Single Image Super-Resolution

Yujing Song

University of Minnesota, Morris

UMM CSci Senior Seminar Conference, November 17 2018

What is Image Super-Resolution?

- Single image super-resolution, SR
- From low-resolution to high-resolution

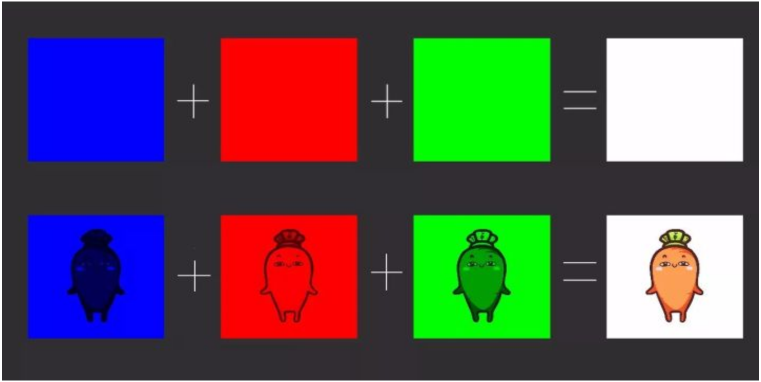
Outline

Images

What makes an image?

- Pixels
- Channels

Channel



Pixels

0	1	2	2	0
0	1	1	0	0
2	2	0	2	1
0	1	1	0	2
1	2	2	1	1

Red Component

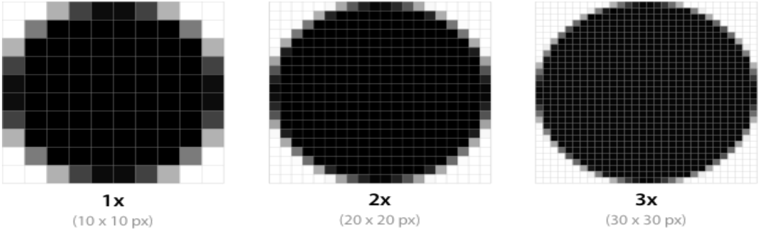
1	0	2	1	1
1	2	2	1	0
1	1	0	2	1
0	0	0	1	1
2	1	1	2	0

Green Component

0	2	2	0	2
1	1	1	1	1
0	0	1	2	2
0	0	1	2	2
0	2	2	2	0

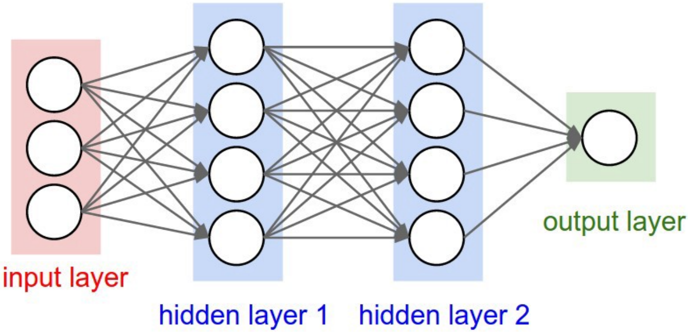
Blue Component

Image Upscaling

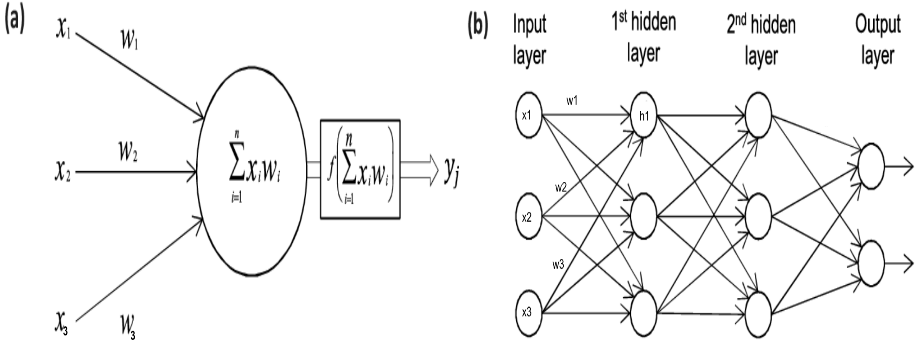


- Bicubic interpolation
- Sparse-coding base method (SC)

Basic structure of neural network



Basic structure of neural network

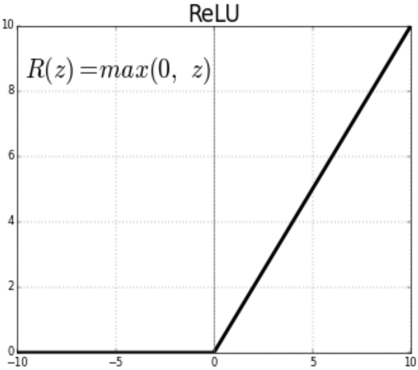


Activation function

Rectified linear unit (ReLU):

$$R(z) = max(0, z)$$

To introduced the non-linearity in neural networks



Loss function

Loss function is used to measure the degree of inconsistency between the predicted value and the true value:

- Peak signal-to-noise ratio (PSNR)

$$PSNR = 10log_{10}\frac{{MAX_I}^2}{MSE}$$

- Mean squared error (MSE)

Mean squared error

- n: number of pixels
- X: pixels of original image
- $\widehat{X}$: pixels of high-resolution image

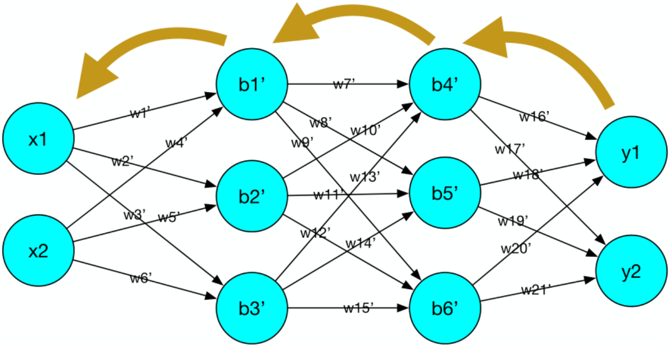
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{X}_i)^2$$

Stochastic gradient descent

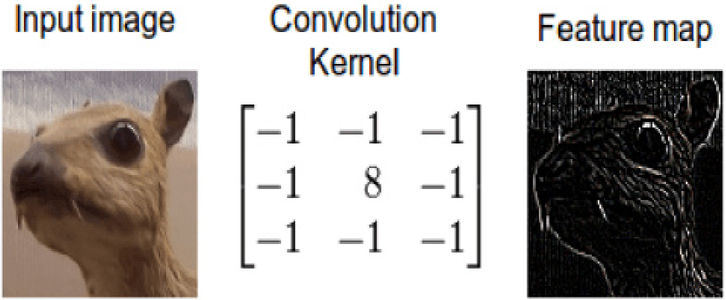
- W_{j+1}: new weight
- W_j: the weight from last iteration W_j
- η: learning rate

$$W_{j+1} = W_j + \eta \cdot \frac{\partial L}{\partial W_j}$$

Back-propagation



Kernel


Input image
Convolution Kernel
Feature map
$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

RGB and Kernel

0	1	2	2	0
0	1	1	0	0
2	2	0	2	1
0	1	1	0	2
1	2	2	1	1

Red Component

1	0	2	1	1
1	2	2	1	0
1	1	0	2	1
0	0	0	1	1
2	1	1	2	0

Green Component

0	2	2	0	2
1	1	1	1	1
0	0	1	2	2
0	0	1	2	2
0	2	2	2	0

Blue Component

1	0	1
1	1	0
1	0	1

Red Kernel

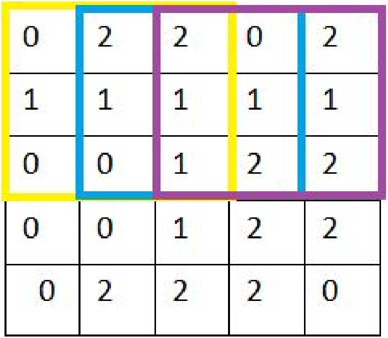
-1	-1	-1
0	1	0
0	1	1

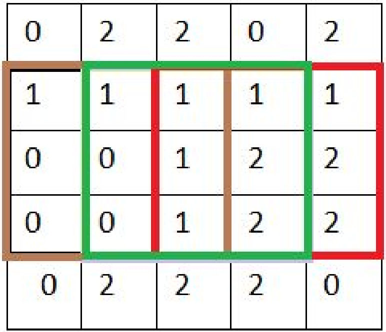
Green Kernel

1	1	0
-1	0	1
-1	1	-1

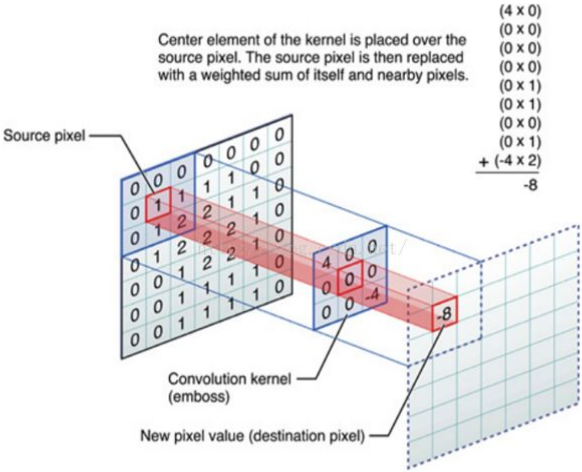
Blue Kernel

Sliding windows

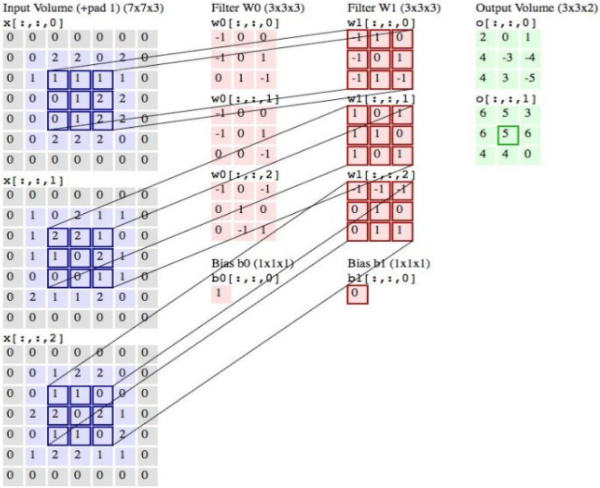




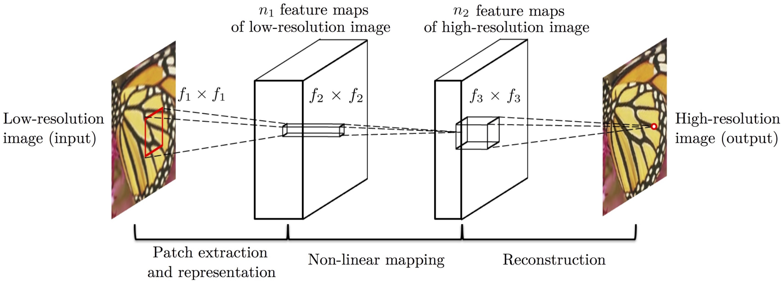
Convolution for a single channel



Convolution for multiple channels



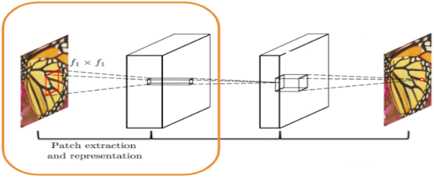
Super-resolution by convolutional neural network



Patch extraction and representation

- Y: upscaled low-resolution image
- W_1: filters in the first layer
- B_1: biases in the first layer
- $F_1(Y)$: feature maps in the first layer

$$F_1(Y) = max(0, W_1 * Y + B_1)$$



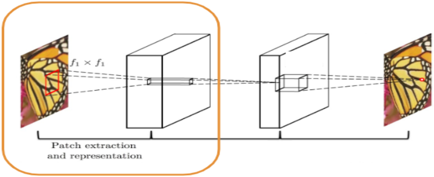
Patch extraction and representation

Suppose the input image Y:

- Size: $A \times B$
- n_1 filters
- Kernel size: $f_1 \times f_1$

After the first step operation:

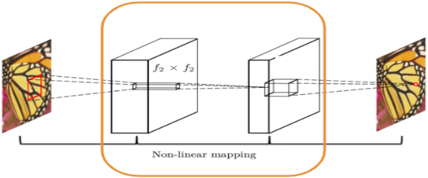
- n_1 feature maps
- Without negative pixels



Non-linear mapping

- W_2: filters in second layer
- B_2: biases in second layer
- $F_2(Y)$: feature maps in second layer

$$F_2(F_1(Y)) = max(0, W_2 * F_1(Y) + B_2)$$



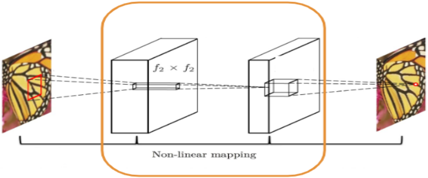
Non-linear mapping

In the second operation step:

- n_2 filters
- Kernel size: 1×1

Notice:

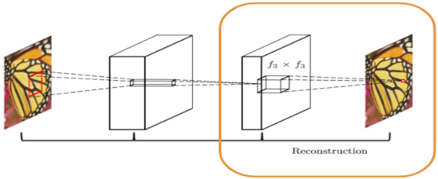
- Kernel size is 1×1
- Same with fully connected layer



Reconstruction

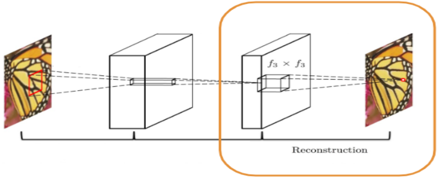
- W_3: filters in third layer
- B_3: biases in third layer
- $F(Y)$: high-resolution output

$$F(F_2(F_1(Y))) = W_3 * F_2(Y) + B_3$$



Reconstruction

- n_2 kernels
- Kernel size: $f_3 \times f_3$
- Filters act like an averaging filters



Training

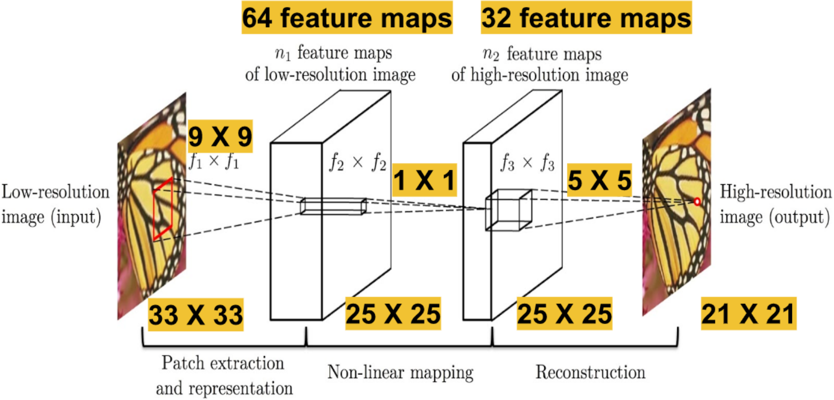
Training set:

- $395,909$ images from ImageNet
- 5 million of 33×33 sub-images

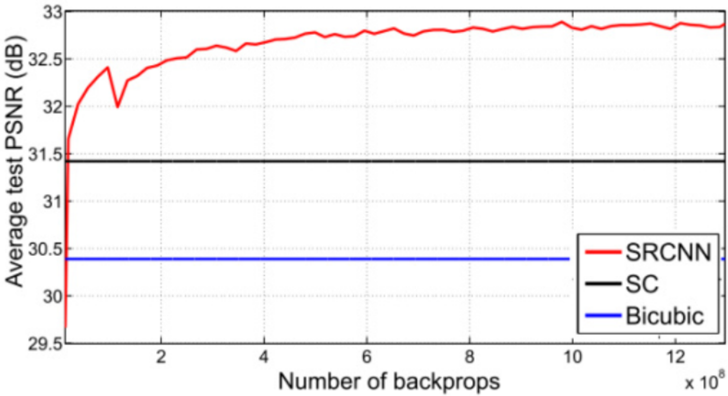
Training kernels:

- MSE as loss function
- Stochastic gradient descent

Training process



Result comparison

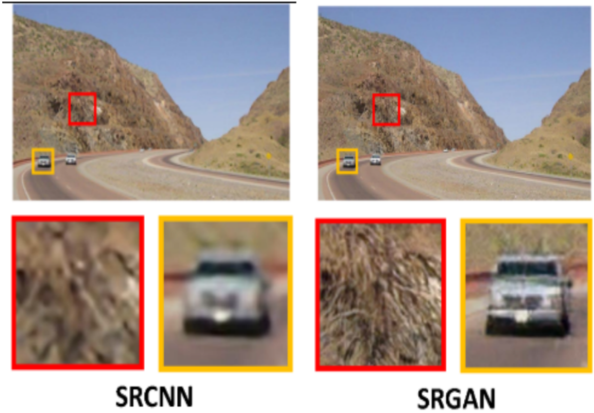


Super-resolution by generative adversarial network

Generative adversarial network:

- Generator network: creator
- Discriminator network: classifier

What is improved?



Conclusion

- Neural network
- Convolutional neural network
- Solving super-resolution problem

Acknowledgement

Thank you to my advisor Peter Dolan and Elena Machkasova for guidance and feedback

Discussion

Questions?