



# Towards Multi-modal Anatomical Landmark Detection for Ultrasound-Guided Brain Tumor Resection with Contrastive Learning

Soorena Salari<sup>1</sup>(✉), Amirhossein Rasoulilian<sup>1</sup>, Hassan Rivaz<sup>2</sup>, and Yiming Xiao<sup>1</sup>

<sup>1</sup> Department of Computer Science and Software Engineering,  
Concordia University, Montreal, Canada

{soorena.salari, ah.rasoulilian, yiming.xiao}@concordia.ca

<sup>2</sup> Department of Electrical and Computer Engineering,  
Concordia University, Montreal, Canada  
hassan.rivaz@concordia.ca

**Abstract.** Homologous anatomical landmarks between medical scans are instrumental in quantitative assessment of image registration quality in various clinical applications, such as MRI-ultrasound registration for tissue shift correction in ultrasound-guided brain tumor resection. While manually identified landmark pairs between MRI and ultrasound (US) have greatly facilitated the validation of different registration algorithms for the task, the procedure requires significant expertise, labor, and time, and can be prone to inter- and intra-rater inconsistency. So far, many traditional and machine learning approaches have been presented for anatomical landmark detection, but they primarily focus on mono-modal applications. Unfortunately, despite the clinical needs, inter-modal/contrast landmark detection has very rarely been attempted. Therefore, we propose a novel contrastive learning framework to detect corresponding landmarks between MRI and intra-operative US scans in neurosurgery. Specifically, two convolutional neural networks were trained jointly to encode image features in MRI and US scans to help match the US image patch that contain the corresponding landmarks in the MRI. We developed and validated the technique using the public RESECT database. With a mean landmark detection accuracy of  $5.88 \pm 4.79$  mm against  $18.78 \pm 4.77$  mm with SIFT features, the proposed method offers promising results for MRI-US landmark detection in neurosurgical applications for the first time.

**Keywords:** Deep learning · Anatomical landmark · Contrastive learning · Inter-modality · Neurosurgery · Intraoperative ultrasound

## 1 Introduction

Gliomas are the most common central nervous system (CNS) tumors in adults, accounting for 80% of primary malignant brain tumors [1]. Early surgical treatment to remove the maximum amount of cancerous tissues while preserving the

eloquent brain regions can improve the patient’s survival rate and functional outcomes of the procedure [2]. Although the latest multi-modal medical imaging (e.g., PET, diffusion/functional MRI) allows more precise pre-surgical planning, during surgery, brain tissues can deform under multiple factors, such as gravity, intracranial pressure change, and drug administration. The phenomenon is referred to as brain shift, and often invalidates the pre-surgical plan by displacing surgical targets and other vital anatomies. With high flexibility, portability, and cost-effectiveness, intra-operative ultrasound (US) is a popular choice to track and monitor brain shift. In conjunction with effective MRI-US registration algorithms, the tool can help update the pre-surgical plan during surgery to ensure the accuracy and safety of the intervention.

As the true underlying deformation from brain shift is impossible to obtain and the differences of image features between MRI and US are large, quantitative validation of automatic MRI-US registration algorithms often rely on homologous anatomical landmarks that are manually labeled between corresponding MRI and intra-operative US scans [3]. However, manual landmark identification requires strong expertise in anatomy and is costly in labor and time. Moreover, inter- and intra-rater variability still exists. These factors make quality assessment of brain shift correction for US-guided brain tumor resection challenging. In addition, due to the time constraints, similar evaluation of inter-modal registration quality during surgery is nearly impossible, but still highly desirable. To address these needs, deep learning (DL) holds the promise to perform efficient and automatic inter-modal anatomical landmark detection.

Previously, many groups have proposed algorithms to label landmarks in anatomical scans [4–9]. However, almost all earlier techniques were designed for mono-modal applications, and inter-modal landmark detection, such as for US-guided brain tumor resection, has rarely been attempted. In addition, unlike other applications, where the full anatomy is visible in the scan and all landmarks have consistent spatial arrangements across subjects, intra-operative US of brain tumor resection only contains local regions of the pathology with non-canonical orientations. This results in anatomical landmarks with different spatial distributions across cases. To address these unique challenges, we proposed a new contrastive learning (CL) framework to detect matching landmarks in intra-operative US with those from MRI as references. Specifically, the technique leverages two convolutional neural networks (CNNs) to learn features between MRI and US that distinguish the inter-modal image patches which are centered at the matching landmarks from those that are not. Our approach has two major novel contributions to the field. First, we proposed a multi-modal landmark detection algorithm for US-guided brain tumor resection for the first time. Second, CL is employed for the first time in inter-modal anatomical landmark detection. We developed and validated the proposed technique with the public RESECT database [10] and compared its landmark detection accuracy against the popular scale-invariant feature transformation (SIFT) algorithm in 3D [11].

## 2 Related Work

Contrastive learning has recently shown great results in a wide range of medical image analysis tasks [12–18]. In short, it seeks to boost the similarity of feature representations between counterpart samples and decrease those between mismatched pairs. Often, these similarities are calculated based on deep feature representations obtained from DL models in the feature embedding space. This self-supervised learning set-up allows robust feature learning and embedding without explicit guidance from fine-grained image annotations, and the encoded features can be adopted in various downstream tasks, such as segmentation. A few recent works [19–21] explored the potential of CL in anatomical landmark annotation in head X-ray images for 2D skull landmarks. Quan et al. [19, 20] attempted to leverage CL for more efficient and robust learning. Yao et al. [21] used multiscale pixel-wise contrastive proxy tasks for skull landmark detection in X-ray images. With a consistent protocol for landmark identification, they trained the network to learn signature features within local patches centered at the landmarks. These prior works with CL focus on single-modal 2D landmark identification with systematic landmark localization protocols and sharp image contrast (i.e., skull in X-ray). In contrast, our described application is more challenging due to the 3D nature, difficulty in inter-modal feature learning, weaker anatomical contrast (i.e., MRI vs US), and variable landmark locations. In CL, many works have employed the InfoNCE loss function [22, 23] in attaining good outcomes. Inspired by Yao *et al.* [21], we aimed to use InfoNCE as our loss function with a patch-based approach. To date, CL has not been explored in multi-modal landmark detection, a unique problem in clinical applications. In this paper, to bridge this knowledge gap, we proposed a novel CL-based framework for MRI-US anatomical landmark detection.

## 3 Methods and Materials

### 3.1 Data and Landmark Annotation

We employed the publicly available EASY-RESECT (REtroSpective Evaluation of Cerebral Tumors) dataset [10] (<https://archive.sigma2.no/pages/public/dataset%20Detail.jsf?id=10.11582/2020.00025>) to train and evaluate our proposed method. This dataset is a deep-learning-ready version of the original RESECT database, and was released as part of the 2020 Learn2Reg Challenge [24]. Specifically, EASY-RESECT contains MRI and intra-operative US scans (before resection) of 22 subjects who have undergone low-grade glioma resection surgeries. All images were resampled to a unified dimension of  $256 \times 256 \times 288$  voxels, with an isotropic resolution of  $\sim 0.5\text{mm}$ . Between MRI and the corresponding US images, matching anatomical landmarks were manually labeled by experts and 15–16 landmarks were available per case. A sample illustration of corresponding inter-modal scans and landmarks is shown in Fig. 1. For the target application, we employed the T2FLAIR MRI to pair with intra-operative US since low-grade gliomas are usually more discernible in T2FLAIR than in T1-weighted MRI[10].

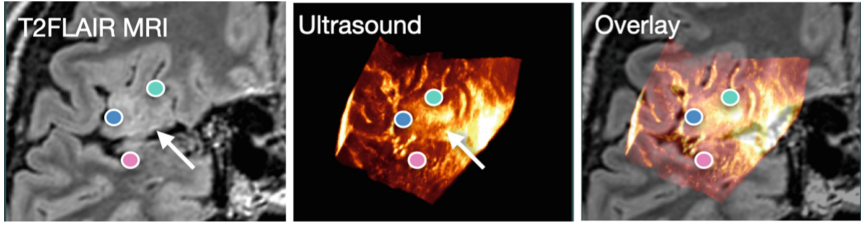


Fig. 1: Sample corresponding landmarks on co-registered T2FLAIR MRI and US. The arrows point to the brain tumor region.

### 3.2 Contrastive Learning Framework

We used two CNNs with identical architectures in parallel to extract robust image features from MRI and US scans. Specifically, these CNNs are designed to acquire relevant features from MRI and US patches, and maximize the similarity between features of corresponding patches while minimizing those between mismatched patches. Each CNN network contains six successive blocks, and each block consists of one convolution layer and one group norm, with Leaky ReLU as the activation function. Also, the convolution layer of the first and last three blocks of the network has 64 and 32 convolutional filters, respectively, and a kernel size of 3 is used across all blocks. After the convolution layers, the proposed network has two multi-layer perceptron (MLP) layers with 64 and 32 neurons and Leaky ReLU as the activation function. These MLP layers compress the extracted features from convolutional layers and produce the final feature vectors. The resulting CNN network is depicted in Fig. 2.

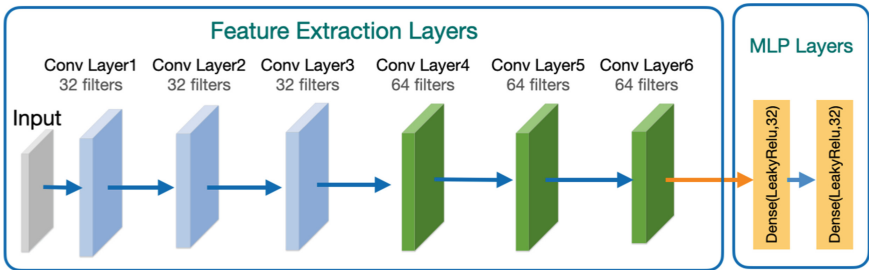


Fig. 2: The proposed CNN for feature encoding from MRI and US scans.

### 3.3 Landmark Matching with a 2.5D Approach

Working with 3D images is computationally expensive and can make the model training unstable and prone to overfitting, especially when the size of the database is limited. Therefore, instead of a full 3D processing, we decided to implement a 2.5D approach [25] to leverage the efficiency of 2D CNN in the

CL framework for the task. In this case, we extracted a series of three adjacent 2D image patches in one canonical direction (x-, y-, or z-direction), with the middle slice centred at the true or candidate landmarks in a 3D scan to provide slight spatial context for the middle slice of interest. To construct the full 2.5D formulation, we performed the same image patch series extraction in all x-, y-, and z-directions for a landmark, and this 2.5D patch forms the basis to compute the similarity between the queried US and reference MRI patches. Note that the setup of CL requires three types of samples, anchor, positive sample pairs, and negative sample pairs. Specifically, the anchor is defined as the 2.5D MRI patch centred at a predefined landmark, a positive pair is represented by an anchor and a 2.5D US patch at the corresponding landmark, and finally, a negative pair means an anchor and a mismatched 2.5D US patch. Note that during network training, instead of 2.5D patches, we compared the 2D image patch series in one canonical direction between MRI and US, and 2D patch series in all three directions were used. During the inference stage, the similarity between MRI and US 2.5D patches was obtained by summing the similarities of corresponding 2D image patch series in each direction, and a match was determined with the highest similarity from all queried US patches. With the assumption that the brain shift moves the anatomy within a limited range, during the inference time, we searched within a range of  $[-5, 5]$  mm in each direction in the US around the reference MRI landmark location to find the best match. Note that this search range is an adjustable parameter by the user (e.g., surgeons/clinicians), and when no match is found in the search range, an extended search range can be used. The general overview of the utilized framework for 2D image patch extraction is shown in Fig. 3.

### 3.4 Landmark Matching with 3D SIFT

The SIFT algorithm [11] is a well-known tool for keypoint detection and image registration. It has been widely used in multi-modal medical registration, such as landmark matching for brain shift correction in image-guided neurosurgery [8, 26]. To further validate the proposed CL-based method for multi-modal anatomical landmark detection in US scans, we replicated the procedure using the 3D SIFT algorithm as follows. First, we calculated the SIFT features at the reference landmark's location in MRI. Then, we acquired a set of candidate SIFT points in the corresponding US scan. Finally, we identified the matching US landmark by selecting the top ranking candidate based on SIFT feature similarity measured with cosine similarity. Note that, for SIFT-based landmark matching, we have attempted to impose a similar spatial constraint like in the CL-based approach. However, as the SIFT algorithm pre-selects keypoint candidates based on their feature strengths, with this in consideration, we saw no major benefits by imposing the spatial constraint.

## 4 Experimental Setup

### 4.1 Data Preprocessing

For CL training, both positive and negative sample pairs need to be created. All 2D patch series were extracted according to Sect. 3.3 with a size of  $42 \times 42 \times 3$  voxels. These sample pairs were used to train two CNNs to extract relevant image features across MRI and US leveraging the InfoNCE loss.

### 4.2 Loss Function

We used the InfoNCE loss [23] for our CL framework. The loss function has been widely used and demonstrated great performance in many vision tasks. Like other contrastive loss functions, InfoNCE requires a similarity function, and we chose commonly used cosine similarity. The formulas for InfoNCE ( $\mathcal{L}_{InfoNCE}$ ) and cosine similarity ( $CosSim$ ) are as follows:

$$\begin{aligned} \mathcal{L}_{InfoNCE} &= -\mathbb{E} \left[ \log \frac{\exp(\alpha)}{\exp(\alpha) + \sum \exp(\alpha')} \right]; \\ \alpha &= s[F_\theta \circ X_i^A, G_\beta \circ X_i^P]; \\ \alpha' &= s[F_\theta \circ X_i^A, G_\beta \circ X_j^N], \end{aligned} \quad (1)$$

$$s[v, w] = \text{Cos Sim}(v, w) = \frac{\langle v \cdot w \rangle}{\|v\| \cdot \|w\|} \quad (2)$$

where  $F_\theta$  and  $G_\beta$  are the CNN feature extractors for MR and US patches.  $X_i^A$  and  $X_i^P$  are the cropped image patches around the corresponding landmarks in MR and US scans, respectively, and  $X_j^N$  is a mismatched patch in the US image to that cropped around the MRI reference landmark. Here,  $F_\theta \circ X_i^A$ ,  $G_\beta \circ X_i^P$ , and  $G_\beta \circ X_j^N$  give the extracted feature vectors for MR and US patches.

### 4.3 Implementation Details and Evaluation

To train our DL model, we made subject-wise division of the entire dataset into 70%:15%:15% as the training, validation, and testing sets, respectively. Also, to improve the robustness of the network, we used data augmentation for the training data by random rotation, random horizontal flip, and random vertical flip. Furthermore, an AdamW optimizer with a learning rate of 0.00001 was used, and we trained our model for 50 epochs with a batch size of 256.

In order to evaluate the performance of our technique, we used the provided ground truth landmarks from the database and calculated the Euclidean distance between the ground truths and predictions. The utilized metric is as follows:

$$\text{Mean landmark identification error} = \frac{1}{N} \sum_{i=1}^N \|x_i - x'_i\| \quad (3)$$

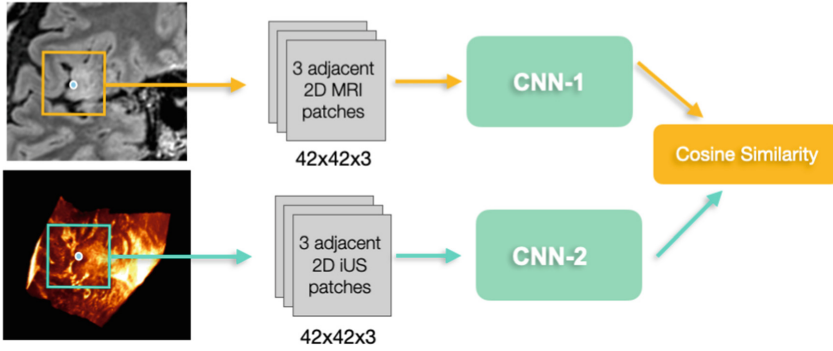


Fig. 3: An overview of the framework for image feature learning.

where  $x_i$  and  $x'_i$ , and  $N$  are the ground truth landmark location, model prediction, and the total number of landmarks per subject, respectively.

## 5 Results

Table 1 lists the mean and standard deviation of landmark identification errors (in mm) between the predicted position and the ground truth in intra-operative US for each patient of the RESECT dataset. In the table, we also provide the severity of brain shift for each patient. Here, tissue deformation measured as mean target registration errors (mTREs) with the ground truth anatomical landmarks is classified as small (mTRE below 3 mm), median (3–6 mm), or large (above 6 mm). The results show that our CL-based landmark selection technique can locate the corresponding US landmarks with a mean landmark identification error of  $5.88 \pm 4.79$  mm across all cases while the SIFT algorithm has an error  $18.78 \pm 4.77$  mm. With a two-sided paired-samples t-test, our method outperformed the SIFT approach with statistical significance ( $p < 1e-4$ ). When reviewing the mean landmark identification error using our proposed technique, we also found that the magnitude is associated with the level of brain shift. However, no such trend is observed when using SIFT features for landmark identification. When inspecting landmark identification errors across all subjects between the CL and SIFT techniques, we also noticed that our CL framework has significantly lower standard deviations ( $p < 1e-4$ ), implying that our technique has a better performance consistency.

## 6 Discussion

Inter-modal anatomical landmark localization is still a difficult task, especially for the described application, where landmarks have no consistent spatial arrangement across different cases and image features in US are rough. We tackled the challenge with the CL framework for the first time. As the first step

Table 1: Landmark identification errors (mean $\pm$ std) per case in mm. Our proposed CL-based algorithm achieved a mean landmark identification error of **5.88 $\pm$ 4.79** mm across all cases while the SIFT algorithm obtained an error of **18.78 $\pm$ 4.77** mm. The level of brain shift is listed beside the patient ID.

Patient ID	Proposed CL	SIFT Algorithm	Patient ID	Proposed CL	SIFT Algorithm
1 (Small)	1.80 $\pm$ 0.78	12.16 $\pm$ 4.75	15 (Medium)	3.07 $\pm$ 1.39	19.33 $\pm$ 4.49
2 (Medium)	6.16 $\pm$ 1.40	17.84 $\pm$ 8.50	16 (Medium)	6.42 $\pm$ 0.92	12.48 $\pm$ 4.84
3 (Large)	7.79 $\pm$ 0.55	13.37 $\pm$ 6.08	17 (Large)	8.13 $\pm$ 0.72	18.57 $\pm$ 6.27
4 (Small)	3.59 $\pm$ 0.73	13.97 $\pm$ 5.50	18 (Medium)	4.19 $\pm$ 0.87	13.71 $\pm$ 5.815
5 (Large)	10.65 $\pm$ 1.13	20.71 $\pm$ 6.12	19 (Medium)	3.97 $\pm$ 0.93	27.70 $\pm$ 16.49
6 (Medium)	2.20 $\pm$ 0.93	28.11 $\pm$ 11.90	21 (Medium)	6.01 $\pm$ 0.77	24.40 $\pm$ 13.73
7 (Small)	1.96 $\pm$ 0.96	24.07 $\pm$ 8.02	23 (Large)	6.97 $\pm$ 1.03	22.50 $\pm$ 6.72
8 (Small)	2.56 $\pm$ 0.79	19.30 $\pm$ 6.09	24 (Small)	1.33 $\pm$ 0.49	14.91 $\pm$ 6.09
12 (Large)	23.77 $\pm$ 0.96	22.01 $\pm$ 6.64	25 (Large)	9.94 $\pm$ 2.43	15.37 $\pm$ 5.42
13 (Medium)	6.18 $\pm$ 1.43	13.86 $\pm$ 6.99	26 (Small)	2.95 $\pm$ 0.88	17.93 $\pm$ 10.15
14 (Medium)	3.39 $\pm$ 0.69	21.67 $\pm$ 6.46	27 (Medium)	6.42 $\pm$ 0.76	19.20 $\pm$ 8.65

towards more accurate inter-modal landmark localization, there are still aspects to be improved. First, while the 2.5D approach is memory efficient and quick, 3D approaches may better capture the full corresponding image features. This is partially reflected by the observation that the quality of landmark localization is associated with the level of tissue shift. However, due to limited clinical data, 3D approaches caused overfitting in our network training. Second, in the current setup, we employed landmarks in pre-operative MRIs as references since its contrast is easier to understand and it allows sufficient time for clinicians to annotate the landmarks before surgery. Future exploration will also seek techniques to automatically tag MRI reference landmarks. Finally, we only employed US scans before resection since tissue removal can further complicate feature matching between MRI and US, and requires more elaborate strategies, such as those involving segmentation of resected regions [27]. We will explore suitable solutions to extend the application scenarios of our proposed framework as part of the future investigation. As a baseline comparison, we employed the SIFT algorithm, which has demonstrated excellent performance in a large variety of computer vision problems for keypoint matching. However, in the described inter-modal landmark identification for US-guided brain tumor resection, the SIFT algorithm didn't offer satisfactory results. This could be due to the coarse image features and textures of intra-operative US and the differences in the physical resolution between MRI and US. One major critique for using the SIFT algorithm is that it intends to find geometrically interesting keypoints, which may not have good anatomical significance. In the RESECT dataset, eligible anatomical landmarks were defined as deep grooves and corners of sulci, convex points of gyri, and vanishing points of sulci. The relevant local features may be



hard to capture with the SIFT algorithm. In this sense, DL-based approaches may be a better choice for the task. With the CL framework, our method learns the common features between two different modalities via the training process. Besides better landmark identification accuracy, the tighter standard deviations also imply that our DL approach serves a better role in grasping the local image features within the image patches.

## 7 Conclusions

In this project, we proposed a CL framework for MRI-US landmark detection for neurosurgery for the first time by leveraging real clinical data, and achieved state-of-the-art results. The algorithm represents the first step towards efficient and accurate inter-modal landmark identification that has the potential to allow intra-operative assessment of registration quality. Future extension of the method in other inter-modal applications can further confirm its robustness and accuracy.

**Acknowledgment.** We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and Fonds de Recherche du Québec Nature et technologies (FRQNT).

## References

1. Holland, E.C.: Progenitor cells and glioma formation. *Curr. Opin. Neurol.* **14**(6), 683–688 (2001)
2. Dolecek, T.A., Propp, J.M., Stroup, N.E., Kruchko, C.: CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the united states in 2005–2009. *Neuro-oncology* **14**(suppl\_5), v1–v49 (2012)
3. Xiao, Y., et al.: Evaluation of MRI to ultrasound registration methods for brain shift correction: the CuRIOUS2018 challenge. *IEEE Trans. Med. Imaging* **39**(3), 777–786 (2019)
4. Yao, Q., Xiao, L., Liu, P., Zhou, S.K.: Label-free segmentation of COVID-19 lesions in lung CT. *IEEE Trans. Med. Imaging* **40**(10), 2808–2819 (2021)
5. Ghesu, F.C., Georgescu, B., Mansi, T., Neumann, D., Hornegger, J., Comaniciu, D.: An artificial agent for anatomical landmark detection in medical images. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9902, pp. 229–237. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46726-9\\_27](https://doi.org/10.1007/978-3-319-46726-9_27)
6. Zhu, H., Yao, Q., Xiao, L., Zhou, S.K.: You only learn once: universal anatomical landmark detection. In: de Bruijne, M., et al. (eds.) *MICCAI 2021*. LNCS, vol. 12905, pp. 85–95. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-87240-3\\_9](https://doi.org/10.1007/978-3-030-87240-3_9)
7. Tripathi, A., et al.: Unsupervised landmark detection and classification of lung infection using transporter neural networks. *Comput. Biol. Med.* **152**, 106345 (2023)
8. Toews, M., Wells, W.M., III.: Efficient and robust model-to-image alignment using 3d scale-invariant features. *Med. Image Anal.* **17**(3), 271–282 (2013)

9. Salari, S., Rasoulia, A., Battie, M., Fortin, M., Rivaz, H., Xiao, Y.: Uncertainty-aware transformer model for anatomical landmark detection in paraspinal muscle MRIs. In: Medical Imaging, Image Processing, vol. 12464, pp. 238–244. SPIE (2023)
10. Xiao, Y., Fortin, M., Unsgård, G., Rivaz, H., Reinertsen, I.: Retrospective evaluation of cerebral tumors (RESECT): a clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries. *Med. Phys.* **44**(7), 3875–3882 (2017)
11. Rister, B., Horowitz, M.A., Rubin, D.L.: Volumetric image registration from invariant keypoints. *IEEE Trans. Image Process.* **26**(10), 4900–4910 (2017)
12. You, K., Lee, S., Jo, K., Park, E., Kooi, T., Nam, H.: Intra-class contrastive learning improves computer aided diagnosis of breast cancer in mammography. In: Medical Image Computing and Computer Assisted Intervention-MICCAI: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part III, pp. 55–64. Springer (2022). [https://doi.org/10.1007/978-3-031-16437-8\\_6](https://doi.org/10.1007/978-3-031-16437-8_6)
13. Cheng, L.-H., Sun, X., van der Geest, R.J.: Contrastive learning for echocardiographic view integration. In: Medical Image Computing and Computer Assisted Intervention-MICCAI: 25th International Conference, Singapore, v September 2022, Proceedings, Part IV, pp. 340–349. Springer (2022). [https://doi.org/10.1007/978-3-031-16440-8\\_33](https://doi.org/10.1007/978-3-031-16440-8_33)
14. Bhattacharya, D., et al.: Supervised contrastive learning to classify paranasal anomalies in the maxillary sinus. In: Medical Image Computing and Computer Assisted Intervention-MICCAI, et al.: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part III, pp. 429–438. Springer (2022). [https://doi.org/10.1007/978-3-031-16437-8\\_41](https://doi.org/10.1007/978-3-031-16437-8_41)
15. Emre, T., Chakravarty, A., Rivail, A., Riedl, S., Schmidt-Erfurth, U., Bogunović, H.: Tinc: temporally informed non-contrastive learning for disease progression modeling in retinal OCT volumes. In: Medical Image Computing and Computer Assisted Intervention-MICCAI: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part II, pp. 625–634. Springer (2022). [https://doi.org/10.1007/978-3-031-16434-7\\_60](https://doi.org/10.1007/978-3-031-16434-7_60)
16. Liu, T., Liu, W., Yu, L., Wan, L., Han, T., Zhu, L.: Joint prediction of meningioma grade and brain invasion via task-aware contrastive learning. In: Medical Image Computing and Computer Assisted Intervention-MICCAI: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part III, pp. 355–365. Springer (2022). [https://doi.org/10.1007/978-3-031-16437-8\\_34](https://doi.org/10.1007/978-3-031-16437-8_34)
17. Pan, Y., Gernand, A.D., Goldstein, J.A., Mithal, L., Mwinyelle, D., Wang, J.Z.: Vision-language contrastive learning approach to robust automatic placenta analysis using photographic images. In: Medical Image Computing and Computer Assisted Intervention-MICCAI: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part III, pp. 707–716. Springer (2022). [https://doi.org/10.1007/978-3-031-16437-8\\_68](https://doi.org/10.1007/978-3-031-16437-8_68)
18. Hang, W., Huang, Y., Liang, S., Lei, B., Choi, K.-S., Qin, J.: Reliability-aware contrastive self-ensembling for semi-supervised medical image classification. In: Medical Image Computing and Computer Assisted Intervention-MICCAI: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part I, pp. 754–763. Springer (2022). [https://doi.org/10.1007/978-3-031-16431-6\\_71](https://doi.org/10.1007/978-3-031-16431-6_71)
19. Quan, Q., Yao, Q., Li, J., Zhou, S.K.: Which images to label for few-shot medical landmark detection?. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20606–20616 (2022)

20. Quan, Q., Yao, Q., Li, J., et al.: Information-guided pixel augmentation for pixel-wise contrastive learning, arXiv preprint [arXiv:2211.07118](https://arxiv.org/abs/2211.07118) (2022)
21. Yao, Q., Quan, Q., Xiao, L., Kevin Zhou, S.: One-shot medical landmark detection. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 177–188. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-87196-3\\_17](https://doi.org/10.1007/978-3-030-87196-3_17)
22. Gutmann, M., Hyvärinen, A.: Noise-contrastive estimation: a new estimation principle for unnormalized statistical models. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, pp. 297–304 (2010)
23. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding, arXiv preprint [arXiv:1807.03748](https://arxiv.org/abs/1807.03748) (2018)
24. Hering, A., et al.: Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Trans. Med. Imaging* (2022)
25. Pirhadi, A., Salari, S., Ahmad, M.O., Rivaz, H., Xiao, Y.: Robust landmark-based brain shift correction with a Siamese neural network in ultrasound-guided brain tumor resection. *Inter. J. Comput. Assisted Radiol. Surgery*, 1–8 (2022)
26. Luo, J., et al.: A feature-driven active framework for ultrasound-based brain shift compensation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 30–38. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00937-3\\_4](https://doi.org/10.1007/978-3-030-00937-3_4)
27. Canalini, L., Klein, J., Miller, D., Kikinis, R.: Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery. *Int. J. Comput. Assist. Radiol. Surg.* **14**, 1697–1713 (2019)