(a) Representation-level Augmentation

Projector

Similar State

Dissimilar State

Adapter
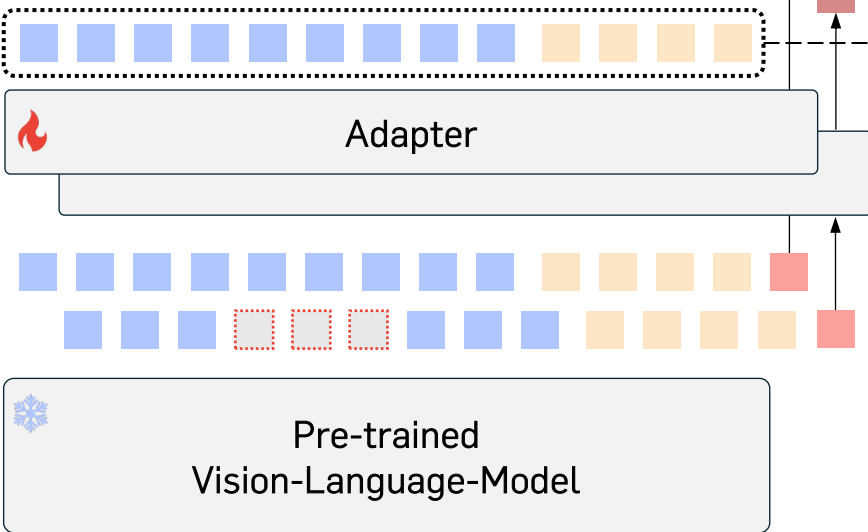
(b) Robot State-aware Contrastive Loss

Similarity w.r.t Proprioceptive State

Pre-trained Vision-Language-Model

"Open the Cabinet Door"

Action Decoder (e.g., DiT)

Control Actions