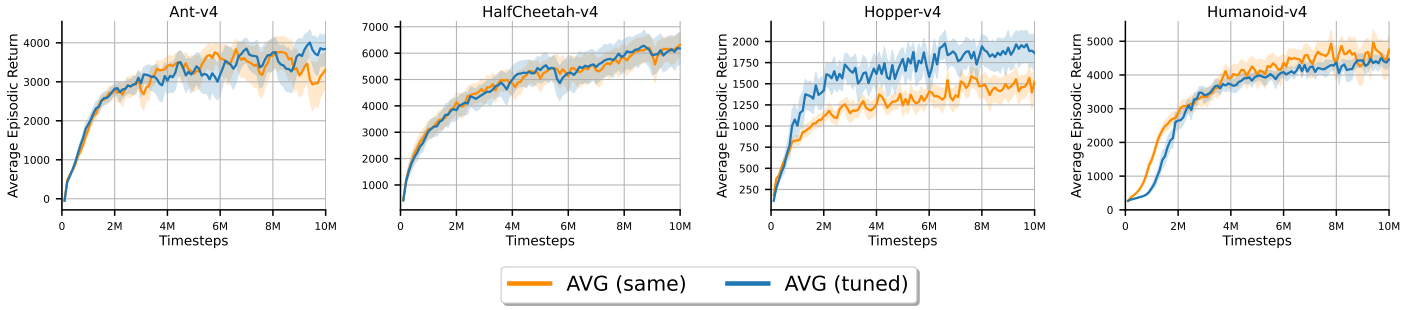

Figure 1: **AVG using the same hyper-parameters on four benchmark tasks.** The orange curves use the same hyper-parameter configuration, while the blue curves use the best hyper-parameters tuned using random search.

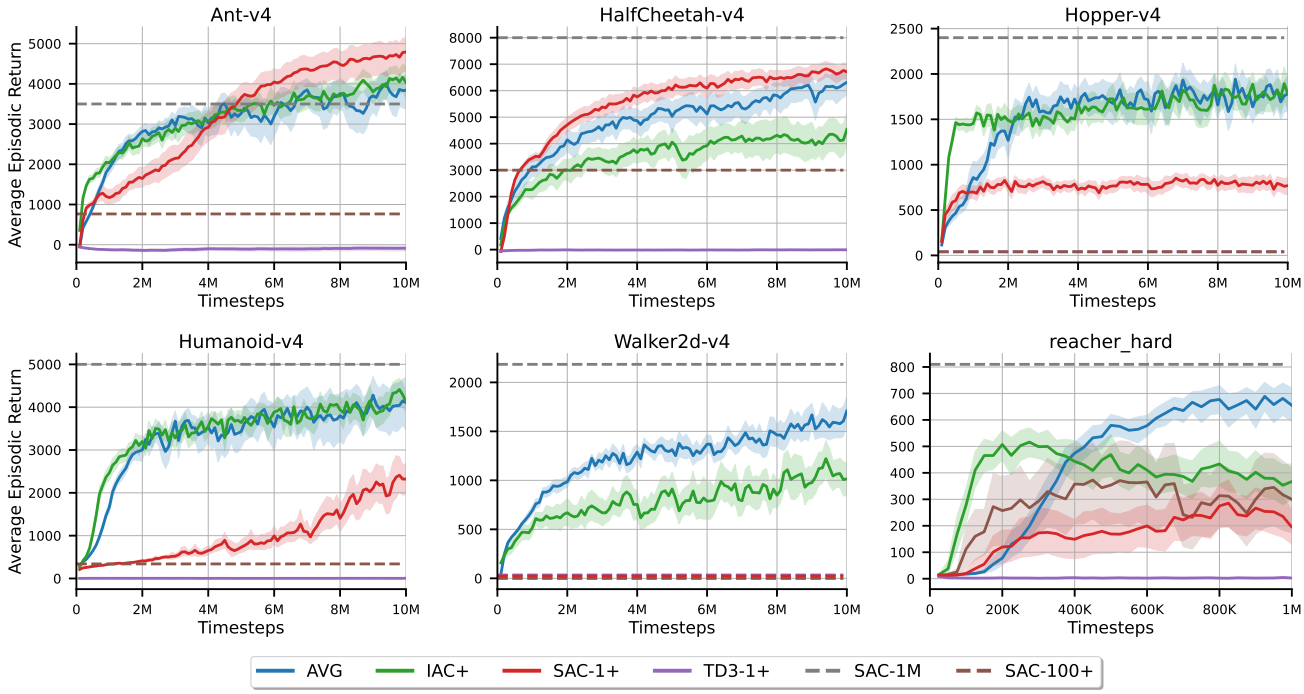

Figure 2: **Impact of normalization and scaling.** Suffix "+" denotes each algorithm plus normalization and scaling. We use random search to tune the hyper-parameters for all variants.

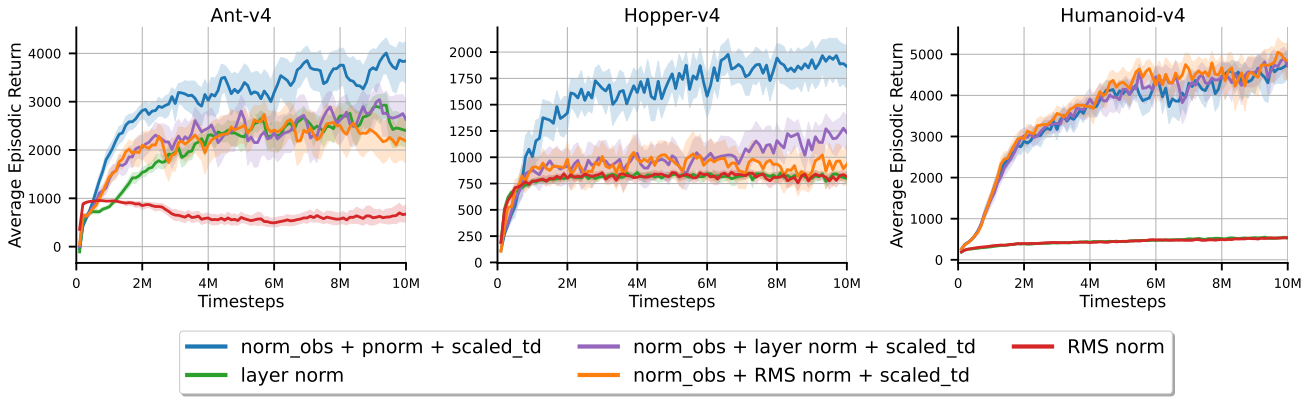

Figure 3: **Choice of normalization.** Each label indicates the normalization and scaling techniques applied to AVG. The solid blue curves depict the performance of our final proposed AVG. We use random search to tune the hyper-parameters for all variants. Note: *norm_obs* is normalized observation, *pnorm* is penultimate normalization, and *scaled_td* is return scaling.

Note: In all our plots, each solid learning curve is an average of 30 independent runs. The shaded regions represent a 95% confidence interval.