photonics

MDPI

Article

Aberration-Conditioned Attention-Driven Centroid Localization: From Simulation Mechanism to Double-Spot Experiment

Zhonghao Zhao [1,2], Jia Hou [1,2], Yuanting Liu [1], Anwei Liu [1] and Zhiping He [1,2,*]

[1] Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China; 18811692673@163.com (Z.Z.); houjia@mail.sitp.ac.cn (J.H.); liuyuanting2121@163.com (Y.L.); liuanwei@mail.sitp.ac.cn (A.L.)
[2] University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: hzping@mail.sitp.ac.cn; Tel.: +86-021-25051697

Abstract

In size, weight, and power (SWaP)-constrained optical systems, such as spaceborne LiDAR, high-precision centroid localization often relies on focal-plane measurements without dedicated wavefront sensors. Under such conditions, the nonlinear coupling between optical aberrations and sensor noise introduces systematic bias that is difficult to mitigate using conventional centroiding methods. To address this issue, we propose a physics-conditioned feature correction framework based on an aberration-conditioned attention mechanism. A hybrid CNN–Transformer architecture is employed to predict and compensate for systematic centroid bias. Specifically, convolutional layers encode the degraded spot morphology, while a multi-head attention mechanism leverages Seidel aberration coefficients to adaptively modulate spatial features for precise regression. Given the unavailability of absolute ground-truth coordinates in empirical scenarios, a physics-consistent simulation framework based on scalar diffraction theory is constructed to generate synthetic data for supervised learning. Simulation results indicate that the proposed method objectively reduces anisotropic systematic bias, achieving a localization root-mean-square error (RMSE) of 0.011 to 0.021 pixels, and maintains stable sub-pixel accuracy even under a 10% empirical prior perturbation. To evaluate generalization performance and engineering reliability, a wedge-based double-spot platform is developed to verify physical consistency via geometric invariance. Experimental results demonstrate a measured spacing standard deviation (SD) of 0.015 to 0.039 pixels. This validates the framework's transferability from theoretical simulation to controlled physical measurements, providing an algorithmic foundation for precision optical metrology in hardware-constrained environments.

Keywords: centroid localization; physics-conditioned attention; seidel aberrations; computational optical metrology; geometric invariance

Received: 26 February 2026
Revised: 16 March 2026
Accepted: 18 March 2026
Published: 20 March 2026

1. Introduction

In applications such as surveying, astronomical observation, and precision optoelectronic measurement, system performance relies on the accurate estimation of beam parameters. Spot centroid localization serves as a fundamental prerequisite for these tasks. Among existing solutions, Shack–Hartmann wavefront sensors (SHWS) are widely adopted. Supported by optimized estimators, hardware acceleration, and neural network matching, modern SHWS have achieved sub-pixel accuracy, high computational efficiency, and expanded dynamic range. These advancements have improved the practicality of SHWS in

compact optical systems [1–4]. However, integrating a microlens array imposes an intrinsic physical trade-off. It divides the incident optical energy and limits the spatial resolution of the retrieved phase [5]. Consequently, in photon-starved or highly compact environments like spaceborne LiDAR, diverting scarce signal photons to a dedicated wavefront-sensing channel is typically an unfavorable engineering trade-off [6].

In this context, performing centroid estimation relying exclusively on standard focal-plane imaging detectors, without additional wavefront sensing hardware, offers a practical hardware-efficient alternative for stringent engineering scenarios. However, achieving high-precision centroid localization in such sensorless configurations remains challenging. Conventional estimators, such as Center of Gravity (CoG) algorithms [5,7], often experience measurable accuracy reduction because asymmetric aberrations violate their centrosymmetric energy assumptions, inducing systematic biases [8–10]. To address this, recent data-driven methods have sought to directly map non-ideal spot morphologies to centroid coordinates. For instance, super-resolution convolutional neural networks (CNNs), deployed on 32×32 focal-plane arrays have achieved sub-pixel position estimation with root-mean-square errors (RMSE) ranging from 0.04 to 0.20 pixels under varying shot noise conditions [11]. Similarly, multi-scale adaptive convolutions have bounded the maximum localization error to approximately 0.15 pixels under non-uniform illumination [12]. Furthermore, to handle severe asymmetric morphological degradation, recent studies have bypassed traditional geometric constraints entirely by treating centroid localization as an end-to-end key-point detection problem, restricting the dynamic centroid deviation to within 0.1 pixels [13].

Nevertheless, while these data-driven architectures effectively suppress empirical noise and morphological degradation, they primarily rely on spatial feature mapping without explicitly incorporating the underlying physical principles of the optical system. Consequently, in addressing aberration-degraded centroid estimation, their performance can be further enhanced by recognizing that the systematic bias depends not only on local spot morphology but also on the deterministic intensity modulation governed by optical aberrations. To capture these underlying physical rules, physics-informed deep learning (PIDL) and Transformer-based attention mechanisms have emerged [14]. Going beyond purely data-driven mapping, modern PIDL frameworks have demonstrated highly accurate recovery of complex optical fields by explicitly embedding physical propagation models [15]. Inspired by this, incorporating structural physical constraints—such as exact Seidel aberration coefficients—to dynamically guide the feature extraction process provides a robust pathway to fundamentally decouple local spatial distortions from systematic centroid bias.

Despite current advances in physics-informed deep learning, existing frameworks primarily focus on global optical field recovery rather than directly correcting centroid systematic biases. The explicit integration of structural physical priors to condition attention mechanisms remains under-explored, and the supervised optimization of these estimators is constrained by the empirical difficulty of acquiring large-scale datasets with absolute sub-pixel ground-truth labels. To address these limitations within SWaP-constrained precision metrology, this study proposes a physics-conditioned computational framework to mitigate centroid localization degradation induced by aberration-noise coupling. The remainder of this manuscript systematically maps this methodology from theoretical formulation to practical deployment. Section 2 details the closed-loop measurement framework, which integrates a scalar diffraction-based forward model for accurate supervision with a prior-guided inverse solver to decouple distorted spot morphology. Following the establishment of theoretical decoupling limits under simulated conditions in Section 3, Section 4 implements a geometric invariance-based evaluation strategy using a wedge-based double-spot platform to validate physical consistency without absolute ground truth. Finally, Section 5 critically analyzes the prerequisites for field

deployment, including degradation parameterization and computational feasibility for edge hardware, before concluding the study in Section 6.

2. Physics-Conditioned Computational Measurement Framework

To address centroid localization degradation caused by aberration–noise coupling in hardware-constrained environments, this section presents a physics-conditioned computational measurement framework. Unlike purely data-driven black-box models, this software-defined approach explicitly incorporates aberration priors to decouple spatial degradation features without requiring additional wavefront sensing hardware. The overall framework, combining a forward physical model with experimental validation, is first introduced, followed by the hybrid network architecture acting as the inverse solver.

2.1. Overall Concept & Model Framework

To verify physical consistency and engineering feasibility, a dual-phase evaluation framework is established, as illustrated in Figure 1. Phase I (Forward Physical Modeling) constructs a physics-based simulation environment based on scalar diffraction theory. Synthetic datasets containing distorted spot images, aberration parameters, and ground truth labels are generated to analyze theoretical performance limits. Phase II (Experimental Validation) implements a wedge-based double-spot system that approximates the configuration of multi-beam engineering scenarios. In such practical setups, the geometrically fixed relative spacing between the two spots serves as a physically invariant reference, compensating for the absence of real-world ground truth and validating the framework's generalization under local field-variant aberrations.

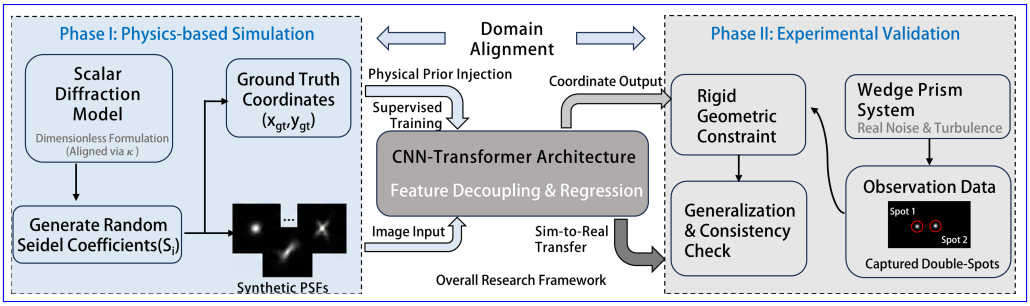

Figure 1. Overall research framework of centroid bias correction. The methodology encompasses a physics-based diffraction simulation environment (Phase I) and a wedge-based double-spot experimental platform (Phase II), which are used to generate accurate ground truth for network training and to evaluate model transfer generalization performance in controlled physical experiments using geometric rigidity constraints, respectively.

Within this framework, a physics-conditioned feature correction network is designed as the core inverse solver for centroid bias regression. As shown in Figure 2, a CNN backbone encodes distorted spot images into translation-invariant spatial features. To decouple aberration and noise effects, an aberration-conditioned attention module is introduced. Physical aberration parameters are projected as conditional tokens and jointly modeled with spatial tokens via multi-head self-attention. By using aberration information as a physical prior to dynamically weight spatial features, this mechanism establishes a nonlinear mapping from degraded spot morphology to systematic bias, enabling high-precision, forward-inference regression.

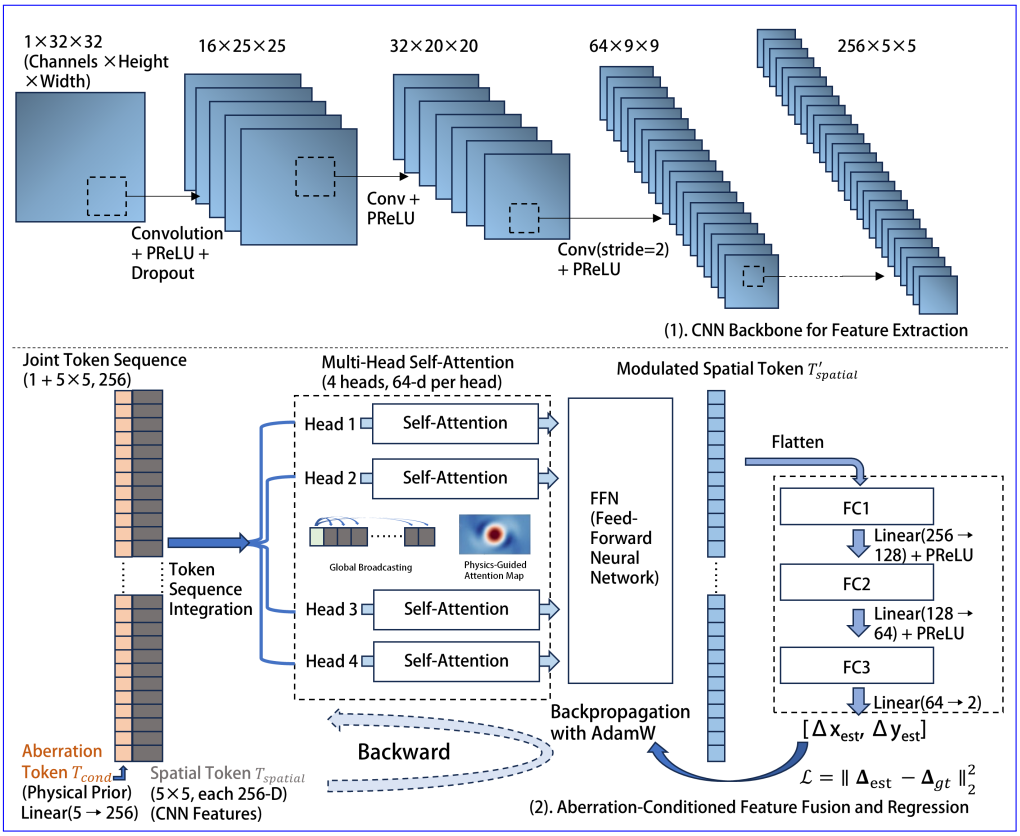

Figure 2. Architecture of the physics-conditioned feature correction network. (**1**) Image feature extraction module: a multi-layer CNN backbone network encodes the input distorted spot and extracts high-dimensional spatial features. (**2**) Aberration-conditioned correction network: aberration parameters are projected as conditional tokens, and multi-head self-attention dynamically weights and decouples CNN-extracted spatial features, enabling physics-driven centroid regression.

2.2. Forward Physical Modeling and Synthetic Data Generation

To ensure that the network learns physically meaningful aberration–spot mappings, a physics-consistent forward model is constructed based on low-order Seidel aberration theory, describing the propagation from pupil wavefront to image-plane spot.

The complex amplitude at the exit pupil of the optical system is expressed as

$$E(x, y) = A(x, y) \exp[j\Phi(\rho, \theta)] \tag{1}$$

Here, $A(x, y) = \exp\left[-\frac{1}{w_0^2}(x^2 + y^2)\right]$ represents the transverse energy distribution. This forward model assumes an untruncated Gaussian illumination to establish a generalized theoretical baseline. The physical pupil radius R serves as the normalization factor for the spatial coordinates, defining the dimensionless radial coordinate as $\rho = \sqrt{x^2 + y^2}/R$, where $\rho \in [0, 1]$.

To directly interface with the neural network without confounding local field parameters, the phase term $\Phi(\rho, \theta)$ encodes the primary structural aberrations as a dimensionless phase polynomial (in radians) defined for a specific local field:

$$\Phi(\rho, \theta) = S_I \rho^4 + S_{II} \rho^3 \cos\theta + S_{III} \rho^2 \cos 2\theta + S_{IV} \rho^2 + S_V \rho \cos\theta \tag{2}$$

In Equation (2), the traditional field-dependent variables (associated with normalized field height h) and the wavenumber scaling factor ($2\pi/\lambda$) have been mathematically absorbed into the effective Seidel phase coefficients $\mathbf{c} = [S_I, S_{II}, S_{III}, S_{IV}, S_V]^{\mathrm{T}}$. Consequently,

these coefficients function strictly as dimensionless phase weights. Seidel polynomials are adopted here because they provide a direct parameterization of the low-order structural aberrations emphasized in this study. A comparative discussion with Zernike-based representations is provided in Section 5.

Based on Fraunhofer diffraction theory, the image-plane intensity distribution $I(u, v)$ is calculated as the squared modulus of the Fourier transform of the complex pupil function. To decouple the optical wave propagation from specific absolute hardware dimensions while maintaining scale equivalence with the experimental setup, the forward model is evaluated within a normalized dimensionless coordinate system.

Specifically, to map this dimensionless model to the physical LiDAR detector, a spatial scaling factor $\kappa = L/(2R)$ is introduced, where L denotes the spatial span of the computational grid. In discrete Fourier optics, when evaluated on an $N \times N$ grid (e.g., $N = 500$), κ determines the dimensionless spatial frequency resolution $\Delta\nu = 1/L$. This resolution subsequently dictates the pixel-level spatial sampling of the point spread function (PSF) on the image plane. By configuring κ, the simulated spatial sampling rate is geometrically constrained to match the physical resolution of the target detector. Finally, the discretized far-field intensity distribution is augmented with a mixed-noise model and randomly translated within ± 5 pixels around the field center to simulate pointing variations. The resulting target is then cropped to a 32×32 pixel observation window, yielding the degraded spot morphology corresponding to the input tensor dimensions of the neural network.

For each noisy spot image, an initial centroid estimate is computed using the standard Center of Gravity (CoG) method over the localized observation window:

$$x_{raw} = \frac{\sum_i \sum_j i \cdot I(i,j)}{\sum_i \sum_j I(i,j)}, \quad y_{raw} = \frac{\sum_i \sum_j j \cdot I(i,j)}{\sum_i \sum_j I(i,j)} \tag{3}$$

where (i, j) denote the discrete two-dimensional pixel coordinates within the cropped observation window. Subsequently, the systematic bias is quantified as the spatial deviation between this initial CoG estimate and the true centroid position (x_{gt}, y_{gt}) established in the simulation, given by:

$$\Delta x_{\text{ref}} = x_{\text{raw}} - x_{\text{gt}}, \quad \Delta y_{\text{ref}} = y_{\text{raw}} - y_{\text{gt}} \tag{4}$$

Because the relationship between the effective phase coefficient vector $\mathbf{c}$ and spot intensity distribution I is highly nonlinear and difficult to solve analytically, I and $\mathbf{c}$ are used as network inputs, and the systematic bias $(\Delta x, \Delta y)$ serves as the regression target. A large-scale synthetic dataset containing 50,000 samples is generated. The effective Seidel phase coefficients are uniformly sampled within the range $[-4\pi, 4\pi]$ rad (equivalent to an optical path difference of $\pm 2\lambda$) to cover system states from near diffraction-limited to strongly aberrated conditions. In addition, Poisson–Gaussian mixed noise with SNR following a $U(5, 30)$ dB distribution is added to account for different illumination levels.

2.3. Physics-Guided Inverse Solver Based on Conditional Self-Attention

To achieve high-precision localization under multi-factor coupling without relying on hardware optical correction, a software-defined inverse solver comprising CNN encoding, attention-based feature fusion, and fully connected regression is designed, as illustrated in Figure 2.

2.3.1. Image Feature Encoder (CNN Backbone)

The CNN backbone extracts local spatial features from spot images. It consists of six cascaded convolutional modules, as shown in Figure 2(1). Each convolutional layer is followed by a parametric ReLU (PReLU) activation function, which allows for adaptive

adjustment of the negative-slope region and enhances sensitivity to weak edge signals and asymmetric trailing features. Dropout is applied to improve generalization. Spatial dimensionality is progressively reduced using strided convolutions, while channel depth is increased. This process suppresses high-frequency noise while preserving the global topological structure of the spot. The resulting high-dimensional feature maps $F \in \mathbb{R}^{C\times H'\times W'}$ are reshaped into a sequence of spatial tokens $T_{spatial}$ to interface with the subsequent Transformer-based module.

2.3.2. Aberration-Conditioned Attention Fusion Module

To enable physics-guided correction in deep feature space, a conditional self-attention mechanism based on token concatenation is introduced (Figure 2(2)). Instead of conventional channel concatenation, a hybrid token sequence is constructed by projecting the Seidel coefficient vector c into an aberration token $T_{cond} \in \mathbb{R}^{1\times C}$ and prepending it to the flattened spatial token sequence $T_{spatial} \in \mathbb{R}^{HW\times C}$. Through multi-head self-attention, the aberration token exploits the global interaction capability of attention to modulate spatial features, generating physics-guided attention maps that assign adaptive weights to different regions. After nonlinear transformation via a feed-forward network (FFN), the aberration-modulated spatial features $T'_{spatial}$ are retained for centroid regression. This design preserves spatial topology while suppressing aberration-induced distortions that interfere with accurate centroid estimation.

2.4. Model Training and Optimization Strategy

After feature fusion, the updated spatial features are flattened and passed through a three-layer fully connected network, producing the predicted centroid bias correction $\Delta x_{pred}, \Delta y_{pred}$. Mean squared error (MSE) is used as the loss function:

$$\mathcal{L} = \frac{1}{N}\sum_{k=1}^{N}\left((\Delta x_{\text{pred}}(k) - \Delta x(k))^2 + (\Delta y_{\text{pred}}(k) - \Delta y(k))^2\right) \tag{5}$$

Rather than predicting absolute centroid coordinates directly, the network estimates the systematic bias $(\Delta x_{\text{pred}}, \Delta y_{\text{pred}})$ of the coarse CoG centroid $(x_{\text{raw}}, y_{\text{raw}})$ relative to the true centroid:

$$\begin{cases} x_{\text{final}} = x_{\text{raw}} - \Delta x_{\text{pred}} \\ y_{\text{final}} = y_{\text{raw}} - \Delta y_{\text{pred}} \end{cases} \tag{6}$$

The final centroid position $(x_{\text{final}}, y_{\text{final}})$ is obtained by correcting the CoG estimate using the predicted bias $(\Delta x_{\text{pred}}, \Delta y_{\text{pred}})$. This strategy reduces learning complexity and allows the network to focus on modeling aberration-induced nonlinear errors.

The model is implemented in PyTorch v2.3.1. This framework accommodates the custom tensor dimension manipulations required for concatenating physical aberration priors via its dynamic computation graph, while providing native support for Transformer-based architectures. The synthetic spot samples are randomly partitioned into mutually exclusive sets to preserve the statistical consistency of Seidel aberrations and signal-to-noise ratios without data leakage.

To prevent the attention mechanism from overfitting to precise physical priors, dynamic Gaussian noise is superimposed onto the input Seidel condition tokens $\mathbf{c}$ during training. The network thereby utilizes these physical parameters as soft contextual guidance rather than rigid deterministic rules, enhancing generalization against empirical prior retrieval errors.

The dataset partitioning, architectural specifications, and training hyperparameters are summarized in Table 1. The network is optimized using the AdamW algorithm alongside a cosine annealing schedule to facilitate early parameter space exploration and stable conver-

gence. Because attention mechanisms often require extended optimization trajectories to spatially stabilize, an early stopping criterion monitored by validation root-mean-square error is applied to mitigate overfitting.

Table 1. Dataset partitioning, neural network architecture, and training hyperparameters.

Parameter/Component	Specification/Value
Dataset Configuration	
Total Synthetic Samples	50,000
Data Split (Train:Val:Test)	35,000:7500:7500 (7:1.5:1.5)
Prior Condition Noise	5% variance (Dynamic Gaussian)
Architecture Details	
Input Tensor Dimension	$1 \times 32 \times 32$ (Grayscale Intensity)
Aberration Prior Dimension	1×5 (Seidel Coefficients)
Transformer Attention	4 Heads, Feedforward Dim: 256
Dropout Rates	0.25 (Conv Layers), 0.5 (FC Layers)
Output Dimension	1×2 ($\Delta x, \Delta y$)
Training Hyperparameters	
Loss Function	Mean Squared Error (MSE)
Optimizer	AdamW ($\beta_1 = 0.9, \beta_2 = 0.999$)
Initial Learning Rate	$\eta_0 = 1 \times 10^{-4}$ (Cosine Annealing)
Weight Decay	1×10^{-4}
Batch Size	128
Maximum Epochs	5000
Early Stopping Criterion	Patience = 200 epochs (Validation RMSE)
Hardware and Framework	NVIDIA GPU, PyTorch v2.3.1

While the hybrid CNN–Transformer architecture introduces higher computational complexity than traditional statistical estimators, its execution is restricted to localized bounding boxes. A quantitative evaluation of the computational cost and deployment feasibility on modern SWaP-constrained hardware is provided in Section 5.

3. Forward Physical Model Evaluation and Decoupling Analysis

In practical SWaP-constrained engineering environments, acquiring absolute ground-truth coordinates is impractical. Therefore, this section evaluates the proposed inverse solver utilizing the forward physical model grounded in scalar diffraction theory. By leveraging the exact ground truth generated by the forward physical model, the framework's capability to explicitly decouple aberration-induced asymmetric errors from random noise is rigorously examined. Furthermore, theoretical performance limits under varying SNRs and aberrations are analyzed to establish baseline benchmarks for the subsequent physical multi-beam experiments.

3.1. Statistical Distribution Characteristics of Centroid Bias and CDF Evaluation

To characterize error distributions, a simulated dataset covering representative aberrations (e.g., coma, astigmatism) and varying SNRs is constructed. Figure 3 presents the statistical centroid localization errors obtained via different methods under simulation conditions.

The selected comparison methods encompass the conventional Center of Gravity (CoG) approach, its noise-suppression variants (e.g., Thresholding CoG, Adaptive Thresholding CoG, and Intensity Weighted CoG), and Gaussian fitting (GFit). These methods represent mainstream paradigms ranging from statistical moment estimation to parametric model fitting. They are included to examine the applicability limits of traditional physical assumptions under asymmetric spot degradation. Consequently, we restrict the comparison to these representative

assumption-based estimators to elucidate the specific benefit of introducing explicit aberration priors, rather than to optimize alternative deep learning backbones.

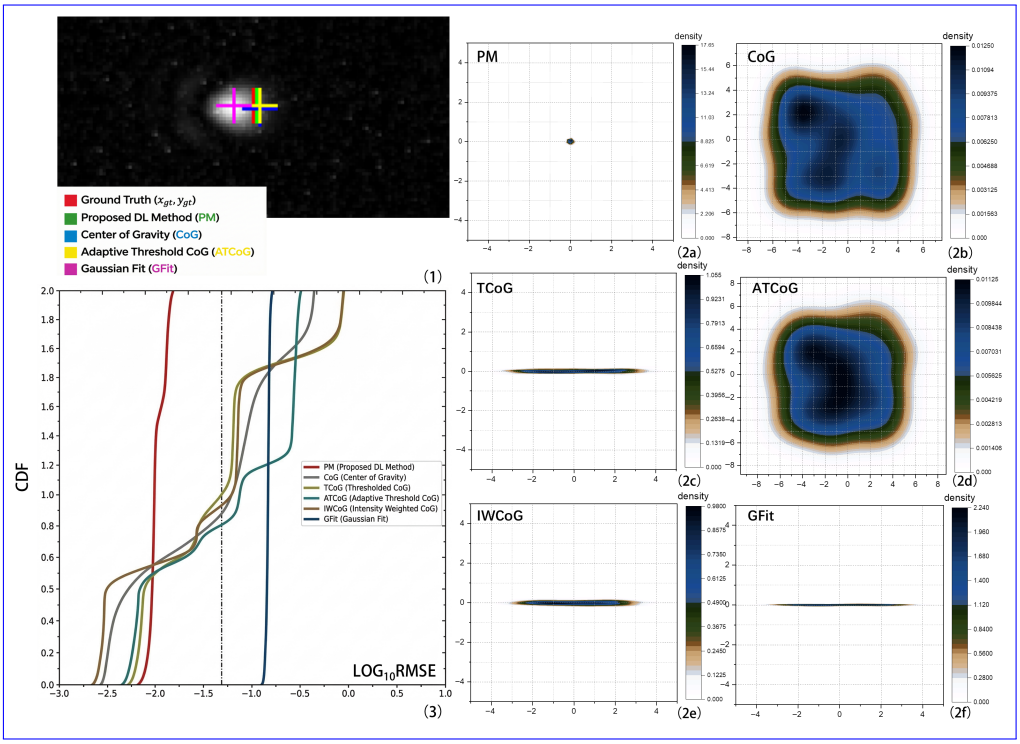

Figure 3. Statistical analysis of localization error characteristics under simulation conditions. (**1**) Visual comparison of centroid localization for typical coma-aberrated spots, where different crosshairs indicate centroid positions estimated by different algorithms; (**2**) two-dimensional kernel density estimation (KDE) of centroid bias, illustrating spatial error distributions along the x- and y-directions for different algorithms: (**2a**) the proposed method (PM), (**2b**) Center of Gravity (CoG), (**2c**) Thresholding CoG (TCoG), (**2d**) Adaptive Thresholding CoG (ATCoG), (**2e**) Intensity Weighted CoG (IWCoG), and (**2f**) Gaussian Fitting (GFit); (**3**) cumulative distribution function (CDF) curves of localization error (RMSE) for different methods evaluated under 18 composite aberration conditions and varying noise levels, with the dash-dotted line indicating the high-precision threshold of RMSE = 0.05 pixels.

Single-frame results in Figure 3(1) reveal distinct responses to aberrations. Conventional moment-based methods exhibit pronounced centroid drift due to energy redistribution along coma-induced tails, while Gaussian fitting fails due to model shape mismatch. In contrast, the proposed method suppresses asymmetric trailing effects via aberration-conditioned attention, yielding estimates closer to the ground truth.

To rigorously analyze the continuous spatial distribution of these localization errors, a two-dimensional Kernel Density Estimation (KDE) [16] is employed. For a given set of N centroid bias predictions $(\Delta x_i, \Delta y_i)$, the estimated probability density function $\hat{f}(\Delta x, \Delta y)$ is defined as:

$$\hat{f}(\Delta x, \Delta y) = \frac{1}{N h_x h_y} \sum_{i=1}^{N} K\left(\frac{\Delta x - \Delta x_i}{h_x}, \frac{\Delta y - \Delta y_i}{h_y} \right) \tag{7}$$

where $K(\cdot, \cdot)$ denotes the two-dimensional standard Gaussian kernel function, and h_x and h_y represent the smoothing bandwidths along the respective axes.

As confirmed by the evaluated KDE distributions in Figure 3(2): baseline error distributions (Figure 3(2b–2f)) show anisotropic, elliptical patterns aligned with aberration principal axes, indicating persistent systematic bias. Conversely, the proposed method (Figure 3(2a)) produces a compact, near-Gaussian distribution centered at zero, confirming

the effective suppression of aberration-induced systematic bias such that residual errors are predominantly stochastic.

Figure 3(3) quantitatively summarizes these advantages evaluated under the 18 composite aberration conditions and varying noise levels. Under a stringent RMSE threshold of 0.05 pixels, the proposed method reaches saturation confidence earlier than the baselines, maintaining localization accuracy within 0.011–0.021 pixels across the evaluated parameter space.

To strictly quantify the physical prior's contribution and evaluate the framework's reliability under deployment-relevant prior uncertainty conditions, where physical priors derived from thermal-structural look-up tables inherently contain measurement uncertainties, a comprehensive sensitivity and ablation analysis is conducted. During the inference phase, controlled Gaussian perturbations are injected into the ground-truth Seidel coefficients to simulate prior estimation errors:

$$\hat{S}_i = S_i \cdot (1 + \gamma \cdot \epsilon) \tag{8}$$

where $\epsilon \sim \mathcal{N}(0, 1)$ is a standard normal random variable, and γ represents the targeted uncertainty level. Results indicate that while the localization RMSE degrades from 0.021 pixels to 0.032 pixels under a 10% prior error ($\gamma = 0.10$), the performance remains lower in RMSE than conventional assumption-based baselines under the same perturbation conditions, indicating that the network tolerates moderate prior retrieval errors.

Furthermore, an uninformed ablation condition was tested where the input Seidel coefficients were forced to zero, simulating a complete failure of the prior estimation module. Under this condition, the network effectively degenerated to pure CNN performance, with the RMSE degrading to 0.152 pixels. This performance trend under varying prior uncertainties demonstrates the contribution of the aberration-conditioned attention mechanism in achieving sub-pixel accuracy.

3.2. Validation of Geometric Invariance for Multi-Beam Engineering Scenarios

To provide a theoretical basis for the physical multi-beam experiments in Section 4, and to verify the robustness of the proposed computational framework under dynamic perturbations, a double-spot simulation scenario isomorphic to the real wedge-based LiDAR receiving system is constructed. In practical engineering applications, maintaining the invariant relative spacing between multiple targets under local field-variant aberrations is a fundamental requirement. Accordingly, a constant physical spacing of 49 pixels is imposed in the image-plane coordinate system, and synchronous random displacements are applied to simulate pointing drift and micro-vibration effects encountered in practice. Representative forward modeling results are shown in Figure 4(1).

Figure 4(2) illustrates the spacing stability over 200 simulated frames. Although the true physical spacing remains constant at 49 pixels, conventional methods exhibit measurable oscillations; for instance, the root-mean-square error (RMSE) for CoG and GFit is 0.3361 pixels and 0.1284 pixels, respectively. In contrast, the proposed method (PM) maintains improved temporal stability with an RMSE of 0.0562 pixels, indicating robustness to local aberration perturbations. The boxplot results in Figure 4(3) further characterize this data dispersion: the proposed method presents a standard deviation of 0.0213 pixels, which is lower than those of the baseline methods (ranging from 0.0754 pixels for GFit to 0.1251 pixels for CoG). Correspondingly, the interquartile range (IQR) associated with the proposed method is narrower and shows no outliers, demonstrating compensation for local aberration variations induced by field-of-view shifts.

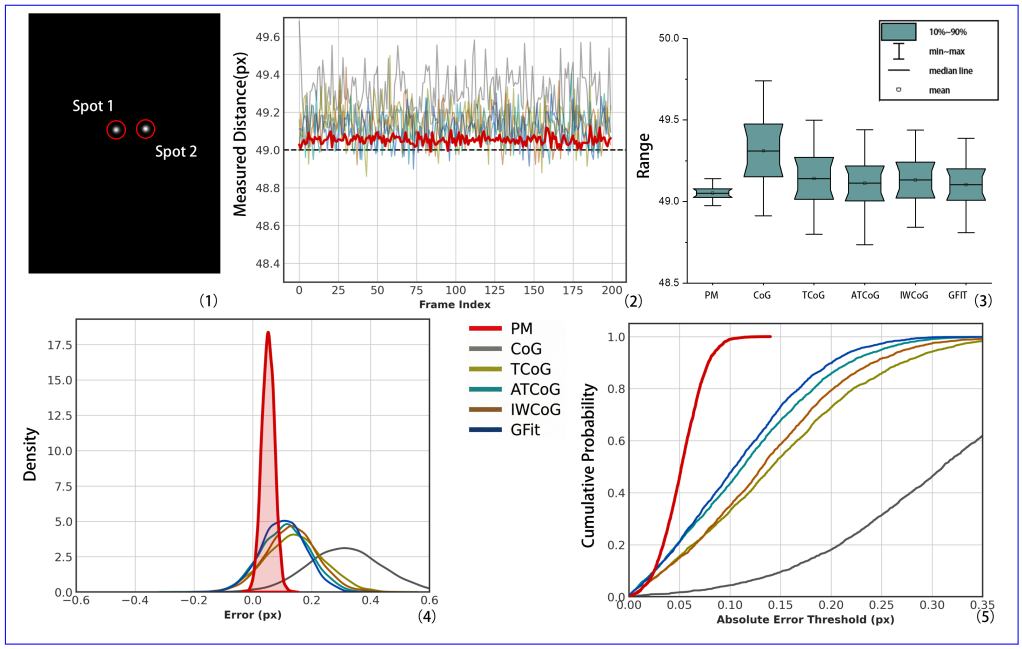

Figure 4. Consistency verification of simulated double-spot spacing: (**1**) simulated double-spot images; (**2**) spacing measurement fluctuations over a 200-frame sequence; (**3**) boxplot statistics of the measured spacing; (**4**,**5**) error probability density distributions and corresponding CDF curves.

The error histogram in Figure 4(4) shows that the proposed method reduces systematic jitter, yielding a compact distribution centered closerto zero. This behavior suggests that single-spot localization accuracy is maintained in the relative spacing measurement. As shown by the cumulative distribution function (CDF) curves in Figure 4(5), the proposed method reaches a 95% confidence level within an error threshold of 0.10 pixels. The CDF curve for the proposed method achieves 100% saturation before the 0.15-pixel threshold. This visual representation demonstrates a restricted error bound and the absence of outliers, compared to the broader distributions (extending to 0.35 pixels) exhibited by conventional estimators. These results indicate that the statistical distribution of double-spot spacing serves as an indirect yet physically meaningful surrogate for evaluating aberration-induced localization bias, thereby justifying its use in real-system experiments where absolute ground truth is unavailable.

4. Double-Spot Centroid Localization Experimental Design and Evaluation Metrics

To assess the generalization capability of the proposed computational framework when transferred from simulation to real physical systems, an evaluation strategy tailored for multi-beam engineering scenarios based on physical geometric invariants is adopted. The stability with which the framework reproduces these invariants serves as an indicator of robustness under complex operating conditions. Unlike the forward simulation scenarios, the spacing constraint in the experiment is strictly determined by device geometry and remains fixed, providing a more rigorous physical benchmark.

4.1. Experimental Design

A wedge-based double-spot LiDAR receiving platform is constructed to validate the proposed bias correction framework under real optical aberrations. As shown in Figure 5, the platform replicates the geometric characteristics of practical engineering setups, wherein an incident laser beam reflects off the front and rear surfaces of an optical wedge to produce two distinct spots on the detector plane.

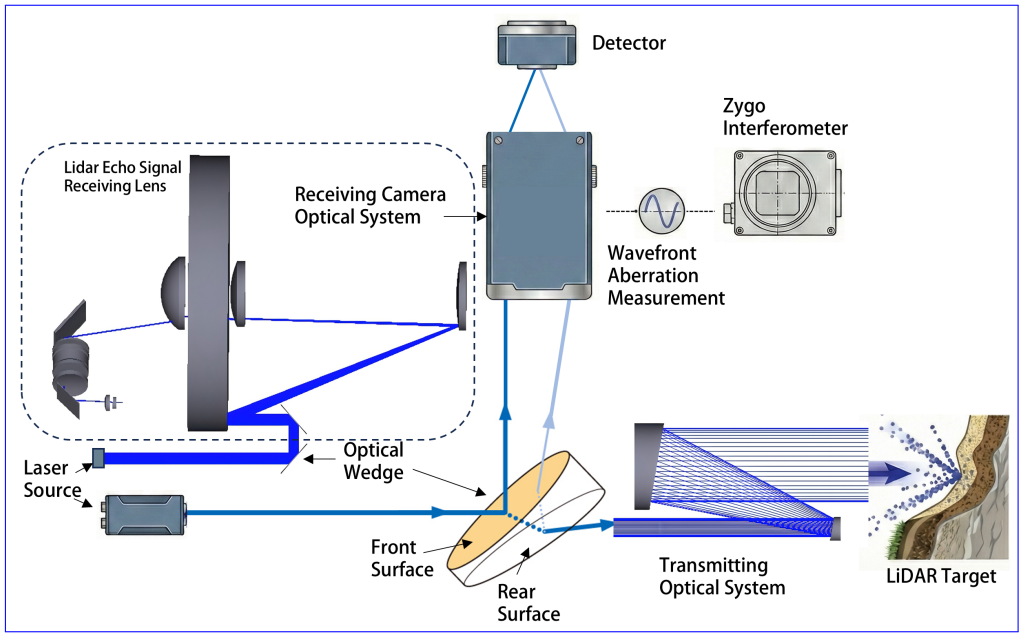

Figure 5. Schematic of the wedge-based double-spot experimental setup. The left panel illustrates the complete optical path. The core validation path corresponds to the wedge-reflected light entering the receiving optical system, where a Zygo (AMETEK, Inc., Berwyn, PA, USA) interferometer provides high-precision wavefront calibration. The transmitted path merely maintains normal system operation.

The system operates with a 1064.4 nm laser (20 mm diameter, 200 µrad divergence). The optical wedge, fabricated from fused silica ($n = 1.458464$), features calibrated angles of 35.84 µrad and 83.94 µrad to provide two spacing configurations. The receiving system utilizes a 600 mm clear aperture and a 2578 mm equivalent focal length, focusing the spots onto a 2048 × 2048 pixel array (5.5 µm pixel pitch) operating at 9 Hz.

Theoretically, the relative spacing between the spots is invariant, dictated by the wedge geometry. However, the autocollimation calibration uncertainty of the wedge angles (0.24 µrad) introduces a static focal-plane uncertainty of approximately 0.33 pixels. Because this baseline physical uncertainty exceeds the evaluated sub-pixel scale, computing an absolute Root Mean Square Error (RMSE) is metrologically impractical. Consequently, the standard deviation (SD) of the spacing is utilized to evaluate algorithmic robustness against system noise. Furthermore, this geometric invariant serves as a differential check: since the two spots correspond to distinct field angles and local Point Spread Functions (PSFs), maintaining consistent spacing provides strong evidence that the algorithm compensates for position-specific aberrations rather than applying a trivial global bias shift.

To ensure reliable physical priors, wavefronts are measured quasi-synchronously using a Zygo interferometer under strictly controlled conditions (vibration isolation, ±0.1 °C). The measured Seidel coefficients at the testing wavelength ($\lambda_{test} = 1064.4$ nm) are normalized to construct the dimensionless condition vector $\mathbf{c}$. The instrument exhibits an RMS repeatability better than $\lambda/100$ (relative error <1%), which is strictly bounded within the network's 10% prior perturbation tolerance (Section 3.1). This static setup establishes a physical benchmark for aberration decoupling; dynamic deployment considerations are discussed in Section 5.

During data acquisition, two continuous measurement sequences of 1738 and 1797 frames were collected (3 min each). The captured spots exhibit strict localized energy distributions without pixel saturation or motion-induced smearing, ensuring high-fidelity preservation of the static optical aberrations and a reliable signal-to-noise ratio for subsequent centroid evaluation.

4.2. Experimental Result Analysis

Using the experimental configuration described above, double-spot image sequences are acquired under real operating conditions to evaluate the performance of the proposed framework. By comparing the proposed approach with conventional centroid localization algorithms, the stability and consistency of the framework during transfer from the forward physical simulation to experiment are assessed. Wavefront calibration results and the corresponding aberration coefficients are shown in Figure 6(1), while representative raw double-spot images are shown in Figure 6(2).

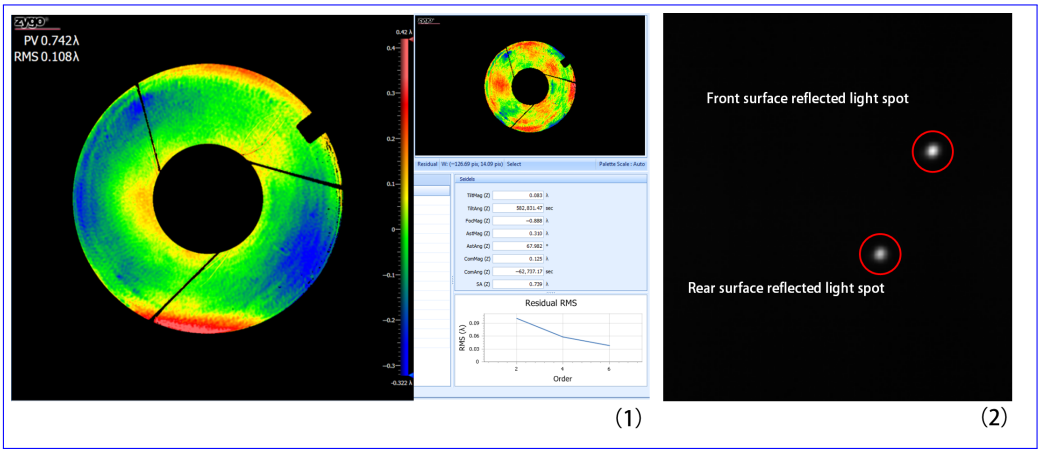

Figure 6. Experimental conditions and raw data acquisition. (**1**) High-precision wavefront measurements and corresponding Seidel aberration coefficients of the receiving optical system, calibrated in situ using a Zygo interferometer; (**2**) representative raw double-spot images formed on the detector by reflections from the front and rear surfaces of the optical wedge.

The temporal traces in Figure 7(1) clearly show that under spacing conditions of approximately 49 pixels and 114 pixels, conventional methods exhibit pronounced low-frequency drift and high-frequency noise, indicating strong sensitivity to environmental perturbations and asymmetric spot deformation. In contrast, the proposed framework maintains high temporal stability across the entire acquisition period, effectively suppressing the influence of local distortion and background noise.

This stability is further reflected in the statistical distributions shown in Figure 7(2,3). While the interquartile ranges of conventional methods vary significantly with field-of-view changes, the proposed framework consistently produces compact boxplots with stable medians and no significant outliers. Notably, these experimental statistics are broadly consistent with the simulated trends shown in Figure 4(2), providing supportive evidence for geometric-stability preservation and transferability from physics-based simulation to real optical measurements.

Quantitatively, the measured double-spot spacing standard deviation (SD) for the proposed framework is confined to 0.015–0.039 pixels, which is of the same sub-pixel order as the simulated double-spot SD of 0.0213 pixels. In contrast, the conventional methods evaluated in the control experiments exhibit a substantially higher spacing SD of approximately 0.1 pixels. This numerical correspondence supports the consistency of the proposed framework across physics-based simulation and controlled optical experiments, indicating that the physically invariant spacing can be preserved under local field-variant aberrations.

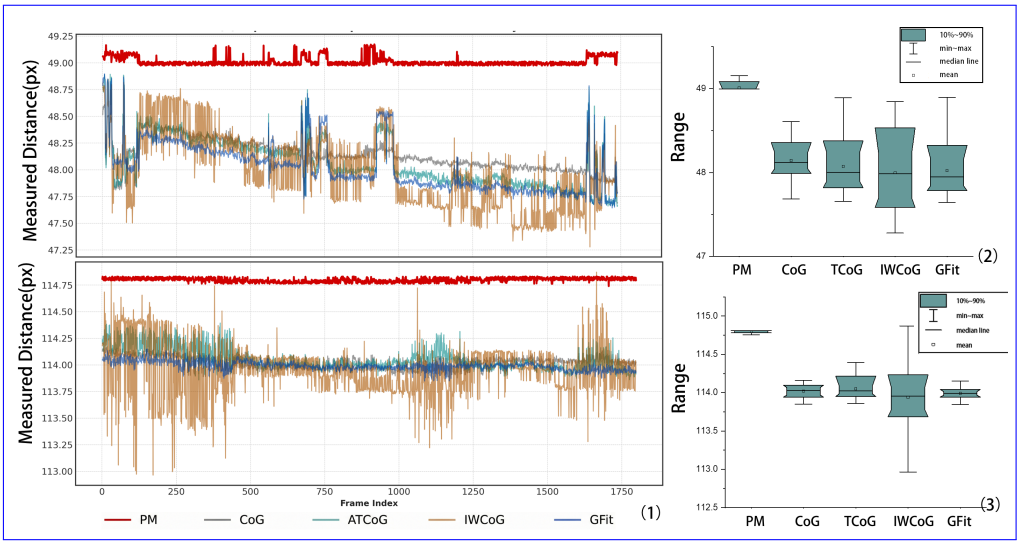

Figure 7. Quantitative performance evaluation of double-spot spacing consistency: (**1**) temporal fluctuations of the measured spacing over image sequences for different algorithms, evaluated under a short-spacing condition (top panel, ∼49 px) and a long-spacing condition (bottom panel, ∼114 px); (**2**) boxplot statistics corresponding to the short-spacing configuration, illustrating the interquartile range and median bias; (**3**) boxplot statistics corresponding to the long-spacing configuration. The proposed method (PM) consistently exhibits compact distribution and minimal drift across both footprint configurations.

5. Discussion

To contextualize the proposed framework, it is insightful to examine recent advancements in sub-pixel spot centroid localization. It should be noted that since detector formats, signal conditions, and evaluation metrics vary across different studies, the following performance figures serve as contextual benchmarks rather than strict one-to-one comparisons. While conventional estimators degrade under asymmetric optical distortions, recent data-driven approaches seek to map non-ideal morphologies directly to centroid coordinates. For instance, super-resolution networks deployed on 32 × 32 focal-plane arrays have achieved estimation root-mean-square errors (RMSE) of 0.04 to 0.20 pixels [11]. Similarly, multi-scale adaptive convolutions bound localization errors to approximately 0.15 pixels under high noise [12], and key-point detection models restrict dynamic centroid deviations to within 0.1 pixels [13]. By explicitly concatenating Seidel physical priors into the feature space via a conditional attention mechanism, our framework achieves a theoretical simulation RMSE of 0.011 to 0.021 pixels and an experimental double-spot relative spacing stability of 0.015 to 0.039 pixels. This sub-pixel accuracy indicates that incorporating structural physical constraints provides a highly effective alternative to purely spatial mapping strategies for suppressing systematic bias.

The precision of this physics-conditioned approach relies on the appropriate parameterization of optical degradations. For the specific low-order aberration regime considered here, Seidel polynomials are explicitly selected over Zernike polynomials to maintain a direct physical correspondence to primary structural aberrations. While Zernike modes are highly advantageous for complete orthogonal wavefront representation, they achieve mathematical orthogonality through radial balancing. Consequently, representing a specific primary aberration with Zernike modes distributes the physical prior across multiple coupled coefficients, complicating the feature space. Utilizing Seidel polynomials avoids this mathematical mixing, providing a more task-aligned conditioning space that enables the attention mechanism to map degraded spot morphology directly to its independent physical origins.

Beyond parameterization, transitioning this framework to real-world dynamic environments requires careful engineering considerations. Our experimental validation utilized a static benchtop platform. The primary objective of this setup was to minimize complex dynamic disturbances, thereby providing a stable baseline to verify the algorithmic decoupling of field-variant optical aberrations. In practical SWaP-constrained deployments, systems often experience combined dynamic perturbations. To bridge this static validation with real-world applications, future systems can operate on a decoupled temporal basis: low-frequency, quasi-static primary aberrations can be retrieved from pre-calibrated look-up tables to serve as physical priors, while the neural network independently accommodates high-frequency dynamic positional shifts during real-time inference. This deployment hypothesis is physically motivated but remains to be validated under controlled dynamic perturbation experiments.

Finally, achieving this sub-pixel precision must not violate SWaP hardware constraints. Evaluated via standard profiling tools, the proposed model contains 2.27 million parameters and requires 0.115 GFLOPs per forward pass. Modern space-grade edge processors routinely provide compute capacities of several TOPS [17,18]. Assuming a conservative baseline of 5 TOPS and applying an order-of-magnitude penalty for memory bandwidth and data transfer overheads, the effective single-spot inference latency is projected between 0.1 ms and 0.5 ms. This yields a sustained throughput of 2 kHz to 10 kHz. These estimates suggest preliminary deployment feasibility on modern edge hardware, pending dedicated profiling on the target embedded platform.

6. Conclusions and Outlook

This paper addresses systematic centroid bias arising from the nonlinear coupling between environmentally driven optical aberrations and noise in SWaP-constrained precision optical measurement systems. To overcome the physical limitations of integrating dedicated wavefront sensors, a software-defined, physics-conditioned computational measurement framework is proposed. At its core, a CNN–Transformer hybrid architecture acts as an inverse solver, in which aberration coefficients are explicitly incorporated as physical conditions within the feature space. This design enables physics-aware feature modulation during extraction and overcomes the accuracy limitations of conventional centroid estimators under complex, asymmetric imaging conditions.

A complete closed-loop workflow, transitioning from forward physical modeling to empirical validation in multi-beam engineering scenarios, is established. Simulation studies based on scalar diffraction theory show that the incorporation of physical priors effectively suppresses anisotropic systematic errors and reshapes the residual error distribution toward an approximately zero-centered stochastic form, achieving a localization RMSE of 0.011 to 0.021 pixels. Furthermore, double-spot LiDAR receiving experiments support the engineering robustness of this decoupling capability under controlled laboratory conditions. By exploiting strict geometric rigidity constraints, the framework yields a measured spacing standard deviation (SD) of 0.015 to 0.039 pixels under empirical conditions without absolute ground truth.

To advance towards practical engineering deployment, future work will extend this framework in two primary directions. First, to address higher-order environmental degradations such as atmospheric turbulence, the mathematical completeness of Zernike polynomials will be systematically evaluated as an alternative prior. Second, while the current experiment establishes an algorithmic upper bound using high-precision interferometric priors, actual hardware-free deployment relies on pre-calibrated thermal-structural look-up tables. Therefore, evaluating and mitigating prior estimation errors in dynamic environments is critical. Coupled with the preliminarily assessed deployment feasibil-

ity on edge-computing devices, these combined efforts will enable efficient, autonomous correction of complex optical degradations on resource-constrained platforms.

Author Contributions: Conceptualization, Z.Z. and J.H.; Methodology, Z.Z.; Software, Z.Z.; Validation, Z.Z.; Formal analysis, Z.Z.; Investigation, J.H. and Y.L.; Resources, J.H., Y.L. and A.L.; Data curation, Z.Z. and A.L.; Writing—original draft preparation, Z.Z.; Writing—review and editing, Z.Z., J.H. and Z.H.; Visualization, Z.Z.; Supervision, Z.H.; Project administration, J.H. and Z.H.; Funding acquisition, Z.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Innovation Program for Quantum Science and Technology, grant number 2021ZD0300304; the Shanghai Municipal Science and Technology Major Project, grant number 2019SHZDZX01; and the National Science Fund for Distinguished Young Scholars, grant number 62125505. The APC was funded by the authors' institution.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the technical staff of the Shanghai Institute of Technical Physics for their support in wavefront measurement and experimental platform maintenance. The authors also thank Xuehan Wang for valuable technical discussions related to this work.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
RMSE	Root Mean Square Error
KDE	Kernel Density Estimation
CDF	Cumulative Distribution Function
PSF	Point Spread Function
CoG	Center of Gravity
SNR	Signal-to-Noise Ratio
SWaP	size, weight, and power
LiDAR	Light Detection and Ranging
SD	Standard Deviation
S-H	Shack-Hartmann

References

1. Xia, A.L.; Ma, C.W. An improved centroid detection method based on higher moment for Shack-Hartmann wavefront sensor. In *Proceedings of the Optoelectronic Imaging and Multimedia Technology*; SPIE: Bellingham, WA, USA, 2010; Volume 7850, pp. 356–361.
2. Kong, F.; Cegarra Polo, M.; Lambert, A. Fpga implementation of shack–Hartmann wavefront sensing using stream-based center of gravity method for centroid estimation. *Electronics* **2023**, *12*, 1714. [CrossRef]
3. Yang, W.; Wang, J.; Wang, B. A method used to improve the dynamic range of Shack–Hartmann wavefront sensor in presence of large aberration. *Sensors* **2022**, *22*, 7120. [CrossRef] [PubMed]
4. Li, X.; Wang, A.; Fan, M.; Yu, L.; Liang, X. Multi-Spectral and Single-Shot Wavefront Detection Technique Based on Neural Networks. *Photonics* **2025**, *12*, 1110. [CrossRef]
5. Hardy, J.W. *Adaptive Optics for Astronomical Telescopes*; Oxford University Press: Oxford, UK, 1998.
6. McManamon, P.F. *LiDAR Technologies and Systems*; SPIE: Bellingham, WA, USA, 2019.
7. Southwell, W.H. Wave-front estimation from wave-front slope measurements. *J. Opt. Soc. Am.* **1980**, *70*, 998–1006. [CrossRef]
8. Hu, L.; Hu, S.; Gong, W.; Si, K. Learning-based Shack-Hartmann wavefront sensor for high-order aberration detection. *Opt. Express* **2019**, *27*, 33504–33517. [CrossRef] [PubMed]
9. Liu, X.; Hu, P.; Luo, W.; Zhang, J.; Zhang, F.; Su, H. Physics-informed deep learning for accurate and efficient wavefront sensing in adaptive optics. *Opt. Commun.* **2025**, *596*, 132458. [CrossRef]
10. Bao, J.; Zhan, H.; Sun, T.; Fu, S.; Xing, F.; You, Z. A window-adaptive centroiding method based on energy iteration for spot target localization. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–13. [CrossRef]

11. Liu, X.; Guo, Y.; Huang, J.; Zhang, L.; Shu, R. Deep-learning-based beam position estimation with a photon-counting camera in free-space optical communications. *Opt. Express* **2025**, *33*, 48740–48749. [CrossRef] [PubMed]
12. Yuan, K.; Li, L. Optimization of laser spot edge extraction and localization based on multi-scale adaptive convolution. *Front. Phys.* **2025**, *13*, 1650714. [CrossRef]
13. Luo, Z.; Guo, Q.; Feng, J.; Li, Y.; Zhang, X. Centroid algorithm for high-dynamic star sensor based on key point detection deep learning algorithms. *Opt. Express* **2025**, *33*, 17203–17232. [CrossRef] [PubMed]
14. Banerjee, C.; Nguyen, K.; Salvado, O.; Tran, T.; Fookes, C. Physics-informed Machine Learning for Medical Image Analysis. *ACM Comput. Surv.* **2025**, *58*, 1–35. [CrossRef]
15. Chai, T.; Liu, X.; Wang, H.; Jin, Y.; Huang, J.; Shi, T.; Jiang, Y. Physics-informed deep learning framework for wavefront sensing via optical beam pattern analysis. *J. Opt. Soc. Am. A* **2026**, *43*, 346–353. [CrossRef] [PubMed]
16. Scott, D. *Multivariate Density Estimation: Theory, Practice, and Visualization*; A Wiley-Interscience Publication; Wiley: Hoboken, NJ, USA, 1992.
17. Friedrich, S.; Wittig, R.; Matus, E.; Grantz, D.; Zeller, M.; Benndorf, J.; Fettweis, G. A 22 nm 10 TOPS Mixed-Precision Neural Network SoC for Image Processing with Energy-Efficient Dilated Convolution Support. In Proceedings of the 2024 IEEE Symposium in Low-Power and High-Speed Chips (COOL CHIPS), Tokyo, Japan, 17–19 April 2024; pp. 1–3. [CrossRef]
18. Renaut, L.; Frei, H.; Nüchter, A. CNN-Based Pose Estimation of a Noncooperative Spacecraft with Symmetries From LiDAR Point Clouds. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *61*, 5002–5016. [CrossRef]