Vessel preserving CNN-based image resampling of retinal images

Andrey Krylov[1], Andrey Nasonov[1], Konstantin Chesnakov[1], Alexandra Nasonova[1], Seung Oh Jin[2], Uk Kang[3], and Min Park[4]

[1] Laboratory of Mathematical Methods of Image Processing
Faculty of Computational Mathematics and Cybernetics
Lomonosov Moscow State University,
kryl@cs.msu.ru
[2] Korea Electrotechnology Research Institute
[3] Biomedical Research Institute, Seoul National University Hospital
[4] Department of Biomedical Science and Family Medicine, Seoul National University College of Medicine

Abstract. High quality resolution enhancement for eye fundus images is an important problem in medical image processing. Retinal images are usually noisy and contain low-contrast details that have to be preserved during upscaling. This makes the development of retinal image resampling algorithm a challenging problem.
The most promising results are achieved with the use of convolutional neural networks (CNN). We choose the popular algorithm SRCNN for general image resampling and investigate the possibility of using this algorithm for retinal image upscaling.
In this paper, we propose a new training scenario for SRCNN including training data preparation and a transfer learning. We demonstrate the improvement of image quality in terms of general purpose image metrics (PSNR, SSIM) and basic edges metrics — the metrics that represent the image quality for strong isolated edges.

1 Introduction

Image resampling is a process of generating a high-resolution (HR) image from a low-resolution (LR) image. It is used in a wide range of practical applications, especially in medical image processing and surveillance systems. Single image upscaling is an underdetermined problem since multiple solutions exist for the input low-resolution image. Thus, different constrains are essential for reducing the diversity of outputs.

One of the requirements for medical image enhancement algorithms is keeping image details. They should preserve non-smooth regions in the image and should not add artifacts like ringing effect or blur. When image details are corrupted, the results of further image analysis become unreliable and a diagnosis may become wrong.

One of the most important parts of retinal images are blood vessels. Thin blood vessels usually have low contrast. Also retinal images suffer from camera shot noise due to low light capturing conditions. Under these conditions, general purpose image resampling algorithms like bilinear, bicubic, Lanczos interpolation are not a good choice. Retinal images resampling algorithms should preserve as many details as possible and reduce noise during resampling.

Edge-directional image upscaling algorithms use the information about image edges to produce an adaptive image interpolation kernel at each pixel. Algorithms EGII [1], ICBI [2] and DCCI [3] use a combination of two directional kernels for pixel interpolation depending on the directions of edges in this pixel. They work well for straight and diagonal edges, but fail at image corners, textured regions with multiple directions and noisy areas.

A group of the state-of-the-art image resampling methods use a mapping transform between low-resolution (LR) image patches and high-resolution (HR) patches. Algorithm NEDI [4] obtains this mapping individually for each pixel from a self-similarity property of natural images at different scales. Learning-based methods obtain this transform from a training dataset containing LR and corresponding HR images. The method SI [5] classifies the LR patch into one of 625 classes and uses individual interpolation kernel for each class. Sparse representation is also used to construct the mapping transform between LR and HR patches [6, 7].

The most promising results are achieved using Convolutional neural networks (CNN) that are also used for finding LR-to-HR mapping function for image upscaling using example-based approach. The mapping function consists of a number of convolutional layers together with non-linear transformation. CNN based upscaling algorithms mostly differs in the number of convolutional layers and their spatial sizes, nevertheless the quality of the resulting HR images mostly depends on sufficiency of the training dataset and matching input image class with image classes from the training data.

Despite the fact that recent CNN models have reported outstanding results, they greatly depend on the training data. If the training data is not sufficient, the results may become unstable: small changes in the input image may result in significant changes in the output images. For example, if the image resampling algorithm has been trained using high-quality image set, it will try to recover the details from noise in case of noisy input image. This effect is strongly unwanted for medical image processing where generation of non-existing structures may produce an incorrect diagnosis. It also results in noise amplification that is highly noticeable in video resampling.

In this paper, we propose a retinal image resampling algorithm that is based on SRCNN algorithm, [8]. It uses the same CNN model but differs from SRCNN in training procedure.

2 Network model

We use the same model for CNN as described in SRCNN algorithm [8]. This model takes the low-resolution input image upscaled using bicubic interpolation with a certain factor and maps it to the high-resolution image by applying three convolutional filters with non-linear activation function.

The first layer of this function is a convolution of the input image with a filter W_1 of size $9 \times 9 \times 64$ + bias B_1 and an application of rectified linear unit (ReLU) after the convolution. Here input is a grayscale image and B_1 is a vector of size 64. In other words, on the first layer we apply 64 convolutions with 9×9 sized filters.

Let Y be the low frequency image which is magnified by the bicubic method. Then the first layer is calculated as

$$F_1(Y) = MAX(0, Y * W_1 + B_1). \tag{1}$$

The first layer extracts low-level structures such as edges of different orientations from the low resolution image.

On the second layer we apply ReLU to the convolution of $F_1(Y)$ with a filter W_2 of size $64 \times 5 \times 5 \times 32$ plus bias $B2$, here $B2$ is 32-dimensional:

$$F_2(Y) = MAX(0, F_1(Y) * W_2 + B_2). \tag{2}$$

It maps the features extracted at the previous step from the low resolution sub-space with corresponding features from the high resolution sub-space.

The third layer is used for image reconstruction. It acts like a weighed averaging of the high-resolution feature patches to single pixel. It is a convolution with $32 \times 5 \times 5$ dimensional filter W_3 plus bias B_3:

$$F_3(Y) = F_2(Y) * W_3 + B_3. \tag{3}$$

All these operations form a convolutional neural network, which we name $F(Y, \Theta)$, where Θ is the network filter and bias coefficients.

3 Training procedure

Training procedure consists in finding the coefficients Θ that minimize the loss between the reconstructed images $F(Y, \Theta)$ and the ground truth high resolution images on the set of LR-HR pairs $(X_i, Y_i), i = 1, \ldots N$ from the training set. We use Mean Squared Error (MSE) as the loss function:

$$L(\Theta) = \sum_{i=1}^{N} \|F(Y_i, \Theta) - X_i\|^2 \tag{4}$$

This leads to higher PSNR values as an objective metric. PSNR is a widely-used metric for quantitatively evaluating image interpolation quality. Though it has weak correlation with human perception, the minimization of the MSE-based loss function produces satisfactory upscaling results, even if they are assessed using other objective metrics, e.g., SSIM, MSSIM.

The loss is minimized using stochastic gradient descent with the standard backpropagation. For the training procedure we have used Caffe package [9]

3.1 Training dataset

The most obvious idea is to take real retinal images and to build the training dataset of high-resolution retinal images and its downsampled versions. But this idea has led to poor results. Retinal images have many flat and noisy areas and few vessels. This makes the training dataset imbalanced.

Another idea is to exclude flat regions from training dataset and to train CNN only on patches containing blood vessels and other structures. But experiments have shown that this dataset is inconsistent too and the resampling quality is still not good enough.

After a series of experiments, we have come to the following algorithm:

1. We take a set of photographic images and perform normal training procedure. We have used a collection of 124 photographic images of nature, buildings and humans (WebShots Premium Collections, October 2007) with average resolution 1600×1200.

2. Then we fix the coefficients in the first and the second layers, take a set of retinal images and optimize the coefficients of the third layer only. We have used the images from DRIVE [10] database. This database consists of 20 retinal images with blood vessel masks. We use these masks to keep only patches containing vessels.

3.2 Handling noise and blur

Comparing to photographic images, retinal images are more noisy and slightly more blurry. To make the results of retinal image resampling better, the training images should be preprocessed.

We apply Zero Component Analysis (ZCA) transformation to the reference images [11, 12]. ZCA enhances important details such as edges and textures and also helps to remove unwanted high-frequency noise by normalizing the variance of the data in the direction of each eigenvector of its covariance matrix. We apply ZCA transformation globally to the whole image, and the output of this transformation is of the same size as input. ZCA preprocessing step is widely used for image classification problems, and we adopt this technique for CNN-based image resampling.

The idea of adding noise to HR images only without affecting LR images has shown the effectiveness for noisy image upscaling [13]. To take into account both blur and noise effects, we apply Gaussian blur with $\sigma = 0.5$ to HR images and then add Gaussian noise with $\sigma = 6$.

4 Evaluation

We have checked our algorithm on DiaretDB database [14]. We take 100 images, downsample them with a factor of 2 using bicubic interpolation and then upsample them using the proposed method. The results are compared with reference images using objective metrics PSNR and SSIM [15]. Also we measure the sharpness value.

4.1 Blur estimation

One of the properties that correlates to subjective image quality is sharpness. We use the edge width estimation algorithm [16] that measures the sharpness of the basic edges [17] — strong edges that are the most suitable for image quality analysis.

Figure 1 shows an example of basic edge detection for a retinal image.

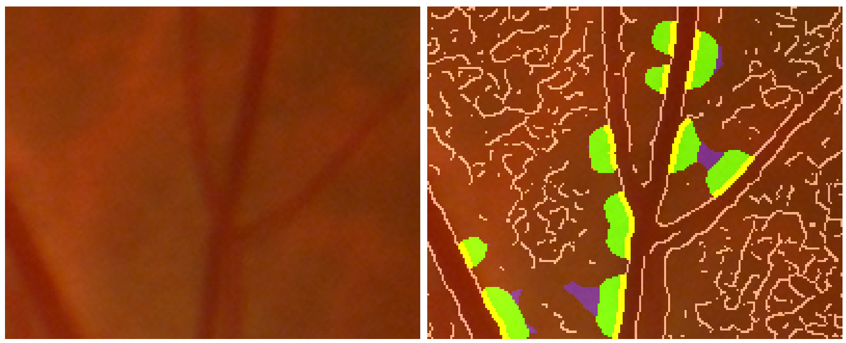

Fig. 1. The basic edges for a retinal image. The basic edges are marked with yellow color and its neighborhood — with green color.

4.2 Results

The results of the application of the SRCNN algorithm with different training strategies to retinal images are shown in Figure 2. Table 1 shows the average metric values calculated over DiaretDB retinal database [14].

It can be seen that the original SRCNN trained by its authors amplifies the noise and reveals its pattern structure. Adding noise to the HR images from the training set results in lower noise level. Applying the ZCA transform and Gaussian blur to the HR images results in increased sharpness. Transfer learning further improves the image quality — its effect appears as slightly better vessel contrast.

5 Conclusion

A novel training procedure for CNN-based upscaling algorithm for retinal images has been proposed. The best results have been achieved using a combination of the special training dataset preparation and the transfer learning. The proposed ideas have been applied to SRCNN algorithm and demonstrated its effectiveness with the retinal images from standard database. It can added that the proposed data preparation procedure can be used for retinal image enhancement with other CNN models.

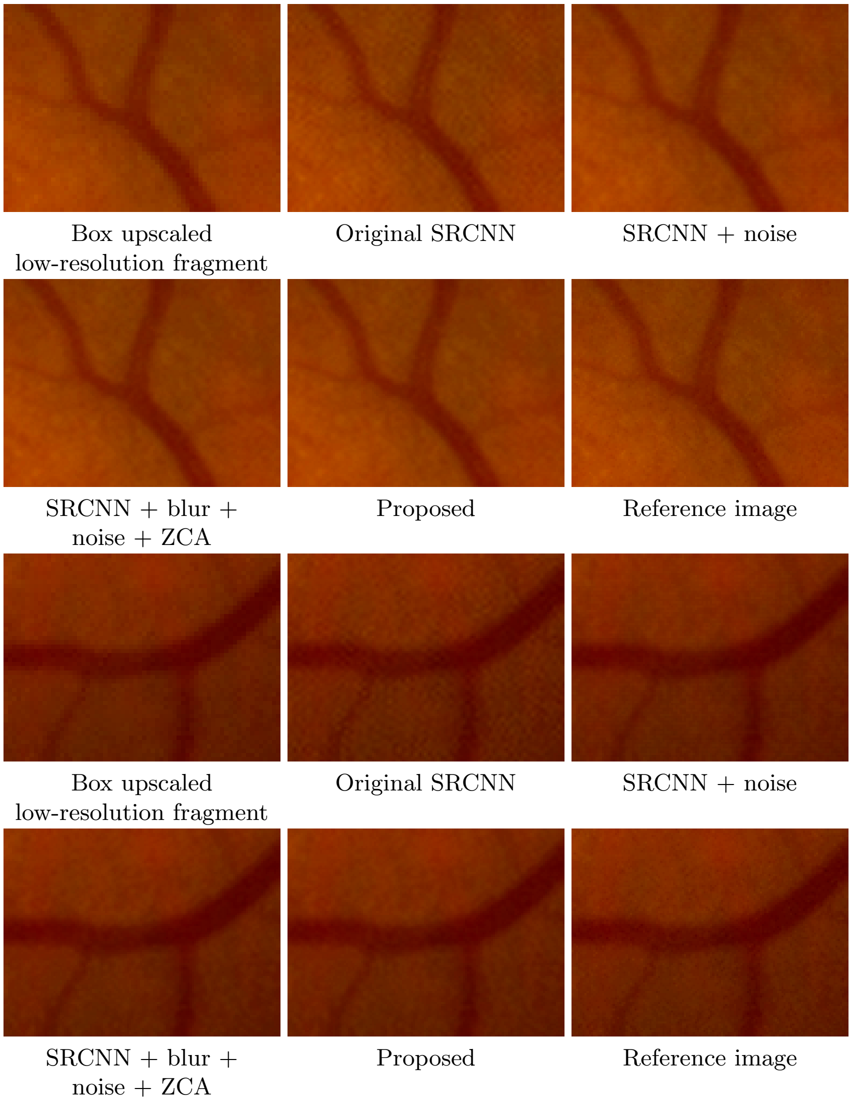

Fig. 2. The result of retinal image upscaling.

References

1. Zhang, L., Wu, X.: An edge-guide image interpolation via directional filtering and data fusion. IEEE Transactions on Image Processing **15** (2006) 2226–2235
2. Giachetti, A., Asuni, N.: Real time artifact-free image interpolation. IEEE Transaction on Image Processing **20**(10) (2011) 2760–2768
3. Zhou, D., Shen, X., Dong, W.: Image zooming using directional cubic convolution interpolation. IET Image Processing **6**(6) (2012) 627–634

Method	PSNR	SSIM	Blur level
Reference	—	—	1.20
Original SRCNN	45.73	0.972	1.32
SRCNN + noise	47.58	0.987	1.29
SRCNN + noise + ZCA	48.42	0.988	1.25
SRCNN + blur + noise + ZCA	48.65	0.988	1.21
Proposed (ZCA, noise, blur, transfer learning)	48.74	0.990	1.19

Table 1. The average metric values for the DiaretDB images for different SRCNN training procedures.

4. Li, X., Orchard, M.: New edge-directed interpolation. IEEE Transactions on Image Processing **10** (2001) 1521–1527
5. Choi, J.S., Kim, M.: Super-interpolation with edge-orientation based mapping kernels for low complex 2x upscaling. IEEE Transactions on Image Processing **25**(1) (2015) 469–483
6. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE Transactions on Image Processing **19**(11) (2010) 2861–2873
7. Dong, W., Zhang, L., Lukac, R., Shi, G.: Sparse representation based image interpolation with nonlocal autoregressive modeling. IEEE Transactions on Image Processing **22**(4) (2013) 1382–1394
8. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. Computer Vision–ECCV 2014 (2014) 184–199
9. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093 (2014)
10. Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., van Ginneken, B.: Ridge based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging (2004) 501–509
11. Krizhevsky, A.: Masters Thesis "Learning multiple layers of features from tiny images". `www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf` (2009)
12. Nasonov, A., Chesnakov, K., Krylov, A.: CNN based retinal image upscaling using zero component analysis. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2017) 27–31
13. Nasonov, A., Chesnakov, K., Krylov, A.: Convolutional neural networks based image resampling with noisy training set. In: International Conference on Signal Processing (ICSP2016), Chengdu, China (2016) 62–66
14. Kälviäinen, R., Uusitalo, H.: DIARETDB1 diabetic retinopathy database and evaluation protocol. In: Medical Image Understanding and Analysis. (2007)
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4) (2004) 600–612
16. Nasonov, A.V., Nasonova, A.A., Krylov, A.S.: Edge width estimation for defocus map from a single image. Lecture Notes in Computer Science **9386** (2015) 15–22
17. Nasonov, A.V., Krylov, A.S.: Edge quality metrics for image enhancement. Pattern Recognition and Image Analysis **22**(1) (2012) 346–353