Initial learning rate	$\alpha = 0.5$
Regularization	$\lambda = 10^{-5}$
Hidden layer size of Ptr-Net	$h = 200$
Filter length of CNN	$l_f = 3, 4, 5$
Number of feature maps	$d_f = 128$
Hidden size of LSTM	$d_r = 200$
Size of embedding	$d_e = 100$
Beam size	$b = 64$
Batch size	128