Avoiding Undesirable Solutions of Deep Blind Image Deconvolution

Antonie Brožová1,2 and Václav Šmídl2
1Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic
2Institute of Information Theory and Automation, Czech Academy of Sciences, Czech Republic

Keywords: Blind Image Deconvolution, Deep Image Prior, No-Blur, Variational Bayes.

Abstract: Blind image deconvolution (BID) is a severely ill-posed optimization problem requiring additional information, typically in the form of regularization. Deep image prior (DIP) promises to model a naturally looking image due to a well-chosen structure of a neural network. The use of DIP in BID results in a significant performance improvement in terms of average PSNR. In this contribution, we offer qualitative analysis of selected DIP-based methods w.r.t. two types of undesired solutions: blurred image (no-blur) and a visually corrupted image (solution with artifacts). We perform a sensitivity study showing which aspects of the DIP-based algorithms help to avoid which undesired mode. We confirm that the no-blur can be avoided using either sharp image prior or tuning of the hyperparameters of the optimizer. The artifact solution is a harder problem since variations that suppress the artifacts often suppress good solutions as well. Switching to the structural similarity index measure from $L_2$ norm in loss was found to be the most successful approach to mitigate the artifacts.

1 INTRODUCTION

Recovery of a sharp, clean image from a degraded one is a difficult task regardless of the type of degradation. This paper is concerned with blur degradation, which may be caused by the relative motion of a camera and a scene, turbulence in the atmosphere, or the focus of a camera. Assuming a spatially invariant blur, a blurred image $d$ can be represented as a convolution (denoted by $\ast$) of a point spread function (PSF) $k$ and an underlying sharp image $x$

$$
d = k \ast x + n,
$$

where $n$ denotes a noise matrix. The deconvolution is basically an inverse operation to the convolution with the aim of recovering the sharp image from the blurred one. The deconvolution is called blind (BID) when not only the sharp image but also the PSF is unknown. The task is then to minimize

$$
\|d - k \ast x\|,
$$

with respect to both $x$ and $k$. To preserve the energy of the image, $k$ is required to contain only nonnegative values and such to 1.

Minimizing (2) is difficult since there can be many local minima other than the ground-truth solution. One notable solution is the trivial no-blur solution reconstructing the observation by the blurred image and Dirac delta PSF. Therefore, it is necessary to add some regularizer to (2) or prior information that helps recover the real sharp image $x$.

Sharp image priors were designed to yield a higher probability of sharp images over blurred ones to steer the optimization algorithm from the no-blur solutions, starting with (Miskin and MacKay, 2000), (Likas and Galatsanos, 2004) and (Molina et al., 2006). Variational Bayes (Tzikas et al., 2009), (Kotera et al., 2017) and Maximum Aposteriori (MAP) (Levin et al., 2011), (Perrone and Favaro, 2016) approaches were mainly discussed and various priors were proposed (Wipf and Zhang, 2014). Total variation (TV) minimizing the L1 norm of horizontal and vertical differences of the sharp image was proposed among first image priors (Chan and Wong, 1998). Later, the strength of super-gaussian priors (Babacan et al., 2009) was discovered. Similarly to TV, they assume that the gradient of the sharp image is sparse and most of its values is centered around zero. The priors do not necessarily need to prefer the sharp image but rather help to avoid the blurred one. Although these traditional methods are quite successful, their efficiency depends on a blur type, and inverse operations often leave the estimates of the sharp images degraded by artifacts (see the left image in Figure 1). A completely different approach to blind image deconvolution is based on deep learning (Huang et al., unknown.
2023), (Zhao et al., 2022), (Asim et al., 2020). Supervised deep learning models usually require training on large datasets, giving them more information than the traditional methods get and, therefore, outperforming them. However, there are real-world scenarios where large datasets are not available and for a long time, the Bayesian methods have been state-of-the-art for these problems. In 2018 (Ulyanov et al., 2018) proposed Deep Image Prior (DIP), stating that the structure of a deep neural network is a regularizer of the problem itself. It was shown that such a neural network learns naturally smooth images faster than noise, which, according to (Shi et al., 2022), is caused by faster learning of low-frequency information. They successfully used it for image denoising, inpainting, and super-resolution. (Ren and et al., 2020) combined the DIP image network with a feedforward neural network (FNN) representing the PSF in 2020 and proposed SelfDeblur. This model deblurs images without any training dataset and outperforms the Bayesian methods. They also propose to use TV regularization, but unless the images are noisy enough, it has no visible benefit. (Kotera et al., 2021) suggests that its success is not caused only by the DIP, but by some interplay between the structure of the network and the optimizer.

DualDeblur (Shin et al., 2021) utilizes DIP and multiple blurry images. (Wang et al., 2019) focus on the PSF and represent it with DIP as well as the sharp image. (Bredell et al., 2023) combined the DIP with Wiener deconvolution. The two conceptions of the ‘prior’ have been combined by (Huo et al., 2023), where the DIP-based model was complemented by the sharp image prior and BID solved as minimization of the variational lower bound.

However, all presented methods report only the average PSNR of the restored images without a detailed analysis of the effect of the components of their method. In this contribution, we analyze the effect of selected variations of DIP deblurring to shed some light on their role in the quality of the restoration.

We demonstrate experimentally that the no-blur solution can be avoided by choosing the optimization hyperparameters well or by the sharp image priors.

4. We illustrate that the suppression of the artifact solutions is a much more demanding task and is greatly influenced by the stochasticity of the DIP-based methods. The most significant improvement seems to be brought by switch to SSIM loss during optimization.

2 BACKGROUND AND MOTIVATION

Here, we will shortly review the analyzed methods, demonstrate their stochastic nature, and formulate the research objectives.

2.1 Deep BID Methods

We now introduce the studied algorithms SelfDeblur and VDIP, and a simplification of the SelfDeblur algorithm that reveals some properties of the MAP approach.

SelfDeblur. (Ren and et al., 2020) combines two generative neural networks: \( G_x \) representing \( x \) and \( G_k \) representing \( k \). The estimates of \( x \) and \( k \) are generated by inputting fixed random arrays \( z_k \) and \( z_x \) into the networks. \( G_x \) is a 5-level U-net (Ronneberger et al., 2015) with skip connections and bilinear upsampling. \( G_k \) is a FNN with one hidden layer. Softmax at the output of \( G_k \) preserves the \( L_1 \) norm of the PSF. The minimized loss is

\[
L_{SDB}(\theta_k, \theta_x) = \text{MSE}(d, G_k(\theta_k|z_k) \circledast G_x(\theta_x|z_x)),
\]

where \( \theta_x \), resp. \( \theta_k \) represents the trainable parameters of \( G_x \), resp. \( G_k \). The two networks are optimized jointly in 3000 epochs using Adam optimizer (Kingma and Ba, 2014) with learning rates (LR) \( \eta_k = 10^{-2} \) for \( G_k \) and \( \eta_k = 10^{-4} \) for \( G_x \). LRs are halved in 2000th, 3000th and 4000th iterations and \( z_x \) is perturbed by Gaussian noise with the standard deviation of 0.001 in every iteration. Although not in the original paper, the code provided by authors contains a switch to SSIM loss (5) after 1000 iterations. (Kotera et al., 2021) reported that this switch sometimes causes the optimization to deteriorate, so it will not be used as default in this method.

SimplerSDB. A slightly simpler model is also used in this paper. \( x \) is represented by \( G_x \) from SelfDeblur, but \( k \) is represented only by an array \( \theta_k \) normalized by softmax function (denoted as \( \sigma(\cdot) \)). The deconvolution is then formulated as the minimization of

\[
L_{SSDB}(\theta_k, \theta_x) = \text{MSE}(d, \sigma(\theta_k) \circledast G_k(\theta_k|z_k)).
\]

(4)
Figure 1: Illustration of stochastic influences. (a) PSNR histograms of nondeterministic computations on two GPUs. (b) PSNR histograms for three different initial values and nondeterministic computations on NVIDIA GeForce RTX 2080 Ti. The vertical lines show the PSNR of a result obtained by deterministic operations from corresponding initial values. The three reconstructed images are displayed above the graph.

The parameters are again optimized by Adam optimizer and $\eta_k = \eta_x = 10^{-2}$. Random perturbations of $z_x$ and learning rate scheduling used in SelfDeblur are turned off. This algorithm, stripped of everything else but DIP, will be called SimplerSDB.

Variational BID: VDIP. The last algorithm used in this paper for comparison is VDIP (Huo et al., 2023) utilizing two priors for the sharp image: sparse one (VDIP-Sparse) and extreme channel prior (Yan et al., 2017) (VDIP-Extreme). $G_i$ and $G_k$ are the same as in SelfDeblur. VDIP not only utilizes the DIP but also adds stronger prior information similar to bayesian methods. Apart from that, the considered loss is MSE-like (3) only in the first 2000 iterations; for the next 3000, it is switched to SSIM (structural similarity index measure) (Wang et al., 2004) loss, which is not mentioned in the paper, but the provided code contains it. The SSIM loss reads as

$$1 - \text{SSIM}(d, G_k(\theta_k|z_k) \ast G_x(\theta_x|z_x)).$$

Furthermore, $G_i$ is pretrained to reconstruct basic contours in the image and the same scheduling as in SelfDeblur is used.

### 2.2 Stochasticity of BID Algorithms

One issue connecting all three DIP-based algorithms is the stochasticity of their output. It is caused by two factors: i) by initialization of $G_i$ and $z_x$, and in some versions also by $G_k$ and $z_k$, and ii) by nondeterministic computations on GPU (convolution and bilinear upsampling). The influence of a GPU is demonstrated by 100 nondeterministic repeated runs of SimplerSDB on NVIDIA GeForce RTX 2080 and NVIDIA TITAN V; one blurred image from the Levin dataset (Levin et al., 2009) was used. Figure 1 (a) shows that there is a significant difference between the two GPUs. The influence of the random initial conditions is demonstrated by 100 runs of SimplerSDB with $G_x$ with nearest neighbor upsampling for three combinations of initial values of parameters $\theta_x$ and the input array $z_x$ on NVIDIA GeForce RTX 2080. Apart from these nondeterministic runs, a deterministic one was carried out for comparison (PyTorch offers a deterministic implementation of convolution); results are displayed in Figure 1 (b). Note that the deterministic computations may lead to a solution very different from the most likely stochastic solution. The three deterministically obtained deblurring results (one for each initial seed) are depicted on the top of Figure 1 (b).

This analysis reveals the sensitivity of SimplerSDB (and inherently all other algorithms based on DIP) not only to stochastic issues such as random initialization but also to computational hardware used in the experiment. This makes a comparison of different methods rather challenging since a naive comparison of novel results with previously published PSNRs may lead to misleading results.

### 2.3 Evaluation Metrics

The stochasticity of the DIP output is well known, and the majority of publications report average PSNR as a comparison metric. Here, we argue that compressing the whole histogram into a single number removes...
Figure 3: Three runs on the Levin dataset performed by SelfDeblur, SimplerSDB, VDIP-Sparse, and VDIP-Extreme. N-B denotes no-blur.

potentially useful information. This is demonstrated in Figure 2 using the histogram of PSNR (in our paper denoted as PSNR-GT) of the tested algorithms in their default setup on the Levin dataset. The mean values of the PSNR are presented in Table 1.

Since the mean value of the SelfDeblur is the highest, the conventional ranking procedure would select it as the best algorithm. While it offers many high-quality results, it also has many results with very low PSNR. On the other hand, VDIP-Sparse has almost no result with PSNR lower than 20 dB, making it a candidate with a low risk of poor solution. This advantage is compensated by the inability to provide excellent solutions with PSNR greater than 30 dB. An interesting area is around 20 dB PSNR, where VDIP-Extreme has fewer solutions then SelfDeblur. This motivates our search for the nature of these results and analyzing which variation of the method influences them.

2.4 Types of Undesired Solutions

The classical literature on BID extensively discusses two types of undesired solutions: i) the no-blur solution, and ii) an artifact solution. The no-blur solution is an estimate when the image is estimated as the blurred one, and the PSF as the Dirac delta function. The loss value of such a solution is zero when no regularization term is added to (2), so it is certainly a valid solution, yet undesirable. The artifact solution is named after visible artifacts corrupting visually the estimated images. Numerical instability of DIP, which was observed in the original paper, may also generate undesirable solutions. These may be easily detected from the value of the loss function or prevented by learning rate scheduling.

We will focus our attention only on the no-blur and artifact solutions. We introduce a modified metric to visualize this distinction. Specifically, we will compute PSNR not only to the ground truth image, \( x_{GT} \), but also to the blurred image, \( x_{NB} \), formally:

\[
PSNR-GT(x) := PSNR(x, x_{GT}),
\]

\[
PSNR-NB(x) := PSNR(x, x_{NB}).
\]

The no-blur solutions can be recognized by PSNR-NB value around 30 dB. Moreover, when comparing deblurring results on a dataset, it is also useful to measure the improved signal-to-noise ratio, which is defined as

\[
ISNR(x) = PSNR-GT(x) - PSNR-GT(x_{NB}).
\]

Plotting ISNR and PSNR-NB in 2D space, Figure 3 extends understanding of histogram from Figure 2. It shows that SelfDeblur and VDIP-Extreme are prone to solutions with artifacts that are in the area with low PSNR-NB and negative ISNR. On the other hand, solutions with ISNR around 0 dB and a high value of PSNR-NB show that SimplerSDB and SelfDeblur sometimes reach the no-blur solution. In the subsequent tests, we analyze which variations of the studied algorithms influence these solutions.

3 THE NO-BLUR SOLUTION

3.1 Model and Initialization of PSF

Good initialization is an important part of the MAP approach, such as SelfDeblur. Both the \( x \) and \( k \) are initialized randomly. In VDIP, an attempt is made to pre-train \( x \) towards the blurred image (at least contours) and \( k \) to a constant array. We now study the effect of the PSF initialization by various strategies on the SimplerSDB.

Initialization of \( k \) as a constant array was found to lead to the no-blur solution more likely than initializing it as a random noise, see Figure 4. On the other
3.2 The Setting of the Optimizer

In the case of SelfDeblur and SimplerSDB, the setting of the optimizer plays an important role. We observed that it is necessary to use the Adam optimizer to reach a reasonably low loss value. Firstly, LRs influence whether we find a good, sharp solution or the no-blur solution.

To see the influence, we deblurred 18 images (4 sharp images from the Levin dataset, 2 from the Kodak dataset (kod,) blurred by 3 PSFs from the Levin dataset) with fixed $\eta_k$. Since the best choice of $\eta_k$ depends on the blurred image, Figure 5 shows results for the images separately. It can be seen that higher values of $\eta_k$ lead the algorithm closer to the no-blur solution. In contrast, too low value of $\eta_k$ leads to solutions with artifacts if the algorithm converges. This behavior is the same for the other tested images. Therefore, carefully slowing down or speeding up the learning of $k$ may help to avoid the no-blur solution or solution with artifacts. We suggest that this is because the initialization of $k$ is very different from the delta function and a lower speed of learning of $k$ does not allow the algorithm to approach the no-blur solution at the beginning of the optimization. On the other hand, with a higher $\eta_k$, the algorithm descends quickly towards the no-blur solution, which the DIP should prefer.

Through hyperparameter search we discovered that increasing the value of $\beta_1^x$ (hyperparameter $\beta_1$ of Adam optimiser of $x$) helps SimplerSDB avoid the no-blur solution, and the effect is the same for SelfDeblur. The effect that $\beta_1^x$ has on the whole dataset is shown in Figure 6. We conjecture that this setting helps preserve the original gradient’s momentum from the initial optimization stages with constant $k$, thus avoiding the sharp local minima of the no-blur. In the case of VDIP, there is no obvious difference for the two values of $\beta_1^x$, but since VDIP does not generate any no-blur solutions, it is not that surprising.

3.3 Sharp Image Priors

Undeniably, the most effective way to avoid the no-blur is the sharp image prior (sparse and extreme channel) incorporated in the VDIP as can be seen from Figure 3. The variational Bayesian approach seems to be important for DIP models since the TV regularisation did not prove successful (Ren and et al., 2020; Kotera et al., 2021). Moreover, the VDIP algorithm is not sensitive to the choice of the optimizer hyperparameters such as $\beta_1^x$ studied in Section 3.2.

3.4 Discussion on No-Blur

Mitigation of the no-blur solution has been the objective of the traditional sharp image priors as well as various optimization tricks. It is not any different in the DIP approach. DIP itself does not prefer sharp images, but carefully setting the optimization hyperparameters helps very well to avoid the blurred one. The best option to avoid the no-blur is the combination of DIP with a sharp image prior in VDIP.
Figure 6: Sensitivity of the solution to the optimizer hyper-parameters $\beta_1$ in terms of PSNR-NB and ISNR for three runs on the Levin dataset.

Figure 7: Effect of pretraining on three runs on the Levin dataset.

4 THE ARTIFACT SOLUTION

Solutions with artifacts are generated mostly by Self-Deblur and VDIP-Extreme.

4.1 Sparse Prior

VDIP-Sparse manages to avoid solutions with artifacts, so it can be concluded that the sparse prior minimizing differences between neighboring pixels in the image estimate helps to avoid them. On the other hand, the prior hinders finding estimates with high PSNR-GT. The extreme channel prior does not seem to have this effect.

4.2 Pretraining

VDIP uses pretraining for initialization of both $x$ and $k$. $G_x$ is pretrained to return the blurred image, but only in 500 iterations, so it learns only the rough contours of the image. The target for $k$ is the constant array, but after the 500 iterations it still reminds more of the initial noise. While VDIP-Sparse behaves the same way when the pretraining is omitted, the performance of VDIP-Extreme improves, and it returns fewer solutions with artifacts when pretraining is not used, see Figure 7. The difference could be explained by the extreme channel prior not being as strong as the sparse one and getting lost on the trajectory between the no-blur and sharp solutions.

Since there is no prior in SelfDeblur and SimplerSDB, it could be expected that they will reach more no-blur solutions when pretrained this way. Even though SelfDeblur struggles to reach low loss value, all reconstructions are no-blur solutions. SimplerSDB, on the other hand, does not converge to no-blur in every run. Considering that the main difference here is the model of the PSF, the FNN may learn some information useful for the no-blur during pretraining, causing all reconstructions to be blurred images. The optimizers were not reset after pretraining in this experiment (following how it was done in VDIP), so the moments used in deblurring were those learned on a path toward the blurred image. Simply
loading a new optimizer for deblurring in SelfDeblur results in a scatterplot quite similar to the one without pretraining and, surprisingly, containing fewer no-blur solutions. This type of pretraining could work similarly to lowering the learning rate of the PSF. In this case, \(G_x\) starts deblurring closer to the solution, but PSF starts from the same initial point and thus decelerates the learning.

4.3 SSIM Loss

Another algorithmic variation used in VDIP (and the codebase of SelfDeblur) is switching to SSIM loss (5) after 2000 iterations. Figure 8 shows that VDIP-Extreme without pretraining requires the use of the combination of MSE and SSIM loss to avoid solutions with artifacts. When VDIP-Extreme is pretrained, there is no significant effect of the switch and it is the same for VDIP-Sparse (pretrained and not pretrained). In the case of SelfDeblur and SimplerSDB, histograms of PSNR-GT move towards higher values, but some solutions with artifacts remain. Surprisingly, the no-blur solutions are eliminated in the case of SimplerSDB. It is difficult to compare these results because each of these algorithms may be at a different stage of optimization at the time the loss function is switched. Overall, the effect of the SSIM loss is significant and positive, mostly for the variants without a sharp image prior.

4.4 Discussion on Artifacts

Artifacts are not visually plausible, making the resulting solution undesirable. DIP was proven to prefer smoother images, so it could be expected to avoid solutions with artifacts and more likely achieve the no-blur solution. Figure 3 shows that DIP-based models still reach them, and almost none of the tested variants avoided them. Even though pretraining could be expected to push the algorithms towards smoother solutions, it did not prove to be true. Moreover, pretraining VDIP-Extreme actually hurts its performance. Eventually, the combination of no pretraining and loss with SSIM helped to get rid of the solutions with artifacts. SSIM loss has a positive effect on all the tested algorithms, nevertheless, we cannot conclude that it helps with the solutions with artifacts every time. The only reliable method is the sparse prior in VDIP-Sparse, which eliminates solutions with artifacts at the cost of losing excellent solutions with a lot of details because these images can contain similar intensity changes as those with artifacts.

5 CONCLUSION

The traditional undesirable solutions of blind image deconvolution, i.e. the no-blur solution and the solutions with artifacts, are also present in DIP-based methods. Similarly to the classical methods, the sharp image prior can effectively avoid the no-blur solution. In the case of a variation of a DIP method without a sparse image prior, optimization tricks in the MAP approach can similarly suppress this solution. The solution with artifacts, traditionally attributed to inversions of poorly conditioned matrices, probably is not caused only by this numerical inaccuracy and remains a difficult task for DIP-based method as well. Even though some variations of the method are more prone to this undesirable solution, such as the switch to the SSIM metric, a reliable solution still remains to be found.
ACKNOWLEDGEMENTS

This work was supported by the grants GA20-27939S and GA24-10400S.

REFERENCES


