A Impact of Length Prediction

The quality of length prediction has an impact on the overall translation/captioning performance. When using the reference target length (during inference), we consistently observed approximately 1 BLEU score improvement over reported results in the tables and figures across different datasets in the paper (see Table 6 for more detailed comparison).

We additionally compared our length prediction model with a simple baseline that uses length statistics of the corresponding training dataset (a non-parametric approach). To predict the target length for a source sentence with length L_s, we take the average length of all the target sentences coupled with the sources sentences of length L_s in the training set. Compared to this approach, our length prediction model predicts target length correctly twice as often (16% vs. 8%), and gives higher prediction accuracy within five tokens (83% vs. 69%)

	IWSLT'16		WMT'16		WMT'14	
	En→	→En	En→	→En	En→	→En
pred	27.01	32.43	29.66	30.30	21.54	25.43
ref	28.15	33.11	30.42	31.26	22.10	26.40

Table 6: BLEU scores on each dataset when using reference length (ref) and predicted target length (pred).