University of Michigan * KAIST † ‡

Hierarchical Novelty Detection for Visual Object Recognition

Kibok Lee* Kimin Lee† Kyle Min* Yuting Zhang* Jinwoo Shin† Honglak Lee*‡

Code: github.com/kibok90/cvpr2018-hnd

CVPR 2018 SALT LAKE CITY • JUNE 18-22

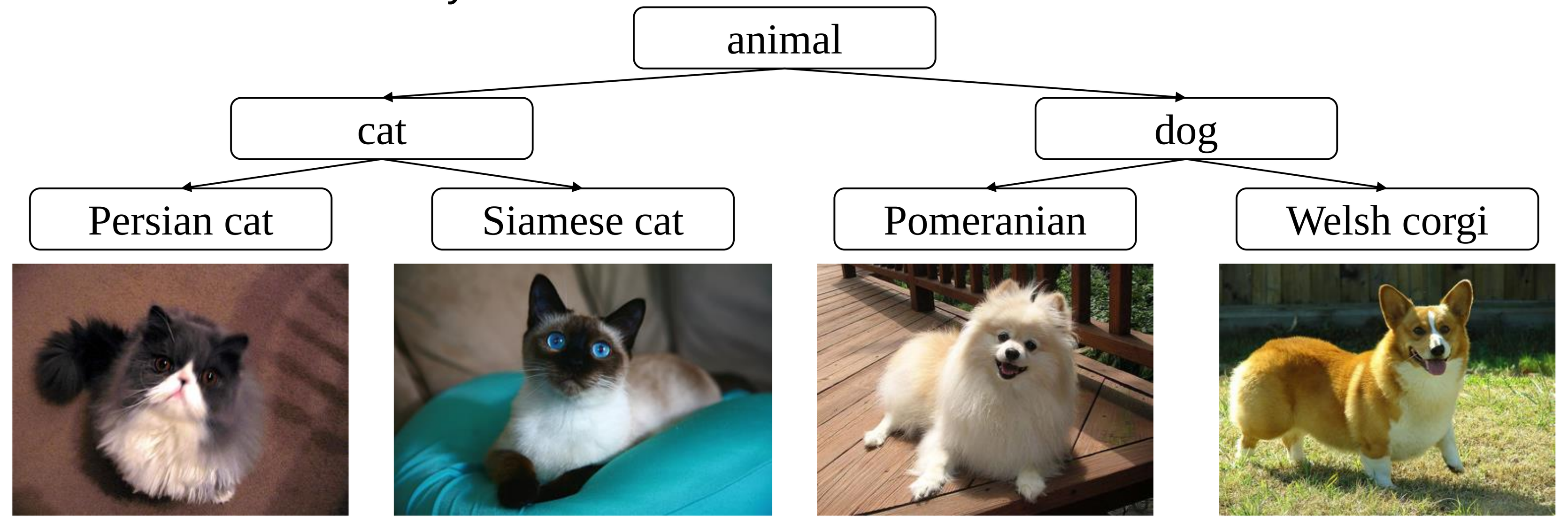
1 Why Hierarchical Novelty Detection?

- ❑ Conventional novelty detection framework does not provide more information than "novelty" of an object.
  - ➢ Our hierarchical novelty detection framework aims to find the most specific class label of *any* data on the hierarchical taxonomy built with known labels.
  - ➢ e.g.,

Test image:

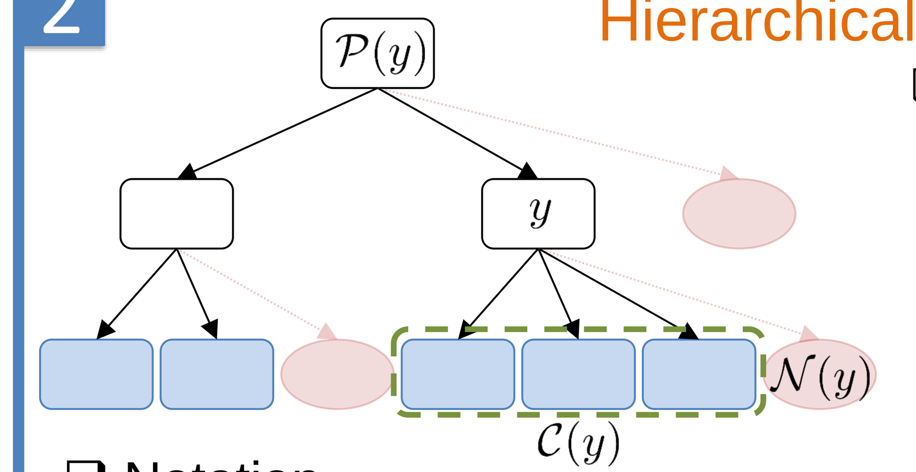
True label:	Siamese cat	Angora cat	Dachshund	Pika
Prior works:	Siamese cat	novel	novel	novel
Ours:	Siamese cat	novel cat	novel dog	novel animal

  - ➢ When the hierarchy of known labels is



- ❑ C.f., Generalized zero-shot learning requires specific semantic information (attributes, word vector, …) of all test classes.
  - ➢ To classify "Angora cat," we need semantic information of "Angora cat."
    - e.g., "Angora cat" = {"silky fur," "domestic," …}

2 Hierarchical Taxonomy



- ❑ Class types
  - ➢ Known leaf class
    - Seen during training
  - ➢ Super class
    - Ancestor of leaf classes, also known
  - ➢ Novel class
    - Unseen during training
    - Expected prediction is the closest super class in the taxonomy in our task

❑ Notation

$\mathcal{T}$	Taxonomy of known classes	$\mathcal{L}(\mathcal{T})$	Set of all known leaf classes
$\mathcal{P}(y)$	Set of parents	$\mathcal{N}(y)$	Set of novel classes whose closest known class is y
$\mathcal{C}(y)$	Set of children	$\mathcal{O}(y)$	Set of known classes except y and its descendants
$\mathcal{A}(y)$	Set of ancestors including itself:		$\mathcal{A}(y)=\{y\}\cup\mathcal{P}(y)\cup\mathcal{P}(\mathcal{P}(y))\cup\dots$

References

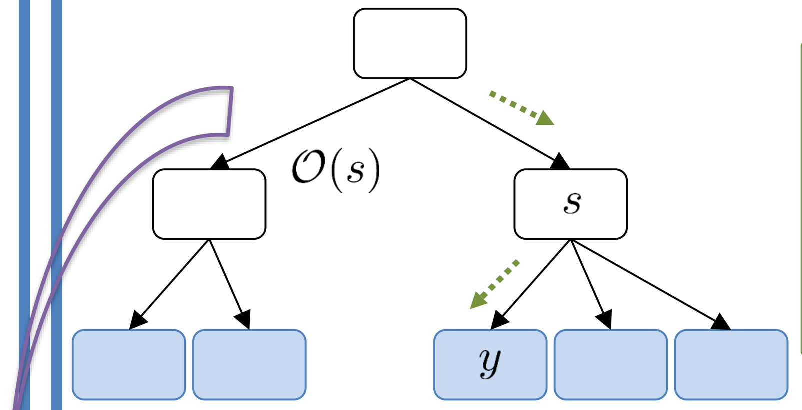
J. Deng, J. Krause, A. C. Berg, and L. Fei-Fei. "Hedging your bets: Optimizing accuracy-specificity trade-offs in large scale visual recognition." In CVPR, 2012.

Z. Akata, S. Reed, D. Walter, H. Lee, and B. Schiele. "Evaluation of output embeddings for fine-grained image classification." In CVPR, 2015.

3 Approach

a. Top-down (TD) method

❑ Recursively classify until arriving at a known leaf class or unconfident



Classification rule: $\hat{y} = \begin{cases} \arg\max_{y'} Pr(y'|x,s;\theta_s) & \text{if confident,} \\ \mathcal{N}(s) & \text{otherwise,} \end{cases}$

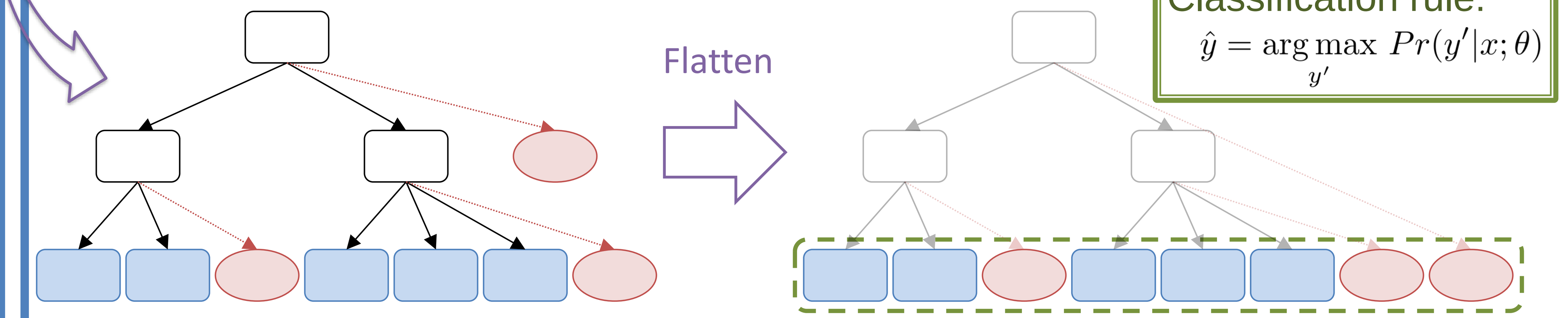
Classification is confident if $D_{KL}(U(\cdot|s)\,\|\,Pr(\cdot|x,s;\theta_s)) \geq \lambda_s$

Training objective: $\min_\theta \mathbb{E}_{Pr(x,y|s)}[-\log Pr(y|x,s;\theta_s)] + \mathbb{E}_{Pr(x,y|\mathcal{O}(s))}[D_{KL}(U(\cdot|s)\,\|\,Pr(\cdot|x,s;\theta_s))]$

b. Flatten method

❑ Represent all probabilities of known leaf and novel classes in a single vector

Add virtual novel classes



Classification rule: $\hat{y} = \arg\max_{y'} Pr(y'|x;\theta)$

b.i. Data relabeling

❑ Fill novel classes by relabeling

Training objective:

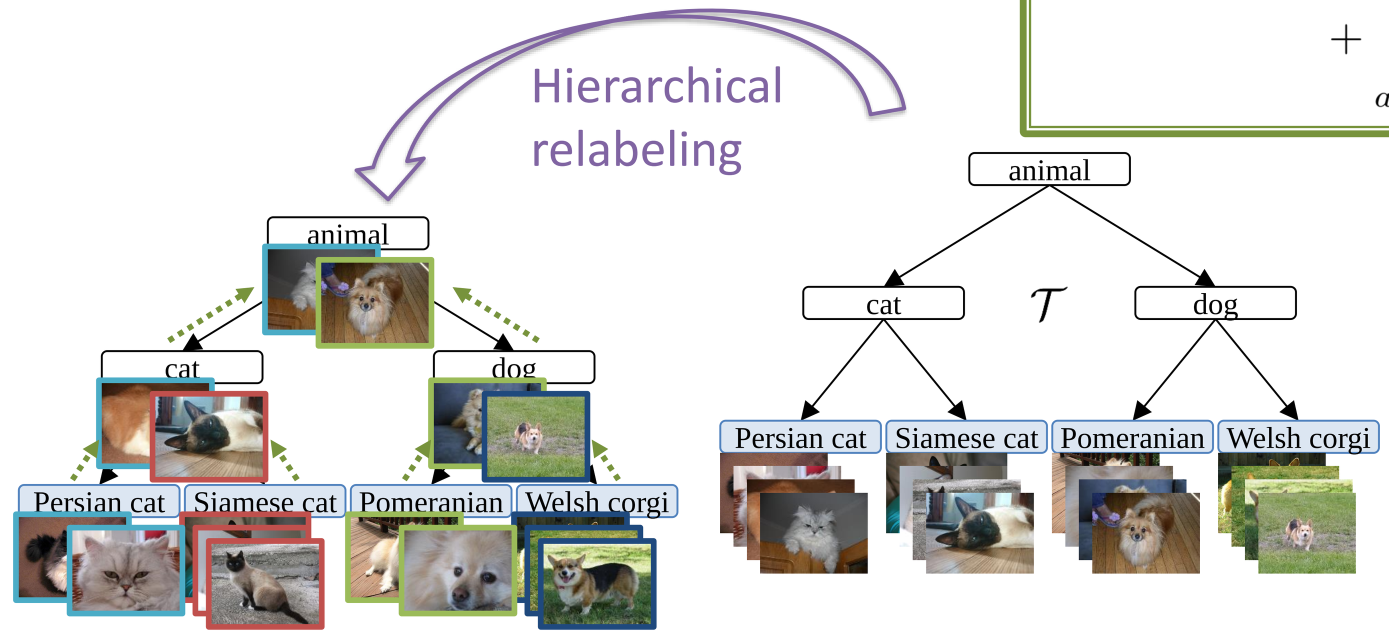
$$\min_\theta \mathbb{E}_{Pr(x,y)}[-\log Pr(y|x;\theta_{\mathcal{T}})]$$

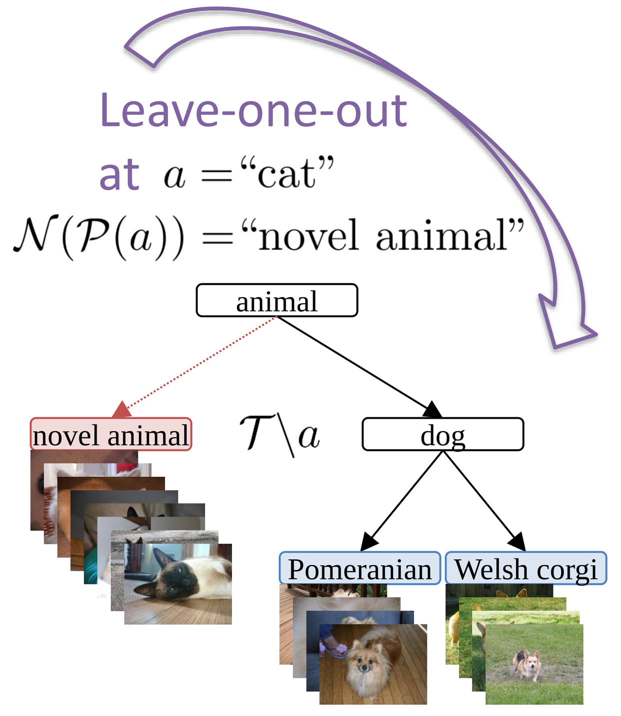
b.ii. Leave-one-out (LOO) strategy

❑ Generate deficient taxonomies

Training objective:

$$\min_\theta \mathbb{E}_{Pr(x,y)}\Big[-\log Pr(y|x;\theta_{\mathcal{L}(\mathcal{T})}) + \sum_{a\in\mathcal{A}(y)} -\log Pr(\mathcal{N}(\mathcal{P}(a))|x;\theta_{\mathcal{T}\setminus a})\Big]$$





c. Combined method (TD+LOO)

❑ Feed the output of TD into LOO
  - ➢ Top-down method leverages hierarchical structure information
  - ➢ Flatten method avoids error aggregation over hierarchy

4 Experimental Results

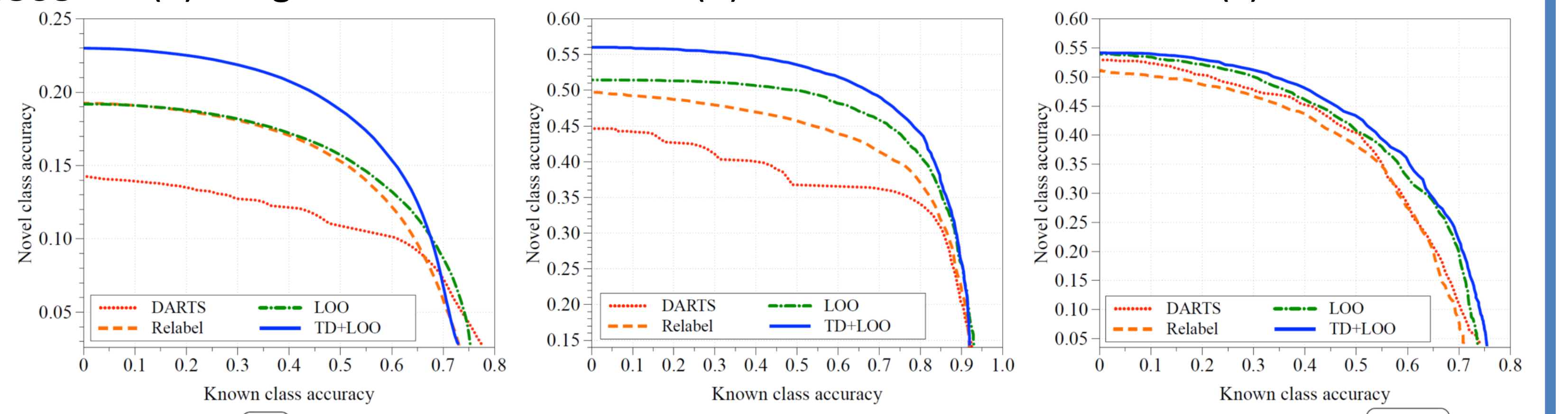
a. Hierarchical novelty detection

- ❑ Compared algorithms
  - ➢ Baseline: DARTS (Deng et al., 2012)
  - ➢ Ours: Relabel, LOO, TD+LOO
- ❑ Datasets
  - ➢ ImageNet: 1k known, 16k novel classes
  - ➢ AwA2: 40 known, 10 novel classes
  - ➢ CUB: 150 known, 50 novel classes
- ❑ Base model: ResNet-101
- ❑ Metrics
  - ➢ Novel class accuracy @ known class accuracy = 50%
  - ➢ Area under known-novel class accuracy curve

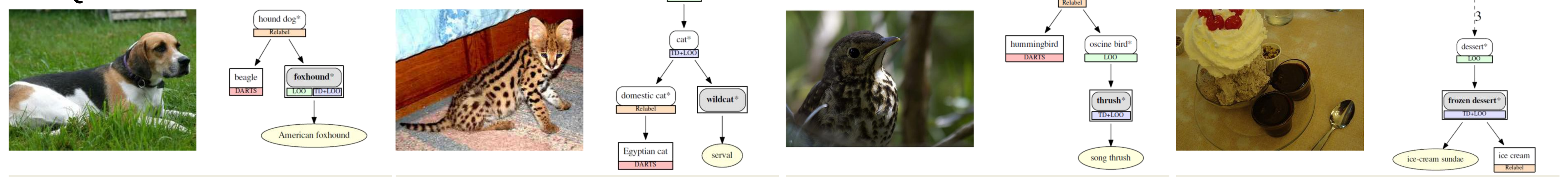
❑ Quantitative results

Method	ImageNet		AwA2		CUB	
	Novel	AUC	Novel	AUC	Novel	AUC
DARTS	10.89	8.83	36.75	35.14	40.42	30.07
Relabel	15.29	11.51	45.71	40.28	38.23	28.75
LOO	15.72	12.00	50.00	43.63	40.78	31.92
TD+LOO	**18.78**	**13.98**	**53.57**	**46.77**	**43.29**	**33.16**

(a) ImageNet (b) AwA2 (c) CUB



❑ Qualitative results



Novel class: American foxhound			
Method	ε	A	Word
GT			foxhound
DARTS	2	N	beagle
Relabel	1	Y	hound dog
LOO	0	Y	foxhound
TD+LOO	0	Y	foxhound

Novel class: serval			
Method	ε	A	Word
GT			wildcat
DARTS	3	N	Egyptian cat
Relabel	2	N	domestic cat
LOO	2	Y	feline
TD+LOO	1	Y	cat

Novel class: song thrush			
Method	ε	A	Word
GT			thrush
DARTS	3	N	hummingbird
Relabel	2	Y	bird
LOO	1	Y	oscine bird
TD+LOO	0	Y	thrush

Novel class: ice-cream sundae			
Method	ε	A	Word
GT			frozen dessert
DARTS	4	Y	food, nutrient
Relabel	1	N	ice cream
LOO	1	Y	dessert
TD+LOO	0	Y	frozen dessert

b. Generalized zero-shot learning

- ❑ Semantic embeddings
  - ➢ Attributes, Word vector, Hierarchical embedding
- ❑ Compared hierarchical embeddings
  - ➢ Baseline: Path (Akata et al., 2015)
    - Distance between classes on hierarchy
  - ➢ Ours: Top-down (TD)
    - Expected output of our top-down model
- ❑ Datasets
  - ➢ AwA1,2: 40 known, 10 novel classes
  - ➢ CUB: 150 known, 50 novel classes
- ❑ Base model: ResNet-101
- ❑ Metrics
  - ➢ Unseen class accuracy (ZSL)
  - ➢ Area under seen-unseen curve (GZSL)

❑ Quantitative results

Embedding			AwA1		AwA2		CUB	
Att	Word	Hier	Unseen	AUC	Unseen	AUC	Unseen	AUC
√			65.29	50.02	63.87	51.27	50.05	23.60
	√		51.87	39.67	54.77	42.21	27.28	11.47
√	√		67.80	52.84	65.76	53.18	49.83	24.13
		Path	42.57	30.58	44.34	33.44	24.22	8.38
√		Path	67.09	51.45	66.58	53.50	50.25	23.70
	√	Path	52.89	40.66	55.28	42.86	27.72	11.65
√	√	Path	68.04	53.21	67.28	54.31	**50.87**	24.20
		TD	33.86	25.56	31.84	24.97	13.09	7.20
√		TD	66.13	54.66	66.86	57.49	50.17	30.31
	√	TD	56.14	46.28	59.67	49.39	29.05	16.73
√	√	TD	**69.23**	**57.67**	**68.80**	**59.24**	50.17	**30.31**

(a) AwA1 (b) AwA2 (c) CUB