Food Recognition Using Statistics of Pairwise Local Features

Shulin Yang
University of Washington

Mei Chen, Dean Pomerleau, Rahul Sukthankar
Intel Labs Pittsburgh

Motivation

1. A challenging computer vision problem compared to other forms of object recognition:

- Large variation in appearance
- Deformable internal structure

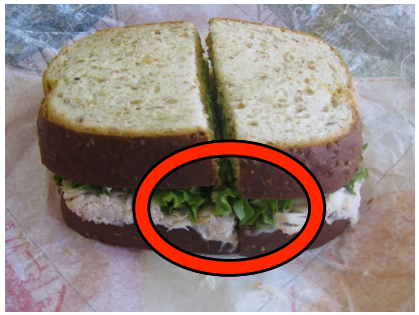
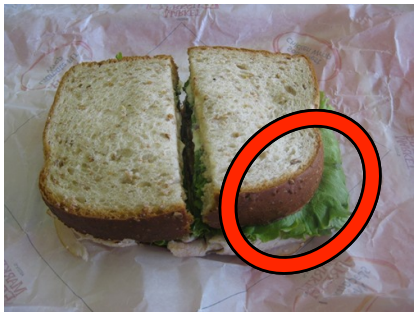
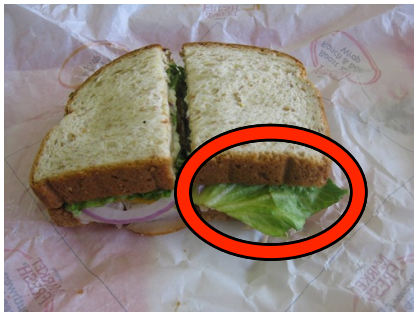
2. Useful for health and wellness applications such as diet monitoring, nutrition research, obesity control.

Challenges for Food Recognition

Main challenges of food recognition

- How to identify food ingredients in images

- How to represent and quantify the highly varied spatial relationships of the ingredients

Related Work

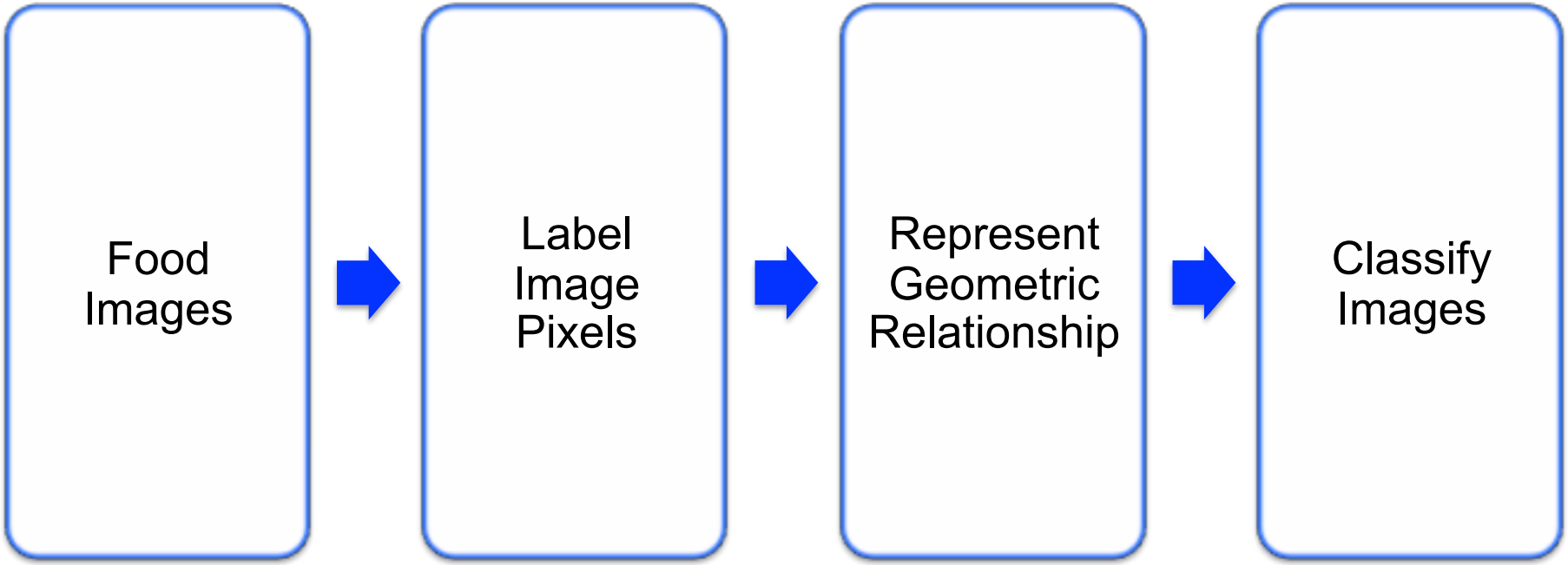
Bolle, et al. (1996): VeggieVision- an automatic produce ID to help with the produce checkout process

Shroff, et al. (2008): A wearable computing system to recognize food for calorie monitoring

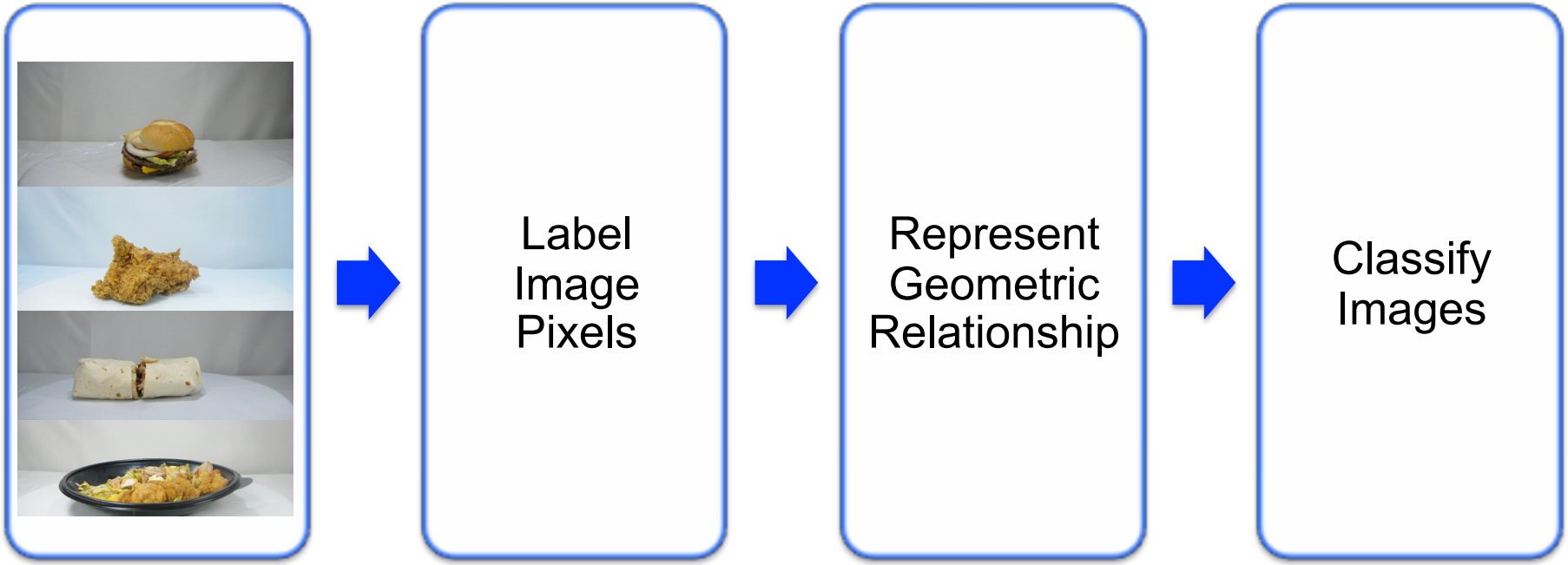
Yang, et al. (2009): PFID- a dataset for food recognition

Wu and Yang (2009): Analyzed eating videos to recognize food 160 food items and estimate caloric intake

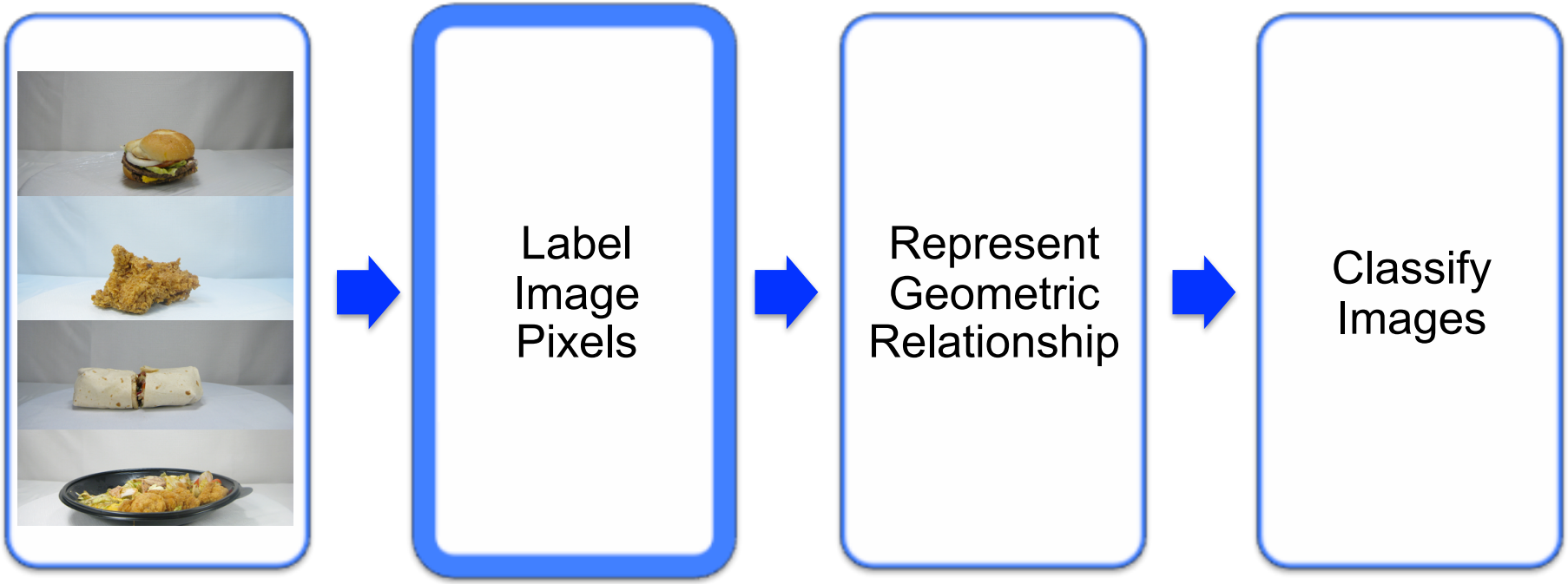
Our Framework



Our Framework

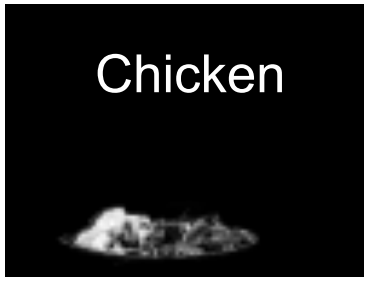


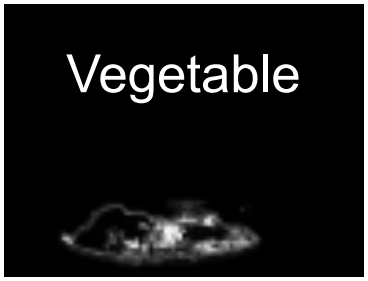
Our Framework

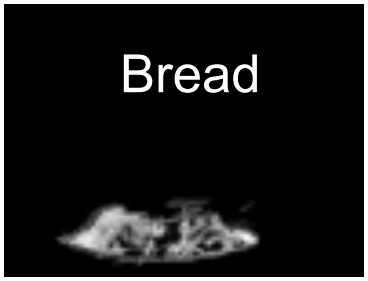


Labeling Image Pixels

- Nine food ingredients: bread, beef, chicken, pork, vegetable, tomato, cheese, egg, others
- Soft Labels: a vector of likelihoods that a pixel belongs to each food ingredient type

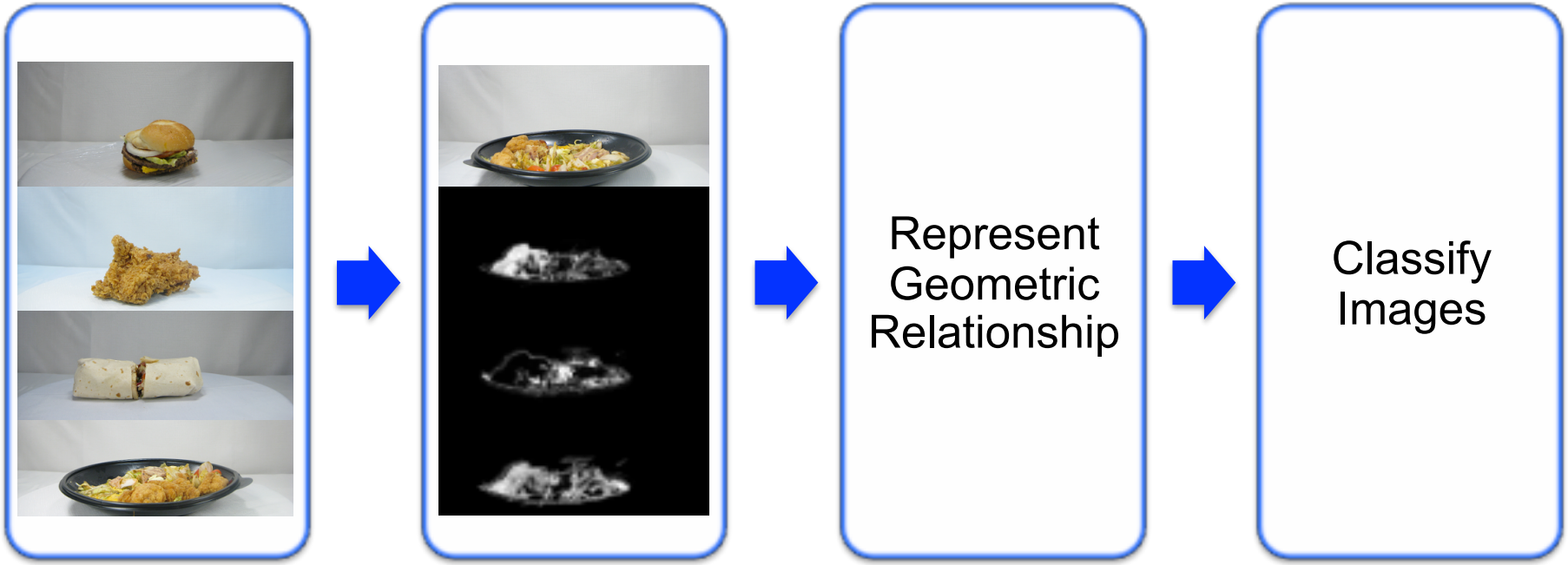






- Soft labels generated using Semantic Texton Forest (STF) algorithm (Shotton, et al. CVPR2008)

Our Framework



Our Framework

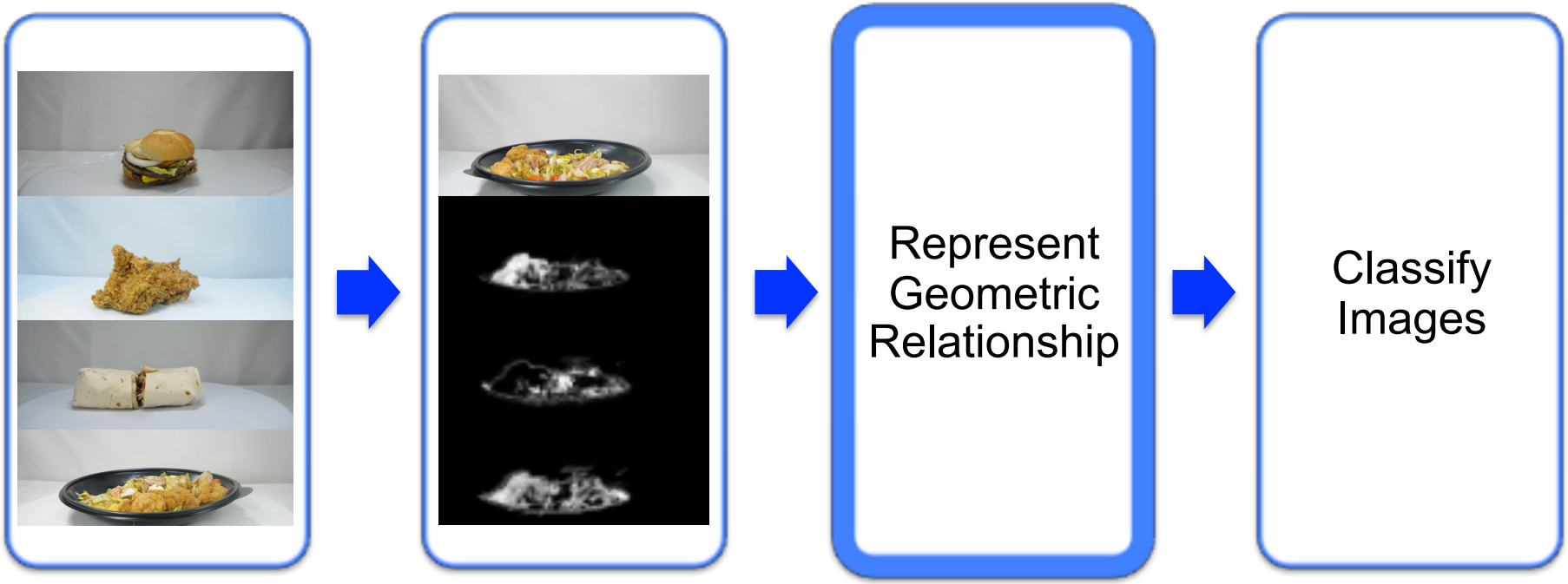
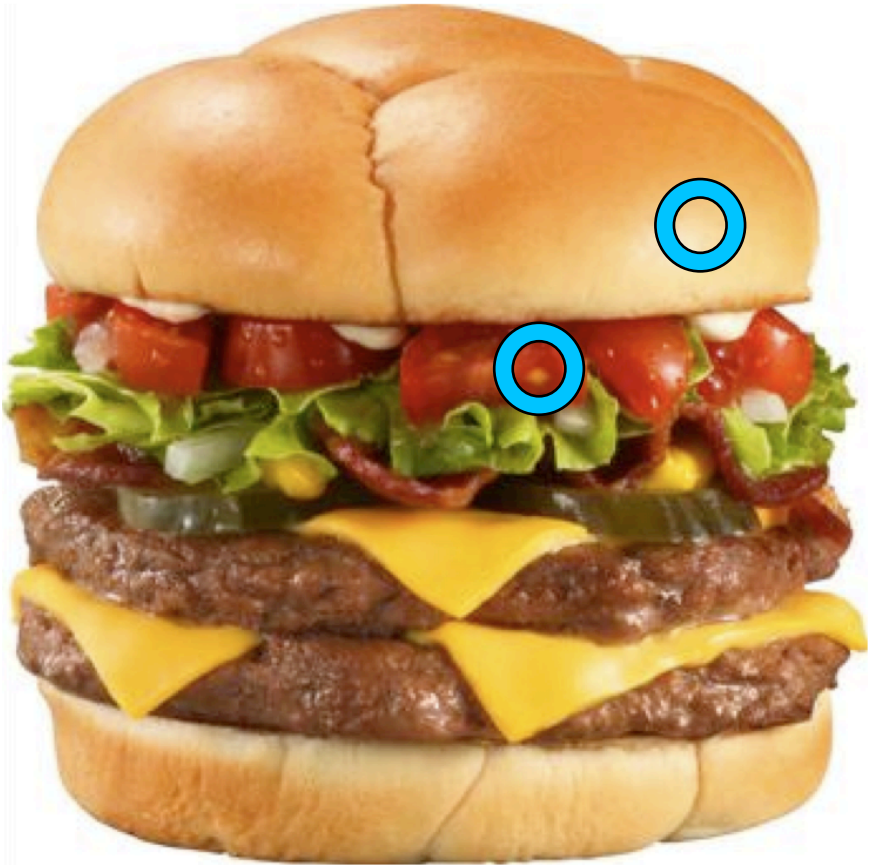


Image Representation – Pairwise Features

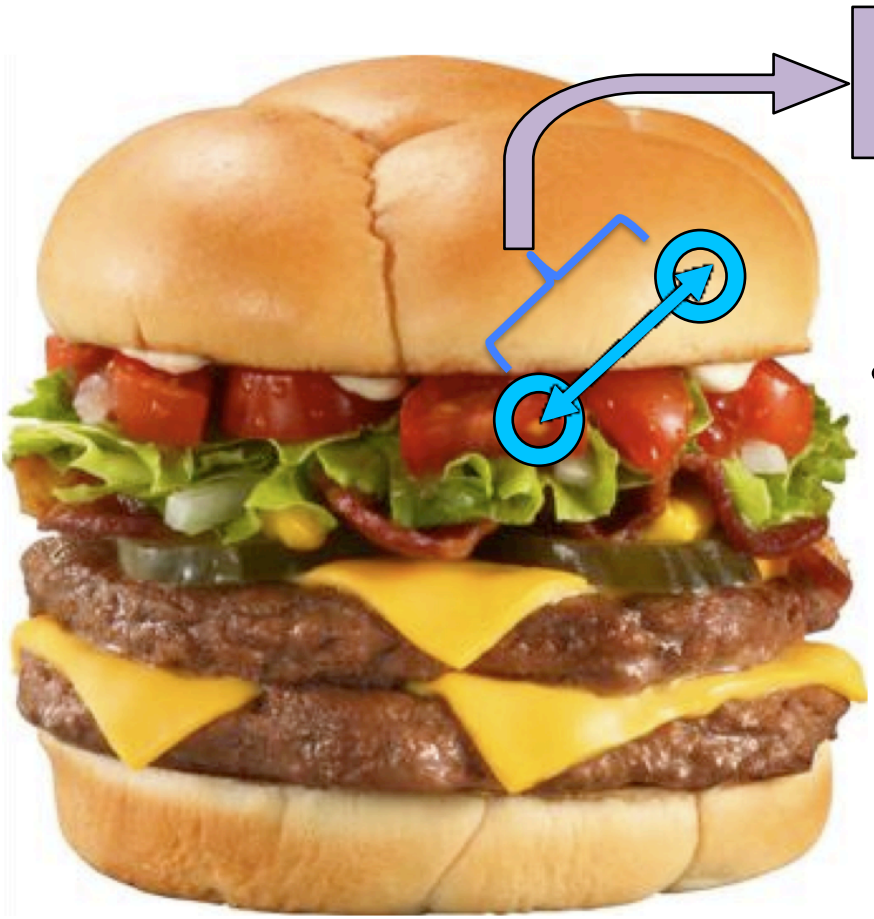
- Using pairwise feature distribution to capture the spatial relationship among pixels of food ingredients

Image Representation – Pairwise Features



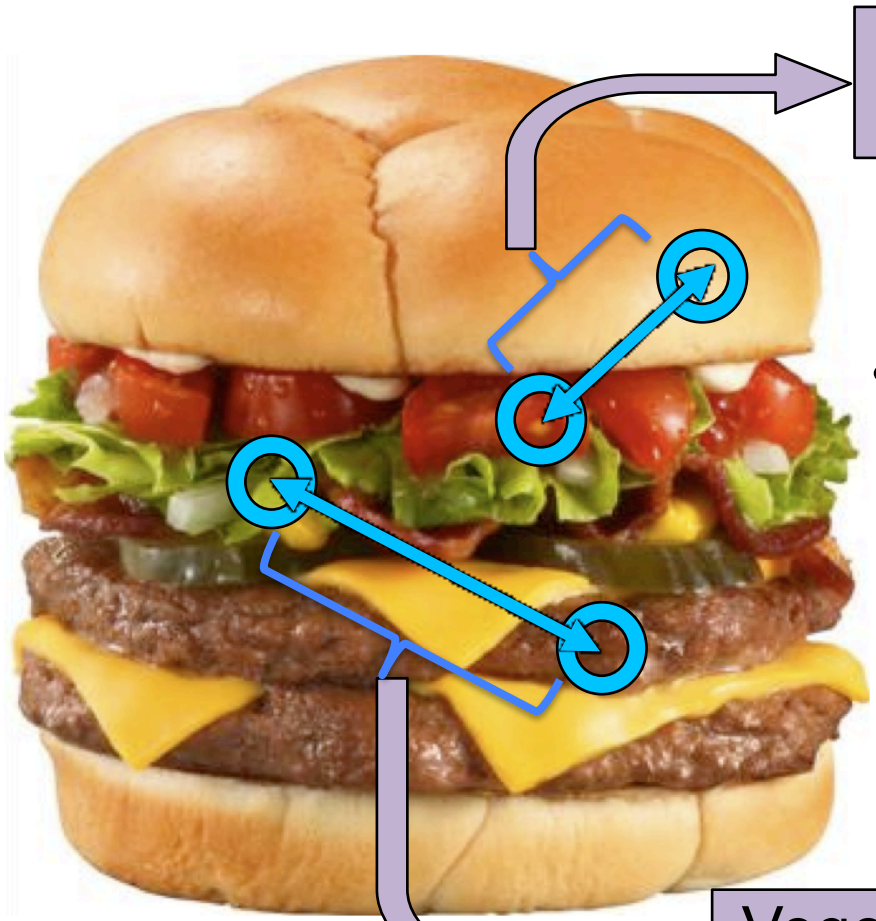
- Using pairwise feature distribution to capture the spatial relationship among pixels of food ingredients

Image Representation – Pairwise Features



- Using pairwise feature distribution to capture the spatial relationship among pixels of food ingredients

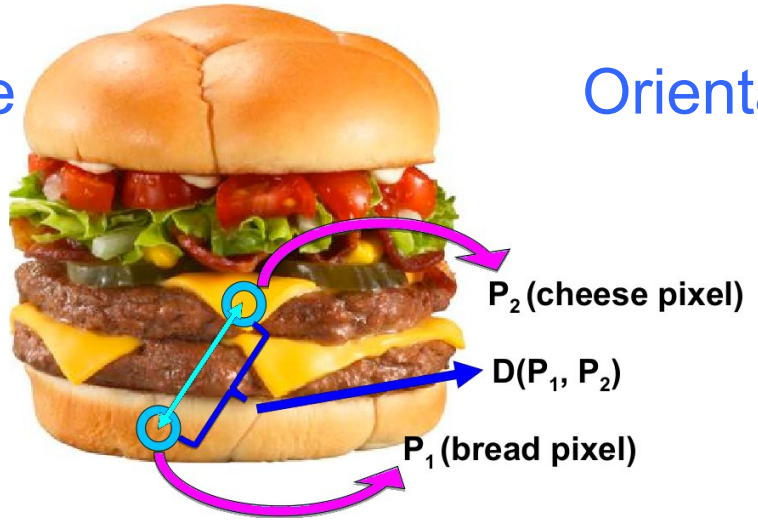
Image Representation – Pairwise Features



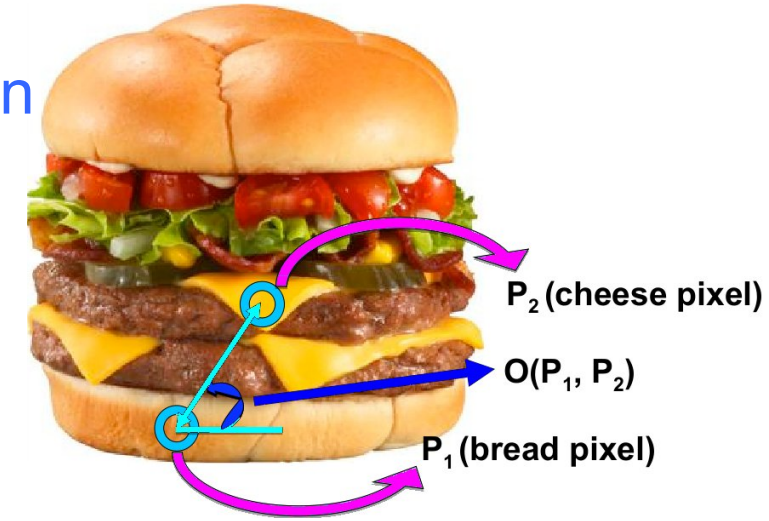
- Using pairwise feature distribution to capture the spatial relationship among pixels of food ingredients

Image Representation – Pairwise Features

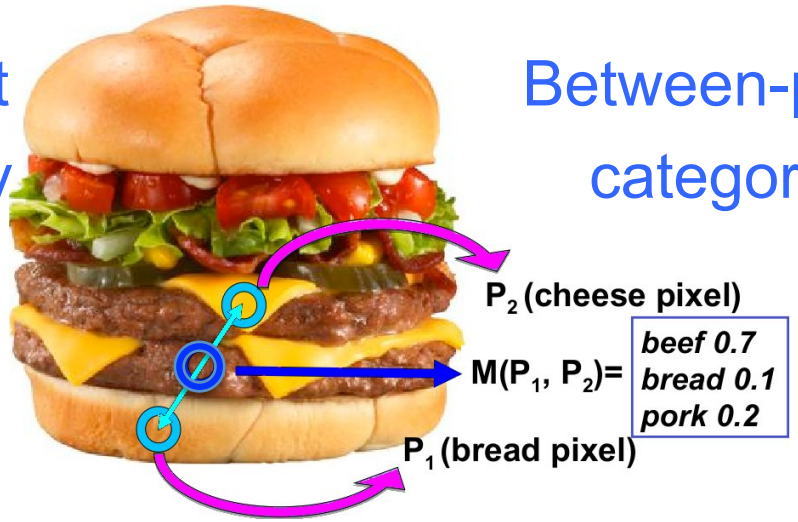
Distance



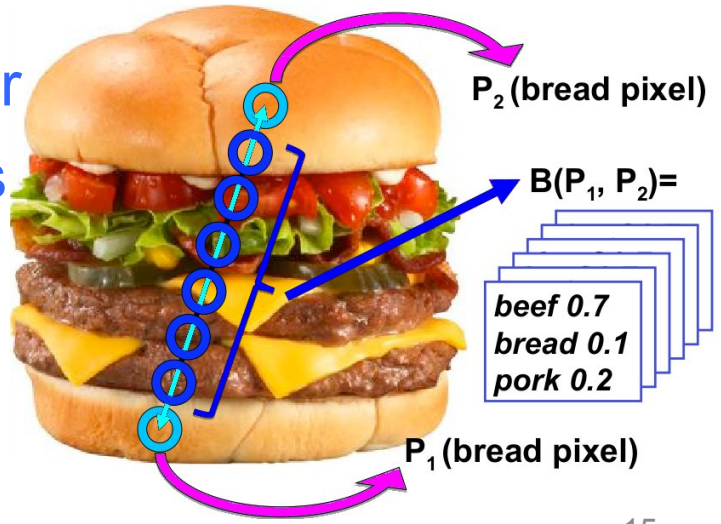
Orientation



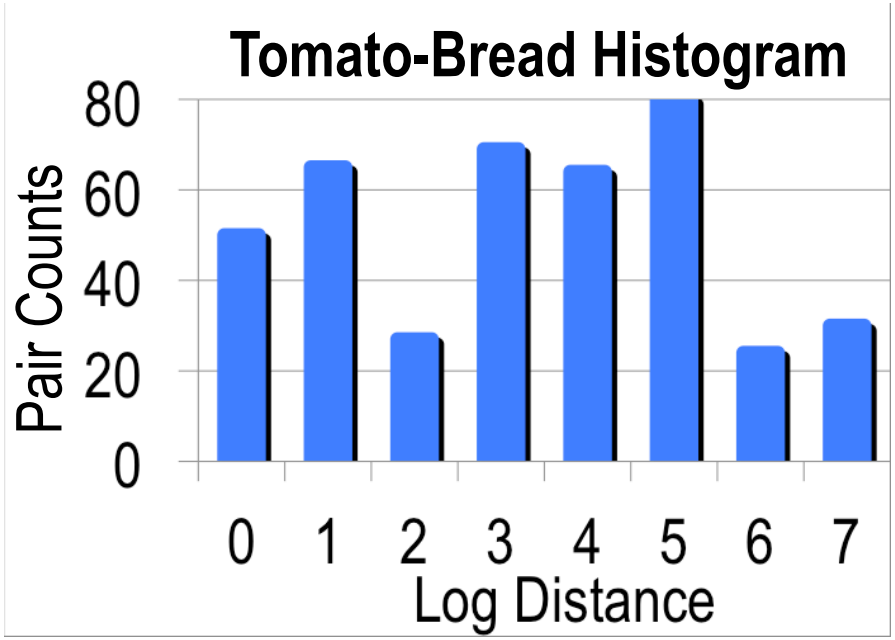
Midpoint category



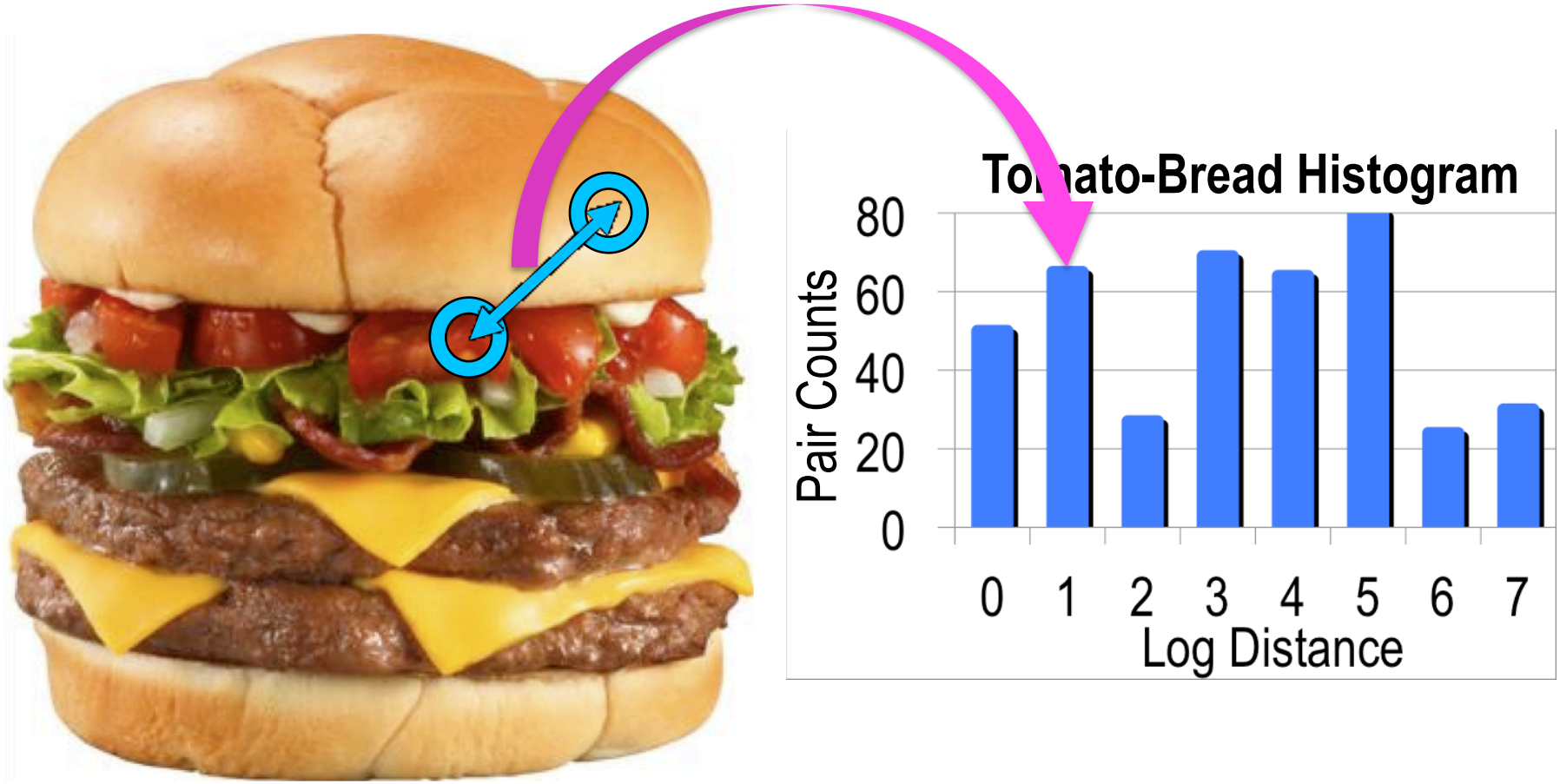
Between-pair categories



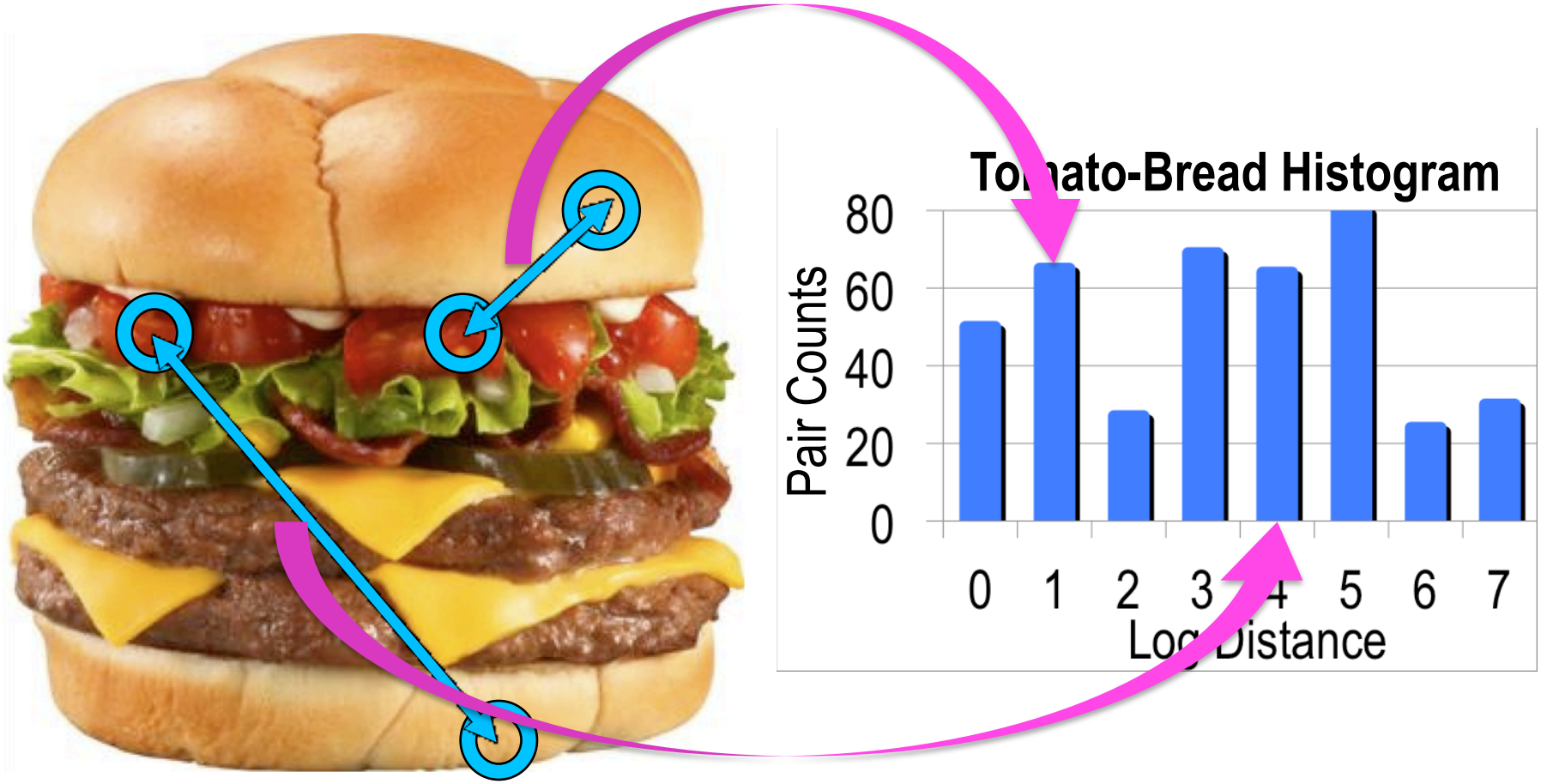
Geometric Relationship Representation



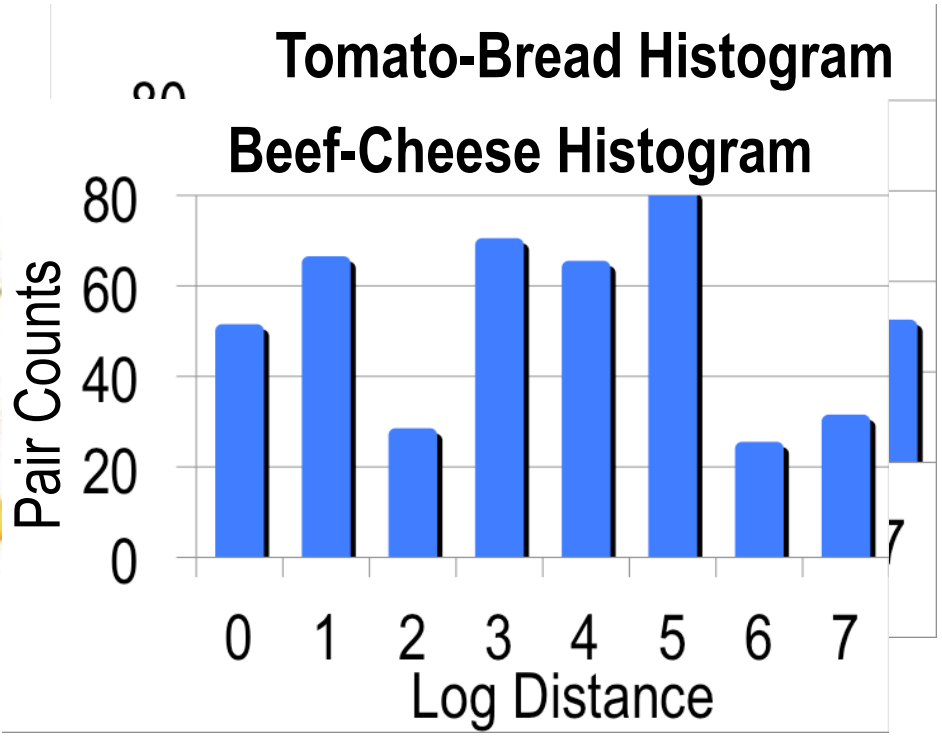
Geometric Relationship Representation



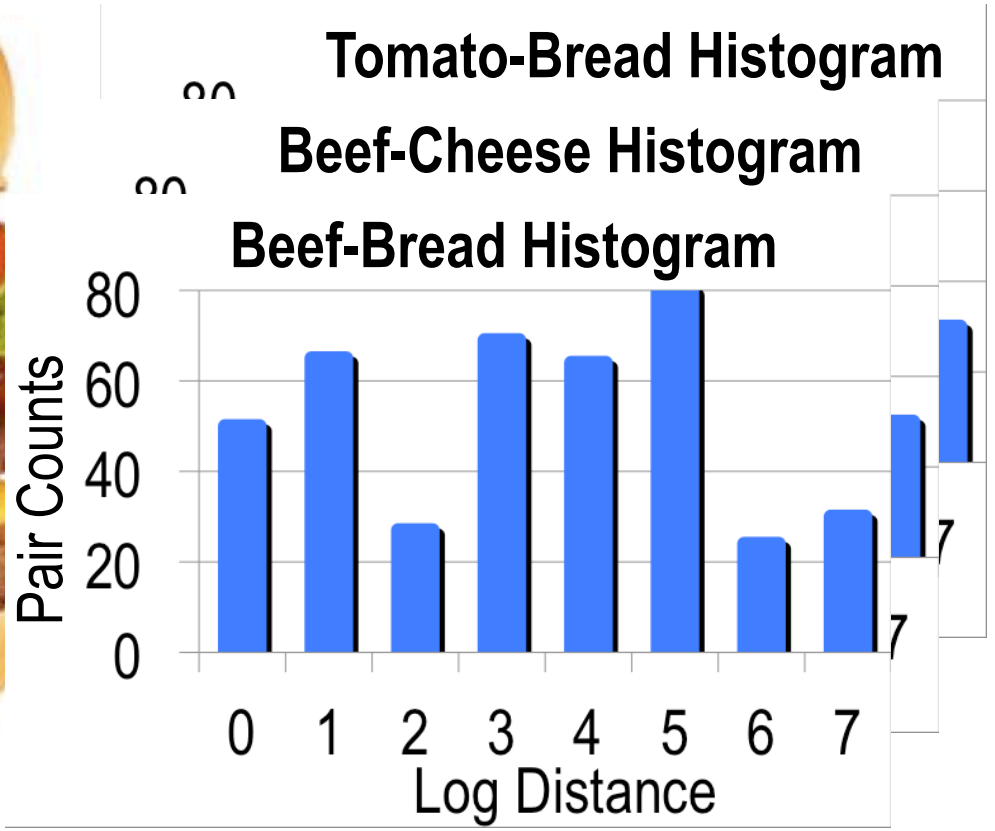
Geometric Relationship Representation



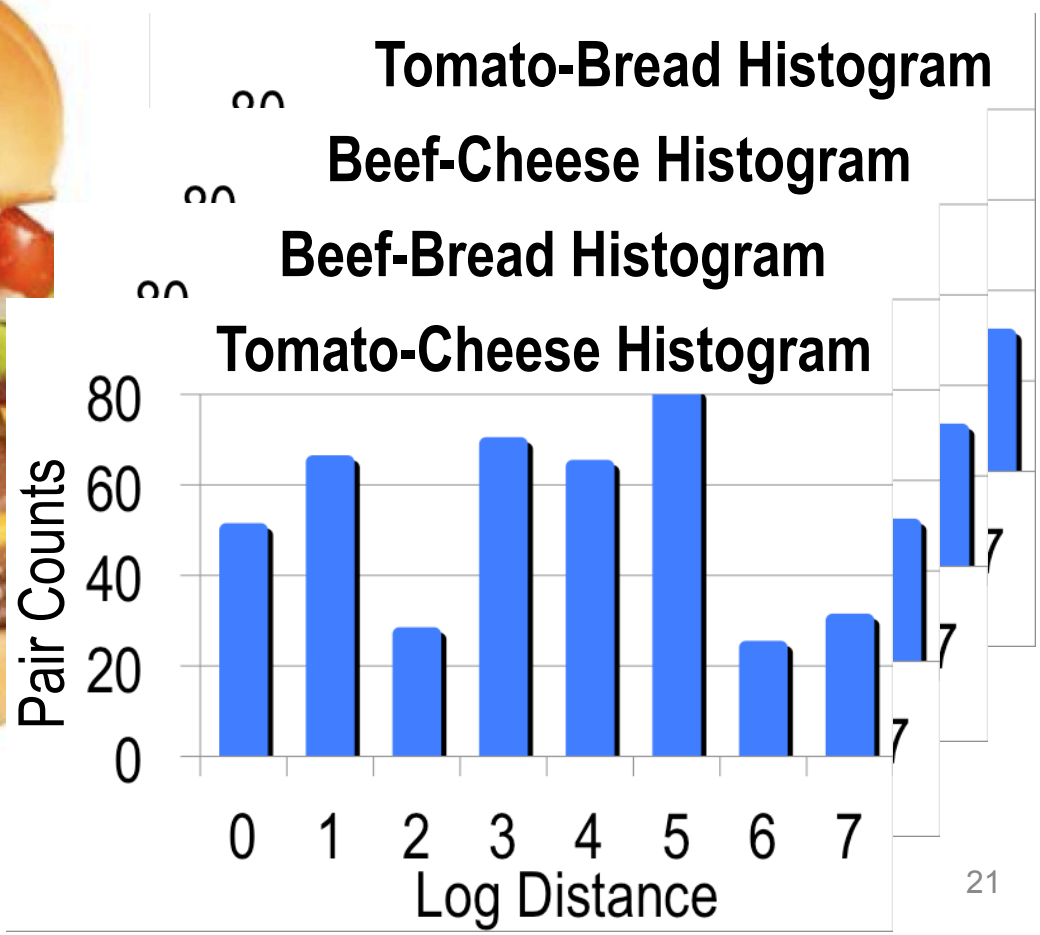
Geometric Relationship Representation



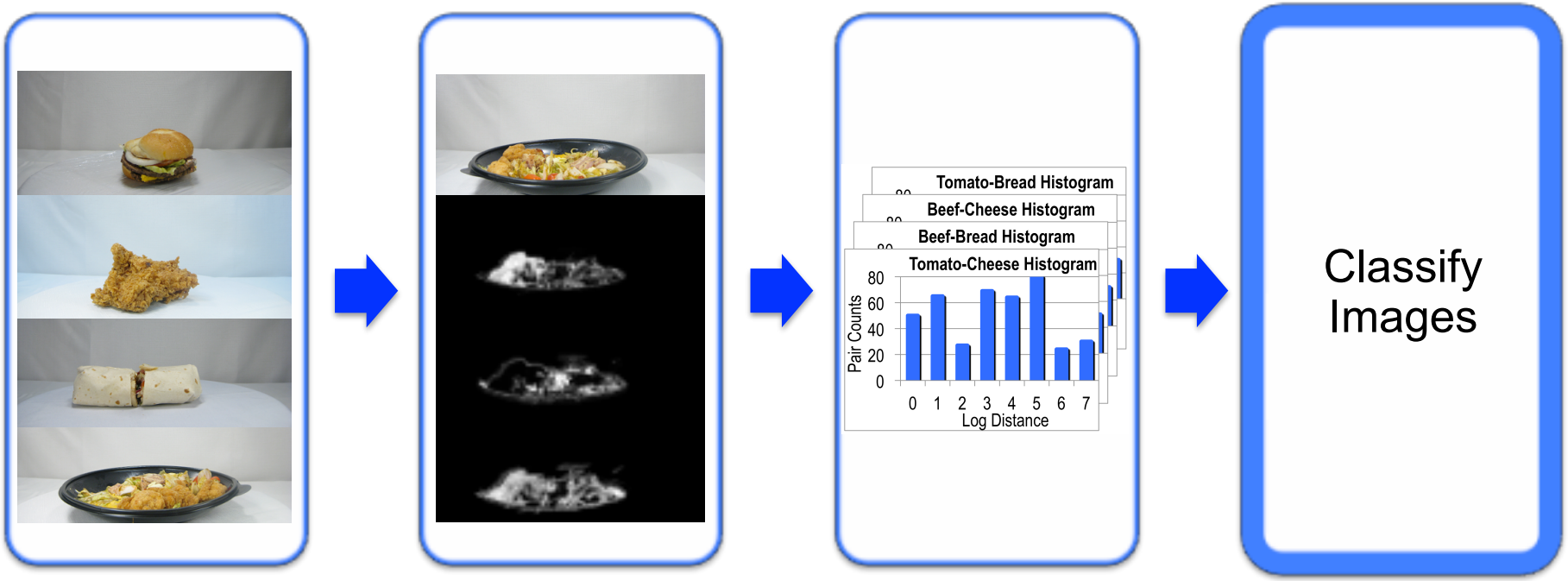
Geometric Relationship Representation



Geometric Relationship Representation



Our Framework



Classification

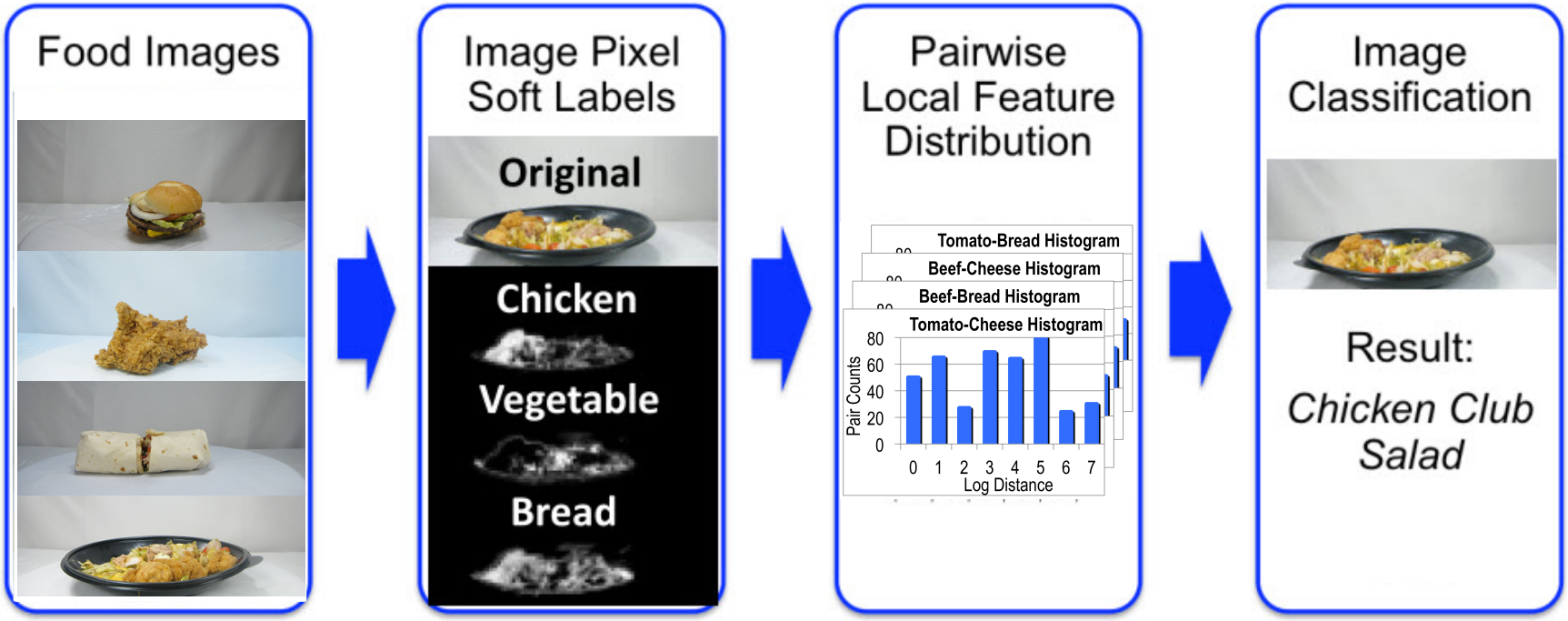
- SVM using pre-computed χ^2 kernel

  χ^2 distance for histogram dissimilarity

$$d_{\chi^2}(x, y) = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}$$

- Image dissimilarity is the sum of χ^2 distances of all their histograms

The Complete Framework



Experiment – Dataset

Dataset: PFID (Pittsburgh Fast-Food Image Dataset)

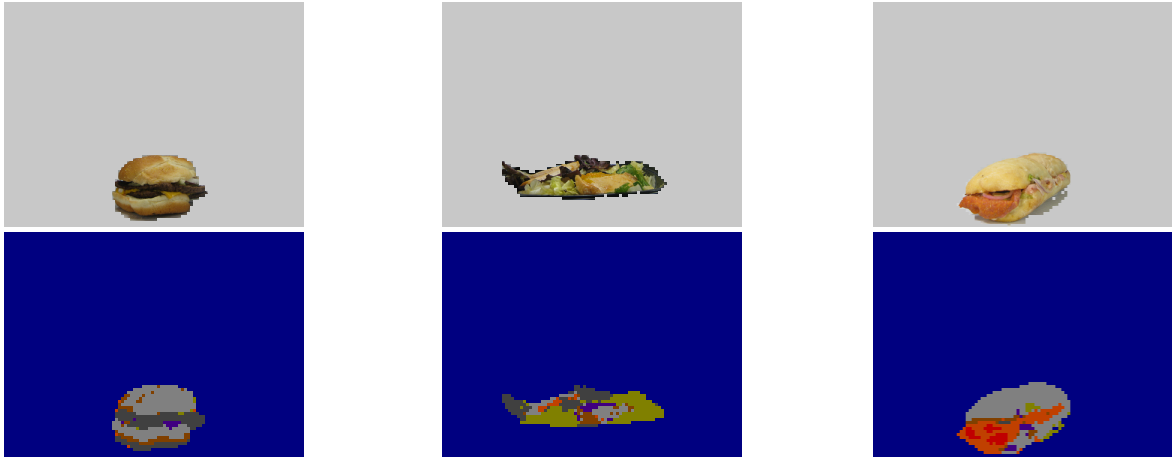
- Available at http://pfid.intel-research.net/baseline.php
- Fast food from 13 chain restaurants, such as Spicy Chicken Sandwich from Wendy's, and Italian B.M.T. Sandwich from Subway
- 61 categories × 3 instances × 6 views = 1098 in total
- Background segmented

Experiment – Soft Labels

Soft label all images using STF

- Train on 16 manually labeled images

- Generate soft labels for the other 1098 images



Experiment – Classification

Training and testing data

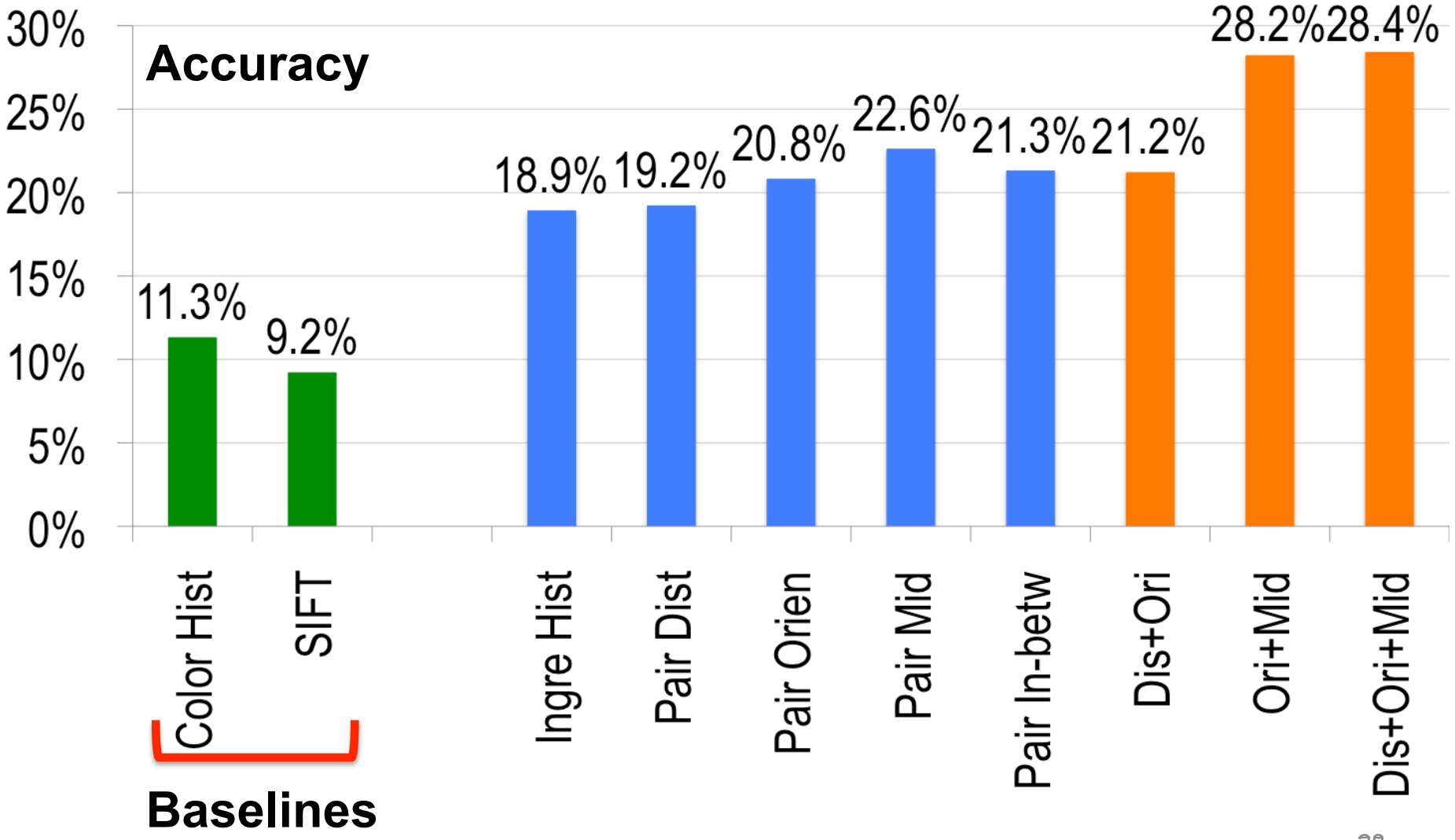
- 732 (61×2×6) images for training
- 366 (61×1×6) images for testing
- Three-fold cross-validation

Two classification tasks

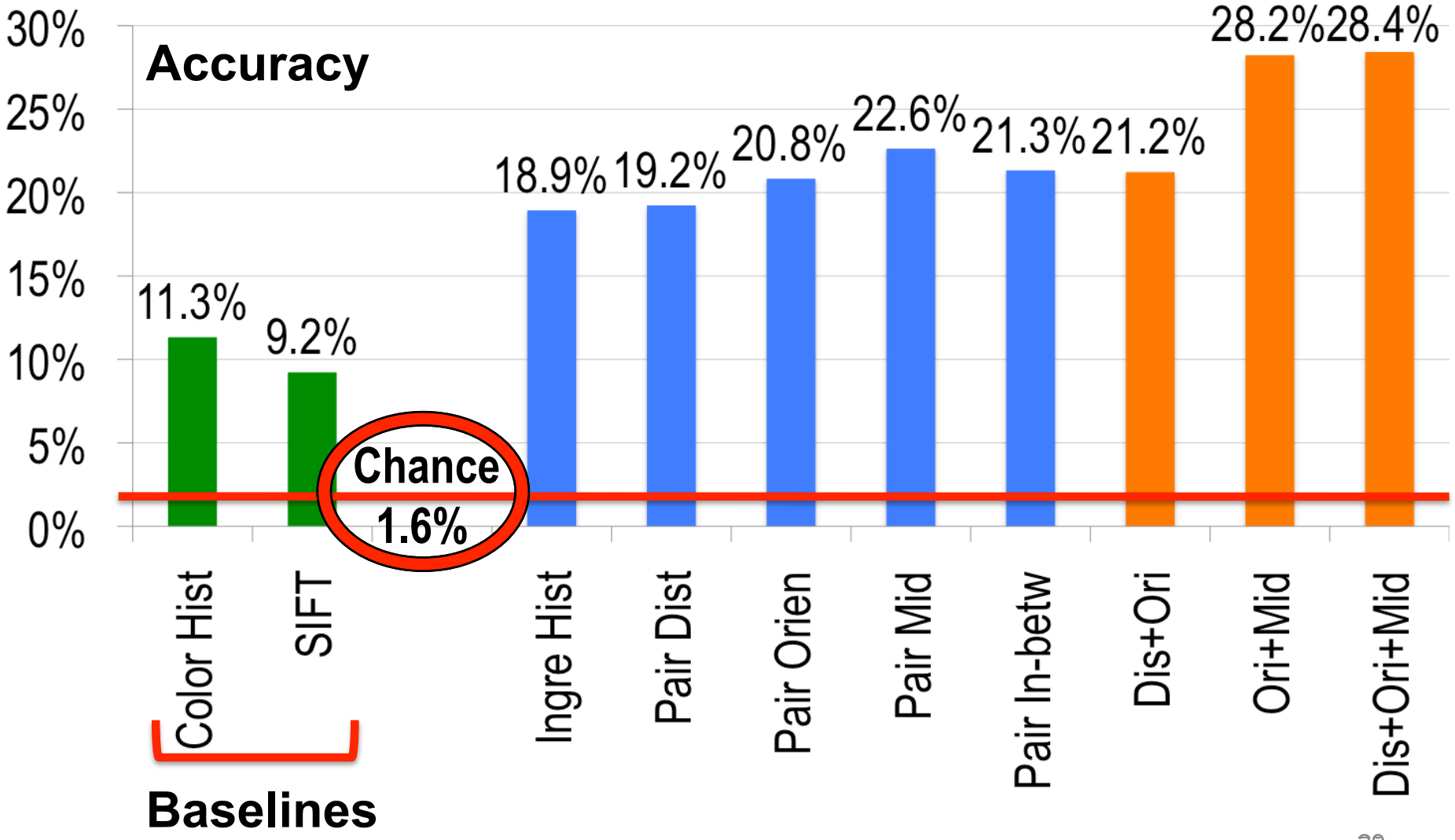
- On 61 food categories
- On 7 major groups of food

Sandwich		Salad		Taco	
Chicken		Bread		Donut	
Bagel					

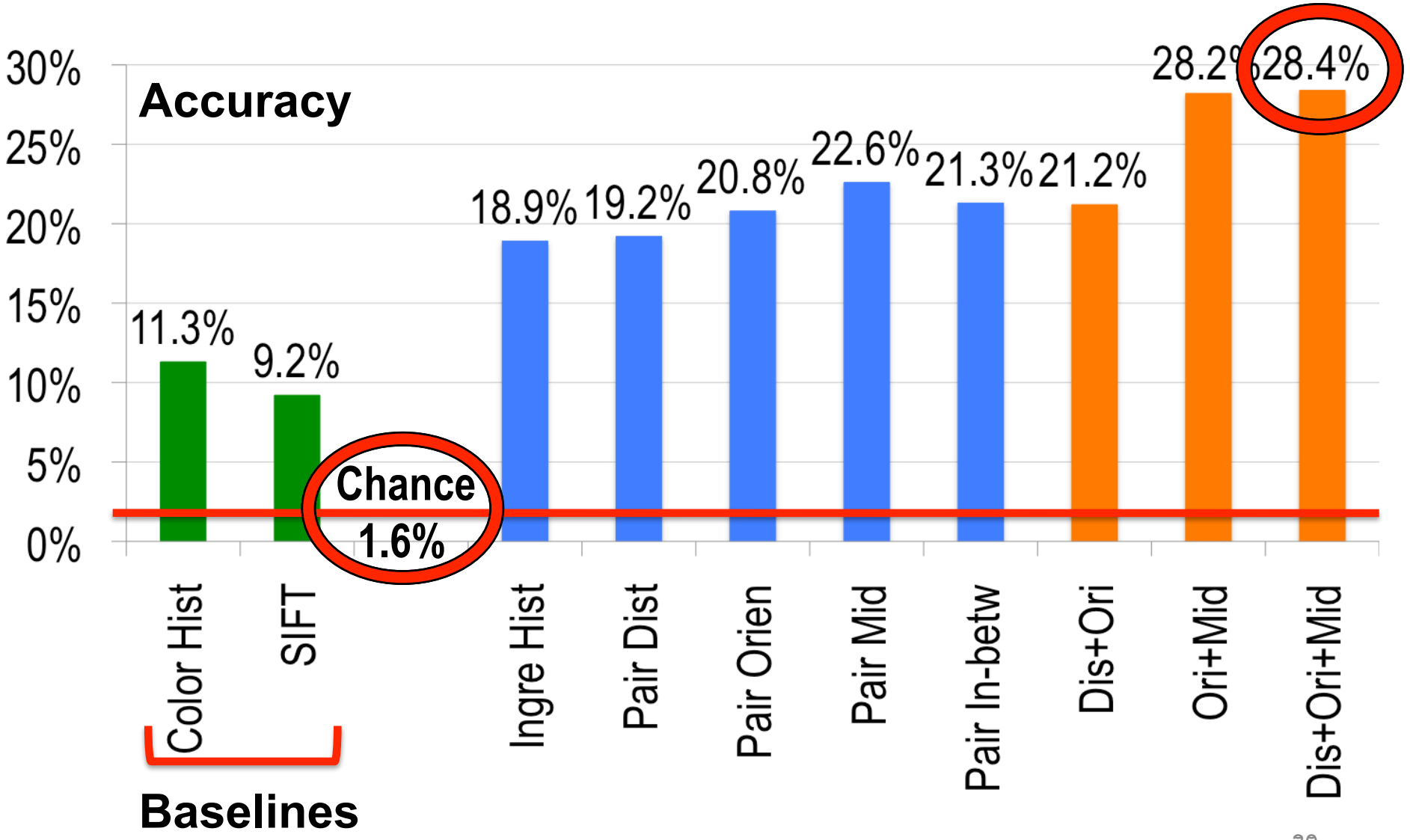
Classification Results on 61 Categories



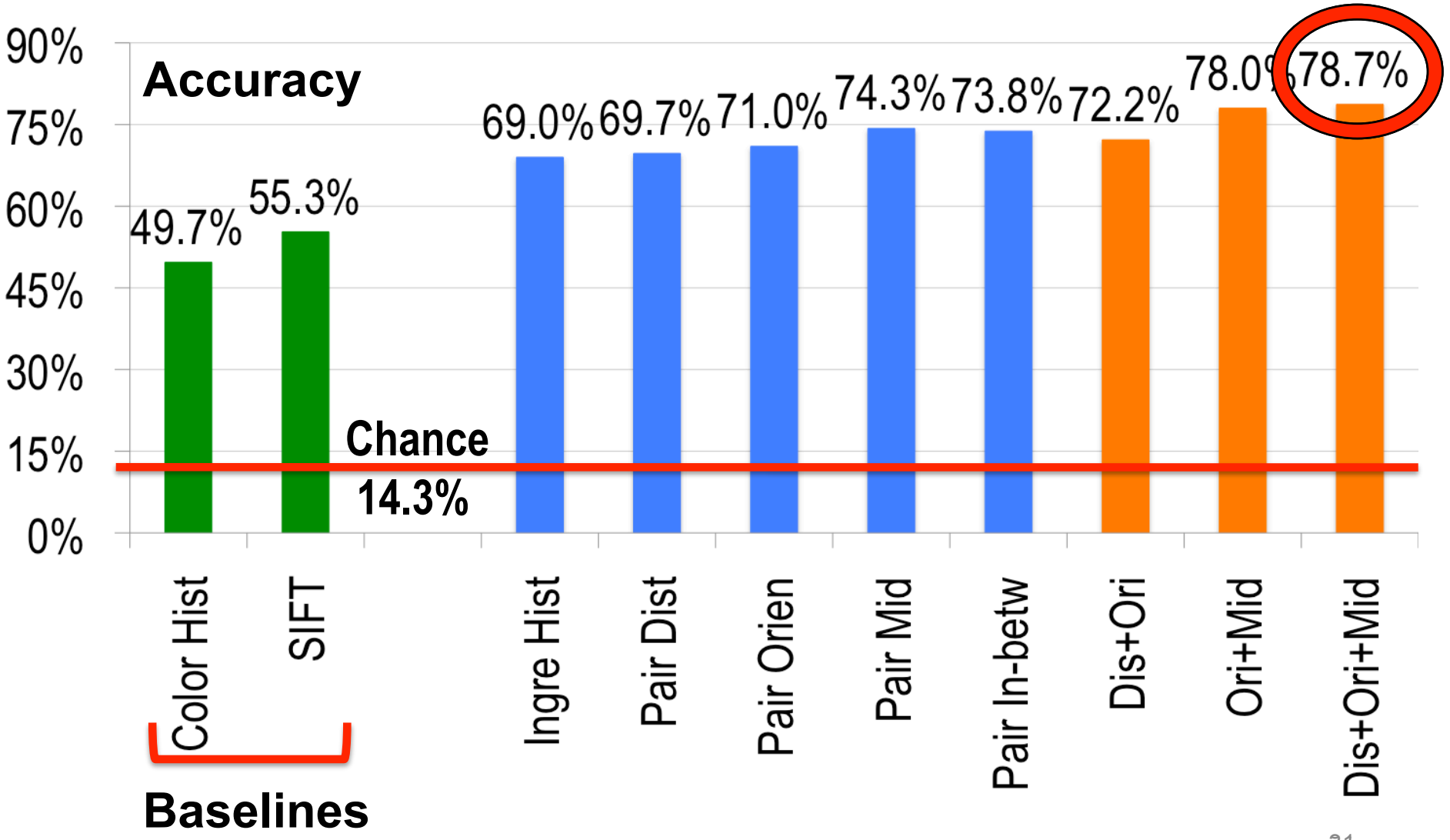
Classification Results on 61 Categories



Classification Results on 61 Categories



Classification Results on 7 Major Groups



Pairwise Features for Scene Recognition

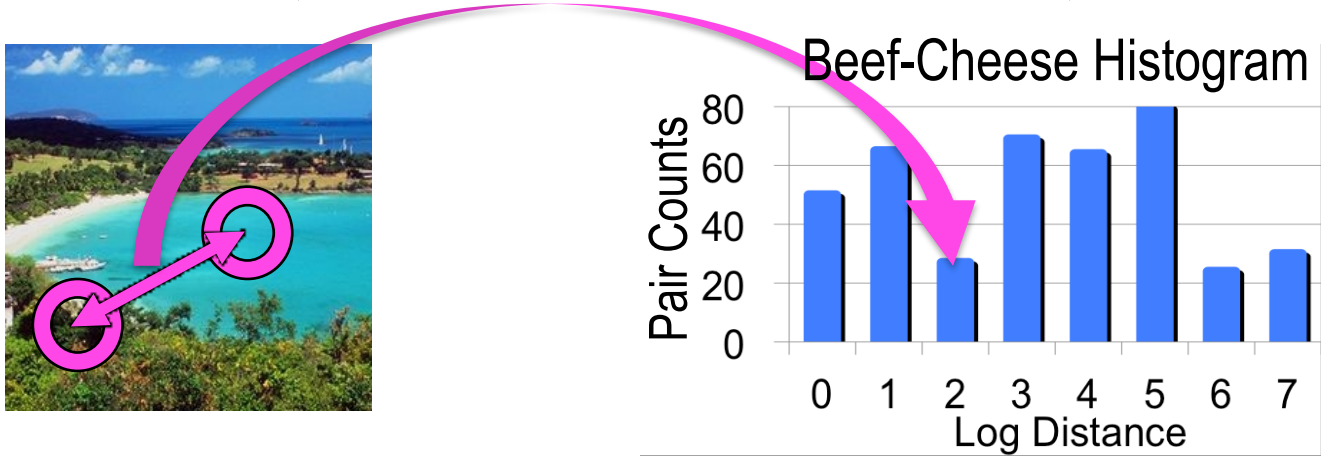
Scene images are similar to the food images:

- Composed of some basic scene "ingredients"
- Geometric relationships between ingredients important but highly variable

Test on the MIT outdoor scene dataset
8 categories including coast, mountain, highways, etc.

Pairwise Feature for Scene Recognition

- 10 basic ingredients for outdoor scenes: sky, water, rocks, trees, road, building, etc. Manually labeled 24 images for soft labeling training
- Combine the Pairwise features (Distance and Orientation) with GIST features (Oliva and Torralba, IJCV2001).



- Preliminary result shows that the accuracy is boosted by 3% from 85% to 88%

Conclusion

1. Developed new technique to capture shape regularities using geometric statistics computed over pairwise features

2. Applied the technique to food recognition problem

3. Outperformed baseline approaches significantly on food recognition

Future work:

- Design and test new pairwise feature
- Generalize the approach for other domains
- Use it for real food recognition applications

Acknowledgements

We would like to thank:

Professor Linda Shapiro and the Multimedia Group in UW for suggestions and advice,

Lei Yang for providing SIFT+SVM baseline code,

The PFID project for help with the food recognition dataset.

Q&A?