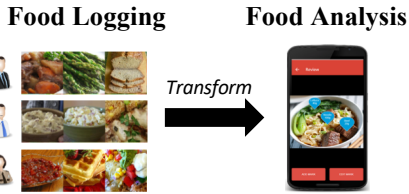


Lei Meng¹, Long Chen², Xun Yang¹, Hanwang Zhang³, ¹National University of Singapore ²Zhejiang University
 Dacheng Tao⁴, Chunyan Miao³, Tat-Seng Chua¹ ³Nanyang Technological University ⁴The University of Sydney

Food Recognition in Visual Food Logging



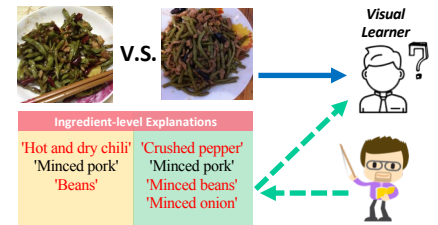
- From food logging to personalized services
- Food recognition from images is a key

Visual v.s. Non-visual



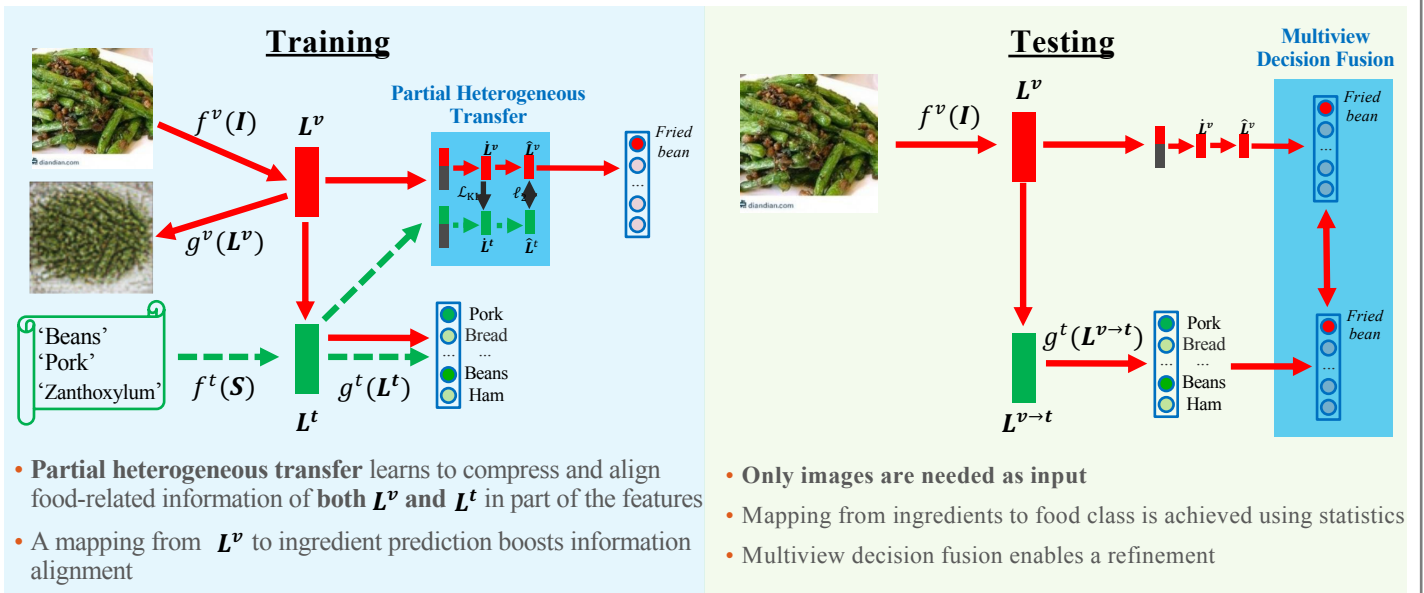
- Different food may have *subtle difference* hidden in minor ingredients, sauces, and cooking methods

Using Ingredients as Privileged Information



- Use ingredients to distinguish food
- Align visual factors to ingredients

Food Recognition Under LUPI Paradigm



- **Partial heterogeneous transfer** learns to compress and align food-related information of both L^v and L^t in part of the features
- A mapping from L^v to ingredient prediction boosts information alignment

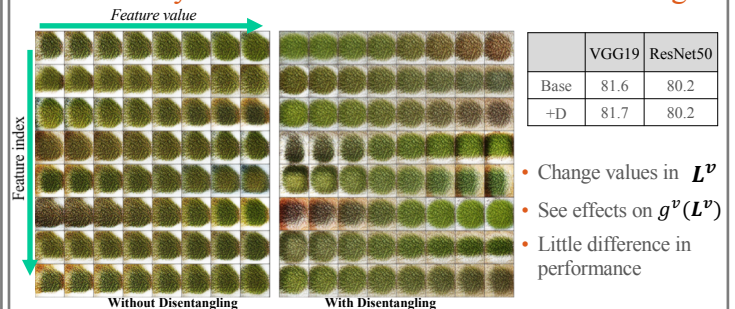
- Only images are needed as input
- Mapping from ingredients to food class is achieved using statistics
- Multiview decision fusion enables a refinement

Performance Comparison

Model	VireoFood-172		Ingredient-101	
	Top-1	Top-5	Top-1	Top-5
resnet18	77.1	93.1	79.6	92.9
resnet50	80.2	93.8	82.1	94.3
vgg16_bn	80.4	95.3	80.7	93.4
vgg19_bn	81.6	95.7	81.3	93.7
ARCH-D [7]	82.1	95.9	83.7	96.2
WRN50-2 [38]	82.5	96.1	84.6	96.5
WiSeR [21]	82.8	95.6	85.1	96.6
ATNet_vgg19_bn	85.3	96.5	86.4	96.8
ATNet_resnet50	85.0	96.2	86.7	96.6
ATNetWRN50-2	86.1	96.6	87.3	96.7
ATNetWiSeR	86.2	96.4	87.1	96.5

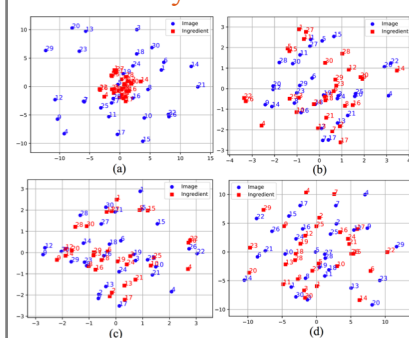
- Our ATNet improves existing methods by a large margin
- ATNet boosts the performance of resnet50 and vgg19_bn to the same level, showing the **effectiveness of feature alignment**
- It **generalizes well** to VGG and ResNet models

Case Study 1: Effects of Partial Feature Disentangle



- Change values in L^v
- See effects on $g^v(L^v)$
- Little difference in performance

Case Study 2: Performance in Feature Alignment



- PCA Visualization
- L^v and L^t
 - After KL loss of alignment
 - After L2 loss of alignment
 - Align with DeepCoral
- Partial heterogeneous transfer first aligns magnitude (b), then the distribution (c)
 - Better than DeepCoral in feature alignment

Case Study 3: Multiview Decision Fusion

