Compressive imaging and deep learning based image reconstruction methods in the "SURPRISE" EU project

EUROPEAN WORKSHOP ON ON-BOARD DATA PROCESSING (OBDP2021), 14-17 JUNE 2021

COMPRESSIVE IMAGING AND DEEP LEARNING BASED IMAGE RECONSTRUCTION METHODS IN THE "SURPRISE" EU PROJECT

Enrico Magli[(1)], Tiziano Bianchi[(1)], Donatella Guzzi[(2)], Cinzia Lastri[(2)], Vanni Nardino[(2)], Lorenzo Palombi[(2)], Valentina Raimondi[(2)] , Davide Taricco[(1)] , Diego Valsesia[(1)]

[(1)]Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
[(2)]CNR-IFAC, Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy

ABSTRACT

Compressive sensing techniques can greatly reduce the complexity of imaging instruments used in space applications, by reducing sampling requirements and power consumptions without sacrificing image resolution. In this paper, we present the compressive imaging technique developed within the EU project "SURPRISE". A suitable CS model adapted to the optical layout of the instrument enables image reconstruction at a resolution higher than the sensing elements. Both classical CS reconstruction techniques and deep learning-based methods are investigated, showing that the latter achieves a better quality of the reconstructed images for different compression ratios. Interestingly, deep learning-based techniques also reduce the complexity of CS reconstruction, since most of the computations are performed in the training phase that can be executed offline.

1. INTRODUCTION

Increasing amounts of data are being generated by imaging sensors in space applications, due to the evolution of image acquisition technologies and the need to obtain high-resolution imagery for improved scientific and commercial exploitation. As reported in [1] along with more data, come increased requirements on the computation and memory resources available on the onboard processing units, which must be able to handle very high data rates with limited power consumption, and on the transmission system, which acts as a bottleneck and prevents full exploitation of the acquired information.

Signal processing techniques can help to address these issues allowing information extraction from the acquired data and images directly onboard satellites. This allows not only to screen the acquired images and select those corresponding to events of interest, but also to generate alarms directly on the satellite, which can be communicated to the ground stations with very low latency, bypassing the image transmission and processing chain.

One of the most important signal processing techniques that can have a role in this scenario is CS [2]. CS represents a signal (or image) through a small set of linear measurements, under the assumption that the signal is "sparse" in some domain, i.e., it has only few significant coefficients, while most are close to zero. In this case, nonlinear reconstruction algorithms can recover the signal exactly or with high fidelity.

CS has been employed for efficient computational imaging, e.g., reducing the number of detectors as in the single-pixel camera (see [3]). Moreover, it has been shown [4] that it is possible to extract information directly from the acquired random projections, enabling the processing of the data directly onboard, without performing the computationally heavy reconstruction process which typically takes place at the ground segment. Finally, in [5] it has been shown that CS can also be used as a cryptosystem, where the sensing matrix serves the role of an encryption key. These features are very interesting for space applications where CS technology potentially enables the design of a class of innovative computational imaging instruments perfectly suited to "big data from space", providing native compression and encryption, enabling onboard data analysis, while at the same time relieving the requirements on the onboard processing system.

This paper describes the compressing imaging instrument developed in the "SURPRISE" EU project, discussing the acquisition model and the main reconstruction techniques. As an important contribution, it is shown that deep learning-based reconstruction methods can sensibly improve the quality of the reconstructed images with respect to conventional techniques, which enables simpler sampling architectures for the onboard instrument. Moreover, the complexity of deep learning methods mainly lies in the training of the network: once the reconstruction network is properly trained, the reconstruction algorithm is usually faster than the traditional optimization methods used in CS, which is an additional advantage of the proposed architecture.

2. COMPRESSIVE SENSING

CS is a mathematical framework that allows to solve under-determined systems and that, applied in the field of image processing, allows to represent signals in a compact (compressed) way. This framework tries to overcome the Shannon-Nyquist theorem by using a

number of data "samples" as close as possible to the number of degrees of freedom of the signal. So, the acquired data is much less than in the classical case, making the signal much easier to manipulate and transmit. Nowadays a very large amount of data is collected, but only a small portion is actually used. Most of it, in fact, is collected and then discarded via lossy compression. For example, for an image of megapixel resolution, compressed in JPEG 2000, just a few kilobytes of data are retained. CS attempts to avoid acquiring large amounts of data just to immediately discard most of them through the compression process.

CS relies on two main concepts: sparsity and incoherence. A signal is said to be sparse when most of its components are equal to zero. In particular a signal $x \in \mathbb{R}^n$ is k-sparse if it has at most k non-zero entries, i.e. $\|x\|_0 \leq k$.
In real life, signals are often sparse in a domain different from the original one (e.g., image pixels are not sparse, but their transformed representation may be). This means that x will be considered sparse also if it has a sparse representation in some basis Ψ:

$$x = \Psi c \quad \text{with } c \text{ sparse} \tag{1}$$

Moreover, signals are rarely exactly sparse, so it is required to approximate the concept of sparsity with the concept of "compressibility". So, typically, once data are acquired, compressing a signal is equivalent to approximating the signal x with a signal $\hat{x} \in \Sigma_k$, defined as the set of all k-sparse signals, such that the approximation error is minimum:

$$\sigma_k(x)_p = \min_{\hat{x} \in \Sigma_k} \|x - \hat{x}\|_p \tag{2}$$

i.e., $\hat{x}$ is obtained by retaining just $k \ll n$ components of x (or c), that are the ones carrying almost all the signal energy.
Given a sparse or compressible signal x, with CS it is possible to obtain the compressed data y directly. The sensing operation can be easily written as:

$$y = \Phi x \tag{3}$$

with $y \in \mathbb{R}^{m \times 1}, x \in \mathbb{R}^{n \times 1}$ and $\Phi \in \mathbb{R}^{m \times n}, m \ll n$.
Recalling the model in (1), the sensing problem becomes $y = \Phi\Psi c$ and can be solved for a certain sparse vector $\hat{c}$ provided that $\Phi\Psi$ satisfies suitable properties; in this context the notion of incoherence is relevant.
Incoherence refers to the fact that the sensing matrix should have a very dense representation in the representation basis Ψ. Coherence is defined as:

$$v(\Phi, \Psi) = \sqrt{n} \max_{1 \leq kj \leq n} \left|\langle \Phi_k, \Psi_j \rangle\right|, \qquad v \in [1, \sqrt{n}]$$

where Φ_k and Ψ_j are the columns vector of Φ and Ψ. So, if the coherence between the sensing matrix and the representation basis is low, then it will promote CS operations. It can be proved that random matrices Φ have a very low coherence with any fixed basis Ψ.

3. CS RECONSTRUCTION ALGORITHMS

In this section the state of the art of reconstruction algorithms for CS, based on traditional and deep learning approaches, are presented.

3.1. Traditional reconstruction algorithms

Let us consider the sensing process:

$$y = \Phi x + \eta \tag{4}$$

affected by the noise $\eta \in \mathbb{R}^m$ whose ℓ_2 norm is bounded by some $\varepsilon > 0$ and where $x \in \mathbb{R}^n$, $\Phi \in \mathbb{R}^{m \times n}$ and $m < n$. If we assume x sparse in the original domain, i.e., $\Psi = I$, the reconstruction of the original signal sampled with CS, can be obtained solving the ℓ_o problem:

$$\hat{x} = \min \|\hat{x}\|_0 \qquad \text{s.t.} \quad \|y - \Phi x\|_2 < \varepsilon \tag{5}$$

where $\|\hat{x}\|_0$ is the ℓ_o pseudo norm, i.e. the number of non-zero entries of $\hat{x}$.
The recovery of the signal is guaranteed under some conditions [6], but the problem is NP-hard and performing an exhaustive search of the sparsest solution is not feasible.
Usually, this problem can be solved using several different approaches, including *greedy algorithms*: fast with low-sparsity signal, but they are sensitive to noise. The main representatives are Matching Pursuit [7], Orthogonal Matching Pursuit [8], Compressed Sampling Matching Pursuit [9], Regularized Orthogonal Matching Pursuit [10] and Subspace Pursuit [11]; *iterative thresholding algorithms*: fast, they recover measurements by soft and hard thresholding throw iterations. Iterative Hard Thresholding based algorithms includes Message Passing [12], Sequential Sparse Matching Pursuits [13] and Belief Propagation [14]. Under Iterative Soft Thresholding algorithms, Iterative Shrinkage/Thresholding Algorithm (ISTA) [15] and its faster FISTA [16] can be found; *convex relaxation algorithms*: signal reconstruction by linear programming with higher complexity then greedy approach. The most relevant examples are Basis Pursuit [17], Least Absolute Shrinkage and Selection Operator [18] and basis pursuit denoising [19]; *non-convex relaxation algorithms*: faster than previous solutions but prone to find local optima instead of the global one. In this category Focal Underdetermined System Solver [20], Iterative Re-weighted Least Squares [21] and Bayesian CS [22] can be found.

A convenient solution, especially in the case of image reconstruction, is to replace the minimization of the ℓ_o norm with the minimization of the Total Variation (TV) pseudo-norm [23]. This approach is particularly useful for natural images that exhibit compressible gradients i.e. the most of the energy is packed in very few of the image gradients. In the TV formulation, the reconstruction problem becomes:

$$\hat{X} = \arg\min_X \mathrm{TV}(X) + \lambda \|y - \Phi x\|_2^2 \qquad (6)$$

where X is a two-dimensional image, x is the vectorized version of the image X and the isotropic TV is defined as:

$$\mathrm{TV}(X) = \sum_{i,j} \sqrt{\left|X_{i+1,j} - X_{i,j}\right|^2 + \left|X_{i,j+1} - X_{i,j}\right|^2}$$

The TV reconstruction method is fast and widely used, so it has been chosen to represent traditional reconstruction algorithms.

3.2. Deep learning-based reconstruction algorithms

In recent years, the generalization capability of deep neural networks has shown to considerably improve the performance of previous state-of-the-art technologies in many fields, including CS reconstruction.
In [24] a stacked denoise auto-encoder is proposed to improve signal reconstruction performance using the statistical dependency between the different entries of the signal. A reconstruction network based on fully connected layers had been presented in [25] and [26]. In [27] authors presented a more advanced CS algorithm based on Convolutional Neural Network (CNN) named ReconNet which performs a regression to reconstruct the original image blocks starting from the CS measurements. In [28] a fully convolutional solution is presented for CS reconstruction. In [29] authors use a CNN to implement the iterative shrinkage-thresholding algorithm (ISTA-Net). Lastly in [30] a scalable convolutional neural network (SCSNet) is presented to improve the image quality reconstructed with a greedy algorithm using a hierarchical reconstruction network.
The two best performing frameworks are ISTA-Net/ISTA-Net$^+$ and CS/SCSNet. The former has been chosen over the latter, because CS/SCNET perform best using a sensing matrix adapted to the dataset, whereas ISTA-Net/ISTA-Net$^+$ have proven to perform better using generic sensing matrices.

3.3. ISTA-Net algorithm

ISTA-Net [29] framework consists of mapping each classic ISTA algorithm update step into a deep network architecture in which there is a fixed number of phases that correspond to iteration in the traditional algorithm. In [29] the authors propose a general nonlinear transformation to sparsify images and, in particular, a combination of two convolutional operators separated by a ReLU. Moreover, all the parameters in this algorithm are learned end-to-end, rather than being hand-crafted.
The two equations corresponding to the update steps in ISTA are replaced by two neural network modules. The first one, called $r^{(k)}$, corresponds to the evaluation of reconstructed signal at step k:

$$r^{(k)} = x^{(k-1)} - \rho^{(k)} \Phi^{(\mathrm{T})}\left(\Phi_x^{(k-1)} - y\right). \qquad (7)$$

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, given by

$$x^{(k)} = \tilde{\mathcal{F}}^{(k)}\left(soft\left(\mathcal{F}^{(k)}\left(r^{(k)}\right), \theta^{(k)}\right)\right). \qquad (8)$$

Also in this case, the module is a particular case of proximal mapping associated to the non-linear transform $\mathcal{F}^{(k)}(\cdot)$, based on the assumptions that each element $x^{(k)} - r^{(k)}$ has an independent distribution with common zero mean and variance σ^2. 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 NN 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:

$$\mathcal{L}_{\text{total}}(\Theta) = \mathcal{L}_{\text{discrepancy}} + \mathcal{L}_{\text{constraint}} \qquad (9)$$

where $\mathcal{L}_{\text{discrepancy}}$ imposes that the reconstructed signal is close to training data whereas $\mathcal{L}_{\text{constraint}}$ imposes that the backward transform approximates the reverse of the forward transform.
Since it provides an excellent trade-off between reconstruction accuracy and flexibility, the ISTA-Net framework has been chosen as the starting point for the development of deep learning-based reconstruction algorithms in this paper.

4. SURPRISE INSTRUMENT

This section describes the proposed SURPRISE instrument, and the corresponding image acquisition model that is needed to successfully reconstruct the images from the subsampled measurements. An overview of the instrument is presented in Figure 1.
The main components of the instrument are:

- *scanning system,* in order to enable whiskbroom operation;
- *spatial light modulator* (SLM) which modulates the scene according to a binary pattern, referred to as *sensing mask*; it can be implemented with a digital micromirror device (DMD);

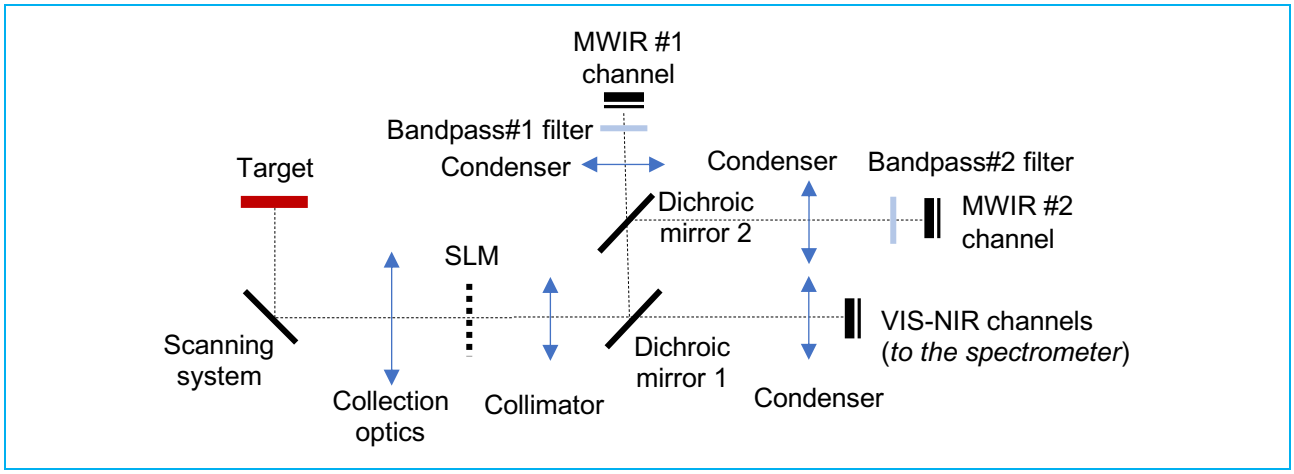

Figure 1: Schematic of the optical layout of the SURPRISE instrument

- *set of dichroic mirrors* used to split VIS-NIR light from IR light and the latter, subsequently, into the spectral bands of interest for the two MWIR channels;
- *set of optical condensers* that perform the signal spatial integration and focus it on the relevant channel's detector;
- *spectrometer and detection system* used to implement the VIS-NIR channels;
- *single element MWIR detectors* and relevant bandpass filters used to implement the two MWIR channels.

The SURPRISE instrument is a super-resolved whiskbroom imager, serially acquiring the scene one macro-pixel at a time. The term macro-pixel refers to the image captured in the instrument's instantaneous field of view (IFOV). Due to the coded aperture design exploiting a spatial light modulator, each macro-pixel can be reconstructed into a block of SxS micro-pixels, where the value of S depends on the design super-resolution factor.

4.1. Simplified Acquisition model

We first present a simplified model to highlight the terminology in the compressive instrument. In particular, we can start by modelling the measurements acquired by a macro-pixel in band i, for a single spatial location P in the whiskbroom acquisition procedure:

$$y_i^P = \Phi_i^P x_i^P \qquad (10)$$

In this model we have:

- y_i^P : column vector of m measurements acquired by the detector in band i
- x_i^P : column vector of n image micro-pixels
- Φ_i^P : sensing matrix. It depends on the binary sensing mask used on the SLM. Notice it might be a function of wavelength.

The sensing matrix can be modelled as an $m \times n$ matrix where each row has different coefficients, corresponding to each serially acquired measurement. A detector macro-pixel acquires $1 \leq m \leq n$ measurements. When $m > 1$, multiple measurements are acquired by serially changing the sensing mask on the SLM. The ratio between the number of measurements m and the number of micro-pixels n is called *compression ratio.* On the other hand, n is the super-resolution factor, i.e., how many micro-pixels can be reconstructed from a macro-pixel measurements.

While ideally the Φ_i^P matrix should be a binary 0/1 matrix, in practice due to nonideality of the SLM it is full (no zero entries) and its entries are binary-valued with a low value equal to α and a high value equal to β. The exact values of the α and β depend on the SLM calibration and can change as a function of wavelength, so the same matrix cannot be reused to model the acquisition of all detectors in all bands. A first approximation can be considered that fixes $\alpha = 0$ and $\beta = 1$ and an unknown rescaling factor that can be measured at calibration time. This approximation is expected to be realistic and simulating system performance using it should yield reliable results. Second order effects are expected due to α being a small positive constant instead of zero, and α, β being stochastic, possibly time-varying, quantities with associated uncertainties, albeit small.

It is also worth noticing that implementation of the SLM will be performed with a DMD with a number of micromirrors typically higher than the number of micro-pixels to be imaged. This allows for binning micromirrors into blocks, which helps reduce the uncertainty on the values of α, β.

The following equation presents an example with a super-resolution factor of 4 and a compression ratio of 75%, i.e., three measurements are serially acquired by a macro-pixel corresponding to four micro-pixels.

$$\begin{bmatrix} y_{i1}^{P} \\ y_{i2}^{P} \\ y_{i3}^{P} \end{bmatrix} = \begin{bmatrix} \alpha & \alpha & \beta & \beta \\ \beta & \beta & \alpha & \beta \\ \alpha & \beta & \beta & \alpha \end{bmatrix} \begin{bmatrix} x_{i1}^{P} \\ x_{i2}^{P} \\ x_{i3}^{P} \\ x_{i4}^{P} \end{bmatrix} \tag{11}$$

However, it is important to consider a global image model in order to effectively exploit spatial correlation in the image reconstruction algorithms. The measurements acquired by means of whiskbroom scanning of a scene with the proposed instrument can be modelled as:

$$y_i = \Phi_i x_i \tag{12}$$

where the sensing matrix Φ_i is now a sparse block-diagonal matrix. The following example considers a scene with 4 spatial locations, each composed of 4 micro-pixels, a compression ratio of 75% and a super-resolution factor of 4.

$$\begin{bmatrix} y_{i1}^{P1} \\ y_{i2}^{P1} \\ y_{i1}^{P1} \\ y_{i2}^{P1} \\ y_{i1}^{P2} \\ y_{i2}^{P2} \\ y_{i1}^{P2} \\ y_{i2}^{P2} \\ y_{i1}^{P3} \\ y_{i2}^{P3} \\ y_{i3}^{P3} \\ y_{i1}^{P4} \\ y_{i2}^{P4} \\ y_{i3}^{P4} \end{bmatrix} = \begin{bmatrix} \alpha & \beta & \alpha & \beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta & \beta & \alpha & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha & \beta & \alpha & \beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha & \alpha & \beta & \beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta & \beta & \alpha & \beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha & \beta & \beta & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta & \alpha & \beta & \alpha & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha & \beta & \alpha & \beta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha & \alpha & \beta & \beta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha & \beta & \beta & \alpha \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta & \alpha & \alpha & \beta \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha & \beta & \alpha & \beta \end{bmatrix} \begin{bmatrix} x_{i1}^{P1} \\ x_{i2}^{P1} \\ x_{i3}^{P1} \\ x_{i4}^{P1} \\ x_{i1}^{P2} \\ x_{i2}^{P2} \\ x_{i3}^{P2} \\ x_{i4}^{P2} \\ x_{i1}^{P3} \\ x_{i2}^{P3} \\ x_{i3}^{P3} \\ x_{i4}^{P3} \\ x_{i1}^{P4} \\ x_{i2}^{P4} \\ x_{i3}^{P4} \\ x_{i4}^{P4} \end{bmatrix} \tag{13}$$

Notice that the SLM sensing mask is changed for every imaged spatial location.

4.2. Full acquisition model

The overall instrument can be modelled as a linear system, acquiring measurements of the scene. In this full model, we take into account the effects of the optics as well as the noise introduced by the detectors. For a single spatial location, the model can be written as:

$$y_{ij}^{P} = \mathbf{1} H \phi_{ij}^{P} x_i^{P} + \eta_{ij}^{P} \tag{14}$$

for the j-th measurement, where H is the linear operator modelling the effect of the point spread function (PSF) of the optical system, $\phi_{ij}^{P} = diag(\Phi_{ij}^{\mathrm{p}})$ is a diagonal matrix whose entries are the entries of the j-th row of Φ_i^{p} and $\mathbf{1}$ is a row vector of ones that integrates the result onto a single measurement.

This model now accounts for both the spatial modulation introduced by the SLM and the effects of the optical system, which allows to provide a more accurate reconstruction of the image. Moreover, the term η_i^P accounts for sensor noise, including thermal noise, shot noise, readout noise.

Finally, an even more complex model can take into account distortion effects due to optical elements (e.g., the dichroic filter and spectral filter). In this model, the matrix H would be a product of different degradation operators:

$$H = H_1 H_2 \dots H_R$$

where H_R represents the PSF, and each of the other matrices H_i models a specific type of degradation.

5. RESULTS

The performance of the SURPRISE instrument has been evaluated by running preliminary experiment on both natural images and simulated hyperspectral images. In all cases, the reconstruction quality achieved by the algorithms has been evaluated using the PSNR, which is a common metric adopted to evaluate the quality of a reconstructed image respect the original one. This metric is defined as the ratio between the squared maximum possible value of a signal and the power of the noise that affects the fidelity of its representation after sensing/reconstruction. It is usually expressed in terms of the logarithmic decibel scale (dB). Typical values of PSRN are in a range of 20 – 40 dB. Moreover, being expressed in terms of the logarithmic decibel scale, an improvement of 0.25 dB can be considered significant. In our specific case, after evaluating each test image individually by calculating the PSNR between the original and the reconstructed image, the average PSNR will be calculated and reported in appropriate tables.

ISTA-Net+ has been trained on a training set of 431 natural pictures, while for TV this phase is not necessary. It is noted that, while these images are not representative of an Earth observation scenario, it is possible to retrain ISTA-Net for a specific Earth observation dataset. In practice, one can keep the baseline ISTA-Net network as a pre-trained network and fine-tune its parameters on a new dataset. For the training, we used a learning rate of 0.001 with the Adam optimizer. The neural network has been trained for 200 epochs.

For both algorithms, a binary matrix with entries $0 – 1/B$ has been used as sensing matrix ϕ, where the dimension of the sensing mask is BxB. Dividing by B is performed in order to normalize ϕ and the same sensing matrix is used also in TVAL to have a fair comparison. In all experiments, the sensing mask will have a size of 32×32. The sensing matrix is obtained starting from a binary mask with the same size as the sensing mask. Specifically, we generate a binary random matrix of dimensions 32×32 from which we extract each time the groups of micro-pixels we are interested in. Then, we flatten the entries of a micropixel and insert them, columns-shifted, in different rows of the sensing matrix

ϕ. This allows us to obtain a ϕ which has, for every row, only few non-zero elements and hence a sparse structure, according to equation (13). When using macro-pixels having 2×2, 4×4 and 8×8 size it may happen that we are unable to extract enough measurements from a single 32×32 mask to fill the sensing matrix as just explained, especially when we want to obtain a large number of measurements M. In those cases, we will generate multiple masks to be able to get the right number of measurements. Finally, the sensing is independent for every spatial location, composed of multiple micro-pixels associated to on macro-pixel. However, for the reconstruction we consider macro-pixels with several spatial location to take advantage of the correlation of the image. Finally, in this stage we assumed no degradations from the optical systems, i.e., $H = I$ in equation (14).

5.1. Performance on natural images

Here, we compare the performance of both ISTA-Net and TVAL on a dataset composed of 11 natural images. The main goal is to understand how the algorithms performance varies with the number of macro-pixels contained in an image, starting from a 32×32 size mask, and the super-resolution factor. The number of macro-pixels in a 32×32 sensing mask changes as a function of the number of micro-pixels that compose it. In our case we have studied macro-pixels having 2×2, 4×4 and 8×8 size with several different masks for obtaining the number of measurements we are interested in. The results will be shown for different total number of measurements M that are needed for each configuration.
The results are reported in Tables 1-4 and show that ISTA-Net sensibly outperforms TVAL for all compression ratios. Moreover, the gain of ISTA-Net over TVAL is usually larger for compression rates in the range 25%-50%.

M=64		
Micro-Pixels & Mask	**ISTA-Net+ Average PSNR**	**TVAL Average PSNR**
4×4, 1 mask	23.06	20.08
8×8, 4 masks	23.18	20.19

Table 1: Average PSNR with a CS ratio of the 6.25%

M=128		
Micro-Pixels & Mask	**ISTA-Net+ Average PSNR**	**TVAL Average PSNR**
4×4, 2 masks	25.43	21.25
8×8, 8 masks	25.97	22.83

Table 2: Average PSNR with a CS ratio of the 12.50%

M=256		
Micro-Pixels & Mask	**ISTA-Net+ Average PSNR**	**TVAL Average PSNR**
2×2, 1 mask	29.22	23.56
4×4, 4 masks	29.11	23.95
8×8, 16 masks	29.93	25.80

Table 3: Average PSNR with a CS ratio of the 25%

M=512		
Micro-Pixels & Mask	**ISTA-Net+ Average PSNR**	**TVAL Average PSNR**
2×2, 2 masks	33.38	26.40
4×4, 8 masks	34.12	27.31
8×8, 32 masks	34.88	29.58

Table 4: Average PSNR with a CS ratio of the 50%

5.2. Simulation of SURPRISE Instrument

For this simulation, we have 8 images which have been simulated in MATLAB by IFAC-CNR for testing the 4×4 super-resolution. They have the following features: 388 rows, 28 columns, 10 bands. Moreover, the spatial resolution is 250 m and the spectral one is 40 nm. Furthermore, the original pictures have been acquired with an along track dimension bigger than the across one.

Image	**Average PSNR**
Argentario_020712	24.20
Argentario_020914	27.63
Firenze_021219	32.80
Firenze_100714	26.54
SanRossore_120810	31.99
Umbria100325	30.56
Venezia_010607	30.77
Venezia_010607	29.67

Table 5: Average PSNR for Instrument Simulation.

Since ISTA-Net+ has been trained to work with pictures having two dimensions, for every simulated picture, we decided to separate the 10 bands and process the obtained images independently. Regarding the performance evaluation, the results are shown in Table 5 and they represent the average PSNR measured on all 10 bands.
We can observe that the reported results are quite heterogeneous. In three cases, the performances are lower than those achieved previously with the same configuration, but on totally different images such as natural ones, probably due to the presence of numerous details for all the bands.

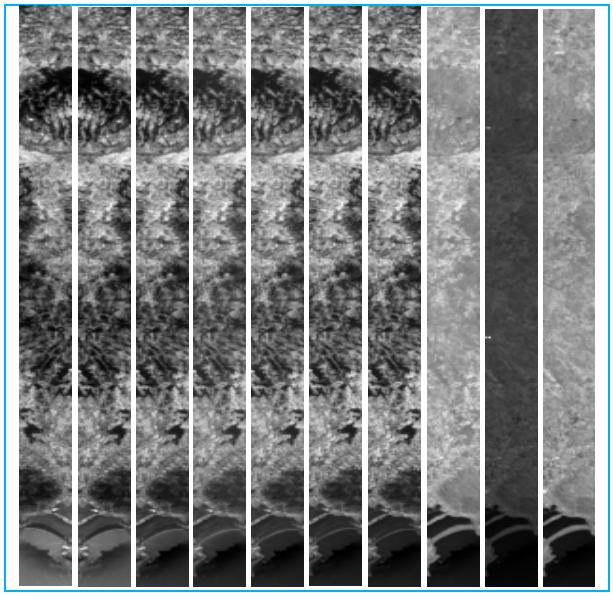

Figure 2: Original bands image for Argentario_020712.

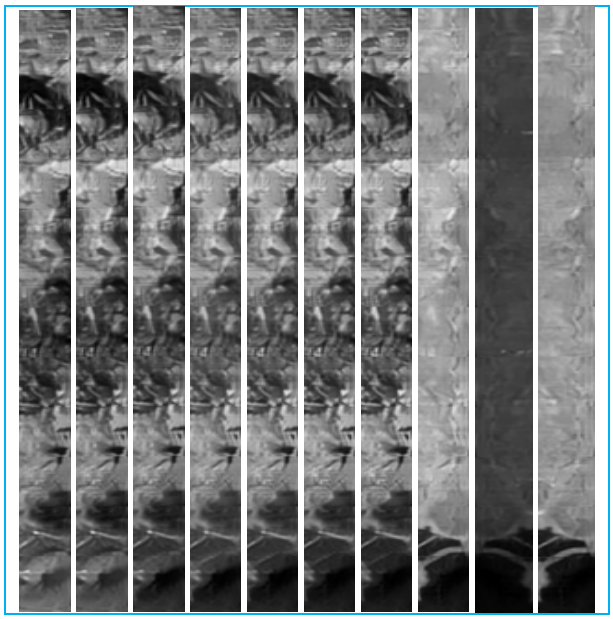

Figure 3: Reconstructed bands image for Argentario_020712.

In the remaining cases, however, the performance is better than in the case of natural images, this is probably due to the fact that the images are very sparse and with few details. We report in Figure 2 the 10 original bands pictures for the image Argentario_020712 and in Figure 3 the reconstructed ones.

6. CONCLUSIONS

This paper has addressed several key problems in CS reconstruction as applied to computational multispectral imaging systems. The results show that methods based on deep learning are better than conventional methods for CS reconstruction, as they achieve better reconstruction accuracy. Moreover, such methods are relatively robust to the partial lack of knowledge of the system parameters, as well as robust to the scene content. Among all possible configurations and super-resolution factors, the best results seem to be obtained when high super-resolution factors are sought. In this case, the image is comprised by fewer macroblocks, and a larger number of measurements are acquired for each macroblock. It has to be noted that, given the sequential nature of the CS process, this configuration tends to require a longer acquisition time for a given scene. A possible drawback of those method is the training time and the need of a training set. However, the required training set is not necessarily specific to the sensor under analysis, and a rather generic set of natural images will suffice. Moreover, since the neural network is run at the ground segment, where the reconstruction takes place, the complexity is not a significant hurdle.

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] R. Vitulli, P. Armbruster and D. Merodio-Codinachs, "Big data starts on-board," *Proc. of BiDS 2016,* 2016.

[2] E. J. Candes e M. B. Wakin, «An introduction to compressive sampling,» *IEEE signal processing magazine,* vol. 25, n. 2, p. 21–30, 2008.

[3] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly e R. G. Baraniuk, «Single-pixel imaging via compressive sampling,» *IEEE signal processing magazine,* vol. 25, n. 2, pp. 83-91, 2008.

[4] M. A. Davenport, P. T. Boufounos, M. B. Wakin e R. G. Baraniuk, «Signal processing with compressive measurements,» *IEEE Journal of Selected Topics in Signal Processing,* vol. 4, n. 2, pp. 445-460, 2010.

[5] T. Bianchi, V. Bioglio e E. Magli, «Analysis of one-time random projections for privacy preserving compressed sensing,» *IEEE Transactions on Information Forensics and Security,* vol. 11, n. 2, pp. 313-327, 2016.

[6] E. Candes e T. Tao, «Decoding by Linear Programming,» *IEEE Transactions on Information Theory,* vol. 51, n. 12, pp. 4203-4215, 2005.

[7] S. F. Cotter e B. D. Rao, «Sparse channel estimation via matching pursuit with application to equalization,» *IEEE Transactions on Communications,* n. 50, pp. 374-7, 2002.

[8] J. Wang, S. Kwon, P. Li e B. Shim, «Recovery of sparse signals via generalized orthogonal matching pursuit: a new analysis,» *IEEE Transaction on Signal Processing,* n. 64, pp. 1076-1089, 2016.

[9] D. Needell e J. Tropp, «CoSaMP: iterative signal recovery from incomplete and inaccurate samples,» *Communications of the ACM,* n. 12, pp. 93-100, 2010.

[10] M. Sajjad, I. Mehmood e S. W. Baik, «Sparse coded image super-resolution using K-SVD trained dictionary based on regularized orthogonal matching pursuit,» *Bio-Medical Materials and Engineering,* n. 26, pp. 1399-1407, 2015.

[11] W. Dai e O. Milenkovic, «Subspace pursuit for compressive sensing signal reconstruction,» *IEEE Transactions on Information Theory,* n. 5, pp. 2230-2249, 2009.

[12] D. L. Donoho, A. Maleki e A. Montanari, «Message Passing Algorithms for Compressed Sensing,» *PNAS,* n. 106, 2009.

[13] R. Berinde e P. Indyk, «Sequential Sparse Matching Pursuit,» in *47th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, 2009.

[14] Y. Kabashima, «A CDMA multiuser detection algorithm on the basis of belief propagation,» *Journal of Physics A: Mathematical and General,* vol. 36, pp. 11111-21, 2003.

[15] A. Chambolle, R. A. DeVore, N. Lee e B. J. Lucier, «Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage,» *IEEE Transaction on Image Processing,* vol. 7, n. 3, pp. 319-35, 1998.

[16] A. Beck e M. Teboulle, «A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems,» *SIAM,* vol. 2, n. 1, p. 183–202, 2009.

[17] T. Goldstein e C. Studer, «Phasemax: convex phase retrieval via basis pursuit,» *IEEE Transaction on Information Theory,* n. 4, pp. 2675-2689, 2018.

[18] R. Tibshirani, «1 Regression Shrinkage and selection via the Lasso,» *Journal of the Royal Statistical Society,* vol. 73, pp. 267-288, 2011.

[19] M. Zibulevsky e M. Elad, «L1-L2 optimization in signal and image processing,» *IEEE Signal Process Magazine,* n. 27, pp. 76-88, 2010.

[20] K. Cao, X. Zhou e Y. Cheng, «7 Improved focal underdetermined system solver method for radar coincidence imaging with model mismatch,» *Journal of Electronic Imaging,* n. 26, 2017.

[21] R. Zhao, X. Lai, X. Hong e Z. Lin, «A matrix-based IRLS algorithm for the least Lp-norm design of 2-d firfilters,» *Multidimensional Systems and Signal Processing,* n. 2, pp. 1-15, 2017.

[22] G. Oliveri, M. Salucci, N. Anselmi e A. Massa, «Compressive sensing as applied to inverse problems for imaging: theory applications, current trends, and open challenges,» *IEEE Antennas and Propagation Magazine,* n. 59, pp. 34-46, 2017.

[23] Y. Wang, J. Yang, W. Yin e Y. Zhang, «A new alternating minimization algorithm for total variation image reconstruction,» *SIAM Journal on Imaging Sciences,* vol. 1, n. 3, pp. 248-272, 2008.

[24] A. Mousavi, A. B. Patel e R. G. Baraniuk, «A deep learning approach to structured signal recovery,» in *Annual Allerton Conference on Communication, Control, and Computing*, 2015.

[25] A. Adler, D. Boublil, M. Elad e M. Zibulevsky, «A deep learning approach to blockbased compressed sensing of images,» in *ICASSP*, 2017.

[26] M. Iliadis, L. Spinoulas e A. K. Katsaggelos, «Deep fully-connected networks for video compressive,» *Digital Signal Processing,* vol. 72, pp. 9-18, 2018.

[27] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche e A. Ashok, «ReconNet: Non-iterative reconstruction of images from compressively sensed measurements,» in *CVPR*, 2016.

[28] A. Mousavi e R. G. Baraniuk, «Learning to invert: Signal recovery via deep convolutional networks,» in *ICASSP*, 2017.

[29] J. Zhang e 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.

[30] W. Shi, F. Jiang, S. Liu e D. Zhao, «Scalable Convolutional Neural Network for Image Compressed Sensing,» in *CVPR*, Long Beach, 2019.