

High-accuracy spectral reconstruction via a physics-fusion cross-attention framework in computational micro-spectrometry

TAO MA, XIN YE*, JINGJUN WU*, FENG TANG, XUELEI LI

Research Center of Laser Fusion, China Academy of Engineering Physics,
Mianyang 621900, China

School of Mathematics and Physics, Extreme Condition Material Properties Science
and Technology Center, Southwest University of Science and Technology,
Mianyang 621010, China

*Corresponding authors: xyecaep@mail.ustc.edu.cn (XY), jingjunwu163@163.com (JW)

Miniaturized spectrometers are gaining significant attention in scientific and industrial fields due to their compact footprint, low cost, and on-site analysis capability. Spectral reconstruction algorithms are crucial for computational spectrometers to achieve accurate signal demixing and recovery. However, these algorithms often struggle under low-light or noisy conditions. While deep learning-based methods show promise, they remain insufficient for computational spectrometers employing band-stop encoding arrays, which offer high optical throughput but introduce strong spectral interference and complex measurement challenges. This work proposes a novel spectral reconstruction framework that explicitly integrates physical priors through a multi-head cross-attention mechanism. This approach dynamically models the interaction between the measurement matrix and observed signals, enabling effective multimodal feature fusion and enhanced extraction of discriminative spectral features. An experimental band-stop computational micro-spectrometer system was built to acquire real data, and data augmentation strategies were applied to improve model generalization. Comparative experiments with various algorithms on both datasets validate the effectiveness of fusing physical prior information via the multi-head cross-attention mechanism. This study presents a new paradigm for spectral reconstruction in band-stop computational micro-spectrometers, offering an effective solution for interference-resistant spectral analysis in computational spectrometry.

Keywords: physics-fusion, spectral reconstruction, computational spectrometers, robust super-resolution reconstruction, band-stop filter array.

1. Introduction

Spectral analysis serves as a fundamental characterization tool in scientific research and industry by decoding material-specific “fingerprint” signatures [1]. The field of

nano-optics has witnessed significant development in recent years. For example, in terahertz device design, various studies on broadband and tunable absorbers have emerged. A notable instance is the work by HAN and CHEN, who proposed a tunable broadband terahertz absorber based on a single-layer graphene metasurface. Through patterned graphene design and electrical tuning, the device achieved broadband absorption exceeding 90% in the approximately 1.14–3.31 THz range, while demonstrating polarization-insensitive performance and angular stability. This provides experimental evidence for the potential of graphene-based tunable broadband absorbers in imaging, sensing, and photonic applications [2]. Computational micro-spectrometers based on metasurfaces and other nano-devices are accelerating the advancement of computational spectroscopy. Their compact size and portability enable new real-time detection capabilities in astronomy [3], physics [4], and chemical analysis [5], demonstrating significant potential. While dispersive spectrometers achieve higher resolution at the expense of reduced optical throughput [6], Fourier-transform spectrometers maximize photon utilization but suffer from slow acquisition speeds due to their inherent step-scan mechanism [7]. In contrast, computational micro-spectrometers reconstruct high-dimensional spectra from low-dimensional signals via the compressed sensing model $T_{m \times n} x_n = I_m$ [8]. Here, T denotes the system sensing matrix, x represents the spectral vector to be reconstructed, and I is the measured one-dimensional response signal. The structure of the computational micro-spectrometer system is illustrated in Fig. 1. Incident light is transmitted through the filter array, and the resulting intensity signal is collected by a CMOS detector. An algorithm is then employed to recover the spectral information from this modulated intensity signal. The m in compressed sensing model denotes the number of filter units and n represents the spectral channels. When $m < n$, indicating an underdetermined system of equations, the sensitivity of the solution to perturbations scales with the condition number of matrix T [9].

High-condition numbers induce ill-conditioning, where measurement noise amplifies reconstruction distortion. Hence, noise-robust reconstruction algorithms are critically important for computational micro-spectrometers.

To overcome resolution and optical throughput limitations in conventional spectrometers, recent studies employ band-stop filtering strategies coupled with a reconstruction

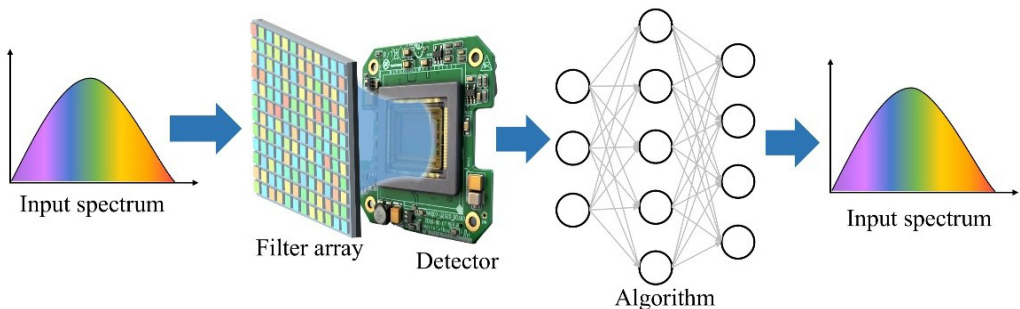


Fig. 1. Diagram of the computational micro-spectrometer system.

algorithm. For example, Tang *et al.* implemented metasurface-based band-stop filtering and achieved spectral reconstruction via least-squares optimization [10]. This type of spectrometer holds an advantage in low-light conditions due to its characteristic of high optical throughput. When paired with an appropriate spectral reconstruction algorithm, it can reconstruct the incident spectrum even in high-noise environments.

Deep learning demonstrates significant potential for spectral reconstruction, particularly in computational speed and noise suppression [11, 12]. Subsequent studies integrated physical models to enhance performance. WAN *et al.* combined the sensing matrix pseudoinverse with U-net under dual spectral-intensity loss constraints [13]. BROWN *et al.* used fully connected networks to efficiently reconstruct random filter profiles [14]. SU *et al.* transformed the alternating direction method of multipliers (ADMM) solver for spectral reconstruction equations into a neural network architecture, mitigating the black-box limitation [15]. Recent research has also explored integrating Transformer architectures with physical imaging models to solve imaging inverse problems. For instance, TransFNet has showcased enhanced reconstruction performance in speckle correlation imaging by incorporating a physics-constrained Transformer structure, though its design remains primarily tailored to that specific imaging scenario [16]. Traditional iterative methods require prior information and often involve time-consuming parameter tuning, while data-driven neural networks struggle with weak band-stop features and noisy conditions. Thus, developing noise-robust spectral reconstruction networks remains a fundamental challenge for ultra-compact spectroscopic systems.

To overcome the dual limitations of noise sensitivity and calibration errors, this study proposes a physics-fusion cross-attention (PF-CAtt) framework specifically designed for band-stop computational micro-spectrometers utilizing guided-mode resonance filters (GMRF). This micro-spectrometer employs a 12×12 metasurface-based filtering array [17]. While the narrow band-stop filtering design ensures high luminous energy utilization, it simultaneously causes encoded optical signals to become susceptible to drowning in background noise. The band-stop filtering characteristic leads to a very high condition number in the sensing matrix T . Therefore, any attempt to use traditional iterative solvers or deep learning methods is rendered infeasible. PF-CAtt innovatively employs multi-head cross-attention to integrate the system sensing matrix with response signals, constructing a reconstruction framework that effectively addresses this challenge. The mechanism of cross-attention captures latent relationships between the sensing matrix and the response signals, significantly improving the recognition of faint features. Requiring only small training datasets, PF-CAtt robustly reconstructs 761-channel visible spectra from noise-corrupted signals, establishing a novel paradigm for low-light detection in ultra-compact spectroscopic systems. The number of reconstructed spectral channels far exceeds the number of metasurface units, achieving super-resolution spectral reconstruction. Unlike traditional optical resolution, which is constrained by the Rayleigh criterion and determined by hardware physics, the super-resolution discussed in this work aims to overcome the physical sampling limitation through computational means. By leveraging algorithmic priors to solve the underde-

terminated inverse problem, it mathematically recovers finer spectral details than hardware sampling allows, thereby realizing a software-level enhancement of spectral resolution.

2. Experimental setup

To accurately acquire and analyze the spectral data required for the experiment, this study designed and constructed a complete detection system for a computational micro-spectrometer. The core of this system lies in the integration of an optical encoding array with a computational reconstruction algorithm, aiming to capture and resolve high-dimensional spectral information through a non-scanning, compact architecture. The system primarily consists of a guided-mode resonance-based encoding chip, a commercial spectrometer for capturing ground-truth spectra, and a detector for acquiring signals modulated by the encoding chip. The various components of the system are detailed below.

2.1. Micro-spectrometer based on band-stop filters

The metasurface configuration, depicted in Fig. 2(a), features a silicon dioxide (SiO_2) substrate. A layer of silicon nitride (Si_3N_4) is deposited on this substrate, followed by the fabrication of a periodic nanopillar array on the Si_3N_4 film by electron beam lithography.

Guided-mode resonance (GMR) is a phenomenon occurring in periodic optical structures characterized by a strong resonance arising from phase matching between the specific wavelength and the guided modes. This resonance enables the manipulation of light at the nanoscale [18, 19]. These GMR metasurfaces exhibit high diffraction efficiency and ultra-narrow linewidths [20–22]. Crucially, its fabrication process is CMOS compatible, allowing the mass production of micro-spectrometers [23, 24]. Due to its subwavelength scale structure and flexible design, transmission properties can be precisely engineered [25]. Precise control of nanopillar geometry, specifically height,

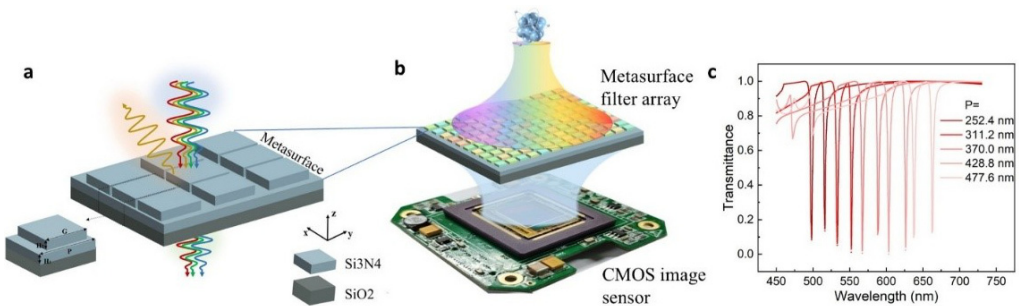


Fig. 2. Diagram of the metasurface-based micro-spectrometer (a) GMRF metasurface unit. (b) Schematic of the spectrometer featuring an individual 12×12 GMRF bandstop filter array integrated on top of a CMOS imaging sensor. (c) Transmission curves of some metasurface units.

width, and periodicity, enables band-stop filtering responses in the visible spectrum featuring tailored full-width-at-half-maximum (FWHM) characteristics and resonance dip positions [26–30]. Figure 2(b) illustrates the metasurface-based micro-spectrometer architecture, where incident light transmitted through the metasurface is modulated and subsequently captured by a CMOS detector, yielding the response signal. The simulation results in Fig. 2(c) demonstrate the designed characteristic of the structure, featuring a dual resonance-dip configuration that enables narrowband filtering of incident light.

2.2. Experiments and data acquisition

To acquire the experimental data, a computational micro-spectrometer system based on an encoding array was constructed. This system was employed to capture the encoded response data and the corresponding ground-truth spectra, which served as labels for the neural network.

2.2.1. Calibration

The calibration setup schematic is shown in Fig. 3(a). The light from a xenon source was modulated by a monochromator to generate quasi-monochromatic light. The quasi-monochromatic light was split by a beam splitter; one beam entered a commercial spectrometer to record spectral ground truth in real time, while the other beam transmitted through the encoding metasurface, ultimately being detected by a CMOS camera.

Across the visible spectrum, spectral scanning was performed using the monochromator with increments of 0.5 nm during the 400–780 nm range (*i.e.*, 761 steps). This

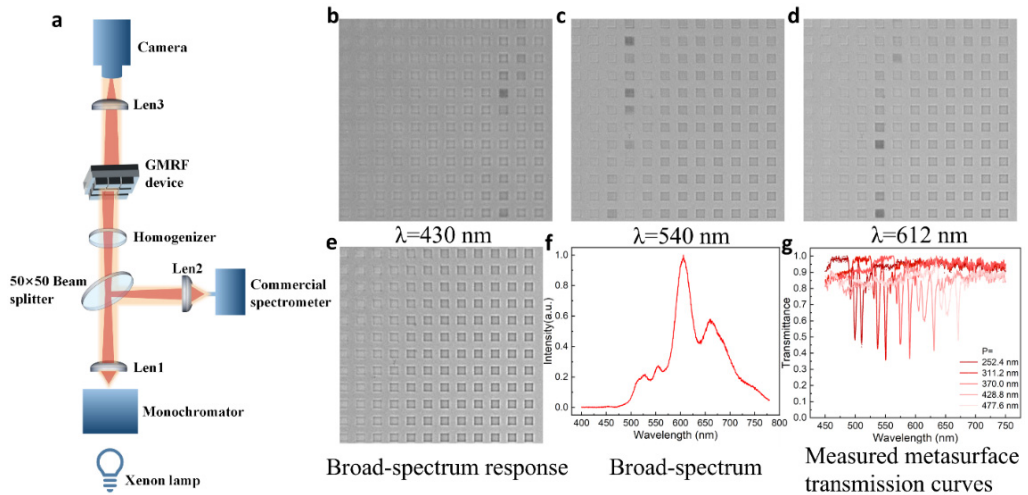


Fig. 3. Spectrometer calibration. (a) Spectrometer calibration setup. (b–d) Response of the metasurface array under illumination with different wavelengths. (e, f) Incident broadband spectrum and the corresponding response of the spectrometer. (g) Experimentally measured transmission curve of the metasurface unit in the spectrometer.

yielded a raw system sensing matrix T_0 with dimensions 761×144 , representing the spatial response of the detector to every monochromatic light. Each row of matrix T_0 corresponds to the intensity distribution of a specific monochromatic wavelength on the CMOS focal plane. Figures 3(b)–(d) display the CMOS images recorded for different monochromatic lights transmitted through the filter. The corresponding 144 filter response signals for each quasi-monochromatic light are obtained by extracting the pixel values at these subunit locations. To eliminate nonuniformity of the intensity across wavelengths, the calibrated sensing matrix was preprocessed to reduce calibration errors, yielding the final sensing matrix T . Detailed procedures are documented in Section 3.

Each column in this matrix represents the transmittance curve of a specific filter unit across the visible spectrum. As shown in Fig. 3(g), key characteristics of these transmittance profiles, including extinction ratios and resonance dip positions, differ from ideal theoretical curves, posing significant challenges for spectral reconstruction.

2.2.2. Data acquisition

Training datasets, generated using a spectrometer-based metasurface imaging system, were developed to optimize the neural network model. The experimental optical setup is shown in Fig. 3(a). The monochromator in the calibration system was replaced with colored optical filters. Light transmitted through different color filter combinations generated different modulated broadband spectra. A commercial spectrometer recorded spectral ground truth as neural network labels, while the CMOS camera captured modulated spectral responses serving as network inputs. Figures 3(e) and (f) show representative incident broadband spectra and corresponding CMOS responses, respectively.

A measured data set of 1005 samples was collected, each pairing a spectral ground-truth curve with its corresponding response signal. The data set was partitioned into training and test sets in an 8:2 ratio, establishing the foundation for neural network training.

2.3. Data augmentation and noise simulation

The collected data underwent cleaning and augmentation. Data expansion was specifically achieved by applying perturbation in the principal component analysis (PCA) space. Furthermore, noise simulation was implemented on the augmented dataset using the K-nearest neighbors (KNN) machine learning method. Ultimately, a dataset of substantial size and dimensionality was obtained. To quantitatively validate the distributional consistency of the PCA + KNN augmentation strategy, we analyzed the geometric properties of the generated spectra in the feature space. The distribution of pairwise cosine similarities (Mean = 0.63, Std = 0.24) confirms that the augmented dataset maintains substantial variance and effectively covers the feature space without suffering from mode collapse. Furthermore, the exceptionally high nearest neighbor

similarity (Mean > 0.99) indicates that the augmented data forms a dense and smooth manifold, suggesting that the generated samples are locally consistent and adhere to the underlying continuous structure of the spectral data.

2.3.1. Spectral dataset augmentation

The size of the spectral training dataset obtained through actual measurements is relatively small. To guard against potential overfitting in deep learning that may arise from using such a limited dataset, this study employed a systematic data augmentation strategy combining PCA with a cosine similarity-based screening method to expand the spectral dataset, thereby generating a sufficiently large and well-distributed augmented dataset.

The augmentation process consisted of two main parts: redundancy removal through spectral screening and spectral generation via perturbation in PCA space. Ultimately, this procedure yielded an augmented spectral dataset containing 4000 entries. Quantitative analysis confirms distributional consistency: the augmented dataset maintains a global mean intensity of 0.625 (± 0.237), aligning with physical laws. Furthermore, the high average nearest-neighbor cosine similarity (0.9955) demonstrates that the generated spectra strictly preserve the morphological characteristics of real-world signals.

2.3.2. Noise simulation

Following the spectral data augmentation, simulated response data can be obtained through the system sensing matrix T and the linear model

$$Tx = I$$

However, such simulation yields only the ideal linear response and does not account for various noise sources and nonlinear instrument errors present in real measurements, such as systematic and environmental noise. Therefore, it is necessary to construct a model capable of learning the noise distribution from actual measurements to generate realistic noise perturbations for the augmented spectra.

This study proposes a generation method based on conditional K-nearest neighbors (KNN) noise fitting. The core premise is that spectra with similar response baselines should exhibit similar noise patterns. Consequently, realistic noise can be generated by searching for and blending the noise patterns from measured samples with similar responses via KNN. Detailed procedures are documented in Section 5.

The 4000 spectra generated via PCA augmentation maintain distributional characteristics consistent with the real measured data in a statistical sense. Simultaneously, the conditional KNN-based noise fitting for the response baseline achieves noise transfer from real measurements to the augmented data. This ensures the generated noise is not only physically realistic but also preserves local similarity and continuity. The final augmented dataset is capable of significantly enhancing the robustness and generalization performance of subsequent regression models in noisy environments.

3. Results and discussion

3.1. Neural network-based reconstruction algorithm

A spectral reconstruction algorithm that integrates physical priors through multi-head cross-attention is proposed, with its architecture illustrated in Fig. 4. This framework combines physical modeling with deep learning. The sensing matrix T , which represents the prior spectral response characteristics of the GMR encoding metasurface, is processed through dedicated convolutional layers for feature extraction. The 761 channels are mapped to 8 channels via convolution, while the detector acquired intensity response vector I is expanded to 8 channels through convolutional upsampling. Multi-head cross-attention then splits these inputs into 4 heads along the channel dimension and fuses them with matrix T , generating 8-channel integrated features. This multi-head mechanism enables deep fusion of physical priors with raw observational data. The integrated features are subsequently fed into a convolutional neural network backbone based on the U-net architecture. In the proposed multi-head cross-attention mechanism, the query (Q) is projected from the response signal I , while the key (K) and value (V) are derived from the sensing matrix T . This architectural design embodies a specific physical interpretation: the observed signal actively queries the system's response characteristics. By computing the attention weights between the signal and the sensing matrix, the model dynamically highlights the specific components of the system matrix that are most relevant to the current observation, thereby enhancing the interpretability and reconstruction accuracy of the network.

The selection of a 7-layer depth for the U-net backbone was determined by the spatial constraints of the input spectral dimension ($N = 144$). Since each encoder block

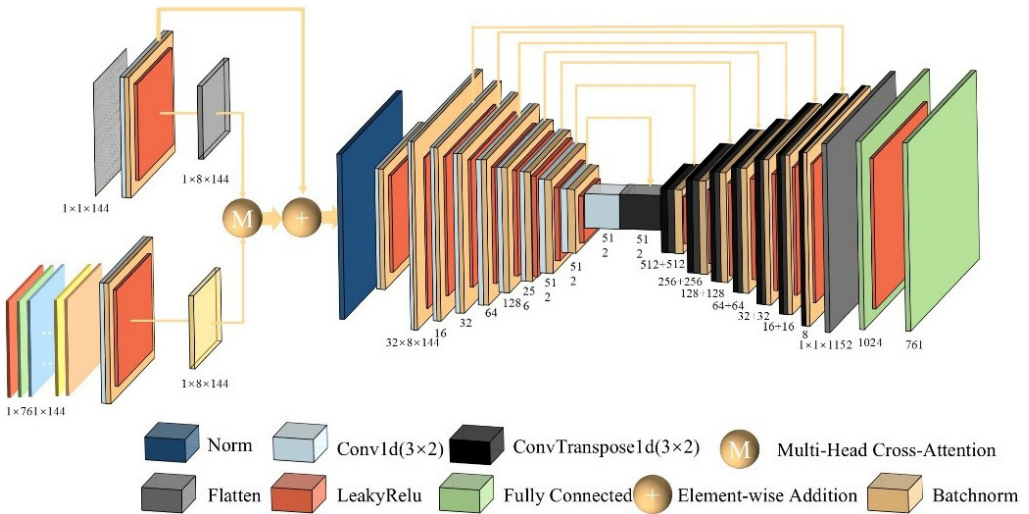


Fig. 4. Algorithm architecture. Schematic (a) Schematic of the spectral reconstruction neural network. (b) Multi-head cross-attention mechanism.

utilizes a stride-2 convolution for downsampling, a 7-layer architecture reduces the feature map size to the theoretical minimum (approaching a scalar representation at the bottleneck) without causing spatial collapse. This depth was chosen to maximize the receptive field and global feature abstraction capability while strictly adhering to the geometric limitations of the input data size.

The network employs a symmetric encoder-decoder architecture, with both paths containing seven convolutional blocks. Each block consists sequentially of a convolutional layer, a batch normalization layer, and a LeakyReLU activation function. Batch normalization regularizes feature distributions to effectively suppress overfitting, while LeakyReLU mitigates vanishing gradient issues and enhances training stability. The encoder progressively compresses spatial dimensions through down-sampling operations while extracting high-level abstract features. On the contrary, the decoder restores spatial resolution via upsampling operations with skip connections, incorporating shallow-layer details to optimize reconstruction accuracy [31].

The feature maps from the decoder output undergo flattening to form a 1D vector, which is then fed into a fully-connected layer for dimensionality mapping. This ultimately outputs a reconstructed spectral vector $\hat{x}$ spanning the 400–780 nm wavelength range.

This work begins from theoretical principles, designing the algorithm architecture based on a mathematical model, as outlined above, the spectral reconstruction task can be formulated as an underdetermined inverse problem:

$$Tx = I$$

where $I \in \mathbb{R}^{144}$ is the system measurement, $x \in \mathbb{R}^{761}$ is the high-resolution spectrum to be recovered, an $T \in \mathbb{R}^{144 \times 761}$ denotes the system transfer matrix. Since $144 \ll 761$, the inverse mapping from I to x is not unique. Therefore, a valid reconstruction model must explicitly incorporate the physical matrix T to constrain the solution space.

To address this challenge, we introduce a cross-attention–fusion U-net architecture. The model first encodes the measurement I and the sensing matrix T into feature spaces:

$$F_b = \varphi_I(I)$$

$$F_T = \varphi_T(T)$$

where φ_I and φ_T are learnable convolutional encoders. Instead of simply concatenating the two features, we design a multi-head cross-attention module that establishes an explicit dependency between the measurement and the physical system:

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

with $Q = W_q F_I$, $K = W_k F_T$, and $V = W_v F_T$.

This formulation allows each spectral component in I to adaptively query the most relevant structural information in T . The resulting attention matrix can be interpreted as a learnable inverse operator that approximates a physics-informed regularizer:

$$\hat{x} = \arg \min_x \|Tx - I\|^2 + \Psi_\theta(x; T)$$

where Ψ_θ represents the implicit prior encoded by cross attention. In this way, the proposed model incorporates the physical measurement process directly into feature interaction, significantly improving the identifiability of the solution in an underdetermined setting.

Finally, the fused representation is decoded by a multi-scale U-net, which captures both local peak structures and global spectral trends, ensuring that the reconstructed spectrum is physically plausible and structurally consistent with the ground truth.

Consistent training parameters were applied across all datasets. The model was trained using the Adam optimizer with an initial learning rate of 0.001, a batch size of 32,

T a b l e 1. Reconstruction performance of different algorithms on experimentally measured and augmented datasets.

Dataset	Method	MSE
Measurement dataset	Reg NNLS	0.0140
	ADMM-LASSO	0.0190
	PD-net	0.0067
	Transformer1D	0.00179
	BiLSTM	0.00122
	TCN	0.00454
	Res-CNN	0.00966
	MLP	0.00154
	U-net (ablation)	0.000982
	PF-CAtt (Ours)	0.000755
KNN-made dataset	Reg NNLS	0.0162
	ADMM-LASSO	0.0165
	PD-net	0.0052
	Transformer1D	0.00191
	BiLSTM	0.00123
	TCN	0.00159
	Res-CNN	0.00199
	MLP	0.00165
	U-net (ablation)	0.00102
	PF-CAtt (Ours)	0.000727

and a total of 1,500 epochs. A step-based learning rate scheduling strategy was adopted, where the rate was halved if the loss showed no improvement over 300 consecutive epochs. To ensure a fair comparison, all algorithms were executed on the same CPU (AMD Ryzen 7735H) and GPU (NVIDIA 4060), with the results depicted in Table 1.

Table 1 presents a comparative analysis of results obtained under the experimentally measured dataset and its augmented counterpart, all trained under identical parameter settings.

3.2. Comparative model setup

To validate the effectiveness of the proposed PF-CAtt in solving the underdetermined linear inverse problem $Tx = I$ for spectral reconstruction, we designed a series of systematic comparative experiments. The compared models encompass a range of representative methods, from traditional 1D sequence models to deep convolutional architectures, and include an ablation model with the multi-head attention mechanism removed, in order to analyze the independent contribution of this core innovative mod-

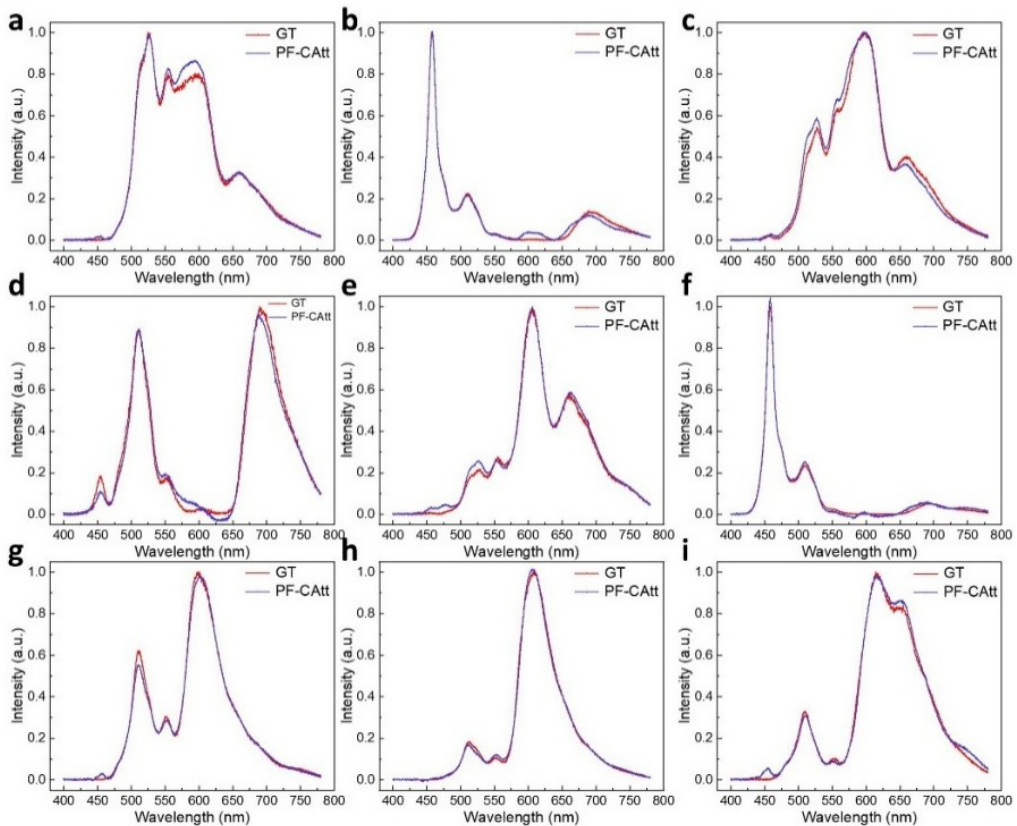


Fig. 5. Spectral reconstruction demonstration. (a–i) Modulated spectra generated using combinations of colored optical filters.

ule. In addition to deep learning methods, we also compared the performance of traditional iterative algorithms for this spectral reconstruction task, including regularized non-negative least squares (Reg NNLS) and a sparsity-promoting solver using the alternating direction method of multipliers (ADMM-Lasso).

The PF-CAtt model proposed in this paper builds upon a U-net-style encoder-decoder architecture by further introducing a multi-head cross-attention module, which enables explicit interaction between the input signal x and the sensing matrix T in the feature space. Specifically, the model first extracts feature from the input and the matrix encoding separately using one-dimensional convolutional encoders. It then computes cross-modal attention with x as the query and T as the key/value, thereby achieving a structured fusion of the physical imaging mechanism. Subsequently, the fused features are fed into a deep U-net to accomplish pointwise recovery of the spectral sequence.

Furthermore, we selected several typical one-dimensional sequence modeling methods as baselines, including a transformer encoder, bidirectional LSTM (BiLSTM), temporal convolutional network (TCN), a 1D residual CNN, and a multi-layer perceptron (MLP). All these models directly process the input sequence of length 144 and output the reconstructed 761-dimensional spectrum. They represent common sequence regression frameworks in recent years and serve as important controls for determining whether our proposed architecture fully leverages physical constraint information. Among the comparative algorithms, we also selected previously proposed neural networks that incorporate physical priors, such as PD-net [13].

Demonstrating robust spectral reconstruction capabilities, Fig. 5 presents the performance of PF-CAtt on experimental data, accurately recovering complex spectral profiles despite significant noise interference and system deviations inherent in real-world measurements.

3.3. Results and discussion

The experimental results in Table 1 demonstrate that our algorithm PF-CAtt achieves higher-precision spectral reconstruction, showing improvements of 23.12% and 28.73% over the ablation model on the experimentally measured and augmented datasets, respectively. Its performance gain is even more significant compared to other deep learning algorithms designed for processing one-dimensional data. On both the experimentally measured and augmented datasets, PF-CAtt achieves an accuracy improvement exceeding 90% over traditional iterative algorithms.

In computational spectrometers, reconstruction performance is fundamentally limited by fabrication-induced deviations in GMR filters. The measured filter curves in Fig. 3(g) exhibit significant deviations from the theoretical curves designed in Fig. 2(c). These discrepancies arise from nanofabrication accuracy limitations. Correlation heat maps of measured filter transmittance functions in Fig. 6(c) reveal a low correlation between the curves, partially demonstrating the potential of this architecture for spectral reconstruction [32].

Noise interference from the CMOS camera also poses severe challenges to reconstruction accuracy. The structural compactness of the metasurface coupled with weakly differentiated features across unit cells, exacerbates the susceptibility of noise in intensity measurements. Despite the use of homogenizing lenses to optimize illumination uniformity, the beam intensity distribution remains imperfectly uniform. Non-uniform illumination causes deviations in the metasurface's transmission response from ideal conditions, adversely affecting inverse spectral reconstruction. Stray light from surface reflections, coupled with inherent CMOS shot noise and dark current noise, collectively degrade the SNR of detected signals [33]. Figure 6(a) shows the incident broadband spectrum, while Fig. 6(b) presents the corresponding simulated noise-free and

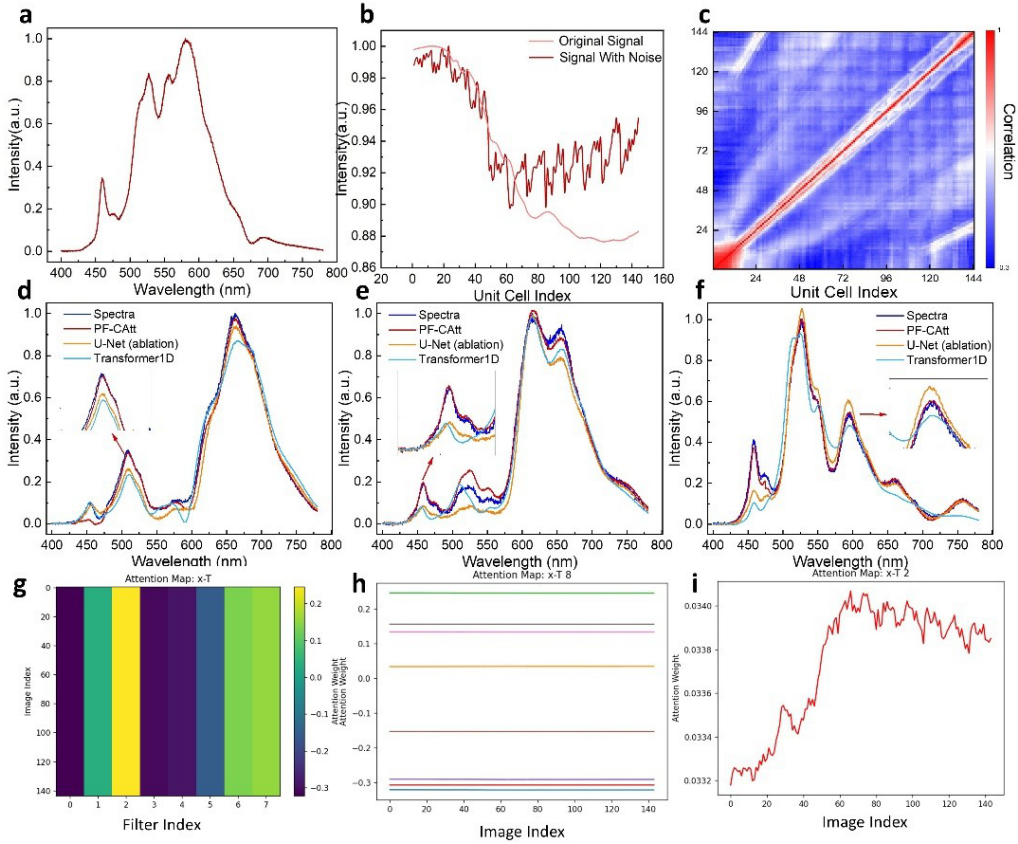


Fig. 6. Comprehensive characterization of the performance of spectral reconstruction. (a) Incident spectrum. (b) Simulated noise-free versus experimentally measured intensity responses. (c) Correlation heat maps of measured filter transmittance functions. (d–f) Comparative reconstruction results of different algorithms using experimental data. (g) Weight distribution of the attention module, showing the response of each of its eight heads to individual metasurface units. (h) Visualization of the attention weights across all eight heads. (i) Detailed weight map for the third attention head (corresponding to one of the heads in panel h).

experimentally measured noisy intensity response curves. Minute-camera noise severely perturbs these low-dynamic-range response curves, adversely affecting spectral reconstruction.

The design of the filter array prioritizes high transmittance, producing a sensing matrix T with a condition number approximately 1.9×10^5 . Within the system $Tx = I$ underdetermined, a large condition renders the reconstruction highly sensitive to noise in the response signal I [34], inducing significant distortion in the recovered spectrum x . Designed filter curves achieve 92.8% average optical transmittance, while measured curves achieve 89.8%. Corresponding average encoding information ratios defined as the proportion of filtered spectral components relative to total transmittance are 7.2% and 10.2%, respectively. In certain scenarios, encoded information becomes completely obscured by background noise, presenting a substantial reconstruction challenge.

Furthermore, the input data exhibits homogenized response characteristics where the transmittance curves of different filter units differ only through minor resonance dip positional shifts. This results in less than 5% response variation across channels for identical incident spectra, creating weak feature discrimination amidst high background noise, which presents a challenge for high-accuracy spectral reconstruction.

Previous studies have integrated physical priors into neural networks through various approaches, such as pioneering application of neural networks to spectral reconstruction [35] and dual intensity-spectrum loss constraints based on the sensing matrix [13]. These methods demand high data accuracy and perform well when sensing matrix characteristics are well-defined and measurement errors are minimal. However, in this system, characterized by multiple optical components, as previously discussed, the precise construction of the T sensing matrix is challenging. Consequently, applying these approaches introduces computational burdens rather than benefits. As evidenced in Table 1, strongly physics-constrained algorithms significantly underperform on the experimental data set. Similarly, constrained by the large condition number of the sensing matrix, traditional iterative algorithms perform poorly in the spectral reconstruction task. The term “strong constraint” here refers to the direct integration of physical constraints by inverting the sensing matrix and multiplying it with the response signal. If the sensing matrix is measured accurately and has a low condition number, this inversion-based fusion can indeed be beneficial. Conversely, in our system—designed to resolve the trade-off between spectral resolution and optical throughput—the sensing matrix is highly ill-conditioned, making it unsuitable for inversion. Furthermore, noise introduced during the calibration process would also compromise the effectiveness of such a prior-integration approach. The multi-head cross-attention mechanism effectively adapts to high-condition-number sensing matrices, enabling the successful fusion of physical priors.

As evident from the experimental results in Table 1, PF-CAtt significantly outperforms all other models across evaluation metrics. This advantage is particularly pronounced in the recovery of high-frequency spectral details and the consistency of the overall spectral shape. The superiority primarily stems from the following key points.

The cross-modal multi-head attention significantly enhances the model's efficiency in utilizing the sensing matrix T . In underdetermined problems, the input sequence x alone provides insufficient constraints. The attention module dynamically captures the contribution of T across different spectral bands, thereby enabling superior recovery of the structural details in the true spectrum.

The multi-scale convolutional structure of U-net further strengthens the simultaneous modeling of both local and global information. Compared to pure sequence models (*e.g.*, LSTM, Transformer), U-net holds advantages in modeling both low-frequency trends and local smoothness, resulting in more stable performance for spectral regression tasks. To effectively learn the fused input features, the U-net framework employs a seven-layer encoder-decoder configuration. This deep convolutional architecture exhibits strong representational capacity.

The performance of the ablation model (U-net) shows an obvious decline, proving that the attention module is the key factor driving performance improvement. Experiments indicate that removing the cross-attention weakens the model's ability to express the system's physical characteristics, causing the reconstruction results to rely more heavily on the convolutional features themselves. This makes it difficult to effectively handle the inherent solution multiplicity in underdetermined systems.

Furthermore, models such as transformer, TCN, CNN, and MLP, although commonly employed for general 1D sequence regression tasks, all underperform our proposed method in this specific task. This indicates that such models lack the structured modeling capability for the physical relationship between T and x , consequently hindering their ability to achieve precise spectral reconstruction.

In summary, the proposed PF-CAtt, by integrating a multi-head cross-attention mechanism into a deep U-net structure, achieves effective encoding of and feature interaction with the sensing matrix T , demonstrating significant advantages in underdetermined spectral reconstruction scenarios. The comparative and ablation experiments validate the core role of the attention module in performance enhancement and confirm that the proposed model design is grounded in solid physical understanding and engineering feasibility.

Figures 6(d)–(f) illustrate the reconstruction results for three representative spectra from the measured dataset. In the challenging case of rapidly varying spectral features, the PF-CAtt algorithm faithfully recovers abrupt changes, delivering a markedly lower mean-squared error (MSE) and thus significantly outperforming the competing methods.

The weight distribution maps in Figs. 6(h) and (i) illustrate the operation of the attention module. The visualization shows that the eight distinct heads attend differently to the 144 metasurface units, effectively validating the designed mechanism's capability to learn varied, task-relevant features. Multi-head process multidimensional information from dual data streams [36], establishing connections between physical priors and optical response data. This multi-head attention mechanism effectively consolidates input data with physical priors, enhancing feature separability and significantly improving algorithmic robustness. The integration of the T sensing matrix essentially

embeds differentiable optical sensing models within the network, guiding the learning of physically constrained spectral mapping relationships.

4. Conclusions

In this work, an optical platform for a band-stop computational micro-spectrometer was constructed to acquire an experimental dataset specifically tailored for spectral reconstruction. On this dataset, the proposed method achieves reconstruction accuracy improvements of 94.6% and 23.12% compared with conventional iterative reconstruction methods and data-driven neural networks, respectively. To mitigate overfitting and enhance data diversity, principal component analysis (PCA) and a K-nearest neighbors (KNN)-based strategy were introduced to perform effective spectral augmentation and high-fidelity simulation of realistic noise. On the augmented dataset, the proposed algorithm demonstrates an accuracy improvement of 95.4% over traditional iterative approaches and achieves gains of at least 28.73% compared with existing deep learning methods designed for one-dimensional spectral data. These results clearly validate the effectiveness of the proposed multi-head cross-attention mechanism in fusing physical prior information for spectral reconstruction. The proposed PF-CAtt framework offers a new direction for the development of interference-resistant spectral reconstruction algorithms in computational micro-spectrometry.

Credit authorship contribution statement

Tao Ma: writing original draft, visualization, software, methodology, investigation, formal analysis, data curation, conceptualization. **Xin Ye:** writing review and editing, supervision, project administration, funding acquisition, data curation, conceptualization. **Jingjun Wu:** supervision, investigation, project administration. **Feng Tang:** writing review and editing, investigation, data curation. **Xuele Li:** investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request. The source code can also be made available via a GitHub repository upon reasonable request.

Funding

This research was funded by the National Natural Science Foundation of China (No. 12304352), the Start-up Foundation for Ph.D. of Southwest University of Science and Technology (No. 25zx7150).

References

- [1] BAO J., BAWENDI M.G., *A colloidal quantum dot spectrometer*, *Nature*, **523**(7558), 2015: 67–70. <https://doi.org/10.1038/nature14576>
- [2] HAN J.Z., CHEN R.S., *Tunable broadband terahertz absorber based on a single-layer graphene meta-surface*, *Optics Express*, **28**(20), 2020: 30289–30298. <https://doi.org/10.1364/OE.403631>
- [3] KULESA C., *Terahertz spectroscopy for astronomy: From comets to cosmology*, *IEEE Transactions on Terahertz Science and Technology*, **1**(1), 2011: 232–240. <https://doi.org/10.1109/TTHZ.2011.2159648>

- [4] CAMPBELL P., MOORE I.D., PEARSON M.R., *Laser spectroscopy for nuclear structure physics*, Progress in Particle and Nuclear Physics, **86**, 2016: 127–180. <https://doi.org/10.1016/j.pnpnp.2015.09.003>
- [5] OZAKI Y., *Near-infrared spectroscopy—its versatility in analytical chemistry*, Analytical Sciences, **28**(6), 2012: 545–563. <https://doi.org/10.2116/analsci.28.545>
- [6] REDDING B., CAO H., *Using a multimode fiber as a high-resolution, low-loss spectrometer*, Optics Letters, **37**(16), 2012: 3384–3386. <https://doi.org/10.1364/OL.37.003384>
- [7] KUEHL D., GRIFFITHS P.R., *Dual-beam Fourier-transform infrared spectrometer*, Analytical Chemistry, **50**(3), 1978: 418–422. <https://doi.org/10.1021/ac50025a016>
- [8] YANG Z., ALBROW-OWEN T., CUI H., ALEXANDER-WEBBER J., GU F., WANG X., WU T.-C., ZHUGE M., WILLIAMS C., WANG P., ZAYATS A.V., CAI W., DAI L., HOFMANN S., OVEREND M., TONG L., YANG Q., SUN Z., HASAN T., *Single-nanowire spectrometers*, Science, **365**(6457), 2019: 1017–1020. <https://doi.org/10.1126/science.aax8814>
- [9] CHANDRASEKARAN S., IPSEN I.C.F., *On the sensitivity of solution components in linear systems of equations*, SIAM Journal on Matrix Analysis and Applications, **16**(1), 1995: 93–112. <https://doi.org/10.1137/S0895479892231255>
- [10] TANG F., WU J., ALBROW-OWEN T., CUI H., CHEN F., SHI Y., ZOU L., CHEN J., GUO X., SUN Y., LUO J., JU B., HUANG J., LIU S., LI B., YANG L., MUNRO E.A., ZHENG W., JOYCE H.J., CHEN H., CHE L., DONG S., SUN Z., HASAN T., YE X., YANG Y., YANG Z., *Metasurface spectrometers beyond resolution-sensitivity constraints*, Science Advances, **10**(49), 2024: eadr7155. <https://doi.org/10.1126/sciadv.adr7155>
- [11] YANG Z., ALBROW-OWEN T., CAI W., HASAN T., *Miniaturization of optical spectrometers*, Science, **371**(6528), 2021: eabe0722. <https://doi.org/10.1126/science.abe0722>
- [12] JIA Y., FAN Z., QIAN C., DEL HOUGNE P., CHEN H., *Dynamic inverse design of broadband metasurfaces with synthetical neural networks*, Laser & Photonics Reviews, **18**(10), 2024: 2400063. <https://doi.org/10.1002/lpor.202400063>
- [13] WAN Y., FAN X., XU B., HE Z., *Reconstructive spectrum analyzer with high-resolution and large-bandwidth using physical-model and data-driven model combined neural network*, Laser & Photonics Reviews, **17**(7), 2023: 2201018. <https://doi.org/10.1002/lpor.202201018>
- [14] BROWN C., GONCHAROV A., BALLARD Z.S., FORDHAM M., CLEMENS A., QIU Y., RIVENSON Y., OZCAN A., *Neural network-based on-chip spectroscopy using a scalable plasmonic encoder*, ACS Nano, **15**(4), 2021: 6305–6315. <https://doi.org/10.1021/acsnano.1c00079>
- [15] SU X., LIU S., WU H., CHEN P., YANG J., WU J., *Fast spectrum reconstruction based-on interpretable method with physical model involved*, Optics & Laser Technology, **182**, 2025: 112079. <https://doi.org/10.1016/j.optlastec.2024.112079>
- [16] TANG C., WANG F., TANG Z.W., CHU Y.P., SITU G.H., *TransFNet: A physics-enhanced transformer for speckle correlation imaging*, Optics Communications, **591**, 2025: 132126. <https://doi.org/10.1016/j.optcom.2025.132126>
- [17] WU J., WEI C., CUI H., CHEN F., HU K., LI A., PAN S., YANG Y., MA J., YANG Z., ZHENG W., ZHU R., *High-sensitivity, high-resolution miniaturized spectrometers for ultraviolet to near-infrared using guided-mode resonance filters*, Molecules, **29**(23), 2024: 5580. <https://doi.org/10.3390/molecules29235580>
- [18] MAGNUSSON R., WANG S.S., *New principle for optical filters*, Applied Physics Letters, **61**(9), 1992: 1022–1024. <https://doi.org/10.1063/1.107703>
- [19] INOUE J., URA S., KINTAKA K., *Guided-mode resonance filter for micro-optic spectrometer*, 2020 IEEE 70th Electronic Components and Technology Conference (ECTC), Orlando, FL, USA, 2020: 1812–1817. <https://doi.org/10.1109/ECTC32862.2020.00283>
- [20] ROSENBLATT D., SHARON A., FRIESEM A.A., *Resonant grating waveguide structures*, IEEE Journal of Quantum Electronics, **33**(11), 1997: 2038–2059. <https://doi.org/10.1109/3.641320>
- [21] FEHREMBACH A.-L., TALNEAU A., BOYKO O., LEMARCHAND F., SENTENAC A., *Experimental demonstration of a narrowband, angular tolerant, polarization independent, doubly periodic resonant grating filter*, Optics Letters, **32**(15), 2007: 2269–2271. <https://doi.org/10.1364/OL.32.002269>

- [22] URA S., NAKATA M., YANAGIDA K., INOUE J., KINTAKA K., *Narrowband-stop reflector using guided-mode resonance in waveguide cavity for WDM optical interconnects*, 2016 IEEE 66th Electronic Components and Technology Conference (ECTC), Las Vegas, NV, USA, 2016: 2270–2276. <https://doi.org/10.1109/ECTC.2016.109>
- [23] COLBURN S., ZHAN A., BAYATI E., WHITEHEAD J., RYOU A., HUANG L., MAJUMDAR A., *Broadband transparent and CMOS-compatible flat optics with silicon nitride metasurfaces [Invited]*, *Optical Materials Express*, **8**(8), 2018: 2330–2344. <https://doi.org/10.1364/OME.8.002330>
- [24] LI N., XU Z., DONG Y., HU T., ZHONG Q., HSING FU Y., ZHU S., SINGH N., *Large-area metasurface on CMOS-compatible fabrication platform: driving flat optics from lab to fab*, *Nanophotonics*, **9**(10), 2018: 3071–3087. <https://doi.org/10.1515/nanoph-2020-0063>
- [25] GAO X., MA Q., GU Z., CUI W.Y., CUI T.J., CHAN C.H., *Reconfigurable spoof localized surface plasmonic for frequency detections*, *Laser & Photonics Reviews*, **17**(11), 2023: 2300267. <https://doi.org/10.1002/lpor.202300267>
- [26] HUANG L., JIN R., ZHOU C., LI G., XU L., OVERVIG A., DENG F., CHEN X., LU W., ALÙ A., MIROSHNICHENKO A.E., *Ultrahigh-Q guided mode resonances in an all-dielectric metasurface*, *Nature Communications*, **14**(1), 2023: 3433. <https://doi.org/10.1038/s41467-023-39227-5>
- [27] LUO S.S., ZHANG Z.F., HE X., ZHANG Z.J., LI X., FU M.C., YANG J.B., *Design of reflective tunable structural color metasurface based on guided-mode resonance filter and Sb_2S_3* , *Photonics*, **10**(7), 2023: 752. <https://doi.org/10.3390/photonics10070752>
- [28] DE CEGLIA D., GANDOLFI M., ANTONIETTA VINCENTI M., TOGNAZZI A., FRANCESCHINI P., CINO A.C., AMBROSIO G., BARATTO C., LI B., CAMACHO-MORALES R., NESHEV D.N., DE ANGELIS C., *Transient guided-mode resonance metasurfaces with phase-transition materials*, *Optics Letters*, **48**(11), 2023: 2961–2964. <https://doi.org/10.1364/OL.486733>
- [29] FINCO G., BIDESKAN M.Z., VERTCHENKO L., BELIAEV L.Y., MALUREANU R., LINDVOLD L.R., TAKAYAMA O., ANDERSEN P.E., LAVRINENKO A.V., *Guided-mode resonance on pedestal and half-buried high-contrast gratings for biosensing applications*, *Nanophotonics*, **10**(17), 2021: 4289–4296. <https://doi.org/10.1515/nanoph-2021-0347>
- [30] MAGNUSSON R., LEE K.J., HEMMATI H., KO Y.H., WENNER B.R., ALLEN J.W., ALLEN M.S., GIMLIN S., WEIDANZ D.W., *The guided-mode resonance biosensor: Principles, technology, and implementation*, *Proceedings of the SPIE*, Vol. 10510, *Frontiers in Biological Detection: From Nanosensors to Systems X*, 2018: 105100G. <https://doi.org/10.1117/12.2288795>
- [31] XU G., FENG C., MA F., *Review of medical image segmentation based on UNet*, *Journal of Frontiers of Computer Science and Technology*, **17**(8), 2023: 1776–1792. <https://doi.org/10.3778/j.issn.1673-9418.2301044>
- [32] ZHUO B., JIANG D., DI Y., *Test-statistic correlation and data-row correlation*, *Statistics & Probability Letters*, **167**, 2020: 108903. <https://doi.org/10.1016/j.spl.2020.108903>
- [33] ZHANG W.-W., CHEN Q., *Noise characteristics of electron multiplying charge coupled devices*, *Acta Photonica Sinica*, **38**(4), 2009: 756–760.
- [34] DEWAELE N., VANNIEUWENHOVEN N., *Which constraints of a numerical problem cause ill-conditioning?*, *Numerische Mathematik*, **156**(4), 2024: 1427–1453. <https://doi.org/10.1007/s00211-024-01427-6>
- [35] KIM C., PARK D., LEE H.-N., *Convolutional neural networks for the reconstruction of spectra in compressive sensing spectrometers*, *Proceedings of the SPIE*, Vol. 10937, *Optical Data Science II*, 2019: 109370L. <https://doi.org/10.1117/12.2509548>
- [36] BIAN J.L., ZHANG X., ZHANG X.Y., XU D.L., WANG G.H., *MCANet: shared-weight-based Multihead-Cross-Attention network for drug–target interaction prediction*, *Briefings in Bioinformatics*, **24**(2), 2023: bbad082. <https://doi.org/10.1093/bib/bbad082>

*Received December 22, 2026
in revised form January 31, 2026*