Policy Loss and Entropy During Training