Supplementary Information for Analysis and Division of Full-Band Errors in Optical Surfaces

Running Title: Full-Band Error Analysis and Division

Author:

Yuan Liu[a,b,c], Xiaokun Wang[a,b,c,*], Zhongkai Liu[a,b,c,*], Yukun Wang[a,b,c], Mengxue Cai[a,b,c], Donglin Xue[a,b,c] and Xuejun Zhang[a,b,c]

[a] *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 3888 Dongnanhu Road, Changchun, Jilin 130033, China*

[b] *University of Chinese Academy of Sciences, Beijing 100049, China*

[c] *State Key Laboratory of Advanced Manufacturing for Optical Systems, 3888 Dongnanhu Road, Changchun, Jilin 130033, China*

[*] *jimwxk@sohu.com*

S1. Classification and Impact of Full-Band Errors on Optical Surfaces

Considering that some readers may not be experts in the field of optical testing, we hope that Section S1 will help readers establish the basic concepts related to frequency band errors as quickly and clearly as possible, thereby better understanding the content of this work.

Due to certain inherent surface defects in optical surfaces, surface manufacturing residuals introduced during the actual manufacturing process, and minor variations introduced by the testing environment, there are varying degrees of deviation between the physical distribution of the surface sag and the theoretical surface sag. These deviations manifest spatially as different spatial frequencies. According to the different distributions of spatial frequencies, the optical surface error can be divided into different frequency band errors—low-frequency (low-spatial frequency, LSF) error, mid-frequency (mid-spatial frequency, MSF) error, and high-frequency (high-spatial frequency, HSF) error, shows in **Fig. S1**. These three are collectively referred to as the optical surface frequency band error, namely the full-band error.

Low-frequency error typically refers to errors with longer spatial period lengths, generally corresponding to the surface error of the optical component, i.e., traditional aberrations, which can be described by Zernike aberrations or Seidel aberrations. It represents the deviation between the surface morphology of the optical component and its theoretical design morphology, and directly affects the imaging quality, wavefront aberration, and system stray light of the optical system; mid-frequency error typically refers to errors with medium spatial period lengths, generally

corresponding to waviness errors such as tool marks introduced during the manufacturing process. Its impact on the optical system is mainly manifested in small-angle scattering, which may generate flare and affect system contrast; high-frequency error typically refers to errors with shorter spatial period lengths, generally corresponding to microscopic surface features or roughness, such as scratches and pits. Its impact on the optical system is mainly manifested in large-angle scattering, affecting the clarity and sharpness of imaging.

To further clarify from the perspective of downstream optical performance, the LSF error primarily induces wavefront distortion and focal spot broadening. Conversely, the extracted MSF and HSF errors collectively constitute surface roughness in a broad sense, dominating the light scattering characteristics of the surface. Specifically, based on the spectral division logic of this framework, two special cases require distinct physical definition: first, randomly scattered surface defects (e.g., scratches and pits) are categorized as HSF features because they induce broadband, large-angle scattering with random orientations; second, highly regular microstructures with minute periods (e.g., computer-generated holograms (CGHs) and gratings) are classified as MSF features because they are inherently introduced by periodic fabrication processes and generate directional, small-angle diffraction or scattering.

Therefore, when evaluating the quality of optical surfaces or analyzing the performance of optical systems, obtaining an accurate frequency band error division standard is highly important. This helps guide the accurate acquisition of the full-band error of optical surfaces, enabling subsequent separate analyses of errors in different frequency bands to investigate their respective impacts on the optical system, while also facilitating targeted reduction in subsequent manufacturing processes.

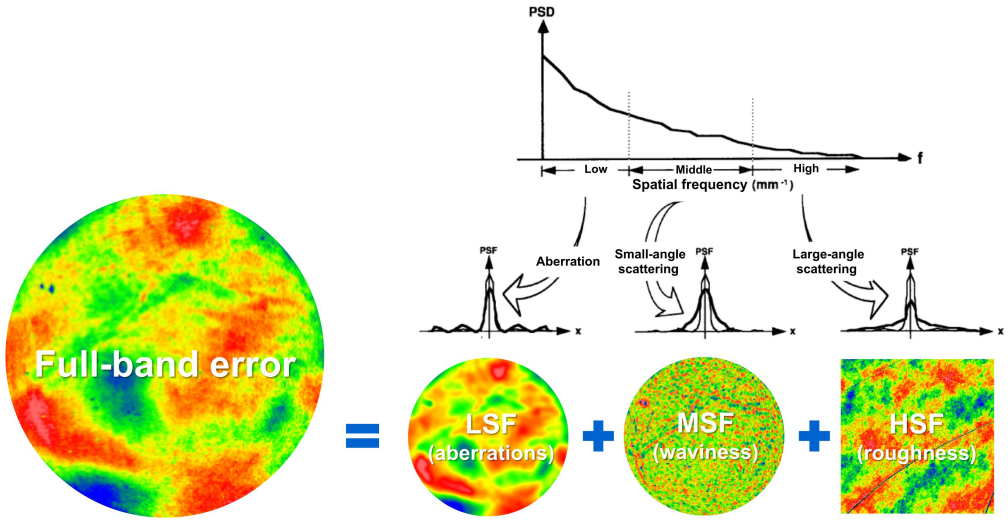

Fig. S1 Classification of full-band errors on optical surfaces.

S2. Architecture and Hyperparameters of the Low-Frequency DPN

The specific architecture of the low-frequency part follows our previous work[1]. For the readers' convenience, the detailed network architecture and hyperparameters of the DPN low-frequency branch are listed here.

Network Architecture: The DPN is a deep learning architecture evolved from the traditional convolution-pooling paradigm, specifically designed to extract the features of specific frequency band errors. The entire network adopts an encoder-bottleneck-decoder structure. The encoder section of the network contains three stages of strided convolution (Conv2d), aimed at extracting global features and performing downsampling; during each downsampling process, the network cascades a Residual Block and a Channel Attention module to enhance the capability of capturing the target frequency band features, and the feature map dimensions are sequentially reduced from 16×256×256 to 32×128×128, ultimately reaching 64×64×64. Following this is a bottleneck layer with feature dimensions denoted as 64×64×64×3, which sequentially integrates a Spatial Attention module, a Residual Block, and a Channel Attention module. Here, the spatial attention mechanism can selectively focus on specific frequency band features by adjusting the convolution kernel size to achieve feature enhancement. In the decoder section, a sub-pixel upsampling layer is used to replace the traditional deconvolution layer, effectively suppressing the checkerboard artifacts introduced by traditional deconvolution. The decoder also contains three upsampling processes, and each module integrates a Residual Block and a Spatial Attention mechanism to ensure the detail and smoothness of the reconstructed image. The feature maps are progressively enlarged here, sequentially restored to 32×128×128, 16×256×256, and 5×512×512. Finally, the network adjusts the feature map size to 512×512 through a 2D convolutional layer, and outputs the final prediction image of the specific frequency band error through a Tanh activation function.

Training Details and Hardware Configuration: The DPN model was trained using the PyTorch framework. Key training parameters are consistent with those described in the main manuscript, including the use of an AdamW optimizer with an initial learning rate of 1e-3, a cosine annealing schedule over 80 epochs, and a batch size of 64. To substantiate the claim that the model can be deployed on standard computers, all training and testing procedures reported in this paper were conducted on a single, commercially available laptop with the following specifications:

- **CPU:** Intel® Core™ Ultra 7 155H
- **RAM:** 32 GB

It is important to note that neither the integrated GPU (Intel® Arc™ Graphics) nor the NPU of the processor were utilized for the performance tests. All performance metrics reported, including the average extraction time of 0.04 seconds per surface, were achieved using only the CPU.

S3. Architecture, Parameter Selection, and Ablation Study of the Mid-Frequency DPN

S3.1. Network Architecture Design

This convolutional neural network consists of a Learnable Spectral Gating (LSG) module and an encoder-decoder backbone architecture. The input data is first transformed into the frequency domain via a 2D Real Fast Fourier Transform (RFFT). The separated real and imaginary parts are concatenated and fed into a lightweight network composed of two 3×3 convolutional layers and a Sigmoid activation function to generate a frequency domain mask for adaptive spectral filtering, which is subsequently restored to the spatial domain via an inverse transform (IRFFT). The filtered spatial domain features enter the encoder, which contains four cascaded feature extraction blocks (sequentially consisting of 3×3 convolution, Batch Normalization, and ReLU activation). The first three modules perform spatial downsampling using convolutions with a stride of 2, while the fourth module has a stride of 1. The number of feature map channels is sequentially expanded to 16, 32, 64, and 128. The decoder section consists of three symmetrical reconstruction blocks. It sequentially uses bilinear interpolation to upscale the spatial resolution by a factor of 2, and performs feature smoothing and dimensionality reduction through 3×3 convolution, Batch Normalization, and ReLU, progressively restoring the channel number to 64, 32, and 16. Finally, the network directly maps and outputs the prediction result through a single 3×3 convolutional layer.

The models in this ablation study were trained using the PyTorch framework. The dataset was randomly divided into training, validation, and test sets with proportions of 70%, 15%, and 15%, respectively. Model optimization utilized the Adam optimizer with a batch size of 32, an initial learning rate of 1e-3, and a weight decay of 1e-4 introduced to prevent overfitting. The learning rate decay adopted a Cosine Annealing strategy, and the total training epochs were set to 40 epochs to prevent overfitting on dense mid-frequency textures. The network's parameter updates were driven by a joint loss function, which specifically included: a spatial domain Mean Square Error (MSE) loss, a frequency domain amplitude MSE loss, a spatial gradient loss with a weight coefficient of 0.1, and a frequency domain mask distance regularization loss (Mask Weighted Loss) with a weight coefficient of 1e-5. In this ablation configuration, the convolution kernel size of the Learnable Spectral Gating module was specified as 3×3, and the bias of its output layer was initialized to 3.0. Furthermore, to substantiate the claim that the model can be deployed on a standard computer, all training and testing procedures reported herein were conducted on a commercially available laptop, with specific performance and settings identical to those in Section S2.

S3.2 Parameter Selection Experiments and Ablation Study

To determine the optimal configuration for the Learnable Spectral Gating (LSG) and the physics-aware deep learning (DPN) network, we conducted systematic parameter selection experiments and an ablation study (as shown in **Table S1**). The entire optimization process was divided into two stages: hyperparameter search and frequency domain receptive field exploration, aiming to ensure the optimal balance between network performance and physical interpretability.

Table S1. Parameter selection experiments and ablation study

Network Configuration	Average RMSE (nm)	Rel Acc (%)	SSIM	PSNR (dB)
net45 (bias=2, gamma=1e-4)	0.94	84.76	0.8839	35.23
net46 (bias=0.5, gamma=0.2)	0.35	84.50	0.8859	34.62
net47 (Alternative spectral weight calculation)	0.44	82.89	0.8713	33.85
net48 (bias=3, gamma=1e-5)	0.45	84.57	0.8839	35.64
net49 (bias=2.5, gamma=5e-5)	0.59	85.08	0.8869	35.58
net50 (bias=5, gamma=0)	0.61	82.93	0.8757	35.48
net51 (w/o gradient loss)	0.51	84.46	0.8869	35.77
net52 (w/o frequency domain loss)	0.40	84.48	0.8886	35.36
net53 (L1 loss replaces distance weight)	0.58	85.44	0.8990	36.21
net54 (3×3 kernel)	**0.37**	**86.86**	**0.9052**	**36.47**
net55 (5×5 kernel)	0.51	83.37	0.8795	34.95
net56 (3×3, w/o LSG)	0.88	86.15	0.9019	36.85
net57 (3×3, w/o gradient loss)	0.35	84.52	0.8881	35.53
net58 (3×3, w/o frequency domain loss)	0.56	84.97	0.8915	36.06
net59 (3×3, Instance Norm replaces Batch Norm)	0.46	81.62	0.8851	36.07
net60 (3×3, gradient weight = 1.0)	0.51	83.33	0.8786	36.08

net61 (3×3, gradient weight = 0.01)	0.73	83.24	0.8767	34.58
net62 (3×3, frequency domain weight = 0.1)	0.82	85.97	0.8998	35.88
net63 (3×3, learning rate = 1e-4)	0.67	80.58	0.8499	32.95
net64 (3×3, channel capacity ×2)	0.45	86.54	0.9035	36.50

In the initial stage of optimization (net45-net53), we conducted an exhaustive grid search on the initialization bias (bias), mask regularization weight (gamma), and various physical constraint losses of the LSG under the basic 1×1 frequency-domain convolution kernel architecture. Experiments indicate that the bias setting determines the pass rate of spectral information in the early stages of training. An excessively low bias (e.g., net46, bias=0.5) leads to an initially overly closed mask, thereby restricting the network's exploration of valid mid-frequency information; conversely, an excessively high bias (e.g., net50, bias=5) introduces excessive high-frequency noise. Combined with the spectral distance regularization term (gamma), net48 (bias=3.0, gamma=1e-5) exhibited the optimal balance across overall metrics. Furthermore, the ablation results for the loss functions (net51-net53) confirmed that removing either the gradient loss (net51) or the frequency domain amplitude loss (net52) leads to a decline in both relative accuracy (Rel Acc) and structural similarity (SSIM); meanwhile, using the L1 loss to replace the spectral distance weight (net53) fails to effectively penalize high-frequency redundant information. This fully demonstrates the necessity of mixed dual-domain physical constraints in driving the network to learn mid-frequency features.

After establishing the optimal physical constraint hyperparameters (bias=3.0, gamma=1e-5), we further explored the impact of the frequency-domain receptive field scale on the extraction of anisotropic mid-frequency textures (net54, net55). Traditional 1×1 convolution essentially degrades into an independent point-by-point filter in the frequency domain, unable to perceive the topological correlation between adjacent frequency components, as shown in **Fig. S2**. When we expanded the frequency-domain convolution kernel to 3×3 (net54), the network performance experienced a significant leap: the RMSE decreased to 0.37 nm, the Rel Acc substantially increased to 86.86%, and the SSIM reached 0.9052. This quantitative result strongly supports the physical hypothesis presented in the main text—because mid-frequency error possesses randomness and anisotropy, its spectral energy exhibits a topological distribution accompanied by continuous side lobes. The 3×3 convolution kernel introduces a local receptive field in the frequency domain, enabling the LSG module to comprehensively consider the energy continuity of its neighborhood when evaluating a specific frequency component, thereby precisely preserving weak yet valid texture features. Conversely, when the convolution kernel was further enlarged to 5×5 (net55), the excessively large frequency-domain receptive field caused the

high-frequency spectral noise to be overly smoothed and erroneously introduced, triggering a degradation in accuracy. Therefore, the 3×3 LSG was established as the optimal scale. These ablation results indicate that the observed physics-constrained correction behavior is not produced by network smoothing alone, but arises from the combined effect of the LSG-based spectral perception, dual-domain physical losses, and the regularization inherent in supervised network training.

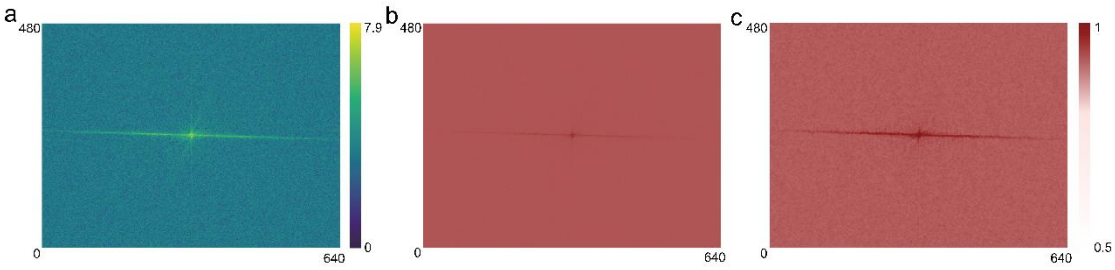

Fig. S2 Necessity of 3×3 Convolutional Kernels. **a** Spectrum of a random sample. **b** Mask generated with a 1×1 kernel. **c** Mask generated with a 3×3 kernel.

S4. Composition of the Interferometer (Low-Frequency Error) Dataset

Table S2. Composition of interferometer data samples

Classification Category	Specific Composition
Optical Component Type	Approximately 70% mirrors (covering various processing stages), 16% lenses, and 14% other types of optical components (e.g., gratings, CGH, freeform optical elements, etc.)
Error Distribution	Average RMS ratio of low-frequency error (amplitude): 85.29% Average RMS² ratio of low-frequency error (energy): 73.95% Average PSD curve integral ratio of low-frequency error: 62.32%
Interferometer Model	Zygo GPI, Verifire™ MST, and Verifire™ HDX interferometers, with proportions of approximately 30%, 35%, and 35%, respectively.

S5. Composition of the Profiler (Mid-Frequency Error) Dataset

Table S3. Composition of optical profiler data samples

Classification Category	Specific Composition
Mid-Frequency Machining Texture Type	Includes textures generated by manufacturing processes such as grinding, polishing, magnetorheological finishing (MRF), ion beam figuring (IBF), single-point diamond turning (SPDT), and bonnet polishing, encompassing various tool paths including longitudinal, transverse, raster, spiral, and screw patterns.
Error Distribution	Average RMS ratio of mid-frequency error (amplitude): 65.09% Average RMS^2 ratio of mid-frequency error (energy): 46.21% Average PSD curve integral ratio of mid-frequency error: 45.32%
Optical Profiler Model	Zygo Newview™ 7200 and Newview™ 9000 3D Optical Surface Profiler, with proportions of approximately 90% and 10% respectively.

S6. Dataset Preprocessing Pipeline

To ensure the clarity and reproducibility of our study, this section provides a comprehensive, step-by-step procedure for the entire dataset preprocessing and normalization pipeline, detailing how the raw surface error data was standardized in terms of both spatial scale and height range.

To ensure the stability of the neural network's training and the generalizability of the results, we performed a series of standardized preprocessing steps on the collected raw surface error data (which includes both the "actual surface error" and its corresponding "low-frequency error"). The entire procedure, aimed at unifying the spatial scale and numerical range of all data, consists of three main parts: surface error height (z-axis) normalization, surface error size (x,y-axes) normalization, and bad pixel restoration.

S6.1. Surface Error Size Normalization

For the interferometer, because the raw data originates from different models, the pixel resolutions and effective circular aperture areas vary. To unify the input dimensions, we designed a two-step scaling pipeline to standardize the effective area of all surface data into a circle with a diameter of 510 pixels (embedded within a 512×512 matrix). The specific steps are as follows:

1. **Preliminary Scaling:** First, the actual surface error data and the low-frequency error data are uniformly scaled to 512×512 pixels using interpolation algorithms (bicubic for the actual surface error to preserve details, and bilinear for the low-frequency error).
2. **Secondary Fine-Scaling:** To address cases where the edges of some actual surface error data remain incompletely cropped after preliminary scaling, we first detect the actual effective (non-NaN value) data range within a standard

circular area with a radius of 255 pixels. If this effective range does not fill the 510-pixel diameter area, a magnification factor is calculated, and nearest-neighbor interpolation is used to further scale this area to a circular region with a 255-pixel radius, ensuring the data dimensions are perfectly consistent.

For the profiler, although different models have varying pixel resolutions and effective aperture areas, the issues to be considered during size scaling are completely different from those of the interferometer: First, the profiler data is a complete rectangle, eliminating the need to address the issue of edge information collapse as required for interferometer data; second, the profiler data contains a vast amount of high-frequency error information. If interpolation scaling were adopted as with the interferometer data, it would cause distortion in the high-frequency error, making it impossible to obtain a usable high-frequency error component subsequently. Therefore, for different profiler data, we uniformly crop them to a size of 640×480 pixels. This is because the majority of profiler data are 640×480 pixels, and the remaining data are larger than 640×480 pixels.

S6.2. Surface Error Height Normalization

To eliminate the influence of height range variations among different surface errors on model training, we normalized the error height to the [-1, 1] interval through linear mapping.

The processing steps for the actual surface error data are as follows and **Fig. S3**:

1. Calculate the maximum, minimum, and median (statistical median) values of each surface error height.
2. Subtract the median value from the entire surface error height map.
3. Based on the new maximum absolute value, normalize the data to the [-1, 1] range, and record the normalization ratio to facilitate subsequent restoration of the original scale range.

Data processing for different frequency band errors follows the same steps. Crucially, to preserve the true physical scale relationship, the corresponding frequency band error must be normalized using the exact same normalization factor calculated from its corresponding actual surface error. This ensures that the amplitude of the frequency band error component maintains the same proportion relative to the total error, providing the network with physically meaningful inputs.

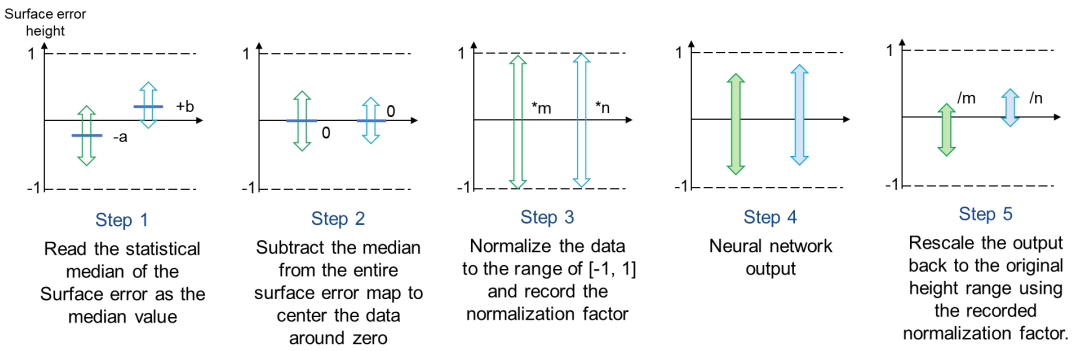

Fig. S3 The workflow for surface error height normalization and restoration. This figure illustrates the key steps for normalizing the height values of both the actual surface error (green) and the corresponding low-frequency error (blue). (Step 1) shows the initial PV ranges. (Step 2) shows the data after being centered by subtracting the median value. (Step 3) is the crucial normalization step where the actual error is scaled to the [-1, 1] range, and its calculated normalization factor is applied to the frequency band errors to preserve the true physical scale ratio. (Steps 4-5) illustrate the conceptual process of restoring the original physical scale using the stored factor.

S6.3. Bad Pixel Restoration

Due to various conditions during the testing process, "bad pixels" (pixels with a NaN value) often appear in the results. We use fillmissing interpolation to fill and restore these points to ensure data integrity. For the low-frequency error dataset, we set all background values outside the 255-pixel radius circular area to 0 to prevent NaN values from causing network training to fail. For the mid-frequency error dataset, however, this step is not necessary.

The complete dataset preprocessing and normalization process is shown in **Fig. S4**. Through the above procedures, we obtained a fully standardized dataset suitable for training the neural network. While the selected interpolation methods were found to be optimal for the dataset in this study, researchers are encouraged to evaluate different methods to determine the most suitable approach for their specific datasets.

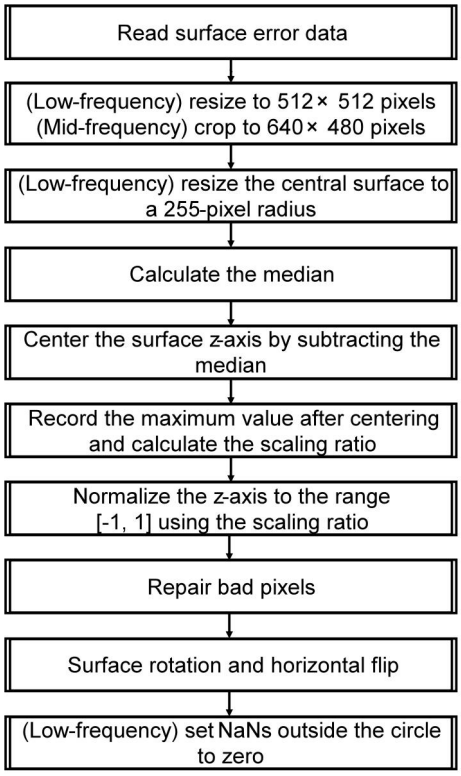

Fig. S4 Flowchart of the dataset preprocessing and normalization pipeline. This chart outlines the entire procedure for transforming raw surface error data into a standardized format suitable for neural network training. The key stages include: (1) a two-step spatial scaling process to unify pixel dimensions, (2) height normalization to the [-1, 1] range, (3) restoration of bad pixels, and (4) final data cleaning and orientation alignment.

S6.4. Dataset splitting and Batch-level isolation strategy

To maximize the model's generalizability while strictly preventing data leakage, the following two-fold strategy was implemented during the partitioning of the training, validation, and test sets:

1. **Cross-process and cross-component mixing:** To force the network to learn surface features with physical diversity, data from different manufacturing processes (e.g., mechanical polishing, magnetorheological finishing) as well as various materials and geometries were randomized and mixed across all datasets. This strategy effectively enhances the model's generalizability when encountering multi-source and multi-type testing data.
2. **Batch-level isolation:** Multiple local measurements of the same optical component across different fields of view represent highly correlated data. If randomized directly during dataset partitioning, it would lead to significant "data leakage," resulting in artificially inflated performance metrics on the test set. Consequently, all data partitioning in this study was performed using the original measurement folders (representing a unique optical component or a single

manufacturing batch) as the minimum independent units. This ensures that all detection data from the same optical device are restricted to a single set, preventing the occurrence of highly correlated data across different sets.

Following the automated partitioning process, manual cross-verification was conducted to eliminate the risk of data leakage, ensuring the authenticity and objectivity of the final evaluation results.

S7. Network Architectures of Baseline Models for Comparative Experiments

This paper employs traditional Convolutional Neural Networks (CNNs), U-Net, Transformer, and PINNs as baseline models to comparatively evaluate the performance of the proposed DPN. To provide a clear basis for comparison, this section details the network architectures of the aforementioned baseline models.

Fair benchmark protocol: To ensure fairness in the comparative experiments, all baseline models and the proposed DPN were trained and evaluated using the same training/validation/test split, preprocessing pipeline, manually filtered reference labels, and evaluation metrics. The unified preprocessing and label regime defines the common input-output setting of this engineering detection task and was applied identically to all methods. The hyperparameters of each model were selected according to validation-set performance. Except for the inherent architectural and loss-function differences of each method, no additional data, preprocessing operation, or label information was provided to any specific model.

As a baseline comparison model, the classic CNN adopts the symmetrical encoder-decoder architecture common in image-to-image regression tasks, without introducing any skip connections. For low-frequency error extraction, the network focuses on the downsampling compression of large-scale features. The encoder consists of 3 downsampling blocks, sequentially containing a 3×3 convolutional layer with a stride of 1, Batch Normalization (BN), a ReLU activation function, and a 2×2 max pooling layer with a stride of 2. Subsequently, the features enter a bottleneck layer composed of a 3×3 convolution, BN, and ReLU. The decoder utilizes three 4×4 transposed convolution blocks with a stride of 2 to progressively upsample the feature maps to the original resolution, and finally outputs through a 3×3 convolutional layer and calculates the regression loss. For mid-frequency error extraction, to adapt to the dense distribution of mid-frequency features and improve computational efficiency, the network streamlines the initial number of channels (16→32→64, bottleneck layer 128). Specifically, its spatial downsampling is entirely dominated by 3×3 convolutions with a stride of 2, while the stride of the 2×2 max pooling layer is

strictly limited to 1. This means the pooling operation is only used to enhance local extremum features (e.g., tool mark peaks and valleys) without additionally compressing the spatial resolution, avoiding the excessive loss of mid-frequency morphological information. The decoder progressively restores the number of channels (64→32→16) through three 4×4 transposed convolutions with a stride of 2, and ultimately directly outputs the linear regression prediction result via a 3×3 convolution.

U-Net adopts the classic symmetrical U-shaped fully convolutional architecture, featuring double convolution blocks (DoubleConv) composed of cascaded 3×3 convolutions (padding of 1), BN, and ReLU as the main body, and fuses multi-scale features through skip connections. For low-frequency error extraction, the network employs a deeper 4-level downsampling structure to obtain a massive global receptive field (the number of channels increases progressively as 64→128→256→512→1024). The decoder gradually restores the resolution through bilinear interpolation combined with DoubleConv blocks, and finally generates a continuous and smooth low-frequency morphology through a 1×1 convolution and a Tanh activation function. For mid-frequency error extraction, to prevent overly deep pooling from irreversibly erasing local high-frequency textures, this paper constructs a lightweight variant (Light U-Net). This structure reduces the downsampling to 3 levels, and streamlines the base channel number sequence to 16→32→64→128, effectively limiting the over-expansion of the receptive field; simultaneously, the output layer discards the Tanh activation and directly adopts a linear 1×1 convolution to output the prediction result, which is more conducive to fitting minute mid-frequency errors and enhancing computational efficiency.

As an architecture based on self-attention, Transformer establishes global spatial dependencies through image patch splitting and the multi-head self-attention mechanism. For low-frequency error extraction, the network adopts a high-capacity configuration to characterize the complex global surface morphology. First, the image is projected into a 256-dimensional embedding vector through a 16×16 convolution kernel (stride of 16), and the encoder stacks 6 standard Transformer blocks (including 8-head attention). After the decoding stage reshapes the sequence into a 2D feature map, it employs 4 continuous upsampling stages (bilinear interpolation combined with 3×3 convolution, BN, and ReLU) to progressively decrease the number of channels (64→32→16), ultimately outputting via convolution and Tanh. This progressive interpolation naturally possesses a low-pass filtering effect, which aligns with the attributes of low-frequency error. For mid-frequency error extraction, to prevent excessive smoothing of high-frequency details, the network's embedding dimension is reduced to 128, and only 4 Transformer blocks are stacked (including 4-head attention). Crucially, the decoder abandons progressive interpolation and instead employs a single-step large-stride transposed convolution (16×16 convolution kernel, stride of 16) to directly map the sequence proportionally into a 64-channel feature map at the original resolution. Subsequently, it utilizes GELU activation and

3×3 convolution for the output, preserving the mid-frequency morphology to the greatest extent.

To verify the impact of physical constraints on the extraction of errors in different frequency bands, this paper introduces PINNs as a baseline model incorporating physical priors. PINNs uniformly introduce spatial derivative penalty terms based on Sobel and Laplace operators into the loss function to ensure the continuity of the slope and curvature of the predicted error. Regarding the main network architecture: For low-frequency error extraction, a high-capacity architecture integrating attention mechanisms is adopted. The encoder contains 3 downsampling stages (including 3×3 convolutions with a stride of 2, residual blocks, and channel attention), with channels expanded to 64, and the bottleneck layer cascades spatial attention and residual modules; the decoder employs 3 sub-pixel upsampling layers for efficient upsampling, followed by residual blocks and spatial attention to refine features, finally outputting via a 3×3 convolution and Tanh. For mid-frequency error extraction, to prevent high-capacity networks from causing the degradation of high-frequency details, the network degrades into a standard vanilla encoder-decoder (vanilla ED). The encoder discards all attention and residual modules, containing only 2 downsampling stages (3×3 convolution, Tanh, and 2×2 average pooling, channels 32→64), and the bottleneck layer is a single 3×3 convolution. The decoder corresponds to 2 bilinear interpolation upsampling stages, paired with 3×3 convolutions to restore the channels to 32 and ultimately output. Herein, the use of average pooling instead of max pooling is intended to better preserve the overall energy distribution of the mid-frequency periodic manufacturing textures.

To verify the performance of the DPN against models integrating frequency-domain architectures, this paper introduces the Physics-Informed Fourier Neural Operator (PI-FNO) and the U-shaped Adaptive Fourier Neural Operator (U-AFNO) as advanced baseline architectures.

PI-FNO is designed to capture global surface morphology by parameterizing operator mappings directly in the Fourier domain. The backbone network initially projects the single-channel input into a 16-channel feature map via a 1×1 convolution (width = 16), followed by cascading four identical spectral convolution blocks (SpectralConv2d). Within each block, the input is mapped to the frequency domain via Fast Fourier Transform (FFT). Matrix multiplication is performed exclusively on the truncated low-frequency modes, while the remaining high-frequency components are zeroed out. The output is then reconstructed back to the spatial domain via Inverse Fast Fourier Transform (IFFT), added to a parallel 1×1 residual convolution path, and activated by GELU. Finally, the features are mapped directly to the single-channel prediction via a 1×1 convolution. For low-frequency error extraction, to precisely capture macroscopic large-period low-frequency morphology while preventing ringing artifacts caused by global mode truncation, the retained frequency modes are strictly limited to 24×24 (modes1 = 24, modes2 = 24). For mid-frequency error extraction, to filter out high-frequency random roughness noise and accurately capture densely

distributed periodic manufacturing marks, the retained frequency modes are adjusted to 32×32, ensuring the complete preservation of valid mid-frequency morphological information.

U-AFNO combines the multi-scale local receptive field of a classic U-shaped architecture with the adaptive global frequency token mixing mechanism of AFNO. The encoder consists of 3 downsampling stages, each sequentially containing a 3×3 convolution with a stride of 2 (padding of 1), BN, and GELU, progressively expanding the channels as 1→32→64→128 (base dimension = 32). The decoder employs 3 bilinear interpolation upsampling stages combined with 3×3 convolutions (stride of 1) to restore the resolution and channels (128→64→32→32). Multi-scale local features from the first two stages of the encoder are fused via skip connections, and the final error map is output through a linear 3×3 convolution. For low-frequency error extraction, the bottleneck layer integrates an AFNO block with 128 channels. The feature map is reshaped into 8 independent blocks (numblocks = 8) for block-diagonal complex weight interaction in the Fourier domain, and an adaptive Soft-Shrinkage thresholding mechanism (sparsity threshold = 0.01) is introduced to dynamically filter out redundant energy. This is cascaded with a standard 3×3 convolution and a second AFNO layer to generate a continuous low-frequency morphology. For mid-frequency error extraction, the backbone network and bottleneck configurations remain identical to the low-frequency task. The extraction relies entirely on the dynamic Soft-Shrinkage mechanism within the bottleneck AFNO to isolate minute mid-frequency mark energy from high-frequency measurement noise, which is highly conducive to fitting local extremum textures without compressing spatial resolution.

S8. Robustness of the Normalized Spatial Frequency Boundary Across Different Manufacturing Processes

To intuitively demonstrate the robustness of the normalized spatial frequency boundary ($f_{norm} = 8$) across different manufacturing processes, we present the actual filtering results of typical samples processed by different techniques.

Different manufacturing processes inherently generate distinct surface features. For instance, magnetorheological finishing (MRF) often leaves regular parallel tool marks, small lap polishing produces concentric zones, and ion beam figuring (IBF) may introduce random mid-frequency ripples. Despite these significant differences in error amplitude and spectral energy distribution, the transition point of the frequency band error morphological features (i.e., the transition from global macroscopic low-frequency morphology to local microscopic mid-frequency manufacturing textures) stably converges near $f_{norm} = 8$.

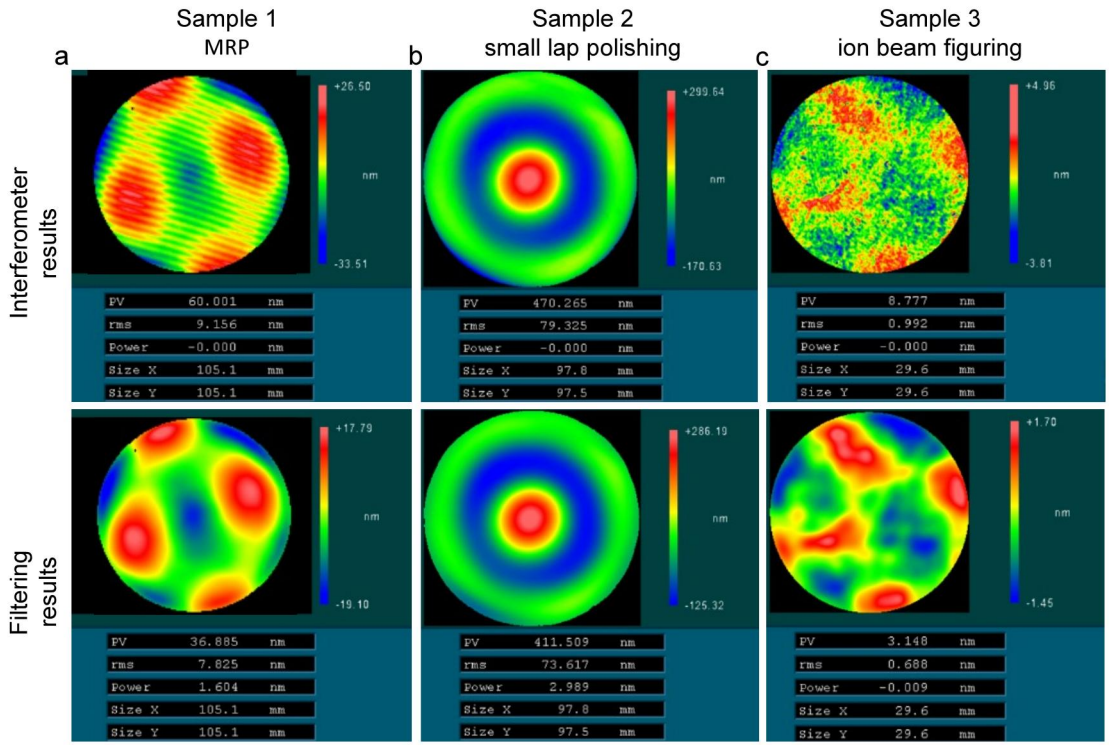

Fig. S5 Robustness of the normalized spatial frequency boundary across different manufacturing processes. **a** Magnetorheological finishing (MRF). **b** Small lap polishing. **c** Ion beam figuring (IBF). Based on the normalized spatial frequency division, the low-mid frequency cutoff wavelengths for the three samples are 13.138 mm, 12.225 mm, and 3.7 mm, respectively. The filtering results effectively retain the complete macroscopic low-frequency morphology.

As shown in **Fig. S5**, based on this normalized spatial frequency division, the low-frequency morphological features of samples from MRF, small lap polishing, and IBF are completely and accurately retained in the filtering results, while their respective distinct mid-frequency manufacturing textures are effectively separated. This fully verifies that the proposed boundary is highly robust to multi-source manufacturing processes.

S9. Downstream Optical Performance Gains and Engineering Application Analysis

To clearly demonstrate the engineering value of high-precision full-band error decoupling in practical optical manufacturing and metrology, this section quantitatively analyzes the downstream translation benefits of DPN extraction results from two dimensions: macroscopic wavefront quality assessment and microscopic surface scatter quantification.**S9.1. Wavefront Quality and Deterministic Polishing Baseline**

In practical large-aperture interferometry, raw data is frequently corrupted by local high-frequency spikes and measurement noise induced by environmental disturbances or dust. Directly calculating the peak-to-valley (PV) value from raw data often leads to significant evaluation distortion, thereby misleading downstream deterministic polishing processes. The DPN robustly isolates these high-frequency spikes and measurement noise, extracting a high-fidelity macroscopic low-frequency morphology.

To verify the reliability of the reconstructed wavefront, we extracted five typical sets of raw interferometer measurement data alongside the low-frequency morphology predicted by the DPN, and introduced global fitting using the classic first 36 Zernike polynomials as a reference baseline. As shown in **Table S4** (for intuitive comparison, the raw interferometric PV values for all samples are normalized), because the crosstalk from local noise spikes is effectively suppressed, the PV values of the DPN predictions are highly consistent with the reference PV baselines obtained via Zernike polynomial fitting. This demonstrates that the DPN provides a highly stable and reliable macroscopic wavefront evaluation baseline for optical systems, effectively avoiding process stagnation or convergence failure caused by local artifacts.

Table S4. Comparison of DPN reconstructed wavefront quality and traditional PV evaluation (normalized)

Sample	Raw Interferometer PV	DPN Predicted Low-Freq PV	Zernike Polynomial Fitting PV
1	2.0000	1.4672	1.3910
2	2.0000	1.3666	1.2754
3	2.0000	1.5667	1.5110
4	2.0000	1.1383	1.0970
5	2.0000	1.3274	1.3185

S9.2. Surface Scatter Quantification (Scatter Reduction) and Process Tracing

The small-angle and large-angle light scattering levels of high-end optical systems (e.g., extreme ultraviolet (EUV) systems) are highly dependent on mid-frequency textures and high-frequency roughness errors on the component surface, respectively. By performing real-time, high-precision decoupling on high-resolution data from a white-light optical profiler, the DPN simultaneously outputs highly accurate mid-frequency and high-frequency roughness error components.

As shown in **Table S5**, the extracted high-frequency root mean square (RMS) component can be used directly to quantify the microscopic high-frequency scattering level of the surface. Meanwhile, the separated mid-frequency RMS component not only serves to evaluate mid-frequency scatter reduction, but its preserved minute texture morphology also clearly reflects machining conditions such as spindle speed and feed rate, assisting engineers in process tracing and parameter optimization.

Table S5. Separation results of mid- and high-frequency scattering metrics from profiler data using DPN

Sample	Raw Profiler RMS (nm)	DPN Predicted Mid-Freq RMS (nm)	DPN Predicted High-Freq RMS (nm)
1	1.1750	0.9497	0.5080
2	4.3087	2.8755	2.5120
3	0.7891	0.2240	0.7398
4	4.7637	3.5538	2.5756
5	0.6468	0.1687	0.6114

S10. Failure Cases and Limitations

Although the DPN framework demonstrates excellent decoupling precision across the vast majority of full-band error extraction tasks, it remains subject to reconstruction limitations or failure risks under specific scenarios due to signal feature overlapping and sampling limits in extreme conditions. This section summarizes three typical cases combined with actual measurement data:

S10.1. Periodic manufacturing textures, extreme machining conditions and surface defects

As shown in **Fig. S6**, when the interferometer input contains extremely strong and dense global periodic manufacturing textures, the decoupling performance of the DPN varies depending on the texture feature scales. For periodic textures with relatively moderate feature scales (**Fig. S6c, d**), the DPN robustly eliminates all high-frequency texture components, retaining only the continuous, smooth, and complete low-frequency error. However, when the input surface contains periodic tool marks of varying scales (**Fig. S6a, b**) or represents an extreme experimental condition for special machining verification (**Fig. S6e**), the DPN-predicted low-frequency results may retain some large-scale periodic manufacturing textures because their local spatial envelopes closely approach the macroscopic low-frequency morphology. Objectively speaking, under such special conditions as **Fig. S6e**, discussing the traditional low-frequency error morphology loses its physical meaning.

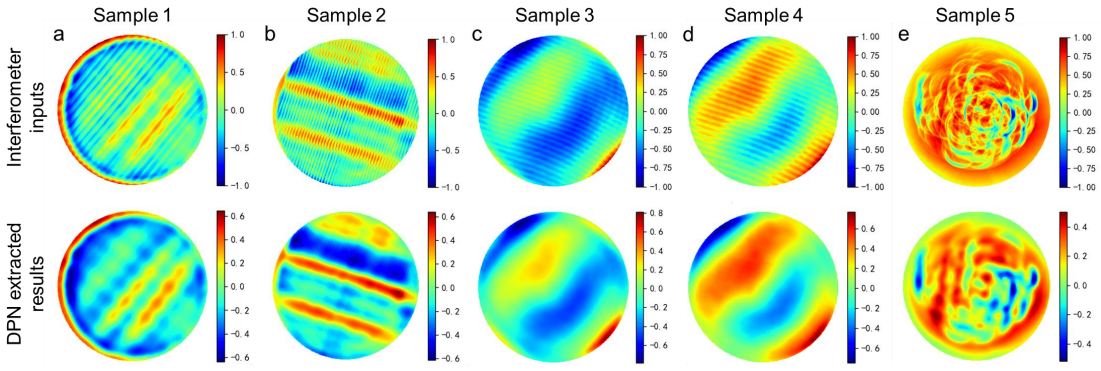

Fig. S6 Challenging cases in low-frequency error extraction. **a, b** DPN-predicted results retaining some large-scale periodic manufacturing textures. **c, d** DPN effectively eliminates all periodic manufacturing textures, preserving only the low-frequency morphological features of the surface. **e** An extreme machining condition (e.g., a verification experiment for a special machining method) and its corresponding DPN low-frequency error extraction result.

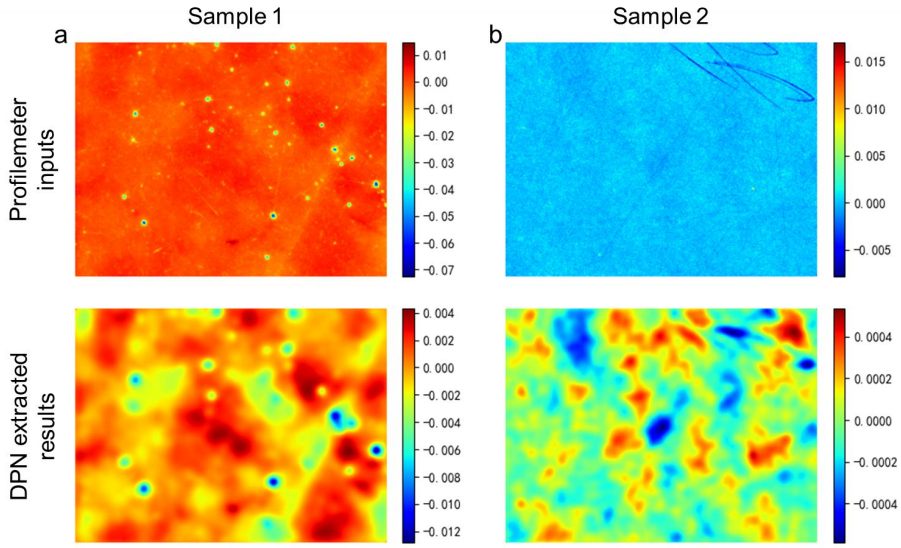

Fig. S7 Challenging cases in mid-frequency error extraction. **a** DPN-predicted result retaining partial pit defects. **b** DPN-predicted result completely eliminates the scratch located in the upper right corner of the field of view.

When processing profiler measurement data, surfaces are often accompanied by manufacturing defects (e.g., pits and scratches), which the DPN might occasionally miss. As shown in **Fig. S7**, the DPN perfectly eliminates the continuous scratch located in the upper right corner of the field of view (**Fig. S7b**). However, when the measured surface contains large pit defects (**Fig. S7a**), the predicted results partially retain these distinct pit features because the local spatial gradients at the edges of some pits physically overlap with mid-frequency texture features to a certain extent.

S10.2. Resolution and spatial sampling rate

The DPN imposes certain requirements on the input. Although the convolution-based DPN supports inputs of various sizes, the length and width of the input matrix must be integer multiples of 8 due to the network's upsampling and downsampling structural design. Fortunately, the routine measurement resolutions of commercial interferometers and profilers (e.g., 640×480, 1696×1696, and 3296×3296 for interferometers; 640×480 and 1000×1000 for profilers) generally satisfy this requirement. If non-compliant data occurs, pre-processing such as cropping or interpolation scaling is required. Furthermore, if a mismatched low-magnification objective or an excessively low-resolution instrument is used during measurement,

valid frequency band error information may be lost according to the Nyquist sampling theorem. This can even induce severe frequency aliasing, thereby misleading the physical spectrum perception process within the DPN. Therefore, the validity of this method relies on appropriate instrument, lens, and sampling rate settings during initial data acquisition.

S10.3. Normalized spatial frequency

The normalized spatial frequency is currently established based on all available interferometer and profiler measurement data. Although the current dataset encompasses a diverse range of manufacturing processes and surface topologies to the greatest extent possible, the DPN may occasionally yield sub-optimal predictions (as discussed in point (1) above). Furthermore, the dataset may not yet fully cover certain specialized topologies or specific machining methods (e.g., computer-generated holograms (CGHs) or metasurfaces containing strong artificial periodic structures). Consequently, under these specific circumstances, the normalized spatial frequency may deviate from the optimal band division point or even face the risk of failure.

References

[1**]** Liu, Y. et al. Data-driven precise extraction and division of spatial frequency band errors. *Opt. Laser Technol*. (in the press).