Synthetic Sub-Aperture Phase Augmentation for Demosaicing 2×2 Shared Microlens Sensors

Jeisung Lee* and Wonil Song

System LSI Business, Samsung Electronics Co., Ltd., Hwaseong, Republic of Korea
{jeisung.lee, w1.song}@samsung.com

Abstract. Modern high-resolution mobile sensors often adopt clustered color filter array (CFA) layouts such as Quad and Hexadeca Bayer, while some premium sensors use quad-pixel phase-detection or 2×2 on-chip-lens (OCL) architectures with shared microlenses for dense phase-detection autofocus (PDAF). While effective for PDAF, this structure can introduce grid-aligned phase artifacts during demosaicing, which are difficult to model without calibrated sub-aperture streams, PSF supervision, or sensor-specific optical calibration. We propose *Synthetic Sub-Aperture Phase Augmentation* (SPA), a physics-inspired data generation framework for artifact-aware demosaicing. SPA uses compact phase kernels to approximate focus-dependent sub-aperture shifts in synthetic RAW, requiring no measured point spread functions (PSFs) or sensor-specific calibration. Experiments on synthetic and real RAW data from Hexadeca and Quad Bayer sensors show consistent suppression of zippering, color bleeding, and grid-aligned artifacts. We also introduce spatial lattice imbalance (SLI) and Nyquist energy ratio (NER) for 2×2 artifact analysis and validate robustness to spatial phase variations.

Keywords: demosaicing · shared microlens · phase modeling · synthetic data · phase-detection autofocus

1 Introduction

Recent mobile CMOS image sensors (CIS) have rapidly progressed toward ultra-high resolutions exceeding 200 megapixels [22]. To balance resolution and sensitivity under fixed pixel pitch constraints [11, 20], modern sensors adopt clustered Color Filter Array (CFA) layouts such as Quad (2×2), Nona (3×3), and Hexadeca (4×4) Bayer patterns [20, 23]. While pixel binning in these layouts improves low-light performance [20, 23], full-resolution operation requires demosaicing from sparsely distributed chroma samples arranged in same-color $n \times n$ clusters, significantly increasing reconstruction difficulty.

In parallel, premium mobile sensors increasingly employ quad-pixel phase-detection or 2×2 on-chip-lens (OCL) architectures with shared microlenses for dense phase-detection autofocus (PDAF). We refer to this structure as a **Q-cell lens**. Unlike dual-pixel (2PD) sensors that capture horizontal disparity only,

* Corresponding author.

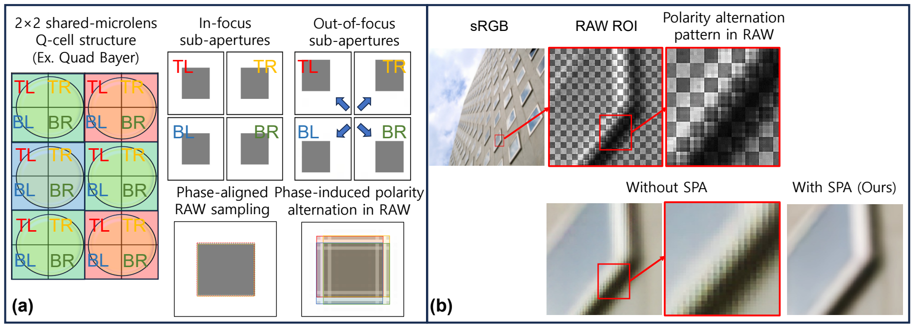

Fig. 1: Phase-induced RAW polarity alternation and resulting grid-aligned demosaicing artifacts. (a) A 2×2 shared-microlens Q-cell structure assigns four same-color pixels to different sub-aperture phases (TL, TR, BL, and BR). Under defocus, phase-dependent sub-aperture shifts induce a polarity alternation pattern within the RAW sampling grid. (b) Without phase-shift modeling, this structured periodic bias propagates to checkerboard artifacts after demosaicing. SPA synthesizes such sub-aperture phase variations during training and suppresses the resulting Q-cell artifacts.

Q-cell lenses introduce four directional sub-aperture phase shifts within each 2×2 cluster. These 2×2 shared-microlens designs provide dense PDAF coverage while preserving high-resolution readout. In this work, we focus on such shared-microlens sensors, where each 2×2 same-color unit produces four sub-aperture responses.

The interaction between clustered CFA sampling and Q-cell phase shifts introduces a previously underexplored artifact domain. As illustrated in Fig. 1, defocus induces lateral displacement among sub-apertures, creating periodic polarity alternation patterns directly in the RAW domain. Conventional demosaicing methods, which assume intra-cluster photometric consistency, propagate this structured bias into grid-aligned checkerboard artifacts and color misalignment in RGB space.

Fig. 2 further illustrates the optical intuition: defocus and chief-ray-angle (CRA)-induced directional bias create structured RAW periodicity across the 2×2 unit.

Relation to Prior Work. Dual-pixel research models defocus-induced PSFs for depth estimation and deblurring [1, 21, 26], primarily addressing view-dependent blur in the image domain. Clustered CFA demosaicing approaches [7, 14, 16, 25, 30] focus on sparse reconstruction under clustered sampling assumptions. However, these methods generally assume photometric consistency within each same-color cluster and do not account for systematic directional bias introduced by shared microlens phase shifts. Our setting differs: Q-cell phase shifts introduce structured directional imbalance before demosaicing, producing color-domain artifacts rather than depth or blur reconstruction errors.

Practical Challenge. Recent work on learning lens blur fields has shown that spatially varying, sensor-specific PSFs can be acquired and modeled for

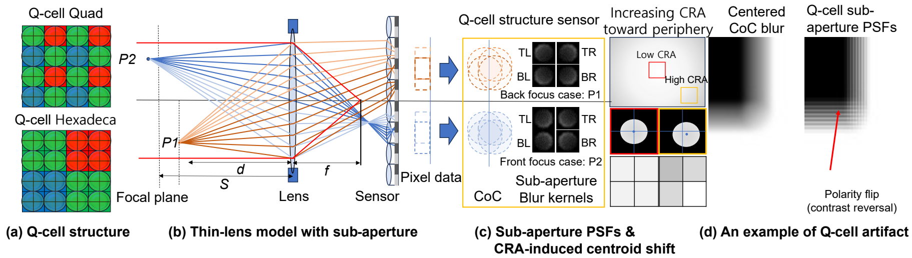

Fig. 2: Conceptual illustration of sub-aperture imbalance in Q-cell sensors. Defocus and CRA-induced directional bias produce structured RAW periodicity within each 2×2 cluster.

calibration-based optical restoration [18]. However, such calibration is difficult to scale to deployment-oriented Q-cell demosaicing because sensor modules, lens stacks, focus positions, field locations, and color channels may require separate characterization. Our setting instead provides the full-resolution Q-cell RAW mosaic from the sensor/image signal processor (ISP) pipeline, not calibrated sub-aperture streams or ground-truth PSFs. Since the phase cues in each 2×2 shared-microlens sample are entangled with CFA sampling, lens shading, pixel-response nonuniformity, optical crosstalk, and focus-/field-dependent variations, we exploit them through physics-inspired, calibration-free synthetic modeling rather than explicit field-dependent PSF estimation.

Our Approach. We propose *Synthetic Sub-Aperture Phase Augmentation* (SPA), a physics-inspired RAW generation framework for artifact-aware demosaicing. SPA approximates four-directional sub-aperture shifts for Q-cell configurations and injects them into synthetic RAW, enabling phase-aware training without measured PSFs or sensor-specific calibration. Rather than increasing backbone capacity or estimating explicit phase maps, SPA exposes the model to the phase-induced RAW structures that cause Q-cell artifacts. Although our main validation focuses on 2×2 shared-microlens Q-cell sensors, the formulation is defined by sub-aperture sampling and directional kernels, making it adaptable to other PDAF layouts.

Contributions.

- We analyze how sub-aperture phase shifts in shared-microlens sensors interact with clustered CFA sampling to produce structured color-domain artifacts—a failure mode not addressed by existing demosaicing methods.
- We propose SPA, a physics-inspired synthetic forward model that enables phase-aware training without measured PSFs or sensor-specific optical calibration.
- We introduce SLI and NER, two complementary metrics for quantifying 2×2 periodic artifacts in the spatial and spectral domains, providing evaluation tools beyond perceptual quality scores.

2 Related Work

2.1 Demosaicing under Bayer and Clustered CFAs

Deep learning has substantially advanced Bayer demosaicing [8, 15, 19, 27], including end-to-end mobile ISP pipelines such as PyNet [9,10,13]. Clustered CFA layouts (Quad, Nona, Hexadeca) increase chroma sparsity and require larger receptive fields and robust cross-channel interpolation. Accordingly, specialized architectures such as DPN [14], multi-scale attention models [25], DRNet [30], PyNet-Q×Q [7], and KLAP [16] have been proposed.

These approaches address sparse clustered sampling but generally assume photometric consistency (up to noise and blur) within each same-color cluster, without modeling systematic directional bias from shared-microlens phase shifts.

2.2 Dual-Pixel and Sub-Aperture Phase Modeling

Dual-pixel (2PD) sensors capture left/right sub-aperture views and have been widely studied for depth estimation and defocus deblurring [1, 21, 26]. Physics-based modeling of defocus-dependent PSFs under thin-lens assumptions has also been validated [2]. Quad-pixel sensors have also been analyzed from a multi-view or plenoptic perspective for view synthesis and refocusing [5].

Prior work primarily targets disparity, depth, or image-domain blur reconstruction. In contrast, we study how omnidirectional sub-aperture shifts interact with clustered CFA sampling to create structured RAW periodicity and color-domain demosaicing artifacts, an interaction not previously analyzed explicitly for demosaicing.

2.3 Synthetic RAW Generation

Unprocessing pipelines [4] enable realistic synthetic RAW generation by inverting ISP operations, but they do not simulate sub-aperture phase displacement. Dual-pixel simulation approaches [2] model two-view horizontal disparity under defocus conditions.

Thus, existing synthetic RAW pipelines do not expose demosaicers to the phase-dependent within-cluster imbalance observed in Q-cell RAW data. SPA fills this gap by adding Q-cell-specific sub-aperture phase kernels and CRA-inspired directional imbalance before CFA mosaicing (Sec. 3).

3 Method

3.1 Sub-Aperture Phase-Shift Modeling for Synthetic RAW

Motivation and training philosophy. Under defocus, Q-cell sensors produce spatially shifted sub-aperture views (TL, TR, BL, BR), violating the local smoothness assumptions commonly used in standard demosaicing. Same-color pixels within each 2×2 cluster become phase-misaligned, creating artificial periodic variations that resemble genuine high-frequency content. Conventional demosaicers often propagate this structured bias into checkerboard patterns, color misalignment, and edge artifacts.

Because standard unprocessing [4] models noise and ISP inversion but not this phase behavior, we introduce **Synthetic Sub-Aperture Phase Augmentation (SPA)**. SPA injects physically inspired sub-aperture blur into synthetic RAW, enabling networks to distinguish natural structure from phase-induced periodicity and suppress artifacts while preserving defocus blur, without explicit alignment or sensor-specific calibration.

Optical assumption. We use a thin-lens approximation and the Butterworth formulation validated for 2PD PSFs [2] to capture the principal defocus-induced disparity, while not explicitly modeling higher-order aberrations. To accommodate ring-like and softened-disk profiles from partial aperture masking, we sample kernels from soft-ring to near-pillbox shapes. Comparable performance on pillbox-based tests (supplementary) indicates that suppression primarily depends on directional energy imbalance rather than exact PSF shape. These kernels are sampled as a broad physics-inspired family rather than fitted to a measured sensor PSF; no sensor-specific optical calibration is used.

Inverse ISP and dataset preparation. Following Brooks et al. [4], we invert tone mapping, gamma, color correction, and white balance to obtain scene-referred linear RGB, with randomly sampled ISP parameters; dataset details appear in Sec. 4.

Base defocus kernel. For scene distance d, the circle-of-confusion (CoC) radius is computed as:

$$r = \frac{q}{2}\frac{s_0}{s}\frac{d-s}{d}, \quad s_0 = \frac{fs}{s-f}, \quad q = \frac{f}{F}, \tag{1}$$

where f is focal length, F is f-number, s is focus distance, and d is scene depth.

The signed CoC parameter r encodes front-/back-focus polarity, while the radial support is determined by $|r|$. The symmetric base kernel is constructed as a softened annular disk:

$$\mathbf{K}_c = \mathcal{N}\Big(G_\sigma * \Big[\widetilde{B}(x,y) \circ C(|r|)\Big]\Big), \tag{2}$$

where $C(|r|)$ denotes a disk of radius $|r|$, G_σ is a Gaussian smoothing kernel, $\mathcal{N}$ denotes unit-sum normalization, and

$$B(x,y) = \frac{\left(\frac{\rho}{D_0}\right)^{2n}}{1+\left(\frac{\rho}{D_0}\right)^{2n}}, \quad \rho = \sqrt{(x-x_0)^2+(y-y_0)^2}. \tag{3}$$

Here, $B(x,y)$ is a radially increasing Butterworth profile that suppresses the kernel center. It is linearly rescaled to $\widetilde{B}(x,y) \in [\beta, 1]$ before disk masking, producing profiles ranging from softened annular shapes to near-pillbox disks. In our implementation, $D_0 = (k-1)/\eta$ for kernel size $k = 2|r| + 1$, where η is the Butterworth cutoff factor controlling the radial transition.

Four-directional sub-aperture approximation. Inspired by dual-pixel PSF formulations [2], we construct a four-directional approximation suitable for Q-cell layouts. Rather than performing a full optical derivation, we combine orthogonal directional attenuation fields to reproduce the structured phase imbalance observed in practice.

Define linear attenuation fields $D_x(i,j) = \frac{j}{k-1}$ and $D_y(i,j) = \frac{i}{k-1}$. Directional kernels are formed via:

$$\begin{aligned} \mathbf{K}_{\mathrm{L}} &= \mathbf{K}_c \odot D_x, \quad \mathbf{K}_{\mathrm{R}} = \mathrm{flip}_x(\mathbf{K}_{\mathrm{L}}), \\ \mathbf{K}_{\mathrm{T}} &= \mathbf{K}_c \odot D_y, \quad \mathbf{K}_{\mathrm{B}} = \mathrm{flip}_y(\mathbf{K}_{\mathrm{T}}). \end{aligned} \tag{4}$$

Diagonal kernels are defined as:

$$\begin{aligned} \mathbf{K}_{\mathrm{TL}} &= \tfrac{1}{2}(\mathbf{K}_{\mathrm{T}} + \mathbf{K}_{\mathrm{L}}), \quad \mathbf{K}_{\mathrm{TR}} = \tfrac{1}{2}(\mathbf{K}_{\mathrm{T}} + \mathbf{K}_{\mathrm{R}}), \\ \mathbf{K}_{\mathrm{BL}} &= \tfrac{1}{2}(\mathbf{K}_{\mathrm{B}} + \mathbf{K}_{\mathrm{L}}), \quad \mathbf{K}_{\mathrm{BR}} = \tfrac{1}{2}(\mathbf{K}_{\mathrm{B}} + \mathbf{K}_{\mathrm{R}}). \end{aligned} \tag{5}$$

All kernels are energy-normalized. Front and back focus are distinguished by the sign of r: reversing the focus direction swaps the attenuation fields, $(D_x, D_y) \to (1-D_x, 1-D_y)$, which mirrors the sub-aperture energy distribution (cf. Fig. 2, P1 vs. P2).

CRA-inspired imbalance. We approximate local chief-ray angle (CRA) asymmetry without encoding absolute sensor coordinates. For each patch, we sample a tilt parameter $\delta \sim \mathrm{Uniform}(-0.15, 0.15)$:

$$\alpha_x = 1 + \delta, \quad \alpha_y = 1 - \delta. \tag{6}$$

Before computing the cardinal kernels, we apply this tilt to the attenuation fields:

$$D'_x = \alpha_x D_x, \quad D'_y = \alpha_y D_y.$$

We then use D'_x and D'_y in (4)–(5), so all kernels inherit a controlled, approximately energy-consistent imbalance. Independent patch-wise sampling exposes the network to local patterns that may arise from lens-shading asymmetry or microlens misalignment, without explicit field modeling or optical calibration.

Sub-aperture synthesis and target formation. Let $\mathbf{I}_{\mathrm{lin}}$ denote the scene-referred linear RGB image obtained by inverse ISP, and let $\mathbf{I}_{\mathrm{tgt}}$ denote the symmetric-blur training target. For each view $v \in \{\mathrm{TL}, \mathrm{TR}, \mathrm{BL}, \mathrm{BR}\}$:

$$\mathbf{I}^{(v)}_{\mathrm{lin}} = \mathbf{K}_v * \mathbf{I}_{\mathrm{lin}}, \quad \mathbf{I}_{\mathrm{tgt}} = \mathbf{K}_c * \mathbf{I}_{\mathrm{lin}}. \tag{7}$$

The target preserves symmetric defocus blur while removing directional imbalance. No geometric alignment or disparity estimation is performed.

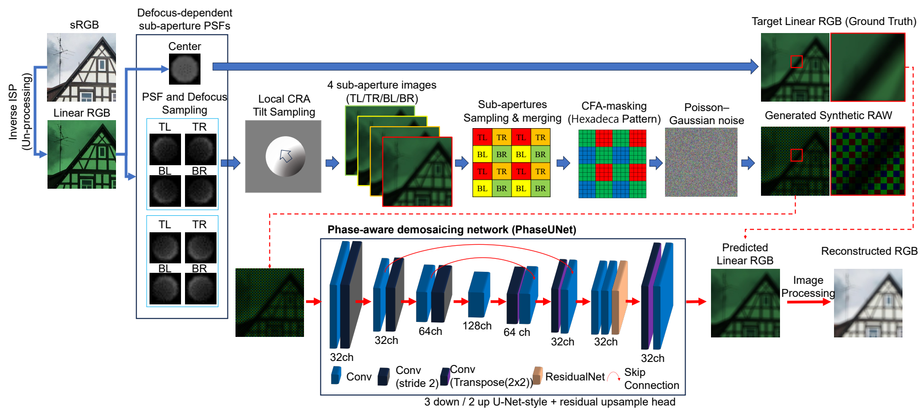

Fig. 3: SPA: Phase-shift-aware synthetic RAW generation and lightweight demosaicing network. *Top:* sRGB images are inverted to scene-referred linear RGB via unprocessing. A randomized family of defocus-dependent sub-aperture kernels is sampled per patch. CRA-inspired directional imbalance is introduced through patch-wise local angular perturbation. The four blurred sub-aperture views are interleaved into a clustered CFA RAW mosaic and corrupted with Poisson–Gaussian noise. *Bottom:* A compact U-Net-style network reconstructs full-resolution RGB while implicitly suppressing sub-aperture phase-shift artifacts.

RAW formation. The four views are interleaved into a clustered layout:

$$\mathbf{I}_{\text{merged}}(i,j) = \begin{cases} \text{TL}, & i \bmod 2 = 0,\ j \bmod 2 = 0, \\ \text{TR}, & i \bmod 2 = 0,\ j \bmod 2 = 1, \\ \text{BL}, & i \bmod 2 = 1,\ j \bmod 2 = 0, \\ \text{BR}, & i \bmod 2 = 1,\ j \bmod 2 = 1, \end{cases} \tag{8}$$

followed by CFA masking and Poisson–Gaussian noise addition.

3.2 Network Architecture and Training Objective

Lightweight design philosophy. We use a compact U-Net-style demosaicing network ($\sim$0.9M parameters, base width C=32) comprising an encoder–decoder with additive skip connections and a residual refinement head (Fig. 3; full specifications in the supplementary). Its small size isolates the benefit of phase-aware data generation from increased model capacity.

Loss function. We use an ℓ_1 reconstruction loss:

$$\mathcal{L} = \|\hat{\mathbf{I}} - \mathbf{I}_{\text{tgt}}\|_1. \tag{9}$$

4 Experiments

4.1 Experimental Setup

Synthetic training data generation. We generate synthetic Q-cell RAW from DIV2K [3] and Flickr2K [17] using the SPA pipeline (Sec. 3.1). Images are tiled into 512×512 patches at 128-pixel stride (approximately 80k patches); training uses random 128×128 crops, 90°/180°/270° rotations, and horizontal/vertical flips.

Defocus parameter sampling. We sample the circle-of-confusion radius up to $|r| = 25$ pixels to cover both front- and back-focus cases. Softened disk PSFs follow the Butterworth formulation in prior 2PD modeling [2] with stochastic sampling per patch: $n \sim \text{Uniform}\{3, 9\}$, $\eta \sim \text{Uniform}(1.5, 2.5)$, and $\beta \sim \text{Uniform}(0.1, 0.2)$.

Following prior 2PD modeling and empirical 2×2 microlens-array (MLA) Q-cell behavior, we cap the effective CoC at 25 pixels under typical shooting distances, prioritizing structured-artifact coverage over sensor-specific calibration.

Directional and noise modeling. CRA-inspired imbalance is sampled per patch via $\delta \sim \text{Uniform}(-0.15, 0.15)$, introducing controlled anti-correlated scaling between orthogonal attenuation fields. This conservative range avoids unrealistic distortion. SPA is applied with 50% probability to balance phase-aware and neutral cases. Poisson–Gaussian noise follows [4], with identical ISP parameters across methods for fair sRGB comparison.

4.2 Evaluation Protocols

We evaluate controlled synthetic benchmarks and real Q-cell RAW captures.

Synthetic sRGB benchmarks. We evaluate on Kodak, McM, Set5, Set14, and Urban100 under Hexadeca (4×4) and Quad (2×2) CFAs. ISP parameters are randomly sampled but fixed across methods per image. Training loss is computed in linear RGB space, while synthetic benchmark metrics are computed after applying the same forward ISP and sRGB rendering to predictions and targets.

To reflect practical capture conditions, defocus probabilities are biased toward near-focus: 50% on-focus ($r = 0$), 30% slight ($0 < |r| \leq 2$), 10% moderate ($2 < |r| \leq 5$), and 10% severe ($|r| > 5$). We report PSNR/SSIM/LPIPS [29] averaged over five ISP seeds.

Real Q-cell RAW captures. We collected RAW images from two independent prototype Q-cell sensors under diverse shooting conditions. The Hexadeca dataset consists of 116 real captures, primarily natural scenes obtained under typical handheld shooting conditions. The Quad Bayer dataset contains 52 RAW captures, mainly controlled chart scenes acquired under varying focus distances ranging from near-focus to moderate defocus.

Because prototype RAW files cannot be released under NDA and ground truth is unavailable, we use qualitative inspection (checkerboard artifacts, color polarity reversal, and texture stability), CLIPIQA [24], and the proposed SLI and NER structural metrics. Crop locations were selected once before method comparison and fixed across all methods.

No-reference metric selection. CLIPIQA has the highest per-dataset rank correlation with LPIPS in our evaluation (mean Spearman $\rho = 0.74$); MUSIQ [12] and MANIQA [28] show lower consistency (supplementary). CLIPIQA assesses perceptual quality, whereas SLI and NER measure spatial sub-lattice imbalance and Nyquist-aligned spectral energy, respectively. Together they provide complementary evidence when full-reference evaluation is unavailable, but do not replace validation against paired ground truth.

4.3 Implementation Details

We train all models for 100 epochs with AdamW (learning rate 10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 10^{-5}) and batch size 8. Loss excludes a 24-pixel boundary. Tiled inference uses 512×512 patches with 64-pixel overlap, discards overlapping boundaries, and stitches central crops; no test-time augmentation is applied.

4.4 Compared Methods

We compare against PyNet-Q×Q [7] and KLAP [16] using official pretrained weights, as well as NAFNet [6] retrained on our data and DPN [14] retrained and extended to Hexadeca. KLAP-M is excluded because it requires per-image fine-tuning, which is incompatible with our zero-shot real RAW evaluation.

Retrainable baselines use the same source datasets, optimizer, loss, crop size, noise model, and non-phase augmentations, but are trained only on conventional synthetic RAW without SPA. PyNet-Q×Q is reported only under its released Q×Q pretrained configuration, which matches our 4×4 clustered CFA evaluation; we do not reconfigure or retrain it for the separate 2×2 Quad setting. All methods are tested under zero-shot inference on real Q-cell RAW. We additionally report an ablation using the identical PhaseUNet backbone trained without SPA.

5 Results

5.1 Synthetic Q-cell RAW Evaluation

Performance Across CFA Patterns We first evaluate the proposed phase-aware training strategy on synthetic Q-cell RAW data generated with controlled phase shifts. Table 1 reports results averaged over five benchmarks for both Hexadeca (4×4) and Quad (2×2) CFA layouts.

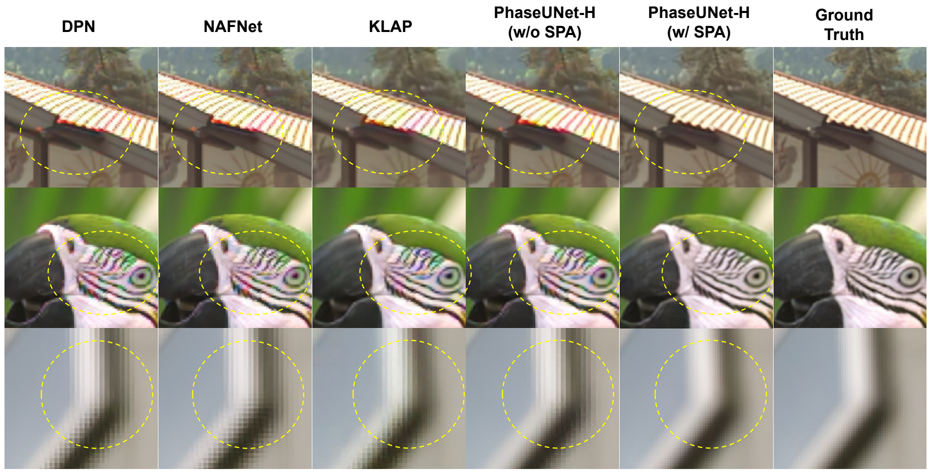

Fig. 4: Qualitative comparison on synthetic Hexadeca RAW. Phase-unaware methods exhibit structured checkerboard residuals and chromatic distortions aligned with the CFA grid. The proposed SPA-aware model suppresses grid-aligned artifacts while preserving natural texture and edge transitions. Zoomed-in regions highlight artifact reduction.

We denote models trained on Hexadeca and Quad CFA as PhaseUNet-H and PhaseUNet-Q, respectively (reported as a single row in Table 1 since the architecture is identical). Despite having only 0.90M parameters, PhaseUNet outperforms significantly larger baselines (NAFNet, KLAP) across both Hexadeca and Quad CFAs. This indicates that gains arise from structured phase-consistent modeling rather than model scaling.

Consistent improvements across CFA geometries demonstrate that SPA captures intrinsic sub-aperture phase behavior rather than overfitting to a specific sampling layout.

Figure 4 confirms that models trained without phase-aware data propagate sub-aperture inconsistencies into structured checker and chromatic artifacts, whereas SPA-aware training suppresses these periodic distortions without sacrificing texture detail. This supports the hypothesis that phase modeling, rather than increased capacity, is the primary factor for robust Q-cell demosaicing.

Impact of Defocus Severity We further analyze performance under increasing defocus levels. As defocus magnitude increases, sub-aperture phase misalignment becomes more pronounced, leading to structured checkerboard artifacts and chromatic distortions aligned with the CFA grid.

Table 2 shows progressively larger gains as defocus severity increases. SPA reduces LPIPS by 48%, 85%, and 95% under slight, moderate, and severe defocus, with PSNR gains of +2.53, +6.86, and +13.87 dB, respectively. All variants use the same PhaseUNet-H backbone, isolating SPA from model capacity; SPA

Table 1: Synthetic RAW evaluation across CFA types. Results averaged over five datasets. Phase-weighted defocus mix: 50% on-focus, 30% slight, 10% moderate, 10% severe. Params/MACs in M @ 128×128.

Model	Hexadeca (4 × 4) PSNR↑	SSIM↑	LPIPS↓	Quad (2 × 2) PSNR↑	SSIM↑	LPIPS↓	Params	MACs
DPN	35.09	0.942	0.0747	35.42	0.945	0.0698	5.58	10775
NAFNet	35.31	0.946	0.0719	35.68	0.948	0.0673	17.11	3985
KLAP	33.64	0.922	0.0965	34.12	0.928	0.0892	17.80	4016
PyNet-Q×Q	28.35	0.836	0.2057	–	–	–	1.07	4408
PhaseUNet (Ours)	**36.60**	**0.954**	**0.0434**	**37.15**	**0.958**	**0.0401**	**0.90**	**3010**

Table 2: Effect of phase-aware training across defocus levels (Hexadeca, averaged over 5 datasets).

Defocus	Variant	PSNR↑	SSIM↑	LPIPS↓
Slight ($0 < \|r\| \leq 2$)	w/o SPA	32.91	0.941	0.083
	w/ SPA	**35.44**	**0.954**	**0.043**
	Δ	*+2.53*	*+0.013*	*−48%*
Moderate ($2 < \|r\| \leq 5$)	w/o SPA	34.76	0.932	0.192
	w/ SPA	**41.62**	**0.985**	**0.028**
	Δ	*+6.86*	*+0.053*	*−85%*
Severe ($\|r\| > 5$)	w/o SPA	33.46	0.827	0.332
	w/ SPA	**47.33**	**0.994**	**0.015**
	Δ	*+13.87*	*+0.168*	*−95%*

mixing-ratio analysis is provided in the supplementary material. The mixing-ratio study shows that higher SPA ratios improve larger-CoC cases, whereas near-focus performance mildly degrades when phase-neutral samples become too rare; the 50% setting provides a balanced trade-off. Although severe-defocus PSNR is inflated by the heavily low-pass target, the monotonic trend remains consistent. Because the target retains the same symmetric defocus blur, the reduction in checker and false-color artifacts cannot be attributed to additional smoothing; it results from removing structured inter-aperture polarity imbalance.

This indicates that phase misalignment becomes the dominant reconstruction error source as blur severity grows. Consistent artifact suppression under strong defocus is further confirmed on real Q-cell RAW in Fig. 6, where peripheral regions with pronounced CRA-induced imbalance exhibit visible checkerboard removal.

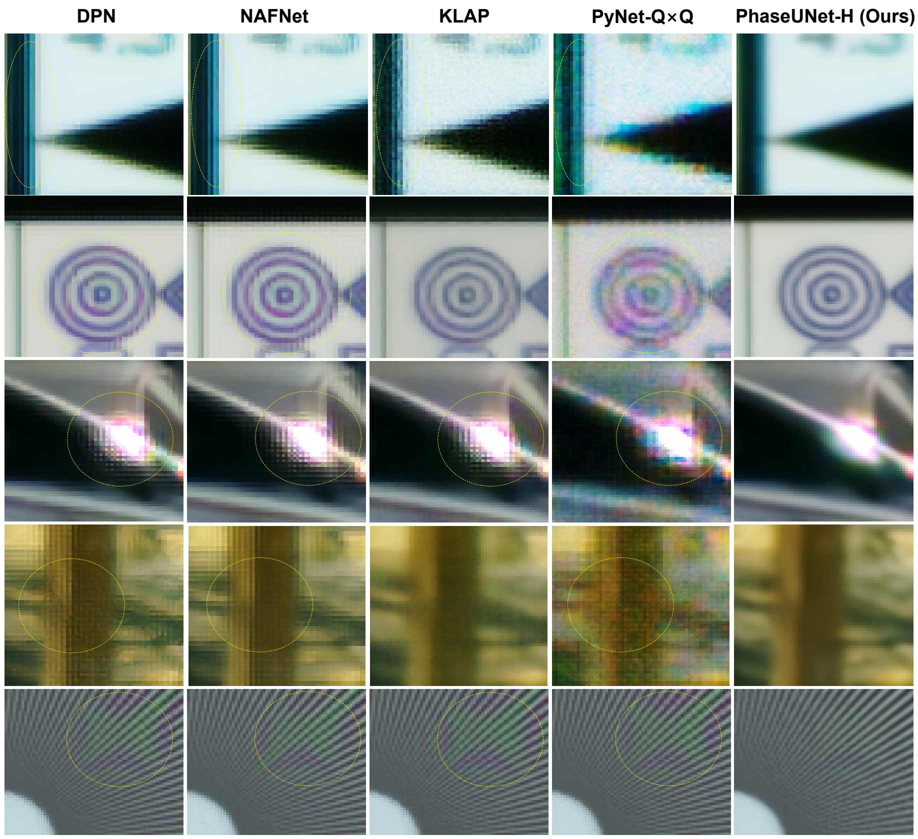

Fig. 5: Real Hexadeca RAW results. Baseline methods exhibit grid-aligned checker residuals and false-color artifacts in peripheral and defocused regions due to sub-aperture phase inconsistency. The proposed SPA-aware model suppresses structured artifacts while preserving natural blur transitions and color stability.

5.2 Real Q-cell RAW Evaluation

Visual Comparison In Hexadeca captures (Fig. 5), baseline methods amplify sub-aperture phase inconsistencies in peripheral and defocused regions, resulting in grid-aligned checker residuals and chromatic polarity shifts. The SPA-based model reduces these structured distortions and preserves stable color transitions.

In Quad-Bayer captures with 2×2 shared microlens structures (Fig. 6), strong lateral illumination creates uneven sub-aperture energy distribution, producing directional periodic artifacts. Similarly, CRA-induced oblique incidence shifts the effective optical response relative to the pixel grid, leading to consistent 2×2 structured patterns. Baseline models propagate these imbalances into visible checker artifacts, whereas SPA-trained reconstruction suppresses phase-induced periodic structures while maintaining perceptual coherence.

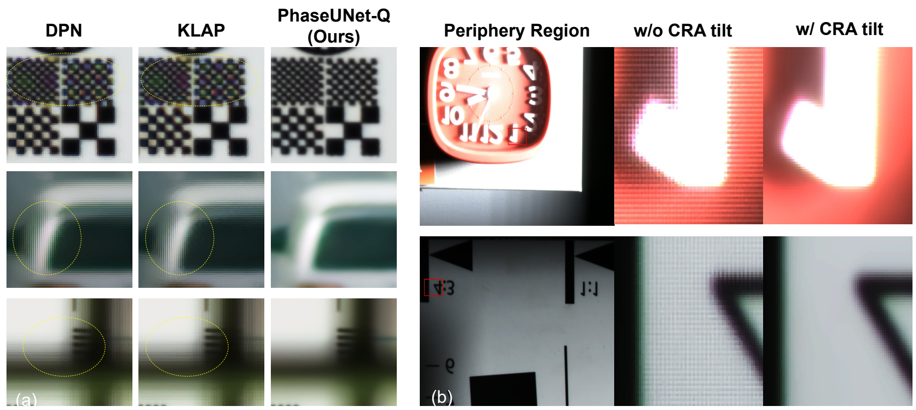

Fig. 6: Generalization to Quad-Bayer CFA and CRA-aware peripheral robustness. (a) SPA retrained under Quad-Bayer layout suppresses phase-shift-induced periodic artifacts on real RAW, confirming generalization beyond the original CFA configuration. (b) Without CRA tilt modeling, off-axis illumination produces directional energy imbalance and 2×2 periodic artifacts that propagate through demosaicing. Incorporating CRA tilt during training reduces these peripheral artifacts.

Perceptual Quality Assessment We further report CLIPIQA as a no-reference perceptual metric (Table 3). PhaseUNet achieves the highest score on Quad captures and remains competitive on Hexadeca without sensor-specific calibration. This indicates that phase-consistent synthetic training transfers effectively to real optical asymmetries.

In addition to perceptual quality, we quantify structured 2×2 periodic artifacts using two complementary metrics. SLI and NER are primarily intended for within-dataset method comparison, since their absolute scale can vary with crop content and CFA-specific spectral normalization.

Table 3: Quantitative comparison on real Q-cell RAW datasets. CLIPIQA (↑) measures perceptual quality, while SLI and NER (↓) quantify 2×2 periodic artifacts. Metrics are averaged over 30 fixed crop locations sampled from defocused and peripheral regions of each dataset and shared across all methods.

	Hexadeca			Quad		
Model	CLIPIQA↑	SLI↓	NER↓	CLIPIQA↑	SLI↓	NER↓
DPN	0.376	0.00461	0.05067	0.409	0.00205	2.92934
NAFNet	0.368	0.00397	0.07274	0.401	0.00326	2.55741
KLAP	0.356	0.00594	0.03957	0.394	0.00139	1.90869
PyNet-Q×Q	0.257	0.01025	0.01204	–	–	–
PhaseUNet (Ours)	**0.390**	**0.00297**	**1.2e-4**	**0.467**	**0.00056**	**7e-5**

Spatial Lattice Imbalance (SLI). Let μ_k denote the mean intensity of the k-th sub-lattice within a 2×2 grouping and μ the global mean. SLI is defined as

$$\mathrm{SLI} = \frac{1}{4}\sum_{k=1}^{4} |\mu_k - \mu| \,. \tag{10}$$

This measures spatial imbalance caused by sub-aperture phase inconsistency.

Nyquist Energy Ratio (NER). Let $P(\mathbf{f})$ denote the FFT power spectrum. For an $H \times W$ patch, let $\mathbf{f}_c = (W/2,\, H/2)$ denote the DC center of the shifted spectrum. The set Ω_{N} consists of eight Nyquist-aligned locations: the four axis points $(\pm W/2,\, 0)$, $(0,\, \pm H/2)$ and four diagonal points $(\pm W/2,\, \pm H/2)$ relative to $\mathbf{f}_c$, each measured within a small square window of radius r_p pixels. The normalizer $\mathcal{M}$ is a mid-frequency annulus $r_{\min} \le \|\mathbf{f} - \mathbf{f}_c\| \le r_{\max}$ excluding a DC neighborhood. NER is defined as

$$\mathrm{NER} = \frac{\sum_{\mathbf{f} \in \Omega_{\mathrm{N}}} P(\mathbf{f})}{\sum_{\mathbf{f} \in \mathcal{M}} P(\mathbf{f})}. \tag{11}$$

This quantifies the relative strength of structured spectral energy at 2×2 periodic frequencies.

Across both Hexadeca and Quad configurations, PhaseUNet consistently achieves the lowest SLI and NER values. In particular, the proposed method reduces NER dramatically in the Quad setting (from $\sim$2.9 to 7×10^{-5}). This reflects near-complete elimination of spectral peaks at 2×2 Nyquist frequencies: baseline spectra exhibit sharp energy concentration at these locations due to periodic sub-aperture polarity alternation, whereas the proposed method produces smooth spectral decay consistent with natural image statistics (see supplementary for parameter details and FFT visualizations).

6 Conclusion

We presented Synthetic Sub-Aperture Phase Augmentation (SPA), a calibration-free synthetic RAW generation framework for phase-aware demosaicing in shared-microlens CFA sensors. Through synthetic experiments and real Q-cell RAW evaluations on Hexadeca and Quad Bayer sensors, we showed that sub-aperture imbalance produces grid-aligned checker and false-color artifacts, and that SPA suppresses these structured distortions while preserving natural defocus characteristics.

SPA is defined by sub-aperture sampling and directional kernels rather than a particular backbone or CFA geometry. Although our main validation focuses on 2×2 shared-microlens Q-cell sensors, this formulation makes the augmentation principle adaptable to other PDAF layouts; we include a synthetic 1×2 two-photodiode generalization check in the supplementary material. Additional real-sensor validation on broader PDAF architectures remains limited by data availability. Future work will explore spatially adaptive, field-dependent modeling and broader sensor validation.

References

1. Abuolaim, A., Brown, M.S.: Defocus deblurring using Dual-Pixel data. In: ECCV (2020)
2. Abuolaim, A., Delbracio, M., Kelly, D., Brown, M.S., Milanfar, P.: Learning to reduce defocus blur by realistically modeling Dual-Pixel data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2289–2298 (2021)
3. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 126–135 (2017)
4. Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: Unprocessing images for learned raw denoising. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11036–11045 (2019)
5. Chataignier, G., Vandame, B., Vaillant, J.: Deep learning applied to Quad-Pixel plenoptic images. In: Computational Optics 2021. vol. 11875, p. 1187505. SPIE (2021)
6. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: European conference on computer vision. pp. 17–33. Springer (2022)
7. Cho, M., Lee, H., Je, H., Kim, K., Ryu, D., No, A.: PyNET-Q×Q: An efficient PyNET variant for Q×Q Bayer pattern demosaicing in CMOS image sensors. IEEE Access **11**, 44895–44910 (2023)
8. Gharbi, M., Chaurasia, G., Paris, S., Durand, F.: Deep joint demosaicking and denoising. ACM Transactions on Graphics (ToG) **35**(6), 1–12 (2016)
9. Ignatov, A., Malivenko, G., Timofte, R., Tseng, Y., Xu, Y.S., Yu, P.H., Chiang, C.M., Kuo, H.K., Chen, M.H., Cheng, C.M., et al.: PyNet-V2 mobile: Efficient on-device photo processing with neural networks. In: 2022 26th International Conference on Pattern Recognition (ICPR). pp. 677–684. IEEE (2022)
10. Ignatov, A., Van Gool, L., Timofte, R.: Replacing mobile camera ISP with a single deep learning model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 536–537 (2020)
11. Jang, D., Park, D., Cha, S., Kwon, H., Kim, M., Lee, S., Lee, H., Kim, S., Lee, N., Han, J., et al.: 0.8μm-pitch CMOS image sensor with dual conversion gain pixel for mobile applications. In: IISW (2019)
12. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: MUSIQ: Multi-scale image quality transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5148–5157 (2021)
13. Kim, B.H., Song, J., Ye, J.C., Baek, J.: PyNET-CA: Enhanced PyNET with channel attention for end-to-end mobile image signal processing. In: European Conference on Computer Vision. pp. 202–212. Springer (2020)
14. Kim, I., Song, S., Chang, S., Lim, S., Guo, K.: Deep image demosaicing for submicron image sensors. Journal of Imaging Science and Technology **63**(6), 060410–1–060410–12 (2019). `https://doi.org/10.2352/J.ImagingSci.Technol.2019.63.6.060410`
15. Kokkinos, F., Lefkimmiatis, S.: Deep image demosaicking using a cascade of convolutional residual denoising networks. In: Proceedings of the European conference on computer vision (ECCV). pp. 303–319 (2018)
16. Lee, H., Park, D., Jeong, W., Kim, K., Je, H., Ryu, D., Chun, S.Y.: Efficient unified demosaicing for Bayer and non-Bayer patterned image sensors. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12750–12759 (2023)

17. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 136–144 (2017)
18. Lin, E.Y.H., Wang, Z., Lin, R., Miau, D., Kainz, F., Chen, J., Zhang, X., Lindell, D.B., Kutulakos, K.N.: Learning lens blur fields. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–12 (2025)
19. Liu, L., Jia, X., Liu, J., Tian, Q.: Joint demosaicing and denoising with self guidance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2240–2249 (2020)
20. Oh, Y., Kim, M., Choi, W., Choi, H., Jeon, H., Seok, J., Choi, Y., Jung, J., Yoo, K., Park, D., et al.: A 0.8 μm Nonacell for 108 megapixels CMOS image sensor with FD-shared dual conversion gain and 18,000 e-full-well capacitance. In: 2020 IEEE International Electron Devices Meeting (IEDM). pp. 16–2. IEEE (2020)
21. Punnappurath, A., Abuolaim, A., Afifi, M., Brown, M.S.: Modeling defocus-disparity in Dual-Pixel sensors. In: 2020 IEEE International Conference on Computational Photography (ICCP). pp. 1–12. IEEE (2020)
22. Scott-Thomas, J.: Trends and developments in state-of-the-art CMOS image sensors. In: 2023 International Image Sensor Workshop (2023)
23. Singh, M., Singh, T.: Double binnable RGB, RGBW and LMS color filter arrays. Electronic Imaging **36**(16), 164-1–164-6 (2024). `https://doi.org/10.2352/EI.2024.36.16.COLOR-164`
24. Wang, J., Chan, K.C., Loy, C.C.: Exploring CLIP for assessing the look and feel of images. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 2555–2563 (2023)
25. Wu, X., Fan, Z., Zheng, J., Wu, Y., Zhang, F.: Learning to joint remosaic and denoise in Quad Bayer CFA via universal multi-scale channel attention network. In: European Conference on Computer Vision. pp. 147–160. Springer (2022)
26. Xin, S., Wadhwa, N., Xue, T., Barron, J.T., Srinivasan, P.P., Chen, J., Gkioulekas, I., Garg, R.: Defocus map estimation and deblurring from a single Dual-Pixel image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2228–2238 (2021)
27. Xing, W., Egiazarian, K.: End-to-end learning for joint image demosaicing, denoising and super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3507–3516 (2021)
28. Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J., Yang, Y.: MANIQA: Multi-dimension attention network for no-reference image quality assessment. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1191–1200 (2022)
29. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595 (2018)
30. Zheng, B., Li, H., Chen, Q., Wang, T., Zhou, X., Hu, Z., Yan, C.: Quad Bayer joint demosaicing and denoising based on dual encoder network with joint residual learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 7552–7561 (2024)