Planned Approach — CALVIN

(1.) Train a network that takes in input current images & outputs actions. So the hope is to use this network as an

$$s_t \longrightarrow \boxed{N_e} \longrightarrow a_t$$

encoding of an episode. So every episode would have its own network representation.

(2.) Now, from the play data, we'll take inputs ⌐ Goal image
                                                └ Lang. Instructⁿ

$$I_a \longrightarrow \boxed{N_w} \longrightarrow \text{weights of the episode representat}^n$$
$$L_a \longrightarrow$$

that we trained above.

The hope is to train an N_w that can predict the weights of a novel representatⁿ of novel goal.

This concept of hypernetwork is chosen in the hope to ease the burden of a single network having to do everything.

Here we don't just have a funct^ (network). But we also have a parent network that "guides" as to how to use that funct^ (or what part of funct^ must be used).