

Improved Deep Multi-Patch Hierarchical Network For Handling Saturation In Image Deblurring

Mahendra B.M. ^{a,*}, Savita Sonoli ^b, Abhishek Ameta ^a

^a Department of Electronics and Communication Engineering, RV College of Engineering, Bengaluru, Karnataka, India

^b Department of Electronics and Communication Engineering, R Y M Engineering College, Ballari, Karnataka, India

ARTICLE INFO

Keywords:

Artifact
Blur
CNN
Convolution
Deblurring
GoPro
Hierarchical network
Linear model
Multi-patch
Network-in-Network
PSNR
Ringing
Saturation
SSIM

ABSTRACT

The most active study topic in the field of computer vision is dealing with saturation in image deblurring. The Deep Multi Patch Framework (DMPHN) is primarily utilized in dynamic scene deblurring, and its fine-to-coarse hierarchical representation results in lower processing costs. The contribution of the framework in gathering and processing local residual information of blur at the coarser levels in the hierarchy makes the dynamic scene deblurring successful. The existing framework uses the traditional CNN with the convolutional filter, which is a generalized linear model (GLM) that goes along with a nonlinear activation function to scan the input. When the samples of latent concepts are not linearly separable then the linear CNN cannot abstract the good representations. Because saturation in an image contradicts the linear properties of the blur model, the present Deep Multi Patch architecture with linear CNN is unable to collect the necessary local information of blur in an image. In CNN, the higher-level layer maps to the larger regions of the input image by combining the lower-level concepts from each local patch. Due to the framework's incapacity to retain information at lower levels, noticeable ringing artifacts and poor deblurring performance resulted. We designed a nonlinear function approximator for this section, which was influenced by the good abstraction capabilities of "Network-in-Network". The nonlinear features of saturated pixels and blur are successfully captured by this upgraded DMPHN with a nonlinear function approximator. Compared to the current DMPHN model, the suggested approach greatly reduced the number of trainable parameters while improving the time cost. On the GoPro, Synthetic, and Natural Saturated Blurred Image datasets, the upgraded DMPHN network achieves superior deblurring performance with better PSNR and SSIM than the state-of-the-art.

1. Introduction

Motion blurring is a process of degradation of the image due to different conditions for example camera shake, relative motion between the camera and object, depth variations (Few to mention) during image capturing. The process of generating a sharp image by removing blur from a degraded image is called Image Deblurring. Image deblurring is the most basic step yet very important in computer vision applications. The blurring process is mathematically modeled as

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

Where y is a degraded or blurred image, x is a latent or sharp image, k is blur kernel, n is noise added and $*$ is the convolutional operation. From a couple of decades, several types of research have been made on image deblurring and results are mostly successful but still, some of the problems have to be looked into deeper insight. Out of which the saturated pixels and their effect in image deblurring are a key

problem to be addressed [1–11]. When the image is captured in low light conditions with extended exposure time, a bright spot in the image will be saturated by clipping pixel intensity to its maximum value [12]. These saturated pixels violate the linear blur model assumed by many existing algorithms and such pixels cause noticeable ringing artifacts in deblurred images as shown in Fig. 1. The blurred images with saturated pixels are modeled as [13],

$$y = c(x * k) + n \quad (2)$$

Where y is a degraded or blurred image, c is the clipping function, x is a latent or sharp image, k is blur kernel, n is noise added and $*$ is the convolutional operation.

Handling Saturated pixels in image deblurring have acquired great interest in computer vision applications. Many researchers have made major contributions in handling such outliers exclusively [2,13,15–27]. In the field of Image deblurring, many approaches have been developed

* Correspondence to: Department of ECE, RV College of Engineering, RV Vidya Nikethana, Mysore Road, Bengaluru 560059, Karnataka, India.

E-mail addresses: [mahan.8787@gmail.com](mailto:mahen.8787@gmail.com) (Mahendra B.M.), [savita.chitriki@gmail.com](mailto:savitachitriki@gmail.com) (S. Sonoli), abhishek.ameta222@gmail.com (A. Ameta).



Fig. 1. Output of [14]. Ringing artifacts are visible near saturated region highlighted in red and green box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in different directions by considering various problems of interest. The Non-Blind and Blind Deblurring are the two main approaches, where the known blur kernel in the Non-Blind Deblurring process and the unknown blur kernel in the Blind Deblurring approach. Most of the images captured are naturally blurred and there will be no information about the blur kernel. Hence most of the approaches come under the Blind Deblurring category. These two categories are further considered the key problems (outliers) such as uniform and non-uniform motion blur, Saturated pixels, Non-Gaussian noise, and are few to mention. The conventional methods for dynamic scene blind image deblurring use various regularizations or constraints to estimate the motion blur filters, which involve non-convex nonlinear optimization. Moreover, the blur patterns which are assumed spatially uniform are restrictive and lead to poor deblurring results for complex blur patterns. With the advantage of its learning capabilities, neural network (CNN) approaches overcome the difficulties from conventional blind deconvolution methods to some extent. There was no exclusion with CNN-based methods such that finding blur kernel was crucial before recovering latent image. Unless designing the blur model carefully, blur kernel estimation is so delicate and sensitive in presence of saturation and noise, which leads to the introduction of artifacts in resulting images. There is a possibility of finding latent images directly without restrictions on the blur kernel model (end to end) in multi-scale CNN. The scale recurrent [4] and multi-scale network architecture [18] with deep CNN under a coarse-to-fine scheme have been introduced to make use of deblurring clues at different processing levels. Parameters are shared across the scales in Scale recurrent structure, which results in a struggle to recover deblurring quality and a costly runtime due to a huge number of training parameters since the large-sized filters used by the structure. Multi-scale strategies suffer from expensive computation due to their upsampling operations. Deep Multi-Patch Hierarchical Network (DMPHN) is an end-to-end approach [11] that addressed the aforementioned problem through a fine-to-course hierarchical representation. A Multi-patch network reduces the computational cost by avoiding upsampling operations, instead, the input blurry image is divided into patches at multiple levels. Along with that, a multi-patch model with a stacked version is introduced to handle the performance saturation concerning depth. With the advantage of computational cost, Fig. 3 illustrates the higher-level layer in DMPHN maps to the larger portions of the input image by combining the lower-level concepts from each local patch. First, errors propagate from lower level to higher level structures in the architecture as the deeper layers with 3×3 sized filters



(a) Blurred Input. **(b) Deblurred Output.**

Fig. 2. Deblurred output of DMPHN [11] with protruded ringing artifacts.

learn abstract features (features contaminated by outliers). Second, by suppressing the fine image structures, the 3×3 filters in the deeper layers lose high-frequency components. Because of these mistakes, it is hard for the network to figure out which features are the best, which leads to ringing effects and bad deblurring Fig. 2. To address this issue, we present Improved DMPHN, whose performance is higher compared to current state-of-the-art deblurring methods. Our contributions are summarized below:

- Instead of using an expensive 3×3 -size filter, we used a 1×1 -size filter at the beginning and end of the encoder and decoder of a multipatch network respectively.
- Other than the GoPro dataset, we trained the proposed model with synthetic and real-world images with saturated pixels provided by [28], which makes the model stronger in learning required features.

We avoid learning abstract features and blurring fine image structures by using 1×1 sized filters instead of 3×3 sized filters in the proposed network model. It also reduces the number of model parameters and increases the network's nonlinearity. The model is more general and better at deblurring after being trained with synthetic and real-world images with various types and levels of blur and saturation.

The experimental results show the clear benefits of the proposed Improved DMPHN (Name IDMPHN) model in handling saturated pixels effectively in motion deblurring with a significant reduction in ringing artifacts and support deblurring of 720-pixels images in real-time (at 30 frames per second). The paper contributes a brief framework of the associated effort of inception modules and Deep Multi-Patch Hierarchical Network in Section 3. The implementation method containing the reformed basic model and the overall architectural structure of the network is described in Section 4. Estimation of our experimental results, aiming at investigating qualitative and quantitative estimations are given in Section 5., and our study is concluded in Section 6.

2. Literature review

In recent years, the conventional deblurring methods have concentrated on three main features, i.e., probabilistic noise model [1,15,16,29], strong priors of natural images [2,3,30,31], and estimation of blur kernel [17,32–34]. Recently, learning-based approaches (CNNs) are evolved tremendously in the image deblurring process due to their self-learning capabilities, computational speed, and robustness. Some stages in the conventional methods are replaced with learning-based methods [35–37], and some techniques completely avoid blur kernel estimation by learning an end-to-end mapping [4,18,38].

Image deblurring falls into two groups based on the knowledge of the blur kernel. The first group is non-blind deblurring [13,15,19,27], where the blur kernel is known and the solution is a purely inverse problem. The second group is blind deblurring [1–3,5–7,10,14–17,21,26,29–33,38], where there is no clue of blur type and this method is more accurate but ill-posed. Statistical and learning methods are

the two approaches in Blind deblurring. Bayesian probability models and handcrafted image priors are utilized for framing statistical methods. Some of the statistical-based approaches have been implemented using CNN's [22,35,37,39]. Many recent efforts have been made on motion deblurring using an end-to-end approach and have developed a deblurring framework by using multiscale and scale-recurrent networks [4,18]. [40] Proposed a novel multi-scale channel attention network (MSCAN) for successful blind deblurring of dynamic single image situations. [41] developed a technique for image deblurring that uses RAW photos directly and looked into novel neural network structures that aid in RAW-based learning. Built a new dataset with both RAW and processed sRGB photographs, and created a new model to take use of RAW images' special properties. [42] proposed an exemplar-based, end-to-end trainable deblurring technique. More specifically, the deblurring is assisted by the utilization of the exemplar's most pertinent features. [43] The Defocus Image Deblurring Auxiliary Learning Net (DID-ANet), a novel network architecture created for single image defocus deblurring. A new method termed spatial pyramid pooling channel attention (SPPCA) is proposed for more powerful network representation. Most of these multi-scale architectures have the shortcoming of more processing time due to the huge number of parameters. Nested skip connections and parameter-selective sharing techniques are introduced in [44] to improve the efficiency of large-scale recurrent networks. To get around these problems with multi-scale and scale recurrent networks, the Deep multi-patch hierarchical network [11] was constructed.

Few statistical-based approaches have been developed for handling saturation in image deblurring. [15] extended the Bayesian Framework and suggested a statistical model analyze the effect of outliers. [16] obtained improved deblurring results by replacing the piecewise function in [15] with a continuous probability function. [17] discussed the effect of outliers on the performance of the system in kernel approximation and suggested a confidence function to eliminate the effect of outliers while estimating the blur kernel. [20] substituted the confidence function introduced in [17] with the data fidelity term in the loss function. [39] estimated the blur kernel by incorporating inliers identified successively. All these [17,20,39] methods are resulting in ringing artifacts in the deblurred images. [19] approximated the blur kernel by considering light streaks and this method fails when there are no light streaks in the image. [45] developed a novel blind motion deblurring technique for light streaks in blurry photos. The main concept is to use the shape of the light streak as a cue to estimate the blur kernel and describe the non-linear blur generated by outliers like the Huber's M-estimation in blind deconvolution. [23] proposed an edge-aware SR Network for motion deblurring and dataset synthesis approach to handle outliers. [24] considered saturated and unsaturated pixels together during the deblurring process to preserve most of the edge information in the image and used an effective MAP-based optimization framework but the MAP optimization does not work well in presence of large outliers. [26] addressed the problem with [24] by introducing a method that identifies and discards the outliers directly during updation of both kernel estimation and latent image. Discarding outliers in [26] may lose the important information of edges and leads to ringing effects in the final latent image. [40] proposed a unique local intensity-based prior, the patch-wise minimal pixels (PMP) prior, is proposed. It is quite adept at distinguishing between sharp and blurry images. Hence there is a trade-off between outlier detection, rejection, and carrying it through the deblurring process is a challenging task and there is room for improving the performance of existing models. The new model is aimed in this direction, and the following are the main distinctions between the proposed technique and other recent approaches in development for similar reasons.

- Under the coarse-to-fine scheme, most networks use a lot of training parameters because their filters are big. So, the multi-scale [18], scale-recurrent [44], and existing DMPHN [11] methods have a long run time and make it hard to improve the quality of deblurring.

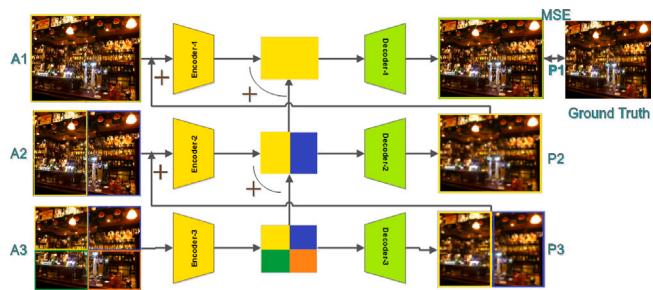


Fig. 3. The suggested system makes advantage of the encoder and decoder architecture's generic structure [11].

- Increasing the network depth for very low-resolution input in multi-scale approaches and the DMPHN network by sub dividing the blurred image into increasingly smaller grids has no effect on the deblurring performance.

In this study, we address the above challenges with the multi-scale, scale-recurrent, and DMPHN architectures.

- Unlike multi-scale and scale-recurrent network models, which need expensive inference, our technique uses a residual-like design, which requires relatively small-size filters, resulting in fewer network parameters.
- Instead of a multi-scale architecture, we use a multi-patch setup so that we can use feature map concatenations. This lets the lower level focus on local information and use the leftover information to make information for the coarser grid, which improved the performance.
- In contrast to the DHMPHN architecture, we utilized a 1×1 -size filter at the beginning and end of the encoder and decoder of a multipatch network rather of a costly 3×3 -size filter. It prevents the acquisition of abstract attributes and the blurring of fine visual structures. It minimizes the number of model parameters and increases the nonlinearity of the network.

3. Proposed method

The proposed system used the general framework of the encoder and decoder architecture shown in Fig. 3. The modifications made in this work to the encoder-decoder architecture is a Nonlinear function approximator i.e. 1×1 convolution filter [46,47] is included in encoder and decoder to capture the nonlinear properties of saturated pixels and blur. Outline of the modifications made to the basic encoder-decoder architecture is explained in this section alongside the overall architecture of the proposed network. In this work, the modified network design is primarily founded on a method called Deep Multi-Patch Hierarchical Network [11].

3.1. Micro neural network

It is evident that in CNN, the higher-level layer maps to the larger regions of the input image by combining the lower-level concepts from each local patch. Hence it is required to make good abstraction at lower levels to make higher levels work better which is possible by Micro neural networks and explained in the following section.

Network-in-Network (NIN) is a method suggested by [46] shown in Fig. 4 to improve the representational capabilities of neural networks. The NIN act as a micro neural network that enhances the discrimination ability of inlier and outlier (more complex structures of saturated pixels) properties for local patches within the receptive field. In traditional CNN, the convolutional filter is a generalized linear model (GLM) that goes along with a nonlinear activation function to scan the input.

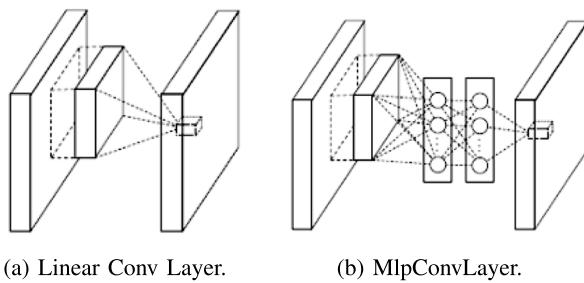


Fig. 4. The NIN (Network-in-Network) approach proposed by [46]

When the samples of latent concepts are not linearly separable then the linear CNN cannot abstract the good representations. The Conventional CNN is designed in such a way that the model can abstract the good representations by utilizing a set of over-complete filters to capture all the variations of the latent concepts. But the utilization of numerous filters for this purpose may put additional load on subsequent layers. In CNN, the higher-level layer map to the larger regions of the input image by combining the lower-level concepts from each local patch. Hence it is required to make good abstraction at lower levels to make higher levels work better. With these advantages of NIN as a micro neural network is combined into existing Convolution Neural Network architecture in hunt of better abstractions for all the levels of features. Inspired by “Network-in-Network” for classification tasks, the existing DMPHN modeled with conventional CNN layers is modified by introducing micro neural networks to capture the nonlinear properties of saturated pixels and motion blur to achieve improved image deblurring results with less computational complexity.

3.2. Enhanced encoder–decoder

The enhanced encoder-decoder architecture is introduced in this section. The differences between enhanced encoder-decoder architecture and architecture proposed in [11] are shown.

The contribution of the DMPHN framework [11] in gathering and processing local residual information of blur at the coarser levels in the hierarchy makes the dynamic scene deblurring successful. The problem with the framework is, First, errors propagate from lower level to higher level structures in the architecture as the deeper layers with 3×3 sized filters learn abstract features (features contaminated by outliers). Second, by suppressing the fine image structures, the 3×3 filters in the deeper layers lose high-frequency components. Third, it is unable to capture the appropriate local information of blur around saturation regions since the blur model is non-linear around the saturation regions represented in Eq. (2). This loss of information due to the inability of framework introduces prominent ringing artifacts and limits deblurring performance shown in Fig. 2.

We solved this problem by introducing a Micro Neural Network explained in Section 3 to the building blocks of the encoder-decoder. In contrary to existing DMPHN [11], we replace the convolutional filter with the kernel size of 3×3 with a 1×1 kernel performing as a more general function approximator at the output and input stages of each encoder and decoder respectively. Convolutional layers have historically been used as feature detectors in classification tasks. However, stacking multiple convolutional layers on top of each other endows the network with the ability to abstract details in deeper layers [48]. Although this property is useful for classification and other similar tasks, it is not suitable for image deblurring because finer details of an image must be preserved for a good reconstruction. To extract all of the necessary features from the image, simply having large number filters of the same size is not enough. Inspired from inception [46] layers of GoogLeNet [47], we employ two sets of convolutional filters(3×3 and 1×1). To prevent learning abstract features in the later layers

of our model by limiting expressivity, we restrict the filter sizes of convolution layers to 1×1 [49]. The 1×1 convolution increases nonlinearity to improve accuracy. It also cuts down on the number of network parameters. [47]. For small image patches, the features that would be extracted by 1×1 filter will be highly local and may not have a general overview of the image. This helps capture smaller, more complex features in the image. For large image patches, the features that would be extracted by 3×3 would be generic, and spread across the image. This helps capture the very basic components but not the specific components in the image. Since each level of our network focuses on different levels of blur or different sized image patches, a combination of 3×3 and 1×1 convolutions gives the best results. The success revealed in the above lines represents the effectiveness of our IDMPHN.

3.2.1. Encoder-decoder structure

The architecture of our IDMPHN network is shown in Fig. 5. Every stage of the architecture comprises one encoder with fifteen convolutional layers, six residual links, and six Rectified Linear Units (ReLU). Every encoder will follow by one decoder whose architecture is similar to the layers of the encoder except deconvolution layers are introduced to generate deblurred images in place two convolution layers. The architectural difference between the proposed DMPHN design in [11] and our IDMPHN architecture is shown in red.

4. Network architecture

This section highlights that our network design is constructed on the model proposed by [11]. The complete structure of our IDMPHN design is shown in Fig. 6. Following the similar approach, the blurry input image is equally divided into Four, Two, and One patch(es) from level-Three to level-One, correspondingly, as shown in Fig. 6. The intermediate feature map of each patch at each level is achieved from the corresponding encoder. At each level, the features extracted are different with more depth at lower levels, the information at all the levels is gated to the highest level. Hence, the top level contains all the information received from the lower levels, the residual links help in providing the information at the upper levels. It is important to note that each feature map of every patch is of the same size for convolutional operations from the lower level to the higher level of the architecture. For element-wise addition, feature maps from lower levels are adjusted according to the size of upper-level feature maps. Let A_1 be the initial blurry input image and the j th patch at the i th level is A_{ij} . Furthermore, E_i is an Encoder and D_i is Decoder at level i , B_{ij} is the output features of encoder E_i for input patch A_{ij} , and P_{ij} represents the output patches from the output of decoder D_i .

The process of deblurring begins at level 3. Input blurred image A_1 is divided into 4 patches i.e. $A_{3,j}$, $j = 1$ to 4. These non-overlapped image patches are given to encoder E_3 . The output of the encoder is a convolutional feature represented as

$$B_{3,i} = E_3(A_{3,i}), j \in \{1 \text{ to } 4\} \quad (3)$$

The adjacent features for each patch are concatenated in a spatial sense to obtain new features, represented as $B_{3,j}^*$, which is of the same size as the convolutional features represented at level 2:

$$B_{3,i}^* = B_{3,2i-1} \oplus B_{3,2i}, j \in \{1 \text{ to } 2\}, \quad (4)$$

where the symbol \oplus represents the concatenation operator. The decoder D_3 gives out an output image patch $P_{3,j}$ for the respective input $B_{3,j}$,

$$P_{3,i} = D_3(B_{3,i}^*), j \in \{1 \text{ to } 2\}, \quad (5)$$

Next, the process will shift to the succeeding level i.e. level 2. The input to encoder E_2 is the sum of $P_{3,j}$ and $A_{2,j}$, and the output of E_2 i.e. $B_{2,j}$ is added with $B_{2,j}^*$. The output of encoder E_2 is,

$$B_{2,i} = E_2(A_{2,i} + P_{3,i}) + B_{3,i}^* j\epsilon(1 \text{ to } 2), \quad (6)$$

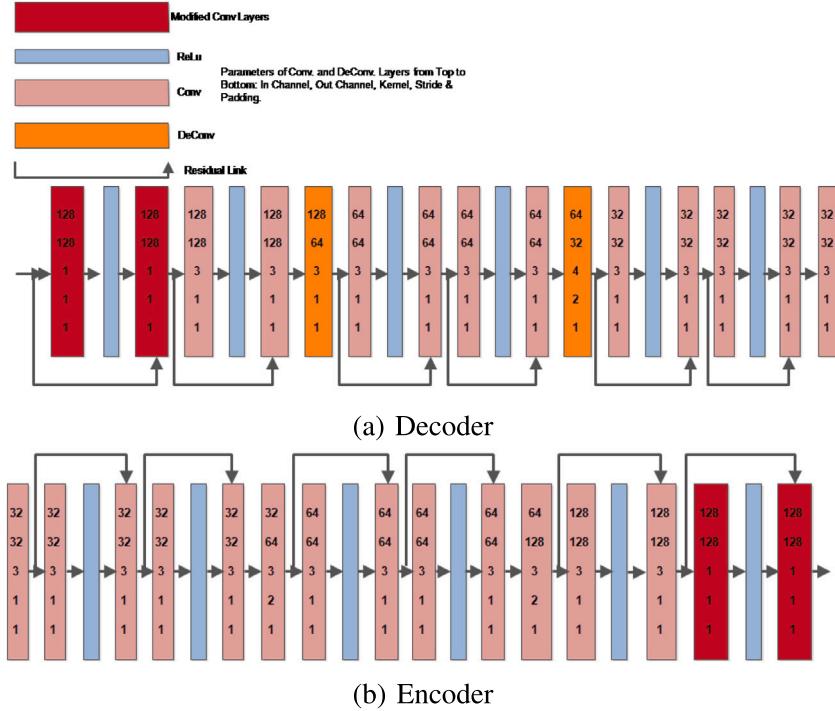


Fig. 5. Layer configurations for proposed (a) decoder and (b) encoder architectures. The difference in architecture between [11] and ours is noted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$B_2^* = B_{2,1} \oplus B_{2,2}$$

The output image of decoder D2 is given by:

$$P_2 = D_2(B_2^*)$$

The final deblurred output P1 at level 1 is given by

$$B_1 = E_1(A_1 + P_2) + B_2^*$$

$$P_1 = D_1(B_1)$$
(10)

The Mean Square Error (MSE) loss is tabulated at level 1 output. The loss function of IDMPHN is given as

$$L = \frac{1}{n} \sum_{i=1}^n (P_i - G_i)^2 \quad (11)$$

Where G is the ground truth sharp image, P is the output latent image at level 1, n is the total number of pixels in an image, G_i and P_i are i th pixel in G and P respectively.

The multi-patch hierarchical network [11] follows the residual learning principles: the image statics captured at intermediate outputs at different levels outputs P_i . Therefore, the loss function is calculated only at level 1.

5. Experiment and evaluation

This section presents the details of implementation and performance evaluation of the proposed method compared to the existing methods qualitatively and quantitatively.

5.1. Implementation

The code is written in Python and tested on a Personal Computer with an NVIDIA GeForce 920MX graphics card using the PyTorch library. The training dataset is created by cropping photos into 256X256 pixels at random. Following that, the obtained patches from

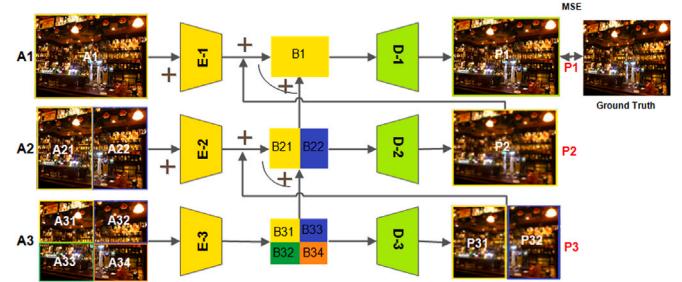


Fig. 6. Structure of the proposed network architecture.

the cropped photos are fed into each level's inputs. The batch size has been set at six. The weights are optimized using Adam optimizer for 2 epochs. The initial learning rate is set to 0.0001 and the decay rate to 0.1. The image is normalized to the range [0, 1], and then 0.5 is subtracted. Due to CPU memory constraints, the deepest Stacking model trained and evaluated is (1-2-4).

5.2. Experiments:

This section assesses the proposed system's performance on both synthetic and real-world images and compares it to current state-of-the-art systems. We compare our method to state-of-the-art algorithms by running large numerical tests on two benchmark datasets with saturated pixels [18,28]. Then, to demonstrate the usefulness of our approach, we use some hard real-world instances with large saturated regions.

5.2.1. GoPro dataset from [18]

The GoPro dataset consists of 3214 pairs of blurred and clean images captured at 720×1280 resolution extracted from 33 sequences. The blurred images are developed by producing varied blur through the averaging of varying numbers (7–13) of successive latent frames. We

Table 1

The results of our model compared with the existing DMPHN model. Both the models are trained with the GoPro dataset [18]. Our proposed model appears to deblur better than DMPHN [11].

Architecture	Metric	Disc	Disc-sat	Motion	Motion-sat	Uniform	Uniform-sat	Non-uniform
DMPHN [11]	PSNR	20.892	17.9175	22.6417	21.9675	16.1484	14.8310	14.3106
	SSIM	0.8415	0.7900	0.8797	0.8949	0.6243	0.6215	0.4026
Ours	PSNR	21.4967	20.1730	22.4360	21.6712	17.8645	16.3862	14.5832
	SSIM	0.8545	0.8519	0.8776	0.8824	0.6902	0.6830	0.4877

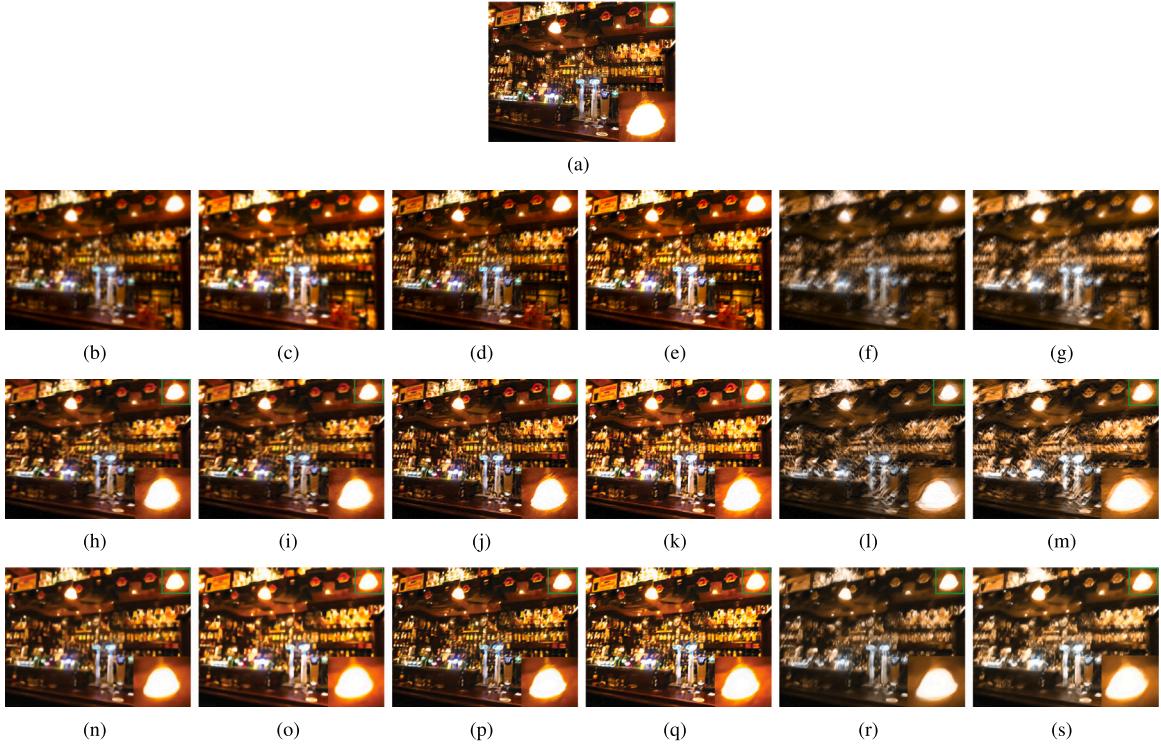


Fig. 7. [a]: Ground Truth, [b-g]: Input with varied blur types, [h-m]: DMPHN output and [n-s]: Ours output.

trained our model with 100 pairs of randomly chosen images out of 3214 pairs, due to hardware limitations. Our method tested on synthetically blurred images of varied blur types and compared with the existing DMPHN [11] model trained with 2103 image pairs, where our model showed better results in handling different blur types with and without additional saturation both quantitative and qualitatively since the model can learn and capture the blur along with saturation regions at initial stages, shown in Fig. 7 and Table 1. Our IDMPHN model outperforms the existing DMPHN model for all blur types, but somewhat worse for motion blur, because the DMPHN model is exclusively trained and tailored to deal with motion blur using a bigger training set [18]. Even with a low training set, our model handles motion blur comparably well to DMPHN, and training with a larger dataset would improve deblurring quality, as shown in Section 5.2.2.

5.2.2. Saturated dataset from [28]

The dataset provided by [28] contains 30 sharp latent saturated images, 4 synthetic uniform blur kernels. Three different blur types are used to generate 3 different datasets with the help of these 30 sharp latent images and are as follows,

Disc blur + saturation + noise. Disc kernel with radius 7 shown in Fig. 10(a) is applied on sharp images first, then each pixel intensities are multiplied by scale factor 1.3, finally, Gaussian noise with zero mean and Variance of 0.05 is added.

Motion blur + saturation + noise. Motion kernel with length=10 and angle=45° shown in Fig. 10(b) is applied on sharp images first, then each pixel intensities are multiplied by scale factor 1.3 to increase the saturation level, finally, Gaussian noise with zero mean and Variance of 0.05 is added.

Synthetic uniform blur+ saturation + noise.: The sharp image is blurred by convolving with one particular uniform blur kernel out of 4 uniform blur kernels provided in dataset [28] shown in Fig. 10(c), then each pixel intensities are multiplied by scale factor 1.3, finally, Gaussian noise with zero mean and Variance of 0.05 is added.

Dataset with each blur type mentioned above consists of 30 pairs of blurred and sharp images and altogether 90 blurred and 30 sharp images. Our model is trained and tested with the individual type of blur dataset mentioned in Section 5.2.2 and results are compared with state-of-the-art methods. Figs. 11 and 12 illustrate the qualitative results of two separate disc blurred images, Image-1 and Image-2, respectively. Table 3 exhibit the quantitative results. It is observed that the proposed model is working better for two different challenging blurred images with highly saturated regions. For further verification of the deblurring capacity of our model, the testing is done on two more blur types which are Motion-Sat and Uniform-Sat blur. Figs. 8 and 9 exhibit the qualitative effects of motion and uniform blurred pictures, respectively. The quantitative comparison in Table 4 demonstrates that the model performs better in terms of motion and uniform blur.

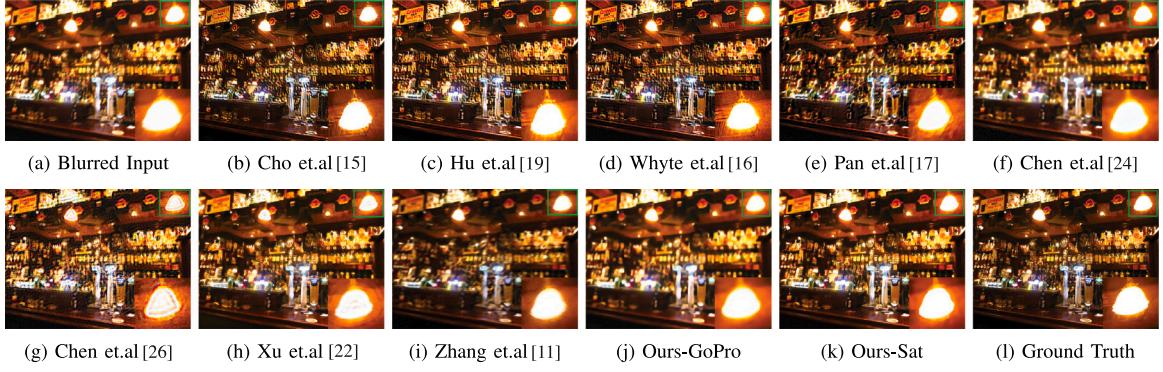


Fig. 8. Deblurred results of motion blurred image.

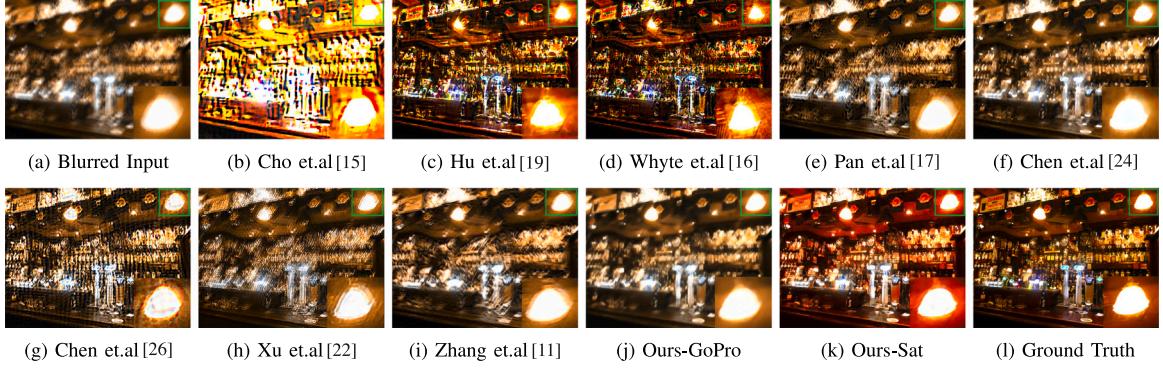


Fig. 9. Deblurred results of uniform blurred image.

5.3. Generalization of the proposed model:

To verify the generalization of our model in deblurring the varied levels of saturated images, we trained our model with 87 blurred and 27 clean images mentioned in Section 5.2.2 and tested with challenging examples with abundant saturated pixels chosen from the literature, consolidated by [28] which includes Non-Uniform and Real blurred images. Table 5 shows the quantitative results of our model for non-uniform blur inputs. Fig. 13 represents deblurred images of Non-Uniform blurred input. Fig. 14 represents the deblurred images of naturally blurred input images. The state-of-the-art approaches [2, 11, 15–20, 22, 24, 26] are less effective due to the side effects caused by saturated pixels, and their deblurred results contain ringing artifacts, and some features are not recovered properly, as it is visible in Fig. 13 & 14 .

5.4. Processing speed & number of trainable parameters:

Compared to the previous DMPHN model [11], the suggested model reduced the time cost by around 1%. For 100 epochs with a batch size of six, the time cost is estimated for both models using GoPro and Lai's dataset. Thus, the present DMPHN model has already been demonstrated to be several times quicker than the other techniques. Aside from DMPHN, our approach performs far better than any other method. To compensate for the lower time cost, our model features 62.7% less trainable parameters than the previous DMPHN model [11].

Table 2

Table 2
Number of trainable parameters.

Model	Total number of trainable parameters
DMPHN	5424393
Proposed network	2021529

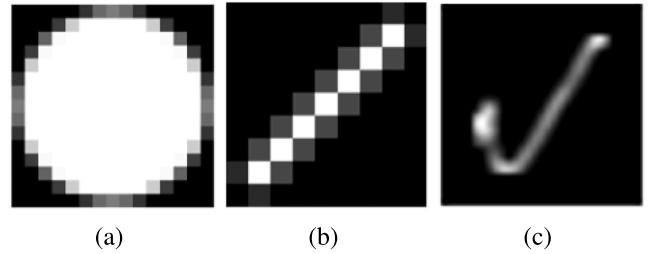


Fig. 10. Synthetic Blur Kernels. 10(a). Disc blur with radius is 7, 10(b). Motion blur with $L=10$ & $\theta=45$, 10(c). Uniform Kernel from [28]. The kernel images are magnified and cropped for display purposes.

(1×1 convolution layer) to encoder-decoders building blocks of the existing DMPHN framework. Unlike the existing DMPHN framework, the improved DMPHN framework captures the appropriate nonlinear properties of saturated pixels along with the blur at the coarser levels. The enhanced feature discrimination capability of these Micro networks within the receptive fields of local patches avoids the propagation of the error due to saturated pixels through the layers which reduce prominent ringing artifacts in the deblurred images. When compared to state-of-the-art results, the upgraded DMPHN network provides faster processing speed with less network parameters and

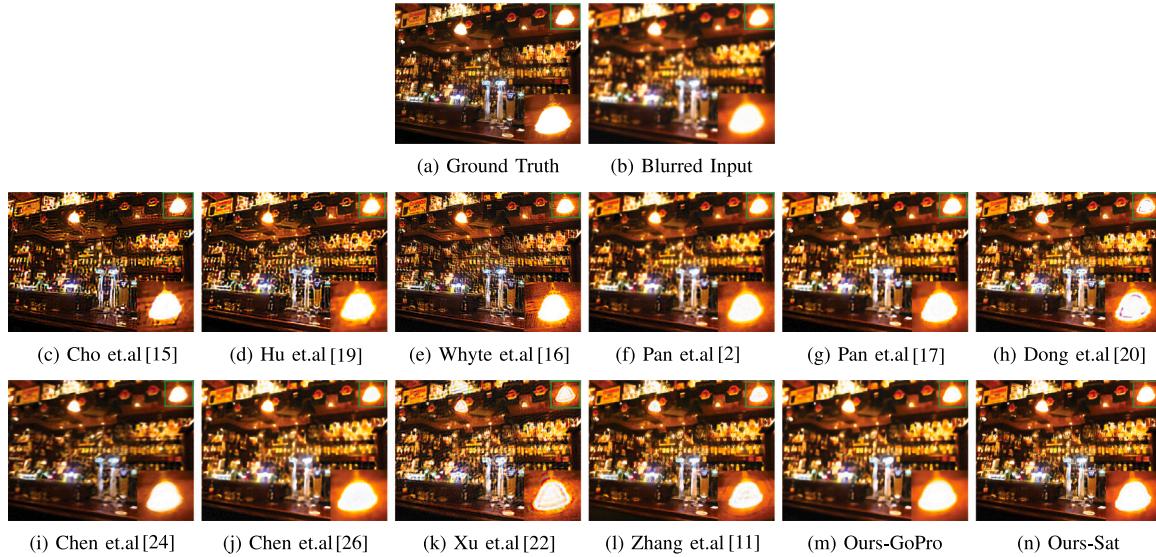
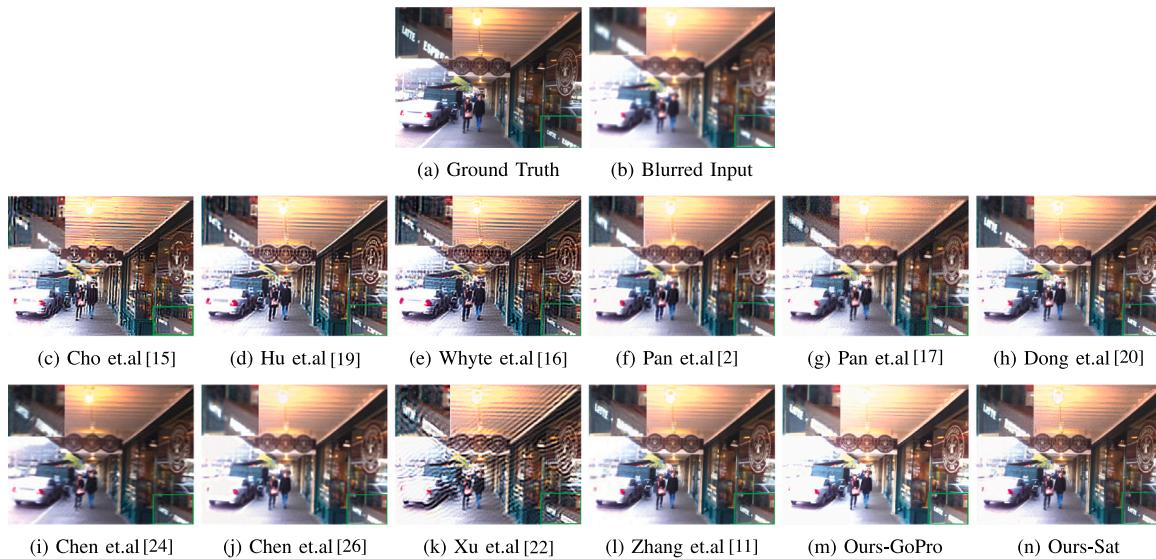
6. Conclusion

In this paper, we addressed the challenging problem of handling saturation in image deblurring by proposing the Improved Deep Multi patch framework. We introduced a nonlinear function approximator

Table 3

The results of our model, which was trained exclusively with 28 pairs of disc blurred images. The model is validated using Images 1 and 2 shown in Figs. 11(b) and 12(b) respectively.

Input	Metric	Cho [15]	Hu [19]	Whyte [16]	Pan [2]	Pan [17]	Dong [20]	Chen [24]	Chen [26]	XU [22]	Zhang [11]	Ours-GoPro	Ours-sat
Image-1	PSNR	13.8858	15.5102	15.0748	18.5626	20.8374	22.4365	19.3166	16.9585	22.5926	17.9175	20.1730	22.6854
	SSIM	0.5891	0.7210	0.6239	0.7660	0.8787	0.8954	0.8313	0.6774	0.8892	0.7900	0.8519	0.9042
Image-2	PSNR	8.7276	19.5122	17.4466	18.8281	19.8150	25.5471	23.0641	17.3327	17.7651	17.2132	23.9675	26.9978
	SSIM	0.3414	0.7507	0.6369	0.6576	0.6165	0.8627	0.7865	0.5634	0.8059	0.7380	0.8497	0.8816

**Fig. 11.** Deblurred results of Image-1 blurred by Disc Blur Kernel.**Fig. 12.** Deblurred results of Image-2 blurred by Disc Blur Kernel.**Table 4**

The results of cutting-edge models and our own. The first row shows the results of training our model with dataset 5.2.2, and the second row shows the results of training our model with dataset 5.2.2.

Blur Type	Metric	Cho [15]	Hu [19]	Whyte [16]	Pan [2]	Chen [24]	Chen [26]	Xu [22]	Zhang [11]	Ours-GoPro	Ours-sat
Motion-sat	PSNR	15.2614	16.3038	15.0748	15.3685	22.3674	21.0030	15.4520	21.9675	21.6712	22.4275
	SSIM	0.6700	0.7740	0.6239	0.6400	0.8920	0.8220	0.6862	0.8949	0.8824	0.8793
Unif-sat	PSNR	7.7900	12.6072	12.2336	14.5509	15.8813	14.6981	13.8313	14.8310	16.3862	21.5673
	SSIM	0.2959	0.5569	0.4799	0.5513	0.6624	0.5293	0.5545	0.6215	0.6830	0.8750

Table 5

Quantitative results for the non-uniform blurred input image shown in Fig. 13(b). Our model is trained with dataset 5.2.2+5.2.2+5.2.2 and 29 clean images mentioned in Section 5.2.2.

Blur type	Metric	Hu [19]	Pan [17]	Dong [20]	Chen [24]	Chen [26]	Xu [22]	Zhang [11]	Ours-GoPro	Ours-sat
Non-uniform	PSNR	13.3720	14.0471	13.7805	14.0518	13.1036	11.3121	14.3106	14.0832	14.1812
	SSIM	0.3734	0.3932	0.2546	0.3974	0.3222	0.4623	0.4026	0.3877	0.5152

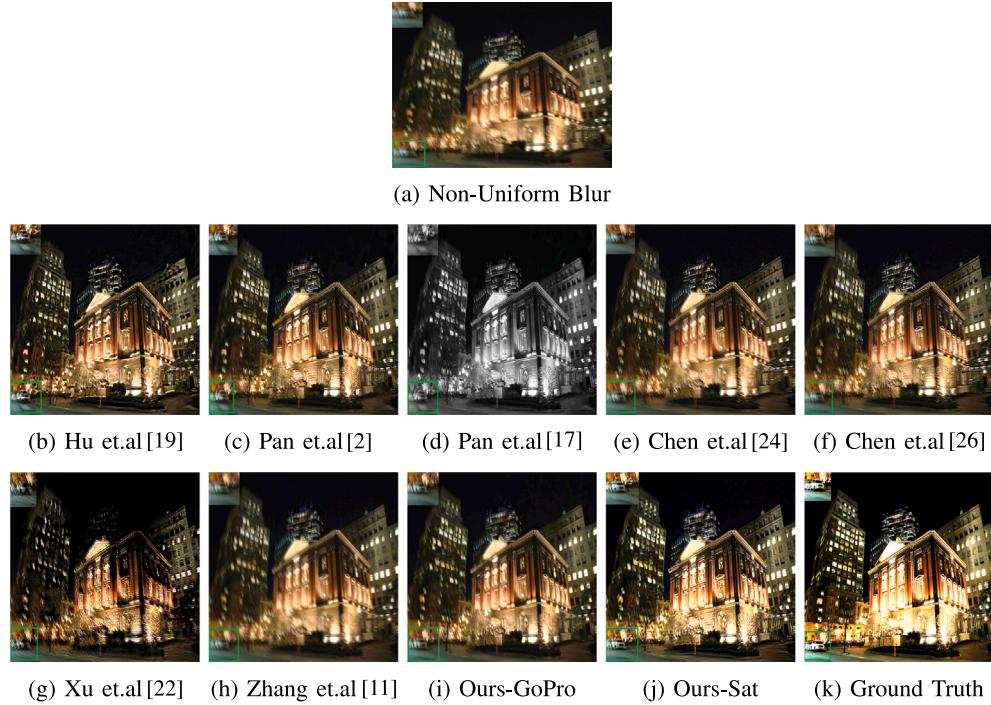


Fig. 13. Results of non-uniform blurred input image.

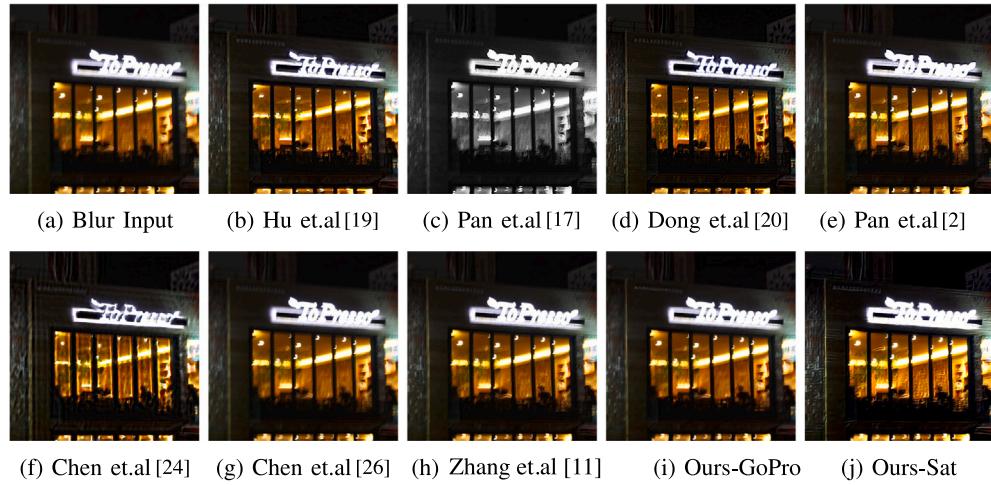


Fig. 14. Deblurred results of naturally blurred image.

better performance with increased PSNR and SSIM on GoPro, synthetic, and hard real-world examples.

7. Future scope

It is no longer beneficial to divide the blurred image into smaller grids within the network. The finest levels cannot contribute to the residuals of coarser levels because coarser levels have lower empirical losses on training data. IDMPHN's deblurring performance is enhanced

to a good extent with the help of 1×1 convolutional filters introduced at the beginning and end stages of each decoder and encoder, respectively, though there is a lack of local features in narrower grids. Furthermore, the IDMPHN performance can be improved by stacking many network models horizontally rather than vertically, as demonstrated in Fig. 15. The model's performance can be improved further by training the network using a suitable data set and fine-tuning network parameters.

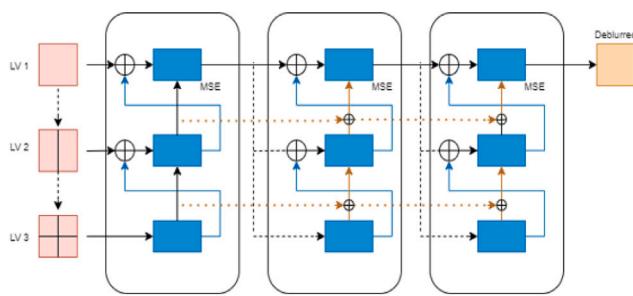


Fig. 15. Deblurring is much easier if the network is deeper horizontally by stacking several network models.

CRediT authorship contribution statement

Mahendra B.M.: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Savita Sonoli:** Project administration, Supervision, Writing – review & editing. **Abhishek Ameta:** Conceptualization, Formal analysis, Investigation, Methodology, Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors are thankful to the management of Raashreeya Sikshana Samithi Trust, Bengaluru, and Veerasaiva Vidyavardhaka Sangha, Ballari, for encouraging and helping them.

References

- [1] Shan Qi, Jia Jiaya, Agarwala Aseem. High-quality motion deblurring from a single image. *Acm Trans Graph (Tog)* 2008;27(3):1–10.
- [2] Pan Jinshan, Hu Zhe, Su Zhixun, Yang Ming-Hsuan. Deblurring text images via L0-regularized intensity and gradient prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. p. 2901–8.
- [3] Pan Jinshan, Sun Deqing, Pfister Hanspeter, Yang Ming-Hsuan. Blind image deblurring using dark channel prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. p. 1628–36.
- [4] Tao Xin, Gao Hongyun, Shen Xiaoyong, Wang Jue, Jia Jiaya. Scale-recurrent network for deep image deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. p. 8174–82.
- [5] Chen Liang, Fang Faming, Lei Shen, Li Fang, Zhang Guixu. Enhanced sparse model for blind deblurring. In: European conference on computer vision. Springer; 2020. p. 631–46.
- [6] Chen Liang, Fang Faming, Wang Tingting, Zhang Guixu. Blind image deblurring with local maximum gradient prior. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019. p. 1742–50.
- [7] Kupyn Orest, Martyniuk Tetiana, Wu Junru, Wang Zhangyang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF international conference on computer vision, 2019. p. 8878–87.
- [8] Liu Jun, Yan Ming, Zeng Tieyong. Surface-aware blind image deblurring. *IEEE Trans Pattern Anal Mach Intell* 2019;43(3):1041–55.
- [9] Xu Li, Zheng Shicheng, Jia Jiaya. Unnatural l0 sparse representation for natural image deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013. p. 1107–14.
- [10] Yan Yanyang, Ren Wenqi, Guo Yuanfang, Wang Rui, Cao Xiaochun. Image deblurring via extreme channels prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. p. 4003–11.
- [11] Zhang Hongguang, Dai Yuchao, Li Hongdong, Koniusz Piotr. Deep stacked hierarchical multi-patch network for image deblurring. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019. p. 5978–86.
- [12] Zhang Xuemei, Brainard David H. Estimation of saturated pixel values in digital color imaging. *J Opt Soc Amer A* 2004;21(12):2301–10.
- [13] Ren Wenqi, Zhang Jiawei, Ma Lin, Pan Jinshan, Cao Xiaochun, Zuo Wangmeng, et al. Deep non-blind deconvolution via generalized low-rank approximation. *Adv Neural Inf Process Syst* 2018;31.
- [14] Levin Anat. Blind motion deblurring using image statistics. *Adv Neural Inf Process Syst* 2006;19.
- [15] Cho Sunghyun, Wang Jue, Lee Seungyong. Handling outliers in non-blind image deconvolution. In: 2011 international conference on computer vision. IEEE, 2011, p. 495–502.
- [16] Whyte Oliver, Sivic Josef, Zisserman Andrew. Deblurring shaken and partially saturated images. *Int J Comput Vis* 2014;110(2):185–201.
- [17] Pan Jinshan, Lin Zhouchen, Su Zhixun, Yang Ming-Hsuan. Robust kernel estimation with outliers handling for image deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. p. 2800–8.
- [18] Nah Seungjun, Hyun Kim Tae, Mu Lee Kyoungh. Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. p. 3883–91.
- [19] Hu Zhe, Cho Sunghyun, Wang Jue, Yang Ming-Hsuan. Deblurring low-light images with light streaks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. p. 3382–9.
- [20] Dong Jiangxin, Pan Jinshan, Su Zhixun, Yang Ming-Hsuan. Blind image deblurring with outlier handling. In: Proceedings of the IEEE international conference on computer vision, 2017. p. 2478–86.
- [21] Wang Bodhi, Liu Guixiong, Wu Junfang. Blind deblurring of saturated images based on optimization and deep learning for dynamic visual inspection on the assembly line. *Symmetry* 2019;11(5):678.
- [22] Xu Li, Ren Jimmy S, Liu Ce, Jia Jiaya. Deep convolutional neural network for image deconvolution. *Adv Neural Inf Process Syst* 2014;27.
- [23] Chang Meng, Yang Chenwei, Feng Huajun, Xu Zihai, Li Qi. Beyond camera motion blur removing: How to handle outliers in deblurring. *IEEE Trans Comput Imaging* 2021;7:463–74.
- [24] Chen Liang, Zhang Jiawei, Lin Songnan, Fang Faming, Ren Jimmy S. Blind deblurring for saturated images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021. p. 6308–16.
- [25] Dong Jiangxin, Pan Jinshan. Deep outlier handling for image deblurring. *IEEE Trans Image Process* 2021;30:1799–811.
- [26] Chen Liang, Fang Faming, Zhang Jiawei, Liu Jun, Zhang Guixu. Oid: Outlier identifying and discarding in blind image deblurring. In: European conference on computer vision. Springer; 2020. p. 598–613.
- [27] Abhishek KP, Mahendra BM. Deep convolutional neural network to handle saturation in image deblurring. In: 2021 international conference on smart generation computing, communication and networking (SMART GENCON). IEEE; 2021, p. 1–5.
- [28] Lai Wei-Sheng, Huang Jia-Bin, Hu Zhe, Ahuja Narendra, Yang Ming-Hsuan. A comparative study for single image blind deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. p. 1701–9.
- [29] Bar Leah, Kiryati Nahum, Sochen Nir. Image deblurring in the presence of impulsive noise. *Int J Comput Vis* 2006;70(3):279–98.
- [30] Fergus Rob, Singh Barun, Hertzmann Aaron, Roweis Sam T, Freeman William T. Removing camera shake from a single photograph. In: ACM SIGGRAPH 2006 papers. 2006, p. 787–94.
- [31] Krishnan Dilip, Fergus Rob. Fast image deconvolution using hyper-Laplacian priors. *Adv Neural Inf Process Syst* 2009;22.
- [32] Xu Li, Jia Jiaya. Two-phase kernel estimation for robust motion deblurring. In: European conference on computer vision. Springer; 2010, p. 157–70.
- [33] Hu Zhe, Yuan Lu, Lin Stephen, Yang Ming-Hsuan. Image deblurring using smartphone inertial sensors. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. p. 1855–64.
- [34] Joshi Neel, Szeliski Richard, Kriegman David J. Psf estimation using sharp edge prediction. In: 2008 IEEE conference on computer vision and pattern recognition. IEEE; 2008, p. 1–8.
- [35] Sun Jian, Cao Wenfei, Xu Zongben, Ponce Jean. Learning a convolutional neural network for non-uniform motion blur removal. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. p. 769–77.
- [36] Xiao Lei, Wang Jue, Heidrich Wolfgang, Hirsch Michael. Learning high-order filters for efficient blind deconvolution of document photographs. In: European conference on computer vision. Springer; 2016, p. 734–49.
- [37] Zhang Kai, Zuo Wangmeng, Gu Shuhang, Zhang Lei. Learning deep CNN denoiser prior for image restoration. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. p. 3929–38.
- [38] Kupyn Orest, Budzan Volodymyr, Mykhailych Mykola, Mishkin Dmytro, Matas Jiří. Deblurgan: Blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. p. 8183–92.
- [39] Gong Dong, Yang Jie, Liu Lingqiao, Zhang Yanning, Reid Ian, Shen Chunhua, et al. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. p. 2319–28.

- [40] Wan Shengdao, Tang Shu, Xie Xianzhong, Gu Jia, Huang Rong, Ma Bin, et al. Deep convolutional-neural-network-based channel attention for single image dynamic scene blind deblurring. *IEEE Trans Circuits Syst Video Technol* 2020;31(8):2994–3009.
- [41] Liang Chih-Hung, Chen Yu-An, Liu Yueh-Cheng, Hsu Winston. Raw image deblurring. *IEEE Trans Multimed* 2020.
- [42] Li Yaowei, Pan Jinshan, Luo Ye, Lu Jianwei. Deep ranking exemplar-based dynamic scene deblurring. *IEEE Trans Image Process* 2022;31:2245–56.
- [43] Ma Haoyu, Liu Shaojun, Liao Qingmin, Zhang Juncheng, Xue Jing-Hao. Defocus image deblurring network with defocus map estimation as auxiliary task. *IEEE Trans Image Process* 2021;31:216–26.
- [44] Gao Hongyun, Tao Xin, Shen Xiaoyong, Jia Jiaya. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019. p. 3848–56.
- [45] Zhang Xinxin, Wang Ronggang, Chen Da, Zhao Yang, Gao Wen. Handling outliers by robust M-estimation in blind image deblurring. *IEEE Trans Multimed* 2020;23:3215–26.
- [46] Lin Min, Chen Qiang, Yan Shuicheng. Network in network. 2013, arXiv preprint [arXiv:1312.4400](https://arxiv.org/abs/1312.4400).
- [47] Szegedy Christian, Liu Wei, Jia Yangqing, Sermanet Pierre, Reed Scott, Anguelov Dragomir, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. p. 1–9.
- [48] Gatys Leon A, Ecker Alexander S, Bethge Matthias. A neural algorithm of artistic style. 2015, arXiv preprint [arXiv:1508.06576](https://arxiv.org/abs/1508.06576).
- [49] Divakar Nithish, Venkatesh Babu R. Image denoising via CNNs: An adversarial approach. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017. p. 80–7.