

Douglas-Rachford Networks: Learning Both the Image Prior and Data Fidelity Terms for Blind Image Deconvolution

Raied Aljadaany Dipan K. Pal Marios Savvides
 Dept. of Electrical and Computer Engineering.
 Carnegie Mellon University
 {raljadaa, dipanp, marios}@andrew.cmu.edu

Abstract

Blind deconvolution problems are heavily ill-posed where the specific blurring kernel is not known. Recovering these images typically requires estimates of the kernel. In this paper, we present a method called Dr-Net, which does not require any such estimate and is further able to invert the effects of the blurring in blind image recovery tasks. These image recovery problems typically have two terms, the data fidelity term (for faithful reconstruction) and the image prior (for realistic looking reconstructions). We use the Douglas-Rachford iterations to solve this problem since it is a more generally applicable optimization procedure than methods such as the proximal gradient descent algorithm. Two proximal operators originate from these iterations, one for the data fidelity term and the second for the image prior. It is non-trivial to design a hand-crafted function to represent these proximal operators which would work with real-world image distributions. We therefore approximate both these proximal operators using deep networks. This provides a sound motivation for the final architecture for Dr-Net which we find outperforms the state-of-the-art on two mainstream blind deconvolution benchmarks. While doing so, we also find that Dr-Net is one of the fastest algorithms according to wall-clock times.

1. Introduction

The Blind Deconvolution Problem. Blind deconvolution problems are interesting inverse problems in image processing. A large part of their challenge is the fact that the kernel that caused the corruption is unknown. Assume a corrupted image y can be generated via convolving a clear image x with a kernel k . This can be written as:

$$y = k * x + \epsilon \quad (1)$$

where ϵ is an additive zero-mean white Gaussian noise and $*$ is the convolution operation. The problem of recovering the clean image is an ill-posed inverse problem. One



Figure 1. From the left to the right, blurry image [34], recovered image by Kupyn *et. al.* [25] and recovered image by our proposed Dr-Net. Dr-Net recovers sharper images with finer details.

approach to solve it is by assuming some prior (or a set of) on the image space and k is provided or estimated. Thus, the clean image can be approximated by solving the following optimization problem:

$$x^* = \arg \min_x \|y - k * x\|_2^2 + g(x) \quad (2)$$

where $\|y - k * x\|_2^2$ is the data fidelity term and g is an operator that defines some prior on the image space also called the image prior (e.g l_1 norm is used to promote sparsity). A good prior is important to recover a feasible and high-quality solution. Image priors are common in signal and image processing tasks such as inverse problems [30, 21] and these communities have spent considerable effort in hand designing suitable priors for signals [1, 42, 47]. However, the optimization problem in Eq. 2 is useful only if good estimates are available for both the image prior and the blurring kernel. Indeed, it has been shown that image recovery algorithms (based on the optimization problem in Eq. 2) fail when the solution space invoked by the assumed prior does not contain good approximations of the real data [13]. This also happens when the estimated kernel is not accurate [6].

Addressing the need for knowing the blurring kernel.

The data fidelity term in Eq. 2 in general can be denoted by $f(y, x, k)$ to emphasize its dependence on k . In accordance with the blind deconvolution problem, k is not known which makes $f(y, x, k)$ difficult to estimate let alone optimize. A reasonable and at times useful assumption addressing this is to assume that k is a random variable as in [4]. Now, the data fidelity term can be computed via marginalizing over the random variable k which makes it as a function of y and x exclusively. This eliminates the need of knowing k . The term $f(x, y)$ now only measures how likely it is to obtain the corrupted image y given a clean image x independent of k , which makes it a good candidate as an objective to be maximized. This approach however, presents a major issue as it requires that the prior density function of k to be known. This issue is one of the problems that our study addresses. Nonetheless, in the general case, the overall optimization problem including an image prior term can be written as the following:

$$x^* = \arg \min_{\hat{x}} f(y, x) + g(x) \quad (3)$$

The need for learning both the data fidelity and prior functions. There are two main challenges in utilizing the previous optimization formulation for deblurring and image recovery in general. 1) It is not trivial to correctly model the image prior using a hand-crafted function. Expressivity of the prior is critical in effective recovery of the image. Indeed, a number of previous works have proposed learning the image prior and found significant benefits. If the prior is incorrect or not expressive enough, the image recovered can potentially have major artifacts [55]. 2) A lesser addressed problem is that of modelling the prior distribution of the blurring kernel (e.g de-focusing kernels have distribution that is different from motion kernels) along with the exact noise distribution. This in turn leads to difficulties in modelling the data fidelity function itself. Current approaches assume useful functions as data fidelity terms despite limited expressivity. Some approaches represent the kernel distribution with the Laplacian distribution [7] since the kernel is assumed to have a sparse representation (e.g motion kernels). Nonetheless, a clean image will not be recovered correctly when the assumed prior over the blurring kernels is not expressive enough. Even in cases where the prior distribution of the kernel is known, it can get difficult to find a closed form of $f(y, x)$. For instance, in cases when the prior distribution is *not* a conjugate prior of the likelihood distribution. This makes hand-crafting a good objective for data fidelity a difficult task.

Our approach to learn the data fidelity and image prior. In this paper, we address these two problems simultaneously by modelling the proximal operators resulting from the data fidelity term and the image prior term with deep networks. To the best of our knowledge, this is the first study to model *both* the image prior and the data fidelity



Figure 2. From the left to the right, blurry image [34], recovered image by Kupyn *et. al.* [25] and recovered image by our proposed Dr-Net. Dr-Net recovers sharper images with finer details.

terms with deep networks making this one of the main contributions of this study. Learning the parameters of those networks lets us learn the data fidelity and the image prior functions indirectly.

The case for Douglas-Rachford iterations. In the usual case, it is difficult to find a closed form solution of Eq. 3. Iterative approaches such as gradient decent methods and proximal decent methods are usually used to solve these problems. However, both gradient decent based and proximal decent based methods require some conditions (e.g one or both the prior term and the data fidelity term need to be differentiable) for the optimization problem in Eq. 3. Importantly, most of these conditions are not required for the Douglas-Rachford algorithm making it a more generally applicable optimization procedure. This is our primary motivation to use Douglas-Rachford iterations to solve these problems. The Douglas-Rachford iterations applied to Eq. 3 lead to proximal operators for both the prior term and the data fidelity term. As discussed, for real-world data, it is difficult to know the exact form of both these terms. Our main contribution is to therefore use deep networks to model *both proximal operators* while having a straight-forward inference mechanism (a simple forward pass). We are motivated by the universal approximation theory [8] which states that the neural networks can model a very large class of functions. The final architecture learns both the prior and data fidelity terms in Eq. 3 from the corrupted data only without making any assumption about the prior distribution of the images or the blurring kernel. This framework leads to a large network whose overall architecture is inspired from and motivated by the Douglas-Rachford iterations. Indeed, as we find in our ablation studies, correct and sufficient network design following these iterations is critical to high-performing architectures.

Contributions. We make three main contributions. 1) We propose a network architecture for blind deconvolution inspired from the Douglas-Rachford optimization al-

gorithm called Dr-Net. 2) We replace the proximal operator in both the data fidelity term and the prior term in Douglas-Rachford algorithm with two different networks which firmly satisfy the non-expansive condition. This condition helps the network to be stable during the training and improve the performance during the testing. A phenomenon we confirm in our experiments. Further and unlike previous work, we motivate the use of multi-channel deep networks as solving the DR iterations while assuming the image as a non-linear weighted sum of an image basis. We show that this provides more theoretical backing to modelling optimization iterations with multi-channeled networks, which was previously lacking. 3) We evaluate the proposed model on the benchmark datasets and show that Dr-Net obtains state-of-the-art results in blind deblurring while being the fastest according to our benchmarked wall-clock times.

2. Related Work

Image Priors. Several methods have been proposed to using hand-designed image priors. Some of these methods assume that the image prior can be represented by Gaussian distribution or a heavy tailed distribution like the Laplacian with respect to a linear transformation [20, 3]. A few other approaches have been proposed to compute the prior from the data such as Principle Component Analysis (PCA) [54] and dictionary learning [42]. However, these data driven approaches are built by making some assumptions on the prior. For instance, low-rank PCA projection assumes that original image is low rank and transforms the image non-linearly into that space. Dictionary learning techniques assume a similar but slightly more expressive prior. However, such priors may not be applied for blind image deconvolution as they might produce a trivial solution without high-frequency details being intact.

Iterative Blind De-convolution. A few approaches have been proposed to recover images assuming that the kernel is *unknown* and therefore estimating it first [43, 14, 35]. These approaches add additional terms to the optimization problem in Eq. 2 which usually represents a prior on the kernel. Following this, the new optimization problem is solved with respect to both the recovered image and the estimated blurring kernel in an iterative fashion [24]. However, Eq. 2 becomes non convex if it is optimized for both the estimated image and kernel simultaneously [6]. Further, inaccurate estimation of the blurring kernels causes artifacts in the recovered images. Additionally, it is not trivial to assume an accurate prior distribution of the blurring kernel without any explicit knowledge. Nonetheless, images will not be recovered faithfully when the assumed prior is not correct which highlights the significance of accurate kernel modeling.

Image Priors via Deep Learning. There have been several recent approaches that employ deep networks to learn or model the image prior. In [52], the authors used a ConvNet trained on a large dataset. Recently, [40] used ADMM

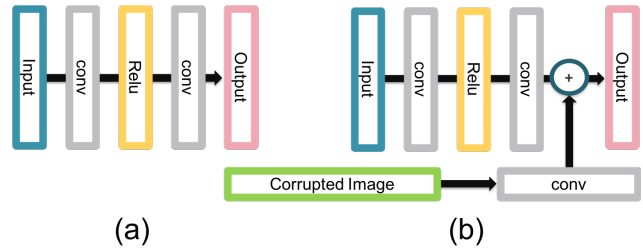


Figure 3. (a) shows the graphical representation of Γ_g while (b) shows Γ_f . These networks model the proximal operators in the Douglas-Rachford iterations for the prior term and the data fidelity term in Eq. 3. They form the main components of the proposed Dr-Block.

in an iterative fashion to solve Eq. 3 where the proximal operator of the prior term is replaced with a deep network trained using the GAN loss. The authors in [32] combined the idea of replacing the proximal operator with a denoising framework [19] and denoising convolutional neural networks [59]. However, all these approaches are not designed for *blind* deconvolution, since it takes the blurring kernel as an input in order to recover deblurred images. Our work on the other hand, focuses on the case when the blurring kernel is *not known apriori*.

Learning the Data Fidelity Term. In [10], they use Gaussian mixture model (GMM) to learn the data term for image denoising. In this approach, they assume that the GMM can approximate the data fidelity function when the noise cause is unknown. However as a major limitation, the blurring kernel is needed as an input besides the corrupted image. Here the prior on the image is that it is sparse with respect to the derivative, which is a hand crafted prior. On the other hand, we learn both the data and the prior term via deep networks, and importantly knowledge of the blurring kernel is not needed.

Blind Deblurring via Deep learning. Recently, [25] proposed DeblurGAN, which is based on the conditional GAN and the content loss for blind motion deblurring. Furthermore, the proposed approach in [34] removes these motion blurs from an image by using a multi-scale convolutional neural network. Another method which deals with motion blurring is [7], where a ConvNet was used to estimate the Fourier coefficients of the motion kernel and sharper reconstructions in the frequency domain were recovered. However, the relation between the *overall architectures* in these approaches and the image recover optimization problem in Eq. 3 is still unclear. In this work, the overall network architecture is motivated based on the application of the Douglas-Rachford iterations to solve Eq. 3.

3. Douglas-Rachford Networks for Inverse Problems

In this section, we first briefly review the proximal operator and traditional Douglas-Rachford splitting for optimiza-

tion. We then continue to present our proposed method for blind image deconvolution.

3.1. Proximal Operators

Let $h : R^n \rightarrow R$ be a function. The proximal operator of the function h with the parameter β is defined as

$$\text{prox}_{h,\beta}(x) = \arg \min_z \beta \|z - x\|_2^2 + h(z) \quad (4)$$

Proximal operators are useful in proximal algorithms [37] such as alternating direction method of multipliers (ADMM) [5], proximal gradient decent method [3] and the Douglas-Rachford algorithm [12]. These algorithms are considered special cases of fixed point algorithms [17]. It is also interesting to note that $\text{prox}_{h,\beta}(x^*) = x^*$ if x^* is a minimizing value of $h(x)$, which is another connection between fixed point algorithms and proximal algorithms. Further, when $\text{prox}_{h,\beta}$ is applied repeatedly, it will find a fixed point only if $\text{prox}_{h,\beta}$ is firmly non-expansive [2]. $\text{prox}_{h,\beta}$ is said to be firmly non-expansive when the following holds $\forall x, y$

$$\|\text{prox}_{h,\beta}(x) - \text{prox}_{h,\beta}(y)\|_2^2 \leq \|x - y\|_2^2 \quad (5)$$

This condition is also related to the Lipschitz continuity condition for the proximal operator with the Lipschitz constant being 1 under the Euclidean distance metric. We will take this feature into our consideration when we design Dr-Net Framework in the coming sections.

3.2. Douglas-Rachford Splitting

The Douglas-Rachford algorithm is an iterative scheme to minimize optimization problems where the objective function is split as the sum of two functions as in Eq. 3 [12]. It is also a generalization of the famous proximal gradient descent method (PGDM) [3]. However, PGDM requires that one of the functions in Eq. 3 to be differentiable, while this condition is not required in Douglas-Rachford splitting.

Douglas-Rachford splitting has been applied to solve nonlinear convex problems [27] before it was improved to deal with non-smooth convex problems [9]. Moreover, [49] show that Douglas-Rachford algorithms have a global convergence rate for specific classes of structured non-convex optimization problems. Analyzing the convergence rates of Douglas-Rachford (DR) based algorithms is beyond the scope of this paper. Nonetheless, [28] do present some results on convergence for the interested reader. Applying DR to Eq. 3 leads to the following update steps:

$$\begin{aligned} q^k &= \text{prox}_{f,\beta}(x^k) \\ z^k &= \text{prox}_{g,\beta}(2q^k - x^k) \\ x^{k+1} &= x^k + \lambda_k(z^k - q^k) \end{aligned} \quad (6)$$

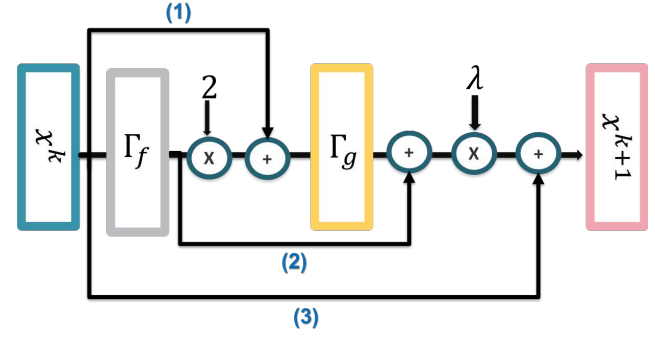


Figure 4. The graphical representation of the proposed Dr-Block which is inspired from the Douglas-Rachford iterations. The networks Γ_f and Γ_g represent the proximal operators in the iterations. Each block computes one iteration of the updates. When multiple of these are cascaded, they form the proposed Dr-Net. Also, the numbers in the blue brackets correspond to the nearest skip connection. The significance of these connections are investigated in the ablation study.

where λ_k is the step size and $\beta > 0$. If $\lambda_k = 2$ for all iterations, this approach is known as the Peaceman-Rachford splitting [39].

3.3. Dr-Net Framework

Our goal is to map the previous Douglas-Rachford update steps to a deep network architecture that consists of a *fixed* number of iteration (layers). We aim to utilize the power of deep network based image recovery combined with the Douglas-Rachford splitting method. This allows the network to perform well on a number of benchmarks as we find in our experiments.

Dr-Net models both the image prior *and* the data fidelity proximal operators using deep networks whose parameters are learned from data. This results in improved performance as compared to other hand-crafted approaches which solve the Eq. 3 formulation. The architecture of the network is based on the updating steps of Eq. 6. The deep networks (specifically, a convolutional neural network) model the prox operators and further satisfy the firmly non-expansive condition. We use a ConvNet for our image based application as the spacial reciprocity property of ConvNets have been shown to be very useful, especially when dealing with 2D visual data [44]. Although the Douglas-Rachford algorithm applies the same proximal operators (the same function) for all the iterations in Eq. 6, there is much to gain from having different proximal operators for every iteration as shown in studies utilizing deep networks for iterative algorithms such as ADMM-Net [46], ISTA-Net [57]. Thus, we define the proximal operators to be different for every iteration as different sub-networks.

The Douglas-Rachford Block. The updating steps of Eq. 6 show that two distinct proximal operators are needed. The first one, $\text{prox}_{f,\beta}(x)$ relates to the data fidelity term

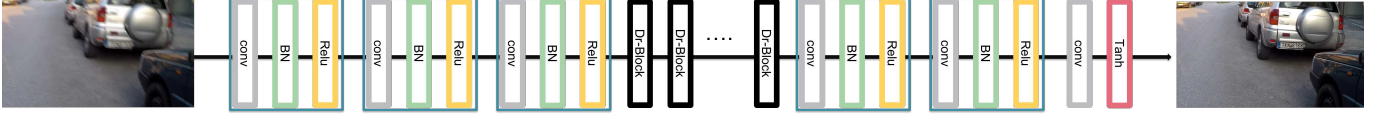


Figure 5. The proposed Dr-Net coupled with the pre and post processing networks. The core components for recovery here are multiple instances of the proposed Dr-Blocks.

while $prox_{g,\beta}(x)$ relates to the image prior term. Instead of setting by hand the regularization terms of f and g we use CNNs to learn their corresponding proximal operators. Thus, we represent each proximal operator with a ConvNet, namely $\Gamma_f(x)$ and $\Gamma_g(x)$. This network consists of two convolutional layers separated by a ReLU. Importantly, the kernel weights of the convolution layers are projected into the unit ball (the length of the vectorized filter weight ≤ 1). This ensures that the network satisfies the firmly non-expansive condition as we show in the following subsection. Since the data fidelity proximal operator $prox_{f,\beta}(x)$ is a function of both the corrupted image and the previous update step, the network $\Gamma_f(x)$ adds the corrupted image after a layer of convolution to the output of the network as shown in Fig. 3.

Non-expansive networks. The non-expansive condition for a function h states that under the Euclidean metric,

$$\|h(x) - h(y)\|_2^2 \leq \|x - y\|_2^2 \quad (7)$$

During the course of development of our networks, we found that enforcing the non-expansive conditions onto them improved the performance and stability of convergence. Enforcing this condition for convolutional layers only required the projection of each filter weight into the unit norm ball *i.e.* the norm ≤ 1 . We present the set of following results demonstrating that an entire ConvNet is firmly non-expansive under this constraint.

Lemma 3.1 *If $h(x)$ 1) is the vanilla rectified linear unit operation (ReLU) function, or 2) be a function that convolves x with a filter that has weights projected into the unit norm ball or 3) be a function that adds a constant to x , then $h(x)$ is firmly non-expansive.*

Lemma 3.2 *Let both $h(x)$ and $g(x)$ be firmly non-expansive, then $h(g(x))$ is firmly non-expansive.*

Lemmas 3.1 and 3.2 together imply that ConvNets with max unit norm filter weights are firmly non-expansive¹. We empirically find that non-expansive networks have more stable training.

The Dr-Net framework. Since the proximal operators are replaced with CNNs, the optimum weights of the convolution layers must be found through an optimization framework as in Eq. 8. Here the equality constraints would define

¹We provide proofs in the supplementary.

the network skip connections. Fig. 4 shows a graphical representing of the architecture in Eq. 8.

$$\begin{aligned} \min_{\Theta, \lambda} L(x_{gt}, x^S) \quad \text{s.t} \quad & q^k = \Gamma_{f,\theta}^k(x^k) \\ & z^k = \Gamma_{g,\theta}^k(2q^k - x^k) \\ & x^{k+1} = x^k + \lambda_k(z^k - q^k) \end{aligned} \quad (8)$$

Here L is the loss function, x_{gt} (ground-truth) is the clear image, x^S is the recovered image returned by the network after S iterations(layers) and $k = 0, \dots, S$. Note that the constraints in this optimization problem represent the Douglas-Rachford iterations. More importantly, the iterations address only single channel networks which are not expected to have satisfactory performance in practice. Modern neural networks on the other hand are multi-channelled. We now address this gap between the theoretical motivation and the practical architecture which has largely been ignored in prior arts [34, 25, 7].

Towards multi channel Dr-Net. We now extend the DR iterations and motivate a multi-channel deep network. Our approach towards this is to assume that the input image is the non-linear sum of some elements in a basis *i.e.* $x = h(\sum_i \alpha_i w_i)$ where w_i is the i -th element of the basis, h is a reasonable non-linearity and α is the weight vector. The assumption of an image being a combination of a basis is one that has been widely used in PCA [54], dictionary learning [42] and wavelet bases [11]. The optimization problem in Eq. 3 can be solved in parallel for each element i . Applying the Douglas-Rachford iterations to the new expression for x we arrive at parallel iterations and updates one for each i . Each of these parallel updates can be explained by a distinct channel in a distinct convolution layer in the deep network, whereas each iteration of all these updates is modelled by each layer in the deep network. This provides a coherent framework to theoretically motivate the more practical multi-channel deep-networks. We provide more details of this connection in the supplementary material.

Pre-processing Network. A good initial image x^0 helps to improve the performance of the proposed approach. The effect of the initial estimate is less apparent when the problem of interest is convex. However, due to the non-convex nature of our problem (training neural networks is non-convex) the initial estimate will have a more significant effect. We utilize another network that uses the corrupted im-

# of Dr-Blocks	PSNR (dB)	Time (sec)
3	27.26	49.1
5	30.40	57.4
7	30.71	75.8
9	30.93	96.1

Table 1. Effect of number of Dr blocks on performance.

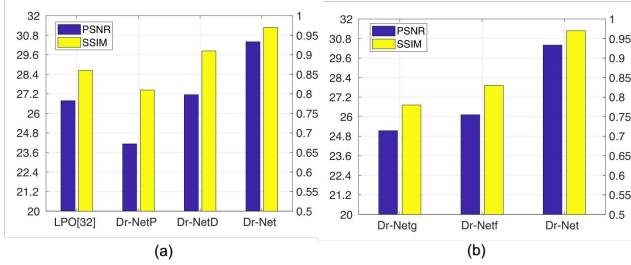


Figure 6. (a) Performance of our proposed Dr-Net and ablated networks without learning the image prior or data fidelity term. (b) Performance of our proposed Dr-net and ablated networks after eliminating one of the proximal operators. In both figures, the left axis represents PSNR in dB while the right one is SSIM.

Pre-Net	Post-Net	PSNR (dB)	Parameters (million)
✓	✓	30.40	6.7
✓		28.20	6.3
	✓	28.57	6.3
		27.12	5.9

Table 2. Pre and post networks ablation study. ✓ means that the correspond network is applied.

age y to provide the initial estimate x_0 and found that this solution works sufficiently well in practice. The proposed initialization network has a pyramid structure and it is consisted of three stages. Each stage consists of a convolution layer with 3×3 filters followed by a batch normalization layer and ReLU layer. The three layers have $\frac{N}{4}$, $\frac{N}{2}$, and N filters respectively where N is 256 in our experiments.

Post-processing Network. Once the updating steps in Eq. 8 are applied for S iterations, the output tensor size is the same image spatial size however with N channels. Thus, we need to add a sub-network which converts the tensor into an image. We call this component the post processing network. This sub-network is similar to the pre-processing network with 3×3 filters, however with $\frac{N}{2}$, $\frac{N}{4}$ and c channels (here c is either 1 or 3). Finally, the output is regulated by a Tanh layer to force the output to between 1 and -1. The final architecture of the proposed network is illustrated in Fig. 5.

3.4. The Loss Function

The loss function for Dr-Nets is formulated as a combination of the classical l_2 loss and GAN loss as shown in 9.

$$L(x, y) = \|x - y\|_2^2 + \mu L_{GAN}(x, y) \quad (9)$$

We find that $\mu = 0.01$ works well in practice. The l_2 error is known to be a good metric for image recovery, however

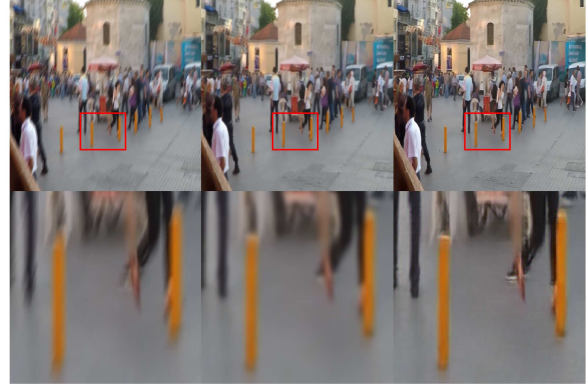


Figure 7. From the left to the right, blurry image [34], recovered image by our approach when MSE loss is used and recovered image by our proposed Dr-Net when GAN loss is added. GAN loss helps to recover sharper images.

the recovered images usually suffer from blurry artifacts. Indeed, the l_2 error tends to ignore high frequency components in the image.

To remedy this, we propose adding the GAN loss [16] to the loss function to more faithfully reconstruct high frequency elements. GANs have been well studied and used in practice to better model the space of real-world images with the help of a discriminator network. This discriminator forces the generator to only operate in the space of real-world images thereby better modelling it [26]. Moreover, GAN loss has been used in several images recovery tasks such as image super-resolution [26] since GAN loss forces the generator to recover photo-realistic textures from corrupted images. Among several techniques [31, 33] that are related to conditional GANs, we chose the Wasserstein-GAN with its gradient penalty [18] owing to its more robust tolerance towards the balance between the discriminator and generator leading to more stable training. Finally, we apply a discriminator network similar to [18]. Fig. 7 shows the advantage of adding GAN loss to the MSE loss where the blurry artifacts is reduced when GAN loss is used.

4. Ablation studies on Douglas-Rachford Nets

We begin the empirically evaluation of Dr-Nets through an extensive ablation study. For this, we created a testing data-set by convolving 68 filters that were generated by the proposed approach in supplementary with 68 images from [41]. The kernels used for training were generated randomly and were distinct from the one used in testing.

Network Architecture. The configuration of the proposed network is outlined in Fig. 5 with the number of Dr-Blocks (iterations) set to 5. All convolutional layers use filters of the size 3×3 . Since we want to have the sizes of all feature maps in the network to be similar to the input image size, we perform sufficient zero-padding.

Training Details. For training, we use 800 images from

(1)		✓		✓		✓		✓
(2)			✓	✓			✓	✓
(3)					✓	✓	✓	✓
PSNR (dB)	28.31	28.21	28.65	28.43	29.24	29.76	29.49	30.40

Table 3. Skip connection ablation study. ✓ means that the correspond skip connection in Fig. 4 is used.

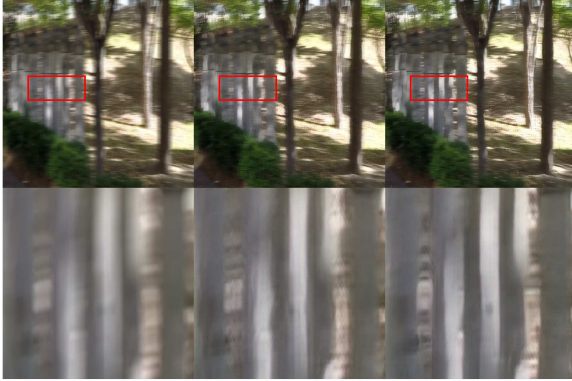


Figure 8. From the left to the right, blurry image [34], recovered image by Kupyn *et. al.* [25] and recovered image by our proposed Dr-Net. Dr-Net recovers sharper images with finer details.

the DIV2K dataset [50] as our training data. For each batch, we randomly sample 16 patches of the size 128×128 from 16 images (one from each image). We augment the training data by scaling, rotation and flipping. In total, we generate about 300 thousands patches. The input to the network were the blurred patches, and the ground truth output was set to be the clean versions of the patches. Note that there is no explicit kernel estimation in this process. The blurry images are generated using our proposed method in the supplementary. We utilize Pytorch [38] as our code base. Training for all ablation studies was conducted for 30 epochs (as opposed to 300 epochs for benchmarking on evaluation datasets) using Adam [22] on Pascal Titan-X GPUs. Each model took about 51 hours for completion. The learning rate was decayed exponentially from $1e-1$ to $1e-4$ (the learning rate is multiplied by $1e-1$ after one fourth of the epochs number) for both generator and discriminator for the 30 epochs.

4.1. Learning the prior and the data fidelity terms

In Dr-NetP, we replace the prior term with a standard hyper-Laplacian prior as in [10]. This is a strong hand-crafted prior since natural images are considered to be sparse with respect to the both the horizontal and vertical derivative. Thus, the proximal operator or Γ_g in Eq. 8 is represented with a soft threshold operator [3] w.r.t the bases in [10]. In Dr-NetD, we replace the data fidelity term with $\|y - Kx\|_2^2$ where K is estimated via [56]. This means that the proximal operator or Γ_f in Eq. 8 is represented with $[KK^T + \beta I]^{-1}(Ky + \beta x)$ where β is a tuning parameters and is found by grid search as in [32]. Also, we compared

against LPO [32] where the proximal operator of the prior term is replaced with a denoising convolutional neural networks (DnCNN) and K is also estimated via [56]. Thus, the proximal operator or Γ_g in Eq. 8 is represented with DnCNN and Γ_f is similar to the one in Dr-NetD. The results of this study are shown in Fig. 6(a). It is clear that learning *both* the prior and the data fidelity terms have improved the performance in both PSNR and SSIM compared with nets where one term is learned while the other is not.

4.2. Empirical Analysis of the Dr-Nets

Effect of Depth vs Performance. We explore the trade-off between the performance and compute time as the number of Dr-Blocks is varied. Table. 1 presents the results and suggests 5 Dr blocks is a reasonable choice for our study.

Effect of Pre and Post-Nets. We explore the advantage of using both pre and post networks. The results are shown in Table. 2. Although the number of parameters increases slightly (compared to the actual size) when both pre and post nets are employed, the performances improves significantly.

Effect of eliminating one of the proximal operators. Next, we explore the importance of exactly following the Douglas-Rachford iterations. Specifically, we ask if indeed we need both the proximal operators in Eq. 8 or we could eliminate one of the them. To test this, we created two ablated versions of Dr-Net. In the first one we replace Γ_g in Eq. 8 with a short circuit (simply remove Γ_g from Eq. 8) and call it Dr-Netg. We do the same with Γ_f and call it Dr-Netf. The results are shown in Fig. 6(b).

Effect of eliminating skip connections defined by the Douglas-Rachford iterations. Finally, we explore the importance of the skip connections which comes from the updating steps in Eq. 6 where there is three skip connections. In this study we observe the effects of removing or adding these connections as shown in Table. 3. Lowest reconstruction error is achieved when all the skip connections are applied as suggested by Eq. 6. We present additional ablation studies investigating the number of channels and blur kernel modeling in the supplementary.

5. Evaluation on Blind Deconvolution

To verify the efficacy of Dr-Net for blind image recovery, we extensively evaluate the proposed methods on two datasets, 1) GoPro test dataset [34] and 2) Kohler dataset [23]. To test the efficacy of the non-expansive condition, we trained another network where we replace projecting the convolution filter weights into the unit ball (firmly non-

Method	Gong	Whyte	Xu	Sun	Pan	Liu	Nah	Kupyn	Zhang	Tao	Dr-Net(IN)	Dr-Net
PSNR (dB)	26.06	24.53	20.30	25.31	23.52	25.75	28.49	28.7	29.19	<i>30.26</i>	28.20	30.35
SSIM	0.8632	0.8458	0.7407	0.8511	0.8336	0.8654	0.9165	0.958	0.9306	0.9342	0.902	0.961

Table 4. Peak signal-to-noise ratio (PSNR) and the structural similarity measure (SSIM) on the GoPro test dataset [34]. Dr-Net (our proposed approach) outperforms other recent algorithms to obtain state-of-the-art on this test set both in terms of PSNR and SSIM. Best and the second best performance indicated by **bold** and *italics*. Due to space constraints the citations are provided in GoPro test set description in the experiments section.

Method	Whyte	Xu	Sun	Nah	Kupyn	Tao	Dr-Net(IN)	Dr-Net
PSNR (dB)	27.03	27.47	25.22	26.48	25.86	26.75	25.12	<i>27.20</i>
SSIM	0.809	0.811	0.773	0.807	0.802	<i>0.837</i>	0.792	0.865

Table 5. Peak signal-to-noise ratio (PSNR) and the structural similarity measure (SSIM) of the Kohler dataset [23] when our approach is applied against the state of the art algorithms. Best and the second best performance indicated by **bold** and *italics*.

Method	Gong	Whyte	Xu	Sun	Pan	Nah	Kupyn	Zhang	Tao	Dr-Net(IN)	Dr-Net
Time (sec)	1500	700	3800	1500	2500	15	2.9	1.4	1.6	1.9	1.2

Table 6. Wall-clock run times (in seconds) for all algorithms on recovering an image of size of 720×1280 pixels. Our proposed method, Dr-Net attains the fastest processing time compared to previous state-of-the-arts.

expansive condition) with layer called the instance normalization layer [51]. This model, called Dr-Net(IN), is a useful baseline against the non-expansive criteria. We follow the same training producer for both models ².

GoPro test dataset: The GoPro test dataset is generated by taking the average of several frames from videos that are captured with high frame rate cameras. Averaging these frames creates blurry images due to the preexisting motion. This procedure has the advantage of providing near perfect ground truths where there is almost no blur, while the averaging provides the motion (and sometimes optical) blurred images. One important point to note is that since this is a real-world dataset, the same blurring kernel in the averaged blurry image is *not* applied homogeneously throughout the image. In other words, some part of the image might have more blur than the rest, thereby the blur is spatially heterogeneous. We compare our models with state of the art models, Gong [15], Whyte [53], Xu [55], Sun [45], Pan [36], Liu [29], Nah [34], Kupyn [25], Zhang [58] and Tao [48].

Results. Table. 4 showcases the results of this experiment. We find that our approach Dr-Net with the non-expansive constraint outperforms all the other methods in term of PSNR and SSIM. Further, we observe that the non-expansive Dr-Net outperforms even Dr-Net(IN) (with instance normalization) which provides even more justification for incorporating the firmly non-expansive condition apart from theoretical justification. Finally, Fig. 1, 2 and 8 show some visual results of our approach compared with the recent work of [25] on this dataset. We provide additional examples in the supplementary.

Kohler dataset: This dataset consists of 48 images that are generated by convolving 12 kernels with 4 images. Importantly, and in contrast to the GoPro test set, the convolution of the blurring kernels with the entire image ensures

equal and spatially homogeneous blur at *all* parts of the image. In [23], they record and analyze real camera motion and generate motion kernels which simulate that motion. For this test we compare against Whyte [53], Xu [55], Sun [45], Nah [34], Kupyn [25] and Tao [48].

Results. Table. 5 showcases the results of this experiment. We find that Dr-Net attains a high PSNR but fails to obtain state-of-the-art in terms of PSNR. Xu [55] obtains a high 27.47 dB. Nonetheless, Dr-Net obtains state-of-the-art in terms of SSIM with 0.865 with Xu [55] following closely at 0.811. Note that Xu [55] obtains a low 20.29 dB on the GoPro test set whereas Dr-Net obtains 29.21 dB. This is probably because Xu [55] requires a single blur kernel estimate for the entire image. For spatially homogeneous blur such as the blur in the Kohler test set, this is well-suited. However, GoPro is a real-world test set with spatially heterogeneous blur, thereby forcing a single kernel estimate for the entire image is not the ideal approach. This leads to the poor performance of Xu [55] on GoPro. This also helps demonstrate that Dr-Net does not suffer from this problem and can deal with spatially heterogeneous blur well (it obtains state-of-the-art on the GoPro test set). Further, we better understand the limitations of other approaches when we incorporate wall-clock run times into account. Table. 6 shows the wall-clock run times of all methods. Clearly, Dr-Net attains the fastest processing time of just 1.2 secs for a 720×1280 image, compared to about 3,800 secs for Xu [55] which retains state-of-the-art in terms of PSNR for the Kohler dataset.

6. Conclusion

We find that Douglas-Rachford iterations within Dr-Net can solve blind image deconvolution problems. We introduce novel aspects such as modelling both data fidelity and image prior proximal operators with ConvNets. Dr-Net obtains SOTA results according to SSIM while being the fastest according to wall clock times. DR iterations applications to other areas within deep learning seem promising.

²We found that the models do not converge when neither of the weight normalization or instance normalization is used.

References

- [1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Transactions on image processing*, 1(2):205–220, 1992.
- [2] H. H. Bauschke, S. M. Moffat, and X. Wang. Firmly nonexpansive mappings and maximally monotone operators: correspondence and duality. *Set-Valued and Variational Analysis*, 20(1):131–153, 2012.
- [3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.
- [4] J.-C. Belfiore and E. Viterbo. Approximating the error probability for the independent rayleigh fading channel. In *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, pages 362–362. IEEE, 2005.
- [5] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.
- [6] P. Campisi and K. Egiazarian. *Blind image deconvolution: theory and applications*. CRC press, 2016.
- [7] A. Chakrabarti. A neural approach to blind motion deblurring. In *European Conference on Computer Vision*, pages 221–235. Springer, 2016.
- [8] T. Chen and H. Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995.
- [9] P. L. Combettes and J.-C. Pesquet. A douglas–rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, 2007.
- [10] J. Dong, J. Pan, D. Sun, Z. Su, and M.-H. Yang. Learning data terms for non-blind deblurring. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 748–763, 2018.
- [11] D. L. Donoho and J. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455, 1994.
- [12] J. Eckstein and D. P. Bertsekas. On the douglas–rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, 1992.
- [13] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.
- [14] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *ACM transactions on graphics (TOG)*, volume 25, pages 787–794. ACM, 2006.
- [15] D. Gong, J. Yang, L. Liu, Y. Zhang, I. D. Reid, C. Shen, A. Van Den Hengel, and Q. Shi. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *CVPR*, volume 1, page 5, 2017.
- [16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [17] A. Granas and J. Dugundji. *Fixed point theory*. Springer Science & Business Media, 2013.
- [18] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems*, pages 5767–5777, 2017.
- [19] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, et al. Flexisp: A flexible camera image processing framework. *ACM Transactions on Graphics (TOG)*, 33(6):231, 2014.
- [20] P. Hoeher, S. Kaiser, and P. Robertson. Two-dimensional pilot-symbol-aided channel estimation by wiener filtering. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 3, pages 1845–1848. IEEE, 1997.
- [21] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE transactions on pattern analysis and machine intelligence*, 32(6):1127–1133, 2010.
- [22] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [23] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *European Conference on Computer Vision*, pages 27–40. Springer, 2012.
- [24] D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE signal processing magazine*, 13(3):43–64, 1996.
- [25] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. *CVPR*, 2018.
- [26] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, volume 2, page 4, 2017.
- [27] J. Lieutaud. *Approximation d’opérateurs par des méthodes de décomposition*. PhD thesis, 1969.
- [28] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.
- [29] S. Liu, J. Pan, and M.-H. Yang. Learning recursive filters for low-level vision via a hybrid neural network. In *European Conference on Computer Vision*, pages 560–576. Springer, 2016.
- [30] S. Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [31] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley. Least squares generative adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2813–2821. IEEE, 2017.
- [32] T. Meinhardt, M. Möller, C. Hazirbas, and D. Cremers. Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. In *IEEE International Conference on Computer Vision*, pages 1781–1790, 2017.
- [33] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

- [34] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, volume 1, page 3, 2017.
- [35] J. Pan, W. Ren, Z. Hu, and M.-H. Yang. Learning to deblur images with exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
- [36] J. Pan, D. Sun, H. Pfister, and M.-H. Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, 2016.
- [37] N. Parikh, S. Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.
- [38] A. Paszke, S. Gross, S. Chintala, and G. Chanan. Pytorch, 2017.
- [39] D. W. Peaceman and H. H. Rachford, Jr. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for industrial and Applied Mathematics*, 3(1):28–41, 1955.
- [40] J. Rick Chang, C.-L. Li, B. Póczos, B. Vijaya Kumar, and A. C. Sankaranarayanan. One network to solve them all—solving linear inverse problems using deep projection models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5888–5897, 2017.
- [41] S. Roth and M. J. Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205, 2009.
- [42] R. Rubinstein, A. M. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.
- [43] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *Acm transactions on graphics (tog)*, volume 27, page 73. ACM, 2008.
- [44] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [45] J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 769–777, 2015.
- [46] J. Sun, H. Li, Z. Xu, et al. Deep admm-net for compressive sensing mri. In *Advances in Neural Information Processing Systems*, pages 10–18, 2016.
- [47] J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [48] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8174–8182, 2018.
- [49] A. Themelis and P. Patrinos. Douglas-rachford splitting and admm for nonconvex optimization: tight convergence results. 2018.
- [50] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1110–1121. IEEE, 2017.
- [51] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Instance normalization: The missing ingredient for fast stylization. corr. *arXiv preprint arXiv:1607.08022*, 2016.
- [52] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. *arXiv preprint arXiv:1711.10925*, 2017.
- [53] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. *International journal of computer vision*, 98(2):168–186, 2012.
- [54] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
- [55] L. Xu, S. Zheng, and J. Jia. Unnatural l0 sparse representation for natural image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1107–1114, 2013.
- [56] X. Xu, J. Pan, Y.-J. Zhang, and M.-H. Yang. Motion blur kernel estimation via deep learning. *IEEE Transactions on Image Processing*, 27(1):194–205, 2018.
- [57] J. Zhang and B. Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1828–1837, 2018.
- [58] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R. W. Lau, and M.-H. Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2521–2529, 2018.
- [59] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.