A Convolutional Neural Network with Two-Channel Input for Image Super-Resolution

Conference Paper · July 2019

DOI: 10.1109/IJCNN.2019.8852039

2 authors:

Purbaditya Bhattacharya
Helmut Schmidt University / University of the Federal Armed Forces Hamburg

Udo Zölzer
Helmut Schmidt University / University of the Federal Armed Forces Hamburg

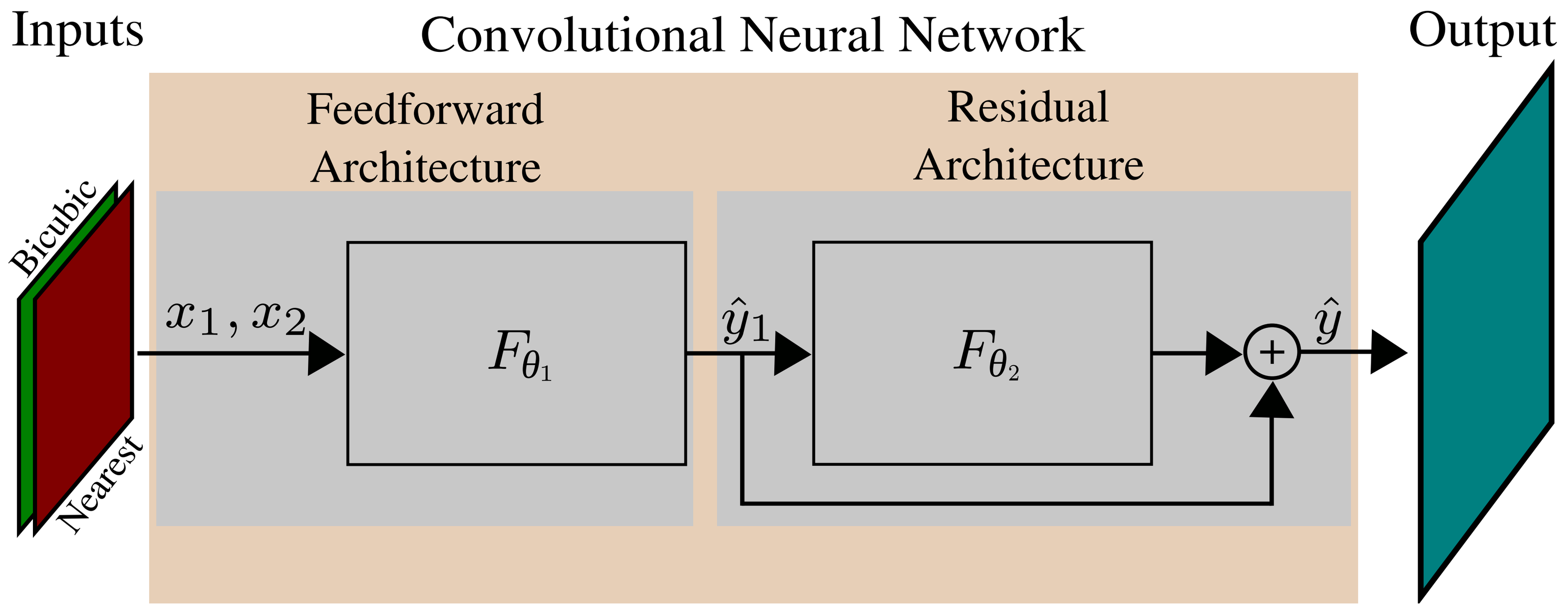
A Convolutional Neural Network with Two-Channel Input for Image Super-Resolution

Purbaditya Bhattacharya and Udo Zölzer
Dept. of Signal Processing and Communications, Helmut-Schmidt-University, Hamburg
bhattacp@hsu-hh.de

Abstract

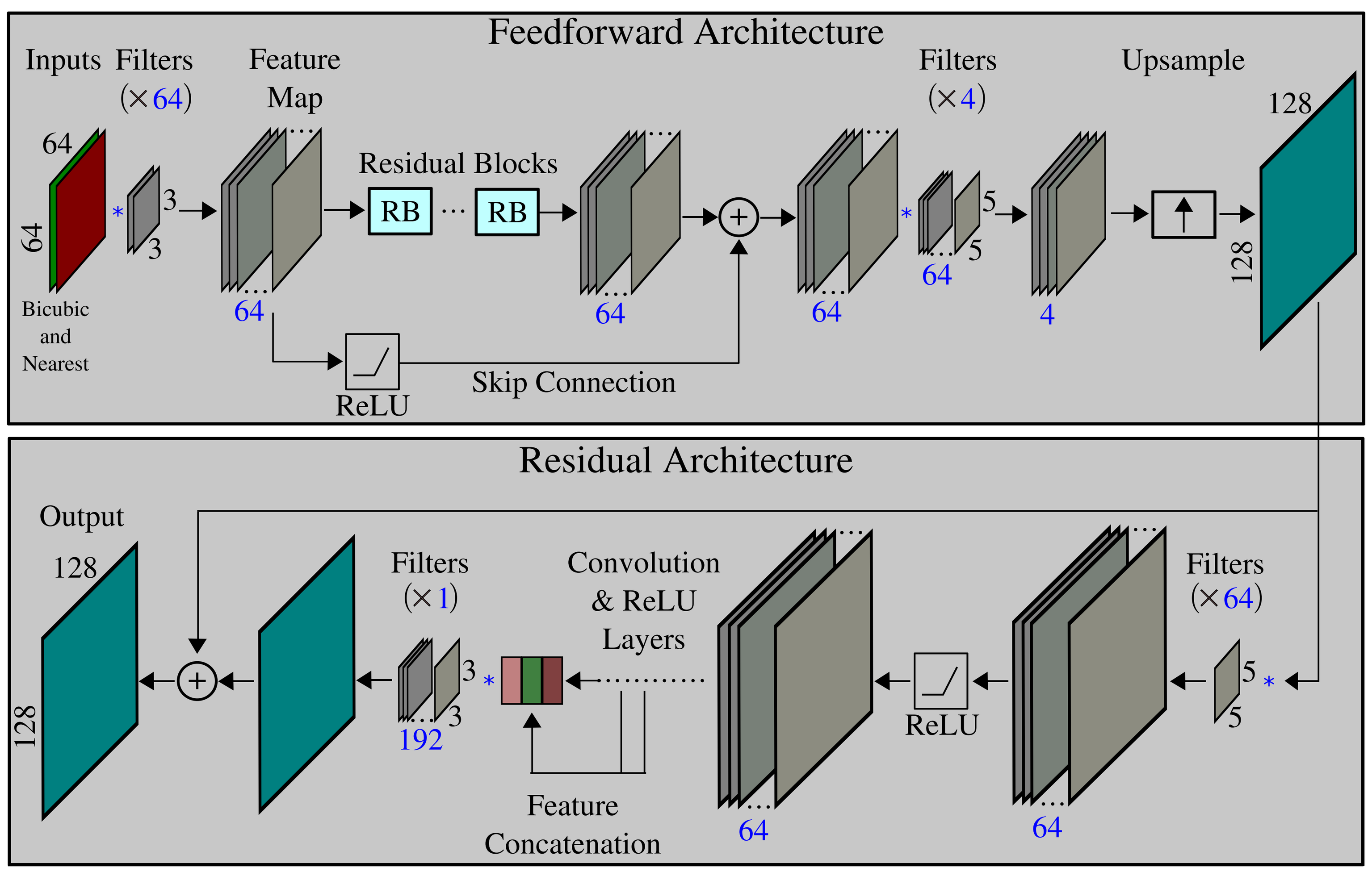
A convolutional neural network with two input channels is proposed for image super-resolution. Initially, the original image is downsampled with bicubic and nearest neighbour interpolation methods and the low-resolution image pair is used as the input to train the network. Additionally, the input channels are randomly swapped as an augmentation method during training. The proposed network contains a feedforward architecture followed by a residual network architecture.

Model Overview



$$\hat{y}_1 = F_{\theta_1}(x_1, x_2)$$
$$\hat{y} = \hat{y}_1 + F_{\theta_2}(\hat{y}_1)$$

Network Architecture

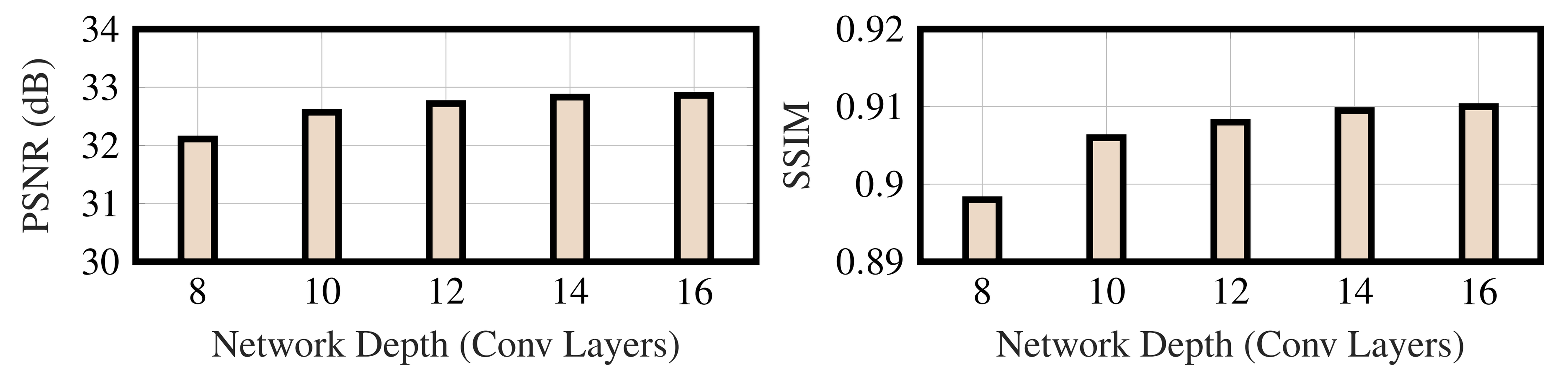


- The nearest neighbour patches provide a high-resolution (HR) image specific prior, constrains the solution possibilities, and reduce the ill-posedness of the super-resolution (SR) problem.
- The feedforward architecture contains one long skip connection, multiple residual blocks (RB) with short skip connections and a subpixel upsampling layer.
- The upsampling layer provides a coarse initial high resolution image and the following residual network structure obtains a smooth super-resolved image.
- The final network has 16 convolution layers yielding approximately 550k parameters to train.

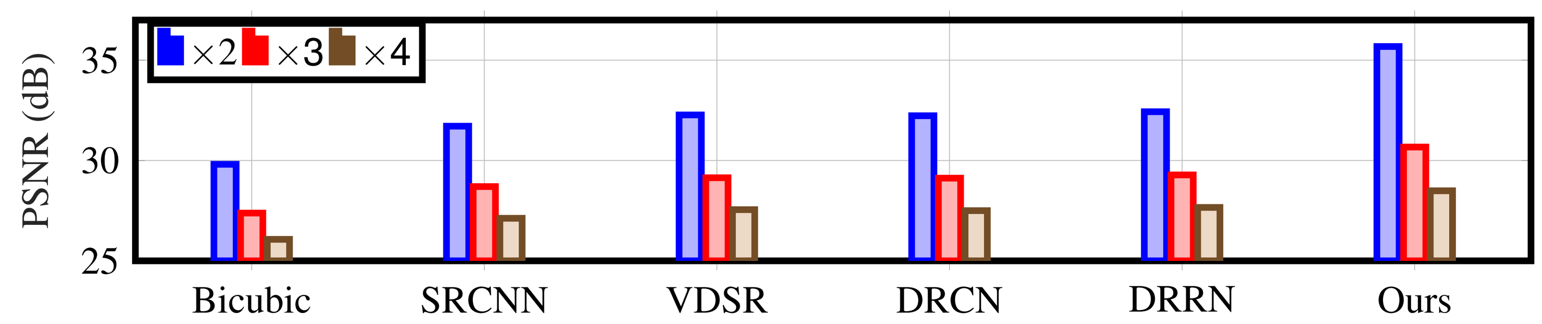
Network Performance

- Separate networks are trained for multiple SR scales.
- A combined loss function is used to train the network.

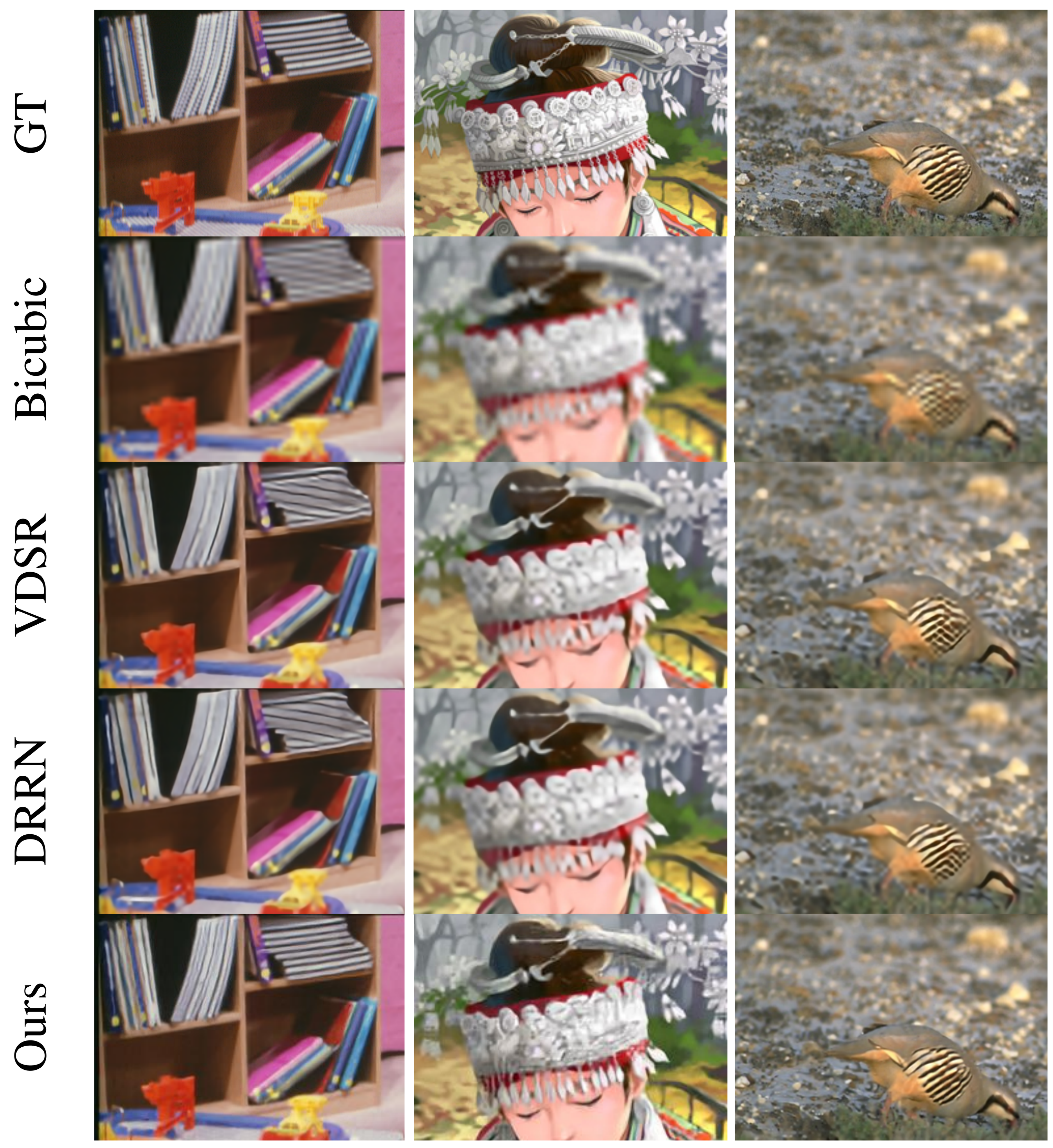
$$\theta^* = \underset{\theta}{\text{argmin}}\ E_{y,\hat{y}}\left[\sum_{p=1}^{2} \alpha_p L_p(y,\hat{y})\right],\ L_p(y,\hat{y}) = \frac{1}{N}\|y-\hat{y}\|_p^p$$



Average PSNR (dB) and SSIM values for a SR scale of $\times 4$ on benchmark dataset Set5, with different network depths.



Performance evaluation of our method for different SR scales ($\times 2$, $\times 3$, and $\times 4$) on the combined benchmark datasets Set5, Set14, and BSDS100, with respect to different single image super-resolution (SISR) methods.



The ground truth (GT) and examples of non-blind super-resolution for a scale of $\times 4$ with bicubic interpolation, very deep super resolution (VDSR) network, deep recursive residual network (DRRN) and our method.

- Relatively shallower networks are able to produce large enhancements with this approach.
- The proposed multi-channel SR method performs better than SISR methods and can be studied further to address blind SR problems.