ACM Multimedia Asia 2024

Adaptive Feature Inheritance and Thresholding for Ingredient Recognition in Multimedia Cooking Instructions

Yixin Zhang[1], Yoko Yamakata[2] and Keishi Tajima[1]

[1]Kyoto University, [2]The University of Tokyo

Research Background

- Recipe as **Multimedia Instructions**

  - Typical multimedia instructional data
  - Have much content on the Internet.
  - Step-by-step instruction format
  - Rich information for multi-modal analysis.

Source: https://cookpad.com/jp/recipes/24255562

Challenges and Problems

- Recognition difficulty

  - The appearance of ingredients are changing during cooking process

Challenges and Problems

- Recognition confidence varies by recipe context

  - Some ingredients are harder to recognize
    - Main ingredients: Easy to recognize
      - Clear visual features
      - Abundant training data
    - Seasonings: Hard to recognize
      - Subtle visual presence
      - Limited training examples

  - The same ingredient might be more visible in one recipe than in another


Easy   vs   Hard

vs

Potatoes in different recipes

Limitations of Traditional Methods

- **Existed Multi-label Recognition Method are limited**
  - Mainly focus on appearance-static objects
  - Mainly look at the final dish image, or analyze text information
  - Can not handle appearance changing of ingredients during cooking

- **Fixed Threshold Issues**
  - Standard threshold (e.g., 0.4) too high
  - Misses subtle ingredients
  - Can not handle variety in recipes

Fixed Thresholds θ on different ingredients

Cannot detect Ingre A
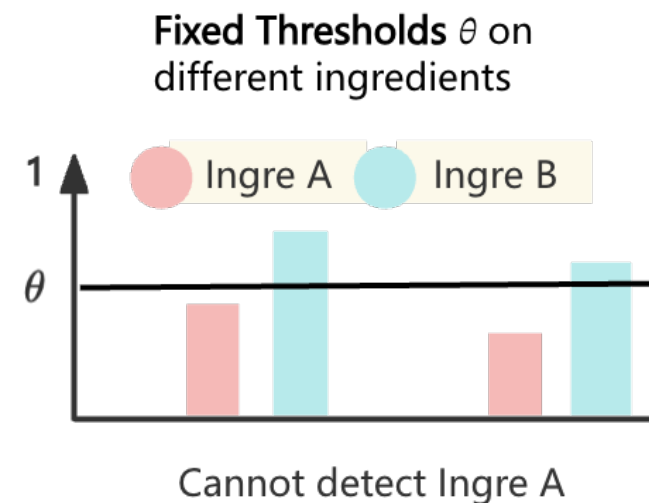
Key Observations

- Recognition Patterns
  - Higher confidence in early steps
  - Decreasing confidence over time
  - Since appearance of ingredient are changing

Easy, higher confidence

Hard, lower confidence

- Ingredient Inheritance
  - Ingredients flow between steps
  - Context helps recognition

Ingredient candidates are the same, and inheritance relationships exist among steps

Key Observations

- Natural predefined label candidate set
  - Since each recipe has **its own ingredient list**
  - **Much smaller** than full dataset labels
  - Smaller recognition space

Title

Cooking Instructions

Ingredient List

Source: https://cookpad.com/jp/recipes/24255562

Our Approach

- Contextual Feature Inheritance
  - Why Learn from Previous Steps?
    - Ingredients change appearance
    - Earlier steps: clearer visual features
    - Maintain inheritance relationships

  - Feature Computation
    - Weight based on step similarity
    - Adaptive feature inheritance
    - Enhance recognition with context
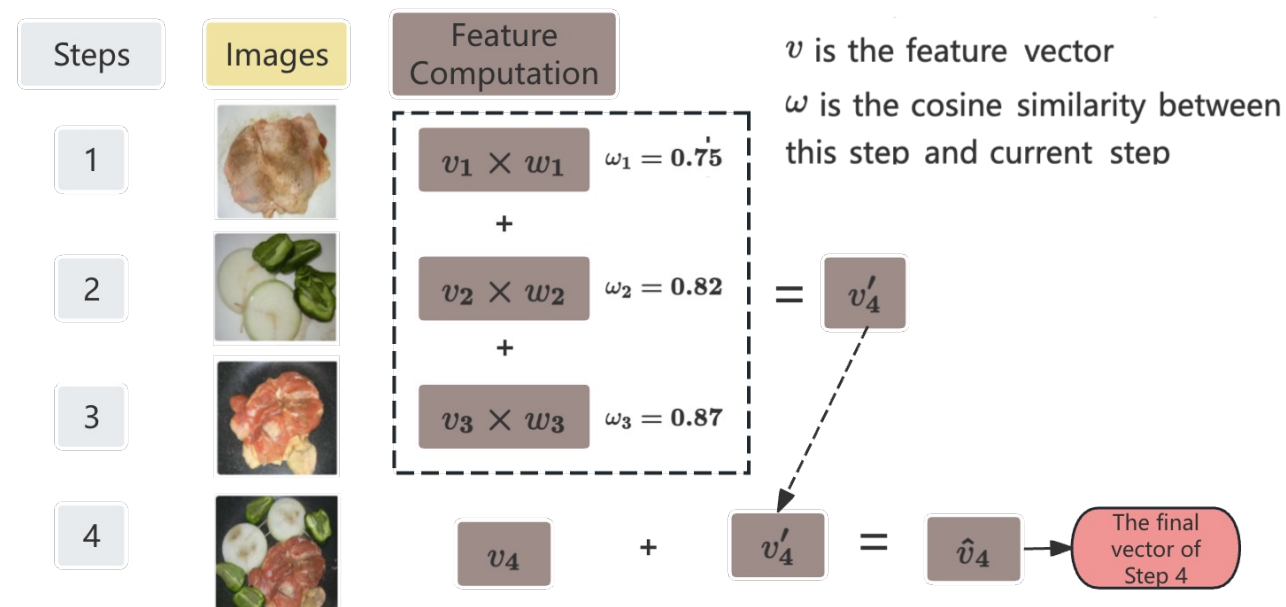
Our Approach

- Adaptive Thresholding

  - Final step contains all ingredients
  →Ingredient Estimation at the Final Step.
  → Recipe-specific thresholds

  - Recognition difficulty varies by recipe
  - Each ingredient needs different threshold
  →Handle varied recognition difficulty
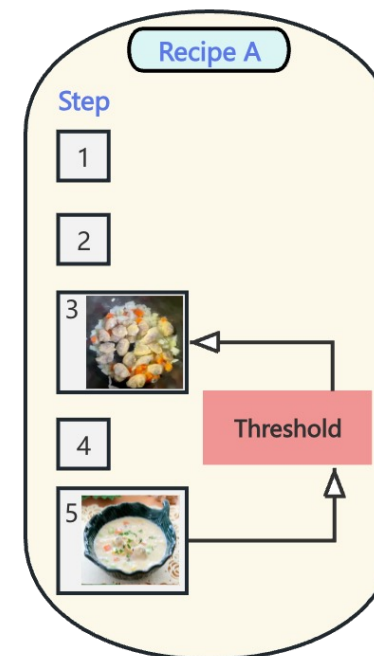
Different Thresholds θ on different ingredients

Performance Results

- Comparison with Baselines

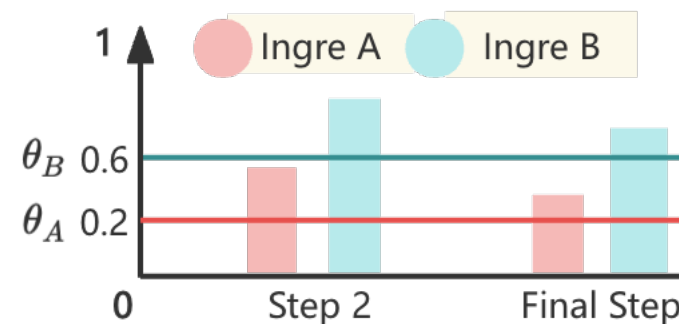| Method	Data	Precision	Recall	F1
Baseline	Image-only	0.4653	0.5536	0.5056
D-Mixup	ImageText	0.5455	0.5714	0.5581
Curriculum Learning	ImageText	0.5781	0.6305	0.6031
Ours	Image-Text +Context	0.6293	0.7384	0.6792

- Data Types:
    - image-only: Using only image data
    - ImgTxt: Using both image and text data
    - ImgTxt-C: Using image, text, and contextual information

Performance Results

- Ablation Studies

Method	Precision	Recall	F1
CLIP	0.5970	0.7219	0.6535
CLIP + Context (Ours)	0.6293	0.7384	0.6792

Threshold Type	Precision	Recall	F1
Fixed (0.15)	0.5501	0.5871	0.5679
Fixed (0.25)	0.5122	0.5315	0.5216
Adaptive	0.6293	0.7384	0.6792

Inheritance + context architecture provides the best performance

- Fixed thresholds struggle with varying ingredient visibility
- Our adaptive approach significantly improves recognition accuracy

Conclusion

- An ingredient recognition approach for cooking processes
- Feature inheritance mechanism for handling parallel tasks
- Adaptive thresholding for ingredient-specific recognition
- Large-scale multimedia recipe dataset

- Cooking Domain Applications
  - Smart kitchen assistance
  - Dietary tracking systems
  - Cooking education platforms

- Beyond Cooking
  - Movie actor detecting
  - Manufacturing process monitoring
  - Chemical reaction tracking
  - …

Thanks for listening!