

RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes (Supplementary Notes)

Semih Yagcioglu, Aykut Erdem, Erkut Erdem and Nazli Ikizler-Cinbis

Hacettepe University Computer Vision Lab

Dept. of Computer Engineering, Hacettepe University, Ankara, TURKEY

`semih.yagcioglu@hacettepe.edu.tr, {aykut, erkut, nazli}@cs.hacettepe.edu.tr`

1 Supplementary Notes

In the following we provide a few prediction results from the baseline models for each task.

Recipe: Grans-Green-Tomato-Chutney

Step 1: Ingredients: 2.5kg green tomatoes, roughly chopped 0.5kg onions, finely sliced 4 tsp / 30g salt 1L malt vinegar 0.5kg soft light brown sugar 250g sultanas, 1 l...



Step 2: Finely slice your onions and washed green tomatoes, cutting out any bad bits. Add to a large bowl and stir. Add the 4 teaspoons of salt, stir again and then cover with food wrap or a large plate and leave overnight. This will draw out lots of the tomato juices and help enhance the flavours...



Step 3: The next day...Place the litre of vinegar into a large pan. Add the 500g of light brown soft sugar and stir over a medium heat until all the sugar has dissolved. Bring to the boil...



Step 4: ...
:

Step 9: While the jars cool, write some labels showing the date, content and maker. Once cool, add the lids and stick on the labels. ...



Textual Cloze
Style Question

Question Choose the best text for the missing blank to correctly complete the recipe

Ingredients. _____ . Drain and Add the Tomatoes and Onions. Preparing Your Jars.

Answer

A. Sultanas B. Spicy Tomato Chutney. C. Cover and Slice. D. Enjoy.

Hasty Student: **Cover and Slice**
Neural Baseline (Text only): **Sultanas**
Neural Baseline (Multimodal): **Sultanas**

Figure 1: Sample groundtruth and model prediction results for a textual cloze style question (context, question and answer triplet) taken from the RecipeQA test set (Question Id: 1000-12665-0-3-4-6). Here, the context is comprised of step descriptions and images where the questions are generated using the step titles in the recipe. The correct answer is in green. The answers selected by the neural models are correct, marked as green whereas Hasty Student's prediction is wrong and marked as red.

Context Modalities: Titles and Descriptions of Steps

Recipe: Peppermint-Patty-Pudding-Shot

Step 1: Gather Ingredients To make peppermint patty pudding shots you will need: 1 small box of chocolate pudding 3/4 cup of milk 3/4 cup of peppermint schnapps 1 tub of cool whip Crushed peppermint candy ...

Step 2: Mixing of Ingredients First whisk together milk and pudding. Once that is combined add in the peppermint schnapps. Then fold in the cool whip...

Step 3: Prep for Serving I then scoop the pudding into small plastic cups with lids. I buy them from a local Chinese restaurant, they are the perfect size. Throw these in the freezer until you are ready to serve. ...

Step 4: Serve Pull them out of the freezer and sprinkle with the crushed peppermint. You can either lick them out of the cup or eat with a spoon :) I hope you enjoy them as much as we did at Christmas! ...

Visual Cloze
Style Question

Question Choose the best image for the missing blank to correctly complete the recipe



Answer

A. B. C. D.

Hasty Student: C
Neural Baseline (Multimodal): C

Visual Coherence
Style Question

Question Select the incoherent image in the following sequence of images



Answer

A. B. C. D.

Hasty Student: C
Neural Baseline (Multimodal): B

Visual Ordering
Style Question

Question Choose the correct order of the images to make a complete recipe



Answer A. (i)-(ii)-(iii)-(iv) B. (iii)-(i)-(iv)-(ii) C. (ii)-(iv)-(iii)-(i) D. (i)-(iii)-(iv)-(ii)

Hasty Student: B
Neural Baseline (Multimodal): A

Figure 2: Sample visual cloze, visual coherence and visual ordering style question (context, question and answer triplet) taken from the RecipeQA test set (Question Ids: 2000-13317-0-1-2-3, 3000-13317-0-1-2-3, 4000-13317-0-1-2-3). Here, the context is comprised of step titles and descriptions where the questions are generated using the images in the recipe. The correct answers are shown with green frames or in green. Wrong answers are marked as red.