Structure-Preserve Expansion for Medical Image Registration with Minimal Overlap

Zaiyuan Liu[1⋆], Yongsheng Pan[1,2,3⋆], and Yong Xia[1,2,3(✉)]

[1] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[2] Ningbo Institute of Northwestern Polytechnical University, Ningbo 315048, China
[3] Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, China
yxia@nwpu.edu.cn

Abstract. Medical image registration relies on the overlapping regions between two images to calculate transformation parameters, thus posing a significant challenge for image registration with limited overlap. To overcome this challenge, this study proposes an image expansion solution by generating more overlapping regions to improve the registration performance between images with minimal overlap. As this is the first study to expand images for registration, we trained a generative network from scratch to avoid chaotic structures in the expanded regions. We proposed the Sequential Structure-Preserve Expansion (SSPE) framework to realize the expansion-based registration, where each image is present by a sliding scope and its expansion can be observed by sliding the scope. When given the current image and a sliding step, SSPE utilizes a generative network to predict the scope content of the next sliding position. Specially, we also bring in the gradient matching to maintain anatomical structures in the predicted scope. The performance of SSPE is evaluated on a public dataset of total-body CT images, which proves that our SSPE is significantly efficient in solving the registration difficulties caused by insufficient overlapping regions. The codes of our framework are made available at https://github.com/YongshengPan/Structure-Preserve-Expansion, and we will also publish software for user-friendly access and testing.

Keywords: CT · Registration· Generative adversarial network.

1 Introduction

Medical image registration has widespread applications in medical imaging [3, 11, 8], especially in fields such as radiotherapy planning, disease diagnosis, and research [14]. It aligns two or more images so that they can be compared and analyzed within the same coordinate system [23]. Most current registration algorithms rely on overlapping regions between two images to calculate transformation parameters (such as translation, rotation, etc.) [22, 1]. Some existing

⋆ Z. Liu and Y. Pan contributed equally. Corresponding author: Y. Xia.

literature and benchmarks have verified that at least 30% overlap is required to maintain measurable registration performance [4, 13, 25]. For image registration with limited overlap, it often poses a significant challenge. This often occurs when capturing images from different parts of the body, even for the same individual. As in Fig. 1, when diagnosing cardiocerebrovascular diseases, the limited apertures of computed tomography (CT)[10, 24] may only allow capture of the brain region and the chest region separately. The minimal overlap between the brain and chest regions significantly increases the difficulty of registration.

In this study, inspired by successive applications in the generation of medical images [15, 17, 16, 18], we propose an image expansion solution to overcome this challenge, which aims to generate more overlapping regions by expanding partial-scope images (as illustrated in Fig. 1), thus improving the registration performance. Although image expansion techniques have already been developed in natural image processing [20], they often fail in medical images. The typical characteristic of these natural image expansion applications is the chaotic structure in the expanded regions, which hinders the good quality of registration. Therefore, we have to develop an image expansion model from scratch for medical images to ensure that the expanded images not only have larger overlapped regions in space but also preserve the real anatomical structures.

Based on previous experiences with natural image expansion, we proposed the Sequential Structure-Preserve Expansion (SSPE) framework to realize expansion-based registration by image-to-image translation. In this framework, we assume that each image is present by a slidable scope, and its expansion can be observed by sliding the scope. When given the current image and a sliding step, SSPE utilizes a generative network to predict the scope content of the next sliding position. Specially, to maintain anatomical structures in the predicted scope, a structure maintain constraint is designed to encourage the generative network to focus more on anatomical structures. The performance of SSPE is evaluated on a public dataset of total-body CT images, where the experimental results prove that our SSPE is significantly efficient in solving the registration difficulties caused by insufficient overlapping regions.

2 Method

2.1 Problem Formulation

A total-body volume $\mathbf{X}$ could be separated as a sequence of multiple volume parts with a slide scope with center θ,

$$\mathbf{X} = \{\mathbf{X}_\theta; \theta = 1, 2, \cdots\}, \tag{1}$$

The progressive expansion from a given $\mathbf{X}_\theta$ is visualized in Fig. 2. Let $\mathbb{G}_\delta$ be a progressive model that can estimate the next scope $\mathbf{X}_{\theta+1}$ from current scope $\mathbf{X}_\theta$ of an interval distance δ from $\mathbf{X}_\theta$, namely,

$$\mathbf{X}_{\theta+\delta} = \mathbb{G}_\delta(\mathbf{X}_\theta). \tag{2}$$

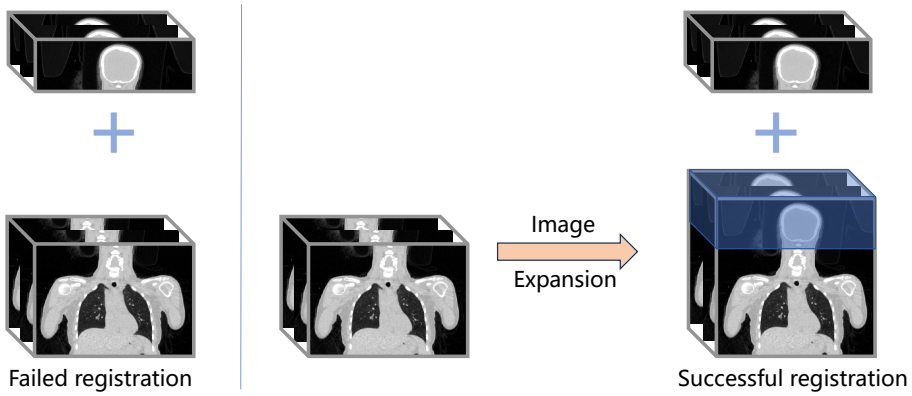

Fig. 1. Problem claim and our solution. Failed registration (left) between two parts with minimal overlap and successful registration based on expanding overlap region.

Then, any other scope $\mathbf{X}_{\theta_1}$ can be represented by gradually applying $\mathbb{G}_\delta$ to the initial scope $\mathbf{X}_{\theta_0}$ if $\theta_1 - \theta_0 = k\delta$ $(k \in N)$,

$$\mathbf{X}_{\theta_1} = \mathbb{G}_\delta(\mathbf{X}_{\theta_1-\delta}) = \mathbb{G}_\delta(\mathbb{G}_\delta(\mathbf{X}_{\theta_1-2\delta})) = \cdots = \overset{k}{\mathbb{G}}_\delta(\mathbf{X}_\theta). \tag{3}$$

2.2 Network Structure

We suppose $\mathbb{G}_\delta$ be the generative network in a generative adversarial network (GAN) [9] that is trained with the assistance of a concurrently trained discriminator network. As this is the first study to expand medical images, we make a full comparison of different backbones, including ResUNet [6], TransUNet [2], and TransResUNet [19]. These networks have a base feature dimension of 16 and a downsampling depth of 4.

ResUNet Combining the strengths of U-Net and ResNet, ResUNet offers improved gradient flow and feature learning through residual connections, which help mitigate the vanishing gradient problem and enhance training stability [12]. Its symmetric encoder-decoder structure with skip connections preserves spatial details, making it ideal for tasks like medical image segmentation.

TransUNet In image extrapolation, the Transformer [21] enhances the model's ability to capture long-range dependencies and global contextual relationship across an image. By self-attention, it can focus on distant regions of the image, helping generate coherent and consistent structures during extrapolation. This is useful for maintaining anatomical consistency in medical imaging tasks, where capturing complex spatial relationships is critical for accurate tissue generation. Integrating the Transformer blocks within the U-Net/ResUNet structure may

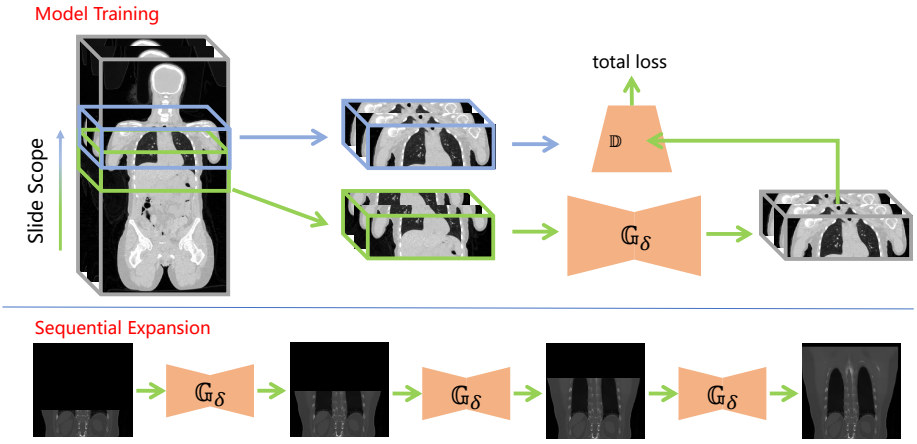

Fig. 2. Framework to sequentially expand medical images for registration with minimal-overlap regions.

enhance the model's ability to capture long-range dependencies, which are denoted TransUNet/TransResUNet.

Adversarial Network The adversarial network is a sequence of convolutional blocks, progressively increasing the feature dimensions while reducing spatial resolution. The network starts with a base feature dimension of 16 and a downsampling depth of 4. A reflective padding strategy is used to minimize boundary artifacts, and the final output is a single-channel volumetric map rather than a single scalar value, allowing for a more detailed and localized comparison with the ground truth. The discriminator is optimized using the Adam optimizer with a learning rate of 1×10^{-3}.

2.3 Training

To reduce memory consumption, we slide a scope of $96 \times 256 \times 256$ to extract a sequence of volume parts from each CT volume. To ensure that the model can access comprehensive data while preventing boundary overflow after shifting by a gap, the starting coordinate θ is randomly selected from the range $[0, D-96-\delta)$, where D is the tomogram count of this volume. The first volume part is defined as $X_\theta = X[\theta - 48 : \theta + 48, 0 : 256, 0 : 256]$, by sliding a distance of $\delta = 46$, the second volume part is obtained as $X_{\theta+\delta} = X[\theta - 2 : \theta + 94, 0 : 256, 0 : 256]$. The first volume part X_θ is fed into the generative network $\mathbb{G}_\delta$, producing the output $\hat{X}_{\theta+\delta}$ through model training. Subsequently, both $\hat{X}_{\theta+\delta}$ and $X_{\theta+\delta}$ are fed into the discriminator $\mathbb{D}$ for further processing and evaluation.

The model is trained via a combination of reconstruction loss and adversarial loss. The reconstruction loss optimizes for coarse image agreement and is imple-

mented as an ℓ_1 loss [5] imposed on the full output of $\mathbb{G}_\delta$. The full equation is below:

$$\mathcal{L}_{rec} = \mathbb{E}_{(X_\theta, X_{\theta+\delta})\sim p_{\text{data}}} [\|X_{\theta+\delta} - \mathbb{G}_\delta(X_\theta)\|_1] \tag{4}$$

For the adversarial loss, which refines the coarse prediction, we use the equation below:

$$\mathcal{L}_{adv} = \mathbb{E}_{(X_\theta, X_{\theta+\delta})\sim p_{\text{data}}} \left[\frac{1}{2} (\|\mathbb{D}(\mathbb{G}_\delta(X_\theta)) - 1\|_1 + \|\mathbb{D}(X_{\theta+\delta})\|_1)\right] \tag{5}$$

In addition to the above losses, we also introduce a gradient matching loss [7] to improve the fine-grained alignment of feature representations between the generator's output and the ground truth. This loss encourages the generator to produce outputs whose gradient maps match those of the real data. The gradient matching loss is calculated as:

$$\mathcal{L}_{gm} = \mathbb{E}_{x\sim p_{\text{data}}} \left[\|f_\Delta(X_\theta) - f_\Delta(\mathbb{G}_\delta(X_{\theta+\delta}))\|_1^1\right] \tag{6}$$

where $f_\Delta(\cdot)$ represents the gradient map extracted from x by a Sobel operator.

The total loss for training the generative network is:

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \mathcal{L}_{adv} + \mathcal{L}_{gm} \tag{7}$$

As X_θ and $X_{\theta+\delta}$ has an overlap in the middle, we assign a different weight $\lambda = 4$ for non-overlapping portions, to draw the model's attention towards generating the non-overlapping sections. We denote the models with gradient matching loss and different weight λResGAN, λTransGAN, and λTransResGAN to distinguish with the original models ResGAN, TransGAN, TransResGAN.

3 Experiment

Dataset. We download the Healthy-Total-Body-CTs dataset from The Cancer Imaging Archive (TCIA) for our model training and evaluation. This dataset provides low-dose whole-body CT images from 30 healthy adults (age $\geq$ 18) imaged on the uEXPLORER total-body PET/CT system at UC Davis. Fifteen participants were scanned at three timepoints (0, 90, and 180 minutes post-PET tracer injection), and the remaining 15 at six timepoints (adding 360, 540, and 720 minutes). CT scans at 90 minutes used 140 kVp and 50 mAs, while other timepoints used 5 mAs. Notably, CT images were reconstructed into a $512 \times 512 \times 828$ image matrix with $0.9766 \times 0.9766 \times 2.344\,\text{mm}^3$ voxel size. Access requires a TCIA Restricted License Agreement due to potential risks of facial reconstruction. We randomly select 120 subjects for training models and the rest remain for evaluation.

Our model is implemented in Python with PyTorch and CUDA12. The generative network and adversarial network are trained jointly using the Adam optimizer with a learning rate of 10^{-3}. During training, we randomly crop a volume part of each CT scan and a slide scope with a down distance of 46 layers,

ensuring that they do not exceed the boundaries. The training was conducted on an NVIDIA 4070 Ti SUPER GPU, with 1400 training iterations for all six models. The input size is (96, 256, 256), with a batch size of 1 and a memory requirement of 15.55 GB.

Evaluation Metrics. We evaluate the performance of our 3D thoracic CT extrapolation models using four key metrics: mean absolute error (MAE), peak signal-to-noise ratio (PSNR), root mean square error (RMSE), and structural similarity index (SSIM). MAE quantifies the average absolute pixel-wise difference, reflecting overall accuracy. PSNR assesses image fidelity by comparing signal strength to noise, with higher values indicating better quality. RMSE emphasizes larger errors by computing the square root of the mean squared differences. SSIM measures structural and perceptual similarity to the ground truth, with values closer to 1 indicating higher visual consistency. Together, these metrics provide a comprehensive evaluation of both quantitative accuracy and perceptual quality.

4 Result Analysis

As we aim to achieve registration among medical images with a minimal-overlap region via image expansion, our result analysis includes the effect of image expansion and minimal-overlap registration.

4.1 Image Expansion

The performance of six models for 3D CT expansion is compared in Table 1, where the metrics are reported as mean ± standard deviation. Three conclusions can be drawn from this table. First, ResGAN, TransGAN, and TransResGAN achieve similar metrics while λResGAN, λTransGAN, and λTransResGAN achieve similar metrics. This suggests that structure of the generative network is not a major impact factor. Second, adding gradient matching loss and assigning different weights to the overlap and non-overlap partials leads to an obvious discrepancy, demonstrating that adding gradient matching loss and assigning a different weights can help preserve the tissue structure and draw the model's attention towards generating the non-overlapping sections. Third, λResGAN achieves the lowest MAE (4.13 ± 2.09) and RMSE (8.88 ± 3.83), with a high SSIM (0.88 ± 0.04), while λTransGAN records the highest PSNR (30.19 ± 4.70), suggesting that a generative network with more complex structures cannot be beneficial to the generated results.

Fig. 3 presents a qualitative comparison of six models for 3D CT expansion, showcasing sagittal and coronal views of the chest region. The "GT" column displays the ground truth (complete CT scan), while the other columns show the results of applying ResGAN, TransGAN, TransResGAN and λResGAN, λTransGAN, λTransResGAN for four successive expansions. Two observations can be seen in Fig. 3. First, the volumes generated by ResGAN, TransGAN,

Table 1. Performance comparison of all models. All scores are reported as mean ± standard deviation.

Model	MAE ↓	PSNR ↑	RMSE ↓	SSIM ↑
ResGAN	8.23 ± 8.03	26.87 ± 5.68	14.50 ± 11.09	0.81 ± 0.11
TransGAN	7.74 ± 8.09	27.36 ± 5.88	14.07 ± 11.59	0.82 ± 0.12
TransResGAN	8.01 ± 7.69	26.77 ± 5.31	14.42 ± 10.91	0.81 ± 0.11
λResGAN	**4.13** ± 2.09	30.18 ± 4.52	**8.88** ± 3.83	**0.88** ± 0.04
λTransGAN	4.27 ± 2.27	**30.19** ± 4.70	8.96 ± 3.97	**0.88** ± 0.04
λTransResGAN	4.76 ± 2.26	29.57 ± 4.61	9.56 ± 4.05	0.87 ± 0.04

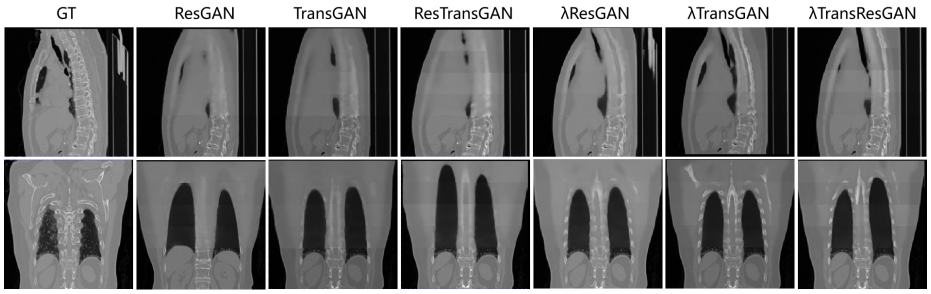

Fig. 3. Comparison of images expanded by six generative models. The top and bottom rows are the sagittal and coronal views of the same individual, with Ground Truth (GT) on the left. Our specialized λResGAN, λTransGAN, and λTransResGAN generate images with tissue details closer to GT than ResGAN, TransGAN, TransResGAN, confirming its superiority.

TransResGAN without gradient matching are more blurred than those generated by λResGAN, λTransGAN, λTransResGAN, again demonstrating the ability of gradient matching to preserve tissue structure. Second, as the number of expansion iterations increases, distortions also accumulate. λResGAN appears to preserve the most realistic structures. This is reasonable, as the input at each iteration is the output of the previous one, leading to residual accumulation.

4.2 Minimal-Overlap Registration

As described in the above subsection, our specialized λResGAN, λTransGAN, and λTransResGAN achieved similar performance in image expansion, where λResGAN shows a little better performance. Therefore, we select λResGAN as a representative to inspect the registration ability of image expansion. We evaluate its performance under two different conditions that frequently appear in medical image analysis, including (1) same person registration, (2) different person registration. The results of these registration conditions are illustrated in Fig. 4, where we consider only the linear registration for simplification and stability. The following conclusions can be observed from Fig. 4.

(1) Based on increased overlap area from image expansion, all these four registration conditions could be well-achieved while failures are reduced without increased overlap. This demonstrates the correctness and effectiveness of the technique line, image expansion for minimal-overlap registration.
(2) The images achieved excellent alignment with imperceptible position or rotation shift for same person registration, but remained a little perceptible shift for different person registration. This is mainly caused by the individual difference that cannot be overcome by linear registration. Individual difference regions are almost evenly distributed in symmetric directions, which suggests that different individuals are well aligned by their centers.
(3) Besides the sample shown in Fig. 4, we also tested on some other subjects and have similar observations as above. This demonstrates the robustness and generalizability of our approach in handling diverse imaging scenarios.

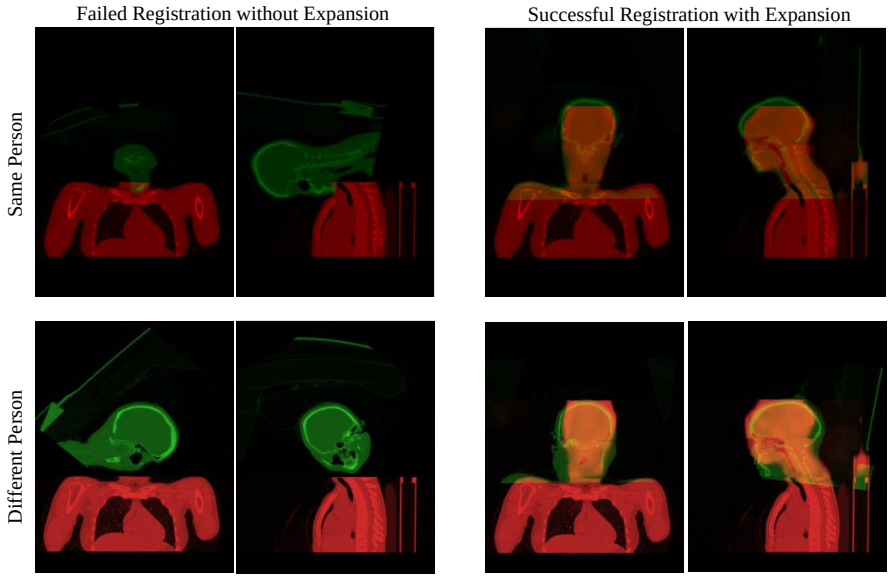

Fig. 4. Failed registration(left) and successful registration after image expansion (right).

5 Conclusion

In this study, we proposed an image expansion solution, Sequential Structure-Preserve Expansion (SSPE), to overcome the challenge of registration with minimal overlap. It generated more overlapping regions by expanding the partial-scope images with minimal overlap regions, thus improving the registration performance. We developed and trained image expansion models from scratch for

medical images preserving real anatomical structures to ensure that the expanded images provide good quality for registration. Our future work will expand to multiple modalities and publish user-friendly software.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 6240012686, 62171377, and 92470101, in part by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang, China, under Grant 2025C01201(SD2), in part by the Ningbo Clinical Research Center for Medical Imaging under Grant 2021L003 (Open Project 2022LYKFZD06),in part by Guangdong Basic and Applied Basic Research Foundation under Grant 025A1515011592, the in part by the Shenzhen Science and Technology Program under Grant JCYJ20220530161616036, and in part by the Fundamental Research Funds for the Central Universities under Grant D5000230376.

Disclosure of Interests. The authors declare no relevant competing interests.

References

1. Chand, K., Fritsch, T., Oster, S., Ulbricht, A., Bruno, G.: Review on image registration methods for the quality control in additive manufacturing. Progress in Additive Manufacturing pp. 1–27 (2025)
2. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
3. Chen, J., Liu, Y., Wei, S., Bian, Z., Subramanian, S., Carass, A., Prince, J.L., Du, Y.: A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. Medical Image Analysis p. 103385 (2024)
4. Choi, S., Zhou, Q.Y., Koltun, V.: Robust reconstruction of indoor scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5556–5565 (2015)
5. Cohen, J.P., Luck, M., Honari, S.: Distribution matching losses can hallucinate features in medical image translation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I. pp. 529–536. Springer (2018)
6. Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C.: Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing **162**, 94–114 (2020)
7. Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M., Felsberg, M.: Roma: Robust dense feature matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19790–19800 (2024)
8. Gong, Y., Liu, H., Li, L., Tian, J., Li, H.: Deep learning-based medical image registration algorithm: Enhancing accuracy with dense connections and channel attention mechanisms. Journal of Theory and Practice of Engineering Science **4**(02), 1–7 (2024)

9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
10. Gupta, H., Jin, K.H., Nguyen, H.Q., McCann, M.T., Unser, M.: Cnn-based projected gradient descent for consistent ct image reconstruction. IEEE transactions on medical imaging **37**(6), 1440–1453 (2018)
11. Jena, R., Sethi, D., Chaudhari, P., Gee, J.: Deep learning in medical image registration: Magic or mirage? Advances in Neural Information Processing Systems **37**, 108331–108353 (2025)
12. Jha, D., Smedsrud, P.H., Johansen, D., De Lange, T., Johansen, H.D., Halvorsen, P., Riegler, M.A.: A comprehensive study on colorectal polyp segmentation with resunet++, conditional random field and test-time augmentation. IEEE journal of biomedical and health informatics **25**(6), 2029–2040 (2021)
13. Khoury, M., Zhou, Q.Y., Koltun, V.: Learning compact geometric features. In: Proceedings of the IEEE international conference on computer vision. pp. 153–161 (2017)
14. Maurer, C.R., Fitzpatrick, J.M.: A review of medical image registration. Interactive image-guided neurosurgery **1**, 17–44 (1993)
15. Pan, Y., Liu, M., Lian, C., Xia, Y., Shen, D.: Spatially-constrained fisher representation for brain disease identification with incomplete multi-modal neuroimages. IEEE transactions on medical imaging **39**(9), 2965–2975 (2020)
16. Pan, Y., Liu, M., Xia, Y., Shen, D.: Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. IEEE transactions on pattern analysis and machine intelligence **44**(10), 6839–6853 (2021)
17. Pan, Y., Xia, Y.: Ultimate reconstruction: understand your bones from orthogonal views. In: 2021 IEEE 18th international symposium on biomedical imaging (ISBI). pp. 1155–1158. IEEE (2021)
18. Pan, Y., Ye, Y., Zhang, Y., Xia, Y., Shen, D.: Draw sketch, draw flesh: Whole-body computed tomography from any x-ray views. International Journal of Computer Vision pp. 1–22 (2024)
19. Reza, S., Amin, O.B., Hashem, M.: Transresunet: Improving u-net architecture for robust lungs segmentation in chest x-rays. In: 2020 IEEE Region 10 Symposium (TENSYMP). pp. 1592–1595. IEEE (2020)
20. Teterwak, P., Sarna, A., Krishnan, D., Maschinot, A., Belanger, D., Liu, C., Freeman, W.T.: Boundless: Generative adversarial networks for image extension. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10521–10530 (2019)
21. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
22. Velesaca, H.O., Bastidas, G., Rouhani, M., Sappa, A.D.: Multimodal image registration techniques: a comprehensive survey. Multimedia Tools and Applications **83**(23), 63919–63947 (2024)
23. Wang, H., Ni, D., Wang, Y.: Recursive deformable pyramid network for unsupervised medical image registration. IEEE Transactions on Medical Imaging **43**(6), 2229–2240 (2024)
24. Wang, T., Xia, W., Lu, J., Zhang, Y.: A review of deep learning ct reconstruction from incomplete projection data. IEEE Transactions on Radiation and Plasma Medical Sciences **8**(2), 138–152 (2023)

25. Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T.: 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1802–1811 (2017)