

Wiener Guided DIP for Unsupervised Blind Image Deconvolution

Gustav Bredell¹ Ertunc Erdil¹ Bruno Weber² Ender Konukoglu¹
¹Computer Vision Laboratory, ETH Zürich ²University of Zürich
 {gustav.bredell, ertunc.erdil, ender.konukoglu}@vision.ee.ethz.ch

Abstract

Blind deconvolution is an ill-posed problem arising in various fields ranging from microscopy to astronomy. Its ill-posed nature demands adequate priors and initialization to arrive at a desirable solution. Recently, it has been shown that deep networks can serve as an image generation prior (DIP) during unsupervised blind deconvolution optimization, however, DIP’s high frequency artifact suppression ability is not explicitly exploited. We propose to use Wiener-deconvolution to guide DIP during optimization in order to better leverage DIP’s ability for blind image deconvolution. Wiener-deconvolution sharpens an image while introducing high-frequency artifacts, which are reproduced by DIP with a delay compared to low-frequency features and sharp edges, similar to what has been observed for noise. We embed the computational process in a constrained optimization problem together with an automatic kernel initialization method and show that the proposed method yields higher performance and stability across multiple datasets.

1. Introduction

Blurred images can occur due to a number of factors ranging from camera motion to the optical set-up. In many cases generating a blur image, I_B , can be modeled as the convolution of a sharp image, I_S , with a kernel k as shown in Eq. (1)

$$I_S * k + \eta = I_B, \quad (1)$$

where η is noise and $*$ is convolution operator. If k is known then a deconvolution can be performed to recover the original sharp image and this is referred to as non-blind deconvolution. However, often in practice k is not known and needs to be estimated together with I_S . The ill-posed nature of this *blind deconvolution* problem requires strong priors on the sharp image and/or kernel to avoid having estimates that do satisfy Eq. (1), but are not the sharp version of the given blurred image. One possible formulation of the optimiza-

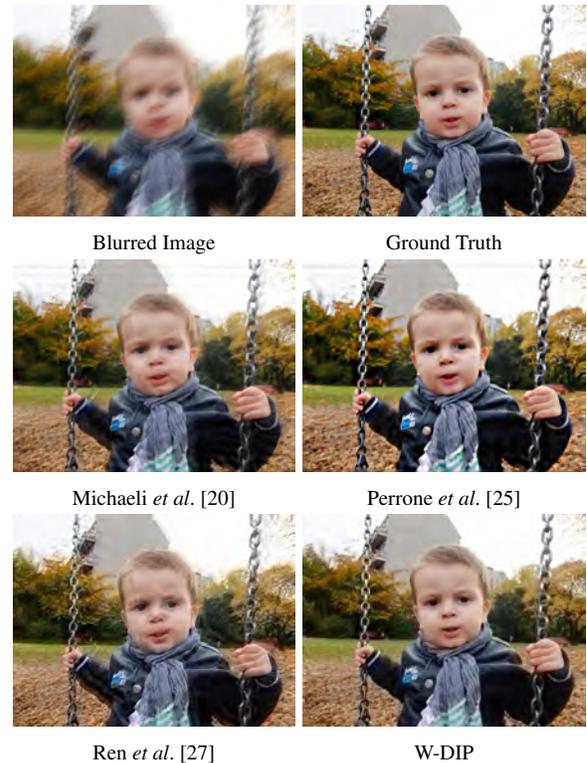


Figure 1: Motivating example showing that while the existing methods produces artifact and blurry results especially around the trees in the background, W-DIP produces the most similar result to the ground truth image.

tion is Eq. (2).

$$\min_{I_S, k} \|I_S * k - I_B\|^2 + R(I_S) + R(k) \quad (2)$$

where $R(I_S)$ and $R(k)$ are priors on the image and kernel, respectively. A vast variety of priors have been suggested to regularize the problem in Eq. (1) in such a way that a solution in a space of interest is found [11, 5, 15, 36, 40, 22]. Common priors include a sparsity constraint on the image gradients and a L_2 -norm on k to avoid the trivial solution. On top or instead of these common priors, recent works suggest more advanced terms such as dark channel prior [23]

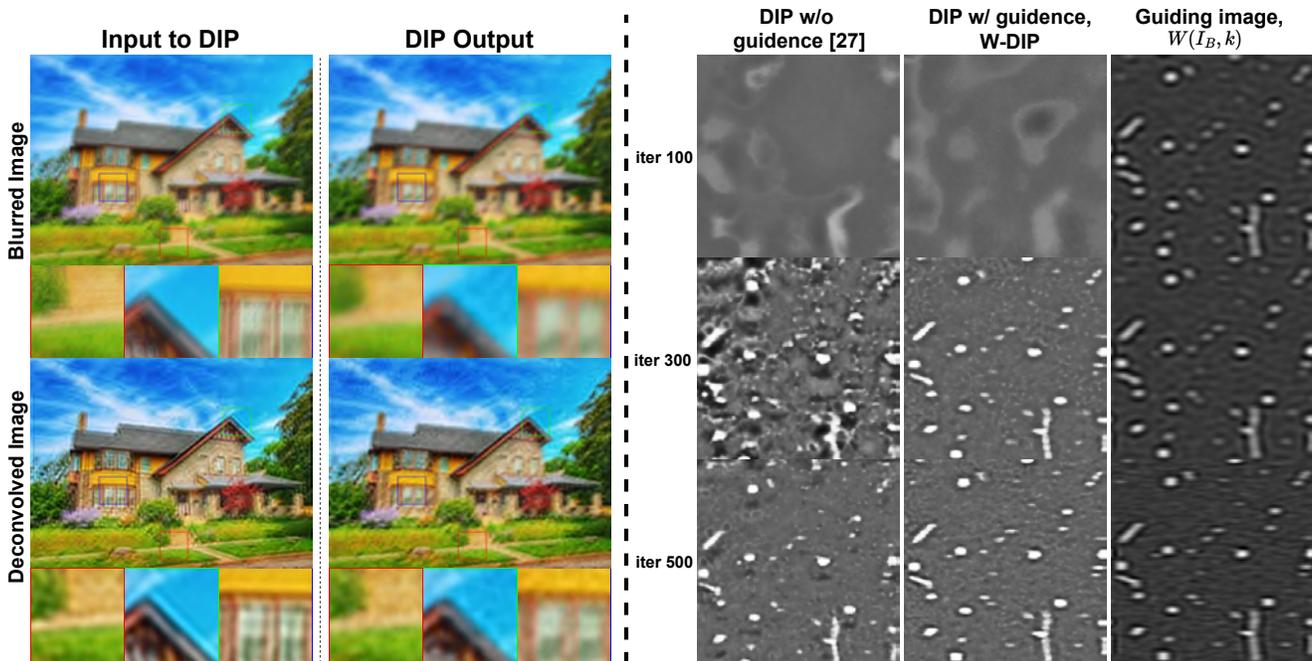


Figure 2: (Left) Visual motivation, best visible when zoomed in, where both a blurred image, top row, and an image that was deconvolved with a standard Gaussian (not the ground truth kernel), bottom row, are reconstructed by DIP. It can be seen that the reconstruction of the blurry image stays blurry, whereas for the deconvolved image high-frequency artifacts are removed leading to a sharper image. (Right) The optimization for blind deblurring using DIP is shown without, Ren *et al.* [27], and with guidance, W-DIP. The deconvolved image used as guidance is shown in the right most column. It can be seen that the guidance prevents a sub-optimal solution as the iteration number increases. In addition, the guiding image’s quality improves along with that of the sharp image estimate.

or edge-based patch priors [30]. While these carefully constructed priors are powerful, they are still limited in fully encapsulating the properties of sharp images.

Another approach that has become popular in recent years is learning the mapping from a blurred image to its sharp counter-part. If a large dataset with matching sharp images and their respective blurred version are available in a domain, a network can be trained to learn a successful mapping within the domain [12, 39, 31]. There are also methods that combine the learned mappings with classical priors as additional losses [38], or learn the priors and incorporate them into classical blind deconvolution frameworks [17, 1]. However, the success of these approaches are often limited to the domain of training images.

A recent interesting insight by Ulyanov *et al.* [32] shows that neural network architectures have an intrinsic image generation prior, referred to as “Deep Image Prior” (DIP), that can be exploited for image restoration tasks. It was shown that certain convolutional neural networks can reproduce noise, high frequency parts of an image, with an iteration lag compared to low frequency parts of the image. In previous works, the suppression of high-frequency artifacts by DIP has been used to improve the restoration performance for inverse problems [19, 2].

However the current adaption of DIP for blind deconvolution does not explicitly leverage the ability of high frequency artifact suppression. DIP was used in blind image deconvolution by Ren *et al.* [27], by having a convolutional neural network (CNN) that generate the estimate for the sharp image, i.e., DIP, and a fully connected network (FCN) that estimates the kernel. By jointly optimizing both networks, impressive blind deconvolution results were obtained. While authors noted basic regularization terms can be added in the article, the best results are possibly obtained using only the data-fitting term in Eq. (2) as this version was released in the published code. The major drawback in using DIP only through a data-fitting term or by just adding inappropriate regularization terms, is that DIP alone does not produce a sharp image from a blurred image given as input, as we demonstrate with an example on the left side of Fig. 2. On the contrary, DIP fits faster to blurred images compared to sharp images as shown in [9]. Therefore, DIP-based blind deconvolution without a suitable prior suffers from two major issues: (i) performance fluctuations across multiple runs and initialization, (ii) sensitivity to critical hyper-parameters that are difficult to predetermine, in particular the size of the kernel to be estimated as we further elaborated in Tab. 2. Kotera *et al.* [9] proposed to improve

upon [27] by initializing the kernel generator with a Gaussian and using multi-scale processing, whereas Li *et al.* [16] proposed an architecture change as well as sampling approach. However, both methods do not explicitly leverage the strength of DIP in removing high frequency artifacts for blind image deconvolution.

In this work we want to leverage the key abilities of DIP by exploiting the ability of high-frequency artifact suppression for blind image deconvolution. To the left of Fig. 2 we show in the bottom row that high frequency artifacts introduced by Wiener-deconvolution [34] are initially suppressed by DIP. Thus the deconvolved image can serve as a useful guidance for the image generator. The advantage of obtaining a sharpened version of the blurry image with Wiener-deconvolution, in contrast to other image sharpening methods such as unsharp masking [26], is that the sharpening is linked to the kernel that generated the blurry image.

Contributions: Building on this insights, we propose Wiener Guided Deep Image Prior (W-DIP), a blind deconvolution method that utilizes Wiener deconvolution of the blurry image to guide the image generator (DIP) through a constraint optimization model. With the guidance, we explicitly exploit the high frequency suppression ability of DIP to improve both the stability and performance of blind image deconvolution. To achieve this, we define an auxiliary kernel to apply Wiener-deconvolution to given blurry image and enforce the deconvolved image to be similar to the sharp image generated by the network. This constraint allows the image generator network to have a realistic sharp target image early on during optimisation. Additionally, the auxiliary kernel is constrained to be similar to the generated kernel to enforce the generated kernel to deconvolve the blurred image such that it is sharp. We solve the constrained optimization problem using Half Quadratic Splitting (HQS) method and optimize the network parameters and the auxiliary kernel in alternating fashion.

In Fig. 1, we present a motivating visual example to show the improved deblurring performance achieved by W-DIP compared to the existing methods in the literature and to the right of Fig. 2 the increased stability of the method is visually shown.

Our contribution is thus two-fold:

- We show that by using Wiener-deconvolution during optimization the high-frequency suppression ability of DIP can be utilized to obtain a highly stable and accurate blind image deblurring.
- We show that initialization of the auxiliary kernel is crucial and propose an initialization strategy that can be automatically adapted to any dataset.

We investigate our contributions by comparing our performance to that of state-of-the-art blind deconvolution algorithms on five different datasets: Levin [14], Sun [30],

Lai [13], synthetic microscopy[28] as well as a dataset that contains real-life blurry images without a known kernel [13]. In addition, we perform an ablation study to investigate the contribution of each element in our framework.

2. Method

2.1. Background

The first use of DIP for blind image deconvolution was proposed by Ren *et al.* [27] and achieved state-of-the-art performance. It consists of a CNN f_θ for generating the estimate of the sharp image and FCN g_ϕ for generating the estimate of the kernel. The optimization problem for the network parameters θ and ϕ is defined as

$$\min_{\theta, \phi} \|f_\theta(z_I) * g_\phi(z_k) - I_B\|^2 + TV(f_\theta(z_I)) \quad (3)$$

in [27] where z_I and z_k are random vectors sampled from $\mathcal{U}(0, 1)$. However, in the published code it has been suggested to remove the TV term and switch from the MSE loss to the structural similarity index measure (SSIM) after 1k iterations which is also confirmed by Kotera *et al.* [9].

2.2. Proposed method

A visualization of the proposed method can be seen in Fig. 3a. As motivated in Sec. 1, we constrain the sharp image estimate generated by f_θ with the deconvolved image obtained after Wiener-deconvolution. The constraint is given in Eq. (4)

$$\begin{aligned} \min_{\theta, \phi} \|f_\theta(z_I) * g_\phi(z_k) - I_B\|^2 \\ \text{subject to } \|f_\theta(z_I) - W(I_B, g_\phi(z_k))\|^2 = 0 \end{aligned} \quad (4)$$

where $W(I_B, k) = \mathcal{F}^{-1}\left\{\frac{|\mathcal{F}\{k\}|^2}{|\mathcal{F}\{k\}|^2 + C} * \frac{\mathcal{F}\{I_B\}}{\mathcal{F}\{k\}}\right\}$, \mathcal{F} is Fourier transform, and C is a constant.

To allow a flexible kernel initialization we re-write Eq. (4) by introducing an auxiliary kernel k with size $n \times n$ and using HQS method as follows:

$$\begin{aligned} \min_{\theta, \phi, k} \|f_\theta(z_I) * g_\phi(z_k) - I_B\|^2 \\ \text{subject to } \|f_\theta(z_I) - W(I_B, k)\|^2 = 0 \\ k = g_\phi(z_k) \end{aligned} \quad (5)$$

We convert the constrained optimization problem using the unconstrained one using the method of Lagrange multipliers

$$\begin{aligned} \min_{\theta, \phi, k} \|f_\theta(z_I) * g_\phi(z_k) - I_B\|^2 + \alpha \|f_\theta(z_I) - W(I_B, k)\|^2 \\ + \beta \sum_{i=1}^n \sum_{j=1}^n (g_\phi(z_k)_{i,j} - k_{i,j})^2 + \lambda \|k\|^2. \end{aligned} \quad (6)$$

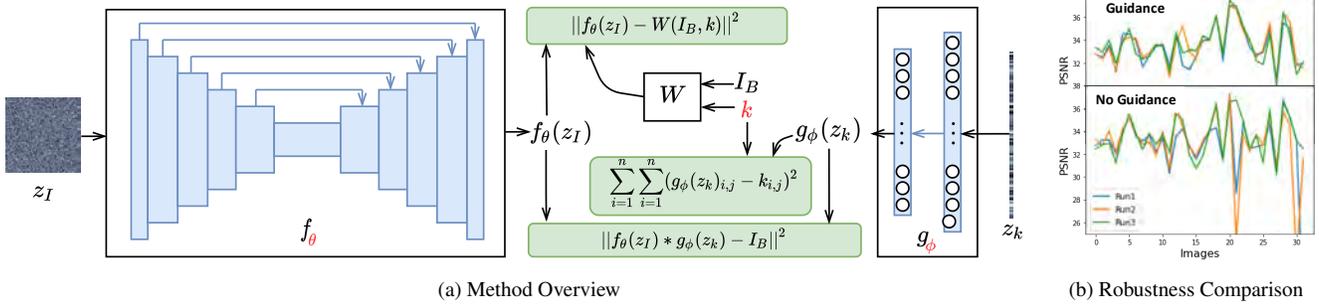


Figure 3: a.) Outline of the proposed method W-DIP. The generative network f_θ generates estimate of sharp image while g_ϕ generates kernel estimate. Additionally, an auxiliary kernel, k , and blur image are passed through Wiener deconvolution to enforce kernel to generate sharp images. The red variables are optimised to minimize Eq. (6). In b.) the performance of DIP blind deconvolution with (W-DIP) and without Wiener-deconvolution guidance is shown for three runs each on the Levin *et al.* [14] dataset. It can be seen that Wiener-deconvolution guidance stabilizes performance.

where α and β are penalty weights. Note that, in addition to the constraints in Eq. (5), we add regularization on k with weight λ . Also, note that we use sum of squared error, not mean square error (MSE) when including the second constraint in Eq. (5) to Eq. (6) since the kernel cost is size dependent, i.e. it tends to be lower for larger kernel sizes and higher for the smaller ones since kernels sum up to 1.

Finally, we optimize the network parameters (θ, ϕ) and the auxiliary variable k in an alternating fashion as suggested by HQS to minimize the loss function in Eq. (6), while keeping α, β and λ fixed. The parameters (θ, ϕ) are updated by keeping k fixed and vice versa:

$$\begin{aligned} \min_{\theta, \phi} & \|f_\theta(z_I) * g_\phi(z_k) - I_B\|^2 + \alpha \|f_\theta(z_I) - W(I_B, k)\|^2 \\ & + \beta \sum_{i=1}^n \sum_{j=1}^n (g_\phi(z_k)_{i,j} - k_{i,j})^2 \end{aligned} \quad (7)$$

$$\begin{aligned} \min_k & \alpha \|f_\theta(z_I) - W(I_B, k)\|^2 \\ & + \beta \sum_{i=1}^n \sum_{j=1}^n (g_\phi(z_k)_{i,j} - k_{i,j})^2 + \lambda \|k\|^2 \end{aligned} \quad (8)$$

Note that since the kernel size estimate is normally larger than the actual kernel to avoid restricting the kernel to an insufficient size, k and $g_\phi(z_k)$ could represent the same kernel, but a shifted version to each other. Although both kernel estimates are desirable and satisfy the first and the second terms in Eq. (6), the kernel matching term weighted by β produces large loss due to the translation between the kernels. To circumvent this we align the kernels before computing the kernel matching term. The process is described in more detail in the supplementary.

2.3. Kernel Initialization

The initialization of kernel k is an important parameter, since the starting point of an ill-posed optimization task often restrict the solution space. In the case that no prior knowledge of k is available it might be useful to obtain an estimate from the given blurry image I_B . We propose to achieve this by looking at the power spectral density (PSD) of I_B . The more centrally concentrated the PSD is, the larger the blur kernel may have been according to the convolution theorem. If no prior information is available on the shape of the k , we initialize it with a Gaussian and determine the variance of the Gaussian by looking at the PSD of an image as shown in Eqs. (9).

$$\sum_{i=1}^Q \sum_{j=1}^Q |\mathcal{F}(y)|_{i,j}^2 = T, \quad G_y = \mathcal{F}^{-1}[\mathcal{N}(0, Q^2)] \quad (9)$$

$$\sigma_k = \frac{1}{N} \sum_{i=1}^N \frac{\sigma(G_i)}{k_{size_i}}, \quad k_y = \mathcal{N}(0, \sigma_k \times k_{size_y}) \quad (10)$$

For each image we calculate the PSD and compute the distance from the center, Q , where a certain threshold, T , of the mass of the PSD is contained. We then construct a Gaussian in the Fourier Domain with the calculated distance as variance and apply an inverse Fourier transform to the constructed Gaussian. This results in an image-specific G_y that is the size of the image y . To facilitate optimization, by reducing the dimension of the kernel, G_y is cropped to the estimate of the kernel size, however the size of G_y will stay the same in the Fourier domain. To reduce the influence of image specific features an average variance σ_k over the dataset consisting of N images is calculated and in addition the variance is normalized by the kernel size estimate k_{size_i}

for the respective image as is shown in Eq. (10). The initial k_y is then constructed by scaling σ_k by the kernel size estimate as shown in Eq. (10).

3. Experimental Results

Datasets: We evaluate W-DIP on four widely used natural image datasets and a synthetic microscopy image dataset that we created. Levin *et al.* [14] introduced a dataset with 4 sharp images that was each convolved with 8 different kernels leading to 32 blurred images. Sun *et al.* [30] used the same blur kernels, but extended the sharp images from 4 to 80 leading to a total of 640 images. The images are also greyscale as in Levin *et al.* [14] but they are significantly larger. Lai *et al.* [13] consists of five different image categories, namely: manmade, saturated, text, people and natural. Each category contains five large sharp colour images corresponding to the category name and was blurred with large kernels ranging from 31×31 to 75×75 in size. This leads to a total of 100 blurred images. In addition, we qualitatively evaluate on real-world blurry images provided by Lai *et al.* [13], where the blurring process is unknown. Lastly, we introduce a dataset of synthetic microscopy images that was blurred with synthetic kernels. Four sharp synthetic images were taken from simulation of Schneider *et al.* [28] and cropped to 255×255 . They were blurred with one kernel from Levin *et al.* [14] and three other kernels, that can be found in the supplementary. The kernels are of similar size and shape of the point spread functions expected in two-photon microscopy [3]. In total this leads to 16 blurred images.

Implementation Details: The proposed method is implemented in PyTorch [24]. We used Adam [7] for both optimization steps in alternating optimization each with different learning rates and schedulers. For the first optimization step in Eq. (7), we set the initial learning rate to $1e^{-4}$ and decreased it by multiplying with 0.5 at the iterations 2K, 3K, and 4K following the implementation by Ren *et al.* [27]. In the second optimization step in Eq. (8), we used a slightly lower learning rate and a scheduler that depends on the kernel size n . The reason is that initial estimates of the sharp image generated by f_θ are worse than that obtained by Wiener-deconvolution since networks are randomly initialized and are given noise as input, thus θ and ϕ should rather adapt to k and $W(I_B, k)$. As the estimate from the first optimization gets more reliable, the adaptation speed of k is increased by increasing its learning rate. The larger the estimated kernel size n and thus blur of the image, the longer it takes Eq. (7) to arrive at a reliable solution, thus we include n as a parameter for learning rate adaptation. We start the second optimization step with the initial learning rate $1e^{-6}$, increase it by a factor of 10 after $70 \times (n/10)^{th}$ iteration and two more times after the next two $50 \times (n/10)^{th}$ iterations. The training took place

on the in-house GPU cluster mainly consisting of GeForce GTX TITAN X with 12GB memory.

For all evaluations we ran the optimization for 5k iterations, except for the microscopy images where 2k iterations was sufficient to reach peak performance. Furthermore, we set the weight parameters to: $\alpha = 1e^{-3}$, $\beta = 1e^{-4}$, $\lambda = 1e^{-3}$. These parameters were kept constant across all experiments unless otherwise stated and were determined by optimizing performance on four of Levin *et al.* [14] images. We set C in Wiener-deconvolution to a fixed value of 0.025 in all experiments. In addition, for all experiments of Ren *et al.* [27] and W-DIP, except on the Sun *et al.* [30] dataset due to its large size, the experiments were run three times to account for stochasticity. Mean values across the three runs are reported. We do not perform non-blind deconvolution for Ren *et al.* [27] and W-DIP as it has been shown that the generated image by the convolutional neural network is comparable to that of the non-blind deconvolved version [27]. We do not present any comparison to Kotera *et al.* [9] because the code is not publicly available and the evaluation metrics used in the paper are different than the ones that are commonly used in the literature.

Evaluation Metrics: To evaluate the performance of the deconvolution algorithms we make use of the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) between the estimated sharp image and the ground truth. For the Levin *et al.* [14] dataset we also include the Error Ratio (E-Ratio) metric [15]. In addition, for the microscopy images the vessel structures are of interest from a biological point of view. The structure of these vessels can be obtained by segmentation with Otsu thresholding [21] after a preprocessing step where extreme pixel values were removed. The overlap of the segmentation of the ground truth and estimated sharp image was then quantified with the Dice coefficient [29].

3.1. Framework Investigation

3.1.1 Ablation Study

To ensure that each component of the proposed framework is necessary we conduct an ablation study, where the weights for α , β and λ are set to zero individually. In addition, we set the initialization of k to uniform to evaluate the importance of its initialization. We perform the ablation study on a natural image dataset [14] as well as on the constructed microscopy dataset [28] as can be seen in Tab. 1.

It can be seen that the strongest loss in performance in both datasets occurs when the weight α , which controls the contribution of the image matching term, is set to zero. This drop empirically demonstrates the importance of the guidance of the Wiener-deconvolution in the proposed method. The performance loss is less pronounced when β , which controls the contribution of the kernel matching term, is set to zero. Even though it is subtle, the drop is observed for all

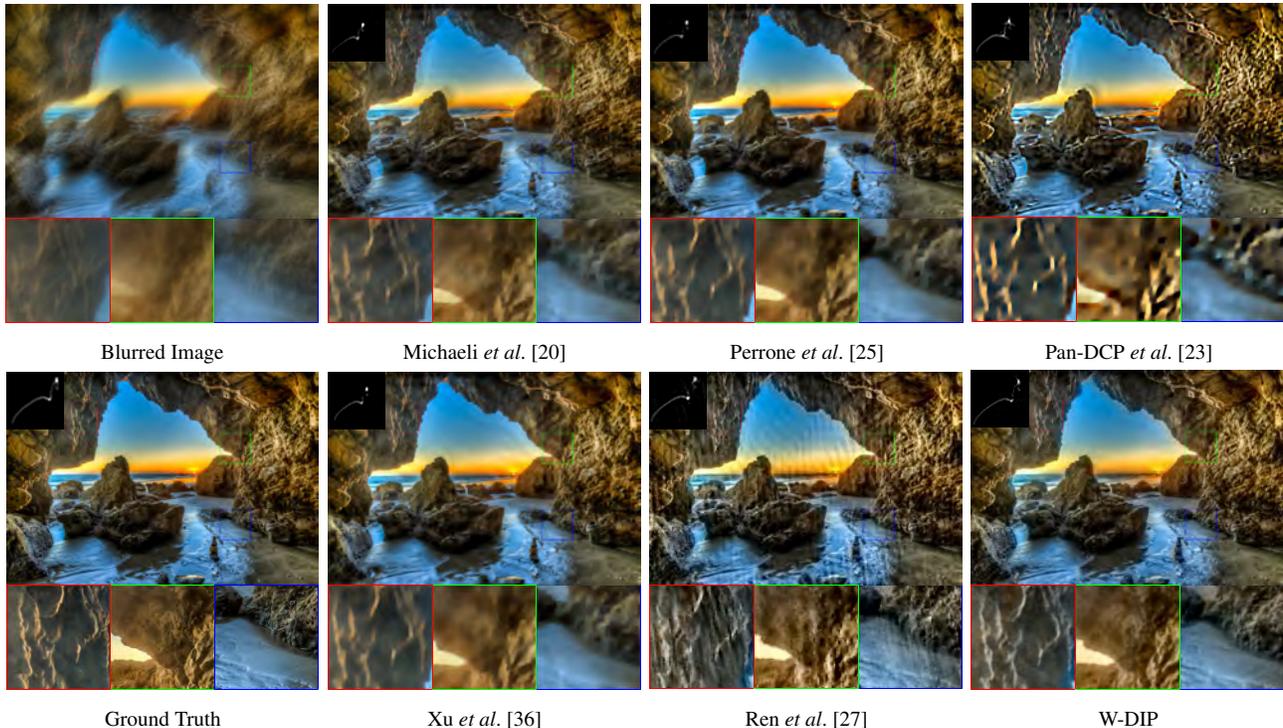


Figure 4: Visual results from the Lai *et al.* [13] dataset to compare W-DIP with five baselines. The results of W-DIP is sharper compared to other methods which is especially visible around the rocks.

the metrics, suggesting that this term also plays an important role in the method. For both microscopy and natural images, we see that the influence of the L2-norm on k in the second optimization step is minimal, nevertheless we include the regularization to ensure that the trivial delta function as kernel is avoided. Lastly, we also compare Uniform initialization of k ($k_0 = Uni$) with the proposed initialization and observe that the proposed initialization improves performance in all metrics.

3.1.2 Robustness Study

To investigate if the Wiener guidance can increase the stability of blind deconvolution while using DIP we conducted two experiments. In the first experiment we ran both Ren *et al.* [27] and our method for three runs each on the Levin *et al.* [14] dataset, utilizing different random seeds in each run. The performance fluctuations for both methods can be seen in Fig. 3b. It can be seen that the performance fluctuation is considerably less for our method compared to not using Wiener guidance as was done in Ren *et al.* [27]. This is also reflected in the average variance of the PSNR over the three runs. Whereas Ren *et al.* [27] has a PSNR variance of 7.18 our method has a variance of 2.71.

In a second experiment we simulated a realistic scenario where the estimated kernel size used in the optimization

Method	PSNR	SSIM	DICE%
W-DIP	33.58±2.7	0.9288± 3.9e ⁻⁴	-
$\alpha = 0$	32.74±8.7	0.9086±4.1e ⁻³	-
$\beta = 0$	33.44±4.4	0.9265±1.6e ⁻³	-
$\lambda = 0$	33.55±2.6	0.9298±3.6e⁻⁴	-
$k_0 = Uni$	32.99±9.3	0.9137±4.9e ⁻³	-
W-DIP	20.80±7.9	0.3977±2.8e⁻²	76.1±7.7
$\alpha = 0$	18.97±18	0.3502±3.6e ⁻²	65.9±12
$\beta = 0$	20.5±9.7	0.3833±2.9e ⁻²	74.97±9.2
$\lambda = 0$	20.71±9.0	0.3972±2.8e ⁻²	75.86±8.2
$k_0 = Uni$	20.39±9.6	0.3867±3.2e ⁻²	72.16±9.4

Table 1: Ablation study of the proposed framework on Levin *et al.* [14], top section, and microscopy dataset, bottom section. To evaluate the contribution of each component in our framework we investigate the change in performance when the component is left out, by setting the respective weight to zero. The variance is reported to the right of the mean.

is considerably larger than the actual kernel. A tight fitting box around the four kernels of the microscopy dataset would be 13×13 , 35×35 , 21×21 and 27×27 . Here, we extended these kernels with zero padding to 15×15 , 41×41 , 31×31 and 37×37 , respectively. In Tab. 2 the

performances for Ren *et al.* [27], W-DIP and W-DIP fine-tuned on the microscopy dataset are shown. Fine-tuning was done by optimizing the weight (α, β, λ) of our method on two of the blurry microscopy images and we obtained the weights $\alpha = 1e^{-2}$, $\beta = 1e^{-1}$, $\lambda = 1e^{-2}$. It can be seen when the kernel size estimate is “good”, referring to the first set of smaller kernel sizes, that DIP without Wiener guidance performs well, however if the kernel size estimate is too large, “bad”, a performance decrease is observed. For our method the performance decrease in the case of an inaccurate kernel size estimate is significantly less. In addition, fine-tuning the weights (α, β, λ) on two of the blurry images, can further improve the results.

Method	Size Estimate	PSNR	SSIM	DICE%
Ren <i>et al.</i> [27]	Bad k_{size_y}	19.19	0.354	67.5
	Good k_{size_y}	21.21	0.4143	83.27
	Difference	-2.02	-0.0603	-15.77
W-DIP	Bad k_{size_y}	20.8	0.3977	76.1
	Good k_{size_y}	21.69	0.4141	82.9
	Difference	-0.886	-0.0164	-6.8
W-DIP FT	Bad k_{size_y}	21.84	0.407	82.9
	Good k_{size_y}	22.12	0.4194	84.55
	Difference	-0.278	-0.0124	-1.65

Table 2: The performance of DIP deblurring without Wiener-deconvolution guidance, Ren *et al.* [27], our method and our method fine-tuned (W-DIP FT) on the microscopy dataset are shown for different kernel size estimates. “Good” and “Bad” indicates an accurate and too large kernel size estimate, respectively.

3.2. Quantitative and Qualitative Results

3.2.1 Levin *et al.* [14] Dataset

We evaluate our algorithm on the Levin *et al.* [14] dataset and compare it against other state-of-the-art unsupervised blind deconvolution methods [11, 5, 15, 36, 30, 40, 23, 37, 4, 27]. The results can be seen in Tab. 3. Yang and Ji [37] did not report their average E-Ratio thus this metric is missing.

Compared to other blind deconvolution methods W-DIP outperforms the baselines both in PSNR and especially with regard to Error Ratio. Since the second optimization step only requires optimization over a small parameter space, the computational cost is only marginally larger than that of Ren *et al.* [27] and still lower than other methods [23, 37]. The lower SSIM score compared with other baselines such as Pan *et al.* [23] can be explained by the SSIM metric that favors over-smoothed images as can be seen in the supplementary. Whereas the strong smoothing might be beneficial for small images, we see in the Lai *et al.* [13] dataset that it is not beneficial for more challenging datasets, where we

Method	PSNR	SSIM	E-Ratio	Time
k known	35.42	0.9576	1.000	-
Cho&Lee [5]*	30.29	0.8973	1.860	1.395
Levin <i>et al.</i> [15]*	30.90	0.9173	1.759	78.26
Krishnan <i>et al.</i> [11]*	29.87	0.8680	2.549	8.940
Xu <i>et al.</i> [36]*	31.27	0.9148	1.608	1.184
Zuo <i>et al.</i> [40]*	32.42	0.9353	1.334	11.00
Pan <i>et al.</i> [23]*	32.88	0.9386	1.232	295.2
Sun <i>et al.</i> [30]*	33.10	0.9385	1.363	191.0
Chen <i>et al.</i> [4]*	31.73	0.9236	1.407	65.20
Yang&Ji [37]*	32.04	0.912	-	354.0
Ren <i>et al.</i> [27]	33.01	0.9134	1.207	220.0
W-DIP	33.50	0.9270	1.087	280.6

Table 3: Quantative evaluation on Levin *et al.* [14] dataset, while comparing to other state-of-the art blind deconvolution approaches. The asterisk next to a method indicates that non-blind deconvolution [15] was performed with the found kernel.

outperform all baselines also with regard to the SSIM metric.

3.2.2 Lai *et al.* [13] Dataset

Whereas the Levin *et al.* [14] dataset is limited to 32 images and relatively small kernels ranging from 13×13 to 27×27 as well as small greyscale images, 255×255 , the Lai *et al.* [13] dataset contains large kernels and large colour images. To deal with colour images we utilize the same approach as Ren *et al.* [27] by splitting the images into the YCbCr channels respectively and only optimize for the Y channel. In addition, the dataset is split into distinct image categories, which we also inspect individually as can be seen in Tab. 4. We compare against other state-of-the-art blind deconvolution baselines [5, 35, 36, 20, 25, 22, 23, 37, 6, 27].

Whereas other baselines such as Yang and Ji [37] perform exceptionally well for a specific category, such as “Natural” in this case, but show a weaker performance in another category for example “Saturated”, our method produce state-of-the art results in every category and thus also leads to the best average performance when compared over all image categories. One example image can be seen in Fig. 4. The figure shows that our method creates the least amount of artifacts for the given image.

3.2.3 Sun *et al.* [30]

We also perform experiments on a large dataset by Sun *et al.* [30] containing 640 large images. The results shows that W-DIP is more stable, by producing a lower variance across all images, while also achieving a higher mean PSNR and SSIM as shown in Tab. 5. A qualitative comparison can be



Figure 5: Real-world blur example from the Lai *et al.* [13] dataset comparing our method with three baselines.

Method	Manmade	Saturated	Text	People	Natural	Average
Cho&Lee [5]*	16.35/0.3890	14.05/0.4927	14.87/0.4429	19.90/0.5560	20.14/0.5198	17.06/0.4801
Xu&Jia [35]*	19.23/0.6540	14.79/0.5632	18.56/0.7171	<u>25.32/0.8517</u>	<u>23.03/0.7542</u>	20.18/0.7080
Xu <i>et al.</i> [36]*	17.99/0.5986	14.53/0.5383	17.64/0.6677	14.40/0.8133	21.58/0.6788	19.23/0.6593
Michaeli <i>et al.</i> [36]*	17.43/0.4189	14.14/0.4914	16.23/0.4686	23.35/0.6999	20.70/0.5116	18.37/0.5181
Perrone <i>et al.</i> [25]*	17.41/0.550	14.24/0.5107	16.94/0.592	22.77/0.7347	21.04/0.676	18.48/0.6130
Pan-DCP <i>et al.</i> [23]*	18.59/0.5942	16.52/0.6322	17.42/0.6193	24.03/0.7719	22.60/0.6984	19.89/0.6656
Pan-L0 <i>et al.</i> [22]*	16.92/0.5316	14.62/0.5451	16.87/0.6030	23.36/0.7822	20.92/0.6622	18.54/0.6248
Yang&Ji [37]*	19.99/0.599	17.04/0.605	20.35/0.762	27.22/ 0.861	24.33/0.692	21.79/0.704
Kaufman&Fattal [6]	18.94/0.517	15.18/0.599	17.85/0.717	27.05/0.833	22.05/0.586	20.22/0.650
Ren <i>et al.</i> [27]	19.39/ 0.6660	<u>18.55/0.7066</u>	<u>23.21/0.7754</u>	<u>27.49/0.8172</u>	21.81/0.6585	22.09/0.7247
W-DIP	<u>19.51/0.6653</u>	18.66/0.7217	<u>22.48/0.7832</u>	28.35/0.8586	<u>23.47/0.7562</u>	22.49/0.7570

Table 4: Quantitative evaluation on Lai *et al.* [13] dataset. Performance of the algorithms are shown per image category with the PSNR value left and SSIM value right of the slash, respectively. The asterisk next to a method indicates that non-blind deconvolution [10, 33] was performed with the found kernel. Bold and underlined metrics are the best and second best result, respectively.

found in the supplementary. Liu *et al.* [18] did not report their variance nor SSIM, thus these metrics are missing.

Method	PSNR	SSIM
Liu <i>et al.</i> [18]	28.5	-
Ren <i>et al.</i> [27]	28.8±7.4	0.7383±8.6e ⁻³
W-DIP	29.07±6.7	0.7465±6.3e⁻³

Table 5: Mean PSNR and SSIM values with their respective variance on the Sun *et al.* [30] dataset.

3.2.4 Real-World Images With Unknown Blur

For real-world application the blurring operation is often not known. More complex blurring operations such as non-uniform blurring can occur [8]. To investigate the performance of W-DIP on these challenging images we compare the performance with baselines [20, 23, 27] in Fig. 5. It can be seen that our result is close to that of Ren *et al.* [27], while outperforming other baselines on this image. More qualitative comparisons can be found in the supplementary. In addition, we investigated to what extent specifically non-uniform blurring can be addressed with our method. Initial

experiments show that our method can also potentially handle small spatial varying blur and is shown in the supplementary.

4. Conclusion and Limitations

In this paper, we proposed an unsupervised blind image deconvolution method by guiding optimization of DIP with Wiener-deconvolution to obtain more stable and enhanced deblurring performance. We performed experiments on five different datasets and ablation studies have been used to validate the proposed components of our framework. The results show that W-DIP improves existing methods in terms of both image quality and training stability. Despite the fact that we achieve significant improvement in terms of stability, the computational time is high and could potentially be addressed by hardware improvements as well as future research into sophisticated training paradigms for networks. Furthermore, in this work we did not quantitatively investigate the performance on non-uniform blurred images. This could be a possible challenge for the current method, since uniform blurring operations are used in the data-fitting term and for deconvolution. Future work could try and extend the approach to different blur operations.

References

- [1] Muhammad Asim, Fahad Shamshad, and Ali Ahmed. Blind image deconvolution using deep generative priors. *IEEE Transactions on Computational Imaging*, 6:1493–1506, 2020.
- [2] Pasquale Cascarano, Andrea Sebastiani, Maria Colomba Comes, Giorgia Franchini, and Federica Porta. Combining weighted total variation and deep image prior for natural and medical image restoration via adm. *arXiv preprint arXiv:2009.11380*, 2020.
- [3] Emmanuelle Chaigneau, Amanda J Wright, Simon P Poland, John M Girkin, and R Angus Silver. Impact of wavefront distortion and scattering on 2-photon microscopy in mammalian brain tissue. *Optics express*, 19(23):22755–22774, 2011.
- [4] Liang Chen, Faming Fang, Tingting Wang, and Guixu Zhang. Blind image deblurring with local maximum gradient prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1742–1750, 2019.
- [5] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. In *ACM SIGGRAPH Asia 2009 papers*, pages 1–8, 2009.
- [6] Adam Kaufman and Raanan Fattal. Deblurring using analysis-synthesis networks pair. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5811–5820, 2020.
- [7] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [8] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan 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.
- [9] Jan Kotera, Filip Šroubek, and Václav Šmídl. Improving neural blind deconvolution. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1954–1958. IEEE, 2021.
- [10] Dilip Krishnan and Rob Fergus. Fast image deconvolution using hyper-laplacian priors. *Advances in neural information processing systems*, 22:1033–1041, 2009.
- [11] Dilip Krishnan, Terence Tay, and Rob Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR 2011*, pages 233–240. IEEE, 2011.
- [12] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018.
- [13] Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. A comparative study for single image blind deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1709, 2016.
- [14] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1964–1971. IEEE, 2009.
- [15] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Efficient marginal likelihood optimization in blind deconvolution. In *CVPR 2011*, pages 2657–2664. IEEE, 2011.
- [16] Ji Li, Yuesong Nan, and Hui Ji. Un-supervised learning for blind image deconvolution via monte-carlo sampling. *Inverse Problems*, 38(3):035012, 2022.
- [17] Lerenhan Li, Jinshan Pan, Wei-Sheng Lai, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. Blind image deblurring via deep discriminative priors. *International journal of computer vision*, 127(8):1025–1043, 2019.
- [18] Jun Liu, Ming Yan, and Tiejiong Zeng. Surface-aware blind image deblurring. *IEEE transactions on pattern analysis and machine intelligence*, 43(3):1041–1055, 2021.
- [19] Gary Mataev, Peyman Milanfar, and Michael Elad. Deepred: Deep image prior powered by red. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.
- [20] Tomer Michaeli and Michal Irani. Blind deblurring using internal patch recurrence. In *European conference on computer vision*, pages 783–798. Springer, 2014.
- [21] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
- [22] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. l_0 -regularized intensity and gradient prior for deblurring text images and beyond. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):342–355, 2016.
- [23] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, 2016.
- [24] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.
- [25] Daniele Perrone and Paolo Favaro. Total variation blind deconvolution: The devil is in the details. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2909–2916, 2014.
- [26] Andrea Polesel, Giovanni Ramponi, and V John Mathews. Image enhancement via adaptive unsharp masking. *IEEE transactions on image processing*, 9(3):505–510, 2000.
- [27] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3341–3350, 2020.
- [28] Matthias Schneider, Sven Hirsch, Bruno Weber, Gábor Székely, and Bjoern H Menze. Joint 3-d vessel segmentation and centerline extraction using oblique hough forests

- with steerable filters. *Medical image analysis*, 19(1):220–249, 2015.
- [29] Th A Sorensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. *Biol. Skar.*, 5:1–34, 1948.
- [30] Libin Sun, Sunghyun Cho, Jue Wang, and James Hays. Edge-based blur kernel estimation using patch priors. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2013.
- [31] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8174–8182, 2018.
- [32] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.
- [33] Oliver Whyte, Josef Sivic, and Andrew Zisserman. Deblurring shaken and partially saturated images. *International journal of computer vision*, 110(2):185–201, 2014.
- [34] Norbert Wiener et al. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*, volume 8. MIT press Cambridge, MA, 1964.
- [35] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *European conference on computer vision*, pages 157–170. Springer, 2010.
- [36] Li Xu, Shicheng Zheng, and Jiaya 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.
- [37] Liuge Yang and Hui Ji. A variational em framework with adaptive edge selection for blind motion deblurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10167–10176, 2019.
- [38] Ada Zhen, Robert L Stevenson, et al. Gan based image deblurring using dark channel prior. *Electronic Imaging*, 2019(13):136–1, 2019.
- [39] Shuai Zheng, Zhenfeng Zhu, Jian Cheng, Yandong Guo, and Yao Zhao. Edge heuristic gan for non-uniform blind deblurring. *IEEE Signal Processing Letters*, 26(10):1546–1550, 2019.
- [40] Wangmeng Zuo, Dongwei Ren, David Zhang, Shuhang Gu, and Lei Zhang. Learning iteration-wise generalized shrinkage-thresholding operators for blind deconvolution. *IEEE Transactions on Image Processing*, 25(4):1751–1764, 2016.