CalibFPA: A Focal Plane Array Imaging System based on Online Deep-Learning Calibration

Alper Güngör[*,†], M. Umut Bahceci[*], Yasin Ergen, Ahmet Sözak, O. Oner Ekiz,
Tolga Yelboga, Tolga Çukur *Senior Member*

Abstract—Compressive focal plane arrays (FPA) enable cost-effective high-resolution (HR) imaging by acquisition of several multiplexed measurements on a low-resolution (LR) sensor. Multiplexed encoding of the visual scene is often attained via electronically controllable spatial light modulators (SLM). To capture system non-idealities such as optical aberrations, a system matrix is measured via additional offline scans, where the system response is recorded for a point source at each spatial location on the imaging grid. An HR image can then be reconstructed by solving an inverse problem that involves encoded measurements and the calibration matrix. However, this offline calibration framework faces limitations due to challenges in encoding single HR grid locations with a fixed coded aperture, lengthy calibration scans repeated to account for system drifts, and computational burden of reconstructions based on dense system matrices. Here, we propose a novel compressive FPA system based on online deep-learning calibration of multiplexed LR measurements (CalibFPA). To acquire multiplexed measurements, we devise an optical setup where a piezo-stage locomotes a pre-printed fixed coded aperture. We introduce a physics-driven deep-learning method to correct for the influences of optical aberrations in multiplexed measurements without the need for offline calibration scans. The corrected measurement matrix is of block-diagonal form, so it can be processed efficiently to recover HR images with a user-preferred reconstruction algorithm including least-squares, plug-and-play, or unrolled techniques. On simulated and experimental datasets, we demonstrate that CalibFPA outperforms state-of-the-art compressive FPA methods. We also report analyses to validate the design elements in CalibFPA and assess computational complexity.

Index Terms—Focal plane array, spatial light modulator, deep learning, calibration, reconstruction.

I. Introduction

Conventional imaging systems leverage high-resolution (HR) sensors with millions of pixels, therefore requiring complex, expensive manufacturing processes [1], [2]. In recent years, multiplexed imaging has come forth as a cost-efficient framework to image HR scenes with low-resolution (LR) sensors for numerous imaging modalities operating across various wavelengths [1]–[7]. In this framework, a spatial light modulator (SLM), typically placed in between the imaged object and the LR sensor, performs spatial encoding of the

A. Güngör, M. U. Bahceci, A. Sozak, T. Yelboga are with ASELSAN, Ankara, Turkiye.

Y. Ergen and O. O. Ekiz are with Nanodev Scientific, Ankara, Turkiye.

O. O. Ekiz is also with dept. of Materials Eng., Ostim Technical University, Ankara, Turkiye.

T. Çukur is with Dept. of Electrical-Electronic Eng., National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkiye.

[*]These authors contributed equally.

[†]Corresponding author (email: alpergungor@windowslive.com).

scene [2]–[4], [8]–[11]. The SLM contains a dense grid of sub-pixels that can be electronically controlled to mask or unmask the incident light. As such, multiplexed encoding of a given scene can be performed by programming the SLM to create time-varying masking patterns [1]. The resultant multiplexed measurements can then be used to solve an inverse problem for reconstructing the HR image of the visual scene [4].

Multiplexed imaging holds particular promise for performant imaging in relatively high-frequency bands where manufacturing HR sensors is challenging [2], [4]–[6]. While systems can be implemented via single-pixel sensors (i.e., photo-detectors), recent studies have adopted LR albeit multi-pixel focal plane array (FPA) sensors for their favorable performance/cost trade-off [4], [10], [12]. Yet, in compressive FPA systems, measurements can be corrupted by system non-idealities including optical aberrations, distortion and vibration, which in turn compromise image quality [4], [10], [12]–[14]. These non-idealities are particularly problematic towards higher wavelengths, such as those in thermal imaging [15].

A fundamental framework to cope with system non-idealities is to conduct offline calibration scans to measure the system matrix that describes the relationships between individual pixels in the HR SLM and the LR FPA, and then solve an inverse problem involving the system matrix for reconstruction (Fig. 1). In a common albeit slow approach, calibration measurements are taken while a single-pixel of the SLM is unmasked to propagate the incident scene, and systematically traversed across grid locations [5], [14]. Calibration measurements can also be taken by using random masking patterns illuminating multiple grid locations at once, and the system matrix can then be recovered from these simultaneous measurements [14]. Note that these offline calibration methods rely on electronically-switchable (ES) SLMs for precise illumination control, yet they yield suboptimal encoding when using fixed coded aperture filters [10], [16]. Furthermore, given typically large system matrix dimensions (e.g., a $256^2 \times 1024^2$ for a 1024×1024 SLM and a 256×256 FPA), offline calibration requires prolonged scans that must be repeated over time to account for system drifts due to nuisance factors (e.g., vibration, thermal expansion/contraction). Lastly, image reconstruction based on dense system matrices in the offline framework introduces significant computational burden.

To address these limitations, here we introduce a novel multiplexed imaging system based on online calibration, CalibFPA, that corrects FPA measurements for influences from optical aberrations without requiring a measured system matrix (Fig. 1). CalibFPA projects the incident scene through a fixed
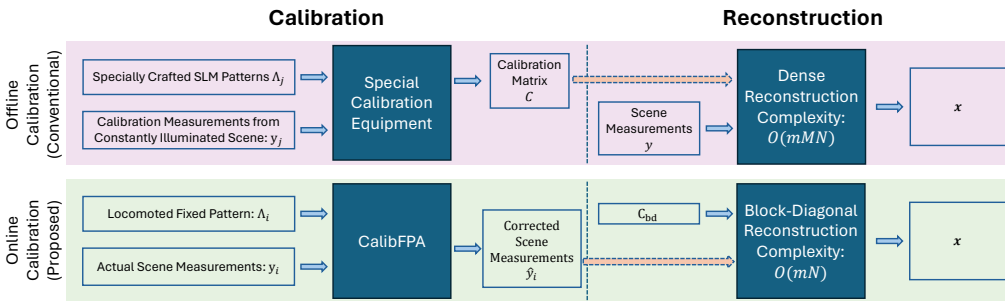
Fig. 1: In the offline calibration framework, dedicated calibration scans are performed where specially tailored SLM patterns Λ_i (e.g., single-point or random masks) and a calibration scene of constant illumination are used to capture a dense system matrix $\mathbf{C}$ that reflects the effects of downsampling, SLM patterns and optical aberrations. Afterwards, an image $\mathbf{x}$ is reconstructed from multiplexed measurements of the actual visual scene $\mathbf{y}$ via an algorithm of $O(mMN)$ complexity (m : number of SLM patterns, M : LR sensor size, N : HR image size). In contrast, the proposed online calibration in CalibFPA directly performs multiplexed measurements on the actual visual scene via a locomoted fixed SLM pattern, and uses a deep neural network to output a corrected set of measurements $\hat{\mathbf{y}}$ for optical aberrations. A system matrix copmrising the effects of downsampling and SLM patterns can be analytically formed. Lastly, a block-diagonal reconstruction algorithm of $O(mN)$ complexity can recover an image.

coded aperture and focuses the multiplexed scene onto the LR sensor via a relay lens (Fig. 2). Multiple measurements are taken while the coded aperture is locomoted by several sub-pixels to create different SLM patterns. Compared to systems with ES-SLMs, our proposed SPI system is compact due to the fixed coded aperture, and it enables the joint design of the piezo-stage with the coded aperture [16]. For calibration, a deep neural network performs online correction of multiplexed measurements for optical aberrations due to the spatial misalignment and transfer function of the relay lens (Fig. 3), which are modeled through a point spread function (PSF) [15]. Afterwards, corrected measurements are processed with a user-preferred reconstruction algorithm that does not require inversion of a dense system matrix. As such, CalibFPA avoids the need for measuring the system matrix via an offline calibration scan and alleviates computational burden during subsequent reconstruction. Our contributions are as follows:

- We propose a compact, low-cost compressive FPA system that acquires multiplexed measurements via piezo-stage-driven locomotion of a fixed coded aperture.
- We introduce a novel deep-learning calibration for online correction of compressive FPA measurements against optical aberrations to improve efficiency in scanning and reconstruction.
- We propose a physics-driven network architecture with radius and SLM coding blocks to ensure reliable calibration across a broad range of parameters.

II. RELATED WORK

A. Compressive FPA Imaging

A powerful imaging framework that avoids expensive HR sensors is single-pixel imaging (SPI) [2], [3]. In SPI, multiplexed encoding of HR spatial information is performed via an SLM and the encoded scene is then imaged using a single photo-detector. The SLM selectively masks or unmasks incident light at each pixel, so the encoded scene is expressed as the multiplication of the original scene with a Bernoulli-type matrix. Use of a single-pixel sensor inherently limits unwanted influences from system non-idealities. Thus, leveraging regularization priors, an inverse problem can be solved to recover the HR image of the scene from raw, multiplexed LR

measurements [17]. SPI systems typically operate at several kHz to permit real-time imaging. Yet, the use of single-pixel measurements inevitably hampers imaging efficiency.

To improve imaging efficiency, recent studies have considered using a multi-pixel LR sensor instead of a photo-detector [5], [9], [14], [18]. A prominent approach combines HR ES-SLMs for spatial encoding with LR FPAs for sensing [4]. An alternative approach uses coded apertures along with dispersive prisms to encode information in the spectral domain [10], [19]. Joint encoding across spatial and spectral dimensions has also been proposed [13], [20]. The presence of multiple pixels in FPA systems renders them more susceptible to system non-idealities. Thus, reliable reconstruction characteristically requires the solution of an inverse problem where system non-idealities are effectively accounted for [21]–[23].

Previous studies have commonly adopted the offline calibration framework that measures a system matrix via additional calibration scans and then compensates system non-idealities via the measured system matrix during reconstruction (Fig. 1). A conventional offline method is point-scan calibration where ES-SLMs are used to mask/unmask a single HR pixel per measurement. Since separate measurements must be taken for each combination of HR SLM and LR FPA pixels, this approach is experimentally burdening [5]. An offline method for accelerated calibration employs compressed sensing (CS) on sparse calibration measurements [14]. Instead of a single HR pixel, a random subset of HR pixels are masked/unmasked per measurement, and a CS algorithm is then used to recover the entire system matrix. That said, both point-scan and CS methods for offline calibration require the use of an electronically controllable SLM, which is not applicable on fixed coded apertures [16]. Furthermore, the system matrix capturing the relations between individual pixels in the SLM and the FPA is large and dense [4], [5]. This incurs excessive computational burden during image reconstruction, potentially compromising real-time processing capabilities of the imaging system [4], [5], [24]. A recent study has proposed to improve computational efficiency via a block-wise reconstruction approach [24]. The HR pixels in the SLM are split into blocks spanning the size of LR pixels, and a locally-constrained calibration matrix is then measured. Afterwards, reconstruction is attained by processing measurements in each block separately. While
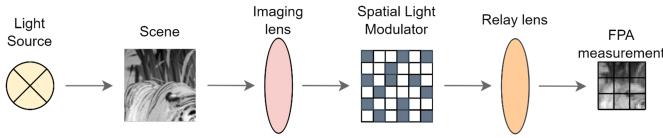
Fig. 2: Multiplexed imaging system in CalibFPA. An imaging lens focuses incident light coming from the scene onto a high-resolution (HR) spatial light modulator (SLM). A relay lens then focuses the encoded scene onto a low-resolution (LR) focal plane array (FPA) for measurement.

promising results have been reported, the locality constraints on the system matrix reduce the sensitivity of this approach to the global context in HR images.

Here, we introduce a novel compressive FPA system based on online calibration to cope with optical aberrations from the relay lens without introducing excessive experimental or computation load due to the use of a dense system matrix (see Fig. 1.). To our knowledge, CalibFPA is the first online calibration method for compressive FPA imaging in the literature. Unlike ES-SLM, CalibFPA utilizes a fixed coded aperture and locomotes it via a piezo-stage for super-resolving the spatial dimensions. Unlike offline calibration, CalibFPA avoids the need to acquire and store dense system matrices during reconstruction. Instead, it performs online calibration via a deep network to enable efficient reconstructions. Moreover, unlike block-wise methods with locality constraints, CalibFPA processes the entire set of measurements collectively to improve sensitivity to global information. These unique technical attributes enable CalibFPA to produce high-quality reconstructions of HR scenes efficiently.

B. Learning-based Methods for Computational Imaging

Learning-based approaches have emerged for image formation tasks in computational imaging [25]–[28]. Many approaches have focused on image reconstruction given acquired data and a system matrix, typically measured via offline calibration. Common methods split the reconstruction task into sub-problems involving data consistency and image denoising, prominent examples being plug-and-play methods that perform denoising via pretrained networks [26], [29], unrolled methods that perform end-to-end training of a cascaded network containing data-consistency and denoising blocks [25], [30], and deep equilibrium methods that can enable an unbounded number of cascades in unrolled networks for fast inference [31], [32]. Unlike reconstruction methods devised to recover images from raw measurements given a system matrix, CalibFPA aims to correct measurements for optical aberrations so that a subsequent reconstruction can recover images without a separate calibration of the system matrix.

Recent approaches for learning-based image restoration from convolutional corruptions such as blind deblurring and single-image super-resolution (SISR) utilize end-to-end trained networks for direct estimation of high-quality images from corrupted measurements. Blind methods perform deconvolution to remove unwanted blur on corrupted input images to improve their quality [33], [34]. In contrast, CalibFPA employs deconvolution to correct LR measurements in SLM-based multiplexed FPA imaging by accounting for the physical properties of the imaging system (i.e., the SLM pattern and the lens PSF

whose shape is controlled via a radius parameter), and this physics-driven network can improve calibration reliability. An alternative framework for HR imaging based on LR measurements is based on SISR methods that do not use any special spatial encoding for measurements [34]–[36]. Unlike SISR methods, CalibFPA utilizes multiple measurements gathered with different SLM patterns for multiplexed encoding, and HR images are then recovered from these measurements [1].

III. THEORY

A. Compressive FPA based on Offline Calibration

In a compressive FPA system, an incident scene is first focused onto a coded aperture for multiplexed encoding of HR information onto LR measurements (Fig. 2). The SLM pattern is expressed as a diagonal matrix $\mathbf{\Lambda} \in R^{N \times N}$, with entries "1" for transparent pixels and "0" for opaque pixels. After encoding, a relay lens focuses the scene onto the LR FPA for sensing. Assuming linear shift invariance, the optical effect of the imaging lens can be modeled via a PSF h_I and that for the relay lens via a PSF h_R [24], [37]:

$$\mathbf{y}_i = \mathbf{D}(h_R \circledast (\mathbf{\Lambda}_i(h_I \circledast \mathbf{x}))) + \mathbf{n}_i, \qquad (1)$$

where $\mathbf{x} \in R^N$ is the HR image vector, $\mathbf{y}_i \in R^M$ are LR measurements, $\mathbf{n}_i \in R^M$ is the noise vector, $\circledast$ is the convolution operator, and $\mathbf{D} \in R^{M \times N}$ is the box-downsampling matrix. M is the LR sensor size and N is the HR image size, and i denotes the measurements taken with the ith coded aperture pattern, $\mathbf{\Lambda}_i$. For a total of m patterns, the resultant super-resolution factor corresponds to $N/M = s$ and the compression ratio is $CR = m/s$ [38].

The cumulative effects of downsampling, SLM patterns and lens PSFs can be described via a system matrix to express the forward model in Eq. (1) as follows:

$$\mathbf{y}_i = \mathbf{C}_i \mathbf{x} + \mathbf{n}_i, \qquad (2)$$

where $\mathbf{C}_i \in R^{M \times N}$ denotes the component of the system matrix that captures the relation between the HR scene and the LR sensor for the ith pattern [39], [40]. In the offline calibration framework, the system matrix components are measured via separate calibration scans at the expense of prolonged experiment times. In a common approach, a uniformly illuminated calibration scene is used to acquire calibration measurements independently for each SLM pattern. These measurements are then aggregated to form the overall system matrix $\mathbf{C} = [\mathbf{C}_1^T \cdots \mathbf{C}_i^T \cdots \mathbf{C}_m^T]^T \in R^{mM \times N}$. Once $\mathbf{C}$ is measured, image reconstruction can then be formulated as:

$$\arg\min_{\mathbf{x}} R(\mathbf{x}) \text{ s.t. } \|\mathbf{C}\mathbf{x} - \mathbf{y}\| \le \epsilon \qquad (3)$$

$R(\cdot)$ is a regularizer that reflects prior information on the image $\mathbf{x}$ and ϵ is a parameter associated with the noise level. Meanwhile, $\mathbf{y} = [\mathbf{y}_1^T \cdots \mathbf{y}_i^T \cdots \mathbf{y}_m^T]^T \in R^{mM}$ denote LR measurements of the visual scene. Solution of Eq. (3) typically requires multiple matrix-vector multiplication operations with the dense $\mathbf{C}$ matrix, resulting in $O(mMN)$ complexity.
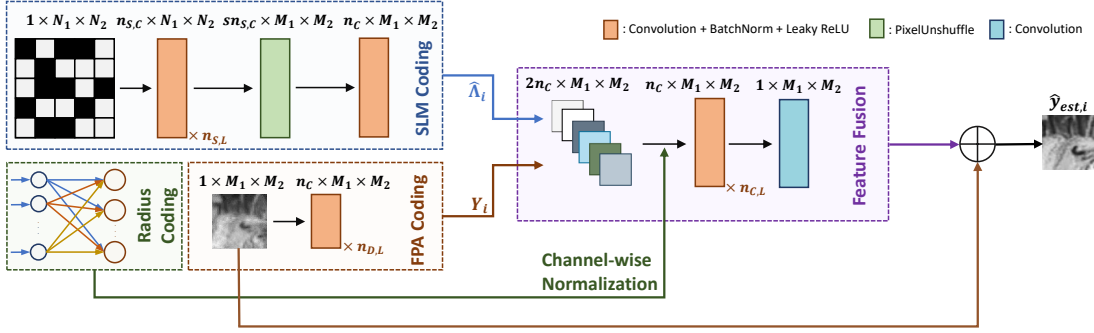
Fig. 3: CalibFPA comprises a network architecture with four blocks. The radius coding block receives the Airy disk radius of the relay lens, and computes radius-dependent latent variables via a multi-layer perceptron. The SLM coding block receives the HR SLM pattern and convolutionally encodes it onto LR feature maps. The FPA coding block receives multiplexed LR measurements and convolutionally encodes them onto LR feature maps. The SLM and FPA feature maps are concatenated and weighted according to latent variables from the radius coding block. Afterwards, a feature fusion block equipped with a long-range skip connection from the LR measurements maps the resultant feature map onto corrected measurements.

B. Proposed Method: CalibFPA

Online Calibration: CalibFPA accounts for undesired influences on measured data $\mathbf{y}_i$ due to optical aberrations from the relay lens in order to output corrected data $\hat{\mathbf{y}}_i$. Here, we consider an ideal imaging lens with $h_I = \delta$ as its effects on the system are often relatively moderate [14]. The relay lens, on the other hand, typically induces significant blur on the encoded scene. Its PSF can be described as an Airy disk function with radius $r = 1.22\lambda f/d$ where λ is the wavelength of the light, f is the focal length, d is the diameter of the lens aperture, and f/d corresponds to the F/# of the relay lens [15]. Since $\mathbf{D}$ is a box-downsampling operator, the role of $\mathbf{D}$ in Eq. (1) can be expressed as a strided convolution with box-filter h_B and stride of s:

$$\mathbf{y}_i = h_B \circledast_{s-strided} (h_R \circledast (\mathbf{\Lambda}_i\mathbf{x})) + \mathbf{n}_i, \quad (4)$$

$$= \delta \circledast_{s-strided} (h_B \circledast h_R \circledast (\mathbf{\Lambda}_i\mathbf{x})) + \mathbf{n}_i, \quad (5)$$

where δ is the two-dimensional (2D) Dirac-delta function. Online calibration aims to compensate the original measurements for the relay lens effects in order to obtain a corrected set of measurements $\hat{\mathbf{y}}_i$ that approach the ground truth $\mathbf{y}_{ref,i}$:

$$\mathbf{y}_{ref,i} = \delta \circledast_{s-strided} (h_B \circledast (\mathbf{\Lambda}_i\mathbf{x})). \quad (6)$$

Here, we make the key observation that $\mathbf{y}_{ref,i}$ and $\mathbf{y}_i$ are related through a convolution, so estimating $\mathbf{y}_{ref,i}$ from $\mathbf{y}_i$ does not require knowledge of the HR scene. Thus, a deconvolution based on h_R and $\mathbf{\Lambda}_i$ can be adopted to perform the correction.

CalibFPA employs a deep architecture to perform online calibration. The network inputs are LR measurements $\mathbf{y}_i \in R^{1 \times M_1 \times M_2}$, diagonal elements of the coded aperture pattern reshaped into 2D format $\mathbf{\Lambda}_{2D,i} \in R^{1 \times N_1 \times N_2}$ and a one-hot encoding of the $r \in R$ value of the Airy disk, with $M_1M_2 = M; N_1N_2 = N$. The network then comprises four main blocks for performing correction.

1) Radius Coding: Since precise estimation of the Airy disk radius with data from an LR sensor is difficult, we pre-defined 9 radius intervals: $1.5 < r < 2.5, 2.5 < r < 3.5, \cdots, 9.5 < r < 10.5$. A one-hot encoding vector $\mathbf{o}_i \in R^9$ is used to indicate the specific interval that r belongs to. The one-hot

vector is input to a multi-layer perceptron (MLP) to derive latent variables $\mathbf{r}_i \in R^{2n_C}$:

$$\mathbf{r}_i = \text{MLP}(\mathbf{o}_i). \quad (7)$$

These latent variables are later used to normalize the scale of feature maps in the feature fusion block as in Eq. (12). Thus, radius coding enables multi-tasking where a single network corrects measurements for all intervals.

2) SLM Coding: To encode the HR SLM pattern, a network with $n_{S,L}$ convolutional layers and $n_{S,C}$ channels is used:

$$\mathbf{SLM}_i = c_{s,n_{S,L}}(\cdots(c_{s,1}(\mathbf{\Lambda}_{2D,i}))), \quad (8)$$

where $c_{\cdot,L}(\cdot)$ represents convolution followed by batch normalization and leaky rectified linear unit (ReLU) activation function. As $\mathbf{SLM}_i \in R^{n_{S,C} \times N_1 \times N_2}$ is $s_1 \times s_2$ (with $N_1/M_1 = s_1, N_2/M_2 = s_2$) times higher in dimensionality compared to $\mathbf{Y}_i$, a pixel unshuffler is used to rearrange the input feature tensor such that the spatial dimensions are reduced while the number of channels are increased [35]:

$$\mathbf{SC}_i = \text{PixelUnshuffler}(\mathbf{SLM}_i) \in R^{(sn_{S,C}) \times M_1 \times M_2}. \quad (9)$$

Finally, the number of channels is reduced to n_C via a convolution layer:

$$\hat{\mathbf{\Lambda}}_i = c_{reduce}(\mathbf{SC}_i) \in R^{n_C \times M_1 \times M_2}. \quad (10)$$

3) FPA Coding: LR measurements are encoded via a network of $n_{D,L}$ convolutional layers, n_C channels:

$$\mathbf{Y}_i = c_{D,n_L}(\cdots(c_{D,1}(\mathbf{y}_i))) \in R^{n_C \times M_1 \times M_2}. \quad (11)$$

4) Feature Fusion: Encoded SLM and FPA feature maps, $\mathbf{Y}_i$ and $\mathbf{SLM}_i$, are first concatenated and then scaled with the latents from the radius coding:

$$\mathbf{DC}_i = \text{Concat}(\mathbf{Y}_i, \mathbf{SLM}_i)/\mathbf{r}_i, \quad (12)$$

where $\mathbf{DC}_i \in R^{2n_C \times M_1 \times M_2}$. Resultant feature maps are then projected across $n_{C,L}$ convolutional layers with n_C channels. The final output is computed based on the computed feature maps residually combined with the input measurements with a final convolutional layer without any activation, denoted as $c_F(\cdot)$ that reduces the number of channels to 1:

$$\hat{\mathbf{y}}_i = \mathbf{y}_i + c_F(c_{C,n_{C,L}}(\cdots(c_{C,1}(\mathbf{DC}_i)))). \quad (13)$$

Algorithm 1 ADMM-based algorithm for FPA reconstruction

Initialize $\mathbf{z}_0^{(j)}$ and $\mathbf{d}_0^{(j)}$ for $j = 0, 1$, choose μ, set $n \leftarrow 0$
while Stopping criterion is not satisfied **do**
   $\mathbf{x}_{n+1} \leftarrow (\mathbf{I} + \mathbf{C}_{bd}^T \mathbf{C}_{bd})^{-1}(\mathbf{C}_{bd}^T(\mathbf{z}_n^{(0)} + \mathbf{d}_n^{(0)}) + \mathbf{z}_n^{(1)} + \mathbf{d}_n^{(1)})$
   $\mathbf{z}_{n+1}^{(0)} \leftarrow \Psi_{\iota_{E(\epsilon, \mathbf{I}, \mathbf{y})}}(\mathbf{C}_{bd}\mathbf{x}_{n+1} - \mathbf{d}_n^{(0)})$
   $\mathbf{z}_{n+1}^{(1)} \leftarrow f_R(\mathbf{x}_{n+1} - \mathbf{d}_n^{(1)}; \mu)$
   $\mathbf{d}_{n+1}^{(0)} \leftarrow \mathbf{d}_n^{(0)} + \mathbf{z}_{n+1}^{(0)} - \mathbf{C}_{bd}\mathbf{x}_{n+1}$
   $\mathbf{d}_{n+1}^{(1)} \leftarrow \mathbf{d}_n^{(1)} + \mathbf{z}_{n+1}^{(1)} - \mathbf{x}_{n+1}$
   $n \leftarrow n + 1$
end while

---

5) Training loss: Training is performed via an ℓ_1-loss:

$$\arg \min_\theta \|\hat{\mathbf{y}}_i - \mathbf{y}_{ref,i}\|_1, \tag{14}$$

where $\hat{\mathbf{y}}_i = g_\theta(\mathbf{\Lambda}_i, \mathbf{y}_i, r_i)$, g_θ is the overall network with parameters θ, and $\mathbf{y}_{ref,i}$ denotes reference data under ideal optical settings (i.e., $h_I = \delta, h_R = \delta$). The network architecture used $n_{D,L}=2$, $n_{S,L}=1$, $n_{C,L}=6$, $n_C=32$, $n_{S,C}=4$ selected based on validation performance.

Image Reconstruction: In CalibFPA, image reconstruction is performed via a user-preferred algorithm based on corrected measurements and a system matrix $\mathbf{C}_{bd}$ of block-diagonal form. Since measurements have been corrected for optical aberrations, $\mathbf{C}_{bd}$ can be analytically expressed to describe the effects of downsampling and SLM patterns [39], [40]. Each block in $\mathbf{C}_{bd}$ belongs to an LR-pixel in the FPA sensor, so $\mathbf{C}_{bd}$ contains M blocks of dimensions $m \times s^2$. In the block for a given LR-pixel, the row dimensions correspond to separate snapshots, and the ith row contains values of the vectorized SLM pattern Λ_i for the $s \times s$ HR-pixels within the LR-pixel. Thus, reconstructions can be efficiently performed via matrix-vector multiplications of $O(mN)$ complexity to solve:

$$\arg \min_{\mathbf{x}} R(\mathbf{x}) \text{ s.t. } \|\mathbf{C}_{bd}\mathbf{x} - \hat{\mathbf{y}}\| \leq \epsilon \tag{15}$$

Here, we demonstrate CalibFPA based on a least-squares reconstruction as a traditional method, and a plug-and-play reconstruction and an unrolled reconstruction as learning-based methods [25]–[28]. Plug-and-play and unrolled reconstructions are implemented via an ADMM algorithm summarized in Alg. 1, where $\mathbf{z}_n^{(0)}, \mathbf{z}_n^{(1)}$ are dual variables, $\mathbf{d}_n^{(0)}, \mathbf{d}_n^{(1)}$ are Lagrange multipliers for data consistency and regularization steps at iteration n, $\Psi_{\iota_{E(\epsilon, \mathbf{I}, \mathbf{y})}}(\cdot)$ is the proximal operator for data consistency, and $f_R(\cdot)$ is the Moreau proximal operator for the regularization prior $R(\cdot)$. Plug-and-play and unrolled methods primarily differ in the training procedure of $f_R(\cdot)$. Detailed descriptions of the reconstruction methods are provided in Supp. Sec. I-A–I-C, respectively.

IV. METHODS

A. Optical Setup

To experimentally demonstrate CalibFPA, we built the in-house imaging system illustrated in Fig. 4. A custom SWIR LED light sphere was used for illumination. A Thorlabs AC508-250-C-ML and AC508-100-C-ML lens pair focused the incident scene onto the coded aperture, which was lithographically printed on a soda lime substrate at METU MEMS
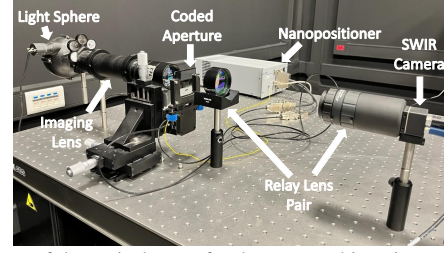

Fig. 4: A photo of the optical setup for the proposed imaging system. A scene illuminated with a SWIR LED is first focused onto a coded aperture via an imaging lens, and then focused onto the FPA sensor via a relay lens.

TABLE I: Calibration performance of raw measurements (uncorrected), CalibFPA variants (w/o SLM, Strided, w/o Rad) and a benchmark variant (CalibFPA-r) for $s=5\times5$. pSNR as mean±std across the validation set for separate ranges of Airy disk radius (r) values. Actual r values used to simulate measurements and assumed r values input to CalibFPA variants were matched.

r	Raw	w/o SLM	Strided	w/o Rad	**CalibFPA**	CalibFPA-r
1.5-2.5	33.8±0.8	37.0±0.9	47.7±2.8	46.4±2.8	48.4±3.3	**49.0±3.7**
2.5-3.5	30.9±0.9	33.8±0.9	44.9±3.2	44.2±3.4	46.0±3.9	**46.3±4.1**
3.5-4.5	29.2±0.9	32.4±0.9	43.0±3.3	42.6±3.7	44.1±4.1	**44.4±4.3**
4.5-5.5	27.9±1.0	31.7±0.9	41.7±3.3	41.7±3.9	42.8±4.2	**43.0±4.3**
5.5-6.5	26.9±1.0	31.1±0.9	40.4±3.3	40.6±3.9	41.5±4.1	**41.6±4.2**
6.5-7.5	26.0±1.1	30.4±0.9	39.2±3.3	39.5±3.9	40.2±4.1	**40.3±4.1**
7.5-8.5	25.2±1.1	29.7±0.9	38.0±3.4	38.3±3.9	38.9±4.1	**39.0±4.1**
8.5-9.5	24.6±1.1	29.0±0.9	36.9±3.4	37.0±3.8	37.6±4.0	**37.7±4.1**
9.5-10.5	24.0±1.2	28.3±0.9	35.7±3.5	35.8±3.7	36.4±4.0	**36.6±4.0**

Center (Ankara, Turkiye). The coded aperture followed a random pattern of transparent and opaque HR pixels with a pitch of 15μm, and dimensions larger than the HR imaging grid (see Supp. Fig. 1 for the used aperture). Each 5×5 HR pixel block in the coded aperture corresponded to a single LR pixel on the FPA. Within each block, 20 transparent and 5 opaque pixels were selected randomly for an open ratio of $p=0.8$. A closed-loop piezo-stage with PI P622.2CD nanopositioner and PI E-727 Digital controller was used to locomote the coded aperture. Between consecutive measurements, the coded aperture was moved by 15μm (i.e., a single HR pixel) in the x-y plane. Following scene encoding, a Thorlabs AC508-250-C-ML and Edmund Optics 50mm SWIR fixed focal length lens pair was used to relay the scene onto the FPA sensor. The relay lens pair were spaced apart for $6\times$ magnification with a sensor pixel pitch of 2.5μm. An FPA sensor was used comprising Sensors Unlimited 320CSX and a high-frame rate SWIR camera with a resolution of 320×256 and pixel pitch of 12.5μm. Hence, the pitch ratio of the camera to the coded aperture resulted in a super-resolution factor of $s=5\times5$.

Alignment markers were used to align the coded aperture and the FPA given the mismatch between the respective pixel resolutions. Opaque markers of size 3×3 LR pixels with a single transparent LR pixel in the middle were placed in the top-left, top-right, bottom-left, and bottom-right parts of the coded aperture. Alignment was achieved by adjusting the tilt, decenter and position of the objective, the relay lens group and the coded aperture to maximize the ratio of incident light at the central versus peripheral pixels in each marker.

TABLE II: Calibration performance for $s=5\times5$, $p=0.8$, $4.5 < r < 8.5$ under difference in input SNR between training and test set. Rows denote the test SNR, columns denote the training SNR.

Input SNR	40	50	60	70
40	**34.8±2.6**	31.9±2.2	31.4±2.3	31.2±2.4
50	37.3±3.1	**38.8±3.4**	38.2±3.2	38.0±3.1
60	37.5±3.1	40.1±4.0	**40.3±4.1**	**40.3±4.1**
70	37.6±3.1	40.2±4.1	**40.6±4.3**	**40.6±4.4**

TABLE III: Plug-and-play reconstruction performance as a function of test input SNR (rows) and training input SNR (columns) for $s=5\times5$, $m=5$, $p=0.8$.

Input SNR	Metric	40	50	60	70
40	pSNR	**17.8±2.7**	14.1±1.6	13.7±1.6	13.6±1.7
	SSIM	**41.3±11.5**	16.0±5.8	14.2±5.6	13.8±5.7
50	pSNR	**24.2±4.2**	23.1±3.9	22.0±3.5	21.7±3.5
	SSIM	**70.2±12.1**	66.5±11.0	57.1±10.7	53.3±11.4
60	pSNR	25.0±4.3	**25.9±4.6**	25.8±4.7	25.8±4.7
	SSIM	72.7±11.9	77.3±10.8	**77.8±11.4**	77.6±10.7
70	pSNR	25.1±4.3	26.0±4.6	26.1±4.7	**26.2±4.7**
	SSIM	72.9±11.8	77.6±10.7	78.5±10.5	**78.7±10.5**

TABLE IV: Calibration performance for $s=5\times5$, $p=0.8$, input SNR=60dB under mismatch between assumed and actual r values. Rows indicate the actual r range used to simulate measurements and columns indicate the assumed r range input to CalibFPA.

r	4.5-5.5	5.5-6.5	6.5-7.5	7.5-8.5
4.5-5.5	**42.2±4.0**	39.0±3.0	33.0±1.9	28.2±1.4
5.5-6.5	38.7±3.3	**41.0±3.9**	38.1±3.0	32.2±1.8
6.5-7.5	34.3±2.3	37.8±3.3	**39.7±3.9**	37.2±3.1
7.5-8.5	31.3±2.0	33.6±2.4	36.7±3.3	**38.4±3.8**

TABLE V: Plug-and-play reconstruction performance as a function of actual (rows) and assumed (columns) r values for $s=5\times5$, $m=5$, $p=0.8$, input SNR=60dB.

r	Metric	4.5-5.5	5.5-6.5	6.5-7.5	7.5-8.5
4.5-5.5	pSNR	**26.4±4.7**	23.7±3.8	17.2±1.7	13.8±1.1
	SSIM	**80.0±9.9**	69.6±7.5	40.5±10.2	26.4±10.2
5.5-6.5	pSNR	24.4±3.9	**26.1±4.7**	23.2±3.7	16.5±1.6
	SSIM	69.3±8.4	**78.8±10.3**	67.7±7.8	37.5±10.2
6.5-7.5	pSNR	19.4±2.0	24.1±3.9	**25.7±4.7**	22.6±3.7
	SSIM	44.7±9.7	68.5±9.1	**77.2±10.7**	66.1±8.5
7.5-8.5	pSNR	16.7±1.4	19.0±2.0	23.3±3.7	**25.1±4.6**
	SSIM	32.4±10.0	43.2±9.8	64.8±9.1	**75.1±11.3**

B. Datasets

1) Visible-band data: For model training and evaluation, a simulated dataset in the visible band was generated with measurement pairs $\hat{y}_i, y_{ref,i}$. Reference data y_{ref} were generated by assuming ideal optical settings and zero noise, i.e., by multiplying the HR image of the visual scene with C_{bd}. In simulations, incident scenes were taken as cropped natural images from the DIV2K dataset [41]. Randomly generated coded apertures Λ_i and a randomly selected $1.5 < r < 10.5$ value for the Airy disk were assumed for each scene. The ratio of transparent to opaque pixels in Λ_i was set as 4 to yield an open ratio of $p=0.8$ similar to the coded apertures used in the optical setup. A discrete PSF of size 81×81 was used to implement the Airy disk. White Gaussian noise was added at a standard deviation corresponding to a given input SNR level, taken as the ratio of maximum pixel intensity to standard deviation (std) of noise.

2) SWIR-band data: A simulated dataset was generated based in the SWIR-band with measurement pairs $\hat{y}_i, y_{ref,i}$ [42]. Similar simulation settings were assumed with the case of visible-band data, albeit incident scenes were taken as scenes provided in [42].

3) Experimental data: An experimental dataset was collected on the optical setup described in Sec. IV-A. A visual scene containing black-colored digits 1-5 in varying sizes against a gray background was used (Supp. Fig. 2). A total of $m=49$ snapshots were used.

C. Competing Methods

We demonstrated CalibFPA against several competing online calibration methods. For all online calibration methods, reconstructions of corrected measurements were performed via least-squares, plug-and-play or unrolled algorithms based on C_{bd}. A single image super-resolution method for direct reconstruction, and reconstructions based on an offline calibration method were also considered.

1) Raw: Original measurements were not subjected to any correction and directly used in reconstruction.

2) Lucy-Richardson: Original measurements were deconvolved using the LR PSF for the Airy disk using 20 iterations [43], and corrected measurements were reconstructed.

3) Blind deconvolution: A deep-learning-based blind deconvolution method was used to correct measurements [33]. A separate network was trained for each Airy disk radius. Corrected measurements were reconstructed.

4) SwinIR: A deep-learning-based blind deconvolution method based on the Swin transformer architecture was used to correct measurements [34]. A separate network was trained for each Airy disk radius. Corrected measurements were reconstructed.

5) Single image super-resolution (SISR): A deep-learning-based SISR method was employed to recover HR images by upsampling LR measurements acquired without using any SLM patterns [35]. SISR was implemented based on the Swin transformer architecture [34].

6) Offline calibration: To attain a benchmark on image quality, the original uncorrected measurements were reconstructed based on the dense system matrix C. To ensure tractability of the reconstructions, analyses were restricted to 120×120 test images.

D. Implementation Details

All calibration methods were implemented in PyTorch. Training for online calibration methods was conducted via the loss function in Eq. (14). A total of (5513, 200, 199) measurement pairs were simulated for the training, validation, and test sets, respectively. Due to memory considerations, images were cropped to 180×180 for the training/validation sets, albeit to 360×360 for the test set. Training was performed via the Adam optimizer with a batch size of 64, an initial learning rate of 10^{-3}, number of epochs 1000. The learning rate was scaled by a factor of 0.999 after each epoch. For each competing method, hyperparameters were selected to maximize validation performance.

Image reconstruction methods were also implemented in PyTorch. The least-squares method was not trained. The plug-and-play method employed an off-the-shelf denoiser, taken as the pretrained DnCNN network in PyTorch [44]. The unrolled method was trained via ℓ_1-loss between the reconstructed images and reference incidents scenes. All reconstruction methods employed the ADMM algorithm, and $(\mathbf{I} + \mathbf{C}^T\mathbf{C})^{-1}$ was precomputed for each method for speeding up inference. Each online calibration method can elicit a different level of residual noise in the corrected measurements it passes onto the reconstruction algorithm. Thus, ϵ in the ADMM algorithm was selected independently for each calibration method based on calibration performance on the validation set.

Demonstrations were performed on both simulated and experimental data. As the ground-truth images are known for simulated data, quantitative assessments were reported via peak SNR (pSNR) and structural similarity index (SSIM) metrics. SSIM values were listed following a factor of 100 scaling for ease of legibility. Except for the Airy disk radius analysis, results were averaged over $4.5 < r < 8.5$ with $m = 5$ prior to reporting. For estimating r that CalibFPA expects as input for experimental data, multiple least-squares reconstructions were performed using measurements corrected for nine separate intervals of r, and the interval that yielded the lowest artifacts among the candidates was visually selected. Since ground-truth is unavailable in experimental settings, only qualitative assessments were performed.

V. Results

A. Ablation Studies for CalibFPA

We first validated the design elements in CalibFPA through a set of ablation studies on the simulated visible-band dataset. To assess the importance of SLM coding, we formed a "w/o SLM" variant that omitted the coded aperture. To assess the importance of pixel unshuffling in downsampling, we formed a "Strided" variant that replaced the unshuffling block with a strided convolution block. To assess the importance of radius coding block, we formed a "w/o Rad" variant that omitted the radius coding block. To examine the efficacy of multi-task learning across a range of Airy disk radius values, we formed a benchmark "CalibFPA-r" variant that was trained separately at each r value to capture an upper performance bound. Table I lists performance metrics for variant models at $s=5\times5$. As expected, we find a notable loss in the quality of measurements with the raw method, particularly towards higher Airy disk radii. While all variants significantly outperform "Raw" uncorrected measurements, CalibFPA achieves the highest overall performance that approaches the performance of the benchmark "CalibFPA-r". On average, CalibFPA attains improvements of 10.3dB over "w/o SLM", 1.0dB over "Strided", 1.1dB over "w/o Rad". These results suggest that SLM coding, pixel unshuffling and radius coding all contribute to the calibration performance of the proposed method. Note that there is only moderate performance difference between CalibFPA and CalibFPA-r, suggesting that CalibFPA effectively multi-tasks across r values without requiring training of separate networks for each radius value.

TABLE VI: Calibration performance of as a function of r for $s=5\times5$, $p=0.8$, input SNR=60dB. pSNR as mean±std across the test set.

r	Raw	Lucy-Rich.	Blind	SwinIR	**CalibFPA**
1.5-2.5	34.9±1.1	35.2±1.1	38.6±1.2	38.7±1.2	**48.0±3.5**
2.5-3.5	31.9±1.2	32.5±1.2	35.4±1.3	35.4±1.2	**45.4±4.0**
3.5-4.5	30.1±1.2	31.1±1.2	33.8±1.3	33.9±1.3	**43.5±4.1**
4.5-5.5	28.8±1.3	30.3±1.2	33.1±1.3	33.2±1.3	**42.2±4.0**
5.5-6.5	27.7±1.3	29.3±1.2	32.4±1.3	32.5±1.3	**40.9±3.9**
6.5-7.5	26.8±1.4	29.8±1.2	31.7±1.2	31.7±1.3	**39.7±3.8**
7.5-8.5	26.0±1.4	29.5±1.2	30.9±1.3	31.0±1.3	**38.4±3.8**
8.5-9.5	25.3±1.4	28.6±1.3	30.1±1.3	30.2±1.3	**37.1±3.7**
9.5-10.5	24.7±1.5	27.6±1.3	29.4±1.3	29.5±1.3	**35.9±3.6**

TABLE VII: Least-squares reconstruction performance as a function of r for $s=5\times5$, $m=25$, $p=0.8$, input SNR=60dB.

r	Metric	Raw	Lucy-Rich.	Blind	SwinIR	**CalibFPA**
1.5-2.5	pSNR	17.5±1.1	17.6±1.1	19.5±1.1	19.5±1.1	**24.7±2.5**
	SSIM	35.8±11.6	36.3±11.7	45.1±12.3	44.9±12.3	**68.6±6.2**
2.5-3.5	pSNR	15.3±1.1	15.6±1.1	17.1±1.1	17.1±1.1	**24.4±2.6**
	SSIM	26.8±10.1	28.1±10.4	35.4±11.6	35.6±11.8	**66.7±5.7**
3.5-4.5	pSNR	14.1±1.1	14.6±1.1	15.9±1.1	16.0±1.1	**24.1±2.6**
	SSIM	21.8±8.8	24.1±9.4	30.8±10.9	30.9±11.0	**64.9±5.3**
4.5-5.5	pSNR	13.4±1.1	13.9±1.1	15.4±1.1	15.4±1.1	**23.8±2.6**
	SSIM	18.4±7.8	21.6±8.8	28.7±10.6	28.7±10.6	**63.1±5.0**
5.5-6.5	pSNR	12.8±1.1	13.4±1.1	14.9±1.1	14.9±1.1	**23.5±2.6**
	SSIM	15.7±7.0	19.2±8.1	26.2±9.8	26.7±10.2	**61.1±4.7**
6.5-7.5	pSNR	12.3±1.1	13.5±1.1	14.4±1.1	14.4±1.1	**23.1±2.6**
	SSIM	13.5±6.3	20.2±8.3	24.1±9.3	24.1±9.3	**58.2±4.3**
7.5-8.5	pSNR	11.8±1.1	13.2±1.1	13.8±1.1	13.8±1.1	**22.6±2.6**
	SSIM	11.8±5.7	19.3±7.8	21.6±8.7	21.8±8.6	**54.6±3.9**
8.5-9.5	pSNR	11.5±1.1	12.7±1.1	13.3±1.1	13.3±1.0	**22.1±2.6**
	SSIM	10.4±5.3	17.5±7.3	19.3±7.8	19.6±8.1	**50.9±3.4**
9.5-10.5	pSNR	11.3±1.1	12.2±1.1	12.8±1.0	12.8±1.0	**21.4±2.6**
	SSIM	9.3±5.0	15.5±6.7	17.2±7.1	17.8±7.7	**46.4±3.0**

B. Reliability of CalibFPA

We conducted a set of analyses on the visible-band dataset to assess the reliability of correction and encoding strategies in CalibFPA. Results on the effects of differences in input SNR values between training and test sets, differences in r values between training and test sets, mismatch between actual r values used to obtain measurements and assumed r values input to CalibFPA, and locomoted versus random SLM patterns are given below.

1) Training-test Differences in Input SNR: To assess reliability against input SNR variations, models trained on data at a given input SNR level were evaluated on data at different SNR levels. Table II lists calibration performance, and Table III lists plug-and-play reconstruction performance. Naturally, training-test differences in input SNR induce performance losses. Yet, for a ±10dB variation in test SNR relative to the training SNR, we find that the average performance loss is 0.9dB pSNR in calibration, and 0.6dB pSNR, 5.3% SSIM in reconstruction.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2024.3477312

8

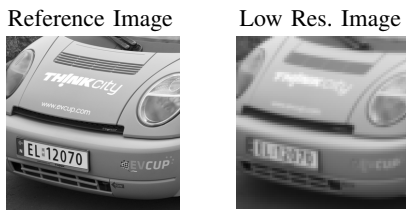Reference Image   Low Res. Image

Fig. 5: A representative image pair depicting the high-resolution reference and low-resolution image from the simulated visible-band dataset.

TABLE VIII: Plug-and-play reconstruction performance as a function of r for $s=5\times5$, $m=5$, $p=0.8$, input SNR=60dB.

r	Metric	Raw	Lucy-Rich.	Blind	SwinIR	SISR	**CalibFPA**
1.5-2.5	pSNR	18.3±1.6	18.0±1.5	19.8±1.8	19.8±1.8	24.2±4.6	**27.8±4.7**
	SSIM	44.9±11.8	42.9±11.7	51.9±11.4	51.8±11.4	69.5±15.3	**83.7±8.3**
2.5-3.5	pSNR	16.5±1.4	16.3±1.3	17.6±1.5	17.6±1.5	24.4±4.7	**27.2±4.8**
	SSIM	36.7±12.0	35.6±11.9	43.1±12.0	43.3±12.0	70.2±15.3	**82.4±9.0**
3.5-4.5	pSNR	15.6±1.3	15.5±1.2	16.6±1.4	16.6±1.4	24.0±4.5	**26.7±4.8**
	SSIM	32.4±12.0	32.0±11.8	39.2±12.1	39.4±12.2	68.5±15.5	**81.1±9.6**
4.5-5.5	pSNR	15.0±1.3	15.0±1.2	16.3±1.3	16.3±1.3	24.5±4.8	**26.4±4.7**
	SSIM	29.6±12.0	29.8±11.8	37.9±12.3	37.9±12.2	70.3±15.3	**80.0±9.9**
5.5-6.5	pSNR	14.6±1.3	14.5±1.2	15.9±1.3	15.9±1.3	24.5±4.8	**26.1±4.7**
	SSIM	27.3±11.9	27.5±11.6	36.4±12.2	36.5±12.2	70.2±15.3	**78.8±10.3**
6.5-7.5	pSNR	14.2±1.3	14.7±1.2	15.6±1.3	15.6±1.3	24.5±4.7	**25.7±4.7**
	SSIM	25.3±11.7	29.5±11.7	34.8±12.3	34.6±12.3	70.0±15.3	**77.2±10.7**
7.5-8.5	pSNR	13.9±1.3	14.5±1.2	15.2±1.3	15.2±1.3	24.3±4.6	**25.1±4.6**
	SSIM	23.7±11.6	29.0±11.7	32.6±12.2	32.8±12.2	68.7±15.5	**75.0±11.3**
8.5-9.5	pSNR	13.7±1.3	14.1±1.2	14.8±1.2	14.8±1.2	24.3±4.6	**24.5±4.5**
	SSIM	22.3±11.5	27.2±11.6	30.8±12.1	30.8±12.2	68.1±15.6	**72.4±12.0**
9.5-10.5	pSNR	13.5±1.3	13.7±1.2	14.4±1.2	14.5±1.2	**24.2±4.6**	23.6±4.4
	SSIM	21.2±11.3	25.2±11.5	28.8±12.0	29.3±12.2	68.0±15.8	**69.3±12.7**

These results suggest that CalibFPA offers notable reliability against input SNR variations between training-test sets.

2) Training-test Differences in r: To assess reliability against r variations, models trained on data at a given r were evaluated on data at different r values. Supp. Table I lists calibration performance, and Supp. Table II lists plug-and-play reconstruction performance. For a ±1 unit variation in test relative to training r values, we find an average performance loss of 2.2dB pSNR in calibration, and 2.1dB pSNR, 9.8% SSIM in reconstruction. These results suggest that CalibFPA offers a modest degree of reliability against r variations between training-test sets.

3) Inaccuracy in r: Inaccurate estimation of the Airy disk radius might influence the accuracy of measurement corrections in CalibFPA, so we evaluated the influences of mismatch between assumed and actual r values. Table IV lists calibration performance, and Table V lists plug-and-play reconstruction performance. On average, we find that a one-unit mismatch in r (equivalent to 1 HR pixel) leads to performance losses of 2.4dB pSNR in calibration, and 2.3dB pSNR, 10.1% SSIM in reconstruction. Hence, identifying the correct radius interval can be critical to attain optimal performance.

4) Locomoted versus Random SLM Patterns: CalibFPA mechanically moves the fixed coded aperture by 1 HR pixel in the

TABLE IX: Calibration performance as a function of input SNR for $s=5\times5$, $p=0.8$, $4.5 < r < 8.5$.

SNR	Raw	Lucy-Rich.	Blind	SwinIR	**CalibFPA**
40	27.0±1.6	28.7±1.3	29.9±1.4	29.9±1.4	**34.8±2.6**
50	27.3±1.7	29.6±1.3	31.7±1.5	31.7±1.5	**38.8±3.4**
60	27.3±1.7	29.7±1.3	32.1±1.5	32.1±1.5	**40.3±4.1**
70	27.3±1.7	29.8±1.3	32.1±1.5	32.2±1.5	**40.6±4.4**

TABLE X: Plug-and-play reconstruction performance as a function of input SNR for $s=5\times5$, $p=0.8$, $4.5 < r < 8.5$.

SNR	Metric	Raw	Lucy-Rich.	Blind	SwinIR	SISR	**CalibFPA**
40	pSNR	13.4±1.1	13.5±1.1	13.8±1.1	13.8±1.1	**23.9±4.3**	17.8±2.7
	SSIM	17.0±7.6	15.5±6.4	21.9±9.7	21.4±9.5	**66.7±15.1**	41.3±11.5
50	pSNR	13.7±1.2	14.5±1.2	15.3±1.3	15.3±1.3	**24.4±4.6**	23.1±3.9
	SSIM	22.1±10.8	27.3±11.2	32.1±11.9	32.1±11.9	**69.2±15.2**	66.5±11.0
60	pSNR	14.4±1.3	14.7±1.2	15.7±1.4	15.7±1.4	24.4±4.7	**25.8±4.7**
	SSIM	26.5±12.0	28.9±11.7	35.4±12.4	35.4±12.4	69.8±15.3	**77.8±10.7**
70	pSNR	13.8±1.2	14.7±1.2	15.8±1.4	15.8±1.4	24.3±4.6	**26.2±4.7**
	SSIM	22.5±11.1	29.1±11.8	35.9±12.4	36.0±12.4	68.9±15.5	**78.7±10.5**

transverse plane to devise SLM patterns for multiplexed measurements. An alternative approach would be to devise random patterns via electronically controllable SLMs. To validate the proposed encoding strategy, variant models were trained based on multiplexed measurements acquired using random SLM patterns, and using SLM patterns moved by 1, 2, or 3 pixels. Supp. Table III lists plug-and-play reconstruction performance for variant models. In general, variants based on locomoted and random SLM patterns perform similarly, indicating the effectiveness of the encoding approach in CalibFPA.

C. Parameter Sensitivity of Competing Methods

Next, we conducted analyses on the visible-band dataset to examine the performance of competing methods across a broad range of imaging parameters (see Supp. Table IV for analysis settings). Results on the effects of Airy disk radius (r), input SNR, mismatch between assumed and actual SLM positions, and comparisons to offline calibration are given below.

1) Airy Disk Radius (r): We first evaluated the effect of r on calibration performance. The simulated imaging system assumed $s=5\times5$, $p=0.8$, and input SNR=60 dB. Table VI lists calibration performance for corrected measurements. Calibration performance expectedly drops with increasing r for all methods. Yet, CalibFPA achieves 8.4dB higher pSNR than the closest competing method on average, and the performance benefits are larger for higher r.

We also evaluated the effect of r on reconstruction performance. Table VII lists least-squares reconstruction performance for $m = 25$. On average, CalibFPA attains an improvement of 8.1dB pSNR, 31.8% SSIM over the top-contending method. Table VIII and Supp. Table V list plug-and-play and unrolled reconstruction performances for $m = 5$. In general, plug-and-play and unrolled reconstructions yield similar performance for a given online-calibration method. Compared to the second-best calibration method, CalibFPA attains improvements of 9.6dB pSNR, 40.4% SSIM on average based on plug-and-play reconstructions. Note that blind deconvolution methods (Blind, SwinIR) attempt to correct measurements for optical aberrations without any assumptions on

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2024.3477312
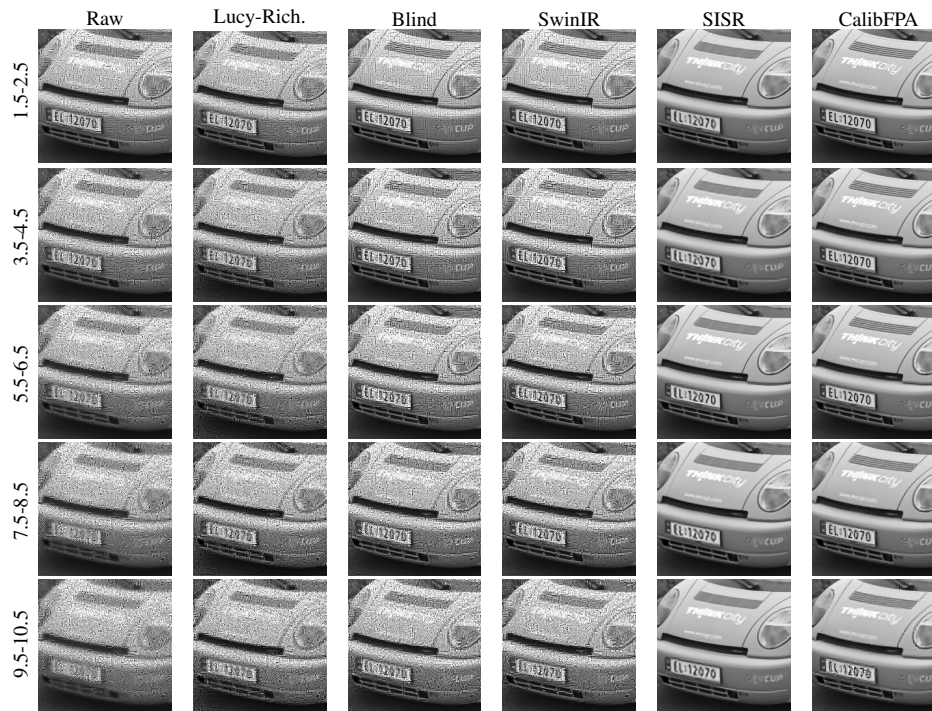
9

Fig. 6: Plug-and-play reconstructions of a visual scene from the simulated dataset for $m = 5$ snapshots and varying Airy disk radii. For SISR, LR images captured without any SLM pattern were super-resolved for a direct reconstruction.

the underlying SLM patterns, so they can confound influences from optical aberrations and SLM patterns, resulting in unwanted errors in corrected measurements. Such measurement errors can become more notable towards larger r values as the calibration problem becomes progressively ill-posed, and compromise image quality in subsequent reconstructions based on blind methods. In contrast, CalibFPA employs a physics-driven architecture informed about r and SLM patterns, so it can offer more accurate correction of optical aberrations and thereby higher reconstruction performance. Compared to SISR, CalibFPA achieves improvements of 1.6dB pSNR, 8.5% SSIM over SISR. CalibFPA yields higher performance than SISR across all r values, except for the range 9.5-10.5 where SISR yields modestly higher pSNR. Note that at relatively high r, the lens blur becomes a predominant factor that limits image quality, so it is expected that differences between SLM-based and super-resolution imaging become less notable.

Figure 5 displays reference visual scenes from the simulated dataset, and Fig. 6 shows respective images reconstructed via the plug-and-play method for $m = 5$ across varying r values. Additional results based on a distinct, relatively more complex visual scene are depicted in Supp. Figs. 3 and 4. In all cases, CalibFPA achieves visible improvements in spatial acuity and noise/artifact suppression against competing online calibration methods that suffer from strong stripe artifacts. While SISR shows similar effectiveness in noise suppression, CalibFPA yields higher spatial acuity than SISR as evident from the depiction of detailed image structure.

2) Input SNR: Next, we evaluated the effect of input SNR on calibration performance as listed in Table IX, and on plug-and-play and unrolled reconstruction performances as listed in Table X and Supp. Table VI, respectively. In general, all methods yield elevated performance towards higher input

TABLE XI: Calibration performance as a function of misalignment due to random SLM shifts for $s=5\times5$, $p=0.8$, input SNR=60dB. CalibFPA-0.5 and CalibFPA-1.0 were trained on data with respective amounts of random shifts.

Shift	CalibFPA	CalibFPA-0.5	CalibFPA-1.0
No Shift	**40.3±4.1**	39.8±3.9	39.5±3.8
-0.5/0.5	**39.7±3.7**	39.5±3.7	39.3±3.7
-1.0/1.0	38.4±3.0	38.5±3.2	**38.6±3.3**

SNR levels. Compared to the second-best calibration method, CalibFPA yields average performance improvements of 8.0dB in calibration, 9.4dB pSNR and 39.9% SSIM in plug-and-play reconstruction, and 8.8dB pSNR and 36.5% SSIM in unrolled reconstruction. In reconstructions, CalibFPA outperforms SISR across input SNRs larger than 50dB, albeit SISR yields similar or higher performance for lower input SNRs. The poor performance of plug-and-play reconstructions based on online calibration at low input SNRs can be attributed to the fact that the pre-trained denoiser network was originally trained on natural images with limited additive noise. Furthermore, recent studies suggest that multiplexed imaging with SLM patterns performs suboptimally under relatively high noise levels compared against imaging without SLM coding [38]. Thus, the suboptimal performance of online calibration methods against SISR at low input SNRs can also be attributed to the relatively limited SNR efficiency of SLM-based imaging at high noise levels.

3) Misalignment in SLM Position: Despite the use of alignment markers, residual sub-pixel shifts between the coded aperture and the FPA sensor can be present in an optical setup. Since these shifts reflect a difference between the assumed and actual position of the SLM pattern, they can compromise the accuracy of measurement corrections. We conducted an analysis to assess reliability against residual pixel shifts
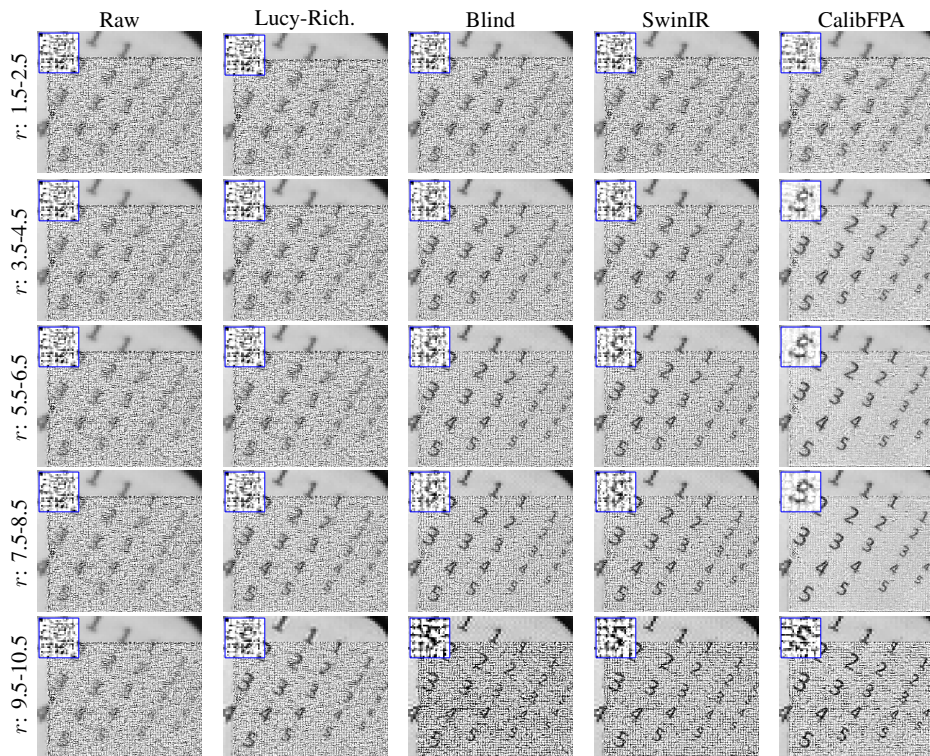
Fig. 7: Least-squares reconstruction of experimental data using competing methods for $m = 49$ snapshots and varying Airy disk radii.

TABLE XII: Plug-and-play reconstruction performance as a function of misalignment due to random SLM shifts for $s=5\times5$, $p=0.8$, input SNR=60dB.

Shift	Metric	CalibFPA	CalibFPA-0.5	CalibFPA-1.0
No Shift	pSNR	**25.8±4.7**	25.4±4.5	25.0±4.3
	SSIM	**77.8±10.7**	76.8±10.7	76.2±10.7
-0.5/0.5	pSNR	24.9±4.1	**24.9±4.2**	24.8±4.2
	SSIM	74.0±10.1	74.0±10.5	**74.1±10.8**
-1.0/1.0	pSNR	22.5±3.0	23.0±3.4	**23.3±3.6**
	SSIM	61.5±9.3	64.1±9.0	**66.7±10.0**

TABLE XIII: Plug-and-play reconstruction performance as a function of number of snapshots for $s=5\times5$, $p=0.8$, input SNR=60dB, $4.5 < r < 8.5$.

m	Metric	Offline-ideal	Offline-realistic	**CalibFPA**
1	pSNR	**21.8 ± 3.4**	21.7 ± 3.4	21.7 ± 3.4
	SSIM	**61.3 ± 13.5**	61.2 ± 13.1	60.7 ± 13.5
5	pSNR	**24.9 ± 3.3**	22.5 ± 2.5	22.9 ± 3.4
	SSIM	**78.4 ± 7.8**	67.8 ± 8.4	71.3 ± 9.6
10	pSNR	**27.3 ± 2.8**	23.6 ± 2.0	23.4 ± 3.1
	SSIM	**83.7 ± 5.1**	70.3 ± 10.2	74.2 ± 7.0
15	pSNR	**28.1 ± 2.3**	23.2 ± 1.7	23.0 ± 3.0
	SSIM	**83.7 ± 7.5**	68.5 ± 12.3	72.7 ± 6.2

within ranges of 0.5 and 1 HR pixels separately. Note that CalibFPA was trained under the assumption of no shifts between assumed and actual SLM positions. For comparison, we also trained variant models, CalibFPA-0.5 and CalibFPA-1.0, on simulated data with random spatial shifts in ranges of 0.5 and 1 pixels, respectively. Table XI lists calibration performance, and Table XII lists plug-and-play reconstruction performance for CalibFPA and variants. Corresponding results on two learning-based calibration methods, Blind and SwinIR, are listed in Supp. Tables VII-X. Naturally, all online

calibration methods examined incur moderate performance losses in calibration and reconstruction under shifts in SLM position. Yet, CalibFPA achieves similar levels of performance improvement over competing methods for the no-shift and shift scenarios. Furthermore, CalibFPA performs competitively with its variants directly trained on data with random spatial shifts, suggesting a degree of reliability against mismatches in SLM position.

4) Comparison to Offline Calibration: Finally, we evaluated CalibFPA against the offline calibration method that relies on measured system matrices. In theory, the measured system matrix would precisely reflect optical aberrations from the relay lens, so an "Offline-ideal" variant was formed using system matrices with precise knowledge of r while simulating measured data. In practice, however, a degree of imprecision is expected, so an "Offline-realistic" variant was formed under moderate differences between r values for the system matrix versus actual scene measurements, as also assumed for CalibFPA by specifying r in intervals as opposed to exact values. Table XIII lists plug-and-play reconstruction performance for CalibFPA and offline calibration variants across different numbers of snapshots. While all methods have elevated performance towards higher m, improvements become marginal beyond m=10 for CalibFPA and "Offline-realistic" that yield similar performances. These results suggest that online calibration can offer similar utility to conventional offline calibration in compressive FPA imaging.

D. Analyses on SWIR-band Data

Next, we conducted analyses to examine the performance of competing methods on a distinct dataset comprising SWIR-band data. First, within-domain performance was assessed
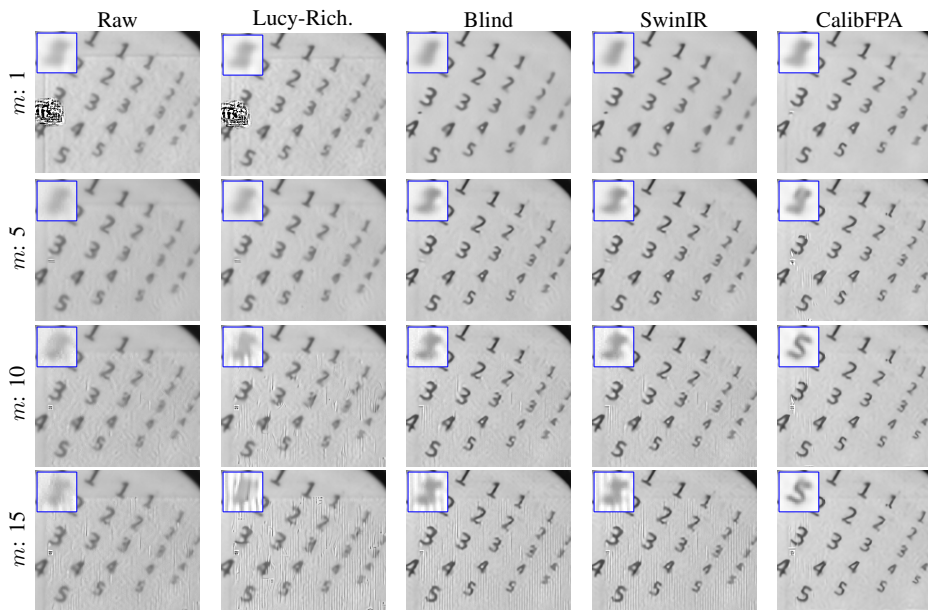
Fig. 8: Plug-and-play reconstruction of experimental data using competing methods for $5.5 < r < 6.5$ and varying number of snapshots. Zoomed-in displays of the region containing the smallest digit "5" in the scene are overlaid onto the upper-left section of each image.

by evaluating models trained on SWIR-band data. Supp. Table XI lists calibration performance of competing methods. On average, CalibFPA outperforms the second-best method by 11.3dB pSNR. Meanwhile, Supp. Table XII lists plug-and-play reconstruction performance of calibration methods along with direct reconstruction performance of SISR. On average, CalibFPA attains improvements of 14.4dB pSNR, 64.8% SSIM over the closest calibration method, and 1.6dB pSNR, 6.4% SSIM over SISR. Note that these improvement levels are generally consistent with those on visible-band data.

We then examined cross-domain performance by training models on visible-band data and evaluating them on SWIR-band data. Supp. Table XIII lists calibration performance of competing methods. In general, we find that cross-domain performance levels and the improvements that CalibFPA offers over competing methods are highly consistent with the within-domain case reported in Supp. Table XI. These results suggest that CalibFPA shows a reasonable degree of reliability against domain shifts in the data distribution.

E. Analyses on Experimental Data

We also demonstrated CalibFPA on experimental data captured using the in-house optical system operating in the SWIR regime. Qualitative evaluations were conducted as ground truth is unavailable in experimental settings. Fig. 7 displays images for varying r based on a least-squares reconstruction with $m = 49$ snapshots. Lucy-Richardson shows a degree of spatial blurring and noise amplification, and Blind and SwinIR suffer from streak artifacts and noise particularly towards larger radii. In contrast, CalibFPA yields improved image quality in reconstructions with higher spatial acuity for the numerical digits in the scenes, along with improved suppression of streaking artifacts and noise.

Next, we examined the performance of competing methods when coupled with the plug-and-play reconstruction for varying numbers of snapshots as shown in Fig. 8. All methods

show a moderate degree of improvement in spatial acuity for larger m, including direct reconstruction of raw data. Yet, Lucy-Richardson, Blind and SwinIR yield notable streaking artifacts and blurring across the visual scene. Instead, CalibFPA achieves higher quality reconstructions with visibly higher acuity and lower artifacts.

F. Runtime Analysis

In compressive FPA, the frame rate is influenced by the frame integration time for measurements, as well as the total runtime of calibration and reconstruction stages for subsequent image formation. Note that the integration time is a design choice that can be tuned based on application-specific requirements (e.g., desired SNR levels). Here we employed 16 ms integration time in experiments that would support up to 60 frames-per-second (FPS). For assessment of runtime, inference times per frame were measured on a Tesla V100 GPU for $m = 5$, 360×360 image size. In the calibration stage, inference times were 2.6 ms for Lucy-Richardson, 20.2 ms for Blind, 80.2 ms for SwinIR, and 0.9 ms for CalibFPA. In the reconstruction stage, inference times were 116.2 ms for plug-and-play, 33.1 ms for unrolled methods. Meanwhile, SISR resolved images directly from measurements in 52 ms. In comparison, CalibFPA combined with unrolled reconstruction can attain a runtime of 34 ms, thereby offering 30 FPS.

We also compared the runtime of CalibFPA against the offline calibration method. Given GPU memory limitations for the offline method, this analysis was conducted for an image size of 120×120, $m = 5$. The offline method involves a separate calibration scan of 230400 ms to measure $\mathbf{C}$, followed by a reconstruction stage with inference time of 33.6 ms for plug-and-play and 25.3 ms for unrolled methods. In comparison, CalibFPA does not require calibration scans, and it has a total calibration-reconstruction runtime of 19.1 ms for plug-and-play and 8.8 ms for unrolled methods.

VI. Discussion

Here we reported a low-cost imaging system that spatially encodes HR scenes via a locomoted coded aperture, and captures multiplexed measurements via an LR FPA. This system allows for a compact physical footprint while still enabling recovery of HR images of incident scenes. Unlike offline calibration, CalibFPA can mitigate biases from optical aberrations of the relay lens without the need to conduct separate scans to measure the system matrix. To compensate for the relay lens PSF, CalibFPA uniquely employs an online deep-leaning calibration that corrects LR measurements from the FPA. The benefits of online calibration might be particularly evident at higher wavelengths, e.g., in mid-wave and long-wave infrared imaging applications [14].

Reconstruction of images from multiplexed FPA measurements involves solution of an ill-posed inverse problem. To account for system nonidealities, previous techniques employ offline calibration to measure the system matrix, and expressing the inverse problem based on this matrix [5]. However, dimensions of the system matrix are often large enough to render the inverse problem computationally intractable [4], [14]. Instead, CalibFPA uses online calibration to correct compressive FPA measurements on the fly, and these corrected measurements are reconstructed without the need to process a system matrix. This enables expression of the forward model in block-diagonal form, and an ADMM-based algorithm then efficiently reconstructs HR images.

We comprehensively evaluated the performance of CalibFPA against offline and online calibration methods. CalibFPA achieved substantially higher performance than online calibration based on traditional or deep methods, regardless of reconstruction algorithm. Meanwhile, CalibFPA demonstrated a reasonable degree of reliability against mismatches between assumed and actual Airy disk radii and SLM shift values. Finally, CalibFPA performed comparably with computationally-burdening offline calibration, especially when realistic inaccuracies are considered in the system matrix. We also evaluated CalibFPA against a deep-learning-based single image super-resolution (SISR) method. Compared to SISR, performance benefits of CalibFPA are especially evident at lower Airy disk radius and higher input SNR values. Towards larger radius values where the lens PSF introduces excessive blurring, the influence of the sensor resolution on spatial acuity diminishes, so similar performance can be expected from SLM-based imaging methods such as CalibFPA versus SISR. Meanwhile, towards lower input SNRs, multiplexed encoding where light from a subset of pixels is blocked via SLM patterns can yield lower SNR efficiency than SISR that receives light from all pixels.

Here we reported experimental results on a visual scene comprising black-colored digits against a uniform background. Compared to competing calibration methods, we find elevated visual acuity and lower artifacts in images based on CalibFPA, for both least-squares and plug-and-play reconstructions. These results suggest that CalibFPA can help boost performance in compressive FPA imaging. Yet, our optical setup was based on a sphere source that projected light through the scene for illumination, so the scene was created by printing a binary patterns on a transparent medium, restricting the complexity of the visual scene. It remains important future work to demonstrate CalibFPA on a broader set of experimental data containing more complex visual scenes.

Several technical limitations can be addressed to further enhance the utility of CalibFPA. Here, a plug-and-play reconstruction was used based on a pretrained convolutional denoiser. While plug-and-play methods facilitate implementation by decoupling the regularization prior from the forward model, performance improvements can be viable when sizable training sets are available. In particular, transformer-based denoisers can be used to elevate sensitivity to long-range context in HR images [36], [45]. Here, an unrolled reconstruction was used based on a fixed number of iterations in the network architecture that might limit performance. Adopting deep equilibrium models [31], [32] can help increase the depth of the unrolled network without compromising computational efficiency during reconstruction. Here, the calibration network in CalibFPA was trained to optimize calibration performance independently from the reconstruction stage, even when learning-based reconstruction methods were used. In principle, joint training of networks for the calibration and reconstruction stages can help improve overall performance and efficiency in compressive FPA imaging. SLM patterns might also be optimized in conjunction with the networks for further improvements [46], [47]. Here, deterministic networks were used for calibration and reconstruction stages. The representational diversity in reconstructed images might be improved by adopting generative models [48]–[50]. Here, CalibFPA was used to reconstruct single static frames of incident scenes. For dynamic scenes, the calibration and image reconstruction stages of CalibFPA can be modified to jointly process multiple frames via recurrent architectures [51], [52]. Lastly, here we assumed a linear-shift-invariant forward model, yet shift-variant models might allow for potentially improved correction fidelity.

VII. Conclusion

In this study, we introduced a novel compressive FPA system for online calibration of measurements against optical aberrations from the relay lens and subsequent reconstruction of HR images from corrected measurements. CalibFPA employs a physics-driven deep network to correct measurements on the fly, so that image reconstruction can be performed on block-diagonal matrices for the corrected measurements and the system matrix capturing the effects of downsampling and SLM patterns. As such, CalibFPA improves computational efficiency of image reconstruction over offline calibration methods that involve dense matrices. Comprehensive demonstrations on simulated and experimental data clearly indicate that CalibFPA holds great promise for HR imaging with low-cost compressive FPA sensors.

References

[1] G. M. G. M. P. Edgar and M. J. Padgett, "Principles and prospects for single-pixel imaging," *Nat. News*, no. 13, pp. 1749–4893, Dec 2018.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2024.3477312

13

[2] G. M. Gibson, S. D. Johnson, and M. J. Padgett, "Single-pixel imaging 12 years on: a review," *Opt. Express*, vol. 28, no. 19, pp. 28 190–28 208, 2020.

[3] M. F. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

[4] H. Chen *et al.*, "FPA-CS: Focal plane array-based compressive imaging in short-wave infrared," in *IEEE CVPR*, 2015, pp. 2358–2366.

[5] A. Mahalanobis *et al.*, "Recent results of medium wave infrared compressive sensing," *Appl. Opt.*, vol. 53, no. 34, pp. 8060–8070, 2014.

[6] R. I. Stantchev *et al.*, "Real-time terahertz imaging with a single-pixel detector," *Nat. Comm.*, vol. 11, no. 1, p. 2535, 2020.

[7] W. Li *et al.*, "Dual-color terahertz spatial light modulator for single-pixel imaging," *Light: Sci. App.*, vol. 11, no. 1, p. 191, 2022.

[8] G. Paunescu, P. Lutzmann, and D. Wegner, "Compressive sensing for active imaging in swir spectral range," in *Elec.-Opt. Remote Sens. XII*, vol. 10796, 2018, pp. 65–71.

[9] J. Wang, M. Gupta, and A. C. Sankaranarayanan, "Lisens- a scalable architecture for video compressive sensing," in *IEEE ICCP*, 2015, pp. 1–9.

[10] G. R. Arce *et al.*, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 105–115, 2013.

[11] L. Song *et al.*, "High-accuracy image formation model for coded aperture snapshot spectral imaging," *IEEE Trans. Comput. Imaging*, vol. 8, pp. 188–200, 2022.

[12] A. A. Wagadarikar *et al.*, "Spectral image estimation for coded aperture snapshot spectral imagers," in *Image Recon. Incomplete Data V*, vol. 7076, 2008, pp. 9–23.

[13] H. Arguello, H. F. Rueda, and G. R. Arce, "Spatial super-resolution in code aperture spectral imaging," in *Compress. Sens.*, vol. 8365, 2012, pp. 43–48.

[14] X.-P. Jin *et al.*, "Long-distance mid-wave infrared super-resolution compressive imaging," *Opt. Laser Tech.*, vol. 157, p. 108740, 2023.

[15] Z. Wu and X. Wang, "Focal plane array-based compressive imaging in medium wave infrared: modeling, implementation, and challenges," *App. Opt.*, vol. 58, no. 31, pp. 8433–8441, 2019.

[16] A. Sözak *et al.*, "A thermal and electro-magnetic hybrid moems microscanner with integrated spatial light modulation and metalens for resolution enhancement of infrared imaging systems," *IEEE Sensors J.*, vol. 22, no. 21, pp. 20 538–20 545, 2022.

[17] R. Baraniuk *et al.*, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, pp. 253–263, 2008.

[18] O. F. Kar, A. Güngör, and H. E. Güven, "Real-time compressive focal plane array imaging," *Imaging App. Opt.*, p. CTu2A.3, 2019.

[19] S. Yang *et al.*, "Mid-wave infrared snapshot compressive spectral imager with deep infrared denoising prior," *Remote Sens.*, vol. 15, no. 1, p. 280, 2023.

[20] H. Arguello and G. R. Arce, "Rank minimization code aperture design for spectrally selective compressive imaging," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 941–954, 2013.

[21] C. Li, W. Yin, and Y. Zhang, "User's guide for tval3: Tv minimization by augmented lagrangian and alternating direction algorithms," *CAAM report*, vol. 20, no. 46-47, p. 4, 2009.

[22] B. Shi, Y. Wang, and D. Li, "Provable general bounded denoisers for snapshot compressive imaging with convergence guarantee," *IEEE Trans. Comput. Imaging*, vol. 9, pp. 55–69, 2023.

[23] Y. Chen, Y. Wang, and H. Zhang, "Prior image guided snapshot compressive spectral imaging," *IEEE Trans. Pattern Anal. Mach. Intelli.*, 2023.

[24] L. Zhang *et al.*, "High-resolution fast mid-wave infrared compressive imaging," *Opt. Letters*, vol. 46, no. 10, pp. 2469–2472, 2021.

[25] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-Based Deep Learning Architecture for Inverse Problems," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 394–405, 2019.

[26] U. S. Kamilov, H. Mansour, and B. Wohlberg, "A plug-and-play priors approach for solving nonlinear imaging inverse problems," *IEEE Sig. Process. Let.*, vol. 24, no. 12, pp. 1872–1876, 2017.

[27] O. F. Kar, A. Güngör, and H. E. Güven, "Learning based regularization for spatial multiplexing cameras," in *IEEE GlobalSIP*, 2019, pp. 1–5.

[28] U. S. Kamilov *et al.*, "Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 85–97, 2023.

[29] B. Askin *et al.*, "PP-MPI: A deep plug-and-play prior for magnetic particle imaging reconstruction," in *MLMIR*, 2022, pp. 105–114.

[30] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Sig. Process. Mag.*, vol. 38, no. 2, pp. 18–44, 2021.

[31] A. Güngör *et al.*, "Deq-mpi: A deep equilibrium reconstruction with learned consistency for magnetic particle imaging," *IEEE Trans. Med. Imaging*, pp. 1–1, 2023.

[32] D. Gilton, G. Ongie, and R. Willett, "Deep equilibrium architectures for inverse problems in imaging," *IEEE Trans. Comput. Imaging*, vol. 7, pp. 1123–1133, 2021.

[33] S. Nah, T. Hyun Kim, and K. Mu Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *IEEE CVPR*, 2017, pp. 3883–3891.

[34] J. Liang *et al.*, "SwinIR: Image restoration using swin transformer," *arXiv preprint arXiv:2108.10257*, 2021.

[35] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE CVPR*, 2016, pp. 1874–1883.

[36] A. Güngör *et al.*, "TranSMS: Transformers for super-resolution calibration in magnetic particle imaging," *IEEE Trans. Med. Imaging*, vol. 41, no. 12, pp. 3562–3574, 2022.

[37] J. P. Dumas *et al.*, "Computational imaging with a highly parallel image-plane-coded architecture: challenges and solutions," *Opt. Express*, vol. 24, no. 6, pp. 6145–6155, 2016.

[38] O. F. Kar *et al.*, "A performance analysis on the optimal number of measurements for coded compressive imaging," in *IEEE GlobalSIP*, 2018, pp. 101–105.

[39] A. Güngör, O. F. Kar, and H. E. Güven, "A matrix-free reconstruction method for compressive focal plane array imaging," in *IEEE ICIP*, 2018, pp. 1827–1831.

[40] O. F. Kar *et al.*, "An efficient parallel algorithm for single-pixel and FPA imaging," in *SPIE Comput. Imaging III*, 2018, p. 106690J.

[41] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *IEEE CVPR Workshops*, July 2017.

[42] A. D. Miron, "Multi-modal, multi-domain pedestrian detection and classification: Proposals and explorations in visible over stereovision, fir and swir," Ph.D. dissertation, INSA de Rouen; Universitatea Babeș-Bolyai (Cluj-Napoca, Roumanie), 2014.

[43] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, p. 745, 1974.

[44] *DnCNN-Pytorch Github Repository, Web Page: https://github.com/SaoYan/DnCNN-PyTorch.*

[45] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv:2010.11929*, 2021.

[46] M. R. Kellman *et al.*, "Physics-based learned design: Optimized coded-illumination for quantitative phase imaging," *IEEE Trans. Comput. Imaging*, vol. 5, no. 3, pp. 344–353, 2019.

[47] P. van der Meulen *et al.*, "Coding mask design for single sensor ultrasound imaging," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 358–373, 2019.

[48] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *NeurIPS*, vol. 33, pp. 6840–6851, 2020.

[49] A. Güngör *et al.*, "Adaptive diffusion priors for accelerated MRI reconstruction," *Med. Image Anal.*, vol. 88, p. 102872, 2023.

[50] M. Özbey *et al.*, "Unsupervised medical image translation with adversarial diffusion models," *IEEE Trans Med Imaging*, 2023.

[51] O. F. Kar, A. Güngör, and H. E. Güven, "Real-time compressive video reconstruction for spatial multiplexing cameras," in *IEEE GlobalSIP*, 2019, pp. 1–5.

[52] R. F. Marcia and R. M. Willett, "Compressive coded aperture video reconstruction," in *Eur. Signal Process. Conf.* IEEE, 2008, pp. 1–5.