Method	$|\theta|$	Train	Test
Encoders (w/o Cross Sentence Attention)			
300D LSTM Encoder	3.0M	83.9	80.6
600D Gated BiLSTM + intra-att	12M	90.5	85.5
300D Gumbel LSTM TreeLSTM	2.9M	91.2	85.6
300D DISAN	2.4M	91.1	85.6
300D Residual Stacked Encoders	9.7M	89.8	85.7
600D Gumbel TreeLSTM	10M	93.1	**86.0**
300D CAFE (w/o Inter-Attention)	3.7M	87.3	<u>85.9</u>
Cross Sentence Attention (Single Models)			
100D LSTM with attention	250K	85.3	83.5
300D mLSTM	1.9M	92.0	86.1
200D DecompAtt	380K	89.5	86.3
200D DecompAtt + Intra-Att	580K	90.5	86.8
300D NTI-SLSTM-LSTM	3.2M	88.5	87.3
300D re-read LSTM	2.0M	90.7	87.5
BiMPM	1.6M	90.9	87.5
448D DIIN	4.4M	91.2	88.0
600D ESIM	4.3M	92.6	88.0
150D CAFE (Sum+2x200D MLP)	750K	88.2	87.7
200D CAFE (Sum+2x400D MLP)	1.4M	89.4	88.1
300D CAFE (Sum+2x600D MLP)	3.5M	89.2	<u>88.3</u>
300D CAFE (AvgMax+300D HN)	4.7M	89.8	**88.5**
Cross Sentence Attention (Ensemble Models)			
600D ESIM + 300D TreeLSTM	7.7M	93.5	88.6
BiMPM	6.4M	93.2	88.8
448D DIIN	17.0M	92.3	88.9
300D CAFE (Ensemble)	17.5M	92.5	**89.3**