Deep Learning Models for Bone Suppression in Chest Radiographs

Maxim Gusarev, Ramil Kuleev, Adil Khan
Department of Computer Science
Innopolis University
Innopolis, Russia
Email: m.gusarev, r.kuleev, a.khan@innopolis.ru

Adin Ramirez Rivera
Institute of Computing
University of Campinas
Campinas, Brazil
Email: adin@ic.unicamp.br

Asad Masood Khattak
College of Technological Innovation
Zayed University
Abu Dhabi, UAE
Email: asad.khattak@zu.ac.ae

***Abstract*—Bone suppression in lung radiographs is an important task, as it improves the results on other related tasks, such as nodule detection or pathologies classification. In this paper, we propose two architectures that suppress bones in radiographs by treating them as noise. In the proposed methods, we create end-to-end learning frameworks that minimize noise in the images while maintaining sharpness and detail in them. Our results show that our proposed noise-cancellation scheme is robust and does not introduce artifacts into the images.**

***Index Terms*—lung cancer, bone suppression, autoencoder, convolution neural network, deep learning.**

I. Introduction

Chest radiography (CXR) is the most commonly used diagnostic imaging technique for identifying chest diseases such as tuberculosis, pneumonia, and lung cancer. This is because CXR is the most cost-effective, routinely available, and dose-effective diagnostic tool. It also has the ability to reveal certain unsuspected pathologic alterations. Among different chest diseases, lung cancer is responsible for more than 900000 deaths each year, making it the leading cause of cancer-related deaths in the world [1].

Accurate suppression of bony structures is essential for computer-aided diagnosis tasks that include various size measurements, detection, and classification of lung nodules [2] and interstitial lung disease.

There are several reasons why automated analysis of chest radiographs can benefit from suppression of bony structures. A study [3] showed that most lung cancer lesions that are missed on frontal chest radiographs are located behind ribs, and that inspection of a soft tissue image can improve detection performance by humans. Automated lung cancer detection schemes also suffer from false positives caused by superposition of bony structures [4]. Small subtle abnormalities occur often in radiographs of tuberculosis-affected people, and it is expected that the detection of these abnormalities can be improved by rib suppression.

One way to reduce the visual clutter of overlying anatomy in CXRs is dual-energy subtraction (DES) imaging [5]. DES radiography involves capturing two radiographs with the use of two X-ray exposures at two different energy levels. These two radiographs are then combined to form a subtraction image that highlights either soft-tissue or bone components. However, only a few hospitals use a DES system because of the required specialized equipment.

Another way to suppress bones in CXRs is using an image processing technique that does not require specialized equipment for DES. The commercial software ClearRead Bone Suppress of Riverain Technologies [6] is a tool for bone suppression in CXRs that has been approved by FDA and CFDA. Other methods are based on features extracted by analysis of huge DES image datasets [1, 7, 8, 9] or on subtracting the segmented bone structure from the original image [4, 10, 11, 12, 13, 14]. Huge DES datasets are difficult to receive, and intermediate segmentation results require manual or semi-automatic delineation. The presence of such kinds of dependencies motivates us to do research on other effective and dependency-free methods of bone suppression.

In this paper, we propose two architectures to perform bone suppression from CXRs. The architectures are supposed to denoise bones from images instead of subtraction of bone shadows or feature extraction analysis. Thus, they are completely different from what was already shown for bone suppression purposes.

The first model is a family of convolutional autoencoders, while the second one is a family of a simple convolutional neural network (CNN). Our learning framework for both models is driven by a denoising reconstruction function, instead of the, commonly used, identity. In our proposal, we explore a loss function that exploits the structural similarities in the image while maintaining the reconstruction as the main goal. Experimentally, we find the best configuration for our proposed models. Moreover, we found that our models are capable of transforming standard CXR images into soft-tissue images. Our final trained models will have a better lung disease detection rate.

II. Previous Work

We distinguish two categories of existing methods for bone suppression: supervised and unsupervised. The suppression problem becomes a regression-prediction problem in the supervised category. Extracting the useful information and recognizing the characteristic structures from the CXR to estimate the corresponding soft-tissue or bone components are the key problems for the regressors [4, 7, 15]. In most

978-1-4673-8988-4/17/$31.00

cases, regressors are trained or optimized by teaching DES radiographs to estimate the soft-tissue or bone images.

On the other hand, the unsupervised methods for bone suppression do not require teaching DES with radiographs, but these methods need segmentation and the border locations of bony structures. The bone-free images are reconstructed with the blind-source separation approach or the gradient images modified according to the intermediate results. The effectiveness of unsupervised methods highly depends on the accuracy of segmentation and the border locations of bony structures.

A. Supervised Methods

Loog and Van Ginneken [8] proposed a general filter framework based on regression which has been applied to the suppression of bony structures. In their next work [4], they attempted to learn a complex non-linear filter directly from the pixel data. The set of input features did not consist of raw pixel values but of a set of Gaussian derivatives. It gave promising results but was not further evaluated on a clinical problem.

Suzuki et al. [1, 7, 16] employed a massive artificial neural network to predict the bone image from a standard chest radiograph which can be subtracted to yield an image similar to a soft-tissue image. They reported a reduction of the contrast of the ribs and showed that this technique increased the visibility of nodules.

Recently, Yang et al. [9] presented a cascade of convolutional networks with three convolutional layers, trained with 404 dual-energy chest exams to estimate, and subtract the bony image from the input image to obtain a virtual soft tissue image. The authors use a multi-scale approach and estimate the gradient of the bone images successively from coarse to fine scales. The authors show that using a large number of filters leads to improved results. The soft tissue images produced are visually highly convincing, and the technique can also be applied to radiographs from different sources [17]. In Section IV-A, we show that much less data is required for training a convolutional network for bone suppression.

B. Unsupervised Methods

None of the supervised methods combined suppression of normal structure with the evaluation on a texture classification problem. Hogeweg et al. [11] presented a rib subtraction technique based on PCA modeling of rib profiles, and evaluated it using the performance of a system classifying image patches containing textural abnormalities. Fitting these profile models to the data and subtracting them resulted in reasonably convincing rib suppression, but it was applied to the posterior ribs only. In their next work, Hogeweg et al. [12] proposed blind source separation techniques together with outlier identification to estimate an intensity model of the unwanted structures and subsequently remove it from the original image. The method was evaluated on three tasks (removing ribs, clavicles, and catheters in chest radiographs) and showed a reduction of the conspicuity.

Simkó et al. [10] suppressed clavicles by creating a bone model from a gradient map smoothed along the clavicle border direction, after which a clavicle free image was created by subtraction of the model. They showed promising results of clavicle suppression on reducing false positives in a nodule detection task, but applied it only to the lateral part (diaphysis) of the clavicle.

The decomposition of the thorax presented by von Berg et al. [14] comprises the automatic delineation of bone contours with the help of dynamic programming and a suppression method. According to the segmentation accuracy, suppression worked relatively better for posterior rib parts and for the central ribs. The method responded very robustly to segmentation errors.

We treat bony structures in CXRs as noise, and the bone suppressed image as a clear denoised image. We should, therefore, use some image denoising techniques. Jain and Seung [18] proposed image denoising using convolutional neural networks. It was observed that using a small sample of training images, one can achieve performance at par or better than state-of-the-art methods based on wavelets and Markov random fields.

Denoising autoencoders are a recent addition to image denoising literature. They easily outperform conventional denoising methods and are less restrictive for specification of noise generative processes. Vincent and Larochelle [19] introduced stacked denoising autoencoders as an extension to classic autoencoders. They form a deep network by feeding the output of one denoising autoencoder to the one below it. Reviewed by Gondara [20], stacked denoising autoencoders showed the reliable results even on small training datasets.

Agostinelli et al. [21] experimented with adaptive multi-column deep neural networks for image denoising, built using combination of stacked sparse autoencoders. This system was shown to be robust for different noise types. Denoising autoencoders constructed using convolutional layers have better image denoising performance for their ability to exploit strong spatial correlations.

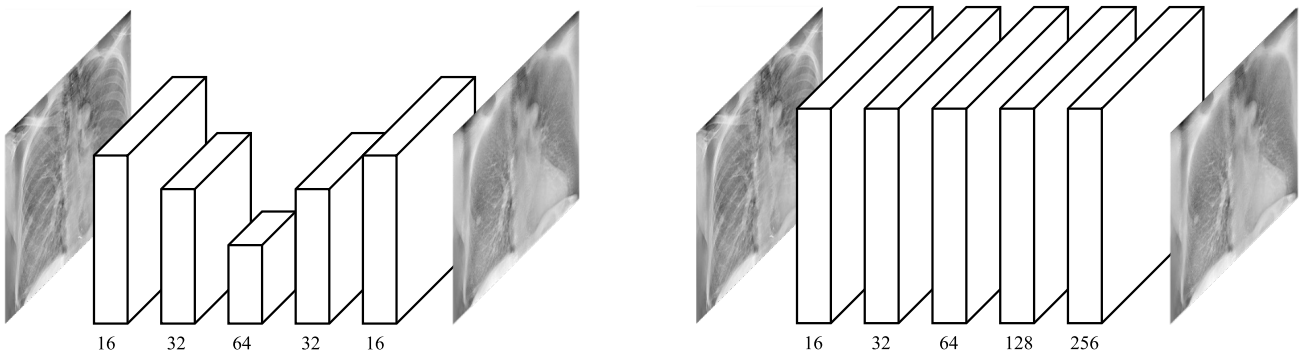
Our methods comprise denoising properties of stacked autoencoders and learning properties of deep convolutional networks. We avoid dependency on intermediate segmentation results, and use a small sample of DES images for training. We present a novel image comparison technique that makes a strong impact on convergence of a learning process and overall performance of bone suppression methods.

III. Proposed Framework

The goal of our framework is to feed a CXR image to the network and produce a bone suppressed image. Moreover, we propose a loss function that teaches the network (in an end-to-end fashion) to perform custom denoising on the input images.

A. Autoencoder-like Convolutional Model

Our first model is a stacked denoising autoencoder (AE). Model output lacks noise defined by user, and, in case of



(a) (b)

Fig. 1: The architectures of convolutional models: (a) Autoencoder-like model, where image size is twice decreased at every encoding layer and twice increased at every decoding one. And a (b) Multilayer neural model, where image size is preserved.

CXRs, the model produces the original image without bones. Overall, the model is a stack of convolutional autoencoders with encoder and decoder sharing the same but mirrored weights. Unlike usual denoising autoencoders, the noise is not normally distributed, and is represented by bony structures. We define our architecture as AE-like, because we are doing a reconstruction not of an original image, but of a bone suppressed, or denoised, one. The model consists of 3 autoencoders, each encoding an image into 16, 32, and 64 neurons, respectively. Figure 1(a) of hourglass shape shows that image size is decreased twice at every encoding layer and increased twice at corresponding decoding layers. Firstly, we optimize the parameters of our model by minimizing mean squared error between the model output and a corresponding soft-tissue image alone, and, secondly, we minimize mean squared error (MSE) along with maximizing multi-scale structural similarity (MS-SSIM) [22] of produced image and a soft-tissue image (the ground truth). We detail the loss functions in Section III-C.

B. Multilayer Convolutional Neural Model

The second model is actually a family of CNNs with 3, 4, 5, and 6 layers. The model has no shared weights, and the hypothesis is that a feature learning process becomes the most effective when provided with certain volume of learning space (number of layers and neurons). In other words, we should discover how much space it needs with the help of the family of 4 CNN models.

There are 16 and 32 hidden neurons at the 1st and the 2nd layers, respectively. Then each model of the family has twice more neurons at its last but one layer than the previous model had. The last layer is an output layer that produces a set of bone suppressed images. The 4-layer network has 64 neurons at the 3rd layer, the 5-layer network has 128 neurons at 4th layer, and the 6-layer network has 256 neurons at the 5th layer. The 6-layer network is shown in Figure 1(b). Image size is preserved through all the layers of CNNs. We optimize the parameters of models by minimizing MSE and maximizing MS-SSIM of the model output and the ground truth.

C. Loss Functions

Our goal is to train a network that learns how to remove the bone structures from the CXR without introducing artifacts to it. To achieve this goal, our approach is to reduce the reconstruction error from the denoised image to its original counterpart, and to maximize a structural index to maintain sharpness.

We have found that a correct choice of a loss function has a strong effect on the reconstruction results. Thus, we explore different configurations to find the best loss function for our denoising framework. For an error function $\mathcal{E}$, the loss function of an image I can be written as

$$\mathcal{L}^{\mathcal{E}}(I) = \frac{1}{N}\sum_{i \in I} \mathcal{E}(i), \tag{1}$$

where N is the number of pixels in the image I, and i is an index of a pixel in the image I.

MSE, also called ℓ_2, is de facto standard error function for reconstruction. It penalizes larger errors, but is more tolerant to small ones. MSE treats pixels separately and is not influenced by their local neighborhoods. Since MSE does not capture the intricate characteristics of the human visual system (HVS), we define the loss function for $\mathcal{E} \coloneqq \ell_2$ as

$$\mathcal{L}^{\ell_2}(I) = \frac{1}{N}\sum_{i \in I} \big(I(i) - G(i)\big)^2, \tag{2}$$

where i is the index of a pixel, I is the original image, and G is the ground truth image of I.

Due to the limitations of MSE, another popular reference-based index is the structural similarity index (SSIM). It evaluates images accounting for sensitivity of the

HVS to changes in local structure [23]. SSIM for pixel i is defined as

$$\text{SSIM}(i) = \frac{2\mu_I(i)\mu_G(i) + C_1}{\mu_I(i)^2 + \mu_G(i)^2 + C_1} \cdot \frac{2\sigma_{IG}(i) + C_2}{\sigma_I(i)^2 + \sigma_G(i)^2 + C_2}, \tag{3}$$

$$= l(i) \cdot cs(i), \tag{4}$$

where $\{\mu_I(i), \sigma_i(i)\}$ and $\{\mu_G(i), \sigma_G(i)\}$ are the pairs of mean and standard deviation for a local neighborhood centered at pixel index i of the given image, I, and its ground truth, G, respectively; $\sigma_{IG}(i)$ is the covariance of both local neighborhoods; C_1 and C_2 are two variables to avoid division by zero. Note that we omitted the dependence of means and standard deviations on pixel i. Means and standard deviations are computed with a Gaussian filter with standard deviation $\sigma_{\mathcal{G}}$. Since SSIM is inspired in the HVS, its terms represent the comparisons between luminance, l, contrast and structure, cs, measurements in the local neighborhoods.

Wang et al. [22] proposed MS-SSIM, a multi-scale version of SSIM that weighs SSIM computed at different scales according to the sensitivity of the HVS, and experimental results have shown the superiority of SSIM-based indexes over ℓ_2. Given a dyadic pyramid of M levels, MS-SSIM is defined as

$$\text{MS-SSIM}(i) = l_M^{\alpha}(i) \cdot \prod_{j=1}^{M} cs_j^{\beta_j}(i), \tag{5}$$

where l_M and cs_j are the luminance, contrast and structure terms defined in (4) at scale M and j, respectively. For convenience, we set $\alpha = \beta_j = 1$, for $j = \{1, \ldots, M\}$, as suggested by Wang et al. [22]. The loss function based on MS-SSIM can be then written setting $\mathcal{E}(i) = 1 - \text{MS-SSIM}(i)$, such as

$$\mathcal{L}^{\text{MS-SSIM}}(I) = \frac{1}{N} \sum_{i \in I} 1 - \text{MS-SSIM}(i). \tag{6}$$

Zhao et al. [24] showed that SSIM and MS-SSIM do not perform as well as mean absolute error, also called ℓ_1. They also concluded that the combination of MS-SSIM and ℓ_1 yields the best trade-offs. However, MS-SSIM preserves the contrast in high-frequency regions better than the other loss functions they experimented with. On the other hand, ℓ_1 preserves colors and luminance (an error is weighed equally regardless of the local structure) but does not produce quite the same contrast as MS-SSIM.

The MSE function is convex and differentiable, which are much better properties for optimization purpose than ℓ_1 function has. Being insensitive to small changes, ℓ_2 is balanced by MS-SSIM loss that pays attention primarily to local neighborhood changes. We chose the combination of MS-SSIM and ℓ_2 as our loss function:

$$\mathcal{L}^{\text{Mix}} = \alpha \cdot \mathcal{L}^{\text{MS-SSIM}} + (1 - \alpha) \cdot \mathcal{L}^{\ell_2}, \tag{7}$$

where we omitted the dependence on image I for all loss functions, and we set $\alpha = 0.84$ as Zhao et al. [24] empirically derived.

IV. Experiments and Results

A. Data Acquisition and Preprocessing

Our data is represented by 35 pairs of chest radiograph and its soft-tissue versions. Most of the soft-tissue images are results of dual-energy subtraction. All the data was acquired from different online sources. We provide this dataset for educational purposes by request.

Using combined affine transformations we augmented 4000 image pairs from 35 initial ones. The transformations include rotations, horizontal and vertical shifts, shear, zoom, intensity shifts, and horizontal flips. Images were cropped to square form and resized to the size of 440×440 pixels. We apply the contrast limited adaptive histogram equalization (CLAHE) for local contrast enhancement, so local details can therefore be enhanced even in regions that are darker or lighter than most of the image. Images were standardized by centering to the mean and scaling to unit variance. Pixel values were scaled to the $[0, 1]$ range.

B. Deep Learning Properties

The type of each layer of our architectures is convolutional. Size of the filters used for convolution remains the same for all of the models and layers: 5×5. The strides of AE-like model are $[1, 2, 2, 1]$, and the strides of CNN family models are $[1, 1, 1, 1]$. The activation function of each layer is Rectified Linear Unit (ReLU). Further we describe details common for all of our models.

1) Training Phase: All the weights values are initialized by normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.02$. We use the Adam algorithm as a learning algorithm.

The initial learning rate is 0.001. We chose the small number of epochs of training for an experiment purpose, and learning rate therefore remains unchanged. It will be decreased by 25% each 100 epochs when the training continues.

Training data is loaded into a queue which keeps a minimum of 4000 images after dequeue assuring the good shuffle quality. Images are read in batches of 5 images, and 800 batches are processed during an epoch.

The stopping criteria is represented by a fixed number of epochs: for the experiment purpose we train models for only 150 epochs.

2) Test Phase: Test dataset consists of 10 pairs of images that are not derived from each other at the stage of augmentation. Train dataset consists of the rest of the images acquired at the data augmentation phase in Section IV-A.

3) Implementation: We used GPU version of TensorFlow framework [25] for data processing, model implementation and training, and Keras framework [26] for data augmentation. NVIDIA GeForce GTX 980M GPU with 4 GB of VRAM was used as a main computing unit.

C. Training and Testing

The results of application of different models described below to a test image are presented at Fig. 2. The loss function values of the whole test dataset are shown at Fig. 3.

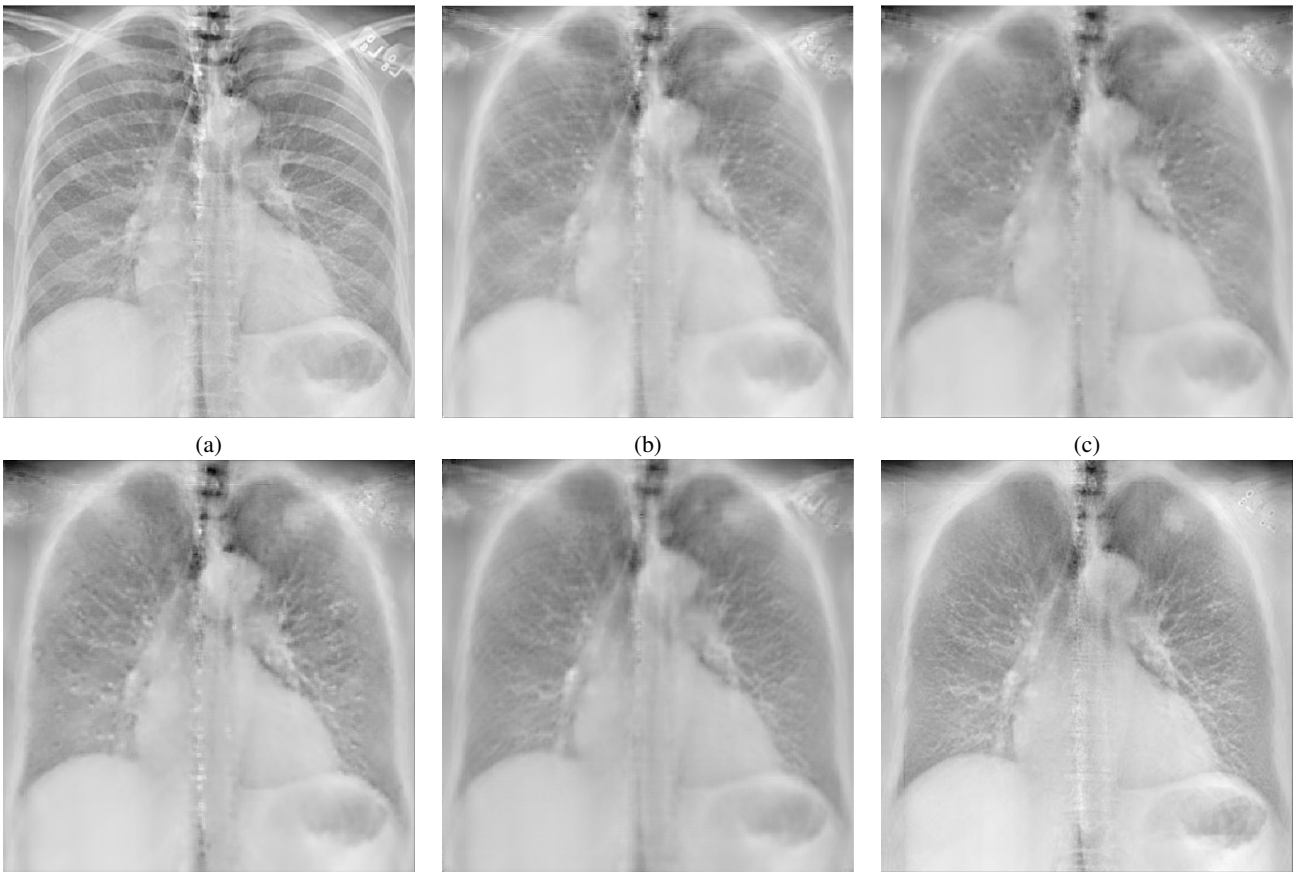


Fig. 2: Results of our proposed models when applied to (a) the original image. Images produced by applying the CNN with (b) 4 layers, (c) 5 layers, and (d) 6 layers, and (e) the AE-like model. Compare them against (f) the soft-tissue image (ground truth).

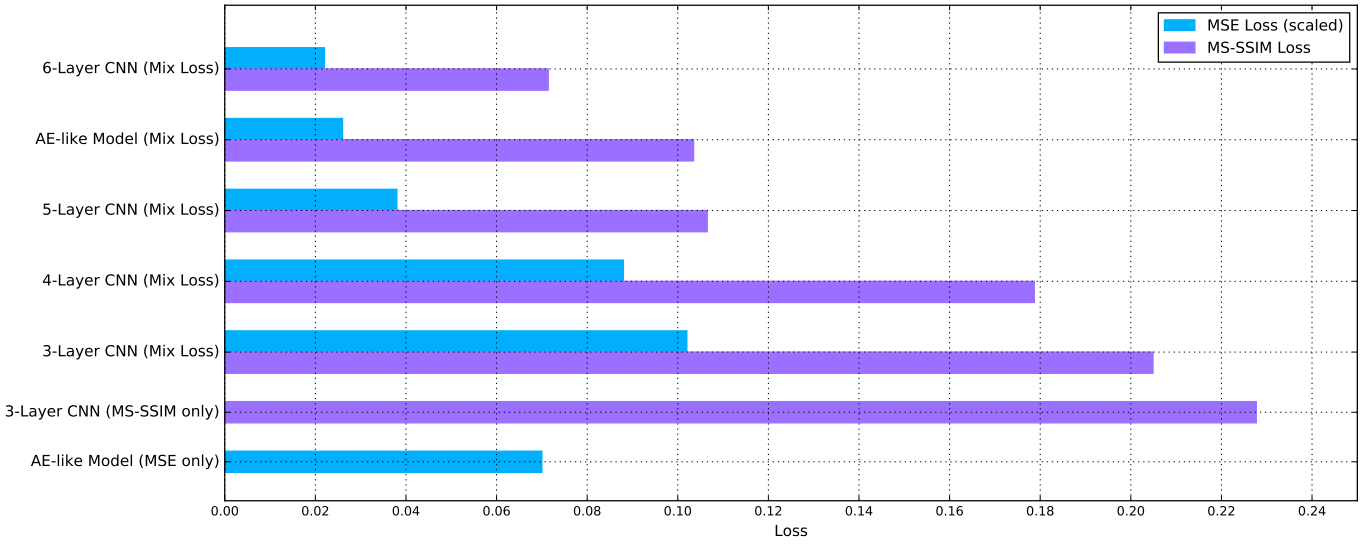


Fig. 3: Losses of different models. MSE loss values are 20 times scaled. MS-SSIM loss preserves its values. Smaller is better.

First of all, we ran the AE-like model with MSE loss (2). The loss showed 0.0035. The produced images had hardly noticeable bone shadows. The bones were suppressed but most of the lung nodules were blurred. The images lacked sharpness.

The simple convolutional neural network with 3 layers and 16 hidden neurons at the first layer, and 32 hidden neurons at the second layer was the first in the family of tested CNNs. MS-SSIM loss (6) was chosen to be tested next. SSIM-based index provided sharp images but reached MS-SSIM loss of only 0.2277. The produced images were clearly readable, but had all the bone structures present.

We considered the problem of choosing another loss function to be the most important. The decision was made to combine the best of both worlds: MS-SSIM index and MSE. We tested AE-like model again but this time with combination of loss functions (7). The model reached MS-SSIM loss of 0.1035 and MSE of 0.0013. The produced images were much more clearer. The bones were hardly noticeable—as shown in Fig. 2(e).

The same combination of MSE and MS-SSIM loss functions was used for training the 3-layer CNN. The MS-SSIM loss was 0.2049, and MSE was 0.0051. The bones were still present, and overall image was slightly blurred. The AE-like model performed better than 3-layer CNN.

We decided to add a convolutional layer and increase the number of its hidden neurons. We increased the depth of the network to give the CNN more space for feature learning. The 4-layer CNN with 64 neurons at the 3rd layer was trained and produced an image set with MS-SSIM loss of 0.1787 and MSE of 0.0044—as shown in Fig. 2(b).

The 5-layer CNN with 128 neurons at the 4th layer was trained and produced an image set with MS-SSIM loss of 0.1065 and MSE of 0.0019. It was an important leap comparing to AE-like model. The bones were hardly noticeable while lung nodules reproduction was the best of all the experiments—as shown in Fig. 2(c). The model was still unacceptable for lung disease analysis because of some missing nodules and blurred areas. With 5-layer CNN we achieved the same results as with the AE-like model.

The 6-layer CNN with 256 neurons at the 5th layer was trained and produced an image set with MS-SSIM loss of 0.0714 and MSE of 0.0011. The images were completely bone suppressed and sharp, and all the soft-tissue elements were preserved—as shown in Fig. 2(d). The 6-layer model is ready to be used for lung disease analysis.

V. Conclusion

In this paper we evaluated two models: AE-like and CNN, and three different loss functions to maintain the content of the image (while keeping it sharp). Both models gave some workable results only after applying a correct loss functions combination. A deeper research in the area of image comparison with optimization friendly properties should result in more effective and precise metrics.

The successful use of an autoencoder-based model with a denoising loss function shows that extracted features can be used to reproduce images without not only normally distributed noise, but any components that are treated as a noise.

The consecutive increase of number of convolutional layers and number of neurons at these layers shows the dependency of image production quality on learning space provided to the model. The later experiments will define the parameters needed for such dependence to effectively converge.

Acknowledgement

This research work was supported by Zayed University Research Initiative Fund # R17057.

References

[1] S. Chen and K. Suzuki, "Bone suppression in chest radiographs by means of anatomically specific multiple massive-training ANNs combined with total variation minimization smoothing and consistency processing," in *Computational Intelligence in Biomedical Imaging*, 2014.

[2] F. Li, R. Engelmann, L. L. Pesce, K. Doi, C. E. Metz, and H. Macmahon, "Small Lung Cancers: Improved Detection by Use of Bone Suppression ImagingComparison with Dual-Energy Subtraction Chest Radiography 1," *Radiology*, vol. 261, 2011.

[3] P. Shah, J. Austin, C. White, P. Patel, L. Haramati, G. Pearson, M. Shiau, and Y. Berkmen, "Missed non-small cell lung cancer: radiographic findings of potentially resectable lesions evident only in retrospect," Weill Cornell Medical Center, New York-Presbyterian Hospital, NY, USA, Tech. Rep., 2003.

[4] M. Loog, B. van Ginneken, and A. M. R. Schilham, "Filter learning: Application to suppression of bony structures from chest radiographs," *Medical Image Analysis*, 2006.

[5] P. Vock and Z. Szucs-Farkas, "Dual energy subtraction: Principles and clinical applications," 2009.

[6] Riverain Technologies, "Bone Suppression," 2017. [Online]. Available: http://www.riveraintech.com

[7] K. Suzuki, H. Abe, H. MacMahon, and K. Doi, "Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN)," *IEEE Transactions on Medical Imaging*, 2006.

[8] M. Loog and B. Van Ginneken, "Bony Structure Suppression in Chest Radiographs," 2006.

[9] W. Yang, Y. Chen, Y. Liu, L. Zhong, G. Qin, Z. Lu, Q. Feng, and W. Chen, "Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain," *Medical Image Analysis*, 2017.

[10] G. Simkó, G. Orbán, P. Máday, and G. Horváth, "Elimination of clavicle shadows to help automatic lung nodule detection on chest radiographs," in *4th European Con-*

ference of the International Federation for Medical and Biological Engineering, 2009, p. 488491.

[11] L. E. Hogeweg, C. Mol, P. A. de Jong, and B. van Ginneken, "Rib suppression in chest radiographs to improve classification of textural abnormalities," in *Proc. SPIE 7624, Medical Imaging 2010: Computer-Aided Diagnosis, 76240Y (March 09, 2010)*, 2010.

[12] L. Hogeweg, C. I. Sanchez, and B. Van Ginneken, "Suppression of translucent elongated structures: Applications in chest radiography," *IEEE Transactions on Medical Imaging*, 2013.

[13] J. Von Berg, S. Young, H. Carolus, R. Wolz, A. Saalbach, A. Hidalgo, A. Giménez, and T. Franquet, "A novel bone suppression method that improves lung nodule detection Suppressing dedicated bone shadows in radiographs while preserving the remaining signal," 2015.

[14] J. von Berg, C. Levrier, H. Carolus, S. Young, A. Saalbach, and P. Laurent, "Decomposing the bony thorax in X-ray images," in *2016 IEEE 13th International Symposium on Biomedical Imaging*, 2016.

[15] S. Chen and K. Suzuki, "Bone suppression in chest radiographs by means of anatomically specific multiple massive-training ANNs combined with total variation minimization smoothing and consistency processing," in *Computational Intelligence in Biomedical Imaging*, 2014.

[16] K. Suzuki, H. Abe, F. Li, and K. Doi, "Suppression of the contrast of ribs in chest radiographs by means of massive training artificial neural network," 2004.

[17] B. Van Ginneken, "Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning," *Radiological Physics and Technology*, vol. 10, 2017.

[18] V. Jain and H. S. Seung, "Natural Image Denoising with Convolutional Networks," 2008.

[19] P. Vincent and H. Larochelle, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion Pierre-Antoine Manzagol," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.

[20] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," 2016.

[21] F. Agostinelli, M. R. Anderson, and H. Lee, "Adaptive Multi-Column Deep Neural Networks with Application to Robust Image Denoising," 2013.

[22] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-Scale Structural Similarity For Image Quality Assessment," 2003.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004.

[24] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Neural Networks for Image Processing," 2016.

[25] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, and G. Research, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," 2015.

[26] F. Chollet, "Keras," 2015. [Online]. Available: https://github.com/fchollet/keras