Bone Scan Denoising by Self-supervised learning network

Si Young Yie[1,2,3], Seung Kwan Kang[1,4], Donghwi Hwang[1,4], Hong Yun Choi[1] and Jae Sung Lee[1,4,5,6]

[1]Department of Nuclear Medicine, Seoul National University College of Medicine, Seoul, Korea
[2]Interdisciplinary Program in Bioengineering, Seoul National University, Seoul, Korea
[3]Integrated Major in Innovative Medical Science, Seoul National University, Seoul, Korea
[4]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea
[5]Institute of Radiation Medicine, Medical Research Center, Seoul National University College of Medicine, Seoul, Korea
[6]Brightonix Imaging Inc., Seoul, Korea

Introduction

Bone scan is a common clinical practice performed to evaluate skeletal lesions or metastases of cancer. It is widely used because of its low cost, but suffers from the low sensitivity and high noise level of gamma camera. To reduce the noise levels, it may be helpful to increase radiotracer dose or scan time. However, increasing radiotracer dose results in high radiation exposure to the patient. Also, the longer scan time increases the likelihood of patient movement and reduces the throughput of the gamma camera. Vendor-supplied filters, such as half-time filters, are commonly used to reduce noise levels in gamma camera. However, the half-time filters applied to low-count images (<50% of full count) sometimes result in significant blocky artifact. Noise2Noise (N2N) [1] is a self-supervised training model that doesn't require large number of corresponding noisy and clean image pairs to train a denoising network. The aim of this study is to explore the feasibility of self-supervised denoising (N2N) in bone scans by comparing its performance with the conventional Noise2Clean (N2C) approach.

Methods

For training and evaluating the denoising networks, ^{99m}Tc-MDP or DPD bone scan data of 133 patients were acquired (27 males and 106 females, age = 55.3±13.0 years, acquisition time = 854±124 sec, 30 for training and 103 for testing) using a GE Discovery 670 scanner. From the list-mode data, we generated 10 data bins with 10% time of the full scan duration. The matrix and pixel sizes of the images were 1024 × 256 and 2.21 × 2.21 mm^2, respectively.

We used two-dimensional U-Net [2] to train and test denoising models: N2N and N2C. The dimensions of input and output data were 256 × 256. In network training, the input is a 10% time image, the N2N target is an independent 10% time image, and the N2C target is the full-count image. For each training method, we trained a one-channel input (1CH) network and a two-channel input (2CH) network. The 1CH network uses either one of the anterior or posterior side of the bone scan as an input. On the other hand, the 2CH network uses both sides of the bone scan as a single input. The Mean Square Error was used as the loss function to optimize the network. To evaluate the performance of denoising network, we calculated the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) and High Frequency Error Norm (HFEN) with respect to the full-count bone scan image as follows:

$$PSNR(x,y) = 10\log_{10}\left(\frac{MAX_y^2}{MSE_{x,y}}\right) \quad (1)$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \quad (2)$$

$$HFEN(x,y) = \frac{\|LoG(x)-LoG(y)\|_2}{\|LoG(y)\|_2} \quad (3)$$

In equations (1) ~ (3), x and y each represents the tested images and the reference image. In equation (1), MAX is the maximum value of the reference image, and MSE is the mean square error between the reference and tested images. In equation (2), μ and σ are mean and variance or covariance of the compared images. C_1 and C_2 are stability terms which we use $C_1 = (0.01MAX_y)^2$ and $C_2 = (0.03MAX_y)^2$ as proposed in [3]. In equation (3), LoG is the Laplacian of Gaussian filter with 7×7 kernel size and 5mm standard deviation.

Results

Figures 1 and 2 show the input images and denoised images using N2N, N2C filters. PSNR, SSIM and HFEN of noisy input and denoised outputs with respect to full-count bone scan image are summarized in Table 1.

The N2N denoising network showed equivalent performance to N2C network. PSNR difference between N2N and N2C denoising methods was within 0.2dB. SSIM difference between the denoising methods was within 0.03, and HFEN difference was within 0.02. Overall, the 1CH network outperformed the 2CH network with regard to HFEN.

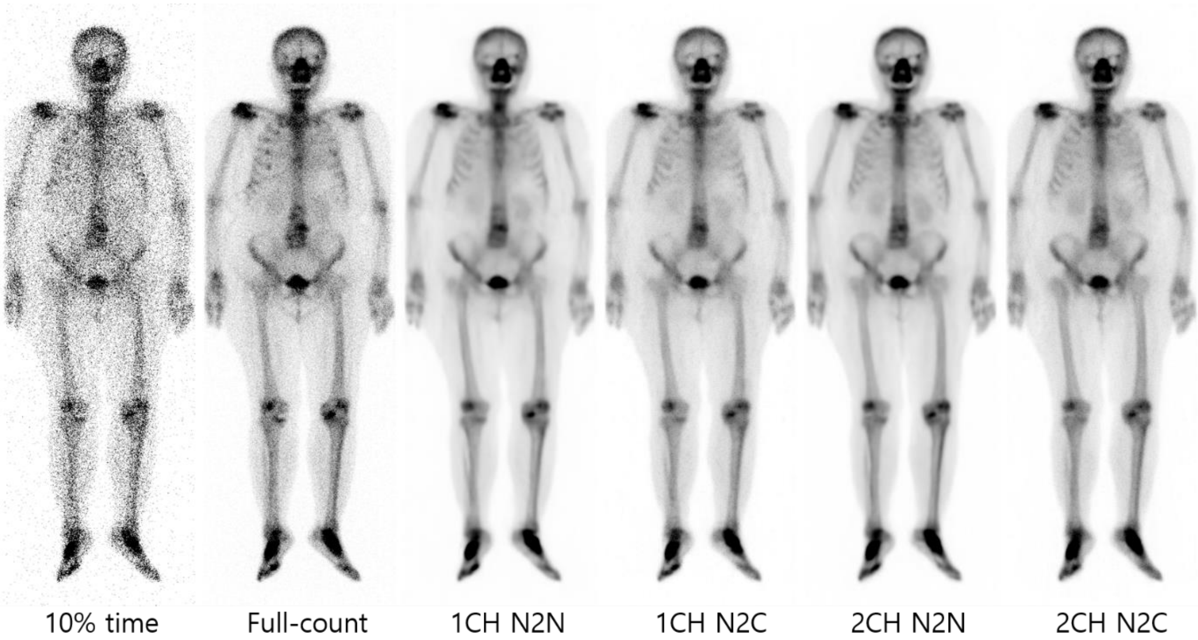

Figure 1. Input image of a 58-years old man with 10% scan time and denoised images using N2N, N2C filters

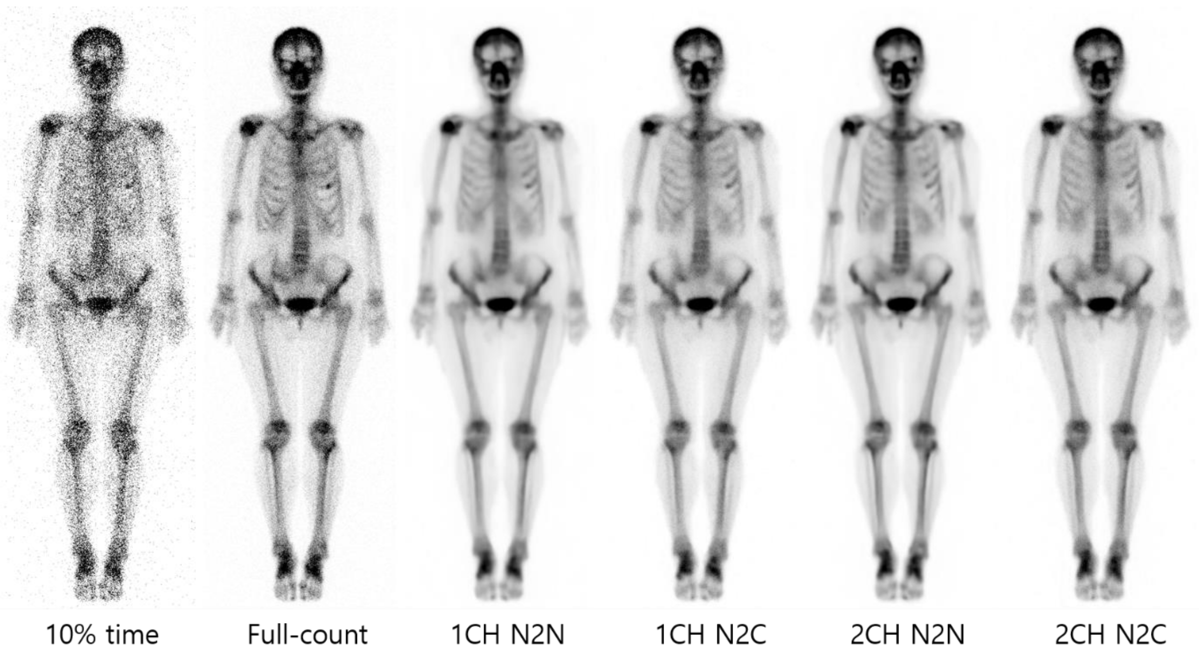

Figure 2. Input image of an 84-years old woman with 10% scan time and denoised images using N2N, N2C filters

Table 1. PSNR, SSIM and HFEN mean of noisy input and denoised outputs with respect to the full-count bone scan image

Metric	Input	1CH N2N	1CH N2C	2CH N2N	2CH N2C
PSNR (dB)	30.5	38.1	38.2	38.4	38.5
SSIM	0.753	0.893	0.915	0.905	0.916
HFEN	0.796	0.447	0.446	0.455	0.466

Conclusion

Self-supervised denoising methods were useful for reducing the noise in the low-count bone scan images. N2N method which requires only paired low-count data for network learning showed remarkable denoising performance. The self-supervised denoising method will be useful for reducing bone scan time or radiation dose.

References

[1] Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M, et al. Noise2Noise: learning image restoration without clean data. arXiv. 2018;1803.04189.

[2] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Pro Med Image Comput Comput Assist Interv. 2015;234-41.

[3] Wang Z, Bovik A, Sheikh H, Simoncelli E. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Trans Image Process. 2004; 600-12.