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

1. Dataset curation details

1.1. Details on volumetric data sourcing

Our dataset is curated from real-world volumetric data, published in various fields of academic research. In Tab. 1 we provide details such as categorization, application domain or context, primary instrument of measurement, and licensing. We also provide the count of unique volumes we were able to source, as well as the count of the volumes we kept after sub-selection.

1.2. Use of personal data and human subjects

Since "Needles & Haystacks" includes two sources of volumetric data from human subjects, we investigate the ethics compliance of the original published works. Importantly, both medical datasets [9, 27] have obtained corresponding ethical waivers, and have taken material steps to anonymize the data by removing personally identifiable features. *AMOS* [9] reports approval by the Research Ethics Committees of Longgang District People's Hospital (reference number: 2021077), as well as Longgang District Central Hospital (reference number: 2021ECJ012). *TotalSegmentator* [27] reports approval by The Ethics Committee Northwest and Central Switzerland (EKNZ BASEC Req-2022–00495).

1.3. Definition of $A_{subvolume}$

To sample a subvolume, we choose a translation T_O to some origin point O, around which we choose a rotation R_O. Then, we choose offset T_V and a non-uniform scaling factor S_V to sample and rescale a parallelepiped centered at O into a cubic shape of $D \times D \times D$ voxels. We express this transformation using an affine matrix $A_{subvolume}$

$$A_{subvolume} = S_V T_V^{-1} T_O R_O T_O^{-1}, \tag{1}$$

the inverse of which allows to sample subvolume $V = V_{original}\left(A_{subvolume}^{-1}\right)$ from the original volume $V_{original}$.

1.4. Visualizations of sampled registration tasks

We provide further illustrations for the registration tasks within our dataset, sampled from each unique source and organized by category: "Materials Characterization" (see Fig. 4), "Life Sciences" (see Fig. 5), and "Paleo-, Archeo- and Anthropology" (see Fig. 6).

1.5. Curation bias and benchmark ranking

The representative strength of methodology ranking in our benchmark hinges on the implicit factors such as the dataset composition, the size of the registration search space, as well as the choice of parameters for registration task validation. Limited by the data availability within published academic research, our approach is inherently biased to unintentionally emphasize some distributions of volumetric data, together with their characteristics: from favorable, well-structured and near-noiseless, to highly challenging, repetitive, noisy, thin-structured and disjoint. Therefore, an alternative approach to dataset composition, even if guided by the same principles, may well emphasize different strengths and weaknesses in S2V algorithms, which would lead to *some* variability in the rankings. Nonetheless, it is unlikely that a *significantly* different methodology ranking can be achieved in the *general-purpose* setting. Note that this would require that, *at the same time*, a sufficiently large set of application domains with distinct characteristics was overlooked during the curation of this dataset, *and* that the challenges of such data were sufficiently distinct to upend the major trends in algorithm performance.

2. LoFTR-S2V implementation details

2.1. Architecture of LoFTR-S2V

While we provide a brief overview of architectural adaptations that allow us to apply the detector-free image matching approach from LoFTR [25] in the main paper, the limitations on paper length prevent us from going into a more exhaustive discussion. Therefore, we discuss some additional details in this subsection. We assume at least a brief familiarity of the reader with [25], and focus our discussion only on the *adaptations* of the original approach. The overview is presented in Fig. 1.

Feature extraction. While the original LoFTR is trained on co-visible pairs of 2D images, LoFTR-S2V must accom-

Table 1. Sources of data with counts of available ($N_{available}$) and kept (N_{kept}) compatible 3D volumes.

Source	Category	Application	Instrument	$N_{available}$	$N_{kept}\downarrow$	License
TotalSegmentator [27]	Life Sciences	Medical imaging	CT, MRI	926	343	CC BY 4.0
AMOS [9]	Life Sciences	Medical imaging	CT, MRI	904	298	CC BY 4.0
Gibson *et al*., 2018 [4]	Paleo-, Archeo- & Anthropology	Cultural heritage	μ-CT	125	125	CC0 1.0
COAL-2 [29]	Materials Characterization	Carbon sequestration	μ-CT	2420	98	CC BY 4.0
Soil Structure Library [28]	Materials Characterization	Geology/soil studies	segmented μ-CT	654	80	CC BY 4.0
Saur, Aubourg & Moonen, 2021 [21]	Materials Characterization	Geology/rock studies	μ-CT	256	77	ODC-BY
μ-CT samples of concrete [5, 11, 18, 23]	Materials Characterization	Materials science/concretes	μ-CT	277	77	CC BY 4.0
Mehdikhani *et al*., 2019 [16]	Materials Characterization	Materials science/polymers	μ-CT	175	72	CC BY 4.0
Quenum, Zenyuk & Ushizima, 2023 [20]	Materials Characterization	Battery research	μ-CT	120	64	CC0 1.0
MRCCM [1]	Materials Characterization	Geology/rock studies	μ-CT	99	60	ODC-BY
11 Sandstones [17]	Materials Characterization	Geology/rock studies	μ-CT	88	56	ODC-BY
ICL μ-CT Images and Networks [2, 3]	Materials Characterization	Geology/rock studies	segmented μ-CT	83	54	CC BY 4.0
Iglauer *et al*., 2022 [8]	Materials Characterization	Geology	μ-CT	24	24	OGL
Stull *et al*., 2017 [24]	Paleo-, Archeo- & Anthropology	Paleontology	μ-CT	23	23	CC0 1.0
μ-CT test samples: a fly, a walnut [15]	Life Sciences	Lab samples	μ-CT	10	10	CC BY 4.0
μ-CT test samples: a dowel, a pawn [15]	Paleo-, Archeo- & Anthropology	Lab samples	μ-CT	10	10	CC BY 4.0
Semple, Peakall & Tatarnic, 2020 [22]	Life Sciences	Zoology/morphology	μ-CT	8	8	CC0 1.0
Lutter, 2023 [14]	Materials Characterization	Lab samples	nano-CT	4	4	CC BY 4.0
Hoffmann Barfod *et al*., 2016 [7]	Paleo-, Archeo- & Anthropology	Cultural heritage	μ-CT	4	4	CC BY 4.0
Kolibáč *et al*., 2023 [12]	Life Sciences	Systematics & phylogeny	μ-CT	4	4	CC BY 4.0
Pritchard & Nesbitt, 2017 [19]	Paleo-, Archeo- & Anthropology	Archeology	μ-CT	4	4	CC0 1.0
Kairišs & Bukejs, 2021 [10]	Life Sciences	Paleontology	μ-CT	3	3	CC BY 4.0

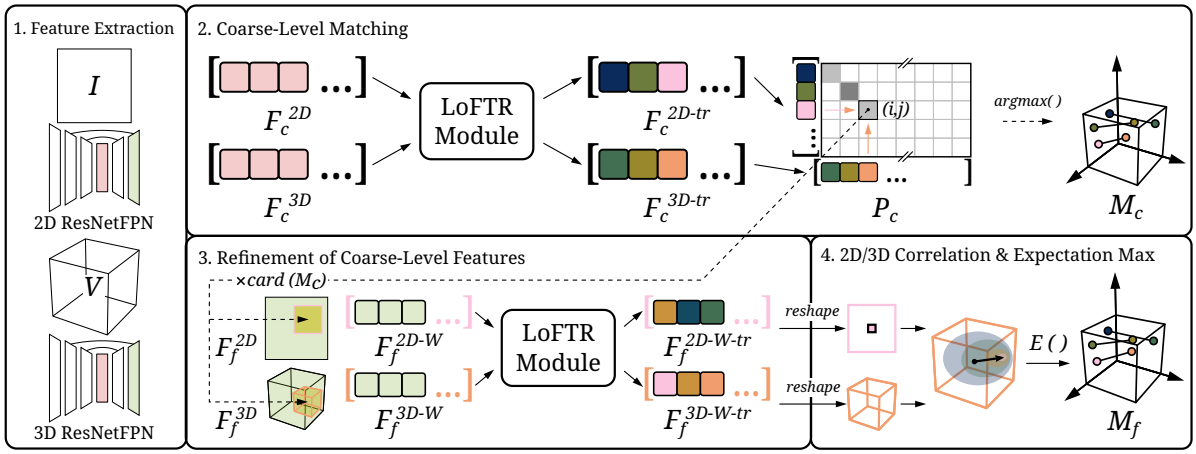

Figure 1. LoFTR [25] architecture adapted for S2V registration. Non-differentiable operations in dashed.

modate for matching between 2D images and 3D volumes. This requires that two separate feature extraction modules, one a 2D and one a 3D, are used instead of a single 2D convolutional neural network (CNN) based on Feature Pyramid Network (FPN) [13] and ResNet-18 [6]. Keeping the original 2D ResNetFPN unchanged, we also define a 3D ResNetFPN for 3D feature extraction. To this end, we substitute the 2D operations with their 3D counterparts, and keep to the same layer topology as the original. Both FPNs extract 3-level features, from which the coarsest-level features, correspondingly F_c^{2D} and F_c^{3D}, form the basis for coarse matching, while the fine level features F_f^{2D} and F_f^{3D} are used for refinement of coarse-level matches.

Positional encoding. To enhance the extracted features with positional dependence, the original approach calls for addition of position encoding in sinusoidal format. While the definition of positional encoding remains identical in the 2D case, a 3D formulation is required for 3D features. We define it as follows:

$$\mathcal{PE}^i_{x,y,z} = f(x,y,z)^i := \begin{cases} \sin(\omega_k \cdot x), & i = 6k \\ \cos(\omega_k \cdot x), & i = 6k+1 \\ \sin(\omega_k \cdot y), & i = 6k+2 \\ \cos(\omega_k \cdot y), & i = 6k+3 \\ \sin(\omega_k \cdot z), & i = 6k+4 \\ \cos(\omega_k \cdot z), & i = 6k+5 \end{cases}, \tag{2}$$

where $\omega_k = \frac{1}{10000^{3k/d)}}$, d is the number of channels, and i is an index for feature channels, similar to the origi-

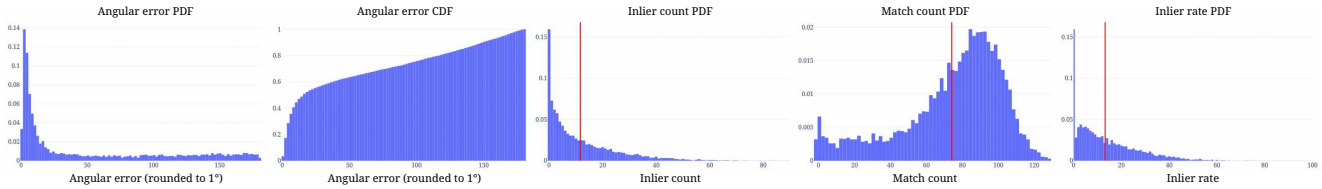

Figure 2. PDF and CDF visualizations for distributions of angular error, as well as feature matching statistics on the test set. Highlighted in red are the average count of inlier (11.65) and total predicted (74.11) matches, average inlier rate (13.21 %).

nal formulation. The position-dependent features $F_c^{2D} = F_c^{2D} + \mathcal{PE}_{x,y}^i$ and $F_c^{3D} = F_c^{3D} + \mathcal{PE}_{x,y,z}^i$ are then flattened and propagated as 1D feature vectors through the original LoFTR module for coarse-level matching. Note that LoFTR module is compatible with differently-sized feature vectors, and needs no adaptations to accommodate for imbalance in sizes between F_c^{2D} and F_c^{3D}.

Coarse-level matching & supervision. Transformed features F_c^{2D-tr} and F_c^{3D-tr} are matched identically to the 2D-2D case, with the use of the dual-softmax operator and mutual nearest neighbor search. Importantly, coarse-level supervision requires a ground-truth matrix of correspondences $\mathcal{M}_c^{gt}$ at 1/8th of the original resolution. While the original recovers these correspondences using both camera poses and depths, for S2V registration $\mathcal{M}_c^{gt}$ is recovered trivially as a mutual nearest neighbor between the coarse $D \times D$ grid of I (grid vertices initialized at voxel centers), transformed with ground-truth $\hat{A}(\hat{R}, \hat{T})$, and the coarse $D \times D \times D$ grid of V. Coarse-level matching completes with a selection of a set of good matches $\mathcal{M}_c$, for which, given the much higher ambiguity of S2V registration, we lower the confidence threshold θ_c to 0.001.

Fine-level matching & supervision. The original approach samples support windows of size $w \times w$ from fine-level 2D features to refine select coarse matches to subpixel-accurate coordinates. Importantly, the search for the subpixel location takes place in only one of the windows, and the central location of the other serves as the reference point. During refinement, the match is therefore fixed in the space of one image, and moving in the space of the other. The nuance of replicating such refinement between 2D and 3D spaces, is that the extracted correspondences must always be refined in 3D space. We sample the 2D windows $w \times w$ from the F_f^{2D} at corresponding positions, enhance them with 2D positional encoding, flatten and propagate them through the LoFTR module to get $F_f^{2D-W-tr}$. Similarly, 3D windows $w{\times}w{\times}w$ are sampled from F_f^{3D}, encoded and transformed to recover $F_f^{3D-W-tr}$. Both $F_f^{2D-W-tr}$ and $F_f^{3D-W-tr}$ are reshaped into their corresponding 2D $w{\times}w$ and 3D $w{\times}w{\times}w$ shapes. Now, we sample $F_f^{2D-W-tr}$ at its central position ($w/2$,$w/2$), and correlate it with all positions within $F_f^{3D-W-tr}$ to produce a scalar 3D volume of shape $w \times w \times w$. Similar to the original approach, we propagate the produced correlation volume through a soft-argmax, and compute a 3D spatial expectation as the centroid of the input tensor in spatial dimensions. At inference time, an argmax operator on the output returns the refined endpoint of the correspondence in the 3D space. This way we acquire the final set of fine matches $\mathcal{M}_f$.

2.2. Training of LoFTR-S2V

LoFTR-S2V is trained from scratch on our entire dataset. We train on 24 GPUs[1] in batches of 24, otherwise keeping the training schedule similar to the original. We train with Adam-W and weight decay of 0.1 with gradient clipping, initial learning rate set to 0.006, and a warm up for 3 epochs from 0.1 of the initial learning rate, which we decay every 3 epochs by the factor of 0.5. During training, we solve for the pose with Least-Squares [26] and observe the training metrics for the entire validation set, until convergence after 14 epochs, in approx. 192 hours.

2.3. Visualizations of LoFTR-S2V results

In the main paper, we report on the summary S2V registration performance of LoFTR-S2V, as well as on its improvement after refinement with local optimization (LO). To evaluate solutions we use angular error, which we aggregate into mean Average Accuracy (mAA) metric. Here, we provide extended details on the distribution of angular error for LoFTR-S2V on the entire test set, before refinement with LO. Similar analysis is provided for all new submissions to the benchmark. In Fig. 2 we visualize the distribution of angular errors, as well as the distributions of matching statistics, such as the counts of inlier and total correspondences, and inlier rates.

Finally, to offer a better qualitative understanding of the results with LoFTR-S2V, we provide additional visualization of registration results on the validation set in Fig. 3. For balanced representation, we demonstrate a wide range of results, from lowest to highest possible angular errors.

[1] NVIDIA GeForce RTX 2080 Ti

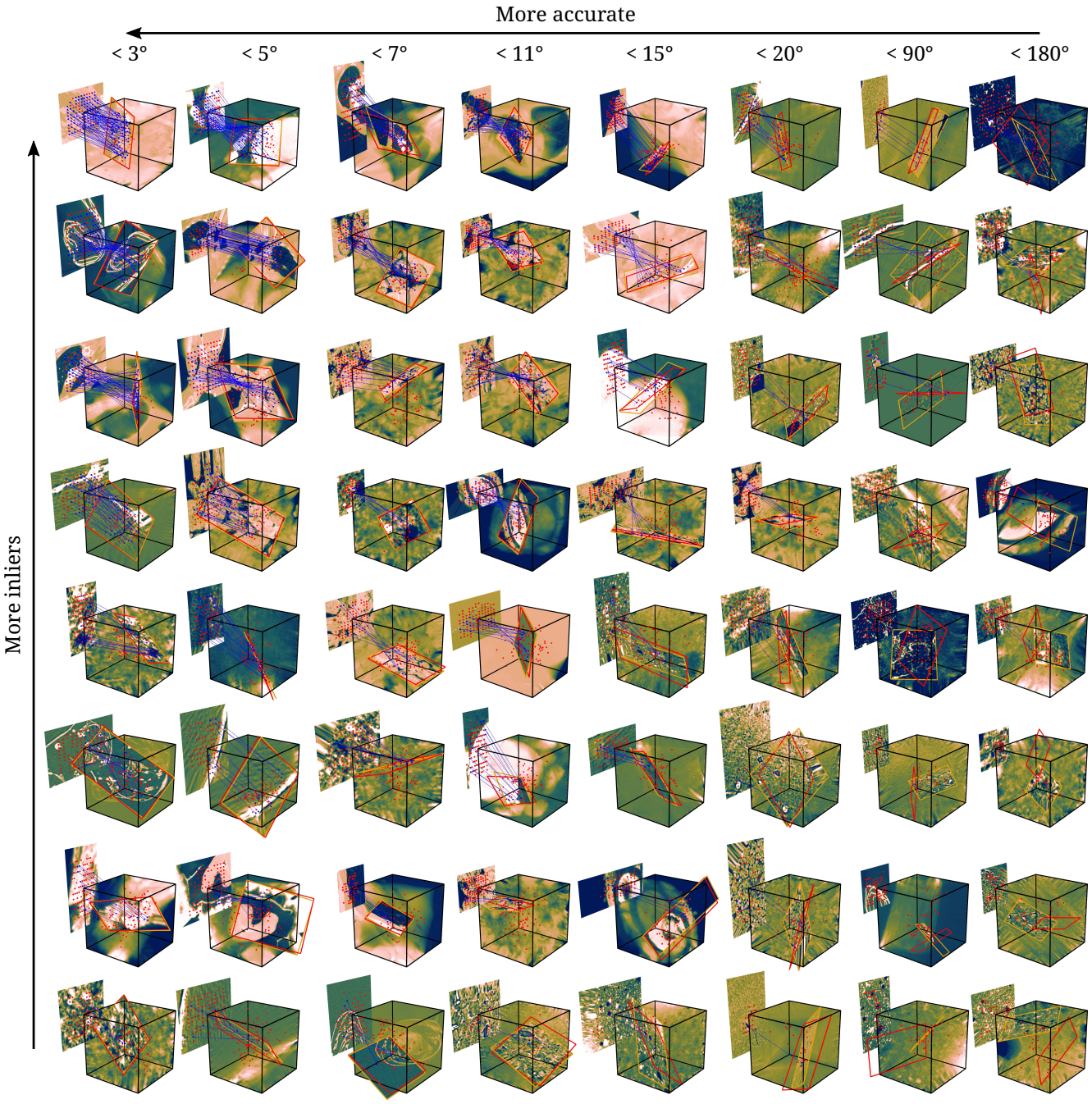

Figure 3. Qualitative results on the validation set for LoFTR-S2V. Ground-truth in red, solution in yellow. True inlier matches in blue, red points are outlier endpoints. Samples sorted left to right by increasing angular error, top to bottom by decreasing count of inlier matches.

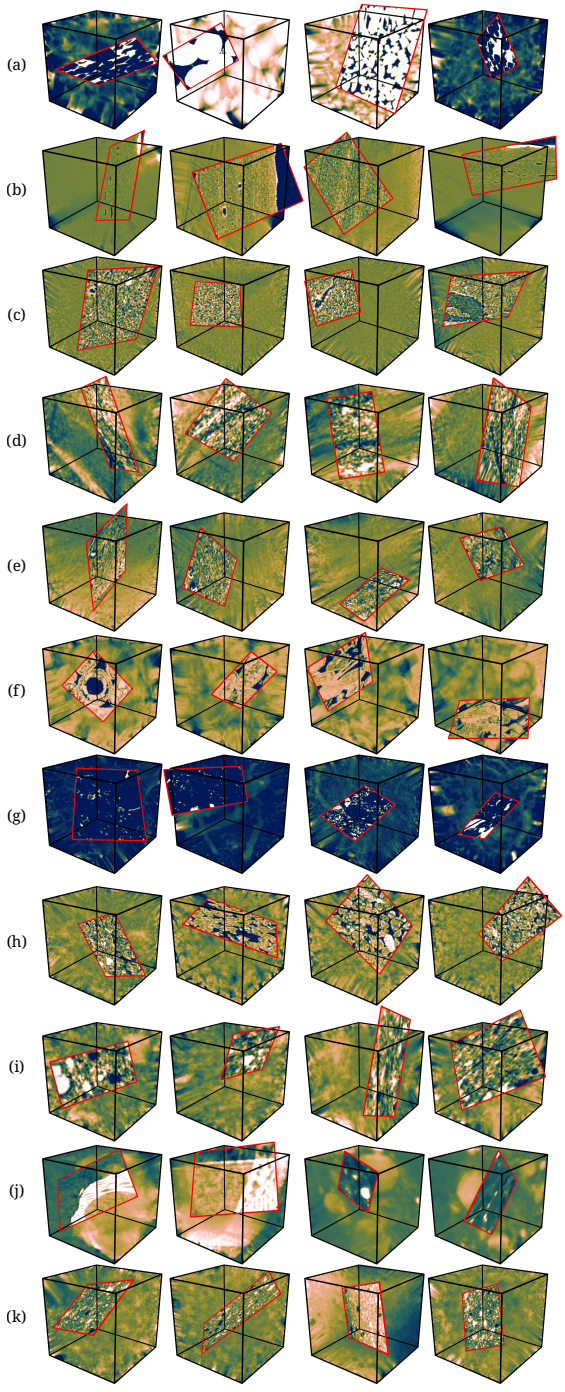

Figure 4. Registration tasks from "Materials Characterization" category. Sourced from geological studies of rocks (a,f,h,i) and soils (g), carbon storage in coal (c), studies of concretes (k) and polymers (b) in materials science, battery research (e) and carbon sequestration (d).

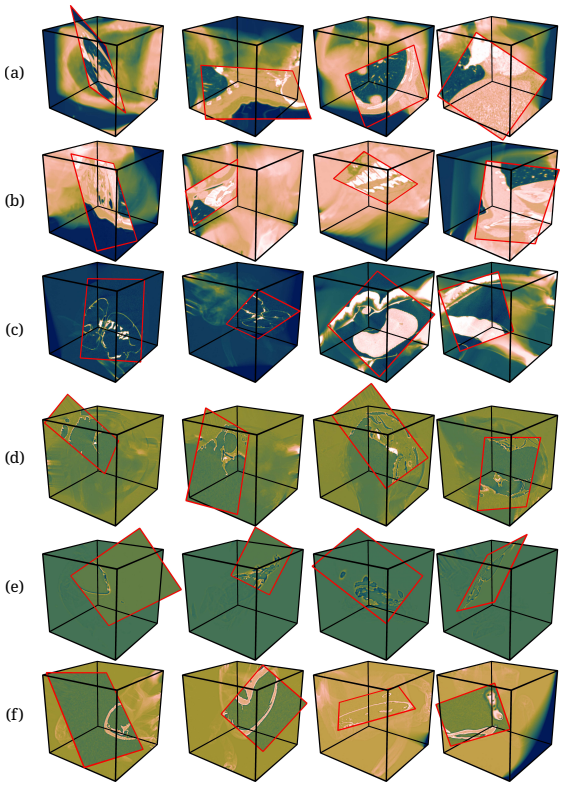

Figure 5. Registration tasks from "Life Sciences" category. Sourced from medical imaging data (a,b), lab samples (c), and biological studies of insects (d,e,f).

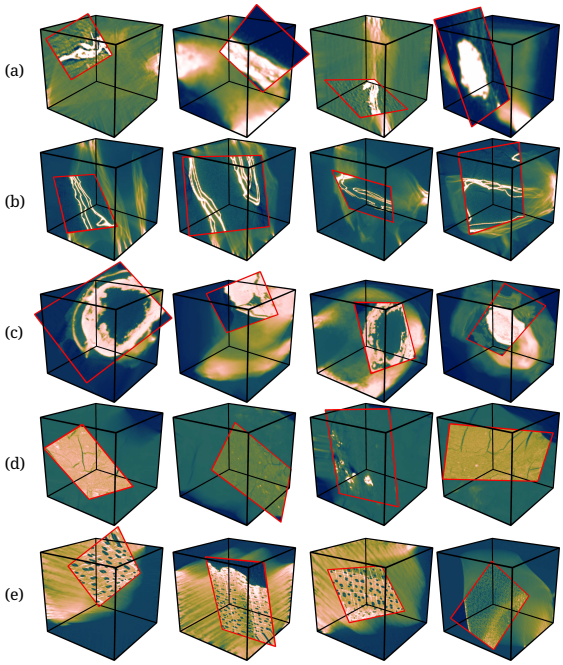

Figure 6. Registration tasks from "Paleo-, Archeo- & Anthropology" category. Sourced from various introspective studies of cultural artifacts (a,b), fossils (c,d), and lab samples (e).

References

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

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

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

[4] 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.

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

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, June 2016.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature Pyramid Networks for object detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, July 2017.

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

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

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

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.