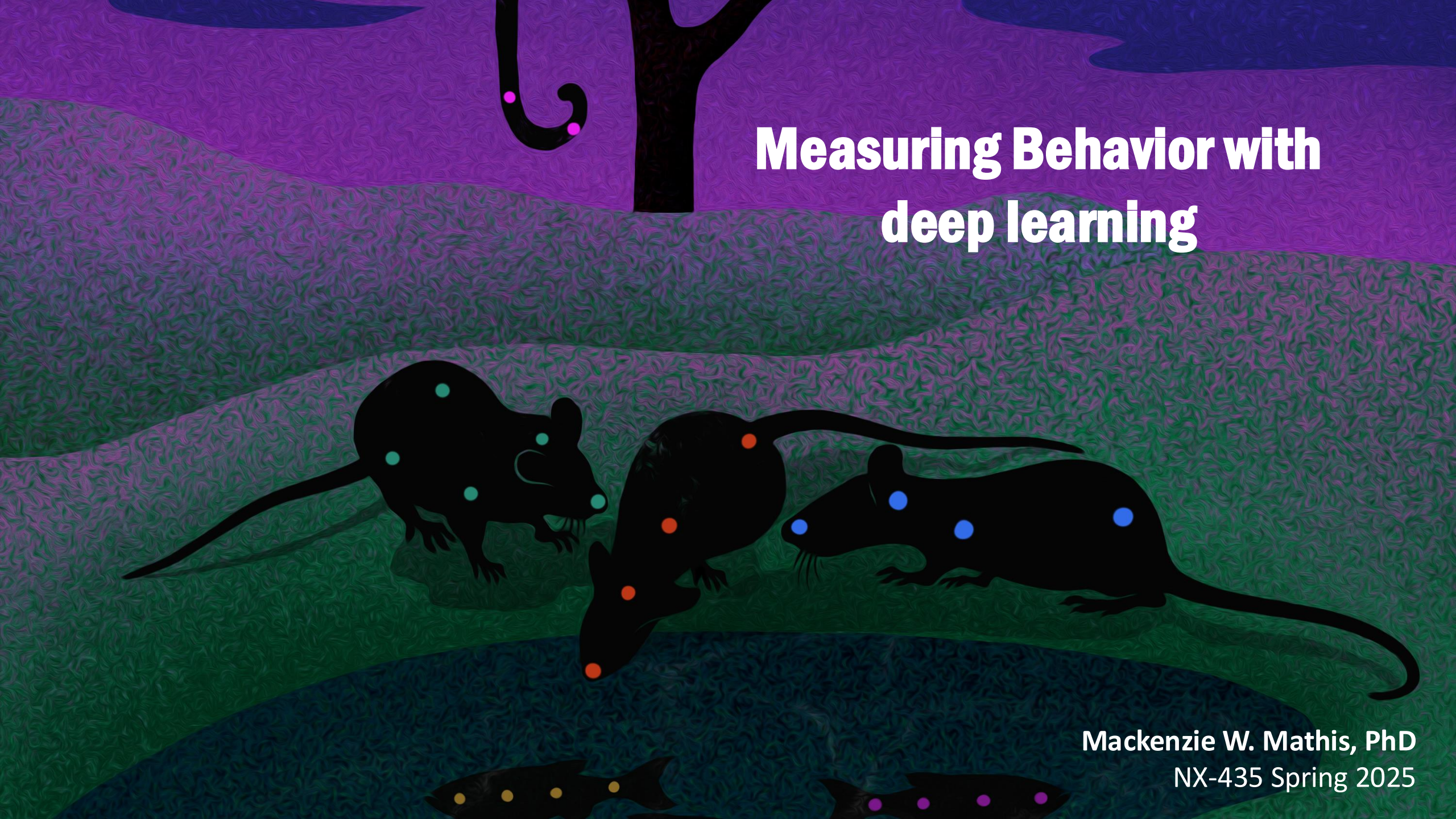
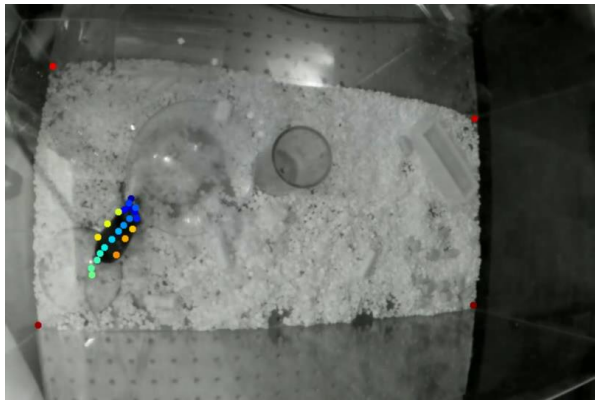


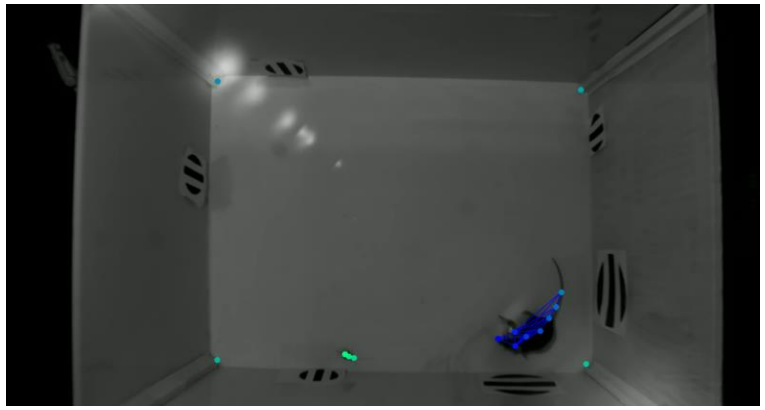
Measuring Behavior with deep learning

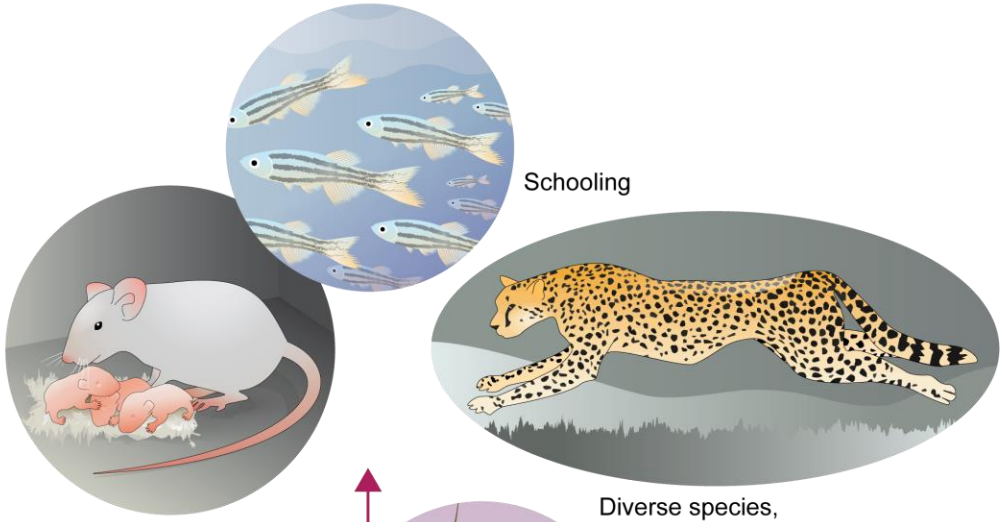
Mackenzie W. Mathis, PhD

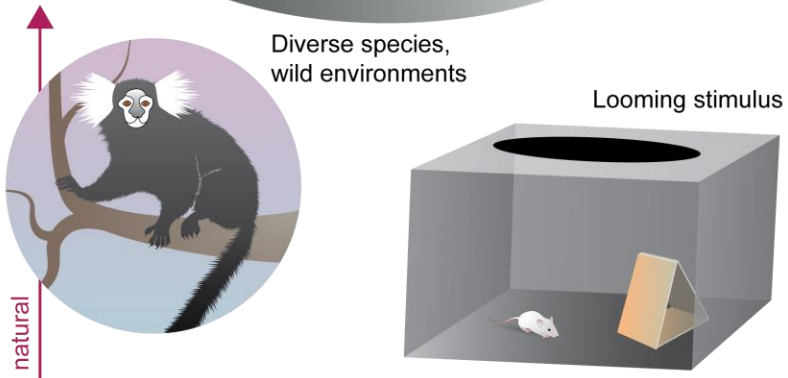
NX-435 Spring 2025

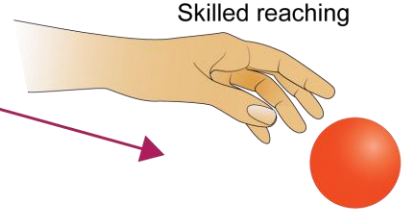
Various behavioral assays in systems neuroscience

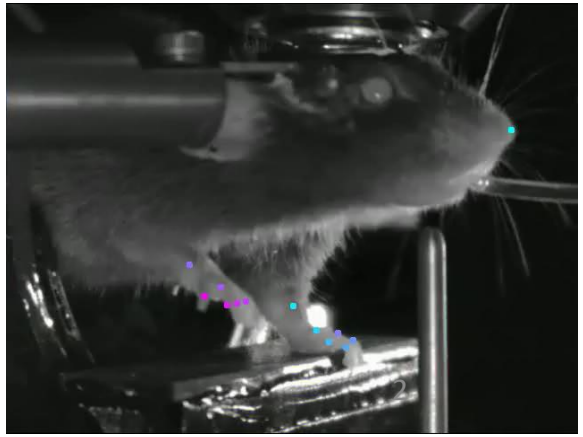




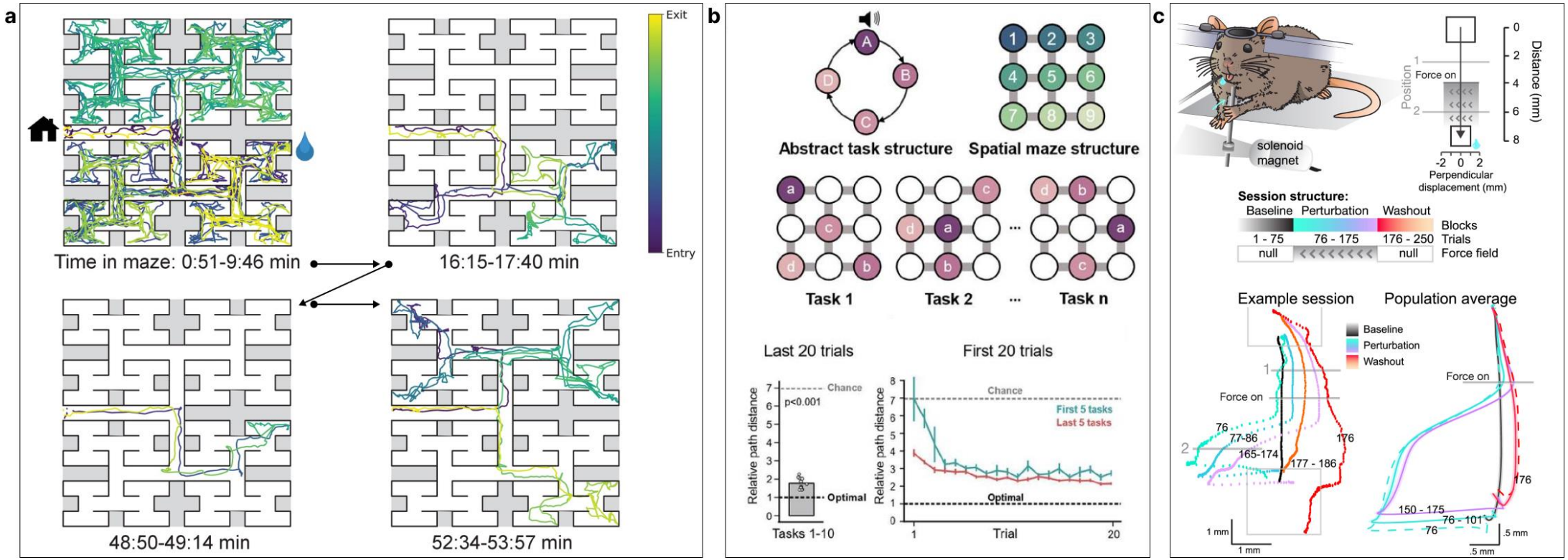








Various behavioral assays in systems neuroscience



Mathis 2025; arXiv

Rapid Learning in animals: from few-shot to updating of internal model-based learning.
(a) Adapted from Rosenberg et al. 2021
(b) Adapted from El-Gaby et al. 2024: Task design: animals learned to navigate between 4 sequential goals on a 3×3 spatial grid-maze. Reward locations changed across tasks but the abstract structure, 4 rewards arranged in an ABCD loop, remained the same.
(c) Adapted from Mathis et al. (7):

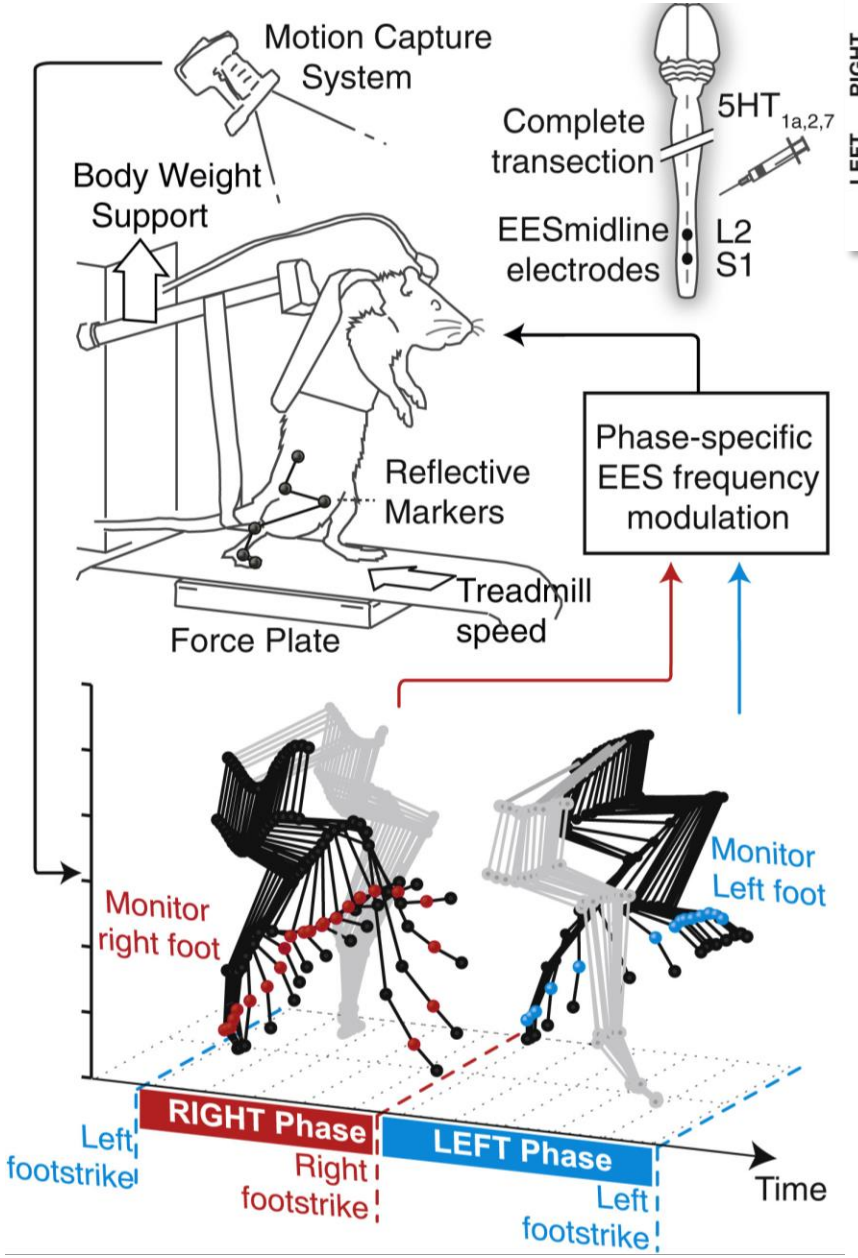
Gunnar Johansson -- Video by James Maas, (Cornell University, 1971)

https://www.youtube.com/watch?v=1F5ICP9SY

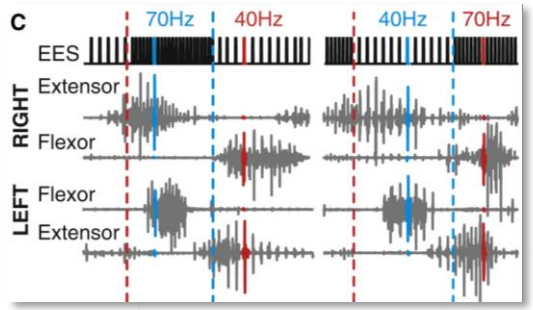
Measuring movement

E. Muybridge, 1887 (zoopraxiscope)

Ota et al. 2015 Sci Reports



Moraud et al. 2016 Neuron

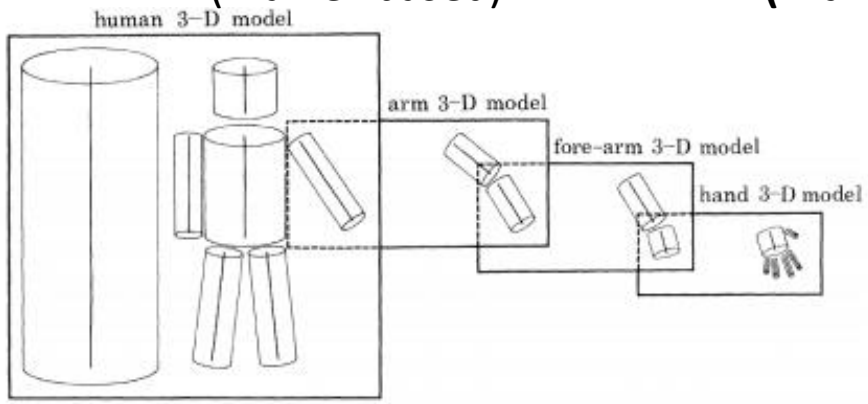


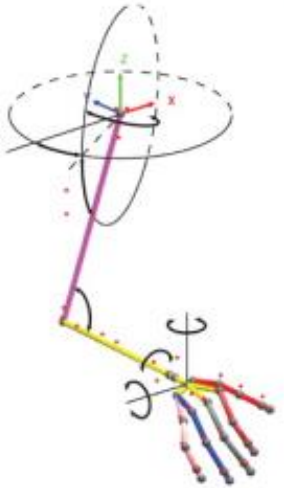
NeuroSoft, Lacour Lab EPFL

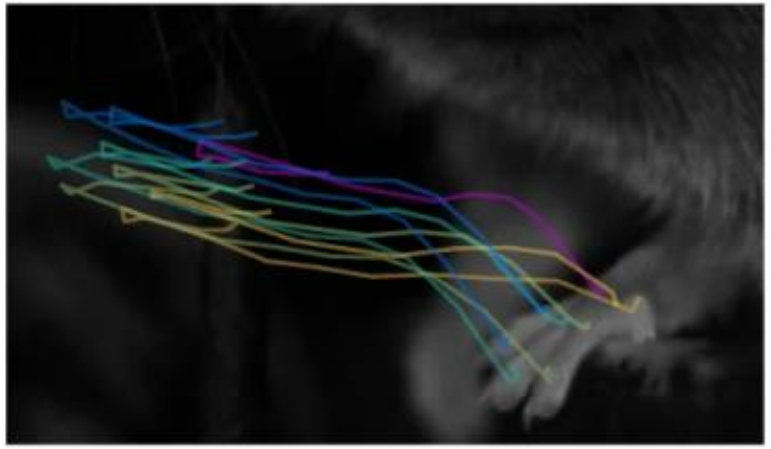
EPFL

1887 | 1978 | 2010 | **2014** | 2018

Muybridge | David Marr | Black & Donoghue (marker based) | **Deep learning to human pose (Markerless)** | Markerless pose to animals







- Hausmann, Vargas, Mathis, Mathis
  Current Opinion in Neurobiology 2021

Schematic Overview of Markerless Motion Capture, aka Pose Estimation

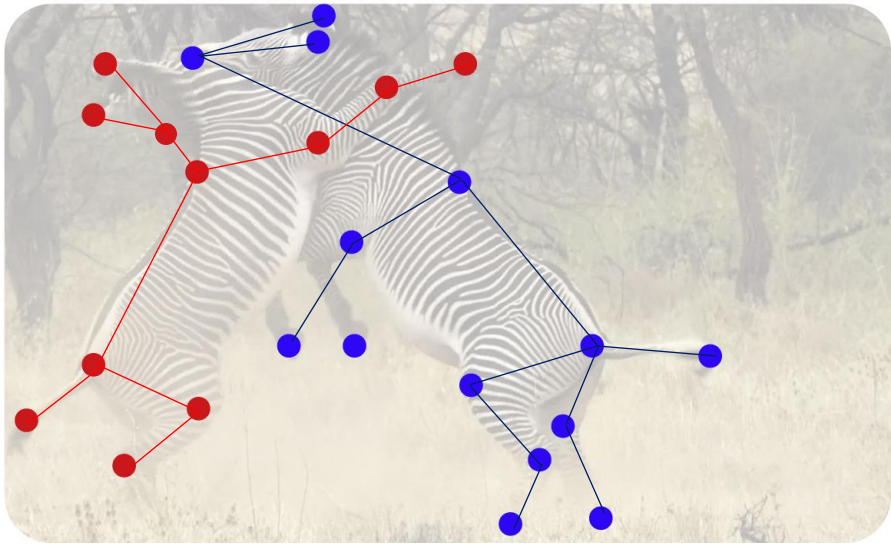
Pixel Representation

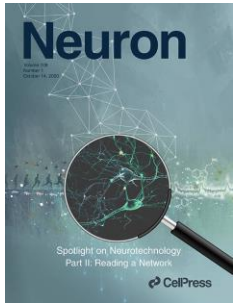
Pose Estimation Algorithm

Subject 1 Keypoints

Subject 2 Keypoints

Keypoint Representation





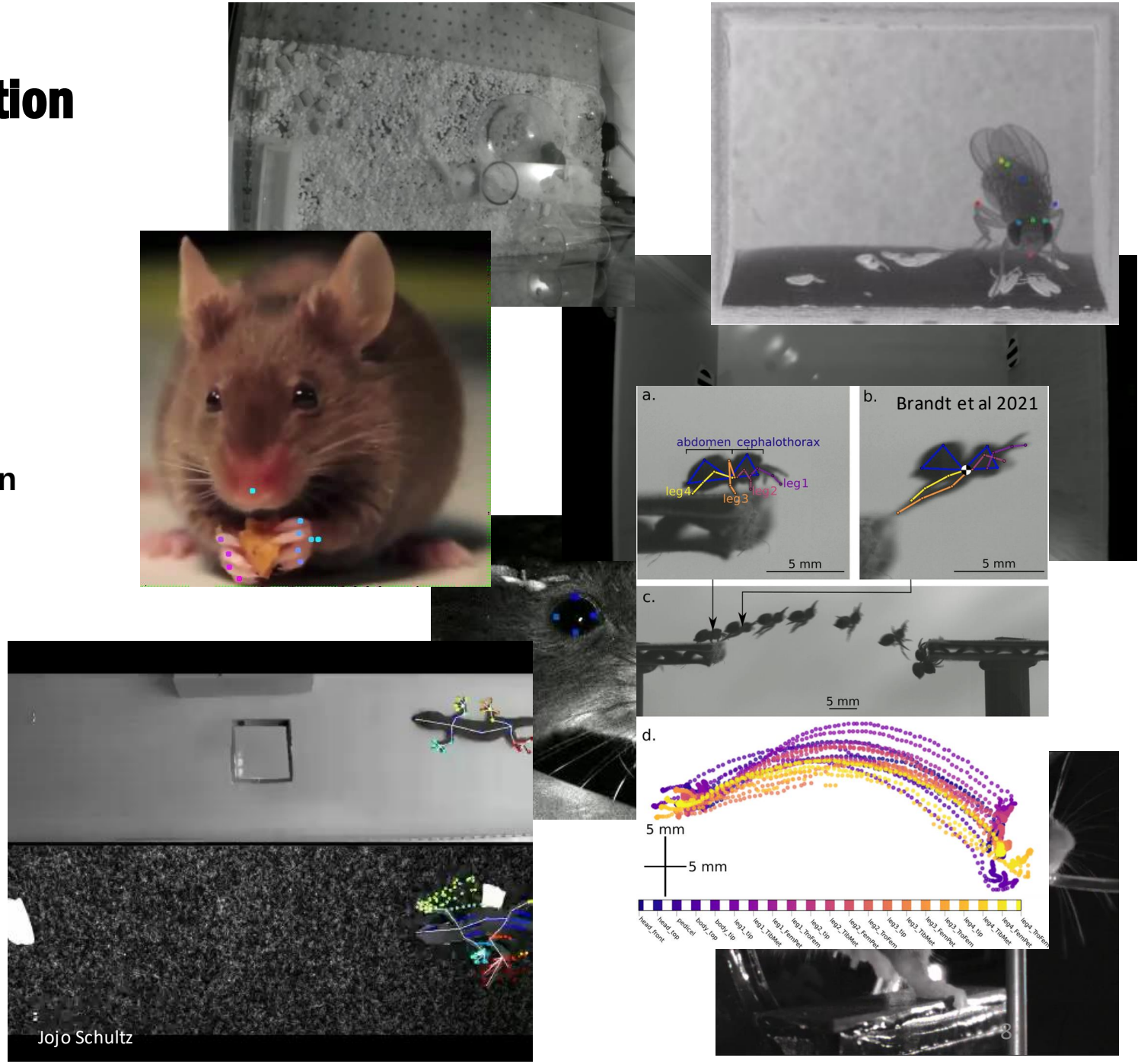
Primer

A Primer on Motion Capture with Deep Learning: Principles, Pitfalls, and Perspectives

Alexander Mathis,[1,2,3,*] Steffen Schneider,[3,4] Jessy Lauer,[1,2,3] and Mackenzie Weygandt Mathis[1,2,3,*]
[1]Center for Neuroprosthetics, Center for Intelligent Systems, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland
[2]Brain Mind Institute, School of Life Sciences, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland
[3]The Rowland Institute at Harvard, Harvard University, Cambridge, MA, USA
[4]University of Tübingen and International Max Planck Research School for Intelligent Systems, Tübingen, Germany
*Correspondence: alexander.mathis@epfl.ch (A.M.), mackenzie.mathis@epfl.ch (M.W.M.)
https://doi.org/10.1016/j.neuron.2020.09.017

Challenges for pose estimation in the laboratory

- animals have highly different bodies (i.e., can't leverage a skeleton or pose prior across all species)
- not practical for individuals to label >10,000 frames for training (i.e., human benchmark dataset sizes)
- fast real-time video analysis
- Multi-animal tracking, where animals can look truly identical
- Robust, plug-N-play solutions?



- Mathis & Mathis 2020
  Current Opinion in Neurobiology

Deep learning in the laboratory: leveraging transfer learning

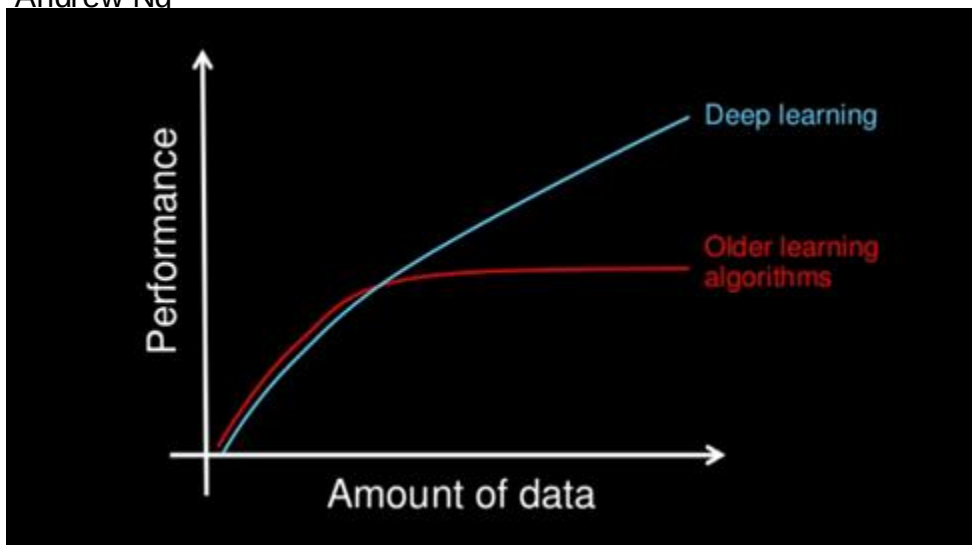
DeepPose
DeeperCut
OpenPose
Conv. PoseMachines
…
HRNet

deep neural networks

image → Predictor → pose

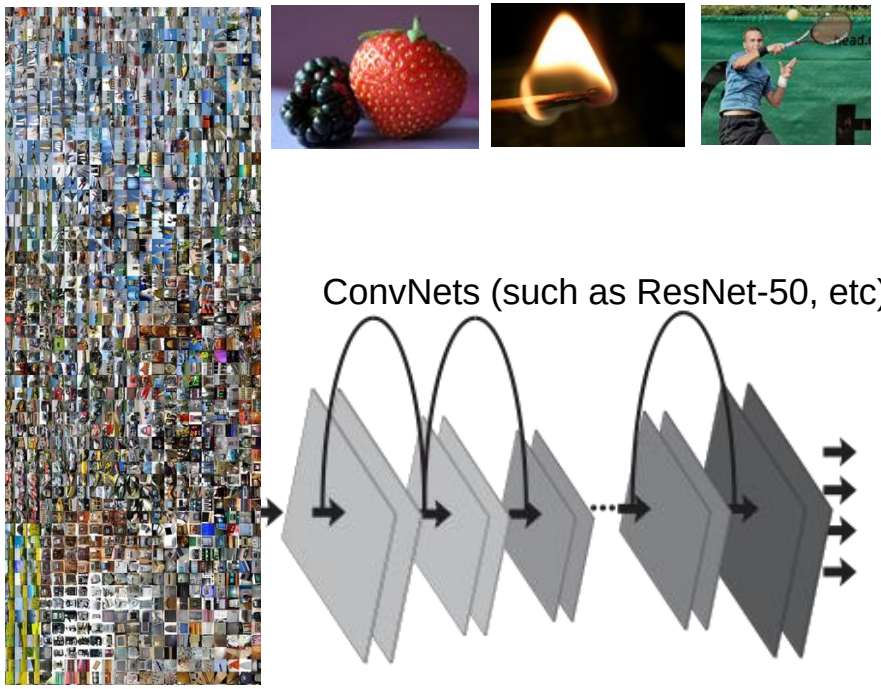
train **A lot of labeled images (>10^6 joints!)**

Andrew Ng



DATA hungry algorithms… how to bring this to the lab?

Transfer Learning: take a trained network and ask it to learn a new task



…

ConvNets (such as ResNet-50, etc)

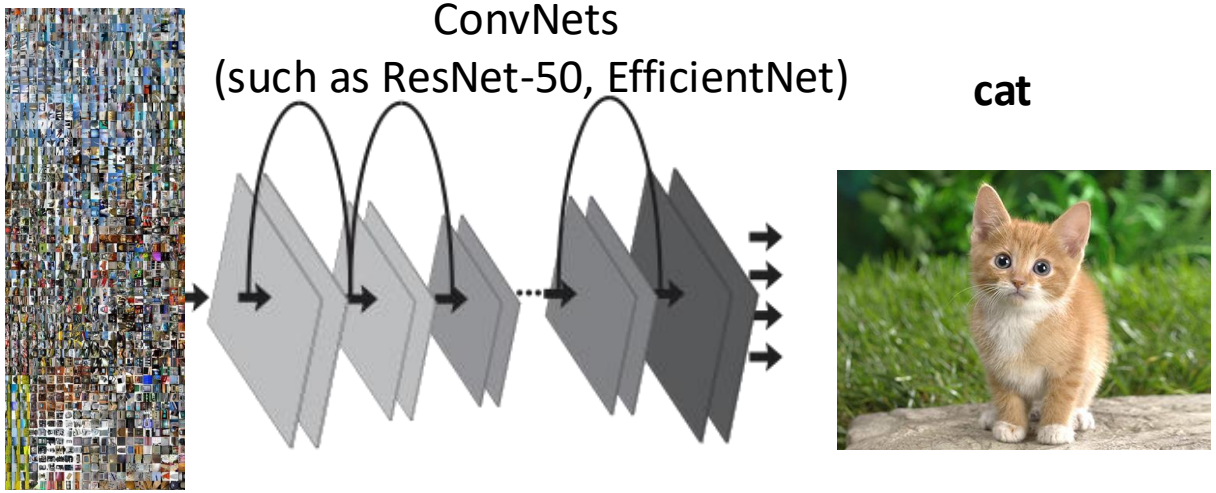
cat

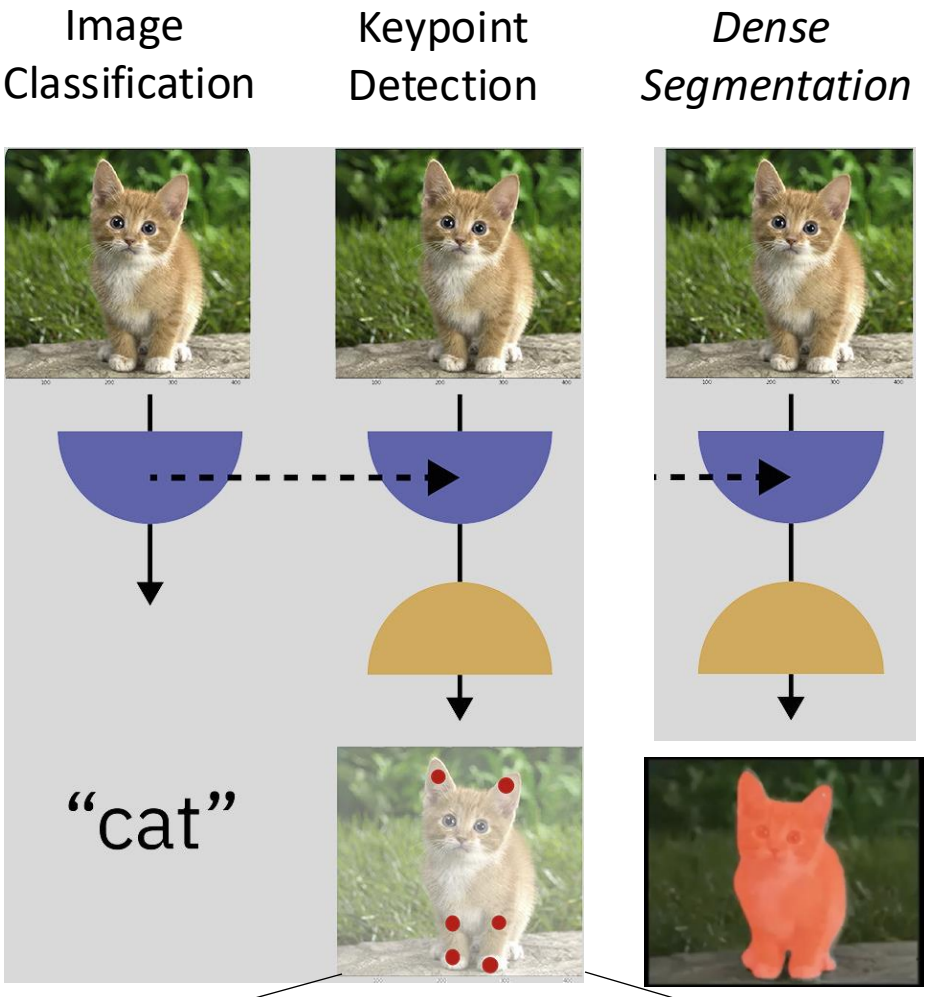
IMAGENET

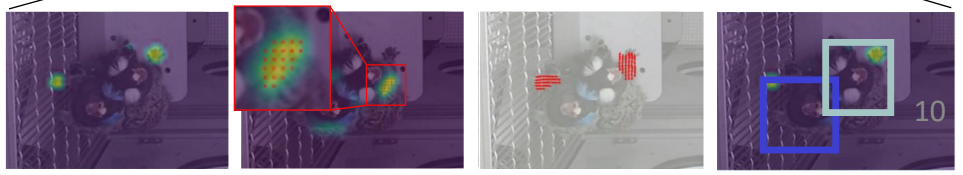
Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) **ImageNet Large Scale Visual Recognition Challenge**. *International Journal of Computer Vision*, 2015.

High performance pose estimation using transfer learning

ImageNet-based transfer learning

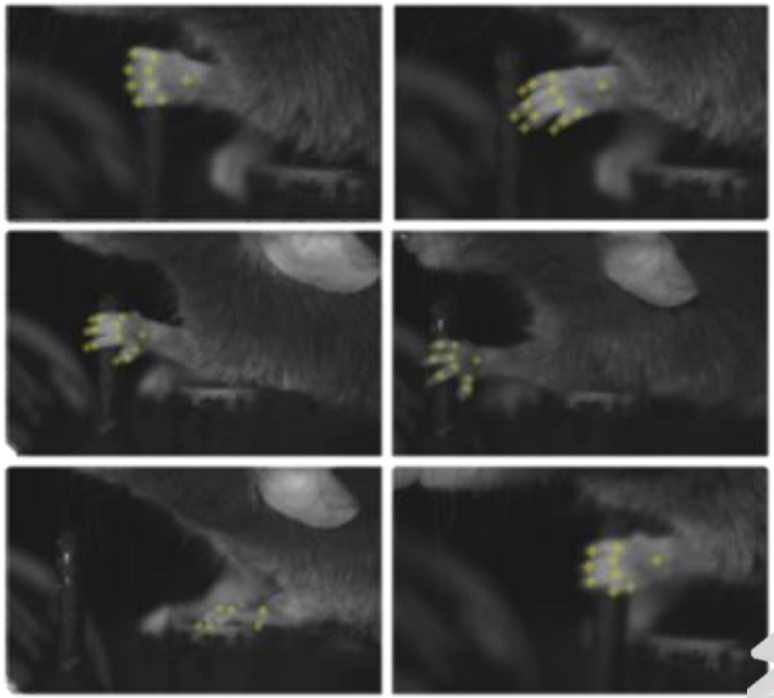


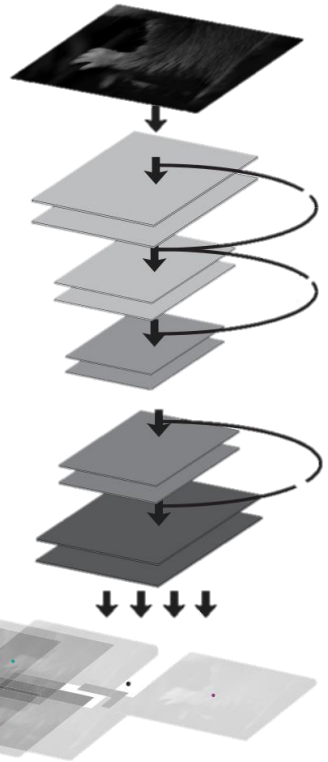


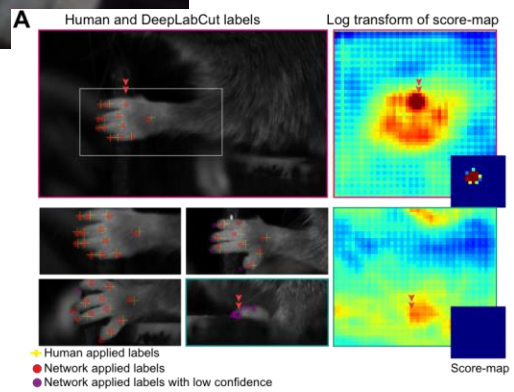


Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) **ImageNet Large Scale Visual Recognition Challenge**. *International Journal of Computer Vision*, 2015.

Measuring Movement: DeepLabCut for efficient markerless tracking of keypoints with deep learning

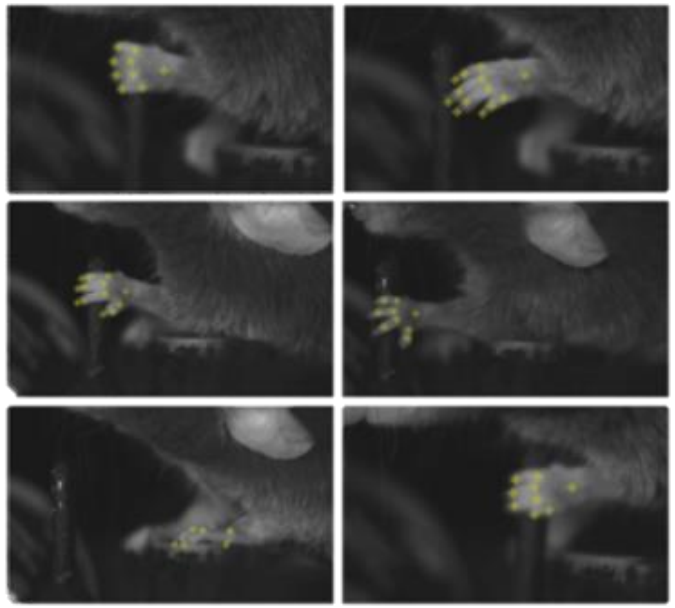


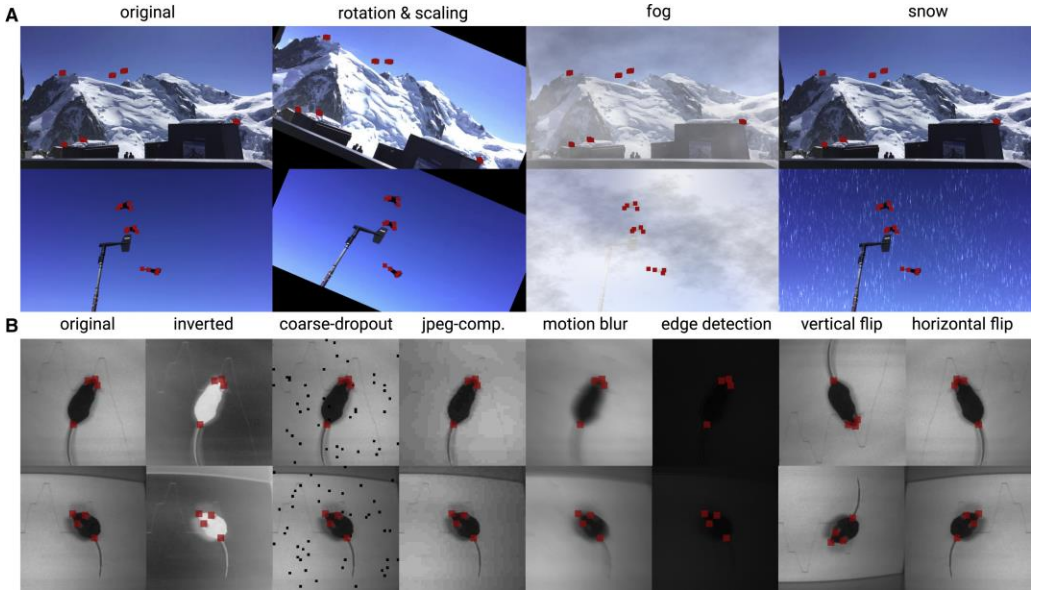


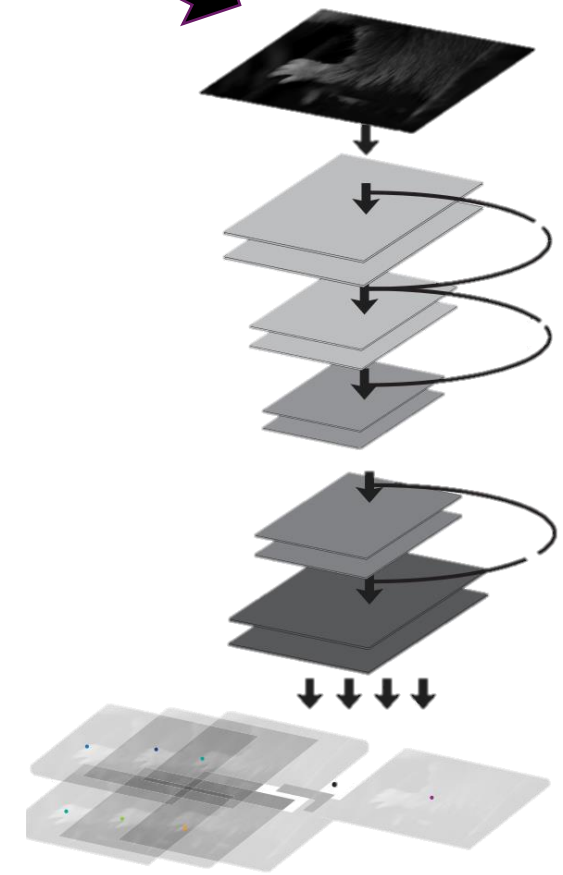


A. Mathis et al. Nature Neurosci, 2018

Data augmentation during training





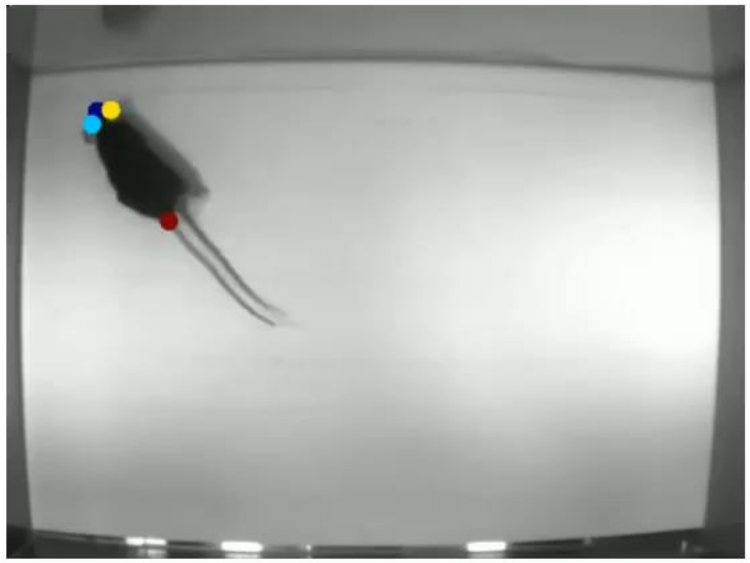


Key Features:

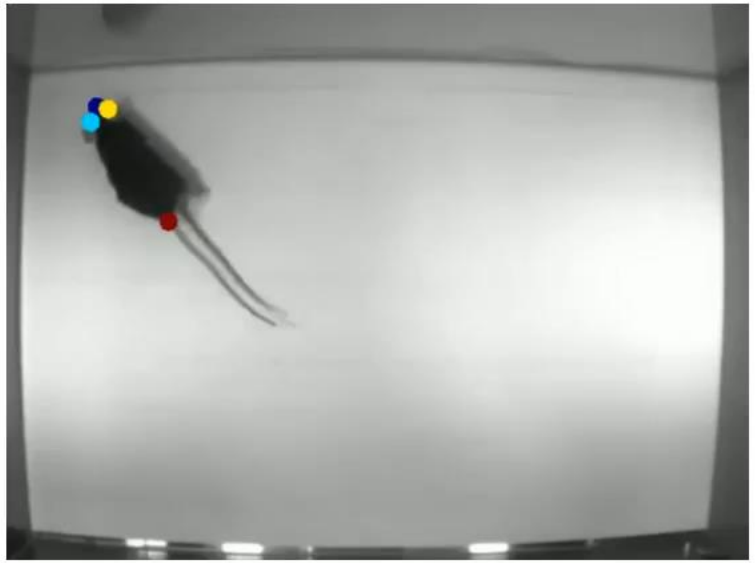
- Data augmentation
- Model architecture
- Optimization

Mathis, Schneider, Lauer & Mathis, 2020 Neuron

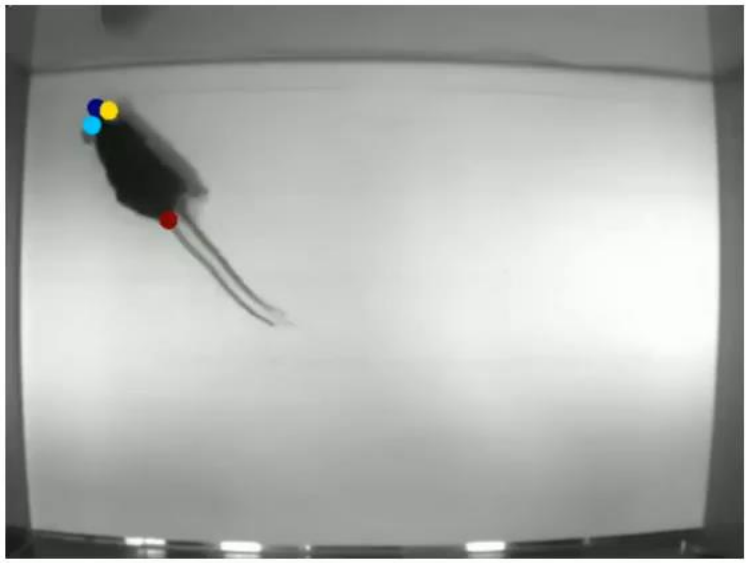
Data Augmentation matters: how to get the most out of your data! (code and videos in the Primer)



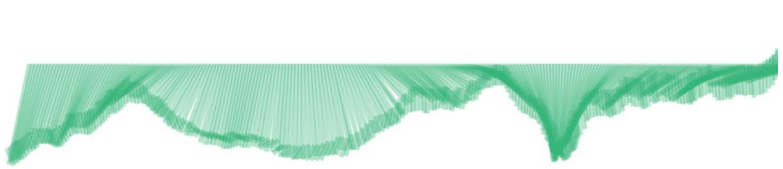
tensorpack

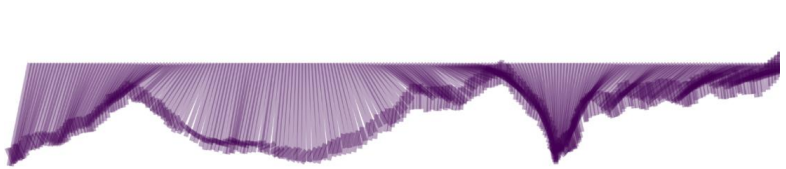


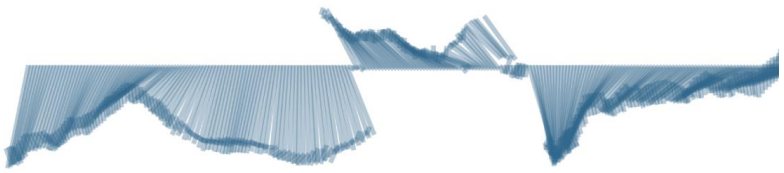
imgaug



scale-crop



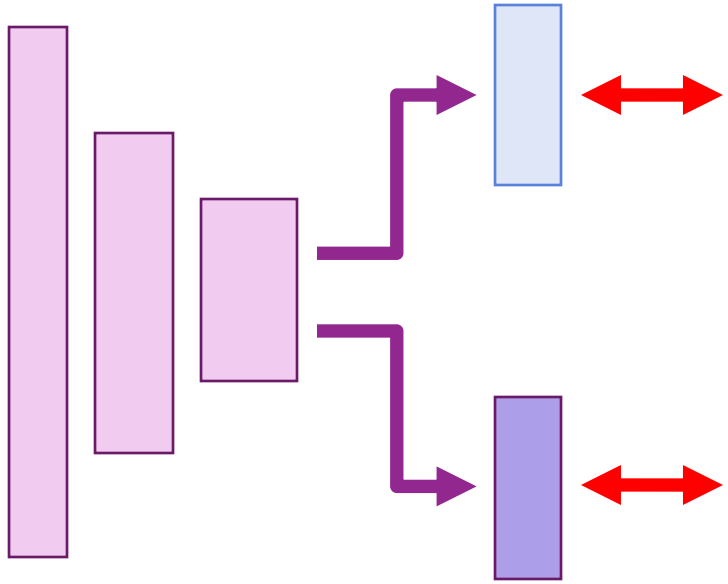




Mathis, Schneider, Lauer & Mathis 2020
Neuron

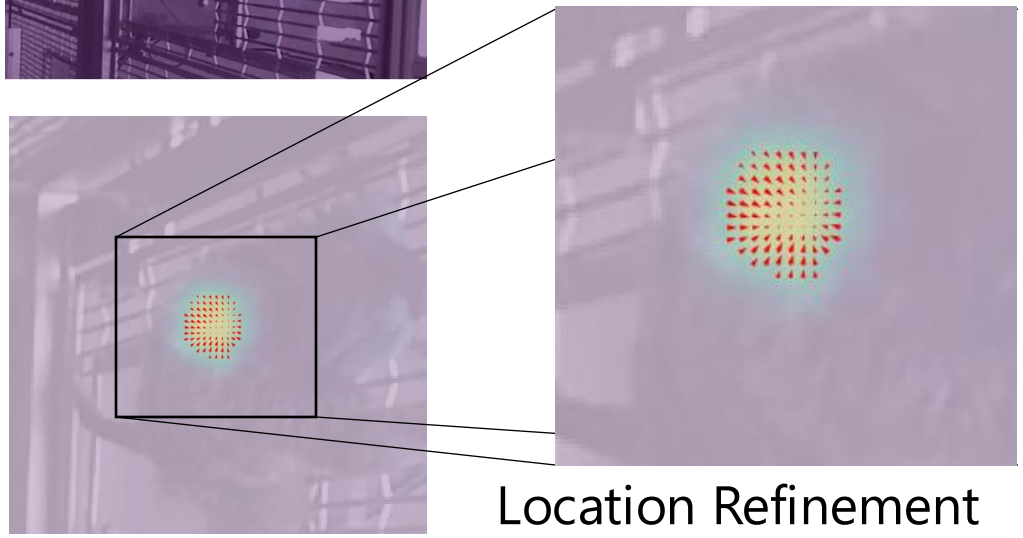
Multi-Task Deep Convolutional Network

- Minimize the combined loss; segmentation + vector field
- Transfer learning (ConvNet pre-trained on object recognition)
- Perform augmentation during training



deconvolution layers (3 per bodypart)

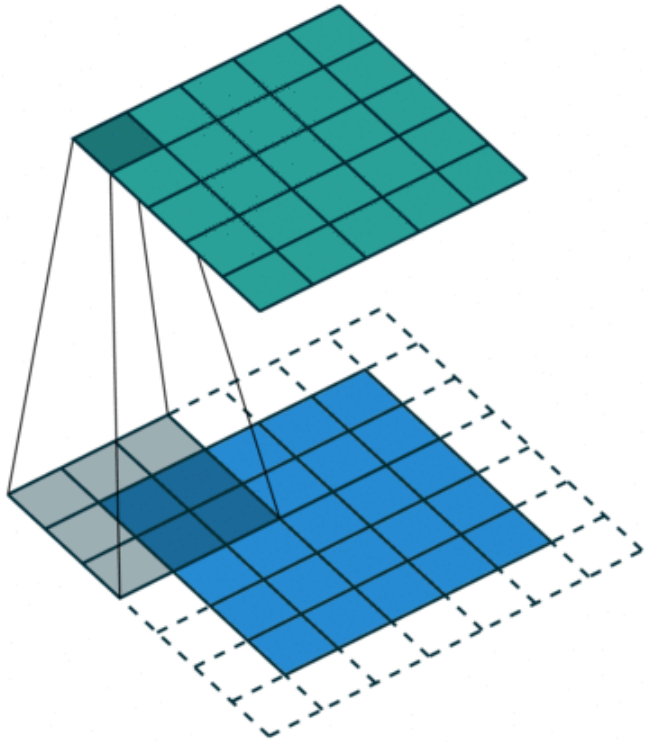
Detection (scoremap)
Cross entropy loss



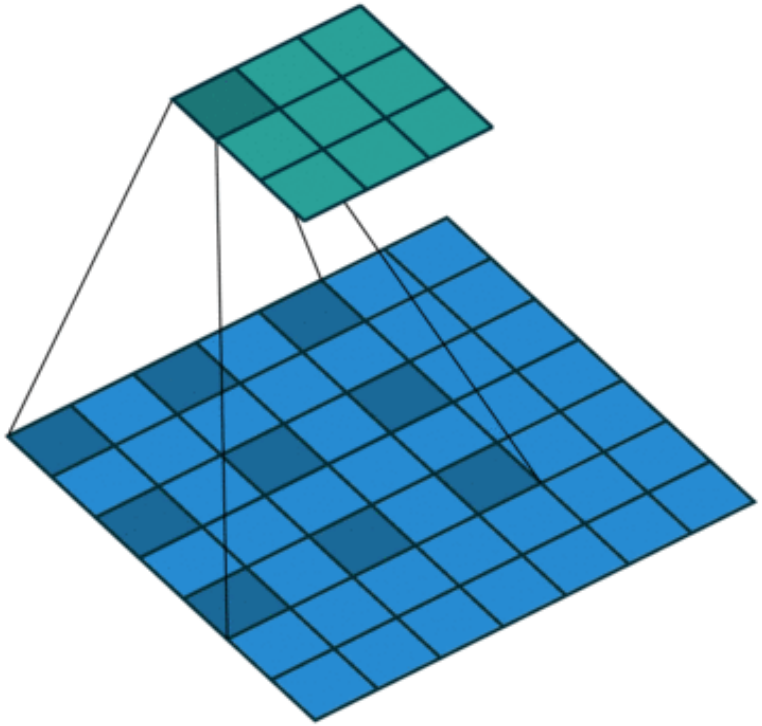
Location Refinement
Smooth L_1 loss

NN Primer: Convolutions

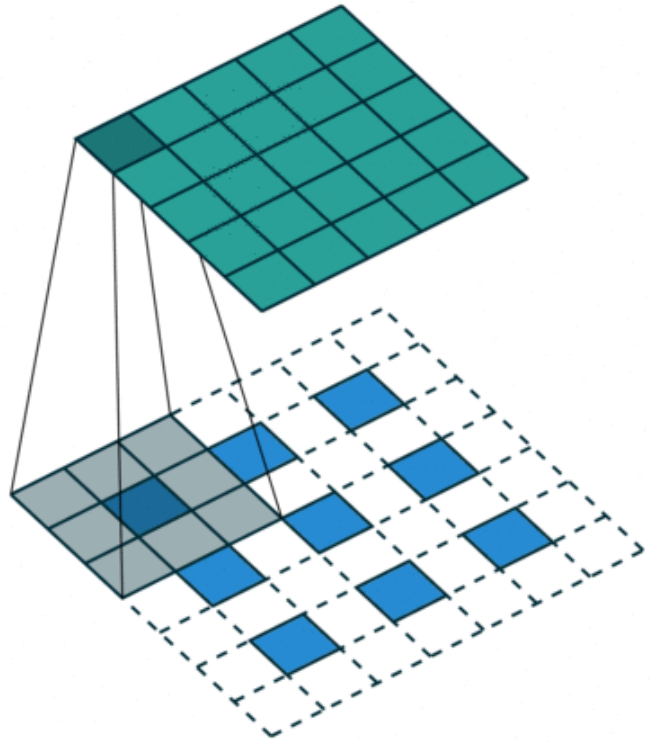
Basic convolution (“same”)



Strided convolution



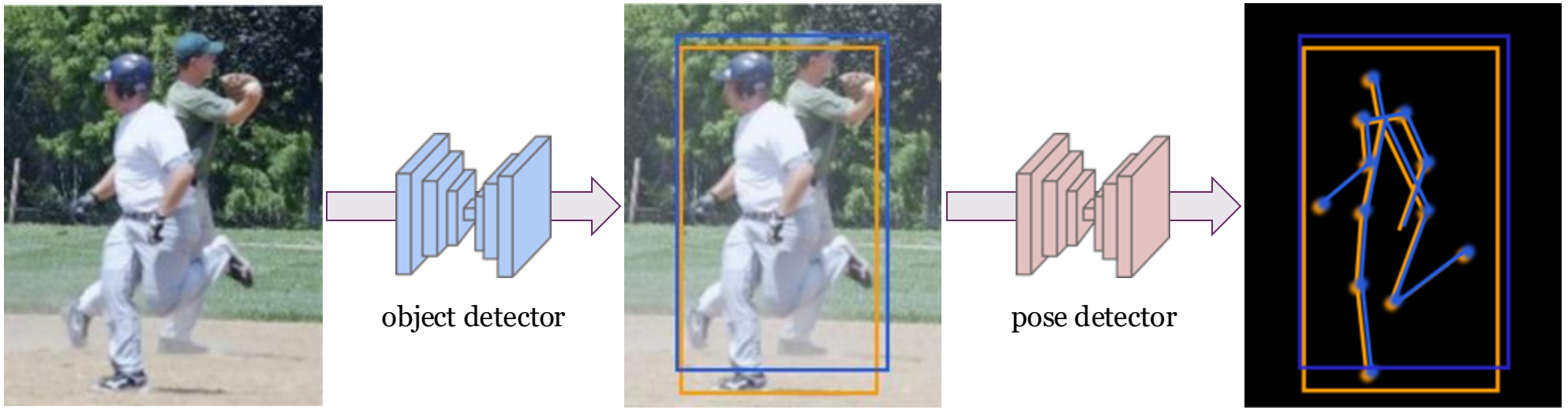
Strided **de**convolution



- https://github.com/vdumoulin/conv_arithmetic

Multi-instance pose estimation

Top-down Approach (TD)



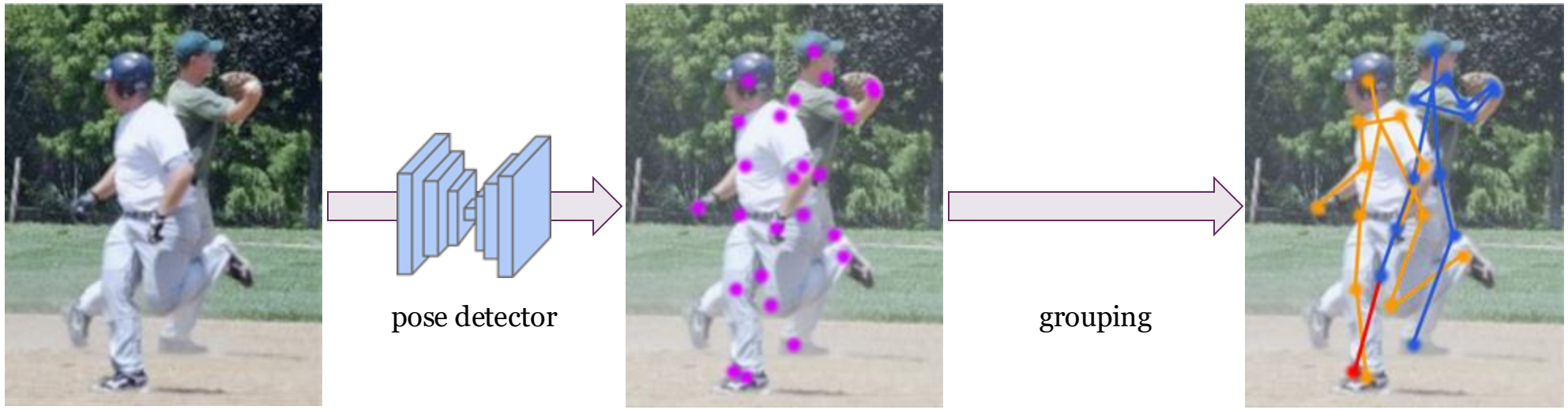
Li et al., CVPR 2019

\+ more accurate in less crowded scenes

\- ambiguous individuals in the same bounding box

Multi-instance pose estimation

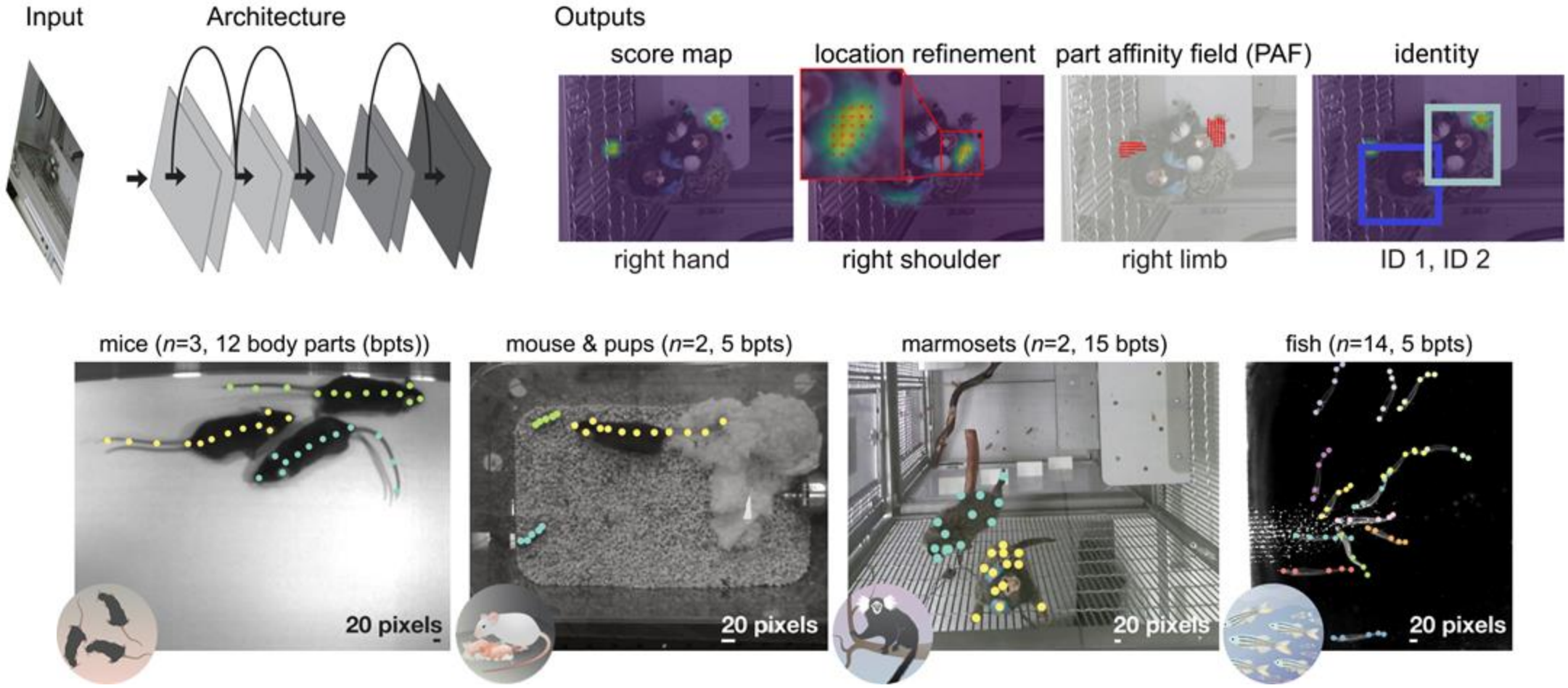
Bottom-up Approach (BU)



17

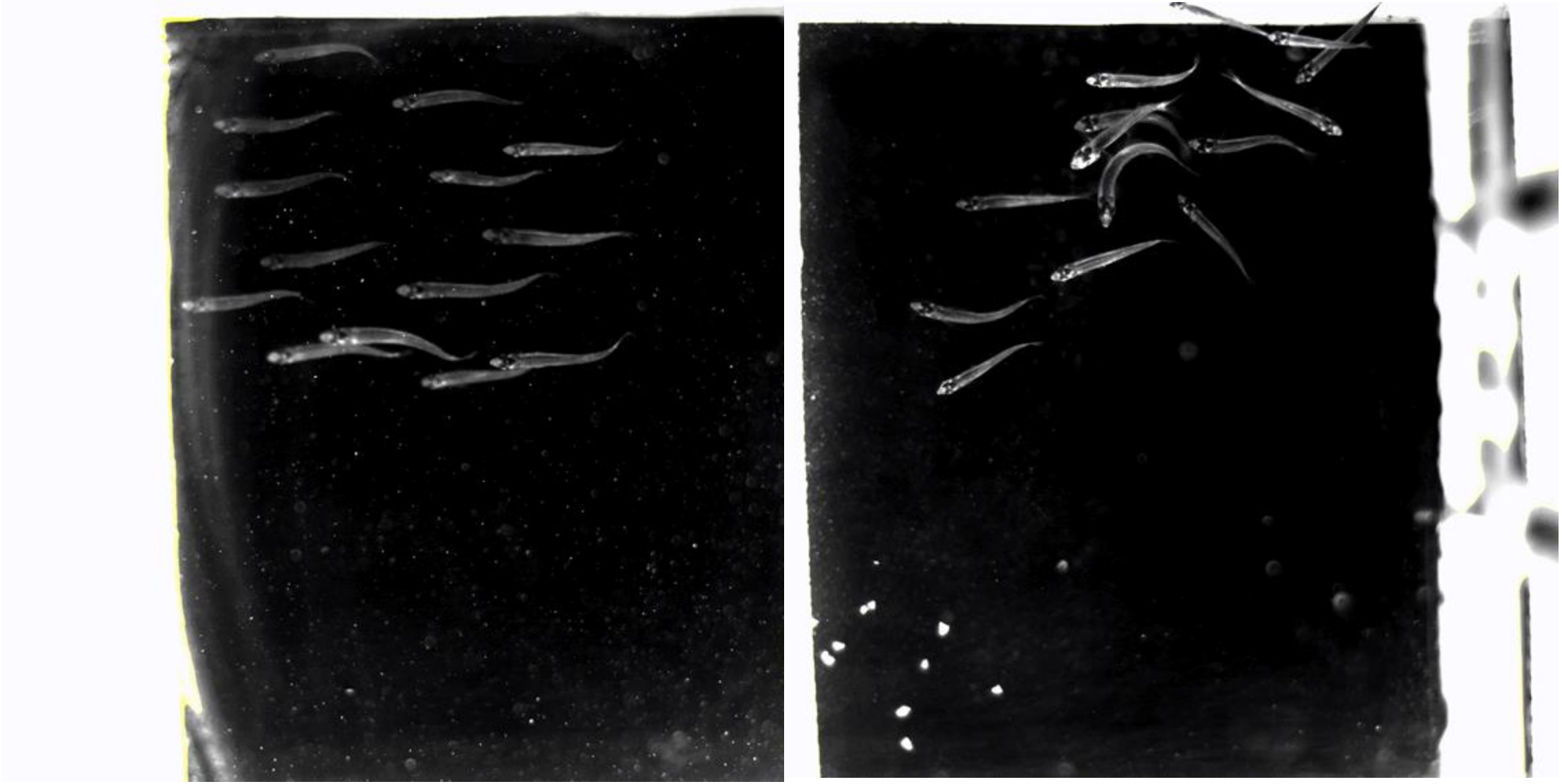
+ no need for detector
+ more accurate in crowded scenes
- grouping key points is a difficult problem (that often no longer relies on visual information)
- lack of precision

Multi-animal pose estimation & identification



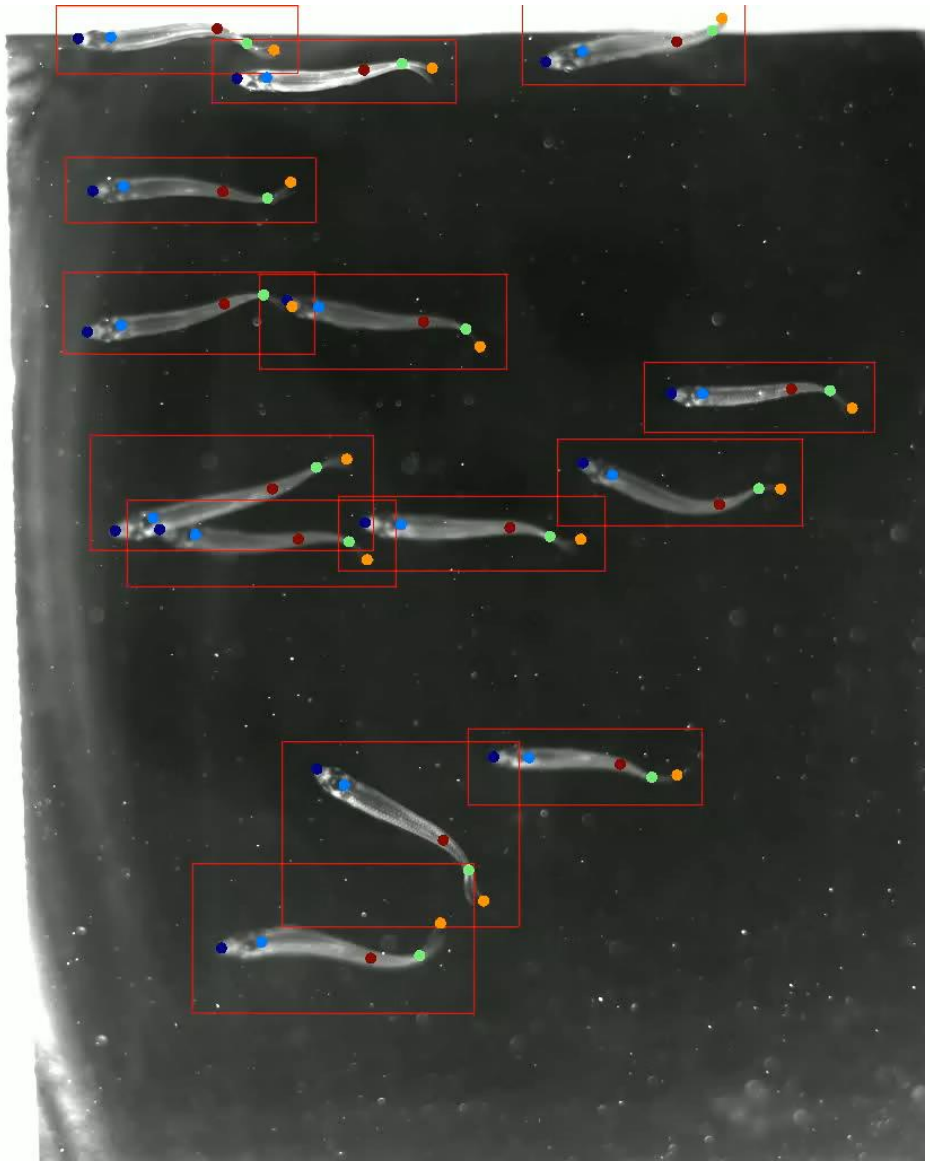
- Nature Methods 2022

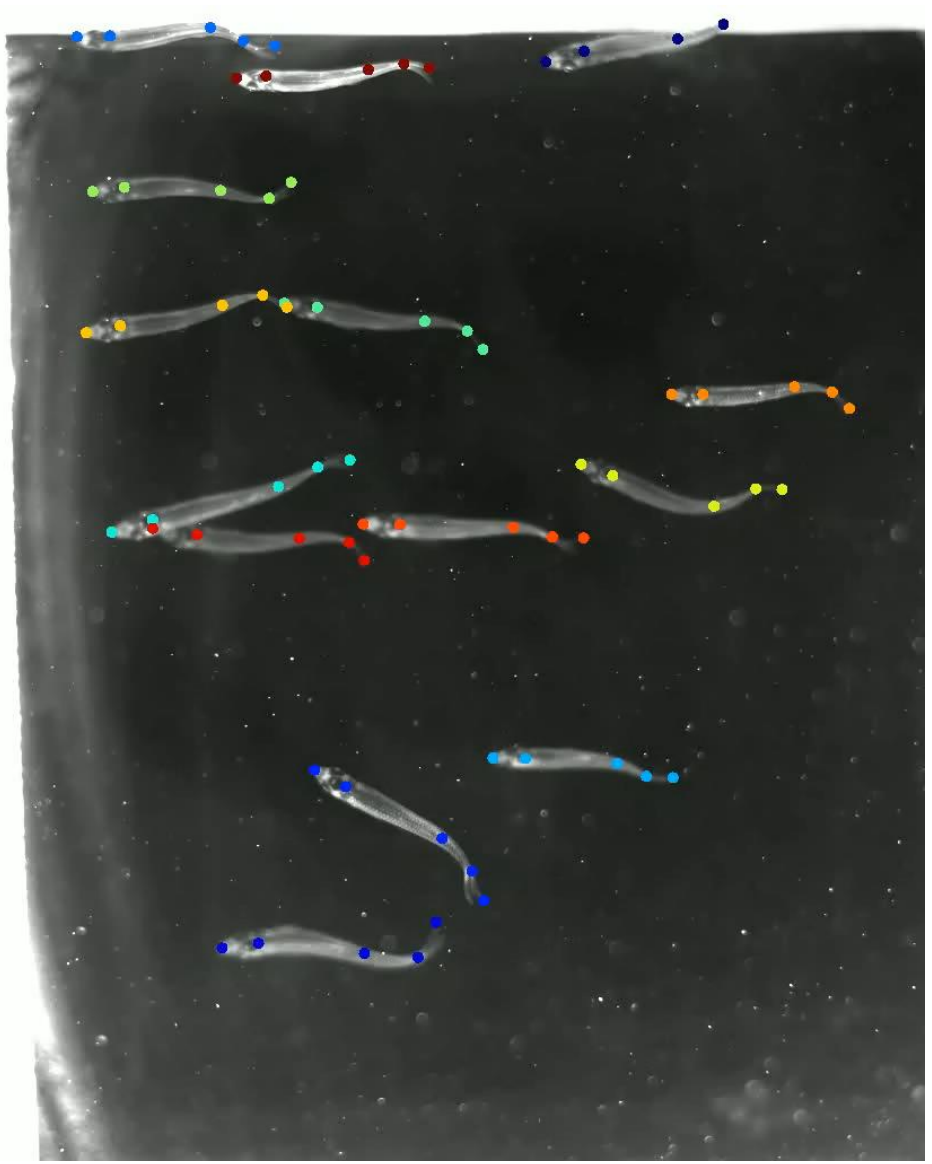
Multi-animal pose estimation & identification



Videos courtesy of
Prof. Valentina Di Santo (now U of Stockholm) and George Lauer, Harvard University

Multi-animal pose estimation & identification

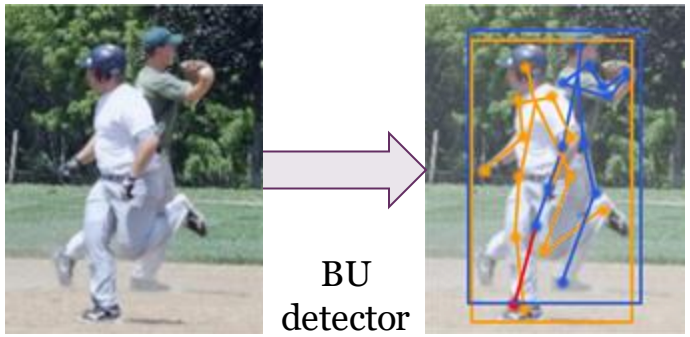




Bottom-Up Conditioned Top-Down Approach (BUCTD)

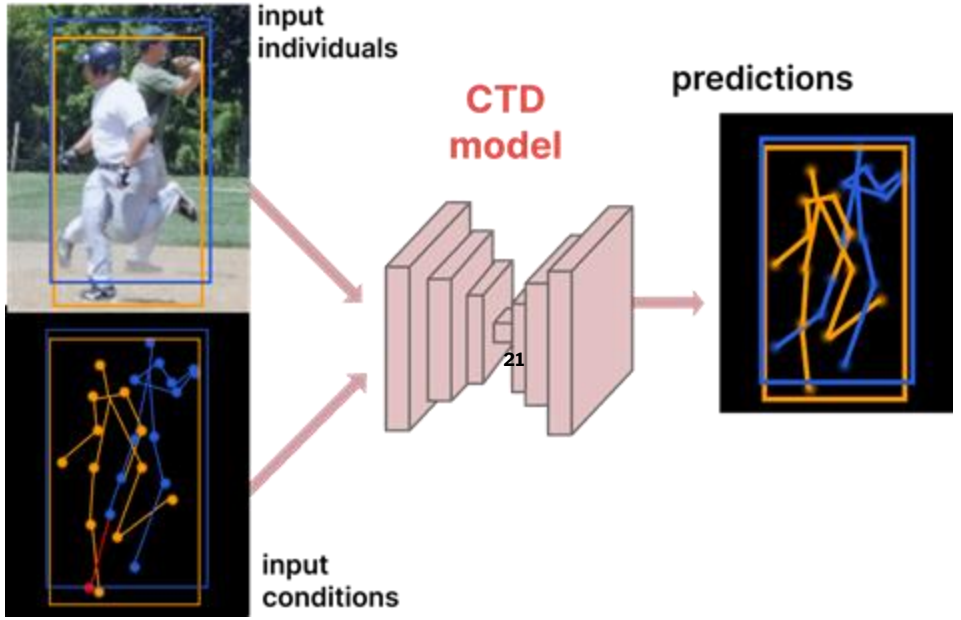
Hybrid approach leveraging the strengths of TD and BU approaches to overcome the detection information bottleneck and ambiguity

Stage1: Object and pose detection



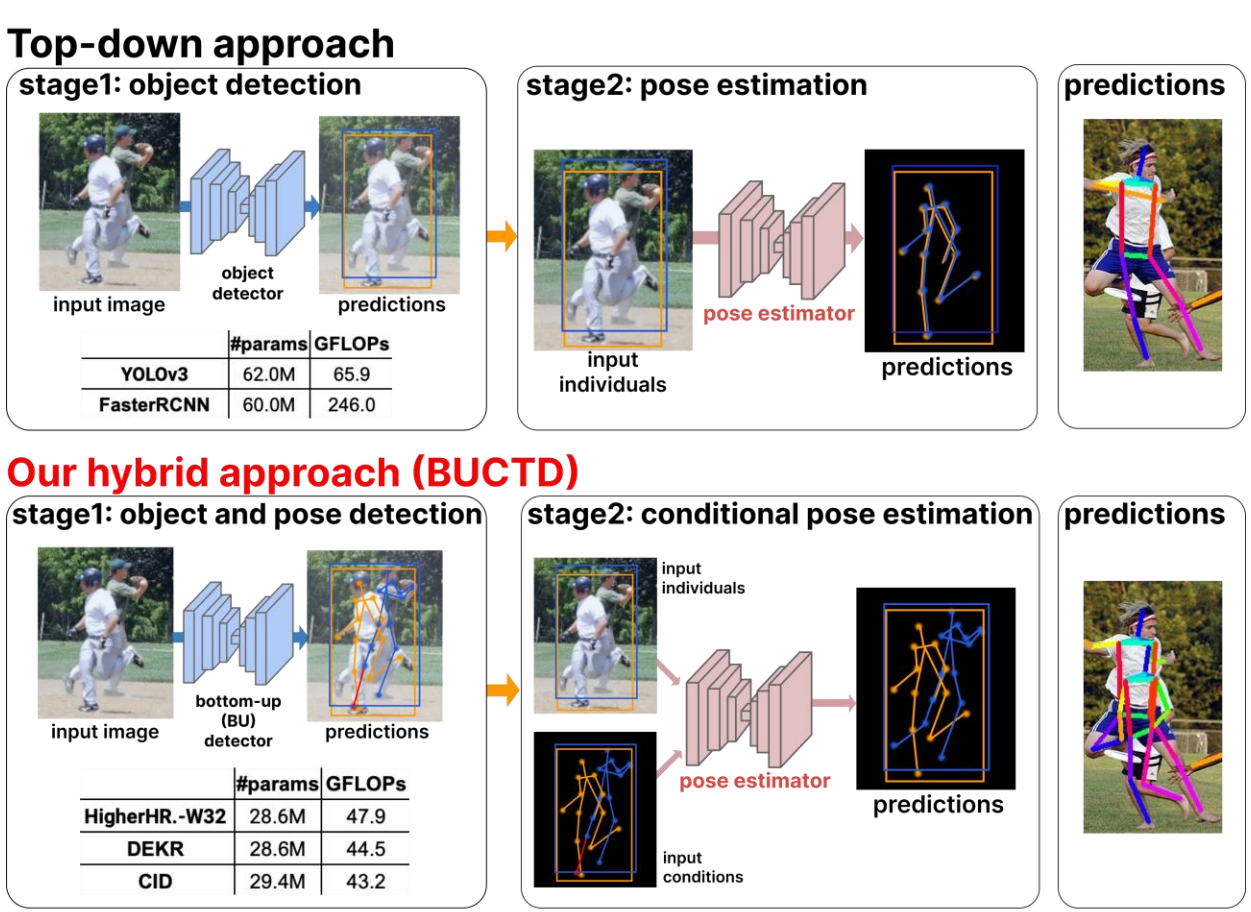
- get predicted pose from BU/single-stage pose estimator
- compute the individual bbox from pose

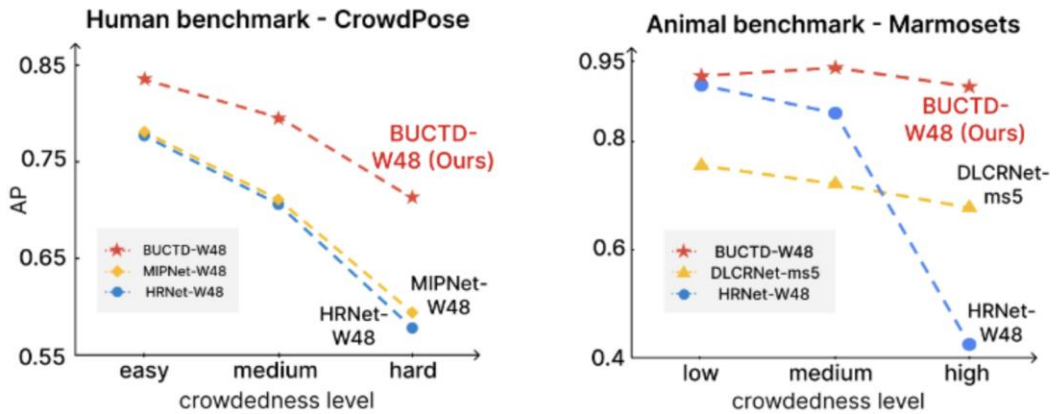
Stage2: Conditional Top-down (CTD) pose estimation



Zhou*, Stoffl*, M.W. Mathis, A. Mathis, ICCV 2023

Bottom-Up Conditioned Top-Down Approach (BUCTD)





SOTA on CrowdPose, OCHuman and four animal datasets

Zhou*, Stoffl*, M.W. Mathis, A. Mathis, ICCV 2023

An animal pose benchmark (horses) for robustness

owse State-of-the-Art

wse > Computer Vision > Pose Estimation > Animal Pose Estimation

Animal Pose Estimation

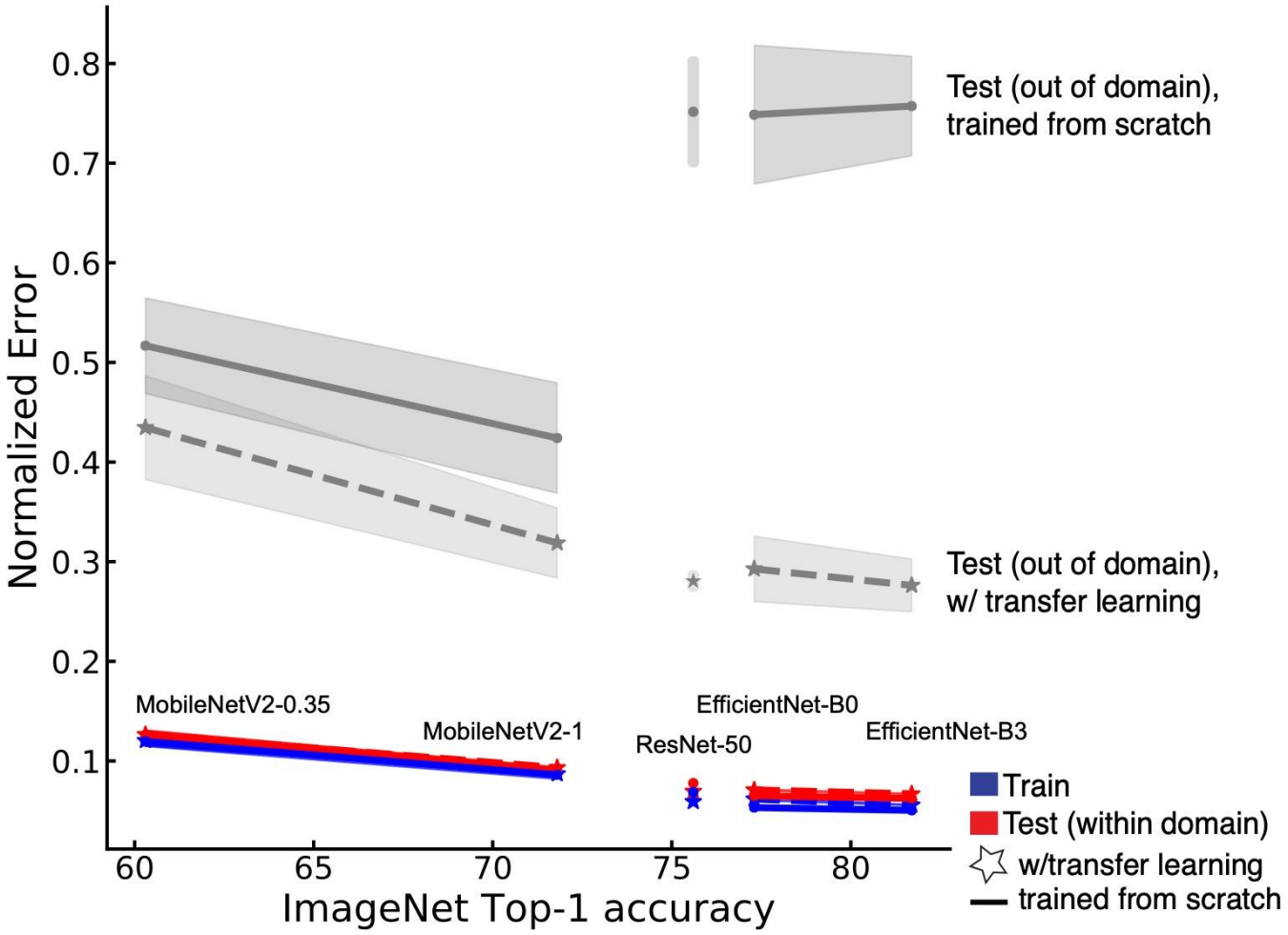
2 papers with code · Computer Vision

Subtask of Pose Estimation

Animal pose estimation is the task of identifying the pose of an animal.

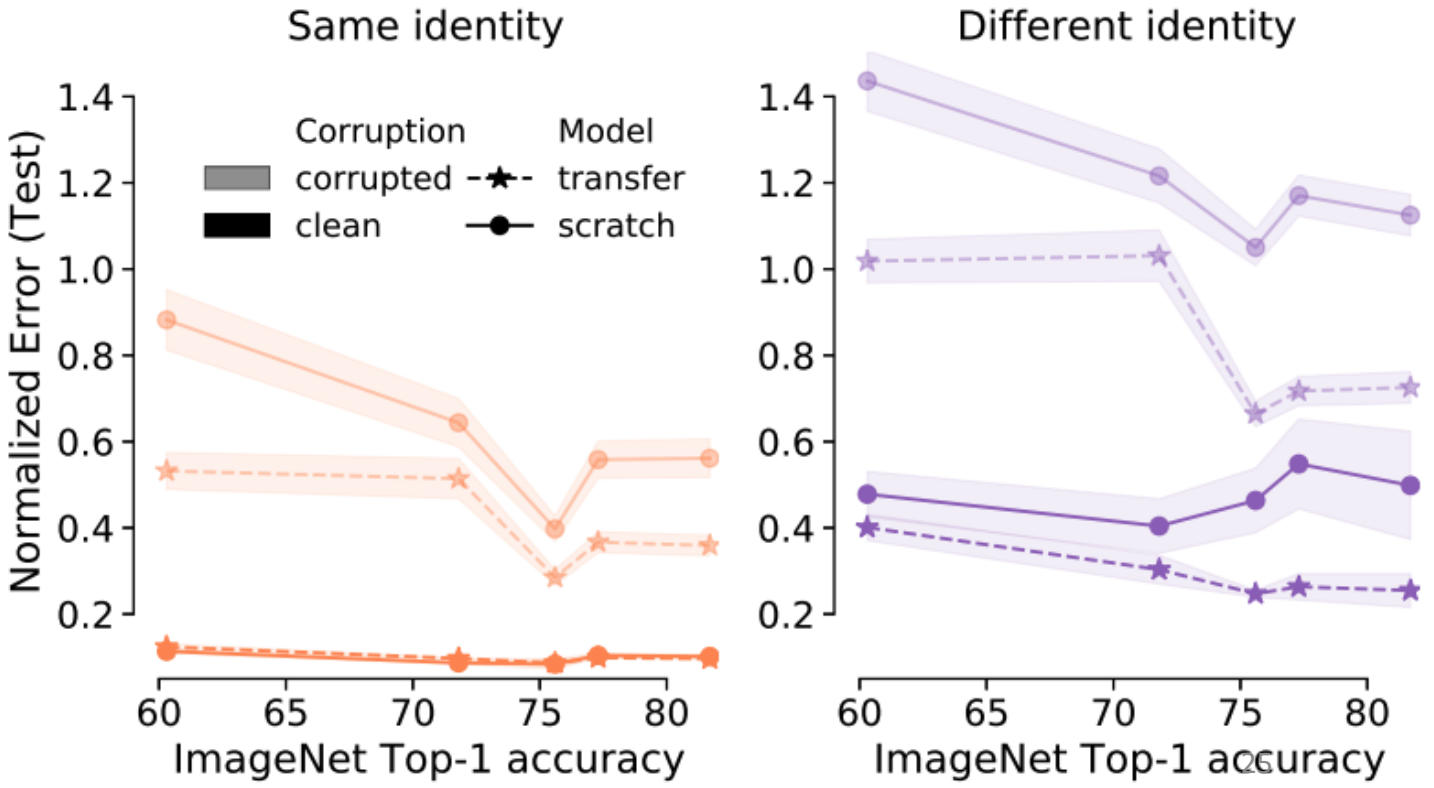
EPFL

Transfer learning (using pretrained ImageNet models), gives a 2X boost on out-of-domain data vs. from scratch training



- Principle: more powerful architectures generalize better!

Horse-C: an animal pose estimation corruption benchmark for robustness



Horse-C aimed to contrast the domain shift inherent in the Horse-10 dataset with domain shift induced by common image corruptions

- Mathis, Biasi et al. WACV 2021

Transfer learning enables pose estimation with less data

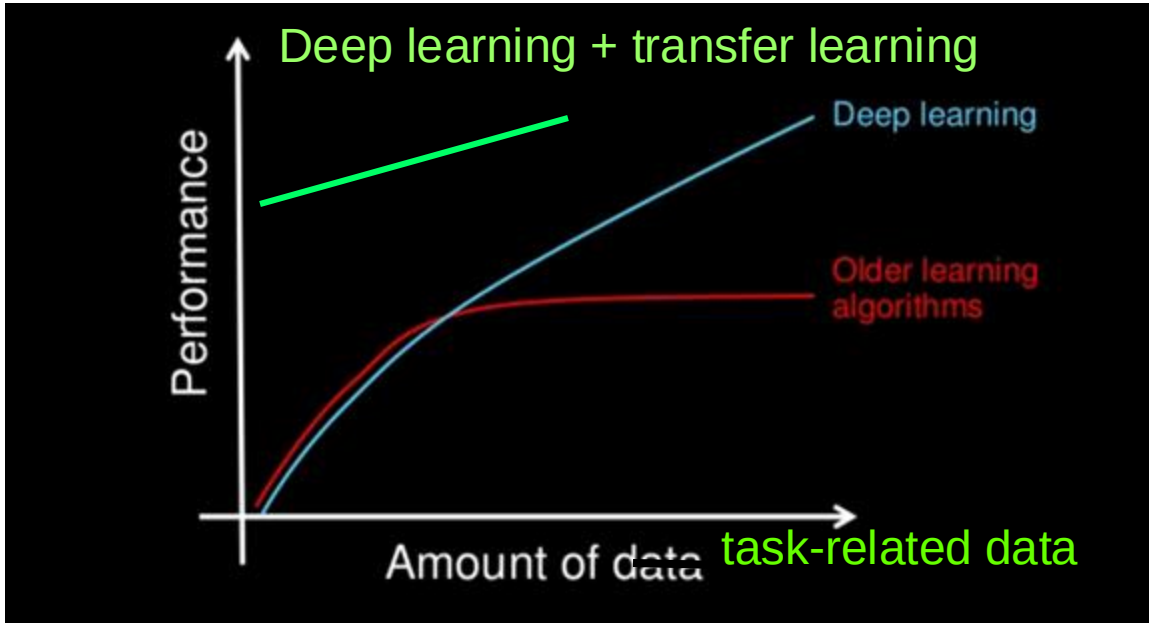
Foundational
deep neural networks

image → Predictor → pose

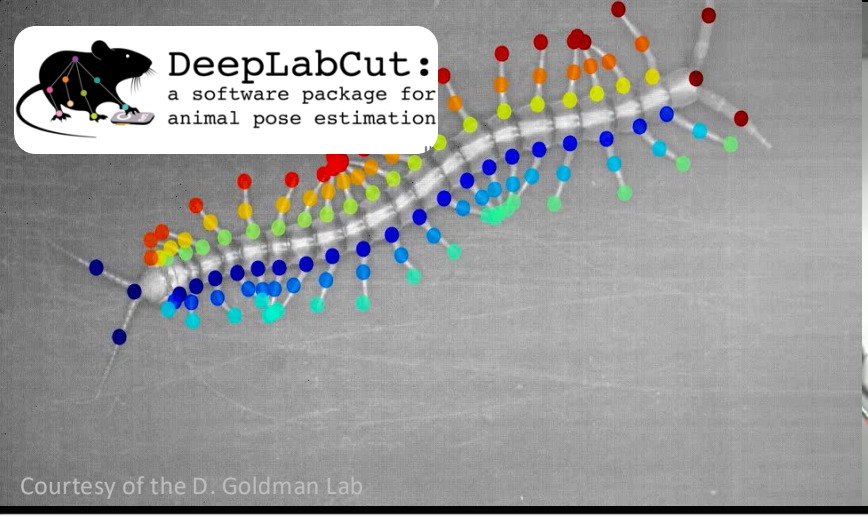
train

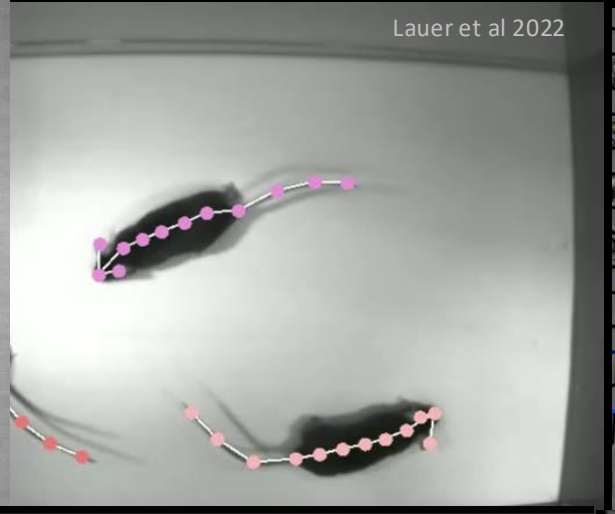
A lot of labeled images (>10^6 joints!)

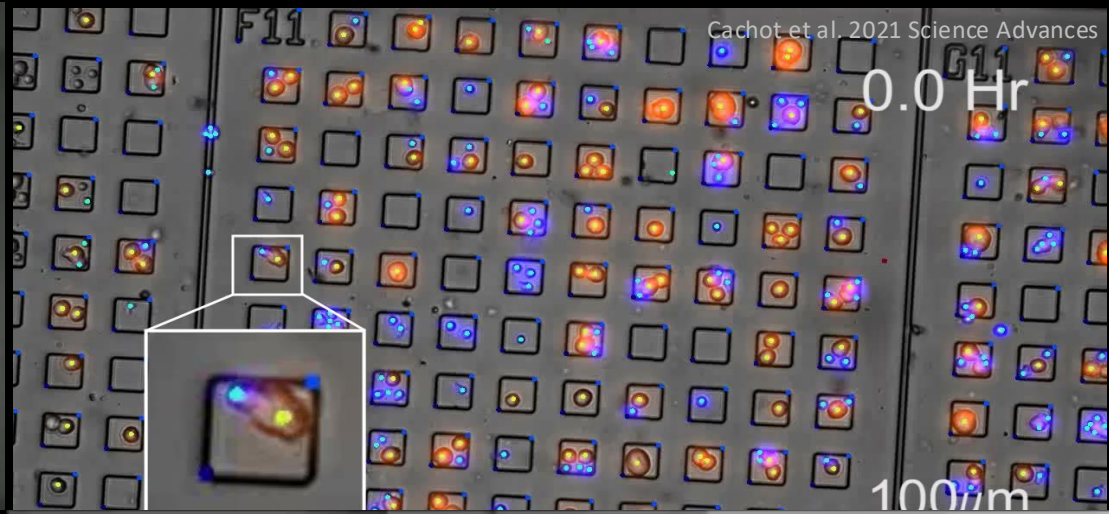
Only a few examples (10-200) for most applications

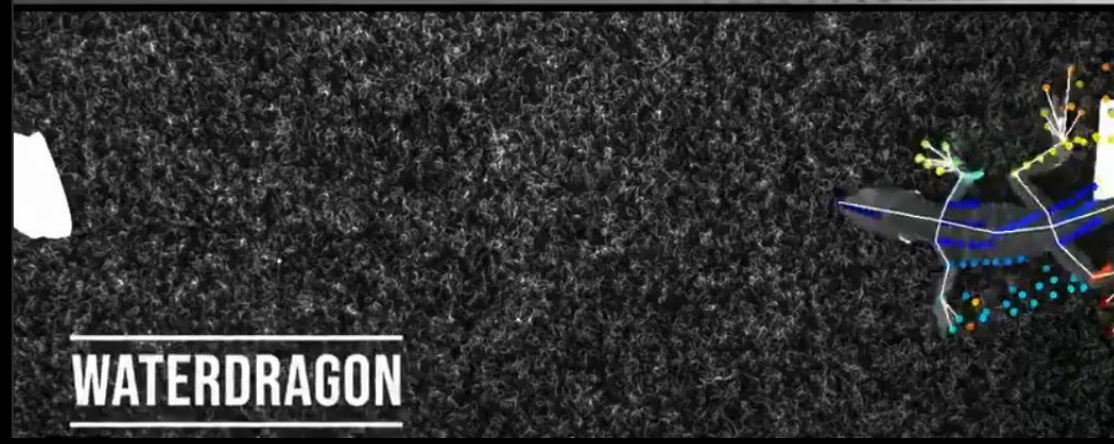


Andrew Ng



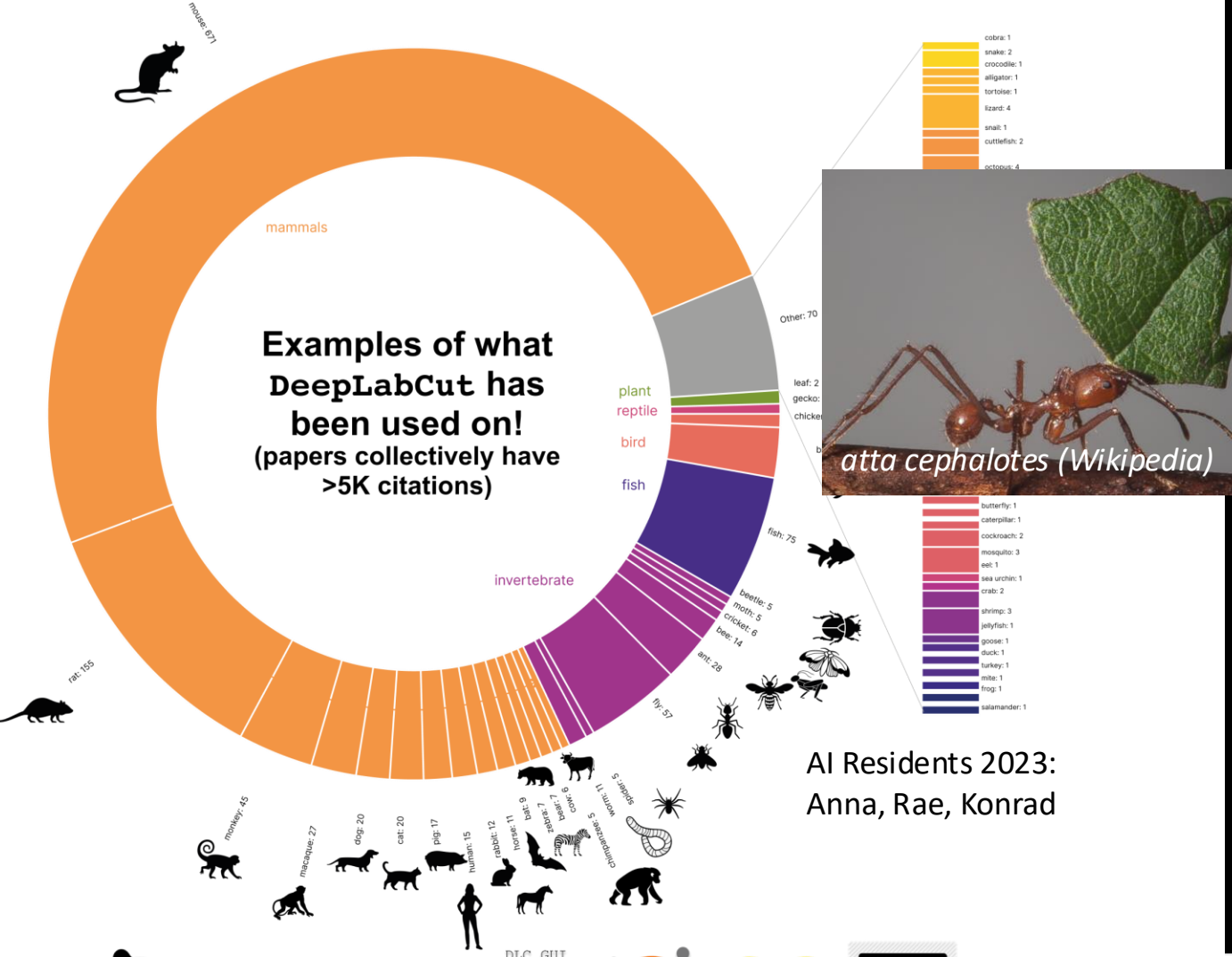


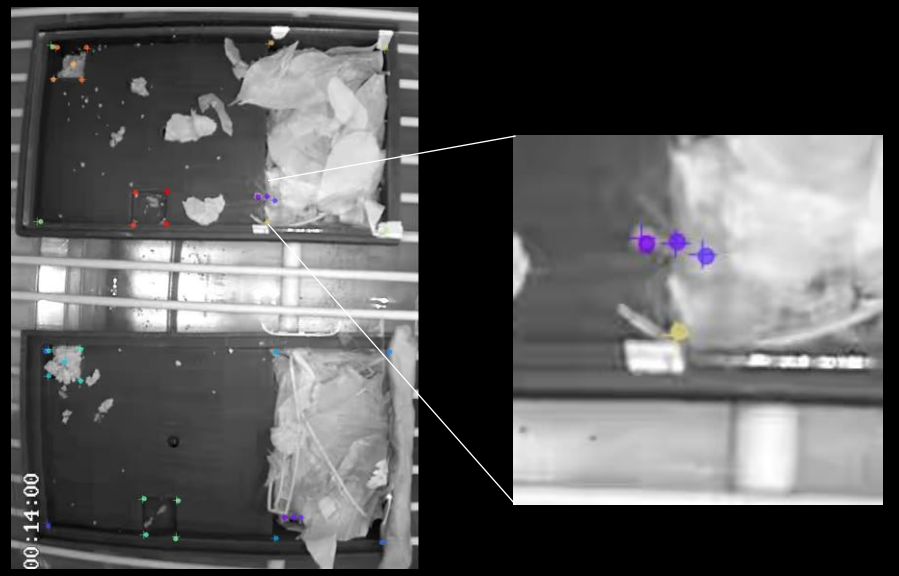


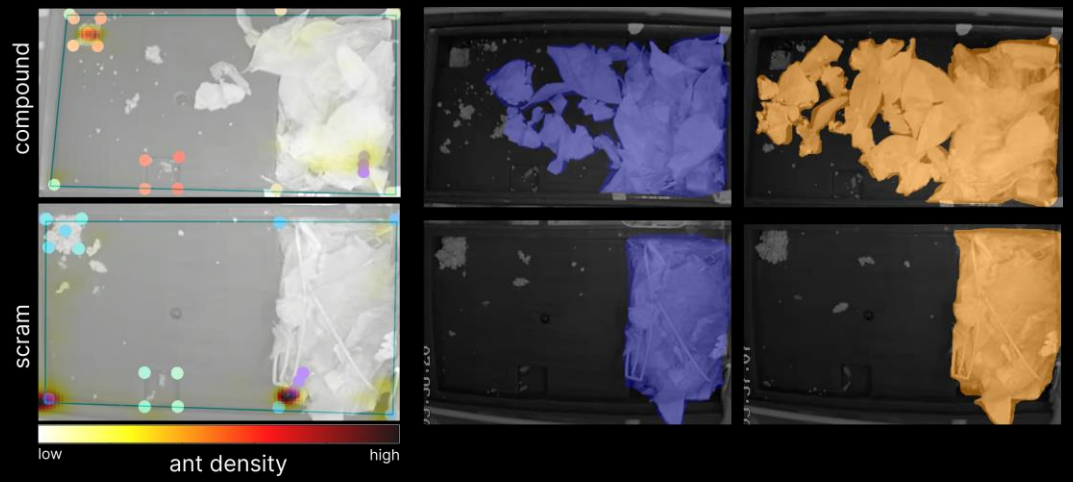


Nature Neuro 2018, Nature Prot. 2019, Neuron 2020, WACV 2021, ICRA, 2021, CVPR-W 2021, Nature Methods 2022, ICCV 2023, Nature Communications 2024

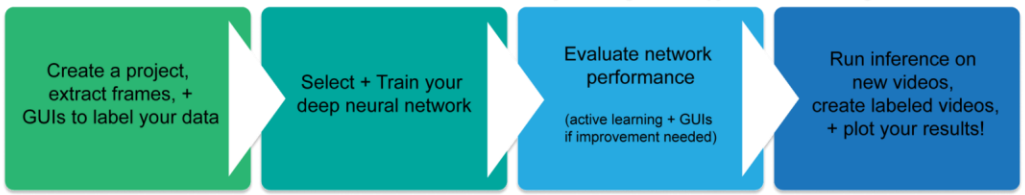
800,000+ downloads | >12K monthly

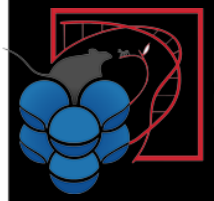






AI Residents 2023:
Anna, Rae, Konrad

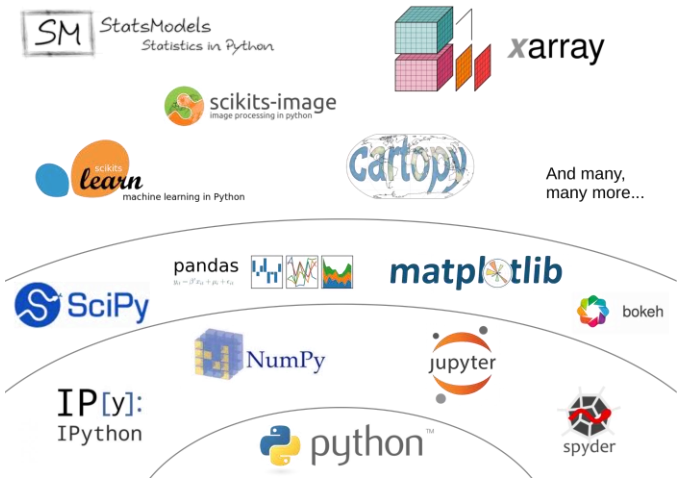




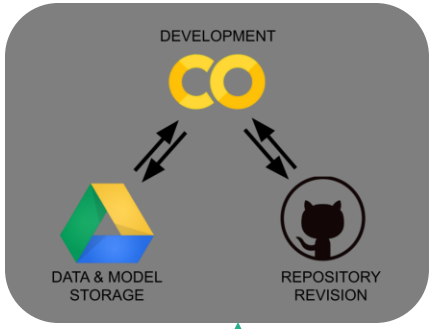
BERGER LAB

Gilbert, Glastad et al. in press Cell 2025

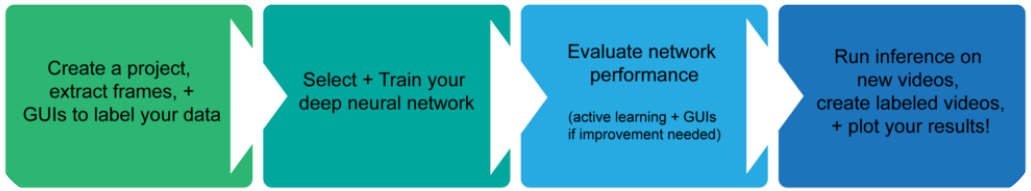
Built on the open source python stack:



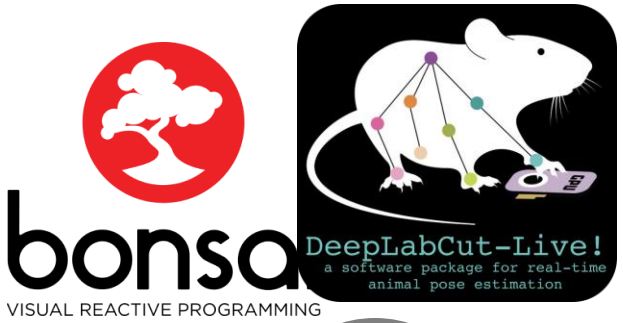
User testing/dev & deployment:

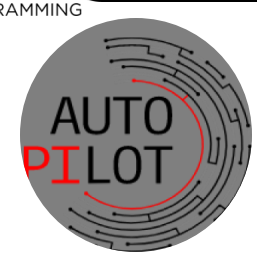


use our Project Manager GUI, Jupyter Notebooks, Google Colab, or terminal!



Real-time specific tools:

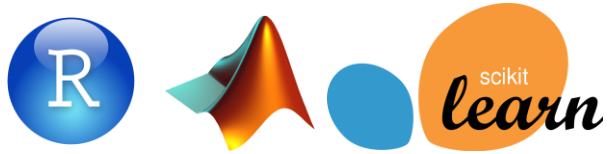




Computer Vision:

Larger scale pipeline computing:

Post- pose estimation tools:



Classifiers: SVMs, Random Forrest, ANNs
- B-SOID, ETH-DLC Analyzer, simba

Models: HMMs, decision-trees, ANNs

Ethograms: BORIS, BENTO, AmadeusGPT, Keypoint-MoSeq,

Clustering: CEBRA, MotionMapper, JAABA

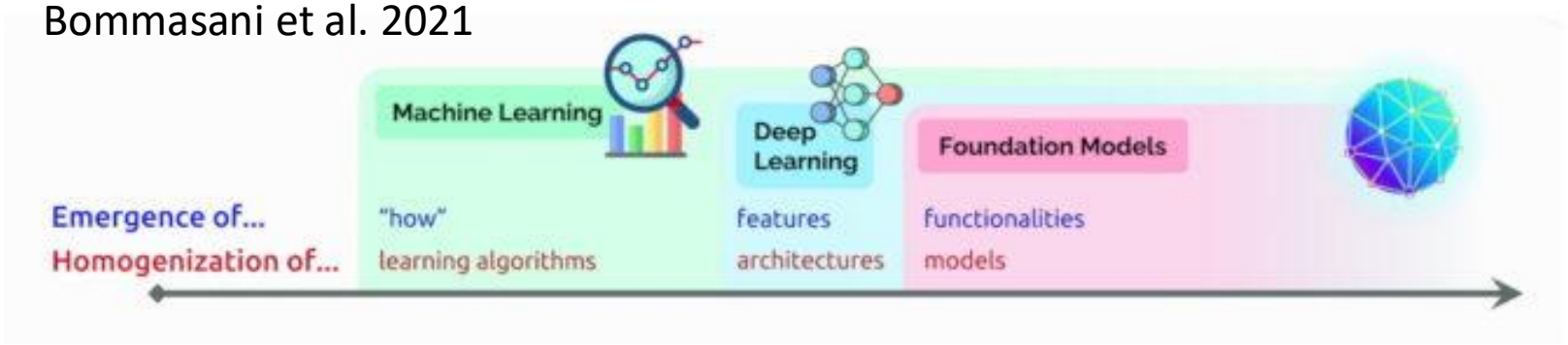
Motor analysis: DLC2Kinematics

Many, many models are trained on (closed-source) animal datasets ....

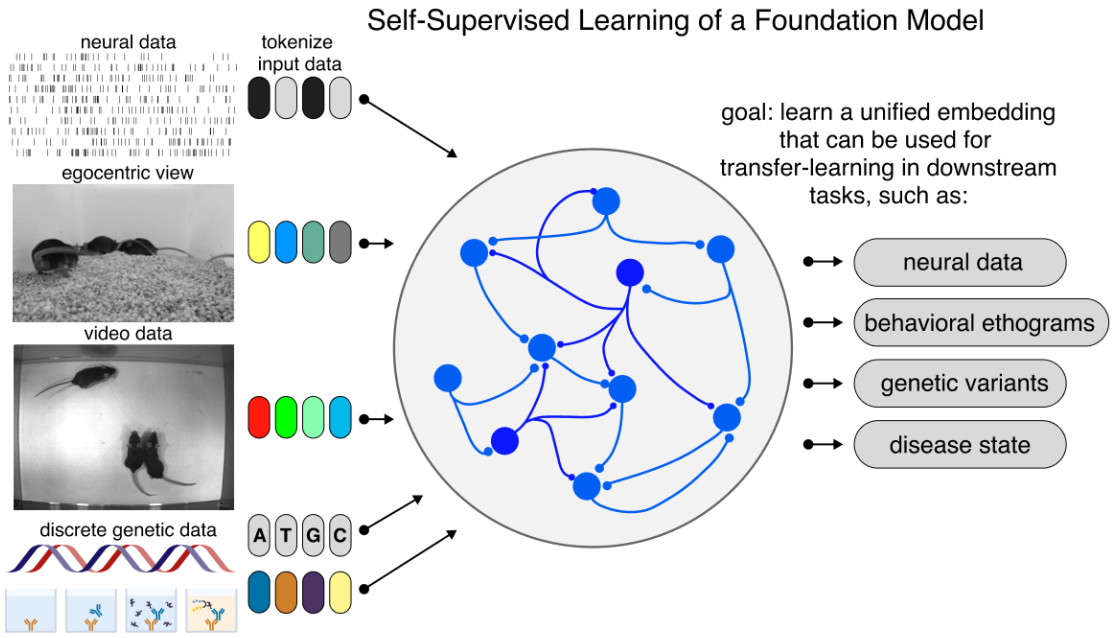
But what if we could combine this collective knowledge into better foundational models?

Foundational models

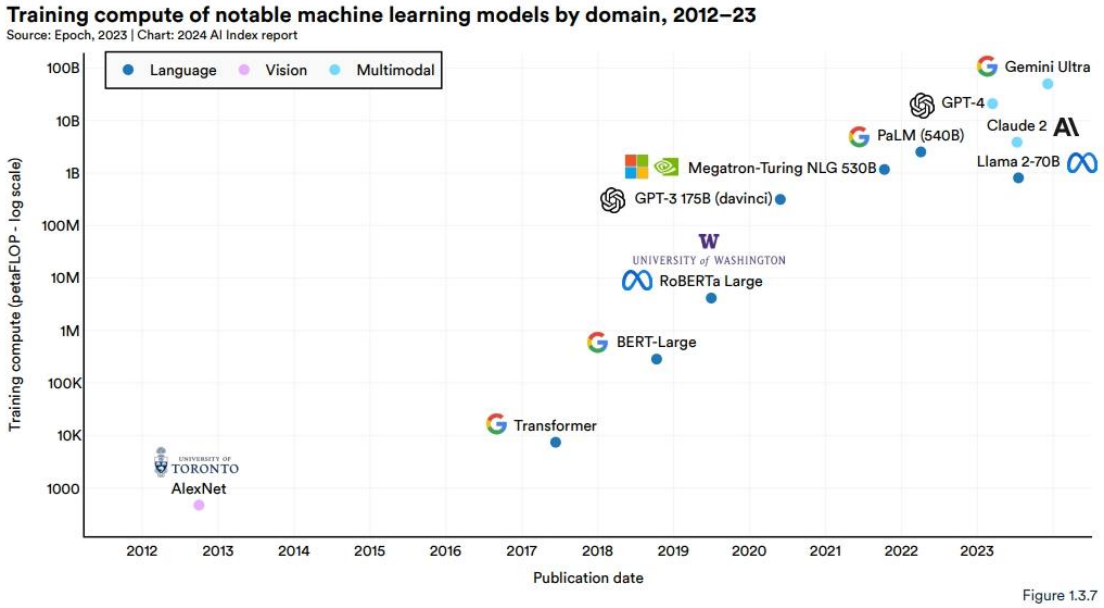
Bommasani et al. 2021



2024 AI Index report from the Stanford Institute for Human-Centered Artificial Intelligence, 149 foundation models were published in 2023, more than double the number released in 2022



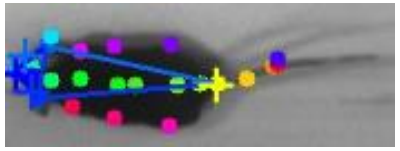
Mathis 2025 arXiv

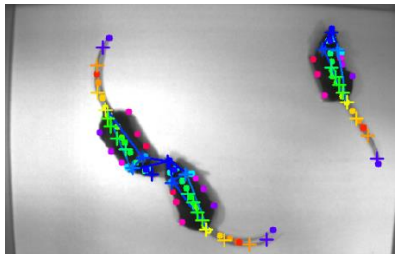


NVIDIA

New animal pose datasets ....

SuperAnimal- TopView Rodent (5K)





Typical Lab Setting
SuperAnimal TopView

26 keypoints

SuperAnimal- Quadruped (80K)

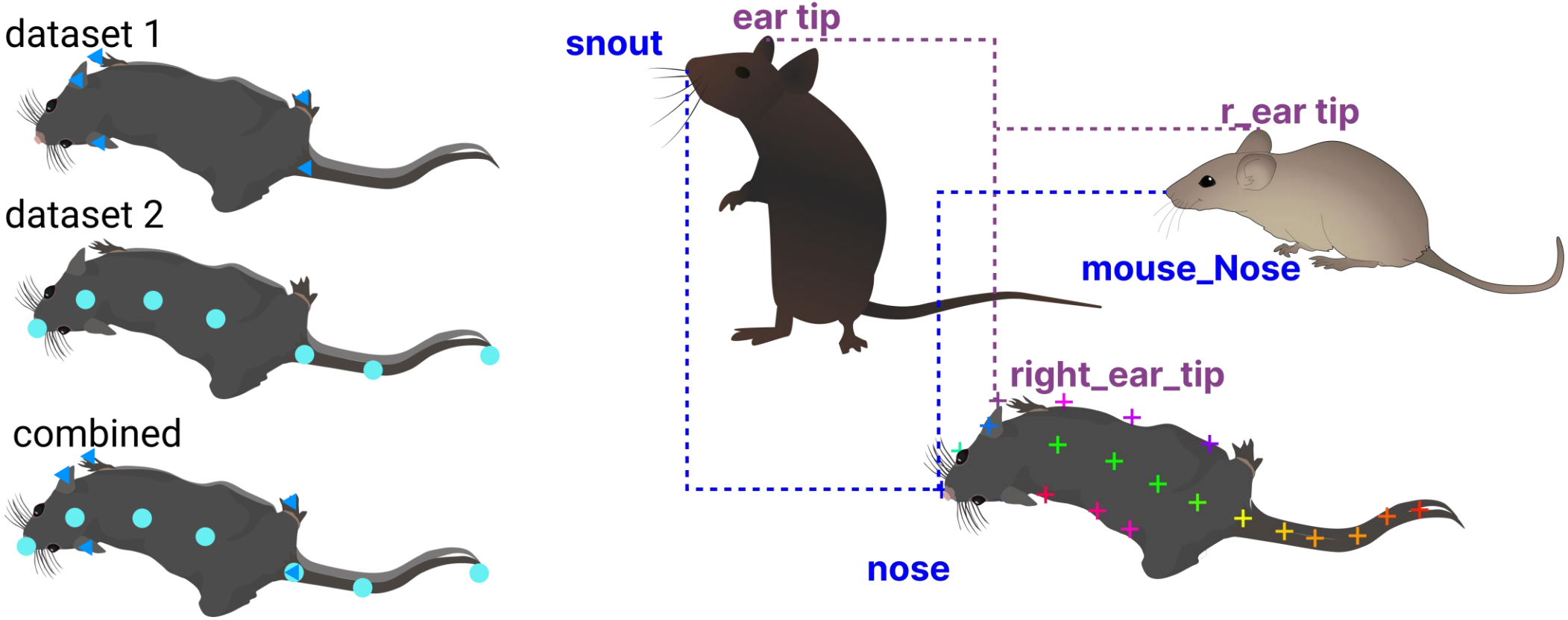
In-the-wild setting
SuperAnimal Quadruped

39 keypoints

Better foundational models for behavioral analysis

Challenge 1:

- Users do not define semantically similar keypoints, or even the same keypoints per animal



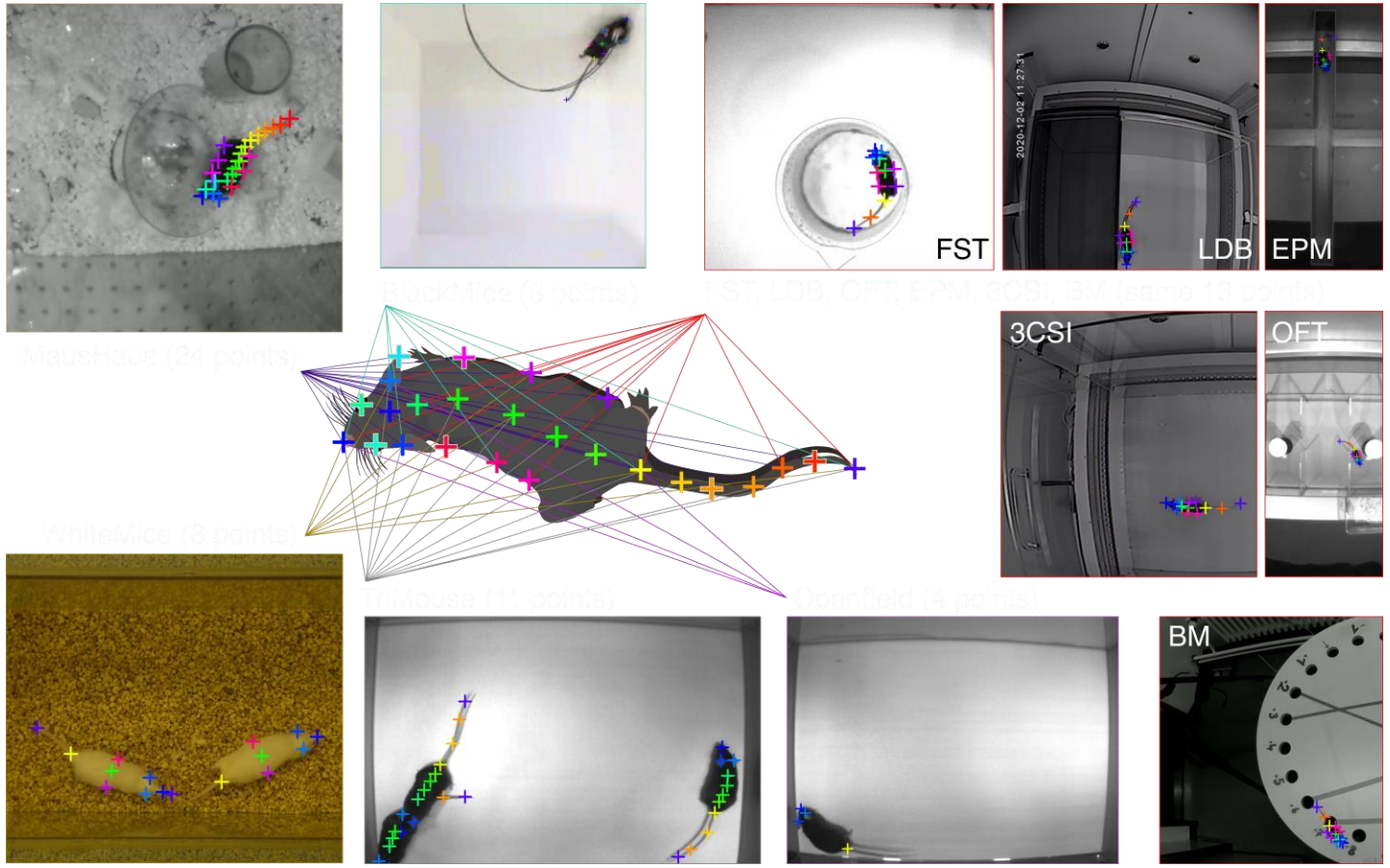
- *Pose estimation is a good video dimensionality reduction step*
- *This can be generalized to semantic behavioral labeling*

Better foundational models for behavioral analysis

Challenge 1:

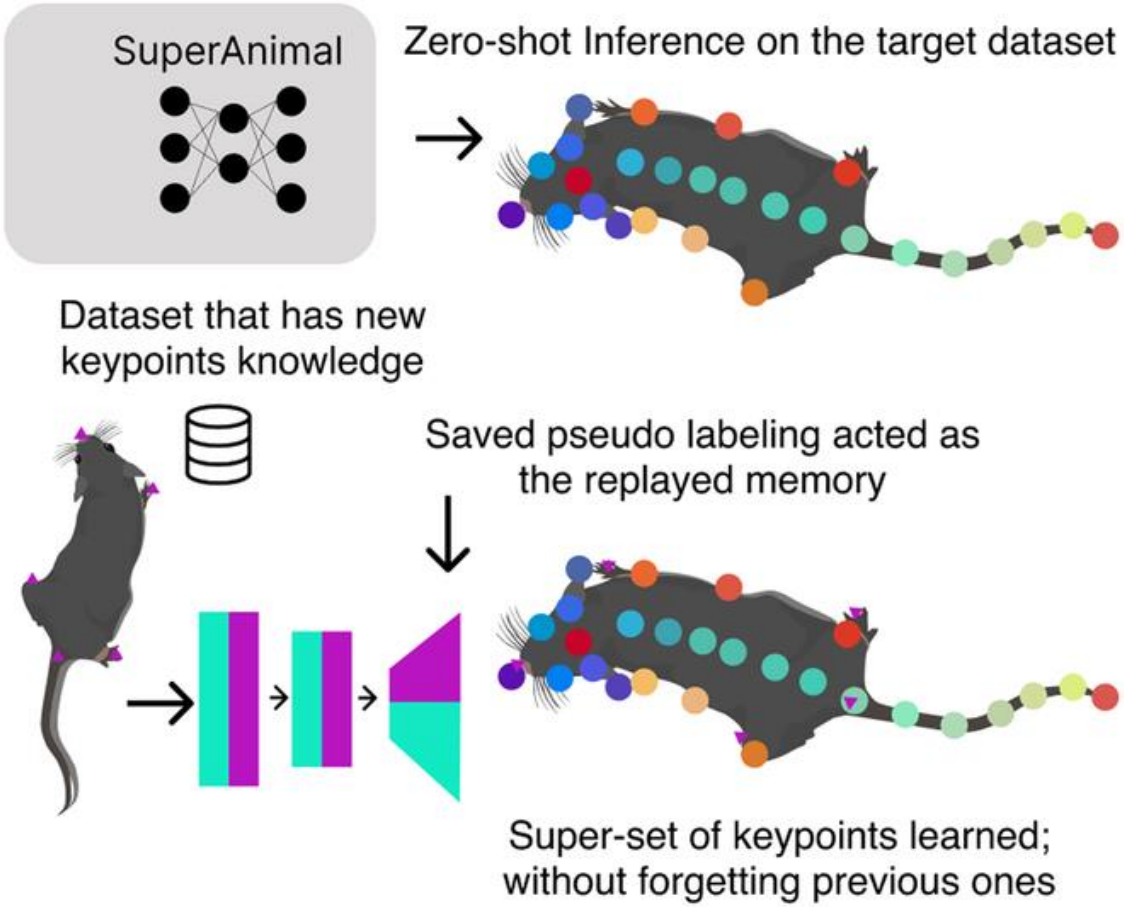
- Users do not define semantically similar keypoints, or even the same keypoints per animal

Solution: learning keypoint mappings, gradient masking

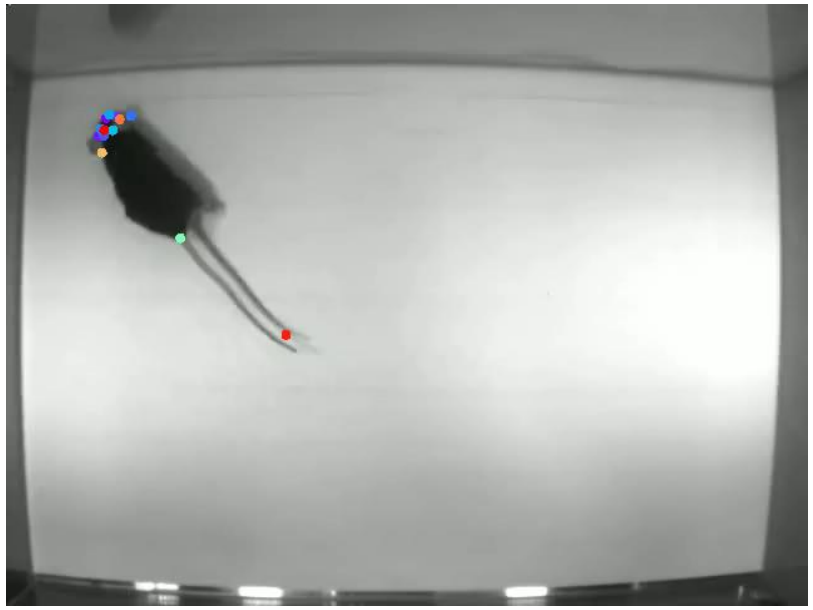


- *Pose estimation is a good video dimensionality reduction step*
- *This can be generalized to semantic behavioral labeling*

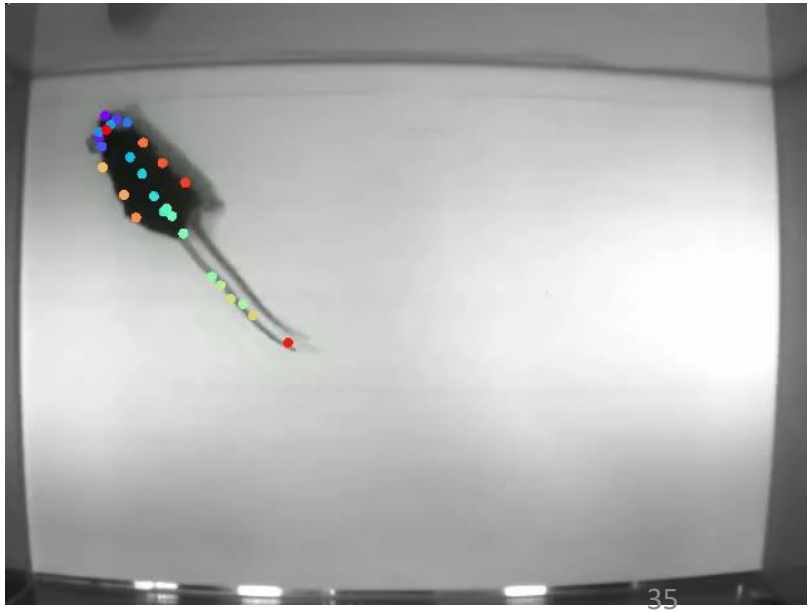
Memory-replay self-supervised fine-tuning boosts performance



Without memory replay



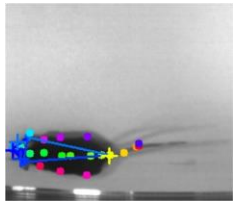
With memory replay



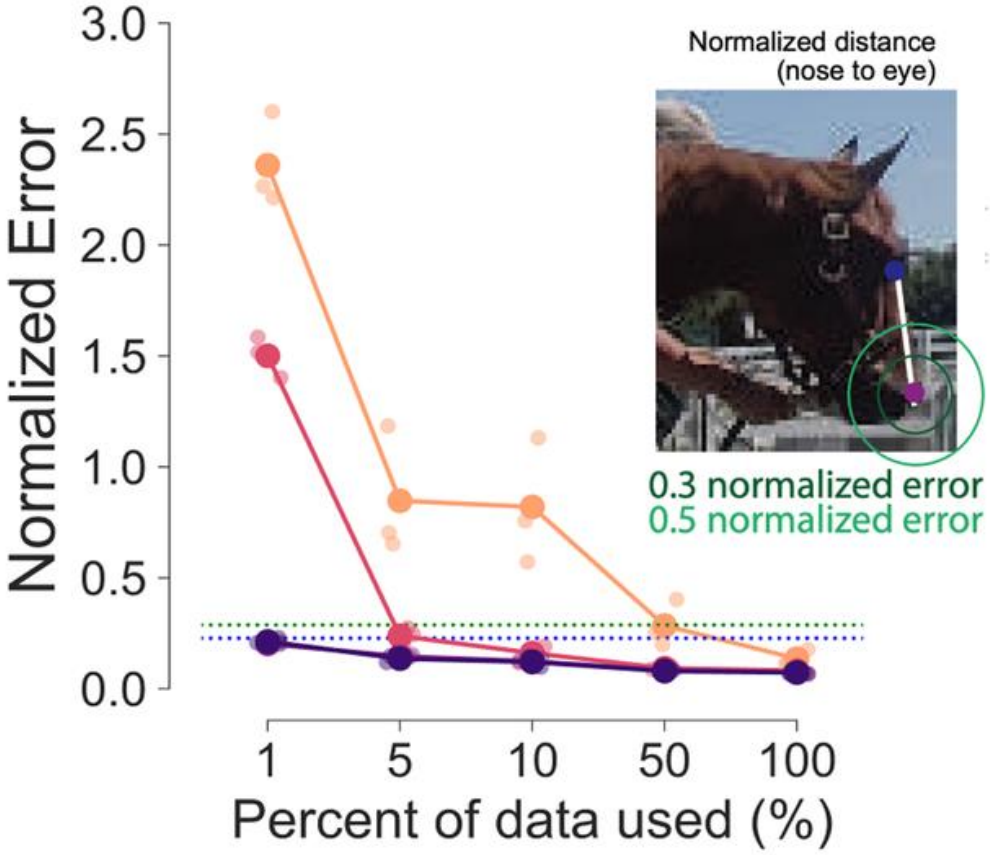
SuperAnimal models zero- and few-shot outperform ImageNet pretrained models (in low data regime)

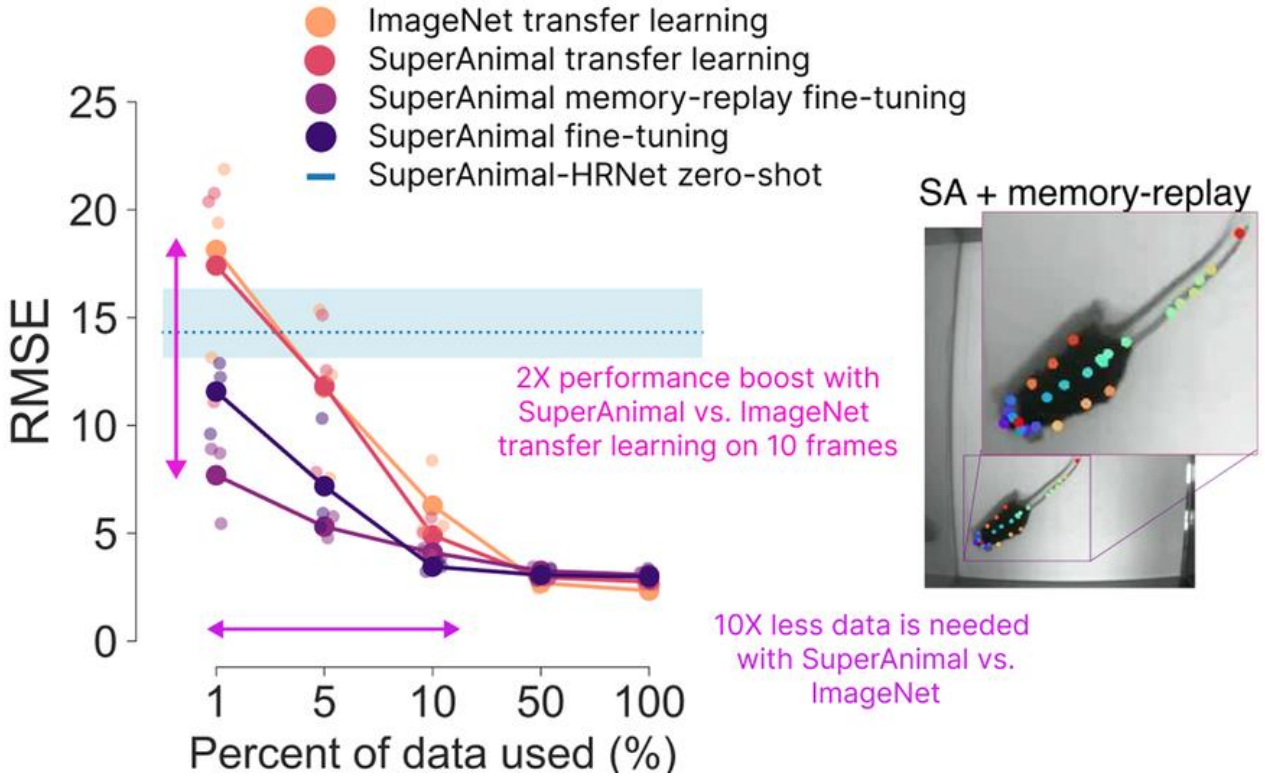
ImageNet transfer learning
SuperAnimal transfer learning
SuperAnimal memory-replay fine-tuning
SuperAnimal fine-tuning
SuperAnimal-HRNet zero-shot
AP-10K-HRNet zero-shot

GT 22 keypoints



GT 4 keypoints

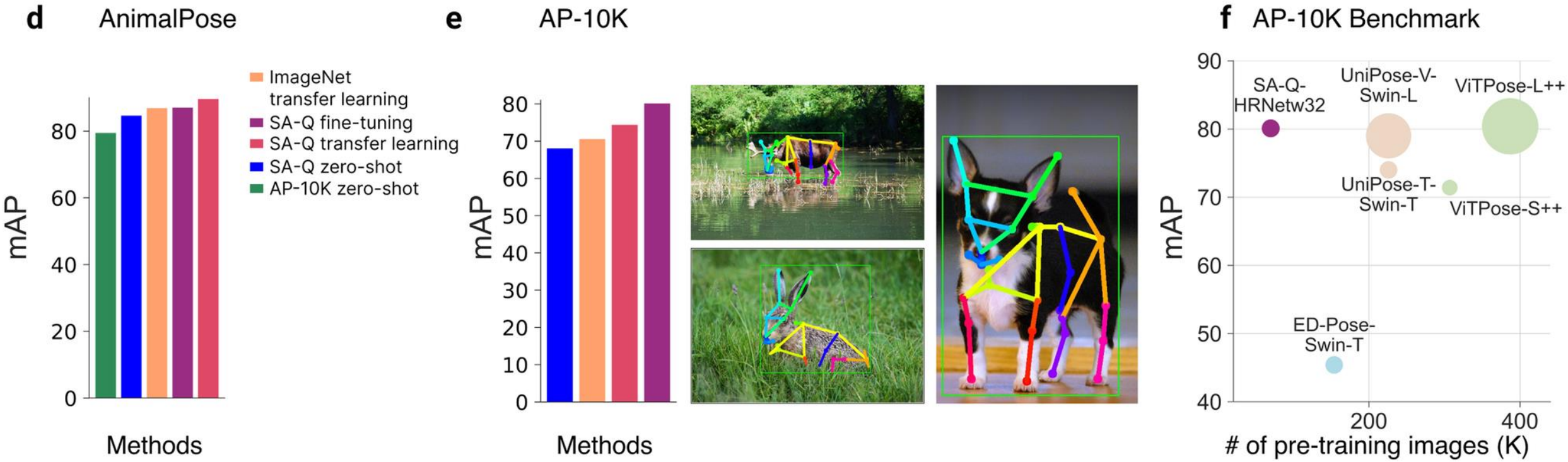




- Horse-10 Benchmark data

Openfield (DLC)

SuperAnimal models fine-tuned outperform ImageNet pretrained models on animal pose CV benchmarks



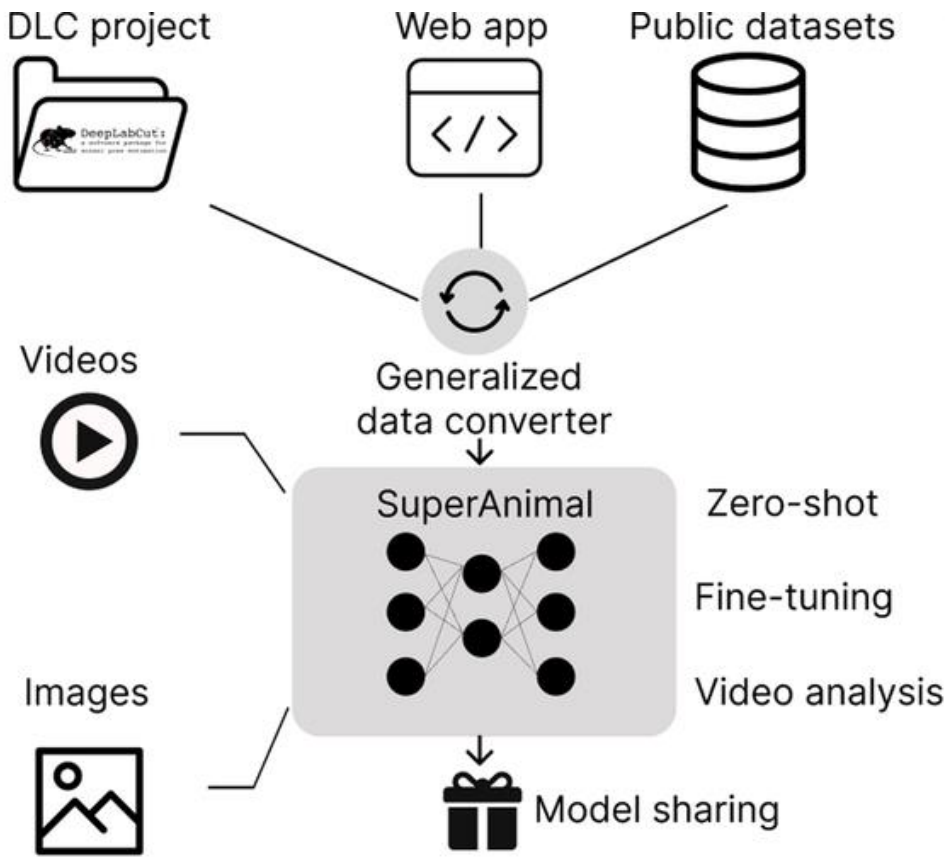
Challenge 2:

- We need more open source data (labeled data is even better, but can also use unlabeled data)

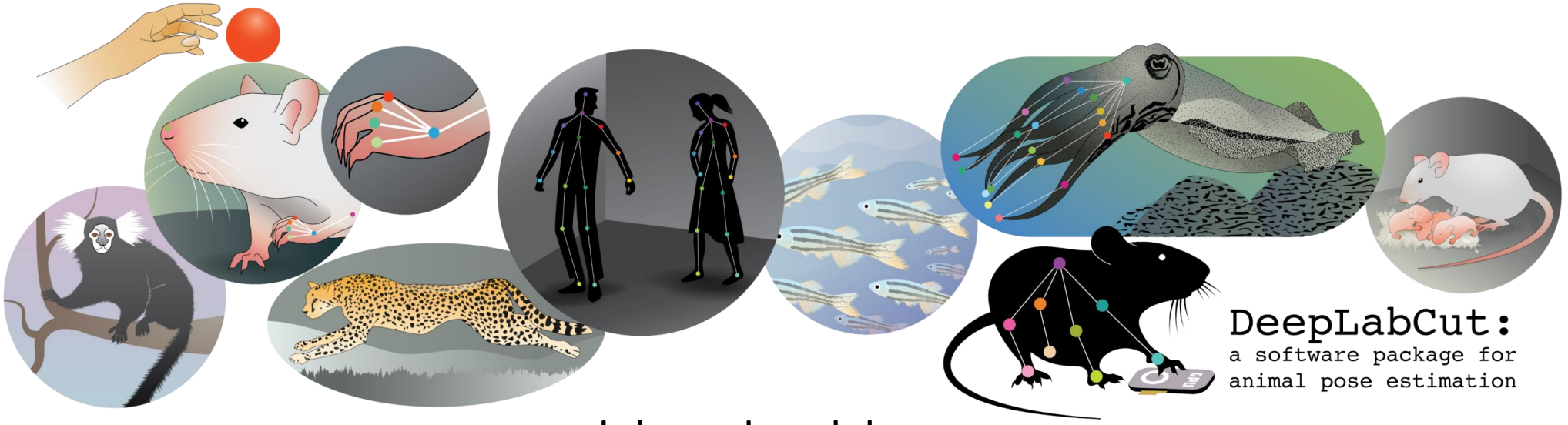
Solution: online crowd sourcing, robust generalized data converters/standards, web infrastructure

Challenge 3:

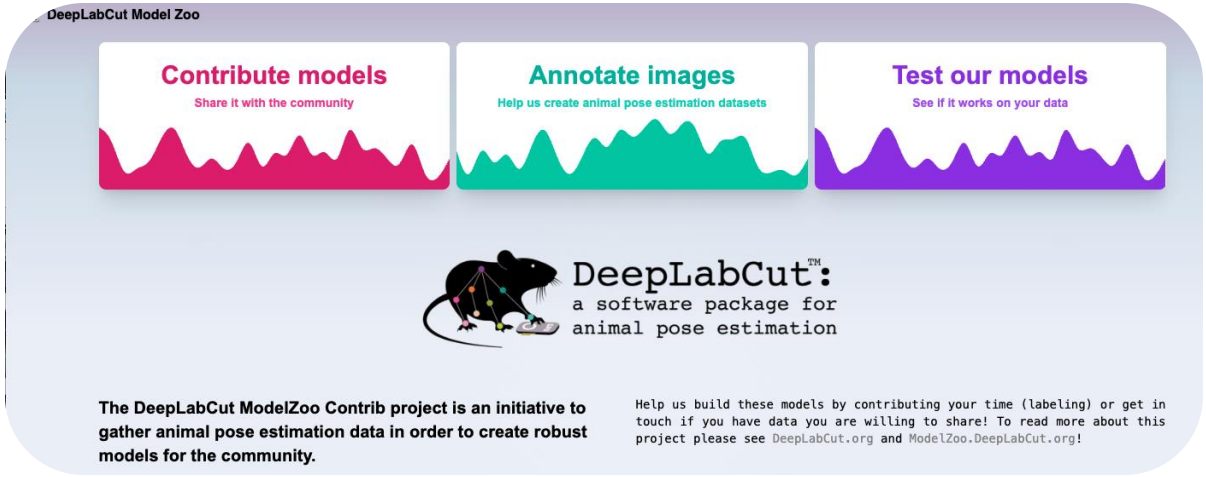
- People need to easily use such tools for zero-shot and for fine-tuning …

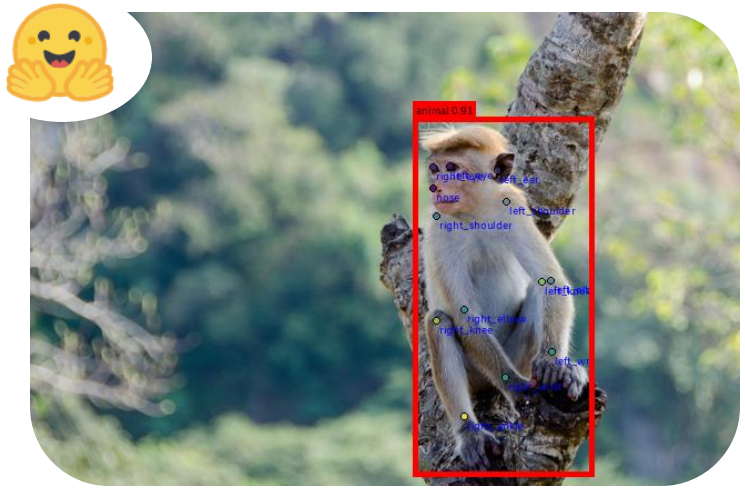


ModelZoo: model deployment & data curation



modelzoo.deeplabcut.org

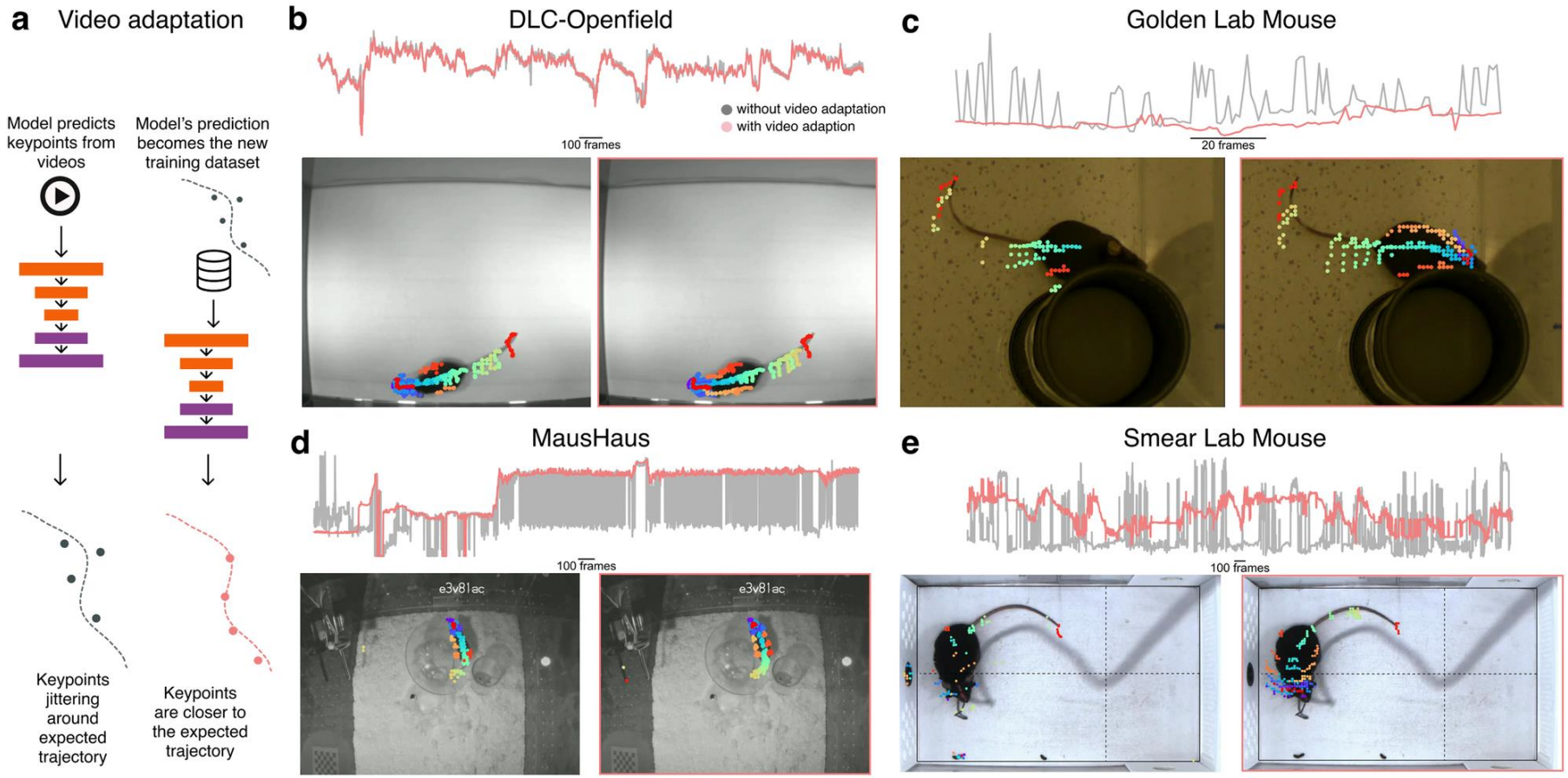




contrib.deeplabcut.org

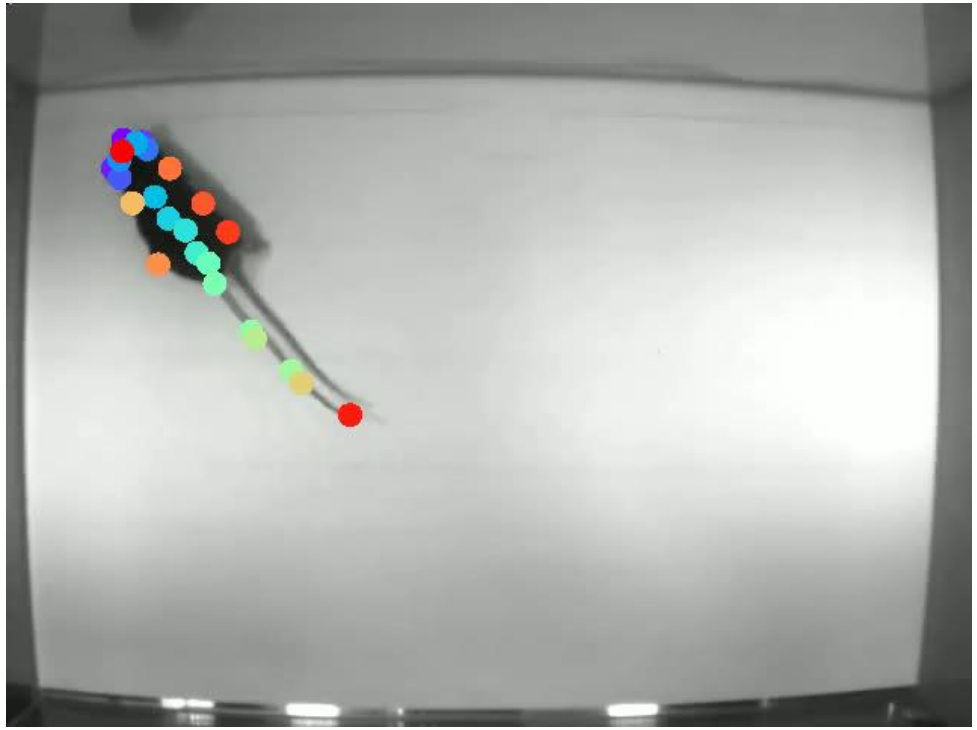
*https://huggingface.co/spaces/DeepLabCut

Real-time video Adaptation on your data

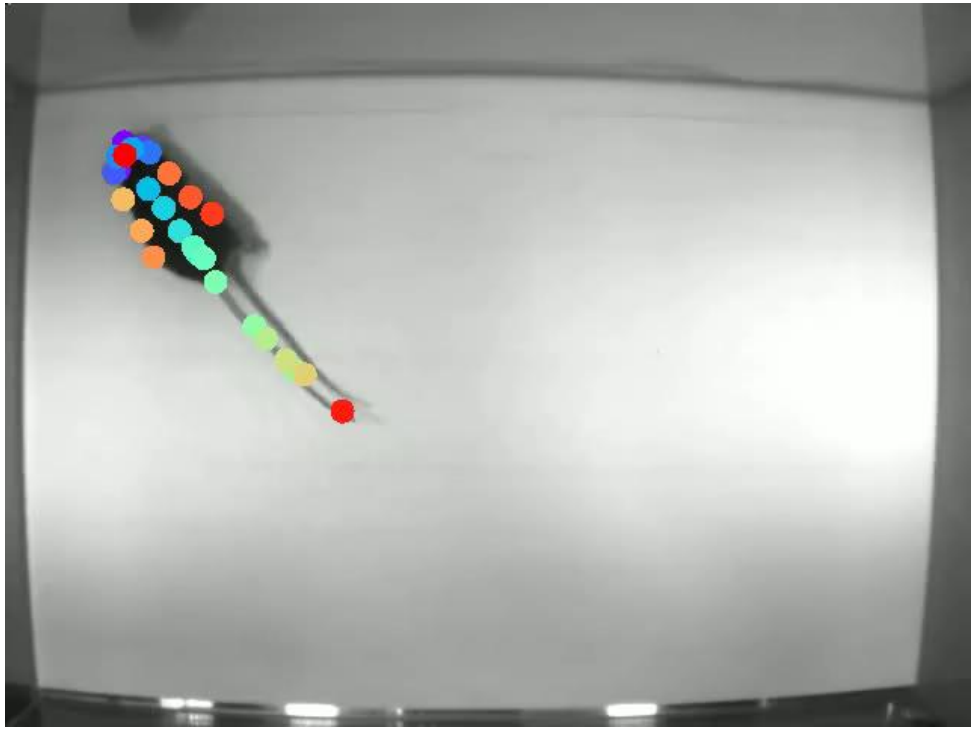


Rapid test-time video adaptation

Zero-shot video

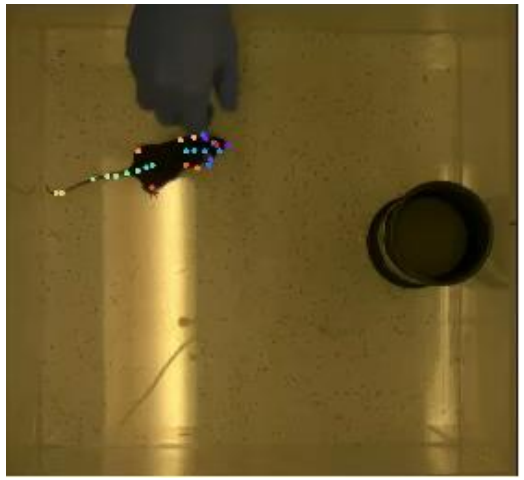


Zero-shot+ video adaptation



~1.3 real time for 1K iters (GPU)

Towards building a foundation model for animal pose estimation



Shaokai Ye et al. 2022, 2024

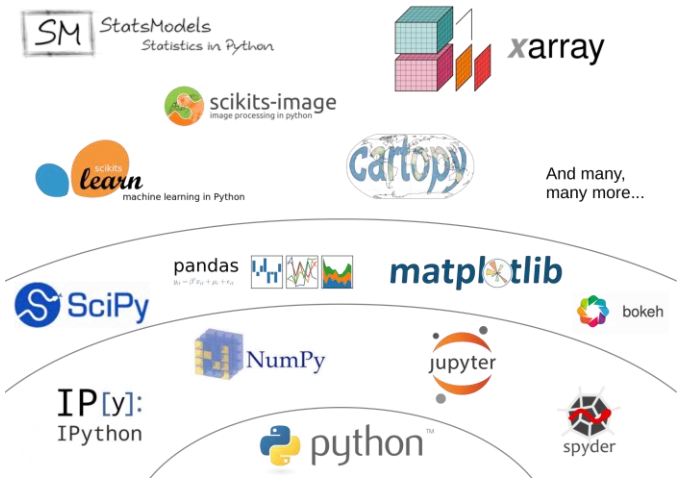
Advances:

- Zero training from scratch could be required (huge energy savings & time/compute!)
- Zero-shot inference, or only tens of images for rapid fine-tuning required
- *(networks: gradient masking, memory replay, semantic mapping)*
- Zero-shot video inference, or 1.3x video inference w/test time aug.
- Tops OOD pose benchmarks

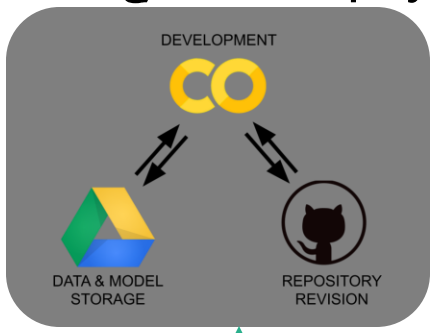
Still (more) challenges:

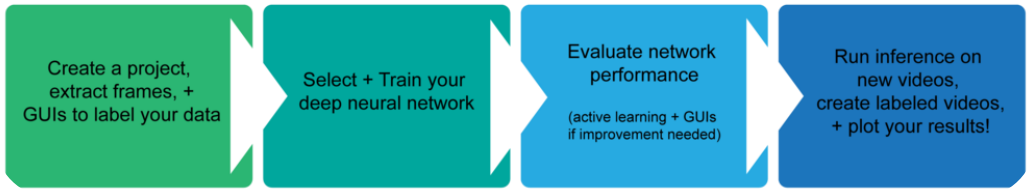
- TopView rodents & quadrupeds are not all animals in neuroscience
- Do we build centralized models, or groups build their own SuperAnimals?
- good data sharing practices // central resources?

Built on the open source python stack:

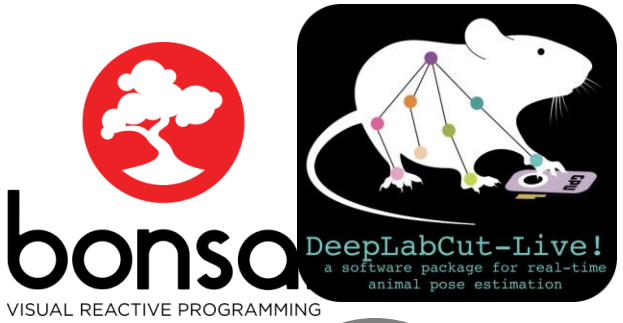


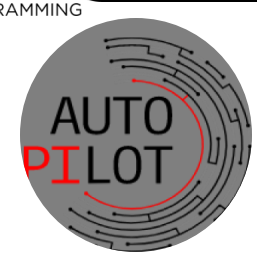
User testing/dev & deployment:





Real-time specific tools:

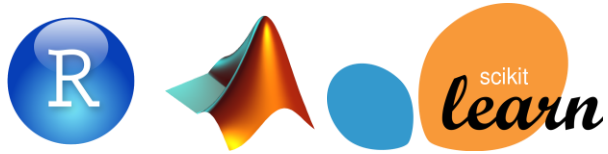




Computer Vision:

Larger scale pipeline computing:

Post- pose estimation tools:



Classifiers: SVMs, Random Forrest, ANNs
- B-SOID, ETH-DLC Analyzer, simba

Models: HMMs, decision-trees, ANNs

Ethograms: BORIS, BENTO, AmadeusGPT, Keypoint-MoSeq,

Clustering: CEBRA, MotionMapper, JAABA

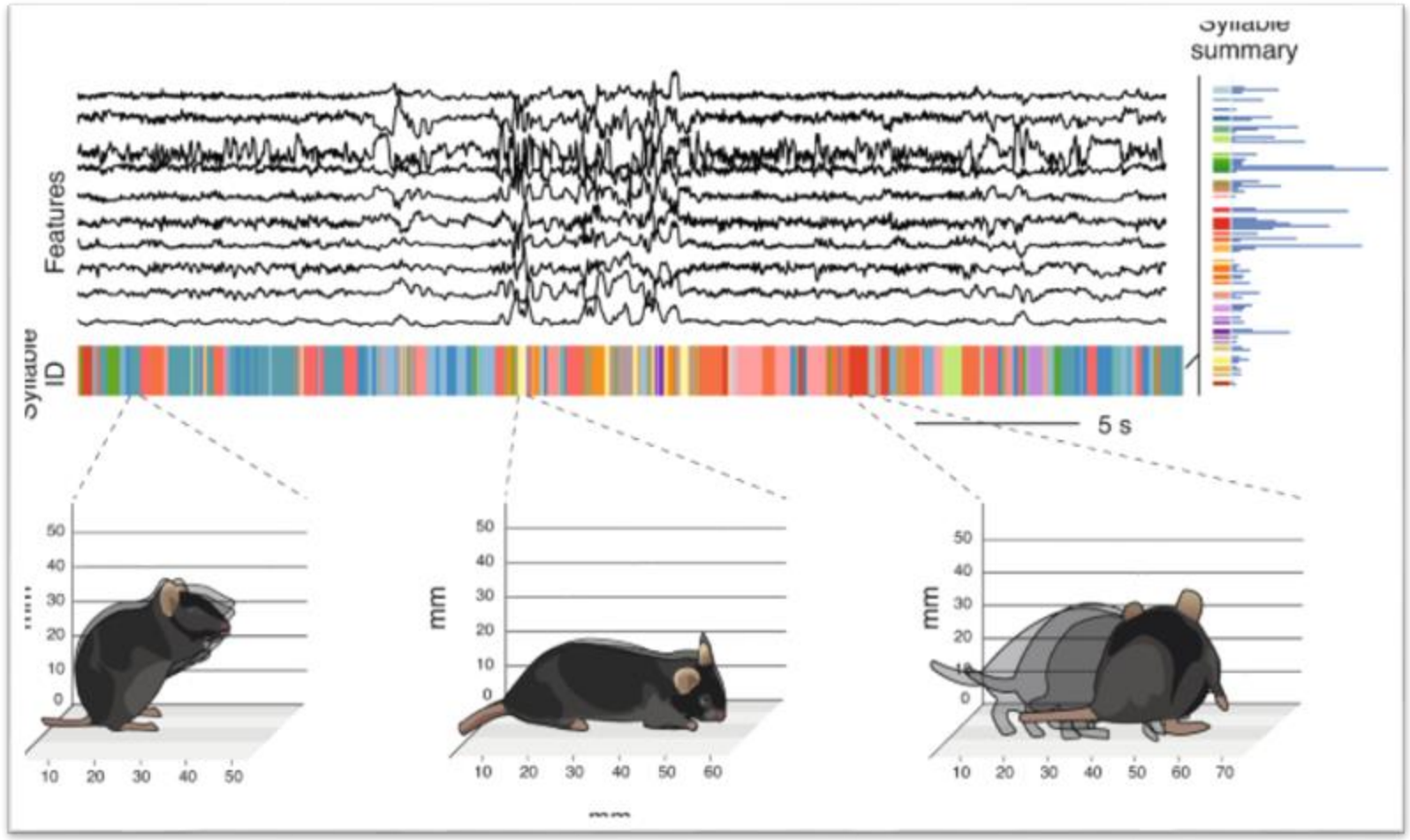
Motor analysis: DLC2Kinematics

Post- pose estimation tools

Classifiers: SVMs, Random Forrest, ANNs
- B-SOID, ETH-DLC Analyzer, simba

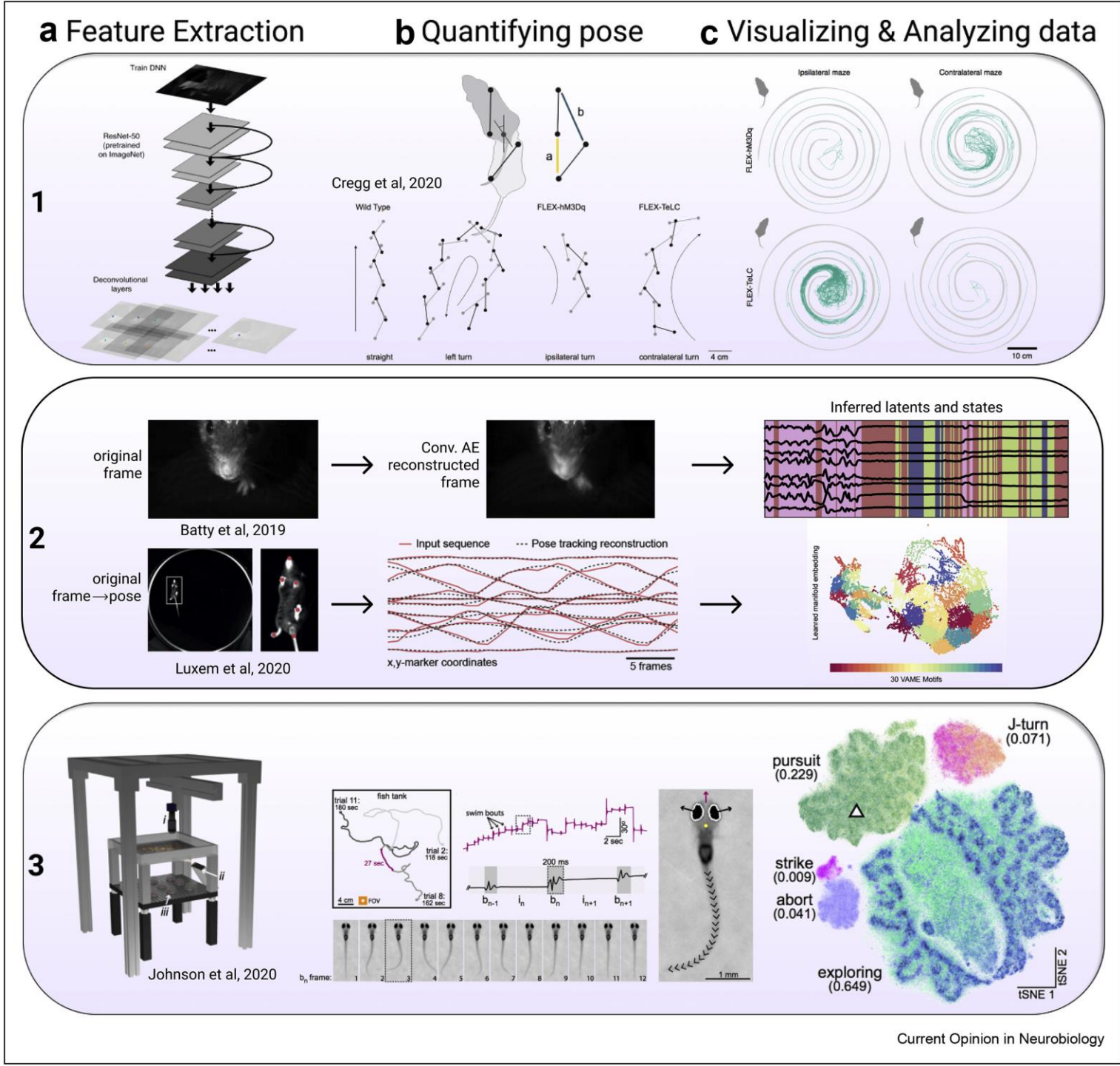
Models: HMMs, decision-trees, ANNs
Ethograms: BORIS, BENTO, AmadeusGPT, Keypoint-MoSeq,
Clustering: CEBRA, MotionMapper, JAABA
Motor analysis: DLC2Kinematics

- Keypoints are an excellent way to reduce the dimensionality of the video (from hundreds-thousands of pixels to key “pixels” over time
- How can we analyze this time-series data?



Behavioral analysis tools (many!)

- feature extraction (pose estimation)
- quantification (quality & cleaning)
- clustering, time series modeling, ethograms



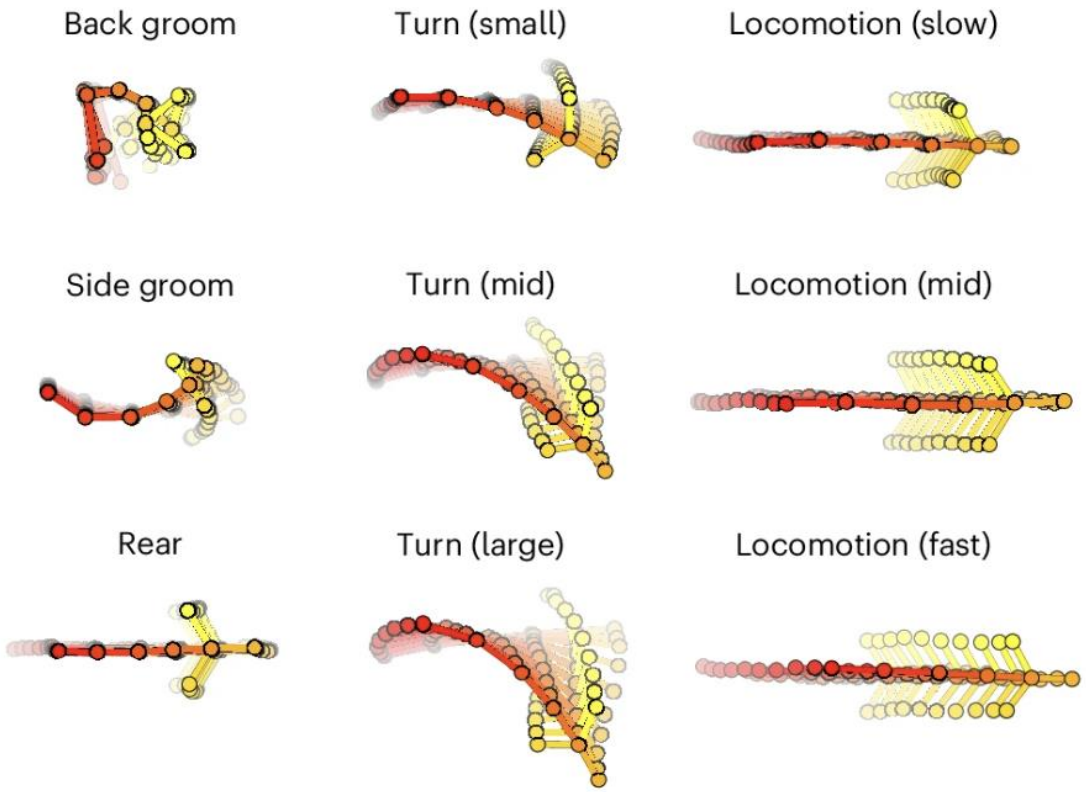
Behavioral analysis tools
Keypoint-MoSeq

Article | Open access | Published: 12 July 2024

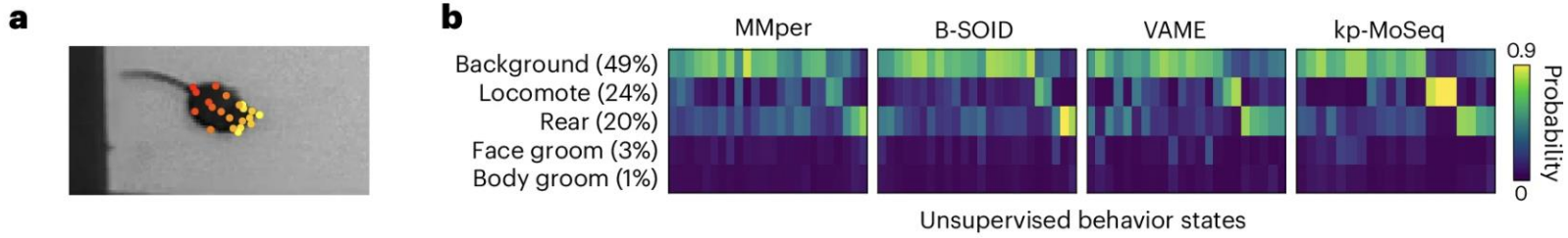
Keypoint-MoSeq: parsing behavior by linking point tracking to pose dynamics

Caleb Weinreb, Jonah E. Pearl, Sherry Lin, Mohammed Abdal Monium Osman, Libby Zhang, Sidharth Annapragada, Eli Conlin, Red Hoffmann, Sofia Makowska, Winthrop F. Gillis, Maya Jay, Shaokai Ye, Alexander Mathis, Mackenzie W. Mathis, Talmo Pereira, Scott W. Linderman & Sandeep Robert Datta

Nature Methods **21**, 1329–1339 (2024) | Cite this article

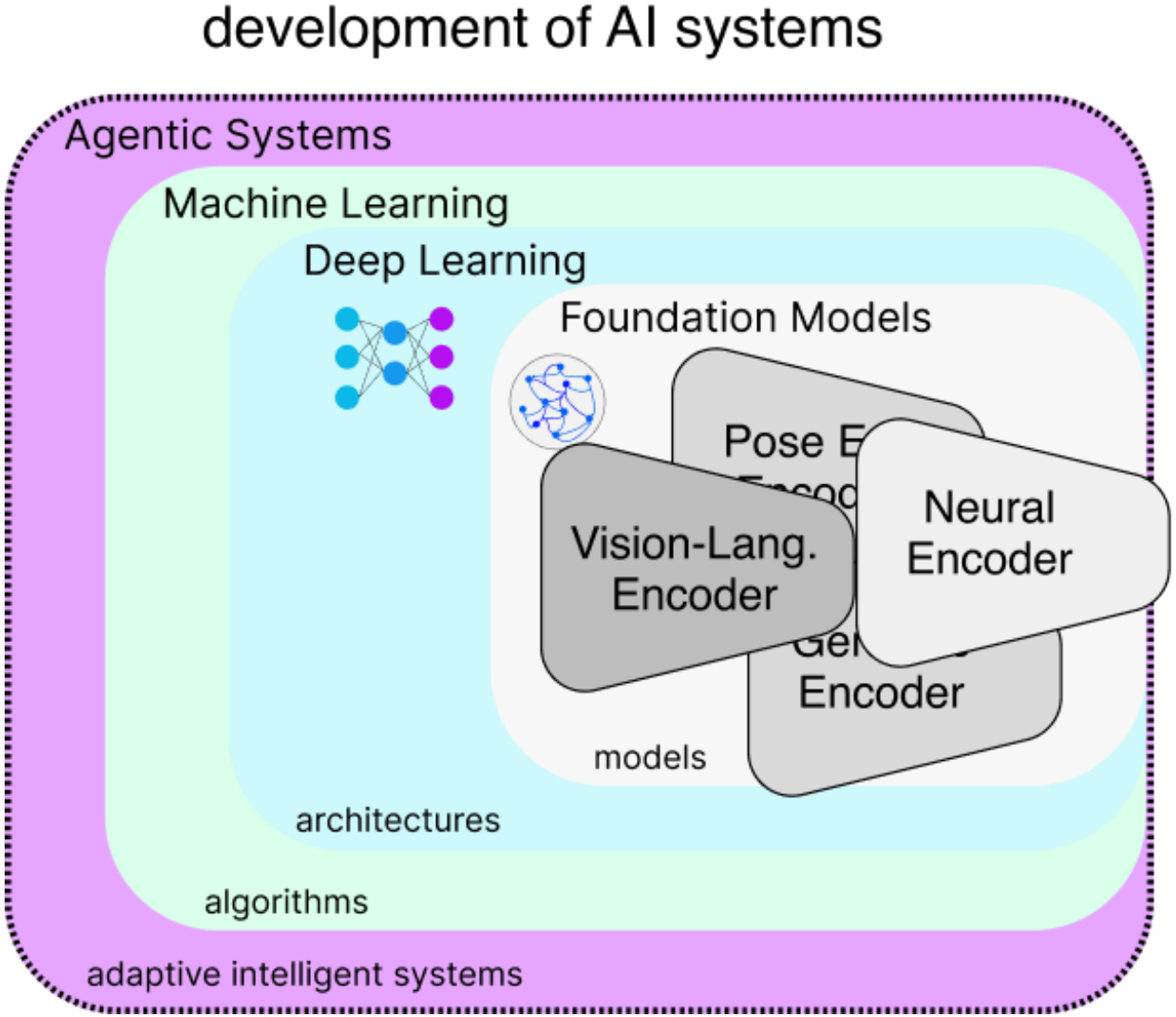


From: Keypoint-MoSeq: parsing behavior by linking point tracking to pose dynamics



Beyond deep learning, and foundation models ...

Agentic Systems

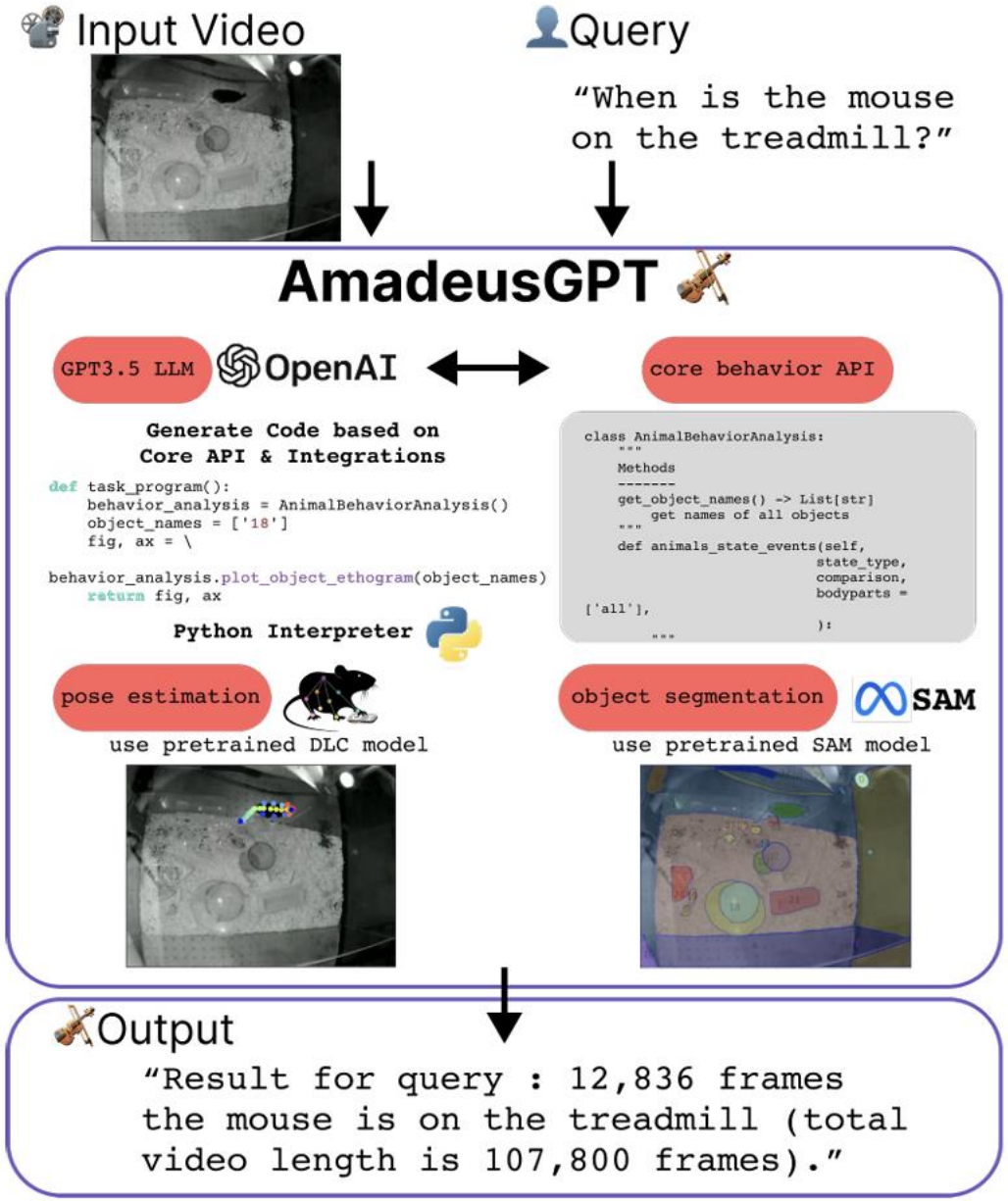


AmadeusGPT: a natural language interface for interactive animal behavioral analysis

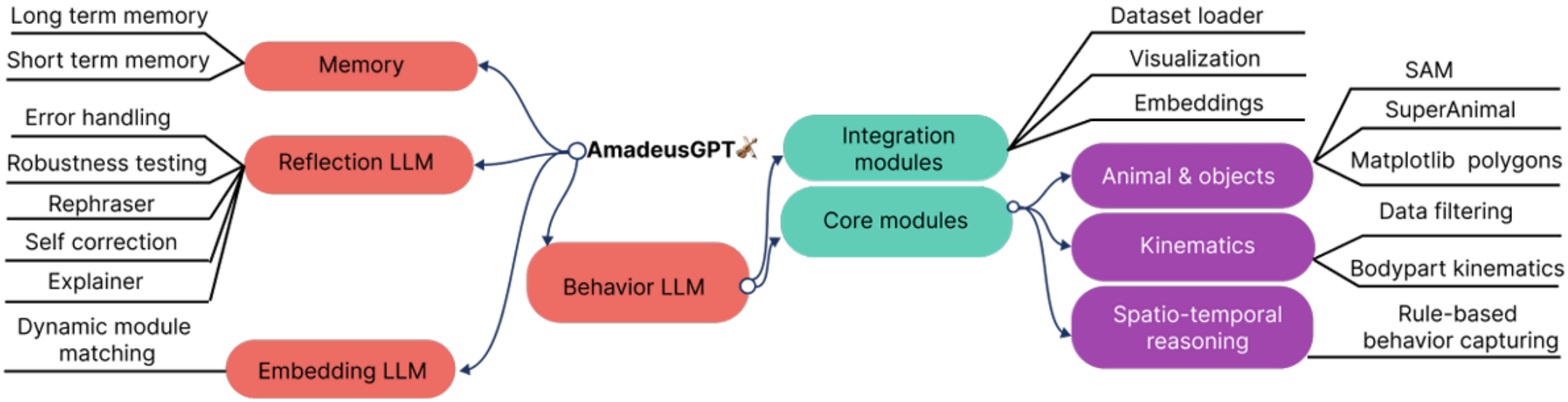
Shaokai Ye, Jessy Lauer, Mu Zhou, Alexander Mathis, Mackenzie W. Mathis NeurIPS 2023

Highlights:

- AmadeusGPT leverages LLMs, such as GPT3.5 or 4
- Its an “OS”: a systems architecture approach to combing LLMs for encoding, rephrasing, and explaining results
- Leverages SOTA models, such as SAM (MetaAI) and SuperAnimals (DeepLabCut)
- Can match human-level performance at quantifying animal behavior



System design & plugin system for AmadeusGPT



Measuring behavior with multi-modal models & natural language

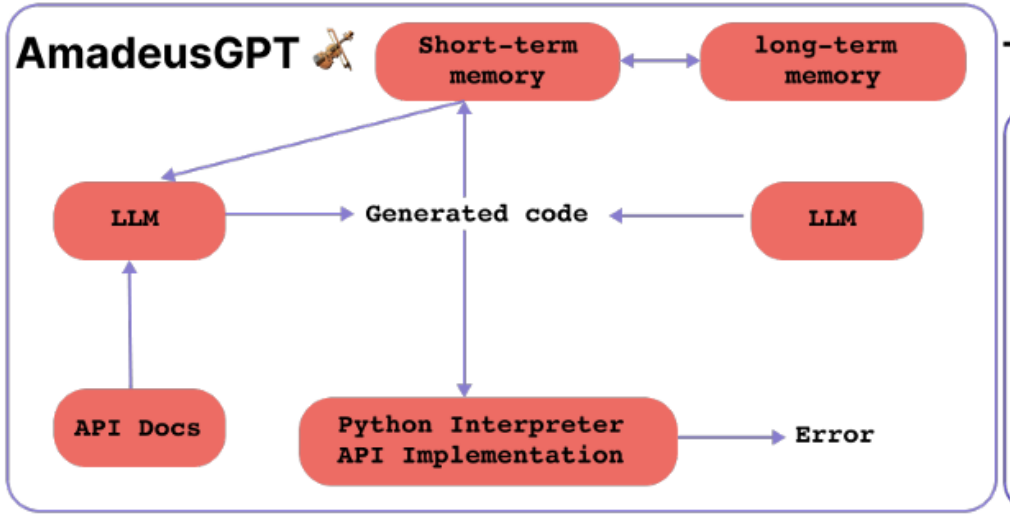
(a) Short questions leading to active refinement and decomposition into to multiple task programs

Queries →

Query 1
"When is the mouse on the treadmill?"

Query 2
"Plot the corresponding object ethogram"

Query 3
"Plot the trajectory of the center of animal"



→ runs computer vision models to extract pose and objects

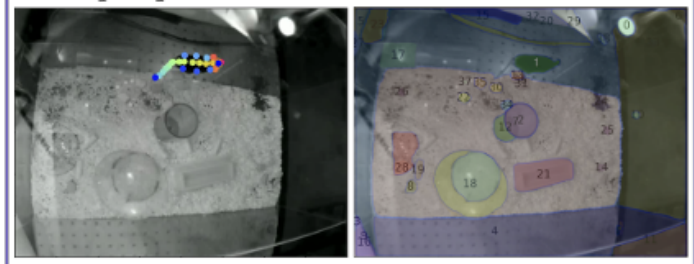
: User specifies Object 18 is the "treadmill"

Chat Output:

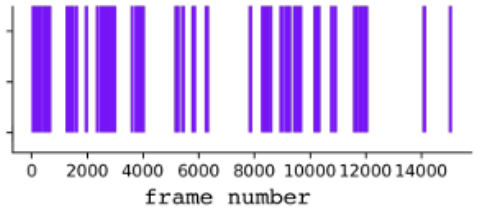
Result for query : 12,836 frames the mouse is on the treadmill (total video length is 107,999 frames).

Visual Outputs:

For query 1:



For query 2:
Ethogram for Object + Animal overlap



Code Outputs:

```
def task_program():
    behavior_analysis = AnimalBehaviorAnalysis()
    object_18_events = \
        behavior_analysis.animals_object_events('18',\
        'overlap', bodyparts=['all'])
    object_18_mask = \
        Event.events2onemask(object_18_events)
    frames_on_object_18 = np.sum(object_18_mask)
    return frames_on_object_18
```

For query 2:

```
def task_program():
    behavior_analysis = AnimalBehaviorAnalysis()
    object_names = ['18']
    fig, ax = behavior_analysis.plot_object_ethogram(object_names)
    return fig, ax
```

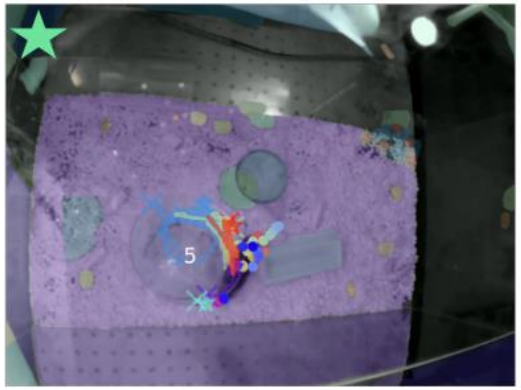
For query 3: **1.** The error message suggests that the bodypart "center" is not defined in the data. Check if "center" is a valid bodypart for the animal in the data. **2.** If "center" is not a valid bodypart, try using a different bodypart that is supported by the data. **3.** Once you have identified a valid bodypart, update the task_program() function to use that bodypart instead of "center".

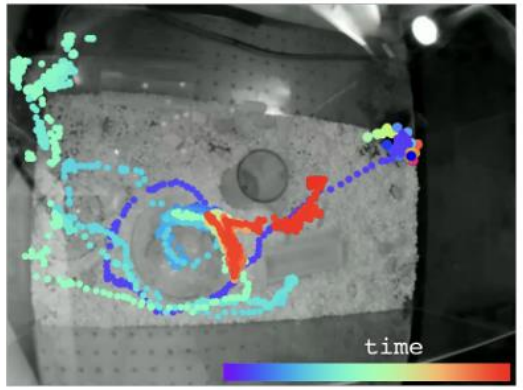
AmadeusGPT benchmarked on several common scenarios in neuroscience

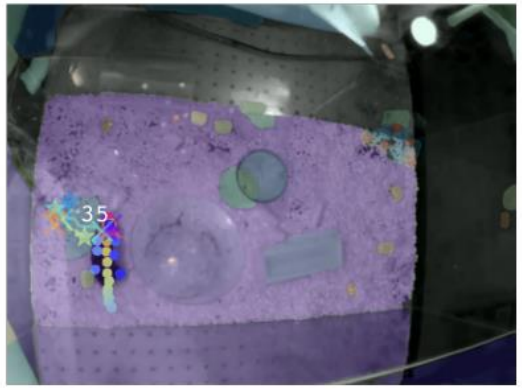
★ 👤: Query:
"When is the animal on the treadmill, which is object 5?"

👤: Query:
"Plot the trajectory of the animal."

👤: Query:
"When is the animal close to the object 35, if I define close as less than 50 pixels?"

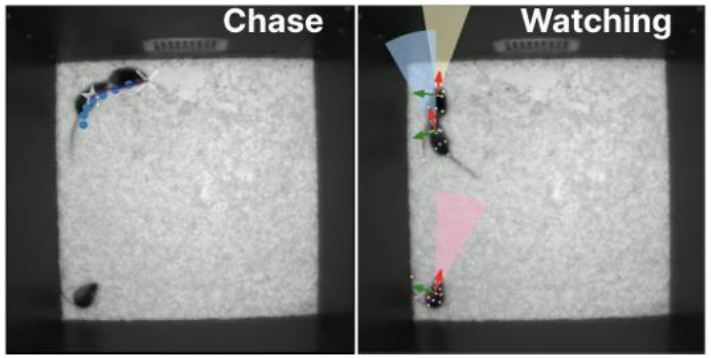






★
```
def task_program():
    behavior_analysis = AnimalBehaviorAnalysis()
    overlap_object_5_events = behavior_analysis.animals_object_events('5', 'overlap', bodyparts=['all'])
    fig, ax = behavior_analysis.plot_trajectory(bodyparts=['all'], events=overlap_object_5_events)
    return fig, ax
```

(c) MABE Challenge dataset



Official definition

Chase: Mice are moving above 15 cm/sec, with closest points less than 5 cm apart, and angular deviation between mice is less than 30 degrees, for at least 80% of frames within at least one second. Merge bouts less than 0.5 seconds apart.

Watching: Mice are more than 5 cm apart but less than 20 cm apart, and gaze offset of one mouse is less than 15 degrees from body of other mouse, for a minimum duration of 3 seconds. Merge bouts less than 0.5 seconds apart.

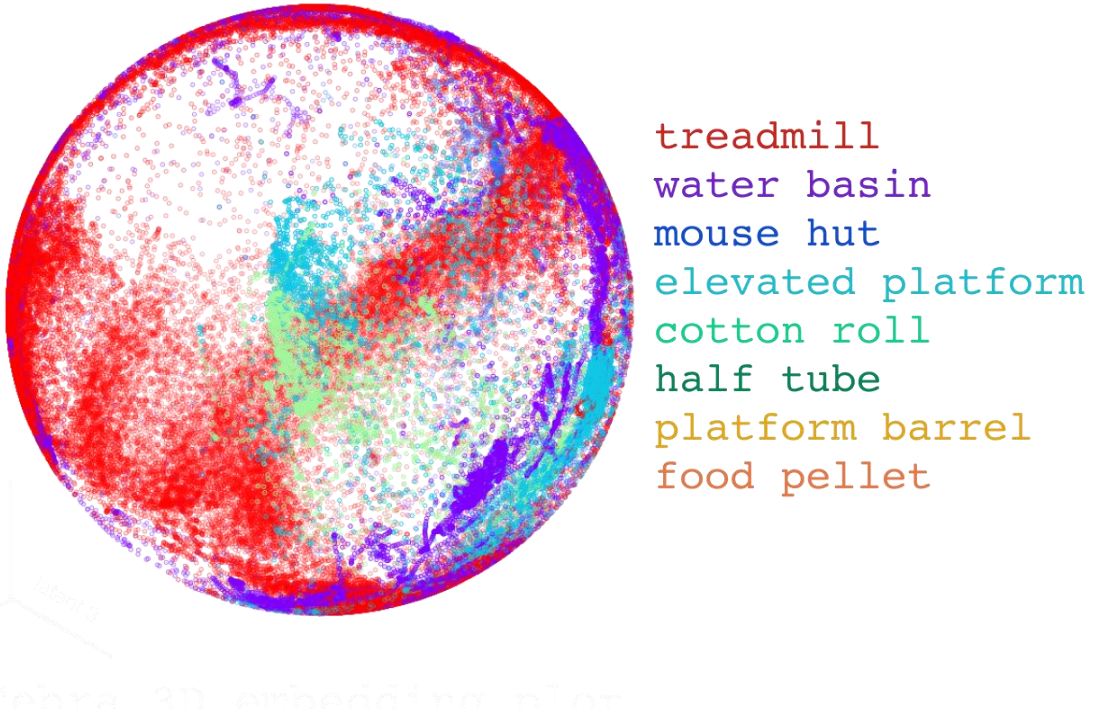
👤 Our prompt

"Define <|chases|> as a social behavior where closest distance between this animal and other animals is less than 40 pixels and the angle between this and other animals have to be less than 30 and this animal has to travel faster than 2 pixels. When do chases happen?"

"Define <|watching|> as a social behavior where distance between animals is less than 260 pixels and larger than 50 and head angle between animals is less than 15. The smooth_window_size is 15. When does watching happen?"

AmadeusGPT:
SuperAnimals, SAM & CEBRA for behavioral analysis

Annotated MausHaus masks



Cebra 3D embedding plot

- Ye et al. Neurips 2023

Left click to draw a polygon. Right click to confirm the drawing. After that, click finish drawing button. Refresh the page if you need a new ROI Or if the ROI canvas does not display

Share

AmadeusGPT

Mathis Laboratory of adaptive motor control | EPFL

Elevated plus maze (EPM) is a widely used behavioral test. The mouse is put on an elevated platform with two open arms (without walls) and two closed arms (with walls). In this example we used a video from https://www.nature.com/articles/s41386-020-0776-y.

- ⬅️ On the left you can see the video data auto-tracked with DeepLabCut and keypoint names (below). You can also draw ROIs to ask questions to AmadeusGPT about the ROIs. You can drag the divider between the panels to increase the video/image size.
- We suggest you start by clicking 'Generate Response' to our demo queries.
- Ask additional questions in the chatbox at the top of the page. Note that you need to scroll down to see the outputs!
- Here are some example queries you might consider: 'The <|open arm|> is the ROI0. How much time does the mouse spend in the open arm?' (NOTE here you can re-draw an ROI0 if you want. Be sure to click 'finish drawing') | 'Define head_dips as a behavior where the mouse's mouse_center and neck are in ROI0 which is open arm while head_midpoint is outside ROI1 which is the cross-shape area. When does head_dips happen and what is the

Streamlit python™ jupyter

https://github.com/AdaptiveMotorControlLab/AmadeusGPT

AmadeusGPT | Mathis Laboratory of adaptive intelligence | EPFL

downloads 3k | downloads/month 641 | pypi package 0.0.2 | Star 192

We turn natural language descriptions of behaviors into machine-executable code.

Installation | Home Page | News | Reporting Issues | Discussions!

Summary

- there are many behaviors people use in systems neuroscience, and therefore they require custom solutions to measure behavior
- pose estimation is the computer vision task of measuring the geometric configuration of keypoints (joints)
- to build high performance models, transfer learning is powerful approach transfer learning is the ability to take a pretrained encoder model and use in a downstream task (i.e. ImageNet or SuperAnimal backbones)
- CNNs are a standard model for this task setting, but transformers are also used
- primer on basics of convolutions and decoders
- “out of distribution” data is very common in neuroscience, so we need robust solutions
- data is also sparse, so we need to get creative and train models with disjoint data; take-home (1) is that even in systems neuro you need to innovate in other areas; (2) technically, this meant masking gradients and video adaptation algorithms
- good algorithms are not enough; in science you need to make your code usable!
- behavioral analysis is a rapidly growing area in neuroscience // many options exist
- towards the future of behavioral analysis: LLMs as “operating systems” for analysis