

# Image Annotation with Relevance Feedback Using a Semi-Supervised and Hierarchical Approach

Cheng-Chieh Chiang, Ming-Wei Hung, Yi-Ping Hung, Wee Kheng Leow

Cheng-Chieh Chiang

Assistant Professor  
Department of Information Technology  
Takming University of Science and Technology  
Taipei, Taiwan

*e-mail: kevin@csie.ntnu.edu.tw*

2008/01/23

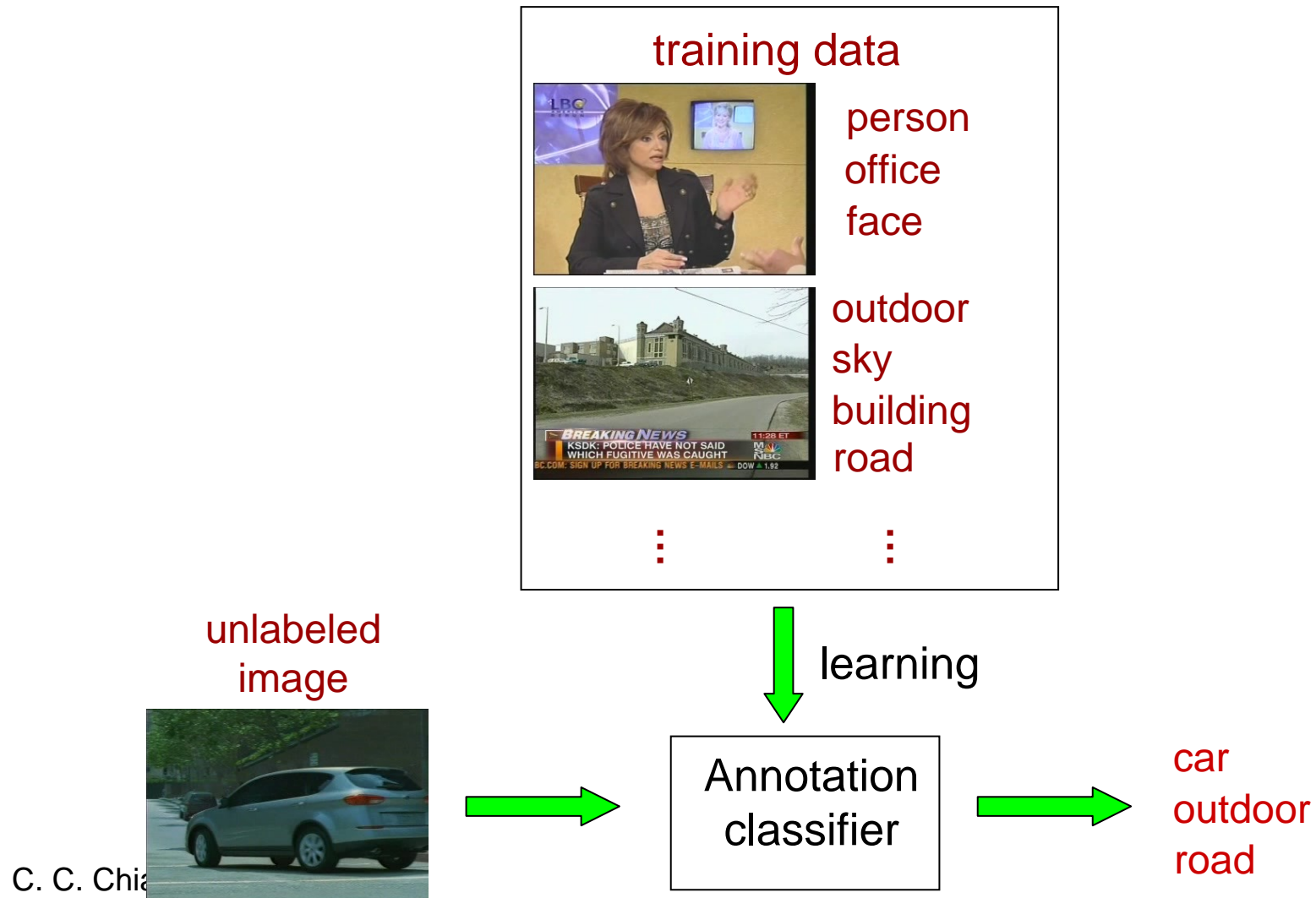
# Outline

- Introduction
- Overview
- Image Feature Extraction
- Classifier Training
- Confidence Value
- Experiments
- Conclusion and Future Work

# Introduction

- Features for image understanding
  - Low-level feature: color, texture, shape, etc.
  - High-level feature: ??
- Semantic gap
  - Between low-level features and high-level concepts for an image
- Image annotation
  - Annotate predefined labels to images
  - Labels are associated with semantic features of images

# Image Annotation



Training Phase



feature extraction

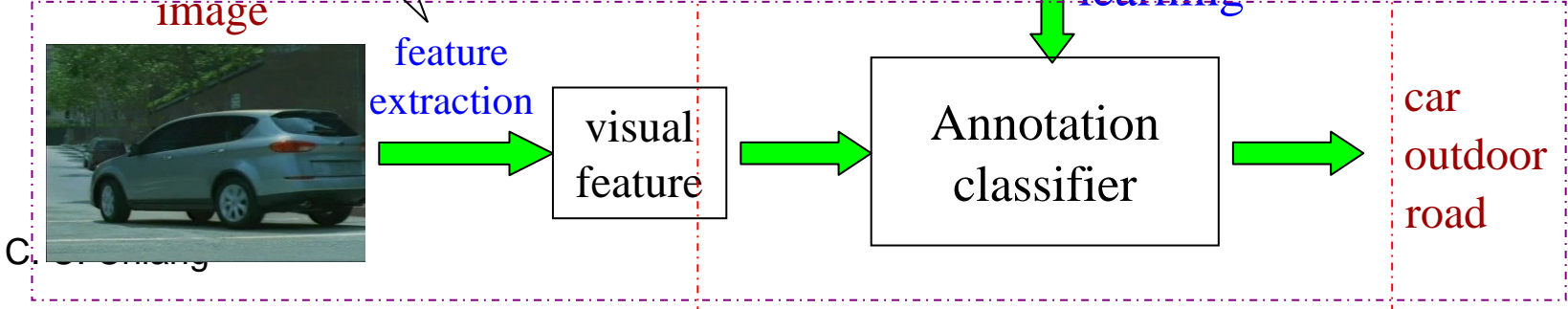
the set of visual feature,  $V$

the set of labels,  $W$

learning

Annotation Phase

unlabeled image



# Challenge

- Divergence of semantic concept

- Label: “sky”



**Hierarchical Classifier**

single label at one time

- Ambiguity of semantic concept

- Different descriptions by different people



flower, leaves

plants, petals

**User Feedbacks**

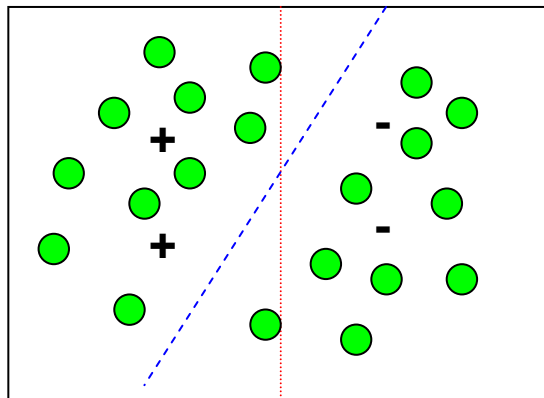
- Less training data

- Difficult to prepare a huge number of labeled images

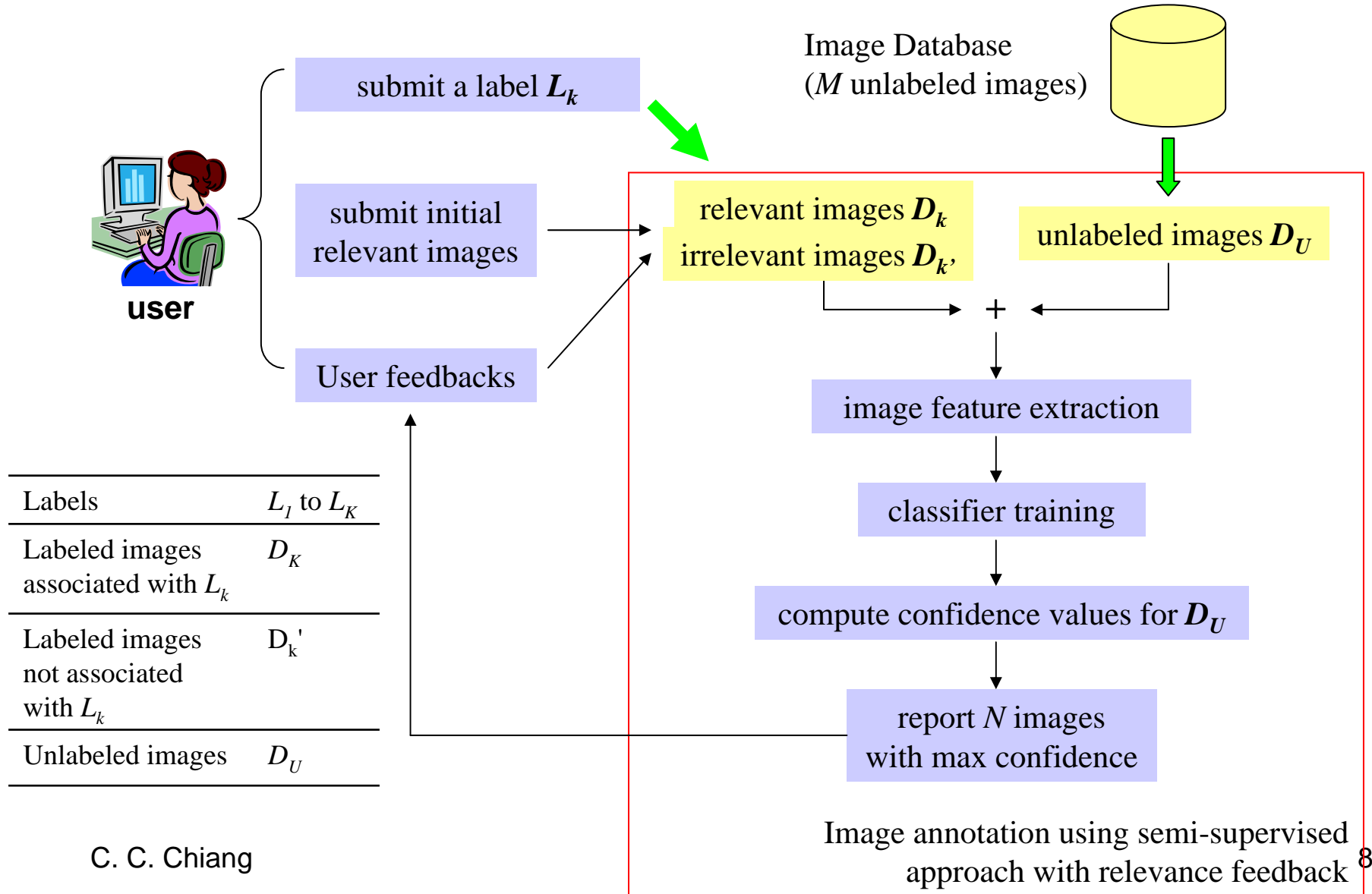
**Semi-Supervised Learning**

# Semi-Supervised Learning

- Labeled data are hard to obtain and unlabeled data are abundant
- Both labeled and unlabeled data are used for training

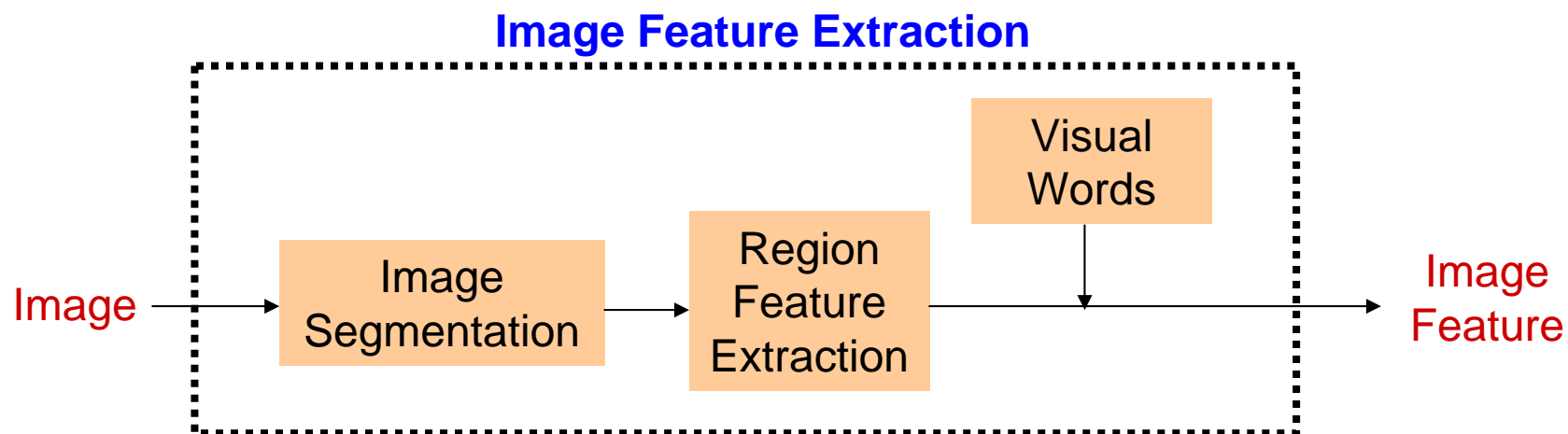


# Overview

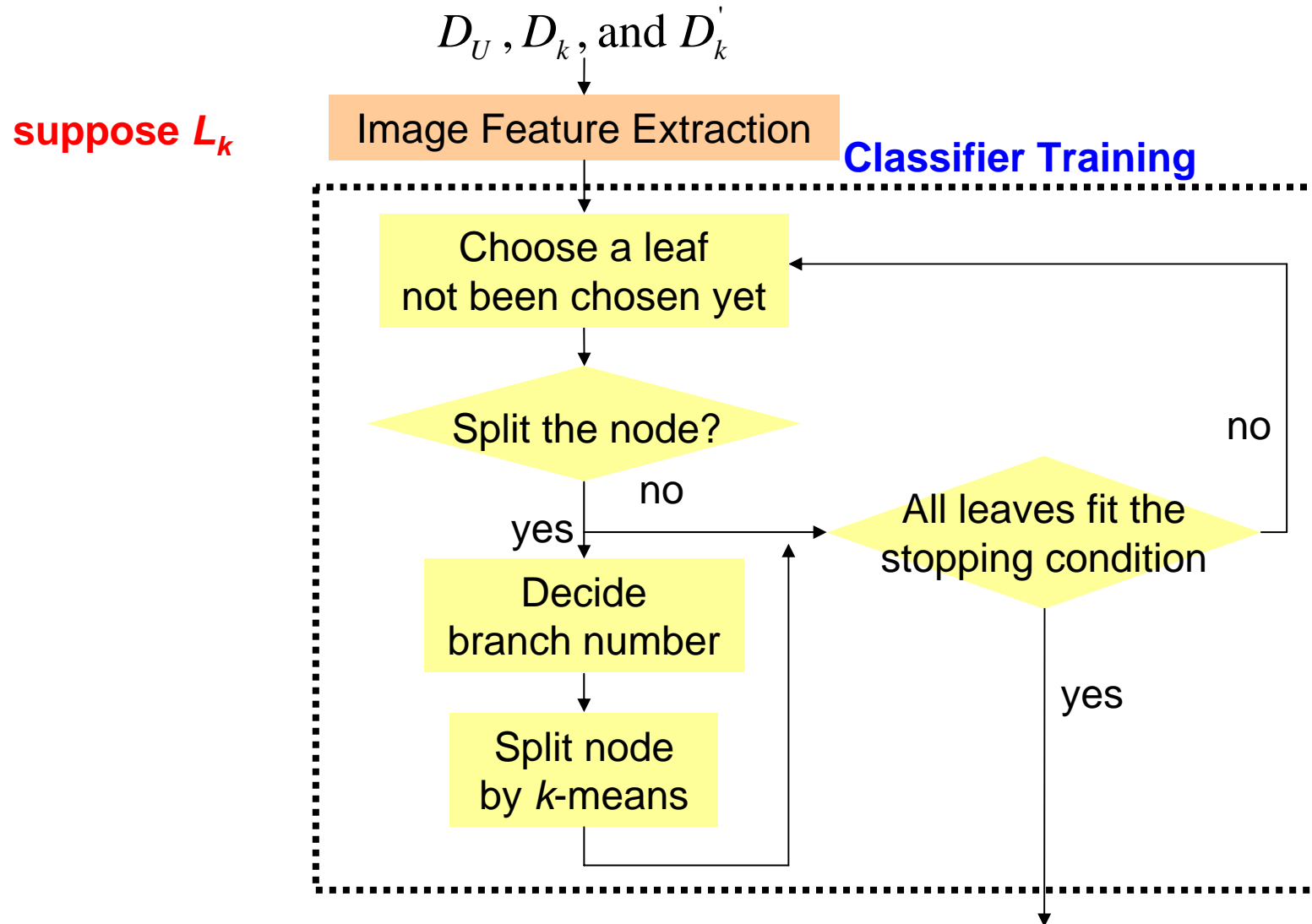


# Image Feature Extraction

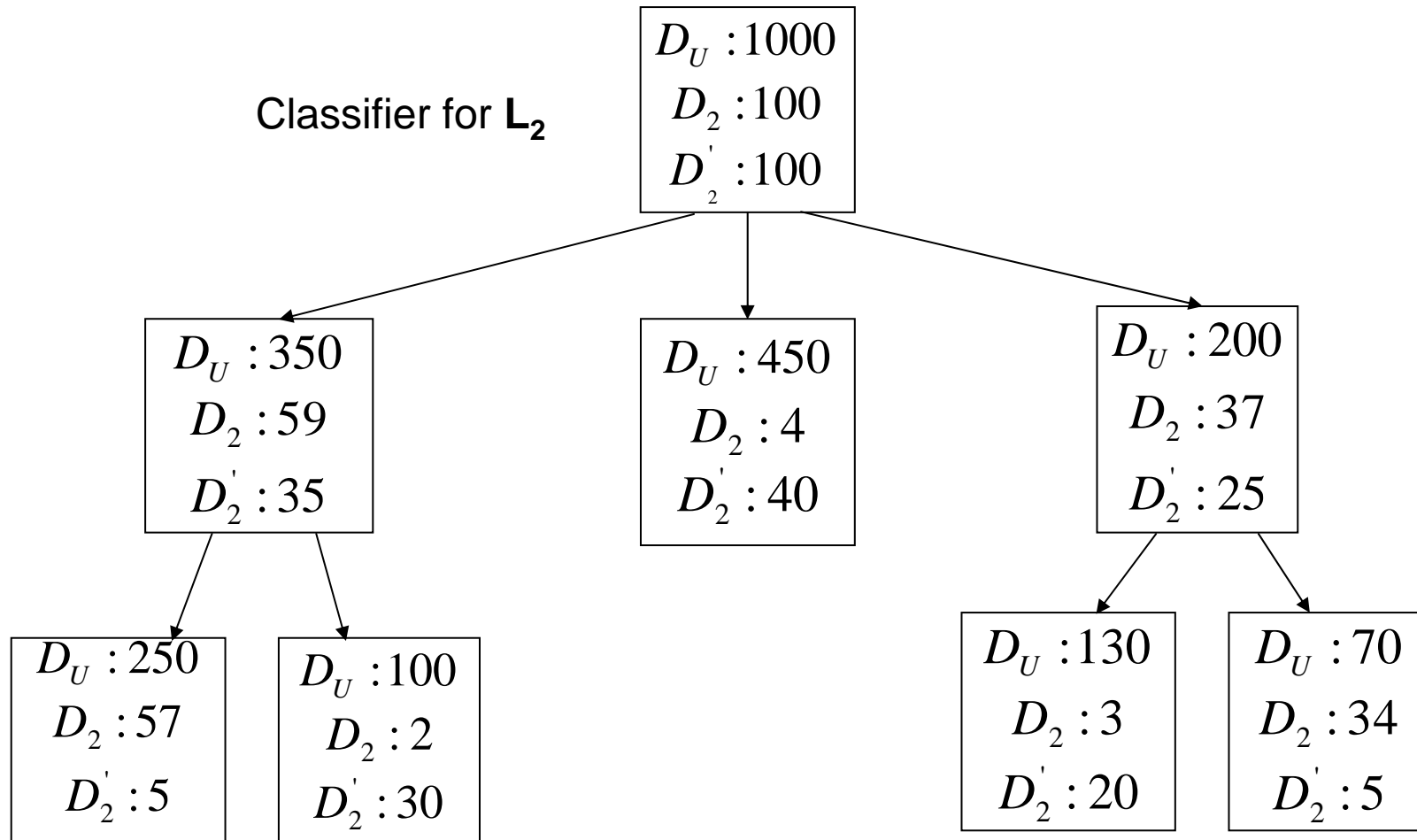
- We adopt the visual word model for image representation



# Classifier Training



# Classifier Training



1. When to stop splitting?

2. How to decide the branch number at a node? 11

# Classifier Training

- When to stop splitting?
  - when the ratio of labeled data in the node is large or small enough
- How to decide the branch number  $z$ ?
  - Try a range of  $z$ , e.g., 2~6
  - Choose the  $z$  with the minimal score
  - The score function is computed using entropy

# Confidence Value

- Given

- a test image  $I_{new}$
- a classifier for  $L_k$

- Compute confidence value of  $I_{new}$  associated with  $L_k$

- $\gamma(L_k, N_{root}^k | I_{new})$

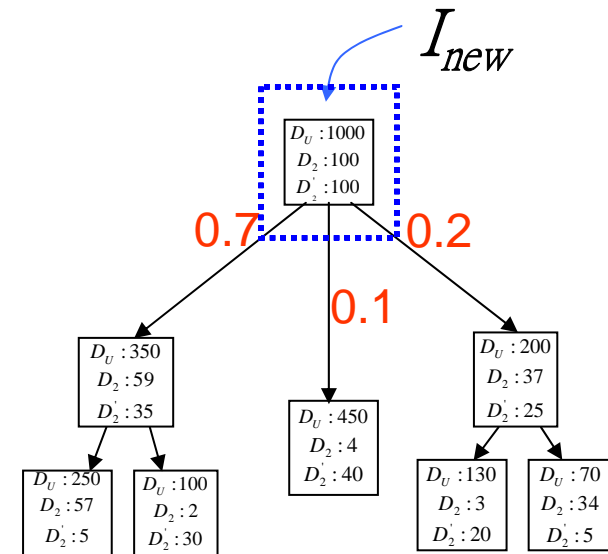
$\gamma(L_k, N_i^k | I_{new})$  is the confidence value of  $I_{new}$

associated with the sub-concept of  $L_k$  represented by  $N_i^k$

$p(N_i^k | I_{new})$  is the possibility that  $I_{new}$  belongs to the node  $N_i^k$

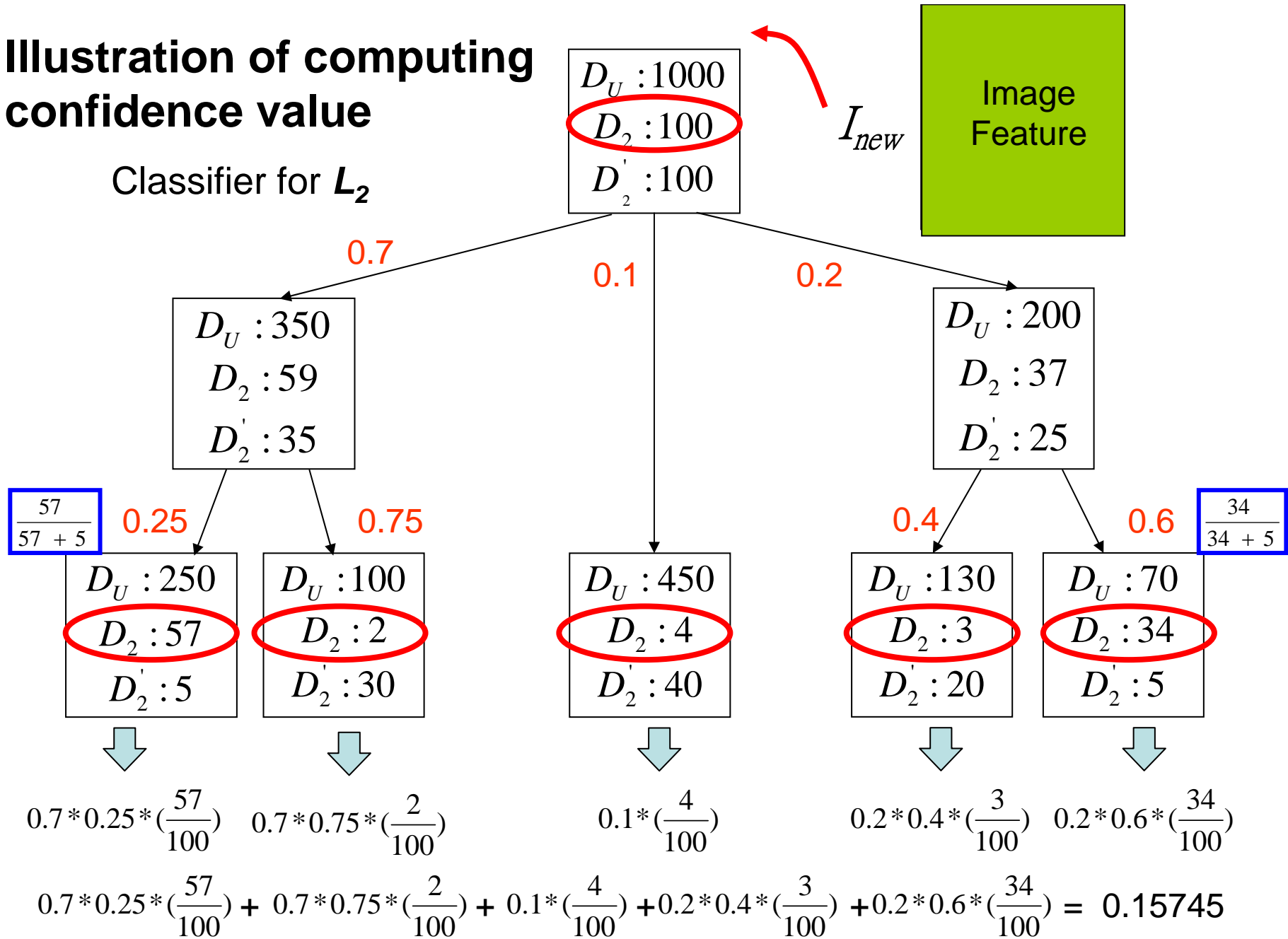
$$\gamma(L_k, N_i^k | I_{new}) = \begin{cases} \sum_{j=1}^{z_i^k} \gamma(L_k, Nc_{ij}^k | I_{new}) p(Nc_{ij}^k | I_{new}), & \text{if } N_i^k \text{ is not a leaf node} \\ \frac{d_i}{d_{root}}, & \text{if } N_i^k \text{ is a leaf node} \end{cases}$$

C. C. Chiang



# Illustration of computing confidence value

Classifier for  $L_2$



# Experiments

- Datasets
  - Barnard's set (Corel Photo)
  - 374 labels
  - 5000 images



plants  
close-up  
flowers  
leaf



petals  
plants



polar  
snow  
tundra  
bear



petals  
lily  
plants  
leaf



vegetation  
garden  
flowers  
plants



bear  
ice  
polar  
snow

# Experiments

- 15 labels
  - mountain, sky, water, clouds, trees, people, grass, buildings, plane, bear, snow, flowers, rocks, street, stone
- In out experiments, we roughly select 200 images for each label
- 1970 images from Barnard's set

Table 2: The labels used in the experiments and their original numbers of associated images.

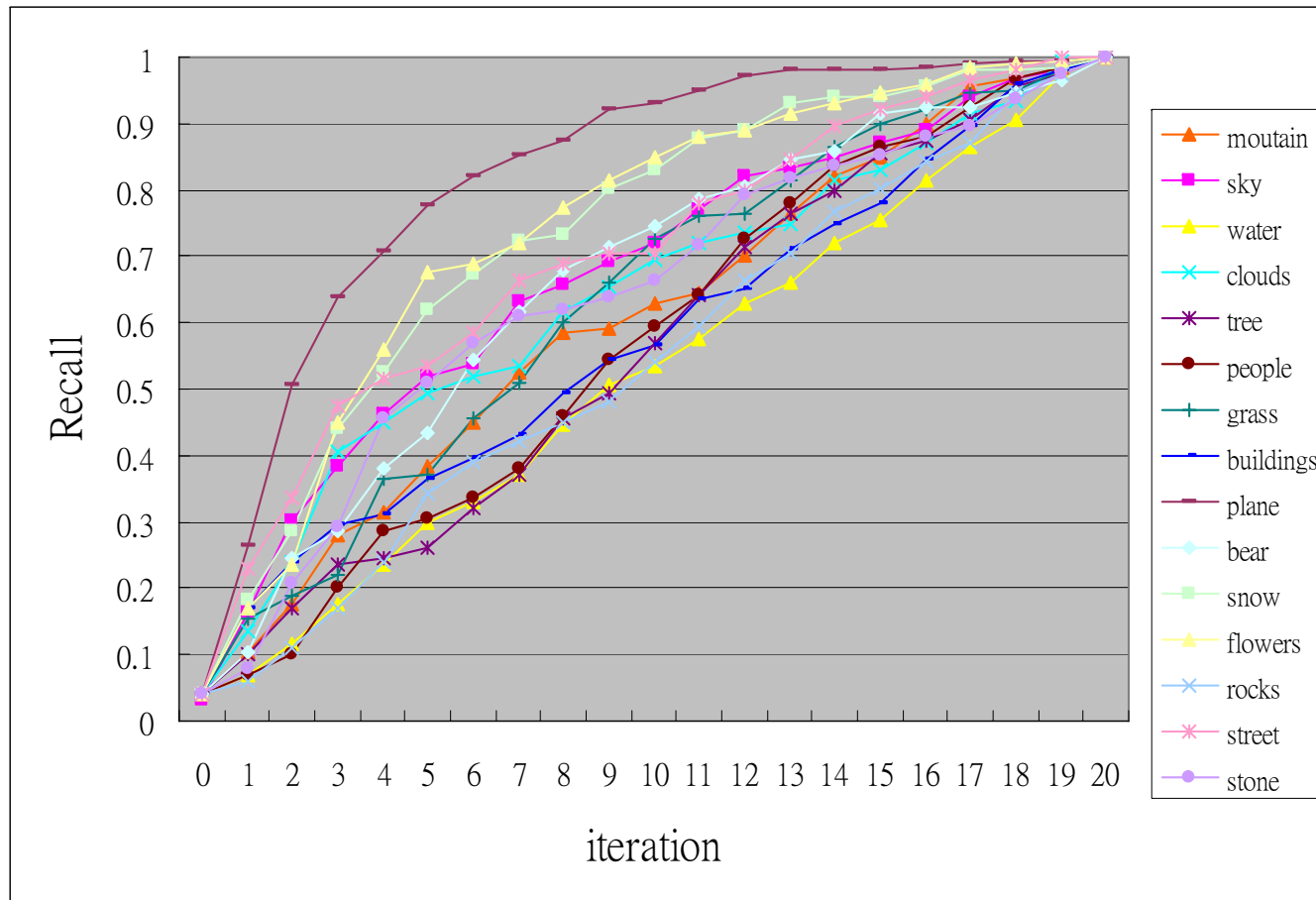
Label	# of images	label	# of images	label	# of images
Water	1120	sky	988	tree	948
People	744	grass	497	buildings	462
mountain	345	snow	298	flowers	296
clouds	280	rocks	250	stone	232
street	229	plane	224	bear	220

# Experiments

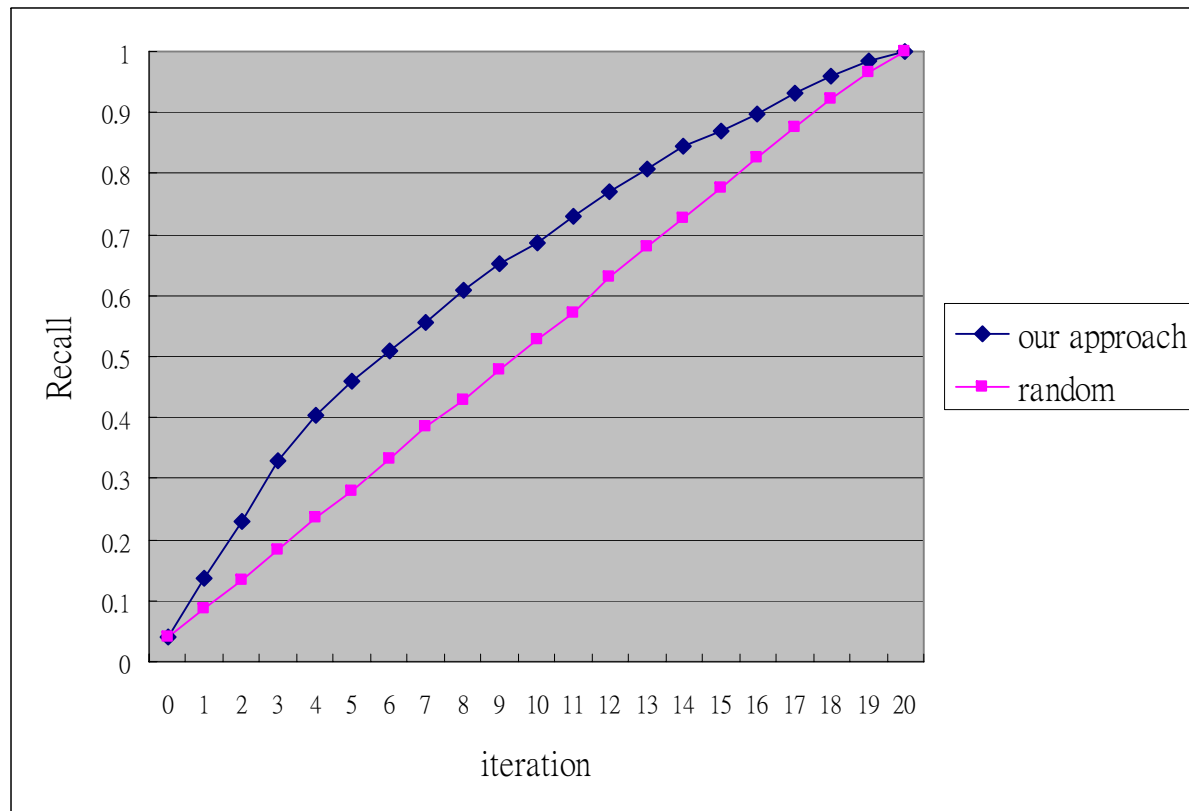
- Performance Evaluation
  - precision =  $A/B$
  - recall =  $A/C$ 
    - **A** is the number of images correctly annotated with the given word
    - **B** is the number of images automatically annotated with the given word
    - **C** is the number of images having that word in ground-truth annotation
  - $F_1$  value
    - $$\frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

# Experiments

The recalls of 15 labels



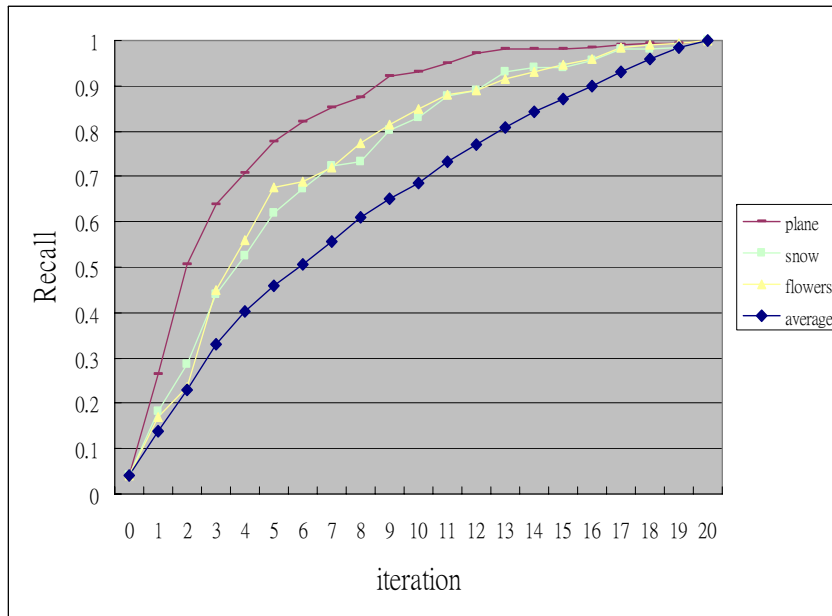
# Experiments



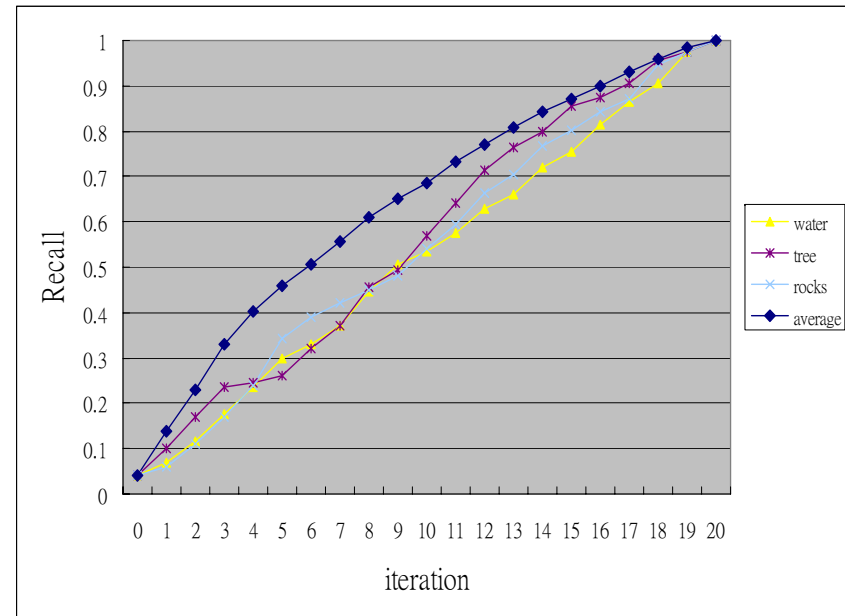
- our approach
  - average of recalls of 15 labels that using our approach to choose  $N$  images for user
- random
  - average of recalls of 15 labels that randomly choosing  $N$  images for user

# Experiments

## Best three cases



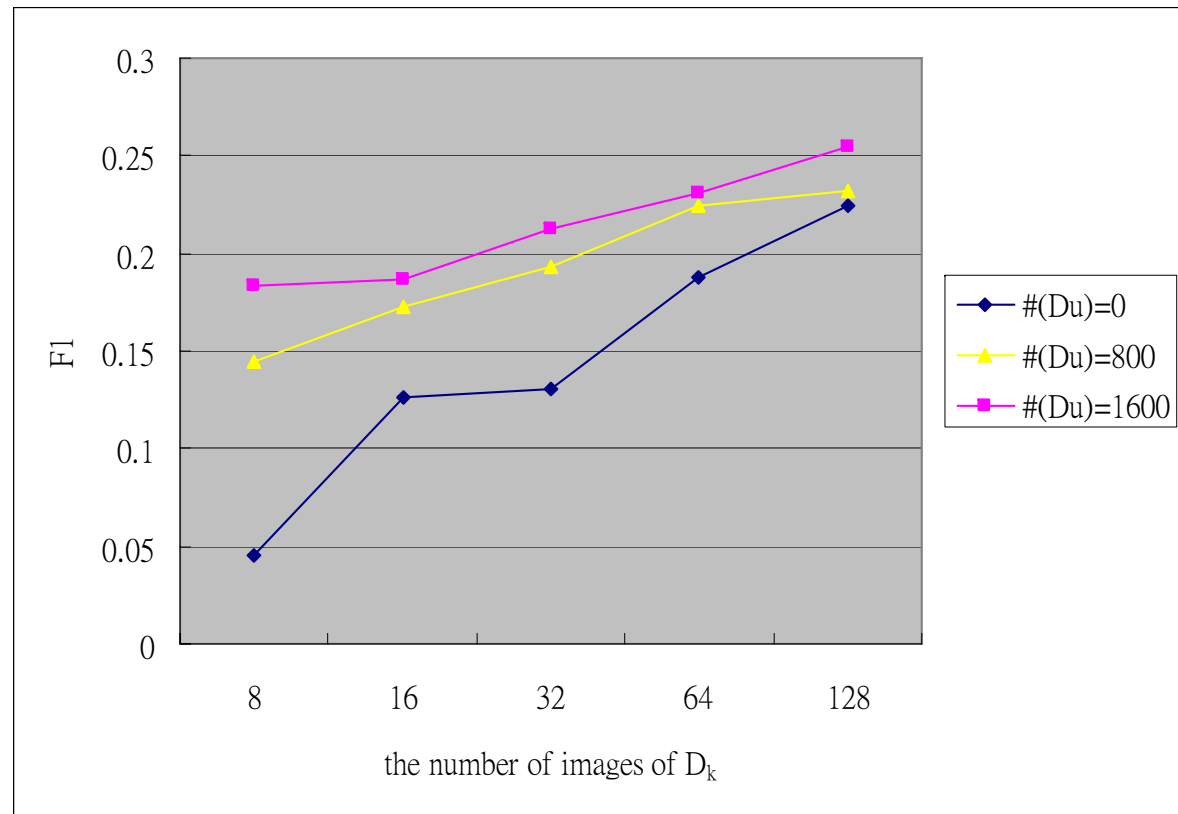
## Worst three cases



# Experiments

The  $F_1$  values computed from classifiers with different number of training images

- $\#(D_u)$  is the number of unlabeled images



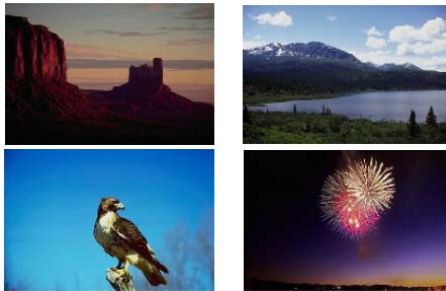
# Conclusion

- We proposed a method for image annotation with user feedback
  - Semi-supervised learning
    - utilizing unlabeled images to help clustering
  - Hierarchical classifiers
    - each of them is associated to one label

# Conclusion & Future Work

- Divergence of semantic concept

- Label: “sky”



## Hierarchical Classifier

group several labels instead  
one single label at one time

- Ambiguity of semantic concept

- Different descriptions by different people



flower, leaves

plants, petals

## Ontological Information

## User Feedbacks

- Less training data

- Difficult to prepare a huge number of labeled images

## Semi-Supervised Learning

# Future Work

- Prior information: ontological structure
- A better computation for the confidence value
- Larger dataset
- Application to image retrieval
- Contest of TRECVID or CLEF

**Thanks for Your Attention!!**