

# Supplementary Material

## Retrieval Augmented Recipe Generation

Guoshan Liu<sup>1,2\*</sup>, Hailong Yin<sup>1,2\*</sup>, Bin Zhu<sup>3</sup>, Jingjing Chen<sup>1,2†</sup>, Chong-Wah Ngo<sup>3</sup>, Yu-Gang Jiang<sup>1,2</sup>

<sup>1</sup>Shanghai Key Lab of Intelligent Information Processing, School of Computer Science, Fudan University

<sup>2</sup>Shanghai Collaborative Innovation Center on Intelligent Visual Computing

<sup>3</sup>Singapore Management University

{gsliu24, hlyin23}@m.fudan.edu.cn, {chenjingjing, ygj}@fudan.edu.cn

{binzhu, cwngo}@smu.edu.sg

### 1. Details of Training Data Organization

As illustrated in Figure 1, we present an example of how the training data is organized in our proposed retrieval augmented framework.

### 2. Retrieved Results

Our SDRA method utilizes a variety of retrieved recipes to enhance model performance. Figure 2, and Figure 4 respectively illustrate some of the training data alongside their corresponding retrieved data. It can be seen that there are significant overlaps between the retrieved ingredients and the ground truth ingredients, which means the retrieved ingredients can provide additional information to enhance the model’s diversity based on the original ingredients. For the retrieved instructions, they provide additional information based on many relevant contents to the ground truth. For example, in Figure 2, concerning the food “award winning soft chocolate chip cookies,” “three out of four retrieved results precisely matched the preparation process for “chocolate chip cookies.” Steps like “Preheat oven to 350 degrees f (175 degrees c).” and “Blend in the dry ingredients, then fold in the chocolate chips.” align with the ground truth instructions, as indicated by the BOLD parts in the diagram, ensuring that the additional information effectively provides content relevant to the original query. Additionally, steps from Retrieved recipe 4 like “Allow cookies to cool for 1 minute on baking sheets before transferring to wire racks to cool completely.” add detailed descriptions to the cookie-making process, enhancing the post-preparation flow and providing the model with more detailed and comprehensive information to complement and diversify the instructions found in the ground truth. Similarly, Figure 4 demonstrates that the retrieved recipes add useful contextual information

\*Equal contribution.

†Jingjing Chen is the corresponding author.

Table 1. Ablation study of the independence of retrieved ingredients and retrieved instructions. “SDRA(Matched)” refers to the situation where retrieved ingredients and instructions are matched, while “SDRA(Independent)” indicates that retrieved ingredients and instructions are mutually independent.

Methods	BLEU	SacreBLEU	ROUGE L
LLaVA-FT	28.32	5.88	38.18
+SDRA(Matched)	28.83	6.17	38.30
<b>+SDRA(Independent)</b>	<b>29.23</b>	<b>6.21</b>	<b>38.43</b>

related to the original queries to the Ground Truth. This guides the model to generate recipes more effectively, perfectly leveraging retrieval augmentation technology for enhanced predictions. This adds fundamental procedural details to the more complexly seasoned ground truth instruction, enabling the model to accurately generate responses to queries about the food image using the given context.

### 3. Ablation of Stochastic Diversified Retrieval Augmentation (SDRA)

As described in Section 4.3.1, to investigate whether the ingredients and instructions added before  $Q_{instructions}$  in Recipe demonstration  $R$  need to come from the same top retrieval results, i.e., having the same  $K$  value, we compared experiments where ingredients and instructions were sourced from the same top  $K$  retrieval results, specifically concatenating the top  $K$  ingredients and top  $K$  instructions in sequence before  $Q_{instructions}$ . Table 1 indicates that the model with independent ingredients and instructions performs better, as our method SDRA relies on a broader range of knowledge and diversified retrieval information settings.

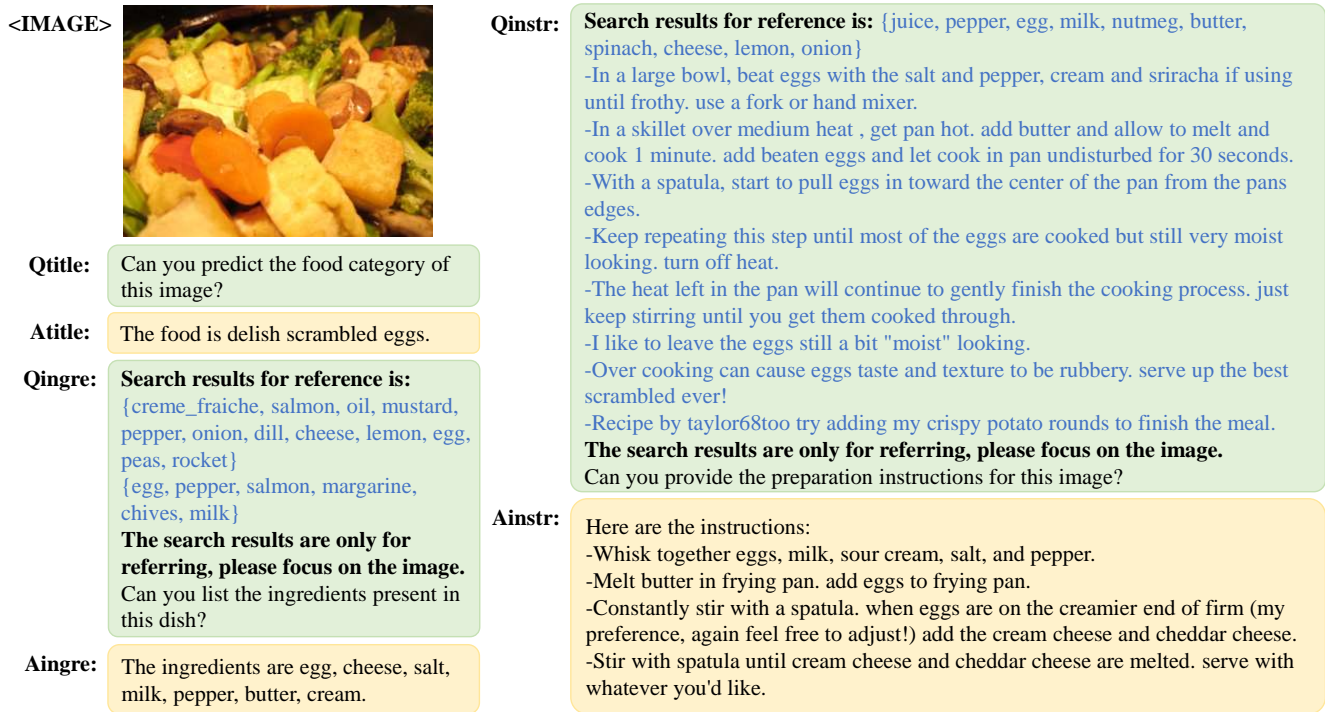


Figure 1. Our training data organization. **BOLD** serves as a fixed prompt, the blue font highlights the three groups of ingredients and one group of instructions retrieved, and the rest of the content represents the ground truth.

#### 4. Ablation of Self-consistency Ensemble Voting

To demonstrate the impact of self-consistent ensemble voting, Table 2 examines the results of our model when using cosine similarity, BLEU, SacreBLEU, and ROUGE L to calculate the mutual agreement among the generated recipe candidates. Note that  $S = 1$  refers to no voting adopted, which is the baseline for our model with Self-consistency Ensemble Voting when  $S > 1$ . In other words, the input prompt is concatenated with the top 1 retrieved ingredients and instruction before  $Q_{instructions}$  during inference. To ensure the fairness of the experiments, these experiments start with the SDRA(top 50) model. The experiments with different voting groups indicate using the top  $S$  retrieved instructions and ingredients for  $S$  separate predictions as described in Section 3.3, and then selecting the score of the prediction with the highest score of mutual agreement as the final output. As  $S$  is increased from 1 to 11, the voting results based on Cosine Similarity as the scoring metric show a steady improvement across all metrics. The results based on the other three scoring metrics also exhibit fluctuating improvements across various indicators, and after  $S=7$ , there is a trend of slower growth, no further increase, or even a decrease in some metrics.

Table 2. Ablation study of Self-consistency Ensemble Voting. Four scoring metrics are examined to evaluate the mutual agreement among generated recipes, including cosine similarity, BLEU, SacreBLEU, and ROUGE L. “S” refers to the number of generated recipes for ensemble voting. “Sum” is the sum of 3 evaluation metrics.

Scoring metric	Number	BLEU	SacreBLEU	ROUGE L	Sum
Cosine Similarity	S=1	29.23	6.21	38.43	73.87
	S=3	29.68	6.31	38.68	74.67
	S=5	30.12	6.39	38.66	75.17
	S=7	30.07	6.41	38.84	75.32
	S=9	30.11	6.42	38.91	75.44
	S=11	30.11	6.42	38.93	75.47
BLEU	S=1	29.23	6.21	38.43	73.87
	S=3	29.25	6.27	38.88	74.40
	S=5	29.47	6.35	38.93	74.75
	S=7	29.50	6.36	39.09	74.95
	S=9	29.53	6.37	39.09	74.99
	S=11	29.52	6.37	<b>39.13</b>	75.02
SacreBLEU	S=1	29.23	6.21	38.43	73.87
	S=3	28.79	6.21	39.11	74.11
	S=5	29.30	6.30	38.83	74.43
	S=7	29.33	6.32	39.01	74.66
	S=9	29.33	6.32	39.10	74.75
	S=11	29.30	6.35	39.10	74.74
ROUGE L	S=1	29.23	6.21	38.43	73.87
	S=3	32.21	6.62	36.99	75.82
	S=5	32.92	6.81	36.90	76.63
	S=7	33.08	6.82	36.80	76.70
	S=9	33.26	6.84	36.67	76.77
	S=11	<b>33.53</b>	<b>6.88</b>	36.60	<b>77.01</b>



**Ground truth**

Title: award winning soft chocolate chip cookies  
 Ingredients: Egg, sugar, extract, vanilla, butter, chips, baking\_soda, walnuts, flour.  
 Instructions:  
 -Preheat oven to 350 degrees f (175 degrees c).  
 -Sift together the flour and baking soda, set aside.  
 -In a large bowl, cream together the butter, brown sugar, and white sugar.  
 -Beat in the instant pudding mix until blended. stir in the eggs and vanilla.  
 -Blend in the flour mixture. finally, stir in the chocolate chips and nuts.  
 -Drop cookies by rounded spoonfuls onto ungreased cookie sheets.  
 -Bake for 10 to 12 minutes in the preheated oven.  
 -Edges should be golden brown.

**Retrieved recipe 1**

Title: oatmeal chocolate coconut chewy  
 Ingredients: **Egg, sugar, walnuts, extract**, milk, coconut, **baking\_soda, chips, butter, salt, flour**, oats.  
 Instructions:  
 -Preheat oven to 375 degrees f (190 degrees c).  
**-In a large bowl, cream together the butter, brown sugar and white sugar until smooth.**  
**-Beat in the eggs one at a time, then stir in the milk and vanilla.**  
 -Combine the flour, baking soda and salt; stir into the sugar mixture until well blended.  
**-Stir in the oats, chocolate chips, walnuts and coconut** until evenly distributed. drop by rounded tablespoons onto ungreased cookie sheet.  
**-Bake 10 to 12 minutes in the preheated** oven for a chewy cookie or 14 minutes for a firmer cookie.  
 -Cool for 1 minute on the cookie sheet and then remove to wire rack. cool completely and then store in tightly sealed container.

**Retrieved recipe 2**

Title: the perfect chocolate chip cookie  
 Ingredients: **Egg, sugar, oil, extract, chips, baking\_soda, salt, flour.**  
 Instructions:  
**-Preheat oven to 350 degrees f (175 degrees c).**  
**-Stir together the flour, baking soda and salt; set aside.**  
**-In a large bowl, cream together the vegetable oil, brown sugar and white sugar until smooth.**  
**-Beat in the eggs one at a time, then stir in the vanilla and almond extracts.**  
**-Blend in the dry ingredients, then fold in the chocolate chips.**  
**-Drop by rounded spoonfuls onto ungreased cookie sheets.**  
**-Bake for 8 to 10 minutes in the preheated oven.**  
 -Allow cookies to cool on baking sheet for 5 minutes before removing to a wire rack to cool completely.

**Retrieved recipe 3**

Title: anna's chocolate chip cookies  
 Ingredients: **Egg, sugar, extract, baking\_soda**, chips, **butter, salt, flour.**  
 Instructions:  
 -Preheat the oven to 375 degrees f (190 degrees c).  
**-In a large bowl, cream together the butter and sugar until smooth.**  
**-Beat in the vanilla and eggs one at a time.** combine the flour, baking soda and salt; stir into the sugar mixture.  
**-Finally, mix in the chocolate chips.**  
**-Drop by tablespoonfuls onto ungreased cookie sheets.**  
**-Bake for 8 to 10 minutes in the preheated oven,** or until edges are golden. remove from baking sheet to cool on wire racks.

**Retrieved recipe 4**

Title: light chocolate chip cookies  
 Ingredients: **Egg, sugar, extract, water, baking\_soda, chips, butter, salt, flour.**  
 Instructions:  
**-Preheat oven to 350 degrees f (175 degrees c).**  
**-In a medium bowl, cream the butter with the brown and white sugars. stir in the vanilla, egg white, and water.**  
**-Sift together the flour, baking soda, and salt; stir into the creamed mixture.**  
 -Mix in the chocolate chips.  
**-Drop dough by heaping spoonfuls onto ungreased cookie sheets. -Bake for 8 to 10 minutes in the preheated oven.**  
 -Allow cookies to cool for 1 minute on baking sheets before transferring to wire racks to cool completely.

Figure 2. Comparison between retrieval results and ground truth. BOLD represents the relevant and similar parts between the retrieved results and the ground truth. This is the ground truth for “award winning soft chocolate chip cookies” and its corresponding retrieved recipes.

### 5. More Qualitative Examples

Figure 5 displays additional qualitative results. In these three cases, our predictions closely align with the ground truth “GT”, whereas other models—“LLaVA-FT”, “Inverse Cooking” [1], and “FoodLLM” [2]—exhibited various hallucinations. For instance, in Figure 5 (a), “LLaVA-FT” incorrectly predicted “coconut” as “pudding”, “Inverse Cooking” identified the food as “cheesecake”, and “FoodLLM” mistakenly labeled blueberries as “chocolate”, though their overall instruction predictions were reasonably accurate. In Figure 5 (b), “LLaVA-FT” and “FoodLLM” made slight errors in ingredient prediction but were generally close to “GT”, however, their instruction predictions did not closely match “GT”, leading to lower test scores, whereas “Inverse Cooking” provided overly simplistic and less detailed predictions. In Figure 5 (c), both “LLaVA-FT” and “FoodLLM” misidentified “meat” as “ground beef”, affecting the accuracy of their instruction predictions, while “Inverse Cooking” was more accurate, though it lacked the detail of our model, such as “4 (1/2-inch-thick) patties”.

Figure 3 demonstrates the comparative results of our best model during the inference process, with and without the

addition of retrieval information. It is evident that the predictions made without retrieval information are less accurate and comprehensive than those made with it. For instance, in Figure 3(a), our method precisely predicts “cook and stir” instead of merely “cook,” in comparison to the results without added retrieval information. In Figure 3(b), our method’s predictions almost perfectly match the ground truth, further underscoring the importance of incorporating retrieval information.

Figure 6 shows the four generated recipe candidates, with prediction 1 being the output with the highest confidence. It overlaps with the ground truth in ‘shaker with ice’ and ‘a chilled cocktail glass.’ It can be observed that Prediction 3 and Prediction 4’s ‘vodka, blue curacao, and pineapple juice’ received supplementary information from their corresponding retrieved ingredients and instructions, such as ‘blue curacao’ and ‘vodka, curacao, and soda’. However, the description in the retrieved information differs significantly from the ground truth, resulting in a lower confidence score for the predictions. This demonstrates that the self-consistency voting mechanism effectively selects the best recipe.



Ours(no RAG): heat oil in large skillet on medium-high heat. add chicken; cook 5 min. or until done, stirring occasionally. add broccoli; cook 3 min. or until crisp-tender, stirring occasionally. add remaining ingredients; cook 2 min. or until heated through, stirring occasionally.

Ours: heat oil in large skillet on medium-high heat. add chicken; **cook and stir** 5 min. or until done. add onions; cook and stir 3 min. or until crisp-tender. add broccoli; cook and stir 3 min. or until crisp-tender. add remaining ingredients; **mix well. bring to boil. reduce heat to medium-low**; cover. simmer 15 min. or until rice is tender.

GT: heat oil in large skillet on medium-high heat. add meat and garlic; **cook and stir** 4 min. or until meat is cooked through. add broth and vegetables; **mix well. bring to boil.** stir in rice and dressing; cover. **reduce heat to low**; simmer 5 min. remove from heat. let stand 5 min. fluff with fork.

(a)



Ours(no RAG): in a medium bowl, whisk together the olive oil, lemon juice, oregano, salt, and pepper. toss with the quinoa to coat. in a medium saucepan, bring the broth to a boil. stir in the quinoa. cover, reduce heat to low, and simmer 15 minutes. remove from heat, and let stand, covered, 5 minutes more. fluff with a fork, and serve.

Ours: **bring quinoa and water to a boil in a saucepan. reduce heat to medium-low, cover, and simmer until quinoa is tender and water has been absorbed, 10 to 15 minutes.** set aside to cool. whisk olive oil, lemon juice, white wine vinegar, garlic, salt, and pepper together in a bowl...**toss to coat.** refrigerate until chilled, at least 1 hour.

GT: **bring broth and quinoa to a boil in a saucepan. reduce heat to medium-low, cover, and simmer until quinoa is tender and water has been absorbed, 15 to 20 minutes.**...set aside to cool, ... mix tomatoes, parsley, kalamata olives, onion, feta cheese, olive oil, vinegar, and garlic into quinoa...and **toss to coat.** chill in refrigerator, 1 to 4 hours.

(b)

Figure 3. Qualitative results of whether the best model uses retrieval information for inference. The yellow highlights indicate the parts where our method provides more accurate predictions on the best model compared to the method without retrieval information.

## 6. Consistency between the Generated Ingredients VS Retrieved Ingredients

We compared the retrieved ingredients with the ingredients predicted by the model. As shown in Figure 7, in the first case, it can be observed that the model prediction partly overlaps with the ground truth. However, the retrieved ingredient list misled the model by introducing ‘cheese’ and ‘vanilla’, which are not part of the ground truth. The retrieved ‘walnuts’ were beneficial, helping the model predict ‘walnuts’ correctly. In the second case, the retrieved ingredients misled the prediction by introducing incorrect ingredients (‘chicken’, ‘pepper’), which were not part of the ground truth. Nonetheless, the model was able to predict other ingredients like ‘flour’, ‘oil’, ‘soy sauce’, and ‘breadcrumbs’ accurately. In both cases, the retrieved ingredients significantly influenced the model’s prediction. When the retrieval stage introduced irrelevant ingredients (like chicken and pepper in the second case), these errors propagated into the prediction. However, the model was still able to predict a portion of the correct ingredients, particularly when they overlapped with the retrieved ones. The results underscore the importance of accurate ingredient retrieval, as incorrect retrievals can lead to faulty predictions, even if the model has the capacity to predict well when provided with accurate inputs.

## 7. Ingredients Extracted from Instructions and Directly Predicted

Figure 8 shows a comparison between ingredients extracted from generated instructions and those directly predicted

dicted. In the first case, ingredients like ‘greens’ and ‘tomato’, which are visually prominent in the image, are successfully predicted. However, ‘jicama’ is detected in the instructions and appears in the instruction ground truth, but is missing from the directly predicted ingredients. This might be because the model, when generating instructions, learns common cooking pairings and automatically adds ingredients missed during ingredient prediction. Additionally, the retrieved instructions may have included ‘jicama’. It is worth noting that the ingredient ‘pineapple’ in the extracted ingredients was identified as ‘fruit’. While this is correct in a broad or human sense, it would be considered incorrect when calculating classification metrics. In the second case, the extracted ingredients are almost entirely accurate, while the predicted ingredients are fewer but still quite accurate. This could be because the model uses more contextual information when generating instructions, which helps it infer ingredients that were not captured during standalone ingredient prediction. Overall, the ingredient predictions are fairly accurate, and there is a significant overlap between the extracted and predicted ingredients.

## 8. Retrieval Failure Case

To explore retrieval failure cases, we analyzed some examples of unsuccessful retrievals. Figure 9 provides an example of a retrieval failure. The prediction is influenced by retrieved information, like “cut into bars,” and mixing nuts into the batter early. In the ground truth, nuts are sprinkled on top at the end. Figure 10 presents two retrieved instructions for the salad and their respective predictions. It can be seen that the completeness of the preparation



**Ground truth**  
 Title: mango shrimp  
 Ingredients: Jalapeno, mango, turmeric, masala, seeds, clove, juice, yogurt, shrimp, cilantro, cashews, butter, salt, ginger.  
 Instructions:  
 -In a blender, combine the cashews, ginger, garlic, jalapeno, mustard seeds, garam masala and turmeric and pulse until the ginger and garlic are finely chopped.  
 -Add the mango nectar and lemon juice and puree until smooth, scraping down the side of the bowl.  
 -Add the yogurt and 1 tablespoon of salt and pulse to blend, pour the mixture into a large bowl.  
 -Add the shrimp and toss to coat, refrigerate for 2 hours, stirring once or twice, light a grill.  
 -Thread the shrimp on pairs of skewers, being sure to leave on some of the marinade.  
 -Brush with the melted butter and sprinkle with salt.  
 -Oil the grate and grill the shrimp over high heat, turning occasionally, until lightly charred and cooked through, about 8 minutes. serve the shrimp with the cilantro and yogurt sauce.

**Retrieved recipe 1**  
 Title: shrimp and corn chowder  
 Ingredients: Salsa, cumin, **juice**, plum, milk, **clove, shrimp**, pepper, cilantro, fish, onion, scallion, oil, corn, **salt**.  
 Instructions:  
 -In a large, shallow glass or stainless-steel bowl, **toss the shrimp with two-thirds of the minced garlic, the scallions, lime juice and 1 teaspoon of salt.**  
 -Cover with plastic wrap and **refrigerate for at least 1 hour or for up to 3 hours.**  
 -Simmer over moderately low heat until very flavorful, about 20 minutes.  
 ...  
 -Return the puree and the strained broth to the saucepan and bring to a simmer.  
 -**Add the shrimp and its marinade and cook over moderate heat until the shrimp are just opaque throughout, about 2 minutes.**

**Retrieved recipe 2**  
 Title: sichuan peppercorn shrimp  
 Ingredients: Chili, **jalapeno, juice, clove**, scallion, oil, **salt**, peppercorn.  
 Instructions:  
 -In a small skillet, toast the peppercorns over moderate heat until fragrant, about 30 seconds; let cool. transfer the peppercorns to a mortar or spice grinder and grind to a powder.  
 -Put the shrimp in a bowl, **toss with 1 teaspoon of the ground peppercorns and season with salt.** in a medium skillet, heat 1 tablespoon of the vegetable oil.  
 ...  
 -**Add the shrimp and lime juice and stir until the shrimp are just cooked through, 1 minute.**  
 -**Season with salt and transfer to a bowl.** garnish with the sliced scallion, drizzle with the chile oil and serve.

**Retrieved recipe 3**  
 Title: almond-crust chicken wings  
 Ingredients: Almonds, vinegar, cumin, **clove**, chicken, pepper, mayonnaise, lemon, paprika, oil.  
 Instructions:  
 -In a large bowl, **combine the 1/3 cup of olive oil with the paprika, cumin, cayenne and garlic.**  
 -**Stir in the vinegar and season with salt and pepper.** add the wings and almonds and toss.  
 -Spread the wings and almonds on a large baking sheet in a single layer and roast for about 25 minutes, until cooked through. meanwhile, in a mini food processor, **combine the mayonnaise with the lemon zest and juice, add the remaining 1/4 cup of olive oil and process until smooth; season the aioli with salt and pepper.**  
 -Turn on the broiler and broil the chicken wings, turning once until they are lightly crisp, 2 to 3 minutes.  
 -Transfer the wings and almonds to a platter and serve with the lemon aioli.

**Retrieved recipe 4**  
 Title: chicken souvlaki  
 Ingredients: **Juice**, chicken, **clove**, pepper, lemon, wine, oregano, onion, oil, **salt**, bay\_leaves.  
 Instructions:  
 -Combine the oil, lemon juice, wine, garlic, oregano, lemon zest, 1 teaspoon salt and the pepper in a large nonreactive bowl and **whisk until blended and the salt is dissolved.**  
 -Taste for seasoning, **adding salt as necessary; the mixture should be highly seasoned.**  
 -Add the chicken and turn to coat. let marinate, at room temperature, for 30 minutes, turning occasionally. soak the bay leaves in a bowl of cold water for 20 minutes.  
 -Preheat the grill to high. when ready to cook, break the onion quarters into individual layers. drain the bay leaves. remove the chicken cubes from the bowl, reserving whatever marinade is left, and thread onto the skewers, placing a piece of onion and bay leaf between each and dividing evenly.

(a) The example of the ground truth for “mango shrimp” and its corresponding retrieved recipes.



**Ground truth**  
 Title: chocolate cherry drops  
 Ingredients: Egg, sugar, walnuts, extract, cocoa, baking\_soda, cherries, butter, salt, flour.  
 Instructions:  
 -Preheat oven to 350 degrees f (175 degrees c).  
 -Grease cookie sheets. in a large bowl, cream together the butter and sugar until smooth.  
 -Blend in the egg and vanilla. combine the flour, cocoa, baking soda and salt; stir into the creamed mixture.  
 -Finally, mix in the chopped cherries and walnuts.  
 -Drop by rounded spoonfuls onto the prepared cookie sheets.  
 -Bake for 8 to 10 minutes in the preheated oven.  
 -Allow cookies to cool on baking sheet for 5 minutes before removing to a wire rack to cool completely.

**Retrieved recipe 1**  
 Title: high altitude banana chocolate chip cookies  
 Ingredients: **Egg, sugar, extract, butter**, banana, baking\_powder, chips, **baking\_soda, salt, flour**.  
 Instructions:  
 -Preheat the oven to 375 degrees f (190 degrees c).  
 -Sift together the flour, baking powder, baking soda and salt, set aside.  
 -**In a large bowl, cream together the butter, sugar and brown sugar.**  
 -**Beat in the eggs, one at a time, then stir in the vanilla and mashed banana.**  
 -**Mix in the dry ingredients until just blended, then fold in chocolate chips.**  
 -Drop by rounded spoonfuls onto prepared cookie sheets.  
 -Bake for 11 to 13 minutes in the preheated oven.  
 -**Allow cookies to cool on baking sheet for 5 minutes before removing to a wire rack to cool completely.**

**Retrieved recipe 2**  
 Title: cracked sugar cookies ii  
 Ingredients: **Egg, sugar**, shortening, **extract**, baking\_powder, **salt, flour**.  
 Instructions:  
 -**In a large bowl, cream together the shortening and 1 1/2 cups sugar until smooth.**  
 -**Beat in the eggs one at a time then stir in the vanilla.**  
 -**Combine the flour, baking powder and salt; stir into the creamed mixture until well blended.**  
 -Cover dough and refrigerate for at least 1 hour. preheat oven to 350 degrees f (175 degrees c). roll dough into 1 inch balls and roll the balls in the remaining sugar.  
 -Place cookies 2 inches apart onto ungreased cookie sheets.  
 -**Bake for 8 to 9 minutes in the preheated oven,** or until just barely golden.  
 -**Allow cookies to cool on the baking sheet for 2 minutes before removing to wire racks to cool completely.**

**Retrieved recipe 3**  
 Title: chocolate chocolate chip cookies i  
 Ingredients: **Egg, sugar, walnuts, extract, cocoa, baking\_soda, chips, butter, salt, flour.**  
 Instructions:  
 -**Preheat oven to 350 degrees f (175 degrees c).**  
 -In large bowl, beat butter, sugar, eggs, and vanilla until light and fluffy.  
 -**Combine the flour, cocoa, baking soda, and salt; stir into the butter mixture until well blended.**  
 -Mix in the chocolate chips and walnuts.  
 -**Drop by rounded teaspoonfuls onto ungreased cookie sheets.**  
 -**Bake for 8 to 10 minutes in the preheated oven, or just until set.**  
 -Cool slightly on the cookie sheets before transferring to wire racks to cool completely.

**Retrieved recipe 4**  
 Title: crunchy chocolate chippers  
 Ingredients: **Egg, sugar, oil, extract**, cereal, margarine, milk, chips, **baking\_soda, salt, flour**, oats.  
 Instructions:  
 -**Preheat oven to 350 degrees f (175 degrees c).**  
 -Grease cookie sheets. **stir together the flour, baking soda and salt; set aside.**  
 -In a large bowl, **cream together the margarine, brown sugar and white sugar until smooth.**  
 -Beat in the eggs, one at a time, then stir in the vanilla, vegetable oil and milk.  
 -**Blend in the flour mixture** before stirring in the oats, frosted corn flakes and chocolate chips, drop by heaping spoonfuls onto the prepared cookie sheets.  
 -**Bake for 8 to 10 minutes in the preheated oven. allow cookies to cool on baking sheet for 5 minutes before removing to a wire rack to cool completely.**

(b) The example of the ground truth for “chocolate cherry drops” and its corresponding retrieved recipes.

Figure 4. Comparison between retrieval results and ground truth.



LLaVA-FT: beat cream cheese and sugar in large bowl with mixer until blended. add cool whip; mix well. spread onto bottom of 13x9-inch pan. refrigerate 1 hour. meanwhile, beat pudding mix and milk with whisk 2 min. ; pour over cream cheese layer. refrigerate 1 hour. spread remaining cool whip over pudding layer; refrigerate 1 hour. top with remaining ingredients. refrigerate 1 hour.

Inverse cooking: ..Beat cream cheese, sugar and vanilla in large bowl with mixer until blended. Add eggs. 1 at a time, mixing on low speed after each just until blended. Pour into 9-inch springform pan sprayed with cooking spray...Meanwhile, beat pudding mix and milk in medium bowl with whisk 2 min...Top cheesecake with berries just before serving.

FoodLMM: ...Line a 9-inch square pan with foil, with ends of foil extending over sides...Beat cream cheese, sugar and vanilla in large bowl with electric mixer on medium speed until well blended. Add cool whip; mix well. Spread over crust...Refrigerate 1 hour. Arrange berries in rows over cream cheese mixture. Drizzle with melted chocolate...

Ours: ...line 13x9-inch pan with foil, with ends of foil extending over sides. mix graham crumbs and butter; press onto bottom of prepared pan. beat cream cheese, sugar and vanilla with mixer until blended. add eggs. 1 at a time, mixing on low speed after each just until blended. pour over crust. bake 40 to 45 min. or until center is almost set...refrigerate 4 hours. top with whipped topping just before serving...

GT: ...line 13x9-inch pan with foil, with ends of foil extending over sides. mix graham crumbs, 3 tsp. sugar and butter; press firmly onto bottom of pan...beat cream cheese, 1 cup sugar and vanilla in large bowl with mixer until well blended...add eggs, 1 at a time, mixing on low speed after each just until blended. pour over crust. bake 40 min. or until center is almost set...refrigerate 4 hours...top with cool whip; sprinkle with coconut to resemble a cross. fill in corners with berries to resemble the dominican republic flag.

(a)



LLaVA-FT: preheat oven to 400 degrees f (200 degrees c). grease 12 muffin cups or line with paper muffin liners. combine flour, baking powder, and pepper in a large bowl...stir the milk mixture into the flour mixture until just combined. fold in the cheese and spinach until evenly combined. spoon the batter into prepared muffin cups. bake in the preheated oven until golden and the tops spring back when lightly pressed, about 20 minutes.

Inverse cooking: Preheat oven to 400 degrees f. In a large bowl, combine the ricotta, spinach, eggs, flour, salt and pepper. Stir in the ricotta cheese. Fill greased muffin cups two-thirds full. Bake for 20 minutes or until golden brown.

FoodLMM: Preheat oven to 400 degrees f (200 degrees c). Lightly grease a 12 cup muffin pan. In a large bowl, beat together eggs, milk, oil, salt, and pepper. Mix in cheese, spinach, and feta...Stir in chives. Spoon batter into prepared muffin pan. Bake in preheated oven for 20 minutes, or until a toothpick inserted into a muffin comes out clean.

Ours: ...grease 12 muffin cups or line with paper muffin liners. combine flour, baking powder, baking soda, and salt in a large bowl. mix egg, milk, oil, and cheddar cheese in a bowl; stir into flour mixture until just moistened... fill prepared muffin cups 2/3 full with batter. bake in preheated oven until a toothpick inserted into the center of a muffin comes out clean, about 35 minutes.

GT: ...lightly grease 12 cup muffin cups. mix the flour, baking powder, baking soda, and salt together in a mixing bowl. stir the melted butter, egg, milk, spinach, and cheddar cheese together in a large mixing bowl until evenly blended. slowly stir in the flour mixture to form a batter. spoon about 2 tablespoons into each muffin cup. bake in preheated oven until a toothpick inserted into the center of a muffin comes out clean, about 35 minutes.

(b)



LLaVA-FT: Preheat the oven to 375 degrees f. Place the brioche buns on a baking sheet and bake for 5 minutes or until golden brown. In a mixing bowl, combine the ground beef, cheddar cheese, green sauce, and parsley. Place a small amount of the meat mixture on top of each bun. Serve immediately.

Inverse cooking: Heat grill to medium heat. Mix meat and 2 tbsp. Barbecue sauce; shape into 4 (1/2-inch-thick) patties. Grill 5 to 6 min. On each side or until done (160 degrees f). Top with cheese; grill 1 min...Fill buns with cheeseburgers, bacon and remaining barbecue sauce.

FoodLMM: Mix ground beef, bread crumbs, egg, onion, garlic, salt, pepper, and 1/2 cup of the cheese. Shape into 4 patties. Heat oil in a large skillet over medium-high heat. Add patties; cook 5 minutes on each side or until well browned. Remove from skillet; cover to keep warm. Add onions to skillet; cook 5 minutes or until tender. Stir in ketchup, mustard, and remaining 1/2 cup cheese. Serve patties topped with onion mixture.

Ours: heat grill to medium heat. mix meat, our favorite chimichurri just until blended; shape into 6 (1/2-inch-thick) patties. grill 5 min. on each side or until done (160 degrees f). top with remaining cheese; grill 1 min. or until melted. fill buns with cheeseburgers and remaining chimichurri.

GT: heat grill to medium-high heat. mix meat and 3 tsp. our favorite chimichurri just until blended; shape into 6 (1/2-inch-thick) patties. grill 5 to 6 min. on each side or until done (160f). top with singles; grill 1 min. or until melted. fill buns with cheeseburgers and remaining chimichurri.

(c)

Figure 5. Additional qualitative results. The ingredients in generated recipes that overlap with ground truth (“GT”) are highlighted in yellow, while details in the instructions that match the GT are shown in blue. Otherwise, the incorrect generation results are displayed in red. Best viewed in color.

steps also influences the comprehensiveness of the prediction, and the predicted ingredients tend to overlap with the ingredients mentioned in the retrieved instructions. These results clearly show that when the retrieved instruction is more comprehensive and accurate, the resulting prediction will also be better.

## 9. Broader Ingredient Categories

To avoid situations where ingredients in the prediction and ground truth are very similar but are counted as completely incorrect due to different terminology in the ingredients list—such as ‘pasta’ and ‘spaghetti’—we designed an alternative penalty calculation method. We identified similar ingredients in the list and grouped them into 42 broader categories, as shown in Table 3. We apply a reduced penalty (0.5 weight) for ingredients within the same category and a

full penalty (1 weight) for those across categories.

## References

- [1] Amaia Salvador, Michal Drozdal, Xavier Giró-i Nieto, and Adriana Romero. Inverse cooking: Recipe generation from food images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10453–10462, 2019. 3
- [2] Yuehao Yin, Huiyan Qi, Bin Zhu, Jingjing Chen, Yu-Gang Jiang, and Chong-Wah Ngo. Foodlmm: A versatile food assistant using large multi-modal model. *arXiv preprint arXiv:2312.14991*, 2023. 3


	GT: combine ingredients in a cocktail <b>shaker with ice</b> . shake vigorously and strain <b>into a chilled cocktail glass</b> . <b>garnish with a cherry</b> .	Confidence scores
	Prediction 1: combine all ingredients in a <b>shaker with ice</b> . shake and strain <b>into a chilled cocktail glass</b> . <b>garnish with a cherry</b> .	0.64
	Prediction 2: combine all ingredients <b>in a shaker with ice</b> . shake and strain <b>into a chilled cocktail glass</b> .	0.52
	Prediction 3: fill a highball glass with ice. add the <b>vodka, blue curacao, and pineapple juice</b> . stir. <b>garnish with a cherry</b> and a <b>maraschino</b> cherry.	0.59
	Prediction 4: fill a tall glass with ice. add the <b>vodka, blue curacao, and pineapple juice</b> . stir. <b>garnish with a cherry</b> and a <b>maraschino</b> cherry.	0.58
Retrieved ingredients 1: curacao, juice, schnapps, vodka.		
Retrieved instruction 1: pour all ingredients except pineapple juice <b>into a cocktail shaker</b> , and mix well. pour <b>into a cocktail glass</b> filled with ice and add pineapple juice to taste. mix well.		
Retrieved ingredients 2: lime, liqueur, vodka, juice, curacao, ice.		
Retrieved instruction 2: pre-chill vodka in freezer ... pre-chill large/9 ounce wine glass in freezer. add about 4-6 ice cubes into the wine glass. add <b>vodka, raspberry liqueur, blue curacao, lime cordial, and cranberry cocktail</b> . stir or shake gently and serve.		
Retrieved ingredients 3: rum, lemonade, curacao, vodka.		
Retrieved instruction 3: fill a highball glass with ice and add vodka and <b>coconut rum</b> . fill with <b>lemonade</b> . splash <b>blue curacao</b> . garnish with <b>pineapple leaf and lemon wheel</b> , if desired.		
Retrieved ingredients 4: lime, vodka, curacao, sprite, ice.		
Retrieved instruction 4: fill a 6-8 oz glass with crushed ice. add <b>vodka, curacao and soda</b> . give the <b>lime</b> a squeeze over the drink. garnish with a slice of <b>lime</b> and serve.		

Figure 6. Qualitative results of Ablation of Self-consistency Ensemble Voting. **BOLD** indicates the parts that overlap with the ground truth, while red text highlights the hallucinated outputs. The retrieved ingredients and instructions correspond to each prediction, and the confidence score is calculated from a 4x4 cosine similarity matrix when S=4, with each row averaged to obtain the final score.



	Retrieved ingredients : <b>chocolate, sugar, milk</b> , pecans, coconut, <b>butter, flour, caramel</b>
	Prediction : <b>butter</b> , cream, <b>sugar</b> , cheese, vanilla, egg, <b>chocolate, walnuts</b>
	Ground truth : <b>chocolate, walnuts</b> , baking_soda, oats, <b>flour, milk</b> , salt, <b>caramel, butter, sugar</b>
	Retrieved ingredients : <b>egg, breadcrumbs</b> , milk, <b>chicken, pepper</b> , lemon, <b>oil, flour</b>
	Prediction : <b>flour</b> , cornstarch, <b>salt, pepper, chicken, oil</b> , garlic, <b>mayonnaise, soy_sauce</b>
	Ground truth : scallion, spinach, <b>oil, egg, flour, soy_sauce, breadcrumbs</b> , seeds, ginger, <b>salt, mayonnaise</b> , pork, clove

Figure 7. Retrieved and predicted ingredients. **BOLD** represents the overlap of ingredients between both Retrieved ingredients and Predicted ingredients with the ground truth ingredients. The ingredients in red font are the ones incorrectly predicted due to errors in the retrieved ingredients.

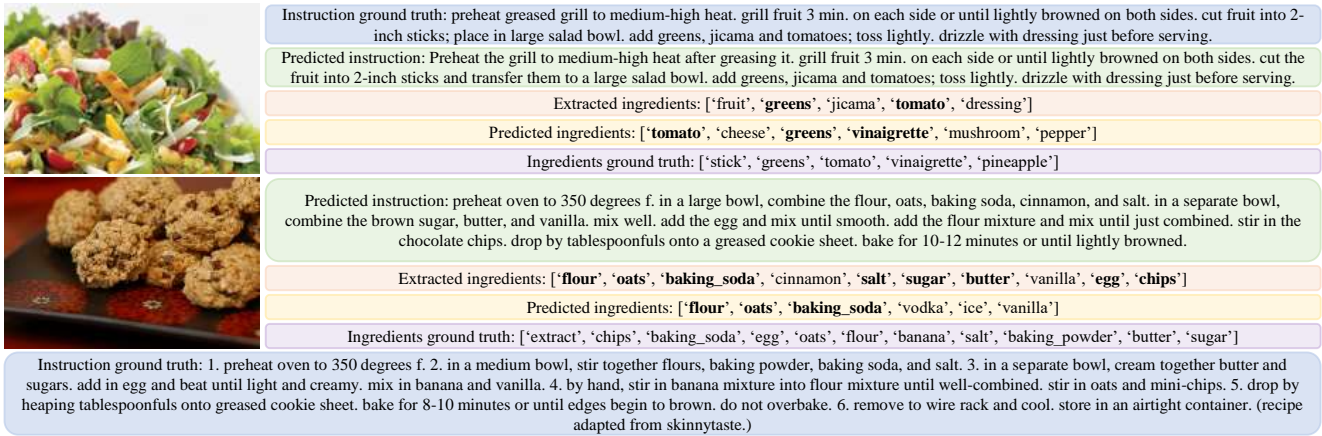


Figure 8. Predicted and ground truth ingredients, instructions, and ingredients extracted from the predicted instructions. **BOLD** represents the overlap of ingredients between both Extracted ingredients and Predicted ingredients with the ground truth ingredients.

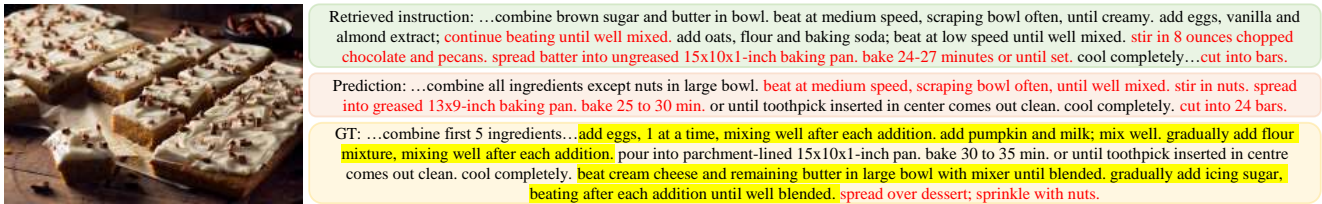


Figure 9. Retrieval failure examples. Red text shows prediction errors influenced by the retrieved instruction, and yellow highlights indicate missing content in the prediction.

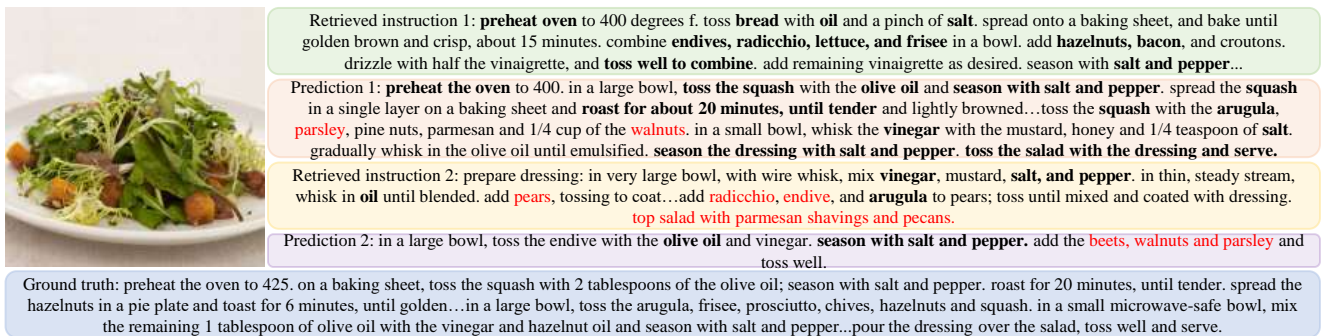


Figure 10. Two sets of retrieval results for "butternut squash salad with hazelnuts" and their corresponding predictions. **BOLD** represents the parts that overlap with the ground truth, while red text indicates those that do not match the ground truth.



<b>Broader Category</b>	<b>Ingredients</b>
Pasta and Derivatives	spaghetti, pasta, penne, linguine, fettuccine, tortellini, ravioli, vermicelli, lasagna_sheet, orecchiette, fusilli, conchiglie, cavatelli, spaghettini, manicotti
Cheese	cheese, medium_cheddar, parmesan_rind, parmigiano, gorgonzola, mascarpone, ricotta_salata, queso_fresco, feta
Meat	beef, chicken, pork, lamb, veal, bacon, ham, sausage, rib, tenderloin, fillets, steak, roast, duck, meatballs, sirloin, liver, tender_quick
Fish and Seafood	tuna, salmon, shrimp, lobster, oyster, cod, scallops, sardines, anchovies, crabmeat, mussels, trout, mackerel, sole, haddock
Seasonings	salt, pepper, chili, oregano, basil, rosemary, thyme, parsley, cumin, coriander, paprika, mustard, dill, chives, cinnamon, nutmeg, allspice, marjoram, bay_leaf, curry, saffron, tarragon, cardamom, ginger, garlic, horseradish, vanilla, extract
Vegetables	onion, celery, carrot, potato, zucchini, cucumber, lettuce, spinach, broccoli, cabbage, kale, squash, arugula, leek, fennel, asparagus, artichoke, beet, radish, tomato, eggplant, pumpkin
Fruits	apple, banana, berries, grapes, melon, cantaloupe, lemon, lime, orange, peach, pineapple, pear, mango, strawberry, kiwi, watermelon
Nuts and Seeds	peanuts, almonds, cashews, walnuts, pecans, hazelnuts, macadamias, sesame
Oils and Fats	oil, butter, margarine, shortening, lard, ghee
Dairy Products	milk, cream, yogurt, buttermilk, cheese
Flour and Grains	flour, cornmeal, oats, quinoa, barley, wheat, rice
Beans and Soy Products	lentils, chickpeas, kidney_bean, soybeans, tofu, edamame, peas
Sweeteners and Sugars	sugar, honey, molasses, syrup, stevia, fructose
Baking Ingredients	baking_soda, baking_powder, yeast, vanilla, cocoa, gelatin, cornstarch
Sauces and Condiments	ketchup, mayonnaise, soy_sauce, worcestershire_sauce, teriyaki_sauce, barbecue_sauce, salad_dressing, vinaigrette, gravy, mustard, hot_sauce, ranch_dressing, marinara_sauce, pesto_sauce, tartar_sauce
Drinks	tea, coffee, wine, beer, lemonade, milk, brandy, rum, vodka, gin, cider, cola
Alcoholic Beverages	brandy, rum, vodka, gin, liqueur, champagne, vermouth, tequila
Processed and Seasoned Foods	bacon, sausage, hot_dog, ham, salami, prosciutto
Processed Grain Foods	bread, cracker, chips, pie_crust, pancake, waffle, biscuit
Dried Seafood	nori, kelp, bonito_flakes, anchovy
Nut Candies and Pastries	peanuts, almonds, cashews, pecans, macadamias, hazelnuts, walnuts
Baking and Desserts	cake, cookie, brownie, muffin, pudding, pancake, waffle
Dry Goods and Grains	rice, oats, quinoa, barley, bulgur, millet, couscous
Broth and Seasoning Liquids	broth, stock, bouillon, gravy, miso
Tea and Coffee Beverages	coffee, espresso
Jams and Preserves	jam, jelly, marmalade, preserves
Candies and Sweets	candy, chocolate, fudge, caramel, marshmallow
Berries	strawberries, blueberries, raspberries, blackberries, cranberries
Tropical Fruits	mango, papaya, pineapple, banana, coconut
Citrus Fruits	orange, lemon, lime, grapefruit, tangerine
Leafy Vegetables	spinach, kale, lettuce, arugula
Root Vegetables	potato, carrot, beet, radish, turnip
Spices and Seasoning Powders	cumin, coriander, turmeric, paprika
Mushrooms and Fungi	mushroom, truffle, morel
Alcoholic Beverages	wine, beer, brandy, vodka, rum, gin, tequila
Legumes and Soy Products	kidney_bean, chickpeas, lentils, soybeans, tofu
Soy Sauce and Asian Condiments	soy_sauce, teriyaki_sauce, hoisin_sauce
Honey and Syrups	honey, molasses
Concentrated Sauces	ketchup, mustard, mayonnaise, barbecue_sauce, salad_dressing
Breads and Baked Goods	bread, baguette, bagel, muffin
Pasta Sauces	pesto_sauce, marinara_sauce, alfredo_sauce, bolognese_sauce
Grains and Cereals	rice, oats, quinoa, barley, bulgur

Table 3. Broader categories and their respective ingredients.