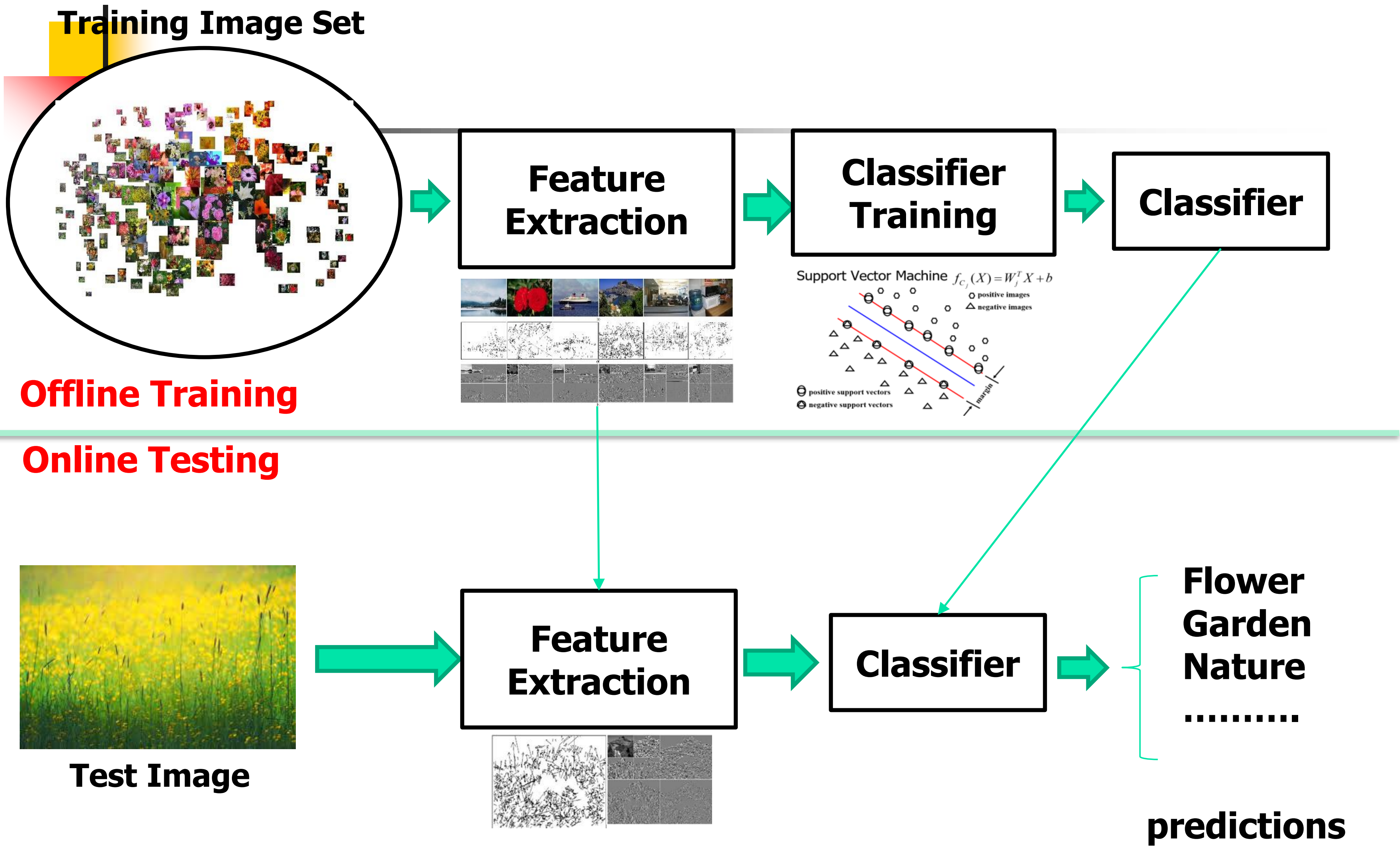
Hierarchical Image Classification over Concept Ontology

Jianping Fan
Dept of Computer Science
UNC-Charlotte

Course Website:
http://webpages.uncc.edu/jfan/itcs5152.html

Pipeline for Traditional Image Classification System



Feature Extraction

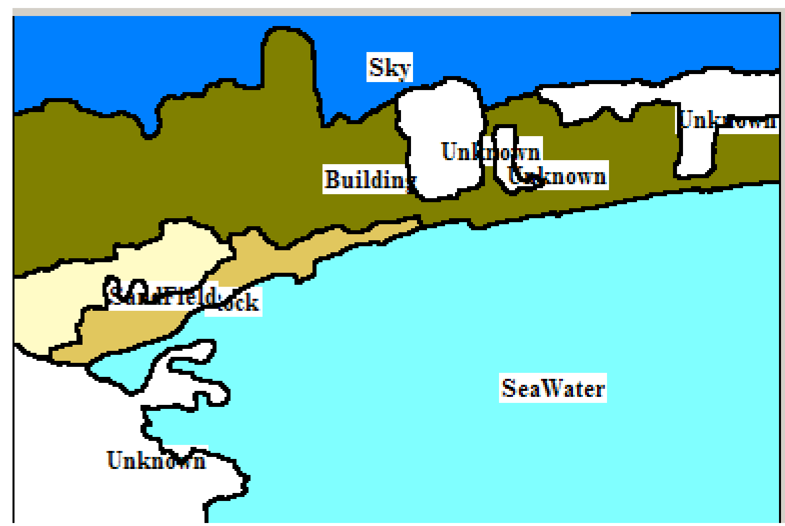
- Image-based approach
- Grid-based approach
- Object-based approach
- Bag-of-Visual-Word
- Deep Networks

Feature Extraction

- **Object-Based Approach**

Input Image

Salient Objects

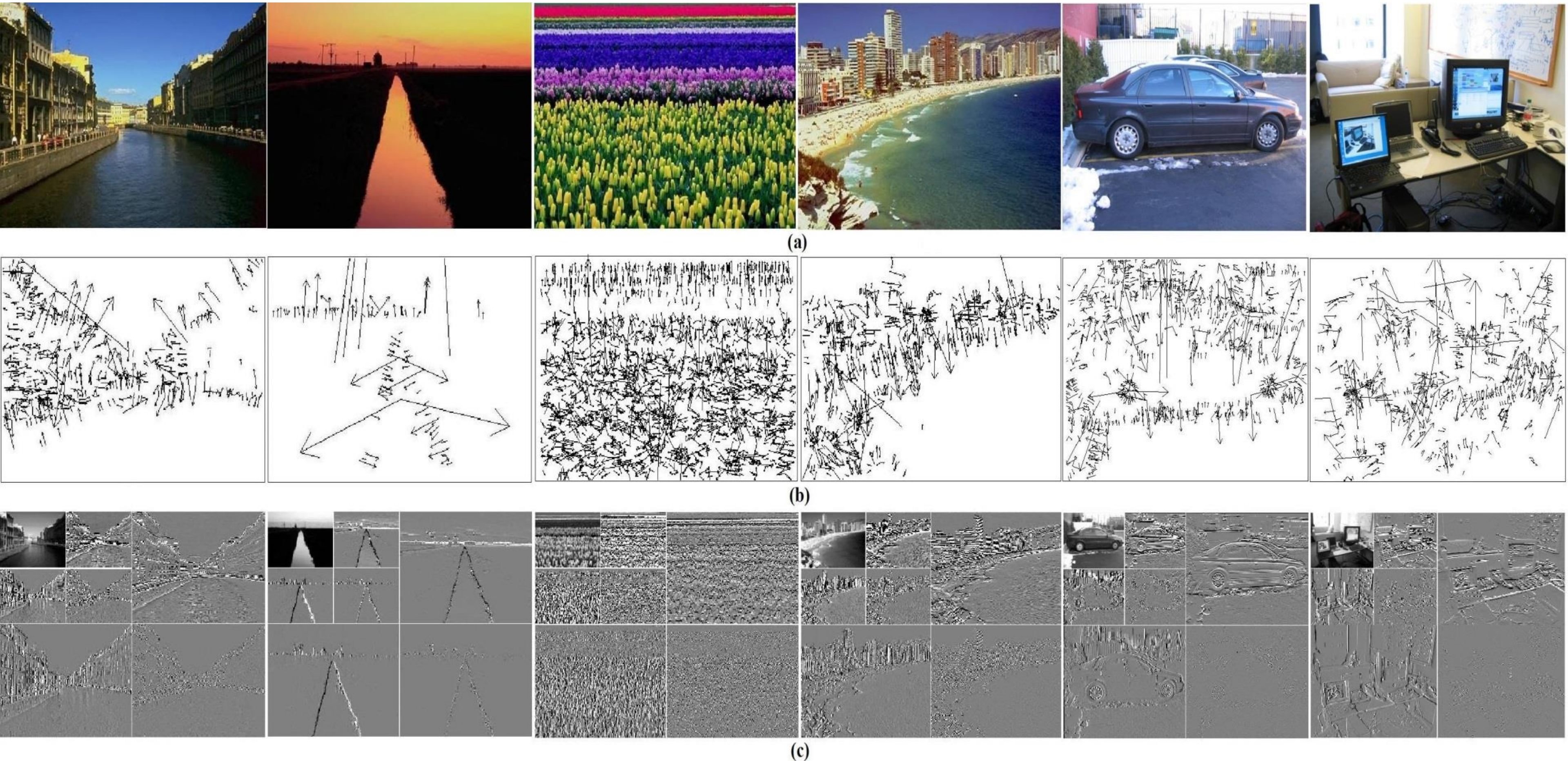


Color histogram, Tamura texture, Locations …….

Visual Features

Feature Extraction

- **Image-Based Approach**



Feature Extraction

- **Grid-Based Approach**

Feature Extraction

- **Bag-of-Visual-Word Approach**

Examples for visual words

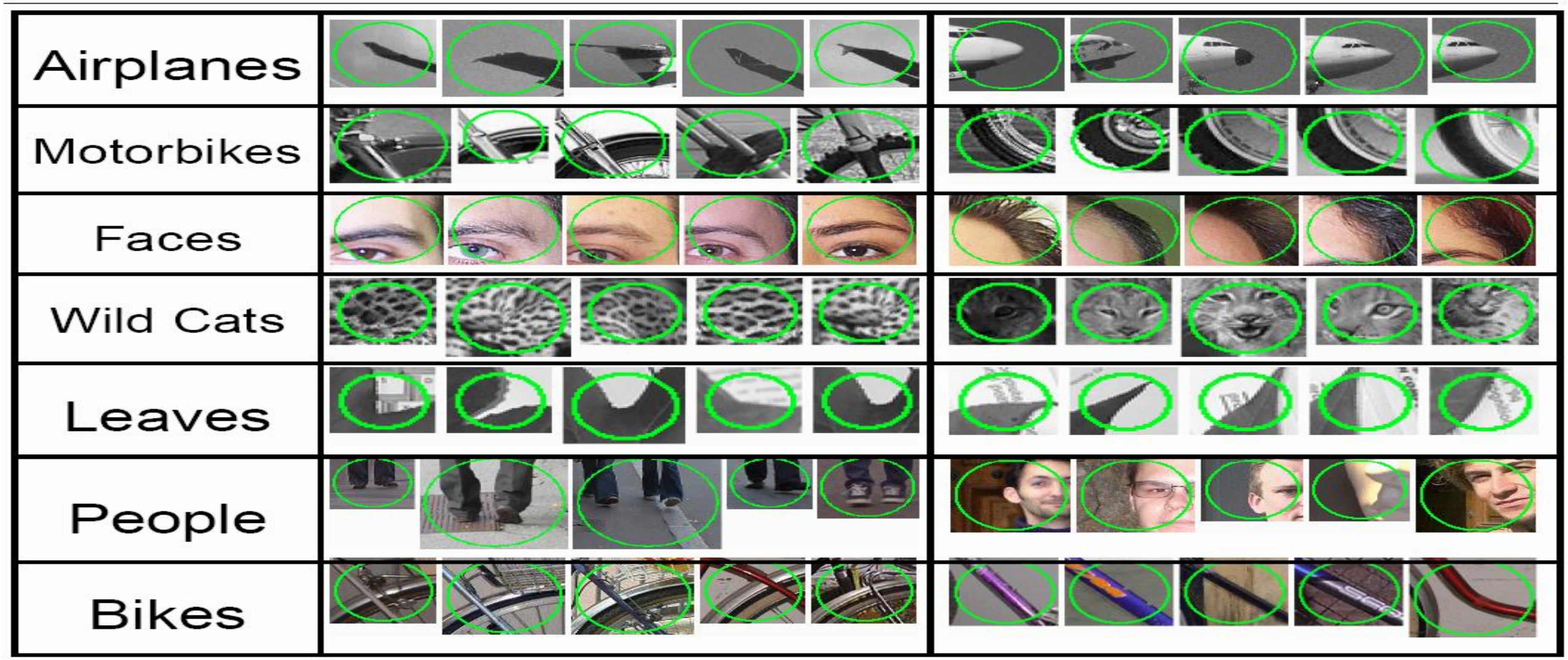


Image Representation

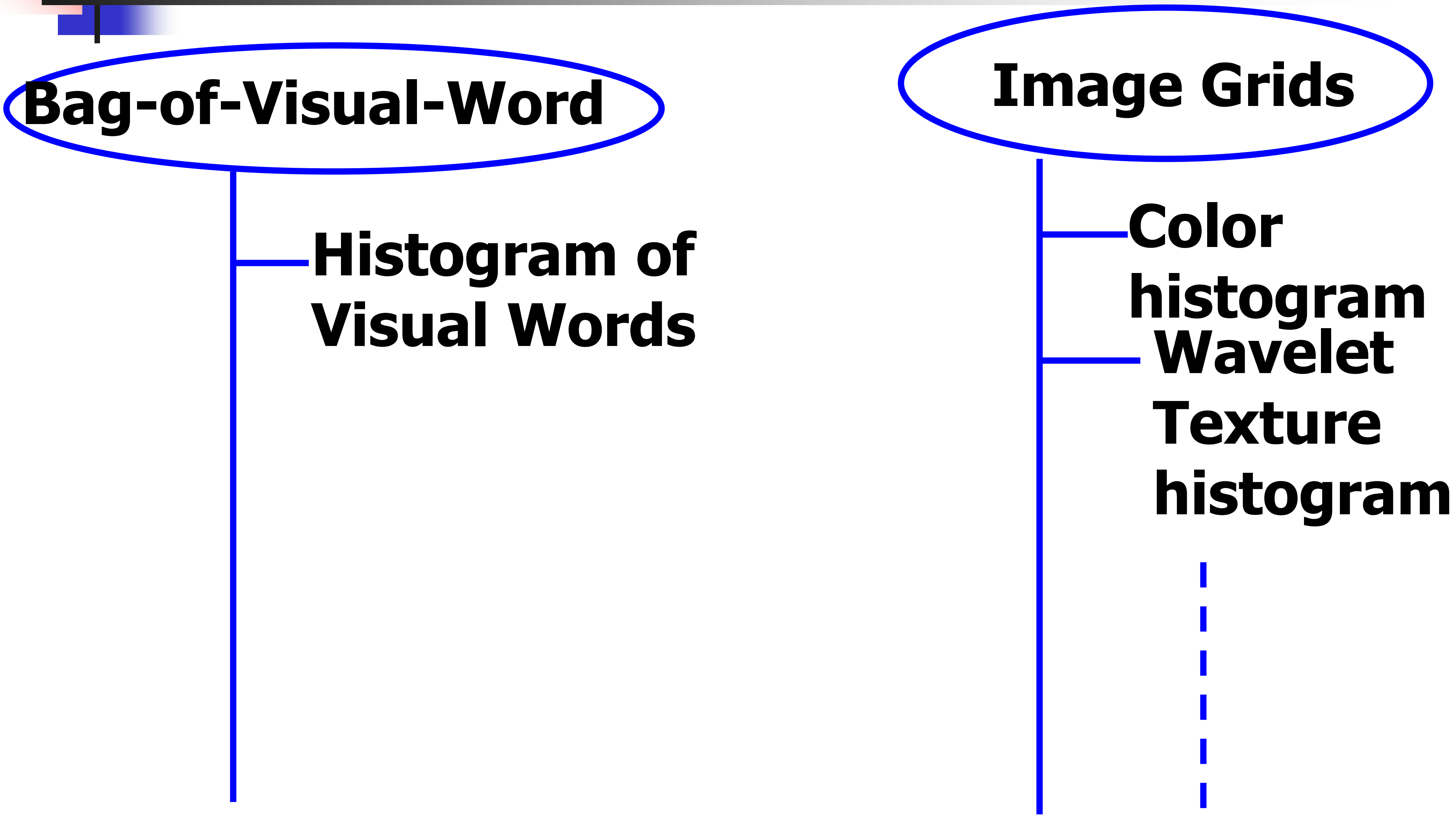
Salient Objects

- **Color histogram**
- **Tamura Texture**
- **Shape**
- ⋮

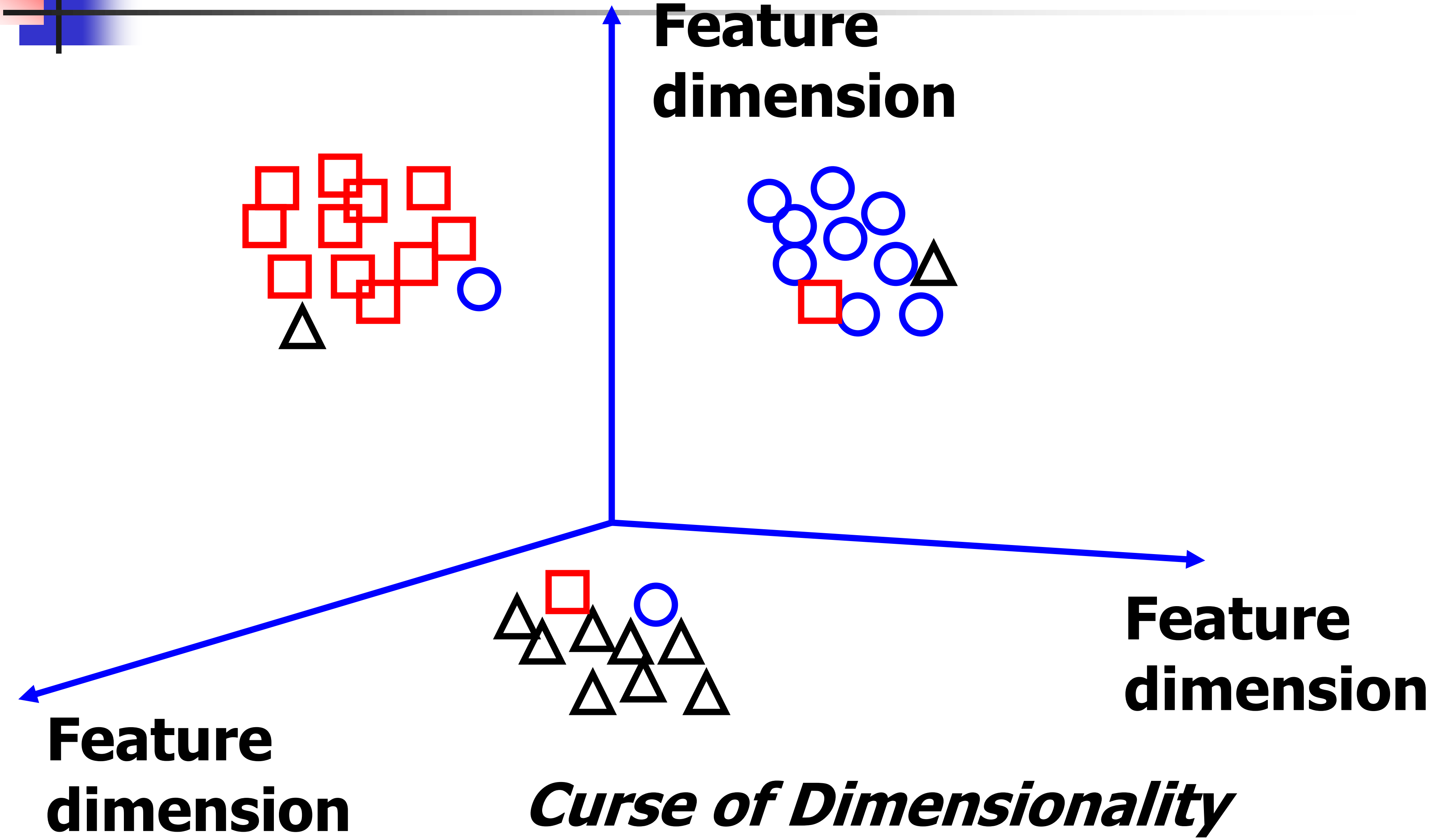
Images

- **Color histogram**
- **Wavelet Texture histogram**
- ⋮

Image Representation



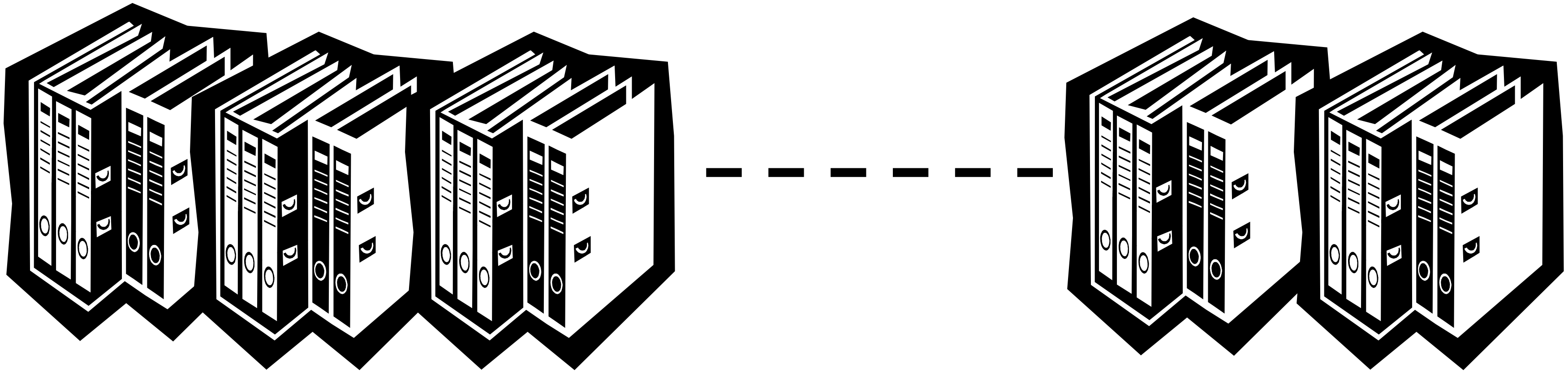
Feature-Based Image Representation



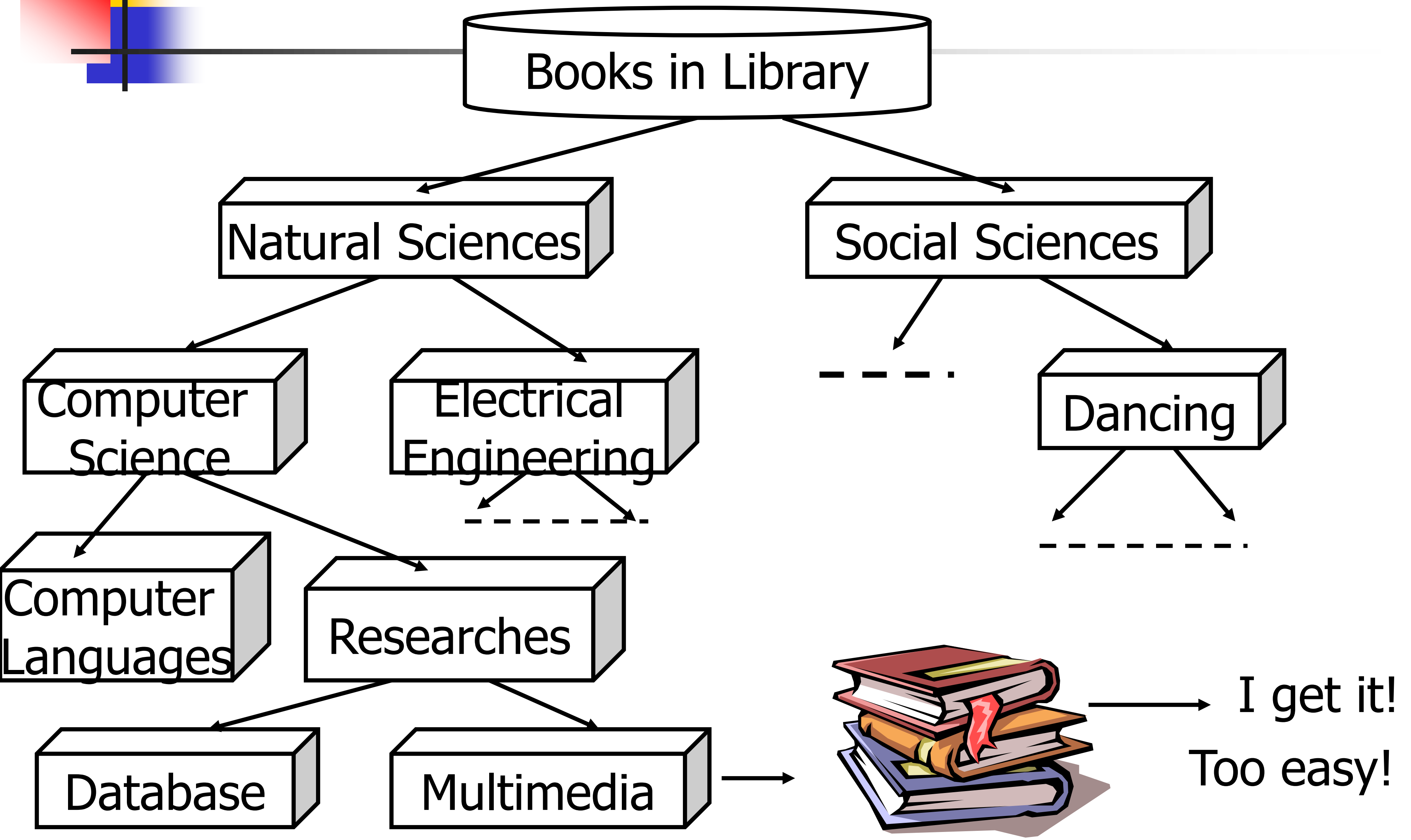
Curse of Dimensionality

Organization of Large-Scale Categories

Library: 12,000,000 books



Organization of Large-Scale Categories



Organization of Large-Scale Categories

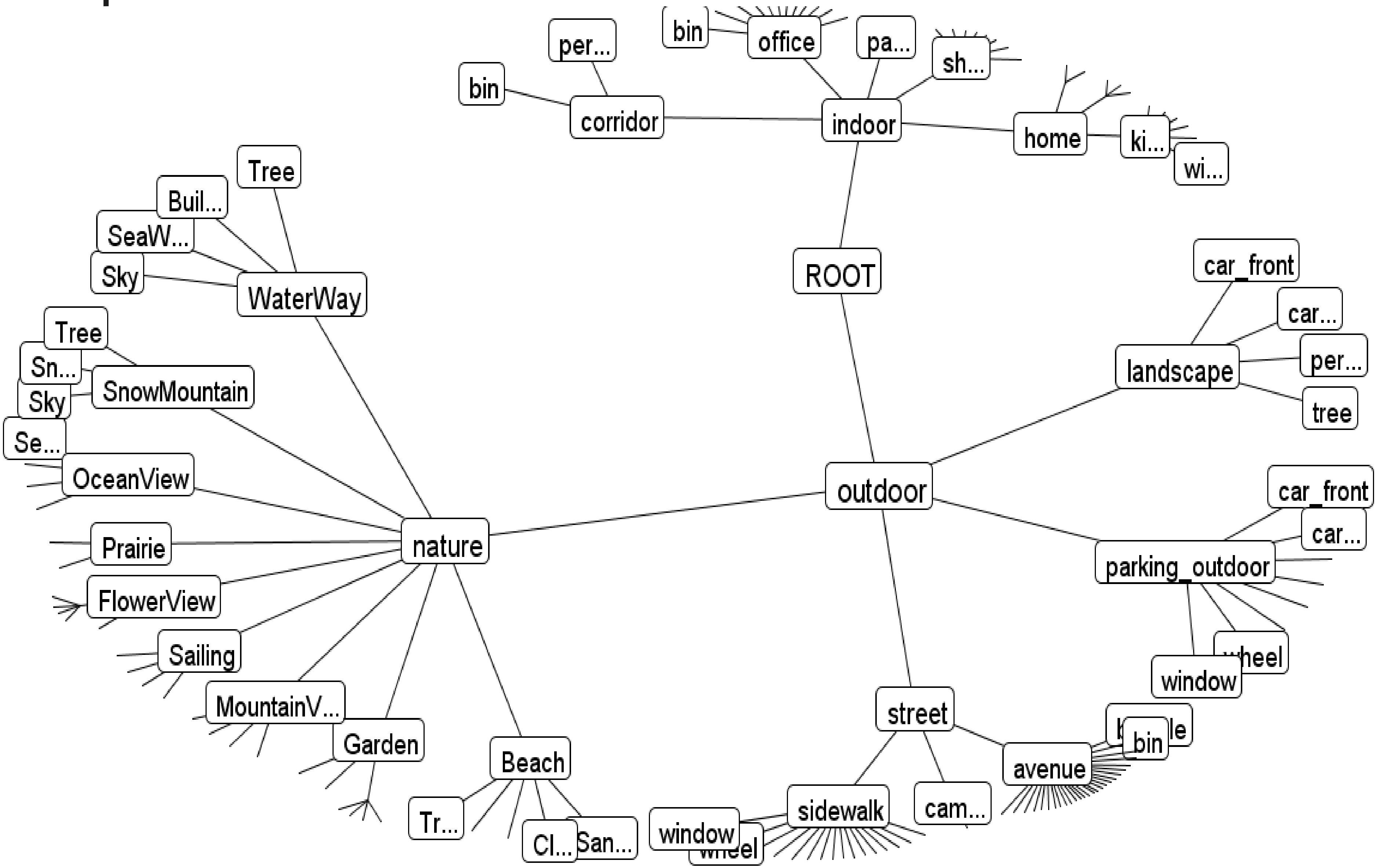
- **What is the solution?**

Let's go back to principle!

Concept Hierarchy or Ontology

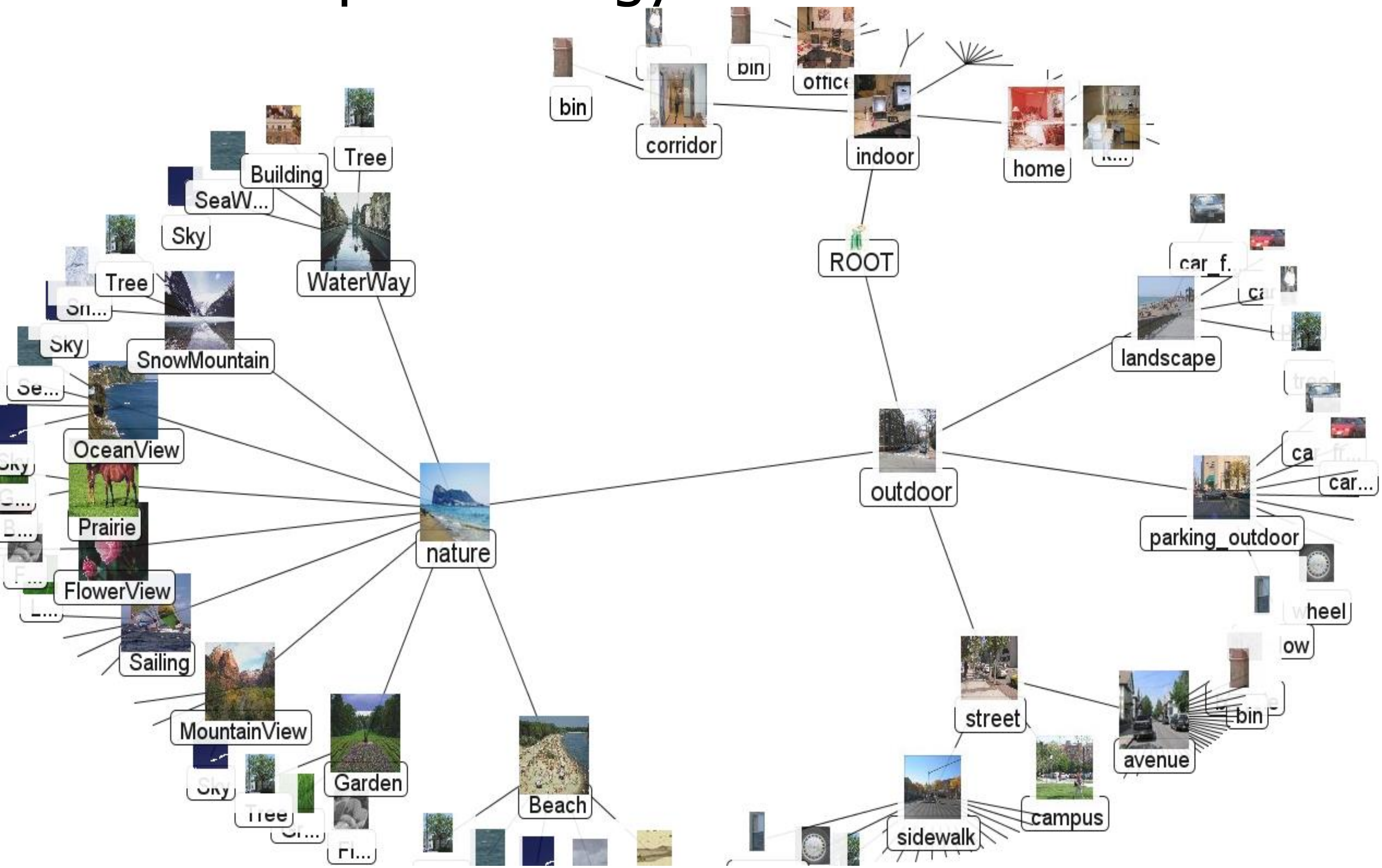
Hierarchical Concept Organization

- Concept Ontology

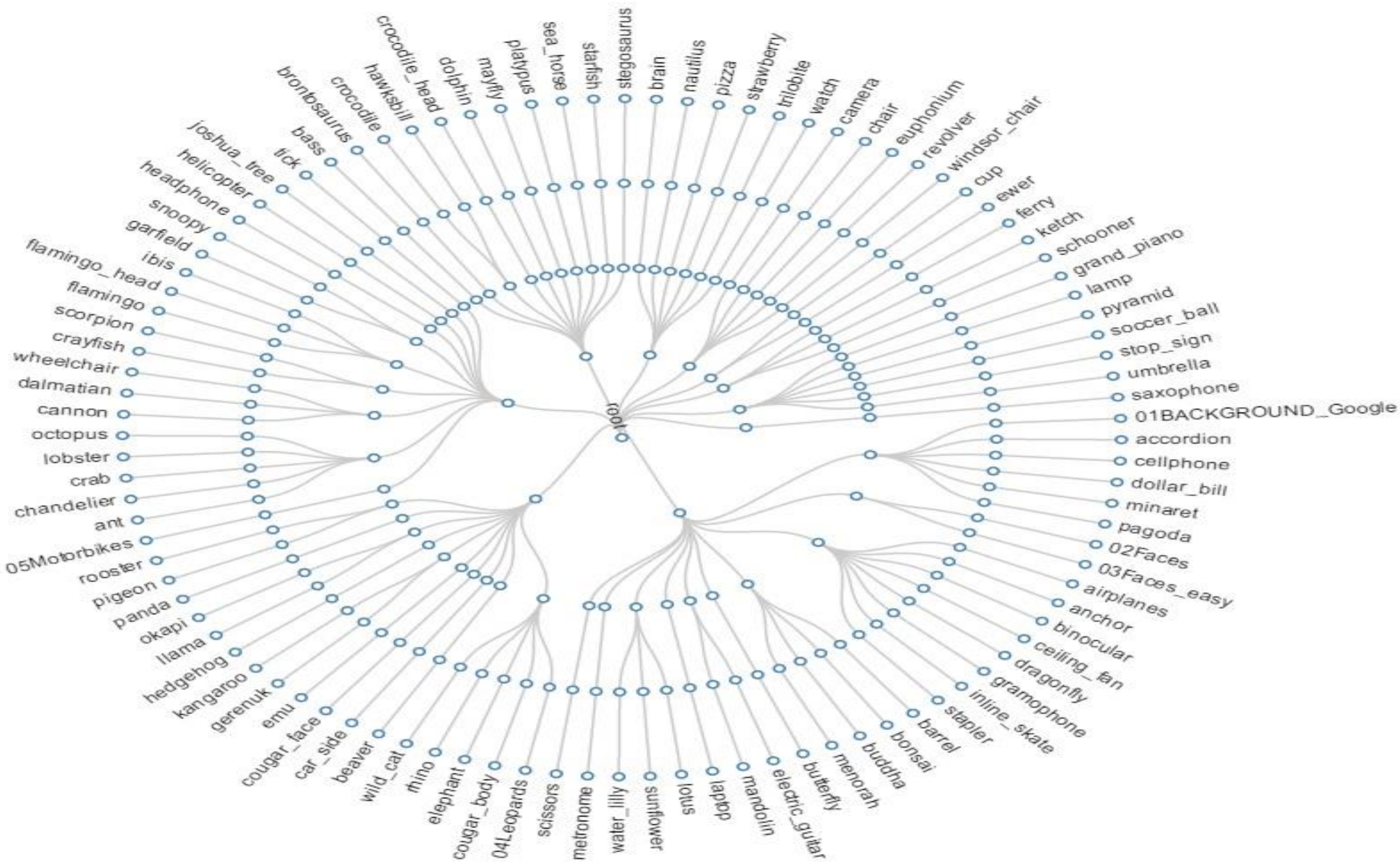


Hierarchical Concept Organization

- Concept Ontology

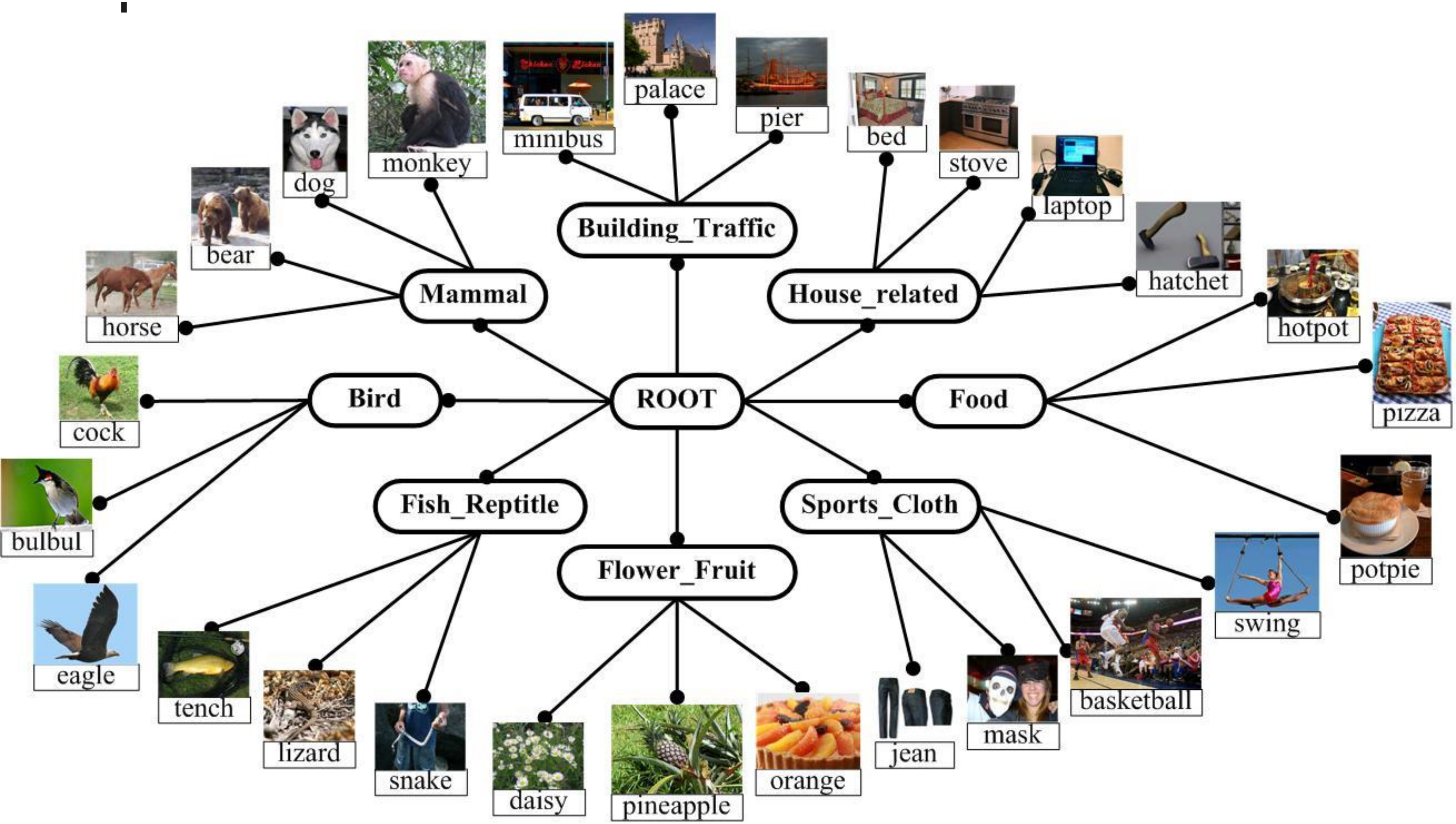


Hierarchical Concept Organization



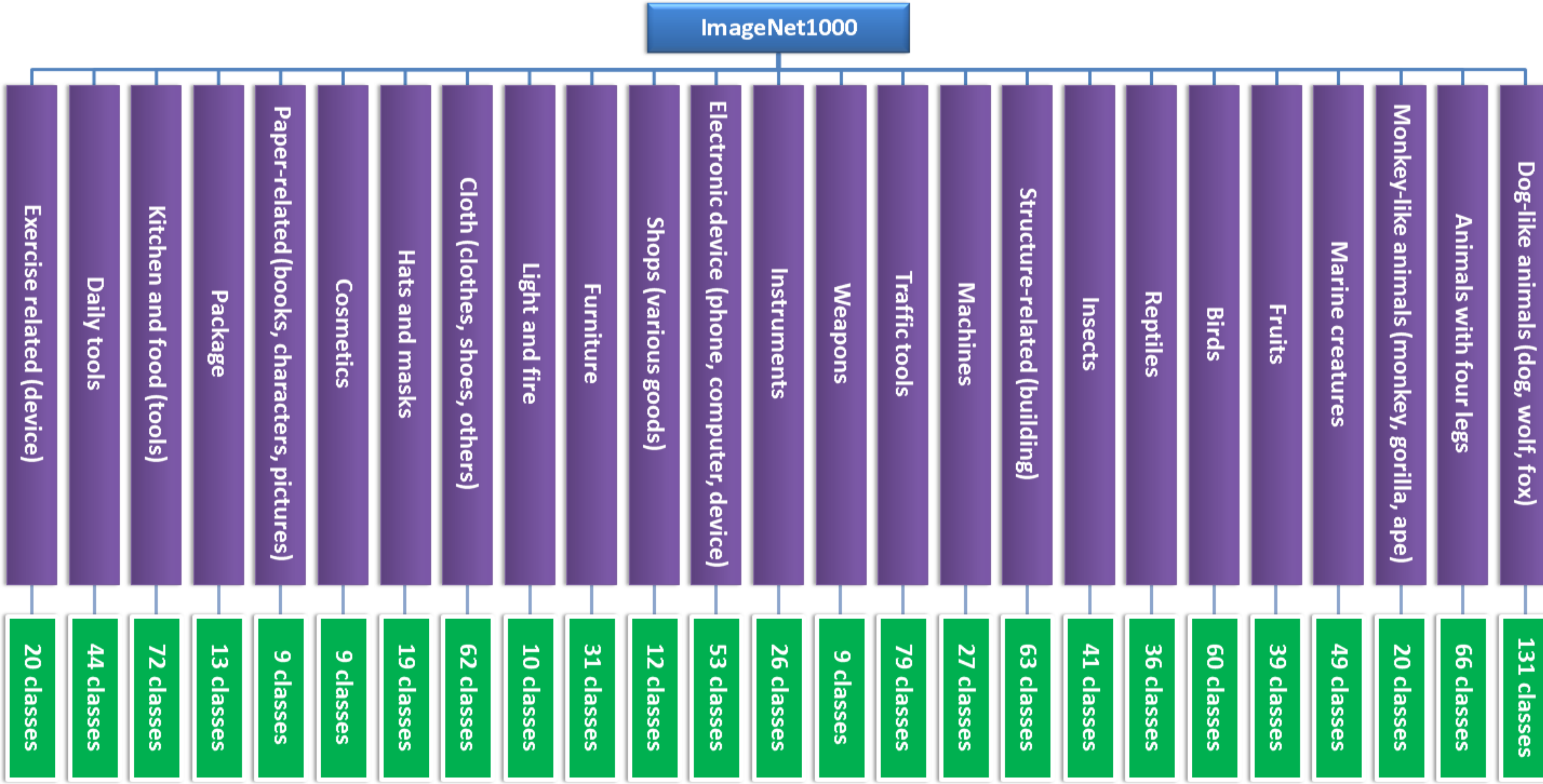
CalTech101

Hierarchical Concept Organization

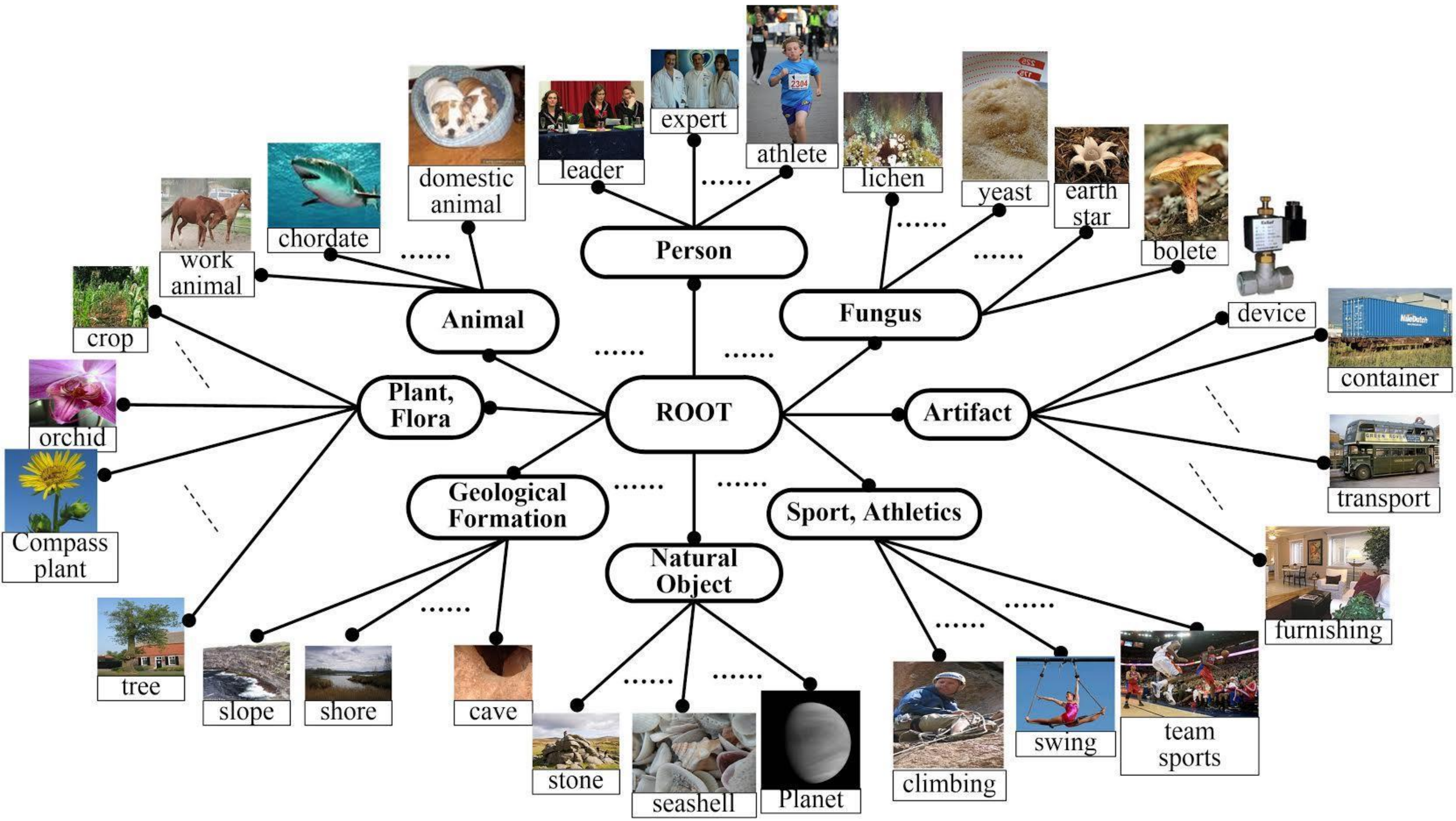


Two-Layer Ontology for ImageNet1K

Hierarchical Concept Organization

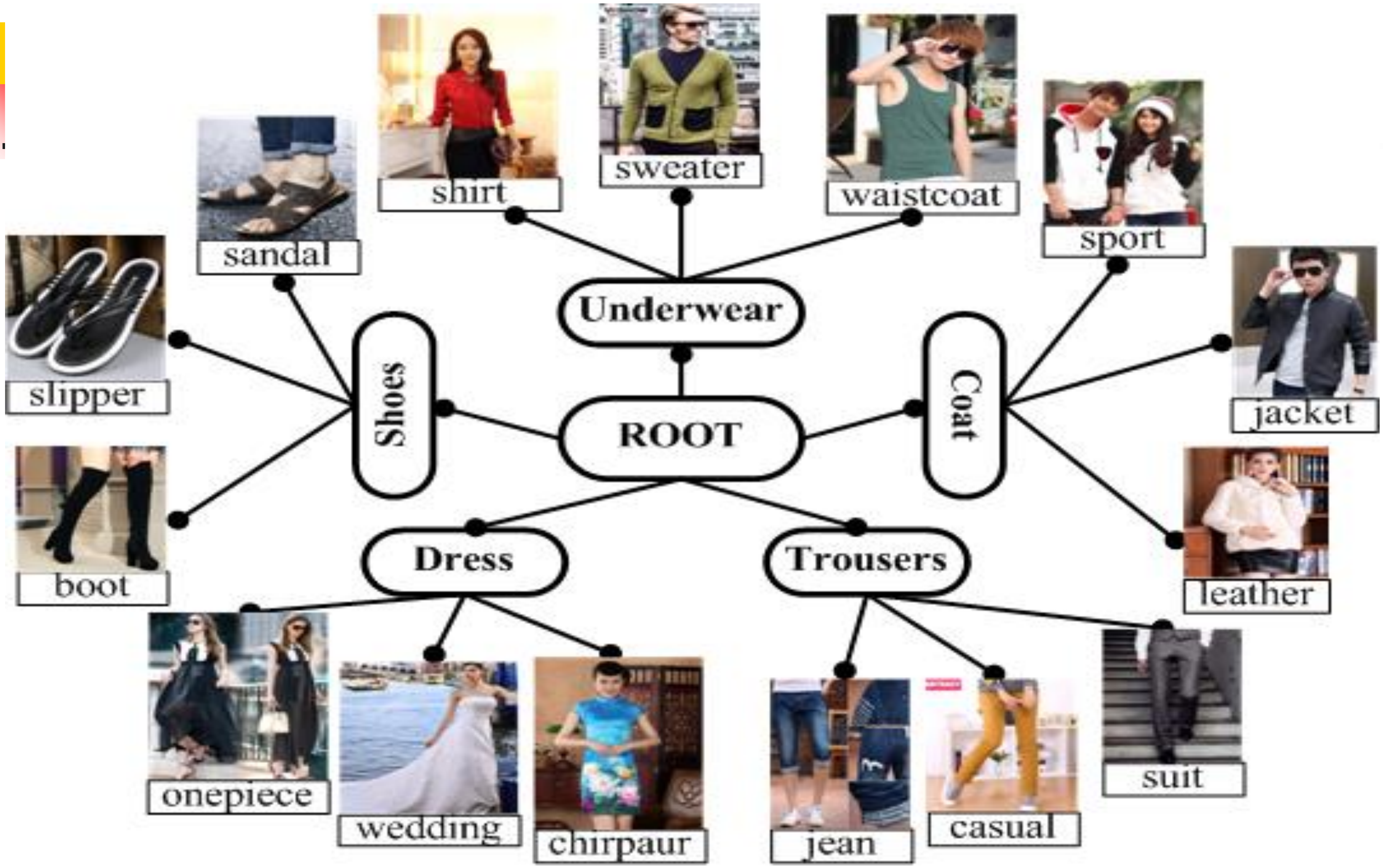


Hierarchical Concept Organization



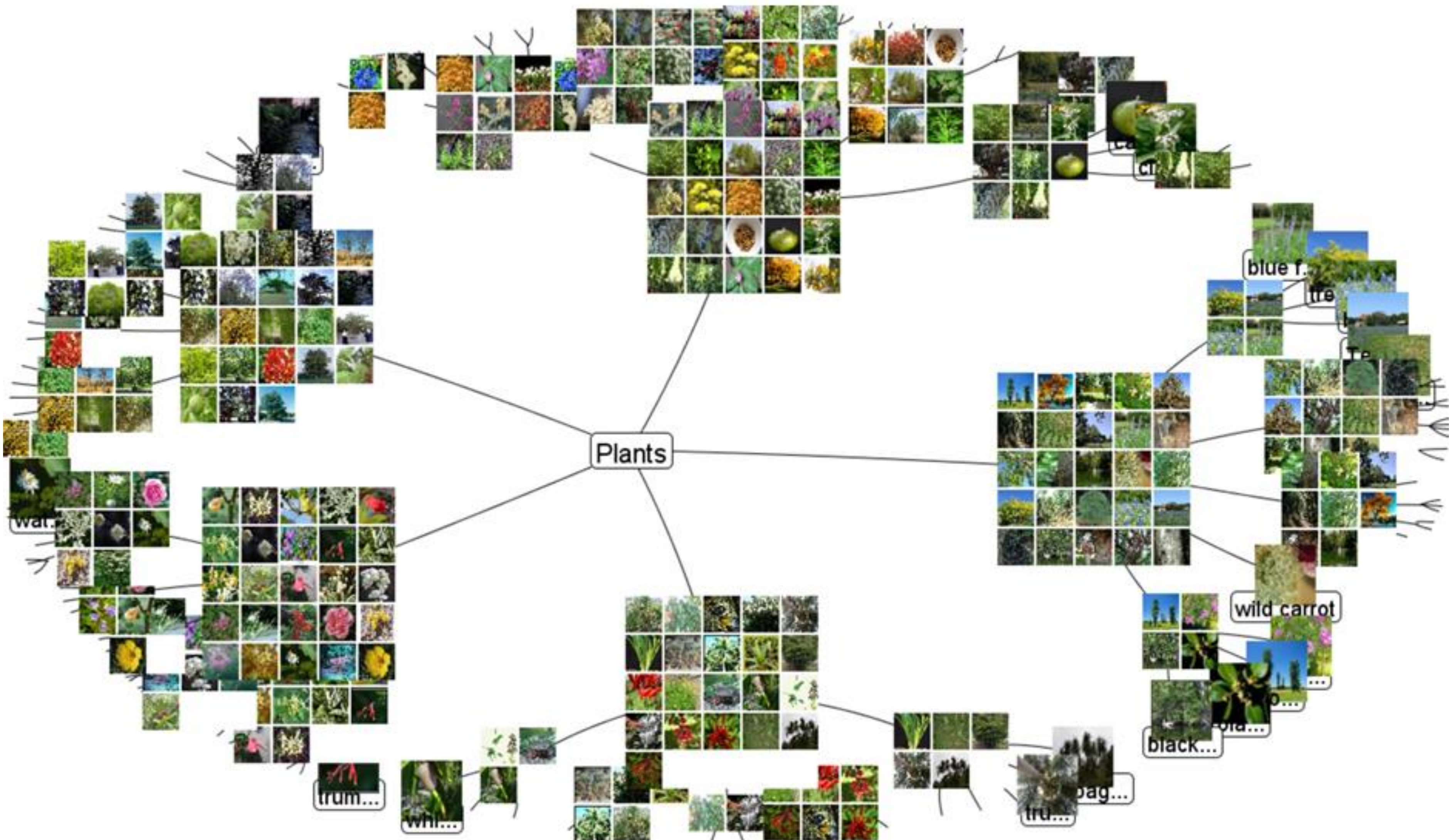
Two-Layer Ontology for ImageNet10K

Hierarchical Concept Organization



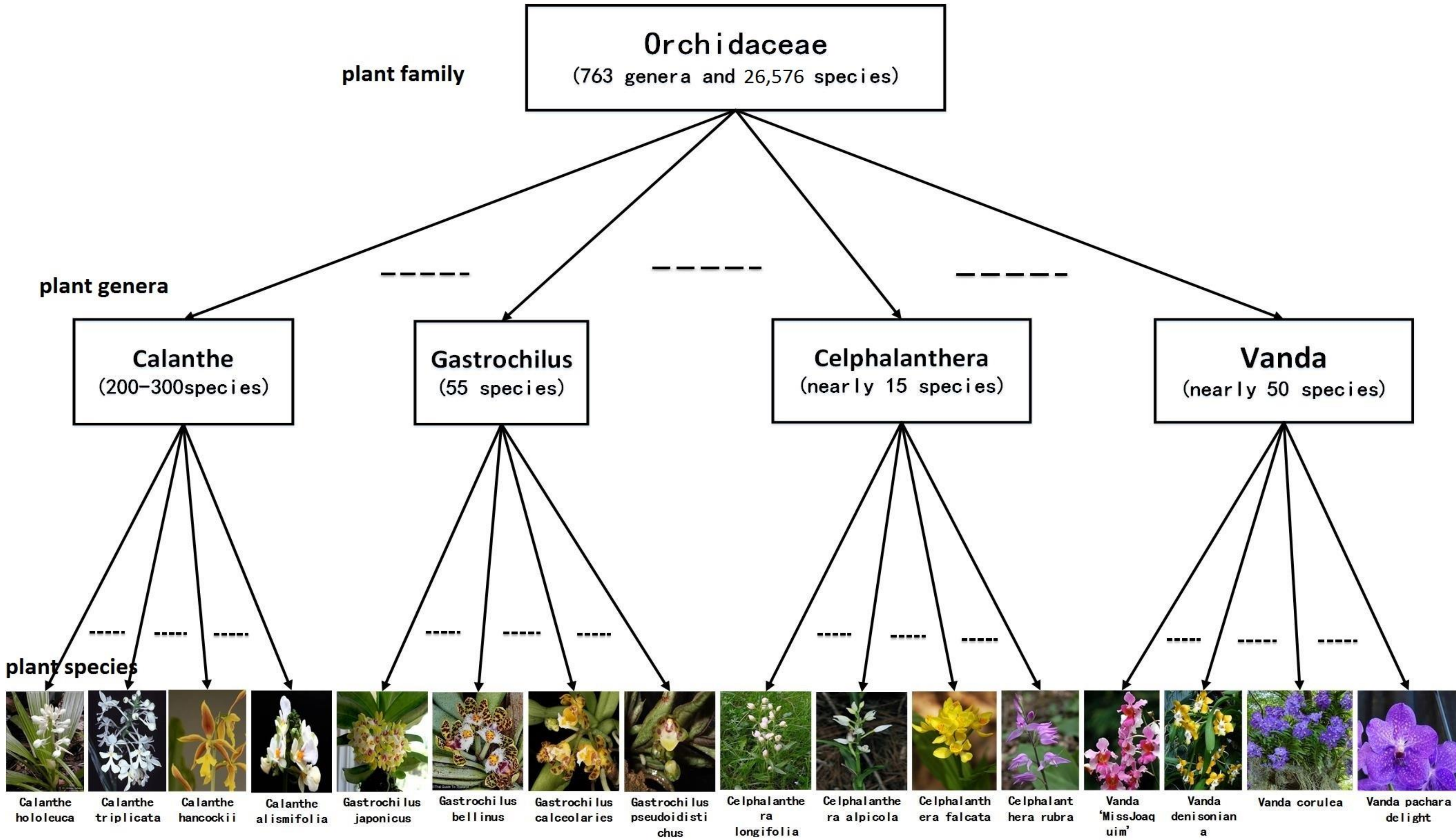
Two-Layer Ontology for Taobao Products

Hierarchical Concept Organization



Plant Ontology

Hierarchical Concept Organization



Two-Layer Ontology for Orchidaceae

Hierarchical Concept Organization

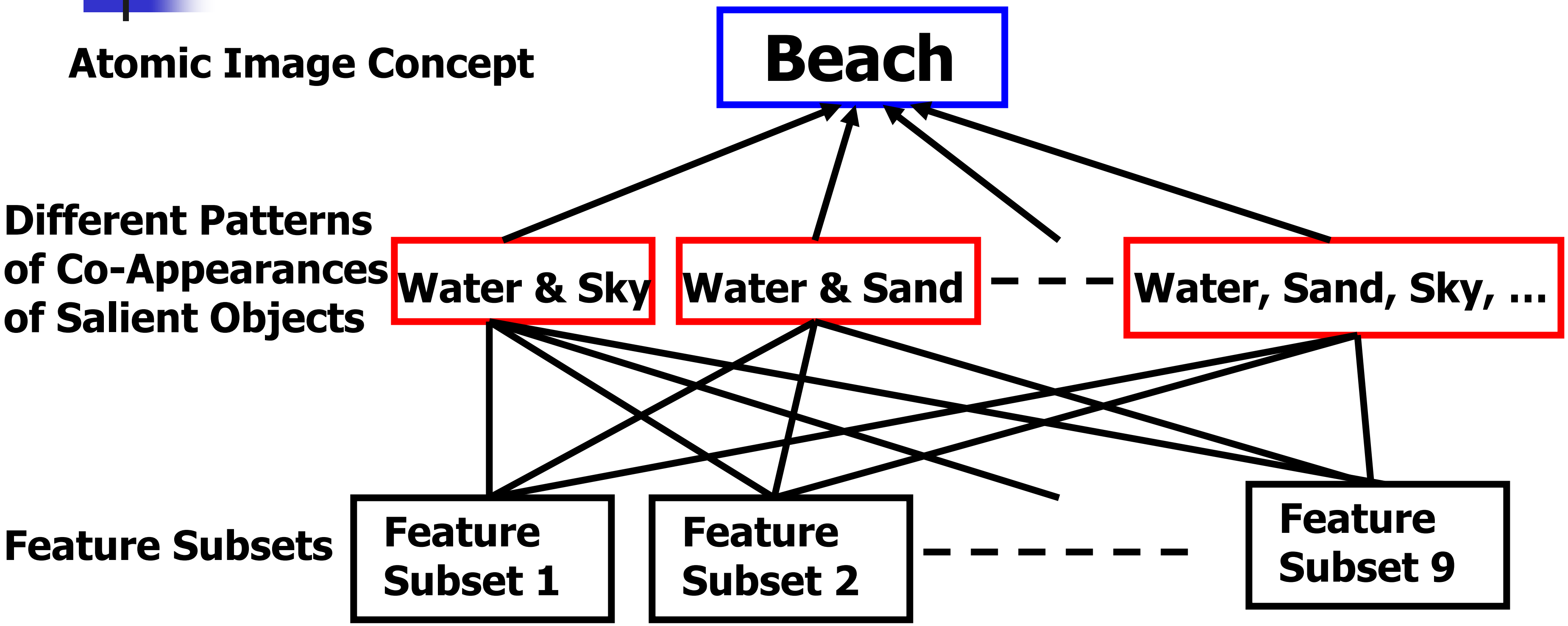
- **Multi-Modal Features**
- **Concept Ontology**

1. How to integrate multiple multi-modal features for classifier training?

2. How to support hierarchical classifier training over concept ontology?

Multi-Level Image Annotation:
Annotating Images at Different Semantic Levels

Classifier Training for Atomic Image Concepts

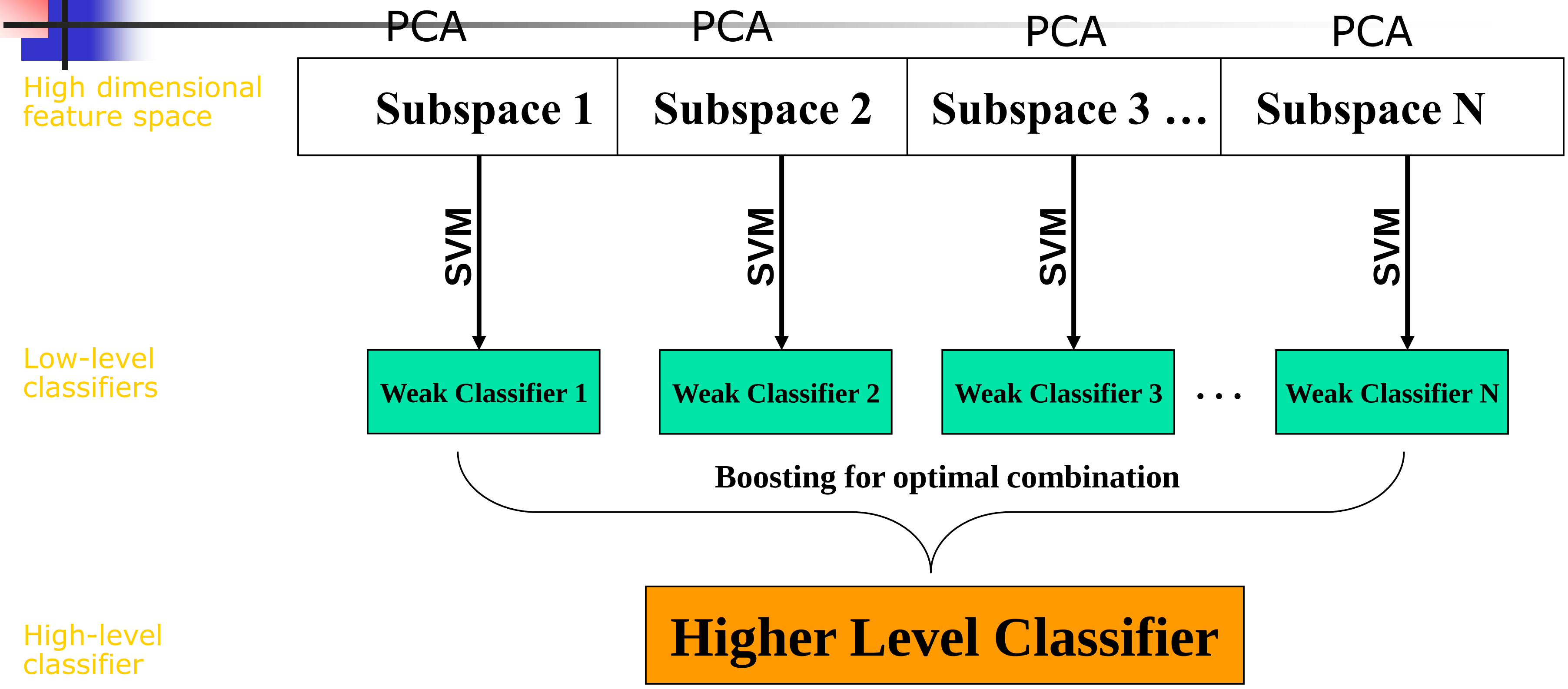


$$M = \sum_{i=2}^{n} C_n^i = 2^n - n - 1$$

Classifier Training for Atomic Image Concepts

- Curse of Dimensionality
  - # samples needed increase with # dimensions (generally exponentially) .
  - Human labeling is expensive
  - Some features are redundant
- Proposal
  - **Joint SVM boosting and feature selection**

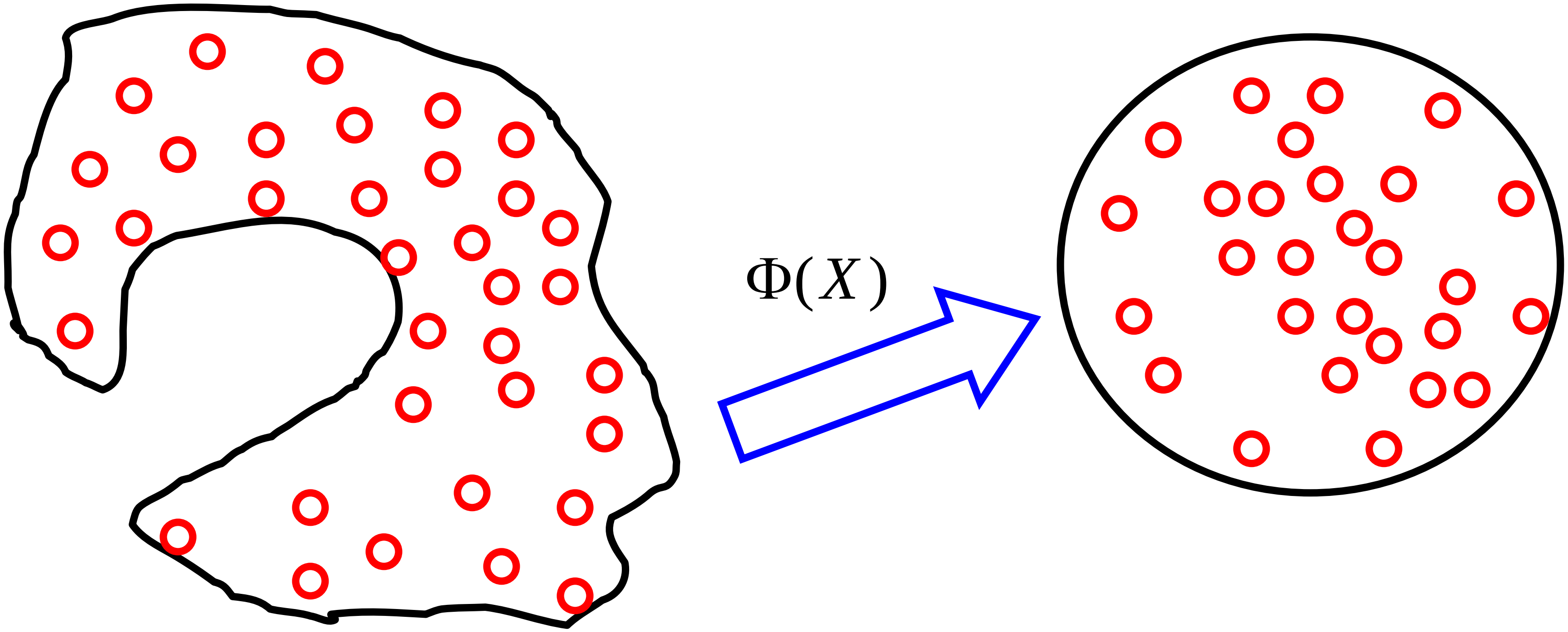
Boosting SVM Classifier Training



- Less training samples due to dimension reduction
- Reuse training results on low-level concepts
- More selection opportunities compared to filter and wrapper

Classifier Training for Atomic Image Concepts

- Kernel-Based Data Warping



Kernel Function: $\mathrm{K}(X_i, X_j) = \Phi(X_i)^T \Phi(X_j)$

Classifier Training for Atomic Image Concepts

- **Kernel for Color Histogram**
  - **Statistical Image Similarity**

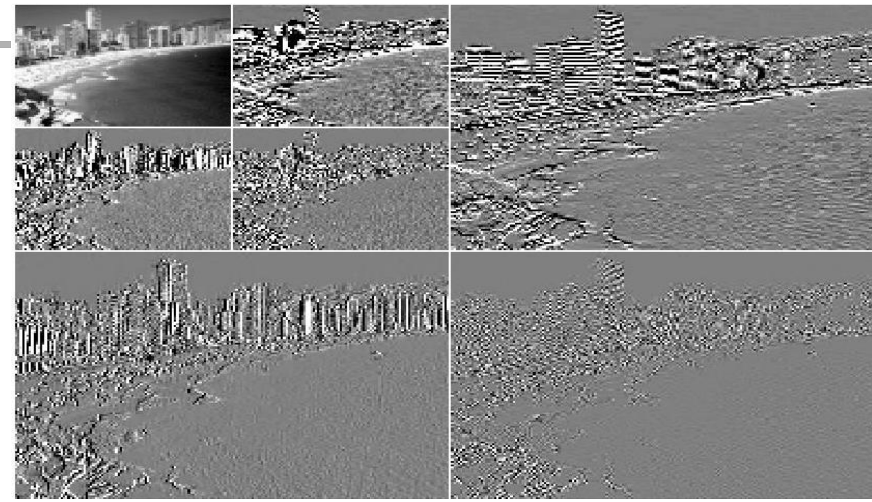
$$\chi^2(u,v) = \frac{1}{2}\sum_{i=1}^{N}\frac{(u_i - v_i)^2}{u_i + v_i}$$

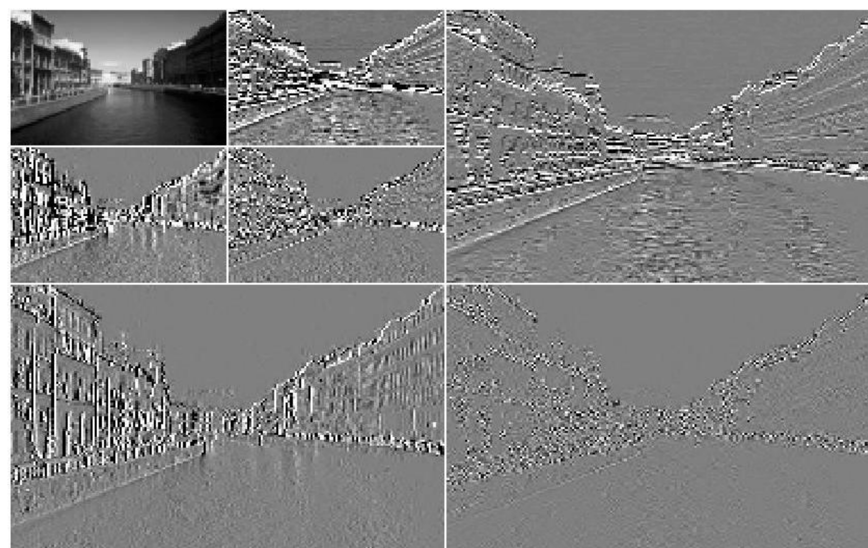
  - **Kernel**

$$\mathrm{K}(u,v) = e^{-\chi^2(u,v)/\sigma}$$

Classifier Training for Atomic Image Concepts

Wavelet Filter Bank Kernel





Classifier Training for Atomic Image Concepts

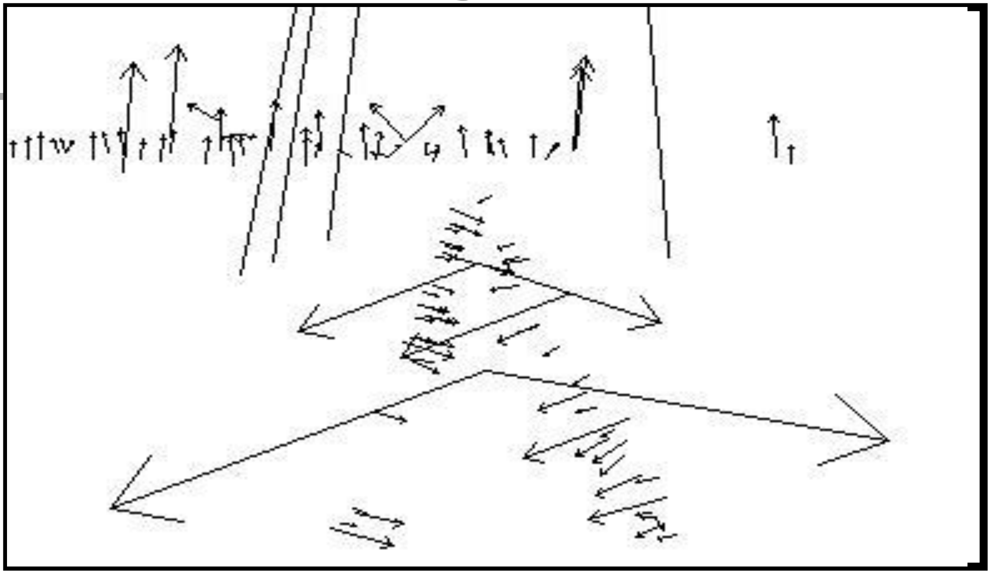
- **Wavelet Filter Bank Kernel**

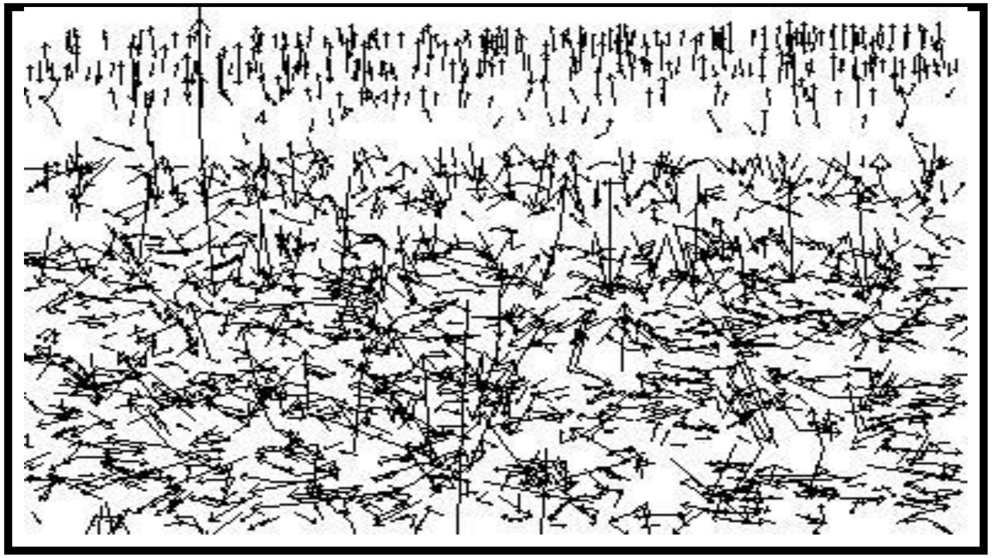
$$\mathrm{K}(x,y) = e^{-\sum_{i=1}^{n} \chi_i^2 (h_i(x), h_j(y))/\sigma_i}$$

$$= \prod_{i=1}^{n} e^{-\chi_i^2 (h_i(x), h_i(y))/\sigma_i}$$

Classifier Training for Atomic Image Concepts

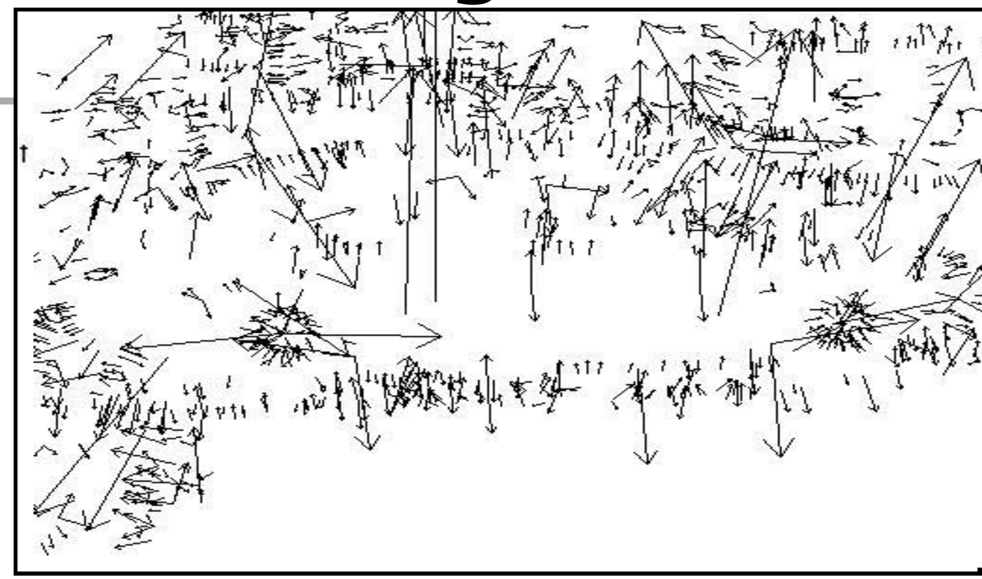
- **Interest Point Matching Kernel**

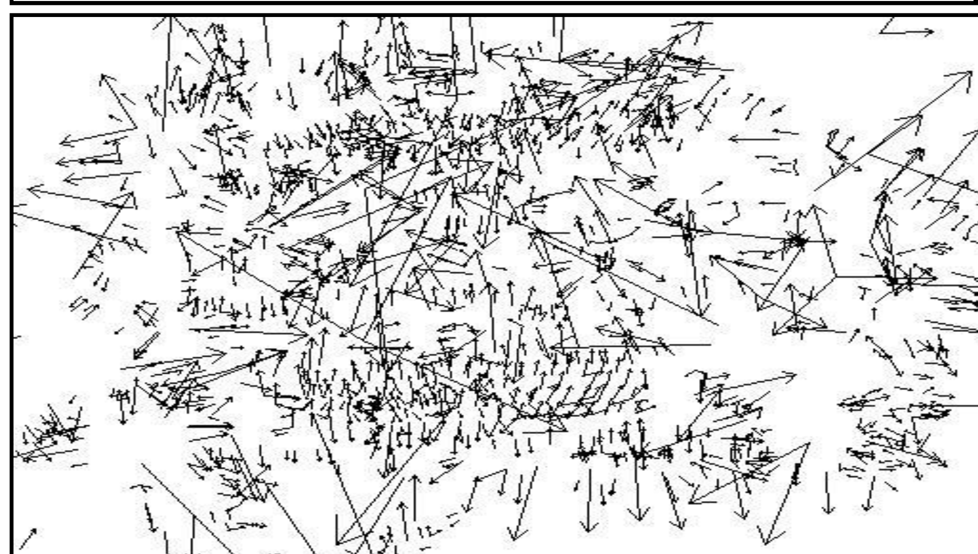





Classifier Training for Atomic Image Concepts

- **Interest Point Matching Kernel**





Classifier Training for Atomic Image Concepts

- **Multiple Kernel Learning**

$$\overset{\wedge}{\mathrm{K}}(x,y)=\sum_{i=1}^{\kappa}\alpha_i \mathrm{K}_i(x,y) \qquad \sum_{i=1}^{\kappa}\alpha_i=1$$

- **SVM Image Classifier**

$$f(X)=\sin g\left(\sum_{l=1}^{M}\beta_l Y_l \overset{\wedge}{\mathrm{K}}(X_l,X)+b\right)$$

Classifier Training for Atomic Image Concepts

- **Dual Problem**

$$\min \frac{1}{2}\left(\sum_{h=1}^{\kappa}\|w_h\|_2\right)^2 + c\sum_{l=1}^{M}\xi_l$$

Subject to:

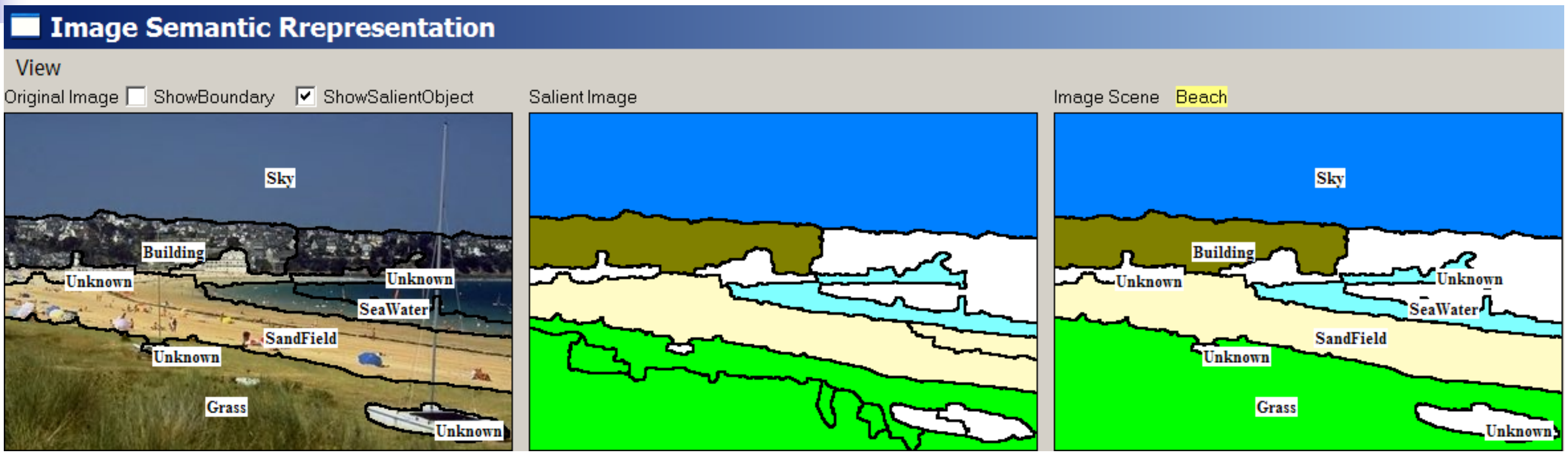
$$\forall_{l=1}^{M} : \xi_l \geq 0, Y_i\left(\sum_{h=1}^{\kappa}\left\langle w_k, \Phi(X_l)\right\rangle + b\right) \geq 1 - \xi_l$$

Classifier Training for Atomic Image Concepts

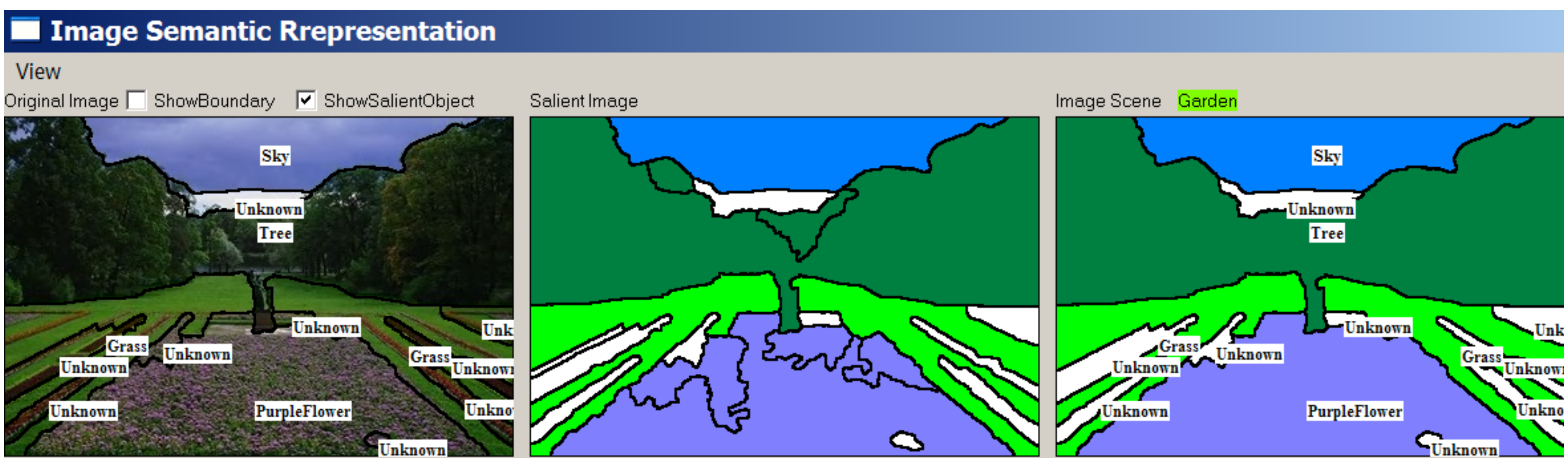
Classifier Training for Atomic Image Concepts

- **Some Results**

Beach Scene

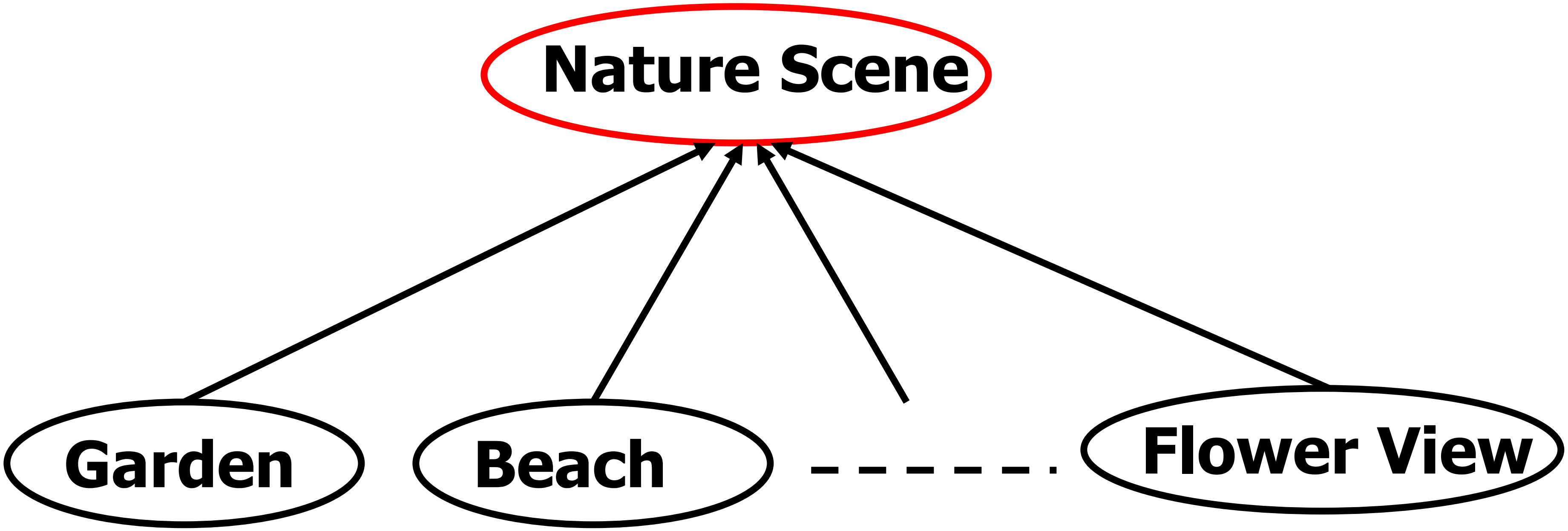


Garden Scene



High-Level Image Concept Modeling

- Inter-Concept Similarity Modeling



Nature Scene: Larger Hypothesis Space & Large Variations of Visual Properties!

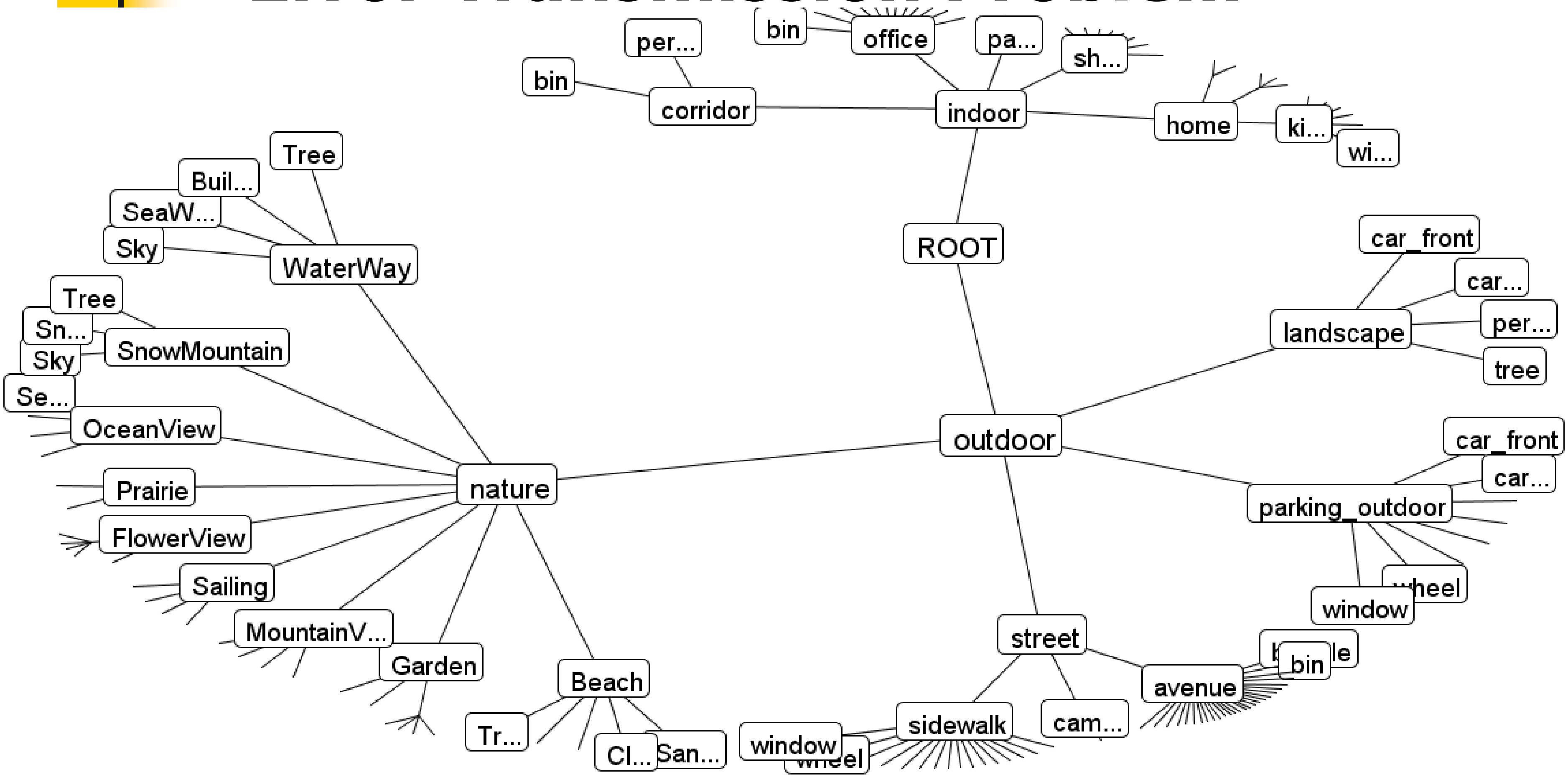
Garden, Beach, Flower view: Different but share common visual properties!

Classifier Training for High-Level Image Concepts

- **Challenging Problems**
  - Error Transmission Problems
  - Training Cost Issue
  - Knowledge Transferability and Task Relatedness Exploitation

Classifier Training for High-Level Image Concepts

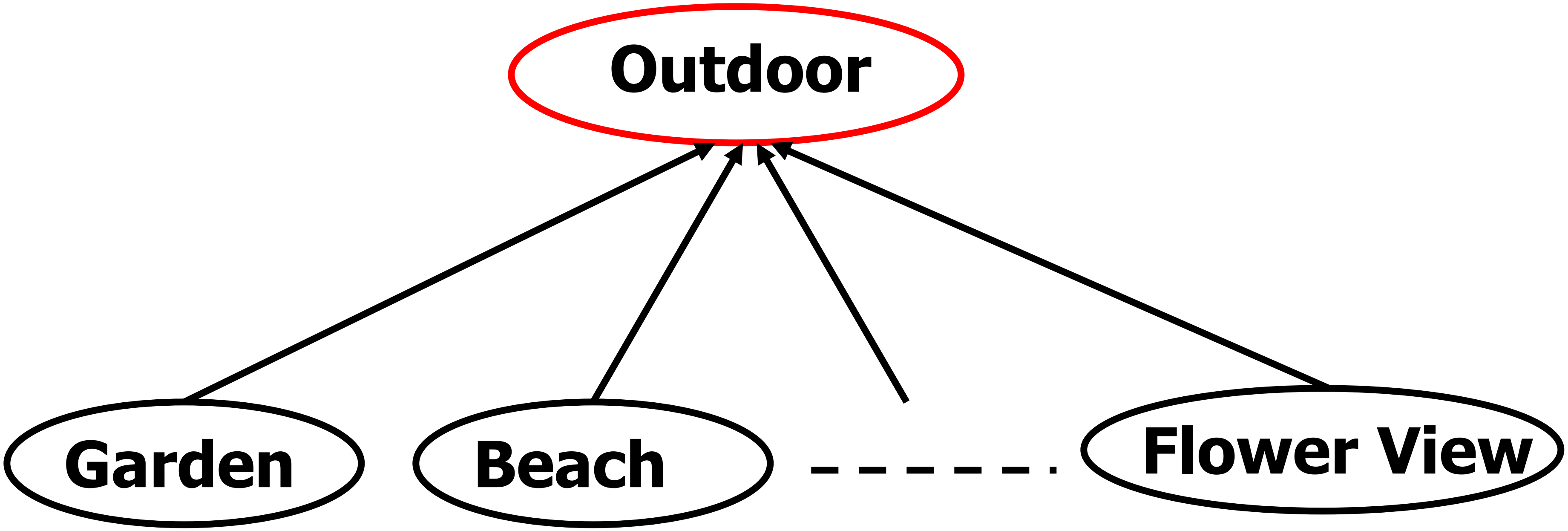
Error Transmission Problem



The classifiers for low-level image concepts cannot recover the errors for the classifiers of high-level image concepts!

Classifier Training for High-Level Image Concepts

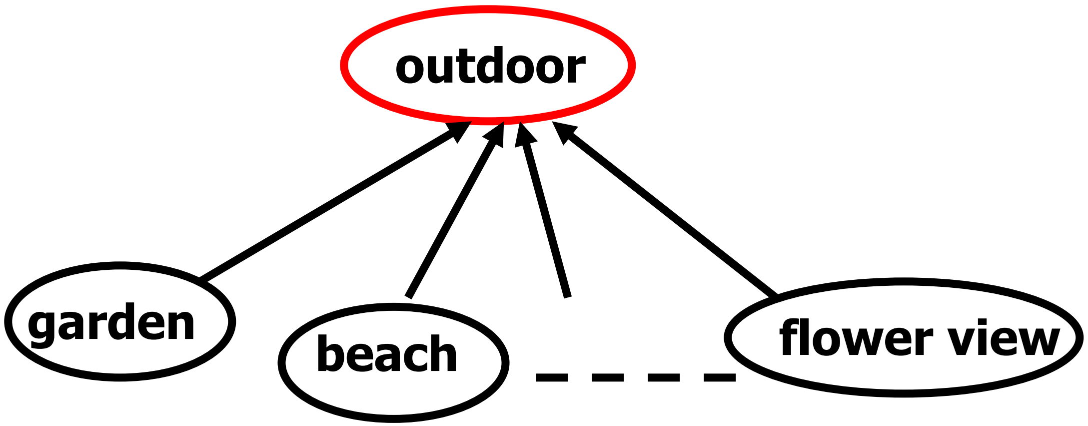
- **Error Transmission Problem**



Errors for the classifiers of atomic image concepts may be transmitted to the classifiers for the high-level image concepts!

Classifier Training for High-Level Image Concepts

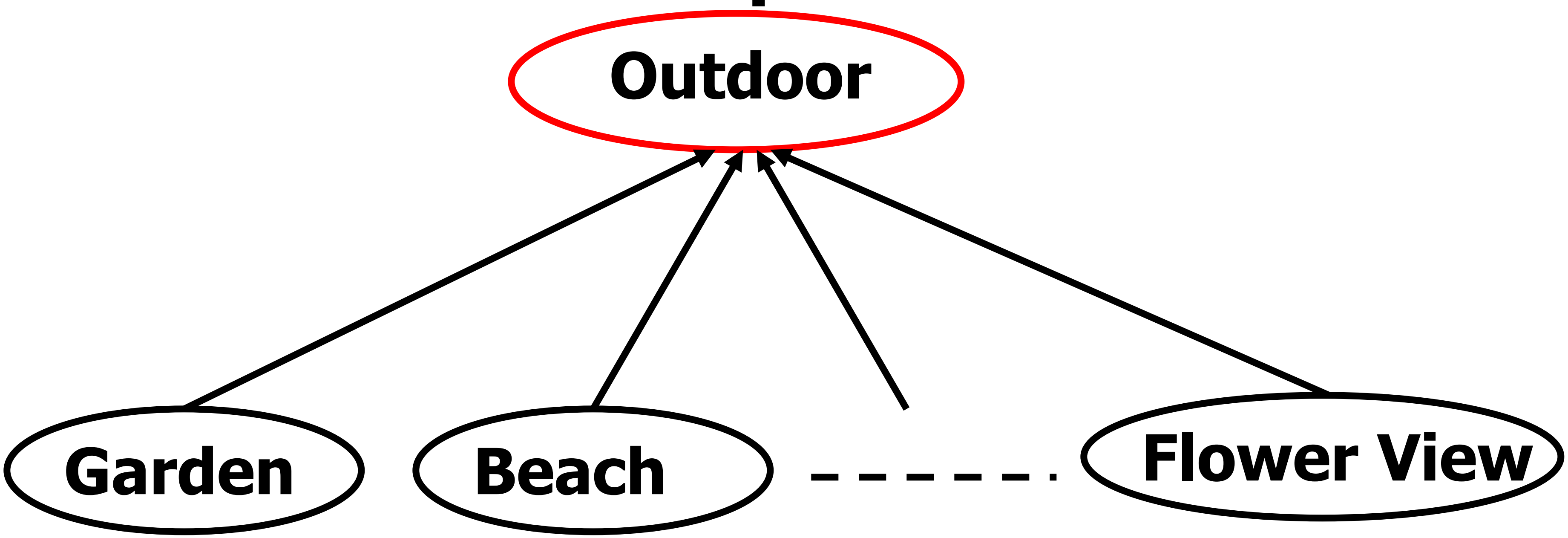
- **Training Cost Issue**
  - Multiple Hypotheses



  - Large Diversity of Contents

Classifier Training for High-Level Image Concepts

- **Knowledge Transferability & Task Relatedness Exploitation**



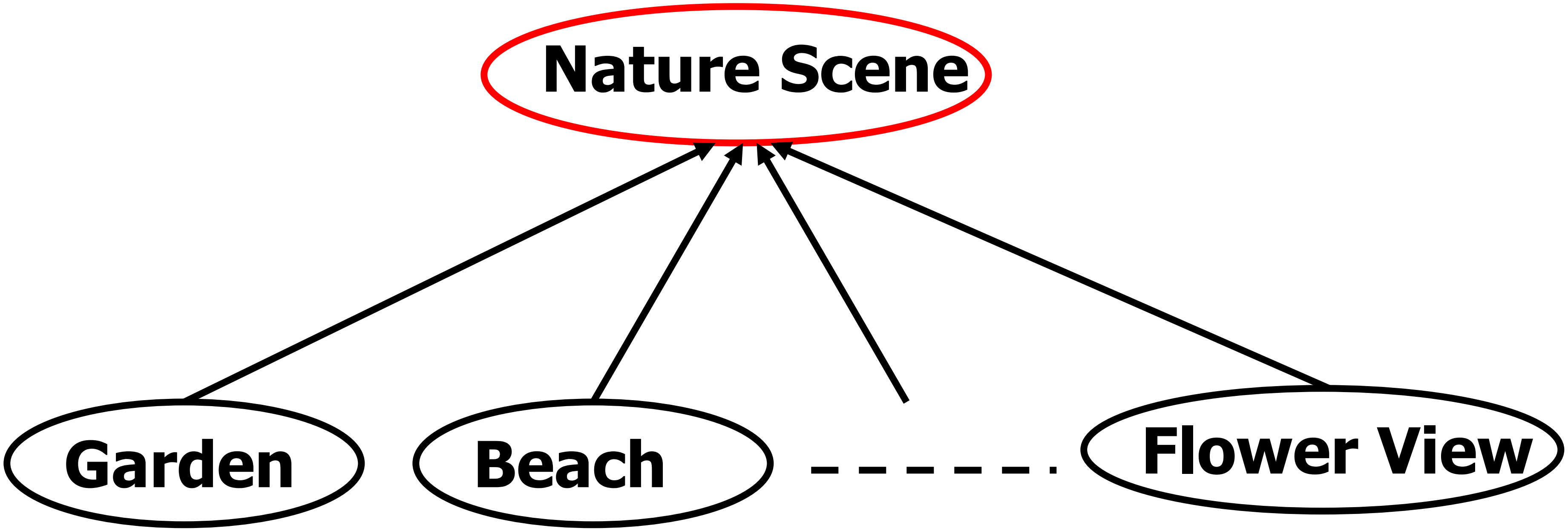
They are different but strongly related!

Classifier Training for High-Level Image Concepts

- Multi-Task Learning
  - **Which tasks are strongly related?**
  - **How to quantify the task relatedness?**
  - **How to integrate such task relatedness for training large-scale related image classifiers?**

Classifier Training for High-Level Image Concepts

- Related Learning Tasks

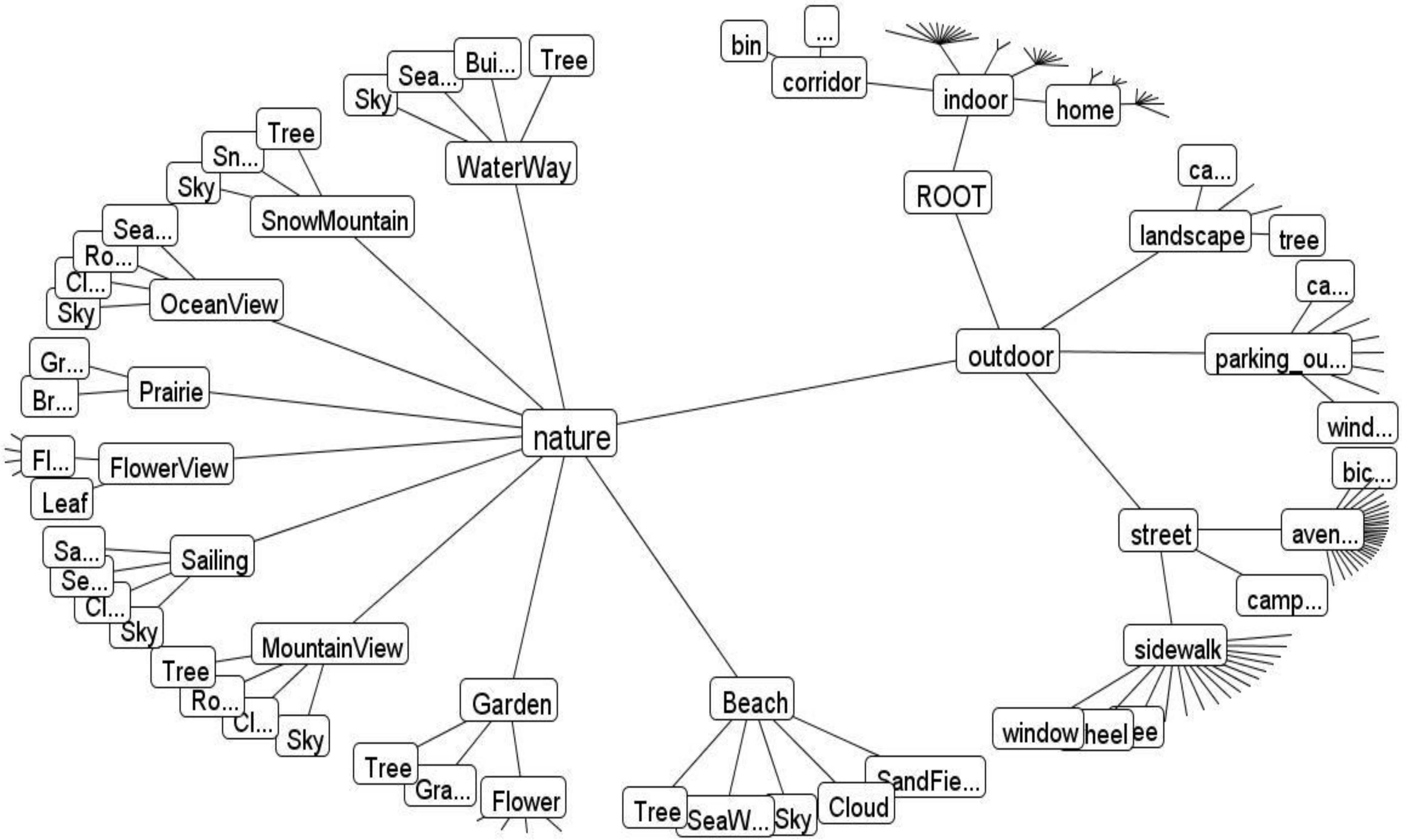


They are different but strongly related!

Concept Ontology can provide a good environment for multi-task learning!

Classifier Training for High-Level Image Concepts

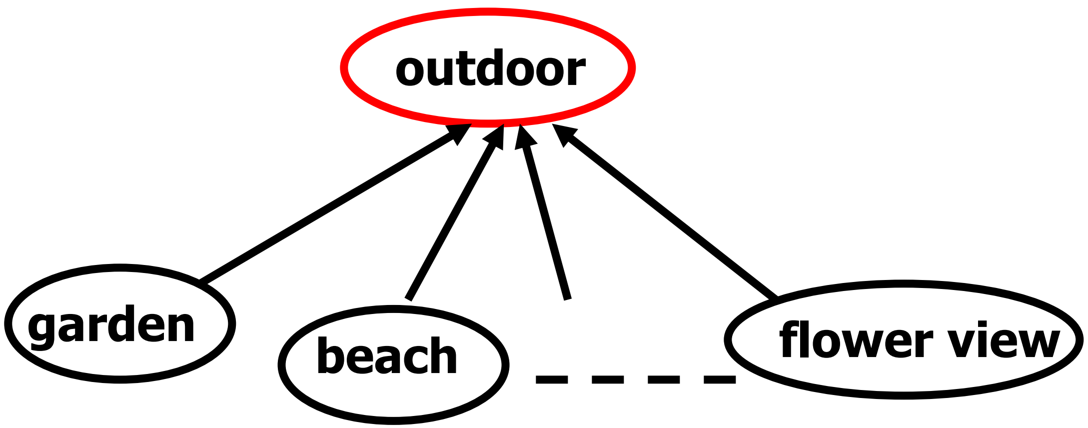
- Related Learning Tasks



Classifier Training for High-Level Image Concepts

- **Relatedness Modelling**

$$f_{C_j}(X) = W_j^T X + b \qquad W_j = W_0 + V_j$$



W_0**: Common Prediction Structure**

Classifier Training for High-Level Image Concepts

- **Joint Objective Function**

$$\min\left\{\sum_{j=1}^{C}\sum_{i=1}^{N}\xi_{ij}+\frac{\beta_1}{C}\sum_{j=1}^{C}\left\|V_j\right\|^2+\beta_2 C\left\|W_0\right\|^2\right\}$$

Subject to:

$$\forall_{j=1}^{C}\forall_{i=1}^{N}: Y_{ij}(W_0+V_j)\bullet X_{ij}+b\geq 1-\xi_{ij}$$

Classifier Training for High-Level Image Concepts

- **Dual Problem**

$$\max\left\{\sum_{j=1}^{C}\sum_{i=1}^{N}\alpha_{ij}-\frac{1}{2}\sum_{j=1}^{C}\sum_{i=1}^{N}\sum_{h=1}^{C}\sum_{l=1}^{N}\alpha_{ih}Y_{ih}\alpha_{jl}Y_{jl}K_{jh}(X_{ih},X_{jl})\right\}$$

Subject to:

$$\forall_{i=1}^{N}\forall_{j=1}^{C}:0\leq\alpha_{ij}\leq C,\sum_{j=1}^{C}\sum_{i=1}^{N}\alpha_{ij}Y_{ij}=0$$

Classifier Training for High-Level Image Concepts

- **Biased Classifier Training**

$$\min \left\{ \frac{1}{2} \left\| W - W_0 \right\|^2 + \alpha \sum_{l=1}^{m} [1 - Y_l (W^T \bullet X_l + b)] \right\}$$

- **Dual Problem**

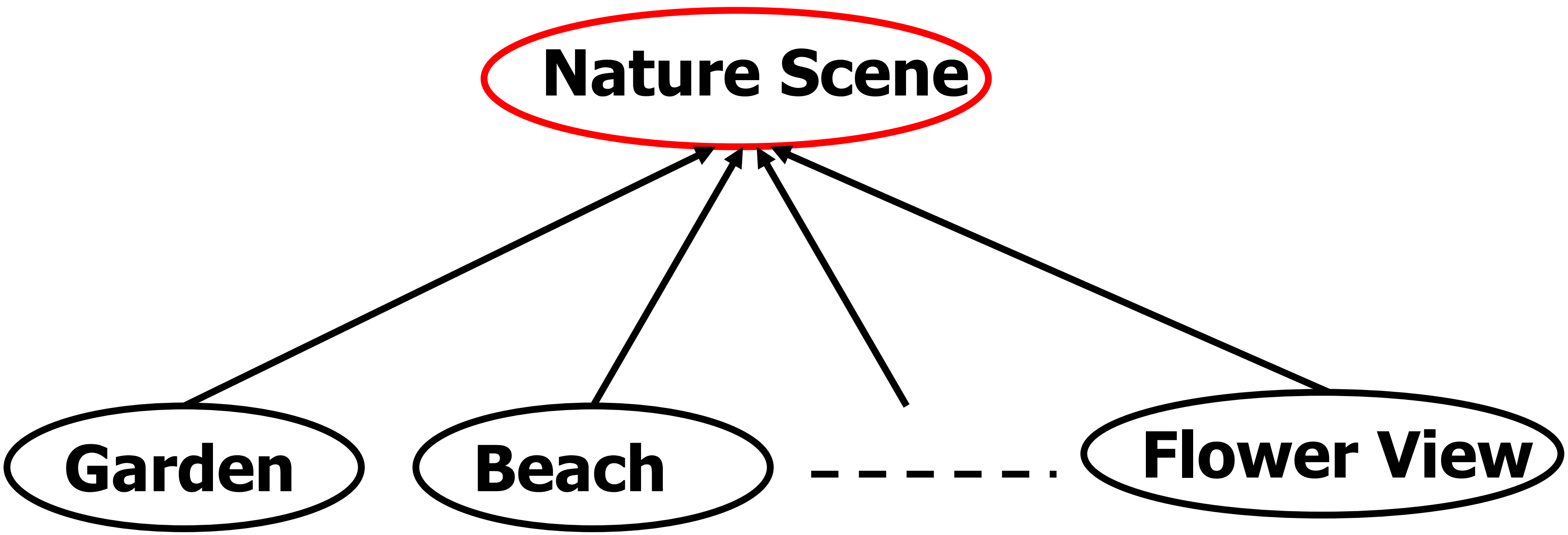
$$\min \left\{ \frac{1}{2} \sum_{l=1}^{m} \sum_{h=1}^{m} \alpha_l \alpha_h Y_l Y_h X_l^T X_h - \sum_{l=1}^{m} \alpha_l (1 - Y_l W_0^T X_l) \right\}$$

Subject to:

$$\forall_{l=1}^{m} : 0 \le \alpha_l \le C, \sum_{l=1}^{m} \alpha_l Y_l = 0$$

Classifier Training for High-Level Image Concepts

- Common Prediction Structure



Common Visual Properties

Classifier Training for High-Level Image Concepts

- **Hierarchical Boosting**

$$H_{C_k}(X) = \sum_{j=1}^{C+1} p_j(X) f_{C_j}(X)$$

$$p_j(X) = \frac{\exp(f_{C_j}(X))}{\sum_{j=1}^{C+1} \exp(f_{C_j}(X))}$$

Classifier Training for High-Level Image Concepts

- **Biased Classifier for Parent Node**

$$W = W_0 + \sum_{l=1}^{m} \alpha_l Y_l X_l$$

$$f_{C_k}(X) = W^T X + b$$

Classifier Training for High-Level Image Concepts

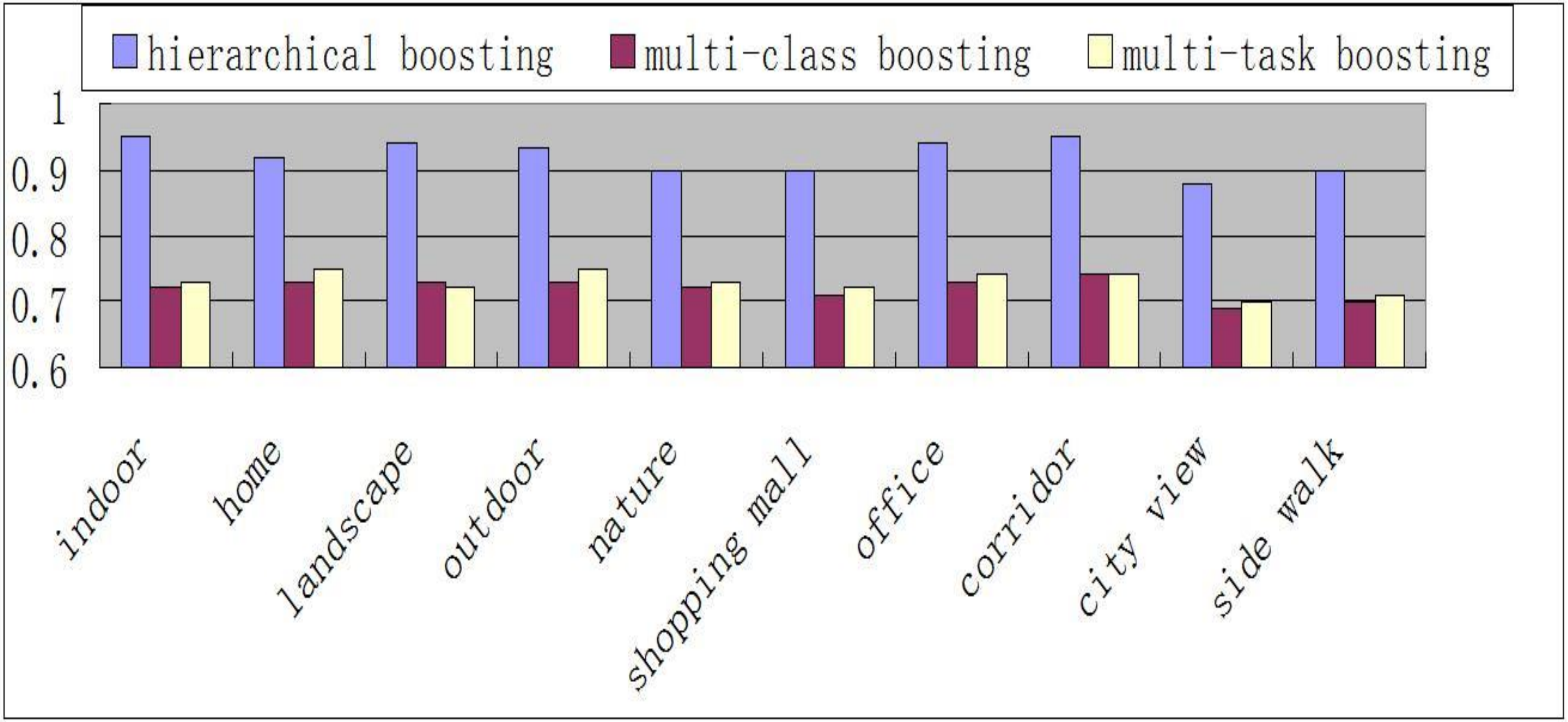
- **Hierarchical Boosting to Generate Classifier for Parent Node**

$$H_{C_k}(X) = \sum_{j=1}^{C+1} p_j(X) f_{C_j}(X)$$

$$p_j(X) = \frac{\exp(f_{C_j}(X))}{\sum_{j=1}^{C+1} \exp(f_{C_j}(X))}$$

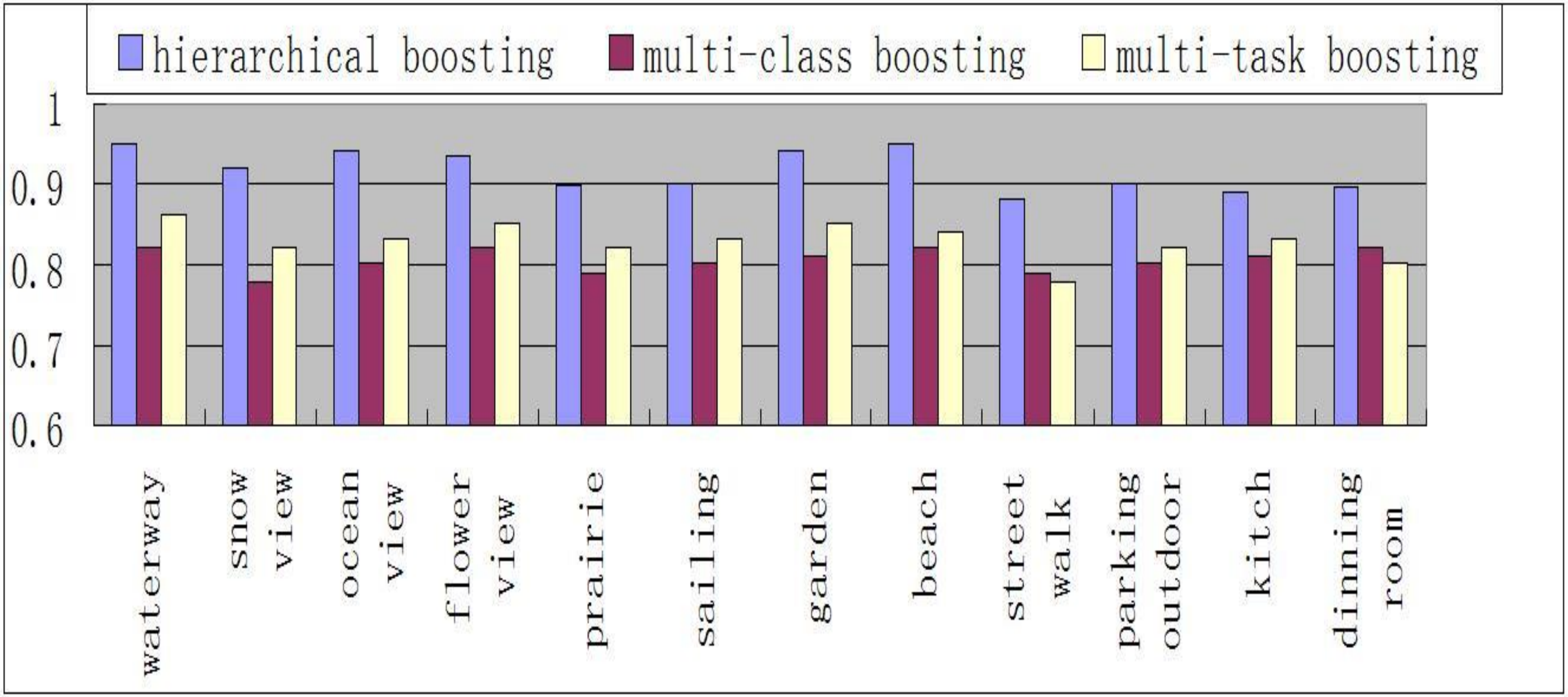
Classifier Training for High-Level Image Concepts

- **Performance Evaluation**



Classifier Training for High-Level Image Concepts

- **Performance Evaluation**

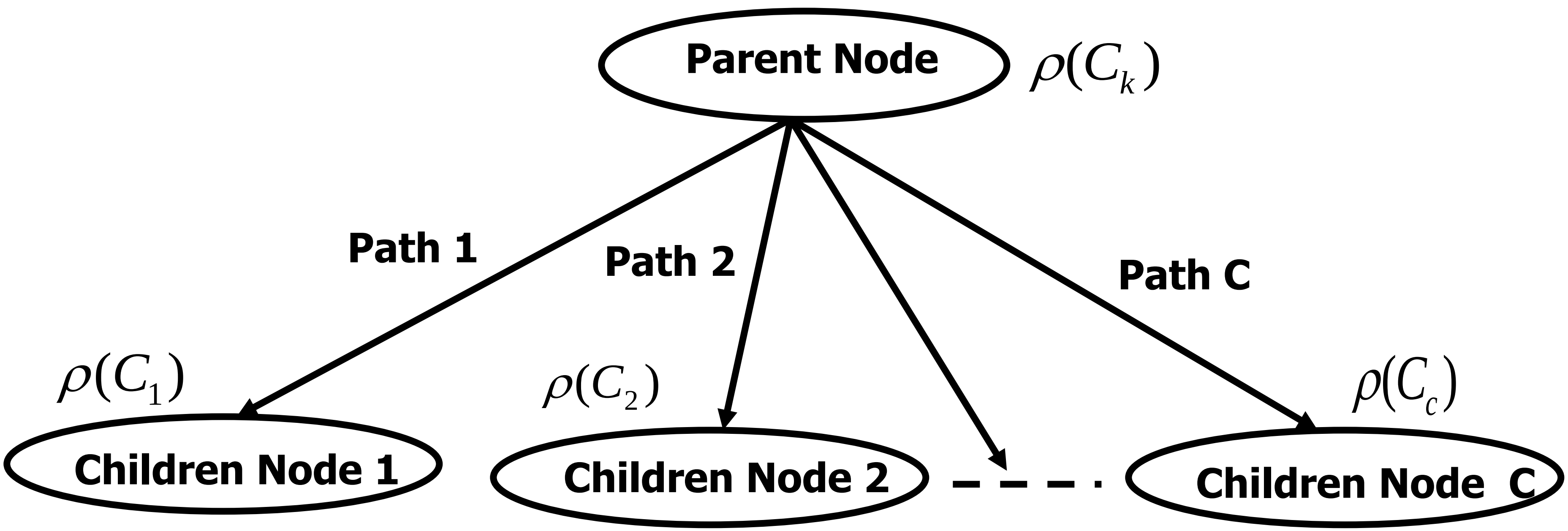


Classifier Training for High-Level Image Concepts

- **Advantages of Hierarchical Boosting**
  - **Handling inter-concept similarity via multi-task learning**
  - **Reducing training cost**
  - **Enhancing discrimination power of the classifiers**

Hierarchical Image Classification

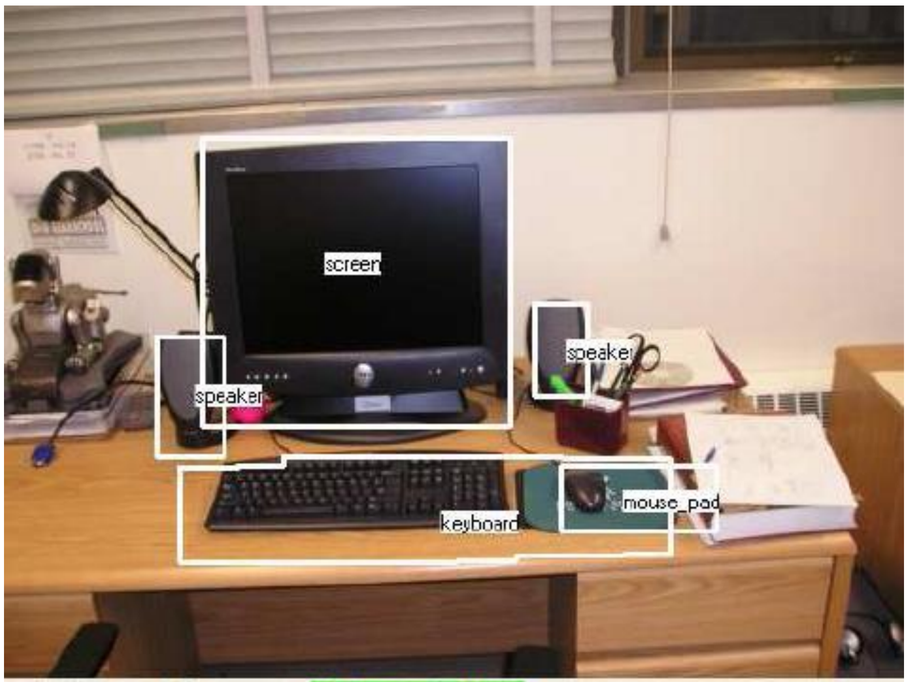
- Overall Probability $\varphi(C_k)$



$$\varphi(C_k) = \rho(C_k) + \max\{\rho(C_j) \mid j = 1, \ldots C\}$$

Hierarchical Image Classification

- **Some Results**



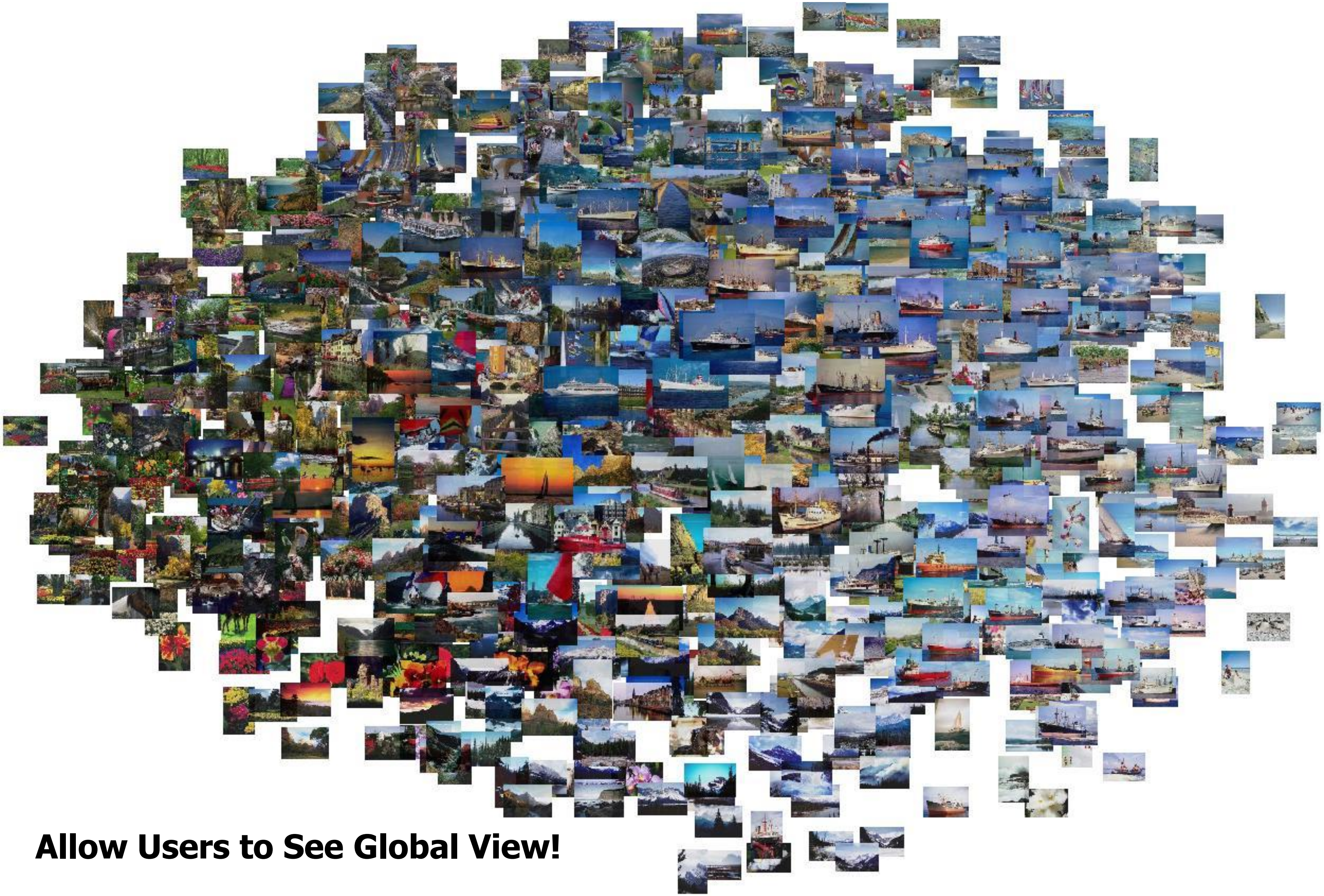
Multi-Level Concepts: office => indoor

Salient Objects: screen; keyboard; speaker; mouse_pad;

Multi-Level Concepts: sidewalk => street => outdoor

Salient Objects: person; car_left; wheel;

Classification Result Evaluation

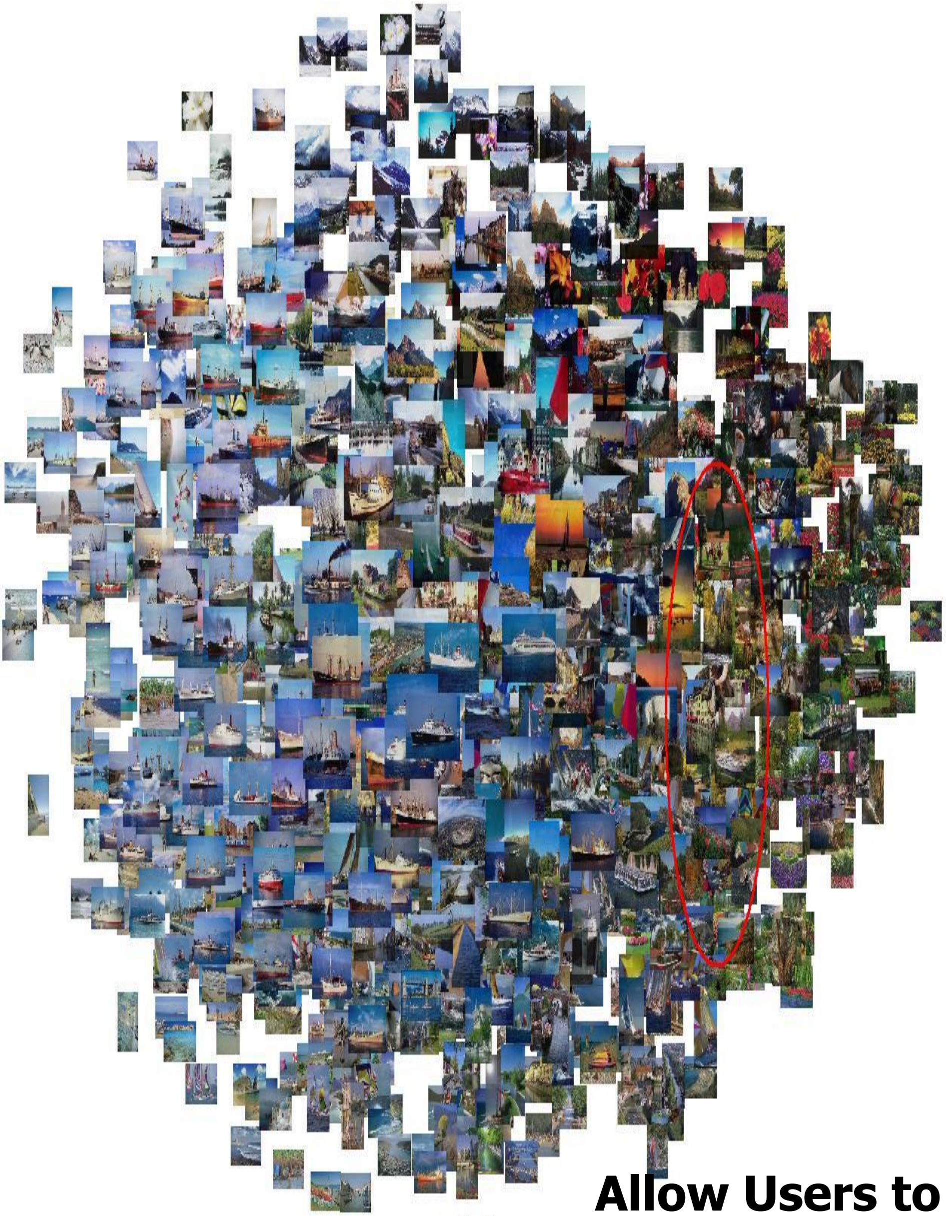


Allow Users to See Global View!

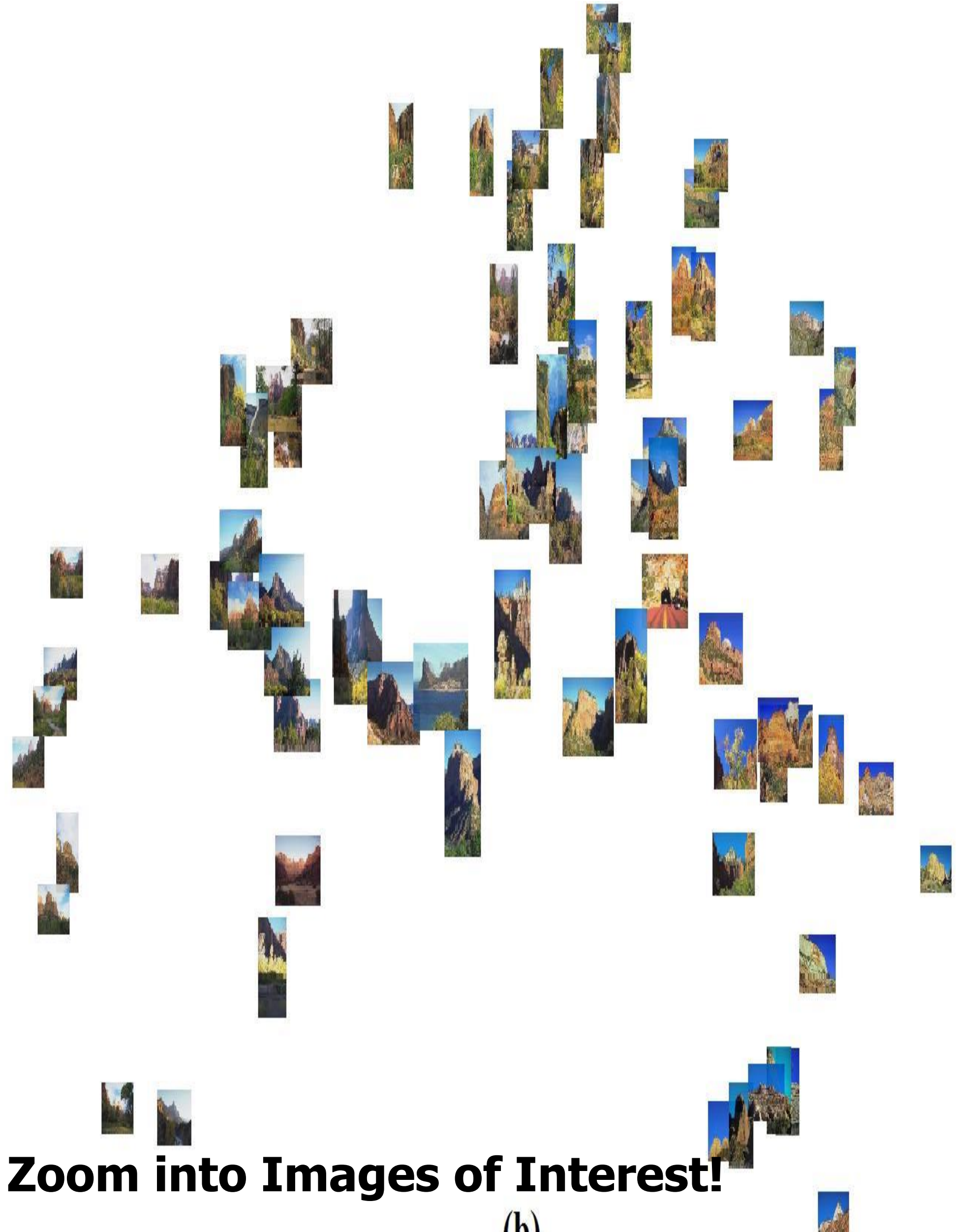
Classification Result Evaluation

Allow Users to See Similarity Direction!

Classification Result Evaluation



(a)



(b)

Allow Users to Zoom into Images of Interest!

Classification Result Evaluation: **Red Flower**

Allow Users to Select Query Example Interactively!

(a) (b)

Classification Result Evaluation: **Sunset**

Allow Users to Look for Particular Images!

(a) (b)

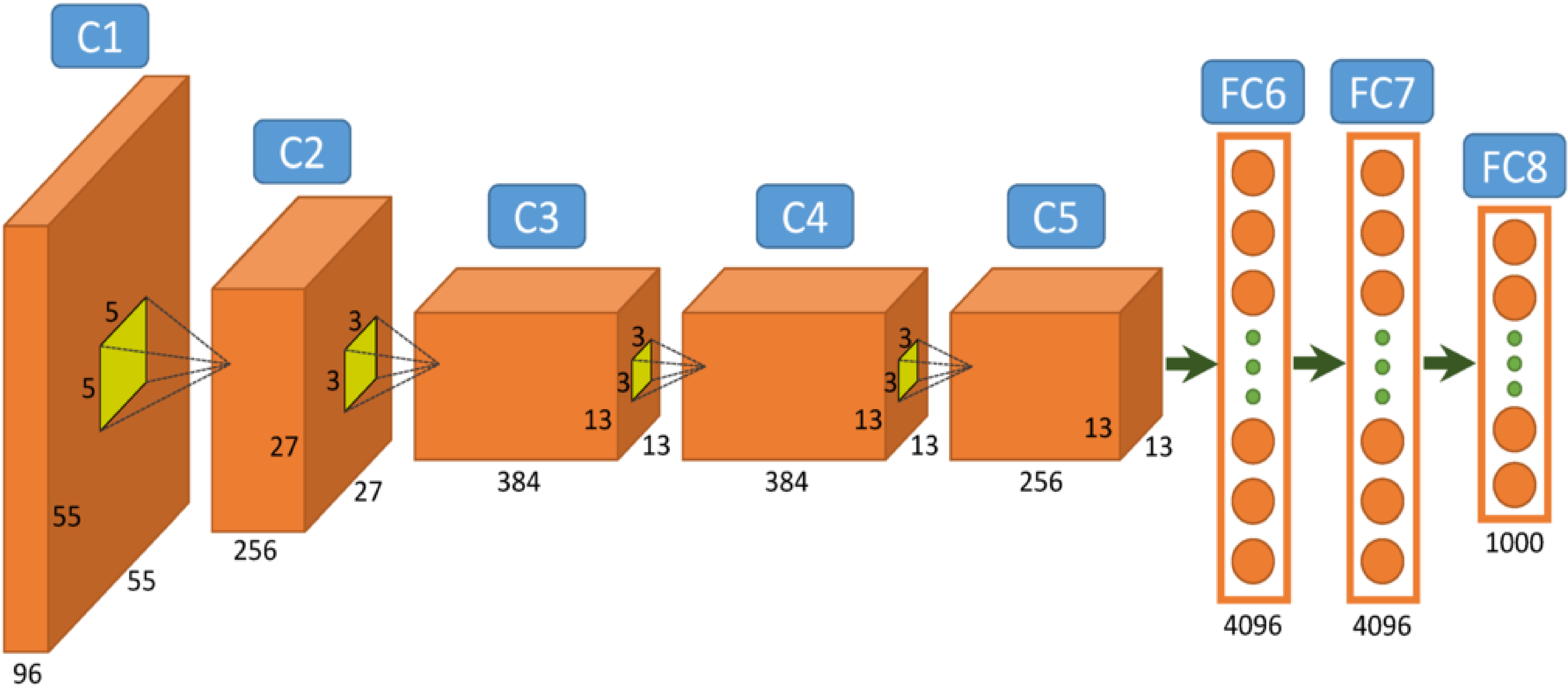
Some Interesting Observations

- **Concept Ontology for identifying inter-related learning tasks**
- **Multi-task learning for inter-related tasks**
- **Hierarchical learning over concept ontology**

Deep Multi-Task Learning over Concept Ontology

- Ontology-driven task group generation
- Deep multi-task learning for each task group
- Hierarchical deep multi-task learning over concept ontology

Traditional Deep Learning



Traditional Deep Learning

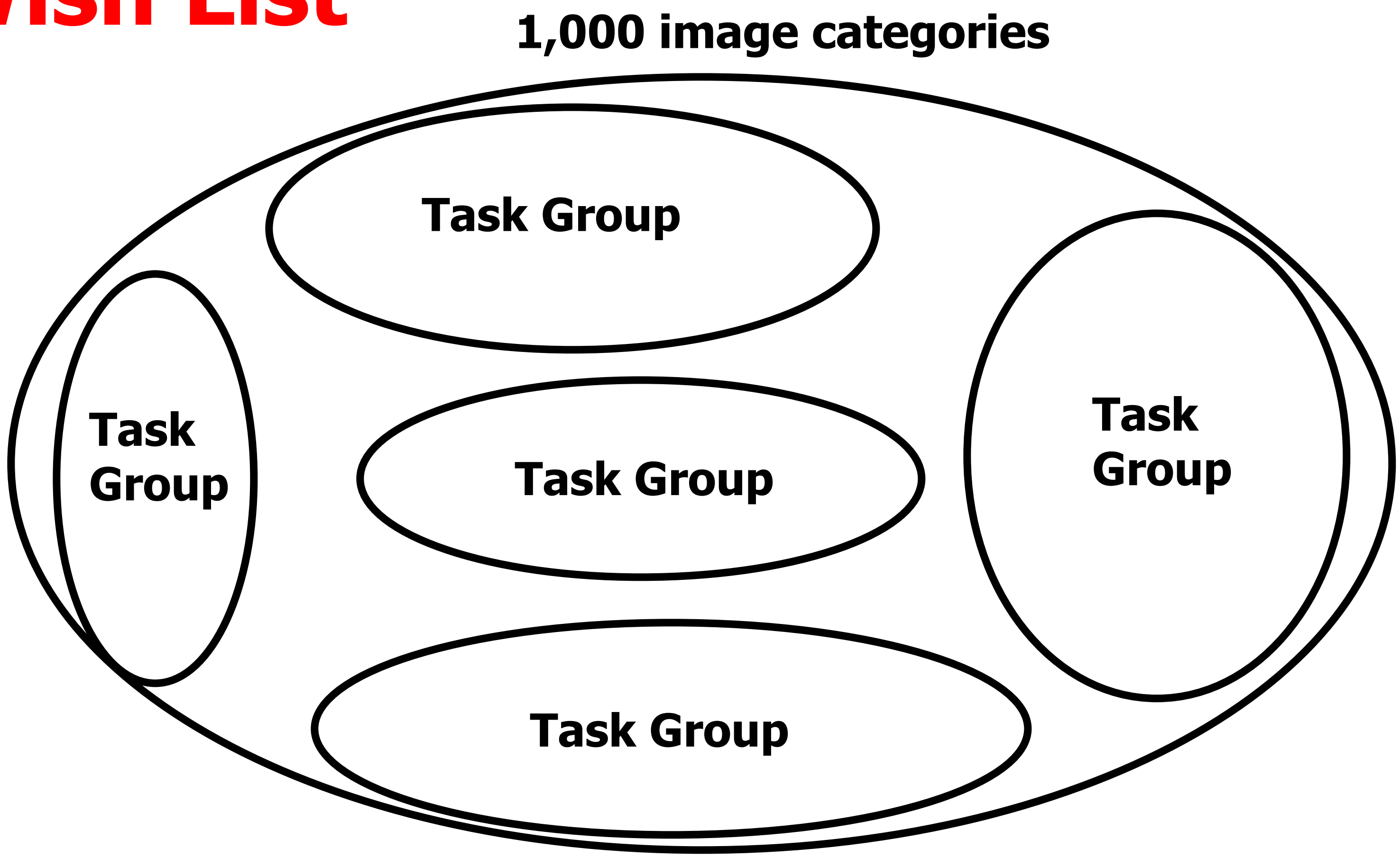
- **Problem of AlexNet**
  - **Inter-class visual correlations are completely ignored! They are assumed to be independent!**
  - **The differences of their learning complexities are completely ignored!**

⇩

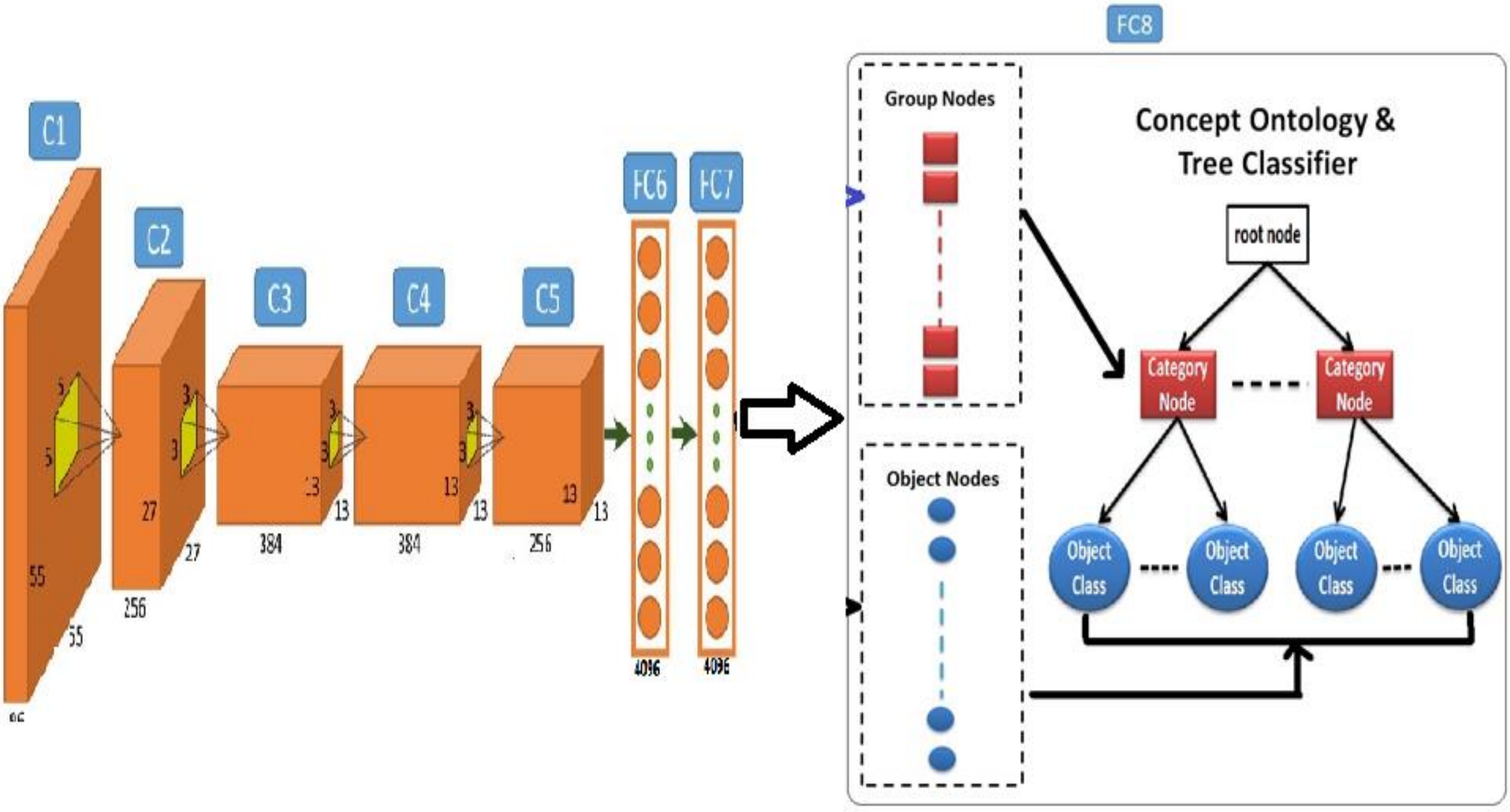
  - **Back-propagation for optimization may pay more attentions on hard ones!**

Traditional Deep Learning

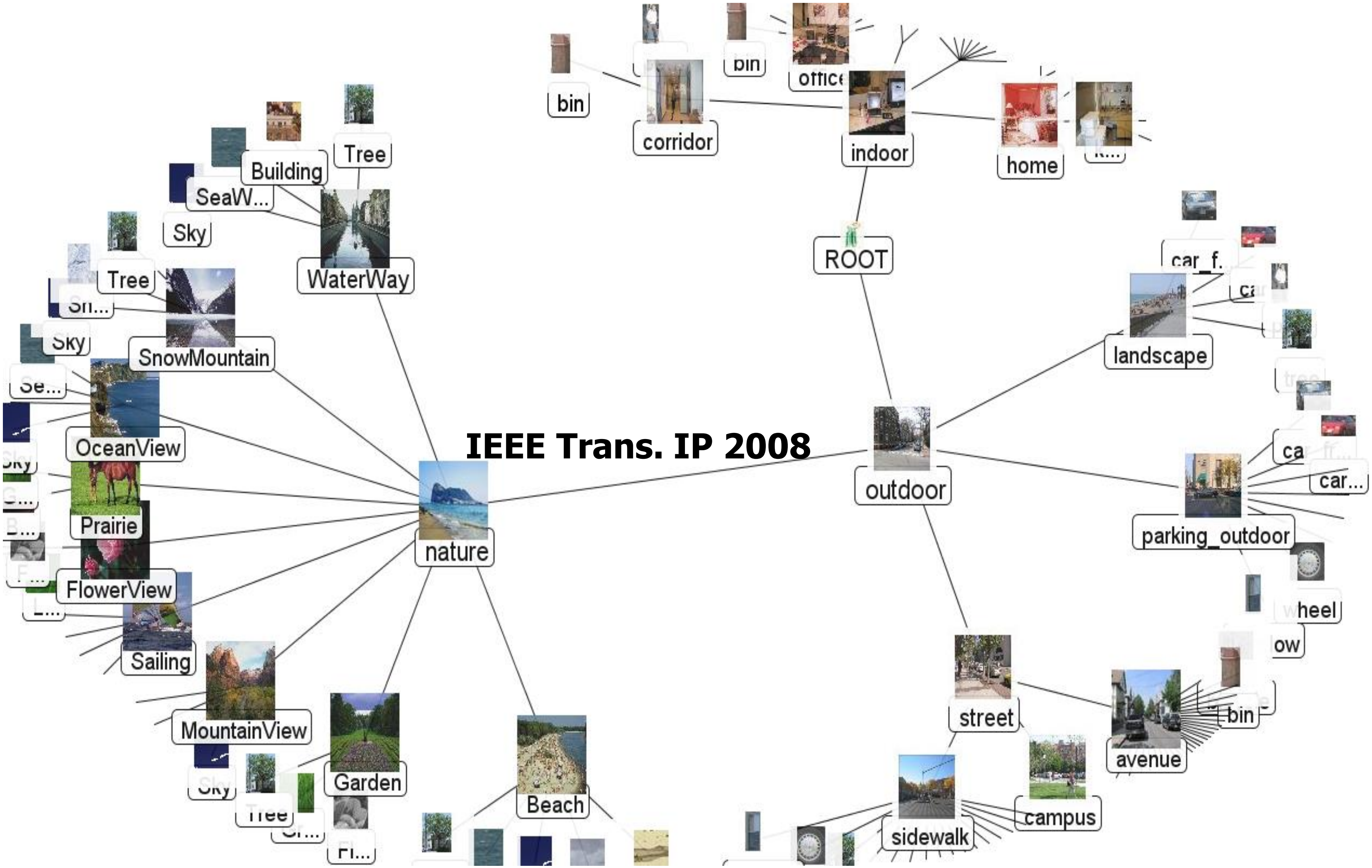
- **Wish List**



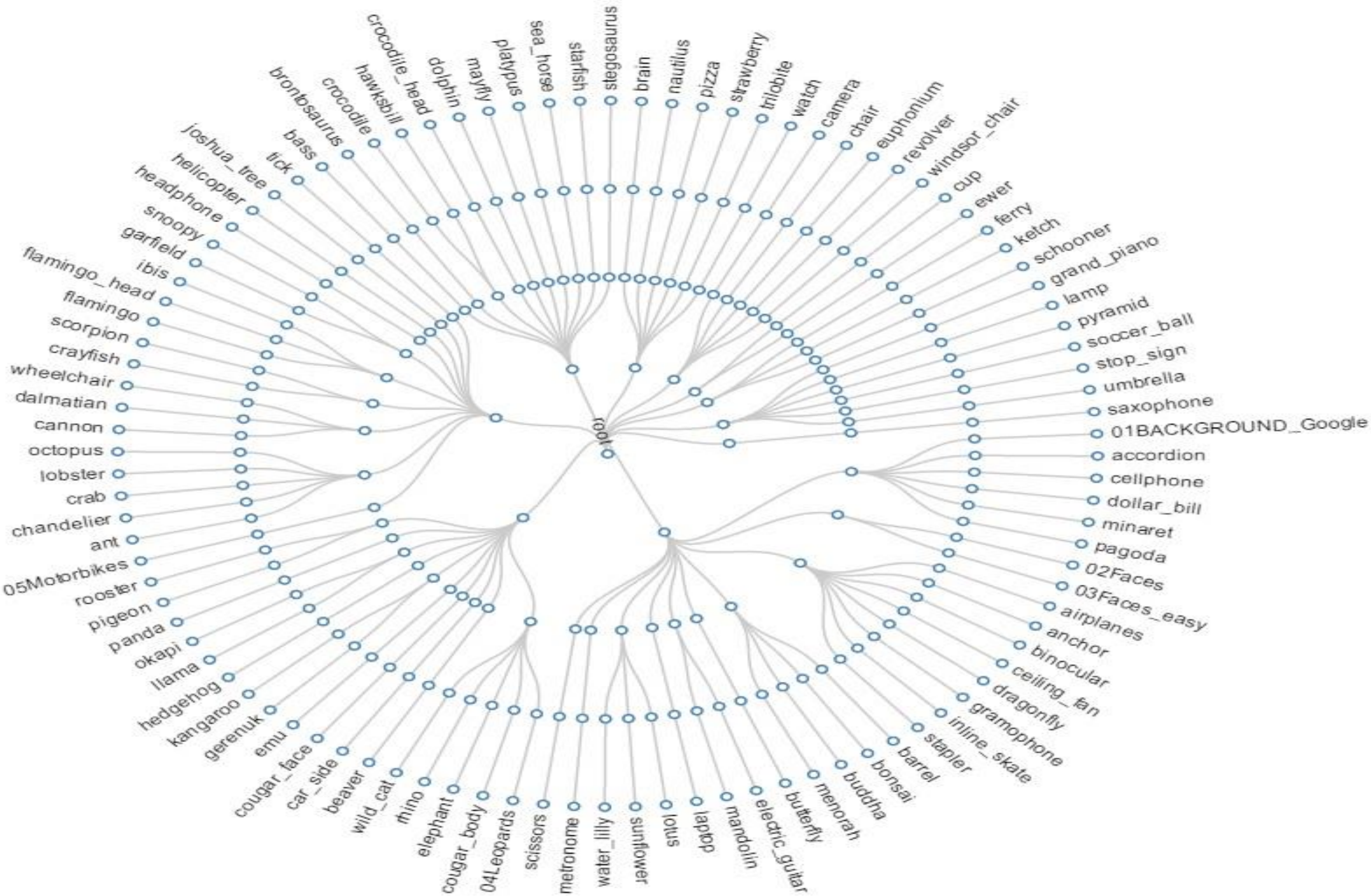
Deep Multi-Task Learning over Concept Ontology



Ontology for Task Group Generation

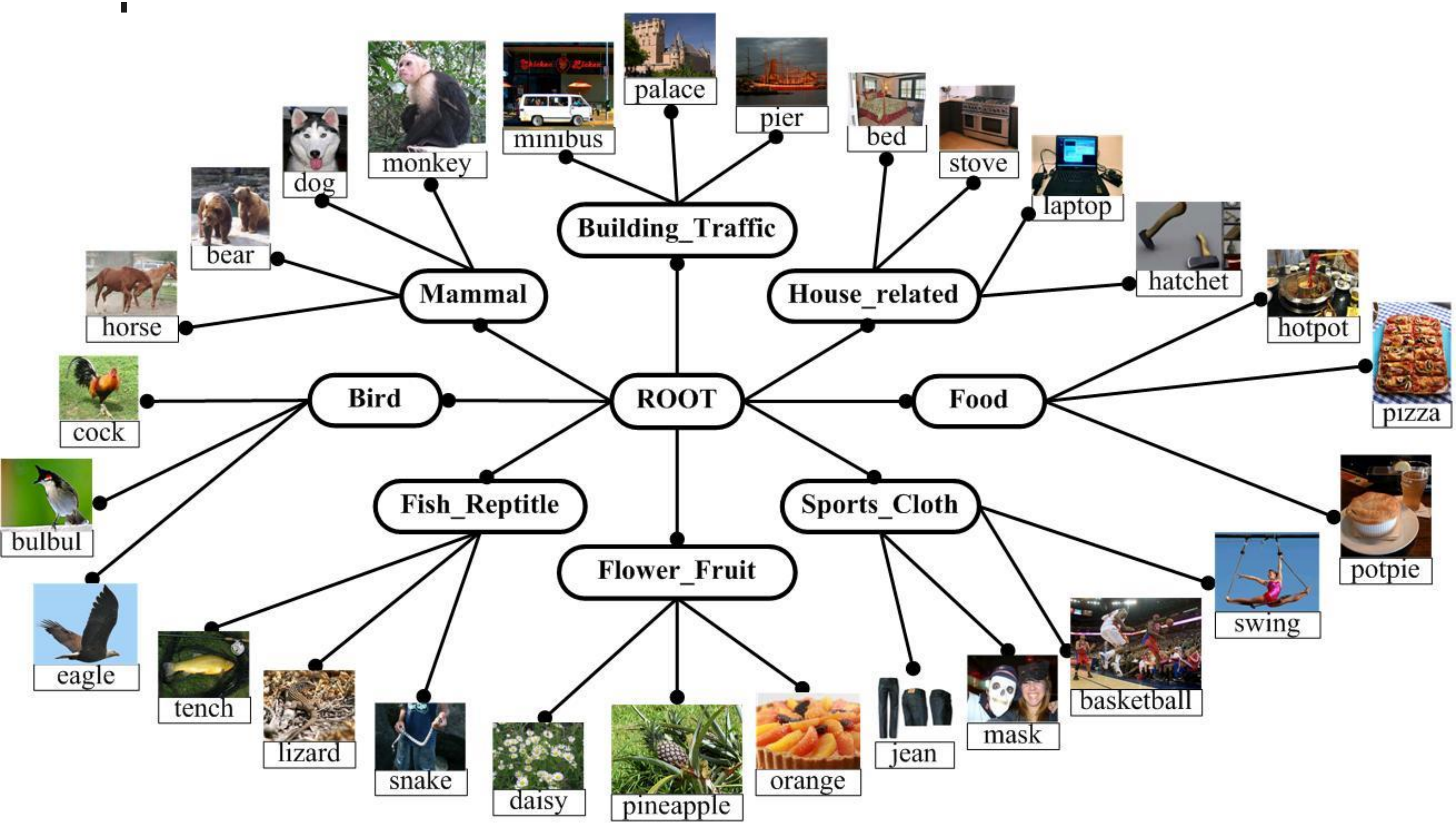


Ontology for Task Group Generation



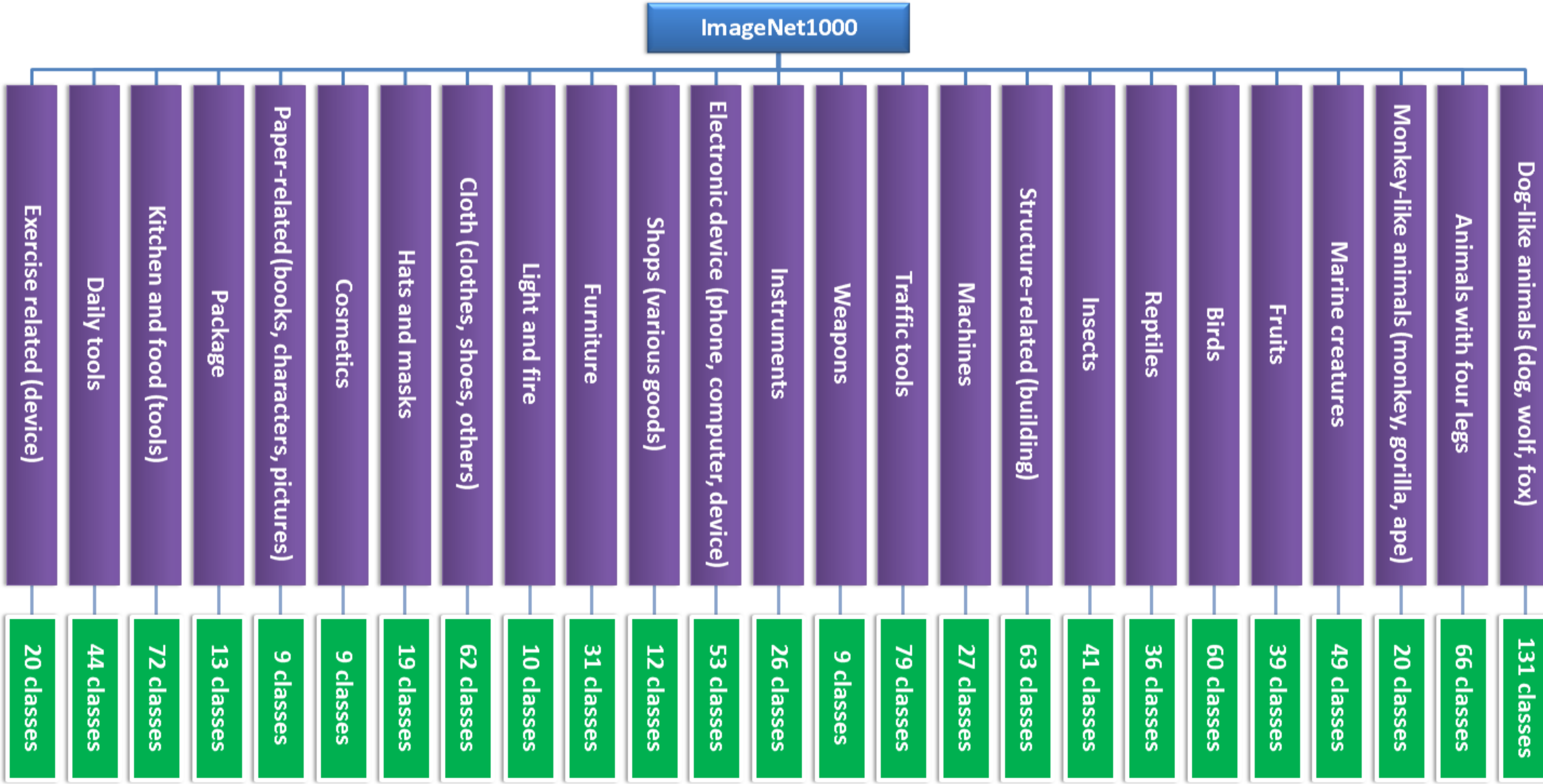
CalTech101

Ontology for Task Group Generation

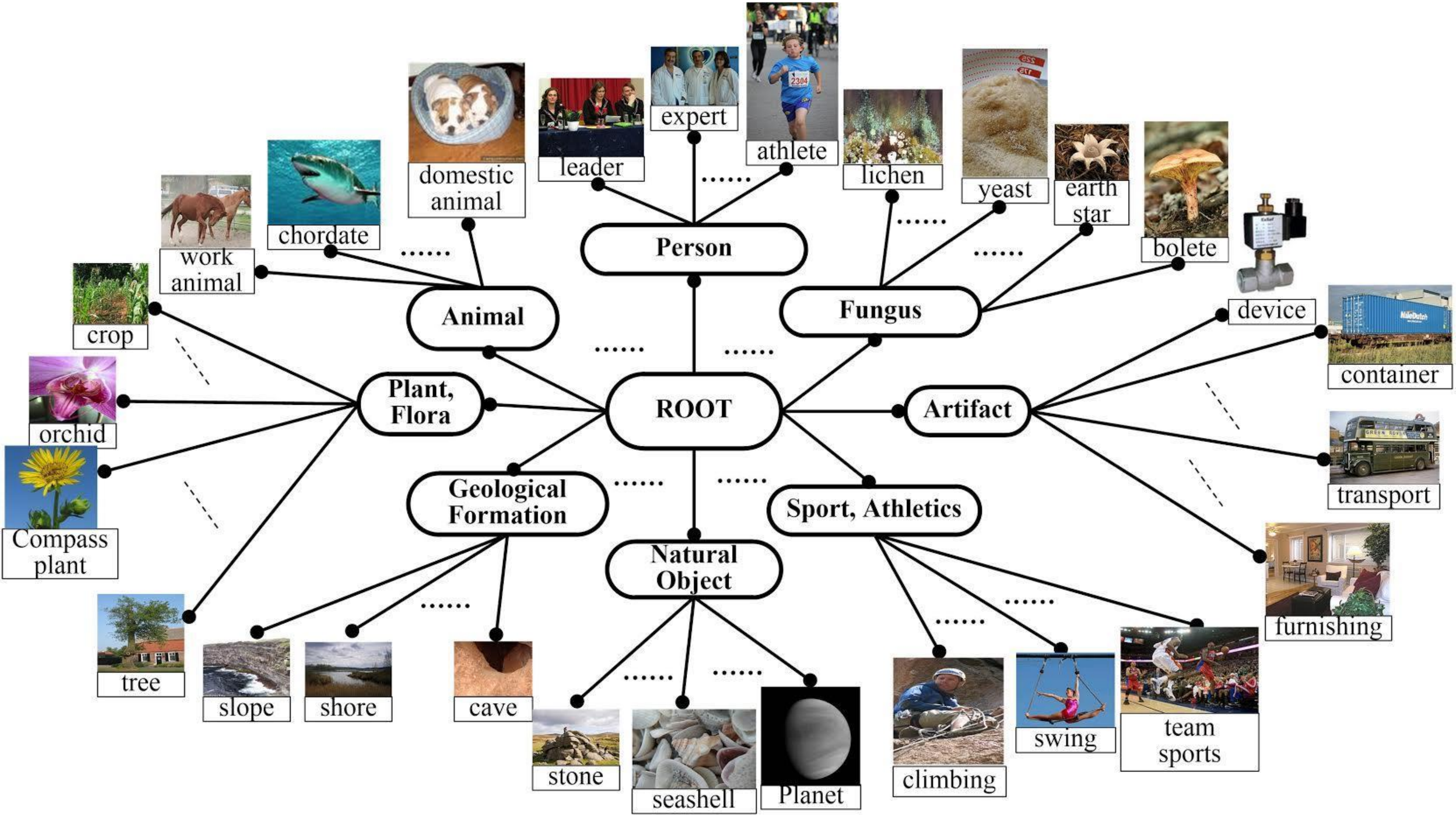


Two-Layer Ontology for ImageNet1K

Ontology for Task Group Generation

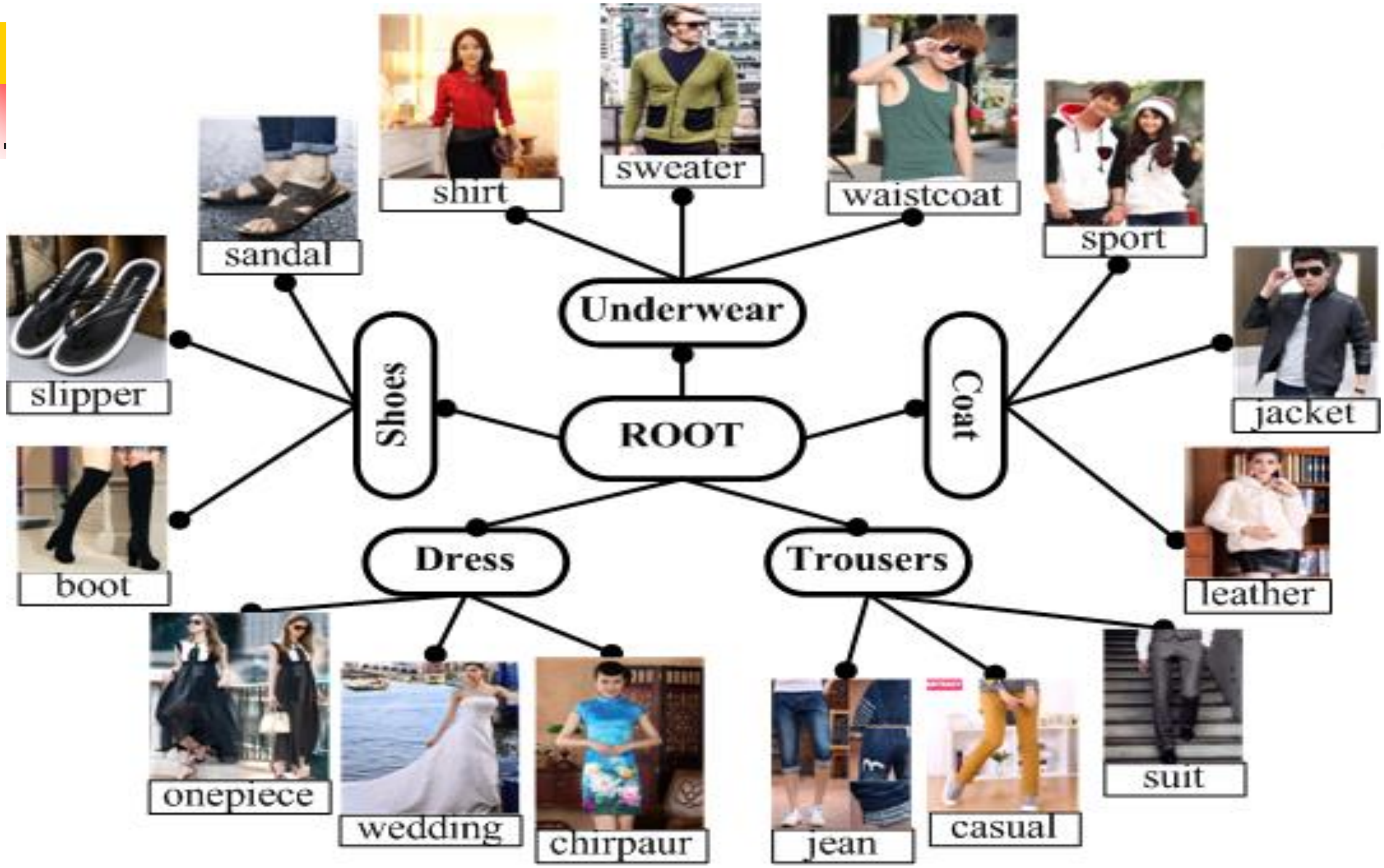


Ontology for Task Group Generation



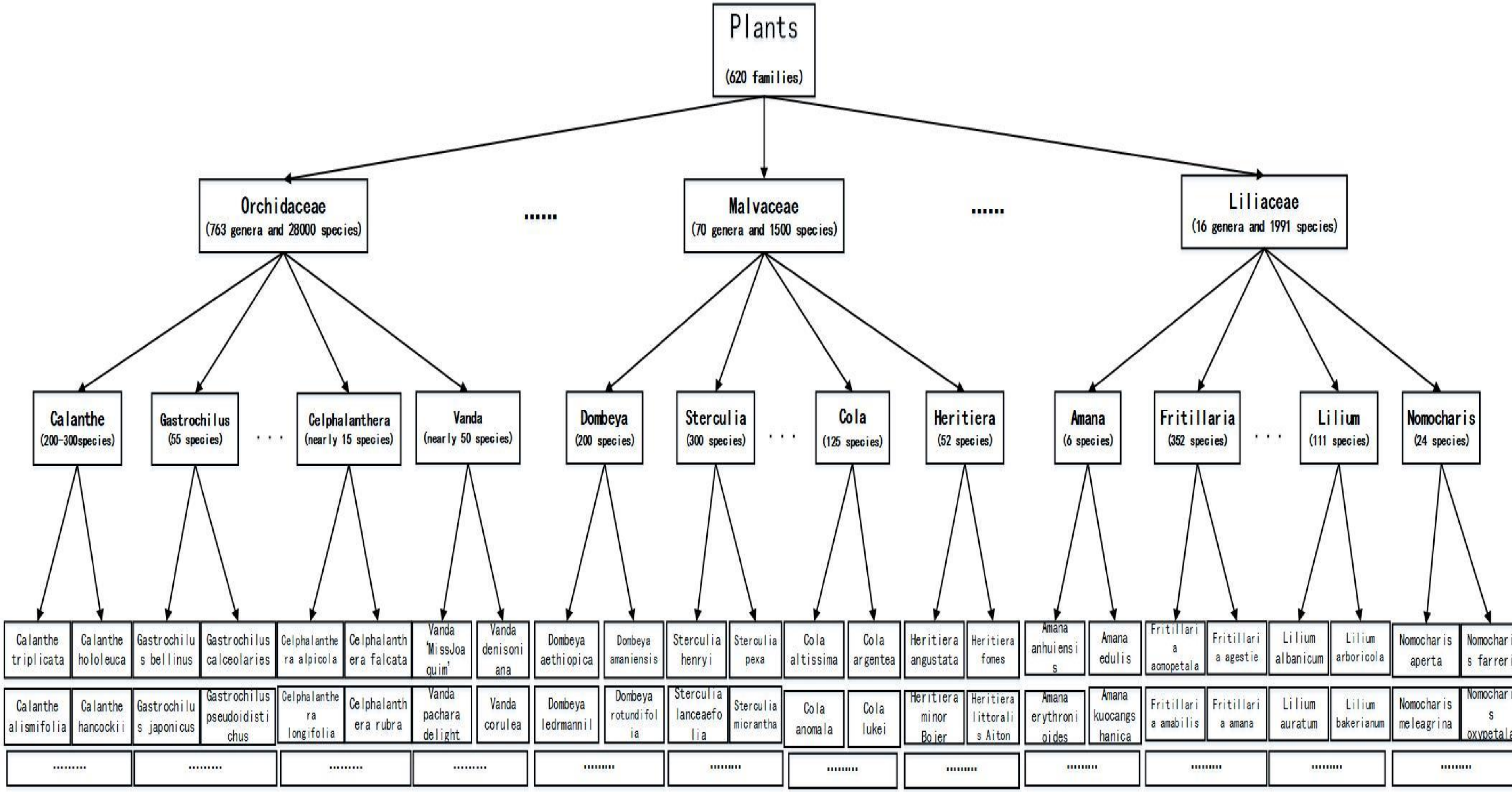
Two-Layer Ontology for ImageNet10K

Ontology for Task Group Generation



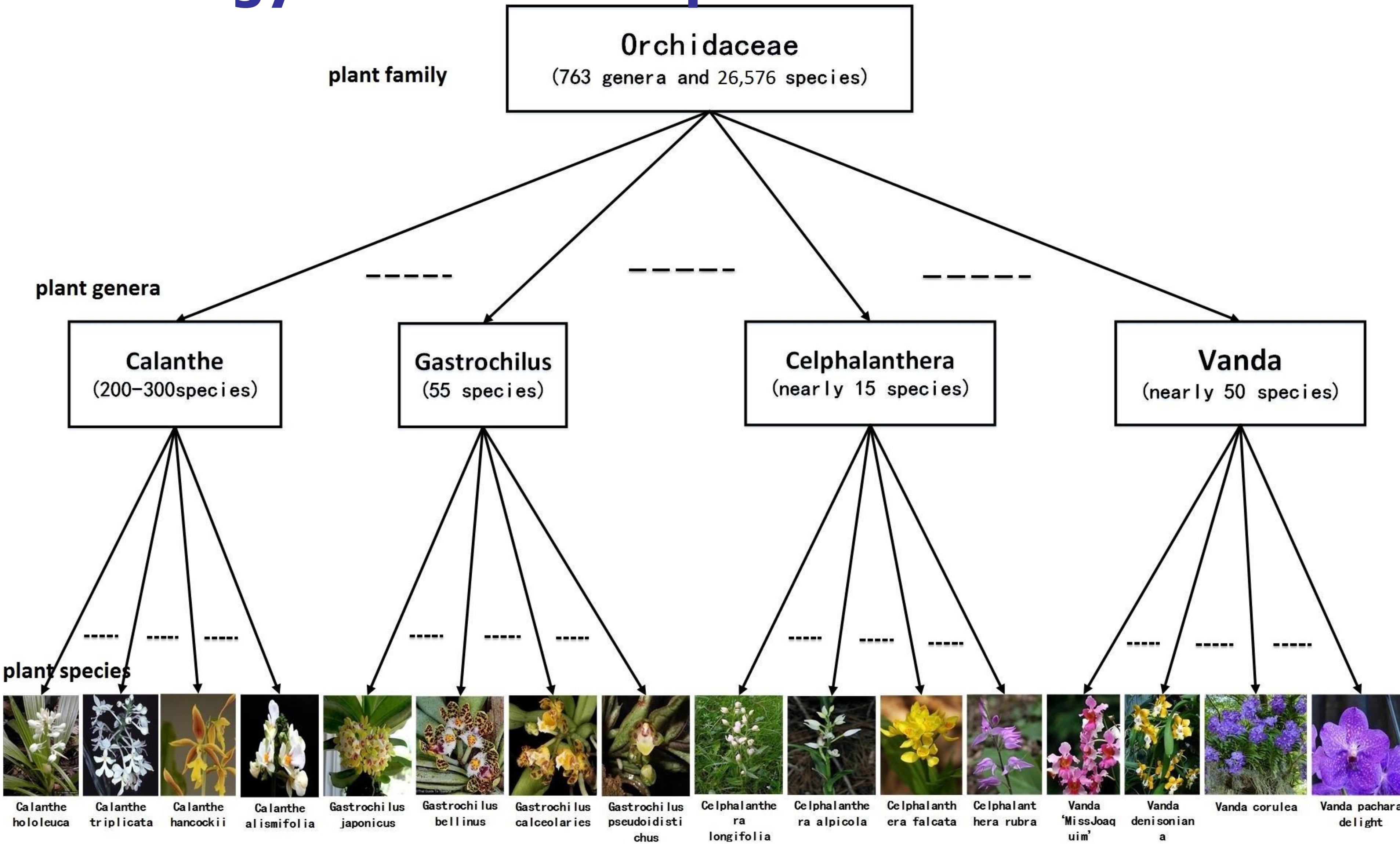
Two-Layer Ontology for Taobao Products

Ontology for Task Group Generation



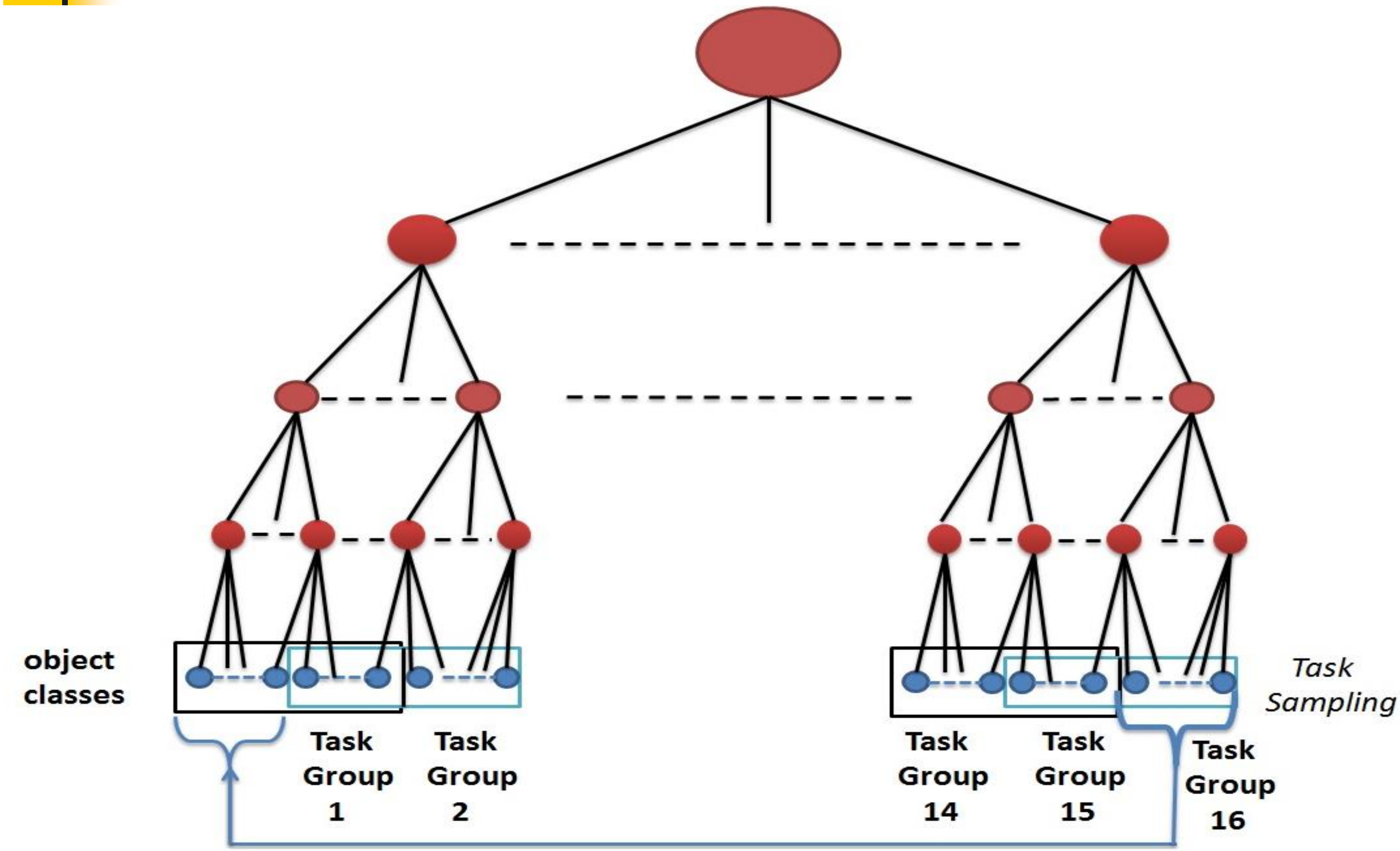
Plant Species Ontology

Ontology for Task Group Generation

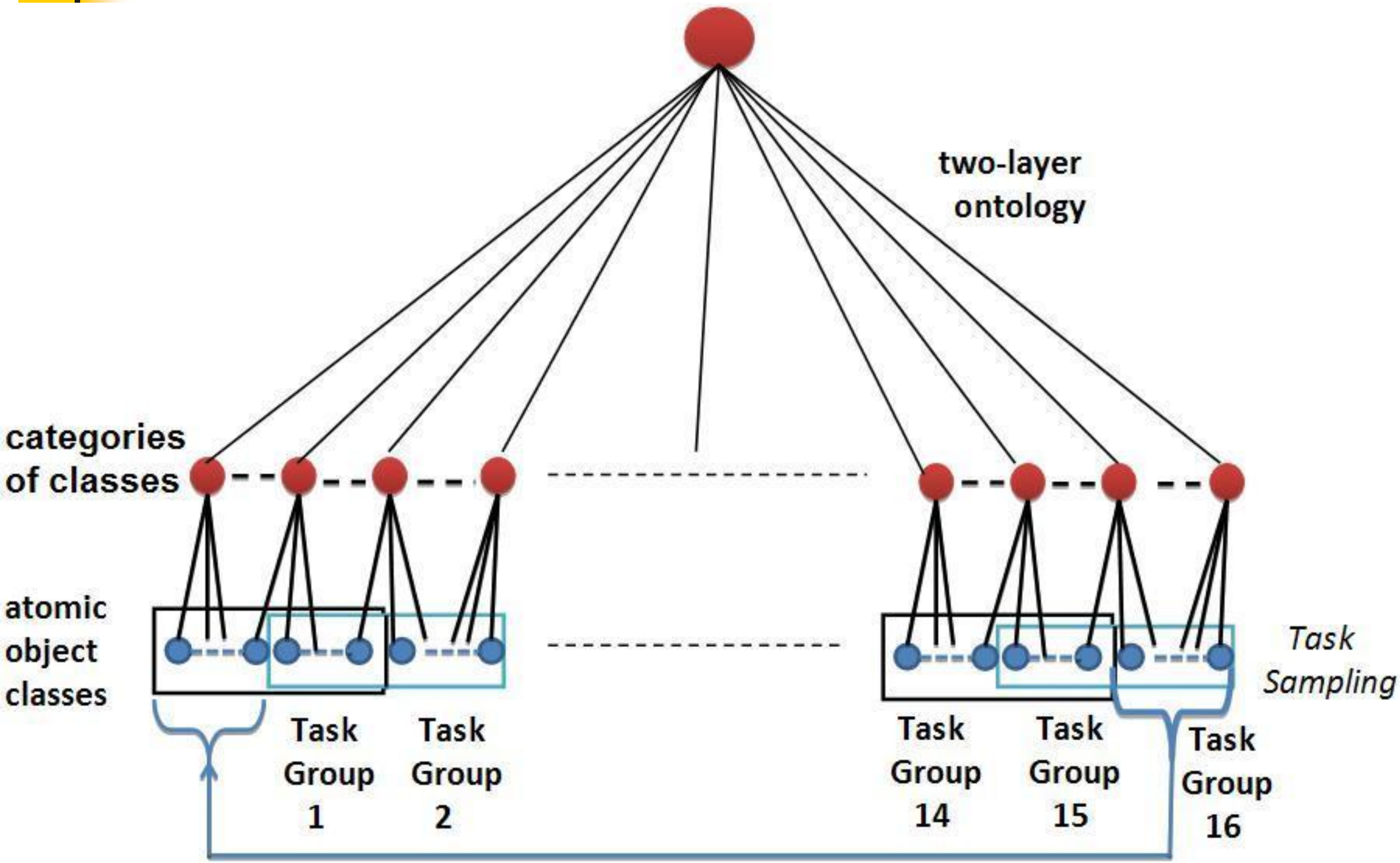


Two-Layer Ontology for Orchidaceae

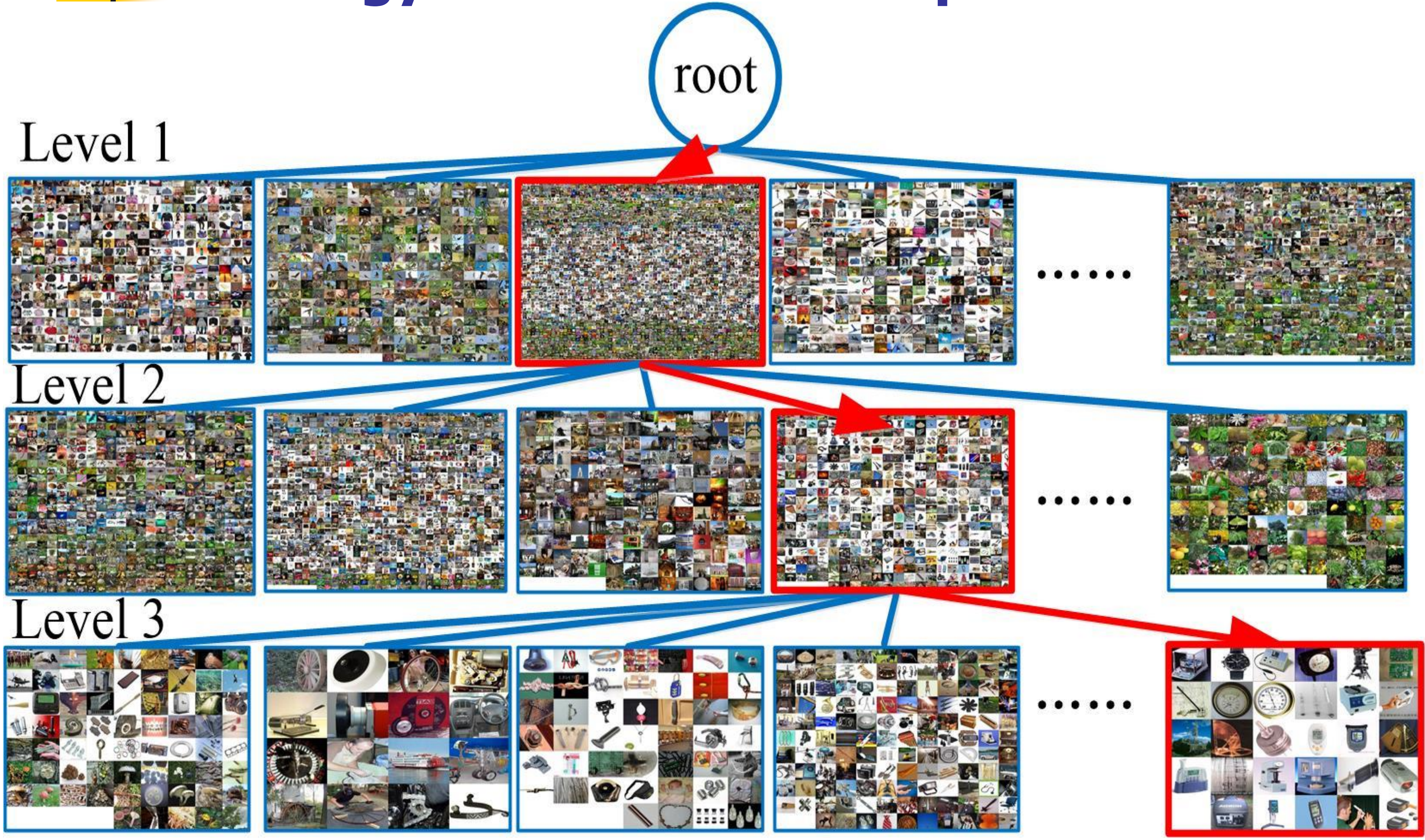
Ontology-Driven Task Group Generation



Ontology-Driven Task Group Generation

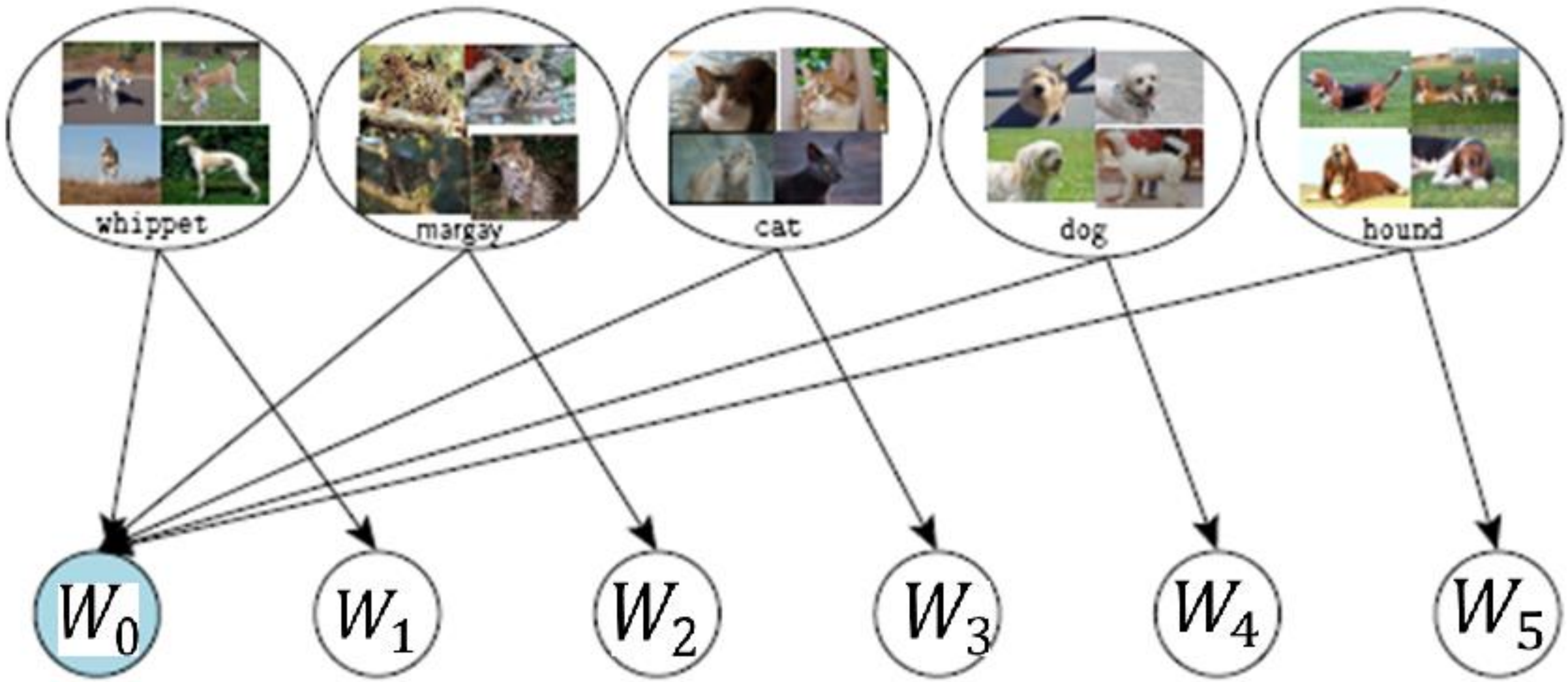


Ontology-Driven Task Group Generation



Learning Base CNNs for Each Task Group

- **Deep Multi-Task Learning:** $F_j(\mathbf{x}) = (W_0 + W_j)^T \mathbf{x} + \mathbf{b}$



Learning Base CNNs for Each Task Group

- **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:

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

Learning Base CNNs for Each Task Group

- **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:

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

Learning Base CNNs for Each Task Group

- **Deep Multi-Task Learning**

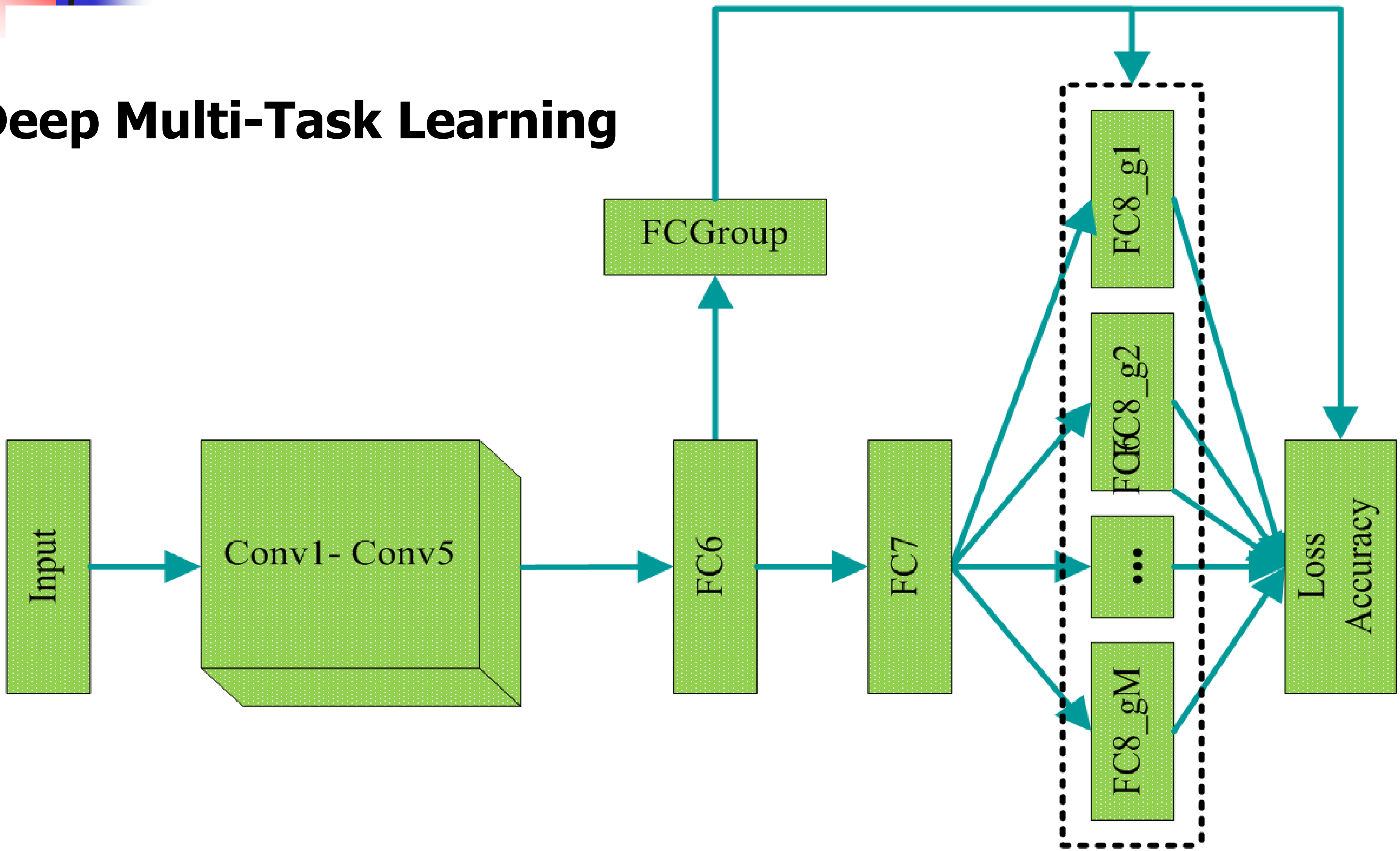
$$\alpha^* = \frac{1}{2\delta_1}\left(\Re + \frac{\delta_2}{\delta_1}\left(\Re\left(L \bigotimes I\right)\Re\right)^{-1}\Re Y\beta^*\right)$$

Multi-Task Classifiers at Sibling Leaf Nodes

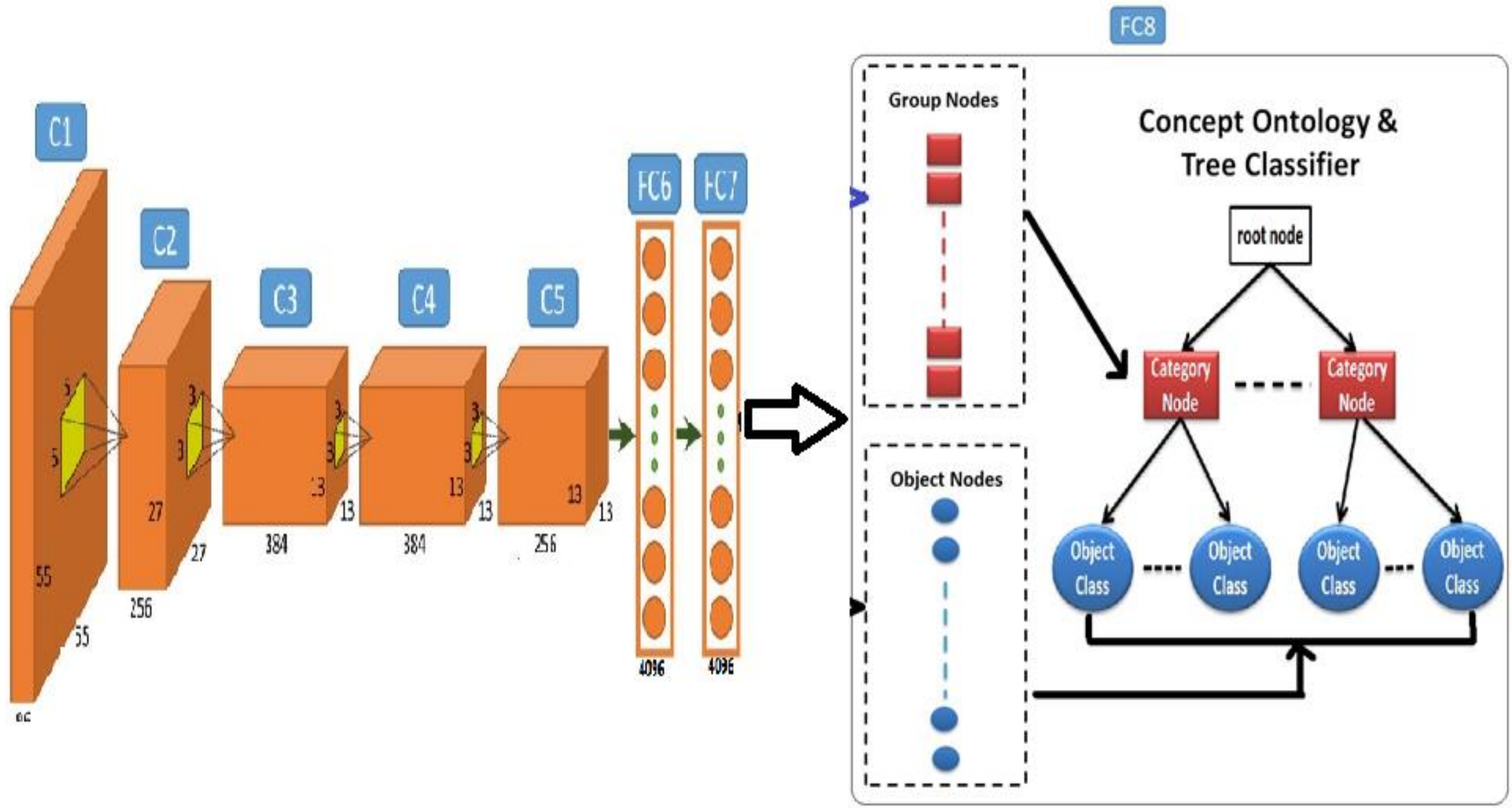
$$\forall_{j=1}^{B}: \quad f_{c_j}^1(x)\mid_{F_{c_j}^1} = \sum_{l=1}^{R}\alpha_j^{l*}\kappa(x_j^l, x) + b_j^*, \quad c_j \in c_h$$

Learning Base CNNs for Each Task Group

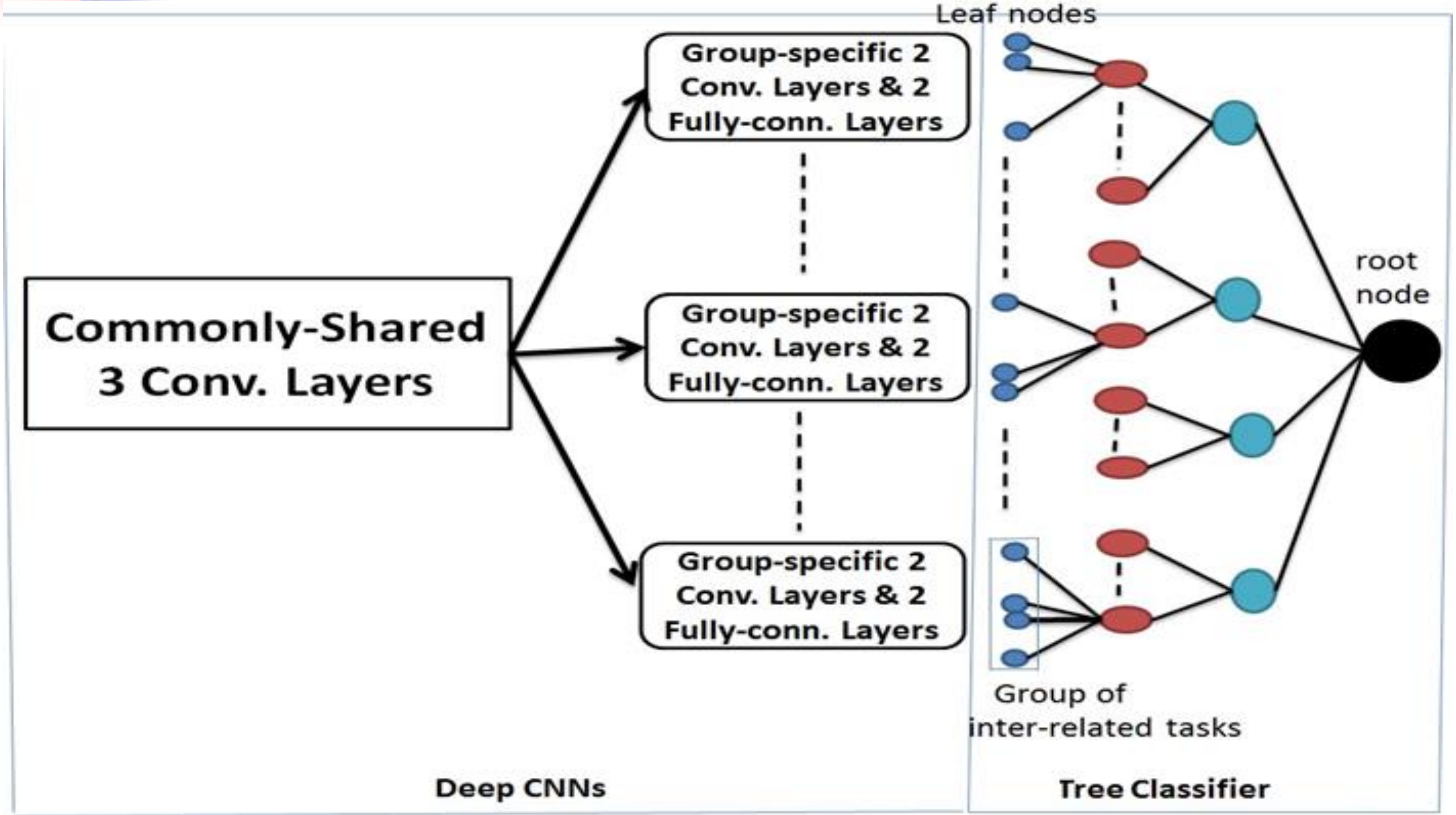
Deep Multi-Task Learning



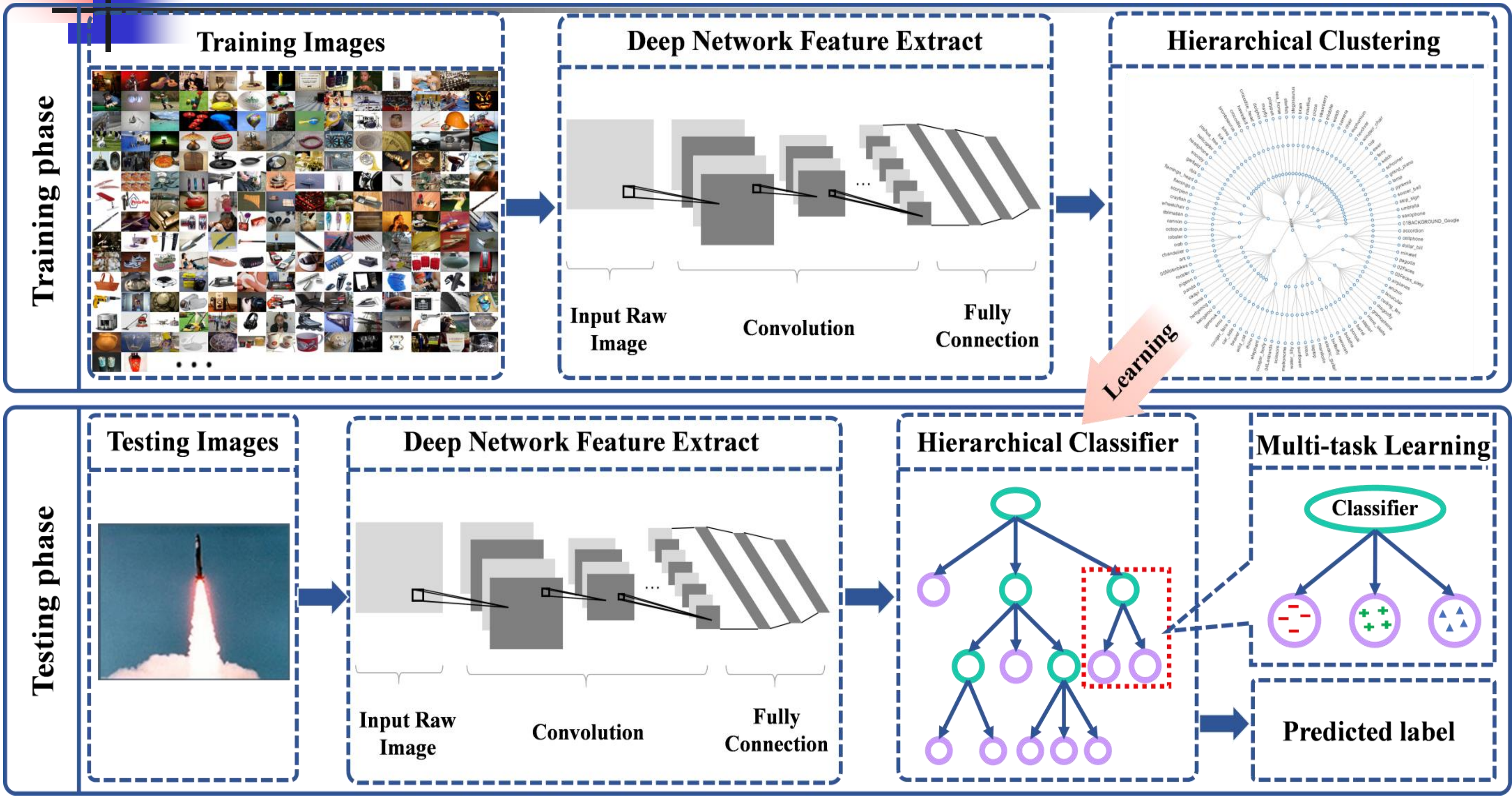
Hierarchical Deep Multi-Task Learning



Hierarchical Deep Multi-Task Learning



Hierarchical Deep Multi-Task Learning



Hierarchical Deep Multi-Task Learning

- **Controlling Inter-Level Error Propagation**

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

subject to:

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

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

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

Hierarchical Deep Multi-Task Learning