Learn2Reg 2024: New Benchmark Datasets Driving Progress on New Challenges

Lasse **Hansen** [1] Wiebke **Heyer** [2] Christoph **Großbröhmer** [2] Frederic **Madesta** [3,4] Thilo **Sentker** [3,4] Wang **Jiazheng** [5] Yuxi **Zhang** [5] Hang **Zhang** [6] Min **Liu** [5] Junyi **Wang** [7] Xi **Zhu** [7] Yuhua **Li** [8] Liwen **Wang** [8] Daniil **Morozov** [9,10] Nazim **Haouchine** [9] Joel **Honkamaa** [11] Pekka **Marttinen** [11] Yichao **Zhou** [12] Zuopeng **Tan** [12] Zhuoyuan **Wang** [13] Yi **Wang** [13] Hongchao **Zhou** [14] Shunbo **Hu** [14] Yi **Zhang** [15] Qian **Tao** [15] Lukas **Förner** [16] Thomas **Wendler** [16] Bailiang **Jian** [10, 17] Christian **Wachinger** [10, 17] Jin **Kim** [18] Dan **Ruan** [18] Marek **Wodzinski** [19,20] Henning **Müller** [21,22] Tony C.W. **Mok** [23] Xi **Jia** [24] Jinming **Duan** [24] Mikael **Brudfors** [25] Seyed-Ahmad **Ahmadi** [26] Yunzheng **Zhu** [27] William **Hsu** [27] Tina **Kapur** [9] William M. **Wells** [9] Alexandra **Golby** [9] Aaron **Carass** [28] Harrison **Bai** [28] Yihao **Liu** [29] Perrine **Paul-Gilloteaux** [30] Joakim **Lindblad** [31] Nataša **Sladoje** [31] Andreas **Walter** [32] Junyu **Chen** [28] Reuben **Dorent** [9,33] Alessa **Hering** [34,*] Mattias P. **Heinrich** [2,*]

1 EchoScout GmbH, Lübeck, DE **2** Institute of Medical Informatics, Universität zu Lübeck, Lübeck, DE **3** Institute of Applied Medical Informatics, University Medical Center Hamburg-Eppendorf, Hamburg, DE **4** Institute of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, DE **5** School of Artificial Intelligence and Robotics, Hunan University, Changsha, CN **6** Cornell University, New York, US **7** University of Electronic Science and Technology of China, Chengdu, CN **8** Mechanical Engineering Department, Tianjin University, Tianjin, CN **9** Harvard Medical School, Brigham and Women's Hospital, Boston, US **10** Technical University of Munich, Munich, DE **11** Aalto University, Espoo, FI **12** Canon Medical Systems (China) Co., Ltd., Beijing, CN **13** Smart Medical Imaging, Learning and Engineering (SMLE) Lab, School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University, Shenzhen, CN **14** School of Information Science and Engineering, Linyi University, Linyi, CN **15** Department of Imaging Physics, Delft University of Technology, Delft, NL **16** Clinical Computational Medical Imaging Research, Department of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Augsburg, Augsburg, DE **17** TUM Klinikum Rechts der Isar, Munich, DE **18** University of California, Los Angeles, US **19** AGH University of Krakow, Krakow, PL **20** Sano Centre for Computational Medicine, Krakow, PL **21** Institute of Informatics, University of Applied Sciences Western Switzerland, Sierre, CH **22** Medical Faculty, University of Geneva, Geneva,CH **23** Hong Kong University of Science and Technology, Hong Kong, HK **24** School of Computer Science, University of Birmingham, Birmingham, UK **25** NVIDIA Ltd., London, UK **26** NVIDIA GmbH, Munich, DE **27** Medical & Imaging Informatics, Department of Radiological Science, University of California, Los Angeles, US **28** Johns Hopkins University, Baltimore, US **29** Vanderbilt University, Nashville, US **30** Nantes Université, CHU Nantes, CNRS, Inserm, BioCore, US16, SFR Bonamy, F 44000 Nantes, FR **31** Uppsala Universitet, Uppsala, SE **32** Aalen University, Aalen, DE **33** National Institute for Research in Computer Science and Control, Paris, FR **34** Radboud University Medical Center, Nijmegen, NL

* These authors contributed equally to this work and share senior authorship.

Abstract

Medical image registration is critical for clinical applications, and fair benchmarking of different methods is essential for monitoring ongoing progress in the field. To date, the Learn2Reg 2020-2023 challenges have released several complementary datasets and established metrics for evaluations. Building on this foundation, the 2024 edition expands the challenge's scope to cover a wider range of registration scenarios, particularly in terms of modality diversity and task complexity, by introducing three new tasks, including large-scale multi-modal registration and unsupervised inter-subject brain registration, as well as the first microscopy-focused benchmark within Learn2Reg. The new datasets also inspired new method developments, including invertibility constraints, pyramid features, keypoints alignment and instance optimisation. Visit Learn2Reg at `https://learn2reg.grand-challenge.org`.

Keywords

Data Challenges, (Bio-) Medical Image Registration

Article informations

1. Introduction

This paper serves as the official summary and documentation of the Learn2Reg 2024 challenge, held in conjunction with MICCAI 2024. As an overview paper, it introduces three new benchmark datasets addressing previously underrepresented registration scenarios, documents the methods developed by participating teams, and provides a comparative analysis of their performance across tasks. While this paper offers a comprehensive cross-task perspective on methods and results, individual task organizers may publish separate, in-depth analyses focusing on specific research questions, extended evaluations, or foundational model development. Our primary contributions in this challenge overview include: (1) establishing new publicly available benchmarks for multimodal intraoperative registration, large-scale unsupervised brain alignment, and histopathology microscopy registration; (2) documenting and comparing diverse methodological approaches across these tasks; and (3) identifying current limitations and future directions for medical image registration research through cross-task insights.

The motivation behind the Learn2Reg 2024 benchmark stems from the imperative need to enhance the robustness and versatility of medical image registration methodologies. Earlier challenges revealed limitations in dataset scale, modality diversity, and task complexity, highlighting the need for more comprehensive benchmarks. The 2024 edition addresses these gaps by introducing new tasks that include large-scale multi-modal and cross-domain registration, encouraging the development of more generalizable and adaptive algorithms.

One of the key new tasks is ReMIND2Reg, which focuses on the multimodal alignment of intraoperative ultrasound (iUS) and MRI for brain tumor surgery. Compared to the earlier CuRIOUS2020 task, ReMIND2Reg provides a much larger dataset, increasing from a few dozen to several hundred scan pairs. Moreover, while CuRIOUS2020 targeted the alignment of pre-dura opening 3D iUS to pre-operative MRI, ReMIND2Reg tackles the more challenging alignment of post-resection iUS with pre-operative MRI. This setting reflects real-world surgical challenges and introduces greater anatomical variability, thereby promoting the development of more precise and robust registration methods for neuro-surgical applications.

The second task introduced is LUMIR, which focuses on large-scale inter-subject brain alignment in T1-weighted MRI and is designed without label supervision. It directly addresses a key limitation of earlier datasets such as OASIS, where the reliance on automatically generated labels risked algorithm overfitting and reduced generalizability. By removing label supervision, LUMIR encourages the development of more generalized and robust models that can
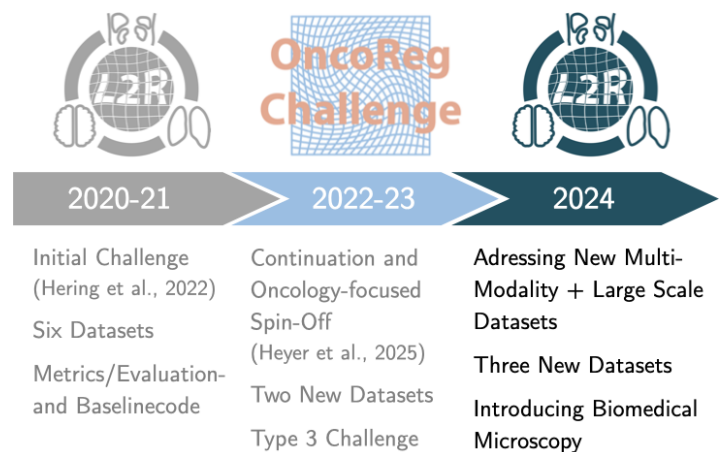
better adapt to variations in brain structures.

Figure 1: Learn2Reg 2024 continues the Learn2Reg Challenge series, which began in 2020, by introducing three new diverse datasets and expanding the challenge's scope to cover a wider range of biomedical imaging applications.

Lastly, The COMULISglobe task introduces the first Learn2Reg benchmark focused on histopathology microscopy, marking a major expansion beyond radiology-based tasks. It represents a significant expansion of Learn2Reg's scope, catering to a wider array of clinical needs and fostering the advancement of cutting-edge techniques such as invertibility constraints, pyramid features, and keypoints alignment. These developments are crucial for achieving superior registration performance and advancing medical imaging applications. For example, they enable informative multimodal characterization of complex interactions in the tumor microenvironment, leading to improved personalized medicine strategies through accurate tracking of morphological changes over time.

2. Related Works

2.1 Medical Image Registration Challenges

Several notable medical image registration challenges have significantly advanced the field. The EMPIRE10 challenge focused on lung CT registration, offering comprehensive evaluations including landmark distances and Jacobian determinants (Murphy et al., 2011). The CuRIOUS challenge addressed intra-operative ultrasound and MRI registration (Xiao et al., 2019), while ANHIR specialized in histological image registration (Borovec et al., 2020), each attracting multiple participating teams and employing diverse evaluation metrics. The Continuous Registration Challenge, co-hosted with WBIR 2018, aimed to unify multiple registration tasks but was limited by its integration requirements, specifically the mandatory use of the SuperElastix C++ API framework (Marstal et al., 2019). More recently, the ACROBAT challenges in 2022 and 2023 addressed whole-

slide image registration for breast cancer tissue, utilizing large datasets and providing valuable insights into algorithm performance (Weitz et al., 2024). These challenges have collectively advanced the development and benchmarking of medical image registration algorithms, each contributing unique datasets and frameworks that have influenced Learn2Reg tasks and evaluation.

2.2 Learn2Reg to Date

Since its inception, the Learn2Reg challenge has served as a key platform for advancing medical image registration through open benchmarking and community collaboration. Over multiple editions, the challenge has brought together researchers worldwide to tackle diverse registration problems across various imaging modalities and anatomical sites.

The Learn2Reg series began with the Learn2Reg Tutorial at MICCAI 2019 in Shenzhen, which introduced the community to emerging deep learning approaches for medical image registration (Learn2Reg Organisers, 2019). This foundational event set the stage for the challenge's collaborative and multidisciplinary focus.

The 2020 and 2021 editions of Learn2Reg were the initial official challenges, introducing six datasets covering multiple anatomies—including brain, abdomen, and thorax—and imaging modalities such as MRI, CT, and ultrasound. Both editions were held as community challenges with workshops at the respective MICCAI conferences, following a Type 1 challenge format with training images and labels provided, while test labels remained withheld for evaluation. These editions introduced the challenge's open-source evaluation and ranking code, as well as baseline methods, enabling standardized and reproducible benchmarking. Together, they attracted over 20 international teams submitting more than 65 methods, offering comprehensive insights into the performance of conventional and deep-learning registration approaches across multiple clinically relevant tasks (Hering et al., 2022).

In 2022 and 2023, Learn2Reg held two further challenge workshops at MICCAI, focusing on continuation and expansion of the initial benchmarks. These editions introduced the NLST dataset (CT Lung Screening Registration Data) and pushed towards Type 3 challenges for selected tasks, where both training and test images, along with their labels, remain hidden during evaluation to better simulate real-world conditions.

In parallel, the OncoReg spin-off challenge was launched at MICCAI 2023, introducing the ThoraxCBCT dataset with longitudinal lung CBCT and FBCT scans from clinical settings. OncoReg targets oncological applications to advance AI-driven registration for monitoring lung cancer progression and improving radiotherapy planning (Heyer et al., 2025).

The 2024 edition continues this trajectory, expanding into new imaging domains including intraoperative multimodal imaging and microscopy, while addressing increasingly complex registration problems.

3. Materials and Methods

3.1 Challenge Organisation

The Learn2Reg 2024 challenge was organised by a dedicated core team led by Mattias Heinrich (University of Lübeck), who coordinated all aspects of the event, including website maintenance, evaluation infrastructure, and the associated workshop. The challenge consisted of three tasks, each developed and maintained by a separate expert team: ReMIND2Reg (Harvard Medical School, led by Reuben Dorent), LUMIR (Johns Hopkins University, led by Junyu Chen), and COMULISglobe SHG/BF (COMULISglobe community, led by Andreas Walter, Hochschule Aalen, (Walter et al., 2021; Rudraiah et al., 2024; Walter et al., 2020)). These tasks addressed diverse challenges in biomedical image registration and are described in detail in later sections.

Returning to a type 2 challenge format, Learn2Reg 2024 followed a two-phase structure hosted on the Grand Challenge platform[1]. During the validation phase, participants were given access to training and validation data to develop and test their methods locally. In the subsequent test phase, participants submitted their final algorithms, which were evaluated on hidden test datasets. The evaluation was performed using a standardized, open-source framework[2], ensuring fairness, reproducibility, and transparency. Performance was assessed using task-specific quantitative metrics, with results published on public leaderboards. The ranking scheme is based on the ranking scheme of the Medical Decathlon (statistical comparisons of method performance across metrics, aggregated into overall task scores) (Antonelli et al., 2022).

For the first time, Learn2Reg was part of the WBIR (Workshop on Biomedical Image Registration), which was officially aligned with MICCAI and held as an on-site half-day workshop at MICCAI 2024 in Marrakesh. Top-performing teams were invited to present their methods and findings during this event.
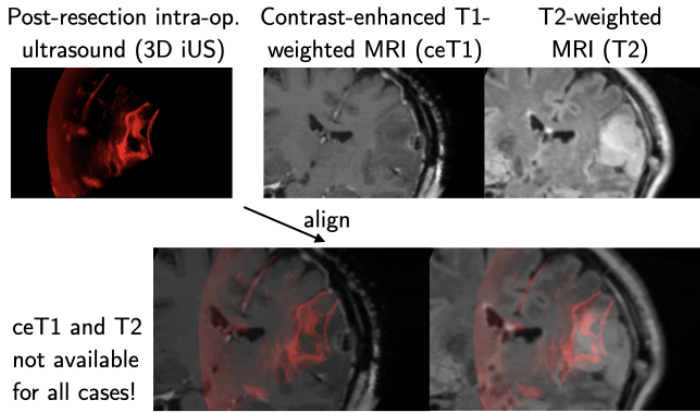
3.2 Tasks

3.2.1 ReMIND2Reg

ReMIND2Reg focuses on a clinically critical but technically complex task: registering preoperative brain MRI to intraoperative ultrasound images after tumor resection. Intraoperative ultrasound (iUS) offers real-time imaging capabilities
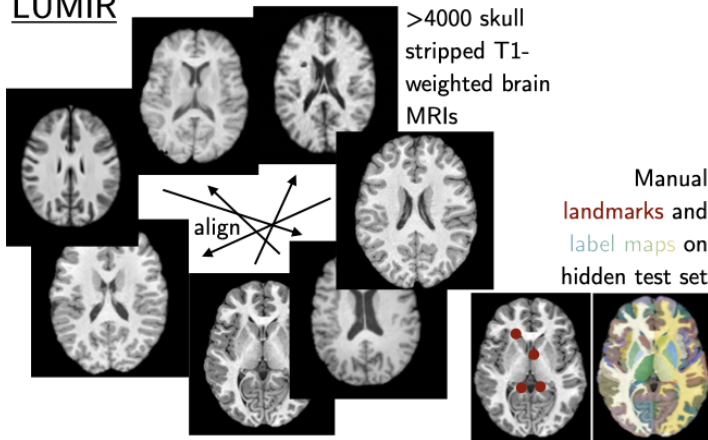
---

1. `https://learn2reg.grand-challenge.org/`
2. `https://github.com/MDL-UzL/L2R`

ReMIND2Reg

Post-resection intra-op. ultrasound (3D iUS)

Contrast-enhanced T1-weighted MRI (ceT1)

T2-weighted MRI (T2)

align

ceT1 and T2 not available for all cases!

LUMIR

>4000 skull stripped T1-weighted brain MRIs

align

Manual landmarks and label maps on hidden test set

COMULISglobe SHG/BF

Second-harmonic generation (SHG) microscopy
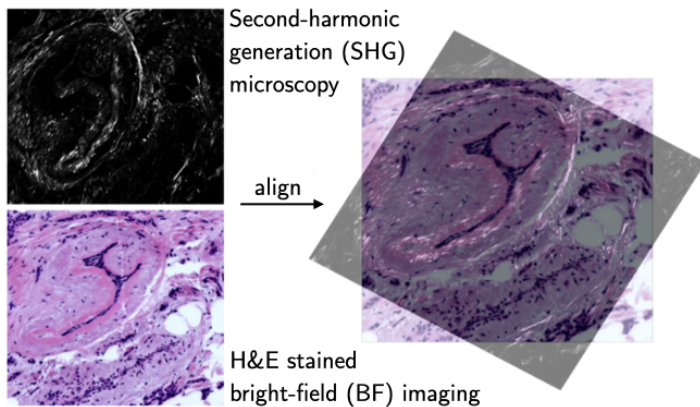
align

H&E stained bright-field (BF) imaging

Figure 2: Learn2Reg 2024 tasks at a glance. This overview illustrates the diversity of the three new tasks in Learn2Reg 2024. ReMIND2Reg challenges participants with intraoperative ultrasound-to-MRI brain registration, addressing brain shift and flexible input modalities. LUMIR focuses on large-scale, inter-subject brain alignment without label supervision, using over 4,000 scans from 10 input sources. COMULISglobe SHG/BF introduces the first histology-based task in the Learn2Reg series, aligning complementary breast tissue modalities—second harmonic generation (SHG) and brightfield (BF) microscopy. All tasks include hidden test data and are evaluated using manual landmarks, label maps, or both.

during surgery but differs significantly in appearance and resolution from preoperative MRI, complicating direct multimodal registration. The surgical process itself induces brain shift and structural deformation, making static navigation systems inaccurate as surgery progresses.

The challenge task is to register post-resection 3D iUS volumes to either contrast-enhanced T1-weighted (ceT1) or T2-weighted MRI, depending on availability for each case. Submitted methods must be able to handle missing modalities and account for large deformations, resections, and noise inherent in iUS data. The dataset is derived from the ReMIND collection (Juvekar et al., 2023) at Brigham and Women's Hospital. Specifically, the ReMIND2Reg dataset includes 99 training cases with 99 3D iUS, 93 ceT1, and 62 T2 scans, 4 validation cases with 4 3D US, 4 ceT1, and 4 T2, and 10 test cases with 10 3D US, 10 ceT1 and 10 T2.

All data have been resampled to 0.5 mm isotropic spacing and MRI modalities co-registered where applicable. The MRI sequences used include contrast-enhanced T1 and T2-SPACE, while ultrasound was acquired using tracked 2D sweeps reconstructed into 3D volumes. Manual landmark annotations created by multiple expert raters serve as the ground truth for validation and testing. For evaluation, the target registration error (TRE) were computed. Participants submit Docker containers for test evaluation.

Full data and acquisition protocols are described in (Juvekar et al., 2023). The dataset is shared under a CC BY 4.0 license and is available via TCIA[3].

3.2.2 LUMIR

LUMIR (Large Scale Unsupervised Brain MRI Image Registration) addresses the alignment of T1-weighted brain MRIs from different subjects in an unsupervised setting. The task builds on previous editions of the Learn2Reg challenge and aims to overcome limitations observed when training with anatomical label maps, which can result in overfitting to labels and non-smooth transformations. By omitting label supervision during training, the challenge promotes the development of models that generalize better and produce realistic, diffeomorphic deformations. This shift also allows fair comparison with classical optimization-based registration algorithms.

Participants are given access to a preprocessed large-scale dataset of 3,384 training images compiled from ten public neuroimaging datasets (OpenBHB) and an additional set from the AFIDs-OASIS dataset. Each MRI has been resampled to 1 mm isotropic resolution and standardized to a shape of $160 \times 224 \times 192$ voxels. Label maps and anatomical landmarks are retained for evaluation only and are not made available during training. Validation data includes

---

3. https://doi.org/10.1101/2023.09.14.23295596

40 images from OpenBHB and 10 with manually placed fiducials from AFIDs-OASIS. Testing is based on 590 hidden images, with evaluation pairs defined by the organizers. Among them, 20 images are from AFIDs-OASIS with manually annotated landmarks, and 110 images were annotated in-house by a team of physicians led by an experienced radiologist (H. Bai) following the AFIDs protocol. The remaining test images are accompanied by label maps of 133 anatomical structures automatically generated using the SLANT method (Huo et al., 2019).

Participants are required to submit deformation fields aligning specified image pairs. The performance is assessed via Dice similarity and Hausdorff-95 distance on segmentation labels, target registration error on anatomical landmarks, and the fraction of non-diffeomorphic volumes (NDV) (Liu et al., 2024b) to measure deformation plausibility. Baseline code and evaluation scripts are provided online[4]. An introductory video is also available[5].

This section documents LUMIR within the Learn2Reg 2024 challenge framework. As a multi-task overview, we focus on describing the task design, documenting participating methods, presenting comparative rankings, and identifying cross-task insights, particularly regarding the impact of dataset scale on classical versus deep learning approaches. The value of including LUMIR lies in enabling direct comparison with ReMIND2Reg and COMULISglobe, and providing a complete record of all Learn2Reg 2024 tasks. A complementary in-depth analysis with extensive ablation studies, zero-shot generalization experiments, and foundational model evaluation is provided in a dedicated paper (Chen et al., 2025). Relevant datasets and prior methods are further detailed in (Dufumier et al., 2022; Taha et al., 2023; Marcus et al., 2007; Liu et al., 2024b).

3.2.3 COMULISglobe SHG/BF

COMULISglobe SHG/BF addresses the registration of histopathology images acquired using two fundamentally different microscopy techniques: second-harmonic generation (SHG) and bright-field (BF) imaging of H&E-stained tissue. SHG provides high-resolution structural information of collagen fibers without the need for staining, while BF microscopy reveals cellular and tissue structure via conventional staining. Registering these modalities is essential for correlating structural and pathological insights and is particularly valuable in cancer research where collagen architecture plays a key role.

This task involves registering co-located SHG and BF images of breast and pancreatic cancer tissue sections. Tissue was sectioned at $5\,\mu m$ thickness, stained, and mounted

before imaging. SHG images were acquired at three axial depths and processed with maximum-intensity projection to compensate for focal plane inconsistencies. BF imaging was performed using commercial and custom-built platforms at 40x magnification.

Participants work with 156 training samples, 10 for validation, and 40 for testing. Manual landmark annotations are available for validation and test sets. The challenge evaluates registration accuracy using target registration error, and participants submit displacement fields as their output. This task is particularly relevant for enabling virtual staining and rapid multimodal analysis in digital pathology, where SHG microscopy is gaining traction as a label-free diagnostic alternative.

Further details on the underlying imaging and dataset are available in (Keikhosravi et al., 2020). More information on the goals of the COMULISglobe initiative and the broader community can be found at the COMULISglobe website[6].

3.2.4 NLST

In addition to the three newly introduced tasks in Learn2Reg 2024, this paper also documents the ongoing NLST task, first launched in 2022 and expanded in 2023.

The NLST (Thoracic Image Registration for Lung Cancer) task targets a highly impactful clinical application: aligning intra-patient thoracic CT scans to facilitate lung nodule tracking and support therapy planning. Accurate registration of baseline and follow-up low-dose CT scans is essential in lung cancer screening programs to assess the progression of pulmonary nodules and metastatic lesions. Manual identification of corresponding structures is laborious and error-prone, particularly in the lung where many anatomical features appear similar. Deformable registration algorithms can significantly streamline this process and enable automated propagation of findings across time points.

Beyond longitudinal screening, the task also includes a radiotherapy sub-cohort with 4DCT scans, where precise alignment is needed for treatment planning and the mapping of structures such as the trachea, esophagus, spinal cord, and heart. These organs-at-risk must be accurately registered to intraoperative cone-beam CT scans to ensure safe and effective dose delivery.

The challenge consists of two phases. In Phase 1, participants developed and tested algorithms locally using 110 training and 22 validation cases, all sourced from the National Lung Screening Trial (NLST) (Team, 2011) [7]. Each case consists of a pair of thoracic CT scans acquired either longitudinally (screening cohort) or across different respi-

---

4. https://github.com/JHU-MedImage-Reg/LUMIR_L2R

5. https://cloud.imi.uni-luebeck.de/s/tm9DEFiDa9XH35k/
   download/Learn2Reg_LUMIR_JChen.mp4

6. https://www.comulis.eu/

7. https://wiki.cancerimagingarchive.net/display/NLST/
   National+Lung+Screening+Trial

ratory phases (radiotherapy cohort). Automatic keypoint correspondences (Heinrich et al., 2015) and lung masks are provided to support model development. In Phase 2, the top-performing teams were invited to submit training Docker containers for remote execution on a hidden, extended dataset comprising 1300 training, over 100 validation, and more than 200 test cases. This second part of the dataset includes additional annotations not publicly released.

The evaluation framework measures registration performance using multiple metrics: target registration error (TRE) on lung nodule centers and manually annotated geometric landmarks (e.g. at bifurcations), deformation plausibility via the standard deviation of log Jacobian determinants, and runtime. Robustness is also assessed by considering the 30% worst-performing cases per submission.

All data are preprocessed to standardized voxel spacing and affine-aligned. The CT data originate from over 30 U.S. medical centers participating in the NLST and from NSCLC radiotherapy patients with fan-beam 4DCT scans. Segmentations were created using CIRRUS lung screening software, with nodule locations derived from the center of gravity of annotated lesions. The dataset reflects diverse demographics and scanning equipment, adding to its clinical realism.

The NLST task exemplifies the challenges of real-world thoracic image registration and is designed to evaluate both accuracy and generalisability of methods under clinically relevant constraints.

4. Results and Discussions

This section presents the results and key insights from each of the three Learn2Reg 2024 tasks as well as from NLST. For each task, we summarize the test setting, describe the participating teams and their approaches, highlight the top-performing methods, and discuss relevant findings, limitations, and future directions. The final subsection offers a cross-task analysis, including observations on team participation across tasks and overall challenge outcomes.

4.1 Tasks

4.1.1 ReMIND2Reg

During the validation phase of the ReMIND2Reg task, a total of thirteen teams submitted results, along with two baseline methods provided by the challenge organizers. The task's complexity was reflected in the exceptionally high number of validation entries, with 253 submissions indicating that teams required extensive iterative optimization to handle the challenging multimodal ultrasound-to-MRI registration problem. In the test phase, participation reduced to five teams plus one baseline method. This reduction high-

lights the technical difficulty of registering post-resection intraoperative ultrasound to preoperative MRI, particularly given the presence of brain shift and tissue deformation. As a baseline, niftyreg was provided for the ReMIND2Reg test phase. The submitting teams included junyi-wang, VROC, next-gen-nn, IGRTS, and SamPLe.

niftyreg, which employed the NiftyReg open-source software package[8] implementing a symmetric block-matching registration approach. This method used iterative block-matching to establish point correspondences between ultrasound and MRI images, followed by least trimmed squares (LTS) regression to determine transformation parameters. To identify informative regions, the algorithm divided images into uniform blocks and selected those with the highest intensity variance. Block similarity was assessed using normalized cross-correlation, and matching was performed based on these scores. To ensure inverse consistency, the method averaged directional transformation matrices at each iteration, maintaining symmetry throughout the registration process (Drobny et al., 2018).

junyi-wang developed a coarse-to-fine pipeline combining imaging style transfer with traditional registration

---

8. https://github.com/KCL-BMEIS/niftyreg/wiki

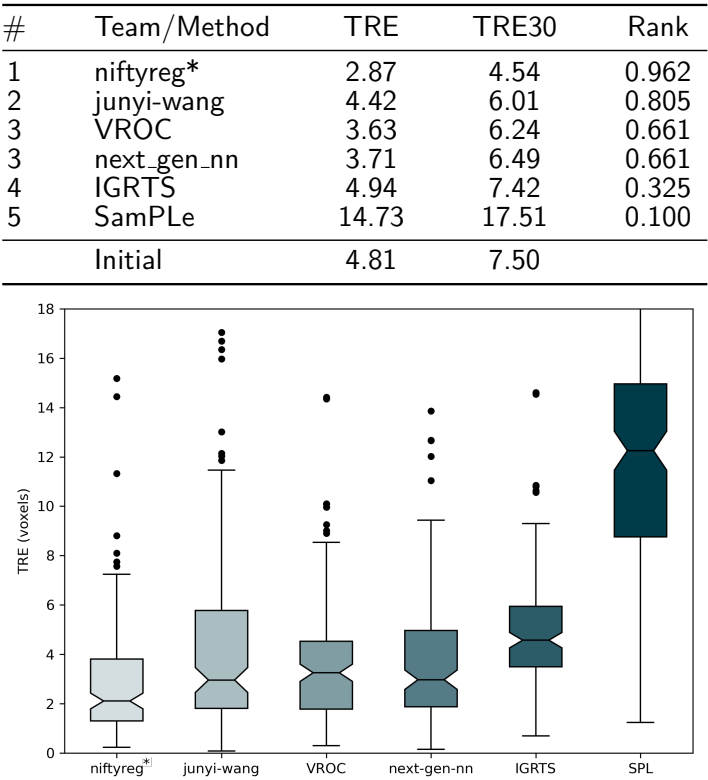| #	Team/Method	TRE	TRE30	Rank
1	niftyreg*	2.87	4.54	0.962
2	junyi-wang	4.42	6.01	0.805
3	VROC	3.63	6.24	0.661
3	next_gen_nn	3.71	6.49	0.661
4	IGRTS	4.94	7.42	0.325
5	SamPLe	14.73	17.51	0.100
	Initial	4.81	7.50	

Figure 3: Results for the ReMIND2Reg task. Target Registration Error (TRE) on manual landmarks and TRE30 (the TRE computed on the 30% of landmarks with the largest initial errors) are reported for all methods. * Baseline method provided by organizers.

techniques. Their approach used 3D CycleGAN to generate synthetic T1-style images from ultrasound data, enabling unified signal distributions across modalities. The registration process employed hierarchical block matching via NiftyReg for affine transformation estimation, followed by a pretrained SynthMorph model for local deformation inference around resection areas. **VROC** employed its fully differentiable variational registration framework (Werner et al., 2014; Ehrhardt et al., 2015)[9], combining task-specific preprocessing with rigid registration optimization. For the ReMIND2Reg task, VROC applied Gaussian smoothing and threshold-based clipping of ultrasound intensity values during preprocessing. The registration pipeline performed initial vector field estimation via global normalized cross-correlation (NCC) loss optimization, followed by fine-tuning using normalized gradient fields (NGF) loss optimization to handle the challenging multimodal ultrasound-to-MRI registration scenario.

next-gen-nn addressed the ReMIND2Reg challenge using an unsupervised multimodal registration method based on Multilevel Correlation Balanced Optimization (MCBO), building upon the ConvexAdam framework (Siebert et al., 2024). The approach was specifically designed to handle two key difficulties: the absence of supervised labels and the presence of large deformations between pre-operative MRI and intra-operative ultrasound images. For paired multimodal moving images (ceT1 and T2), the method adopted maximum fusion strategies to optimize registration performance across the challenging modality gap. **IGRTS** developed a topological higher-order Markov Random Field approach for deformable registration. Their method utilized multiscale multimodal Gabor attribute vectors to characterize MRI and ultrasound images, with the 3D deformation field's topological framework constructed through Jacobian determinants derived from higher-order cliques. The optimization employed a multi-resolution QPBO (Quadratic Pseudo-Boolean Optimization) strategy to handle the complex higher-order MRF energy function.

SamPLe implemented a cross-modal feature matching approach using 3D keypoint descriptors specifically designed for MR-ultrasound registration. Their method trained a Siamese convolutional neural network on 3D patches extracted at keypoint locations using supervised contrastive learning with triplet loss. The approach detected keypoints using 3D SIFT, enforced repeatability across modalities, and learned cross-modal discriminant features robust to ultrasound texture changes and speckle noise. Final registration was performed by solving a least-squares problem using matched keypoint correspondences.

The results demonstrate the substantial difficulty of the

---

9. https://github.com/IPMI-ICNS-UKE/vroc

ReMIND2Reg task, with most methods showing modest improvements over the initial misalignment (TRE: 4.81 mm). Notably, the classical optimization-based baseline niftyreg achieved the best overall performance with a TRE of 2.87 mm and TRE30 of 4.54 mm, ranking first among all submissions. This finding highlights that well-tuned classical methods can still outperform learning-based approaches for challenging multimodal registration tasks with limited training data. Among the learning-based submissions, junyi-wang ranked second with a TRE of 4.42 mm, followed by VROC and next-gen-nn tied for third place with TREs of 3.63 mm and 3.71 mm respectively. IGRTS ranked fourth with a TRE of 4.94 mm, while SamPLe showed the poorest performance with a TRE of 14.73 mm, indicating substantial registration failures. The wide variation in performance across methods (TRE ranging from 2.87 to 14.73 mm) suggests that robustness remains a significant concern. Several methods achieved TREs that were only marginally better than, or in some cases worse than, the initial misalignment, indicating that the task remains largely unsolved for many approaches. The presence of outliers and failure cases across most methods emphasizes the need for more robust algorithms capable of handling the substantial anatomical deformations, tissue resection, and multimodal appearance differences inherent in this clinical scenario. The strong performance of the classical niftyreg baseline suggests that the explicit modeling of correspondences through block-matching and the iterative refinement process may be more effective than end-to-end learning approaches for this particular challenge. The substantial drop in participation from validation (13 teams) to test phase (5 teams) reflects the technical difficulty of the task. This pattern differs from other tasks in the challenge and underscores the need for continued research into multimodal intraoperative registration methods that can reliably handle brain shift and tissue deformation in clinical settings. Future work should focus on improving robustness through better handling of outliers, incorporating anatomical priors specific to brain tumor surgery, and developing hybrid approaches that combine the strengths of classical optimization with learning-based feature extraction and similarity measures.

4.1.2 LUMIR

During the validation phase of the LUMIR task, a total of eighteen teams submitted results, with four baseline methods provided by the challenge organizers. In the test phase, the number of submitting teams reduced to eleven, while the organizers expanded the baseline comparisons to include four additional methods, providing eight baselines in total. This comprehensive baseline coverage enabled thorough benchmarking across diverse registration approaches. Notably, LUMIR attracted strong participation. The high

| #	Team/Method	TRE	DSC	HD95	NDV	Rank
1	honkamj	3.09	0.785	3.04	0.0025	0.814
2	next_gen_nn	3.12	0.777	3.28	0.0001	0.781
3	MadeForLife	3.07	0.778	3.29	0.0121	0.737
4	zhuoyuanw210	3.14	0.773	3.23	0.0045	0.723
5	LYU1	3.13	0.778	3.25	0.0150	0.722
6	DutchMasters	3.11	0.770	3.26	0.0030	0.702
7	uniGradICONiso*	3.14	0.760	3.40	0.0002	0.668
8	VFA*	3.14	0.777	3.15	0.0704	0.667
9	lukasf	3.14	0.764	3.42	0.2761	0.561
10	Bailiang	3.16	0.774	3.33	0.0222	0.526
11	TransMorph*	3.14	0.762	3.46	0.3621	0.518
12	TimH	3.19	0.730	3.57	0.0000	0.487
13	deedsBCV*	3.10	0.696	3.94	0.0002	0.423
14	uniGradICON*	3.24	0.742	3.57	0.0001	0.402
15	LoRA-FT	3.24	0.736	3.73	0.0033	0.384
16	SynthMorph*	3.23	0.722	3.61	0.0000	0.361
17	VROC	3.23	0.760	3.63	0.0475	0.351
18	ANTsSyN*	3.48	0.703	3.69	0.0000	0.265
19	VoxelMorph*	3.53	0.714	4.07	1.2167	0.157
	Initial	4.38	0.555	4.91		
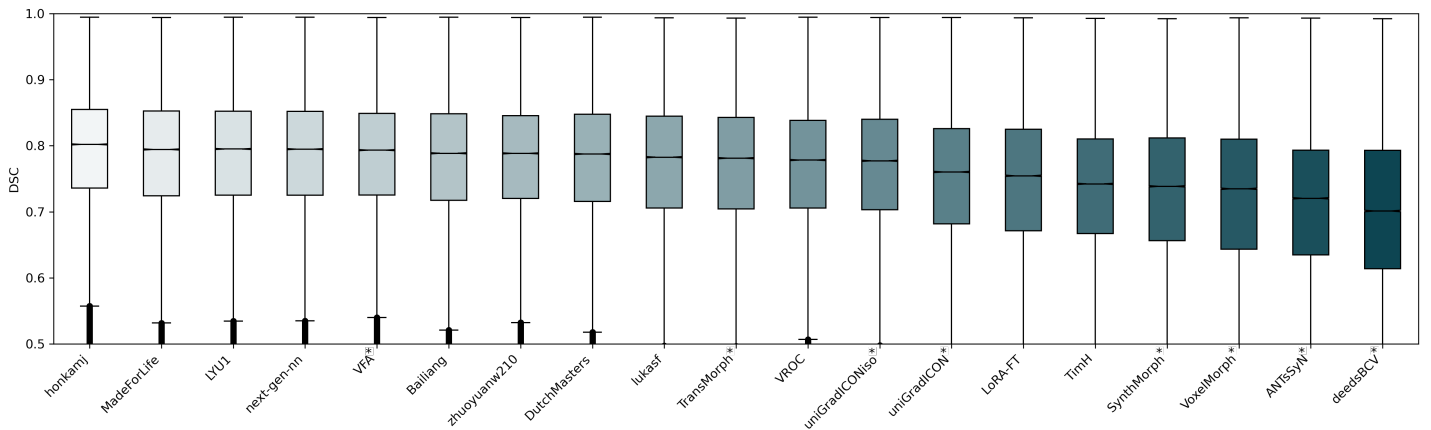
Figure 4: Results for the LUMIR task. The table presents final test set results for all participating teams and baseline methods. Reported metrics include Target Registration Error (TRE) on manual landmarks, Dice-Sørensen coefficient (DSC) and 95th percentile Hausdorff distance (HD95) on brain label maps, as well as the non-diffeomorphic volume (NDV) of displacement fields. The accompanying box plot provides a visual overview of DSC performance distribution across all methods. * Baseline method provided by organizers.

level of engagement reflects the strong community interest in large-scale, unsupervised brain MRI registration tasks.

A distinctive aspect of the LUMIR challenge was its extensive post-challenge analysis, documented in detail in "Beyond the LUMIR challenge: The pathway to foundational registration models" (Chen et al., 2025). This companion paper provides comprehensive validation with particular emphasis on foundational registration models, zero-shot evaluation, and out-of-domain generalization tasks. The present section summarizes the key findings and participating methods, while readers are referred to the companion paper for in-depth technical analysis and broader implications for developing generalizable brain MRI registration frameworks. The LUMIR challenge included eight baseline

methods spanning classical optimization-based approaches and modern deep learning frameworks. **ANTsSyN** employed symmetric, time-varying diffeomorphic registration using cross-correlation within the space of diffeomorphic maps (Avants et al., 2008). **deedsBCV** utilized discrete optimization with Markov random field strategies and minimum spanning trees to handle sliding motions efficiently (Heinrich et al., 2012). **SynthMorph** implemented a contrast-agnostic learning strategy trained exclusively on synthetic data generated from label maps (Hoffmann et al., 2021). **TransMorph** incorporated Vision Transformer architecture with a three-stage Swin Transformer encoder for hierarchical feature extraction (Chen et al., 2022). **uniGradICON** and **uniGradICONiso** represented foundation registration mod-

els using gradient inverse consistency regularization, with the latter including instance-specific optimization (Tian et al., 2024). **VFA** employed dual-stream encoders with vector field cross-attention decoders for pixel-level correspondence retrieval (Liu et al., 2024a). **VoxelMorph** served as the canonical U-Net-based learning framework for fast deformable registration (Balakrishnan et al., 2019). The top-performing methods in LUMIR employed diverse architectural strategies, ranging from specialized ConvNet designs to foundation model adaptations.

Bailiang utilized a dual-stream encoder architecture with registration-specific components, employing coarse-to-fine optimization pyramids and progressive registration to handle large deformations through iterative composition of deformation fields. Correlation layers were used to establish voxel-wise correspondences between moving and fixed images (Jian et al., 2024).

next-gen-nn implemented a robust backbone network (Zhang et al., 2024a) with several architectural enhancements, including co-attention mechanisms (Chen et al., 2021), large kernel convolutions, and increased convolutional channels in the deformation prediction stage. The method was further refined using image-guided bilateral filtering to improve registration accuracy.

honkamj introduced a novel approach focusing on mathematical guarantees, implementing inverse consistency, symmetry, and topology preservation by construction rather than through loss terms. The method employed an auxiliary network formulation where $f(A, B) = u(A, B) \circ u(B, A)^{-1}$, ensuring cycle-consistency across a multi-resolution pipeline. Features were extracted separately before deformation computation, with symmetric halfway deformations composed at the highest resolution (Honkamaa and Marttinen, 2024).

LoRA-FT addressed the computational challenges of foundation model adaptation by applying Low-rank Adaptation (LoRA (Hu et al., 2022)) to uniGradICON foundation models (Tian et al., 2024). This approach decomposed weight updates into low-rank matrices, reducing trainable parameters by orders of magnitude while enabling rapid domain adaptation without catastrophic forgetting or requiring anatomical label supervision.

MadeForLife employed the GroupMorph (Tan et al., 2024) structure with dual encoders and specialized grouping update modules (GUMs). The method decomposed deformation fields into multiple subfields with different receptive fields, facilitated by grouping modules for feature correlation and contextual fusion modules for inter-group communication, enabling finer-grained optimization.

lukasf refined a pre-trained TransMorph (Chen et al., 2022) model using the Fisher Adam (FAdam) optimizer (Hwang, 2024) for improved convergence stability and incorporated gradient correlation in the similarity measure to enhance deformation smoothness and anatomical alignment.

LYU1 implemented a pyramid network architecture with shared encoder and auxiliary decoder components. The fusion pyramid decoder operated across five scales, using multi-scale feature fusion blocks (MSFB) to iteratively combine coarse deformation fields. The method progressively refined deformations by upsampling and applying fields to moving feature maps at each scale, removing redundant information through the MSFB processing.

TimH employed the ConstrICON framework (Greer et al., 2023) to achieve multistep inverse consistency through architectural design rather than penalty terms. The approach used three residual U-Nets in a recursive TwoStep-Consistent (TSC) operator structure, parameterizing transformations via Lie groups. Stationary velocity fields were smoothed using average pooling, and selective Jacobian determinant regularization was applied to mitigate folding.

VROC utilized a classical variational registration approach with task-specific preprocessing, including Gaussian smoothing and multi-Otsu thresholding to generate intensity-based label maps. The method performed initial deformable registration using normalized cross-correlation (NCC) forces on label maps, followed by Demons forces refinement on original intensity images for precise alignment (Werner et al., 2014; Ehrhardt et al., 2015).

DutchMasters introduced a novel instance optimization algorithm addressing multi-objective optimization challenges in registration. The method applied gradient projection techniques to handle conflicting updates between similarity and regularization objectives, projecting conflicting gradients into common spaces when cosine similarity indicated gradient conflicts, while preserving non-conflicting gradient updates (Zhang et al., 2024b).

zhuoyuanw210 employed a dual-stream pyramid encoder paired with a GPU-accelerated motion decomposition Transformer (ModeTv2) decoder (Wang et al., 2024). The dual-stream architecture extracted hierarchical features from both moving and fixed images, generating deformations from coarse to fine across multiple scale levels.

The results demonstrate strong performance across both learning-based and classical approaches, with honkamj achieving the highest overall rank (0.814), followed by next-gen-nn (0.781) and MadeForLife (0.737) completing the top three. Notably, the top-performing methods achieved remarkably similar performance across all metrics, with the leading submissions showing narrow ranges in Target Registration Error (honkamj: 3.09, MadeForLife: 3.07, next-gen-nn: 3.12), Dice scores (0.773-0.785), and Hausdorff distances (3.04-3.29). This tight clustering of results suggests that multiple architectural strategies can achieve comparable accuracy for unsupervised brain MRI registration when properly optimized. The LUMIR challenge highlighted the significant impact of dataset scale on method performance.

With over 4,000 brain MRI scans available for training, a clear trend emerged favoring deep learning approaches over classical optimization-based methods. Non-deep learning methods including VROC (rank: 0.351), ANTsSyN (rank: 0.265), and deedsBCV (rank: 0.423) consistently ranked in the lower half of submissions, contrasting with other Learn2Reg 2024 tasks where classical methods remained competitive with learning-based approaches. Foundation model approaches showed mixed results. While uniGradI-CONiso (rank: 0.668) with instance optimization performed reasonably well, the base uniGradICON model (rank: 0.402) ranked lower, emphasizing the importance of task-specific adaptation strategies. LoRA-FT's parameter-efficient adaptation approach (rank: 0.384) demonstrated the potential for foundation model fine-tuning, though further optimization may be needed to compete with top-performing specialized architectures. The challenge successfully demonstrated the viability of unsupervised approaches for large-scale brain MRI registration, with all methods substantially outperforming the unregistered baseline (initial TRE: 4.38, DSC: 0.555). The removal of label supervision during training encouraged the development of more generalizable models while maintaining competitive registration accuracy, as evidenced by the consistently strong Dice scores (> 0.73) among the top ten methods.

4.1.3 COMULISglobe

During the validation phase of the COMULISglobe task, a total of ten teams, including two baseline methods provided by the challenge organizers, submitted over 120 valid entries on the public leaderboard. The high number of submissions and overall strong performance demonstrated meaningful progress in aligning SHG and BF microscopy images, despite the substantial initial misalignments. In the test phase, the number of submitting teams dropped to five, with mainly the top-performing validation participants proceeding. This potentially reflects a broader issue in challenge design: teams with less competitive validation results may be discouraged from submitting to the test phase, introducing a performance bias and limiting the diversity of approaches evaluated in the final benchmark. Addressing this remains an open challenge in the design of fair and representative benchmarking competitions.

One of the two baseline methods in the test phase was the affine variant of **greedy**, a fast, CPU-based deformable image registration algorithm developed by Paul Yushkevich at the Penn Image Computing and Science Lab. This variant focuses on efficient affine registration using a non-symmetric, gradient descent-based optimization, prioritizing speed over symmetry. While greedy shares conceptual roots with ANTs' SyN (Avants et al., 2008), its streamlined design and optimized computations enable rapid alignment suitable for large-scale tasks. Serving as a classical baseline, greedy provides robust reference for comparing the effectiveness of more complex registration methods in the COMULISglobe challenge task[10].

The second baseline method, **globalign**, performs multimodal image alignment by globally maximizing mutual information (MI) using a novel cross-mutual information function (CMIF) computed efficiently in the frequency domain (Öfverstedt et al., 2022). Designed primarily for low-degree-of-freedom transformations such as rigid alignment, globalign has demonstrated robust and fast performance across various multimodal biomedical image datasets, including histological images. Its algorithmic efficiency enables rapid estimation of global optima, contrasting with typical local MI maximization approaches. For this challenge, the default configuration—featuring a faster but less comprehensive search—was selected as the baseline. Adjustments to the hyperparameters may lead to improved results. The method's open-source implementation is publicly available[11].

One of the three submissions, **VROC**, is a classical variational image registration framework that supports flexible choices of similarity measures, regularization strategies, and transformation models (Werner et al., 2014; Ehrhardt et al., 2015). For this challenge task of multimodal registration of BF and SHG images, VROC combines handcrafted preprocessing with affine alignment and multi-resolution variational refinement. A custom-trained U-Net is used to synthesize SHG-like representations from BF inputs, reducing the modality gap. During inference, VROC performs extensive hyperparameter tuning, including affine step sizes, smoothness parameters, and optimization schedules, optionally preceded by a pre-registration step. The optimization uses normalized cross-correlation (NCC) as similarity metric, and incorporates a Demons-based strategy (Thirion, 1998) for deformable refinement. This highly engineered pipeline enables robust adaptation to the modality-specific challenges of SHG/BF registration. The general VROC framework is publicly available on GitHub[12].

next-gen-nn, employs a hybrid pipeline that combines learning-based feature extraction with classical optimization. The feature backbone, adapted from COMIR (Pielawski et al., 2020), is pretrained to extract modality-invariant representations, facilitating consistent feature alignment across SHG and BF inputs. Prealignment is performed using XFeat (Potje et al., 2024), a sparse feature matcher optimized for multimodal pairs. During training, the mutual information between image pairs is used both as the objective function and as a criterion for selecting optimal

---

10. `https://github.com/hansenlasse/L2R-COMULIS-baseline-greedy`
11. `https://github.com/MIDA-group/globalign`
12. `https://github.com/IPMI-ICNS-UKE/vroc`

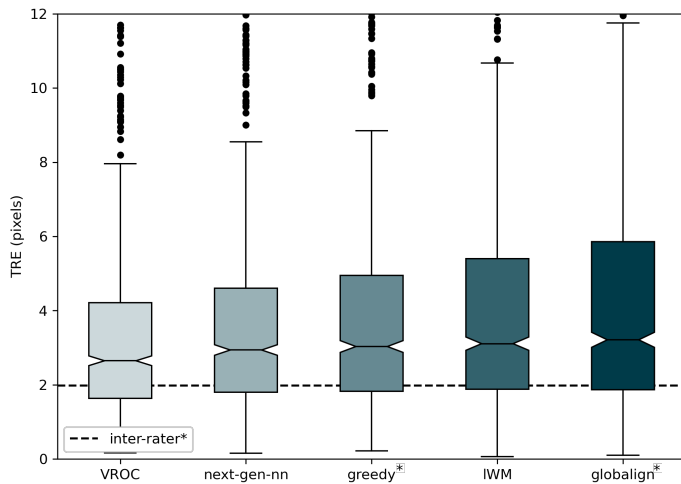| #	Team/Method	TRE	TRE30	Rank
1	VROC	2.64	2.86	0.90
2	next_gen_nn	2.93	3.18	0.77
2	greedy*	3.03	3.29	0.77
3	IWM	3.10	3.42	0.46
4	globalign*	3.21	3.64	0.28
	Initial	69.6	102.	

Figure 5: Results for the COMULISglobe SHG/BF task. *The mean inter-rater error was computed on a random subset of test cases. While thus not directly comparable to the evaluated methods, it provides a useful reference for the level of alignment that may be achievable. * Baseline method provided by organizers.

All methods were evaluated on the 40 test cases using the TRE (Target Registration Error) and TRE30 (the TRE computed on the 30% of landmarks with the largest initial errors) metrics, based on manually annotated landmarks ($\sim$25 per case). An inter-rater (three raters) error, computed on a random subset of test cases, had a mean TRE of 2.00 pixels, serving as a rough upper bound for achievable performance. The results (cf. 5) show that all evaluated methods achieved substantial improvements over the unaligned baseline (mean TRE: 69.6). The top-ranked method, VROC, achieved the lowest mean TRE (2.64) and TRE30 (2.86), followed closely by next-gen-nn (TRE: 2.93) and greedy (TRE: 3.03), which obtained nearly identical overall ranks. IWM and globalign completed the test phase with mean TREs of 3.10 and 3.21, respectively. Despite differences in methodological paradigms, the final TRE scores among the top submissions fall within a narrow range, suggesting that both learning-based and classical pipelines can achieve similar levels of accuracy when appropriately tuned. However, robustness remains a challenge: failure cases—defined as those with mean TRE greater than 15 pixels—occurred across all methods, with counts of 3, 1, 4, 4, and 7 for VROC, next-gen-nn, greedy, IWM, and globalign, respectively.

The COMULISglobe SHG/BF task attracted fewer submissions compared to other Learn2Reg tasks, highlighting a potential engagement gap between the classical medical image registration and the biomedical (particularly histology) imaging communities. Notably, the majority of participants came from traditional medical image registration backgrounds, with minimal contribution from the histology and broader biomedical imaging communities, including active COMULISglobe members. Bridging these communities and fostering greater collaboration should be a priority for future editions of the challenge to better address the unique requirements of multimodal microscopy registration.

This division was also reflected in an initially introduced second COMULISglobe task—3D-CLEM registration—based on data from (Krentzel et al., 2023) which was ultimately dropped due to its high complexity. It involved automatic registration of multimodal 3D microscopy data, specifically aligning high-resolution electron microscopy (FIB-SEM and SBF-SEM) with super-resolution fluorescence light microscopy. The task benefited from domain knowledge and presented significant challenges related to large volumes and sparse correspondences, making it less accessible to the broader registration community. We assume it would have been of great interest to microscopy specialists, a group that was underrepresented in the challenge.

While most methods effectively solved the predominantly rigid alignment components of the task, opportunities remain for improving deformable registration performance.

deformation fields. Fine-tuning is performed using a similar design approach to ConvexAdam (Siebert et al., 2024), a recent automatic registration framework.

The **IWM** method combines classical feature matching with instance-specific deformable registration, forming a hybrid pipeline (Wodzinski and Müller, 2024; Wodzinski et al., 2024). The process begins with modality normalization to minimize appearance differences between BF and SHG images. Sparse keypoint correspondences are then extracted using the SuperPoint detector (DeTone et al., 2018) and matched via the SuperGlue (Sarlin et al., 2020) algorithm, which are both considered high-capacity learning-based matching models. To compensate for their lack of rotation and scale invariance, the method performs exhaustive matching across a range of initial rotations and scales, selecting the transformation with the highest number of valid matches. This provides a robust affine initialization. The final stage involves dense, multi-resolution deformable registration via instance optimization rather than deep learning. This choice was motivated by the limited availability of training data and the lack of pre-trained foundation models for microscopy registration, favoring a flexible, optimization-based approach with strong generalization properties.

Robustness continues to be a key challenge, as evidenced by approximately 10% failure cases across all methods. Addressing these limitations will be critical for the practical application of multimodal microscopy registration in biomedical research and clinical workflows.

4.1.4 NLST

The NLST task presented a highly realistic and challenging benchmark for thoracic CT registration, incorporating both lung cancer screening and radiotherapy cases. The new and extended evaluation from 2023 benefited from a substantially enhanced dataset, with over 100 landmarks per case annotated independently by two medical experts, offering a rigorous basis for assessment. The top-performing method, **LapIRNv2** (Mok and Chung, 2021), achieved the best overall rank (0.800), closely followed by the organizer-provided baseline (**corrfield** (Heinrich et al., 2015)) (rank: 0.670) and **Birmingham** (Jia et al., 2022, 2023) (rank: 0.553). These methods demonstrated low target registration errors (TRE: 1.425–1.63 mm) and excellent robustness on difficult cases (TRE30: 1.923–2.200 mm). While deep learning approaches dominated the top ranks, classical keypoint-based strategies remained competitive. Notably, the lightweight learning-based MONAI baseline[13] also performed well (rank: 0.452), highlighting the potential of efficient architectures. While teams such as UKE and UCLA reported higher TREs, all submissions significantly outperformed the unregistered baseline (initial TRE: 10.40 mm). Importantly, lung nodule TRE, a clinically critical metric, was included in the final ranking and aligned well with landmark-based scores (here the Top-4 had roughly 50% lower errors than MONAI and UCLA with $TRE_{nodules}$: 1.135–1.434 mm), supporting its continued use in longitudinal studies. All leading methods executed within clinically acceptable runtimes of several seconds, and the task's realistic anatomical variability underlined the importance of robustness and generalizability in clinical deployment.

#	Team	TRE	SDLogJacDet	Rank
1	LapIRNv2	1.441	0.042	0.800
2	corrfield	1.638	0.038	0.670
3	Birmingham	1.425	0.045	0.553
4	UKE	2.435	0.062	0.489
5	MONAI	2.042	0.039	0.452
6	UCLA	2.934	0.071	0.352
	initial	10.40		

Table 1: Summary of NLST 2023 Results. TRE is the mean target registration error on anatomical landmarks in millimeters.

---

13. `https://github.com/Project-MONAI/tutorials/blob/main/3d_registration/learn2reg_nlst_paired_lung_ct.ipynb`

4.2 Learn2Reg 2024

The 2024 edition of the Learn2Reg challenge demonstrated active engagement from a wide range of research teams across all three new tasks. Notably, two teams—VROC and next-gen-nn—submitted algorithms consistently across all three challenge tracks, demonstrating the versatility of their methods and enabling valuable cross-task insights. Both teams shared second place on the ReMIND2Reg task. In the LUMIR challenge, next-gen-nn also secured second place, while VROC did not place among the top ten. In the CO-MULISglobe SHG/BF challenge, VROC achieved first place, followed by next-gen-nn in second place. These placements are based solely on participant submissions and do not include organizer-provided baselines. While no official overall ranking was announced, the consistent high performance of next-gen-nn across all tasks suggests strong generalizability of their approach. The two teams differed notably in their methodological strategies. VROC employed a consistent, predominantly optimization-based pipeline across all tasks. Their approach combined task-specific preprocessing (e.g., Gaussian smoothing, modality mapping) with affine or rigid registration using modality-adapted similarity metrics such as NCC, NGF, or MSE. For deformable alignment, they incorporated Demons-like forces and relied on extensive case-wise hyperparameter tuning, guided by a surrogate loss based on NCC. In contrast, next-gen-nn adopted a more adaptive strategy depending on the task: for ReMIND2Reg, they chose a fully optimization-based approach, while for COMULISglobe SHG/BF, they applied optimization-based fine-tuning on top of deep learning results.

Several general findings emerged from the 2024 edition of the Learn2Reg challenge. While participation was generally strong and engagement high, most teams focused their final test submissions on the LUMIR task, a brain MRI registration challenge. This mirrors trends from previous editions, where tasks using the OASIS data also attracted the majority of participants. The relatively lower participation in more complex tasks like ReMIND2Reg, characterized by multi-modality and varying fields of view, suggests these remain challenging and potentially less accessible. In contrast, the LUMIR task yielded consistently good results across nearly all submissions, indicating it may be more motivating or tractable for participants. The COMULISglobe task—focusing on registration of histology images—attracted fewer teams. This may be due to Learn2Reg's traditional emphasis on classical medical image registration, making biomedical image registration a more niche and specialized subfield within the challenge.

The 2024 challenge also highlighted the absence of a standardized, robust registration framework analogous to nnUNet in segmentation. ConvexAdam (Siebert et al., 2024) represents a step toward this goal and was adopted by

next-gen-nn in several tasks, highlighting promising research avenues focused on adaptable and reusable registration frameworks.

Baseline methods, particularly classical optimization-based tools, continue to perform strongly in some tasks. For instance, NiftyReg (Modat et al., 2014) achieved the best results on ReMIND2Reg, while Greedy (Yushkevich, 2020) shared second place with next-gen-nn on the CO-MULISglobe SHG/BF task. The LUMIR challenge was an exception, where classical baselines such as ANTs SyN were outperformed by deep learning-based methods. Notably, LUMIR had the largest dataset of all three tasks, with over 4,000 scans, which likely highlights the connection between dataset size and the success of deep learning approaches. Overall, while deep learning approaches demonstrate strong performance and continue to advance the field, the results suggest a degree of stagnation in optimization-based image registration and deep learning methods suited for smaller datasets, emphasizing the need for novel approaches that can generalize robustly across diverse clinical scenarios.

5. Conclusion and Outlook

The Learn2Reg 2024 challenge provided valuable insights into the current state of medical image registration through its three distinct tasks: ReMIND2Reg, LUMIR, and CO-MULISglobe SHG/BF. These tasks addressed a range of clinically relevant scenarios, including intraoperative ultrasound-to-MRI alignment in brain tumor surgery, large-scale unsupervised brain MRI registration, and multimodal histopathology image analysis. The outcomes highlighted significant progress in various methodologies, from advanced deep learning approaches to refined classical optimization techniques.

A key observation from Learn2Reg 2024 was the influence of dataset characteristics on method performance. For instance, the LUMIR task, benefiting from an extensive dataset of over 4,000 brain MRI scans, generally favored deep learning methods. Conversely, in tasks with more constrained data or unique challenges, such as the multi-modal and varying field-of-view aspects of ReMIND2Reg, or the specialized nature of COMULISglobe in histology, classical optimization-based methods like NiftyReg and Greedy demonstrated strong performance and even achieved leading results in some instances. This variability underscores the ongoing need for flexible registration techniques that can perform effectively across different data scales and modality combinations.

Despite the notable advancements, several limitations and areas for further development were identified. The field still lacks a widely applicable and robust registration framework that can adapt across diverse clinical scenarios and imaging modalities without extensive task-specific adjustments. While certain frameworks show promise in adaptability, consistency in performance remains a challenge. Additionally, the differing levels of participation across tasks, particularly the lower engagement in more specialized challenges like COMULISglobe, suggests a need for better integration and outreach to broader biomedical imaging communities. Building stronger connections between traditional medical image registration and other fields, such as histology, is crucial for addressing a wider spectrum of clinical needs. The occurrence of failure cases, even among top-performing methods, also points to the persistent need for enhanced robustness in real-world applications, aiming for algorithms that are more resilient to anomalies and noise.

Moving forward, Learn2Reg plans to build on these findings. Future editions will aim to broaden the scope of biomedical imaging tasks, actively seeking to involve new communities and incorporate novel modalities to address a wider range of registration problems. A strong focus will be placed on developing more generalizable and robust registration frameworks that require less manual tuning and can adapt more seamlessly to new data distributions and imaging conditions. We will also continue to refine our challenge design to encourage broader participation and foster interdisciplinary collaboration, ensuring Learn2Reg remains a valuable platform for advancing medical image registration research and its practical applications.

(LLMs) to support the writing and editing of this manuscript. All content was reviewed and finalised by the authors.

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

Conflicts of Interest

We declare we do not have any conflicts of interest.

Data availability

All data of this challenge paper, except for the hidden test data, are publicly available through the Learn2Reg 2024 challenge website at `https://learn2reg.grand-challenge.org/learn2reg-2024/`.

References

Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. The medical segmentation decathlon. *Nature communications*, 13(1):4128, 2022.

Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.

Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019.

Jiří Borovec, Jan Kybic, Ignacio Arganda-Carreras, Dmitry V Sorokin, Gloria Bueno, Alexander V Khvostikov, Spyridon Bakas, Eric I-Chao Chang, Stefan Heldmann, Kimmo Kartasalo, et al. ANHIR: Automatic non-rigid histological image registration challenge. *IEEE Transactions on Medical Imaging*, 39(10):3042–3052, 2020.

Junyu Chen, Eric C Frey, Yufan He, William P Segars, Ye Li, and Yong Du. TransMorph: Transformer for unsupervised medical image registration. *Medical Image Analysis*, 82: 102615, 2022.

Junyu Chen, Shuwen Wei, Joel Honkamaa, Pekka Marttinen, Hang Zhang, Min Liu, Yichao Zhou, Zuopeng Tan, Zhuoyuan Wang, Yi Wang, Hongchao Zhou, Shunbo Hu, Yi Zhang, Qian Tao, Lukas Förner, Thomas Wendler, Bailiang Jian, Benedikt Wiestler, Tim Hable, Jin Kim, Dan Ruan, Frederic Madesta, Thilo Sentker, Wiebke Heyer, Lianrui Zuo, Yuwei Dai, Jing Wu, Jerry L. Prince, Harrison Bai, Yong Du, Yihao Liu, Alessa Hering, Reuben Dorent, Lasse Hansen, Mattias P. Heinrich, and Aaron Carass. Beyond the LUMIR challenge: The pathway to foundational registration models, 2025. URL `https://arxiv.org/abs/2505.24160`.

Xiang Chen, Yan Xia, Nishant Ravikumar, and Alejandro F Frangi. CAR-Net: unsupervised Co-attention guided registration network for joint registration and structure learning. *arXiv preprint arXiv:2106.06637*, 2021.

Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperPoint: Self-Supervised Interest Point Detection and Description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. URL `https://arxiv.org/abs/1712.07629`.

David Drobny, Tom Vercauteren, Sébastien Ourselin, and Marc Modat. Registration of MRI and iUS data to compensate brain shift using a symmetric block-matching based approach. In *International Workshop on Point-of-Care Ultrasound*, pages 172–178. Springer, 2018.

Benoit Dufumier et al. OpenBHB: A large-scale multisite brain MRI dataset for age prediction and debiasing. *NeuroImage*, 263:119637, 2022.

Jan Ehrhardt, Alexander Schmidt-Richberg, René Werner, and Heinz Handels. Variational registration: A flexible open-source ITK toolbox for nonrigid image registration. In *Bildverarbeitung für die Medizin 2015: Algorithmen-Systeme-Anwendungen. Proceedings des Workshops vom 15. bis 17. März 2015 in Lübeck*, pages 209–214. Springer, 2015.

K Eliceiri, B Li, and A Keikhosravi. Multimodal Biomedical Dataset for Evaluating Registration Methods (Full-Size TMA Cores). Zenodo, 2021. `https://doi.org/10.5281/zenodo.4550300`.

Hastings Greer, Lin Tian, Francois-Xavier Vialard, Roland Kwitt, Sylvain Bouix, Raul San Jose Estepar, Richard Rushmore, and Marc Niethammer. Inverse consistency by construction for multistep deep registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 688–698. Springer, 2023.

Mattias P Heinrich, Mark Jenkinson, Manav Bhushan, Tahreema Matin, Fergus V Gleeson, Michael Brady, and Julia A Schnabel. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis*, 16(7):1423–1435, 2012.

Mattias P Heinrich, Heinz Handels, and Ivor JA Simpson. Estimating large lung motion in copd patients by symmetric regularised correspondence fields. In *International conference on medical image computing and computer-assisted intervention*, pages 338–345. Springer, 2015.

Alessa Hering, Lasse Hansen, Tony CW Mok, Albert CS Chung, Hanna Siebert, Stephanie Häger, Annkristin Lange, Sven Kuckertz, Stefan Heldmann, Wei Shao, et al. Learn2Reg: Comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Transactions on Medical Imaging*, 42(3):697–712, 2022.

Wiebke Heyer, Yannic Elser, Lennart Berkel, Xinrui Song, Xuanang Xu, Pingkun Yan, Xi Jia, Zi Li, Tony CW Mok, BoWen LI, et al. OncoReg: Medical Image Registration for Oncological Challenges. *arXiv preprint arXiv:2503.23179*, 2025.

Malte Hoffmann, Benjamin Billot, Douglas N Greve, Juan Eugenio Iglesias, Bruce Fischl, and Adrian V Dalca. SynthMorph: learning contrast-invariant registration without acquired images. *IEEE Transactions on Medical Imaging*, 41(3):543–558, 2021.

Joel Honkamaa and Pekka Marttinen. Sitreg: Multi-resolution architecture for symmetric, inverse consistent, and topology preserving image registration. *Machine Learning for Biomedical Imaging*, 2:2148–2194, 2024. ISSN 2766-905X. . URL https://melba-journal.org/2024:026.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022.

Yuankai Huo, Zhoubing Xu, Yunxi Xiong, Katherine Aboud, Prasanna Parvathaneni, Shunxing Bao, Camilo Bermudez, Susan M Resnick, Laurie E Cutting, and Bennett A Landman. 3D whole brain segmentation using spatially localized atlas network tiles. *NeuroImage*, 194:105–119, 2019.

Dongseong Hwang. FAdam: Adam is a natural gradient optimizer using diagonal empirical Fisher information. *arXiv preprint arXiv:2405.12807*, 2024.

Xi Jia, Joseph Bartlett, Tianyang Zhang, Wenqi Lu, Zhaowen Qiu, and Jinming Duan. U-net vs transformer: Is u-net outdated in medical image registration? In Chunfeng Lian, Xiaohuan Cao, Islem Rekik, Xuanang Xu, and Zhiming Cui, editors, *Machine Learning in Medical Imaging*, pages 151–160, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-21014-3.

Xi Jia, Alexander Thorley, Alberto Gomez, Wenqi Lu, Dipak Kotecha, and Jinming Duan. Fourier-net+: Leveraging band-limited representation for efficient 3d medical image registration. *arXiv preprint arXiv:2307.02997*, 2023.

Bailiang Jian, Jiazhen Pan, Morteza Ghahremani, Daniel Rueckert, Christian Wachinger, and Benedikt Wiestler. Mamba? catch the hype or rethink what really helps for image registration. In *International Workshop on Biomedical Image Registration*, pages 86–97. Springer, 2024.

P Juvekar, R Dorent, et al. The Brain Resection Multimodal Imaging Database (ReMIND). *Scientific Data*, 2023. https://doi.org/10.1101/2023.09.14.23295596.

A Keikhosravi, B Li, Y Liu, and KW Eliceiri. Intensity-based registration of bright-field and second-harmonic generation images of histopathology tissue sections. *Biomedical Optics Express*, 11(1):160–173, 2020.

Daniel Krentzel, Matouš Elphick, Marie-Charlotte Domart, Christopher J Peddie, Romain F Laine, Cameron Shand, Ricardo Henriques, Lucy M Collinson, and Martin L Jones. Clem-reg: An automated point cloud based registration algorithm for correlative light and volume electron microscopy. *BioRxiv*, pages 2023–05, 2023.

Learn2Reg Organisers. Learn2reg tutorial 2019 at miccai. https://learn2reg.github.io, 2019. Accessed: 2025-05-30.

Yihao Liu, Junyu Chen, Lianrui Zuo, Aaron Carass, and Jerry L Prince. Vector field attention for deformable image registration. *Journal of Medical Imaging*, 11(6):064001–064001, 2024a.

Yihao Liu et al. On finite difference Jacobian computation in deformable image registration. *International Journal of Computer Vision*, 2024b.

Daniel S Marcus et al. 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.

Kasper Marstal, F Berendsen, Niels Dekker, Marius Staring, and Stefan Klein. The continuous registration challenge:

Evaluation-as-a-service for medical image registration algorithms. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1399–1402. IEEE, 2019.

Marc Modat, David M Cash, Pankaj Daga, Gavin P Winston, John S Duncan, and Sébastien Ourselin. Global image registration using a symmetric block-matching approach. *Journal of medical imaging*, 1(2):024003–024003, 2014.

Tony CW Mok and Albert CS Chung. Conditional deformable image registration with convolutional neural network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 35–45. Springer, 2021.

Keelin Murphy, Bram Van Ginneken, Joseph M Reinhardt, Sven Kabus, Kai Ding, Xiang Deng, Kunlin Cao, Kaifang Du, Gary E Christensen, Vincent Garcia, et al. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE Transactions on Medical Imaging*, 30 (11):1901–1920, 2011.

Johan Öfverstedt, Joakim Lindblad, and Nataša Sladoje. Fast computation of mutual information in the frequency domain with applications to global multimodal image alignment. *Pattern Recognition Letters*, 159:196–203, 2022.

Nicolas Pielawski, Elisabeth Wetzer, Johan Öfverstedt, Jiahao Lu, Carolina Wählby, Joakim Lindblad, and Nataša Sladoje. CoMIR: Contrastive multimodal image representation for registration. *Advances in neural information processing systems*, 33:18433–18444, 2020.

Guilherme Potje, Felipe Cadar, André Araujo, Renato Martins, and Erickson R Nascimento. XFeat: Accelerated features for lightweight image matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2682–2691, 2024.

Pavitra Sokke Rudraiah, Rafael Camacho, Julia Fernandez-Rodriguez, Dror Fixler, Jan Grimm, Florian Gruber, Matúš Kalaš, Christopher Kremslehner, Claudia Kuntner, Daniela Kuzdas-Wood, Joakim Lindblad, Julia G. Mannheim, Martina Marchetti-Deschmann, Perrine Paul-Gilloteaux, Paula Sampaio, Peter Sandbichler, Anna Sartori-Rupp, Nataša Sladoje, Paul Verkade, Andreas Walter, and Samuele Zoratto. Correlated multimodal imaging in life sciences: lessons learnt. *Frontiers in Biomaterials Science*, 3:1338115, 2024.

Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning Feature Matching with Graph Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4938–4947, 2020. URL `https://arxiv.org/abs/1911.11763`.

Hanna Siebert, Christoph Großbröhmer, Lasse Hansen, and Mattias P Heinrich. ConvexAdam: Self-Configuring Dual-Optimisation-Based 3D Multitask Medical Image Registration. *IEEE Transactions on Medical Imaging*, 2024.

Alaa Taha et al. Magnetic resonance imaging datasets with anatomical fiducials for quality control and registration. *Scientific Data*, 10(1):449, 2023.

Zuopeng Tan, Lihe Zhang, Yanan Lv, Yili Ma, and Huchuan Lu. GroupMorph: Medical image registration via grouping network with contextual fusion. *IEEE Transactions on Medical Imaging*, 2024.

National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5):395–409, 2011.

Jean-Philippe Thirion. Image matching as a diffusion process: An analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

Lin Tian, Hastings Greer, Roland Kwitt, François-Xavier Vialard, Raúl San José Estépar, Sylvain Bouix, Richard Rushmore, and Marc Niethammer. *unigradICON*: A foundation model for medical image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 749–760. Springer, 2024.

Andreas Walter, Perrine Paul-Gilloteaux, Birgit Plochberger, Ludek Sefc, Paul Verkade, Julia G. Mannheim, Paul Slezak, Angelika Unterhuber, Martina Marchetti-Deschmann, Manfred Ogris, Katja Bühler, Dror Fixler, Stefan H. Geyer, Wolfgang J. Weninger, Martin Glösmann, Stephan Handschuh, and Thomas Wanek. Correlated multimodal imaging in life sciences: expanding the biomedical horizon. *Frontiers in Physics*, 8:47, 2020.

Andreas Walter, Gerard J Kleywegt, and Paul Verkade. Correlative multimodal imaging: Building a community. In *Methods in cell biology*, volume 162, pages 417–430. Elsevier, 2021.

Haiqiao Wang, Zhuoyuan Wang, Dong Ni, and Yi Wang. ModeTv2: GPU-accelerated motion decomposition transformer for pairwise optimization in medical image registration. *arXiv preprint arXiv:2403.16526*, 2024.

Philippe Weitz, Masi Valkonen, Leslie Solorzano, Circe Carr, Kimmo Kartasalo, Constance Boissin, Sonja Koivukoski,

Aino Kuusela, Dusan Rasic, Yanbo Feng, et al. The ACROBAT 2022 challenge: Automatic registration of breast cancer tissue. *Medical Image Analysis*, 97:103257, 2024.

René Werner, Alexander Schmidt-Richberg, Heinz Handels, and Jan Ehrhardt. Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: A comparison and evaluation study. *Physics in Medicine & Biology*, 59(15):4247, 2014.

Marek Wodzinski and Henning Müller. Automatic Registration of SHG and H&E Images with Feature-Based Initial Alignment and Intensity-Based Instance Optimization: Contribution to the COMULIS Challenge. In *International Workshop on Biomedical Image Registration*, pages 346–357. Springer, 2024.

Marek Wodzinski, Niccolo Marini, Manfredo Atzori, and Henning Müller. DeeperHistReg: Robust whole slide images registration framework. *arXiv preprint arXiv:2404.14434*, 2024.

Yiming Xiao, Hassan Rivaz, Matthieu Chabanas, Maryse Fortin, Ines Machado, Yangming Ou, Mattias P Heinrich, Julia A Schnabel, Xia Zhong, Andreas Maier, et al. Evaluation of MRI to ultrasound registration methods for brain shift correction: the CuRIOUS2018 challenge. *IEEE Transactions on Medical Imaging*, 39(3):777–786, 2019.

Paul A. Yushkevich. Greedy: A Fast Diffeomorphic Image Registration Algorithm. `https://github.com/pyushkevich/greedy`, 2020. Accessed: 2025-06-04.

Hang Zhang, Xiang Chen, Renjiu Hu, Dongdong Liu, Gaolei Li, and Rongguang Wang. MemWarp: Discontinuity-Preserving Cardiac Registration with Memorized Anatomical Filters. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 671–681. Springer, 2024a.

Yi Zhang, Yidong Zhao, and Qian Tao. Improving Instance Optimization in Deformable Image Registration with Gradient Projection. *arXiv preprint arXiv:2410.15767*, 2024b.