Alma Mater Studiorum
University of Bologna

Artificial Intelligence
Machine Learning for Computer Vision
Alessandro Dicosola [Matr. 935563]

Super Resolution
Survey Any-Scale Deep Network (ASDN [1])

Contents

1 Introduction **1**

2 Background **2**

2.1 Deep Bi-Dense Network (DBDN[2]) 3

2.2 Residual Channel Attention Network (RCAN[3]) 4

2.3 Channel-wise and Spatial Feature Modulation Network (CSFM[4]) 5

2.4 Deep Laplacian Pyramid Network (LapSRN[5]) and Multi-scale Deep Laplacian Pyramid Network (MS-LapSRN [6]) 6

2.4.1 Features 6

2.4.2 Architecture 7

2.4.3 Loss 7

2.4.4 Results 8

2.5 Magnification-Arbitrary Network (MetaSR [7]) 9

1 Introduction

In computer vision the **super resolution** (hereinafter SR) task is an ill-posed problem for reconstructing a low resolution (hereinafter LR) image to an higher one (hereinafter SR - super resolution image); in order to do so an high resolution image is used (hereinafter HR).

Many method proposed trying to define the images in the features space and apply linear and non-linear operators in order to learn and reconstruct the high resolution image: using random forest, linear and non-linear regressor, manifold embedding[8].

Due to advancement of deep learning, SR task started to be achieved using deep neural networks: from the simplest one, SRCNN[9] with only three convolution for extracting and processing features used then for reconstructing the SR image to more complex ones that use short,long and dense skip connections [10][2] and recursive blocks [11][12], attention mechanism [3][4], meta-learning [7], multi-levels [5][6].

Studies done in deep learning has lead research to understand that deeper and wider networks performs better than shallow and narrow ones as well as using residual learning; skip connections (short,long and dense) are used for overcoming the exploding/vanishing gradient problem and at the same time they allow to let networks converge faster; dense connection allow also to reuse features from previous stages and attention mechanism allow networks to focus on most important features (channel-wise and spatial-wise).

SR task achieved with deep learning trying to learn a mapping between LR to HR images given a specific scale. The studies done up to now always used fixed integer scales (2x,3x,4x,. . .) but none had explored decimal scales but Meta-SR[7] (using meta-learning).

Therefore **Any-Scale Deep Network** (ASDN) tried to explore decimal scale in SR task learning the mapping in an end-to-end way.

2 Background

Here will be explored the most important networks referenced by ASDN [1].

2.1 Deep Bi-Dense Network (DBDN[2])

2.2 Residual Channel Attention Network (RCAN[3])

2.3 Channel-wise and Spatial Feature Modulation Network (CSFM[4])

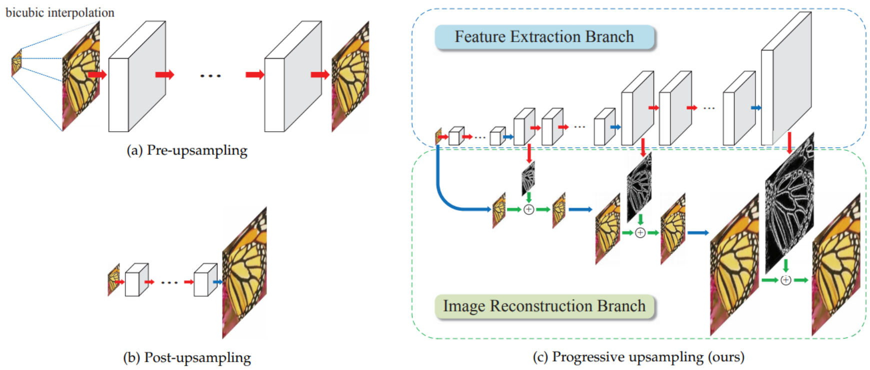

Fig. 1. **Comparisons of upsampling strategies in CNN-based SR algorithms.** Red arrows indicate convolutional layers. Blue arrows indicate transposed convolutions (upsampling), and green arrows denote element-wise addition operators. (a) Pre-upsampling based approaches (e.g., SRCNN [9], VDSR [11], DRCN [12], DRRN [13]) typically use the bicubic interpolation to upscale LR input images to the target spatial resolution before applying deep networks for prediction and reconstruction. (b) Post-upsampling based methods directly extract features from LR input images and use sub-pixel convolution [14] or transposed convolution [15] for upsampling. (c) Progressive upsampling approach using the proposed Laplacian pyramid network reconstructs HR images in a coarse-to-fine manner.

Figure 1: LapSRN architecture.

2.4 Deep Laplacian Pyramid Network (LapSRN[5]) and Multi-scale Deep Laplacian Pyramid Network (MS-LapSRN [6])

2.4.1 Features

LapSRN in order to be able to reconstruct the SR image uses:

- **Charbonnier loss** because the L2 loss is not able to capture the underlying mapping of LR images to many HR images.
- **progressive reconstruction** of the SR image using the Laplacian Pyramid Framework [13].
- **residual learning**: learn the summation between the laplacian extracted by *features extraction branch* and the upscaled LR image in the *image reconstruction branch*

MS-LapSRN improve LapSRN introducing:

- **parameter sharing** across pyramid level and within pyramid levels in order to reduce the amount of parameters which increase with the scale (the greater the scale the deeper the network since there are more levels).
- **local skip connections** for avoiding the vanishing/exploding gradient prolem with the increase in the depth of the network.
- **multi scale training**: the previous network was trained for each scale different networks.

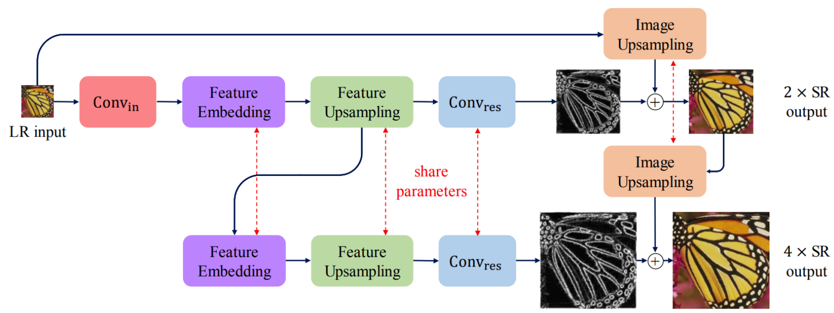

Fig. 3. **Detailed network architecture of the proposed LapSRN.** At each pyramid level, our model consists of a feature embedding sub-network for extracting non-linear features, transposed convolutional layers for upsampling feature maps and images, and a convolutional layer for predicting the sub-band residuals. As the network structure at each level is highly similar, we share the weights of those components across pyramid levels to reduce the number of network parameters.

Figure 2: MS-LapSRN architecture.

2.4.2 Architecture

In *LapSRN* [Figure 1] the **feature extraction branch** extract laplacian representations which are used for training (in an end-to-end fashion) the **image reconstruction branch** in order to reconstruct the SR image at the last level (due to the laplacian pyramid framework training on an higher scales lead to have also a network capable to resolve SR task for lower scale).

The residual learning allow the network to focus on high-frequency information (edges) instead of low-frequency ones.

In the *MS-LapSRN* [Figure 2] the feature extractor ($Conv_{in}$, Feature Embedding, Feature Upsampling, $Conv_{res}$) has always the same function: extract meaningful information from an input image in order to create a residual whose spatial dimension in 2x then the input one; therefore is logic to use same weights for doing so.

The *Feature embedding* has **R** recursive block [11] [12] which contains distinct **D** convolutional layers.

The structure of the recursive block [Figure 3b] use a skip connection directly connected to the input and a pre-activation inside the residual path [14].

2.4.3 Loss

Each level s has its own loss function and target ground truth.

The **Charbonnier loss** function is :

$$L_S = \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{L} \rho \times \left(\left(y_l^{(i)} - x_l^{(i)} \right) - r_l^{(i)} \right)$$

where S is the target upsampling factor, $\rho = \sqrt{x^2 + \epsilon^2}$ (Charbonnier penalty function which is a differentiable L1 norm), x_l is the upscaled LR image at level

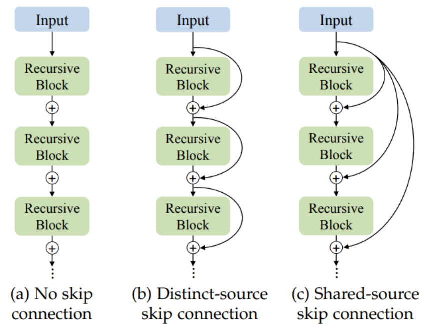

Fig. 4. **Local residual learning.** We explore three different ways of local skip connection in the feature embedding sub-network of the LapSRN for training deeper models.

(a) Differences between local skip connections.

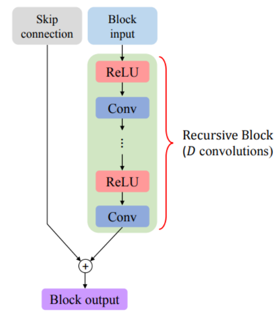

Fig. 5. **Structure of our recursive block.** There are D convolutional layers in a recursive block. The weights of convolutional layers are distinct within the block but shared among all recursive blocks. We use the pre-activation structure [41] without the batch normalization layer.

(b) Recursive block used in MS-LapSRN.

l, y_l is the downscaled HR target image (y) at level l and r_l is the residual at level l.

The final loss is the sum of the loss for each level trained using images with different scales (**multi scales training**).

2.4.4 Results

Here will be presented only the results of **MS-LapSRN**[6] since it's an evolution of LapSRN.

2.5 Magnification-Arbitrary Network (MetaSR [7])

References

[1] J. Shen, Y. Wang, and J. Zhang, "Asdn: A deep convolutional network for arbitrary scale image super-resolution," *ArXiv*, vol. abs/2010.02414, 2020.

[2] Y. Wang, J. Shen, and J. Zhang, "Deep bi-dense networks for image super-resolution," *CoRR*, vol. abs/1810.04873, 2018.

[3] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," *CoRR*, vol. abs/1807.02758, 2018.

[4] Y. Hu, J. Li, Y. Huang, and X. Gao, "Channel-wise and spatial feature modulation network for single image super-resolution," *CoRR*, vol. abs/1809.11130, 2018.

[5] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," *CoRR*, vol. abs/1704.03915, 2017.

[6] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *CoRR*, vol. abs/1710.01992, 2017.

[7] X. Hu, H. Mu, X. Zhang, Z. Wang, T. Tan, and J. Sun, "Meta-sr: A magnification-arbitrary network for super-resolution," *CoRR*, vol. abs/1903.00875, 2019.

[8] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1, pp. I–I, 2004.

[9] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *CoRR*, vol. abs/1501.00092, 2015.

[10] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *CoRR*, vol. abs/1812.10477, 2018.

[11] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," *CoRR*, vol. abs/1511.04491, 2015.

[12] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2790–2798, 2017.

[13] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *CoRR*, vol. abs/1603.05027, 2016.