


# Physical parameter regression from black hole images using a multiscale adaptive neural network\*

Jialei Wei (魏佳蕾)<sup>1</sup> Ao Liu (刘敖)<sup>1†</sup> Dejiang Li (李德江)<sup>1</sup> Cuihong Wen (温翠红)<sup>1‡</sup> 

<sup>1</sup>College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China

**Abstract:** High-precision regression of physical parameters from black hole images generated by General Relativistic Ray Tracing (GRRT) is essential for investigating spacetime curvature and advancing black hole astrophysics. However, owing to limitations in observational resolution, high observational costs, and imbalanced distributions of positive and negative samples, black hole images often suffer from data scarcity, sparse parameter spaces, and complex structural characteristics. These factors pose significant challenges to conventional regression methods based on simplified physical models. To overcome these challenges, this study introduces the Multiscale Adaptive Network (MANet), a novel regression framework grounded in deep learning. MANet integrates an Adaptive Channel Attention (ACA) module to selectively enhance features in physically informative regions. Meanwhile, a Multiscale Enhancement Feature Pyramid (MEFP) is employed to capture fine-grained spatial structures, such as photon rings and accretion disks, while alleviating information loss due to downsampling. Experimental evaluations on GRRT-simulated datasets demonstrate that MANet substantially improves parameter estimation accuracy and generalization capability in high-dimensional parameter spaces, outperforming existing baseline approaches. This framework presents a promising avenue for high-precision parameter regression in Event Horizon Telescope (EHT) data analysis and broader astrophysical imaging applications characterized by sparse and noisy data.

**Keywords:** black hole image analysis, parameter regression, deep neural network

**DOI:** 10.1088/1674-1137/adf542 **CSTR:** 32044.14.ChinesePhysicsC.49125105

## I. INTRODUCTION

In recent years, the Event Horizon Telescope (EHT) successfully imaged the supermassive black holes Sagittarius A\* and M87 [1–3], providing valuable observational evidence for the study of strong gravitational fields and general relativity. These images reveal the black hole shadow, photon rings, and accretion disk radiation structures, which contain information about spacetime geometry, plasma conditions, and gravitational dynamics [4–8]. Inverting physical parameters—such as black hole spin, accretion disk thickness, and electron temperature—from such images aids in the reconstruction of electromagnetic processes near the black hole and offers important tests of general relativity and insights into black hole formation and evolution [9–11].

With the development of General Relativistic Ray Tracing (GRRT) simulations [12–14], researchers can

generate high-fidelity synthetic black hole images for modeling purposes. However, the limited observational resolution and high cost of data acquisition have resulted in sparse datasets and a highly nonlinear parameter space, both of which pose significant challenges to accurate regression. Traditional regression methods predominantly rely on physics-based models that compare GRRT-simulated images with observational data to infer black hole parameters. Common approaches include forward modeling combined with grid search, where parameter spaces are exhaustively sampled to identify best-fit models [15], and Bayesian inference methods, which incorporate prior knowledge and probabilistic frameworks to estimate posterior distributions of parameters [16, 17]. Although these methods provide interpretability, they depend on idealized assumptions about accretion disk structure, magnetic field configuration, and radiation mechanisms, which may not fully capture the complexities of real observa-

Received 6 June 2025; Accepted 22 July 2025; Published online 23 July 2025

\* Supported by the Natural Science Foundation of Hunan Province under grant No. 2023JJ30384; the National Natural Science Foundation of China under Grants No. 12374408; the University Students' Innovation and Entrepreneurship Training Program Project

<sup>†</sup> E-mail: liuao@hunnu.edu.cn

<sup>‡</sup> E-mail: cuihongwen@hunnu.edu.cn



Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Article funded by SCOAP<sup>3</sup> and published under licence by Chinese Physical Society and the Institute of High Energy Physics of the Chinese Academy of Sciences and the Institute of Modern Physics of the Chinese Academy of Sciences and IOP Publishing Ltd

tions [18–21]. Moreover, the computational cost of dense sampling in high-dimensional parameter spaces limits the efficiency and scalability of these methods. Under conditions of limited resolution and noisy data, such approaches are sensitive to initialization and may converge to local optima or produce nonphysical results. The nonlinear coupling of black hole image features further complicates explicit modeling, restricting the ability to capture detailed structures and maintain regression robustness [22].

Deep learning approaches have demonstrated strong capabilities in nonlinear modeling and feature extraction across various computer vision tasks, such as classification, recognition, and regression [23–27]. Convolutional neural networks (CNNs) can learn complex mappings between images and parameters from data, improving prediction accuracy and efficiency [28]. Compared with traditional physics-based methods, deep learning reduces modeling bias and improves performance in the presence of observational noise and limited data, making it particularly suitable for black hole image-based parameter estimation [29–31].

However, the particular characteristics of black hole images pose challenges for training and generalization. Key physical structures such as photon rings [32] are subtle and sensitive to noise, complicating feature extraction [33, 34]. Additionally, current GRRT-based synthetic datasets are relatively small for deep network training, increasing the risk of overfitting and limiting the ability to capture complex physical phenomena. Therefore, it is essential to design an efficient neural network architecture adapted to the specific characteristics of black hole imaging data to enhance regression performance.

To address these challenges, we propose the Multiscale Adaptive Network (MANet), which enhances modeling capacity and prediction robustness for black hole image parameter regression. MANet adopts a ResNet backbone and integrates two key modules in intermediate layers [35]: a Multiscale Enhancement Feature Pyramid (MEFP) and an Adaptive Channel Attention (ACA) mechanism. MEFP fuses spatial features across multiple scales to better represent multilayered structures such as photon rings and accretion disks, mitigating information loss caused by downsampling. ACA dynamically adjusts channel weights to enhance feature responses in important physical regions, filtering redundant information and improving feature discrimination. Ablation studies indicate that both modules contribute to performance improvements independently, and their combination yields further gains. Experiments under simulated observational noise show that MANet maintains stable performance and generalization. Even with limited GRRT training data, MANet effectively captures complex physical features, demonstrating its applicability in data-constrained scenarios.

This paper is organized as follows: Sec. II introduces the proposed MANet, detailing its overall architecture and two key components: the ACA mechanism and MEFP. MANet aims to overcome the challenges to high-precision physical parameter regression posed by limited observational resolution, costly observations leading to sparse and uneven data, and the intrinsic complexity of black hole images. Sec. III presents the experimental settings and comprehensive analyses. It includes the dataset and preprocessing pipeline, evaluation metrics, and a comparison of different learning rate scheduling strategies. Visual interpretations of attention-induced activations are provided to examine how the model focuses on key relativistic structures such as photon rings and accretion disks. In addition, ablation studies are conducted to quantify the contributions of individual modules, while robustness evaluations under simulated observational noise further assess the model's stability in astrophysically realistic scenarios. Sec. IV concludes the paper and outlines future directions.

## II. PROPOSED METHOD

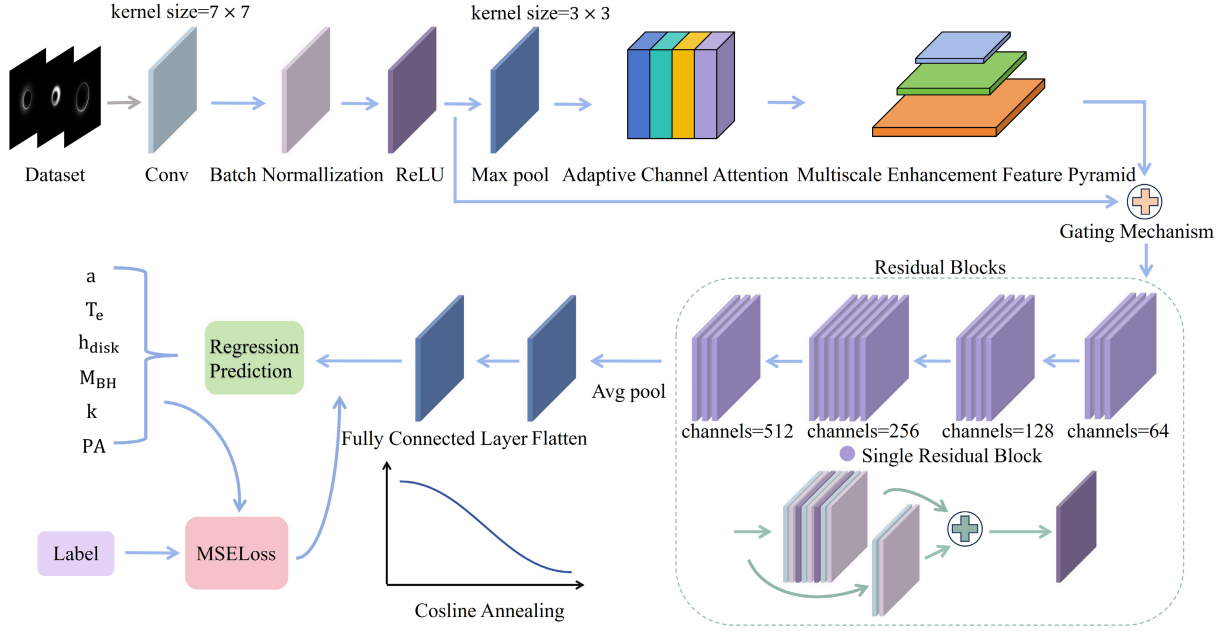
To accurately invert physical parameters from black hole images, we propose MANet, a deep convolutional architecture built upon a ResNet backbone. As illustrated in Fig. 1, MANet retains the hierarchical feature extraction capabilities of ResNet, while embedding two specialized modules: ACA and MEFP. The ACA module adaptively highlights informative feature channels, while MEFP captures multiscale spatial structures crucial for robust physical parameter regression. The enhanced features are fused with the original residual representation via a learnable channel-wise gating mechanism, enabling the model to adaptively balance raw and enriched information for robust parameter regression.

### A. Adaptive channel attention

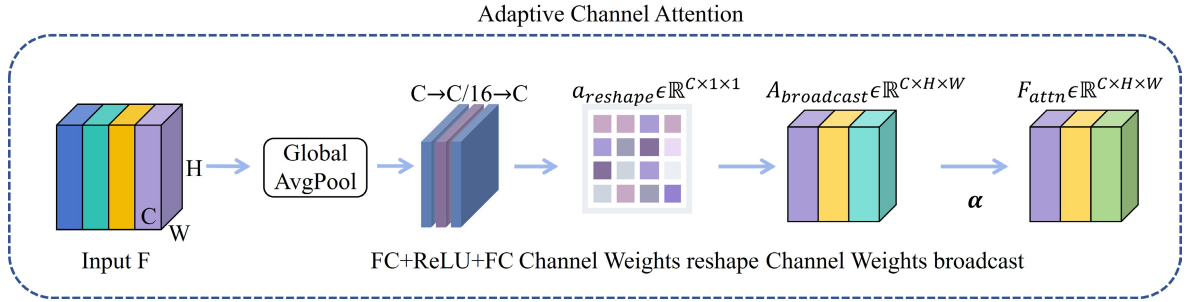
In the context of black hole image parameter regression, the complexity and variability of physical features—such as relativistic beaming-induced brightness asymmetries and jet boundaries—require the model to selectively enhance channels that encode critical diagnostic information. This motivates the introduction of an adaptive channel attention module that adaptively learns to assign dynamic importance weights to feature channels based on global contextual information extracted via global average pooling. Figure 2 illustrates the internal structure of the ACA module.

Given an input feature map  $F \in \mathbb{R}^{C \times H \times W}$ , we compute a channel-wise descriptor by averaging over spatial dimensions:

$$z = \text{GlobalAvgPool}(F) \in \mathbb{R}^C. \quad (1)$$



**Fig. 1.** (color online) Architecture of the proposed MANet for physical parameter regression from black hole images.



**Fig. 2.** (color online) Architecture of the ACA module.

This descriptor is passed through a fully connected layer to produce channel-wise attention weights:

$$a = \text{FC}(z) \in \mathbb{R}^C. \quad (2)$$

After reshaping to  $\mathbb{R}^{C \times 1 \times 1}$ , the attention weights are broadcast across spatial dimensions and applied to the input feature map:

$$F_{\text{attn}} = F \cdot (a \cdot \alpha), \quad (3)$$

where  $\alpha \in [0, 1]$  is a learnable scaling factor controlling the attention strength. This adaptive reweighting enables the model to dynamically emphasize channels corresponding to physically meaningful regions, improving interpretability and robustness under complex astrophysical conditions.

The attention-enhanced feature map  $F_{\text{attn}}$  is then passed through the MEFP module to extract rich spatial in-

formation across multiple scales, producing  $F_{\text{MEFP}}$ .

To balance the original features and the enhanced multiscale features, a learnable gating parameter  $g \in \mathbb{R}^C$  is introduced. The final output  $F_{\text{gated}}$  is computed as

$$F_{\text{gated}} = \sigma(g) \cdot F_{\text{MEFP}} + (1 - \sigma(g)) \cdot F, \quad (4)$$

where  $\sigma(\cdot)$  denotes the sigmoid activation applied channel-wise. This gating mechanism allows the network to adaptively fuse the original and enhanced features, ensuring that the most informative representation is emphasized for accurate black hole parameter regression.

## B. Multiscale enhancement feature pyramid

Black hole images generated by GRRT simulations contain structural features spanning multiple spatial scales. The photon ring appears as a thin, high-contrast boundary, whereas the accretion disk manifests broader, smoother intensity variations. Additionally, relativistic beaming and gravitational lensing create complex emis-

sion patterns distributed across different receptive fields. When extracting features of certain key physical parameters that are not clearly observable—such as black hole spin and disk inclination—it is crucial to capture information at both fine and coarse spatial scales.

To address this challenge, we propose the MEFP module. Given an input feature map  $F_{\text{attn}} \in \mathbb{R}^{C \times H \times W}$  produced by the preceding attention mechanism, where  $C$  is the number of channels and  $H, W$  denote spatial dimensions, MEFP employs three parallel convolutional branches with distinct kernel sizes to capture multiscale information. The multi-scale feature extraction design of MEFP is shown in Fig. 3.

The first branch applies a  $1 \times 1$  convolution with weights  $W_1$  and bias  $b_1$ , followed by a nonlinear activation  $\sigma(\cdot)$  (e.g., ReLU), producing

$$F_1 = \sigma(W_1 * F_{\text{attn}} + b_1), F_1 \in \mathbb{R}^{C' \times H \times W}. \quad (5)$$

This branch preserves local spatial details critical for detecting sharp edges such as the photon ring.

The second branch captures medium-scale context by sequentially applying a  $1 \times 1$  convolution  $W_2^{(1)}, b_2^{(1)}$  for channel reduction, followed by a  $3 \times 3$  convolution  $W_2^{(3)}, b_2^{(3)}$  with padding 1 to maintain spatial dimensions:

$$F_2 = \sigma(W_2^{(3)} * \sigma(W_2^{(1)} * F_{\text{attn}} + b_2^{(1)}) + b_2^{(3)}), F_2 \in \mathbb{R}^{C' \times H \times W}. \quad (6)$$

This branch extracts features related to mid-scale spatial patterns, such as the photon ring's thickness and local asymmetries.

The third branch extracts coarse-scale features by applying a  $1 \times 1$  convolution  $W_3^{(1)}, b_3^{(1)}$ , followed by a  $5 \times 5$  convolution  $W_3^{(5)}, b_3^{(5)}$  with padding 2:

$$F_3 = \sigma(W_3^{(5)} * \sigma(W_3^{(1)} * F_{\text{attn}} + b_3^{(1)}) + b_3^{(5)}), F_3 \in \mathbb{R}^{C' \times H \times W}. \quad (7)$$

This branch captures large receptive field information encompassing broad emission patterns and global spatial context of the accretion disk.

As the convolutional branches use kernel-size-specific zero-padding (0 for  $1 \times 1$ , 1 for  $3 \times 3$ , and 2 for  $5 \times 5$ ), all output feature maps  $F_1, F_2, F_3$  retain identical spatial dimensions  $H \times W$ . This alignment enables their direct concatenation along the channel dimension:

$$F_{\text{concat}} = \text{Concat}[F_1, F_2, F_3] \in \mathbb{R}^{3C' \times H \times W}. \quad (8)$$

Finally, a  $1 \times 1$  convolution with weights  $W_f$  and bias  $b_f$  fuses the concatenated features and reduces the channel dimension back to  $C$ :

$$F_{\text{MEFP}} = \sigma(W_f * F_{\text{concat}} + b_f), F_{\text{MEFP}} \in \mathbb{R}^{C \times H \times W}. \quad (9)$$

By integrating information across multiple receptive fields, MEFP mitigates the loss of critical spatial details caused by fixed kernel sizes and downsampling. It preserves fine-grained structures like the sharp photon ring edges while simultaneously capturing broader accretion disk morphology. This enriched feature representation enhances the model's ability to regress black hole physical parameters accurately, reflecting the complex spacetime geometry and accretion physics encoded in the images.

### III. EXPERIMENTS AND ANALYSIS

#### A. Dataset and pre-processing

To construct a dataset tailored for black hole image parameter regression, we employ GRRT simulations to generate high-fidelity synthetic images that faithfully capture radiation from accretion flows under strong gravitational fields. GRRT enables precise treatment of light propagation effects such as gravitational lensing and Doppler beaming, ensuring that the simulated images are consistent with realistic observational conditions, such as those of the M87 black hole.

Each image in our dataset simulates a 230 GHz observation with a field of view of  $128 \times 128$  microarcseconds and a  $256 \times 256$  pixel resolution, reflecting commonly adopted settings in GRRT simulations related to the EHT. The observer inclination is fixed at  $163^\circ$ , and the total

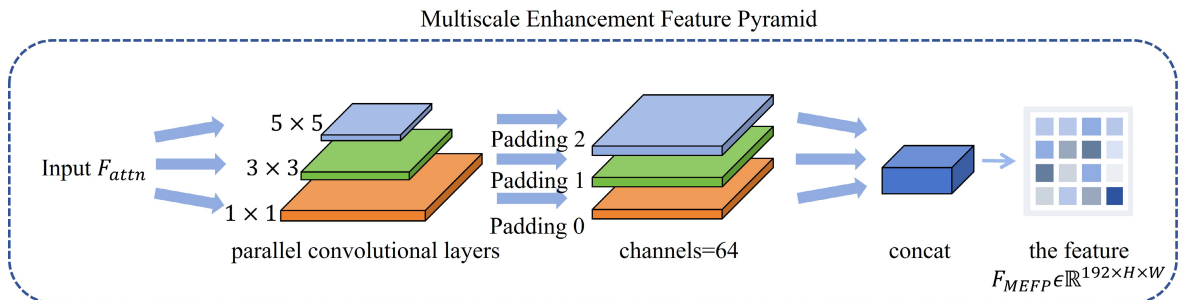


Fig. 3. (color online) Architecture of the MEFP module.



flux of each image is normalized to 0.5 Jy for consistency.

To ensure broad coverage of the physical parameter space, we uniformly sample six key parameters: black hole spin, mass, electron temperature, disk thickness, Keplerian velocity factor, and position angle, while randomly setting the disk rotation direction. In total, we generate 2157 images, each annotated with corresponding ground-truth physical parameters, forming a comprehensive dataset suitable for training and evaluating regression models. The ranges of these physical parameters are listed in Table 1, which provides an overview of the sampling space for GRRT-based simulation.

**Table 1.** Physical parameters sampled for GRRT-based black hole image simulation and their value ranges.

Parameter	Symbol	Range
Black hole spin	$a$	$[-1, 1]$
Black hole mass	$M_{\text{BH}}$	$[2 \times 10^9 M_{\odot}, 10^{10} M_{\odot}]$
Electron temperature	$T_e$	$[10^9 \text{ K}, 10^{12} \text{ K}]$
Accretion disk thickness	$h_{\text{disk}}$	$[0.1, 0.8]$
Keplerian factor	$k$	$[0, 1]$
Position angle	$PA$	$[0^\circ, 360^\circ]$

To ensure effective learning in black hole image parameter regression, it is critical to address the disparity in scale and units across physical parameters. Directly feeding raw values into the model may lead to instability or gradient inefficiency due to magnitude imbalances. To mitigate this issue, we adopt a standardization strategy that normalizes each parameter dimension to a standard normal distribution with zero mean and unit variance. This follows established practices in scientific data regression and ensures scale consistency across parameters.

Given the  $k$ -th physical parameter set  $x_1, x_2, \dots, x_n$ , we compute the empirical mean and standard deviation as

$$\mu_k = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma_k = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_k)^2}. \quad (10)$$

Each sample is then standardized by

$$Z_r = \frac{x_r - \mu_r}{\sigma_r}, \quad r \in \{1, 2, \dots, 6\}. \quad (11)$$

This transformation not only improves training stability and convergence by eliminating scale-induced biases but also enhances the model's ability to learn joint representations across heterogeneous physical quantities. Importantly, the transformation is reversible—original phys-

ical values can be recovered from predicted outputs using stored normalization statistics, preserving interpretability and physical significance.

In our implementation, normalization is applied independently to each parameter category to avoid dominance by parameters with larger magnitudes. This parameter-wise standardization ensures balanced gradient flow and improves the model's capacity to extract cross-scale physical features relevant to GRRT-simulated black hole images.

To facilitate model training and evaluation, the dataset is partitioned into training, validation, and test subsets. The dataset comprises 2157 GRRT-simulated black hole images annotated with corresponding physical parameters.

Following standard machine learning protocols, we allocate 1725 samples (80.0%) for training, 216 samples (10.0%) for validation, and 216 samples (10.0%) for testing. The training set is used for iterative model optimization, the validation set guides hyperparameter tuning, and the test set provides an independent assessment of the model's generalization performance.

To prevent distributional shifts and ensure statistical consistency across subsets, we adopt stratified random sampling. This ensures that critical physical parameters—such as black hole mass and accretion disk thickness—are uniformly distributed across the splits, thereby minimizing potential evaluation bias and preserving the representativeness of each subset. This ensures fair evaluation of model generalization and supports reliable performance benchmarking in subsequent experiments.

## B. Evaluation indicators

In black hole image parameter regression tasks, accurately assessing model performance and interpretability requires appropriate evaluation metrics. The coefficient of determination, denoted as  $R^2$ , is widely adopted as the primary metric because it effectively quantifies the proportion of variance in the target parameters explained by the model [36, 37].

Mathematically,  $R^2$  is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (12)$$

where  $y_i$  is the true value of the  $i$ -th sample,  $\hat{y}_i$  is the predicted value from the model, and  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  is the mean of the true values over all samples.

This metric provides an interpretable measure of goodness-of-fit, with values closer to 1 indicating better explanatory power of the model for the variability of physical parameters.

### C. Learning rate scheduling strategy selection: cosine annealing

In GRRT-simulated black hole image regression tasks, we compared four prevalent learning-rate schedules—constant rate, step decay, exponential decay and cosine annealing—and observed that cosine annealing substantially outperforms the others in validation  $R^2$ . The initial gentle decay phase of cosine annealing preserves a high learning rate for extensive global exploration of the intricate, multiscale spacetime features in black hole images, thereby avoiding premature convergence to suboptimal solutions. Its mid-training rapid reduction concentrates parameter updates in promising regions, accelerating convergence and improving optimization efficiency. Finally, the late slow taper facilitates precise fine-tuning, enhancing the accuracy of subtle spacetime curvature estimations.

Under a unified experimental setup with an initial learning rate of  $9 \times 10^{-4}$ , 500 training epochs, and identic-

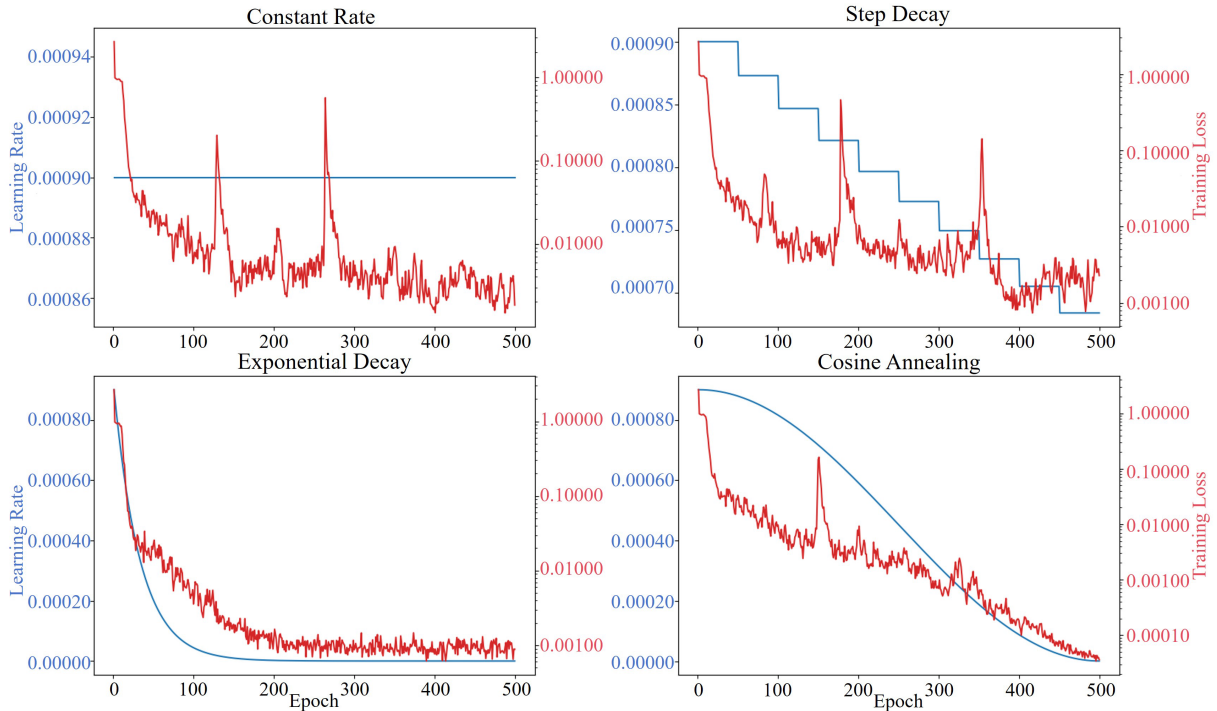
al train/validation splits, Table 2 summarizes the  $R^2$  performance of each learning rate schedule across six target parameters and Fig. 4 illustrates the learning rate and loss dynamics for parameter  $a$ , shown here as a representative case.

As shown in Table 2, cosine annealing consistently achieves the highest  $R^2$  scores across all six physical parameters. For  $T_e$  and  $h_{\text{disk}}$ —parameters that are particularly sensitive to high-frequency image details—the relative improvements reach 1.3% and 1.1%, respectively, compared to the best alternative. Moreover, for  $PA$ , which encodes orientation-related information, cosine annealing surpasses step decay by more than 10%. These improvements demonstrate that the cosine learning rate schedule is better suited to the intricate, multi-modal optimization landscape characteristic of GRRT-based regression tasks.

Figure 4 shows the training dynamics for parameter  $a$  as a representative example. Among the four learning rate scheduling strategies compared, the final loss values are

**Table 2.**  $R^2$  Scores for various learning rate schedules in black hole parameter regression.

	$a$	$T_e$	$h_{\text{disk}}$	$M_{\text{BH}}$	$k$	$PA$
Constant Rate	0.9866	0.8253	0.8910	0.9980	0.9691	0.8987
Step Decay	0.9881	0.9836	0.9653	0.9987	0.9583	0.8565
Exponential Decay	0.9582	0.9943	0.9841	0.9986	0.9534	0.8182
Cosine Annealing (ours)	<b>0.9906</b>	<b>0.9969</b>	<b>0.9927</b>	<b>0.9993</b>	<b>0.9725</b>	<b>0.9602</b>



**Fig. 4.** (color online) Training dynamics of learning rate and loss under different scheduling strategies (shown for parameter  $a$ ). Loss is plotted on a logarithmic scale to better illustrate convergence across several orders of magnitude.

as follows: constant rate achieves a loss of  $1.838 \times 10^{-3}$ , step decay  $2.285 \times 10^{-3}$ , exponential decay  $8.843 \times 10^{-4}$ , while cosine annealing achieves the lowest loss of  $3.618 \times 10^{-5}$ . This indicates that cosine annealing enables more effective optimization and better convergence in black hole parameter regression tasks.

#### D. Visual interpretation of the attention mechanism

To qualitatively assess how the adaptive channel attention mechanism influences spatial feature localization in black hole image regression, we employed Gradient-weighted Class Activation Mapping (Grad-CAM) [38–40] to generate activation heatmaps. These visualizations highlight the input regions contributing most to the model's predictions, offering insight into how attention reshapes feature weighting in alignment with physical structures. This is particularly relevant for GRRT-simulated black hole images, which exhibit bright compact features such as the photon ring and inner accretion disk, surrounded by diffuse backgrounds.

As shown in Fig. 5, comparing activation patterns between the baseline and attention-enhanced models reveals meaningful differences. The baseline model already demonstrates reasonable localization of the black hole structure, typically activating circular or wide ring-shaped regions. However, with the introduction of the attention mechanism, the model produces more structured and spatially selective heatmaps. The broad activation ring becomes thinner and more sharply delineated, occasionally fragmenting into arc-like segments that more precisely trace the morphology of the photon ring and the high-

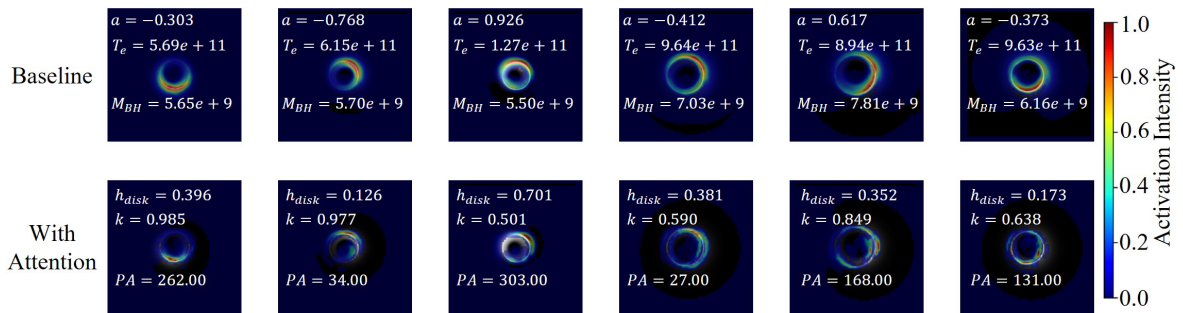
emission regions of the accretion disk.

These localized, high-gradient zones encode key physical information such as black hole spin, viewing angle, and accretion dynamics. By concentrating attention on these critical structures, the model enhances its ability to capture physically informative representations, leading to improved regression accuracy and stronger alignment between learned features and the underlying spacetime geometry.

#### E. Ablation experiments

To evaluate the synergistic contribution of MEFP and the attention module to the performance of black hole image parameter regression, we conducted a series of ablation experiments. Starting from the complete network, we successively removed the MEFP module and the attention mechanism to construct reduced variants. All experiments were carried out under the same training configuration and data partition as described in the previous sections, with the coefficient of determination  $R^2$  used as the evaluation metric across six physical parameters derived from GRRT-simulated black hole images. The results of these experiments are summarized in Table 3, enabling a controlled assessment of how each architectural component influences the model's overall regression performance.

As presented in Table 3, both the MEFP module and the channel attention mechanism independently contribute to performance gains across all six regression targets. When the MEFP is added to the baseline ResNet50, the  $R^2$  score for electron temperature  $T_e$  increases from



**Fig. 5.** (color online) Activation heatmaps comparing baseline and attention-enhanced models for black hole image parameter regression. Each column presents two heatmaps derived from the same input image and identical ground truth parameters. For readability, the upper image (baseline model) displays three of the parameters, while the lower image (with attention model) displays the remaining three. A color scale representing relative activation intensity is included to facilitate interpretation of the spatial focus of each model.

**Table 3.**  $R^2$  Performance comparison of MANet and ablated variants on simulated black hole datasets.

	$a$	$T_e$	$h_{\text{disk}}$	$M_{\text{BH}}$	$k$	$PA$
ResNet50	0.9359	0.9284	0.9708	0.9986	0.9556	0.8569
ResNet50+MEFP	0.9887	0.9819	0.9871	0.9992	0.9603	0.9181
ResNet50+ACA	0.9864	0.9759	0.9912	0.9990	0.9561	0.9459
MANet (ours)	<b>0.9906</b>	<b>0.9969</b>	<b>0.9927</b>	<b>0.9993</b>	<b>0.9725</b>	<b>0.9602</b>

0.9284 to 0.9819—a significant gain of 0.0535. Similarly,  $h_{\text{disk}}$  improves from 0.9708 to 0.9871, highlighting the module’s strength in capturing hierarchical spatial cues. This suggests that MEFP is particularly effective for parameters governed by multiscale morphological gradients, as commonly found in GRRT-generated accretion flows.

The attention module, on the other hand, delivers the most pronounced improvements in parameters like spin  $a$  and position angle  $PA$ . For example, the  $R^2$  for  $PA$  rises from 0.8569 to 0.9459—an absolute increase of 0.089—indicating its ability to enhance directional signals and suppress irrelevant background emissions through adaptive channel-wise reweighting.

When both modules are integrated into the full MCNet architecture, the model achieves consistent and superior results across all parameters. Notably, the  $R^2$  values reach 0.9906 for  $a$ , 0.9969 for  $T_e$ , and 0.9602 for  $PA$ , outperforming all ablated variants. This synergy between multiscale feature abstraction and attentive refinement enables the model to better disentangle complex emission patterns, leading to more robust and physically accurate parameter estimations under realistic image conditions.

#### F. Robustness evaluation under simulated observational noise

In practical astronomical observations, imaging data inevitably suffer from various noise sources owing to instrument limitations and environmental factors. To realistically simulate these unavoidable noise disturbances and further evaluate the robustness of the proposed framework under real-world imaging conditions, we conducted additional experiments by introducing Gaussian noise to the test images. Gaussian noise is a common artifact in image acquisition, particularly in high-sensitivity astronomical observations, where signal instability and instrumental interference are prevalent. In this study, we injected zero-mean Gaussian noise with standard deviations of 10 and 20—referred to as Gauss10 and Gauss20, respectively—into the test set to simulate varying levels of signal degradation typically encountered in practical scenarios [41–44].

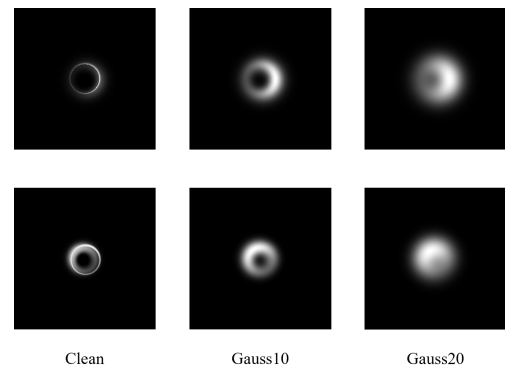
Figure 6 presents representative examples before and after noise injection, while Table 4 reports the  $R^2$  scores

for different models across six physical parameters under each noise condition.

Experimental results, as shown in Table 4, indicate that under moderate noise conditions (Gauss10), the proposed MANet maintains high regression accuracy across all parameters, demonstrating robustness against mild perturbations. As the noise level increases to Gauss20, MANet continues to outperform the baseline model in most cases, demonstrating robust generalization. However, the regression accuracy of the disk height  $h_{\text{disk}}$  declines significantly, with the  $R^2$  score dropping to  $-0.0060$ , suggesting that this physical parameter is particularly sensitive to noise interference. This sensitivity is mainly due to the inherently subtle and fine-grained morphological features associated with  $h_{\text{disk}}$  in the GRRT images, which become increasingly obscured by higher noise levels, as visually illustrated in Fig. 6. By contrast, parameters such as the electron temperature  $T_e$  and black hole mass  $M_{\text{BH}}$  exhibit relative robustness, maintaining high  $R^2$  scores even under severe noise conditions. These observations highlight the intrinsic difficulty of accurately estimating parameters characterized by delicate image features in noisy observational environments.

#### IV. CONCLUSION

In this study, we propose MANet, a ResNet-based regression network that integrates MEFP and the ACA module to improve the regression of black hole image parameters. Traditional approaches, such as forward modeling and Bayesian inference, typically rely on ex-



**Fig. 6.** Representative examples of black hole images before and after Gaussian noise injection

**Table 4.** Comparison of model parameter estimation  $R^2$  scores under simulated observational noise.

	$a$	$T_e$	$h_{\text{disk}}$	$M_{\text{BH}}$	$k$	$PA$
ResNet50 (10 $\mu\text{as}$ )	0.9444	0.8876	0.9538	0.9927	0.9754	0.8376
<b>ours</b> (10 $\mu\text{as}$ )	<b>0.9641</b>	<b>0.9976</b>	<b>0.9775</b>	<b>0.9960</b>	<b>0.9807</b>	<b>0.8998</b>
ResNet50 (20 $\mu\text{as}$ )	0.6868	0.8862	$-0.0949$	0.9742	0.9491	0.8507
<b>ours</b> (20 $\mu\text{as}$ )	<b>0.7029</b>	<b>0.9862</b>	<b><math>-0.0060</math></b>	<b>0.9744</b>	<b>0.9580</b>	<b>0.8601</b>



haustive sampling of the parameter space or computationally intensive probabilistic computations, which can be inefficient and poorly scalable in high-dimensional or data-scarce settings. By contrast, the MEFP module captures fine-grained physical structures through multiscale feature fusion, while the ACA module highlights critical regions, such as the photon ring and brightness asymmetries, using a dynamic attention mechanism. Leveraging deep learning, MANet learns complex nonlinear mappings directly from data, improving estimation accuracy while significantly reducing inference time. Experimental results show that both MEFP and ACA independently contribute to performance gains, and their combination further enhances accuracy and robustness. Even under conditions of limited data and simulation noise, MANet exhibits strong generalization ability and stability, demonstrating the effectiveness of combining multiscale feature fusion with adaptive attention for high-precision astrophysical regression tasks. In the current study, we focused on the regression of six representative parameters. In future research, we plan to incorporate temporal features and uncertainty quantification into the network

architecture to improve reliability and interpretability, further enhancing the model's applicability to real observational scenarios.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Science Data Bank at <https://www.doi.org/10.5281/zenodo.15846647>

## APPENDIX A: ADDITIONAL GRAD-CAM VISUALIZATIONS

In addition to the six examples presented in the main text, this appendix provides supplementary Grad-CAM visualizations of several additional samples. These expanded examples offer a more comprehensive understanding of the attention mechanism's behavior, including cases with clear focus as well as ambiguous or challenging scenarios. Figure A1 illustrates these additional Grad-CAM maps, which further demonstrate the interpretability and robustness of the model's attention mechanism by



**Fig. A1.** (color online) Additional Grad-CAM visualizations illustrating the attention maps of diverse samples. Typical cases and challenging scenarios are included to further demonstrate the interpretability and robustness of the model.

highlighting both typical and difficult cases.

## References

- [1] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **875**, L1 (2019)
- [2] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **930**, L12 (2022)
- [3] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **930**, L14 (2022)
- [4] J. Dexter, E. Agol, P. C. Fragile *et al.*, *Astrophys. J.* **717**, 1092 (2010)
- [5] B. Hamil and B. C. Lütftüoğlu, *Chin. Phys. C* **48**, 055102 (2024)
- [6] S. Yuan, C. Luo, Z. Hu *et al.*, *Chin. Phys. C* **49**, 025103 (2025)
- [7] J. Sun, Y. Liu, W. Qian *et al.*, *Chin. Phys. C* **47**, 025104 (2023)
- [8] R. Narayan and E. Quataert, *Nature* **615**, 597 (2023)
- [9] T. Johannsen, *Class. Quantum Grav.* **33**, 124001 (2016)
- [10] S. Chen, J. Jing, W. Qian *et al.*, *Sci. China Phys. Mech. Astron.* **66**, 260401 (2023)
- [11] Z. Younsi, D. Psaltis, and F. Özel, *Astrophys. J.* **942**, 47 (2023)
- [12] J. Dexter and E. Agol, *Astrophys. J.* **696**, 1616 (2009)
- [13] C. Chan, D. Psaltis, F. Özel *et al.*, *Astrophys. J.* **799**, 1 (2015)
- [14] C. Böhmer, E. Jenson, and R. Lazkoz, *Universe* **9**(4), 166 (2023)
- [15] R. Lu, A. E. Broderick, F. Baron *et al.*, *Astrophys. J.* **788**, 120 (2014)
- [16] J. Kim, D. P. Marrone, C. Chan *et al.*, *Astrophys. J.* **832**, 156 (2016)
- [17] D. O. Chang, M. D. Johnson, P. Tiede *et al.*, *Astrophys. J.* **974**, 143 (2024)
- [18] P. C. Fragile, *The Physics of Accretion onto Black Holes*, (Springer New York, 2015), p.87
- [19] M. Mościbrodzka, C. F. Gammie, J. C. Dolence *et al.*, *Astrophys. J.* **706**, 497 (2009)
- [20] B. You, X. W. Cao, and Z. Yan, *Science* **381**, 961 (2023)
- [21] K. Ebisawa and T. Kawaguchi, *Advances in Space Research* **38**, 2862 (2006)
- [22] S. Guo, Y. Huang, E. Liang *et al.*, *Astrophys. J.* **975**, 237 (2024)
- [23] Y. L. Cun, Y. Bengio, G. Hinton *et al.*, *Nature* **521**, 436 (2015)
- [24] S. Khan, M. Naseer, M. Hayat *et al.*, *ACM Comput. Surv.* **54**, 41 (2022)
- [25] X. Wei, Y. Song, O. M. Aodha *et al.*, *TPAMI* **44**(12), 8927 (2022)
- [26] M. A. Savelonas, C. N. Veinidis, and T. K. Bartsokas, *Remote Sens.* **14**(23), 6017 (2022)
- [27] N. Le, V. S. Rathour, K. Yamazaki *et al.*, *Artif. Intell. Rev.* **55**, 2733 (2022)
- [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *NIPS 2012* **25**, 1097 (2012)
- [29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016)
- [30] K. Akiyam *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **875**, L4 (2019)
- [31] R. Patel, B. Roachell, S. Caño-Lores *et al.*, *Comput. Sci. Eng.* **24**, 42 (2022)
- [32] M. D. Johnson, A. Lupsasca, A. Strominger *et al.*, *Sci. Adv.* **6**, eaaz1310 (2020)
- [33] D. Psaltis, L. Medeiros, P. Christian *et al.*, *Phys. Rev. Lett.* **125**, 141104 (2020)
- [34] Z. Xu, B. Wu, and W. Yang, *Chin. Phys. C* **44**, 095106 (2020)
- [35] K. He, X. Zhang, S. Ren *et al.*, in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2016), p.770
- [36] D. Chicco, M. J. Warrens, and G. Jurman, *PeerJ Comput. Sci.* **7**, e623 (2021)
- [37] S. Lathuilière, P. Mesejo, X. Alameda-Pineda *et al.*, *TPAMI* **42**, 2065 (2020)
- [38] R. R. Selvaraju, M. Cogswell, A. Das *et al.*, in *Proceedings of the 2017 IEEE International Conference on Computer Vision* (IEEE, 2017), p.618
- [39] M. M. Musthafa, T. R. Mahesh, V. K. Vinoth *et al.*, *BMC Med. Imaging.* **24**(1), 107 (2024)
- [40] F. M. Talaat, S. A. Gamel, R. M. El-Balka *et al.*, *Cancers* **16**(21), 3668 (2024)
- [41] T. V. Martínez, N. B. Perez, and M. Brüggen, *A&A* **691**, 18 (2024)
- [42] B. J. Brewer and D. Stello, *Mon. Not. Roy. Astron. Soc.* **395**, 2226 (2009)
- [43] A. liu, Z. Zhang, S. Chen *et al.*, *ApJS* **279**, 10 (2025)
- [44] S. Aigrain and D. Foreman-Mackey, *Annu. Rev. Astron. Astrophys.* **61**, 329 (2023)