A hybrid interior point - deep learning approach for Poisson image deblurring

Mathilde Galinier, Marco Prato, Emilie Chouzenoux, Jean-Christophe Pesquet

2020 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 21–24, 2020, ESPOO, FINLAND

A HYBRID INTERIOR POINT - DEEP LEARNING APPROACH FOR POISSON IMAGE DEBLURRING

M. Galinier[*]*, M. Prato*[†]

Università di Modena e Reggio Emilia
Dipartimento di Scienze Fisiche,
Informatiche e Matematiche
Via Campi 213/b, 41125 Modena, Italy
first.last@unimore.it

E. Chouzenoux[‡]*, J.-C. Pesquet*[§]

Université Paris-Saclay
Centrale-Supélec, Inria
Centre de Vision Numérique
91190 Gif-sur-Yvettes, France
first.last@centralesupelec.fr

ABSTRACT

In this paper we address the problem of deconvolution of an image corrupted with Poisson noise by reformulating the restoration process as a constrained minimization of a suitable regularized data fidelity function. The minimization step is performed by means of an interior-point approach, in which the constraints are incorporated within the objective function through a barrier penalty and a forward–backward algorithm is exploited to build a minimizing sequence. The key point of our proposed scheme is that the choice of the regularization, barrier and step-size parameters defining the interior point approach is automatically performed by a deep learning strategy. Numerical tests on Poisson corrupted benchmark datasets show that our method can obtain very good performance when compared to a state-of-the-art variational deblurring strategy.

Index Terms— Interior point method, proximal algorithms, deep unfolding, neural network, Poisson image restoration.

1. INTRODUCTION

Image restoration is a challenging problem that aims at recovering an accurate estimate of an original image from degraded observations. The degradations, often unavoidable in practical situations, may arise in various forms, such as low resolution, motion blur, or noise. In this paper, we focus on the particular task of deblurring images corrupted by Poisson noise, which has applications in many fields such as astronomy, medicine, and fluorescence microscopy (see e.g. [1] and references therein).

If we consider a discrete version of an object $\bar{x} \in \mathbb{R}^N$ observed as an image $y \in \mathbb{R}^M$ through an optical system modeled by the linear operator $H \in \mathbb{R}^{M\times N}$ and corrupted by a Poisson noise process $\mathcal{D}$, then the image formation model can be written as:

$$y = \mathcal{D}(H\bar{x}). \tag{1}$$

We will focus on the case when H models the 2D convolution of the image with a kernel assumed to be known from a physical model or prior identification step.

An approximation to the solution $\bar{x}$ in Problem (1) can be computed by means of variational methods, where one seeks to minimize a cost function expressed as the sum of a data-fitting term and a regularization function, which encodes a priori knowledge about $\bar{x}$ and penalizes unfeasible solutions. Furthermore, some additional prior information about the sought variable $\bar{x}$ can be incorporated through constraints limiting the domain of search. In this paper, we choose the regularization function to be a smoothed variant of the Total Variation (TV) [2]. The formulation of the data-fidelity term standard methods relies upon the Maximum Likelihood approach [3], which leads, in the Poisson noise case, to the Kullback–Leibler (KL) divergence [4, 5] or alternatively, to the Anscombe transform [6, 7, 8, 9]. The latter, defined as a variance–stabilizing transformation, approximately converts Poisson data into a signal in which the noise can be treated as additive Gaussian noise. Because it is easy to compute and most importantly, it is defined for non–negative values of its variables, unlike the KL divergence, we opt here for the Anscombe transform approach. Therefore, the regularized minimization problem reads:

$$\min_{x\in\mathcal{C}} \frac{1}{2}\|\phi(Hx) - \phi(y)\|_2^2 + \lambda \mathrm{TV}(\mathrm{x}) \tag{2}$$

where $\lambda \in]0, +\infty[$ is a regularization parameter, $\mathcal{C}$ is a subset of $\mathbb{R}^N$ that we will set equal to the hypercube $[0,1]^N$, and ϕ

[*]M.G., as fellow of the INdAM Doctoral Programme in Mathematics and/or Applications, is supported by Marie Sklodowska-Curie Actions (INdAM-DP-COFUND-2015), grant number 713485.

[†]M. P. is partially funded by the ECSEL JU programme under the PRYSTINE Project Grant Agreement No. 783190.

[‡]E.C. acknowledges support from the ERC StG MAJORIS ERC-2019-STG-850925.

[§]J.-C. Pesquet is member of Institut Universitaire de France.

is the Anscombe transform, defined as $\phi(z) = 2\sqrt{z + 3/8}$ where the square root is applied on each component of the vector $z \in \mathbb{R}^M$.

The resulting constrained minimization problem can be addressed by an Interior Point Method (IPM), which converts (2) into an unconstrained problem penalized by a logarithmic barrier function. An approximation to the solution can then be obtained by using iterative solvers such as a proximal forward–backward algorithm [10, 11]. In such a case, the overall scheme relies on a significant number of parameters, namely the regularization parameter, the barrier parameter and step-size sequences. Existing approaches offer a way to choose such parameters based on statistical considerations or heuristic analysis, but may respectively lead to a substantial increase of the computational cost or a loss in terms of efficiency and versatility of the restoration scheme. To overcome those issues, we propose to learn the above-mentioned parameters through the powerful framework of Deep Neural Networks (DNN). The proximal IPM is thus unfolded over a finite number of iterations and the parameters to be learned are untied across the layers of the network, in a similar fashion that what was proposed in [12] in the case of image deblurring subject to Gaussian noise. The resulting architecture combines the benefits of the variational–based method and deep learning techniques. Our proposed algorithm is all the more appealing as very few existing works use machine learning approaches to address the problem of deblurring images corrupted by Poisson noise. GradNet [13], combining a Bayesian MAP estimator of noise level and a CNN-based gradient descent for deblurring, as well as the cascaded architecture for training the fidelity term in [14], are two interesting strategies for deblurring-and-denoising problems which however are not specific to Poisson noise. On the contrary, the dictionary learning approach developed in [15] offers an appealing way of deblurring images corrupted by Poisson noise, but does not exploit the potential of deep learning.

This paper is organized as follows: in Section 2, we describe the proximal interior point optimization method. In Section 3, we present the proposed neural network architecture. Finally, Section 4 is dedicated to the numerical experiments.

2. PROXIMAL INTERIOR POINT METHOD

2.1. Interior point approach

Let us denote by $f(\cdot, y) = \frac{1}{2}\|\phi(\cdot) - \phi(y)\|_2^2$ the data-fidelity term and $h(\cdot, y, \lambda)$ the cost function to be minimized: for every $x \in \mathbb{R}^N$,

$$h(x, y, \lambda) = f(Hx, y) + \lambda \mathrm{TV(x)} \tag{3}$$
$$= \frac{1}{2}\|\phi(Hx) - \phi(y)\|_2^2 + \lambda \sum_{i=1}^{N} \sqrt{\frac{(D_V x)_i^2 + (D_H x)_i^2}{\delta^2} + 1}$$

where $\delta > 0$ is a smoothing parameter for the total variation, and $D_V \in \mathbb{R}^{N\times N}$ and $D_H \in \mathbb{R}^{N\times N}$ are the vertical and horizontal gradient operators, respectively.

Problem (2) does not have a closed form solution, hence an iterative solver must be used. Several resolution approaches are available, either based on projected gradient strategies, ADMM, primal–dual schemes, or interior point techniques [16, 17]. Here, we focus on the method described in [12], where the interior point framework is combined with a proximal forward–backward strategy. This method first amounts to replacing the initial constrained optimization problem by a sequence of unconstrained problems:

$$\min_{x\in\mathbb{R}^N} h(x, y, \lambda) + \mu_k \mathcal{B}(x) \tag{4}$$

where $\{\mu_k\}_{k\in\mathbb{N}} \subset]0, +\infty[$ are the so-called barrier parameters, which vanish along the minimization process, and $\mathcal{B} : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is the logarithmic barrier function defined, for every $x \in \mathbb{R}^N$, as:

$$\mathcal{B}(x) = \begin{cases} -\sum_{j=1}^{N} \left(\ln(x_i) + \ln(1 - x_i)\right) & \text{if } x \in]0, 1[^N \\ +\infty & \text{otherwise.} \end{cases}$$

The forward–backward proximal interior point method in [12] can then be written as Algorithm 1, where $\nabla_1 h$ denotes the partial gradient of h with respect to the first variable and the proximity operator is defined as [11]:

$$\mathrm{prox}_{\gamma\mu\mathcal{B}}(x) = \underset{u\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|x - u\|^2 + \gamma\mu\mathcal{B}(u)$$

for all $(\gamma, \mu) \in]0, +\infty[^2$ and $x \in \mathbb{R}^N$.

Algorithm 1: Forward–backward proximal IPM

Let $x_0 \in \mathrm{int}\, \mathcal{C}$, $\underline{\gamma} > 0$ and $\{\gamma_k\}_{k\in\mathbb{N}} \subset [\underline{\gamma}, +\infty[$
for $k = 0, 1, \ldots$ **do**
  $x_{k+1} = \mathrm{prox}_{\gamma_k\mu_k\mathcal{B}}(x_k - \gamma_k \nabla_1 h(x_k, y, \lambda))$
end

2.2. Unfolding of forward–backward proximal IPM

In IPMs, the sequences of the barrier and step-size parameters, $\{\mu_k\}_{k\in\mathbb{N}}$ and $\{\gamma_k\}_{k\in\mathbb{N}}$ are usually chosen according to some heuristic rules, so that the convergence of the method to a minimizer of the objective function is guaranteed. However, such handcrafted variational methods do not necessarily capture perceptual image quality well, nor lead to satisfactory reconstructions. As for the regularization parameter λ, existing techniques based on statistical considerations do offer a way of selecting it, but may lead to a significant increase of the computational cost.

To overcome those limitations, Algorithm 1 is unfolded over a finite number of iterations K so that the stepsize, barrier and regularization parameters are learned by means of a data-driven approach.

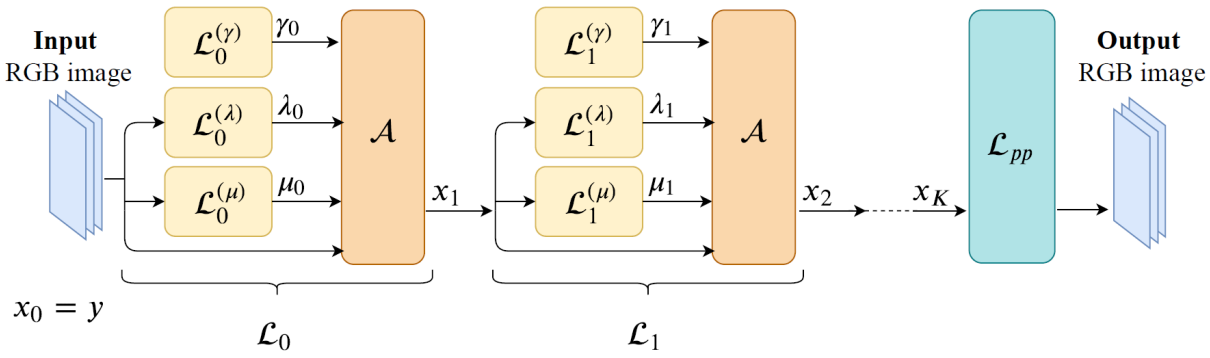

Fig. 1: iRestNet architecture [12]. $\mathcal{A}$ stands for the update rule (5)

3. IRESTNET ARCHITECTURE

3.1. Overview

Unfolding Algorithm 1 allows us to incorporate Deep Neural Networks to the overall scheme, and thus to determine an optimal setting for the above-mentioned parameters through a supervised learning process. In order to increase the flexibility of the model, the parameter λ is furthermore untied across the network, leading to the update rule at a given iteration $k \in \{0, \dots, K-1\}$:

$$x_{k+1} = \text{prox}_{\gamma_k \mu_k \mathcal{B}}(x_k - \gamma_k \nabla_1 h(x_k, y, \lambda_k)). \tag{5}$$

Each unfolded iteration $k \in \{0, \dots, K-1\}$ is represented by a layer $\mathcal{L}_k$, which is built as the combination of three hidden structures, $\mathcal{L}_k^{(\mu)}$, $\mathcal{L}_k^{(\gamma)}$ and $\mathcal{L}_k^{(\lambda)}$, followed by the update (5). Structures $\mathcal{L}_k^{(\mu)}$, $\mathcal{L}_k^{(\gamma)}$ and $\mathcal{L}_k^{(\lambda)}$ infer the parameters μ_k, γ_k and λ_k, respectively. After the K^{th} layer, we make use of a post-processing layer, denoted as $\mathcal{L}_{\text{pp}}$. The resulting architecture, depicted in Figure 1, was originally presented in the paper [12] under the name iRestNet and will be referred to as such in the remaining of this article.

3.2. Hidden structures

The structures of $\mathcal{L}_k^{(\mu)}$, $\mathcal{L}_k^{(\gamma)}$ and $\mathcal{L}_k^{(\lambda)}$ were designed to best infer the corresponding parameters. First, the stepsize is chosen to be estimated as:

$$\gamma_k = \mathcal{L}_k^{(\gamma)} = \text{Softplus}(a_k), \quad k \in \{0, \dots, K-1\}, \tag{6}$$

where a_k is a scalar parameter learned during the training of the network, and the Softplus function

$$\text{Softplus}(z) = \ln(1 + \exp(z)), \qquad \forall z \in \mathbb{R} \tag{7}$$

is a strictly positive activation function that can be seen a smooth approximation of the ReLU function.

Secondly, the hidden structure $\mathcal{L}_k^{(\mu)}$, built to determine the barrier parameter, is composed of two consecutive convolutional and average pooling layers, followed by a fully connected layer, as illustrated in Figure 2.

Finally, the regularization parameter λ_k at a given iteration $k \in \{0, \dots, K-1\}$ is obtained by the following expression:

$$\lambda_k = \mathcal{L}_k^{(\lambda)} = \frac{\text{Softplus}(b_k)\widehat{\sigma}\left(\phi(y)\right)}{\eta(x_k) + \text{Softplus}(c_k)} \tag{8}$$

where b_k and c_k are scalar parameters learned by the network, $\eta(x_k)$ is the standard deviation of the concatenated spatial gradients of x_k, $[(D_V x_k)^\top (D_H x_k)^\top]$, and $\widehat{\sigma}\left(\phi(y)\right)$ is an approximation of the noise level in the Anscombe transform of the blurred image. The noise level is approximated according to the method formulated in [18]:

$$\hat{\sigma}\left(z\right) = \text{median}(|W_{Hz}|)/0.6745, \quad \forall z \in \mathbb{R}^N \tag{9}$$

where $|W_{Hz}|$ is the vector collecting the absolute value of the diagonal coefficients of the first level Haar wavelet decomposition of z. This architecture offers the advantage of not requiring any prior knowledge about the noise level and thus can be applied to input images with different noise standard deviations.

As for the post-processing layer L_{pp}, it is made of 9 dilated convolutional layers with filters of size 3×3. The dilation factors of the convolutions from the first layer to the last one are set respectively to 1, 2, 3, 4, 5, 4, 3, 2 and 1, and the number of feature maps in each middle layer is set to 64. Furthermore, a ReLU activation function as well as a batch normalization are used after each convolution. Finally, a skip connection between the input and the output of L_{pp} is added, making of this post-processing structure a residual neural network. The last activation function is chosen to be the Sigmoid function.

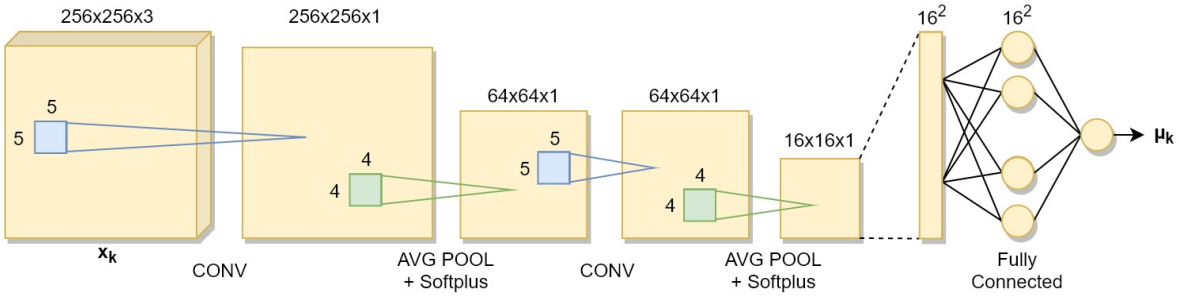

Fig. 2: Architecture of hidden structure $\mathcal{L}_k^{(\mu)}$

4. NUMERICAL EXPERIMENTS

4.1. Problem formulation and experimental setting

We consider the non-blind color image deblurring problem (1), where $N = 3n$ is the number of pixels over the 3 RGB channels, $y = \{y^{(j)}\}_{1\leq j\leq 3} \in \mathbb{R}^{3n}$ is the blurred color image, $\bar{x} = \{\bar{x}^{(j)}\}_{1\leq j\leq 3} \in \mathbb{R}^{3n}$ is the ground-truth and $H \in \mathbb{R}^{3n\times 3n}$ is a linear operator that models the circular convolution of a known blur kernel with each channel of the color image. We set the smoothing parameter δ to 0.01 in all experiments, which appears as an appropriate order of magnitude, and the number of unfolded iterations K to 40.

4.2. Datasets

To train our version of iRestNet, we use a training set composed of 1200 RGB images, among which 1000 stem from the COCO training set and the remaining 200 come from the Berkeley segmentation (BSD500) training set. The validation set is made of 100 images taken from BSD500 and the performance of the proposed method is evaluated both on the 200 RGB images of the BSD500 test set, and on the 30 RGB images of Flickr30 test set used in [19]. The test images are center-cropped using a window of size 256×256.

Blurred images are created by convolving the images with the following 25×25 blur kernels:

- **GaussA:** Gaussian kernel with a standard deviation of 1.6 pixels. Models atmospheric turbulence.
- **GaussB:** Gaussian kernel with a standard deviation of 3 pixels.
- **MotA/B:** Eight/third test kernel from [20] (real-world camera shake kernels)
- **Square:** Square uniform kernel of size 7×7.

Finally, Poisson noise has been added to the blurred images by means of the Python routine *noise.py* available online [1].

[1] https://github.com/scikit-image/scikit-image/blob/master/skimage/util/noise.py

4.3. Training

An iRestNet network is trained for each degradation model. A greedy approach is used to train the first 30 layers. Those layers are trained one after the other, so that the number of layers is not limited by the hardware. In concrete terms, the images of the training set are randomly cropped using a 256×256 window, blurred with the given kernel, corrupted with Poisson noise, and then passed through the first layer $\mathcal{L}_0$ in minibatches of 10 images. The training of $\mathcal{L}_0$ stops after 40 epochs, and its outputs are saved and used as inputs to train $\mathcal{L}_1$. The same process is repeated over the first 30 layers, the output of each layer $\mathcal{L}_k$ becoming the input of the next one $\mathcal{L}_{k+1}$. The rest of the network $\mathcal{L}_{pp} \circ \mathcal{L}_{39} \circ \cdots \circ \mathcal{L}_{30}$ is trained as one block over 500 epochs. The learning rate is initialized to 0.001 for all the layers, and multiplied by 0.9 every 50 epochs. In order to accelerate the training, the weights of $\mathcal{L}_{k+1}$ are initialized with the weights of $\mathcal{L}_k$, for $k \in \{1, \ldots, K-1\}$.

Training our version of iRestNet is performed using Pytorch [21] with an Adam optimizer [22]. The training loss to be minimized in the learning process is chosen to be the Structural Similarity index (SSIM) [23], which measures the perceptual difference between two images. All trainings are conducted using a GeForce GTX 1080 GPU and take approximately 4 days to be completed, for each blur kernel. Processing a test image through the trained network only takes about 1.4 sec.

4.4. Compared method and evaluation metrics

For the assessment of image quality, we make use of the SSIM. The reconstruction given by our proposed method is compared with a solution to Problem (2) obtained by the scaled gradient projection algorithm (SGP) [24, 25], which is a variable metric forward–backward method. We used the original Matlab routines which are available at www.unife.it/prin/software and were adapted to address the box-constrained minimization of (3). To the best of our knowledge, for the problem of image restoration under Poisson noise, there are no machine nor deep learning methods, for which codes are available online. Thus we only made comparisons with SGP.

Groundtruth | Blurred: 0.781 | iRestNet: 0.943 | SGP_B: 0.933 | SGB_D: 0.905

Fig. 3: Visual results and SSIM obtained with the different methods on one image from Flickr30 test set degraded with GaussA.

4.5. Results

In Tables 1–2 we report the SSIM values of the reconstructions obtained with iRestNet and SGP for the two test datasets considered. While for iRestNet, the untied regularization parameters λ_k are learned through the training of the network, for SGP, the unique parameter λ can be determined in several ways. In all our experiments, we first compare iRestNet with a version of SGP in which λ is selected as follow: 20 different values of the regularization parameter are tested, and we retain the one leading to the highest SSIM. This method ensures that the chosen λ is the best one according to our evaluation criterion. However, it implies that this version of SGP, denoted SGP_B, requires the knowledge of the groundtruth. In order to compare iRestNet with algorithms that are not based on the groundtruth, we also show the results obtained by selecting λ according to some automatic rule. In particular, we generated satisfactory reconstruction by setting λ according to Morozov's discrepancy principle [26] (SGP_D), where the noise level has been estimated with (9).

	GaussA	GaussB	MotA	MotB	Square
Blurred	0.695	0.545	0.406	0.576	0.570
iRestNet	**0.909**	**0.760**	0.955	**0.978**	**0.898**
SGP_B	0.883	0.720	**0.965**	0.972	0.873
SGP_D	0.829	0.697	0.897	0.918	0.845

Table 1: SSIM results on the BSD500 test dataset.

	GaussA	GaussB	MotA	MotB	Square
Blurred	0.743	0.576	0.399	0.618	0.605
iRestNet	**0.931**	**0.809**	0.955	**0.978**	**0.917**
SGP_B	0.917	0.780	**0.974**	**0.978**	0.914
SGP_D	0.869	0.755	0.933	0.949	0.877

Table 2: SSIM results on the Flickr30 test dataset.

Except for MotA, the mean SSIM achieved with iRestNet over all the experiments is greater than those obtained with the SGP methods. In particular, iRestNet provides SSIMs higher than the best SGP ones we could get by having access to the groundtruth. In the particular instance where iRestNet is not as good as SGP_B, it anyway leads to the second highest average SSIM, thus providing better results than SGP_D. Furthermore, iRestNet offers the advantage of not requiring any tuning of the parameters of the variational problem, as those are learned during the training of the network. Figure 3 shows an example of reconstructions obtained with the different methods for a Flickr30 image and the GaussA blur kernel. It is worth mentioning that no image was taken from Flickr30 for training iRestNet. Therefore, the results on the Flickr30 test set show how well the performance of the trained networks are transferable on test sets with statistics that are different from those of the training set.

5. CONCLUSION

This paper extends the scope of the hybrid interior point/deep learning approach developed in [12] to the case of images corrupted by Poisson noise, by exploiting the properties of the Anscombe transform. The resulting scheme corresponds to a sequence of neural network layers associated with forward–backward steps, in which the parameters are learned by maximizing a loss function evaluated on a set of training images. The proposed approach is able to provide satisfactory restorations and always outperforms the compared variational approach with automatically set parameters. Future work will involve the inclusion of a variable metric in the forward–backward step [27, 28], as well as a further generalization to other nonlinear problems.

6. ACKNOWLEDGMENTS

Mathilde Galinier and Marco Prato are members of the Italian INdAM Research group GNCS, which is kindly acknowledged.

7. REFERENCES

[1] M. Bertero, P. Boccacci, and V. Ruggiero, *Inverse Imaging with Poisson Data*, IOP Publishing, Bristol, UK, 2018.

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *J. Phys. D.*, vol. 60, no. 1–4, pp. 259–268, 1992.

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, 1984.

[4] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, 1951.

[5] M. Bertero, H. Lantéri, and L. Zanni, "Iterative image reconstruction: a point of view," in *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)*, Y. Censor, M. Jiang, and A. K. Louis, Eds., pp. 37–63. Birkhauser-Verlag, Pisa, Italy, 2008.

[6] F. J. Anscombe, "The transformation of poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3–4, pp. 246–254, 1948.

[7] S. Harizanov, J.-C. Pesquet, and G. Steidl, "Epigraphical projection for solving least squares Anscombe transformed constrained optimization problems," in *Scale Space and Variational Methods in Computer Vision*, A. Kuijper, K. Bredies, T. Pock, and H. Bischof, Eds., Berlin, Heidelberg, 13, pp. 125–136, Springer Berlin Heidelberg.

[8] M. Mäkitalo and A. Foi, "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 99–109, 2011.

[9] Y. Marnissi, Y. Zheng, E. Chouzenoux, and J.-C. Pesquet, "A variational Bayesian approach for image restoration. application to image deblurring with Poisson–Gaussian noise," *IEEE Trans. Comput. Imaging*, vol. 3, no. 4, pp. 722–737, 2017.

[10] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward–backward splitting," *Multiscale Model. Sim.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[11] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2010.

[12] C. Bertocchi, E. Chouzenoux, M.-C. Corbineau, J.-C. Pesquet, and M. Prato, "Deep unfolding of a proximal interior point method for image restoration," *Inverse Probl.*, vol. 36, no. 3, pp. 034005, 2020.

[13] M. Jin, S. Roth, and P. Favaro, "Noise-blind image deblurring," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3834–3842.

[14] Jiangxin Dong, Jinshan Pan, Deqing Sun, Zhixun Su, and Ming-Hsuan Yang, "Learning data terms for non-blind deblurring," in *Computer Vision – ECCV 2018*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, Eds., Cham, 2018, pp. 777–792, Springer International Publishing.

[15] L. Ma, L. Moisan, J. Yu, and T. Zeng, "A dictionary learning approach for poisson image deblurring," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1277–1289, 2013.

[16] N. Komodakis and J. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, 2015.

[17] D. Bertsekas, *Convex optimization theory. Supplementary Chapter 6 on convex optimization algorithms*, Athena Scientific, Belmont, MA, 2 december 2013 edition, 2009.

[18] S. Mallat, *A wavelet tour of signal processing*, Elsevier, 1999.

[19] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proceedings of Advances in Neural Information Processing Systems*, Montreal, 2014, pp. 1790–1798.

[20] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 2009, pp. 1964–1971.

[21] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, Eds., pp. 8024–8035. Curran Associates, Inc., 2019.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

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

[24] S. Bonettini, R. Zanella, and L. Zanni, "A scaled gradient projection method for constrained image deblurring," *Inverse Probl.*, vol. 25, no. 1, pp. 015002, 2009.

[25] S. Bonettini and M. Prato, "New convergence results for the scaled gradient projection method," *Inverse Probl.*, vol. 31, no. 9, pp. 095008, 2015.

[26] V. A. Morozov, "The error principle in the solution of operational equations by the regularization method," *URSS Comp. Math. Math. Phys.*, vol. 8, no. 2, pp. 63–87, 1968.

[27] E. Chouzenoux, J.C. Pesquet, and A. Repetti, "Variable metric forward—backward algorithm for minimizing the sum of a differentiable function and a convex function," *Journal of Optimization Theory and Applications*, vol. 162, no. 1, pp. 107–132, Jul. 2014.

[28] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero, "Efficient gradient projection methods for edge-preserving removal of Poisson noise," *Inverse Probl.*, vol. 25, no. 4, pp. 045010, 2009.