IMAGE SUPER-RESOLUTION USING DEEP CONVOLUTIONAL NETWORKS (ARXIV PREPRINT ARXIV:1501.00092)
&
LEARNING A DEEP CONVOLUTIONAL NETWORK FOR IMAGE SUPER-RESOLUTION (ECCV2014)

Chao Dong[1], Chen Change Loy[1], Kaiming He[2], and Xiaoou Tang[1]

[1]Department of Information Engineering, The Chinese University of Hong Kong

[2]Microsoft Research

A student presentation for a course: Deep Learning and Its Applications
Sami Varjo, University of Oulu, 2015

Super Resolution

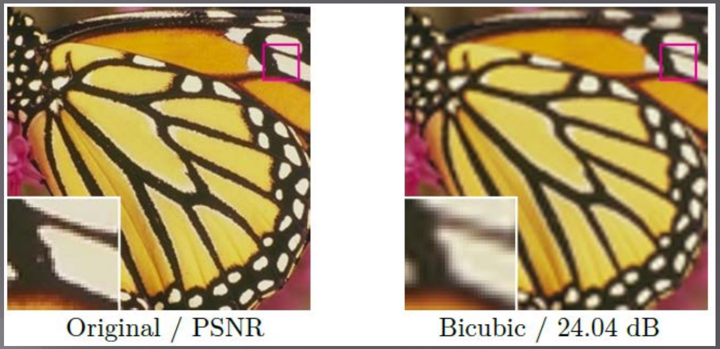
- Reminder:
  - Spatial resolution, # of pixels (or OTF/PSF)
  - Brightness resolution, # of gray levels (HDR data)
  - Temporal resolution, # of frames

- Generally the goal is to get out more than the utilized sensor is cabable of

- Two main categories
  - multiple input images
  - single image

Single Image Superresolution

- Fancy term for "smart" image up scaling
- Some approaches:
  - Bicubic interpolation
  - Blind deconvolution
  - Learned Wavelet Coefficients
  - Generalized interpolation (to eigen decomposition)
  - Sparse dictionaries or sparse coding (SC)
  - etc...

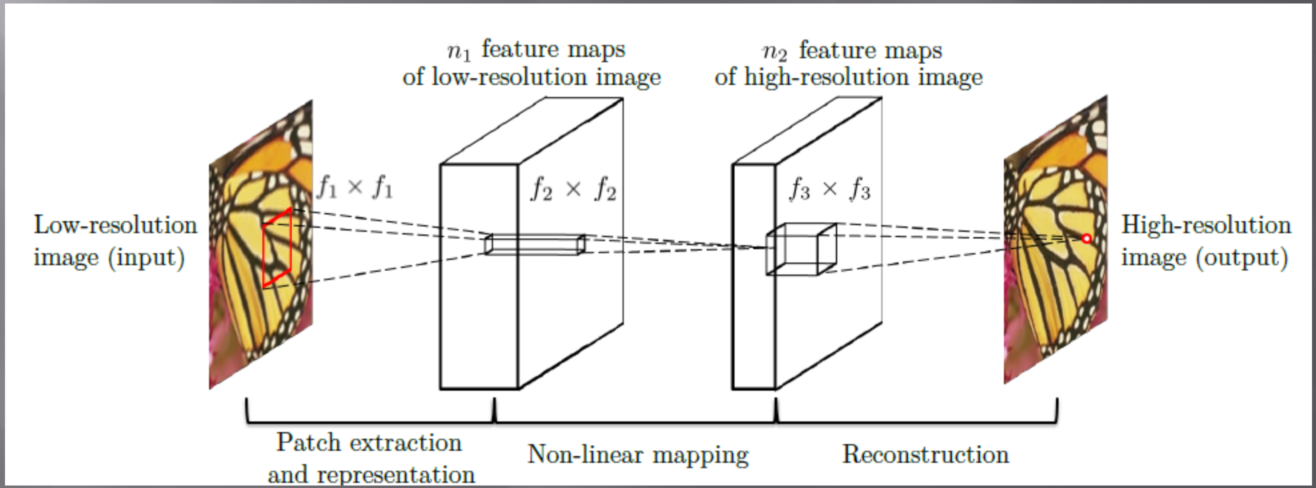


- Convolutional neural networks (CNN) have been utilized previously for denoising

Super-Resolution Convolutional Neural Network (SRCNN)

- Goal end to end mapping i.e. LR in SR out with CNN

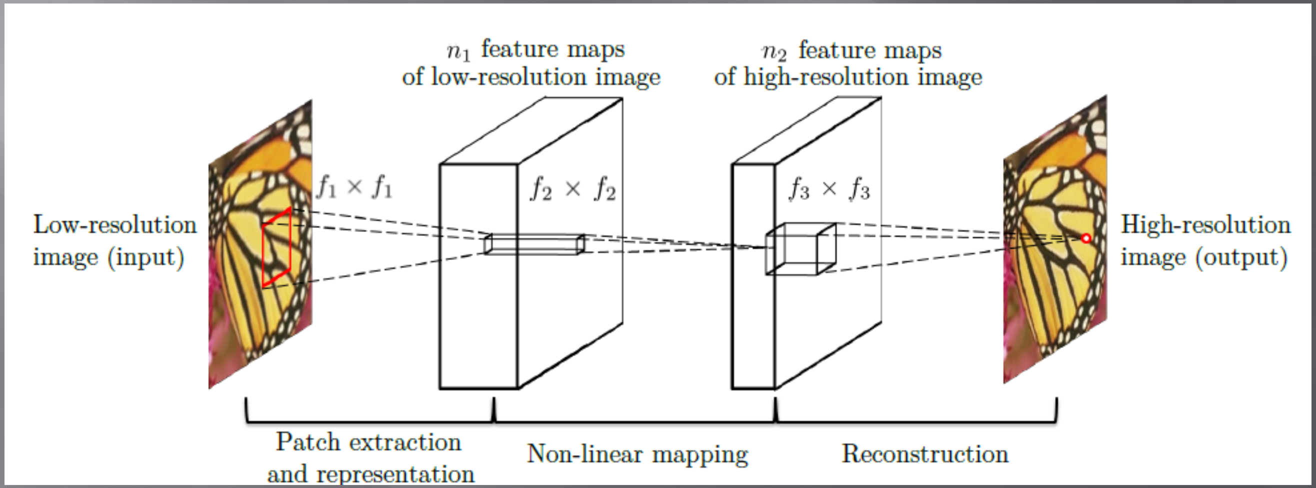
1. Up scale image to target size using bicubic interpolation
2. Patch extraction and representation (first layer)
3. Non-linear mapping (ReLU)
4. Reconstruction



A Matlab implementation available at (no learning part)
http://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html

Super-Resolution Convolutional Neural Net-work (SRCNN)

- Two convolutional layers with ReLUs followed with a reconstruction layer (e.g. $f_1 = 9, f_2 = 1, f_3 = 5, n_1 = 64, n_2 = 32$)
- No border padding (smaller image out, but no border effects).
- Simple structure → speed



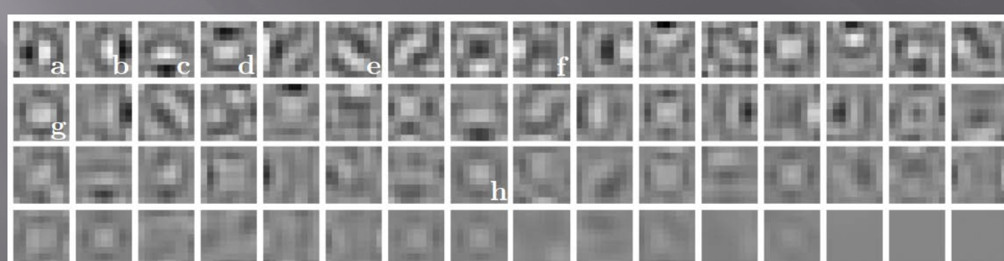
A Matlab implementation available at (no learning part)
http://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html

Training

- Training image size 33x33 91 images → 24800 sub-images Imagenet → 5 million sub-images
- LR-input by Gaussian blurring + subsampling
- Mean Squared Error minimized using stochastic gradient descent

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} ||F(\mathbf{Y}_i;\Theta) - \mathbf{X}_i||^2,$$

- Learning rate 10^{-4} for first two laryers and 10^{-5} for the third layer
- cuda-convnet and Caffe utilized (similar performance)

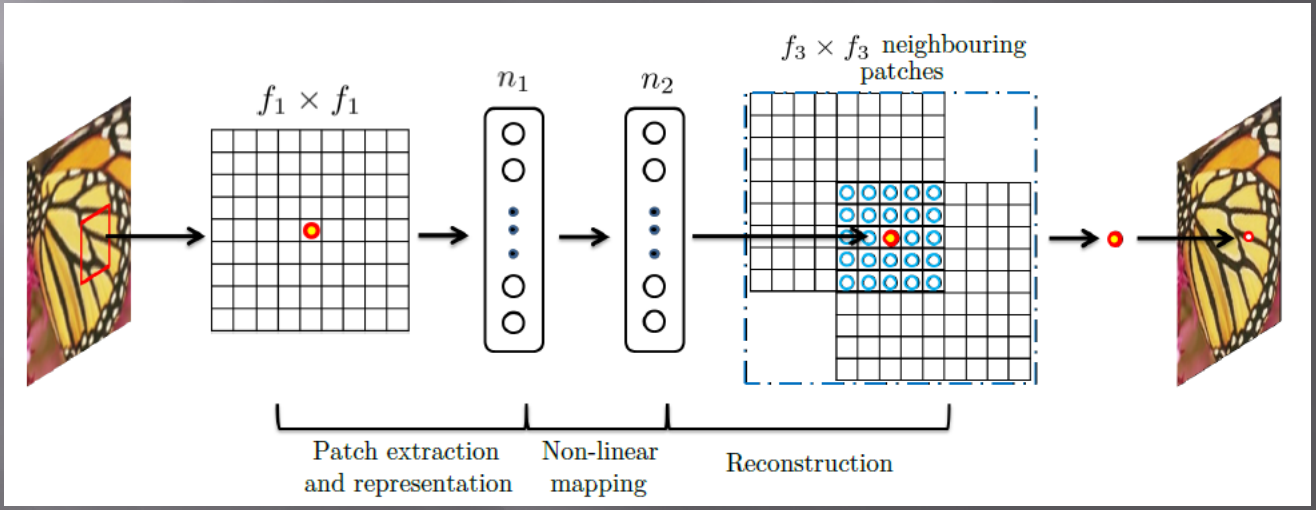


First layer filters trained on ImageNet for x3

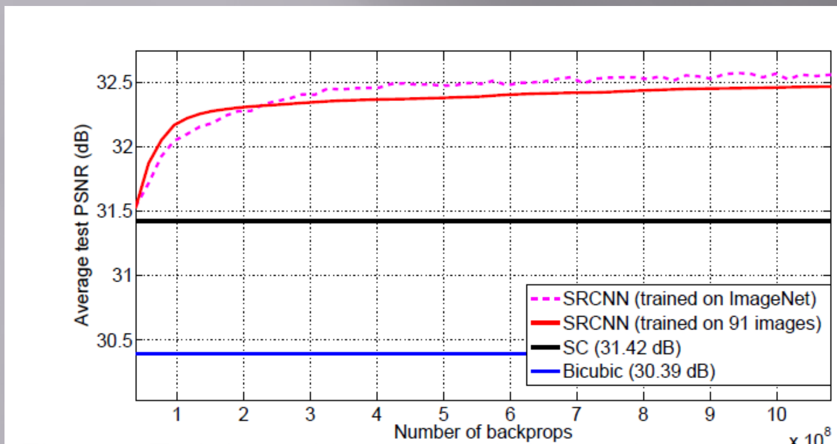
Relation to sparse-coding

SC:

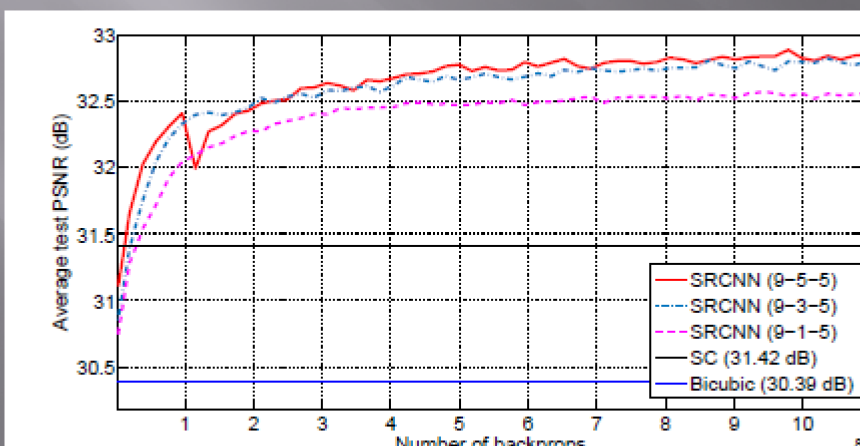
- Patches with size of f_1 x f_1 extracted from image
- Find a n_2 sparse set of coefficients in a n_1 sized dictionary
- Reconstruct a high resolution patch from corresponding HR patches with found weights



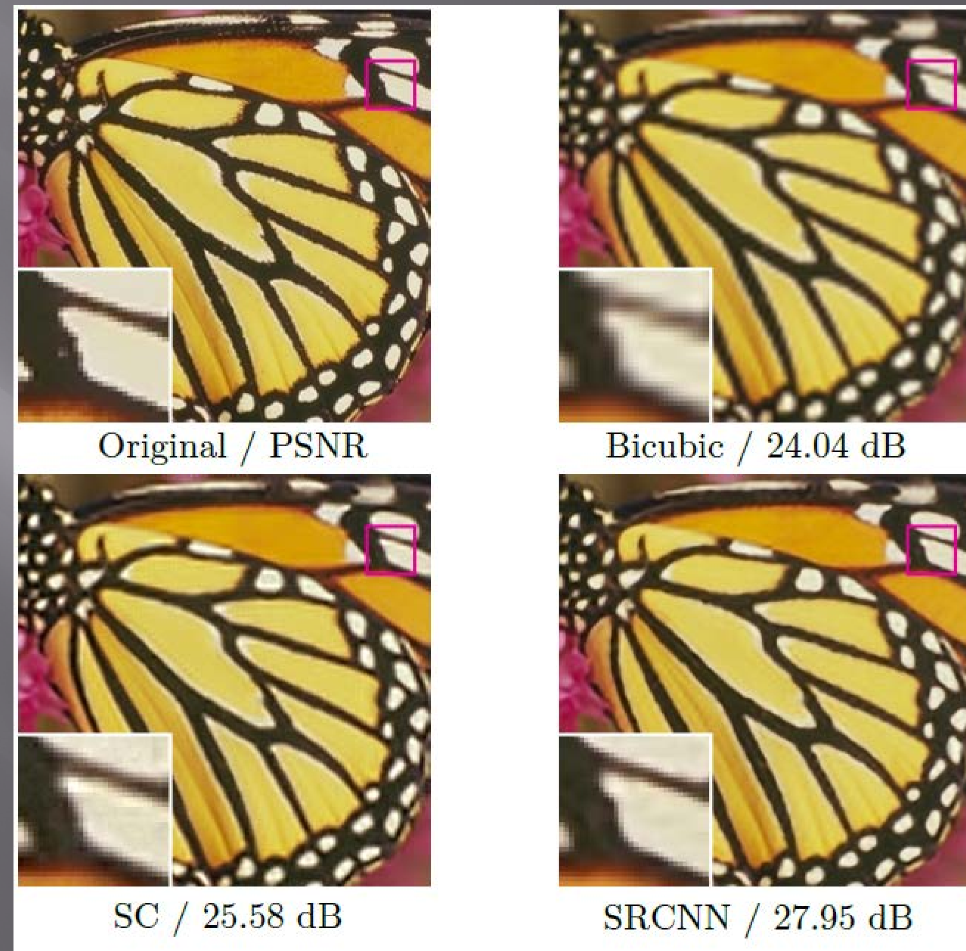
Results



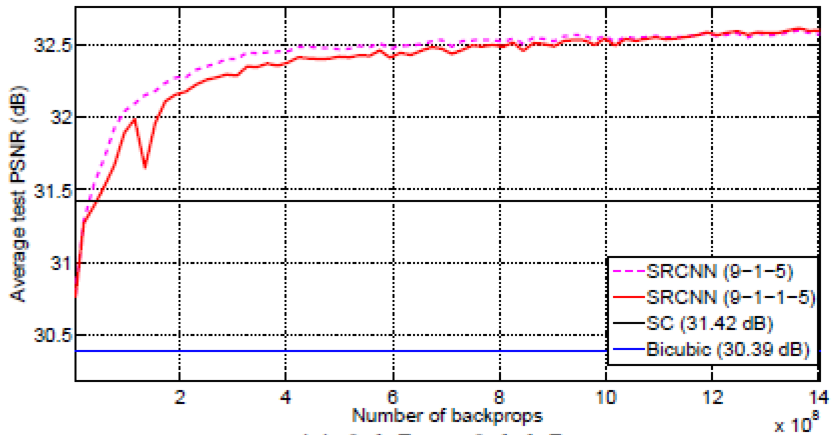
Having large training set effects results surprisingly little.



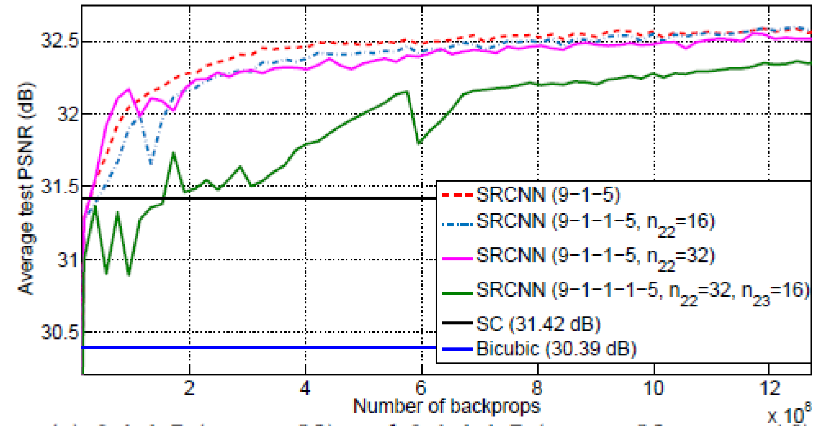
Larger 2nd stage filter is better.



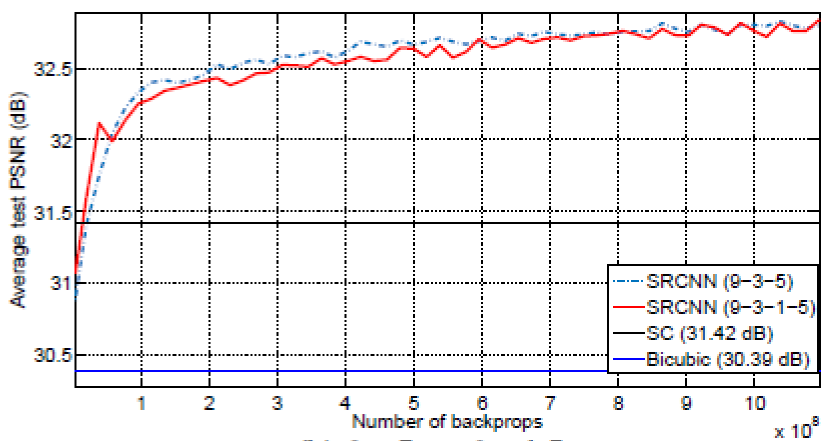
Results



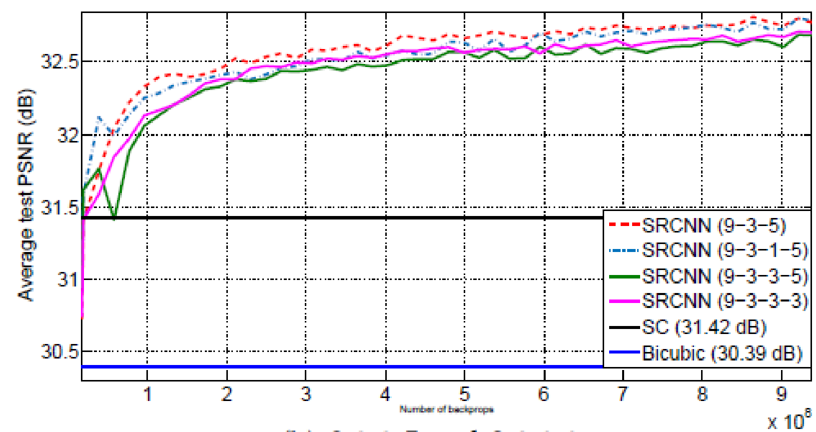
(a) 9-1-5 vs. 9-1-1-5



(a) 9-1-1-5 ($n_{22} = 32$) and 9-1-1-1-5 ($n_{22} = 32, n_{23} = 16$)



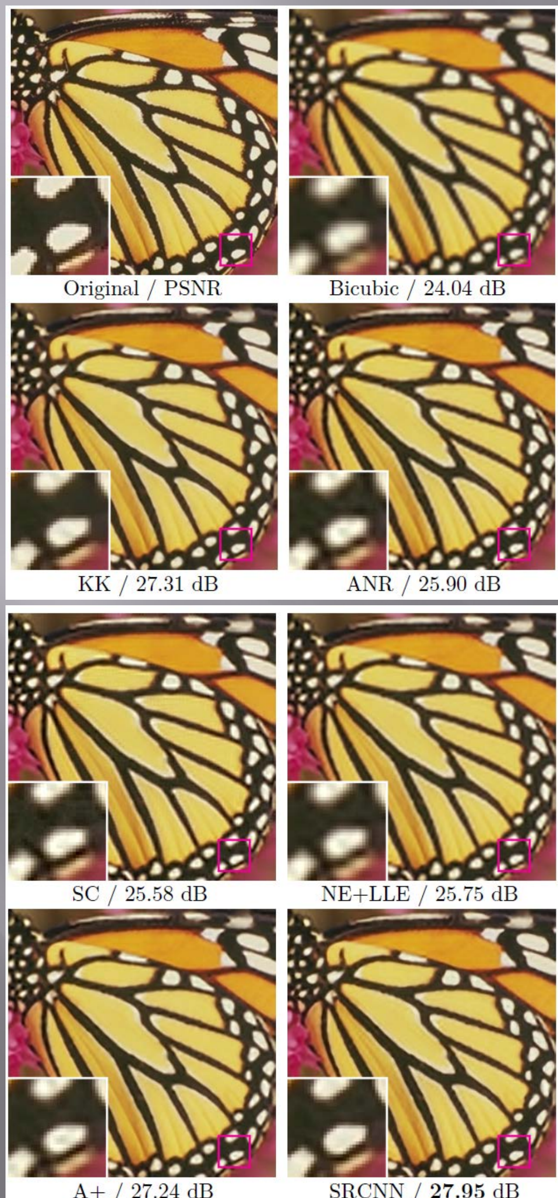
(b) 9-3-5 vs. 9-3-1-5

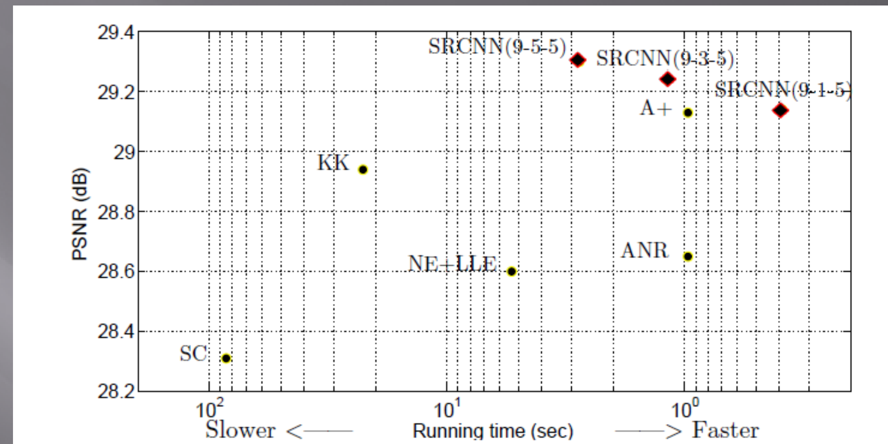


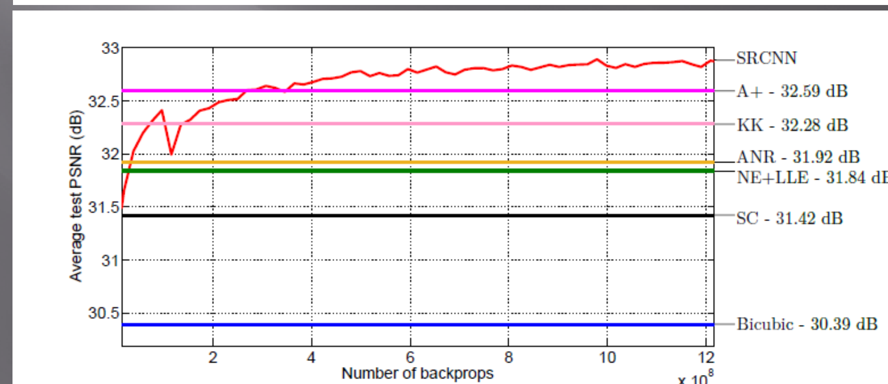
(b) 9-3-3-5 and 9-3-3-3

Added convolutional layers do not improve results here

Results







Comparison to state of art:

A+	Adjusted Anchored Neighbourhood Regression
ANR	Anchored Neighbourhood Regression
KK	Sparse Regression
NE+LLE	Neighbour embedding + locally linear embedding method
SC	Sparce Coding

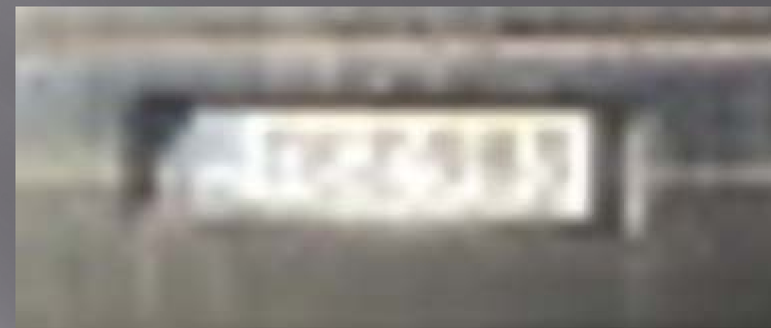
Results


Original / PSNR
Bicubic / 22.49 dB
SC / 23.40 dB
NE+LLE / 23.34 dB
KK / 23.67 dB
ANR / 23.39 dB
A+ / 23.76 dB
SRCNN / **24.29** dB

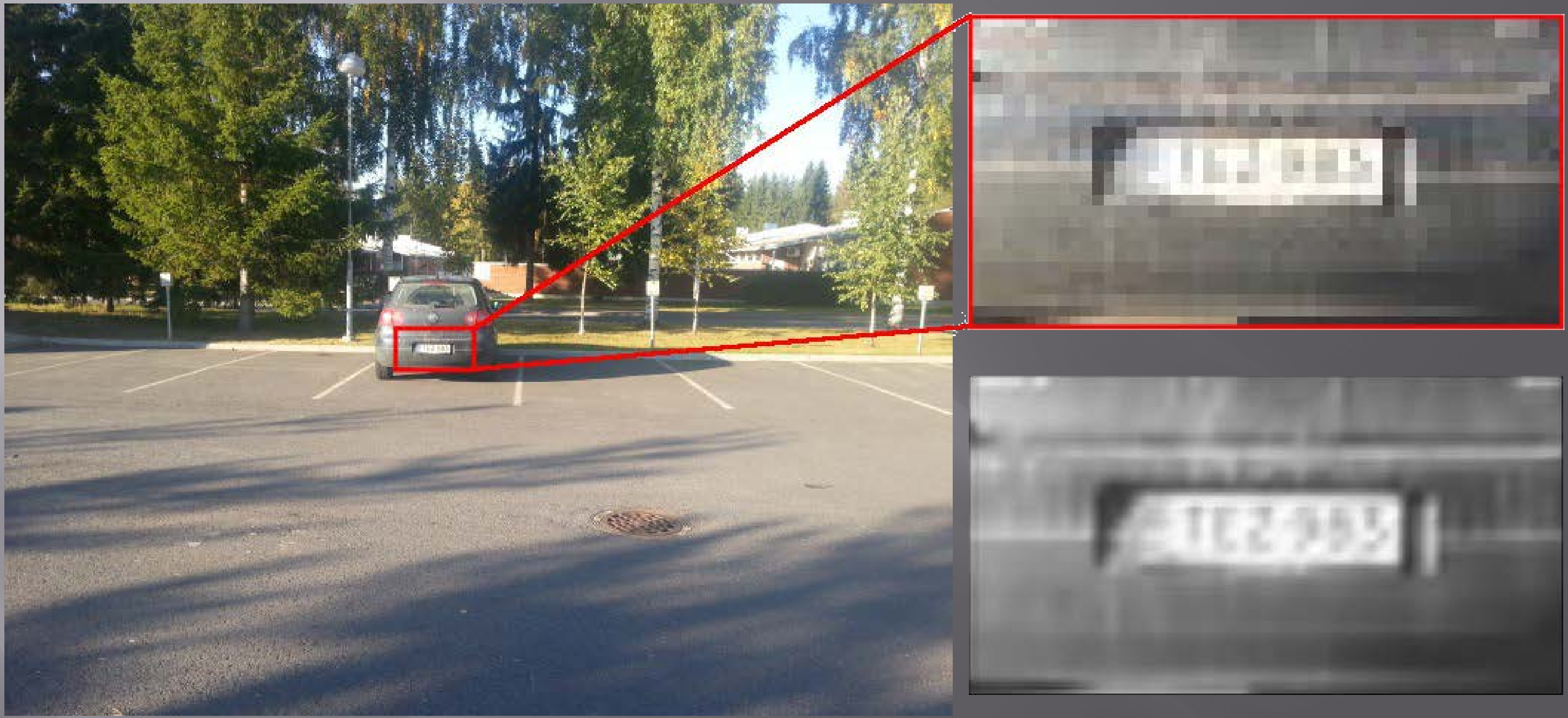
Own Experiment (multi-image)

.
.
20 images
.

Own Experiment Result (single images)



Own Experiment Result (multi-image)

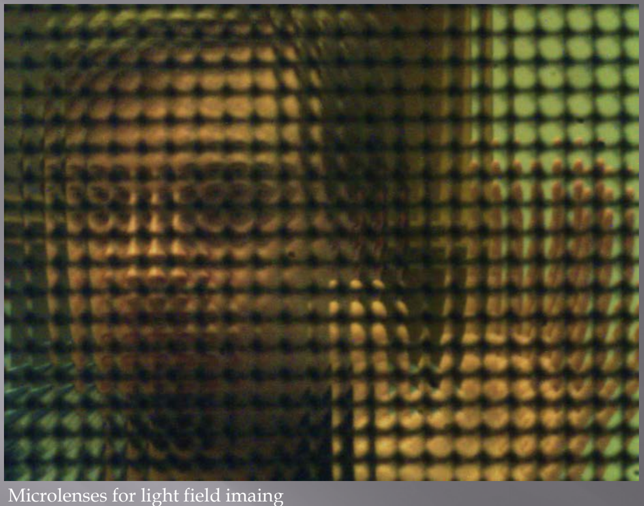


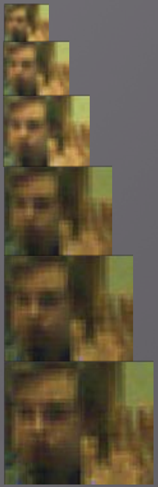
Fourier-Mellin aligned images → 3x upscale →
SRCNN → Re-align → median of 20 images

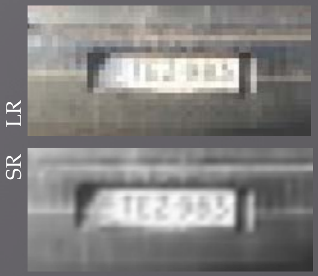
(TCZ-955 appears to actually exist)

Conclusion

- end-to-end mapping between low- and high-resolution images, with little pre/post-processing
- structure is intentionally designed with simplicity → fast
- fully feed-forward and does not need to solve any optimization problem → fast
- deep-learning-based SR method and the sparse-coding-based SR are similar
- SRCNN beats state of art in single image SR in quality (and seems to actually work…)







Thank You for Your Interest