Extended QANet and application on SQuAD 2.0

CS224N Winter 2019

Larry Jin, Yimin Liu, Su Jiang

Stanford ERE

INTRODUCTION

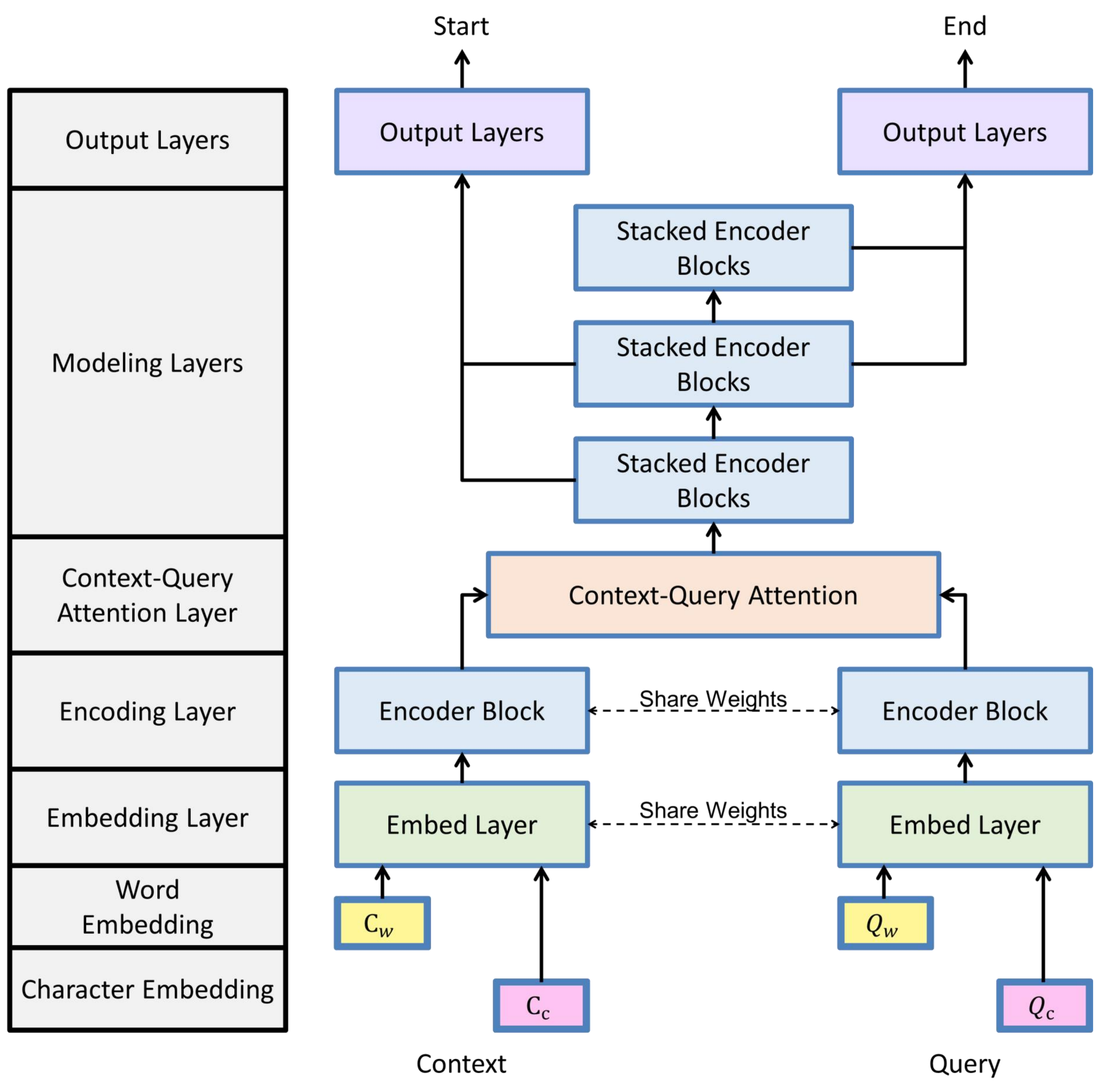
The topic of this project is extending the recent proposed QANet [1] and apply these methods on SQuAD 2.0 [2]. In this project, we propose a framework that combines the strength of transformer and RNN to conduct fast machine comprehension with consideration to sequential logic in the context.

DATASET AND METHOD

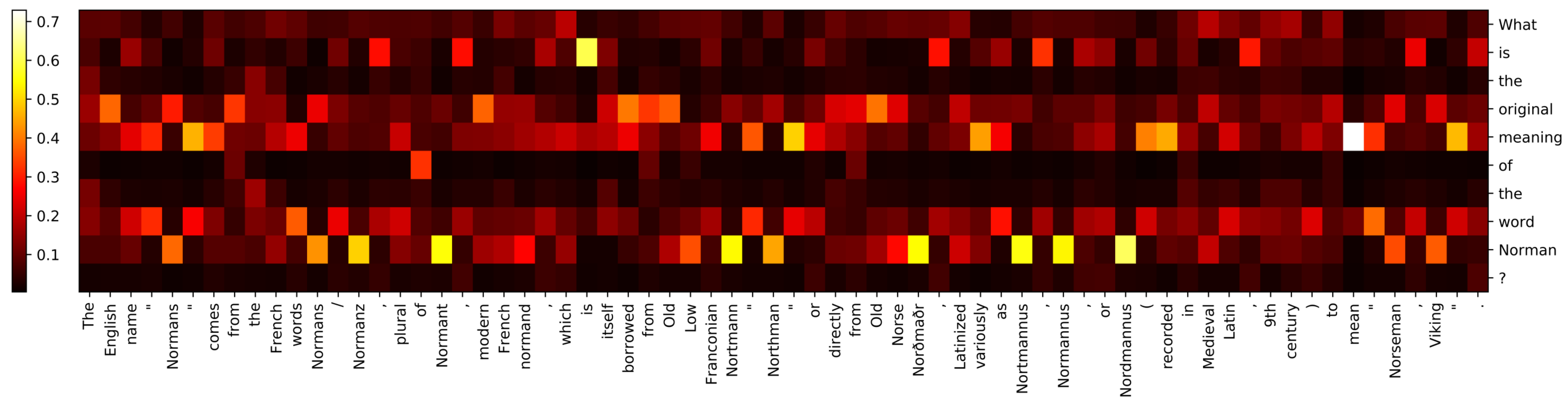
We apply the proposed method on SQuAD 2.0 [2].

Based on the QANet architecture, we made following modifications:

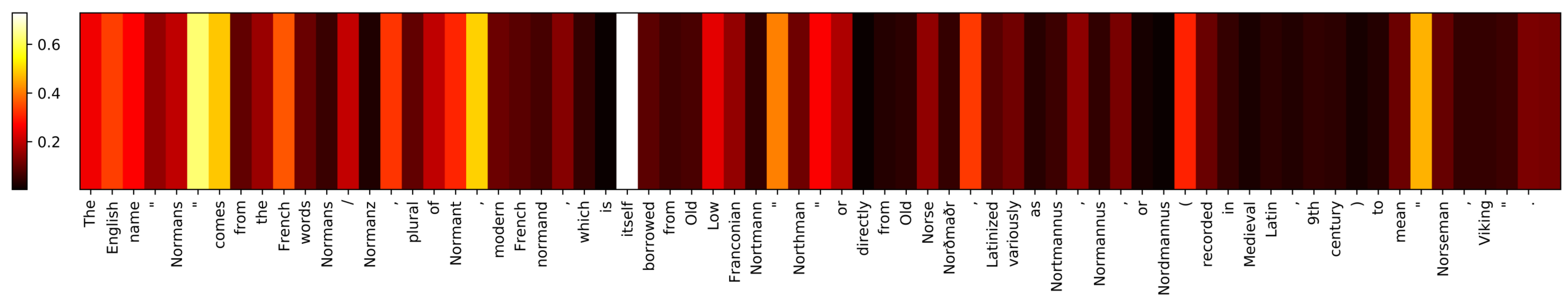
- add a RNN based contextual embedding layer in addition to the word-level and character-level embeddings.
- simplify the encoder blocks for both encoding and modeling layers with less stacks of ConvNet.
- replace the 1D ConvNet with GRU in the encoder block.



RESULTS



Context-to-query Attention (C2Q)



Query-to-context Attention (Q2C)

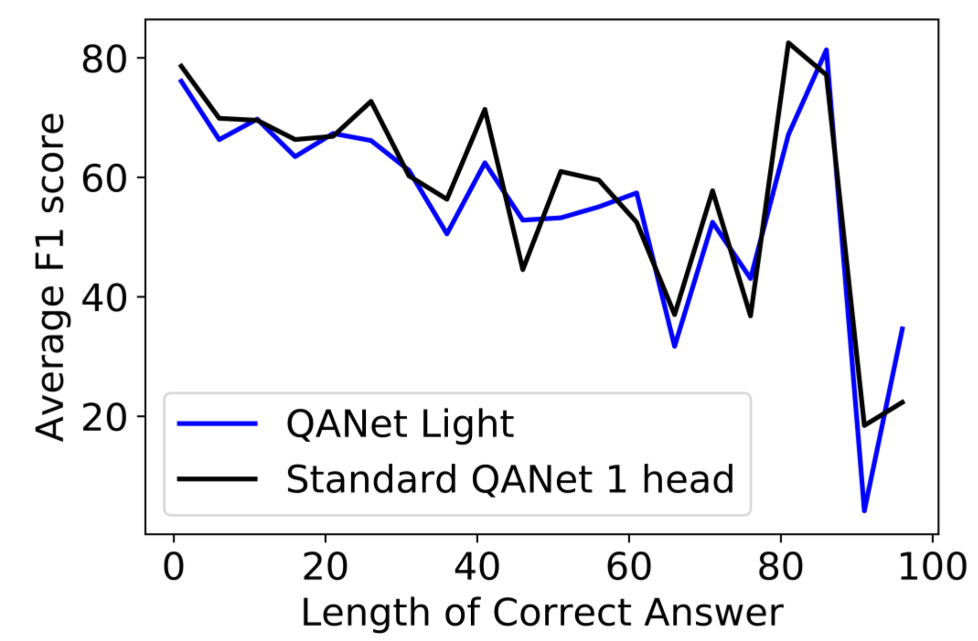
Model	EM	F1
Std QANet 8 hds	63.07	66.60
Std QANet 1 hd	63.15	66.72
QANet Light	60.56	64.30
QANet GRU	59.57	62.99
Ensemble model	64.26	67.60

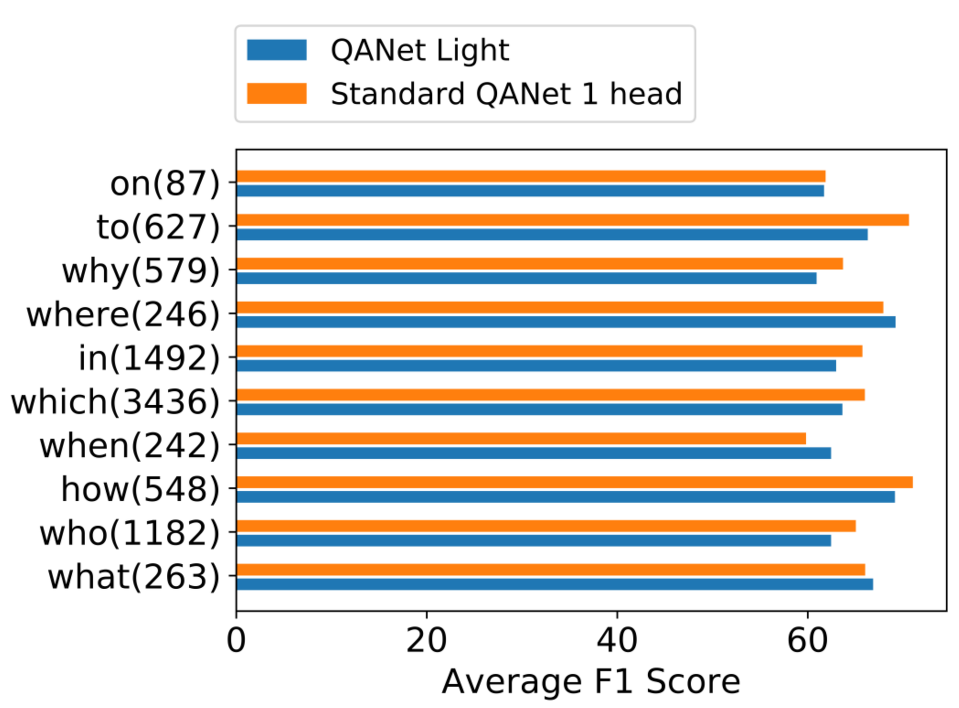
The table shows the EM and F1 results for all of the methods we applied in this project.

- Standard QANet 1 head performs best
- Applying ensemble model based on all of the methods achieve the highest F1 score and EM score.

- For context-to-query attention and query-to-context attention analysis, we show a example where the model correctly predicts the answer.
- The Q2C attention shows the words that have highest attention regarded to the query reasonably represent the key meaning of the query.
- The C2Q shows the relevant context words for query word may offer some reference to the meaning of the query word.

- As the increase of answer length, the F1 score decrease, which means the models are still not good at predicting longer deeper logic.
- Both methods get a high F1 score on 'to', 'how', 'where' problem. The Standard QANet 1 head result has a higher F1 score.





CONCLUSIONS

- Our implementation of QANet achieved near SoTA accuracy on SQuAD 2.0 dataset
- Ablation study conducted to analyze the functionality of important layers
- Insights drawn form visualization of Q2C and C2Q attentions
- Accuracy study conducted over question types and length of answer

ACKNOWLEDGEMENTS

Stanford Center for Computational Earth & Environmental Science (CEES)

REFERENCES

[1] Yu, A. W., Dohan, D., Luong, M. T., Zhao, R., Chen, K., Norouzi, M., & Le, Q. V. (2018). Qanet: Combining local convolution with global self-attention for reading comprehension. arXiv preprint arXiv:1804.09541.

[2] Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.