MIRecipe: A Recipe Dataset for Stage-Aware Recognition of Changes in Appearance of Ingredients

Yixin Zhang[1], Yoko Yamakata[2] and Keishi Tajima[1]
[1]Kyoto University and [2]The University of Tokyo, Japan

RESEARCH BACKGROUND

- In recent years, user-submitted recipe sites have become popular.
- Recipes with text-image paired instructions.
- Recipe website: Haodou

source: https://www.haodou.com/recipe/1190778

RESEARCH PROBLEM

Example: when user want to figure out one specific instructional cooking step:

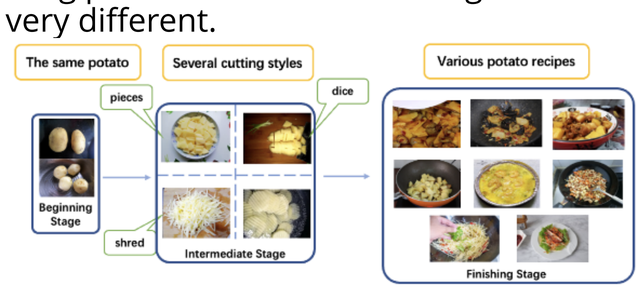
- **Foodstuffs are omitted** in some instructional steps.
  - difficult to understand.
- However, those omitted entities in text descriptions are sometimes shown in the **attached images**.

If we want to **supplement food in text**, we need to **recognize food** in instructional images.

- Therefore, we need to **recognize food** in instructional images.

- Basic image recognition method:
  - Good for entities like tools → applying to tool recognition
  - Difficult for ingredient (ingredient shapes are changing as being cooked) → need adjusting

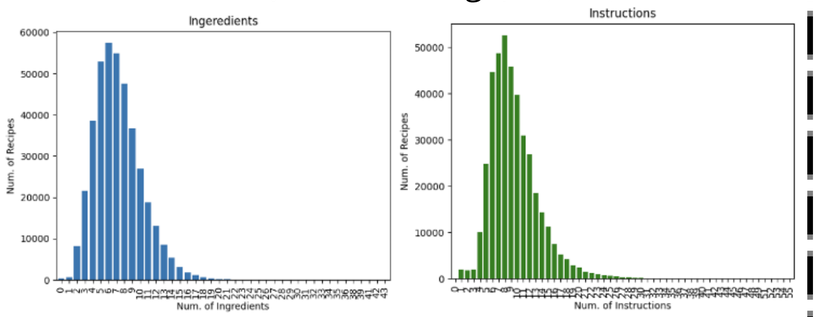
- Images located in different positions in the cooking procedure of the same ingredient may be very different.



DATASET

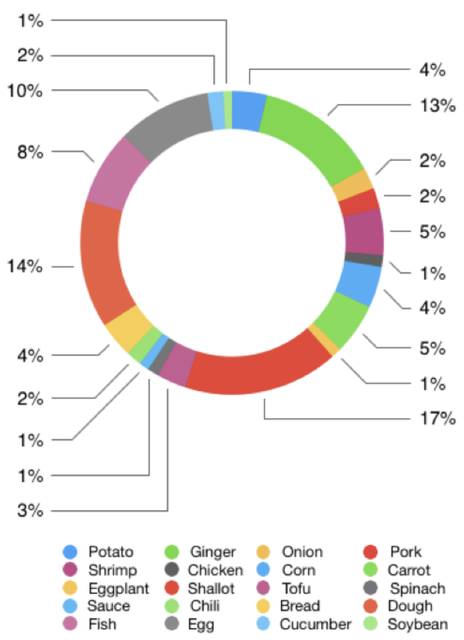
Recipe Dataset Statistics (up to now)

- #Recipes: 398,597
- #Ingredient Classes: 35,319
- #Instructions (text and image): 3,745,544



- #Ingredient classes used in this experiments: 20
  - high frequencies of occurrence
  - #Images of 20 classes: 35,401

Potato, ginger, onion, pork, shrimp, chicken, corn, carrot, eggplant, shallot, tofu, spinach, sauce, chili, bread, dough, fish, egg, cucumber, soybean



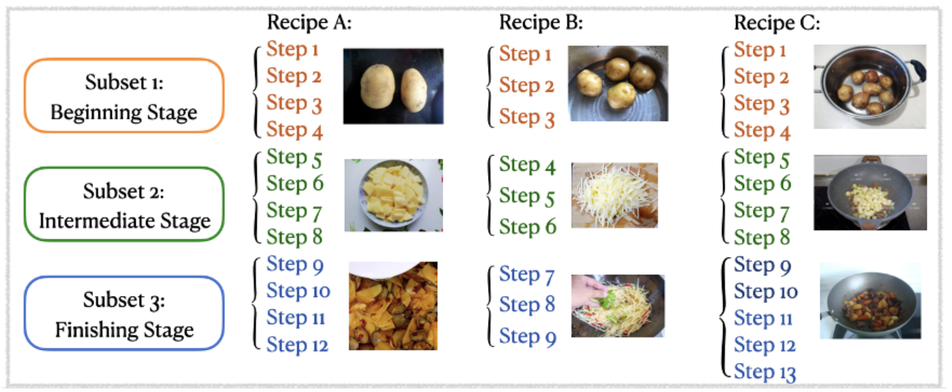
STAGE-AWARE INGREDIENT RECOGNITION METHOD

Image Classification

- We compute the relative position of the steps in the whole recipe
  - E.g., Step No.4 out of 15 steps: 0.267

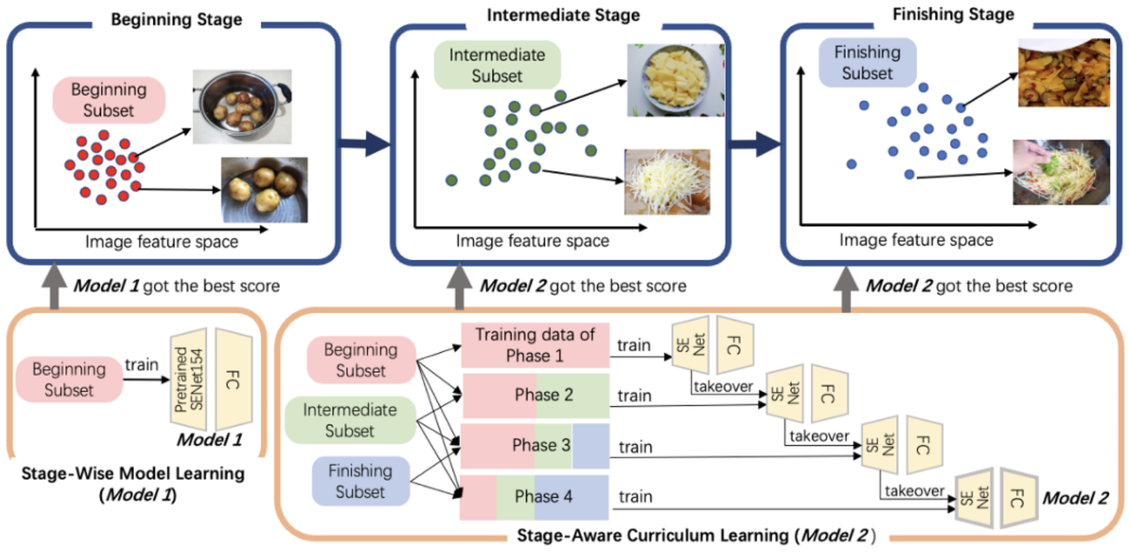
$$RelativePosition = \frac{StepNum.}{TotalStepNum.}$$

- Images are divided into 3 subsets according to their relative positions in recipes.



Stage-Aware Ingredient Recognition

- Stage-Aware Recipe Image Recognition For Food Changing in Appearance
- Overview of the baseline method and proposed methods:



EXPERIMENT AND EVALUATION

Table 2: Accuracy of Stage-wise Model Learning

Training \ Test Subset	Beginning	Intermediate	Finishing	All
Beginning	**64.16%**	54.72%	40.63%	49.91%
Intermediate	55.63%	**60.59%**	49.51%	47.33%
Finishing	42.79%	50.66%	**52.83%**	47.01%
All	50.17%	51.74%	46.28%	49.63%

Table 3: Accuracy of Stage-aware Curriculum Learning

Training Pattern\ Test Subset	Beginning	Intermediate	Finishing
Pattern 1	**64.16%**	54.72%	40.63%
Pattern 2	61.79%	58.84%	55.31%
Pattern 3	60.10%	**62.61%**	**58.34%**

Table 4: Comparison of the Proposed and Baseline Methods

Plan		Top-1 acc.
Baseline (SENet154)		49.63%
Stage-Wise Model Learning ($m(i) = i$)	Beginning Subset	**64.16%**
	Intermediate Subset	60.59%
	Finishing Subset	52.83%
	Average	59.19%
Curriculum Learning Pattern 3 (Model 2)	Beginning Subset	60.10%
	Intermediate Subset	**62.61%**
	Finishing Subset	**58.34%**
	Average	60.35%

Table 5: Comparison of Our Methods Based on SENet154 with Other Standard Methods

Plan		Top-1 acc.	Top-3 acc.	Top-5 acc.
Ours	Stage-Wise	59.19%	81.21%	89.47%
	Model 2	60.35%	83.76%	90.91%
Baseline	SENet154	49.63%	76.51%	86.93%
	Resnet50	46.41%	74.13%	84.35%
	VGG16	43.67%	72.39%	85.01%
	AlexNet	31.77%	64.06%	77.42%

Table 6: Final Accuracy of Our method

	Model Selection	Accuracy
Beginning Subset	Model 1	64.16%
Intermediate Subset	Model 2	62.61%
Finishing Subset	Model 2	58.34%
Average	61.70%	

CONCLUSION

- **Contributions:**
  1. We construct a recipe dataset which **contains both instructional text and image data.**
  2. We develop a recognition method for **ingredients whose appearance changes** with the cooking progress

- Future work
  - Since We only focused on the single-label recognition in this work, experiments with multi-label data is also an important remaining issue for future work.