Needles & Haystacks: Dataset and Benchmark for Domain-Agnostic Image-Based Rigid Slice-to-Volume Registration

Anton Frolov Florian Kleiner Christiane Rößler Volker Rodehorst
Bauhaus-Universität Weimar
{firstname}.{lastname}@uni-weimar.de

Abstract

*We address domain-agnostic slice-to-volume (S2V) registration, the alignment of 2D sliced/tomographic images into 3D volumes without prior knowledge of structure, shape, or orientation. While S2V registration is well-studied in medical imaging, which often relies on auxiliary information (*e.g. *landmarks, segmentation masks, pre-defined orientations, canonical/atlas volumes), applications such as micro-structure characterization in materials science lack such domain-specific aids. This leaves the task inherently ill-posed due to noise, unstructured regions, repetitive patterns, rotational and translational symmetries. To address this challenge, we present "Needles & Haystacks,"[1] a novel multi-domain algorithm development dataset with* $158,436$ *unique registration problems and ground-truth solutions, based on diverse and openly licensed real-world volumetric data. Additionally, we provide an online platform with* $8,461$ *test problems for reproducible evaluation of competing methods. We also propose strong baseline solutions with public implementations and highlight opportunities for further algorithmic advancements.*

1. Introduction

The problem of S2V registration is best known from applications in medical imaging [7, 13]. These applications include correlative multi-instrument analysis (*e.g.*, cross-referencing 3D computer tomographic (CT) scans with 2D tissue classifications from histology), localization of 2D imaging instruments in image-guided surgeries, and volumetric reconstruction from 2D imaging modalities. Medical S2V deals with well-studied anatomical subjects, making it feasible to directly harness the existing domain knowledge during registration. This knowledge can be expressed in the form of landmarks and segmentation masks, which inform robust initializations. Additionally, favorable indirect conditions, such as prevalent standard orientations and limited variability in subject appearance, allow for robust and accurate modeling. At the same time, practical realities of medical S2V often require accommodating inescapable non-rigid deformations and temporal changes between the registered modalities. This constitutes a significant challenge and is a primary focus of state-of-the-art medical S2V research.

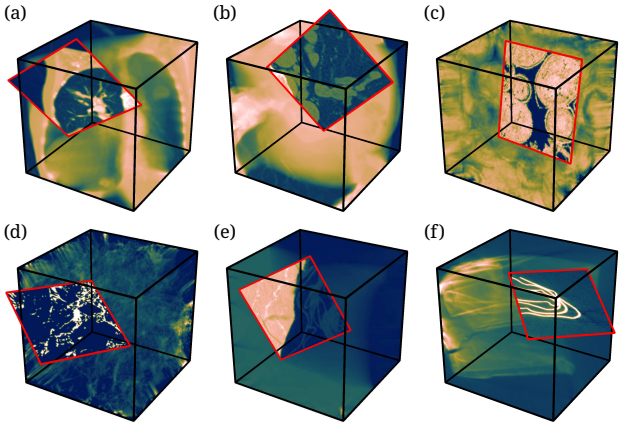

Figure 1. Examples of S2V registration tasks in "Needles & Haystacks". The dataset encompasses numerous domains: (a,b) de-contextualized anonymous medical data [21, 58], (c) rocks [1], (d) soils [59], (e) fossils [41], (f) archeological artifacts [17], *etc*. Ground-truth slice position highlighted in red. Volumes visualized by average projection with batlowW color map [8].

However, applications of S2V registration extend beyond medical imaging, offering insights across disciplines such as materials science, geology, biology, studies of cultural artifacts, industrial testing. Similar to medical S2V, spatial allocation of 2D images in the space of 3D volumes allows strengthening the analysis through bi-directional cross-referencing of often complementary optical, spectroscopic, ultrasound, radiographic measurements, *etc*. Unlike in medical imaging, applications of this kind often deal with subjects of unpredictable shape, structure and orientation, which poses a serious challenge. In practice, this often

[1] Project page: https://xaf-cv.github.io/nh-rs2v/

necessitates manual initialization by human experts, and, whenever technically feasible, the labor-intensive introduction of physical fiducials. Conversely to medical imaging, modeling of non-rigid deformations is less relevant, as firm materials, *e.g*. sandstone, wood, are generally less prone to deformation, while soft materials are usually sufficiently uncomplicated to immobilize. It is therefore clear that such S2V applications are distinct from the ones typically encountered in medical imaging. We describe this class of S2V as *domain-agnostic* image-based S2V registration. An accurate and robust fully automatic solution to domain-agnostic S2V is in high demand and offers significant potential for high-impact research.

Domain-agnostic S2V is an ill-posed and extremely hard problem, as individual 2D slices contain significantly less information than entire 3D volumes. This initial imbalance conditions the problem to have multiple plausible solutions, which is further exacerbated by ambiguities from rotational and translational symmetries, measurement noise and poorly-structured regions. This is similar to challenging ambiguities in wide-baseline image matching tasks [22]. In recent years, wide-baseline image matching has been revolutionized through introduction of novel data-driven methodologies. The development has been made possible to a significant extent by public availability of datasets, evaluations and challenges. To the best of our knowledge, no commonly recognized datasets or benchmarks have been proposed in the field of domain-agnostic S2V. This is likely due to the practical challenges associated with curation of such datasets. Collecting large and diverse datasets with accurate ground truth for real-world S2V problems is technically challenging and costly, and often involves legal (privacy, data sharing policies) and commercial (business interests) obstacles. We aim to address this deficiency by contributing a novel large-scale, open source, openly licensed dataset. Instead of physically sourcing real-world S2V problems, we opt to synthetically sample a diverse multitude of registration tasks from abundantly available, openly licensed 3D volumes previously published in the relevant scientific fields. This approach ensures that the dataset is openly licensed, promoting transparency, reproducibility, and reusability [60]. Our approach to dataset creation is conceptually similar to that of major datasets in image matching [9, 28], and can be repurposed in the future to produce even larger or more specialized datasets. With our principled synthetic sampling strategy, we acquire highly challenging yet feasible registration tasks with error-free ground-truth. Having accurate ground-truth allows us to define clear and informative metrics for evaluating registration solutions, favoring accurate and robust methodologies. The scale of our dataset indirectly favors sufficiently fast algorithms, although speed is not our primary performance factor. To enable clear ranking of S2V methodologies, we provide an online evaluation platform containing $8,461$ test problems with non-public ground-truth.

Finally, to gauge the intrinsic challenge of our dataset, we implement and deploy massive randomized solution search with classic optimization-based techniques. Additionally, we adapt, train and evaluate a modern high-performing deep-learning architecture for local image feature matching to the task of domain-agnostic S2V. These complementary techniques provide strong baseline performance but also clearly demonstrate the significant potential for improvement. Our implementations are open source and can serve as a basis for future development of novel S2V methodologies.

In summary, this paper contributes:

- a novel open source and openly licensed dataset for domain-agnostic image-based rigid S2V registration, which is well-positioned to simplify and to stimulate the development of novel methodologies,
- an online benchmark, that adequately represents the significant challenges of such registration tasks,
- baseline methodologies that demonstrate the feasibility of the registration and can serve as reference for future methodologies.

2. Related work

2.1. Domain-agnostic S2V in medical imaging

Over the last decades, the problem of S2V registration has received a lot of coverage in the context of medical image registration. While practical applications of medical S2V generally tend to have a different set of requirements and expectations, select methods in *rigid image-based (intrinsic)* registration, in the taxonomy of [33], are directly applicable in the context of domain-agnostic S2V. Often an initial approximation of a more complex registration task, this formulation allows *e.g*. to acquire a strong initial guess for subsequent non-rigid S2V registration. This class of registration problems has been traditionally approached as a dissimilarity minimization problem [52], the comprehensive survey on which is provided in [13].

DL-based methodologies have been proposed to overcome some of the fundamental limitations of optimization-based medical image registration [7]. In the context of medical S2V, some examples include applications of deep convolutional neural networks (CNNs) [19] to perform direct regression on the manifold of special Euclidean group in 3 dimensions $\mathbf{SE}(3)$, as well as self-attention-style layers [62] and iterative transformer models [61] for further accuracy improvement. Notably, models from [19, 61, 62] are conceptually applicable only to registration of individual slices in the space of a *single* canonical volume (atlas), which has

limited applicability to our formulation of domain-agnostic S2V, where arbitrary 2D slices are registered into arbitrary 3D volumes.

Conversely, a recent development [34] has proposed to induce point-wise correspondences between registered 2D slices and 3D volumes through the use of a correlation layer, such as in image matching [54], and was trained on 16 *separate* 3D MRI volumes with $3,957$ individual 2D ultrasound slices. Since methodology from [34] can be trivially extended to arbitrary volumetric data, it has direct relevance for domain-agnostic S2V. To evaluate our data contributions, we derive and train a complete S2V generalization of [54], and demonstrate the strength of the approach in the context of domain-agnostic S2V.

2.2. Image matching and domain-agnostic S2V

Applications of image matching (IM) [22], such as 3D reconstruction, camera re-localization and SLAM share a number of similar challenges with domain-agnostic S2V registration. Both methodologies attempt to recover a 6 DoF pose through spatial allocation of 2D data within a 3D context, even though 3D IM traditionally deals with sparse projective imaging, while S2V registration is performed within densely sampled sliced/tomographic 3D context. In both scenarios, the challenge is to disambiguate the effects of multiple confounding factors that lead to the possibility of multiple plausible solutions. This is often the case due to simultaneous effects of viewpoint and illumination changes in 3D IM, and due to the effects of local symmetries and 2D/3D information imbalance in S2V. Poorly structured, repetitive and noisy regions are a common challenge in both tasks.

In recent years, image matching has witnessed revolutionary developments based on end-to-end deep learning, most notably the introduction of highly repeatable and robust salient point detectors and descriptors [6, 10, 36, 44], novel context-aware descriptor matchers [29, 47], and even detection-free matchers [54, 65]. To a significant extent, recent advancements made with end-to-end deep learning have been made possible through the introduction of large-scale publicly available datasets and benchmarks, *e.g*. HPatches [3], ETH SfM [49], ScanNet [9], MegaDepth [28], Image Matching Benchmark/Challenge [22], *etc*. Remarkably, both [9] and [28] have applied a synthetic sampling approach to produce accurate ground-truth for IM tasks on the basis of real-world 3D reconstructions, *i.e*. registered RGB-D sequences in case of [9] and multi-view stereo reconstructions in case of [28]. Our work adopts a conceptually similar approach by sampling domain-agnostic S2V tasks with error-free ground-truth from existing 3D volumetric reconstructions.

3. Methodology

3.1. Assumptions of rigidity and monomodality

We assume rigid and monomodal registration, where 2D slices align with 3D volumes without deformation and within the same imaging modality. While real-world applications of domain-agnostic S2V registration are likely to be cross-modal, *e.g*. correlative multi-instrument microstructure characterization, our monomodality assumption is well-justified for several reasons. First, within-mode registration already poses an extremely hard and *unsolved* challenge due to the inherent 2D–3D information imbalance. Second, cross-modal registration often follows domain adaptation [56] steps (*e.g*., image-to-image translation [46]) that reduce the problem to within-mode registration in learned [39] or transformed spaces. Finally, the assumption allows us to sample massive amounts of error-free ground-truth without the need to collect physical measurements or to perform potentially imperfect simulations of specific imaging modalities.

3.2. Registration task formulation

Rigid S2V registration is defined [13] as the problem of finding the affine transformation A that aligns a 2D slice I with a 3D volume V. To perform this alignment, we allocate the initial pose of I to the $Z = 0$ plane in 3D space. Under rigid motion, A is parameterized by a rotation $R \in \mathbf{SO}(3)$ and a translation T. We expressed it as:

$$A(R,T) = TRST_c^{-1}, \tag{1}$$

where T_c is translation to the slice center, and S is a known lateral scaling of the slice. S is fixed to some ground-truth $\hat{S}$, which we vary across registration tasks to simulate variation in intersection scale. Assuming monomodality, the registration task thus requires recovering the correct solution $\left\{\hat{R},\hat{T}\right\}$, such that $I \approx I_{\hat{R},\hat{T}}$, where $I_{R,T}$ denotes $V\left(A^{-1}\left(R,T\right)\right)$ for the sake of simplicity, and is a slice sampled from V at a solution candidate $\{R,T\}$.

3.3. Dataset curation

The goal of our dataset is to enable development and robust evaluation of novel S2V methodologies in a domain-agnostic setting. To this end, we curate diverse volumetric data with balanced domain representation, and sample a large collection of feasible registration tasks in a principled way that maximizes internal diversity and challenge.

Volumetric data collection. Our approach leverages openly licensed real-world volumetric data from published academic research. Although 3D imaging is costly and resource-intensive, global efforts towards open scientific practices have resulted in online sharing of volumetric data on online platforms such as Zenodo [12], Digital Rocks

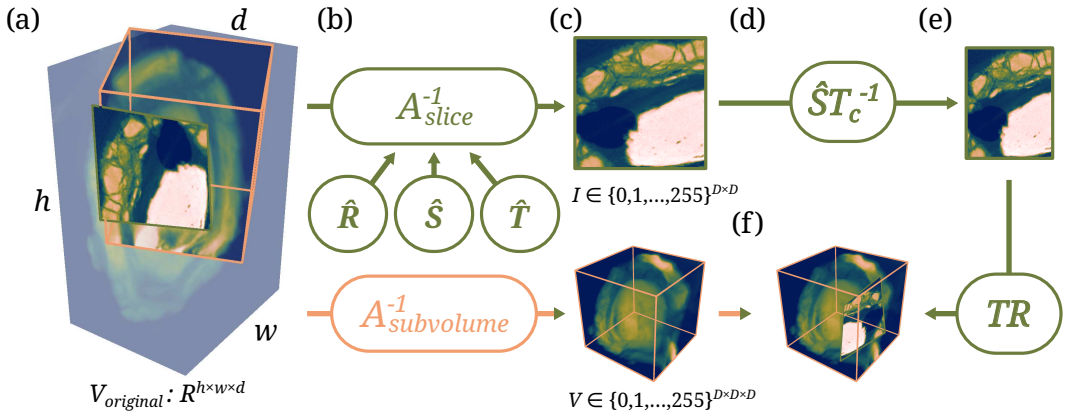

Figure 2. Registration task sampling scheme. From a volume $V_{original}$ (a) randomized matrices $A^{-1}_{subvolume}$ and A^{-1}_{slice} (b) sample a cubic volume V and a square slice I (c). A^{-1}_{slice} is parametrized with ground-truth non-uniform scaling $\hat{S}$, rotation $\hat{R}$ and translation $\hat{T}$ in relation to V. The registration task is to recover, given I, V and $\hat{S}$, the correct parameters $\hat{R}$ and $\hat{T}$, such that I, when pre-transformed (d) into the coordinate system of V (e), aligns into its true position within V (f). Illustrated with a μ-CT of a fossil [53].

Portal [42], datadryad.org and figshare.com. While only a fraction of all collections is licensed non-restrictively with permissive usage rights, we limit our search to openly licensed collections, such that the curated dataset can inherit the permissive licensing[2]. Additionally, our limiting factors include usage rights, volumetric resolution and quality, which lead us to reject some of the candidate collections. Based on these criteria, we identify a multitude of independent sources of volumetric data, each with distinct properties and characteristics. We observe that these allow for a high-level categorization into three branches of science, which we label as "Life Sciences" [21, 24, 26, 32, 50, 58], "Materials Characterization" [1, 2, 11, 16, 20, 25, 31, 35, 37, 38, 43, 48, 51, 59, 64] and "Paleo-, Archeo- & Anthropology" [15, 17, 32, 41, 53]. The significance of such categorization for us lies in the fact that sources which belong to the same category are considerably more alike, *e.g.* in terms of measurement subjects and instruments, between-item variability, structure and noise patterns. To ensure technical feasibility and efficiency of subsequent steps, we manually validate and homogenize the collected volumetric data in various ways, *e.g.* by the means of transforming, cropping to content and splitting large volumes into smaller chunks. All volumes are converted into 8-bit unsigned integer format after contrast adjustment. The initial imbalance in the number of available volumes $N_{available}$, both within and between categories, necessitates us to further sub-select the collection in order to mitigate these imbalances. We approach this by selecting individual volumes evenly and randomly between the sources and categories. In practice, we select an individual volume from a source with $N_{available}$ volumes with a probability proportional to $\frac{1}{N_{cat}} \frac{1}{N_{available}}$, where the number of all categories

[2]The dataset is released under CC BY 4.0

Table 1. Collected volumetric data categorization with counts of available $N_{available}$ and kept N_{kept} volumes, split counts.

Category	Sources	$N_{available}$	N_{kept}	Train / Val / Test
Life Sciences	6	1855	666	342 / 21 / 21
Materials Characterization	11	4200	666	591 / 34 / 34
Paleo-, Archeo- & Anthropology	5	166	166	146 / 10 / 10
Total	22	6221	1498	1079 / 65 / 65

$N_{cat} = 3$. Finally, kept volumes in our selection stem approx. to $44,46\,\%$ from life sciences, to $44,46\,\%$ from materials characterization and to $11,08\,\%$ from anthropology. For the purpose of registration task sampling, we split our selection into train, validation and test sets with the proportions of approx. $90\,\%$ / $5\,\%$ / $5\,\%$, such that all three sets contain strictly no overlap of volumetric data. This split is performed source-wise, such that all data sources are represented in all sets. A brief overview of data collection results is presented in Tab. 1, and the details in Supp. Mat., Sec. 1.1, ethical considerations for medical data in Sec. 1.2.

Synthetic sampling of registration tasks. The collected 3D volumes contain rich amount of structure, which can be used in a multitude of independent registration tasks. Such partial reuse of volumetric data requires a principled sampling strategy which can yield unique, meaningful and feasible registration tasks. We approach this by sampling pairs of subvolumes V and intersecting slices I with randomized matrices $A_{subvolume}$ and A_{slice} (see Fig. 2). $A_{subvolume}$ applies arbitrarily strong rotation, as well as some translation, cropping, and weak shearing, to sub-select a parallelepiped from the original arbitrarily-shaped volume. This sub-selected region is rescaled into a cubic volume $V \in \{0,1,...255\}^{D \times D \times D}$. We provide a formal definition for $A_{subvolume}$ in the Supp. Mat., Sec. 1.3. Note

that $A_{subvolume}$ is prone to sampling outside the boundaries of the original volume. In order to make the best use of the volumetric data, we do not prohibit out-of-bounds sampling entirely. Instead, we allow such regions to be filled via interpolation, which prevents an unfair advantage for the registration algorithm in the form of clearly-defined boundaries. A single intersecting slice I is sampled for each V by randomizing parameters $\hat{R}$, $\hat{T}$ and $\hat{S}$, which define some ground-truth $\hat{A}$ (see Eq. (1)). The first and second diagonal coefficients of $\hat{S}$ are chosen in the range $(0.5, 1.5)$. Then, $A_{slice} = A^{-1}_{subvolume}\hat{A}$ sub-selects a rectangle from the original volume, which is rescaled into a square $I \in \{0, 1, ...255\}^{D\times D}$. We randomize all non-rotational sampling parameters uniformly, while rotations are randomized for uniform cover on the $\mathbf{SO}(3)$. As a result, a registration task with ground-truth solution $\left\{\hat{R}, \hat{T}\right\}$ has been sampled, for which I, V, and $\hat{S}$ are the given inputs.

Validation and feasibility of registration tasks. Our randomized sampling approach generally produces unique and challenging registration tasks, yet some of the sampled tasks have low practical value for our purposes. This is the case when sampled slices and subvolumes contain too little or too ambigous an overlap of information, and cannot meaningfully contribute to training or evalution of registration algorithms. One example is when too large of a portion of the subvolume is sampled outside the boundaries of the original volume, which results in poor and uncharacteristic 3D structure. Another unreasonable outcome is when the intersection between the slice and the volume is too small, which imposes purely numerical limits on the precision of the registration. Additionally, some of the registration tasks are infeasible to solve when the slices are sampled from monotonous or featureless regions, or from regions overpowered by noise to the extent that the underlying structure is suppressed. To ensure practicality of our dataset, we implement a systematic validation process, the goal of which is to filter out such tasks. We only accept such registration tasks which simultaneously satisfy the requirements that: (1) out-of-bounds regions in V do not exceed $25\,\%$, (2) I intersects V at least to $75\,\%$ of its area, and (3) the ground-truth solution $\left\{\hat{R}, \hat{T}\right\}$ is a well-defined local minimum to an L2 dissimilarity minimization task $\underset{R,T}{\operatorname{argmin}}\, M(R, T)$,

$$M(R, T) = \sum_{(i,j)}^{\{1,...,D\}^2} \frac{|I(i, j) - I_{R,T}(i, j))|^2}{D^2}. \qquad (2)$$

All intensities are linearly scaled such that V is normalized by $3 \times \sigma$, where σ is the standard deviation of intensities in V. We evaluate Jacobian $\mathcal{J}(M)$ and Hessian $\mathcal{H}(M)$ at the ground-truth $\left\{\hat{R}, \hat{T}\right\}$, and accept it as a stable local minimum if the conditions hold:

$$\begin{cases} |M_{\hat{R},\hat{T}}| & < C_1, \\ \max(|\mathcal{J}(M_{\hat{R},\hat{T}})|) & < C_2, \\ \lambda_{min}(\mathcal{H}(M_{\hat{R},\hat{T}})) & > C_3, \end{cases} \qquad (3)$$

where $\lambda_{min}(\mathcal{H})$ is the smallest eigenvalue of the Hessian. Coefficients C_1, C_2 and C_3 are hand-chosen empirically. We start from 1.0 for C_1 and C_2, and 0 for C_3, and progressively strengthen the conditions, such that for all sources of volumetric data the clearly unreasonable tasks are systematically rejected, while a sufficient amount of reasonable tasks is accepted (see Fig. 3). Finally, we arrive at $C_1 = 0.1$, $C_2 = 0.5$ and $C_3 = 0.01$. Effectively, our approach guarantees that the accepted tasks can be solved at least by minimizing Eq. (2) when initialized sufficiently close to the ground-truth solution.

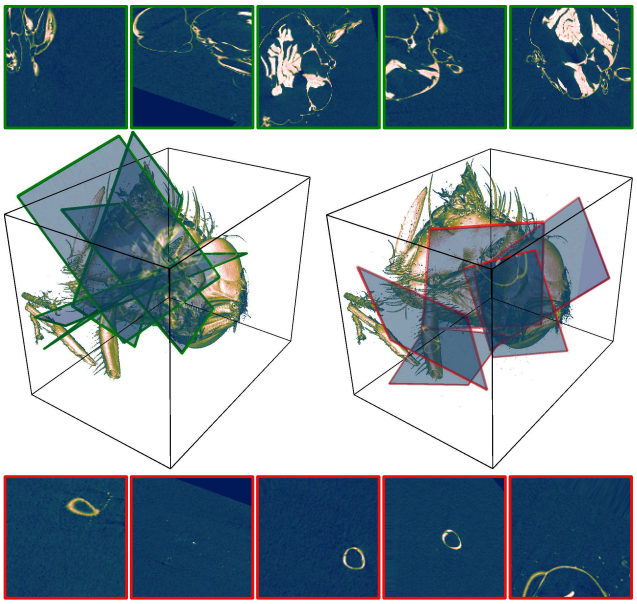

Figure 3. Registration task feasibility test outcomes. Sampled slices visualized in the coordinate space of the original volume. Sampled slices that satisfy the test (top/left, green) demonstrate clearly defined structure, while the rejected problems (bottom/right, red) are subject to rotational and translational ambiguities. μ-CT scan of a fly [32], isosurface rendering.

Dataset split. We apply our validation process after randomly sampling a large number of problems per volume. We then reject the infeasible, and keep at most 125 problems per volume. As a result, we produce a collection of $166,897$ unique S2V registration problems (see Fig. 1, more examples in Supp. Mat., Sec. 1.5). Finally, we end up with a train/validation/test split of $149,895$ / $8,541$ / $8,461$ registration problems. We also provide an additional, smaller train/validation split, in order to support fast cycles of validation, where just $10\,\%$ of problems are available per

Table 2. Optimization results (mAA). LO marks 1 pass of SLSQP for the best candidate. N_{rand} is the count of random initializations.

N_{rand}	Subplex [45] + LO			COBYLA [40] + LO			LBFGS [30]			SLSQP [27]		
	@ 5°	@ 10°	@ 20°	@ 5°	@ 10°	@ 20°	@ 5°	@ 10°	@ 20°	@ 5°	@ 10°	@ 20°
1	**0.53 %**	**0.57 %**	**0.70 %**	0.22 %	0.25 %	0.33 %	0.33 %	0.33 %	0.35 %	0.53 %	0.54 %	0.63 %
2	0.81 %	0.88 %	0.98 %	0.46 %	0.54 %	0.73 %	0.79 %	0.83 %	0.94 %	**0.99 %**	**1.04 %**	**1.23 %**
8	2.15 %	2.21 %	2.58 %	1.75 %	1.79 %	1.90 %	2.72 %	2.79 %	2.97 %	**3.51 %**	**3.56 %**	**3.80 %**
16	4.17 %	4.21 %	4.69 %	3.51 %	3.56 %	3.80 %	3.90 %	4.05 %	4.42 %	**5.83 %**	**5.96 %**	**6.33 %**
64	11.16 %	11.46 %	12.34 %	9.78 %	10.04 %	10.71 %	13.11 %	13.36 %	14.06 %	**16.51 %**	**16.85 %**	**17.64 %**
128	16.45 %	16.81 %	17.97 %	14.28 %	14.43 %	15.31 %	19.34 %	19.93 %	20.99 %	**23.86 %**	**24.41 %**	**25.81 %**
256	23.46 %	23.95 %	25.28 %	20.07 %	20.15 %	21.13 %	26.69 %	27.07 %	28.17 %	**31.97 %**	**32.73 %**	**34.11 %**
512	30.33 %	30.73 %	32.07 %	27.35 %	27.69 %	28.82 %	34.52 %	34.91 %	35.95 %	**40.18 %**	**40.91 %**	**42.13 %**
1024	39.34 %	39.93 %	41.40 %	34.14 %	34.59 %	35.94 %	41.07 %	41.55 %	42.66 %	**46.91 %**	**47.45 %**	**48.61 %**
2048	47.13 %	47.48 %	48.85 %	42.21 %	42.75 %	44.12 %	47.61 %	48.11 %	49.32 %	**53.05 %**	**53.60 %**	**54.62 %**

volume, which results in a split of $16,152 \;/\; 912$.

3.4. Evaluation metrics for registration accuracy

We evaluate solutions $\{R, T\}$ against the ground truth $\left\{\hat{R}, \hat{T}\right\}$ through *angular pose error*, similar to *e.g.* [22,63]. This approach allows for scale-independent comparison between poses that is independent from metric units of distance. We express translations in unit-norm vector form, while rotations are represented by single-cover unit-norm quaternions r and $\hat{r}$. Then, the total angular pose error is $e = \max\left(\left\langle t, \hat{t}\right\rangle, \langle r, \hat{r}\rangle\right)$, where $\langle\cdot,\cdot\rangle$ is the internal product. We accumulate angular errors for entire sets of solutions into the widely-used *mean Average Accuracy (mAA)* [63] metric, which is common in image matching evaluations [22]. mAA is calculated by integrating over the normalized histogram of errors with a bin resolution of $1°$ up to a threshold θ_{max}: $\text{mAA}(\theta_{\text{max}}) = \sum_{\theta=1}^{\theta_{\text{max}}} \frac{\text{A}(\theta)}{\theta_{\text{max}}}$, where $\text{A}(\theta)$ is the accuracy at error threshold θ. mAA reports on the cumulative error distribution, which has an advantage over the plain accuracy metric in that methodologies which produce the lowest possible errors are favored. The chosen θ_{max} defines the upper limit on the acceptable error of registration, *i.e.* all errors above the threshold are considered to be equally incorrect. We rank methodologies by reporting on the mAA at thresholds of $5°$, $10°$ and $20°$, which can inform applications with diverging accuracy requirements.

4. Evaluation

To assess the challenges posed by our curated dataset and benchmark, we evaluate the performance of two distinct methodologies for S2V registration. First, we examine classic optimization-based registration in a massive randomized search campaign. Rapid algorithm development cycles are enabled here by validating on a reduced validation set. Second, we explore modern deep learning-based approach inspired by recent advances in end-to-end image matching. To this end, we adapt and train LoFTR-S2V, a detector-free learned local feature matcher. We validate alternatives for robust registration based on the predicted correspondences, and apply local optimization (LO) steps to refine the prediction. For both methodologies, the best-performing variants are submitted to the test set benchmark as baseline references for future research. Our evaluation serves as an example of how both hand-crafted and deep-learned methodologies can be developed and evaluated using our dataset and benchmark. It also highlights the significant potential for improvement, as well as the key advantages of learning-based registration.

Table 3. Evaluation of dissimilarity criteria with random search (at $65,536$ initializations).

	@ 5°	@ 10°	@ 20°↑
ZNCC	5.13 %	5.44 %	6.34 %
MSE	5.26 %	5.49 %	6.51 %
MAE	**8.27 %**	**8.34 %**	**9.14 %**

4.1. Optimization-based registration

Classic optimization-based S2V registration is defined [13] as a minimization problem over some dissimilarity criterion M, which for our case takes the form of:

$$\underset{R,T}{\text{argmin}}\; M\left(I_{R,T}, I\right). \tag{4}$$

We represent R and T as a set of 7 parameters $P : \{r_1, r_2, r_3, r_4, t_1, t_2, t_3\}$, where R is expressed through a quaternion vector r and T through a translation vector t. These parameters are bounded by $r_i \in (-1, 1)$, $t_i \in (0, 1.)$. Our assumption of within-mode registration allows us to limit our choice of M to the most simple options, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), and Zero-mean Normalized Cross-Correlation (ZNCC). To choose the best dissimilarity criterion, we randomly sample a large number $(65,536)$ of initializations, evaluate the dissimilarity, and minimize Eq. (4)

with SLSQP [27] for the initialization with lowest dissimilarity. This allows to establish (Tab. 3) MAE as the best performing measure of dissimilarity. Next, we validate classic local solvers, both gradient-based LBFGS [30] and SLSQP [27], and derivative-free Subplex [45] and COBYLA [40] (Tab. 2). We solve Eq. (4) for N_{rand} randomized initializations in a CPU-parallel way, and reduce for the best solution. Our implementations are based on SciPy [57] and NLopt [23] solvers. By progressively increasing N_{rand} from 1 to $2,048$, we observe that the accuracy of final solutions increases with diminishing returns in relation to N_{rand}, while the solving time depends linearly on N_{rand}. We stop at $N_{rand} = 2,048$, since at this level the solving time reaches impracticable levels, *i.e.* per registration task it takes $36s$ for Subplex, $35s$ for COBYLA, $116s$ for LBFGS and $60s$ for SLSQP on an average consumer workstation[3]. Correspondingly, the total solving time for all problems in the test set for the best performing method (SLSQP, $N_{rand} = 2,048$) reaches approx. 141 hours, while further significant improvements in accuracy demand further exponential increase in solving time. Despite these high computational costs, SLSQP with $N_{rand} = 2,048$ offers just $54.62\,\%$ mAA @ $20°$, which stands to show that a significant portion of registration tasks is too challenging when considering only the local dissimilarity energy landscape. Another important observation here is that the same method achieves $53.05\,\%$ mAA @ $5°$, which is an indication that nearly all problems within $20°$ angular error can be successfully refined with local optimization.

4.2. Registration with learned local features

To the best of our knowledge, there exist no public implementations for end-to-end learned domain-agnostic S2V registration. Note that implementations from medical S2V [19, 61, 62] are not compatible with registration of arbitrary 2D slices into arbitrary 3D volumes. Therefore, we derive and train from scratch our own adaptation of the image-matching architecture known as LoFTR [54]. Our approach follows local feature matching with robust registration and refinement of the predicted pose with local optimization.

S2V registration by local feature matching. Image matching models usually predict sets of 2D-2D point-wise correspondences. In S2V registration we aim instead to associate 2D and 3D data. We propose that a set of point-wise correspondences $\mathcal{M}$, $card(\mathcal{M}) = N$ is estimated, with two sets of matching 3D key-points $C_I : \{c_{I_0}, c_{I_1}, ..., c_{I_N}\}$ and $C_V : \{c_{V_0}, c_{V_1}, ..., c_{V_N}\}$, where all points from C_I belong to the XY plane at $Z = 0$. The solution $\{R, T\}$ can then be recovered through minimization of

$$\underset{R,T}{\operatorname{argmin}} \frac{1}{N} \sum_{i}^{N} \left| c_{V_i} - A(R,T) c_{I_i} \right|^2 , \tag{5}$$

which can be solved in a closed form [18, 55], as long as $N \geq 3$, and the key-points are non-collinear.

[3] Intel(R) Core(TM) i9-7920X CPU @ 2.90GHz, NVIDIA GeForce RTX 2080 Ti

Table 4. Robust registration results (mAA). T for time / task.

	@ 5°	@ 10°	@ 20° ↑	T, s
Least-squares [55]	0.53 %	2.24 %	6.98 %	**0.294**
MAGSAC [5]	12.79 %	20.92 %	29.62 %	0.392
RANSAC [14]	12.65 %	22.98 %	33.99 %	0.591
GC-RANSAC [4]	13.68 %	24.13 %	35.64 %	0.297
Least-squares [55] + LO	23.64 %	23.90 %	25.47 %	0.549
MAGSAC [5] + LO	43.49 %	43.72 %	44.31 %	0.600
RANSAC [14] + LO	49.96 %	50.32 %	51.56 %	0.786
GC-RANSAC [4] + LO	**50.88 %**	**51.43 %**	**52.72 %**	**0.491**

Architecture of LoFTR-S2V. The key contribution of [54], the eponymous Local Feature Transformer (LoFTR) module, has enabled efficient use of transformers for image matching with outstanding results. We harness this approach for S2V registration, which requires a number of architectural changes to the original model. Our adaptation, which we name LoFTR-S2V, affects implementations of feature extraction, positional encoding, coarse- and fine-level matching, supervision and loss. First, we accommodate for mixed 2D-3D input by extracting features separately with two independent backbones, as opposed to using a single 2D backbone with shared weights. Then, we derive a definition for 3D positional encoding to be used with the extracted 3D features. This allows us to flatten both 2D and 3D features into 1D vectors, and perform coarse-level matching identically to the original model. Finally, we generalize the refinement of coarse-level matches to 3D case. We supervise both coarse- and fine-level matching with ground-truth correspondences produced by transforming points on the slice with $\hat{R}$ and $\hat{T}$. Aside from these changes, we adhere to the original approach. More details on architecture, training and raw performance of LoFTR-S2V are provided in the Supp. Mat., Sec.2.

Robust registration. Since high outlier rate is expected in the predicted $\mathcal{M}$, we approach Eq. (5) as a robust estimation task. To this end, we evaluate RANSAC [14], GC-RANSAC [4] and MAGSAC [5] (see Tab. 4), and establish that GC-RANSAC offers consistent advantage at all accuracy thresholds.

Refinement with local optimization. Finally, we allow the solutions recovered with Eq. (5) to be refined with a single pass of SLSQP in the fashion of Eq. (4), which leads to a significant improvement, in particular at finer accuracy thresholds. After refinement, LoFTR-S2V achieves

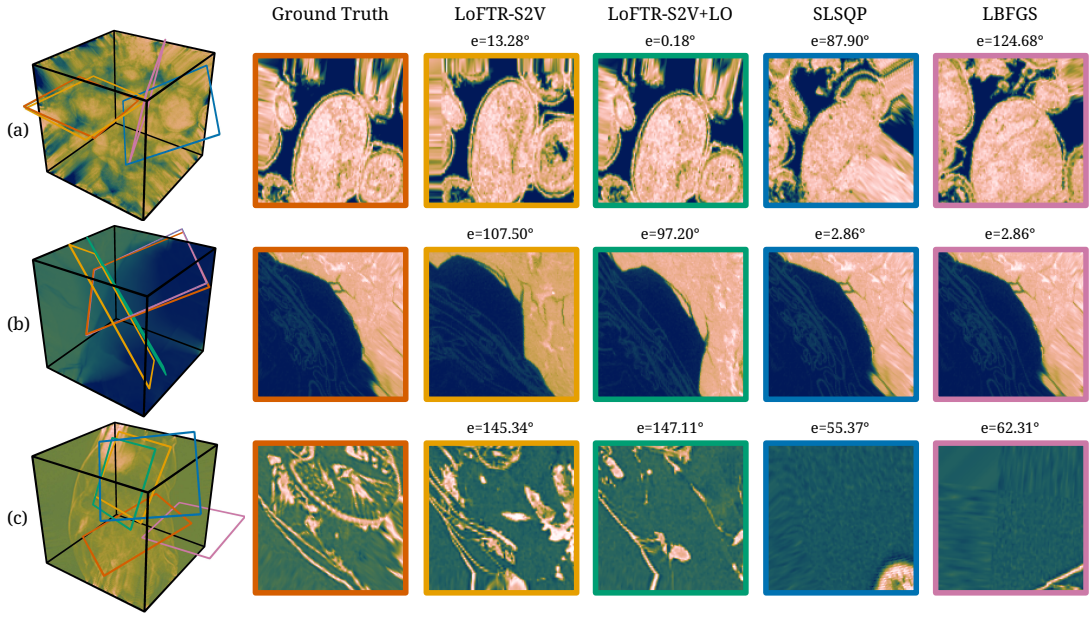

Figure 4. Qualitative results. (a) LO allows to refine coarsely-fit results from LoFTR-S2V, optimization-based methods fail; (b) LoFTR-S2V fails, optimization-based methods converge; (c) all methods fail. μ-CT of sandstone (a) [1], fossils (b) [41] and (c) [24].

Table 5. Test set results (mAA). LO marks 1 pass of SLSQP.

	@ 5°	@ 10°	@ 20°↑
SLSQP [27] (2,048)	50.52 %	51.17 %	52.58 %
LoFTR-S2V	13.51 %	25.13 %	36.99 %
LoFTR-S2V + LO	**52.08 %**	**52.57 %**	**53.78 %**

results comparable in terms of accuracy to the best performing optimization baselines (see Fig. 4 for qualitative examples). At the same time, LoFTR-S2V takes approx. $0.297s$ and $0.491s$ (with LO) per registration task, which is over 202 and 122 (with LO) times faster than exhaustive optimization-based search.

4.3. Evaluation results

Having validated our methodologies, we report the results for the complete test/benchmark set in Tab. 5. Despite the significant speed advantage with LoFTR-S2V, the dataset has demonstrated significant challenge for both hand-crafted and end-to-end learned methodologies. We provide additional considerations for limitations of our approach and representativeness of our benchmark ranking in Supp. Mat., Sec. 1.5.

5. Conclusions

In this paper, we introduced a novel dataset and benchmark for domain-agnostic S2V registration. Designed to address the challenges of rigid S2V across diverse scientific domains, the dataset provides $158,436$ unique registration tasks derived from real-world volumetric data, alongside accurate ground-truth solutions. Complementing this, we developed an online evaluation platform and strong baseline methodologies, including an optimization-based approach, as well as one based on deep learned local feature matching. Our work highlights the significant challenges of domain-agnostic S2V, while demonstrating the feasibility and potential for future advancements. Our contributions lower the barrier of entry into S2V algorithm development, enabling training and standardized evaluation of innovative algorithms. Moreover, our findings underscore the potential of learned local feature matching for efficient and accurate registration, paving the way for methods that can generalize across diverse applications in materials science, medical imaging, geology, and beyond. While our focus was on monomodal rigid registration, extending this work to cross-modal and non-rigid scenarios presents exciting opportunities. By providing a robust foundation for research and a clear baseline for future methods, we hope our contributions will catalyze further breakthroughs in domain-agnostic S2V registration and related tasks, impacting a wide range of scientific and industrial applications.

6. Acknowledgments

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 518560113. We thank the reviewers for their thoughtful feedback and recommendations, which has allowed us to improve the presentation of our contributions.

References

[1] Naif Alqahtani, Peyman Mostaghimi, and Ryan Armstrong. A Multi-Resolution Complex Carbonates Micro-CT dataset (MRCCM). Available: digitalrocksportal.org/projects/362, 2021. 1, 4, 8

[2] Matthew Andrew. *Reservoir condition pore scale imaging of multiphase flow using X-ray microtomography*. PhD thesis, 2014. 4

[3] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 3

[4] Daniel Barath and Jiri Matas. Graph-cut RANSAC. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018. 7

[5] Daniel Barath, Jiri Matas, and Jana Noskova. MAGSAC: marginalizing sample consensus. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019. 7

[6] Axel Barroso-Laguna, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. Key.Net: Keypoint Detection by Handcrafted and Learned CNN Filters. In *Int. Conf. Comput. Vis.*, 2019. 3

[7] Junyu Chen, Yihao Liu, Shuwen Wei, Zhangxing Bian, Shalini Subramanian, Aaron Carass, Jerry L. Prince, and Yong Du. A survey on deep learning in medical image registration: new technologies, uncertainty, evaluation metrics, and beyond. July 2023. 1, 2

[8] Fabio Crameri. Scientific colour maps. Available: doi.org/10.5281/ZENODO.1243862, 2023. 1

[9] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. ScanNet: Richly-annotated 3d reconstructions of indoor scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 2, 3

[10] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, June 2018. 3

[11] H. Dong. *Micro-CT Imaging and Pore Network Extraction*. PhD thesis, 2008. 4

[12] European Organization For Nuclear Research and OpenAIRE. Zenodo. Available: zenodo.org/, 2013. 3

[13] Enzo Ferrante and Nikos Paragios. Slice-to-volume medical image registration: A survey. *Med. Image Anal.*, 39:101–123, 2017. 1, 2, 3, 6

[14] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, jun 1981. 7

[15] Adam Gibson, Kathryn Piquette, Uwe Bergmann, William Christens-Barry, Graham Davis, M. Endrizzi, Shuting Fan, Sina Farsiu, Anthony Fitzgerald, Jennifer Griffiths, Cerys Jones, Guorong Li, Phillip Manning, Charlotte M-J, Roberta Mazza, David Mills, Peter Modregger, Peter Munro, Alessandro Olivo, and Melissa Terras. An assessment of multimodal imaging of subsurface text in mummy cartonnage using surrogate papyrus phantoms. *Heritage Science*, 6, 02 2018. 4

[16] Martin Hampe. Micro-CT dataset of a steel fibre reinforced concrete sample. Available: doi.org/10.5281/ZENODO.10785684, 2024. 4

[17] Gry Hoffmann Barfod, John Møller Larsen, Achim Lichtenberger, and Rubina Raja. Revealing text in a complexly rolled silver scroll from Jerash with computed tomography and advanced imaging software. *Sci. Rep.*, 5(1):17765, Dec. 2015. 1, 4

[18] Berthold KP Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Amer. A*, 4(4):629–642, 1987. 7

[19] Benjamin Hou, Nina Miolane, Bishesh Khanal, Matthew C. H. Lee, Amir Alansary, Steven McDonagh, Jo V. Hajnal, Daniel Rueckert, Ben Glocker, and Bernhard Kainz. *Computing CNN Loss and Gradients for Pose Estimation with Riemannian Geometry*, pages 756–764. Springer International Publishing, Cham, 2018. 2, 7

[20] Stefan Iglauer, Hamed Akhondzadeh, Hussein Abid, Adriana Paluszny, Alireza Keshavarz, Muhammad Ali, Ausama Giwelli, Lionel Esteban, Joel Sarout, and Maxim Lebedev. microCT scan images of hydrogen storage in coal. Available: doi.org/10.5285/84502681-F445-4A01-9DAD-C561A94E7C87, 2022. 4

[21] Yuanfeng Ji, Haotian Bai, Chongjian GE, Jie Yang, Ye Zhu, Ruimao Zhang, Zhen Li, Lingyan Zhanng, Wanling Ma, Xiang Wan, and Ping Luo. Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Adv. Neural Inform. Process. Syst.*, volume 35, pages 36722–36732. Curran Associates, Inc., 2022. 1, 4

[22] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image Matching across Wide Baselines: From Paper to Practice. *Int. J. Comput. Vis.*, 2020. 2, 3, 6

[23] Steven G. Johnson. The NLopt nonlinear-optimization package. Available: github.com/stevengj/nlopt, 2007. 7

[24] Kristaps Kairišs and Andris Bukejs. The original micro-CT scans of Monolepta rappsilberi, holotype, no T-I-K-30 [CIR], overall. Available: doi.org/10.5281/ZENODO.4900039, June 2021. 4, 8

[25] Florian Kleiner. Micro-CT scans of hydrated alite and belite from 7 to 84 days. Available: doi.org/10.5281/ZENODO.10778209, Mar. 2024. 4

[26] Jiří Kolibáč, Kateřina Rosová, Jan Simon Pražák, Jörg U. Hammel, and Jakub Prokop. The first larva of the cucujiform superfamily Cleroidea from the Mesozoic and its ecological implications (Coleoptera). *Arthropod Syst. Phylo.*, 81:289–301, 2023. 4

[27] Dieter Kraft. Algorithm 733: TOMP–Fortran modules for optimal control calculations. *ACM Transactions on Mathematical Software*, 20(3):262–281, Sept. 1994. 6, 7, 8

[28] Zhengqi Li and Noah Snavely. MegaDepth: Learning single-view depth prediction from internet photos. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018. 2, 3

[29] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. LightGlue: Local feature matching at light speed. In *Int. Conf. Comput. Vis.*, 2023. 3

[30] Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Math. Program.*, 45:503–528, 1989. 6, 7

[31] Fabian Lutter. *Elementsensitive Bildgebung - Einsatz chromatischer Pixelarrays in Röntgen nano-CT*. PhD thesis, Universität Würzburg, 2023. 4

[32] Fabian Lutter and Martin Hampe. µCT datasets of a fly, a walnut, a wooden dowel and wooden meeple. Available: doi.org/10.5281/ZENODO.10784902, 2024. 4, 5

[33] J.B. Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Med. Image Anal.*, 2(1):1–36, 1998. 2

[34] Viktoria Markova, Matteo Ronchetti, Wolfgang Wein, Oliver Zettinig, and Raphael Prevost. *Global Multi-modal 2D/3D Registration via Local Descriptors Learning*, pages 269–279. Springer Nature Switzerland, Cham, 2022. 3

[35] Mahoor Mehdikhani, Ilya Straumit, Larissa Gorbatikh, and Stepan V. Lomov. Detailed characterization of voids in multidirectional carbon fiber/epoxy composite laminates using X-ray micro-computed tomography. *Composites Part A: Applied Science and Manufacturing*, 125:105532, 2019. 4

[36] Anastasiya Mishchuk, Dmytro Mishkin, Filip Radenovic, and Jiri Matas. Working hard to know your neighbor's margins: Local descriptor learning loss. In *Adv. Neural Inform. Process. Syst.*, Dec. 2017. 3

[37] Rodrigo Neumann, Mariane Andreeta, and Everton Lucas-Oliveira. 11 sandstones: raw, filtered and segmented data. Available: digitalrocksportal.org/projects/317, 2020. 4

[38] Max Patzelt and Martin Hampe. µCT-scans of a selection of concrete samples made of cement and coarse aggregates without sand fraction. Available: doi.org/10.5281/ZENODO.10784827, 2024. 4

[39] Nicolas Pielawski, Elisabeth Wetzer, Johan Öfverstedt, Jiahao Lu, Carolina Wählby, Joakim Lindblad, and Nataša Sladoje. CoMIR: Contrastive multimodal image representation for registration. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Adv. Neural Inform. Process. Syst.*, volume 33, pages 18433–18444. Curran Associates, Inc., 2020. 3

[40] M. J. D. Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. In S. Gomez and J.-P. Hennart, editors, *Advances in Optimization and Numerical Analysis*, volume 275 of *Mathematics and Its Applications*, pages 51–67. Springer, 1994. 6, 7

[41] Adam C. Pritchard and Sterling J. Nesbitt. Data from: A bird-like skull in a Triassic diapsid reptile increases heterogeneity of the morphological and phylogenetic radiation of Diapsida. Available: doi.org/10.5061/DRYAD.F5Q10, 2017. 1, 4, 8

[42] M. Prodanovic, M. Esteva, M. Hanlon, G. Nanda, and P. Agarwal. Digital Rocks Portal: a sustainable platform for imaged dataset sharing, translation and automated analysis. In *AGU Fall Meeting Abstracts*, volume 2015, pages MR43A–02, Dec. 2015. 4

[43] Jerome Quenum, Iryna Zenyuk, and Daniela Ushizima. 3D microCT of lithium metal battery after charge and discharge. Available: doi.org/10.6078/D1FM8J, 2023. 4

[44] Jerome Revaud, Philippe Weinzaepfel, César Roberto de Souza, and Martin Humenberger. R2D2: repeatable and reliable detector and descriptor. In *Adv. Neural Inform. Process. Syst.*, 2019. 3

[45] Thomas Harvey Rowan. *Functional stability analysis of numerical algorithms*. PhD thesis, Department of Computer Science, University of Texas at Austin, Austin, TX, 1990. 6, 7

[46] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022. 3

[47] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 3

[48] Hugo Saur, Charles Aubourg, and Peter Moonen. X-ray micro-CT images of calcareous shale samples. Available: digitalrocksportal.org/projects/370, 2021. 4

[49] Johannes Lutz Schönberger, Hans Hardmeier, Torsten Sattler, and Marc Pollefeys. Comparative Evaluation of Hand-Crafted and Learned Local Features. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 3

[50] Thomas L. Semple, Rod Peakall, and Nikolai J. Tatarnic. A comprehensive and user-friendly framework for 3D-data visualisation in invertebrates and other organisms. *Journal of Morphology*, 280(2):223–231, Jan. 2019. 4

[51] Shiva Shirani, Ana Cuesta, Alejandro Morales-Cantero, Isabel Santacruz, Ana Diaz, Pavel Trtik, Mirko Holler, Alexander Rack, Bratislav Lukic, Emmanuel Brun, Inés R. Salcedo, and Miguel A. G. Aranda. Dataset for "4D nanoimaging of early age cement hydration" Nature Communications paper. Available: doi.org/10.5281/ZENODO.7030106, 2023. 4

[52] Guoli Song, Jianda Han, Yiwen Zhao, Zheng Wang, and Huibin Du. A review on medical image registration as an optimization problem. *Current Medical Imaging*, 13(3):274–283, 2017. 2

[53] Gregory W. Stull, Neil F. Adams, Steven R. Manchester, Dan Sykes, and Margaret E. Collinson. Data from: Revision of icacinaceae from the early eocene london clay flora based on X-ray micro-CT. Available: doi.org/10.5061/DRYAD.706DF, 2017. 4

[54] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with transformers. *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021. 3, 7

[55] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(4):376–380, 1991. 7

[56] Henry O Velesaca, Gisel Bastidas, Mohammad Rouhani, and Angel D Sappa. Multimodal image registration techniques: a comprehensive survey. *Multimed. Tools Appl.*, 83(23):63919–63947, Jan. 2024. 3

[57] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods*, 17:261–272, 2020. 7

[58] Jakob Wasserthal, Hanns-Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, Daniel T. Boll, Joshy Cyriac, Shan Yang, Michael Bach, and Martin Segeroth. TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023. 1, 4

[59] U. Weller, L. Albrecht, S. Schlüter, and H.-J. Vogel. An open *Soil Structure Library* based on X-ray CT data. *SOIL*, 8(2):507–515, 2022. 1, 4

[60] Mark D Wilkinson, Michel Dumontier, I Jsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S Grethe, Jaap Heringa, Peter A C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J Lusher, Maryann E Martone, Albert Mons, Abel L Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*, 3(1):160018, Mar. 2016. 2

[61] Junshen Xu, Daniel Moyer, P. Ellen Grant, Polina Golland, Juan Eugenio Iglesias, and Elfar Adalsteinsson. *SVoRT: Iterative Transformer for Slice-to-Volume Registration in Fetal Brain MRI*, pages 3–13. Springer Nature Switzerland, Cham, 2022. 2, 7

[62] Pak-Hei Yeung, Moska Aliasi, Aris T Papageorghiou, Monique Haak, Weidi Xie, and Ana IL Namburete. Learning to map 2D ultrasound images into 3D space with minimal human annotation. *Med. Image Anal.*, 70:101998, 2021. 2, 7

[63] Kwang Moo Yi, Eduard Trulls, Yuki Ono, Vincent Lepetit, Mathieu Salzmann, and Pascal Fua. Learning to find good correspondences. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2666–2674, 2018. 6

[64] Jing Yu, Ryan Armstrong, Peyman Mostaghimi, and Aaron Uthaia Kumaran. Micro-CT images of a coal sample (coal-2). Available: figshare.com/articles/dataset/X-Ray_micro_CT_of_coal_sample_2_used_in_4D_PET_imaging_of_11C_CO2_flow_dynamics_in_coal_/23949450, 8 2023. 4

[65] Xiaoming Zhao, Xingming Wu, Jinyu Miao, Weihai Chen, Peter C. Y. Chen, and Zhengguo Li. ALIKE: Accurate and lightweight keypoint detection and descriptor extraction. *IEEE Trans. Multimedia*, 25:3101–3112, 2023. 3