

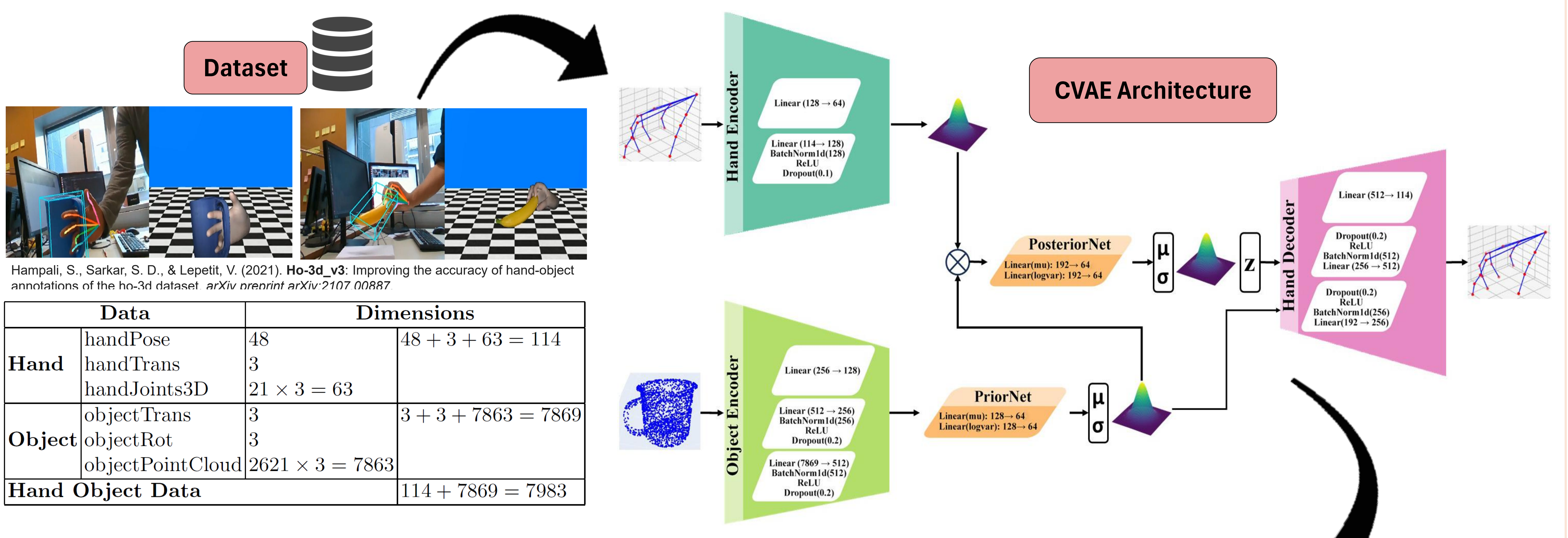
ABSTRACT

- Brain Imaging studies show that when planning grasping action, certain parts of the **sensorimotor cortex** associate **object properties** with the **type of grasp** that would match the object.
- Curious to see if such patterns can also emerge in Artificial Neural Networks, a Conditional Variational Autoencoder (CVAE) was trained on the HO-3D_v3 dataset to **reconstruct grasping hand poses** conditioned on the object.
- Principal Component Analysis (PCA) on the latent space of the trained CVAE model revealed **abstract hand representations** that were distinctly **organized by the type of object**.

RESEARCH QUESTIONS

- Can **Artificial Neural Networks** develop **latent representations** that mimic how the human brain **links object properties to grasp types**?
- Would the **latent space of a CVAE** trained on **hand-object interaction data** show **affordance-relevant organizations** like what is observed in the human sensorimotor cortex?

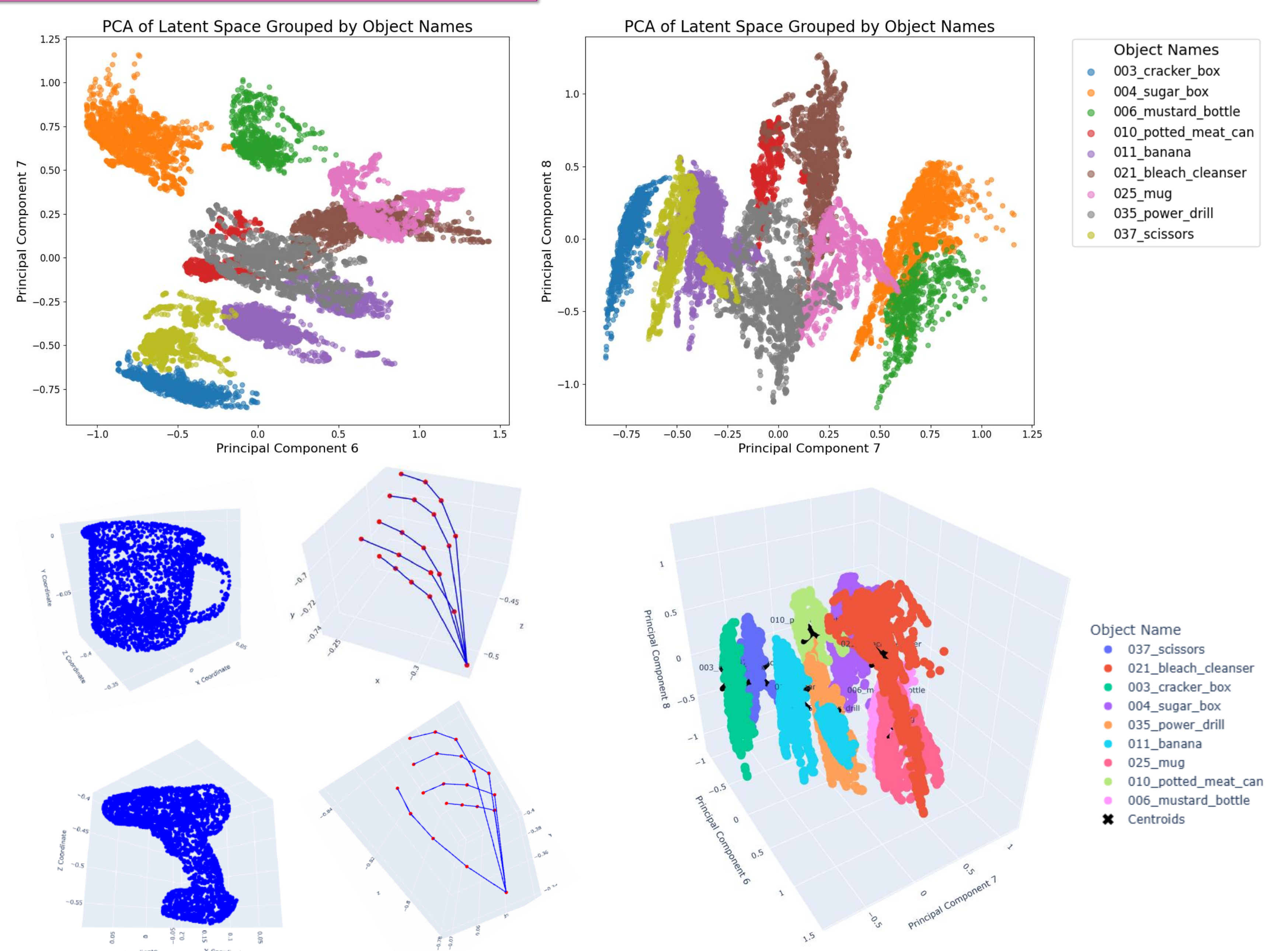
DATASET AND MODEL TRAINING



LATENT SPACE ANALYSIS

Findings:

- CVAE effectively reconstructed grasping hand poses based on the objects.
- PCA revealed a structured latent space: the first **8 components** explained over **97%** of the variance.
- Distinct **object-specific clusters** were observed in the latent space, notably **PCs 5-8**.
- Clusters** reflect **functional similarities** in **grasp types** rather than purely visual object features (e.g., scissors and banana overlap due to similar grip).
- These representations could improve robotic control by enabling real-time motor planning in the latent space for complex humanoid hand actions, bypassing the need for pixel and joint-level computations.



ACKNOWLEDGMENTS

This work was funded by **Innovate UK** (grant number 10089807 for the Horizon Europe project **PRIMI** Grant agreement n. 101120727).