Journal of Nuclear Medicine, published on July 19, 2019 as doi:10.2967/jnumed.119.226613

Denoising of Scintillation Camera Images using a Deep Convolutional Neural Network: A Monte Carlo Simulation Approach

David Minarik[1], Olof Enqvist[2, 3], Elin Trägårdh[4, 5]

[1]Radiation Physics, Skåne University Hospital, Malmö, Sweden, [2]Eigenvision AB, Malmö, Sweden, [3]Department of Electrical Engineering, Chalmers University of Technology, Göteborg, Sweden, [4]Clinical Physiology and Nuclear Medicine, Skåne University Hospital, Malmö, Sweden, [5]Wallenberg Centre for Molecular Imaging, Lund University, Sweden.

Corresponding author: David Minarik, Radiation Physics, Skåne University Hospital, Carl Bertil Laurells gata 9, 205 02 Malmö, Sweden. David.minarik@med.lu.se

Financial support: The work was made possible by research grants from the Knut and Alice Wallenberg Foundation, the Swedish Federal Government under ALF agreement, Lund University and Region Skåne.

Word count: 5000

Short running title: Denoising nuclear medicine images

ABSTRACT

Rationale: Scintillation camera images contain a large amount of Poisson noise. We have investigated whether noise can be removed in whole-body bone scans using convolutional neural networks (CNN) trained with sets of noisy and noiseless images obtained by Monte Carlo simulation.

Methods: Three CNNs were generated using three different sets of training images: simulated bone scan images, images of a cylindrical phantom with hot and cold spots, and a mix of the first two. Each training set consisted of 40,000 noiseless and noisy image pairs. The CNNs were evaluated with simulated images of a cylindrical phantom and simulated bone scan images. The mean squared error (MSE) between filtered and true images was used as difference metric and the coefficient of variation (COV) was used to estimate noise reduction. The CNNs were compared to Gaussian and median filters. A clinical evaluation was performed where the ability to detect metastases for CNN- and Gaussian-filtered bone scans with half the number of counts were compared with standard bone scans.

Results: The best CNN reduced COV with on average 92%, and the best standard filter reduced COV with 88%. The best CNN gave an MSE that was on average 68% and 20% better than the best standard filters, for the cylindrical and bone scan images, respectively. The best CNN for the cylindrical phantom and bone scans were the dedicated CNNs. No significant differences for the ability to detect metastases between standard, CNN- and Gaussian-filtered bone scans were found.

Conclusion: Noise can be removed efficiently regardless of noise level with little or no resolution loss. The CNN filter enables reducing the scanning time by half and still obtain good accuracy for bone metastases assessment.

Keywords: nuclear medicine; artificial intelligence; machine learning; image enhancement; Monte Carlo

INTRODUCTION

Scintillation camera images are inherently noisy due to the specifics of the imaging process. Several filtering methods to remove noise exist, which range from simple convolution with small filter kernels to more complex filtering using wavelets or statistical methods (*1-4*). However, the trade-off for most types of denoising filters is resolution loss.

Convolutional neural networks (CNN), a machine learning algorithm, have been shown to work well in denoising photographic images. These types of images generally contain only a small amount of noise, making it easy to generate sets of noisy and pristine photographic images to train a CNN. Scintillation camera images, however, suffer from a larger amount of Poisson noise. Thus, it is more challenging to acquire sets of noisy and pristine image sets for training purposes.

Machine learning is believed to change radiology and nuclear medicine in the future (*5*). Recent publications using machine learning algorithms for classification and segmentation purposes exist (*6-10*). CNNs have been used to obtain standard-dose CT and PET images from low-dose data (*11,12*) and for image enhancement by determining scatter correction parameters (*13*) and CNN augmented emission based attenuation correction (*14*) in positron emission tomography (PET). Recently, Gong et al, used computer simulated PET images to pre-train a denoising CNN, and then fine-tuned the CNN with patient data (*15*). The same group have also implemented a CNN in the reconstruction process of PET data (*16*).

This study investigates whether noise can be removed in scintillation camera images using a CNN that has been trained with sets of noisy and noiseless images obtained by Monte Carlo simulation and whether the types of training images affect the results. If a CNN is to be used for any type of medical image enhancement, it is vital that no true information is removed or that false information is added to the images. The CNN should only recreate information that has been lost in the imaging process. The question is whether a single CNN can be trained and used on multiple types of images, or whether specialized CNNs that are trained and used on only one type of images (e.g., bone scans) are needed.

The aim of this study was to generate different CNNs by utilizing different sets of training images and to evaluate the performance of the CNNs on simulated images of cylindrical phantom and simulated whole-body bone scan images. We have compared the CNN-filtered images with images filtered with Gaussian and median filters. As a proof of concept, we performed a pilot clinical evaluation where bone scans with half the number of counts were filtered using the CNN and Gaussian filter and then compared with standard bone scans for the diagnosis of bone metastases.

MATERIALS AND METHODS

Training Images

Noiseless whole-body bone scan images were generated with the SIMIND Monte Carlo program and the XCAT anthropomorphic phantom (*17,18*). Three different phantoms were used, and 60 different simulations were generated with random numbers and sizes of bone metastases ranging from one to more than 50. The method used to create and simulate phantoms with different tumor burden is described in (*8*). Both anterior and posterior views were simulated, and the simulations included all physically degenerative effects such as attenuation and scatter in the phantom, scatter in and penetration of the collimator as well as depth dependent resolution. The SIMIND program was set up to mimic a Siemens Symbia gamma camera using a 256x1024 matrix with a pixel size of 2.21 mm and a 15 % energy window centered over the 140-keV peak. Next, 10,000 noiseless training images were created by randomly extracting a 256x256 patch from either the anterior view or the posterior view of a simulated image, applying a random shearing operation on the patch (in order to generate images with various body shapes and sizes), and multiplying the patch with a random number so that the total number of counts in the corresponding whole body image would range from 0.5 to 3 million counts (bone scan guidelines recommend at least 1.5 million counts (*19*)). The patches were then downsized to a matrix size of 128x128.

A second training set was created with the Simind Monte Carlo program using a simple cylindrical phantom with a homogeneous activity distribution. Three phantoms with a length of 30 cm, a long axis of 15 cm and short axes of 5, 10, or 15 cm were simulated with the same camera setup as the bone simulation. 15 projections evenly sampled in an arc of 0 to 90 degrees, with the starting angle parallel to the short axis, were simulated for each phantom to create 45 different images in total. For each cylinder, three different simulations with small ellipsoids were performed in the same way with the ellipsoid in the middle of the cylinder.

Next, 10,000 noiseless training images were created by randomly selecting a projection from one of the three simulations. One or more hotspots or cold spots were then added at random places in the image of the cylinder by randomly choosing an image from the ellipsoid simulations, multiplying it with a random number, translating it in a random manner, and finally adding or subtracting it from the cylinder image. The cylinder images were then multiplied with a random number to mimic a range of intensity levels that is normally seen in nuclear medicine images. Finally, a random affine transformation was applied, which included translation, rotation, shearing, and scaling. Examples of the sets of noiseless and noisy images are displayed in Fig. 1.

CNN

Denoising CNN (*20*) has shown good results in denoising ordinary color photographs, and hence, the same network structure was used in this work. This type of network has three parts. The first layer is a convolutional layer with 64 3x3 filters and a rectified linear unit activation. There are also 19 convolutional layers, which each have 64 filters with a size of 3x3, batch normalization, and rectified linear unit activation. Finally, a convolutional layer with a single 3x3 filter produces the output image.

The CNNs were trained using a quadratic loss function in MatLab^{tm} (Mathworks, Natick, Massachusetts, USA). Three different CNNs were evaluated: one where only bone scan images were used for training (further called B-CNN), a second using only cylinder images (C-CNN), and a third with a mix of images (M-CNN). Each CNN was trained using 10,000 different images. Four noisy patches per noiseless image were used for training, leading to a total of 40,000 training images. Random Poisson noise was added to the images corresponding to the intensity level of the images. Training was performed on an nVidia GeForce RTX 2080 TI GPU.

Evaluation

A simulation of a cylindrical phantom was used, which consists of a 15 cm high cylinder with a radius of 15 cm, with the camera placed perpendicular to the height direction of the cylinder. Several hot spheres with radii of 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.2, and 1.4 cm and four cold spheres with radii of 0.5, 1.0, 1.5, and 2.0 cm were placed in the middle of the cylinder in the height direction (Fig. 2). The activity ratio of hot spot to background was 10:1, which yields a maximum contrast of around 2.4 for the large hot spot, when accounting for attenuation and overlapping background activity. Ten images with a total number of counts ranging from 100,000 to 1 million were created, and 40 noise realizations per image were generated. All the noisy images were filtered with the different CNNs and with 6 different Gaussian filters with full width at half maximum (FWHM) of 3, 5, 7, 9, 11, and 13 mm, respectively. Four different median filters were also used with quadratic kernels of 9, 25, 49 and 81 pixels.

For each noisy and filtered image, the coefficient of variation (COV) was calculated for pixel values in a circular region-of-interest (109 pixels) placed in the homogeneous part of the middle of the phantom. The mean squared error (MSE) (eqn. 1) was calculated for each filtered and noisy image with the noiseless image as reference.

$$MSE = \frac{1}{NM}\sum_1^N \sum_1^M \left(I_{noisless}(m,n) - I_{noisy/filtered}(m,n)\right)^2 \qquad (1)$$

The CNNs were compared using 40 simulated bone scans that were not part of the training data. For each of the 40 bone scans, 10 images were created, which had different noise levels ranging from a total number of counts in the posterior view of 0.1 million to 1 million. All images were filtered with the CNNs, the Gaussian and median filters. The MSE was calculated for each filtered and noisy image with the noiseless image as reference. An example is shown in Fig. 3.

Clinical Evaluation

In a pilot clinical study, we compared M-CNN filtered and Gaussian filtered (FWHM 7mm) half time imaging whole-body bone scans with standard scans. Images acquired with half the acquisition time were generated by binominal subsampling (*21*). Bone scans from 39 patients (3 women, 36 men) clinically referred to Skåne University Hospital, Malmö, Sweden for assessment of bone metastases, were evaluated. The median age was 76 years (range 52-92). Patients were injected with 600 MBq ^{99m}Tc-hydroxydiphosphonate. Accumulation time was 2-4 hours. All patients were scanned on a Siemens Symbia gamma camera with a low energy high resolution collimator; scan time of 15 cm/min and a 256x1024 matrix with a pixel size of 2.21 mm. One nuclear medicine physician (Observer A) and one resident in nuclear medicine (Observer B) interpreted the images in a random and blinded fashion and assessed if bone metastases were present or not. They were given only one set of images (anterior and posterior views) at a time and were not aware of their interpretation of the other image sets. After two months, Observer A reinterpreted the images. Presence of bone metastases were assessed separately for anterior and posterior views. The standard bone scans were considered reference method. Sensitivity and specificity for detecting bone metastases for the CNN- and the Gaussian-filtered images were calculated, as well as the area under the receiver operating characteristic (ROC) curve. Differences between the standard and the CNN- and the Gaussian-filtered images were assessed with McNemar using IBM SPSS version 25 (IBM, Armonk, NY, USA).

Examples of bone scans and filtered equivalents are shown in Fig. 4.

The institutional review board at Lund University (#2019-00644) approved this retrospective study and the requirement to obtain informed consent was waived.

RESULTS

The calculation results of the COV are displayed in Table 1, and the results of the MSE evaluation are presented in Tables 2 and 3. The best CNN reduced the COV with on average 92 %, whilst the best standard filter (mean filter with 81 pixels) reduced the COV with 88 %. The best CNN gave an MSE that was on average 68 % and 20 % better than the best standard filters (mean filter with 9 pixels and Gaussian filter with a FWHM of 7 mm), for the cylindrical and bone scan images respectively. The results showed that there is a small difference between images denoised with the different CNNs. The noise reduction in the specific case shown in Table 1 is more than tenfold for the C-CNN and slightly less for the M-CNN. The C-CNN gave the lowest MSE in the evaluation with the cylindrical phantom. For the bone scan evaluation, the best results were obtained for the B-CNN. The CNN created with a mix of images was similar to the best CNN for both cylindrical phantom and bone scans. Among the conventional filters, a median filter with a 3x3 neighborhood and a Gaussian filter with a FWHM of 7 mm produced the best MSE results for the cylinder and bone scans, respectively.

Clinical Evaluation

The sensitivities, specificities and areas under the ROC curve are found in Table 4. There were no significant differences between standard and CNN-filtered images (p=0.99 and p=1.0

for Observer A and p=0.25 for Observer B), nor between standard and Gaussian-filtered images (p=0.45 and p=0.69 for Observer A and p=1.0 for Observer B).

DISCUSSION

We have trained and applied CNNs on simulated bone scans and images of cylindrical phantoms. We have also used the CNNs to denoise real bone scans to verify the feasibility of using CNNs trained on Monte Carlo images to remove noise and performed a pilot clinical evaluation. We have shown that it is possible to remove noise with an almost total reduction of noise with very little or no resolution loss. The C-CNN even outperformed a median filter with a 9x9 neighborhood while still maintaining the resolution.

Other statistical filtering methods and other types of methods require some form of optimization of input parameters, which may not be valid for all intensity ranges encountered, such as those in bone scans. In contrast, a trained bone scan CNN can handle all types of scanning situations if the noise distribution in the training images follows that of real images and the range of intensities in the images reflects the whole spectra of observed intensities. This is demonstrated in Fig. 4, which shows original bone scans, mathematically generated images representing the same image acquired with 75%, 50%, 25% and 10% of the imaging time, and the corresponding CNN filtered and Gaussian filtered images. There are some small structural differences between the CNN filtered images, but the level of noise is almost the same, which is not case for the Gaussian filtered images. In the 10% CNN-filtered image, hot spots in the upper spine appear, that are not shown in the images with higher count rates. This indicate that using the CNN-filter on images with only 10% of the counts might not be optimal, but this needs to be established in future clinical studies.

It seems that the type of images used to train the CNN matters to some extent. The differences are most clearly seen in Fig. 2, where the images filtered with the bone CNN are mottled. However, it seems that if a variety of images are used (e.g., a mix of different types of images), the result can be almost as good as if the CNN were trained using a specific set of images.

As with other filtering methods, it is necessary to establish whether the filtered images provide any benefits or improvements regarding the confidence about what the reading physicians see in the images. It is also necessary to establish that no false information is added to the images. In our pilot clinical evaluation, we showed that it is possible to reduce the scanning time by half and then apply the CNN filter and still obtain very high accuracy for assessment of bone metastases. It should be noted that the observers were used to interpreting original bone-scan images, not CNN- or Gaussian filtered images. Since filtered images look different than original bone scans, they are not expected to increase reader confidence at this early stage. Whether CNN-filtered images can provide better accuracy in the detection of metastases compared with standard bone scans and if the acquisition time can be further reduced need to be evaluated in larger clinical studies, with a better ground truth than the standard bone scan.

It is crucial to have a good model for generating data used for training the CNN. Any error may lead to bias in the clinical images. Our study shows that it is possible to use Monte Carlo simulated images for training the CNNs, if the training data is carefully generated. An alternative could be to use real bone scan images with very high statistics as a substitute for noiseless images. However, in order to receive reasonably noiseless images, the imaging time needs to be very long (more than 1 hour) and few patients can lay completely still for such a long time. Therefore, we propose using Monte Carlo simulated images instead.

CONCLUSION

Noise can be removed very efficiently by using a CNN trained with noisy and noiseless simulated images, regardless of the noise level and with very little or no resolution loss. The CNN filter makes it possible to reduce the scanning time by half and still obtain good accuracy for detecting bone metastases, but this needs to be confirmed in large clinical studies.

DISCLOSURES

The authors declare that they have no competing interests. This work was made possible by research grants from the Knut and Alice Wallenberg Foundation, the Swedish Federal Government under ALF agreement, and from Region Skåne. We thank Anders Persson for interpretation of images.

KEY POINTS

QUESTION: Can a noise reducing convolutional neural network be trained with Monte Carlo simulated gamma camera images.

PERTINENT FINDINGS: The convolutional neural networks trained with Monte Carlo simulated images were able to reduce noise by a factor of ten while still maintaining the resolution.

IMPLICATIONS FOR PATIENT CARE: The results indicate that noise in planar nuclear medicine images can be removed very efficiently with very little or no resolution loss, thus enhancing the image quality and enabling shorter scanning times with preserved accuracy for detecting bone metastases.

REFERENCES

1. Wesolowski CA, Yahil A, Puetter RC, Babyn PS, Gilday DL, Khan MZ. Improved lesion detection from spatially adaptive, minimally complex, Pixon reconstruction of planar scintigraphic images. *Comput Med Imaging Graph.* 2005;29:65-81.

2. Gonzalez RC, Woods RE. *Digital image processing*. New York, NY: Pearson; 2018:161-252.

3. Dutta J, Leahy RM, Li Q. Non-local means denoising of dynamic PET images. *PLoS One.* 2013;8:e81390.

4. Christian BT, Vandehey NT, Floberg JM, Mistretta CA. Dynamic PET denoising with HYPR processing. *J Nucl Med.* 2010;51:1147-1154.

5. Forsting M. Machine learning will change medicine. *J Nucl Med.* 2017;58:357-358.

6. Wang H, Zhou Z, Li Y, et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from (18)F-FDG PET/CT images. *EJNMMI Res.* 2017;7:11.

7. Perk T, Bradshaw T, Chen S, et al. Automated classification of benign and malignant lesions in (18)F-NaF PET/CT images using machine learning. *Phys Med Biol.* 2018;63:225019.

8. Anand A, Morris MJ, Kaboteh R, et al. Analytic validation of the automated Bone Scan Index as an imaging biomarker to standardize quantitative changes in bone scans of patients with metastatic prostate cancer. *J Nucl Med.* 2016;57:41-45.

9. Anand A, Morris MJ, Kaboteh R, et al. A preanalytic validation study of automated Bone Scan Index: Effect on accuracy and reproducibility due to the procedural variabilities in bone scan image acquisition. *J Nucl Med.* 2016;57:1865-1871.

10. Betancur J, Rubeaux M, Fuchs TA, et al. Automatic valve plane localization in myocardial perfusion SPECT/CT by machine learning: Anatomic and clinical validation. *J Nucl Med.* 2017;58:961-967.

11. Xiang L, Qiao Y, Nie D, An L, Wang Q, Shen D. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. *Neurocomputing.* 2017;267:406-416.

12. Chen H, Zhang Y, Zhang W, et al. Low-dose CT via convolutional neural network. *Biomed Opt Express.* 2017;8:679-694.

13. Gao F, Shah V, Sibille L, Zuehlsdorff S. An AI system to determine reconstruction parameters and improve PET image quality. *J Nucl Med.* 2018;59:31.

14. Hwang D, Kim KY, Kang SK, et al. Improving the accuracy of simultaneously reconstructed activity and attenuation maps ssing deep learning. *J Nucl Med.* 2018;59:1624-1629.

15. Gong K, Guan J, Liu C, Qi J. PET Image denoising using a deep neural network through fine tuning. *IEEE Trans Radiat Plasma Med Sci.* 2018:1-1.

16. Gong K, Guan J, Kim K, et al. Iterative PET image reconstruction using convolutional neural network representation. *IEEE Trans Med Imaging.* 2018; Sept 12.

17. Ljungberg M, Strand SE. A Monte Carlo program for the simulation of scintillation camera characteristics. *ComputMethods Programs Biomed.* 1989;29:257-272.

18. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging research. *Med Phys.* 2010;37:4902-4915.

19. Van den Wyngaert T, Strobel K, Kampen WU, et al. The EANM practice guidelines for bone scintigraphy. *Eur J Nucl Med Mol Imaging.* 2016;43:1723-1738.

20. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans Image Process.* 2017;26:3142-3155.

21. Wiuf C, Stumpf MPH. Binomial Subsampling. *Proceedings: Mathematical, Physical and Engineering Sciences.* 2006;462:1181-1195.

Figures

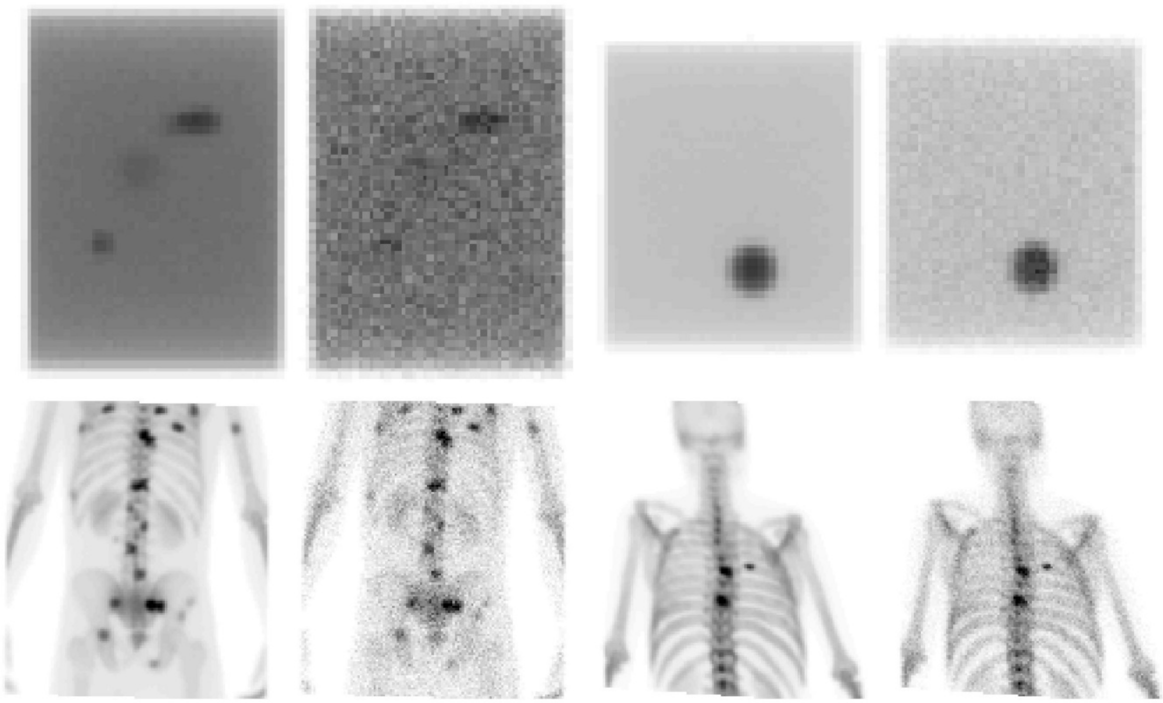

FIGURE 1. Top row: two examples from the training batch with cylindrical phantoms. Bottom row: two examples from the training batch with bone scans. Images to the left are noiseless and images to the right are noisy.

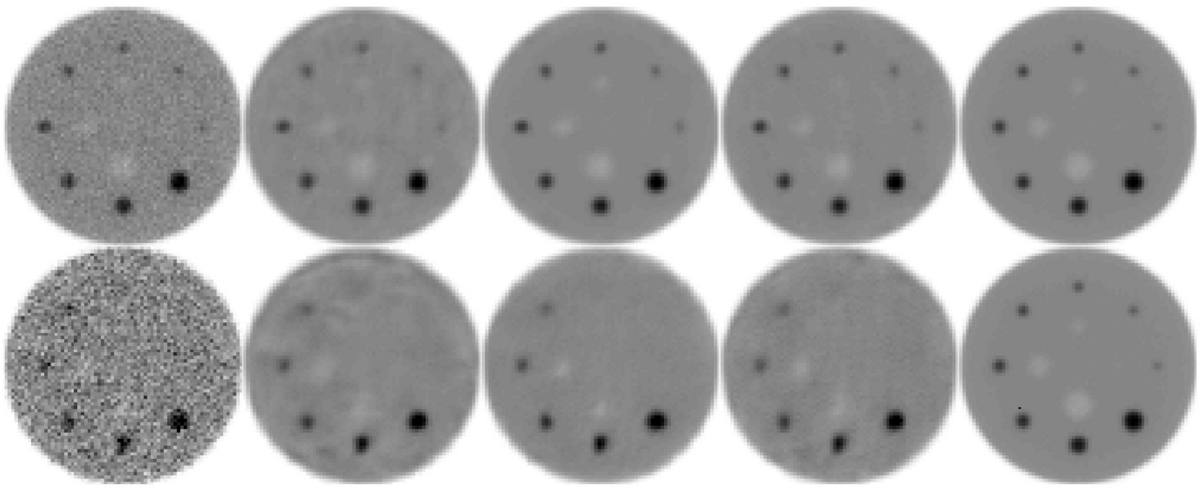

FIGURE 2. Simulated cylindrical phantom. Images on the top row have 1 million counts, and those on the lower row have 100,000 counts. From left to right: noisy image; images filtered with B-CNN, C-CNN, and M-CNN; and the true image.

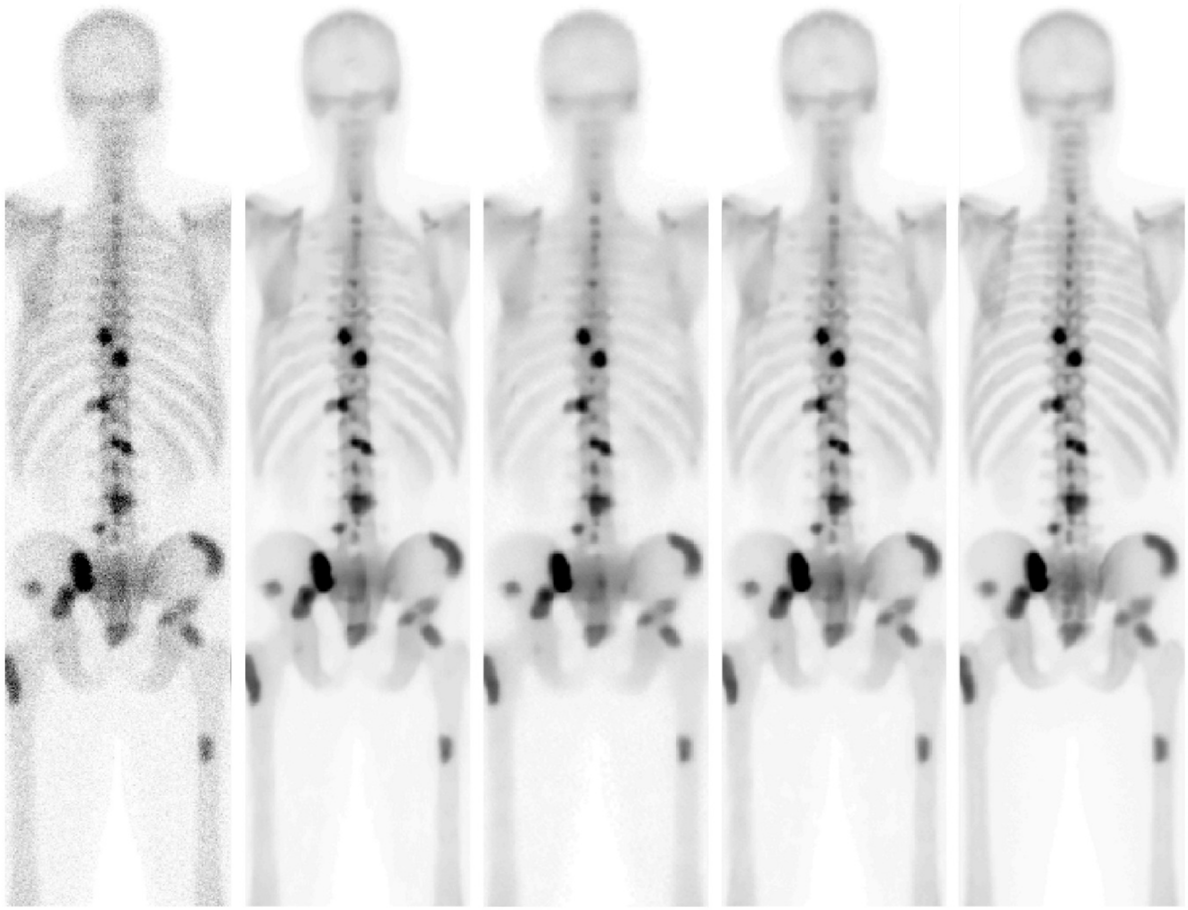

FIGURE 3. Posterior view of a simulation with a total of 0.6 million counts. From left to right: noisy image; noisy image filtered with B-CNN, C-CNN, and M-CNN; and true image

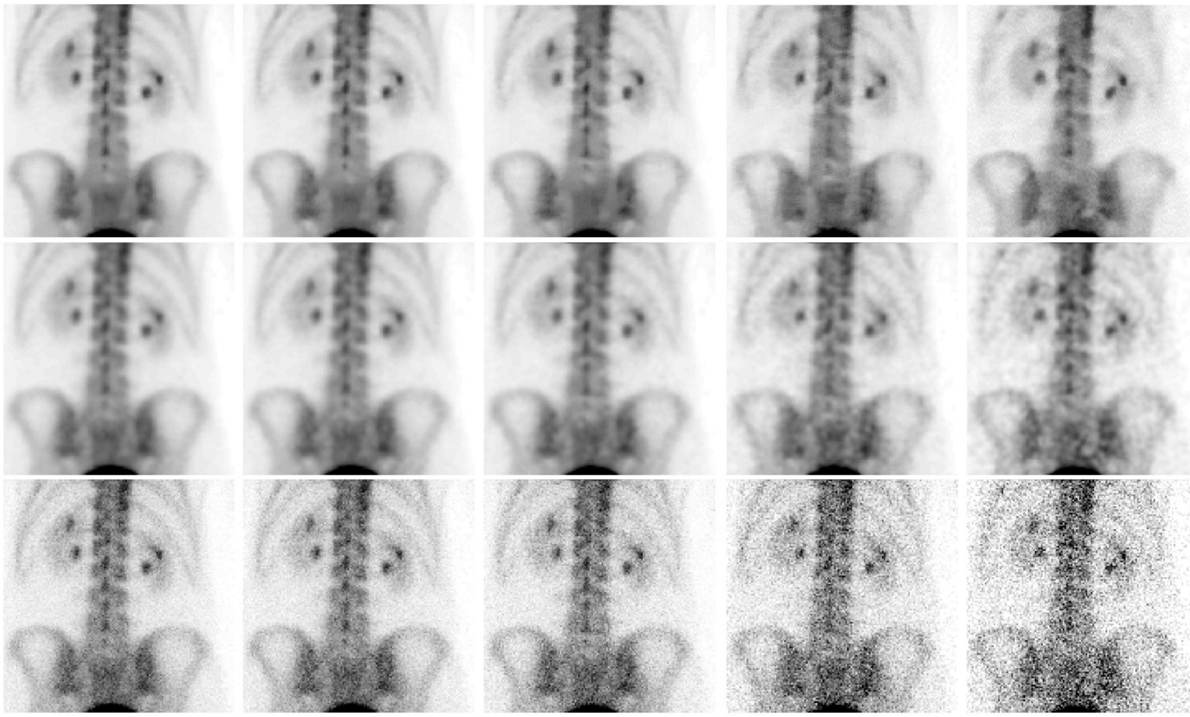

FIGURE 4. Posterior views of a bone scan. From top to bottom: images filtered with M-CNN, images filtered with 7 mm FWHM Gaussian filter and unfiltered images. From left to right: original images, images with 75%, 50%, 25% and 10% of the original counts.

Tables

TABLE 1. Mean COV in % of 40 noise realizations

	Total counts in millions																			
	1		**0.9**		**0.8**		**0.7**		**0.6**		**0.5**		**0.4**		**0.3**		**0.2**		**0.1**	
Image type	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Orig	6.41	0.43	6.80	0.40	7.08	0.44	7.50	0.55	8.25	0.53	8.91	0.59	9.88	0.61	11.67	0.77	14.33	1.04	10.17	1.64
B-CNN	1.71	0.30	1.85	0.26	1.77	0.30	1.94	0.31	2.05	0.34	2.23	0.43	2.33	0.45	2.51	0.42	2,78	0.60	3.01	0.94
M-CNN	0.98	0.16	1.02	0.16	1.00	0.17	1.02	0.12	1.09	0.15	1.16	0.17	1.24	0.15	1.36	0.14	1.50	0.17	2.60	0.69
C-CNN	0.46	0.11	0.48	0.09	0.53	0.20	0.53	0.08	0.56	0.07	0.60	0.11	0.64	0.07	0.78	0.10	0.95	0.18	1.56	0.27
Gaus3	6.29	0.42	6.67	0.40	6.95	0.44	7.38	0.54	8.09	0.52	8.76	0.59	9.69	0.60	11.44	0.75	14.04	1.02	19.78	1.64
Gaus5	4.00	0.31	4.28	0.28	4.40	0.31	4.69	0.34	5.10	0.36	5.59	0.41	6.15	0.45	7.24	0.46	8.95	0.66	12.64	1.17
Gaus7	2.50	0.29	2.73	0.26	2.74	0.29	2.95	0.27	3.19	0.33	3.52	0.41	3.86	0.44	4.50	0.39	5.62	0.52	8.07	1.09
Gaus9	1.88	0.28	2.08	0.26	2.05	0.28	2.24	0.26	2.42	0.32	2.66	0.42	2.94	0.42	3.41	0.37	4.26	0.51	6.17	1.06
Gaus11	1.49	0.26	1.65	0.26	1.61	0.27	1.79	0.25	1.92	0.29	2.11	0.40	2.33	0.37	2.72	0.34	3.39	0.50	4.92	0.98
Gaus13	1.24	0.24	1.39	0.25	1.34	0.26	1.50	0.24	1.61	0.28	1.75	0.38	1.95	0.34	2.30	0.33	2.84	0.50	4.12	0.91
Med9	2.55	0.36	2.73	0.39	2.77	0.35	2.98	0.35	3.17	0.46	3.47	0.48	3.98	0.47	4.61	0.58	5.62	0.79	8.26	1.29
Med25	1.48	0.28	1.62	0.34	1.61	0.30	1.71	0.32	1.82	0.36	2.00	0.40	2.32	0.44	2.72	0.46	3.29	0.61	4.80	0.79
Med49	0.99	0.20	1.08	0.25	1.12	0.26	1.13	0.25	1.18	0.30	1.29	0.29	1.58	0.36	1.82	0.43	2.25	0.49	3.28	0.72
Med81	0.77	0.19	0.79	0.20	0.89	0.23	0.83	0.20	0.92	0.24	0.96	0.24	1.22	0.32	1.36	0.37	1.71	0.39	2.49	0.52

TABLE 2. MSE between noisy and CNN-filtered cylidrical phantom images with noiseless image as reference. The values represent the mean MSE of 40 noise realizations.

	Total counts in millions																			
	1		**0.9**		**0.8**		**0.7**		**0.6**		**0.5**		**0.4**		**0.3**		**0.2**		**0.1**	
Image type	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Orig	61.57	1.64	68.01	1.46	76.85	1.92	87.17	1.74	102.04	1.70	121.82	2.52	153.81	3.03	204.72	4.98	305.88	8.41	610.56	14.50
B-CNN	11.10	0.51	11.65	0.49	12.45	0.66	13.17	0.49	14.42	0.61	16.01	0.97	18.11	0.94	20.86	1.38	25.61	1.50	39.68	2.47
M-CNN	6.57	0.34	6.95	0.39	7.54	0.43	8.16	0.48	9.06	0.52	10.39	0.83	12.43	0.93	15.11	0.83	20.71	1.25	36.36	2.17
C-CNN	5.84	0.35	6.22	0.44	6.81	0.46	7.53	0.47	8.36	0.51	9.62	0.78	11.55	0.86	14.25	0.71	20.40	1.12	40.08	2.78
Gaus3	59.42	1.58	65.64	1.41	74.18	1.86	84.14	1.68	98.50	1.64	117.58	2.44	148.46	2.92	197.58	4.79	295.22	8.14	589.30	13.96
Gaus5	27.19	0.84	29.79	0.77	33.34	0.91	37.37	0.76	43.42	0.92	51.09	1.38	63.82	1.50	83.86	2.16	123.75	3.76	245.17	6.14
Gaus7	22.12	0.70	23.12	0.73	24.66	0.72	26.27	0.68	28.92	0.85	31.84	1.15	37.26	1.31	45.31	1.64	61.48	2.10	112.09	3.99
Gaus9	31.19	0.78	31.67	0.81	32.67	0.78	33.63	0.83	35.32	0.93	36.90	1.15	40.30	1.38	45.02	1.66	54.58	1.74	85.06	3.35
Gaus11	48.87	0.91	49.06	0.92	49.81	0.90	50.46	1.03	51.68	1.06	52.57	1.20	54.93	1.51	58.01	1.79	64.29	1.71	84.41	3.07
Gaus13	69.29	1.01	69.33	1.01	69.95	0.98	70.45	1.16	71.41	1.19	71.94	1.25	73.75	1.62	75.98	1.92	80.60	1.78	95.22	3.01
Med9	17.58	0.86	19.08	0.91	20.92	0.86	23.24	0.99	26.20	0.98	29.97	1.40	36.50	1.55	46.13	2.26	65.22	2.52	123.37	5.07
Med25	32.22	1.55	33.15	1.50	34.05	1.48	35.23	1.48	36.43	2.19	38.38	1.66	41.37	2.14	46.94	2.87	56.36	2.98	88.63	4.29
Med49	72.97	2.43	73.62	1.92	73.81	2.45	75.03	1.94	74.95	2.76	76.09	2.85	78.27	3.15	82.50	3.58	89.85	4.23	116.64	5.18
Med81	117.12	2.68	117.55	2.44	117.19	2.55	117.62	2.42	118.11	2.71	118.92	2.68	120.67	3.12	124.07	3.51	131.24	3.51	159.07	5.13

TABLE 3. MSE between noisy and CNN-filtered simulated bone scan images with noiseless image as reference. The values represent the mean MSE of 40 simulared bone scans.

	Total counts in millions																			
	1		**0.9**		**0.8**		**0.7**		**0.6**		**0.5**		**0.4**		**0.3**		**0.2**		**0.1**	
Image type	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Orig	3.73	0.10	4.15	0.11	4.67	0.13	5.34	0.15	6.24	0.17	7.50	0.19	9.42	0.25	12.73	0.31	19.15	0.50	38.43	2.34
B-CNN	0.25	0.02	0.27	0.03	0.30	0.03	0.33	0.03	0.38	0.03	0.44	0.03	0.55	0.03	0.78	0.06	1.33	0.14	3.71	0.45
M-CNN	0.27	0.02	0.29	0.02	0.32	0.02	0.36	0.03	0.41	0.03	0.48	0.03	0.59	0.03	0.82	0.06	1.38	0.14	3.86	0.38
C-CNN	0.35	0.02	0.40	0.02	0.46	0.02	0.54	0.01	0.66	0.02	0.83	0.03	1.11	0.03	1.65	0.06	2.96	0.14	8.04	0.63
Gaus3	1.04	0.03	1.15	0.03	1.29	0.04	1.48	0.04	1.73	0.05	2.08	0.05	2.62	0.07	3.58	0.09	5.54	0.16	12.09	0.71
Gaus5	0.42	0.03	0.45	0.03	0.50	0.03	0.56	0.03	0.65	0.03	0.76	0.03	0.95	0.03	1.31	0.04	2.11	0.13	5.28	0.36
Gaus7	0.38	0.06	0.40	0.06	0.43	0.06	0.46	0.06	0.51	0.06	0.57	0.06	0.67	0.06	0.88	0.06	1.39	0.12	3.65	0.30
Gaus9	0.53	0.13	0.54	0.12	0.56	0.12	0.58	0.13	0.61	0.13	0.65	0.13	0.71	0.12	0.86	0.12	1.23	0.12	3.04	0.30
Gaus11	0.79	0.22	0.80	0.21	0.81	0.21	0.83	0.22	0.85	0.22	0.87	0.22	0.92	0.21	1.03	0.21	1.33	0.18	2.87	0.34
Gaus13	1.14	0.33	1.15	0.32	1.16	0.32	1.17	0.33	1.18	0.33	1.20	0.33	1.23	0.32	1.33	0.32	1.57	0.27	2.92	0.42
Med9	0.78	0.04	0.86	0.04	0.96	0.04	1.10	0.04	1.28	0.04	1.55	0.05	1.97	0.06	2.73	0.08	4.45	0.15	10.64	0.57
Med25	0.59	0.09	0.63	0.09	0.68	0.09	0.76	0.09	0.85	0.09	1.00	0.10	1.22	0.10	1.66	0.11	2.71	0.11	6.75	0.29
Med49	0.89	0.22	0.92	0.22	0.97	0.22	1.03	0.23	1.11	0.23	1.24	0.24	1.42	0.25	1.79	0.26	2.69	0.19	6.39	0.36
Med81	1.50	0.45	1.54	0.46	1.59	0.45	1.65	0.46	1.73	0.46	1.85	0.47	2.03	0.47	2.41	0.50	3.31	0.41	7.10	0.56

TABLE 4. Performance of filtered images, using standard bone scans as reference method. Observer A interpreted the images twice, two months apart.

	Observer A		Observer B	
	CNN filter	**Gaussian filter**	**CNN filter**	**Gaussian filter**
Sensitivity	97.8%; 95.0%	95.6%; 95,0%	100%	100%
Specificity	97.0%; 94.7%	84.8%; 89.5%	93.6%	100%
Area under the ROC curve	0.97; 0.95	0.90; 0.92	0.97	1.0