Towards Image representation for Visual Recognition

韓　先花 (Xian-Hua Han)
主任研究员
日本国立研究法人:产业综合技术研究所

Outline

- **Visual Recognition (Image recognition and Understanding)**

- **Three frameworks for Image representation**

  1) **Traditional representation:**

     Color statistics, HOG, LBP-based descriptor etc.

  2) **Bag of Feature model**

     (Super-vector, SC-based, Fisher vector)

  3) **Deep Learning (end to end learning)**

- **Applications**

Image Recognition and understanding

Datasets Public

Caltech101
Caltech256

ImageNet
Pascal VOC

Visual Genome

Object detection &Localization

Image Categorization

Object detection Localization

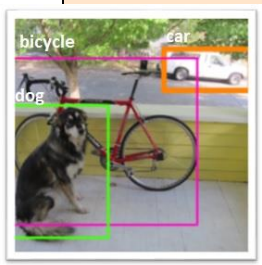
Image Captioning

Purpose

图像全体属于哪一个事先定义的category

图像中有什么物体，在什么位置

给整幅图像或者任意区域赋予说明文



不仅要确定图像中有什么物体，在什么位置，而且需要自动标注物体之间的关系

Image Recognition and understanding

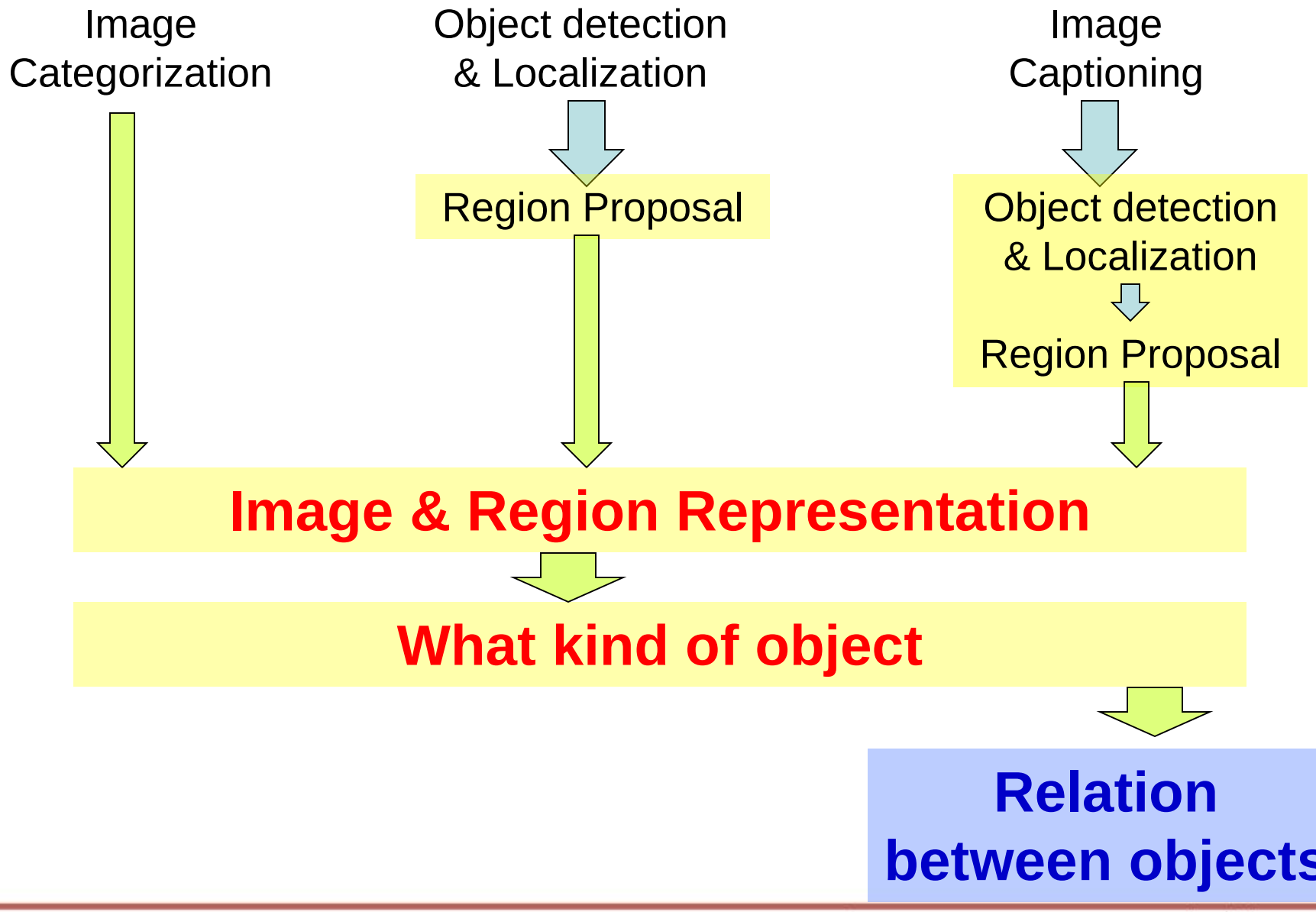


Image Representation

图像

Raw Feature:
RGB, local patch Gradient/Orientation

Coding
Uniformly binary quantization

2005年以前

Global pooling:
CH, LBP, LTP, HOG

最近10年間

Region-based Pooling:
SIFT, SURF etc.

Coding
Pooling

BOF
Super- vector
Fisher Vector

近三，四年

End to End learning:
Deep learning

Learning Coding parameter directly from Raw image

Image Representation: Our works

2005年以前 → 最近10年間 → 近三，四年

图像

Low-level

Global pooling: CH, LBP, LTP, HOG

Local Features Set

Research Line

Selection and fusion of different features

Tensor-based global/local Feature fusion

Middle-level

BOF
Super- vector
Fisher Vector

1. Time-consuming
2. Ignore the relation of local features

Improvement of Coding methods

High-Level

End to End learning: Deep learning

基于张量統計解析的多类特徵以及局部特徵集的融合

多view

颜色特征
形状特征
纹理特征
位置(空間)信息

传统特征融合选择方法

直接连接基于统计分析的维数压缩

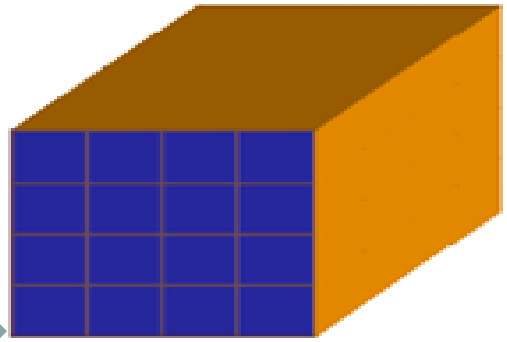
分别进行维数压缩，然后连接

局部特征集

State-of-the-Art融合方法

Bag-of-Features模型

张量表示



张量统计分析

多种特征的有效融合

Extraction of Core Information

开发的多重线性统计分析算法

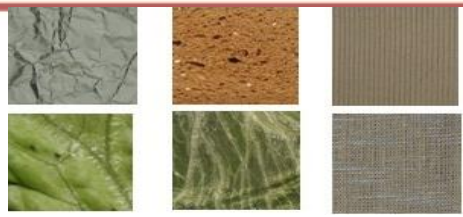
1. GND-ICA　2. 多重线性流形学习

优点：保持特征内部构造和不同特征间的关系进行统计分析

基于多重線形統計解析的多类特徵以及局部特徵集的融合

一般物体识别

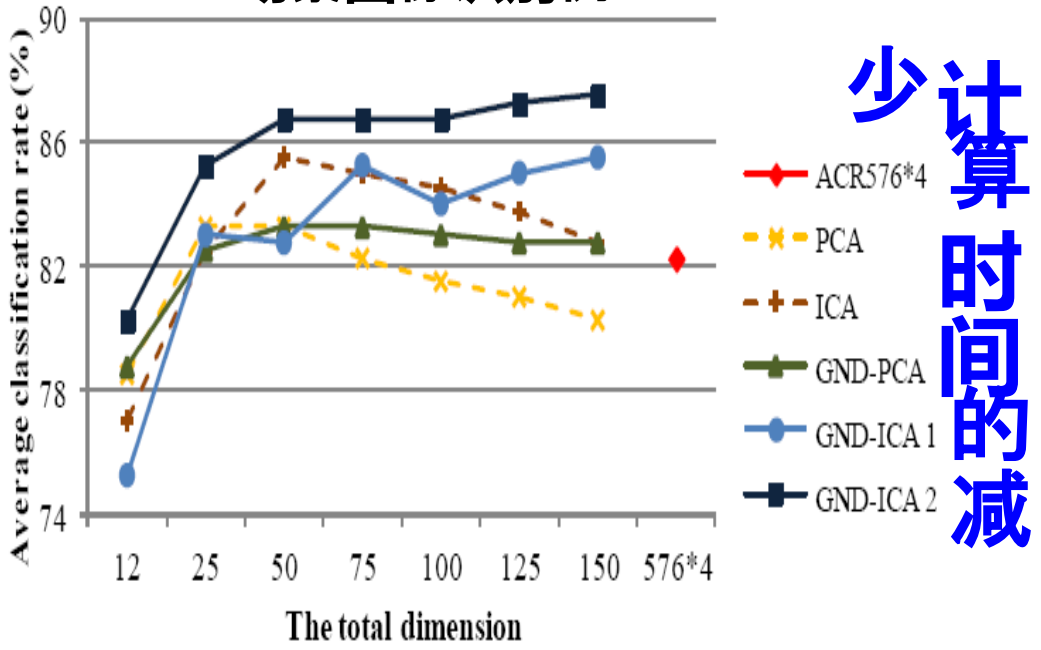
纹理图像分类



3D物体识别

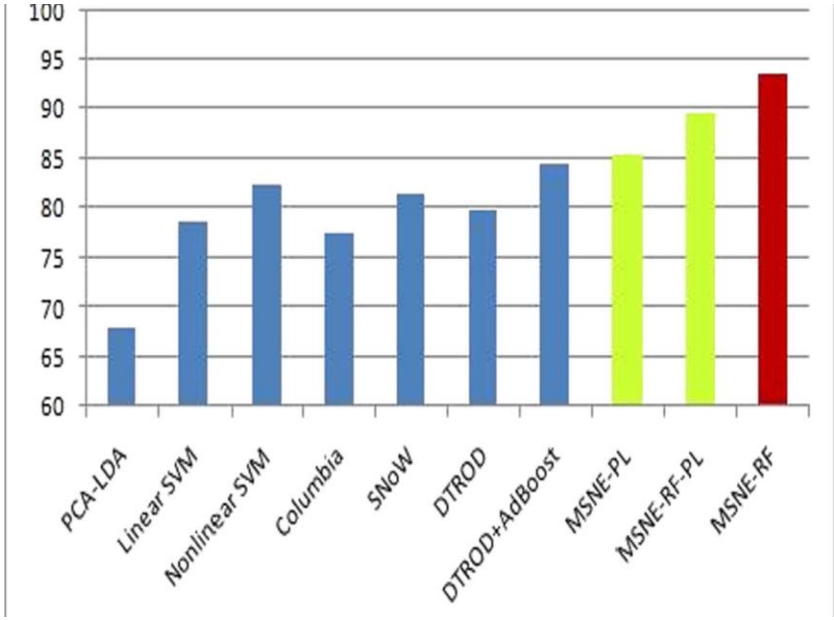
场景识别

场景图像识别例



少计算时间的减

3D物体识别例



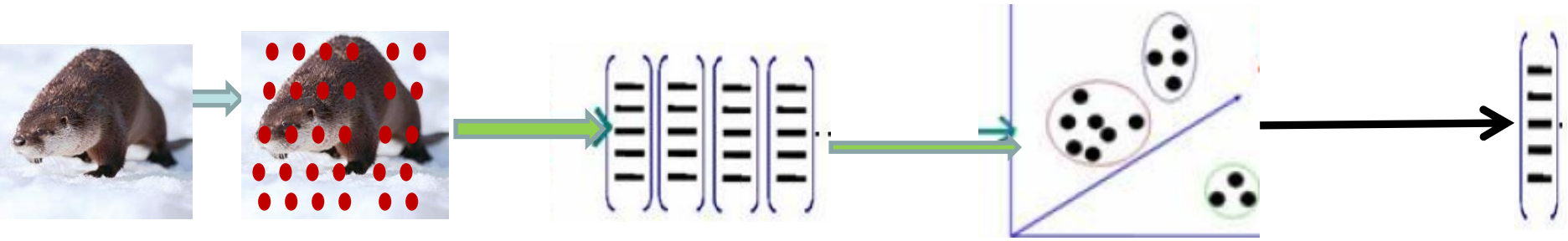
Danni Ai, Guifang Duan, Xian-Hua Han, Yen-wei Chen NEUROCOMPUTING

Xian-Hua Han, Yen-Wei Chen Ruan Xiang, IEEE Transaction on Image processing

Bag-of-Features模型的改良以及高阶統計量的提取

Bag-of-Features: 4 steps

图像 → Keypoint 检测 → 局部特征提取 → Coding → Pooling



传统方法：

Regular Grid
Corner检测

SIFT: 特征的量化
信息丢失

K-Means: 只用一个Visual word来表示局部特征

平均
Max

低阶统计量

开发算法：

Regular Grid

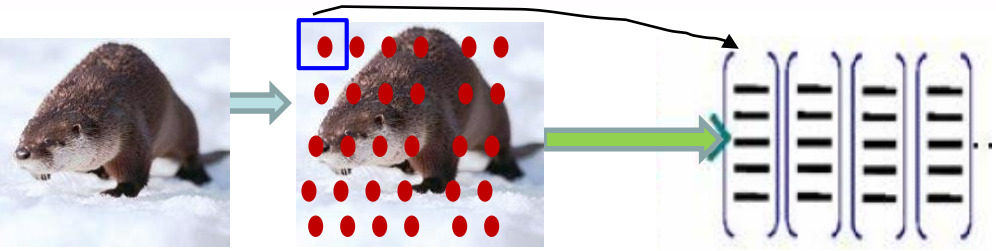
Texton 和WLD:
无信息丢失

稀疏编码
高斯混合模型 (GMM)

K-support

高阶统计量

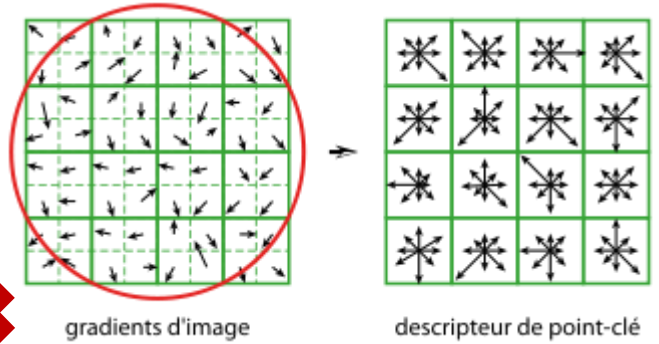
Bag-of-Features模型的改良以及高阶统计量的提取

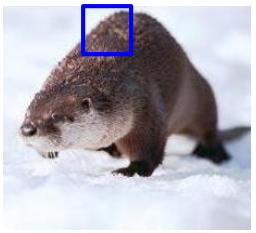


Efficient for distinguishing objects ○

Ignoring tiny information variation for fine-grained ✖

SIFT: 特征的量化（信息丢失）





Raw Patch: Texton

Transform to Excitation domain: Texton (**WLD**)

Texton 和WLD:
无信息丢失

Transformation to excitation domain

Weber Law: **Just-noticeable difference between two stimuli is proportional to the magnitude of the stimuli**

$$\frac{|X'-X|}{X}=\frac{\nabla X}{X}=K$$

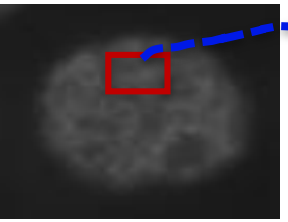
Two Stimulus : X, X',

$$|X'-X|>K\times X$$

X' is different from X

Bag-of-Features模型的改良以及高阶統計量的提取

Focused Pixel

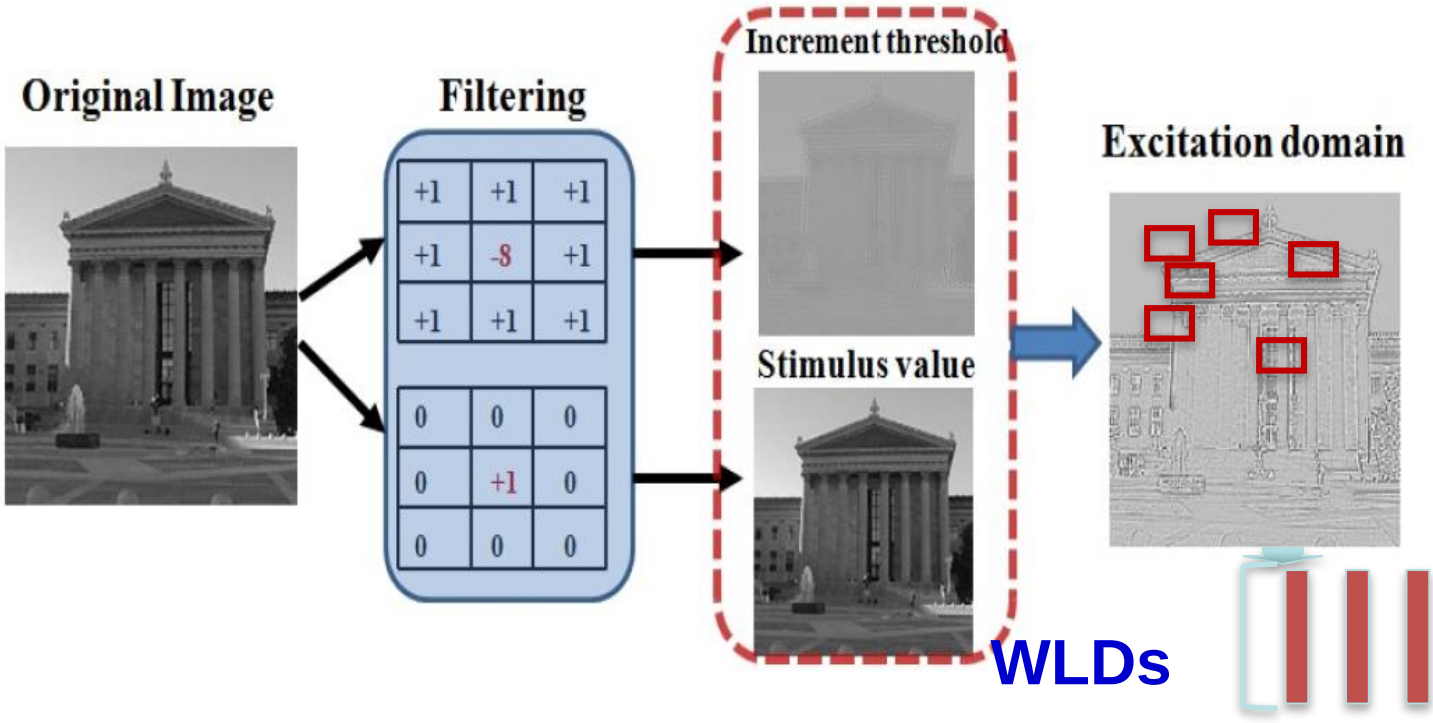


200	80	70
23	100	90
56	30	120

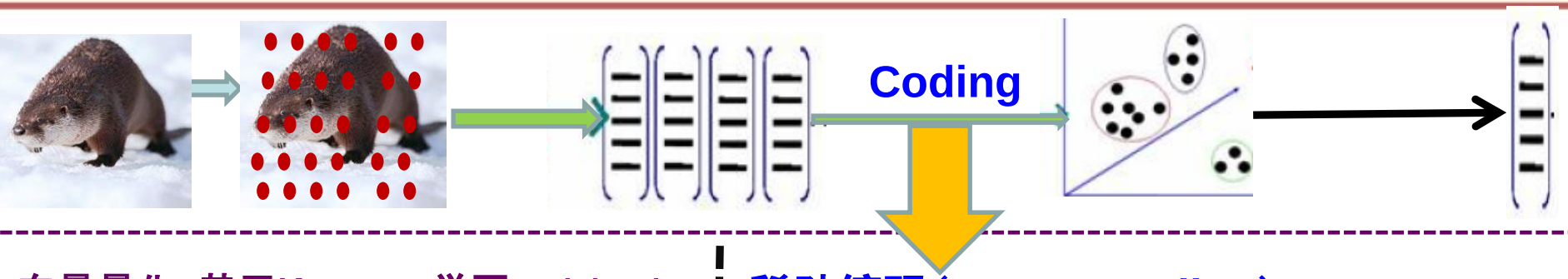
$$\triangle I_c = \sum_{i=0}^{p-1}(\triangle I_c^i) = \sum_{i=0}^{p-1}(I_c^i - I_c)$$

$$H_{ratio}(x_c) = \frac{\triangle I_c}{I_c}$$

$$v_c = \xi(x_c) = \arctan[H_{ratio}(x_c)]$$



Bag-of-Features模型的改良以及高阶統計量的提取



向量量化：基于K-means学习codebook

$$\min \sum_{m=1}^{M} \left\| \mathbf{x}_m - \mathbf{V}\mathbf{u}_m \right\|^2$$

$$Card(\mathbf{u}_n) = 1, |\mathbf{u}_n| = 1, \mathbf{u}_n \geq 0, \forall n$$

只用一个Visual word来j近似局部特征：信息丢失

Pooling方法：Average pooling

$$\mathbf{z} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{u}_m$$

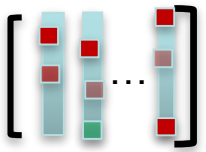
U的均值作为最终图像表示

稀疏编码（Sparse coding）：

$$\min \sum_{m=1}^{M} \left\| \mathbf{x}_m - \mathbf{V}\mathbf{u}_m \right\|^2 + \lambda \left| \mathbf{u}_m \right|_1$$

$$\left\| \mathbf{v}_l \right\| \leq 1, \forall l = 1, 2, \mathrm{K}\ , L$$

多个稀疏系数向量



多个Visual word的线性组合来表示局部特征：近似精度高

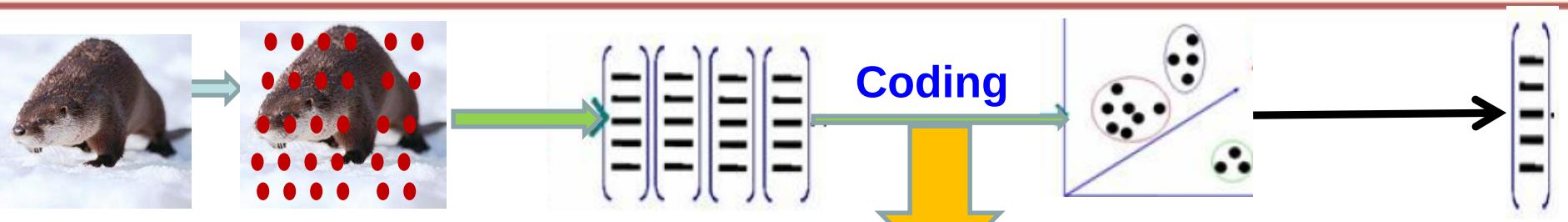
Pooling方法：Max pooling

U的最大值

$$z_j = \max\{|u_{1j}|, |u_{2j}|, \ldots, |u_{Mj}|\}$$

可能导致极端的噪声作为最终的图像表示

Bag-of-Features模型的改良以及高阶統計量的提取

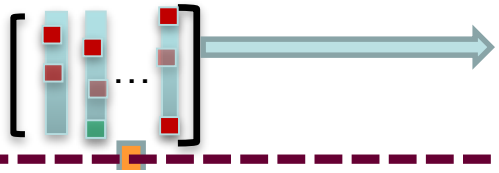


稀疏编码：

$$\min \sum_{m=1}^{M} \left\| \mathbf{x}_m - \mathbf{V}\mathbf{u}_m \right\|^2 + \lambda \left| \mathbf{u}_m \right|_1$$

$$\left\| \mathbf{v}_l \right\| \le 1, \forall l = 1,2,\mathrm{K}\ ,L$$

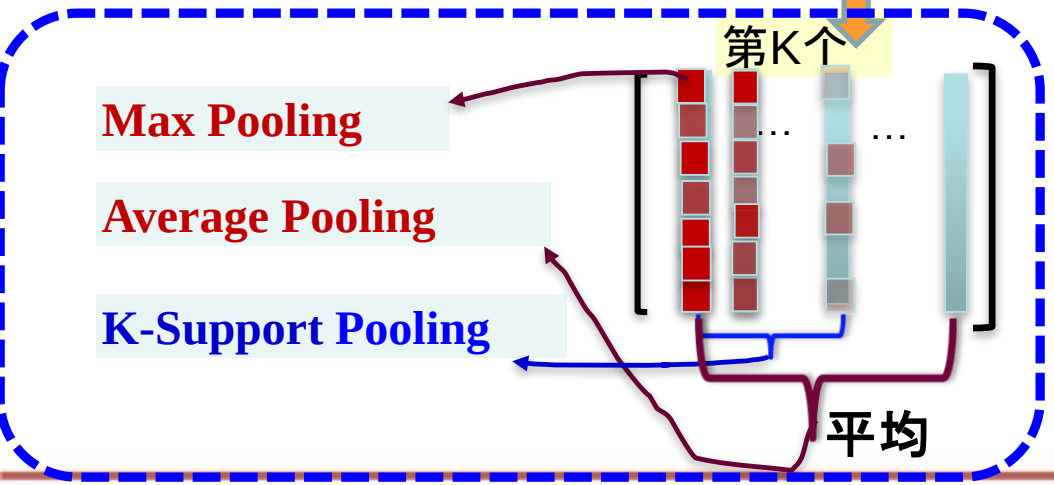
多个稀疏系数向量



Pooling方法：用来融合多个稀疏系数向量、得到同样长度的向量来表示任意图像

統合ベクトル

将行的各个元素根据从大到小排序



K-Support Pooling

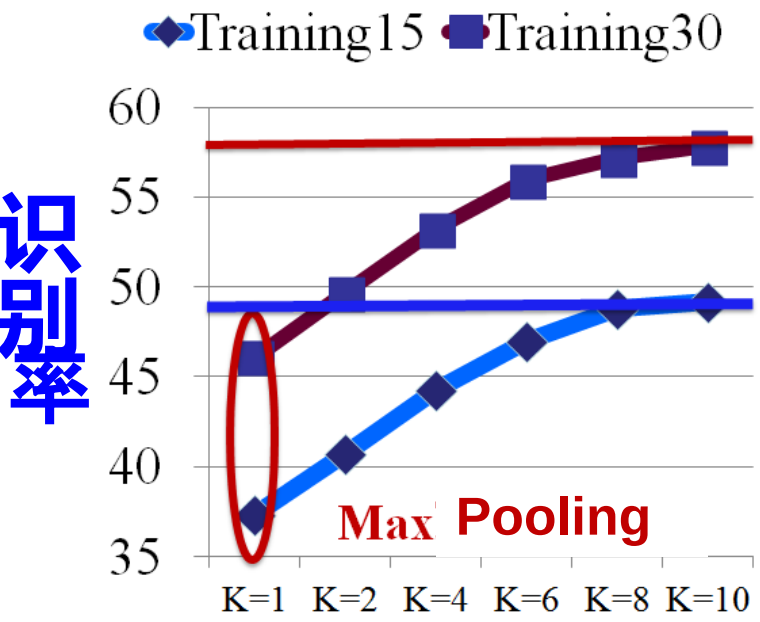
$$\mathbf{T}^j = descend\{\mathbf{U}^j = [u_1^j, u_2^j, \Lambda\ \Lambda\ , u_M^j]\}$$

$$z_j = \frac{1}{K} \sum_{k=1}^{K} t_l^j$$

Bag-of-Features模型的改良以及高阶統計量的提取

Caltech101数据库

1. 101物体类别和其他 (102 class)
2. 30~800张图像/class
3. Benchmark实验条件

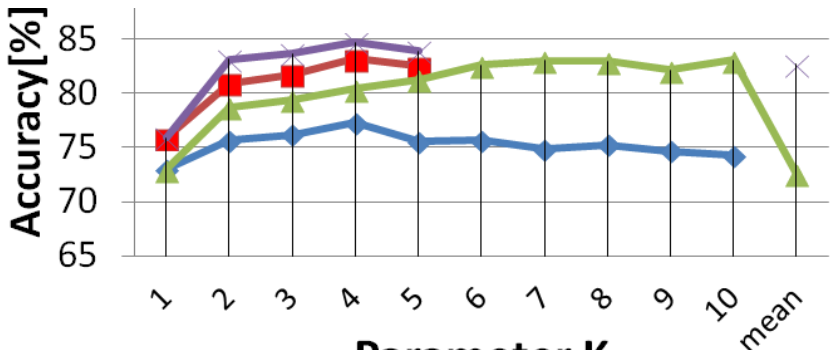


FoodLog系统的开发;食事生活的管理

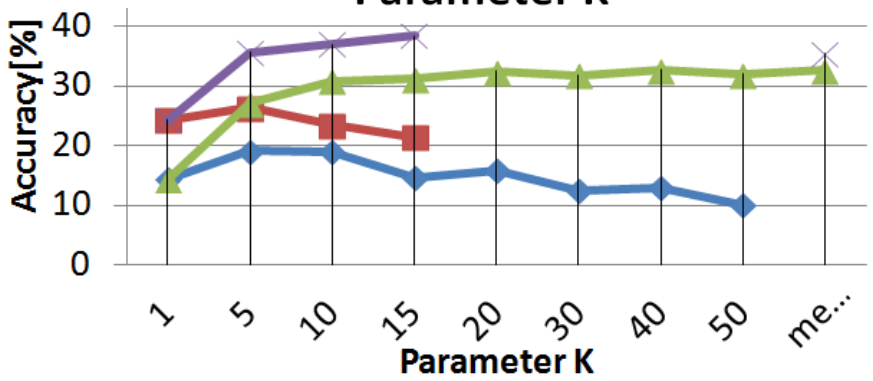
→ 食物内容的自動認識

1. RFID (Ritsumeikan Food Image Database): 10 Class、300张图像
2. PFID(Pittsburgh Fast Food Image database): 61 Class 、 2000多张图像

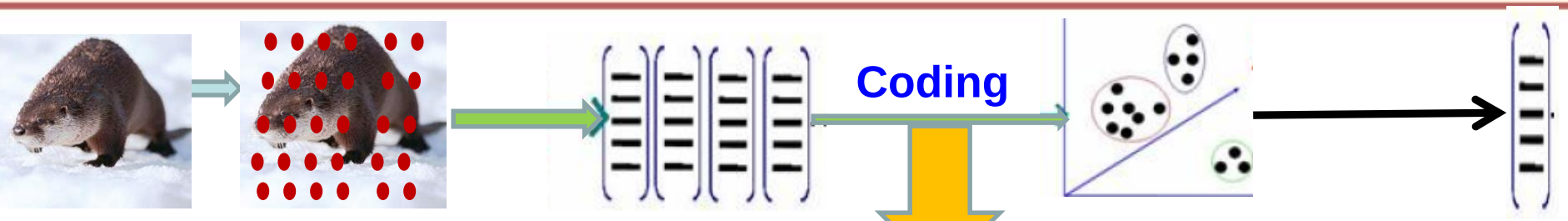
RFID



PFID



Bag-of-Features模型的改良以及高阶統計量的提取



K-means，稀疏编码：

$$\min_{V} \sum_{m=1}^{M} \left\| \mathbf{x}_m - \mathbf{V}\mathbf{u}_m \right\|^2 + \lambda \left| \mathbf{u}_m \right|_1$$

学习Phase

仅一个参数：Codebook V

Coding

基于Codebook V来计算编码系数向量，再融合（Pooling）

Pooling

低阶统计量

高斯混合模型（GMM）

$$P(\mathbf{X}/\lambda) = \sum_{k=1}^{K} w_k N(\mathbf{X}, \mu_k, \Sigma_k)$$

三种参数：权重、均值、标准方差

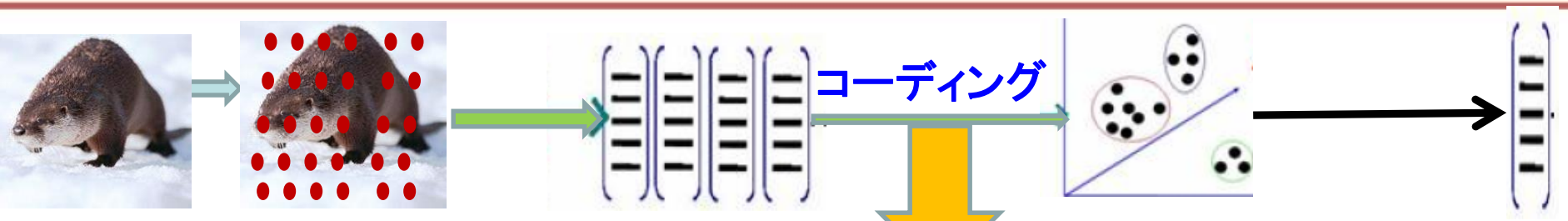
Coding

各种参数的梯度信息

Pooling

高阶统计量

Bag-of-Features模型的改良以及高阶統計量的提取



高斯混合模型(GMM):

$$P(\mathbf{X}/\lambda) = \sum_{k=1}^{K} w_k N(\mathbf{X}, \mu_k, \Sigma_k)$$

三种参数:权重、均值、标准方差

Coding

各种参数的梯度信息

Pooling

高阶统计量

定义: $\mathbf{G}_\lambda^{\mathbf{X}} = \nabla_\lambda \log P_\lambda(\mathbf{X})$

w_k 的梯度: $\frac{\partial P(\mathbf{X}|\lambda)}{\partial \alpha_k} = \sum_{t=1}^{T}[\gamma_t(k) - w_k]$

低阶统计量(order0)

μ_k, Σ_k 的梯度:

高阶统计量

$$\frac{\partial P(\mathbf{X}|\lambda)}{\partial \mu_k^d} = \sum_{t=1}^{T} \gamma_t(k)[\frac{x_t^d - \mu_k^d}{(\sigma_k^d)^2}]$$

order1

$$\frac{\partial P(\mathbf{X}|\lambda)}{\partial \sigma_k^d} = \sum_{t=1}^{T} \gamma_t(k)[\frac{(x_t^d - \mu_k^d)^2}{(\sigma_k^d)^3} - \frac{1}{\sigma_k^d}]$$

order2

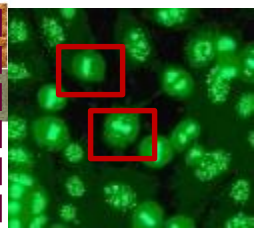
Bag-of-Features模型的改良以及高阶統計量的提取

Texture图像识别（素材分类）：Benchmark 数据库

Challenge：

1. 素材：11种类
2. 摄影条件：Different Scale, Illumination, Orientation

Hep-2细胞的自动识别

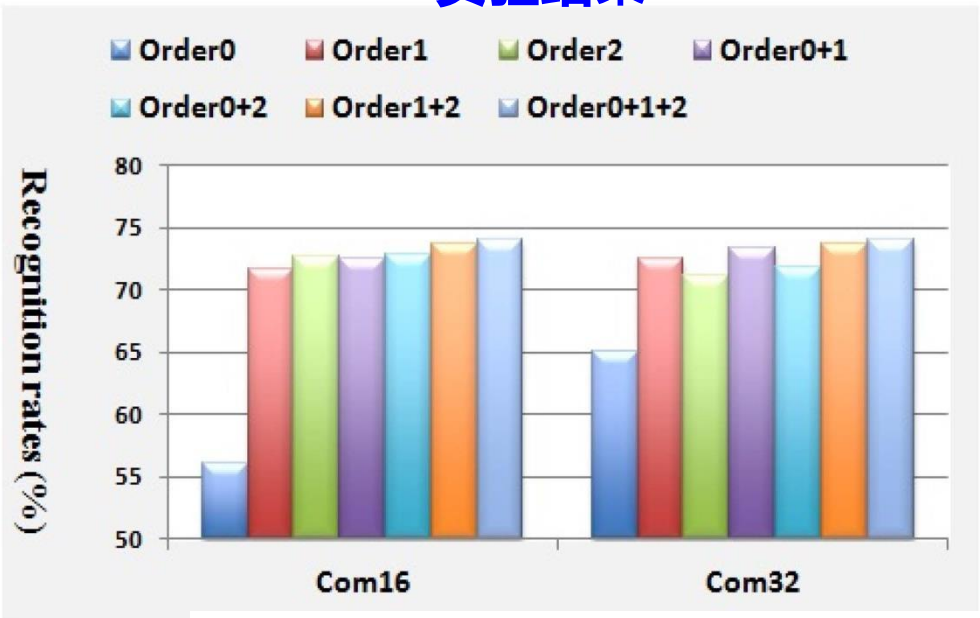


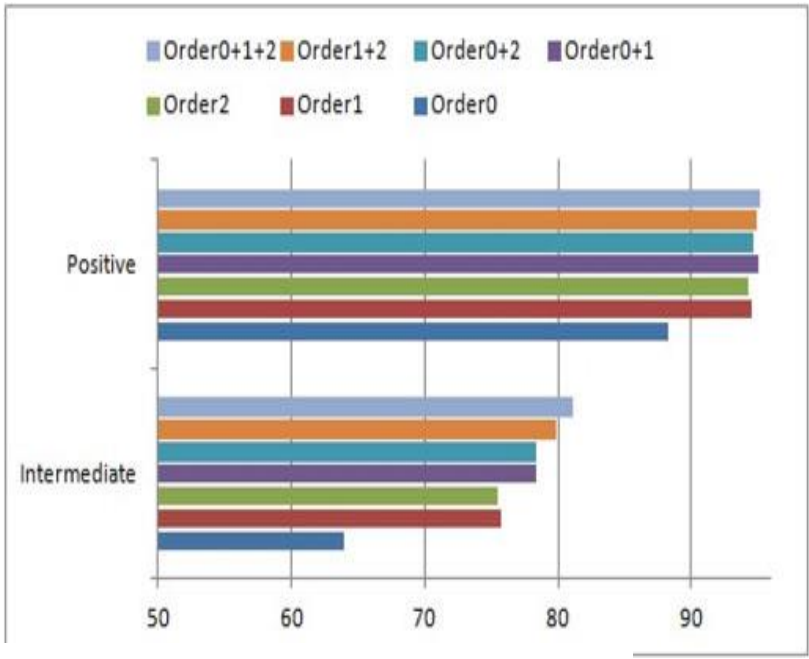
Staining Pattern

关联

疾病

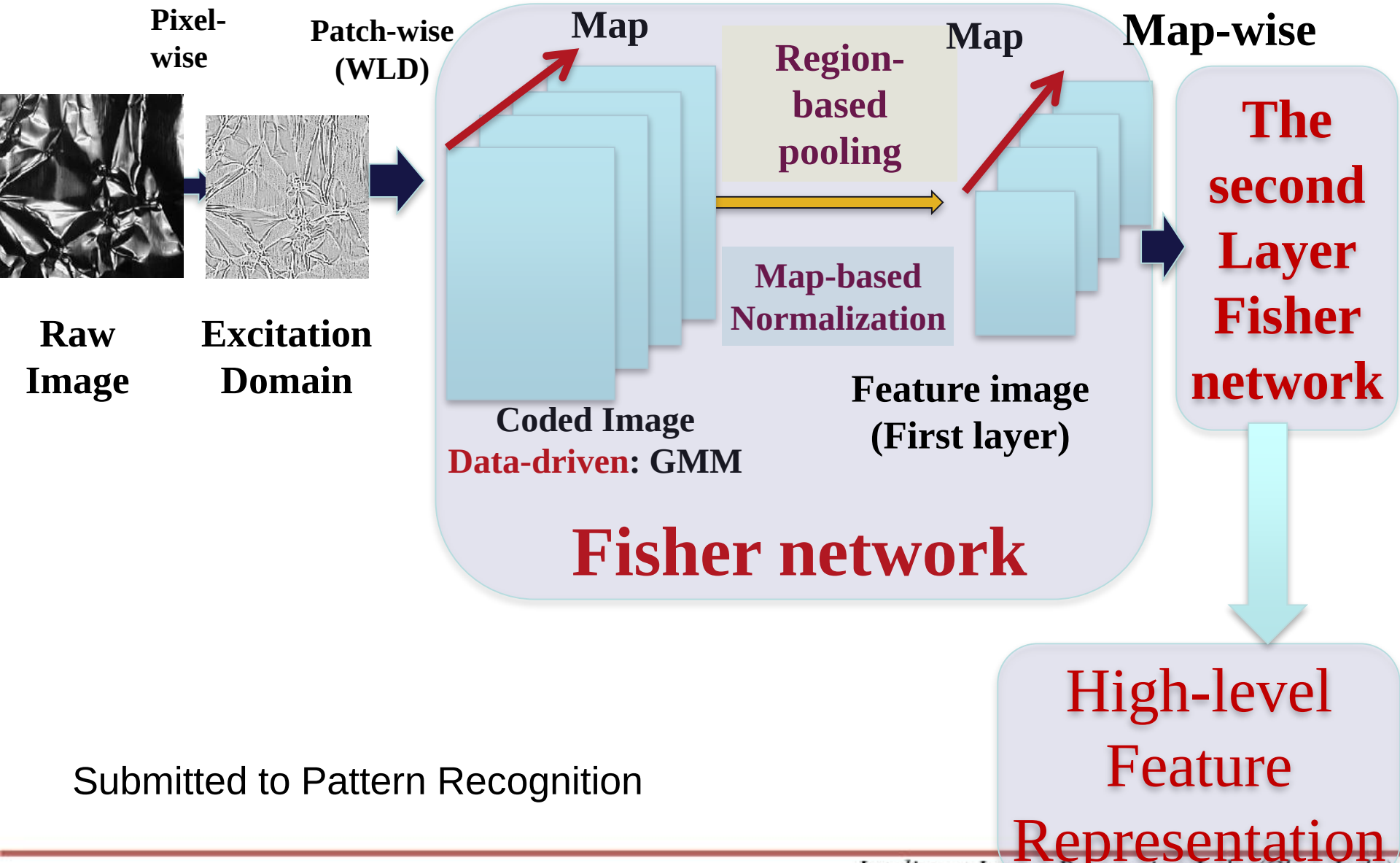
实验结果





IEEE Transaction on Biomedical Engineering/ Cybernetics

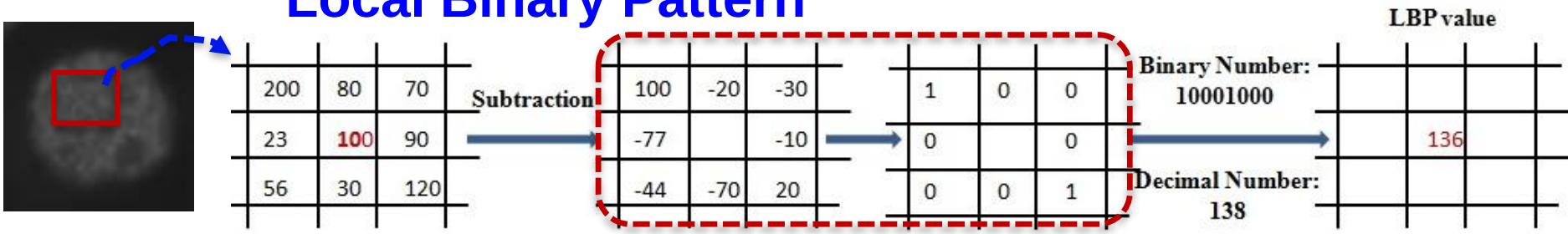
Multi-layer BOF Model



Submitted to Pattern Recognition

Coding methods for Local features

Local Binary Pattern



Reduce coding error

Integrate Spatial or orientation invariance

Local ternary Pattern

Rotation invariant co-occurrence LBP

Pairwise Rotation Invariant Co-Occurrence LBP

Proposed methods:

1) Data-driven threshold based on Weber's law (WLTP)
2) Integrating Spatial and orientation structure (RICoWLP)

High-order statistics of spatial structure

Appeared to Pattern Recognition Letters

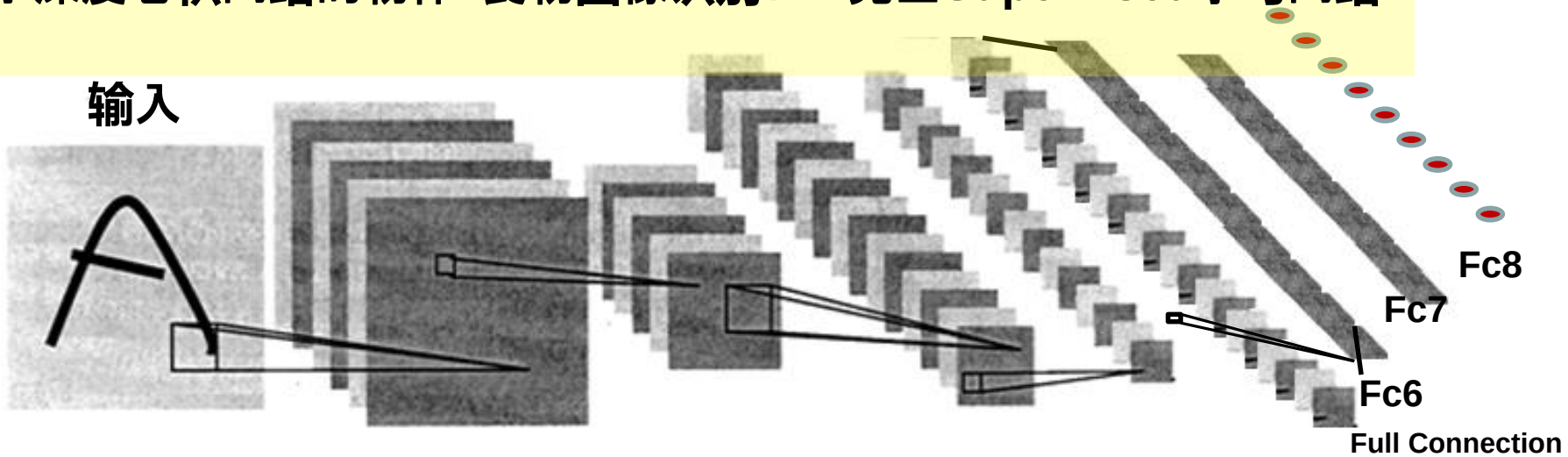
基于深度学习的图像理解

深层学习：　1．神经网络关联的人工智能构造技术的总称
2.模仿 脑（神经细胞）的信息处理过程的学习算法
3. 以具有超多层的大规模网络结构为特征

应用：　图像·物体识别、信息检索、音声识别、机器人工学口等研究领域

识别精度大幅度的提高

基于深度卷积网络的物体·食物图像识别：　完全Supervised学习网络



Conv1:
(1)Convolution
(2) Pooling
(3) Normalization

Conv2: (1)Convolution
(2) Rectification
(3) Pooling
(4) Normalization

Conv3
(1)Conv
(2)Rect

Conv4
(1)Conv
(2)Rect

Conv5
(1)Conv
(2)Rect
(3) Pooling

基于深度学习的图像理解

- **传统图像表示方法**

Fixed/engineered features (fixed kernel)

hand-crafted Feature Extractor (SIFT,HOG等)

等间隔量化

事先决定，不是data-driven

Low-level Features

Unsupervised

Learning for middle-level features (K-means，Sparse Coding，GMM等)

Middle-level Features

"Simple" Trainable Classifier

Possible

(1)直接从图像中学习Low-level特征

(2) High-level特征的学习(Multi-Layer: 深度网络)

K-means

Feature Learning (特征学习)

- **End-to-end learning / Feature learning / Deep learning**

Trainable Feature Extractor

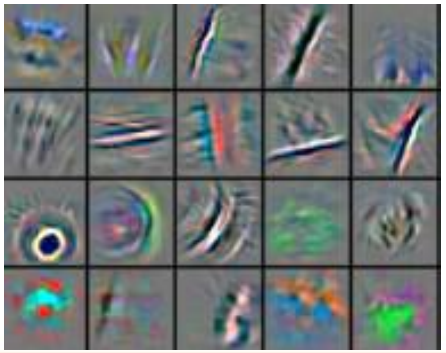
Low-Level 特征 → Middle-Level 特征 → High-Level 特征 → Trainable Classifier

从图像中直接学习卷积核系数

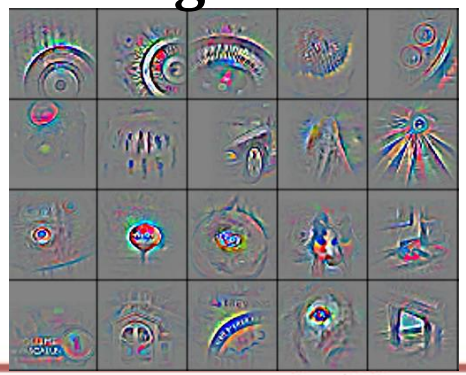
多层非线性特征变换(深度学习)

Low-Level

Middle-Level



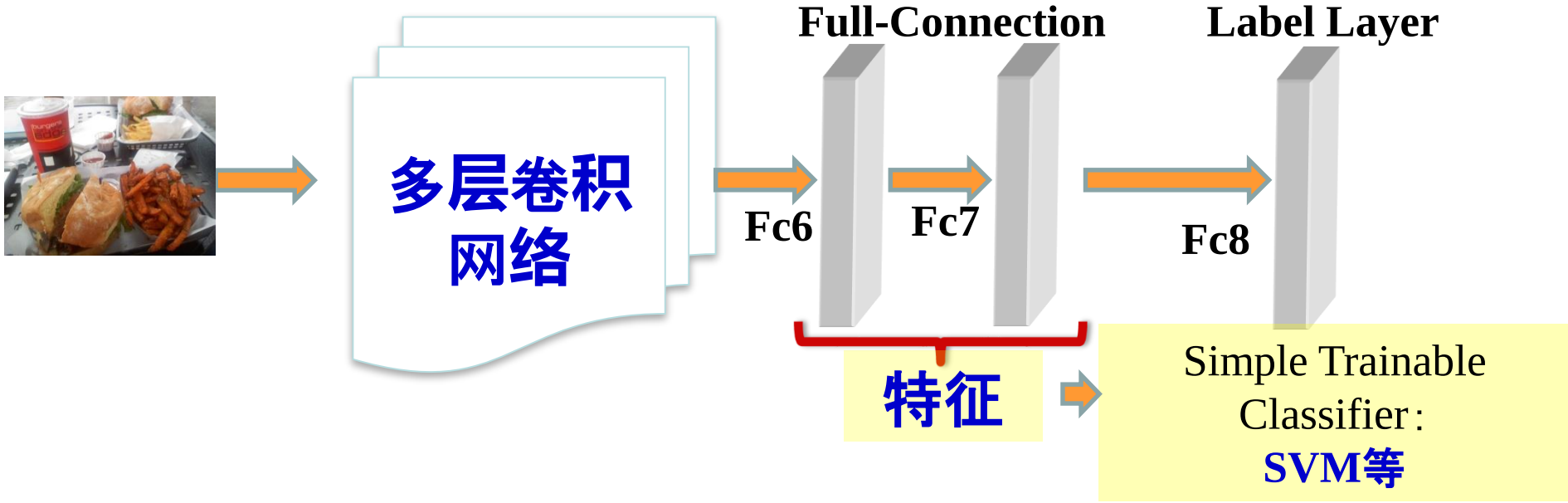
High-Level



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2012]

基于深度卷积网络的食品识别

1. Feature Extraction using the already learned network



2. 转移学习(Transfer Learning)

(1) 有一定数量的标注图像库

(2) 已有的学好网络参数可以作为初始值，利用可获得数据再学习

食品识别:图像库

Databese	RFID (Ritsumeikan Food Image Database)	Food101 ECCV2014论文
Image Example		
Class	10	101
Image Number	300 (小規模)	750*101+250*101(大規模)
Usage of the previous learned network	**Feature Extraction**	**Feature Extraction**/ **Transfer Learning**
Simple Trainable Classifier	Linear Support Vector Machine	

食事認識の実験条件

1. 用いた学習済みモデル（ネットワーク）

Imagenetデータベースを用いた学習されたモデル：５層畳み込み、３層full Connection

2. 特徴抽出器として

A. Full Connectionの第六層（fc6）または七層（fc7）のニューロン反応強度を特徴として
B.分類器： Linear SVM

3. 転移学習によるネットワークの再学習

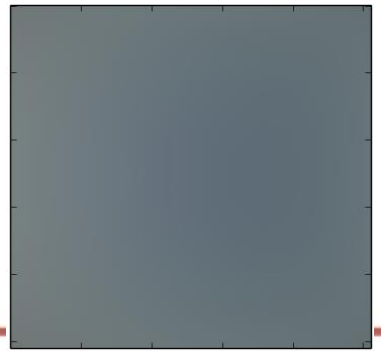
A. 学習済みネットワークの最終層（fc8）： ニューロン数 1000→101
B. Fc8層の学習率は他の層の10倍に設定する

4. 前処理：平均画像を引く

Food101

平均画像

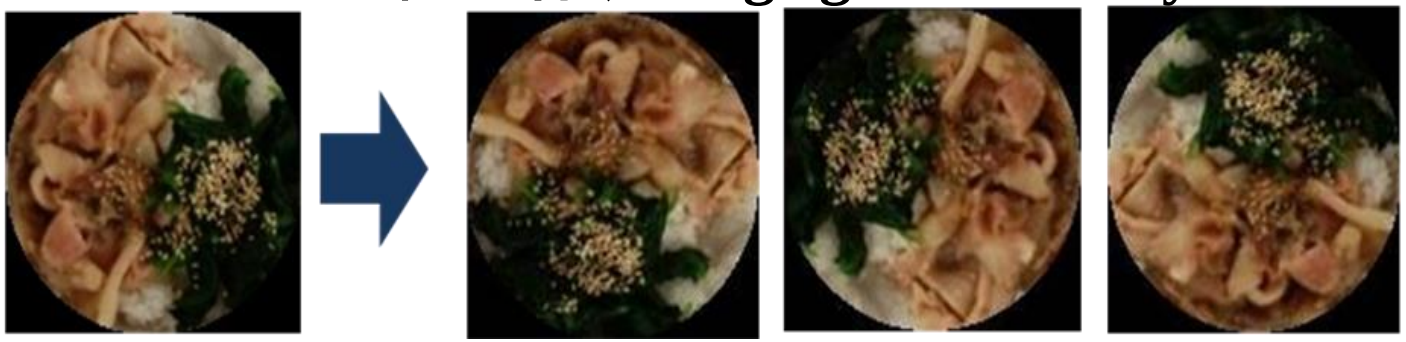
Imagenet



RFID

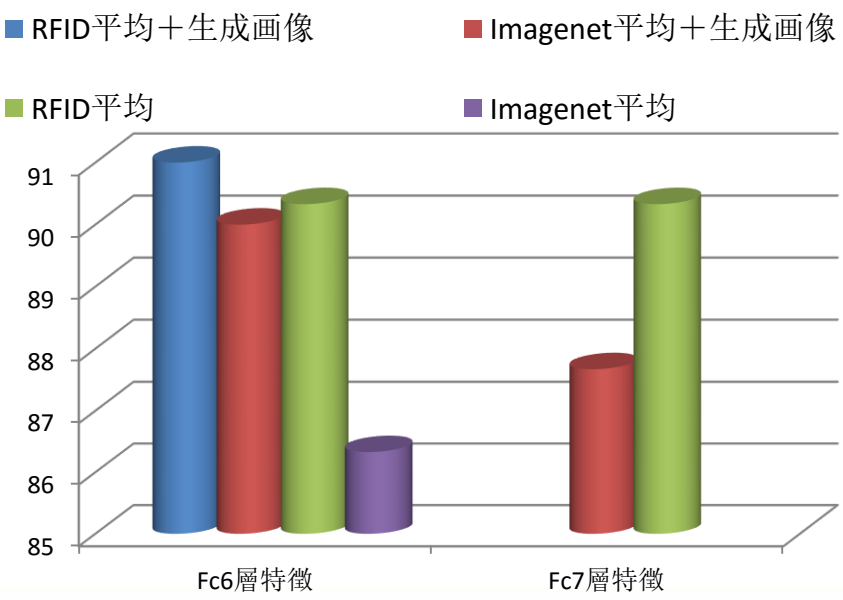
食品识别：特征提取

RFID：300副图像、Image generation by Rotation transform

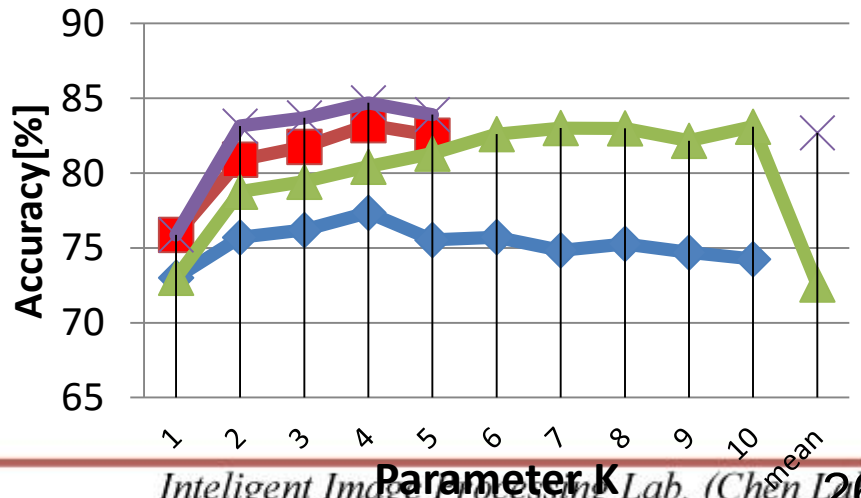


SVM分类器：Leave-one-out Strategy

结果



基于BOF的最好精度：约85%



食品识别：转移学习

Databese：Food101

1） 2014年公开

2） 101类食品图像

3） ECCV2014论文的识别精度：50.76%

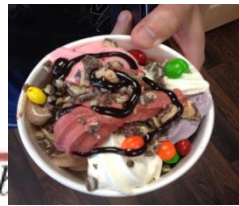
无关图像



	方法	识别精度（%）
The State-of-the-art 法	BOF	28.51
	IFV	38.88
	ECCV2014	50.76
深度学习	特征提取	50.56
	转移学习	**58.72**



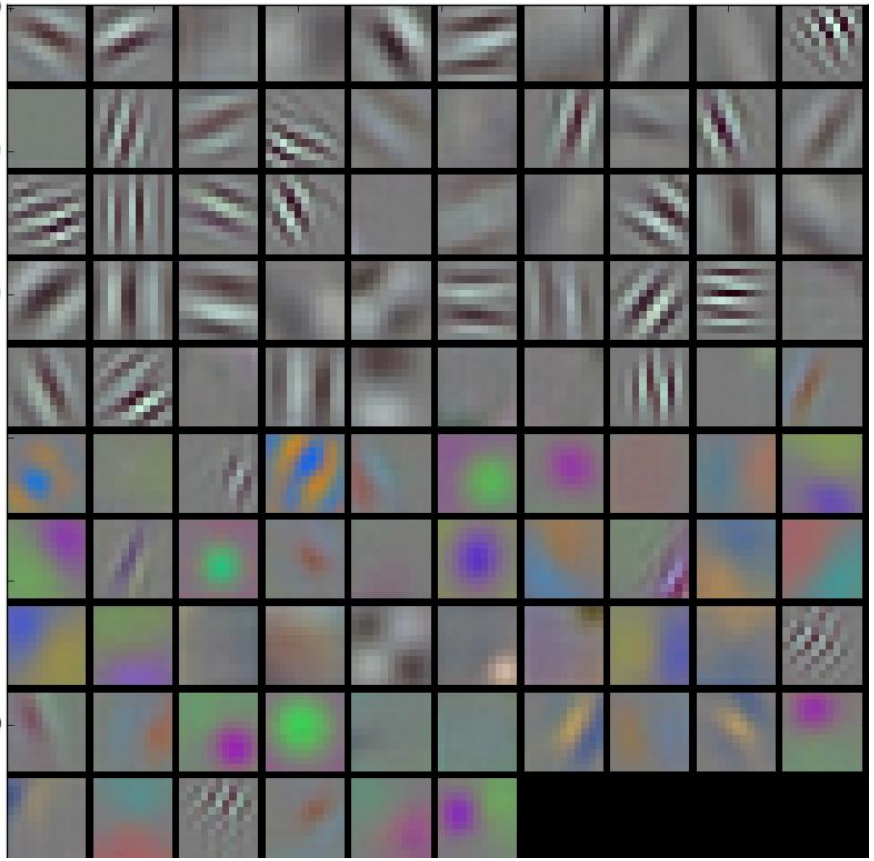


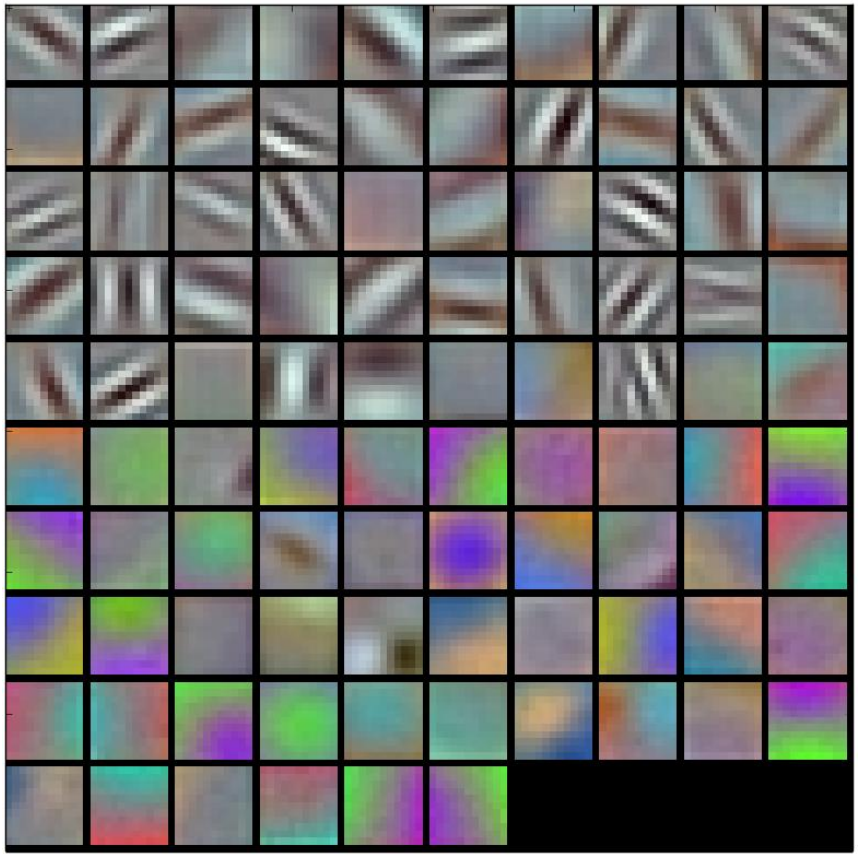
卷积核显示

第一层的卷积核：96个

Imagenet学习网络



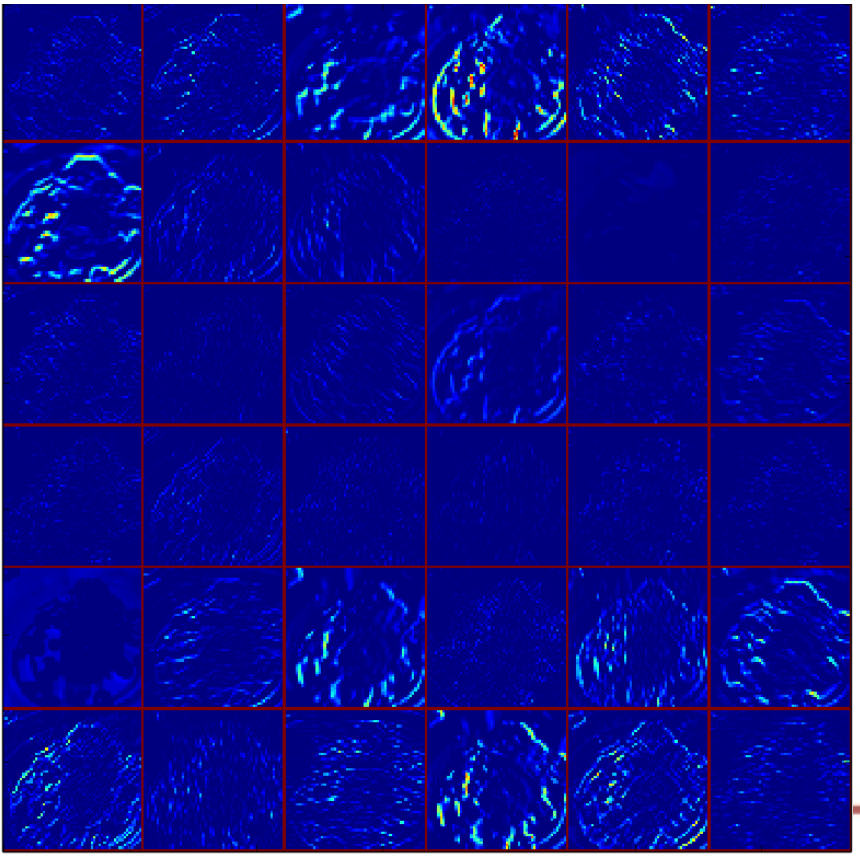
基于Food101的转移学习网络



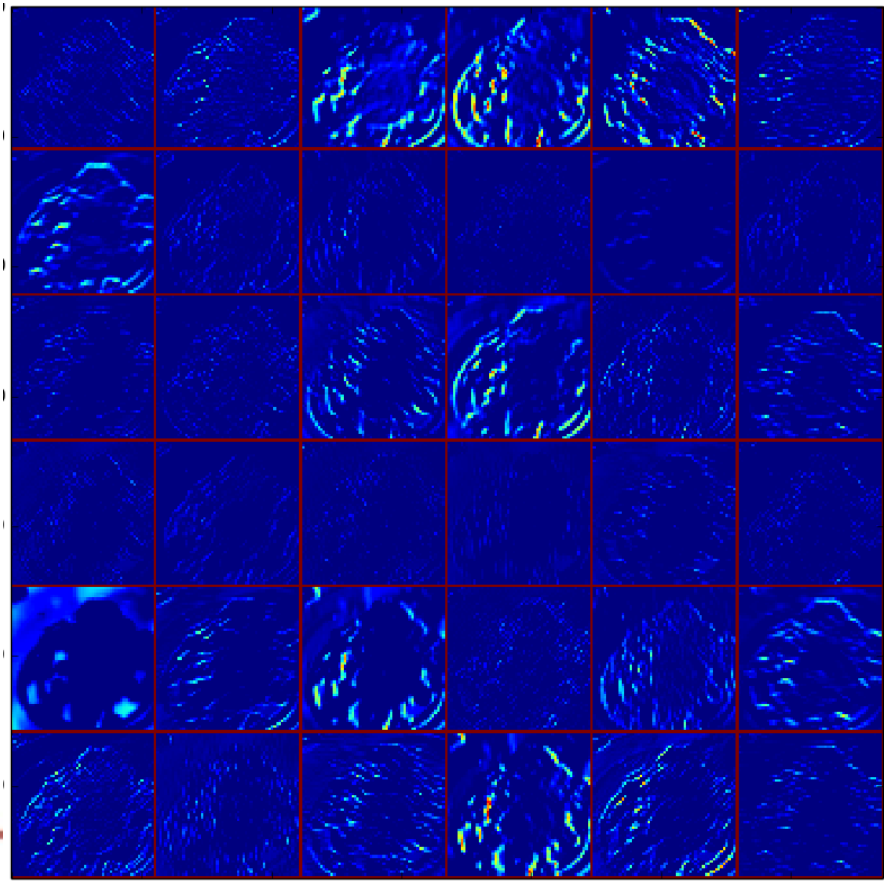
RITSUMEIKAN 转换图像显示

基于第二层卷积核的转换图像（1～36）

Imagenet学习网络



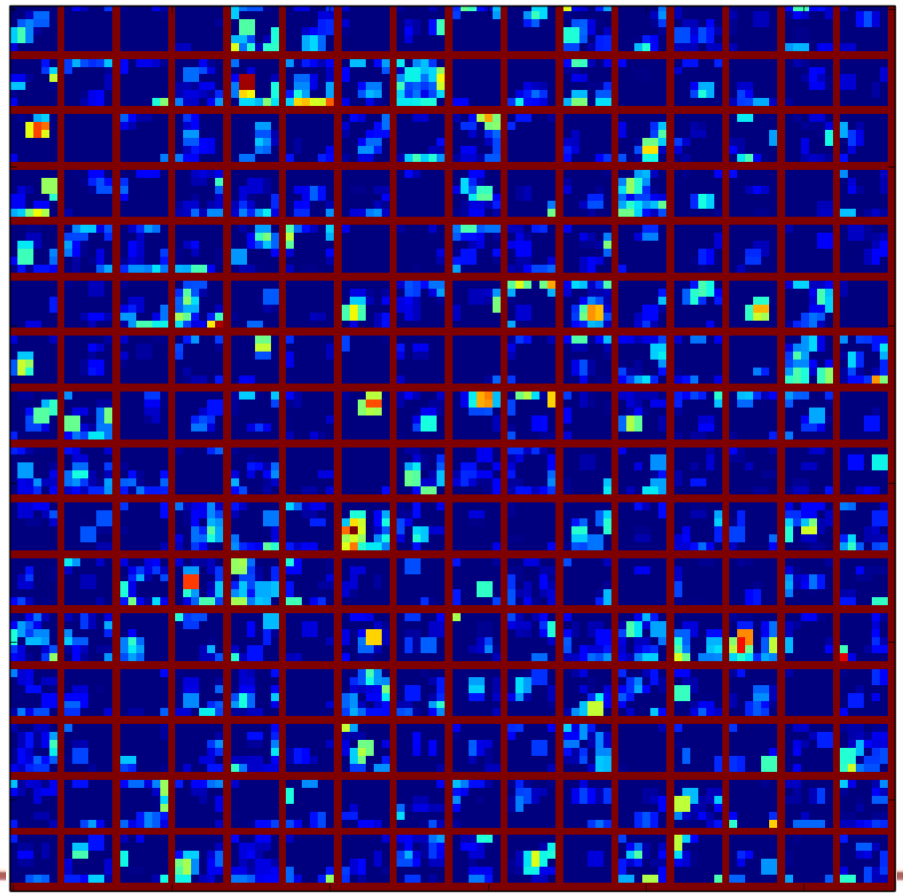
基于Food101的转移学习网络



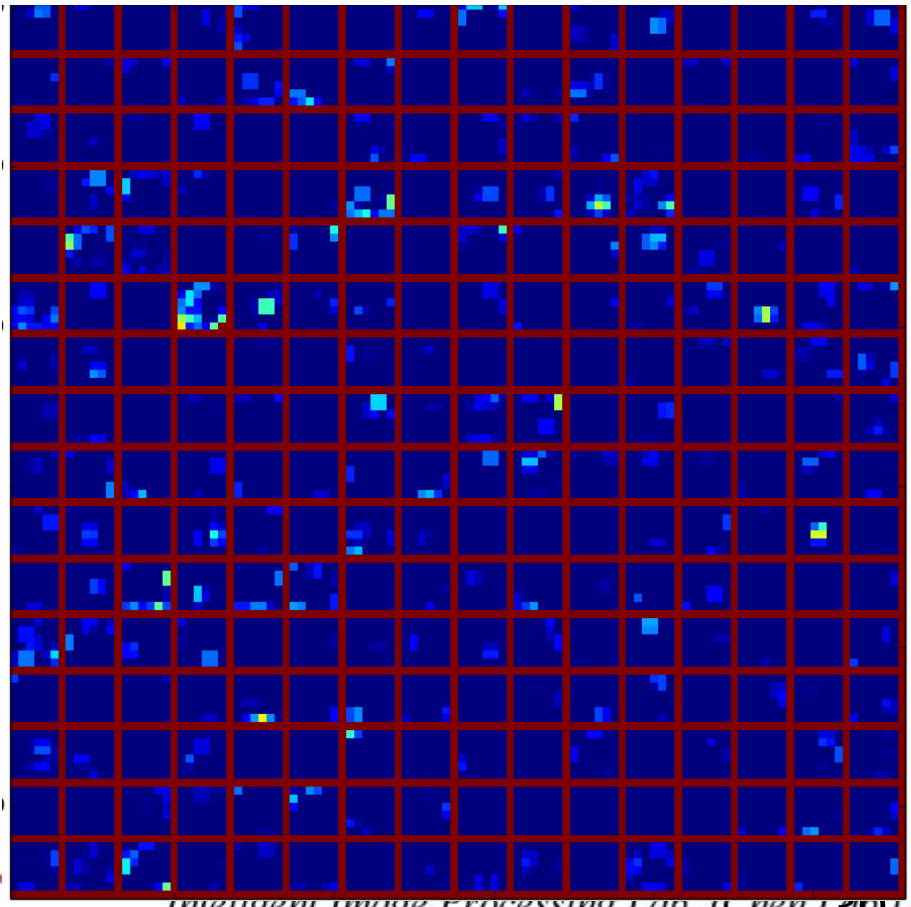
转换图像显示

第五层pooling后的转换图像（256）

Imagenet学习网络



基于Food101的转移学习网络



Summary

- **Image representation**

1) Traditional representation:

Color statistics, HOG, LBP-based descriptor etc.

2) Bag of Feature model

(Super-vector, SC-based, Fisher vector)

3) Deep Learning (end to end learning)