Deep Post-Processing for Sparse Image Deconvolution

Matthieu Terris*, Abdullah Abdulaziz*, Arwa Dabbech*, Ming Jiang‡, Audrey Repetti*†, Jean-Christophe Pesquet§ and Yves Wiaux*
*Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK
†AMS Department, Heriot-Watt University, Edinburgh EH14 4AS, UK
‡Signal Processing Laboratory 5 (LTS5), École Polytechnique Fédérale de Lausanne, CH-1015, Lausanne, Switzerland
§CentraleSupélec, Inria, Université Paris-Saclay, Center for Visual Computing, 91190 Gif sur Yvette, France

***Abstract*—Variational-based methods are the state-of-the-art in sparse image deconvolution. Yet, this class of methods might not scale to large dimensions of interest in current high resolution imaging applications. To overcome this limitation, we propose to solve the sparse deconvolution problem through a two-step approach consisting in first solving (approximately and fast) an optimization problem followed by a neural network for "Deep Post Processing" (DPP). We illustrate our method in radio astronomy, where algorithms scalability is paramount due to the extreme data dimensions. First results suggest that DPP is able to achieve similar quality to state-of-the-art methods in a fraction of the time.**

I. Introduction

We consider the following deconvolution problem

$$\boldsymbol{y} = \boldsymbol{k} * \boldsymbol{x} + \mathbf{e}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ is the observed image, $\boldsymbol{k} \in \mathbb{R}^M$ is a blurring kernel, $\boldsymbol{x} \in \mathbb{R}^N$ is the unknown image of interest assumed to be sparse in a given basis, and $\mathbf{e}$ is a realization of an additive white Gaussian noise with variance σ^2. A wide class of methods have been designed to solve (1), including state-of-the-art variational approaches. However, reaching near-optimal solutions in high dimension comes at the cost of significant computation time, despite the parallelization and distribution capabilities of the underlying optimization algorithms. Recently, Neural Networks (NN), in particular U-nets [1] have shown promising results on imaging problems such as deconvolution or CT reconstruction [2] with extremely fast forward pass. Yet, NN still suffer from robustness issues in more complex problems, in particular when deblurring is attempted on a blurring kernel on which it has not been trained. To alleviate this issue, we propose a two-step approach consisting in first solving an optimization problem and then refining its solution with a NN. The optimization problem is kept simple and the algorithm is stopped after a fixed number of iterations to guarantee speed of the first step. The aim of the DPP step is not anymore to deconvolve but rather to correct the regularization artefacts introduced in the first step. We showcase the interest of our method in terms of imaging quality and scalability with an application to radio-astronomical imaging. Our preliminary results suggest that adequate preprocessing allows for the generalization of the network when not trained on the kernel of interest.

II. Methodology illustrated in radio astronomy

Radio interferometry is a high precision imaging technique probing the radio emissions in the universe. In essence, radio data are complex measurements corresponding to noisy and under-sampled Fourier coefficients of the radio image. In the image domain, the data model can be formulated as in (1), where $\boldsymbol{k}$ is known and called the Point Spread Function (PSF). As a benchmark, we consider the state-of-the-art method SARA, a reweighted ℓ_1 minimization algorithm [3].

This work was supported by the UK Engineering and Physical Sciences Research Council (EP/M008843/1 and EP/M019306/1) and the Swiss-South Africa Joint Research Program (IZLSZ2_170863/1), and used the Cirrus UK National Tier-2 HPC Service at EPCC (http://www.cirrus.ac.uk) funded by the University of Edinburgh and EPSRC (EP/P020267/1).

Our dataset is derived from a high sensitivity simulated radio map of size $4.10^5 \times 4.10^5$, provided in the SKA data challenge [4], from which we extract 1100 random patches of size $N = 512 \times 512$ as ground truth, where 1000 patches are used for training and the remaining for testing. Observed images are obtained by blurring each patch with a realistic PSF of the MeerKAT telescope of size $M = 256 \times 256$, with additive zero-mean Gaussian noise.

We study two strategies for the first step: Wiener filtering and the approximate solving of the ℓ_1 minimization problem

$$\underset{\boldsymbol{x} \in \mathbb{R}_+^N}{\text{minimize}} \ \frac{1}{2}\|\boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{\Psi}^\top \boldsymbol{x}\|_1, \tag{2}$$

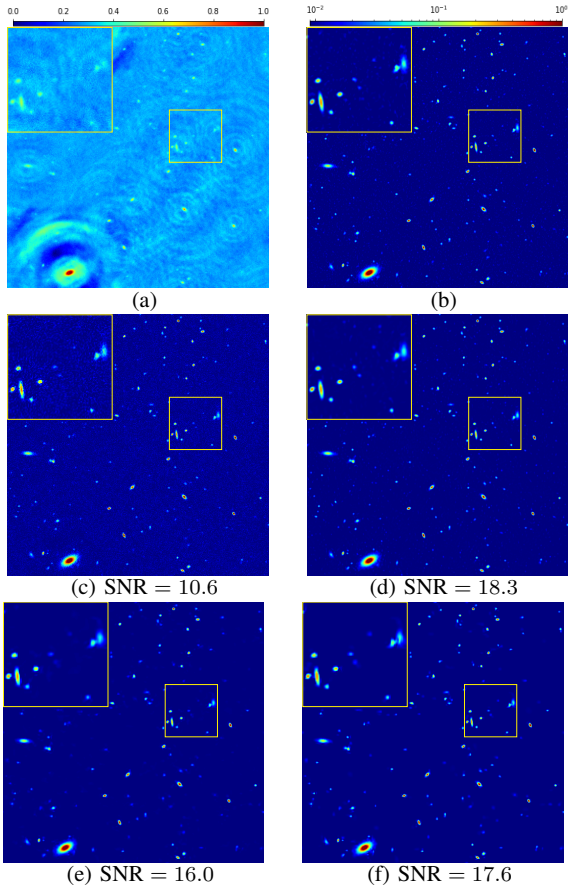
where $\lambda > 0$ is a regularization parameter and $\boldsymbol{\Psi}$ a sparsity basis. We solve (2) using FISTA [5]. In order to maintain processing speed, we stop the algorithm when reaching 200 iterations. The problem is thus not solved precisely due to a loosely tuned parameter λ and the small number of iterations allowed.

We propose two input datasets to be fed to the network: (i) the estimate obtained for the Daubechies wavelet basis Db4, dubbed $\ell_{1,+}$; and (ii) the estimate obtained as the average of the solutions to (2) achieved for different wavelet bases Db2, Db4, Db6 and Db8, dubbed Sparsity Averaging (SA). Our target output data is the log scale of the ground truth and the input is kept in linear scale; both are normalized in [-500, 500]. The chosen network is the U-net [1] and is trained by minimizing the mean squared error on 100 epochs using stochastic gradient decent (batch size 3) with a decreasing step size from 10^{-2} to 2×10^{-6}. The training is done on two NVIDIA Tesla V100 GPUs.

To evaluate the performance of DPP, we use the signal-to-noise ratio (SNR) that is $\mathrm{SNR}_{\mathrm{dB}} = 20 \log \left(\frac{\|\boldsymbol{x}\|_2}{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2} \right)$, where $\hat{\boldsymbol{x}}$ is the recovered image and $\boldsymbol{x}$ is the ground truth. We first consider the *PSF-aware* case, where the network is tested on a dataset blurred with the same PSF that was used for training. Reconstructions are given in Fig. 1. One can notice the noisy but deblurred aspect of the preprocessed data. Fig. 2 shows the SNR values for different noise levels and indicates the comparable performance of our method to that obtained by SARA [3]. We also consider a *PSF-unaware* case, where the network is tested on a dataset blurred with PSFs on which it has not been trained. These new PSFs show the same structure as the one used for training but the lobes and the width of the beam differ (see Fig. 4). We refer to 'DPPa' and 'DPPu' as DPP in the PSF-aware and PSF-unaware cases respectively. Reconstruction results are given in Fig. 3. Table I summarises both timing and SNRs.

III. Conclusions

Our results suggest that DPP is able to achieve an imaging quality similar to state-of-the-art optimization algorithms in a fraction of the time, thus opening the door to a significant improvement in scalability. Furthermore, our approach could be of potential interest to upgrade historical data in radio astronomy, processed by the traditional CLEAN algorithm for decades.



(a) (b) (c) SNR = 10.6 (d) SNR = 18.3 (e) SNR = 16.0 (f) SNR = 17.6

Fig. 1: Deconvolution in the PSF-aware case for $\sigma = 0.1$: (a) observed image of the sky in linear scale. All in logscale: (b) the groundtruth, (c) the reconstruction from the SA preprocessing, (d) the deblurred image by SA-DPPa. (e) shows the deblurred image by SARA without reweighting and (f) with reweighting.

Algorithm	t. (s)	SNR (dB)
SARA	$8.2\,10^3$	18.1
SARA, no reweighting	200	16.9
SA-DPPa	31	18.3
Wiener-DPPa	< 1	13.7
SA-DPPu	31	16.2
Wiener-DPPu	< 1	11.4

TABLE I: Comparison of the different methods on a sample of 100 test images for $\sigma = 0.1$. In the PSF-unaware case, each image from the test set is blurred with 9 different PSFs. The computation time is experienced on a GPU and assumes a parallel implementation of SA. Both SNR and computation times are averaged over the test set.

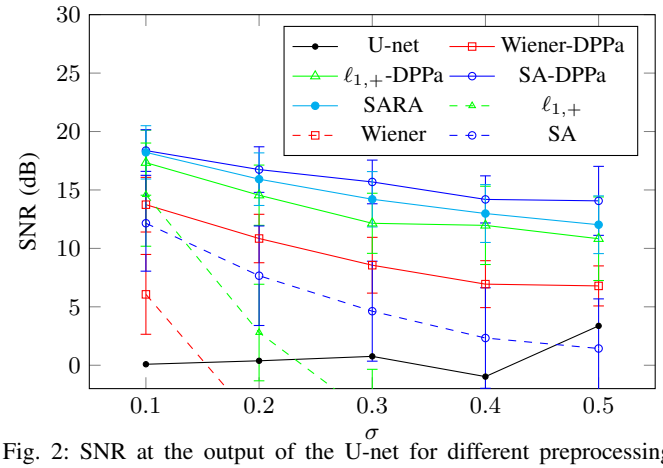
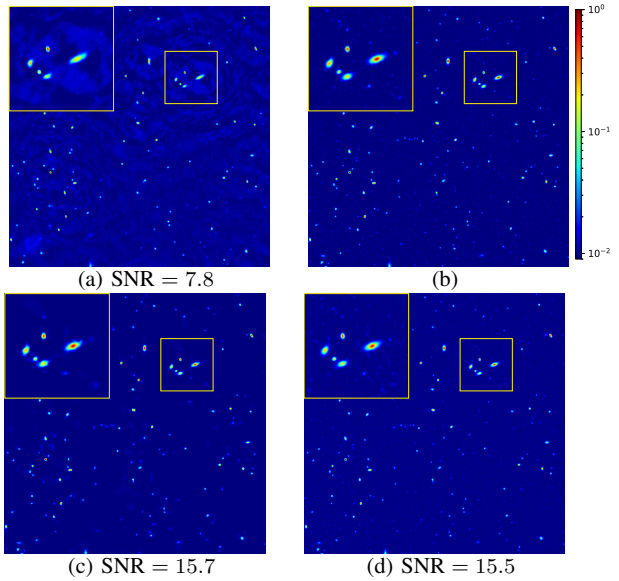


Fig. 2: SNR at the output of the U-net for different preprocessing strategies on 100 test images in the PSF-aware case. Dashed lines show the performance of preprocessing and black line shows the results of the non-preprocessed U-net. SARA is shown in cyan.



(a) SNR = 7.8 (b) (c) SNR = 15.7 (d) SNR = 15.5

Fig. 3: Deconvolution results (in logscale) in the PSF-unaware case, for $\sigma = 0.1$. (a): deblurred image by the U-net *without* preprocessing; (b) groundtruth; (c) deblurred by SARA; (d) deblurred by SA-DPPu.

Fig. 4: Samples of PSFs (zoom); the leftmost one is used for training and in PSF-aware experiments. Others are used in PSF-unaware cases.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[2] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[3] R. E. Carrillo, J. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (sara): a novel algorithm for radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 2012.

[4] A. Bonaldi *et al.*, "Square kilometre array science data challenge 1," *arXiv preprint arXiv:1811.10454*, 2018.

[5] A. Chambolle and C. Dossal, "On the convergence of the iterates of fista," *Journal of Optimization Theory and Applications*, vol. 166, no. 3, p. 25, 2015.