Robot observation

Action prediction

Self-supervised encoder

Visual representation

Nearest-Neighbor matches

Robot execution