CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.2020-0016

Vol. **1**, No. xx, pp. 1-22
xxx 2020

REVIEW ARTICLE

Model Meets Deep Learning in Image Inverse Problems

Na Wang and Jian Sun*

Xi'an Jiaotong University, Xi'an, 710049, China.

Received 22 April 2020; Accepted 16 June 2020

Abstract. Image inverse problem aims to reconstruct or restore high-quality images from observed samples or degraded images, with wide applications in imaging sciences. The traditional methods rely on mathematical models to invert the process of image sensing or degradation. But these methods require good design of image prior or regularizer that is hard to be hand-crafted. In recent years, deep learning has been introduced to image inverse problems by learning to invert image sensing or degradation process. In this paper, we will review a new trend of methods for image inverse problem that combines the imaging/degradation model with deep learning approach. These methods are typically designed by unrolling some optimization algorithms or statistical inference algorithms into deep neural networks. The ideas combining deep learning and models are also emerging in other fields such as PDE, control, etc. We will also summarize and present perspectives along this research direction.

AMS subject classifications: 00-XX

Key words: Image inverse problem, model-driven deep learning, statistical model, optimization model.

1 Introduction

Image inverse problem [1] attempts to reconstruct / restore high-quality images from a few observed samples or degraded images captured by different imaging equipments. It has wide applications in medical imaging [2], compressive sensing image reconstruction [3], natural image restoration [4]. It is a fundamental and challenging task because it requires to well regularize the inverse process to have a better solution in the image

*Corresponding author. *Email addresses:* NaWang2018@stu.xjtu.edu.cn (N. Wang), jiansun@xjtu.edu.cn (J. Sun)

space. The specific tasks that involve image inverse problems include image denoising, image deblur, image dehaze, image restoration, image inpainting, super-resolution, etc. The research on image inverse problem has been popularized for a long history, based on traditional signal processing methods [5], regularization-based methods [6–8], and deep learning methods [2, 9, 10].

1.1 Traditional methods

Traditionally, image inverse problems are mainly tackled by different model-based methods. The essential contributions of these methods are the design of different regularizers or priors from various perspectives, e.g., regularized variational methods, sparse representation methods and statistical methods. The *variational methods* are based on energy models taking image as continuous function, and regularized by total variation (TV) [4, 6, 11], generalized total variation (GTV) [7, 12], non-local regularizers [8, 13] or wavelet domain regularization [14, 15]. *Sparse representation-based method* assumes that the image or image patches can be represented by a sparse combination of basis in a dictionary [8, 16–20]. *Statistical methods* in image inverse problem generally restore / reconstruct image by maximum a posteriori estimation using a Bayesian framework, and different natural image priors are designed based on statistical models such as Gaussian scale mixture [21], Markov random field [22] or conditional random field [23].

These traditional mathematical models are commonly built based on physical mechanism of imaging and image degradation process, with the expert-designed image prior and regularizer to constrain the image space. Due to the modeling of psychical mechanism, these models are explainable and without relying on a large set of training data as the deep learning approach that will be introduced as follows.

1.2 Deep learning methods

Deep learning [24] approach has been recognized as a state-of-the-art tool in artificial intelligence with wide applications in face recognition [25], machine translation [26], chess and Go competition [27], and medical image analysis [28].

In recent years, deep neural networks have been introduced to solving image inverse problems [1, 2, 29]. These deep learning methods directly learn a mapping from the degraded image to the high-quality image taking advantage of the nonlinearity and high capacity of deep neural network. Various networks have been devised for image super-resolution, denoising [30], inpainting [30], medical image reconstruction. For image super-resolution, Dong et al. [9] firstly uses a deep convolutional neural network (CNN) to learn an end-to-end mapping between the low/high-resolution images, and then Kim et al. [31] uses a very deep CNN achieving higher accuracy. To reconstruct or restore realistic images, [32] proposed to use generative adversarial network (SRGAN) and perceptual loss to introduce visually realistic textures [33, 34]. Then, [35] combined the per-pixel loss and perceptual loss for Image super-resolution task. For medical image reconstruction, classical deep CNN structures such as U-net [10], Encoder-Decoder [36] and some newly deep CNN structures [2, 37] have been used to learn a mapping from preliminary reconstructed results to high-quality reconstruction results.

All of these deep learning-based methods rely on a black-box deep neural network for mapping to the targeting high quality image. These network architectures consist of commonly utilized network layers, such as convolution, ReLU, batch normalization, etc. Obviously, these network layers are standard layers designed for general tasks, which are not specifically designed considering the domain knowledge or imagining model in image inverse problems. Due to this limitation, it relies on large training dataset and good training tricks to learn the network to properly map input degraded image / samples to the desired high quality images.

We naturally ask the following two questions. (1) Can we design deep network topology in an interpretable way such that intermediate and final outputs are explainable? (2) Can we embed the physical knowledge in the image inverse problems into the network structure design, in order to merge the domain knowledge in the design of the network structure?

1.3 Model meets deep learning: Model-driven deep learning

In recent years, there is a popular trend of research that deep network has been deeply involved with the domain knowledge and mathematical models in image inverse problems [38–40]. On the one hand, these models have inspired us to design novel network architectures with novel network layers and their connections, which are specifically designed for image inverse problems. On the other hand, deep network has been introduced to learn more powerful models, such as image prior or regularizers, for better solving image inverse problems. This category of deep-learning methods, defined as "*model-driven deep learning*", has been firstly proposed by Xu and Sun in a perspective paper of [41].

Model-driven deep learning [41] is a mixed data-driven and model-driven approach combining their respective advantages. The methods reviewed in this paper are mainly based on unfolding some optimization algorithms [38, 42, 43] or statistical inference algorithms [44, 45] in image inverse problems. Recently, the ideas that take advantages of model and deep learning are actively emerging in diverse fields. Basically, the main procedures of the model-driven deep learning in [41] are as follows. First, we formulate a parameterized model family (*model hypothesis space*) based on objectives of tasks, domain knowledge / priors. Then, iterative optimization or inference algorithm can be designed according to the model family. The algorithm inherits the uncertainty such as the hyper-parameters from model hypothesis space, forming an *algorithm hypothesis space*; finally, the parametric algorithm is unfolded into a deep structure, which can be regarded as a hierarchical nonlinear transformation, stacked by various mathematical operations in a deep architecture, dubbed as *model-driven deep network*. The network parameters include model parameters and algorithm parameters, which can be optimized in a data-driven manner through error back-propagation.

Model-driven deep learning approach [38,42,44,46–48] can be dated back to the first works that discriminatively learn the sparse codes in [42] and Markov random field model parameters in [44]. In [42], the iterative thresholding algorithm for solving the sparse codes has been unfolded as a deep architecture, for which the parameters can be optimized to speed up the approximation of sparse codes in very few number of iterations. Almost in parallel, Sun and Tappen [44] proposed to estimate statistical parameters of MRF model as natural image prior by optimizing a bilayer model using a fixed number of gradient descent iteration as approximation, and the gradient descents are unfolded as a deep architecture, and MRF parameters are learned by back-propagation over this deep architecture.

Since the popularity of deep learning after 2013, there are several typical model-driven deep learning methods that significantly advanced the research in image inverse problems. In fast MRI and compressive sensing, ADMM-Net [38] is a benchmarking work that expands the iterative process of ADMM into a deep network to learn all the model uncertainties such as regularizer, sparse transform, etc., which firstly bridge the compressive sensing and deep learning in an elegant way. For image super-resolution, deep plug-and-play method proposed in [40] extends existing plug-and-play framework by variable splitting technique allowing for plugging any learnable super-resolver. The approach in [43] replaces the proximal operator of regularizer by deep network for image denoising [39]. In parallel to model-driven deep learning approach, some interesting works that combine deep networks with differential equations [50, 51] and control systems [52,53], bearing similarity with the basic idea of model-driven deep learning.

In the following sections, we will review typical works that combine model and deep learning approach in image inverse problem, in the categories of optimization-inspired deep learning and statistics-inspired deep learning in Sections 2 and 3 respectively. We will finally conclude with discussions and perspectives on this research direction.

2 Optimization model-driven deep learning

The image inverse problems can be generally modeled as an energy function minimization problem, and the optimal solution is the desired restored or reconstructed image. It is commonly challenging to handcraft image prior / regularizer, and hyper-parameters in model/algorithm. Moreover, the optimization algorithm commonly needs hundreds of iterations to get the desired solution. To alleviate these challenges, model-driven deep learning method based on algorithm unfolding enables to learn model and algorithm uncertainties by data-driven approach and also speed-up the optimization using a fixed number of iterations. Three typical works along this research direction are LISTA [42], ADMM-Net [38], and learning proximal operators [43].

2.1 LISTA: Iterative thresholding algorithm as a learnable network

Sparse coding is a fundamental tool in image representation with potential applications in image restoration [16], classification [54], etc. Iterative thresholding-based optimization algorithms, such as ISTA (iterative shrinkage-thresholding algorithm) [55], FIST (fast iterative shrinkage-thresholding algorithm) [56] have been effective in optimizing this sparse coding model. The landmarking paper [42] first proposed a discriminant learning method to optimize sparse coding model that can achieve good solution in a fixed number of steps.

For an input vector $x\in R^n$ and a given dictionary $W\in R^{n\times m}$, which is over-complete, i.e., $m>n$. The objective of sparse coding is to find the optimal sparse coding vector $\alpha^*\in R^m$ as the minimum of energy function $E(Z;x,W)=\frac{1}{2}\|x-W\alpha\|_2^2+\lambda\|\alpha\|_1$. The iteration of ISTA algorithm is

$$\alpha_{k+1}=h_\tau(W_e x+S\alpha_k),\quad \alpha_0=0,\quad k=1,\cdots,K, \tag{2.1}$$

where $\tau=\lambda/L$, $W_e=\frac{1}{L}W^T$, $S=I-\frac{1}{L}W^TW$, L a constant as an upper bound on the largest eigenvalue of W^TW and function h_τ is defined as an element-by-element soft threshold function $h_\tau(V)_i=sign(V_i)(|V_i|-\tau)_+$.

Then a parameterized non-linear sparse "encoder" can be designed to fast approximate the optimal sparse code. The basic idea is to map the iterative algorithm in Eq. (2.1) to be a non-linear parametric deep structure as shown in Fig. 2.1 with fixed depth, in which the parameters W_e,S can be trained to approximate the optimal sparse code. The deep structure of the encoder will be expressed as $\alpha=f(X;\Theta)$, where Θ represents the set of trainable parameters in sparse encoder. The training objective of the encoder is to minimize the loss function $L(\Theta)=\sum_i\|f(x_i;\Theta)-\alpha_i^*\|_2^2$, where α_i^* is the desired sparse codes pre-computed. This learnable version of ISTA is dubbed LISTA.

The original sparse optimization algorithm ISTA commonly needs dozens or even hundreds of iterations to obtain the optimal sparse codes, LISTA can approximate the sparse codes only by iterating a few steps, which is much faster than ISTA. In term of accuracy, using LISTA as sparse encoder, the approximate solution generated by the learned LISTA can achieve similar accuracy for restoration or classification as the original ISTA. Motivated by LISTA, Sprechmann et al. [57] proposed to extend the idea to the robust principal component analysis (RPCA). Similar to the learned ISTA method, it builds the architecture of the encoders based on the iterations of RPCA algorithms, which can approximate online RPCA in a very fast way.

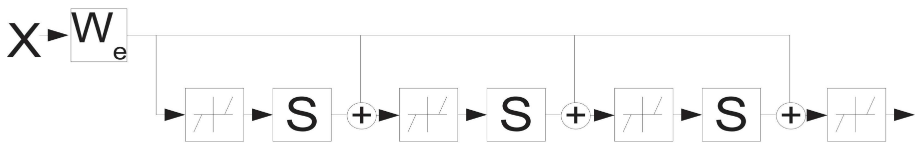


Figure 1: Neural network architecture of LISTA [42] unfolded by ISTA algorithm.

2.2 ADMM-Net: Model-driven compressive sensing network

The main purpose of the above learnable sparse coding methods is designed to accelerate the optimization speed of the original sparse coding model. The deep compressive sensing method, pioneered in ADMM-Net [38], can not only speed-up the optimization speed, but also learn the uncertainties in regularization model, e.g., the transform domain and sparse regularizers in the transformed domain, which can help solve the challenges in designing a good image prior or regularizer term for image compressive sensing. It is the first model-driven deep network for solving the compressive sensing problem in both magnetic resonance imaging (MRI) and natural image compressive sensing. We will introduce the ADMM-Net for compressive sensing MRI in this paper, and leave the details on the extended version of ADMM-CSNet for general image compressive sensing in [3].

Let $x\in\mathbb{C}^N$ be a reconstructed MRI image and $y\in\mathbb{C}^{N'}(N'<N)$ be subsampled data in k-space (i.e., Fourier space). According to compressive sensing theory [58], CS-MRI reconstruction can be generally modeled as:

$$\hat{x}=\underset{x}{\operatorname{argmin}}\Big\{\frac{1}{2}\|Ax-y\|_2^2+\sum_{l=1}^{L}\lambda_l g(D_l x)\Big\}, \tag{2.2}$$

where $A=PF$ is a measurement matrix, composed of sampling matrix P and Fourier transform F. For the traditional compressive sensing method, the two challenging problems are, first how to design the transform D_l and sparsity regularizer $g(\cdot)$ in the transform domain. Second, how to speed up the optimization procedures, which commonly takes at least tens of iterations traditionally. ADMM-Net enables to adaptively learn the transform, the sparse regularizer in the transform domain, and the hyper-parameters of iterative optimization algorithm.

As far as we know, the alternating direction method of multipliers (ADMM) has become the most widely used solver, which has been proven to be efficient and generally applicable with convergence guarantee [59, 60]. For the Eq. (2.2), by introducing auxiliary variables $z=\{z_1,z_2,\cdots,z_L\}$ to replace $D_l x, l\in\{1,\cdots,L\}$, the augmented Lagrangian function can be obtained. Then variables can be split into three subgroups, which can be alternately optimized by solving three simple subproblems:

$$\begin{cases} x^{(n)}=F^T(P^TP+\sum\limits_l \rho_l F D_l^T D_l F^T)^{-1}\Big[P^Ty+\sum\limits_l \rho_l F D_l^T(z_l^{(n-1)}+\beta_l^{(n-1)})\Big], \\ z_l^{(n)}=S\Big(D_l x^{(n)}+\beta_l^{(n-1)};\dfrac{\lambda_l}{\rho_l}\Big), \\ \beta_l^{(n)}=\beta_l^{(n-1)}+\eta_l(D_l x^{(n)}-z_l^{(n)}), \end{cases} \tag{2.3}$$

where β_l is a penalty coefficient, $S(\cdot)$ is a non-linear proximal operator related to $g(\cdot)$ and η_l is the update rate. Eq. (2.3) is the ADMM iterations for solving $\hat{x}$.

Then ADMM algorithm is unfolded into a T-layer deep neural network (ADMM-Net). According to the corresponding formula, the topology of ADMM-Net is shown

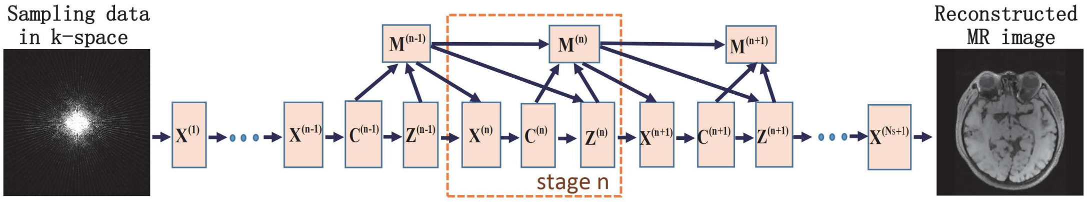


Figure 2: ADMM-Net [38] takes sampled k-space data and outputs reconstructed MRI image.

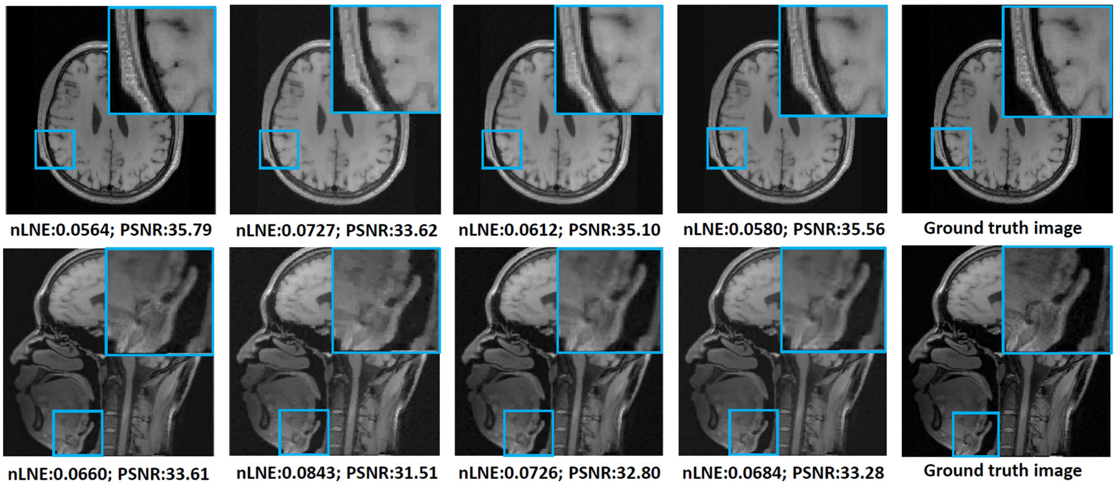


Figure 3: An example of CS-MRI [38]. From top to bottom is the comparison of 20% and 30% sampling rates respectively. From left to right is the comparison of ADMM net [38], RecPF [61], PANO [62] and FDLCP [63].

in Fig. 2, which is composed of T stages. Each stage contains four network layers: reconstruction layer $X^{(n)}: X^{(n)} = x^{(n)}$, convolution layer $C^{(n)}: C_l^{(n)} = D_l^{(n)} x^{(n)}$, nonlinear transformation layer $Z^{(n)}: Z_l^{(n)} = z_l^{(n)}$, multiplier update layer $M^{(n)}: M_l^{(n)} = \beta_l^{(n)}$, each of which is explainable and with mathematical operations related to CS-MRI, compared with standard network layers.

Fig. 3 shows the visual comparisons of 20% and 30% sampling rates in different methods on the brain data. It clearly shows that the ADMM network can achieve better reconstruction quality and preserve the fine image details without obvious artifacts.

Inspired by ADMM-Net, the conventional iterative algorithm, D-AMP, has been unrolled into the Learned D-AMP (LDAMP) in [64] achieving higher accuracy and shorter running time. On the other hand, based on ADMM network, differentiable Linearized ADMM, a learning-based method, is proposed in [65] for solving the constrained optimization problem.

2.3 Learning proximal operators by deep network

A more general approach for learning prior/regularizer in image inverse problem is to directly substitute its proximal operator by a deep network in a variable-splitting iterative optimization algorithm. For a general image inverse problem solving:

$$\hat{x}=\underset{x}{\operatorname{argmin}}\left\{\frac{1}{2}\|Ax-y\|_2^2+f(x)\right\}. \tag{2.4}$$

If we use splitting-based optimization algorithm, e.g., half-quadratic splitting [66] or ADMM, by introducing an auxiliary variable z to substitute x in Eq. (2.4), the sub-problem for optimizing z in iterative optimization can be derived by the proximal operator of regularizer $f(\cdot)$

$$v^*=prox_{\lambda f}(x)=\underset{v}{\operatorname{argmin}} f(v)+\frac{1}{2\lambda}\|x-v\|^2. \tag{2.5}$$

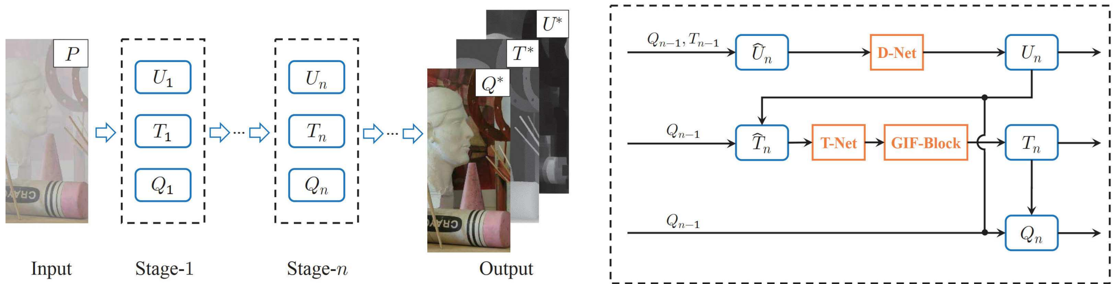
Obviously, the proximal operator is determined by the regularizer $f(\cdot)$, acting as a regularization on the solution, i.e., reconstructed/restored image. Inspired by the idea that denoising algorithm can be interpreted as a proximal operator, [67, 68] proposed to replace the proximal operator by denoising methods such as NLM [69] or BM3D [70]. With the popularity of deep neural networks, Meinhardt et al. [39] proposed to replace the proximal operator of the regularization by a denoising neural network in the primal-dual hybrid gradient (PDHG) method.

Yang and Sun [71] proposed a proximal dehaze-net for single image dehazing by learning the proximal operators of dark channel prior and transmission map prior using convolutional neural networks. Based on the haze imaging model [72–75], it uses dark channel [73] and transmission priors [72] to formulate single image dehazing as an energy model:

$$E(Q,T)=\frac{\alpha}{2}\sum_{c\in\{r,g,b\}}\|Q^c\circ T+1-T-P^c\|_F^2+\frac{\beta}{2}\|Q^{dk}\circ T+1-T-P^{dk}\|_F^2+f(T)+g(Q^{dk}), \tag{2.6}$$

where $P=I/A$ and $Q=J/A$ represent the scaled hazy image and latent haze-free image respectively, T is the media transmission, $*^c$ is a color channel, and $*^{dk}$ are dark channels, and f,g are regularization terms respectively on transmission map and dark channel.

Using half-quadratic splitting (HQS) algorithm to solve Eq. (2.6) by introducing an auxiliary variable U to substitute the dark channel Q^{dk} of latent haze-free image, the augmented energy function can be derived. Then, minimizing the augmented energy function can be achieved by solving three sub-problems for alternately updating U, T and Q. Given Q_{n-1} and T_{n-1} at iteration $n-1$, then U, T and Q can be updated sequentially as $U_n=prox_{\frac{1}{b_n}g}(\hat{U}_n)$, $T_n=prox_{\frac{1}{c_n}f}(\hat{T}_n)$ and $\vec{Q}_n=\frac{\alpha(\vec{P}+\vec{T}_n-1)\circ\vec{T}_n+\gamma D^T\vec{U}_n}{\alpha\vec{T}_n\circ\vec{T}_n+\gamma diag(D^TD)}$. Instead of using hand-designed regularizations, deep CNNs are used to learn proximal operators $prox_{\frac{1}{b_n}f}$ and $prox_{\frac{1}{c_n}f}$ for updating U_n and T_n in each stage n: $U_n\triangleq$ D-Net$(\hat{U}_n,P)$,



(a) Multi-stage network for image dehazing (b) Network structure for the n–th stage

Figure 4: Proximal Dehaze-Net [71]. (a),(b) illustrate the deep network architecture and each network stage respectively.

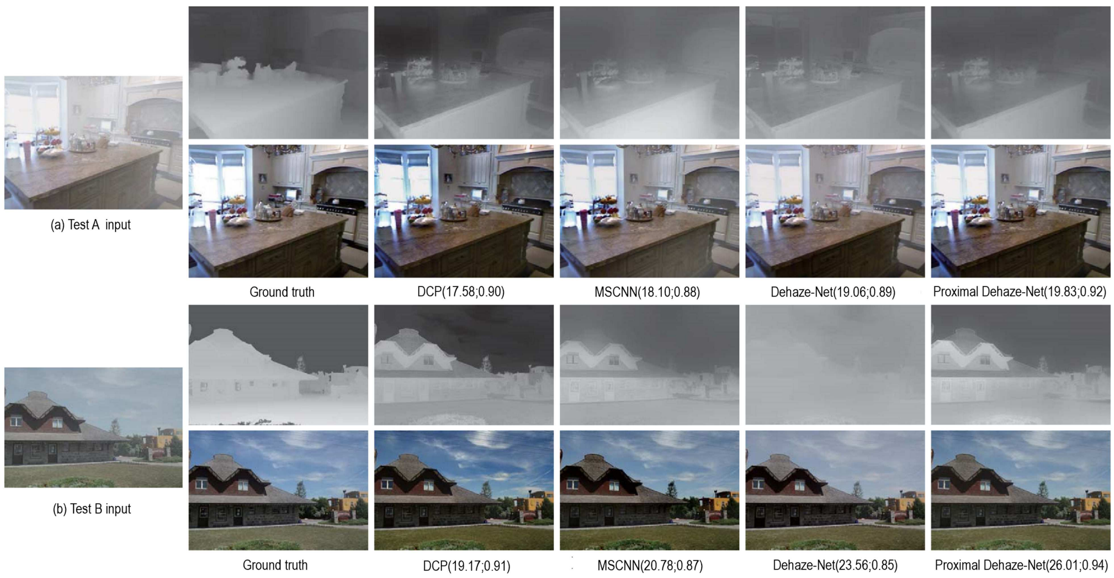


Figure 5: Dehazing results of Proximal Dehaze-Net [71] on two examples.

$T_n \triangleq \text{GIF-Block}(\text{T-Net}(\hat{T}_n, P))$. The structure of the Proximal Dehaze-Net is shown in Fig. 4. The training loss is defined as the sum of pixel-wise L_1 distances between the outputs of proximal dehaze-net $\{Q^*, T^*, U^*\}$ and the ground truths $\{Q^{gt}, T^{gt}, U^{gt}\}$. Please see [71] for technical details, and some dehazing results are shown in Fig. 5.

Learning regularizer by implicitly learning its proximal operator using deep network is an effective way for learning to regularize in image inverse problems. First, the deep network for proximal operator is obviously more powerful than the hand-crafted image regularizer such as total variation, sparse regularizer, etc. Second, the proximal operator can be specifically learned end-to-end for specific task by unfolding the iterative optimization to be a deep network with the network substituting proximal operator as a building block of the whole network.

2.4 Plug-and-play deep networks in optimization framework

Plug-and-play (PnP) is an another flexible framework to bridge iterative optimization algorithm with deep networks/related models (e.g., BM3D). Different to the end-to-end training framework in the above optimization model inspired deep networks, the PnP approach plugs the existing algorithms or pre-trained deep networks into the iterative optimization algorithm to substitute some mappings, e.g., proximal operator. This approach is simple, but the plugged networks/models are not end-to-end trained in a principled way.

An example of the early PnP method is the Plug-and-Play ADMM (PnP-ADMM) proposed in [67], which enabled to use existing denoising algorithms as priors / regularizers in ADMM model. Since then, PnP has been studied extensively with great success. A parameterized PnP-ADMM was proposed in [76]. It used a deep learning-based strategy for model-based iterative reconstruction (MBIR) to simultaneously address the challenges in prior design and MBIR parameter selection. Based on the Plug-and-Play proximal gradient method (PnP-PGM), [77] used the plug-and-play prior to transform denoisers to super-resolution solvers. [40] further proposed a deep plug-and-play super-resolution framework to solve single image super-resolution (SISR), and [78] developed a new algorithm based on plug-and-play prior for solving nonlinear imaging inverse problems. For visual tracking task, [79] presented Plug-and-Play Correlation Filters (PPCF) that iteratively integrate different off-the-shelf CF trackers.

The PnP framework, as an extension of traditional optimization framework, has attracted attentions in theoretical analysis of its convergence. Sreehari et al. [80] studied the sufficient conditions of denoising algorithms to ensure the convergence of PnP approach. Chan et al. [81] proved that the PnP-ADMM with a bounded denoiser converges to a fixed point under a continuation scheme. Theoretical convergence analyses of PnP-PGM and PnP-ADMM were provided in [82] and [83]. Xie et al. [84] proved that there exists a set of learnable parameters for Differentiable Linearized ADMM generating globally converged solutions. Recently, Ryu et al. [85] theoretically analyzed convergence of PnP-PGM and PnP-ADMM under a certain Lipschitz condition on the denoisers. It also proposed spectral normalization for training deep learning-based denoisers to satisfy the Lipschitz condition.

3 Statistical model-driven deep learning

The above optimization models usually take unknown variables as deterministic variables without considering their uncertainties. The statistical models instead use statistical distribution to describe the unknown variables. They commonly rely on maximum likelihood, variational inference, or EM (Expectation Maximization) algorithm to estimate the parameters of distributions of unknown variables. In applications such as computer vision and speech recognition, Markov random field [86–88], conditional random field [23], and corresponding EM algorithm [89] and variational inference algorithms [90] are popular statistical models and algorithms for modeling statistical dependency of high-dimensional random variables. In recent years, statistical models have been extensively combined with deep learning, deriving novel statistical model-driven deep learning methods [44, 45].

3.1 Markov random field model inspired deep network

Markov random field model (MRF) provides an effective framework for establishing the statistical distribution of images or signals. The Markov Random Field (MRF) model can be defined as a graph $G = <V, E>$, where V represents the node set of graph, $X = \{\mathbf{x}_v\}$, $v \in V$ represents the random field defined on the node set, and E represents the edges connecting graph nodes. In an early work published in 2011, Sun and Tappen [44] proposed a non-local range MRF (NLR-MRF) statistical model as a prior of natural image, and its corresponding discriminative parameter learning method can be taken as a statistical model-driven deep learning method. The statistical model for NLR-MRF is defined as

$$p(\mathbf{x};\Theta) = \frac{1}{Z(\Theta)} \prod_{c \in C} \prod_{i=1}^{N} \phi(f_i * \mathbf{x}_c;\Theta), \tag{3.1}$$

where $f_i *$ denotes the non-local range convolution using non-local range filters f_i, ϕ is the potential function defined based on Gaussian mixture model or student-T distribution, modeling the heavy-detailed distribution of natural images in transform domain, please refer to [44] for details of the model. A challenge is how to estimate the model parameters Θ.

Given the model in Eq. (3.1) as prior, the reconstructed image is usually derived by Maximum A Posteriori (MAP) estimation

$$\mathbf{x}^* = \underset{x}{\operatorname{argmin}} \{E(\mathbf{x}|\mathbf{y},\Theta) = E_{data}(\mathbf{y}|\mathbf{x}) + E_{prior}(\mathbf{x};\Theta)\}, \tag{3.2}$$

where $\mathbf{x},\mathbf{y}$ respectively denote the high-quality image to recover and degraded image, E_{prior} is derived from the prior distribution of NLR-MRF

$$E_{prior}(\mathbf{x};\Theta) = -\sum_p \sum_i \log \phi((F_i \mathbf{x})_p;\Theta). \tag{3.3}$$

The parameters Θ are learned by minimizing the loss function

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} L(\mathbf{x}^K(\mathbf{y},\Theta), t), \quad \text{where } \mathbf{x}^K(\mathbf{y},\Theta) = GradDesc_K\{E(\mathbf{x}|\mathbf{y},\Theta)\}, \tag{3.4}$$

and t is ground-truth image to recover. Using gradient descent algorithm to minimize $E(\mathbf{x}|\mathbf{y},\Theta)$ and taking ϕ as Student-t distribution, the n-th gradient descent formula for

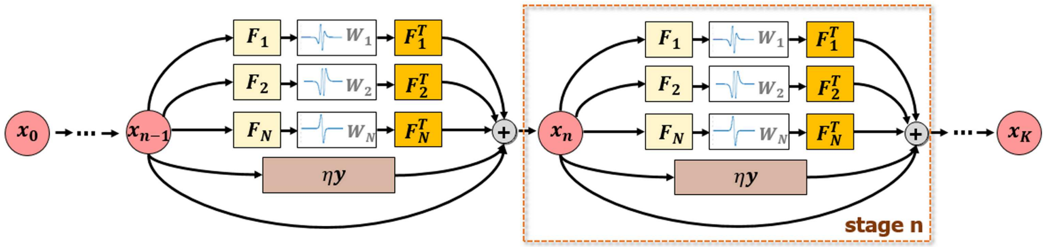


Figure 6: Deep network architecture unfolded by NLR-MRF model [44].

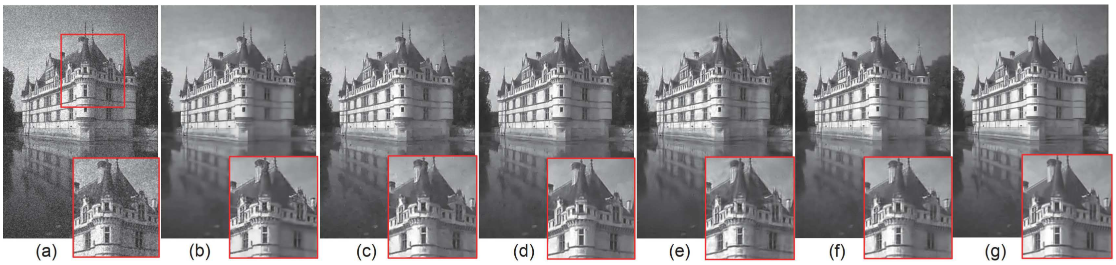


Figure 7: Comparison of denoising results of different algorithms [44]. (a) Noise image (PSNR= 25). (b) FOE (PSNR= 28.67). (c) ARF (PSNR= 28.94). (d) NLR-MRF [44] (PSNR= 29.39). (E) BLS-GSM [92] (PSNR= 29.03). (f) KSVD [17] (PSNR= 29.05). (g) BM3D [70] (PSNR= 29.60).

$GradDesc_K$:

$$\mathbf{x}^{n+1} = (1-\eta)\mathbf{x}^n + \eta\mathbf{y} - \eta\sum_{i=1}^{N}\alpha_i F_i^N W_i F_i \mathbf{x}^n, \quad n = 0,\cdots,K-1, \tag{3.5}$$

where $\mathbf{x}^0 = \mathbf{y}, F_i^N W_i F_i \mathbf{x}$ can be decomposed into convolution, nonlinear transformation and convolution operation. Please see [44] for definitions of W_i. Then, as shown in Fig. 6, the K steps of gradient descent can be unfolded into a K-stage deep network, and all the unknown parameters of model in Eq. (3.1) are parameters of this deep network to be learned by end-to-end training. Fig. 7 shows typical image denoising results. Compared with the classical MRF model (such as FOE [22], ARF [91]), NLR-MRF has better denoising results. The experiments also show that improving the number of network stages will improve the image denoising performance, but saturates after a fixed number of stages, e.g., 4 stages.

3.2 Conditional random field model inspired deep network

Zheng et al. [45] unfold the iterative process of mean-field inference algorithm of conditional random field (CRF) model into a deep neural network. In image semantic segmentation, given the observed image $\mathbf{I}$, CRF models the class label of image pixels as a

Algorithm 1 Mean-field method for solving CRF [45]

$Q_i \leftarrow \frac{1}{Z_i}\exp(U_i(l))$ for all i ▷ Initialization
while not converged **do**
  $\widetilde{Q}_i^{(m)}(l) \leftarrow \sum_{j\neq i} k^m(\mathbf{f}_i,\mathbf{f}_j)Q_j(l)$ for all m ▷ Message passing
  $\check{Q}_i(l) \leftarrow \sum_m w^{(m)}\widetilde{Q}_i^{(m)}(l)$ ▷ Linear combination of message passing outputs
  $\hat{Q}_i(l) \leftarrow \sum_{l'\in\mathcal{L}} \mu(l,l')\check{Q}_i(l)$ ▷ Compatibility transform
  $Q_i(l) \leftarrow U_i(l) - \hat{Q}_i(l)$ ▷ Adding unary values
  Exponentiate and normalize Q_i for all i ▷ Normalizing
end while

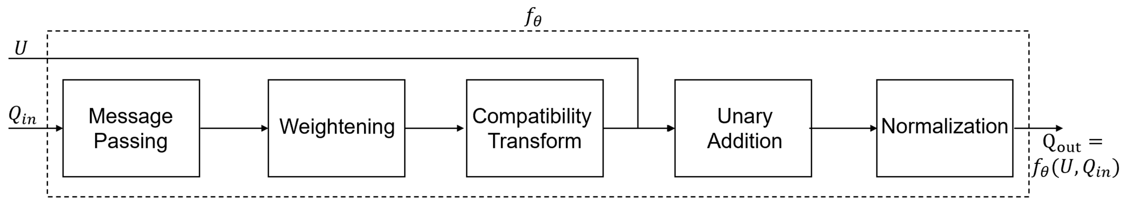


Figure 8: Deep Network Architecture unfolded by the mean-field [45].

random field with distribution

$$P(\mathbf{X}=\mathbf{x}|\mathbf{I}) = \frac{1}{Z(\mathbf{I})}\exp(-E(\mathbf{x}|\mathbf{I})), \quad \text{where } E(\mathbf{x}) = \sum_i \Psi_u(x_i) + \sum_{j<i}\Psi_p(x_i,x_j), \tag{3.6}$$

which is conditioned on the observed image I, and Ψ_u and Ψ_p are unary and pairwise terms [45]. This optimization problem is a combinatorial optimization problem. To facilitate the fast computation, the CRF distribution $P(X)$ can be approximated by the mean field, which can be written as the product of an independent marginal distribution $Q(x) = \prod_i Q_i(x_i)$. Then the iterative algorithm of approximate mean field inference can be used to build a deep RNN (recurrent neural network). The iteration of the mean field algorithm in Algorithm 1 consists of the following steps of message passing, weighting, compatibility transform, unary addition and normalization. Each step can be formulated as a stack of common CNN layers. As shown in Fig. 8, the iterative process of the mean field algorithm is unfolded into a deep RNN, dubbed CRF-RNN.

As an RNN, the parameters of Algorithm 1 can be trained by back propagation method. These parameters include weighted parameter $w^{(m)}$, compatibility matrix $\mu(l,l')$. CRF-RNN can be applied to semantic segmentation to smooth the segmentation probabilistic maps, which can be taken as a post processing restoration step [45]. The CRF-RNN can be connected to the end of a segmentation network and takes the segmentation confidence as input, and outputs the restored smooth segmentation probability maps. Specifically, using the FCN network of [93] to be concatenated with CRF-RNN, Fig. 9 shows some

Figure 9: Segmentation results of CRF-RNN [45].

visual results. It can be observed that segmentation results after adding CRF-RNN are better aligned with object boundaries.

3.3 Summary

This section mainly introduces the statistical model-driven deep learning methods inspired by Markov random field model and conditional random field model. These two models bridges the gaps between statistical parameter estimation/statistical inference of MRF/CRF with the deep learning approach. This enables to learn statistical model parameters in a discriminative data-driven way, in coupling with the unfolded iterative inference as a deep network. This idea is obviously versatile and can be generalized to more general statistical models. For example, in [94], the spatial mixed distribution model was proposed, and its EM algorithm is used to design a deep network for clustering different objects.

4 Other related research

The above mentioned model-driven deep learning networks are based on statistical and optimization models. Obviously, there are many different areas such as control, physics based on partial differential equations, that mathematical models play essential roles. Bridging mathematical models in different areas and deep learning in a model-driven approach has been widely investigated in recent years.

Deep learning and PDE. Deep learning has been a new tool or source of inspiration for the research of partial differential equation in recent years. First, the data-driven deep learning method can be used to learn the formulation of PDEs. For example, PDEs have been learned through the data-driven deep learning approach in the pioneering works [50, 95]. Second, the numerical solution of high-dimensional PDEs can be solved by deep neural network. For example, the PDEs are modeled as backward stochastic differential equations in [51], and the gradient of unknown solution is approximated by deep neural network. Third, PDEs or dynamical systems can be used to explain or inspire the design of new deep learning models. For example, the research in [96] established the relationship between the differential dynamic system and the deep residual network, which provides a new theoretical tool for the analysis of deep neural network. Based on Feynman KAC formula, the ensemble deep residual network with better robustness was designed in [97]. Based on nonlinear reaction diffusion equation, a flexible learning framework, i.e., Trainable Nonlinear Reaction Diffusion (TNRD), was proposed in [49] and all its parameters are learned from training data by learning-based approach. The above research progresses show that artificial intelligence algorithm is becoming a new tool for modeling or computations in the field of applied mathematics. At the same time, differential equation has also become a new idea for analyzing and designing novel deep neural network. Differential equation and deep learning are merging interactively and promoting each other.

Deep learning and control. The modern control theory in the control research is based on linear system, nonlinear system, time-varying system, stochastic control system, etc, which has sound fundamentals in mathematical modeling and theoretical analysis. With the rapid development of robot and autonomous vehicle, data plays a more and more important role in control. How to combine model-driven control method and data-driven learning method has become a research frontier in control research [98]. Typical developments include the combination of control theory models with the data-driven reinforcement learning methods [52,53] or deep learning-based dynamic system [99,100]. The research in [53] constructs the benchmark problem set of deep reinforcement learning for continuous control, including cart-pole swing-up, 3D humanoid locomotion etc., which provides platform for the research of reinforcement learning in control. A neural Lyapunov control method proposed in [99] learn the controller and Lyapunov function through deep learning methods to ensure the stability of the nonlinear system and obtain a larger attractive field compared with other existing methods. The model-based optimal control and optimal system are also helpful to analyze and understand data-driven deep learning methods. For example, the paper [101] reviewed and discussed how to establish deep learning theory from the perspective of optimal control and dynamic system. In summary, the combination of deep learning and control models is an important research direction in robot control, autonomous driving, flight control and other related fields. It is expected to overcome the shortcomings of the classical model-based control theory in practice and the shortcomings of the pure data-driven reinforcement learning method lacking model and theoretical guidance for designing strategies, trajectories, etc.

5 Discussion and perspective

This paper reviewed the model-driven deep learning in image inverse problems, including the optimization model-driven deep learning and statistical model-driven deep learning, and further introduced the closely related work of deep learning methods in PDE, control, etc. The model-driven deep learning method establishes a bridge between traditional mathematical modeling and modern data-driven learning, and brings the artificial intelligence techniques into traditional mathematical models in diverse applications such as imaging science, physics, control and other fields. It has potential to enforce the modern deep learning approach to gradually become a basic tool for solving natural and social science problems. However, to achieve these above expectations, researchers on model-driven deep learning need to solve some theoretical, methodological, and application challenges.

Theoretical analysis of model-driven deep learning. Although model-driven deep learning integrates mathematical modeling of domain knowledge and data-driven learning in a principled way, it is interesting to theoretically analyze its advantages from the learning theory perspective by analyzing the hypothetical function space deduced by model-driven deep network. For example, compared with the standard deep neural network, are the model-driven deep networks relying on fewer parameters and less data to achieve the same or higher approximation and generalization ability? It is necessary to systematically analyze and compare the basic mathematical operations of model-driven and standard deep networks in a compositional function space. Intuitively, model-driven deep learning should be advantageous because the mathematical operations in its network architectures are derived by directly modeling the task itself, therefore should be more efficient to be learned to achieve the desired solution.

Investigation of model-driven deep learning in more diverse research fields. Nowadays, deep learning has been in a trend of becoming tools in various fields. In addition to the above mentioned applications such as imaging sciences, control, physics, model-driven deep learning should be more extendable to more extensive fields, such as molecular chemistry, astronomy, mechanics, geological sciences, etc., which deserve us to investigate. But in different applications, how to leverage domain knowledge to develop new model-driven deep learning approach is a cutting-edge and challenging research direction. For example, in geophysics for the inverse problem of petroleum exploration, it has specific physical domain knowledge. The modeling of these domain knowledge and incorporation of corresponding models to derive novel model-driven deep learning method is of great importance in geophysical applications.

Model-driven deep learning in an open and dynamic environment. In the image inverse problems, the model-driven deep learning is modeled in a static perspective without considering the dynamic environmental factors in practical applications. In many applications, such as communications and autonomous driving, the external environment is constantly changing and the input data are online and dynamic. In such open and dynamic environment, especially for control problem, it is a challenging task to combine the physical model and sequentially supplied data in the context of complex environment to design effective model-driven deep learning models.

In a summary, the interaction between modeling and deep learning is a promising research direction with the popularity of big data and artificial intelligence (AI). On the one hand, the models may inspire, analyze, stabilize, and even reformulate the deep network architectures and related theories. On the other hand, the deep network and AI techniques may revolutionize the applied mathematics with topics including inverse problems, imaging sciences, optimization, control, etc.

Acknowledgments

This work was supported by NSFC (11971373, 11690011, U1811461, 61721002) and National Key R&D Program 2018AAA0102201.

References

[1] Kyong Hwan Jin, Michael T Mccann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[2] Jo Schlemper, Jose Caballero, Joseph V. Hajnal, Anthony N. Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, 2018.

[3] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. ADMM-CSNET: A deep learning approach for image compressive sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(3):521–538, 2020.

[4] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[5] B. S. Reddy and B. N. Chatterji. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8):1266–1271, 1996.

[6] Leonid Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60:259–268, 1992.

[7] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total generalized variation. *Siam Journal on Imaging Sciences*, 3(3):492–526, 2010.

[8] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Non-local sparse models for image restoration. *2009 IEEE 12th International Conference on Computer Vision*, pages 2272–2279, 2009.

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016.

[10] Dongwook Lee, Jaejun Yoo, and Jong Chul Ye. Deep residual learning for compressed sensing MRI. *International Symposium on Biomedical Imaging*, pages 15–18, 2017.

[11] Mila Nikolova. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

[12] Paul Rodriguez and Brendt Wohlberg. Efficient minimization method for a generalized total variation functional. *IEEE Transactions on Image Processing*, 18(2):322–332, 2009.

[13] Weisheng Dong, Guangming Shi, Xin Li, Yi Ma, and Feng Huang. Compressive sensing via nonlocal low-rank regularization. *IEEE Transactions on Image Processing*, 23(8):3618–3632, 2014.

[14] Cédric Vonesch and Michael Unser. A fast multilevel algorithm for wavelet-regularized image restoration. *IEEE Transactions on Image Processing*, 18(3):509–523, 2009.

[15] Jeroen Verhaeghe, Dimitri Van De Ville, Ildar Khalidov, Yves DÁsseler, Ignace Lemahieu, and Michael Unser. Dynamic PET reconstruction using wavelet regularization with adapted basis functions. *IEEE Transactions on Medical Imaging*, 27(7):943–959, 2008.

[16] Julien Mairal, Michael Elad, and Guillermo Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008.

[17] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006.

[18] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. *Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[19] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. *International Conference on Machine Learning*, pages 689–696, 2009.

[20] Weisheng Dong, Lei Zhang, and Guangming Shi. Centralized sparse representation for image restoration. *International Conference on Computer Vision*, pages 1259–1266, 2011.

[21] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. *International Conference on Computer Vision*, pages 479–486, 2011.

[22] Stefan Roth and Michael J. Black. Fields of experts: a framework for learning image priors. *Computer Vision and Pattern Recognition*, 2:860–867, 2005.

[23] John Lafferty, Andrew Mccallum, and Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *International Conference on Machine Learning*, pages 282–289, 2001.

[24] Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[25] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015.

[26] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv: Computation and Language*, 2016.

[27] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.

[28] Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Journal of the American Medical Association*, 316 22:2402–2410, 2016.

[29] Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE Transactions on*

Medical Imaging, 37:1322–1332, 2018.
[30] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In *Neural Information Processing Systems*, 2012.
[31] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. *Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.
[32] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, et al. Photo-realistic single image super-resolution using a generative adversarial network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2016.
[33] Jian Sun, Zongben Xu, and Harry Shum. Image super-resolution using gradient profile prior. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
[34] Joan Bruna, Pablo Sprechmann, and Yann LeCun. Super-resolution with deep convolutional sufficient statistics. *arXiv: Computer Vision and Pattern Recognition*, 2015.
[35] Justin Johnson, Alexandre Alahi, and Fei-Fei Li. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, 2016.
[36] Hu Chen, Yi Zhang, Mannudeep K. Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, 36(12):2524–2535, 2017.
[37] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. Accelerating magnetic resonance imaging via deep learning. *2016 IEEE 13th International Symposium on Biomedical Imaging*, pages 514–517, 2016.
[38] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Deep ADMM-Net for compressive sensing MRI. In *Neural Information Processing Systems*, 2016.
[39] Tim Meinhardt, Michael Moeller, Caner Hazirbas, and Daniel Cremers. Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. *2017 IEEE International Conference on Computer Vision*, pages 1799–1808, 2017.
[40] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. *Computer Vision and Pattern Recognition*, pages 1671–1681, 2019.
[41] Zongben Xu and Jian Sun. Model-driven deep-learning. *National Science Review*, v.5(1):26–28, 2018.
[42] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *International Conference on International Conference on Machine Learning*, 2010.
[43] Christoph Vogel and Thomas Pock. A PRIMAL DUAL NETWORK FOR LOW-LEVEL VISION PROBLEMS. In *German Conference on Pattern Recognition*, 2017.
[44] Jian Sun and Marshall F. Tappen. Learning non-local range markov random field for image restoration. *Computer Vision and Pattern Recognition*, pages 2745–2752, 2011.
[45] Shuai Zheng, Sadeep Jayasumana, Bernardino Romeraparedes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip H. S. Torr. Conditional random fields as recurrent neural networks. *2015 IEEE International Conference on Computer Vision*, pages 1529–1537, 2015.
[46] Jian Sun and Marshall F. Tappen. Separable Markov random field model and its applications in low level vision. *IEEE Transactions on Image Processing*, 22:402–407, 2013.
[47] Jian Sun and Zongben Xu. Color image denoising via discriminatively learned iterative shrinkage. *IEEE Transactions on Image Processing*, 24:4148–4159, 2015.
[48] Pablo Sprechmann, Alexander M. Bronstein, and Guillermo Sapiro. Learning efficient sparse and low rank models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

37:1821–1833, 2012.

[49] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 39(6):1256–1272, 2017.

[50] Zichao Long, Yiping Lu, Xianzhong Ma, and Bin Dong. PDE-Net: Learning PDEs from data. *International Conference on Machine Learning*, pages 3208–3216, 2018.

[51] Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, 115(34):8505–8510, 2018.

[52] Anusha Nagabandi, Gregory Kahn, Ronald S. Fearing, and Sergey Levine. Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. *International Conference on Robotics and Automation*, pages 7559–7566, 2018.

[53] Yan Duan, Xi Chen, Rein Houthooft, John Schulman, and Pieter Abbeel. Benchmarking deep reinforcement learning for continuous control. *International Conference on Machine Learning*, pages 1329–1338, 2016.

[54] Jianchao Yang, Kai Yu, Yihong Gong, and Thomas S. Huang. Linear spatial pyramid matching using sparse coding for image classification. *Computer Vision and Pattern Recognition*, pages 1794–1801, 2009.

[55] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[56] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *Siam Journal on Imaging Sciences*, 2(1):183–202, 2009.

[57] Pablo Sprechmann, Alexander M. Bronstein, and Guillermo Sapiro. Learning robust low-rank representations. *ArXiv*, abs/1209.6393, 2012.

[58] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2004.

[59] Stephen Boyd, Neal Parikh, Eric Kingwah Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3:1–122, 2011.

[60] Yu Wang, Wotao Yin, and Jinshan Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78:29–63, 2015.

[61] Yin Zhang Junfeng Yang and Wotao Yin. A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288297, 2010.

[62] Xiaobo Qu, Yingkun Hou, Fan Lam, Di Guo, Jianhui Zhong, and Zhong Chen. Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator. *Medical Image Analysis*, 18(6):843–856, 2014.

[63] Zhifang Zhan, Jianfeng Cai, Di Guo, Yunsong Liu, Zhong Chen, and Xiaobo Qu. Fast multi-class dictionaries learning with geometrical directions in MRI reconstruction. *arXiv: Computer Vision and Pattern Recognition*, 2015.

[64] Christopher A. Metzler, Ali Mousavi, and Richard G. Baraniuk. Learned D-AMP: Principled neural network based compressive image recovery. *Neural Information Processing Systems*, pages 1772–1783, 2017.

[65] Xingyu Xie, Jianlong Wu, Guangcan Liu, Zhisheng Zhong, and Zhouchen Lin. Differentiable linearized ADMM. *arXiv: Learning*, pages 6902–6911, 2019.

[66] Kuanhong Cheng, Juan Du, Huixin Zhou, Dong Zhao, and Hanlin Qin. Image super-

resolution based on half quadratic splitting. *Infrared Physics & Technology*, 105:103193, 2020.

[67] Singanallur Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948, 2013.

[68] Felix Heide, Markus Steinberger, Yunta Tsai, Mushfiqur Rouf, Dawid Pajk, Dikpal Reddy, et al. FlexISP: a flexible camera image processing framework. *International Conference on Computer Graphics and Interactive Techniques*, 33:231:1–231:13, 2014.

[69] Thomas Brox, O. Kleinschmidt, and Daniel Cremers. Efficient nonlocal means for denoising of textural patterns. *IEEE Transactions on Image Processing*, 17(7):1083–1092, 2008.

[70] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[71] Dong Yang and Jian Sun. Proximal Dehaze-Net: A prior learning-based deep network for single image dehazing. In *European Conference on Computer Vision*, 2018.

[72] Raanan Fattal. Single image dehazing. In *International Conference on Computer Graphics and Interactive Techniques*, volume 27, page 72, 2008.

[73] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *Computer Vision and Pattern Recognition*, 2009.

[74] Robby T. Tan. Visibility in bad weather from a single image. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[75] R. Tan, Niklas Pettersson, and Lisa Petersson. Visibility enhancement for roads with foggy or hazy scenes. *2007 IEEE Intelligent Vehicles Symposium*, pages 19–24, 2007.

[76] Ji He, Yan Yang, Yongbo Wang, Dong Zeng, Zhaoying Bian, Hao Howard Zhang, Jian Sun, Zongben Xu, and Jianhua Ma. Optimizing a parameterized plug-and-play ADMM for iterative low-dose CT reconstruction. *IEEE Transactions on Medical Imaging*, 38:371–382, 2019.

[77] Alon Brifman, Yaniv Romano, and Michael Elad. Turning a denoiser into a super-resolver using plug and play priors. *2016 IEEE International Conference on Image Processing*, pages 1404–1408, 2016.

[78] Ulugbek S. Kamilov, Hassan Mansour, and Brendt Wohlberg. A plug-and-play priors approach for solving nonlinear imaging inverse problems. *IEEE Signal Processing Letters*, 24:1872–1876, 2017.

[79] Qianru Chen, Risheng Liu, Xin Fan, and Haojie Li. Plug-and-play correlation filters for visual tracking. 2018.

[80] S. Sreehari, Singanallur Venkatakrishnan, Brendt Wohlberg, Gregery T. Buzzard, Lawrence F. Drummy, Jeff P. Simmons, and Charles A. Bouman. Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Transactions on Computational Imaging*, 2:408–423, 2016.

[81] Stanley H. Chan, Xiran Wang, and Omar A. Elgendy. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3:84–98, 2017.

[82] Yu Sun, Brendt Wohlberg, and Ulugbek Kamilov. An online plug-and-play algorithm for regularized image reconstruction. *IEEE Transactions on Computational Imaging*, 5:395–408, 2019.

[83] Stanley H. Chan. Performance analysis of plug-and-play ADMM: A graph signal processing perspective. *IEEE Transactions on Computational Imaging*, 5:274–286, 2019.

[84] Xingyu Xie, Jianlong Wu, Zhisheng Zhong, Guangcan Liu, and Zhouchen Lin. Differen-

tiable linearized ADMM. 2019.

[85] Ernest K. Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Plug-and-play methods provably converge with properly trained denoisers. *International Conference on Machine Learning*, 2019.

[86] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[87] Xiangyang Lan, Stefan Roth, Daniel P. Huttenlocher, and Michael J. Black. Efficient belief propagation with learned higher-order Markov random fields. *European Conference on Computer Vision*, pages 269–282, 2006.

[88] Pushmeet Kohli and M. Pawan Kumar. Energy minimization for linear envelop MRFs. *Computer Vision and Pattern Recognition*, 2010.

[89] Todd K. Moon. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, 13(6):47–60, 1996.

[90] Matthew D. Hoffman, David M. Blei, Chong Wang, and John Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

[91] Adrian Barbu. Training an active random field for real-time image denoising. *IEEE Transactions on Image Processing*, 18(11):2451–2462, 2009.

[92] Javier Portilla, Vasily Strela, Martin J. Wainwright, and Eero P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, 2003.

[93] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[94] Klaus Greff, Sjoerd Van Steenkiste, and Jurgen Schmidhuber. Neural expectation maximization. *Neural Information Processing Systems*, pages 6691–6701, 2017.

[95] Risheng Liu, Zhouchen Lin, Wei Zhang, and Zhixun Su. Learning PDEs for image restoration via optimal control. *European Conference on Computer Vision*, pages 115–128, 2010.

[96] E Weinan, Chao Ma, and Lei Wu. Barron spaces and the compositional function spaces for neural network models. *arXiv: Learning*, 2019.

[97] Bao Wang, Zuoqiang Shi, and Stanley Osher. Resnets ensemble via the Feynman-Kac formalism to improve natural and robust accuracies. *Neural Information Processing Systems*, pages 1655–1665, 2019.

[98] Mengzhe Huang, Weinan Gao, Yebin Wang, and Zhongping Jiang. Data-driven shared steering control of semi-autonomous vehicles. *IEEE Transactions on Human-Machine Systems*, 49(4):350–361, 2019.

[99] Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):1–10, 2018.

[100] Yachien Chang, Nima Roohi, and Sicun Gao. Neural Lyapunov control. *Neural Information Processing Systems*, pages 3240–3249, 2019.

[101] Guanhorng Liu and Evangelos A. Theodorou. Deep learning theory review: An optimal control and dynamical systems perspective. *arXiv: Learning*, 2019.