

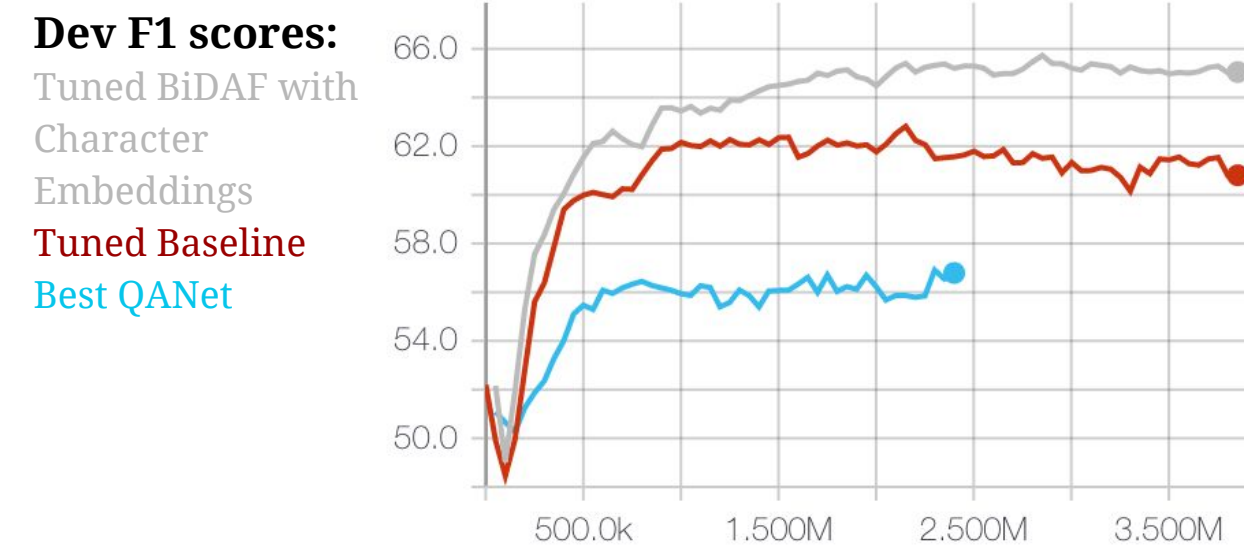
## Introduction

- **Goal:** Build a system for question answering as well as identify unanswerable questions in the SQuAD 2.0 dataset.
- **Question answering:** Identifying a sequence in a context paragraph that correctly answers an input query.

## Approach

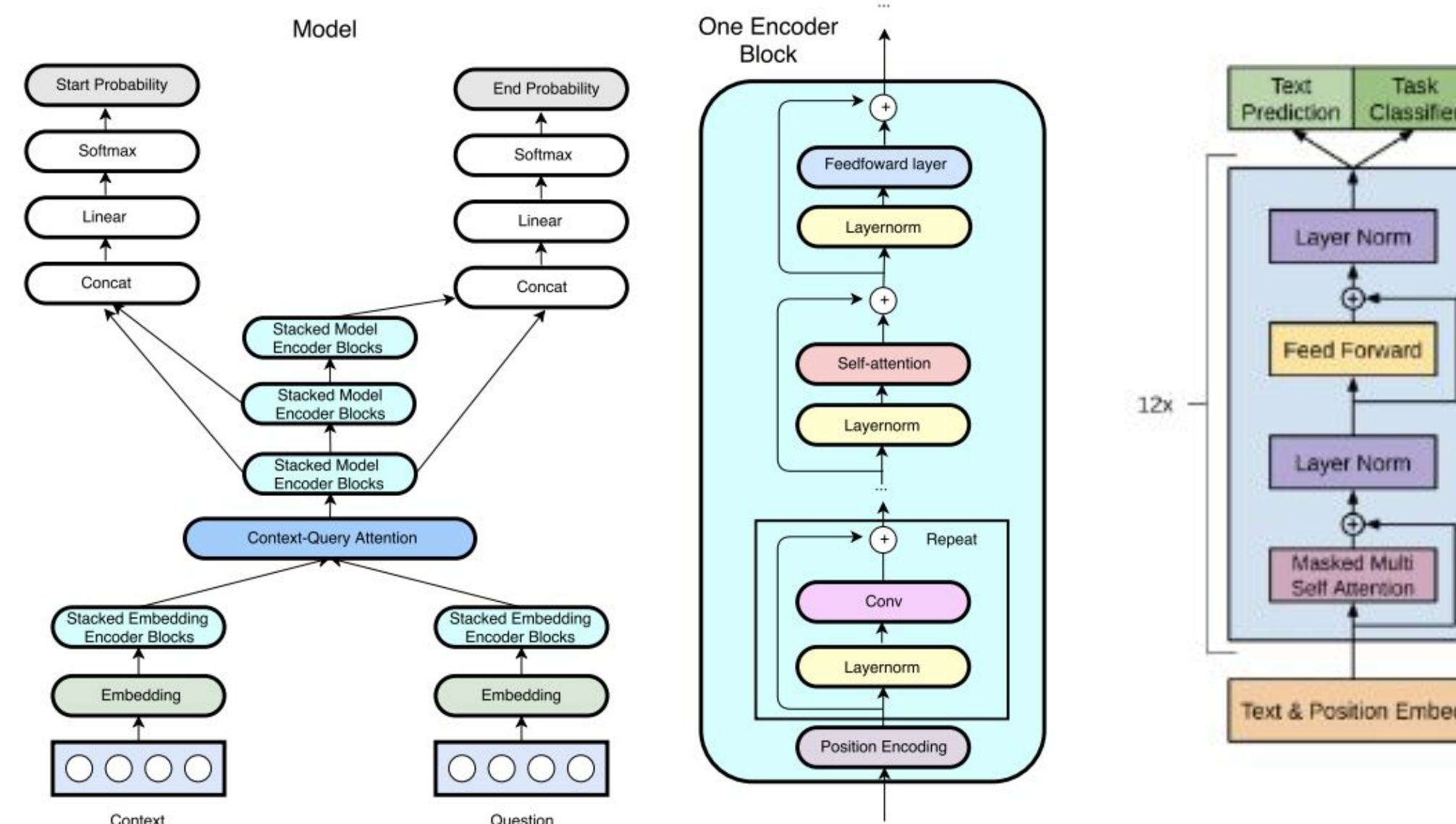
- **Evaluation Metric:** F1 and EM scores on the Dev and Test set.
- **Data:** SQuAD 2.0, a dataset of over 150,000 questions, both answerable and unanswerable.
- **Word Embeddings:** Stanford's pretrained GloVe embeddings.
- **Experimental Details:**
  - Adam Optimizer with learning rate warmup and weight decay.
  - Character-level embedding ablation.
  - Gradient accumulation for larger effective batch sizes.
  - Variation in number of heads, hidden size, encoder repetition.
  - Stochastic Depth Dropout.
  - Weight initialization ablation testing.

## Results



Model	F1 Score	EM
Baseline BiDAF	58	55
Tuned BiDAF	62.814	59.267
Tuned BiDAF + Char Embeddings	65.731	62.225
QANet	55.64	52.26
QANet + Char Embeddings	56.91	53.18

## Model



### QANet Model:

- Embedding Layer
- Embedding Encoder
- Context-Query Attention Layer
- Model Encoder Layer
- Output Layer

### Answer Verifier:

- Embedding Layer
- Transformer Layer
- Output Layer
- Option to pre-train as a language model

## Challenges/Conclusions

- The model was too large to fit onto any of the GPUs provided with a reasonable batch size. As a result, training took much longer than anticipated. This defeated the purported speed gains of the non-sequential architecture.
- The long training cycle (~12 hours) made debugging and tuning very difficult. We often had to try multiple changes at once which made it difficult to narrow down our problems.
- We found the transformer architecture to be extremely fragile and small tunes to the hyper parameters made drastic positive and negative changes.
- Future work would include incorporating the verifier with our BiDAF and QANet models, with a possible ensemble approach.

## Error Analysis

- **Misidentified Specifier:**

**Question:** What sort of motion did Newcomen's steam engine continuously produce?

**Context:** In 1781 James Watt patented a steam engine that produced continuous rotary motion.

**Answer:** N/A

**Prediction:** rotary
- **Incorrect Answer Length:**

**Question:** What is the Chinese name for the Yuan dynasty?

**Context:** The Yuan dynasty (Chinese: 元朝; pinyin: Yuán Cháo), officially the Great Yuan (Chinese: 大元; pinyin: Dà Yuán; Mongolian: Yehe Yuan Ulus[a])...

**Answer:** Yuán Cháo

**Prediction:** 元朝; pinyin: Yuán Cháo), officially the Great Yuan

### References Cited

1. Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V. Le. Qanet: Combining local convolution with global self-attention for reading comprehension. CoRR, abs/1804.09541, 2018.
2. Minghao Hu, Furu Wei, Yuxing Peng, Zhen Huang, Nan Yang, and Ming Zhou. Read + verify: Machine reading comprehension with unanswerable questions. CoRR, abs/1808.05759, 2018.