

Original Research Paper

Regularization Method for Solving Denoising and Inpainting Task Using Stacked Sparse Denoising Autoencoders

Pavel Vyacheslavovich Skribtsov and Sergey Olegovich Surikov

PAWLIN Technologies Ltd, Dubna, Russia

Article history

Received: 06-10-2015

Revised: 01-01-2016

Accepted: 13-01-2016

Corresponding Author:

Pavel Vyacheslavovich Skribtsov,
PAWLIN Technologies Ltd,
Dubna, Russia
Email: monblan.pro@yandex.ru

Abstract: This article offers a regularization method for training stacked sparse denoising autoencoders aimed at designing model description of objects used for image denoising and inpainting. The offered regularization method allows increasing the generalizing ability of model description, which results in greater stability of denoising methods using it with regard to variation of the noise type. This makes the offered method vital for the tasks where noise or image distortion types cannot be known beforehand. Response speed of the offered algorithm enables to use it for dataflow processing. Absence of the need to formalize the physical nature of noises allows applying the approach to processing images received from various sensors, including sensors beyond the visible spectrum, multispectral and other sensors. The article shows the results of applying the offered regularization method in the denoising and inpainting task as exemplified by FERET face image base.

Keywords: Image Denoising, Image Inpainting, Stacked Sparse Denoising Autoencoder, RPROP, Regularization Method, Image Preprocessing, Sensor Invariant Preprocessing

Introduction

The problem of distorting an initial signal received from the sensors by some types of noises is one of the problems arising during the design of computer vision systems. To solve this problem image denoising and inpainting methods are applied; their task is to restore the original image from its distorted version. The image denoising problem occurs in case if the image is distorted by adding any type of noise to it (for example, white Gaussian noise, which is common for many kinds of sensors), whereas the inpainting problem arises in case if it is necessary to restore separate missing image pixels or remove some complex pattern overlay (for example, a text superimposed on the image).

Related Data

The denoising task may be formulated (Xie *et al.*, 2012) as follows: Let $x \in R^n$ be a noisy image and $y \in R^n$ - an original image matching it. In this case the noise contamination process may be represented in the form: $\zeta: R^n \rightarrow R^n$ Equation 1:

$$x = \zeta(y) \quad (1)$$

In this case the denoising task aim is formulated as Equation 2:

$$f = \arg \min_f E_y \|f(x) - y\|_2^2 \quad (2)$$

This formula shows that the task is to find function f , representing the best approximation of ζ^{-1} .

There are two different approaches to denoising-in one case the distorted image is converted into another representation space (for example, using transformation to wavelet domain, as in (Xu *et al.*, 2009)), where the original image can be more easily separated from the overlaid noise (Xu *et al.*, 2009; Portilla *et al.*, 2003). Another approach lies in analyzing the image statistics directly in the initial representation space. KSVD method exemplifies the implementation of this approach (Elad and Aharon, 2006).

The existing inpainting methods may be divided into two categories-blind inpainting and non-blind inpainting. Non-blind inpainting techniques are used when a priori information about the missing domains of the image to

be inpainted is provided to the algorithm input; blind inpainting methods are applied in case if such information is unavailable and the method must automatically identify the distorted image domain. The existing non-blind inpainting methods demonstrate high performance in removing the superimposed text, simple overlaid images, etc. (Criminisi *et al.*, 2004; Bertalmio *et al.*, 2000). On the other side, blind inpainting is much more complicated problem and until recently the efficient methods to solve it existed only for the case of simple impulse noise (Dong *et al.*, 2011; Wang *et al.*, 2013). Usage of deep neural networks, in particular of Stacked Sparse Denoising Autoencoders (SSDA) became a breakthrough in solving the blind inpainting problem (Xie *et al.*, 2012). SSDA architecture (Xie *et al.*, 2012) demonstrates the advantage as compared to the standard KSVD denoising method (Elad and Aharon, 2006) based on application of visual dictionaries. Xie *et al.* (2012) indicate the connection of such possibilities with the deep and consequently, more non-linear image processing scheme that had proved the advantages in a number of other tasks as compared to the 2D 'flat' architectures. The ability of the SSDA method to train on the subset of visual examples sharing common features and to use this for improvement of denoising characteristics is especially important for solving denoising tasks. Autoencoders are applied in the denoising and inpainting methods by two different ways. In one of them the autoencoder is trained to obtain the descriptive noise model and thus, the denoising method based on its usage appears to be stable to the variation of image class it is applied to. In the other one the autoencoder is learned on the images of a definite class, which makes it resistant to the variation of the type of noise.

The SSDA method is limited by the fact that the result of its operation strongly relies on the quality of training the applied autoencoders, in particular to achieve the required degree of generalization in its process. In addition, the SSDA method is not resistant to those types of noise, samples of which had not been presented to the autoencoder in the course of its training. To overcome these limitations a number of methods were developed, which improve the performance quality of this algorithm. In particular, Shcherbakov and Batishcheva (2014) offered to use two algorithms to improve the blind inpainting performance by means of SSDA-multiple iterative image feeding to the autoencoder input (at first the initial image is supplied to the autoencoder, then they obtained autoencoder output is again supplied to the input and so on), as well as the metaheuristic search in the obtained images to find optimal representation of the missing domain of the initial image. Agostinelli *et al.* (2013) suggested an

algorithm that increases denoising resistance to various types of noise using linear combination of autoencoders, each of which is initially trained for its own type of noise contamination.

Training and use of SSDA is rather computationally intensive task. Therefore, priority lines of improving noise reduction techniques based on the SSDA use are to improve the performance of each of the autoencoders used as opposed to the methods based on the increase in the total number and/or complication of the denoising algorithm structure. One of the possibilities used for this is to apply regularization methods in the learning process. In particular, the authors of the SSDA method show the efficiency of sparsity regularization method based on Kullback-Leibler metric (KL-regularization) (Ng, 2011) in their article (Wang *et al.*, 2013).

This paper offers a new regularization method used for SSDA training when solving denoising and inpainting tasks.

Methods

General Structure of the Algorithm

The denoising and inpainting algorithm investigated in this article consists of two steps:

- The training image set of objects of a certain class is formed and SSDA is trained using it
- The trained SSDA is used for denoising and/or inpainting. For this purpose the noise contaminated image is passed through the trained SSDA

SSDA Structure

The simplest option of the Denoising Autoencoder (DAE) (Ng, 2011) is a three-layered feedforward neural network containing input and output neural layers, dimensions of these layers being equal and also one hidden layer of neurons usually of lesser dimensions (Fig. 1a).

Let $y_i, i=1, \dots, N$ be original images, x_i - noise contaminated images matching them.

The output signal of the hidden layer of the denoising autoencoder is given by the expression Equation 3:

$$z = F_a(W_E * x + B_E) \quad (3)$$

Where:

- X = An autoencoder input signal
- F_a = A diagonal nonlinear operator of activation functions (usually of sigmoid ones)
- W_E = A synaptic weight matrix of the neural network part called encoder of the autoencoder
- B_E = An encoder shift vector

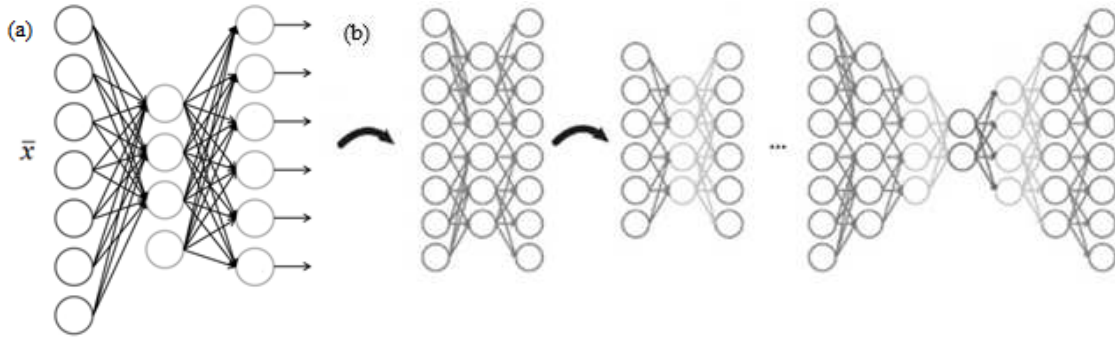


Fig. 1. (a) Structure of the simplest denoising autoencoder (b) Illustration of stacked autoencoder formation mechanism

Then the denoising autoencoder output is given by the following expression Equation 4:

$$\hat{y} = F_a(W_D z + B_D) \quad (4)$$

To solve a denoising or inpainting task the autoencoder is trained using various optimization methods, minimizing the reconstruction error:

$$E = \arg \min_{\theta} \sum_{i=1}^N \|y_i - \hat{y}(x_i)\| \quad (5)$$

Improvement of the denoising autoencoder is a stacked denoising autoencoder-a neural network obtained by combining autoencoders, each of which is trained on the output of the previous one (Fig. 1b).

Applied SSDA Training Algorithm and the Suggested Regularization Method

Let SSDA training is carried out using the following error function:

$$E(w) = E_r(w) + E_{wd}(w) \\ = \frac{1}{2} \sum_{i=1}^N \|\bar{y}(\bar{x}_i, \bar{w}) - \bar{x}_i\|^2 + \lambda \sum_{k=1}^K \|w_k\|^2 \quad (6)$$

where, $\bar{y}(\bar{x}_i, \bar{w})$ -the network output signal vector with \bar{x}_i input signal vector fed to it, w_k -k-th weight coefficient of the neural network.

This function consists of two members, the first of which $E_r(w)$ is a reconstruction error, with the Euclidean distance between the input and output signal vectors used as this error and the second $E_{wd}(w)$ is a regularization term implementing the Weight Decay regularization method (Moody *et al.*, 1995).

The idea of the regularization method suggested in this article is in the following: Weight correction values corresponding to the first (reconstruction error) and the second (Weight decay regularization term) terms of the expression (9) are computed independently, with the RPROP algorithm (Riedmiller and Braun, 1993;

Riedmiller, 1994) being used for the first term and usual gradient descent for the second.

Let us denote Δ_k -weight correction value with number k . Weight correction value update rules (7) and the weights themselves (8), (9) in our case look as follows:

$$\Delta_k^{(t)} = \begin{cases} \eta^+ \Delta_k^{(t-1)} & , \text{if } \frac{\partial E_r^{(t-1)}}{\partial w_k} \bullet \frac{\partial E_r^{(t)}}{\partial w_k} > 0 \\ \eta^- \Delta_k^{(t-1)} & , \text{if } \frac{\partial E_r^{(t-1)}}{\partial w_k} \bullet \frac{\partial E_r^{(t)}}{\partial w_k} < 0 \\ \Delta_k^{(t-1)} & , \text{otherwise} \end{cases} \quad (7)$$

where, $0 < \eta^- < 1 < \eta^+$:

$$\Delta w_k^{(t)} = \begin{cases} -\Delta_k^{(t)} & , \text{if } \frac{\partial E_r^{(t)}}{\partial w_k} > 0 \\ +\Delta_k^{(t)} & , \text{if } \frac{\partial E_r^{(t)}}{\partial w_k} < 0 \\ 0 & , \text{otherwise} \end{cases} \quad (8)$$

$$w_k^{(t+1)} = w_k^{(t)} + \Delta w_k^{(t)} - \frac{\partial E_{wd}}{\partial w_k} = w_k^{(t)} + \Delta w_k^{(t)} - 2\lambda w_k^{(t)} \quad (9)$$

There is an exception to the rule (8): If a partial differential changes the sign, i.e., the previous step was too big and local minimum was missing, the previous weight update with the reverse sign shall be used:

$$\Delta w_k^{(t)} = -\Delta w_k^{(t-1)} , \text{if } \frac{\partial E_r^{(t-1)}}{\partial w_k} \bullet \frac{\partial E_r^{(t)}}{\partial w_k} < 0 \quad (10)$$

The peculiarity of the proposed method of training is as follows. In case the weight coefficient value w_k gets in a sufficiently small neighborhood of a local minimum of error function E_r , the value of $\Delta w_k^{(t)}$ as follows from the expressions (7) and (8) starts decreasing due to the E_r error sign fluctuation. Accordingly, in the expression (9) for this weight coefficient variation value the contribution of regularization term $-2\lambda w_k$ increases, since

its value depends only on the weight value w_k and is not limited by the RPROP step.

This results in additional decrease of the weight coefficient values which get “stuck” in the local minimums of error function E_r and exert no significant impact on the final solution of the optimization problem. As shown by the experimental research, such approach allows achieving greater degree of the representation model sparsity, which is apparent from the weight coefficients of the hidden layer of the autoencoder (Table 1) and in leads to the increased generalization level of the descriptive model obtained as a result of the descriptive model training. This increases the resistance of the descriptive model to various distortions of the input signal and enables to use it efficiently for solving denoising and inpainting tasks.

Results

To assess the performance of the proposed method numerical denoising experiments were carried out using FERET face database (Phillips *et al.*, 2000).

The numerical experiments were designed as follows:

- For the experiments training and test sets of face images were formed from the FERET database, which included 1,085 and 275 faces, respectively
- The test set images were made noise contaminated with the white Gaussian noise having various parameter values

- Autoencoder was learned by the training set using the above method of training
- The denoising results were assessed according to the metrics of peak signal to noise ratio $PSNR = 10 \log_{10}(255^2 / \delta_e^2)$, where δ_e^2 - root mean square error. PSNR is one of the standard metrics used to evaluate denoising results (Xie *et al.*, 2012)

In the course of preliminary experiments an optimal configuration of the autoencoder was selected, it included one hidden layer containing 512 neurons and optimal value of coefficient λ that determines the extent of Weight Decay regularization impact on the error function $\lambda = 0.05$.

Visual examples of the algorithm operation are given in Table 2.

PSNR curves versus Gaussian noise level σ before and after denoising are shown in Fig. 2a.

Inpainting experiments were performed additionally. In these experiments the superimposed text was removed from the image. In this case the pre-trained SSDA of the denoising experiment was used.

Visual examples of inpainting are given in Table 3.

To assess the inpainting performance the PNSR measure was used. Scatter diagram of the averaged PSNR of images after inpainting versus the same value before inpainting is shown in Fig. 2b.

Table 1. Examples of the hidden layer autoencoder traits obtained by using ordinary method of RPROP training using Weight Decay” regularization and by the proposed regularization method

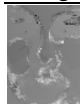




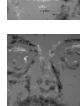





Ordinary method of RPROP Training using “Weight Decay” regularization	The proposed regularization method
	
	
	
	
	
	
	

Table 2. Visual examples of the denoising results for the white Gaussian noise versus the noise parameters













































$\sigma = 25, PSNR = 20.35$		
$\sigma = 50, PSNR = 14.97$		
$\sigma = 75, PSNR = 12.17$		
$\sigma = 100, PSNR = 10.48$		

Table 3. Inpainting example. The upper-most image is an original, it is followed by top-to-bottom intermittent examples of the distorted images and the corresponding images obtained by means of inpainting

Original				
				
				
				
				
				
				
				
				

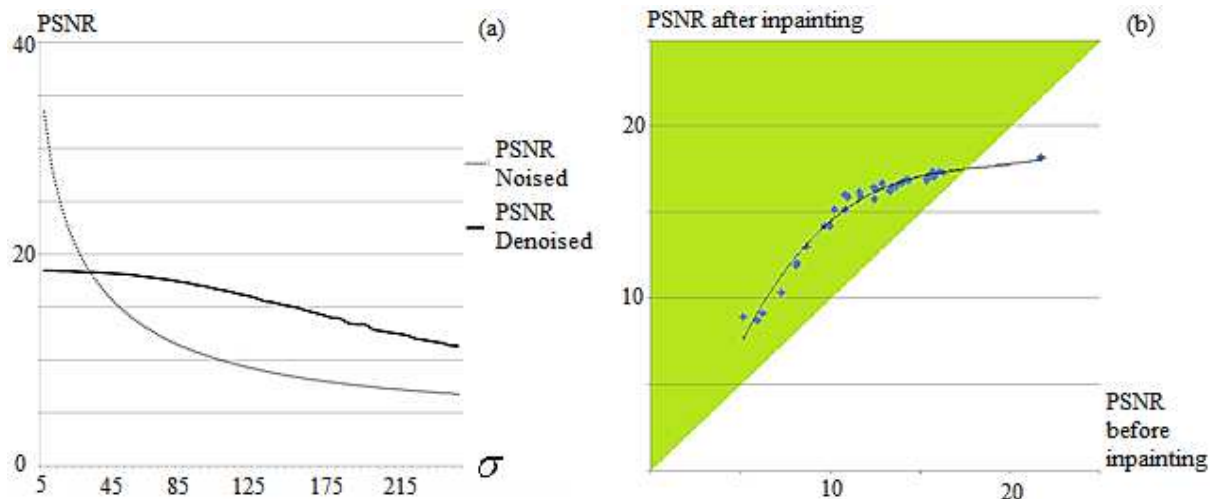


Fig. 2. (a) PSNR curves versus Gaussian noise level σ before and after denoising. (b) Scatter diagram of the averaged PSNR value for images after inpainting versus its value before it. Green zones denotes the area where the PSNR value after inpainting happens to be higher than before it, i.e., the area where inpainting gives positive results

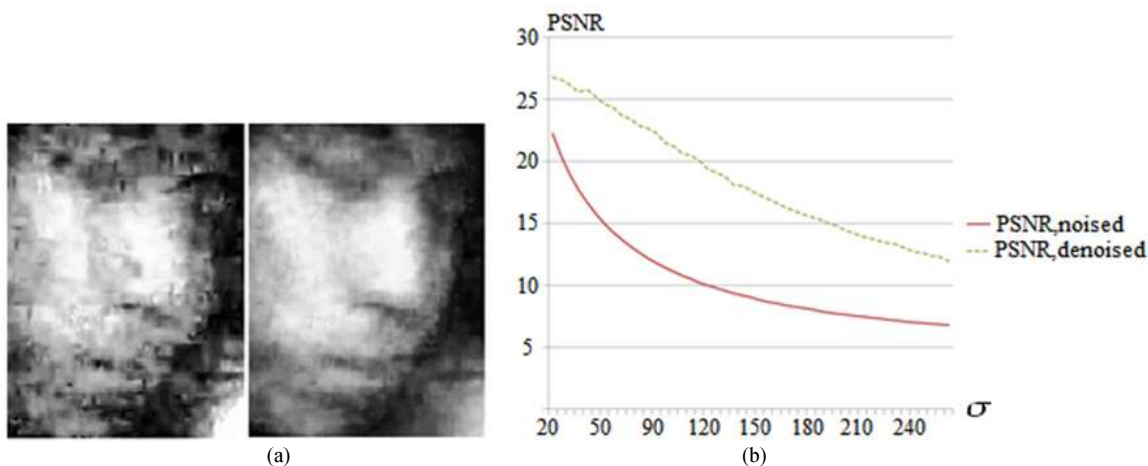


Fig. 3. (a) An example of denoising the image obtained in the nearest IR range using SSDA trained by the proposed regularization method; (b) PSNR curve versus Gaussian noise level σ before and after the experiment with noise contamination of image passed through the autoencoder

In addition, the experiments were carried out to verify the efficiency of the proposed regularization method in the denoising task for the in-stream video obtained in the nearest IR range. The example of the experimental results is shown in Fig. 3a.

Discussion

Quality of the proposed method performance assessed by PSNR metrics appears to be lower as compared to that proposed in (Xie *et al.*, 2012) (for example, for the initial images with $PSNR \approx 8$ denoising method according to Xie *et al.* (2012) gives increase of this value to $PSNR \approx 24$, whereas the method studied herein gives $PSNR \approx 15$). At the same time it should be

noted that the quality of denoising by the proposed method is limited by the reconstruction quality obtained by means of the autoencoder (PSNR metrics of the distortions introduced by the autoencoder itself for the noise-free images (≈ 18.67) is approximately equal to the metrics corresponding to the denoising quality achieved in the experiments (Fig. 2a)). The main distortion introduced by the autoencoder into the output images is a removal of high-frequency component, which is not significant during elaboration of recognition algorithms (for example, face recognition algorithms). To evaluate distortions relating to the autoencoder training quality, two experiments were carried out. In the first of them noise was superimposed on the images preliminarily

passed through the pre-trained autoencoder. The denoising quality considerably increased in this experiment (Fig. 4). Thus, performance of the proposed method may be increased by improving image reconstruction quality obtained with the autoencoder. To assess the potential of such improvement another experiment was performed aimed at finding the dependence of PSNR of autoencoder distortions on the volume of the set used to train it. In this case PSNR was assessed by the test set containing 282 images. The experimental result is shown in Fig. 4. As it is seen, autoencoder distortions may be considerably decreased by using the training set of greater volume.

In addition, preliminary experiments demonstrated that in order to decrease the impact of distortions introduced by the autoencoder it is possible to apply approach with training certain models for different zones of the object-this approach will be described in more detail in the further works.

PSNR curves versus Gaussian noise level σ before and after denoising (Fig. 2a) enables to conclude about the scope of application of the studied denoising method. It is efficient (i.e., results in PSNR value increase) in case of fairly severe noise contamination of images. This is also connected with the fact that signal reconstruction by means of SSDA introduces distortion in the signal by itself.

According to the curve (Fig. 2b) a similar conclusion may be drawn with regard to the application of the investigated method to the inpainting task, as well.

Furthermore, visual assessment of the inpainting quality shows that besides drawbacks relating to the distortion by the high-frequency component of images, the proposed method is advantageous as compared to (Xie *et al.*, 2012). Particularly, in the inpainting example of Fig. 5 the proposed method completely removes text from the images, whereas the method described in (Xie *et al.*, 2012) leaves distortions.

Denoising and inpainting approach proposed herein has an important advantage as compared to the method described in (Xie *et al.*, 2012)- it does not require preliminary recognition of the type of noise. This is connected with the fact that in the course of the autoencoder training a descriptive model of objects of a certain class is created, which is resistant to the input signal distortions rather than the descriptive noise model, as it is done in (Xie *et al.*, 2012). Thus, the studied approach is relevant for the tasks where a definite noise or image distortion model cannot be known beforehand (for example, removal of watermarks when recognizing scanned photos from the documents, elimination of abrasion marks and other film defects when scanning old photos, etc.). Furthermore, since the investigated method is not apparently locked to the definite noise models of specific sensors, it can be applied to a wide class of applications, including for denoising and inpainting of images obtained by means of sensors operating beyond the visible range, multispectral images, etc.

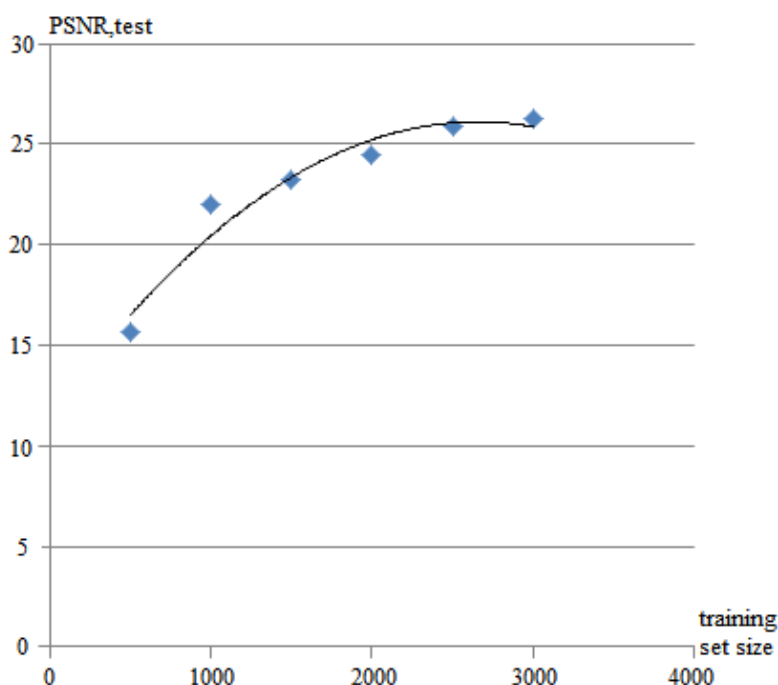


Fig. 4. PSNR of distortions introduced by the autoencoder on its training set volume

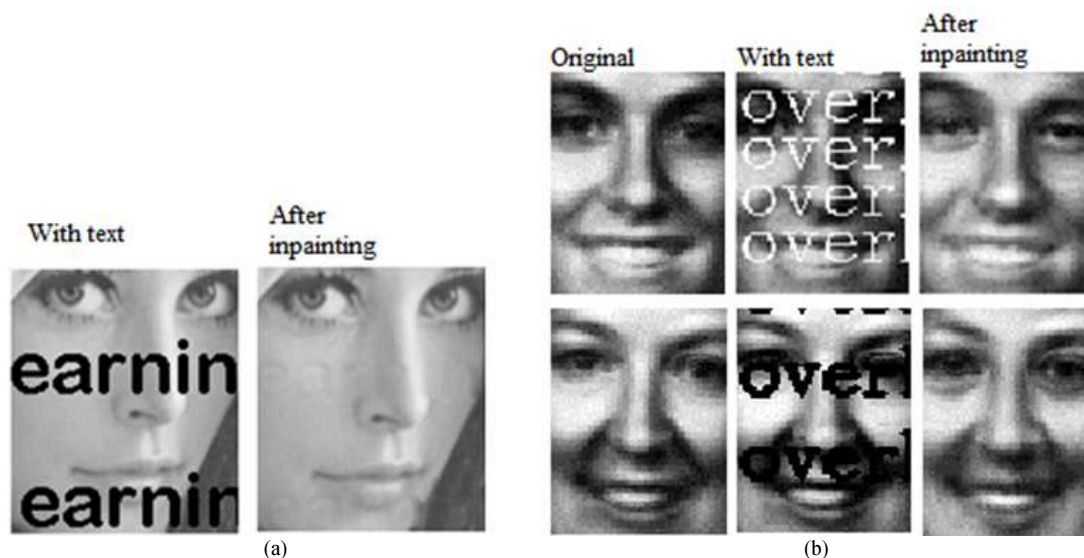


Fig. 5. Visual comparison of the inpainting quality obtained in (Xie *et al.*, 2012) (a) and by the proposed approach (b)

Conclusion

A new regularization method is proposed to be applied for training stacked sparse denoising autoencoders aimed at designing object description model used for image denoising and inpainting. The proposed regularization method allows increasing generalizing ability of the description model, which results in greater invariance of the denoising methods using it with regard to the noise type variation. The proposed method is relevant for the tasks where a noise model or types of possible image distortions cannot be known in advance.

We see a further improvement of our method in training separate models for different distinguishable zones of an object in order to decrease the impact of distortions introduced by the autoencoder.

Acknowledgement

The authors gratefully acknowledge the support provided for this research by the Ministry of Education and Science of the Russian Federation.

Funding Information

The research was financed by the Ministry of Education and Science of the Russian Federation under the Grant Contract dated 24 November, 2014, (grant number 14.579.21.0070, contract unique identifier RFMEFI57914X0070), the grant was assigned for the research into “2D- and 3D-data fusion and processing software for object detection and recognition”.

Author's Contributions

Pavel Skribtsov: Designed the research plan, developed regularization procedure and autoencoder's training algorithm, organized collection of training and validation datasets, discussed the results.

Sergey Surikov: Organized the study, data collection, analysis and writing of the manuscript, participated in discussion.

Ethics

The authors have no conflicts of interest in the development and publication of current research.

References

- Agostinelli, F., M.R. Anderson and H. Lee, 2013. Adaptive multi-column deep neural networks with application to robust image denoising. *Adv. Neural Inform. Process. Syst.*, 26: 1493-1501.
- Bertalmio, M., G. Sapiro, V. Caselles and C. Ballester, 2000. Image inpainting. *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, Jul. 23-28, ACM Press, New Orleans, pp: 417-424. DOI: 10.1145/344779.344972
- Criminisi, A., P. Pérez and K. Toyama, 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.*, 13: 1200-1212. DOI: 10.1109/TIP.2004.833105
- Dong, B., H. Ji, J. Li, Z. Shen and Y. Xu, 2011. Wavelet frame based blind image inpainting. *Applied Comput. Harmonic Anal.*, 32: 268-279. DOI: 10.1016/j.acha.2011.06.001

- Elad, M. and M. Aharon, 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15: 3736-3745. DOI: 10.1109/TIP.2006.881969
- Moody, J., S. Hanson, A. Krogh and J.A. Hertz, 1995. A simple weight decay can improve generalization. *Adv. Neural Inform. Process. Syst.*, 4: 950-957.
- Ng, A.Y., 2011. Sparse autoencoder. CS294A Lecture notes.
- Phillips, P.J., H. Moon, S.A. Rizvi and P.J. Rauss, 2000. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Patt. Anal. Mach. Intell.*, 22: 1090-1104. DOI: 10.1109/34.879790
- Portilla, J., V. Strela, M.J. Wainwright and E.P. Simoncelli, 2003. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Process.*, 12: 1338-1351. DOI: 10.1109/TIP.2003.818640
- Riedmiller, M., 1994. Rprop I. Rprop-description and implementation details. Citeseer.
- Riedmiller M. and H. Braun, 1993. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proceedings of the IEEE International Conference on Neural Networks*, Mar. 28-Apr. 01, IEEE Xplore Press, San Francisco, pp: 586-591. DOI: 10.1109/ICNN.1993.298623
- Shcherbakov, O. and V. Batishcheva, 2014. Image inpainting based on stacked autoencoders. *J. Phys.: Conf. Ser.*, 536: 012-020. DOI: 10.1088/1742-6596/536/1/012020
- Wang, Y., A. Szelam and G. Lerman, 2013. Robust locally linear analysis with applications to image denoising and blind inpainting. *SIAM J. Imag. Sci.*, 6: 526-562. DOI: 10.1137/110843642
- Xie, J., L. Xu and E. Chen. 2012. Image denoising and inpainting with deep neural networks. *Adv. Neural Inform. Process. Syst.*, 25: 341-349.
- Xu, J., K. Zhang, M. Xu and Z. Zhou. 2009. An adaptive threshold method for image denoising based on wavelet domain. *Proceedings of SPIE, the International Society for Optical Engineering, (SOE' 06), SPIE*, pp: 74954M-74954M.