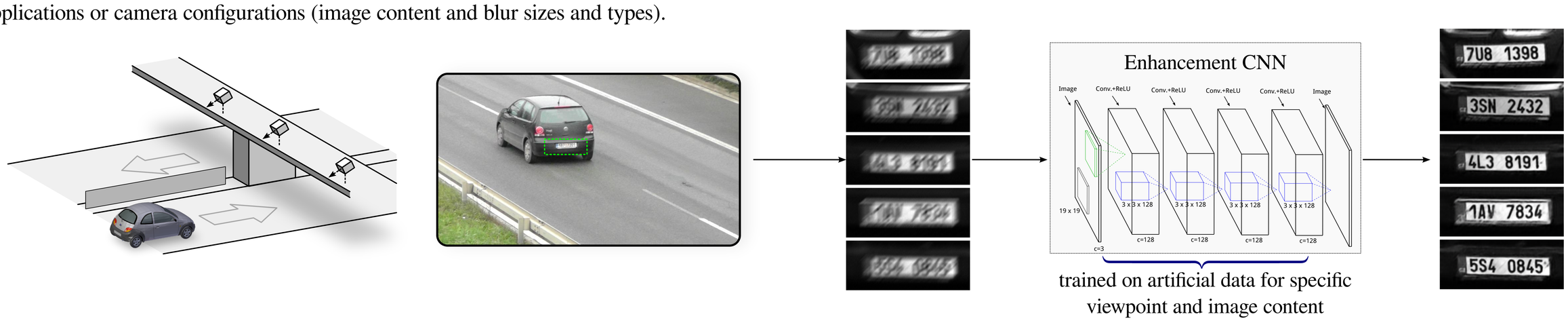
CNN FOR LICENSE PLATE MOTION DEBLURRING

Pavel Svoboda, Michal Hradiš, Lukáš Maršík, Pavel Zemčík
Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic

BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

We explore direct blind deconvolution with convolutional neural networks on images from a real-life traffic surveillance system, where the blur kernels are partially constrained. The neural networks trained on artificial generated data provide superior reconstruction quality compared to traditional blind deconvolution methods. Custom CNNs can be easily trained for other specific applications or camera configurations (image content and blur sizes and types).



1. Image restoration

Degradation - linear blur and noise
$y = x * k + n$

Inverse problem - MAP estimate?
$p(x, k|y) \propto p(y|x, k)p(x)p(k)$

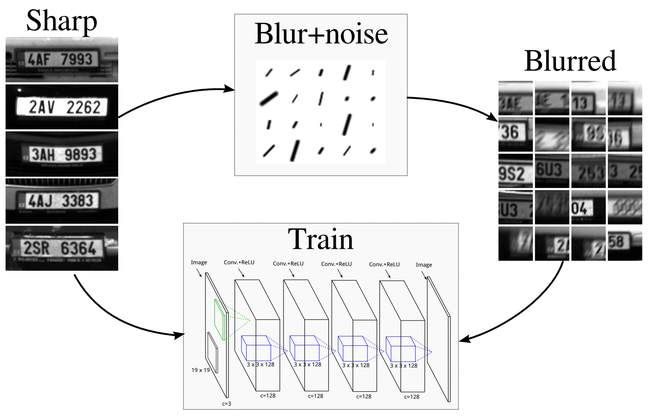
Not convex, too simple image priors.
$(\hat{x}, \hat{k}) = \text{argmin}||x * k - y||^2 + |g_x(x)|^\alpha + |g_y(x)|^\alpha$

How to handle saturation, non-uniform blur, compression, non-linearities, non-gaussian noise?

Our approach (Who needs k?)
$\hat{x} = \mathrm{E}[p(x|y)] \qquad \hat{x} = \mathrm{CNN}(y)$

2. Training data

It is relatively easy to generate artificially degraded images.



We generate data for a specific camera - blur lengths and orietations, license plate orientations and scales.

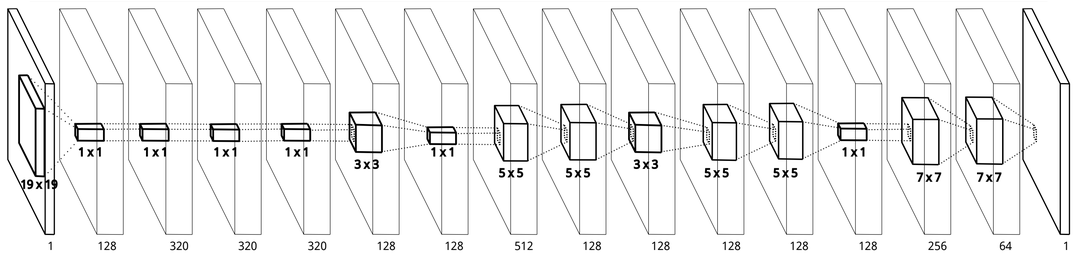
It is easy to include complex degradations including non-linear sharpenning filters and compression.

3. Architecture

The same as for text deblurring (Hradiš et al.; 2015)

- 15 convolution+ReLU layers (last is linear)
- Inputs and outputs are RGB images
- Outputs MAP estimates of the original images
- No padding, crops 25px borders
- 2M weights (9 MB), 2Mflops per pixel
- Processes 1Mpx image in ~4s on GTX 780
- One LP in ~0.2s

Layer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Filter size	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
Channel count	128	320	320	320	128	128	512	128	128	128	128	128	256	64	3



4. Training

- Stochastic gradient descent with momentum
- L2 loss as objective function
- Initialization with uniform distribution
- 400K iterations
- Minibatch 54 crops 66x66 (output 16x16)
- 3 days on GTX 980

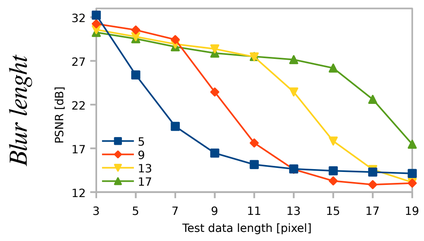
Results

Two static cameras
Blur directions 37°-57° and 59°-79°

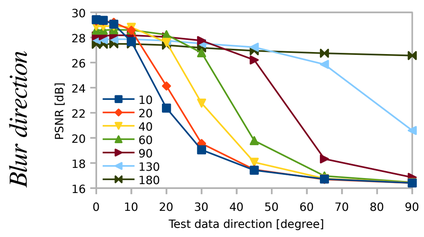
LP crops 264x128px
140K sharp training/validation images

Motion blur range (lenght-direction)

Networks were trained and tested for different blur lengths

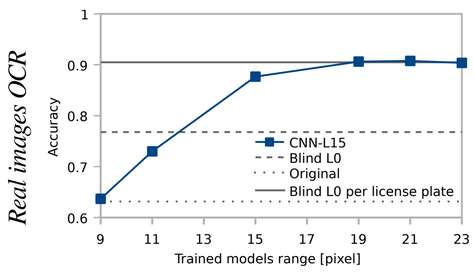


Networks were trained and tested for different blur orientation ranges

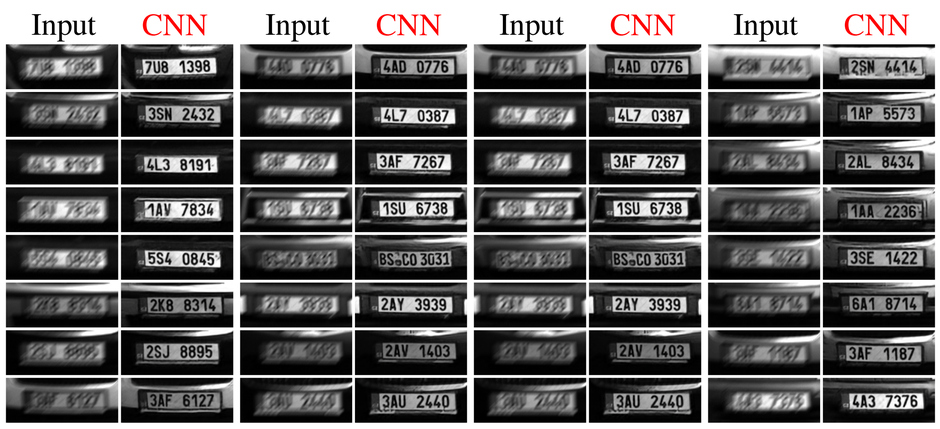


Evaluation on real images

- 721 images, longer exposure (6-12ms)
- Manually annotated LP characters
- Evaluating OCR accuracy
- Existing LP-specific OCR engine
- Baseline blind L0-regularized deconvolution (Pan et al.; 2014)
- Blur orientation range 50°

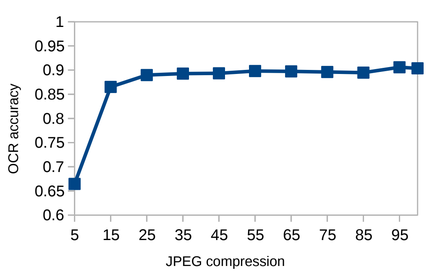


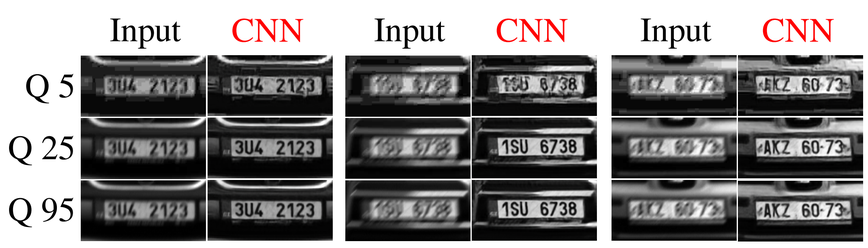
- **Error improved from 37% to 9%**
- L0 deconvolution only imporoved character error down to 23%



Robustness to JPEG compression

We tested a network trained for motion blur on images with additional JPEG compression to asses robustness to additional degradations. The network was able to maintain OCR accuracy down to quality 25.





References

- Michal Hradiš, Jan Kotera, Pavel Zemčík, and Filip Šroubek, "Convolutional neural networks for direct text deblurring," in BMVC, 2015.
- Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang, "Deblurring Text Images via L0-Regularized Intensity and Gradient Prior," in CVPR 2014.

Download from
www.fit.vutbr.cz/~ihradis/CNN-Deblur

This research is supported by the ARTEMIS joint undertaking under grant agreement no 621439 (ALMARVI). This work was supported by The Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project "IT4Innovations National Supercomputing Center – LM- 2015070".