

Inputs

Observations



Instructions

"Pick up LEGO blocks and place it into the bin with the matching color."

State



Traditional WAM

Model size



heavy

Tokens inputs



dense

Denosing timesteps

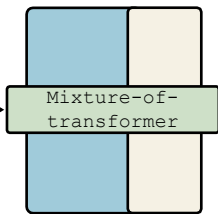


more

Inference speed



slow



Correct action

Efficient-WAM



lightweight



sparse



fewer



fast

Low-cost future imagination



O_{t+1}



O_{t+2}

...



O_{t+k}

Low inference latency

$\Pi_{0.5}$

113

ms/chunk

motus

3215

ours

~100



Performance in simulation



$\Pi_{0.5}$

82.7

%



motus

88.7



ours

86.7

Performance in real world



$\Pi_{0.5}$

53.8

%



motus

63.8



ours

66.3