Optimization and analysis of Deep Learning algorithm for Single Image Super Resolution

Mattia Ceccarelli

Alma Mater Studiorum - Università di Bologna

12/05/2020

Single Image Super Resolution

What is Super Resolution?

Two main groups of techniques :

- SR microscopy, which aim is to overcome the diffraction limits with a complex laboratory setup.
- SR algorithm, which aim is to enhance the spatial resolution of an image for different purpose such as object detection, segmentation and improve its visual quality

Super resolution procedure

- Start from a prior-known High Resolution (HR) image;
- Rescale the image obtaining its Low Resolution (LR) counterpart;
- Feed a super-resolution model with the LR image trying to obtain as output its HR version;
- With the trained model we can re-apply the up-sampling procedure and further increase the HR quality*.

* This is even more efficient if we think about a **zoom** procedure.

Deep Learning model

- Neural Networks are mathematical models commonly used in data analysis.
- From a theoretical point-of-view we can define a Neural Network as **a series of non-linear multi-parametric functions**.
- Neural Networks are considered as **Universal Function Approximators**.

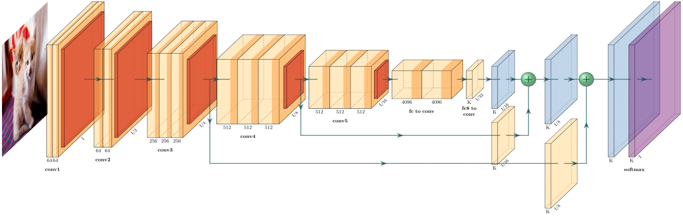


Figure: Neural Network model scheme. The model is composed by a series of layers made by interconnected computational units. The connections between layers can be serials or parallels and each layer models a pre-determined mathematical function.

Deep Learning

Neural Network models

- The supervised training of neural network happens by feeding the model with labeled example, that are couples of inputs and expected outputs.
- Then, the output is compared with the label by means of a **cost function**, which is a measure of the error made by the network. The objective is to modify the **parameters** of the model to minimize the cost function.
- Commonly, we think about Neural Network models as tools to **reduce** problems dimensionality, i.e. starting from high-dimensional data (e.g $w \times h \times c$ image) the model predicts a class (single number).
- In Super Resolution we have no classes but the desired output is also an image. Single Image Super Resolution tasks start from an image and the aim is to reconstruct its HR counterpart.

State-of-art libraries

Why another library?

Leaders on this topic are the multiple open-source Python libraries available on-line as Tensorflow, Pytorch and Caffe.

Pros

- Simplicity in writing complex models in a minimum number of code lines;
- Simplicity of the Python language;
- Extremely efficient in GPU environments;

Cons

- Steep learning curve for complex applications or algorithm modifications;
- They are not designed for CPU performances;

Only a small part of the research community uses deeper implementation in C++ or other low-level programming languages. About them it should be mentioned the darknet project of Redmon J. et al. which created a sort of standard in object detection applications using a pure Ansi-C library.

Byron

Build YouR Own Neural Network

Byron

- The library is written in **pure C++** with the support of the modern standard 17.
- The library is **optimized for image processing** (probably the most common task in biomedical research).
- Starting from the darknet project backbone, Byron introduces **numerous improvements and fixes**.
- The library is also **completely wrapped** using Cython to enlarge the range of users also to the Python ones.

Standard up/down scaling algorithms

Single Image Super Resolution

Interpolation algorithms:

- **Nearest**;
- **Bicubic**;
- **Lanczos**;

The best compromise between computation cost and efficiency is given by the Bicubic interpolation algorithm. Given a pixel, the interpolation function evaluates the 4 pixels around it applying a filter given by the equation:

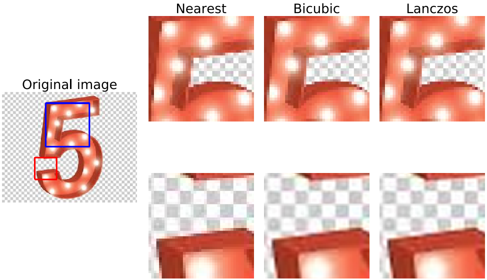


Image Quality

Single Image Super Resolution

The most common image quality evaluator is given by our eyes.
Quantitative evaluations:

PSNR - Peak Signal to Noise Ratio

- It establishes the compression lossless of an image;
- Equation:

$$PSNR = 20 \cdot \log_{10}\left(\frac{\max(I)}{\sqrt(MSE)}\right)$$

where $\max(I)$ is the maximum value which can be taken by a pixel in the image and MSE is the Mean Square Error evaluated between the original and reconstructed images.

SSIM - Structural SIMilarity index

- It evaluates the structural similarity between two images taking into account also the visible improvement seen by human eyes.
- Equation:

$$SSIM(I,K) = \frac{1}{N}\sum_{i=1}^{N}\frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{({\mu_x}^2 + {\mu_y}^2 + c_1)({\sigma_x}^2 + {\sigma_y}^2 + c_2)}$$

where N is the number of arbitrary patches which divide the image and c_1 and c_2 parameters are fixed to avoid mathematical divergence.

	PSNR (dB)	SSIM ($\in [-1,1]$)
Nearest	25.118	0.847
Bicubic	27.254	0.894
Lanczos	26.566	0.871

EDSR - Enhanced Deep Super Resolution

Super Resolution Models

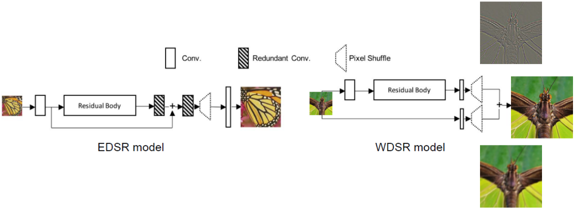
Architecture

Layer	Channels input/output	Filter dimensions	Number of Parameters
Conv. input	3/256	3×3	6912
Conv. (residual block)	256/256	3×3	589824
conv. (pre-shuffle)	256/256	3×3	589824
Conv. (upsample block)	256/1024	3×3	2359296
Conv. output	256/3	3×3	6912

Average Time on 510×339 image: 576.92 s.
Winner of the **NTIRE 2017**. More than **3 million** of parameters!

WDSR - Wide Deep Super Resolution

Super Resolution Models



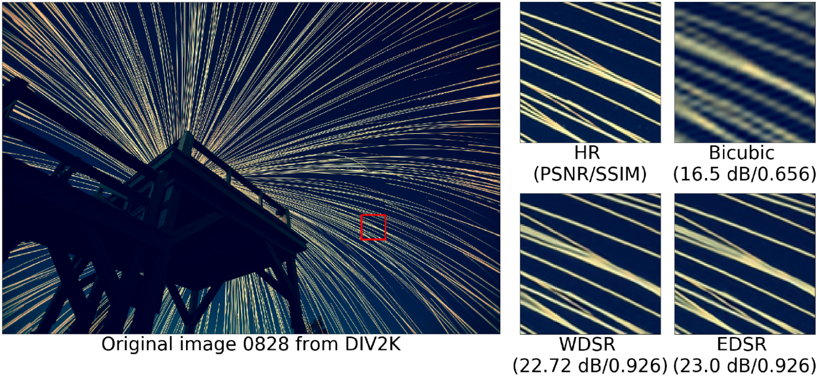
Architecture

Layer	Channels input/output	Filter dimensions	Number of Parameters
Conv. input 1	3/32	3×3	864
Conv. 1 (residual block)	32/192	3×3	55296
conv. 2 (residual block)	192/32	3×3	55296
Conv. (pre-shuffle)	32/48	3×3	13824
Conv. input 2 (pre-shuffle)	3/48	5×5	3600

Average Time on 510×339 image: 46.35 s.
Winner of the **NTIRE 2018**. $\sim$ 100K parameters, **less than** 10% **of EDSR model**!

Image Results

Single Image Super Resolution



The model is trained on the **DIV2K**(*DIVerse 2K resolution high quality images*) dataset.
The dataset contains 800 high-resolution images as training set and their corresponding low-resolution ones, obtained by different down-sampling methods and different scale factors (2, 3, and 4).

Image Results

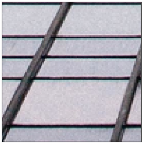
Single Image Super Resolution

The model is trained on the **DIV2K**(*DIVerse 2K resolution high quality images*) dataset. The dataset contains 800 high-resolution images as training set and their corresponding low-resolution ones, obtained by different down-sampling methods and different scale factors (2, 3, and 4).

Image Results

Single Image Super Resolution

Original image 0845 from DIV2K



HR
(PSNR/SSIM)

Bicubic
(21.5 dB/0.669)

WDSR
(24.58 dB/0.801)

EDSR
(25.04 dB/0.811)

The model is trained on the **DIV2K**(*DIVerse 2K resolution high quality images*) dataset.
The dataset contains 800 high-resolution images as training set and their corresponding low-resolution ones, obtained by different down-sampling methods and different scale factors (2, 3, and 4).