

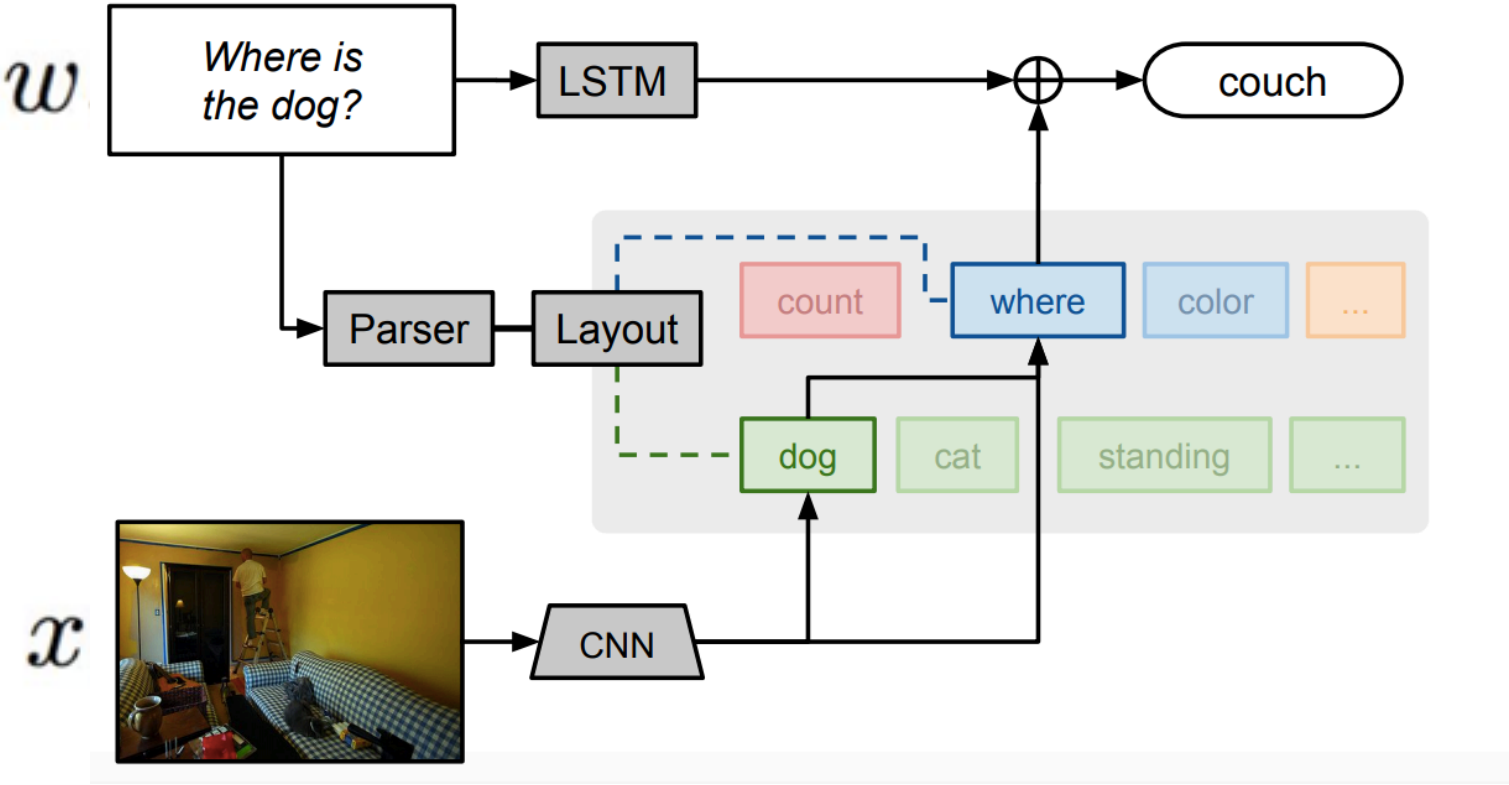
Neural Module Networks for Reasoning Over Text

Nitish Gupta , Kevin Lin , Dan Roth , Sameer Singh & Matt Gardner

Presented by:
Jigyasa Gupta

Neural Modules

- Introduced in the paper “**Deep Compositional Question Answering with Neural Module Networks**” by Jacob Andreas, Marcus Rohrbach, Trevor Darrell, Dan Klein for Visual QA task



Model = collection of modules + network layout predictor

Output = distribution over answers

$$p(y | w, x; \theta)$$

Motivation : Compositional Nature of VQA

Each question requires different # of reasoning steps and different *kinds* of reasoning

“Is this a truck?”

1. Classify
 - Convolutional

“How many objects are to the left of the toaster?”

1. Find toaster
2. Look left
3. Find objects
4. Count
 - Convolutional
 - Recurrent

Motivation : Compositional Nature of VQA

Questions are composed of “reasoning modules” and might share substructures

Where is the dog?

What color is the dog?

Where is the cat?

Motivation: Combine Both Approaches

Representational Approaches

Neural network structures are not universal, but at least modular in applications

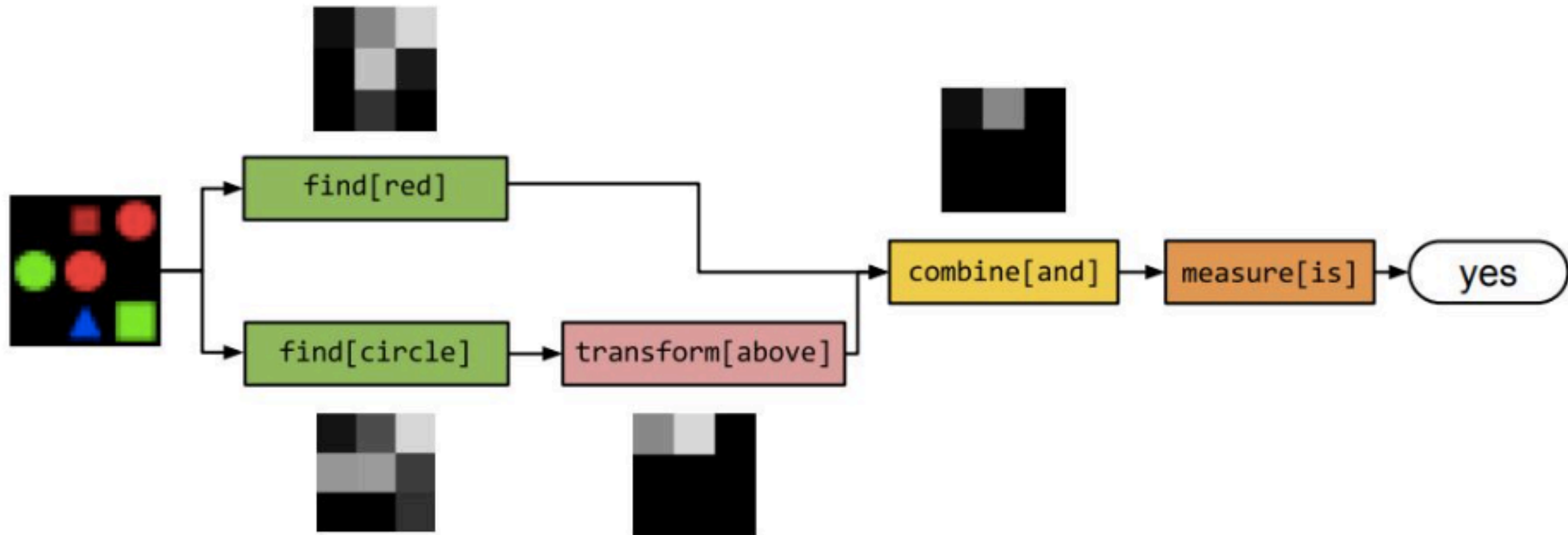
Compositional Approaches

Logical expressions and functional programs provide computational structure

Proposed Approach

1. Predict computational structure from question
2. Construct modular neural network from this structure

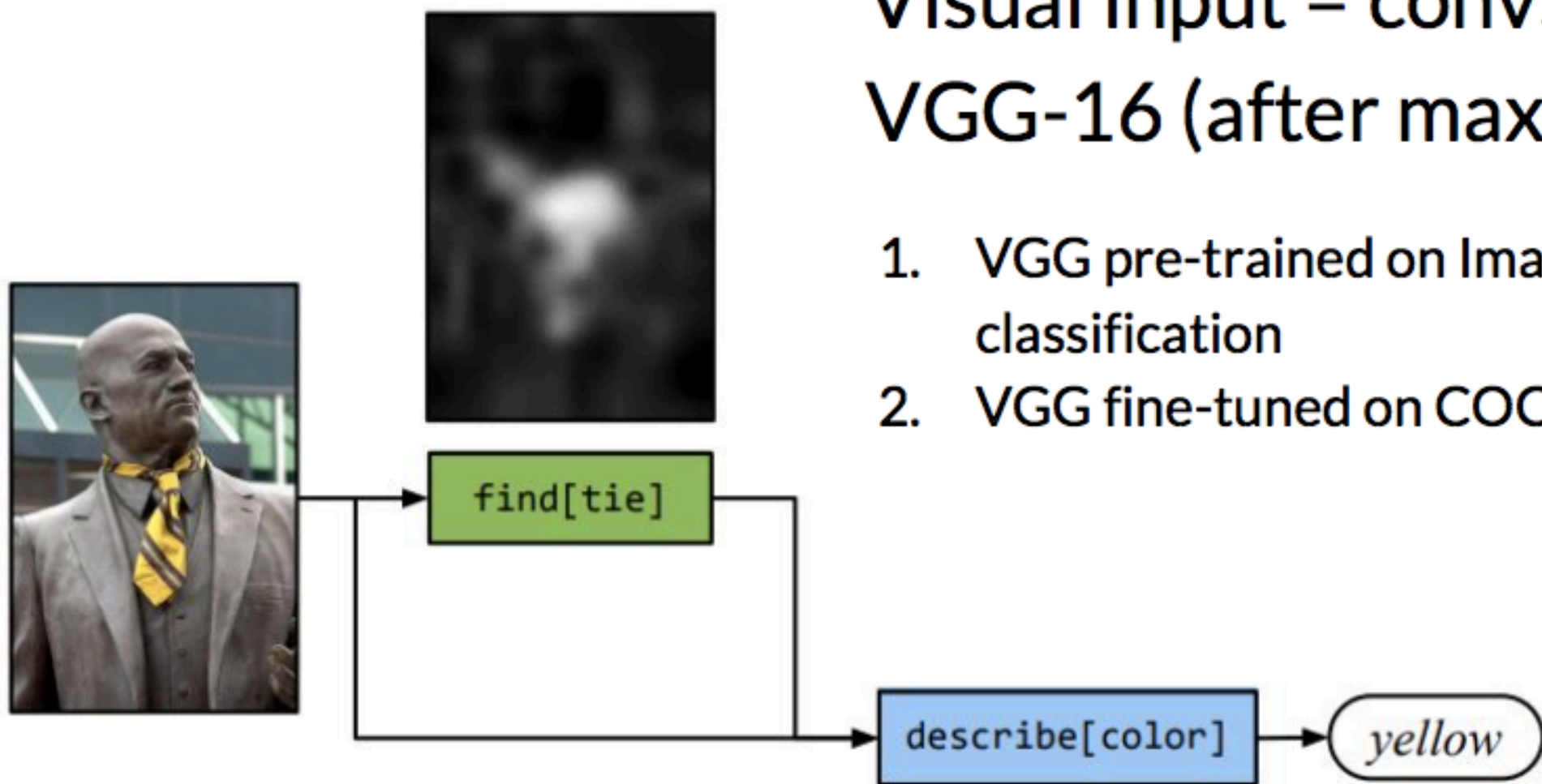
Is there a red shape above a circle?



What color is his tie?

Visual input = conv5 layer of VGG-16 (after max-pooling)

1. VGG pre-trained on ImageNet classification
2. VGG fine-tuned on COCO for captioning



Modules

Data Types:

1. Images
2. Attentions
3. Labels

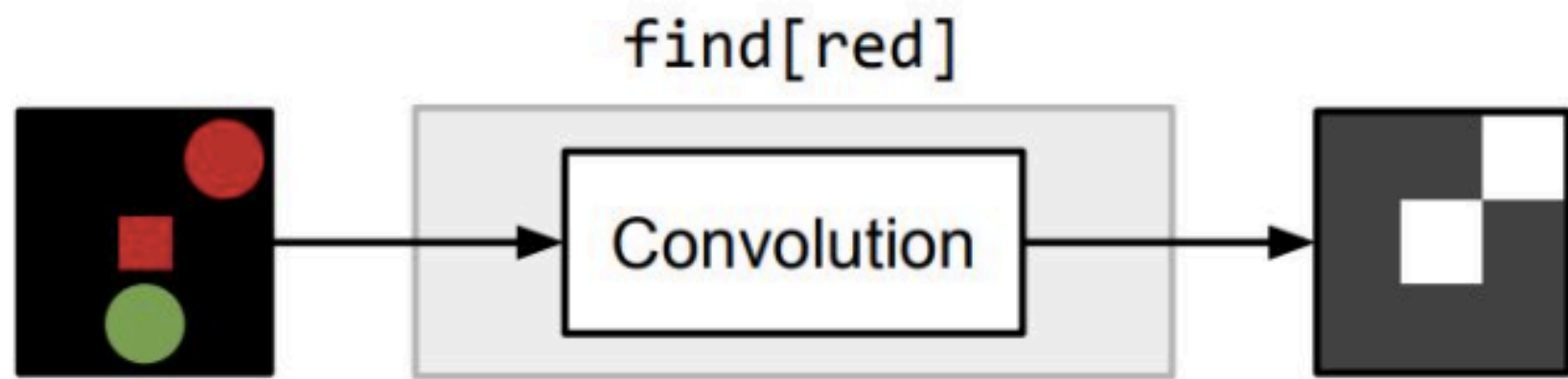
Module Types:

- Attention (Find)
- Re-Attention (Transform)
- Combination
- Classification (Describe)
- Measurement

TYPE[INSTANCE](ARG₁, ...)

Find Module

Image \rightarrow Attention

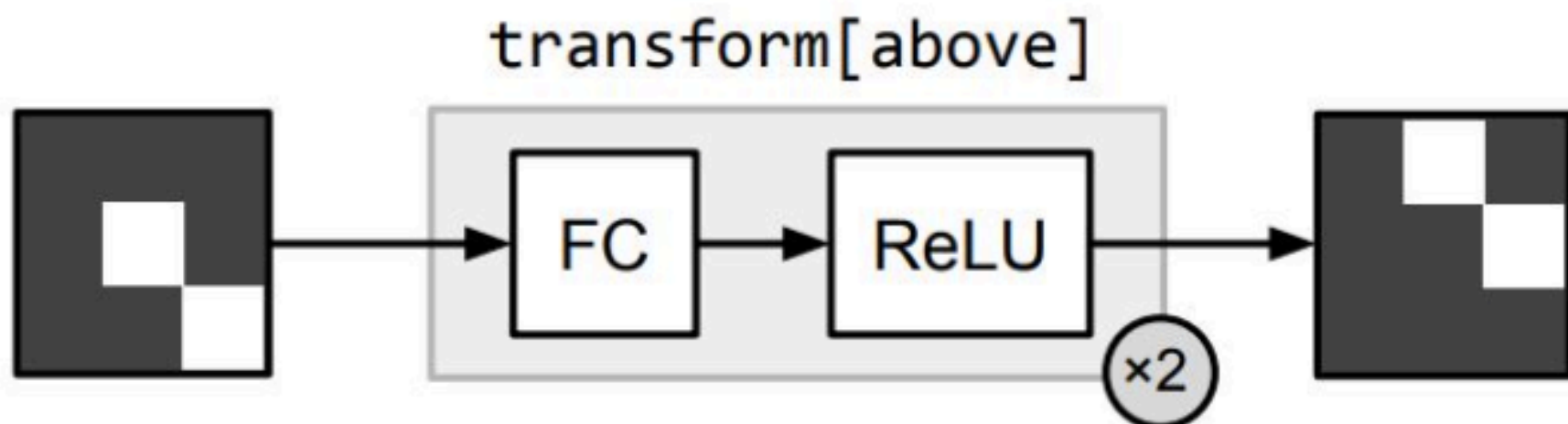


Convolve every position in input image w/ weight vector (distinct for each instance) to produce attention heatmap

`find[dog]` = matrix with high values in regions containing dogs

Transform Module

Attention \rightarrow **Attention**

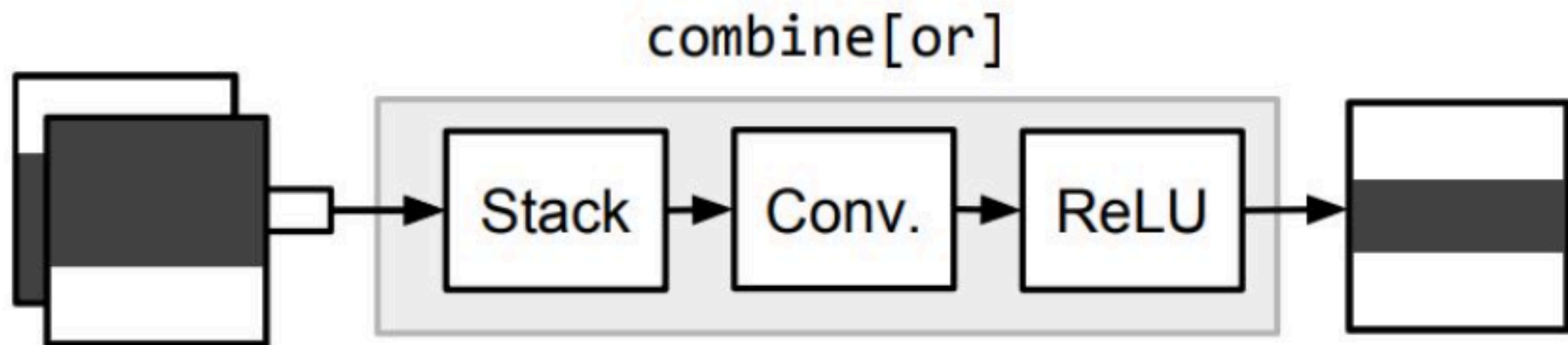


Fully connected mapping from one attention to another (MLP w/ ReLUs)

transform[not]: move attention away from active regions

Combine Module

Attention x Attention \rightarrow Attention

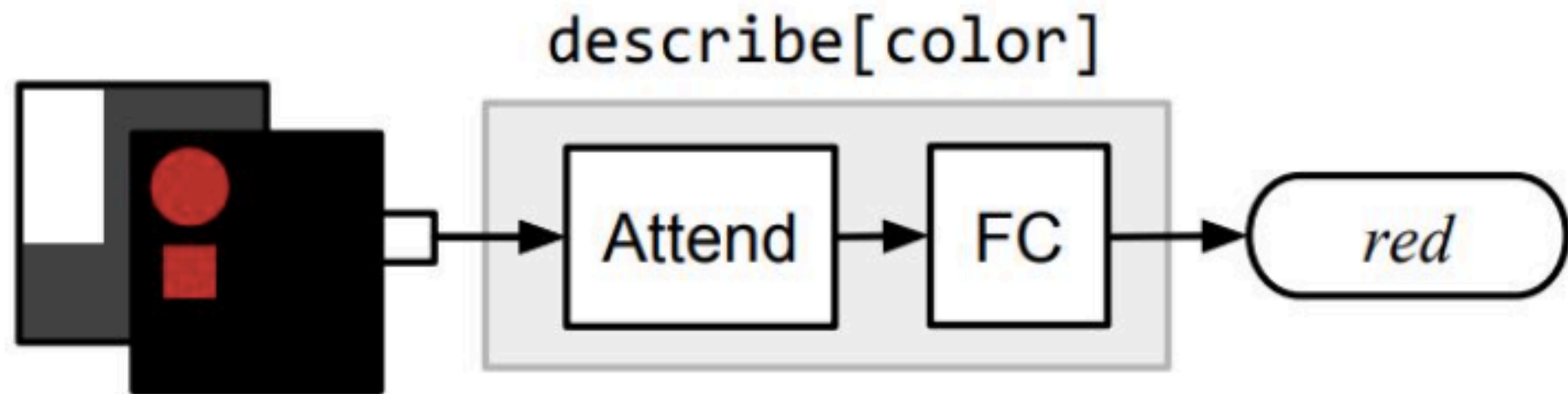


Merge two attentions into one

`combine[and]`: activate regions active in both inputs

Describe Module

Image \times Attention \rightarrow Label

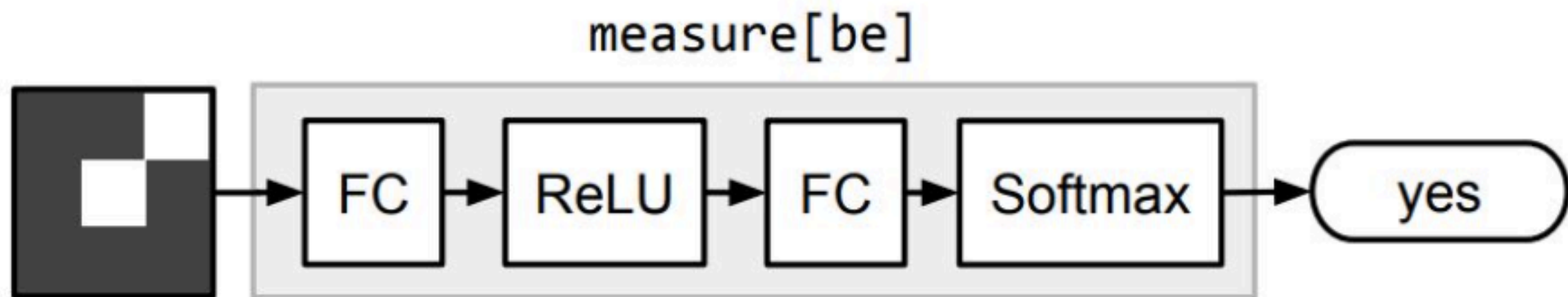


Average image features weighted by attention, then pass averaged feature vector through FC layer

`describe[color]` = representation of colors in the region attended to

Measure Module

Attention \rightarrow **Label**



Map attention to a distribution over labels

Can evaluate existence of detected object or count sets of objects

Question Parsing

Question → dependency representation

Filter dependencies to those connected by *wh*-word or connecting verb

“What is standing in the field?”

what(stand)

“What color is the truck?”

color(truck)

“Is there a circle next to a square?”

is(circle, next-to(square))

Question Tree

leaves:

find

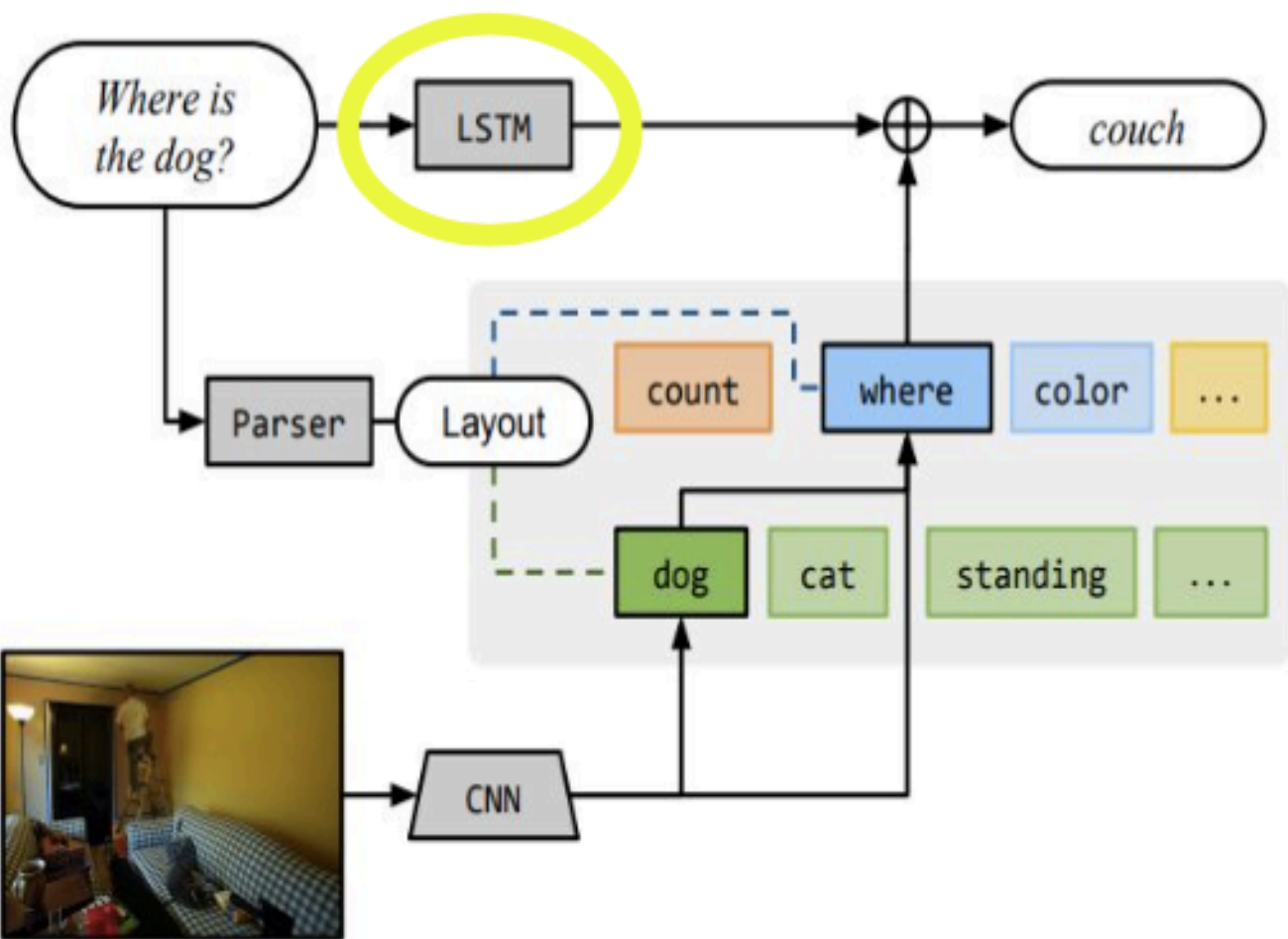
internal nodes:

transform, combine

root nodes:

describe, measure

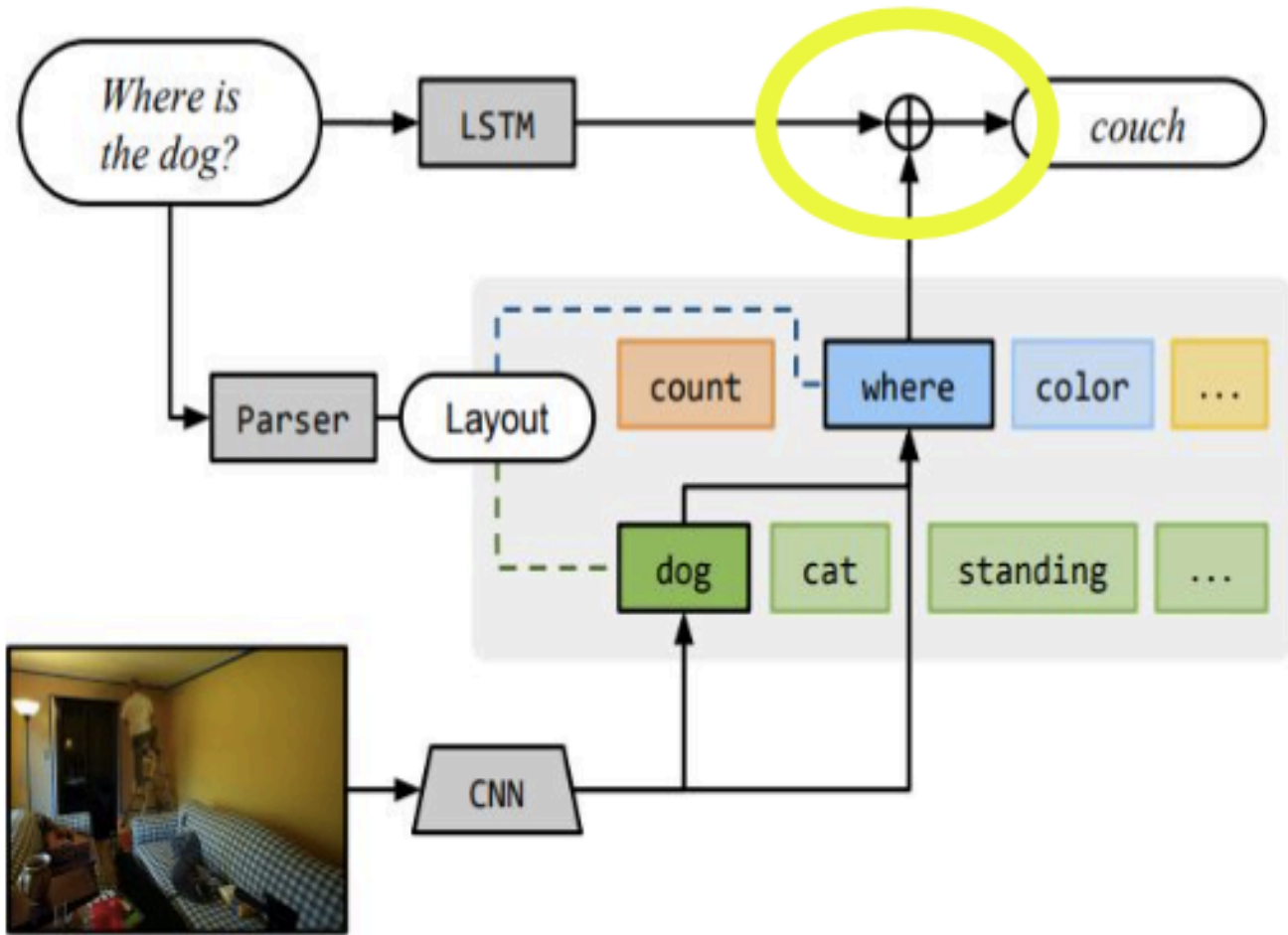
Question Encoding



Full model = NMN +
LSTM question encoder

- LSTM models attributes lost in parsing: "underlying syntactic, semantic regularities in the data"
- Single-layer LSTM w/ 1000 hidden units

Predicting an Answer



1. Pass final hidden state of LSTM through FC layer
2. Add output elementwise to representation produced by root module of NMN
3. Apply ReLU, another FC layer, and get distribution over possible answers

$$p(y | w, x; \theta)$$

DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs

Reasoning	Passage (some parts shortened)	Question	Answer	BiDAF
Subtraction (28.8%)	That year, his Untitled (1981) , a painting of a haloed, black-headed man with a bright red skeletal body, depicted amid the artists signature scrawls, was sold by Robert Lehrman for \$16.3 million, well above its \$12 million high estimate.	How many more dollars was the Untitled (1981) painting sold for than the 12 million dollar estimation?	4300000	\$16.3 million
Comparison (18.2%)	In 1517, the seventeen-year-old King sailed to Castile. There, his Flemish court In May 1518, Charles traveled to Barcelona in Aragon.	Where did Charles travel to first, Castile or Barcelona?	Castile	Aragon
Selection (19.4%)	In 1970, to commemorate the 100th anniversary of the founding of Baldwin City, Baker University professor and playwright Don Mueller and Phyllis E. Braun, Business Manager, produced a musical play entitled The Ballad Of Black Jack to tell the story of the events that led up to the battle.	Who was the University professor that helped produce The Ballad Of Black Jack, Ivan Boyd or Don Mueller?	Don Mueller	Baker
Addition (11.7%)	Before the UNPROFOR fully deployed, the HV clashed with an armed force of the RSK in the village of Nos Kalik, located in a pink zone near Šibenik, and captured the village at 4:45 p.m. on 2 March 1992. The JNA formed a battlegroup to counterattack the next day.	What date did the JNA form a battlegroup to counterattack after the village of Nos Kalik was captured?	3 March 1992	2 March 1992

Count (16.5%) and Sort (11.7%)	Denver would retake the lead with kicker Matt Prater nailing a 43-yard field goal , yet Carolina answered as kicker John Kasay ties the game with a 39-yard field goal Carolina closed out the half with Kasay nailing a 44-yard field goal In the fourth quarter, Carolina sealed the win with Kasay's 42-yard field goal .	Which kicker kicked the most field goals?	John Kasay	Matt Prater
Coreference Resolution (3.7%)	James Douglas was the second son of Sir George Douglas of Pittendreich, and Elizabeth Douglas, daughter David Douglas of Pittendreich. Before 1543 he married Elizabeth , daughter of James Douglas, 3rd Earl of Morton. In 1553 James Douglas succeeded to the title and estates of his father-in-law .	How many years after he married Elizabeth did James Douglas succeed to the title and estates of his father-in-law?	10	1553
Other Arithmetic (3.2%)	Although the movement initially gathered some 60,000 adherents , the subsequent establishment of the Bulgarian Exarchate reduced their number by some 75% .	How many adherents were left after the establishment of the Bulgarian Exarchate?	15000	60,000
Set of spans (6.0%)	According to some sources 363 civilians were killed in Kavadarci , 230 in Negotino and 40 in Vatasha .	What were the 3 villages that people were killed in?	Kavadarci, Negotino, Vatasha	Negotino and 40 in Vatasha
Other (6.8%)	This Annual Financial Report is our principal financial statement of accountability. The AFR gives a comprehensive view of the Department's financial activities ...	What does AFR stand for?	Annual Financial Report	one of the Big Four audit firms

NEURAL MODULE NETWORKS FOR REASONING OVER TEXT

- Use Neural Module Networks (**NMNs**) to answer compositional questions against a paragraph of text.
- Require multiple steps of reasoning : discrete, symbolic operations (as shown in DROP dataset)
- **NMNs are**
 - Interpretable
 - Modular
 - Compositional

Example

Who kicked the longest field goal in the second quarter?

Question Parser

```
relocate(find-max-num(filter(find())))
```

Program Executor

find	filter	find-max-num	relocate
field goal	in the second quarter		Who kicked

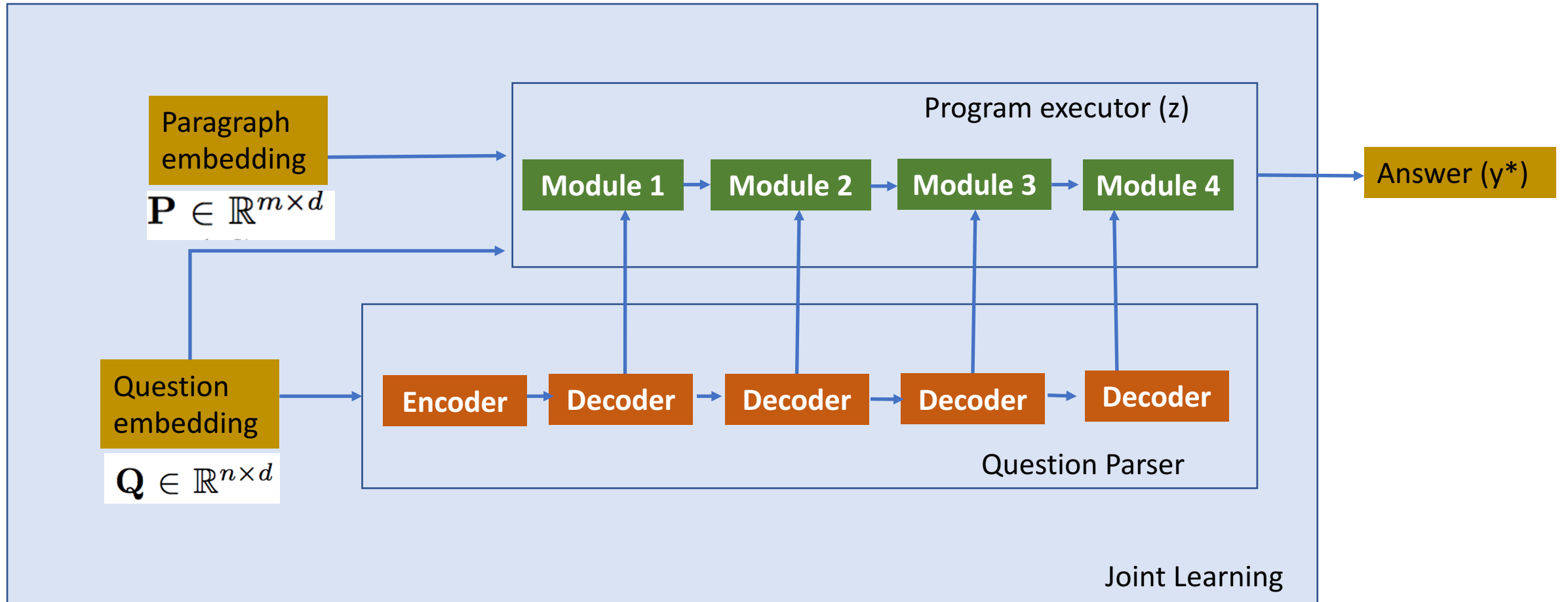
Answer: Connor Barth

In the first quarter, Buffalo trailed as Chiefs QB Tyler Thigpen completed a 36-yard TD pass to RB Jamaal Charles. The Bills responded with RB Marshawn Lynch getting a 1-yard touchdown run. In the second quarter, Buffalo took the lead as kicker Rian Lindell made a 21-yard and a 40-yard field goal. Kansas City answered with Thigpen completing a 2-yard TD pass. Buffalo regained the lead as Lindell got a 39-yard field goal. The Chiefs struck with kicker Connor Barth getting a 45-yard field goal, yet the Bills continued their offensive explosion as Lindell got a 34-yard field goal, along with QB Edwards getting a 15-yard TD run. In the third quarter, Buffalo continued its poundings with Edwards getting a 5-yard TD run, while Lindell got himself a 48-yard field goal. Kansas City tried to rally as Thigpen completed a 45-yard TD pass to WR Mark Bradley, yet the Bills replied with Edwards completing an 8-yard TD pass to WR Josh Reed. In the fourth quarter, Edwards completed a 17-yard TD pass to TE Derek Schouman.

NMN components

- Modules : differentiable modules that perform reasoning over text and symbols in a probabilistic manner
- Contextual token representations : q as $\mathbf{Q} \in \mathbb{R}^{n \times d}$ p as $\mathbf{P} \in \mathbb{R}^{m \times d}$
 - n and m are number of tokens in ques and para, d = size of embedding (bidirectional - GRU or pre trained BERT)
- Question Parser : encoder decoder model with attention to map question into executable program
- Learning: $J = \sum_{\mathbf{z}} p(y^* | \mathbf{z}) p(\mathbf{z} | q)$
 - likelihood of the program under the question-parser model $p(\mathbf{z} | q)$
 - for any given program z , likelihood of the gold-answer $p(y^* | z)$

NMN components



$$J = \sum_{\mathbf{z}} p(y^* | \mathbf{z}) p(\mathbf{z} | q)$$

Learning Challenges

- **Question Parser :**

- Free form real world questions : diverse grammar and lexical variability

- **Program Executor**

- No intermediate feedback available for modules. Errors gets propagated

- **Joint Learning:**

- supervision only at gold level, difficult to learn question parser and program executor jointly

Modules

Module	In	Out	Task
find	Q	P	For question spans in the input, find similar spans in the passage
filter	Q, P	P	Based on the question, select a subset of spans from the input
relocate	Q, P	P	Find the argument asked for in the question for input paragraph spans
find-num	P	N	} Find the number(s) / date(s) associated to the input paragraph spans
find-date	P	D	
count	P	C	Count the number of input passage spans
compare-num-lt	P, P	P	Output the span associated with the smaller number.
time-diff	P, P	TD	Difference between the dates associated with the paragraph spans
find-max-num	P	P	Select the span that is associated with the largest number
span	P	S	Identify a contiguous span from the attended tokens

find(Q) \rightarrow P

For question spans in the input, find similar spans in the passage

- Similarity matrix between question and para tokens embedding

$$S_{ij} = \mathbf{w}_f^T [\mathbf{Q}_{i:}; \mathbf{P}_{j:}; \mathbf{Q}_{i:} \circ \mathbf{P}_{j:}] \quad \bar{\mathbf{w}}_f \in \mathbb{R}^{3d}$$

- Normalize S to get attention matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$
- Compute *expected* paragraph attention

$$P = \sum_i Q_i \cdot A_{i:} \in \mathbb{R}^m$$

Output para
attention map


Input
question
attention map

find(Q) → P : Example

Question: Which player scored the longest rushing TD?

Program: `span(relocate(find-max-num(find)))`

Question attention map is available from the encoder – decoder of parser



question-attention: Which player scored the longest rushing TD?

passage_attention: Coming off their Thursday night home win over the Packers, the Cowboys flew to Ford Field for a Week 14 interconference duel with the Detroit Lions. In the first quarter, Dallas trailed early as Lions RB T.J. Duckett **getting a 32-yard TD run**, along with kicker Jason Hanson getting a 19-yard field goal. In the second quarter, the Cowboys got on the board with RB Marion Barber **getting a 20-yard TD run**. Detroit would answer with Hanson kicking a 36-yard field goal, while RB Kevin Jones **getting a 2-yard TD run**. The Cowboys ended the half with QB Tony Romo completing an 8-yard TD pass to Barber. In the third quarter, the Lions replied with Jones **getting a 3-yard TD run** for the only score of the period. In the fourth quarter, Dallas came back and took the lead with Barber **getting a 1-yard TD run** and Romo completing a 16-yard TD pass to TE Jason Witten. With the win, the Cowboys improved to 12-1 and clinched the NFC East crown for the first time since 1998.

filter(Q, P) \rightarrow P

Based on the question, select a subset of spans from the input

- Weighted sum of question-token embedding

$$\mathbf{q} = \sum_i Q_i \cdot \mathbf{Q}_{i:} \in \mathbb{R}^d$$

- Compute a locally-normalized paragraph-token mask

$$M_j = \sigma(\mathbf{w}_{\text{filter}}^T [\mathbf{q}; \mathbf{P}_{j:}; \mathbf{q} \circ \mathbf{P}_{j:}]) \quad M \in \mathbb{R}^m \quad \mathbf{w}_{\text{filter}}^T \in \mathbb{R}^{3d}$$

- Output is a normalized masked input paragraph attention

$$P_{\text{filtered}} = \text{normalize}(M \circ P).$$

filter(Q, P) → P : Example

Who kicked the longest field goal in the second quarter?

Question Parser

relocate(find-max-num(filter(find())))

Program Executor

find	filter	find-max-num	relocate
field goal	in the second quarter		Who kicked

Answer: Connor Barth

In the first quarter, Buffalo trailed as Chiefs QB Tyler Thigpen completed a 36-yard TD pass to RB Jamaal Charles. The Bills responded with RB Marshawn Lynch getting a 1-yard touchdown run. In the second quarter, Buffalo took the lead as kicker Rian Lindell made a 21-yard and a 40-yard field goal. Kansas City answered with Thigpen completing a 2-yard TD pass. Buffalo regained the lead as Lindell got a 39-yard field goal. The Chiefs struck with kicker Connor Barth getting a 45-yard field goal, yet the Bills continued their offensive explosion as Lindell got a 34-yard field goal, along with QB Edwards getting a 15-yard TD run. In the third quarter, Buffalo continued its poundings with Edwards getting a 5-yard TD run, while Lindell got himself a 48-yard field goal. Kansas City tried to rally as Thigpen completed a 45-yard TD pass to WR Mark Bradley, yet the Bills replied with Edwards completing an 8-yard TD pass to WR Josh Reed. In the fourth quarter, Edwards completed a 17-yard TD pass to TE Derek Schouman.

relocate(Q, P) \rightarrow P

Find the argument asked for in the question for input paragraph spans

- Weighted sum of question-token embedding with attention map

$$\mathbf{q} = \sum_i Q_i \cdot \mathbf{Q}_{i:} \in \mathbb{R}^d$$

- Compute a paragraph-to-paragraph attention matrix

$$\mathbf{R}_{ij} = \mathbf{w}_{\text{relocate}}^T [(\mathbf{q} + \mathbf{P}_{i:}); \mathbf{P}_{j:}; (\mathbf{q} + \mathbf{P}_{i:}) \circ \mathbf{P}_{j:}] \quad \mathbf{w}_{\text{relocate}} \in \mathbb{R}^{3d}$$

- Output attention is a weighted sum of the rows R weighted by the input paragraph attention

$$P_{\text{relocated}} = \sum_i P_i \cdot \mathbf{R}_i$$

find-num(P) \rightarrow N and find-date(P) \rightarrow D

Find the number(s) / date(s) associated to the input paragraph spans

- Extract numbers and dates as a pre-processing step, eg [2, 2, 3, 4]
- Compute a token-to-number similarity matrix

$$\mathbf{S}^{\text{num}} \in \mathbb{R}^{m \times N_{\text{tokens}}} \text{ as, } \mathbf{S}^{\text{num}}_{i,j} = \mathbf{P}_{i:}^T \mathbf{W}_{\text{num}} \mathbf{P}_{n_j:} \quad \mathbf{W}_{\text{num}} \in \mathbb{R}^{d \times d}$$

$$\mathbf{A}^{\text{num}}_{i:} = \text{softmax}(\mathbf{S}^{\text{num}}_{i:})$$

- Compute an expected distribution over the number tokens

$$\bar{T} = \sum_i P_i \cdot \mathbf{A}^{\text{num}}_{i:} \quad \bar{T} = [0.1, 0.4, 0.3, 0.2]$$

- Aggregate the probabilities for number-tokens ,
- Example : {2, 3, 4} with N = [0.5, 0.3, 0.2]

find-num(P) → N : xample

Question: Which group was smaller, the 25 to 44 year olds or the 45 to 64 year olds?

Program: span(compare-num-lt(find, find))

number-distribution: 7.1 10.3 12 15.3 18 22 24 25 44 45 55.2 64 65 100 160.7 173.2

question-attention: Which group was smaller, the 25 to 44 year olds or the 45 to 64 year olds?

passage-attention: In the city, the population was spread out with 12.0% under the age of 18, 55.2% from 18 to 24, 15.3% from 25 to 44, 10.3% from 45 to 64, and 7.1% who were 65 years of age or older. The median age was 22 years. For every 100 females, there were 160.7 males. For every 100 females age 18 and over, there were 173.2 males.

Question: Which event happened first, when king Chulalongkorn enacted two decrees banning the capture and sale of Kha slaves or when Siam abolished the tributes collected from vassal states?

Program: span(compare-date-lt(find, find))

Answer: banning the capture and sale of Kha slaves

date-distribution: 1874/-1/-1 1884/-1/-1 1868/-1/-1 1883/-1/-1 1899/-1/-1

question-attention: Which event happened first, when king Chulalongkorn enacted two decrees banning the capture and sale of Kha slaves or when Siam abolished the tributes collected from vassal states?

passage-attention: Before the Monthon reforms initiated by king Chulalongkorn, Siamese territories were divided into three categories: Inner Provinces forming the core of the kingdom, Outer Provinces that were adjacent to the inner provinces and tributary states located on the border regions. The area of southern Laos that came under Siamese control following the Lao rebellion and destruction of Vientiane belonged to the later category, maintaining relative autonomy. Lao nobles who had received the approval of the Siamese king exercised authority on the Lao population as well as the Alak and Laven-speaking tribesmen. Larger tribal groups often raided weaker tribes abducting people and selling them into slavery at the trading hub of Champasak, while themselves falling prey to Khmer, Lao and Siamese slavers. From Champasak the slaves were transported to Phnom-Penh and Bangkok, thus creating a large profits for the slavers and various middlemen. In 1874 and 1884, king Chulalongkorn enacted two decrees banning the capture and sale of Kha slaves while also freeing all slaves born after 1868. Those abolitionist policies had an immediate effect on slave trading communities. In 1883, France attempted to expand its control in Southeast Asia by claiming that the Treaty of Hué extended into all Vietnamese vassal states. French troops gradually occupied the Kontum Plateau and pushed the Siamese from Laos following the Franco-Siamese War. A new buffer zone was thus created on the west bank of Mekong, as the area lacked the presence of the Siamese military local outlaws flocked the newly created safe haven. In 1899, Siam abolished the tributes collected from vassal states, replacing them with a new tax collected from all able bodied men, undermining the authority of Lao officials.

count(P) \rightarrow C

Count the number of input passage spans

- Count([0, 0, 0.3, 0.3, 0, 0.4]) = 2
- Module first scales the attention using the values [1, 2, 5, 10] to convert it into a matrix $P_{\text{scaled}} \in \mathbb{R}^{m \times 4}$

$$c_v = \sum \sigma(\bar{F} F(\text{countGRU}(\bar{P}_{\text{scaled}}))) \in \mathbb{R}.$$
$$p(c) \propto \exp(-(c - c_v)^2 / 2v^2) \quad \forall c \in [\hat{0}, 9]$$

Normalized-passage-attention where passage lengths are typically 400-500 tokens. Hence scaling the attention using values >1 helps the model in differentiating amongst small values.

Pretraining this module by generating synthetic data of attention and count values helps

Question: How many rushing touchdowns were scored in the game?

Program: `count(find)`

Answer: 5

count-distribution: 0 1 2 3 4 5 6 7 8 9

question-attention: How many rushing touchdowns were scored in the game?

passage-attention: Coming off their Thursday night home win over the Packers, the Cowboys flew to Ford Field for a Week 14 interconference duel with the Detroit Lions. In the first quarter, Dallas trailed early as Lions RB T.J. Duckett getting a 32-yard TD run, along with kicker Jason Hanson getting a 19-yard field goal. In the second quarter, the Cowboys got on the board with RB Marion Barber getting a 20-yard TD run. Detroit would answer with Hanson kicking a 36-yard field goal, while RB Kevin Jones getting a 2-yard TD run. The Cowboys ended the half with QB Tony Romo completing an 8-yard TD pass to Barber. In the third quarter, the Lions replied with Jones getting a 3-yard TD run for the only score of the period. In the fourth quarter, Dallas came back and took the lead with Barber getting a 1-yard TD run and Romo completing a 16-yard TD pass to TE Jason Witten. With the win, the Cowboys improved to 12-1 and clinched the NFC East crown for the first time since 1998.

compare-num-lt(P1, P2) → P

Output the span associated with the smaller number

- $N1 = \text{find_num}(P1)$, $N2 = \text{find_num}(P2)$
- Computes two soft boolean values, $p(N1 < N2)$ and $p(N2 < N1)$

$$p(N_1 < N_2) = \sum_i \sum_j \mathbb{1}_{N_1^i < N_2^j} N_1^i N_2^j \quad p(N_2 < N_1) = \sum_i \sum_j \mathbb{1}_{N_2^i < N_1^j} N_2^i N_1^j$$

- Outputs a weighted sum of the input paragraph attentions

$$P_{out} = p(N_1 < N_2) * P_1 + p(N_2 < N_1) * P_2.$$

Question: Which group was smaller, the 25 to 44 year olds or the 45 to 64 year olds?

Program: span(compare-num-lt(find, find))

Answer: 45 to 64

compare-num-lt passage-attention: In the city, the population was spread out with 12.0% under the age of 18, 55.2% from 18 to 24, 15.3% from 25 to 44, 10.3% from 45 to 64, and 7.1% who were 65 years of age or older. The median age was 22 years. For every 100 females, there were 160.7 males. For every 100 females age 18 and over, there were 173.2 males.

number-distribution: 7.1 10.3 12 15.3 18 22 24 25 44 45 55.2 64 65 100 160.7 173.2

question-attention: Which group was smaller, the 25 to 44 year olds or the 45 to 64 year olds?

passage-attention: In the city, the population was spread out with 12.0% under the age of 18, 55.2% from 18 to 24, 15.3% from 25 to 44, 10.3% from 45 to 64, and 7.1% who were 65 years of age or older. The median age was 22 years. For every 100 females, there were 160.7 males. For every 100 females age 18 and over, there were 173.2 males.

time-diff(P1, P2) → TD

Difference between the dates associated with the paragraph spans

- Module internally calls the find-date module to get a date distribution for the two paragraph attentions, D1 and D2

$$p(t_d) = \sum_{i,j} \mathbb{1}_{(d_i - d_j = t_d)} D_1^i D_2^j.$$

find-max-num(P) \rightarrow P, find-min-num(P) \rightarrow P

Select the span that is associated with the largest number

- Compute an expected number token distribution T using find-num
- Compute the expected probability that each number token is the one with the maximum value, $T^{\max} \in \mathbb{R}^{\text{ntokens}}$
- Reweight the contribution from the i-th paragraph token to the j-th number token

$$\bar{P}_i = \sum_j T_j^{\max} / T_j \cdot P_i \cdot \mathbf{A}^{\text{num}}_{ij}.$$

Question: Which player scored the longest rushing TD?

Program: `span(relocate(find-max-num(find)))`

max-num-distribution: 1 2 3 8 12 14 16 19 20 32 36 1998

input-num-distribution: 1 2 3 8 12 14 16 19 20 32 36 1998

question-attention: Which player scored the longest rushing TD?

passage_attention: Coming off their Thursday night home win over the Packers, the Cowboys flew to Ford Field for a Week 14 interconference duel with the Detroit Lions. In the first quarter, Dallas trailed early as Lions RB T.J. Duckett [getting a 32-yard TD run](#), along with kicker Jason Hanson getting a 19-yard field goal. In the second quarter, the Cowboys got on the board with RB Marion Barber [getting a 20-yard TD run](#). Detroit would answer with Hanson kicking a 36-yard field goal, while RB Kevin Jones [getting a 2-yard TD run](#). The Cowboys ended the half with QB Tony Romo completing an 8-yard TD pass to Barber. In the third quarter, the Lions replied with Jones [getting a 3-yard TD run](#) for the only score of the period. In the fourth quarter, Dallas came back and took the lead with Barber [getting a 1-yard TD run](#) and Romo completing a 16-yard TD pass to TE Jason Witten. With the win, the Cowboys improved to 12-1 and clinched the NFC East crown for the first time since 1998.

$\text{span}(P) \rightarrow S$

Identify a contiguous span from the attended tokens

- Only appears as the outermost module in a program.
- Outputs two probability distributions, P_s and $P_e \in \mathbb{R}^m$, denoting start and end of a span
- This module is implemented similar to the count module

Auxiliary supervision

- unsupervised auxiliary loss to provide an inductive bias to the execution of find-num, find-date, and relocate modules
- provide heuristically-obtained supervision for question program and intermediate module output for a subset of questions (5–10%).

Unsupervised auxiliary loss for IE

- find-num, find-date, and relocate modules perform information extraction
- Objective increases the sum of the attention probabilities for output tokens that appear within a window $W = 10$

$$H_{\text{loss}}^{\text{n}} = - \sum_{i=1}^m \log \left(\sum_{j=0}^{N_{\text{tokens}}} \mathbb{1}_{n_j \in [i \pm W]} \mathbf{A}^{\text{num}}_{ij} \right)$$

$$H_{\text{loss}} = H_{\text{loss}}^{\text{n}} + H_{\text{loss}}^{\text{d}} + H_{\text{loss}}^{\text{r}}.$$

Question Parse Supervision

- Heuristic patterns to get program and corresponding question attention supervision for a subset of the training data (10%)
 1. *what happened first SPAN1 or SPAN2?*
`span(compare-date-lt(find(), find()))`: with `find` attentions on SPAN1 and SPAN2, respectively. Use `compare-date-gt`, if *second* instead of *first*.
 2. *were there fewer SPAN1 or SPAN2?*
`span(compare-num-lt(find(), find()))`: with `find` attentions on SPAN1 and SPAN2, respectively. Use `compare-num-gt`, if *more* instead of *fewer*.
 3. *how many yards was the longest {touchdown / field goal}?*
`find-num(find-max-num(find()))`: with `find` attention on *touchdown / field goal*. For *shortest*, the `find-min-num` module is used.
 4. *how many yards was the longest {touchdown / field goal} SPAN ?*
`find-num(find-max-num(filter(find())))`: with `find` attention on *touchdown / field goal* and `filter` attention on all SPAN tokens.

Intermediate Module Output Supervision

- Used for find-num and find-date modules
- For a subset of the questions (5%)
- Eg : “how many yards was the longest/shortest touchdown?”
 - Identify all instances of the token “touchdown”
 - Assume the closest number to it should be an output of the find-num module.
 - Supervise this as a multi-hot vector N^* and use an auxiliary loss

Dataset

20, 000 questions for training/validation, and 1800 questions for testing (25% of DROP)

Automatically extracted questions in the scope of model based on their first n-gram.

Based on the manual analysis we classify these questions into different categories, which are:

Date-Compare e.g. *What happened last, commission being granted to Robert or death of his cousin?*

Date-Difference e.g. *How many years after his attempted assassination was James II coronated?*

Number-Compare e.g. *Were there more of cultivators or main agricultural labourers in Sweden?*

Extract-Number e.g. *How many yards was Kasay's shortest field goal during the second half?*

Count e.g. *How many touchdowns did the Vikings score in the first half?*

Extract-Argument e.g. *Who threw the longest touchdown pass in the first quarter?*

RESULTS

Model	F1	EM
NAQANET	62.1	57.9
TAG-NABERT+	74.2	70.6
NABERT+	75.4	72.0
MTMSN	76.5	73.1
OUR MODEL (w/ GRU)	73.1	69.6
OUR MODEL (w/ BERT)	77.4	74.0

(a) Performance on DROP (pruned)

RESULTS – Questions Type

Question Type	MTMSN	Our Model (w/ BERT)
DATE-COMPARE (18.6%)	85.2	82.6
DATE-DIFFERENCE (17.9%)	72.5	75.4
NUMBER-COMPARE (19.3%)	85.1	92.7
EXTRACT-NUMBER (13.5%)	80.7	86.1
COUNT (17.6%)	61.6	55.7
EXTRACT-ARGUMENT (12.8%)	66.6	69.7

(b) Performance by Question Type (F1)

Effect of Auxiliary Supervision

Supervision Type		w/ BERT	w/ GRU
H_{loss}	MOD-SUP		
✓	✓	77.4	73.1
✓		76.3	71.8
	✓	—*	57.3

(a) **Effect of Auxiliary Supervision:** The auxiliary loss contributes significantly to the performance, whereas module output supervision has little effect. **Training diverges without H_{loss} for the BERT-based model.*

Incorrect Program Predictions.

- *How many touchdown passes did Tom Brady throw in the season?* - `count(find)`
 - Correct answer requires a simple lookup from the paragraph.
- *Which happened last, failed assassination attempt on Lenin, or the Red Terror?* `date-compare-gt(find, find)`
 - Correct answer requires natural language inference about the order of events and not symbolic comparison between dates.
- *Who caught the most touchdown passes?* - `relocate(find-max-num(find))`.
 - Require nested counting which is out of scope

Future Work

- Design additional modules
 - How many languages each had less than 115, 000 speakers in the population?
 - Which quarterback threw the most touchdown passes?
 - How many points did the packers fall behind during the game?
- Use complete dataset of DROP : In current system, training model on the questions for which modules can't express the correct reasoning harms their ability to execute their intended operations
- Opens up avenues for transfer learning where modules can be independently trained using indirect or distant supervision from different tasks
- Combining black-box operations with the interpretable modules so that can capture more expressivity

Review Comments - Pros

- Interesting idea [Atishya, Rajas, Keshav, Siddhant, Lovish]
- Interpretable and modular [Atishya, Rajas, Siddhant, Lovish, Vipul]
- Better than BERT for symbolic reasoning [Keshav]
- Auxiliary loss formulation seems a very novel idea[Vipul]
- Question parser has new role: parse to return composition of modules.[Pawan]

Review comments - Cons

- Difficult to understand module description [Atishya, Siddhant]
- Auxillary loss not generalizable [Atishya, Rajas]
- Contribution of each module not studied [Atishya, Rajas, Siddhant, Lovish, Pawan]
- Only 22% of DROP dataset used [Rajas, Keshav, Lovish]
- Compositional reasoning queries like “Who is the mother of PM of India?” are not handled. [Keshav]
- Endless amount of modules required to achieve full reasoning capability[Vipul]

Review comments - Extensions

- Study on the contribution of each module [Atishya]
- Pre-train all the modules by collecting data using specific heuristics [Atishya, Rajas]
- RL framework to predict whether a given question can be sufficiently reasoned [Rajas]
- Module to predict open-predicates of the type $PM(\text{India}, x)$ & $Mother(x, y)$. [Keshav, Vipul]
- Train multi purpose modules (to predict *citizen of* and *president of* relationships) [Vipul]
- Combine end-to-end neural system and NMN [Keshav]
- Learn new modules from dataset automatically ; learn new SPARQL template from data) [Siddhant, Pawan]
- Curriculum learning [Siddhant]
- Metalearning to automatically determine the modules [Lovish]