Injecting Image Priors into Learnable Compressive Subsampling

Martino Ferrari†, Olga Taran†, Taras Holotyak†, Karen Egiazarian‡ and Slava Voloshynovskiy†*
†University of Geneva, Geneva, Switzerland
Email: martino.ferrari@etu.unige.ch, {olga.taran,taras.holotyak,svolos}@unige.ch
‡Tampere University of Technology, Tampere, Finalnd
Email: karen.egiazarian@tut.fi

***Abstract*—Many medical (computerized tomography, magnetic resonance imaging) and astronomy imaging problems (Square Kilometre Array), spectroscopy and Fourier optics attempt at reconstructing high quality images in the pixel domain from a limited number of samples in the frequency domain. In this paper, we extend the problem formulation of learnable compressive subsampling [1] that focuses on the learning of the best sampling operator in the Fourier domain adapted to spectral properties of training set of images. We formulate the problem as a reconstruction from a finite number of sparse samples with a prior learned from the external dataset or learned on-fly for the image to be reconstructed. The proposed methods are tested on diverse datasets covering facial images, medical and multi-band astronomical applications using the mean square error and SSIM as a perceptual measure of reconstruction. The obtained results demonstrate some interesting properties of proposed methods that might be of interest for future research and extensions.**

***Index Terms*—Compressive sensing, learnable compressive subsampling, support learning, reconstruction, deep priors.**

I. Introduction

The problem of recovering a high dimension signal $\mathbf{x} \in \mathbb{R}^N$ from its low dimensional observation $\mathbf{a} = \mathbf{W}\mathbf{x} + \mathbf{e}$ with $\mathbf{a} \in \mathbb{R}^n$ and $n << N$ corrupted by noise $\mathbf{e}$ is of great importance for many imaging applications. In one of the most popular formulation known as compressive sampling, it has received a tremendous attention in many publications [2]. In such a formulation, the signal $\mathbf{x}$ is assumed to possess some structure that can be expressed either via sparsity in some orthonormal or overcomplete (shallow) representations or via some properties of latent space in more recent deep representations [3]. In the classical compressive sampling, the measurement matrix $\mathbf{W} \in \mathbb{C}^{n \times N}$ is generally generated randomly, that might be very inefficient for practical applications in terms of storage and computations. The recent work [2] extends the classical formulation of compressive sensing using the sub-sampling structured matrices of the form $\mathbf{W} = \mathbf{P}_\Omega \mathbf{\Psi}$, with an orthonormal operator $\mathbf{\Psi} \in \mathbb{C}^{N \times N}$ and sampling operator $\mathbf{P}_\Omega : \mathbb{C}^N \rightarrow \mathbb{C}^n$. The main idea behind the proposed extension consists in learning of sampling operator $\mathbf{P}_\Omega$ on the training set of images to minimize the average reconstruction error as a cost function. It should be pointed out that due to the physical imaging constraints there is no freedom in the selection of the operator $\mathbf{\Psi}$, whereas the only possible adaptation to data is via the operator $\mathbf{P}_\Omega$. The solution to the above adaptation problem leads to a natural conclusion that the highest sampling rate of the operator $\mathbf{P}_\Omega$ should be concentrated in the region of high energetic image components. In turns, it suggests that the optimal sampling geometry $\mathbf{P}_\Omega$ computed on average for a set of natural imaging should be in the region of low frequencies possessing the highest energy. Additionally, in contrast to the compressive sensing based on non-linear reconstruction algorithms, the authors in [1] consider a linear decoder of form $\hat{\mathbf{x}} = \mathbf{\Psi}^* \mathbf{P}_\Omega^T \mathbf{a}$, where * denotes the complex conjugate.

In some previous works, a problem of optimal on-fly construction of a sampling operator $\mathbf{P}_\Omega$ with an iterative reconstruction was considered in [4], [5]. No training on the external dataset was assumed whereas the process of adaption was based on a fact that many natural images possess dominating energy frequency components located along some spatial frequency directions. A small number of samples in $\mathbf{P}_\Omega$ were used to estimate these directions, mainly in low frequency part of spectrum, and the remaining budget of sampling components was adapted accordingly. In this case, the sampling operator was not trained on average as above but was adapted to the properties of each image. One can consider this approach as an analogue of k-best selection with the given number of sampling components to be k.

Being intuitive, the above approaches do not take into account more powerful priors about the image statistics that can be learned by modern machine learning methods.

In this respect, our paper extends the previous works in the following ways:

1) given a training set of images, we train unsupervised models in several sub-bands in the frequency domain and ensure the correct classification of parameters of these models just from several samples, thus reducing the sampling rate and enabling the reconstruction of details in images from the training models;
2) we design an optimized partitioning onto sub-bands' and sampling rate for each sub-band;
3) we consider an alternative algorithm based on recently proposed deep image prior model [3] that does not require any pre-training and a regularized reconstruction

This work was partially supported by the SNF project No. 200021-165672.
"*" - the corresponding author.

ISBN 978-90-827970-1-5

is performed via imposing constrains on parameters of deep network;

4) we also investigate the impact of alignment quality (accuracy of prior) for the trained models.

Notations. We use capital bold letters, $\mathbf{W} \in \mathbb{C}^{n\times N}$, to denote real or complex valued matrices, small bold letters, $\mathbf{x} \in \mathbb{R}^N$, to denote real or complex valued vectors, capital letters, $\mathrm{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \in \mathbb{R}^{N\times m}$, to denote a set of real or complex valued vectors. The sign $*$ denotes the complex conjugate. The estimate of $\mathbf{x}$ is denoted as $\hat{\mathbf{x}}$. $\mathbf{c}_{re}$ and $\mathbf{c}_{im}$ denote respectively the real and imaginary parts of the complex vector, $\mathbf{c} = \mathbf{c}_{re} + i\mathbf{c}_{im} \in \mathbb{C}^M$. All vectors have finite length, explicitly defined wherever appropriate.

II. Proposed approaches

In this paper we will consider *Learnable Compressive Sub-sampling based on Multi-Band Coding* (section **II-A**) and *Learnable Compressive Sub-sampling with Reconstruction based on Deep Image Prior* (section **II-B**) that can be applied depending on the available priors and sampling data. Our estimations are obtained in the following form:

$$\hat{\mathbf{x}} = \mathbf{\Psi}^* \left(\mathbf{P}_{\Omega}^T \mathbf{a} + \mathbf{P}_{\Omega^C}^T \mathbf{P}_{\Omega^C} \mathbf{\Psi} \mathbf{b} \right), \tag{1}$$

where $\mathbf{a}$ is the sampled observation in the transform domain, Ω^C is the complementary support set of Ω and $\mathbf{b}$ is the prior information obtained from the training set.

A. Learnable Compressive Sub-sampling based on Multi-Band Coding

The main idea is simple, given a set of m well aligned training signals $\mathrm{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \in \mathbb{R}^{N\times m}$, we generate **(a)** a codebook capable to well represent new signals and **(b)** a sub-sampling scheme that maximizes the correct code identification within each codebook.

The training algorithm is illustrated in Fig. 1. The training is split into four main stages: **(1)** change of basis and vectorization; **(2)** sub-bands partition; **(3)** codebook generation for each sub-band and **(4)** selection of the sub-sampling operators for each sub-band.

1) Change of basis and vectorization: Due to the fact that very often in real cases the signal acquisition is done not in the signal but in some transform domain, first of all we transform the training signals $\mathrm{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \in \mathbb{R}^{N\times m}$ from the signal domain to the transform domain using the operator $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$. Additionally, since the transform domain is supposed to be complex, for simplification, we split the resulting signals into *real* and *imaginary* parts:

$$\mathbf{\Psi}\mathrm{X} = \mathbf{\Psi}_{re}\mathrm{X} + i\mathbf{\Psi}_{im}\mathrm{X}, \tag{2}$$

where $\mathbf{\Psi}_{re}$ and $\mathbf{\Psi}_{im}$ represent real and imaginary parts, respectively.

In our analysis we use 2D Discrete Fourier Transform (2D-DFT) to obtain the transform representation of a set of 2D real images, i.e., the training set. In the transform domain, we perform a vectorization of each obtained signal in a zig-zag manner [6] with subsequent splitting into *real* and *imaginary* parts.

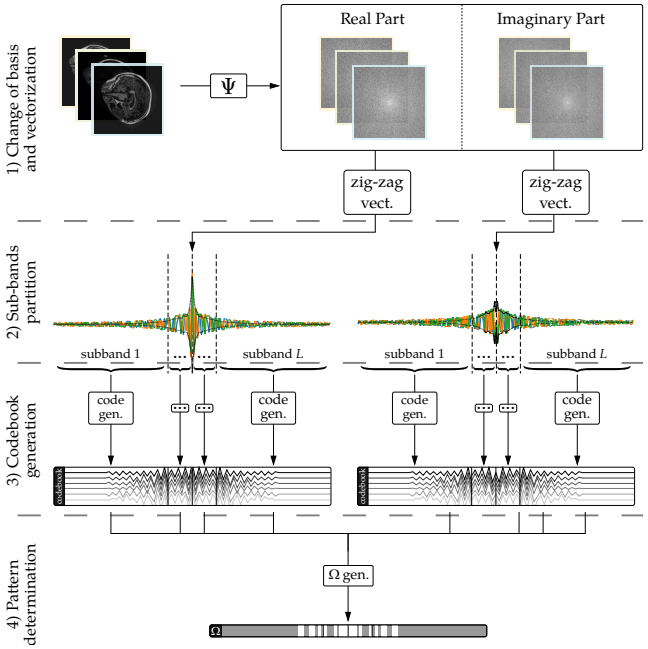

Fig. 1. Training process: from the training set we generate sub-bands in the transform domain, train codebooks and sampling operators for each sub-band.

2) Sub-Bands partition: Dividing the transformed signals into sub-bands reduces the dependency on the size of the training set, improves the robustness to geometrical alignment, increases the reconstruction quality and avoids overfitting.

Taking into account that in the transform domain the signals have not equal energy in each sub-band, the size and position of the sub-bands have a big impact on the final results. For such a reason we choose to adapt the sub-bands division to the structure of the transformed signals. In particular, we divide each transformed signals into L sub-bands in such a way that each sub-band contains on average the same amount of energy:

$$\bar{e}(j) = \frac{1}{m}\sqrt{\sum_{i=1}^{m} |\mathbf{\Psi}\mathbf{x}_i(j)|^2}, \tag{3}$$

where $j \in \{-J, ..., J\}$ denotes index of discrete samples in the DFT domain, $|.|$ denotes the magnitude of DFT, $\mathbf{\Psi}\mathbf{x}_i(j)$ is the value of the transformed i-th training signal at the frequency j.

Once the energy is estimated per frequency sample, it is possible to cumulate it as:

$$c(t) = \sum_{j=-J}^{t} \bar{e}(j). \tag{4}$$

Based on the obtained cumulated energy, we perform the splitting into L equal-energy sub-bands (see Fig. 2) and we mark the l-th sub-band of the training set as:

$$(\mathbf{\Psi}\mathrm{X})^l = \{(\mathbf{\Psi}\mathbf{x}_1)^l, \ldots, (\mathbf{\Psi}\mathbf{x}_m)^l\}.$$

ISBN 978-90-827970-1-5 © EURASIP 2018

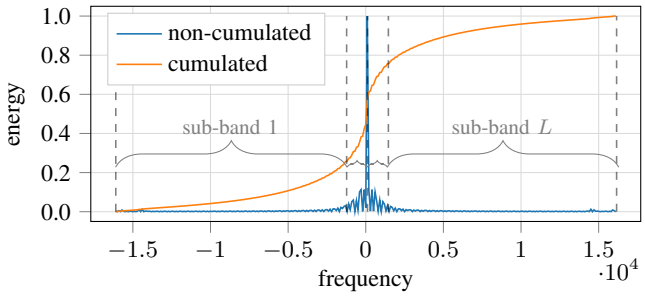

Fig. 2. Flatten and cumulated energy with bands' division accordingly to (4).

3) Codebook generation: We perform codebook generation for each of the L sub-bands. We generate the codebooks simply using the *K-means* algorithm. However, instead of generating a single codebook of complex type, we generate two independent codebooks $\mathcal{C}^l_{re}$ and $\mathcal{C}^l_{im}$ for the real and imaginary parts, respectively.

4) Sub-band sub-sampling: The final step of the training process is a generation of the sampling scheme per sub-band. Given the codebooks, $\mathcal{C}^l_{re}$ and $\mathcal{C}^l_{im}$, for the l-th sub-band, we seek the index set Ω_l that maximizes the correct codewords identification in each sub-band. Due to the fact that if the distance between pairs of codewords in the selected Ω_l is close to 0 then it is impossible to distinguish between them and the selection will be random. At the same time, if the distance between each pair in the selected Ω_l is sufficiently big then the probability of correct selection is close to 1. For this reason we seek such a sampling pattern that maximizes the overall distance between each pair of codewords in the codebooks. This can be expressed in the next form:

$$\begin{aligned}\Omega_l = \arg\max_{\Omega_l} & \sum_{\mathbf{c}_a,\mathbf{c}_b \in \mathcal{C}^l_{re}} \|\mathbf{P}_{\Omega_l}\mathbf{c}_a - \mathbf{P}_{\Omega_l}\mathbf{c}_b\|_2^2 + \\ & \sum_{\mathbf{c}_c,\mathbf{c}_d \in \mathcal{C}^l_{im}} \|\mathbf{P}_{\Omega_l}\mathbf{c}_c - \mathbf{P}_{\Omega_l}\mathbf{c}_d\|_2^2,\end{aligned} \tag{5}$$

where $\mathbf{c}_a$ and $\mathbf{c}_b$ are all pairs of codewords from the codebook $\mathcal{C}^l_{re}$ and $\mathbf{c}_b$ and $\mathbf{c}_d$ are the corresponding pairs of codewords from the codebook $\mathcal{C}^l_{im}$.

Since for a large number of codewords this operation is quite complex, of order $\mathcal{O}(n^2)$, therefore it may be convenient to approximate it as:

$$\Omega_l = \arg\max_{\Omega_l} \left(\mathbf{P}_{\Omega_l}\sigma_{\mathcal{C}^l_{re}} + \mathbf{P}_{\Omega_l}\sigma_{\mathcal{C}^l_{im}}\right), \tag{6}$$

where $\sigma_{\mathcal{C}^l_{re}}$ and $\sigma_{\mathcal{C}^l_{im}}$ are the frequency wise standard deviation of the real and imaginary codebooks, respectively.

5) Encoding and Decoding: The *encoding* of a new signal $\mathbf{x} \in \mathbb{R}^N$ can be performed as described below. The first step is the sub-sampling per sub-band:

$$\mathbf{a}^l = \mathbf{P}_{\Omega_l}(\mathbf{\Psi}\mathbf{x})^l. \tag{7}$$

Then a pair of codewords for real and imaginary parts is assigned to each sub-sample $\mathbf{a}^l$ from the respective codebooks $\mathcal{C}^l_{re}$ and $\mathcal{C}^l_{im}$ via:

$$\begin{aligned}\mathbf{c}^l_{re} &= \arg\min_{\mathbf{c}^l_{re}} \|\mathrm{Re}(\mathbf{a}^l) - \mathbf{P}_{\Omega_l}\mathbf{c}^l_{re}\|_2^2 \\ \mathbf{c}^l_{im} &= \arg\min_{\mathbf{c}^l_{im}} \|\mathrm{Im}(\mathbf{a}^l) - \mathbf{P}_{\Omega_l}\mathbf{c}^l_{im}\|_2^2,\end{aligned} \tag{8}$$

where $\mathrm{Re}(\mathbf{a}^l)$ and $\mathrm{Im}(\mathbf{a}^l)$ denote the real and imaginary part, respectively.

To improve both identification rate and reconstruction it is possible to include the magnitude and phase error into the minimization problem:

$$\begin{aligned}\mathbf{c}^l_{re}, \mathbf{c}^l_{im} = \arg\min_{\mathbf{c}^l_{re},\mathbf{c}^l_{im}} & \|\mathrm{Re}(\mathbf{a}^l) - \mathbf{P}_{\Omega_l}\mathbf{c}^l_{re}\|_2^2 + \\ & \|\mathrm{Im}(\mathbf{a}^l) - \mathbf{P}_{\Omega_l}\mathbf{c}^l_{im}\|_2^2 + \\ & \||\mathbf{a}^l| - |\mathbf{P}_{\Omega_l}(\mathbf{c}^l_{re} + i\cdot\mathbf{c}^l_{im})|\|_2^2 + \\ & \|\arg(\mathbf{a}^l) - \arg(\mathbf{P}_{\Omega_l}(\mathbf{c}^l_{re} + i\cdot\mathbf{c}^l_{im}))\|_{2\pi},\end{aligned} \tag{9}$$

where $|.|$ and $\arg(.)$ denote respectively the magnitude and phase of a complex vector, and $\|.\|_{2\pi}$ is the mean angular difference normalized in the range $(0, 2\pi)$. The optimization of this formula will minimize the error not only between the real and the imaginary part, but also between phase and magnitude. However, solving this problem is computationally much more expensive ($\mathcal{O}(n^2)$) than the previous problem (8) ($\mathcal{O}(n)$). To reduce the computation complexity it is possible to apply the minimization problem to smaller subset of the codebook, using the formula (8) to select few pairs of codes from both datasets.

The reconstruction, or *decoding*, of the signal $\hat{\mathbf{x}}$ is computed by processing the output of the encoder as follows:

$$\hat{\mathbf{x}} = \mathbf{\Psi}^* \left\{ \begin{array}{c} (\mathbf{P}^T_{\Omega_1}\mathbf{a}^1 + \mathbf{P}^T_{\Omega_1^C}\mathbf{P}_{\Omega_1^C}(\mathbf{c}^1_{re} + i\mathbf{c}^1_{im})), \\ \dots, \\ (\mathbf{P}^T_{\Omega_L}\mathbf{a}^L + \mathbf{P}^T_{\Omega_L^C}\mathbf{P}_{\Omega_L^C}(\mathbf{c}^L_{re} + i\mathbf{c}^L_{im})) \end{array} \right\}, \tag{10}$$

where $\{.\}$ denotes the concatenation of the vectors and Ω_l^C is the complementary support set of Ω_l. The formula (10) can be expressed as well in the form of the equation (1), where $\Omega = \{\Omega_1, \dots \Omega_L\}$, Ω^C is its complementary set and $\mathbf{b} = \{\mathbf{c}^1_{re} + i\mathbf{c}^1_{im}, \dots, \mathbf{c}^L_{re} + i\mathbf{c}^L_{im}\}$.

B. Learnable Compressive Sub-sampling with Reconstruction based on Deep Image Prior

In a more general approach, one can consider the reconstruction problem as an inverse problem:

$$\begin{aligned}(\hat{\mathbf{z}}, \hat{\theta}) = \arg\min_{\mathbf{z},\theta} & \|\mathbf{a} - \mathbf{P}_\Omega \mathbf{\Psi} f_\theta(\mathbf{z})\|_2^2 + \\ & \alpha\Omega_z(\mathbf{z}) + \beta\Omega_\theta(\theta),\end{aligned} \tag{11}$$

where $f_\theta(\mathbf{z})$ is a parametrized non-linear regressor expressed in a form of the deep network with the parameters of layers' filters $\theta = (\mathbf{W}_1, \cdots, \mathbf{W}_L)$, $\Omega_z(.)$ and $\Omega_\theta(.)$ denote the regularization operators for the latent variable $\mathbf{z}$ and network parameters θ, respectively, α and β stand for Lagrangian multipliers. The reconstruction is given as $\hat{\mathbf{x}} = f_{\hat{\theta}}(\hat{\mathbf{z}})$.

ISBN 978-90-827970-1-5 © EURASIP 2018

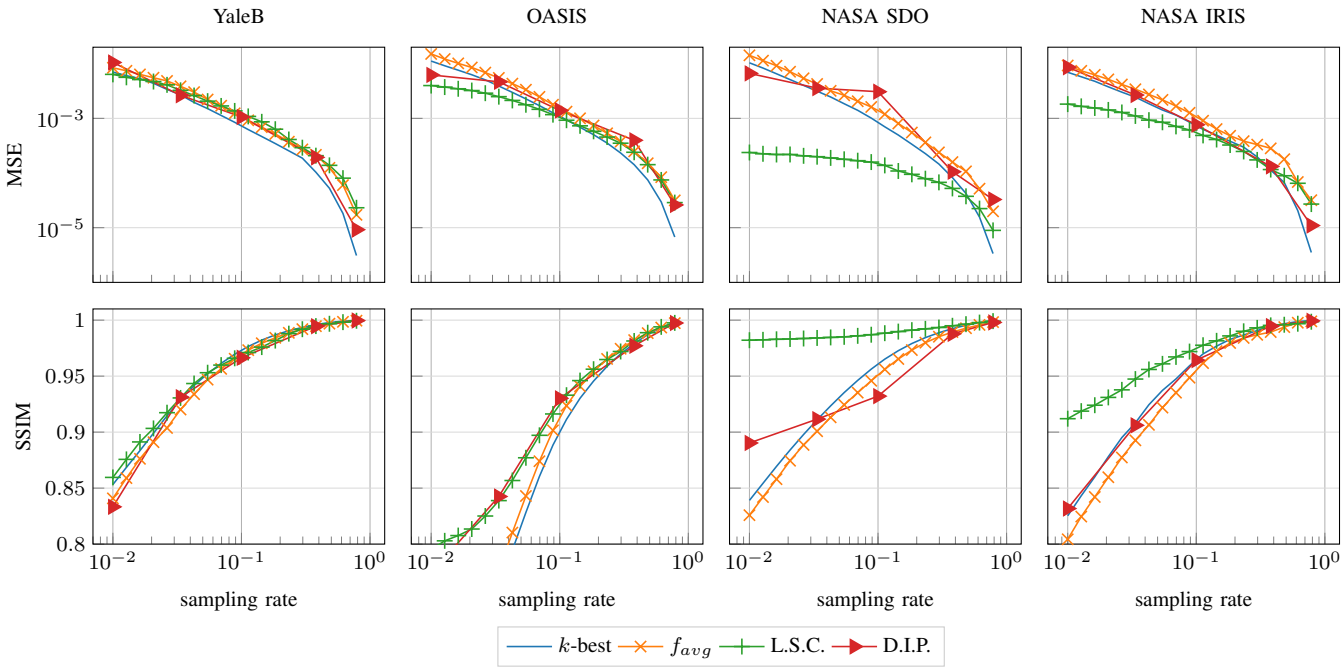

Fig. 3. MSE and SSIM at different sampling rate for the *YaleB*, aligned *OASIS*, *NASA SDO* and *NASA IRIS* datasets.

This can be achieved by an alternative minimization. However, it requires to be trained on the external training dataset. Recently Ulyanov *et. al.* [3] suggested a simplified version with a fixed $\mathbf{z}$ for all the data vectors:

$$\hat{\theta} = \arg\min_{\theta} \|\mathbf{a} - \mathbf{P}_\Omega \mathbf{\Psi} f_\theta(\mathbf{z})\|_2^2 + \beta\Omega_\theta(\theta), \tag{12}$$

where the optimization is only performed with respect to θ that essentially represents one step of the equation (11).

Prior injection: it is also possible to improve more the results by using the prior information of the image (e.g., a dataset mean value). The injection of the prior information can be done as follows:

$$\begin{aligned}\hat{\theta} = \arg\min_{\theta} &\|\mathbf{a} - \mathbf{P}_\Omega \mathbf{\Psi} f_\theta(\mathbf{z})\|_2^2 \\ &+ \alpha \|f_\theta(\mathbf{z}) - \mathbf{b}\|_2^2 + \beta\Omega_\theta(\theta),\end{aligned} \tag{13}$$

where $\mathbf{b}$ is the prior information extracted from the training dataset $\mathbf{X}$ and α stands for Lagrangian multiplier.

The *decoding* scheme is very simple:

$$\hat{\mathbf{x}} = \mathbf{\Psi}^*(\mathbf{P}_\Omega^T \mathbf{a} + \mathbf{P}_{\Omega^C}^T \mathbf{P}_{\Omega^C} \mathbf{\Psi} f_{\hat{\theta}}(\mathbf{z})), \tag{14}$$

where Ω^C is the complementary set of Ω and $f_{\hat{\theta}}(\mathbf{z})$ is the result of reconstruction based on the trained neural network.

III. Results

To evaluate the performance of the proposed approaches we use several databases, namely, *YaleB* Faces [7], *OASIS* MRIs (63$^{\text{th}}$ slice, raw and aligned) [8], a subset of 1500 images of size 256×256 pixels extracted from the *NASA Solar Dynamic Observatory* dataset [9] and 540 cropped images of size 256×256 pixels from the *NASA IRIS* dataset [10]. All used images were converted to gray-scale and the DFT was used as a transform $\mathbf{\Psi}$.

We compare the performance of the proposed approaches with the *adaptive sampling strategy* (further denoted as k-best) proposed in [11] and with the *learned average approximation* introduced in [1] (further denoted as f_{avg}).

Fig.3 shows the dependence of the reconstruction error on the sampling rate. As a measure of the reconstruction error we use the *Mean Square Error* (MSE) and *Structural SIMilarity* (SSIM) index [12].

The results obtained for the proposed *Learnable Compressive Sub-sampling based on Multi-Band Coding* (L.S.C.) and the *Learnable Compressive Sub-sampling with Reconstruction based on Deep Image Prior* (D.I.P.) as well as the results for the k-best and f_{avg} methods are illustrated for four used datasets. It is easy to see that in the case of the *YaleB* dataset all methods have approximately the same performance. At the same time, for both *NASA* datasets *L.S.C.* shows significant improvements especially for low sampling rates. As for *D.I.P.* it should be pointed out that, in general, its performance is quite close to the results obtained with the k-best and f_{avg} methods with small improvements at low rates.

For a more detailed analysis of the performance of the proposed methods the reconstructed images for different sampling rates are shown for each dataset in Fig. 4. To be noted that the visual quality of recovering the original image using only few samples (of the order of 10^{-3}) based on k-best or f_{avg} is very poor and it is quite difficult to recognize the general shape of the object. Increasing the sampling rate leads to improvement of the reconstruction quality and the level of details for all considered methods.

Additionally, it should be noted that the obtained results

ISBN 978-90-827970-1-5 © EURASIP 2018

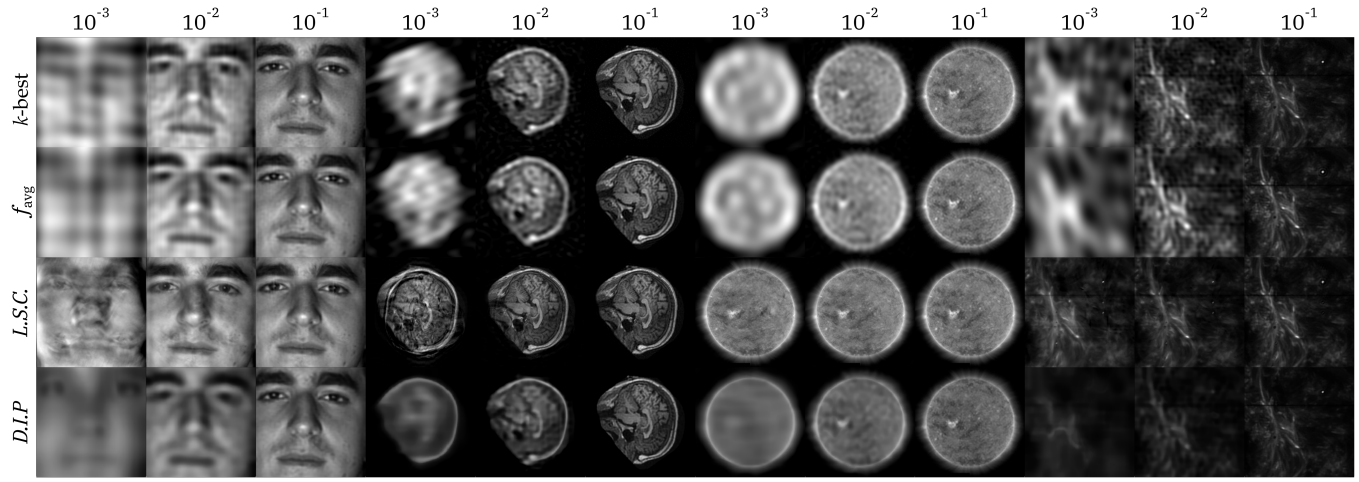

Fig. 4. Example of reconstruction for sampling rate varying from 10^{-3} to 10^{-1} for the datasets (from left to right): *YaleB*, *OASIS*, *NASA SDO* and *NASA IRIS*.

and performed analysis allow to affirm the higher risk of over-fitting of the *D.I.P.* method. This algorithm converges sufficiently quickly to the acquired image. However, as it can be seen in Fig. 4 its potential for image recovery with only few samples is larger than the one of the k-best and f_{avg} methods. For small sampling rates *D.I.P.* allows to capture very well low frequency components and contours.

As the main disadvantage of the proposed approaches, it should be mentioned that the proposed *Learnable Compressive Sub-sampling based on Multi-Band Coding* (L.S.C.) and the *Learnable Compressive Sub-sampling with Reconstruction based on Deep Image Prior* both require a good geometrical alignment between the training dataset and the image to be reconstructed. In the case of the *YaleB* Faces and *OASIS* MRIs datasets where the image alignment is not as good as for the considered *NASA* datasets, the obtained performance of the proposed approaches does not outperform the k-best and f_{avg}.

Conclusions

This paper addresses the problem of learning the best sampling operator in a transform domain adapted to spectral properties of the training set of images. We first formulated it as a reconstruction problem from a finite number of sparse samples with a prior learned either from the external dataset or on fly from the image to be reconstructed. We then proposed two methods that have both low training complexity and, very low encoding and decoding complexity of the signal of interest. Moreover, one of the proposed approaches, based on the deep image prior extension, represents a novel approach in deep image priors and does not require any training on the external datasets besides the mean vector. We evaluated the performance of the proposed approaches on four datasets and compared their efficiency with two well known methods in this domain. The proposed *Learnable Compressive Sub-sampling based on Multi-Band Coding* shows significant improvements in the reconstruction from only few samples on the considered *NASA* datasets and has approximately equal performance compared with the reviewed state-of-the-art methods on the other used datasets.

Acknowledgment

Our research work uses the MRI dataset from OASIS. The OASIS project is supported in part by grants P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584.

References

[1] L. Baldassarre, Y.-H. Li, J. Scarlett, B. Gözcü, I. Bogunovic, and V. Cevher, "Learning-based compressive subsampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 809–822, 2016.

[2] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Birkhäuser Basel, 2013, vol. 1, no. 3.

[3] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *arXiv preprint arXiv:1711.10925*, 2017.

[4] I. Prudyus, S. Voloshynovskiy, and T. Holotyak, "Robust image restoration matched with adaptive aperture formation in radar imaging systems with sparse antenna arrays," in *Signal Processing Conference (EUSIPCO 1998), 9th European*, Sept 1998, pp. 1–4.

[5] ——, "Adaptive aperture formation matched with radiometry image spatial spectrum," in *IEEE International Microwave and Radar Conference*, Krakow, Poland, 1998.

[6] G. K. Wallace, "The jpeg still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, Apr. 1991. [Online]. Available: http://doi.acm.org/10.1145/103085.103089

[7] A. Georghiades, P. Belhumeur, and Kriegman, "The yale face database b," 2001.

[8] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults," *Journal of cognitive neuroscience*, vol. 22, no. 12, pp. 2677–2684, 2010.

[9] W. D. Pesnell, B. J. Thompson, and P. C. Chamberlin, "The solar dynamics observatory (SDO)," vol. 275, no. 1, pp. 3–15. [Online]. Available: https://doi.org/10.1007/s11207-011-9841-3

[10] "Nasa interface region imaging spectrograph," http://iris.lmsal.com/search.

[11] B. Bah, A. Sadeghian, and V. Cevher, "Energy-aware adaptive bi-lipschitz embeddings," *CoRR*, vol. abs/1307.3457, 2013. [Online]. Available: http://arxiv.org/abs/1307.3457

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.