IMAGES AND RECIPES
Retrieval in the cooking context

Friday 16th April, 2018

Micael Carvalho*, Rémi Cadène*,

David Picard, Laure Soulier, and Matthieu Cord
Sorbonne Université

An extended version of this work will appear at SIGIR 2018

Why this work?

- Cooking is a fundamental human activity
- Machine learning is evolving fast: data and technique

What do we do?

- Artificial intelligence-oriented approach for cooking
- We find recipes from pictures, and pictures from recipes

Introduction
○●○○○○○○○○

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○

Task 1: Image to Recipe retrieval

Introduction
○○●○○○○○○○

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○

Task 2: Recipe to Image retrieval

Query Recipe

Ingredients	Instructions
butter	1. Heat butter in 2 qt saucepan over low heat until melted
garlic cloves	2. Add garlic.
all - purpose flour	3. Stir in flour and salt.
kosher salt	4. Cook, stirring constantly until bubbly.
milk	5. Remove from heat and stir in milk and broth.
chicken broth	...
mozzarella cheese	6. Cook uncovered at 350F 20-30 minutes until nice and
parmesan cheese	bubbly.
onion	7. Let stand 10 minutes before cutting.
...	

Retrieved Image

Query Recipe

Ingredients	Instructions
dashi stock	1. Transfer dashi to a small soup pot over medium-low
hot water	heat.
miso	2. Meanwhile, stir together hot water and miso until miso
firm tofu	is dissolved.
green onion	3. Pour watery miso mixture into the pot.
	4. Add cubed tofu.
	5. Bring the pot to a simmer.
	6. To serve, sprinkle sliced green onions and a pinch of
	katsuobushi on top.

Retrieved Image

Introduction
○○○●○○○○○

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○

How to tackle these tasks?

Introduction
○○○○●○○○○

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○
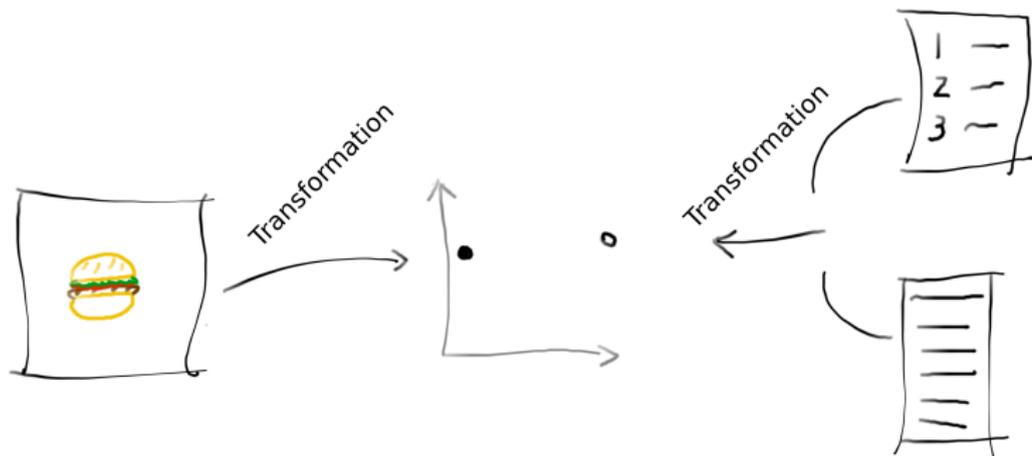
What can we do with these points?

Transformation

Transformation

Common Representation

Introduction
○○○○○●○○○○

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○

SCIENCES
SORBONNE
UNIVERSITÉ

Dataset composed of pairs image-recipe

	ingr (ingredients)	instr (cooking instructions)	image
Pizza	1) *pizza dough* 2) *hummus* 3) *arugula* 4) *cherry / grape tomatoes* 5) *pitted greek olives* 6) *crumbled feta cheese*	1) *Cut the dough into two 8-ounce sized pieces.* 2) *Roll the ends under to create round balls.* 3) *Then using a well-floured rolling pin, roll the dough out into 12-inch circles.* 4) *Place the dough circles on sheets of parchment paper.*	
Pecan Pie	1) *unsalted butter* 2) *eggs* 3) *condensed milk* 4) *sugar* 5) *vanilla extract* 6) *chopped pecans* 7) *chocolate chips*	1) *Preheat the oven to 375 degrees F.* 2) *In a large bowl, whisk together the melted butter and eggs until combined.* 3) *Whisk in the sweetened condensed milk, sugar, vanilla, pecans, chocolate chips, butterscotch chips, and co-conut.*	

Crawled from recipe websites by Salvador et al., CVPR 2018

Introduction
○○○○○○○●○○

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○

Back to the common representation

Common Representation

Introduction
○○○○○○○●○

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○
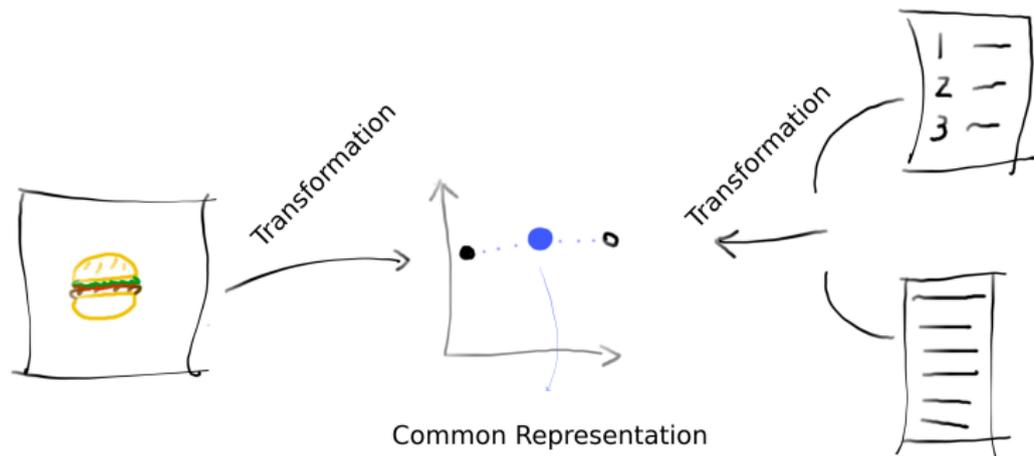
Transformation? Artificial Neural Networks

Common Representation

Introduction
○○○○○○○○○●

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○

How does it look like?

Introduction
ooooooooo

Problem-driven proposal
●ooooo

Experiments
ooooooo
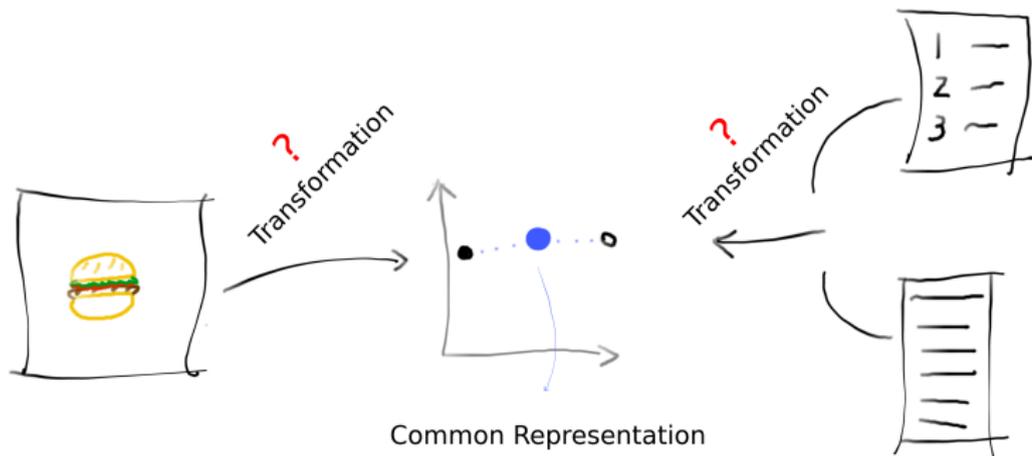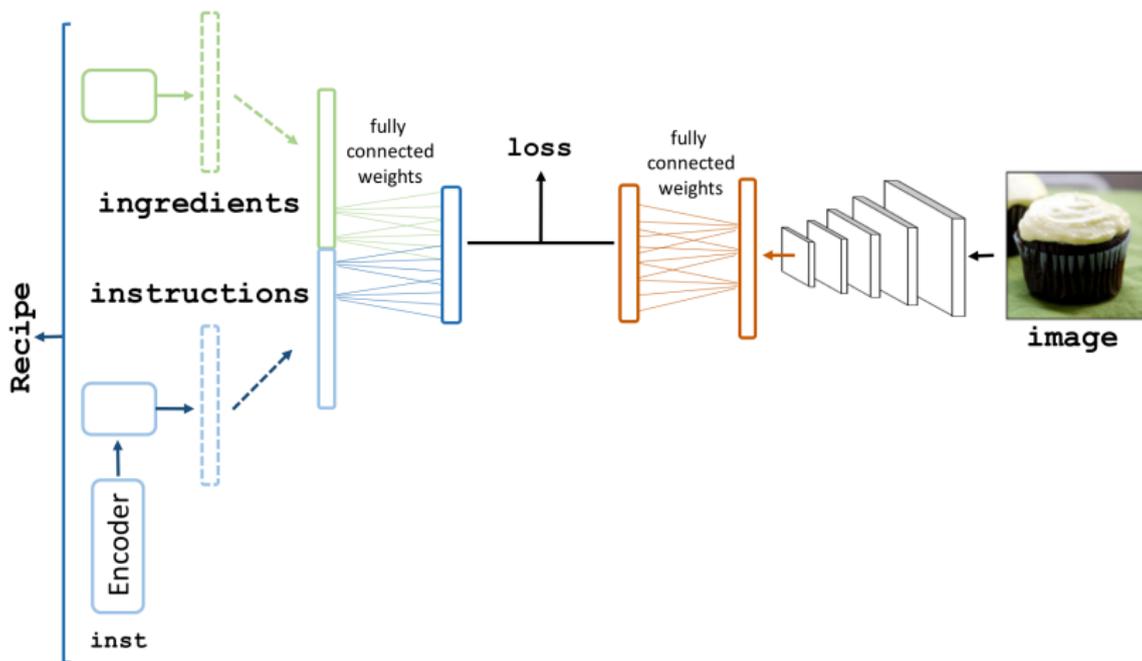
Challenge – part 1

Challenge: What is the distance between… ?

1275 VS 1317

It's *easier* to find distances between numbers than between images

Introduction
○○○○○○○○○

Problem-driven proposal
●○○○○○

Experiments
○○○○○○○

Challenge – part 1

SCIENCES
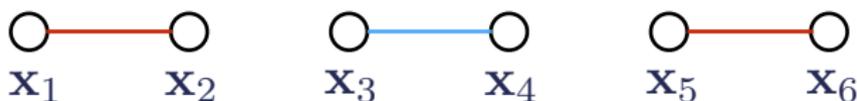SORBONNE
UNIVERSITÉ

Challenge: What is the distance between... ?

Introduction
○○○○○○○○○

Problem-driven proposal
○●○○○○○

Experiments
○○○○○○○

Pairwise/contrastive

Trained on the paired data $\{(\mathbf{x}_i, \mathbf{x}_j, y_{i,j})\}$, with the cost function

$$y_{i,j} D_{i,j}^2 + (1 - y_{i,j})[\alpha - D_{i,j}]_+^2$$

$$y_{i,j} \in \{0, 1\}, \qquad D_{i,j} = ||f(\mathbf{x}_i) - f(\mathbf{x}_j)||_2, \qquad [\cdot]_+ = max(0, \cdot)$$

[☺] Approaches positive pairs and distances negative pairs by α;

[☹] Forces positive examples to have distance 0;

[☹] (...) Other problems, lets just agree it's *not optimal*.

Introduction
ooooooooo

Problem-driven proposal
oo●oooo

Experiments
ooooooo

Challenge – part 2

Challenge: What is the distance between... ?

Challenge – part 2

Challenge: What is the distance between… ?

Introduction
oooooooooo

Problem-driven proposal
ooo●ooo

Experiments
ooooooo

Challenge – part 2

Challenge: What is the distance between... ?

Introduction
ooooooooo

Problem-driven proposal
ooo●oo

Experiments
ooooooo

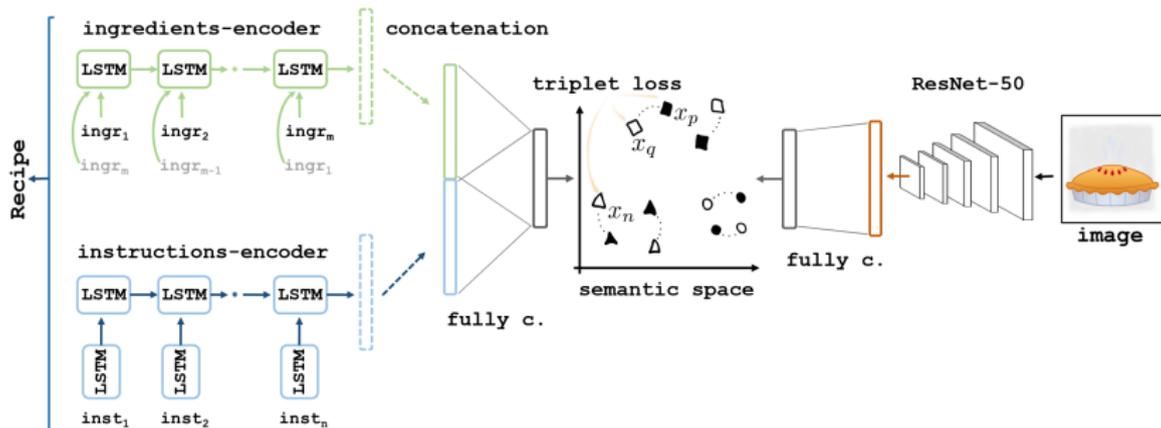SCIENCES
SORBONNE
UNIVERSITÉ

Triplet

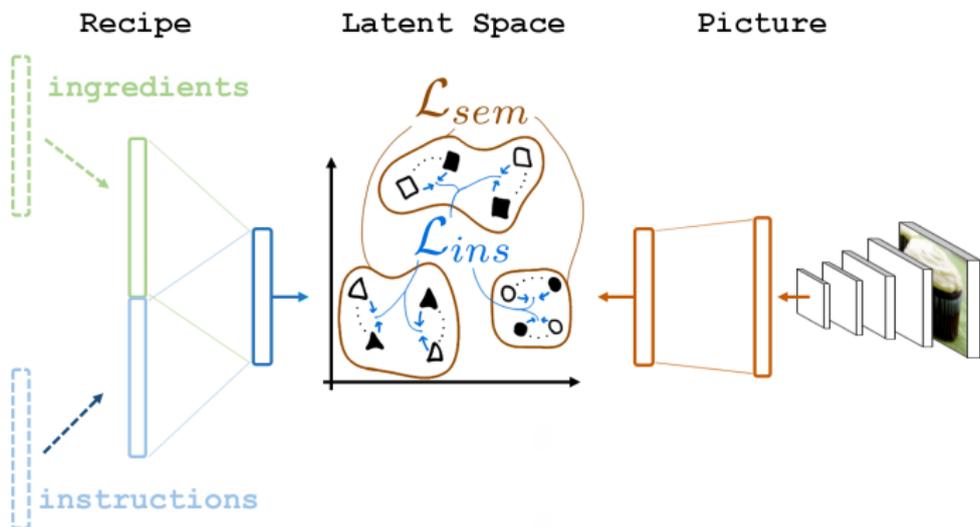Trained on $\{(x_a, x_p, x_n)\}$, with the cost function

$$[D^2_{ia,ip} - D^2_{ia,in} + \alpha]_+$$

[☺] Approaches positive examples and distances negative examples;

[☺] Pushes away *the negative example* and closer *the positive example* if the negative one is inside $D^2_{ia,ip} + \alpha$;

Introduction
ooooooooo

Problem-driven proposal
oooo●o

Experiments
ooooooo

Constructing latent semantic spaces

Adding semantic information

Semantic-based loss $\mathcal{L}_{sem}$ added to organize the feature space. It is used to train the parameters of the network

Experiments

Introduction
○○○○○○○○○

Problem-driven proposal
○○○○○○

Experiments
●○○○○○○

Dataset composed of pairs image-recipe

	ingr (ingredients)	instr (cooking instructions)	image
Pizza	1) pizza dough 2) hummus 3) arugula 4) cherry / grape tomatoes 5) pitted greek olives 6) crumbled feta cheese	1) Cut the dough into two 8-ounce sized pieces. 2) Roll the ends under to create round balls. 3) Then using a well-floured rolling pin, roll the dough out into 12-inch circles. 4) Place the dough circles on sheets of parchment paper.	
Pecan Pie	1) unsalted butter 2) eggs 3) condensed milk 4) sugar 5) vanilla extract 6) chopped pecans 7) chocolate chips	1) Preheat the oven to 375 degrees F. 2) In a large bowl, whisk together the melted butter and eggs until combined. 3) Whisk in the sweetened condensed milk, sugar, vanilla, pecans, chocolate chips, butterscotch chips, and coconut.	

Crawled from recipe websites by Salvador et al., CVPR 2018

Nearly 1 million recipes, about 800,000 images

Introduction
○○○○○○○○○

Problem-driven proposal
○○○○○○

Experiments
○●○○○○○

Quality of semantic space

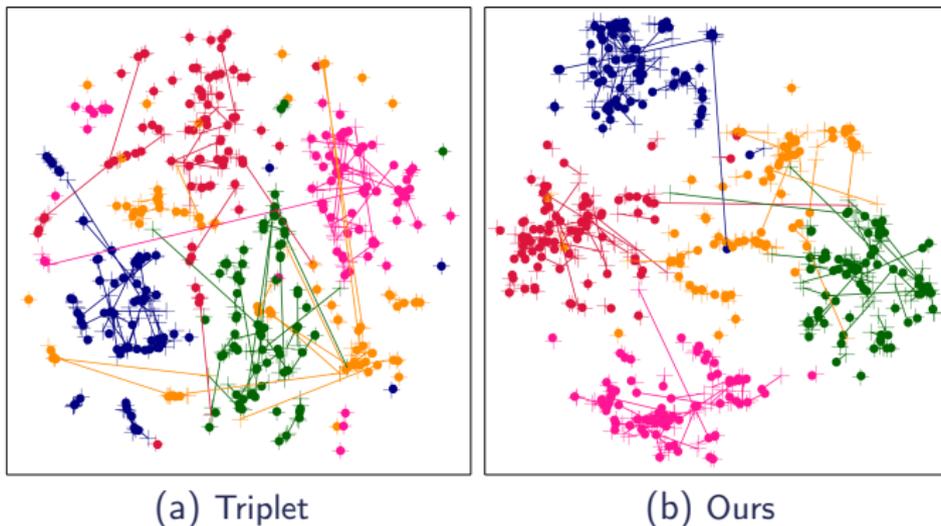SCIENCES
SORBONNE
UNIVERSITÉ

(a) Triplet   (b) Ours

Figure 1: **t-SNE visualization.** Image (resp. Recipe) points are denoted with the + (resp. ●) symbol. Matching pairs are connected with a trace. Blue points are associated to the cupcake class, orange to hamburger, pink to green beans, green to pork chops, and red to pizza.

Measuring the retrieval effectiveness

	im2recipe @ 1k			recipe2im @ 1k		
	MedR	R@1	R@10	MedR	R@1	R@10
CCA [1]	15.7	14.0	43.0	24.8	9.0	35.0
PWC [1]	5.2	24.0	65.0	5.1	25.0	65.0
PWC++ (pairwise, ours)	3.3 ± 0.4	25.8 ± 1.6	67.1 ± 1.4	3.5 ± 0.5	24.8 ± 1.1	67.1 ± 1.2
Ours	$\mathbf{1.0} \pm 0.1$	$\mathbf{39.8} \pm 1.8$	$\mathbf{77.4} \pm 1.1$	$\mathbf{1.0} \pm 0.1$	$\mathbf{40.2} \pm 1.6$	$\mathbf{78.7} \pm 1.3$

Table 1: **State-of-the-art comparison.** MedR means Median Rank (lower is better). R@K means Recall at K (between 0% & 100%, higher is better). Mean and std values over 10 bags of 1k pairs each are reported

[1] Salvador et al., "Learning Cross-modal Embeddings for Cooking Recipes and Food Images," CVPR'17.

Introduction
●●●●●●●●●
Problem-driven proposal
●●●●●●
Experiments
●●●●●●●●

Qualitative analysis – Query

Crunchy Onion Potato Bake

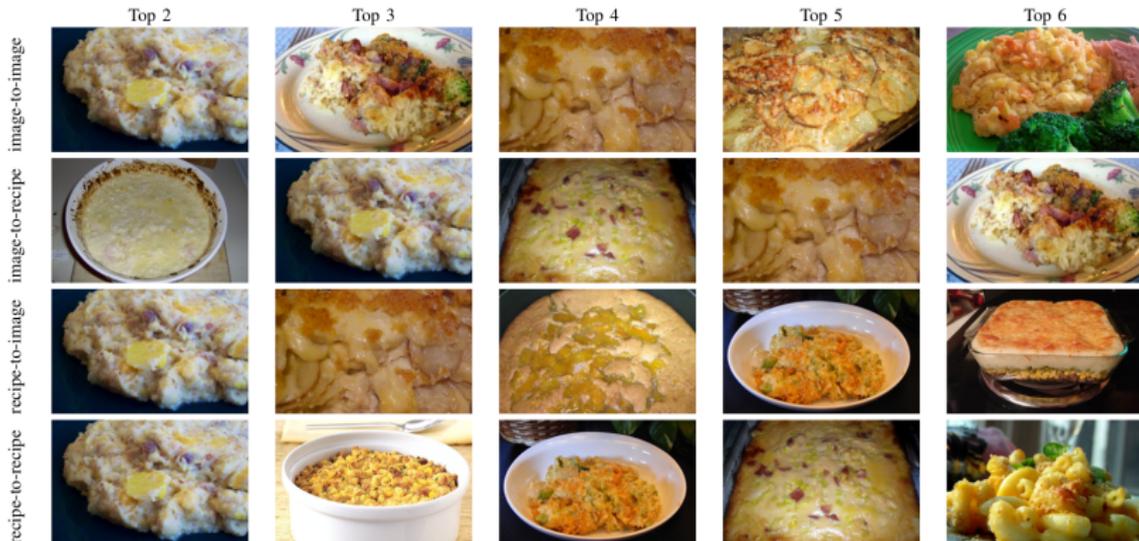| Ingredients	Cooking instructions	Image
Milk, Water, Butter, Mashed potatoes, Corn, Cheddar cheese, French-fried onions	*Preheat oven to 350 degrees Fahrenheit. Spray pan with non stick cooking spray. Heat milk, water and butter to boiling; stir in contents of both pouches of potatoes; let stand one minute. Stir in corn. Spoon half the potato mixture in pan. Sprinkle half each of cheese and onions; top with remaining potatoes. Sprinkle with remaining cheese and onions. Bake 10 to 15 minutes until cheese is melted. Enjoy !*	

Introduction
○○○○○○○○○

Problem-driven proposal
○○○○○○

Experiments
○○○○●○○

Qualitative analysis – Top retrieved items for the query

Introduction
○○○○○○○○○

Problem-driven proposal
○○○○○○

Experiments
○○○○○●○

Qualitative analysis – Looking for ingredients
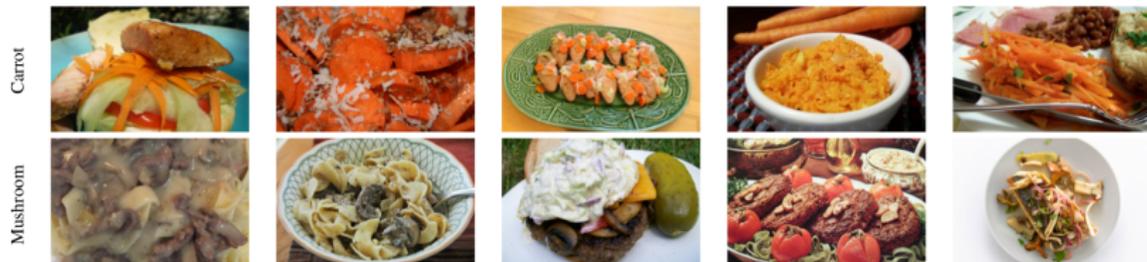

TABLE V
EXAMPLES OF IMAGES IN THE TOP 20 WHEN SEARCHING FOR THE INGREDIENT *Carrot* (TOP ROW) OR *Mushroom* (BOTTOM ROW).

Introduction
○○○○○○○○○

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○●
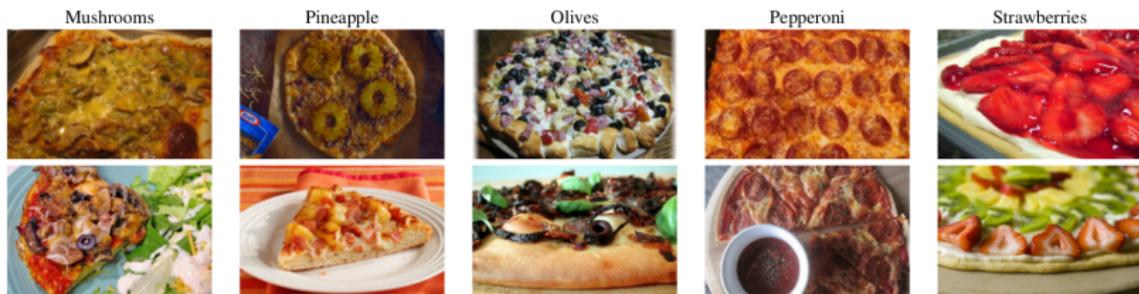
Qualitative analysis – Looking for ingredients on pizza
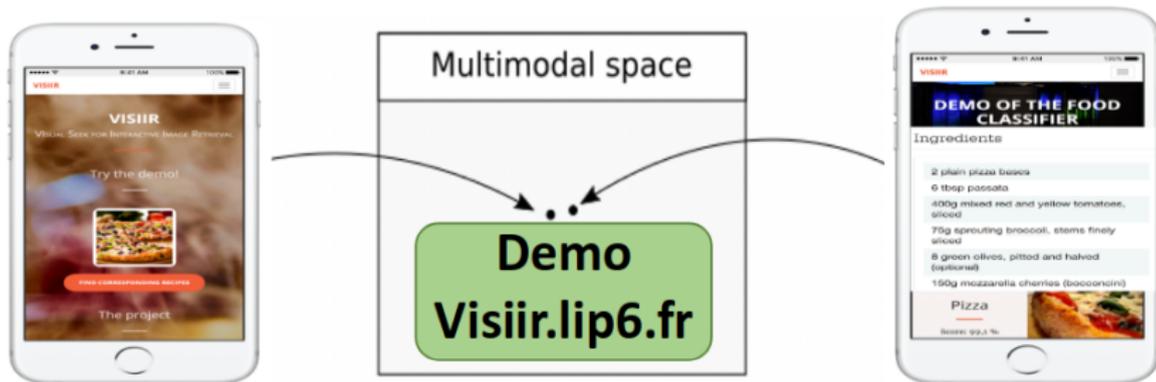

TABLE VI
EXAMPLES OF IMAGES IN THE TOP 20 RESULTS WHEN SEARCHING FOR A SPECIFIC INGREDIENT WITHIN THE CLASS *Pizza*.

Try one of our demos

Introduction
○○○○○○○○○

Problem-driven proposal
○○○○○○

Experiments
○○○○○○○

Thank you

Micael Carvalho*, Rémi Cadène*,
David Picard, Laure Soulier, and Matthieu Cord

micael.carvalho@lip6.fr

micaelcarvalho.com

Laboratoire d'Informatique de Paris 6; Sorbonne Université, Paris, France