Cross-Modal Recipe Embeddings by Disentangling Recipe Contents and Dish Styles

Yu Sugiyama and Keiji Yanai

The University of Electro-Communications, Tokyo

Introduction

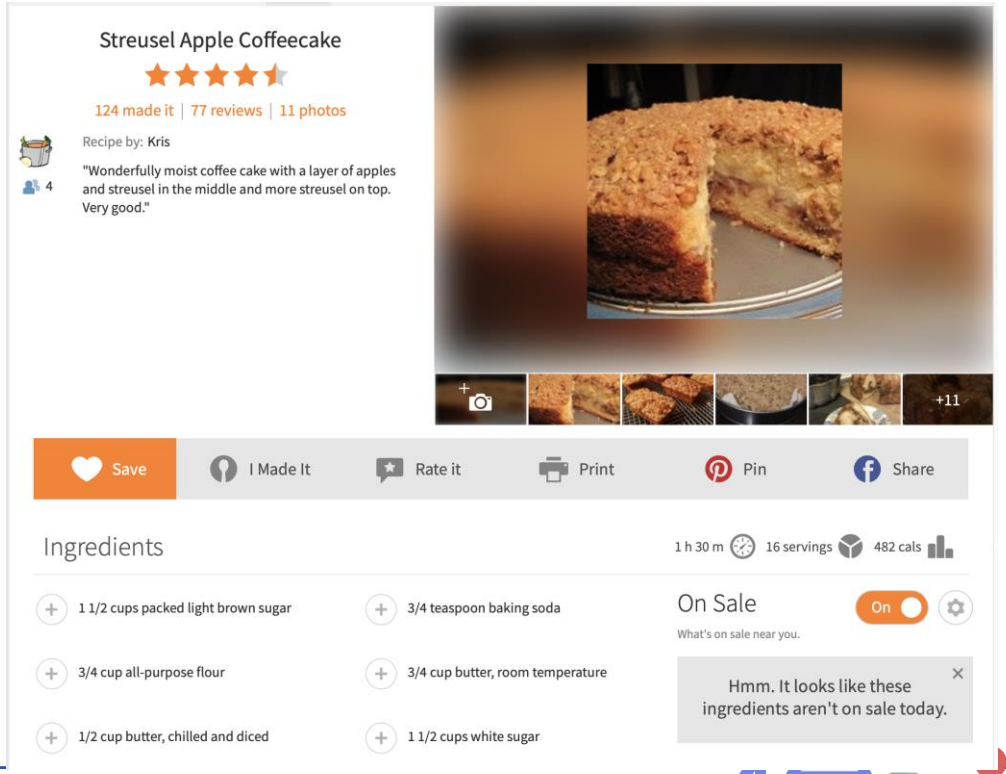
- Cooking recipe sharing site are widely used
- Recipe data has Text and Images
- We propose a RDE-GAN(Recipe Disentangled Embedding GAN)

allrecipes

https://www.allrecipes.com

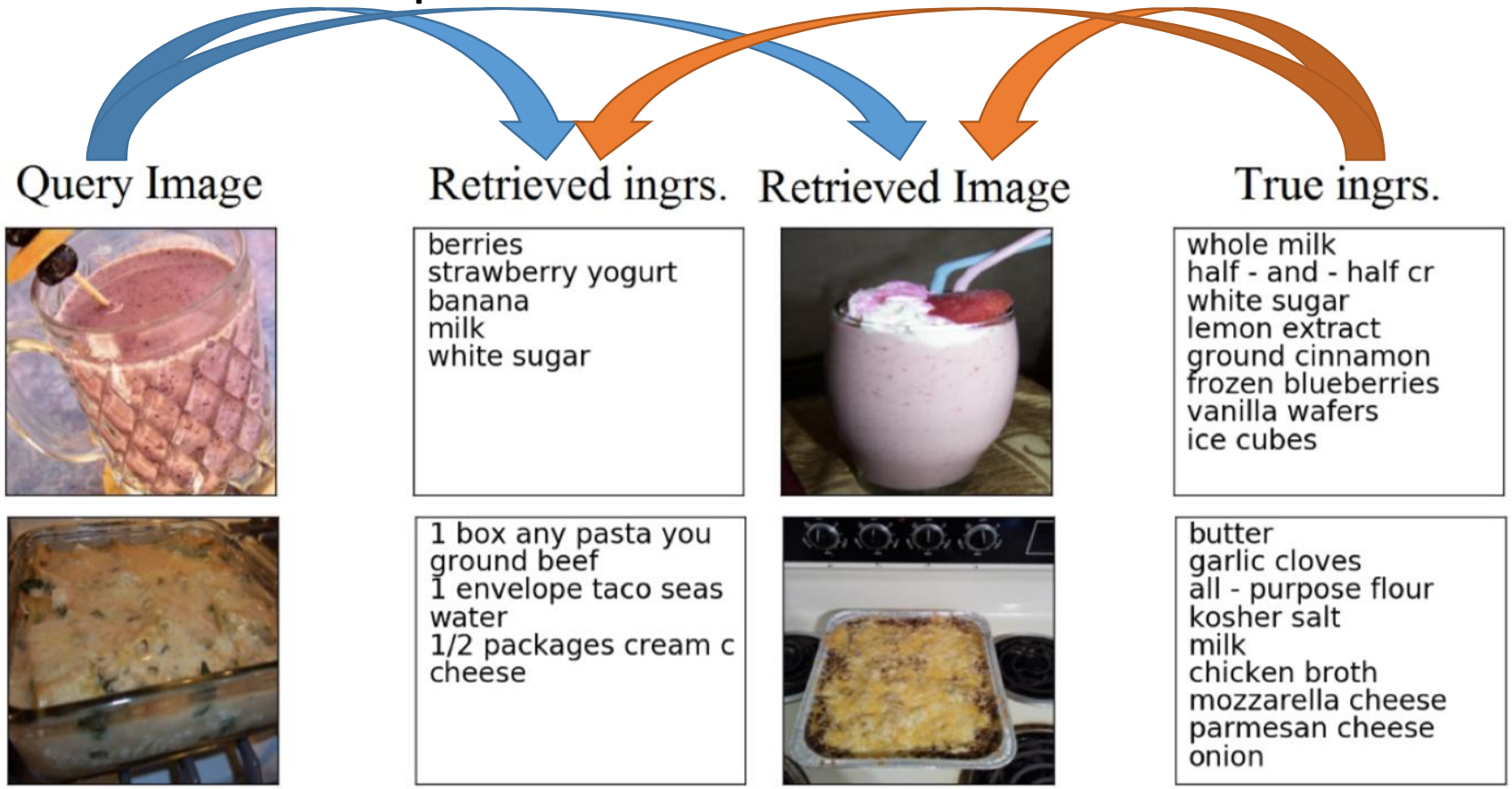
cookpad

https://cookpad.com



Background

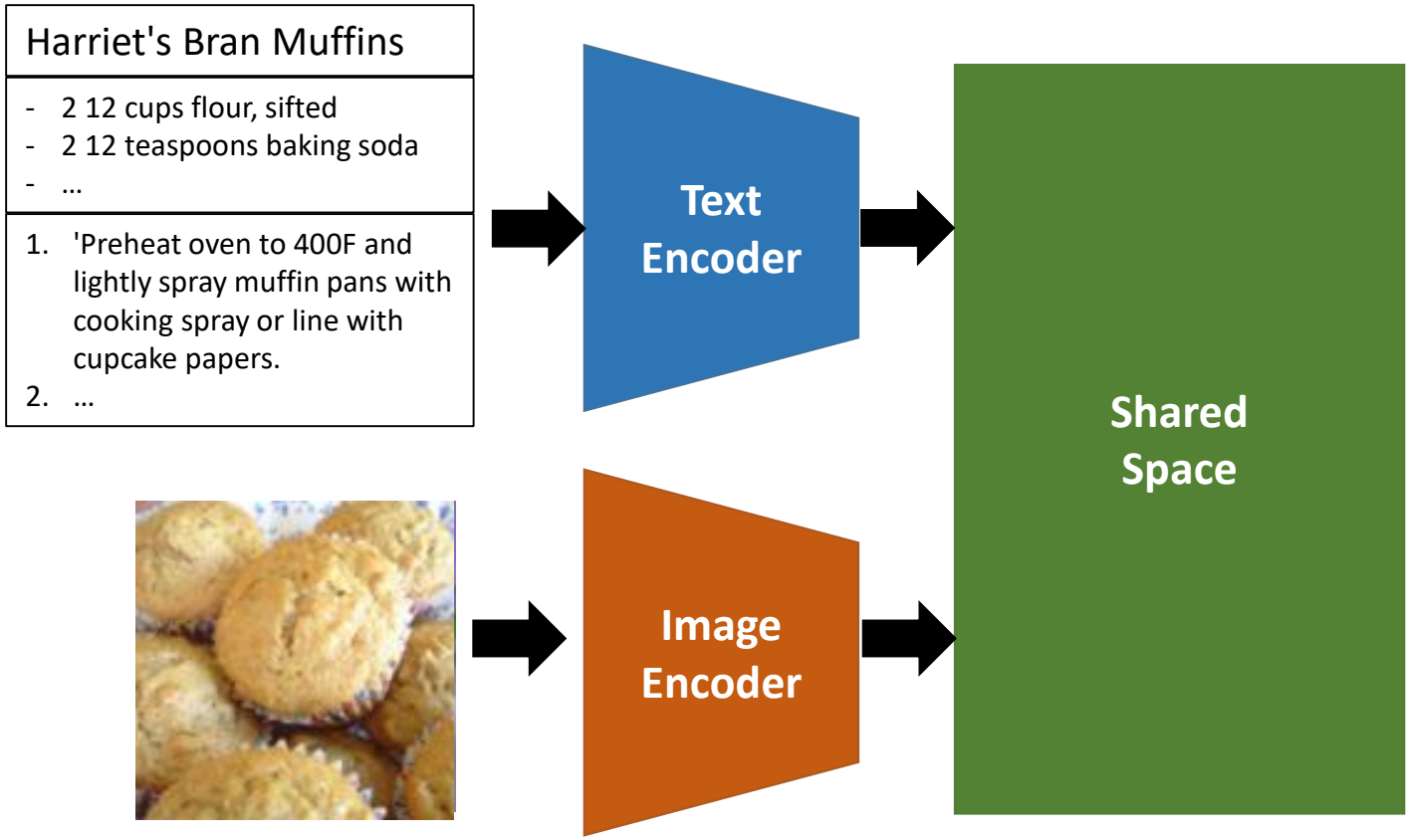
- After Recipe 1M+, text and image cross-modal learning becomes hot topic



Recipe1M+[Salvador, CVPR2017]

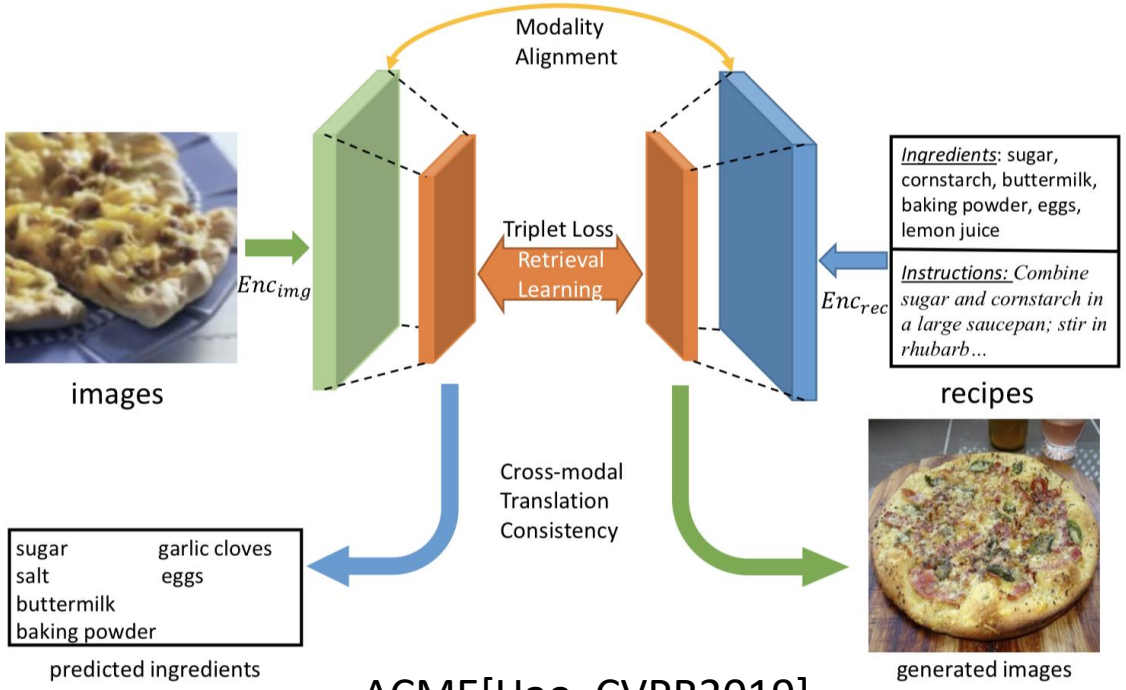
Background: Multi-modal retrieval

- Embed text and image into shared space
- Retrieval by distance in shared space



Related Work

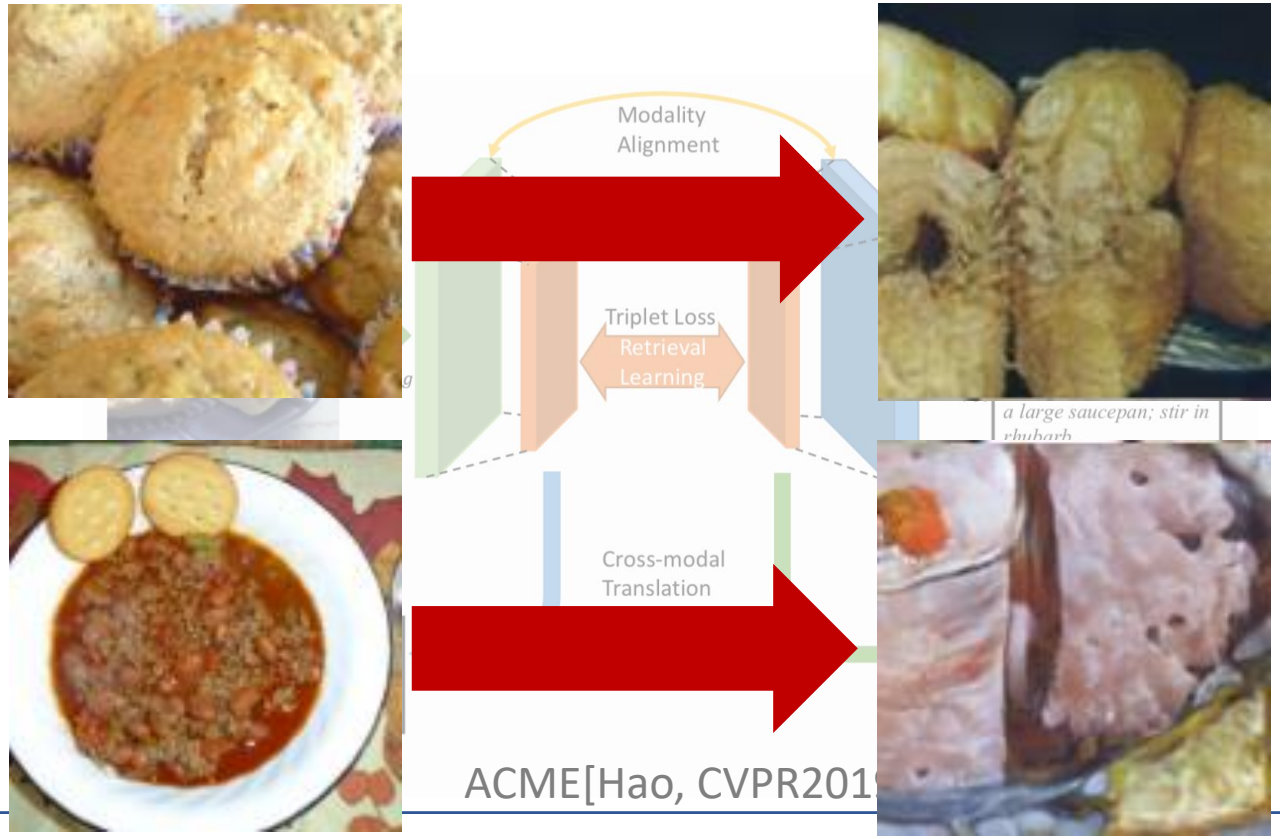
- ACME[Hao, CVPR2019]
  - *Adversarial Cross-Modal Embedding*
  - Image generator improved cross-modal retrieval



ACME[Hao, CVPR2019]

Related Work

- ACME[Hao, CVPR2019]
  - *Adversarial Cross-Modal Embedding*
  - Image generator improved cross-modal retrieval



ACME[Hao, CVPR201

Proposed Method: Overview

- Recipe images have…
  - recipe information: useful for recipe retrieval
  - non-recipe information: non-useful for recipe retrieval

recipe information

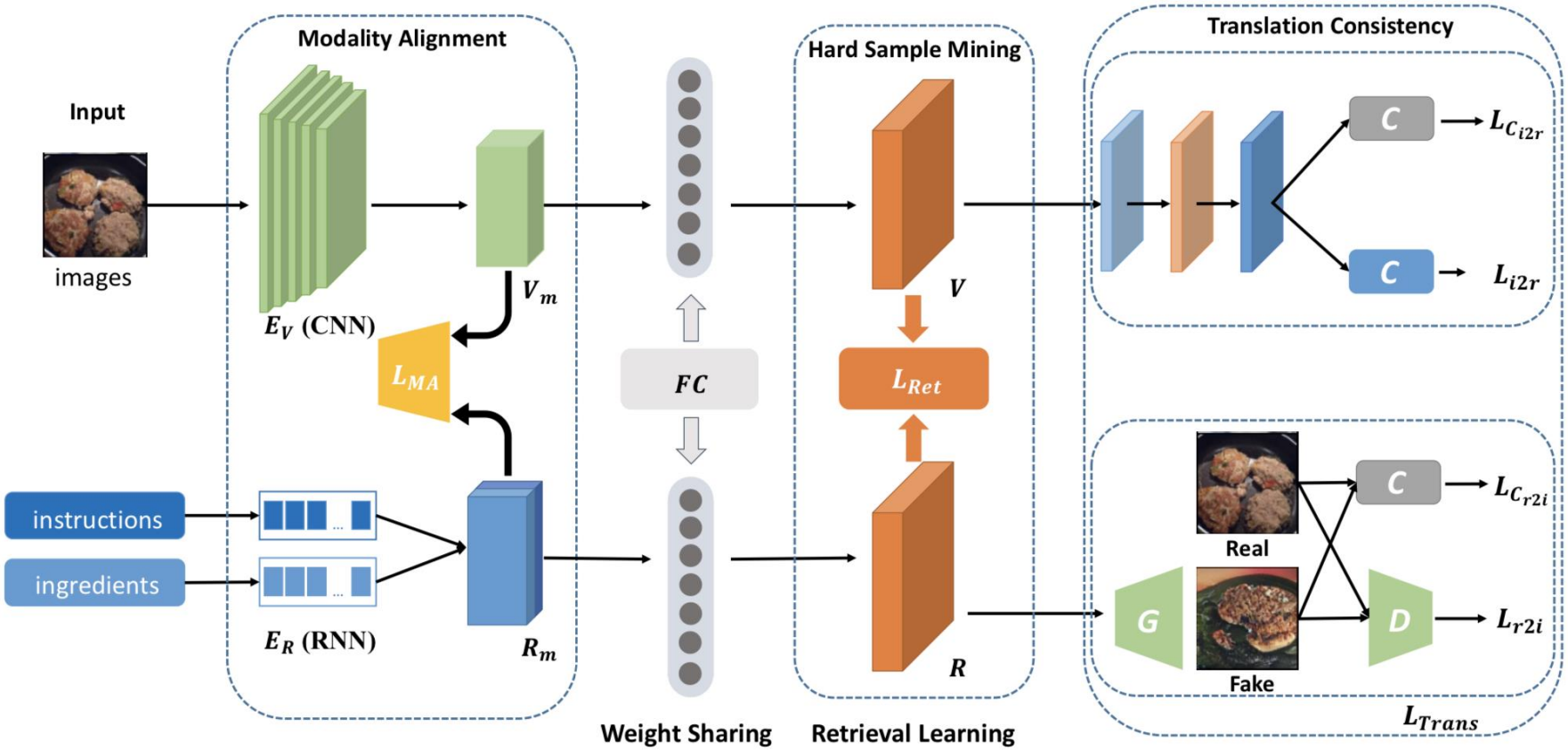
- chili
- soup
- tomato base
- stew
- …

non-recipe information

- round plate
- white plate
- centered picture
- photo in bright location
- …

Proposed Method: Overview

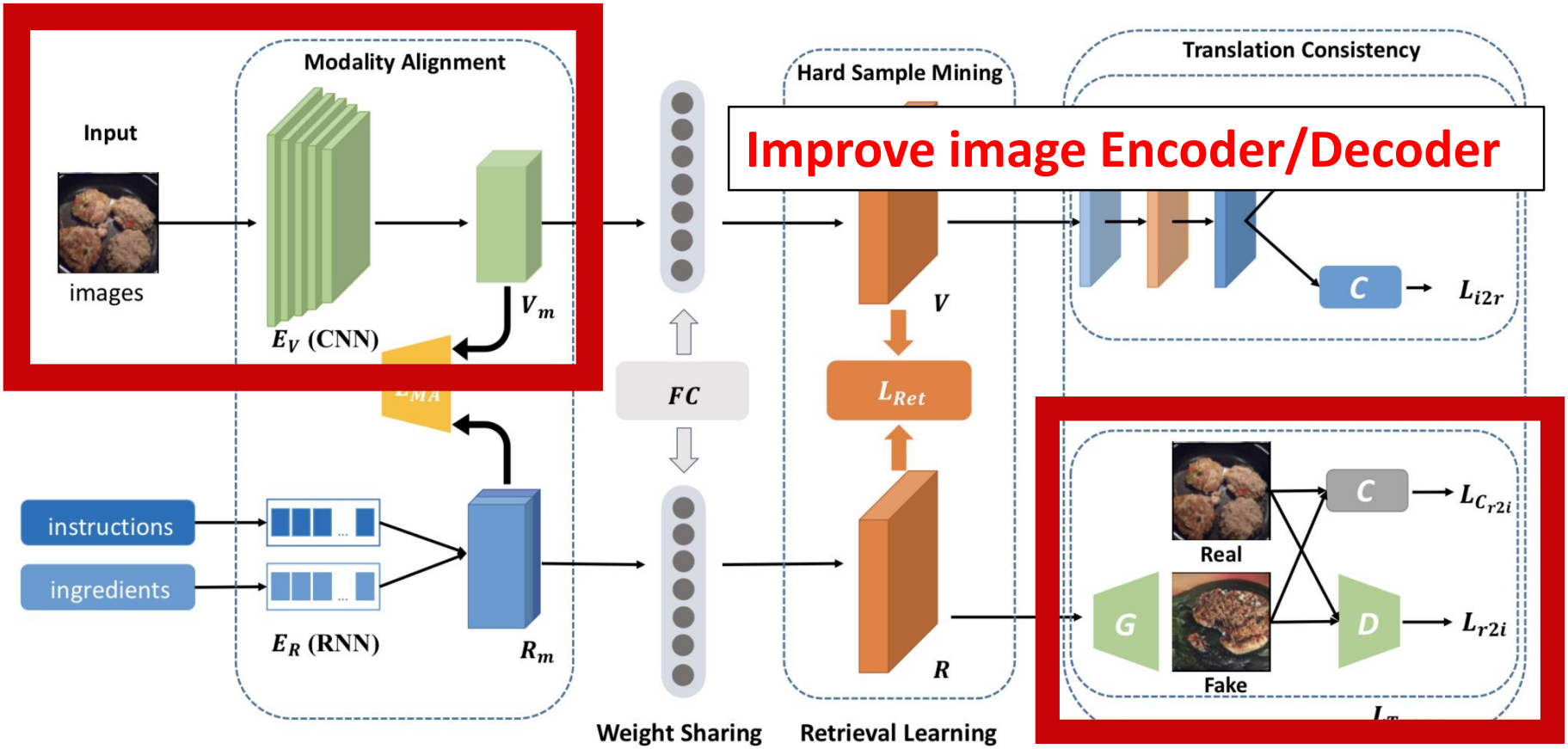
- Improve ACME, disentangle image features



ACME[Hao, CVPR2019]

Proposed Method: Overview

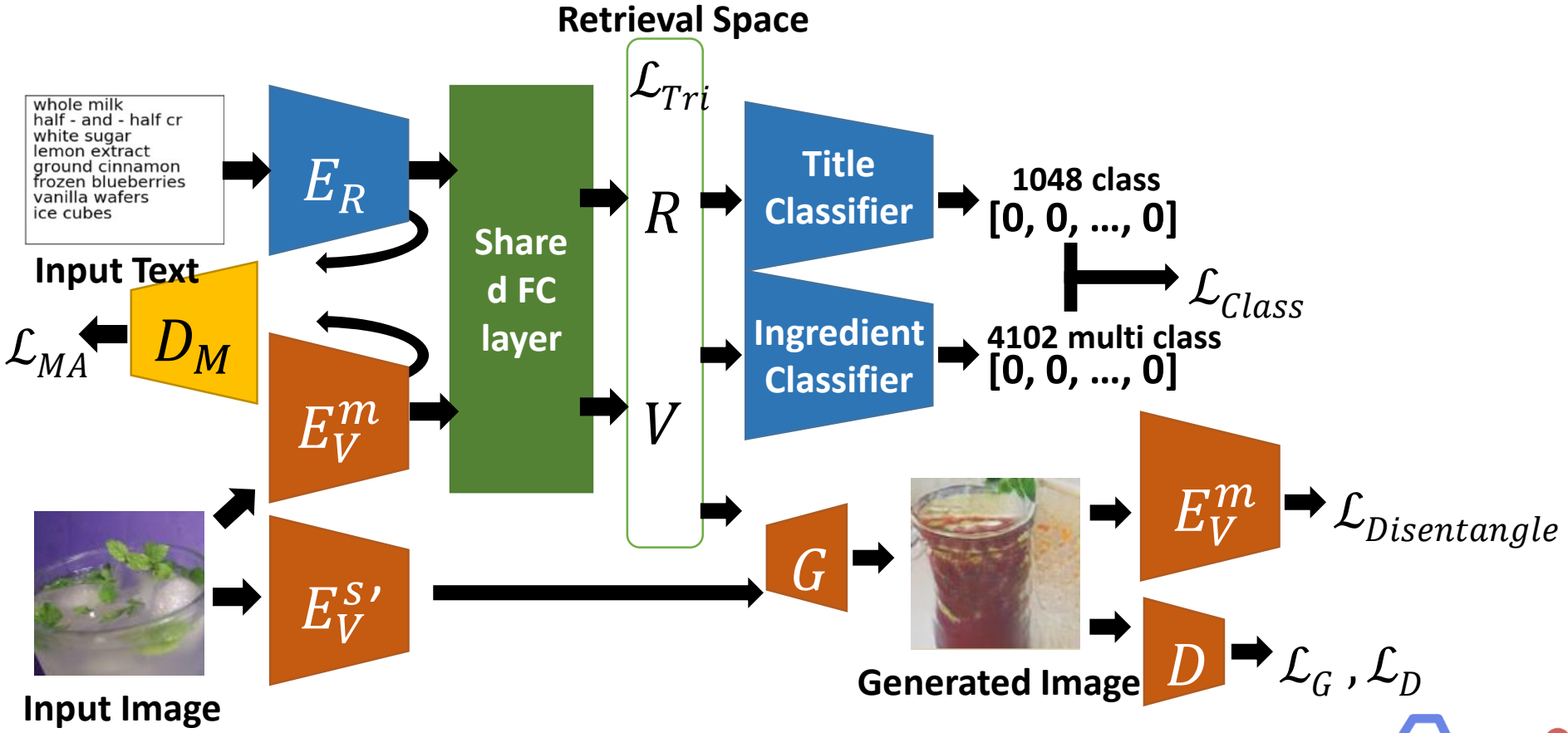
- Improve ACME, disentangle image features



ACME[Hao, CVPR2019]

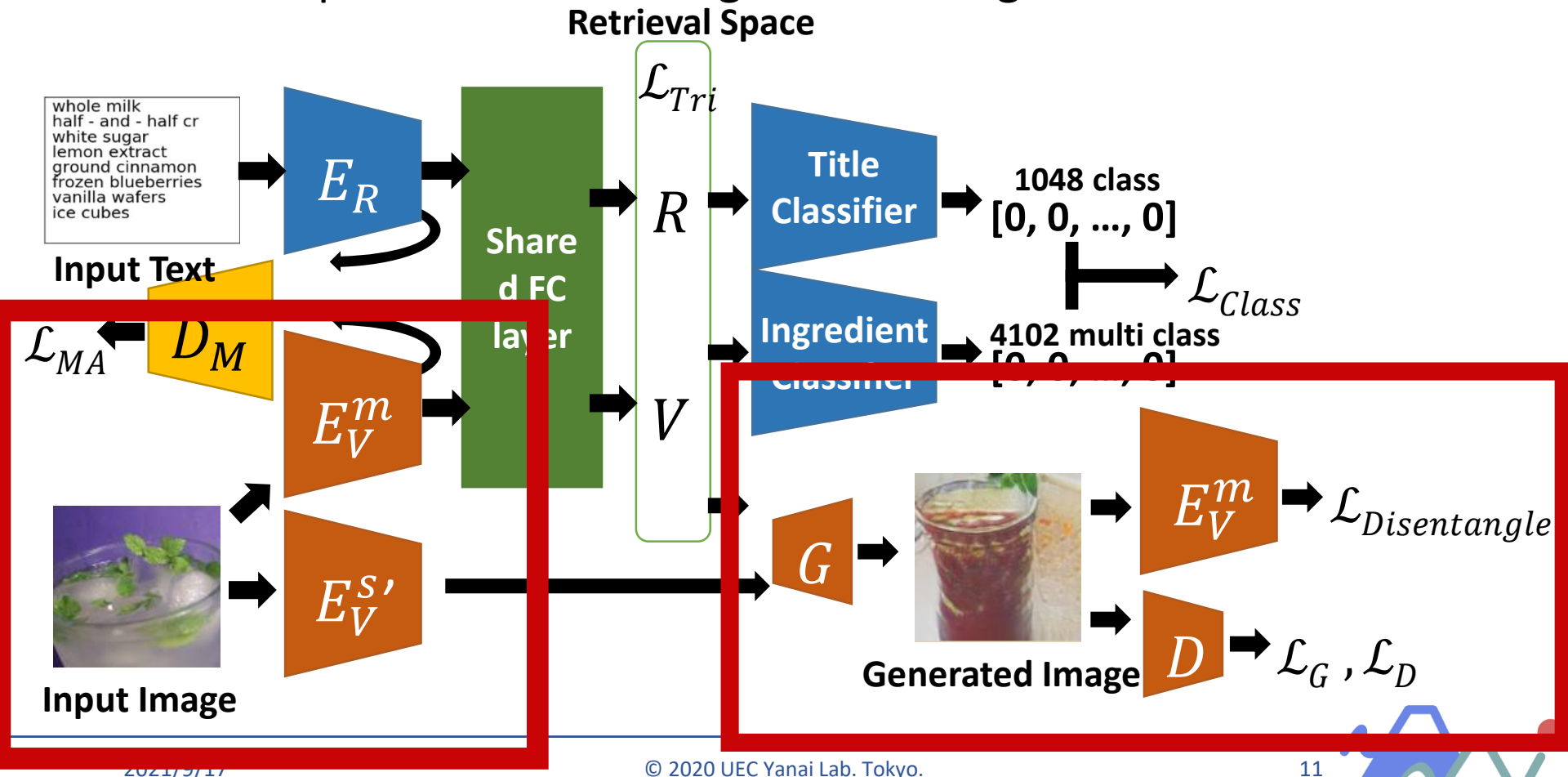
Proposed Method: RDE-GAN

- RDE-GAN separates image information
  - recipe information for learning retrieval
  - non-recipe information for generate image



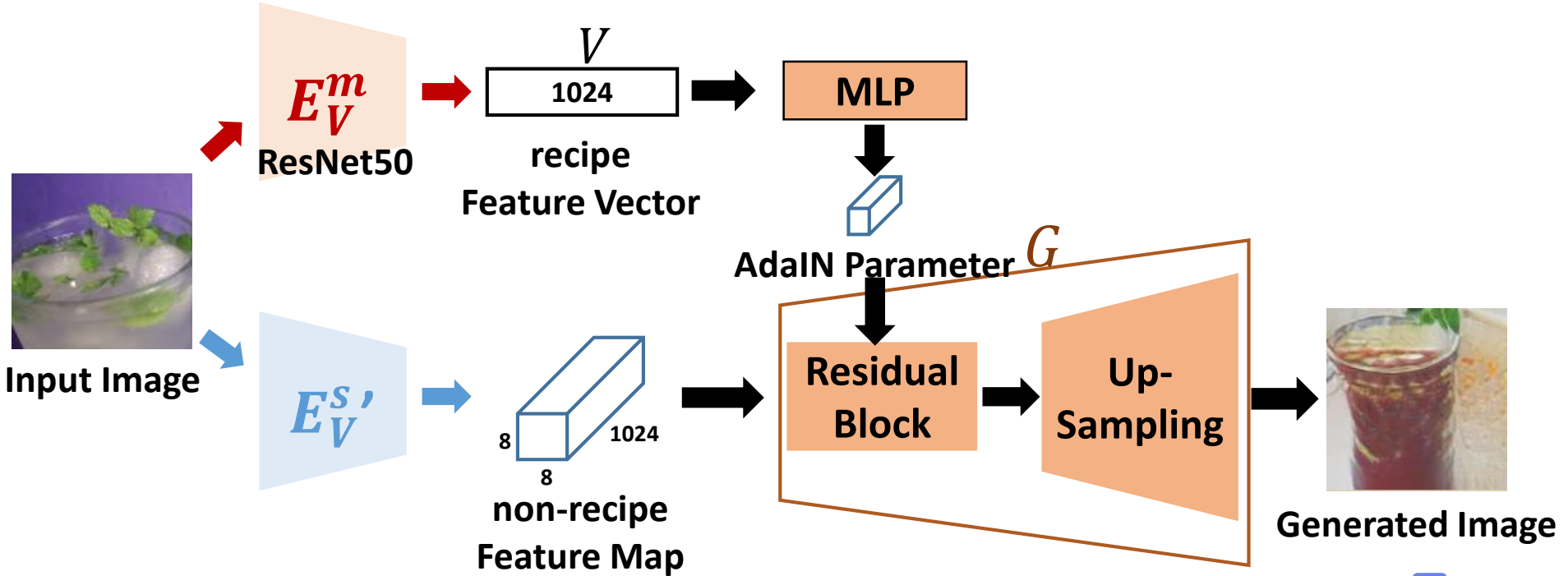
Proposed Method: RDE-GAN

- RDE-GAN separates image information
  - recipe information for learning retrieval
  - non-recipe information for generate image



Proposed Method – Encode Images

- Use different network for recipe feature and non-recipe feature
- Shallow network for non-recipe feature
- Deep network for recipe feature



Proposed Method: RDE-GAN

We propose **RDEGAN**(Recipe Disentangled Embedding GAN)

RDE-GAN disentangles recipe feature and non-recipe feature

1. Use only recipe feature for retrieval

2. Use recipe and non-recipe feature for image generate

Experiments – Recipe Retrieval Evaluation

- We test with median rank(MedR) and recalls
- Ours performed state-of-the-art with large test set

Num		Image2Recipe			Recipe2Image		
	method	MedR↓	R@1↑	R@10↑	MedR↓	R@1↑	R@10↑
1k	JE	5.2	25.6	65.0	5.1	25.0	65.0
	ACME	**1.0**	51.8	**87.5**	**1.0**	56.3	85.9
	Ours	**1.0**	**59.4**	87.4	**1.0**	**61.2**	**87.2**
10k	JE	41.9	-	-	-	-	-
	ACME	6.7	22.9	57.9	6.0	24.4	59.0
	Ours	**3.5**	**36.0**	**64.4**	**3.0**	**38.2**	**65.8**

Experiments – Retrieval Example

Query Text	#1	#2	#3
Ainsley Harriott's baked mussels recipe - 350 g (12.3oz) Cherry tomatoes - 1 Red onion, finely chopped - … 1. Pre-heat the oven 200 degrees C/400 degrees F gas mark 6. 2. Place the cherry tomatoes in a large roasting tin and scatter over the onions. 3. …			
I Cant Believe Its Cutty: Victory Is Mine Cocktail - 2 ounces Cutty Sark Prohibition Edition - 3/4 ounce Marie Brizard Apry - … 1. Combine liquid ingredients in one small shaker tin, and only the egg white (no ice) in another. 2. Shake vigorously to froth egg whites and other ingredients. 3. …			

Experiments – Retrieval Example

Query Image **#1** **#2**

Chicken Mushroom Barley Soup

- 4 garlic cloves
- 1 leek
- …

1. Mince garlic; slice leek longways then chop.
2. Place a standard dutch oven(or soup pot if you don't have a dutch " 'oven) on the stove top.
3. …

Tonjiru (Miso Soup with Veggies and Pork)

- 100 g pork
- 1/2 carrot
- …

1. Chop all vegetables and pork into small pieces.
2. (Since I have a small child, I cut them into a child-sized pieces, ' 'but you can cut as you wish.)
3. …

Ultimate Green Beans

- 2 slices bacon, diced
- 1/2 white onion, minced
- …

1. Cook the bacon in a large, deep skillet over medium-high heat until crisp, about 10 minutes.
2. Remove the bacon with a slotted spoon and drain on a paper towel-lined plate
3. …

Herbed Green Beans and Mushrooms

- 2 lbs fresh green beans, trimmed
- 3 cups water
- …

1. Bring salted water to boil.
2. Add trimmed green beans cook until tender but still crisp.
3. …

Experiments – Retrievals

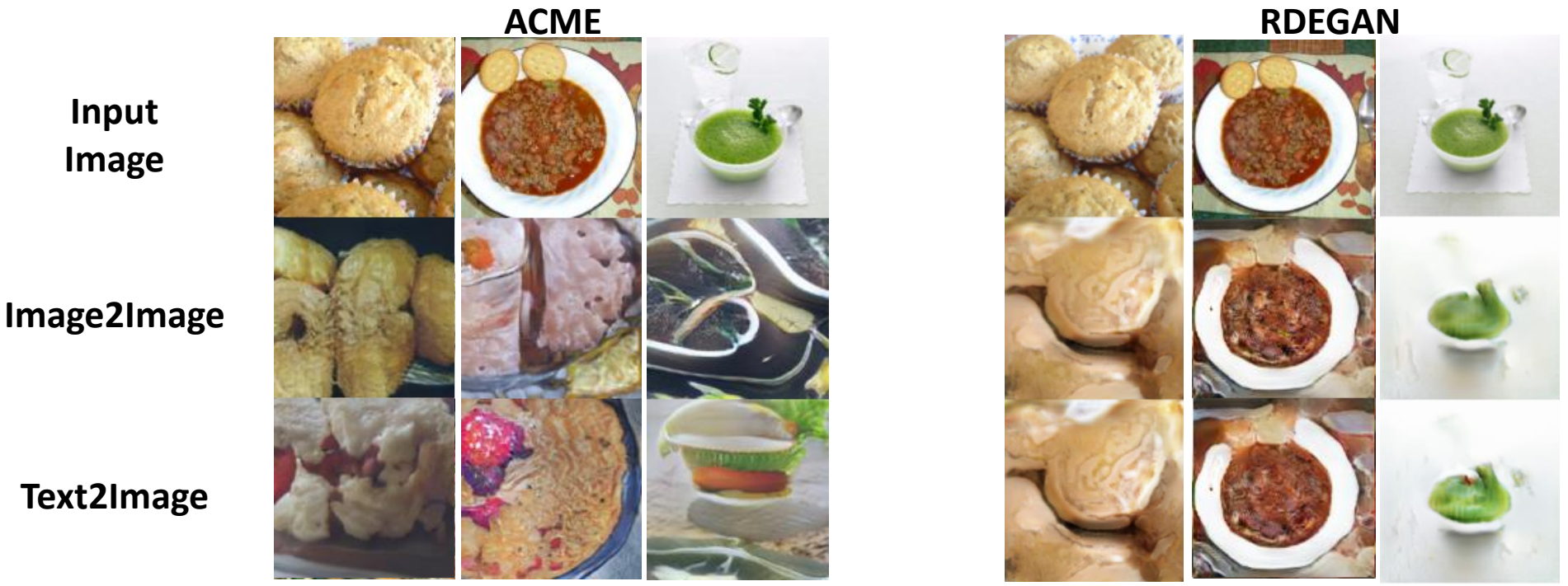
- RDE-GAN performed state-of-the-art retrieval score

- The retrievals are improved with disentangle image features
  - It is important separate image feature into recipe feature and non-recipe feature

- Recipe feature denotes feature only for retrieval
  - Improve the retrieval

Experiments – Image Generation

- We compare generated image
  - ACME
  - RDEGAN(ours)
- Ours generates better images



Experiments – Image Generation

- Evaluate generated images with FID
  - Lower score is better

- Our method performed better score in FID

Method	FID↓@ Image2Image	FID↓@ Text2Image
ACME	183.8	182.9
RDEGAN(ours)	**158.9**	**158.6**

Experiments – Gradual Change

- 3 experiments for test encoded space

- Test1: Change only non-recipe feature with Image A to Image B

- Test2: Change only recipe feature with Image A to Image B

- Test3: Change only recipe feature with Text A to Text B

Experiments – Gradual Change

- Test1: Change only non-recipe feature with Image A to Image B
- Non-recipe feature represents dish shapes, table color, rough shape, etc.

Experiments – Gradual Change

- Test1: Change only non-recipe feature with Image A to Image B
- Non-recipe feature represents dish shapes, table color, rough shape, etc.

Input Image

Non-recipe feature

Target Image

Experiments – Gradual Change

- Test2: Change only recipe feature with Image A to Image B

- Recipe feature represents ingredients color, texture, number of ingredients, etc.

Input Image

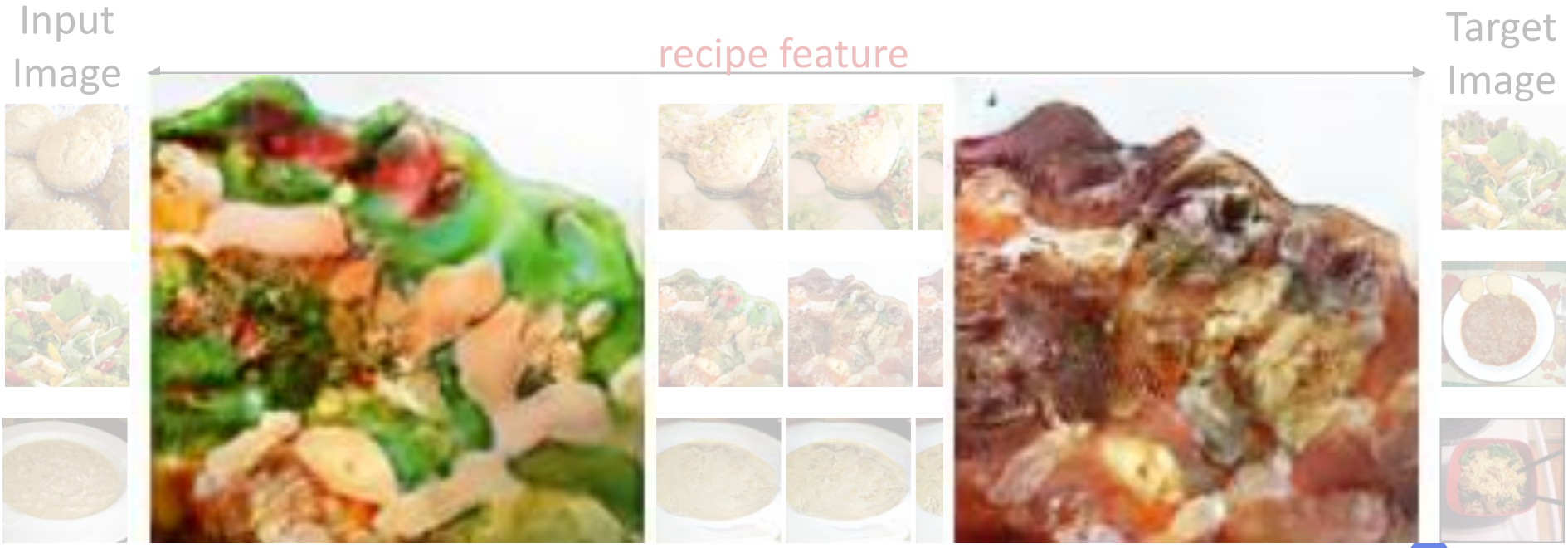
recipe feature

Target Image

Experiments – Gradual Change

- Test2: Change only recipe feature with Image A to Image B

- Recipe feature represents ingredients color, texture, number of ingredients, etc.



Experiments – Gradual Change

- Test3: Change only recipe feature with Text A to Text B
- recipe feature with text represents ingredients, texture, etc.

Harriet's Bran Muffins		Rob and Lisa's Island-Escape Salad
- 2 12 cups flour, sifted - …		- 1/2 of a medium pineapple, cored - …
Rob and Lisa's Island-Escape Salad		**Four Alarm Chili**
- 1/2 of a medium pineapple, cored - …		- 2 lbs ground sirloin - …
Chili Blanco		**Udon-Beef Noodle Bowl**
- 1 tablespoon vegetable oil - …		- 8 ounces uncooked udon noodles - …

Experiments – Gradual Change

- Test3: Change only recipe feature with Text A to Text B
- recipe feature with text represents ingredients, texture, etc.

Harriet's Bran Muffins

- 2 12 cups flour, sifted
- …

Rob and Lisa's Island-Escape Salad

- 1/2 of a medium pineapple, cored
- …

Chili Blanco

- 1 tablespoon vegetable oil
- …

Rob and Lisa's Island-Escape Salad

- 1/2 of a medium pineapple, cored
- …

Four Alarm Chili

- 2 lbs ground sirloin
- …

Udon-Beef Noodle Bowl

- 8 ounces uncooked udon noodles
- …

Experiments – Gradual Change

- Our method generate images better than ACME
  - non-recipe feature improved generator

- recipe feature represents color, texture, etc.
  - Same in Image2Image and Text2Text
- non-recipe feature represents rough shape, dish, etc.

Conclusion

[Conclusion]

- RDE-GAN improved retrieval accuracy
- RDE-GAN can control separately rough shape and ingredients colors

[Future work]

- Improve quality and size of synthesized images
- Apply this method into other domains