DOI:10.34918/85105

Seweryn KALISZ[1,*], Michał MARCZYK[1,2]

Chapter 5. METHODS FOR REDUCTION OF BLUR EFFECT IN THE DENOISING AUTOENCODER MODEL FOR RIB SUPPRESSION IN CXR IMAGES

5.1. Introduction

The development of bioinformatics tools related to image analysis is of high interest to researchers, medical doctors, and engineers since selected technologies could improve clinical performance and efficiency and influence the development of personalized medicine [1]. One of the most widely used medical imaging methods is X-ray imaging, which is characterized by low cost, short processing time, low radiation dose, and high availability [1, 2]. The chest X-ray (CXR) is frequently the first imaging study obtained during a broad range of conditions and remains central to the screening, diagnosis, and management of the disease [2, 3]. Chest radiography provides images of soft tissues such as the heart, airways, blood vessels, and lungs, and additionally the bones of the spine and chest. Overlapping of structures present in the image can complicate the visual interpretation of the detection of abnormalities in radiographs by medical doctors leading to false negative results [4]. A major reason is the presence of bone structures on the radiograph during soft-tissue diagnosis. There exists a technique that enables the separation of soft tissue from bone called dual--energy subtraction (DES) radiography. Some studies showed that suppressing rib shadows on soft-tissue images by using the DES technique has improved the speed and accuracy of radiologists in diagnosing pulmonary nodules [5]. However, compared to traditional imaging, DES requires a higher radiation dose and could result in noisy images with visual abnormalities due to the patient moving [6]. These

[1] Department of Data Science and Engineering, Faculty of Automatic Control, Silesian University of Technology, Gliwice, Poland.
[2] Yale Cancer Center, Yale School of Medicine, 06511 New Haven, CT, USA.
* Corresponding author: Seweryn.Kalisz@polsl.pl.

problems, together with lower availability compared to traditional CXR, have interested researchers in the field of image processing to search for solutions allowing the reduction of the presence of bone structures on X-ray images resulting in soft--tissue reconstruction.

Deep learning has become the technique of choice for image analysis tasks in recent years and has had a huge impact on the field of medical imaging [3, 7]. A common method for suppressing ribs from CXR images is based on autoencoder architecture. In literature, three [8] or four [9] layer convolutional denoising autoencoder architectures can be found. Models based on autoencoder architecture are often characterized by a blurring effect on decoded images, which is mostly caused by the occurrence of distance metrics in the loss function [25]. Additionally, several modifications of convolutional neural networks were tested [10, 11], also in the wavelet domain [12]. There are also approaches based on conditional generative adversarial networks minimizing the pairwise image difference and adding Haar 2D wavelet decomposition to improve model convergence [13].

This work aims to review and evaluate methods to improve image quality after the suppression of bone structures by reducing the blurring effect on the example of convolutional denoising autoencoder, composed of 4 layers, and introduced earlier [15]. First, we checked the modification of the image resizing scheme through layers of the autoencoder. Next, we introduce bridge connections between the coder and encoder layers. Finally, we modified the α parameter of the loss function, which was responsible for establishing the weights between the mean square error index and the multi-scale structural similarity index. Also, we have tested the combination of these methods.

5.2. Materials and methods

5.2.1. Data

The Bone Suppression dataset contains 35 pairs of standard CXR images and corresponding soft-tissue images created using DES technology. These data were gathered from a variety of online sources and made available by Innopolis University researchers [8]. There are 11 female pairs and 24 male pairs, each with a different

resolution (the largest being 660 x 775 pixels and the smallest being 424 x 465 pixels). JPEG images had a quality compression ratio ranging from 85 to 100 (high-quality images). Several image pairs have arrows or captions superimposed on the CXR image as artifacts. It can be noted that the images are characterized by varying brightness.

5.2.2. Image preprocessing

Considering the different sizes of the images in the dataset all images were cropped to a square to maintain the correct proportion of the structures. The images were then resized to 512 x 512 pixels resolution. Next, the contrast of each image has been enhanced by using Contrast Limited Adaptive Histogram Equalization with a clip limit equal to 2 and grid size equal to 8 x 8 pixels, which specifies the area within which the contrast is increased so that the input and output histograms for the region are almost identical. This operation allows an increase of details in the biological structures shown in medical images [15].

Data augmentation operations were applied to increase the robustness of the model and to increase the size of training data. The resulting images allow reflection of patients' behavior (affine transformations) and image property differences between X-ray scanners (color transformations). In this work, the following operations were used [16]: (i) translation (± 10 pixels); (ii) rotation (± 5 degrees); (iii) scaling (± 10%); brightness change (± 20%); contrast change (± 20%). The parameters of each operation were chosen based on the recent literature [17, 18].

5.2.3. Model architecture

Convolutional denoising autoencoder is the unsupervised learning algorithm that learns to map corrupted data to uncorrupted data by minimizing the loss function between pairs of images [19]. The ribs in the images are treated as noise (corrupted data) which we want to obtain for the soft-tissue image (uncorrupted data). The architecture used is symmetric and consists of 4 layers with several filters (32, 32, 48, 48) and a window size equal to 4 x 4 pixels on the encoder and decoder part. The input image should be grayscale, characterized by dimensions of 512 x 512 pixels, and be normalized to a floating point value in the range <0;1>. Rectified Linear Unit

(ReLU) in each layer was used as an activation function. The padding property of the convolutional block has been set to 'same' for all layers.

The loss function minimizes Mean Square Error (MSE) and maximizes Multi-Scale Structural Similarity Index Measure (MS-SSIM) [20]. MS-SSIM can be defined by the equation:

$$MS-SSIM(I,G) = [l_M(I,G)]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(I,G)]^{\beta_j} [s_j(I,G)]^{\gamma_j} \quad (1)$$

where *l, c,* and *s* are the luminance, contrast and structure terms at scale *M* and *j*. The exponents *α,* β, and γ are used to adjust the relative importance of different components. Two discrete non-negative signals are represented by *I* and *G*. *I* is a model output and *G* is a ground truth image (soft-tissue image).

The mathematical description of the loss function is given by the following equations:

$$L = \alpha \cdot L^{MS-SSIM} + (1-\alpha) \cdot L^{MSE} \quad (2)$$

$$L^{MS-SSIM}(I) = \frac{1}{N} \sum_{i \in I} 1 - MS-SSIM(I(i), G(i)) \quad (3)$$

$$L^{MSE}(I) = \frac{1}{N} \sum_{i \in I} (I(i) - G(i))^2 \quad (4)$$

where *N* is several pixels in *I*, *i* is the index of pixels in *I* and *α* is a parameter. The value of parameter *α* was set to 0.84 in the base model, which was empirically determined as the best value for image analysis in [21]. The training was run for 300 epochs in all cases. The initial learning rate was 0.001 for the first 100 epochs. Then it decreased twice every 50 epochs. The model was implemented in Python 3.9.12 using Tensorflow 2.5.0 with CUDA 11.2 and cuDNN 8.2.1 libraries. Figure 1 presents the architecture of the base model. Parameters ABCD in convolutional layers specifying the strides of the convolution along the height and width of the input image.

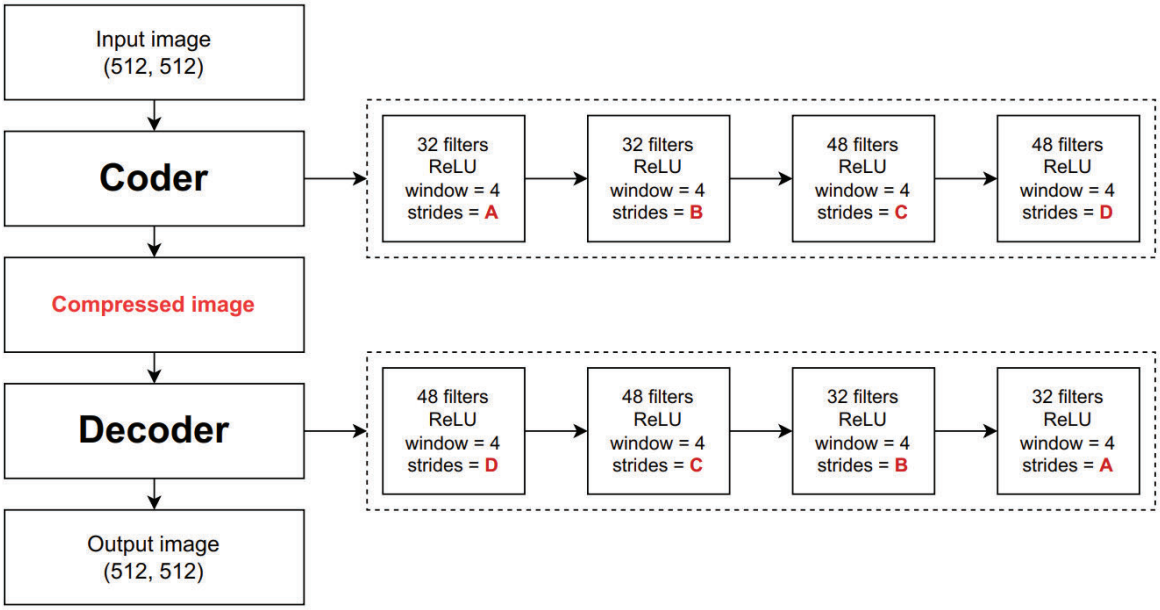

Fig. 1. The base architecture of convolutional denoising autoencoder for bone suppression in CXR images

Rys. 1. Podstawowa architektura konwolucyjnego autoenkodera odszumiającego do tłumienia kości w obrazach rentgenowskich klatki piersiowej

5.2.4. Image resizing

The stride parameter defines the step size for sampling when scanning the input layer to run convolutional operations [22]. In this study, we tested the effect of changing the strides parameter in individual layers, which also changed the compressed images. The compressed image at the input to the decoder is 64 x 64 pixels or 128 x 128 pixels, where it is then increased to its original size by increasing the size of the stride. Table 1 shows the verified configurations. The name Base refers to the configuration used in previously published work.

Table 1

Experimental stride parameters

Configuration Name	A	B	C	D	Compressed Image
Base	1	2	2	2	64 x 64
Strides 1122	1	1	2	2	128 x 128
Strides 1212	1	2	1	2	128 x 128
Strides 1221	1	2	2	1	128 x 128
Strides 2121	2	1	2	1	128 x 128
Strides 2112	2	1	1	2	128 x 128
Strides 2211	2	2	1	1	128 x 128

5.2.5. Bridge Connections

Each convolutional layer learns image features which are called feature maps, which are passed forward using bridge connections (shortcut connections, skip connections) to layers that are not directly adjacent. Compared to the ResNet architecture, the connections are symmetrical between the encoder and decoder. The element-wise addition of the feature maps from a shortcut connection and the connected deconvolutional layer follows [23]. The introduction of symmetric skip connections between the encoder and decoder exhibits two main advantages: (i) improving the results, in deep networks, composed of many layers, where many details are lost during image resizing, (ii) solving the optimization challenge posed by gradient vanishing, resulting in increased performance as the network grows deeper [24]. In this work, the effect of bridging connections between the extreme layers of the autoencoder was examined using different configurations shown in Fig. 2.

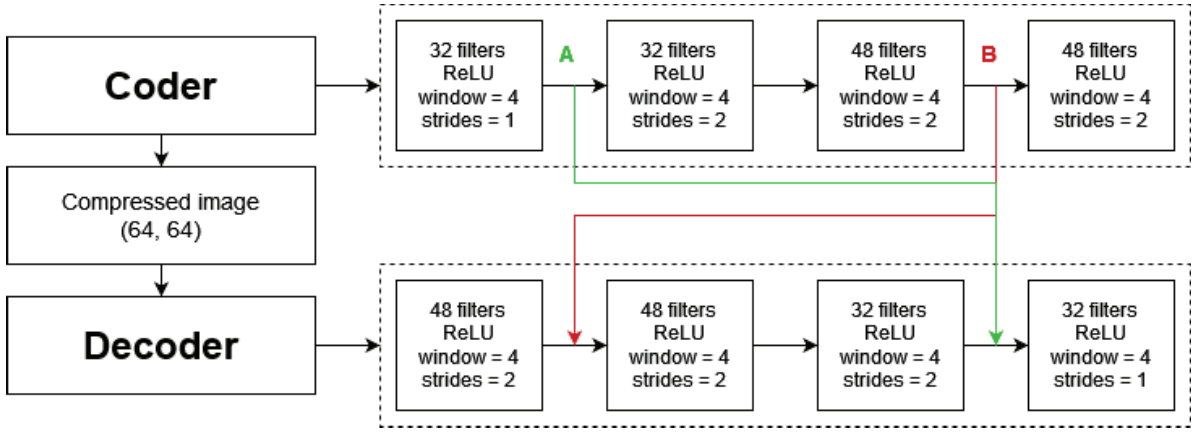

Fig. 2. The base architecture of convolutional denoising autonecoder extended with bridge connections
Rys. 2. Podstawowa architektura konwolucyjnego autonekodera denoisingowego z uwzględnieniem połączeń mostkowych

5.3. Experiments and results

The Bone Suppression dataset was divided into a training set containing 25 pairs of images, a validation set, and a test set containing 5 pairs each. New images were generated for the training and validation set by applying an augmentation technique with the same operations for each image in the pair. After this operation, the training set consists of 1025 images, while the validation set consists of 205 images which is 20% of the training set size. The batch size was set to 16. Adam was chosen as the optimizer in the model training. The models were trained for 300 epochs. Each model was trained using the same initial parameters previously generated at random. The

model with the best validation loss was selected for each run. Calculations were performed on a desktop computer with an Intel Core I5-10500 CPU, 64 GB Ram, and RTX 3060 12GB graphics card. A single training took approximately 2 hours.

The value of the loss function, Peak signal-to-noise ratio (PSNR), and Structural Similarity Index Measure (SSIM) were used as quality indicators to evaluate the model outcome in comparison to the ground truth image (soft-tissue image measured using the DES method) on 5 pairs of an independent test set data. The values obtained are presented in % and refer to the base architecture introduced earlier [14].

5.3.1. Image resizing

The basic architecture resizes the input image three times by half of the current size, and the compressed image is 64 x 64 pixels. The effect of changing the compressed image size to 128 x 128 pixels in various configurations was tested. The greatest improvement for loss function and image quality metrics was obtained for the following configuration: Strides 2121. Detailed results for all configurations are shown in Fig. 3.

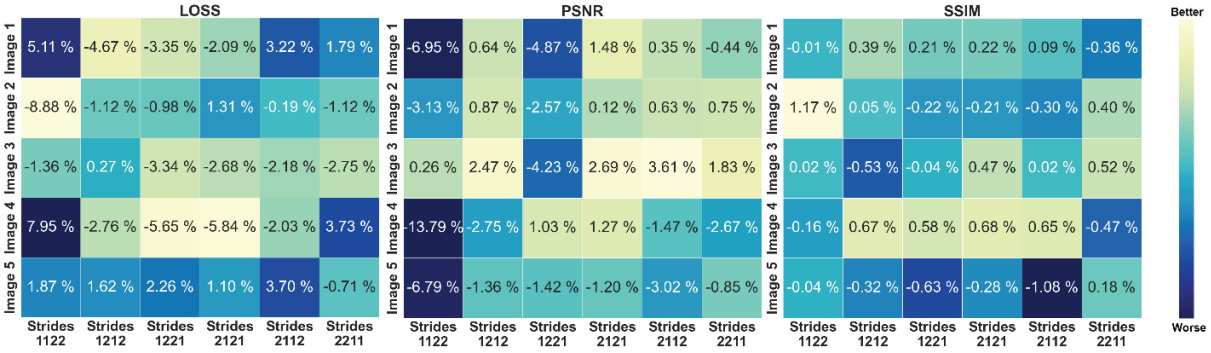

Fig. 3. Heatmaps with values of the loss function and two measures of image quality, calculated between model output and soft-tissue image: Peak signal-to-noise ratio (middle) and Structural Similarity Index Measure (right) for different strides configuration

Rys. 3. Mapy ciepła przedstawiające wartości funkcji straty oraz dwóch wskaźników jakości obrazu: PSNR oraz SSIM dla różnych konfiguracji parametru „strides”

5.3.2. Loss Function

Parameter α in the loss function is responsible for establishing the weights between the mean square error index and the multi-scale structural similarity index. The parameter value was changed from 0 to 1 with a step of 0.1 using the Base configuration shown in Table 1. The best results were obtained for a parameter α equal to 0.5. The value of the loss function for test images presented in the heatmap was calculated for a baseline parameter α equal to 0.84 for each of the trained models. Figure 4 shows the results obtained.

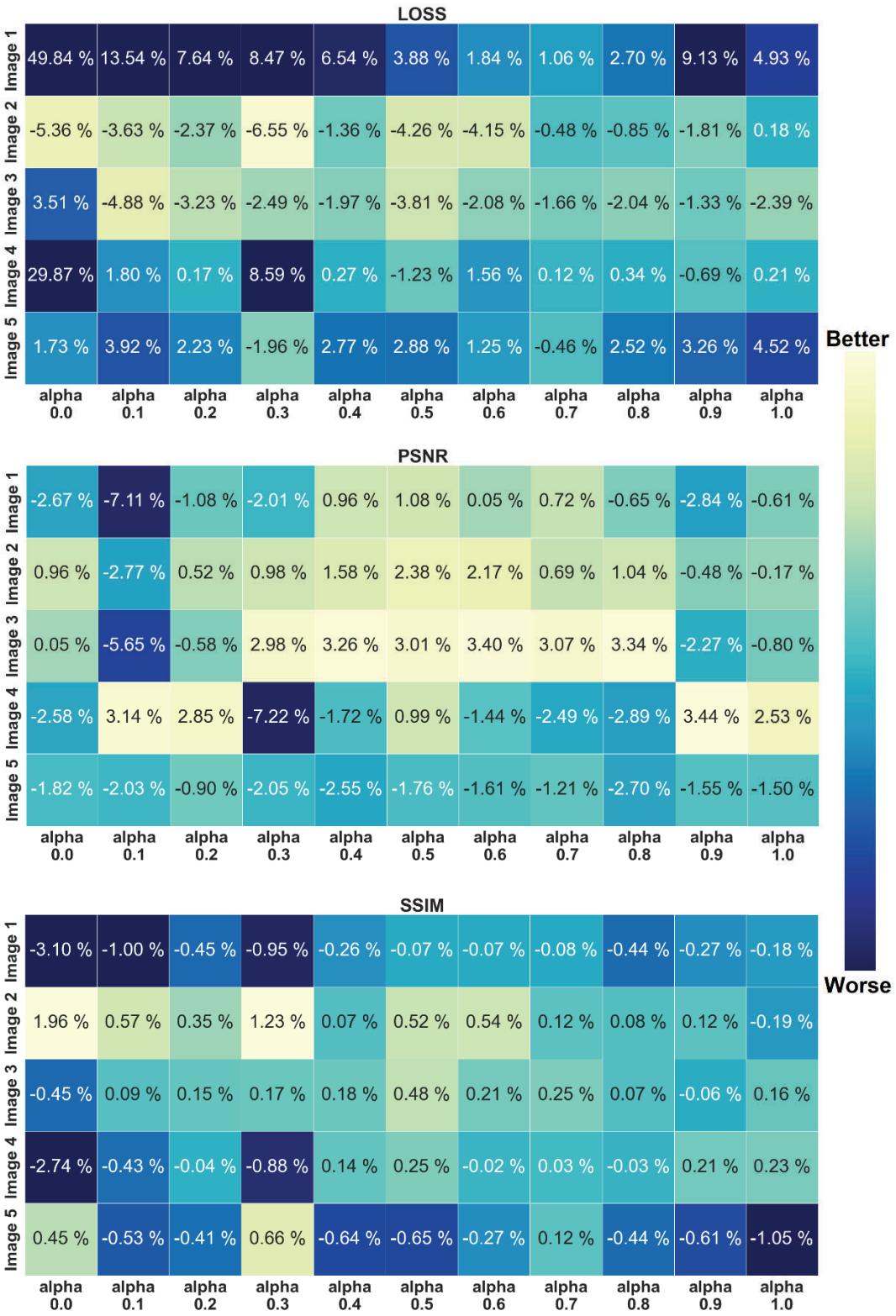

Fig. 4. Heatmaps with values of the loss function and two measures of image quality, calculated between model output and soft-tissue image: Peak signal-to-noise ratio (middle) and Structural Similarity Index Measure (bottom) for various α values in range <0;1>

Rys. 4. Mapy ciepła przedstawiające wartości funkcji straty oraz dwóch wskaźników jakości obrazu: PSNR oraz SSIM dla zmian parametru α w przedziale <0;1>

5.3.3. Bridge Connections

Bridge connections between the symmetrical encoder and decoder layers were introduced into the basic architecture. In this work three configurations were tested: (i) the connection between the outermost layers (connection A in Fig. 2), (ii) the connection between innermost layers (connection B in Fig. 2), (iii) the connection between the outermost layers and the innermost layers (A+B). The best image quality results were obtained for the variant characterized by 2 skipped connections. The detailed results are shown in Fig. 5.

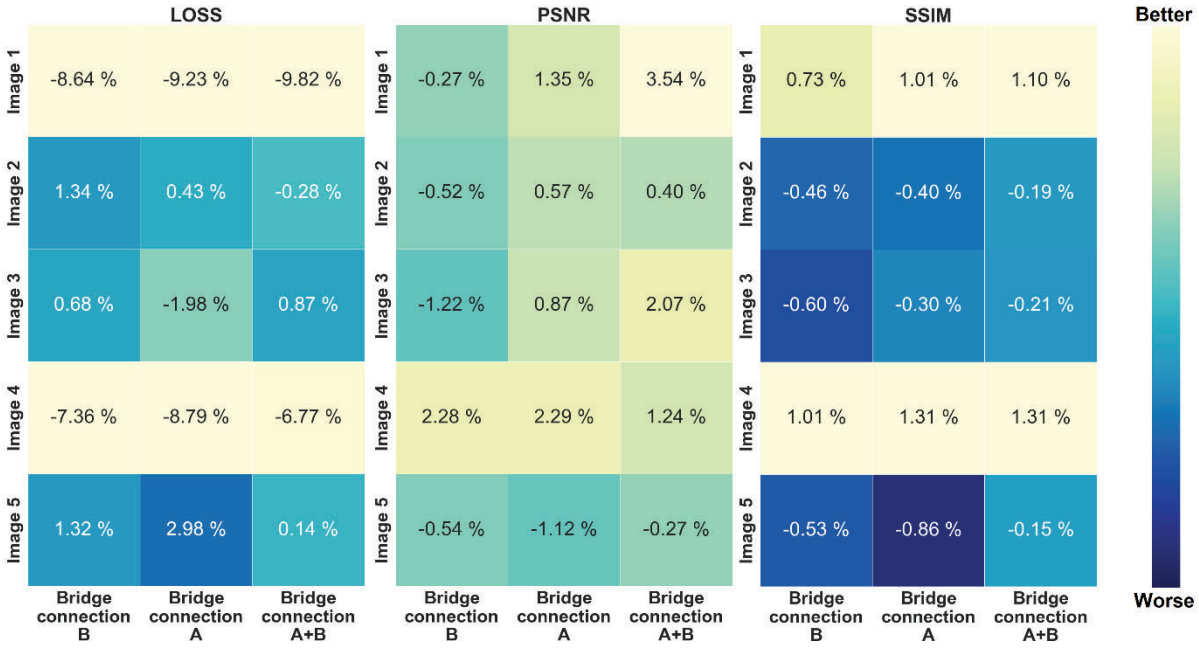

Fig. 5. Heatmaps with values of the loss function and two measures of image quality, calculated between model output and soft-tissue image: Peak signal-to-noise ratio (middle) and Structural Similarity Index Measure (right) for different bridge connection configurations

Rys. 5. Mapy ciepła przedstawiające wartości funkcji straty oraz dwóch wskaźników jakości obrazu: PSNR oraz SSIM dla różnych konfiguracji połączeń mostkowych

5.3.4. Combination of Multiple Methods

The configurations with the best results were selected from the previously described methods. An architecture was created with $\alpha = 0.5$, a strides 2121 configuration, and the use of the bridge connections shown in Fig. 2 (A+B). The applied operations allowed for a significant improvement of the used indices. The values of the obtained results are shown in Fig. 6, while the comparison of the sample images obtained is shown in Fig. 7.

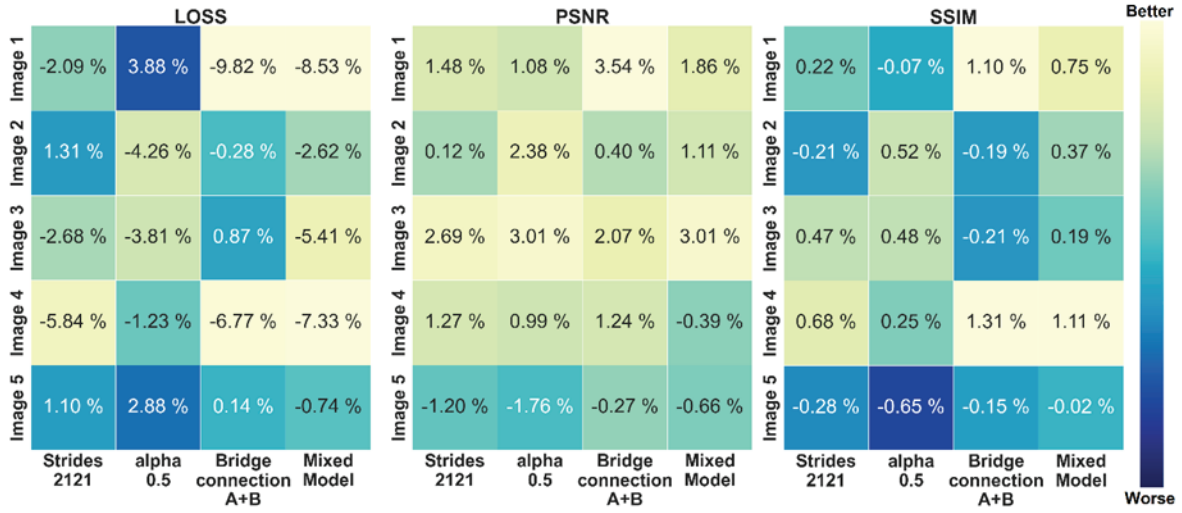

Fig. 6. Heatmaps with values of the loss function and two measures of image quality, calculated between model output and soft-tissue image: Peak signal-to-noise ratio (middle) and Structural Similarity Index Measure (right) for models characterized by the best parameters described in 1.3.1–1.3.4

Rys. 6. Mapy ciepła przedstawiające wartości funkcji straty oraz dwóch wskaźników jakości obrazu: PSNR oraz SSIM dla modeli charakteryzujących się najlepszymi parametrami opisanymi w 1.3.1–1.3.4

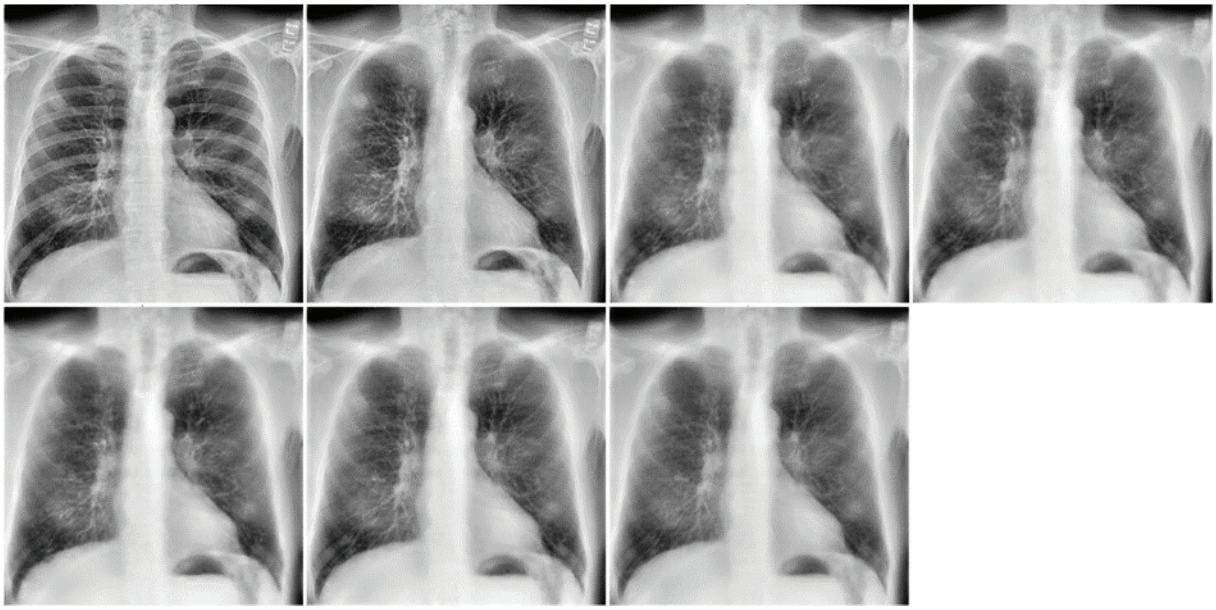

Fig. 7. Comparison of each model based on the output image. Top row: The preprocessed CXR image no. 4 from a testing set (left), soft-tissue ground truth image (middle left), the output of base architecture (middle-right), and the output of the model with strides 2121 configuration (right). Bottom row: output of architecture with α equals 0.5 (left), the output of architecture with two skipped connections (middle), the output of architecture with all presented modifications (right)

Rys. 7. Porównanie obrazów wyjściowych dla każdego z opisanych modeli. Górny rząd: Obraz wejściowy nr. 4 (pierwszy z lewej), obraz tkanek miękkich DES (drugi z lewej), obraz architektury bazowej (drugi z prawej), obraz architektury o konfiguracji strides 2121 (pierwszy z prawej). Dolny rząd: obraz architektury o parametrze α równym 0.5 (z lewej), obraz architektury złożonej z dwóch połączeń mostkowych (środkowy), obraz architektury złożonej ze przedstawionych modyfikacji (z prawej)

5.4. Conclusions

The proposed convolutional denoising autoencoder architecture was updated to eliminate the blur effect, that occurs frequently. A change in the configuration of the 'strides' parameter increases the image size on the input of the decoder, which decreases the overall image compression rate, keeping more important details of the image that are necessary for proper image reconstruction. Adding the bridge connections improves the quality of output images by transferring feature maps between symmetrical encoder and decoder layers. Changing the alpha parameter is needed to tune the loss function for the task-specific images; CXR images in our case. We have tested these methods separately, however, the combination of them brought the highest increase in the quality of the rib-suppressed image in comparison to the base model. An additional advantage of the described methods is the non-impact number of network parameters. The final designed architecture consists of only 4 convolutional layers on the encoder and decoder side. It is a rather simple structure with a small number of parameters, which distinguishes it from existing deep-learning models [9].

The rib-suppressed images presented in Fig. 7 confirm the improved quality of the output image, for each of the methods presented and the final model. In our opinion, these results are sufficient enough to be used as a medical diagnostic support system in a medical facility. An example of the use of the final images obtained with suppressed ribs may be a classification of different lung diseases or tracking disease progression over time.

Future work on the proposed algorithm will focus on separating the region of interest, which in the case of CXRs is the lung region. An architecture consisting of two independent neural networks is planned to be applied. The first network should be responsible for the segmentation of the lung region and could be based on a U-net architecture [26]. The second model will be based on an architecture presented here but trained on segmented lung images.

Acknowledgments

This work was financed by the Silesian University of Technology grant no. 02/070/BK22/0033 for maintaining and developing research potential (SK, MM).

Bibliography

1. C.M. Jones et al.: Chest radiographs and machine learning – Past, present and future. Journal of Medical Imaging and Radiation Oncology 65, 538-544 (2021).
2. S. Raoof et al.: Interpretation of Plain Chest Roentgenogram. Chest 141, 545–558 (2012).
3. E. Sogancioglu, B. Ginneken, K. van Leeuwen, K. Murphy: Deep Learning for Chest X-ray Analysis: A Survey. Medical Image Analysis 72, 102125 (2021).
4. P.K. Shah et al.: Missed non-small cell lung cancer: radiographic findings of potentially resectable lesions evident only in retrospect. Radiology 226, 235–241 (2003).
5. F. Manji, J. Wang, G. Norman, Z. Wang, D. Koff: Comparison of dual energy subtraction chest radiography and traditional chest X-rays in the detection of pulmonary nodules. Quantitative Imaging in Medicine and Surgery 6, (2016).
6. I. Sirazitdinov et al.: Evaluation of Deep Learning Methods for Bone Suppression from Dual Energy Chest Radiography. In Artificial Neural Networks and Machine Learning – ICANN 2020, I. Farkaš, P. Masulli, S. Wermter (eds.), (Springer International Publishing, Cham, 2020), pp. 247–257.
7. G. Litjens et al.: A survey on deep learning in medical image analysis. Medical Image Analysis 42, 60-88 (2017).
8. M. Gusarev, R. Kuleev, A. Khan, A.R. Rivera, A.M. Khattak: Deep learning models for bone suppression in chest radiographs. In 2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). (2017), pp. 1–7.
9. S. Rajaraman, G. Cohen, L. Spear, L. Folio, S. Antani, DeBoNet: A deep bone suppression model ensemble to improve disease detection in chest radiographs. PLOS ONE 17, e0265691 (2022).
10. W. Yang et al.: Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. Med Image Anal 35, 421–433 (2017).
11. N. Matsubara, A. Teramoto, K. Saito, H. Fujita: Bone suppression for chest X-ray image using a convolutional neural filter. Physical and Engineering Sciences in Medicine 43, 97–108 (2020).
12. Y. Chen et al.: Bone Suppression of Chest Radiographs With Cascaded Convolutional Networks in Wavelet Domain. IEEE Access 7, 8346-8357 (2019).
13. J. Liang, Y.-X. Tang, Y.-B. Tang, J. Xiao, R.M. Summers: Bone suppression on chest radiographs with adversarial learning. In Medical Imaging 2020: Computer-

Aided Diagnosis. (International Society for Optics and Photonics, 2020), vol. 11314, pp. 1131409.
14. S. Kalisz, M. Marczyk: Autoencoder-based bone removal algorithm from x-ray images of the lung. In 21st International Conference on Bioinformatics and Bioengineering (BIBE). (2021), pp. 1–6.
15. N. Salem, H. Malik, A. Shams: Medical image enhancement based on histogram algorithms. Procedia Computer Science 163, 300–311 (2019).
16. A. Buslaev et al.: Albumentations: Fast and Flexible Image Augmentations. Information 11, 125 (2020).
17. I. Sirazitdinov, M. Kholiavchenko, R. Kuleev, B. Ibragimov: Data Augmentation for Chest Pathologies Classification. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). (2019), pp. 1216–1219.
18. M. Elgendi et al.: The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective. Frontiers in Medicine 8, (2021).
19. L. Gondara: Medical Image Denoising Using Convolutional Denoising Autoencoders. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW). (2016), pp. 241–246.
20. Z. Wang, E.P. Simoncelli, A.C. Bovik: Multiscale structural similarity for image quality assessment. In The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. (IEEE, 2003), vol. 2, pp. 1398–1402.
21. H. Zhao, O. Gallo, I. Frosio, J. Kautz: Loss functions for neural networks for image processing. arXiv preprint arXiv:1511.08861, (2015).
22. L. Zaniolo, O. Marques: On the use of variable stride in convolutional neural networks. Multimedia Tools and Applications 79, (2020).
23. L.F. Dong, Y.Z. Gan, X.L. Mao, Y.B. Yang, C. Shen: Learning Deep Representations Using Convolutional Auto-Encoders with Symmetric Skip Connections. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). (2018), pp. 3006–3010.
24. X.-J. Mao, C. Shen, Y.-B. Yang: Image Restoration Using Convolutional Auto--encoders with Symmetric Skip Connections. arXiv preprint arXiv: 1606.08921, (2016).
25. A. Dosovitskiy, T. Brox: Generating Images with Perceptual Similarity Metrics based on Deep Networks. arXiv preprint arXiv: 1602.02644, (2016).
26. O. Ronneberger, P. Fischer, T.Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation. LNCS. 9351. 234-241. (2015).

METHODS FOR REDUCTION OF BLUR EFFECT IN THE DENOISING AUTOENCODER MODEL FOR RIB SUPPRESSION IN CXR IMAGES

Abstract

The most widespread, low-cost, and highly available diagnostic method for detecting abnormalities in the cardiopulmonary system is chest radiography. In many cases, the presence of bone structures on the image makes the correct diagnosis much more difficult. There exist few computational tools for rib suppression on x-ray images of the lungs and the denoising autoencoder model seems to be perfect for this task. However, mostly due to the considerable compression of image size in the model, the blurring effect of pathological structures occurs. This work aims to improve the quality of the lung images obtained from a rib bone suppression algorithm based on a denoising autoencoder. The Bone Suppression dataset, consisting of 35 pairs of chest x-ray images and corresponding soft-tissue images, was used to develop the model. Three methods were proposed to reduce the blurring effect: (i) modification of the image resizing scheme through layers of the autoencoder; (ii) modification of a parameter of the loss function, which was responsible for establishing the weights between the mean square error index and the multi-scale structural similarity index; (iii) introduction of bridge connections between the coder and encoder layers. Different scenarios of proposed corrections were examined, but in general, all tested methods showed increased performance in comparison to the baseline model in terms of peak signal-to-noise ratio and structural similarity. Our results show that these methods significantly reduce the denoising autoencoder blurring effect, improving the quality of output image with suppressed ribs, which could potentially influence the medical diagnosis of the patients.

Keywords: x-ray imaging, blurring effect reduction, bone suppression, autoencoders, deep learning