Uncovering Interpretable Point-Wise Image Correspondences from Deep Segmentation Features

Eli Duval(✉) and Frederik Maes

Dept. ESAT/PSI, KU Leuven, Belgium
eli.duval@kuleuven.be

Abstract. Establishing accurate point-wise correspondences between different images is crucial for various tasks in medical image computing, such as motion tracking, shape analysis and spatial normalization. Existing approaches mainly rely on deformable registration or (semi-) supervised landmark detection and typically require dedicated training or annotated correspondences. In this work, we propose a training-free framework for correspondence estimation that leverages internal representations from pretrained segmentation networks. We hypothesize that deep learning segmentation models implicitly encode rich anatomical structure within their feature maps. By computing voxel-wise similarities in this feature space, interpretable point-wise correspondences are obtained without need for additional training or landmark annotations. We evaluate our approach and the impact of design options, such as the granularity of the segmentation, the choice of feature maps and the similarity measure, for one-shot landmark detection in different applications involving different anatomies and modalities, demonstrating competitive performance in terms of mean target registration error (mean TRE: intra-subject cardiac MRI: 2.85 mm; intra-subject lung CT: 2.76 mm; inter-subject brain MRI: 3.08 mm). We further illustrate the potential of the method for joint estimation of coupled landmarks by incorporating geometric constraints and for unsupervised salient keypoint discovery across a sequence of images. Our results show that accurate, interpretable, label- and training-efficient correspondence estimation is feasible by repurposing segmentation-derived representations in a generic manner.

Keywords: Segmentation model · Correspondence prediction · Landmark detection · Interpretable · Multi-scale

1 Introduction

Deep learning medical image segmentation has achieved remarkable success across modalities and anatomies. However, conventional segmentation models do not explicitly provide precise point-wise correspondences between images, which are essential for downstream tasks that involve local analysis of anatomical shape changes over time or shape differences between subjects. Such point-wise correspondences are typically obtained by deformable image registration.

Current state-of-the-art approaches, such as the popular deep learning framework VoxelMorph [2], learn and predict dense voxel-wise deformations based on image similarity. However, this dense optimization problem remains ill-posed, requiring careful regularization, and the predicted correspondences are often difficult to interpret. Alternatively, landmark detection methods can be used that predict sparse correspondences for predefined anatomical points. Conventional approaches rely on supervised training with annotated landmarks, which are costly to obtain. Recently, weakly supervised approaches have been proposed [7,18,19,14,8,1] that address the lack of ground-truth correspondences by learning distinctive landmark representations through contrastive learning, with correspondences derived from synthetic transformations. While these methods are label-efficient, they still require dedicated training and their performance depends on how well their transformations reflect true anatomical variability.

To overcome these limitations of label and training costs, we propose a novel approach for establishing point-wise correspondences between images that does not require dedicated training of a correspondence-specific model. Our key insight is that since state-of-the-art segmentation models perform robustly across diverse image appearances, their learned representations capture rich and stable information about the underlying anatomical structures. We hypothesize that these representations implicitly encode multi-scale information that can be exploited for uncovering correspondences between images by exploring voxel-wise similarities in the feature space of the segmentation model, resulting in inherently interpretable results. To our knowledge, leveraging pretrained segmentation networks for correspondence estimation without additional training has not been explored in medical imaging computing. Our approach is conceptually related to recent computer-vision work [17], which shows that correspondences can emerge from pretrained diffusion models. The proposed framework is generic, does not rely on predefined landmarks or anatomical priors and can be applied across imaging modalities and anatomical regions. We demonstrate the potential of our approach for one-shot landmark detection for three applications involving different modalities and anatomies, providing quantitative validation of several design choices and comparison against state-of-the-art. We further show that geometric constraints can be incorporated for joint landmark estimation and we also present an extension of the method to enable unsupervised discovery of salient keypoints across a sequence of images.

2 Methodology

Point-wise correspondence estimation We propose a training-free framework that leverages the features of a pretrained deep learning segmentation model to uncover point-wise correspondences between two images. The segmentation model used is assumed to be applicable to the images considered, but is trained completely independently. Each image is processed by the segmentation network and feature maps from multiple layers of the network are extracted for each image separately. These maps are spatially upsampled to the original image

resolution to account for the internal downsampling operations in the network. Similarities between the extracted features are computed layer by layer at the voxel level and then aggregated by summation into a single similarity score. For points x in image X, point y in image Y and network layers L with upsampled features $f^X_{l\uparrow}$ and $f^Y_{l\uparrow}$ in layer l, their similarity $s(x, y)$ is thus defined as:

$$s(x, y) = \sum_{l \in L} S_l(f^X_{l\uparrow}(x), f^Y_{l\uparrow}(y)) \tag{1}$$

with S_l the similarity measure adopted for layer l. Computing similarities between a specified point x in X and every point in Y produces a similarity map. This map is converted into a correspondence probability distribution $p(y|x)$ using softmax normalization with temperature τ:

$$p(y|x) = \frac{\exp(s(x, y)/\tau)}{\sum_{y' \in Y} \exp\left(s(x, y')/\tau\right)} \tag{2}$$

This probability distribution expresses the likelihood of each point in image Y to correspond with the point x in image X based on similarity in their deep segmentation features. Since $\sum_{y \in Y} p(y|x) = 1$, it is implicitly assumed that x has indeed a corresponding point in image Y. For any given query point x in source image X, the corresponding point $y(x)$ in target image Y is then estimated as the point with the highest correspondence likelihood: $y(x) = \arg\max_{y \in Y} p(y|x)$. The correspondence can be easily interpreted by visualizing $p(y|x)$ as a probability map overlaid on Y, providing direct insight into the correspondence uncertainty: a sharp peak in $p(y|x)$ around $y(x)$ indicates a confident match, whereas a broad plateau reflects ambiguity. An overview of the framework is shown in Figure 1.

The correspondence estimation framework as outlined above is generic and makes no assumptions about the nature of the images, the points of interest, or the segmentation model itself. The implementation requires different design choices, such as: which layers of the segmentation model are selected for feature extraction and similarity computation, how are the feature maps upsampled to voxel resolution, and which similarity measure is employed. Different choices are considered and validated in the experiments.

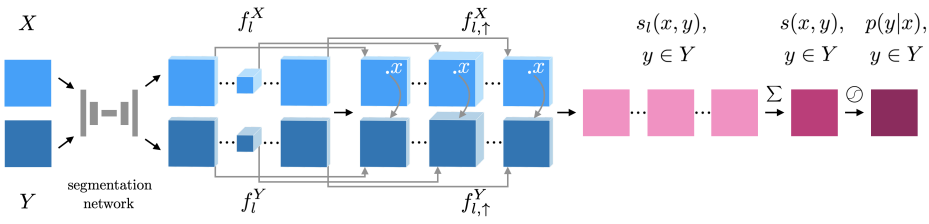

Fig. 1. Overview of the proposed framework: feature maps f_l for images X and Y are extracted from a trained segmentation network and upsampled to the image resolution, yielding $f_{l,\uparrow}$. For a query point x in source image X, deep feature similarities $s_l(x, y)$ with all points y in target image Y are computed layer per layer and combined into a single similarity score $s(x, y)$. Finally, this score is converted to the correspondence probability distribution $p(y|x)$.

Unsupervised salient keypoint discovery The same framework can be applied to enable the discovery of salient keypoints across a sequence of images $X_1, X_2, .., X_n$ by analyzing the structure of their correspondence probability distributions. The underlying intuition is that salient keypoints should have stable and unambiguous correspondences across multiple images. For each consecutive image pair (X_i, X_{i+1}) in the sequence, the correspondence probability distributions $p(x_{i+1}|x_i)$ are computed for each point x_i in X_i. To compute the correspondence probabilities across the entire sequence, the probabilities are propagated through the sequence using marginalization:

$$p(x_n|x_1) = \sum_{x_{n-1} \in X_{n-1}} ... \sum_{x_3 \in X_3} \sum_{x_2 \in X_2} p(x_n|x_{n-1})...p(x_3|x_2).p(x_2|x_1) \quad (3)$$

The correspondence ambiguity for any point x_1 in X_1 is assessed using the entropy of this distribution: $h(x_1) = -\sum_{x_n \in X_n} p(x_n|x_1) \log p(x_n|x_1)$. The entropy map $h(x_1)$ can be overlaid on X_1 for visual interpretation. Lower entropy values are indicative of keypoints with stable correspondences across the sequence.

3 Experiments

Datasets We evaluated the proposed correspondence estimation framework for landmark detection in three different settings, using publicly available datasets for which expert annotated landmarks are available: 2D intra-subject cardiac MRI, 3D intra-subject lung CT, and 3D inter-subject brain MRI. Cardiac MRI data were obtained from the ACDC dataset [3], consisting of 2D cine MRI images, with annotations of the 2 ventricular insertion points on the end-diastole (ED) and end-systole (ES) frames obtained from [12] (74 slices from 20 patients, median size 222×224, pixel size 1.56 mm). Lung CT data were obtained from the Lung250M-4B benchmark [4], consisting of intra-subject inspiratory and expiratory CT pairs with 300 correspondences annotated between each pair (10 subjects, median size $275 \times 225 \times 291$, voxel size 1 mm). Brain MRI data were taken from the OASIS dataset [13], with ground-truth annotations of 32 anatomical fiducials obtained from [16] for images of 30 subjects and for the MNI template [6]. For comparison with [15], correspondences were estimated from the template to each of the 30 subjects after affinely aligning all images with the template using FreeSurfer [5] (size $193 \times 229 \times 193$, voxel size 1 mm).

Segmentation models For each of these settings, vanilla nnU-Net [10,11] segmentation models were trained, using the standard nnU-Net configuration with minor application-specific adjustments: the cardiac MRI employed the 2D architecture trained for 100 epochs, whereas lung CT and brain MRI used the full-resolution 3D architecture without mirroring augmentation. Across all applications, the networks had 11 layers, with 32 channels in the most shallow layers and 512 (2D) or 320 (3D) in the deepest, resulting in 2496 (2D) or 1920 (3D) features in total. The models were trained and validated on separate images as the images used for correspondence estimation, but from the same source,

which also provided the segmentations used during training. Segmentation performance was verified with the Dice Similarity Coefficient (DSC). To evaluate the impact of the nature of the segmentation model on correspondence estimation, two segmentation models were trained for each application with different levels of anatomical granularity (coarse versus more fine-grained), providing different levels of anatomical detail to the segmentation model: for cardiac MRI, the left ventricular cavity (mean DSC 93.72%) versus the left and right ventricular cavities and myocardium (91.27%); for lung CT, the lungs mask (99.11%) versus the veins and arteries (92.04%); for brain MRI, the left and right hemispheres (96.33%) versus 35 subcortical regions (89.71%).

One-shot landmark detection In one-shot landmark detection, a single image is provided with annotated landmarks and the task is to estimate their correspondences in similar images. Performance is evaluated with the target registration error (TRE, in mm), defined as the Euclidean distance between predicted and ground-truth correspondences (in 2D for cardiac MRI, 3D for lung CT and brain MRI), and the success detection rate (SDR, in %), measured as the proportion of predictions within a specified distance threshold of the ground truth. We evaluated different design choices: the selection of network layers used for feature extraction (single layer, decoder layers, encoder layers, all layers), the type of interpolation used to upsample the feature vectors to voxel resolution (nearest-neighbor, linear), and the similarity measure used (inverse Euclidean distance, inner product, cosine similarity). For simplicity, in Eqn. 1, the same similarity measure S_l was used for all layers.

Illustrations on downstream tasks The proposed framework performs correspondence estimation independently for each landmark, but we also demonstrate coupled estimation of multiple landmarks by embedding geometric constraints. Additionally, we show the potential of the approach for unsupervised salient keypoint discovery across a series of images. Both tasks are illustrated on cardiac MRI and only evaluated visually, as no ground truths are available.

Implementation details The framework itself was implemented in PyTorch (version 2.8) using nnU-Net pip package (version 2.5.1) on Cuda (version 12.8). Experiments were conducted on an NVIDIA GeForce TX 3090 GPU (24 GB). The best configuration of design choices was used in all experiments, except when mentioned otherwise. The hyperparameter τ in Eqn. 2 was set to 0.01. The SDR metric used the distance thresholds of 3mm and 6mm. Mean correspondence prediction runtimes were 11.7 s for cardiac MRI (2 landmarks), 124.3 s for lung CT (300 landmarks), and 26.5 s for brain MRI (32 landmarks), scaling with image size and number of query points.

4 Results and discussion

Figures 2(a-b) show representative results of landmark predictions and correspondence probability maps for one-shot landmark detection in the three settings considered. The probability maps make the predictions interpretable by highlighting the anatomical consistency and spatial (un)certainty of the results.

Table 1. Evaluation of different configurations of the proposed landmark detection framework across the three applications on TRE (mm, mean ± std) and SDR (%, mean) with distance thresholds 3mm and 6mm. The optimal configuration uses all layers, linear interpolation, cosine similarity and fine-grained segmentations. Variants differ from the optimal setting by modifying a single aspect as indicated in the table. Statistically significant differences compared to the optimal variant are indicated with * (Wilcoxon signed-rank test, $p < 0.05$).

	Cardiac MRI			**Lung CT**			**Brain MRI**		
	TRE (mm)	SDR (%)		TRE (mm)	SDR (%)		TRE (mm)	SDR (%)	
		≤3	≤6		≤3	≤6		≤3	≤6
Optimal config.									
	2.85 ± 2.43	62.2	89.2	2.76 ± 6.55	82.8	92.5	3.08 ± 3.43	66.4	90.3
Layer selection									
Best	3.33 ± 3.82	57.8	90.5	8.94 ± 9.71*	7.1*	33.8*	5.35 ± 7.31*	39.6*	80.9*
Worst	111.05 ± 69.00*	3.4*	4.7*	118.64 ± 57.46*	0.8*	1.0*	75.90 ± 26.84*	0.6*	0.8*
Encoder	2.78 ± 2.53	62.8	90.5	3.33 ± 10.31	82.2	92.0	4.86 ± 6.08*	54.8*	78.7*
Decoder	3.33 ± 2.72*	51.4*	88.5	5.87 ± 12.22*	54.4*	74.4*	4.29 ± 4.54*	51.9*	79.1*
Interpolation method									
Nearest	3.81 ± 3.08*	50.5*	80.4*	8.36 ± 16.92*	48.2*	68.1*	5.17 ± 5.12	41.8*	71.7*
Similarity metric									
Inv. Euc.	26.08 ± 26.08*	21.0*	35.8*	10.42 ± 19.51*	32.5*	46.9*	16.94 ± 9.73*	1.4*	7.2*
In. prod.	6.64 ± 4.25*	20.3*	48.7*	116.09 ± 51.26*	0.1*	0.3*	22.11 ± 21.38*	7.1*	19.1*
Segmentation granularity									
Coarse	3.30 ± 2.74*	54.7	85.8	7.29 ± 11.23*	55.6*	69.1*	4.46 ± 6.98*	62.6*	83.5*

The maps indicate that correspondence distributions are highly concentrated around the true correspondence location, implying that the framework can accurately uncover anatomical correspondences in a generic way, based on features derived from a pretrained segmentation model.

Table 1 summarizes landmark detection performance in terms of TRE and SDR for each application considered and for different design choices and segmentation granularities. Optimal performance was consistently obtained for all three applications by using features from all layers combined, upsampled using linear interpolation and with cosine similarity as similarity metric. The choice of network layers used for feature extraction has a substantial effect on performance. Using features from a single layer leads to inconsistent results and which layers performs best and worst varies per application. Combining features from multiple layers significantly improves accuracy and robustness, with the selection of all layers consistently performing best. As can be seen in Figure 2b, different levels of scale and abstraction are captured throughout the network: earlier layers capture more low-level, structural details, while the last layers encode more semantic information related to the segmentation learned during training. Combining this complementary information yields a rich multi-scale representation for landmark detection. Secondly, the interpolation method used for spatially upsampling of the feature maps influences correspondence precision. While nearest-neighbor interpolation already provides reasonable performance, linear interpolation consistently improves results by producing smoother and more precise feature representations. Lastly, the similarity metric plays a critical role. Cosine similarity achieves the best performance, while inverse Euclidean distance and inner product were found to be not robust. This suggests

that feature vector direction rather than magnitude is more relevant for capturing anatomical similarity in our framework. While even coarse segmentations enable reasonable performance relative to image scale, the use of a more fine-grained segmentation model consistently yielded significantly lower TRE and higher SDR across the three applications. This demonstrates that segmentation models inherently learn representations relevant to correspondence estimation and that richer anatomical supervision results in more informative features.

To further contextualize the performance of the proposed approach, we compare its target registration error with previously reported work evaluated on the same datasets and correspondences. On cardiac MRI, our approach outperforms the supervised landmark detection model of [12] (TRE of 3.53 mm). For lung CT, comparison with three registration models evaluated in [4] shows that our framework outperforms both the unsupervised models VoxelMorph [2] (TRE 7.98 mm) and VoxelMorph+ [9] (TRE 4.31 mm), while the semi-supervised VoxelMorph++ [9] performs more accurate (TRE 2.26 mm). On brain MRI, our approach outperforms the one-shot landmark detection model CABLD [15] (TRE 3.89 mm). Overall, these results shows that our approach reaches competitive performance without requiring any dedicated training or annotations.

Figure 3(a) illustrates incorporation of task-specific geometric priors into the landmark detection framework for downstream shape analysis, such as wall thickness measurement in cardiac MRI. Constrained correspondence estimation in this case was performed by initializing landmark pairs on the endo- and epicardial contours on ED and detecting the corresponding landmarks on ES, whereby each pair of landmarks must lie along identically oriented radial lines from the center of the left ventricular cavity. Candidate landmark pairs are jointly evaluated based on their combined similarity and the best pair is selected. As shown in Figure 3(a), the coupled predictions improve anatomical plausibility compared to independent predictions.

Figures 3(b-c) illustrate the use of the framework for unsupervised salient keypoint discovery over an entire cardiac cine MRI sequence, propagating correspondence probabilities from ED over all intermediate frames to ES. The entropy map in Figure 3(b) was computed for the ED image within the segmentation masks as predicted by the segmentation model . Keypoints with minimal entropy were selected while enforcing a minimum spatial separation between them. Beyond visual accuracy of the predicted correspondences, the entropy distribution reveals meaningful anatomical patterns: regions with low entropy correspond to geometrically distinctive and stable structures, such as the ventricular insertion areas and the epicardium, whereas higher entropy appears in strongly deforming regions, like the expanding myocardium. These findings suggest that entropy provides a meaningful proxy for anatomical saliency, enabling automatic identification of stable and variable regions. This highlights a key advantage of the proposed framework: it not only predicts correspondences but also offers interpretable uncertainty information that can guide downstream analysis.

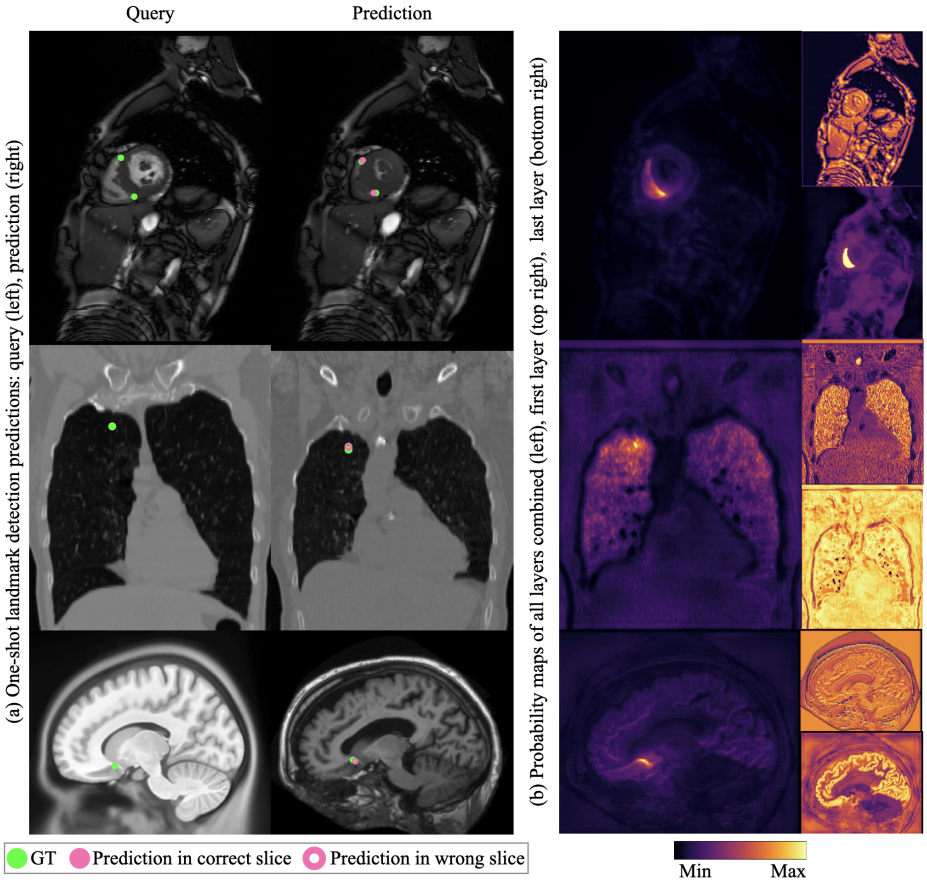

Fig. 2. (a) Example one-shot landmark detections with average performance for each application considered: cardiac MRI, lung CT, brain MRI. Queried landmarks are indicated on the left images and the detected and ground-truth correspondences on the right images. (b) Probability maps of the combination of all layers, the first layer only and the last layer only for the examples in (a) (for one of the two points in cardiac MRI).

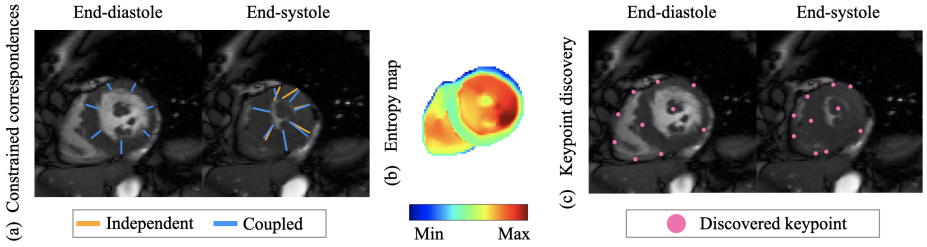

Fig. 3. (a) Constrained correspondence estimation on cardiac MRI with independent versus coupled results. Unsupervised keypoint discovery on cine cardiac MRI with (b) the corresponding entropy map and (c) the discovered salient keypoints.

5 Conclusion

We propose a generic and training-free framework for interpretable correspondence estimation by leveraging multi-scale deep features from a pretrained segmentation model. Experiments across cardiac MRI, lung CT and brain MRI show that these features implicitly encode rich anatomical information that can be directly repurposed for correspondence estimation with state-of-the-art performance across various anatomies and modalities, without requiring task-specific training or additional annotations. By producing entire correspondence probability distributions rather than only the correspondence itself, the framework is inherently interpretable and enables uncertainty-aware analysis of anatomical variability. As illustrated, the correspondence framework can serve as a generic foundation for diverse tasks such as shape and motion analysis.

Acknowledgments. This work is supported by the Flemish Government under the 'Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen' programme. Data was provided in part by Open Access Series of Imaging Studies (Principal Investigators: D. Marcus, R. Buckner, J. Csernansky J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Bai, X., Bai, F., Huo, X., Ge, J., Lu, J., Ye, X., Shu, M., Yan, K., Xia, Y.: UAE: Universal Anatomical Embedding on multi-modality medical images. Medical Image Analysis **103**, 103562 (2025)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging **38**(8), 1788–1800 (2019)
3. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Transactions on Medical Imaging **37**(11), 2514–2525 (2018)
4. Falta, F., Großbröhmer, C., Hering, A., Bigalke, A., Heinrich, M.: Lung250M-4B: a combined 3D dataset for CT-and point cloud-based intra-patient lung registration. Advances in Neural Information Processing Systems **36**, 54819–54832 (2023)
5. Fischl, B.: FreeSurfer. NeuroImage **62**(2), 774–781 (2012)
6. Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., Collins, D.: Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage **47**, S102 (2009)
7. Grewal, M., Wiersma, J., Westerveld, H., Bosman, P.A., Alderliesten, T.: Automatic landmark correspondence detection in medical images with an application to deformable image registration. Journal of Medical Imaging **10**(1), 014007 (2023)
8. Han, L., Chen, B., Zhang, Z., Guo, L., Xu, S.: SCS: Spatially Consistent Self-supervised approach for one-shot anatomical landmark detection. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2025). pp. 1–5 (2025)

9. Heinrich, M.P., Hansen, L.: VoxelMorph++ going beyond the cranial vault with keypoint supervision and multi-channel instance optimisation. In: International Workshop on Biomedical Image Registration. pp. 85–95. Springer (2022)
10. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
11. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnU-Net revisited: A call for rigorous validation in 3D medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2024). pp. 488–498. Springer (2024)
12. Koehler, S., Sharan, L., Kuhm, J., Ghanaat, A., Gordejeva, J., Simon, N.K., Grell, N.M., André, F., Engelhardt, S.: Comparison of evaluation metrics for landmark detection in CMR images. In: Bildverarbeitung für die Medizin: Proceedings, German Workshop on Medical Image Computing, Heidelberg, June 26-28, 2022. pp. 198–203. Springer (2022)
13. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience **19**(9), 1498–1507 (2007)
14. Miao, J., Chen, C., Zhang, K., Chuai, J., Li, Q., Heng, P.A.: FM-OSD: foundation model-enabled one-shot detection of anatomical landmarks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2024). pp. 297–307. Springer (2024)
15. Salari, S., Harirpoush, A., Rivaz, H., Xiao, Y.: CABLD: Contrast-Agnostic Brain Landmark Detection with Consistency-Based Regularization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 20991–21002 (2025)
16. Taha, A., Gilmore, G., Abbass, M., Kai, J., Kuehn, T., Demarco, J., Gupta, G., Zajner, C., Cao, D., Chevalier, R., et al.: Magnetic resonance imaging datasets with anatomical fiducials for quality control and registration. Scientific Data **10**(1), 449 (2023)
17. Tang, L., Jia, M., Wang, Q., Phoo, C.P., Hariharan, B.: Emergent correspondence from image diffusion. Advances in neural information processing systems **36**, 1363–1389 (2023)
18. Yan, K., Cai, J., Jin, D., Miao, S., Guo, D., Harrison, A.P., Tang, Y., Xiao, J., Lu, J., Lu, L.: SAM: Self-supervised learning of pixel-wise anatomical embeddings in radiological images. IEEE Transactions on Medical Imaging **41**(10), 2658–2669 (2022)
19. Yao, Q., Quan, Q., Xiao, L., Kevin Zhou, S.: One-shot medical landmark detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2021). pp. 177–188. Springer (2021)