Optimization of a Computational Imaging System by Statistical Regularization Based on a Deep Learning Method

Román Alejandro Jácome Carrascal

Master thesis to qualify for the title of Master of Applied Mathematics

Advisor

Henry Arguello Fuentes

PhD. in Electrical and Computer Engineering

Universidad Industrial de Santander

Faculty of Sciences

Department of Physics

Bucaramanga

2024

Dedicatory

To Karen, my love, my inspiration, and my life partner.

To my family, for their love, guidance and support.

Acknowledgments

I would like to thank Karen, my love, who has stood by me on the long working nights and has cheered me up in the difficult moments with her happiness.

To my family who has supported me in all my decisions during this important stage of my life.

I want to thank my advisor, Professor Henry Arguello, for his constant support and guidance in my academic and personal growth. To all the members of the HDSP group who have helped me with the research projects, and have eased the long work hours with their kind and pleasant company.

Table of Contents

List of Figures

Abstract

Title: Optimization of a Computational Imaging System by Statistical Regularization Based on a Deep Learning Method *

Autor: Eng. Roman Alejandro Jacome Carrascal **

Keywords: Computational Imaging, End-to-End Optimization, Inverse Problems, Regularized Optimization.

Description: 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. Although the training performance of the whole network is acceptable, the encoder layers can be flawed, leading to deficient OCE designs. This flawed performance in the encoder arises from factors such as the network's loss function 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 is defined 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 optimization, where a set of regularization functions are used to overcome the flawed optimization of the encoder. In fact, our regularization does not require additional knowledge from the encoder and can be applied to most optical sensing instruments in computational imaging. Accordingly, the regularization exploits some prior knowledge

---

* Master Thesis

** Faculty of Sciences. Department of Physics. Advisor: Henry Arguello Fuentes, Ph.D. in Electrical and Computer Engineering.

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. Moreover, the proposed regularization was employed for high-level computer vision tasks in generative models showing its efficiency also in this new application.

Resumen

Título: Optimización de un Sistema de Codificación Óptico-Computacional Mediante Regularización Estadística Basado en un Método de Aprendizaje Profundo *

Autor: Román Alejandro Jácome Carrascal **

Palabras Clave: Imágenes computacionales, optimización de extremo a extremo, problemas inversos, optimización regularizada.

Descripción: La codificación óptica es una técnica esencial en la imagen computacional (IC) que permite la detección de señales de alta dimensión a través de proyecciones codificadas post-procesadas para decodificar la señal subyacente. En la actualidad, los elementos de codificación óptica (OCE) se optimizan de extremo a extremo (E2E), donde un conjunto de capas (codificador) de una red neuronal profunda modela el OCE mientras que el resto de la red (decodificador) realiza una tarea computacional determinada. Aunque el rendimiento de entrenamiento de toda la red es aceptable, la capa del codificador óptico pueden ser defectuosas, dando lugar a diseños de OCE deficientes. Este rendimiento defectuoso en el codificador se debe a factores como que la función de pérdida de la red no considera las capas intermedias por separado, ya que se desconoce la salida en esas capas. En segundo lugar, el codificador sufre la desaparición del gradiente, ya que el codificador se define en las primeras capas. En tercer lugar, la estimación adecuada del gradiente en estas capas está restringida a satisfacer limitaciones físicas. En este trabajo, proponemos una optimización de extremo a extremo regularizado la salida intermedia de la red, en la que se utiliza un conjunto de funciones de regularización para superar la optimización defectuosa del codificador óptico. De hecho, nuestra

---

* Trabajo de grado

** Facultad de Ciencias. Escuela de Física. Director: Henry Arguello Fuentes, Doctorado en Ingeniería Eléctrica y Computación.

regularización no requiere conocimientos adicionales del codificador y puede aplicarse a la mayoría de los instrumentos de detección óptica en imagen computacional. En consecuencia, la regularización explota algunos conocimientos previos sobre la tarea computacional, las propiedades estadísticas de la salida del codificador (medidades codificadas) y del sistema de adqusición. En concreto, propusimos tres tipos de regularizadores: El primero se basa en las divergencias estadísticas de las medidas comprimidas, el segundo depende sólo de la varianza de las medidas, y el último es un regularizador estructural que promueve el bajo rango y escacez del conjunto de medidas. Hemos validado el procedimiento de entrenamiento propuesto en dos sistemas de imagen computacional representativos, la cámara de píxel único (SPC) y el sistema de única cáptura de imágenes espectrales con aperturas codificadas (CASSI), mostrando una mejora significativa con respecto a los diseños no regularizados. Además, la regularización propuesta se empleó para tareas de visión por computador de alto nivel en modelos generativos mostrando su eficacia también en esta nueva aplicación.

Research Products

Contributions of the thesis

- We provide insights on optimality criteria for designing optical coding elements via end-to-end optimization for different computational tasks. Our insights suggest that contracting the distribution of the coded measurements allows better performance for reconstruction.

- We proposed a set of regularization functions over the optical coding elements that improve the performance of the network for several computational tasks and with different optical systems.

    - Statistical divergence functions where we aim to approximate the distribution of the coded measurements to a prior distribution.

    - Variance regularization by minimizing or maximizing the variance of the measurements.

    - 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. Also sparsity in a given basis, e.g., wavelet along the measurement set to promote smoothness, reducing the data variability.

- We extensively validate the proposed design methodology with both simulation and real acquisition scenarios, where our method outperforms the non-regularization optimization.
- Further applications beyond computational imaging scenarios of the proposed regularization functions are shown for computer vision tasks.

Publications

The developments of this thesis have been disseminated in various international journals and conferences. **Journal papers:**

1. **Jacome, Roman**, Pablo Gomez., and Henry Arguello. Middle Output Regularized End-to-End Optimization for Computational Imaging. In *Optica (2023)* .
2. Urrea, Sergio, **Roman Jacome**, Salman M. Asif, Henry Arguello, and Hans Garcia. DoDo: Double DOE Optical System for Multishot Spectral Imaging. Submitted to *IEEE Transactions on Computational Imaging*

Conference papers:

1. **Jacome, Roman,** Alejandra Hernandez-Rojas, and Henry Arguello. Probabilistic regularization for end-to-end optimization in compressive imaging. *Computational Optical Sensing and Imaging. Optica Publishing Group, 2022.*
2. Martinez, Emmanuel, **Roman Jacome**, Alejandra Hernandez-Rojas, and Henry Arguello. LD-GAN: Low-Dimensional Generative Adversarial Network for Spectral Image

Generation with Variance Regularization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023,*

3. **Jacome, Roman**, Henry Arguello, Alejandra Hernandez-Rojas, and Paul Goyes-Penafiel. Divergence-Based Regularization for End-to-End Sensing Matrix Optimization in Compressive Sampling Systems. *In SIGNAL 2023*

4. Urrea, Sergio, **Roman Jacome**, Salman M. Asif, Henry Arguello, and Hans Garcia. Optical Solutions for Spectral Imaging Inverse Problems with a Shift-Variant System. *IEEE/CVF International Conference on Computer Vision 2023*

1. Problem statement

The coding elements in optical-computational imaging systems play a key role in the fidelity of the acquired data and the quality of their reconstruction. For this reason, a great variety of works have endeavored to propose different methods for the design of these coding elements. Currently, the best-performing method in the state of the art is end-to-end optimization. In this method, the optical system is modeled as a layer of a neural network which is optimized in conjunction with a reconstruction or decoding network.

The layer that models the optical system (optical encoder) is the first layer of the model and also the number of parameters of the computational decoder is much larger than the parameters of the optical layer, the training of the optical coding element suffers from gradient fading during the training stage, this results in the suboptimal design of the optical elements and the quality of the reconstruction relies mostly on the training of the computational decoder.

Therefore, in this thesis, we propose the use of regularizing functions that allow an optimal optimization of the coding elements, in order to improve the overall performance of the optical computational system in terms of the reconstructed image quality. For the formulation of these regularization functions, it is proposed to study the statistical properties of the set of encoded measurements i.e., the output of the optical layer.

2. Objectives

General Objective

To optimize a computational optical coding system from a regularizing function based on statistical properties of the system in end-to-end deep learning methods.

Specific Objectives

1. To model a computational optical coding system using a differentiable parameterization to be incorporated as layers in an end-to-end deep neural network, where the coding elements are the trainable variables.

2. To determine a regularization function based on statistical features of the coding system, which allows an improvement of the performance of the decoding neural network.

3. To validate the performance obtained with the proposed regularizing function for the computational optical coding system in the decoding task, comparing it with the non-optimized system and with the optimized system without the regularization.

4. To evaluate the design of the designed optical coding element obtained with the proposed method by decoding real measurements obtained with a laboratory prototype.

3. Introduction

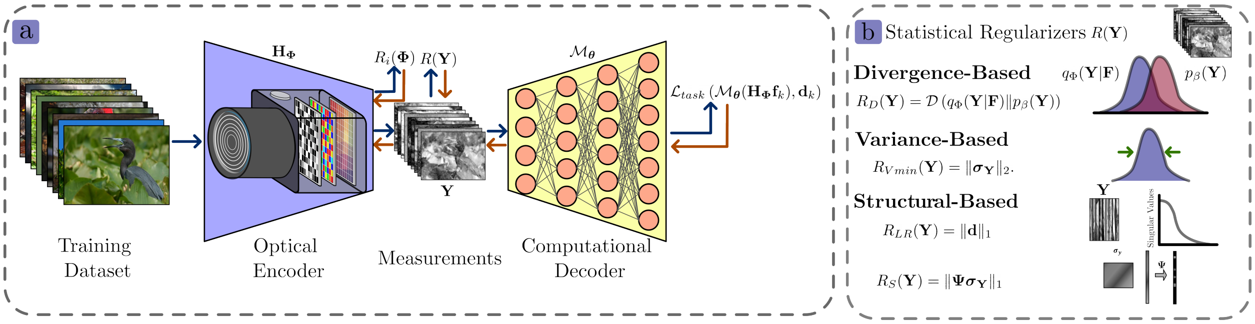
The joint operation of optical systems and computational algorithms in computational imaging (CI) has allowed the acquisition of D-dimensional signals, $D > 2$, such as spectral imaging Arce et al. (2014), polarization state Fu et al. (2015), depth imaging Chang and Wetzstein (2019a), temporal imaging León-López and Fuentes (2020), and angular views in light fields Hirsch et al. (2014). 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) Caroli et al. (1987), phase using diffractive lenses Peng et al. (2015), polarization using micro-polarizers Fu et al. (2015) or spectral information employing dispersive elements Wagadarikar et al. (2008). 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 Caroli et al. (1987); Gottesman and Fenimore (1989) or compressive sensing theory Candes and Wakin (2008), such as the restricted isometry property Correa et al. (2016); Arguello and Arce (2014). Additionally, in the design of diffractive lenses, methods have been proposed to reduce chromatic aberrations and geometries Mait et al. (2018) to improve CI systems. Moreover, these elements have been designed for the encoding of spectral information Heide et al. (2016); Jeon et al. (2019). 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 LeCun et al. (2015), and a large number of databases available, the end-to-end optimization method Arguello et al. (2023) 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). Hence, 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 Vargas et al. (2021), classification, and depth estimation Bacca et al. (2021), compressive spectral image fusion Jacome et al. (2022, 2021), extended chromatic field of view and super-resolution Sitzmann et al. (2018), or monocular depth estimation Chang and Wetzstein (2019b) among others, optimization of the OE can be subpar because of several reasons. For instance, the OE parameters can be only optimized with respect to the loss function computed with the output of the 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 the training stage.

To overcome the OCE training issues, we propose middle output regularized end-to-end, 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 some criteria to better optimize 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. In particular, the Gaussian distribution (widely used in variational autoencoders Kingma and Welling (2013)) and Laplacian distribution (employed in regression tasks Meyer (2021)) 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 obtain better task performance. 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 Bengio et al. (2013). Contractive representations have been used in traditional autoencoders Rifai et al. (2011). 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 to any optical architecture and can be adapted for any computational task. An illustration of the proposed regularized E2E optimization method is shown in Figure 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) Duarte et al. (2008) for imaging, and the coded aperture snapshot spectral imager (CASSI) Wagadarikar et al. (2008), for spectral imaging. Finally, a compressive seismic acquisition system

Figure 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 stage.



was also employed Mosher et al. (2017). The OCE design of these systems has been addressed with the E2E framework. For instance, in the SPC, authors in Higham et al. (2018) design the CA for single-pixel video, also in Bacca et al. (2021) the author employs the E2E framework to optimize the CASSI CA, and in compressive seismic acquisition for the design of geometry settings Hernandez-Rojas and Arguello (2022).

4. 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 $m \ll n$. Here, we focus on linear computational imaging systems. In this case, the the E2E optimization framework, the sensing procedure is modeled as a differentiable linear operator, i.e.,

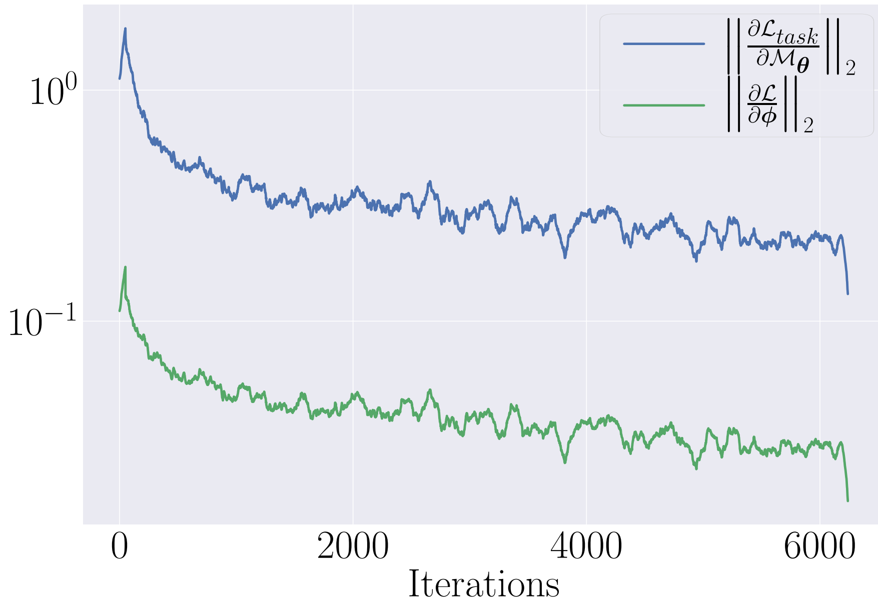
$$\mathbf{y} = \mathbf{H}_{\mathbf{\Phi}}\mathbf{f} + \boldsymbol{\omega}, \tag{1}$$

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

$$\{\boldsymbol{\theta}^\star, \mathbf{\Phi}^\star\} = \underset{\boldsymbol{\theta},\mathbf{\Phi}}{\arg\min}\, \mathcal{L}(\boldsymbol{\theta}, \mathbf{\Phi}) = \underset{\boldsymbol{\theta},\mathbf{\Phi}}{\arg\min}\, \frac{1}{K}\sum_{k=1}^{K} \mathcal{L}_{task}\left(\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{H}_{\mathbf{\Phi}}\mathbf{f}_k), \mathbf{d}_k\right) + \rho R_i(\mathbf{\Phi}), \tag{2}$$

where $\{\mathbf{f}_k\}_{k=1}^K$ is the training dataset, $\mathcal{L}_{task}(\cdot)$ is the loss function of desired tasks, $\mathbf{d}_k$ corresponds to the expected output, e.g., classification labels Bacca et al. (2020), ground truth image Jacome et al. (2022), depth maps Chang and Wetzstein (2019a) etc. Usually, the OCE is constrained to a set of feasible values from the physical limitations of the elements. To impose this constraint, a regularization function $R_i(\mathbf{\Phi})$ is added to the loss function, where ρ is the regularization parameter. This regularization can also induce the desired properties on

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



the OCE, such as transmittance in CA, number of shots, etc., (Arguello et al., 2023, Table II). The 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

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

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 because 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 the norm of the loss function gradient with respect to the $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$ 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 related to the intermediate output of the E2E (the coded measurements), which is not taken into account independently on training, and the optimization is only performed with respect to the CD output. Therefore, 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, Bacca et al. (2021) proposes to minimize a regularization based on concentrating the eigenvalues of the sensing matrix $\mathbf{H}_\Phi$ following the function $\|\mathbf{H}_\Phi^T\mathbf{H}_\Phi\mathbf{f} - \mathbf{f}\|_2$. Similarly Bacca et al. (2022a) 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 to 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 $\mathbf{H}_\Phi$. These functions aim to obtain an approximation of the desired image only with the invertibility properties of the sensing matrix. However, this kind of invertibility is not common 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 are based on the recovery problem 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.

5. Proposed regularization functions

In this thesis, it is proposed a new type of regularization function for E2E optimization, promoting some properties on the distribution of the measurements. The optimization problem (2) becomes

$$\begin{aligned}\{\boldsymbol{\theta}^{\star}, \boldsymbol{\Phi}^{\star}\} &= \underset{\boldsymbol{\theta},\boldsymbol{\Phi}}{\arg\min}\, \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Phi}) \\ &= \underset{\boldsymbol{\theta},\boldsymbol{\Phi}}{\arg\min}\, \frac{1}{K}\sum_{k=1}^{K} \mathcal{L}_{task}\left(\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{H}_{\boldsymbol{\Phi}}\mathbf{f}_k), \mathbf{d}_k\right) + \rho R_i(\boldsymbol{\Phi}) + \mu R(\mathbf{Y}),\end{aligned} \tag{4}$$

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., $\mathbf{Y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_K^T]^T$.

5.1. Divergence-based regularization

This type of regularization function is based on the idea behind variational auto-encoders Kingma and Welling (2013). Particularly, this regularization aims to approximate the probability distribution of the measurements set denoted as the posterior distribution $q_{\boldsymbol{\Phi}}(\mathbf{Y}|\mathbf{F})$, 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

$$R_D(\mathbf{Y}) = \mathcal{D}\left(q_{\Phi}(\mathbf{Y}|\mathbf{F}) \| p_{\beta}(\mathbf{Y})\right), \tag{5}$$

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 Kingma and Welling (2013), generative adversarial networks Nguyen et al. (2017), self-supervised learning Hung et al. (2019) 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 of using the KL divergence is that it has a closed-form solution when P and Q are Gaussian or Laplacian distributions (see Kingma and Welling (2013); Metzler et al. (2020)). 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 $\boldsymbol{\mu}_\mathbf{Y} \in \mathbb{R}^m$ and variance $\boldsymbol{\sigma}_\mathbf{Y} \in \mathbb{R}^m_+$ are computed pixel-wise across the training batch. For the Gaussian case, the KL divergence-based regularizer is defined as

$$R_{KL-G}(\mathbf{Y}) = \log\left(\frac{\sigma_\mathbf{Y}}{\sigma_p}\right) - \frac{\boldsymbol{\sigma}_\mathbf{Y}^2 + (\boldsymbol{\mu}_\mathbf{Y} - \mu_p)^2}{2\sigma_p^2} + \frac{1}{2}, \tag{6}$$

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

$$R_{KL-L}(\mathbf{Y}) = \log\left(\frac{\sigma_\mathbf{Y}}{\sigma_p}\right) - \frac{\sigma_p + e^{\left(\frac{-|\mu_p - \boldsymbol{\mu}_\mathbf{Y}|}{\sigma_p}\right)} + |\mu_p - \boldsymbol{\mu}_\mathbf{Y}|}{\sigma_p} - 1. \tag{7}$$

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.

5.2. 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 Bardes et al. (2022), sparse-coding Evtimova and LeCun (2021). Here, we extrapolate these criteria of optimal low-dimensional representation basis to a compressive sensing system, thus giving more interpretability of the designed OCE by the E2E optimization. The proposed regularization function is given by

$$R_{Vmin}(\mathbf{Y}) = \|\boldsymbol{\sigma}_{\mathbf{Y}}\|_2. \tag{8}$$

For this regularization, we control how much the variance is minimized by tuning the hyperparameter μ on (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

$$R_{Vmax}(\mathbf{Y}) = \|\sigma_{max} - \boldsymbol{\sigma}_{\mathbf{Y}}\|_2, \tag{9}$$

where σ_{max} is a maximum variance reference. This hyperparameter can also be tuned.

5.3. Structural regularization

This type of regularization is based on the common priors of compressed sensing recovery: low-rank and sparsity Candes and Wakin (2008); Fazel et al. (2008). 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 $\mathbf{Y} = \mathbf{UDV}^T$ 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 $\mathbf{d} = [d_i, \dots, d_K]$ where $d_i = \mathbf{D}_{(i,i)}$ for $i = 1, \dots, K$. Thus, our low-rank regularization is the following

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

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

$$R_S(\mathbf{Y}) = \|\mathbf{\Psi}\boldsymbol{\sigma}_{\mathbf{Y}}\|_1 \tag{11}$$

where $\mathbf{\Psi}$ is the representation basis. In this work, we consider the Haar wavelet, which has shown good results in promoting smoothness on signals Selesnick and Figueiredo (2009).

6. Compressive Imaging Sensing Models

To validate the proposed deep optical design, it was employed two flagship CI optical architectures, the CASSI and SPC. Additionally, beyond optics systems, we employ a compressive seismic acquisition scheme.

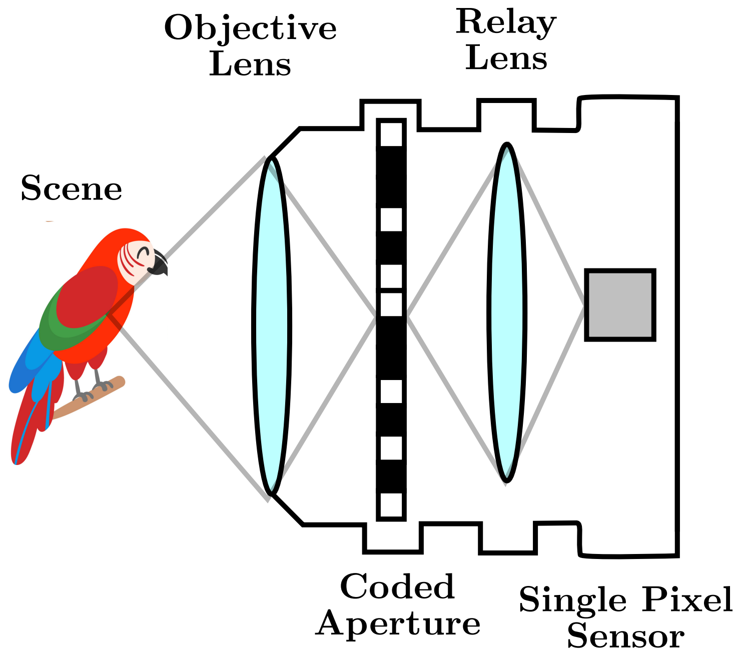
6.1. Single Pixel Camera

The first optical architecture is the single pixel camera (SPC) Duarte et al. (2008), this architecture is widely used in compressive imaging systems. This system employs an imaging lens that spatially introduces light, which is previously modulated by a CA, and then integrates the encoded image into a single-pixel detector. An illustration of the SPC system is depicted in Figure 14. The CA can be implemented with spatial light modulators (SLM) Osorio Quero et al. (2021), such as a digital micro-mirror device (DMD)Galvis et al. (2015), that selectively redirects parts of the light beam Jerez et al. (2018). The SPC uses a CA $\mathbf{\Phi}^k_{(i,j)}$ 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 $\mathbf{\Phi}^k_{(i,j)}$ is a binary pattern whose spatial distribution determines the reconstruction performance. Mathematically, the CA effect over the scene can be represented as:

$$\hat{\mathbf{F}}^k_{(i,j)} = \mathbf{F}_{(i,j)}\mathbf{\Phi}^k_{(i,j)}, \tag{12}$$

After that, the modulated scene $\hat{\mathbf{F}}$ is focused in a single spatial point by the condenser

Figure 3. Single pixel camera scheme. A scene is codified by the CA and this coded field is integrated into a single pixel sensor.

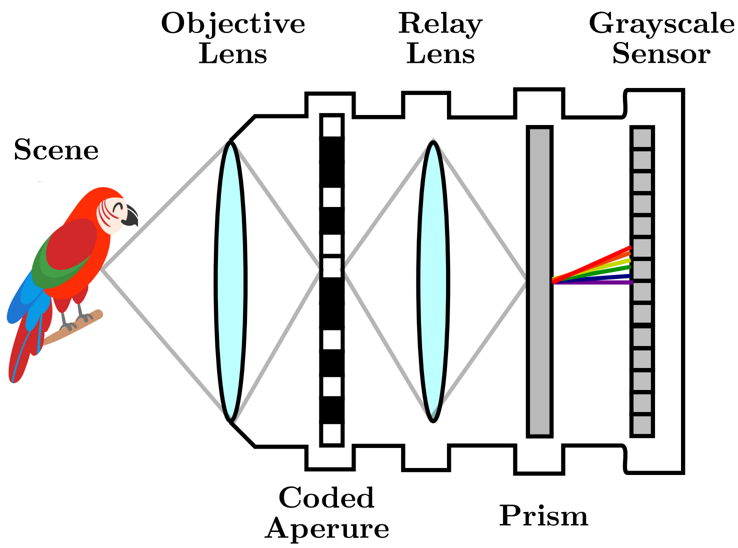


lens, and captured by a single-pixel detector. The resulting sensing matrix $\hat{\mathbf{H}}_{\phi_s} \in \mathbb{R}^{C \times MN}$ contains the vectorization of the CA at each snapshot c in his rows. The aperture codes implemented for the sensing matrix, are the design parameter from the proposed regularizers. The acquisition system is modeled as

$$\mathbf{y} = \hat{\mathbf{H}}_{\phi_s}\mathbf{f} + \mathbf{n}_s, \tag{13}$$

where, $\mathbf{y} = [y_1, ..., y_C]^T$ is the compressed measurements, $\mathbf{f} \in \mathbb{R}^{MN}$ is the vectorized image and $\mathbf{n}_c$ is additive Gaussian noise.

Figure 4. CASSI optical system scheme. A spectral scene is spatially modulated by the CA which is dispersed by the prism. Finally, a grayscale sensor integrates the measurements



6.2. Coded Aperture Snapshot Spectral Imager

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. An illustration of the CASSI system is shown in 4. Therefore, the discrete model of the CASSI measurements $\mathbf{y}_c$ can be formulated as

$$\mathbf{y}_{c_{(i,j)}} = \sum_{\ell=1}^{L} \mathbf{\Phi}_{c_{(i,j)}} \mathbf{F}_{(i,j-\ell,\ell)}, \tag{14}$$

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

$$\mathbf{y}_c = \mathbf{H}_{\mathbf{\Phi}_c}\mathbf{f} + \mathbf{n}_c, \tag{15}$$

where $\mathbf{y}_c \in \mathbb{R}^{M(N+L-1)}$ is the compressed measurements, $\mathbf{H}_{\mathbf{\Phi}_c} \in \mathbb{R}^{M(N+L-1)\times MNL}$ is the CASSI sensing matrix, $\mathbf{f} \in \mathbb{R}^{MNL}$ 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 parameter to be designed is the CA $\mathbf{\Phi}_c$

6.3. Compressive seismic acquisition

The cross-spread is a fundamental seismic acquisition geometry involving one linear arrangement of shot points and receivers perpendicular to each other Yilmaz (2008)Liner (2016). To mathematically represent the seismic data acquired by a cross-spread, let $\mathcal{X} \in \mathbb{R}^{I_1\times I_2\times I_3}$ be a data cube where each dimension represents I_1 time samples, I_2 receivers, and I_3 number of shots. Since economic limitations and environmental constraints, the observed seismic field data is irregular and incomplete along the receiver dimension, leading to a recovery task. To simulate the undersampled data, let $\boldsymbol{\phi} \in \{0,1\}^{I_2}$ be a sampling vector with dimensions equal to the number of receivers. The entries of $\boldsymbol{\phi}$, denoted as ϕ_i, define whether the information is acquired. If $\phi_i = 0$, the receiver is removed; otherwise, $\phi_i = 1$, and it is acquired. The diagonalization of the sampling vector derives the diagonal sampling matrix as $\mathbf{H}_{\boldsymbol{\phi}} = \text{diag}(\boldsymbol{\phi})$. Once $\mathbf{H}_{\boldsymbol{\phi}}$ is built, the undersampled measurements are obtained via n-mode product($\times_n$) defined in Lathauwer et al. (2000)

$$\mathcal{Y} = \mathcal{X} \times_2 \mathbf{H}_{\boldsymbol{\phi}}, \tag{16}$$

where (16) represents the 2-mode product between the full data $\mathcal{X}$ and $\mathbf{H}_{\boldsymbol{\phi}}$. The undersampled measurements $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ contains the removed receivers as columns in zero for each shot.

A conventional relation that determines the number of acquired receivers by the sensing matrix is the transmittance, calculated as

$$\delta_{\boldsymbol{\phi}} = \sum_{i=1}^{M} \frac{\phi_i}{I_2}. \tag{17}$$

For instance, when $\delta_{\boldsymbol{\phi}} = 0.7$, the 70 % of the total receivers are acquired. The E2E optimization is mathematically expressed as

$$\{\hat{\boldsymbol{\phi}}, \hat{\theta}\} = \underset{\phi,\theta}{\arg\min}\, \mathcal{L}\left(\mathcal{N}_\theta\left(\mathcal{X} \times_2 \mathbf{H}_{\boldsymbol{\phi}}\right), \mathcal{X}\right) + \rho R\left(\boldsymbol{\phi}\right), \tag{18}$$

where the regularization $R\left(\boldsymbol{\phi}\right) = \left(\delta_0 - \delta_{\boldsymbol{\phi}}\right)^2$ controls the transmittance to converge to a desired value δ_0, and ρ represents a weight parameter.

7. Simulation Results

7.1. Simulation Settings

7.1.1. Datasets and pre-processing. Three datasets were employed to train the networks. For the SPC experiments, we employed the Fashion MNIST dataset Xiao et al. (2017) which contains 60000 images of 10 classes of clothes. We split this data set into 50000 for training and 10000 for testing. All images were resized to have 32×32. For the CASSI experiments, the ARAD spectral images dataset was used Arad et al. (2022), where the images were resized to have a size of $128 \times 128 \times 31$, and 900 images were used for training 100 for testing. For the compressive seismic acquisition, we employed the synthetic dataset SEAM Phase II built by the SEG Advanced Modeling Program (SEAM) during its second project, named "SEAM Phase II–Land Seismic Challenges". The Foothills model is focused on mountainous regions with sharp topography at the surface and high geological complexity at depth, which makes this data set a challenge for seismic data reconstruction Regone et al. (2017). The seismic survey covers a rectangular patch of 1.5×1.2 km with a total sampled depth of 4100 ms. The training and testing datasets comprise 381 images of 128×128.

7.1.2. Metrics. To measure the reconstruction quality the following metrics were used the peak-signal-to-noise-ratio (PSNR) Horé and Ziou (2010), the structural similarity index measure (SSIM) Wang et al. (2004),

1. PSNR: Measured in dB, is defined as the logarithm of the ratio between the maximum

possible power of a signal and the power of corrupting noise that affects the fidelity of its representation so that a higher value indicates superior quality of reconstruction. And is expressed as

$$\mathrm{PSNR}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \left(\frac{\max(\mathbf{x})^2}{\mathrm{RMSE}(\mathbf{x}, \hat{\mathbf{x}})^2} \right). \tag{19}$$

2. SSIM: As aforementioned, this metric measures the quality of the estimated image in terms of the degradation of the structural information instead of absolute errors. It is implemented on various windows of the image, denotes $\mathbf{f}_X$ and $\hat{\mathbf{f}}_Y$ a window of $S \times S$ of the ground truth image and the estimated image, respectively, then, the SSIM metric is defined as

$$\mathrm{SSIM}(\mathbf{x}_X, \hat{\mathbf{x}}_Y) = \frac{(2\mu_X \mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)}, \tag{20}$$

where μ, σ are the mean value and variance of the window, respectively, σ_{XY} are the covariance of $\mathbf{f}_X$ and $\hat{\mathbf{f}}_Y$, $c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$ are two variables to stabilize the division with weak denominator, L is the number of quantization levels of the image, k_1 and k_2 are hyperparameters, usually 0.01 and 0.03 respectively. For SSIM values close to 1 the quality of the estimated image is better.

3. To evaluate the performance on the classification task, we employ the accuracy metric

defined as

$$A = \frac{1}{C}\sum_{c=1}^{C}\frac{\mathrm{TP}_c}{\mathrm{Total}_c},$$

where C is the number of the classes and TP are the True Positive

7.1.3. Computational Decoders. In particular, we perform classification and recovery tasks, where for the first we use a MobilNet-V2 network Sandler et al. (2018) which is a lightweight model widely used for classification. For the recovery task, a U-Net model with five convolution blocks was used for each downsampling and upsampling process.

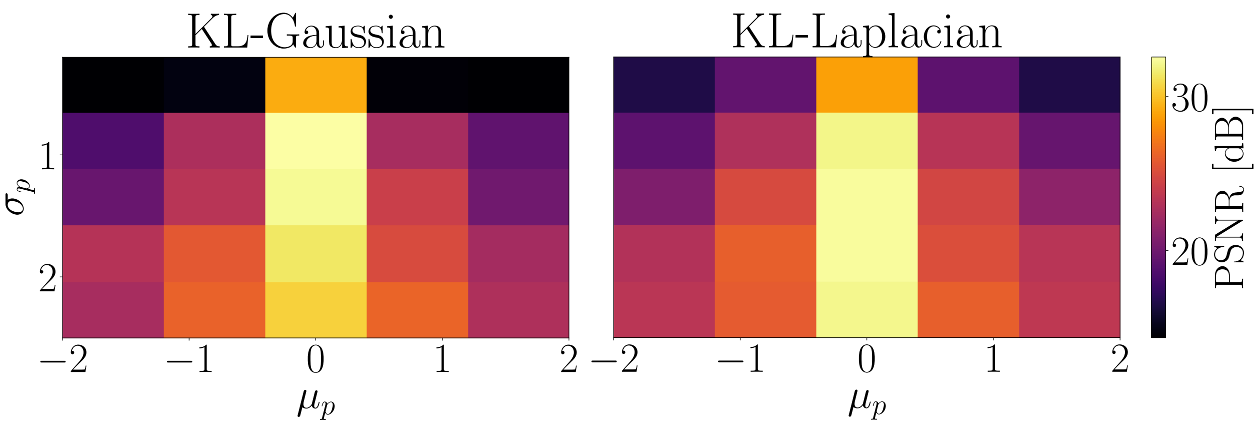
7.1.4. Training settings. For all the experiments, we trained the E2E network for 100 epochs, halving the learning rate every 40 epochs. For the CASSI CA binary constraint, the polynomial regularization in Bacca et al. (2021) was employed i.e., $R(\mathbf{\Phi}) = \sum_{ij}(1-\mathbf{\Phi}_{ij})^2(\mathbf{\Phi}_{ij})^2$. For the SPC CA constraint, we consider values $\{-1, 1\}$ that in practice can be achieved by following the procedure in detail in (Bacca et al., 2020, Appendix), which allows better signal-to-noise ratio (SNR). Then, the physical constraint regularizer is $R(\mathbf{\Phi}) = \sum_{ij}(1-\mathbf{\Phi}_{ij})^2(1+\mathbf{\Phi}_{ij})^2$. The parameter of the physical constraint regularizer ρ was dynamically updated during training as suggested in Bacca et al. (2021).

7.2. Compressed Sensing Experiments

In 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 $\mathbf{H}_{\Phi}$. Here, we use a compression ratio of $10\,\%$.

KL-Divergence: First, we analyze the effect of the mean and variance of the prior

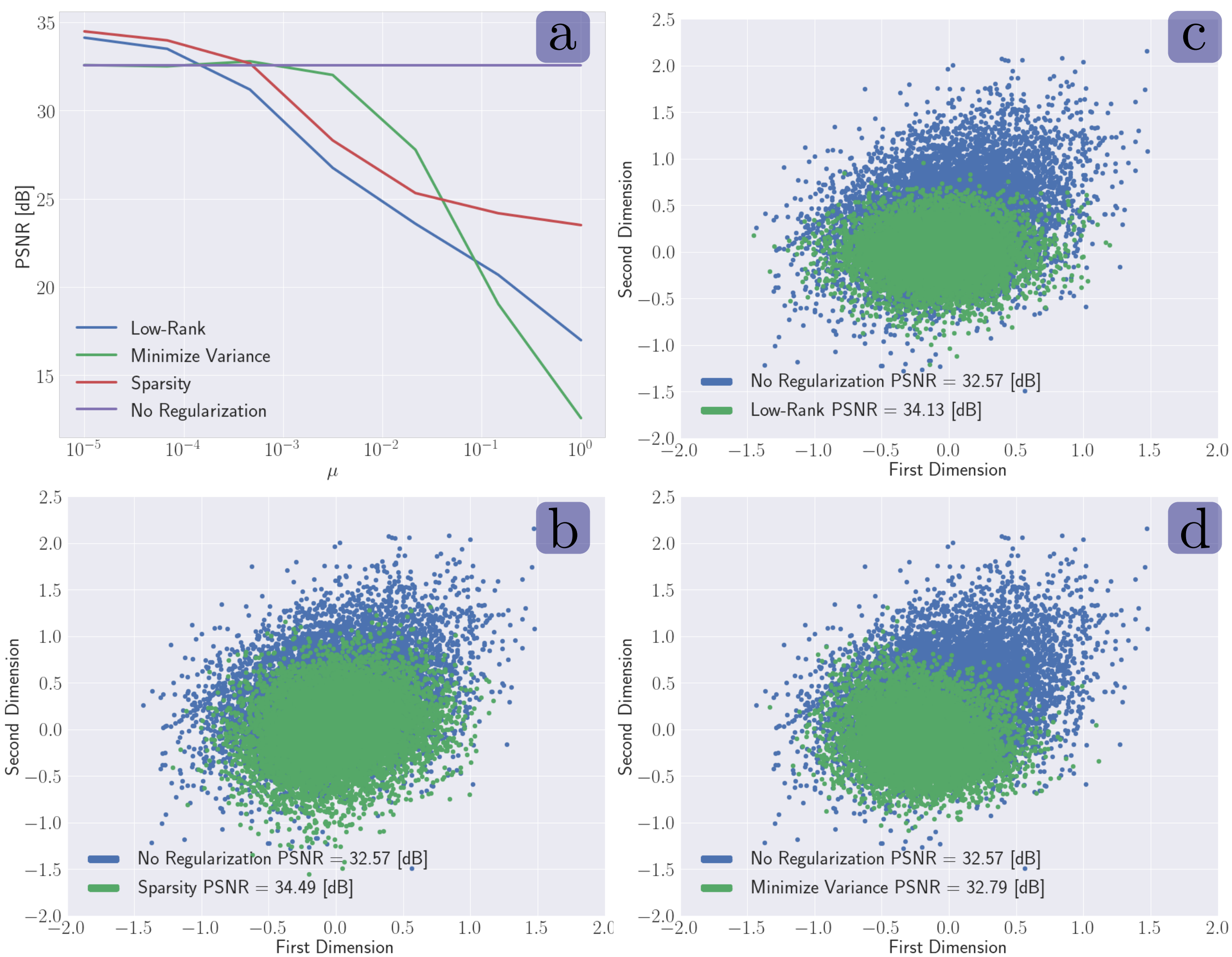
Figure 5. Recovery performance for the CS scenario employing the KL-D regularizers with the Gaussian (left) and Laplacian (right) cases



distribution (μ_p, σ_p) on 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 5 , where optimal reconstruction PSNR values are obtained at variances close to 1.0 and for means close to 0. These results suggest that 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.

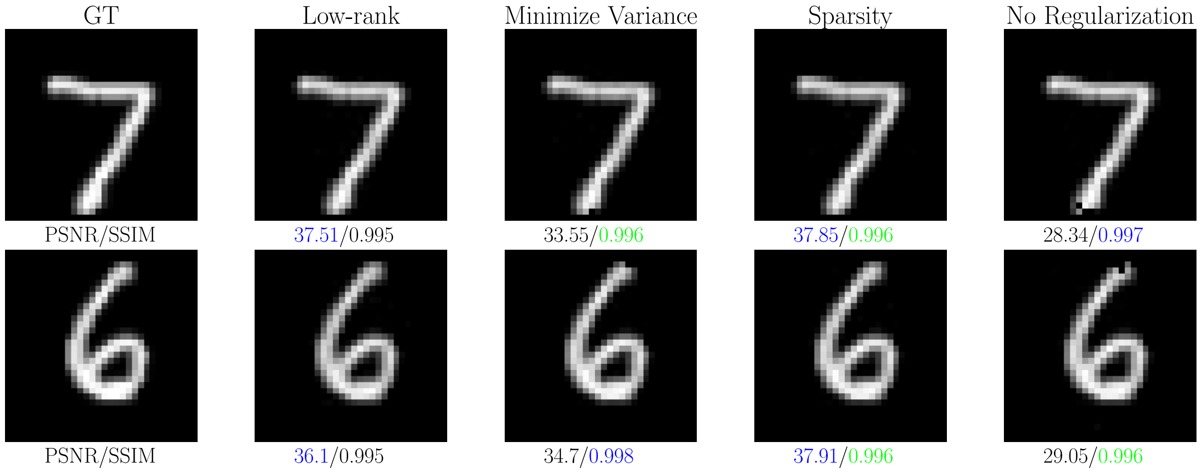
Variance and structural regularizers Then, we analyze the performance of the E2E network for the variance minimization and structural regularization. In Fig 6(a) the recovery performance is shown depending on the regularization parameter in (4) where optimal values for the regularization suggest a trade-off between how much concentrate the distribution and the recovery performance. In particular, significant recovery improvements are shown with the low-rank and sparsity experiments with respect to the baseline (no regu-

Figure 6. 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 data set for b) low rank, c) minimize variance, and d) sparsity.



larization E2E). Fig 6(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. Additionally, Fig. 7 shows some visual reconstructions from two images of the test set, validating that the proposed regularization outperforms the baseline E2E method. Here, the best overall performance was obtained by the sparsity regularization.

Figure 7. Visual reconstruction results for the CS scenario using the variance and the structural regularizations and the baseline E2E designs. The blue values correspond to the best results and the green to the second best.

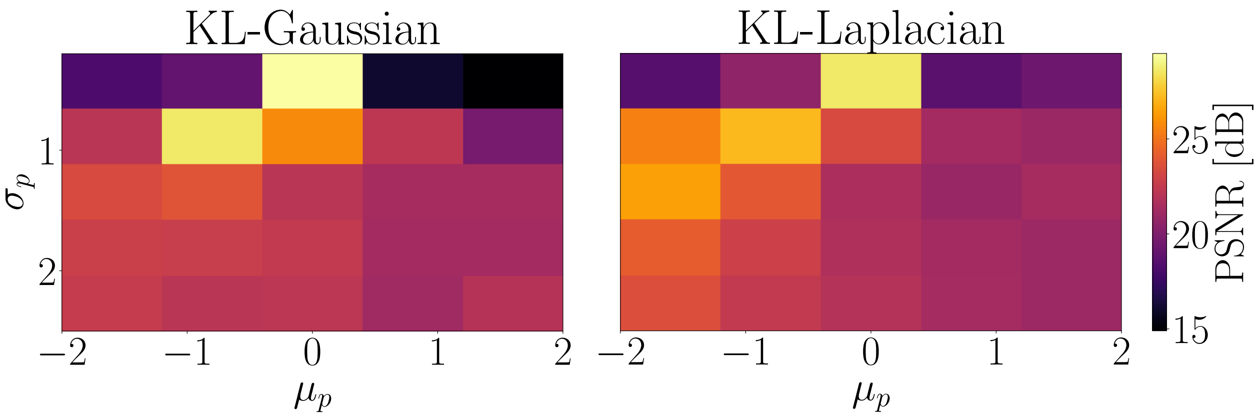


This is due that in CS, where the sensing matrix is not constrained to the physical conditions of a particular system, the E2E network becomes a commonly used autoencoder, and then applying sparsity regularization on the low-dimensional representation yields sparse autoencoders Ng et al. (2011) which is widely employed method to improve representation performance.

7.3. 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 Bacca et al. (2021), 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,

Figure 8. Recovery performance for the SPC system the KL regularizers with the Gaussian (left) and Laplacian (right) cases.

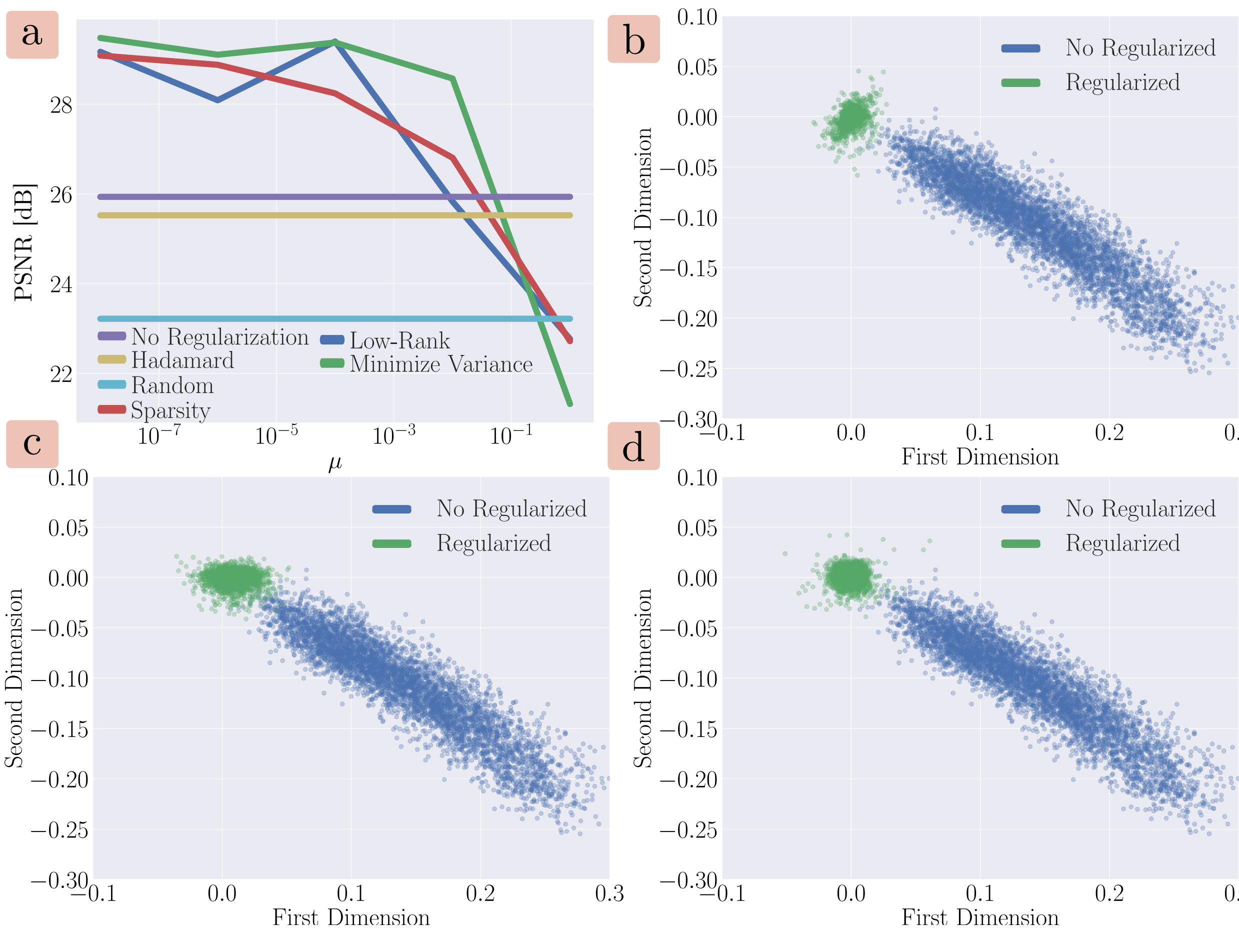


we employed the Fashion MNIST dataset. For all experiments, we used a compression ratio of 0.1.

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 Ronneberger et al. (2015) with five downsampling and five upsampling blocks. The results of this experiment are shown in Fig. 8. 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. 5, the best 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. The results of this

Figure 9. 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)



experiment are compared with the baseline E2E (non-regularized training), traditional CA aperture design based on Hadamard matrices Duarte et al. (2008), and random CA obtained with a Bernoulli distribution. The reconstruction performance measured in PSNR of this experiment is shown in Fig. 9(a). The results suggest that, in most cases, the proposed regularized outperforms the baseline E2E design, Hadarmard, and random settings. Later, the distribution of the first two SPC snapshots for all images in the test data set was

Figure 10. Visual reconstruction results for the SPC scenario using the variance and the structural regularizations, the baseline E2E designs, random coding, and coding based on Hadamard matrices. The blue values correspond to the best results and the green to the second best.

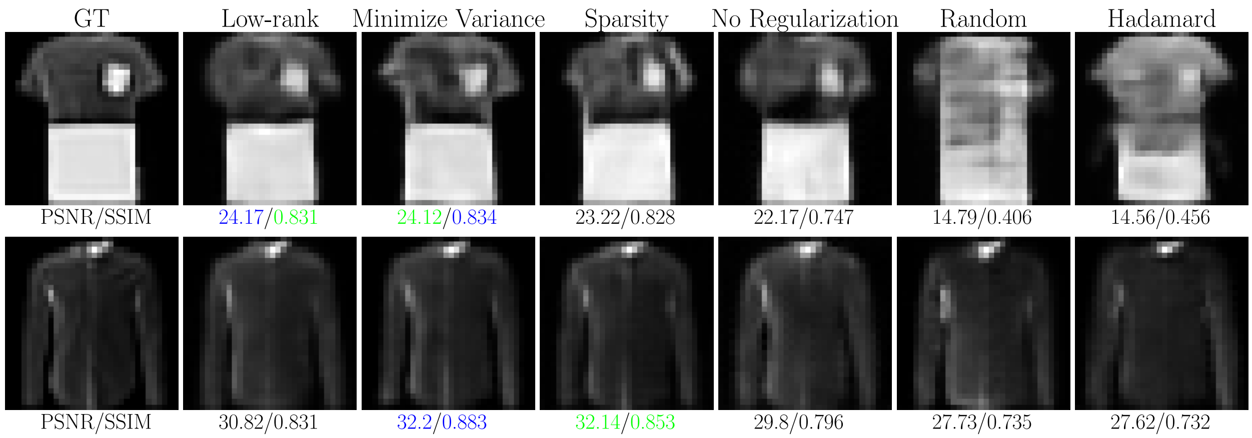
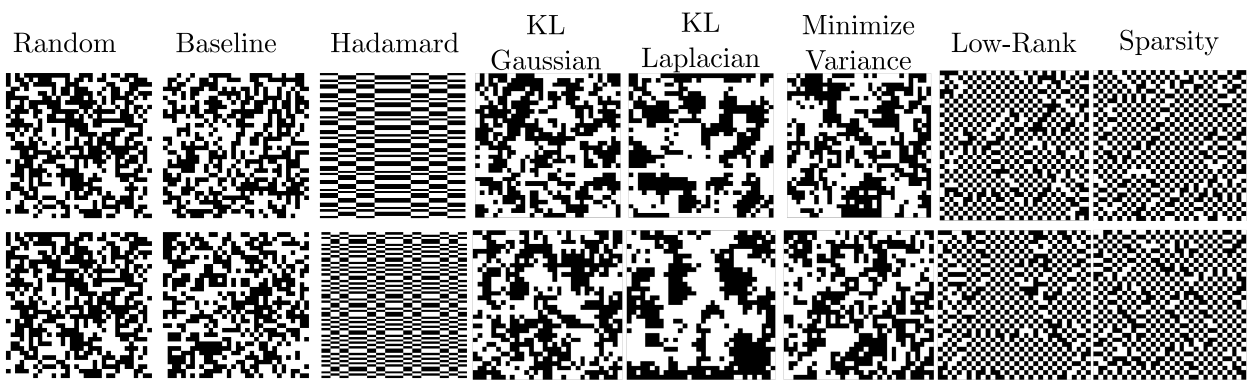
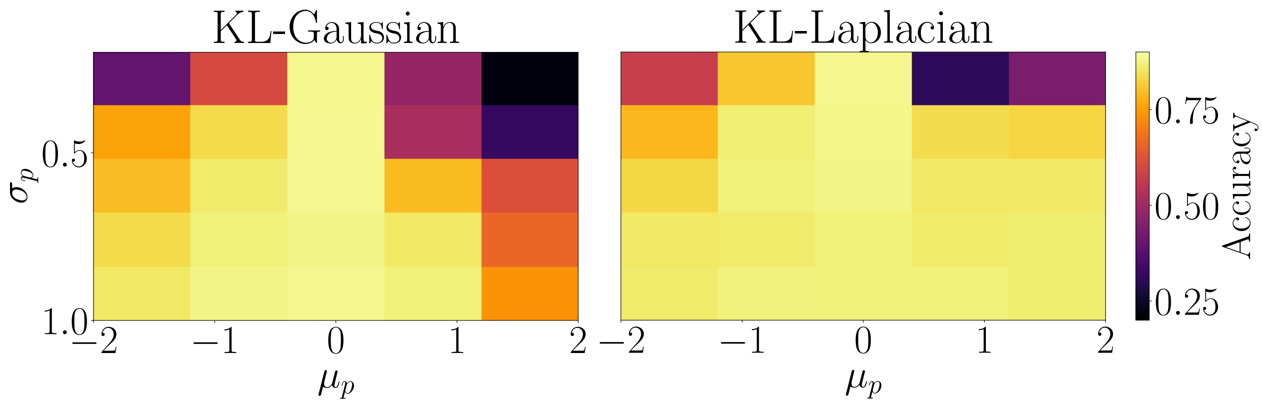


Figure 11. Optimized CA of two SPC snapshots employing the proposed regularization functions for the recovery tasks. Additionally, the E2E baseline, random, and Hadamard CAs are shown for comparison purposes.



plotted for the best-performing setting of each regularizer, variance minimization Fig. 9 (b), sparsity. 9 (c), and low-rank Fig. 9(d). Each scatter plot also shows the distribution obtained by the non-regularized E2E sensing matrix design. In all cases, the resultant distribution employing the regularizers is more concentrated than the non-regularized validating that for the reconstruction task, we perform better by reducing the variance of the measurements.

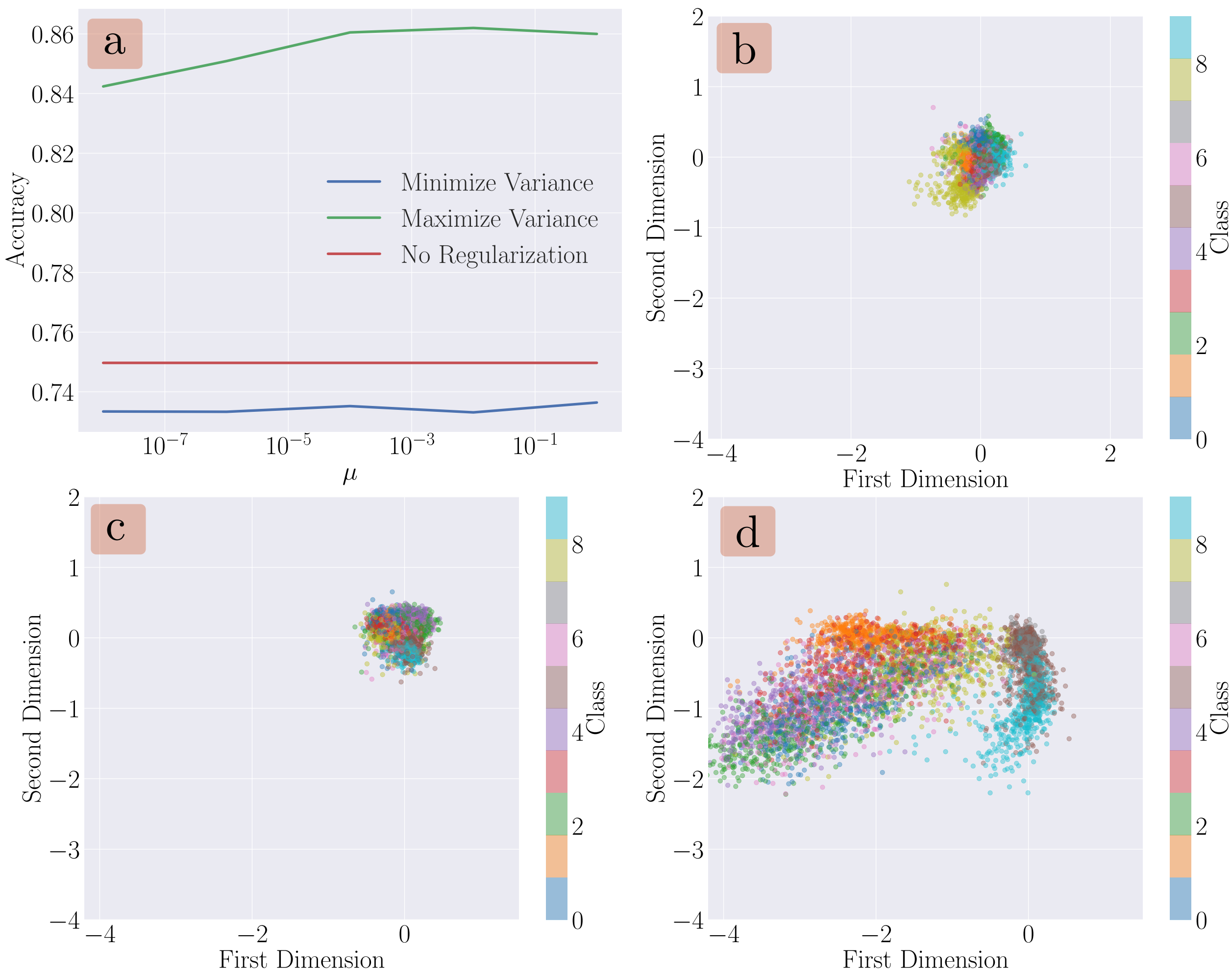
Figure 12. Classification performance for the SPC system the KL regularizers with the Gaussian (left) and Laplacian (right) cases.



Additionally, in Fig. 10 are shown some visual reconstruction of two images of the test set, where the reconstruction for these particular examples shows that the proposed design methodology improves upon the non-regularized E2E, the random coding and the Hadarmad-based CA. Here, the best-performing regularization is the variance minimization, mainly due to the physical constraint that the OCE has values {-1,1} yields that with a proper design, data variance can be effectively reduced. Finally, the resulting CA for employing the regularization functions is shown in 11 along with the CA of the comparison methods. Notably, the CA structure is highly affected by the regularization functions showing for instance that the structural regularizer converges to an almost uniform pattern, while the variance and KL-based regularizer tends to form clusters into the CA.

Classification experiments: Here, we evaluate the proposed regularization functions on the classification high-level task. The CD is a Mobilnet-V2 Sandler et al. (2018), which is a lightweight classification network. In this scenario, the same values were used in the experiment in Fig. 8 of μ_p and σ_p. The results are shown in Fig. 12, where an opposi-

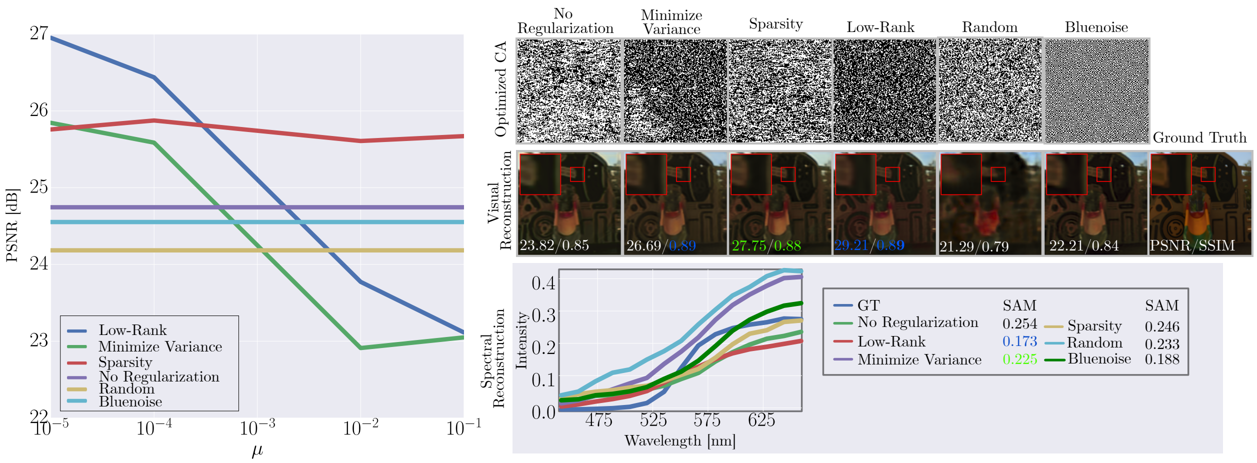
Figure 13. Classification accuracy performance for the SPC system with the minimum and maximum variance regularizers for different regularization parameters μ_p (a) measurements distribution comparison for the non-regularized. Measurements distribution of the non-regularized design (b), minimum variance (c), and maximum variance (d).



te performance is obtained compared to the recovery case. Higher variance provides 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 $\sigma_{max} = 5$ as we saw better

Figure 14. 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). Blue values correspond to the best results, and green to the second best.

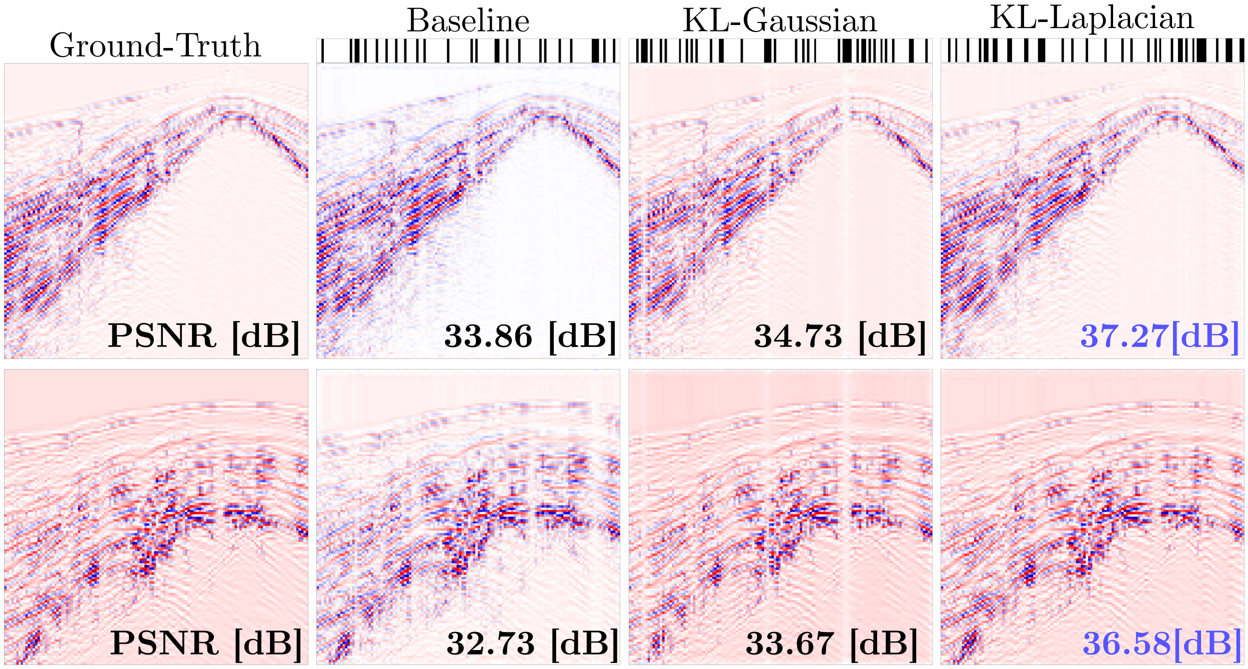


performance with this setting. The results of this experiment are compared with the baseline E2E (non-regularized training). The accuracy performance of this experiment is shown in Fig 13(a). These results show that the variance maximization regularizer outperforms the baseline regularization and the variance minimization. Additionally, the regularization R_{Vmin} 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 performant setting of each regularizer, the non-regularized design Fig. 13(b), variance minimization Fig. 13(c) and variance maximization Fig. 13(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.

7.4. CASSI experiments

Here, we aim to design the CA of the CASSI with the proposed regularization functions. The regularizers R_{Vmin}, R_S, and R_{LR} were used for this scenario since we want to recover the spectral image from the compressed measurements. Also, as a comparison, random CA and blue noise CA Correa et al. (2016) as non-data-driven designs. Then, we first evaluate the performance with respect to the regularization parameter μ_p compared with the non-regularization design. This parameters was varied from 10^{-8} to 10^{0} in a logarithmic scale. The results in Fig 14(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. 14(b) is shown the optimized CA for the non-regularized and regularized design. Remarkably, the low-rank design convergence to a uniform sampling pattern is a highly desired criterion in compressive imaging sensing matrix design Correa et al. (2016); Arguello and Arce (2014). Fig 14(c) shows a visual reconstruction of a test image with its corresponding PSNR and SSIM reconstruction values. Finally in Fig 14(d) the reconstruction of a red spectral signature is plotted with the corresponding SAM value. These last results show that the best results correspond also to the low-rank design. The low-rank regularizer performs better in this scenario since this prior is a very suitable prior for spectral images since this kind of data contains highly redundant information which can be effectively represented via low-rank approximation.

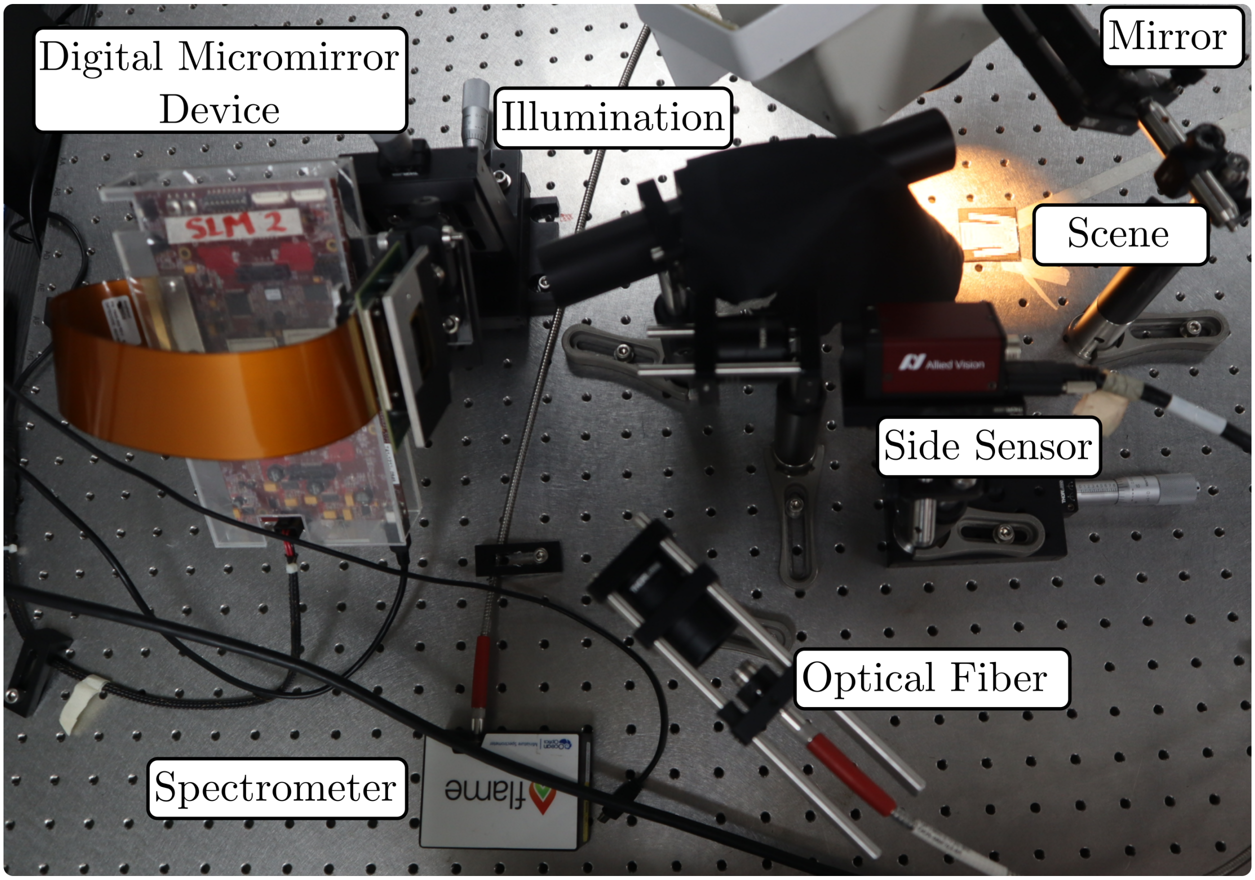
Figure 15. Visual results of the reconstructed seismic data for the non-regularized model and the models trained with the KL-Gaussian and KL-Laplacian regularizers.



7.5. Compressive seismic acquisition setting

For this experiment, the transmittance value was set to $\delta_0 = 0.6$. The CD network is a convolutional neural network with 5 convolutional layers with 128 filters each. Here we set for both regularizers $\mu_p = 0.5$ and $\sigma_p = 1.6$. Figure 15 shows the reconstruction of two seismic test data, where the best results are obtained by the KL-Laplacian regularization. Nevertheless, the KL-Guassian model outperforms the non-regularized model. Also, it is shown the subsampling vector for each model.

Figure 16. SPC acquisition system validation of proposed method



8. Experimental Validation

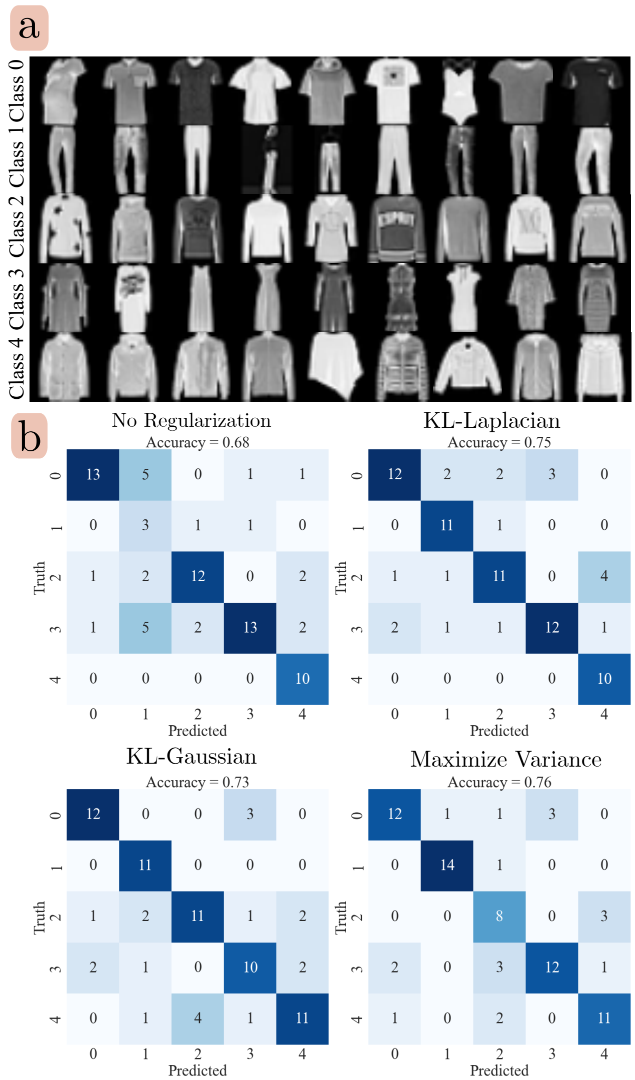
To perform the experimental validation of the proposed OCE design method, the SPC, and the CASSI systems were implemented. or the SPC, we focus only on the classification task since this system has proven to be very suitable for this task Bacca et al. (2020, 2022c). For the CASSI the computational task is reconstruction, therefore both tasks are also experimentally validated.

8.1. 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, a Flame Vis-Nir spectrometer was used, which has a spectral range from 350 to 1077 [nm], as shown in Fig. 16.

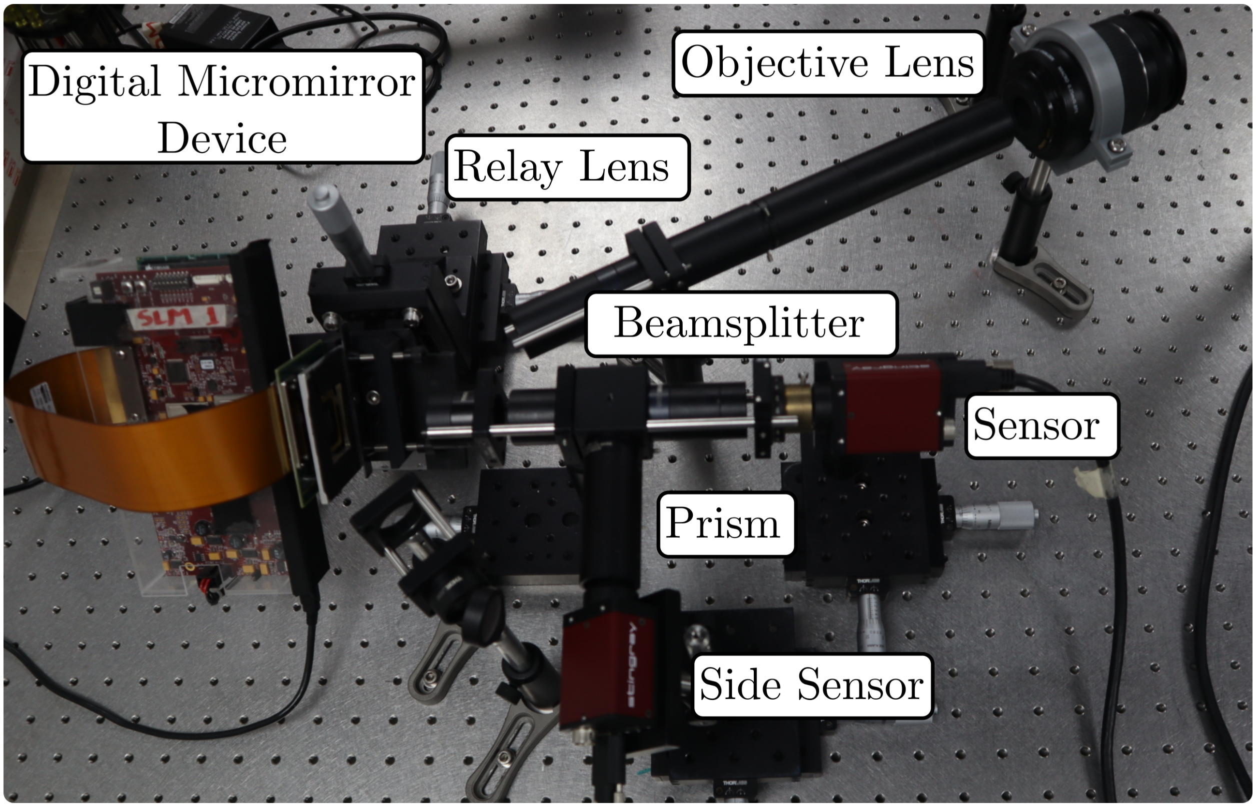
We employed this architecture to validate the performance of the proposed method. For this experiment, fifteen scenes of the first five classes of the Fashion MNIST dataset were acquired utilizing the implemented SPC system. A re-training of the network was performed with the calibrated and captured CA and using only the images from the first five classes of the Fashion MNIST dataset. From this, some of the examples acquired are shown in Fig. 17 a) were used as a test to evaluate the performance of the proposed method for every one of the regularizers. Fig. 17 b) shows the confusion matrix for the non-regularized design, the KL-Laplacian, the KL-Gaussian, and the maximize variance regularization. The results

Figure 17. 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.



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.

Figure 18. Experimental prototype of the CASSI acquisition System



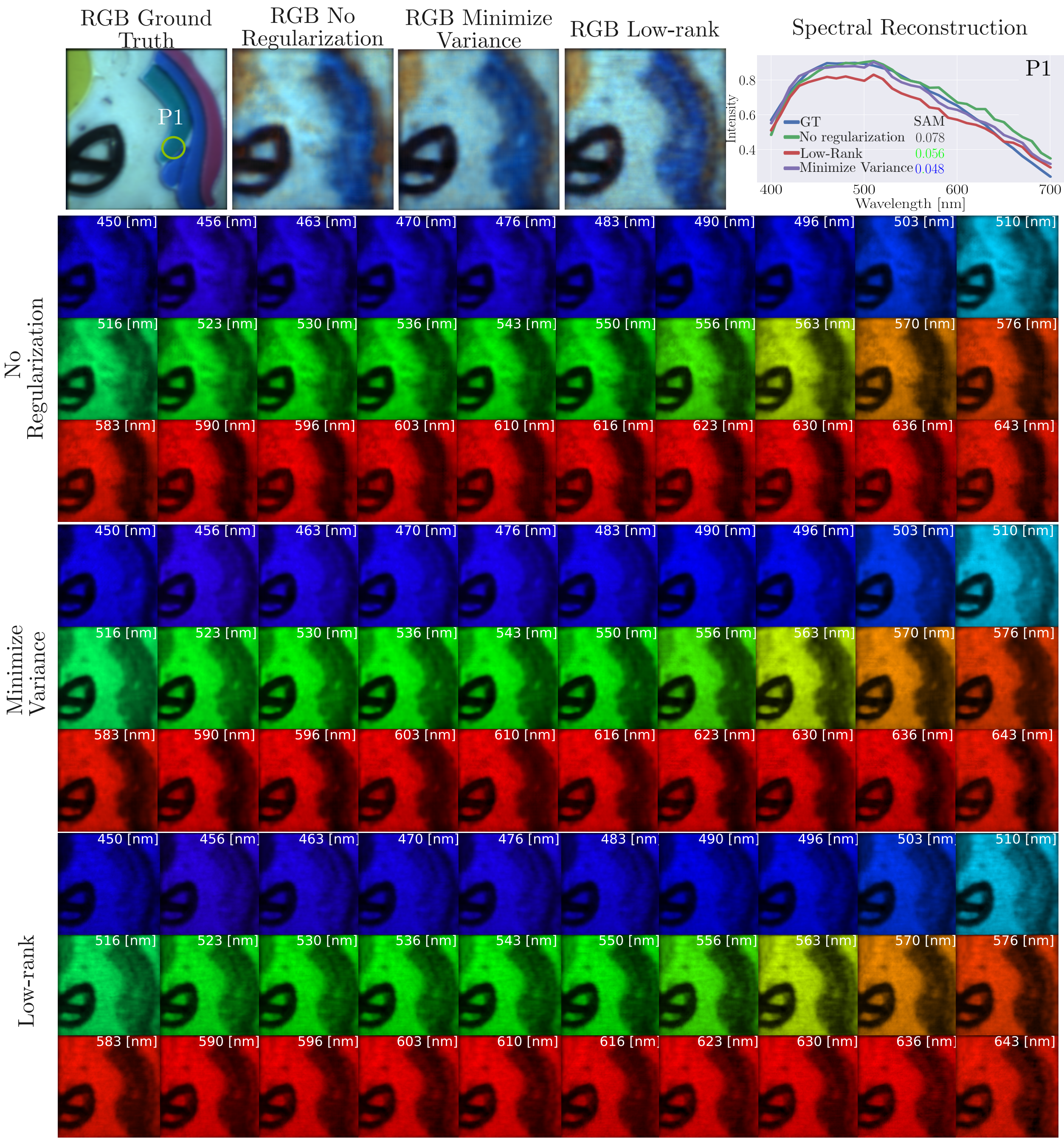
8.2. CASSI Implementation

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 19, 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. The sparsity regularization is not employed in this experiment because its best performance was lower than the low-rank design, therefore, we only employ the low-rank design to compare the results of the structural-based regularization with the variance-based regularization design. 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.

Figure 19. Reconstruction of real data with the CASSI system with the low-rank, minimize variance, and no-regularization designs. The 30 reconstructed spectral bands and a spectral signature reconstruction of a region of interest. The blue SAM value refers to the best performance and the green to the second best.



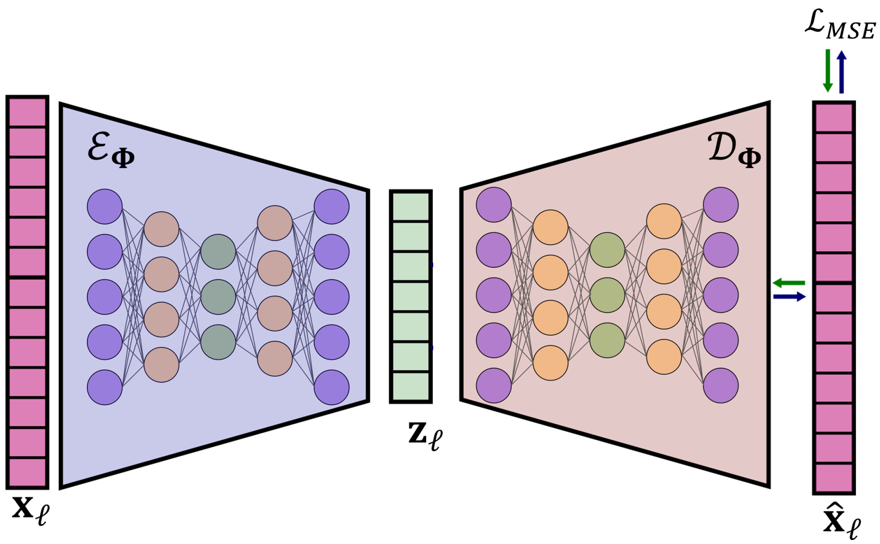
9. Extension to High-Level Tasks In Generative Adversarial Networks

The proposed regularization functions within this thesis have been employed to design physical encoders in computational imaging. However, this codification strategy is highly suitable in several computer-vision applications. Particularly, architectures based on the autoencoder Hinton and Salakhutdinov (2006), low-dimensional and meaningful representation of the high-dimensional data is obtained. This compressed representation from autoencoders has been employed in a wide range of applications including semi-supervised learning, clustering, and anomaly detection Bank et al. (2023). The core idea behind autoencoder is that by employing a deep neural network, called an encoder, denoted by $\mathcal{E}_\Phi$, where Φ is the trainable parameters, we can find a low-dimensional representation of the image dataset $\mathcal{X} = \{\mathbf{x}_\ell\}_{\ell=1}^{T}, \mathbf{x}_\ell \in \mathbb{R}^n$. The low-dimensional representation of the ℓ training sample is defined as $\mathbf{z}_\ell = \mathcal{E}_\Phi(\mathbf{x}_\ell), \mathbf{z}_\ell \in \mathbb{R}^m$ where $m \ll n$. Then, another network $\mathcal{D}_\theta$ that aims to recover the input image $\mathbf{x}_\ell$ from $\mathbf{z}_\ell$. The training of the autoencoder networks is performed as follows

$$\{\Phi, \theta\} = \underset{\Phi,\theta}{\arg\min} \sum_{\ell=1}^{T} \mathcal{L}_{MSE}(\mathcal{D}_\theta(\mathcal{E}_\Phi(\mathbf{x}_\ell)), \mathbf{x}_\ell), \tag{21}$$

where $\mathcal{L}_{MSE}(\cdot)$ denotes the mean squared error loss function. An illustration of this neural network architecture is shown in Fig. 20. Here, it is of wide interest to impose certain structures on the latent representation $\mathbf{z}_\ell$ for certain applications. Thus, we proposed to employ the variance regularization function for spectral image generation using low-dimensional

Figure 20. Autoencoder architecture



representation from an autoencoder.

9.1. Low-Dimensional Approach for Spectral Image Generation Motivation

Spectral images (SI) are a collection of images acquired at different wavelengths of an electromagnetic field, which conforms to a 3D data cube. This information allows for estimating the unique characteristics and distribution of the different materials within a scene. Hence, spectral image information is valuable in medical applications Li et al. (2016), remote sensing Govender et al. (2007), art conservation Fischer and Kakoulli (2006), among others. Spectral images can be classified into two groups depending on their spatial and spectral resolution; multispectral images, which have high spatial resolution and low spectral resolution, and hyperspectral images, which have a low spatial resolution but high spectral resolution.

The recent advances in data-driven DL methods have opened new frontiers for SI processing, acquisition, and its applications Ozdemir and Polat (2020). Some examples are in hyperspectral, and multispectral fusion Xie et al. (2019); Jacome et al. (2021); Yang et al. (2018); Jacome et al. (2022), classification Jácome et al. (2021); Bacca et al. (2022b), or recovery methods for snapshot compressive spectral imaging (CSI) recovery Monroy et al. (2021); Arguello et al. (2023). However, one of the main reasons for the great success of DL in a wide range of applications is that the models can extract the intrinsic structure of large datasets LeCun et al. (2015). Further, the SI datasets are limited in the number of available SIs due to the expensive and long acquisition times Hagen and Kudenov (2013). Thus the performance of the DL methods is still restricted to the available data.

To address this issue, data augmentation (DA) strategies are employed Shorten and Khoshgoftaar (2019), where geometrical transformations (flipping, rotation, contrast modification, etc.) are applied to the SIs to generate new training samples. Recent approaches employ generative adversarial networks (GANs) Goodfellow et al. (2020) to synthesize new samples as DA based on learning the probability distribution of the SI dataset and generating new SI samples from the learned distribution.

Despite the great success of GANs in RGB image generation, this type of network suffers when the data distribution is of high dimensionality as those of SI. Thus, we train an autoencoder network to obtain a low-dimensional representation of the SI dataset. Then, a GAN will be trained adversarially using the LD image dataset to generate new LD image samples. Finally, the generated LD image samples are decoded through the AE network to obtain generated SIs.

9.2. Low Dimensional Generative adversarial networks

Consider a SI dataset $\{\mathbf{x}_\ell\}_{\ell=1}^{T}$, with T samples, and $\mathbf{x}_\ell \in \mathbb{R}^{MNL}$, where M, N are the spatial dimensions and L the number of spectral bands. Denoting the dataset distribution as $p_{\mathrm{x}}(\mathbf{x})$. Then, a generative network $\mathcal{G}$ will be optimized to obtain a distribution p_g from data $\mathbf{x}_\ell$ which will probably achieve that $p_g \approx p_{\mathrm{x}}$. A prior distribution is assumed on input noise variables denoted as $p_{\mathbf{n}}(\mathbf{n})$ where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ and m is the random variable dimension, usually a Gaussian distribution, which are mapped to the desired generated images. Then, a discriminative network $\mathcal{T}$ is also defined. This network will receive the generated samples or dataset samples. The training of a GAN Goodfellow et al. (2020), called adversarial training, can be represented as

$$\{\hat{\mathcal{G}}, \hat{\mathcal{T}}\} = \arg\min_{\mathcal{T}} \arg\max_{\mathcal{G}} \mathbb{E}_{\mathbf{x}\sim p_{\mathrm{data}}(\mathbf{x})} \left[\log\left(\mathcal{W}(\mathbf{x})\right)\right] + \mathbb{E}_{\mathbf{n}\sim p_{\mathbf{n}}(\mathbf{n})} \left[\log\left(1 - \mathcal{W}\left(\mathcal{G}(\mathbf{n})\right)\right)\right]. \tag{22}$$

In the adversarial training, each network $\{\mathcal{G}, \mathcal{T}\}$ compete to achieve their goals: $\mathcal{G}$ will generate fake samples, and $\mathcal{T}$ will predict if the received samples are real or fake. However, achieving proper convergence of GAN through adversarial training will be difficult, while more high dimensionality has the generated images and with limited training samples as in SI. Thus, our proposed approach is the following First, we trained an SI autoencoder following the optimization problem in (21). Then, we use a GAN that will be optimized with respect to the LD representation from the SI autoencoder. Then, the adversarial optimization problem from 22 is re-formulated as

$$\{\hat{\mathcal{G}}, \hat{\mathcal{T}}\} = \arg\max_{\mathcal{G}} \arg\min_{\mathcal{T}} \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})} \left[\log\left(\mathcal{T}(\mathcal{E}_\Phi(\mathbf{x}))\right)\right] + \mathbb{E}_{\mathbf{n}\sim p_{\boldsymbol{n}}(\mathbf{n})} \left[\log\left(1 - \mathcal{T}\left(\mathcal{G}(\mathbf{n})\right)\right)\right]. \quad (23)$$

Thus, in this case, is of high importance what properties should both, the low-dimensional features of the original SI dataset $\mathbf{z}_\ell = \mathcal{E}_\theta(\mathbf{x}_\ell)$ and the generated synthetic low-dimensional samples $\mathbf{e} = \mathcal{G}(\mathbf{n})$. Thus we applied the variance regularization for both the autoencoder and the GAN.

9.3. Statistical Regularization for the AE and GAN

Towards improving the LD representations of the AE and the diversity of generated LD images by the GAN, we propose a variance minimization regularizer in the AE training that allows a compact representation of the SI dataset in the LD space, which improves AE recovery performance. Then, we employed a variance maximization on the generated LD space for the GAN training to produce diverse data and more quality on the generated SI dataset. First, define the set $\mathbf{Z} = [\mathbf{z}_1^T, \ldots, \mathbf{z}_T^T] \in \mathbb{R}^{m\times T}$ that contains the LD representation of the AE and define the set of the GAN generated LD images $\mathbf{B} = [\mathbf{e}_1^T, \ldots, \mathbf{e}_T^T] \in \mathbb{R}^{m\times T}$. We compute the variance of the $\mathbf{A}$ and $\mathbf{B}$ as mentioned in section 5.2. Thus, for the AE training, we have

$$\{\Phi, \theta\} = \arg\min_{\Phi,\theta} \sum_{\ell=1}^{T} \mathcal{L}_{MSE}(\mathcal{D}_\theta(\mathcal{E}_\Phi(\mathbf{x}_\ell)), \mathbf{x}_\ell) + \mu_{ae} R_{Vmin}(\mathbf{Z}), \quad (24)$$

where R_{Vmin} is defined as in (8). The regularized GAN optimization problem is given by

$$\{\hat{\mathcal{G}}, \hat{\mathcal{T}}\} = \arg\max_{\mathcal{G}} \ \arg\min_{\mathcal{T}} \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})} \left[\log\left(\mathcal{T}(\mathcal{E}_{\Phi}(\mathbf{x}))\right)\right] + \mathbb{E}_{\mathbf{n}\sim p_{\boldsymbol{n}}(\mathbf{n})} \left[\log\left(1 - \mathcal{T}\left(\mathcal{G}(\mathbf{n})\right)\right)\right] - \mu_{gan} R_{Vmin}(\mathbf{B}), \tag{25}$$

Note that the negative sign in the variance regularization function is because we want to maximize the variance of the generated LD representation of the SI. In both optimization problems, the AE and the GAN, the parameters μ_{ae} and μ_{gan} are regularization hyperparameters that control how much we concentrate the AE LD space or how much we increase the variability of the generated images by the GAN.

9.4. Spectral Image Applications

To validate the performance of our proposed method, we address the following DL-based SI tasks:

CSI Recovery: Here, we employ the CASSI system described in Section 6.2. Here, the CA was set randomly distributed as we only wanted to analyze the network performance.

Single Image Spatial-Spectral Image Super-Resolution: A well-known task for SI is to recover a high spatio-spectral resolution SI from a low spatio-spectral resolution SI Yan et al. (2018). In this task, the SI can be downsampled spatially and spectrally by a decimation matrix $\mathbf{D} \in \mathbb{R}^{\frac{MN}{k_s}\left(\frac{L}{k_l}\right)\times MNL}$, where k_s and k_l represent the decimation factor for the spatial and spectral resolution of the SI, respectively. The low spatio-spectral resolution image is represented as $\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{n}$.

SI Recovery from RGB Images: Another task that has gathered significant attention from the research community is the mapping from RGB image to SI Arad et al. (2022). This task consists of recovering an SI from an RGB image with a count with less spectral information considering a known spectral response function $\mathbf{R} \in \mathbb{R}^{MN3 \times MNL}$. Then, the RGB image can be represented as $\mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{n}$.

Since the mentioned tasks are ill-posed problems, the objective is to recover the SI $\mathbf{x}$ from the observed data $\mathbf{y} = \mathbf{T}\mathbf{x}$, where $\mathbf{T}$ could represent the sensing matrix $\mathbf{H}$, the decimation matrix $\mathbf{D}$ or the spectral response function $\boldsymbol{R}$, according to the selected problem. Then, we can solve the DL-based SI computational tasks through the optimization problem

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{T} \sum_{\ell=1}^{T} \|\mathbf{x}_\ell - \mathcal{M}_{\boldsymbol{\beta}}(\mathbf{y}_\ell)\|_2^2, \tag{26}$$

These are ill-posed inverse problems that are challenging. Therefore, we aim to increase the performance of the network $\mathcal{M}_\beta$ for each case by adding synthetic samples generated by the LD-GAN.

9.5. Numerical Experiments

In this section, we perform several experiments with the AE network with different channels for the LD representation. For the GAN architecture, the deep convolutional generative adversarial network Radford et al. (2015) was adapted to the spatial size of the employed dataset, and the experiments are performed with respect to the LD image dataset obtained from the AE network against the entire-sized SI dataset. All experiments were performed on GPU with an NVIDIA RTX 3090 graphic card. The dataset employed for all the experiments is the ARAD 1K Arad et al. (2022), which was preprocessed, reducing the spatial resolution to 256×256 and keeping the 31 spectral bands. This dataset contains 900 samples for training and 50 samples for testing. We extract unique patches with a spatial resolution of 128×128 obtaining a total of 3600 patches for training and 200 patches for testing. Recovery performance is measured with the peak-noise-to-signal ratio (PSNR) and the structural similarity index measure (SSIM) Wang et al. (2004).

9.5.1. Statistical Regularization Experiments. To validate the effectiveness of the proposed regularized training in the AE and GAN, we perform a hyperparameter study of the regularization parameters μ_{gan} and μ_{ae}. The parameters $\mu_{ae} = \mu_{gan} = \{0, 1e^{-5}, 1e^{-3}\}$ were changed for each task. The augmented SI dataset was 100 % of the original dataset. Fig. 21 shows the performance of the mentioned experiment. The highest performance in each task is obtained at higher regularization parameters showing the effectiveness of this training method.

To visualize the effect of maximizing the variance of the generated LD representation in the GAN training, in Fig. 22, the three first principal components of 3000 synthesized SIs were computed for the non-regularized GAN and the proposed LD-GAN with $\mu_{ae} = \mu_{gan} = 0.001$ showing that the last one has more variability than the first one.

Figure 21. SI recovery performance in terms of PSNR and SSIM employing the proposed LD-GAN with different values of the regularization parameters μ_{gan} and μ_{ae} for each SI application.

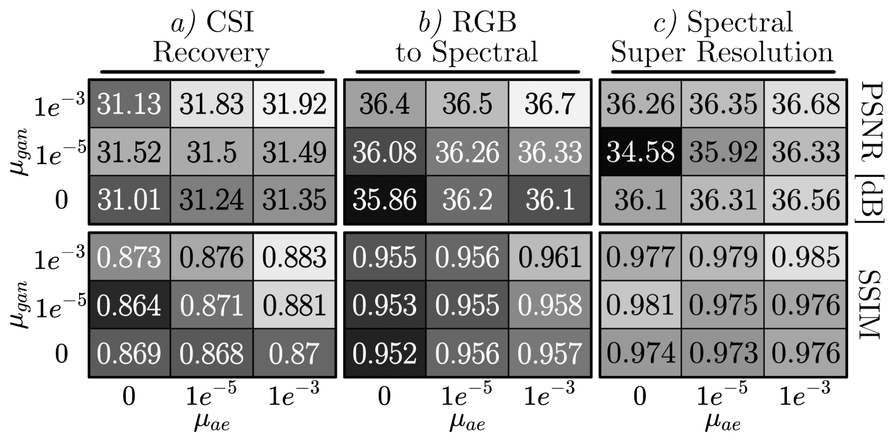
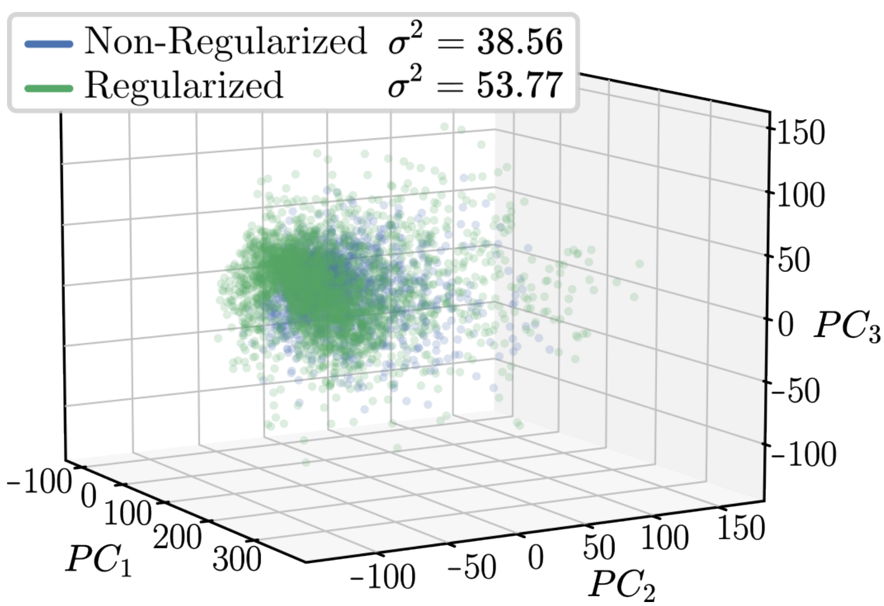


Figure 22. Three principal components of the generated dataset and its respective variance with the regularized training and the non-regularized. PC_1, PC_2, and PC_3 refer to the first, second, and third principal components of the dataset respectively.



10. 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., they 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 best classification performance, a wider distribution is desired. We validate the design of optical coding elements through regularized E2E optimization in different optical architectures, showing improvement with respect to the non-regularized design and other traditional non-data-driven approaches such as blue noise coding and Hadamard sensing. We present extensive simulation results for both computational tasks, whose performance was also validated by real scenarios with data acquired with the physical implementation of the designed systems. While here we analyzed three types of regularization individually, it remains an open question, and future work, on how to combine these functions to promote more complex priors and structures on the set of measurements by designing the OCE. Several computational imaging systems can harness the proposed coding design method such as spectral-depth imaging Baek et al. (2021), light-field Vargas et al. (2021), among others.

Additionally, beyond the sensing matrix design, these regularizations can also be used in high-level tasks such as generative models Martinez et al. (2023) where the variance of the generated samples is maximized to have high-diversity synthetic samples.

References

Arad, B., Timofte, R., Yahel, R., Morag, N., Bernat, A., Cai, Y., Lin, J., Lin, Z., Wang, H., Zhang, Y., et al. (2022). Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 863–881.

Arce, G. R., Brady, D. J., Carin, L., Arguello, H., and Kittle, D. S. (2014). Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115.

Arguello, H. and Arce, G. R. (2014). Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Transactions on Image Processing*, 23(4):1896–1908.

Arguello, H., Bacca, J., Kariyawasam, H., Vargas, E., Marquez, M., Hettiarachchi, R., Garcia, H., Herath, K., Haputhanthri, U., Ahluwalia, B. S., et al. (2023). Deep optical coding design in computational imaging: a data-driven framework. *IEEE Signal Processing Magazine*, 40(2):75–88.

Bacca, J., Galvis, L., and Arguello, H. (2020). Coupled deep learning coded aperture design for compressive image classification. *Optics express*, 28(6):8528–8540.

Bacca, J., Gelvez-Barrera, T., and Arguello, H. (2021). Deep coded aperture design: An end-to-end approach for computational imaging tasks. *IEEE Transactions on Computational Imaging*, 7:1148–1160.

Bacca, J., Gelvez-Barrera, T., and Arguello, H. (2022a). Invariant coded aperture design for compressive imaging. In *Adaptive Optics and Applications*, pages JTh2A–9. Optica Publishing Group.

Bacca, J., Hernandez-Rojas, A., and Arguello, H. (2022b). Deep coding patterns design for compressive near-infrared spectral classification. In *2022 30th European Signal Processing Conference (EUSIPCO)*, pages 548–552.

Bacca, J., Marquez, M., and Arguello, H. (2022c). Single pixel near-infrared imaging for spectral classification. In *Computational Optical Sensing and Imaging*, pages CW1B–2. Optica Publishing Group.

Baek, S.-H., Ikoma, H., Jeon, D. S., Li, Y., Heidrich, W., Wetzstein, G., and Kim, M. H. (2021). Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2651–2660.

Bank, D., Koenigstein, N., and Giryes, R. (2023). Autoencoders. *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, pages 353–374.

Bardes, A., Ponce, J., and Lecun, Y. (2022). Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In *ICLR 2022-10th International Conference on Learning Representations*.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828.

Candes, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30.

Caroli, E., Stephen, J., Di Cocco, G., Natalucci, L., and Spizzichino, A. (1987). Coded aperture imaging in x-and gamma-ray astronomy. *Space Science Reviews*, 45(3):349–403.

Chang, J. and Wetzstein, G. (2019a). Deep optics for monocular depth estimation and 3d object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10193–10202.

Chang, J. and Wetzstein, G. (2019b). Deep optics for monocular depth estimation and 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10193–10202.

Correa, C. V., Arguello, H., and Arce, G. R. (2016). Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging. *JOSA A*, 33(12):2312–2322.

Duarte, M. F., Davenport, M. A., Takhar, D., Laska, J. N., Sun, T., Kelly, K. F., and Baraniuk, R. G. (2008). Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91.

Evtimova, K. and LeCun, Y. (2021). Sparse coding with multi-layer decoders using variance regularization. *arXiv preprint arXiv:2112.09214*.

Fazel, M., Candes, E., Recht, B., and Parrilo, P. (2008). Compressed sensing and robust recovery of low rank matrices. In *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 1043–1047. IEEE.

Fischer, C. and Kakoulli, I. (2006). Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications. *Studies in Conservation*, 51(sup1):3–16.

Fu, C., Arguello, H., Sadler, B. M., and Arce, G. R. (2015). Compressive spectral polarization imaging with coded micropolarizer array. In *Compressive Sensing IV*, volume 9484, pages 59–65. SPIE.

Galvis, L., Arguello, H., and Arce, G. R. (2015). Coded aperture design in mismatched compressive spectral imaging. *Appl. Opt.*, 54(33):9875–9882.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

Gottesman, S. R. and Fenimore, E. E. (1989). New family of binary arrays for coded aperture imaging. *Applied optics*, 28(20):4344–4352.

Govender, M., Chetty, K., and Bulcock, H. (2007). A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water Sa*, 33(2).

Hagen, N. A. and Kudenov, M. W. (2013). Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):1 – 23.

Heide, F., Fu, Q., Peng, Y., and Heidrich, W. (2016). Encoded diffractive optics for full-spectrum computational imaging. *Scientific reports*, 6(1):1–10.

Hernandez-Rojas, A. and Arguello, H. (2022). 3d geometry design via end-to-end optimization for land seismic acquisition. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 4053–4057. IEEE.

Higham, C. F., Murray-Smith, R., Padgett, M. J., and Edgar, M. P. (2018). Deep learning for real-time single-pixel video. *Scientific reports*, 8(1):2369.

Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507.

Hirsch, M., Wetzstein, G., and Raskar, R. (2014). A compressive light field projection system. *ACM Transactions on Graphics (TOG)*, 33(4):1–12.

Horé, A. and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369.

Hung, W.-C., Jampani, V., Liu, S., Molchanov, P., Yang, M.-H., and Kautz, J. (2019).

Scops: Self-supervised co-part segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 869–878.

Jacome, R., Bacca, J., and Arguello, H. (2021). Deep-fusion: An end-to-end approach for compressive spectral image fusion. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 2903–2907. IEEE.

Jacome, R., Bacca, J., and Arguello, H. (2022). D 2 uf: Deep coded aperture design and unrolling algorithm for compressive spectral image fusion. *IEEE Journal of Selected Topics in Signal Processing*.

Jácome, R., López, C., Garcia, H., and Arguello, H. (2021). Deep learning-based object classification for spectral images. In Orjuela-Cañón, A. D., Lopez, J., Arias-Londoño, J. D., and Figueroa-García, J. C., editors, *Applications of Computational Intelligence*, pages 147–159, Cham. Springer International Publishing.

Jeon, D. S., Baek, S.-H., Yi, S., Fu, Q., Dun, X., Heidrich, W., and Kim, M. H. (2019). Compact snapshot hyperspectral imaging with diffracted rotation. *Association for Computing Machinery (ACM)*.

Jerez, A., Garcia, H., and Arguello, H. (2018). Single pixel spectral image fusion with side information from a grayscale sensor. In *2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (ColCACI)*, pages 1–6.

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Lathauwer, L. D., Moor, B. D., and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21:1253–1278.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

León-López, K. M. and Fuentes, H. A. (2020). Online tensor sparsifying transform based on temporal superpixels from compressive spectral video measurements. *IEEE Transactions on Image Processing*, 29:5953–5963.

Li, C., Balla-Arabé, S., and Yang, F. (2016). Embedded multi-spectral image processing for real-time medical application. *Journal of Systems Architecture*, 64:26 – 36. Real-Time Signal Processing in Embedded Systems.

Liner, C. L. (2016). *Elements of 3D Seismology.* Society of Exploration Geophysicists.

Mait, J. N., Euliss, G. W., and Athale, R. A. (2018). Computational imaging. *Advances in Optics and Photonics*, 10(2):409–483.

Martinez, E., Jacome, R., Hernandez-Rojas, A., and Arguello, H. (2023). Ld-gan: Low-dimensional generative adversarial network for spectral image generation with variance regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 265–275.

Metzler, C. A., Ikoma, H., Peng, Y., and Wetzstein, G. (2020). Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1375–1385.

Meyer, G. P. (2021). An alternative probabilistic interpretation of the huber loss. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 5261–5269.

Monroy, B., Bacca, J., and Arguello, H. (2021). Deep low-dimensional spectral image representation for compressive spectral reconstruction. In *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE.

Mosher, C. C., Li, C., Janiszewski, F. D., Williams, L. S., Carey, T. C., and Ji, Y. (2017). Operational deployment of compressive sensing systems for seismic data acquisition. *The Leading Edge*, 36(8):661–669.

Ng, A. et al. (2011). Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19.

Nguyen, T., Le, T., Vu, H., and Phung, D. (2017). Dual discriminator generative adversarial nets. *Advances in neural information processing systems*, 30.

Osorio Quero, C. A., Durini, D., Rangel-Magdaleno, J., and Martinez-Carranza, J. (2021). Single-pixel imaging: An overview of different methods to be used for 3d space reconstruction in harsh environments. *Review of Scientific Instruments*, 92(11):111501.

Ozdemir, A. and Polat, K. (2020). Deep learning applications for hyperspectral imaging: a systematic review. *Journal of the Institute of Electronics and Computer*, 2(1):39–56.

Peng, Y., Fu, Q., Amata, H., Su, S., Heide, F., and Heidrich, W. (2015). Computational imaging using lightweight diffractive-refractive optics. *Optics express*, 23(24):31393–31407.

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Regone, C., Stefani, J., Wang, P., Gerea, C., Gonzalez, G., and Oristaglio, M. (2017). Geologic model building in seam phase ii — land seismic challenges. *The Leading Edge*, 36:738–749.

Rifai, S., Mesnil, G., Vincent, P., Muller, X., Bengio, Y., Dauphin, Y., and Glorot, X. (2011). Higher order contractive auto-encoder. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 645–660. Springer.

Ronneberger, O., Fischer, P., and Brox, T. (2015). 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*, pages 234–241. Springer.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.

Selesnick, I. W. and Figueiredo, M. A. (2009). Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors. In *Wavelets XIII*, volume 7446, pages 107–121. spie.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48.

Sitzmann, V., Diamond, S., Peng, Y., Dun, X., Boyd, S., Heidrich, W., Heide, F., and Wetzstein, G. (2018). End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)*, 37(4):1–13.

Vargas, E., Martel, J. N., Wetzstein, G., and Arguello, H. (2021). 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*, pages 2692–2702.

Wagadarikar, A., John, R., Willett, R., and Brady, D. (2008). Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 47(10):B44–B51.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

Xiao, H., Rasul, K., and Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

Xie, Q., Zhou, M., Zhao, Q., Meng, D., Zuo, W., and Xu, Z. (2019). Multispectral and hyperspectral image fusion by ms/hs fusion net. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1585–1594.

Yan, Y., Zhang, L., Li, J., Wei, W., and Zhang, Y. (2018). Accurate spectral super-resolution from single rgb image using multi-scale cnn. In *Pattern Recognition and Computer Vision: First Chinese Conference, PRCV 2018, Guangzhou, China, November 23-26, 2018, Proceedings, Part II 1*, pages 206–217. Springer.

Yang, J., Zhao, Y., and Chan, J. (2018). Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. *Remote Sensing*, 10:800.

Yilmaz, O. (2008). *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*, volume 1. Society of Exploration Geophysicists.