End-to-End Learning for Image Burst Deblurring

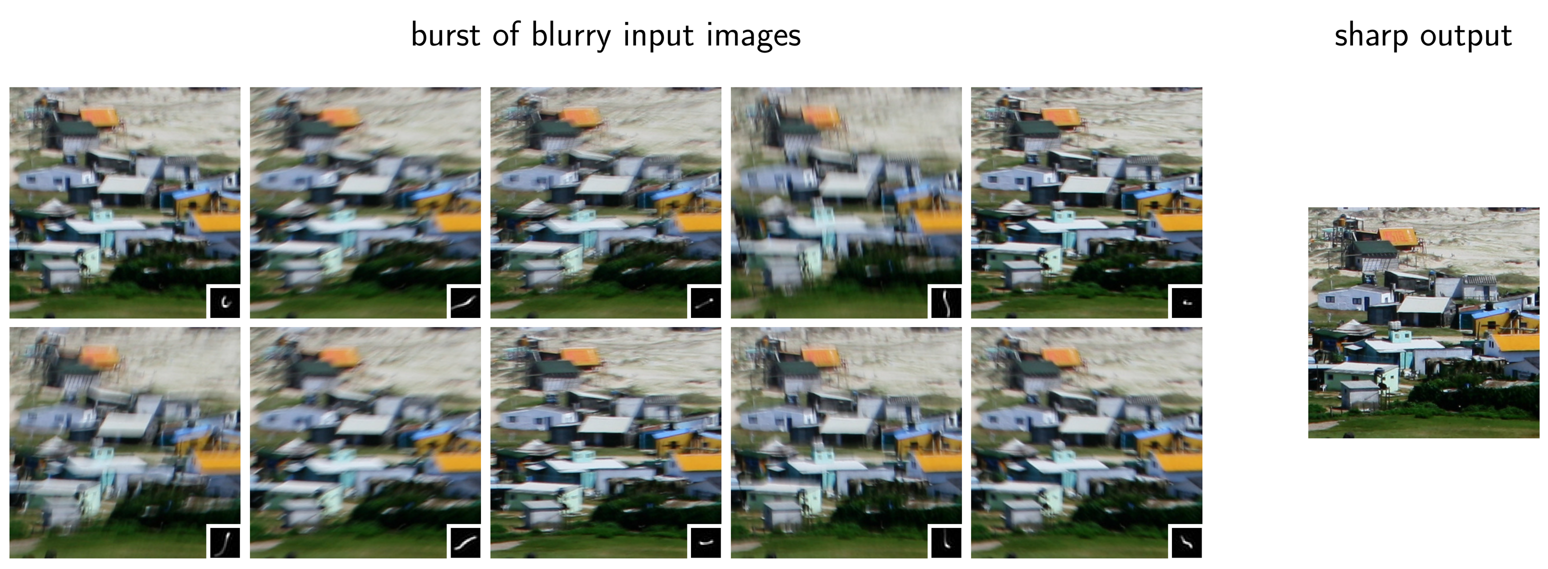
P. Wieschollek[1,2], B. Schölkopf[1], H. P.A. Lensch[2], M. Hirsch[1]

[1] Max Planck Institute for Intelligent Systems, Tübingen, [2] University of Tübingen

In a nutshell: We describe a learning-based approach to multi-frame blind deconvolution using a neural network enabling information sharing across images. Our system is trained end-to-end using standard back-propagation on a set of artificially generated training examples, enabling competitive performance in multi-frame blind deconvolution, both with respect to quality and runtime.

Multi-frame Blind Deblurring Problem

Multi-frame blind deconvolution aims at recovering a sharp image from a series of blurry photos.



Multiple photos of the same scene facilitate the reconstruction of a single sharp image

Given a burst of observed images $Y_1, Y_2, \dots, Y_N \in \mathcal{I}$ capturing the same scene, our image model reads

$$Y_t = k_t * X + \varepsilon_t, \quad \text{for all } t \tag{1}$$

where $*$ denotes the convolution operator, k_t blur kernel for observation Y_t and ε_t additive zero-mean Gaussian noise.

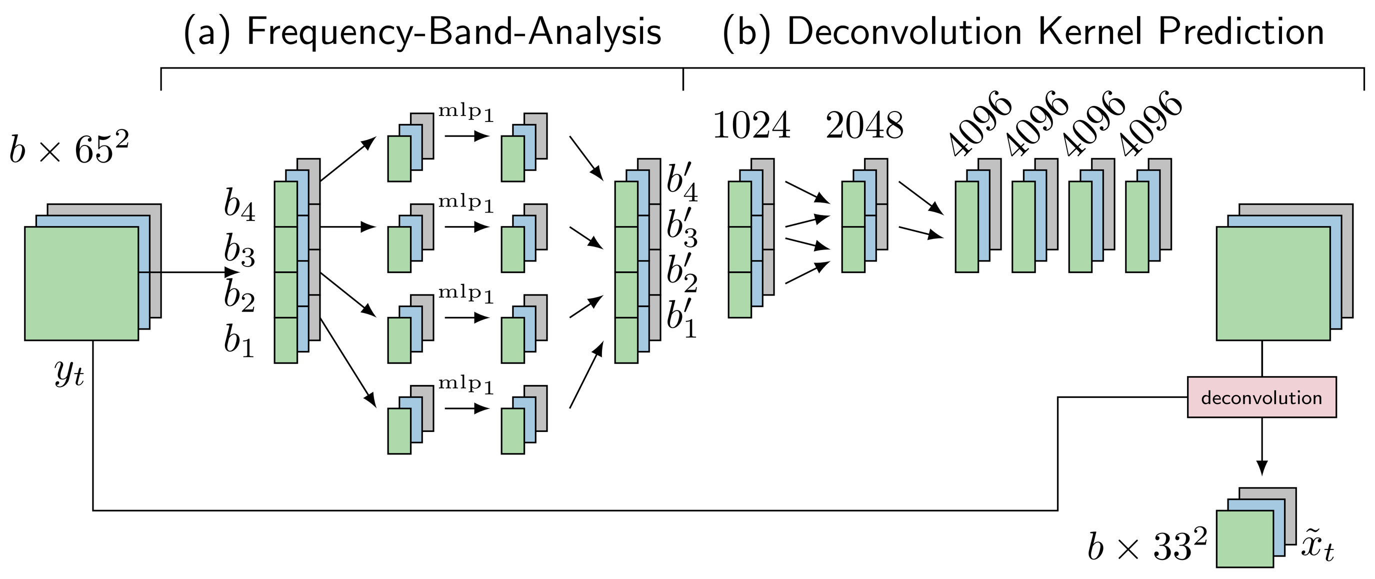
Goal Predict a single sharp image $\hat{X}$ satisfying all constraints in (1).

Network Architecture

Our neural network mapping

$$\pi^{(\theta)} \colon \mathcal{I}_p^k \to \mathcal{I}_p, \qquad (y_1, y_2, \dots, y_k) \mapsto \hat{x} = \pi^{(\theta)}(y_1, y_2, \dots, y_k).$$

operates on overlapping image patches $y_t \in \mathcal{I}_p$ using a sliding window approach. We directly minimize the L2-loss. The architecture $\pi^{(\theta)}(\cdot)$ consists of several stages:



Our network is applied to all image patches from a burst in parallel.
This allows us to share information between different photos of the same scene.

a) Frequency Band Analysis Following the work of Chakrabarti [1] we separate the Fourier spectrum in 4 different bands by computing the discrete Fourier transform of the observed patch at three different sizes ($17 \times 17, 33 \times 33, 65 \times 65$) using different sample sizes in b_1, b_2, b_3. Hereby, band b_4 represents a low-pass band containing all coefficient $\max |z| \leq 4$ from band b_3.
In addition, we allow each band separately to interact across all images in one burst to support early information sharing.

b) Deconvolution Pairwise merging of the bands b_1, b_2, b_3, b_4 from the frequency band analysis (Step a) into b'_1, b'_2, b'_3, b'_4 is performed using fully connected layers with ReLU activation which entails a successive dimensionality reduction. The last layer directly produces a 4225 dimensional prediction of the filter coefficients of the deconvolution kernel, which is then applied to each input patch.

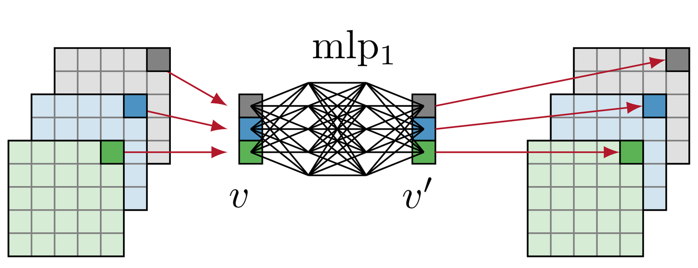
c) Image Fusion Information fusion across entire bursts is done via a modified version of the recently proposed *Fourier-Burst-Accumulation* (FBA) [2] method. The vanilla FBA algorithm applies the following weighted sum to a Fourier transform $\hat{\alpha}$ of the patch α:

$$u(\hat{\alpha}) = \mathcal{F}^{-1}\left(\sum_{i=1}^{k} w_i(\zeta)\hat{\alpha}_i(\zeta)\right)(x), \qquad w_i(\zeta) = \frac{|\hat{\alpha}_i(\zeta)|^p}{\sum_{j=1}^{k} |\hat{\alpha}_j(\zeta)|^p}, \tag{2}$$

where w_i denotes the contribution of frequency ζ of a patch α_i. Contrary to [2] we propose to learn these parameters $w_i(\zeta)$ in a neural network layer by making use of information from all observed images within one burst.

Value-Sharing

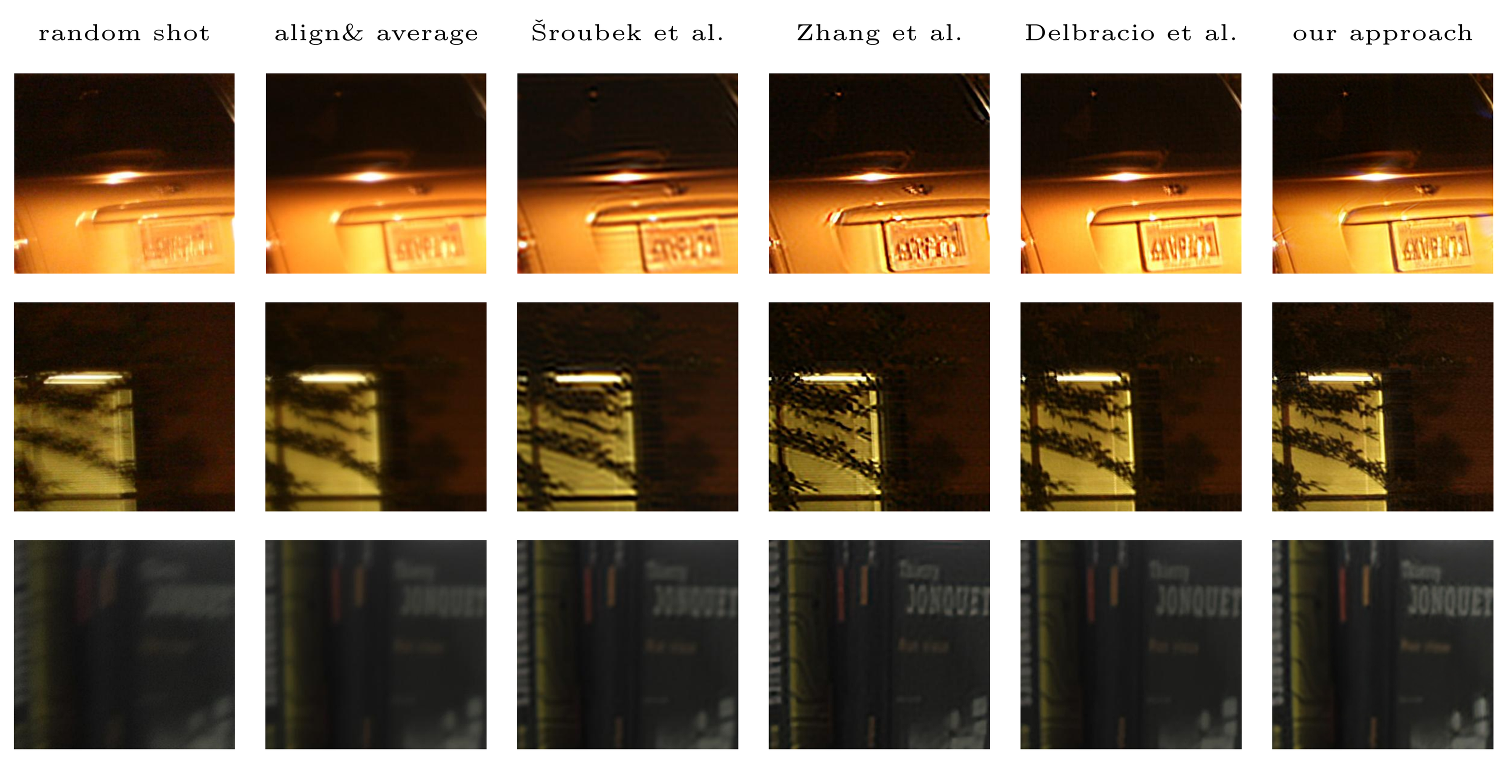
In addition to support weight-sharing across all burst images, our architecture enables spread early intermediate information across the entire burst, which essentially embeds a fully connected neural network for each coefficient $(f_{ij})_t$ with weight sharing.



We interpret each coefficient across one burst as a single vector. A transformed version (multiple 1×1 convolutions) of this excerpt will be placed at the same location in the output patch again.

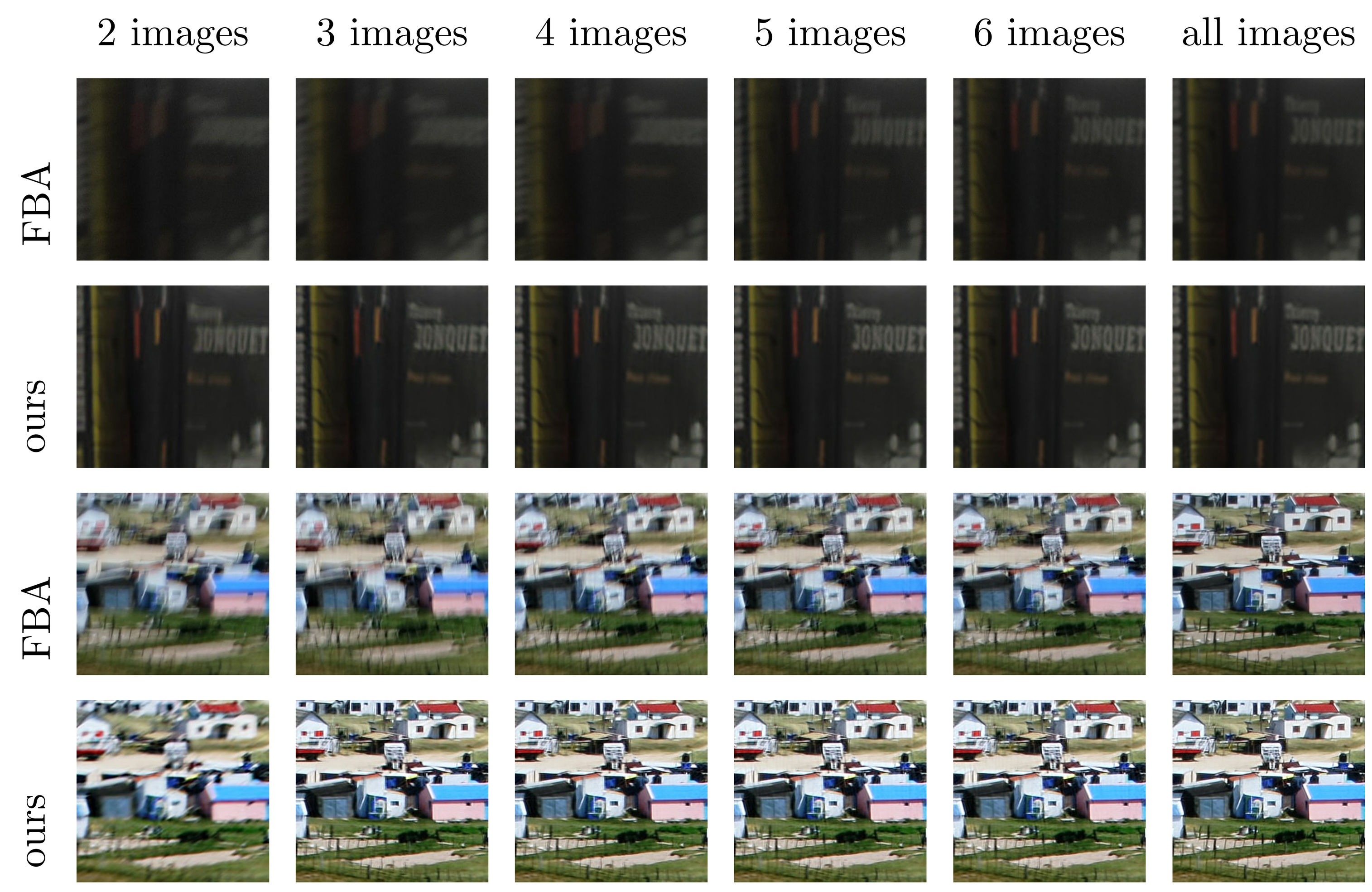
Results

Comparison to state-of-the-art methods We compare the restored images with other state-of-the-art multi-image blind deconvolution algorithms such as the multi-channel blind deconvolution method of Šroubek et al. [4], the sparse-prior method of [5] and the FBA method proposed in [2].



Our trained neural network featuring the FBA-like averaging yields comparable if not superior results compared to previous approaches [4, 5, 2]. In direct comparison to the FBA results, our method is better in removing blur due to our additional prepended deconvolution module.

Effect of our deconvolution module FBA and our algorithm are compared on bursts with increasing number of images of increasing quality. The individual images are sorted according to their PSNR starting with the most blurried images. The blurred input images are taken from [2].



Effect of end-to-end training One might ask, how our trained neural network compares to an approach that applies the methods of Chakrabarti [1] and Delbracio and Sapiro [2] subsequently, each in a separate step. Without end-to-end training ringing-artifacts (left) are clearly visible on the blue roof. They are significantly dampened (right) after training.

Spatially-varying blur Our patch-based approach is able to handle spatially-varying blur. We took recorded camera trajectories and generated input images using the Efficient Filter Flow model of [5]. Our results (right) are consistently sharper than [2] (left) and demonstrate that our approach is also able to correct for spatially-varying blur.

References

[1] Ayan Chakrabarti. *A Neural Approach to Blind Motion Deblurring.*, ECCV 2016.
[2] Delbracio, Mauricio and Sapiro, Guillermo. *Burst deblurring: Removing camera shake through fourier burst accumulation.*, CVPR 2015.
[3] Tsung-Yi Lin et al. *Microsoft COCO: Common Objects in Context*.
[4] Šroubek, Filip and Milanfar, Peyman. *Robust multichannel blind deconvolution via fast alternating minimization*, 2012.
[5] Zhang, Haichao and Wipf, David and Zhang, Yanning. *Multi-image blind deblurring using a coupled adaptive sparse prior*, CVPR, 2013.
[5] Hirsch, Michael and Sra, Suvrit and Schölkopf, Bernhard and Harmeling, Stefan. *Efficient filter flow for space-variant multiframe blind deconvolution*, 2010.

goo.gl/GxxS8R