

Tassilo Wald, PhD

Research Scientist | Self-Supervised Learning, Vision-Language Models & Representation Learning

Heidelberg, Germany | tassilo.wald@gmail.com | [Homepage](#) | [Google Scholar](#) (1,900+ citations, h-index 17) | [LinkedIn](#)

SUMMARY

Foundation-model researcher working across self-supervised learning, vision-language models, and representation learning, with research carried out on 2D and 3D image data. Pre-trained multimodal foundation models on 2M+ 3D volumes at up to 96 GPUs per run and built a 3D VLM that improved report generation by +20 Macro-F1 over the prior state of the art, now used by Microsoft's Mayo Clinic collaborators. Open-source releases (nnU-Net architectures, OpenMind, nnssl) are reference tooling across the field.

TECHNICAL SKILLS

- **Foundation models:** self-supervised learning (MAE, contrastive, JEPA, DINO), vision-language pre-training, multimodal fusion (3D vision + text), SFT / LLM fine-tuning, longitudinal and temporal modeling, image analysis (segmentation, classification), interactive segmentation
- **Scale & systems:** PyTorch, Lightning, DDP, NCCL, multi-node Slurm HPC (96 GPUs/run), mixed precision, reproducible large-scale training pipelines
- **Data:** natural images, 3D volumetric data (CT, MRI, PET), video, paired image-text corpora, structured tabular data, DICOM / NIfTI, data curation at 1 TB+ (OpenMind) / 2M+ volume scale (The Human Radiome Project)
- **Engineering:** Git & CI/CD, Docker, W&B / MLflow, HuggingFace
- **AI-assisted development:** Claude Code, Codex, agentic workflows for experimentation and research throughput

EXPERIENCE

Microsoft Research

Cambridge, UK

Research Intern

May 2025 - Aug 2025

- Developed a 3D vision-language foundation model trained on paired volumetric imaging and text reports, establishing joint image-text representations for 3D understanding. [ECCV 2026, first author]
- Fine-tuned (SFT) an LLM decoder with the 3D foundation model for report-generation yielding a VLM that beat the prior best by +20 Macro-F1. The resulting models are used by Microsoft's Mayo Clinic collaborators in clinical deployment.

German Cancer Research Center (DKFZ)

Heidelberg, Germany

Doctoral Researcher, Medical Image Computing | Advisor: Prof. Klaus H. Maier-Hein

Apr 2020 - Jun 2026

- **Scale & leadership.** Co-led The Human Radiome Project, comprising >80 researchers at DKFZ and 6 industrial and academic partner institutions. Led The Human Radiome Project pre-training team, training a multimodal foundation model on 2M+ 3D volumes spanning every imaging modality, scaling to 96 GPUs per run on a HPC system. Directly supervised 6 team members. [under review at Nature Biomedical Engineering]
- **Industry translation.** Led an industry collaboration with the AI startup Floy, delivering a large, pre-trained brain-MRI foundation model now in their production pipeline. [CVPR 2025 Highlight]
- **Self-supervised learning.** Built the first unified benchmark for 3D SSL, showing MAE pre-training outperforms newer methods under matched conditions. Released OpenMind (115k volumes / 1 TB), now the field's reference dataset and benchmark, plus nnssl, an open reference implementation of the major 3D SSL methods. Organized the MICCAI 2025 challenge built on it. [CVPR 2025 Highlight, ICCV 2025, Challenge under Review at Nature Communications]
- **Architecture & rigor.** Led nnU-Net Revisited (810 citations), contributing a new CNN architecture and demonstrating that prior transformer claims fail under rigorous validation. The follow-up, Primus, delivers a transformer that finally outperforms the strong CNN baselines. [MICCAI 2024, TMLR 2026]
- **Representation analysis.** Showed that spatial alignment confounds standard similarity measures between independently trained networks and introduced a Representational Similarity Benchmark. [NeurIPS 2024 w/ Google DeepMind, ICLR 2025]
- **Longitudinal & temporal modeling.** Co-authored LesionLocator, a zero-shot framework for universal object segmentation and tracking across timepoints in 3D whole-body imaging, and LongiSeg, an open framework exploiting temporal context across follow-up scans for longitudinal segmentation. [CVPR 2025, MICCAI 2024]
- **Downstream validation.** Detection models validated on 800+ patients across five external cohorts; cross-technique annotation raised lesion-wise F1 by 2.5 to 9.6 points on every external dataset, finding sub-centimeter targets missed by expert readers. [Neuro-Oncology Advances 2022, European Radiology Experimental 2025]

- Temporal feature aggregation for CNN-based video object detection. [IEEE Intelligent Vehicles 2021]

OPEN SOURCE

- **nnU-Net** (+8.9k Stars on Github): Maintainer since 2023. The standard framework for 3D dense prediction.
- **OpenMind** (+90k Downloads last month): 115,000-volume / 1 TB 3D dataset and benchmark for 3D SSL research.
- **nnssl** (+210 Stars on Github): Authored nnssl, a reference repository of relevant SSL methods for 3D vision.

EDUCATION

PhD, Computer Science | DKFZ / Heidelberg University | *grade 1.0 (highest grade), magna cum laude* defended Jun 2026

M.Sc., Electrical Engineering & Information Technology | Karlsruhe Institute of Technology (KIT) | *grade 1.3* 2016 - 2019

Thesis (1.0): Combination of Temporal and Spatial Information Extraction Within a CNN for Improving Object Detection. Specialization: Machine Learning, Control Theory

SELECTED PUBLICATIONS

First or equal-contribution author (*). Full list: [Google Scholar](#)

- Ulrich C. *, Wald T. *, Gotkowski K. *, Kirchhoff Y. *, Knopp M. *, Rokuss M. * et al. "nnFoundation: 3D Foundation Models for Radiology." Preprint, arXiv:2609.26924, 2026.
- Wald T. et al. "Comprehensive Language-Image Pre-training for 3D Medical Image Understanding." ECCV 2026. (Microsoft Research)
- Wald T. *, Roy S. *, Isensee F. * et al. "Primus: Enforcing Attention Usage for 3D Medical Image Segmentation." TMLR 2026.
- Wald T. *, Ulrich C. *, Suprijadi J. * et al. "An OpenMind for 3D Medical Vision Self-Supervised Learning." ICCV 2025.
- Wald T. *, Ulrich C. * et al. "Revisiting MAE Pre-training for 3D Medical Image Segmentation." CVPR 2025 (Highlight).
- Wald T. *, Klabunde M. *, Schumacher T. * et al. "ReSi: A Comprehensive Benchmark for Representational Similarity Measures." ICLR 2025.
- Wald T., ..., Jaini P. et al. "Decoupling Semantic Similarity from Spatial Alignment for Neural Networks." NeurIPS 2024. (with Google DeepMind)
- Isensee F. *, Wald T. *, Ulrich C. *, Baumgartner M. * et al. "nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation." MICCAI 2024. (810 citations)
- Wald T. *, Hamm B. * et al. "Enhancing deep learning methods for brain metastasis detection through cross-technique annotations on SPACE MRI." European Radiology Experimental 9(1), 2025.

CHALLENGES & AWARDS

- Extraordinary Contributions Award, CVPR 2025, ALLDATA track, Foundation Models for Interactive 3D Biomedical Image Segmentation workshop, for nnInteractive
- MICCAI Challenge (1st place / 13): Mediastinal Lymph Node Quantification (LNQ) Challenge, 2023
- MICCAI Challenge (1st place / 20): Abdominal Multi-Organ Segmentation (AMOS) Challenge, 2022
- MICCAI Challenge (2nd place): HNTS-MRG Challenge, Task 2, mid-RT segmentation, 2024
- MICCAI Challenge (Overall Top Performer): ABUS Challenge, 3D breast ultrasound, 2023
- HiDA Hackathon (1st place): AI-HERO Hackathon for Energy-Efficient AI, 2022

SERVICE

- [Invited talk at Cohere](#). Other speakers in this series include e.g. Lucas Beyer
- Organizer, [Heidelberg.ai](#), since 2023. Local AI Research Community of >3000 Members
- Organizer of MICCAI Challenge, SSL3D Challenge at MICCAI, 2025 [Challenge under Review at Nature Communications]
- Outstanding Reviewer, CVPR 2025. Regular reviewer for NeurIPS, ICLR, ICML, CVPR, ICCV/ECCV