Image inpainting based on stacked autoencoders

O Shcherbakov[1] and V Batishcheva[2]

[1]Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, 49 Kronverkskiy pr., Saint Petersburg 197101, Russia
[2]Saint-Petersburg State University, 7-9 Universitetskaya nab., St.Petersburg 199034, Russia

E-mail: scherbakovolegdk@yandex.ru, elokkuu@gmail.com

Abstract. Recently we have proposed the algorithm for the problem of image inpaiting (filling in occluded or damaged parts of images). This algorithm was based on the criterion spectrum entropy and showed promising results despite of using hand-crafted representation of images. In this paper, we present a method for solving image inpaiting task based on learning some image representation. Some results are shown to illustrate quality of image reconstruction.

1. Introduction
Image inpainting problem (filling in missing or damaged regions in images) presenting its own theoretical interest has also wide field of applications. Such applications could be, for example, image recovery after lossy image transmission or communication and different kinds of painting restorations. Some partial solutions for this task have already been developed. There are methods based on wavelet representation of images [1], variational analysis [2], structural features [3], interpolation models, algorithmic probability (ALP) [4] and etc. Mostly, these methods use particular hand-crafted image representation. However under ALP approach the image inpainting task has been stated for generalized image representation [4], and similar approach has been investigated for representation learning [5]-[6], but not in the context of image inpainting. Thus the problem of learning representations is of interest here.

One of the approaches to automatic feature learning from unlabeled data is stacked autoencoder (AE) that presents a special type of artificial neural networks (ANN). This approach is interesting for image inpainting task because architecture of AE allows not only encoding some features in its hidden units but also reconstructing an input pattern when decode mode of AE is on. However, this feature of AE does not allow one to solve image inpainting task directly. Nevertheless, this approach to feature learning can be extended to perform image inpainting.

2. Autoencoders for learning representation
Autoencoders present the unsupervised learning algorithm that sets output values to be equal to the inputs. In other words, AE tries to learn approximation to identity function so as output vector is going to be similar to input vector. Such identity function demands some constraints on network structure. One of these constraints is to limit the number of hidden units which forces network to learn compressed representation of the input.

Thereby AE has to reconstruct full input vector by means of given hidden unit activations. If input vector values were completely random, then compression and reconstruction tasks would not be easy. However, if there are some regularities in the input data (e.g. some features of patterns are correlated),

then this network will be able to discover some of these regularities. Such approach seems very similar to principal component analysis (PCA), but it already can partially help in solving inpainting problem.

In this work we use stacked AEs which provide architecture for learning deep representations. In Figure 1 b) one can see that every hidden layer of the network except the last one might be considered as input vector of the next AE. This structure allows the stacked AE to learn more abstract features and to train each AE one after another. Also this feature makes the stacked AE less trivial comparing to PCA. Training the stacked AE is reduced to minimization of simple linear form for each simple AE.
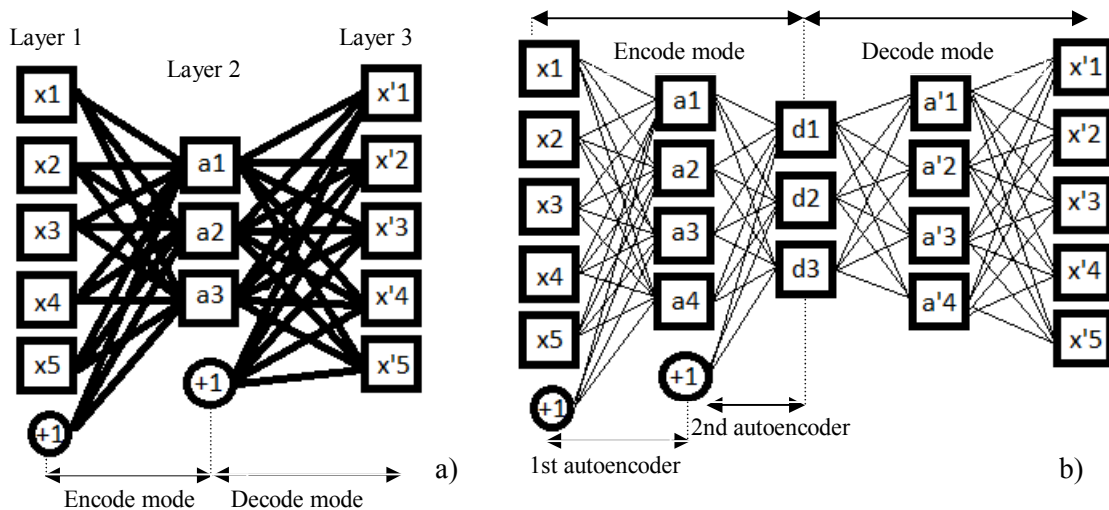

Figure 1. AE (a) and stacked AE (b) structure. Figure (a) shows simple AE structure that works similarly to PCA. Figure (b) shows multi-layer structure of stacked AE which is composed of two simple AEs.

3. Proposed algorithms

In this paper we propose the next two algorithms for image inpainting based on stacked autoencoders. The first one represents the iterative procedure which can be described by several steps:

1. Stacked AE are trained using training set of images
2. Corrupted image (e.g. image with occluded region) is passed to the input of the trained stacked AE
3. Network's response to the input stimulus (reconstructed image) is calculated
4. Corrupted areas of the input image are replaced by reconstructed ones
5. The image with replaced areas becomes the input vector of the stacked AE
6. Steps 3-5 are repeated until stopping criterion (e.g. number of iterations) is met

The second algorithm is based on the assumption that reconstructed region of the image can be described as fixed point (also known as fixpoint or invariant point) of some very complex and multidimensional function. Therefore, that function may have more than one such fixed point. The idea is to find optimal fixpoint using metaheuristic search such as genetic algorithms (GA). Fitness function for each point individual can be calculated as standard deviation (SD) between input and reconstructed by stacked AE images.

4. Experimental results

For experimental evaluation we used handwritten digits database and FEI University face database. Stacked AE was trained on the training set consisting of different digits or different faces. Test images were created by cutting some region of interest from an image. Image inpainting algorithms were tested with different parameters such as number of images in training set, number of hidden layers in AE, sizes of hidden layers, learning rate, number of training epochs and etc. Figure 2 shows some results of inpainting for digits and figure 3 shows inpainting results for faces.
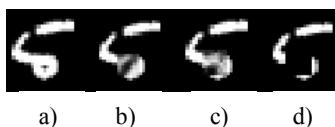
a)   b)   c)   d)

Figure 2. Inpainting results for handwritten digits images. (a) original image. (b) iterative algorithm inpainting result. (c) GA inpainting result. (d) corrupted image.
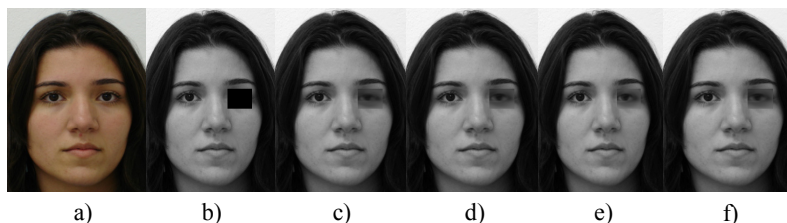

a)   b)   c)   d)   e)   f)

Figure 3. Inpainting results for frontal facial images. (a) original image. (b) corrupted image. (c) iterative algorithm inpainting result with 3 hidden layers in AE. (d) GA inpainting result with 3 hidden layers in AE. (e) iterative algorithm inpainting result with 5 hidden layers in AE. (f) GA inpainting result with 5 hidden layers in AE.

5. Conclusion

We have considered the approach to image inpainting based on stacked autoencoders. This approach was implemented in two image inpainting algorithms and tested on two types of images handwritten digits images and frontal facial images. The results show that inpainting using GA approach has slight advantage compared to iterative inpainting procedure. Also increasing amount of hidden layers in stacked AE allowed gaining different kinds of inpainting results which might be caused by more abstract features learnt by the network on deeper layers.

Acknowledgements

References

[1]   Chan T. F., Shen J., Zhou H.-M. A 2006 Total Variation Wavelet Inpainting Model with Multilevel Fitting Parameters *Proc. SPIE 6313, Advanced Signal Processing Algorithms, Architectures, and Implementations XVI*.

[2]   Arias P., Caselles V., Sapiro G. A 2009 Variational Framework for Non-local Image Inpainting *Proc. EMMCVPR'09*. 345–358.

[3]   Shibata T., Iketani A., Senda Sh. 2011 Fast and Structure-preserving Inpainting Based on Probabilistic Structure Estimation MVA2011 *IAPR Conference on Machine Vision Applications, Nara, JAPAN* 22–25.

[4]   Potapov A., Scherbakov O., Zhdanov I. 2013 Practical algorithmic probability: an image inpainting example *Proc. SPIE 9067, Sixth International Conference on Machine Vision*.

[5]   Potapov A. S. 2012 Using multiple video-information representations in automatic image-analysis systems *Journal of Optical Technology* 79 689–692.

[6]   Potapov A. S., Malyshev I. A., Puysha A. E., Averkin A. N. 2010 New paradigm of learnable computer vision algorithms based on the representational MDL principle *Proc. SPIE* vol 7696, 769606.