Overcoming the Domain Gap in Contrastive Learning of Neural Action Representations

Semih Günel, Florian Aymanns, Sina Honari, Pavan Ramdya and Pascal Fua

Computer Vision Lab, EPFL, Lausanne, Switzerland
Neuroengineering Lab, EPFL, Lausanne, Switzerland

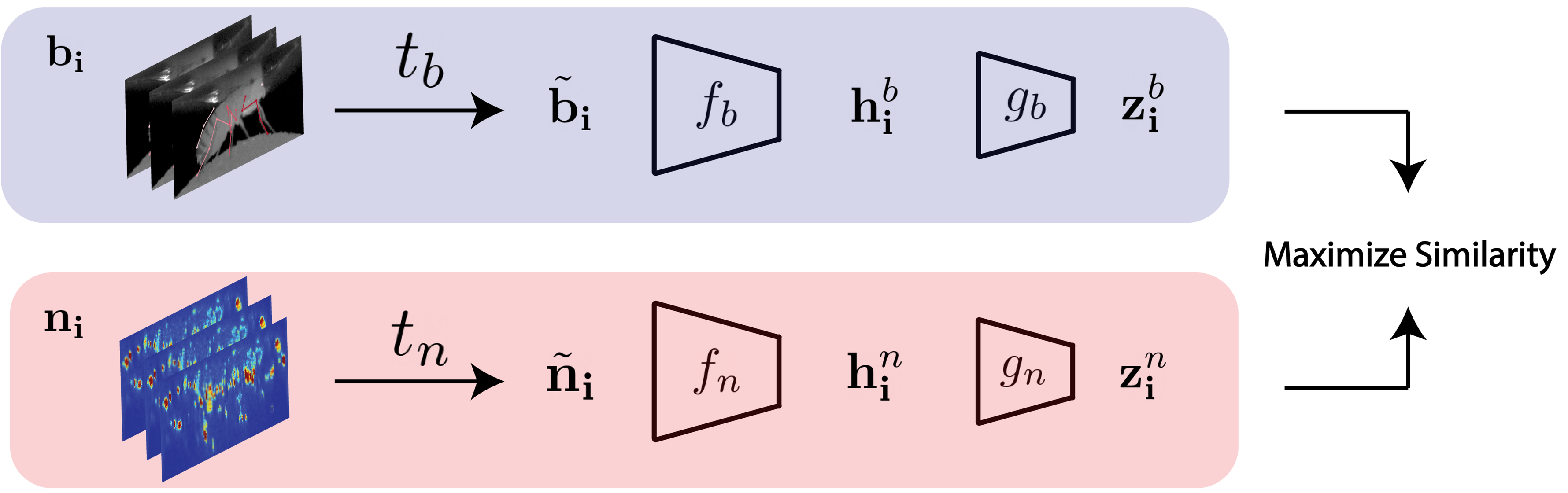
Problem:

- Relating behavior to brain activity is a fundamental goal in neuroscience, with practical applications in building robust brain-machine interfaces. However, the domain gap between individuals is a major issue that prevents the training of general models that work on unlabeled subjects.
- Since 3D pose data can now be reliably extracted from multi-view video sequences without manual intervention, we propose to use it to guide the encoding of neural action representations together with a set of neural and behavioral augmentations exploiting the properties of microscopy imaging. To reduce the domain gap, during training, we swap neural and behavioral data across animals that seem to be performing similar actions

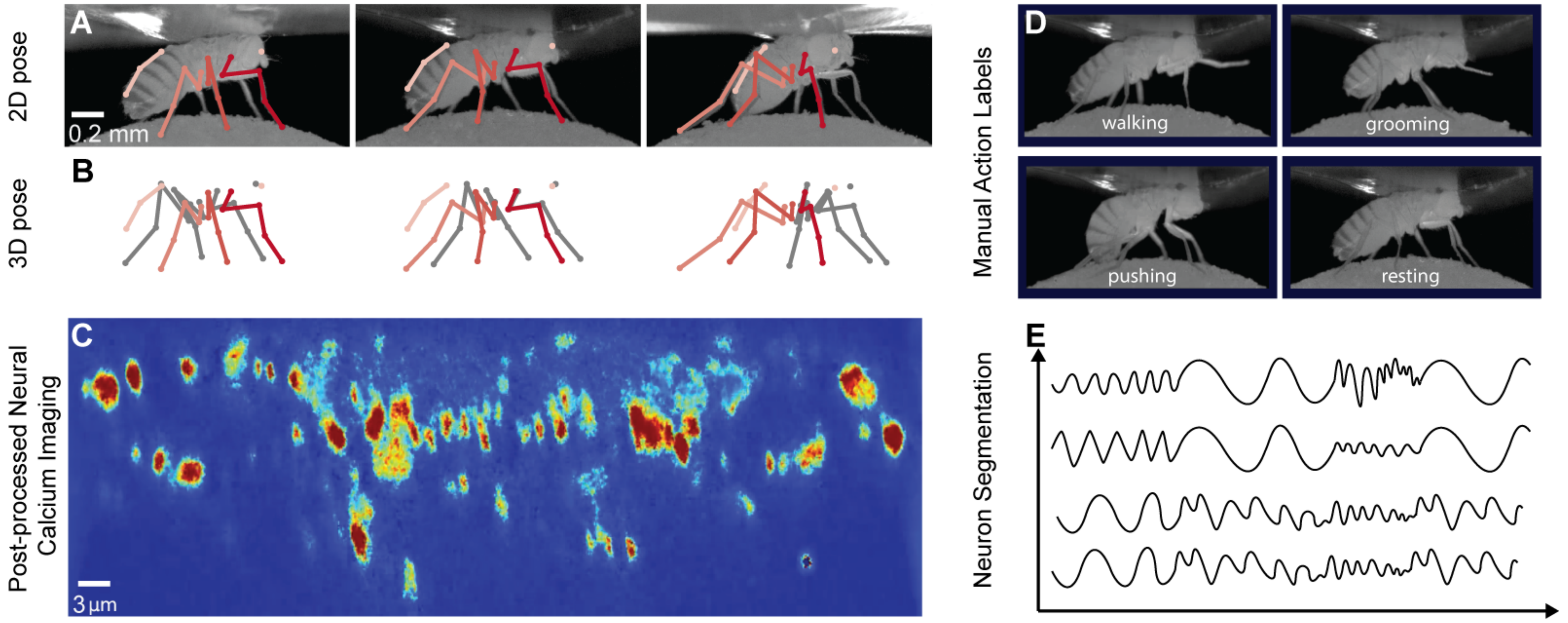
Method:

- The loss function maximizes the mutual information between two modalities, 3D pose and neural microscopy recordings.

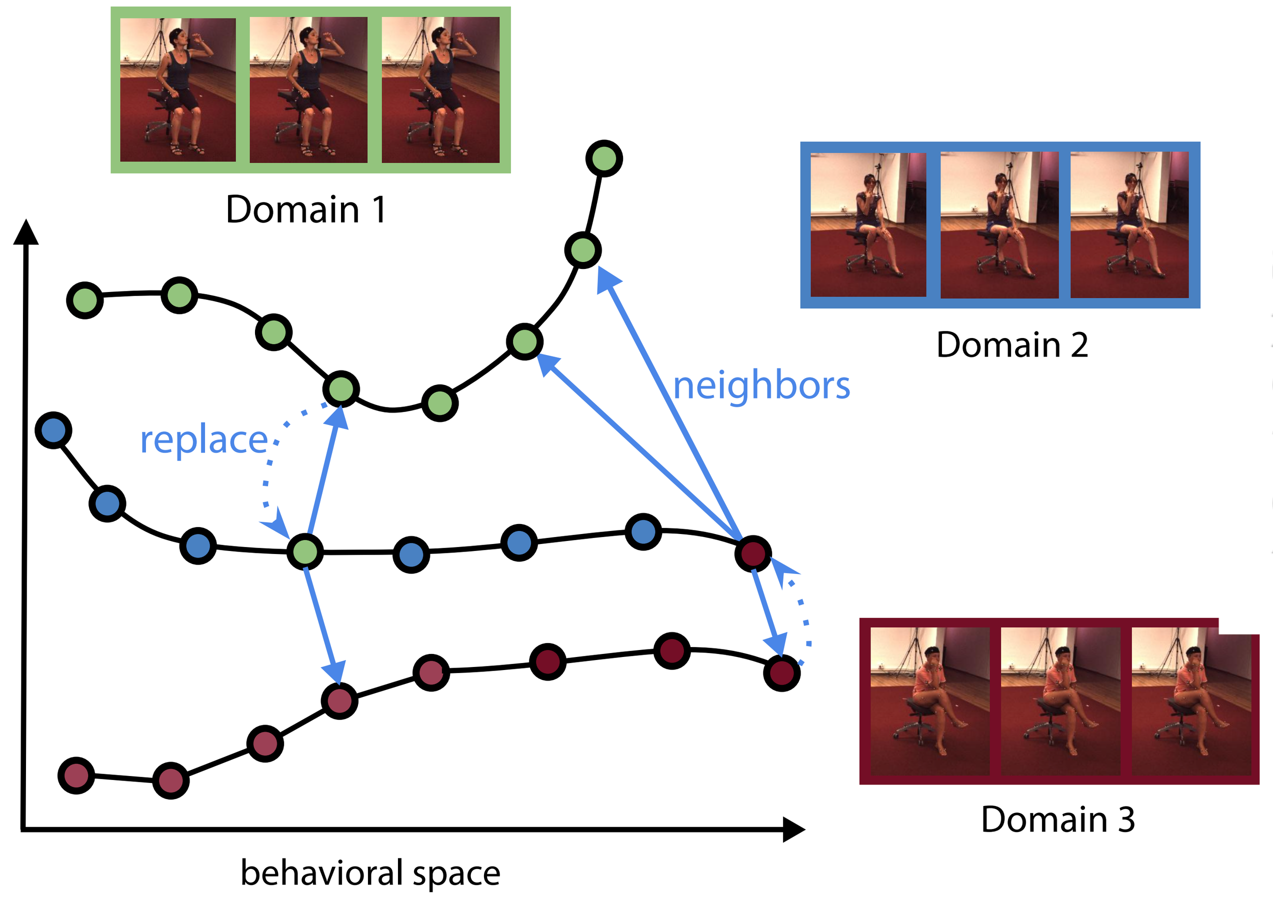
$$\mathcal{L}_{NCE}^{n \to b} = -\sum_{i=1}^{N} \log \frac{\exp\left(\langle \mathbf{z}_b^i, \mathbf{z}_n^i \rangle / \tau\right)}{\sum_{k=1}^{N} \exp\left(\langle \mathbf{z}_b^k, \mathbf{z}_n^i \rangle / \tau\right)}. \quad (1)$$



Our Dataset:



Method:



- A tethered fly (Drosophila melanogaster) is recorded using six multi-view infrared cameras and a two-photon microscope. The dataset includes **40 animals over 364 trials, resulting in 20.7 hours of recordings with 7,480,000 behavioral images and 1,197,025 neural images.**

Swapping Augmentation. Given a set of consecutive 3D poses $\mathbf{b}_\mathbf{i}^s$, for each $k \in \mathbf{i}$, we stochastically replace $\mathbf{b}_k^s$ with one of its nearest pose neighbors in the set of domains $\mathcal{D}_\mathcal{S}$, where $\mathcal{S}$ is the set of all animals. To do this, we first randomly select a domain $\hat{s} \in \mathcal{S}$ and define a probability distribution $\mathbf{P}_{\mathbf{b}_k^s}^{\hat{s}}$ over the domain $\mathcal{D}_{\hat{s}}$ with respect to $\mathbf{b}_k^s$,

$$\mathbf{P}_{\mathbf{b}_k^s}^{\hat{s}}(\mathbf{b}_l^{\hat{s}}) = \frac{\exp(-\|\mathbf{b}_l^{\hat{s}} - \mathbf{b}_k^s\|_2)}{\sum_{\mathbf{b}_m^{\hat{s}} \in \mathcal{D}_{\hat{s}}} \exp(-\|\mathbf{b}_m^{\hat{s}} - \mathbf{b}_k^s\|_2)}. \quad (2)$$

Results:

- Our swapping augmentation strongly improves action recognition compared to the regression and multi-modal contrastive (SimCLR) baselines. Uni-modal neural contrastive training cannot generalize across subjects, due to the large domain gap in neural modality.

Tasks →		Single-Animal ↑		Multi-Animal ↑	
Percentage of Data →		0.5	1.0	0.5	1.0
Random Guess		16.6	16.6	16.6	16.6
Neural (Linear)	Raw	29.3	32.5	18.4	18.4
Neural (MLP)	Raw	–	–	18.4	18.4
SimCLR [56]	Self-Supervised	54.3	57.6	46.9	50.6
Regression (Recurr.)	Self-Supervised	53.6	59.7	49.4	51.2
Regression (Conv.)	Self-Supervised	52.6	59.6	50.6	55.8
BehaveNet [27]	Self-Supervised	54.6	60.2	50.5	56.8
Ours	Self-Supervised	**57.9**	**63.3**	**54.8**	**61.9**
SimCLR [56] + MMD	Domain Ada.	53.6	57.8	50.1	53.1
SimCLR [56] + GRL	Domain Ada.	53.5	56.3	49.9	52.3
Regression (Conv.) + MMD	Domain Ada.	54.5	60.7	52.6	55.4
Regression (Conv.) + GRL	Domain Ada.	55.5	60.2	51.8	55.7

Table 1

Method	Identity Recog. (0.5, Accuracy)	Identity Recog. (1.0, Accuracy)
Random Guess	12.5	12.5
Behavior (Linear)	88.6	89.7
Neural (Linear)	100.0	100.0
SimCLR [56]	69.9	80.3
Regression (Recurrent)	89.5	91.8
Regression (Convolution)	88.7	92.5
BehaveNet [27]	80.2	83.4
Ours	**12.5**	**12.5**
SimCLR + MMD [67]	18.4	21.2
SimCLR + GRL [68]	16.7	19.1

Table 2