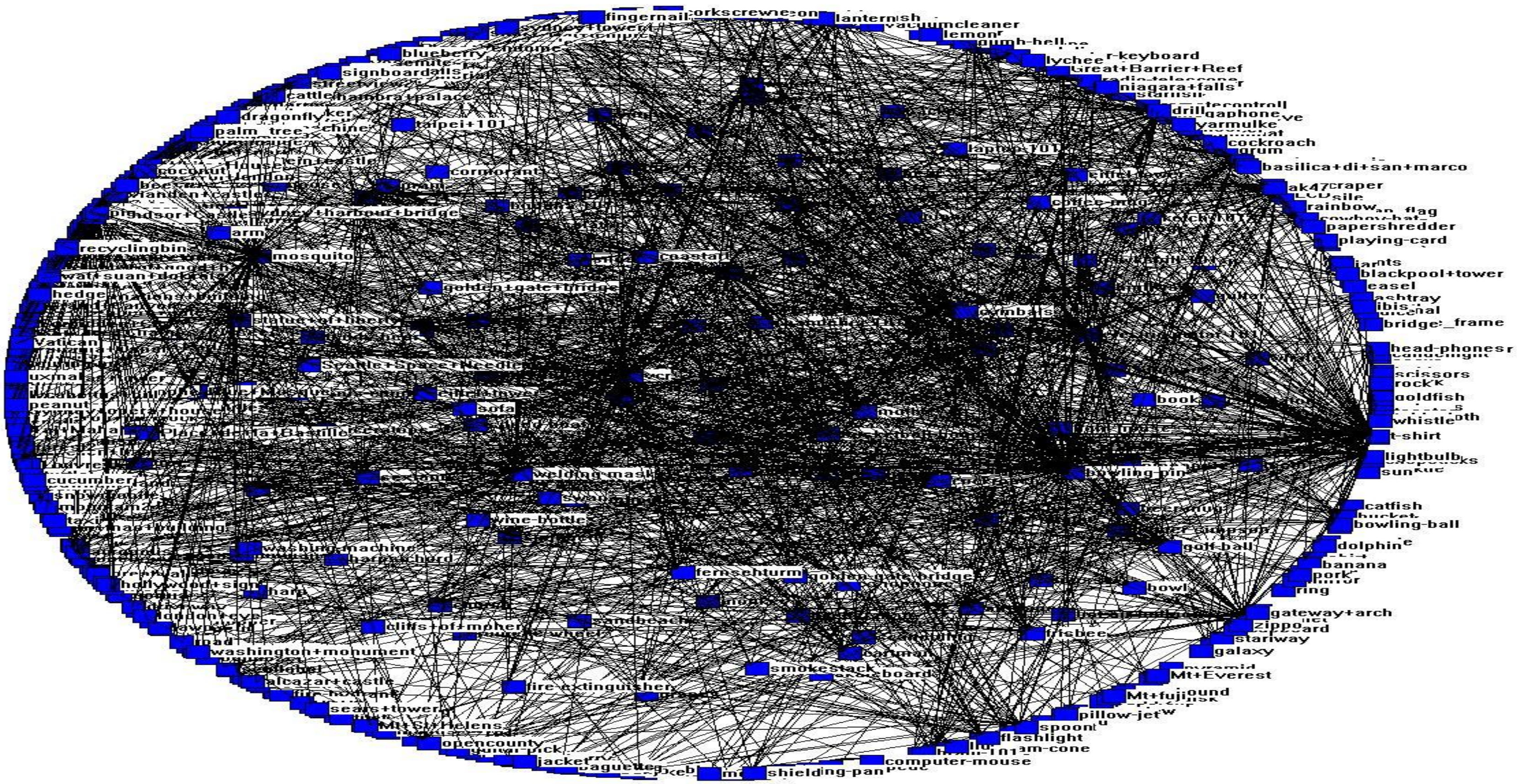
Hierarchical Image Classification over Visual Tree

Jianping Fan
Dept of Computer Science
UNC-Charlotte, NC 28223
www.cs.uncc.edu/~jfan

1. Problems to be addressed

- **Large-Scale Visual Recognition**



Inter-Object Visual Correlations rather than independency

1. Problems to be addressed

- **Large-Scale Visual Recognition: Challenges**
  - **We need to learn large amounts of classifiers for large-scale visual recognition!**
  - **Some object classes and image concepts are visually-related and hard to be discriminated!**
  - **Some object classes and image concepts may have huge inner-concept visual diversity!**

1. Problems to be addressed

- **Large-Scale Visual Recognition: Challenges**
  - **Huge inner-concept visual diversity**

    ---simple models may not work, but using complex models may overlap with others!
  - **Huge inter-concept visual similarity**

    ---training complexity will increase for distinguishing visually-related concepts!
  - **Huge computational cost**

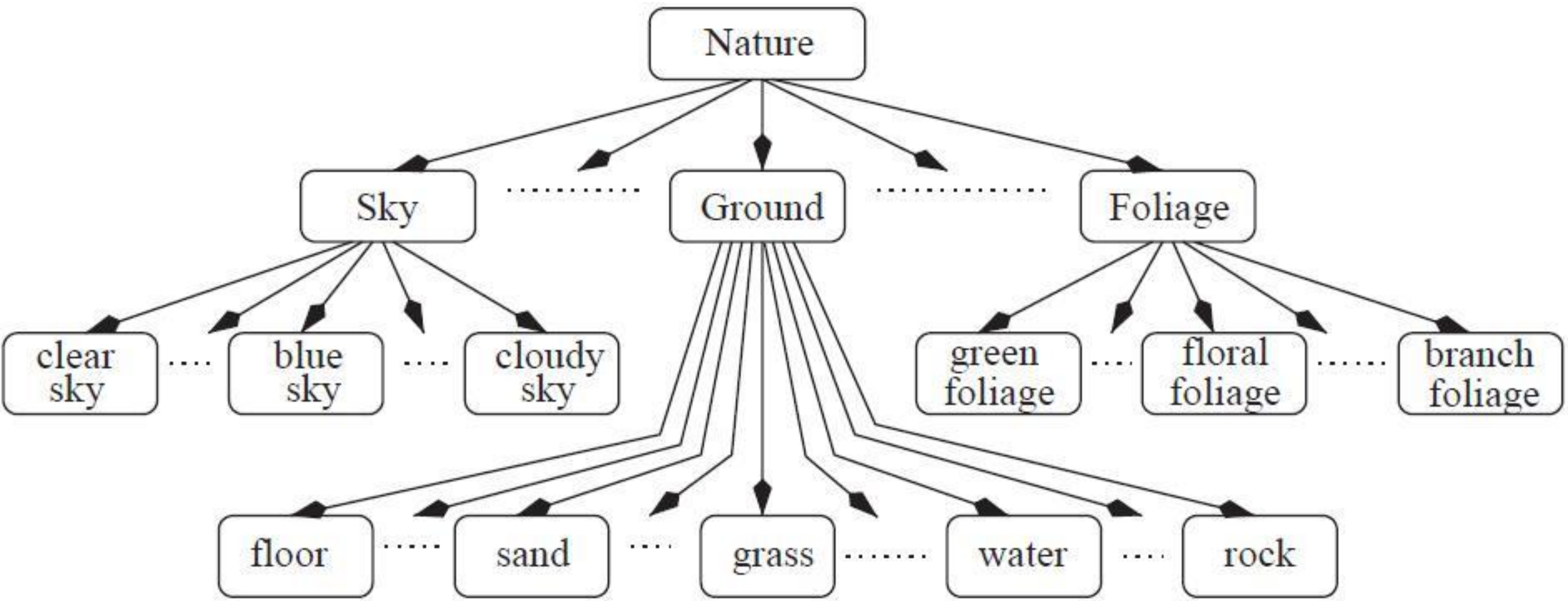
    ---thousands of inter-related classifiers should be trained jointly!

1. Problems to be addressed

- **Large-Scale Visual Recognition**
  - **How to leverage social images for classifier training?**
  - **How to leverage inter-class correlations for dictionary learning & classifier training?**
  - **How to deal with inter-level error propagation?**

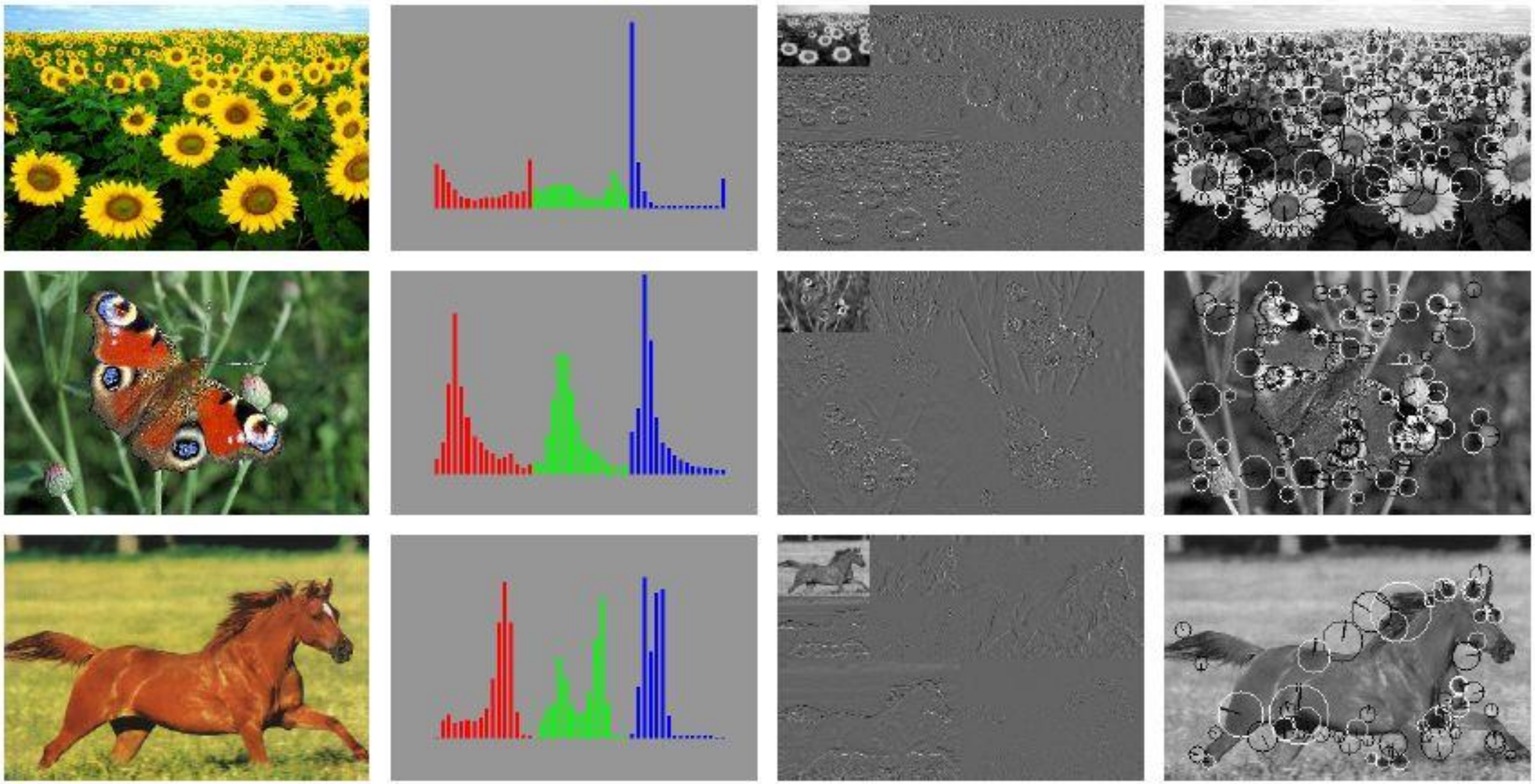
2. Collecting Large-Scale Training Images

- **Flickr Images & Other Image Sites**
  - **Keywords for image crawling**



2. Collecting Large-Scale Training Images

- **Visual Feature Extraction**



2. Collecting Large-Scale Training Images

- **Synonymous Concepts: Visual Similarity**

CVPR2010

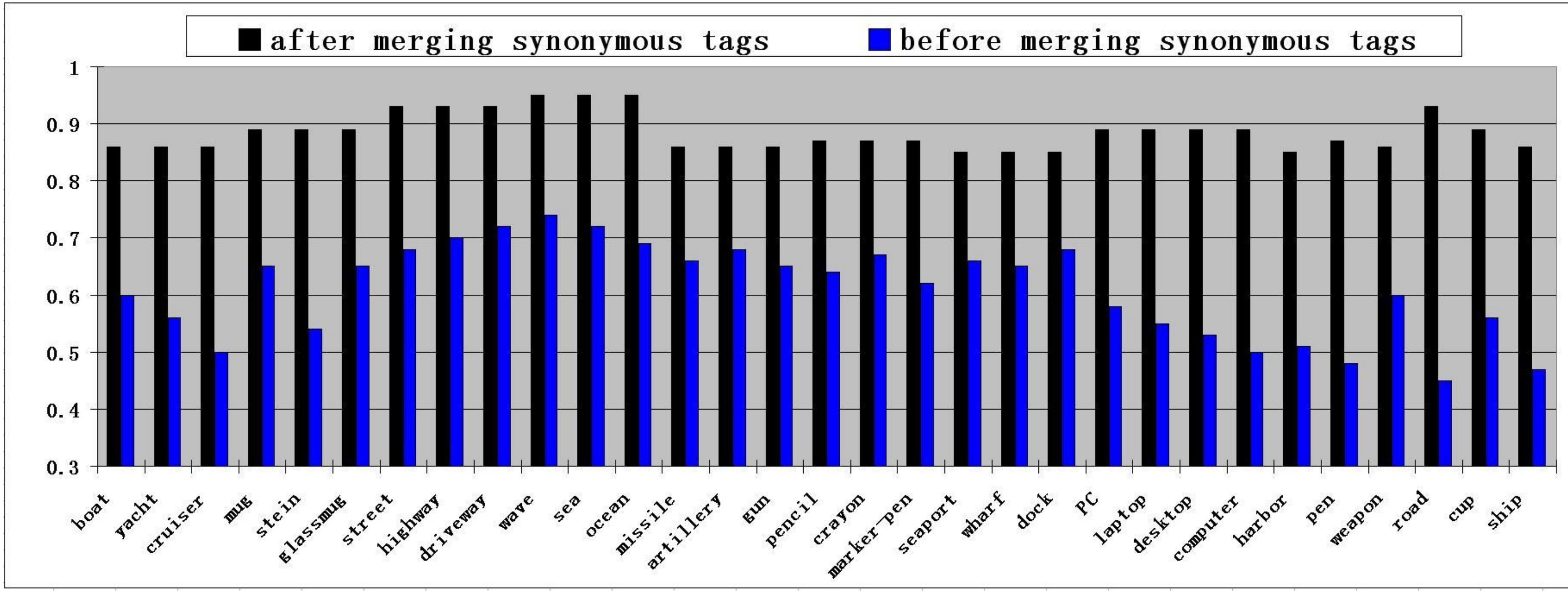
(a) Auto

(b) Automobile

(c) Car

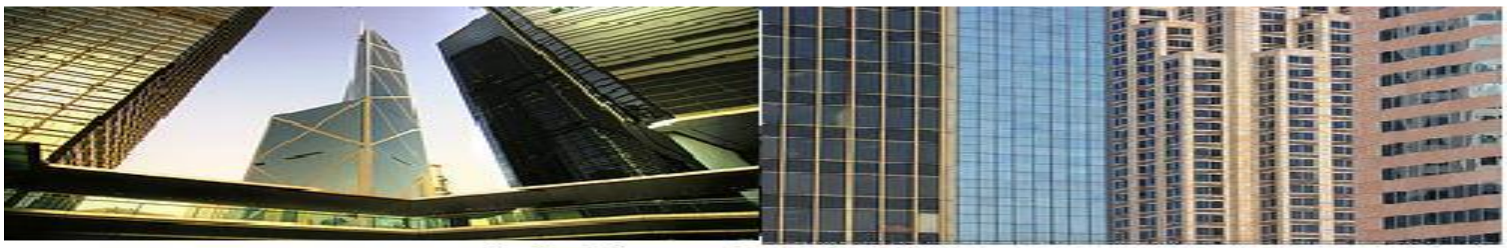
2. Collecting Large-Scale Training Images

- **Synonymous Concepts: Visual Similarity**



2. Collecting Large-Scale Training Images

- **Ambiguous Concept: Visual Diversity**



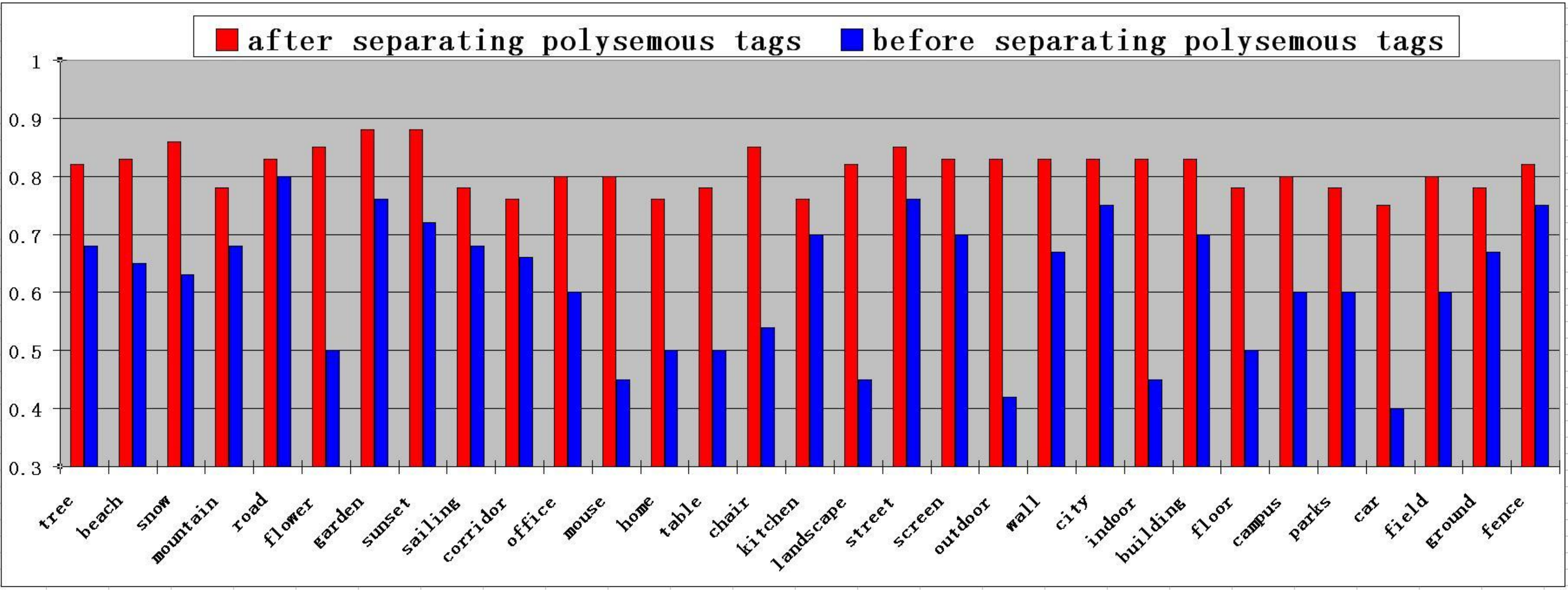
(a) Bank Office

(b) River Bank

(c) Cloud Bank

2. Collecting Large-Scale Training Images

- **Ambiguous Concept: Visual Diversity**



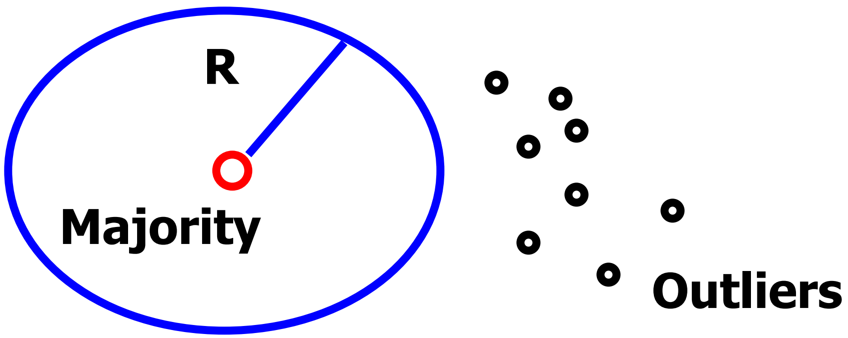
2. Collecting Large-Scale Training Images

- **Junk Image Filtering**

IEEE Trans. CSVT 2009

$$\forall_{j=1}^{N}: \quad \|\phi(x_j) - \mu^{\phi}\|^2 \leq R^2$$

$$\mu^{\phi} = \sum_{j=1}^{N} \phi(x_j)$$



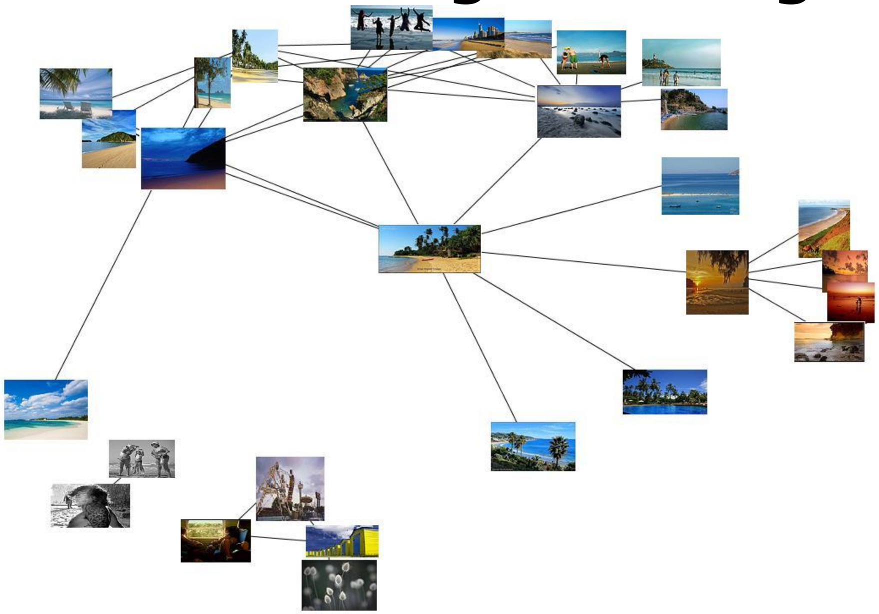
$$min\left\{R^2 + \frac{C}{N}\sum_{j=1}^{N}\xi_j\right\}$$ **subject to:** $$\forall_{j=1}^{N}: \quad \|\phi(x_j) - \mu^{\phi}\|^2 \leq R^2 + \xi_j, \qquad \xi_j \geq 0$$

Decision function:

$$f(x) = R^2 - \sum_{i,j}^{N} \alpha_i \alpha_j \kappa(x_i, x_j) + 2\sum_{j}^{N} \alpha_j \kappa(x_j, x) - \kappa(x, x)$$

2. Collecting Large-Scale Training Images

- **Junk Image Filtering**

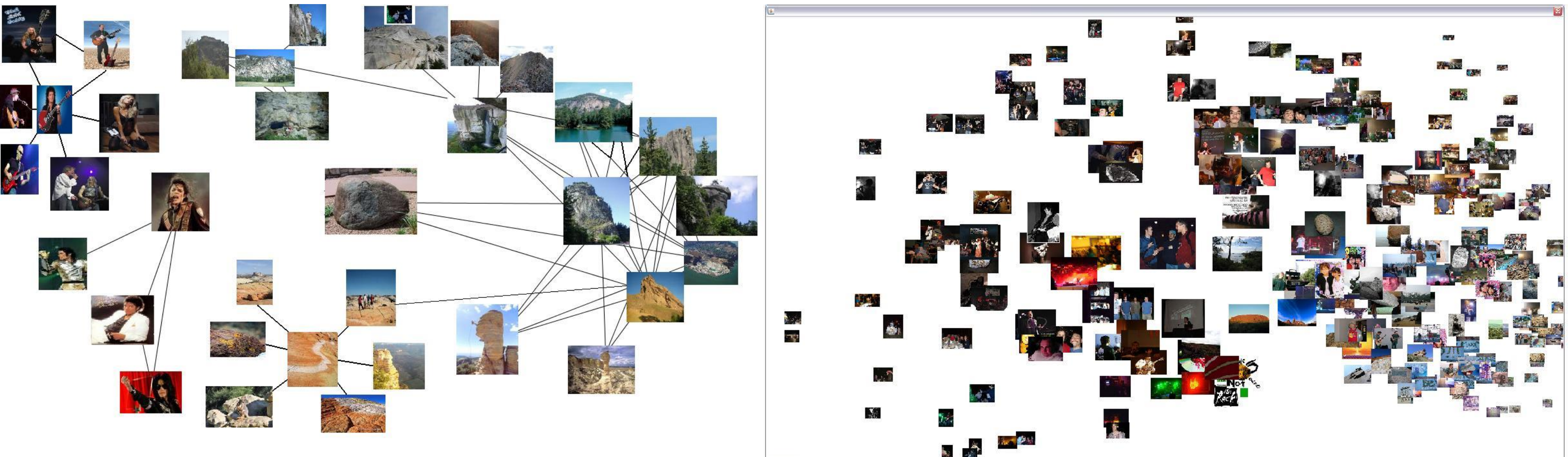


(a)

(b)

2. Collecting Large-Scale Training Images

- **Junk Image Filtering**



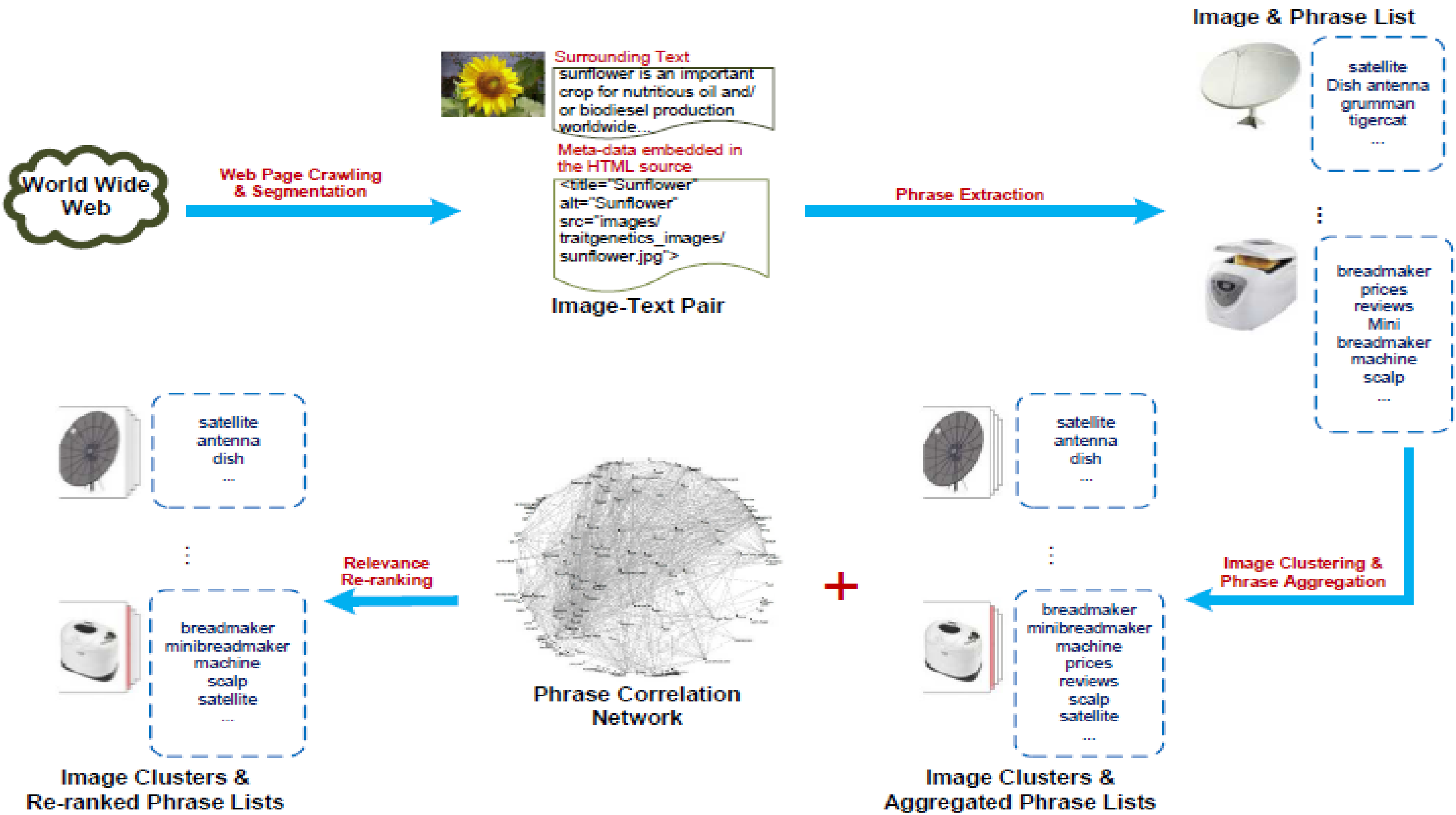
(a) (b)

- Most text terms are weakly related or even irrelevant to web images in the same webpage

Noise image

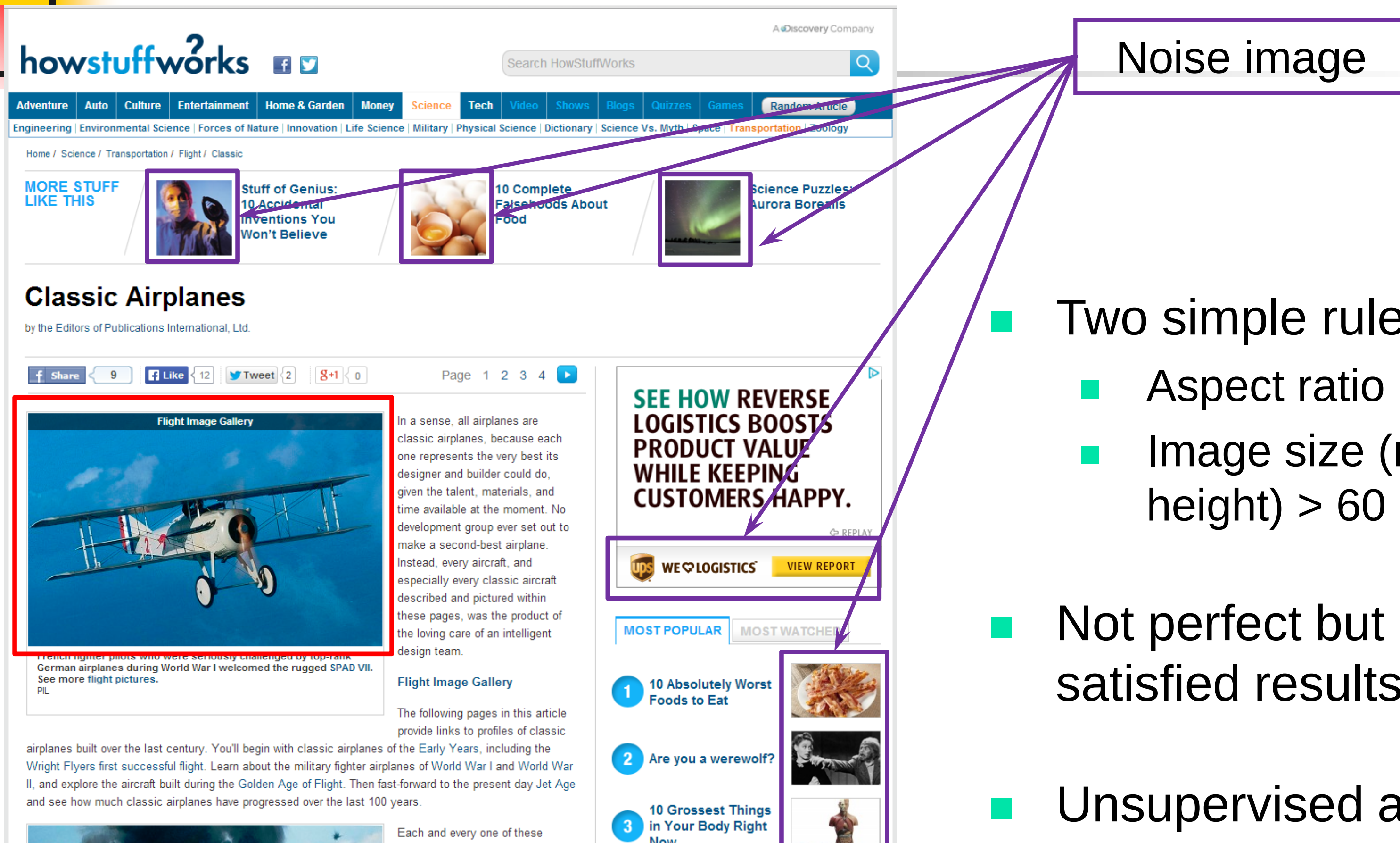
tiger
big cat
Southeast Asia
Russian
Chinese
Bengal
Siberian
Indochinese
South Chinese
....

Text-Image Alignment for Web Image Indexing



WWW2010, PR2014

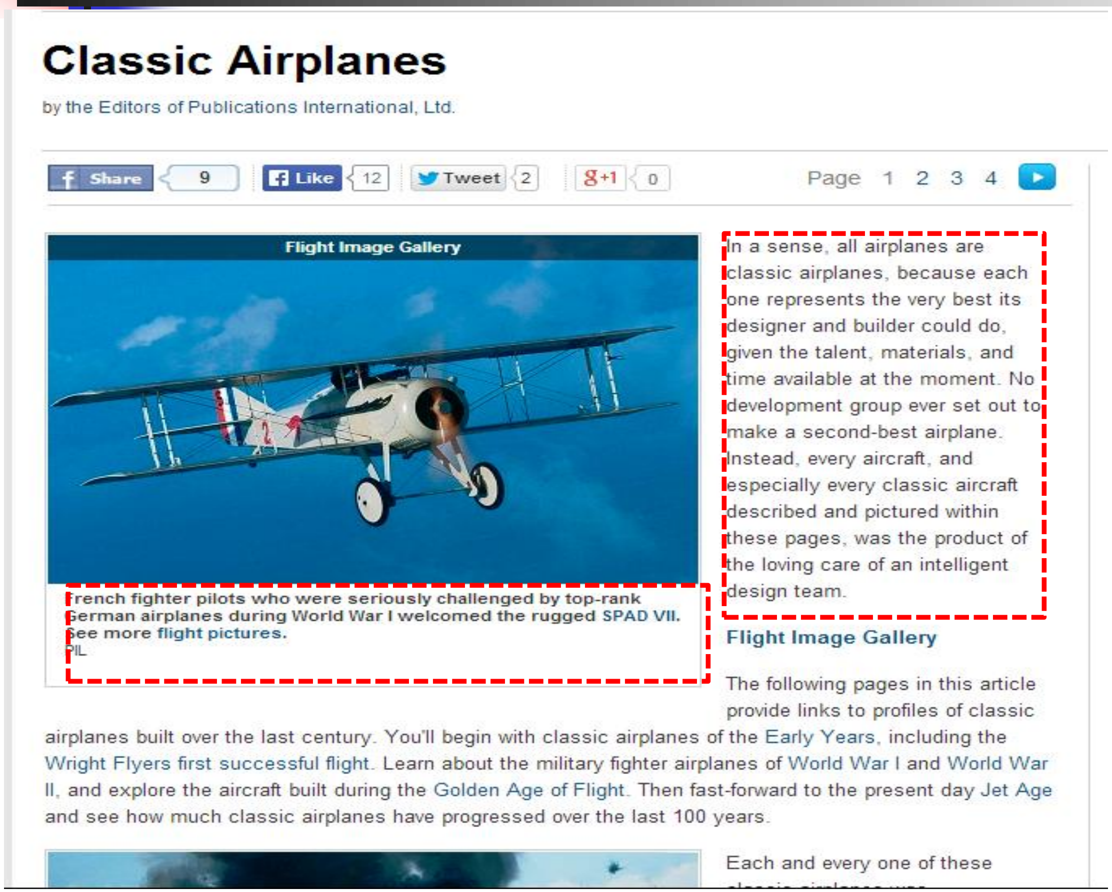
Informative Image Extraction



- Two simple rules:
  - Aspect ratio (>0.2 or <5)
  - Image size (min(width, height) > 60 pixel)
- Not perfect but produce satisfied results
- Unsupervised and computationally efficient

Webpage Segmentation

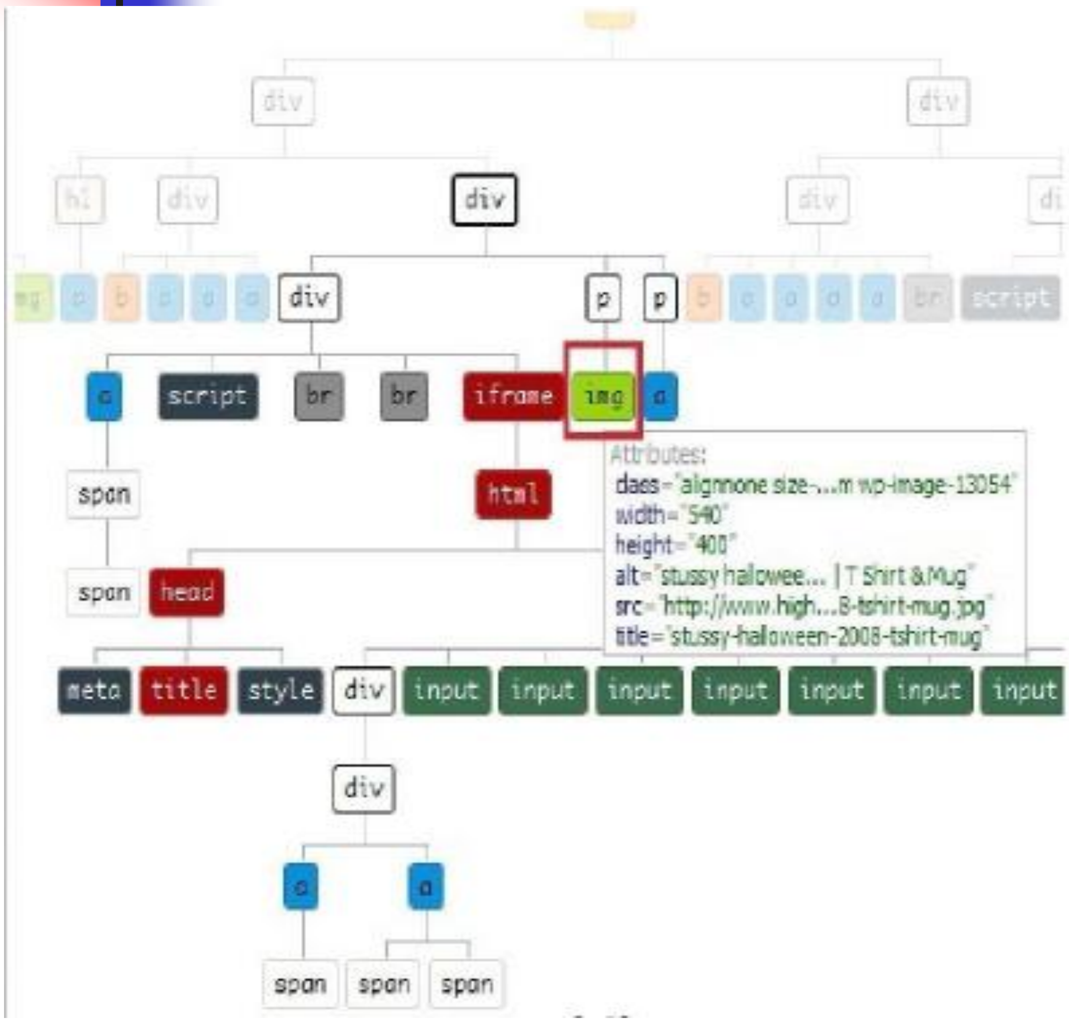
- Surrounding Text Extraction



- Visual-based algorithm
  - precise but expensive
  - [Cai et al. MSR-TR'03]
- DOM (Document Object Model) based method
  - computationally efficien

Webpage Segmentation

- Surrounding Text Extraction



- Visual-based algorithm
  - precise but expensive
  - [Cai et al. MSR-TR'03]
- DOM (Document Object Model) based method
  - computationally efficient

Text-Image Alignment for Web Image Indexing

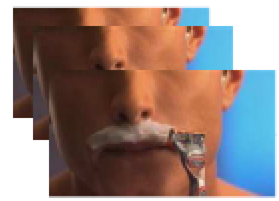
Cluster No.: 3598, 10 duplicates

Phrase list 1: sterilization equipment, water, sterilizer, china mainland
Phrase list 2: autoclave, sterilizer, water, china mainland, manufacturer
Phrase list 3: retort, heating, sterilizer, water, china mainland, manufacturer
Phrase list 4: sterilizer, water, china mainland, manufacturer
Phrase list 5: sterilization equipment, water, sterilizer, china mainland, manufacturer
....
Aggregation: sterilizer, sterilization equipment, water, retort, manufacturer, ...

Cluster No.: 6244, 13 duplicates

Phrase list 1: cimarron, roper, saddle, roper saddle, horse, ...
Phrase list 2: cimarron, roper, saddle, roper saddle,...
Phrase list 3: saddle, roper, roper saddle, horse, sale
Phrase list 4: roper saddle, saddle, cimarron, horse
....
Aggregation: saddle, roper, roper saddle, cimarron, ...

Cluster No.: 16263, 33 duplicates



Phrase list 1: face, area, drive stick, rule safety
Phrase list 2: face, grip, play tennis, tennis racket
Phrase list 3: face, , tennis racket, maintenance
Phrase list 4: face, shaver, tennis preparation tip,.
....
Aggregation: face, shaver, gillete, ...

Cluster No.: 29906, 8 duplicates

Phrase list 1: pisa feb, pisa, leaning tower, location, photo
Phrase list 2: pisa, leaning tower, location, photo
Phrase list 3: pisa, location, leaning tower, photo
Phrase list 4: pisa, leaning tower, photo....
Aggregation: pisa, learning tower, pisa feb, location, ...

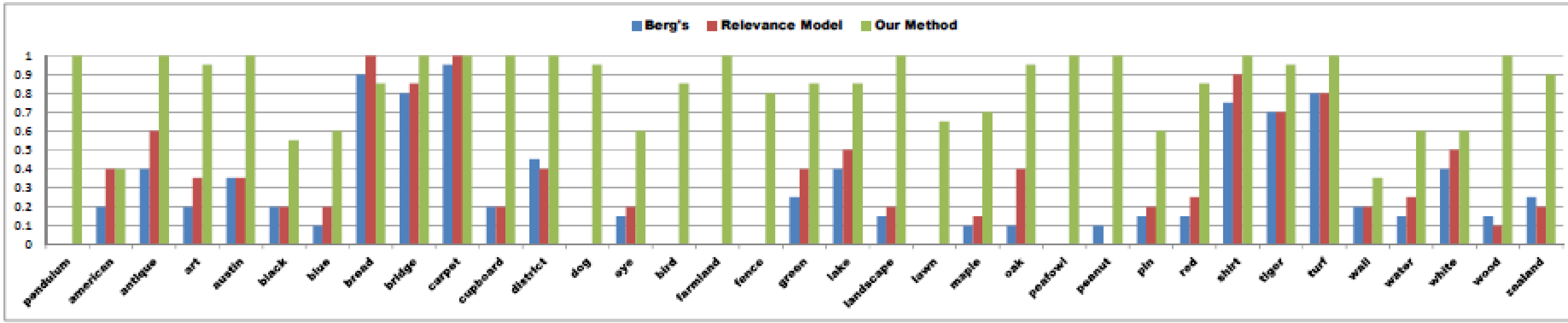
Cluster No.: 35950, 27 duplicates

Phrase list 1: venture snowmobile, indonesia
Phrase list 2: venture snowmobile, arctic, snowmobile, ...
Phrase list 3: venture snowmobile, snowmobile
Phrase list 4: venture snowmobile, snowmobile manufacture
Aggregation: venture snowmobile, snowmobile, ...

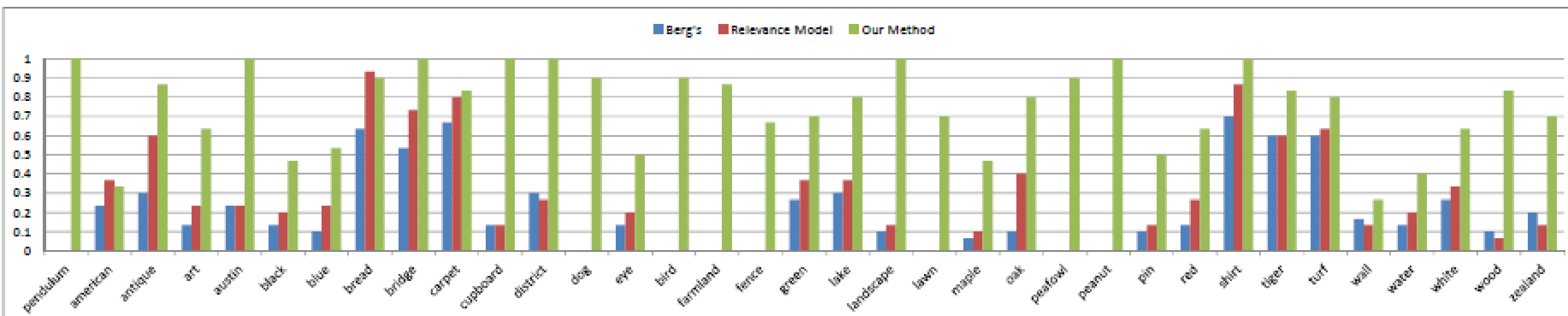
Near-duplicates share similar semantics!

WWW2010, PR2014

Text-Image Alignment for Web Image Indexing



(b) Pre@20, top 20 images are evaluated.



(c) Pre@30, top 30 images are evaluated.

WWW2010, PR2014

2. Collecting Large-Scale Training Images

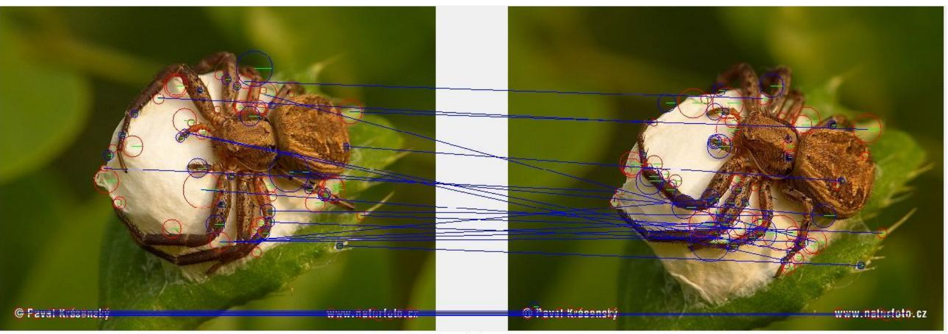
- **Duplicate Detection**

Duplicates may mislead classifier training tools!

CVPR2012

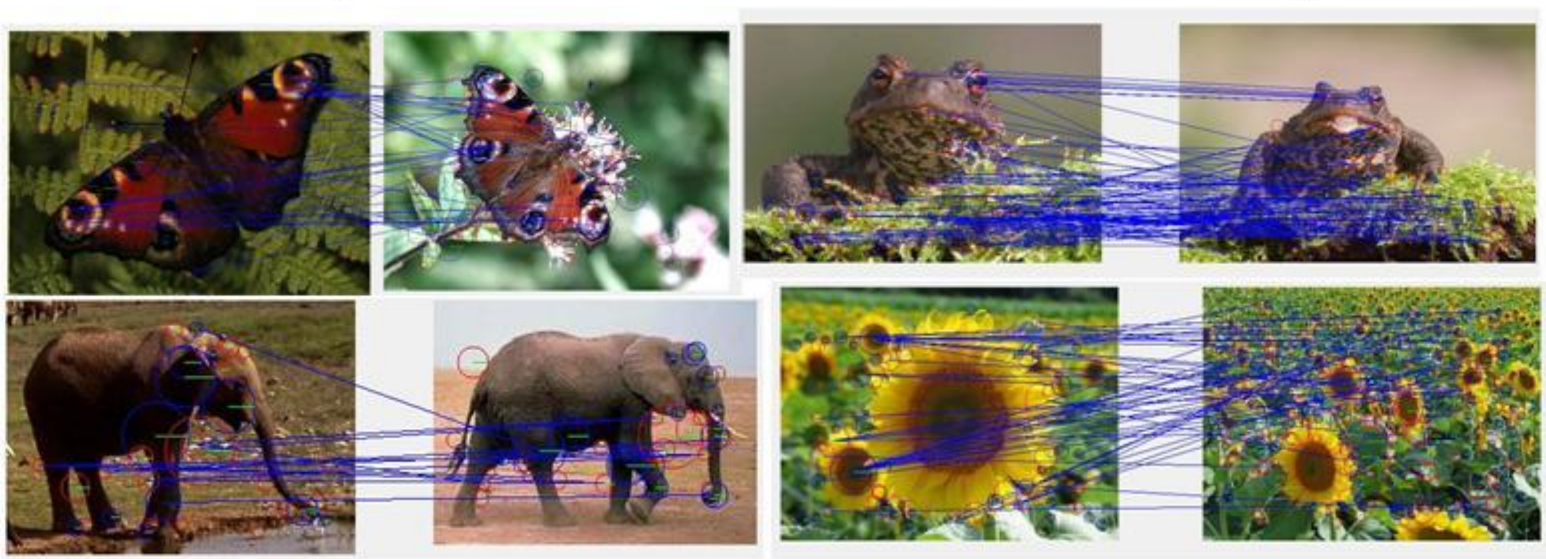
2. Collecting Large-Scale Training Images

- **Duplicate Detection**



(a)

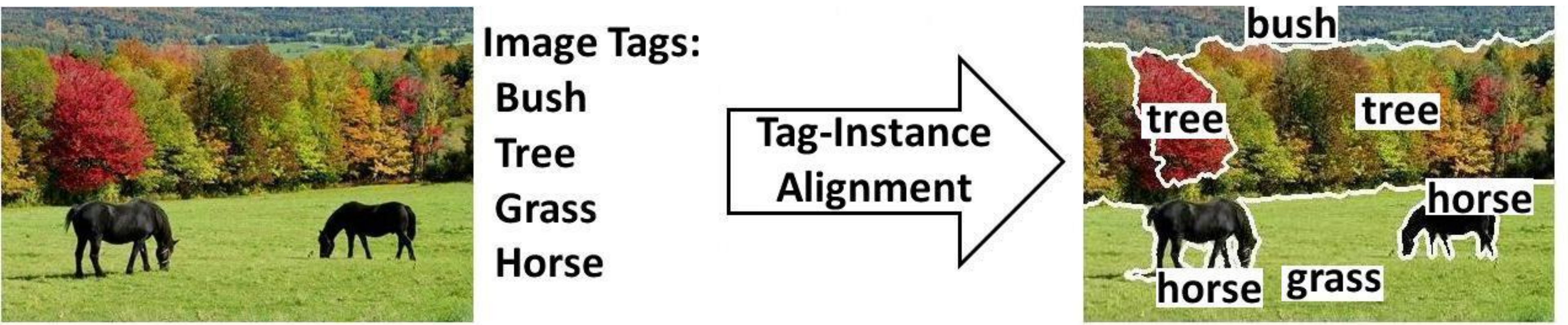
(b)



2. Collecting Large-Scale Training Images

- **Automatic Tag-Instance Alignment**

ACM MM 2010



(a) Original Image (b) Multiple Image Instances

- **Missing Tag Prediction**

CVPR 2012

3. Visual Concept Network

ACM MM2009

- **Why we need visual concept network?**

  ---concept ontology, object co-occurrence network, ….

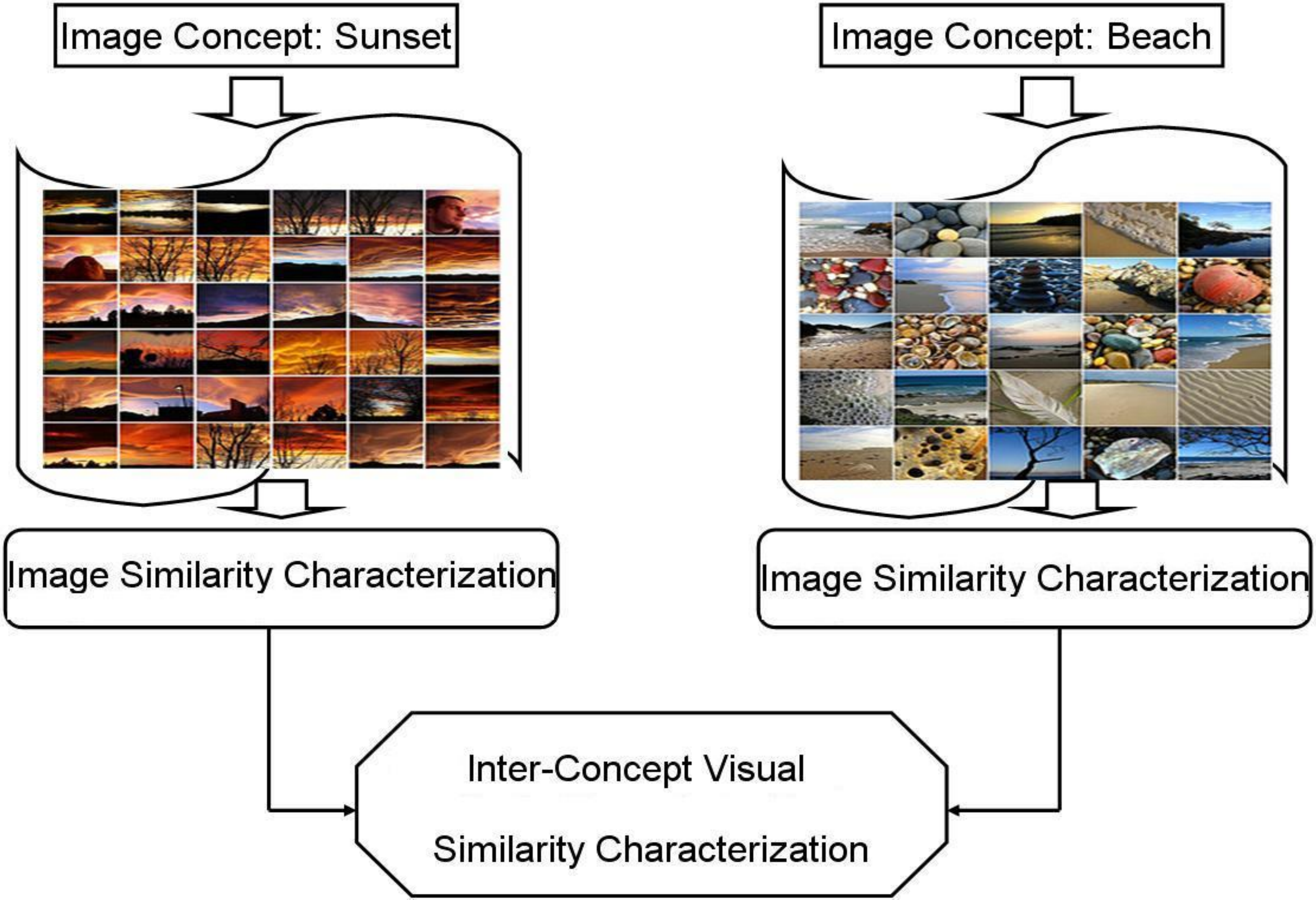
- **Common space: classifier training & concept detection**

  ---visual feature space rather than label space or concept space

We need to characterize inter-concept visual correlations rather than others!

Inter-related learning task determination

3. Visual Concept Network



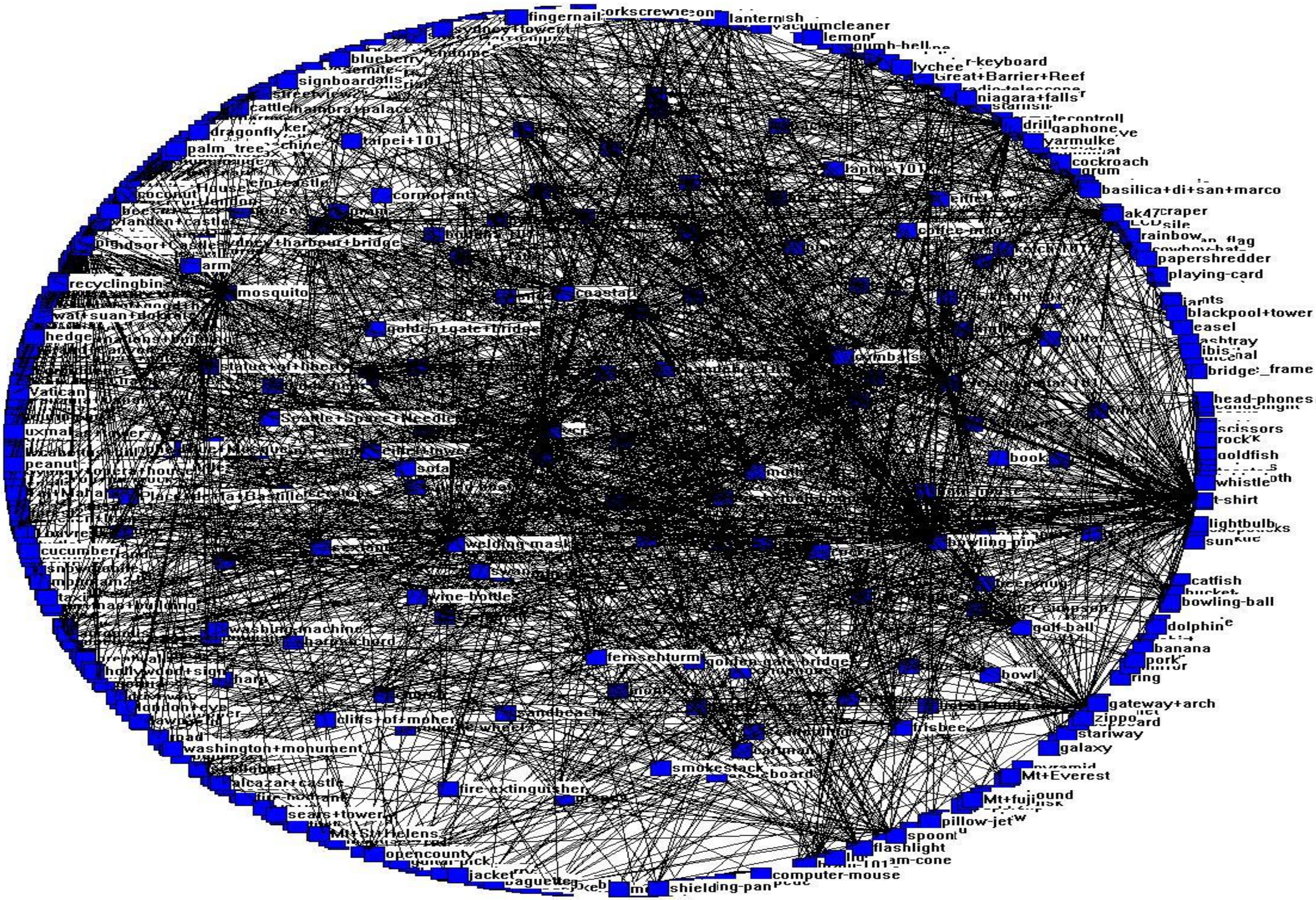
3. Visual Concept Network

concept pair	γ	concept pair	γ	concept pair	γ	concept pair	γ
urbanroad-streetview	0.99	cat-dog	0.81	kerb-saucer	0.28	tweezer-corn	0.19
frisbee-pizza	0.80	dolphin-cruiser	0.73	fridge-vest	0.29	journal-grape	0.19
moped-bus	0.75	habor-outview	0.71	stick-cupboard	0.29	sheep-greatwall	0.26
monkey-humanface	0.71	guitar-violin	0.71	mushroom-moon	0.32	whistle-watermelon	0.28
lightbulb-firework	0.69	mango-broccoli	0.69	cannon-ruler	0.41	snake-ipod	0.31
porcupine-lion	0.68	bridge-warship	0.68	tombstone-crab	0.42	helicopter-city	0.63
doorway-street	0.65	statue-building	0.68	pylon-highway	0.61	LCD-container	0.65
windmill-bigben	0.63	cat-lion	0.66	beermug-bar	0.62	sailboat-cruiser	0.66

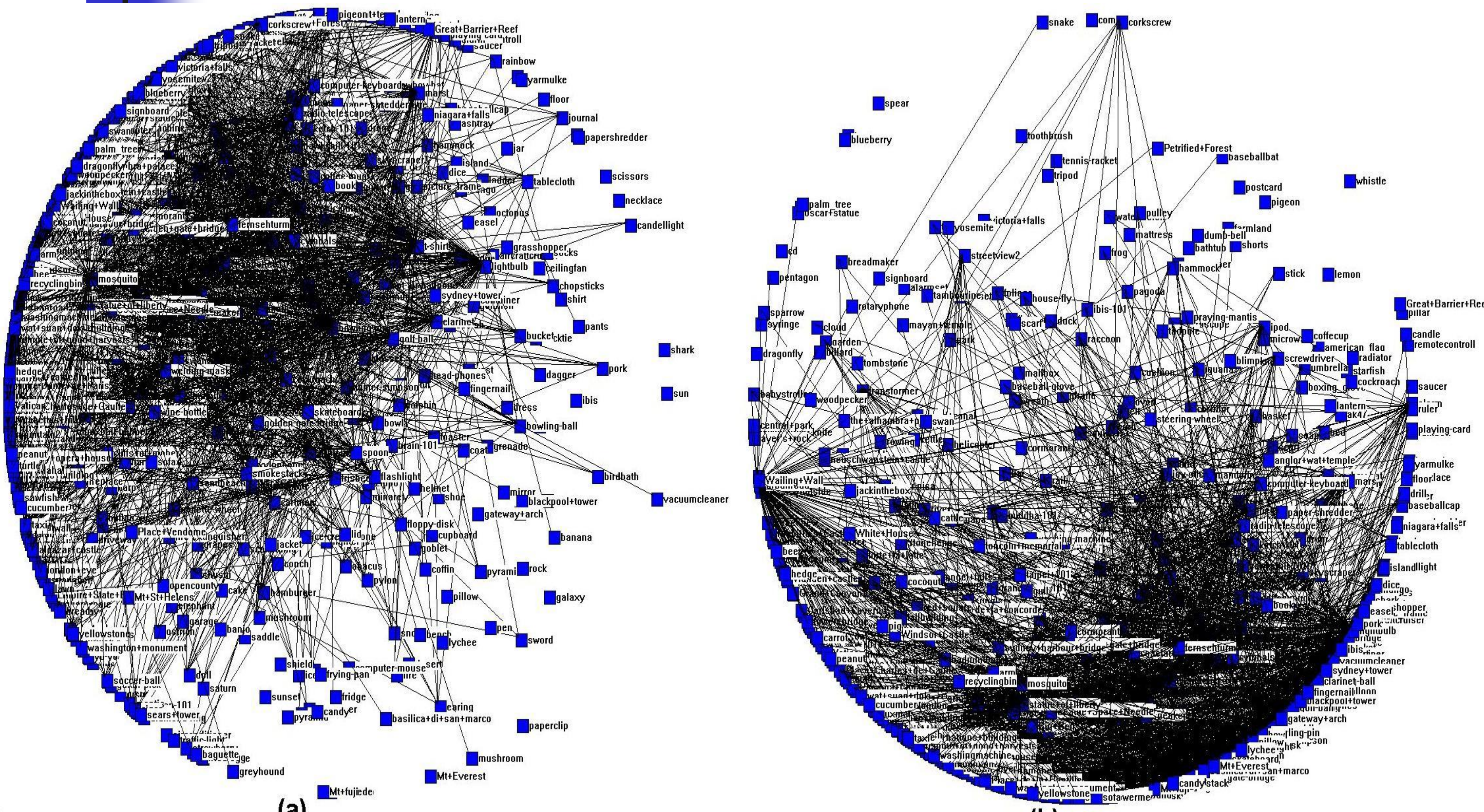
3. Visual Concept Network

sky		0.62	0.60	0.44	0.38	0.44	0.32	0.44	0.38	0.46	0.46	0.43	0.51
cloud	0.62		0.62	0.45	0.38	0.45	0.33	0.46	0.39	0.47	0.46	0.44	0.52
wave	0.60	0.62		0.44	0.36	0.44	0.30	0.44	0.36	0.47	0.46	0.44	0.53
maillot	0.44	0.45	0.44		0.59	0.62	0.57	0.61	0.56	0.58	0.60	0.55	0.60
pajama	0.38	0.38	0.36	0.59		0.62	0.59	0.61	0.57	0.58	0.60	0.54	0.58
short pants	0.44	0.45	0.44	0.62	0.62		0.61	0.65	0.60	0.61	0.63	0.57	0.62
oilskin	0.32	0.33	0.30	0.57	0.59	0.61		0.58	0.56	0.54	0.58	0.51	0.53
pullover	0.44	0.46	0.44	0.61	0.61	0.65	0.58		0.62	0.59	0.62	0.55	0.63
stole	0.38	0.39	0.36	0.56	0.57	0.60	0.56	0.62		0.52	0.55	0.48	0.56
bookcase	0.46	0.47	0.47	0.58	0.58	0.61	0.54	0.59	0.52		0.78	0.69	0.70
china cabinet	0.46	0.46	0.46	0.60	0.60	0.63	0.58	0.62	0.55	0.78		0.70	0.71
medicine chest	0.43	0.44	0.44	0.55	0.54	0.57	0.51	0.55	0.48	0.69	0.70		0.64
mailbox	0.51	0.52	0.53	0.60	0.58	0.62	0.53	0.63	0.56	0.70	0.71	0.64	

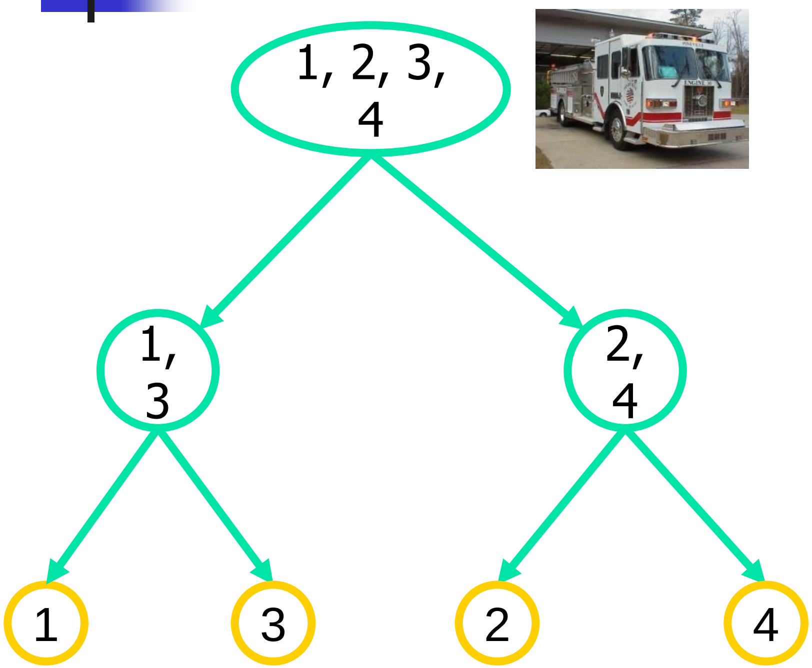
3. Visual Concept Network



3. Visual Concept Network



Label Tree for Efficient Classification



Label 1: cat
Label 3: dog

Label 2: mini van
Label 4: fire truck

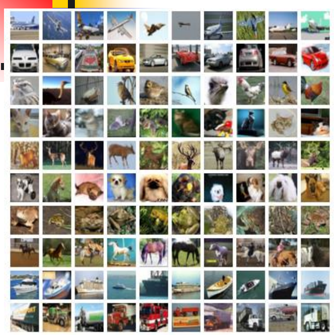
[Bengio et al. NIPS'2010]

Number of dot products needed in the label tree:
1 + 1 = 2

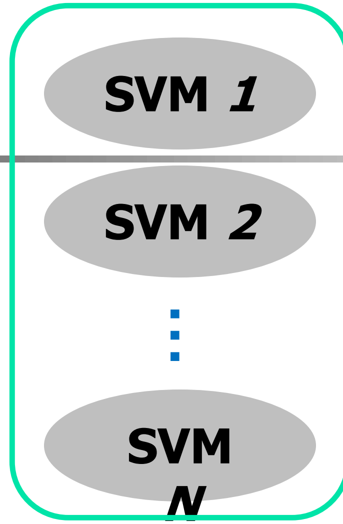
Number of dot product needed in a flat approach:
1 + 1 + 1 + 1 = 4

It is a fire truck!

Construction of Label Tree

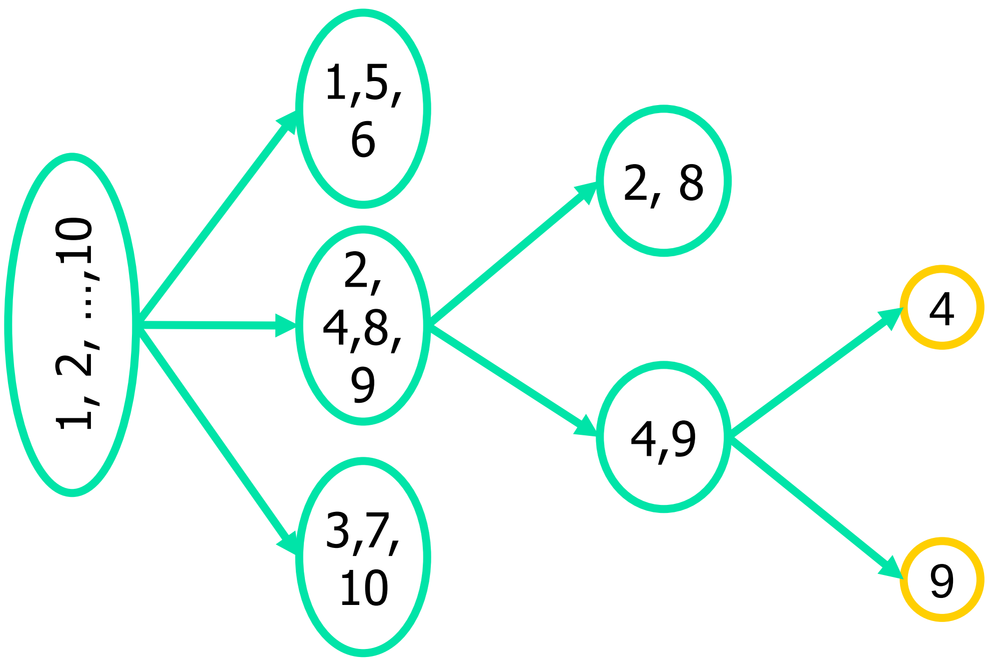


Testing Samples



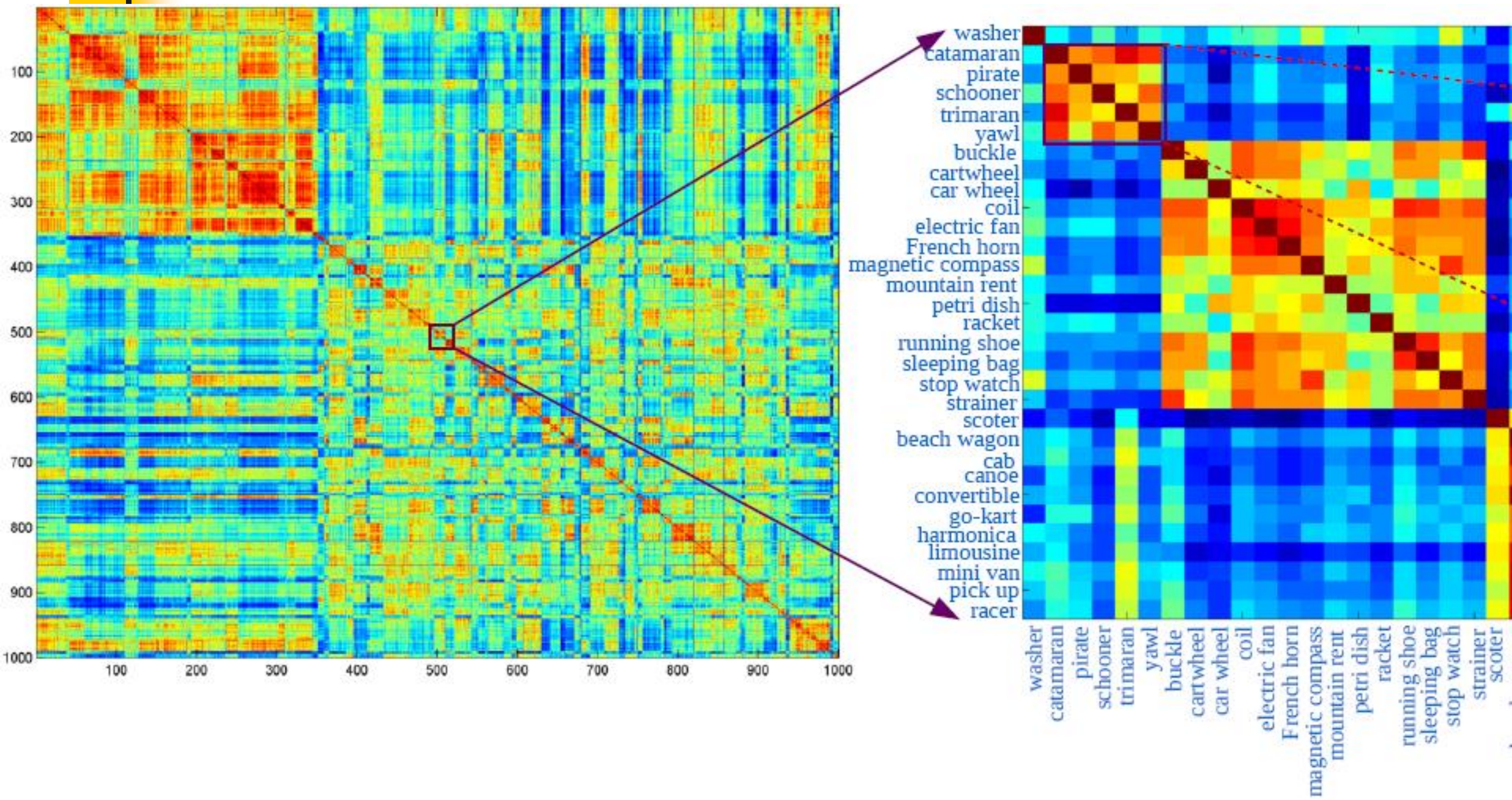
83.0	0.0	3.0	3.0	1.0	2.0	4.0	0.0	2.0	2.0
0.0	95.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	1.0
6.0	0.0	71.0	4.0	1.0	2.0	0.0	6.0	2.0	8.0
5.0	1.0	5.0	68.0	5.0	0.0	1.0	5.0	2.0	8.0
3.0	0.0	2.0	4.0	76.0	0.0	3.0	3.0	7.0	2.0
1.0	6.0	1.0	0.0	0.0	92.0	0.0	0.0	0.0	0.0
1.0	0.0	0.0	3.0	1.0	0.0	92.0	1.0	0.0	2.0
0.0	1.0	7.0	9.0	4.0	4.0	0.0	72.0	2.0	1.0
1.0	0.0	3.0	2.0	15.0	1.0	1.0	4.0	70.0	3.0
3.0	1.0	8.0	8.0	2.0	0.0	5.0	1.0	5.0	67.0

Confusion Matrix



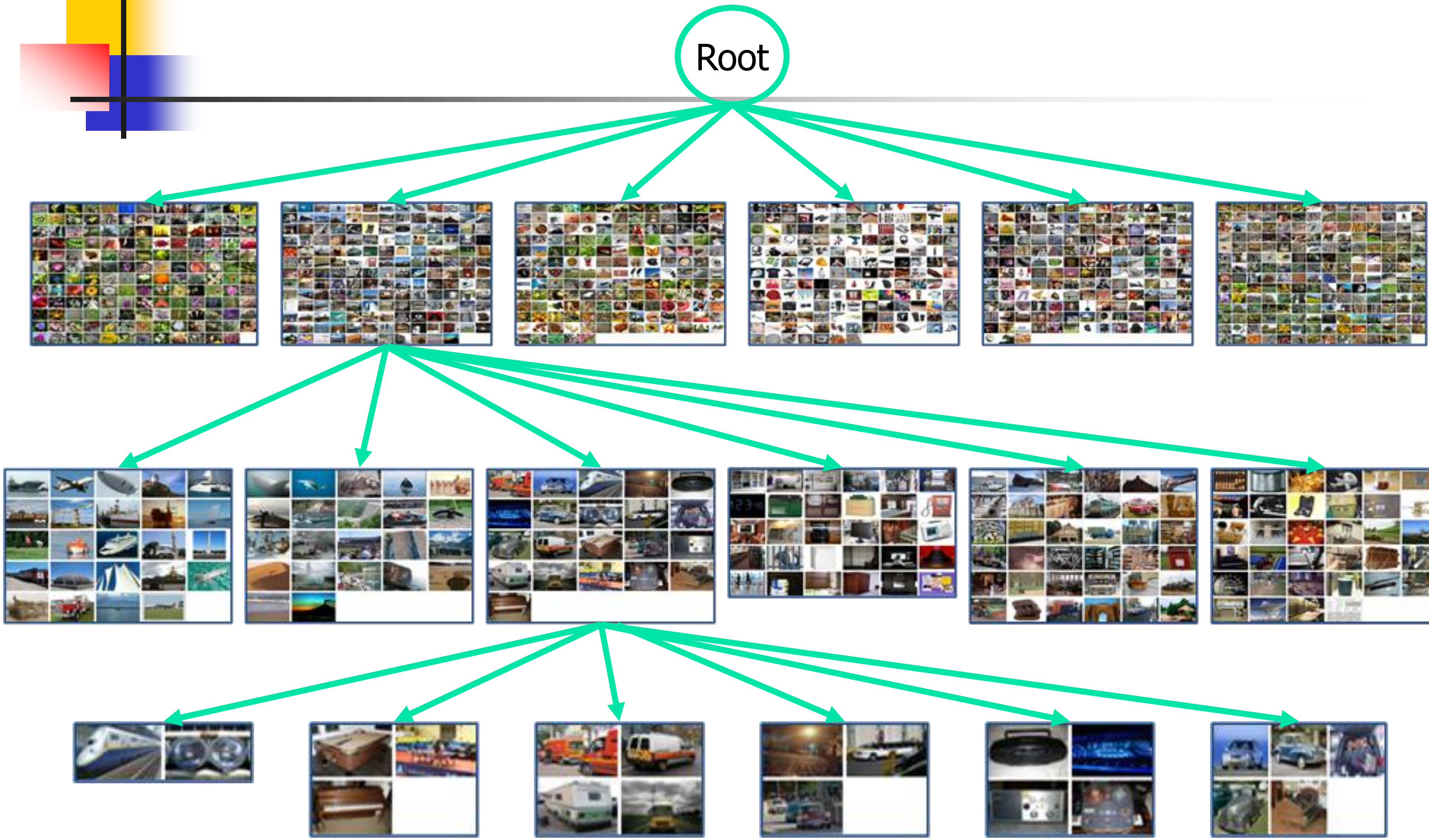
- Training N one-vs-rest SVMs is very expensive
- The SVMs could be unreliable
  - Huge sample imbalance
  - Negative samples could mislead the classifier training

Visual Similarity Matrix

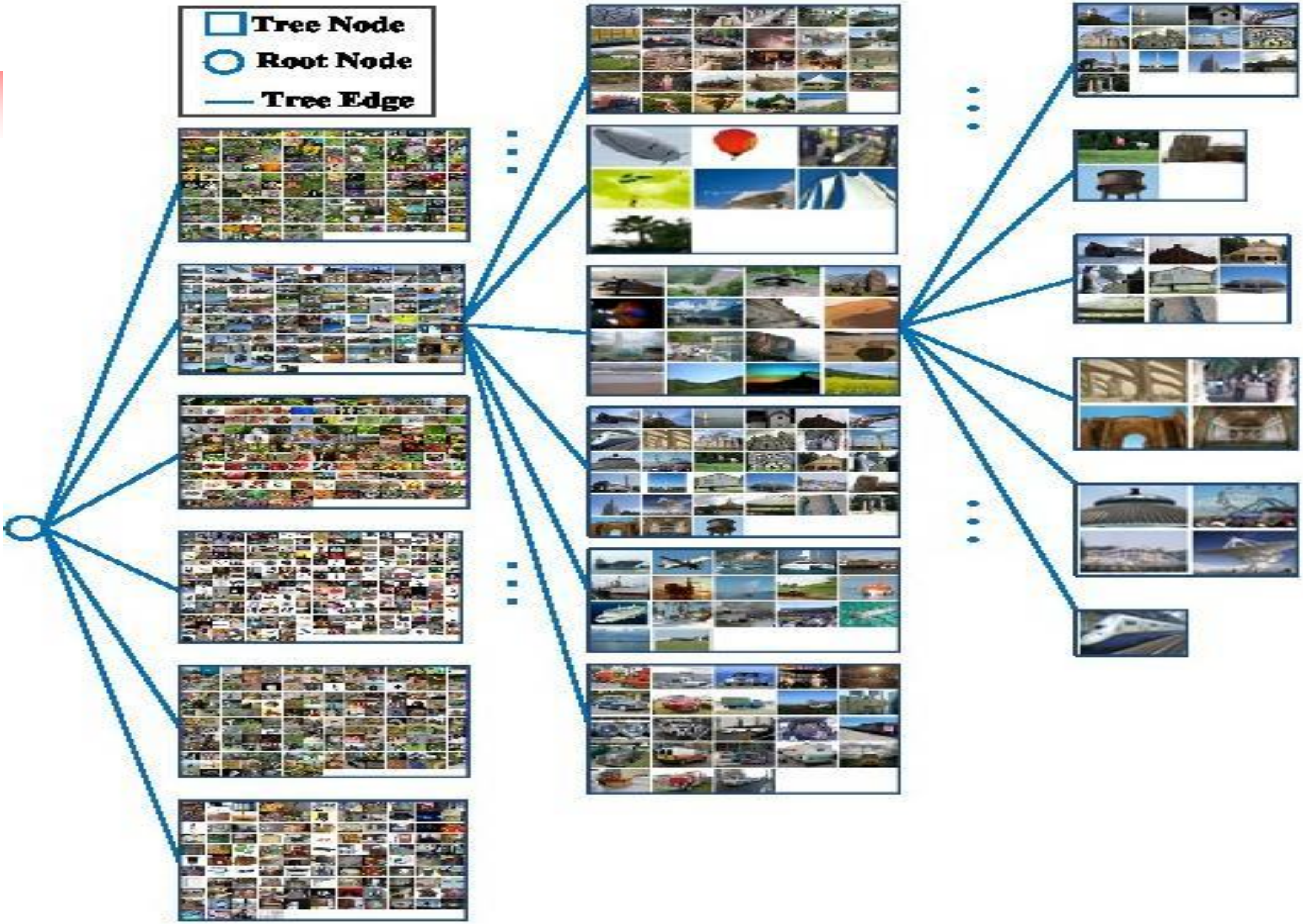


Result is based on ImageNet data set of 1000 categories

4. Visual Tree Construction: Hierarchical Clustering



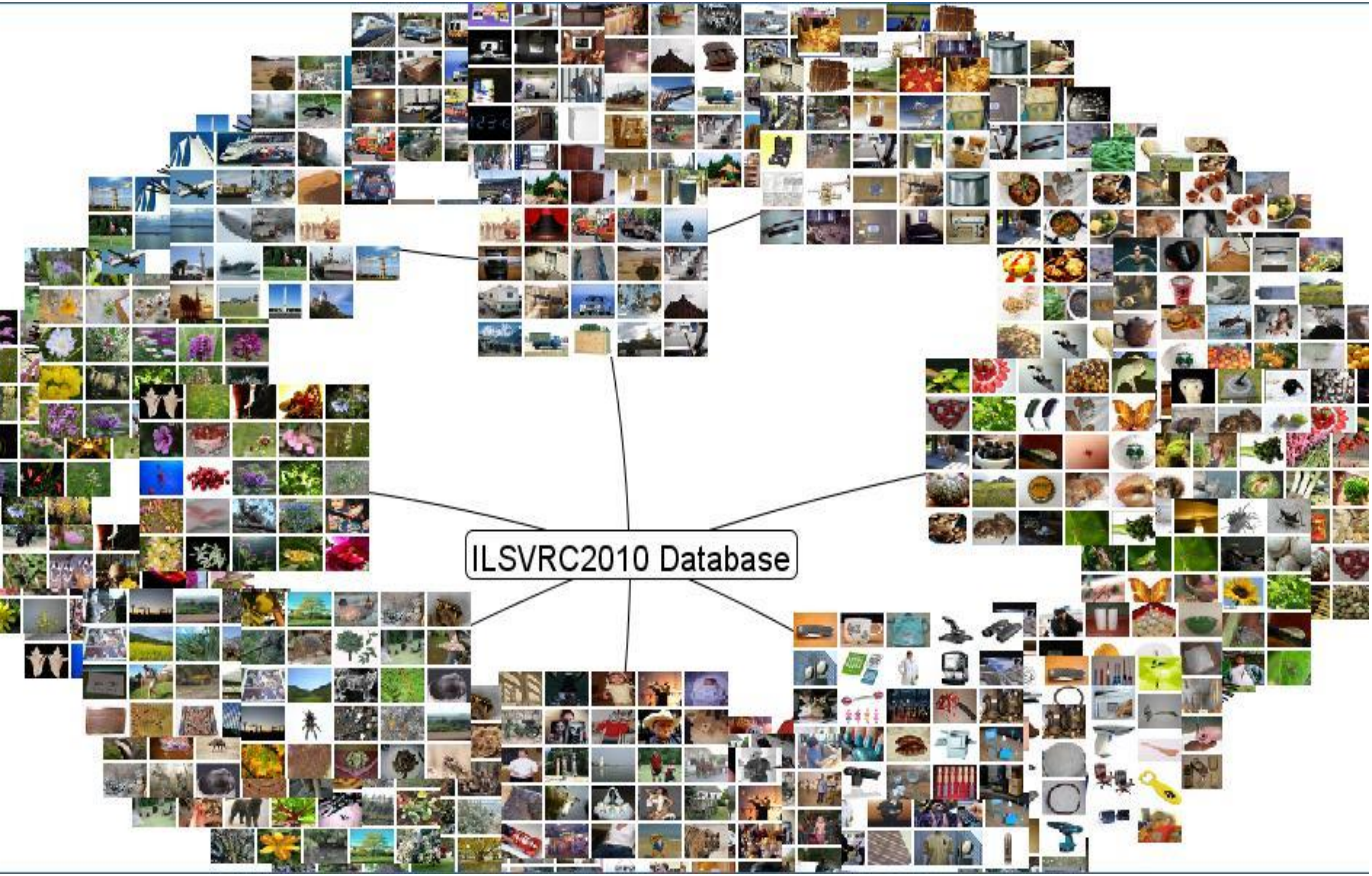
4. Visual Tree Construction: Hierarchical Clustering



4. Visual Tree Construction: Hierarchical Clustering

Visual_kmeans tree with $B=6$, $H=4$.
The leaf nodes are not shown.

4. Visual Tree Construction: Hierarchical Clustering

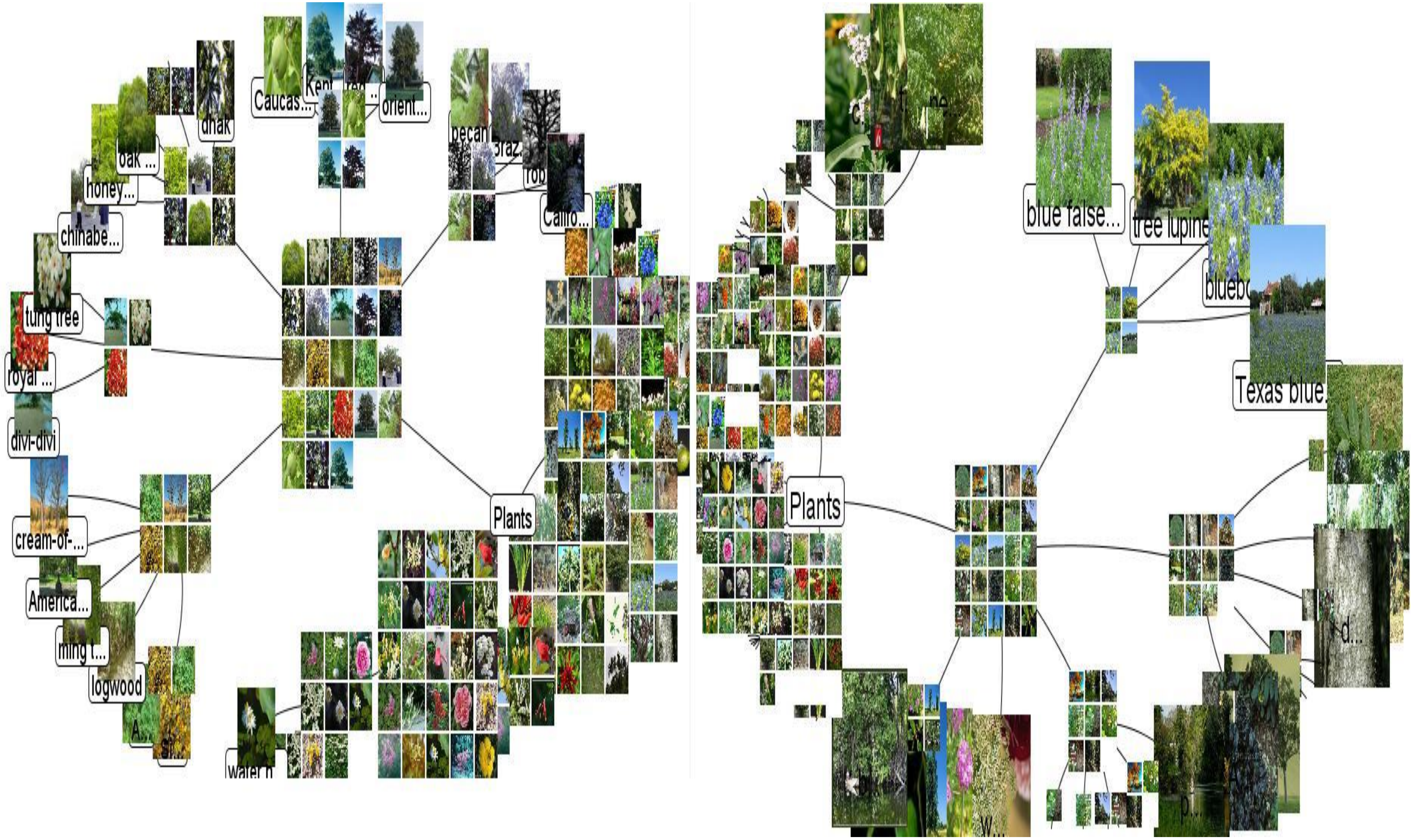


4. Visual Tree Construction: Hierarchical Clustering

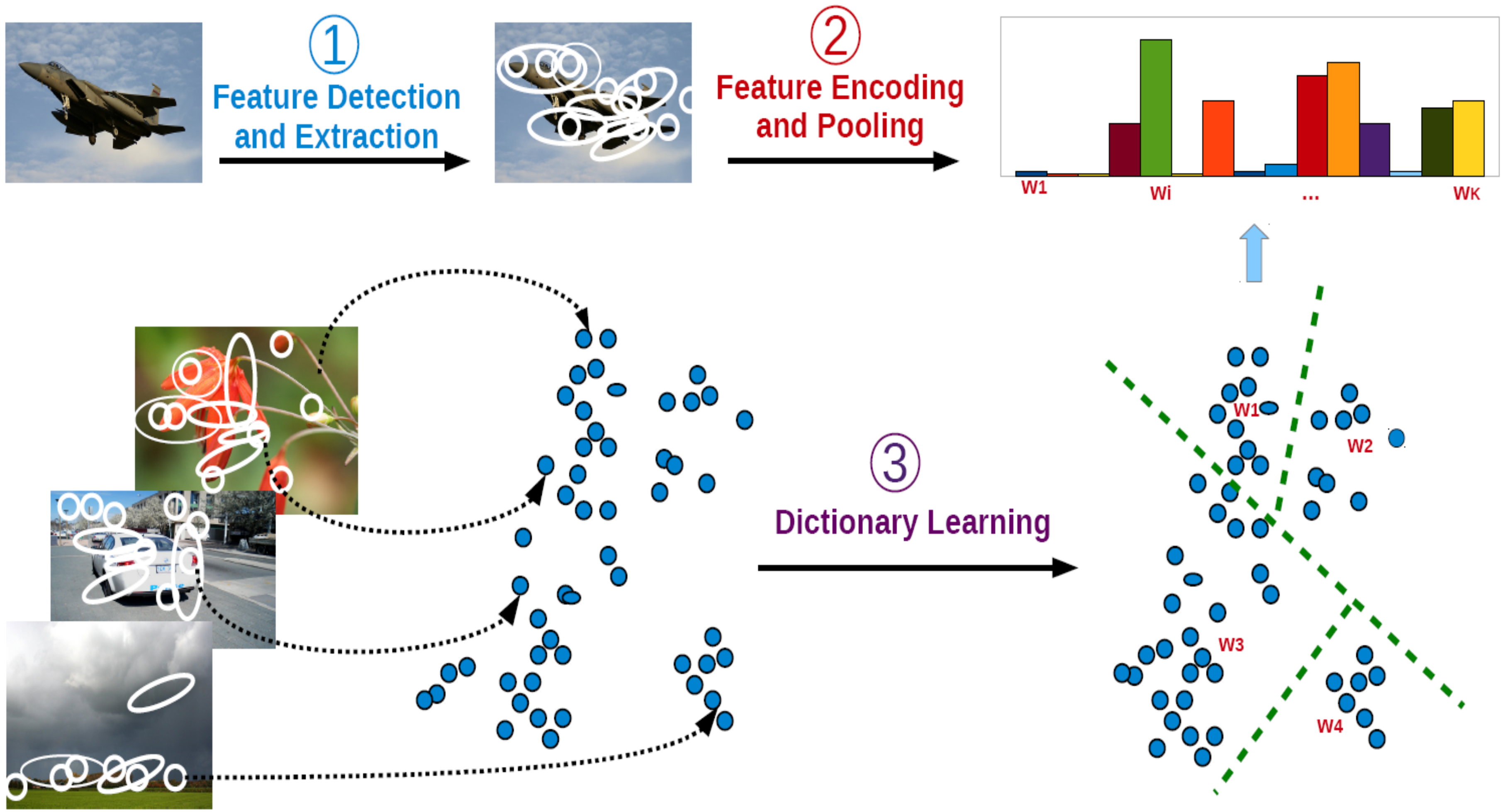
(a)

(b)

4. Visual Tree Construction: Hierarchical Clustering



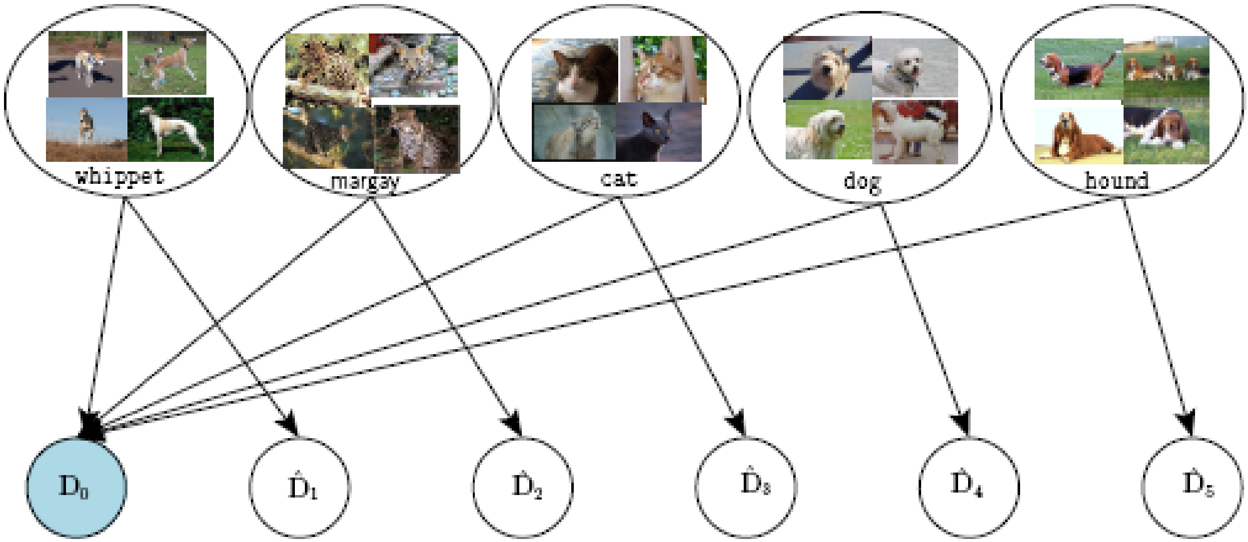
Bag-of-Words (BoW)



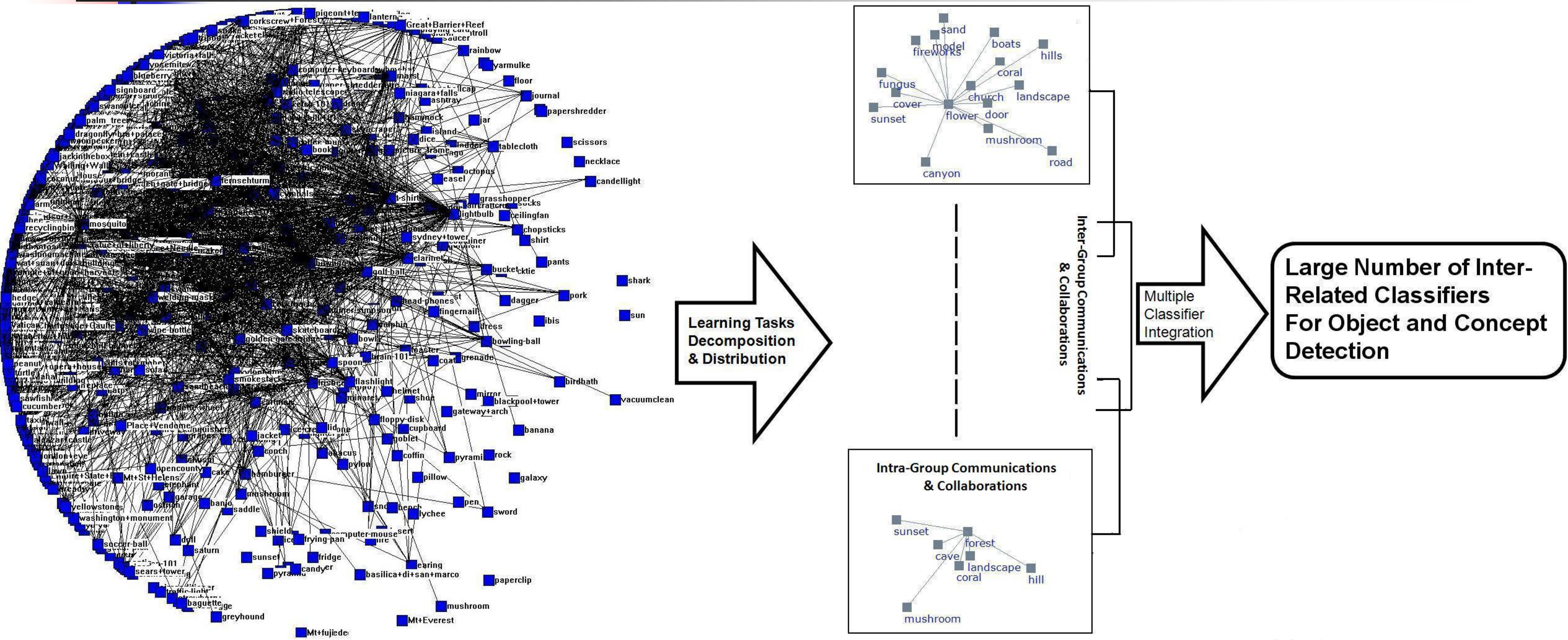
5. Joint Dictionary Learning for Discriminative Image Representation

- To distinguish visually-similar categories, dictionaries with strong discrimination is critical
- Joint dictionary learning

CVPR 2012
TPAMI2013



7. Large-Scale Classifier Training



(a) Concept Network for Task and Communication Organization (b) Training Groups and Inter-Group Communication (c) Classifier Integration

IEEE Trans. IP 2011, IEEE Trans. PAMI 2014, PR 2013

7. Large-Scale Classifier Training

- **Inference Model Selection for Classifier Training**

$$f_{C_j}(x) = W_j^{tr}\Phi_j(x) + \sum_{C_t \in \Theta_j} \gamma_t \cdot V_t^{tr}\Phi_t(x), \qquad \sum_{C_t \in \Theta_j} \gamma_t = 1$$

If the given image concept C_j is visually-related with the image concept C_t (i.e., C_j is linked with C_t on the visual concept network), $V_t \neq 0$. If the given image concept C_j is visually-irrelevant with the image concept C_t (i.e., C_j is not linked with C_t on the visual concept network), $V_t = 0$.

$$J = \frac{1}{2}(\|W_j\|^2 + \sum_{t=1}^{|\Theta_j|} \lambda_t \|V_t\|^2) + \rho_0 \sum_{t=1}^{|\Theta_j|} \sum_{i=1}^{n_j} \xi_{ti} + \sum_{t=1}^{|\Theta_j|} \rho_t \sum_{i=1}^{n_t} \eta_{ti}$$

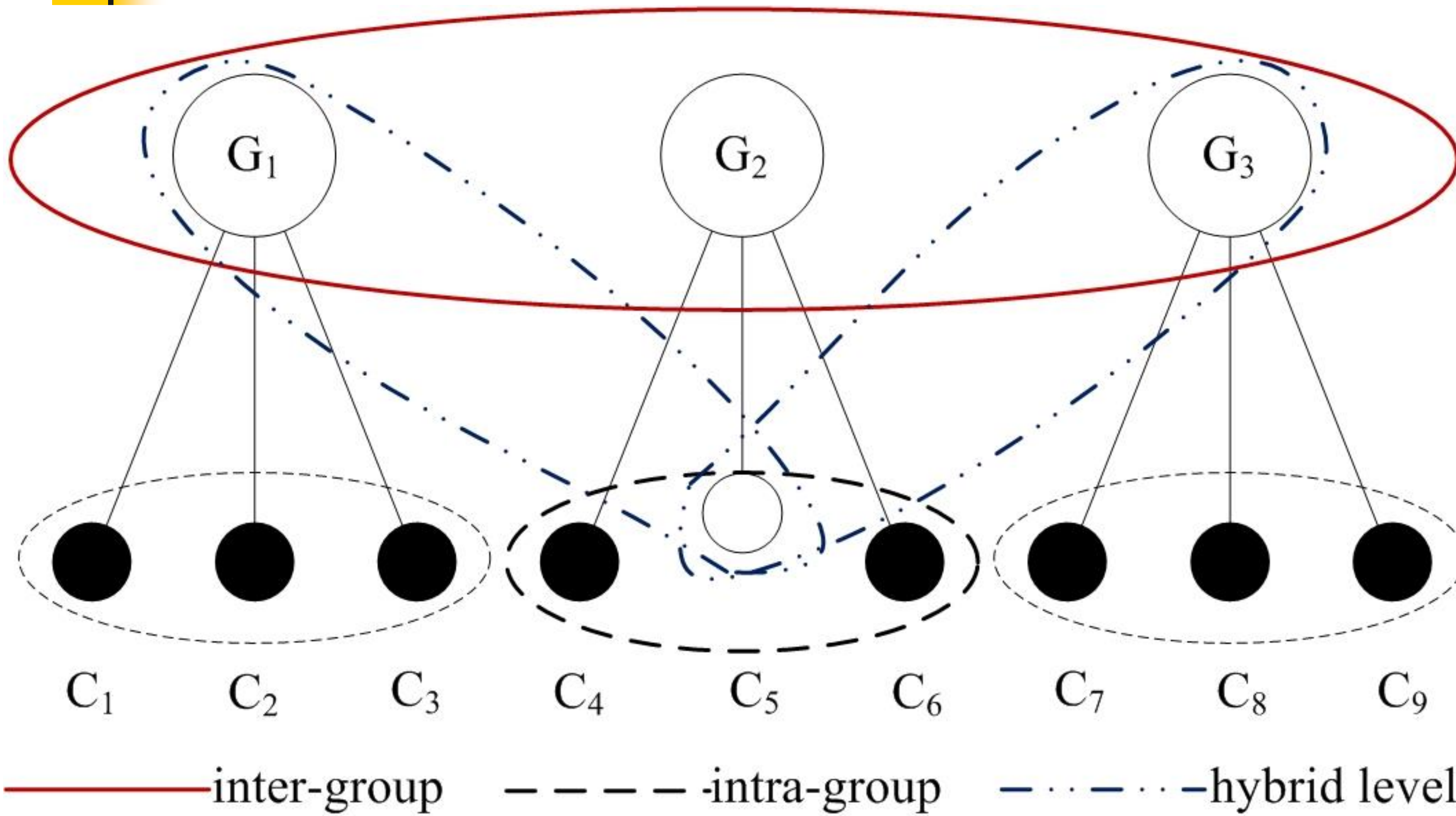
7. Large-Scale Classifier Training

- **Inference Model Selection for Classifier Training**

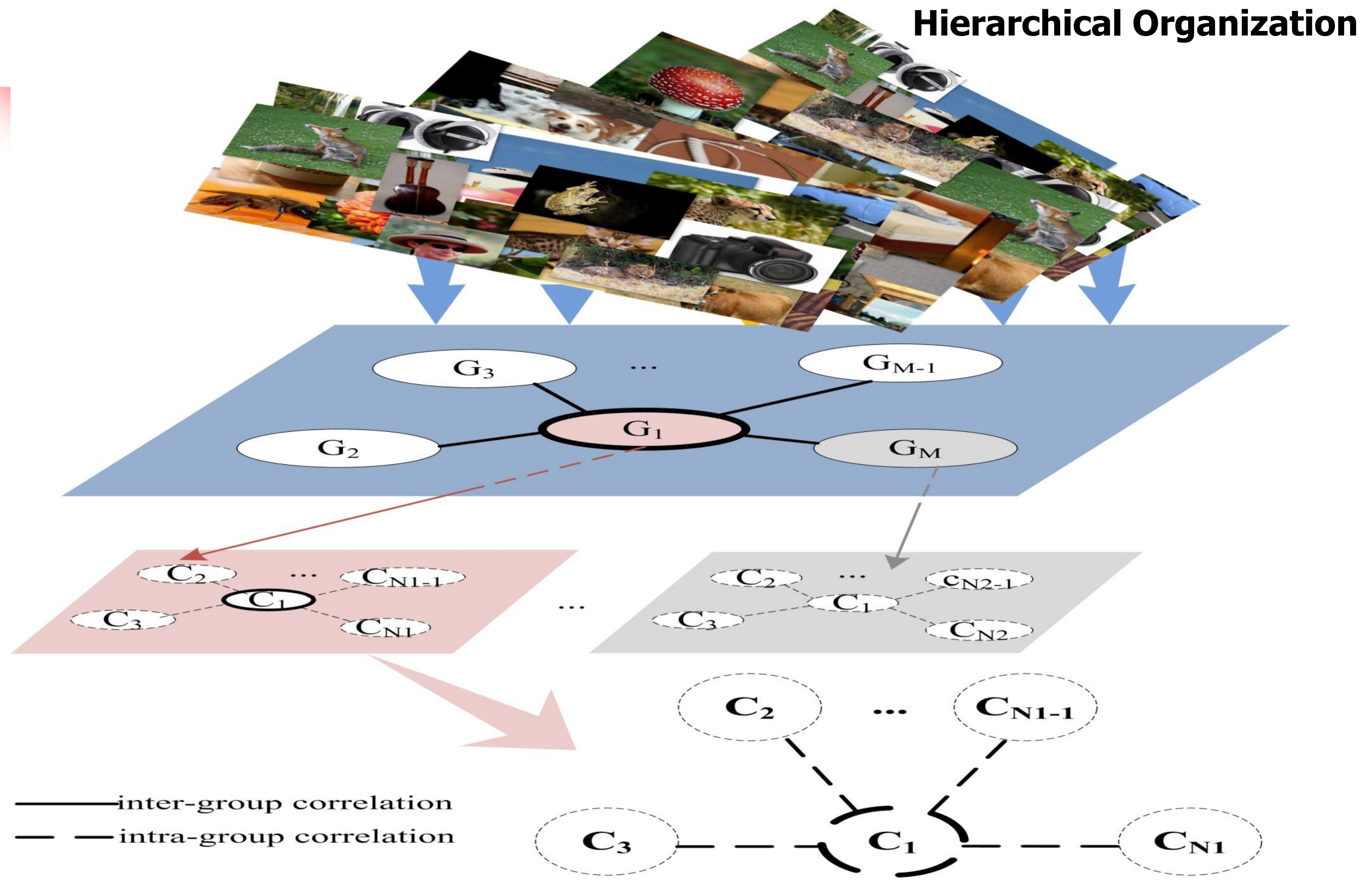
$$f_{C_j}(x) = \sum_{h,t=1}^{|\Theta_j|} \gamma_t \kappa_s(t,h) \left(\sum_{i=1}^{n_j} \beta_{hi} \kappa(x_{ji}, x) - \sum_{i=1}^{n_h} \overline{\beta_{hi}} \kappa(x_{hi}, x) \right)$$

$$+ \sum_{t=1}^{|\Theta_j|} \frac{\gamma_t}{\lambda_t} \left(\sum_{i=1}^{n_j} \beta_{ti} \kappa(x_{ji}, x) - \sum_{i=1}^{n_t} \overline{\beta_{ti}} \kappa(x_{ti}, x) \right)$$

7. Large-Scale Classifier Training

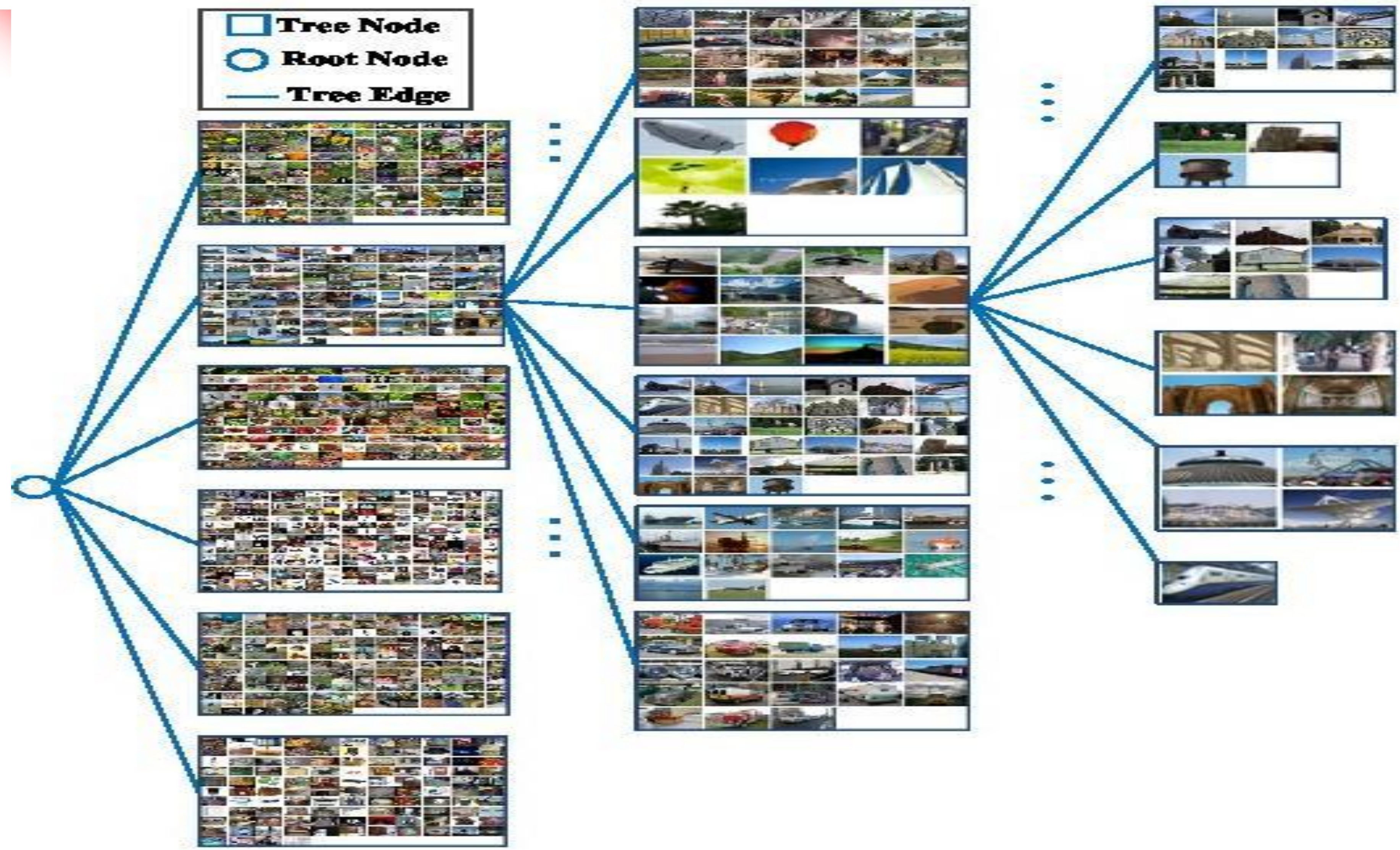


7. Large-Scale Classifier Training



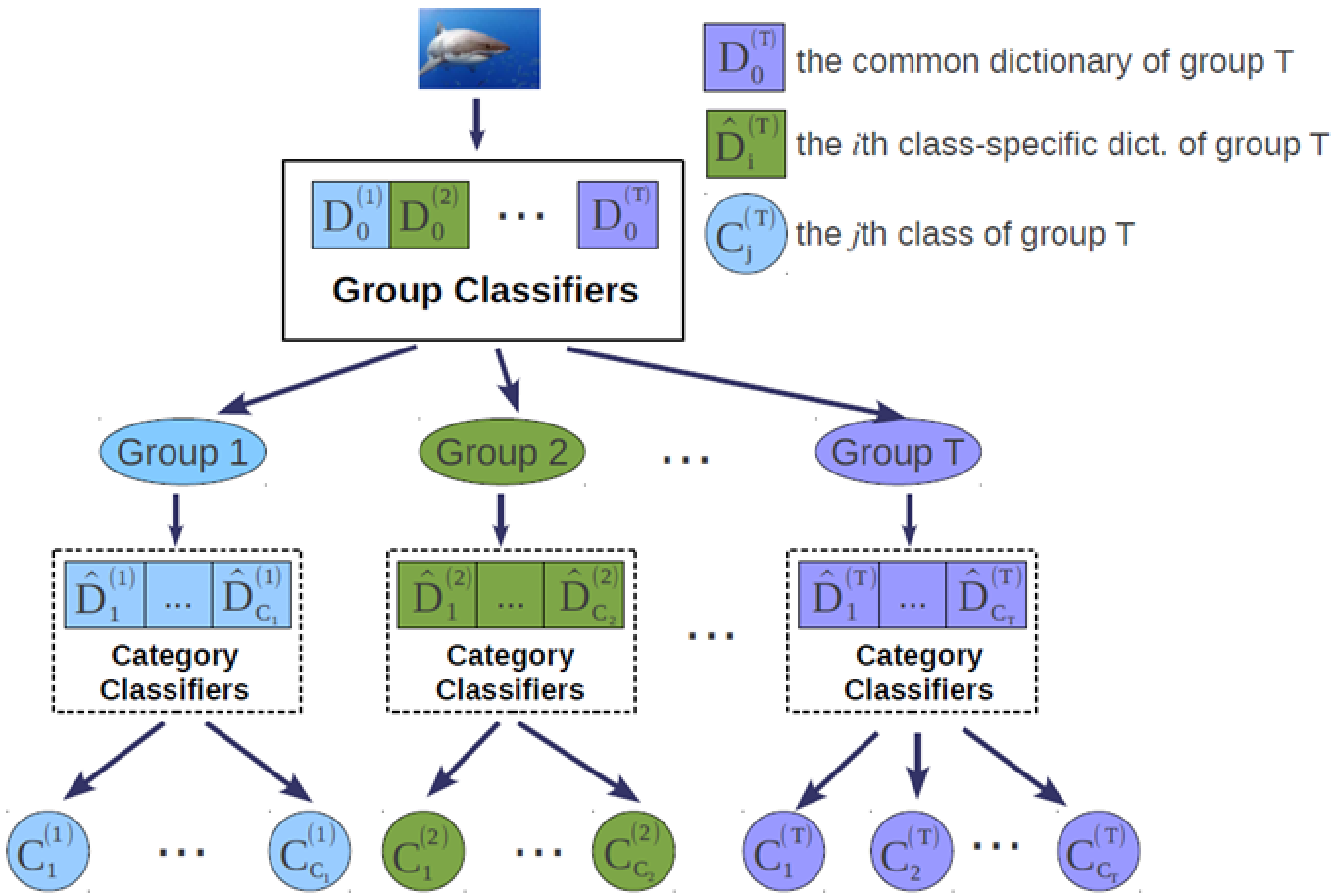
7. Large-Scale Classifier Training

Hierarchical Organization



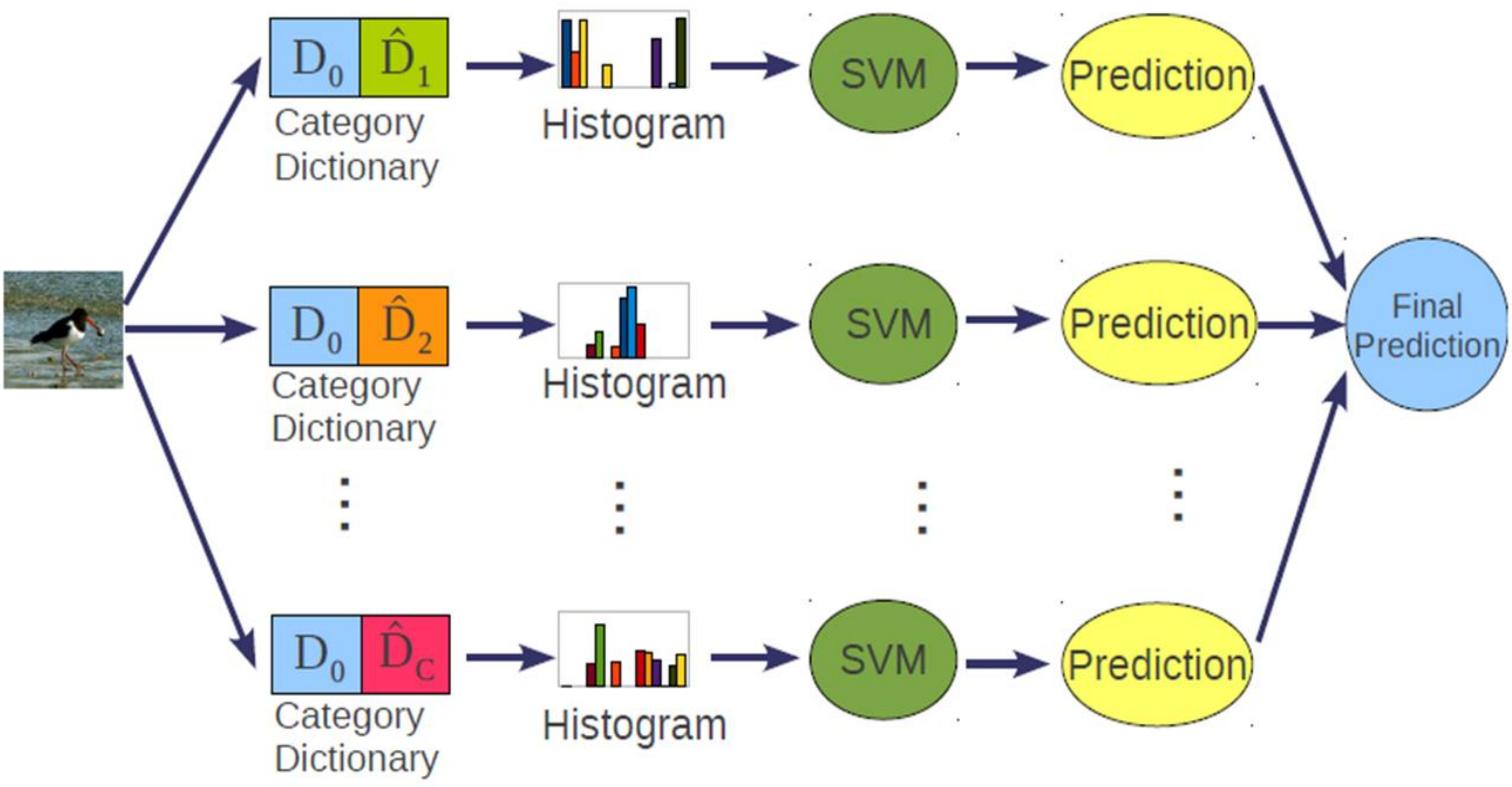
7. Large-Scale Classifier Training

Hierarchical Classification Scheme

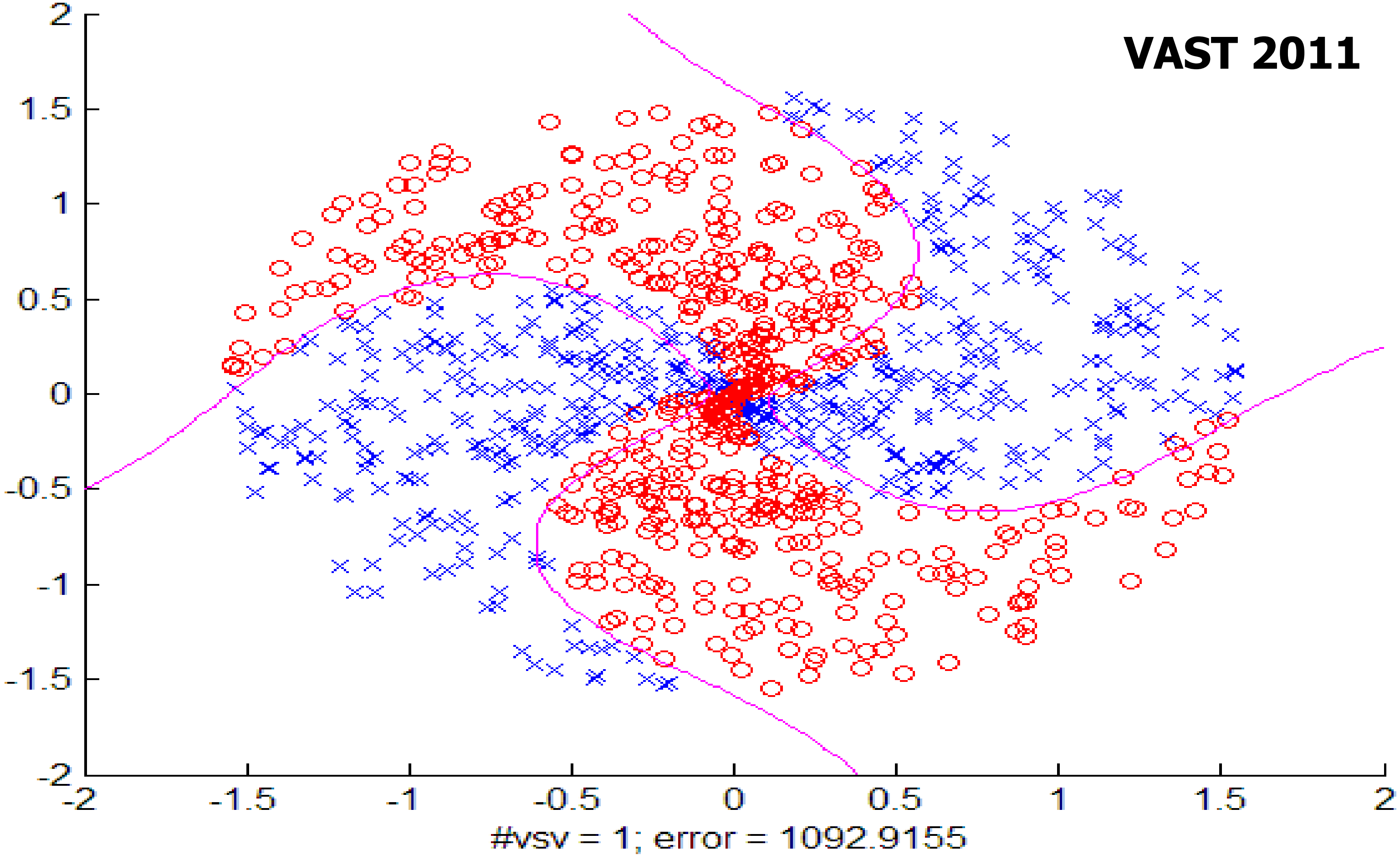


7. Large-Scale Classifier Training

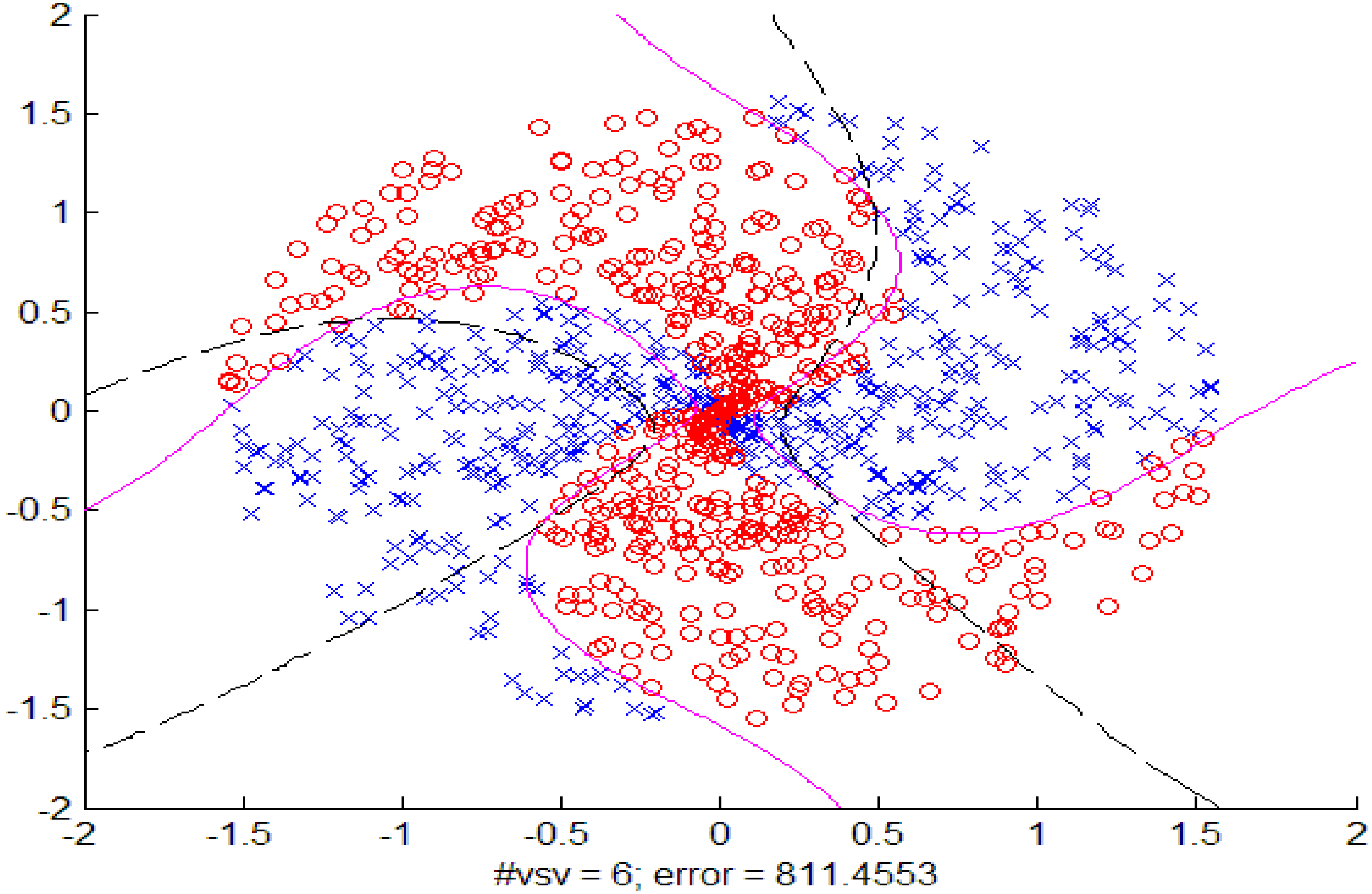
Flat Organization



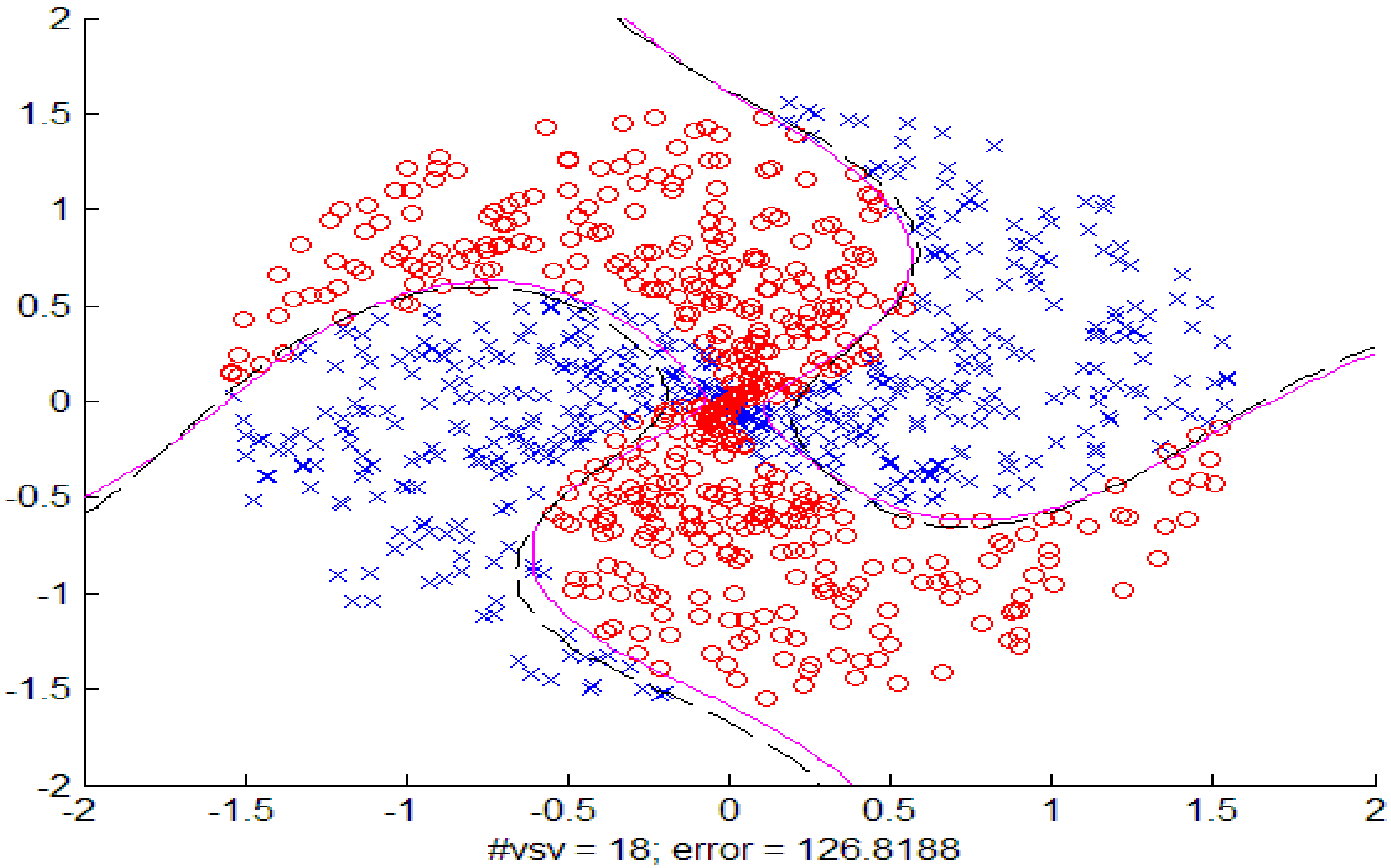
8. Interactive Classifier Assessment



8. Interactive Classifier Assessment

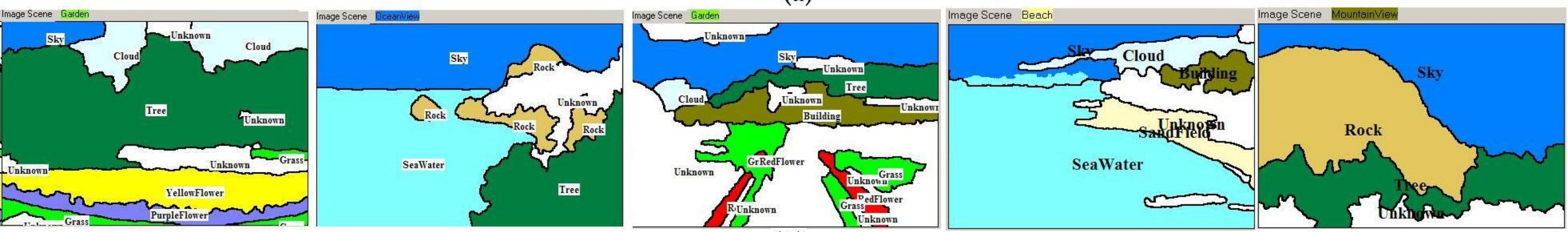


8. Interactive Classifier Assessment

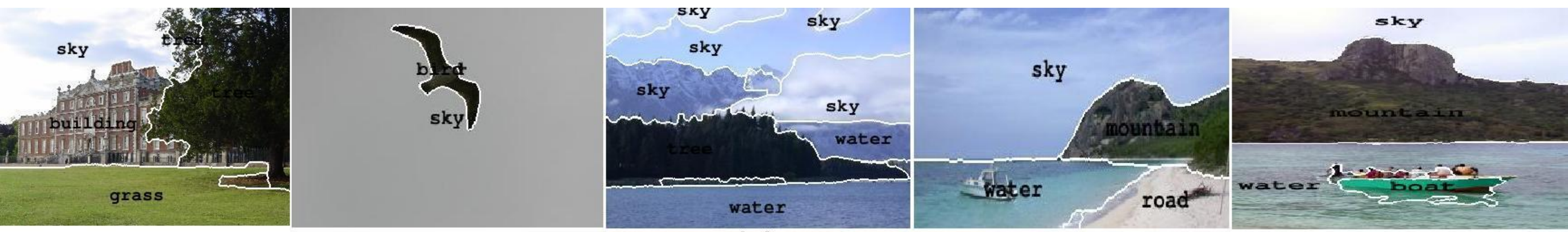


9. Some Experimental Results

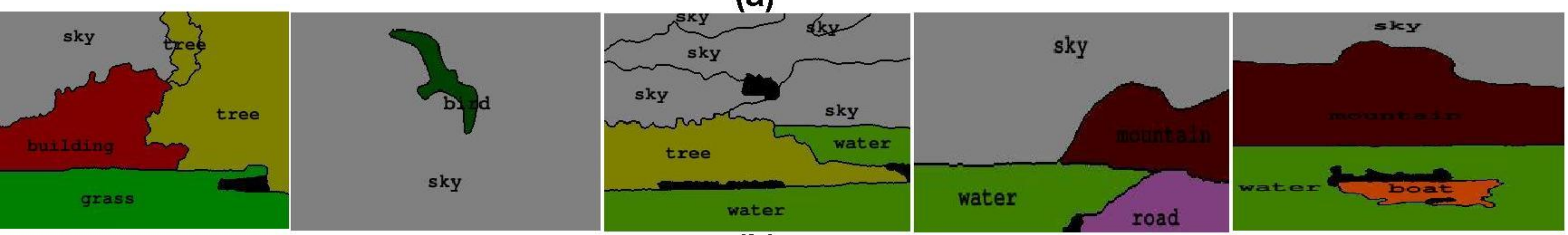
(a)



(b)

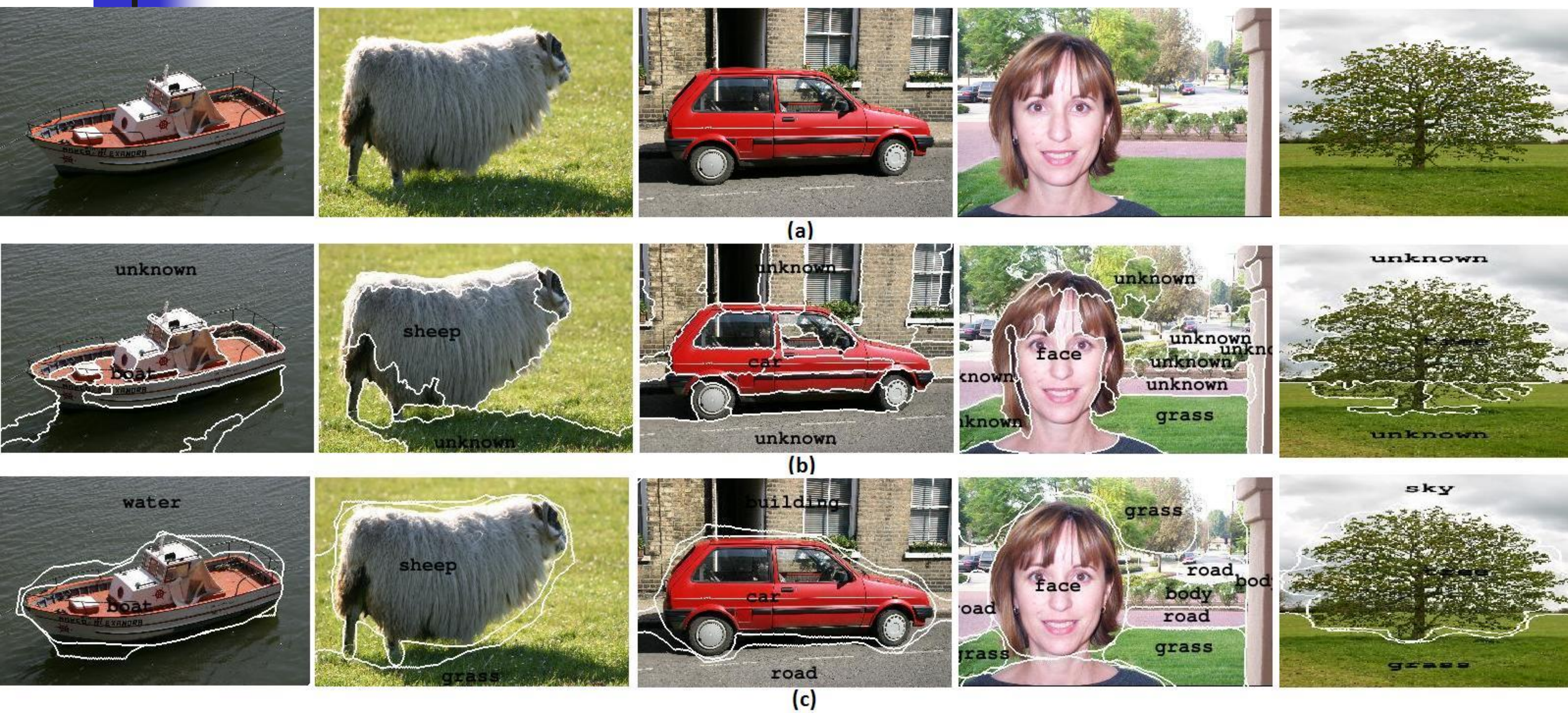


(a)

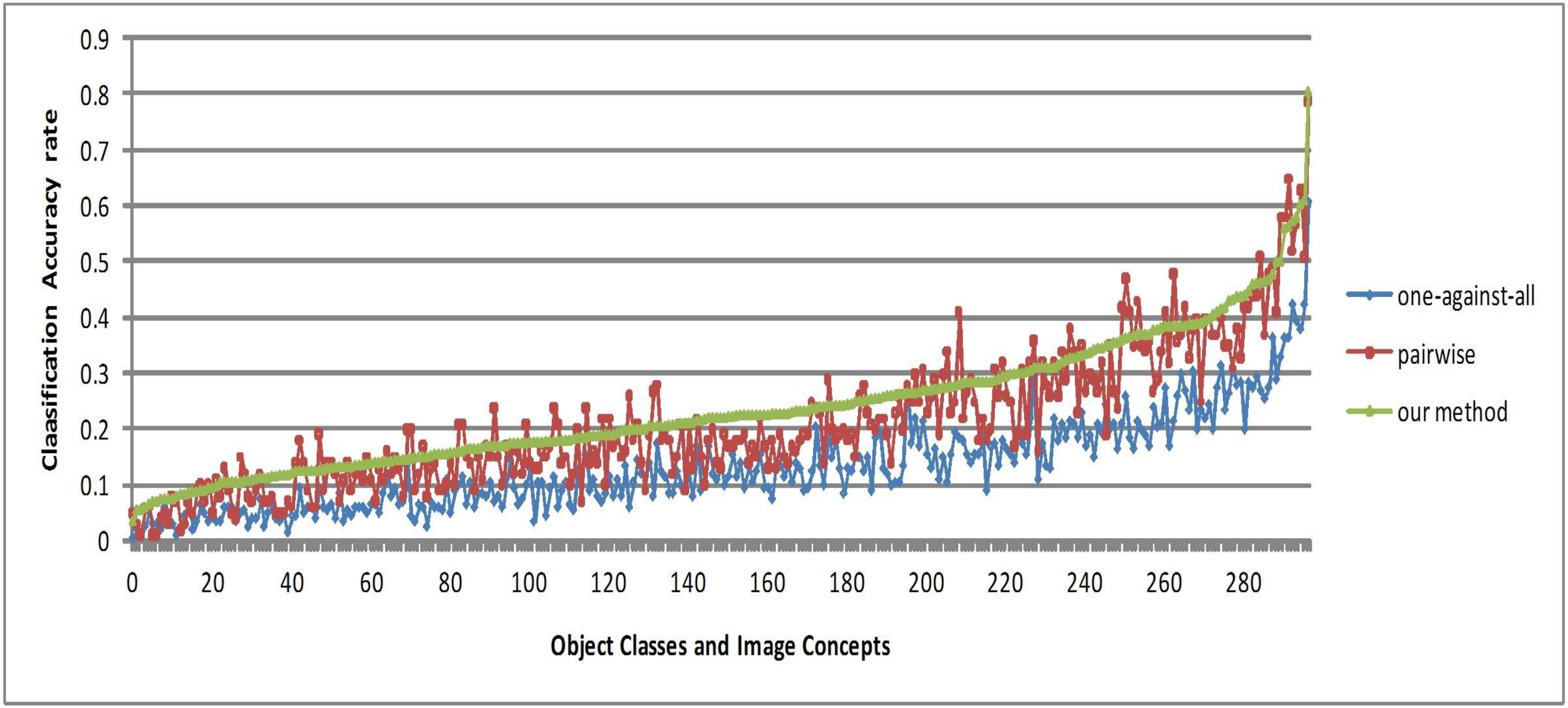


(b)

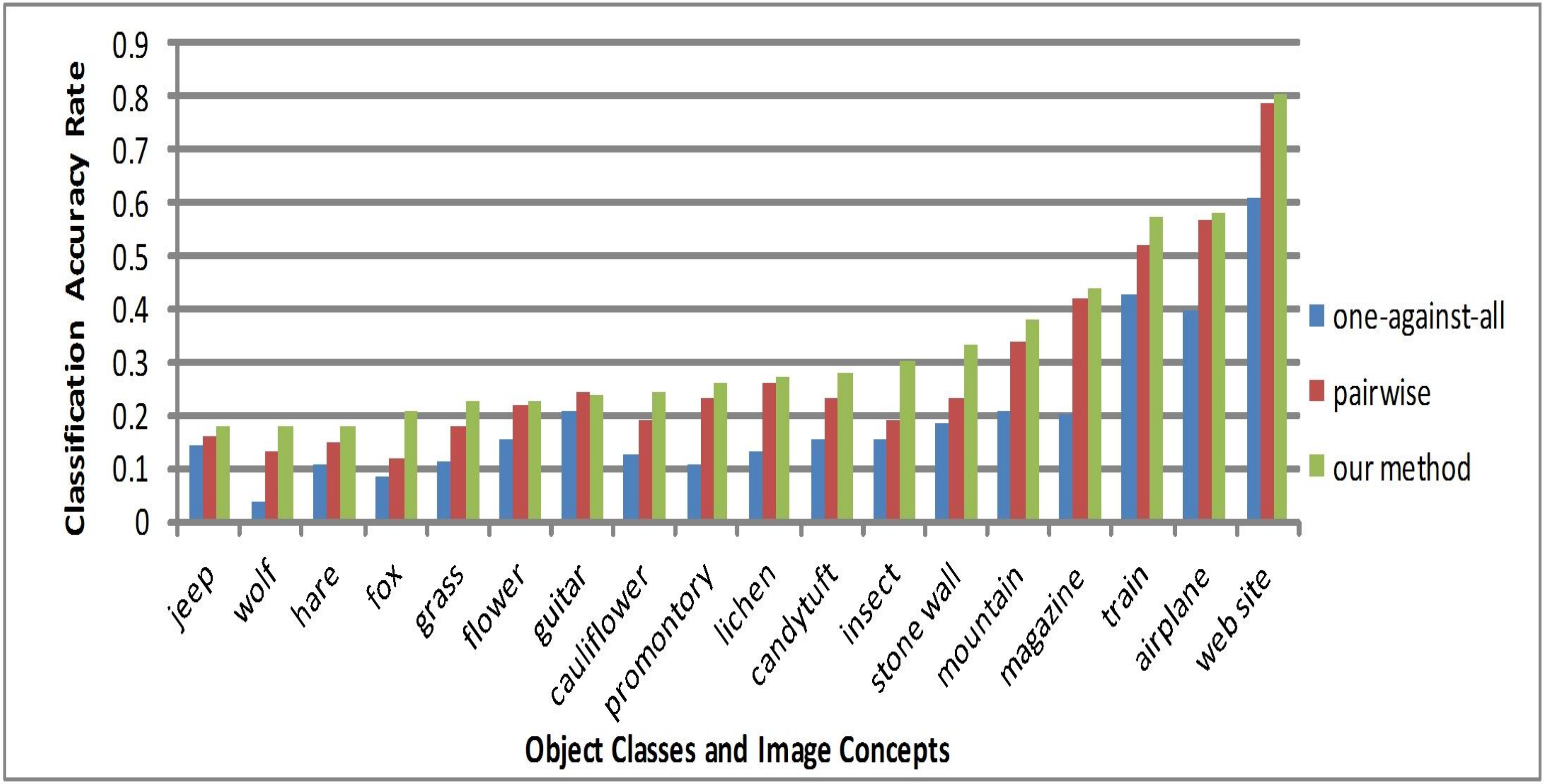
9. Some Experimental Results



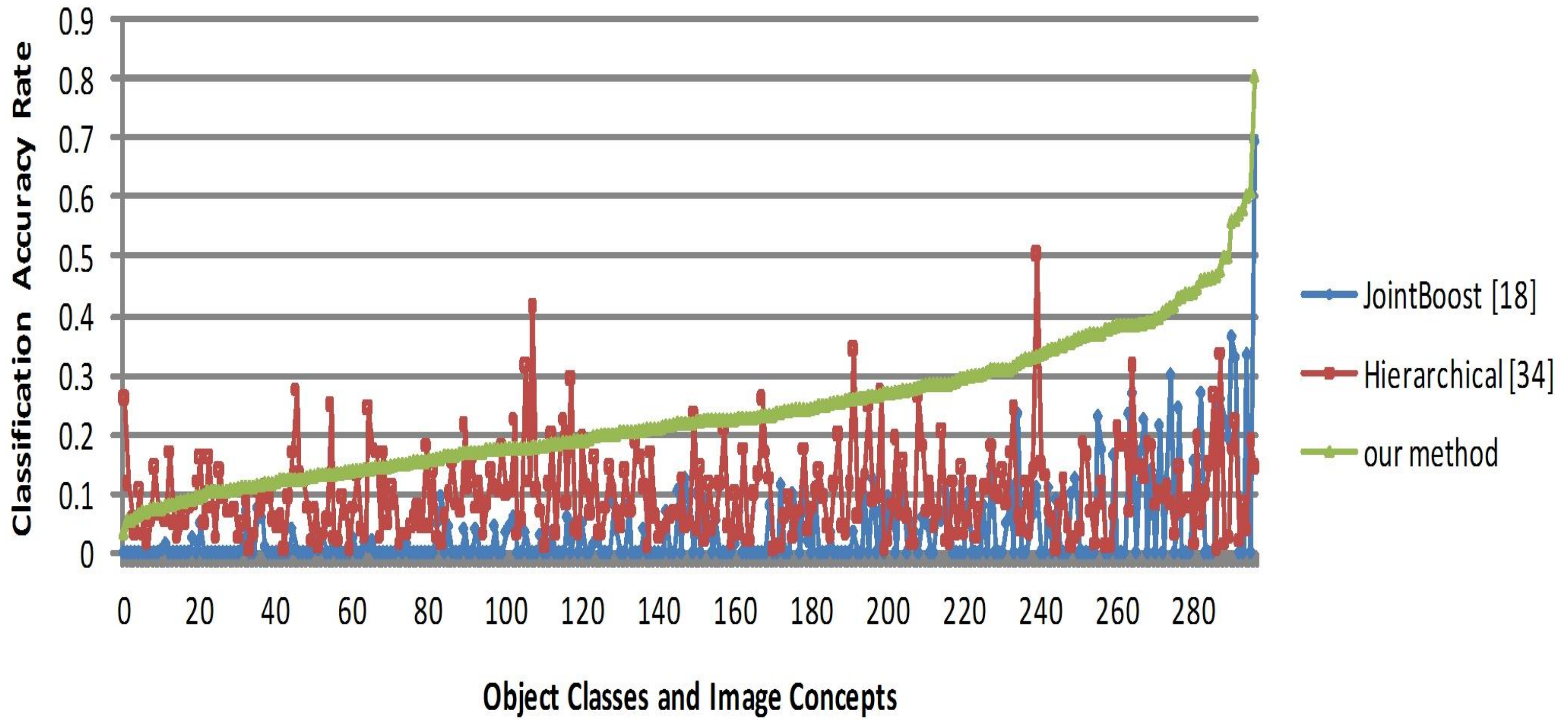
9. Some Experimental Results



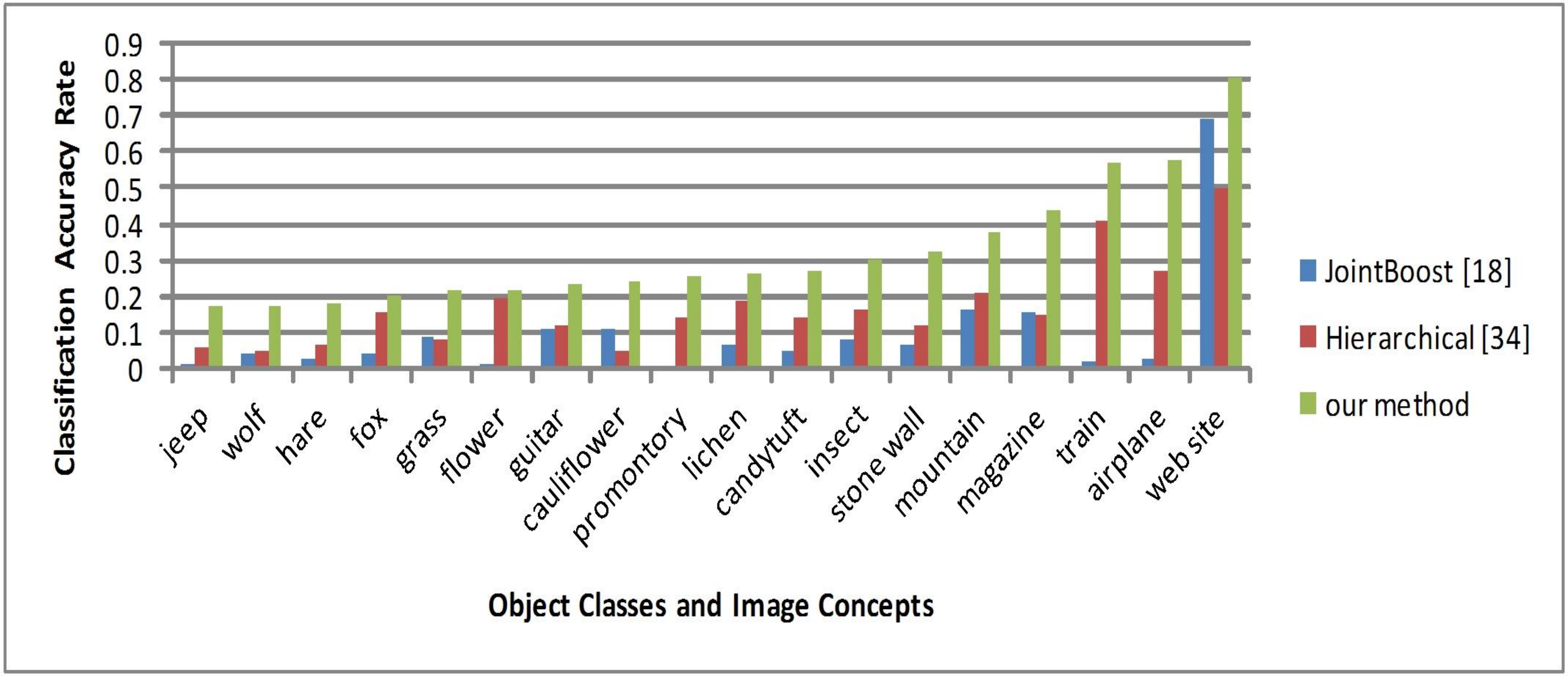
9. Some Experimental Results



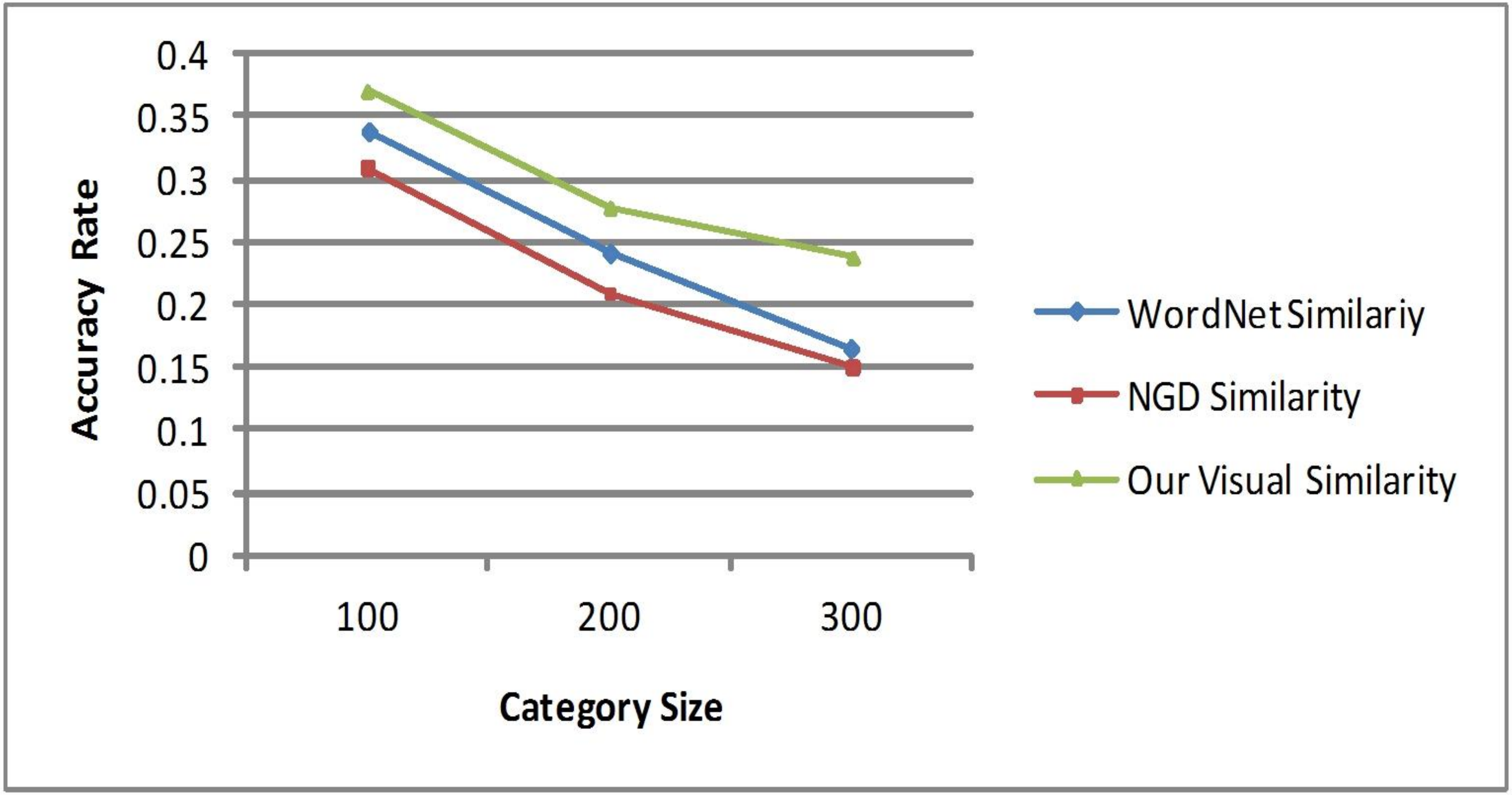
9. Some Experimental Results



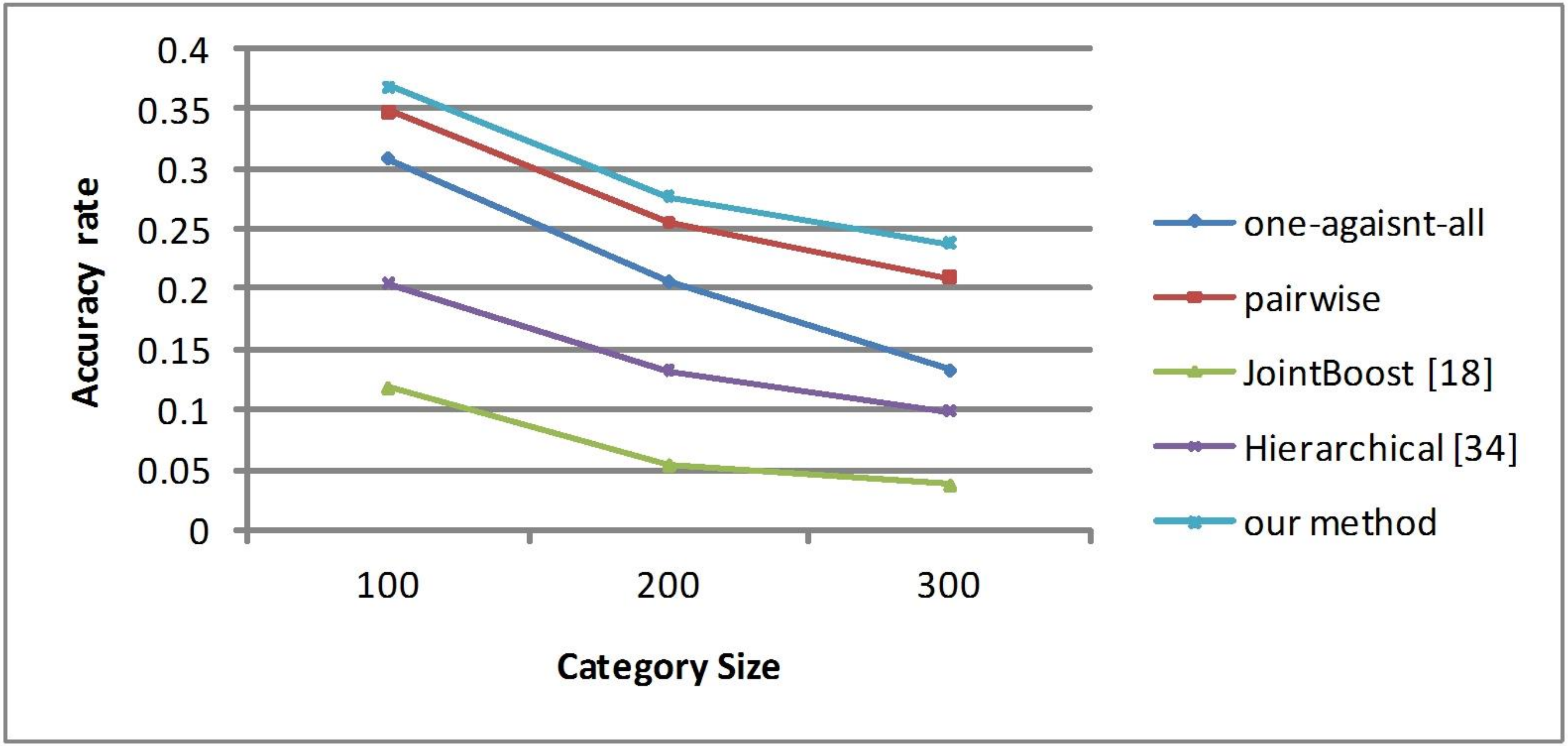
9. Some Experimental Results



9. Some Experimental Results

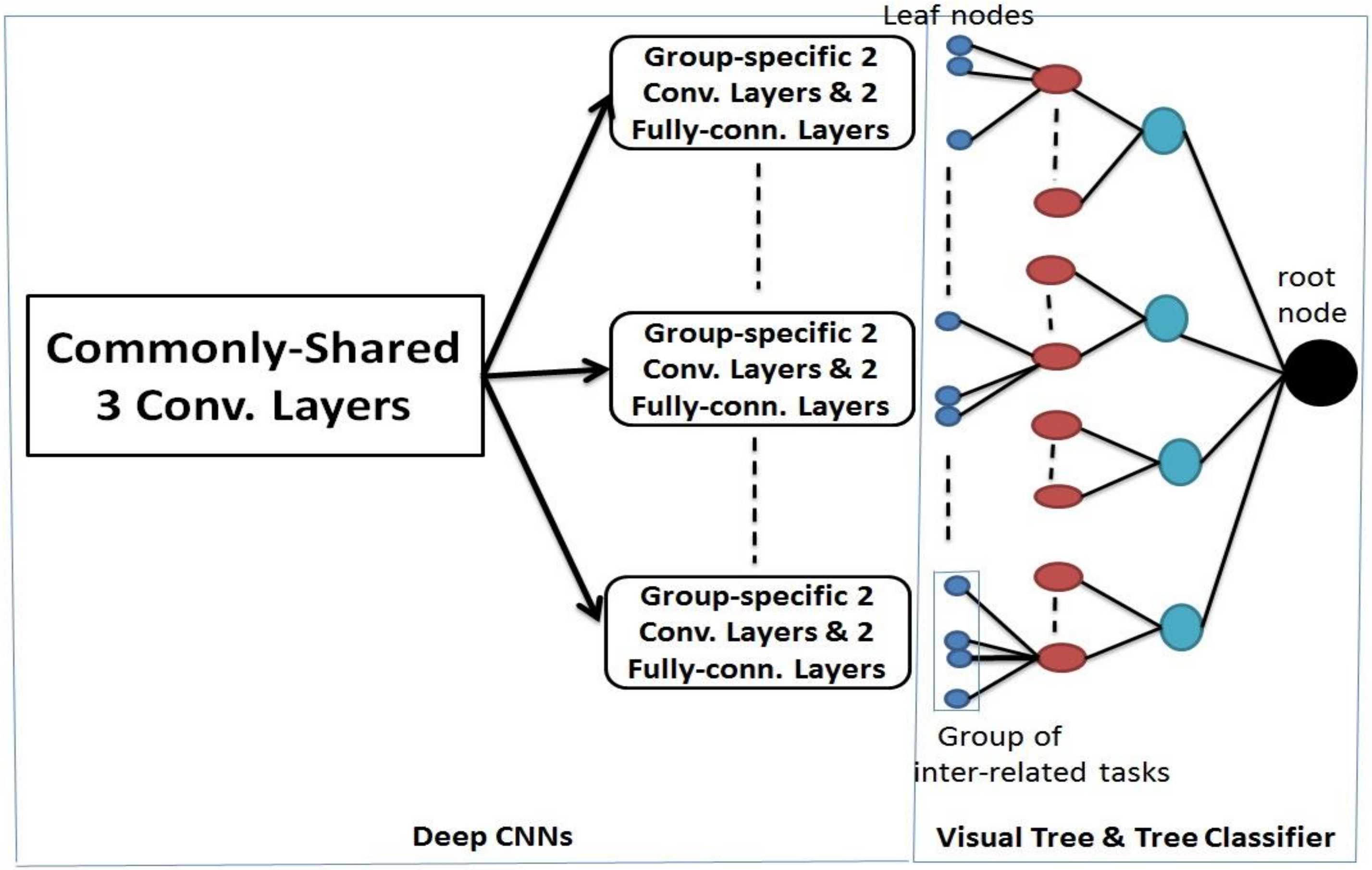


9. Some Experimental Results



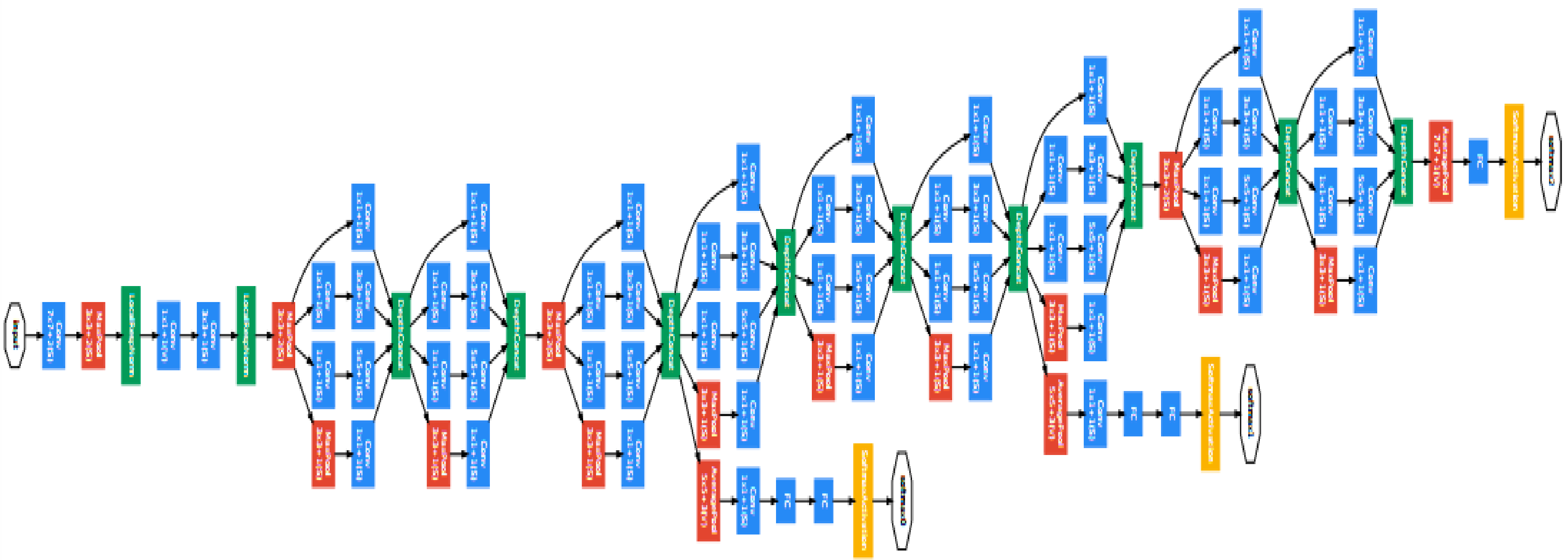
Hierarchical Deep Multi-Task Learning (HD-MTL) over Visual Tree

- **HD-MTL**



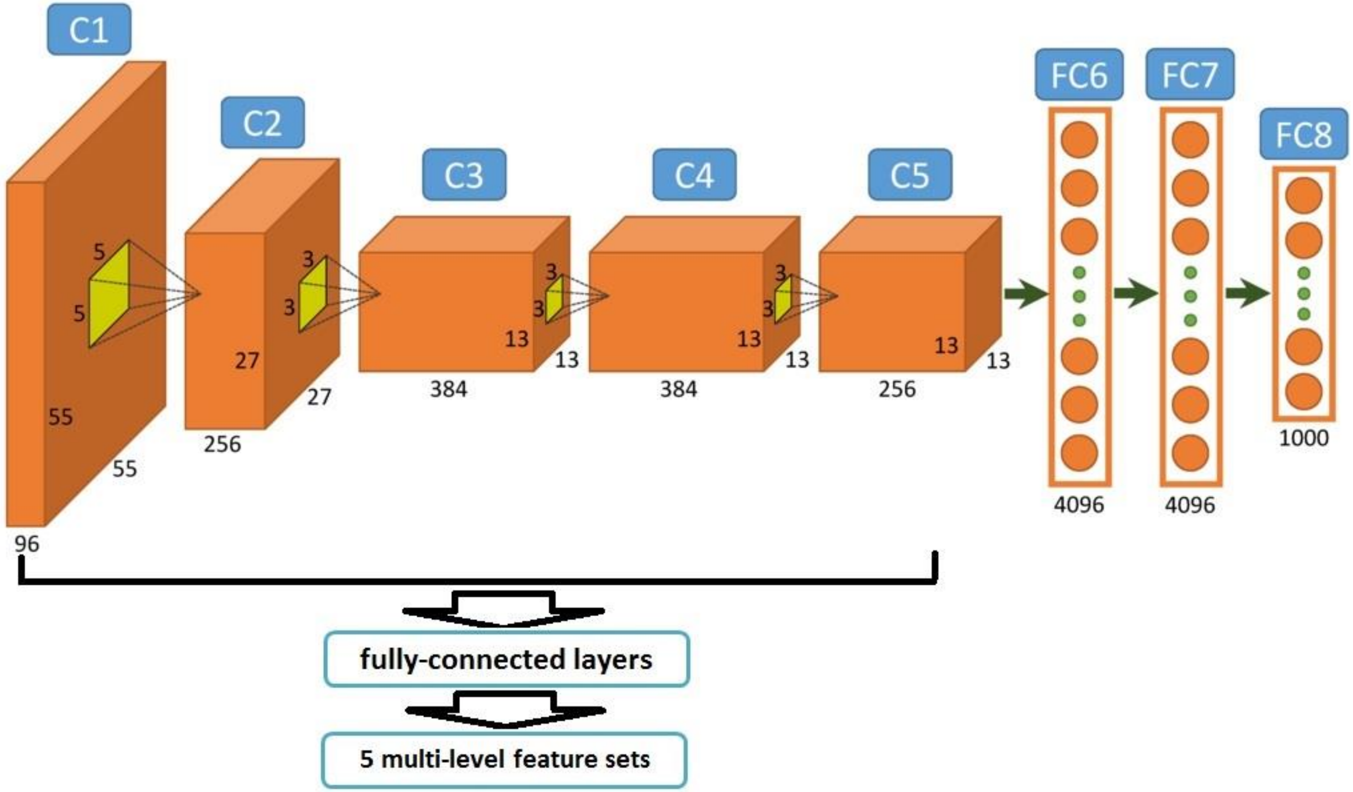
2. Multi-Level Deep Feature Extraction

- **GoogleNet**



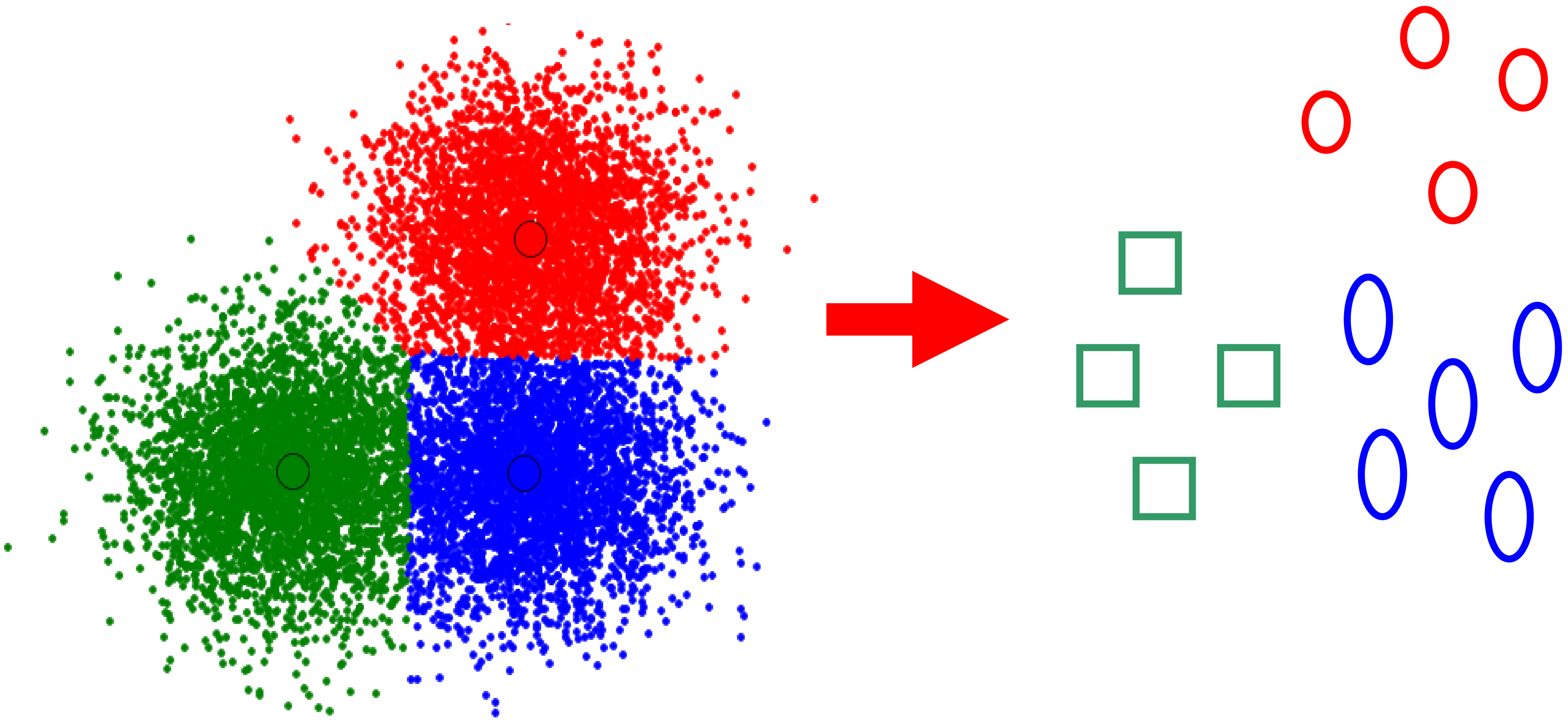
2. Multi-Level Deep Feature Extraction

- **Deep CNNs for Feature Extraction**



3. Visual Tree Construction

- **Object Class Representation**



4. Visual Tree Construction: Hierarchical Clustering

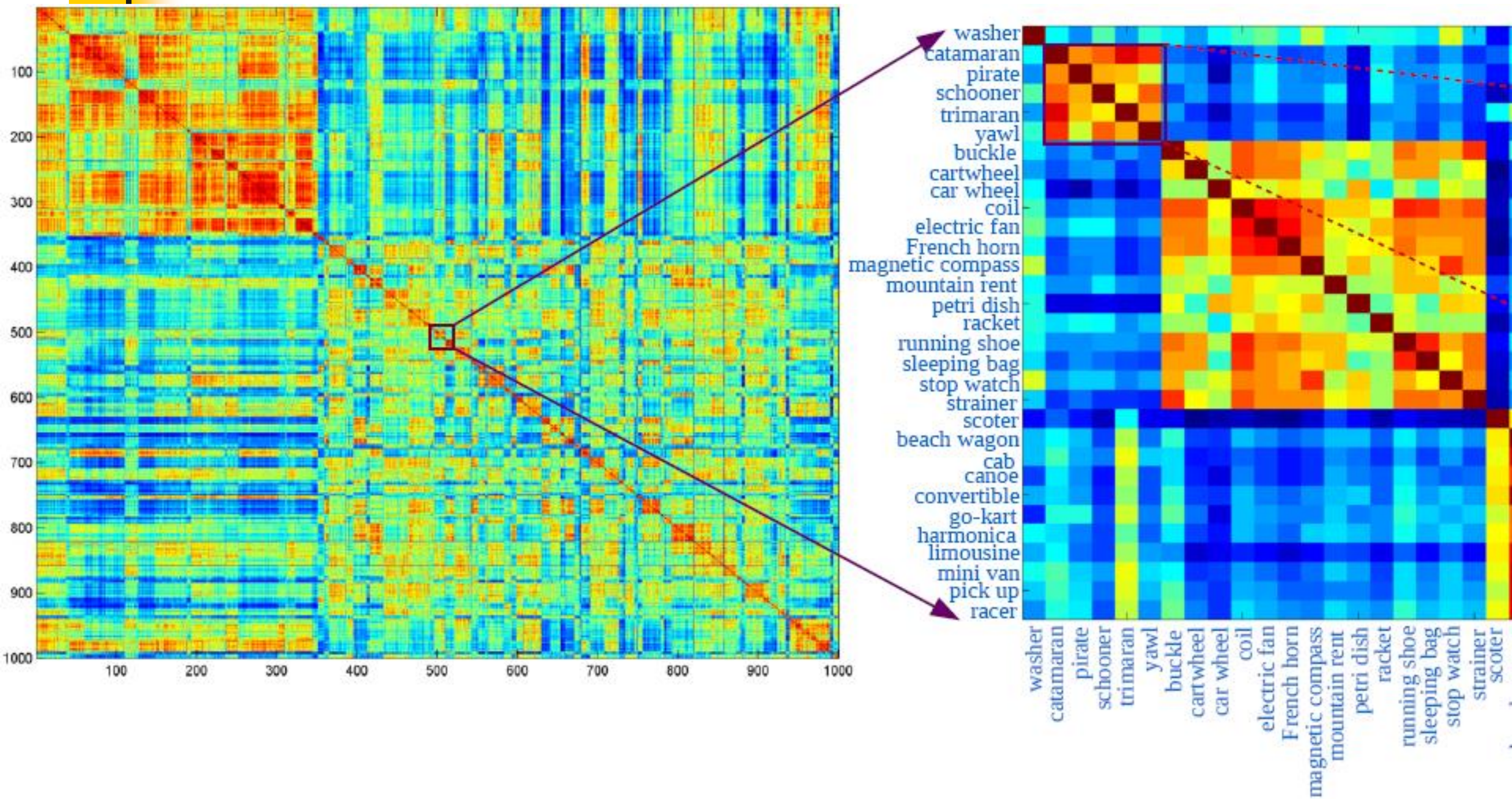
- **Feature Subset Selection**

$$F^{c}_{best} = max\left\{\Phi^{c}_{t} = \frac{1}{\sum_{i=1}^{M}\sum_{j=1}^{M}\kappa^{t}(i,j)},\quad F_t \in \mathbb{F}\right\}$$

- **Node Partitioning**

$$min\left\{\psi(c,B) = \sum_{l=1}^{B}\frac{\sum_{i\in G_l}\sum_{j\in G^c/G_l}\kappa_t(i,j)}{\sum_{i\in G_l}\sum_{j\in G_l}\kappa_t(i,j)}\right\}$$

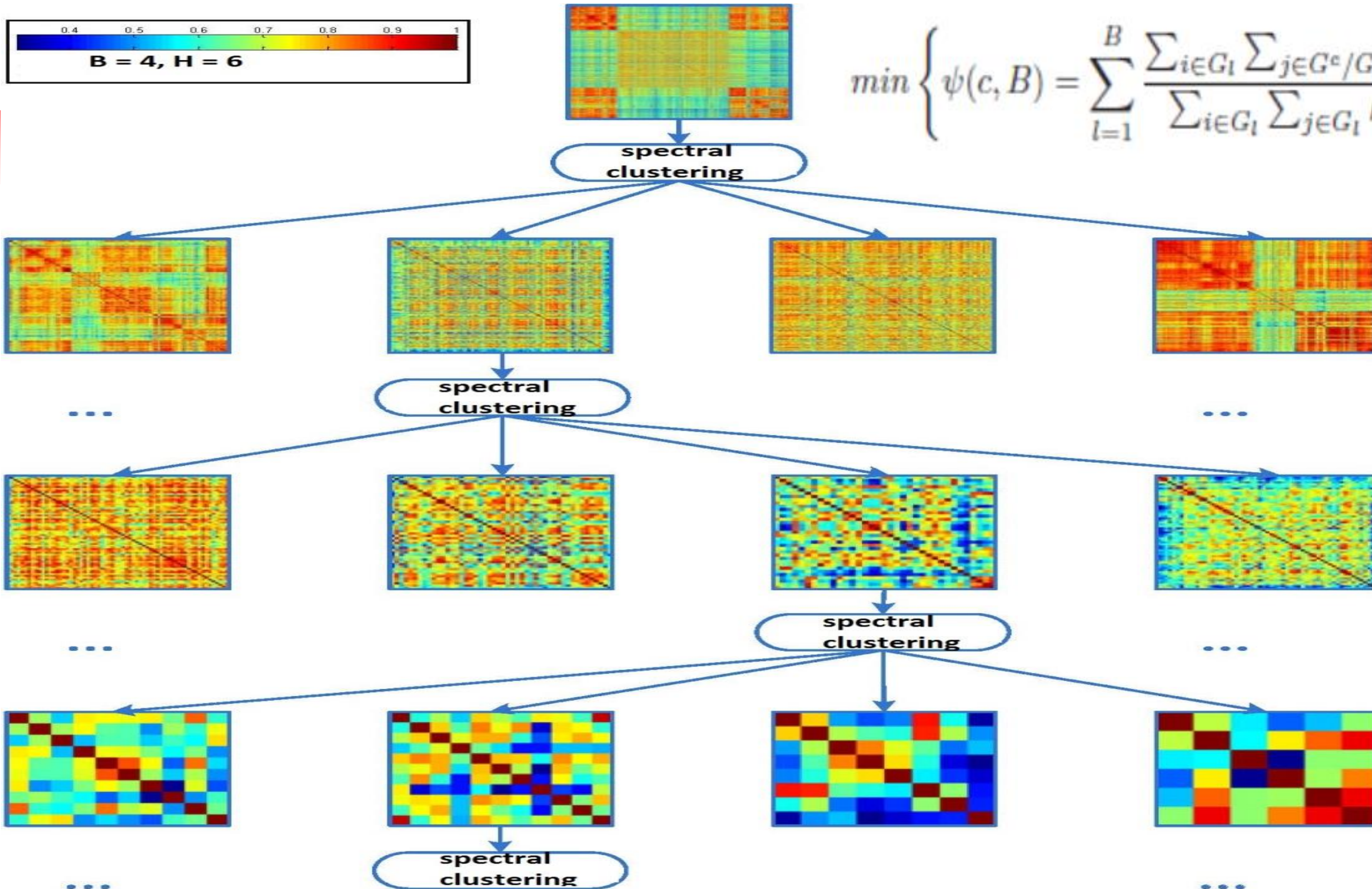
4. Visual Tree Construction: Hierarchical Clustering



Result is based on ImageNet data set of 1000 categories

4. Visual Tree Construction: Hierarchical Clustering

$$min\left\{\psi(c,B)=\sum_{l=1}^{B}\frac{\sum_{i\in G_l}\sum_{j\in G^c/G_l}\kappa_t(i,j)}{\sum_{i\in G_l}\sum_{j\in G_l}\kappa_t(i,j)}\right\}$$

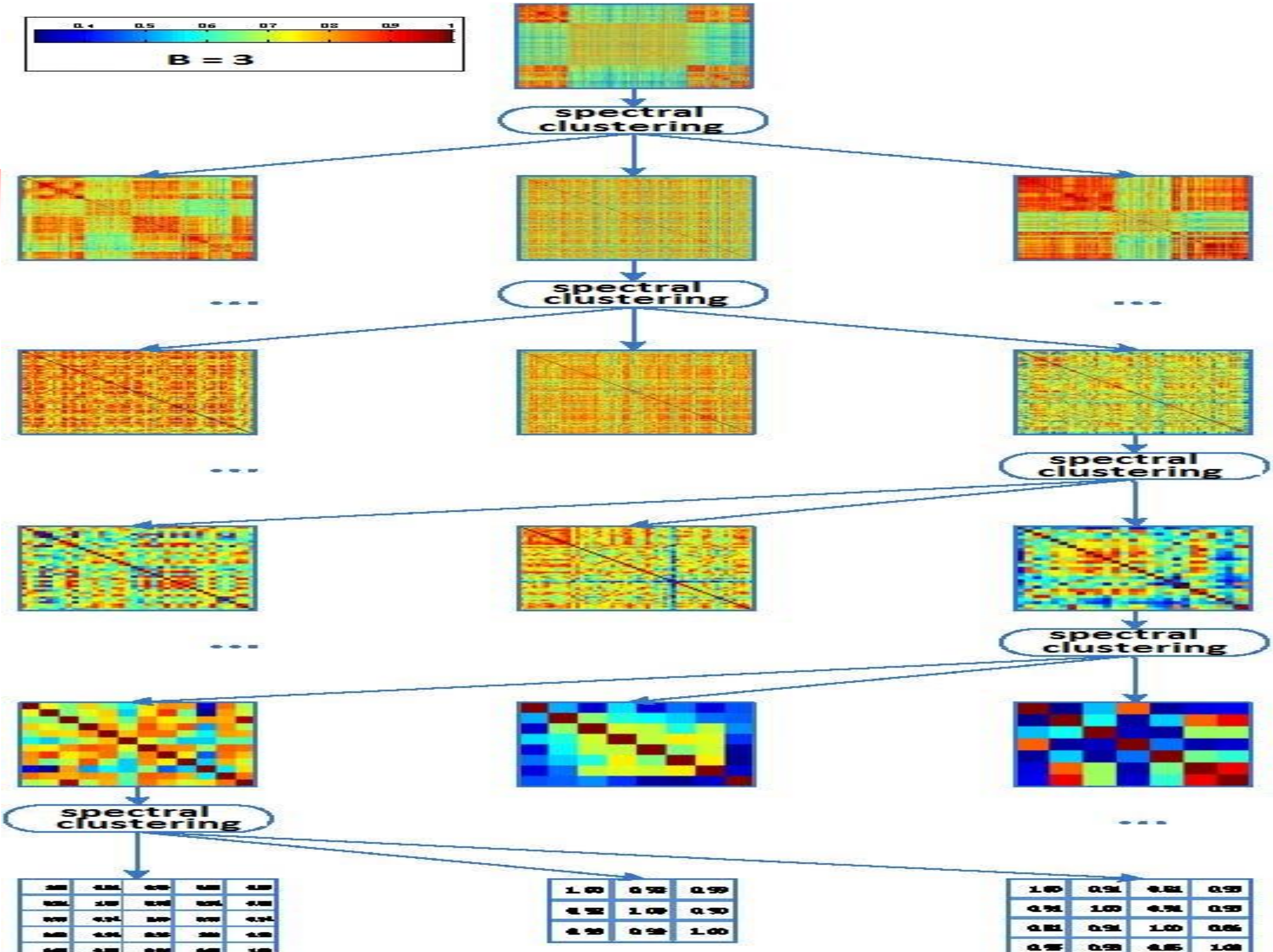


1.00	0.94
0.94	1.00

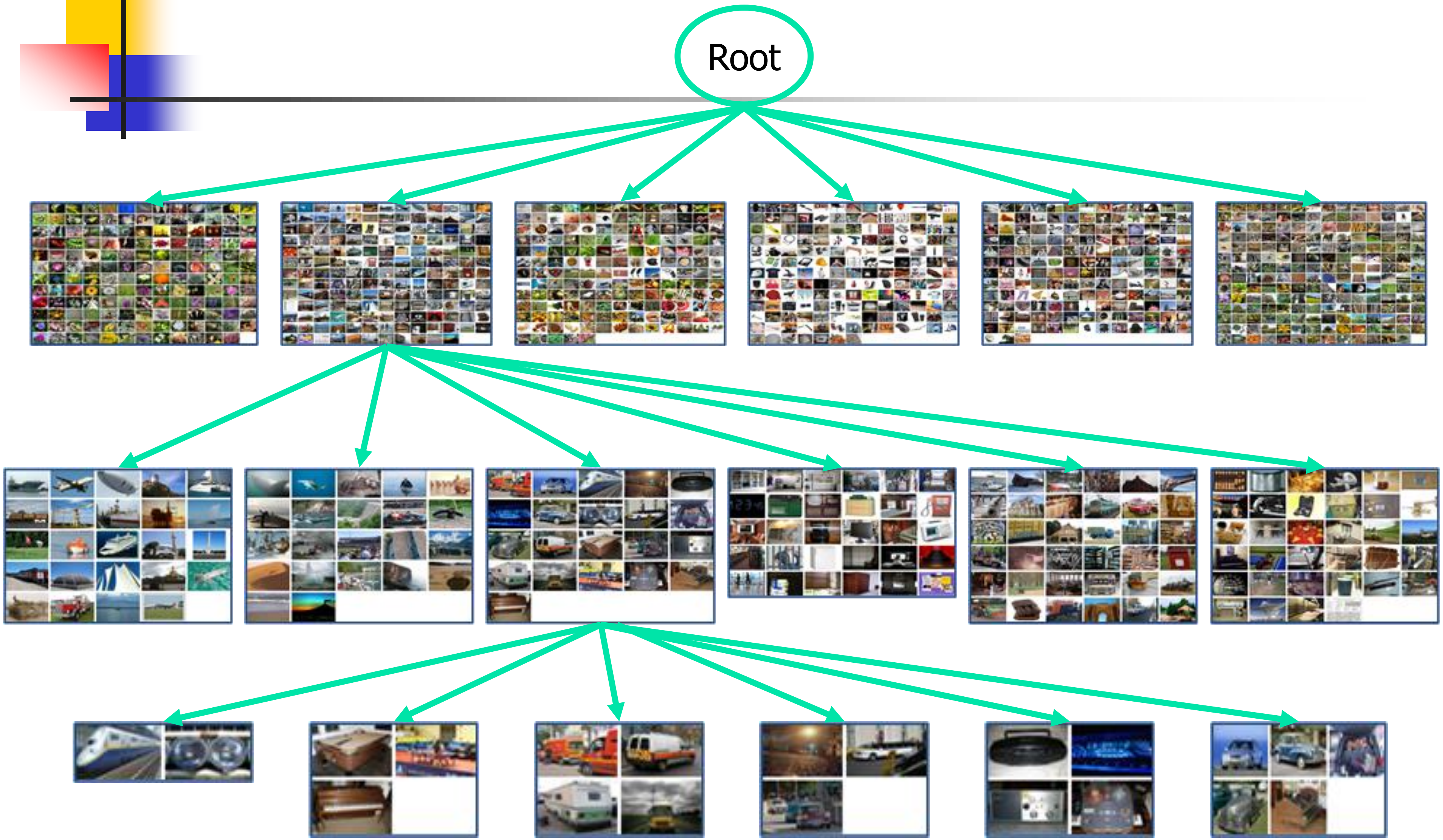
1.00	0.91	0.94
0.91	1.00	0.85
0.94	0.85	1.00

1.00	0.93	0.94
0.93	1.00	0.93
0.94	0.93	1.00

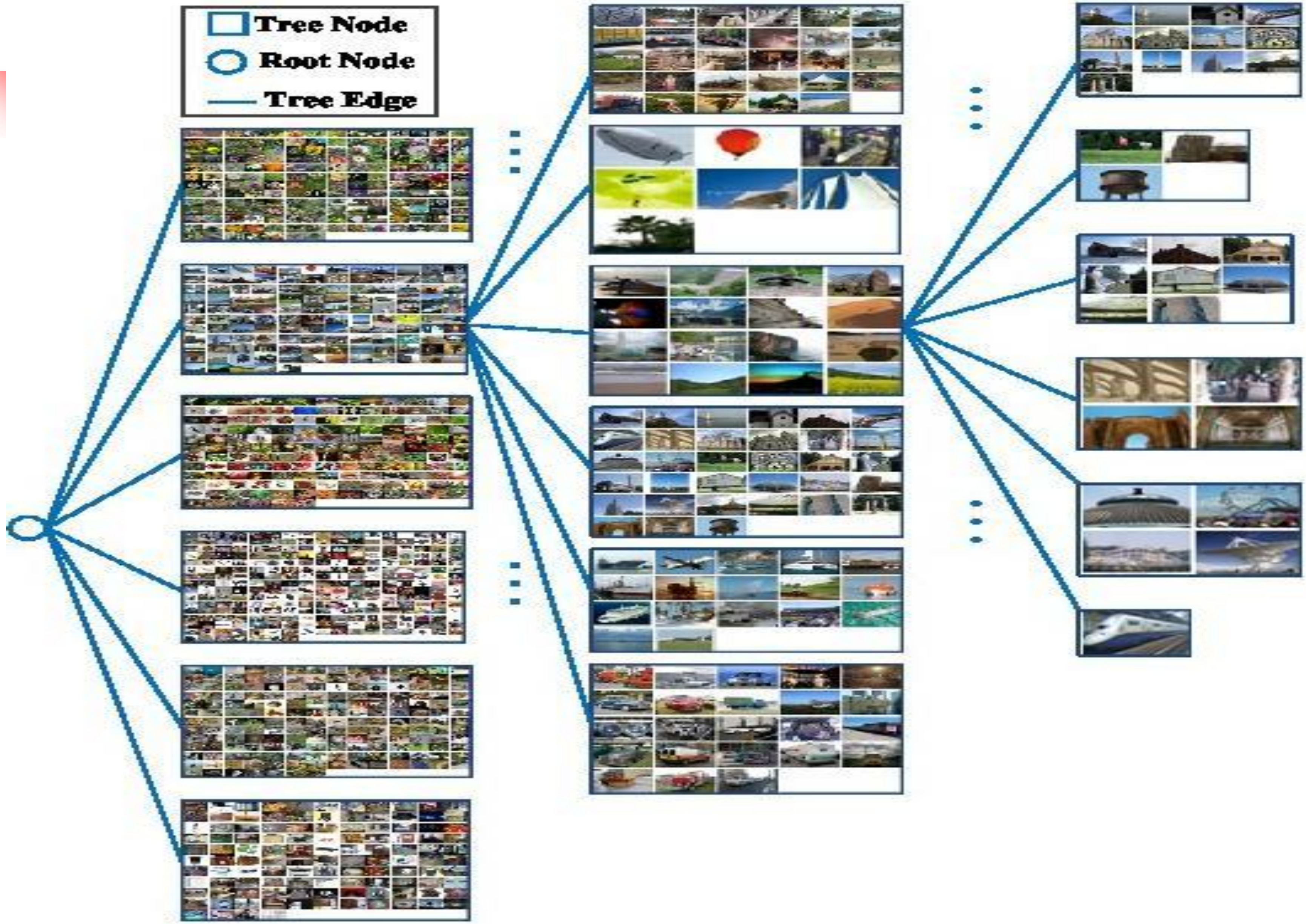
1.00	0.91	0.99
0.91	1.00	0.90
0.99	0.90	1.00


0.4 0.5 0.6 0.7 0.8 0.9 1
B = 3
spectral clustering
spectral clustering
spectral clustering
spectral clustering
spectral clustering

4. Visual Tree Construction: Large-Scale Object Classes



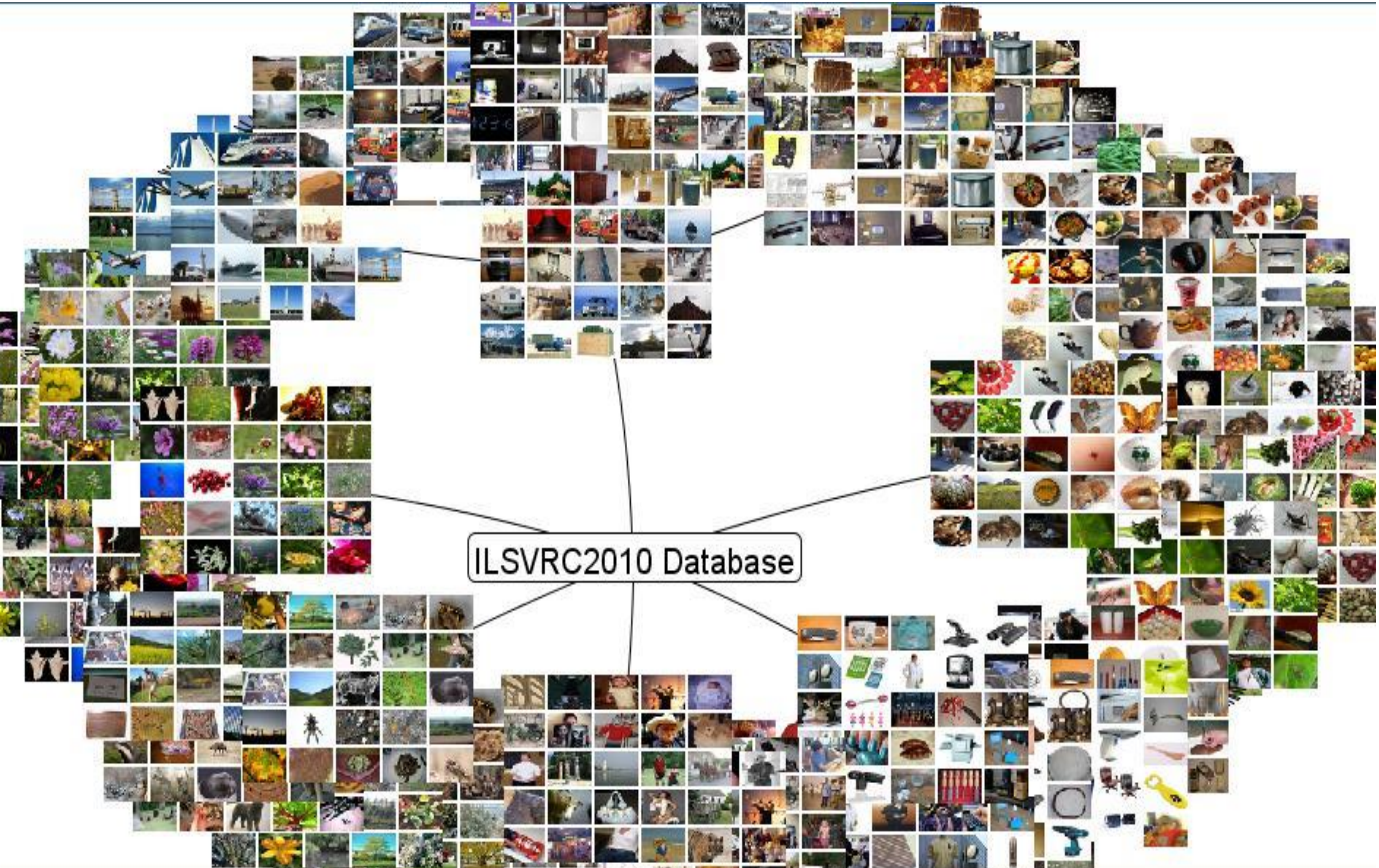
4. Visual Tree Construction: Large-Scale Object Classes



4. Visual Tree Construction: Large-Scale Object Classes

Visual_kmeans tree with $B=6$, $H=4$.
The leaf nodes are not shown.

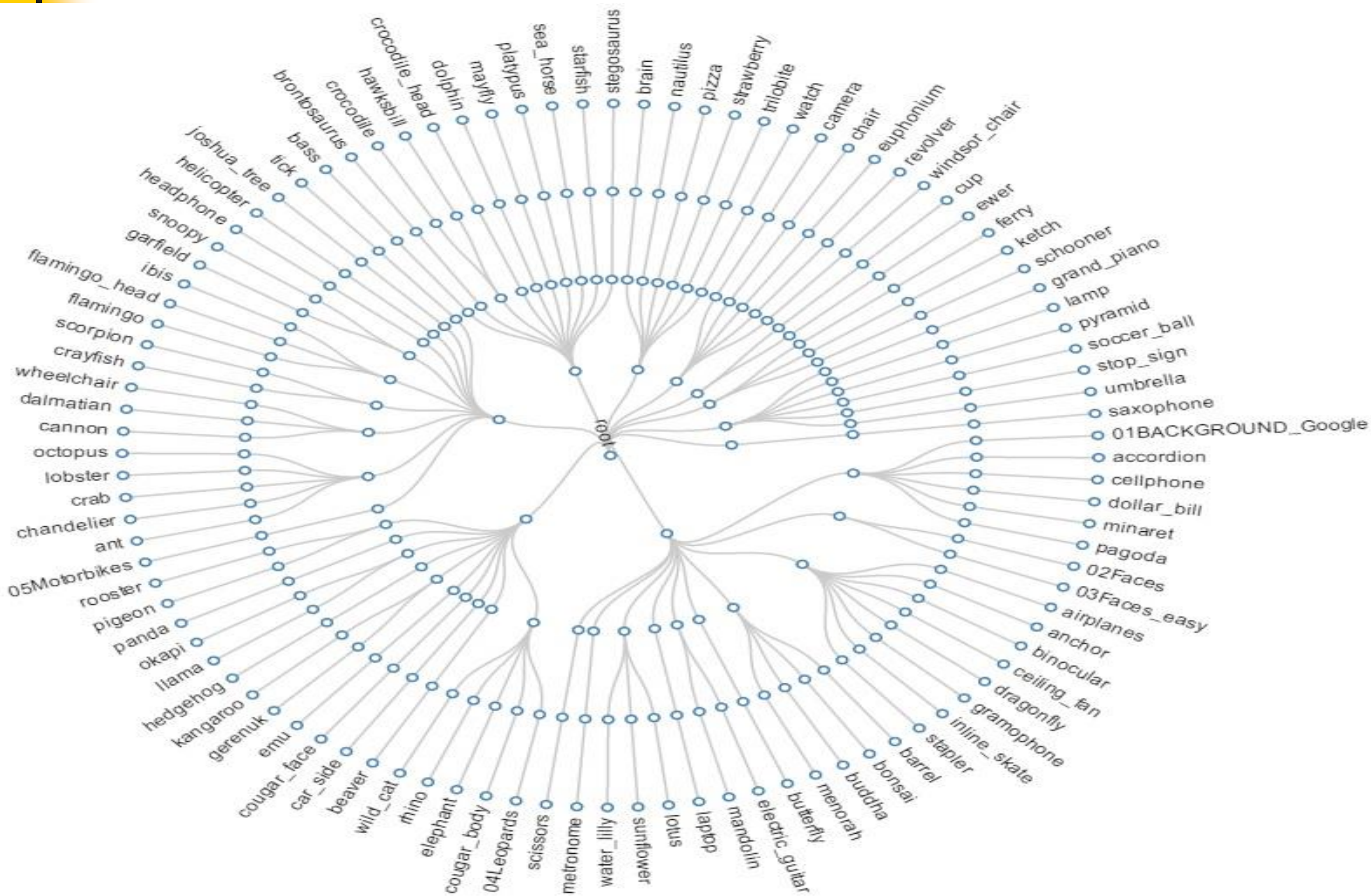
4. Visual Tree Construction: Large-Scale Object Classes



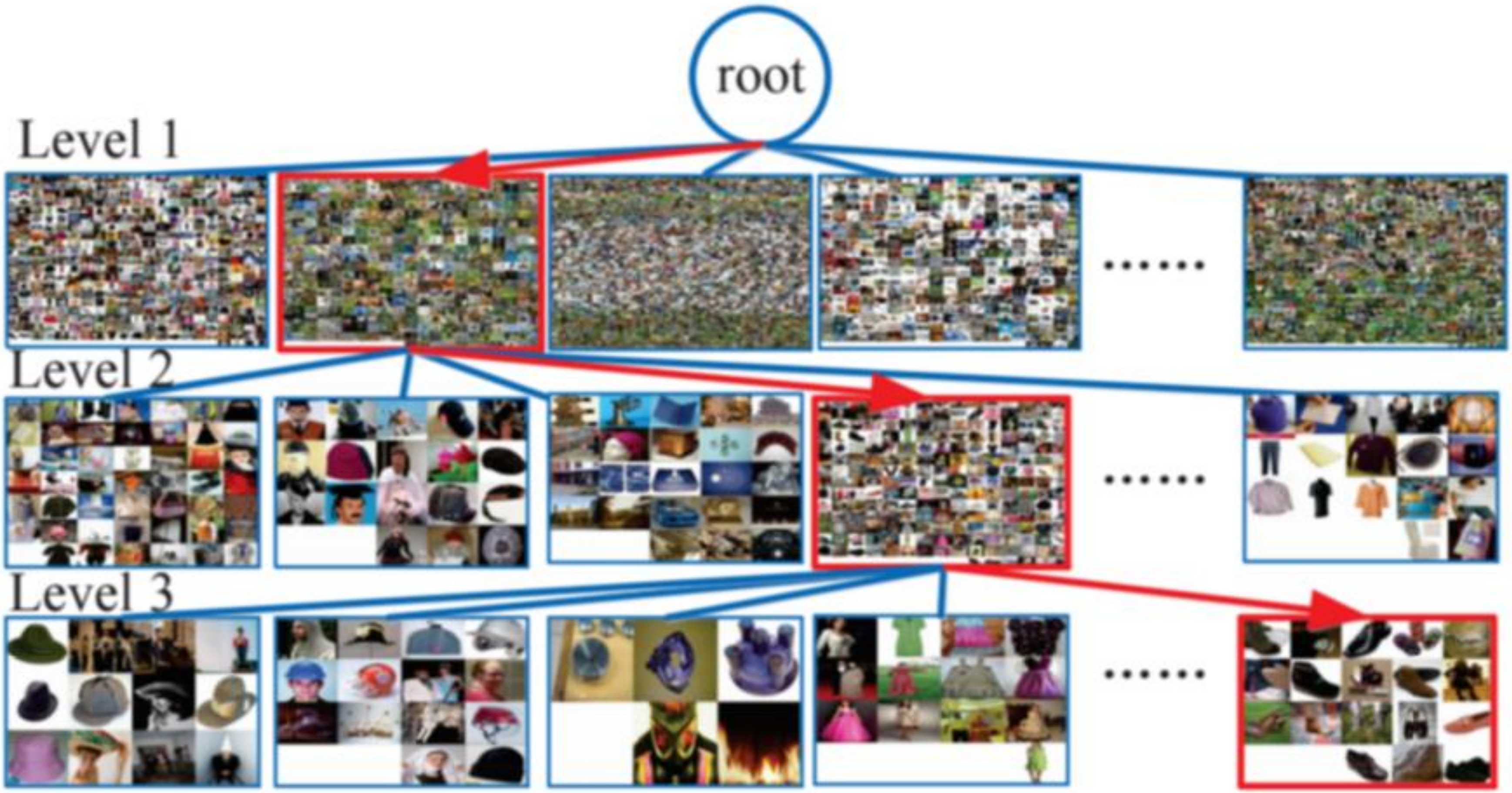
4. Visual Tree Construction: Large-Scale Object Classes

(a) (b)

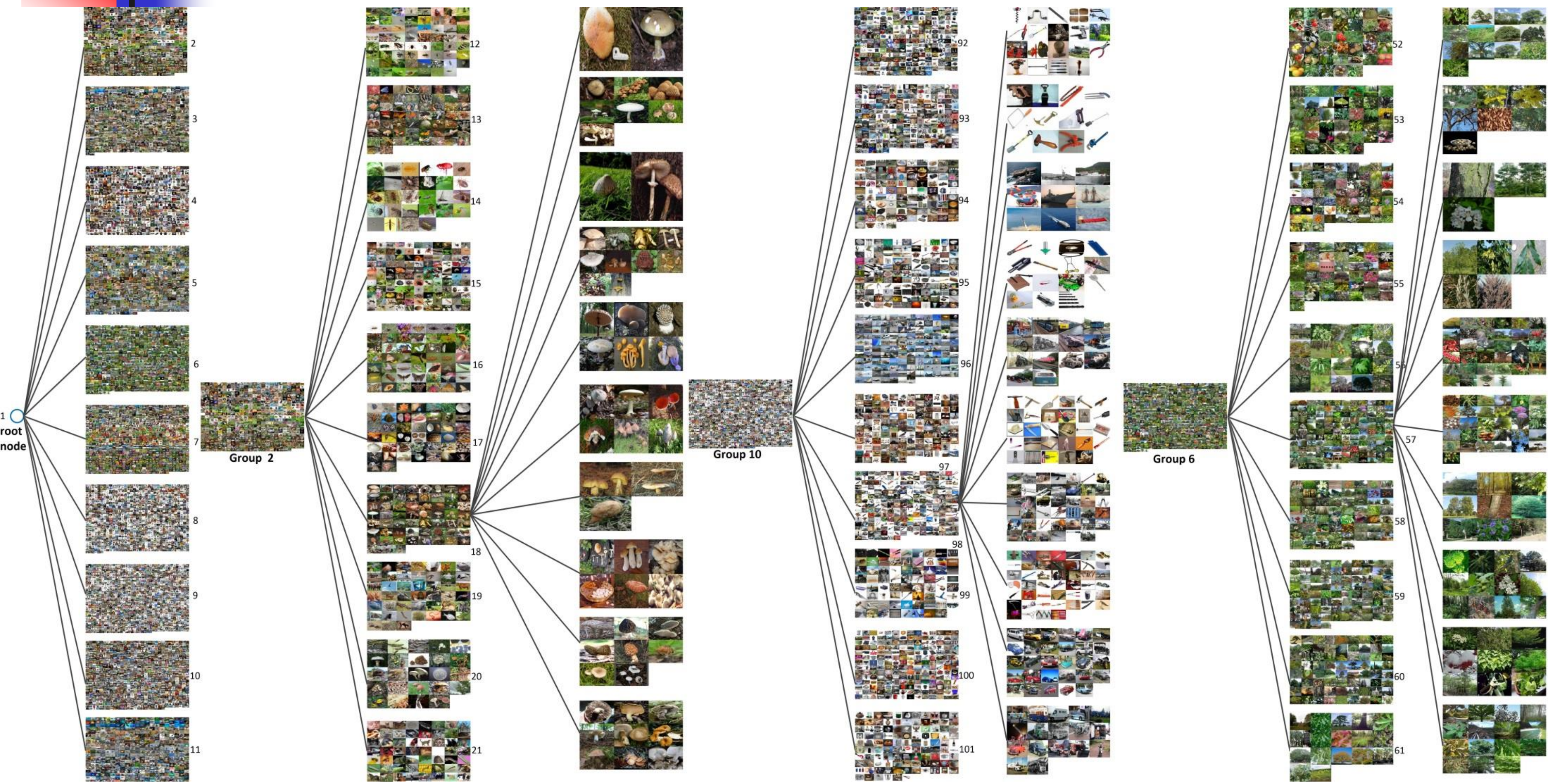
4. Visual Tree for CalTech101



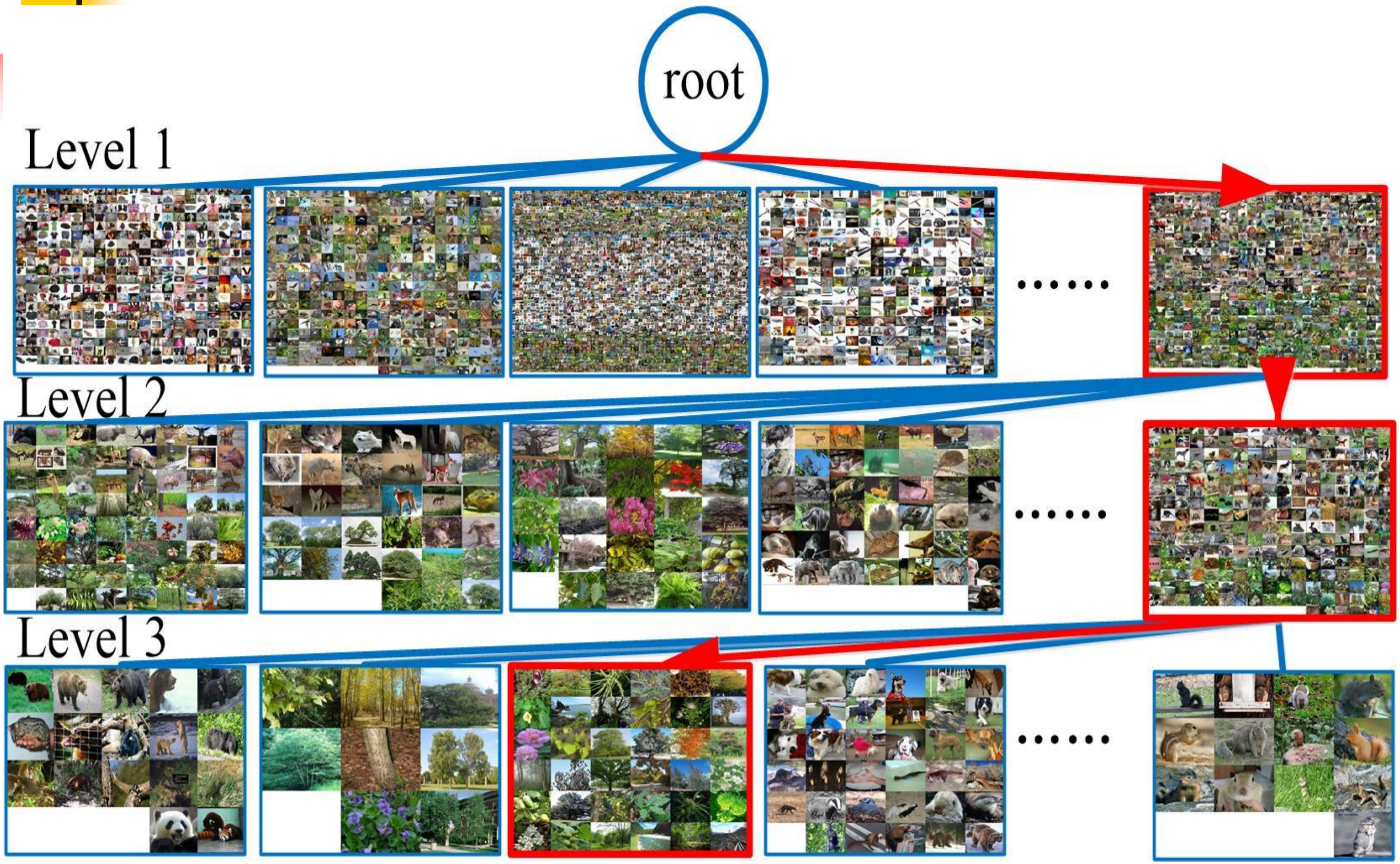
4. Visual Tree for ImageNet 10K



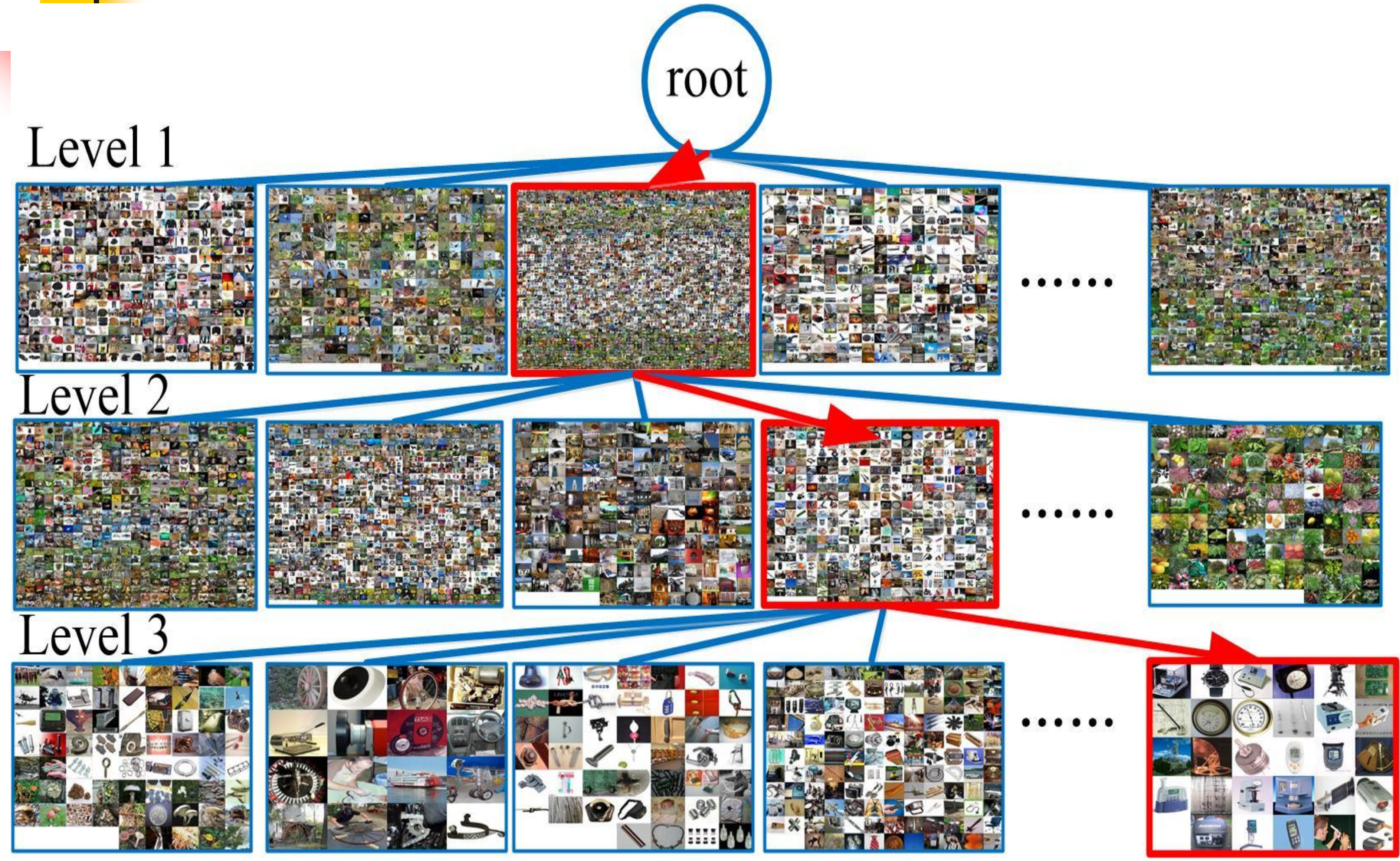
4. Visual Tree for ImageNet 10K



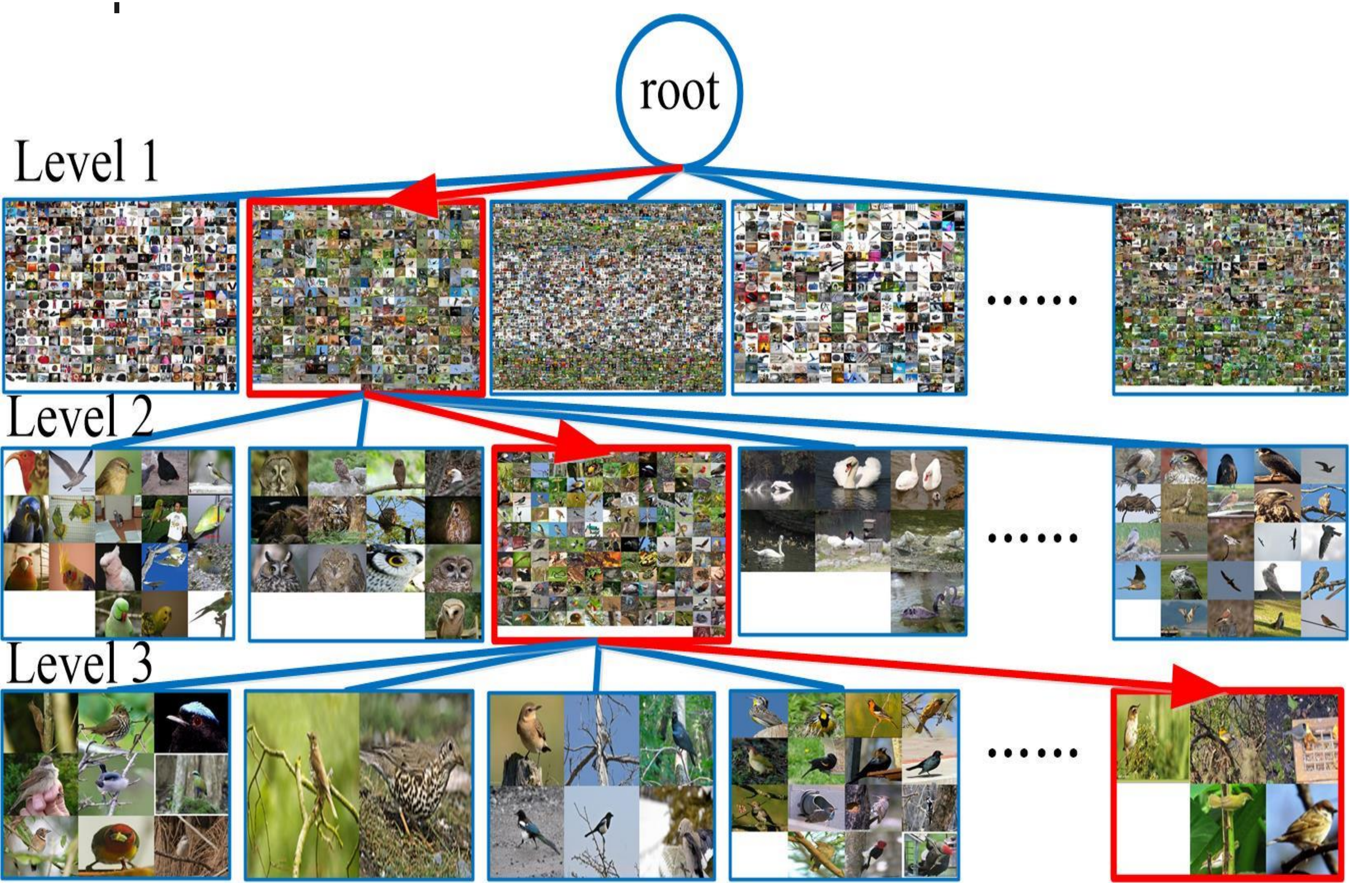
4. Visual Tree Construction



4. Visual Tree Construction

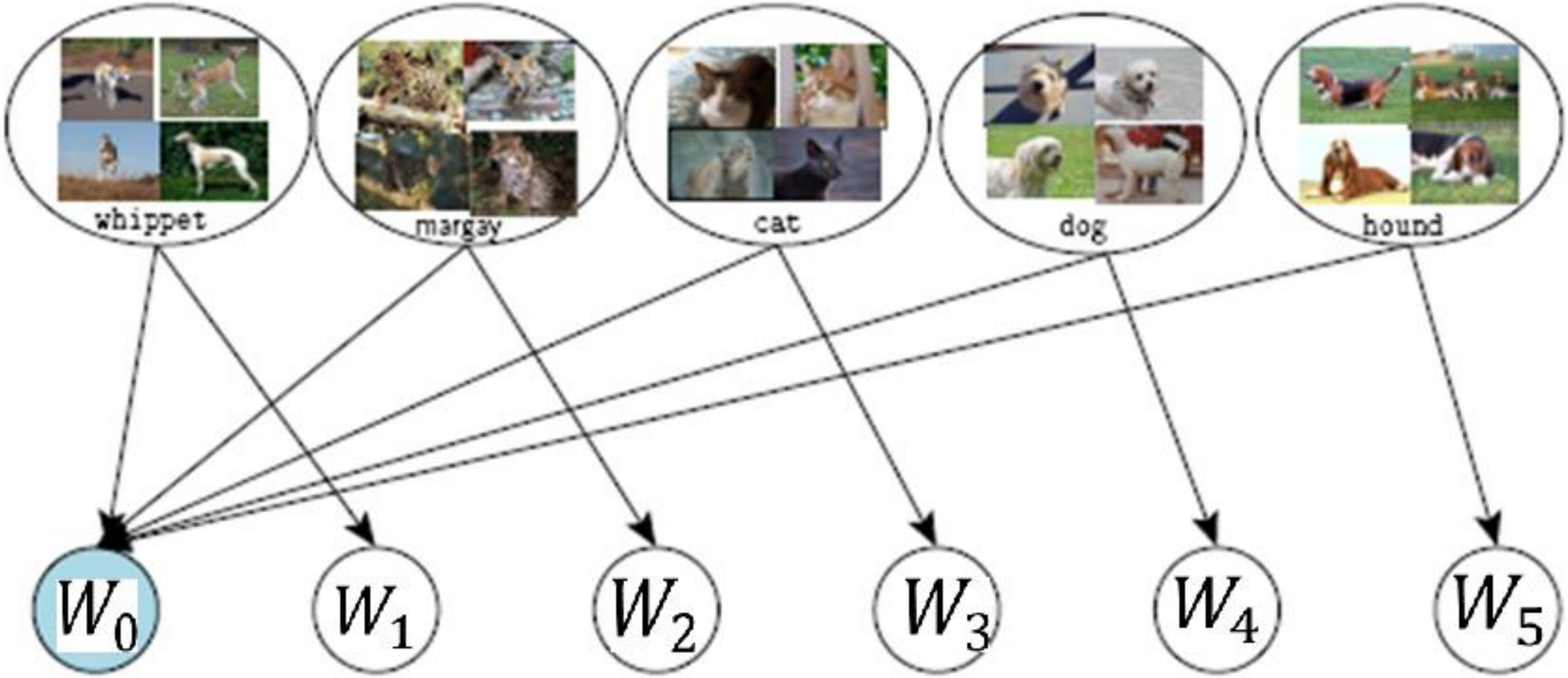


4. Visual Tree Construction



5. Hierarchical Deep Multi-Task Learning

- **Deep Multi-Task Learning**



5. Hierarchical Deep Multi-Task Learning

- **Deep Multi-Task Learning**

$$min\left\{C\sum_{l=1}^{R}\sum_{j=1}^{B}\xi_j^l+\delta_1 Tr\left(WW^T\right)+\frac{\delta_2}{2}Tr\left(WLW^T\right)\right\}$$

subject to:

$$\forall_{l=1}^{R}\forall_{j=1}^{B}:\ \ y_j^l\left(W_j^T\cdot x_j^l+b\right)\geq 1-\xi_j^l,\ \ \xi_j^l\geq 0$$

5. Hierarchical Deep Multi-Task Learning

- **Deep Multi-Task Learning**

$$min \left\{ \sum_{j=1}^{B} \sum_{l=1}^{R} \beta_l^j - \frac{1}{2\delta_1} \beta^T Y \Re \left(\Re + \frac{\delta_2}{\delta_1} \Re \left(L \bigotimes I \right) \Re \right)^{-1} \Re Y \beta \right\}$$

subject to:

$$\forall_{l=1}^{R} \forall_{j=1}^{B} : \sum_{l=1}^{R} \beta_l^j \cdot y_l^j = 0, \quad 0 \leq \beta_l^j \leq 1$$

5. Hierarchical Deep Multi-Task Learning

- **Deep Multi-Task Learning**

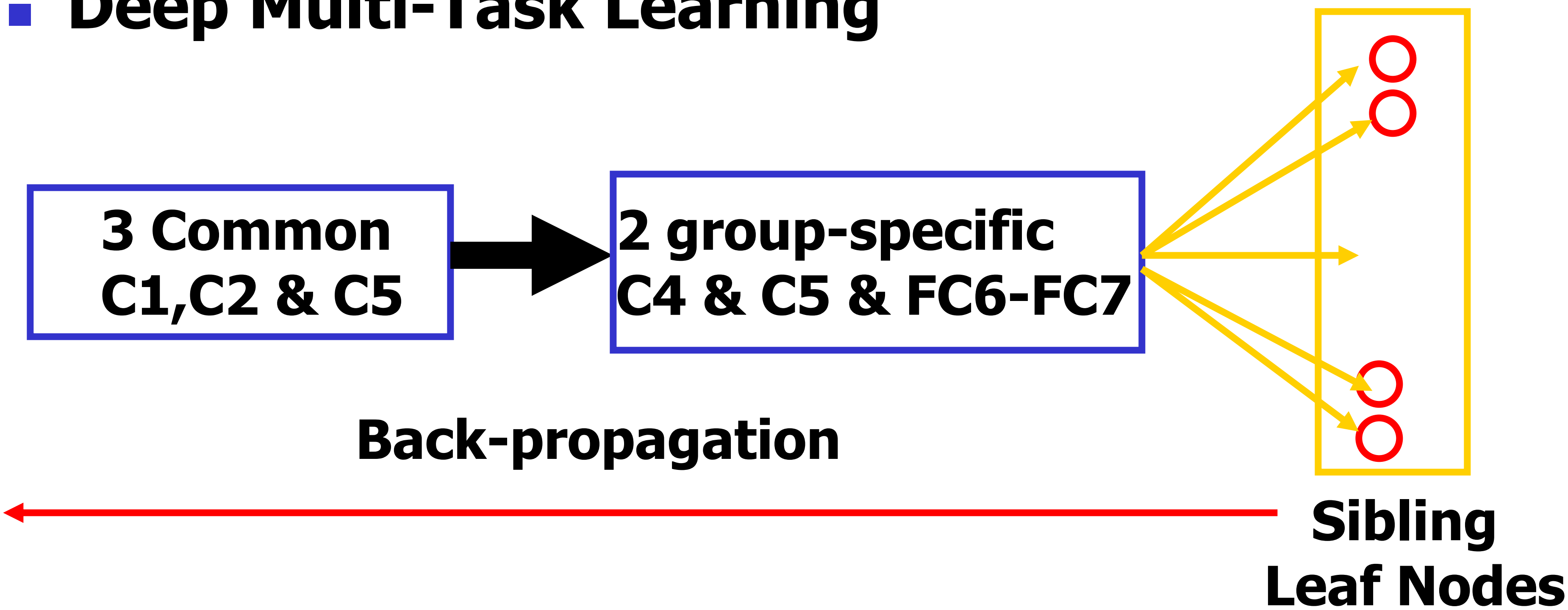
$$\alpha^* = \frac{1}{2\delta_1}\left(\Re + \frac{\delta_2}{\delta_1}\left(\Re\left(L \bigotimes I\right)\Re\right)^{-1}\Re Y \beta^*\right)$$

Multi-Task Classifiers at Sibling Leaf Nodes

$$\forall_{j=1}^{B}: \quad f_{c_j}^{1}(x)\,|_{F_{c_j}^{1}} = \sum_{l=1}^{R} \alpha_j^{l*}\kappa(x_j^l, x) + b_j^*, \quad c_j \in c_h$$

5. Hierarchical Deep Multi-Task Learning

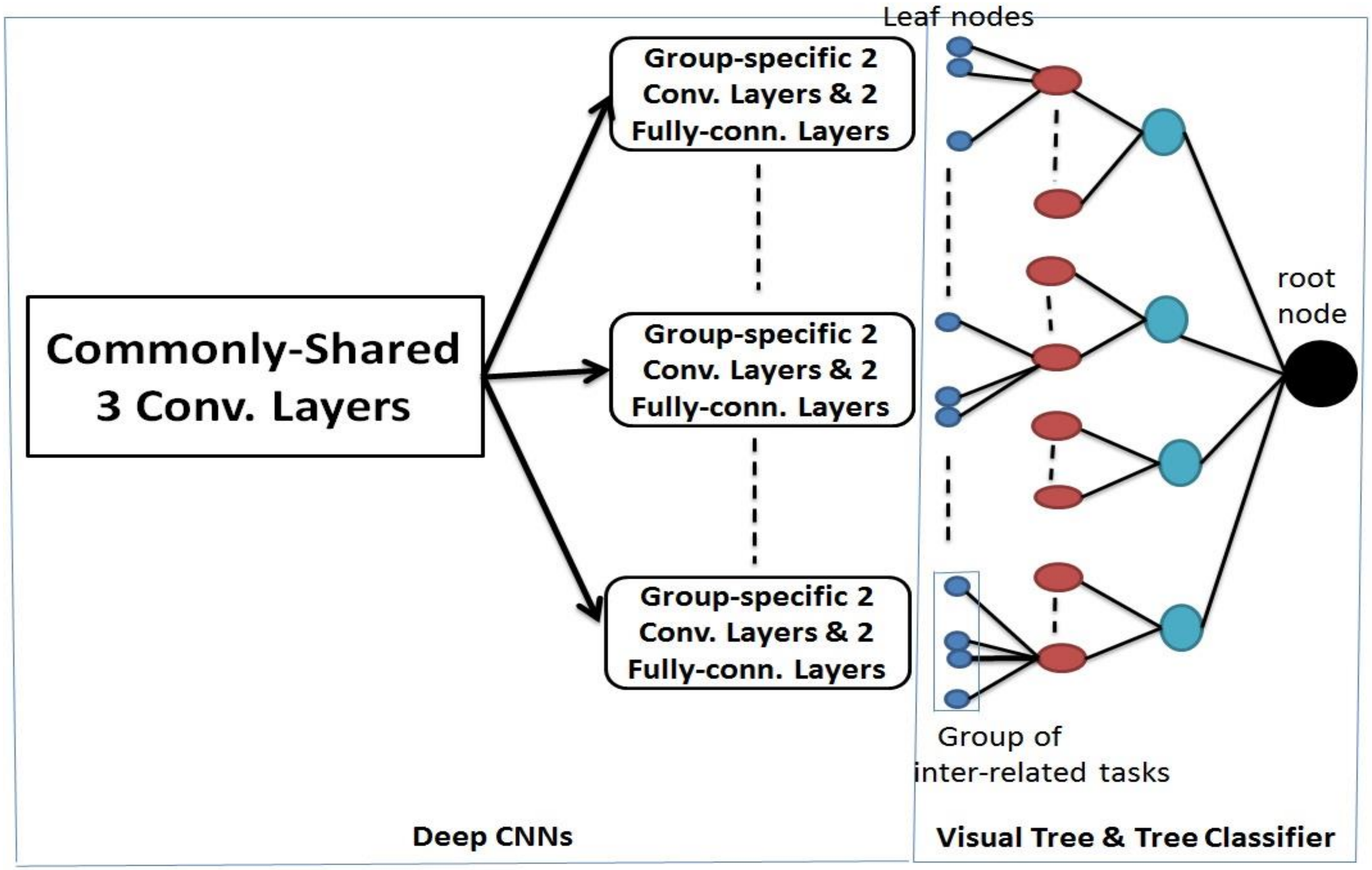
- **Deep Multi-Task Learning**



$$\frac{\partial \pounds(W,X,Y)}{\partial W} = W(\delta_1 I+\delta_2 L)-\mu \sum_{l=1}^{R}\sum_{j=1}^{B} I\{y_j^l\}\left[x_j^l - \sum_{j=1}^{B} x_j^l\right]$$

5. Hierarchical Deep Multi-Task Learning

- **Hierarchical Deep Multi-Task Learning**



5. Hierarchical Deep Multi-Task Learning

- **Hierarchical Deep Multi-Task Learning**

$$min\left\{C\sum_{m=1}^{R}\sum_{h=1}^{B}\xi_j^m+\gamma_1 Tr\left(WW^T\right)+\frac{\gamma_2}{2}Tr\left(WLW^T\right)\right\}$$

subject to:

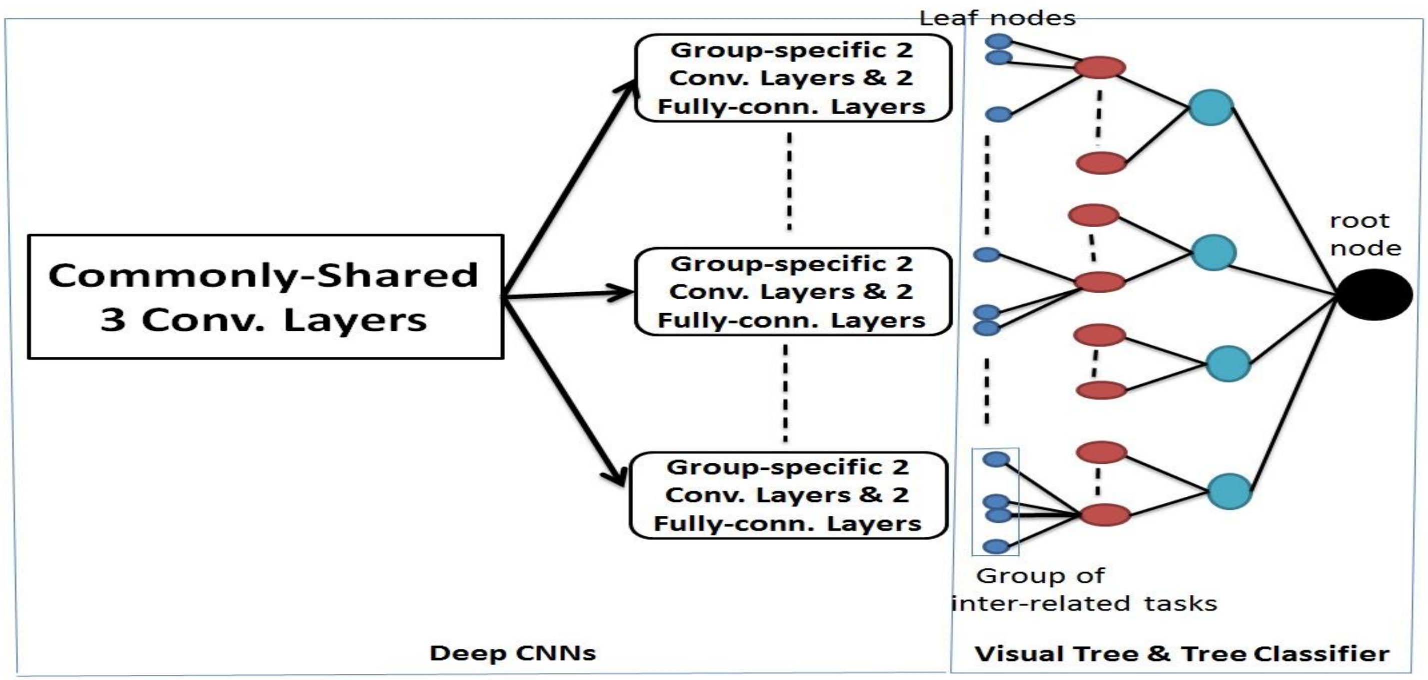
$$\forall_{m=1}^{R}\forall_{h=1}^{B}:\ y_h^m\left(W_h^T\cdot x_h^m+b\right)\geq 1-\xi_h^m,\ \xi_h^m\geq 0,\ c_h\in c_k$$

$$\forall_{h=1}^{B}:\quad f_{c_h}^{l+1}(x)\mid_{F_{c_h}^{l+1}}-f_{c_j}^{l}(x)\mid_{F_{c_j}^{l}}\geq 0$$

$$\forall_{h=1}^{B}:\quad f_{c_h}^{l+1}(x)\mid_{F_{c_h}^{l+1}}=\sum_{j=1}^{B}\eta_j f_{c_j}^{l}(x)\mid_{F_{c_j}^{l}}$$

5. Hierarchical Deep Multi-Task Learning

- **Hierarchical Deep Multi-Task Learning**



Back-propagation

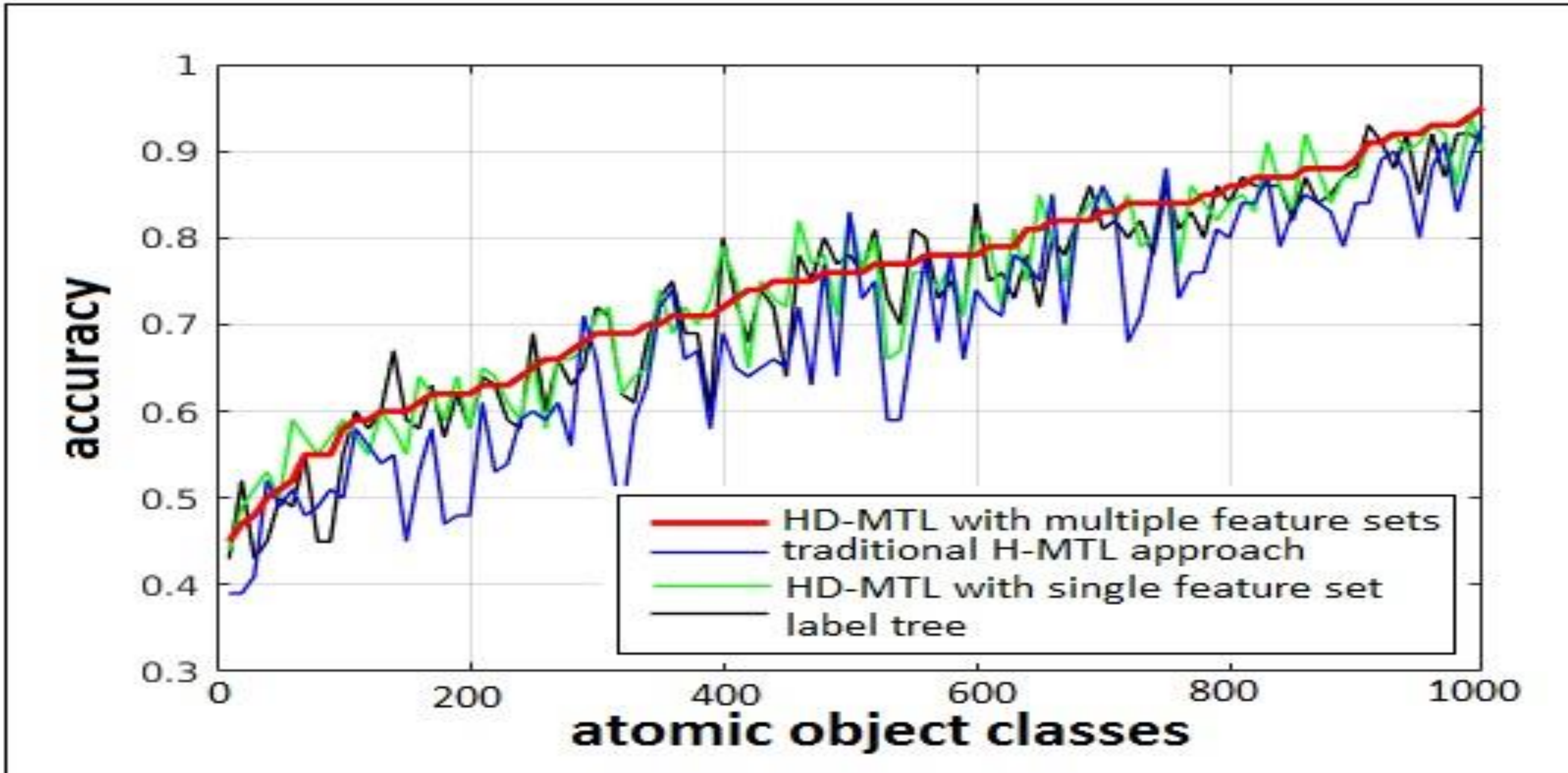
$$\frac{\partial \Psi(W,X,Y)}{\partial W} = \Omega(\gamma_1 I + \gamma_2 \Xi) - \nu \sum_{m=1}^{N} \sum_{h=1}^{L} I\{y_h^m\} \left[x_h^m - \sum_{h=1}^{L} x_h^m \right] + \lambda$$

5. Hierarchical Deep Multi-Task Learning

- **Back Propagation**
  - **Errors from High-Level Node**
    - **Node classifier for itself**
    - **Node Classifiers for lower-level nodes which treat it as their ancestors**
    - **Weights of deep networks**
  - **Errors from Leaf Node**
    - **Node classifier for itself**
    - **Weights of deep networks**

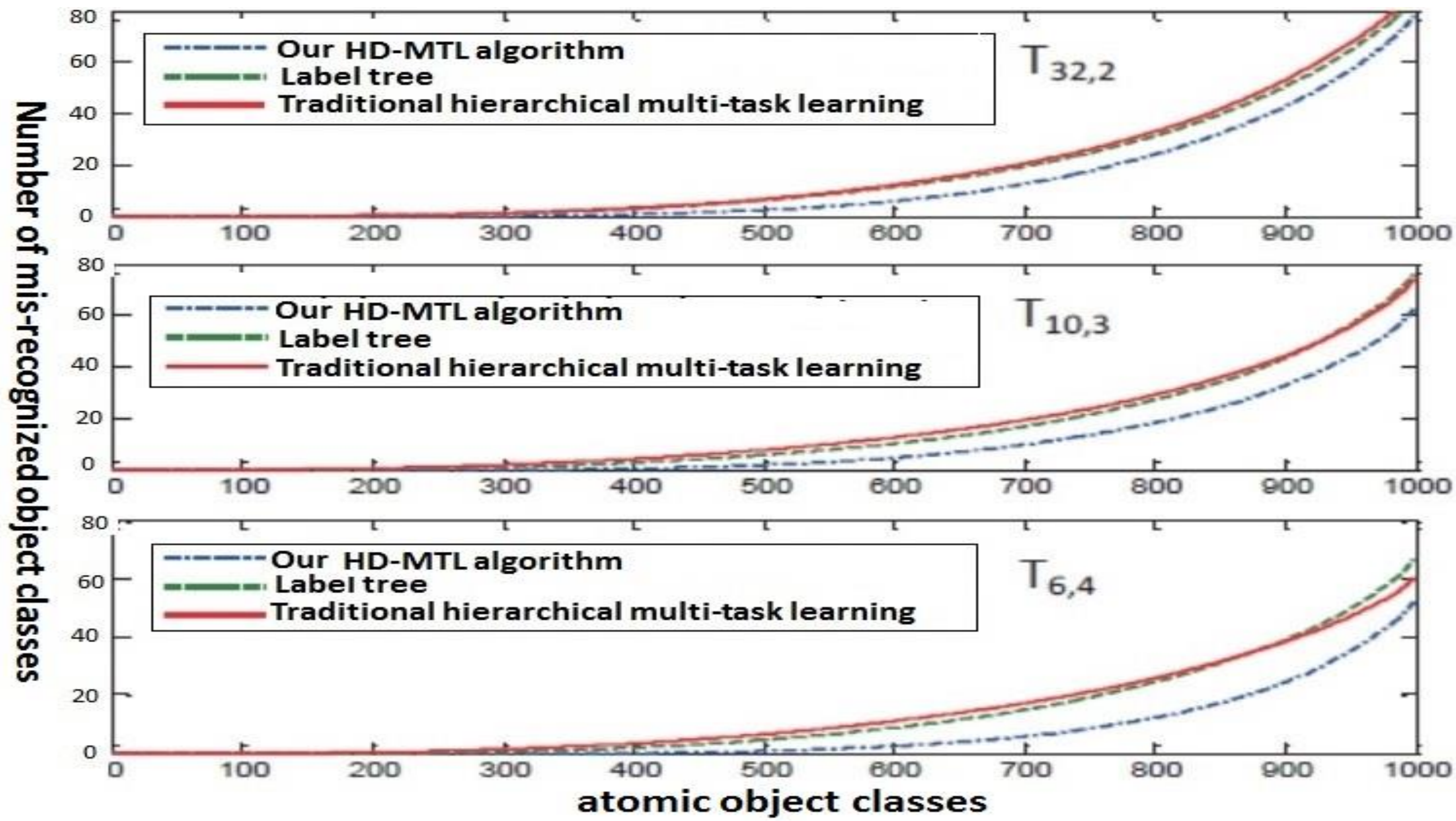
6. Some Experimental Results

- **Impacts of Feature Subset Selection**



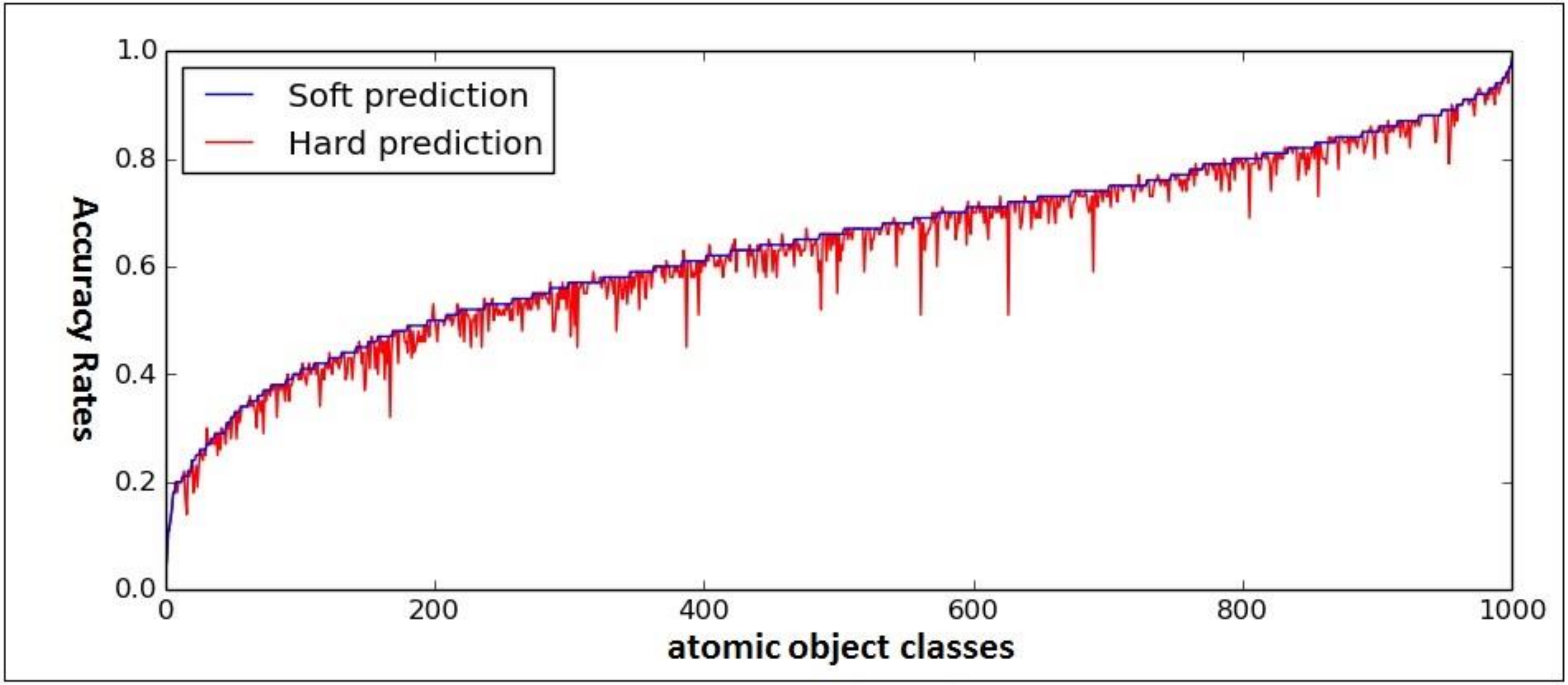
6. Some Experimental Results

- **Impacts of Feature Subset Selection**



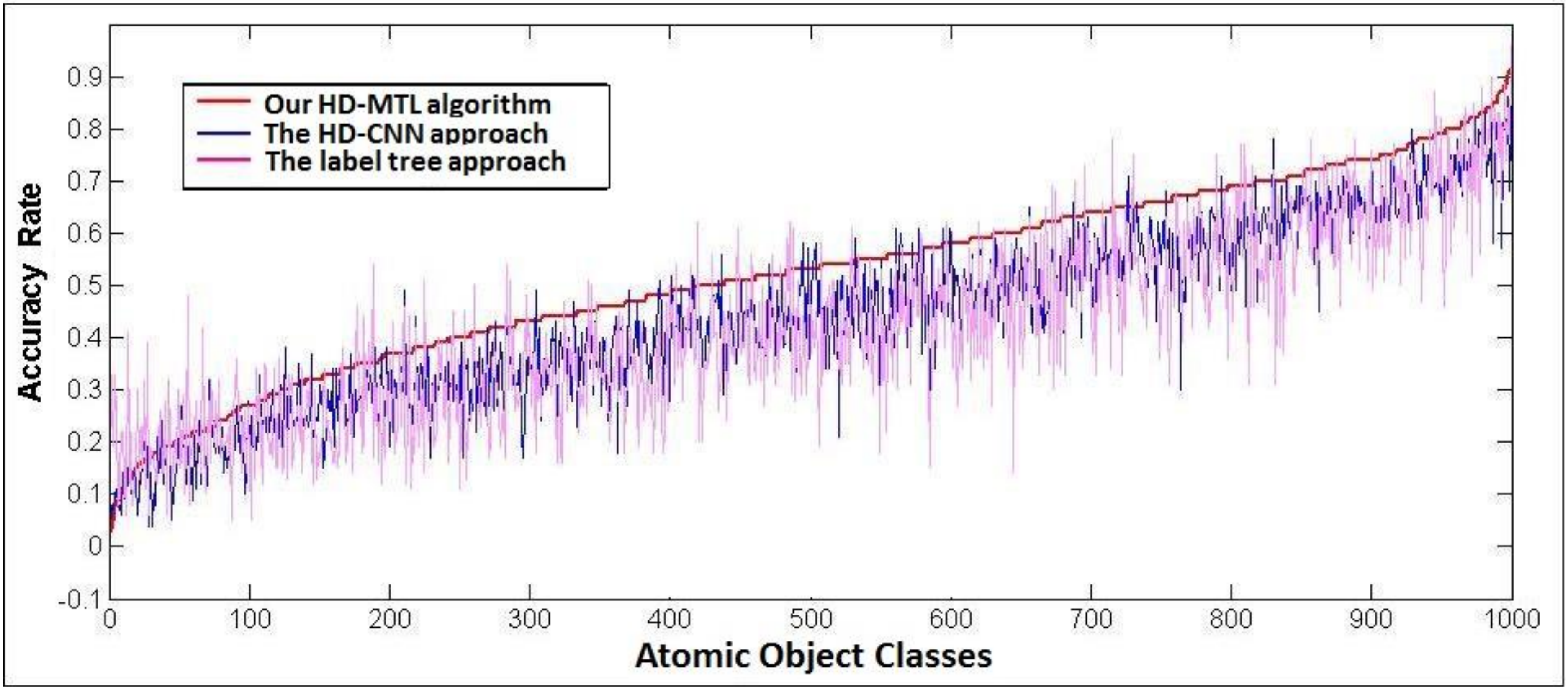
6. Some Experimental Results

- **Impacts of Soft Prediction**



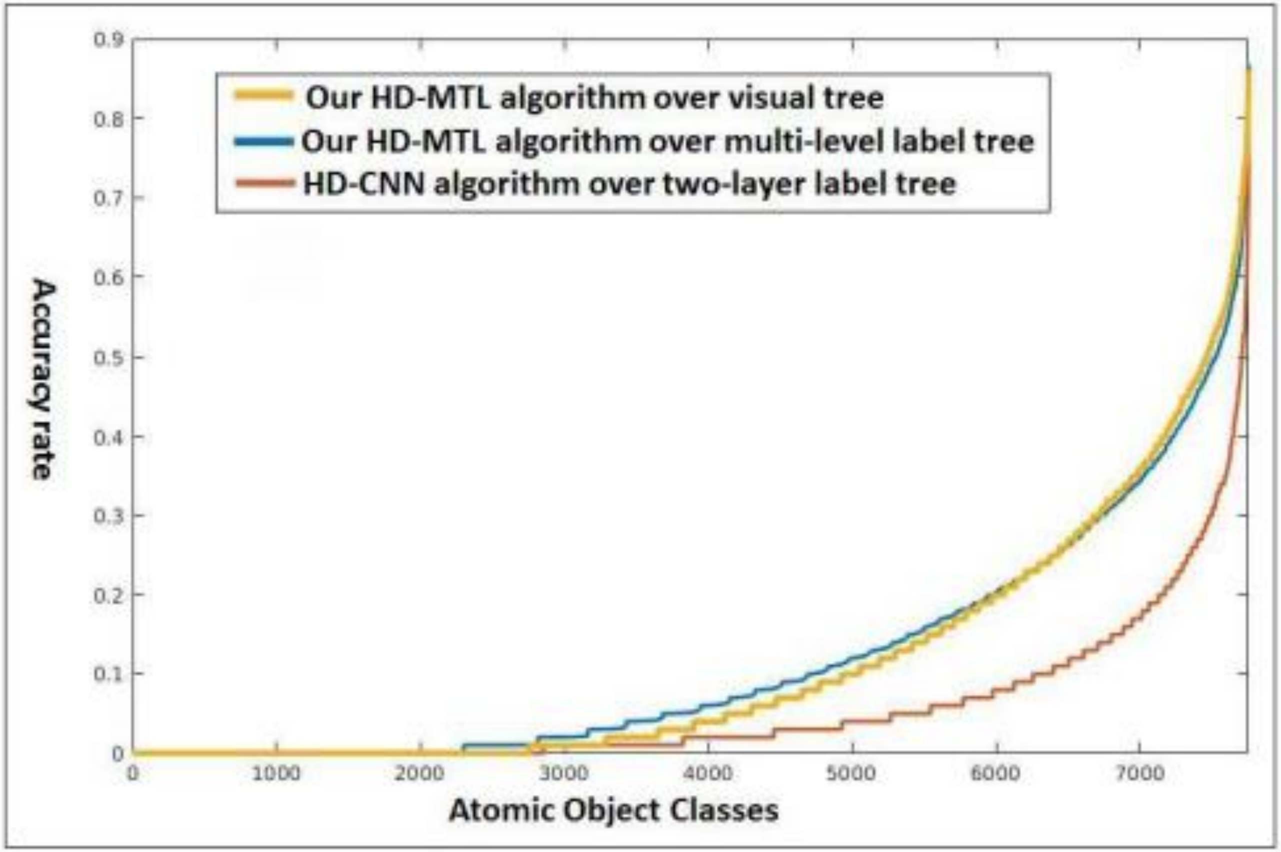
6. Some Experimental Results

- **Impacts of Deep Multi-Task Learning**



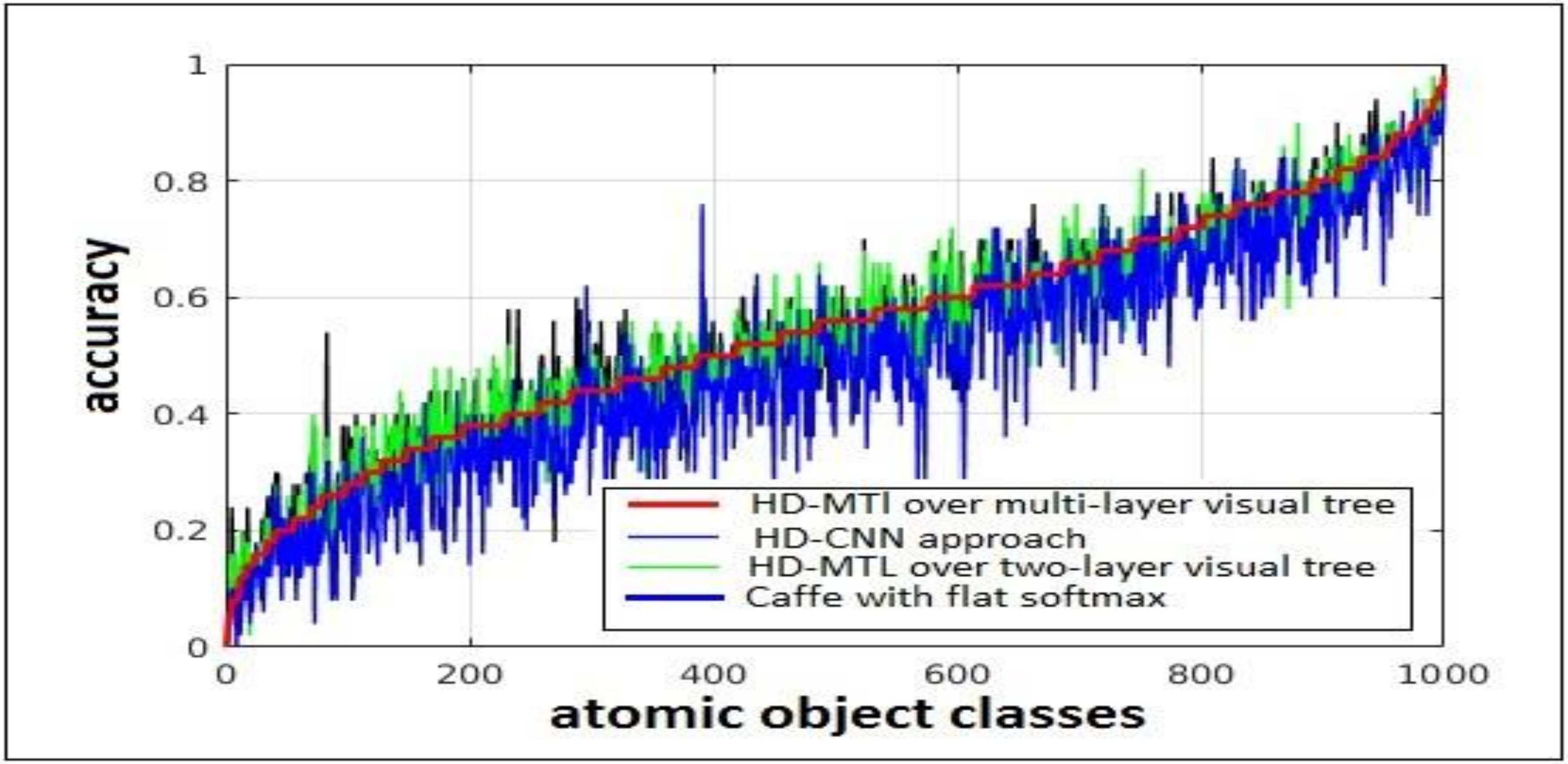
6. Some Experimental Results

- **Impacts of Deep Multi-Task Learning**



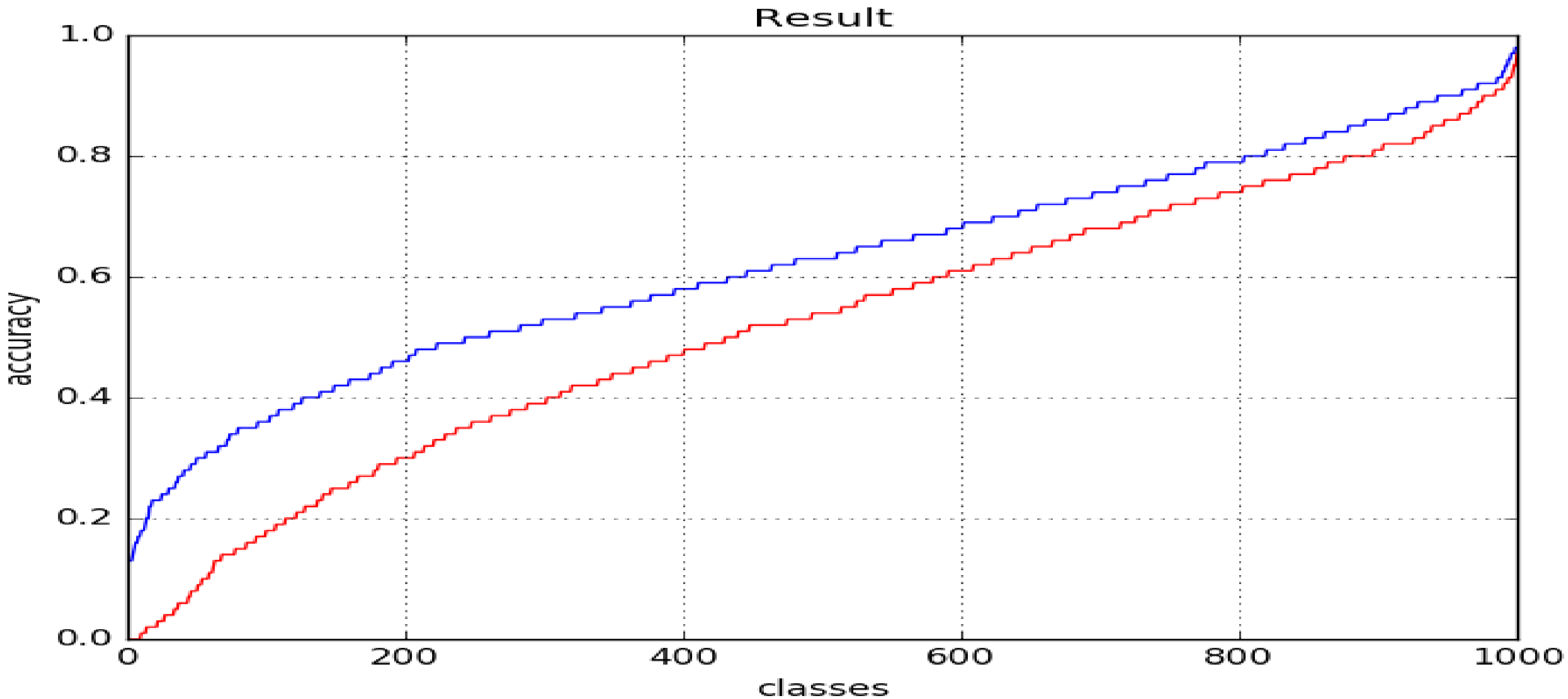
6. Some Experimental Results

- **Impacts of Visual Tree**



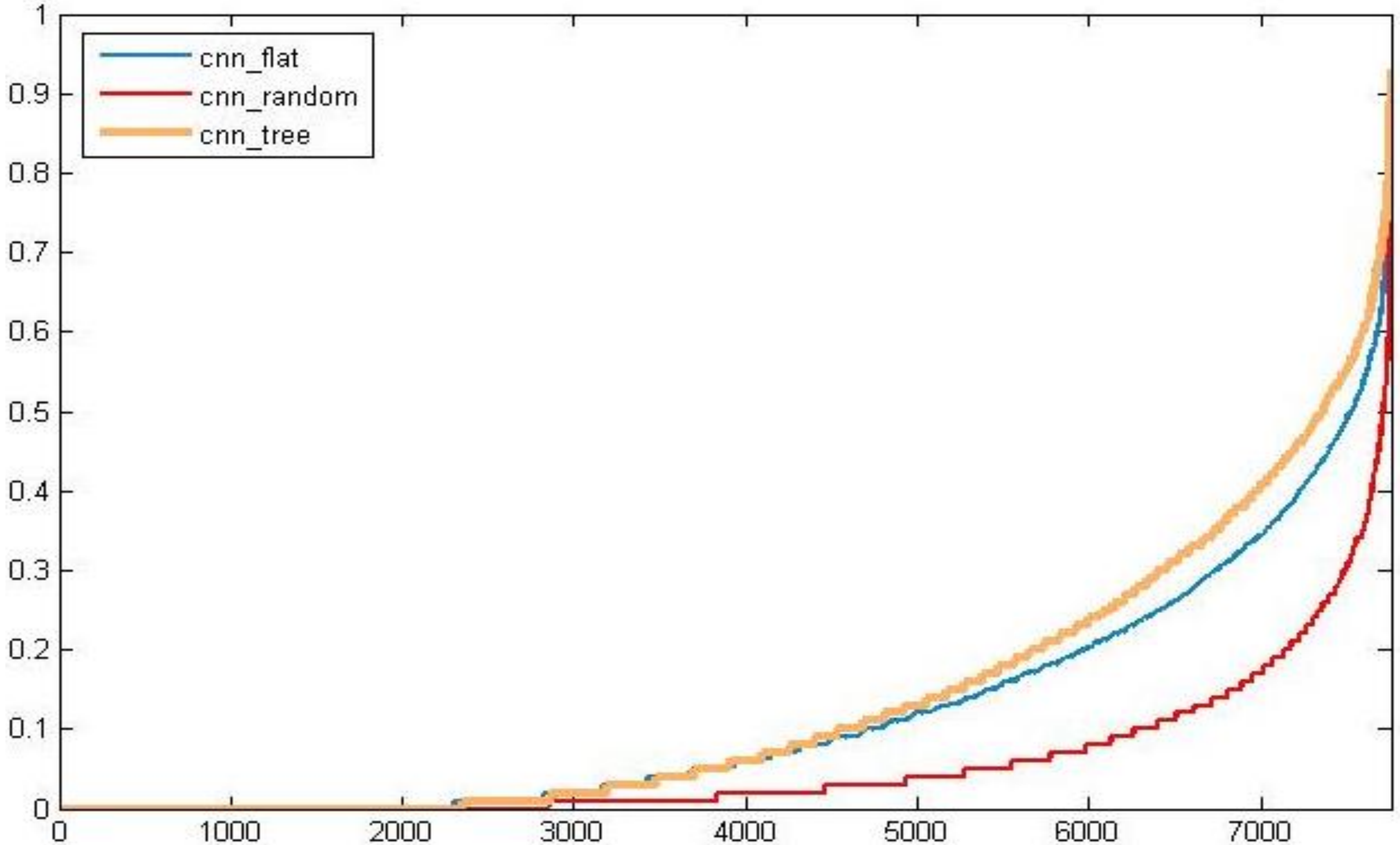
6. Some Experimental Results

- **Impacts of Visual Tree**

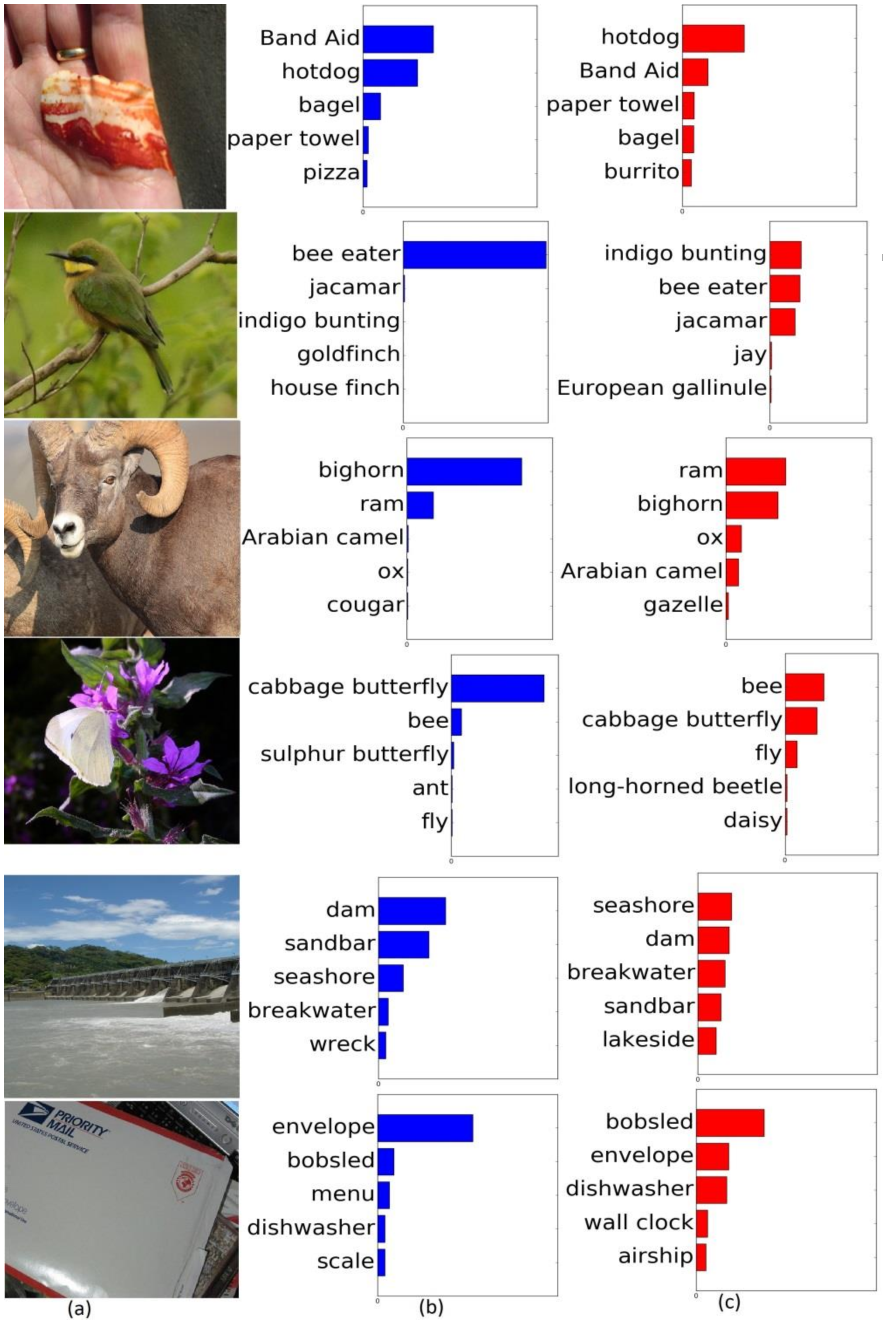


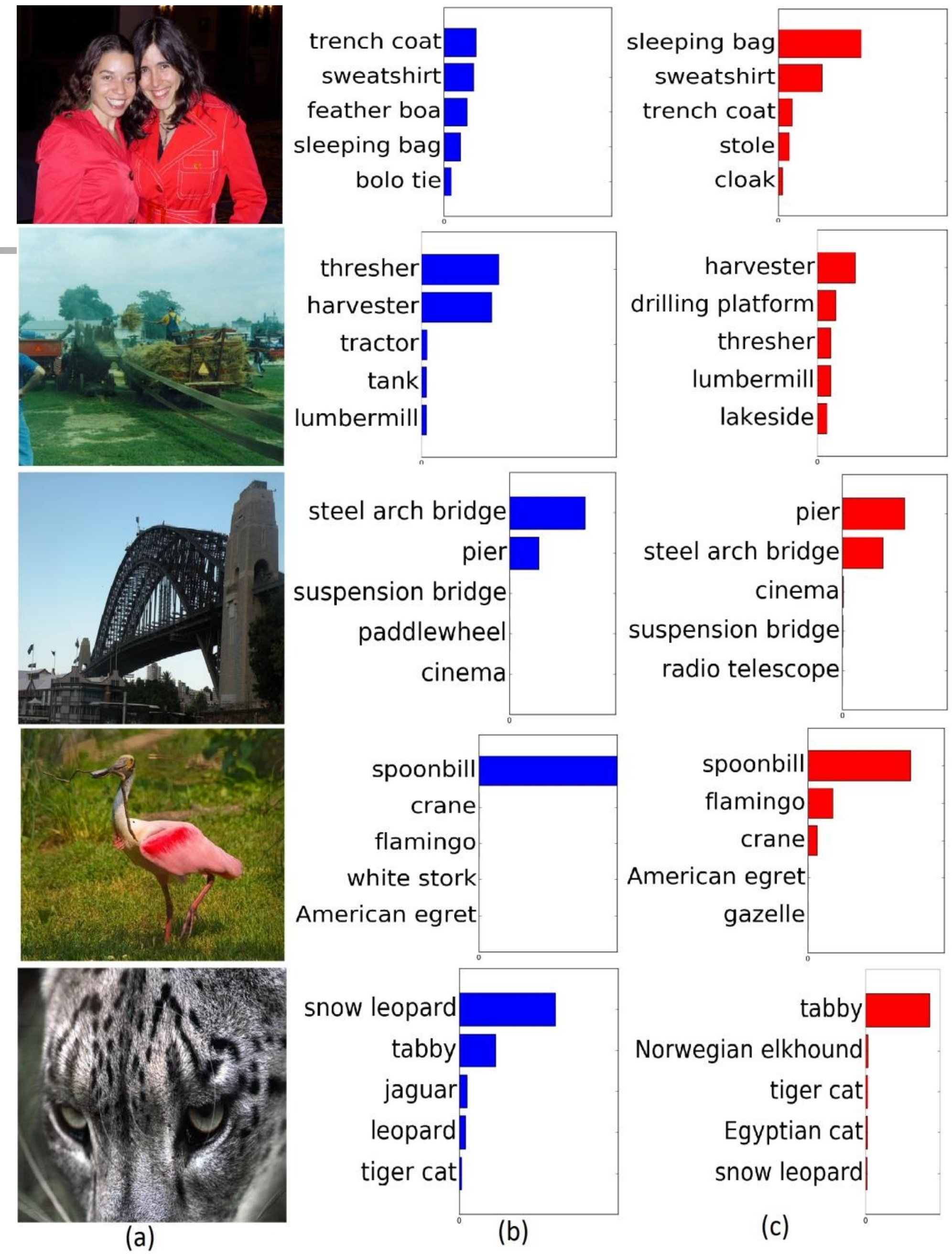
6. Some Experimental Results

- **Impacts of Visual Tree**



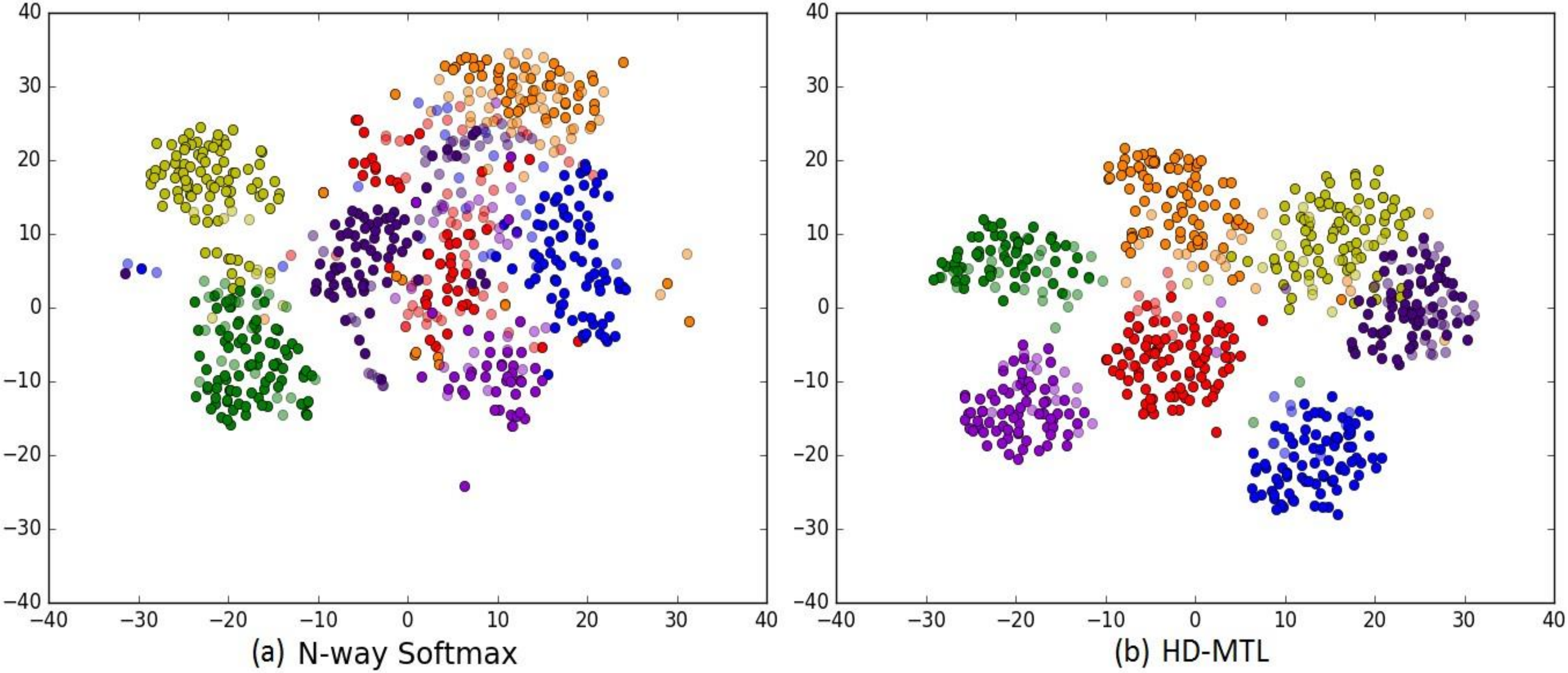
■ Prediction Confidence Enhancement





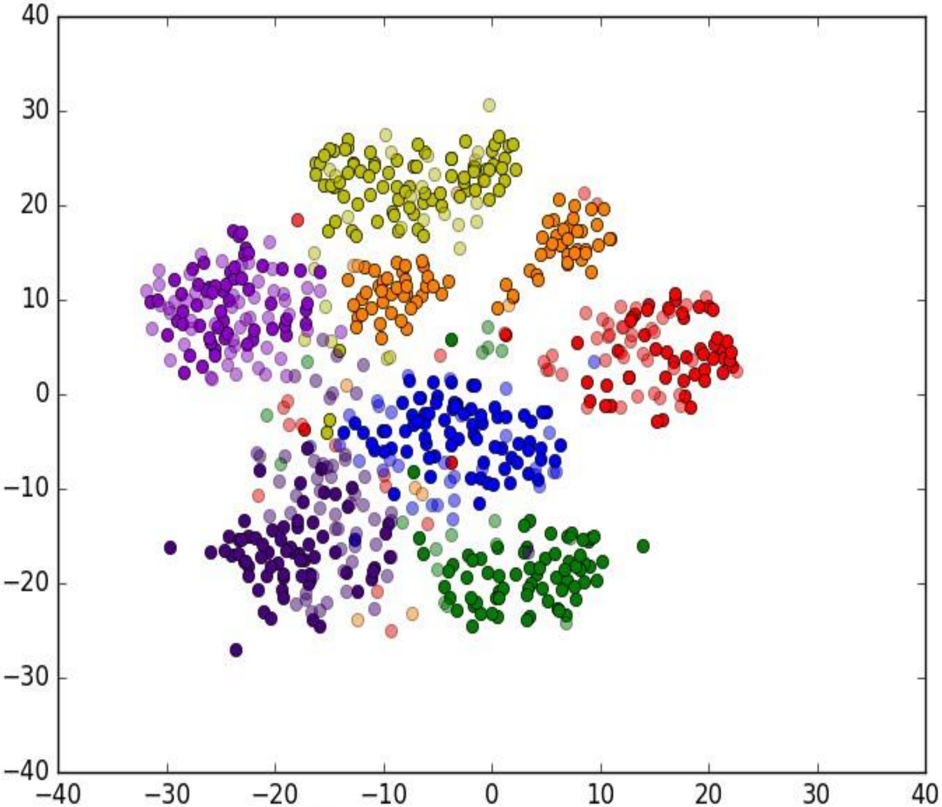
6. Some Experimental Results

- **Impacts of Deep Multi-Task Learning**

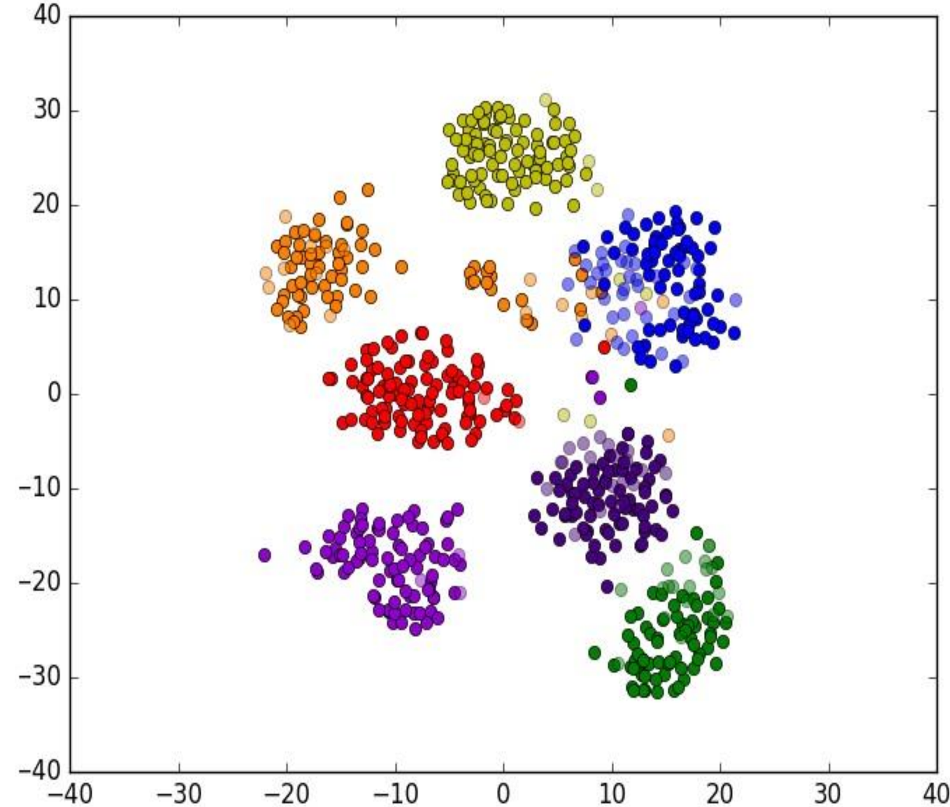


6. Some Experimental Results

- **Impacts of Deep Multi-Task Learning**



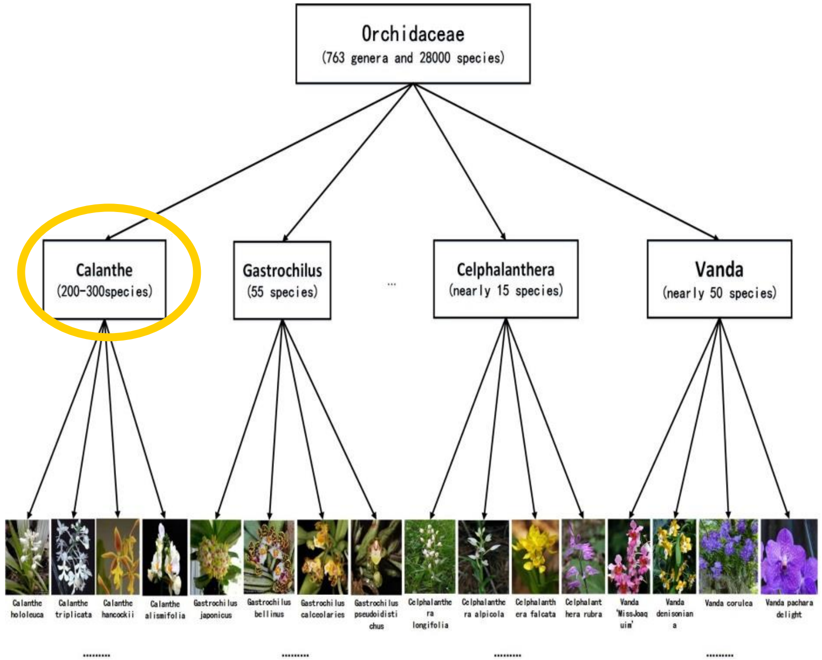
(a) N-way Softmax



(b) LMM

Concept Ontology *vs.* Visual Tree

- **Early Stop**



- **Semantic Interpretation**