

ORIGINAL RESEARCH

Open Access



# Ultra high speed SPECT bone imaging enabled by a deep learning enhancement method: a proof of concept

Boyang Pan<sup>1†</sup>, Na Qi<sup>2†</sup>, Qingyuan Meng<sup>2†</sup>, Jiachen Wang<sup>1</sup>, Siyue Peng<sup>1</sup>, Chengxiao Qi<sup>1</sup>, Nan-Jie Gong<sup>3\*</sup>  and Jun Zhao<sup>2\*</sup>

<sup>†</sup>Boyang Pan, Na Qi and Qingyuan Meng have contributed equally to this work

\*Correspondence: nanjie.gong@gmail.com; petcenter@126.com

<sup>1</sup>RadioDynamic Healthcare, Shanghai, China

<sup>2</sup>Department of Nuclear Medicine, Shanghai East Hospital, Tongji University School of Medicine, No. 150 Jimo Road, Pudong New District, Shanghai, China

<sup>3</sup>Vector Lab for Intelligent Medical Imaging and Neural Engineering, International Innovation Center of Tsinghua University, No. 602 Tongpu Street, Putuo District, Shanghai, China

## Abstract

**Background:** To generate high-quality bone scan SPECT images from only 1/7 scan time SPECT images using deep learning-based enhancement method.

**Materials and methods:** Normal-dose (925–1110 MBq) clinical technetium 99 m-methyl diphosphonate (99mTc-MDP) SPECT/CT images and corresponding SPECT/CT images with 1/7 scan time from 20 adult patients with bone disease and a phantom were collected to develop a lesion-attention weighted U<sup>2</sup>-Net (Qin et al. in *Pattern Recognit* 106:107404, 2020), which produces high-quality SPECT images from fast SPECT/CT images. The quality of synthesized SPECT images from different deep learning models was compared using PSNR and SSIM. Clinic evaluation on 5-point Likert scale (5 = excellent) was performed by two experienced nuclear physicians. Average score and Wilcoxon test were constructed to assess the image quality of 1/7 SPECT, DL-enhanced SPECT and the standard SPECT. SUVmax, SUVmean, SSIM and PSNR from each detectable sphere filled with imaging agent were measured and compared for different images.

**Results:** U<sup>2</sup>-Net-based model reached the best PSNR (40.8) and SSIM (0.788) performance compared with other advanced deep learning methods. The clinic evaluation showed the quality of the synthesized SPECT images is much higher than that of fast SPECT images ( $P < 0.05$ ). Compared to the standard SPECT images, enhanced images exhibited the same general image quality ( $P > 0.999$ ), similar detail of 99mTc-MDP ( $P = 0.125$ ) and the same diagnostic confidence ( $P = 0.1875$ ). 4, 5 and 6 spheres could be distinguished on 1/7 SPECT, DL-enhanced SPECT and the standard SPECT, respectively. The DL-enhanced phantom image outperformed 1/7 SPECT in SUVmax, SUVmean, SSIM and PSNR in quantitative assessment.

**Conclusions:** Our proposed method can yield significant image quality improvement in the noise level, details of anatomical structure and SUV accuracy, which enabled applications of ultra fast SPECT bone imaging in real clinic settings.

**Keywords:** SPECT, Bone, Image quality enhancement, Deep learning

## Background

Single-photon emission computed tomography (SPECT), registered with anatomical imaging CT (together known as SPECT/CT), combines the advantage of molecular-level functional SPECT images and the precise anatomic details of CT images. It has been proved powerful in bone disease diagnosis and is widely used for the detection of bone metastases. Considering the as low as reasonable achievable (ALARA) principle, dose reduction is important. From the imaging technique, reducing radiation dose and shortening scanning time have the same physical essence. The latter also contributes to the patient experience in SPECT/CT examination and reducing the motion artifact. However, short scanning time also means high imaging noise, low image quality and losing diagnostic value. Advanced methods have been proposed to reduce the scanning time or injected dose without degrading the image quality both from the hardware and software sides. In the 1990s, a multi-head gamma camera was introduced to shorten the scan time. In the 2000s, the invention of cadmium zinc telluride (CZT) detector further improved imaging efficiency [1]. Iftikhar Ali et al. [2] adopted ordered-subset expectation maximization with resolution recovery (OSEM-RR) algorithm to keep the image quality for half-time SPECT myocardial perfusion imaging. Aju P. Pazhenkottil [3] proposed a dose-saving algorithm to depress the radiation in SPECT/CT examination.

In recent years, the deep learning method has made massive achievements on multiple imaging tasks like lesion detection [4–6], disease diagnosis [7, 8] and image augmentation [9–11] for different imaging modalities. Some CNN models succeeded to transform low-quality images into high-resolution images with less noise and clear boundaries. For SPECT images, Liu et al. [12] used a 3D coupled U-Net (CU-Net) to suppress imaging noise and artifacts in SPECT myocardial perfusion imaging (MPI). Shiri et al. [13] explored the performance of deep learning method in two time-reduction MPI-SPECT imaging, namely reduction of the acquisition time per projection and reduction of the number of angular projections.

In this work, we collect anonymous clinical SPECT/CT image pairs, enhance the fast acquired image using deep neural network and evaluate the quality of augmented image quantitatively. Compared to previous work, the innovation of this work lies in: (1) we use clinical SPECT image rather than simulated data used in other works; (2) we combine the SPECT image and corresponding CT images which bear richer anatomical information; and (3) we investigate deep learning performance on ultra-high-speed SPECT (1/7 scan time of normal examination), and the noise level is below the standard for making valid clinical diagnosis.

In this paper, we propose a method based on U<sup>2</sup>-Net architecture but can integrate multi-scale and multi-modality features from both low-dose SPECT and corresponding CT images. We incorporated a lesion attention loss function to enhance the sensitivity of our model to reconstruct lesion regions with more accurate SUV measures. Compared to the state-of-the-art methods, our proposed method achieves the best image quality with the highest peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). Furthermore, we evaluate our proposed method on real clinical data and phantom data. It is demonstrated that synthesized SPECT image from fast SPECT with only 1/7 of standard scan time is reliable and provides high agreement and similar diagnostic confidence

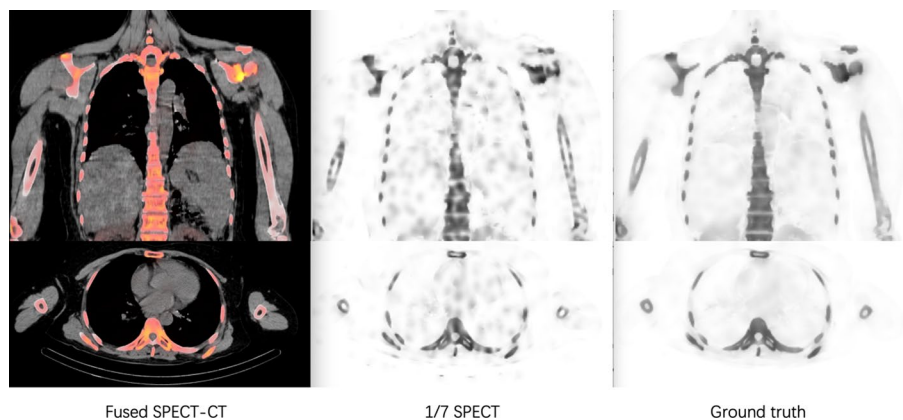
as standard SPECT imaging in clinical routine, which has great value in real clinical applications.

## Material and methods

### Subjects and image acquisition

Patients received systemic bone imaging after the injection of 25–30 mCi (925–1110 MBq) technetium 99 m-methyl diphosphonate ( $^{99m}\text{Tc}$ -MDP) at Shanghai East Hospital. This study was approved by the Institutional Review Board, and all patients signed informed consent before the examination. The SPECT/CT data were collected using Siemens Symbia Intevo T16 with two continuous protocols: one standard scan with 20 s per projection (referred as the standard SPECT) and one fast scan with 3 s per projection (referred as the 1/7 SPECT). Sixty projections were performed each scan. Projection data were then reconstructed based on CT attenuation map of ordered-subset conjugate gradient (OSCG) algorithm enhanced with 2 subsets and 28 iterations without post-smoothing. Low-dose CT images were collected at 130 kV and 10 valid mAs and reconstructed using a smooth attenuation-correction kernel B31s with 3 mm slice thickness. The resolution of each SPECT images is  $256 \times 256$ , and 200 images were collected each scan. Each SPECT voxel represented a  $1.95 \text{ mm} \times 1.95 \text{ mm} \times 1.95 \text{ mm}$  space. The resolution of CT images is  $512 \times 512$ , and 131 slices were collected for each patient. Each CT voxel represented a  $0.97 \text{ mm} \times 0.97 \text{ mm} \times 3 \text{ mm}$  space. All patients were informed to stay still during the examination to keep the quality and correspondence of images. Unmatched 1/7 SPECT and standard SPECT were discarded. Twenty matched groups (11 males, mean age: 56 years, age range: 26–75) of fast and standard SPECT/CT images were collected for further research. Ten subjects were used for training the proposed deep learning model, while the rest 10 subjects were set for testing the synthesized images. One example of training data is shown in Fig. 1.

National Electrical Manufacturers Association (NEMA) International Electrotechnical Commission (IEC) Body Phantom Set was applied as the model, which is a hollow cylinder made of plexiglass with 6 spheres of different diameters (10, 13, 17, 22, 28, 37 mm), and the volume of the hollow cylinder is 9700 ml. The center of the spheres all



**Fig. 1** One example of training dataset. The left images are coronal and axial views of fused SPECT and CT images. The middle images are coronal and axial views of 1/7 SPECT. The right images are coronal and axial views of the standard SPECT

locates on a circle 5 cm from the center of the cylinder and a plane 70 mm away from the upper surface of the cylinder. SPECT/CT quantitative tomography images with 20 s, 3 s/view and total of 60 views were performed at a specific activity of 12:1 after one hour of instilling. 200 slices of SPECT images and 131 slices of CT images were collected. The first 80 slices of SPECT images and corresponding 53 CT images were inserted to the training dataset, and the rest matched images were treated as testing samples.

### Image preprocess

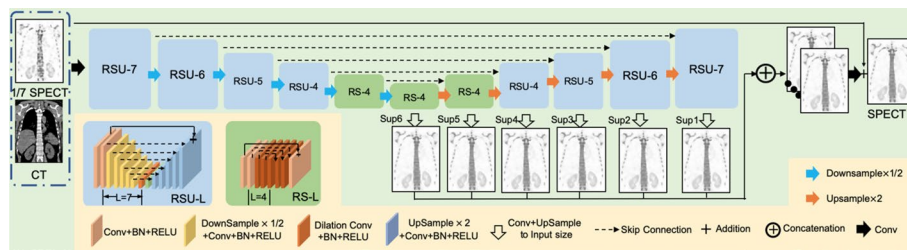
The simultaneous SPECT and CT acquisition modes facilitate the integration of input data from the two modalities. SPECT image provides diagnostic information at the expense of the anatomic features which can be supplemented by the corresponding CT image. Hence, we propose to combine 1/7 SPECT image and CT image as the input and take the standard SPECT as the network output. To facilitate the combination of SPECT and CT images, each collection of CT images was reshaped into a  $200 \times 256 \times 256$  matrix which has the same shape with SPECT. To flatten the difference of voxel values, all the input images and the output were divided by their own average. Then corresponding SPECT and CT slices were concatenate in the first dimension before sending to the proposed U<sup>2</sup>-Net architecture.

The ablation study in the experiment shows the effectiveness of the combination compared with using only 1/7 SPECT as input.

### Residual U-block and U<sup>2</sup>-Net

The 1/7 SPECT varies greatly from the standard SPECT both from the presence of bone structure and voxel value of lesions. So, both local and global contextual features are important for this image synthesis task. U<sup>2</sup>-Net was originally proposed for salient object detection (SOD). The neural network architecture is different from modern CNN designs, such as AlexNet [14], ResNet [15], GoogLeNet [16], DenseNet [17] and VGG [18]. These networks were originally built for image classification tasks which prefer to use small convolutional filters with a size of 1\*1 or 3\*3 to extract features. U<sup>2</sup>-Net [19] is a simple yet powerful deep network architecture. It contains a novel two-layer nested U-shaped structure [20]. The proposed residual U-block (RSU) consists of a mixture of different-sized receive domains that helps capture contextual information on different scales more efficiently. It also uses pooling operations to increase the overall architecture depth without affecting the computational cost much. There are three major components in RSU which are input convolutional layer, U-Net-like symmetric encoder-decoder structure of 'L' height and residual connection to fuse local and multi-scale features using summation. The RSU module with height = 7 is also shown in Fig. 2.

Based on RSU blocks alone, U<sup>2</sup>-Net was developed. It consists of a 6-stage encoder, a 5-stage decoder and a graph fusion module attached to the decoders at different stages. (i) In the encoder stages, RSU-7, RSU-6, RSU-5 and RSU-4 are used, respectively, in which '7', '6', '5' and '4' denote the heights (L) of RSU blocks. As the resolution of feature maps in the middle part of U<sup>2</sup>-Net is relatively low, further downsampling of these feature maps would cause loss of useful context. Hence, we use dilated convolution to replace the pooling and upsampling operations, and this kind of special block is referred to as 'RS-L' which is also shown in Fig. 2 with height = 4. (ii) In the decoder stages, it has



**Fig. 2** Illustration of proposed U<sup>2</sup>-Net architecture. It consists of 4 residual U-block (RSU) encoders with height of L, 3 bottom residual block (RS) and 4 symmetric RSU decoder. Skip connections are used to save spatial information along matched encoders and decoders. 1/7 SPECT and corresponding CT image are used as the input. The output of the network is images convoluted and upsampled from the second and third RS and follow-up decoders

similar structures to their symmetrical encoder stages starting from RS-4. There are 5 stages in total in which each decoder stage takes the concatenation of the upsampled feature maps from its previous stage and those from its symmetrical encoder stage as the input as shown in Fig. 2. (iii) The last graph fusion module is used for generating final synthesized SPECT images. First, this U<sup>2</sup>-Net generates six side output synthesized SPECT images Sup(6), Sup(5), Sup(4), Sup(3), Sup(2), Sup(1), which are upsampled to have the same size as input 1/7 SPECT image. Then, these outputs are fused with a concatenation operation and input to a 1\*1 convolution layer followed by a long skip connection with 1/7 SPECT to generate the final synthesized SPECT image.

**Lesion attention loss and deep supervision**

To ensure the accuracy of the synthesized image value and distinguishability of the structure and important ROIs, we adapt a combination loss function of the structural similarity index (SSIM) loss and the L1 loss. The total loss for each output SPECT at different decoder stages is

$$L = L_1 + \alpha L_{ssim}$$

where  $\alpha$  is a fixed weight ( $\alpha = 0.5$ ) that balances the SSIM loss and L1 loss.

To further improve the performance on lesion regions, we add lesion attention masks to emphasize the loss in these areas. The lesion masks were contoured on standard SPECT by physicians. So, the improved lesion attention loss is defined as

$$\ell = L + \beta L * M$$

where  $\beta$  is a fixed weight ( $\beta = 100$ ) that balances the lesion region loss and whole image loss.  $M$  is the lesion mask.

We also use deep supervision strategy in the training process to speed up the training process and acquire stable media layers. The total loss for training the U<sup>2</sup>-Net is defined as

$$\mathcal{L}_{total} = \sum_{i=1}^N w_{side}^i \ell_{side}^i + w_{final} \ell_{final}$$

where  $\ell_{\text{side}}^i$  ( $N=6$ , as Sup1, Sup2, Sup3, ..., Sup6 in Fig. 2) is the loss of the side output and  $\ell_{\text{final}}$  is the loss of final output of the network.  $w_{\text{side}}^i$  and  $w_{\text{final}}$  control the weights of each component in the total loss. In the testing process, final output is the only part for synthesizing SPECT images.

### Implementation details

The proposed method is implemented using PyTorch 1.6.0 and trained on four NVIDIA GEFORCE 3090 (24 GB). The network is trained for 100 epochs, and the batch size is set to 4 by using axial slice as inference plain. VGG net is used as the discriminator. Adam optimizer is used with the learning rate of 0.0002 for both the generator and discriminator and divided by 10 after 80 epochs.

### Quantitative assessment

To quantitatively evaluate the performance of synthesized images, PSNR and SSIM are used as evaluation metrics. PSNR for synthesized image is defined as

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{\text{MAX}_{\text{gt}}^2}{\text{MSE}} \right)$$

where  $\text{MAX}_{\text{gt}}$  is the maximum pixel value of ground truth standard SPECT. MSE is the mean square error of synthesized images compared to the standard SPECT.

SSIM for synthesized image is defined as

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where  $\mu_x$  and  $\sigma_x^2$  are average value and variance of input synthesized image.  $\mu_y$  and  $\sigma_y^2$  are the average value and variance of input standard SPECT.  $\sigma_{xy}$  is covariance of two images.  $c_1$  and  $c_2$  are small constants. SSIM is calculated using scikit-image package.

### Clinical assessment

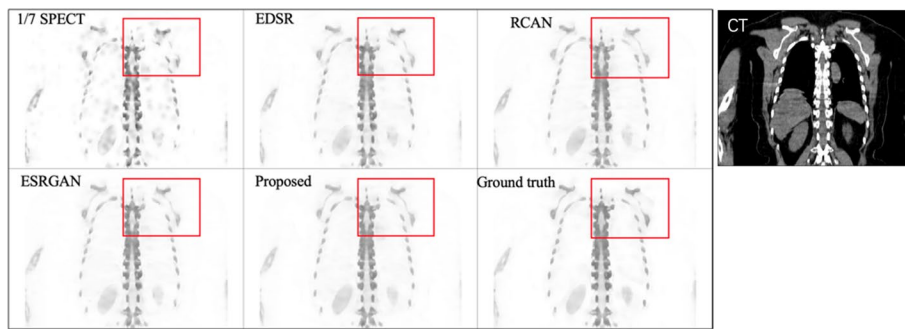
Two readers independently grade 1/7 SPECT, synthesized SPECT by proposed method and standard SPECT in terms of general image quality, detail of 99mTc-MDP distribution, presence of artifacts and general diagnostic confidence using 5-point Likert scale (1 for unacceptable image quality; 2 for suboptimal image quality; 3 for acceptable image quality; 4 for good image quality; and 5 for excellent image quality). Readers are blinded to meta-information of compared images.

Average scores of each kind of image are compared. Paired t test is used to identify significant differences between each criterion.

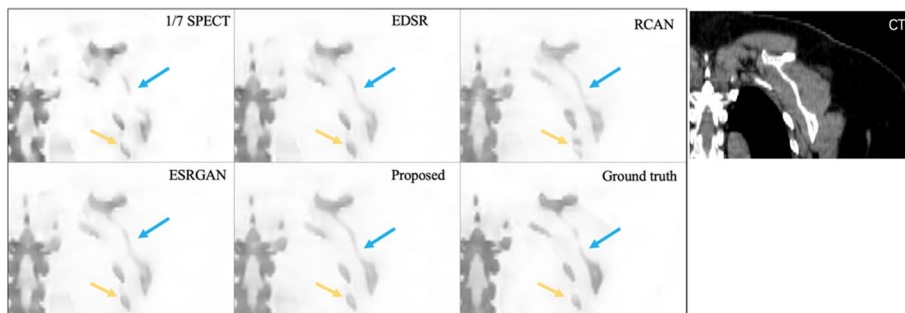
### Phantom study

Half phantom images are used to training the model. The rest half including the center position of the phantom are used for testing. The images of the sphere centers are used to distinguish different spheres. SUVmax and SUVmean are measured and compared for each recognizable sphere. SUV is defined as





**Fig. 3** Visual results of different methods on synthesizing standard SPECT from 1/7 SPECT. Please zoom in for a better comparison. The results are images from the same case using 1/7 scan time constructed with the OSCG algorithm, enhanced by EDSR network, RCAN network, ESRGAN network, proposed U<sup>2</sup>-Net, respectively, and the standard SPECT image



**Fig. 4** Zoomed visual results of different methods on synthesizing standard SPECT from 1/7 SPECT. The results are zoomed images from the same case using 1/7 scan time constructed with the OSCG algorithm, enhanced by EDSR network, RCAN network, ESRGAN network, proposed U<sup>2</sup>-Net, respectively, and the standard SPECT image

$$\text{SUV} = \text{pixel value} * \text{weight in grams} / \text{total dose in MBq.}$$

PSNR and SSIM are calculated for 1/7 SPECT and generated SPECT.

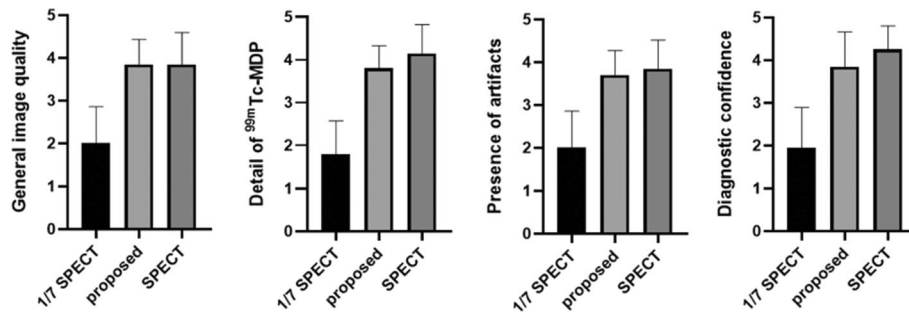
## Results

### Comparison with other networks

The proposed U<sup>2</sup>-Net model is compared with four widely used deep learning architectures, i.e., EDSR [21], RCAN [22] and ESRGAN [23], respectively.

The visualization results for one testing case for different methods are shown in Fig. 3. A zoomed-in region is also shown in Fig. 4 for better comparison. Figure 5 provides an axial view of image difference between different methods and the standard SPECT.

The quantitative results are presented in Table 1. In general, our result reaches the best performance based on both qualitative and quantitative evaluation. The noise in the chest region as shown in the 1/7 SPECT image has been successfully removed in our proposed result. This is the effect of using U<sup>2</sup>-Net as the neural network architecture since it provides abundant contextual information from different scales for improving the anatomical structural details. This is also the effect of adding the CT image as the input of the network due to the clear anatomy feature provided by CT images which



**Fig. 5** An axial view of image difference between different methods and the standard SPECT. **a–f** The original image of 1/7 SPECT, EDSR, ESRGAN, RCAN, proposed method and the standard SPECT in the grayscale of minimum = 0, maximum = 120,000. **g–k** The difference between the standard and corresponding image (standard–corresponding image, minimum = – 20,000, maximum = 20,000)

**Table 1** Average SSIM and PSNR comparison for different methods

	1/7 SPECT	EDSR	RCAN	ESRGAN	Proposed
SSIM	0.765	0.778	0.781	0.772	0.788
PSNR	37.7	38.6	40.5	40.1	40.8

results in clear boundaries in the synthesized standard SPECT image. Our proposed method also increases the sharpness with consistent details and achieves the highest PSNR and SSIM score at the same time (PSNR = 40.8, SSIM = 0.788).

**Clinic evaluation**

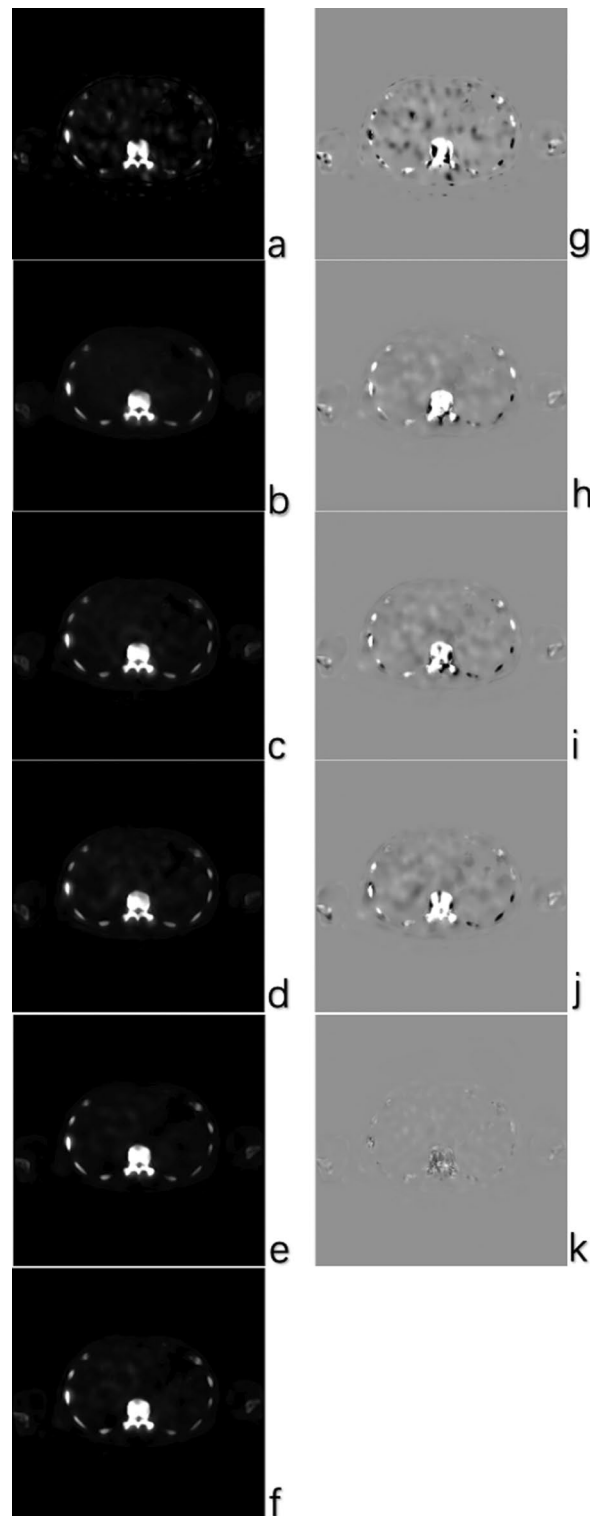
The assessment results are shown in Fig. 6. The average scores for the 1/7 SPECT, proposed method and the standard SPECT are 2, 3.85 and 3.85 in terms of general image quality, 1.8,3.8 and 4.15 for details of contrast, 2, 3.75 and 3.9 for presence of artifacts and 1.95, 3.85 and 4.15 for diagnostic confidence. We can see the average grades of the proposed method are much better than those of 1/7 SPECT. Significant difference was found using Wilcoxon test ( $P < 0.05$ ) for all evaluation metrics. Compared to the standard SPECT, our method achieved the same general image quality ( $P > 0.999$ ), similar detail of <sup>99m</sup>Tc-MDP ( $P = 0.0510$ ), presence of artifacts ( $P = 0.3434$ ) and the diagnostic confidence ( $P = 0.1265$ ).

**Phantom study**

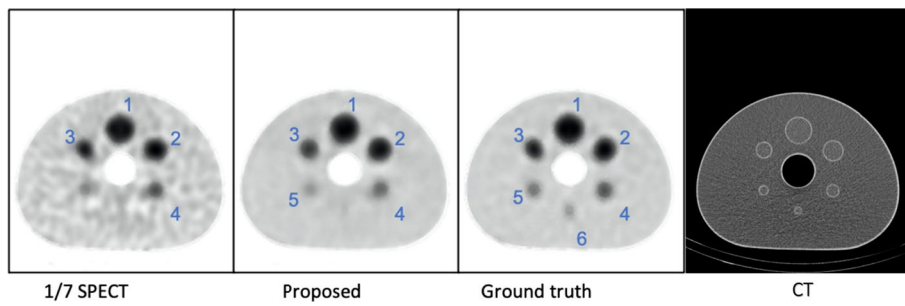
The phantom images acquired from 1/7 of standard scan time are also enhanced using our method. The images of the center part of the phantom are shown in Fig. 7. On 1/7 SPECT, we can only identify 4 spheres that exhibited evident contrast against background signals. On the generated image and the ground truth image, we can distinguish 5 and 6 spheres separately. All these spheres are numbered in the figure for further quantitative evaluation.

SUVmax and SUVmean are calculated for spheres 1 to 4. The quantitative results are shown in Table 2. SUVmax and SUVmean of each sphere in enhanced image are much closer to the ground truth image compared with the 1/7 SPECT image.





**Fig. 6** Clinical assessment results of image quality for 1/7 SPECT, proposed method and the standard SPECT. The average scores for the 1/7 SPECT, proposed method and the standard SPECT are 2, 3.85 and 3.85 in terms of general image quality, 1.8, 3.8 and 4.15 for details of contrast, 2, 3.75 and 3.9 for presence of artifacts and 1.95, 3.85 and 4.15 for diagnostic confidence. 1/7 SPECT is significantly different from SPECT image enhanced by proposed method under Wilcoxon test ( $P < 0.05$ ) for all evaluation metrics. Compared to the standard SPECT, proposed method achieved the same general image quality ( $P > 0.999$ ), similar detail of  $^{99m}\text{Tc}$ -MDP ( $P = 0.125$ ), presence of artifacts ( $P = 0.531$ ) and the diagnostic confidence ( $P = 0.1875$ )



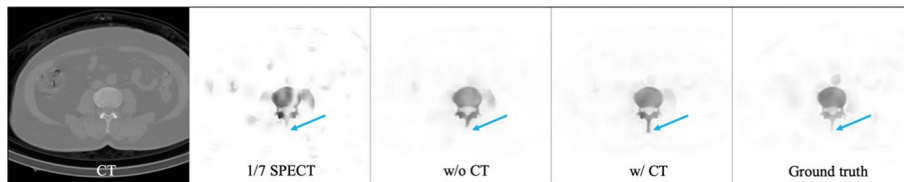
**Fig. 7** Visualization results of phantom images from different methods. On 1/7 SPECT, only 4 spheres can be identified that exhibited evident contrast against background signals. On the generated image and the ground truth image, 5 and 6 spheres can be distinguished separately

**Table 2** Quantitative results for different spheres

	SUVmax			SUVmean		
	1/7 SPECT	Proposed	Ground truth	1/7 SPECT	Proposed	Ground truth
Sphere 1	12.314	13.591	13.756	9.529	10.216	10.379
Sphere 2	13.176	12.4	12.231	8.405	8.095	7.923
Sphere 3	9.733	9.385	9.292	7.366	7.275	7.284
Sphere 4	4.897	4.62	4.735	4.093	3.869	3.885

**Table 3** Average SSIM and PSNR comparison for 1/7 SPECT and enhanced image

	1/7 SPECT	Proposed
SSIM	0.87	0.94
PSNR	32.12	33.54



**Fig. 8** Visual results of w/ and w/o CT as input of the proposed model. Enhanced image with combined CT input preserved more structure than those without CT input and 1/7 SPECT

SSIM and PSNR results for 1/7 SPECT and enhanced image using the standard SPECT as reference are shown in Table 3. Proposed method performed better than 1/7 SPECT on both criteria.

### Ablation study

To verify the effectiveness of using both 1/7 SPECT and CT image as input that boosts the image quality, ablation studies are conducted on w/ or w/o CT, while another implementation set remains the same. The visual results are shown in Fig. 8, in which we can observe that the bone structure pointed by the blue arrow has been recovered in the

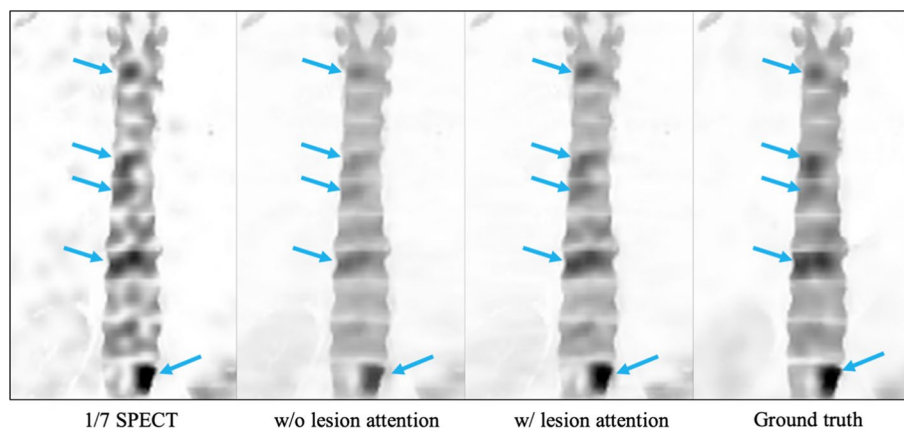
result with both 1/7 SPECT and CT as input, while the result w/o CT as input fails to recover this bone structure due to the missing information in the original 1/7 SPECT image. Ablation study of w/o and w/ lesion attention loss is also shown in Fig. 9. We can see that the synthesized SPECT using lesion attention has more accurate SUV values and better contrast compared to w/o lesion attention as pointed by the blue arrow.

## Discussion

The deep learning-based imaging enhancement method has been applied to many modalities like low-dose PET/MRI [24], compressive sensing MRI [9] and low-dose CT [25]. For SPECT images, Olia et al. [26] implemented a generative adversarial network to predict non-gated normal-dose SPECT-MPI images in the projection domain at different reduced dose levels. Song et al. [27] investigated a 3D residual convolutional neural network (CNN) model to predict standard-dose images from 1/4-dose gated SPECT-MPI images. However, no research has employed an artificial intelligence algorithm on SPECT bone image augmentation to the best of our knowledge. Compared with previous SPECT study, Olia et al. and Song et al. use simulated low-dose SPECT reconstructed from partial list mode data, and our work uses continuous scan to acquire data pairs which is closer to clinical settings. Besides, we explored a more complicated network architecture which could extract smaller features.

This study evaluated 1/7 SPECT/CT image augmentation using the proposed deep learning method on a clinical set. The results showed that an undiagnosable SPECT image could be enhanced by deep learning method to be comparable with the standard SPECT in terms of visual effect. Quantitative results showed that PSNR and SSIM of synthesized SPECT images were much better than the original 1/7 SPECT and the proposed method reached the best performance compared with other advanced deep learning methods.

Organ movements during acquisition series may lead to non-perfect alignment of CT to SPECT and fast scan to the standard scan. Our result shows this phenomenon is not severe in organs visible in SPECT images like kidney. Our result shows the boundary of these soft tissues remains clear. The main problem of soft tissue quality is still noise level.



**Fig. 9** Visual results of w/ and w/o lesion attention as input of the proposed model. Enhanced image with lesion attention loss holds clearer SUV variation within tissues than those without lesion attention

Clinic assessment exhibits that the clinical value of synthesized images is much better than this of low-dose SPECT images and comparable to the standard SPECT images. The average score for deep learning enhanced image is obviously higher 3 and close to 4 in 5-point Likert scale, which proved the image quality meets the standard for clinic use. Dietze et al. [28] compared different methods for SPECT image reconstruction including filtered back projection (FBP), clinical used method, convolution neural network (CNN)-based method and Monte Carlo-based method. They concluded that images generated by CNN-based method reached similar image quality as those reconstructed by the Monte Carlo method but spent far less time. This result also demonstrated the potential of CNN-based imaging enhancement method in clinical use.

Compared to other widely used network architectures, our proposed U<sup>2</sup>-Net reached the best PSNR and SSIM performance. U<sup>2</sup>-Net has been verified useful in image segmentation task [19], but the augmentation task is quite different in predicting the exact value of each pixel. Our visual results (Figs. 3, 4) proved U<sup>2</sup>-Net-based method has the best capability to handle small structures which could be discontinuous or invisible on 1/7 SPECT images. Since no bone scan has been enhanced by deep learning method, we compared our results with other SPECT study. Li et al. [29] compared SSIM and PSNR of DL-enhanced phantom images on different dose levels. They received 0.9470 SSIM of simple Jaszczak phantom from 60-view projection data and 0.9117 from 30-view projection data. It is higher than our clinical result and comparable to our phantom result. High noise level in background is the main reason accounting for the relatively low SSIM result. If we set the background all to 0, we got SSIM of about 0.87 for 1/7 SPECT and about 0.96 for proposed method which is on the same level as Li's result.

The ablation study showed the effectiveness of combining both SPECT and corresponding CT as the input, which makes it possible to recover the bone structures that were missing on 1/7 SPECT. Associating a functional image and the matching structural image as the input of the neural network for image augmentation task has been proved effective in low-dose PET image recovery task [24]. Our result showed the possibility of structure recovery in fast SPECT that not presented on the standard SPECT. This means the input CT image may not only provide the structure information but also contribute to the pixel value which make the result misleading. One solution to this issue could be use the binarized gradient map of CT image which only provide structure boundaries. Increasing the training dataset may also help to solve this problem by generating a more precise model.

The ablation study also proved the usefulness of adding lesion region loss in the training process, which retrieved the contrast within structures much better than those without lesion region loss. The SUV value in lesion area is more important in the clinical criteria. Ly et al. [30] evaluated the SUV of low-dose PET, the standard PET and PET synthesized using deep learning method. The simulated image has equal SUV performance as the low-dose image, but far less than full-dose PET image. Our results exhibit that the SUV distribution with lesion attention is much more precise than those without lesion attention, which means the strategy also contributed to the accuracy of SUV.

Our experiment proved the feasibility of greatly accelerating SPECT imaging without sacrificing image quality using deep learning method. The fast SPECT imaging technique ameliorates patient experience and reduces motion artifacts. Methods have been

proposed to speed up SPECT imaging [1, 2, 12, 13]. However, traditional methods like using advanced detectors or using iterative algorithms could only increase the imaging efficiency by a factor of two, and our methods could apply to ultra-low-dose SPECT imaging with only 1/7 scan time. Compared to other DL-based methods applying to MPI-SPECT, our method improved the imaging quality of bone SPECT. We also made it possible to make the image quality of synthesized SPECT comparable to that of the standard acquisitions.

In this work, we also applied our method to the phantom data. Chrysostomou et al. [31] used deep learning method to reconstruct low-projection SPECT imaging. They proved SSIM, MSE and PCC of enhanced phantom images were better than that of simulated low-projection SPECT images. In our study, we scan the real phantom and measured SSIM, PSNR, SUVmax and SUVmean and reach the same results as the previous work.

Our method could also enhance the image quality of low-dose SPECT images for both shortening the scan time and reducing the amounts of injected radionuclides, resulting in the reduction of received signals for SPECT scanners, which is the direct factor affecting SPECT image quality. Ramon et al. [32] evaluated the diagnostic accuracy of CNN-enhanced low-dose MPI-SPECT image and draw a conclusion that DL denoising can achieve additional dose reduction without sacrificing the diagnostic accuracy in SPECT-MPI compared with iteration algorithms.

One limitation of our study is that our method is tested on a single SPECT/CT system, single contrast agent and the same imaging reconstruction method. Further, we did not study the impact on image quality in special patient groups such as obese patients, young or old patients and patients with other metabolic diseases. Nevertheless, this is the first study to evaluate the performance of deep learning method on fast bone SPECT using clinical data. Further study in a larger population is required to explore the optimum weights for different parts of the total loss and cut off the extra channels in the network to speed up the calculation. We believe this method could also be applied to greatly reducing injection dose if further validated.

## Conclusion

In this paper, we focus on the synthesis of standard SPECT from a ultra fast SPECT scan with only 1/7 scan time. We propose to apply a novel U<sup>2</sup>-Net-based framework that aggregates features from both fast SPECT scan and corresponding CT image as well as extract both local and global information from multi-scale features. We demonstrated the proposed method was able to achieve an acquisition time reduction by 7 times. The comprehensive results showed the proposed method can yield significant image quality improvement both in the noise level, anatomical structure clearness and SUV accuracy which enables application in real clinical settings.

## Abbreviations

ALARA	As low as reasonable achievable
CNN	Convolution neural network
CU-Net	Coupled U-Net
CZT	Cadmium zinc telluride
DL	Deep learning
FBP	Filtered back projection

MPI	Myocardial perfusion imaging
NEMA	National electrical manufacturers association
OSEM-RR	Ordered-subset expectation maximization with resolution recovery
PSNR	Peak signal-to-noise ratio
RS	Residual block
RSU	Residual U-block
SOD	Salient object detection
SSIM	Structural similarity index
SUV	Standard uptake value
SUVmax	Maximum standardized uptake value
SUVmean	Mean standardized uptake value
<sup>99m</sup> Tc-MDP	Technetium 99 m-methyl diphosphonate

#### Acknowledgements

Not applicable.

#### Author contributions

BP built and trained the network and drafted the manuscript. NQ analyzed the data and gave the clinical result. QM collected the data and helped with the clinical analysis. JW helped build the network and revised the paper. SP revised the paper. CQ revised the paper. JZ designed the data collection process and supervised data collection and clinical analysis process. NG designed the study concept, network building process and supervised technical analysis process. All authors read and approved the final manuscript.

#### Funding

This work was supported by Project of Science and Technology Commission of Shanghai Municipality (19DZ1930703) and Shanghai Pujiang Program (2020PJD038).

#### Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

Institutional Review Board approval from Ethical Review of Medical Ethics Committee of Shanghai East Hospital EC. D (BG). 009. 02.1 was obtained.

##### Consent for publication

All patients signed informed consent before the examination.

##### Competing interests

Author NG is stockholder of RadioDynamic Healthcare.

Received: 18 February 2022 Accepted: 29 May 2022

Published online: 13 June 2022

#### References

- Nkoulou R, Pazhenkottil AP, Kuest SM, et al. Semiconductor detectors allow low-dose-low-dose 1-day SPECT myocardial perfusion imaging. *J Nucl Med*. 1978;2011(52):1204–9.
- Ali I, Ruddy TD, Almgrahi A, Anstett FG, Wells RG. Half-time SPECT myocardial perfusion imaging with attenuation correction. *J Nucl Med*. 1978;2009(50):554–62.
- Pazhenkottil AP, Herzog BA, Husmann L, et al. Non-invasive assessment of coronary artery disease with CT coronary angiography and SPECT: a novel dose-saving fast-track algorithm. *Eur J Nucl Med Mol Imaging*. 2010;37:522–7.
- Lung K, Chang C, Weng S, Lin H, Shuai H, Cheng W. ROSNet: Robust one-stage network for CT lesion detection. *Pattern Recognit Lett*. 2021;144:82–8.
- Rowe SP, Macura KJ, Mena E, et al. PSMA-based [<sup>18</sup>F]DCFPyL PET/CT is superior to conventional imaging for lesion detection in patients with metastatic prostate cancer. *Mol Imag Biol*. 2016;18:411–9.
- Bagher-Ebadian H, Janic B, Liu C, et al. Detection of dominant intra-prostatic lesions in patients with prostate cancer using an artificial neural network and mr multi-modal radiomics analysis. *Front Oncol*. 2019;9:1313–1313.
- Liu M, Cheng D, Wang K, Wang Y. Alzheimer's disease neuroimaging initiative, the Alzheimer's disease neuroimaging initiative. Multi-modality cascaded convolutional neural networks for alzheimer's disease diagnosis. *Neuroinformatics (Totowa, NJ)*. 2018;16:295–308.
- Sivaranjini S, Sujatha CM. Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimed Tools Appl*. 2020;79:15467–79.
- Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Trans Med Imaging*. 2019;38:167–79.
- Cui J, Gong K, Guo N, et al. PET image denoising using unsupervised deep learning. *Eur J Nucl Med Mol Imaging*. 2019;46:2780–9.



11. Hong X, Zan Y, Weng F, Tao W, Peng Q, Huang Q. Enhancing the image quality via transferred deep residual learning of coarse PET sinograms. *IEEE Trans Med Imaging*. 2018;37:2322–32.
12. Liu J, Yang Y, Wernick MN, Pretorius PH, King MA. Deep learning with noise-to-noise training for denoising in SPECT myocardial perfusion imaging. *Med Phys (Lanc)*. 2021;48:156–68.
13. Shiri I, Sabet KA, Arabi H, et al. Standard SPECT myocardial perfusion estimation from half-time acquisitions using deep convolutional residual neural networks. *J Nucl Cardiol*. 2020;28:2761–79.
14. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. *Commun ACM [serial online]*. 2017;60:84–90.
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015.
16. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. 2014.
17. Huang G, Liu Z, Laurens V, et al. Densely Connected Convolutional Networks. *IEEE Computer Society*; 2016.
18. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *International conference on learning representations*, 2015.
19. Qin X, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M. U<sup>2</sup>-Net: going deeper with nested U-structure for salient object detection. *Pattern Recognit*. 2020;106: 107404.
20. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.
21. Lim B, Son S, Kim H, Nah S, Lee KM. Enhanced Deep Residual Networks for Single Image Super-Resolution. 2017.
22. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. 2018.
23. Wang X, Yu K, Wu S, et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. 2018.
24. Chen KT, Gong E, de Carvalho Macruz FB, et al. Ultra-low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiology*. 2019;290:649–56.
25. Chen H, Zhang Y, Kalra MK, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging*. 2017;36:2524–35.
26. Olia NA, Kamali-Asl A, Tabrizi SH, et al. Deep learning-based noise reduction in low dose SPECT myocardial perfusion imaging: quantitative assessment and clinical performance. 2021
27. Song C, Yang Y, Wernick MN, Pretorius PH, King MA. Low-dose cardiac-gated SPECT studies using a residual convolutional neural network. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*; IEEE; 2019. p. 653–656.
28. Dietze MMA, Branderhorst W, Kunnen B, Viergever MA, de Jong HWAM. Accelerated SPECT image reconstruction with FBP and an image enhancement convolutional neural network. *EJNMMI Phys*. 2019;6:1–12.
29. Li S, Ye W, Li F. LU-Net: combining LSTM and U-Net for sinogram synthesis in sparse-view SPECT reconstruction. *Math Biosci Eng*. 2022;19(4):4320–40. <https://doi.org/10.3934/mbe.2022200>.
30. Ly J, Minarik D, Jögi J, Wollmer P, Trägårdh E. Post-reconstruction enhancement of [18F] FDG PET images with a convolutional neural network. *EJNMMI Res*. 2021;11:48–10.
31. Chrysostomou C, Koutsantonis L, Lemesios C, Papanicolaos CN. Deep Convolutional Neural Network for Low Projection SPECT Imaging Reconstruction. 2021.
32. Ramon AJ, Yang Y, Pretorius PH, Johnson KL, King MA, Wernick MN. Improving diagnostic accuracy in low-dose SPECT myocardial perfusion imaging with convolutional denoising networks. *IEEE Trans Med Imaging*. 2020;39:2893–903.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---