# Robust Super-Resolution via Deep Learning and TV Priors

Marija Vella[⋆] Colin Rickman[♯] João F. C. Mota[⋆]
[⋆] Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh, UK
[♯] Institute of Biological Chemistry, Biophysics and Bioengineering, Heriot-Watt University, Edinburgh, UK

## I. Introduction

In several applications, from biological microscopy and healthcare to consumer electronics, imaging sensors produce images with sub-optimal resolution. To super-resolve them, one has to make assumptions about the unobserved pixels; and different methods make different types of assumptions. Super-resolution (SR) methods can be divided into interpolation-based (e.g., [1]), sparsity-based, and learning-based.

Sparsity-based methods assume that images have parsimonious representations in some domain, e.g., in the edge-space [2]. As, in general, they are based on convex optimization, such methods are robust and have strong theoretical guarantees. Currently, however, the best performing methods are learning-based (e.g., [3]–[5]). These consist of deep neural networks (DNNs) that map low-resolution (LR) to high-resolution (HR) images and whose weights are learned by solving a nonconvex optimization problem using extensive training data. Although learning-based methods have outstanding performance, they are unstable, lack theoretical foundations, and suffer from generalization issues, i.e., may fail to super-resolve an image that differs significantly from the images in the training set.

In this paper, we propose a robust method that combines the advantages of both sparsity-based and learning-based methods. The method has as inputs the LR image and the (super-resolved) output of a DNN. It can thus be viewed as a post-processing step. Experiments using the DNNs in [4], [5] show that our scheme not only systematically improves the outputs of those networks, but also mitigates generalization problems. For example, for upscaling factors of 2 and 4, it leads to a gain of $0.8 - 2.6$ dB in the average PSNR.

## II. Our Method

**Problem statement.** Let $X^\star \in \mathbb{R}^{M\times N}$represent a HR image and let $x^\star \in \mathbb{R}^n$ be its vectorization, with $n = MN$. Here, $X^\star$ is either a grayscale image or, as in [4], [5], represents the luminance channel of a color image in YCrCb space. Suppose we have access to a LR version of $x^\star$, denoted $b \in \mathbb{R}^m$, where $m < n$, and which can be related to $x^\star$ linearly: $b = Ax^\star$, with $A \in \mathbb{R}^{m\times n}$ being a subsampling matrix. Our goal is to reconstruct $x^\star$ from $b$ and $A$, i.e., to super-resolve $b$.

**Our method.** As the equation $b = Ax$ has an infinite number of solutions, to be able to reconstruct $x^\star$ one has to make assumptions about its structure. A common assumption in many imaging tasks is that $x^\star$ has a small number of edges, which can be captured by a small 2D TV-norm $\|x^\star\|_{\mathrm{TV}} = \|Dx^\star\|_1$, where $D$ is a circulant matrix that computes horizontal and vertical gradients. As mentioned, however, such a sparsity-based principle, although stable and well understood, is outperformed by learning-based algorithms. We thus propose to reconstruct $x^\star$ using both TV minimization and information from the output $w \in \mathbb{R}^n$ of a DNN that super-resolves $x^\star$ (by using $b$). Fig. 1 illustrates our scheme, which is inspired by the framework of $\ell_1$-$\ell_1$ minimization [6]. Given the LR image $b$, and a HR image $w$ computed by a DNN, it integrates $b$ and $w$ via TV-TV minimization:

$$\min_x \ \|x\|_{\mathrm{TV}} + \beta\|x - w\|_{\mathrm{TV}} \quad \text{s.t.} \quad Ax = b\,, \tag{1}$$

with $\beta \geq 0$. As suggested in [6], we set $\beta = 1$. Problem (1) is convex and can be solved, e.g., using ADMM [7].

## III. Experimental Results

**Experimental setup.** We instantiated the DNNs in Fig. 1 with two pretrained networks: the coupled deep autoencoder (CDA) [4], and the cascaded sparse coding based network (CSCN) [5]. CDA consists of two autoencoders that learn intrinsic representations of image patches: one of a HR patch, and another of the corresponding LR patch (previously upsampled via bi-cubic interpolation). A one-layer network connects the intrinsic representations. CSCN [5] takes as input a LR patch (previously upsampled via bi-cubic interpolation) and extracts its features with a convolutional layer. The features are then fed to a LISTA [8] network, yielding a sparse representation. The HR patch is reconstructed by multiplying the sparse code by a HR dictionary. We used the Matlab implementation in [9].

For each DNN instantiation, we super-resolved the test images in datasets Set5 [10] and Set14 [11] using upscaling factors of 2 and 4, and compared three SR procedures: simple TV minimization, i.e., (1) with $\beta = 0$ using the TVAL3 solver [12], DNN (CDA or CSCN), and our proposed method, i.e., (1) with $\beta = 1$ and $w$ equal to the output of the DNN. Whole images were considered as inputs to the solvers for $\beta = 0$ and $\beta = 1$.

**Results.** Table I (resp. II) shows the average PSNR and SSIM [13], both computed on the luminance channel, of the super-resolved images using CDA (resp. CSCN). It can be observed that our method achieves better results in both metrics.

Figures 2 and 3 show the results for an example test image, for each of the two DNN instantiations: Fig. 2 for CDA, Fig. 3 for CSCN. These images were obtained by merging the output of the DNN and the bi-cubic upscaled chrominance channels of the original image and converting them to RGB. The region highlighted in the figures illustrates that our method preserves the details of the parrot's eye better than both simple TV minimization and the DNNs implemented by CDA and CSCN.

## IV. Conclusion

We introduced a robust super-resolution (SR) algorithm that takes as input the output of another SR method, and improves it. Our experiments showed that the proposed algorithm leads to $0.8 - 2.6$ dB improvement in the average PSNR over two state-of-the-art SR methods based on deep neural networks.

## Acknowledgment

Work supported by EPSRC EP/S000631/1 via UDRC/MoD.

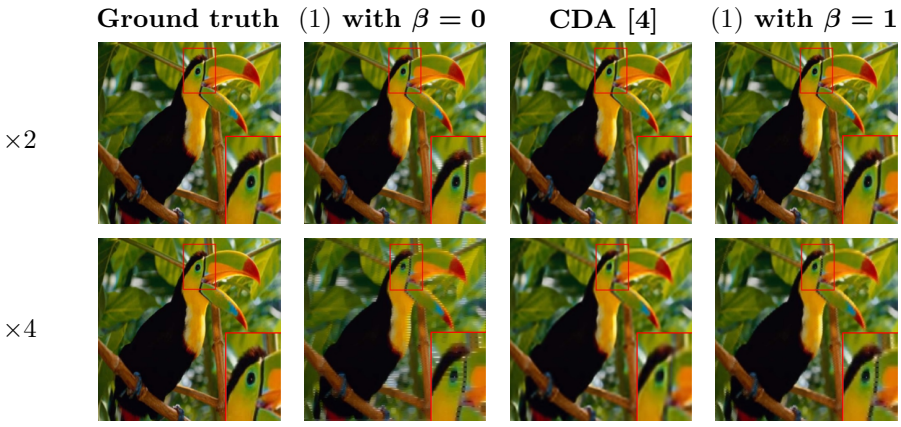


Fig. 2. Results on *Bird image* using CDA [4]. (a) Original (b) (1) with $\beta = 0$ (c) CDA [4] (d) (1) with $\beta = 1$.

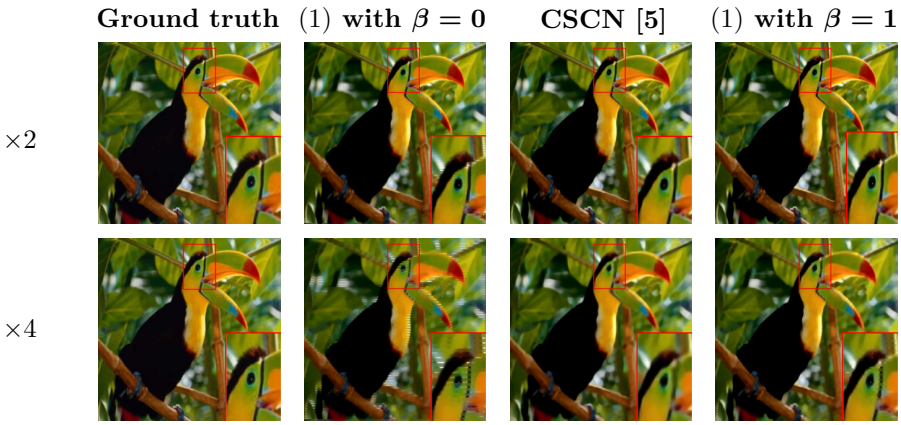


Fig.3. Results on *Bird image* using CSCN [5]. (a) Original (b) (1) with $\beta = 0$ (c) CSCN [5] (d) (1) with $\beta = 1$.

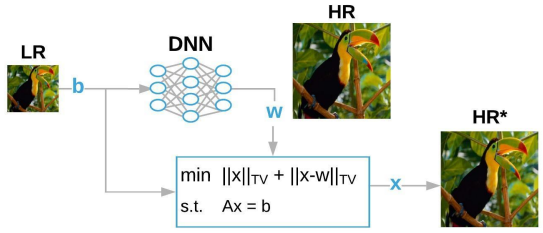


Fig. 1. Block diagram representation of our method.

TABLE I: Average PSNR (SSIM) results using CDA [4].

| Dataset | Scale | (1), $\beta = 0$ | CDA [4] | (1), $\beta = 1$ |
|---|---|---|---|---|
| Set5 | ×2 | 32.89 (0.95) | 36.52 (0.96) | **38.63 (0.98)** |
| | ×4 | 26.08 (0.84) | 30.39 (0.87) | **31.54 (0.91)** |
| Set14 | ×2 | 29.83 (0.94) | 32.44 (0.96) | **34.37 (0.97)** |
| | ×4 | 24.29 (0.79) | 27.54 (0.83) | **28.37 (0.87)** |

TABLE II: Average PSNR (SSIM) results using CSCN [5].

| Dataset | Scale | (1), $\beta = 0$ | CSCN [5] | (1), $\beta = 1$ |
|---|---|---|---|---|
| Set5 | ×2 | 33.04 (0.96) | 36.52 (0.96) | **38.95 (0.98)** |
| | ×4 | 26.21 (0.84) | 30.50 (0.88) | **31.80 (0.92)** |
| Set14 | ×2 | 29.87 (0.95) | 31.48 (0.95) | **34.04 (0.97)** |
| | ×4 | 24.60 (0.81) | 26.22 (0.81) | **27.96 (0.88)** |

## References


[1] M. Unser, A. Aldroubi, M. Eden, "Enlargement or reduction of digital images with minimum loss of information," *IEEE Trans. Im. Proc.*, 4(3), 1995.

[2] F.Shi, J. Cheng, L. Wang, P. Yap, D. Shen, "LRTV: MR image super-resolution with low-rank and total variation regularizations," *IEEE Trans. Med. Im.*, 34(12), 2015.

[3] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CVPR*, 2016.

[4] K. Zeng, J. Yu, R.Wang, C. Li, D.Tao, "Coupled deep autoencoder for single image super-resolution," *IEEE Trans. Cybernetics*, 47(1), 2017.

[5] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, "Deep Networks for Image Super-Resolution with Sparse Prior," *ICCV*, 2015.

[6] J. F. C. Mota, N. Deligiannis, M. R. D. Rodrigues, "Compressed Sensing With Prior Information: Strategies, Geometry, and Bounds," *IEEE Trans. Inf. Th.*, 63(7), 2017.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *FoT Machine Learning*, 3(1), 2011.

[8] K. Gregor, Y. LeCun, "Learning fast approximations of sparse coding," *ICML*, 2010.

[9] CSCN code: https://github.com/huangzehao/SCN_Matlab.

[10] M. Bevilacqua, A. Roumy, C. Guillemot, M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," *Proc. BMVC*, 2012.

[11] R. Zeyde, M. Elad, M. Protter, "On single image scale-up using sparse-representations," *Curves and Surfaces*, 2012.

[12] C. Li, W. Yin, H. Jiang, Y. Zhang, "An efficient augmented lagrangian method with applications to total variation minimization," *Comput. Optimi. Appl.*, 56(3), 2013.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli,"Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Im. Proc.*, 13(4), 2004.