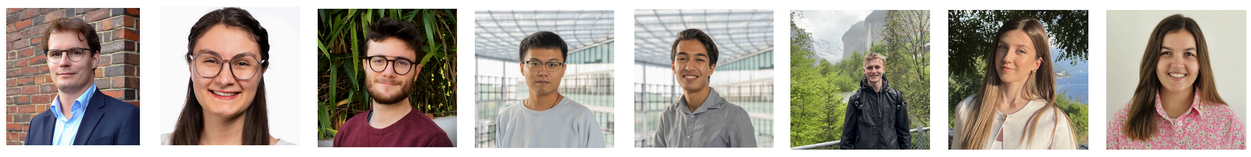
Deep learning for measuring behavior

Mathis Group for Computational Neuroscience and AI

We develop **computer vision** and **machine learning** tools for the analysis and quantification of **behavior** including **pose estimation**, **animal re-identification**, **action segmentation**...

Join us in Geneva!





Lab Alumni:

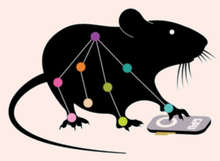

Graduated in 2025 and went to the Zoo in Zurich


Graduated in 2025, now at Cornell Weill for a postdoc


Now a software engineer at Jua.ai

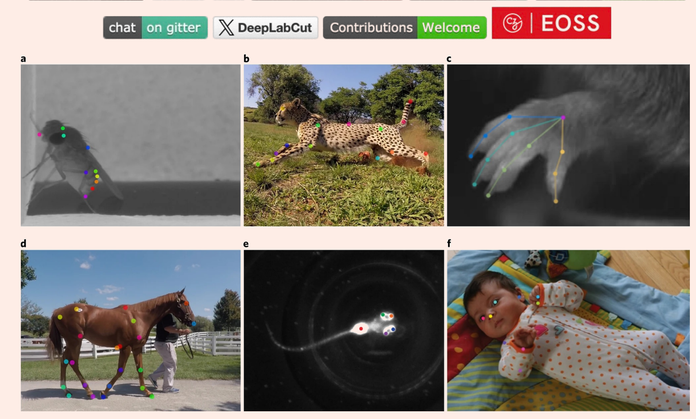

Now a data science engineer at Groupe Mutuel

Pose estimation



DeepLabCut™
a software package for animal pose estimation

DLC: Efficient deep learning for single and multi-animal pose tracking and identification

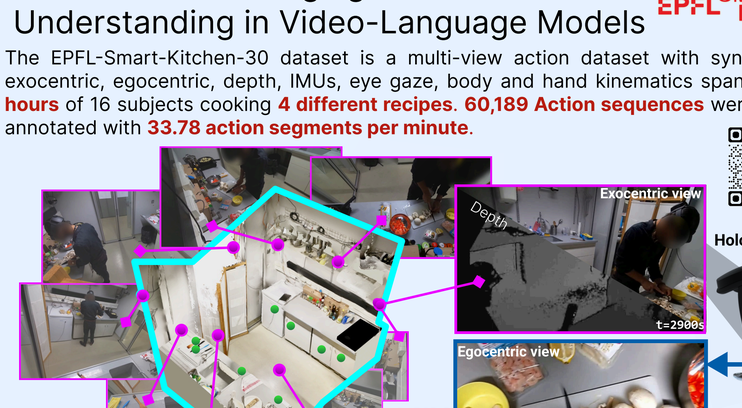


Mathis, A., et al. "*DeepLabCut: markerless pose estimation of user-defined body parts with deep learning.*" Nature neuroscience 21.9 (2018): 1281-1289.

Nath, T., et al. "*Using DeepLabCut for 3D markerless pose estimation across species and behaviors.*" Nature protocols 14.7 (2019): 2152-2176.

Lauer, J., et al. "*Multi-animal pose estimation, identification and tracking with DeepLabCut.*" Nature Methods 19.4 (2022): 496-504.

BUCTD: Overcoming the detection information bottleneck and ambiguity in pose estimation

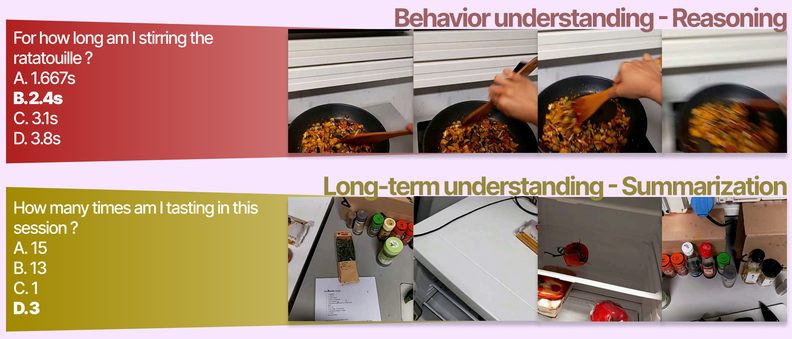


Zhou, M*., Stoffl, L*., Mathis, MW., Mathis, A., "*Rethinking pose estimation in crowds: overcoming the detection information bottleneck and ambiguity*", ICCV 2023

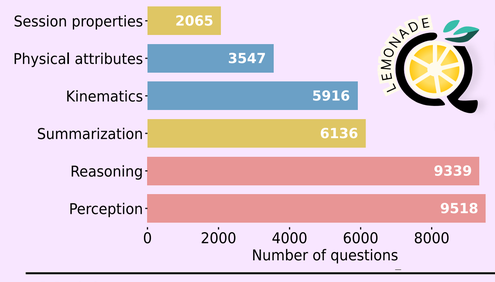
HOISDF: Hand object pose estimaition based on Signed Distance Field

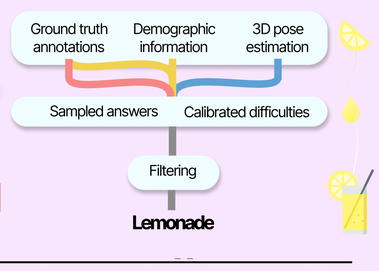
Qi, H., Zhao, C., Salzmann, M., Mathis, A., "*HOISDF: Constraining 3D Hand-Object Pose Estimation with Global Signed Distance Fields*", CVPR 2024

AI4Science with language models



Lemonade (**L**anguage models **E**valuation of **MO**tion a**N**d **A**ction-**D**riven **E**nquiries) is a framework designed to generate millions of unique QA pairs by combining various video clips, question formats, and answer types. Lemonade consists of **36,521 closed-ended QA pairs** linked to egocentric video clips, categorized into three groups and six subcategories.

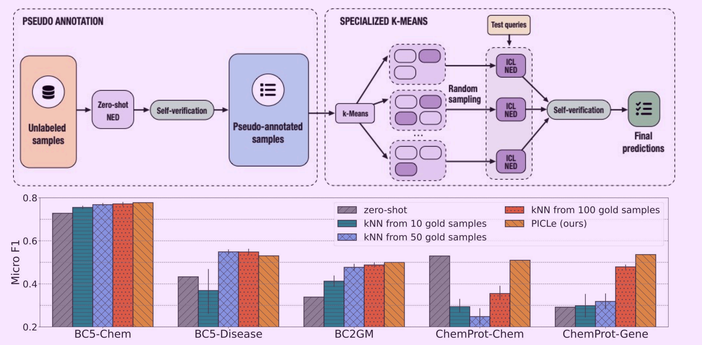




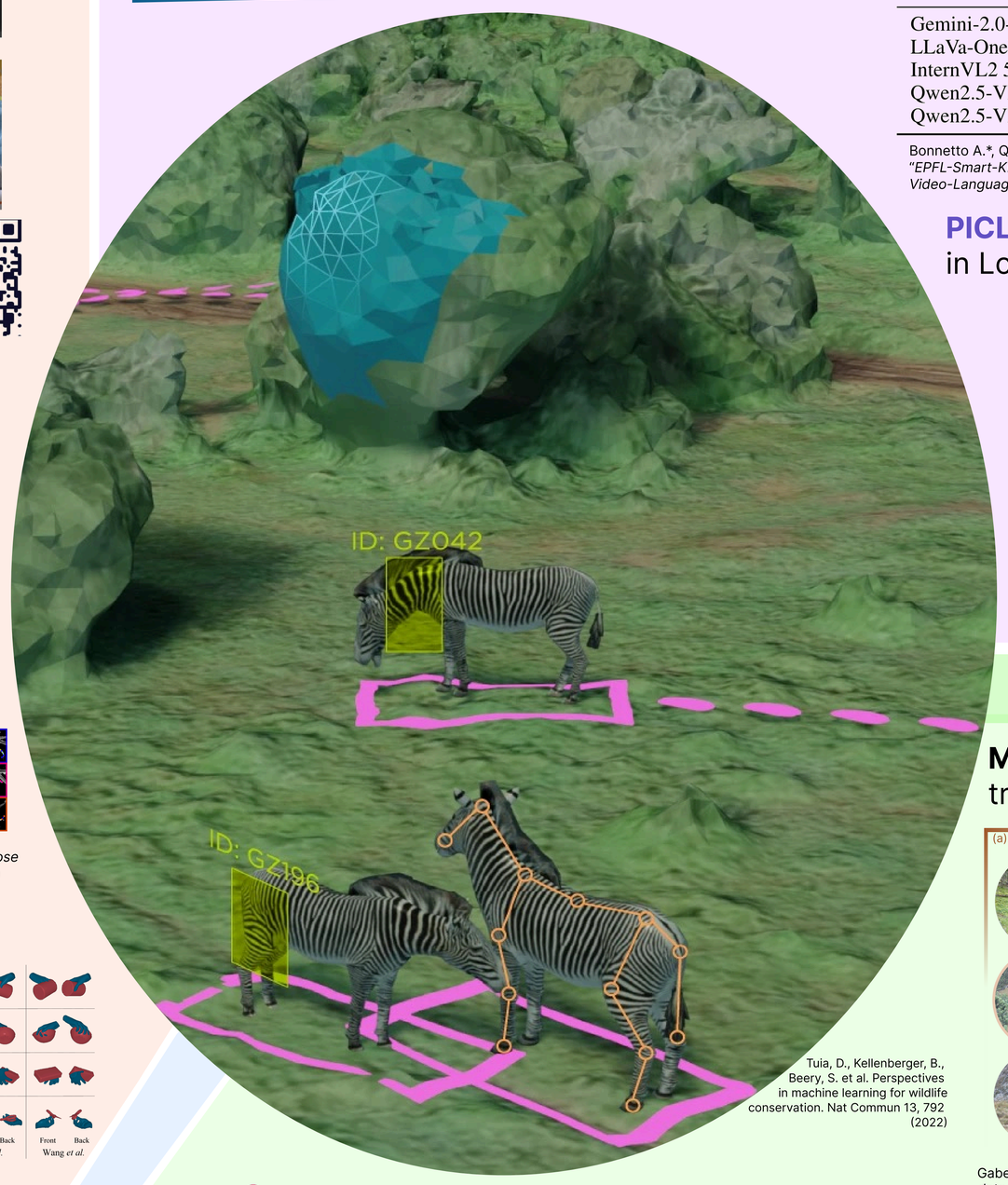
Models	Perc.	Reas.	Sess.	Summ.	Phys.	Kin.	**All**
Human*	62.32	52.38	62.38	44.40	70.78	38.34	54.75
Gemini-2.0-Flash	41.47	41.03	**49.10**	35.09	**40.26**	21.69	37.39
LLaVa-OneVision	**47.23**	41.46	36.76	41.87	34.03	**34.09**	**40.85**
InternVL2 5-8B	43.94	41.52	45.76	33.80	39.89	27.74	38.70
Qwen2.5-VL-32B (4fr.)	43.72	**45.87**	46.25	**42.75**	32.22	28.52	40.67
Qwen2.5-VL-7B	42.86	43.58	48.14	37.66	38.93	25.64	39.30

Bonnetto A.*, Qi H.*, Leong F., Tashkovska M., Rad M., Shokur S., Hummel F., Micera S., Pollefeys M., Mathis A., "*EPFL-Smart-Kitchen: An Ego-Exo Multi-Modal Dataset for Challenging Action and Motion Understanding in Video-Language Models*", NeurIPS 2025

PICLe: Pseudo-annotations for In-Context Learning in Low-Resource Named Entity Detection



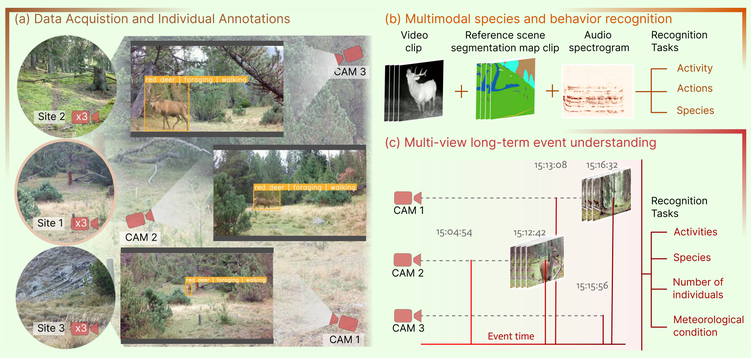
Mamooler S., Montariol S., Mathis A., Bosselut A. *"PICLe: Pseudo-annotations for In-Context Learning in Low-Resource Named Entity Detection"*. NAACL 2025



Tuia, D., Kellenberger, B., Beery, S. et al. Perspectives in machine learning for wildlife conservation. Nat Commun 13, 792 (2022)

Action understanding

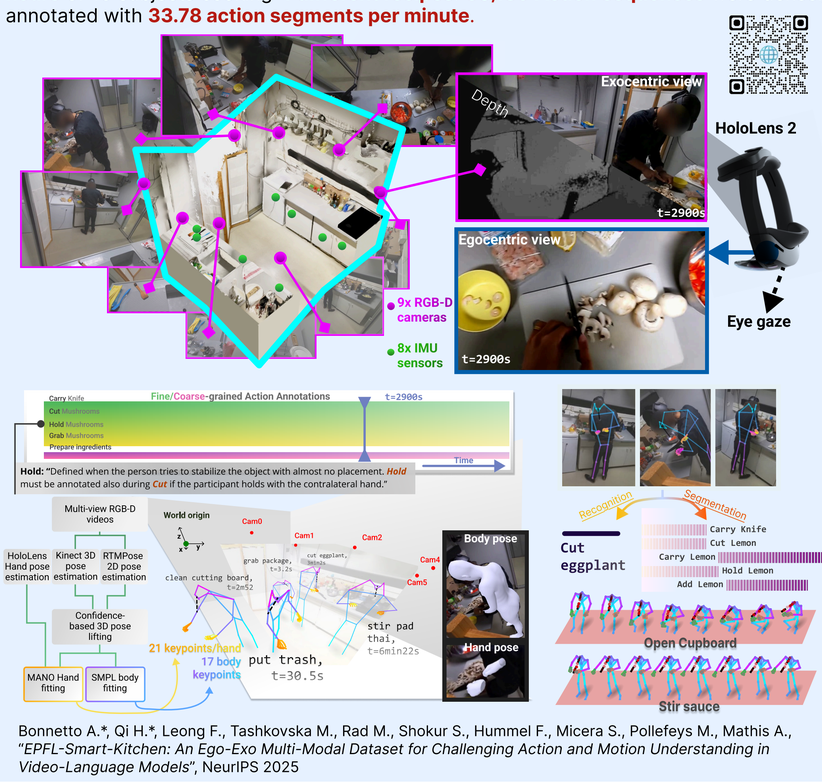
MammAlps: A multimodal and multiview camera trap dataset of wildlife behavior in the Swiss Alps



Gabeff V., Qi H., Flaherty B., Sumbül G., Mathis A., Tuia D. "*MammAlps: A multi-view video behavior monitoring dataset of wild mammals in the Swiss Alps*", Presented as a Highlight at CVPR 2025

EPFL Smart Kitchen: An Ego-Exo Multi-Modal Dataset for Challenging Action and Motion Understanding in Video-Language Models

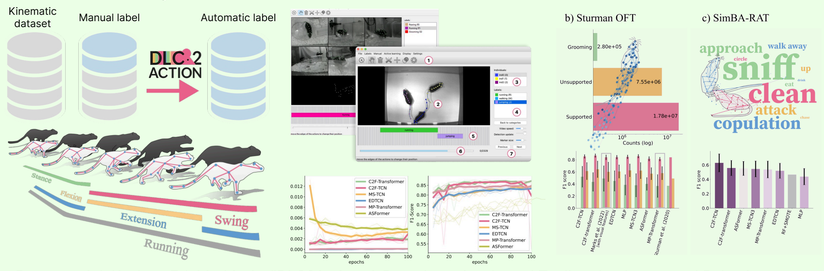
The EPFL-Smart-Kitchen-30 dataset is a multi-view action dataset with synchronized exocentric, egocentric, depth, IMUs, eye gaze, body and hand kinematics spanning **29.7 hours** of 16 subjects cooking **4 different recipes**. **60,189 Action sequences** were densely annotated with **33.78 action segments per minute**.



Bonnetto A.*, Qi H.*, Leong F., Tashkovska M., Rad M., Shokur S., Hummel F., Micera S., Pollefeys M., Mathis A., "*EPFL-Smart-Kitchen: An Ego-Exo Multi-Modal Dataset for Challenging Action and Motion Understanding in Video-Language Models*", NeurIPS 2025

DLC2ACTION What is my cat doing?

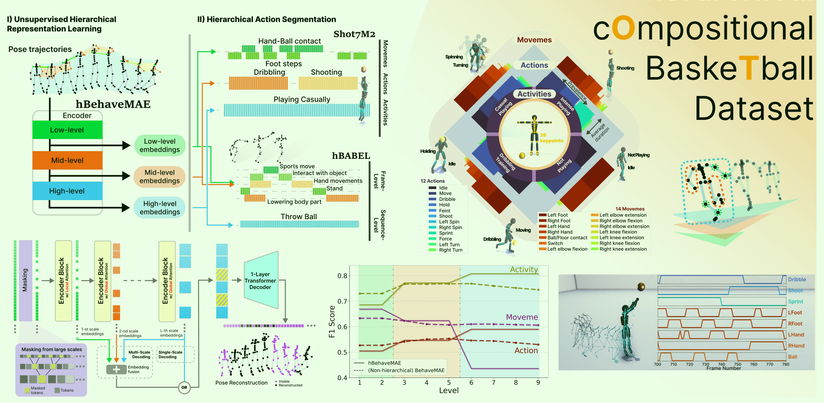
*DLC2Action is an easy-to-use python toolbox to perform **action segmentation** from body kinematics. Supporting diverse pose data formats, the toolbox allows you to **extract kinematic features** and tidily run trainings, evaluations and predictions from a set of supervised **Deep Learning models**. The toolbox is associated with a User Interface to manually annotate actions of interest from videos.*



Bonnetto A.*, Kozlova E.*, Mathis A., "*DLC2Action: A Deep Learning-based Toolbox for Automated Behavior Segmentation*", bioRxiv 2025

hBehaveMAE: Learning the hierarchical structure of behavior

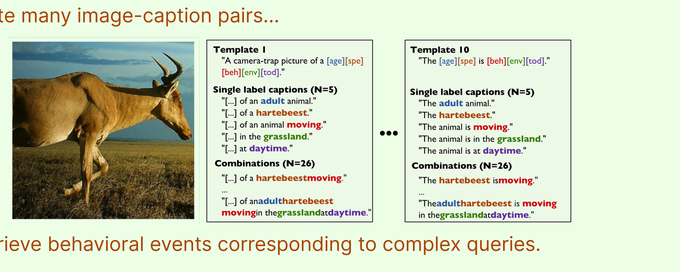
Shot7M2: Synthetic Hierarchical cOmpositional BaskeTball Dataset



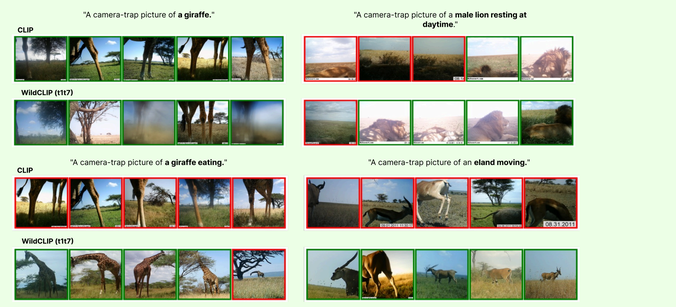
Stoffl L., Bonnetto A., d'Ascoli S., Mathis A., "*Elucidating the Hierarchical Nature of Behavior with Masked Autoencoders*", ECCV 2024

WildCLIP: Retrieving events of interest from camera trap images with domain-adapted vision-language models

We create many image-caption pairs...



... to retrieve behavioral events corresponding to complex queries.



Gabeff, V., Russwurm, M., Tuia, D., & Mathis, A. "*WildCLIP: Scene and animal attribute retrieval from camera trap data with domain-adapted vision-language models*" International Journal of Computer Vision (in press) 2023

Neuro X Institute
BRAIN MIND INSTITUTE

We love open source!

Check out our website!

Selected Collaborators:

Mackenzie Mathis, EPFL
Silvestro Micera, EPFL
Friedhelm Hummel, EPFL
Marc Pollefeys, ETH
Antoine Bosselut, EPFL
Devis Tuia, EPFL

Microsoft AI
Chan Zuckerberg Initiative
THE KAVLI FOUNDATION
Boehringer Ingelheim Fonds
Stiftung für medizinische Grundlagenforschung
Swiss National Science Foundation