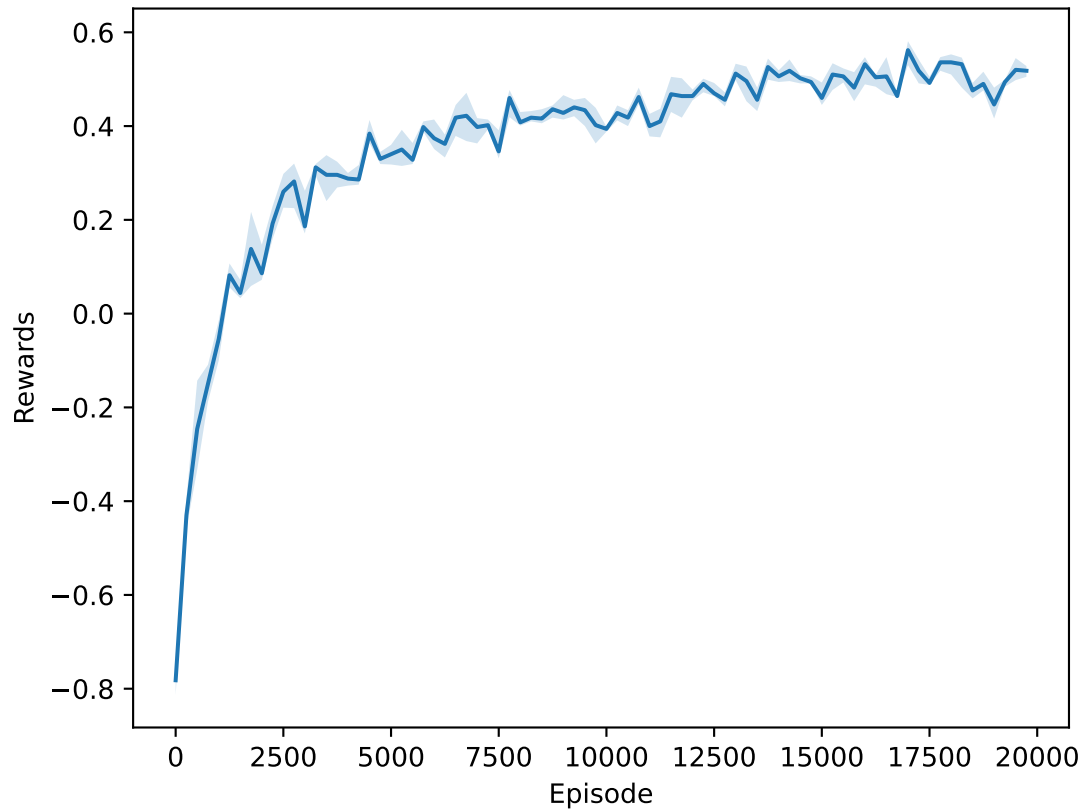
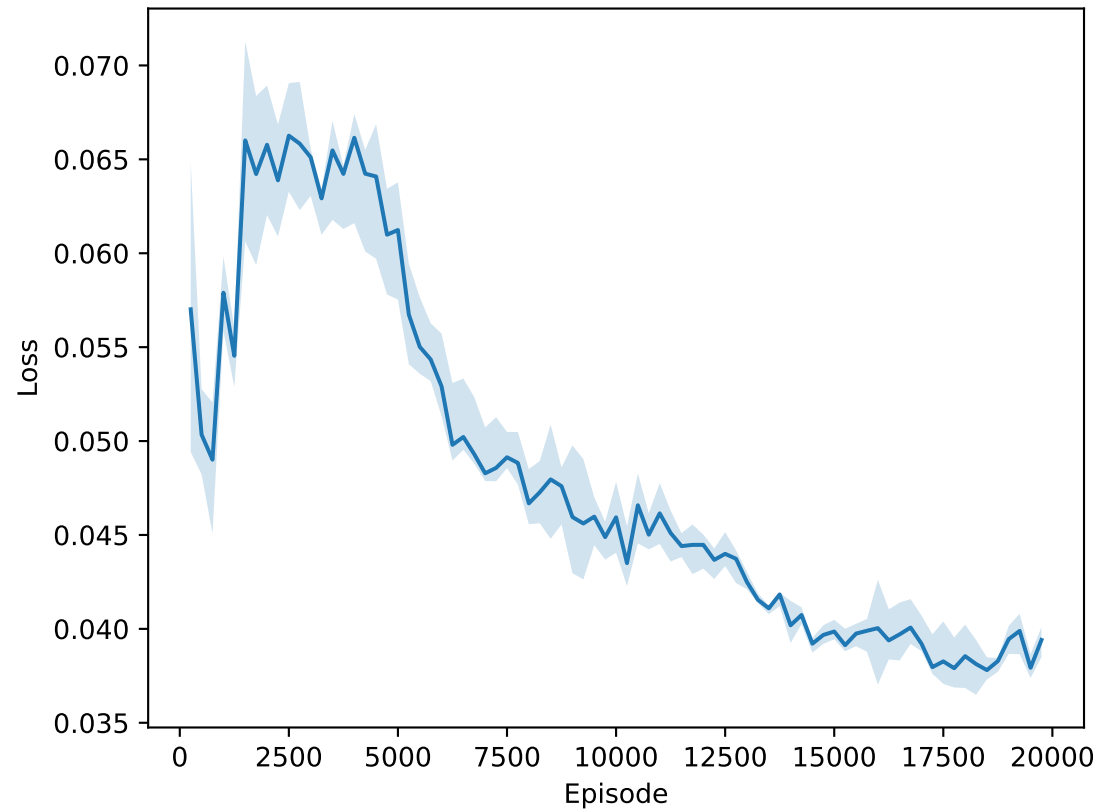


Average rewards during training



Average loss during training

 $\epsilon = 0.1$