TOOPLOOX AI

EMNLP conference

Highlights

Review by **Dominika Basaj**

EMNLP in numbers

- 2100 submitted papers
- 549 accepted papers (24% acceptance rate)
- 72 demo submissions (40% acceptance rate)
- 2500 attendees

Trends

- Critical analysis of ML models
- More difficult tasks and thus datasets
- Language modelling
- Unsupervised machine translation

Critical analysis of ML models

Pathologies of Neural Models make Interpretations Difficult

https://arxiv.org/abs/1804.07781

$$\sum_{(\mathbf{x},y)\in(\mathcal{X},\mathcal{Y})} \log(f(y\mid\mathbf{x})) + \lambda \sum_{\tilde{\mathbf{x}}\in\tilde{\mathcal{X}}} \mathbb{H}(f(y\mid\tilde{\mathbf{x}})), \tag{2}$$

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 ?
Answer	Colorado Springs experiments
Before	did
After	spend Astor money on ?
Confidence	0.78 → 0.91 → 0.52

SNLI	
Premise	Well dressed man and woman dancing in the street
Original	Two man is dancing on the street
Answer	Contradiction
Before	dancing
After	two man dancing
Confidence	0.977 → 0.706 → 0.717

VQA	
Original	What color is the flower ?
Answer	yellow
Before	flower ?
After	What color is flower ?
Confidence	0.847 → 0.918 → 0.745

Critical analysis of ML models

How much reading does reading comprehension require? A critical investigation of popular benchmarks.

https://arxiv.org/pdf/1808.04926.pdf

Best Short Paper

Task	Full	Q-only	P-only	$\Delta(min)$
Key-Value Memory Networks				
CBT-NE	**35.0%**	29.1%	24.1%	−5.9
CBT-CN	**37.6%**	32.4%	24.4%	−5.2
CBT-V	52.5%	**55.7%**	36.0%	+3.2
CBT-P	55.2%	**56.9%**	30.1%	+1.7
Gated Attention Reader				
CBT-NE	**74.9%**	50.6%	40.8%	−17.5
CBT-CN	**70.7%**	54.0%	36.7%	−16.7
CNN	**77.8%**	25.6%	38.3%	−39.5
WdW	**67.0%**	41.8%	52.2%	−14.8
WdW-R	**69.1%**	50.0%	50.6%	−15.6

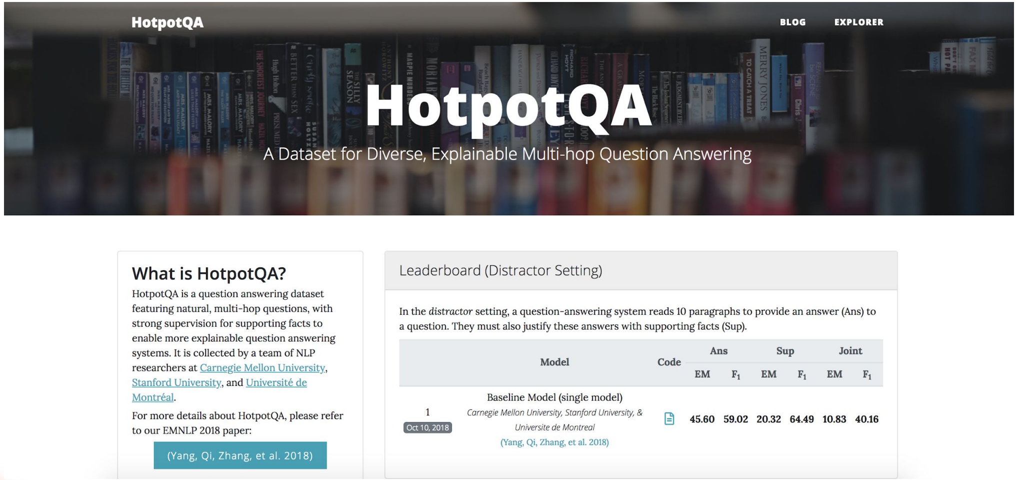
Metric	Full	Q-only	P-only	$\Delta(min)$
EM	**70.7%**	0.6%	10.9%	−59.8
F1	**79.1%**	4.0%	14.8%	−64.3

Table 4: Performance of QANet on SQuAD

More difficult tasks

HotPotQA: A dataset for diverse, explainable multi-hop question answering

https://arxiv.org/pdf/1809.09600.pdf




Paragraph A, Return to Olympus:
[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

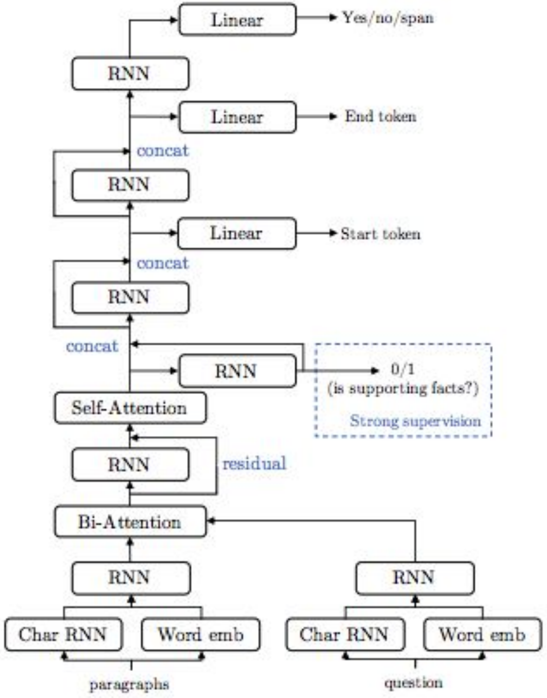
Paragraph B, Mother Love Bone:
[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success.* [8] The album was finally released a few months later.

Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?
A: Malfunkshun
Supporting facts: 1, 2, 4, 6, 7

More difficult tasks

Reasoning Type	%	Example(s)
Inferring the *bridge entity* to complete the 2nd-hop question (Type I)	42	**Paragraph A:** The 2015 Diamond Head Classic was a college basketball tournament ... *Buddy Hield was named the tournament's MVP.* **Paragraph B:** *Chavano Rainier "Buddy" Hield* is a Bahamian professional basketball player for the **Sacramento Kings** of the NBA... **Q:** Which team does the player named 2015 Diamond Head Classic's MVP play for?
Comparing two entities (Comparison)	27	**Paragraph A:** LostAlone were a British rock band ... consisted of *Steven Battelle, Alan Williamson, and Mark Gibson...* **Paragraph B:** Guster is an American alternative rock band ... Founding members *Adam Gardner, Ryan Miller, and Brian Rosenworcel* began... **Q:** Did LostAlone and Guster have the same number of members? (**yes**)
Locating the **answer entity** by checking multiple properties (Type II)	15	**Paragraph A:** Several *current and former members of the Pittsburgh Pirates* ... John Milner, **Dave Parker**, and Rod Scurry... **Paragraph B:** **David Gene Parker**, *nicknamed "The Cobra"*, is an American former player in Major League Baseball... **Q:** Which former member of the Pittsburgh Pirates was nicknamed "The Cobra"?
Inferring about the property of an entity in question through a *bridge entity* (Type III)	6	**Paragraph A:** *Marine Tactical Air Command Squadron 28* is a United States Marine Corps aviation command and control unit based at *Marine Corps Air Station Cherry Point*... **Paragraph B:** *Marine Corps Air Station Cherry Point* ... is a United States Marine Corps airfield located in **Havelock, North Carolina**, USA ... **Q:** What city is the Marine Air Control Group 28 located in?
Other types of reasoning that require more than two supporting facts (Other)	2	**Paragraph A:** ... the towns of Yodobashi, **Okubo, Totsuka, and Ochiai town** *were merged into Yodobashi ward. ... Yodobashi Camera is a store with its name taken from the town and ward.* **Paragraph B:** *Yodobashi Camera* Co., Ltd. is *a major Japanese retail chain specializing in electronics, PCs, cameras and photographic equipment.* **Q:** Aside from Yodobashi, what other towns were merged into the ward which gave the major Japanese retail chain specializing in electronics, PCs, cameras, and photographic equipment it's name?

Baseline model



More difficult tasks

In the *distractor* setting, a question-answering system reads 10 paragraphs to provide an answer (Ans) to a question. They must also justify these answers with supporting facts (Sup).

	Model	Code	Ans		Sup		Joint	
			EM	F_1	EM	F_1	EM	F_1
1 Oct 10, 2018	Baseline Model (single model) *Carnegie Mellon University, Stanford University, & Universite de Montreal* (Yang, Qi, Zhang, et al. 2018)		**45.60**	**59.02**	**20.32**	**64.49**	**10.83**	**40.16**

Leaderboard (Fullwiki Setting)

In the *fullwiki* setting, a question-answering system must find the answer to a question in the scope of the entire Wikipedia. Similar to in the distractor setting, systems are evaluated on the accuracy of their answers (Ans) and the quality of the supporting facts they use to justify them (Sup).

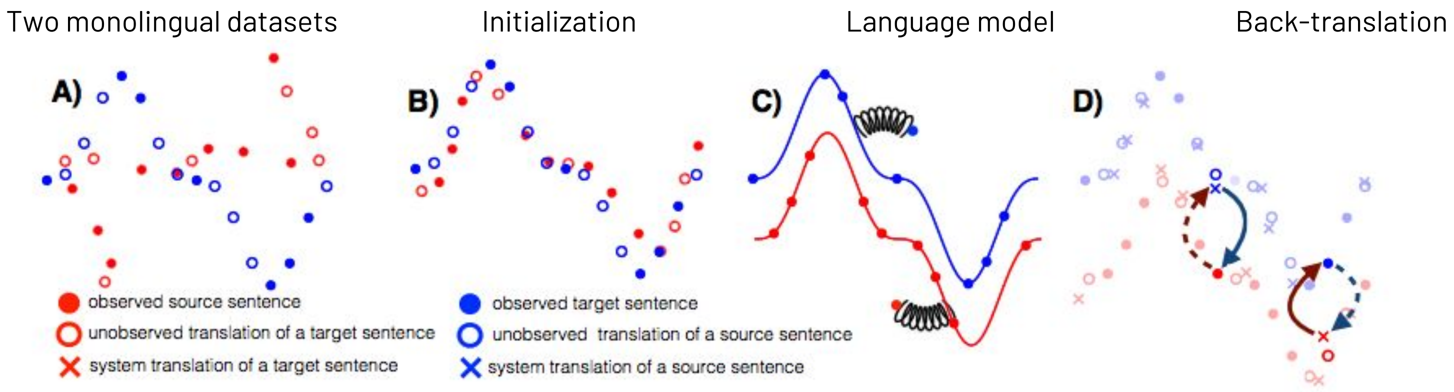
	Model	Code	Ans		Sup		Joint	
			EM	F_1	EM	F_1	EM	F_1
1 Oct 12, 2018	Baseline Model (single model) *Carnegie Mellon University, Stanford University, & Universite de Montreal* (Yang, Qi, Zhang, et al. 2018)		**23.95**	**32.89**	**3.86**	**37.71**	**1.85**	**16.15**

Unsupervised MT

Phrased-Based & Neural Unsupervised Machine Translation

https://arxiv.org/abs/1804.07755

- This work investigates how to learn to translate when having access to only large monolingual corpora in each language.
- We apply these methods to distant and low-resource languages, like English-Russian, English-Romanian and English-Urdu, and report competitive performance against both semi-supervised and supervised baselines.



Unsupervised MT

Phrased-Based & Neural Unsupervised Machine Translation

https://arxiv.org/abs/1804.07755

Algorithm 1: Unsupervised MT

Language models: Learn language models P_s and P_t over source and target languages;
Initial translation models: Leveraging P_s and P_t, learn two initial translation models, one in each direction: $P^{(0)}_{s\to t}$ and $P^{(0)}_{t\to s}$;
for k=1 **to** N **do**
  Back-translation: Generate source and target sentences using the current translation models, $P^{(k-1)}_{t\to s}$ and $P^{(k-1)}_{s\to t}$, factoring in language models, P_s and P_t;
  Train new translation models $P^{(k)}_{s\to t}$ and $P^{(k)}_{t\to s}$ using the generated sentences and leveraging P_s and P_t;
end

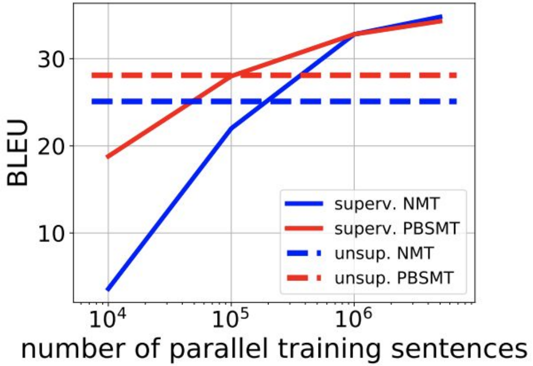


Figure 2: Comparison between supervised and unsupervised approaches on WMT'14 En-Fr, as we vary the number of parallel sentences for the supervised methods.

VQA

4D: Visual QA

16:30 – 18:00

Hall 100 / Hall 400

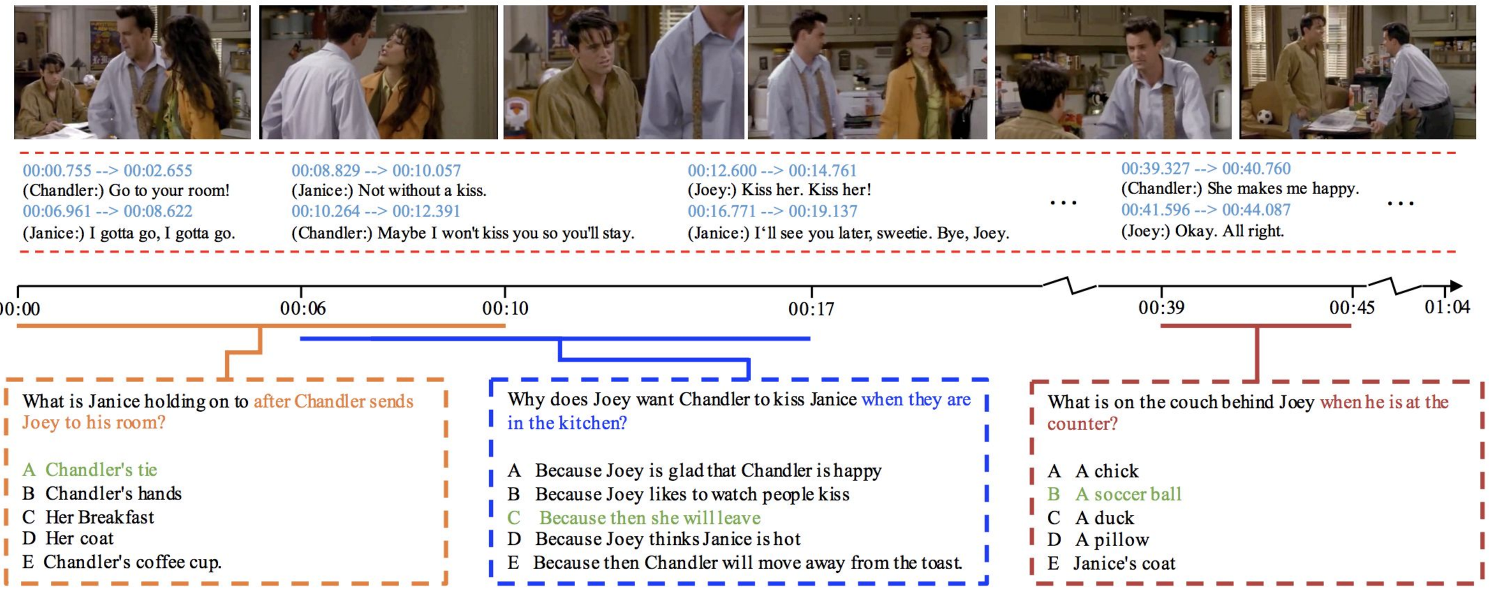
Choose All Remove All

Chair: Carina Silberer

16:30–16:48	**Tell-and-Answer: Towards Explainable Visual Question Answering using Attributes and Captions.** *Qing Li, Jianlong Fu, Dongfei Yu, Tao Mei and Jiebo Luo*
16:48–17:06	**Learning a Policy for Opportunistic Active Learning.** *Aishwarya Padmakumar, Peter Stone and Raymond Mooney*
17:06–17:24	**RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes.** *Semih Yagcioglu, Aykut Erdem, Erkut Erdem and Nazli Ikizler-Cinbis*
17:24–17:42	**TVQA: Localized, Compositional Video Question Answering.** *Jie Lei, Licheng Yu, Mohit Bansal and Tamara Berg*
17:42–18:00	**Localizing Moments in Video with Temporal Language.** *Lisa Anne Hendricks, Oliver Wang, Eli Shechtman, Josef Sivic, Trevor Darrell and Bryan Russell*

Visual QA

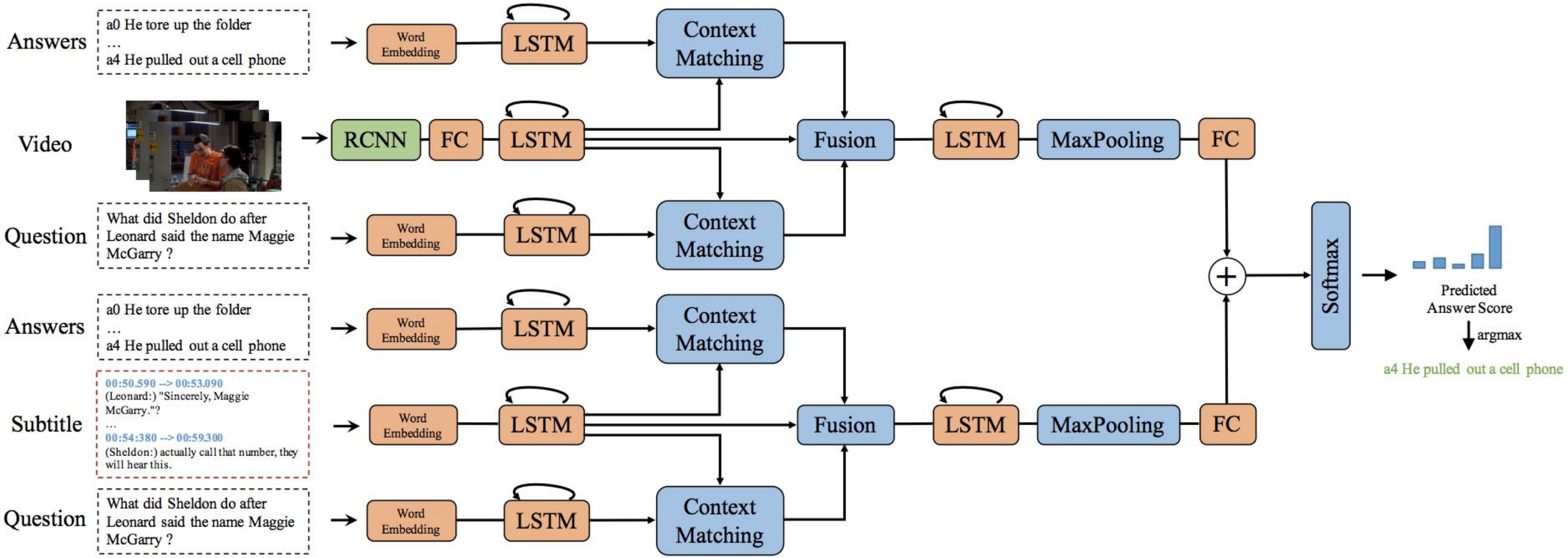
TVQA: Localized, Compositional Video Question Answering http://aclweb.org/anthology/D18-1167



- All questions and answers are attached to 60-90 seconds long clips.
- Some questions can be answered using subtitles or videos alone, while some require information from both modalities

Visual QA

TVQA: Localized, Compositional Video Question Answering http://aclweb.org/anthology/D18-1167

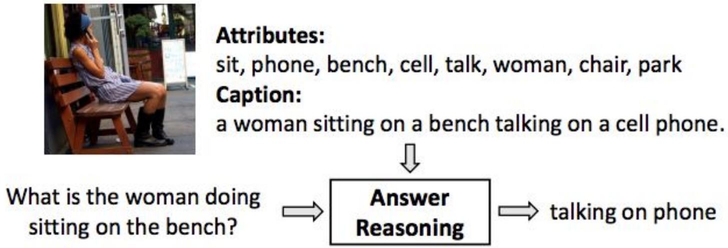


Visual QA

Tell-and-Answer: Towards Explainable Visual Question Answering using Attributes and Captions

http://aclweb.org/anthology/D18-1164

In this work, we propose to break up the end-to-end VQA into two steps: **explaining and reasoning**, in an attempt towards a more explainable VQA by shedding light on the intermediate results between these two steps.



We first extract attributes in the image such as "sit", "phone" and "woman." A caption is also generated to encode the relationship between these attributes, e.g. "woman sitting on a bench."

Visual QA

Tell-and-Answer: Towards Explainable Visual Question Answering using Attributes and Captions

http://aclweb.org/anthology/D18-1164

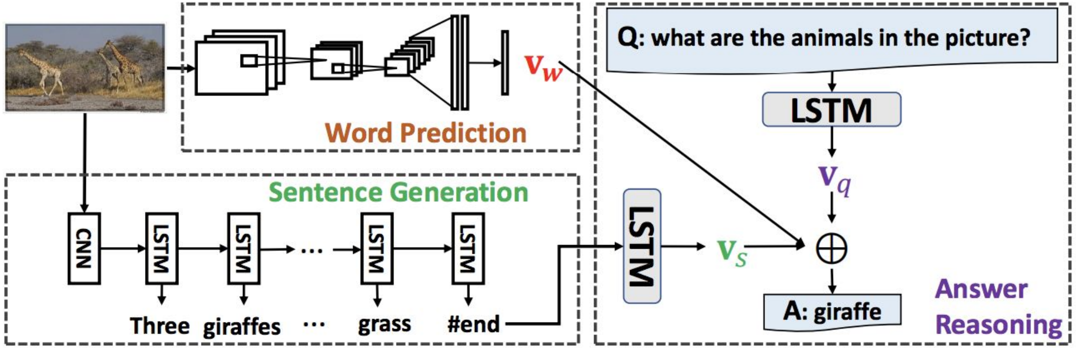


Figure 3: An overview of the proposed framework for VQA with three modules: word prediction (upper left), sentence generation (lower left), answer reasoning (right). **Explaining**: in word prediction, the image is fed into pre-trained visual detectors to extract word-level explanation, which is represented by probability vector $\mathbf{v}_w$; in sentence generation, we input the image to pre-trained captioning model to generate a sentence-level explanation. **Reasoning**: the caption and question are encoded by two different LSTMs into $\mathbf{v}_s$ and $\mathbf{v}_q$, respectively. Then $\mathbf{v}_q$, $\mathbf{v}_w$ and $\mathbf{v}_s$ are concatenated and fed to a fully connected layer with softmax to predict an answer.

Visual QA

RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes

http://aclweb.org/anthology/D18-1166

Text Cloze Style Question	Context Modalities: Images and Descriptions of Steps

Recipe: Last-Minute Lasagna

1. Heat oven to 375 degrees F. Spoon a thin layer of sauce over the bottom of a 9-by-13-inch baking dish.
2. Cover with a single layer of ravioli.
3. Top with half the spinach half the mozzarella and a third of the remaining sauce.
4. Repeat with another layer of ravioli and the remaining spinach mozzarella and half the remaining sauce.
5. Top with another layer of ravioli and the remaining sauce not all the ravioli may be needed. Sprinkle with the Parmesan.
6. Cover with foil and bake for 30 minutes. Uncover and bake until bubbly, 5 to 10 minutes.
7. Let cool 5 minutes before spooning onto individual plates.

Question	Choose the best text for the missing blank to correctly complete the recipe Cover. ______________. Bake. Cool, serve.
Answer	**A. Top, sprinkle** B. Finishing touches C. Layer it up D. Ravioli bonus round

Visual QA

RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes

http://aclweb.org/anthology/D18-1166

	train	**valid**	**test**
# of recipes	15847	1963	1969
...avg. # of steps	5.99	6.01	6.00
...avg. # of tokens (titles)	17.79	17.40	17.67
...avg. # of tokens (descr.)	443.01	440.51	435.33
...avg. # of images	12.67	12.74	12.65
# of question-answers	29657	3562	3567
...textual cloze	7837	961	963
...visual cloze	7144	842	848
...visual coherence	7118	830	851
...visual ordering	7558	929	905

Context Modalities: Titles and Descriptions of Steps

Recipe: Bacon Sushi

Step 1: **What You'll Need** This recipe makes enough bacon sushi to feed 2 - 4 people. 2 x 500g(1 lb.) packages of bacon (I chose an applewood smoked bacon, but any type would work). 3 tbsp. oil. 1 medium onion, finely diced. 1 1...

Step 2: **Cooking the Bacon** The bacon "nori" will have to be partially cooked before it can be rolled with the risotto filling. Preheat the oven to 350 degrees F. Lay half a package of bacon on the rack of the roasting pan, then bak...

Step 3: **Making the Risotto Filling** I once made risotto with sushi rice, since I had no Arborio rice on hand, and I decided that the starchiness was similar in the two. My experiment was a success, and the resulting dish was just as deli...

Step 4: **Jazzing Up the Risotto** Risotto is a wonderfully customizable dish, and a quick search on the internet will result in a multitude of variations. Here are two of my favorites: Asian mushroom risotto. 1 tbsp. oil. 1 package...

Step 5: **Rolling the Sushi** Cover the sushi rolling mat with a large piece of aluminum foil as protection from the risotto and bacon grease. (You don't want your next sushi dinner tasting like bacon. Or maybe you do...) Lay the stri...

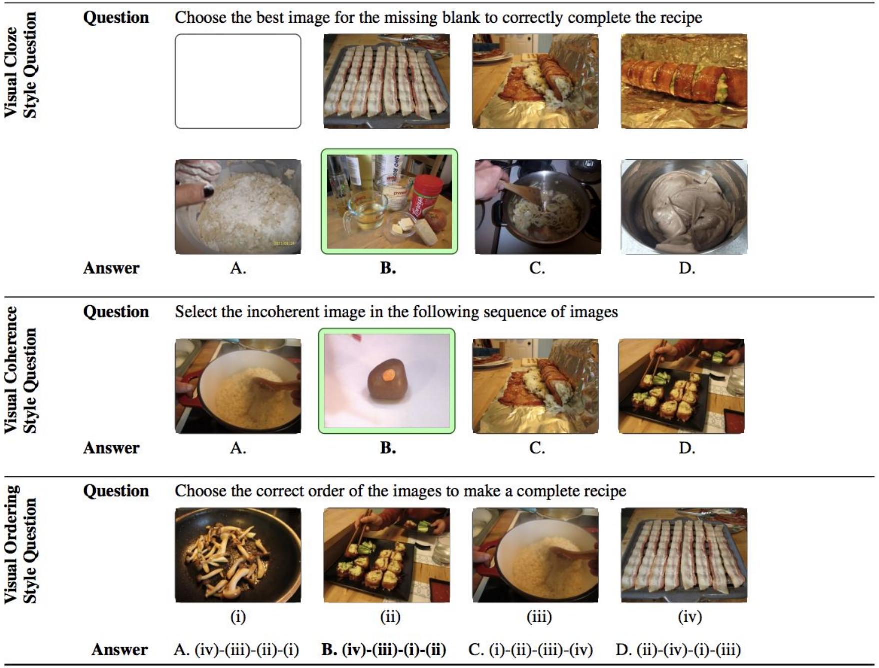
Step 6: **Baking and Slicing** Preheat the oven to 350 degrees F. Place the aluminum foil-covered sushi rolls in the oven and bake for 20 minutes. This will warm all the ingredients and crisp the bacon a little more. It will also melt a...

Step 7: **And You're Done!** Serve the sushi with a light crispy vegetable side dish, such as refreshing cucumber sticks, or a green salad. White wine makes an excellent compliment to the meal, especially if it is the same wine used in ...

Visual QA

RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes

http://aclweb.org/anthology/D18-1166



Thank you!

Questions?

TOOPLOOX

www.tooploox.com