StructuRegNet: Structure-Guided Multimodal 2D-3D Registration

Amaury Leroy[1,2,3(✉)], Alexandre Cafaro[1,2,3], Grégoire Gessain[4],
Anne Champagnac[5], Vincent Grégoire[3], Eric Deutsch[2], Vincent Lepetit[6],
and Nikos Paragios[1]

[1] Therapanacea, Paris, France
`a.leroy@therapanacea.eu`
[2] Gustave Roussy, Inserm 1030, Paris-Saclay University, Villejuif, France
[3] Department of Radiation Oncology, Centre Léon Bérard, Lyon, France
[4] Department of Pathology, Gustave Roussy, Villejuif, France
[5] Department of Pathology, Centre Léon Bérard, Lyon, France
[6] Ecole des Ponts ParisTech, Marne-la-Vallée, France

Abstract. Multimodal 2D-3D co-registration is a challenging problem with numerous clinical applications, including improved diagnosis, radiation therapy, or interventional radiology. In this paper, we present StructuRegNet, a deep-learning framework that addresses this problem with three novel contributions. First, we combine a 2D-3D deformable registration network with an adversarial modality translation module, allowing each block to benefit from the signal of the other. Second, we solve the initialization challenge for 2D-3D registration by leveraging tissue structure through cascaded rigid areas guidance and distance field regularization. Third, StructuRegNet handles out-of-plane deformation without requiring any 3D reconstruction thanks to a recursive plane selection. We evaluate the quantitative performance of StructuRegNet for head and neck cancer between 3D CT scans and 2D histopathological slides, enabling pixel-wise mapping of low-quality radiologic imaging to gold-standard tumor extent and bringing biological insights toward homogenized clinical guidelines. Additionally, our method can be used in radiation therapy by mapping 3D planning CT into the 2D MR frame of the treatment day for accurate positioning and dose delivery. Our framework demonstrates superior results to traditional methods for both applications. It is versatile to different locations or magnitudes of deformation and can serve as a backbone for any relevant clinical context.

Keywords: Multimodal · Registration · 2D-3D · Histopathology · Radiology

1 Introduction

2D-3D registration refers to the highly challenging process of aligning an input 2D image to its corresponding slice inside a given 3D volume [4]. It has received growing

---

attention in medical imaging due to the various contexts where it applies, like image fusion between 2D real-time acquisitions and either pre-operative 3D images for guided interventions or reference planning volumes for patient positioning in radiation therapy (RT). Another important task is the volumetric reconstruction of a sequence of misaligned slices *ex vivo*, enabling multimodal comparison toward improved diagnosis. In this respect, overlaying 3D radiology and 2D histology could significantly enhance radiologists' understanding of the links between tissue characteristics and radiologic signals [9]. Indeed, MRI or CT scans are the baseline source of information for cancer treatment but fail to provide an accurate assessment of disease proliferation, leading to high variability in tumor detection [5,13,17]. On the other hand, high-resolution digitized histopathology, called Whole Slide Imaging (WSI), provides cell-level information on the tumor environment from the surgically resected specimens. However, the registration process is substantially difficult due to the visual characteristics, resolution scale, and dimensional differences between the two modalities. In addition, histological preparation involves tissue fixation and slicing, leading to severe collapse and out-of-plane deformations. (Semi-)automated methods have been developed to avoid time-consuming and biased manual mapping, including protocols with 3D mold or landmarks [10,22], volume reconstruction to perform 3D registration [2,18,19,23], or optimization algorithms for direct multimodal comparison [3,15]. More recently, deep learning (DL) has been introduced but is limited to 2D/2D and requires prior plane selection [20]. On the other hand, successful DL methods have been proposed to address the 2D/3D mapping problem for other medical modalities [6,8,16,21]. However, given the extreme deformation that the tissue undergoes during the histological process, additional guidance is needed. One promising solution is to rely on rigid structures that are supposedly more robust during the preparation. Structural information to guide image registration has been studied with the help of segmentations into the training loop [11], or by learning new image representations for refined mapping [12].

In this paper, we propose to leverage the structural features of tissue and more particularly the rigid areas to guide the registration process with two distinct contributions: (1) a cascaded rigid alignment driven by stiff regions and coupled with recursive plane selection, and (2) an improved 2D/3D deformable motion model with distance field regularization to handle out-of-plane deformation. To our knowledge, no previous study proposed 2D/3D registration combined with structure awareness. We also use the CycleGAN for image translation and direct monomodal signal comparison [25]. Like [14,24], we combine registration with modality translation and integrate the two aforementioned components. We demonstrate superior quantitative results for Head and Neck (H&N) 3D CT and 2D WSIs than traditional approaches failing due to the histological constraints. In addition, we show that StructuRegNet performs better than the state-of-the-art model from [14] on 3D CT/2D MR for the pelvis in RT.

2   Methods

2.1   Histology-to-CT Modality Translation

Our three-step structure-aware pipeline is thoroughly detailed in Fig. 1. For clarity, we focus on the radiology-histology application but the pipeline is versatile
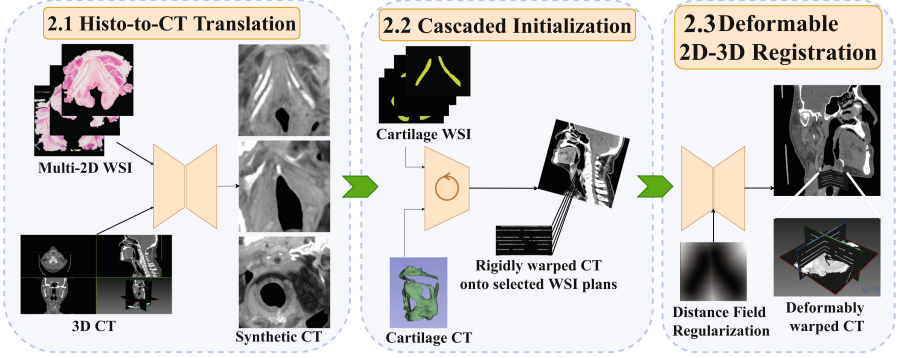
Fig. 1. Overview of StructuRegNet, where the WSIs are first translated into synthetic CTs (2.1) before getting matched with the most similar CT slices thanks to a cascaded structure-aware plane selection (2.2). Finally, the rigid transformation is refined through a deformable network to handle out-of-plane distortions (2.3).

to any 2D/3D setting. The modality transfer is a 2D image-to-image translation problem defined as follows: Given a sequence of n slices $H = \{h_1, ..., h_n\}$ and a volume considered as a full stack of m axial slices $CT = \{ct_1, ..., ct_m\}$, we build a CycleGAN with two generators and two discriminators $G_{H \to CT}, G_{CT \to H}, D_H$ and D_{CT}. With a symmetric situation for $G_{CT \to H}$, $G_{H \to CT}$ outputs a synthetic CT image, which is then processed by D_{CT} along with randomly sampled original input slices with an associated adversarial loss L_{adv}. The cyclical pattern lies in the similarity between the original images and the reconstructed samples $G_{CT \to H} \circ G_{H \to CT}(h_i)$ through a pixel-wise cycle loss L_{cyc}. Finally, we employ two additional metrics: an identity loss L_{Id} to encourage modality-specific feature representation when considering h_i being the input for $G_{CT \to H}$ with an expected identity synthesis; and a structure consistency MIND loss from [7] to ensure style transfer without content alteration. These losses are the classical implementations for CycleGAN and are detailed in the supplementary material.

2.2   Recursive Cascaded Plane Selection

We replace the volume reconstruction step with a recursive dual model. We first rotate and translate the 3D CT to match the stack of slices H, which is crucial as an initialization step to help the 2D-3D network to focus on small out-of-plane deformations and avoid local minima. Then, we perform a precise plane selection and solve the spacing gap by adjusting the z-position of each slice to its most similar CT section. These two steps are performed iteratively until convergence, with a recursive algorithm to reduce computational cost (Fig. 2).

For rigid initialization, the hypothesis is that the histological specimen is cut with an unknown spacing and angle, but the latter is supposed constant between WSIs. A rigid alignment is thus sufficient to reorient moving CT onto fixed H. Based on a theoretical axial slice sequence $Z = (Z_1, ..., Z_m)$, we define
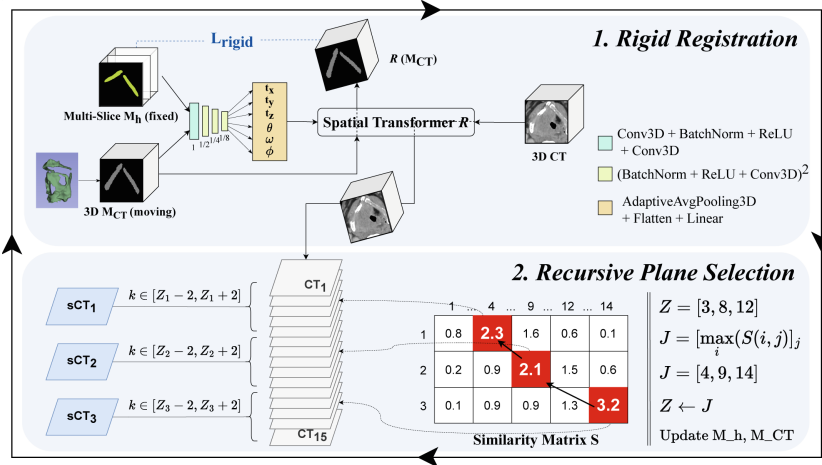
Fig. 2. Cascaded alignment through rigid structure-aware warping followed by recursive plane selection. The deformed CT from 1. is the input for 2., along with sCT and slice sequence Z. The updated Z from 2. is applied to M_h while the rigid deformation is applied to M_{CT} so that new inputs can feed 1. again as iterative refining.

H as a sparse 3D volume the same size as CT, filled in with h_i at $z = Z_i$ and zeros elsewhere (the same applies for the corresponding sCT from the previous module). Because soft tissues undergo too large out-of-plane deformations, we leverage the rigid structures which are supposed not to be distorted or shrunk during the histological process. We extract their segmentation masks M_{ct}, M_h for both modalities (see preprocessing in Sect. 3), concatenate and fed them into an encoder followed by a fully connected layer that outputs six transformation parameters (3 rotations, 3 translations). A differentiable spatial transform R finally warps M_{ct} for similarity optimization with M_h. Similarly to [14], we adopt a loss L_{rigid} masked on empty slices to avoid the introduction of noise at slices within the gradient where no data is provided, and directly train on the Dice Similarity Coefficient (DSC) between rigid areas:

$$L_{\mathrm{rigid}}(M_h, R(M_h, M_{ct})) = \sum_{i \in [1,m]} \mathrm{DSC}(M_{h_{Z_i}}, R(M_h, M_{ct})(M_{ct_{Z_i}})) . \quad (1)$$

Additionally, R also warps CT without gradient backpropagation and is the input with sCT for plane selection. We then introduce a sequence alignment problem, the objective being to update the slice sequence Z of sCT by mapping it to a corresponding sequence J of 2D images from CT. We define S a similarity matrix, where $S(i,j)$ is the total similarity (measured with MI) when mapping $sct_{Z_1}, ..., sct_{Z_i}$ with $ct_{J_1}, ..., ct_{J_j}$ for $i \in [1,n], j \in [1,m]$:

$$S(i,j) = \begin{cases} MI(sct_{Z_i}, ct_{J_j}) & \text{if } i = 1 \\ \max_k S(i-1, k) + MI(sct_{Z_i}, ct_{J_j}) & \text{else} \end{cases}, \quad (2)$$

which means that each row of S will be filled by computing the sum of the MI for the corresponding column j and the maximum similarity from the last row. Like any dynamic programming method, we want to find the optimal sequence J by following the backward path of S building. To do so, we retrieve the new index j that yielded the maximized similarity for each step $J = [\max_i(S(i,j))]_j$, and we update $Z \leftarrow J$ accordingly. In addition, the J sequence cannot be too different from Z as it would induce overlap between ordered WSIs. We thus constrained the possible matching values with $k \in [Z_i - 2, Z_i + 2]$. Based on these rigid registration and plane selection blocks, we build a cascaded module to iteratively refine the alignment where the intermediate warping becomes the new input. We defined the number of iterations as a hyperparameter to reach a good balance between computational time and similarity maximization. This dual model is crucial for initialization but does not take into account out-of-plane deformations and a perfect alignment is not accessible yet. The deformable framework bridges this gap by focusing on irregular displacements caused by tissue manipulation and refining the rigid warping.

2.3   Deformable 2D-3D Registration

Given one fixed multi-slice sCT' and a moving rigidly warped $R(CT)$ from the previous module, we adopt an architecture close to Voxelmorph [1]. Still, the rigidly warped $CT' = R(CT)$ and the plane-adjusted sparse sCT' are fed through two different encoders for independent feature extraction. The architecture is depicted in Fig. 3. Both latent representations are element-wise subtracted. A decoder is connected to both encoders and generates a displacement field Φ the same size as input images but with (x, y, z)-channels corresponding to the displacement in each spatial coordinate. A differentiable sampler D warps CT' in a deformable setting, which is then compared to sCT' through a masked Normalized Cross-Correlation (NCC) loss L_{defo}:

$$L_{\text{defo}}(sCT', D(sCT', CT')) = \sum_{i \in [1,m]} \text{NCC}(sCT'_{Z_i}, D(sCT', CT')(CT'_{Z_i})) . \quad (3)$$

Finally, we add two sources of regularization. Soft tissues away from bones and cartilage are more subject to shrinkage or disruption, so we harness the information from the cartilage segmentation mask of CT to generate a distance transform map Δ defined as $\Delta(\mathbf{v}) = \min_{\mathbf{m} \in M_{CT}} ||\mathbf{v} - \mathbf{m}||_2$. It maps each voxel $\mathbf{v}$ of CT to its distance with the closest point $\mathbf{m}$ to the rigid area M_{CT}. We can then control the displacement field, with close tissue being more highly constrained than isolated areas: $\Phi' = \Phi \odot (\Delta + \epsilon)$, where $\odot$ is the Hadamard product and ϵ is a hyperparameter matrix allowing small displacement even for cartilage areas for which distance transform is null. A second regularization takes the form of a loss $L_{\text{regu}}(\Phi') = \sum_{\mathbf{v} \in \mathbb{R}^3} ||\nabla \Phi'(\mathbf{v})||^2$ on the volume to constrain spatial gradients and thus encourage smooth deformation, which is essential for empty slices which are excluded from L_{defo}. The total loss is a weighted sum of L_{defo} and L_{regu}.
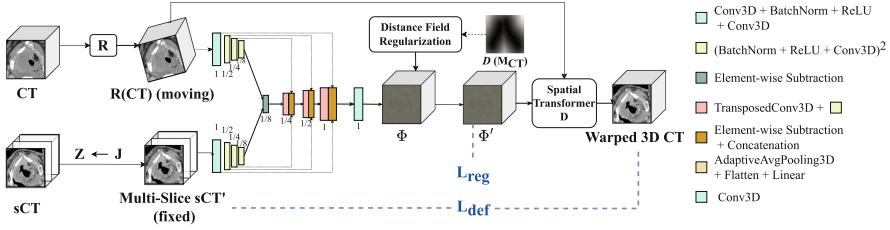
Fig. 3. Deformable 2D-3D registration pipeline, made of two encoders and a shared decoder, with regularization applied on the displacement field Φ thanks to the distance map from CT.

3   Experiments

Dataset and Preprocessing. Our clinical dataset consists of 108 patients for whom were acquired both a pre-operative H&N CT scan and 4 to 11 WSIs after laryngectomy (with a total amount of 849 WSIs). The theoretical spacing between each slice is 5 mm, and the typical pixel size before downsampling is $100K \times 100K$. Two expert radiation oncologists on CT delineated both the thyroid and cricoid cartilages for structure awareness and the Gross Tumor Volume (GTV) for clinical validation, while two expert pathologists did the same on WSIs. They then meet and agreed to place 6 landmarks for each slice at important locations (not used for training). We ended up with images of size 256×256 ($\times 64$ for 3D CT) of 1 mm isotropic grid space. We split the dataset patient-wise into three groups for training (64), validation (20), and testing (24). To demonstrate the performance of our model on another application, we also retrieved the datasets from [14] for pelvis 3D CT/2D MR. It is made of 451 pairs between CT and TrueFISP sequences, and 217 other pairs between CT and T2 sequences. We guided the registration thanks to the rigid left/right femoral heads and computed similarity metrics on the 7 additional organs at risk (anal canal, bladder, rectum, penile bulb, seminal vesicle, and prostate). All masks were provided by the authors and were originally segmented by internal experts.

Hyperparameters. We drew our code from CycleGAN and Voxelmorph implementations with modifications explained above, and we thank the authors of MSV-RegSynNet for making their code and data available to us [1,14,25]. A detailed description of architectures and hyperparameters can be found in the supplementary material. We implemented our model with Pytorch1.13 framework and trained for 600 (800 for MR/CT) epochs with a batch size of 8 (4 for MR/CT) patients parallelized over 4 NVIDIA GTX 1080 Tis.

Evaluation. We benchmarked our method against three baselines: First, to assess the benefit of modality translation over the multimodal loss, we re-used the original 3D VoxelMorph model with MIND as a multimodal metric for optimization. We also modified this approach by masking the loss function to account for the 2D-3D setting. Next, we implemented the modality translation-based MSV-

RegSyn-Net and modified it for our application to measure the importance of joint structure-aware initialization and regularization. Finally, to differentiate the latter contributions, we tested two ablation studies: without the cascaded rigid mapping or without the distance field control. According to the MR/CT application in RT, we compared our model against the state-of-the-art results of MSV-RegSynNet which were computed on the same dataset.

4   Results

4.1   Modality Translation

Three samples from the test set are displayed in Fig. 1. From a qualitative perspective, the densities of the different tissues are well reconstructed, with rigid structures like cartilage being lighter than soft tissues or tumors. The general shape of the larynx also complies with the original radiologic images. We achieve a mean Structural Similarity (SSIM) index of 0.76/1 between both modalities, demonstrating the strong synthesis capabilities of our network compared to MSV-RegSynNet and our ablative study without initialization process, with an SSIM of 0.72 (respectively 0.69). Therefore, the cascaded rigid initialization is crucial and helps the modality translation module in getting more similar pairs of images for eased synthesis on the next pass.

4.2   Registration

We present visual results in Fig. 4. The initialization enables an accurate plane selection as proved by the similar shape of cartilages in (b). Even for some severe
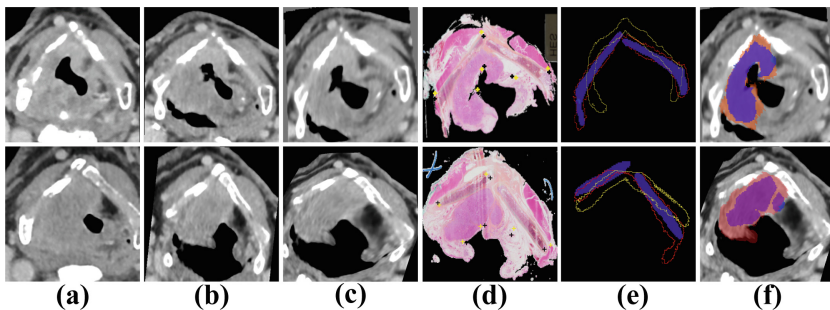

(a)   **(b)**   **(c)**   **(d)**   **(e)**   **(f)**

Fig. 4. Visual results for two samples. (a) Original CT, (b) warped CT after rigid initialization and plane selection, (c) warped CT after deformable registration, (d) original histology with landmarks from pathologist (black) and warped projected landmarks from radiologists (yellow), (e) overlaid cartilage masks after registration of histology (filled blue) and radiology (red for our method, yellow for MSV-RegSynNet), (f) Overlaid contours between warped CT (GTV, red) and WSI (true tumor extent, blue). (Color figure online)

Table 1. Mean and stand. dev. registration performance in terms of Dice Score (%), Hausdorff Distance (mm) and Landmark Error (mm). Inference runtime is in seconds.

Method	Dice	Hausdorff	Landmark	Runtime
VoxelMorph 3D	68.4 ± 0.6	8.53 ± 0.32	6.71 ± 0.16	**1.3**
VoxelMorph 2/3D	71.9 ± 1.7	7.19 ± 0.24	5.99 ± 0.22	1.4
MSV-RegSynNet	76.3 ± 1.4	6.88 ± 0.28	4.98 ± 0.15	2.1
Ours (no init)	77.9 ± 1.9	6.91 ± 0.19	4.73 ± 0.31	2.1
Ours (no regu)	85.1 ± 0.8	4.23 ± 0.27	3.71 ± 0.19	2.8
Ours	**86.9 ± 1.3**	**3.81 ± 0.20**	**3.28 ± 0.16**	2.9
3D CT/2D MR	0.35T TrueFISP → 3D CT		1.5T T2 → 3D CT	
	Dice	Hausdorff	Dice	Hausdorff
MSV-RegSynNet	84.6 ± 0.9	7.25 ± 0.05	86.1 ± 1.0	5.84 ± 0.15
Ours	**84.8 ± 1.1**	**7.12 ± 0.08**	**87.9 ± 1.2**	**5.21 ± 0.09**

difficulties inherent to the histological process like a cut larynx, the model successfully maps both cartilage and soft tissue without completely tearing the CT image thanks to regularization (c-d-e). For quantitative assessment, we computed the DSC as well as the Hausdorff Distance between cartilages, and the average distance between characteristic landmarks disposed before registration(Table 1). Our method outperforms all baselines, proving the necessity of a singular approach to handle the specific case of histology. The popular Voxelmorph framework fails, and the 2D-3D adaptation demonstrates the value of the masked loss function. The superior performance of MSV-RegSynNet advocates for a modality translation-based method compared to a direct multimodal similarity criterion. In addition, the ablation studies prove the benefit of the distance field regularization and more importantly the cascaded initialization. Concerning the GPU runtime, with a 3-step cascade for initialization, the inference remains in a similar time scale to baseline methods and performs mapping in less than $3s$. We also compared against MSV-RegSynNet on its own validation dataset for generalization assessment: we yielded comparable results for the first cohort and significantly better ones for the second, which proves that StructuRegNet behaves well on other modalities and that the structure awareness is an essential asset for better registration, as pelvis is a location where organs are moving. Visuals of registration results are displayed in the supplementary material. Eventually, an important clinical endpoint of our study is to compare the GTV delineated on CT with gold-standard tumor extent after co-registration to highlight systematic errors and better understand the biological environment from the radiologic signals. We show in (f) that the GTV delineated on CT overestimates the true tumor extent of around 31%, but does not always encompass the tumor with a proportion of histological tumor contained within the CT contour of 0.86. The typical error cases are the inclusion of cartilage or edema, which highlights the

limitations and variability of radiology-based examinations, leading to increased toxicity or untreated areas in RT.

5 Discussion and Conclusion

We introduced a novel framework for 2D/3D multimodal registration. StructuRegNet leverages the structure of tissues to guide the registration through both initial plane selection and deformable regularization; it combines adversarial training for modality translation with a 2D-3D mapping setting and does not require any protocol for 3D reconstruction. It is worth noticing that even if the annotation of cartilage was manual, automating this process is not a bottleneck as the difference in contrast between soft tissue and stiff areas is clear enough to leverage any image processing tool for this task. Finally, it is entirely versatile as we designed our experiments for CT-WSI but any 3D radiological images are suitable. We achieve superior results than state-of-the-art methods in DL-based registration in a similar time scale, allowing precise mapping of both modalities and a better understanding of the tumor microenvironment. The main limitation lies in the handling of organs without any rigid areas like the prostate. Future work also includes a study with biomarkers from immunohistochemistry mapped onto radiology to go beyond binary tumor masks and move toward virtual biopsy.

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**(8), 1788–1800 (2019). https://doi.org/10.1109/TMI.2019.2897538
2. Caldas-Magalhaes, J., et al.: The accuracy of target delineation in laryngeal and hypopharyngeal cancer. Acta Oncologica **54**(8), 1181–1187 (2015). https://doi.org/10.3109/0284186X.2015.1006401
3. Chappelow, J., et al.: Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information. Med. Phys. **38**(4), 2005–2018 (2011). https://doi.org/10.1118/1.3560879
4. Ferrante, E., Paragios, N.: Slice-to-volume medical image registration: a survey. Med. Image Anal. **39**, 101–123 (2017). https://doi.org/10.1016/j.media.2017.04.010
5. Geets, X., et al.: Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI. Radiother. Oncol.: J. Eur. Soc. Ther. Radiol. Oncol. **77**(1), 25–31 (2005). https://doi.org/10.1016/j.radonc.2005.04.010
6. Guo, H., Xu, X., Xu, S., Wood, B.J., Yan, P.: End-to-end ultrasound frame to volume registration (2021)
7. Heinrich, M.P., et al.: MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. **16**(7), 1423–1435 (2012). https://doi.org/10.1016/j.media.2012.05.008
8. Jaganathan, S., Wang, J., Borsdorf, A., Shetty, K., Maier, A.: Deep iterative 2D/3D registration. arXiv:2107.10004 [cs, eess], vol. 12904, pp. 383–392 (2021). https://doi.org/10.1007/978-3-030-87202-1_37

9. Jager, E.A., et al.: Interobserver variation among pathologists for delineation of tumor on H&E-sections of laryngeal and hypopharyngeal carcinoma. How good is the gold standard? Acta Oncologica **55**(3), 391–395 (2016). https://doi.org/10.3109/0284186X.2015.1049661

10. Kimm, S.Y., et al.: Methods for registration of magnetic resonance images of ex vivo prostate specimens with histology. J. Magn. Reson. Imaging **36**(1), 206–212 (2012)

11. Kuckertz, S., Papenberg, N., Honegger, J., Morgas, T., Haas, B., Heldmann, S.: Learning deformable image registration with structure guidance constraints for adaptive radiotherapy. In: Špiclin, Ž, McClelland, J., Kybic, J., Goksel, O. (eds.) WBIR 2020. LNCS, vol. 12120, pp. 44–53. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-50120-4_5

12. Lee, M.C.H., Oktay, O., Schuh, A., Schaap, M., Glocker, B.: Image-and-spatial transformer networks for structure-guided image registration (2019). https://doi.org/10.48550/arXiv.1907.09200

13. Leroy, A., et al.: MO-0476 statistical discrepancies in GTV delineation for H&N cancer across expert centers. Radiother. Oncol. **170**, S426–S427 (2022). https://doi.org/10.1016/S0167-8140(22)02370-2

14. Leroy, A., et al.: End-to-end multi-slice-to-volume concurrent registration and multimodal generation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, pp. 152–162. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_15

15. Li, L., et al.: Co-registration of ex vivo surgical histopathology and in vivo T2 weighted MRI of the prostate via multi-scale spectral embedding representation. Sci. Rep. **7**(1), 8717 (2017). https://doi.org/10.1038/s41598-017-08969-w

16. Markova, V., Ronchetti, M., Wein, W., Zettinig, O., Prevost, R.: Global multimodal 2D/3D registration via local descriptors learning (2022)

17. Njeh, C.F.: Tumor delineation: the weakest link in the search for accuracy in radiotherapy. J. Med. Phys./Assoc. Med. Physicists India **33**(4), 136–140 (2008). https://doi.org/10.4103/0971-6203.44472

18. Ohnishi, T., et al.: Deformable image registration between pathological images and MR image via an optical macro image. Pathol. Res. Pract. **212**(10), 927–936 (2016). https://doi.org/10.1016/j.prp.2016.07.018

19. Rusu, M., et al.: Registration of presurgical MRI and histopathology images from radical prostatectomy via RAPSODI. Med. Phys. **47**(9), 4177–4188 (2020)

20. Shao, W., et al.: ProsRegNet: a deep learning framework for registration of MRI and histopathology images of the prostate. arXiv:2012.00991 [eess] (2020)

21. Tian, L., Lee, Y.Z., Estépar, R.S.J., Niethammer, M.: LiftReg: limited angle 2D/3D deformable registration (2023)

22. Ward, A.D., et al.: Prostate: registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil. Radiology **263**(3), 856–864 (2012). https://doi.org/10.1148/radiol.12102294

23. Xiao, G., et al.: Determining histology-MRI slice correspondences for defining MRI-based disease signatures of prostate cancer. Comput. Med. Imaging Graph. **35**(7), 568–578 (2011). https://doi.org/10.1016/j.compmedimag.2010.12.003

24. Xu, Z., et al.: Adversarial uni- and multi-modal stream networks for multimodal image registration. arXiv:2007.02790 [cs, eess] (2020)

25. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv:1703.10593 [cs] (2020)