Convolutional neural networks based image resampling with noisy training set

Andrey Nasonov, Konstantin Chesnakov, Andrey Krylov
Laboratory of Mathematical Methods of Image Processing
Faculty of Computational Mathematics and Cybernetics
Lomonosov Moscow State University
119991, Russia, Moscow, Leninskie Gory, MGU BMK
Email: kryl@cs.msu.ru

Abstract—A new learning model for image resampling with convolutional neural network is proposed. Its main idea is the dataset preparation method for deep learning. The proposed algorithm can work with noisy and noiseless images and provides good quality for wide noise level range. The method was tested using standard datasets and was also applied for retinal image resampling.

Keywords: image resampling, convolutional neural networks

I. INTRODUCTION

Image interpolation is a classical problem of image processing. Its goal is to recover a high-resolution image from a single low resolution image. Image interpolation algorithms are widely used in a large set of devices and applications, such as medical image processing tools, where the input is often a noisy low resolution image.

This problem is ill-posed since multiple solutions exist for a given low-resolution image. So it is an underdetermined inverse problem. Such problems are typically solved by adding constraints based on prior information.

The simplest algorithms are general purpose linear interpolation methods like bilinear, bicubic, and Lanczos interpolation. These methods represent the neighborhood of each pixel as a polynomial function [1]. They are very fast but do not perform edge-directed interpolation. Usually, a combination of artifacts (blur, ringing and aliasing) comes with linear interpolation.

A large class of edge-directional image interpolation algorithms use prior information about the images. These algorithms assume that the edges look similar at different resolutions. The method [2] improves linear interpolation by kernel elongation along edges. It produces excellent results at straight edges but does not perform well for image corners and textured areas containing multiple directions. Algorithms EGII [3], ICBI [4] and DCCI [5] consider two directions at each pixel using a combination of two directional interpolation results. Algorithm NEDI [6] uses self-similarity of the natural images at different scales to calculate the interpolation coefficients. It works well at both edges and corners but corrupts the structures without self-similarity.

Regularization-based algorithms represent the image resampling problem as a computationally expensive functional minimization problem with data-fitting and stabilizer terms [7]–[9]. The stabilizer holds prior information about the image. Total variation stabilizer is usually used. It smooths the image while keeping the edges sharp.

Example-based strategy is used to find the image constraints that cannot be easily expressed by mathematical models. These methods use internal similarities in the same image or learn mapping functions from the patches from low- and high-resolution training images [10], [11]. The method family SI [12] puts the \(3 \times 3\) patch into one of 625 classes and uses individual interpolation kernel for each class.

Convolutional neural network that directly learns end-to-end conversion between low- and high-resolution images [13] also uses the mapping functions. In these methods all data of convolutional layers is fully obtained through learning with little pre/post processing. During image upscaling these methods don’t need to solve any optimization problems. Quality of the resulted high-resolution images depends on the convolutional model scale, neural network size and the number of iterations during learning process.

The result of example-based image upscaling method greatly depends on the training data. If the training data is not sufficient, the result may become unstable: small changes in the input image may result in high changes in the output image. 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 the appearance of non-existing structures may produce incorrect diagnosis. It also results in noise amplification that is highly noticeable in video resampling.

In this work, we create a new learning model for image resampling with convolutional neural network. The main idea of the proposed model is the dataset preparation method for deep learning. Simple taking noisy high-resolution images as the reference images in the training set result in strong instability and poor quality of the interpolated images. In our approach noise is added only to low-resolution images. It results in resampling quality improvement as well as noise reduction.
II. PROPOSED METHOD

A. Network model

We use the same model for CNN as described in SRCNN method [13]. In this scheme, considering a single low-resolution image, we first upscale it to the desired size using bicubic interpolation, but we can still consider this image a low-resolution image. After that we apply 3 layers, where the first layer is a convolution of the input image with a filter of size $9 \times 9 \times 64$ plus bias $B_1$ and an application of rectified linear unit (ReLU) to this 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 $9 \times 9$ sized filters. On the second layer we apply ReLU to the convolution of first layer’s results with a filter of size $64 \times 5 \times 5 \times 32$ plus bias $B_2$, here $B_2$ is 32-dimensional. The third layer is used for image reconstruction. It is a convolution with $32 \times 5 \times 5$ dimensional filter plus bias $B_3$. All these operations form a convolutional neural network, which we name $F(Y, \Theta)$, where $Y$ is a low-resolution input image, $\Theta$ is the network filters’ and biases’ coefficients. So our goal is to recover a high-resolution image from the low-resolution image using this network and the result should be consistent to the noiseless images and images with different levels of Gaussian noise.

B. Training method

We have to find the network parameters $\Theta$ producing the appropriate result for images with and without noise. This is achieved through minimizing the loss between reconstructed images $F(Y, \Theta)$ and ground truth high resolution images. A training image set containing high-resolution images $\{X_i\}$ and their matching low-resolution images $\{Y_i\}$ is used. To get higher PSNR values as an objective metric, we use Mean Squared Error (MSE) as the loss function:

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

Images with a lot of high-frequency information from the target image class are taken as the ground truth images $\{X_i\}$. In order to make the algorithm effective and stable to noisy input images, we use the following method to generate low-resolution images for scale factor $s$:

1. Add Gaussian noise with standard deviation $\sigma_n$.
2. Apply Gaussian filter with radius $\sigma_s = \sigma_0 \sqrt{s^2 - 1}$ for aliasing suppression, $\sigma_0 = 0.3$.
3. Perform image decimation by taking each $n$-th pixel. This process can be formulated as:

$$[Y_i]_{x,y} = [(X_i + n \sigma_n) \ast G_{\sigma_s}]_{sx, sy}.$$

III. EXPERIMENTS

Experiments have been performed for image resampling with a factor of 2. The proposed method has been compared with state-of-the-art image resampling algorithms: A+ [11], original SRCNN [13], DCCI [5], SI-3 [12] and bicubic interpolation. We have used a collection of 124 photographic images of nature, buildings and humans (WebShots Premium Collections, October 2007) as the high-resolution training image set for the proposed method with noise level $\sigma_n = 8$. The same image set has been used to train SI-3 algorithm. The method A+ has been used as provided by its authors without changes.

The sets Set5 and Set14 [13] have been used to demonstrate the effectiveness of the proposed method. Gaussian noise has been added to each image. The standard deviation of the noise is increased uniformly from 0 to 10 with step 0.1.

The dependence of algorithm performance against noise level is shown in Fig. 4 and Fig. 5. It can be seen that original SRCNN algorithm trained on noiseless images produces low quality results for noisy input images while the proposed method outperforms state-of-the-art algorithms for both noisy and noiseless images.

The Fig. 1 contains the average PSNR and SSIM results for images from the test datasets for multiple noise levels. The resampling results for ‘clown’ image with noise $\sigma = 3.4$ are shown in Fig. 2. It can be seen that the proposed algorithm effectively increases the image resolution. At the same time, noise suppression is observed. The cause of this effect is the use of training dataset containing noisy low-resolution image patches with corresponding noiseless high-resolution patches.

A. Retinal image upscaling

The CNN-based approach looks very important to be used for retinal image processing and analysis. The problem of retinal image resampling is often an important preprocessing problem in order to make the evaluation consistent with images in different sizes like in [14] where bicubic interpolation method is used. And the diagnosis results can depend seriously on the resampling method. Below in Fig. 3 two examples of the retina resampling illustrate the difference of the CNN-based and bicubic methods.

It is necessary to mention that the CNN-based retinal image analysis algorithms like [15] started to outperform other classification algorithms. It looks promising to use the CNN-based resampling approach within or together with these algorithms.

IV. CONCLUSION

The proposed algorithm performs high quality image upscaling as well as effective noise suppression. It looks very promising to be used for medical image resampling within CNN-based medical image analysis. The proposed algorithm gives the best result compared to other resampling methods but at the same time it is not fast enough for realtime applications. It is about 10 times slower than other resampling methods using mapping functions. Nevertheless, it can be expected that in a couple of years it will be possible to use it in realtime applications.

ACKNOWLEDGMENT

The work was supported by Russian Science Foundation grant 14-11-00308.
Fig. 1. The average PSNR and SSIM results for the test datasets for resampling with a factor of 2.

REFERENCES


Fig. 2. Application of image upscaling for 'clown' image with noise $\sigma = 3.4$ with a factor of 2.

Fig. 3. The result of retinal image resampling.
Fig. 4. The values of objective metrics against noise level for 'bird' image from Set5. The vertical axis is the metric value (PSNR or SSIM), the horizontal axis is the noise level. The value 0 corresponds to noiseless images while the value 100 corresponds to Gaussian noise with $\sigma = 10$.

Fig. 5. The values of objective metrics against noise level for 'bridge' image from Set14.