Figure 1: The BLEU score plot paints a better picture for pre-training than the cross-entropy plot, since pre-training was able to improve the En-Zh BLEU score to be on par with the score of joint training for 600k steps.