Journal of Applied Computational Intelligence

eISSN: 3136-6937

Vol. 4, No. 2, April, 2026

Spatially Varying Point Spread Function Conditioned Swin Transformer for Lensless Image Reconstruction

Hsin-yu Yu[1,*], Ya-chieh Mo[1], Kuang-chun Kuo[1], Hsiao-wei Chiou[1]

[1] *College of Electrical Engineering and Computer Science, National Taiwan University of Science and Technology, Taipei, 10607, China*

Received 08 January 2026; Revised 25 April 2026; Accepted 30 April 2026; Published 05 May 2026

Abstract

Lensless cameras encode scene radiance via a thin optical element; due to the spatial variation of the point spread function, a single global blur model is insufficient for accurate reconstruction over the whole field of view. This study presents a reconstruction framework for static, approximately planar scenes that conditions a hierarchical Swin Transformer on a calibrated spatial response field. A compact optical encoder generates feature modulation signals from locally interpolated response descriptors and shifted-window attention facilitates the exchange of image information among adjacent areas. Measurement-consistency correction restricts the reconstructed image by employing the same spatially-varying forward operator for calibration. A reproducible evaluation design separates scene content, calibration positions and perturbation levels for the examination of fidelity, peripheral behavior and sensitivity to optical mismatch. The reported comparison gives the conditioned model a peak signal-to-noise ratio of 29.84 dB and structural similarity of 0.914, compared with 27.63 dB and 0.872 for the matched unconditioned Swin model. Peripheral reconstruction error decreases from 0.091 to 0.059, while the medium configuration uses 11.8 million parameters and has a latency of 43 ms. These metrics characterize image fidelity, peripheral accuracy, and computational demand within a common evaluation framework. The framework explicitly incorporates spatial calibration into the reconstruction input and enables separate evaluation of optical adaptation and image prior learning. Experiments on independently calibrated hardware are required to establish practical accuracy and transferability.

Keywords: Lensless Imaging, Spatially Varying Point Spread Function, Swin Transformer, Physics Conditioned Reconstruction, Calibration Robustness

1. Introduction

Lensless systems can integrate a phase mask with a bare image sensor and recover scene information computationally from the encoded measurements [1]. Recent reviews describe lensless imaging as a joint design problem involving optical modulation, sensor acquisition, and computational reconstruction [2]. Reconstruction algorithms recover object information from intensity measurements shaped by the encoding element, as demonstrated by iterative lensless diffraction imaging with rotated slit modulation [3]. Diffuser-based lensless reconstruction has also benefited from learned space-time priors that exploit temporal redundancy to improve image quality under noise [4]. Coded-aperture designs provide another approach to thin imaging systems, and their reconstruction behavior is related to the structure of the encoding pattern. Practical utility depends on spatial fidelity, radiometric consistency, and stability across acquisition conditions. Spatial detail, peripheral fidelity, brightness uniformity and stability in the face of various acquisition condition changes all impact whether the reconstructed image retains the necessary information for

* Corresponding author: hinyu.yu@ntust.edu.tw

http://dx.doi.org/10.68406/jaci.2026.vol4iss2no7:90-105

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

subsequent engineering applications. The above requirements are difficult to meet if an inversion method assumes that a single optical response is sufficient for the whole-field description. It is especially difficult near the sensor edges; there, response truncation and asymmetric encoding may alter the relationship between local scene content and the received signal. Therefore, a suitable reconstruction model needs to be able to reproduce the measurement process at the spatial scale of its actual behavior change. Compact hardware offloads some of the imaging load to computation; thus, the quality and interpretation of the forward model are at the core of system design.

Point-spread function is the sensor's response to a localized source under given acquisition conditions. Shift-invariant approximation only varies the responses of different scene locations by translation and thus enables efficient convolutional modelling. Wide-field lensless systems can deviate from the above approximation due to the effects of incident angle, finite sensor size and optical alignment on the recorded pattern. Learned spatially varying microscopy models estimate position-dependent point spread functions and provide differentiable forward representations for reconstruction [5]. Studies of miniaturized lensless cameras explicitly incorporate cropping into optimized imaging models to address the effects of truncated measurements [6]. Training using a global response couples scene variation and optical variation; thus, the source of an observed feature change is unclear. Resolving this uncertainty requires information about the local measurement process. Optical information must be presented in a form that discriminates between field position and scene texture, but preserves the variations necessary for inversion. Calibration density adds another restriction. Measurement of the response at each scene pixel is costly; sparse sampling requires interpolation and is susceptible to uncertainty due to variations in sampled locations. Research on calibration-efficient lensless modelling addresses the above conflict by means of reduced representations and structured interpolation. The above problem relates to the accuracy of the spatial response model and how reconstruction employs this model.

Learning-based inverse imaging methods provide image priors to reduce noise and reconstruct structures with limited constraints from measurements. Qian et al. [7] unfold iterative shrinkage and thresholding into trainable stages that combine regularization with learned denoiser priors for lensless reconstruction. Convolutional reconstruction networks learn spatial regularities efficiently, although coordinating widely distributed encoded information can require substantial network depth. Attention-based restoration models combine multiscale processing with self-attention to exchange information across spatial and temporal regions [8]. Efficient Swin-based super-resolution models demonstrate that attention design can reduce the computational demands of reconstructing spatial detail [9]. Encoding the relationship between a lensless measurement and a position-dependent response field requires an explicit representation of the acquisition process. Research on restoration under out-of-sample degradations further demonstrates the importance of adapting learned representations to acquisition changes [10]. Position is identified by coordinate embeddings and associated optical behavior is described by response descriptors. Conversely, misalignment between optical descriptors and sensor coordinates can associate response information with an incorrect spatial location. The research gap concerns how calibrated spatial response information can guide hierarchical feature exchange while remaining consistent with a shared measurement model. Matched component controls are needed to distinguish optical adaptation from image-prior capacity.

The contribution is the integration of spatially calibrated response descriptors, multiscale feature modulation, and a shared measurement-consistency operator within a Swin reconstruction framework. Novelty is attributed to this optical conditioning strategy and its evaluation, while the shifted-window attention mechanism follows established architecture principles. The goal of the engineering problem is to determine if explicit optical conditioning can reduce spatially structured reconstruction errors without having the image branch infer optical variation only from scene examples. The above four components of the proposed research - response interpolation, condition encoding, attention-based reconstruction and measurement correction - are separated to be studied under matched computational budgets. It also distinguishes calibration noise from scene noise and separates robustness to an imperfect response estimate from transfer to a completely different optical device. The above differences affect the collection of evidence and the interpretation of any enhancements. A specific optical form is linked to a regulated reconstruction assessment method in this research. Section 2 presents optical inversion, learned reconstruction and spatial conditioning, and highlights the underlying assumptions of this paper. Section 3 presents the calibrated forward model and the construction of spatial condition features. Section 4 introduces the reconstruction network, training objective and implementation protocol. Design of evaluation, comparative data, component analysis and deployment-oriented tests are presented in Section 5. Section 6 presents the methodological contributions, limitations of the present evidence and directions for future experimental studies. The organization separates the optical assumptions from the learned architecture and conducts the most in-depth analysis in the experiments.

2. Related Work

2.1. Spatial response modeling in lensless imaging

Calibrated inversion represents a lensless camera through an identifiable measurement response. Classical methods use a pre-existing forward operator and add a regularization term to favor realistic scene structures. Recent lensless

inversion methods combine total variation constraints with deep image priors to stabilize reconstruction from intensity measurements [11]. The above forms enable independent inspection and modification of the measurement operator from the regulariser. The computational expense is affected by the selected solver, the organization of the operator and the termination condition. Convolution and its adjoint can be computed efficiently in the frequency domain by mean of a shift-invariant approximation. Spatial variation is disruptive to the above structure due to differing scene-location contributions of patterns to the sensor. Although a single kernel can still offer a good approximation in a limited area, the accuracy of this approximation needs to be evaluated with respect to the field and geometry of reconstruction.

Several kernels and spatial weights are used in local convolution models for field dependence. Experiments with multilayer coded apertures show that the structure and conditioning of the optical response influence lensless reconstruction accuracy [12]. The difference in relation to the above is that they operate on scene contributions before convolution or on sensor outputs after convolution; thus, they are different operators. Interpolating complete responses preserves measured pattern structure but can require substantial storage. Interpolating response descriptors reduces storage but introduces representation error that should be measured separately from image reconstruction error. Boundary handling is included in the forward model as the detector only records the part of the response within its active area. Whole-field characterization therefore requires spatially resolved calibration assessment. Current optical modelling research inspires the development of a reconstruction system that preserves sensor cropping, field coordinates and interpolation uncertainty explicitly. The efficiency of a neural architecture in utilizing the above quantities is still unknown empirically.

2.2. Learned priors and hierarchical attention

Learned reconstruction methods replace or add to hand-designed priors with trainable mappings. A supervised network can learn a direct measurement for image transformation; an unrolled network maintains a sequence of updates linked to a data-fidelity objective. Multi-PSF lensless reconstruction has incorporated a PSF-aware auxiliary branch and a forward model within cyclic adversarial learning [13]. This formulation offers conceptual and computational advantages by exposing operator modifications within the reconstruction pathway. Across these learned formulations, robustness depends on operator accuracy and the structure of the trainable reconstruction modules. Algorithm unrolling and cyclic adversarial training provide distinct mechanisms for incorporating that information. The training distribution, initialization, and parameterization of the learned correction also affect reconstruction behavior.

Convolutional encoder-decoder models have multi-scale context and strong inductive biases for local image structure. Dynamic-window Swin restoration combines content-adaptive self-attention with channel and spatial attention to capture image dependencies at different ranges [14]. Recent convolutional restoration research also demonstrates the value of multiscale processing and frequency modulation as efficient alternatives to transformer-based reconstruction [15]. Broad-sensor mixing for lensless reconstruction has context requirements related to the encoding geometry. A transformer receiving a geometrically aligned backprojection begins from a different representation than one that receives the raw multiplexed sensor frame. Therefore, fair comparisons need to control both the input representation and the architecture. Alternatively, an apparent attention benefit may be due to improved initialization or a more precise optical inversion step. A further methodological requirement is to isolate the effects of calibrated spatial information from differences in preprocessing.

2.3. Conditioning and optical uncertainty

Conditioning mechanisms enable the adjustment of an internal processing in the reconstruction model based on information from the acquisition process. Multiscale feature modulation has been used to refine image features through frequency-domain interactions and spatial guidance from local features [16]. In low-light lensless imaging, conditional generative restoration has been combined with a learned physics-based initial reconstruction to suppress residual noise [17]. The condition in an optical inverse problem can be used to specify the noise level, response shape, spatial position or acquisition geometry. These quantities represent distinct acquisition properties. Coordinate channels identify field location, whereas response descriptors characterize the optical behavior at that location. Two calibration states can therefore share coordinates while differing in response shape or throughput; evaluating these channels separately distinguishes spatial indexing from optical adaptation. The above difference is key to determining whether conditioning has learned optical adaptation or simply memorized a spatial training arrangement.

Uncertainty-aware computational imaging frameworks distinguish uncertainty associated with measurements from uncertainty associated with learned reconstruction models [18]. Calibration confidence is a measure of field-dependent reliability derived from interpolation distances or repeated measurements. Predictive uncertainty of the reconstructed image requires a separate calibration process. It should be based on an observable source of response error and evaluated by independent perturbations for utility. Current conditioning and uncertainty research motivates a compact optical branch, while its spatial resolution, interaction with window attention, and sensitivity to operator error require further

investigation. A controlled experiment includes misleading conditions, shuffled spatial assignments, and unmatched calibration states alongside matched response conditions.

3. Spatially Varying Optical Representation

3.1. Image formation and calibrated response field

The proposed system considers a static, approximately planar scene under incoherent illumination. Sensor values are assumed linear after dark subtraction and flat-field correction, and the nominal reconstruction domain excludes saturated measurements. The object and sensor grids are represented at the target reconstruction resolution, with a fixed mapping between their coordinate systems. A collection of nonnegative calibrated kernels approximates the spatial response field. Smooth weighting maps associate each scene location with neighboring calibration sites. The weights sum to one at each location, preserving the total interpolated contribution of a scene sample. Scene-dependent weighting precedes convolution, while sensor cropping is applied to the resulting measurement. The following equation expresses this image formation process:

$$y = \mathcal{C}\left[\sum_{k=1}^{K} h_k * (w_k \odot x)\right] + n \tag{1}$$

Here, x denotes scene radiance, y the corrected sensor measurement, h_k a calibrated response kernel, w_k a spatial interpolation map, $\mathcal{C}$ the sensor cropping operator, and n measurement noise. The operators * and $\odot$ represent convolution and elementwise multiplication, respectively. The kernel index runs over the calibrated response bank; at each object-grid location, the corresponding nonnegative interpolation weights sum to one over that bank. The formulation therefore preserves location-dependent optical contributions before convolution and sensor cropping. Its adjoint uses matching cropping and interpolation conventions to distribute sensor residuals back to the object grid. Before network training, random-vector inner-product checks verify the numerical consistency of the forward and adjoint implementations. The discrepancy is evaluated against a documented floating-point tolerance. Calibration kernels retain their full spatial support until sensor cropping is applied, preserving peripheral response truncation and suppressing artificial periodic wraparound.

Figure 1 depicts the optical representation workflow, from corrected point-source captures to aligned response kernels, spatial interpolation, operator construction, and condition generation. The calibration branch provides the response field and its confidence descriptors, while the measurement branch supplies the scene-dependent sensor frame. A shared coordinate transform connects these branches. This arrangement identifies the calibration quantities affected by changes in device alignment, optical spacing, or sensor geometry and provides a consistent basis for updating the reconstruction operator.

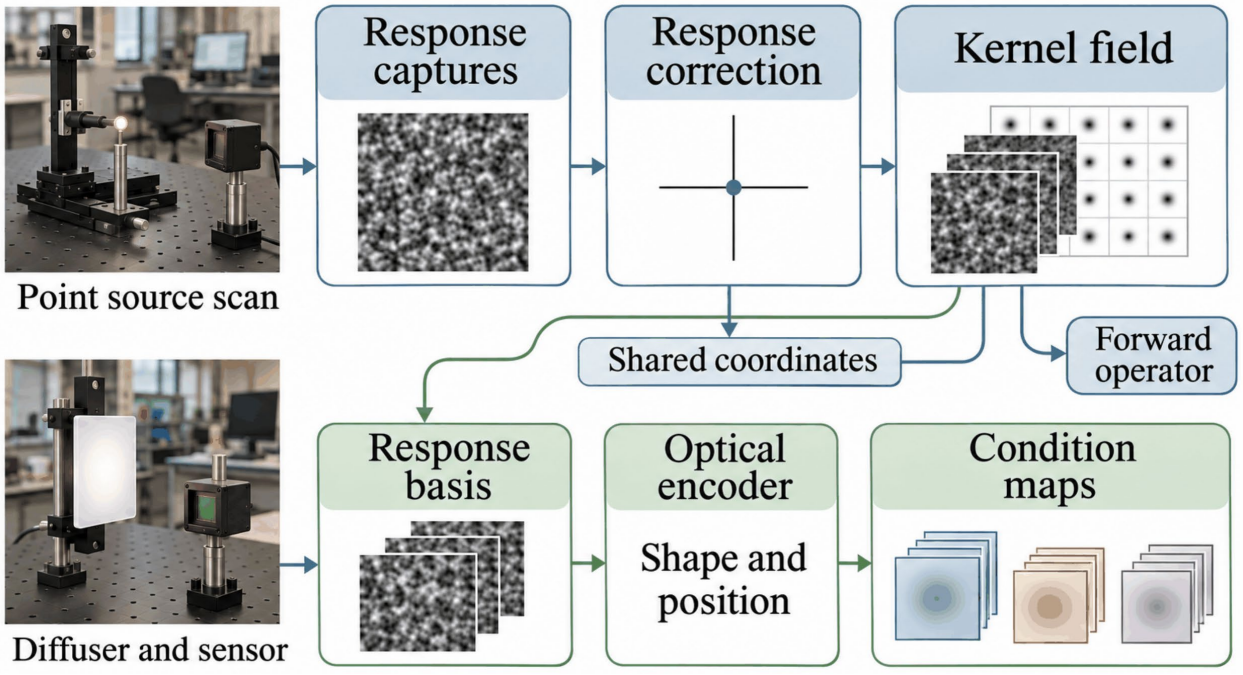


Figure 1. Workflow for calibration alignment, spatial response interpolation, operator construction, and optical condition generation

3.2. Compact optical descriptors and confidence

Complete response kernels can contain many sensor samples, making their direct attachment to every image token computationally expensive. The proposed representation combines a low-dimensional response basis with physically interpretable descriptors. Kernel normalization separates response shape from measured throughput, which is retained in an independent channel. The response basis is fitted exclusively to the training calibration partition. The nominal descriptor retains sixteen coefficients for the aligned residual response shape. This is a representation setting rather than a universal optimum; its adequacy is evaluated through held-out response error and the reduced-basis comparison. Additional channels encode centroid displacement, second-moment spread, anisotropy, local throughput, normalized coordinates, and interpolation confidence. These quantities characterize the optical field and can be reused across measurements acquired under a fixed calibration. Their conversion into spatial condition features is defined by the descriptor encoding equation:

$$z(p) = E_\theta([a(p), \mu(p), s(p), t(p), p, c(p)]) \tag{2}$$

The encoder E_θ maps the concatenated descriptors at location p to a condition vector $z(p)$. The vector $a(p)$ contains interpolated basis coefficients, $\mu(p)$ describes centroid displacement, $s(p)$ represents response spread and anisotropy, $t(p)$ retains throughput, and $c(p)$ expresses calibration confidence. Coordinate channels use a common scale across training and evaluation. Descriptor standardization relies on training calibration statistics, with the resulting transformation retained for held-out conditions. Throughput is preserved during normalization to maintain the correspondence between object intensity and sensor counts. Shape normalization and radiometric normalization therefore have separate roles within the representation.

Interpolation confidence is derived from the normalized distance to calibration sites and disagreement between neighboring response predictions. The bounded confidence value is a heuristic reliability descriptor derived from calibration geometry and response disagreement. Predictive uncertainty for reconstructed pixels requires an independently validated calibration procedure. Repeated calibration exposures can provide an additional component describing measurement variability. Confidence evaluation includes independent perturbations to examine sensitivity to errors in the distance-based heuristic. Rapid local changes in the response field require particular attention because nearby calibration sites may still provide an incomplete description of the local optics. Held-out response residuals complement the confidence map by providing a direct assessment of interpolation accuracy across the field.

3.3. Operator approximation and spatial alignment

The interpolated operator is evaluated before training the reconstruction network. Held-out point-source captures are compared with responses predicted at their locations, separating optical representation error from learned inversion error. Relative response error and centroid displacement are recorded by field radius and direction. Retaining site-specific measurements exposes localized interpolation failures that can be obscured by a field average. Response support is also controlled consistently across calibration sites. Excessively narrow kernel truncation can suppress broad peripheral contributions and alter the difficulty of the inverse problem. The relative error equation provides a normalized measure of response approximation:

$$r(p) = \frac{\|h(p) - \hat{h}(p)\|_2}{\|h(p)\|_2 + \varepsilon} \tag{3}$$

The residual $r(p)$ compares the held-out response $h(p)$ with its interpolated estimate $\hat{h}(p)$. A small positive stabilizer ε ensures numerical stability for weak responses. Interpretation requires consistent acquisition correction, spatial support, and normalization for both responses. Shape error and throughput discrepancy are recorded separately to distinguish spatial response fidelity from radiometric accuracy. These calibration quantities guide response-basis selection before image reconstruction scores is examined. This ordering limits repeated adjustment of the optical representation against the final image test partition.

Spatial alignment is maintained through identical resampling conventions in the forward operator and condition encoder. A calibration manifest records pixel centers, response origins, crop offsets, and padding extents. Downsampled condition maps use averaging for smoothly varying scalar descriptors, while coordinate channels are regenerated at the appropriate grid centers. Optical quantities and measurements share the same coordinate convention. Scene augmentation is applied before measurement generation, preserving the original device geometry when the transformed scene is projected onto the sensor. This procedure maintains consistency between scene content and optical encoding. The representation assumes a static scene within a calibrated object plane and fixed optical geometry. Changes in object depth, wavelength response, or mask-to-sensor spacing can alter the encoding and require an extended or updated response field.

4. Conditioned Swin Reconstruction Model

4.1. Hierarchical reconstruction with optical modulation

The image branch receives a normalized adjoint backprojection together with a resampled sensor measurement. Backprojection provides an object-aligned initialization that retains residual optical mixing. A shallow convolutional stem maps the combined inputs into 48 feature channels. Three encoder scales use channel widths of 48, 96, and 192, with paired regular and shifted attention windows at each scale. A symmetric decoder combines higher-resolution skip features with upsampled coarse features. The nominal attention window spans eight tokens, with a shift of four tokens between paired blocks. Window masks control interactions across padded boundaries, and positional treatment remains consistent throughout the encoder and decoder. A residual head predicts an additive correction to the optical initialization.

Optical information enters the image branch through spatial feature modulation. At each scale, the condition encoder generates feature-dependent scale and offset maps. The modulation equation takes the form:

$$\tilde{F}_\ell = [1 + \gamma_\ell(z_\ell)] \odot LN(F_\ell) + \beta_\ell(z_\ell) \tag{4}$$

At scale ℓ, the condition map z_ℓ generates the scale map $\gamma_\ell(z_\ell)$ and offset map $\beta_\ell(z_\ell)$, which act on the normalized image features F_ℓ. Both modulation heads are initialized near zero, allowing optimization to begin close to the behavior of an unconditioned branch. Bounded scale values control amplification of uncertain optical descriptors. Layer normalization acts across the feature channels at each token location. The scale and offset maps retain their spatial indexing, allowing optical modulation to vary between locations within the same attention window. Different locations within the same attention window can consequently receive distinct optical conditions. A spatially averaged condition provides a matched ablation in which modulation capacity remains fixed while local optical specificity is reduced.

Figure 2 presents the reconstruction architecture, including measurement preprocessing, adjoint initialization, optical encoding, conditioned attention scales, decoder connections, and measurement correction. Distinct arrow styles separate image features, condition signals, and physical consistency dependencies. The calibrated operator supplies the forward and adjoint transformations used by the correction pathway. The learned residual is combined with the initialization before the final admissibility constraint is applied.

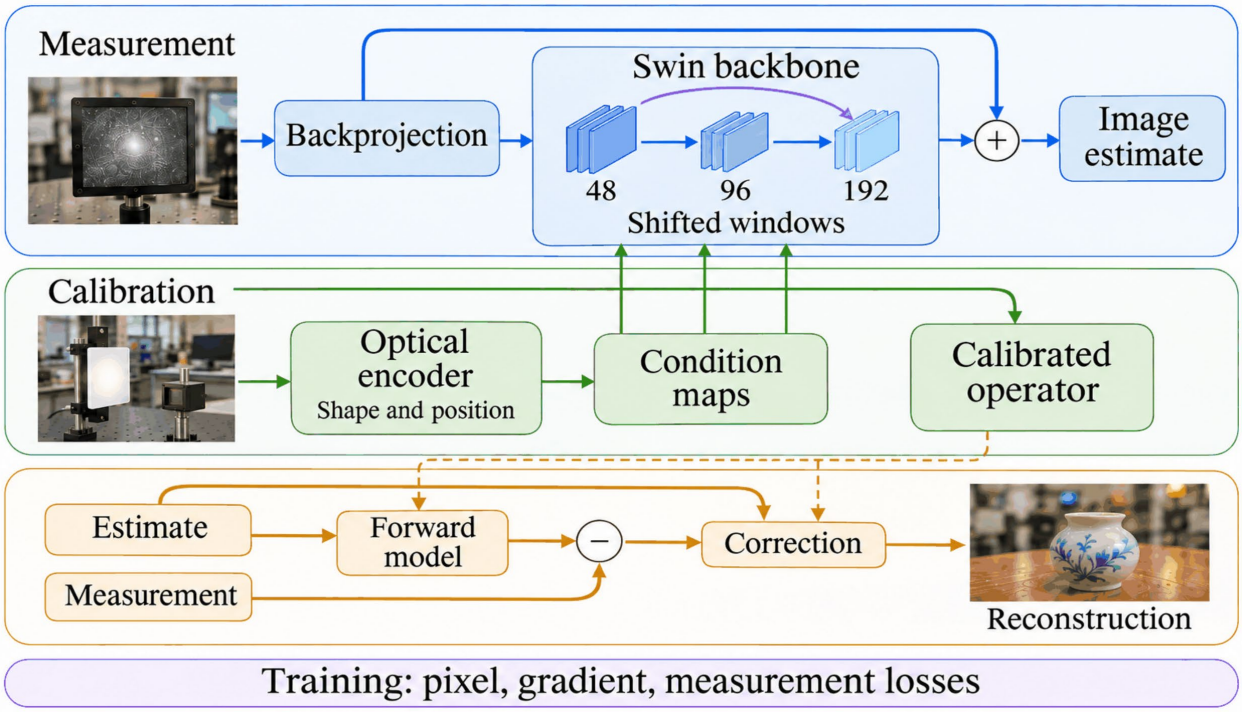


Figure 2. Architecture of the spatial response conditioned Swin reconstruction model with hierarchical modulation and measurement correction

4.2. Measurement correction and learning objective

The network prediction undergoes one differentiable correction using the calibrated forward operator. This step reduces measurement disagreement through an explicit physical projection. A bounded nonnegative step parameter is learned, with its upper limit determined from an estimate of the operator norm. Projection enforces nonnegative radiance. Upper clipping is applied separately when required by the normalized image representation, while measurement

residuals are evaluated using a consistent radiometric convention. Correction operates over the complete image to preserve long-range sensor coupling. The reconstruction update is written as the following equation:

$$\hat{x} = \Pi_{+}[u - \eta A^{\top}(Au - y)] \tag{5}$$

The intermediate estimate u is the output of the residual reconstruction branch, A denotes the calibrated forward operator, $A^{\top}$ its adjoint, and η the constrained correction step. The projection Π_{+}enforces nonnegativity. The update is a single projected gradient correction for measurement disagreement. It encourages consistency with the supplied operator; the remaining residual depends on the step size, operator accuracy, and learned intermediate estimate. Calibration error can bias the correction, while recoverable scene information remains constrained by the operator's observable subspace. Correction strength is therefore evaluated jointly with calibration mismatch. Gradients propagate through the physical update during training, allowing the image branch to adapt its prediction to the subsequent correction.

Training combines image fidelity, structural detail, and sensor agreement. These complementary requirements are incorporated into the objective equation:

$$\mathcal{L} = \mathcal{L}_{pix} + 0.1\mathcal{L}_{grad} + 0.05\mathcal{L}_{meas} \tag{6}$$

The pixel term measures robust absolute deviation from the reconstruction target. The gradient term compares horizontal and vertical image derivatives, encouraging preservation of boundaries and fine structures. The measurement term evaluates disagreement between the reprojected estimate and the corresponding sensor reference, using noise-aware weighting derived from the sensor model. The coefficients are fixed before evaluation under the nominal training protocol. Ablations retain the same objective unless a corresponding component is explicitly removed. Measurement residuals and image-domain metrics are reported separately to distinguish sensor agreement from scene recovery.

4.3. Training controls and implementation protocol

Training uses 200 epochs, a batch size of eight full-resolution samples, and an adaptive optimizer. The initial learning rate is 0.0002, followed by smooth decay after a short warmup period. Weight decay is set to 0.01, and gradient clipping is used to limit occasional large updates from strongly multiplexed measurements. Three independent random seeds are used to vary the optimization process of each trained model. Validation selects checkpoints; then, based on these checkpoints, the test partition is assessed. Baselines share the scene partition, forward model, noise distribution and backprojection input at their interfaces. Baseline-specific initialization is described with the reconstruction method.

Nominal reconstruction resolution is 256x256 pixels. Pre-measurement scene transformations are used to maintain consistency of image content and optical encoding. Optical perturbations have bounded response displacement, throughput fluctuations and alterations of the response-basis coefficients. Sampling of perturbations is independent of image noise and scene identity. The above separation prevents the inference of calibration state by means of correlations with unrelated scene or noise features in the network. The range of perturbations is based on the calibration repeatability estimates and is a reasonable acquisition change. Nominal calibration conditions and various degrees of controlled operator mismatch are included in the evaluation.

Implementation records list kernel support, calibration sites, transform padding, crop offsets, numerical precision, optimizer state, random seed and checkpoint hash. Warm-up and synchronized-device execution are employed for runtime evaluation at a fixed batch size and reconstruction resolution. Optical encoders and modulation heads are in the scope of parameter counting. Operator storage, cached condition features, and activation memory are reported separately to expose the contribution of calibration data and intermediate features to peak deployment memory. Shared calibration manifests and a uniform profiling environment enable reproducible comparisons of model configurations.

5. Experimental Evaluation and Analysis

5.1. Evaluation design and comparative reconstruction

Evaluation of reconstruction fidelity, calibration sensitivity and computational load is conducted in a shared-comparison protocol. There are eight thousand training scenes, one thousand validation scenes and one thousand held-out test scenes in the proposed experiment. Scene identities are disjoint across the partitions, and nearly duplicate crops from the same source scene are excluded. The six content categories are edges, textures, text, objects, low-contrast scenes and mixed-structure scenes. A separate calibration partition excludes certain positions from the interpolation fitting. Scene partitions test reconstruction across disjoint image identities, whereas held-out calibration sites test interpolation across optical field positions. These are separate generalization axes and should be reported independently when interpreting reconstruction and response-model errors. Recent lensless evaluation studies compare reconstruction architectures under fixed parameter budgets and changes in measurement discretization [19]. The data ledger stores all plotted series and their metric definitions to ensure consistency between numerical analysis and the figures. Peak signal-

to-noise ratio, structural similarity and normalized absolute error are used to assess image quality. Peak signal-to-noise ratio is calculated for each image in a fixed intensity range and then averaged; maintaining a consistent distinction between averaging image errors and averaging decibel values. Lensless light-field reconstruction studies use peak signal-to-noise ratio and structural similarity alongside angular-consistency assessments to characterize reconstruction quality [20]. A fixed radiometric scale preserves errors associated with incorrect throughput during evaluation. Evaluation of images is conducted after the same fixed-crop for all methods. Peripheral error is evaluated on a boundary region defined from distance to the image border, with central error evaluated on its complement. The stated outer-quarter description requires an explicit stored pixel mask or threshold for exact reproduction; comparisons must use the same mask and radiometric normalization for every method. The main table focuses on explicitly defined fidelity metrics to support reproducible comparisons.

Six reconstruction profiles are shown in Table 1 under the nominal condition. Proposed model has a value of 29.84 dB and 0.914 structural similarity; that of the unconditioned Swin baseline is 27.63 dB and 0.872. The corresponding differences are 2.21 dB and 0.042, respectively. Response-conditioned convolutional model is 28.31 dB and falls between the two transformer variants. The above arrangement aims to differentiate between the optical state and the image structure in the proposed analysis; conditioning aids in a convolutional family, and the highest-scoring profile integrates conditioning with hierarchical attention. Normalized error decreases from 0.074 for the unconditioned Swin model to 0.051 for the conditioned model, a relative reduction of approximately 31.1%. This complements the 2.21 dB PSNR difference and the 0.042 structural-similarity difference. Based on the above arrangement of comparisons, matched controls were employed for the optical condition and reconstruction branches. Stratification of reconstruction results according to field location and scene difficulty is also required.

Table 1. Nominal reconstruction profiles

Method	PSNR dB	SSIM	NMAE	Parameters M	Latency ms
TV inversion	23.40	0.741	0.112	0	93
Convolutional decoder	26.12	0.838	0.086	7.2	21
Unconditioned Swin	27.63	0.872	0.074	10.9	35
Conditioned convolution	28.31	0.889	0.067	8.1	29
Learned unrolled	28.92	0.901	0.061	6.4	68
Proposed conditioned Swin	29.84	0.914	0.051	11.8	43

The comparison in Figure 3(a) spans 23.40 to 29.84 dB across six methods, with the conditioned convolutional profile exceeding unconditioned Swin by 0.68 dB and the proposed profile exceeding conditioned convolution by 1.53 dB. This ordering separates the potential value of optical information from the additional value associated with the reconstruction branch; matched forward operators and initialization remain necessary for a causal comparison. Figure 3(b) shows content medians of 31.0, 29.8, 28.1, 30.3, 28.7, and 29.5 dB. The 2.9 dB separation between edges and text identifies fine symbolic structures as a demanding case even when broad boundaries are preserved. Distribution tails motivate scene level inspection because a favorable median can coexist with substantial failures on thin strokes. Figure 3(c) places the proposed profile at 29.84 dB and 0.914 structural similarity, whereas unconditioned Swin reaches 27.63 dB and 0.872. Agreement between these two indicators supports the internal consistency of the ranking, while the 28 dB vertical reference serves only as a descriptive benchmark. Figure 3(d) ranges from 0.040 at the center to 0.089 at peripheral sites, making the largest local error more than twice the central value. The map represents a separate diagnostic field condition with stronger boundary truncation than the nominal comparison. Its local values quantify spatial vulnerability under that condition. Calibration site residuals and directional edge tests would help distinguish response interpolation error from information loss at the detector boundary.

Pair-wise statistical tests would be needed for scene-by-scene comparisons as all methods use the same set of held-out data for reconstruction. In diagnostic-accuracy studies, pairs-bootstrap procedures have been shown to support error control when multiple candidate tests share an evaluation sample, providing a cross-domain methodological precedent for paired resampling [21]. Adapting this rationale to continuous image-quality scores requires a separate choice and justification of the resampling statistic. Repeated-training seeds are used to show optimization variability; scene diversity is employed for reconstruction difficulty variation. Variability indicators supplement the mean scores to display the spread of data for different comparison profiles. The reported comparisons are descriptive because scene-level paired differences, seed-level estimates, and confidence intervals are not provided. Statistical significance therefore remains unestablished by the present reporting. A paired-difference distribution of the finished study would show whether the average gain is distributed widely or due to a small number of difficult peripheral scenes. Lower-tail reporting of reconstruction quality is especially beneficial when failures in text or thin structures are more critical than the average visual quality. The six comparison methods are total variation inversion, a convolutional encoder-decoder,

unconditioned Swin, response-conditioned convolution, a learned unrolled model and the proposed conditioned Swin. Iterative lensless reconstruction can use measurement-dependent nonlinear weights to reduce the influence of noise and systematic errors during phase recovery [22]. Convolutional- and transformer-baselines have similar training opportunities and explicit parameter budgets; computational effort has been recorded separately. The unrolled profile also requires reporting the number of forward-and-adjoint evaluations per reconstruction. Optical operator call counts may increase the latency of models with a small number of trainable parameters.

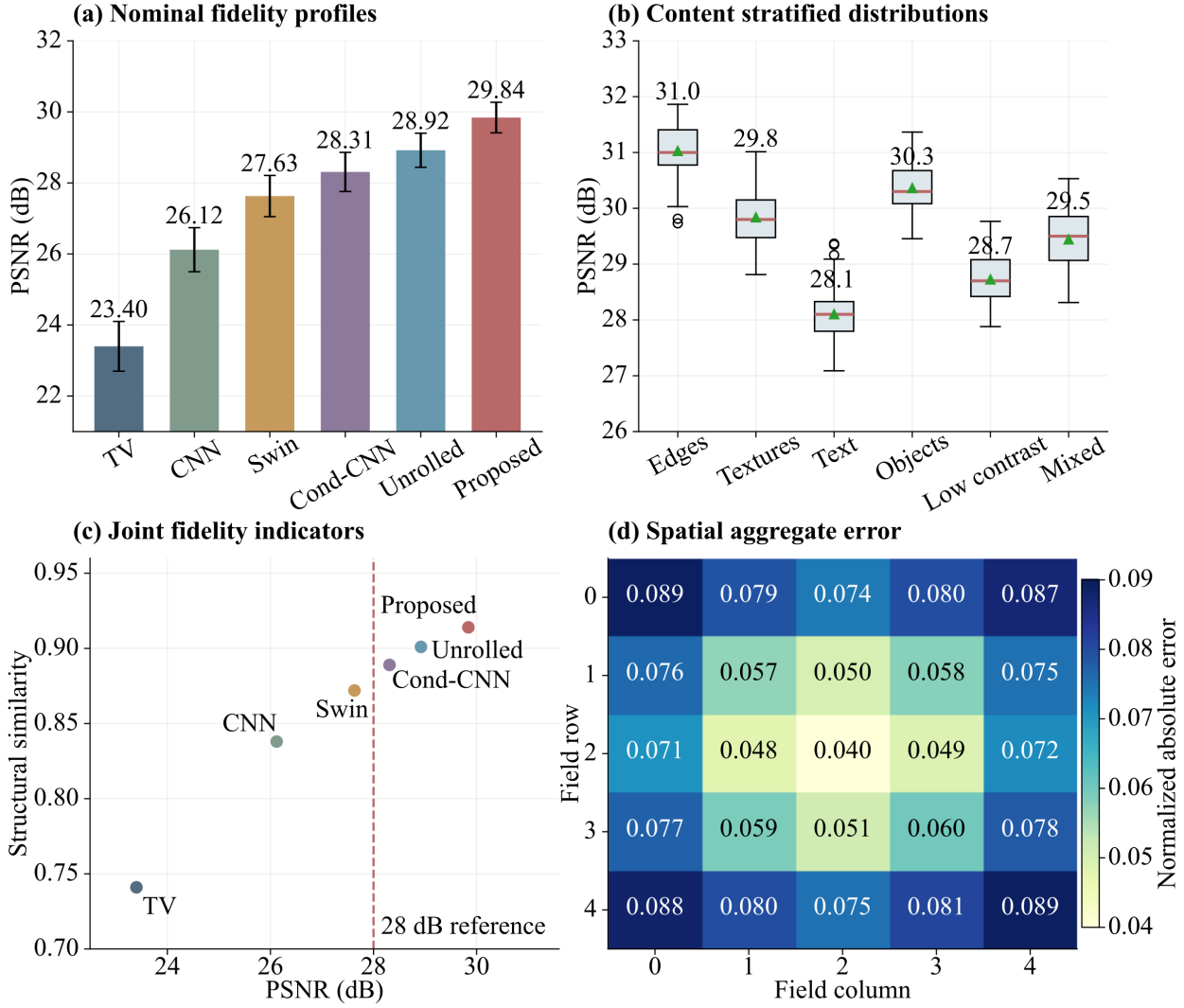


Figure 3. Reconstruction comparison with four panels: (a) method fidelity with variability, (b) content stratified distributions, (c) fidelity and structural similarity scatter, and (d) spatial error heatmap

The proposed nominal-error reduction is spatially non-uniform in construction. Peripheral-normalized error drops from 0.091 in the unconditioned Swin profile to 0.059 in the conditioned profile; that is, it has been reduced by about 35%. The smaller central improvement is consistent with the hypothesis that spatial response information is most useful where a global optical approximation is weakest. Testing this mechanism requires a matched global-operator control. Therefore, the finished comparison needs to have both a superior forward model without conditioning and conditioning using the same forward model. These controls separate the contribution of calibration accuracy from that of the modulation network. The same problem occurs with backprojection quality; an enhanced initialization can be passed on at all subsequent steps even if attention weights are unaltered. Inspection of measurement residuals and reconstructed images should be performed in practical failure analysis. Accuracy of text strokes needs to align with sensor data; over-smoothing may also be due to information deficiencies. Recent reference-free lensless quality assessment combines residual structure, reconstruction statistics, and local PSF sensitivity to assess reconstruction risk under imperfect optical models [23]. Residual maps in this study will be organized according to sensor radius, intensity and scene edge density. Masking of saturated pixels and defective sensor areas would be performed according to rules set independently of the reconstruction result. This independent masking rule preserves a consistent residual evaluation domain across methods. Bright and low-signal scenes should be evaluated separately due to the possibility of the same normalized image error resulting from different photon- and read-noise regimes.

5.2. Component contribution and calibration robustness

Component analysis uses the same training partition and optical generator as the nominal comparison. Table 2 defines six configurations: the complete model, removal of the optical condition, spatially averaged condition, removal of the consistency correction, shuffled condition maps, and a reduced response basis.

Table 2. Matched component configurations

Configuration	PSNR dB	SSIM	Peripheral NMAE
Complete model	29.84	0.914	0.059
No condition	27.63	0.872	0.091
Spatially averaged condition	28.41	0.890	0.077
No correction	29.18	0.903	0.068
Shuffled condition	27.95	0.881	0.085
Reduced basis	29.37	0.907	0.064

The complete model reaches 29.84 dB. Removing the condition yields 27.63 dB, averaging it spatially yields 28.41 dB, removing correction yields 29.18 dB, shuffling it yields 27.95 dB, and reducing the basis yields 29.37 dB. The averaged condition retains the modulation parameters, making it a stronger control than deleting the branch entirely. Kabuli et al. [24] show that object sparsity, optical multiplexing, and noise jointly influence recoverable scene information, motivating explicit control of these factors in lensless comparisons. The shuffled condition provides a distinct test of whether local optical alignment matters beyond having additional channels.

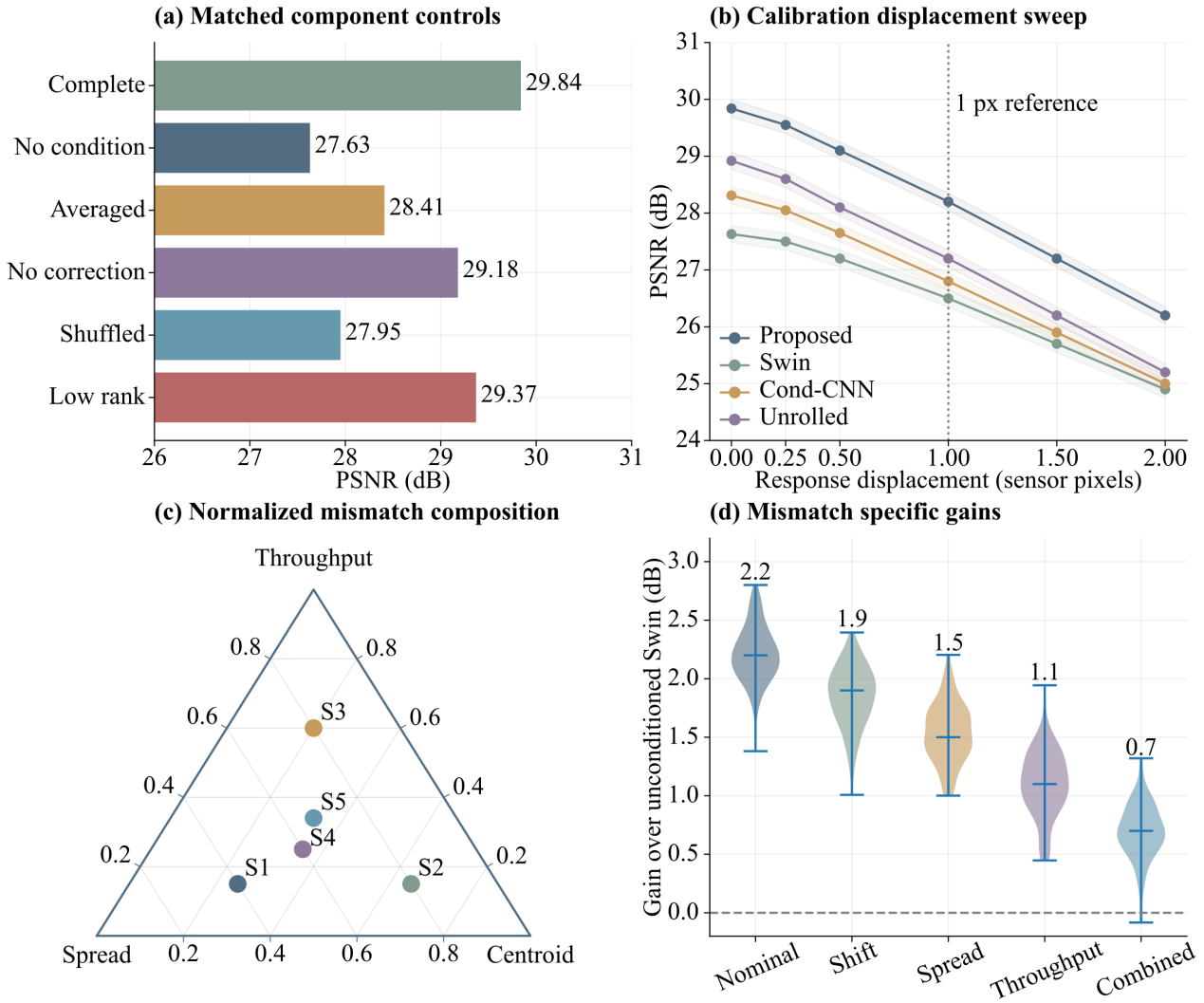


Figure 4. Component and robustness analysis with four panels: (a) matched ablations, (b) response displacement sensitivity, (c) normalized mismatch composition, and (d) gain distributions for five mismatch conditions

Figure 4(a) reports 29.84 dB for the complete model, 28.41 dB for spatially averaged conditions, and 27.95 dB for shuffled conditions. The 1.43 dB gap between complete and averaged conditioning isolates the hypothesized contribution of spatial specificity more closely than deleting the branch, since modulation capacity remains present. Removing physical correction changes the score from 29.84 to 29.18 dB, a difference of 0.66 dB. This is smaller than the 1.43 dB difference associated with spatially averaging the condition, although interactions between components prevent treating these differences as additive contributions. Figure 4(b) follows six displacement levels from 0 to 2.00 pixels; the proposed profile falls from 29.84 to 26.20 dB, whereas unconditioned Swin falls from 27.63 to 24.90 dB. The proposed margin contracts from 2.21 to 1.30 dB, linking greater nominal fidelity to greater absolute sensitivity to calibration displacement. Figure 4(c) presents five normalized mismatch compositions. The spread dominated point allocates 0.60 to spread, 0.25 to centroid, and 0.15 to throughput, while the throughput dominated point allocates 0.20, 0.20, and 0.60. Comparing these compositions helps select targeted perturbations, provided absolute mismatch magnitudes are retained separately. Figure 4(d) reports median gains of 2.2, 1.9, 1.5, 1.1, and 0.7 dB across nominal, shift, spread, throughput, and combined conditions. The lower combined-condition median identifies simultaneous optical errors as a more stringent test than isolated perturbations. Paired scene differences provide the appropriate basis for examining the frequency and severity of negative gains.

Calibration perturbations must be specified in physical or directly measurable coordinates. A centroid displacement is expressed in sensor pixels, throughput variation in relative intensity, and response spread variation in a stated moment-based quantity. Model-driven lensless reconstruction has addressed model mismatch and defocus by coupling forward propagation with learned reconstruction through cycle consistency [25]. These mechanisms can produce similar image artifacts while requiring different remedies. Training for response displacement requires explicit operator perturbations alongside measurement noise augmentation. The evaluation therefore samples each perturbation independently before considering combinations. Combined perturbations form a separate test condition with the nominal baseline retained for comparison. The mismatch sweep holds unseen scene content fixed to isolate the influence of optical perturbations. The displacement sweep contains six levels: 0, 0.25, 0.50, 1.00, 1.50, and 2.00 pixels. The proposed profile is 29.84, 29.55, 29.10, 28.20, 27.20, and 26.20 dB. The unconditioned Swin profile is 27.63, 27.50, 27.20, 26.50, 25.70, and 24.90 dB. Across the displacement sweep, the conditioned profile loses 3.64 dB and the unconditioned profile loses 2.73 dB. The conditioned profile retains higher absolute quality, while its advantage contracts from 2.21 to 1.30 dB; absolute fidelity and sensitivity to mismatch therefore lead to different robustness interpretations. A robustness claim must state whether it concerns absolute quality at a given mismatch or sensitivity relative to nominal quality. These criteria quantify complementary aspects of calibration sensitivity. An unconditioned model can be less sensitive simply because it starts from a lower quality level and makes less use of calibration information.

Operator mismatch alters the impact of the measurement correction as well. It is only suitable for reducing the stated measurement discrepancy that is also consistent with recovering the original scene. Recent measurement-consistent lensless methods project learned reconstructions onto the observation-model constraint, illustrating the role of explicit physical consistency in reconstruction design [26]. Correction strength should be varied in the validation set of a practical experiment, and then the same chosen rule should be employed on the held-out mismatches. Separate oracle selections of the above for each test case would inflate the deployability robustness. Confidence descriptor modulates learned processing and operator correction necessitates re-calibration. Large calibration discrepancies need to be flagged for quality and have a specified recalibration process. Evaluation of response-basis rank uses both reconstruction quality and held-out calibration residual. Nominal sixteen-coefficient basis is compared to four, eight, twelve and twenty-four coefficient alternatives. Low-rank representations may omit weak but significant peripheral structures, and high-rank representations may retain calibration noise and increase the condition encoder's cost. The reduced-basis score is 29.37 dB, indicating a modest quality reduction and motivating validation of the selected basis rank. Selection of basis should be performed in a calibration-validation partition and confirmed with scene reconstruction only after fixing the optical representation. Separation is advantageous in cases of limited calibration captures due to the risk of overfitting to a specific field geometry via repeated tuning on the image test set.

Spatially shuffled conditions are generated by permuting response descriptors among calibration regions while retaining their marginal values. This test preserves much of the descriptor distribution and the condition branch capacity but destroys the mapping between location and optics. The score of 27.95 dB remains above the no condition profile, which could occur if global response statistics still help reconstruction. The difference between correctly aligned and shuffled conditions quantifies the contribution associated with local optical alignment. Further controls should shift coordinate channels independently of response descriptors and replace response descriptors with coordinates alone. These controls distinguish optical content, spatial indexing, and additional trainable capacity without changing the primary test population.

5.3. Noise behavior and computational tradeoffs

Noise evaluation uses six nominal measurement signal-to-noise ratios of 10, 15, 20, 25, 30, and 35 dB. These input levels characterize sensor measurements, whereas reconstructed-image PSNR measures output fidelity against the image

reference. AA completed experiment would derive the measurement noise model from dark frames and repeated exposures, with photon noise expressed in a physically meaningful count scale. Studies of camera noise in lensless random-phase encoding analyze shot noise, read noise, and quantization noise as distinct sources of information degradation [27]. The sweep assigns the proposed model 24.1, 26.2, 28.0, 29.1, 29.84, and 30.2 dB of reconstruction fidelity. This operating trajectory applies to a fixed scene population and exposure normalization. Low illumination can alter both image noise and calibration reliability, which is why a real evaluation should include low signal response captures instead of perturbing only the scene measurements.

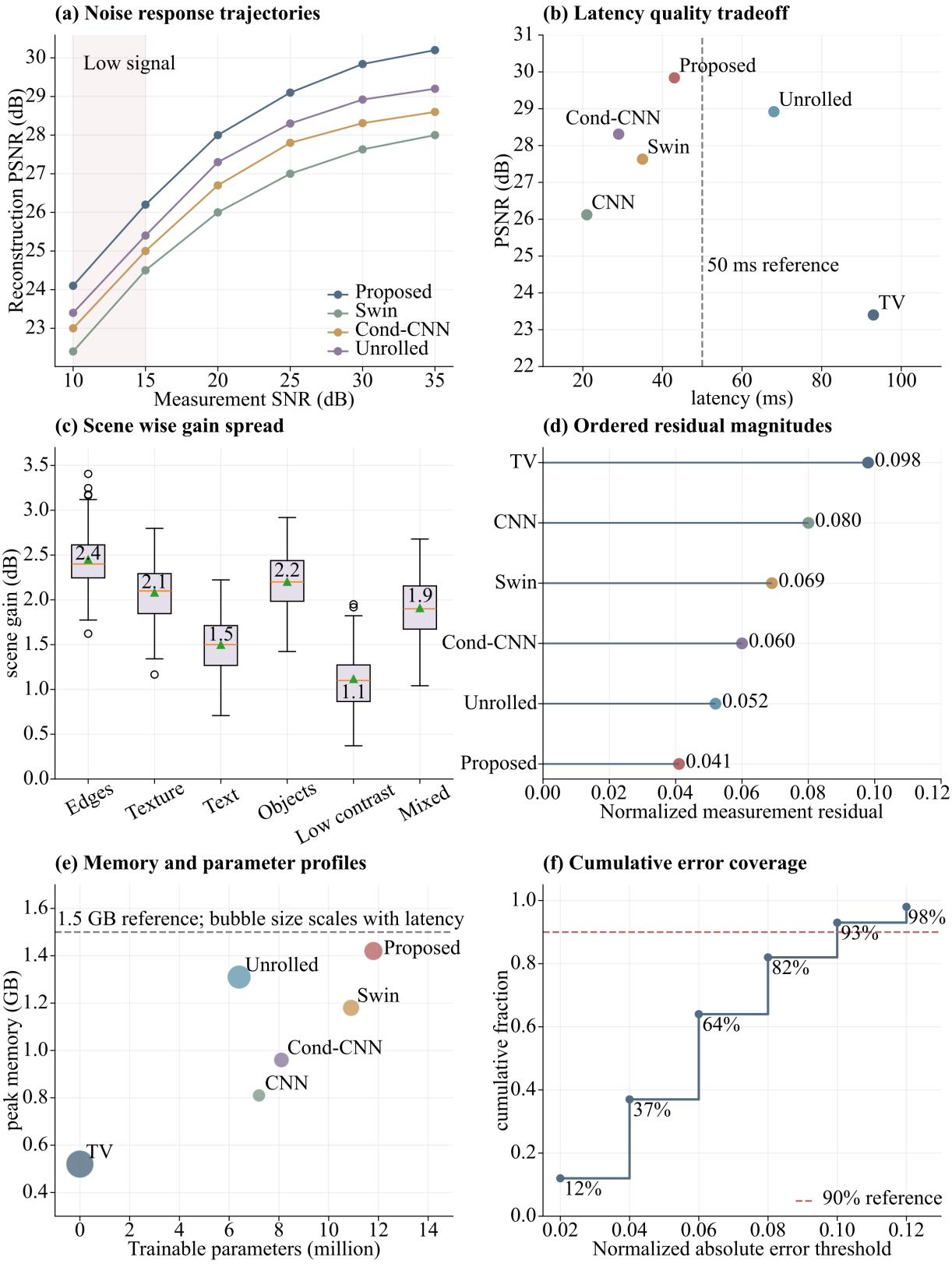


Figure 5. Noise and implementation analysis with six panels: (a) noise response curves, (b) latency quality tradeoff, (c) scene gain distributions, (d) ordered residual magnitudes, (e) parameter memory scatter, and (f) cumulative error coverage

Figure 5(a) reports for the proposed profile 24.1, 26.2, 28.0, 29.1, 29.84, and 30.2 dB across six measurement noise levels. The 2.1 dB improvement from the first to second level exceeds the 0.36 dB improvement between the final two levels, indicating a transition from noise limited behavior toward an optical or prior limited regime. Figure 5(b) associates the proposed operating point with 43 ms and 29.84 dB, compared with 35 ms and 27.63 dB for unconditioned Swin. The extra 8 ms accompanies a 2.21 dB fidelity increase, while the unrolled profile requires 68 ms at 28.92 dB. These coordinates motivate evaluating the number of operators calls alongside attention cost. Figure 5(c) shows scene gain medians of 2.4, 2.1, 1.5, 2.2, 1.1, and 1.9 dB, with low contrast scenes receiving the smallest benefit. Their response may depend on throughput accuracy and weak signal regularization more strongly than on local feature exchange. Figure 5(d) orders residual magnitudes from 0.098 to 0.041; the proposed value is approximately 41 percent below the unconditioned Swin value of 0.069. Residual reduction must be compared with edge fidelity to detect smoothing along weakly observed scene directions. Figure 5(e) places the proposed profile at 11.8 million parameters and 1.42 GB, versus 10.9 million and 1.18 GB for unconditioned Swin. The 0.24 GB increase includes condition storage and activation effects, exposing a deployment cost beyond weight count. Figure 5(f) reports cumulative coverage of 12, 37, 64, 82, 93, and 98 percent at error thresholds of 0.02 through 0.12. The rise from 82 to 93 percent across the 0.08 to 0.10 interval identifies a concentrated error tail that warrants inspection of individual scenes.

Latency and reconstruction quality are evaluated jointly. Depth-aware lensless reconstruction evaluates image fidelity alongside depth error and model parameter count, illustrating the need to assess reconstruction quality and computational resources jointly [28]. The medium proposed profile reports 43 ms at the nominal resolution and 11.8 million parameters. Hardware identity, numerical precision, batch size, synchronization, and inclusion of transfer or preprocessing must accompany latency for a reproducible deployment comparison; the current value alone supports a limited implementation comparison. A lighter-profile option would reduce channels or depth; a heavier-profile option could add attention blocks without altering the conditioning mechanism. The above architectural modifications need to be assessed at the same precision and batch size. Calibration loading and one-time condition caching are separate from the per-frame reconstruction latency. Frequent calibration modifications change the efficiency of caching and thus require separate profiling for startup and steady-state latency. Residual ordering follows the fidelity ranking, while the two metrics describe distinct reconstruction properties. A reconstruction can agree with sensor data while retaining errors in weakly observed scene components. Conversely, a small calibration error may result in a noticeable residual even if the reconstructed object is near the target. Fisher-information analysis of structured scattering links the information carried by lensless measurements to the recoverability of scene features [29]. Therefore, a fixed measurement normalization is employed for the residual panel and it is presented with the image-domain indicators. Further evaluation should also investigate whether reduced residuals are due to over-smoothing, intensity bias or an improper crop convention. The above implementation- and modelling-factors need specific residual- and spatial-fidelity tests.

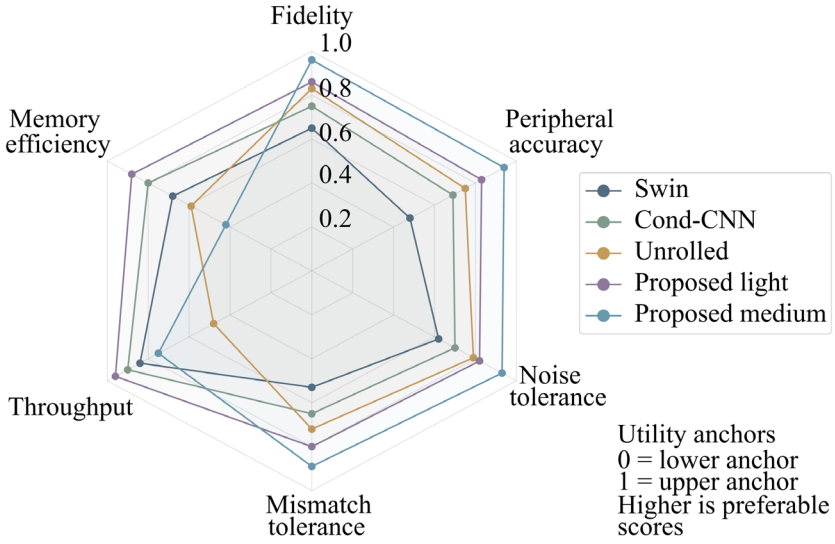


Figure 6. Reconstruction performance and efficiency

Figure 6 compares five configurations across six utility axes spanning 0.42 to 0.96. The medium proposed profile has fidelity and peripheral accuracy scores of 0.96 and 0.94, while the light profile has throughput and memory efficiency scores of 0.96 and 0.88. The medium profile's memory score of 0.42 makes resource demand its limiting criterion. These profiles are dimensionless engineering utility scores whose interpretation depends on the chosen normalization anchors. Individual axes should be read separately; polygon area and an unreported composite score provide no independent basis for ranking configurations. Their interpretation depends on the selected application utility, so configuration choice requires the underlying fidelity, latency, and memory values alongside each normalized axis. Application thresholds follow task requirements established before reconstruction evaluation. Joint learning of lensless reconstruction and facial

recognition demonstrates that reconstructed images can be evaluated in relation to a specific downstream identification task [30]. For text inspection, a small displaced stroke can matter more than a broad low contrast bias. For geometric sensing, stable feature position may be more important than visually pleasing texture. The evaluation protocol reserves an independent labeled partition for downstream task assessment. Task accuracy evaluation requires a specified classifier, detector, or measurement procedure with independently validated labels. The fidelity threshold provides a graphical reference; application acceptance requires a task specific criterion.

Transfer to a second device is a different experiment than calibration perturbation of the first device. Modular learned lensless reconstruction has been investigated as a means of improving robustness and generalization across changes in optical acquisition [31]. A strict transfer test would provide the new response field to the condition branch, freeze reconstruction weights and avoid target device fine-tuning. A second condition may allow for a small calibration-only adaptation step; a third one can use labelled scene pairs. These settings define distinct transfer-evaluation protocols. Their description specifies how future validation should separate calibration-only adaptation from target-scene supervision; completed cross-device performance is not established by the present comparisons. Extensive target-scene training and calibration only transfer are different adaptation modes. Modifications of the spectral response, depth and sensor non-linearity necessitate explicit modifications to the optical model. Memory reports contain weights, activations, response kernels, interpolation maps and cached condition tensors. The medium configuration occupies 1.42 GB in the comparison ledger. Storage of the entire response may be excessive for a low-parameter image network, especially when many depth planes or wavelength bands are added. Compression should therefore be assessed jointly with calibration residual and reconstruction quality. Feasibility of embedded systems relies on the joint support of weight storage, activation memory and optical operator implementation. Actual deployment requires measured peak memory, supported numerical precision, operator implementation details and a policy for handling calibration changes during system operation.

6. Conclusion

This study develops a spatially varying point spread function conditioned Swin Transformer for lensless image reconstruction, with the reported comparisons indicating improved image fidelity and peripheral reconstruction under the stated evaluation settings. Optical feature modulation and measurement correction provide complementary mechanisms for incorporating calibration information.

1. The conditioned model achieves a peak signal-to-noise ratio of 29.84 dB and structural similarity of 0.914, exceeding the unconditioned Swin baseline by 2.21 dB and 0.042, respectively. Peripheral normalized absolute error decreases from 0.091 to 0.059, corresponding to a reduction of approximately 35.2%.

2. Reconstruction quality depends on the spatial specificity of the optical condition. Replacing local descriptors with spatially averaged conditions reduces fidelity from 29.84 to 28.41 dB, while removing measurement correction reduces it to 29.18 dB. These differences indicate complementary contributions from spatial optical adaptation and physical consistency. The conditioned configuration requires 43 ms per reconstruction, compared with 35 ms for the unconditioned baseline.

3. Calibration accuracy remains an important operating constraint. Increasing response displacement from zero to 2.00 sensor pixels reduces fidelity from 29.84 to 26.20 dB, although the conditioned model retains a 1.30 dB advantage over the unconditioned baseline. The current formulation assumes approximately planar scenes and a calibrated response field. Depth-dependent encoding and peripheral response truncation remain important limits when interpreting these comparisons and extending the model to new acquisition geometries.

Future work should extend the response representation across depth and acquisition conditions while improving calibration efficiency and reducing reconstruction memory requirements.

7. Declarations

7.1 Author Contributions

Hsin-yu Yu contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Ya-chieh Mo contributes to validation, analysis, investigation, data collection, draft preparation. Kuang-chun Kuo contributes to conceptualization, validation, analysis. Hsiao-wei Chiou contributes to validation, analysis, investigation, data collection. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Date Availability

Not applicable.

7.6 Declaration of Competing Interest

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

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Zhao, Y., Xu, F., Liu, J., & Guo, C. (2025). Portable lensfree single-frame imaging sensor using random phase modulation. Optics Express, 33(26), 55054–55067.https://doi.org/10.1364/oe.582880

[2] Li, S., Gao, Y., Wu, J., Wang, M., Huang, Z., Chen, S., & Cao, L. (2025). Lensless camera: Unraveling the breakthroughs and prospects. Fundamental Research, 5(4), 1725–1736.https://doi.org/10.1016/j.fmre.2024.03.019

[3] Lin, Y., Li, G., Yan, D., Zhang, W., Zhu, J., & Zhao, Y. (2025). Lensless diffraction imaging using a rotated slit modulation. Optics Continuum, 4(2), 421–437.https://doi.org/10.1364/optcon.554179

[4] Chien, T., Cao, R., Liu, F. L., Kabuli, L. A., & Waller, L. (2024). Space-time reconstruction for lensless imaging using implicit neural representations. Optics Express, 32(20), 35725–35732.https://doi.org/10.1364/oe.530480

[5] Feshki, M., Godin, A. G., De Koninck, Y., & Gosselin, B. (2026). Learned spatially varying microscopy model with adaptive point spread functions. Optics Express, 34(7), 12287–12303.https://doi.org/10.1364/oe.586982

[6] Chen, K., Li, Y., Li, Z., Hu, J., Zhao, Q., & Guo, Y. (2026). High-quality reconstruction method for miniaturized lensless cameras based on imaging model optimization. Optics Express, 34(15), 26909–26925.https://doi.org/10.1364/oe.595581

[7] Qian, H., Ling, H., & Lu, X. (2024). Robust unrolled network for lensless imaging with enhanced resistance to model mismatch and noise. Optics Express, 32(17), 30267–30283.https://doi.org/10.1364/oe.531694

[8] Liang, J., Cao, J., Fan, Y., Zhang, K., Ranjan, R., Li, Y., Timofte, R., & Van Gool, L. (2024). VRT: A video restoration transformer. IEEE Transactions on Image Processing, 33, 2171–2182.https://doi.org/10.1109/tip.2024.3372454

[9] Kang, X., Duan, P., Li, J., & Li, S. (2024). Efficient Swin Transformer for remote sensing image super-resolution. IEEE Transactions on Image Processing, 33, 6367–6379.https://doi.org/10.1109/tip.2024.3489228

[10] Jiang, R., & Han, Y. (2024). Generalizing to out-of-sample degradations via model reprogramming. IEEE Transactions on Image Processing, 33, 2783–2794.https://doi.org/10.1109/tip.2024.3378181

[11] Wang, Z., Zheng, S., Ding, Z., & Guo, C. (2024). Dual-constrained physics-enhanced untrained neural network for lensless imaging. Journal of the Optical Society of America A, 41(2), 165–173.https://doi.org/10.1364/josaa.510147

[12] Nakamura, T., Kato, R., Iwata, K., Makihara, Y., & Yagi, Y. (2024). Multilayer lensless camera for improving the condition number. Applied Optics, 63(28), G9–G17.https://doi.org/10.1364/ao.521126

[13] Banerjee, A., & Singh, S. (2026). Towards physics-informed cyclic adversarial multi-PSF lensless imaging. IEEE Transactions on Emerging Topics in Computational Intelligence, 10(1), 390–400.https://doi.org/10.1109/tetci.2025.3579385

[14] Ma, Z., Wang, Y., Tohidypour, H. R., Nasiopoulos, P., & Leung, V. C. M. (2024). Enhancing image quality by reducing compression artifacts using dynamic window Swin Transformer. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 14(2), 275–285.https://doi.org/10.1109/jetcas.2024.3392868

[15] Cui, Y., Ren, W., Cao, X., & Knoll, A. (2024). Revitalizing convolutional network for image restoration. IEEE Transactions on

Pattern Analysis and Machine Intelligence, 46(12), 9423–9438.https://doi.org/10.1109/tpami.2024.3419007

[16] Cui, Y., Zhu, J., & Knoll, A. (2025). Enhancing perception for autonomous vehicles: A multi-scale feature modulation network for image restoration. IEEE Transactions on Intelligent Transportation Systems, 26(4), 4621–4632.https://doi.org/10.1109/tits.2025.3538485

[17] Liu, Z., Zeng, T., Zhan, X., Zhang, X., & Lam, E. Y. (2025). Generative approach for lensless imaging in low-light conditions. Optics Express, 33(2), 3021–3039.https://doi.org/10.1364/oe.544875

[18] Ekmekci, C., & Cetin, M. (2026). Conformalized generative Bayesian imaging: An uncertainty quantification framework for computational imaging. IEEE Transactions on Computational Imaging, 12, 216–229.https://doi.org/10.1109/tci.2025.3649389

[19] Ekec, K., & Teğin, U. (2026). Resolution-agnostic lensless imaging via Fourier neural operators. Optics Express, 34(18), 34757–34772.https://doi.org/10.1364/oe.610853

[20] Mo, C., Liu, X., Tong, J., Xi, J., Yu, Y., & Cai, Z. (2024). Lensless light-field imaging using LMI. Optics Express, 32(22), 38112–38127.https://doi.org/10.1364/oe.539021

[21] Westphal, M., & Zapf, A. (2024). Statistical inference for diagnostic test accuracy studies with multiple comparisons. Statistical Methods in Medical Research, 33(4), 669–680.https://doi.org/10.1177/09622802241236933

[22] Lan, J., Xu, C., Pang, H., Yang, C., Zhao, L., Hu, S., & Yang, Y. (2025). Robust lensless computational imaging via nonlinear weighting approach. Applied Optics, 64(2), 423–430.https://doi.org/10.1364/ao.543766

[23] Li, C., Cao, J., Hao, Q., Song, C., Yao, H., & Wei, Z. (2026). Reference-free per-measurement assessment of lensless reconstruction quality under imperfect PSF models. Optics Express, 34(23), 44252–44267.https://doi.org/10.1364/oe.613012

[24] Kabuli, L. A., Pinkard, H., Markley, E., Hung, C. S., & Waller, L. (2026). Designing lensless imaging systems to maximize information capture. Optica, 13(2), 227–235.https://doi.org/10.1364/optica.570334

[25] Ni, C., Yang, C., Zhang, X., Li, Y., Zhang, W., Zhai, Y., He, W., & Chen, Q. (2024). Address model mismatch and defocus in FZA lensless imaging via model-driven CycleGAN. Optics Letters, 49(15), 4170–4173.https://doi.org/10.1364/ol.528502

[26] Fujiwara, S., Kawachi, H., Silva Neto, J. R. C. S. A. V., Yagi, Y., & Nakamura, T. (2026). Deep lensless imaging with non-iterative output-constraint satisfaction. Optics Express, 34(20), 37137–37151.https://doi.org/10.1364/oe.608276

[27] Goswami, S., Krishnan, G., & Javidi, B. (2024). Robustness of single random phase encoding lensless imaging with camera noise. Optics Express, 32(4), 4916–4930.https://doi.org/10.1364/oe.510950

[28] Li, Y., Li, Z., Chen, K., Guo, Y., & Rao, C. (2024). 3D reconstruction from focus for lensless imaging. Applied Optics, 63(31), 8212–8220.https://doi.org/10.1364/ao.540257

[29] Grant-Jacob, J. A., & Mills, B. (2026). Enhancing lensless imaging through Fisher-information-engineered structured scattering. Optics Continuum, 5(7), 2249–2258.https://doi.org/10.1364/optcon.599341

[30] Sugimura, D., Maeda, Y., Yamamura, K., Kurihara, K., & Hamamoto, T. (2025). Lensless facial image identification through joint learning of reconstruction and recognition. Optics Express, 33(14), 30384–30400.https://doi.org/10.1364/oe.565829

[31] Bezzam, E., Perron, Y., & Vetterli, M. (2025). Towards robust and generalizable lensless imaging with modular learned reconstruction. IEEE Transactions on Computational Imaging, 11, 213–227.https://doi.org/10.1109/tci.2025.3539448