High-frequency Restoration using Deep Belief Nets for Super-resolution

Toru NAKASHIKA[1,a)] Tetsuya TAKIGUCHI[1,b)] Yasuo ARIKI[1,c)]

1. Introduction

Super-resolution techniques are generally divided into two approaches: example-based methods and statistical methods. Example-based methods [1] simply use (or select in sparce coding [2]) pairs of low-resolution and high-resolution patches for the reconstruction. In this approach, a low-resolved input image is decomposed into patches, each of which is compared with the patches in the database and replaced with the corresponding high-resolved patch. Although this approach produces relatively less-deteriorated images, it is not based on any statistical models and lacks versatility. For the statiscical approach, various methods have been proposed so far: the eigen-space BPLP [3], the MRF-based approach [4], a GMM-based approach [5], and so on. Some of these statistical approaches rely on the training of the correspondence relationships between low-resolved/high-resolved images. Therefore, if one wants to enlarge an image with the desired scale, the relationships between the low and high resolution with that scale need to be trained beforehand.

Meanwhile, Hinton et el. introduced an effective training algorithm of Deep Belief Nets (DBNs) in 2006 [6], and the use of DBNs rapidly spread in the field of signal processing with great success. DBNs are probabilistic generative models that are composed of multiple layers of stochastic latent variables, and have a greedy layer-wise unsupervised learning algorithm. DBNs are not only used for classification tasks, but also for the completion of an image or for collaborative filtering. Eslami et el. adopted a type of DBNs (called ShapeBM) to complete the missing region in an image [7]. Salakhutdinov et el. used 2-layer DBNs (i.e., Restricted Boltzmann Machines; RBMs) for collaborative filtering [8], which has the benefit of the DBNs dealing with missing data.

In this paper, we propose a novel super-resolution method using DBNs to restore the missing high-frequencies (Fig. 1), motivated by the above-mentioned characteristics of DBNs. In our approach, a low-resolved image is first scaled up to the prescribed size by using bicubic interpolation, and the high-frequency information is estimated by inference of trained DBNs. The networks are trained only using high-resolved image patches in a multiple-layer-wise unsupervised manner, so as to find the deep relational connections between spatial frequencies. Thus, we expect that the self-trained DBNs capture the high-order dependencies of low-frequencies and high-frequencies, and complete the high-frequency components of a low-resolved image, assuming that the low-frequency components are the same.

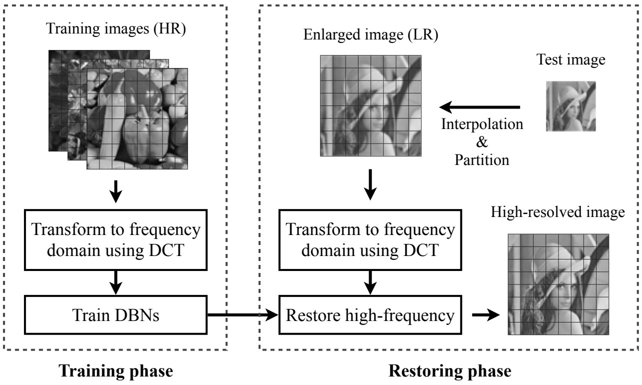


Fig. 1 System flowchart of our proposed method.

2. High-frequency restortion

In this paper, we employ Deep Belief Nets (DBNs) for capturing the co-occurrence relationships among DCT coefficients based on joint probability, expecting that the DBNs can capture even higher-order connections between the frequencies. Once the networks are constructed, the lost high-frequency components can be restored based on the co-occurrence.

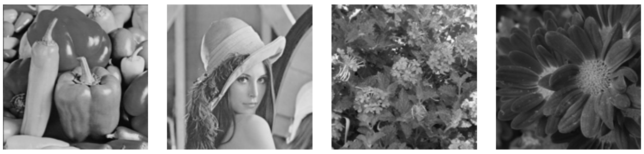
In the literature of RBMs, the joint probability $p(\boldsymbol{v},\boldsymbol{h})$ of real-valued visible units $\boldsymbol{v} = [v_1, \cdots, v_I]^T, v_i \in \mathcal{N}(0,1)$ (note that the training data should be first normalized for each dimension to have zero mean and unit variance) and binary-valued hidden units $\boldsymbol{h} = [h_1, \cdots, h_J]^T, h_j \in \{0,1\}$ is defined as:

$$p(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{Z}\exp(-E(\boldsymbol{v},\boldsymbol{h})) \quad (1)$$

$$E(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{2}|\boldsymbol{v}|^2 - \boldsymbol{c}^T\boldsymbol{h} - \boldsymbol{v}^T\boldsymbol{W}\boldsymbol{h} \quad (2)$$

where, Z is the normalizing constant, and $\boldsymbol{W} \in \mathbb{R}^{I\times J}$, $\boldsymbol{c} \in \mathbb{R}^{J\times 1}$ are a weight matrix between visible units and hidden units, a bias vector of hidden units, respectively.

[1] Department of System Informatics, Kobe University, Rokkodai 1–1, Nada-ku, Kobe, 657–8501 Japan
a) nakashika@me.cs.scitec.kobe-u.ac.jp
b) takigu@kobe-u.ac.jp
c) ariki@kobe-u.ac.jp



(I) (II) (III) (IV)

Fig. 2 Four input images. Only (I) is used for training, and the others for testing.

Since there are no connections between visible units or between hidden units, the conditional probabilities $p(\boldsymbol{h}|\boldsymbol{v})$ and $p(\boldsymbol{v}|\boldsymbol{h})$ form simple equations as follows:

$$p(h_j = 1|\boldsymbol{v}) = \sigma(c_j + \boldsymbol{v}^T \boldsymbol{W}_{:j}) \tag{3}$$

$$p(v_i|\boldsymbol{h}) = \mathcal{N}(\boldsymbol{W}_{i:}\boldsymbol{h}, 1) \tag{4}$$

where $\boldsymbol{W}_{:j}$ and $\boldsymbol{W}_{i:}$ denote the j-th column vector and the i-th row vector, respectively. $\sigma(x)$ indicates sigmoid function, i.e. $\sigma(x) = 1/(1+\exp(-x))$. For the parameter estimation, the log likelihood of visible units is used as an evaluation function. Although the gradient is intractable to compute, contrastive divergence [6] can be used to approximate it.

In the training of DBNs, the hidden units of the current stack are regarded as visible units in the next layer. This procedure is repeated layer-by-layer until the highest layer.

Once the weights of the networks are estimated, the almost-zero-valued high-frequency components are restored given a low-resolved bicubic-interpolated image. At this point, the values of the high-frequency components are inferred by using Eq. (3) and (4).

3. Experiments

For the training of Deep Belief Nets (DBNs), we used image (I) shown in Fig. 2, whose size is 512×512. We partitioned the image into patches to have the size of 16×16, allowing overlaps. Each patch (in total 15625 patches) was transformed by 2-dimensional DCT, normalized, and then fed to DBNs. We trained DBNs with a learning rate of 0.01 for 500 epochs, which have 2-hidden layers, 400 hidden units for the first layer and 200 hidden units for the second layer.

For testing, 3 images (Fig. 2(II)(III)(IV)) were reduced by half ($s = 2$) from 512×512 in the horizontal and vertical directions, and enlarged by two times using our proposed method.

To evaluate the efficacy of our method, we compared it with 2 conventional methods (sparse-coding [2] and GMM [5]) and bicubic interpolation with 2 measures (PSNR and SSIM [9]). Given an original image $\mathbf{Y}$ (high-resolved image) and its processed image $\mathbf{E}$, PSNR and SSIM measure the quality of the processed image. The larger the values of PSNR and SSIM are, the higher the quality of the images is supposed to be. For reference, we also compared within our methods to different architecture of DBNs: 1-layer, 400 hidden units (i.e. RBMs).

Table 1 summarizes the experimental results, comparing our proposed method with bicubic interpolation, GMM, and Sparse-Coding. As shown in Table 1, the proposed method using DBNs performed best for each test image with either measure. Furthermore, 2-hidden-layer DBNs (Proposed(DBNs)) outperformed 1-hidden-layer DBNs (Proposed(RBMs)). The architecture of the deep DBNs captures higher-order dependencies between low and high frequencies better than the other methods including shallow DBNs, and we consider that this ends up with the preferable results.

Table 1 Comparison of super-resolution methods using PSNR and SSIM.

Image	Method	PSNR	SSIM
(II)	Bicubic	36.43	0.8816
	Sparse-Coding	37.71	0.9084
	GMM	37.98	0.9272
	Proposed(RBMs)	38.52	0.9289
	Proposed(DBNs)	**38.60**	**0.9308**
(III)	Bicubic	33.07	0.8015
	Sparse-Coding	34.20	0.8608
	GMM	35.59	0.9154
	Proposed(RBMs)	35.73	0.8639
	Proposed(DBNs)	**37.60**	**0.9067**
(IV)	Bicubic	38.15	0.8977
	Sparse-Coding	39.68	0.9240
	GMM	40.83	0.9460
	Proposed(RBMs)	40.40	0.9452
	Proposed(DBNs)	**41.31**	**0.9548**

4. Conclusion

In this work, we proposed the use of Deep Belief Nets (DBNs) to tackle super-resolution, replacing the task with the completion problem of the missing data. In our approach, the missing high-frequency components in a low-resolved image are restored using self-trained DBNs in the spatial frequency domain. In our experiments, we showed the efficacy of the proposed method, in comparison to conventional methods.

References

[1] Freeman, W.T., Jones, T.R., and Pasztor, E.C., "Example-Based Super-Resolution," IEEE Computer Graphics and Applications, vol. 22 no. 2, pp. 56-65, 2002.

[2] Mario A. T. F. M. E. and Yi Ma, "On the role of sparse and redundant representations in image processing," Proceedings of the IEEE, vol. 98, no. 6, pp. 972-982, 2010.

[3] Kawano, H., Suetake, N., and Maeda, H., "Image Enlargement with Lost High-Frequency Components Estimation Using Clustered Eigenspace-BPLP," Image and Video Technology (PSIVT), pp. 404-407, 2010.

[4] Deepu R. and Subhasis C., "An MRF-Based Approach to Generation of Super-Resolution Images from Blurred Observations," Journal of Mathematical Imaging and Vision, vol. 16, pp. 5-15, 2002.

[5] Y. Ogawa, Y. Ariki, and T. Takiguchi, "Super-resolution by GMM based conversion using self-reduction image," International Conference on ICASSP, pp. 1285-1288, 2012.

[6] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," Neural Computation, vol. 18, pp. 1527—1554, 2006.

[7] S. M. Eslami, N. Heess, and J. Winn, "The Shape Boltzmann Machine: a Strong Model of Object Shape," IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[8] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," in Proc. International Conference on Machine Learning, pp. 791-798, 2007.

[9] Z. Wang, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. on Image Processing, vol. 13, no. 4, pp. 600—612, 2004.