Supplementary Materials

The following content was not necessarily subject to peer review.

E Hyperparameter Details

In this section, we provide hyperparameter values for each MTRL scheme.

E.1 Meta-World MT-PPO

E.2 Meta-World MT-GRPO

E.3 Meta-World MT-PQN

E.4 Meta-World MT-SAC

E.5 Parkour MT-PPO

F Comparison to the original Meta-World

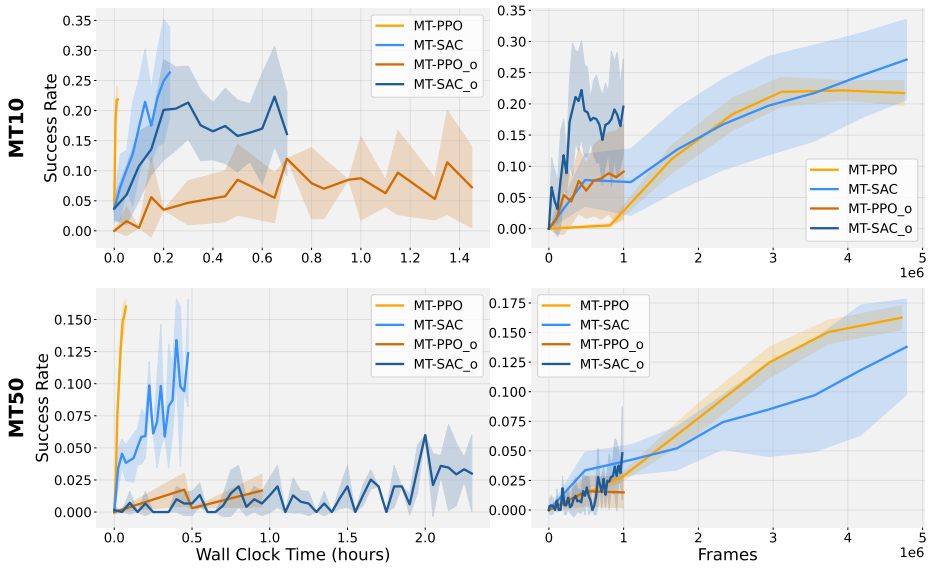

Figure 8: Comparison of our Meta-World to the original (_o) Meta-World using the hyperparameters listed from (Yu et al., 2021)

Description	value	variable_name
Number of environments	5120 / 6400	num_envs
Minibatch size	16384 / 25600	minibatch_size
Horizon length	32	horizon
Mini-epochs	5	mini_epochs
Number of epochs	1526 / 1221	max_epochs
Episode length	150	episodeLength
Discount factor	0.99	gamma
Clip ratio	0.2	e_clip
Policy entropy coefficient	.005	entropy_coef
Optimizer learning rate	5e-4	learning_rate
Optimizer learning schedule	fixed	lr_schedule
Advantage estimation tau	0.95	tau
Value Normalization by task	True	normalize_value
Input Normalization by task	True	normalize_input
Separate critic and policy networks	True	network.separate
CARE-Specific Hyperparameters		
Network hidden sizes	[400,400,400]	care.units
Mixture of Encoders experts	6	encoder.num_experts
Mixture of Encoders layers	2	encoder.num_layers
Mixture of Encoders hidden dim	50	encoder.D
Attention temperature	1.0	encoder.temperature
Post-Attention MLP hidden sizes	[50,50]	attention.units
Context encoder hidden sizes	[50,50]	context_encoder.units
Context encoder bias	True	context_encoder.bias
MOORE-Specific Hyperparameters		
MoE experts	4 / 6	moore.num_experts
MoE layers	3	moore.num_layers
MoE hidden dim	400	moore.D
Activation before/after task encoding weighting	[Linear, Tanh]	moore.agg_activation
Task encoder hidden sizes	[256]	task_encoder.units
Task encoder bias	False	task_encoder.bias
PaCO-Specific Hyperparameters		
Number of Compositional Vectors	5 / 20	paco.K
Network hidden dim	400	paco.D
Network layers	3	paco.num_layers
Task encoder bias	False	task_encoder.bias
Task encoder init	orthogonal	task_encoder.compositional_initializer
Task encoder activation	softmax	task_encoder.activation
Soft-Modularization-Specific Hyperparameters		
MoE experts	2	soft_network.num_experts
MoE layers	4	soft_network.num_layer
State encoder hidden sizes	[256,256]	state_encoder.units
Task encoder hidden sizes	[256]	task_encoder.units
PCGrad Hyperparameters		
Project actor gradient	False	project_actor_gradient
Project critic gradient	True	project_critic_gradient
CAGrad Hyperparameters		
Project actor gradient	False	project_actor_gradient
Project critic gradient	True	project_critic_gradient
Local ball radius for searching update vector	0.4	c
FAMO Hyperparameters		
Regularization coefficient	1e-3	gamma
Learning rate of the task logits	1e-3	w_lr
Small value for the clipping of the task logits	1e-2	epsilon
Normalize the task logits gradients	True	norm_w_grad

Table 3: Hyperparameters used for MTPPO. A '/' indicates the value used for MT10/MT50 respectively and otherwise is identical for each setting.

Description	value	variable_name
Number of environments	4096 / 6400	num_envs
Minibatch size	16384 / 25600	minibatch_size
Episode length	150	episodeLength
Horizon length	32	horizon
Mini-epochs	5	mini_epochs
Number of epochs	1908 / 1221	max_epochs
Episode length	150	
Discount factor	0.99	gamma
Clip ratio	0.2	e_clip
Policy entropy coefficient	.005	entropy_coef
Optimizer learning rate	5e-4	learning_rate
Optimizer learning schedule	fixed	lr_schedule
Advantage estimation tau	0.95	tau
Value Normalization by task	True	normalize_value
Input Normalization by task	True	normalize_input
Separate critic and policy networks	True	network.separate

Table 4: Hyperparameters used for MT-GRPO in MT10 / MT50. A '/' indicates the value used for MT10/MT50 respectively and otherwise is identical for each setting.

Description	value	variable_name
Number of environments	8192	num_envs
Gamma	.99	gamma
Peng's Q(lambda)	.5	q_lambda
Number of minibatches	4	num_minibatches
Episode length	500	episodeLength
Bang-off-Bang	3	binsPerDim
Action Scale	.005	actionScale
Mini epochs	8	mini_epochs Max grad norm
10.0	max_grad_norm	
Horizon	16	horizon
Start epsilon	1.0	start
End epsilon	0.005	end
Decay epsilon	True	decay_epsilon
Fraction of exploration steps	.005	exploration_fraction
Critic learning rate	3e-4	critic_lr
Anneal learning rate	True	anneal_lr
Value Normalization by task	False	normalize_value
Input Normalization by task	False	normalize_input
Use residual connections	True	q.residual_network
Number of LayerNormAndResidualMLPs	2	q.num_blocks
Network hidden dim	256	q.D
Batch norm input	False	q.norm_first_layer

Table 5: Hyperparameters used for MT-PQN in MT10.

Description	value	variable_name
Number of environments	4096	num_envs
Gamma	.99	gamma
Separate critic and policy networks	True	network.separate
Number of Gradient steps per epoch	32	gradient_steps_per_itr
Learnable temperature	True	learnable_ temperature
Use distangeled alpha	True	use_disentangled_alpha
Initial alpha	1	init_alpha
Alpha learning rate	5e-3	alpha_lr
Critic learning rate	5e-4	critic_lr
Critic tau	.01	critic_tau
Batch size	8192	batch_size
N-step reward	16	nstep
Grad norm	.5	grad_norm
Horizon	1	horizon
Value Normalization by task	True	normalize_value
Input Normalization by task	True	normalize_input
Replay Buffer Size	5000000	replay_buffer_size
Target entropy coef	1.0	target_entropy_coef

Table 6: Hyperparameters used for MT-SAC in MT10/MT50. A '/' indicates the value used for MT10/MT50 respectively and otherwise is identical for each setting. MT-SAC is very sensitive to the number of environments and replay ratio in the massively parallel regime.

Description	value	variable_name
Minibatch size	16384	minibatch_size
Horizon length	32	horizon
Mini-epochs	5	mini_epochs
Number of epochs	2000 / 4000	max_epochs
Episode length	800	
Discount factor	0.99	gamma
Clip ratio	0.2	e_clip
Policy entropy coefficient	.005	entropy_coef
Optimizer learning rate	5e-4	learning_rate
Optimizer learning schedule	adaptive	lr_schedule
Advantage estimation tau	0.95	tau
Value Normalization by task	False	normalize_value
Input Normalization by task	False	normalize_input
Separate critic and policy networks	True	network.separate
MOORE-Specific Hyperparameters		
MoE experts	2	moore.num_experts
MoE layers	2	moore.num_layers
MoE hidden dim	256	moore.D
Activation before/after task encoding weighting	[Linear, Linear]	moore.agg_activation
Task encoder hidden sizes	[128]	
Task encoder bias	False	task_encoder.bias
Multihead	False	multi$_{h}ead$
PaCO-Specific Hyperparameters		
Number of Compositional Vectors	5	paco.K
Network hidden dim	400	paco.D
Network layers	3	paco.num_layers
Task encoder bias	False	task_encoder.bias
Task encoder init	orthogonal	task_encoder.compositional_initializer
Task encoder activation	softmax	task_encoder.activation
Soft-Modularization-Specific Hyperparameters		
MoE experts	2	soft_network.num_experts
MoE layers	2	soft_network.num_layer
State encoder hidden sizes	[256,256]	state_encoder.units
Task encoder hidden sizes	[128]	task_encoder.units
PCGrad Hyperparameters		
Project actor gradient	False	project_actor_gradient
Project critic gradient	True	project_critic_gradient
CAGrad Hyperparameters		
Project actor gradient	False	project_actor_gradient
Project critic gradient	True	project_critic_gradient
Local ball radius for searching update vector	0.4	c
FAMO Hyperparameters		
Regularization coefficient	1e-4	gamma
Learning rate of the task logits	5e-3	w_lr
Small value for the clipping of the task logits	1e-3	epsilon
Normalize the task logits gradients	True	norm_w_grad

Table 7: Hyperparameters used for MTPPO in Parkour Benchmark. A '/' indicates the value used for Parkour-easy/Parkour-hard respectively and otherwise is identical for each setting.