Image Deblurring with a Class-Specific Prior

Saeed Anwar, Cong Phuoc Huynh, Fatih Porikli

Abstract—A fundamental problem in image deblurring is to recover reliably distinct spatial frequencies that have been suppressed by the blur kernel. To tackle this issue, existing image deblurring techniques often rely on generic image priors such as the sparsity of salient features including image gradients and edges. However, these priors only help recover part of the frequency spectrum, such as the frequencies near the high-end. To this end, we pose the following specific questions: (i) Does any image class information offer an advantage over existing generic priors for image quality restoration? (ii) If a class-specific prior exists, how should it be encoded into a deblurring framework to recover attenuated image frequencies? Throughout this work, we devise a class-specific prior based on the band-pass filter responses and incorporate it into a deblurring strategy. More specifically, we show that the subspace of band-pass filtered images and their intensity distributions serve as useful priors for recovering image frequencies that are difficult to recover by generic image priors. We demonstrate that our image deblurring framework, when equipped with the above priors, significantly outperforms many state-of-the-art methods using generic image priors or class-specific exemplars.

Index Terms—image deblurring, blind deconvolution, image prior, class prior.

1 Introduction

Image deblurring is an important and long-standing research challenge in low-level vision dating back to 1960s [1]. Blur due to camera shake and camera motion is still a prevalent issue with images captured by hand-held devices, *e.g*. smartphones or tablet computers. With an exponentially increasing amount of image data captured by these devices, there has been continuing research effort in image deblurring in the last decade [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15].

In this paper, we focus our attention on the case of uniform blur, in which a sharp image is convolved with a spatially uniform blur kernel. The goal of blind image deblurring is hence viewed as solving for the latent image $\mathbf{x}$ and the kernel $\mathbf{k}$ given the blurred image $\mathbf{y}$.

By nature, image deblurring is an ill-posed problem, as there exists an infinite number of pairs of latent image $\mathbf{x}$ and kernel $\mathbf{k}$ that result in the same observation $\mathbf{y}$.

To resolve the above ambiguity, previous works have exploited the sparsity of natural image gradients to impose additional constraints on the deblurring problem. This sparsity constraint is commonly stated in terms of the hyper-Laplacian prior [16], [17], the ℓ_0 [9], ℓ_1 [6] and ℓ_2-norms [5], the ℓ_1/ℓ_2 prior [7], a Gaussian [18] or a mixture of Gaussians [3] of the image gradients. A common feature in these works is the presence of a regulariser that minimises the sparsity of the image gradient. As a result, these methods favour images with strong high-frequency components while ignoring other spatial frequencies. For this reason, these methods are not suitable for many object categories with gradual changes in the surface orientation such as faces, animals, cars, etc.

Furthermore, a common symptom of deblurred images is the presence of ringing artifacts. Mosleh *et al*. [11] has proposed a solution to the detection and removal of ringing by generating a set of Gabor filters that reveals existing ringing artifacts in deblurred images and incorporating these filters in a regularisation scheme to suppress the artifacts. Meanwhile, Whyte *et al*. [15] address the issue of ringing reduction in the presence of saturated pixels. We draw a general remark, from the analysis of these works, that the main cause of ringing is the suppression of some spatial frequencies by the blur kernel. The frequencies missing from the blur kernel usually cause the Fourier sum of the remaining waves to overshoot at jumps in image intensity. This is known as the Gibbs phenomenon [19], rendering ringing artifacts to appear near strong edges.

To overcome the above problem, we leverage prior knowledge of the distribution of frequency components specific to each image class, rather than generic gradient sparsity priors. As a natural choice, we analyse images in the Fourier space due to the convenient transformation of the blur model between the spatial and transfer domain. Instead of imposing a general sparsity constraint, we focus on modeling a class-specific prior in each band of the Fourier spectrum. Specifically, we learn a subspace spanned by the filter responses of sharp images in each class to a band-pass filter. Repeating this learning process over multiple band-pass filters, we capture the characteristics of the target image class across a wide range of frequency bands. The spirit of this work is to discover a more comprehensive prior than those based exclusively on edges or high-frequency image gradients. With our learned priors in hand, we perform the deblurring process in a content-aware fashion.

Figure 1 depicts our approach to the restoration of the spatial frequencies attenuated by a blurring kernel. In the first row, we display a sample image from the Cat dataset [20] (first column) and the magnitudes of its Fourier components in three frequency bands (the subsequent columns). The frequency components in each band are obtained by a convolution of the input image with a Butterworth band-pass filter. Although most of the frequency

• *S. Anwar, C. P. Huynh, F. Porikli are with the Research School of Engineering, Australian National University, and Data61/CSIRO.* *This research was supported under Australian Research Council's Discovery Projects funding scheme (project number DP150104645) and an Australian Government RTP Scholarship.*
E-mail: (saeed.anwar, cong.huynh, fatih.porikli)@anu.edu.au

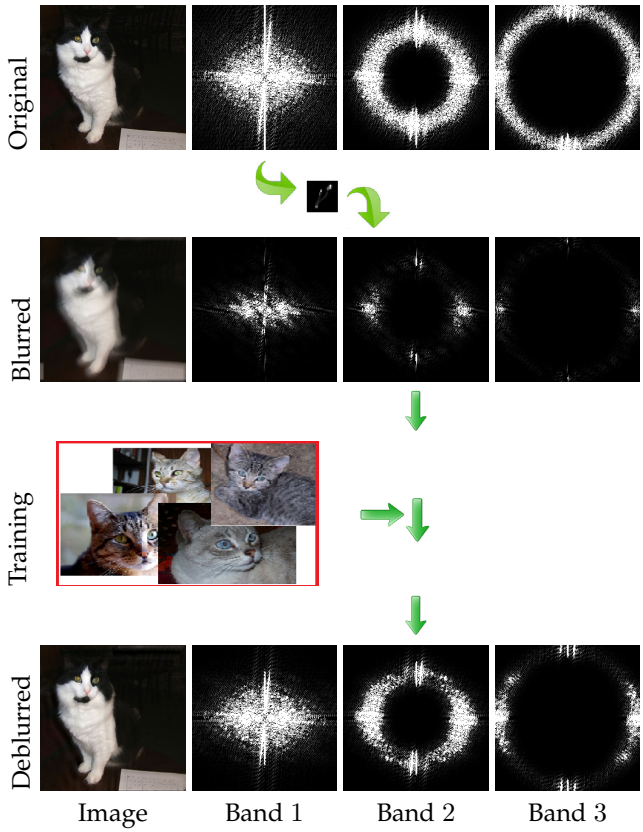


Fig. 1: Recovering spatial frequencies that have been suppressed by a blur kernel using band-pass frequency components from the training data.

components of the blurred image (second row) have been annihilated, the recovered image (shown in the last row) contains many frequency components present in the original (in the third row).

Our method is inspired by previous works on image categorisation using image statistics. In [21], the authors investigated the spectral signature, i.e., the power spectra of the horizontal and vertical image gradients for each image category. The shape of this spectral signature is an indicator of the scale (size) of the primary element in the scene. This study revealed significant variations in the power spectrum across different image categories, which could enable the categorisation of natural and man-made images. In a related work, Geusebroek and Smeulders [22] modeled the spatial statistics using a parametric Weibull distribution for the characterisation of uniform stochastic textures. Building on this model, subsequent works have proposed methods for image categorisation using local texture descriptors [23]. Specific to image deblurring, Levin [24] integrated the statistics of derivative filters into a maximum likelihood method for blind motion deblurring. However, this study was limited to blurs caused only by a one-dimensional box kernel. The other practical limitation is that it requires the segmentation of the image into layers with common blurs.

We advance the above formulation of image statistics for the purpose of image characterisation. In the previous works, image statistics constitute the power spectra of image gradients or derivative filters, which can be viewed as responses to high-frequency filters. In our work, we generalise this notion and consider the distribution of image responses to band-pass filters across all the bands in the frequency spectrum. The novel class prior is based on the following conjectures. Firstly, for every image band, the distribution of band-pass filter responses is characteristic of the image class. Secondly, the band-pass filter responses of images in the same class span a linear subspace. As we shall demonstrate later, these two underpinning conjectures alone are proven to be effective in recovering frequencies suppressed by blur kernels.

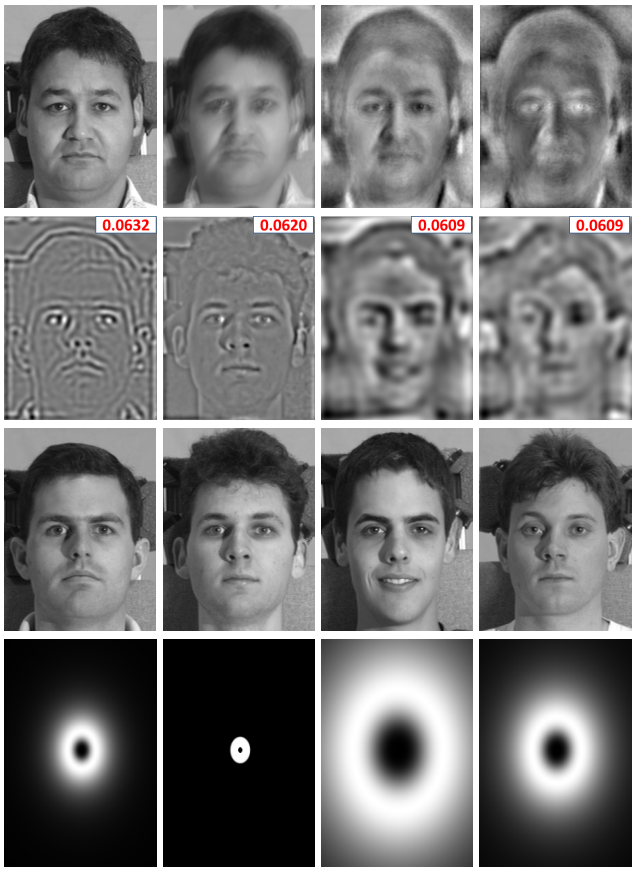


Fig. 2: A visual demonstration of the proposed prior. Top row (from left to right): original (ground-truth) image $\mathbf{x}^*$, input blurred image $\mathbf{y}$, the image reconstructed by the weighted combination of **200** filtered training images, and the absolute difference $\|\mathbf{x} - \mathbf{x}^*\|$. Second row (from left to right): the four most important filtered training images sorted by the descending order of their weights (shown in the inset). Third row: the training images corresponding to those in the second row. Fourth row: the bandpass filters (shown in the frequency domain) involved in the filtered training images in the second row.

To perform blind deblurring, we incorporate the linear subspaces of band-pass filter responses as a class-specific image prior, together with a common ℓ_2-norm kernel prior into a joint objective function. Subsequently, we employ an iterative optimisation approach over several coarse-to-fine image resolutions. In each iteration, the latent image and kernel can be alternately computed as a closed-form solution.

We provide a visual illustration of the relevant training images and bandpass filters selected by the proposed image prior. Figure 2 shows an example blurry image (in the second column of the first row), and the four most relevant filtered training images in the second row, together with their weights (shown in the inset). The corresponding training images are displayed in the third row and the associated bandpass filters are in the fourth row.

It can be seen that, the algorithm selects a variety of frequency components from different training images to compose the latent image, including low-frequency details from the first two training images, and mid-frequency details from the latter two. Noticeably, the latent image constructed from the combination of the bandpass components of the training images (in the third column of the first row) is free of blur, especially near edges. Most of the mid-frequency to high-frequency components have been recovered. The absolute difference image (in the fourth column of the first row) simply shows low to mid-frequency details, which could be recovered by a final non-blind deconvolution step.

1.1 Related Work

Utilising edge information as a form of image sparsity, single image deblurring methods rely on the implicit or explicit extraction of this information for kernel computation. Several approaches [5], [6] enhance the detection and selection of strong edges via various techniques such as bilateral filtering, shock filtering, and gradient magnitude thresholding. Joshi *et al*. [25] predicted the step edges underlying the blurred ones for the estimation of spatially varying sub-pixel point-spread functions (PSF). Cho *et al*. [26] also detected step edges in blurry images and used this information to compute the Radon transform of the blur kernel. A concern about these approaches is that wrong edges can be mistakenly selected based on only local information, due to the possible presence of multiple copies of the same edge induced by a large kernel width. Moreover, object classes with relatively limited texture details such as face and text do not usually benefit from methods using local edge information.

There has been a few notable examples of deconvolution methods that utilise image edge information for the estimation of the blur kernel. The fast deconvolution algorithm is based on the hyper Laplacian prior [16] and a decomposition of the inverse kernels in the frequency domain into a series of 1D kernels of Xu et al. [14]. Whyte et al. [15] proposed a model to effectively reduce the ringing artifacts by simply discarding the saturated pixels, using only the non-saturated ones to estimate the blur kernel. Specific to text image deblurring, Pan et al. [13] proposed an effective L_0 regularisation method. This method works well with smooth surfaces but is less effective for non-uniform and highly textured areas/background. Our method is distinguishable from all the above, as the latter only utilise generic edge priors into account, without considering class-specific spatial priors. Furthermore, these methods do not rely on external training images in additional to the input image.

Another approach is to adopt a probabilistic viewpoint by modelling the posterior probability of the latent image and the kernel. With this view, Fergus *et al*. [3] modelled the distribution of the latent image gradients as a mixture of zero-mean Gaussians and the distribution of the kernel elements as a mixture of exponential distributions. On the other hand, Shan *et al*. [4] opted for a Maximum a Posteriori (MAP) formulation under the assumption of a Gaussian noise model. This formulation eventually leads to an objective function with norm constraints on the latent image to model the gradient sparsity and the smooth local prior of the image, and ℓ_1-norm regulariser on the blur kernels. Improving upon this approach, Levin *et al*. [8] aimed at maximising the posterior distribution with the best kernel while marginalising over all possible latent images. To reduce computational complexity, they tackle an approximate MAP problem with an EM-like iteration strategy.

As an alternative, several methods [27], [28], [29] have employed selective information from image patches and their priors, rather than the whole image. Zoran and Weiss [27] proposed patch-based image prior using GMM model, which is overly expressive *i.e*. models a wide range of phenomena including motion blur and defocus blur, and will eventually accommodate blur, causing imprecise convergence of the solution pair. Building on the idea of [27], Sun *et al*. [28] modelled the patch-based image prior using atomic elements, namely, edges, corners, T-junctions, etc. learned from natural image datasets and synthetic structures. Michaeli and Irani [29] exploited the multi-scale patch recurrence property as a natural image prior to recover the blur kernel.

In addition, several works have started to pursue the "learning to deblur" approach with a large amount of training data [30], [31]. Schuler *et al*. [30] proposed to learn a stack of multiple neural sub-networks, each consisting of three modules to estimate the blur kernel. An interesting finding from Schuler *et al*. [30] is that the deblurring quality for a specific class in the ImageNet dataset improves when the proposed neural network is trained on images from the same class, rather than the entire dataset (Section 6.1). In their work, neural networks were employed as a black-box tool for kernel estimation, and the rationale behind this observation was not much studied and understood. Here, we offer an in-depth study on class-specificity for image deblurring from a classical image processing perspective. We also compare a class-agnostic and a class-specific variant of our algorithm and concur with a similar observation as Schuler *et al*. [30] (see Section 4.3 for details). In a related development, Chakrabarti [31] proposed to learn a neural network which predicts the complex Fourier coefficients of a motion blur kernel as an input to a subsequent non-blind convolution step. Qualitative results shows that it does not cope well with dense texture.

Recently, class-specific information has been employed up to some extent for image deblurring. Joshi *et al*. [32] proposed a method for personal photo enhancement, including deblurring, given examples in a photo collection. This approach requires manual annotation of face regions for the matting and segmentation of faces from input images. HaCohen *et al*. [33] tackled this problem, requiring a dense correspondence between a sharp reference image and its corresponding blurred image. This method produces decent results for complex kernels, but has limited applications due to the strict requirements of the similar content between the

reference and the blurred image. In another work, Sun *et al*. [34] investigated context-specific priors to transfer mid and high frequency details from example scenes for non-blind deconvolution.

Lately, Pan *et al*. [12] introduced a face image deblurring method by selecting the best exemplar from a training set with the closest structural similarity to the blurred image. It requires manual annotations of salient features such as the eyes, mouth and lower contour of the face for each training image. Information at these locations then serves as guidance for deblurring face images. In contrast, our method neither requires manual annotations of the training data nor relies on the similarity between the blurred image and the training image contents. Specific to face image deblurring, our method does not require the person or the background in the blurred image to be present in the training dataset.

2 Problem Formulation

In this section, we introduce our class-specific image prior and incorporate it into an optimisation framework. Given a set of N sharp training images $\{\mathbf{z}_i | i = 1, \ldots, N\}$ and an arbitrary blurred image $\mathbf{y}$ that belongs to the same class, we aim to recover the latent image $\mathbf{x}$ and the kernel $\mathbf{k}$.

2.1 Image prior

We formulate the deblurring problem in the Fourier domain by first obtaining a bank of Butterworth bandpass filters, each of which has a constant magnitude in a certain (2D) frequency band and zero elsewhere. The visual representation of bandpass filters in the frequency domain are concentric circular bands (centered at the origin) with unit values. To filter an image with a Butterworth bandpass filter, we first clip its frequency components in the Fourier domain to the range defined by the bandpass filter (corresponding to the filter's non-zero frequencies). Subsequently, the remaining frequency components are transformed to the spatial domain via an inverse Fourier transform.

The main underlying hypothesis of our novel class-specific prior with the hypothesis is that the frequency components in each band span a sparse linear subspace in the Fourier domain. To this end, we let $\mathcal{F}_{\mathbf{x}}(\omega)$ denote the Fourier coefficient of the 2D image $\mathbf{x}$ at the spatial frequency ω. aving divided the frequency spectrum into a set of M frequency bands, we formulate the linear subspace constraint for band b_j as

$$\mathcal{F}_{\mathbf{x}}(\omega) = \sum_{i=1}^{N} w_{i,j} \mathcal{F}_{\mathbf{z}_i}(\omega), \forall \omega \in b_j, j = 1, \ldots, M, \quad (1)$$

where $w_{i,j}$ is a weight associated with the training image $\mathbf{z}_i$ and the band b_j in the representation of the latent image $\mathbf{x}$. This coefficient correlates to the similarity between the frequency components of the training and the latent image in the band b_j.

In addition, we enforce sparsity on the weight vector $\mathbf{w}_j \triangleq [w_{1,j}, \ldots, w_{N,j}]$ for each band b_j. The sparsity constraint emphasizes the major contributions from a few training images to the representation of the latent image $\mathbf{x}$ for each separate band. Here, we express this constraint as a minimisation of the L_1-norm $\|\mathbf{w}_j\|_1$ due to its well-known robustness. Combining the linear subspace constraint and the sparsity constraint on $\mathbf{w}_j$ over all the frequency bands, we define the prior function $P(\mathbf{x}, \mathbf{w})$ as

$$\begin{aligned} P(\mathbf{x}, \mathbf{w}) \triangleq \gamma \sum_{j=1}^{M} \sum_{\omega \in b_j} |\mathcal{F}_{\mathbf{x}}(\omega) - \sum_{i=1}^{N} w_{i,j} \mathcal{F}_{\mathbf{z}_i}(\omega)|^2 \\ + \tau \sum_{j=1}^{M} \|\mathbf{w}_j\|_1, \end{aligned} \quad (2)$$

where γ and τ are the balance factors of the reconstruction error and the sparsity term, respectively, and $|\cdot|$ denotes the modulus of a complex number.

For each band b_j, we define a corresponding band-pass filter $\mathbf{f}_j$, such as a Butterworth filter [35], whose Fourier transform is a non-zero constant c within b_j and zero elsewhere. With this filter, let us consider the 2D function $\mathbf{g} = \mathbf{x} \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j} (\mathbf{z}_i \otimes \mathbf{f}_j)$, where $\otimes$ denotes the convolution operator. The Fourier transform of this function is

$$\mathcal{F}_{\mathbf{g}}(\omega) = \begin{cases} c \left(\mathcal{F}_{\mathbf{x}}(\omega) - \sum_{i=1}^{N} w_{i,j} \mathcal{F}_{\mathbf{z}_i}(\omega) \right) & \forall \omega \in b_j, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Applying the Parseval's theorem to the function $\mathbf{g}$, we have $\int \mathbf{g}(u)^2 du = \int |\mathcal{F}_{\mathbf{g}}(\omega)|^2 d\omega$. Noting that $\int |\mathcal{F}_{\mathbf{g}}(\omega)|^2 d\omega$ is a multiple of the reconstruction error in Equation 2, we rewrite it as follows

$$P(\mathbf{x}, \mathbf{w}) = \beta \sum_{j=1}^{M} \|\mathbf{x} \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j} (\mathbf{z}_i \otimes \mathbf{f}_j)\|_2^2 + \tau \sum_{j=1}^{M} \|\mathbf{w}_j\|_1, \quad (4)$$

where we use the variable substitution $\beta \triangleq \frac{\gamma}{c^2}$.

2.2 Objective function

In image deblurring, the aim is to minimise the data fidelity term associated with the blur model $\mathbf{y} = \mathbf{x} \otimes \mathbf{k} + \mathbf{n}$, where $\mathbf{n}$ is the image noise. In addition, several deblurring approaches have utilised image gradients to enforce an *a priori* image gradient distribution, *i.e.* natural image statistics [3] and to better capture the spatial randomness of noise [4]. Following these previous approaches, we also exploit the derivative form of the blur model and aim to minimise the error $\|\nabla_d \mathbf{x} \otimes \mathbf{k} - \nabla_d \mathbf{y}\|_2^2$, where ∇_d denotes the gradient operator in the direction $d \in \{x, y\}$.

In addition, we employ a regulariser on the blur kernel using the conventional L_2-norm $\|\mathbf{k}\|_2^2$ as in previous works [5], [36]. Combining all the above components, we arrive at a minimisation of the objective function

$$\begin{aligned} J(\mathbf{x}, \mathbf{w}, \mathbf{k}) = \|\mathbf{x} \otimes \mathbf{k} - \mathbf{y}\|_2^2 + P(\mathbf{x}, \mathbf{w}) \\ + \sum_{d \in \{x, y\}} \|\nabla_d \mathbf{x} \otimes \mathbf{k} - \nabla_d \mathbf{y}\|_2^2 + \alpha \|\mathbf{k}\|_2^2, \end{aligned} \quad (5)$$

where α is the balancing factor for the kernel regulariser.

3 Deblurring Framework

Given $\mathbf{y}$, $\{\mathbf{f}_b | b = 1, \ldots, M\}$ and $\{\mathbf{z}_i | i = 1, \ldots, N\}$, we aim to minimise the objective function in Equation 5 with respect

to the unknowns $\mathbf{x}, \mathbf{w}$ and $\mathbf{k}$. Since a simultaneous minimisation with respect to all the variables is computationally expensive, we adopt an alternating minimisation scheme. In each iteration of this scheme, we solve a sub-problem with respect to one of the variables $\mathbf{x}, \mathbf{w}$ and $\mathbf{k}$, while fixing the others. The following subsections describe the solution to each sub-problem.

3.1 Estimating $\mathbf{w}$ given $\mathbf{x}$ and $\mathbf{k}$

Assuming that $\mathbf{x}$, and $\mathbf{k}$ have been obtained in an earlier iteration, we aim to minimise the objective function $J(\mathbf{x}, \mathbf{w}, \mathbf{k})$ with respect to the weights $w_{i,j}$. Here, we note that $P(\mathbf{x}, \mathbf{w})$ in Equation 5 can be decomposed into separate bands. Therefore, we can break down the above problem into the minimisation of the following function (with respect to $\mathbf{w}_j$) for each band b_j

$$J_{\mathbf{w}_j} = \|\mathbf{x} \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j}(\mathbf{z}_i \otimes \mathbf{f}_j)\|_2^2 + \frac{\tau}{\beta}\|\mathbf{w}_j\|_1, \quad (6)$$

We vectorise the images involved in the above Equation using the following shorthand notation $\tilde{\mathbf{x}}_j = vec(\mathbf{x} \otimes \mathbf{f}_j)$ and $\tilde{\mathbf{z}}_{i,j} = vec(\mathbf{z}_i \otimes \mathbf{f}_j)$. The minimisation of the above cost function can be regarded as an ℓ_1-regularized least-squares problem and can be solved by standard techniques such as the one reported in [37]. The above problem is usually well-formed when the length of $\tilde{\mathbf{x}}_j$ and $\tilde{\mathbf{z}}_{i,j}$ exceeds that of $\mathbf{w}_j$, *i.e.* the number of image pixels is more than the number of training images N.

3.2 Latent image estimation

With the current update of the contributions $\mathbf{w}_j$, $j = 1, \ldots, M$, from the training images to each band, and the kernel $\mathbf{k}$, we now estimate the latent image so as to minimise Equation 5. Similar to the approach above, we only consider the sum of the terms dependent on $\mathbf{x}$

$$\begin{aligned} J_{\mathbf{x}} = \|\mathbf{x} \otimes \mathbf{k} - \mathbf{y}\|_2^2 + \sum_{d \in \{x,y\}} \|\nabla_d \mathbf{x} \otimes \mathbf{k} - \nabla_d \mathbf{y}\|_2^2 \\ + \beta \sum_{j=1}^{M} \|\mathbf{x} \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j}(\mathbf{z}_i \otimes \mathbf{f}_j)\|_2^2. \end{aligned} \quad (7)$$

To this end, we apply the Parseval's theorem to the terms on the right-hand side of Equation 7. This theorem states that the total energy of a function over the spatial domain is equal to that of its Fourier transform over the frequency domain. We also note that the image derivative $\nabla_d \mathbf{x}$ can be expressed as a convolution as $\nabla_d \otimes \mathbf{x}$, where ∇_d is a convolution kernel representing the corresponding derivative operation. With these ingredients, we rewrite Equation 7 in the Fourier transforms of its terms as

$$\begin{aligned} J_{\mathbf{x}} = \int |\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{k}}(\omega) - \mathcal{F}_{\mathbf{y}}(\omega)|^2 d\omega \\ + \sum_{d \in \{x,y\}} \int |\mathcal{F}_{\nabla_d}(\omega)\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{k}}(\omega) - \mathcal{F}_{\nabla_d}(\omega)\mathcal{F}_{\mathbf{y}}(\omega)|^2 d\omega \\ + \beta \sum_{j=1}^{M} \int |\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{f}_j}(\omega) - \sum_{i=1}^{N} w_{i,j}\mathcal{F}_{\mathbf{z}_i}(\omega)\mathcal{F}_{\mathbf{f}_j}(\omega)|^2 d\omega, \end{aligned} \quad (8)$$

where ω represents a spatial frequency, $|\cdot|$ signifies the modulus of a complex number and all the integrals are taken over the entire frequency spectrum.

The Parseval's theorem yields a convenient expression with respect to the Fourier transform of the latent image. Since the function in Equation 8 is a convex function of $\mathcal{F}_{\mathbf{x}}(\omega)$ in the Fourier domain, a local optimisation method can be applied to obtain its global minimum. Also, we note that $\frac{\partial(|z|^2)}{\partial z} = \overline{z}$, where $\overline{z}$ is the conjugate of the complex number z. For brevity, we omit the frequency ω from the following expressions. By the chain rule, we derive the partial derivative with respect to the Fourier transform $\mathcal{F}_{\mathbf{x}}$ as follows

$$\begin{aligned} \frac{\partial J_{\mathbf{x}}}{\partial \mathcal{F}_{\mathbf{x}}} = 2(\mathcal{F}_{\mathbf{k}}\left(\overline{\mathcal{F}_{\mathbf{x}}\mathcal{F}_{\mathbf{k}}} - \overline{\mathcal{F}_{\mathbf{y}}}\right) \\ + \sum_{d \in \{x,y\}} \mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{k}}\left(\overline{\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{x}}\mathcal{F}_{\mathbf{k}}} - \overline{\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{y}}}\right) \\ + \beta \sum_{j=1}^{M} \mathcal{F}_{\mathbf{f}_j}(\overline{\mathcal{F}_{\mathbf{x}}\mathcal{F}_{\mathbf{f}_j}} - \sum_{i=1}^{N} w_{i,j}\overline{\mathcal{F}_{\mathbf{z}_i}\mathcal{F}_{\mathbf{f}_j}})), \end{aligned} \quad (9)$$

where the multiplications on the right-hand side are performed frequency-wise in the Fourier domain.

We rewrite the complex conjugate of $\frac{\partial J_{\mathbf{x}}}{\partial \mathcal{F}_{\mathbf{x}}}$ as follows

$$\begin{aligned} \overline{\left(\frac{\partial J_{\mathbf{x}}}{\partial \mathcal{F}_{\mathbf{x}}}\right)} = 2(|\mathcal{F}_{\mathbf{k}}|^2\mathcal{F}_{\mathbf{x}} - \overline{\mathcal{F}_{\mathbf{k}}}\mathcal{F}_{\mathbf{y}} \\ + \sum_{d \in \{x,y\}} \left(|\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{k}}|^2\mathcal{F}_{\mathbf{x}} - |\mathcal{F}_{\nabla_d}|^2\overline{\mathcal{F}_{\mathbf{k}}}\mathcal{F}_{\mathbf{y}}\right) \\ + \beta \sum_{j=1}^{M} |\mathcal{F}_{\mathbf{f}_j}|^2(\mathcal{F}_{\mathbf{x}} - \sum_{i=1}^{N} w_{i,j}\mathcal{F}_{\mathbf{z}_i})). \end{aligned} \quad (10)$$

By equating the complex conjugate of $\frac{\partial J_{\mathbf{x}}}{\partial \mathcal{F}_{\mathbf{x}}}$ to zero, we obtain the following closed-form solution for the latent image $\mathbf{x}$

$$\begin{aligned} \mathcal{F}_{\mathbf{x}} = (\overline{\mathcal{F}_{\mathbf{k}}}\mathcal{F}_{\mathbf{y}} + \sum_{d} |\mathcal{F}_{\nabla_d}|^2\overline{\mathcal{F}_{\mathbf{k}}}\mathcal{F}_{\mathbf{y}} + \beta \sum_{j=1}^{M} |\mathcal{F}_{\mathbf{f}_j}|^2 \sum_{i=1}^{N} w_{i,j}\mathcal{F}_{\mathbf{z}_i})./ \\ (|\mathcal{F}_{\mathbf{k}}|^2 + \sum_{d} |\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{k}}|^2 + \beta \sum_{j=1}^{M} |\mathcal{F}_{\mathbf{f}_j}|^2), \end{aligned} \quad (11)$$

where the $./$ notation stands for a frequency-wise division in the Fourier domain. The latent image can be obtained by an inverse Fourier transform of the solution to $\mathcal{F}_{\mathbf{x}}$.

3.3 Blur kernel estimation

Once the latent image $\mathbf{x}$ is computed, the next step is to estimate the blur kernel $\mathbf{k}$. Based on Equation 5, this optimisation step involves the following terms

$$J_{\mathbf{k}} = \|\mathbf{x} \otimes \mathbf{k} - \mathbf{y}\|_2^2 + \sum_{d} \|\nabla_d \mathbf{x} \otimes \mathbf{k} - \nabla_d \mathbf{y}\|_2^2 + \alpha\|\mathbf{k}\|_2^2. \quad (12)$$

Again, we leverage the Parseval's theorem and express the above function in the Fourier domain as

$$\begin{aligned} J_{\mathbf{k}} = \int |\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{k}}(\omega) - \mathcal{F}_{\mathbf{y}}(\omega)|^2 d\omega + \alpha \int |\mathcal{F}_{\mathbf{k}}(\omega)|^2 d\omega \\ + \sum_{d \in \{x,y\}} \int |\mathcal{F}_{\nabla_d}(\omega)\mathcal{F}_{\mathbf{x}}(\omega)\mathcal{F}_{\mathbf{k}}(\omega) - \mathcal{F}_{\nabla_d}(\omega)\mathcal{F}_{\mathbf{y}}(\omega)|^2 d\omega. \end{aligned} \quad (13)$$

Algorithm 1 Deblurring with the class-specific prior.

Require:
$\mathbf{y}$: the given blurred image.
$\mathbf{z}_i, i = 1, \ldots, N$: the class-specific training images.
$\mathbf{f}_j, j = 1, \ldots, M$: a set of band-pass filters covering the frequency spectrum.
α, β: the weights of the terms in Equation 5.
ρ: the attenuation factor of the class-specific prior.

1: $\mathcal{F}_{\mathbf{x}} \leftarrow \mathcal{F}_{\mathbf{y}}$.
2: $\mathbf{k} \leftarrow \delta$ (Dirac delta kernel).
3: **while** $size(\mathbf{k}) \leq max\_size$ **do**
4: $\quad\beta \leftarrow \beta_0$.
5: $\quad$**repeat**
6: $\quad\quad$Minimise $J_{\mathbf{w}_j}$ in 6 w.r.t. $\mathbf{w}_j, \forall j$, with solver in [37].
7: $\quad\quad$Update $\mathbf{x}$ according to Equation 11.
8: $\quad\quad\beta \leftarrow \rho\beta$.
9: $\quad\quad$Update $\mathbf{k}$ according to Equation 16.
10: $\quad$**until** the maximum number of iterations is reached or $\mathbf{x}$ and $\mathbf{k}$ change by an amount below a relative tolerance threshold.
11: $\quad\mathbf{k} \leftarrow upsample(\mathbf{k})$ (Initialisation of kernel for the following scale) .
12: **end while**
13: **return** Latent image $\mathbf{x}$ and blur kernel $\mathbf{k}$.

where, as before, the integrals are taken over the entire frequency spectrum.

Since $J_{\mathbf{k}}$ is a quadratic function of $\mathcal{F}_{\mathbf{k}}(\omega)$, we can obtain the minimiser by setting $\frac{\partial J_{\mathbf{k}}}{\partial \mathcal{F}_{\mathbf{k}}}$ to zero. This derivative can be expanded as

$$\begin{aligned}\frac{\partial J_{\mathbf{k}}}{\partial \mathcal{F}_{\mathbf{k}}} &= \mathcal{F}_{\mathbf{x}}\left(\overline{\mathcal{F}_{\mathbf{x}}\mathcal{F}_{\mathbf{k}}} - \overline{\mathcal{F}_{\mathbf{y}}}\right) \\ &+ \sum_{d\in\{x,y\}} \mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{x}}\left(\overline{\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{x}}\mathcal{F}_{\mathbf{k}}} - \overline{\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{y}}}\right) \\ &+ \alpha\overline{\mathcal{F}_{\mathbf{k}}}.\end{aligned} \tag{14}$$

Setting the complex conjugate of the above equation to zero, we obtain the following closed-form solution for $\mathcal{F}_{\mathbf{k}}$ as

$$\begin{aligned}\mathcal{F}_{\mathbf{k}} = (\overline{\mathcal{F}_{\mathbf{x}}}\mathcal{F}_{\mathbf{y}} + \sum_d |\mathcal{F}_{\nabla_d}|^2\overline{\mathcal{F}_{\mathbf{x}}}\mathcal{F}_{\mathbf{y}})./ \\ (|\mathcal{F}_{\mathbf{x}}|^2 + \sum_d |\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{x}}|^2 + \alpha).\end{aligned} \tag{15}$$

For sparse kernels such as motion kernels, which contain mainly high-frequency components, we choose to follow the practice in [8] and include only the image gradient term in the above Equation as its frequency components are more relevant to the kernel spectrum. In that case, the closed-form solution for $\mathbf{k}$ is simplified as

$$\mathbf{k} = \mathcal{F}^{-1}\left(\frac{\sum_{d\in\{x,y\}} |\mathcal{F}_{\nabla_d}|^2\overline{\mathcal{F}_{\mathbf{x}}}\mathcal{F}_{\mathbf{y}}}{\sum_{d\in\{x,y\}} |\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{x}}|^2 + \alpha}\right), \tag{16}$$

where $\mathcal{F}^{-1}(\cdot)$ denotes the inverse Fourier transform.

3.4 Implementation

Our optimisation approach is summarised in Algorithm 1. The algorithm takes, as input, a given blurred image $\mathbf{y}$, a training set of sharp images $\mathbf{z}_i, i = 1, \ldots, N$ and a bank of band-pass filters $\mathbf{f}_j, j = 1, \ldots, M$, which together cover the entire frequency spectrum. With this input, it aims to compute the latent image $\mathbf{x}$ and the blur kernel $\mathbf{k}$.

The algorithm commences with the initialisation of the latent image and the kernel to the given blurred image and the Dirac delta function, respectively. Subsequently, it proceeds in an iterative manner. In each iteration, we minimise the objective function with respect to $\mathbf{w}$, $\mathbf{x}$ and $\mathbf{k}$ in alternating steps, as shown in lines 6, 7 and 9. The update steps for $\mathbf{x}$ and $\mathbf{k}$ are undertaken by fast forward and inverse Fourier transforms according to Equations 11 and 15. After every iteration, $\mathbf{k}$ is centred and normalised so that the sum of its elements is unity. Meanwhile, to solve for $\mathbf{w}$, we minimise the cost function in Equation 6 using the L_1 least-squares solver in [37]. The algorithm terminates when the values of $\mathbf{x}$ and $\mathbf{k}$ do not change by pre-determined tolerance thresholds over two successive iterations.

To improve the stability of the estimates, we progressively increase the kernel size in a coarse-to-fine scheme. Within a fixed kernel scale, we iterate between the estimation steps with respect to $\mathbf{w}$, $\mathbf{x}$ and $\mathbf{k}$ until convergence, before expanding the kernel size to the next scale. The initial kernel size is 3×3 and the expansion factor between two successive scales which we found empirically is $\sqrt{1.6}$.

To initialise the kernel in the next scale, we upsample the kernel estimated in the previous iteration using bicubic interpolation. Since iterations at a finer kernel resolution usually inherit good estimates from those at coarser resolutions before further fine-tuning, we enforce a small number of iterations typically between fifteen and twenty for kernel resolutions of 11×11 and above.

In addition, while we preset the weight α of the kernel regulariser, we adjust the weight β of the class-specific prior incrementally over iterations. The reason for this adjustment is that we initially prefer to obtain as much class information as needed to constrain the space of the latent image. On the other hand, as the iterations proceed, we deliberately decrease the influence of this term so that the estimation is increasingly driven by the data fidelity term. In other words, the resulting latent image and kernel will increasingly gather instance-specific details from the given blurred image, rather than the class prior. This step is taken after the update of $\mathbf{x}$ in every iteration, as shown in line 8.

3.5 Extension to colour images

While Algorithm 1 accepts grayscale images as input, it can be extended to deblur colour images in a straightforward manner. This extension assumes that all the colour channels have been distorted by the same spatially uniform blur kernel. In this case, the variables $\mathbf{w}$ and $\mathbf{x}$ are defined per colour channel $c \in \{R, G, B\}$ as $\mathbf{w}_c$ and $\mathbf{x}_c$, while the kernel $\mathbf{k}$ is the same all the channels. The objective function is then

modified as

$$J(\mathbf{x}_c, \mathbf{w}_c, \mathbf{k}) = \sum_c \Big[\beta \sum_{j=1}^{M} \|\mathbf{x}_c \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j,c}(\mathbf{z}_{i,c} \otimes \mathbf{f}_j)\|_2^2 + \|\mathbf{x}_c \otimes \mathbf{k} - \mathbf{y}_c\|_2^2 + \sum_{d\in\{x,y\}} \|\nabla_d \mathbf{x}_c \otimes \mathbf{k} - \nabla_d \mathbf{y}_c\|_2^2 + \tau \sum_{j=1}^{M} \|\mathbf{w}_{j,c}\|_1\Big] + \alpha\|\mathbf{k}\|_2^2. \tag{17}$$

The solution for $\mathbf{w}_c$ can be derived by minimising the following function per channel

$$J(\mathbf{x}_c, \mathbf{w}_c) = \sum_c \Big[\beta \sum_{j=1}^{M} \|\mathbf{x}_c \otimes \mathbf{f}_j - \sum_{i=1}^{N} w_{i,j,c}(\mathbf{z}_{i,c} \otimes \mathbf{f}_j)\|_2^2 + \tau \sum_{j=1}^{M} \|\mathbf{w}_{j,c}\|_1\Big]. \tag{18}$$

Similarly, the update step for $\mathbf{x}_c$ can be performed for each channel using a similar formula to Equation 11 as

$$\mathcal{F}_{\mathbf{x_c}} = (\overline{\mathcal{F}_{\mathbf{k}}}\mathcal{F}_{\mathbf{y_c}} + \sum_d |\mathcal{F}_{\nabla_d}|^2 \overline{\mathcal{F}_{\mathbf{k}}}\mathcal{F}_{\mathbf{y_c}} + \beta \sum_{j=1}^{M} |\mathcal{F}_{\mathbf{f}_j}|^2 \sum_{i=1}^{N} w_{i,j,c}\mathcal{F}_{\mathbf{z}_{i,c}})./ (|\mathcal{F}_{\mathbf{k}}|^2 + \sum_d |\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{k}}|^2 + \beta \sum_{j=1}^{M} |\mathcal{F}_{\mathbf{f}_j}|^2), \tag{19}$$

Meanwhile, the kernel $\mathbf{k}$ is computed by taking a summation of the both the numerator and denominator over the color channels as

$$\mathcal{F}_{\mathbf{k}} = \sum_c \Big[(\overline{\mathcal{F}_{\mathbf{x_c}}}\mathcal{F}_{\mathbf{y_c}} + \sum_d |\mathcal{F}_{\nabla_d}|^2 \overline{\mathcal{F}_{\mathbf{x_c}}}\mathcal{F}_{\mathbf{y_c}})\Big]./ \sum_c \Big[(|\mathcal{F}_{\mathbf{x_c}}|^2 + \sum_d |\mathcal{F}_{\nabla_d}\mathcal{F}_{\mathbf{x_c}}|^2 + \alpha)\Big]. \tag{20}$$

4 Results and Discussion

In this section, we aim to demonstrate the advantage of incorporating the proposed class-specific prior for the blind deconvolution task. For this purpose, we will provide a detailed performance comparison between our method and a number of state-of-the-art alternatives over several datasets. We commence our analysis on the contribution of the components in our framework to the overall performance. Here, we particularly pay attention to the role of the class-specific prior in our framework. Next, we compare our method to a number of well-known deblurring methods that are not equipped with image class priors, in terms of both quantitative and qualitative results. For completeness, we will illustrate the superiority of our method to existing algorithms that exploit class-specific information or class exemplars, in terms of the visual quality of the results.

4.1 Datasets and experimental settings

We performed the experimental validation on six datasets including the CMU PIE face dataset [38], the car dataset in [39], the cat dataset in [20], the ETHZ dataset of shape classes [40], the Yale-B face database [41] and the INRIA person dataset [42]. For each dataset, we randomly selected half of the images as training data and between 10 and 15 sharp images from the remaining half as ground-truth test images for deblurring. To generate blurred images from the test images, we employed the eight complex ground-truth blur kernels computed by Levin *et al*. [43] from emulated camera shakes. With this input, we compared our proposed algorithm against the state-of-the-art deblurring algorithms with and without using class exemplars under same conditions. The comparison, as will be shown in the following part of the paper, is based on both the visual quality of the recovered image and blur kernel, as well as the numerical accuracy of these two. In this paper, we report the numerical error of the full image and kernel in terms of the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR).

We have implemented our algorithm in MATLAB on an Intel Core^{TM} i7 machine with $16GB$ of memory. In all of our experiments, we set the parameters $M = 90$, $\alpha = 10$ and $\tau = 0.01$, initialise β to $\beta_0 = 50$, and decrease the value of β by a factor of $\rho = 1.3$ in every iteration until it reaches the minimal value of 0.01. In other words, the contribution from the training images is reduced as the algorithm proceeds and becomes negligible in the end. In the last few iterations, the image and kernel estimation is mainly driven by the information from the blurred image.

For a fair comparison[1], we strive to apply the same non-blind deblurring method, *i.e*. Levin *et al*. [18], in the final step of our and prior methods whose source code is available and can be modified. Pan *et al*. [12] and Levin *et al*. [8] already use Levin *et al*.'s method [18] in their original implementation. We also change the default non-blind deconvolution step in Fergus *et al*. [3] and Shan *et al*. [4] to Levin *et al*. [18]. However, the remaining algorithms in our comparison opt for other non-blind deconvolution methods, which cannot be modified in a straightforward manner. For example, Sun *et al*. [28] use Zoran and Weiss [27] as a final non-blind deblurring step. Meanwhile, Zhong *et al*. [44], Cho *et al*. [5] and Xu *et al*. [9] devise their own non-blind deconvolution methods and only provide the binary executables of their algorithms. In addition, Krishnan *et al*. [7] utilised their previous work in [16].

4.2 Ablative studies

We perform an extensive ablative studies to validate the efficacy of our approach in various aspects.

4.2.1 Effectiveness of the proposed prior

Using the data and setting described above, we demonstrate the effectiveness of the proposed class-specific prior within our deblurring framework. In Figure 3, we compare the visual quality of the recovered latent image and the kernel

1. We also found out that Zoran *et al*. [27] performance on class specific algorithms is lower than levin *et al*. [18]

Ground-truth Blurred image Without prior With prior

Fig. 3: Latent images and kernels recovered by our method without (third column) and with the proposed prior (fourth column).

	SSIM		PSNR	
Prior	Without	With	Without	With
Images	0.365	**0.754**	16.87	**25.78**
Kernel	0.707	**0.855**	39.42	**42.66**

TABLE 1: A comparison of the accuracy achieved by our deblurring framework on all the mentioned datasets with and without the proposed prior.

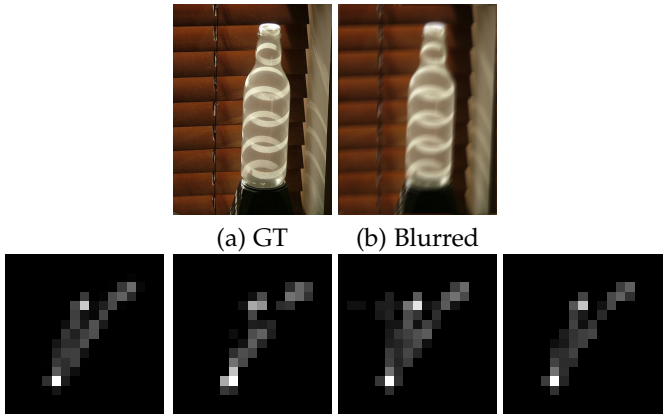
obtained without the prior (in the third column) and with the prior (in the last column). Evidently, the image recovered with the prior does not contain visible artifacts, whereas that obtained without the prior shows severe ringing and multiple false edges. In addition, when inspecting the estimated kernel (better viewed when zoomed in the electronic copy), we observed a noisy one close to the delta kernel (the initial kernel) when we do not include the image prior in our method. This suggests that the method may have not converged without this prior. On the other hand, the kernel is almost identical to the ground-truth with the prior included.

Further, we have quantified the accuracy of the recovered latent image and blurred kernel with and without the use of the class-specific prior. In Table 1, the accuracy is measured in SSIM and PSNR, indicating the similarity between the estimated quantities and the corresponding ground-truth. These results demonstrate that the accuracy of the recovered image and kernel improves significantly (by several orders of magnitude) with the proposed prior. This is consistent with the visual observations above, suggesting that the proposed prior plays an important role in correctly guiding the estimates to the ground-truth.

We have also performed an experiment where the prior only covers the mid and high-frequency bands, and the low-frequency components of the latent image are estimated directly from the input image. Without the low frequency in the prior, the average PSNR for the CMU dataset declines to 25.67 dB, as compared to 30.75 dB (in Table 4, when all the frequency bands are incorporated in the image prior). Hence, incorporating low-frequency bands is beneficial, rather than harmful, to the deblurring task.

	SSIM			PSNR		
	Intensity only	Gradient only	Both	Intensity only	Gradient only	Both
Images	0.678	0.529	**0.754**	23.28	21.16	**25.78**
Kernel	0.819	0.745	**0.855**	41.27	40.01	**42.66**

TABLE 2: Influence of intensity and gradient fidelity terms on the deblurring results.



(a) GT (b) Blurred

(c) True kernel (d) Intensity (40.2/0.837) (e) Gradient (50.6/0.975) (f) Both **(52.7/0.985)**

Fig. 4: Influence of the data fidelity term in the objective function on the kernel estimate. A pair of PSNR/SSIM error metrics is shown for each kernel estimate in the sub-figures (d)–(f). (a) ground-truth image, (b) blurred image, (c) ground-truth kernel, (d) estimated kernel with the intensity term only, (e) estimated kernel with the gradient term only, (f) estimated kernel with both terms.

4.2.2 *Influence of data fidelity terms*

We experimented with different options of the data terms for the estimation of the latent image (while only employing the gradient information for estimating the kernel). These includes the intensity fidelity term, *i.e.* $\|\mathbf{x}\otimes\mathbf{k}-\mathbf{y}\|_2^2$, and the gradient fidelity term, *i.e.* $\|\nabla_d\mathbf{x}\otimes\mathbf{k}-\nabla_d\mathbf{y}\|_2^2$ in Equation 5, or both.

Table 2 shows the accuracy of the deblurred image and kernel estimate across all the datasets under study, in terms of SSIM and PSNR. The highest accuracy is achieved when both data fidelity terms, *i.e.* intensity and gradients, are employed jointly with with the class specific prior, while using each individual fidelity term yields a lower accuracy. For this reason, we employ both the intensity and gradient fidelity terms in our framework.

Figure 4 illustrates example kernels estimated with the above three options. Specifically, the kernels in Figures 4(d)-(f) are recovered using only either the intensity or the gradient fidelity term, and then with both terms, respectively. Among these options, the former two yield kernel estimates with clear structural deviations from the ground-truth shown in Figure 4c. On the other hand, the kernel yielded using both fidelity terms (Figure 4f) is closer to the ground-truth. This implies a more accurate estimation of the intermediate latent image.

4.2.3 Influence of the dataset size

We also examine the variation of deblurring performance with respect to the number of training images. Table 3 shows that the image and kernel estimation accuracy for the CMU PIE dataset improves consistently with the increasing number of training images. Even with only 50 training samples, our method can achieve an average image accuracy of 25.87 dB, outperforming all the other methods on this dataset (More detailed results are given in Table 8).

Training size	50	125	250	500	1000	2000
Image	25.87	26.97	27.92	28.58	30.42	30.75
Kernel	41.15	42.11	42.80	42.99	44.01	44.13

TABLE 3: Deblurring performance (in PSNR) on the CMU PIE dataset for different numbers of training images.

4.2.4 Choice of the training class

We ask the question whether the choice of training class significantly alters the deblurring accuracy. To this end, we experiment with various pairs of training and test object categories. Table 4 shows the accuracy of deblurred images (in PSNR) for various training (along the columns) and test (along the rows) categories. In most cases, the best accuracy is achieved along the diagonal, *i.e.* when the input test image belongs to the same object category as the training dataset. The only exception is that our algorithm achieves the best deblurring accuracy for the Yale-B dataset when being trained on the Cat dataset. This result matches the observation that a number of features of a cat face such as eyes, lips and contours resemble those of a human face. As a consequence, employing cat faces as training data are potentially as beneficial for the deblurring of human faces. Otherwise, the PSNR degrades significantly when the training class differs from the test class. These results demonstrate the impact of choosing the correct training class on the deblurring accuracy.

4.2.5 Schedule of the prior weight β

We have assessed the performance of our algorithm with a fixed weight β over all the iterations. In Table 5, we present the accuracy of the latent image (in PSNR) recovered for the INRIA human dataset, with respect to different constant values of β. The image PSNR suffers severely from an overweighted image prior (when $\beta \geq 1$), and varies slightly with a smaller prior weight, *i.e.* no more than 10^{-1}. The highest PSNR of 16.90 dB is observed for $\beta = 10^{-2}$. However, it is worth noting that this level of accuracy is still

Test \ Train	Bottles	Car	Cat	CMU	Human	Yale
Bottles	23.43	20.11	20.91	20.34	20.41	20.87
Car	21.65	24.51	21.93	20.75	19.93	20.21
Cat	20.92	18.31	30.10	21.66	20.88	20.25
CMU	28.19	27.36	26.68	30.75	26.13	28.15
Human	14.24	15.07	13.96	12.95	18.56	14.92
Yale	27.96	25.49	29.24	29.02	25.71	29.04

TABLE 4: Deblurring performance (in PSNR) on each class of test input (blurred) images when using various (external) training datasets. The PSNR is significantly higher when the training dataset matches the test image category.

β	50	5	1	0.5	10^{-1}	$\mathbf{10^{-2}}$	10^{-3}	10^{-4}
PSNR	9.35	8.57	7.57	10.80	16.55	**16.90**	16.15	16.01

TABLE 5: The average image accuracy (in PSNR) achieved with a constant prior weight β when our algorithm is evaluated on the Person dataset [42].

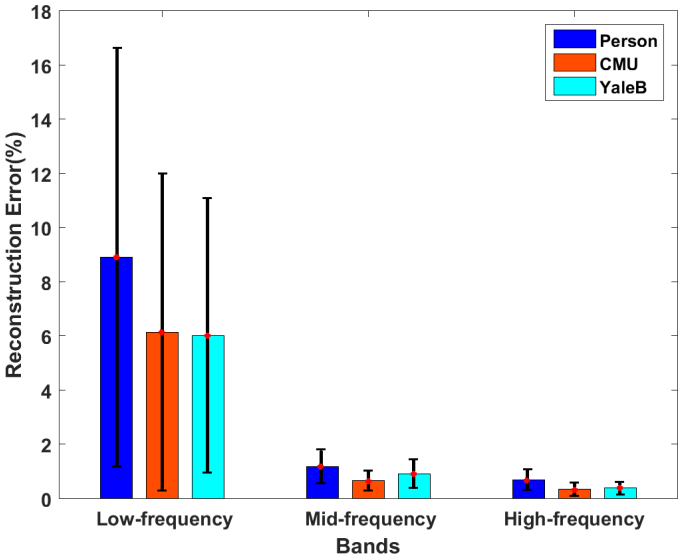


Fig. 5: The relative reconstruction errors (averaged over 80 test images) for the INRIA person [42], the CMU-PIE [38] and the Yale-B [41] datasets.

several orders of magnitude lower than the image PSNR of 18.56 dB, which is reported in the last row and the "Person" column in the PSNR section of Table 8. This comparison demonstrates that the strategy of attenuating β by a factor of $\rho = 1.3$ in every iteration is more effective than using a constant prior weight.

4.2.6 The number of bandpass filters

We also evaluate our algorithm performance with different numbers of bandpass filters *i.e.* M using the same setting for other parameters. In Table 6, we observe that the average image PSNR for the Person dataset [42] varies gradually with respect to different values of M. Since the peak PSNR is achieved at $M = 90$, we employ 90 filters throughout all other experiments.

No. filters	10	30	50	70	**90**	110	130
PSNR	17.31	17.81	17.94	18.12	**18.56**	18.52	18.53

TABLE 6: The average accuracy of the deblurred image (in PNSR) for the Person dataset [42], with respect to different numbers of bandpass filters M.

4.2.7 Reconstruction error of the latent image

To validate the prior, we report the relative error of the latent image reconstructed by the weighted combination of the filtered training images. We evenly divide the 90 bandpass filters into three groups, corresponding to the low-frequency, the mid-frequency and the high-frequency bands, and study the reconstruction error per group. Figure 5 shows the average relative reconstruction error for 80 blurry

	80 filters		90 filters	
	Greyscale	Colour	Greyscale	Colour
Images	18.31	**18.33**	18.56	**18.57**
Kernel	38.68	**39.65**	41.32	**41.78**

TABLE 7: A comparison of the image and kernel accuracy (in PSNR) obtained using greyscale vs. colour input images. The results are reported for the INRIA person dataset [42].

images in the INRIA person [42], the CMU-PIE [38] and the Yale-B [41] datasets, across the above three groups of frequency bands. For each input image, we employ 100 training images from the same class.

Overall, the average errors in most cases are reasonably low (6% and below), except the 9% error for the low-frequency bands in the INRIA Person dataset. This could be explained by the fact that this dataset contains a wider variety of human poses and background than the other datasets. In particular, in the mid-frequency and high-frequency regions, the error mean is 1% or below and one standard deviation above the error mean is lower than 2%, across the datasets. This supports the claim that, with a sufficient number of training images and bandpass filters, we can recover the mid-frequency and high-frequency details of the blurry images with a high level of accuracy.

4.2.8 Grayscale versus colour images

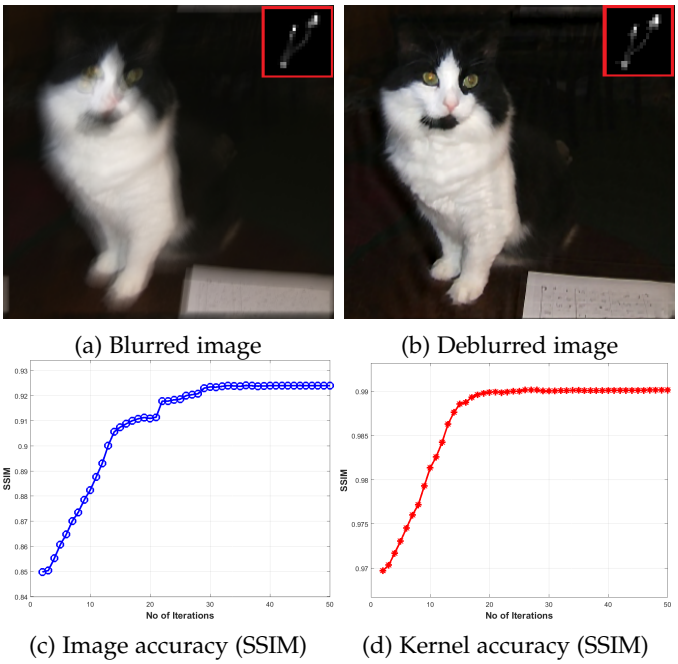
We compare the accuracy of our algorithm when it is run on input colour images as opposed to their grayscale counterparts. As a demonstration, we perform this comparison on the INRIA human dataset [42], using $M = 80$ and $M = 90$ bandpass filters. Table 7 reports the accuracy (in PSNR) of the kernel estimate and the final deblurred image. Under both settings, the kernel PSNR obtained from color input images is higher than that from the grayscale ones. However, there is no clear correlation between the kernel PSNR and the image PSNR as the latter is almost unaffected by the input modality. The explanation is that, although the kernel estimated from colour images is more accurate, it may still lack a number of frequency components in the original image. Therefore, these components could not be recovered from either the grayscale or the colour blurry image directly, but could only be hallucinated using image priors.

4.2.9 Convergence of the algorithm

The objective function in Equation 5 is convex with respect to each of the variables $\mathbf{w}$, $\mathcal{F}_{\mathbf{x}}(\omega)$ and $\mathcal{F}_{\mathbf{k}}(\omega)$. When two of these three variables are fixed, the overall objective function is reduced to those in Equations 6, 8 and 13. Those objective functions are convex with respect to the respective variables to be optimized, because they consist of a quadratic term, and an additional ℓ_1 regularisation term when the weights $\mathbf{w}$ are to be optimized.

Therefore, each alternating minimisation step between lines 5 and 10 of Algorithm 1 is guaranteed to converge to a global minimum for each subproblem. Overall, the algorithm converges to a local minimal solution for the variable triplet $\mathbf{w}$, $\mathbf{x}$ and $\mathbf{k}$.

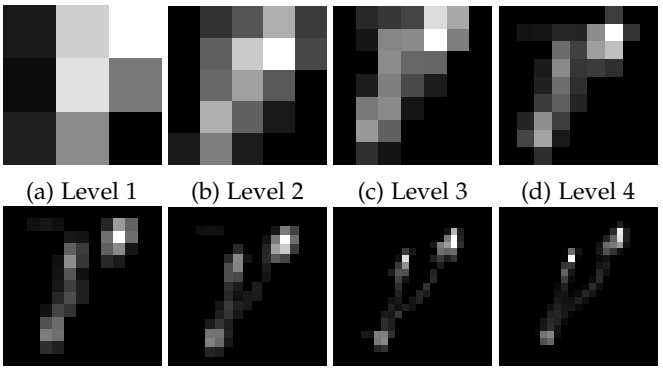
In Figure 6a, we demonstrate the convergence of our algorithm on a sample image. The top row shows the input



(a) Blurred image (b) Deblurred image

(c) Image accuracy (SSIM) (d) Kernel accuracy (SSIM)

Fig. 6: The convergence of the iterative algorithm. The image and kernel similarity between the estimated and the ground-truth are measured in terms of the SSIM.



(a) Level 1 (b) Level 2 (c) Level 3 (d) Level 4

(e) Level 5 (f) Level 6 (g) Final level (h) Ground truth

Fig. 7: Estimated kernels for the sample image in Figure 6 at different scales. As visible, the kernel becomes progressively more similar to the ground-truth at finer resolutions.

(left) and the deblurred image (right). In Figures 6c and 6d, we plot the similarity of the estimated image and kernel to the corresponding ground-truth with respect to the iteration number on the finest scale. Here, the similarity is measured by SSIM. The overall trend is that the estimated image and kernel become increasingly similar to the ground-truth in the long run, and the image and kernel similarity measure plateaus at high values, above 0.92 and 0.99, respectively. In addition, Figure 7 illustrates the progression of the estimated kernel from the coarsest to the finest resolution, for the blurred image in Figure 6a. As shown, the estimated kernel is progressively closer to the ground-truth as the resolution becomes finer.

4.2.10 Running time

One can deduce the complexity of our algorithm in its basic implementation. It is necessary to specify the complexity of each loop, and each step of the algorithm *i.e.* Equation 6, 11 and 16. Let m be the number of pixels of the input image, then the computational load of Equation 6 for N training images is Nm. Equation 11 requires two 2D Fast Fourier Transforms (FFT) and an inverse FFT in each inner iteration/minimisation step. Notice that, the bandpass filters, the derivative filters and responses for the training images are precomputed. Therefore, after simplification and ignoring the constants values, the complexity of Equation 11 is $O(m \log m)$. Similarly, for Equation 16 one can see the the complexity to be same as Equation 11 *i.e.* $O(m \log m)$. The computational complexity of the first pass of the main loop is $O(m \log m + Nm)$, while it takes σ inner iterations and $k_s/3$ outer iterations to give us the final kernel. Hence, the overall complexity of our method is $O(\sigma k_s m \log m + \sigma k_s N m)$, where k_s is the size of the kernel. The execution time for a 320×240 image is 33 seconds with our MATLAB implementation without any code optimization. Notice that, the weight estimation step is highly parallelisable and the 2D FFT operations typically run in real-time at full framerate for much larger video frames in dedicated hardware platforms.

4.2.11 Example real-world images

We also illustrate the qualitative results of our method for real-world examples, where the original sharp images are unavailable. Figure 8 illustrates the qualitative results for two such examples from Pan *et al*. [12]. For deblurring purposes, we employ the same kernel size as Pan *et al*. [12], *i.e.* 35×35 for the first image and 25×25 for the second one. It is noted that the first example contains noisy and/or saturated pixels. For this image, our algorithm recovers finer facial details and hair textures and smoother facial skin than the remaining methods, whereas the others produce ringing artefacts and amplify noise. In the second example, the methods in [6], [12], [28] produce blurry images with ringing artefacts, perhaps due to the sub-optimal selection of edge scales for kernel estimation. Our result is competitive to Krishnan [16], while yielding finer facial details and less ringing artefacts than Levin *et al*. [17].

4.3 Comparisons with generic image deblurring

In this section, we compare the performance of our method to several state-of-the-art deblurring methods that use generic priors on the datasets mentioned earlier. The methods included in our comparison are that of Fergus *et al*. [3], Shan *et al*. [4], Cho and Lee [5], Xu and Jia [6], Krishnan *et al*. [7], Levin *et al*. [8], Cai *et al*. [45], Zhong *et al*. [44], Xu *et al*. [9], Michaeli and Irani [29], Sun *et al*. [28] and Pan *et al*. [12].

We learned a variant of our algorithm on the training examples combined from all the datasets described in section 4.1 and evaluated it on individual classes. We named this variant "class-agnostic" and reported its performance in the second last row of Table 8. We also reported the class-specific variant of our method (last row).

Table 8 presents the average SSIM and PSNR scores for the recovered latent images. Overall, our "class-agnostic" variant outperforms all the other methods under study, and leads the second best method by a significant margin (several dB in terms of PSNR) in most datasets except for Shape [40]. This lead is due to the ability of our method in modelling image signals in a wide range of frequencies. This aspect distinguishes our method from the previous approaches, which mainly employ sparse intensity or gradient priors and, as a consequence, favour reconstructions with uniform regions.

Moreover, our "class-specific" variant, trained on only the relevant object class offers a further performance boost from the class-agnostic variant. This result is an evidence that class-specific examples provides important information for improving the quality of the deblurred image.

Further, Table 9 shows the accuracy of the estimated kernel. Apart from SSIM, we computed the ratio of image MSE when the kernel is estimated (in the blind deconvolution approach) to that achieved in the non-blind approach (with the ground-truth kernel) [43]. Our class-agnostic variant outperforms all the others in terms of ratio of MSE, and is comparable to the best baseline in terms of SSIM. Again, our class-specific variant outperforms all the prior works in both measures. In terms of ratio of MSE, both our approaches outperform the other by at least an order of magnitude on several classes. Since our method captures class-specific information in every frequency band of the latent image, it is capable of coping with a broad range of kernels, irrespective of whether they are sparse or not.

Further, we have evaluated our algorithm with training examples combined from all object classes, and compared its performance to the case of separate training classes. The second last row in Table 8 reports the image SSIM and PSNR using training examples from all object classes. Indeed, including all the object classes in the training data degrades the image accuracy compared to only the correct training class. This result is an evidence that examples within the same class are more beneficial to the deblurring accuracy than those outside the class.

For a comprehensive evaluation, we present the qualitative comparisons for sample images from the datasets under study. As the first example, in Figure 9, we show the deblurring results for a car image from the dataset in [39]. Overall, our method produces the image with the smallest amount of artifacts and the most accurate kernel. Note that the ground-truth image in Figure 9a does not contain much texture except for a small number of edges. Therefore, the methods that amplify edges such as those in [5], [6], [9], [12] receive limited information, thus, cannot handle this case well. Moreover, the methods based on gradient sparsity priors, including those in [7], [8], [47], tend to produce artifacts in the deblurred image. Levin *et al*. [8] (Figure 9g) appears to generate a similar result to our method (Figure 9o). However, a close inspection reveals that [8] contains ringing defects in the deblurred image and undesirable non-zeros in the kernel.

As a second example, we present deblurring results on a challenging image from the INRIA person dataset [42]. As shown in Figure 10, the input image has a low resolution of 64×80 and incurs severe blur due to the large kernel size rel-

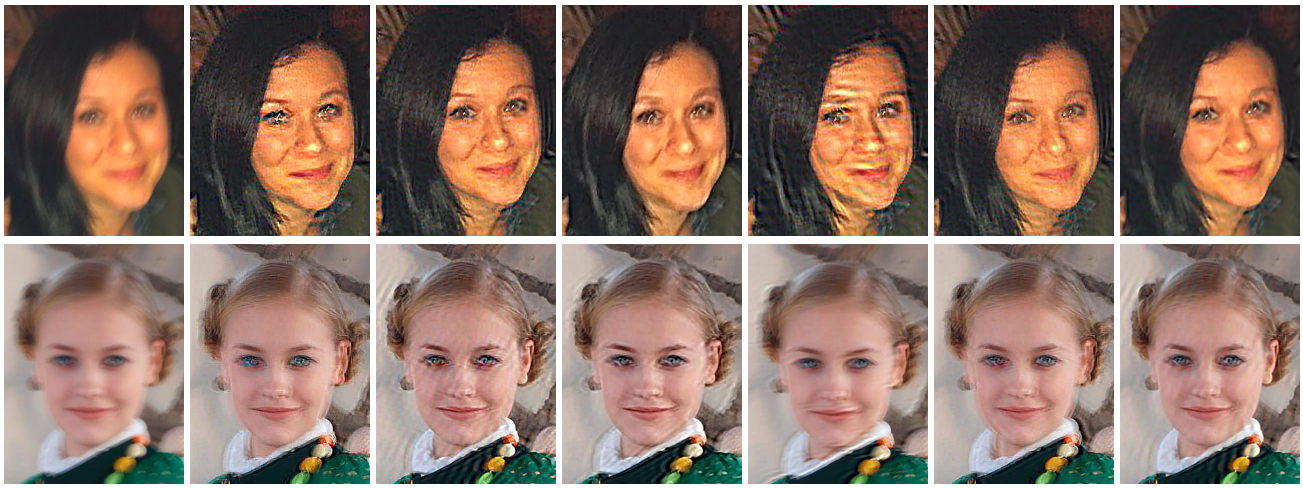


Fig. 8: Deblurring results for real input images from [12], where the one in the first row contains noise and saturated pixels.

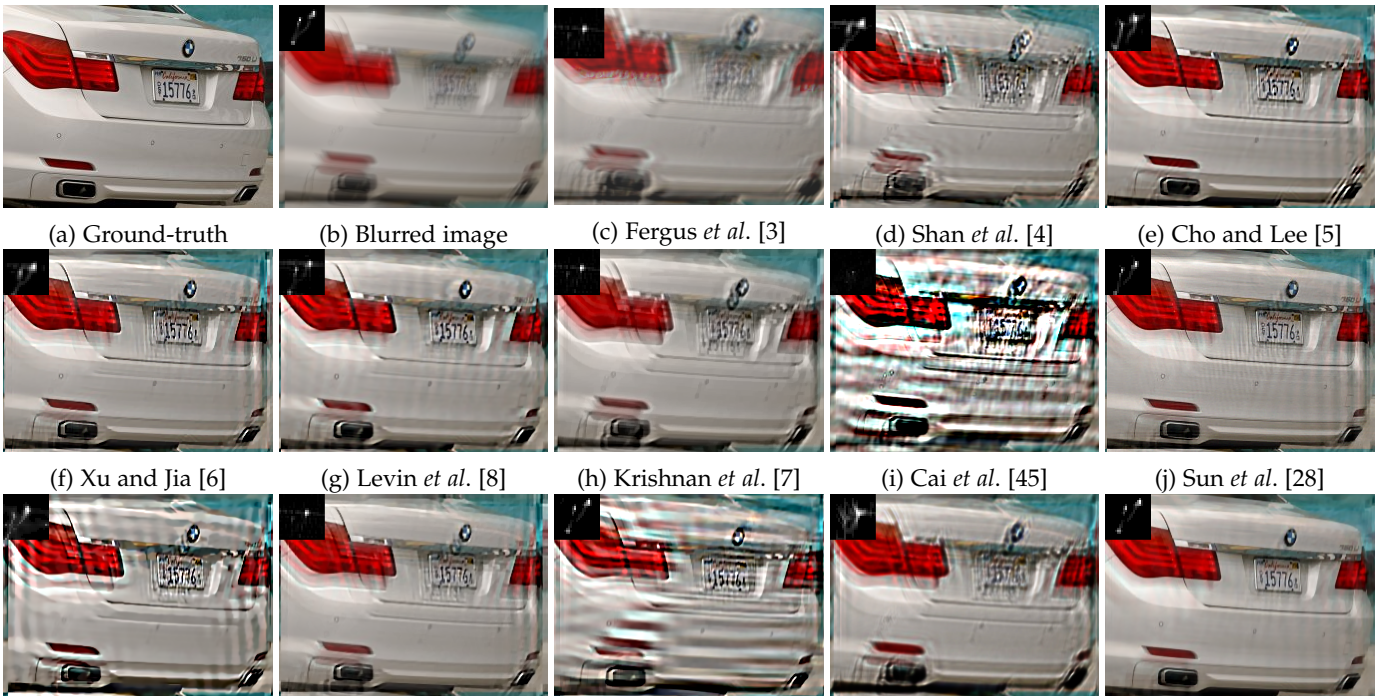
	SSIM (the higher the better)						PSNR (the higher the better)					
Methods	Car [39]	Shape [40]	Cat [20]	CMU [38]	Person [42]	YaleB [41]	Car [39]	Shape [40]	Cat [20]	CMU [38]	Person [42]	YaleB [41]
Fergus [3]	0.411	0.415	0.598	0.559	0.207	0.535	16.99	16.69	19.88	18.26	14.81	19.56
Shan [4]	0.632	0.624	0.742	0.775	0.407	0.773	21.56	21.61	25.20	25.59	17.78	26.42
Cho [5]	0.559	0.595	0.627	0.699	0.293	0.678	19.99	20.47	22.54	24.38	15.05	22.99
Xu [6]	0.631	0.638	0.704	0.739	-	0.681	20.93	21.25	22.73	23.30	-	23.30
Krishnan [7]	0.544	0.544	0.668	0.693	0.296	0.755	19.75	19.73	22.79	23.54	15.41	24.09
Levin [8]	0.500	0.567	0.699	0.758	0.332	0.673	18.09	19.24	23.12	24.31	16.77	25.22
Cai [45]	0.298	0.358	0.292	0.178	-	0.205	13.89	14.86	14.63	11.72	-	12.37
Zhong [44]	0.485	0.520	0.643	0.641	-	0.655	17.23	18.00	20.73	20.93	-	22.16
Michaeli [29]	0.605	0.641	0.735	0.693	0.226	0.660	18.34	17.36	19.47	18.80	13.39	22.59
Sun [28]	0.481	0.669	0.724	0.744	-	0.680	19.06	22.50	23.93	24.78	-	23.74
Class-agnostic	0.760	0.697	0.818	0.860	0.509	0.745	23.56	22.41	27.14	29.35	18.46	28.03
Class-specific	**0.765**	**0.715**	**0.864**	**0.881**	**0.509**	**0.788**	**24.51**	**23.43**	**30.10**	**30.75**	**18.56**	**29.04**

TABLE 8: Accuracy of the deblurred images, measured by SSIM and PSNR. The missing results, indicated by "-", occurs when the respective method is not capable of dealing with the low resolution of the input images. Best results are in bold.

	SSIM (the higher the better)						Ratio of MSE (the lower the better)					
Methods	Car [39]	Shape [40]	Cat [20]	CMU [38]	Person [42]	YaleB [41]	Car [39]	Shape [40]	Cat [20]	CMU [38]	Person [42]	YaleB [41]
Fergus [3]	0.629	0.668	0.653	0.759	0.589	0.778	16.99	26.84	11.70	23.11	11.13	14.34
Shan [4]	0.831	0.816	0.819	0.816	0.778	0.780	8.13	13.59	8.72	13.55	11.83	15.34
Cho [5]	0.845	0.830	0.834	0.820	0.803	0.809	32.19	33.39	43.16	45.15	30.94	26.34
Xu [6]	0.840	0.761	0.837	0.840	-	0.815	8.57	11.81	9.84	13.15	-	17.96
Krishnan [7]	0.724	0.721	0.719	0.787	0.698	0.760	9.41	14.47	9.26	12.55	11.22	14.36
Levin [8]	0.702	0.692	0.750	0.782	0.621	0.762	12.53	15.11	8.80	10.84	8.25	10.46
Cai [45]	0.640	0.627	0.652	0.669	-	0.662	36.32	61.23	74.27	188.91	-	252.09
Zhong [44]	0.704	0.755	0.764	0.774	-	0.748	15.72	21.39	15.23	21.02	-	22.68
Michaeli [29]	0.607	0.577	0.606	0.588	0.572	0.660	11.92	22.43	18.09	30.03	18.88	21.31
Sun [28]	0.829	0.822	0.816	0.826	-	0.813	10.66	10.52	8.02	10.35	-	18.11
Class-agnostic	0.842	0.801	0.823	0.883	0.795	0.784	4.01	8.08	2.20	3.93	7.84	11.34
Class-specific	**0.884**	**0.833**	**0.886**	**0.901**	**0.805**	**0.823**	**3.93**	**7.91**	**1.77**	**2.99**	**7.59**	**10.30**

TABLE 9: The accuracy of the estimated kernel, measured by SSIM and ratio of MSE. The missing results, indicated by "-", occurs when the respective method is not capable of dealing with the low resolution of the input images. Best results are in bold.

ative to the image size. Note that the results for Sun *et al*. [28] and Xu *et al*.'s [9] methods are not available since their original implementations are unable to handle the low image resolution. Since all the edges are significantly distorted, sparsity and gradient priors do not benefit the deblurring task. For this reason, it is difficult for the methods that utilise these priors, such as those in [5], [6], [44], the MAP frameworks [4], [7] and the variational Bayesian ones [3], [8], to explicitly recover sharp edges for kernel computation. In Figure 10, we show that our method successfully recovers several objects with a large resemblance to the ground-truth, namely the foreground and background pedestrians, as well as the bus in the background. In contrast, the mentioned objects are unrecognisable in the other deblurred images. Moreover, our estimated kernel is also the most accurate one among all the methods.



(a) Ground-truth (b) Blurred image (c) Fergus *et al*. [3] (d) Shan *et al*. [4] (e) Cho and Lee [5]

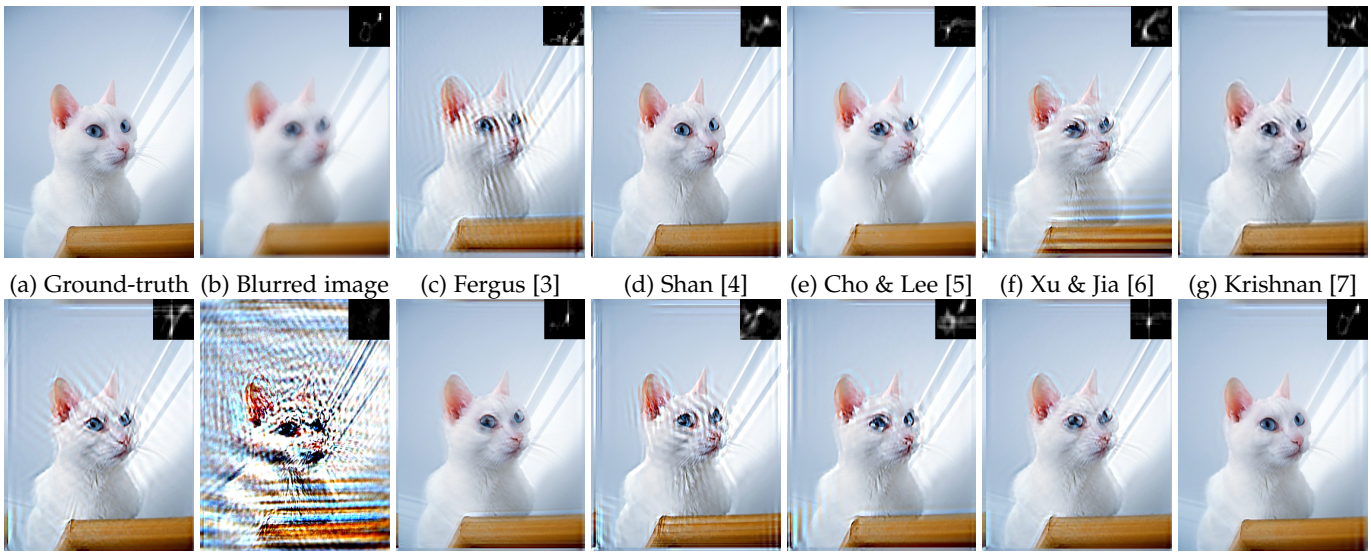
(f) Xu and Jia [6] (g) Levin *et al*. [8] (h) Krishnan *et al*. [7] (i) Cai *et al*. [45] (j) Sun *et al*. [28]

(k) Zhong *et al*. [44] (l) Xu *et al*. [9] (m) Pan *et al*. [12] (n) Pan *et al*. [13] (o) Our method

Fig. 9: Results for a sample image from the Car dataset in [39]. The restored image from our method has more legible text on the license plate compared to the other methods.

Original Blurred Fergus [3] Shan [4] Cho [5] Xu & Jia [6] Krishnan [7] Levin [8] Zhong [44] Ours

Fig. 10: Comparison of several methods on a sample image selected from the INRIA dataset [42]. Our method successfully recovers parts of the image with a significant resemblance to the ground-truth, including the pedestrians and the bus in the background. Our estimated kernel is also the most accurate among all the methods.



(a) Ground-truth (b) Blurred image (c) Fergus [3] (d) Shan [4] (e) Cho & Lee [5] (f) Xu & Jia [6] (g) Krishnan [7]

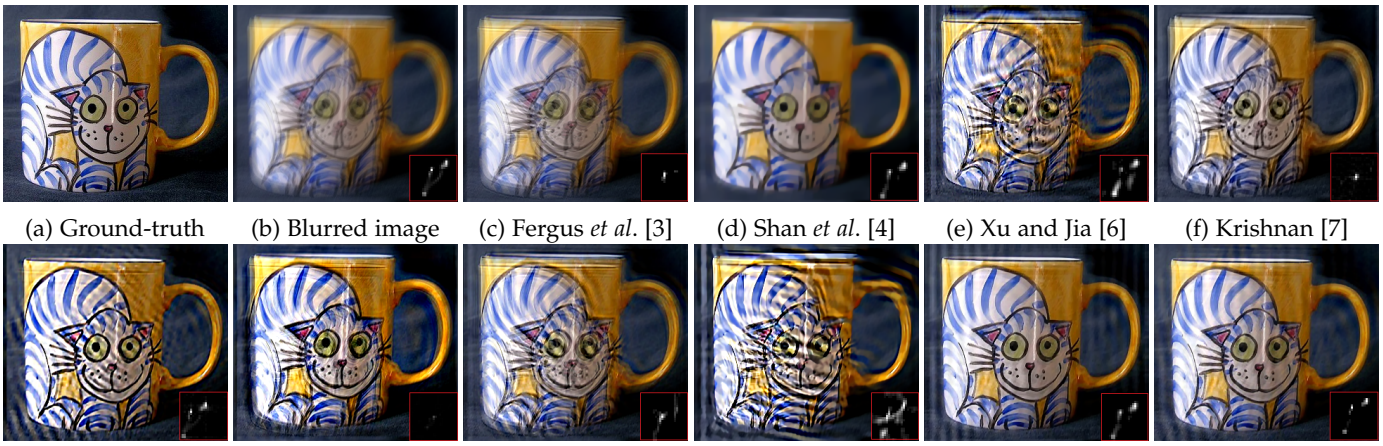
(h) Levin [8] (i) Cai [45] (j) Sun [28] (k) Zhong [44] (l) Xu [9] (m) Pan [12] (n) Our method

Fig. 11: Comparisons on a sample image from the Cat dataset [20]. Our method recovers fine texture around the neck, mouth and whiskers, which cannot be accurately reproduced by the others.

(a) Ground-truth (b) Blurred image (c) Fergus *et al*. [3] (d) Cho and Lee [5] (e) Xu and Jia [6] (f) Krishnan *et al*. [7]

(g) Levin *et al*. [8] (h) Cai *et al*. [45] (i) Sun *et al*. [28] (j) Zhong *et al*. [44] (k) Xu *et al*. [9] (l) Our method

Fig. 12: Comparisons on a sample image with strong edges and a blurred background, selected from the ETHZ Shape Classes dataset [40]. The visual quality, *e.g*. sharpness of the text on the label, reproduced by our method is par to the best one among the other methods, *i.e*. Sun *et al*.'s [28].



(a) Ground-truth (b) Blurred image (c) Fergus *et al*. [3] (d) Shan *et al*. [4] (e) Xu and Jia [6] (f) Krishnan [7]

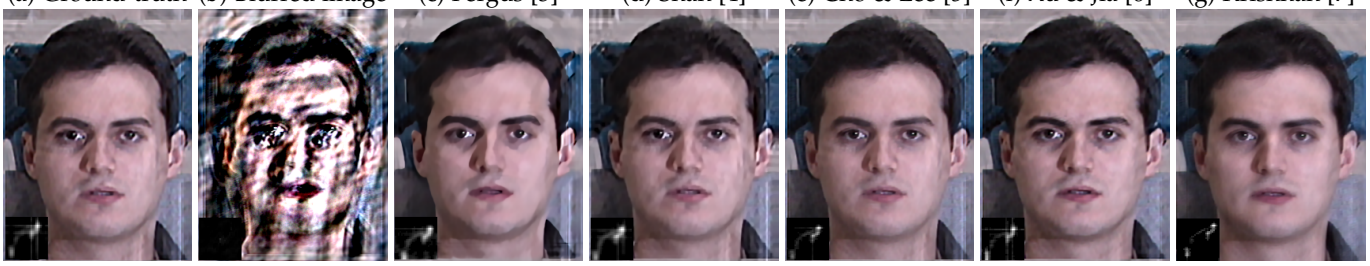
(g) Levin *et al*. [8] (h) Cai *et al*. [45] (i) Sun *et al*. [28] (j) Zhong *et al*. [44] (k) Xu *et al*. [9] (l) Our method

Fig. 13: Comparisons on a sample image with rich textures, selected from the ETHZ Shape Classes dataset [40]. On a magnified view, the image our method recovers is sharper than those generated by most of the methods, and comparable to the best, *i.e*. of Xu *et al*. [9], while exhibiting a less degree of ringing artifacts.

Next, we depict an example from the Cat dataset [20]. The visual quality of our recovered image, as shown in Figure 11n, outperforms all others. Noticeable features restored by our method include the sharpness and the clarity of the cat's eyes. Our method can also recover subtle textures around the neck, mouth and whiskers of the cat while they are not reconstructed in the results produced by the other methods. Further, a magnified view of the results in Figures 11e-11m shows that the methods that rely on edges, *e.g*. Krishnan *et al*. [7], and patches with high-contrast, *e.g*. Sun *et al*. [28], fail to yield an accurate estimation of the kernel. Our method outperforms Pan *et al*.'s [12], which uses class exemplars with additional manually drawn mask input around the contours of the cat's head, the mouth and eyes in the ground-truth image.

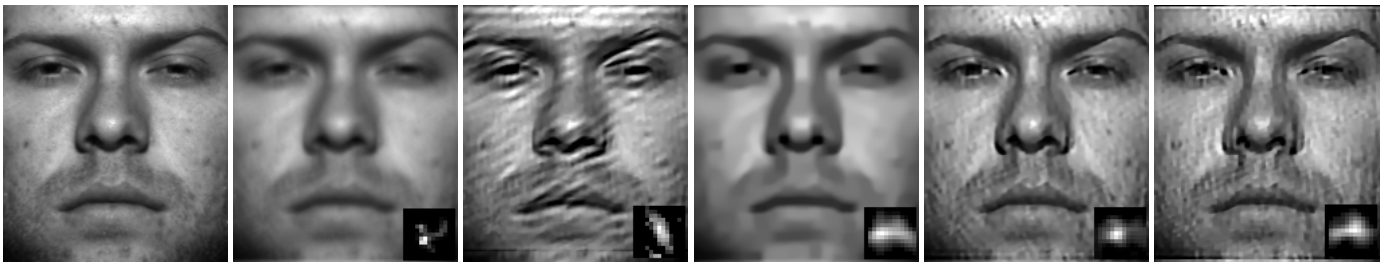
Subsequently, we examine the visual quality of the results achieved by our method and several others on two examples from the ETHZ Shape Classes dataset [40]. Firstly, in Figure 12, we show results for an image containing strong occlusion and text edges and a blurred background with no sharp textures. Again, our method can recover reasonably sharp edges and text in the image. Meanwhile, the methods of Fergus *et al*. [3], Xu and Jia [6], Krishnan *et al*. [7], Levin *et al*. [8], Cai et al. [45], Zhong *et al*. [44] and Xu *et al*. [9] have

(a) Ground-truth (b) Blurred image (c) Fergus [3] (d) Shan [4] (e) Cho & Lee [5] (f) Xu & Jia [6] (g) Krishnan [7]

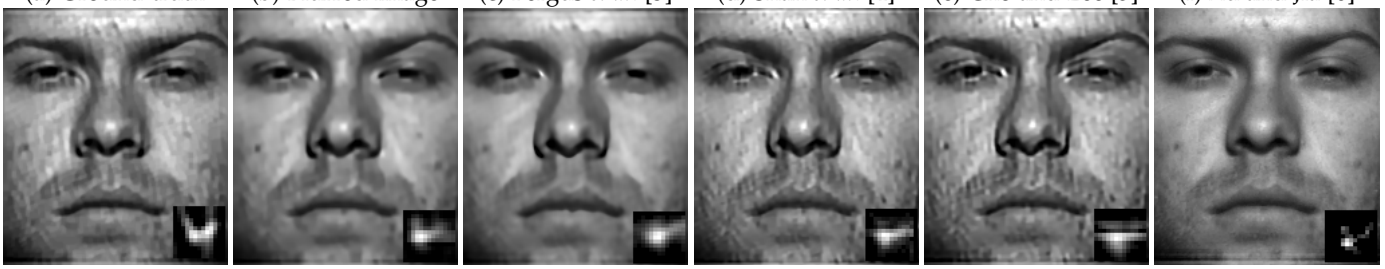


(h) Levin [8] (i) Cai [45] (j) Zhong [44] (k) Xu [9] (l) Sun [28] (m) Pan [12] (n) Our method

Fig. 14: Comparisons on a face image selected from the CMU PIE dataset [38]. Although our deblurred image appears to be similar to those produced some other methods, its intensity profile (on the face) is richer than the other methods.

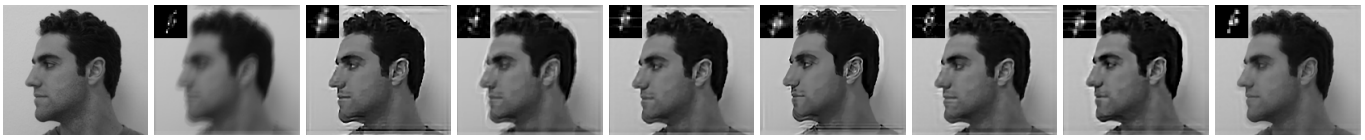


(a) Ground-truth (b) Blurred image (c) Fergus *et al*. [3] (d) Shan *et al*. [4] (e) Cho and Lee [5] (f) Xu and Jia [6]



(g) Krishnan *et al*. [7] (h) Levin *et al*. [8] (i) Zhong *et al*. [44] (j) Sun *et al*. [28] (k) Pan *et al*. [12] (l) Our method

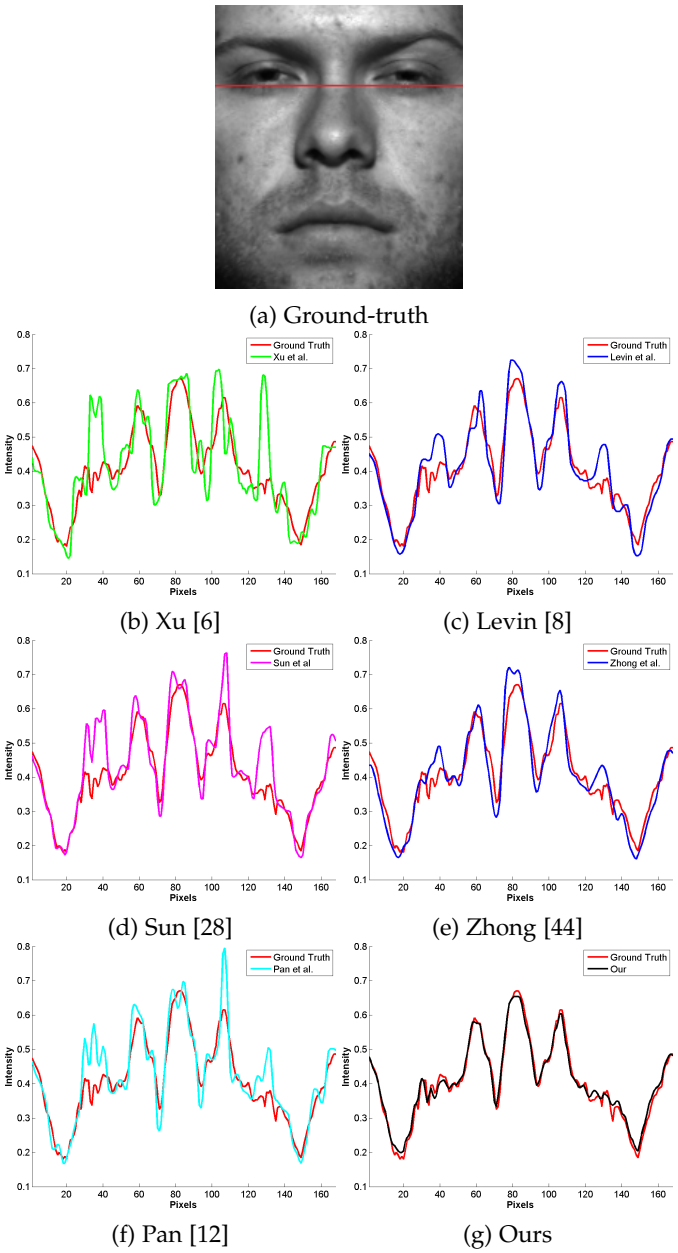
Fig. 15: Comparisons on a sample face image selected from the Yale-B dataset [41]. The image we recover is more natural and contains less ringing and exaggerated contrast artifacts. Our estimated kernel is also the closest to the ground-truth.



Original | Blurred | Xu & Jia [6] 22.01 dB | Krishnan [7] 25.05 dB | Levin [8] 22.67 dB | Zhong [44] 24.66 dB | Sun [28] 25.28 dB | Pan [12] 25.78 dB | Ours 29.51 dB

Fig. 16: Comparisons on a sample image from the FEI dataset [46]. Differences can be better seen in magnified view.

poorly estimated the PSF, which indirectly causes ringing artifacts and multiple false edges in the deblurred image. At close inspection, Cho and Lee's method [5] produces slight ringing on the left occlusion boundary of the bottle and false edges on the white background of the label. Meanwhile, Sun *et al*.'s result in Figure 12i appears to be comparable to ours, although the kernel they recover incurs a downward shift compared to the ground-truth. Secondly, we qualitatively compare deblurring results for an image with rich textures and edge information, as shown in Figure 13. The blurred edges in the input image 13b are of different thicknesses, which potentially causes incorrect estimation of the kernel.



(a) Ground-truth (b) Xu [6] (c) Levin [8] (d) Sun [28] (e) Zhong [44] (f) Pan [12] (g) Ours

Fig. 17: Intensity profiles (corresponding to pixel row 55) of the deblurred images produced by our method and others. The input blurred image is given in Figure 15b. The red trace in each subplot shows the ground-truth profile.

Therefore, methods relying on strong or thick edges such as Fergus *et al*. [3], Xu and Jia [6], Krishnan *et al*. [7], Levin *et al*. [8] result in strong ringing artifacts and incorrect edges. Shan *et al*.'s method [4] suffers less ringing artifacts than the others but the result appears to be over-smoothed. On a magnified view, the image we recover is sharper than those produced by most of the other methods, and comparable to the best of them, *i.e*. of Xu *et al*. [9], while exhibiting a lower level of ringing artifacts.

Further, we examine the results for two benchmark face image datasets. The first one is taken from the CMU PIE face dataset [38]. As shown in Figure 14b, the blurred image is quite challenging due to the large scale and the complex trajectory of the blur-induced motion. As a result, Fergus *et al*. [3], Shan *et al*. [4], Krishnan *et al*. [7], Levin *et al*. [8], Xu *et al*. [9] and Pan *et al*. [12] yields incorrect kernels which are less sparse than the ground-truth one. Although our deblurred image appears to be similar to those of Xu and Jia [6], Zhong *et al*. [44], Xu *et al*. [9] and Sun *et al*. [28], it shows gradual changes in the intensity across the face, as opposed to the flatness on the other deblurred images. This result suggests that our algorithm has recovered a wider range of spatial frequencies than the high frequencies reproduced by the other methods.

Another face example is selected from the Yale-B dataset [41], which contains cropped and well-aligned face images. In Figure 15, note that Xu *et al*.'s result is not available for this example since the image dimensions of less than 200×200 pixels are below the limit that can be handled by their implementation. The methods of [3], [5], [6], [7], [12], [28] emphasize noise and artifacts and estimate the kernel incorrectly. Meanwhile, the images estimated by [4], [8] and [44] are either over-smooth or lack fine details such as hair and speckles.

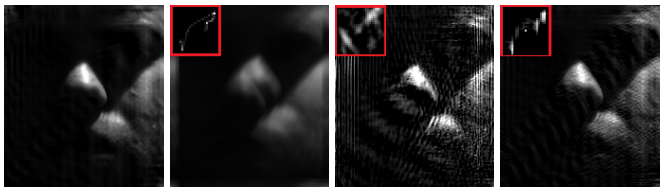
To visually demonstrate that our method recovers the image more accurately, we randomly select a row of pixels from the ground-truth image and compare it with the corresponding row in the recovered image. In Figure 17, it can be observed that our method is the closest to the ground-truth scanline. The failure of the alternative methods is partly due to the lack of strong edges in the blurred input image. On the other hand, by taking the learned frequency spectrum of faces, our method can recover more curvature and finer details in the face, and less ringing artifacts than the others. Our estimated kernel is also the closest to the ground-truth compared to the others.

Furthermore, our method is not restricted to only frontal face images, but can also deblur face images in a different view. In Figure 16, we show results for an image from the FEI dataset [46]. The training data contains images with frontal as well as different viewing angles. Our method yields the highest accuracy (PSNR) using a training dataset comprising images in a similar pose.

4.4 Comparison with exemplar-based methods

For completeness, we compare our algorithm to several state-of-the-art deblurring methods that use class exemplars. Since the implementations of these methods are not available publicly, the comparisons are purely based on the qualitative results reported in the respective papers. In the comparison, we consider the methods of Zhang et al [48], HaCohen *et al*. [33], and those of Pan *et al*. [12], [13].

As shown in Figure 18a), we examine the performance of our method and compare it directly with that of Zhang *et al*. [48] from their paper, which they classified as a failure case. This is a challenging example as most of the pixels are dark and noise prone, and there are almost no salient edge features to estimate the kernel correctly. In Figure 18b, we show a blurred image generated by the ground-truth kernel depicted at its top left corner. In addition, we obtain the deblurring result recovered by Zhang *et al*.'s algorithm [48] directly from their paper and display both the latent image and the kernel in Figure 18c (top-left corner). Similarly, our deblurring result is shown in Figure 18d. As visible,



(a) Groundtruth (b) Blurred (c) [48] (d) Our result

Fig. 18: Comparison with Zhang *et al*. [48]. (a) Ground-truth image, (b) blurred image, (c) deblurring results produced by Zhang *et al*. [48], which is reported as a failure case in their paper, (d) our results.

(a) Original (b) Blurred (c) Reference Image used by (d) (d) HaCohen PSNR: 26.23

(e) Our training images

(f) Our result PSNR: 29.90

Fig. 19: Comparison to HaCohen *et al*.'s [33] on a blurred image taken from their paper.

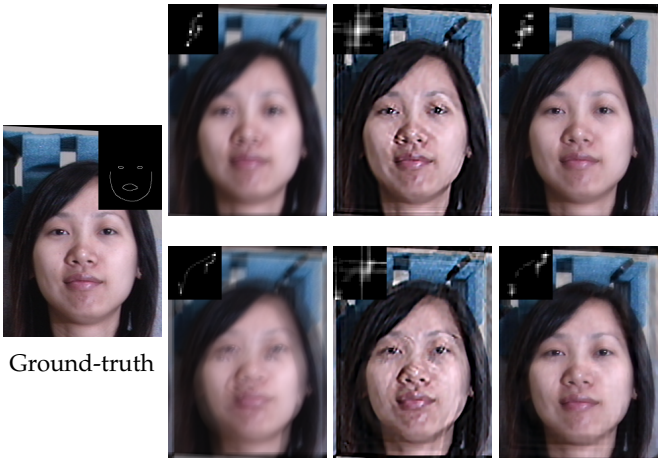
the latent image produced by Zhang *et al*. suffers from ringing artifacts and the estimated kernel does not resemble the sparse structure of the ground-truth. In contrast, our recovered latent image contains much fewer artifacts and is visually closer to the ground-truth. Furthermore, the kernel computed by our algorithm appears to be more similar in shape to the original kernel and much sparser than that produced by [48].

Next, we illustrate the advantage of our method over that by HaCohen *et al*. [33] by an example taken directly from their paper. This method requires a dense correspondence to be established between the blurred input image and a reference image of the same content and structure. Here, it is observed that our result is comparable to the other method. However, the greatest gain from our method is the simplicity of the required input. Our method does not need a reference image with restrictive content and structure and a correspondence map to the blurred image. The only requirement for our input is that the training images belong to the same class. In this example, [33] employed a reference image of the same person, with many matches to the blurred image, whereas our method permits the flexibility to collect training faces images of various individuals and expressions.

More recently, Pan *et al*. [13] introduced a deblurring

(a) Original (b) Blurred (c) Pan [13] (d) Our result

Fig. 20: Deblurring of an image containing foreground text and complex background. (a) Ground-truth (sharp) image, (b) blurred image, (c) deblurring results by Pan *et al*. [13], (d) our results (zoom in to see the differences).



Ground-truth Blurred image Pan *et al*. [12] Our result

Fig. 21: Results for an image provided by Pan *et al*. [12]. First column: ground-truth image, second column: blurred images and original blur kernels (at the top left corners of the images), third column: deblurred images and estimated kernels by Pan *et al*. [12], fourth column: our results.

algorithm for two-tone text images using an L_0-regularised intensity and gradient prior and applied it to the deblurring of non-document text images. We examine the performance of their method on an image from ETHZ shape classes dataset [40]. As shown in Figure 20a, the image contains a cup with large printed text in the foreground and some cluttered background regions, which possess the same features as the non-document text images used in their paper. Comparing our deblurring result in Figure 20d to that of Pan *et al*. [13] in 20c, we observe that our estimated kernel is more similar to the ground-truth kernel than the other. Moreover, in the image recovered by Pan *et al*., ringing artifacts are visible around the edges of the text printed on the cup. In contrast, our algorithm recovers the foreground text with increased sharpness and much less ringing than the former method. Furthermore, it restores the legibility of the background text within the marked red box, which Pan *et al*. [13] failed. This stems from the fact that, the ℓ_0-regularised prior employed by Pan *et al*. simply favours uniform intensity regions, which is insufficient to capture spatial variations caused by illumination and shading in shape images. In this example, our method has aimed to capture this variation via the subspace of frequency bands, and therefore is more successful in restoring the original image.

Subsequently, we compare our method with [12], which

aims at face image deblurring using annotated salient edges of sharp exemplars. For a fair comparison, we use the dataset and the mask annotations provided by the authors. In Figure 21, we depict the deblurring results delivered by both methods for an example ground-truth image (the first column), which are blurred with two different blurred kernels (the second column). Here, we run their implementation with the contour mask obtained from the ground-truth image as shown at the top left corner in the first column of Figure 21. This mask, according to [12], plays a major role in the selection of salient edges as input for their method. The recovered images by [12] and our method are shown in the third and fourth columns, respectively. We also show the estimated kernel at the top-left corners of these images. Comparing these results, our method produces a sharper image with much fewer artifacts. Besides, it can be seen that our kernels exhibit a strong similarity to the ground-truth ones.

5 Conclusion and Future Work

We have introduced a novel class-specific prior that significantly improves the performance of image deblurring. The prior is designed to capture the properties of transform domain coefficients for specific image classes over the entire spectrum of frequency bands. Representing images on the class-specific subspaces, we reconstruct the frequency responses suppressed after the blurring process. Our approach overcomes the limitation of existing methods when dealing with blurred images lacking high-frequency details. We have demonstrated the role of this prior in extensive experimental evaluations. We show that our method outperforms prior deconvolution works that use generic priors and class exemplars both in numerical accuracy and visual quality.

At the current state, our algorithm focuses on deblurring of images containing a single object using a class-specific training dataset. In the future, this work can be extended to deal with multiple objects. This could be achieved *e.g.* first localising and classifying the different objects in the image, and deblurring each object region separately using the training data for the corresponding class. Furthermore, it is worth investigating whether, and if so, to what extent, class-specific training data is required as opposed to generic training data.

Our algorithm is currently limited by the assumption of spatially uniform blur. In the future, we would like to extend our blur model to handle non-uniform blur caused by camera motion, rotation and defocus. This extension requires the geometrical and physical modeling of image formation in the above circumstances.

References

[1] T. Trott, "The effect of motion of resolution," *Photogrammetric Engineering*, vol. 26, pp. 819–827, 1960.

[2] A. Levin, "Blind motion deblurring using image statistics," in *NIPS*, 2006, pp. 841–848.

[3] 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)*, vol. 25, no. 3, 2006, pp. 787–794.

[4] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," in *TOG*, vol. 27, no. 3, 2008, p. 73.

[5] S. Cho and S. Lee, "Fast motion deblurring," in *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5, 2009, p. 145.

[6] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *ECCV*, 2010.

[7] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *CVPR*, 2011, pp. 233–240.

[8] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *CVPR*, 2011.

[9] L. Xu, S. Zheng, and J. Jia, "Unnatural l_0 sparse representation for natural image deblurring," in *CVPR*, 2013.

[10] Y.-W. Tai, X. Chen, S. Kim, S. J. Kim, F. Li, J. Yang, J. Yu, Y. Matsushita, and M. S. Brown, "Nonlinear camera response functions and image deblurring: Theoretical analysis and practice," *TPAMI*, vol. 35, no. 10, pp. 2498–2512, 2013.

[11] A. Mosleh, J. P. Langlois, and P. Green, "Image deconvolution ringing artifact detection and removal via psf frequency analysis," in *ECCV*, vol. 8692, 2014, pp. 247–262.

[12] J. Pan, Z. Hu, Z. Su, and M. Yang, "Deblurring face images with exemplars," in *ECCV*, 2014, pp. 47–62.

[13] J. Pan, Z. Hu, Z. Su, and M. H. Yang, "Deblurring text images via L0 regularized intensity and gradient prior," in *CVPR*, 2014, pp. 2901–2908.

[14] L. Xu, X. Tao, and J. Jia, "Inverse kernels for fast spatial deconvolution," in *ECCV*, 2014, pp. 33–48.

[15] O. Whyte, J. Sivic, and A. Zisserman, "Deblurring shaken and partially saturated images," *IJCV*, vol. 110, no. 2, pp. 185–201, 2014.

[16] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-laplacian priors," in *NIPS*, 2009.

[17] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *TPAMI*, 2011.

[18] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, 2007.

[19] E. Hewitt and R. E. Hewitt, "The gibbs-wilbraham phenomenon: an episode in fourier analysis," *Archive for history of Exact Sciences*, vol. 21, no. 2, pp. 129–160, 1979.

[20] W. Zhang, J. Sun, and X. Tang, "Cat head detection-how to effectively exploit shape and texture features," in *ECCV*. Springer, 2008, pp. 802–816.

[21] A. Torralba and A. Oliva, "Statistics of natural image categories," *Network: computation in neural systems (NCNS)*, vol. 14, no. 3, pp. 391–412, 2003.

[22] J.-M. Geusebroek and A. W. M. Smeulders, "A six-stimulus theory for stochastic texture," *IJCV*, vol. 62, no. 1-2, pp. 7–16, 2005.

[23] J. van Gemert, J.-M. Geusebroek, C. Veenman, C. Snoek, and A. Smeulders, "Robust scene categorization by learning image statistics in context," in *CVPR Workshop*, June 2006, pp. 105–105.

[24] A. Levin, "Blind motion deblurring using image statistics," in *NIPS*. MIT Press, 2007, pp. 841–848.

[25] N. Joshi, R. Szeliski, and D. Kriegman, "Psf estimation using sharp edge prediction," in *CVPR*, 2008, pp. 1–8.

[26] T. S. Cho, S. Paris, B. K. Horn, and W. T. Freeman, "Blur kernel estimation using the radon transform," in *CVPR*, 2011.

[27] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *ICCV*, 2011, pp. 479–486.

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

[29] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in *ECCV*, 2014, pp. 783–798.

[30] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *TPAMI*, vol. 38, no. 7, pp. 1439–1451, 2016.

[31] A. Chakrabarti, "A neural approach to blind motion deblurring," in *ECCV*, 2016, pp. 221–235.

[32] N. Joshi, W. Matusik, E. H. Adelson, and D. J. Kriegman, "Personal photo enhancement using example images," *ACM Trans. Graph*, 2010.

[33] Y. Hacohen, E. Shechtman, and D. Lischinski, "Deblurring by example using dense correspondence," in *ICCV*, 2013.

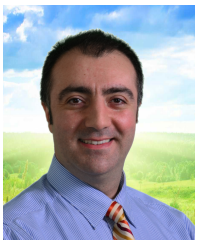
[34] L. Sun, S. Cho, J. Wang, and J. Hays, "Good Image Priors for Non-blind Deconvolution - Generic vs. Specific," in *ECCV*, 2014, pp. 231–246.

[35] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Addison-Wesley Longman Publishing Co., Inc., 1992.

[36] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image deblurring with blurred/noisy image pairs," in *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3. ACM, 2007, p. 1.
[37] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An Interior-Point Method for Large-Scale L1-Regularized Least Squares," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.
[38] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," *Automatic Face and Gesture Recognition*, pp. 46–51, 2002.
[39] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3d object representations for fine-grained categorization," in *ICCVW*, 2013, pp. 554–561.
[40] V. Ferrari, F. Jurie, and C. Schmid, "From images to shape models for object detection," *IJCV*, vol. 87, no. 3, pp. 284–303, 2010.
[41] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *TPAMI*, vol. 23, no. 6, 2001.
[42] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, vol. 2, June 2005, pp. 886–893.
[43] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *CVPR*, 2009.
[44] L. Zhong, S. Cho, D. Metaxas, S. Paris, and J. Wang, "Handling noise in single image deblurring using directional filters," in *CVPR*, 2013.
[45] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Framelet-based blind motion deblurring from a single image," *Image Processing, IEEE Transactions on*, vol. 21, no. 2, pp. 562–572, 2012.
[46] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, 2010.
[47] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in *CVPR*, 2009.
[48] H. Zhang, J. Yang, Y. Zhang, N. M. Nasrabadi, and T. S. Huang, "Close the loop: Joint blind image restoration and recognition with sparse representation prior," in *ICCV*, 2011, pp. 770–777.

Saeed Anwar received a Bachelor degree in Computer Systems Engineering with distinction from University of Engineering and Technology (UET), Pakistan, in July 2008, and a Master degree in Erasmus Mundus Vision and Robotics (Vibot), jointly from Heriot Watt University United Kingdom (HW), University of Girona Spain (UD) and University of Burgundy France in August 2010 with distinction. During his masters, he carried out his thesis at Toshiba Medical Visualization Systems Europe (TMVSE), Scotland. He has also been a visiting research fellow at Pal Robotics, Barcelona in 2011. Since 2014, he is a Ph.D. student at the Australian National University (ANU) and Data61/CSIRO. He has also been working as lecturer/Assistant Professor at the National University of Computer and Emerging Sciences (NUCES), Pakistan. His major research interests are low level vision, image enhancement, image restoration, computer vision and optimization.

Cong Phuoc Huynh is an Adjunct Research Fellow at the Australian National University. He has co-authored a book on imaging spectroscopy for scene analysis and over 20 journal articles and conference papers in computer vision and pattern recognition. He is an inventor of eight patents on spectral imaging. He is a co-recipient of a DICTA Best Student's paper Award in 2013. Previously, he was a computer vision researcher at National ICT Australia (NICTA). He received a BSc degree (Hons) in Computer Science and Software Engineering from the University of Canterbury, New Zealand in 2006, and MSc. and PhD. degrees in Computer Science from the Australian National University (ANU) in 2007 and 2012.



Fatih Porikli is an IEEE Fellow and a Professor in the Research School of Engineering, Australian National University (ANU). He has received his PhD from New York University (NYU) in 2002. Previously he served Distinguished Research Scientist at Mitsubishi Electric Research Laboratories. His research interests include computer vision, deep learning, manifold learning, online learning, and image enhancement with commercial applications in video surveillance, intelligent transportation, satellite, and medical systems. Prof. Porikli is the recipient of the R&D 100 Scientist of the Year Award in 2006. He won 5 best paper awards and received 5 other professional prizes. He authored more than 200 publications and invented 66 patents. He is the co-editor of 2 books. He is serving as the Associate Editor of 6 journals for the past 10 years.