NOIR: Neural Signal Operated Intelligent Robots for Everyday Activities

Ruohan Zhang*[1,4], Sharon Lee*[1], Minjune Hwang*[1], Ayano Hiranaka*[2], Chen Wang[1], Wensi Ai[1], Jin Jie Ryan Tan[1], Shreya Gupta[1], Yilun Hao[1], Gabrael Levine[1], Ruohan Gao[1], Anthony Norcia[3], Li Fei-Fei[1,4], Jiajun Wu[1,4]

*Equally contributed; zharu@stanford.edu

[1]Department of Computer Science, Stanford University

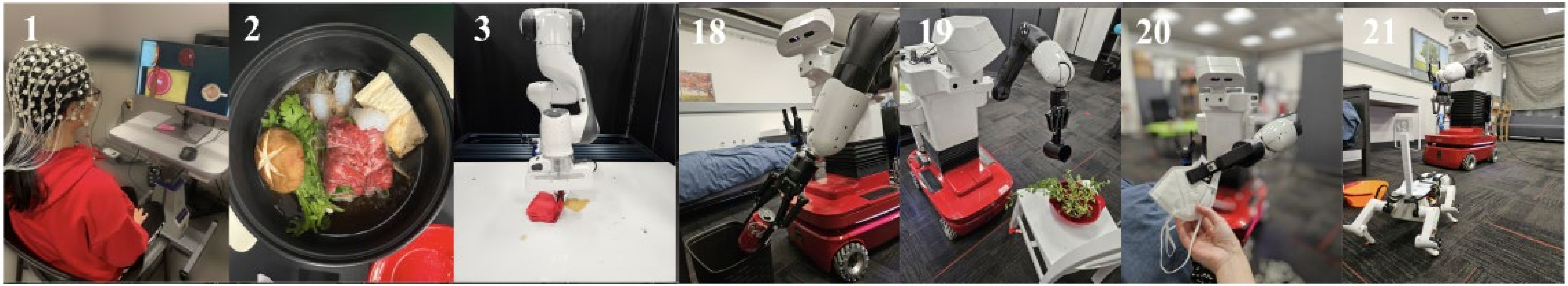
[2]Department of Mechanical Engineering, Stanford University

[3]Department of Psychology, Stanford University

[4]Institute for Human-Centered AI (HAI), Stanford University

CORL 2023

Motivation



- Brain-robot interfaces (BRIs) are a pinnacle achievement in the realm of art, science, and engineering.

  Brain-robot interfaces, BRIs: 脑机接口 or 脑机器人接口
  Brain-Machine Interface, BMI; Brain Computer Interface, BCI: 脑机接口

Contributions:

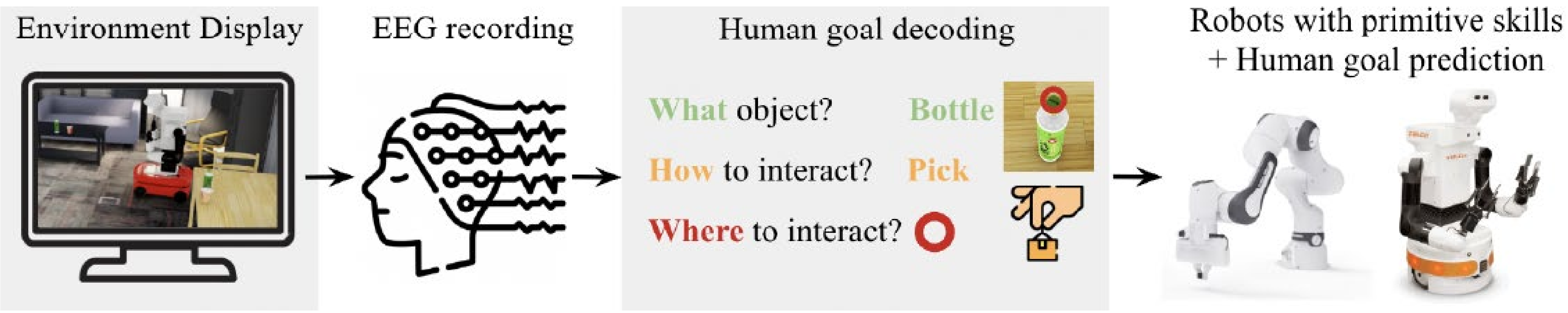
- First, NOIR is general-purpose in its diversity of tasks and accessibility.

  Accomplish 20 daily everyday activities, in contrast to existing BRI systems.
  The system can be used by the general population, with a minimum amount of training.

- Second, the “I” in NOIR means that our robots are intelligent and adaptive.

  The robots are equipped with a library of diverse skills, allowing them to perform low-level actions without dense human supervision.

Motivation



- A modular neural signal decoding pipeline for human intentions.

  Decoding human intended goals (e.g., “pick up the mug from the handle”) from neural signals is extremely challenging.

  Decompose human intention into three components:
  What object to manipulate, How to interact with the object, and Where to interact.

Background

- Steady-state visually evoked potential (SSVEP)

  SSVEP is the brain's exogenous response to periodic external visual stimulus, wherein the brain generates periodic electrical activity at the same frequency as flickering visual stimulus.

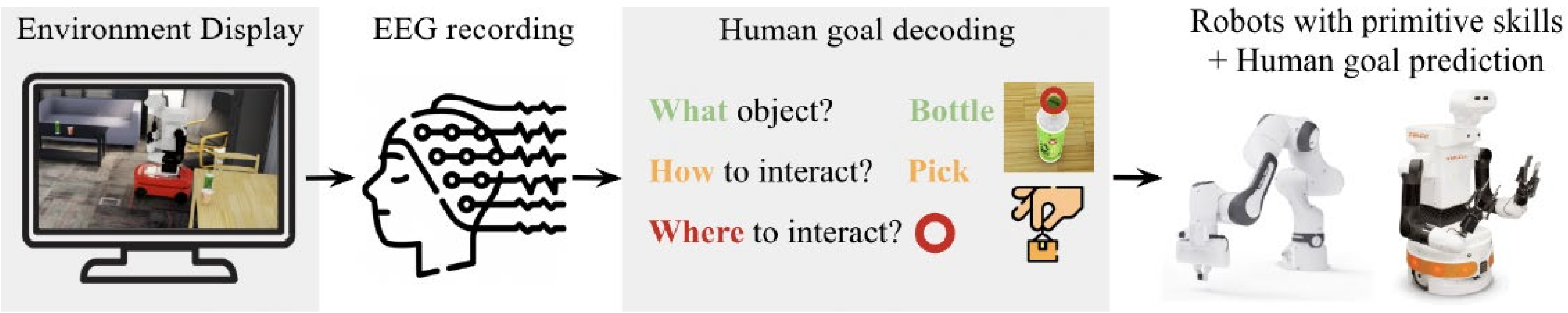
  Prior work: LED lights   This work: segment objects, attach virtual flickering masks to each object.

- Motor Imagery (MI)

  Motor Imagery (MI) differs from SSVEP due to its endogenous nature, requiring individuals to mentally simulate specific actions, such as imagining oneself manipulating an object.

  The decoded signals can be used to indicate a human's intended way of interacting with the object.

Method



- Humans act as planning agents to perceive, plan, and communicate behavioral goals to the robot, while robots use pre-defined primitive skills to achieve these goals.

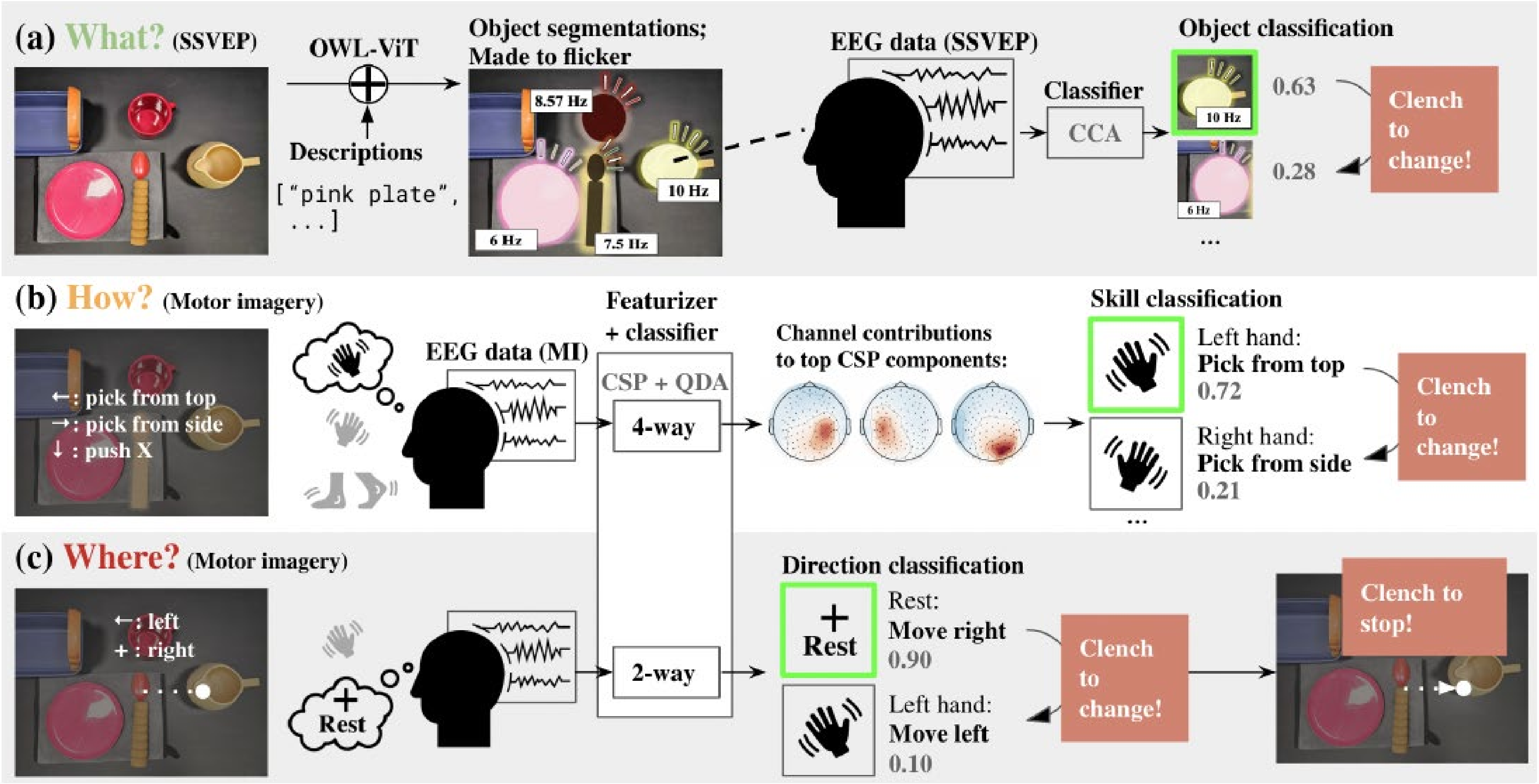
- A general-purpose BRI system is achieved by synergistically integrating two designs together.

  First, we propose a novel modular brain decoding pipeline for human intentions, in which the human intended goal is decomposed into three components: what, how, and where.

  Second, we equip the robots with a library of parameterized primitive skills to accomplish human-specified goals.

Method

- The brain: A modular decoding pipeline



Method

- The robot: Parameterized primitive skills

  The benefits of using these skills are that they can be combined and reused across tasks.

  Moreover, these skills are intuitive to humans.

  Neither the human nor the agent requires knowledge of the underlying control mechanism for these skills, thus the skills can be implemented in any method as long as they are robust and adaptive to various tasks.

Robot	Skill	Parameters
Franka	Reaching	6D goal pose in world
Franka	Picking	3D world pos to pick, gripper orientation (choose from 4)
Franka	Placing	3D world pos to place, gripper orientation (choose from 3)
Franka	Pushing	3D world pos to start pushing, axis of motion (choose from 3)
Franka	Wiping	3D world pos to start wiping
Franka	Drawing	3D world pos
Franka	Pouring	3D world pos, gripper orientation (choose from 3)
Franka	Pulling	3D world pos, gripper orientation (choose from 2), pull direction (choose from 2)
Franka	Grating	3D world pos
Tiago	Navigating	ID of pre-defined positions and poses
Tiago	Picking	ID of the object
Tiago	Placing	ID of the object
Tiago	Pouring	ID of the object
Tiago	Dropping	ID of object to drop the grasped object by

Table 6: Parameterized primitive skills for Franka and Tiago robots.

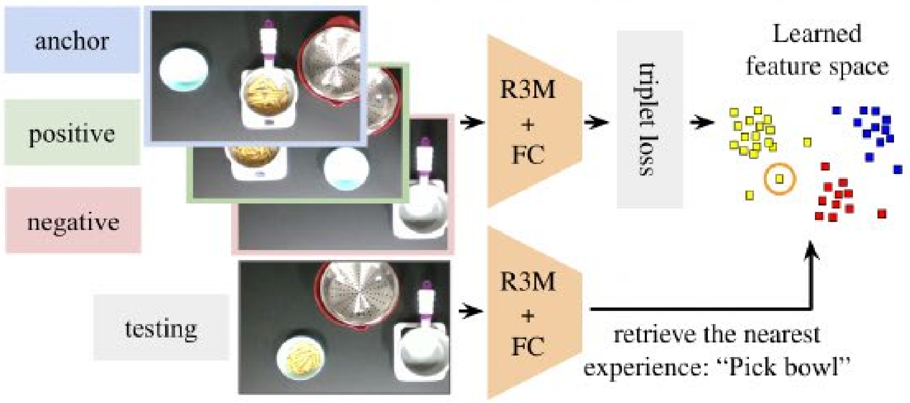
Method

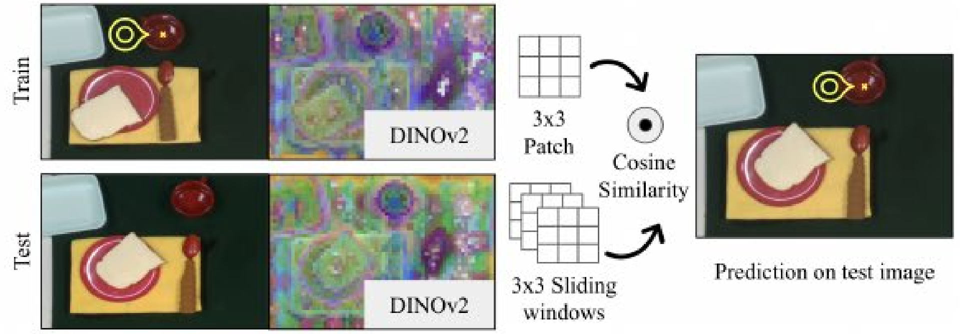
- Leveraging robot learning for efficient BRI

The modular decoding pipeline and the primitive skill library lay the foundation for NOIR. However, the efficiency of such a system can be further improved.

- Retrieval-based few-shot object and skill selection

Inspired by retrieval-based imitation learning, our proposed method learns a latent state representation from observed states. Given a new state observation, it finds the most similar state in the latent space and the corresponding action.





- One-shot skill parameter learning

Parameter selection requires a lot of human effort as it needs precise cursor manipulation through MI. To reduce human effort, we propose a learning algorithm, for predicting parameters given an object-skill pair as an initial point for cursor control.

Experiments

Decoding Stage	Signal	Technique	Calibration Acc.	Task-Time Acc.
Object selection (What?)	SSVEP	CCA (4-way)	-	0.812
Skill selection (How?)	MI	CSP + QDA (4-way)	0.580	0.422
Parameter selection (Where?)	MI	CSP + QDA (2-way)	0.882	0.739
Confirmation / interruption	EMG	Thresholding (2-way)	1.0	1.0

Table 2: Decoding accuracy at different stages of the experiment.

Thanks