

MORE: Middle Output Regularized End-to-End Optimization for Computational Imaging

ROMAN JACOME¹, PABLO GOMEZ², AND HENRY ARGUELLO³

¹Department of Physics, Universidad Industrial de Santander, Bucaramanga, 680002, Colombia

²Department of Electrical Engineering, Universidad Industrial de Santander, Bucaramanga, 680002, Colombia

^{3,*}Department of Computer Science, Universidad Industrial de Santander, Bucaramanga, 680002, Colombia

*Corresponding author: henarfu@uis.edu.co

Compiled May 17, 2024

Optical coding is an essential technique in computational imaging (CI) that allows high-dimensional signal sensing through post-processed coded projections to decode the underlying signal. Currently, the optical coding elements (OCE) are optimized in an end-to-end (E2E) manner where a set of layers (encoder) of a deep neural network model the OCE while the rest of the network (decoder) performs a given computational task. However, while the training performance of the whole network can be acceptable, the encoder layers can be sub-optimal, leading to deficient OCE designs. This sub-optimal performance of the encoder originated from factors such as the loss function of the network does not consider the intermedium layers separately as the output at those layers is unknown. Second, the encoder suffers from the vanishing of the gradient since the encoder takes place in the first layers. Third, the proper estimation of the gradient in these layers is constrained to satisfy physical limitations. In this work, we propose a Middle Output Regularized End-to-end (MORE) optimization, where a set of regularization functions are used to overcome the sub-optimal performance of the encoder. The significant advantage of our regularization is that it does not require additional knowledge of the encoder and can be applied to most optical sensing instruments in computational imaging. Instead, the regularization exploits some prior knowledge about the computational task, the statistical properties of the output of the encoder (measurements), and the sensing model. Specifically, we proposed three types of regularizers: The first one is based on statistical divergences of the measurements, the second depends only on the variance of the measurements, and the last one is a structural regularizer promoting low rankness and sparsity of the set of measurements. We validated the proposed training procedure in two representative computational imaging systems, the single-pixel camera (SPC), and the coded aperture snapshot spectral imager (CASSI), showing significant improvement with respect to non-regularized designs. © 2024 Optical Society of America

<http://dx.doi.org/10.1364/ao.XX.XXXXXX>

1. INTRODUCTION

The joint operation of optical systems and computational algorithms in computational imaging (CI) has allowed the acquisition of high-dimensional signals, $D > 2$ where D is the signal dimension, such as spectral imaging [1], polarization state [2], depth imaging [3], temporal imaging [4], and angular views in light fields [5]. A key in these systems is the optical coding elements (OCE), which allow modulating variables of the incident light wave, such as its amplitude, using coded aperture (CA) [6], phase using diffractive lenses [7], polarization using micro-polarizers [2] or spectral information employing dispersive elements [8]. Consequently, the design of these elements for optimal CI performance has received great attention. Particularly, the design of CA has been extensively studied based on

analytical criteria such as the Hadamard invertibility [6, 9] or compressive sensing theory [10], such as the restricted isometry property [11, 12]. Additionally, in the design of diffractive lenses, methods have been proposed to reduce chromatic aberrations and geometries [13] to improve CI systems. Moreover, these elements have been designed for the encoding of spectral information [14, 15]. Although an increase in the performance of the aforementioned design methods is presented with respect to standard configurations (Bernoulli CA or Fresnel lenses), these are based on structural assumptions of the signal or system, which in some cases are not achieved and do not work well in several scenarios.

With new advances in machine learning algorithms, particularly those of deep learning [16], and a large number of databases

available, the end-to-end optimization method [17] has been proposed where the OCE is optimized taking into account properties of the training dataset. Here, the optical system is modeled as a layer of a neural network whose trainable parameters are the OCE, and this layer is called the optical encoder (OE). The OCE is coupled with a network that performs the decoding task, i.e., reconstruction, classification, segmentation, etc., and is called the computational decoder (CD). This way, the OCE is jointly trained with the inference task, allowing the OCE to adapt according to the training database and the CD. While the whole E2E network has shown an overall good performance in several tasks such as spectral imaging [18], classification, and depth estimation [19], compressive spectral image fusion [20, 21], depth estimation with compressive light field [22], extended chromatic field of view and super-resolution [23], or monocular depth estimation [24] among others, optimization of the OE can be subpar due to several reasons. For instance, the OE parameters can be only optimized with respect to the loss function computed with the output CD network yielding, first in the gradient vanishing on the OCE. Thus, the performance of the E2E network relies more on optimal CD training than on optimal optical codification design. Moreover, the output of the intermediate layer is not considered a variable that needs to be carefully optimized to increase the entire performance of the network. Additionally, the OE is highly constrained to a feasible set of values due to the physical meaning of the OCE, which reduces the degrees of freedom in training.

To overcome the OCE training issues, we propose Middle Output Regularized end-to-end (MORE), where a set of regularization functions performed in the output of the OE are devised. First, the proposed regularization functions can exploit prior knowledge about the task, the dataset, and the OE to optimize the OCE. Also, we give insights into the optimality criteria at the intermediate layers' output based on these outputs' statistical properties such as the mean and variance of the measurements set. We show how the measurement distribution affects the CD performance according to the tasks. Empirically, we demonstrate that if we concentrate on the distribution of the measurements (reducing the data variance), it allows a more compact representation of the data, thus allowing better reconstruction performance. While for the classification task, increasing the variance improves accuracy since the classes are better identified by the CD. Based on these criteria, three types of regularization functions are proposed to promote these properties on the OE. i) Kullback-Leiber divergence regularization, where these functions aim to approximate the distribution of the intermediate output (the OE output) to a prior distribution. Particularly, Gaussian distribution (widely used in variational autoencoders [25]) and Laplacian distribution (employed in regression tasks [26]) priors are employed since the KL-D has closed form solution and can be efficiently implemented. This regularization promotes a given mean and variance value on the measurement distribution by the prior distribution. We study the effect of this prior distribution to find the optimal configuration. Preliminary results on the KL-D regularizer have shown promising results in [27] for recovery tasks and also beyond computational imaging in [28] where we employed this regularization to improve the design of the geometry of acquisition in compressive seismic applications. ii) Variance-based regularization in which the variance of the coded observations is minimized or maximized. This criterion has been studied in self-supervised representation learning, where controlling the variance allows a more compact representation of the data. We minimize the variance for the

reconstruction task and maximize it for classification. iii) Structural regularization, where we exploit low rank in the measurement set by sparsifying the singular values of the measurements, thus concentrating the dataset information in a few linear independent coded measurements. And sparsity in a given basis, e.g., wavelet along the measurement set to promote smoothness, i.e., reduce the data variability. These regularization functions indirectly concentrate on the distribution of the measurements. From a learning representation point-of-view, these regularization functions encourage invariant OE and allow contractive representation in the data manifold, while the recovery loss function enforces accurate image estimation [29]. Contractive representations have been used in traditional autoencoders [30]. However, this criteria has not been proposed for sensing matrix optimization. One of the main advantages of the proposed training methodology is that it can be applied in any optical architecture and can be adapted for any computational task. An overview of the proposed approach is shown in Fig. 1.

Several systems were employed to validate the proposed design criteria's effectiveness. First, the regularization functions were evaluated using a compressive sensing scenario; further real imaging systems were employed, such as the single-pixel camera (SPC) [31] for imaging, and the coded aperture snapshot spectral imager (CASSI) [8], for spectral imaging. We showed that decreasing the variance of the set of measurements for all these systems allows for better reconstruction quality.

The rest of the paper is organized as follows. In section 2, the E2E formulation is presented, section 3 presents the proposed regularization functions to improve the E2E performance. Later in section 4 is shown the mathematical modeling of the compressive imaging systems employed to validate the proposed design. Section 5 presents the numerical experiments. Furthermore, section 6 reports the experimental validation of the proposed design, and finally section 7 presents the conclusion of this work.

2. END-TO-END OPTIMIZATION

In computational imaging, a high-dimensional signal $\mathbf{f} \in \mathbb{R}^n$ is acquired via a low-dimensional coded projection $\mathbf{y} \in \mathbb{R}^m$, with $m \ll n$. In the E2E optimization framework, the sensing procedure is modeled as a differentiable linear operator, i.e.,

$$\mathbf{y} = \mathbf{H}_{\Phi} \mathbf{f} + \boldsymbol{\omega} \quad (1)$$

where $\mathbf{H}_{\Phi} \in \mathbb{R}^{n \times m}$ is the sensing matrix of the system, namely, the OE, Φ is the OCE of the sensing system, e.g., CA or DOE, and $\boldsymbol{\omega}$ is additive noise. The OCE is then optimized jointly with a CD network \mathcal{M}_{θ} with trainable parameters θ as

$$\begin{aligned} \{\theta^*, \Phi^*\} &= \arg \min_{\theta, \Phi} \mathcal{L}(\theta, \Phi) \\ &= \arg \min_{\theta, \Phi} \frac{1}{K} \sum_{k=1}^K \mathcal{L}_{task}(\mathcal{M}_{\theta}(\mathbf{H}_{\Phi} \mathbf{f}_k), \mathbf{d}_k) + \rho R_i(\Phi), \end{aligned} \quad (2)$$

where $\{\mathbf{f}_k\}_{k=1}^K$ is the training dataset, \mathcal{L}_{task} is the loss function of desired tasks, \mathbf{d}_k corresponds to the expected output, e.g., classification labels [32], ground truth image [20], depth maps [3] etc. Usually, the OCE is constrained to a set of feasible values due to the physical limitations of the elements. To impose this constraint, a regularization function $R_i(\Phi)$ is added to the loss function, where ρ is the regularization parameter. This regularization can also induce the desired properties on the OCE, such as transmittance in CA, number of shots, etc., [Table II 17]. The

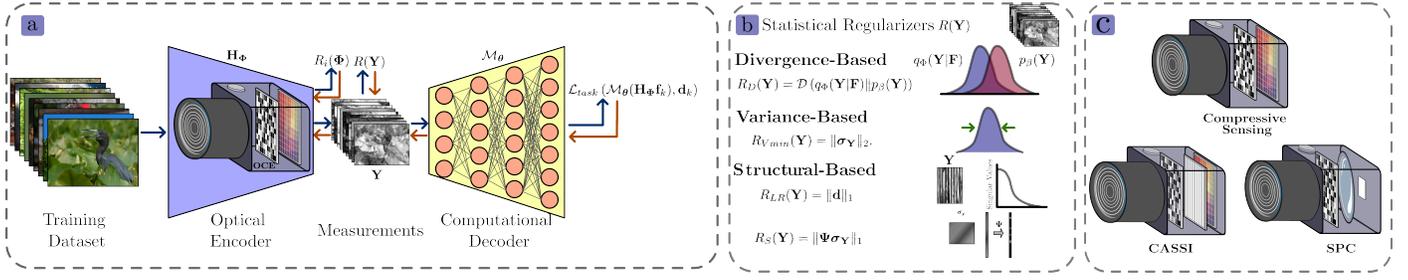


Fig. 1. a) E2E scheme where the OE is optimized jointly with the CD network. b) Proposed regularization functions to improve the design of the OE by inducing statistical priors during the training of the E2E network.

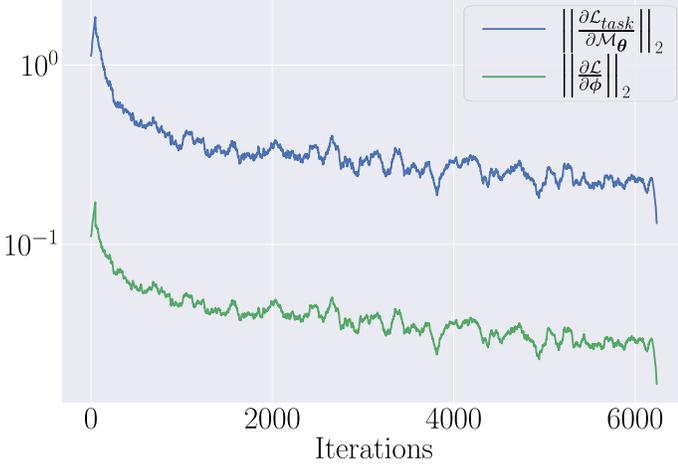


Fig. 2. Norm of the gradient of the CD parameters and the OCE of the OE.

main goal here is that the OCE is updated according to the task loss function and the physical constraint given by the regularization. Particularly, following the chain rule, the gradient of the loss function with respect to the OCE is

$$\frac{\partial \mathcal{L}}{\partial \Phi} = \frac{\partial \mathcal{L}_{task}}{\partial \mathcal{M}_\theta} \frac{\partial \mathcal{M}_\theta}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \Phi} + \rho \frac{\partial R_i(\Phi)}{\partial \Phi} \quad (3)$$

The training of the OCE has two main issues. i) the training is highly conditioned to the physical-limitation regularization function, which decreases the degrees of freedom of the OCE. ii) Gradient vanishing due to the OE being the first layer of the E2E network, most of the optimization is performed over the CD parameters rather than optimizing the optical coding properly. As an illustration of this phenomenon in Fig. 2 is plotted norm of the loss function gradient with respect to the θ and ϕ in a logarithmic scale. This experiment is performed with an SPC as OE; its corresponding OCE is the CA, and the computational task is recovery via a UNET network. Here, a significant difference (almost one order of magnitude) between the OE gradient and the CD parameters gradient. Mainly, this issue is because the intermediate output of the E2E (the coded measurements) is not taken into account independently on training, and the optimization is only performed with respect to the CD output. Thus, we provide new insights into what should be this intermediate output based on the statistical properties of this output. Then, based on this criterion, we propose a set of regularization functions that control the statistical properties of the coded measurements. Other regularization functions have

been proposed to increase the performance of the E2E network. For instance, [19] proposes to minimize a regularization based on concentrating the eigenvalues of the sensing matrix \mathbf{H}_Φ following the function $\|\mathbf{H}_\Phi^T \mathbf{H}_\Phi \mathbf{f} - \mathbf{f}\|_2$. Similarly [33] proposes to minimize the closed-form solution of a regularized $\ell - 2$ optimization problem, i.e. $\arg \min_{\mathbf{f}} \|\mathbf{H}_\Phi \mathbf{f} - \mathbf{y}\|_2 + \gamma \|\mathbf{f}\|_2$, yielding the regularization function $\|\mathbf{f}_k - (\mathbf{H}_\Phi^T \mathbf{H}_\Phi + \gamma \mathbf{I})^{-1} \mathbf{H}_\Phi^T \mathbf{H}_\Phi \mathbf{f}_k\|_2$, thus promoting good invertibility properties on \mathbf{H}_Φ . These functions aim to obtain an approximation of the desired image only with the invertibility properties of the sensing matrix. However, such invertibility is usually not met due to a highly structured matrix and mostly due to the ill-posed nature of the problem. Thus this regularization does not provide better optimization of the OE. Additionally, these regularization functions only apply to recovery and cannot be adapted to other computational tasks. The proposed regularization functions promote a contractive OE, which reduces the variance between training samples' compressed projections. Then, by reducing the variability on the compressed domain, the decoder performs better in the reconstruction. Also, for the classification task, the opposite effect is desired, expanding the distribution of the measurements. In the following section will be detailed the proposed regularization.

3. PROPOSED REGULARIZATION FUNCTIONS

In this paper, we propose a new type of regularization function for E2E optimization, promoting some properties on the distribution of the measurements. The optimization problem Eq. (2) becomes

$$\begin{aligned} \{\theta^*, \Phi^*\} &= \arg \min_{\theta, \Phi} \mathcal{L}(\theta, \Phi) \\ &= \arg \min_{\theta, \Phi} \frac{1}{K} \sum_{k=1}^K \mathcal{L}_{task}(\mathcal{M}_\theta(\mathbf{H}_\Phi \mathbf{f}_k), \mathbf{d}_k) + \rho R_i(\Phi) \\ &\quad + \mu R(\mathbf{Y}) \end{aligned} \quad (4)$$

where μ is the regularization parameter and $\mathbf{Y} \in \mathbb{R}^{K \times m}$ is the matrix containing all the training batch of compressed measurements, i.e., $\mathbf{Y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_K^T]^T$.

A. Divergence-based regularization

This type of regularization function is based on the idea behind variational auto-encoders [25]. Particularly, this regularization aims to approximate the probability distribution of the measurements set denoted by the posterior distribution $q_\Phi(\mathbf{Y}|\mathbf{F})$, where $\mathbf{F} \in \mathbb{R}^{K \times n}$ is a matrix with all the input training images, to a prior distribution $p_\beta(\mathbf{Y})$ where β is the set of parameters that defines the prior distribution. This regularizer is defined as

$$R_D(\mathbf{Y}) = \mathcal{D}(q_\Phi(\mathbf{Y}|\mathbf{F}) \| p_\beta(\mathbf{Y})) \quad (5)$$

where \mathcal{D} denotes the divergence function. Several divergences have been used as loss functions in neural network training. The most common is the Kullback-Leiber divergence, employed in variational-autoencoders [25], generative adversarial networks [34], self-supervised learning [35] among others. Particularly, the KL divergence is defined as follows, given two probability distributions $P(x)$ and $Q(x)$, we have $\mathcal{D}_{KL}(P\|Q) = \int P(x) \log\left(\frac{P(x)}{Q(x)}\right) dx$. One of the main reasons the KL divergence is widely used is that it has a closed-form solution when P and Q are Gaussian or Laplacian distributions (see [25, 36]). In these cases, the parameters for the prior distribution $p_\beta(\mathbf{Y})$ are $\beta = \mu_p, \sigma_p$, where μ_p is the mean value and σ_p is the variance of the distribution. For the distribution of the measurements $q_\Phi(\mathbf{Y}\|\mathbf{F})$. The mean $\mu_{\mathbf{Y}} \in \mathbb{R}^m$ and variance $\sigma_{\mathbf{Y}} \in \mathbb{R}_+^m$ are computed pixel-wise across the training batch. For the Gaussian case, the KL divergence-based regularizer is defined as:

$$R_{KL-G}(\mathbf{Y}) = \log\left(\frac{\sigma_{\mathbf{Y}}}{\sigma_p}\right) - \frac{\sigma_{\mathbf{Y}}^2 + (\mu_{\mathbf{Y}} - \mu_p)^2}{2\sigma_p^2} + \frac{1}{2} \quad (6)$$

and for the Laplacian assumption, the KL divergence-based regularizer is given by

$$R_{KL-L}(\mathbf{Y}) = \log\left(\frac{\sigma_{\mathbf{Y}}}{\sigma_p}\right) - \frac{\sigma_p + e^{\left(\frac{-|\mu_p - \mu_{\mathbf{Y}}|}{\sigma_p}\right)} + |\mu_p - \mu_{\mathbf{Y}}|}{\sigma_p} - 1 \quad (7)$$

The effect of these regularizers depends directly on the values of the mean and variance of the prior distribution. Thus, these are hyperparameters of the regularizers needed to be chosen to obtain the desired behavior.

B. Variance-Based regularization

Another way to control the measurement set distribution is to regularize the variance directly. Here we proposed a variance minimization regularizer. This variance-based regularization criterion has also been used in representation learning for self-supervised task [37], sparse-coding [38]. Here, we extrapolate these criteria of optimal low-dimensional representation basis optimal compressive sensing system, thus giving more interpretability of the designed OCE by the E2E optimization. The proposed regularization function is the following

$$R_{Vmin}(\mathbf{Y}) = \|\sigma_{\mathbf{Y}}\|_2. \quad (8)$$

For this regularization, to control how concentrated we want the distribution of the measurements to be, we tune the hyperparameter μ on Eq. (4). In some downstream tasks, such as classification, where we want to identify the difference from the image of different classes, therefore, if the distribution of the measurements is wider, i.e., greater variance, the CD could better identify the classes. Thus, the variance maximization can be promoted by the following regularization function

$$R_{Vmax}(\mathbf{Y}) = \|\sigma_{max} - \sigma_{\mathbf{Y}}\|_2. \quad (9)$$

where σ_{max} is a maximum variance reference, a hyperparameter that can be tuned.

C. Structural regularization

This type of regularization is based on the common priors of compressed sensing recovery: low-rank and sparsity [10, 39]. Although these priors are employed over the underlining signal \mathbf{f} , here we employ these criteria to achieve the following effects

in the measurement space. The low-rank prior is employed to concentrate the information of the dataset in a few representative measurements, thus reducing the projection manifold and allowing better reconstruction by the CD. To promote the low rankness on the measurement space, we minimize the ℓ_1 norm of the singular values of \mathbf{Y} . Particularly, employing the singular value decomposition (SVD) of the measurement matrix, we obtain $\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ where the matrices $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{K \times K}$ are the left and right singular vector respectively and $\mathbf{D} \in \mathbb{R}^{m \times K}$ is a rectangular diagonal matrix with the singular values in its diagonal. The singular values are denoted by $\mathbf{d} = [d_1, \dots, d_K]$ where $d_i = \mathbf{D}_{(i,i)}$ for $i = 1, \dots, K$. Thus, our low-rank regularization is the following

$$R_{LR}(\mathbf{Y}) = \|\mathbf{d}\|_1 \quad (10)$$

By applying the ℓ_1 norm on the singular values, we promote having few non-zero values on \mathbf{d} and thus reducing the rank.

The second criterion, the sparsity-based regularization, follows the same intuition of its application in imaging inverse problems, where sparsity over a given representation basis (wavelet, DCT, or Fourier) is employed to promote the smoothness of the images. Here we aim to promote smoothness along the coded measurements, thus reducing the variance. Mathematically, the regularizer is

$$R_S(\mathbf{Y}) = \|\Psi\sigma_{\mathbf{Y}}\|_1 \quad (11)$$

where Ψ is the representation basis. In this work, we consider the Haar wavelet, which has shown good results in promoting smoothness on signals [40].

4. COMPRESSIVE IMAGING SENSING MODELS

To validate the proposed deep optical design, we employed two flagship CI optical architectures the CASSI and SPC.

A. Single Pixel Camera

The first optical architecture is the Single Pixel Camera (SPC) [31], this architecture is widely used in compressive imaging systems. This system employs an imaging lens that spatially introduces light, which is previously modulated by an CA, and then integrates the encoded image into a single pixel detector. The CA can be implemented with spatial light modulators (SLM) [41], such as a digital micro-mirror device (DMD)[42], that selectively redirects parts of the light beam [43][44]. The SPC uses a CA $\Phi_{(i,j)}^k$ that spatially modulates all the information from the scene $\mathbf{F}_{(i,j)}$ with the same pattern, where (i, j) index the spatial coordinates, k indexes each captured snapshot. In particular, the CA $\Phi_{(i,j)}^k$ is a binary pattern whose spatial distribution determines the performance of the reconstruction. Mathematically, the CA effect over the scene can be represented as:

$$\hat{\mathbf{F}}_{(i,j)}^k = \mathbf{F}_{(i,j)} \Phi_{(i,j)}^k \quad (12)$$

After that, the modulated scene $\hat{\mathbf{F}}$ is focused in a single spatial point by the condenser lens, and captured by a single-pixel detector. The resulting sensing matrix $\hat{\mathbf{H}}_{\phi_s} \in \mathbb{R}^{K \times MN}$ contains the vectorization of the CA of each snapshot k in his rows. The aperture codes implemented for the sensing matrix, are the design parameter from the proposed regularizers. The acquisition system can be modeled as:

$$\mathbf{y} = \hat{\mathbf{H}}_{\phi_s} \mathbf{f} + \mathbf{n}_s \quad (13)$$

System	Dataset	Task	Computational Decoder
Compressive Sensing	MNIST	Recovery	U-NET
	Fashion MNIST	Recovery	U-NET
Single-Pixel	Fashion MNIST	Classification	MobilNet-V2
	CIFAR-10	Recovery	U-NET
CASSI	ARAD	Recovery	U-NET

Table 1. Experiments to validate the proposed OCE design

Where, $\mathbf{y} = [y_0, \dots, (y_{K-1})]^T$ is the compressed measurements, $\mathbf{f} \in \mathbb{R}^{M \times N}$ is the vectorized image and \mathbf{n}_c is a additive Gaussian noise.

B. CASSI

In the CASSI architecture, the input light source is first focused by an imaging lens to a CA, which codifies the spatial information of the image. Then, the spectral information of the coded field is dispersed through a prism. Finally, the coded and dispersed information impinges on a focal plane array. Therefore, the discrete model of the CASSI measurements \mathbf{y}_c can be formulated as:

$$\mathbf{y}_{c(i,j)} = \sum_{\ell=1}^L \Phi_{c(i,j)} \mathbf{F}(i,j-\ell,\ell), \quad (14)$$

where $\mathbf{F} \in \mathbb{R}^{M \times N \times L}$ and the CASSI measurements, Φ_c represents the CA. The discrete model in Eq. (14) can be expressed in a matrix-vector product in the following expression

$$\mathbf{y}_c = \mathbf{H}_{\Phi_c} \mathbf{f} + \mathbf{n}_c, \quad (15)$$

where $\mathbf{y}_c \in \mathbb{R}^{M(N+L-1)}$ are the compressed measurements, $\mathbf{H}_{\Phi_c} \in \mathbb{R}^{M(N+L-1) \times M \times N \times L}$ the CASSI sensing matrix, $\mathbf{f} \in \mathbb{R}^{M \times N \times L}$ is the vectorization of the high spatial-spectral resolution image, and $\mathbf{n}_c \in \mathbb{R}^{M(N+L-1)}$ is additive noise. Here the design parameter are the CA Φ_c

5. SIMULATION RESULTS

To evaluate the performance of the proposed design methodology, we perform the experiments shown in Table 1. Particularly, we perform classification and recovery tasks, where for the first we employ a MobilNet-V2 network [45] which is a lightweight model widely employed for classification. For the recovery task, it was used a U-Net model with five convolution blocks for each downsampling and upsampling process. For all the experiments, we trained the E2E network for 100 epochs, halving the learning rate every 40 epochs. For the CASSI CA binary constraint, the polynomial regularization in [19] was employed i.e., $R(\Phi) = \sum_{ij} (1 - \Phi_{ij})^2 (\Phi_{ij})^2$. For the SPC CA constraint, we consider values $\{-1, 1\}$ which in practice can be achieved by following the procedure in detail in [Appendix 32] which allows better signal-to-noise-ratio (SNR). Then, the physical constraint regularizer is $R(\Phi) = \sum_{ij} (1 - \Phi_{ij})^2 (1 + \Phi_{ij})^2$. The parameter of the physical constraint regularizer ρ was dynamically updated during training as suggested in [19].

A. Compressed Sensing Experiments

A first experiment to validate the performance of the proposed regularized E2E network, we study a compressive imaging scenario, not imposing a physical and structural meaning on the sensing matrix \mathbf{H}_{Φ} . Here we use a compression ratio of 10%

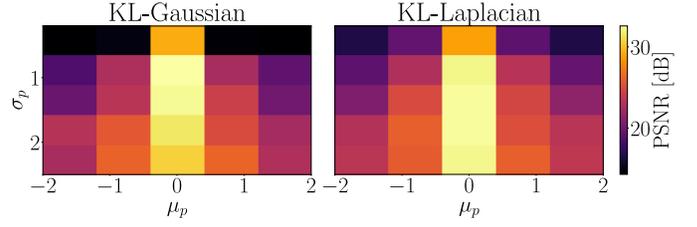


Fig. 3. Recovery performance for the CS scenario employing the KL-D regularizers with the Gaussian (left) and Laplacian (right) cases

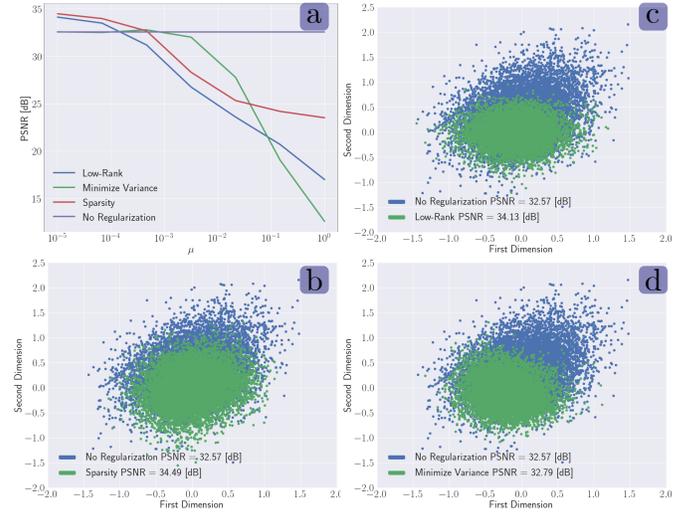


Fig. 4. Recovery performance for the CS scenario employing the variance and structural regularizers compared with the non-regularized E2E network. a) Performance depends on the regularization parameter μ . First and second-pixel distribution of the test dataset for the b) low-rank, c) minimize variance, d) sparsity.

KL-Divergence: First, we analyze the effect of the mean and the variance of the prior distribution (μ_p, σ_p) on the network performance. Here, we vary the μ_p from -2 to 2, and σ_p was changed from 0.1 to 2.0, taking five equispaced values. The results of this experiment are shown in Fig 3, where optimal reconstruction PSNR values are obtained at variances close to 1.0 and for means close to 0. These results suggest better reconstruction performance is obtained by concentrating on the measurement distribution. The main interpretation is that reducing the representation space can improve the CD performance since the variability of the data is reduced.

Variance and structural regularizers Then, we analyze the performance of the E2E network for the variance minimization and structural regularization. In Fig 4(a) is shown the recovery performance depending on the regularization parameter in Eq. (4) where optimal values for the regularization suggest a trade-off between how much concentrate the distribution and the recovery performance. Particularly, significant recovery improvements are shown with the low-rank and sparsity experiments with respect to the baseline (no regularization E2E). Fig 4(b-d) presents the distribution of two pixels of the test set measurement with the trained system, where it depicts the distribution concentration compared with the no-regularized model.

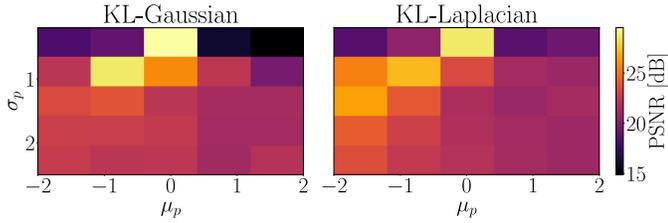


Fig. 5. Recovery performance for the SPC system the KL regularizers with the Gaussian (left) and Laplacian (right) cases.

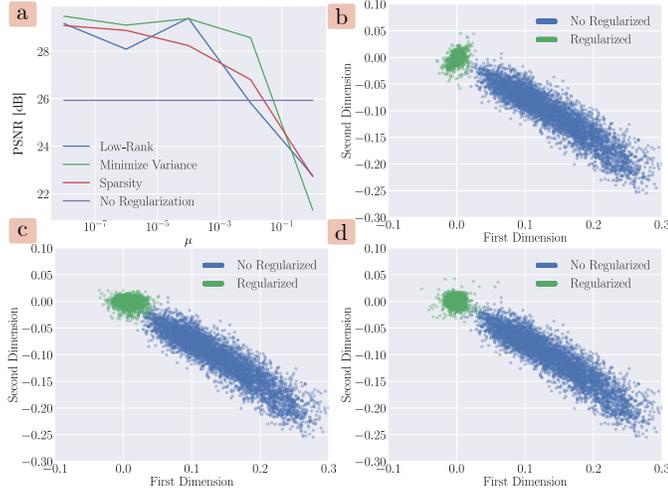


Fig. 6. Recovery PSNR performance for the SPC system with the minimum variance and structural regularizers for different regularization parameter μ_p (a) measurements distribution comparison for the non-regularized design with the sparsity (b), variance minimization regularization (c) and low-rank (d)

B. SPC experiments

For the SPC, we performed experiments on classification and recovery tasks. The classification is performed directly from the compressed measurements without reconstructing the underlying scene. During the training of the E2E network, the parameter of the physical constraint regularizer ρ was dynamically updated during training as suggested in [19], which in the first epochs the ρ is very low, thus not constraining the training of the SL and it is increased to obtain a binary CA. For both the recovery and classification tasks, we employed the Fashion MNIST dataset with 60000 images for training and 10000 for testing. All images were resized to 32×32 .

Recovery experiments: For this experiment, we vary the values of μ_p from -2 to 2, and σ_p was changed from 0.1 to 2.0, taking five equispaced values. The CD in this experiment is a UNET [46] with five downsampling and five upsampling blocks. The results of this experiment are shown in Fig. 5. Here, the performance obtained is similar to that obtained in the CS case, where lower variance yields better reconstruction performance. Also, similar to the results in Fig. 3, the optimal performance is obtained in $\mu_p = 0$, following the concept of batch normalization where the centered output distribution yields more stable training and better performance.

Then, we evaluate the variance and structural regularizers (R_{Vmin} , R_{LR} and R_S) in the recovery task for the SPC architecture. To this end, a study of the hyperparameter μ_p was performed, varying the μ_p from 10^{-8} to 10^0 in a logarithmic scale.

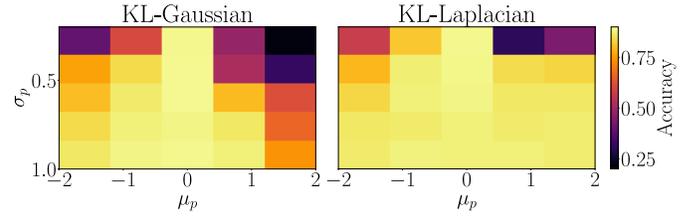


Fig. 7. Classification performance for the SPC system the KL regularizers with the Gaussian (left) and Laplacian (right) cases.

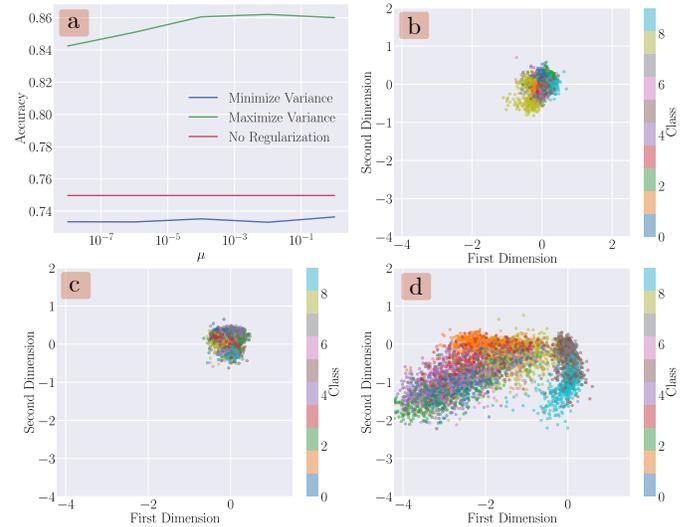


Fig. 8. Classification accuracy performance for the SPC system with the minimum and maximum variance regularizers for different regularization parameters μ_p (a) measurements distribution comparison for the non-regularized. Measurements distribution of the non-regularized design (b), minimum variance (c), and maximum variance (d).

The results of this experiment are compared with the baseline E2E (non-regularized training). The reconstruction performance measured in PSNR of this experiment is shown in Fig. 6(a). The results suggest that in most cases the proposed regularized outperforms the baseline design. Later, the distribution of the first two SPC snapshots was plotted for the best performant setting of each regularizer, the variance minimization Fig. 6(b), sparsity Fig. 6(c) and low-rank Fig. 6(d). Each scatter plot is also shown the distribution obtained by the non-regularized sensing matrix design. In all cases, the resultant distribution employing the regularizers is more concentrated than the non-regularized validating.

Classification experiments: Here, we evaluate the proposed regularization functions on the classification high-level task. The CD is a Mobilnet-V2 [45], which is a lightweight classification network. The same values in the experiment of Fig. 5 of μ_p and σ_p were used in this scenario. The results are shown in Fig. 7, where an opposite performance is obtained compared to the recovery case. Higher variance gives better classification performance.

Then, employ the variance regularization (R_{Vmin} and R_{Vmax}) in the classification task. A study of the hyperparameter μ_p was performed, varying the μ_p from 10^{-8} to 10^0 in a logarithmic scale. The maximum variance value of R_{Vmax} was set to

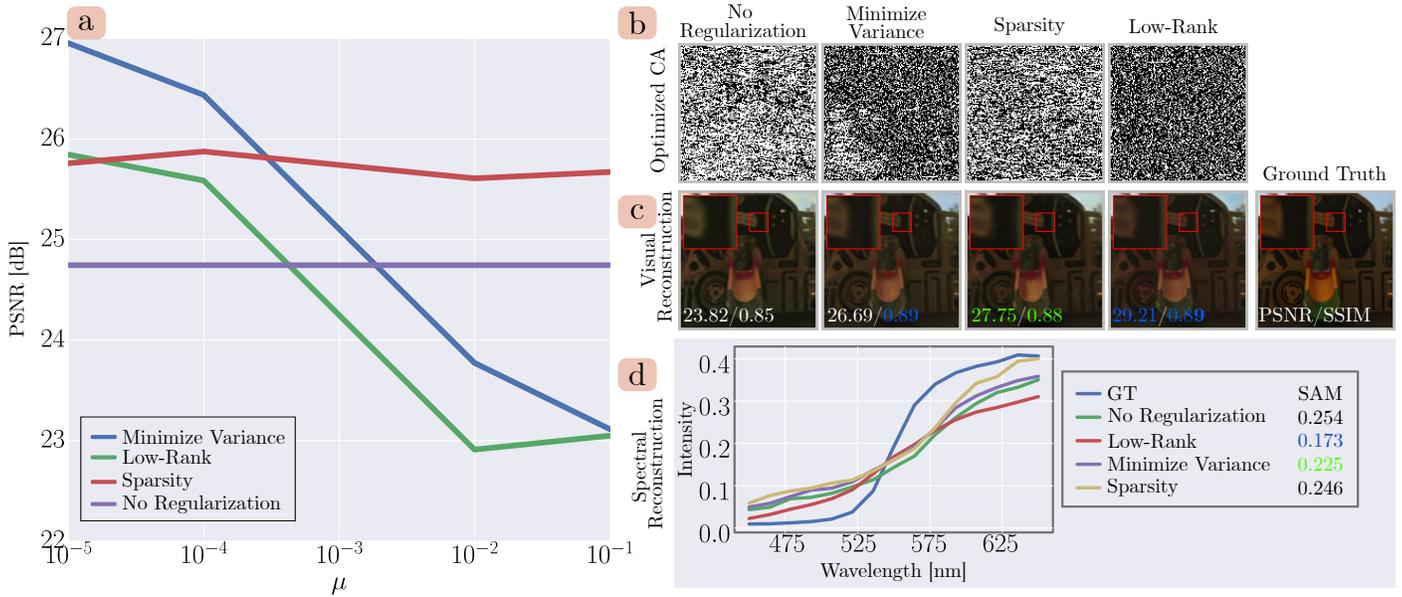


Fig. 9. Reconstruction PSNR performance for CASSI system varying the regularization parameter μ_p (a), the optimized CA for non-regularized and regularized training (b) visual reconstruction (c), and spectral reconstruction (d). The blue highlighted values correspond to the best results and the green to the second best.

$\sigma_{max} = 5$ as we saw better performance with this setting. The results of this experiment are compared with the baseline E2E (non-regularized training). The accuracy performance of this experiment is shown in Fig 8(a). These results show that the variance maximization regularizer outperforms the baseline and the variance minimization regularization. Additionally, the regularization $R_{V_{min}}$ underperforms the baseline validating that more concentrated distribution negatively affects the decoder performance. Then, the distribution of the first two SPC snapshots was plotted for the best performing setting of each regularizer, the non-regularized design Fig. 8(b), variance minimization Fig. 8(c) and variance maximization Fig. 8(d). The colors on the scatter represent the corresponding class of each measurement. While in the baseline and minimize variance distributions, the classes are hardly identified, in the variance maximization design, the measurements of each class are clustered which helps the decoder to classify better the data.

C. CASSI experiments

Here, we aim to design the CA of the CASSI with the proposed regularization functions. Then, we employed the ARAD 1K dataset [47] to train the E2E network where the spectral image size was set to $128 \times 128 \times 31$ by resizing the original dataset spectral image dimension. The regularizers $R_{V_{min}}$, R_S , and R_{LR} were used for this scenario since we want to recover the spectral image from the compressed measurements. Then, we first evaluate the performance with respect to the regularization parameter μ_p compared with the non-regularization design. This parameter was varied from 10^{-8} to 10^0 in a logarithmic scale. The results in Fig 9(a) show that the proposed regularizer improves upon the non-regularized setting where the low-rank is, in this case, the one that provides the best performance. In Fig. 9(b) is shown the optimized CA for the non-regularized and regularized design. Remarkably, the low-rank design converges to a uniform sampling pattern which is a highly desired criterion in compressive imaging sensing matrix design [11, 12]. Fig 9(c) shows a visual reconstruction of a test image with its

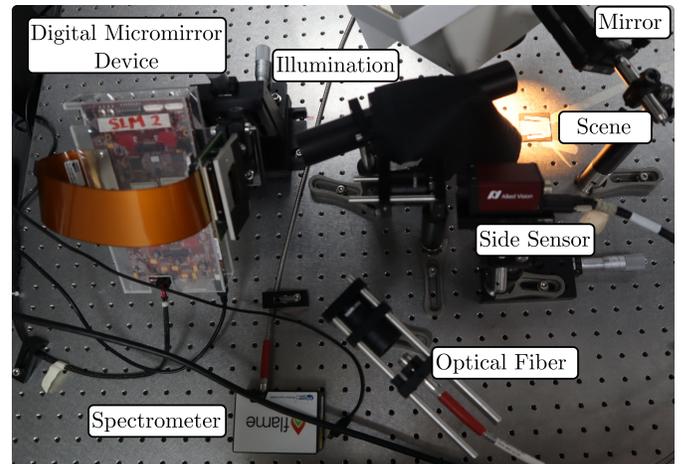


Fig. 10. SPC acquisition system validation of proposed method

corresponding PSNR and SSIM reconstruction values. Finally in Fig 9(d) the reconstruction of a red spectral signature is plotted with the corresponding SAM value. These last results show that the best results corresponds also to the low-rank design.

6. EXPERIMENTAL VALIDATION

To perform the experimental validation of the proposed and theoretically proven regularizers, the following optical systems were implemented.

A. SPC Implementation

The single-pixel system (SPC) was implemented employing a group of lenses that concentrate the light on a single pixel which is focused at the entrance of the optical fiber. The illumination used was a 3900E lamp from Illumination Technology, which has a spectral range of 400-2200 [nm]. For the implementation of

the CA generated by the regularizers, a reference DMD DLP7000 from Thorlabs was used, which has a pitch of 13.6 [μm]. In this case, the binary levels are either 1 or -1. The modulation effect caused by the -1 level can be implemented by acquiring a measurement with a CA of all ones and subtracting it from each captured snapshot. Also, two types of sensors were used, the first of these is the side information sensor, which is a stingray camera F-145, with a pitch size of 6.45 [μm]. On the other hand, to acquire the SPC measurements, it was used a Flame Vis-Nir spectrometer, which has a spectral range from 350 to 1077 [nm], as shown in Fig. 10.

We employed this architecture to validate the performance of the proposed method. For this experiment, three scenes of the first five classes of the Fashion MNIST dataset were acquired utilizing the implemented SPC system. A re-training of the network was performed with the calibrated and captured CA and using only the images from the first five classes of the Fashion MNIST dataset. From this, the examples in Fig. 11 a) were used as a test to evaluate the performance of the proposed method for every one of the regularizers. Fig. 11 b) shows the confusion matrix for the non-regularized design, the KL-Laplacian, the KL-Gaussian, and the maximize variances regularization. The results suggest that the variance maximization regularization has the most accurate classification performance. Additionally, using the other regularization functions there is an improvement with respect to the non-regularized design.

B. CASSI Implementation

On the other hand, the CASSI system was mounted, which consists of an amici prism to perform light scattering at different wavelengths. Additionally, a Thorlabs DLP7000 DMD was used to perform the scene modulation, with the same specifications mentioned above. Additionally, for acquiring this information 2 stingray cameras were used, which were placed at a distance from the image plane of the lenses. Finally, for the spectral illumination of the scene, a TLS Tunable QTH Light Source monochromator was used, which allowed for illuminating the scene in a spectral range of 400-700 [nm], obtaining 31 spectral bands.

In this experiment Fig 13, we performed the acquisition of several scenes by varying the CA implemented in the DMD. These CA were generated from the proposed model by varying the regularizers used, which are minimum variance, low-rank, and without regularizer. From these captures the reconstruction of the scene was performed in a range of spectral bands ranging from 400-700 [nm], where it is observed that the behavior of the proposed model along the spectral range produces less artifact with the proposed design than with the base E2E design. Additionally, a region of interest in the reconstructed images was analyzed, where the mean spectral signature is plotted along with the SAM metric. This result shows that with the proposed CA design, a more accurate spectral reconstruction is obtained.

7. CONCLUSION AND DISCUSSION

We proposed a set of regularization functions over the output of the optical encoder layer within an E2E optimization of optics and image processing framework. These regularizations promote some statistical properties over the coded measurements i.e. concentrate or spread the distribution of the measurements. We found that the optimal distribution depends on the computational task; for the recovery task, a concentrated distribution allows better performance while for optimal classification per-

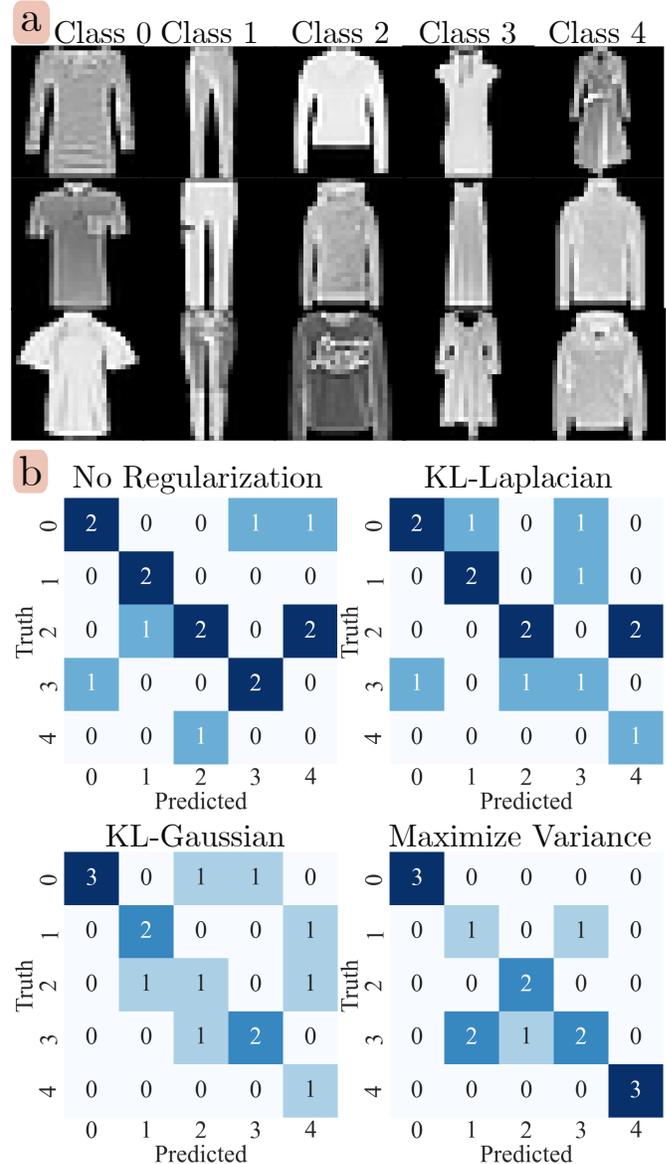


Fig. 11. Validation of the proposed method through the SPC acquisition system, for the classification task. (a) Scenes from the acquired Fashion-MNIST dataset with SPC implementation. (b) Classification confusion matrix for non-regularized design and with the regularization functions.

formance, a wider distribution is desired. We validate the design of the optical coding elements via the regularized E2E optimization in different optical architectures showing improvement with respect to the non-regularized design in both simulation and experimental setups.

Here we employed optical architecture to validate the proposed regularized E2E optimization, however, this methodology can be extended to a general sensing matrix design such as the design of geometry design in a compressive seismic acquisition scenario [28]. Additionally, beyond the sensing matrix design, these regularizations can also be employed in high-level tasks such as generative models [?] where the variance of the generated samples is maximized to have high diversity synthetic samples.

Acknowledgments. This work was supported by project

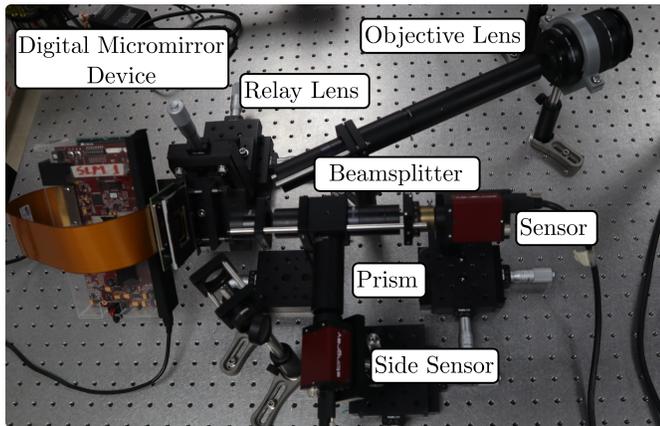


Fig. 12. Experimental prototype of the CASSI acquisition System

110287780575 through the agreement 785-2019 between the Agencia Nacional de Hidrocarburos and the Ministerio de Ciencia, Tecnología e Innovación and Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación Francisco José de Caldas.

Disclosures. The authors declare no conflicts of interest.

REFERENCES

- G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Process. Mag.* **31**, 105–115 (2014).
- C. Fu, H. Arguello, B. M. Sadler, and G. R. Arce, "Compressive spectral polarization imaging with coded micropolarizer array," in *Compressive Sensing IV*, vol. 9484 (SPIE, 2015), pp. 59–65.
- J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3d object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, (2019), pp. 10193–10202.
- K. M. León-López and H. A. Fuentes, "Online tensor sparsifying transform based on temporal superpixels from compressive spectral video measurements," *IEEE Transactions on Image Process.* **29**, 5953–5963 (2020).
- M. Hirsch, G. Wetzstein, and R. Raskar, "A compressive light field projection system," *ACM Transactions on Graph. (TOG)* **33**, 1–12 (2014).
- E. Caroli, J. Stephen, G. Di Cocco, L. Natalucci, and A. Spizzichino, "Coded aperture imaging in x-and gamma-ray astronomy," *Space Sci. Rev.* **45**, 349–403 (1987).
- Y. Peng, Q. Fu, H. Amata, S. Su, F. Heide, and W. Heidrich, "Computational imaging using lightweight diffractive-refractive optics," *Opt. express* **23**, 31393–31407 (2015).
- A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.* **47**, B44–B51 (2008).
- S. R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Appl. optics* **28**, 4344–4352 (1989).
- E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
- C. V. Correa, H. Arguello, and G. R. Arce, "Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging," *JOSA A* **33**, 2312–2322 (2016).
- H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Transactions on Image Process.* **23**, 1896–1908 (2014).
- J. N. Mait, G. W. Euliss, and R. A. Athale, "Computational imaging," *Adv. Opt. Photonics* **10**, 409–483 (2018).
- F. Heide, Q. Fu, Y. Peng, and W. Heidrich, "Encoded diffractive optics for full-spectrum computational imaging," *Sci. reports* **6**, 1–10 (2016).
- D. S. Jeon, S.-H. Baek, S. Yi, Q. Fu, X. Dun, W. Heidrich, and M. H. Kim, "Compact snapshot hyperspectral imaging with diffracted rotation," *Assoc. for Comput. Mach. (ACM)* (2019).
- Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature* **521**, 436–444 (2015).
- H. Arguello, J. Bacca, H. Kariyawasam, E. Vargas, M. Marquez, R. Hettiarachchi, H. Garcia, K. Herath, U. Haputhanthri, B. S. Ahluwalia *et al.*, "Deep optical coding design in computational imaging: a data-driven framework," *IEEE Signal Process. Mag.* **40**, 75–88 (2023).
- E. Vargas, J. N. Martel, G. Wetzstein, and H. Arguello, "Time-multiplexed coded aperture imaging: Learned coded aperture and pixel exposures for compressive imaging systems," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), pp. 2692–2702.
- J. Bacca, T. Gelvez-Barrera, and H. Arguello, "Deep coded aperture design: An end-to-end approach for computational imaging tasks," *IEEE Transactions on Comput. Imaging* **7**, 1148–1160 (2021).
- R. Jacome, J. Bacca, and H. Arguello, "D 2 uf: Deep coded aperture design and unrolling algorithm for compressive spectral image fusion," *IEEE J. Sel. Top. Signal Process.* (2022).
- R. Jacome, J. Bacca, and H. Arguello, "Deep-fusion: An end-to-end approach for compressive spectral image fusion," in *2021 IEEE International Conference on Image Processing (ICIP)*, (IEEE, 2021), pp. 2903–2907.
- E. Martínez, E. Vargas, and H. Arguello, "Fast disparity estimation from a single compressed light field measurement," in *2022 30th European Signal Processing Conference (EUSIPCO)*, (IEEE, 2022), pp. 1991–1995.
- V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," *ACM Transactions on Graph. (TOG)* **37**, 1–13 (2018).
- J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3d object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), pp. 10193–10202.
- D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114* (2013).
- G. P. Meyer, "An alternative probabilistic interpretation of the huber loss," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (2021), pp. 5261–5269.
- R. Jacome, A. Hernandez-Rojas, and H. Arguello, "Probabilistic regularization for end-to-end optimization in compressive imaging," in *Computational Optical Sensing and Imaging*, (Optica Publishing Group, 2022), pp. CW1B–1.
- R. Jacome, H. Arguello, A. Hernandez-Rojas, and P. Goyes-Penafiel, "Divergence-based regularization for end-to-end sensing matrix optimization in compressive sampling systems," *SIGNAL* 2023 Ed. p. 79 (2023).
- Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis machine intelligence* **35**, 1798–1828 (2013).
- S. Rifai, G. Mesnil, P. Vincent, X. Muller, Y. Bengio, Y. Dauphin, and X. Glorot, "Higher order contractive auto-encoder," in *Joint European conference on machine learning and knowledge discovery in databases*, (Springer, 2011), pp. 645–660.
- M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal processing magazine* **25**, 83–91 (2008).
- J. Bacca, L. Galvis, and H. Arguello, "Coupled deep learning coded aperture design for compressive image classification," *Opt. express* **28**, 8528–8540 (2020).
- J. Bacca, T. Gelvez-Barrera, and H. Arguello, "Invariant coded aperture design for compressive imaging," in *Adaptive Optics and Applications*, (Optica Publishing Group, 2022), pp. JTh2A–9.
- T. Nguyen, T. Le, H. Vu, and D. Phung, "Dual discriminator generative adversarial nets," *Adv. neural information processing systems* **30** (2017).
- W.-C. Hung, V. Jampani, S. Liu, P. Molchanov, M.-H. Yang, and J. Kautz,

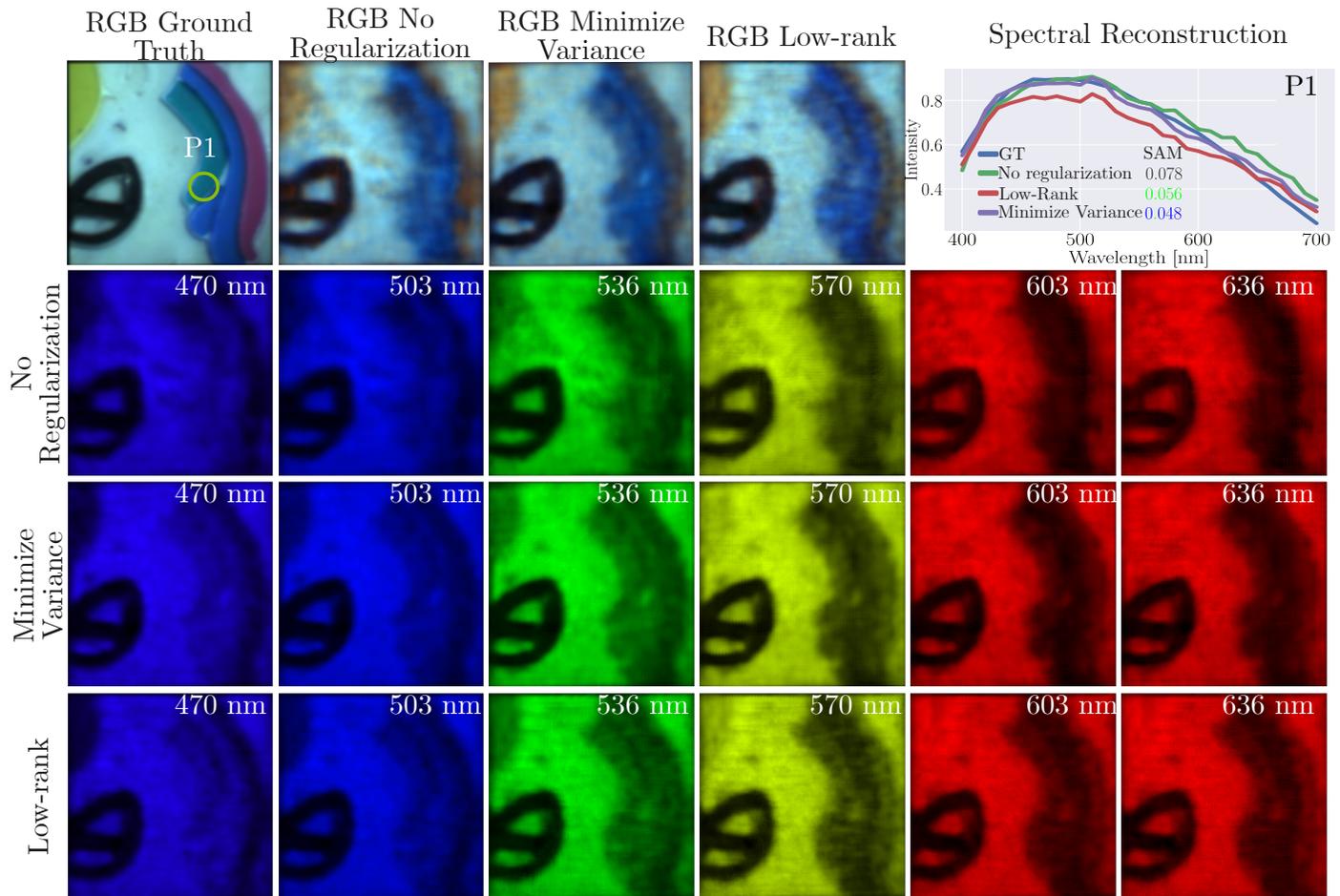


Fig. 13. Reconstruction performance of CASSI system varying the regularization parameter on different wavelengths, and spectral signature reconstruction of a region of interest. The blue SAM value refers to the best performance and the green to the second best.

- "Scops: Self-supervised co-part segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2019), pp. 869–878.
36. C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2020), pp. 1375–1385.
 37. A. Bardes, J. Ponce, and Y. Lecun, "Vicreg: Variance-invariance-covariance regularization for self-supervised learning," in *ICLR 2022-10th International Conference on Learning Representations*, (2022).
 38. K. Evtimova and Y. LeCun, "Sparse coding with multi-layer decoders using variance regularization," arXiv preprint arXiv:2112.09214 (2021).
 39. M. Fazel, E. Candes, B. Recht, and P. Parrilo, "Compressed sensing and robust recovery of low rank matrices," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, (IEEE, 2008), pp. 1043–1047.
 40. I. W. Selesnick and M. A. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," in *Wavelets XIII*, vol. 7446 (spie, 2009), pp. 107–121.
 41. C. A. Osorio Quero, D. Durini, J. Rangel-Magdaleno, and J. Martinez-Carranza, "Single-pixel imaging: An overview of different methods to be used for 3d space reconstruction in harsh environments," *Rev. Sci. Instruments* **92**, 111501 (2021).
 42. L. Galvis, H. Arguello, and G. R. Arce, "Coded aperture design in mismatched compressive spectral imaging," *Appl. Opt.* **54**, 9875–9882 (2015).
 43. R. Stojek, A. Pastuszczak, P. Wróbel, and R. Kotyński, "Single pixel imaging at high pixel resolutions," *Opt. Express* **30**, 22730–22745 (2022).
 44. A. Jerez, H. Garcia, and H. Arguello, "Single pixel spectral image fusion with side information from a grayscale sensor," in *2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (ColCACI)*, (2018), pp. 1–6.
 45. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), pp. 4510–4520.
 46. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, (Springer, 2015), pp. 234–241.
 47. B. Arad, R. Timofte, R. Yahel, N. Morag, A. Bernat, Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang *et al.*, "Ntire 2022 spectral recovery challenge and data set," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2022), pp. 863–881.