.59
NITO	0.40
TopoTransformer	0.79
HPG-Diff	2.79
TopoMamba	0.48

4.3 Ablation Studies

We conduct controlled ablation studies to examine the effectiveness of the key architectural designs in TopoMamba. In addition to removing the physical-field input as a reference, we focus on the components directly associated with our main methodological contributions, including physical fields, the explicit load-support relation representation, and adaptive feature fusion. The explicit relation representation is designed to provide the model with structured mechanical dependencies between loads and supports, while the multi-directional scanning mechanism facilitates long-range dependency modeling along different spatial directions. The adaptive fusion module further enables these directional features to be dynamically integrated according to the underlying mechanical conditions. As shown in Table 4, removing any of these components consistently degrades performance, with a more evident deterioration under the OOD tests. This trend indicates that the proposed designs provide benefits: explicit relation modeling introduces mechanically meaningful structural priors, multi-directional scanning improves long-range spatial dependency modeling, and adaptive fusion enhances the model's ability to adjust feature aggregation across varying load and support configurations. Together, these components contribute to the robustness and generalization capability of TopoMamba under distribution shifts.

Table 4: Ablation study on ID and OOD tests.

Mean CE (%)	TopoMamba (5SIMP)	w/o Physical Fields	w/o Load–support Relations	w/o Adaptive Fusion
ID	0.43	1.24	0.49	0.47
OOD	1.79	10.00	1.83	1.91

5 Conclusion

This paper presents TopoMamba, a load-support relation-guided multi-directional state-space framework for topology prediction. TopoMamba achieves strong predictive performance and generalization with a compact parameterization and low computational cost, outperforming existing state-of-the-art methods on the benchmarks. These results demonstrate the potential of deterministic learning-based models to provide an effective balance between structural quality, generalization, and inference efficiency, and offer a promising direction for accelerating TO in high-dimensional design spaces and computationally intensive iterative settings.

Although our methods perform well on current benchmarks, their generalization to more complex design conditions requires further improvement. Future work should enhance generalization while preserving low inference cost, and extend the framework beyond 2D regular grids toward irregular discretizations, 3D structures.

AI use statement

In this paper, we used generative AI tools, specifically ChatGPT (OpenAI, GPT-6), for language polishing, generating LaTeX table templates, and assisting with code development, including the limited generation of auxiliary code.

We did not use generative AI tools to formulate the research motivation or research ideas, design the algorithms or experiments, collect data, analyze or interpret experimental results,

or derive the conclusions. The algorithms and core implementation were independently developed by the authors, and the remaining required disclosure tasks are not applicable to this paper.

We reviewed all AI-assisted work. All research content was developed by the authors, and all experimental results were produced and verified by the authors. We reviewed all AI-assisted language revisions and LaTeX formatting to ensure that they preserved the intended technical meaning and accurately represented the original experimental results. Any AI-generated code or code suggestions were manually reviewed, modified where necessary, and tested for correctness by the authors before being incorporated into the final implementation.

We take responsibility for the final content of this work, including all text, code, claims, tables, and other artifacts produced or revised with the aid of generative AI.

Ethics statement

This research does not involve human subjects, animal experiments, or sensitive datasets, nor does it develop or deploy applications that pose specific ethical risks. We therefore do not identify any specific ethical concerns arising from this paper.

Reproducibility statement

The methodological details of this paper are provided in Sections 3.1–3.5. The supplementary material complements these sections with an implementation-oriented description of TopoMamba, including the end-to-end data flow and input representations, physical-field construction and normalization, coordinate conventions, the structure and responsibilities of the principal model components, the complete model configuration and parameter count, and the prediction and optional SIMP-refinement settings. Together, the main paper and supplementary material document the model architecture, input-processing workflow, and principal implementation configurations used in the reported experiments.

Acknowledgments

This paper is supported by the National Science Foundation for Distinguished Young Scholars of China (Grant No. 52525505), the National Postdoctoral Program for Innovative Talents (Grant No. BX20250423), Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (Grant No. JYB2025XDXM208).

References

Erik Andreassen, Anders Clausen, Mattias Schevenels, Boyan S. Lazarov, and Ole Sigmund. Efficient topology optimization in MATLAB using 88 lines of code. Structural and Multidisciplinary Optimization, 43:1–16, 2011. doi: 10.1007/s00158-010-0594-7.

Jan-Hendrik Bastek, WaiChing Sun, and Dennis M. Kochmann. Physics-informed diffusion models. In International Conference on Learning Representations, 2025. URL https://openreview.net/forum?id=tpYeermigp.

Martin P. Bendsøe. Optimal shape design as a material distribution problem. Structural Optimization, 1:193–202, 1989. doi: 10.1007/BF01650949.

Martin P. Bendsøe and Noboru Kikuchi. Generating optimal topologies in structural design using a homogenization method. Computer Methods in Applied Mechanics and Engineering, 71(2):197–224, 1988. doi: 10.1016/0045-7825(88)90086-2.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In International Conference on Learning Representations, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In International Conference on Learning Representations, 2022. URL https://openreview.net/forum?id=uYLFoz1vIAC.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. URL https://arxiv.org/abs/2006.11239.

Shufan Li, Harkanwar Singh, and Aditya Grover. Mamba-ND: Selective state space modeling for multi-dimensional data. In Computer Vision – ECCV 2024, pp. 75–92, 2024. URL https://www.ecva.net/papers/eccv_2024/papers_ECCV/html/4811_ECCV_2024_paper.php.

Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. VMamba: Visual state space model. In Advances in Neural Information Processing Systems, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/baa2da9ae4bfed26520bb61d259a3653-Abstract.html.

Aaron Lutheran, Srijan Das, and Alireza Tabarraei. Physics-informed transformer for real-time high-fidelity topology optimization, 2026. URL https://arxiv.org/abs/2604.03522.

François Mazé and Faez Ahmed. Diffusion models beat GANs on topology optimization. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, pp. 9108–9116, 2023. doi: 10.1609/aaai.v37i8.26093. URL https://ojs.aaai.org/index.php/AAAI/article/view/26093.

Z. Nie, T. Lin, H. Jiang, and L. B. Kara. TopologyGAN: Topology optimization using generative adversarial networks based on physical fields over the initial domain. Journal of Mechanical Design, 143(3), 2021. doi: 10.1115/1.4049533. URL https://arxiv.org/abs/2003.04685.

Amin Heyrani Nobari, Lyle Regenwetter, Giorgio Giannone, and Faez Ahmed. NITO: Neural implicit fields for resolution-free and domain-adaptable topology optimization. Transactions on Machine Learning Research, 2025. URL https://openreview.net/forum?id=XHXAvACdgv.

Mohammad Rashed, Duarte F. Valoroso Madeira, Babak Gholami, Caglar Guerbuez, Yunjia Yang, and Nils Thuerey. On the generalization in topology optimization via sensitivity-conditioned Bernoulli flow matching, 2026. URL https://arxiv.org/abs/2606.02179. arXiv preprint arXiv:2606.02179.

Seungyeon Shin, Dongju Shin, and Namwoo Kang. Topology optimization via machine learning and deep learning: A review. Journal of Computational Design and Engineering, 10(4):1736–1766, 2023. doi: 10.1093/jcde/qwad072.

Ole Sigmund. A 99 line topology optimization code written in Matlab. Structural and Multidisciplinary Optimization, 21:120–127, 2001. doi: 10.1007/s001580050176.

Ole Sigmund and Kurt Maute. Topology optimization approaches: A comparative review. Structural and Multidisciplinary Optimization, 48:1031–1055, 2013. doi: 10.1007/s00158-013-0978-6.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2021. URL https://openreview.net/forum?id=St1giarCHLP.

Ivan Sosnovik and Ivan Oseledets. Neural networks for topology optimization. Russian Journal of Numerical Analysis and Mathematical Modelling, 34(4):215–223, 2019. doi: 10.1515/rnam-2019-0018. URL https://arxiv.org/abs/1709.09578.

Krister Svanberg. The method of moving asymptotes—a new method for structural optimization. International Journal for Numerical Methods in Engineering, 24(2):359–373, 1987. doi: 10.1002/nme.1620240207.

Erik A. Träff, Anton Rydahl, Sven Karlsson, Ole Sigmund, and Niels Aage. Simple and efficient gpu accelerated topology optimisation: Codes and applications. Computer Methods in Applied Mechanics and Engineering, 410:116043, 2023. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2023.116043.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Rebekka V. Woldseth, Niels Aage, J. Andreas Bærentzen, and Ole Sigmund. On the use of artificial neural networks in topology optimisation. Structural and Multidisciplinary Optimization, 65:294, 2022. doi: 10.1007/s00158-022-03347-1. URL https://doi.org/10.1007/s00158-022-03347-1.

Chenhongyi Yang, Zehui Chen, Miguel Espinosa, Linus Ericsson, Zhenyu Wang, Jiaming Liu, and Elliot J. Crowley. PlainMamba: Improving non-hierarchical Mamba in visual recognition. In British Machine Vision Conference, 2024. URL https://bmva-archive.org.uk/bmvc/2024/papers/Paper_755/paper.pdf.

Jinbo Yang, Mingyue Yuan, Boyuan Zhang, Yoshifumi Kitamura, and Shikai Jing. HPG-Diff: Hierarchical physics-guided diffusion with differentiable connectivity constraints for topology optimization. Applied Soft Computing, pp. 115899, 2026. doi: 10.1016/j.asoc.2026.115899. URL https://arxiv.org/abs/2607.07233.

Shuai Zheng, Zhenzhen He, and Honglei Liu. Generating three-dimensional structural topologies via a u-net convolutional neural network. Thin-Walled Structures, 159:107263, 2021. ISSN 0263-8231. doi: https://doi.org/10.1016/j.tws.2020.107263.

Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision Mamba: Efficient visual representation learning with bidirectional state space model. In Proceedings of the 41st International Conference on Machine Learning, volume 235 of Proceedings of Machine Learning Research, pp. 62429–62442. PMLR, 2024. URL https://proceedings.mlr.press/v235/zhu24f.html.