Automatic expansion of a food image dataset leveraging existing categories with domain adaptation

Yoshiyuki Kawano, Keiji Yanai

The University of Electro Communications, Tokyo, JAPAN

Background

- Large-scale image dataset
  - up to over 1000
  - Using crowd-sourcing
    - Semi-automatically build

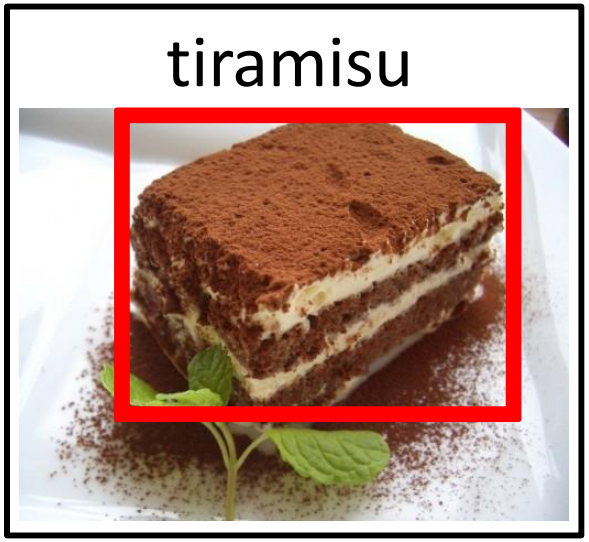
- The issue
  - Costly
  - Performance is not perfect

We aim to reduce cost and improve accuracy
by noise removal of data in advance automatically.

Objective

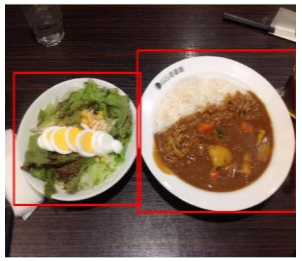
- Expand an existing image dataset
  - Target domain: food image
  - 100 kinds of Japan foods -> 256 kinds of worldwide foods
  - Noise removal

- Annotation
  - Category label
  - Bounding box



Related work

- Food image dataset
  - UEC Food100 [Matsuda. ICME2013]
    - One country (japan) food
    - Expande this dataset.



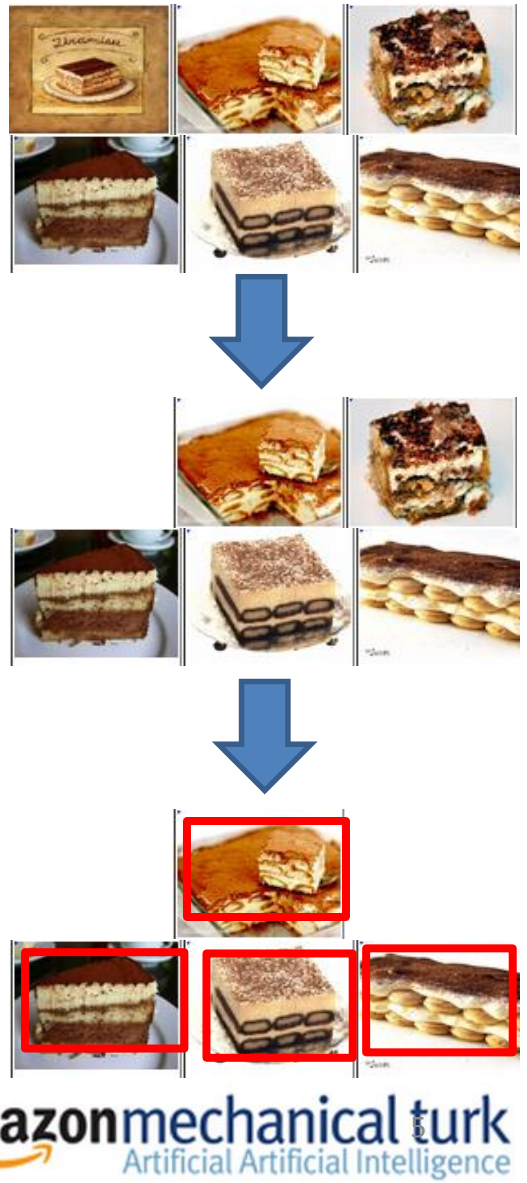
- Large-scale database by using AMT
  - ImageNet (http://image-net.org) [CVPR2009]
    - Web image search + AMT

IMAGENET

- CV+ active learning + crowdsourcing
  - Object detector and AMT [Vijayanarasimhan. CVPR2012]

Overview of the flow

1. Collect images from the Web
   - Web image search using keyword
2. Automatic noise removal by CV
   - Foodness classifier and adaptive SVM
   - Employing existing dataset
3. Manually noise removal by AMT
   - representative image selection task, noise removal task, drawing BB task,



Detail of the flow

(1) pre-processing

Existing food DB

Food group

HOG FV → SVMs

Color FV → SVMs

(2) Query: food name

Web (Flickr, Twitter, Bing)

Web API

(3-a) Apply "Foodness" Classifier (FC)

Target images (Web)

VisualRank

HOG FV → A-SVM

Color FV → A-SVM

(3-b) Apply transfer learning (A-SVM)

Target images (CV)

(4-a) Select representative sample

Sample images

(4-b) Noise removal

Target images

(4-c) Draw bounding boxes

Annotated images

Food photos of the new categories

1. Collect images from the Web

- Prepare a new food name.
  - e.g. “mango pudding”, ”tiramisu”
- Collecting 


  - Image search with a keyword using Web API (Flickr, Twitter, bing)
  - Local language and English

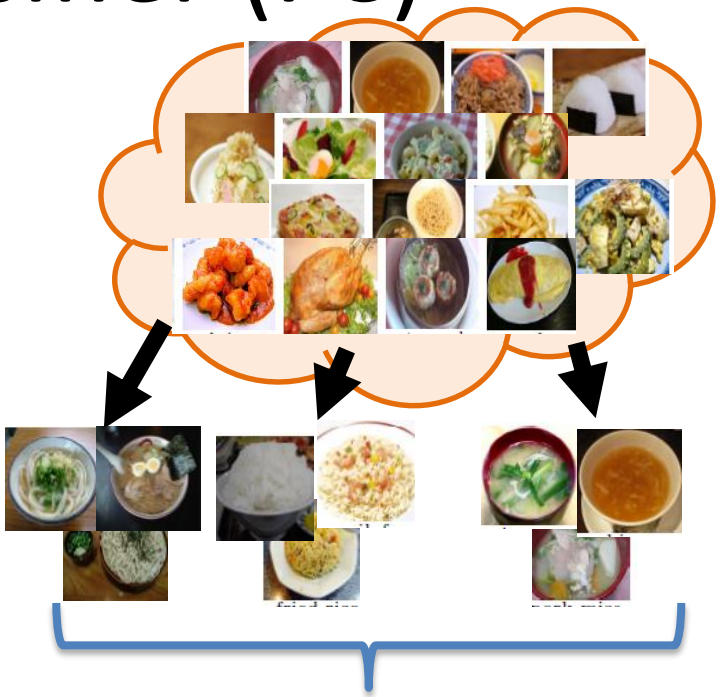
2. Noise removal by CV

- Automatic noise removal by 2-step filtering

1. Foodness Classifier
   - Food or non food image
   - Training: using the existing food image dataset.

2. Adaptive SVM
   - the new category image or not.
   - Training: Higher foodness images

2.1. Foodness Classifier (FC)



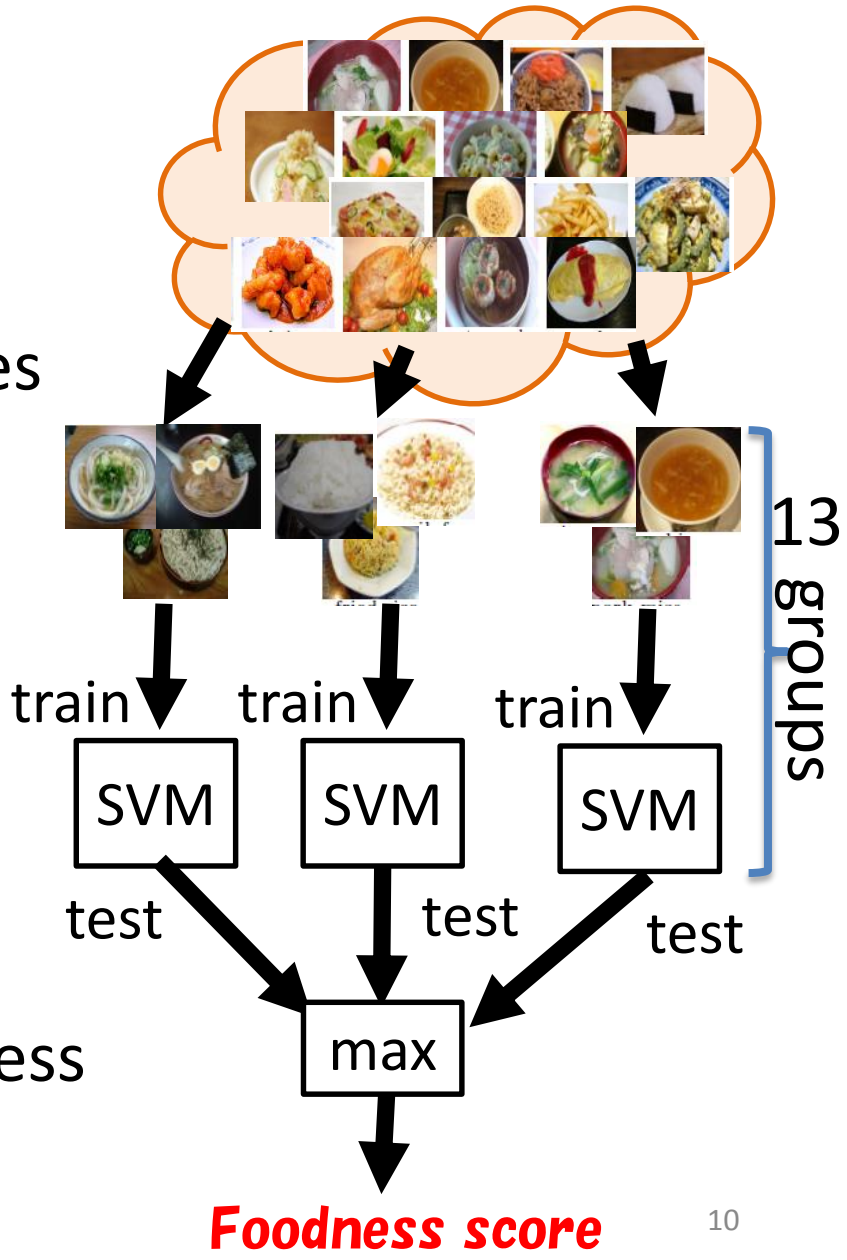
- Objective
  - food / nonfood classification
  - use the existing dataset
- Image features, Classifier
  - HOG and color FV, linear SVMs
- Grouping of the existing food categories
  - multi-class classification
  - Build a confusion matrix
  - Make some category groups based on the matrix.

    (inspired by [Bergano et al. CVPR 2011])

2.1. Foodness Classifier (FC)

- Train
  - positive : each 13 group.
  - negative: non food web images
    - street stall, kitchen,
    - dinner party and restaurant

- Test
  - The maximum value as foodness



2.1 group and some food categories

noodles	udon nooles, dipping noodles, ramen
yellow color	omlet, potage, steamed egg hotchpotch
soup	miso soup, pork miso soup, japanese tofu and vegetable chowder
fried	takoyaki, japanese-stype pancake, fried noodle
deep fried	croquette, sirloin cutlet, fried chicken
salad	green salad, sauteed vegetables, vegetable tempra
bread	sandwiches, raisin bread, roll bread
seafood	sashimi, sashimi bowl, sushi
rice	rice, pilaf, fried rice
fish	grilled salmon, grilled pacific saury, dried fish
boiled and seasoned	sesoned beef with potatoes simmered ganmodoki sesoned beef with potatoes
sauteed	sauteed vegetables, goya chanpuru, kinpira-style sauteed burdock
sauce	stew, curry, stir-fried shrimp in chili sauce

2.1 group and some food categories

noodles	udon nooles, dipping noodles, ramen
yellow color	omlet, potage, steamed egg hotchpotch
	ed salmon, grilled pacific saury, dried fish
boiled and seasoned	sesoned beef with potatoes simmered ganmodoki sesoned beef with potatoes
sauteed	sauteed vegetables, goya chanpuru, kinpira-style sauteed burdock
sauce	stew, curry, stir-fried shrimp in chili sauce

2.1 group and some food categories

noodles	udon nooles, dipping noodles, ramen
yellow color	omlet, potage, steamed egg hotchpotch
fish	grilled salmon, grilled pacific saury, dried fish
boiled and seasoned	sesoned beef with potatoes simmered ganmodoki sesoned beef with potatoes
sauteed	sauteed vegetables, goya chanpuru, kinpira-style sauteed burdock
sauce	stew, curry, stir-fried shrimp in chili sauce

2.1 group and some food categories

noodles	udon nooles, dipping noodles, ramen
yellow color	omlet, potage, steamed egg hotchpotch
soup	miso soup, pork miso soup, japanese tofu and vegetable chowder
fish	grilled salmon, grilled pacific saury, dried fish
boiled and seasoned	sesoned beef with potatoes simmered ganmodoki sesoned beef with potatoes
sauteed	sauteed vegetables, goya chanpuru, kinpira-style sauteed burdock
sauce	stew, curry, stir-fried shrimp in chili sauce

2.1 group and some food categories

noodles	udon nooles, dipping noodles, ramen
yellow color	omlet, potage, steamed egg hotchpotch
boiled and seasoned	sesoned beef with potatoes simmered ganmodoki sesoned beef with potatoes
sauteed	sauteed vegetables, goya chanpuru, kinpira-style sauteed burdock
sauce	stew, curry, stir-fried shrimp in chili sauce

2.2. adaptive SVM

Objective

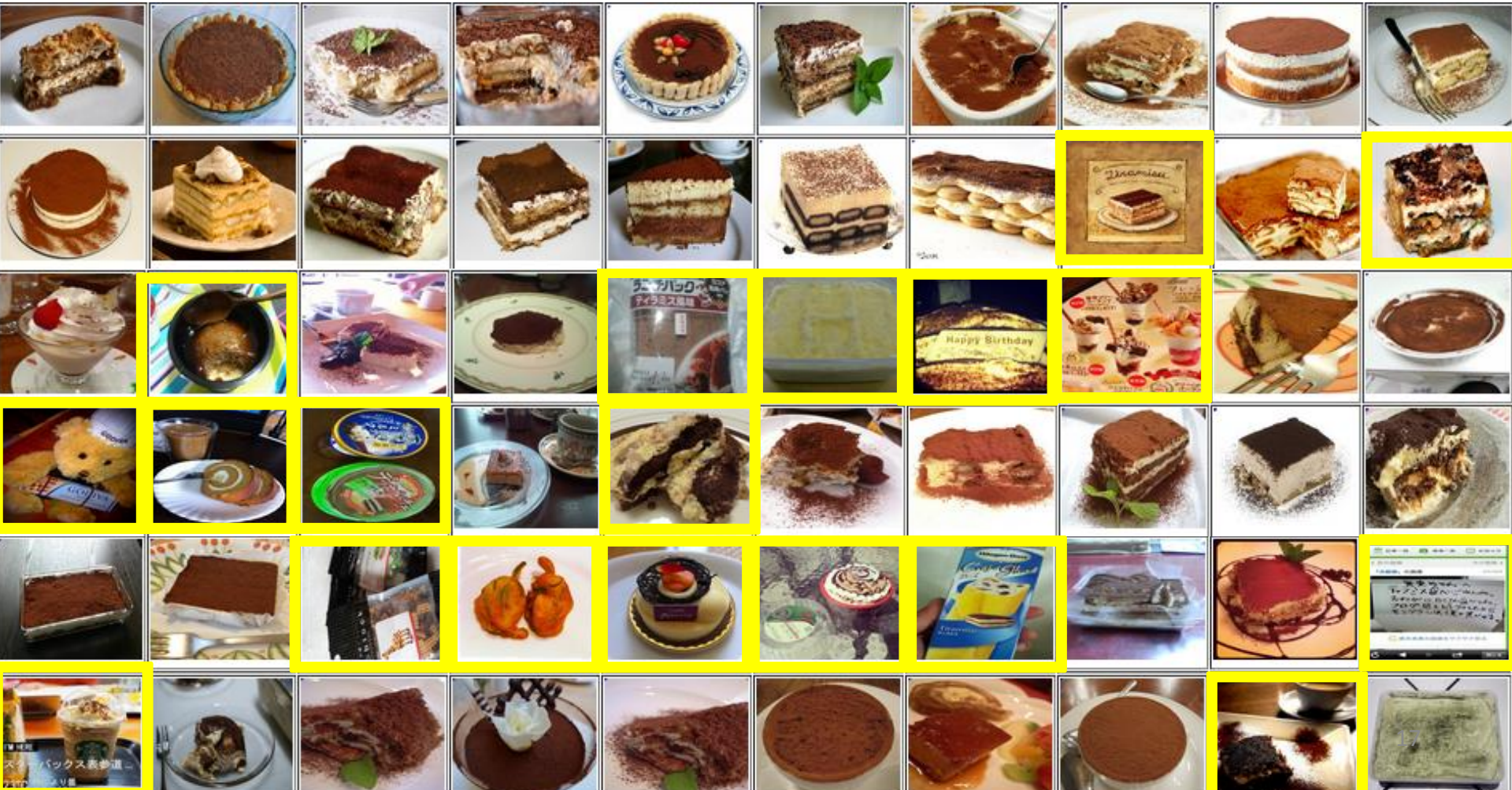
- Classify the new category or not
- [issue] no training data of the new unknown category.
- Employ the existing dataset by transfer learning,

• Use image recognition (same as 2.1 except A-SVM)

- Image features + **adaptive SVM**
- [pseudo-positive]: the images with higher VisualRank value (apply VisualRank for the remaining images)
- [negative]: the images with lower foodness score
- [source domain]: The most related food group

Collecting Web images

- tiramisu



Apply Foodness Classifier

- tiramisu

Apply adaptive SVM

- tiramisu

Some examples of the pair of target foods and source food groups.

target　the most related food group

Okinawa soba




Mango pudding




minestrone




nachos




3. Manually noise removal by AMT

- 3-step crowdsourcing
  - Representative select sample task
    - 0.06$ per HIT, 5 workers
  - Noise removal task
    - 0.03$ per HIT, 5 workers
  - Drawing bounding box task
    - 0.05$ per HIT, 4 workers

Experiment

- Evaluation
  - Performance comparison on food image filtering
  - The final results after AMT and analysis of cost
- Setup
  - Web API (Flickr, twitter ,bing)
  - More than 600 images for each category
  - 5 countries, total 35 foods
- abbreviation
  - Food Classifier: FC
  - Adaptive SVM: A-SVM
  - Noise removal task (AMT): NR Task

Experimental result

Average precision @300

Method	America	Japanese	Chinese	Thailand	Indonesia	Average
Visual Rank (VR)	58.47	54.95	60.66	62.19	58.71	59.00
Food Classifier (FC)	78.00	75.33	77.61	82.85	78.61	78.48
FC + SVM	85.57	83.38	85.09	89.23	82.23	85.10
FC + A-SVM	89.61	87.76	87.76	91.38	84.09	88.12

Experimental result

Precision of food image on dataset

	precision[%]	**gain[%]**
FC	91.10	-
FC + A-SVM	94.19	+ 3.09
FC + A-SVM + NR Task	97.83	+ 3.64

Experimental result

Cost (annotated 100 images)

	Noise removal		**Draw BB**		**Total**
	Recovery ratio[%]	cost[$]	Recovery ratio[%]	cost[$]	cost[$]
FC	-	-	64.2	3.11	3.11
FC + A-SVM	-	-	74.7	2.68	2.68
FC + A-SVM + NR-Task	80.9	0.74	86.7	2.31	3.16

Conclusion

- Proposed a novel framework to expand an existing image dataset automatically
  - Generic classifiers
  - Domain adaptation

- Feature Works
  - make further analysis on the difference between a manually collected dataset and an automatically collected dataset.

3.1. サンプル選択タスク

- タスクの特徴
  - 30枚を与えて、最大7枚、
    一般的なサンプルを選択してもらう
  - リンクを与えるので勉強してもらう
  - $0.06、4人
    - 少し高め、外部リンクで勉強させる、また重要であるため
  - 質の向上
    - サンプル選択をしないとヒットを提出できない
    - 設置したリンクをたどらないとヒットを提出できない

ワーカーの国別人口

- 自由回答によるワーカーの国別人口
  - インド人が圧倒的に多い

india	usa	Canada	French	uk	Asia
1668	394	20	3	3	4

- 今回対象にした食事は、世界の食事
  - インド人は未知の食事も多い

悪意のあるワーカー

- 特定の一人の悪意のあるワーカーの結果

3.3. example of drawing BB task

Obtained representative images

Draw bounding box or mark irrelevant images

3.3. the real exapmle

Draw bounding box around rice and tea in image.

Description

Overview

Draw Bounding-Box around rice and tea in images.
You can draw Bounding-Box by Drag-and-Drop.
A Bounding-Box you draw is displayed as a red rectangle.
If target is curry, an example is shown bellow.

Now, target (rice and tea) is shown below(automatically selected).

You can redraw Bounding-Box just by left-clicking on a image again.
Previous one will disappear.

If there are no rice and tea in a image,
the target is invisible more than half of it and blurred,
the target is not a real food item such as illustraion, animetion and package or
the target is not "ready-to-eat",
check the image as "no rice and tea" by right clicking on it.
In images checked as "no rice and tea", a red cross is drawn as shown below.

You can uncheck a image just by right-clicking on a image again.
You can redo everything except submitting in this HIT.
Don't be so nervous.

Note

1. Images

   We have already filtered no rice and tea images by mturk, but images maybe include some no rice and tea images.

2. When you are not sure

   When you are not sure if you should draw Bounding-Box or check as "no rice and tea", please check as "no rice and tea".

3. About image loading

   You have 10 images in this HIT.
   When there are images that are not loaded, please reload this page after a while.
   If reloading doesn't solve the problem, please skip this HIT.
   Click Here to Reload

4. About "cooking-in-progress"

   "cooking-in-progress" image means "no rice and tea".
   Please draw Bounding-Box only on a image which is "ready-to-eat".
   "Toy foods" also should not be marked.

5. About Dish

   If rice and tea is on a dish in a image, please draw Bounding-Box not including the dish.
   So, when there are multiple dishes which have rice and tea on, you must choose whichever to draw Bounding-Box around too.
   Remember, "One Bounding-Box, one food".
   Don't let a Bounding-Box include a dish.

Examples

Last, I show the good example and bad example for curry.

GOOD example

BAD example

What is rice and tea?

About rice and tea information.
Please you visit on web site Google Image Search, bing or Wikipedia for more detail.

I show the target (rice and tea) images again.

Now please start this HIT

TARGET IS NOW rice and tea

Please answer one question and tell us your nationality (country name)
Fill this form everytime you submit.
Thank you for your contribution.
sample images of rice and tea are useful or useless for this task ? useful so so useless
: nationality
Please provide your opinions to make this HIT more friendly to workers.
We appreciate your feedback!
(this form is optional)

: nationality
Please provide your opinions to make this HIT more friendly to workers.
We appreciate your feedback!
(this form is optional)

Your state:

You have not visit site yet.
draw bounding box: 0 image.
checked noObject: 0 image.
unchecked: 10 images.

Have you read "Note" section above?
YOU WILL BE REJECTED if you are regarded you don't follow the rules in "Note" section.
Before submitting, make sure that you worked correctly.

3.3. the final result

Apple pie

3.1. the result of selecting representative task

- Obtain about 5 images
  - 100 categories100%

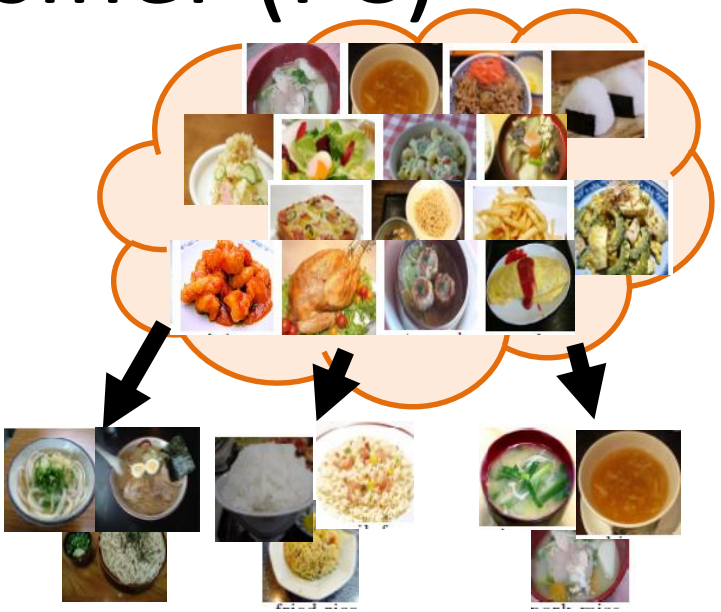
- Worker's evaluation
  - Ask workers in each HIT

	useful	so so	useless
NR Task	89.59	7.90	2.52
BB Task	91.68	7.02	1.31

Basic idea

- Food category
  - Look more similar to each other
  - Visual knowledge on food help to annotations other food.
- Method
  - Effective framework by domain adaptation

2.1. Foodness Classifier (FC)



- objective
  - food / nonfood classification
- Image
  - Image features + linear SVM
  - positive : make 13 group based on confusion matrix. (originally 100 kinds) [Bergamo et al., CVPR2012],
  - negative: gather images using web images
    - street stall, kitchen, dinner party and restaurant
  - Train SVMs for each 13 groups
  - The highest evaluation value is foodness

2.1. Foodness Classifier (FC)

- Objective
  - food / nonfood classification
  - use the existing dataset
- Image features, Classifier
  - HOG and color Patch FV, linear SVMs
- Train and Test
  - positive : make 13 group based on confusion matrix. [Bergamo et al., CVPR2012],
  - negative: non food web images
    - street stall, kitchen, dinner party and restaurant
  - Train SVMs for each 13 groups
  - The highest evaluation value is foodness

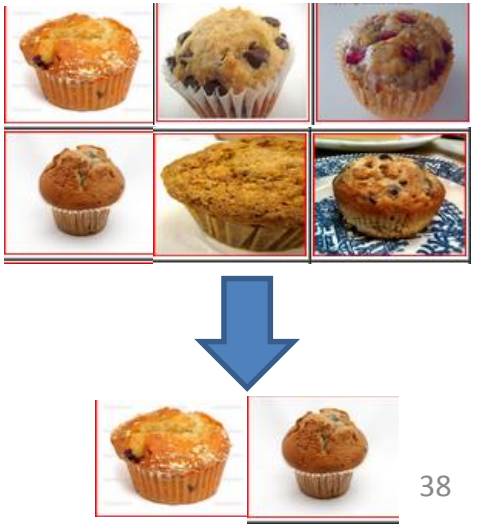
3.1. Selecting representative sample task

- issue

  Worker dose not have knowledge about various kinds of foods.

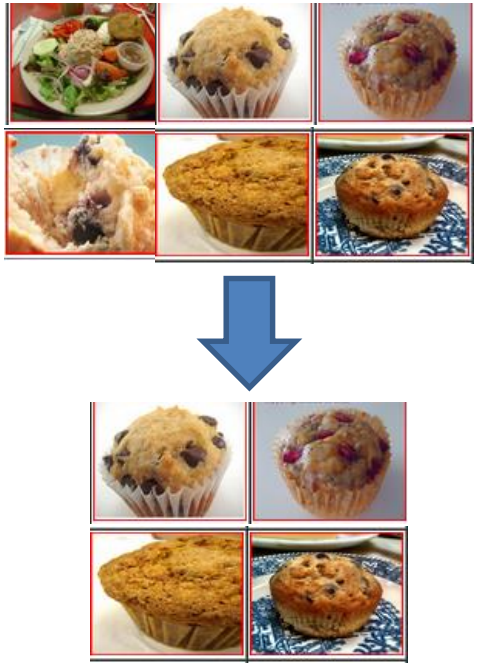
- Objective
  - Obtain highly accurate representative samples for other task.

- Task
  - 1HIT 0.06$, 5 workers
  - Needs to study about the food category



3.2. Noise removal task

- Objective
  - Manually noise removal



- Task
  - 1 HIT $0.03, 5 workers
  - Majority voting
  - Cannot be submit if there are more than 4 unchecked images

3.3. Drawing BB task

- objective
  - Drawing bounding box
  - Manually noise removal
- Task
  - 1 HIT $0.05, 4 workers
  - can mark irrelevant images