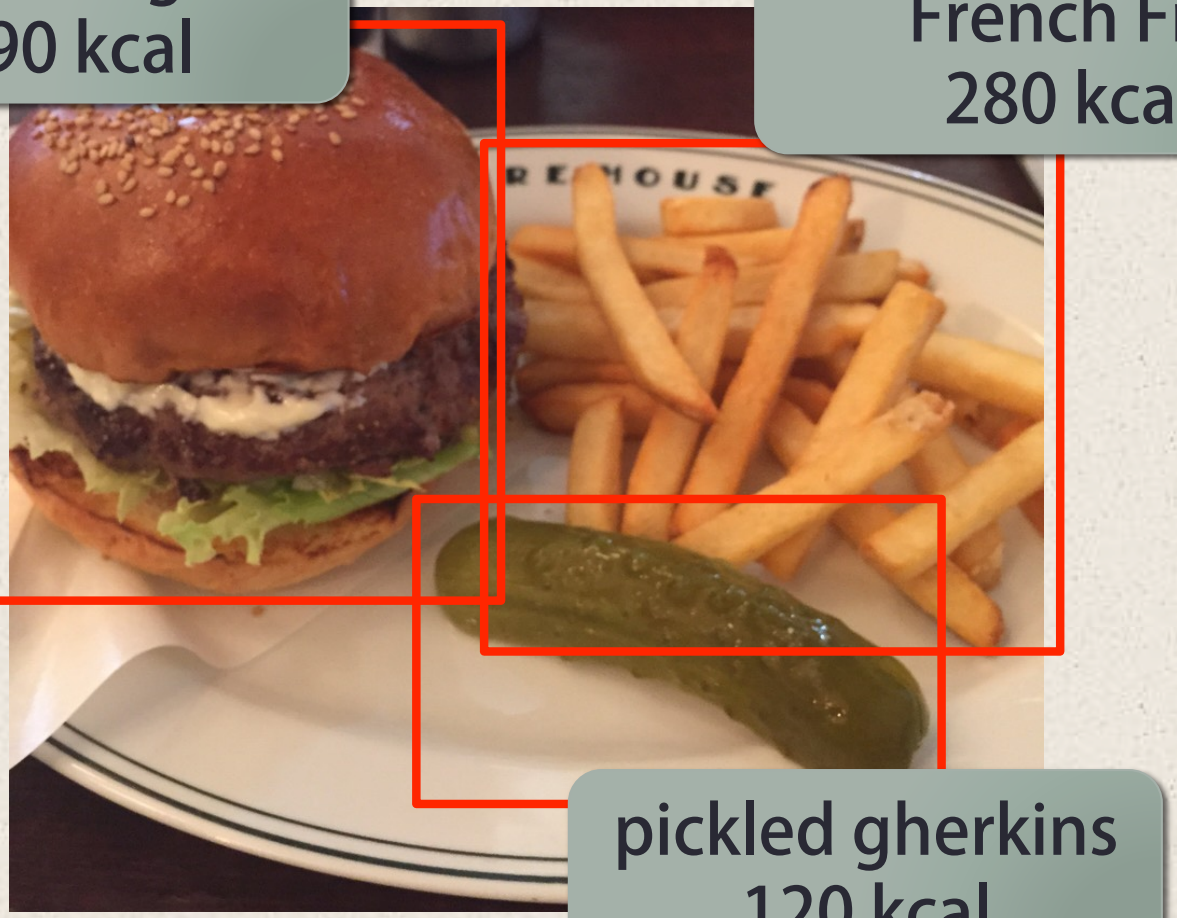
1st International Workshop on
Multimedia Assisted Dietary Management (MADiMa)

Highly Accurate Food/Non-Food Image Classification based on a Deep Convolutional Neural Network

6th, September 2015

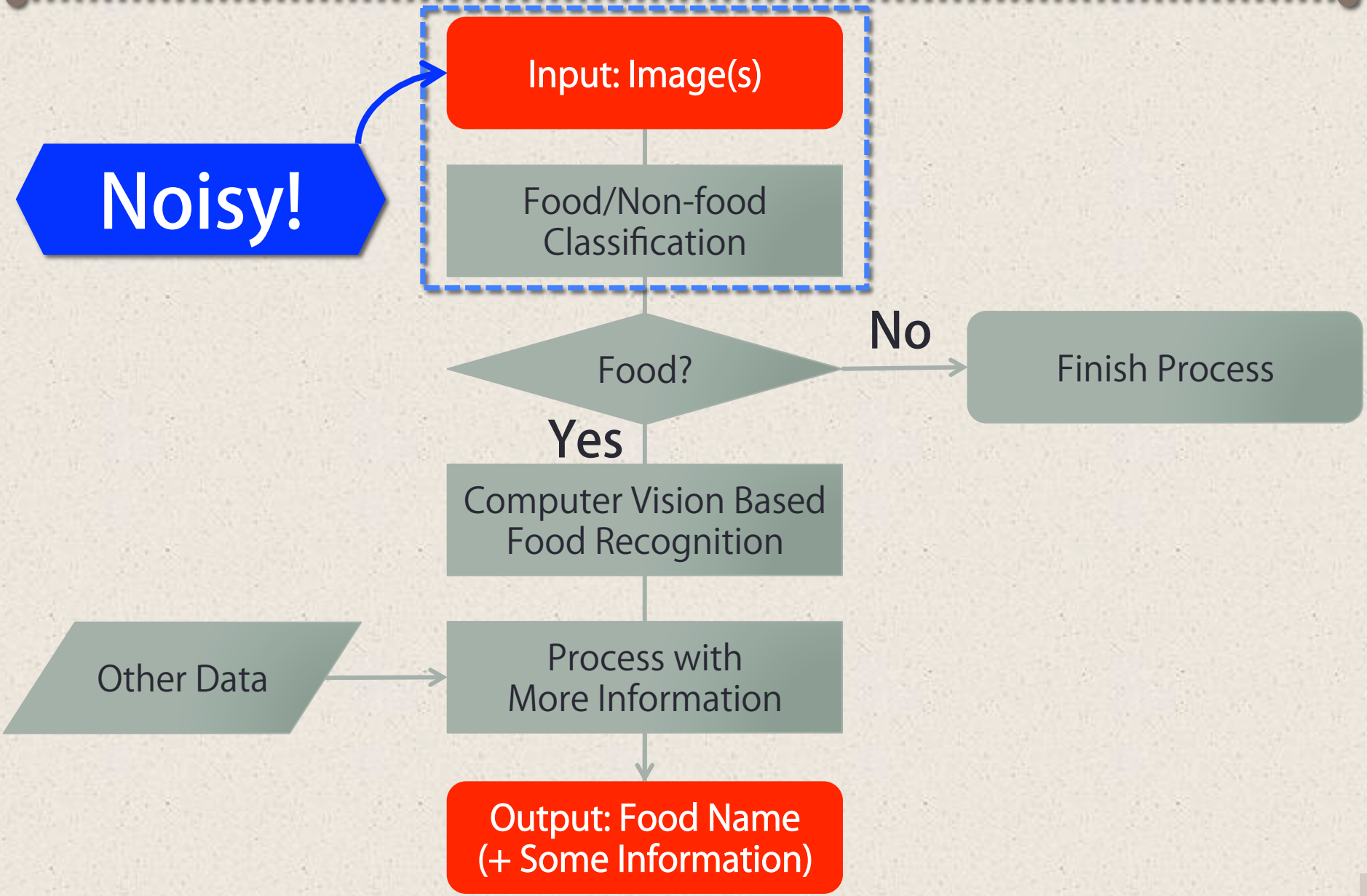
Hokuto Kagaya and Kiyoharu Aizawa
The University of Tokyo

Food Image Analysis



* Displayed calories are not real values.

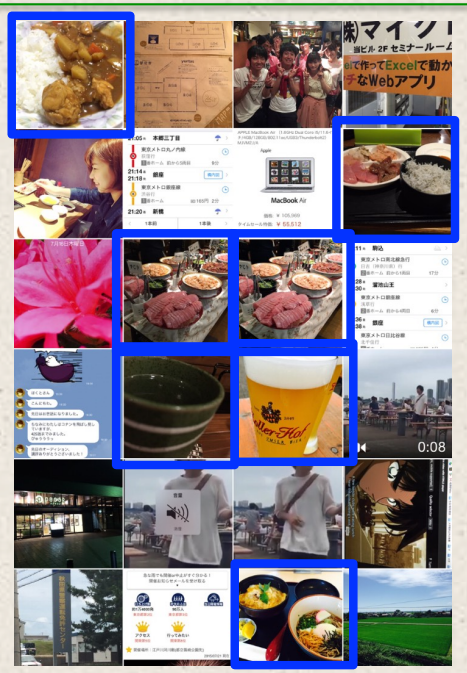
Food Analysis Framework



Noisy Input Images

- Case 1: Personal photo albums
  - Personal photos is usually not arranged
  - We want our photos in personal albums to be classified food or non-food automatically.

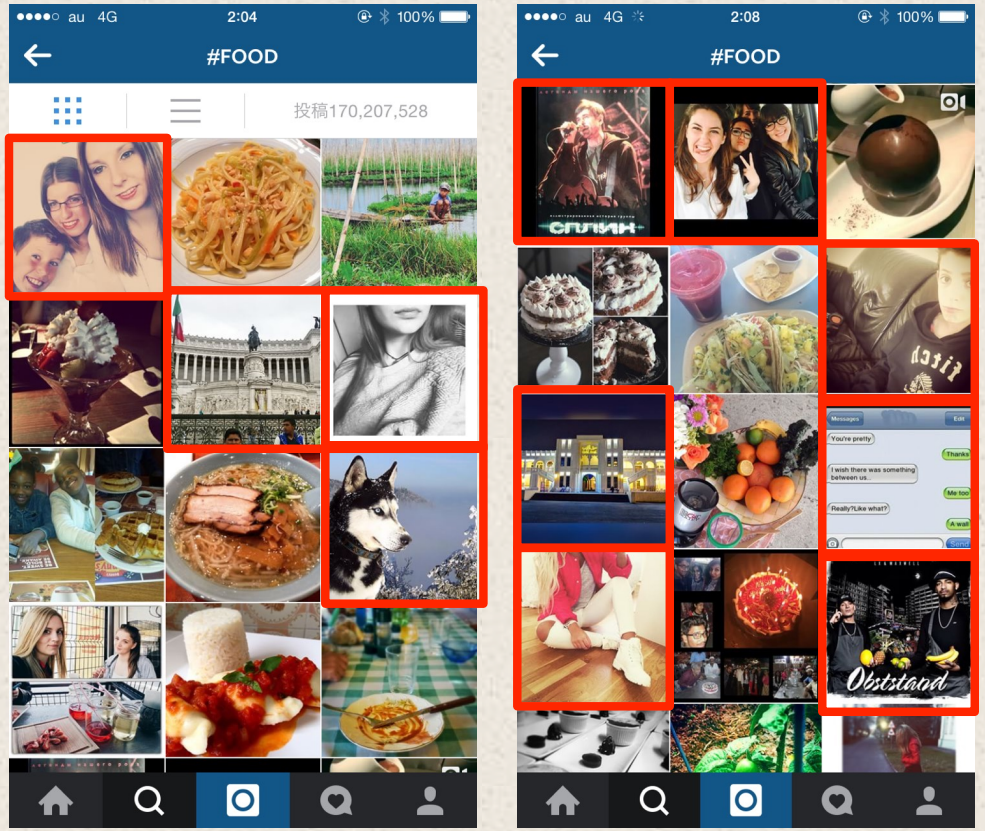
My photo album in iPhone



Noisy Input Images

- Case 2: SNS photos
  - Hash tags are very confusing

The search results
of "#food" in Instagram



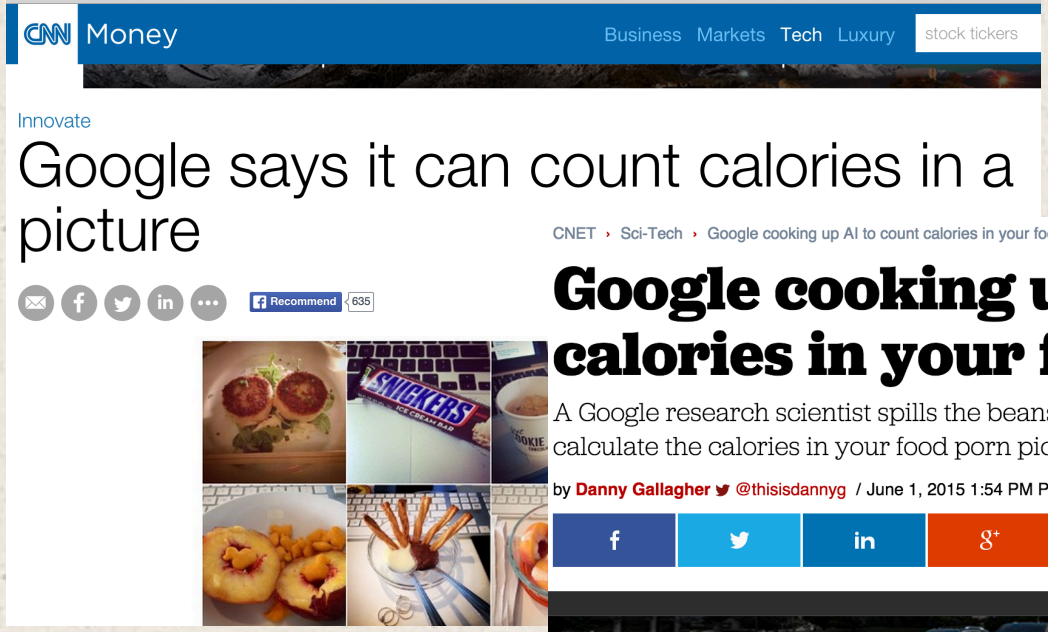
Food Image Analysis

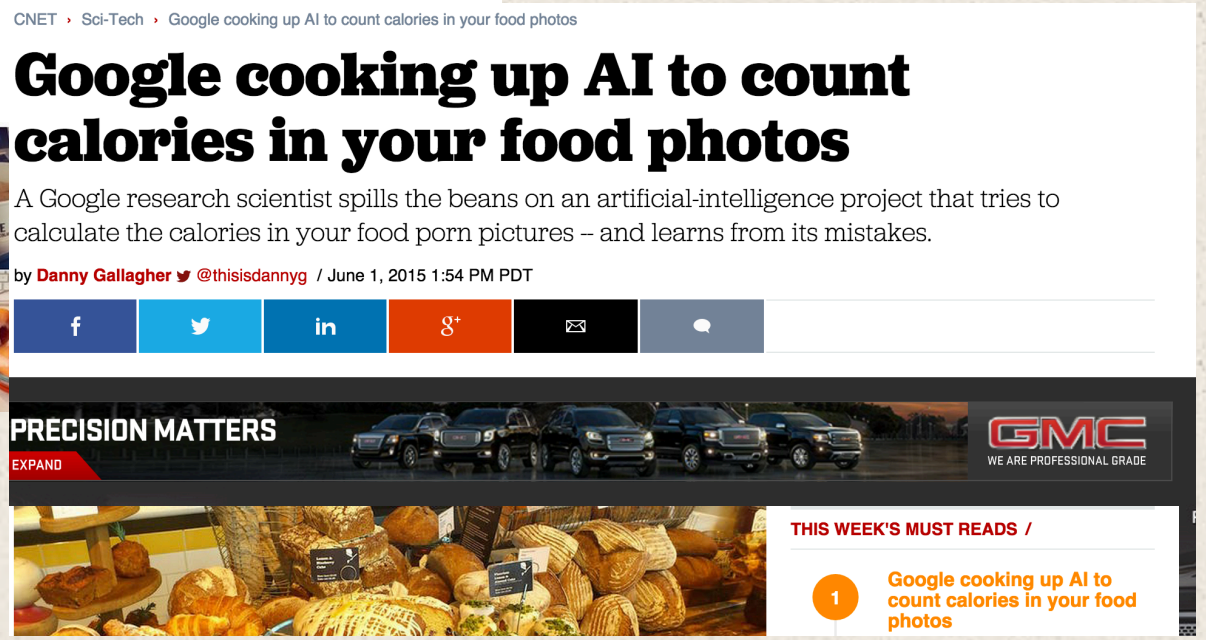
- Health Care, Lifelog, Entertainment (e.g. SNS)
- FoodLog App[1]
  - Assist food recording using image processing

[1] http://app.foodlog.jp/

Food Image Analysis

- Google made a statement about AI to count calories in food porno pictures (May, 2015)





http://money.cnn.com/2015/06/03/technology/google-calories-food/
http://www.cnet.com/news/google-working-on-ai-that-can-count-calories-in-your-food-photos/

Related Work

- Food item recognition from one image
  - Fusing general appearance features[Hoashi+, 2010]
    - Dataset : Japanese foods (collected from web)
    - Used SIFT, Gabor feature, Color feature (MKL-SVM)
  - Original feature specific to food images[Yang+, 2010]
    - Dataset : General USA fast foods (PFID)
    - “pair-wise local features”
  - Deep Learning [Kagaya+, 2014](ours)
    - Dataset: FoodLog Dataset
    - Compared to existing techniques, 10-15% better performance.

Related Work

- Food item recognition from one image
  - Fusing general appearance features[Hoashi+, 2010]
    - Dataset : Japanese foods (collected from web)
    - Used SIFT, Gabor feature, Color feature (MKL-SVM)
  - Original feature specific to food images[Yang+, 2010]
    - Dataset : General USA fast foods (PFID)
    - “pair-wise local features”
  - Deep Learning [Kagaya+, 2014](ours)
    - Dataset: FoodLog Dataset
    - Compared to existing techniques, 10-15% better performance.

[2] H. Hoashi et al., “Image recognition of 85 food categories by feature fusion", IEEE ISM, 2010.
[3] S. Yang et al., “Food recognition using statistics of pairwise local fatures”, CVPR, 2010.
[4] H. Kagaya et al., “Food Detection and Recognition Using Convolutional Neural Network”, ACMMM 2014.

Related Work

- More and more food image dataset
  - PFID[Chen+, 2009]
  - Food-101[Bossard+, 2014]
    - 101 classes * 1,000 images from web
  - UEC Food-256[Kawano+, 2014]
    - 256 classes * 101~729 images from web
  - FoodLog Dataset (ours) [Kagaya+, 2014]

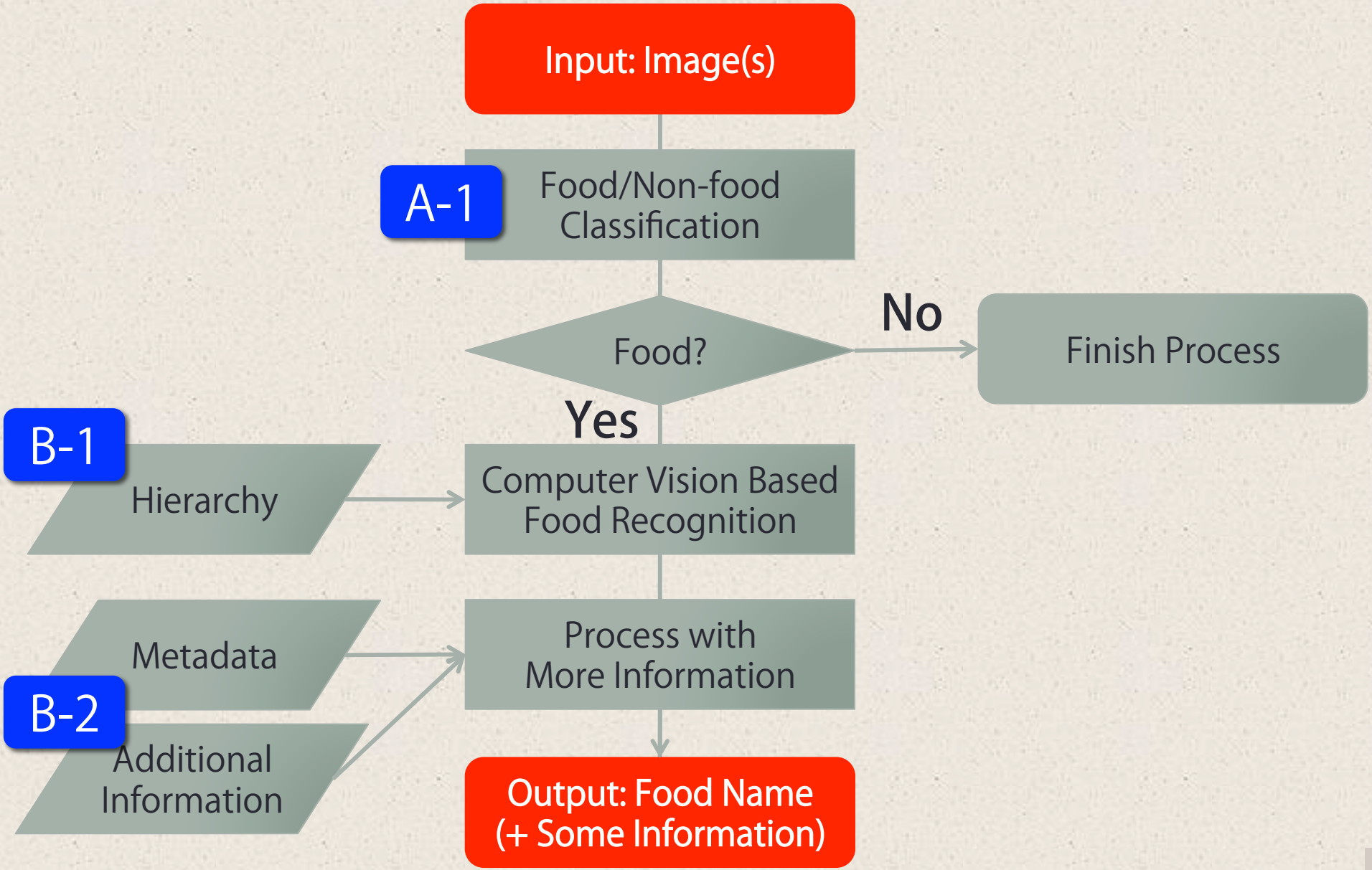
FoodLog App

- Smartphone application for food recording helped using image processing[1]
  - Available in iOS and Android
  - Image retrieval from user's past food record using color feature with spatial pyramid

✔ Personal trends is very important

✘ Many manual processes

Proposed Framework



Problem and Contribution

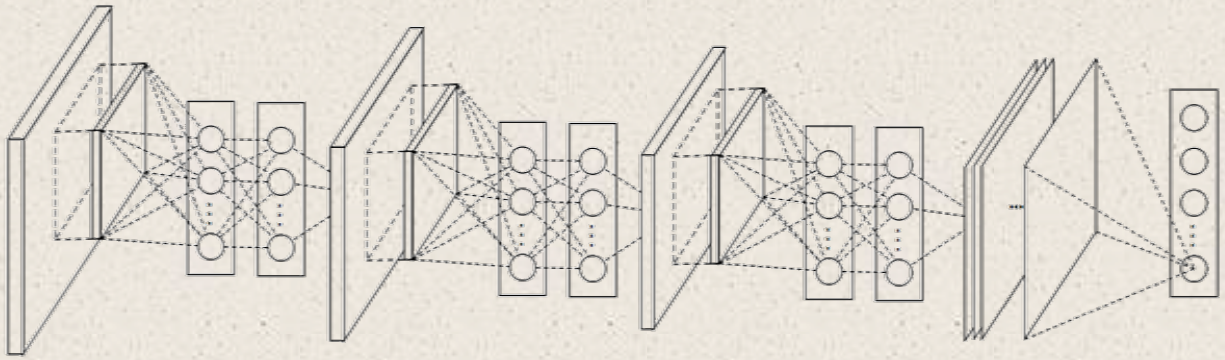
- Food/Non-Food Classification
  - judges whether an image is a food or not

Contribution

- Build two novel datasets
- Evaluate the performance of a deep CNN for food/non-food recognition in the experiment within and across the same dataset.

Classification with CNN-NIN

- Convolutional Neural Network (CNN)[14]
  - state-of-the-art for general image classification
  - Also effective for food recognition[4]
  - Network in Network (NIN)[15]
    - We employ NIN as the model because it performs better than usual AlexNet and it is memory efficient
    - mlpconv layers and global average pooling



[14] Y. LeCun et al., "Gradient-based learning applied to document recognition", Proc. IEEE, 1998.
[15] M. Lin et al., "Network in Network", ICLR, 2014.

Datasets

- Build two datasets for food/non-food classification
  - ① **Instagram Food/Non-Food Dataset (Insta.)**
    - Collect images tagged with "#food" from Instagram[12]
    - Motivated by another application: filtering SNS
  - ② **Food-101/Caltech-256 Dataset (F/C)**
    - Collect images in Food-101[10] as food class and Caltech-256[13] (except for some food images) as non-food class

Classess	#images (Insta.)	#images (F/C)
Food	4,230	25,250
Non-Food	5,428	28,322
Unspecified	342	-
Total	10,000	53,572

[12] http://instagram.com/
[13] G. Griffin et al., "Caltech-256 object category dataset", Caltech Technical Report, 2006.

Classification Protocol

No	Training	Testing
1	Instagram	Instagram
2	Food-101/Caltech-256	Food-101/Caltech-256
3	Food-101/Caltech-256	Instagram
4	Instagram	Food-101/Caltech-256

- No. 1 + 2 (Within a dataset)
  - change train/test ratio from 0.5 to 0.9 and average the result of five trial for each ratio.

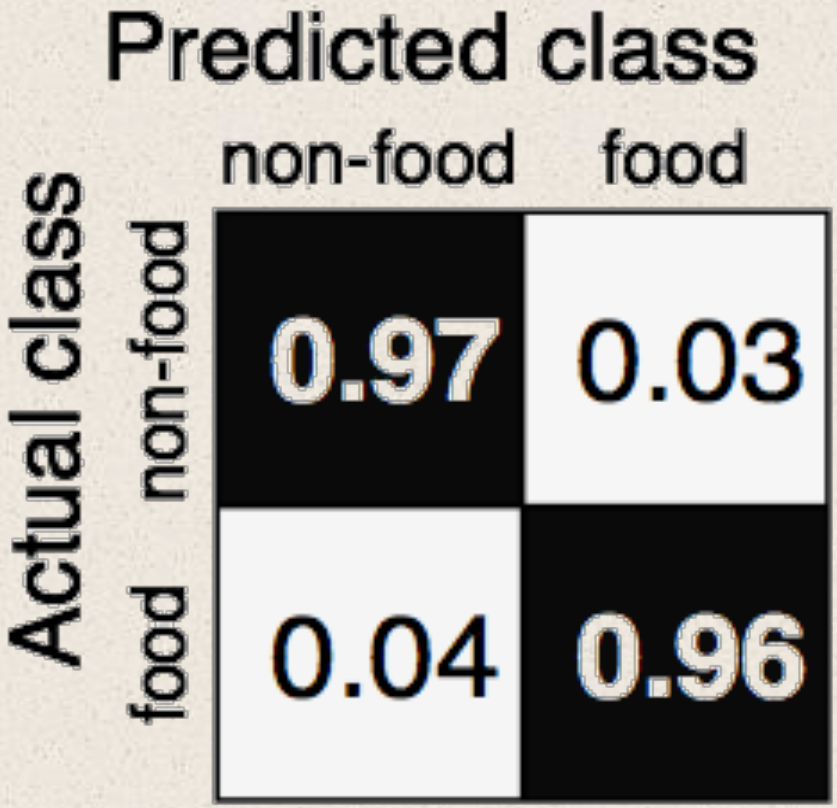
- No. 3 + 4 (Across datasets)
  - is the model general?

Within one dataset

Train: Insta. / Test: Insta.

Accuracy: 94.8%

Train: F/C / Test: F/C



Accuracy: 96.4%

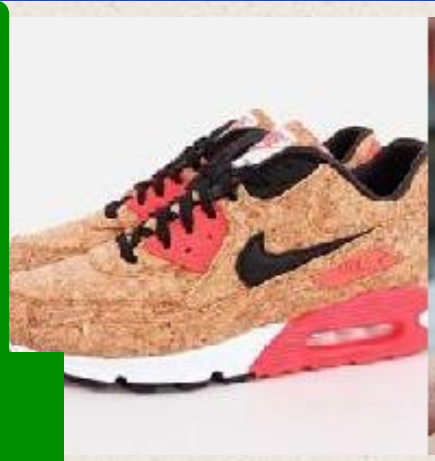
Train / Test ratio is 0.8 / 0.2

Wrong Cases (Instagram)

Ground Truth: Food

Package

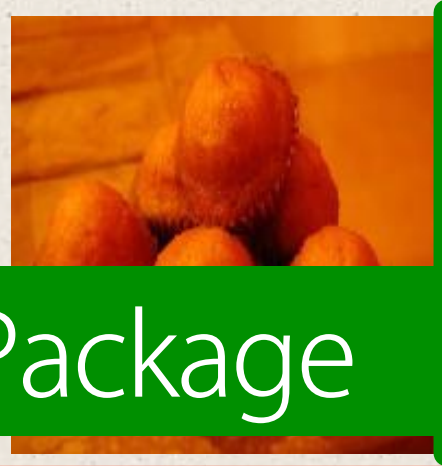
Ground Truth: Non-food



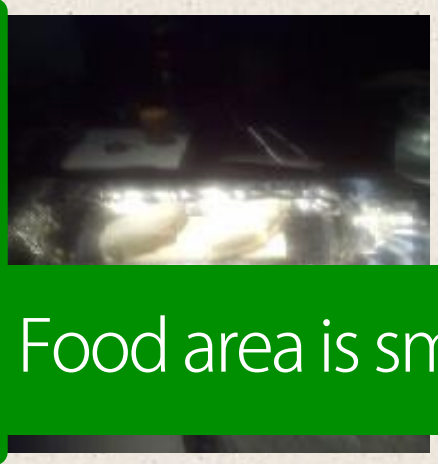
Flower

Wrong Cases (Food/Caltech)

Ground Truth: Food



Package



Food area is small

Ground Truth: Non-food

Living thing

Across datasets

Train: F/C / Test: Insta.

Accuracy: 91.5%

Train: Insta. / Test: F/C

Accuracy: 90.6%

Wrong Cases (across)

*correct in "within" but wrong in "across"

Ground Truth: Food

Ground Truth: Non-food

Comparison with The Previous Methods

- For comparison, we utilized the dataset in [Kagaya+, 2014]
  - Baseline: SIFT/Circle feature/Color feature + SVM[19]
    - It is used the web version of FoodLog

Method	Accuracy
Baseline[19]	89.7±0.73%
Kagaya+, 2014	93.8±1.39%
This study (CNN-NIN)	**99.1± 0.81%**

[19] K. Kitamura et al., "Food log by analyzing food images", ACMMM, 2008

Conclusion & Future Work

- Highly accurate food/non-food image classification for pre-process of food imge analysis
  - Built novel datasets (we will get them open publicly)
  - Used a deep convnet as a state-of-the-art and conducted some experiments
  - Evaluated the performance using two datasets
    - within scheme: 95-96%
    - across scheme: 90-91%
- Future: we really introduce this study to the current FoodLog system

Thank you for your kind attention!

Any question?