

Marcel Simon

ML RESEARCH SCIENTIST · COMPUTER VISION · SELF-SUPERVISED LEARNING

🏠 <https://www.marcelSimon.com> 🌐 MarcelSimon 📄 Marcel Simon 📍 Marcel Simon ✉ Contact via homepage

Summary

Research scientist specialising in self-supervised learning, transformer architectures, and production-scale computer vision systems. PhD Magna Cum Laude (FSU Jena; visiting scholar at UC Berkeley BAIR with Trevor Darrell) → Microsoft Research (millions of daily Bing inferences) → Nota AI (Diffusion Transformer world models, AAAI 2026 oral); 1,600+ citations, h-index 14.

Experience

Machine Learning Research Engineer

Berlin, Germany

NOTA AI GMBH

Feb 2025 – Present

- **Architected a Diffusion Transformer (DiT) world model** for self-supervised learning: the encoder is trained to predict the feature representation of the next video frame, injecting physical and temporal priors without optical flow or labels.
- Designed the full training pipeline end-to-end — architecture, self-supervised objective, evaluation, and analysis — working autonomously in a team of two.
- Achieved **state-of-the-art for DINO-style video distillation** on downstream detection and segmentation benchmarks (ADE20K mIoU: 35.0 → 36.4).
- **Published two peer-reviewed papers in 10 months** (ICML Workshop 2025; AAAI Workshop 2026, oral).
- Mentored an intern on transformer-based SSL methods.

Research Intern → Remote Research Contractor

Redmond, WA / Remote

MICROSOFT RESEARCH (AI FOR EARTH)

Jun 2018 – Oct 2019

- Trained and deployed an animal species recognition model for **Bing Visual Search**, reaching **millions of daily inferences** shortly after production launch.
- Significant contributor to **Microsoft/CameraTraps** (950+ GitHub stars, 250+ forks) — an open-source PyTorch platform for AI-powered wildlife conservation.
- Contributor to **Microsoft/SpeciesClassification** (100+ stars) and **Microsoft/arcticseals**: aerial detection pipeline for Arctic seal surveys (NOAA).

Selected Publications

Next-Frame Prediction as a Reliability-Aware Training Paradigm for Robust Vision Encoders

Oral

AAAI WORKSHOP

2026

Video Self-Distillation for Single-Image Encoders: A Step Toward Physically Plausible Perception

ICML WORKSHOP

2025

The Whole is More Than Its Parts? From Explicit to Implicit Pose Normalization

IEEE TPAMI (IF 24.3, #1 AI JOURNAL)

2020

Generalized Orderless Pooling Performs Implicit Salient Matching

ICCV (A*)

2017

Neural Activation Constellations: Unsupervised Part Model Discovery with Convolutional Networks

~500 citations

ICCV (A*)

2015

Total citations: 1,600+ · h-index: 14 · Full list: scholar.google.de

Education

PhD in Computer Vision — Magna Cum Laude

Jena, Germany

FRIEDRICH-SCHILLER-UNIVERSITY JENA

2014 – 2018

- Supervised by Prof. Joachim Denzler (H-index >50). Research: unsupervised part model discovery and fine-grained visual classification using CNN attention mechanisms.
- **Visiting Research Scholar at UC Berkeley (BAIR/ICSI)**: collaborated with Prof. Trevor Darrell (H-index >180). Invented Alpha-Pooling — published at ICCV 2017.

M.Sc. in Computer Vision

Jena, Germany · Exchange: Univ. of Mississippi, USA

FRIEDRICH-SCHILLER-UNIVERSITY JENA

2011 – 2014

Technical Arsenal

Programming

Python, PyTorch, C++, ONNX, MATLAB

Infrastructure

Docker, SLURM, Azure, HuggingFace Transformers, Firebase

ML / Research

Self-Supervised Learning, Vision Transformers (ViT/DiT), Diffusion Models, Distributed Training, MLOps

Open Source

Microsoft/CameraTraps (950+ ★), cvjena/cnn-models (360+ ★)

Spoken Languages

German (native), English (fluent)