BOZEK LAB

CMMC
CENTER FOR
MOLECULARE MEDICINE
COLOGNE

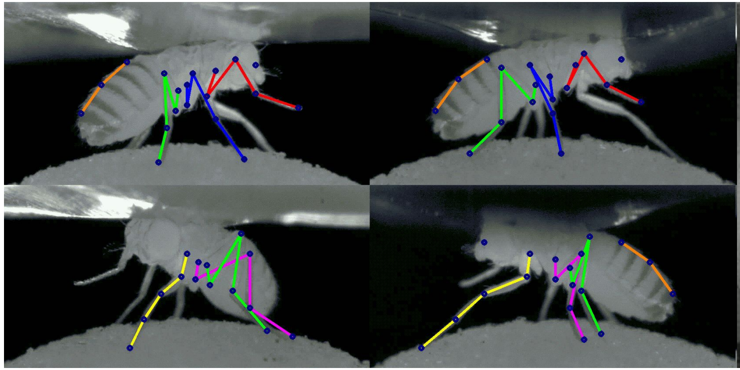
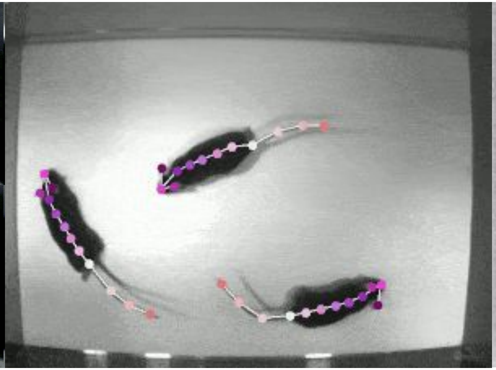
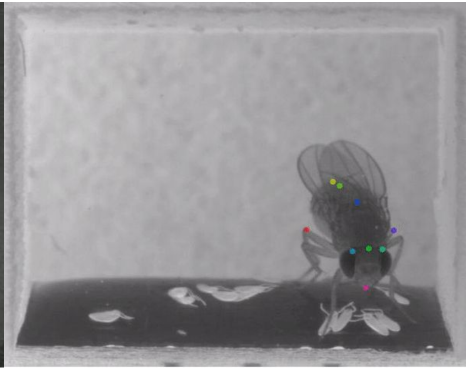
UNIKLINIK
KÖLN

FUNDED BY
Ministry of Culture and Science
of the State of
North Rhine-Westphalia

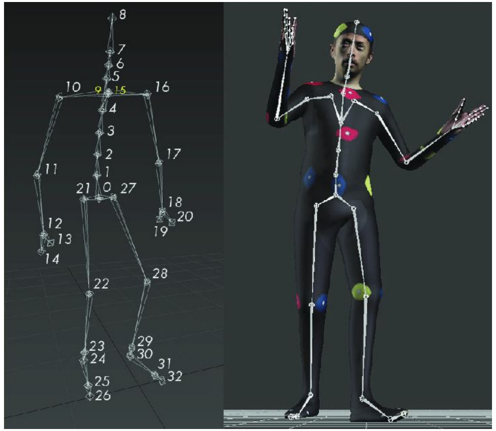
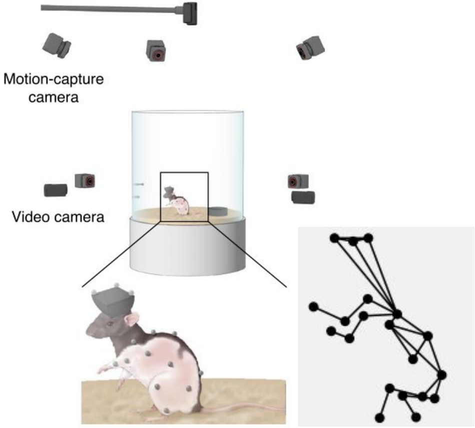
Uncertainty aware SSL on multi-dimensional time series for animal behavior

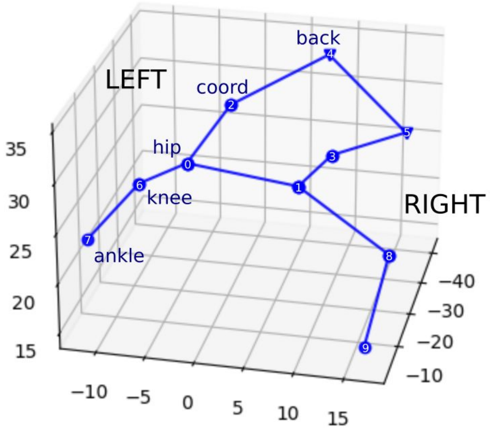
France Rose, Ph.D.
Data Science of Bioimages, Prof. Bozek
University of Cologne, Germany

Video Pose Estimation

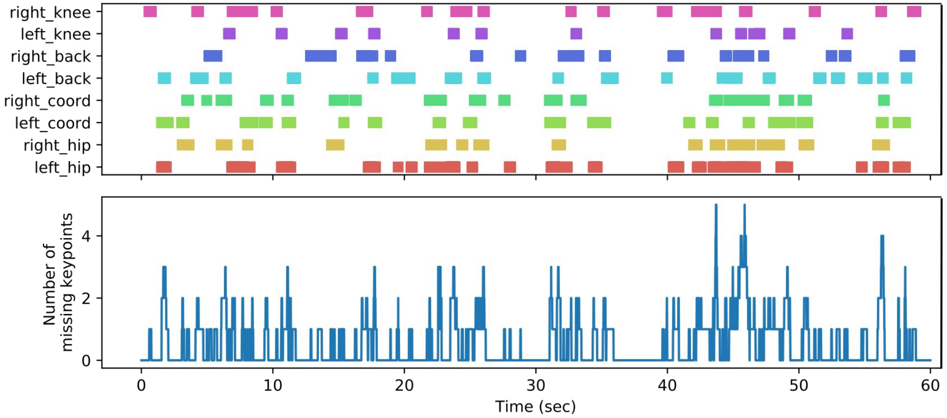




Motion Capture Systems

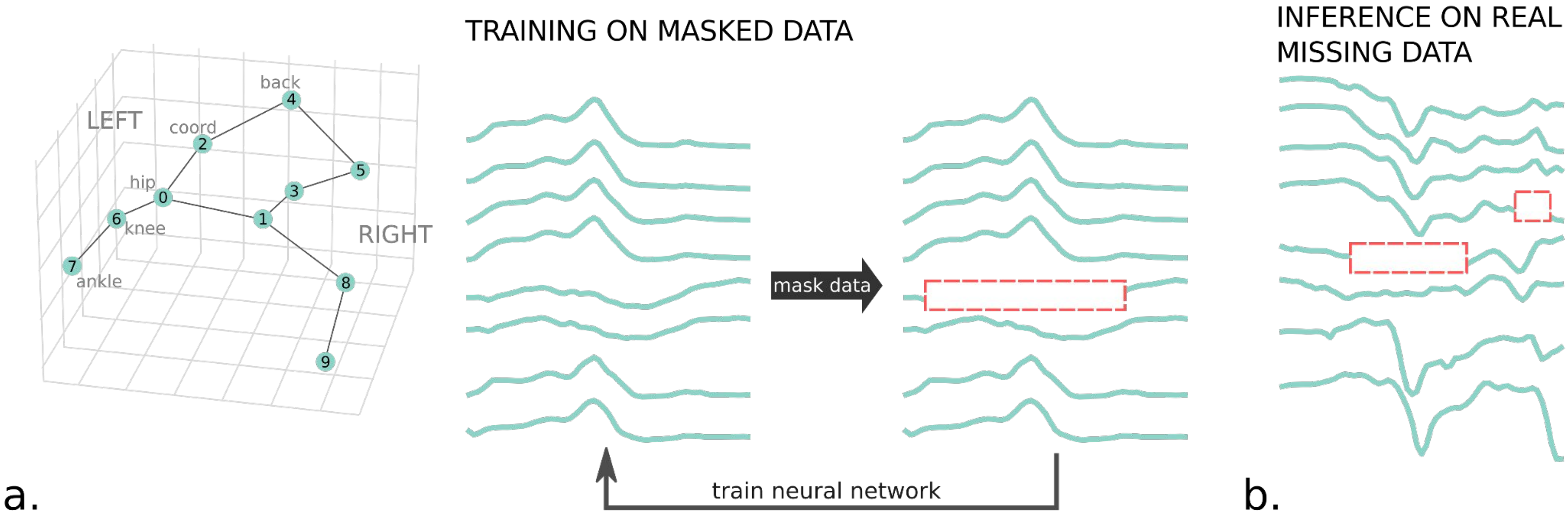


Günel et al. ELife 2019. Mathis et al. *Nat. Neuro.* 2018. Dunn et al. Nat. Met. 2021. Ignatowska-Jankowska et al. BioRxiv 2023.



- Missing keypoints in behavior analysis are dropped
- Existing imputation methods for general time series
- But no method developed or tested at large scale on skeleton data

Unsupervised training and testing scheme

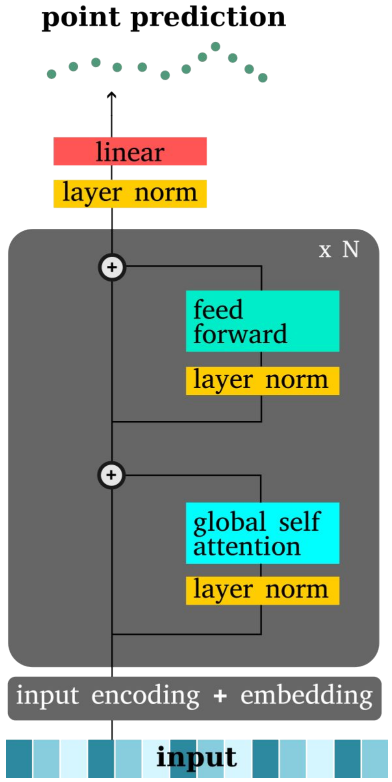


Tested algorithms

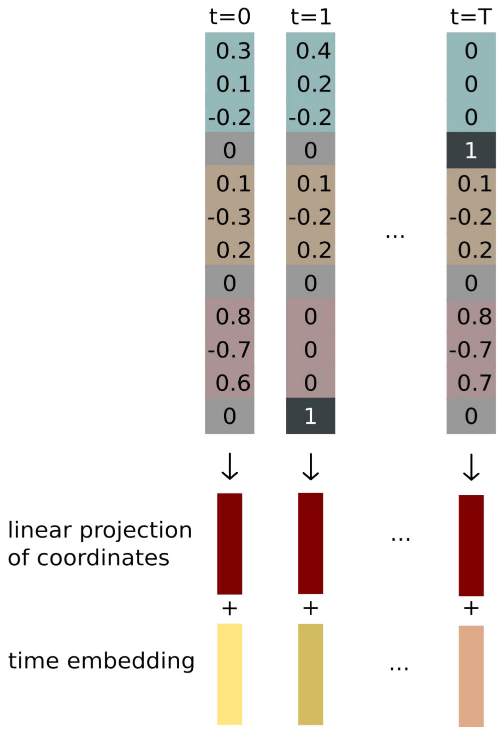
- Linear interpolation (Baseline)
- 5 different Neural Networks
  - Recurrent neural network (GRU)
  - Temporal Convolutional Network (TCN)
  - Graph Convolutional Networks
    - Spatio-temporal GCN
    - Space-Time-Separable GCN
  - Custom Transformer (DISK)

Yan et al. AAAI 2018. Sofianos et al. ICCV 2021. Zerveas et al. SIGKDD 2021. Grigsby et al. arXiv 2021.

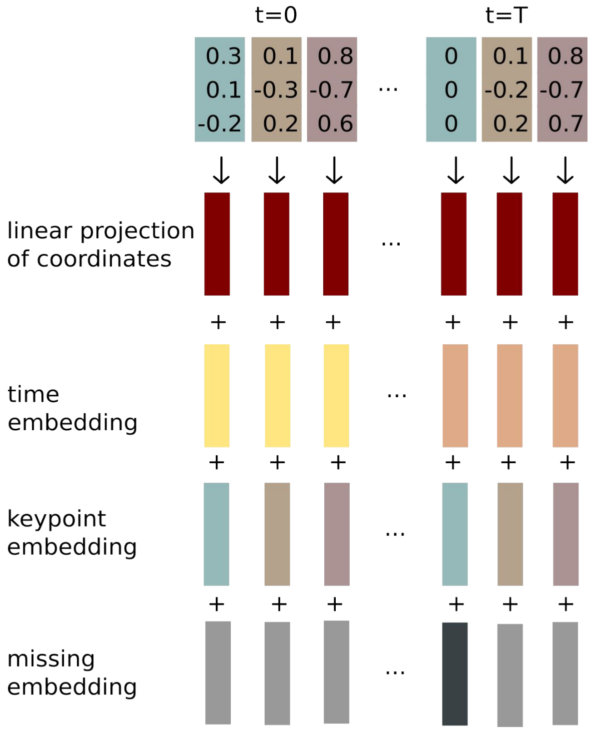
DISK architecture



Usual projection

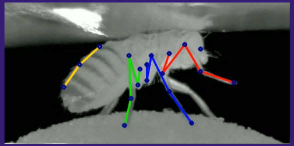


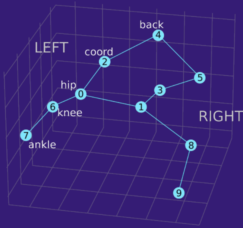
“Flattened” projection

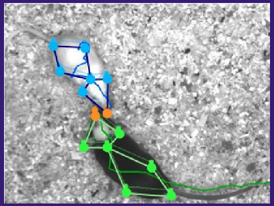


Datasets

- 7 datasets
- 5 species
- 2D and 3D
- 1 to 2 animals

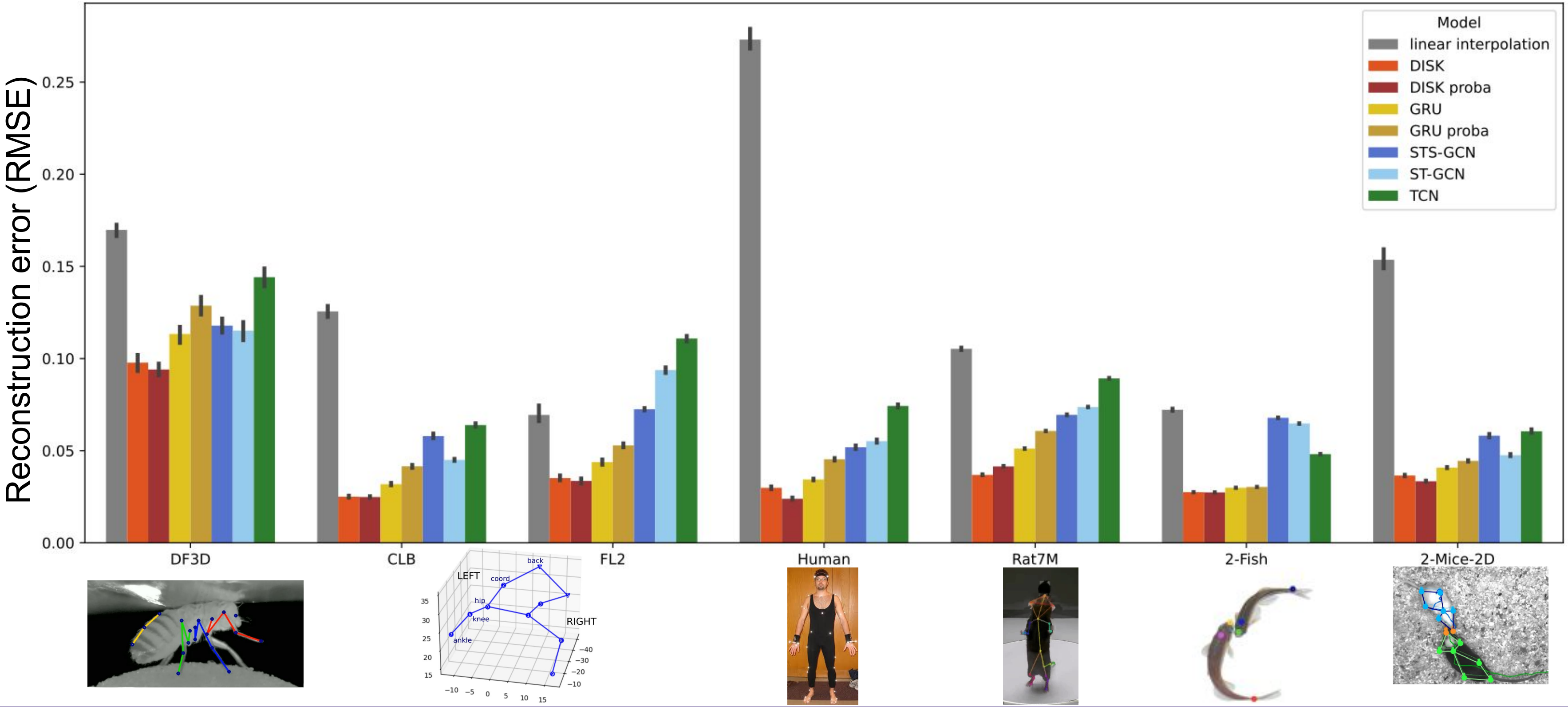




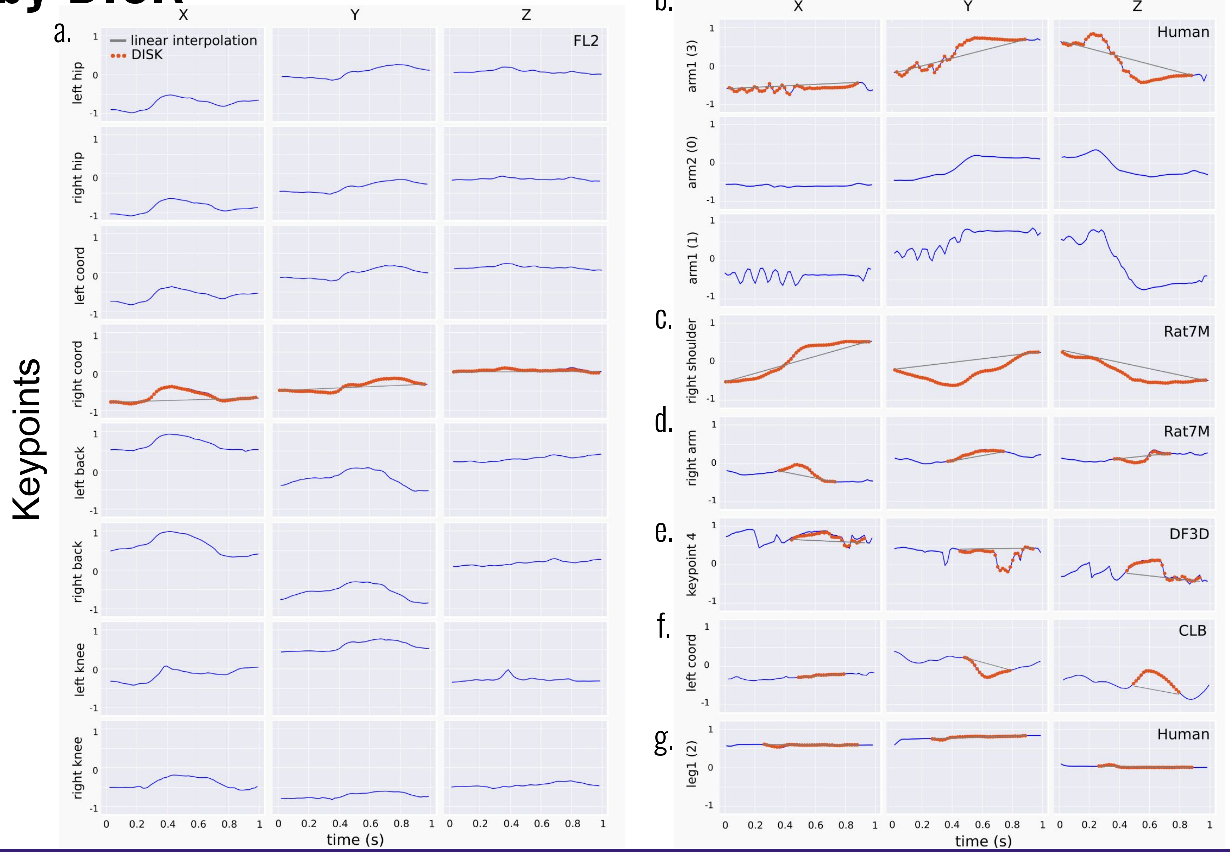


O'Shaughnessy et al. bioRxiv 2024. Dunn et al. Nat. Met. 2021. Günel et al. eLife 2019. Ignatowska-Jankoska et al. bioRxiv 2023. Sun et al. NeurIPS 2021. CMU MoCap database.

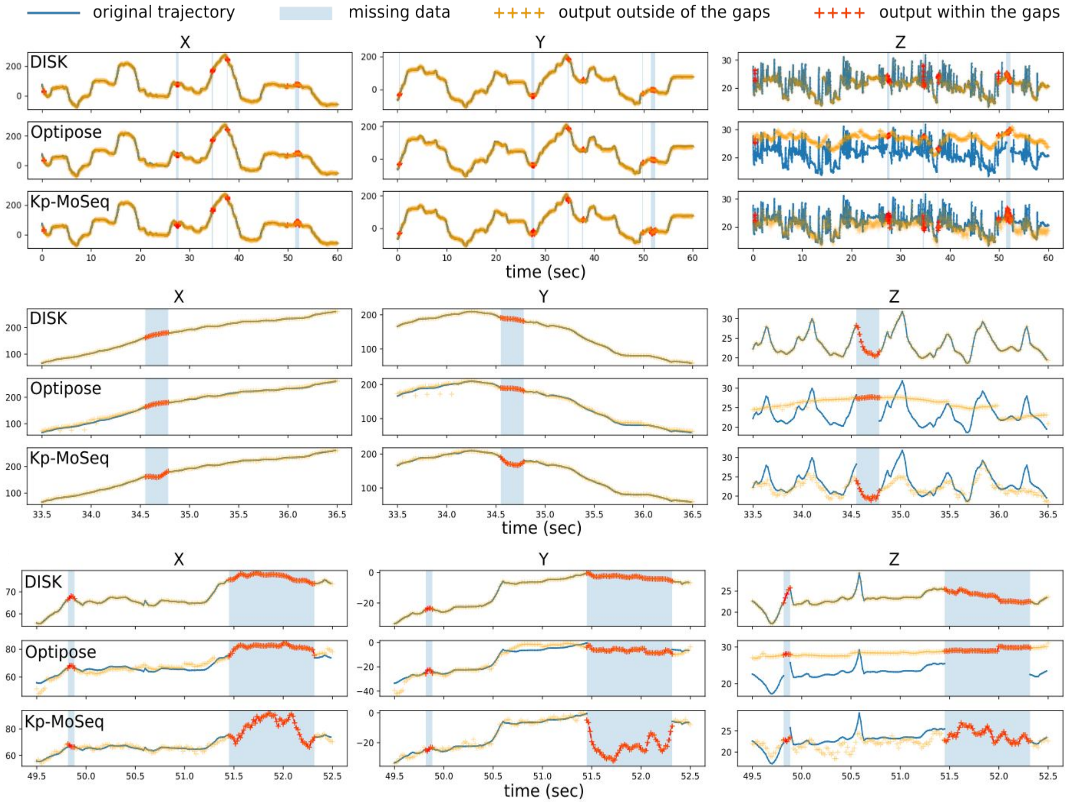
Performance on the 7 datasets

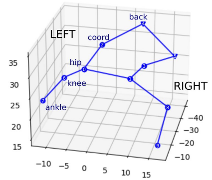


Imputation by DISK



Comparison with methods used in behavior analysis



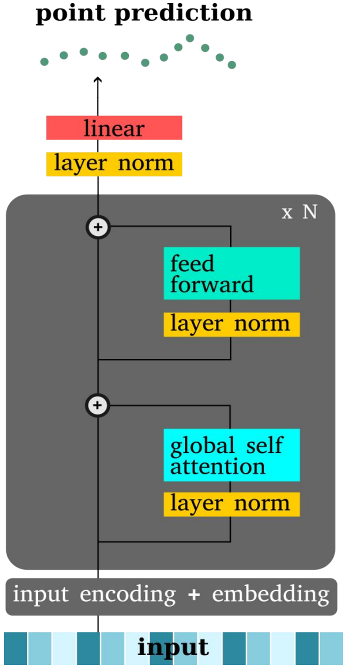


Real gaps, no ground truth

Trusting a black box model?

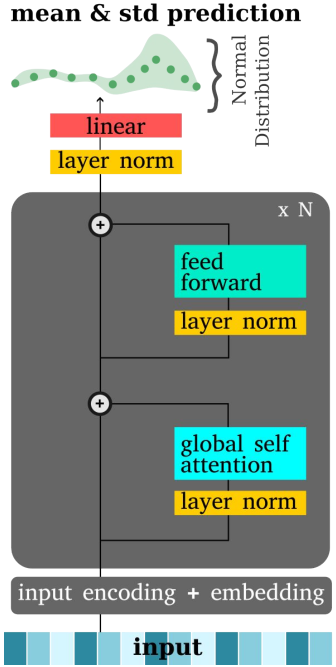
- Estimate the quality of the imputation
- Control the quality of the output dataset

Adding a probabilistic head



Output: $X_{\{k,t\}} \in \mathbb{R}^3$

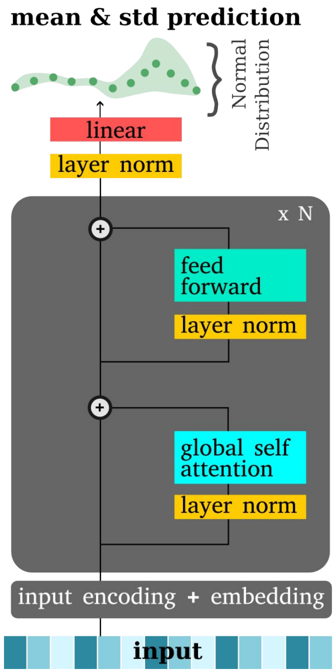
L1-loss

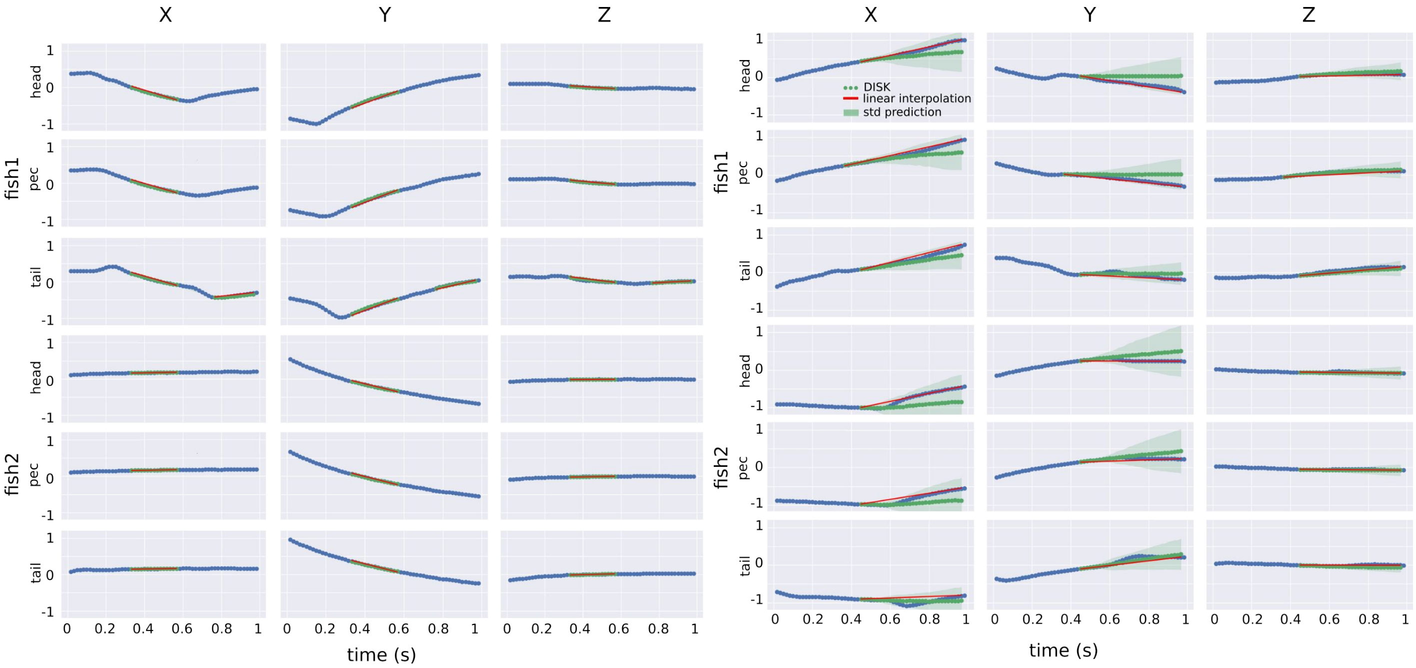


Output: $(\mu \in \mathbb{R}^3, \sigma \in \mathbb{R}^3)_{\{k,t\}}$

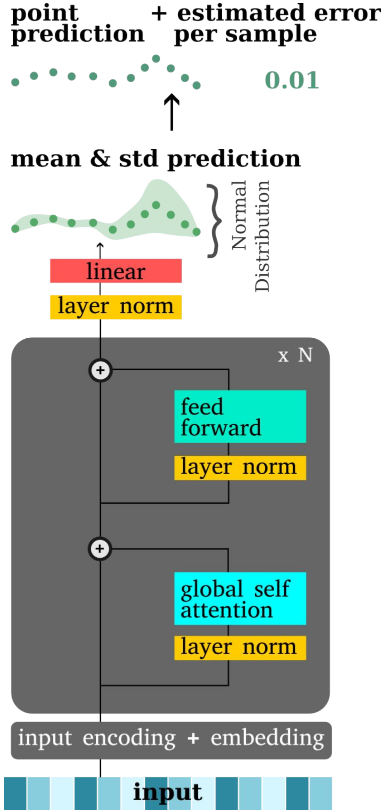
Negative log-likelihood loss:
$\sum_{\{k,t\}} ½(X_{GT} - \mu) / \sigma^2 - \log(\sigma)$

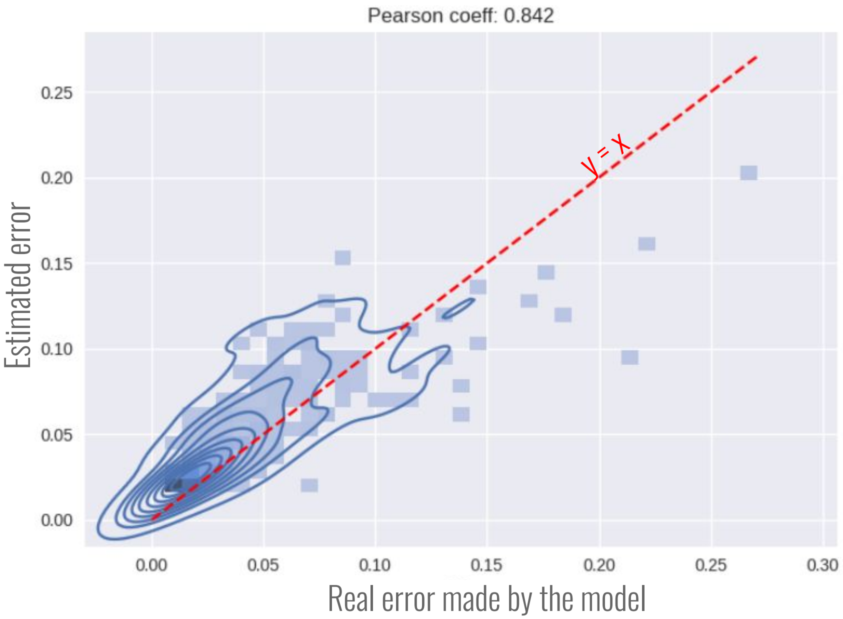
Estimated error on the imputed samples





Estimated error on the imputed samples





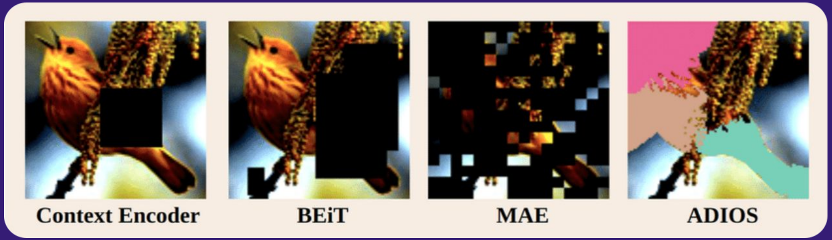
Uncertainty aware models

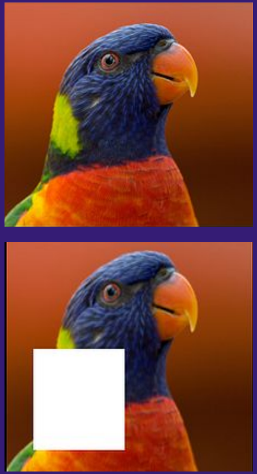
- Other tested approaches:
  - Ensemble
  - Variants of dropout
  - Additional branch to predict the estimated error
- Lower Pearson correlation, uncalibrated estimated error wrt real error
- Probabilistic head works better with transformer than GRU

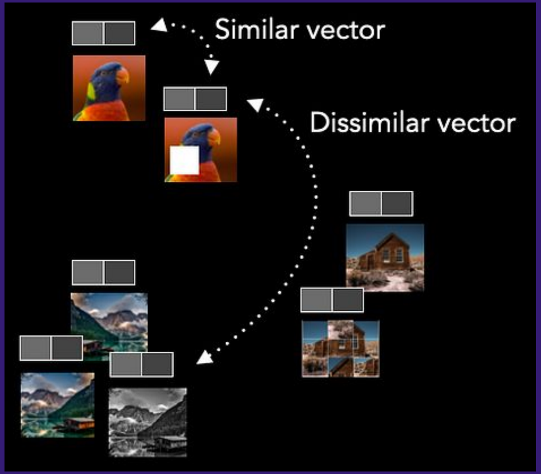
What does DISK learn?

Imputation = masking task in Self-Supervised Learning

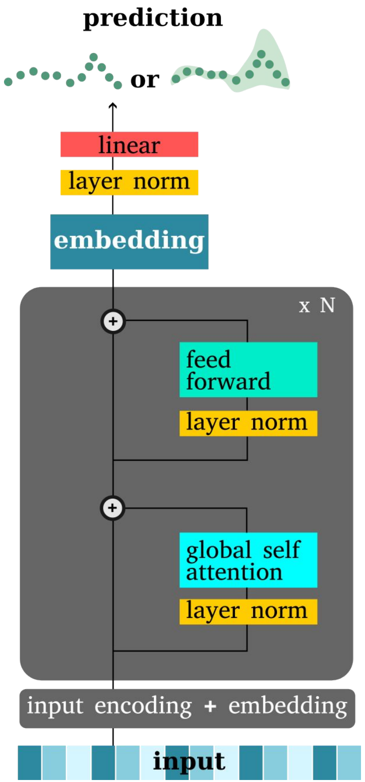
Masked Image Models



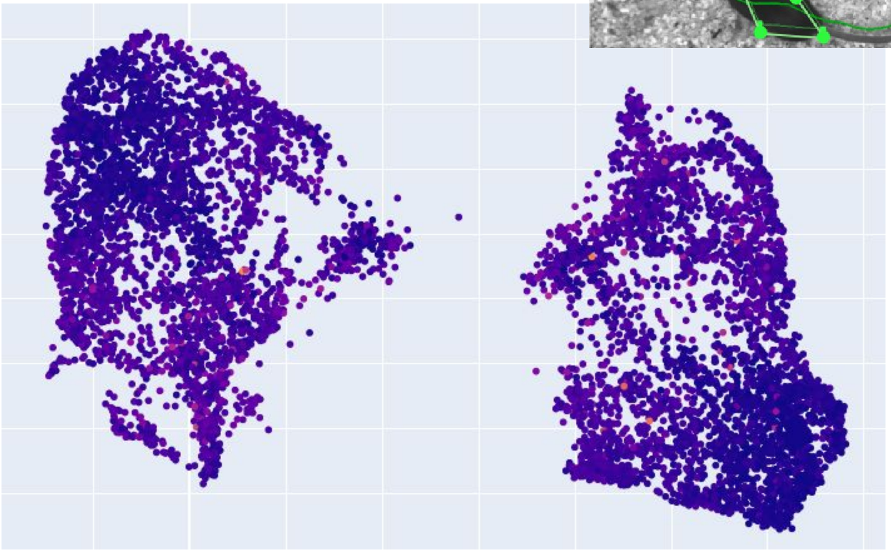


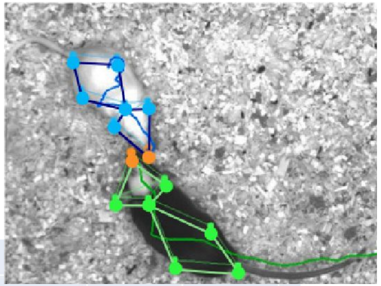


Exploring DISK learned representations

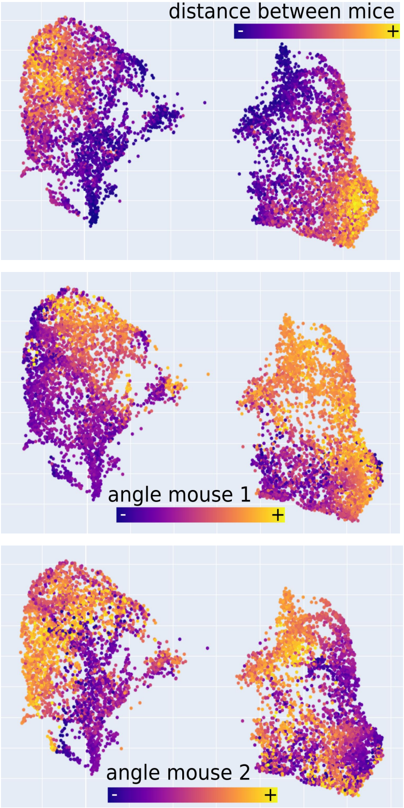


U-map of sequence embeddings

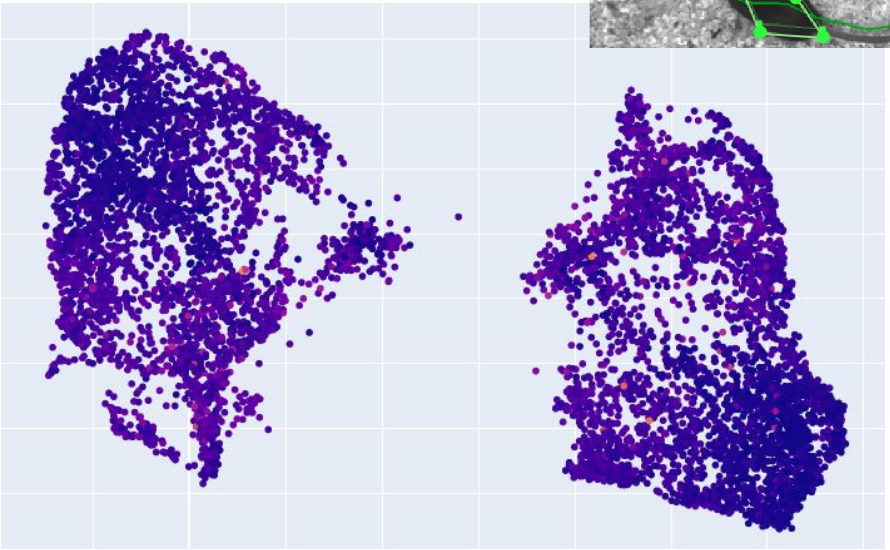


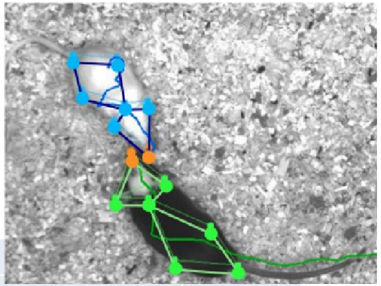


Exploring DISK learned representations

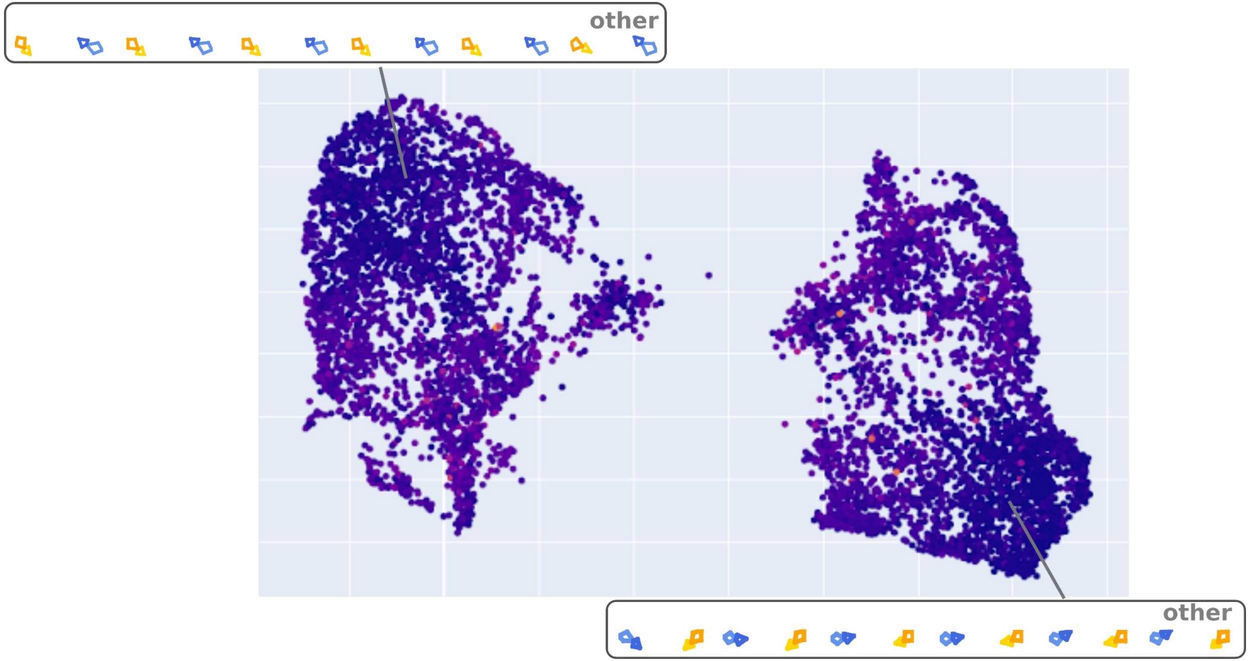


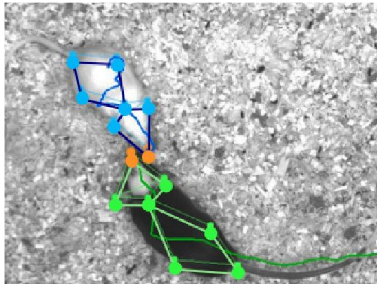
U-map of sequence embeddings



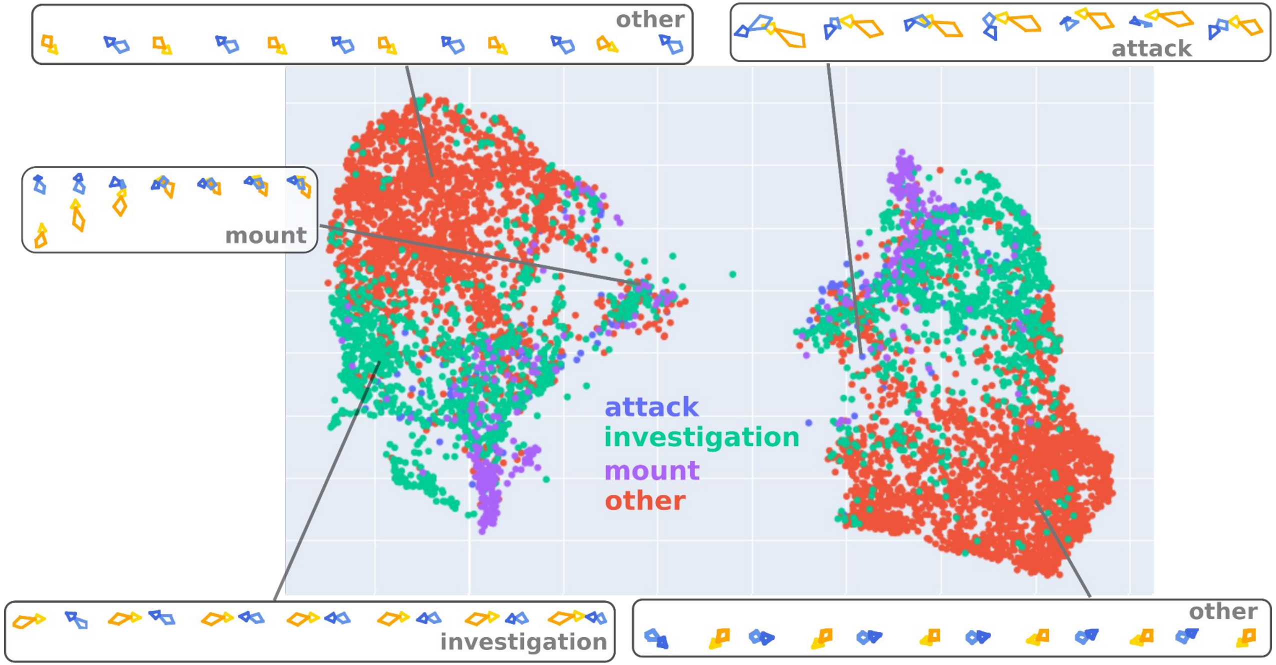


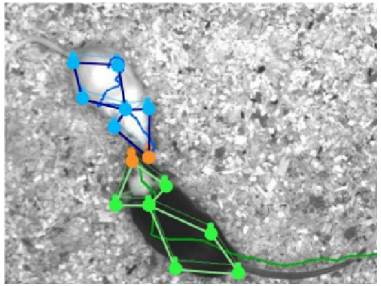
Exploring DISK learned representations





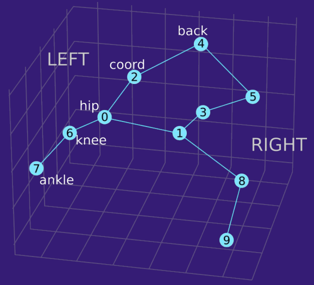
Exploring DISK learned representations





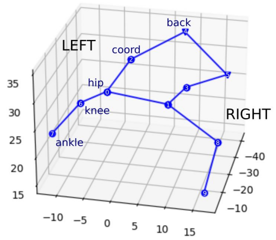
Random Forest on latent vectors
4-action class classification
- balanced accuracy: 0.877
- balanced F1-score: 0.846
- balanced precision score: 0.874

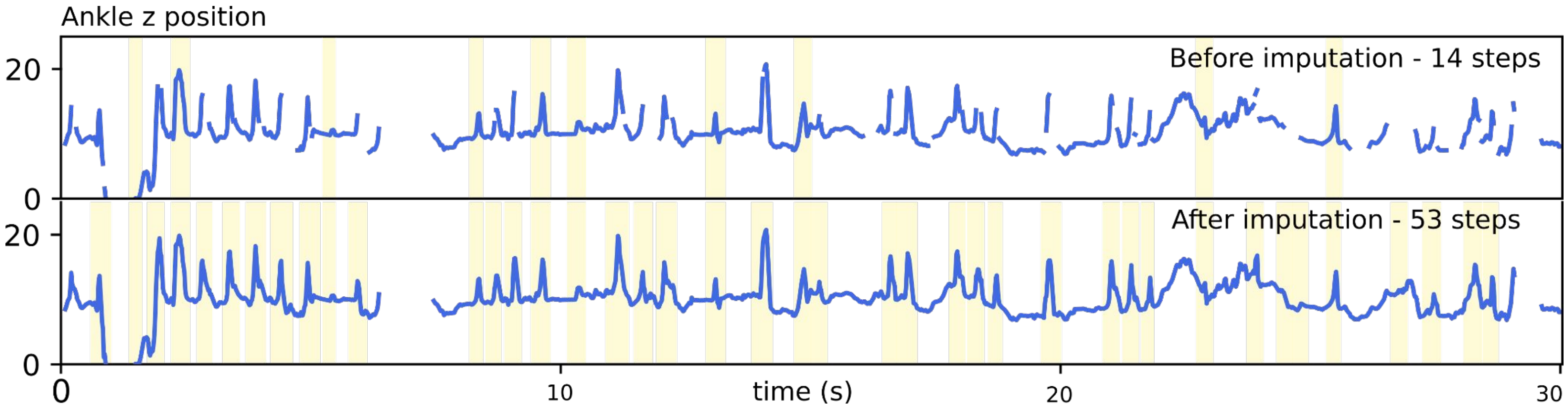
What to do with DISK?



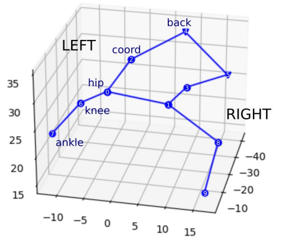
An example:
Step detection in freely moving mice

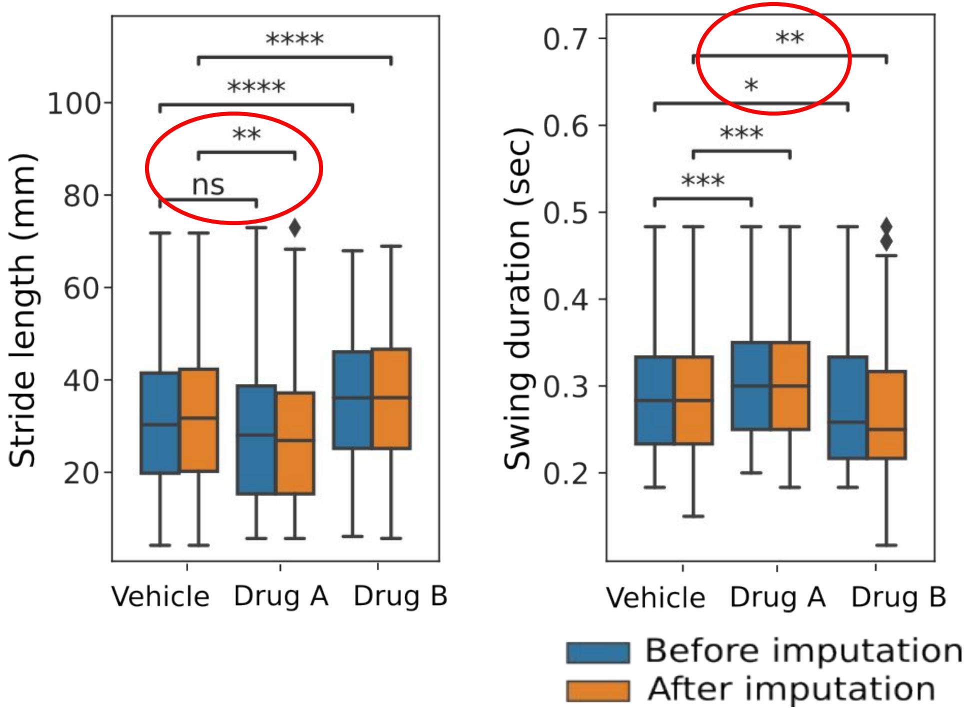
Step detection in 3D Motion Capture mouse data





Insight on pharmacological drug effect





Concluding remarks

Github

Preprint

- DISK is able to impute correctly long gaps for single or multiple missing keypoints.
- An estimated error helps filtering out below-threshold imputed samples.
- Complementary to pose detection, DISK can help analyze fine movements like locomotion.

BOZEK LAB

CMMC
CENTER FOR
MOLECULARE MEDICINE
COLOGNE

UNIKLINIK
KÖLN

FUNDED BY
Ministry of Culture and Science
of the State of
North Rhine-Westphalia

Katarzyna Bozek

France Rose

Timon Blindauer

Monika Michaluk

Talmo D. Pereira

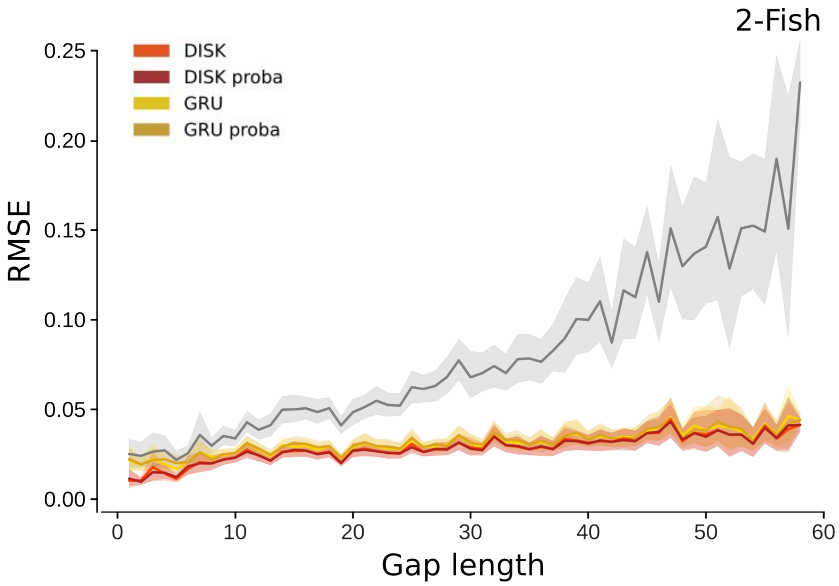
Liam O'Shaughnessy

Greg J. Stephens

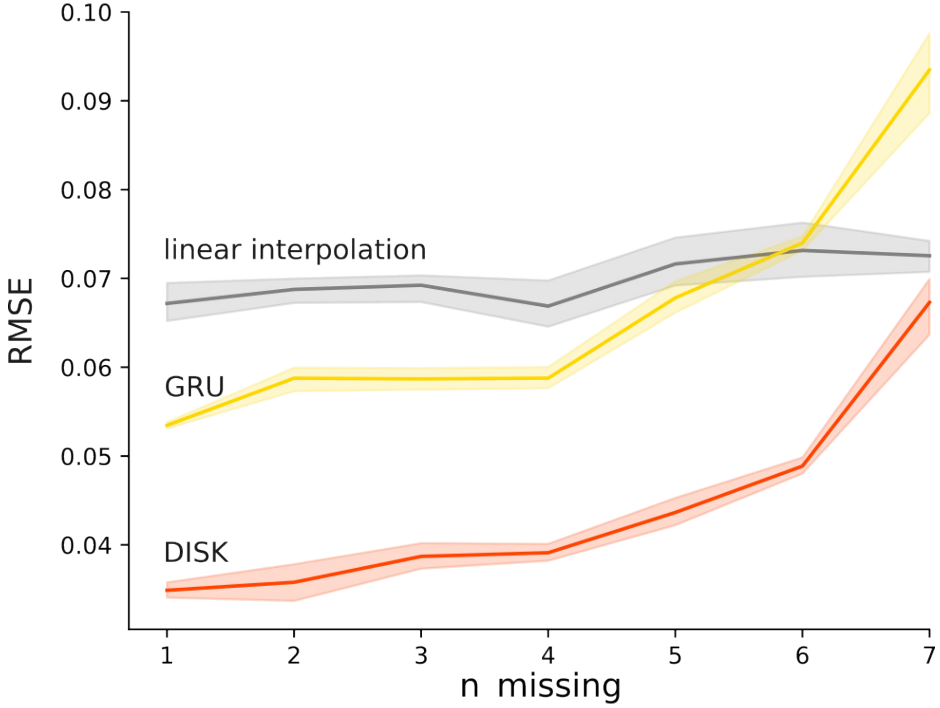
Bogna Ignatowska-Jankowska

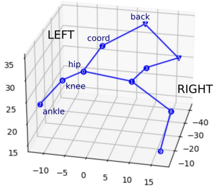
Marylka Y. Uusisaari

Neural methods robust to increasing gap length

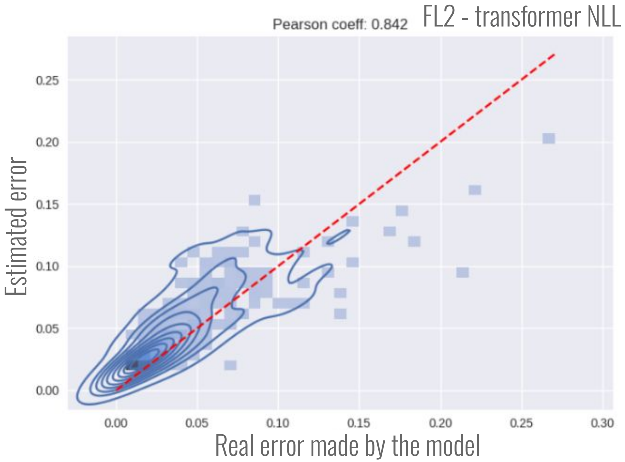


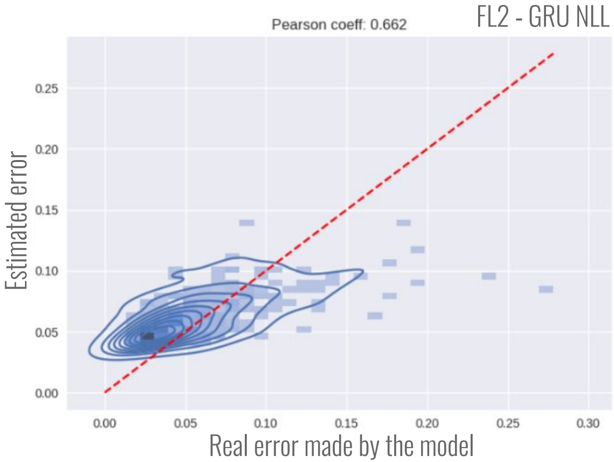
Imputing multiple keypoints simultaneously

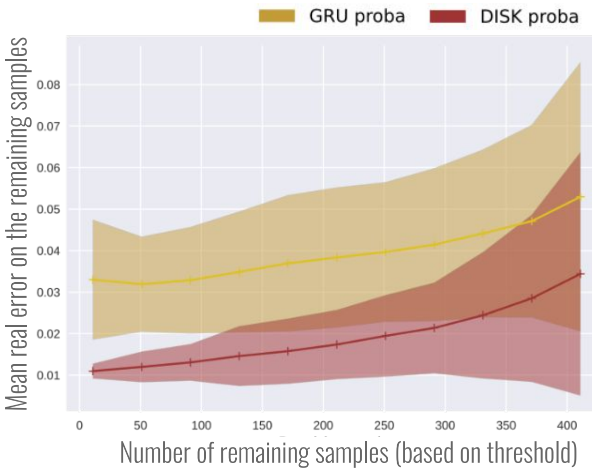




Estimated error on the imputed samples







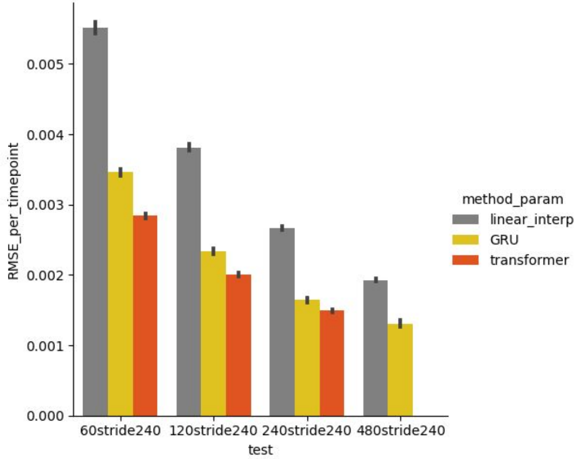
- Good correlation between real and estimated error
- Red line is x=x: slight overestimate of the real error

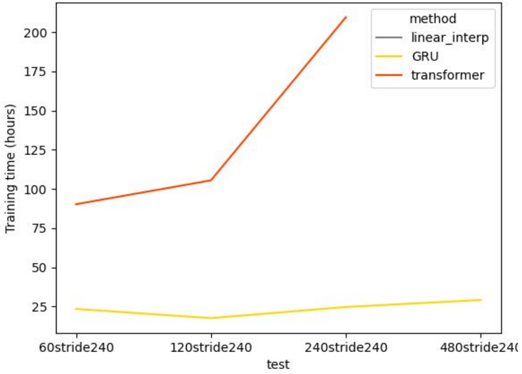
- Use it to threshold and keep only good samples

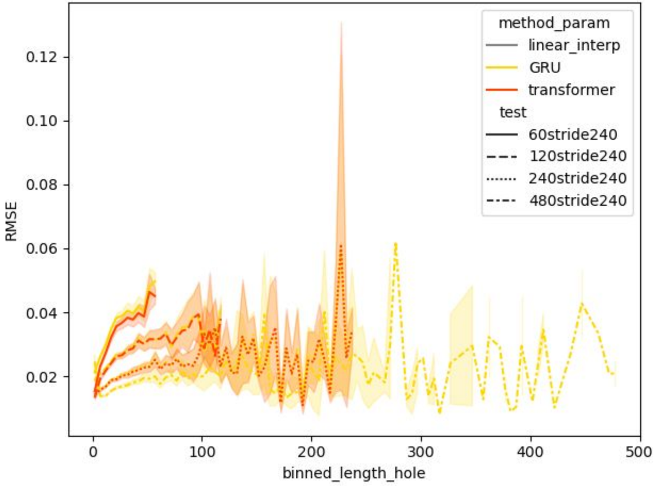
Datasets' properties

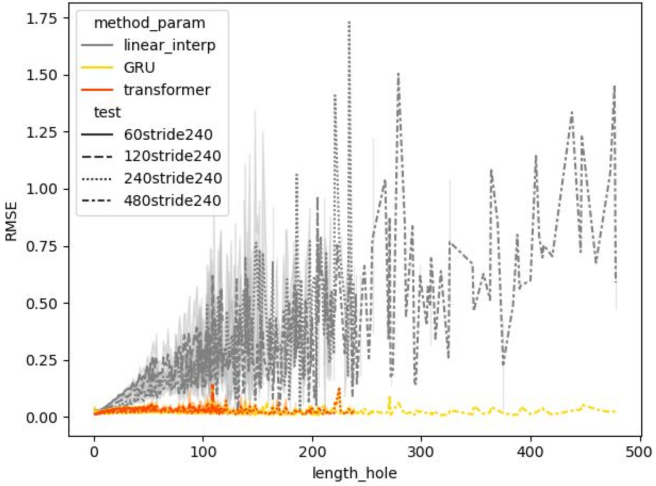
Dataset	N kp	Freq	Stride	Size train / val / test	Missing prop [%]
FL2	8	60	30	4,396 / 422 / 413	24
CLB	8	6	30	8,571 / 983 / 918	16
DF3D	38	100	5	2,095 / 652 / 614	0
Human	20	12	30	8,593 / 823 / 869	0
Rat7M	20	30	30	13,463 / 2,840 / 2,713	44
2-Fish	2×3	60	120	99,029 / 13,327 / 15,705	6
MABe	2×7	30	60	6,820 / 986 / 622	0

Input sequence length



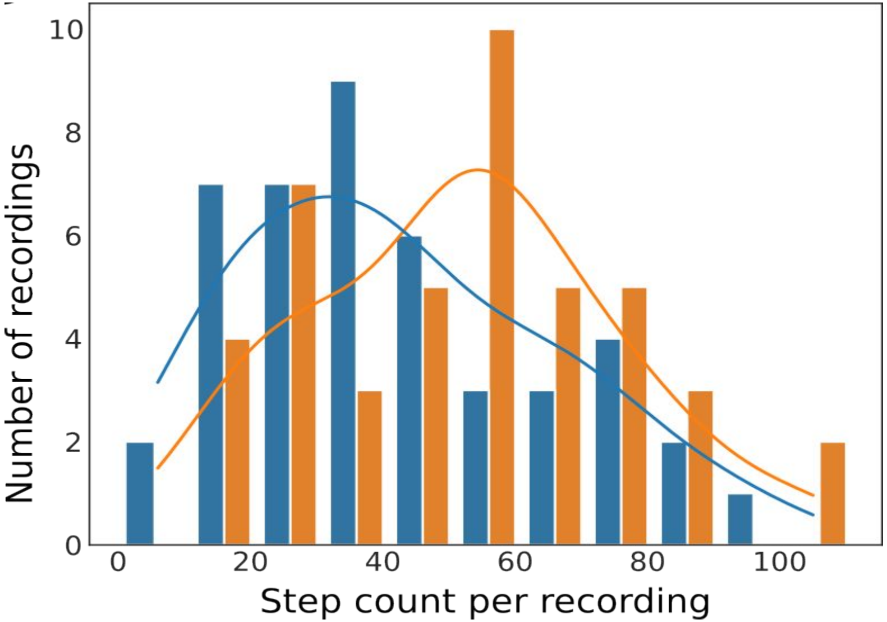


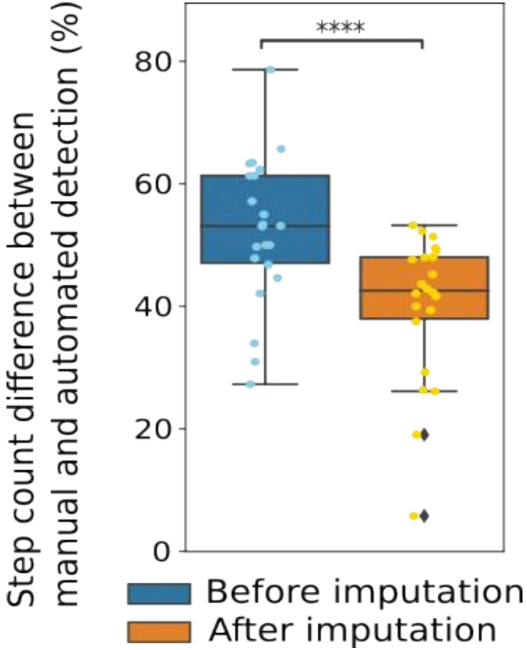




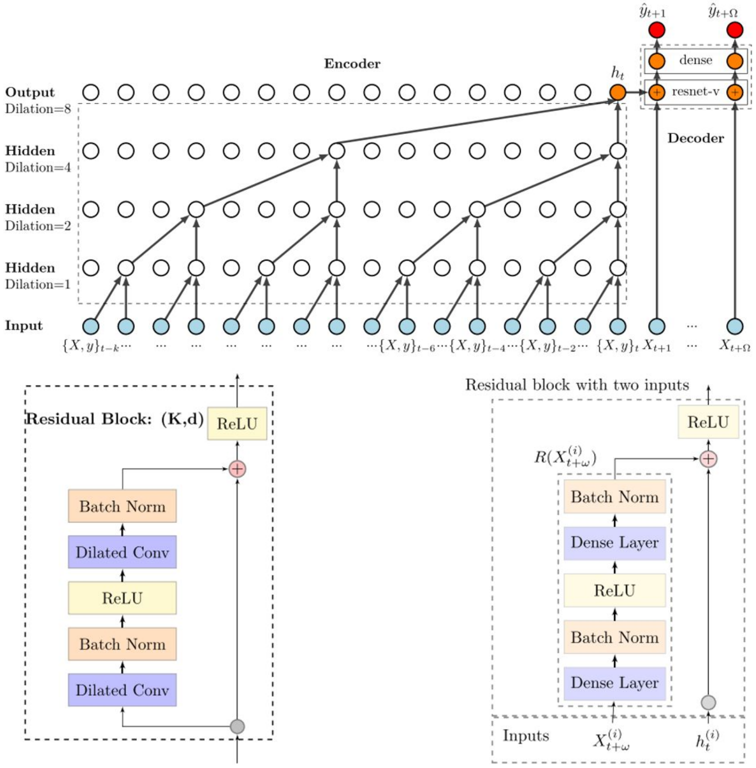
- Increasing input sequence length improves performance (see RMSE per timepoint or RMSE vs length_hole plots)
- Increasing input length is more beneficial to GRU than transformer (Weird!)
- Increased input length + GRU is a better combination (less training time for better performance)

Better step detection with imputed data





TCN



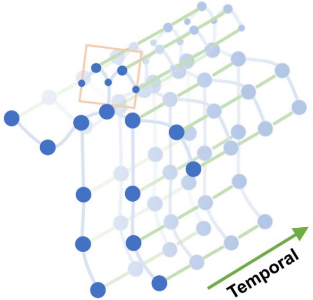
(b) Encoder module

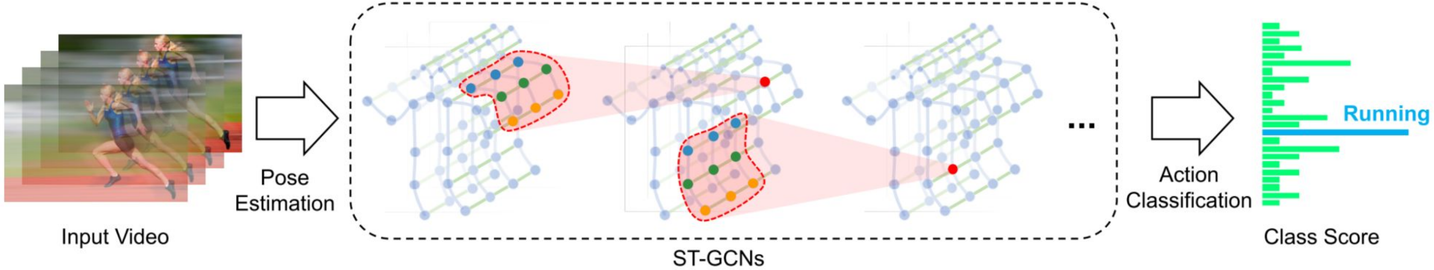
(c) Decoder module

Temporal Convolutional Networks for Action Segmentation and Detection, Lea et al. 2016

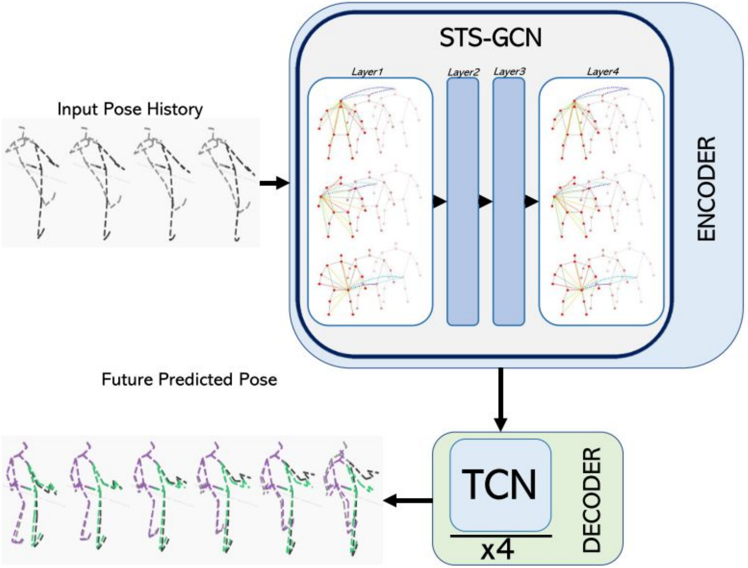
Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition

Sijie Yan, Yuanjun Xiong, Dahua Lin
Department of Information Engineering, The Chinese University of Hong Kong
{ys016, dhlin}@ie.cuhk.edu.hk, bitxiong@gmail.com

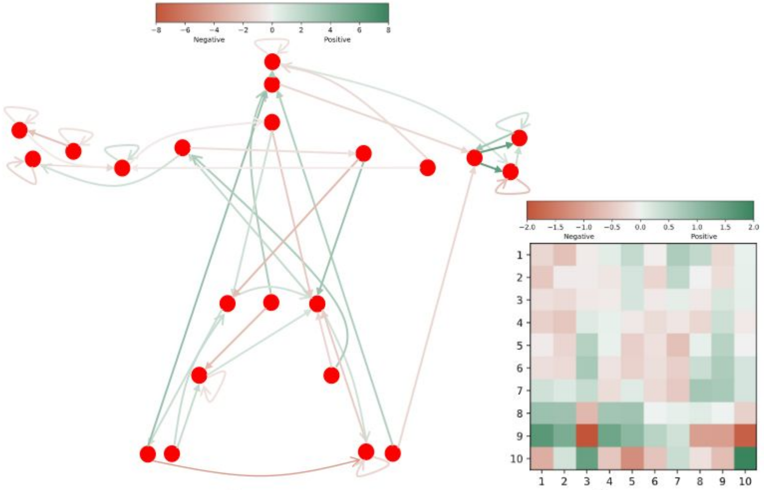




STS-GCN



Separable learnable adjacency matrices in time and space

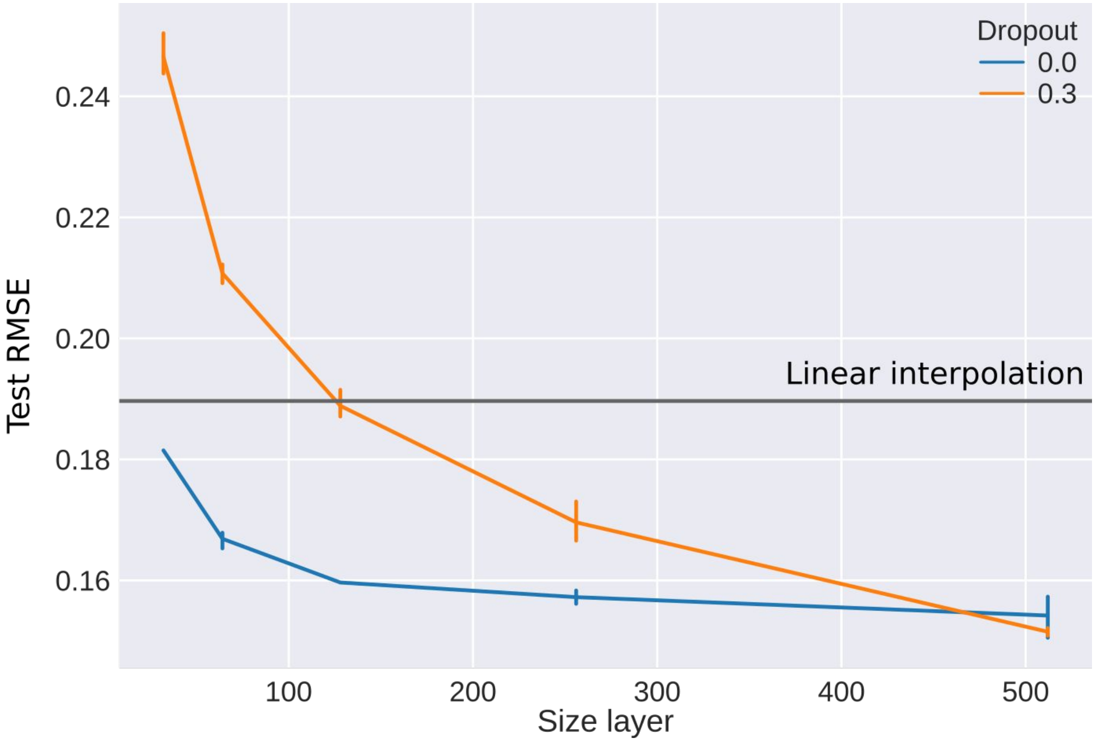


Learnt A^s (layer 1, time 1) Learnt A^t (layer 1, head)

coding GCN. Bottleneck'ing the space-time cross-talk is realized by factoring the space-time adjacency matrix into the product of separate spatial and temporal adjacency matrices $A^{st} = A^s A^t$. A separable space-time graph convolutional layer l is therefore written as follows

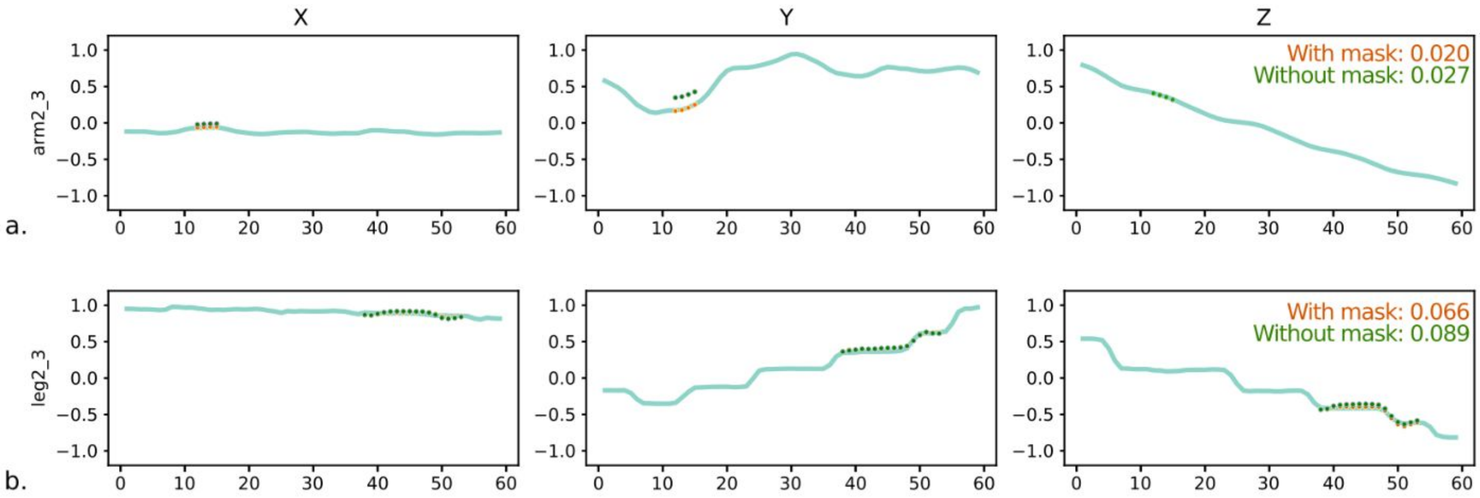
$$\mathcal{H}^{(l+1)} = \sigma(A^{s-(l)} A^{t-(l)} \mathcal{H}^{(l)} W^{(l)}) \qquad (2)$$

Bigger hidden size performs better (DF3D)



Binary input mask guides the network

	FL2	CLB	DF3D	MoCap	Rat7M
Linear interpolation	0.07	0.17	0.20	0.36	0.13
ImputeSkeleton					
With mask	**0.04**	**0.04**	**0.15**	**0.04**	**0.05**
Without mask	0.05	0.05	0.16	0.05	0.07



* Human