IMAGE SUPER-RESOLUTION USING DEEP CONVOLUTIONAL NETWORKS

BY CHAO DONG, ET AL. (2014)

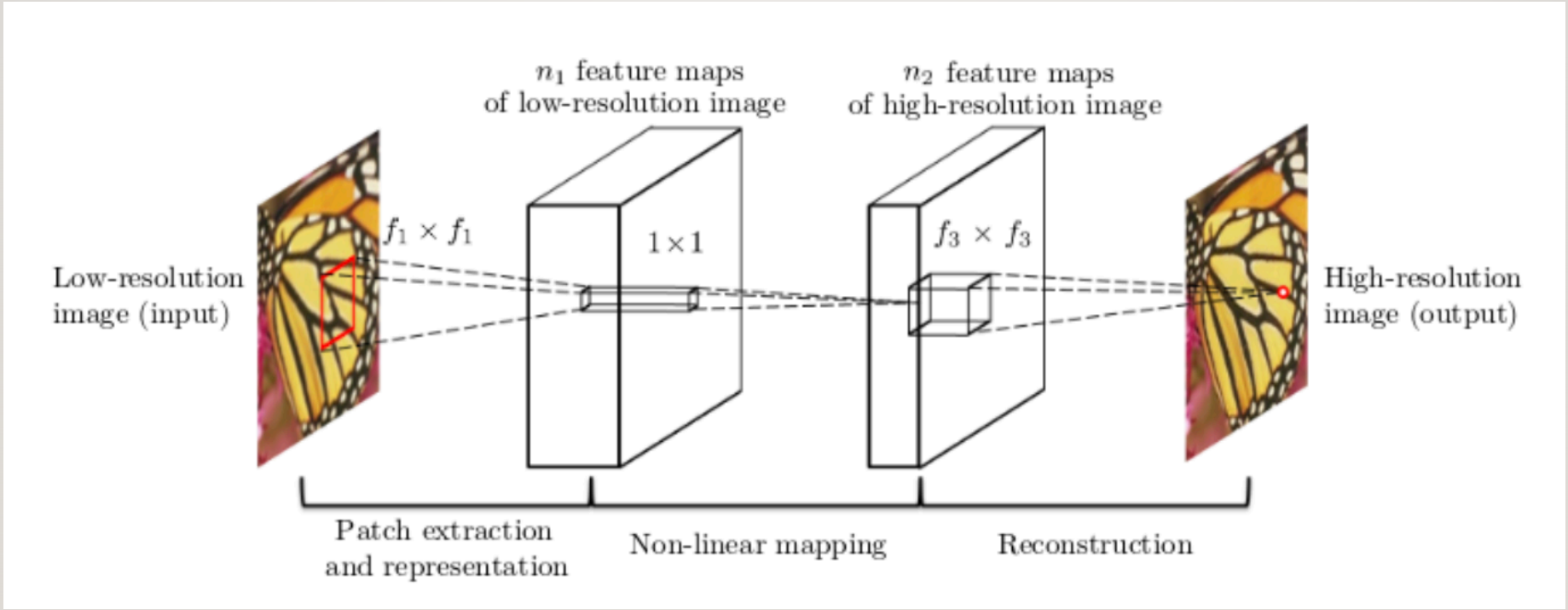
PRESENTED BY: ALAN ZHENG

THE PROBLEM

- Deep CNN that learns the end-to-end mapping of low-resolution to high-resolution images
- Given image X, (m*n*c) pixels in size
- Scale by factor k to image Y, (km*kn*c) in size
- F(X)=Y
- More data from less, k^2 times each pixel
- Resolution images have higher perception to noise
  - Generating a perfect mapping is difficult

PROCESS

- Preprocessing
  - Up-scales LR image to desired HR size
  - Bicubic/spline interpolation
- Feature extraction
  - Extracts a set of feature maps from the up-scaled LR image.
  - Represents each patch of the LR image as a high-dimensional vector, vectors comprise the feature maps
- Non-linear mapping
  - Maps the feature maps representing LR to HR patches.
- Reconstruction
  - Produces the HR image from HR patches.


n_1 feature maps
of low-resolution image
n_2 feature maps
of high-resolution image
$f_1 \times f_1$
1×1
$f_3 \times f_3$
Low-resolution
image (input)
High-resolution
image (output)
Patch extraction
and representation
Non-linear mapping
Reconstruction

$$F_1(\mathbf{Y}) = \max(0, W_1 * \mathbf{Y} + B_1)$$

$$F_2(\mathbf{Y}) = \max(0, W_2 * F_1(\mathbf{Y}) + B_2)$$

$$F(\mathbf{Y}) = W_3 * F_2(\mathbf{Y}) + B_3$$

METRICS

- MSE
  - Mean Sqaured Error

$$MSE = \frac{\sum_{i,j=1}^{km,kn} \|Y_{true} - Y_{pred}\|^2}{km * kn}$$

- PSNR
  - Peak Signal to Noise Ratio

$$PSNR = 10 \log_{10}\left(\frac{R^2}{MSE}\right)$$

- MOS
  - Mean Opinion Score
  - Opinions/user ratings on the same image being processed by different algorithms
- SSIM
  - Structural Similarity Index
  - Takes into account, the image degradation (change) across luminance (*l*), contrast (c) as well as structure (*s*).

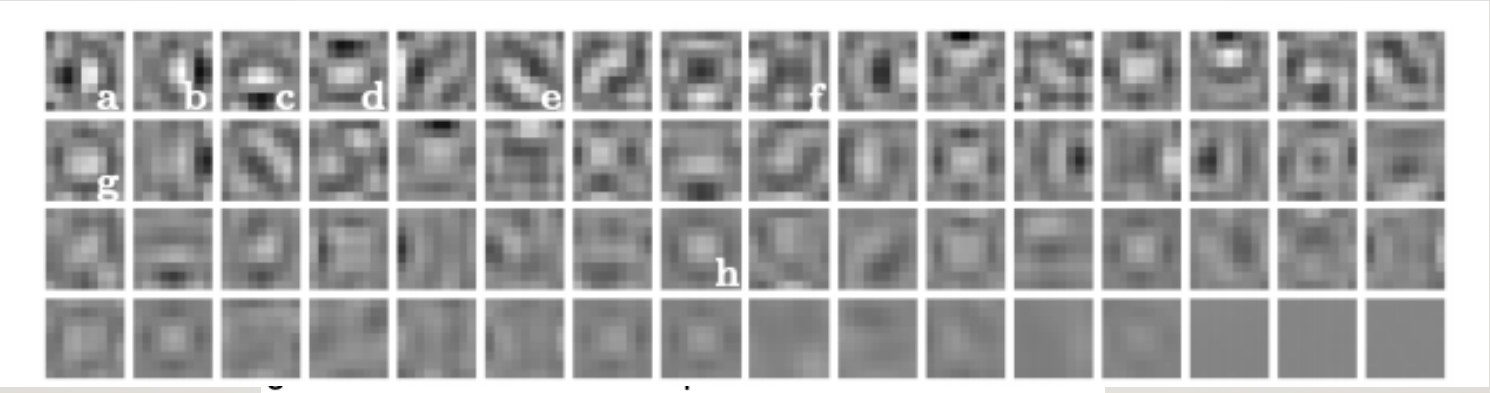
$$SSIM(x,y) = \left[l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma}\right]$$

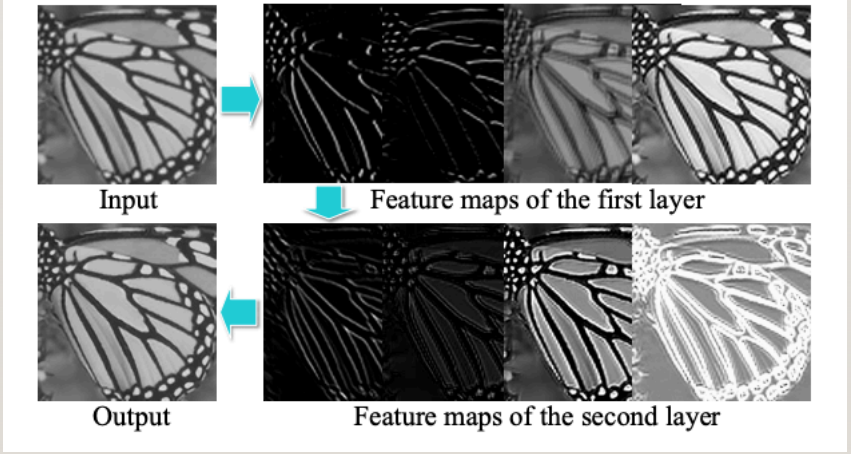
$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \qquad c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \qquad s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$$

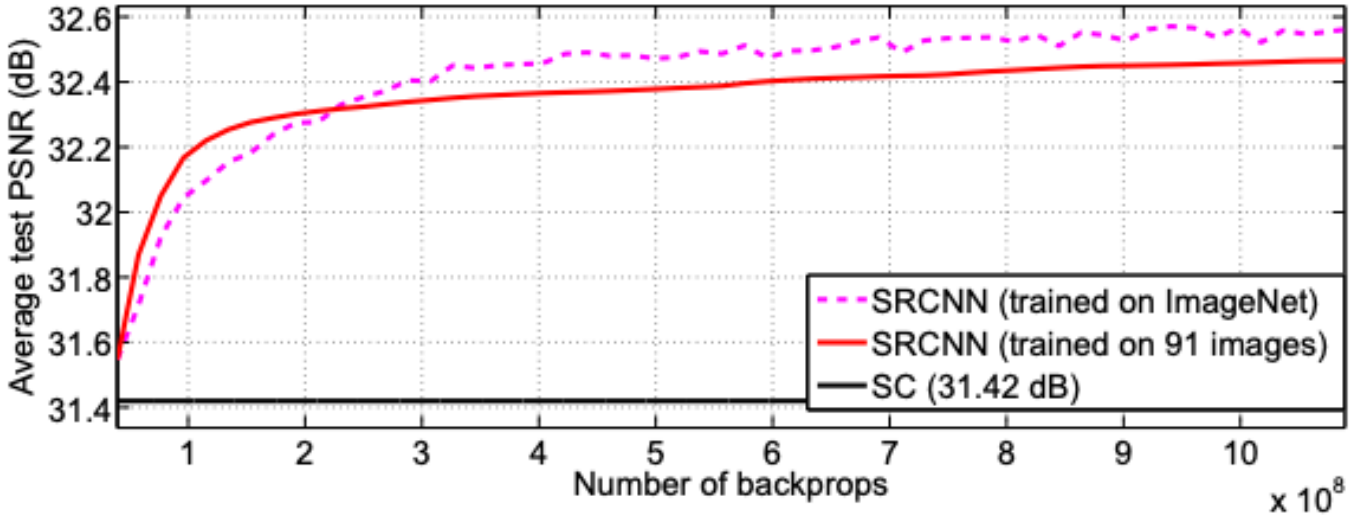
TRAINING

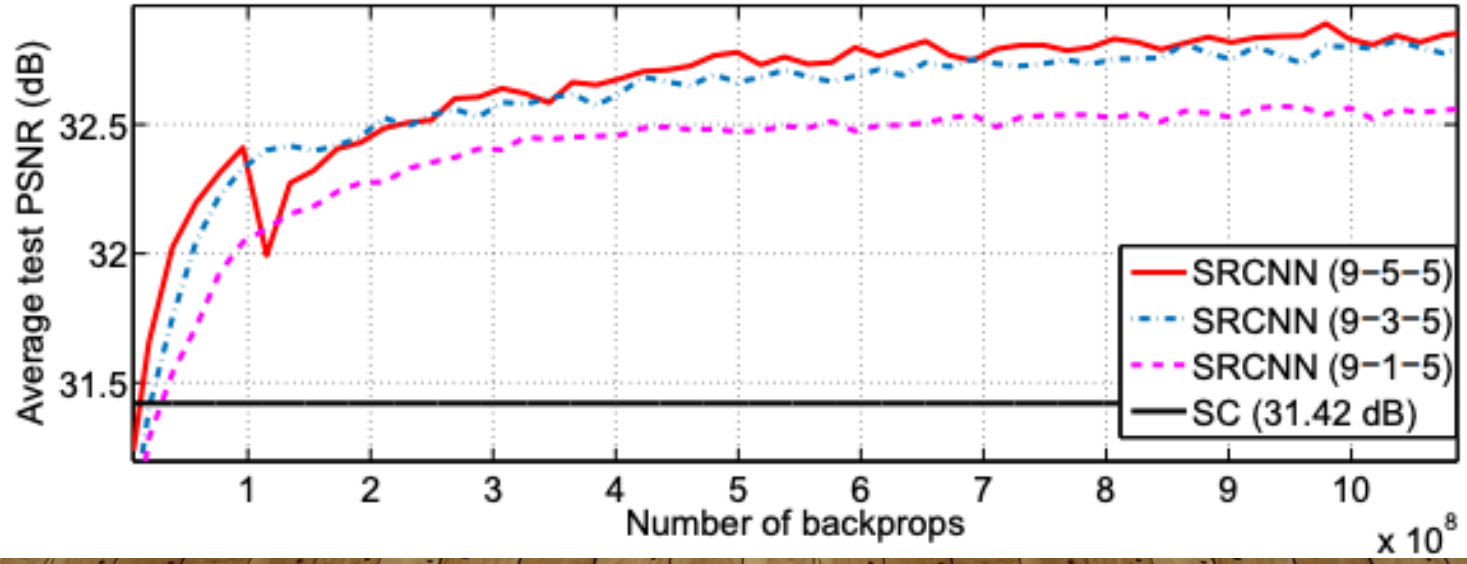
- MSE
- SGD, standard backprop
- cuda-convnet
- Data – ImageNet
  - 91-image set into 24,800 subimages
  - 395,909 image set
- Testing – Set5

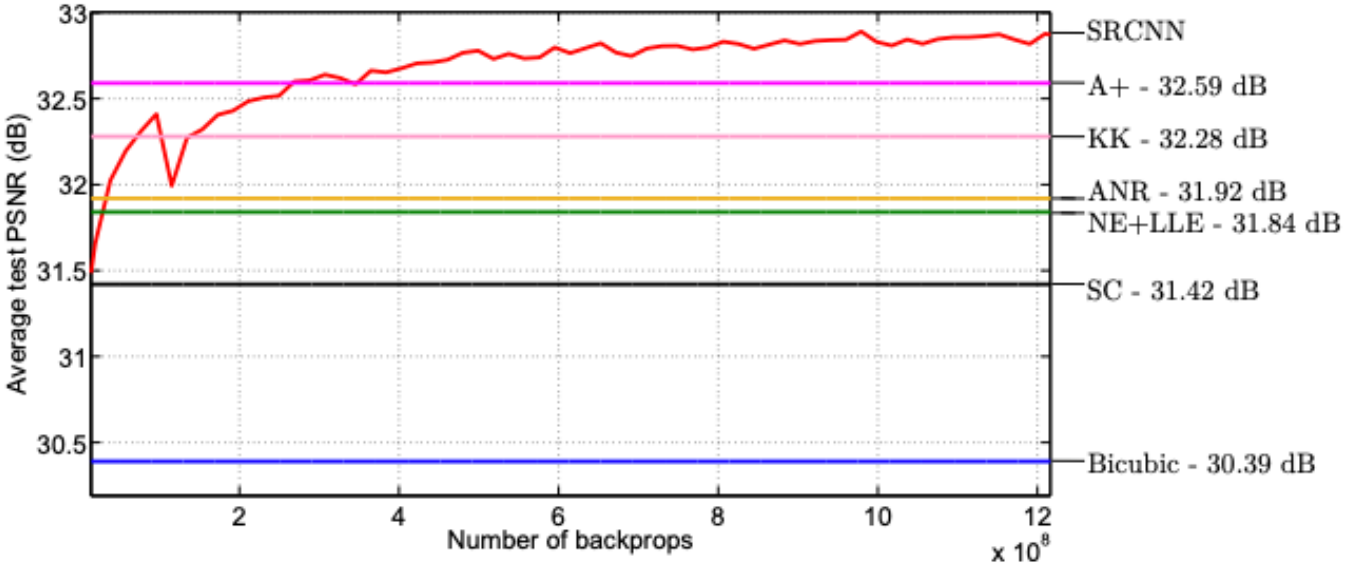
RESULTS

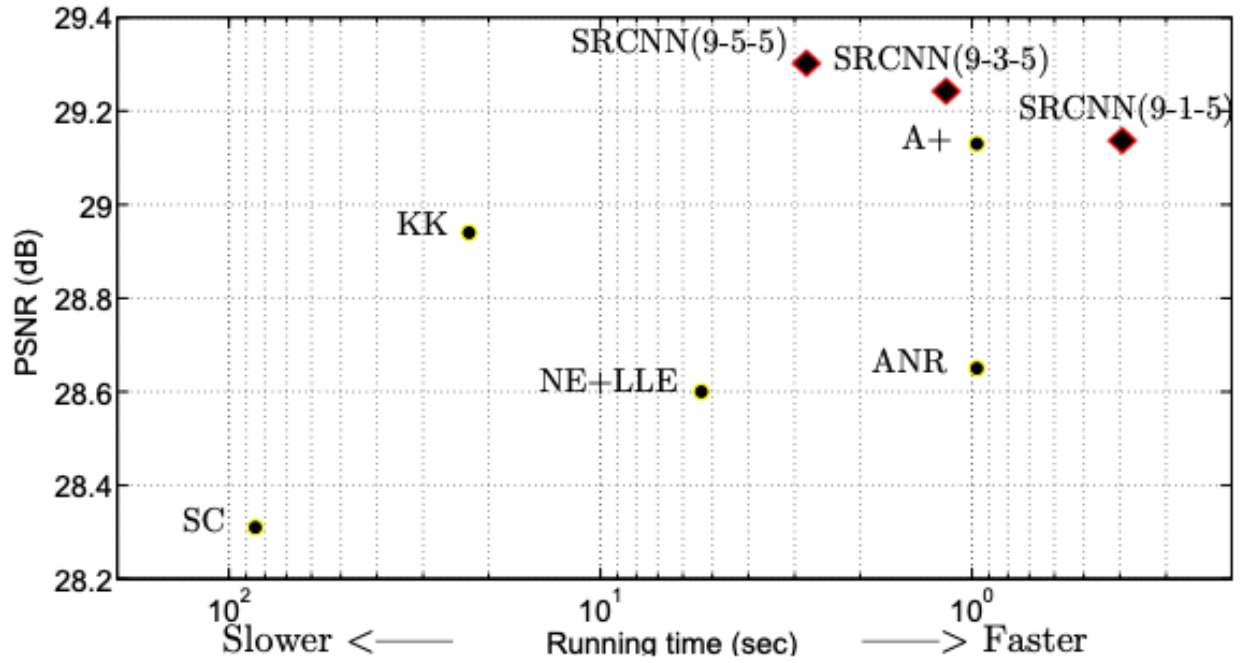

a
b
c
d
e
f
g
h


Input
Feature maps of the first layer
Output
Feature maps of the second layer


Average test PSNR (dB)
32.6
32.4
32.2
32
31.8
31.6
31.4
1
2
3
4
5
6
7
8
9
10
Number of backprops
x 10^8
SRCNN (trained on ImageNet)
SRCNN (trained on 91 images)
SC (31.42 dB)


Average test PSNR (dB)
32.5
32
31.5
1
2
3
4
5
6
7
8
9
10
Number of backprops
x 10^8
SRCNN (9−5−5)
SRCNN (9−3−5)
SRCNN (9−1−5)
SC (31.42 dB)


33
32.5
32
31.5
31
30.5
Average test PSNR (dB)
2
4
6
8
10
12
Number of backprops
x 10^8
SRCNN
A+ - 32.59 dB
KK - 32.28 dB
ANR - 31.92 dB
NE+LLE - 31.84 dB
SC - 31.42 dB
Bicubic - 30.39 dB


29.4
29.2
29
28.8
28.6
28.4
28.2
PSNR (dB)
SRCNN(9-5-5)
SRCNN(9-3-5)
SRCNN(9-1-5)
A+
KK
NE+LLE
ANR
SC
10^2
10^1
10^0
Slower <——
Running time (sec)
——> Faster

EXPERIMENTS ON COLOR CHANNELS

- YCbCr instead of RGB
- Training on different channels

Training Strategies	PSNR of different channel(s)			
	Y	Cb	Cr	RGB color image
Bicubic	30.39	45.44	45.42	34.57
Y only	**32.39**	45.44	45.42	36.37
YCbCr	29.25	43.30	43.49	33.47
Y pre-train	32.19	**46.49**	**46.45**	36.32
CbCr pre-train	32.14	46.38	45.84	36.25
RGB	32.33	46.18	46.20	**36.44**
KK	32.37	44.35	44.22	36.32

FUTURE WORKS

- FSRCNN (Fast SRCNN)
- VDSR (Very Deep SR)
- ESPCN (Efficient Sub-Pixel CNN)
- SRResNet