ReDIP: Rethinking Deep Image Prior for Compressive Spectral Imaging Calibration

Emmanuel Martinez[1], Jorge Bacca[1], Tatiana Gelvez-Barrera[2], and Henry Arguello[1]
[1]Department of Computer Science, Universidad Industrial de Santander, Bucaramanga, Colombia
[2] Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, LVA, Villeurbanne, France

***Abstract*—Learning-based reconstruction methods have shown state-of-the-art (SOTA) performance for compressive spectral imaging (CSI). However, the training setup dependence causes slight variations in the scene's statistical distribution or calibration errors in the sensing matrix producing poor testing reconstruction quality. Therefore, we propose a computational calibration methodology to improve any CSI reconstruction deep neural network (DNN) testing quality and robustness. Specifically, similar to transfer learning, we rethink the deep image prior framework to retrain a SOTA DNN for a particular CSI measurement without retraining the DNN from scratch. Experimental results show an average peak signal-to-noise ratio improvement of 5.5dB when slight variations in the CSI measurement and calibration errors in the sensing matrix are considered. Additionally, even when no variations are considered, the proposed calibration methodology improves in up to 2.3dB the reconstruction quality.**

***Index Terms*—Compressive Spectral Imaging, Deep Neural Networks, Deep Image Prior, Spectral Imaging Calibration.**

I. Introduction

Snapshot compressive spectral imaging (CSI) acquires the spatial-spectral information of a spectral image (SI) in a set of 2D projected measurements through CSI optical systems [1], [2]. The SI is reconstructed from the 2D projected measurements by computational algorithms, covering optimization-based [3], [4] and learning-based [5], [6] approaches. In particular, deep-learning (DL) based methods employ SI datasets to train a deep neural network (DNN), whose weights are optimized concerning a subset of the entire SI dataset, known as the training dataset, achieving state-of-the-art (SOTA) CSI reconstruction quality [7], [8].

The training of DNNs commonly requires high-performance computing equipment, takes a long time to optimize, and, in the case of CSI reconstruction, the performance highly depends on the training dataset statistical distribution and accurate modeling of the CSI optical system. Expressly, the sensing matrix modeling the physical implementation is assumed to be ideal, exactly known and accurately simulated for the training and inference steps [5], [6]. Consequently, when a CSI system is implemented in the laboratory to acquire real 2D projected measurements, the performance of DL-based methods drops due to the acquisition system degradation given by the environmental conditions that are not considered during the training process. Furthermore, even if the system optical elements uncertainty is modeled, the acquisition of the real 2D projected measurements can not be accurately modeled due to the high real-world complexity [9].

In contrast, the deep image prior framework (DIP) [10] reconstructs a SI by considering the DNN as a prior, which is optimized using only the acquired 2D projected measurements, without relying on a training dataset [11]. The DIP is usually divided into three sequential components. (i) The input component corresponds to the 2D projected measurements or random noise. (ii) The central component is formed by the network that produces the reconstructed SI. (iii) The output component calculates the estimated 2D projected measurements from the reconstructed SI through the modeled sensing matrix. Despite the advantage of DIP to reconstruct a SI from the 2D projected measurements without relying on a training dataset, the optimization process is commonly computationally expensive and unstable when adjusted from scratch [12].

Therefore, this paper proposes a computational calibration process where pre-trained CSI reconstruction DNNs can be adjusted with only the 2D projected measurements and the sensing matrix of an implemented CSI optical system setup. The proposed methodology adjusts the SOTA DNN weights for each image input by minimizing a proposed loss function that only considers the real 2D projected measurements and the real sensing matrix, where the reconstructed SI is obtained at the central component of the framework. The methodology is validated over two DNNs using Bayesian frameworks [5], [6], whose training model and pre-trained weights are publicly available. The proposed method improves the CSI reconstruction in more than 5dBs when degradation errors in the sensing matrix are considered. Additionally, even when ideal 2D projected measurements and ideal sensing matrix are employed, the proposed method achieves a gain of more than 2 dB in the SI reconstruction.

II. Learning-based Compressive Spectral Imaging Reconstruction

CSI optical systems are specialized cameras that contain coding and shared optical elements for compressing a SI. The most popular CSI optical system is the coded aperture snapshot spectral imager (CASSI) [2], whose optical path consists of a coded aperture, followed by a dispersive prism and a 2D sensor that receives the shifted modulated light

This work was supported by the Sistema general de Regalias-Colombia under Grant BPIN 2020000100415, with UIS code 8933 and VIE-UIS under project "Institucionalización de los semilleros de investigacion - 3834".
T. Gelvez-Barrera is supported by Labex Celya.

forming the 2D projected measurements. The sensing process can be represented in a linear form as

$$\mathbf{y} = \mathbf{H}\mathbf{f} + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{f} \in \mathbb{R}^n$ is a vector representation of the SI, $\mathbf{H} \in \mathbb{R}^{m \times n}$, $m \ll n$ is the sensing matrix, $\mathbf{y} \in \mathbb{R}^m$ denotes the 2D projected measurements, and $\boldsymbol{\eta} \in \mathbb{R}^n$ denotes the optical system noise, generally modeled with a statistical distribution, such as $\mathcal{N}(\mu, \sigma^2)$ [13].

A. Deep Learning-based Reconstruction Methods

Consider a DNN $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$ with $\boldsymbol{\theta}$ adjustable weights, a dataset $\{\mathbf{f}^{(i)}\}_{i=1}^{M}$ with M SIs, and a simulated sensing matrix $\mathbf{H}$. The 2D projected measurements for each SI can be simulated as $\mathbf{y}^{(i)} = \mathbf{H}\mathbf{f}^{(i)}$, building a paired dataset $\{\mathbf{y}^{(i)}, \mathbf{f}^{(i)}\}_{i=1}^{M}$. The CSI reconstruction DNN can then be trained by solving the optimization problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{M} ||\mathbf{f}^{(i)} - \mathcal{M}_{\boldsymbol{\theta}}(\mathbf{y}^{(i)})||_2^2, \tag{2}$$

where $\hat{\boldsymbol{\theta}}$ are the DNN optimized weigths.

Figure 1 a) depicts the pipeline of the CSI reconstruction process training a DNN with the built dataset. Usually, the training process requires many epochs and long training times.

B. Real Sensing Matrix Modeling

Typically, the sensing matrix $\mathbf{H}$ used during the training step only considers an ideal environment, assuming that the illumination is perfectly continuous and constant over the scene [9]. Nonetheless, obtaining the same sensing behavior in real laboratory implementations is impossible due to environmental conditions and calibration errors, even when carefully manufacturing and aligning the trained parameters of $\mathbf{H}$.

Therefore, a calibration step is carried out, characterizing the camera's light response to model the real sensing matrix. The characterization process can be formulated as follows

$$\mathbf{H}_r = \mathcal{D}(\mathbf{H}), \tag{3}$$

where $\mathcal{D} : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ models the real implementation artifacts, and $\mathbf{H}_r$ denotes the real sensing matrix. In the case, the operator $\mathcal{D}$ can be modeled as a blur degradation in the coded apertures, with various degradation levels.

III. Rethought Deep Image Prior for Compressive Spectral Imaging Calibration

The DIP is an unsupervised framework that considers the DNN $\mathcal{M}_{\boldsymbol{\theta}}(\cdot)$ as prior information, optimized using only the 2D projected measurements $\mathbf{y}$ and the sensing matrix $\mathbf{H}$ by solving

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \; ||\mathbf{y} - \mathbf{H}\mathcal{M}_{\theta}(\mathbf{z})||_2^2, \tag{4}$$

where $\hat{\boldsymbol{\theta}}$ connotes the optimized weights, and the recovered SI is calculated as $\hat{\mathbf{f}} = \mathcal{M}_{\hat{\boldsymbol{\theta}}}(\mathbf{z})$ with $\mathbf{z} \sim \mathcal{N}(0, 1)$.

Fitting the weights $\hat{\boldsymbol{\theta}}$ from scratch requires high iterations and does not guarantee proper reconstructions due to the inherent non-convex nature of the problem [12]. Hence, we rethink

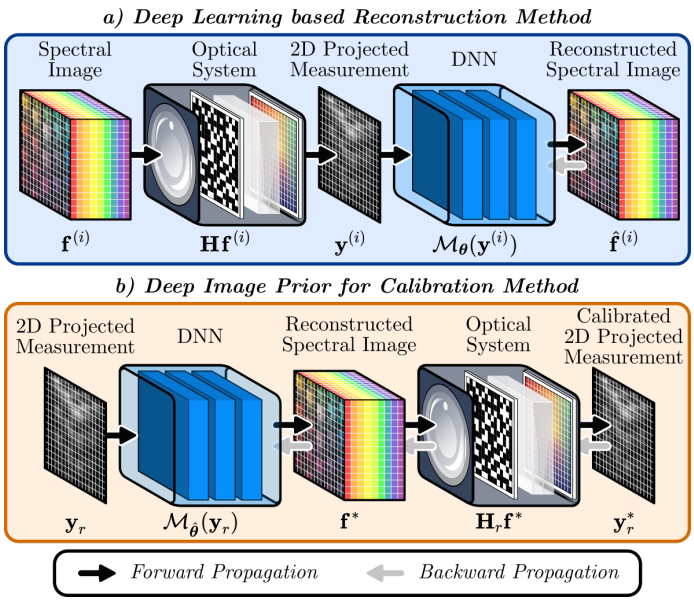

Fig. 1. Pipeline of the proposed calibration method for CSI reconstruction. a) A SOTA DNN is trained with a SI dataset using an ideal sensing matrix $\mathbf{H}$ and ideal calculated 2D projected measurements $\mathbf{y}$. b) The pre-trained SOTA DNN is retrained with the DIP framework, modeling the CSI optical system as an optical encoder layer after reconstructing the SI, where only the real 2D projected measurements $\mathbf{y}_r$ and real sensing matrix $\mathbf{H}_r$ are considered.

the DIP framework to retrain the DNN $\mathcal{M}_{\hat{\boldsymbol{\theta}}}(\cdot)$, previously trained with a simulated sensing matrix $\mathbf{H}$ and simulated 2D projected measurements $\mathbf{y}$, adjusting the DNN weights taking into account a real sensing matrix $\mathbf{H}_r$ and real 2D projected measurements $\mathbf{y}_r$.

Figure 1 b) illustrates the proposed calibration methodology referred to as a ReDIP, where the DNN structure and the sensing matrix are in inverted locations compared to the traditional training setup in Fig. 1 a). Notice that only the real 2D projected measurements $\mathbf{y}_r$ from a single SI is employed in ReDIP, where the retraining acts as a calibration step using a few iterations compared to applying DIP from scratch.

The proposed computational calibration method is modeled with the following optimization problem

$$\boldsymbol{\theta}^* = \arg\min_{\hat{\boldsymbol{\theta}}} \; ||\mathbf{y}_r - \mathbf{H}_r\mathcal{M}_{\hat{\boldsymbol{\theta}}}(\mathbf{y}_r)||_2^2, \tag{5}$$

where $\boldsymbol{\theta}^*$ denotes the optimal weights for calibrating the real 2D projected measurements $\mathbf{y}_r$, and the calibrated reconstructed SI is calculated as $\mathbf{f}^* = \mathcal{M}_{\boldsymbol{\theta}^*}(\mathbf{y}_r)$.

IV. Experiments and Results

A. Experimental Setup

The proposed methodology can be employed for any SOTA DNN. Therefore, we selected two DNNs based on Bayesian frameworks with publicly available implementation and pre-trained weights. These works employed the CASSI system for acquiring a single snapshot with a coded aperture with 0.5 of transmittance.

- **Deep Gaussian Scale Mixture Prior (DGSMP)** [5], based on the maximum posterior estimation. DGSMP

learns the Gaussian scale mixture prior (GSM) to exploit the SI spectral correlations and the GSM local means.

- **Mask Modeling Uncertainty (MMU)** [6], which models the coded aperture uncertainty to treat hardware miscalibration inspired by variational Bayesian learning while performing a bilevel optimization to optimize the SI.

DGSMP and MMU receive the 2D CASSI projected measurements as input and return the reconstructed SI as the output. For the retraining step, the hyperparameters for both DNNs are kept similar to the original implementations, with the following main differences: we use only one batch, corresponding to the real 2D projected measurements, we use the real sensing matrix to retrain the model using the proposed loss function introduced in (5), and we perform 500 gradient iterations.

The experiments employed the KAIST SI dataset [14], containing 30 SIs with a spatial resolution of 2704×3376 and 31 spectral bands. Following the testing setup in [5], [6], we used 10 scenes from the KAIST dataset with a spatial resolution of 256×256 and 28 bands. Specifically, we use each testing scene to retrain the SOTA DNN following the proposed calibration method. In this setup, the 2D CASSI measurement results in a compression level of ≈ 0.04.

The quality of the CSI reconstructions is presented in terms of the peak signal-to-noise ratio (PSNR), the spectral similarity index (SSIM) metric, and the spectral angle mapper (SAM), which were calculated according to the implementation in [11].

B. Calibration Errors Experiment

This experiment analyses the robustness of the proposed ReDIP under calibration errors in the sensing matrix. The calibration errors are emulated by perturbing the coded aperture with different blurring levels. The resulting degraded sensing matrix is then used to acquire the corresponding 2D compressed measurements.

The experiment evaluates four blurring levels denoted by $D_i, i = 0, 1, 2, 3$, where D_0 refers to no blurring, and D_3 refers to the maximum blurring level. For implementation purposes, the blurring levels were applied in the frequency domain following the Matlab function Lowpass with the thresholds $t_r = \{1, 0.8, 0.7, 0.5\}$ for each degradation level, respectively.

Figure 2 depicts a coded aperture with each blurring level and the resulting measurements $\mathbf{y}_r$. Visually, the introduced perturbations in the coded aperture do not appear to affect the measurements. However, when the 2D projected measurements obtained with degraded coded apertures are compared to the obtained with no degradation, the PSNR metric indicates a significant loss in the similarity, which is reflected when reconstructing the underlying SI. Therefore, a precise fit is needed to improve the reconstruction quality.

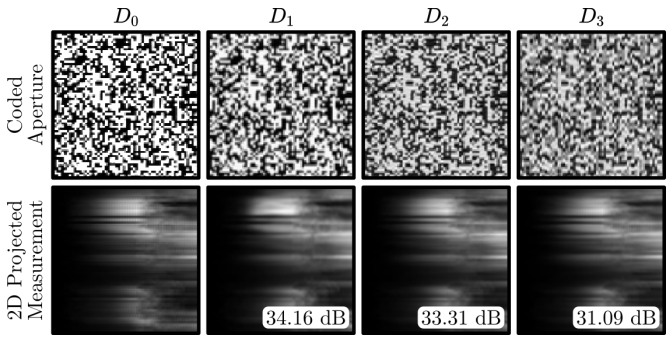

Fig. 2. *(Top)* Visualization of a sub-region of the degraded coded apertures when perturbing with four different degradation levels following (3) *(Bottom)* Visualization of the acquired 2D compressed measurements using the corresponding degraded matrices over the scene #02 in the dataset. The PSNR measures the loss of similarity in the 2D projected measurements.

Table I summarizes the quantitative quality results when using the DGSMP and MMU models across the four degradation levels. The "Baseline" column refers to the performance obtained from the pre-trained SOTA DNNs, and the "ReDIP" column refers to the performance obtained when applying the proposed computational calibration method.

The SI reconstructions obtained with the proposed ReDIP for all tested degradation scenarios achieve a higher performance against the baseline reconstructions. Specifically, for the D_1, D_2 and D_3 degradation levels, the performance outperforms the baseline in up to 7.8dB and 5.6dB for DGSMP and MMU, respectively. On the other hand, in the scenario where ReDIP is employed when there is no degradation, i.e., using D_0, the gain is up to 4.7dB and 2.3dB, respectively. Such results suggest that the developed retraining scheme can improve even the testing results of any SOTA reconstruction method since the weights are iteratively adapted for each specific testing image.

Figure 3 presents an RGB mapping of the SI reconstructions for scenes 02 and 09 for visual comparisons. It can be seen a

TABLE I
AVERAGE AND STANDARD DEVIATION OF THE RECONSTRUCTION QUALITY ACROSS 10 KAIST TESTING SCENES IN TERMS OF PSNR [DB], SSIM AND SAM [RAD].

Degradation	Method	PSNR [dB]		SSIM		SAM [rad]	
		DGSMP	MMU	DGSMP	MMU	DGSMP	MMU
D_0	**Baseline**	30,28 ± 3,09	31,85 ± 2,89	0,92 ± 0,02	0,93 ± 0,01	0,16 ± 0,03	0,14 ± 0,02
	ReDIP	**34,97 ± 4,18**	**34,19 ± 3,95**	**0,96 ± 0,02**	**0,95 ± 0,02**	**0,09 ± 0,02**	**0,12 ± 0,03**
D_1	**Baseline**	21,70 ± 3,09	19,48 ± 2,96	0,71 ± 0,10	0,64 ± 0,09	0,27 ± 0,06	0,48 ± 0,05
	ReDIP	**29,78 ± 3,61**	**26,44 ± 4,03**	**0,89 ± 0,04**	**0,80 ± 0,07**	**0,14 ± 0,04**	**0,27 ± 0,11**
D_2	**Baseline**	22,75 ± 2,96	21,59 ± 2,95	0,77 ± 0,08	0,72 ± 0,07	0,26 ± 0,05	0,37 ± 0,05
	ReDIP	**30,57 ± 3,46**	**27,67 ± 4,25**	**0,91 ± 0,03**	**0,83 ± 0,06**	**0,13 ± 0,04**	**0,24 ± 0,08**
D_3	**Baseline**	21,07 ± 2,90	22,70 ± 2,91	0,73 ± 0,09	0,77 ± 0,07	0,30 ± 0,06	0,26 ± 0,02
	ReDIP	**29,62 ± 3,77**	**28,32 ± 3,97**	**0,90 ± 0,03**	**0,87 ± 0,05**	**0,16 ± 0,05**	**0,22 ± 0,04**

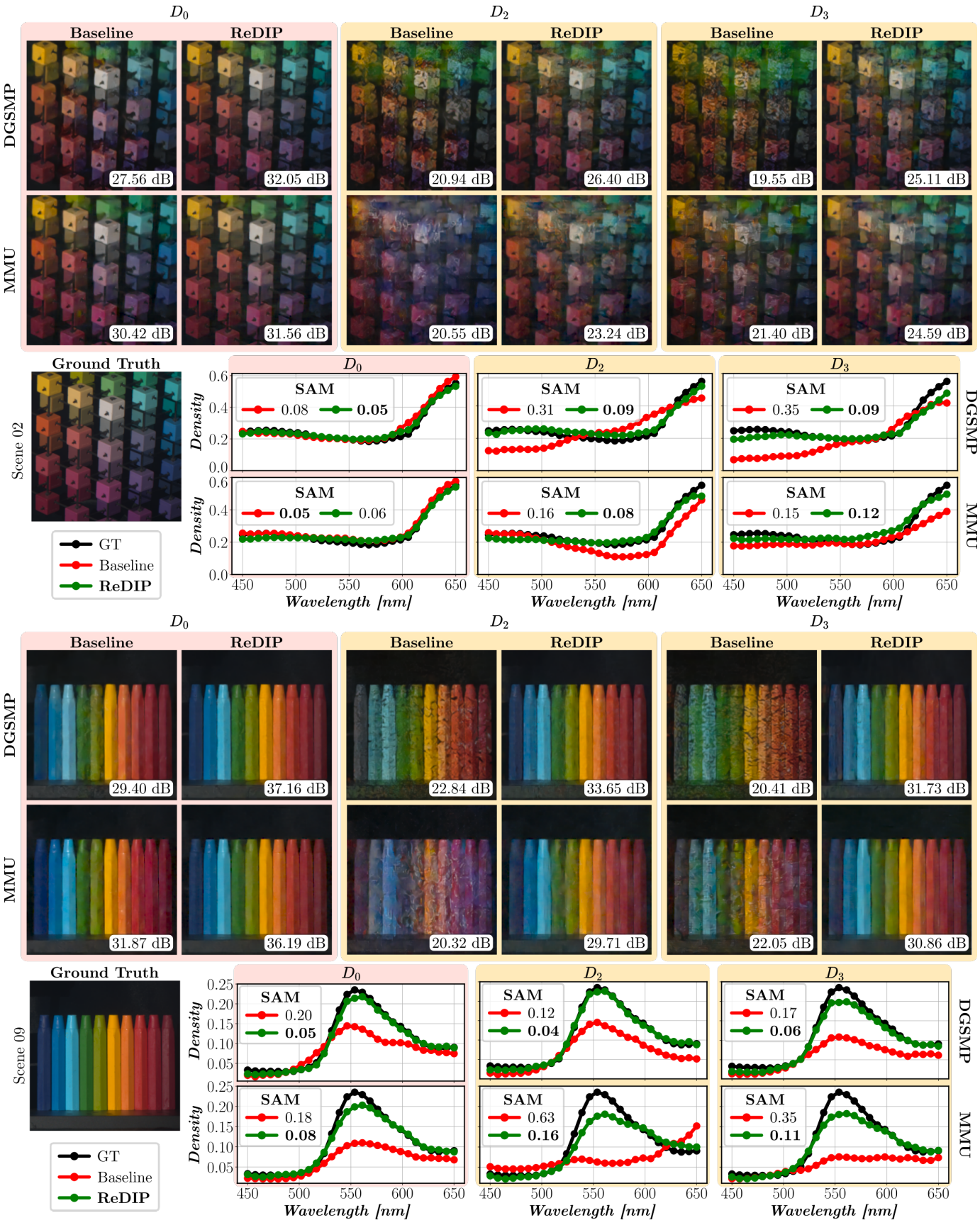

Fig. 3. Qualitative CSI reconstruction results over the testing scenes #02 and #09. *(Top)*: RGB mapping of the SI reconstruction from the baseline (pre-trained DNN) and the proposed method (ReDIP) across three degradation levels D_0, D_2, and D_3, and for both DGSMP and MMU methods. *(Bottom)*: RGB mapping of the ground truth SI with the spectral signature comparison of the pixel located at the middle position for each scenario.

significant improvement with the proposed calibration method, especially under strong degradation scenarios, where the pre-trained DNNs performance drops due to the slight variations in the 2D projected measurements not considered in the training. Additionally, Fig. 3 shows the spectral signature comparison corresponding to the central spatial pixel from the ground truth, the baseline, and the ReDIP reconstructions. Notice that in all the evaluated degradation levels, the spectral signature recovered with the proposed method is more similar to the ground truth than the baseline's spectral signature in terms of the SAM metric. The improvement of the spectral domain can be better seen in the highest degradation scenarios, where the baseline reconstruction entirely deviates from the ground truth spectral signature.

C. Weights Adjustment Analysis

This experiment analyzes visually how the weights of the pre-trained DNN and the obtained quality metrics vary in the reconstruction problem and how the proposed ReDIP calibration methodology improves the results. The experiment was carried out over the baseline DGSMP DNN and the proposed DGSMP ReDIP for D_0 and D_3 degradation levels using scene #02.

To illustrate the influence of the retraining step, one single weight θ from the DNN $\mathcal{M}_{\boldsymbol{\theta}}$ was adjusted manually, exploring how the reconstructed SI $\hat{\mathbf{f}}$ quality and the loss with respect to the measurements $\mathbf{y}$ varies under the degradation scenarios. The default value of the trained weight θ is shown in the dot point in blue in Fig. 4 for the trained scenario D_0; for this case, if the parameters are moved from the optimal, good performance is maintained and when the ReDIP is employed (dark-blue solid line) a better performance is obtained by slightly adjusting the selected weight. However, when the degradation scenario occurs (orange dot line), the performance decreases by maintaining the same weights. Notice that manually adjusting the weights, represented in the orange dot lines, suggests that a better stationary point can be achieved. Therefore, by applying the ReDIP methodology in this degradation scenario, the performance is not only similar to the scenario with no degradation, but the optimal value is shifted a bit from its starting value, which makes sense due to the 2D projected measurements and the sensing matrix changes with respect to the original one.

V. Conclusions

A computational calibration method for SI reconstruction considering the DIP framework was proposed. A qualitative and quantitative improvement over the performance of a pre-trained SOTA DNN concerning real 2D projected measurement and real sensing matrix is achieved in different degradation scenarios. A convergence analysis is done over D_0 and D_3 degradation scenarios, where it can be seen that the adjustment of the weights moves slightly from the baseline, allowing to improve in the SI performance by 5 dB when the degradation is considered and also 5 dB when no degradation is applied.

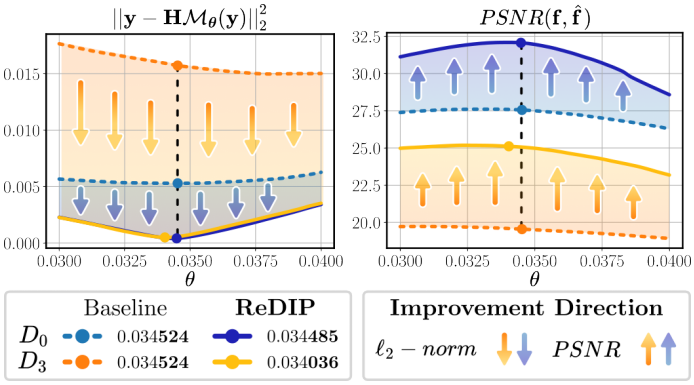

Fig. 4. Influence of the ReDIP calibration method under degradation scenarios D_0 and D_3 with respect to one single parameter θ from the DGSMP model $\mathcal{M}_{\boldsymbol{\theta}}$ with the scene #02. The baseline (blue and orange dot lines) indicates the measurement and reconstruction performance of the trained model for the corresponding dataset. ReDIP (dark-blue and gold solid lines) represents the improvement provided by our proposed method according to the shown direction. Each big dot represents the default achieved parameter value.

References

[1] J. Bacca, E. Martinez, and H. Arguello, "Computational spectral imaging: a contemporary overview," *JOSA A*, vol. 40, no. 4, pp. C115–C125, 2023.

[2] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 105–115, 2013.

[3] J. Bacca, C. V. Correa, and H. Arguello, "Noniterative hyperspectral image reconstruction from compressive fused measurements," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1231–1239, 2019.

[4] T. Gelvez-Barrera, H. Arguello, and A. Foi, "Joint nonlocal, spectral, and similarity low-rank priors for hyperspectral–multispectral image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[5] T. Huang, W. Dong, X. Yuan, J. Wu, and G. Shi, "Deep gaussian scale mixture prior for spectral compressive imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16216–16225, 2021.

[6] J. Wang, Y. Zhang, X. Yuan, Z. Meng, and Z. Tao, "Modeling mask uncertainty in hyperspectral image reconstruction," in *European Conference on Computer Vision*, pp. 112–129, Springer, 2022.

[7] H. Arguello, J. Bacca, H. Kariyawasam, E. Vargas, M. Marquez, R. Hettiarachchi, H. Garcia, K. Herath, U. Haputhanthri, B. S. Ahluwalia, *et al.*, "Deep optical coding design in computational imaging: a data-driven framework," *IEEE Signal Processing Magazine*, vol. 40, no. 2, pp. 75–88, 2023.

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

[9] H. Garcia, J. Bacca, B. Wohlberg, and H. Arguello, "Calibration reinforcement regularizations for optimized snapshot spectral," *Applied Optics*, vol. 7, pp. 1148–1160, 2023.

[10] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9446–9454, 2018.

[11] T. Gelvez-Barrera, J. Bacca, and H. Arguello, "Mixture-net: Low-rank deep image prior inspired by mixture models for spectral image recovery," *arXiv preprint arXiv:2211.02973*, 2022.

[12] J. Bacca, Y. Fonseca, and H. Arguello, "Compressive spectral image reconstruction using deep prior and low-rank tensor representation," *Applied optics*, vol. 60, no. 14, pp. 4197–4207, 2021.

[13] Y. Mejía, H. Arguello, F. Costa, J.-Y. Tourneret, and H. Batatia, "Bayesian reconstruction of hyperspectral images by using compressed sensing measurements and a local structured prior," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3116–3120, IEEE, 2017.

[14] I. Choi, M. Kim, D. Gutierrez, D. Jeon, and G. Nam, "High-quality hyperspectral reconstruction using a spectral prior," tech. rep., 2017.