Behavior is a Language for Cross-Embodiment Control

Matt Hyatt [1] Griffen Lee [1] Nathan Hogg [1] George K Thiruvathukal [1] Neil Klingensmith [1]

[1]Loyola University Chicago

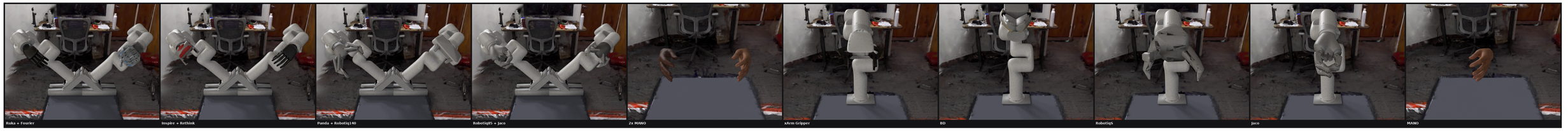


Figure 1. We use Genesis [5] physics engine and render gaussian splat behind the simulator scene, matching our real deployment environment. We test across 4 mounting platforms (L, R, LR, C) with mixed bimanual and single arm robots and across 11 end effectors. Additionally we test with human hands in single and bimanual manipulation.

Simulating Human Motion is more Scalable

- How best to learn robot control from humans behaviors? To implicitly bridge the embodiment gap, we must learn from many different embodiments.
- We hypothesize that learning subtle cues from position and embodiment shape will teach the model to predict motion conditioned by those attributes- allowing adaptation to new robots, or from people!
- **BeLa** is a model cross attention architecture that learns differences in robot motion primitives allowing flexible control across all robot embodiments.

Latent Space

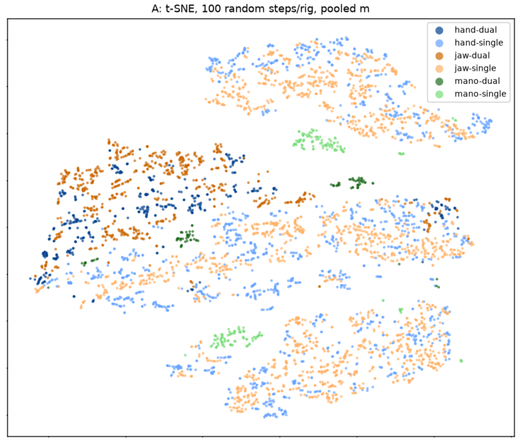


Figure 2. t-SNE visualization of latent vector prior to decoding. Each embodiment occupies its own territory. Notably, the macro shape shows that latents are organized according to how the robot was mounted to the table; bimanual robots and human hands lie together while left, right, and center mounts each lie in distinct groups.

During training, our autoencoder reconstruction loss shows that the robot state and identity is recoverable from this latent, showing that it is possible to implicitly learn morphology differences across actors.

Tokenization

- Our model is designed based on the Perciever-IO transformer [3] and tokenization inspired by [2].
- Our tokenization pipeline consists of primitives derrived from learned embedding "factors": bodypart, property, frame, and Fourier projection of the corresponding value.
- Other embeddings include Entity Tokens.
- Finally, primities are modulated by their "role": observation, goal, prompt.
- BeLa attends to action queries a_q to denoise actions from latents.

Users can pick-and-chose and mix control modes.

Cross-Embodiment Scaling

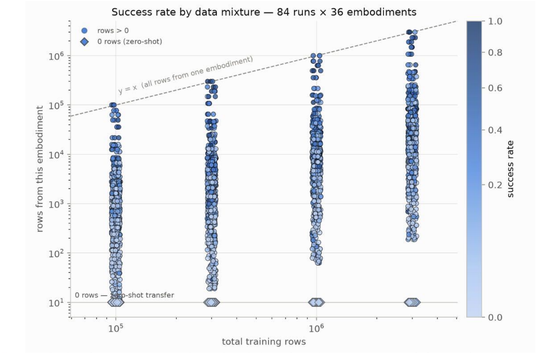


Figure 3. As the aggregate size of training data increases, models can better generalize from all embodiments toward the target embodiment.

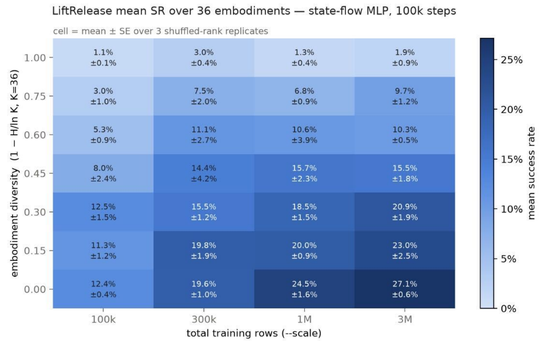
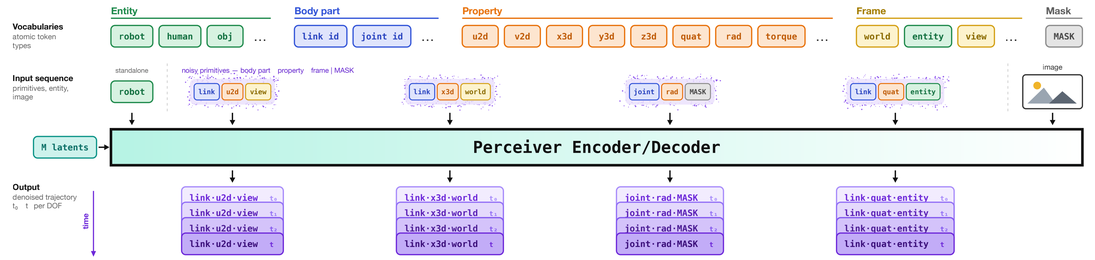


Figure 4. Increasing the diversity of embodiments in the training data set results in increased success rate in the block lifting task.

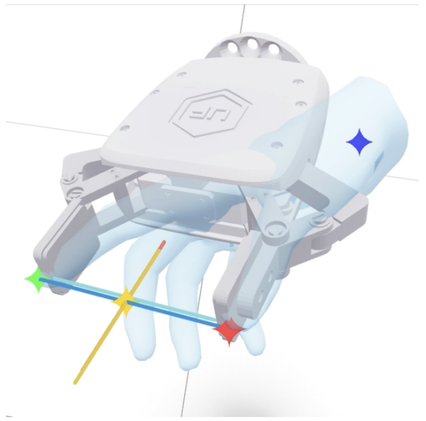
BeLa Architecture



Architecture of BeLa.

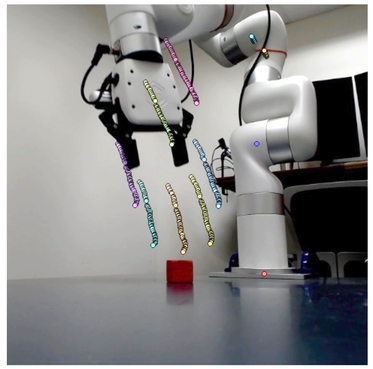
Prior Work & Retargeting

- Previous work learns palm/end effector position [4].
- The best retargeting lines up the thumb with the forefinger, not the palm with the mount.
- Object interaction happens at finger tips which must be implicitly controlled unless they are also predicted as a target.



Deployment

Outside the simulator, we can use BeLa under zero-shot sim2real deployment to successfully complete manipulation tasks.



This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-23-F-0014 in the amount of $ 153,600.

References

[1] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. *arXiv preprint arXiv:2212.08051*, 2022.

[2] Huy Ha, C Karen Liu, and Shuran Song. Transformer transformer: A unified model for motion-conditioned robot co-design. *arXiv preprint arXiv:2607.25798*, 2026.

[3] Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, Olivier Hénaff, Matthew M. Botvinick, Andrew Zisserman, Oriol Vinyals, and João Carreira. Perceiver io: A general architecture for structured inputs outputs, 2022. URL `https://arxiv.org/abs/2107.14795`.

[4] Simar Kareer, Dhruv Patel, Ryan Punamiya, Pranay Mathur, Shuo Cheng, Chen Wang, Judy Hoffman, and Danfei Xu. Egomimic: Scaling imitation learning via egocentric video, 2024. URL `https://arxiv.org/abs/2410.24221`.