Convolutional Neural Networks for Direct Text Deblurring

Michal Hradiš[1]
ihradis@fit.vutbr.cz
Jan Kotera[2]
kotera@utia.cas.cz
Pavel Zemčík[1]
zemcik@fit.vutbr.cz
Filip Šroubek[2]
sroubekf@utia.cas.cz

[1] Faculty of Information Technology
Brno University of Technology
Brno, Czech Republic

[2] Institute of Information Theory and Automation
Czech Academy of Sciences
Prague, Czech Republic

Abstract

In this work we address the problem of blind deconvolution and denoising. We focus on restoration of text documents and we show that this type of highly structured data can be successfully restored by a convolutional neural network. The networks are trained to reconstruct high-quality images directly from blurry inputs without assuming any specific blur and noise models. We demonstrate the performance of the convolutional networks on a large set of text documents and on a combination of realistic de-focus and camera shake blur kernels. On this artificial data, the convolutional networks significantly outperform existing blind deconvolution methods, including those optimized for text, in terms of image quality and OCR accuracy. In fact, the networks outperform even state-of-the-art non-blind methods for anything but the lowest noise levels. The approach is validated on real photos taken by various devices.

1 Introduction

Taking pictures of text documents using hand-held cameras has become commonplace in casual situations such as when digitizing receipts, hand-written notes, and public information signboards. However, the resulting images are often degraded due to camera shake, improper focus, image noise, image compression, low resolution, poor lighting, or reflections. We have selected the restoration of such images as a representative of tasks for which current blind deconvolution methods are not particularly suited for. They are not equipped to model some of the image degradations, and they do not take full advantage of the available knowledge of the image content. Moreover, text contains high amount of small details which needs to be preserved to retain legibility. We propose a deblurring method based on a convolutional network which learns to restore images from data.

In its idealized form, blind deconvolution is defined as the task of finding an original image x, and possibly a convolution kernel k, from an observed image y which is created as

$$y = x * k + n, \tag{1}$$

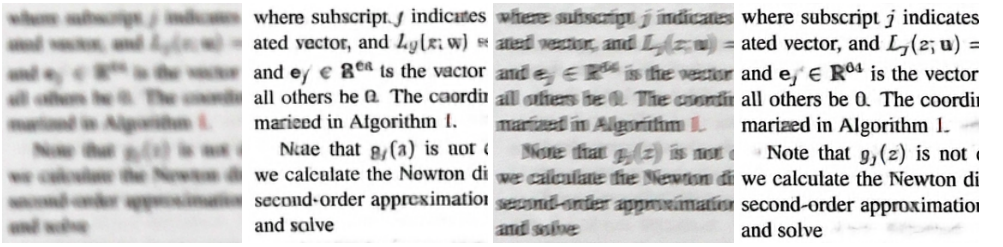


Figure 1: Real photos deblured with a convolutional neural network.

where n is an independent additive noise. This is an ill-posed problem with infinite number of solutions. The problem may remain ill-posed even in the absence of noise and when the blur kernel is known (non-blind deconvolution). Fortunately, real images are not completely random and the knowledge of their statistics can be used to constrain the solution.

A possible solution to the deconvolution problem is a minimization of a suitable cost function, *e.g.*

$$\hat{x} = \arg\min_x \|y - x * k\|_2^2 + r(x). \tag{2}$$

Here the data term $\|y - x * k\|_2^2$ forces the solution $\hat{x}$ to be a plausible source of the observed image. The regularizing prior term $r(x)$ expresses knowledge of the image statistics and forces the solution to be a *"nice"* image. Unfortunately, the process of capturing images with real cameras is much more complex than Eq. (1) suggests. Atmosphere scatters incoming light (haze, fog). Instead of being space-invariant, real blur depends on the position within the image, on object distance, on object motion, and it interacts with chromatic aberrations of the lens. The linear convolution assumption does not hold either. The image can be saturated, cameras apply color and gamma correction, and colors are interpolated from neighboring pixels due to Bayer mask. Many cameras even apply denoising and sharpening filters and compress the image with lossy JPEG compression.

Some of the aspects of the real imaging process can be incorporated into Eq. (2), for example by modification of the data term, but doing so increases computational complexity and is often too difficult (*e.g.* space-variant blur [32], saturation [9], or non-Gaussian noise [4]). It is sometimes not practical to model the imaging process fully and the unmodeled phenomena is left to be handled as a noise.

In our work we propose an alternative approach to deconvolution by directly modeling the restoration process as a general function

$$\hat{x} = F(y, \theta) \tag{3}$$

with parameters θ which can be learned from data. The function $F(y, \theta)$ implicitly incorporates both the data term and the prior term from Eq. (2). The advantage of this data-driven approach is that it is relatively easy to model the forward imaging process and to generate (x, y) pairs. In many applications, it is even possible to get the (x, y) pairs from a real imaging process. Consequently, learning the restoration function $F(y, \theta)$ can be straightforward even for a complex imaging system, as opposed to the extremely hard task of inverting the imaging process by hand.

Convolutional neural networks (CNN) are the state-of-the art function approximators for computer vision problems and are well suited for the deconvolution task – they can naturally

incorporate inverse filters and, we believe, they posses enough computational capacity to model the blind deconvolution process, including strong data priors.

We investigate whether CNNs are able to learn end-to-end mapping for deblurring of text documents. Text is highly structured and the strong prior should allow to deconvolve images locally even in blind setting without the need of extracting information from a large image neighborhood.

CNNs were previously used to learn denoising [16], structured noise removal [11], non-blind deconvolution [26, 35], and sub-tasks of blind deconvolution [27]. Our work is the first to demonstrate that CNNs are able to perform state-of-the-art blind deconvolution (see Figure 1). Locality, data-driven nature, and the ability to capture strong priors makes the approach robust, efficient, easy to use, and easy to adapt to new data.

2 Related work

The key idea of general modern blind deblurring methods is to address the ill-posedness of blind deconvolution by a suitable choice of prior (which then forms the regularizer r in Eq. (2)), for example by using natural image statistics, or by otherwise modifying either the minimized functional or the optimization process. This started with the work of Fergus *et al.* [12], who applied variational Bayes to approximate the posterior. Other authors (e.g. [1, 2, 18, 20, 28, 33, 34]) maximize the posterior by an alternating blur-image approach and, in some cases, use rather ad hoc steps to obtain an acceptable solution. Levin *et al.* [22, 23] showed that marginalizing the posterior with respect to the latent image leads to the correct solution of the PSF, while correct prior alone might not – therefore these ad hoc steps are in fact often crucial for some deconvolution methods to work.

Space-variant blind deconvolution is even more challenging problem, as the PSF also depends on the position and the problem thus has much more unknowns. In such case, the space of possible blurs is often limited, for example to camera rotations. The blur operator can be then expressed as a linear combination of a small number of base blurs, and the blind problem is solved in the space spanned by such basis [15, 19, 32]. Successful image deblurring is further complicated by e.g. saturated pixels [9], the problem of unknown image boundary in convolution [1], non-linear post-processing by the camera, and many more. In short, modeling the whole process is, while perhaps possible, simply too complicated and even state-of-the-art deblurring methods do not attempt to include all degradation factors at once.

Images of text are markedly different from images of natural scenes. One of the first methods specialized for text-image deblurring was [25] where Panci *et al.* modeled text image as a random field and used the Bussgang algorithm to recover the sharp image. Cho *et al.* [7] segment the image into background and characters, for which they use different prior based on properties characteristic for images of text. Even more recent and arguably state-of-the-art approach is the method of Pan *et al.* [24] which uses sparse l0 prior on image gradients and on pixel intensities. Otherwise, these methods follow the established pipeline of first estimating the blur kernel and then deconvolving the blurred image using a non-blind method, which includes all the complications mentioned before.

Neural networks and other learning methods have been used extensively in image restoration. The most relevant to our work are methods witch use CNNs to directly predict high-quality images (as in Eq. (3)). Xu et al. [35] learn CNNs for space-invariant non-blind deconvolution. They initialize first two layers with a separable decomposition of an inverse

filter and then they optimize the full network on artificial data. This network can handle complex blurs and saturation, but it has to be completely re-trained for each blur kernel. Jain and Seung [16] learned a small CNN (15,697 parameters) to remove additive white Gaussian noise with unknown energy. Eigen et al. [11] detect and remove rain drops and dirt by CNN learned on generated data. They report significant quality increase with CNN over a patch-based network. Dong et al. [10] learn a small 3-layer CNN with Rectified Linear Units for state-of-the-art single image super resolution. Interesting and related to our work is the blind deconvolution method by Schuler et al. [27]. They propose to learn a "sharpenning" CNN for blur kernel estimation. In contrast to our work, their network is rather small and they reconstruct the final image with a standard non-blind method.

3 Convolutional networks for blind deconvoution

We directly predict clean and sharp images from corrupted observed images by a convolutional network as in Eq. (3). The architecture of the networks is inspired by the recently very successful networks that redefined state-of-the-art in many computer vision tasks including object and scene classification [29, 30, 37], object detection [13], and facial recognition [31]. All these networks are derived from the ImageNet classification network by Krizhevsky et al. [21]. These networks can be reliably trained even when they contain hundreds of millions weights [29] and tens of layers [30].

The networks are composed of multiple layers which combine convolutions with element-wise Rectified Liear Units (ReLU):

$$\begin{aligned} F_0(y) &= y \\ F_l(y) &= \max(0, W_l * F_{l-1}(y) + b_l), l = 1, \ldots, L-1 \\ F(y) &= W_L * F_{L-1}(y) + b_L \end{aligned} \tag{4}$$

The input and output are both 3-channel RGB images with their values mapped to interval $[-0.5, 0.5]$. Each layer applies c_l convolutions with filters spanning all channels c_{l-1} of the previous layer. The last layer is linear (without ReLU).

As in previous works [3, 11, 26, 35], we train the networks by minimizing mean squared error on a dataset $D = (x_i, y_i)$ of corresponding clean and corrupted image patches:

$$\arg\min_{W,b} \frac{1}{2|D|} \sum_{(x_i,y_i)\in D} ||F(y_i) - x_i||_2^2 + 0.0005||W||_2^2 \tag{5}$$

The weight decay term $0.0005||W||_2^2$ is not required for regularization, but previous work showed that it improves convergence [21]. The optimization method we use is Stochastic Gradient Descent with momentum. We set the size of clean patches x_i to 16×16 pixels, which we believe provides good trade-off between computational efficiency and diversity of mini-batches (96 patches in each mini-batch). Size of the blurred training patches y_i depends on the spatial support of a particular network.

We initialize weights from uniform distribution with variance equal to $\frac{1}{n_{in}}$, where n_{in} is the size of the respective convolutional filter (fan-in). This recipe was derived by authors of Caffe framework [17] from recommendations by Xavier and Bengio [14].

Table 1: CNN architecture – filter size and number of channels for each layer.

Layer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
L15	19×19	1×1	1×1	1×1	1×1	3×3	1×1	5×5	5×5	3×3	5×5	5×5	1×1	7×7	7×7
	128	320	320	320	128	128	512	128	128	128	128	128	256	64	3
L10	23×23	1×1	1×1	1×1	1×1	3×3	1×1	5×5	3×3	5×5					
S	128	320	320	320	128	128	512	48	96	3					
M	196	400	400	400	156	156	512	56	128	3					
L	220	512	512	512	196	196	512	64	196	3					

4 Results

We tested the approach on the task of blind deconvolution with realistic de-focus and camera-shake blur kernels on a large set of documents from the CiteSeerX[1] repository. We explored different network architecture choices, and we compared results to state-of-the-art blind and non-blind deconvolution methods in terms of image quality and OCR accuracy. We purposely limited the image degradations to shift-invariant blur and additive noise to allow for fair comparison with the baseline methods, which are not designed to handle other aspects of the image acquisition process. To validate our approach, we qualitatively evaluated the created networks on real photos of printed documents. The experimentes were conducted using a modified version of Caffe deep learning framework [17].

Dataset. We selected scientific publications for the experiments as they contain an interesting mix of different content types (text, equations, tables, images, graphs, and diagrams). We downloaded over 600k documents from CiteSeerX repository from which we randomly selected 50k files for training and 2k files for validation. We rendered at most 4 random pages from each document using Ghostscript with resolution uniformly sampled form 240-300 DPI. We down-sampled the images by factor of two to a final resolution 120-150 DPI (A4 paper at 150 DPI requires 2.2 Mpx resolution). Patches for training (3M) and validation (35k) were sampled from the rendered pages with a preference to regions with higher total variation.

To make the extracted patches more realistic, we applied small geometric transformations with bicubic interpolation corresponding to camera rotations in the document plane and deviations from the perpendicular document axis (normal distributions with standard deviation of 1° and 4°, respectively). We believe that this amount of variation may correspond to that of automatically rectified images.

We combined two types of blur – motion blur similar to camera shake and de-focus blur. The de-focus blur is an uniform anti-aliased disc. The motion blur was generated by a random walk. The radius of the de-focus blur was uniformly sampled from $[0,4]$ and the maximum size of the motion kernel was sampled from $[5,21]$. Each image patch was blurred with a uniqe kernel. A histogram of kernel sizes is shown in Figure 2 (bottom-right). Gaussian noise with standard deviation uniformly sampled from $[0,\frac{7}{255}]$ was added to the blurred patches which were finally quantized into 256 discrete levels.

CNN architecture. Bigger and deeper networks give better results in computer vision tasks [30] provided enough training data and computing power is available. Figure 2 (top-

[1] http://citeseerx.ist.psu.edu/

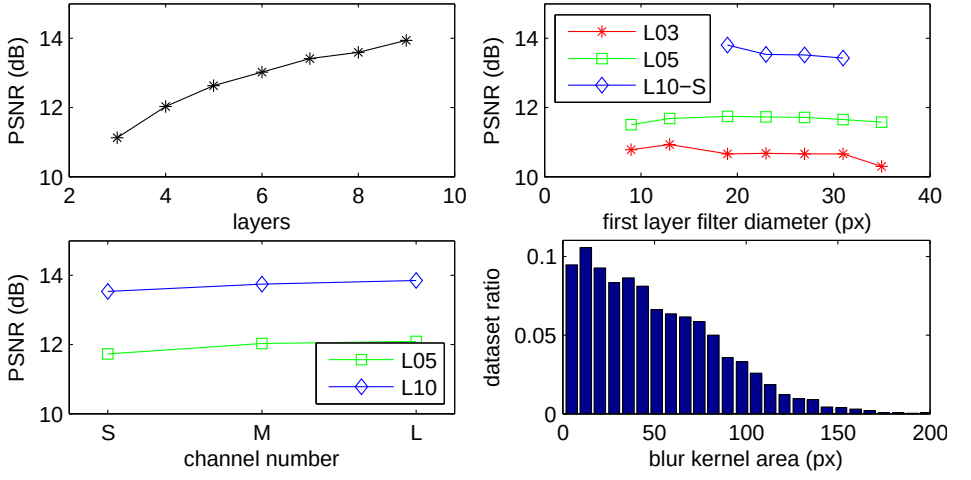


Figure 2: Different CNN architectures. top-left – network depth; top-right – spatial support size; bottom-left – channel number; bottom-right – distribution of blur-kernel sizes in dataset

left) shows Peak Signal to Noise Ration (PSNR) on the validation set for networks with different number of layers (128 filters in each layer; size of first layer filters 23×23, other layers 3×3). The deeper networks consistently outperform shallower networks.

The better results of deeper networks could be due to their higher computational capacity or due to larger spatial support (2px per layer). To gain more insight, we trained multiple networks with filter sizes in the first layer from 9×9 up to 35×35. These networks have 128 filters in the first layer, 3 filters 5×5 in the last layer (RGB), and 256 filters 1×1 in the middle layers. These networks with three layers will be refered to as L03 and those with five layers as L05. Figure 2 (top-right) shows that the spatial support affects performance insignificantly compared the depth of the network. The optimal size of the first layer filters is relatively small: 13×13 for L03 and 19×19 for L05. The same observations hold for a 10 layer network (L10-S from Table 1).

Another way to enlarge a network is to increase the number of channels (filters). Such change affects the amount of information which can pass through the network at the expense of quadratic increase of the number of weights and of computational complexity. Figure 2 (bottom-left) shows the effect of changing number of channels for networks L10 (see Table 1) and L05 (first layer channels 128 (S), 196 (M), 220 (L); other layers 256 (S), 320 (M), 512 (L)). The reconstruction quality increases only slightly with higher number of channels, while the effect of network depth is much stronger.

The largest and deepest network we trained has 15 layers and 2.3M parameters (L15 from Table 1). This network provides the best results (validation PSNR 16.06 dB) still without over-fitting. Compared to convolutional networks used in computer vision tasks, this network is still small and computationally efficient – it requires 2.3M multiply-accumulate operations per pixel. Assuming 50% efficiency, contemporary GPUs, which provide peak single-precision speeds exceeding 4 Tflops, should be able to process a 1Mpx image in 2s. The network was trained for more than a weak on a single GPU.

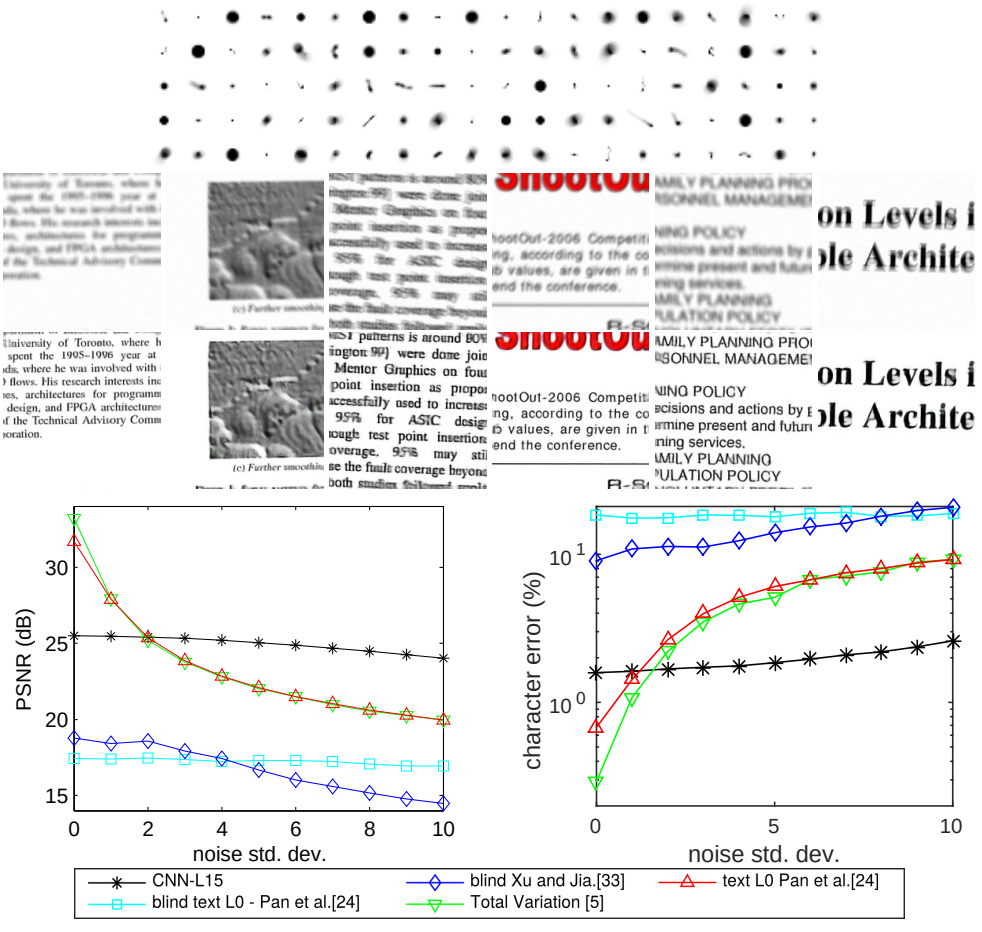


Figure 3: Artificial dataset and results. top – filters from test set; middle - examples from test set with L15 CNN deblurring resuts; bottom-left – deblurring image quality; bottom-right – OCR error on deblurred data

Image quality We evaluated image restoration quality on one hundred 200×200 image patches which were prepared the same way as the training and the validation set. These patches were extracted from unseen documents and the preference for regions with higher total variation was relaxed to only avoid empty regions. A random subset of the regions together with all 100 blur filters is shown in Figure 3. PSNR was measured only on the central 160×160 region.

We selected 4 baselines. Two blind deconvolution methods – the standard method of Xu and Jia [33], and the L0 regularized method by Pan et al. [24] which is especially designed for text. In addition, we selected two non-blind methods as an upper bound on the possible quality of the blind methods – total variation regularized deconvolution [5], and L0 regularized deconvolution which is a part of the blind method of Pan et al. [24]. Optimal parameters of all four baseline methods were selected by grid search for each noise level separately on an independent set of 25 images.

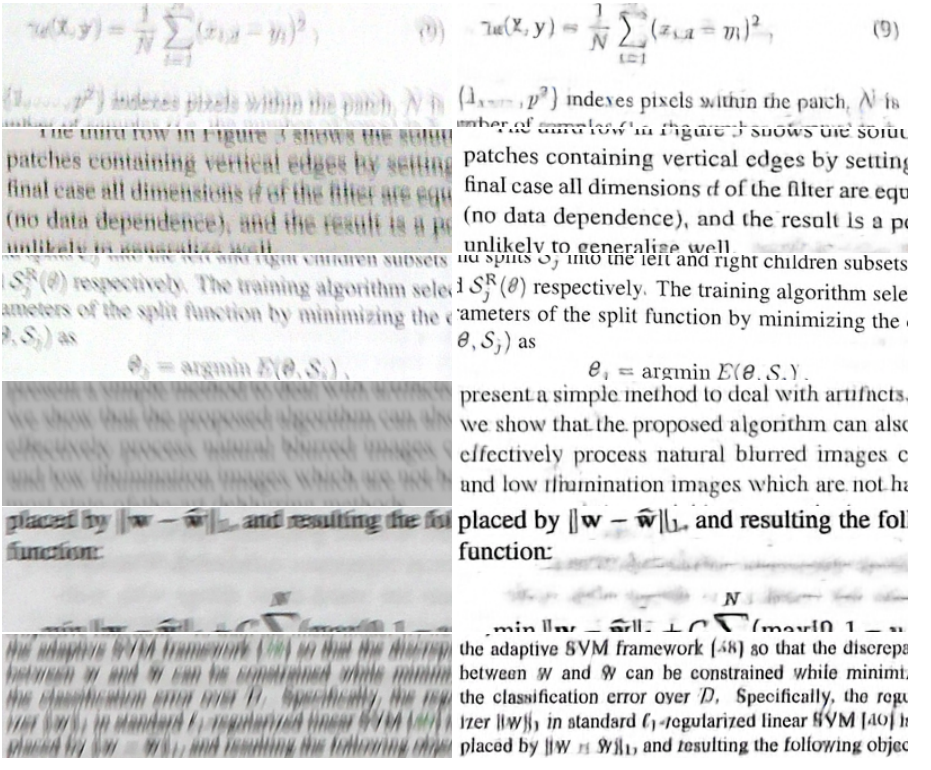


Figure 4: CNN deblurring of challenging real images. More in supplementary material.

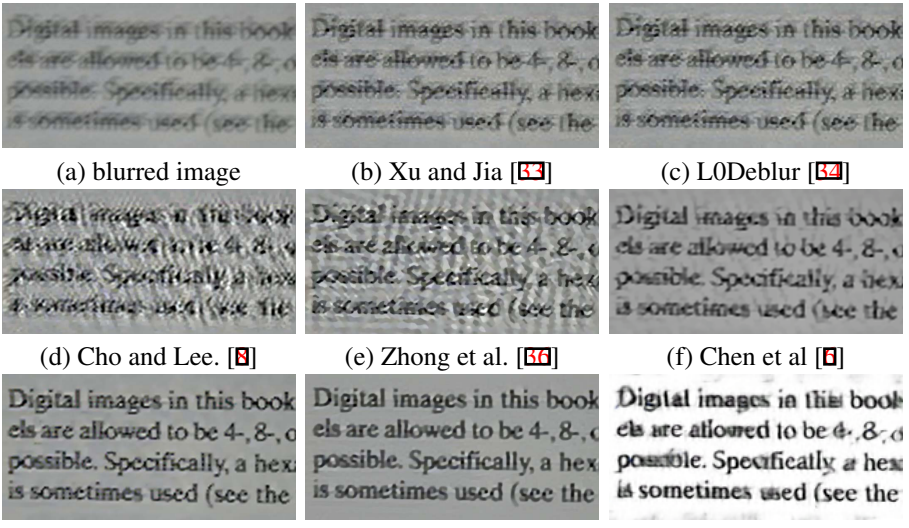
Figure 3 (bottom-left) shows results of all methods for different amounts of noise. L15 CNN clearly outperforms both reference blind methods for all noise levels. The non-blind methods are better at very low noise levels, but their results rapidly degrade with stronger noise – L15 CNN is better for noise with std. dev. 3 and stronger. Surprisingly, the network maintains good restoration quality even for noise levels higher than for what it was trained for.

Optical Character Recognition. One of the main reasons for text deblurring is to improve legibility and OCR accuracy. We evaluated OCR accuracy on 100 manually cropped paragraphs which ware blurred with the same kernels as the test set. The manual cropping was necessary as OCR software uses language models and needs continuous text for realistic performance. The 100 paragraphs were selected with no conscious bias except that width of the selected paragraphs had to be less than 512px. The paragraphs contain in total 4730 words and 25870 chracters of various sizes and fonts[2]. We used ABBYY FineReader 11 to recognize the text and we report mean Character Error Rate[3] after Levenshtein alignment.

Figure 3 (bottom-right) shows results for different noise levels. The OCR results are very similar to the deblurring image quality. Non-blind methods are the best for very low amount of noise, but L15 CNN becomes better for noise level 3 and higher. The blind methods fail

[2] See supplementary material for examples.

[3] Character Error Rate can rise over 100% – we clipped errors at 100% when computing the mean.



(a) blurred image (b) Xu and Jia [33] (c) L0Deblur [34]

(d) Cho and Lee. [8] (e) Zhong et al. [36] (f) Chen et al [6]

(g) Cho et al. [7] (h) Pan et al. [24] (i) CNN deblurring

Figure 5: A real image from [7]. The reference results are from [24]. The input of L15 CNN was downsampled by factor of 3 compared to the other methods.

on many images and the respective mean results are therefore quite low.

Real images To create networks capable of restoring real images, we incorporated additional transformations into the simulation of the imaging process when creating the blurred training data. The additional transformations are: color balance, gamma and contrast transformations, and JPEG compression. This way the network effectively learns to reconstruct the documents with perfect black and white levels. Figure 4 shows that the network was able to restore even heavily distorted photos with high amount of noise. The presented examples are at the limits of the CNN and reconstruction artifacts are visible. The network is not able to handle blur larger than what it was trained for – in such cases the reconstruction fails completely. On the other hand, it is able to handle much stronger noise and even effects like saturation for which it was not trained. Figure 5 shows comparison with other methods on an image by Cho et al. [7].

Discussion The results are extremely promising especially considering that the presented approach naturally handles space-variant blur and that it processes a 1Mpx image in 5s even in our non-optimized implementation. L15 CNN is small (9MB) and it would easily fit into a mobile device.

To some extent, the network performs general deconvolution – it sharpens and denoises images in documents and it produces typical ringing artefacts in some situations. However, much of the observed performance is due to its ability to model the text prior which has to be strong enough to disambiguate very small image patches (spatial support of L15 CNN is only 50×50px). We observed better reconstruction quality for common words (e.g. "the", "and") and relatively worse results for rare characters (e.g. calligraphic fonts, greek alphabet). However, these effects are evident only for severely blurred inputs.

For computational reasons, we considered only relatively small-resolution images with

correspondingly small characters and blur kernels. With the simple CNN structure we use, it would be inefficient to process higher-resolution images – a more complex structure (e.g. the "Inception module" [30]) would be needed.

5 Conclusions

We have demonstrated that convolutional neural networks are able to directly perform blind deconvolution of documents, and that they achieve restoration quality surpassing state-of-the-art methods. The proposed method is computationally efficient and it naturally handles space-variant blur, JPEG compression, and other aspects of imaging process which are problematic for traditional deconvolution methods.

In the future, we intend to broaden the domain on which CNN deblurring can be applied, and we would like to increase the complexity of the network structure to make it more efficient. A natural extension of the presented approach is to combine the deblurring with character recognition into a single network. Character labels could help to learn reconstruction and vice versa. Alternatively, the deblurring task could be used to pre-train an OCR network.

Acknowledgements

This research is supported by the ARTEMIS joint undertaking under grant agreement no 641439 (ALMARVI), GACR agency project GA13-29225S. Jan Kotera was also supported by GA UK project 938213/2013, Faculty of Mathematics and Physics, Charles University in Prague.

References

[1] Mariana S.C. Almeida and Mário A.T. Figueiredo. Blind image deblurring with unknown boundaries using the alternating direction method of multipliers. In *ICIP*, pages 586–590. Citeseer, 2013.

[2] M.S.C. Almeida and L.B. Almeida. Blind and semi-blind deblurring of natural images. *IEEE Transactions on Image Processing*, 19(1):36–52, January 2010. ISSN 1057-7149, 1941-0042. doi: 10.1109/TIP.2009.2031231.

[3] Harold C. Burger, Christian J. Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with BM3D? In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2392–2399, 2012. ISBN 9781467312264. doi: 10.1109/CVPR.2012.6247952.

[4] M. Carlavan and L. Blanc-Feraud. Sparse poisson noisy image deblurring. *Image Processing, IEEE Transactions on*, 21(4):1834–1846, April 2012. ISSN 1057-7149. doi: 10.1109/TIP.2011.2175934.

[5] Stanley H. Chan, Ramsin Khoshabeh, Kristofor B. Gibson, Philip E. Gill, and Truong Q. Nguyen. An augmented Lagrangian method for total variation video restoration. *IEEE Transactions on Image Processing*, 20(11):3097–3111, 2011. ISSN 10577149. doi: 10.1109/TIP.2011.2158229.

[6] Xiaogang Chen, Xiangjian He, Jie Yang, and Qiang Wu. An effective document image deblurring algorithm. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 369–376, 2011. ISBN 9781457703942. doi: 10.1109/CVPR.2011.5995568.

[7] Hojin Cho, Jue Wang, and Seungyong Lee. Text image deblurring using text-specific properties. In *ECCV*, volume 7576 LNCS, pages 524–537, 2012. ISBN 9783642337147. doi: 10.1007/978-3-642-33715-4_38.

[8] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. *ACM Transactions on Graphics*, 28(5):1, 2009. ISSN 07300301. doi: 10.1145/1618452.1618491.

[9] Sunghyun Cho, Jue Wang, and Seungyong Lee. Handling outliers in non-blind image deconvolution. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 495–502, Nov 2011. doi: 10.1109/ICCV.2011.6126280.

[10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a Deep Convolutional Network for Image Super-Resolution. In *ECCV*, pages 184–199. Springer International Publishing, 2014.

[11] David Eigen, DIlip Krishnan, and Rob Fergus. Restoring an Image Taken through a Window Covered with Dirt or Rain. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 633–640, 2013. ISBN 978-1-4799-2840-8. doi: 10.1109/ICCV.2013.84.

[12] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. Removing camera shake from a single photograph. *ACM Transactions on Graphics (TOG)*, 25(3):787–794, 2006.

[13] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Computer Vision and Pattern Recognition*, 2014.

[14] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATSâĂŹ10). Society for Artificial Intelligence and Statistics*, 2010.

[15] Ankit Gupta, Neel Joshi, C. Lawrence Zitnick, Michael Cohen, and Brian Curless. Single image deblurring using motion density functions. In *Computer Vision–ECCV 2010*, pages 171–184. Springer, 2010.

[16] Viren Jain and Sebastian Seung. Natural Image Denoising with Convolutional Networks. In D Koller, D Schuurmans, Y Bengio, and L Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 769–776. Curran Associates, Inc., 2009.

[17] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[18] Neel Joshi, Richard Szeliski, and David Kriegman. PSF estimation using sharp edge prediction. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[19] Neel Joshi, Sing Bing Kang, C. Lawrence Zitnick, and Richard Szeliski. Image deblurring using inertial measurement sensors. In *ACM SIGGRAPH 2010 Papers*, SIGGRAPH '10, pages 30:1–30:9, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0210-4. doi: 10.1145/1833349.1778767.

[20] Dilip Krishnan, Terence Tay, and Rob Fergus. Blind deconvolution using a normalized sparsity measure. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 233–240. IEEE, 2011.

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9, 2012.

[22] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding blind deconvolution algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12): 2354–2367, December 2011. ISSN 0162-8828, 2160-9292. doi: 10.1109/TPAMI.2011.148.

[23] Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1964–1971. IEEE, 2009.

[24] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. Deblurring Text Images via L0-Regularized Intensity and Gradient Prior. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2901–2908. IEEE, June 2014. ISBN 978-1-4799-5118-5. doi: 10.1109/CVPR.2014.371.

[25] G. Panci, P. Campisi, S. Colonnese, and G. Scarano. Multichannel blind image deconvolution using the bussgang algorithm: Spatial and multiresolution approaches. *IEEE Transactions on Image Processing*, 12(11):1324–1337, November 2003. ISSN 1057-7149. doi: 10.1109/TIP.2003.818022.

[26] Christian J. Schuler, Harold Chrisopher Burger, Stefan Harmeling, and Bernhard Scholkopf. A Machine Learning Approach for Non-blind Image Deconvolution. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1067–1074, 2013. ISBN 1063-6919 VO -. doi: 10.1109/CVPR.2013.142.

[27] Christian J Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. Learning to Deblur. *CoRR*, abs/1406.7, 2014.

[28] Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. In *ACM Transactions on Graphics (TOG)*, volume 27, page 73. ACM, 2008.

[29] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1, 2014.

[30] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going Deeper with Convolutions. *CoRR*, abs/1409.4, 2014.

[31] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708. IEEE, June 2014. ISBN 978-1-4799-5118-5. doi: 10.1109/CVPR.2014.220.

[32] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 491–498, June 2010. doi: 10.1109/CVPR.2010.5540175.

[33] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, volume 6311 LNCS, pages 157–170, 2010. ISBN 3642155480. doi: 10.1007/978-3-642-15549-9_12.

[34] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural L0 sparse representation for natural image deblurring. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1107–1114, 2013. ISBN 978-0-7695-4989-7. doi: 10.1109/CVPR.2013.147.

[35] Li Xu, Jimmy SJ. Ren, Ce Liu, and Jiaya Jia. Deep Convolutional Neural Network for Image Deconvolution. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.

[36] Lin Zhong, Sunghyun Cho, Dimitris Metaxas, Sylvain Paris, and Jue Wang. Handling noise in single image deblurring using directional filters. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 612–619, 2013. ISBN 978-0-7695-4989-7. doi: 10.1109/CVPR.2013.85.

[37] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning Deep Features for Scene Recognition using Places Database. *NIPS*, 2014.