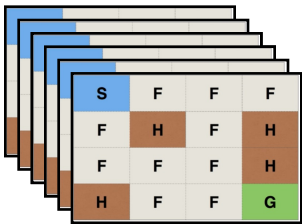
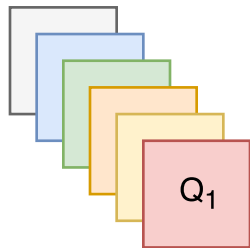
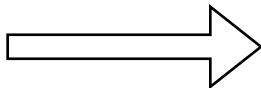


Map configurations
(train)



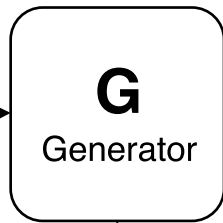
Q-learning



Real
Samples

(Q-table
policies)

z
latent space



Generated
Fake
Q-table
policies

D
Discriminator

*real or
fake policy
?*

Fine tune training

