Extreme Channel Prior Embedded Network for Dynamic Scene Deblurring

Jianrui Cai[1], Wangmeng Zuo[2], Lei Zhang[1]
[1]Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
[2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
{csjcai, cslzhang}@comp.polyu.edu.hk, wmzuo@hit.edu.cn

Abstract

*Recent years have witnessed the significant progress on convolutional neural networks (CNNs) in dynamic scene deblurring. While CNN models are generally learned by the reconstruction loss defined on training data, incorporating suitable image priors as well as regularization terms into the network architecture could boost the deblurring performance. In this work, we propose an Extreme Channel Prior embedded Network (ECPeNet) to plug the extreme channel priors (*i.e.*, priors on dark and bright channels) into a network architecture for effective dynamic scene deblurring. A novel trainable extreme channel prior embedded layer (ECPeL) is developed to aggregate both extreme channel and blurry image representations, and sparse regularization is introduced to regularize the ECPeNet model learning. Furthermore, we present an effective multi-scale network architecture that works in both coarse-to-fine and fine-to-coarse manners for better exploiting information flow across scales. Experimental results on GoPro and Köhler datasets show that our proposed ECPeNet performs favorably against state-of-the-art deep image deblurring methods in terms of both quantitative metrics and visual quality.*

1. Introduction

Reproducing a high quality image faithful to the scene is an essential goal of digital photography. The real images, however, are often blurred during image acquisition due to the effect of many factors such as camera shake, object motion, and out-of-focus [27]. The resulting blurry images will not only degrade the perceptual quality of photos but also degenerate the performance of many image analytic and understanding models [22]. Blind image deblurring, which has been studied extensively in low level vision for decades of years [25], plays an essential role in improving the visual quality of real-world blurry images.

In general, the purpose of blind image deblurring is to recover the latent sharp image $\mathbf{y}$ from its blurry observation $\mathbf{x} = \mathbf{k} \otimes \mathbf{y} + \mathbf{n}$, where $\mathbf{k}$ is an unknown blur kernel (*i.e.*, uniform or non-uniform), $\mathbf{n}$ is an additive white Gaussian noise and $\otimes$ denotes the convolution operator. This inverse problem, however, is severely ill-posed and requires extra information on latent image $\mathbf{y}$ to constrain the solution space. Thus, there are two categories of approaches for utilizing prior knowledge, *i.e.*, optimization-based and deep learning based deblurring methods. Optimization-based approaches explicitly model prior knowledge to regularize the solution space of blur kernel [4, 20, 28, 37, 46] and latent image [6, 2, 30, 42, 24]. In contrast, deep learning based methods [27, 22, 38, 43] implicitly utilize prior knowledge by learning a direct mapping (*e.g.*, convolutional neural network, CNN) from degraded image to latent clean image.

For blind image deblurring problem, optimization-based and deep learning methods respectively have their merits and limitations. Optimization-based methods are flexible in incorporating versatile prior or regularization [4, 20, 46, 30, 42] tailored for blind deblurring, but suffer from the time-consuming optimization procedure and over-simplified assumptions on blur kernel (*e.g.*, spatially invariant and uniform). Moreover, conventional image priors (*e.g.*, total variation [4]) are limited in blind deblurring and prone to ordinary solution of delta kernel. Specific priors, *e.g.*, ℓ_0-norm [41] and normalized sparsity [20], are then suggested for blur kernel estimation. On the other hand, deep learning methods [27, 22, 38, 43] benefiting from end-to-end training and joint optimization can enjoy a fast speed and flexibility in handling spatially variant blur in the dynamic scene. However, deep models learn the direct mapping for blind deblurring, and may be limited in capturing specific priors for blind deblurring. As for dynamic scene deblurring, existing dataset [27] is of relatively small scale, which may be a factor hindering the performance of learned model.

Taking the merits and drawbacks of optimization-based and deep learning based methods into account, one interesting question is to ask whether we can exploit prior model to constrain both the network architecture and learning losses for improved dynamic scene deblurring performance. In this paper, we make the first attempt to address this chal-

(a) Blurry image (b) Our result (c) Ground truth

Figure 1. Deblurring result on GoPro image [27]. **Red box:** zoom-in view of the original local patch. **Green box:** zoom-in view of the dark channel of its corresponding local patch. **Blue box:** zoom-in view of the bright channel of its corresponding local patch.

lenging problem. In particular, based on the effectiveness of image prior in blind deblurring, we propose an Extreme Channel Prior embedded Network (ECPeNet) to help the restoration of latent clean image. The critical component of ECPeNet is a novel trainable extreme channel prior embedded layer (ECPeL), which can aggregate extreme channel and blurry image representations to leverage their respective advantages. By enforcing sparsity on both dark and bright channel of feature maps, we can regularize the solution space of CNN during training, thereby incorporating extreme channel priors into ECPeNet.

Furthermore, existing deep dynamic scene deblurring models [27, 38] usually adopt multi-scale network architecture but only consider the coarse-to-fine information flow. That is, blind deblurring is first performed at the small scale, and then deblurring results (or latent representations) are combined with feature representations of a larger scale for further refinement. However, we show that feature representations of a larger scale actually also benefit the dynamic scene deblurring at a smaller scale. To this end, we present a more effective multi-scale network architecture that works in both coarse-to-fine and fine-to-coarse manners for better exploiting information flow across scales. Experimental results on GoPro and Köhler datasets show that our proposed ECPeNet performs favorably against state-of-the-art deep models. With the above claimed advantages, the ECPeNet can produce visually pleasing results, as shown in Figure 1.

The contribution of this paper is three-fold: (i) We propose a novel trainable structural layer, which can aggregate both data information and prior knowledge together to leverage their respective merits but avoid limitations. To the best of our knowledge, this is the first attempt to plug prior knowledge (*i.e.*, statistical properties) into a deblurring network in an end-to-end manner. (ii) We introduce a new multi-scale network architecture to fully exploit different resolution images for maximizing the information flow. (iii) Extensive experimental results on GoPro [27] and Köhler [18] datasets demonstrate that our ECPeNet outperforms state-of-the-art dynamic scene deblurring methods.

2. Related Work

In this section, we briefly review the recent optimization-based and deep learning based image deblurring methods.

2.1. Optimization-based Deblurring Methods

The optimization-based methods aim to develop effective image priors to favor clean images over the blurry one. Representative priors include sparse gradients [6, 33, 23, 40], hyper-Laplacian prior [19], normalized sparsity prior [20], L_0-norm prior [41], patch recurrence prior [26] and discriminative learned prior [46, 24]. By taking advantages of the aforementioned priors, existing optimization-based methods could deliver competitive results on generic natural images. These approaches, however, cannot be generalized well to handle domain specific images. Thus, specific priors are needed to be introduced for specific images (*e.g.*, a light streak prior [13] for low light images, and a combination of intensity and gradient prior [29] for text images). Recently, Pan *et al*. [30] developed a dark channel prior (DCP) [10] based model to enforce sparsity on the dark channel of latent image and achieved promising result on both generic and specific images. With the success of [30], Yan *et al*. [42] further introduced a bright channel prior (BCP) to solve the corner case image, which contains a large amounts of bright pixels. By plugging the extreme channel prior (a combination of BCP and DCP) into the deblurring model, Yan *et al*. achieved state-of-the-art results on various scenarios.

Although those algorithms demonstrate their effectiveness in image deblurring, the simplified assumptions on the blur model and time-consuming parameter-tuning strategy are two lethal problems to hinder their performance in real-world cases. In this work, we utilize a realistic GoPro

dataset [27] to end-to-end train a new multi-scale network for latent sharp image restoration.

2.2. Deep Learning based Deblurring Methods

Deep learning based methods focus on exploiting external training data to learn a mapping function accords with the degradation process. The powerful end-to-end training paradigm and non-linear modeling capability make CNNs a promising approach to image deblurring. Early CNN-based deblurring methods aim to mimic conventional deblurring frameworks for the estimation of both latent image and blur kernel. Works in [36] and [8] first used networks to predict the non-uniform blur kernel and then utilized a non-blind deblurring method [45] to restore images. Schuler *et al*. [32] introduced a two-stages network to simulate iterative optimization. Chakrabarti *et al*. [3] utilized a network to predict frequency coefficients of blur kernel. However, these methods may fail when the estimated kernel is inaccurate [31]. Therefore, more recent approaches preferred to train kernel estimation-free networks to restore latent images directly. Specifically, Nah *et al*. [27] proposed a multi-scale based CNN to progressively recover the latent image. Tao *et al*. [38] introduced a scale-recurrent network which equipped with a ConvLSTM layer [39] to further ensure information flow between different resolution images. Kupyn *et al*. [22] adopted Wasserstein GAN [9, 1] as an objective function to restore the texture details of latent image. Zhang *et al*. [43] employed spatially variant recurrent neural networks (RNNs) to reduce the computational cost.

While existing deblurring networks have reported impressive results, the limited number of training data and the disappreciation of prior knowledge may become two main factors hampering the performance improvement. To mitigate aforementioned issues, we in this paper introduce the extreme channel prior for CNN and encourage it to constrain the solution space under these priors.

3. Extreme Channel Prior Embedded Network

Given a single blurry image $\boldsymbol{x_i}$, existing CNN-based methods aim at learning a mapping F_Θ to generate an estimation of latent sharp image $\hat{\boldsymbol{y}}_i$, which is required to approximate the ground-truth $\boldsymbol{y_i}$. This procedure can be formulated as:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum\nolimits_i \ell(\hat{\boldsymbol{y}}_i, \boldsymbol{y_i}) \quad s.t. \quad \hat{\boldsymbol{y}}_i = F_\Theta(\boldsymbol{x_i}) \tag{1}$$

where ($\boldsymbol{x_i}$, $\boldsymbol{y_i}$) refer to the i-th image pairs in the training dataset and Θ is the parameter of mapping function.

However, such formulation is limited in capturing image priors specified to blind deblurring, which is generally much different from those for non-blind restoration. Moreover, the existing training image pairs are insufficient to learn an effective mapping function F_Θ. Therefore, we propose ECPeNet that aggregates both blurry image and extreme channel representations to enhance deblurring performance, and Eqn. (1) can be rewritten as:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum\nolimits_i \ell(\hat{\boldsymbol{y}}_i, \boldsymbol{y_i}) \quad s.t. \quad \hat{\boldsymbol{y}}_i = F_\Theta(\boldsymbol{x_i}|\Lambda, \Omega) \tag{2}$$

where Λ and Ω are the extreme channel representations under the constraint of both dark and bright channels priors. By this way, we can embed the image priors specified to blind deblurring into the mapping function F_Θ, and expect it to have ability to generate the higher quality latent sharp image. In the following subsections, we present in detail the proposed ECPeNet.

3.1. Architecture

The overall architecture of our proposed ECPeNet is illustrated in Figure 2. It contains three sub-networks respectively for three scales, and each of them consists of three major components: (i) input and output; (ii) encoder and decoder; (iii) feature mapper. Note that instead of utilizing Rectifier Linear Units (ReLU) [21], we take parametric ReLU (PReLU) [11] as activation function since it can improve the modeling capability with negligible extra computational cost. Unless denoted otherwise, all the convolution filters are set to 3×3, not 5×5 that utilized by most of the other dynamic scene deblurring networks (*e.g*., [27] and [38]). Although a filter of size 5×5 has more parameters than two filters of size 3×3, the one utilizing 3×3 filters is more efficient since additional nonlinearity can be inserted between them [35]. Besides, the stride size for all convolution layers is set to 1 and the number of feature maps in each layer is set to 64, except for the last layer and the proposed ECPeL which are set to 3 and $\{3, 64, 3\}$, respectively. The details of each component are described as follows.

Input and Output. An effective multi-scale network architecture is utilized in this work to restore the latent sharp image from a coarser scale to finer scale. Consider an image pair $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{H \times W \times C}$, where $H \times W$ is the pixel resolution and C is the number of channels, which is set to 3. We first use bicubic interpolation to progressively downsample the image pair with the ratio of $\frac{1}{2}$, and generate 3 scales image pairs with the resolution of $\{H \times W, \frac{H}{2} \times \frac{W}{2}, \frac{H}{4} \times \frac{W}{4}\}$. We then take the blurred image at each scales as input to produce its corresponding sharp images. The sharp one at original resolution is considered as the final output.

Encoder and Decoder. Each scale of encoder consists of 4 convolution layers, the proposed ECPeL and a shuffle operation with factor $\frac{1}{2}$. As for decoders, they basically mirror the architecture of the encoders, except for the factor of shuffle operation [34] is set to 2. The encoder and decoder networks are mainly designed for three purposes.

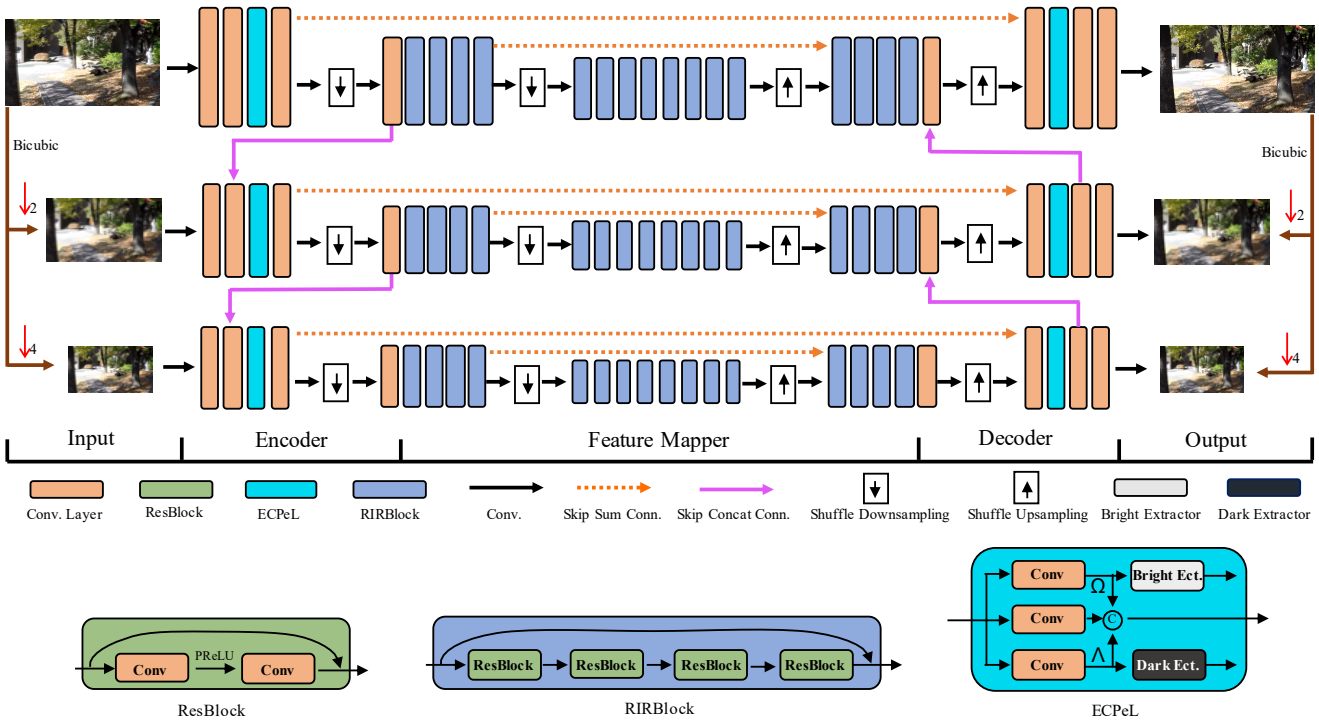

Figure 2. Illustration of our proposed ECPeNet architecture.

Firstly, they progressively transform different scale images to extract shallow features (in encoders) and transform them back to the resolution of inputs (in decoders).

Secondly, they downsample and upsample the shallow features to ensure information flow between different scale images and expand the receptive field. Note that instead of directly concatenating the upsampled coarser scale latent image with the finer scale blurry image [27, 5, 38], we argue that (i) the information of finer scale blurry image is beneficial for estimating the coarser scale latent image and (ii) the concatenation in shallow feature domain can yield a better result than in image (RGB) domain. Thus, the encoder first shuffles features with factor $\frac{1}{2}$ to ensure the same resolution between different scales features, *e.g.*, a features with size $m \times n \times c$ is shuffled to $\frac{m}{2} \times \frac{n}{2} \times 4c$. Then it concatenates the downsampled features with coarser scale features for restoring the coarser scale sharp image. On the contrary, the decoder shuffles features with size $\frac{m}{2} \times \frac{n}{2} \times 4c$ back to $m \times n \times c$ and concatenates them with finer scale features for predicting the finer scale sharp image. Benefiting from this multi-scale architecture, the network can fully exploit different scale images to maximize the information flow between them, resulting in better performance. In Section 4.1, we conduct an ablation study to verify its effectiveness.

Thirdly, they integrate the extreme channel prior into the network via ECPeL. We provide more details of the proposed ECPeL in Section 3.2.

Feature Mapper. The feature mapper module, which aims to refine the shallow features progressively, is an essential part of the latent image restoration. One critical factor for reducing blur artifacts is the size of receptive field. To enlarge the receptive field, we (i) stack a set of convolutional layers to achieve a larger depth network and (ii) utilize the shuffle operations for downsampling and upsampling the features. Considering that a deeper neural network is more challenging to converge, we adopt the residual blocks to speed-up the training procedure [12, 15]. Besides, the feature mapper module utilizes the long skip connection and short skip connection to make full use of hierarchical features in all convolutional layers. A similar skip connection strategy has been utilized in a very recent work [44]. As illustrated in Figure 2, the feature mapper contains 16 residual in residual blocks (RIRBlock), and each of them has 4 residual blocks (ResBlock). The ResBlock consists of 2 convolution layers and a PReLU activation function. Since we utilize the filter of size 3×3, the total parameters of our ECPeNet are almost in the same magnitude as the previous methods. Note that all the weights across different scales feature mapper sub-modules are shared.

3.2. Extreme Channel Prior Embedded Layer

The proposed ECPeL is designed for aggregating both blurry image representation and extreme channel representation to regularize the solution space of CNN. Specifically,

it first learns 3 mapping functions $\mathcal{M}_\theta$, $\mathcal{M}_{[\alpha|\mathcal{D}]}$ and $\mathcal{M}_{[\beta|B]}$ to transform the feature map f^{l-1} from previous layer into 3 new feature maps, including a deeper layer transformed feature f^l, a dark channel prior constrained feature Λ, and a bright channel prior constrained feature Ω. It then adopts a concatenation operation to concatenate those 3 feature maps for the integration of blurry image representation and extreme channel representation. Formally, the proposed ECPeL can be expressed as:

$$\begin{aligned}[\Lambda, f^l, \Omega] &= \textit{ECPeL}(f^{l-1}) \\ f^l &= \mathcal{M}_\theta(f^{l-1}) \\ \Lambda &= \mathcal{M}_{[\alpha|D]}(f^{l-1}) \\ \Omega &= \mathcal{M}_{[\beta|B]}(f^{l-1})\end{aligned} \tag{3}$$

where $[\Lambda, f^l, \Omega]$ denotes the concatenation of the feature maps, and the subscripts $[\alpha|D]$ and $[\beta|B]$ denote that parameters α and β are optimized under the dark and bright channel prior constraint. To add extreme channel prior constraint into a network, the ECPeL utilizes (i) extractors to extract both dark and bright channel of features, and (ii) the ℓ_1-regularization term to enforce sparsity in training.

The extractor $D(\cdot)$ is designed to extract the dark channel of Λ via computing its minimum values in a local patch. Formally, its `forwards` function can be written as follows:

$$\begin{aligned}D(\Lambda)_{[h,w]} &= \Lambda_{[\mathcal{I}_{\mathcal{D}[h,w]}]} \\ \mathcal{I}_{\mathcal{D}[h,w]} &= \text{argmin}_{i^\star \in \Psi_{[h,w,c]}} \Lambda_{[i^\star]}\end{aligned} \tag{4}$$

The extractor $B(\cdot)$ aims to extract the bright channel of Ω by calculating its maximum values in a local patch. And its `forwards` function can be formulated as:

$$\begin{aligned}B(\Omega)_{[h,w]} &= \Omega_{[\mathcal{I}_{\mathcal{B}[h,w]}]} \\ \mathcal{I}_{\mathcal{B}[h,w]} &= \text{argmax}_{i^\star \in \Psi_{[h,w,c]}} \Omega_{[i^\star]}\end{aligned} \tag{5}$$

where $\Psi_{[h,w,c]}$ is the index set of inputs in a sub-window centered at a pixel location $[h, w, c]$, $\mathcal{I}_{\mathcal{D}[h,w]}$ and $\mathcal{I}_{\mathcal{B}[h,w]}$ are the masks that records an index of the minimum and maximum value in a local patch, respectively. The patch sizes for each scale are set to $\{31\times31, 19\times19, 11\times11\}$. A single element $\Lambda_{[h,w,c]}$ and $\Omega_{[h,w,c]}$ of the input may be assigned to several different outputs $D(\Lambda)_{[h,w]}$ and $B(\Omega)_{[h,w]}$.

The `backwards` function of extractors computes partial derivative of the loss function with respect to each input variable Λ_i and Ω_i as follows:

$$\begin{aligned}\frac{\partial L}{\partial \Lambda_i} &= \sum_h \sum_w \sum_c 1\{i = \mathcal{I}_{\mathcal{D}[h,w]}\} \frac{\partial L}{\partial D(\Lambda)_{[h,w]}} \\ \frac{\partial L}{\partial \Omega_i} &= \sum_h \sum_w \sum_c 1\{i = \mathcal{I}_{\mathcal{B}[h,w]}\} \frac{\partial L}{\partial B(\Omega)_{[h,w]}}\end{aligned} \tag{6}$$

where i refers to the pixel location $[h, w, c]$. In words, the partial derivatives $\frac{\partial L}{\partial D(\Lambda)_{[h,w]}}$ and $\frac{\partial L}{\partial B(\Omega)_{[h,w]}}$ are accumulated if i is the argmin and argmax selected for $D(\Lambda)_{[h,w]}$ and $B(\Omega)_{[h,w]}$, respectively. In back-propagation, the partial derivatives $\frac{\partial L}{\partial D(\Lambda)_{[h,w]}}$ and $\frac{\partial L}{\partial B(\Omega)_{[h,w]}}$ are already calculated by the `backwards` function of the loss layer.

With the proposed ECPeL, we can extract the dark and bright channel of shallow features (*i.e.*, $D(\Lambda)$ and $B(\Omega)$), which can be further enforced to be sparse via the objective function. By integrating the constrained features Λ and Ω into the network, the proposed ECPeNet can achieve a better performance while using the same training dataset. The ablation study in Section 4.1 is conducted for the evaluation.

3.3. Loss Function

We utilize the ℓ_1-norm of the reconstruction error as loss function for each scale. More specifically, we can rewrite Eqn. (2) as:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{3} \|\boldsymbol{y_i^j} - F_\Theta(\boldsymbol{x_i^j}|\Lambda^{\boldsymbol{j}}, \Omega^{\boldsymbol{j}})\|_1 \tag{7}$$

where N is the total number of training pairs $(\boldsymbol{x}, \boldsymbol{y})$ and $\boldsymbol{j}$ is the number of scales, which is set to 3 in this paper. $(\cdot)^{\boldsymbol{j}}$ is a symbol referring to the image and feature in the $\boldsymbol{j}$-th scale.

According to the observation by [30, 42], a dark channel of images would be less dark after the blurring process, while a bright channel of images would be no longer bright. The reason is that dark/bright pixel would be averaged with its neighboring high/low intensity pixels during a blurring process. The sparsity regularization term is thus more beneficial for restoring a sharp image than a blurred one. To this end, we introduce a ℓ_1-regularization term to enforce sparsity on both dark and bright channels of shallow features. The objective function can be given by:

$$\begin{aligned}\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{3} &\|\boldsymbol{y_i^j} - F_\Theta(\boldsymbol{x_i^j}|\Lambda^{\boldsymbol{j}}, \Omega^{\boldsymbol{j}})\|_1 \\ &+ \lambda\|D(\Lambda^{\boldsymbol{j}})\|_1 + \omega\|1 - B(\Omega^{\boldsymbol{j}})\|_1\end{aligned} \tag{8}$$

where λ and ω are the trade-off parameters. $D(\cdot)$ and $B(\cdot)$ are the extractors to extract the dark channel and bright channel of features, respectively. With the `forwards` and `backwards` functions, the dark and bright channel extractors can be jointly end-to-end optimized with the network.

4. Experimental Results

In this section, we provide experimental results to show the advantage of our proposed ECPeNet. We implement our framework by using Caffe toolbox [14], and train the model on a PC equipped with an Intel Core i7-7820X CPU, 128G RAM and a single Nvidia Quadro GV100 GPU.

Table 1. Ablation study of extreme channel prior (ECP) constraint and image fully exploitation (IFE) strategy. The average PSNR (dB) on GoPro testing dataset with $150K$ iterations.

	Different combinations of ECP and IFE			
ECP	×	√	×	√
IFE	×	×	√	√
PSNR	28.63	28.86	28.79	28.95

Datasets. We train our proposed ECPeNet on the GoPro training dataset [27], which contains 22 sequences with $2,103$ blurred/clear image pairs. Once the model is trained, we test it on the standard GoPro testing dataset [27] and Köhler [18] dataset. The GoPro testing dataset consists of 11 sequences with $1,111$ image pairs, and the Köhler dataset has 4 latent images and 12 blur kernels. Note that, to simulate the realistic blurring process, the GoPro dataset generates blurred images through averaging adjacent short-exposure frames captured by a high-speed video camera, and the Köhler dataset replays the recorded $6D$ real camera motion trajectory to synthesize blurred images.

Parameter Settings. We crop the GoPro training dataset (linear subset) into $256 \times 256 \times 3$ patches and make use of these patches to train the ECPeNet. The mini-batch size in all the experiments is set to 10, and the trade-off parameters λ and ω are set to $0.1(0.2)$ in this paper. For the model training, we utilize Xavier [7] to initialize all trainable variables. The Adam solver [17] is adopted to optimize the network parameters. The default parameters of Adam solver are set as $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. We fix the learning rate as 10^{-4} and train the network with $600K$ iterations, which takes about 140 hours. Additionally, we randomly rotate and/or flip the image patches for data augmentation. The 1% additive Gaussian noise is also randomly added to the blurred images for robust learning.

4.1. Ablation Study

It is generally agreed that a larger scale training dataset which covers various image contents and blur models will bring benefit to train a robust deep network. The type of scenes and number of images in the current GoPro dataset, however, are barely sufficient to train an efficient network. Rather than enlarging the training dataset, in this work, we propose to integrate the extreme channel prior into CNN and fully exploit different scales images for the performance improvement. Here, we compare our network with several baseline models to verify the effectiveness of the extreme channel prior (ECP) constraint and image fully exploitation (IFE) strategy. Table 1 shows the investigation on the effects of ECP constraint and IFE strategy. It can be seen that the one without both ECP and IFE performs much worse than the ECPeNet in terms of PSNR (28.63 dB $v.s.$ 28.95 dB).

Table 2. Average PSNR (dB), SSIM, MSSIM indices and running time for different methods on the benchmark datasets (running time is measured for an image with the size of $1280 \times 720 \times 3$).

Method	GoPro		Köhler		Time
	PSNR	SSIM	PSNR	MSSIM	
Kim [16]	23.64	0.824	24.68	0.794	1 hr
Sun [36]	24.64	0.843	25.22	0.774	20 min
Nah [27]	29.08	0.914	26.48	0.808	$2.87s$
Tao [38]	30.26	0.934	26.75	0.837	$0.62s$
Kupyn [22]	28.70	0.858	26.10	0.816	$0.59s$
Zhang [43]	29.19	0.931	25.71	0.800	$0.76s$
Proposed	31.10	0.945	26.79	0.839	$0.65s$

By plugging the ECP constrain into the network, we can significantly improve the performance with a 0.23 dB gain. While compared to the networks that utilizes the multi-scale architectures [27, 38] or cascaded refinement [5] strategy, the one adopting the proposed IFE strategy (both coarse-to-fine and fine-to-coarse manners) can has 0.16 dB improvement in terms of PSNR index. Note that we train all these 4 networks with $150K$ iterations for the testing in this ablation study. These comparisons firmly indicate the proposed ECP and IFE benefit for performance improvement.

4.2. Comparisons with State-of-the-art Methods

In this subsection, both quantitative and qualitative evaluations are conducted to verify the proposed ECPeNet on the benchmark datasets.

Quantitative Evaluations. We first compare the proposed ECPeNet with previous state-of-the-art deblurring methods [16, 36, 27, 22, 43, 38] in a quantitative way. The source codes and trained models of the aforementioned methods are publicly available on the authors' websites, except for [16] and [43] whose results have been reported in previous works [27] and [43], respectively. Additionally, we utilize the same training dataset to retrain the network provided by [43] for its evaluation on Köhler dataset. The average PSNR, SSIM, and MSSIM indices for different deblurring methods on GoPro testing and Köhler datasets are shown in Table 2. One can see that on the GoPro testing dataset, the proposed ECPeNet significantly outperforms both the conventional non-uniform deblurring method [16] and these recent developed CNN based methods [36, 27, 22, 43, 38]. Even compared to the previous state-of-the-art method [38], the proposed ECPeNet still has a 0.84 dB lead. While on the Köhler dataset, although the performance of these dynamic scene deblurring networks is comparable, it can be found that our method still has a slight advantage. Meanwhile, the running time by different methods for processing an image of resolution $1280 \times 720 \times 3$ is also listed in Table 2. One can notice that it takes plenty of time for a conventional

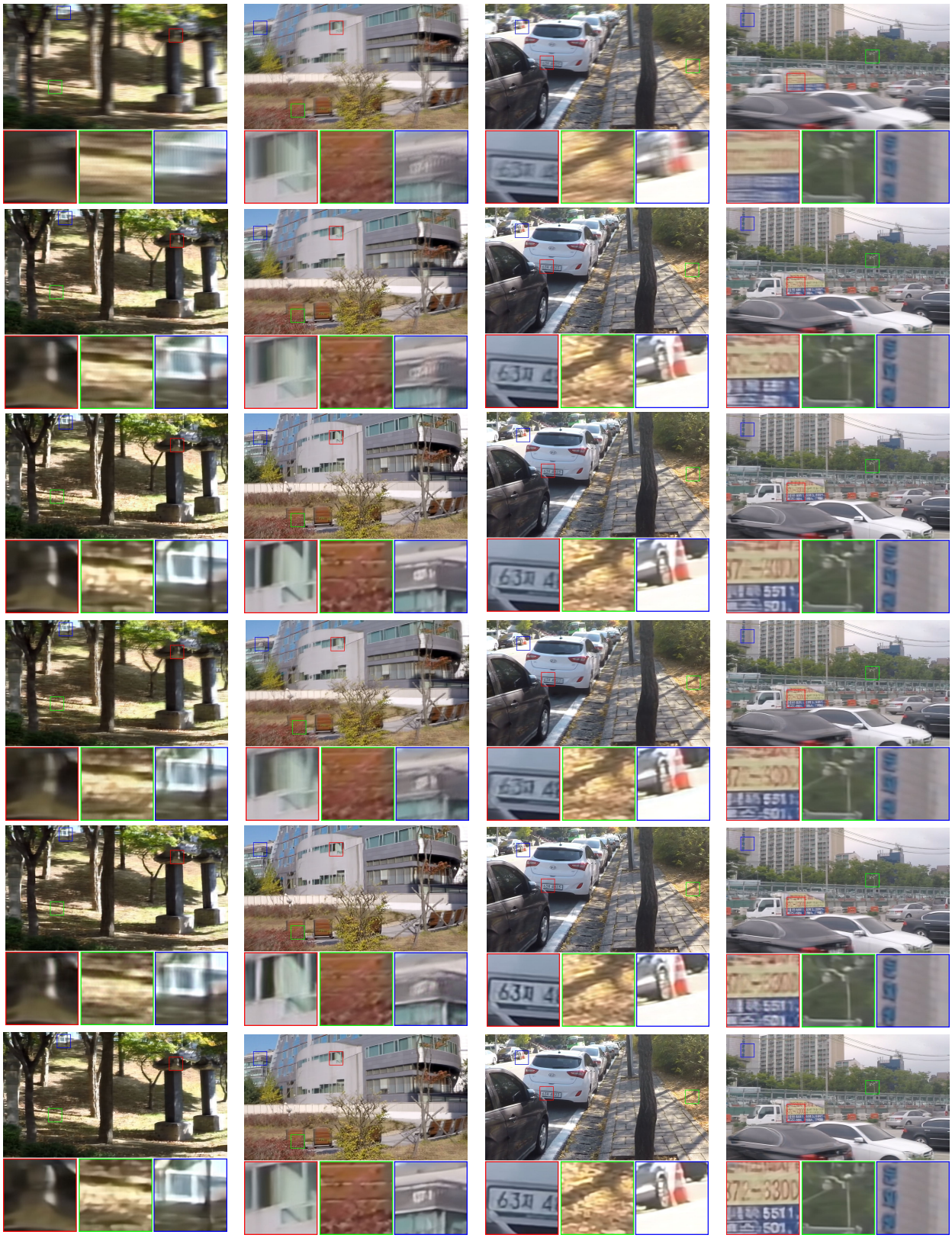

Figure 3. Deblurring results on GoPro dataset [27] by different methods. In the top-down order, we show inputs, results of Nah *et al*. [27], Tao *et al*. [38], Kupyn *et al*. [22], Zhang *et al*. [43], and our results.

(a) Blurry image (b) Nah *et al*. [27] (c) Tao *et al*. [38]

(d) Kupyn *et al*. [22] (e) Zhang *et al*. [43] (f) Proposed

Figure 4. Deblurring results on Köhler dataset [18] by different methods.

method to restore an image because of the time-consuming iterative inference and the CPU implementation. While for these end-to-end training networks, they can achieve much faster speed to process an image on GPU. Considering that these dynamic scene deblurring networks are implemented by different deep learning platforms, the minor difference between them can be neglected within the margin of error.

Qualitative Evaluations. We further compare the visual quality of restored images by our proposed ECPeNet and these recent developed CNN based dynamic scene deblurring networks, including Nah [27], Tao [38], Kupyn [22], and Zhang [43]. Figure 3 shows several blurred images from the GoPro [27] testing dataset and their corresponding deblurring results produced by the above methods. One can see that although these recent developed CNNs could remove the overall motion blur artifacts, the results restored by them are not pleasant enough because of the blurred edges and noticeable artifacts. For example, in the fourth column, all these previous CNN based deblurring networks could not recover the text information (see the red box zoom-in region). While in the second column, the noticeable artifacts exist around the license plate number. By contrast, benefiting from the extreme channel prior constraint, our method can deliver a more visual pleasing result with much fewer artifacts and sharper edges.

To further demonstrate the robustness of our method, the visual comparison results on images from the Köhler [18] dataset are also provided in Figure 4. Again, it can be seen that artifacts and blurred edges in the zoom-in areas (see characters 'B', '70', and '15') are noticeable for these previous CNN based methods. Although results recovered by Kupyn [22] and Tao [43] are sharper than other methods, distortion still exists. Compared with these methods, the ECPeNet can restore image sharpness and naturalness.

5. Conclusion

In this work, we presented a simple yet effective Extreme Channel Prior embedded Network (ECPeNet) with a novel trainable extreme channel prior embedded layer (ECPeL), which aims to integrate extreme (*i.e*., dark and bright) channel priors into a deep CNN for dynamic scene deblurring. By extracting the extreme channels of shallow features and enforcing sparsity on them, ECPeNet can regularize the solution space of the network. Additionally, ECPeNet works in both coarse-to-fine and fine-to-coarse manners to exploit information of blurred images at different resolutions to maximize information flow across scales. Benefiting from the extreme channel prior constraint and effective multi-scale network architecture, the developed ECPeNet outperforms previous dynamic scene deblurring networks by a large margin. Quantitative evaluations on the challenging GoPro dataset showed that the proposed ECPeNet had at least 0.84 dB PSNR gains over the existing state-of-the-arts.

References

[1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. *stat*, 1050:26, 2017. 3

[2] Y. Bahat, N. Efrat, and M. Irani. Non-uniform blind deblurring by reblurring. In *ICCV*, 2017. 1

[3] A. Chakrabarti. A neural approach to blind motion deblurring. In *ECCV*, 2016. 3

[4] T. F. Chan and C.-K. Wong. Total variation blind deconvolution. *IEEE TIP*, 7(3):370–375, 1998. 1

[5] Q. Chen and V. Koltun. Photographic image synthesis with cascaded refinement networks. In *ICCV*, 2017. 4, 6

[6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *ACM transactions on graphics (TOG)*, volume 25, pages 787–794. ACM, 2006. 1, 2

[7] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*, 2010. 6

[8] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. Van Den Hengel, and Q. Shi. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *CVPR*, 2017. 3

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPs*, 2014. 3

[10] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *CVPR*, 2011. 2

[11] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, 2015. 3

[12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 4

[13] Z. Hu, S. Cho, J. Wang, and M.-H. Yang. Deblurring low-light images with light streaks. In *CVPR*, 2014. 2

[14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *ACM MM*, 2014. 5

[15] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 2016. 4

[16] T. H. Kim and K. M. Lee. Segmentation-free dynamic scene deblurring. In *CVPR*, 2014. 6

[17] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2014. 6

[18] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *ECCV*, 2012. 2, 6, 8

[19] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In *NIPs*. 2009. 2

[20] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR*, 2011. 1, 2

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPs*, 2012. 3

[22] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *CVPR*, 2018. 1, 3, 6, 7, 8

[23] A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, 2009. 2

[24] L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, and M.-H. Yang. Learning a discriminative prior for blind image deblurring. In *CVPR*, 2018. 1, 2

[25] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The astronomical journal*, 79:745, 1974. 1

[26] T. Michaeli and M. Irani. Blind deblurring using internal patch recurrence. In *ECCV*, 2014. 2

[27] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 1, 2, 3, 4, 6, 7, 8

[28] J. Pan, Z. Hu, Z. Su, and M.-H. Yang. Deblurring face images with exemplars. In *ECCV*, 2014. 1

[29] J. Pan, Z. Hu, Z. Su, and M.-H. Yang. L_0-regularized intensity and gradient prior for deblurring text images and beyond. *IEEE TPAMI*, 39(2):342–355, 2017. 2

[30] J. Pan, D. Sun, H. Pfister, and M.-H. Yang. Blind image deblurring using dark channel prior. In *CVPR*, 2016. 1, 2, 5

[31] D. Ren, W. Zuo, D. Zhang, J. Xu, and L. Zhang. Partial deconvolution with inaccurate blur kernel. *IEEE TIP*, 27(1):511–524, 2018. 3

[32] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Scholkopf. Learning to deblur. *IEEE TPAMI*, (7):1439–1451, 2016. 3

[33] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *Acm transactions on graphics (TOG)*, volume 27, page 73. ACM, 2008. 2

[34] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *CVPR*, 2016. 3

[35] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. 3

[36] J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *CVPR*, 2015. 3, 6

[37] L. Sun, S. Cho, J. Wang, and J. Hays. Edge-based blur kernel estimation using patch priors. In *ICCP*, 2013. 1

[38] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia. Scale-recurrent network for deep image deblurring. In *CVPR*, 2018. 1, 2, 3, 4, 6, 7, 8

[39] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *NIPs*, 2015. 3

[40] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, 2010. 2

[41] L. Xu, S. Zheng, and J. Jia. Unnatural l_0 sparse representation for natural image deblurring. In *CVPR*, 2013. 1, 2

[42] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao. Image deblurring via extreme channels prior. In *CVPR*, 2017. 1, 2, 5

[43] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R. W. Lau, and M.-H. Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In *CVPR*, 2018. 1, 3, 6, 7, 8

[44] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 4

[45] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *ICCV*, 2011. 3

[46] W. Zuo, D. Ren, S. Gu, L. Lin, and L. Zhang. Discriminative learning of iteration-wise priors for blind deconvolution. In *CVPR*, 2015. 1, 2