Michael Dorkenwald

m.dorkenwald@gmail.com
Google Scholar
mdorkenwald.com

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 and video learning, vision-language models, and efficient foundation models.
Temporary research leave (Jun 2022–Aug 2023); full-time research resumed Aug 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 foundation models.

- Developed a post-pretraining pruning framework for Vision Transformers enabling elastic inference across compute budgets without retraining or labels; accepted at **NeurIPS 2025**.
- Introduced TVBench, a large-scale benchmark for temporal video–language understanding (10K+ clips), revealing temporal reasoning gaps in current models; accepted at **BMVC 2025**.
- Proposed SIGMA, a Sinkhorn-guided masked video modeling approach achieving state-of-the-art semantic and temporal representations; published at **ECCV 2024**.
- Designed PIN, an adapter that equips vision–language models (VLMs) such as Flamingo and BLIP-2 with object localization capabilities using only synthetic data; published at **CVPR 2024**.

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

Researched large-scale self-supervised video representation learning at AWS AI Labs (Kognition) with Davide Modolo and Joseph Tighe, leveraging contrastive and transformer-based architectures.

- Designed a shuffled contrastive learning framework (SCVRL) that jointly captures appearance and motion, improving temporal modeling in large-scale video datasets; 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 a conditional invertible neural network (cINN) framework for stochastic image-to-video synthesis, achieving fine-grained temporal diversity; published at **CVPR 2021**.
- Developed a behavior-driven video synthesis model for controllable human motion generation; published at **CVPR 2021**.
- Co-authored methods for learning object dynamics from unlabeled video, enabling interactive image-to-video synthesis; published at **CVPR 2021** and **ICCV 2021**.

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

Researched GAN-based generative models for video synthesis under Kosta Derpanis, laying foundations for later work in video generation.

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 — SURF Research Bootcamp, TNO (The Hague), and NII Tokyo (2024).

ELLIS Inclusive AI Mentor — mentoring an AI Master's student (since 2023).

Master's Thesis Supervision — N. Barazani, L. Vermeer (2024); J. Tomaszewski, M. Belitsky (2025).

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

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

DAAD Scholarship — funded research internship at Ryerson Vision Lab (2019).

Selected Publications

[1] Walter Simoncini*, **Michael Dorkenwald***, Tijmen Blankevoort, Cees Snoek, Yuki Asano
"Elastic ViTs from Pretrained Models without Retraining".
To appear at **NeurIPS 2025**.

[2] Daniel Cores*, **Michael Dorkenwald***, Manuel Mucientes, Cees G. M. Snoek, Yuki M. Asano
"Lost in Time: A New Temporal Benchmark for VideoLLMs".
To appear 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**, Fanyi Xiao, Biagio Brattoli, Joseph Tighe, Davide Modolo.
"SCVRL: Shuffled Contrastive Video Representation Learning".
Published at **CVPR Workshop 2022**.

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

[7] **Michael Dorkenwald***, Uta Büchler*, Björn Ommer.
"Unsupervised Magnification of Posture Deviations Across Subjects".
Published at **CVPR 2020**.

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 roughly **6k times** since one year on Hugging Face.