Title

Author

Abstract

1 Introduction

Related Work （字不够再凑）

2 Methodology

(under-construction)

FSRCNN

FSRCNN[1] is also called fast-SRCNN. It is designed by Chao Dong, Chen Change Loy, and Xiaoou Tang, who first put forward applying deep learning neural network to super-resolution.

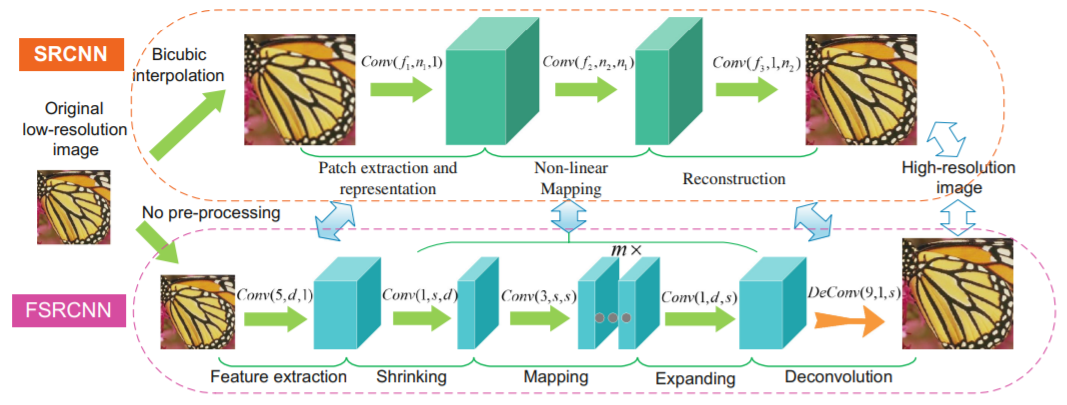


Figure from[1] : The network structure of SRCNN and FSRCNN.

Fast-SRCNN accelerating the current SRCNN[6]. SRCNN upscales LR image to target size by bicubic interpolation before actually entering the network. While FSRCNN just take LR image as the input, and conducts upsampling by the deconvolution layer at the end of the net work. This change leads to 2 main difference: 1. Directly, no need to cost time carrying out Bicubic operation. 2. Conv operation on LR image is less costly, With less computational complexity, FSRCNN can support a deeper network and achieve a better effect.

FSRCNN takes LR image as the input and carry out feature extraction. Then, replacing the non-linear mapping step, FSRCNN takes shrinking, mapping and expanding steps instead. And finally construct the desired HR image with a deconvolution operation.

FSRCNN take PReLU as the activation function and L2 as loss.

ESPCN

ESPCN[2] is also motivated by SRCNN[6], like FSRCNN[1], ESPCN takes LR image as the input and have low computational complexity.

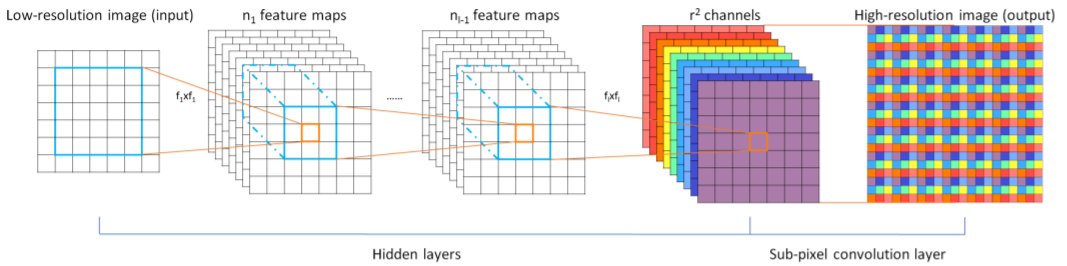


Figure from [2]: The network structure of ESPCN.

The author team implement a more efficient sub-pixel convolution layer for learn. ESPCN takes a LR image as the input. For a network with L layers, ESPCN will learn L-1 upscaling filters for every channel, and finally a deconvolution layer that recovering resolution of the image. The use of deconvolution layer had shown good effect in other in visual field and cause low cost, that’s why both FSRCNN and ESPCN choose to use deconvolution layer. And operation also enable cheaper convolution operation to be use in hidden layers.

ESPCN take tanh as activation function and L2 as loss.

LapSRN

LapSRN’s full name is Laplacian Pyramid Super-Resolution Network[3].

In comparison previous 2 model, LapSRN have deeper network structure, and has many different properties.

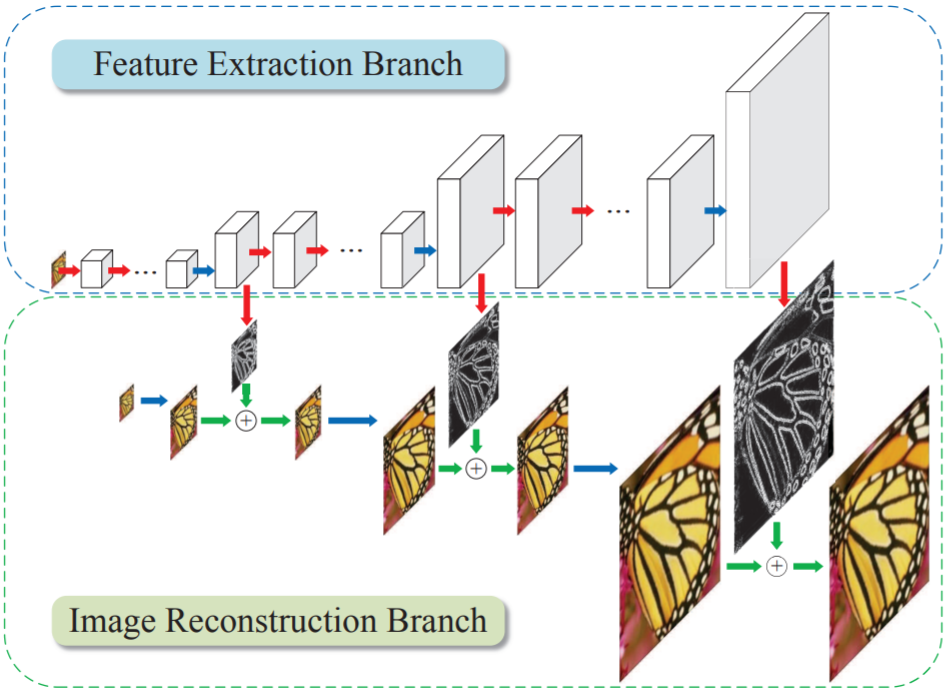


Figure from[3] : The network structure of LapSRN

Despite the detailed network implementation, LapSRN have 3 main characteristic.

1. LapSRN has multiple-level structure, every structure can conduct a 2x upscaling on input image and output of previous level can be taken as input of next level. As a result, LapSRN can support up to 8x upscaling, while most model only support 4x.

2. In each levels, feature extraction conduct first, then upscaling by a deconvolution layer to 2x size. LapSRN also motivated by Residual Learning[7], the upsampled image is then combined (using element-wise summation) with the predicted residual image from the feature extraction branch to produce a high-resolution output image.[3]

3. The author team think L2 loss is not good enough for SR learning and s inevitably

generate blurry predictions. So, LapSRN choose another loss: Charbonnier penalty function (a differentiable variant of L1 norm)[3]. This loss is considered at the end of every level.

EDSR

The design of EDSR is based on the SRResNet[8], which is motivated by ResNet[7] and achieve good performance in solving time/memory issue in SR.[4] What’s more EDSR has won NTIRE2017 SR Challenge.

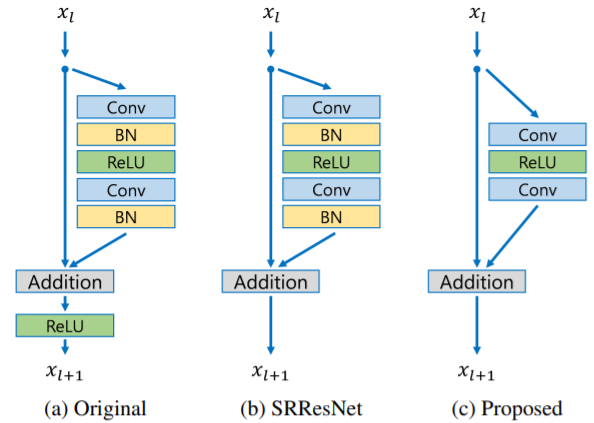


Figure from[4]: Comparison of residual blocks in ResNet, SRResNet and EDSR.

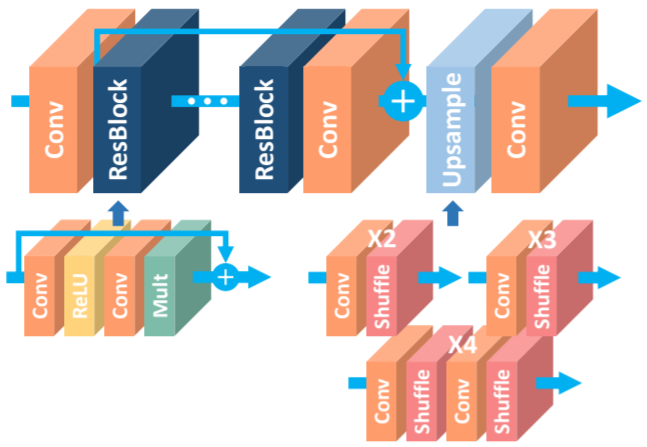


Figure from[4]: The architecture of the proposed single-scale SR network (EDSR).

The author team find the batch normalization layers get rid of range flexibility from networks by normalizing the features. Since SR is low-level computer visual problem. These BN block in original ResNet may not do good for SR. So, EDSR remove these BN layers is better, which further lead to approximately 40% of memory usage saving.[4]

EDSR use L1 loss instead of L2, because author team find L1 loss provides better conver-gence.

Real-ESRGAN

3 Experiments

Implementation Details

Model

Test set

Different method HR -> LR

Compare Strategy

Result

4 Conclusion

References (page-number under-construction)

[1] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In ECCV, 2016.

[2] W. Shi, J. Caballero, F. Huszar, J. Totz, A. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In CVPR, 2016.

[3] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang. Deep laplacian pyramid networks for fast and accurate superresolution. In CVPR, 2017.

[4] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In CVPRW, 2017.

[5] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. arXiv preprint arXiv:2107.10833, 2021.

[6] Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part IV. LNCS, vol. 8692, pp. 184–199. Springer, Heidelberg (2014)

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016.

[8] Ledig, C., Theis, L., Husz´ar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image superresolution using a generative adversarial network. In: CVPR (2017)