Effects of weight decay on the convolutional layer 4500 No weight decay $--- \eta_{inv}^{lr} = 1000$ $--- \eta_{inv}^{lr} = 500$ $--- \eta_{inv}^{lr} = 250$ 1500 1000 No weight decay 7500 2500 5000 η_{inv} forward