Michael Dorkenwald

🏠 mdorkenwald.com | ✉ m.dorkenwald@gmail.com | in m-dorkenwald | mdorkenw

Education

PhD Artificial Intelligence — Amsterdam, Netherlands
ELLIS & QUVA Lab, University of Amsterdam — *June 2022- 2027*

Supervisors: Prof. Cees Snoek (University of Amsterdam) & Prof. Yuki Asano (University of Technology Nuremberg)
Research interests: Self-supervised video learning, multimodal vision-language models, and efficient foundation models.
Due to an accident in June 2022, I had to pause work for 14 months and resumed in August 2023.

BSc & MSc Physics — Heidelberg, Germany
Heidelberg University, Master Grade: 1.2/4.0 (Best: 1.0) — *2015 – 2022*

Supervisors: Prof. Björn Ommer (now LMU Munich)

Broad range of courses in theoretical and experimental physics, mathematics, and specialization courses in machine learning, computer vision, and deep learning. Exchange semester at Monash University in Melbourne with courses in AI.

Research Experience

QUVA Lab ELLIS Doctoral Student — Amsterdam, Netherlands — Aug 2023 – Present

Research at the QUVA Lab, University of Amsterdam, under Cees Snoek and Yuki Asano, focusing on vision-language and generative models.

- Efficient post-pretraining pruning for Vision Transformers, enabling elastic inference across compute budgets without retraining or labels; published at **NeurIPS 2025**.
- Introduced TVBench, a benchmark for temporal video-language understanding with 10k+ clips and revealing shortcomings of existing benchmarks; accepted at **BMVC 2025**.
- Advanced masked video modeling with Sinkhorn-Knopp regularization, yielding state-of-the-art semantic and temporal representations; published at **ECCV 2024**.
- Designed PIN adapter, equipping VLMs (Flamingo, BLIP-2) with object localization using only synthetic data; published at **CVPR 2024**.

Amazon AWS Research Intern — Seattle, USA (remote) — Jul 2021 – Dec 2021

Worked in the AWS Kognition Lab with Davide Modolo and Josephe Tighe on self-supervised video representation learning. I proposed a novel contrastive learning objective that improved temporal modeling in video representations, outperforming recent approaches; published at **CVPR Workshop 2022**.

Ommer Lab Student Researcher — Heidelberg, Germany — Jan 2020 – Feb 2022

Master's research under Andreas Blattmann and Björn Ommer, focusing on generative video modeling.

- Proposed an invertible neural network framework for image-to-video synthesis, introducing a probabilistic residual representation to bridge domain gaps; published at **CVPR 2021**.
- Developed a conditional generative model for controllable human behavior synthesis, enabling posture transfer and behavior editing; published at **CVPR 2021**.
- Contributed to methods for learning object dynamics from unlabeled video data, enabling interactive image-to-video synthesis; co-author on publications at **CVPR 2021** and **ICCV 2021**.

Vision Lab Student Researcher — Toronto, Canada — Sep 2019 – Dec 2019

Explored generative models (e.g., GANs) for video synthesis under the supervision of Kosta Derpanis, gaining experience in deep generative video modeling that informed later research at the Ommer Lab.

Ommer Lab Student Researcher — Heidelberg, Germany — Jan 2018 – Sep 2019

Bachelor's thesis research under Biagio Brattoli and Björn Ommer, using generative modeling to magnify and visualize posture discrepancies in human motion videos of the same action; published at **CVPR 2020** and contributed to an article in **Nature Machine Intelligence**.

Scholarships & Community Service

Workshop Organizer 'Self Supervised Learning: What is Next?' at ECCV 2024.

Invited Talks at the Surf research bootcamp, TNO in the Hague, and the National Institute for Informatics in Tokyo (all 2024).

ELLIS Inclusive AI Mentor where I mentor Matteo Nulli, an AI Master's student since 2023.

Master Thesis Supervisor of Nimrod Barazani, Luc Vermeer (both 2024), Jakub Tomaszewski, and Max Belitsky (both 2025).

Teaching Assistant UvA Foundation Models Course and Oxford MLx Fundamentals (both 2024).

Reviewer for ECCV'22, ECCV'24, ICCV'21, BMVC'21, ICLR'22, TPAMI.

Scholarship of German Academic Exchange Association (DAAD) for my research internship at the Ryerson Vision Lab in 2019.

Selected Publications

[1] Walter Simoncini*, **Michael Dorkenwald***, Tijmen Blankevoort, Cees Snoek, Yuki Asano
"Elastic ViTs from Pretrained Models without Retraining".
Accepted at **Neurips 2025**. Paper will be released soon!

[2] Daniel Cores*, **Michael Dorkenwald***, Manuel Mucientes, Cees G. M. Snoek, Yuki M. Asano
"Lost in Time: A New Temporal Benchmark for VideoLLMs".
Accepted at **BMVC 2025**.

[3] Mohammadreza Salehi*, **Michael Dorkenwald***, Fida Thoker*, Efstratios Gavves, Cees Snoek, Yuki Asano
"SIGMA: Sinkhorn-Guided Masked Video Modeling".
Published at **ECCV 2024**.

[4] **Michael Dorkenwald**, Nimrod Barazani, Cees Snoek, Yuki Asano
"PIN: Positional Insert Unlocks Object Localisation Abilities in VLMs".
Published at **CVPR 2024**.

[5] **Michael Dorkenwald**, Timo Milbich, Andreas Blattmann, Robin Rombach, Kosta Derpanis, Björn Ommer
"Stochastic Image-to-Video Synthesis using cINNs"
Published at **CVPR 2021**.

[6] Andreas Blattmann*, Timo Milbich*, **Michael Dorkenwald***, Björn Ommer
"Behavior-Driven Synthesis of Human Dynamics".
Published at **CVPR 2021**.

Selected Published Coding Projects

Software Proficiency: Python + Science Packages, PyTorch, TensorFlow, Git, LaTeX

[1] "Image-to-Video Synthesis using cINNs". Link [*Python, PyTorch*]
Official implementation of the **CVPR 2021** paper on stochastic image-to-video synthesis with conditional invertible neural networks with **185+ GitHub stars**.

[2] "TVBench: A Temporal Video-Language Benchmark". Dataset Link [*Hugging Face*]
Benchmark dataset for temporal video-language understanding, introduced at **BMVC 2025**. Downloaded average **800 times/month** on Hugging Face.

[3] "SIGMA: Sinkhorn-Guided Masked Video Modeling". Code Link — Checkpoint Link [*Python, PyTorch*]
Official implementation of the **ECCV 2024** paper on Sinkhorn-guided masked video modeling, improving semantic and temporal representation learning.