Transformer Based Question Answering Model

Emma Chen, Jennifer She

Problem

- Study the performance of **attention-based** models (inspired by **Transformer** and **QANet**) in solving the **SQuAD 2.0 Question and Answering Challenge**.
- Reproduce QANet as a competitive alternative to the LSTM-based baseline model **BiDAF**.
- Experiment with variants of QANet (and BiDAF).

Approach

1. **BiDAF + Char Emb:** Augment BiDAF baseline with character embeddings
2. **QANet + Char Emb:** Reproduce QANet
3. **QANet + Char Emb + Trainable Word Emb:** Allow word vectors to be trainable
4. **QANet + Char Emb + Conv Layers:** Replace depthwise separable conv layers with regular convolutions
5. **QANet + Char Emb + Recurrence:** Divide context into segments and feed sequentially into model encoding

QANet + Char Emb + Recurrence Model

Layout of an encoder Block

(From Yu et al. [2018])

(Modified from Yu et al. [2018])
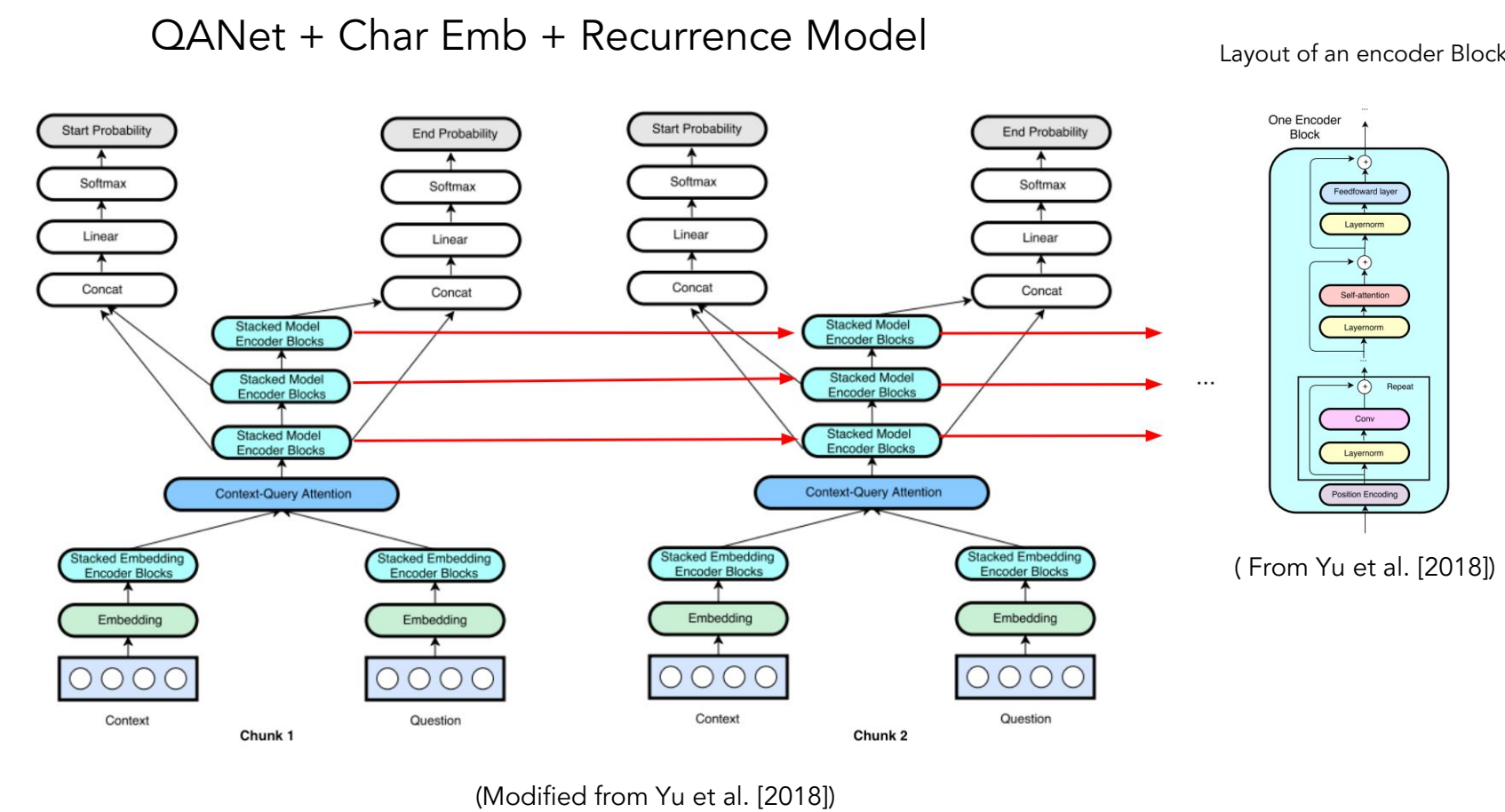
Results

Model	F1	EM
BiDAF + Char Emb	64.9	**61.4**
QANet + Char Emb	**65.1**	61.2
QANet + Char Emb + Trainable Word Emb	54.8	52.3
QANet + Char Emb + Conv Layers	51.8	51.6
QANet + Char Emb + Recurrence (4 Seg)	61.5	57.8

Data/Task

Data: SQuAD 2.0 dataset modified and provided by the course with train, dev and test set
- 140k examples of context-question-answer triples

Task: given a pair of context and question, correctly predict the answer:
- ~⅓ context-question pairs do not have an answer: the model should correctly predict N/A
- ~⅔ context-question pairs have answers: the model should correctly predict the starting + ending position of the answer in the context
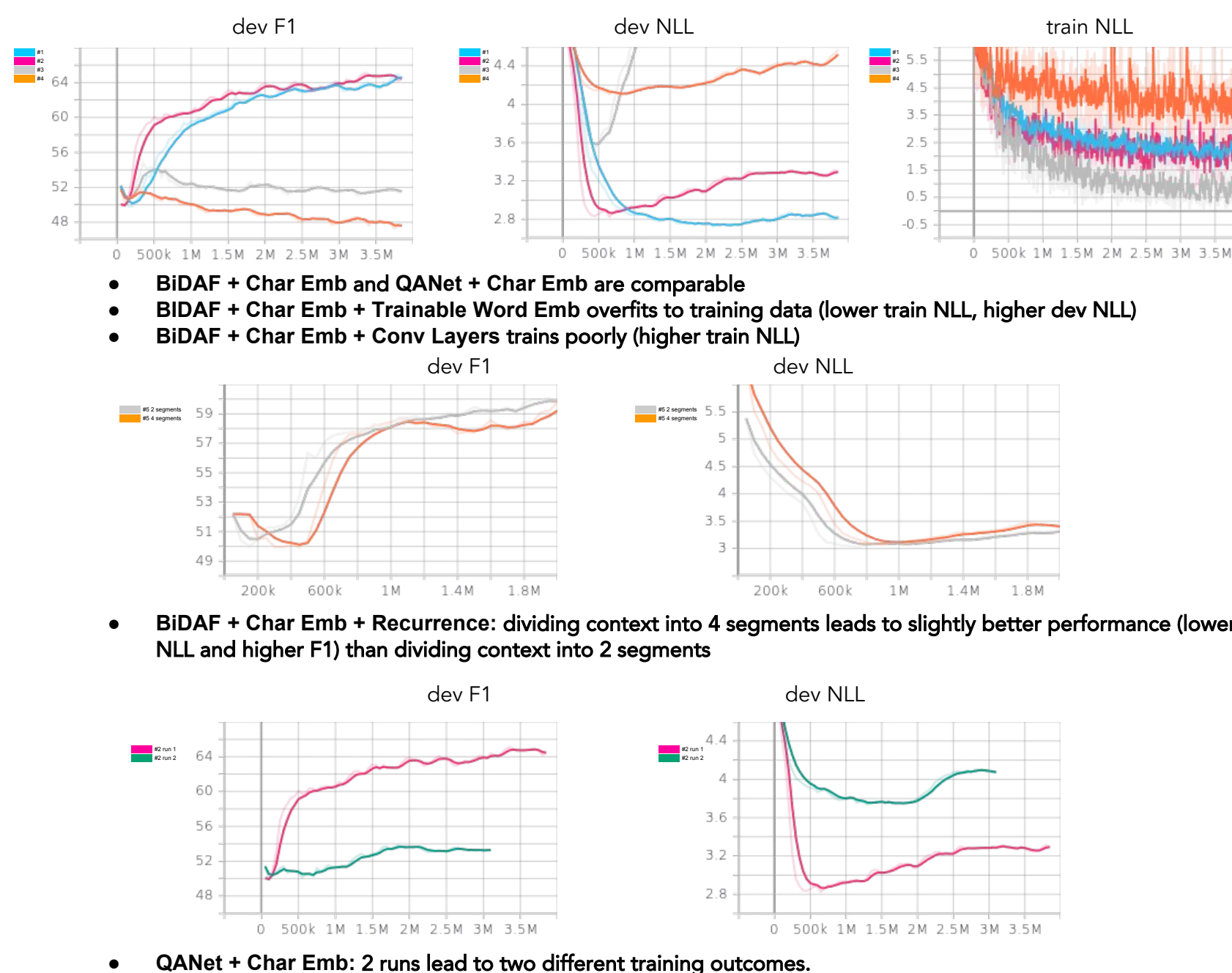
Example:

Question: What castle currently houses the Centre for Contemporary Art?

Context: The 17th century Royal Ujazdów Castle currently houses Centre for Contemporary Art, with some permanent and temporary exhibitions, concerts, shows and creative workshops. The Centre currently realizes about 500 projects a year. Zachęta National Gallery of Art, the oldest exhibition site in Warsaw, with a tradition stretching back to the mid-19th century organises exhibitions of modern art by Polish and international artists and promotes art in many other ways. Since 2011 Warsaw Gallery Weekend is held on last weekend of September.

Answer: Royal Ujazdów Castle

Analysis

dev F1   dev NLL   train NLL

- BiDAF + Char Emb and QANet + Char Emb are comparable
- BiDAF + Char Emb + Trainable Word Emb overfits to training data (lower train NLL, higher dev NLL)
- BiDAF + Char Emb + Conv Layers trains poorly (higher train NLL)

dev F1   dev NLL

- BiDAF + Char Emb + Recurrence: dividing context into 4 segments leads to slightly better performance (lower NLL and higher F1) than dividing context into 2 segments

dev F1   dev NLL

- QANet + Char Emb: 2 runs lead to two different training outcomes.

Question: When did King Harold II conquer England?

Context: In 1066, Duke William II of Normandy conquered England killing King Harold II at the Battle of Hastings. The invading Normans and their descendants replaced the Anglo-Saxons as the ruling class of England. The nobility of England were part of a single Normans culture and many had lands on both sides of the channel. Early Norman kings of England, as Dukes of Normandy, owed homage to the King of France for their land on the continent. They considered England to be their most important holding (it brought with it the title of King—an important status symbol).

Answer: N/A

Prediction: 1066

QANet + Char Emb tends to predict answer to questions that have no answer
—especially "when" and "what" questions that appear easy

Question: Why did OPEC block oil deliveries to the United States?

Context: Question: Why did OPEC block oil deliveries to the United States?

Context: In response to American aid to Israel, on October 16, 1973, OPEC raised the posted price of oil by 70%, to $5.11 a barrel. The following day, oil ministers agreed to the embargo, a cut in production by five percent from September's output and to
…

1973. At their Kuwait meeting, OAPEC proclaimed the embargo that curbed exports to various countries and blocked all oil deliveries to the US as a "principal hostile country".

Answer: American aid to Israel

Prediction: principal hostile country

QANet + Char Emb tends to favor text near question keywords in the context

Conclusions

1. Adding character embeddings improve performance of both BiDAF and QANet
2. **QANet + Char Emb** is competitive with BiDAF + Char Emb
3. Making word embeddings trainable leads to overfitting
4. Replacing depthwise separable convolutions with regular convolutions leads to worse learning
5. Incorporating recurrence into QANet model encoding works reasonably well (also better with smaller segments / more recurrence)

References

C. Chute. squad-transformer. 2018. https://github.com/chrischute/squad-transformer.

Z. Dai, Z. Yang, Y. Yang, W. W. Cohen, J. Carbonell, Q. V. Le, and R. Salakhutdinov. Transformer-xl: Language modeling with longer-term dependency. 2018.

Y.-H. Huang. attention-is-all-you-need-pytorch. 2018. https://github.com/jadore801120/attention-is-all-you-need-pytorch.

M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi. Bidirectional attention flow for machine comprehension. arXiv preprint arXiv:1611.01603, 2016.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems, pages 5998–6008, 2017.

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