

A Model and Training Details

SQUAD: Reading Comprehension The model we use for SQUAD is the Document Reader of DRQA (Chen et al., 2017). We use the open source implementation from <https://github.com/hitvoice/DrQA>. The model represents the words of the paragraph as the concatenation of: GLOVE embeddings (Pennington et al., 2014), other word-level features, and attention scores between the words in the paragraph and the question. The model runs a BiLSTM over the sequence and takes the resulting hidden states at each time step as a word’s final vector representation. Two classifiers then predict the beginning and end of the answer span independently. The span with the highest product of beginning and end probabilities is selected as the answer. In input reduction, we keep the exact same span.

SNLI: Textual Entailment The model we use for SNLI is the Bilateral Multi-Perspective Matching Model (BiMPM) (Wang et al., 2017). We use the open source implementation from <https://github.com/galsang/BiMPM-pytorch>. The model runs a BiLSTM over the GLOVE word embeddings of both the premise and hypothesis. It then computes “matches” between the two sentences at multiple perspectives, aggregates the results using another BiLSTM, and feeds the resulting representation through fully-connected layers for classification.

VQA: Visual Question Answering The model we use for VQA is Show, Ask, Attend and Answer (Kazemi and Elqursh, 2017). We use the open source implementation from <https://github.com/Cyanogenoid/pytorch-vqa>. The model uses a ResNet-152 (He et al., 2016) to represent the image and an LSTM to represent the natural language question. The model then computes attention regions based on the two representations. All representations are finally combined to predict the answer.

Entropy Regularized Fine-Tuning To optimize the objective of Equation 2, we alternate updates on the two terms. We update the model on two batches of regular examples for the first term (maximum likelihood), followed by two batches of reduced examples for the second term (maximum entropy). The batches of reduced examples are randomly sampled from the collection of re-

duced inputs. We use two separate Adam (Kingma and Ba, 2015) optimizers for the two terms. For SQUAD and SNLI, we use a learning rate of $2e^{-4}$ and λ of $1e^{-3}$. For VQA, we use a learning rate of $1e^{-4}$ and λ of $1e^{-4}$.

B More Examples

This section provides more examples of input reduction. We show the original and reduced examples, which are generated on the models both *Before* and *After* fine-tuning with the entropy regularization from Section 4. All examples are correctly classified by the model.

SQUAD

Context	In 1899, John Jacob Astor IV invested \$ 100,000 for Tesla to further develop and produce a new lighting system. Instead, Tesla used the money to fund his Colorado Springs experiments .
Original	What did Tesla spend Astor's money on ? (0.78)
Before	did
After	spend Astor money on ?
Context	The Colorado experiments had prepared Tesla for the establishment of the trans-Atlantic wireless telecommunications facility known as Wardenclyffe near Shoreham, Long Island.
Original	What did Tesla establish following his Colorado experiments ?
Before	experiments ?
After	What Tesla establish experiments ?
Context	The Broncos defeated the Pittsburgh Steelers in the divisional round, 2316, by scoring 11 points in the final three minutes of the game. They then beat the defending Super Bowl XLIX champion New England Patriots in the AFC Championship Game, 2018, by intercepting a pass on New England's 2-point conversion attempt with 17 seconds left on the clock. Despite Manning's problems with interceptions during the season, he didn't throw any in their two playoff games.
Original	Who did the Broncos defeat in the AFC Championship game ?
Before	Who the defeat the
After	Who Broncos defeat AFC Championship game
Context	In 2014, economists with the Standard & Poor's rating agency concluded that the widening disparity between the U.S.'s wealthiest citizens and the rest of the nation had slowed its recovery from the 2008-2009 recession and made it more prone to boom-and-bust cycles . To partially remedy the wealth gap and the resulting slow growth, S&P recommended increasing access to education. It estimated that if the average United States worker had completed just one more year of school, it would add an additional \$105 billion in growth to the country's economy over five years.
Original	What is the United States at risk for because of the recession of 2008 ?
Before	is the risk the
After	What risk because of the 2008
Context	The Central Region, consisting of present-day Hebei, Shandong, Shanxi, the south-eastern part of present-day Inner Mongolia and the Henan areas to the north of the Yellow River, was considered the most important region of the dynasty and directly governed by the Central Secretariat (or Zhongshu Sheng) at Khanbaliq (modern Beijing); similarly, another top-level administrative department called the Bureau of Buddhist and Tibetan Affairs (or Xuanzheng Yuan) held administrative rule over the whole of modern-day Tibet and a part of Sichuan, Qinghai and Kashmir.
Original	Where was the Central Secretariat based ?
Before	the based
After	was Central Secretariat based
Context	It became clear that managing the Apollo program would exceed the capabilities of Robert R. Gilruth's Space Task Group, which had been directing the nation's manned space program from NASA's Langley Research Center. So Gilruth was given authority to grow his organization into a new NASA center, the Manned Spacecraft Center (MSC). A site was chosen in Houston, Texas , on land donated by Rice University, and Administrator Webb announced the conversion on September 19, 1961. It was also clear NASA would soon outgrow its practice of controlling missions from its Cape Canaveral Air Force Station launch facilities in Florida, so a new Mission Control Center would be included in the MSC.
Original	Where was the Manned Spacecraft Center located ?
Before	Where
After	Where was Manned Center located

SNLI

Premise	a man in a black shirt is playing a guitar .
Original	the man is wearing a blue shirt .
Answer	contradiction
Before	blue
After	man blue shirt
Premise	two female basketball teams watch suspenseful as the basketball nears the basket .
Original	two teams are following the movement of a basketball .
Answer	entailment
Before	.
After	movement
Premise	trucks racing
Original	there are vehicles .
Answer	entailment
Before	are
After	vehicles
Premise	a woman , whose face can only be seen in a mirror , is applying eyeliner in a dimly lit room .
Original	a man is playing softball .
Answer	contradiction
Before	man playing
After	a man
Premise	a tan dog jumping over a red and blue toy
Original	a dog playing
Answer	entailment
Before	dog
After	playing
Premise	a man in a white shirt has his mouth open and is adjusting dials .
Original	a man is sleeping .
Answer	contradiction
Before	sleeping
After	man sleeping

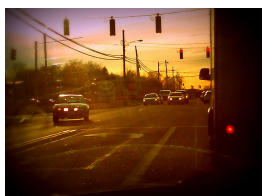
VQA



Original what color is the flower
 Answer yellow
 Before flower
 After what color is flower



Original what letter is on the lanyard
 Answer g
 Before letter is on lanyard
 After what letter is the lanyard



Original are the cars lights on
 Answer yes
 Before cars
 After lights



Original if the man steps forward, will he fall onto the track
 Answer no
 Before if the fall the
 After if the steps forward, fall onto the



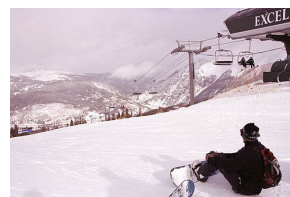
Original do this take batteries
 Answer yes
 Before batteries
 After batteries



Original are there any clouds in the sky
 Answer yes
 Before are clouds
 After there clouds sky



Original what is the man riding
 Answer elephant
 Before man riding
 After what riding



Original what color is the skiers backpack
 Answer black
 Before color
 After what color is the skiers



Original what kind of pants is the girl wearing
 Answer jeans
 Before kind of pants wearing
 After what kind of pants



Original are the giants playing
 Answer yes
 Before are
 After giants playing