Numerical analysis of SLSSIM similarity on medical X-ray image domain

Jonas Brusokas
Institute of Computer Science
Vilnius University
Vilnius, Lithuania
jonas.brusokas@mif.vu.lt

Linas Petkevičius
Institute of Computer Science
Vilnius University
Vilnius, Lithuania
linas.petkevicius@mif.vu.lt

***Abstract*—The X-ray has been adopted and used for various purposes including medical diagnostics. To remove noise created by new low dose X-ray imaging procedures and reduce medical image size, X-ray image reconstruction and lossless compression using deep neural networks are being researched. To enable this, image similarity metrics capable of performing well on X-ray images must be used. In this paper, the requirements for medical X-ray similarity metrics are defined. A new similarity metric is proposed taking into account the quality of structures within different intensity levels. An analysis is given comparing the proposed and other currently known metrics performance on real X-ray images in simulated scenarios.**

***Keywords*—Image similarity metrics, Low dose X-ray imaging, Medical X-ray images**

I. Introduction

The X-ray ever since its inception has been widely adopted and used for various purposes. One of the key uses of X-ray imaging is in medical diagnostics. It allows for a non-invasive method of diagnosing various bone structure defects, infections, arthritis, and most cancers [1]–[3].

Despite wide usage and rapid technological advances the ionizing radiation emitted during X-ray imaging procedures creates real health risks for patients [4]–[6]. Efforts have been made to reduce the risks associated with radiation exposure by creating new procedures for performing X-ray imaging. These procedures (commonly referred to as low-dose) use lower voltage or amperage settings to reduce the amount of radiation emitted thus reducing the risks [7]. Unfortunately, using these types of procedures results in images having artifacts and noises which can reduce diagnostic suitability [8]. There are cases where it might not be necessary to remove the noises in order to give an accurate diagnosis [9]. In most cases, however, it is imperative to remove or reduce the amount of noise on an X-ray image. Conventional approaches that use defined properties of noise distribution do not yield satisfactory results [10], [11].

There have been some successful attempts in using deep neural networks for image reconstruction tasks, outperforming other approaches in removing various noises from images [12]–[14]. There have also been attempts in using deep neural networks for lossless image compression, to make storing images and using them for calculation more efficient [15]. Furthermore, a significant amount of research has recently been conducted on finding ways of improving X-ray imaging procedures, especially on automated means of disease or abnormality detection [2], [3], [16].

One of the key factors to the accuracy and effectiveness of a deep neural network is the objective function. In image reconstruction and compression image similarity metrics are used as objective functions. They compare two images with each other and produce a scalar result denoting their degree of difference [17]. The similarity metric defines what image properties are being evaluated so the selection of an effective metric is crucial.

In this paper a similarity metric for medical X-ray images domain is proposed and compared with other known metrics. In section II essential properties of X-ray images in the medical domain and requirements for a similarity metric are defined. An overview of currently used metrics is made in section III and the definition of the proposed metric is made in section IV. The analysis comparing the performance of the metrics is made in section V.

II. Requirements for medical X-ray image similarity metrics

Medical X-ray images are created and used to enable trained radiologists to analyse the human body, determine and diagnose various irregularities and illnesses. X-ray images can be done over any part of the human body and as such contain various structures including bones, tissue, and organs [1], [2]. Any assessment of X-ray image similarity or quality must take into account properties of the images which enable them to be used for accurate and reliable diagnosis.

Through analysis of literature and working with trained specialists in the field of radiography several important requirements for medical X-ray image similarity metrics have been identified:

R1 The metric must detect emerging noises and artifacts from images created with low dose X-ray procedures. As was previously stated, in most cases to enable low dose procedures, it is important to detect and remove noises from the created images as they hinder diagnosis [8], [13].

R2 The metric must detect perceptual geometric distortions. They can appear on images during imaging, reconstruction or decompression. From a radiologists perspective the distortions may cause difficulties or even render the images completely unusable for diagnostics [18].

R3 The metric must take into account the discrepancies of images within specific ranges of brightness intensity levels. Due to the nature of the X-ray, images contain various human body structures including bones, tissue, organs. They are captured as areas of different brightness intensities [1]–[3]. In this paper we will refer to them as sub-levels. Most image quality assessment approaches attempt to evaluate the broader perceptual view of image quality and focus on more visible areas. In medical X-ray images, all structures captured in these images have significant value and greatly contribute to diagnosis [13]. Taking into account reconstruction or decompression accuracy of not only the whole picture but also structures within different intensity ranges is critical.

The proposed metric (in section IV) and following experiments (in section V) will take these properties into account.

III. Selected metrics

Research into image similarity metrics (in some cases referred to as image quality assessment) has witnessed attention and notable progress over the past decades [19]. Image quality assessment is essential in fields that use image processing [20]. The majority of metrics used evaluate the similarity of the distorted image use the complete reference image [19]. In this paper 3 selected general purpose similarity metrics which are potentially usable in the X-ray image domain will be discussed and compared. It is important to note, that for the rest of the paper metric definitions will be used, where metric value 1 means the compared images are equal, and value 0, means they are completely dissimilar.

A. Mean squared error (MSE)

Mean squared error is considered to be one of the most simple and straight-forward similarity metrics. It is computed by averaging the intensity differences of the two compared images [17]. MSE is known for being quite mathematically convenient for optimization purposes. Although the metric does not correlate well with perceived visual quality [21]. Despite its flaws it still remains widely used for many image processing tasks. For the purposes of this paper, this metrics' performance will be compared to others during analysis. The following normalized form of the metric will be used:

$$MSE^*(X,Y) = 1 - \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2 \ / \ \ M^2 \qquad (1)$$

Here X and Y are compared images, M - maximal pixel intensity, X_i and Y_i - a concrete pixel in the image, n - total number of pixels in image.

B. Structural similarity index (SSIM)

As a response to unsatisfactory perceived visual quality results of MSE and similar metrics a new approach was proposed to construct new similarity metrics. It was based on the idea that images are highly structured and rather than evaluating pixels individually (like in MSE), groups of spatially proximate and related pixels should be taken into account together in order to achieve more accurate representation of perceivable quality [17]. This is based on the assumption that the human visual system is adapted to identify various structures within field of sight and effectively spot any distortions affecting them [17], [22]. A well-known metric based on these principles is the Structural Similarity Index (SSIM) [23]. SSIM metric calculates in account three key features of the images: luminance changes, contrast changes and structural changes [17]. In this paper it will be defined and used in its standard form:

$$\text{SSIM}^*(X,Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \qquad (2)$$

SSIM has seen wide usage for many image processing tasks [23]. It has served as inspiration and basis for many other similarity metrics supplementing additional features to the standard metric [24]–[28].

C. Weighted SSIM (wSSIM)

Weighted SSIM is a general-purpose modification of the SSIM metric mentioned in sub-section III-B. It was created to be used as an objective function in Deep neural networks. It is a composition of SSIM and traditional L_1 loss. Intuitively, SSIM gives the perceptual image assessment and L_1 decreases the metric value for more distorted images and increases for less distorted images [29]. The authors claimed that if used for training, it would put more emphasis on pairs which are performing worse and increase training speed and accuracy. In this paper L_1 will be defined in its normalized form and wSSIM will be defined in the following form:

$$L_1^*(X,Y) = 1 - |X - Y|_{L_1} \ / \ \ (M \cdot N) \qquad (3)$$

$$wSSIM^*(X,Y) = 1 - SSIM^* \cdot (1 - L_1^*) \qquad (4)$$

Here X and Y are compared images, N - number of pixels in the images, M - maximal pixel intensity.

IV. Proposed SLSSIM metric

Substantial research regarding using Deep neural networks as a method for medical image reconstruction and processing is very recent [30], [31]. Although a significant amount of metrics have been developed to suit different domains, there is no definitive similarity metric for comparing medical X-ray images.

As mentioned in section II a metric for medical X-ray image similarity assessment must take into account specific properties of the images. Also, it has become clear from image quality measurement research that handcrafting and tailoring a similarity metric by combining several known metrics can

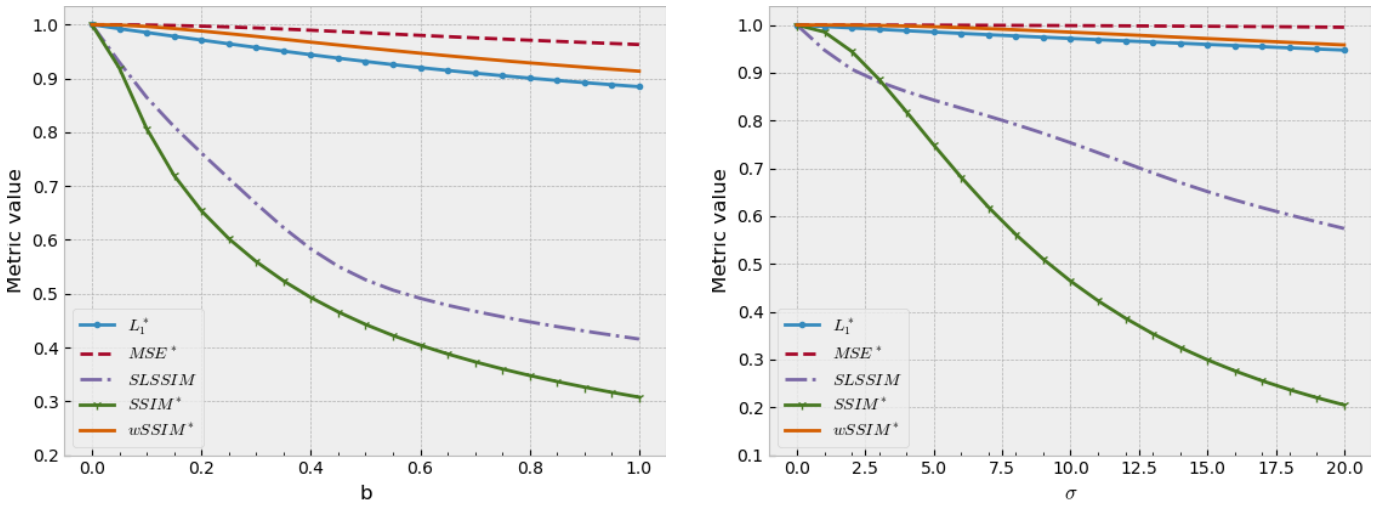

Fig. 1. Results with Poisson Gauss noise augmentation (left), additive Gauss noise augmentation (right)

yield significantly better performance than using any single known metric [32], [34].

In this paper a similarity metric for assessing medical X-ray image similarity is proposed. It is called the sub-level structural similarity index (or SLSSIM). As its name implies it is based on the widely used structural similarity index (or SSIM). Its main differentiating feature is the consideration of human body structures appearing within different sub-levels of the image (as mentioned in requirement R3). It is achieved by calculating SSIM values on discrete sub-levels and not the entire image. This should enable more strict comparison of the structures appearing within the sub-levels. In this paper sub-levels are mathematically defined as:

$$X^{(i)}, Y^{(i)} \in \left[\frac{iM'}{k}, \frac{(i+1)M'}{k}\right) \tag{5}$$

Here X and Y are images, $X^{(i)}, Y^{(i)}$ are sub-levels of the images, N - number of pixels in image, M' - amount of different values that a pixel can have, k - is the number of compared discrete sub-levels. The mathematical definition of SLSSIM is:

$$SLSSIM = \sqrt[k+1]{(1 - L_1^*) \cdot \prod_{i=0}^{k} (SSIM^*(X^{(i)}, Y^{(i)}))} \tag{6}$$

SLSSIM is a root of a product where the members are normalized L_1 calculated from the entire image and SSIM calculated from defined sub-levels. Here, L_1 is used the same way as in subsection III-C, to decrease the metric value further if the image is more distorted and. Also, as all of the product members have values in interval $[0; 1]$ the root is used to prevent a steep downward gradient of the metric. We will be using 8 sub-levels ($k = 8$) for metrics performance analysis.

V. Metrics analysis

Analysis has been carried out to evaluate performance of the selected metrics in real-life scenarios. Tests were conducted using real X-ray images from openly available medical X-ray image data sets. The data sets used for analysis were:

1) RSNA pediatric image set created by the Radiological society of North America, which contains hand X-ray images [35];
2) NIH image set created by the National Institute of Health (of the United States of America) which contains chest X-ray images [36];
3) MURA musculoskeletal image set created by Stanford University which contains elbow, forearm, hand and shoulder X-ray images [37];

For the analysis 3000 images were selected from the data sets (1000 from each) in a random order. All medical images within the data sets were gray-scale, with pixel value ranging within the interval $[0; 255]$.

Image augmentations have been used to simulate noises and distortions in the X-ray images to simulate real-life medical X-ray imaging scenarios. To analyse performance values of metrics were calculated on image pairs that contained reference image – unmodified image from a data set, and augmented image – created from reference image by applying an augmentation. Similarity between the two images was evaluated using the selected and proposed similarity metrics from sections III and IV. The results were aggregated by taking the mean metric value from all the images for each augmentation level.

A. Noise detection results

As defined by requirement R1 in section II a medical X-ray image similarity metric must detect emerging noises. Two image augmentations simulating real life scenarios have been selected to generate augmented images for testing:

- Poisson-Gauss noise augmentation $\mathcal{PG}(a, b)$ – represents artifacts and noises emerging in images obtained via computed tomography X-ray imaging when using X-ray sensors in low dosage configuration [8]. The augmented

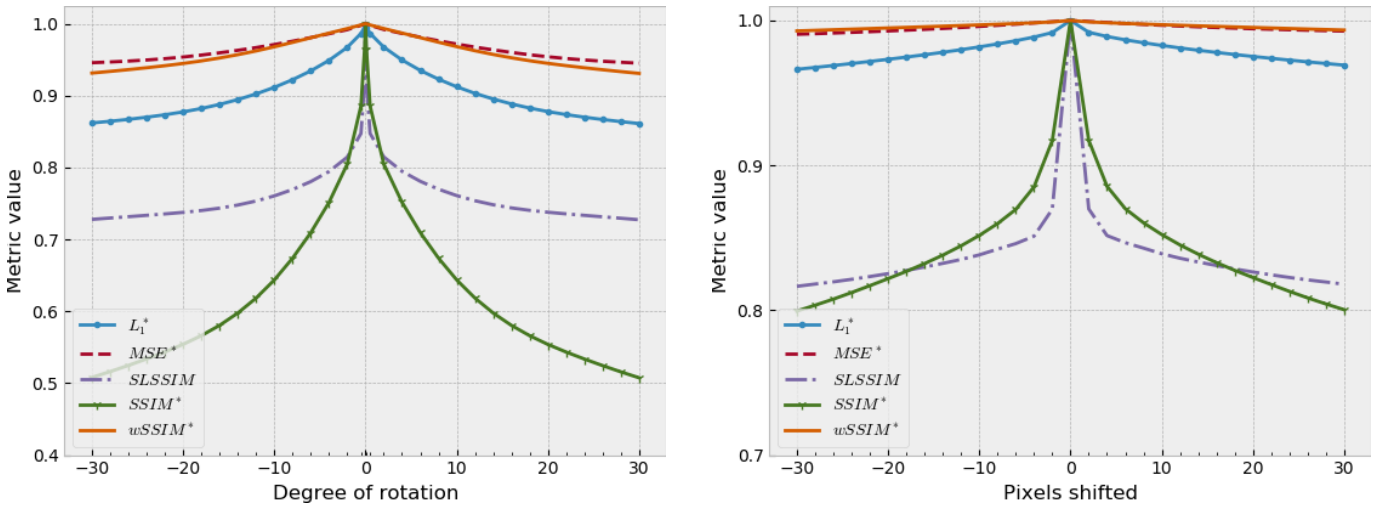

Fig. 2. Results with rotation augmentation (left), Y-axis translation augmentation (right)

image was generated using parameters: $a = 31$ and $b = 0.05k, k \in [0; 20], k \in \mathbb{Z}$. Higher b values result in more noise in the image. There are confirmed results stating that noises generated with parameter $b \approx 0.1$ can occur in low dose imaging [8].

- Additive Gauss noise augmentation $\mathcal{N}(\mu, \sigma)$ – represents general noises and artifacts that can appear in X-ray images. Some research points to its applicability in simulating noises emerging from low dose imaging as well [38]. The augmented image was generated using parameters: $\mu = 0, \sigma \in [0; 20], \sigma \in \mathbb{Z}$. Higher σ values result in more noise in the image. Visually, the noise becomes noticeable at $\sigma \approx 5$.

1) Poisson Gauss noise detection results: On the left side of figure 1 similarity metrics ability to detect emerging Poisson-Gauss noise is observed. Metrics L_1^*, MSE^*, $wSSIM^*$ do not detect the noise even when the noise parameter b is high (images are highly disrupted). Whereas, $SSIM^*$ metric and the proposed metric $SLSSIM$ perform well. When noise parameter $b = 0.1$, metric values $SLSSIM \approx 0.78$ and $SSIM^* \approx 0.66$. These results are acceptable, as visually the noise should still allow for accurate diagnosis.

2) Additive Gauss noise detection results: On the right side of figure 1 similarity metrics performance against emerging additive Gauss noise is witnessed. Similarly to the previous results, metrics L_1^*, MSE^*, $wSSIM^*$ do not detect the noise well, the resulting values do not drop below 0.94. However, in this case the difference between the proposed $SLSSIM$ metric and $SSIM^*$ is higher. When noise parameter $\sigma = 5$, metric values $SLSSIM \approx 0.84$ and $SSIM^* \approx 0.76$. Although these values are still acceptable, the rate at which $SLSSIM$ value decreases as the noise level rises is not as representative of the added disruptions as the $SSIM^*$. It can be stated that $SSIM$ is slightly more accurate in detecting additive Gauss noise.

B. Geometric distortion detection results

As defined by requirement R2 in section II a medical X-ray image similarity metric must also detect perceptual geometric distortions. Two widely used image augmentations have been selected to simulate geometric transformations and generate augmented images for testing:

- Translation augmentation – used to evaluate general metric robustness against translation (shifting of the image). Theoretically, translation can occur during reconstruction or decompression. The augmented image was generated by translating the reference image on the Y axis by y pixels, where values $y = \{-30, -28, ..., 26, 28, 30\}$
- Rotation augmentation – also used to evaluate general metric robustness. The augmented image was generated by rotating the reference image by r degrees, where values $r = \{-30, -28, ..., 26, 28, 30\}$

1) Rotation augmentation detection results: On the left side of figure 2 similarity metrics ability to detect rotations is observed. As an emerging pattern, metrics L_1^*, MSE^*, $wSSIM^*$ do not detect the distortions even when the degree of rotation is very high. $SSIM^*$ and $SLSSIM$ still perform well, with $SSIM^*$ once again being slightly more sensitive to the level of distortion.

2) Translation augmentation detection results: On the right side of figure 1 similarity metrics performance in detecting translations is observed. Once again, metrics L_1^*, MSE^*, $wSSIM^*$ do not detect distortions well and metrics $SSIM^*$, $SLSSIM*$ do so.

C. Analysis results

As observed in the tests L_1^*, MSE^*, $wSSIM^*$ metrics are not suitable for accurately detecting noises or geometric transformations in medical X-ray images. While tests show that $SSIM^*$ is more sensitive to additive Gauss noise, but $SLSSIM$ has an advantage in being the only metric that specifically measures discrepancies in sub-levels of the image.

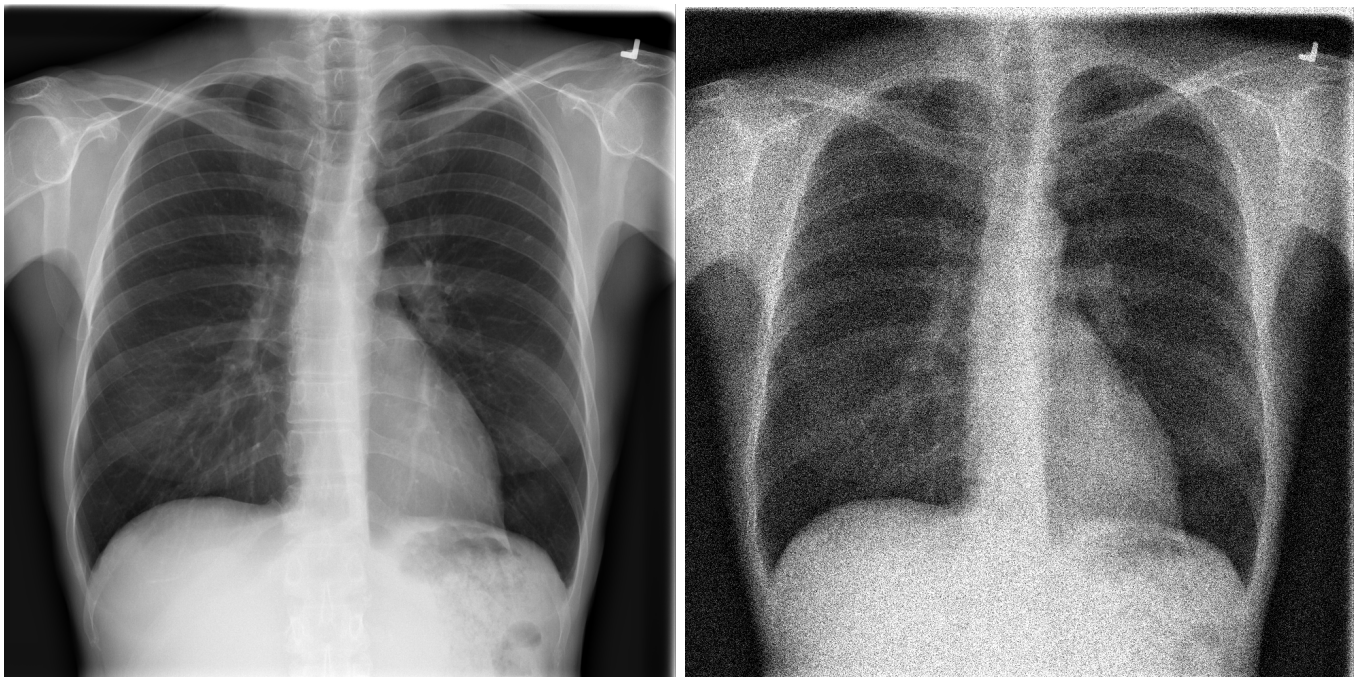

Fig. 3. Sample chest X-ray image from NIH data set (left), same image distorted with Poisson Gauss noise (right)

Fig. 4. Sample wrist X-ray image from MURA data set (left), image distorted with Poisson Gauss noise (center), visualisation of sub-level range $[64; 127]$ of distorted image (right)

Both $SSIM^*$ and $SLSSIM$ metrics are potentially suitable for use with medical X-ray images.

VI. Samples of metrics behaviour on X-Ray images

Sample images from the data sets have been displayed to better visualise the impact that augmentations have on the images. Also, metric values are given for each image pair.

In figure 3 a reference X-Ray image from NIH chest X-ray image data set can be seen on the left. An augmented image distorted by Poisson-Gaussian noise, $a = 31, b = 0.08$ can be seen on the right. It simulates the effect that low dose imaging would have on the image [8]. The metric values calculated from reference and augmented images are: $MSE^* = 0.99$, $L_1^* = 0.93$, $SSIM^* = 0.10$, $wSSIM^* = 0.93$ and $SLSSIM = 0.36$.

Similarly, in figure 4 a reference X-Ray image from the MURA musculoskeletal X-ray image data set can be seen on the left. An augmented image distorted by Poisson-Gaussian noise, with same parameters as before can be seen in the center. The image on the right represents a visualisation of

the augmented image sub-level of intensity range $[64; 127]$. In this particular case, this sub-level represents the bone structure from the image, with tissue and other matter being barely represented. The metric values calculated from reference and augmented images are: $MSE^* = 0.98$, $L_1^* = 0.88$, $SSIM^* = 0.08$, $wSSIM^* = 0.89$ and $SLSSIM = 0.27$. It is important to note that the metrics values taken from images seen in figures 3 and 4 correspond with the results of the analysis in section V.

VII. Conclusions

Although there are many general purpose image similarity metrics used in various domains, only a few of them are suitable for use in comparing medical X-ray images. Based on the analysis performed on multiple medical X-ray image data sets and using different augmentations, L_1^*, MSE^*, $wSSIM^*$ metrics are ineffective in this domain as metric values never drop below 0.8 even when significant distortions are applied. $SSIM^*$ and the proposed $SLSSIM$ can be used to effectively compare X-ray images as both metrics' value is much lower when comparing heavily distorted images, in some cases going below 0.3. Both metrics perform similarly in tested cases. Unlike $SSIM^*$, $SLSSIM$ takes into account differences between human body structures in the images, which should allow diagnosticians to make more accurate diagnoses. The proposed $SLSSIM$ can be used as a basis of an objective function in medical X-ray image reconstruction or compression tasks.

References

[1] U. Eglė, "Diagnostic X-ray hardware," 2011, (In Lithuanian).

[2] M. Woźniak and D. Połap, "Bio-inspired methods modeled for respiratory disease detection from medical images," Swarm Evol. Comput., vol. 41, pp. 69–96, Aug. 2018.

[3] Wozniak, M., Polap, D., Kosmider, L., Napoli, C., and Tramontana, E., "A novel approach toward X-ray images classifier". In 2015 IEEE Symposium Series on Computational Intelligence, pp. 1635-1641, IEEE, 2015.

[4] B. V. Daga, V. R. Shah, and S. V. Daga, Radiodiagnosis, nuclear medicine, radiotherapy and radiation oncology. Jaypee Brothers Medical Publishers, 2013.

[5] A. Berrington de González et al., "Projected Cancer Risks From Computed Tomographic Scans Performed in the United States in 2007," Arch. Intern. Med., vol. 169, no. 22, p. 2071, Dec. 2009.

[6] Y. Y. Kim, H. J. Shin, M. J. Kim, and M.-J. Lee, "Comparison of effective radiation doses from X-ray, CT, and PET/CT in pediatric patients with neuroblastoma using a dose monitoring program.," Diagn. Interv. Radiol., vol. 22, no. 4, pp. 390–4, 2016.

[7] S. P. Raman, M. Mahesh, R. V. Blasko, and E. K. Fishman, "CT scan parameters and radiation dose: Practical advice for radiologists," J. Am. Coll. Radiol., vol. 10, no. 11, pp. 840–846, 2013.

[8] S. Lee, M. S. Lee, and M. G. Kang, "Poisson–Gaussian Noise Analysis and Estimation for Low-Dose X-ray Images in the NSCT Domain," Sensors, vol. 18, no. 4, p. 1019, Mar. 2018.

[9] A. Neverauskiene et al., "Image based simulation of the low dose computed tomography images suggests 13 mAs 120 kV suitability for non-syndromic craniosynostosis diagnosis without iterative reconstruction algorithms," Eur. J. Radiol., vol. 105, pp. 168–174, Aug. 2018.

[10] Jiang Hsieh, Computed Tomography: Principles, Design, Artifacts, and Recent Advances, Second Edition, 2009.

[11] L. L. Geyer et al., "State of the Art: Iterative CT Reconstruction Techniques," Radiology, vol. 276, no. 2, pp. 339–357, Aug. 2015.

[12] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," J. Mach. Learn. Res., vol. 11, pp. 3371–3408, 2010.

[13] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "A Cascaded Convolutional Neural Network for X-ray Low-dose CT Image Denoising," May 2017.

[14] J. Xie, L. Xu, and E. Chen, "Image Denoising and Inpainting with Deep Neural Networks," pp. 341–349, 2012.

[15] G. Toderici et al., "Full Resolution Image Compression with Recurrent Neural Networks," Aug. 2016.

[16] Q. Ke et al., "A neuro-heuristic approach for recognition of lung diseases from X-ray images," Expert Syst. Appl., vol. 126, pp. 218–232, Jul. 2019.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, S. Member, E. P. Simoncelli, and S. Member, "Image Quality Assessment: From Error Visibility to Structural Similarity," vol. 13, no. 4, pp. 1–14, 2004.

[18] T. Černius, "Image reconstruction of x-ray images using deep learning," Vilnius University, 2018, (In Lithuanian).

[19] K. Xu, X. Liu, H. Cai, and Z. Gao, "Full-reference image quality assessment-based B-mode ultrasound image similarity measure," Jan. 2017.

[20] D. M. Chandler, "Seven Challenges in Image Quality Assessment: Past, Present, and Future Research," ISRN Signal Process., vol. 2013, pp. 1–53, 2013.

[21] A. B. Watson and Bernd, Digital images and human vision. MIT Press, 1993.

[22] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?," in IEEE International Conference on Acoustics Speech and Signal Processing, 2002, p. IV-3313-IV-3316.

[23] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," J. Vis. Commun. Image Represent., vol. 22, no. 4, pp. 297–312, May 2011.

[24] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "DeepSim: Deep similarity for image quality assessment," Neurocomputing, vol. 257, pp. 104–114, Sep. 2017.

[25] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A Haar Wavelet-Based Perceptual Similarity Index for Image Quality Assessment," Jul. 2016.

[26] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," in 2012 19th IEEE International Conference on Image Processing, 2012, pp. 1473–1476.

[27] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index," IEEE Trans. Image Process., vol. 23, no. 2, pp. 684–695, Feb. 2014.

[28] Lin Zhang, Lei Zhang, Xuanqin Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," IEEE Trans. Image Process., vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[29] N. Johnston et al., "Improved Lossy Image Compression with Priming and Spatially Adaptive Bit Rates for Recurrent Networks," Mar. 2017.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015.

[31] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image Reconstruction is a New Frontier of Machine Learning," IEEE Trans. Med. Imaging, vol. 37, no. 6, pp. 1289–1296, Jun. 2018.

[32] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, and W. Samek, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment," IEEE Trans. Image Process., vol. 27, no. 1, pp. 206–219, Jan. 2018.

[33] Wozniak, M., Polap, D., Nowicki, R. K., Napoli, C., Pappalardo, G., and Tramontana, E., "Novel approach toward medical signals classifier". In 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1-7, IEEE, 2015.

[34] V. V. Lukin, N. N. Ponomarenko, O. I. Ieremeiev, K. O. Egiazarian, and J. Astola, "Combining full-reference image visual quality metrics by neural network," 2015, p. 93940K.

[35] "RSNA Bone Age data set from Kaggle." [Online]. Available: https://www.kaggle.com/kmader/rsna-bone-age. [Accessed: 01-Mar-2019].

[36] "NIH Chest X-rays data set from Kaggle." [Online]. Available: https://www.kaggle.com/nih-chest-xrays/data. [Accessed: 01-Mar-2019].

[37] “Stanford University MURA Dataset: Towards Radiologist-Level Abnormality Detection in Musculoskeletal Radiographs.” [Online]. Available: https://stanfordmlgroup.github.io/competitions/mura/. [Accessed: 01-Mar-2019].

[38] Hanying Li, T. Toth, S. McOlash, Jiang Hsieh, and N. Bromberg, “Simulating low dose CT scans by noise addition,” in 2002 IEEE Nuclear Science Symposium Conference Record, 2002, vol. 3, pp. 1832–1834.