Received: 16 October 2023 | Revised: 27 March 2024 | Accepted: 4 May 2024

DOI: 10.1002/mp.17129

MEDICAL PHYSICS

RESEARCH ARTICLE

Enhancing voxel-based dosimetry accuracy with an unsupervised deep learning approach for hybrid medical image registration

Keon Min Kim[1,2] | **Minseok Suh**[3,4] | **Haniff Shazwan Muhd Safwan Selvam**[5] | **Teik Hin Tan**[5] | **Gi Jeong Cheon**[3,6,7] | **Keon Wook Kang**[3,4,8] | **Jae Sung Lee**[1,2,3,4]

[1] Interdisciplinary Program in Bioengineering, Seoul National University Graduate School, Seoul, Republic of Korea

[2] Integrated Major in Innovative Medical Science, Seoul National University Graduate School, Seoul, Republic of Korea

[3] Department of Nuclear Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

[4] Institute of Radiation Medicine, Medical Research Center, Seoul National University, Seoul, Republic of Korea

[5] Nuclear Medicine Centre, Sunway Medical Centre, Subang Jaya, Selangor, Malaysia

[6] Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea

[7] Cancer Research Institute & Institute on Aging, Seoul National University, Seoul, Republic of Korea

[8] Bio-MAX Institute, Seoul National University, Seoul, Republic of Korea

Correspondence
Jae Sung Lee, Department of Nuclear Medicine, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea.
Email: jaes@snu.ac.kr

Present address
Keon Min Kim, Interdisciplinary Program in Bioengineering, College of Engineering, Seoul National Graduate School, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea.

Funding information
National Research Foundation of Korea, Grant/Award Number: 2020M2D9A1093989

Abstract

Background: Deformable registration is required to generate a time-integrated activity (TIA) map which is essential for voxel-based dosimetry. The conventional iterative registration algorithm using anatomical images (e.g., computed tomography (CT)) could result in registration errors in functional images (e.g., single photon emission computed tomography (SPECT) or positron emission tomography (PET)). Various deep learning-based registration tools have been proposed, but studies specifically focused on the registration of serial hybrid images were not found.

Purpose: In this study, we introduce CoRX-NET, a novel unsupervised deep learning network designed for deformable registration of hybrid medical images. The CoRX-NET structure is based on the Swin-transformer (ST), allowing for the representation of complex spatial connections in images. Its self-attention mechanism aids in the effective exchange and integration of information across diverse image regions. To augment the amalgamation of SPECT and CT features, cross-stitch layers have been integrated into the network.

Methods: Two different ^{177}Lu DOTATATE SPECT/CT datasets were acquired at different medical centers. 22 sets from Seoul National University and 14 sets from Sunway Medical Centre are used for training/internal validation and external validation respectively. The CoRX-NET architecture builds upon the ST, enabling the modeling of intricate spatial relationships within images. To further enhance the fusion of SPECT and CT features, cross-stitch layers have been incorporated within the network. The network takes a pair of SPECT/CT images (e.g., fixed and moving images) and generates a deformed SPECT/CT image. The performance of the network was compared with Elastix and TransMorph using L1 loss and structural similarity index measure (SSIM) of CT, SSIM of normalized SPECT, and local normalized cross correlation (LNCC) of SPECT as metrics. The voxel-wise root mean square errors (RMSE) of TIA were compared among the different methods.

Results: The ablation study revealed that cross-stitch layers improved SPECT/CT registration performance. The cross-stitch layers notably enhance SSIM (internal validation: 0.9614 vs. 0.9653, external validation: 0.9159 vs. 0.9189) and LNCC of normalized SPECT images (internal validation: 0.7512 vs. 0.7670, external validation: 0.8027 vs. 0.8027). CoRX-NET with the cross-stitch layer achieved superior performance metrics compared to Elastix and

---

TransMorph, except for CT SSIM in the external dataset. When qualitatively analyzed for both internal and external validation cases, CoRX-NET consistently demonstrated superior SPECT registration results. In addition, CoRX-NET accomplished SPECT/CT image registration in less than 6 s, whereas Elastix required approximately 50 s using the same PC's CPU. When employing CoRX-NET, it was observed that the voxel-wise RMSE values for TIA were approximately 27% lower for the kidney and 33% lower for the tumor, compared to when Elastix was used.

Conclusion: This study represents a major advancement in achieving precise SPECT/CT registration using an unsupervised deep learning network. It outperforms conventional methods like Elastix and TransMorph, reducing uncertainties in TIA maps for more accurate dose assessments.

KEYWORDS

^{177}Lu-DOTATATE, deep learning, deformable registration, SPECT/CT, voxel-based dosimetry

1 | INTRODUCTION

Peptide receptor radionuclide therapy (PRRT) with ^{177}Lu-octreotate (i.e., ^{177}Lu-DOTATATE) is a prominent therapeutic option for neuroendocrine tumors (NETs) exhibiting somatostatin receptor type-2 positivity.[1] After ^{177}Lu-DOTATATE PRRT, retrospective dosimetry is required to assess absorbed doses to organs at risk (i.e., kidneys and bone marrow) and dose response for tumor regions.[2–4] In particular, voxel-based dosimetry for therapeutic radiopharmaceuticals, utilizing quantitative images, offers the potential for greater accuracy compared to organ-based dosimetry, which fails to account for morphological variations and tissue heterogeneity.[5] The quantitative assessment of radioactivity in voxel-level can be performed using SPECT, utilizing the gamma rays emitted with energies of 208 and 113 keV from ^{177}Lu. Additionally, accurate voxel-based dosimetry can be conducted by leveraging the density information obtained sequentially from CT scans.[6]

The voxel-based dosimetry procedure, while lacking standardization, typically involves the following steps: (1) SPECT/CT Hybrid imaging scans: sequential imaging scans are performed at multiple time points. (2) Image registration: These sequential images are registered, aligning them accurately. (3) Time-Integrated Activity (TIA) map generation: A TIA map is generated, reflecting the cumulative activity over time. (4) Dose map calculation: Dose maps are computed based on the TIA data and corresponding CT image. In particular, to calculate the TIA at the voxel level, it is necessary to align the SPECT images to ensure spatial consistency throughout the time series. However, SPECT-SPECT image registration poses difficulties due to the limited spatial resolution, absence of anatomical details, noise, and temporal changes in the distribution of activity over time. Utilizing SPECT/CT hybrid imaging was advantageous for SPECT registration by employing registration fields derived from CT-CT registration; however, visual verification was essential to reduce spatial mismatch.[5] It is inherent that errors may arise during TIA fitting if spatial mismatches persist in SPECT images not only for organ level, but also for voxel level.[7–9]

Unfortunately, there is no consensus on the optimal method for image registration between sequential hybrid image datasets for TIA map generation. Typically, affine and deformable registration is conducted between anatomical images (e.g., CT) and these registration parameters are then applied to functional images (e.g., SPECT or PET). The rationale behind this approach lies in the need for high-resolution anatomical information for accurate nonlinear image registration. Nevertheless, misalignment between anatomical, and functional images, both acquired sequentially, can result in registration errors.[10,11] Tumor regions, in particular, are susceptible to registration errors due to the use of non-contrast low-dose CT for SPECT/CT and PET/CT scans. Consequently, voxel-level TIA values exhibit considerable uncertainty, which is subsequently propagated to the dose estimation.

Various methods have emerged for nonlinear medical image registration.[12–14] Conventional approaches often employ iterative optimization algorithms but suffer from slow registration speeds due to extensive computational demands. In contrast, deep learning networks have recently gained prominence as an alternative, thanks to their speed and robustness. Notably, 3D convolutional neural networks and 3D U-Net have been applied for supervised learning in registrations of brain magnetic resonance imaging (MRI), abdominal MRI, and pulmonary CT.[15–17]

Supervised learning, however, necessitates access to true labels (deformation matrices) for each training data, which is not always feasible. To address this limitation, unsupervised learning methods have been introduced for image registration. These include DIRNet, VoxepMorph, and strategies inspired by traditional optical flow concepts.[18–23] Recently, TransMorph, which utilizes the ST known for its exceptional performance in image processing, was proposed for unsupervised medical image

TABLE 1 ^{177}Lu-DOTATATE SPECT/CT image data sets.

	SNUH	SMC
# of Image sets	22 (Total 82 SPECT/CT)	14 (Total 56 SPECT/CT)
Device	GE NM670	Siemens symbia intevo
Protocol	4, 24, 48, 120 h post-injection	0.5, 4, 24, 48 h post-injection
CT image	512 × 512, 161 slices	512 × 512, Variable thickness and slice number
SPECT image	128 × 128, 128 slices (3.87 × 3.87 × 3.87 mm^3)	103 × 103, variable slice number (4.88 × 4.88 × 4.88 mm^3)
Reconstruction	Ordered-Subsets Expectation Maximization (OSEM) (Attenuation correction (AC), resolution recovery (RR))	Ordered-Subset Conjugate Gradient Minimizer (OSCGM) (Attenuation correction (AC), scatter correction (SC))

registration.[24–26] However, studies specifically focused on the registration of serial hybrid images were not found.

In this study, we introduce CoRX-NET, a novel unsupervised deep learning network designed for deformable registration of hybrid medical images, such as SPECT/CT or PET/CT, which is pivotal for voxel-based dosimetry. The CoRX-NET architecture builds upon the ST, enabling the modeling of intricate spatial relationships within images. Its self-attention mechanism facilitates efficient information exchange and integration across different image regions. To further enhance the fusion of SPECT and CT features, cross-stitch layers have been incorporated within the network.[27] In the following sections, we will describe how we have designed, trained, and evaluated this new network model for the deformable registration of ^{177}Lu-DOTATATE SPECT/CT images.

2 MATERIALS AND METHODS

2.1 Datasets

This study comprises two different datasets acquired using different SPECT/CT machines. The brief information for two datasets including imaging protocols is provided in Table 1. The first dataset consists of 22 SPECT/CT image sets obtained from seven patients who underwent ^{177}Lu-DOTATATE therapy at Seoul National University Hospital (SNUH, Republic of Korea). All seven patients were male, with five diagnosed with rectal NET, one with pancreatic NET, and the remaining one with multiple NETs. The averaged activity of administered ^{177}Lu-DOTATATE for 22 sets was 6.98 (±1.03) GBq. Among these sets, 17 include sequential SPECT/CT images obtained at four different time points: 4, 24, 48, and 120 h post-radiopharmaceutical injection. These images were acquired using the GE NM 670 SPECT/CT system. The remaining five sets consists of three sequential SPECT/CT images obtained at three out of four previously mentioned time points. This first dataset was utilized for network training and internal validation.

The second dataset comprises 14 SPECT/CT image sets from 14 patients who received ^{177}Lu-DOTATATE therapy at Sunway Medical Centre (SMC, Malaysia). Among the 14 patients, seven were male, and the remaining seven were female patients. Seven patients were identified with pancreatic NET, four with rectal NET, one with duodenal NET, one with renal NET, and one with sacral paraganglioma. The averaged activity of administered ^{177}Lu-DOTATATE for 22 sets was 7.96 (±0.35) GBq. In this case, four sequential SPECT/CT images were acquired at 0.5, 4, 24, and 48 h after the radiopharmaceutical injection, utilizing the Siemens Symbia Intevo SPECT/CT system. This second dataset was utilized for external validation.

Detail patient demographics for both datasets can be found in Table S1. Ethical approval was obtained from the Institutional Review Boards at both SNUH and SMC, and the study was conducted retrospectively.

2.2 Deep neural networks

The architecture of our proposed unsupervised deep learning network for deformable registration of hybrid imaging, which we've named CoRX-NET, is shown in Figure 1. CoRX-NET consists of two parallel streamlines, both sharing identical structures. Each of these streamlines takes two CT or SPECT images as input (i.e., fixed CT, moving CT, fixed SPECT, and moving SPECT), and leverages the advantages of ST.[24] The network architecture consists of several stages, with each stage including a set of ST blocks. ST is a kind of hierarchical vision Transformer, which utilizes the Shifted Window Multi-Head Self Attention (MSA). The ST blocks consist of conventional MSA and Shifted Window MSA, followed by a multi-layer perceptron. Following the final ST block, continuous up-sampling and a double convolution layers are used to resize the images, ensuring they match the original input size. The structure of each streamline of CoRX-NET closely resembles the existing Transmorph, but we have notably adopted shallower layers and fewer ST blocks, which helps manage the overall network size more effectively.[24] In particular, due to the necessity of incorporating four images of size 128 × 128 × 128 into the network, the network size was inevitably set to be relatively small. The depth of layer is three and number of ST blocks for each layer is two. The head numbers for MSA for four for the first two

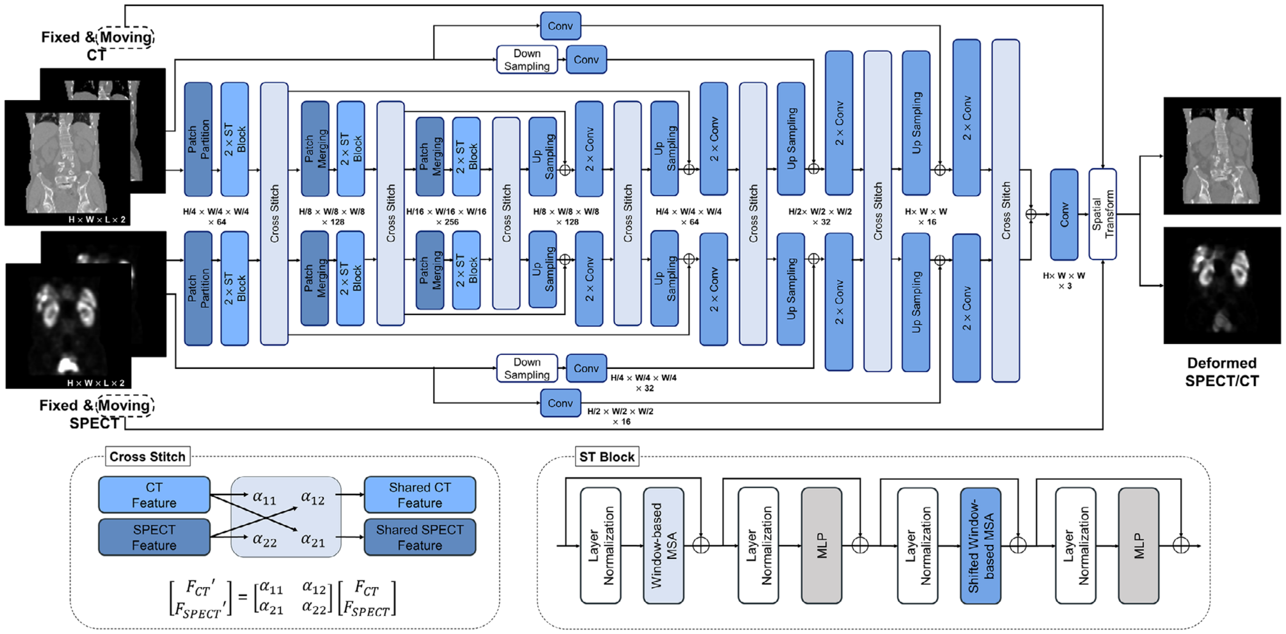

FIGURE 1 The structure of CoRX-NET.

layers and eight for the last layer. The window size used for attention mechanism is 4 × 4 × 4. The number of channels in the first layer is 64, doubled with each subsequent layer.

In this study, we have introduced cross-stitch layers following the ST block and the double convolution layer to facilitate the exchange of information between features extracted from CT and SPECT images.[27] The cross-stitch layers were specifically designed to improve the network performance, especially in tasks that require cross-modality information analysis.[28,29] The cross-stitch layers combine the features in a weighted manner, allowing for the effective integration of information from SPECT and CT images. The cross-stitch layer propagates the shared features from CT and SPECT images into the next layer through a learnable linear combination mechanism, as follows:

$$\begin{bmatrix} F_{CT}' \\ F_{SPECT}' \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} F_{CT} \\ F_{SPECT} \end{bmatrix} \quad (1)$$

where F_{CT} and F_{SPECT} are the features from CT and SPECT before propagating to layers. F_{CT}' and F_{SPECT}' are the features from CT and SPECT after propagating to layers. α_{ij} are linear coefficient of the cross-stitch layer learned during training. The learnable linear coefficients of each cross-stitch layer were trained independently.

2.3 | Network training and evaluation

As mentioned earlier, the SNUH dataset was utilized for network training and internal evaluation. Before the network training, CT images were resampled to match the voxel size and image dimension of the SPECT images, resulting in 128 × 128 × 128 image size with 3.87 × 3.87 × 3.87 mm^3 voxel size considering network size and simultaneous SPECT/CT registration. In contrast to CT images, SPECT images acquired at different time points following radiopharmaceutical injection exhibited varying image intensity level (i.e., count or Bq/mL). However, SPECT images should have similar intensity distribution because similarity measures employed for network training and evaluation, such as mutual information and normalized cross correlation, rely on the image intensity. Therefore, the intensity of each SPECT image was normalized using min-max normalization. For external validation, SMC SPECT/CT image sets were also resampled and cropped to align with the image dimensions of SNUH dataset. However, SPECT images acquired at 0.5 h post-injection often exhibited highly concentrated radioactive distribution in specific organs, such as the bladder. In addition, in most cases, there was insufficient radiopharmaceutical accumulation in target tumor region. Applying min-max normalization in such cases would result in the loss of radioactivity distribution information outside those concentrated areas, rendering it unsuitable for evaluation. Therefore, the images acquired at 0.5 h post-injection were excluded from external validation.

After image processing, two SPECT/CT images from each image set from SNUH dataset were grouped into image pairs, corresponding to a fixed SPECT/CT image and a moving SPECT/CT image. Five of 22 sets from SNUH dataset contain three sequential SPECT/CT acquired in different time and the remains contain four sequential SPECT/CT. The image pairs for network training and evaluation were generated considering all possible combinations within each set to increase the number of training datasets. Therefore, 12 pairs (calculated as 4P2) were generated from each set of 17, and 6 pairs (calculated as 3P2) were generated from each set of five. Total 234 image pairs (calculated as 12×17 + 6×5) were separated to 204 and 30 image pairs for network training and internal evaluation. To augment the number of training dataset, we applied left-right image flipping and ± 5° rotation around the *z*-axis for training dataset, resulting in total 1224 image pairs for network training. In case of SMC dataset, all possible combinations were considered as external validation data sets as in the SNUH datasets. Accordingly, 84 (calculated as 14 × 3P2) image pairs were included in external validation.

The loss function used for network training was the energy function employed in conventional image registration algorithms:

$$Loss = L_{sim}^{CT}\left(I_F^{CT},\ \phi\left(I_M^{CT}\right)\right) + \lambda_{SPECT} L_{sim}^{SPECT}\left(I_F^{SPECT},\ \phi\left(I_M^{SPECT}\right)\right) + \lambda_{Reg} R(\phi) \tag{2}$$

where I_F^{CT}, I_M^{CT}, I_F^{SPECT}, and I_M^{SPECT} are fixed CT, moving CT, fixed SPECT, and moving SPECT, respectively. ϕ is the deformation field, generated by CoRX-NET. L_{sim}^{CT}, L_{sim}^{SPECT}, and $R(\phi)$ are similarity measure for CT, SPECT, and regularization function of ϕ, respectively. λ_{SPECT} and λ_{Reg} are weighting factors for similarity measure for SPECT and regularization function, respectively.

For L_{sim}^{CT}, mean squared loss (MSE) between fixed CT and deformed CT calculated for all points p in image space Ω was used:

$$L_{sim}^{CT}\left(I_F^{CT},\ \phi\left(I_M^{CT}\right)\right) = MSE\left(I_F^{CT},\ \phi\left(I_M^{CT}\right)\right) = \frac{1}{\Omega}\sum_{p\in\Omega}\left(I_F^{CT} - \phi\left(I_M^{CT}\right)\right)^2 \tag{3}$$

For L_{sim}^{SPECT}, local normalized cross correlation (LNCC) was used to quantify the similarity between fixed SPECT and deformed SPECT, where K is local window with size of 9^3[30]:

$$L_{sim}^{SPECT}\left(I_F^{SPECT},\ \phi\left(I_M^{SPECT}\right)\right) = LNCC\left(I_F^{SPECT},\ \phi\left(I_M^{SPECT}\right)\right) = \sum_{p\in K}\frac{\left(\sum_p\left(I_F^{SPECT}(p) - \overline{I_F^{SPECT}}(p)\right)\left(\phi\left(I_M^{SPECT}\right)(p) - \overline{\phi\left(I_M^{SPECT}\right)}(p)\right)\right)^2}{\left[\sum_p\left(I_F^{SPECT}(p) - \overline{I_F^{SPECT}}(p)\right)^2\right]\left[\sum_p\left(\phi\left(I_M^{SPECT}\right)(p) - \overline{\phi\left(I_M^{SPECT}\right)}(p))\right)^2\right]}. \tag{4}$$

For $R(\phi)$, diffusion regularization term was used to avoid folding artifact, where $u(p)$ is spatial gradients[30]:

$$R(\phi) = \sum_{p\in\Omega} \|\nabla u(p)\|^2 \tag{5}$$

λ_{SPECT} and λ_{Reg} are set to 0.1 and $8e^{-4}$ respectively according to our empirical experiments. The AdamW optimizer was used for the network training and the initial learning rate was set to 0.01.

2.4 | Registration performance comparison

The registration performance of CoRX-NET was compared with Elastix and TransMorph.[14,24] In Elasitx, deformable registration using the BSpline transform was performed, with the advanced Mattes mutual information similarity metric. To ensure fair comparison with CoRX-NET, registration was conducted on SPECT/CT images, each with dimensions of 128 × 128 ×128. Registration was first performed on CT images, and the corresponding deformation field was then applied to SPECT images. A maximum number of iterations set for registration was 1000. For TransMorph, vanilla version of TransMorph with 64 channels in the first layer was used.[24] The same loss function, including weighting factors, as used in CoRX-NET was utilized for network training, but only CT images were used as inputs of the TransMorph. For the comparison, L1 difference and Structural Similarity Index Measure (SSIM) were calculated for CT images, while SSIM, and LNCC values were computed for normalized SPECT images.

2.5 | TIA estimation

TIA maps were generated from the registered images to evaluate the dosimetry accuracy using the external dataset. 4-h SPECT/CT images were considered as the fixed images and 24- and 48- h SPECT/CT images were considered as the moving images. Then the 24- and 48-h images were registered to 4-h SPECT/CT images using each method. The mono-exponential function was used to estimate voxel-wise TIA maps. Regardless of the registration method used, there were regions where voxel values monotonically increased over time or TIA values are excessively high. For these regions, TIA was calculated using the trapezoidal rule as follows:

$$\tilde{A}(x, y, z) = \begin{cases} \int_0^\infty e^{\lambda_{eff}\, t} dt, & if\ \lambda_{eff} > \lambda_{phy} \\ \sum_{i=0}^{2} \frac{1}{2}\left(A_i + A_{i+1}\right)\Delta t_i + \int_{t_3}^\infty A_3 e^{-\lambda_{phy} t} dt, & if\ \lambda_{eff} < \lambda_{phy} \end{cases} \tag{6}$$

where A_i is the voxel-wise activity ($A_0 = 0$) of the *i*-th SPECT images acquired at t_i, and $\Delta t_i = t_{i+1}$-t_i. λ_{eff} represents the effective decay factor when fitting with a mono-exponential function. λ_{phy} is the physical decay factor of ^{177}Lu. The voxel-wise RMSE of curve fitting were then compared among the different methods. The voxel-wise TIA was also compared with the organ-based TIA. In brief, the volumes of interest (VOIs) were delineated on original SPECT/CT images and activity in each VOI was estimated. Then, mono-exponential fitting or trapezoidal sum was calculated as Equation (6). This comparison provided insights into which method yielded lower errors in dose assessment.

3 | RESULTS

3.1 | Overall registration performances

Table 2 compares the registration performance of Elastix, TransMorph, and CoRX-NET. The ablation study revealed that the inclusion of cross-stitch layers significantly improved SPECT/CT registration. Notably, the addition of cross-stitch layers resulted in significant enhancements in SSIM and LNCC for normalized SPECT images. CT SSIMs were increased from 0.9213 to 0.9320 for internal validation cases and from 0.9531 to 9533 for external validation cases when cross-stitch layers were embedded. SPECT LNCCs were increased from 0.7511 to 0.7765 for internal validation cases and from 0.8363 to 0.8568 for external validation cases. CoRX-NET with cross-stitch layers achieved the best metrics, except for CT SSIM in the external dataset. Elastix showed inferior SPECT registration metrics (i.e. SPECT SSIMs: 0.9593 for internal and 0.9083 for external validations, SPECT LNCC: 0.6875 for internal and 0.7372 for external validations) compared to CoRX-NET, indicating inaccuracies in conventional iterative SPECT/CT registration. The CoRX-NET with the cross-stitch layer exhibited the smallest percentage of non-positive Jacobian determinant, indicating the smallest occurrence of folding artifacts. The conventional Elastix showed relatively high ratio of non-positive Jacobian determinant (7.56% and 5.75 % for internal and external validations respectively) compared to the deep-learning based approaches (< 1 % for any validations with TransMorph and CoRX-NET). Furthermore, CoRX-NET outperformed TransMorph, despite TransMorph having more than twice the number of parameters. Remarkably, CoRX-NET completed SPECT/CT image registration in under 6 s, while Elastix took approximately 50 s on the same PC's CPU.

3.2 | Internal validations

The registration results for an internal validation case (a 54-year-old male patient with rectal neuroendocrine tumor [NET] and liver metastasis) in the coronal and axial views are presented in Figures 2 and 3, respectively. The figure shows registered SPECT/CT images in coronal and axial views, along with corresponding RGB (Red-Green-Blue)-scaled absolute normalized deformation fields and deformed grids. In the deformation fields, the red, green, and blue scales represent the extent of deformation along the *x*, *y*, and *z* axes, respectively. Conventional Elastix enabled robust CT registration but often resulted in deficiencies, especially

TABLE 2 Quantitative evaluation results of nonlinear registration for internal and external validation using Elatix, TransMorph, and CoRX.

		CT L1 loss	CT SSIM	SPECT SSIM	SPECT LNCC	% of Jacobian(pi) < 0	# of Parameter
Elastix	Internal	0.0130 (±0.0130)	0.9188 (±0.0128)	0.9593 (±0.0138)	0.6875 (±0.0927)	7.56 (±0.9693)	N/A
	External	0.0247 (±0.0253)	**0.9585 (±0.0161)**	0.9083 (±0.0593)	0.7372 (±0.1578)	5.75 (±2.27)	
TransMorph	Internal	0.0096 (±0.0008)	0.9128 (±0.0159)	0.9521 (±0.0134)	0.5999 (±0.1158)	0.0348 (±0.0239)	20,800,563
	External	0.0241 (±0.0050)	0.9519 (±0.0172)	0.9219 (±0.0603)	0.6941 (±0.1695)	0.1903 (±0.1662)	
CoRX w/o cross-stitch	Internal	0.0091 (±0.0008)	0.9213 (±0.0137)	0.9615 (±0.0137)	0.7511 (±0.1004)	0.0309 (±0.0210)	9,557,110
	External	0.0239 (±0.0050)	0.9531 (±0.0170)	0.9331 (±0.0610)	0.8363 (±0.1682)	0.1479 (±0.122)	
CoRX with cross-stitch	Internal	**0.0085 (±0.0007)**	**0.9320 (±0.0134)**	**0.9653 (±0.0113)**	**0.7765 (±0.0943)**	**0.0249 (±0.0166)**	9,557,138
	External	**0.0232 (±0.0053)**	0.9539 (±0.0176)	**0.9360 (±0.0594)**	**0.8568 (±0.1590)**	**0.1016 (±0.0758)**	

Abbreviation: SSIM, structural similarity index measure; LNCC, local normalized cross correlation.

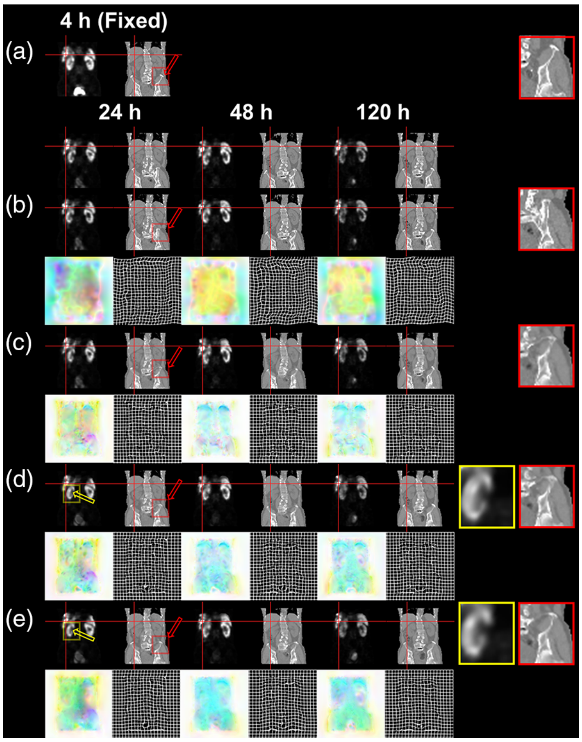

FIGURE 2 The SPECT/CT registration results and corresponding deformation fields using various methods for a case of internal validation in the coronal view are shown ((a) input images, (b) elastix, (c) TransMorph, (d) CoRX-NET without the cross-stitch layers, and (e) CoRX-NET with the cross-stitch layers). The input images, with the 4-h image as the fixed image and the 24-, 48-, and 120-h images as the moving images, are displayed in (a).

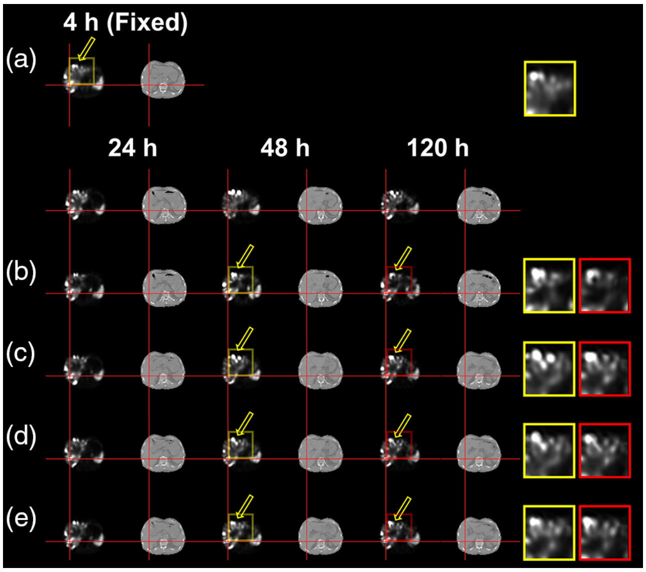

FIGURE 3 The SPECT/CT registration results and corresponding deformation fields using various methods for a case of internal validation in the axial view are shown ((a) input images, (b) elastix, (c) TransMorph, (d) CoRX-NET without the cross-stitch layers, and (e) CoRX-NET with the cross-stitch layers). The input images, with the 4-h image as the fixed image and the 24-, 48-, and 120-h images as the moving images, are displayed in (a).

in some regions such as the pelvic bone, highlighted by red arrows in the 24-h CT image (Figure 2). Deep learning-based methods, on the other hand, exhibited superior CT registration, particularly in the specified region. The red crosshairs in Figures 2 and 3 marked the midsection of a metastatic liver tumor, visible in the SPECT images. Elastix-registered images failed to align accurately with the tumor, and tumor locations across different SPECT scans registered using Elastix showed inconsistencies.

In contrast, CoRX-NET showed notable improvements in registration results, as indicated by crosshair alignment on tumor regions, especially when cross-stitch layers were employed. However, when cross-stitch layers were not included, significant folding artifacts, and mis-registration were observed in registered SPECT images, as highlighted by yellow arrows (Figure 3). These artifacts were notably suppressed by cross-stitch layers, as suggested by the sharper appearance of CoRX-NET's deformation fields without them (Figure 2). TransMorph, renowned for medical image registration, exhibited acceptable CT registration but did not achieve a similar level of performance in SPECT registration, compared to CoRX-NET.

3.3 | External validations

Figures 4 and 5 shows the registration results for an external validation case (a 49-year-old male patient with pancreatic NET and liver metastasis) in the coronal and axial views, respectively. CT images acquired at 4- and 48-h post-injection had greater slice thickness compared to those acquired at 24 h. The red crosshairs in Figures 4 and 5 were placed on a tumor region, as in the internal validation case. Distinct differences in CT registration results were not evident between Elastix and other deep learning-based registration methods. However, significant mismatches in SPECT registration were observed when Elastix was used, highlighted by red arrows in Figures 4 and 5. In contrast, CoRX-Net demonstrated superior registration results in terms of both the shape and position of tumors.

Similar to the internal validation, CoRX-Net with cross-stitch layers achieved the best registration among the deep learning-based methods. In axial SPECT imaging at 24 h post-injection, shape disparities arose with TransMorph or cross-stitch-less CoRX-Net (yellow arrows in Figure 5). Notably, CoRX-Net without cross-stitch layers yielded folding artifacts in the 48-h post-injection SPECT images (yellow arrows in Figure 4). The deformed grids of TransMorph and cross-stitch-less CoRX-Net presented in Figure 4 exhibited under-regularized behavior compared to CoRX-NET with cross-stitch layers.

In Figure 6, for an external validation case (a 54-year-old female patient with rectal NET and liver metastasis), a crosshair was placed on the liver metastasis, right

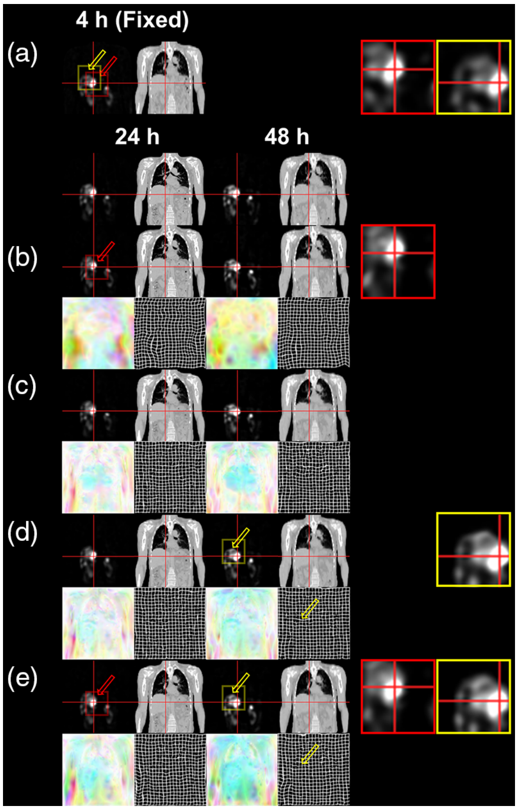

FIGURE 4 The SPECT/CT registration results and corresponding deformation fields using various methods for a case of external validation in the coronal view are shown ((a) input images, (b) elastix, (c) TransMorph, (d) CoRX-NET without the cross-stitch layers, and (e) CoRX-NET with the cross-stitch layers). The input images, with the 4-h image as the fixed image and the 24- and 48-h as the moving images, are displayed in (a).

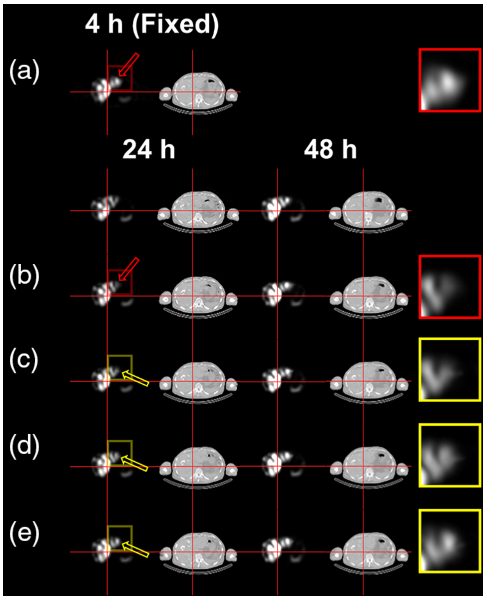

FIGURE 5 The SPECT/CT registration results and corresponding deformation fields using various methods for a case of external validation in the axial view are shown ((a) input images, (b) elastix, (c) TransMorph, (d) CoRX-NET without the cross-stitch layers, and (e) CoRX-NET with the cross-stitch layers). The input images, with the 4-h image as the fixed image and the 24- and 48-h as the moving images, are displayed in (a).

kidney, and thoracic spine metastasis to compare SPECT registration accuracy. Profiles near the liver metastasis in the fixed SPECT image and CoRX-NET-registered image were nearly identical (blue arrows), contrasting with Elastix. CoRX-NET aligned with the fixed image's activity distribution in the kidney and bone metastasis (green and red arrows), while Elastix deviated.

3.4 | TIA estimation

Table 3 shows the TIAs for kidney and tumor regions, along with their corresponding average errors and voxel-wise RMSE using the external validation dataset with 4-h SPECT images as the fixed images. CoRX-NET exhibited the smallest average error with organ-based TIA for tumor regions, while TransMorph exhibited the smallest error for kidneys. TIAs estimated using CoRX-NET were higher than other methods, regardless of the regions. However, smaller errors compared to organ-based TIA do not necessarily imply accurate registration or TIA estimations. It is more important that voxel-wise RMSEs were significantly lower for both regions when using CoRX-NET, compared to other methods.

Figure 7 shows voxel-wise TIA maps and corresponding RMSE maps for an external validation case (a 54-year-old female patient with rectal NET and liver metastasis) generated using Elastix and CoRX-NET with the cross-stitch layers. The results for TransMorph and CoRX-NET without the cross-stitch layers are shown in Figure S1. A red crosshair was placed on the edge of the liver tumor regions where high uncertainty was expected. RMSE values were lower when using CoRX-NET with the cross-stitch layers not only near the red crosshair but also within the spine.

4 | DISCUSSION

In this study, we introduced a novel unsupervised deep learning network designed for SPECT/CT registration, utilizing both CT and SPECT images as input. Through a comprehensive ablation study, we demonstrate that including cross-stitch layers significantly improves network performance. Remarkably, the integration of cross-stitch layers required only an additional 28 network parameters. As shown in Figures 2–5, the utilization of cross-stitch layers effectively suppressed folding artifacts, improving the robustness of network performance. This improvement was evident not only in internal validation but also in external validation, as summarized in Table 2. Despite the training set

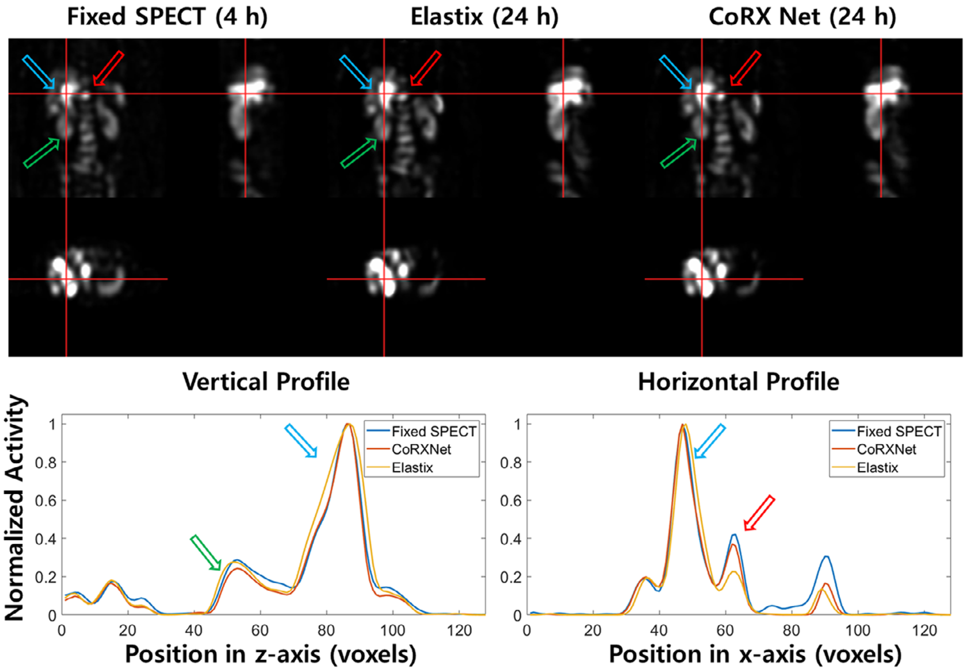

FIGURE 6 The vertical and horizontal line profiles of normalized activities in fixed and registered SPECT images acquired using Elastix and CoRX-NET (with the cross-stitch layers) for an external validation case.

being composed exclusively of male patient data, it was found that the performance differences among various methods for the external dataset, as indicated in Table S2, were independent of patients' sex.

To comprehensively evaluate SPECT registration performance, we conducted both quantitative and qualitative analyses. Quantitatively, we calculated SSIM and LNCC between the fixed and deformed normalized SPECT images. Given that SPECT images acquired at different time points exhibit varying activity levels, comparing SSIM without any normalization would be inappropriate. While normalized SPECT images were employed for SSIM estimation, it is insufficient to compare registration performance, as regional uptake may vary over time. Therefore, we also employed LNCC to account for local context variations.

These quantitative analyses were conducted based on two assumptions. First, during voxel-based dosimetry after PRRT, the process of calculating TIA following the registration of SPECT/CT implicitly assumes that the registered images share the same 3D space.[5,7] This means that, regardless of whether it's SPECT or CT, for registered images taken at different time points, assuming that the tissue or radioactivity distribution in voxels at the same position in 3D space is actually in the same 3D space. In the context of CT images, this assumption holds true given the minimal uncertainty associated with the images themselves. However, for SPECT images, characterized by substantial uncertainty, noise, partial volume effect, and inferior spatial resolution compared to CT images, the validity of this assumption becomes less clear.[5,31] Nevertheless, owing to the inherent properties of SPECT images, there exists a certain margin to assume similarity in the radioactivity distribution between the two registered SPECT images. Despite the limited spatial resolution of SPECT images, which prevents the exact determination of the actual activity distribution, there is a margin allowing us to assume that the trend of activity distribution between two registered SPECT images is similar. To clarify, although the absolute voxel-wise activity remains unknown, observing high activity in a specific area of one SPECT image implies an expectation of similarly elevated activity in the corresponding area of another SPECT image at a different time point. As a result, we assert that assuming localized correlation between the two registered SPECT images is a reasonable proposition. We hypothesized that such correlation could be utilized in SPECT image registration and employed local normalized cross-correlation in the network's training and evaluation.

Second, the need for visual verification was mentioned even after using SPECT/CT hybrid imaging for registration (i.e., applying the deformation field obtained from CT-CT registration to SPECT images).[5] It was considered that the visual verification of registration could be evaluated by visual similarity which is quantitatively evaluated using SSIM in our study. Unfortunately, SPECT images taken at different time points have different radioactivity levels as mentioned earlier. Therefore, SSIM between min-max normalized SPECT image pair was utilized in network evaluation. However, the application of this metric would be imperfect. Although it would be a valid assumption if SPECT images taken

TABLE 3 The comparison of TIA (± standard deviation) between organ-based approach and voxel-based approach with Elastix or CoRX-Net.

		Kidneys	Tumors
Organ-based	**TIA (MBq•h)**	3.47×10^5 ($\pm 1.49 \times 10^5$)	2.15×10^5 ($\pm 3.22 \times 10^5$)
Elastix	**TIA (MBq•h)**	4.07×10^5 ($\pm 1.33 \times 10^5$)	2.20×10^5 ($\pm 3.29 \times 10^5$)
	Average error (%)	27.11 (±39.68)	–13.22 (±23.56)
	Voxel-wise RMSE (kBq•h)	3.63×10^5 ($\pm 1.30 \times 10^5$)	1.76×10^5 ($\pm 1.97 \times 10^5$)
TransMorph	**TIA (MBq•h)**	3.64×10^5 ($\pm 1.14 \times 10^5$)	1.34×10^5 ($\pm 1.17 \times 10^5$)
	Average error (%)	12.63 (±28.16)	–30.27 (±31.57)
	Voxel-wise RMSE (kBq•h)	4.60×10^5 ($\pm 1.71 \times 10^5$)	2.02×10^5 ($\pm 2.28 \times 10^5$)
CoRX w/o cross-stitch	**TIA (MBq•h)**	4.29×10^5 ($\pm 1.30 \times 10^5$)	2.25×10^5 ($\pm 3.39 \times 10^5$)
	Average error (%)	35.90 (±42.23)	–10.83 (±36.61)
	Voxel-wise RMSE (kBq•h)	3.01×10^5 ($\pm 1.01 \times 10^5$)	1.23×10^5 ($\pm 1.58 \times 10^5$)
CoRX with cross-stitch	**TIA (MBq•h)**	4.35×10^5 ($\pm 1.41 \times 10^5$)	2.28×10^5 ($\pm 3.59 \times 10^5$)
	Average error (%)	39.15 (±47.46)	–6.84 (±64.22)
	Voxel-wise RMSE (kBq•h)	2.66×10^5 ($\pm 8.14 \times 10^4$)	1.18×10^5 ($\pm 1.67 \times 10^5$)

Abbreviation: TIA, time-integrated activity.

at different time points exhibit similar radioactivity concentrations across the all organ and voxel levels, this is often not the case in reality. Nevertheless, SSIM would show higher values for well-registered pairs of normalized SPECT images compared to poorly registered ones. Fundamentally, considering the absence of established criteria for quantifying the equivalence between registered SPECT images with different radioactivity concentration, these assumptions and justifications were introduced. Additionally, quantitative evaluation becomes more challenging as there is no true label for SPECT registration. Despite the limitations of the assumptions and evidence provided, we would like to emphasize that, given the task at hand, this type of quantitative analysis was the best available approach at the present time. Notably, CoRX-NET consistently exhibited superior SPECT registration, regardless of the evaluation metric.

Qualitatively, we examined activity profiles for an external validation case (Figure 6). Although CoRX-NET exhibited a profile similar to that of the fixed image, the similarity in the profile alone does not guarantee successful registration. However, at the very least, the distribution along the boundaries of high uptake regions should exhibit spatial similarity. This was the specific aspect we aimed to validate through profile comparison.

In an external validation case, we observed severe artifacts when using CoRX-NET with cross-stitch layers, as highlighted by red arrows in Figure S2. These artifacts were a result of significant discrepancies in radioactivity distributions between the fixed and moving SPECT images. Notably, activity was predominantly concentrated in the kidneys and urinary bladder, with minimal uptake observed in tumor regions in the initial SPECT images. However, substantial tumor uptake was observed in the 48-h SPECT images. When employing min-max normalization for this case, information within the low uptake regions of the fixed SPECT image was lost. Conversely, log-scaling normalization mitigated information loss and consequently improved registration performance in this case. Therefore, the network's generalization capability extended to outlier data without necessitating additional network training.

In this study, SPECT/CT registration through a neural network aimed to ensure that the overall radioactivity distribution between normalized SPECT images was similar as mentioned earlier. This process can enforce that the shape and size of tumors remain constant in SPECT images through all the time points, which may cause error in dose estimation. However, physical tumor size may not significantly reduce during the sequential scans conducted through a week. To account for this issue, upon comparing the activities within the kidneys and tumor regions before and after registration, we have confirmed negligible differences as shown in Figure S3 and Table 3.

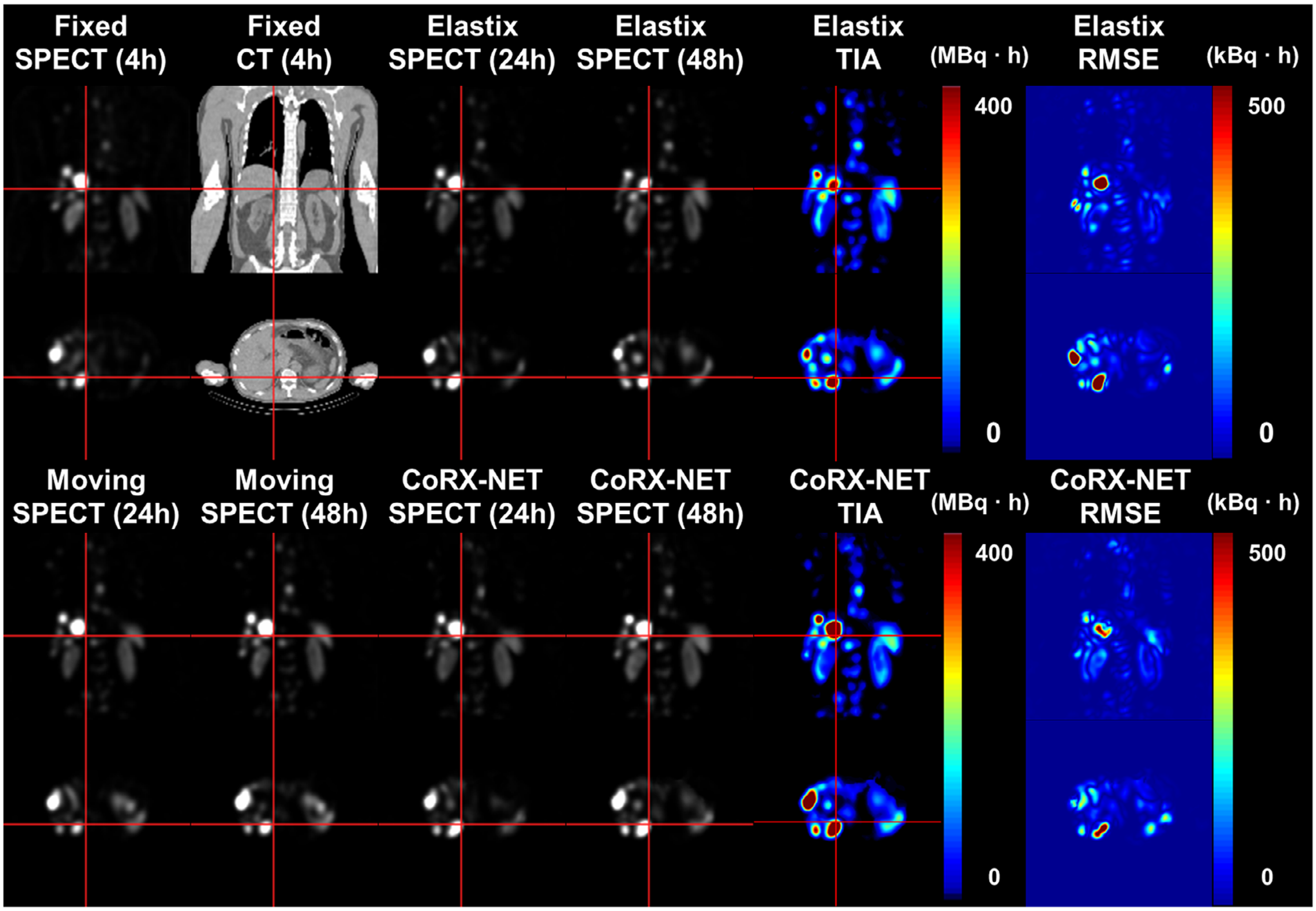

FIGURE 7 SPECT registration results, voxel-wise TIA and corresponding RMSE acquired using Elastix and CoRX-NET with the cross-stitch layers. RMSE, root mean square errors; TIA, time-integrated activity.

As known earlier, in internal dosimetry after radioisotope therapy, there are various sources of uncertainty in dose estimation, including TIA calculation, VOI delineation, and dose calculation methods.[32–35] Notably, in the context of dose assessment, it has been reported that there exists an inherent uncertainty of approximately 20% associated with estimating TIA.[36] Although not quantitatively analyzed, the registration process can also introduce uncertainty into the entire dose assessment procedure.[37] The accuracy of registration, as evidenced in this study, not only influences subsequent segmentation but also contributes to the uncertainty in voxel-wise TIA calculation.[38] As emphasized by the society of nuclear medicine and molecular imaging (SNMMI) Dosimetry Taskforce, in order for dosimetry assessment to gain broader recognition as a routine therapeutic procedure, addressing such uncertainties will be essential. From this perspective, this study has endeavored not only to enhance the accuracy of dosimetry assessment but also to contribute to the standardization of the dosimetry framework, with the ultimate goal of making it more widely applicable and reliable in clinical practice. Ultimately, the optimization of treatment by adjustment of injection dose would be achievable through accurate dosimetry with reliable SPECT/CT registration. Furthermore, establishing the relationship between absorbed dose and dose response through accurate dosimetry will play an important role in increasing treatment efficiency.

Furthermore, although not included in this study, subsequent research could explore the application of CoRX-NET for Lu-177 PSMA dosimetry as well. Additionally, in the context of performing pre-treatment dosimetry for Y-90 radioembolization using Tc-99 m SPECT/CT, registration between Tc-99 m SPECT/CT, and Y-90 PET/CT may be required, and CoRX-NET could facilitate this registration.[39] Incorporating fine-tuning techniques would serve to validate the generalizability of CoRX-NET across diverse image datasets. While the network's robustness was assessed through external validation, additional investigations, such as those mentioned above, could further confirm its robustness.

One of the limitations of CoRX-NET is its restriction on input image size, which is limited by the GPU memory. Consequently, we had to resample SPET/CT images to have 128 × 128 × 128 matrix size, which necessitated additional preprocessing with the potential risk of losing anatomical information. However, the similar approaches are also employed in conventional iterative registration methods to address memory and computation time issue. Furthermore, the final dose maps generated through the SPECT/CT registration are predominantly determined by the radioactivity distribution exhibited by the lower-resolution image, SPECT.

Another limitation of this study is that the image pairs for network training is not entirely independent since each SPECT/CT image was included several times to increase the number of training datasets as mentioned. This dependency would be a possible reason for lower CT registration metrics for external validation (i.e. CT L1 loss and SSIM). Nevertheless, CoRX-NET demonstrated a superior SPECT registration metric compared to the conventional Elastix in the external validation. This superiority extended to the voxel-wise TIA calculation, where CoRX-NET exhibited the smallest voxel-wise RMSE. In other words, for the SPECT/CT registration task, robust results were obtained even for an external dataset that was entirely independent of the training dataset.

5 | CONCLUSIONS

This study marked a significant advancement in achieving precise SPECT/CT registration through the application of an unsupervised deep learning network. Notably, this approach demonstrated its efficacy across both the internal validation and external validation datasets, surpassing the accuracy of conventional methods like Elastix or TransMorph. This enhanced registration process yielded the creation of TIA maps with reduced uncertainties, which in turn indicates the potential for more accurate dose assessments.

ACKNOWLEDGMENTS

This work was supported by grants from the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Science and ICT (grant number 2020M2D9A1093989). The funding source was not involved in the study design, collection, analysis, or interpretation. No other potential conflict of interest relevant to this article was reported.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts to disclose.

DATA AVAILABILITY STATEMENT

Data is available on request due to privacy/ethical restrictions.

REFERENCES

1. Zhang J, Song Q, Cai L, Xie Y, Chen Y. The efficacy of Lu-DOTATATE peptide receptor radionuclide therapy (PRRT) in patients with metastatic neuroendocrine tumours: a systematic review and meta-analysis. *J Cancer Res Clin*. 2020;146(6):1533-1543.
2. Sandstrom M, Garske-Roman U, Granberg D, et al. Individualized dosimetry of kidney and bone marrow in patients undergoing 177Lu-DOTA-octreotate treatment. *J Nucl Med*. 2013;54(1):33-41.
3. Hagmarker L, Svensson J, Ryden T, et al. Bone marrow absorbed doses and correlations with hematologic response during (177)Lu-DOTATATE treatments are influenced by image-based dosimetry method and presence of skeletal metastases. *J Nucl Med*. 2019;60(10):1406-1413.
4. Roth D, Gustafsson J, Warfvinge CF, et al. Dosimetric quantities in neuroendocrine tumors over treatment cycles with Lu-DOTATATE. *J Nucl Med*. 2022;63(3):399-405.
5. Dewaraja YK, Frey EC, Sgouros G, et al. MIRD pamphlet No. 23: quantitative SPECT for patient-specific 3-dimensional dosimetry in internal radionuclide therapy. *J Nucl Med*. 2012;53(8):1310-1325.
6. Ljungberg M, Celler A, Konijnenberg MW, et al. MIRD pamphlet no. 26: joint EANM/MIRD guidelines for quantitative 177Lu SPECT applied for dosimetry of radiopharmaceutical therapy. *J Nucl Med*. 2016;57(1):151-162.
7. Sarrut D, Halty A, Badel JN, Ferrer L, Bardiès M. Voxel-based multimodel fitting method for modeling time activity curves in SPECT images. *Med Phys*. 2017;44(12):6280-6288.
8. Papavasileiou P, Divoli A, Hatziioannou K, Flux GD. The importance of the accuracy of image registration of SPECT images for 3D targeted radionuclide therapy dosimetry. *Phys Med Biol*. 2007;52(24):N539-N548.
9. He B, Frey EC. The impact of 3D volume of interest definition on accuracy and precision of activity estimation in quantitative SPECT and planar processing methods. *Phys Med Biol*. 2010;55(12):3535-3544.
10. Mok GSP, Dewaraja YK. Recent advances in voxel-based targeted radionuclide therapy dosimetry. *Quant Imaging Med Surg*. 2021;11(2):483-489.
11. Lyu YQ, Chen GF, Lu ZL, Chen Y, Mok GSP. The effects of mismatch between SPECT and CT images on quantitative activity estimation—A simulation study. *Z Med Phys*. 2023;33(1):54-69.
12. Yang D, Li H, Low DA, Deasy JO, El Naqa I. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. *Phys Med Biol*. 2008;53(21):6143-6165.
13. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *Neuroimage*. 2009;45(1):S61-S72.
14. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29(1):196-205.
15. Yang X, Kwitt R, Styner M, Niethammer M. Quicksilver: fast predictive image registration—A deep learning approach. *Neuroimage*. 2017;158:378-396.
16. Lv J, Yang M, Zhang J, Wang X. Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: a feasibility study. *Br J Radiol*. 2018;91(1083):20170788.
17. Eppenhof KAJ, Pluim JPW. Pulmonary CT registration through supervised learning with convolutional neural networks. *IEEE Trans Med Imaging*. 2019;38(5):1097-1105.
18. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. *Adv Neur In*. 2015;28:2017-2025.
19. de Vos BD, Berendsen FF, Viergever MA, Staring M, Isgum I. End-to-End unsupervised deformable image registration with a convolutional neural network. *Lect Notes Comput Sc*. 2017;10553:204-212.

20. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging*. 2019;38(8):1788-1800.
21. Yoo I, Hildebrand DGC, Tobin WF, Lee WCA, Jeong WK. ssEMnet: serial-section electron microscopy image registration using a spatial transformer network with learned features. *Lect Notes Comput Sc*. 2017;10553:249-257.
22. Mocanu S, Moody AR, Khademi A. Flowreg: fast deformable unsupervised medical image registration using optical flow. *arXiv preprint*. 2021; arXiv:2101.09639.
23. O'Briain T, Uribe C, Yi KM, Teuwen J, Sechopoulos I, Bazalova-Carter M. FlowNet-PET: unsupervised learning to perform respiratory motion correction in PET imaging. *arXiv preprint*. 2022; arXiv:2205.14147.
24. Chen J, Frey EC, He Y, Segars WP, Li Y, Du Y. TransMorph: transformer for unsupervised medical image registration. *Med Image Anal*. 2022;82(5):102615.
25. Liu Z, Lin YT, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows. *2021 IEEE/CVF International Conference on Computer Vision (ICCV 2021)*. IEEE; 2021:9992-10002.
26. Liu Z, Hu H, Lin YT, et al. Swin transformer V2: scaling Up capacity and resolution. *2022 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2022)*. IEEE; 2022:11999-12009.
27. Misra I, Shrivastava A, Gupta A, Hebert M. Cross-stitch networks for multi-task learning. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*. IEEE; 2016:3994-4003.
28. Elmahdy MS, Beljaards L, Yousefi S, et al. Joint registration and segmentation via multi-task learning for adaptive radiotherapy of prostate cancer. *IEEE Access*. 2021;9:95551-95568.
29. Beljaards L, Elmahdy MS, Verbeek F, Staring M. A cross-stitch architecture for joint registration and segmentation in adaptive radiotherapy. *Proceeding of the Third Conference on Medical Imaging with Deep Learning*. PMLR; 2020:62-74.
30. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008;12(1):26-41.
31. Tran-Gia J, Salas-Ramirez M, Lassmann M. What you see is not what you get: on the accuracy of voxel-based dosimetry in molecular radiotherapy. *J Nucl Med*. 2020;61(8):1178-1186.
32. Brosch-Lenz J, Ke S, Wang H, et al. An international study of factors affecting variability of dosimetry calculations, part 2: overall variabilities in absorbed dose. *J Nucl Med*. 2023;64(7):1109-1116.
33. Gustafsson J, Brolin G, Cox M, Ljungberg M, Johansson L, Gleisner KS. Uncertainty propagation for SPECT/CT-based renal dosimetry in ^{177}Lu peptide receptor radionuclide therapy. *Phys Med Biol*. 2015;60(21):8329-8346.
34. Sarrut D, Halty A, Badel JN, Ferrer L, Bardies M. Voxel-based multimodel fitting method for modeling time activity curves in SPECT images. *Med Phys*. 2017;44(12):6280-6288.
35. Peters SMB, Mink MCT, Prive BM, et al. Optimization of the radiation dosimetry protocol in Lutetium-177-PSMA therapy: toward clinical implementation. *EJNMMI Res*. 2023;13(1):1-10.
36. Finocchiaro D, Gear JI, Fioroni F, et al. Uncertainty analysis of tumour absorbed dose calculations in molecular radiotherapy. *EJNMMI Phys*. 2020;7(1):63.
37. Mora-Ramirez E, Santoro L, Cassol E, et al. Comparison of commercial dosimetric software platforms in patients treated with ^{177} Lu-DOTATATE for peptide receptor radionuclide therapy. *Med Phys*. 2020;47(9):4602-4615.
38. Ao EC, Wu NY, Wang SJ, Song N, Mok GS. Improved dosimetry for targeted radionuclide therapy using nonrigid registration on sequential SPECT images. *Med Phys*. 2015;42(2):1060-1070.
39. Yue JT, Mauxion T, Reyes DK, et al. Comparison of quantitative Y-90 SPECT and non-time-of-flight PET imaging in post-therapy radioembolization of liver cancer. *Med Phys*. 2016;43(10):5779-5790.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kim KM, Suh M, Selvam HSMS, et al. Enhancing voxel-based dosimetry accuracy with an unsupervised deep learning approach for hybrid medical image registration. *Med Phys*. 2024;51:6432–6444. https://doi.org/10.1002/mp.17129