Deep Learning Methods for Satellite Compressive Imaging

Deep Learning Methods for Satellite Compressive Imaging

in practice, the quality of the reconstruction depends on how close to Gaussian the weights are, and the specific reconstruction algorithm employed.

Each algorithm can be seen as an optimizer based on a specific model of the image data. Since realistic models are difficult to write and optimize in closed form, the best reconstruction results are typically obtained using a "learned" model, i.e. a complex non-linear function having a very large number of parameters θ that "encode" the model, whose values are determined from optimization on a set of training images, leading to a specific reconstruction algorithm $R(y, \Phi, \theta)$. Learning such model is best done using deep neural networks (DNN), that is why in this paper we implement a deep learning-based reconstruction.

Notice that, typically, a smaller number of measurements ($m < n$) are acquired with respect to the number of pixels, thus achieving compression of the information. We define the compression ratio (CR***)*** as the ratio between the number of measurements and the number of pixels:

$$CR = \frac{m}{n}. \tag{3}$$

A lower CR is desirable, as this will reduce the data size for the transmission to the ground segment; this yields a "compression" functionality without the need of a specific compression circuitry. On the other hand, this also implies that vector y may not necessarily contain all the information to reconstruct the image x exactly and, even in the case of no noise, one cannot employ $R(y, \Phi, \theta) = \Phi^{-1}y$ because the matrix Φ is not invertible. In practice CS always attains an approximate reconstruction, therefore the CR has to be chosen as a trade-off between conflicting requirements, i.e. the amount of data size reduction and the quality of the reconstructed image. The difficulty to reconstruct an image from far fewer measurements highlights the need for complex reconstruction algorithms and for this reason a deep learning approach is suitable to improves the reconstruction performance.

ISTA-NET ALGORITHM

The ISTA-Net framework proposed in [7] consists of mapping the classic Iterative Shrinkage/Thresholding Algorithm (ISTA) [9] into a DNN. The architecture is made up of a fixed number of steps that correspond to iterations in the traditional algorithm and the network is depicted in Fig. 1.

In the traditional ISTA algorithm, the update step is made up by two equations. These equations correspond in the ISTA-Net architecture to two modules. The first called $r^{(k)}$ corresponds to the evaluation of the reconstructed signal at step k:

$$r^{(k)} = x^{(k-1)} - \rho^{(k)}\Phi^{(\mathrm{T})}(\Phi x^{(k-1)} - y) \tag{4}$$

The only main difference with traditional ISTA is that now ρ can vary across iterations instead of being fixed. The second module is the $x^{(k)}$ module. Also in this case, the module is a particular case of proximal mapping associated to the non-linear transform $F(\cdot)$, based on the assumptions that each element $x^{(k)} - r^{(k)}$ has an independent distribution with common zero mean and variance σ^2.

In this algorithm, the step size $\rho^{(k)}$, the parameters of the forward and backward transform $\mathcal{F}^{(k)}(\cdot)$ and $\tilde{\mathcal{F}}^{(k)}(\cdot)$, as well as the shrinkage threshold $\theta^{(k)}$ are learned as neural network parameters. Like traditional ISTA, ISTA-Net initialization $x^{(0)}$, given any input CS measurement y, is computed as: $x^{(0)} = Q_{\text{init\_}y}$, with $Q_{\text{init\_y}}$ equal to linear mapping matrix.
The loss function is defined as:

$$\mathcal{L}_{\text{total}}(\Theta) = \mathcal{L}_{\text{discrepancy}} + \mathcal{L}_{\text{constraint}} \tag{5}$$

When dealing with natural images, it is possible to optimize the reconstruction process and the result is ISTA-Net^{+} [7]. The substantial difference with ISTA-Net is in module $x^{(k)}$ where we have:

$$x^{(k)} = r^{(k)} + G(\tilde{H}(soft(H(D(r^{(k)})), \theta))) \tag{6}$$

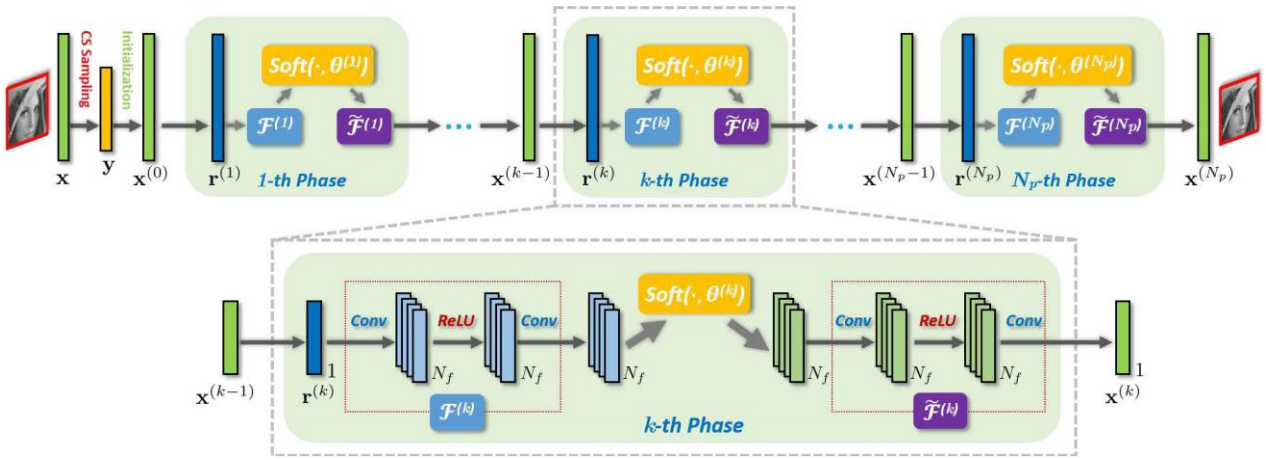

Fig. 1. ISTA-Net **[7]**

Fig. 1 and Fig. 2 show the ISTA-Net block diagram and the ISTA-Net+ $k\text{-}th$ phase. In Fig. 2 it is possible to notice the changes with respect to the k-th phase of the baseline version of the algorithm, like the addition of the recursive term $r^{(k)}$ and the additional filters $\mathcal{D}^{(k)}$ and $\mathcal{G}^{(k)}$ to retrieve some missing high frequencies.

The results shown in [7] greatly improve over the results obtained by traditional iterative algorithms at all compression ratios, and they perform better also for what concerns the computational time. Furthermore in [7] is highlighted how ISTA-Net+ is very flexible as it is possible to let learnable parameters be shared or unshared among the different phases.

In order to study the advanced reconstruction algorithms for the SURPRISE system, it is important to utilize a dataset of satellite images, that includes a large enough collection of training samples. The dataset that we are going to use in this paper is a subset of the data released by the IEEE Geoscience and Remote Sensing society for the 2020 Data Fusion Contest, namely DFC2020 [10]. The data are images acquired by the Sentinel 2 satellite, having 12 spectral channels and a ground sampling distance of up to 10m. Our DFC2020 subset is composed of 5178 images with 256x256 pixels. These are partitioned into 5128 images to be used for algorithm training purposes and 50 for testing. Fig. 3 reports one example from the dataset.

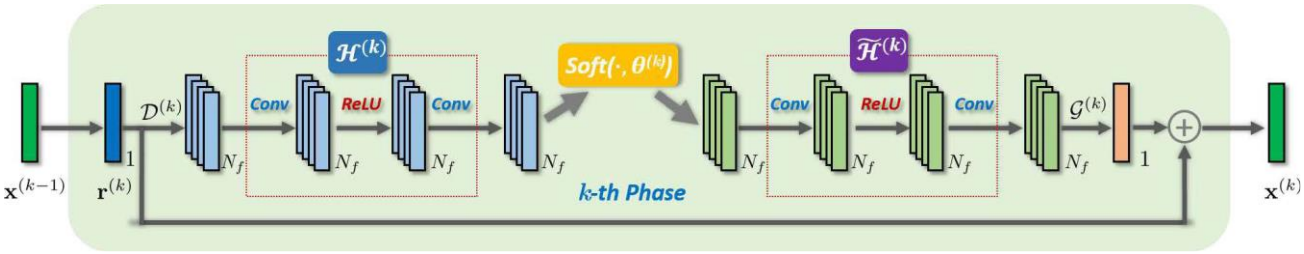

Fig. 2. ISTA-Net+ **[7]**

Fig. 3. DFC2020 dataset, image 47, band 2 **[10]**

CONVOLUTIONAL IMAGE RECONSTRUCTION

The activity described in this section is concerned with improving the quality of the reconstructed image in the areas around the boundary between sensing blocks. Since each block is independently sensed, it is natural to formulate the reconstruction problem independently. This has the advantage of parallelizing the reconstruction process across the multiple blocks comprising the scene. However, it does not account for the existence of correlation among nearby pixels but belonging to different blocks. This results in a reconstruction exhibiting sharp discontinuities at the boundaries between blocks, particularly at compression ratios below 50%.

In order to address this suboptimality, we leverage the fully-convolutional nature of the ISTA-Net neural network used for deep-learning-based reconstruction. We can recall that ISTA-Net is composed of stages, each performing a 2D convolution operation between a progressively more accurate estimate of the reconstructed image and sets of learned filters. Moreover, each stage computes a residual in the measurements domain by simulating the sensing process of each pixel. The convolution operation is well-defined for images of arbitrary dimension. Allowing the learned filters to slide over the entire image comprising the reconstructions of multiple blocks allows to exploit the inter-block dependencies as well as the intra-block ones. Two modifications to the standard ISTA-Net design and training process are required to fully take advantage of this property:

1. An area corresponding to multiple blocks must be provided at training time, instead of the conventional single pixels. Since this approach increases the memory requirements, we found that just using an area of 2x2 blocks was sufficient to learn the inter-block correlation patterns and it not necessary to use an entire scene.
2. Careful simulation of the sensing process which acts independently for each block is needed for the residual computation in the network stages.

We tested the reconstruction process on the DFC2020 dataset with the 32x32 super-resolution configuration. Training used 51280 patches of size 64x64 from the images in the training set to simulate the area corresponding to 2x2 blocks. Table **1** reports the root mean squared error (RMSE) achieved by conventional independent reconstruction and with the fully convolutional reconstruction on the 50 images from the test set. The network processes all the blocks in the image at the same time during testing, in order to maximally avoid border effects. It can be noticed that the convolutional reconstruction provides slightly lower RMSE values. Fig. **4** and Fig. **5** show the reconstruction achieved by the independent and convolutional methods at 25% compression ratio. It can be noticed that the convolutional method is more effective at suppressing blocking artifacts between blocks.

MULTIBAND IMAGE RECONSTRUCTION

In this section, we explore the concept of simultaneous reconstruction of multiple spectral bands rather than independent reconstruction. The SURPRISE instrument acquires the content of the scene at multiple wavelengths. However, some correlation between the scene imaged in different spectral bands might exist and could be fruitfully exploited to improve the reconstruction performance. Existing CS reconstruction algorithms, including the state-of-the-art ones using deep learning such as ISTA-Net, have neglected this issue, focusing on single-band reconstruction. This activity has modified the ISTA-Net algorithm in order to exploit the inter-band correlation in the reconstruction process and benchmarked the reconstruction performance against independent reconstruction.

In particular, ISTA-Net is based on the unrolling of an iterative algorithm into a fixed number of stages. The input to each stage is an estimate of the reconstructed image, which at the first stage is obtained via an optimized linear operator from the CS measurements. The joint reconstruction technique considers the spectral channels as feature channels which are merged into higher-level features by means of a weight matrix applied along the channel dimension by the convolutional layers of each stage. This allows to extract latent features that depend on all the input spectral bands and that are decoded by the last convolutional layer in an equal number of spectral channels. This multiband reconstruction model has a small number of extra trainable parameters with respect to the single-band model due to the weights mixing the channels to extract the latent features. However, this increase in parameters is very small when compared to the total number of parameters of the model and has negligible effects on computational complexity.

For this activity we simulated the acquisition process using three bands in the visible spectrum. The sensing process is independent for each channel while independent and joint reconstruction are compared in the following benchmark for the 32x32 SR configuration with various compression ratios. Moreover, the fully-convolutional reconstruction procedure described in the previous section is used in the following experiments. Table 2 compares the RMSE of the reconstructed images in the DFC2020 test set for the initial reconstruction via linear operator, the single-band technique and the multi-band technique. It can be noticed that, for this dataset, exploiting the inter-band correlation decreases the reconstruction error.

Table 1. Independent reconstruction vs. convolutional reconstruction (RMSE)

		RMSE	
		Independent reconstruction	**Convolutional reconstruction**
CR	**25%**	74.69	72.87
	50%	47.78	47.04
	75%	29.36	28.73

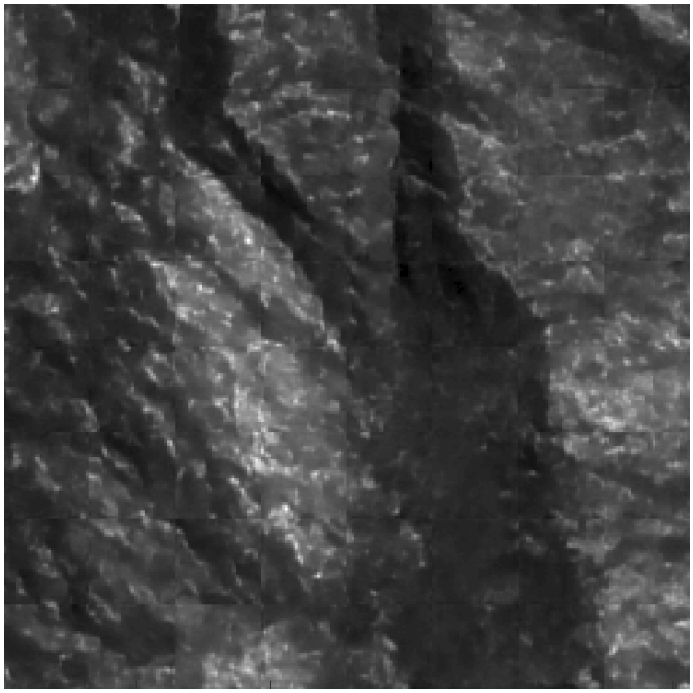

Fig. 4. Independent reconstruction. Sensing matrix 32x32, CR 25%. Notice the blocking artifacts**.**

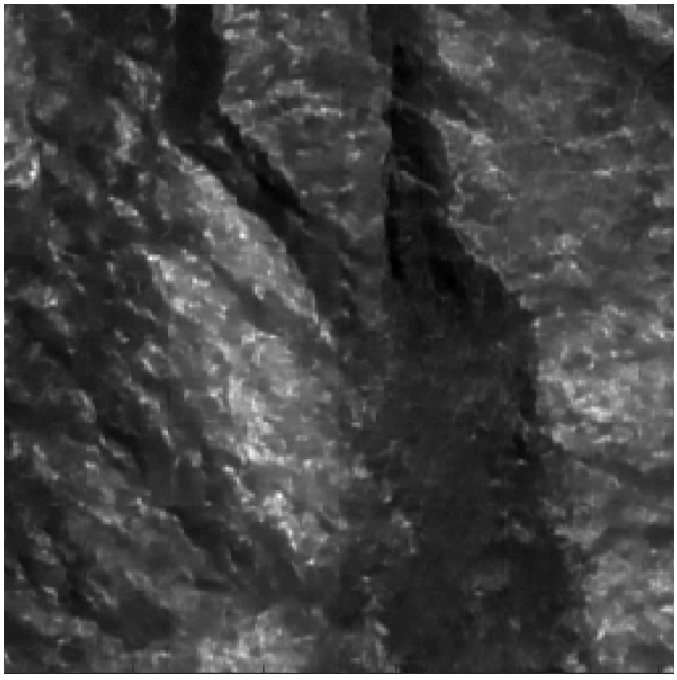

Fig. 5. Convolutional reconstruction. Sensing matrix 32x32, CR 25%. Notice that the blocking artifacts have been significantly suppressed.

Table 2. Single band reconstruction vs. multi band reconstruction (RMSE)

		RMSE	
		Single band reconstruction	**Multi-band reconstruction**
CR	**25%**	72.87	65.74
	50%	47.04	42.43
	75%	28.73	26.64

RECONSTRUCTION ACCURACY IN PRESENCE OF NOISE

In this section we explore the impact of noise in the acquisition process on the quality of reconstruction. Given a sensing process affected by noise $z = \Phi x + \eta = y + \eta, \eta$ can be modelled according to the following distribution:

$$\eta \sim \mathcal{N}(0, ay + b) \tag{7}$$

This model is the heteroskedastic normal approximation of a Poisson-Gaussian noise [11]. The signal dependent component ay models the noise derived from quantum phenomena during photon accumulation, while b describes the other signal-independent noise sources, such as thermal noise.

We decided to simulate a purely signal dependent noise $\eta \sim \mathcal{N}(0, ay)$ on the DFC2020 datasets and to re-train ISTA-Net on the noisy measurements. Three levels of noise were simulated: SNR_z = 50, 75, 100 [dB]. We define SNR_z as the signal-to-noise ratio on the sensed signal affected by noise z. It is obtained with the following equation:

$$SNR_z = 10 \log_{10} \frac{\sum_i y^2(i)}{\sum_j \eta^2(j)} \tag{8}$$

SNR_z is the measure in dB of the ratio between the power of the original sensed signal y and the simulated noise η. Table **3** compares the RMSE for these three levels of SNR_z and the noiseless case. The system was tested for four compression ratios: 25, 50, 75 and 100%. Thanks to the re-training process, the Network is capable to adapt to the different levels of SNR. The impact of the noise on the reconstruction is minimal, and it is noticeable only at very high levels of SNR (e.g., 50 dB). In Fig. **6** we show an example of the reconstructions obtained for noise at 50 dB using a compression rate of 25%.

At this point we compared the results obtained using a purely signal-independent noise for the same level of SNR_z. Again, the test was repeated for SNR_z = 50, 75, 100 [dB]. In the following results we refer to the signal-dependent noise as ay and we use b to refer to the signal-independent noise. As it is shown in Table **4**, the difference in the distribution of the noise seems to affect the quality of the reconstruction in a negligible way.

Table 3. Comparison in reconstruction performance between Linear Reconstruction and ISTA-Net for various levels of compression and SNR for the signal-dependent noise

Root Mean Squared Error (RMSE)								
	Compression Ratio 25.00%		**Compression Ratio 50.00%**		**Compression Ratio 75.00%**		**Compression Ratio 100.00%**	
SNR_z [dB]	**Linear Recon.**	**ISTA-Net**	**Linear Recon.**	**ISTA-Net**	**Linear Recon.**	**ISTA-Net**	**Linear Recon.**	**ISTA-Net**
50	90.06	81.14	78.81	64.50	74.56	60.39	72.27	58.08
75	79.47	71.84	51.58	48.54	31.70	29.39	16.94	15.07
100	79.06	74.07	51.33	47.49	30.36	28.47	5.52	5.51
Noiseless	79.00	72.90	51.30	47.00	30.50	28.70	0.59	0.59

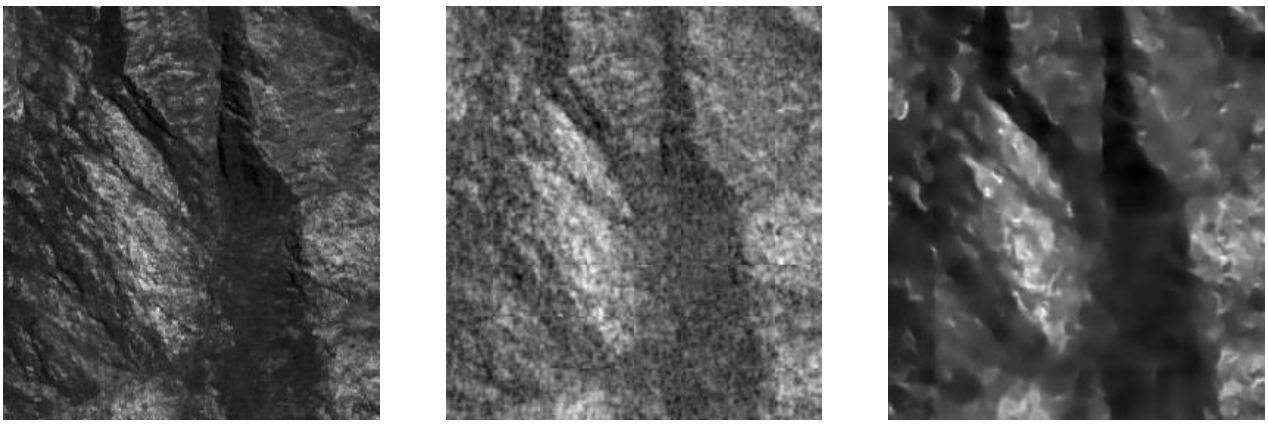

Fig. 6. Comparison between Ground Truth image *(on the left)*, the reconstruction obtained using Linear Reconstruction *(center)* and on the reconstruction obtained using ISTA-Net (*on the right*) with 50 dB of SNR and SNR 25 dB.

Table 4. Comparison in reconstruction performance between signal-dependent and signal-independent noise

Root Mean squared Error (RMSE)									
		CR = 25.00%		**CR = 50.00%**		**CR = 75.00%**		**CR = 100.00%**	
SNR_z **[dB]**	Algorithm	Noise ay	Noise b	Noise ay	Noise b	Noise ay	Noise b	Noise ay	Noise b
50 dB	**Linear Recon.**	90.06	94.83	78.81	74.93	74.56	75.65	72.27	67.42
	ISTA-Net	81.14	88.15	64.50	64.33	60.39	60.64	58.08	56.32
75 dB	**Linear Recon.**	79.47	79.38	51.58	51.58	31.70	30.97	16.94	14.73
	ISTA-Net	71.84	73.08	48.54	47.61	29.39	29.14	15.07	13.95
100 dB	**Linear Recon.**	79.06	79.15	51.33	51.33	30.36	30.36	5.52	4.72
	ISTA-Net	74.07	73.89	47.49	48.37	28.47	28.38	5.51	4.73

CONCLUSIONS

In this paper we describe the development of an improved version of the state-of-the-art deep learning algorithm for CS reconstruction, destined to be used for compressive satellite imaging. Our experiments demonstrate that a data-driven algorithm represents a viable solution to obtain a high-resolution version of a satellite image, taking into consideration some of the effects related to a real-world scenario.

ACKNOWLEDGEMENTS

The SURPRISE project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 870390.

REFERENCES

[1] "SURPRISE: SUper Resolved ComPRessive InStrumEnt," [Online]. Available: https://www.h2020surprise.eu/.

[2] E. Magli, T. Bianchi, D. Guzzi, C. Lastri, V. Nardino, L. Palombi, V. Raimondi, D. Taricco and D. Valsesia, "Compressive imaging and deep learning based image reconstruction methods in the "SURPRISE" EU project," in *Proceedings of the European Workshop on On-Borad Data Processing (OBDP)*, 2021.

[3] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE signal processing magazine,* vol. 25, no. 2, p. 21–30, 2008.

[4] M. A. Davenport, P. T. Boufounos, M. B. Wakin and R. G. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected Topics in Signal Processing,* vol. 4, no. 2, pp. 445-460, 2010.

[5] G. Olivieri, M. Salucci, N. Anselmi and A. Massa, "Compressive sensing as applied to inverse problems for imaging: theory, applications, current trends, and open challenges," *IEEE Antennas and Propagation Magazine,* no. 59, pp. 34-46, 2017.

[6] A. Mousavi and R. G. Baraniuk, "Learning to invert: signal recovery via deep convolutional networks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[7] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[8] E. Candes and T. Tao, "Decoding by Linear Programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203-4215, 2005.

[9] A. Chambolle, R. A. DeVore, N. Lee and B. J. Lucier, "Nonlinear wavelet image processing: Varational problems, compression, and noise removal through wavelet shrinkage," *IEEE Transaction on Image Processing*, vol. 7, no. 3, pp. 319-35, 1998.

[10] C. Robinson, K. Malkin, N. Jojic, H. Chen, R. Qin, C. Xiao and e. al., "Global land-cover mapping with weak supervision: Outcome of the 2020 IEEE GRSS data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021.

[11] A. Foi, M. Trimeche, V. Katkovnik and K. Egiazarian, "Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737-1754, 2008.