Hyperparameter	Meaning
batch_size	Minibatch size
n_steps	The number of steps to run for each environment per update
gamma	Discount factor
learning_rate	Learning rate
n_epochs	Number of epoch when optimizing the surrogate loss
clip_range	Clipping parameter
gae_lambda	Factor for trade-off of bias vs variance for Generalized Advantage Estimator
buffer_size	Size of the replay buffer
tau	The soft update coefficient
train_freq	Update the model every train_freq steps
gradient_steps	How many gradient steps to do after each rollout
action_noise	the action noise type
learning_starts	how many steps of the model to collect transitions for before learning starts
n_quantiles	Number of quantiles for the critic
top_quantiles_to_drop_per_net	Number of quantiles to drop per network

Algorithm	Hyperparameter	Range of Values and Selection Methods
PPO	batch_size	{8, 16, 32, 64, 128, 256, 512}, select a fixed value
	n_steps	([64, 8192], 64), select with 64-step increment within the given range [64, 8192]
	gamma	([0.8, 0.9999], log), select with logarithmically uniform distribution within the given range [0.8, 0.9999]
	learning_rate	([10^{-5}, 1], log), select with logarithmically uniform distribution within the given range [10^{-5}, 1]
	n_epochs	{1, 5, 10, 20}, select a fixed value
	clip_range	[0.1, 0.4], randomly select values within the given range [0.1, 0.4]
	gae_lambda	[0.8, 0.99], randomly select values within the given range [0.8, 0.99]
DDPG	batch_size	{16, 32, 64, 128, 256, 512, 1024, 2048}, select a fixed value
	buffer_size	{10^4, 10^5, 10^6}, select a fixed value
	gamma	([0.85, 0.9999], log), select with logarithmically uniform distribution within the given range [0.85, 0.9999]
	learning_rate	([10^{-5}, 1], log), select with logarithmically uniform distribution within the given range [10^{-5}, 1]
	tau	{0.001, 0.005, 0.01, 0.02, 0.05, 0.08}, select a fixed value
	train_freq	{1, 4, 8, 16, 32, 64, 128, 256, 512}, select a fixed value
	gradient_steps	{1, 4, 8, 16, 32, 64, 128, 256, 512}, select a fixed value
	action_noise(noise_std)	[0, 1], randomly select values within the given range [0, 1]
SAC	batch_size	{16, 32, 64, 128, 256, 512, 1024, 2048}, select a fixed value
	buffer_size	{10^4, 10^5, 10^6}, select a fixed value
	gamma	{0.9, 0.95, 0.98, 0.99, 0.995, 0.999, 0.9999}, select a fixed value
	learning_rate	([10^{-5}, 1], log), select with logarithmically uniform distribution within the given range [10^{-5}, 1]
	learning_starts	{0, 1000, 10000, 20000}, select a fixed value
	tau	{0.001, 0.005, 0.01, 0.02, 0.05, 0.08}, select a fixed value
	train_freq	{1, 4, 8, 16, 32, 64, 128, 256, 512}, select a fixed value
	gradient_steps	{1, 4, 8, 16, 32, 64, 128, 256, 512}, select a fixed value
TQC	batch_size	{16, 32, 64, 128, 256, 512, 1024, 2048}, select a fixed value
	buffer_size	{10^4, 10^5, 10^6}, select a fixed value
	gamma	{0.9, 0.95, 0.98, 0.99, 0.995, 0.999, 0.9999}, select a fixed value
	learning_rate	([10^{-5}, 1], log), select with logarithmically uniform distribution within the given range [10^{-5}, 1]
	learning_starts	{0, 1000, 10000, 20000}, select a fixed value
	tau	{0.001, 0.005, 0.01, 0.02, 0.05, 0.08}, select a fixed value
	train_freq	{1, 4, 8, 16, 32, 64, 128, 256, 512}, select a fixed value
	gradient_steps	{1, 4, 8, 16, 32, 64, 128, 256, 512}, select a fixed value

	n_quantiles	[5, 50], select integers within the range [5, 50]
	top_quantiles_to_drop_per_net	[0, n_quantiles - 1], select integers within the range [0, n_quantiles - 1]
TD3	batch_size	{16, 32, 64, 100, 128, 256, 512, 1024, 2048}, select a fixed value
	buffer_size	{10^4, 10^5, 10^6}, select a fixed value
	gamma	{0.9, 0.95, 0.98, 0.99, 0.995, 0.999, 0.9999}, select a fixed value
	learning_rate	([10^{-5}, 1], log), select with logarithmically uniform distribution within the given range [10^{-5}, 1]
	tau	{0.001, 0.005, 0.01, 0.02, 0.05, 0.08}, select a fixed value
	train_freq	{1, 4, 8, 16, 32, 64, 128, 256, 512}, select a fixed value
	gradient_steps	{1, 4, 8, 16, 32, 64, 128, 256, 512}, select a fixed value
	action_noise(noise_std)	[0, 1], randomly select values within the given range [0, 1]

*The hyperparameter "action_noise" is specified as normal noise, where "noise_std" represents the standard deviation of the noise.