

Cauchyformer: Few-Shot Spectrum Inference for Photonic Metasurfaces via Mixture of Cauchy

Shujie Yang,[†] Xuzhe Zhao,[†] Chunjiang Li, Yuxiao Li, Hongyan Fu, Yansong Tang,* and Kaichen Dong*



Cite This: *ACS Omega* 2026, 11, 49510–49520



Read Online

ACCESS |



Metrics & More

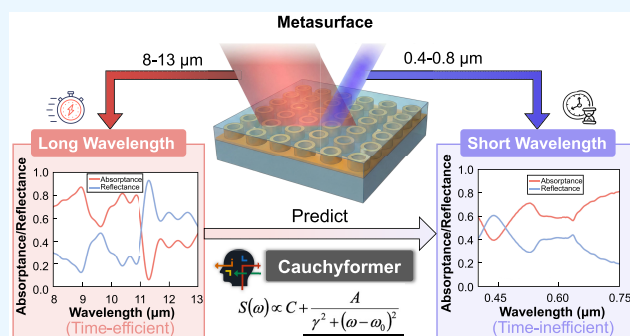


Article Recommendations



Supporting Information

ABSTRACT: Inferring hard-to-access spectra from simpler measurements is a fundamental challenge across photonics, materials science, chemistry, and biology. Intrinsic cross-frequency correlations encode rich information about structure, composition, and dynamics; utilizing these relations enables rapid characterization, screening, and design without expensive iterative simulations or experiments. A representative and highly demanding setting is metasurface photonics, where structural optimization relies on large-scale partial differential equation (PDE) solvers, rendering ultrabroadband full-wave electromagnetic simulations prohibitively expensive. Existing time-series and signal forecasting approaches typically require large training data sets, yet spectral data is often scarce and costly to acquire. Moreover, these data-driven models lack explicit awareness of the underlying spectral physics. In this work, we introduce **Cauchyformer**, a physics-informed framework for accurate few-shot spectra-to-spectra inference. At its core, Cauchyformer employs a Mixture-of-Cauchy (MoC) mechanism that explicitly encodes Cauchy–Lorentz resonances, representing spectra as superpositions of Cauchy basis functions. Combined with patching and bidirectional attention modules to capture local and global dependencies across frequencies and channels, Cauchyformer enables accurate spectrum inference from minimal training data. Experiments on simulated and fabricated metasurfaces demonstrate state-of-the-art performance in fast and reliable prediction of resource-intensive high-frequency spectra from more accessible low-frequency optical responses, even with scarce training samples. Owing to its generic architecture, Cauchyformer can be readily adapted to a wide range of multispectral inference and spectrometry tasks, including material characterization, environmental sensing, and proteomics.



1. INTRODUCTION

Spectrum inference,¹ the process of deducing hard-to-measure spectral responses from accessible spectral features by leveraging the intrinsic consistency of material properties, has emerged as an essential technique across multiple scientific communities. By utilizing cross-frequency correlations that originate from shared underlying physical and chemical mechanisms—often reflecting common dispersion constraints and resonance behaviors across bands—this approach enables the prediction of complex or otherwise inaccessible spectral regions from readily measured optical features. The applications of spectrum inference are vast and inherently interdisciplinary. In biomedical applications,^{2–4} inferring the optical properties of tissues facilitates noninvasive diagnostic techniques with enhanced contrast, i.e., improved separability between healthy and pathological regions by amplifying differences in absorption and scattering signatures, thereby increasing diagnostic sensitivity and specificity. In chemistry,^{5,6} spectrum inference is pivotal for gas detection and the spectroscopic measurement of chemical substances, where identifying unique spectral fingerprints from limited observations is crucial for in situ safety monitoring and real-time

analysis. Moreover, protein profiling^{7–9} can benefit substantially from spectrum inference, as many protein-related spectral measurements (e.g., vibrational or mass spectra) are expensive, time-consuming, and sample-limited. Hence, predicting informative spectral bands from partial or easier-to-acquire readouts can accelerate protein identification, quantification, and high-throughput screening. Across these applications, constructing large-scale data sets is often impractical or prohibitively costly, and thus high-performance data-processing architectures for accurate spectrum inference under limited-data regimes are in urgent need.

Among various fields utilizing spectrum inference, photonic metasurfaces have emerged as a prominent platform due to their ability to manipulate light at subwavelength scales,¹⁰

Received: March 18, 2026

Revised: July 15, 2026

Accepted: July 28, 2026

Published: August 11, 2026



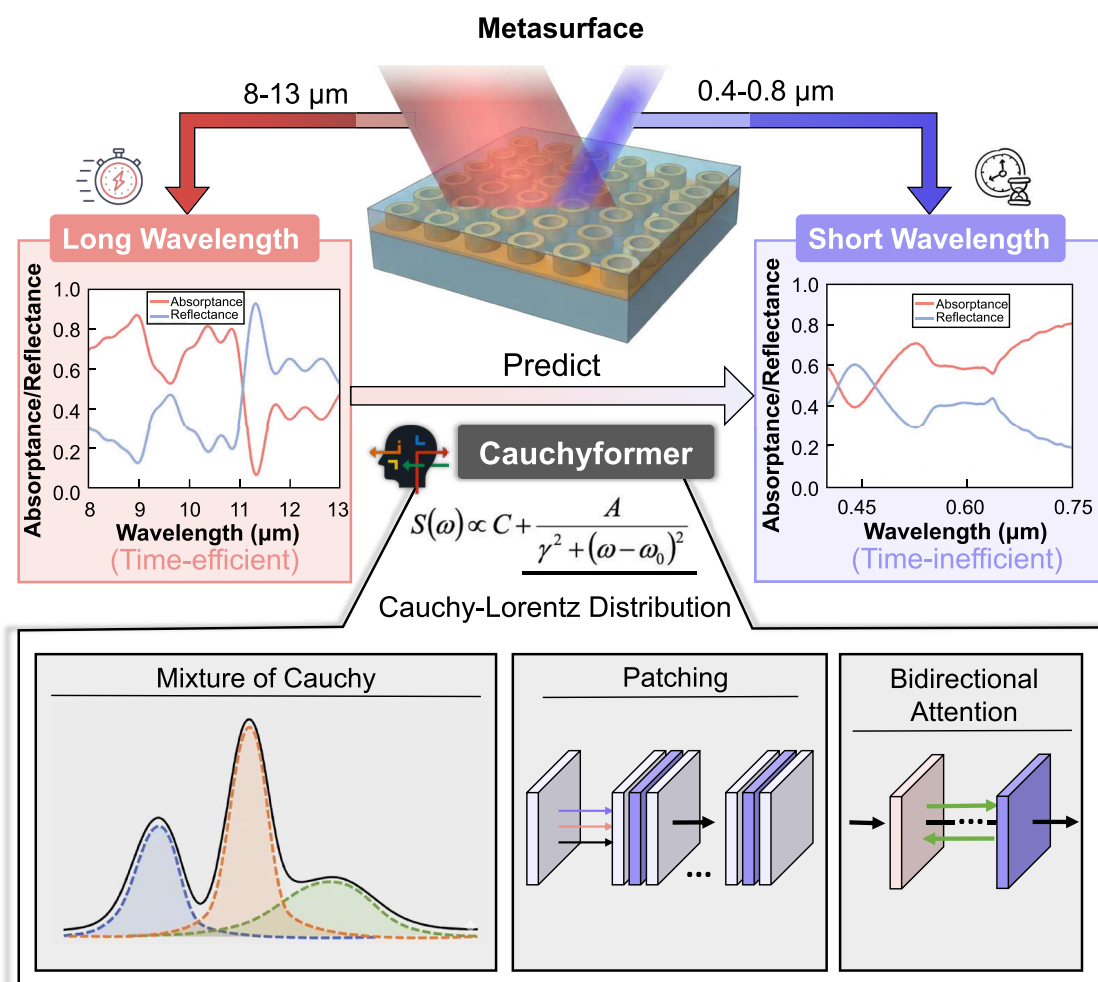


Figure 1. Schematic workflow of Cauchyformer.

enabling broad applications in high-resolution imaging,^{11–13} optical communication,^{14–16} and robotic sensing.^{17–19} Since metasurface functionality is governed by densely packed meta-elements with intricate geometries and arrangements, their design and optimization typically rely on full-wave numerical solvers such as the finite-element method (FEM), finite-difference time-domain (FDTD), and the finite integration technique (FIT), etc. Notably, the required mesh resolution scales with frequency, meaning that the mesh size must decrease as frequency increases. Therefore, high-frequency simulations demand much finer discretization, leading to a substantially larger number of mesh elements with markedly increased runtime and memory. This computational burden becomes especially severe for ultrabroadband characterization of deep-subwavelength light–matter interactions, posing a major barrier to rapid metasurface design and deployment.

In very recent years, the rapid advancement of artificial intelligence has revolutionized sequence modeling, particularly through Transformer-based architectures.²⁰ In the domain of multivariate time series forecasting, innovative models such as **Informer**,²¹ **PatchTST**,²² and **iTransformer**²³ have achieved remarkable performance, showing great potential for scientific applications including spectrum inference. However, the effectiveness of these data-driven approaches is predominantly predicated on the availability of massive, labeled data sets to capture numerical regularities. As a result, the scarcity of high-quality labeled data in scientific applications is identified as the

primary bottleneck for those aforementioned models, and directly applying these data-hungry methods to spectrum inference tasks often leads to overfitting and poor generalization. To overcome this challenge, it is necessary to shift the strategy from expanding data volume to maximizing data efficiency. Since spectral responses are shaped by the collective effects of physical theories,^{24–26} it is essential to develop a methodology that leverages physical information to grasp the intrinsic physical essence embedded within the limited samples, enabling the model to learn the underlying laws rather than merely memorizing surface-level statistics.

In this work, we propose a physics-informed Transformer model, coined as **Cauchyformer**, designed specifically for few-shot spectrum inference. Figure 1 illustrates the schematic workflow of **Cauchyformer**. Our approach builds on the self-attention mechanism of Transformer to capture long-range dependencies across spectral channels,^{27,28} while tailoring the representation to the nuances of photonic physics. By employing a **Mixture of Cauchy (MoC)**, our model represents a spectrum as a sparse superposition of **Cauchy** modes rather than a simple sequence of points, allowing resonance-driven peaks and line shapes to be expressed as **mode coefficients** and embedding physically meaningful priors into the learning process. The MoC plays a role analogous to dictionary learning^{29,30} in representation learning: the spectrum is sparsely coded over Cauchy-mode atoms to capture narrow absorption peaks, whose coefficients reflect metasurface

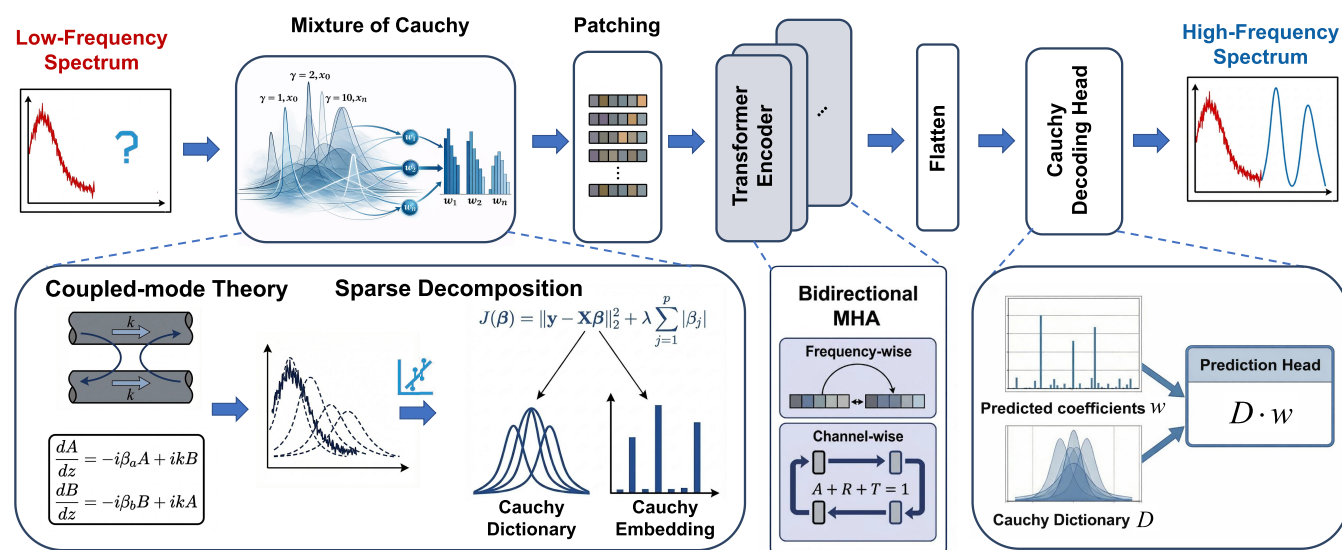


Figure 2. Schematic architecture of Cauchyformer. The model takes a low-frequency spectrum as input, represents it with a Mixture of Cauchy, applies patching, and feeds the tokens into a Transformer encoder with bidirectional multihead attention (frequency-wise attention and channel-wise interaction). Finally, a Cauchy decoding head employs flattened coefficients w , multiplied with the dictionary D to reconstruct and predict the target high-frequency spectrum.

material and structural properties. Mapping spectra into this higher-dimensional MoC space disentangles even minor spectral variations and improves sampling efficiency and training stability in low-data regimes.

Based on the above formulation, we performed a series of experiments to quantify the advantages of Cauchyformer under limited supervision. We first studied cross-band spectrum inference by predicting high-frequency (visible) spectra from low-frequency (mid-infrared) spectra with limited training data, where Cauchyformer outperformed strong Transformer baselines on measure square errors (MSEs) and mean absolute errors (MAEs), especially in few-shot regimes that are common in practical photonic applications. For example, with only 20 training samples, Cauchyformer achieved an MSE of 0.0185 and an MAE of 0.1067, corresponding to reductions of approximately 10% and 7%, respectively, compared with the best-performing baselines. We further validated Cauchyformer on real-world fabricated metasurfaces with measured spectra, demonstrating robust performance under experimental noise and fabrication imperfections. Thanks to the generality of MoC-based spectral representation and inference framework, Cauchyformer not only facilitates streamlined design of metasurfaces, but also serves as a general-purpose building block for fast high-throughput spectroscopy in biosensing, environmental monitoring, and materials informatics, where data is scarce but spectral priors are strong.

2. METHODOLOGY

2.1. Task Formulation

Metasurfaces for tasks such as energy harvesting,^{31,32} electromagnetic cloaking,^{33,34} and image processing^{35,36} are designed to manipulate three optical responses: absorptance (A), reflectance (R), transmittance (T). Different tasks require distinct tuning of the above responses across multiple frequency bands, and metasurface parameters are typically optimized to match band-wise target spectra. In numerical simulations, the required mesh resolution (and thus computational cost) is adjusted according to the working frequency

bands: high-frequency responses demand finer meshes, substantially longer runtimes, and more memory usage, while low-frequency responses are comparatively easier to obtain. Moreover, with experimental spectra typically limited, we evaluate spectrum inference in extremely small-sample settings.

Since low-frequency optical responses are numerically easier to access, **Cauchyformer** takes any two of the three low-frequency optical responses—absorptance, reflectance, and transmittance—as input, denoted as $X_{1;\text{low}}$ and $X_{2;\text{low}}$. The Cauchyformer predicts the corresponding high-frequency spectra $X_{1;\text{high}}$ and $X_{2;\text{high}}$, which form the two variables in the sequence and are inferred simultaneously.

2.2. Cauchyformer

Most Transformer-based forecasting models embed raw input series, which is effective in data-rich domains (e.g., audio,³⁷ traffic,³⁸ and finance^{39,40}) where abundant labeled data are available to capture the characteristics of the series. In photonics, obtaining numerically simulated or experimental data is extremely time-consuming and costly, so large training data sets for spectrum inference are not readily available. Meanwhile, spectral responses are governed by physical laws that can be utilized as strong priors to guide model design.

To tackle this data-scarce problem and leverage physical priors, **Cauchyformer** employs Cauchy Embedding to pre-embed the input spectrum. Leveraging the strong underlying interconnections across frequency bands, spectra are divided into patches to enlarge the local and global receptive fields and to enhance cross-frequency and cross-spectrum dependencies. This enables the model to handle longer sequences, followed by embedding in the frequency direction to describe the properties of variables. The embeddings then interact via self-attention and are processed by feed-forward networks for series representation.

As depicted in Figure 2, **Cauchyformer** includes modules such as Mixture of Cauchy (MoC), patching, Transformer blocks, and a Cauchy Decoding Head. **Cauchyformer** contains L Transformer encoder blocks. Each block incorporates bidirectional attention and a feed-forward network (FFN).

The Transformer Encoder maps the input low-frequency spectrum into latent representations, and the Cauchy Decoding Head reconstructs the predicted spectrum from latent representations. The above process can be formulated as follows:

$$\hat{\mathbf{X}}_{:,high} = f_{\theta}(\mathbf{X}_{:,low}) \quad (1)$$

where $\mathbf{X}_{:,low}$ denotes the low-frequency input channels and $\hat{\mathbf{X}}_{:,high}$ is the predicted high-frequency spectrum.

$$\begin{aligned} \mathbf{Z}_i &= \text{MoC}(\mathbf{X}_{i,low}) \\ \mathbf{H}_{i,j}^{(0)} &= \text{VariateEmbed}(\text{Patch}(\mathbf{Z}_i)) + \mathbf{W}_{pos} \\ \mathbf{H}_{i,j}^{(l+1)} &= \text{EncBlock}(\mathbf{H}_{i,j}^{(l)}) \\ \mathbf{H}_i^{(L)} &= \text{Flatten}(\mathbf{H}_{i,:}^{(L)}) \\ \hat{\mathbf{X}}_{i,high} &= \text{MLP}(\mathbf{D}\mathbf{H}_i^{(L)}) \\ i &= 1, 2, j = 1, \dots, N, l = 1, \dots, L \end{aligned} \quad (2)$$

where i is the index of spectrum channels, j indexes patches, and N is the number of patches. MoC: $\mathbb{R}^n \rightarrow \mathbb{R}^n$ is a physics-informed method for spectral representation based on coupled-mode theory. Patch: $\mathbb{R}^n \rightarrow \mathbb{R}^{N \times s}$ divides embedded tokens into N patches, with s denoting the length of each patch. $\mathbf{W}_{pos} \in \mathbb{R}^d$ is a learnable positional encoding to capture the order of patches in the frequency domain. $\mathbf{H}_{i,:}^{(l)} = \{\mathbf{h}_{i,1}, \dots, \mathbf{h}_{i,N}\} \in \mathbb{R}^{N \times d}$ is the hidden representation of the patches of the i -th channel output by the l -th *EncBlock*, where d is the latent space dimension. VariateEmbed: $\mathbb{R}^s \rightarrow \mathbb{R}^d$ and Flatten: $\mathbb{R}^{N \times d} \rightarrow \mathbb{R}^m$ are implemented by a multilayer perceptron (MLP). After flattening, we multiply the flattened representation with the Cauchy Dictionary $\mathbf{D}$ and leverage another MLP to obtain the predicted high-frequency spectrum $\hat{\mathbf{X}}_{i,high}$. We optimize the model by minimizing the MSE between $\hat{\mathbf{X}}_{i,high}$ and the ground-truth spectrum $\mathbf{X}_{i,high}$.

2.3. Mixture of Cauchy

2.3.1. Coupled-Mode Theory. The spectral response of an ideal lossless photonic resonator would be a train of delta functions determined by material properties and structure-induced photonic modes. In practice, ubiquitous losses (radiative and nonradiative) broaden each resonance with a finite line width that is well-approximated by a Cauchy-Lorentz line shape, which is naturally captured by the coupled-mode theory (CMT).^{24–26}

Consider a metasurface excited by an incident plane wave $|s_+\rangle$. Under CMT, the modal amplitude a and the outgoing wave $|s_-\rangle$ satisfy

$$\begin{cases} \frac{da}{dt} = (i\omega_0 - \gamma)a + \langle \kappa |^* |s_+\rangle \\ |s_-\rangle = c|s_+\rangle + ald \end{cases} \quad (3)$$

where ω_0 is the resonance frequency, $\gamma = \gamma_{rad} + \gamma_{nr}$ is the total decay rate, and κ , c , and d are the coupling coefficients (input-to-mode, direct pathway, and mode-to-output), respectively. For a monochromatic excitation $|s_+\rangle \propto e^{i\omega t}$, the input–output relation can be written as

$$|s_-\rangle = S(\omega)|s_+\rangle = \left[c + \frac{ld\langle \kappa |^*}{i(\omega - \omega_0) + \gamma} \right] |s_+\rangle \quad (4)$$

With energy-conservation and time-reversal constraints, the resulting intensity response exhibits a Lorentzian (or, with interference, a Fano-like) profile. A canonical Lorentzian component takes the form

$$L(\omega; \omega_0, \gamma) = \frac{\gamma^2}{(\omega - \omega_0)^2 + \gamma^2} \quad (5)$$

while geometry-dependent prefactors absorb system-specific coupling constants. The key spectral information is thus concentrated in these Lorentzian peaks.

2.3.2. Cauchy Embedding. Dictionary learning^{29,30,41–44} seeks a set of atoms (basis functions) over which each signal admits a sparse linear representation. Instead of sampling an abstract dictionary from a standard Cauchy distribution, we adopt a physics-parametrized dictionary of Cauchy-Lorentz atoms indexed by center frequency and line width

$$\phi_k(\omega) = \frac{1}{\pi} \frac{\gamma_k}{(\omega - \omega_{0,k})^2 + \gamma_k^2}, \quad k = 1, \dots, n \quad (6)$$

where $\{\omega_{0,k}\}$ tile the frequency axis of interest and $\{\gamma_k\}$ span a set of plausible line widths (e.g., values consistent with radiative (external) and absorptive/nonradiative (internal) loss rates that contribute to spectral broadening). Stacking these atoms yields a Cauchy dictionary

$$\mathbf{D} = [\phi_1, \phi_2, \dots, \phi_n] \in \mathbb{R}^{m \times n} \quad (7)$$

with m frequency samples. Any spectrum (or multichannel spectra) $\mathbf{Y} \in \mathbb{R}^{m \times p}$ is then modeled as a sparse mixture of these modes

$$\mathbf{Y} \approx \mathbf{D}\mathbf{Z}, \quad \mathbf{Z} \in \mathbb{R}^{n \times p} \quad (8)$$

where columns of $\mathbf{Z}$ contain the mode coefficients. We obtain the **Cauchy embedding** (i.e., MoC coefficients) by solving an l_1 -regularized least-squares problem

$$\mathbf{Z} = \underset{\mathbf{Z}}{\text{argmin}} \frac{1}{2} \left\| \mathbf{Y} - \mathbf{D}\mathbf{Z} \right\|_F^2 + \alpha \left\| \mathbf{Z} \right\|_1 \quad (9)$$

with sparsity weight $\alpha > 0$. This embedding explicitly encodes resonance physics—each active coefficient selects a Cauchy-Lorentz mode—thereby denoising the raw spectrum and disentangling overlapping peaks with few training samples. In the network, after Flatten, the latent representation $\mathbf{H}_i^{(L)}$ is interpreted as the MoC coefficient vector for sample i ; multiplying $\mathbf{D}$ and $\mathbf{H}_i^{(L)}$ reconstructs the spectrum as a superposition of Cauchy modes

$$\hat{\mathbf{Y}}_{(:,i)} = \mathbf{D}\mathbf{H}_i^{(L)} \quad (10)$$

which we refer to as the Cauchy-based reconstruction. This Cauchy Embedding not only aligns with the CMT-derived line shapes but also furnishes a compact, interpretable representation that is data-efficient and robust to noise.

2.4. Patching and Bidirectional Attention

We segment the spectral data into contiguous, fixed-length patches to preserve fine-grained local spectral features while enabling efficient, localized representation learning for accurate spectrum inference.

Leveraging the scaled dot-product attention mechanism,²⁰ we employ bidirectional attention to capture cross-spectrum dependencies by attending along both the frequency and channel dimensions. For the j -th patch in the i -th channel, denoted by h_{ij} , each head $h = 1, \dots, H$ maps it to queries, keys, and values $Q_h^{(ij)}, K_h^{(ij)}, V_h^{(ij)} \in \mathbb{R}^{N \times d_k}$, where d_k is the projected dimension. A scaled dot-product attention is then applied to obtain the attention output $O_h^{(ij)}$

$$O_h^{(ij)} = \text{Softmax} \left(\frac{Q_h^{(ij)} (K_h^{(ij)})^T}{\sqrt{d_k}} \right) V_h^{(ij)} \quad (11)$$

By incorporating bidirectional attention, each patch within a spectrum channel not only learns similarities among patches at different frequency bands within the same channel, but also interacts with patches from other channels. This cross-spectral interaction is crucial for spectrum inference due to the intrinsic connections between different channels of the same spectrum. By utilizing these implicit physical relationships, bidirectional attention improves inference accuracy even with scarce training data.

2.5. Cauchy Decoding Head

Cauchyformer predicts the coefficients of the target spectrum in the Cauchy space and reconstructs it using the Cauchy dictionary. This physics-informed decoding preserves the physical structure captured by the Cauchy embedding. Finally, an MLP maps the reconstructed representation to the predicted spectrum, yielding an efficient and interpretable output.

3. RESULTS AND DISCUSSION

3.1. Validation on Numerical Simulation

In many metasurface applications—including radiative cooling and infrared camouflage—two wavelength bands are of particular interest: the atmospheric window (8–14 μm) and the solar band (0.4–2.5 μm). Accordingly, we use the absorptance and reflectance spectra in the long-wavelength band (8–14 μm) as inputs and infer the corresponding short-wavelength spectra in the solar band (0.4–2.5 μm). We employ Lumerical FDTD to generate a data set of 290 metasurface samples, where each sample corresponds to one metasurface design together with its paired absorptance/reflectance spectra computed over 0.4–25 μm (details of the random design generation and simulation settings are provided in the [Supporting Information, Section A](#)). Each spectrum is discretized into 1000 uniformly spaced wavelength points. We use 60 samples for testing and 30 samples for validation; the remaining 200 samples are used for training, and the training set size is varied from 10 to 200. All experiments are conducted on an NVIDIA GeForce RTX 3090; additional data set and implementation details are provided in the [Supporting Information, Section B](#). We also provide performance comparison of selected methods in the [Supporting Information, Section C](#).

Table 1 summarizes the architectural characteristics of the baselines. In addition to Gen-Elim Net,¹ DeepONet,⁴⁵ Deep-RNN,⁴⁶ and Euler-Latent⁴⁷ which are tailored for spectrum inference, we considered representative time-series baselines including Informer,²¹ Reformer,⁴⁸ DLinear,⁴⁹ RLinear,⁵⁰ Crossformer,⁵¹ PatchTST,²² and iTransformer.²³ Notably, Cauchyformer is the only method that combines all listed

Table 1. Comparison of Baseline Characteristics^a

methods	Trm	Enc-only	preprocess	CD	Vanilla Attn	Phys-Info
Gen-Elim Net ¹			✓	✓		✓
Informer ²¹	✓					
Reformer ⁴⁸	✓					
DLinear ⁴⁹			✓			
RLinear ⁵⁰			✓			
Crossformer ⁵¹	✓			✓		
PatchTST ²²	✓	✓	✓		✓	
iTransformer ²³	✓	✓		✓	✓	
DeepONet ⁴⁵				✓		
Deep-RNN ⁴⁶			✓	✓		
Euler-Latent ⁴⁷				✓		✓
Cauchyformer (This work)	✓	✓	✓	✓	✓	✓

^aTransformer (Trm) indicates whether the model is Transformer-based. Encoder-only (Enc-only) signifies the presence of an encoder-only architecture. Pre-process refers to the treatment of input sequences, such as decomposition, convolutional feature extraction, or patching. Channel Dependence (CD) denotes joint modeling of multiple spectral channels. Vanilla Attention (Vanilla Attn) indicates the use of standard attention mechanisms without physics-informed modification. Physics-informed (Phys-Info) indicates the incorporation of explicit physical priors or physics-inspired latent-dynamics formulations.

elements, which motivates its strong performance in the subsequent evaluation.

Experiments are conducted for each training set size ranging from 10 to 200, where each set of experiments are repeated three times. The mean values and error bars for MSE and MAE of each set of experiments are recorded in [Table 2](#), where a lower MSE/MAE indicates a more accurate inference. In each training set size, the best results are highlighted in **bold**, while the second-best results are highlighted with underline.

The experimental results demonstrate that our model **Cauchyformer** maintains consistently strong performance across different few-shot regimes. Particularly, Cauchyformer exhibits commendable inference performance with consistent accuracy when the amount of data is extremely limited (such as with sizes of 10 and 20), which is common in practical photonic applications. In contrast, models that achieve state-of-the-art performance in multivariate time-series forecasting, such as **iTransformer** and **PatchTST**, perform less favorably on spectrum inference, indicating that spectrum inference imposes inherently different modeling requirements compared to conventional time-series forecasting tasks. We also observe that **Crossformer**,⁵¹ which explicitly incorporates channel-dependence modeling, attains competitive performance. This trend suggests that channel dependence is particularly beneficial for spectrum inference, whereas channel independence is often emphasized in multivariate time-series forecasting. Hence, the existing forecasting-oriented architectures may not transfer directly to spectrum inference tasks.

Figure 3 presents representative Cauchyformer's testing results on three metasurface samples within the test set. It is evident that Cauchyformer accurately predicts the trends and peak positions for three types of metasurfaces. Such cutting-edge performance meets the requirements for energy-related metasurface applications like radiative cooling and infrared camouflage, where integrated optical response (e.g., total reflectance/absorptance/transmittance) over the whole inter-

Table 2. Inference Performance of Cauchyformer and Baselines with Different Training Set Sizes

model	Cauchyformer (This work)	iTransformer ²³	Gen-Elim Net ¹	PatchTST ²²	RLinear ⁵⁰	DLinear ⁴⁹	Crossformer ⁵¹	Informer ²¹	Reformer ⁴⁸	DeepONet ⁴⁵	Deep-RNN ⁴⁶	Euler-Latent ⁴⁷
MSE	10	0.0219 (± 0.0001)	0.0371 (± 0.0008)	0.2494 (± 0.0222)	0.1105 (± 0.0698)	0.0382 (± 0.0003)	0.0354 (± 0.0006)	0.0225 (± 0.0005)	0.0226 (± 0.0015)	0.0219 (± 0.0006)	0.0228 (± 0.0020)	0.0219 (± 0.0007)
	20	0.0185 (± 0.0003)	0.0392 (± 0.0033)	0.2058 (± 0.0249)	0.0473 (± 0.0025)	0.0334 (± 0.0001)	0.0313 (± 0.0002)	0.0206 (± 0.0011)	0.0220 (± 0.0016)	0.0217 (± 0.0001)	0.0220 (± 0.0017)	0.0186 (± 0.0005)
	30	0.0195 (± 0.0001)	0.0371 (± 0.0021)	0.2223 (± 0.0233)	0.0420 (± 0.0074)	0.0317 (± 0.0001)	0.0297 (± 0.0006)	0.0207 (± 0.0019)	0.0207 (± 0.0010)	0.0218 (± 0.0002)	0.0218 (± 0.0025)	0.0194 (± 0.0008)
	50	0.0180 (± 0.0004)	0.0346 (± 0.0007)	0.2304 (± 0.0369)	0.0327 (± 0.0030)	0.0267 (± 0.0003)	0.0227 (± 0.0001)	0.0181 (± 0.0004)	0.0186 (± 0.0010)	0.0180 (± 0.0004)	0.0197 (± 0.0014)	0.0185 (± 0.0005)
	100	0.0163 (± 0.0002)	0.0364 (± 0.0012)	0.2320 (± 0.0244)	0.0392 (± 0.0010)	0.0233 (± 0.0003)	0.0196 (± 0.0006)	0.0166 (± 0.0005)	0.0178 (± 0.0011)	0.0177 (± 0.0003)	0.0178 (± 0.0012)	0.0173 (± 0.0004)
	150	0.0147 (± 0.0001)	0.0352 (± 0.0006)	0.1948 (± 0.0059)	0.0282 (± 0.0015)	0.0214 (± 0.0003)	0.0178 (± 0.0001)	0.0153 (± 0.0001)	0.0180 (± 0.0008)	0.0217 (± 0.0001)	0.0178 (± 0.0010)	0.0158 (± 0.0003)
MAE	200	0.0142 (± 0.0000)	0.0364 (± 0.0020)	0.2347 (± 0.0129)	0.0327 (± 0.0046)	0.0206 (± 0.0001)	0.0180 (± 0.0002)	0.0165 (± 0.0025)	0.0171 (± 0.0011)	0.0216 (± 0.0000)	0.0174 (± 0.0009)	0.0151 (± 0.0003)
	10	0.1104 (± 0.0009)	0.1479 (± 0.0022)	0.4001 (± 0.0170)	0.2391 (± 0.0915)	0.1525 (± 0.0009)	0.1448 (± 0.0013)	0.1164 (± 0.0016)	0.1168 (± 0.0039)	0.1147 (± 0.0008)	0.1119 (± 0.0052)	0.1140 (± 0.0018)
	20	0.1067 (± 0.0005)	0.1511 (± 0.0066)	0.3636 (± 0.0227)	0.1576 (± 0.0071)	0.1435 (± 0.0006)	0.1383 (± 0.0003)	0.1144 (± 0.0028)	0.1164 (± 0.0033)	0.1148 (± 0.0007)	0.1096 (± 0.0035)	0.1079 (± 0.0010)
	30	0.1101 (± 0.0004)	0.1484 (± 0.0049)	0.3775 (± 0.0187)	0.1515 (± 0.0136)	0.1398 (± 0.0002)	0.1346 (± 0.0008)	0.1154 (± 0.0046)	0.1135 (± 0.0030)	0.1152 (± 0.0011)	0.1069 (± 0.0058)	0.1114 (± 0.0019)
	50	0.1044 (± 0.0008)	0.1431 (± 0.0016)	0.3836 (± 0.0302)	0.1342 (± 0.0040)	0.1288 (± 0.0004)	0.1178 (± 0.0009)	0.1054 (± 0.0018)	0.1077 (± 0.0034)	0.1156 (± 0.0020)	0.1047 (± 0.0032)	0.1075 (± 0.0011)
	100	0.0980 (± 0.0010)	0.1456 (± 0.0027)	0.3868 (± 0.0198)	0.1526 (± 0.0022)	0.1211 (± 0.0008)	0.1103 (± 0.0015)	0.1009 (± 0.0015)	0.1051 (± 0.0036)	0.1143 (± 0.0002)	0.0982 (± 0.0030)	0.1030 (± 0.0010)
	150	0.0925 (± 0.0002)	0.1427 (± 0.0024)	0.3546 (± 0.0049)	0.1249 (± 0.0023)	0.1165 (± 0.0007)	0.1045 (± 0.0003)	0.0966 (± 0.0004)	0.1061 (± 0.0020)	0.1146 (± 0.0006)	0.0945 (± 0.0025)	0.0966 (± 0.0008)
	200	0.0915 (± 0.0005)	0.1452 (± 0.0047)	0.3885 (± 0.0108)	0.1395 (± 0.0123)	0.1142 (± 0.0002)	0.1050 (± 0.0003)	0.0995 (± 0.0058)	0.1030 (± 0.0029)	0.1144 (± 0.0002)	0.0952 (± 0.0021)	0.0944 (± 0.0007)

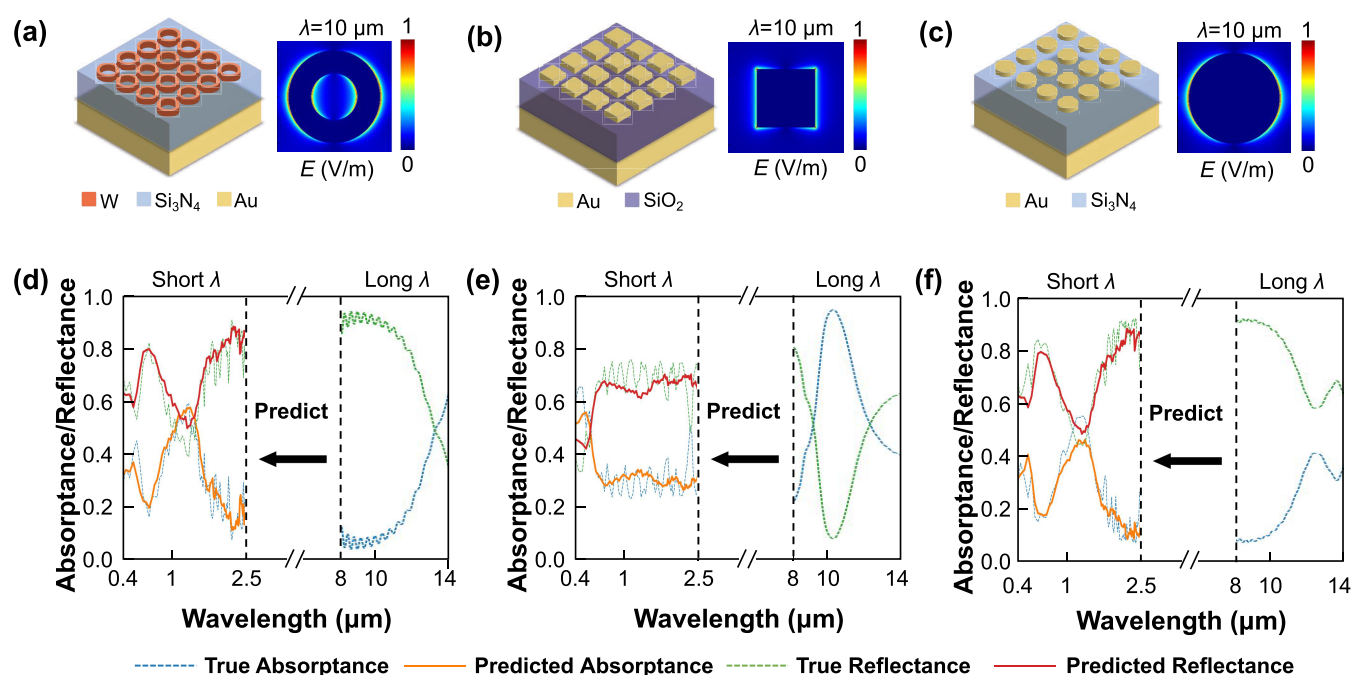


Figure 3. Representative Cauchyformer performance based on simulation results. Panels (a–c) illustrate the schematic designs of three metasurfaces with corresponding electric field distributions of characteristic resonant modes. Panels (d–f) show the spectrum inference results on the three samples in (a–c), respectively.

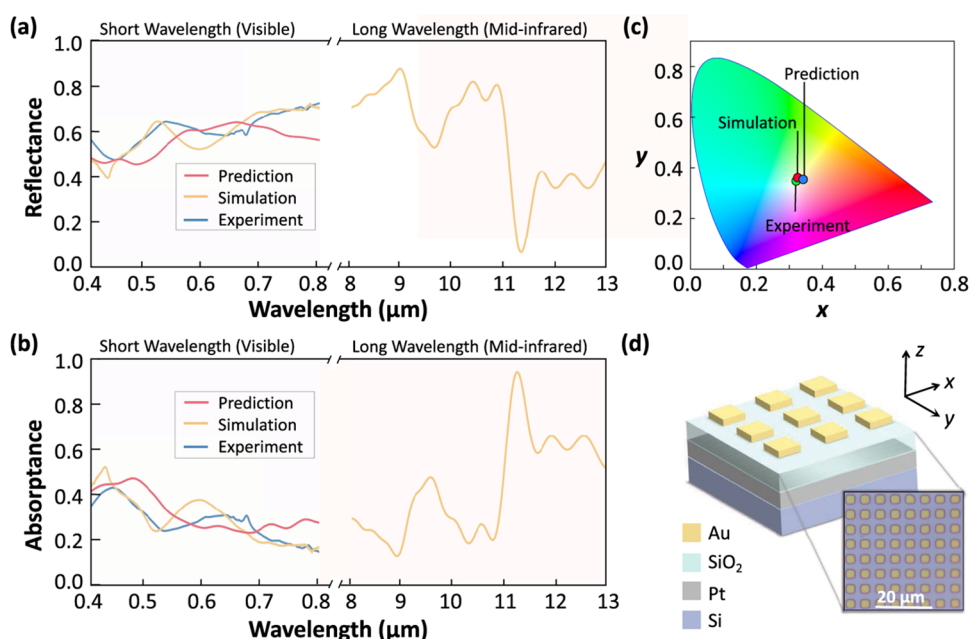


Figure 4. Validation of Cauchyformer on a fabricated metasurface. (a, b) Reflectance and absorbance (0.4–0.75 μm) of a fabricated square-patch Au/SiO₂/Pt metasurface. Predicted spectra by Cauchyformer are compared with experimentally measured and FDTD simulated spectra. Cauchyformer takes the FDTD simulated mid-infrared (8–13 μm) spectra of the same design as input. (c) CIE 1931 chromaticity coordinates computed from (a). (d) Device configuration (Au/SiO₂/Pt/Si) and the corresponding microscopic image.

ested frequency bands are more critical than precise spectral values at each discrete frequency point. The key targets include the number and positions of spectral peaks, the overall spectral shape, and the alignment of spectral features with the targeted operational bands.

3.2. Validation on Fabricated Metasurface

To validate the practical efficacy and generalization capability of the proposed Cauchyformer, we fabricate a trilayer Au/

SiO₂/Pt metasurface with square top-layer patterns (unit size 5 μm, square side length 2.5 μm) using standard photolithography and lift-off techniques (detailed fabrication processes and geometric parameters are provided in the [Supporting Information, Section D](#)). We experimentally measure its optical response with a microspectrophotometry system based on Horiba iHR320 Imaging Spectrometer to assess whether our model, trained only on numerically simulated data, can accurately predict the high-frequency

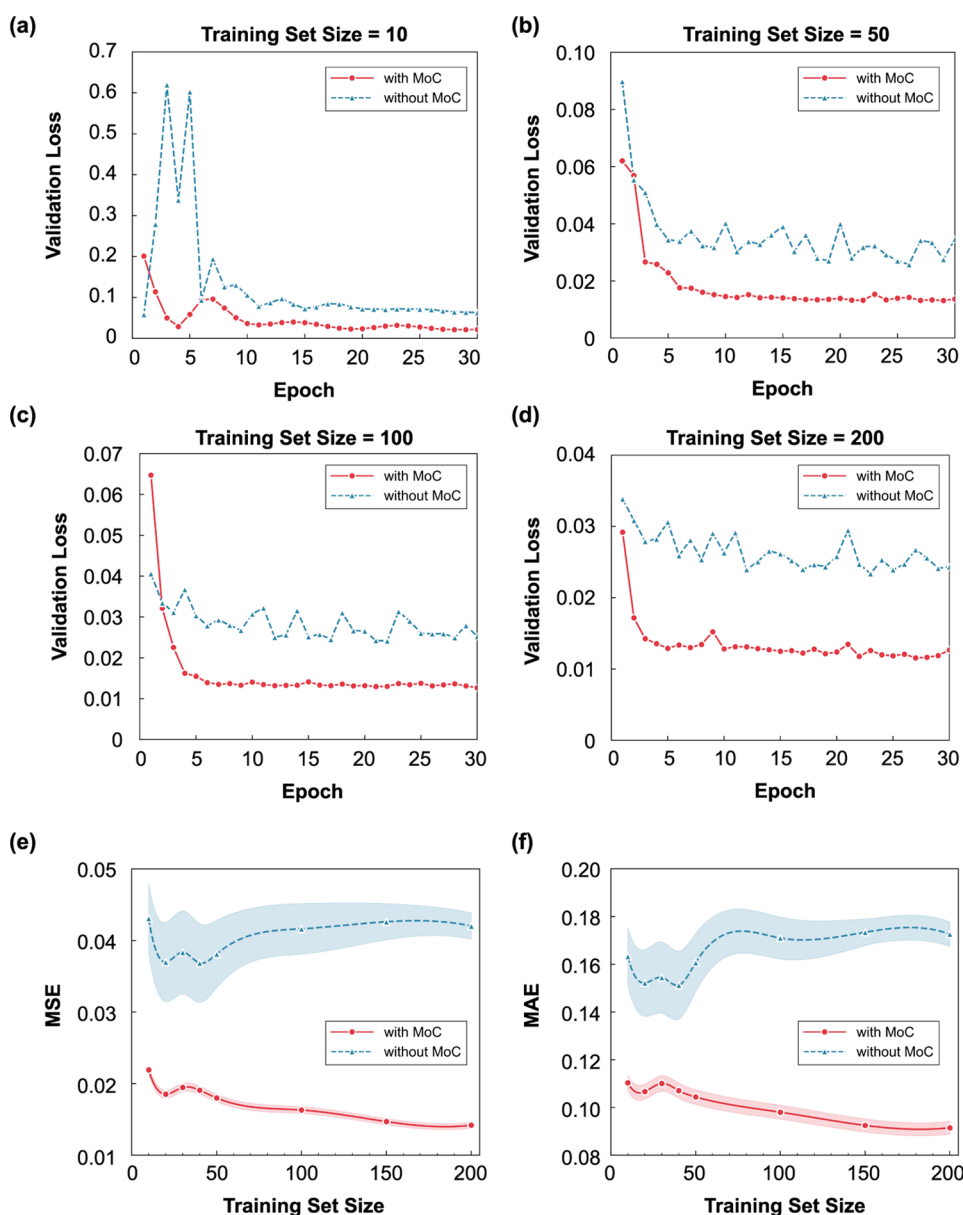


Figure 5. Ablation study of MoC. (a–d) Validation loss trajectories over epochs for training-set sizes $N = 10, 50, 100,$ and 200 , comparing Cauchyformer trained with MoC and without MoC. (e, f) MSE and MAE as functions of training-set size. Lines denote the mean over three random seeds, and shaded regions indicate ± 1 standard deviation.

spectral behavior of real-world metasurfaces. This validation experiment serves as a critical test for the Cauchyformer's robustness as well as its potential to bridge the gap between computational designs and real-world experiments, where material variations and fabrication imperfections inevitably exist. For validation, we use the simulated mid-infrared spectra of the fabricated design ($8\text{--}13\ \mu\text{m}$) as inputs to Cauchyformer and predict the corresponding visible-band optical response ($0.4\text{--}0.75\ \mu\text{m}$). The predicted reflectance/absorptance spectrum in the visible range is compared with both the experimentally measured spectrum and FDTD simulated spectrum of the same sample (Figure 4(a,b)). Note that the fabricated data is strictly excluded from training to avoid data leakage. Beyond spectrum-by-spectrum comparison, we further convert each visible reflectance spectrum into CIE 1931 chromaticity coordinates using the standard color-matching functions,⁵² providing a compact, application-oriented metric

of their perceived colors. As shown in Figure 4(c), the chromaticity point of the predicted spectrum lies close to that of the measurement (and is consistent with the simulated result), indicating that the model preserves not only the spectral trends but also the resulting color appearance for practical applications. Together, these results validate that Cauchyformer serves as a lightweight and powerful surrogate that bridges simulation and experiments for practical metasurface applications.

3.3. Ablation Study

To clarify the role of the MoC module in Cauchyformer, an ablation study is conducted to (1) attribute performance gains specifically to MoC under a controlled setting; (2) assess whether MoC improves optimization stability and generalization, particularly in data-limited regimes. To this end, the effect of the MoC module is evaluated individually by removing it from Cauchyformer and retraining an otherwise

identical model. Meanwhile all nonablated components, optimization hyperparameters, and data splits remain identical for a controlled comparison. The training-set size was swept from severely data-limited to moderately sized regimes ($N = 10, 50, 100, 200$) to evaluate two complementary aspects: training behavior through validation-loss trajectories over epochs (Figure 5(a–d)), and generalization through MSE/MAE on test set (Figure 5(e,f)). Each setting was repeated with three random seeds, and the mean and standard deviation were reported.

The comparison results in Figure 5 clearly show that MoC yields consistent benefits across all N . In the most constrained setting ($N = 10$), MoC leads to a smooth, monotonic decrease and a substantially lower terminal validation loss, whereas the ablated model exhibits sharp validation-loss spikes and pronounced seed sensitivity, indicating unstable optimization. As supervision increases ($N \geq 50$), MoC continues to converge more rapidly to a lower validation-loss plateau than the baseline without MoC, suggesting that the gain is not merely due to regularization but reflects a better inductive bias for spectrum inference. These improvements translate directly to test-time performance in Table 2: MoC achieves lower MSE and MAE for every training-set size, and simultaneously tightens the run-to-run dispersion. Taken together, this ablation study proves that MoC effectively improves both data efficiency and robustness, reducing sensitivity to initialization and scarce supervision.

4. CONCLUSION

In this paper, we present Cauchyformer, a physics-informed Transformer for accurate few-shot spectrum inference. Cauchyformer is built upon a Mixture-of-Cauchy mechanism that draws on the coupled-mode-theory view of resonance spectra, allowing the model to obtain a high-quality representation of spectral responses and perform accurate prediction with extremely limited data. This theory-guided design yields stable optimization and strong generalization in the few-shot regime that is common in scientific applications. The performance is validated not only on simulation benchmarks, but also on a real-world fabricated metasurface, where predictions remain accurate in the presence of material nonidealities and fabrication imperfections. These results demonstrate that incorporating domain theory into modern AI models is a powerful strategy, simultaneously reducing data dependency and enhancing model reliability. Looking forward, several extensions could further broaden the scope of Cauchyformer. First, beyond the symmetric Cauchy–Lorentz mode used in the present implementation, incorporating asymmetric modes may further improve the representation of complex and overlapping resonances. Second, uncertainty quantification could enhance the robustness of Cauchyformer across different materials, fabrication tolerances, and measurement conditions. In conclusion, this work offers a principled direction for bridging the gap between data-driven learning and physics laws, and further transforming AI-for-Science models from data-intensive tools into sample-efficient discovery partners.

■ ASSOCIATED CONTENT

Data Availability Statement

All data supporting the findings of this study are available within the article and its Supporting Information. Additional

data relevant to this work are available from the corresponding authors upon reasonable request.

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.6c02954>.

Experimental and fabrication details, schematic illustrations for the simulated metasurface; hyperparameter settings of Cauchyformer; details of fabrication (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Yansong Tang – Institute of Data and Information, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; Email: tang.yansong@sz.tsinghua.edu.cn

Kaichen Dong – Institute of Data and Information, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; Center of Double Helix, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; Intelligent Passive Thermal Control Center, Research Institute of Tsinghua University in Shenzhen, Shenzhen 518057, China; orcid.org/0000-0001-5334-4243; Email: dkc22@sz.tsinghua.edu.cn

Authors

Shujie Yang – Institute of Data and Information, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; Center of Double Helix, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; Intelligent Passive Thermal Control Center, Research Institute of Tsinghua University in Shenzhen, Shenzhen 518057, China

Xuzhe Zhao – Institute of Data and Information, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; Center of Double Helix, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; Intelligent Passive Thermal Control Center, Research Institute of Tsinghua University in Shenzhen, Shenzhen 518057, China

Chunjiang Li – Institute of Data and Information, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; Center of Double Helix, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; Intelligent Passive Thermal Control Center, Research Institute of Tsinghua University in Shenzhen, Shenzhen 518057, China

Yuxiao Li – Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115, United States; orcid.org/0009-0002-9813-5925

Hongyan Fu – Institute of Data and Information, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acsomega.6c02954>

Author Contributions

[†]S.Y. and X.Z. contributed equally to this work. K.D. conceived the methodology. S.Y. developed the algorithm and performed the model design, implementation, and computational analysis. X.Z. conducted the metasurface

fabrication and performed the experimental measurements and data processing. C.L., Y.L., H.F., Y.T., and K.D. assisted with numerical simulations and data analysis. Y.T. and K.D. supervised the project. All authors discussed the results and contributed to writing and revising the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

S.Y., X.Z., C.L., and K.D. acknowledge support from the National Key R&D Program of China (Grant No. 2023YFB3208700), the National Natural Science Foundation of China (Grant No. 62375151), the Shenzhen Science and Technology Program (Grant No. KJZD20240903095721028), and start-up funding from the Tsinghua Shenzhen International Graduate School, Tsinghua University.

REFERENCES

- (1) Chen, J.; Qian, C.; Zhang, J.; Jia, Y.; Chen, H. Correlating metasurface spectra with a generation-elimination framework. *Nat. Commun.* **2023**, *14*, No. 4872.
- (2) Zhou, Y.; Liu, C.-H.; Wu, B.; Yu, X.; Cheng, G.; Zhu, K.; Wang, K.; Zhang, C.; Zhao, M.; Zong, R.; et al. Optical biopsy identification and grading of gliomas using label-free visible resonance Raman spectroscopy. *J. Biomed. Opt.* **2019**, *24*, No. 095001.
- (3) Ezhov, I.; Scibilia, K.; Giannoni, L.; Kofler, F.; Iliash, I.; Hsieh, F.; Shit, S.; Caredda, C.; Lange, F.; Montcel, B.; et al. Learnable real-time inference of molecular composition from diffuse spectroscopy of brain tissue. *J. Biomed. Opt.* **2024**, *29*, No. 093509.
- (4) Mishra, A.; Aghaee, M.; Tamer, I. M.; Budman, H. Spectroscopic Advances in Real Time Monitoring of Pharmaceutical Bioprocesses: A Review of Vibrational and Fluorescence Techniques. *Spectrosc. J.* **2025**, *3*, No. 12.
- (5) Ibrahim, A.; Ataca, C. Prediction of frequency-dependent optical spectrum for solid materials: a multioutput and multifidelity machine learning approach. *ACS Appl. Mater. Interfaces* **2024**, *16*, 41145–41156.
- (6) Ma, C.; Peng, D.; Bai, X.; Liu, S.; Luo, L. A review of optical fiber sensing technology based on thin film and Fabry-Perot cavity. *Coatings* **2023**, *13*, No. 1277.
- (7) Bandeira, N.; Tsur, D.; Frank, A.; Pevzner, P. A. Protein identification by spectral networks analysis. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6140–6145.
- (8) Pelton, J. T.; McLean, L. R. Spectroscopic methods for analysis of protein secondary structure. *Anal. Biochem.* **2000**, *277*, 167–176.
- (9) Liu, H.; Yang, J.; Wang, M.; Xue, L.; Chou, K.-C. Using Fourier spectrum analysis and pseudo amino acid composition for prediction of membrane protein types. *Protein J.* **2005**, *24*, 385–389.
- (10) Kuznetsov, A. I.; Brongersma, M. L.; Yao, J.; Chen, M. K.; Levy, U.; Tsai, D. P.; Zheludev, N. I.; Faraon, A.; Arbabi, A.; Yu, N.; et al. Roadmap for optical metasurfaces. *ACS Photonics* **2024**, *11*, 816–865.
- (11) Khorasaninejad, M.; Capasso, F. Metalenses: Versatile multifunctional photonic components. *Science* **2017**, *358*, No. eaam8100.
- (12) Decker, M.; Chen, W. T.; Nobis, T.; Zhu, A. Y.; Khorasaninejad, M.; Bharwani, Z.; Capasso, F.; Petschulat, J. Imaging performance of polarization-insensitive metalenses. *ACS Photonics* **2019**, *6*, 1493–1499.
- (13) Kenry; Duan, Y.; Liu, B. Recent advances of optical imaging in the second near-infrared window. *Adv. Mater.* **2018**, *30*, No. 1802394.
- (14) Chen, S.; Liu, W.; Li, Z.; Cheng, H.; Tian, J. Metasurface-empowered optical multiplexing and multifunction. *Adv. Mater.* **2020**, *32*, No. 1805912.
- (15) Nazemosadat, E.; Mazur, M.; Kruk, S.; Kravchenko, I.; Carpenter, J.; Schröder, J.; Andrekson, P. A.; Karlsson, M.; Kivshar, Y. Dielectric broadband metasurfaces for fiber mode-multiplexed communications. *Adv. Opt. Mater.* **2019**, *7*, No. 1801679.
- (16) Hu, J.; Guo, Z.; Shi, J.; Jiang, X.; Chen, Q.; Chen, H.; He, Z.; Song, Q.; Xiao, S.; Yu, S.; et al. A metasurface-based full-color circular auto-focusing Airy beam transmitter for stable high-speed underwater wireless optical communications. *Nat. Commun.* **2024**, *15*, No. 2944.
- (17) Qin, J.; Jiang, S.; Wang, Z.; Cheng, X.; Li, B.; Shi, Y.; Tsai, D. P.; Liu, A. Q.; Huang, W.; Zhu, W. Metasurface micro/nano-optical sensors: principles and applications. *ACS Nano* **2022**, *16*, 11598–11618.
- (18) Beruete, M.; Jáuregui-López, I. Terahertz sensing based on metasurfaces. *Adv. Opt. Mater.* **2020**, *8*, No. 1900721.
- (19) Lee, Y.; Kim, S.-J.; Park, H.; Lee, B. Metamaterials and metasurfaces for sensor applications. *Sensors* **2017**, *17*, No. 1726.
- (20) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. In *Attention Is All You Need*, Advanced in Neural Information Processing System; NIPS, 2017.
- (21) Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115.
- (22) Nie, Y.; Nguyen, N. H.; Sinthong, P.; Kalagnanam, J. In *A Time Series is Worth 64 Words: Long-term Forecasting with Transformers*; Eleventh International Conference on Learning Representations, 2022.
- (23) Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; Long, M. In *iTransformer: Inverted Transformers Are Effective for Time Series Forecasting*; Twelfth International Conference on Learning Representations, 2023.
- (24) Fan, S.; Suh, W.; Joannopoulos, J. D. Temporal coupled-mode theory for the Fano resonance in optical resonators. *J. Opt. Soc. Am. A* **2003**, *20*, 569–572.
- (25) Zhou, M.; Liu, D.; Belling, S. W.; Cheng, H.; Kats, M. A.; Fan, S.; Povinelli, M. L.; Yu, Z. Inverse design of metasurfaces based on coupled-mode theory and adjoint optimization. *Acs Photonics* **2021**, *8*, 2265–2273.
- (26) Passia, M.-T.; Yioultis, T. V.; Kriezis, E. E. A coupled-mode-theory formulation for periodic multi-element metasurfaces in the presence of radiation losses. *J. Appl. Phys.* **2024**, *135*, No. 023102, DOI: 10.1063/5.0179442.
- (27) Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. et al. In *An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale*; International Conference on Learning Representations, 2020.
- (28) Kenton, J. D. M.-W. C.; Toutanova, L. K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding; Proceedings of NAACL-HLT, 2019; pp 4171–4186.
- (29) Xiang, Z.; Xu, H.; Ramadge, P. J. In *Learning Sparse Representations of High Dimensional Data on Large Scale Dictionaries*, Advances in Neural Information Processing Systems; NIPS, 2011.
- (30) Mairal, J.; Ponce, J.; Sapiro, G.; Zisserman, A.; Bach, F. In *Supervised dictionary learning*, Advances in Neural Information Processing Systems; NIPS, 2008.
- (31) Jiang, X.; Yuan, H.; Chen, D.; Zhang, Z.; Du, T.; Ma, H.; Yang, J. Metasurface based on inverse design for maximizing solar spectral absorption. *Adv. Opt. Mater.* **2021**, *9*, No. 2100575.
- (32) Heidari, M. H.; Sedighy, S. Broadband wide-angle polarization-insensitive metasurface solar absorber. *J. Opt. Soc. Am. A* **2018**, *35*, 522–525.
- (33) Liu, M.; Kozyrev, A. B.; Shadrivov, I. V. Time-varying metasurfaces for broadband spectral camouflage. *Phys. Rev. Appl.* **2019**, *12*, No. 054052.
- (34) Buhara, E.; Ghobadi, A.; Ozbay, E. Adaptive visible and short-wave infrared camouflage using a dynamically tunable metasurface. *Opt. Lett.* **2021**, *46*, 4777–4780.
- (35) Li, L.; Ruan, H.; Liu, C.; Li, Y.; Shuang, Y.; Alù, A.; Qiu, C.-W.; Cui, T. J. Machine-learning reprogrammable metasurface imager. *Nat. Commun.* **2019**, *10*, No. 1082.

- (36) Zhang, X.; Zhou, Y.; Zheng, H.; Linares, A. E.; Ugwu, F. C.; Li, D.; Sun, H.-B.; Bai, B.; Valentine, J. G. Reconfigurable metasurface for image processing. *Nano Lett.* **2021**, *21*, 8715–8722.
- (37) Zölzer, U. *Digital Audio Signal Processing*; John Wiley & Sons, 2022.
- (38) Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 4927–4943.
- (39) Yan, H.; Ouyang, H. Financial time series prediction based on deep learning. *Wireless Pers. Commun.* **2018**, *102*, 683–700.
- (40) Kim, S.; Kang, M. Financial Series Prediction Using Attention LSTM. 2019; arXiv:1902.10877, arXiv.org e-Print archive. <https://arxiv.org/abs/1902.10877>.
- (41) Guo, J.; Guo, Y.; Kong, X.; Zhang, M.; He, R. Discriminative analysis dictionary learning *Proc. AAAI Conf. Artif. Intell.* 2016; Vol. 30 DOI: 10.1609/aaai.v30i1.10213.
- (42) Kolouri, S.; Rostami, M.; Owechko, Y.; Kim, K. Joint dictionaries for zero-shot learning *Proc. AAAI Conf. Artif. Intell.* 2018; Vol. 32 DOI: 10.1609/aaai.v32i1.11649.
- (43) Liu, Y.; Chen, Q.; Chen, W.; Wassell, I. Dictionary learning inspired deep network for scene recognition *Proc. AAAI Conf. Artif. Intell.* 2018; Vol. 32 DOI: 10.1609/aaai.v32i1.12312.
- (44) Zhou, F.; Huang, S.; Xing, Y. Deep semantic dictionary learning for multi-label image classification. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 3572–3580.
- (45) Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; Karniadakis, G. E. Learning Nonlinear Operators via DeepONet Based on the Universal Approximation Theorem of Operators. *Nat. Mach. Intell.* **2021**, *3*, 218–229.
- (46) Yan, R.; Wang, T.; Jiang, X.; Huang, X.; Wang, L.; Yue, X.; Wang, H.; Wang, Y. Efficient Inverse Design and Spectrum Prediction for Nanophotonic Devices Based on Deep Recurrent Neural Networks. *Nanotechnology* **2021**, *32*, No. 335201.
- (47) Zhu, E.; Zong, Z.; Li, E.; Lu, Y.; Zhang, J.; Xie, H.; Li, Y.; Yin, W.-Y.; Wei, Z. Frequency Transfer and Inverse Design for Metasurface under Multi-Physics Coupling by Euler Latent Dynamic and Data-Analytical Regularizations. *Nat. Commun.* **2025**, *16*, No. 2251.
- (48) Kitaev, N.; Kaiser, L.; Levskaya, A. In *Reformer: The Efficient Transformer*; International Conference on Learning Representations, 2019.
- (49) Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are transformers effective for time series forecasting? *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 11121–11128.
- (50) Li, Z.; Qi, S.; Li, Y.; Xu, Z. Revisiting Long-Term Time Series Forecasting: An Investigation on Linear Mapping. 2023; arXiv:2305.10721, arXiv.org e-Print archive. <https://arxiv.org/abs/2305.10721>.
- (51) Zhang, Y.; Yan, J. In *Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting*; 11 International Conference on Learning Representations 2022.
- (52) Fairman, H. S.; Brill, M. H.; Hemmendinger, H. How the CIE 1931 color-matching functions were derived from Wright-Guild data. *Color Res. Appl.* **1997**, *22*, 11–23.