

**Jiabo Huang**

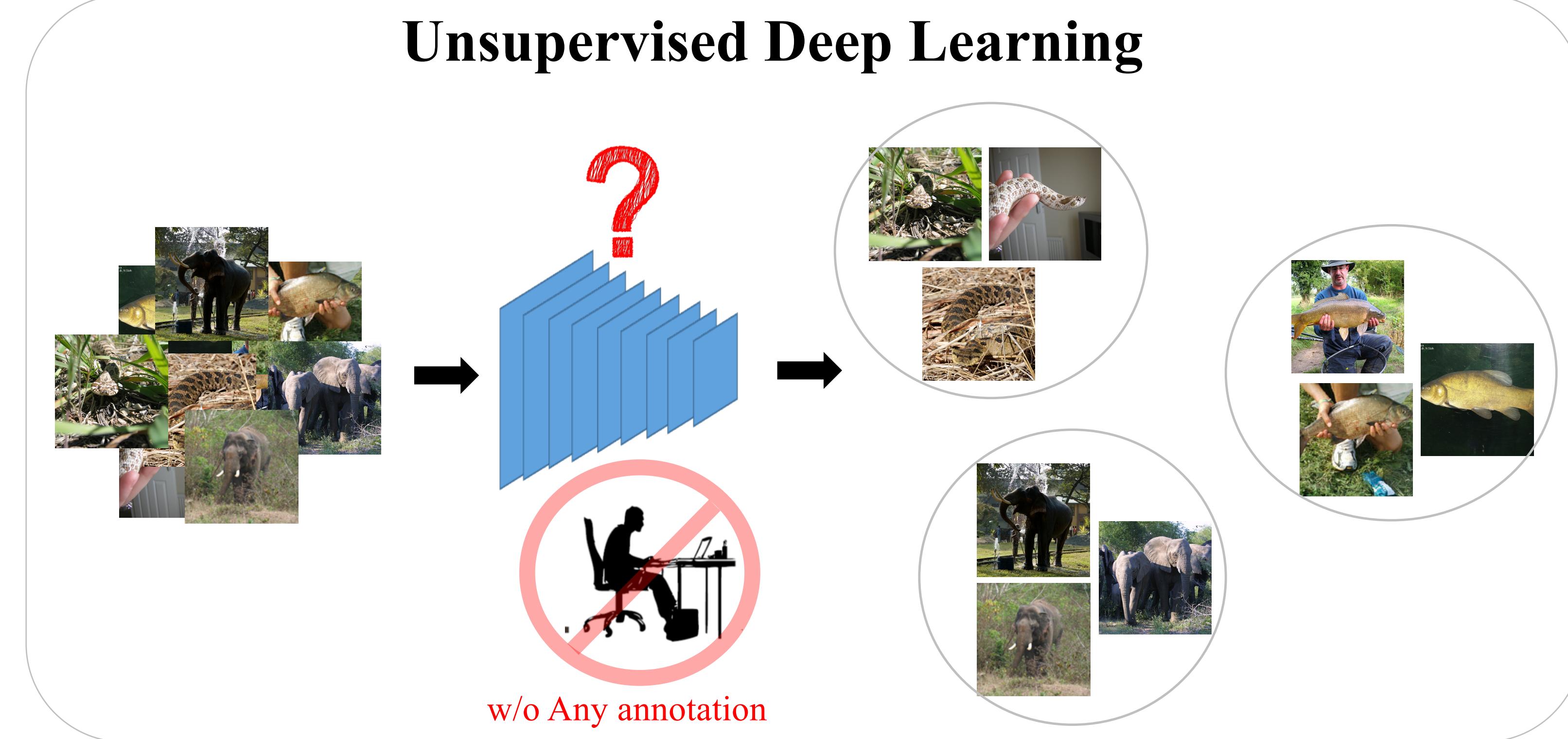
 Queen Mary University of London  
[jiabo.huang@qmul.ac.uk](mailto:jiabo.huang@qmul.ac.uk)
**Qi Dong**

 Queen Mary University of London  
[q.dong@qmul.ac.uk](mailto:q.dong@qmul.ac.uk)
**Shaogang Gong**

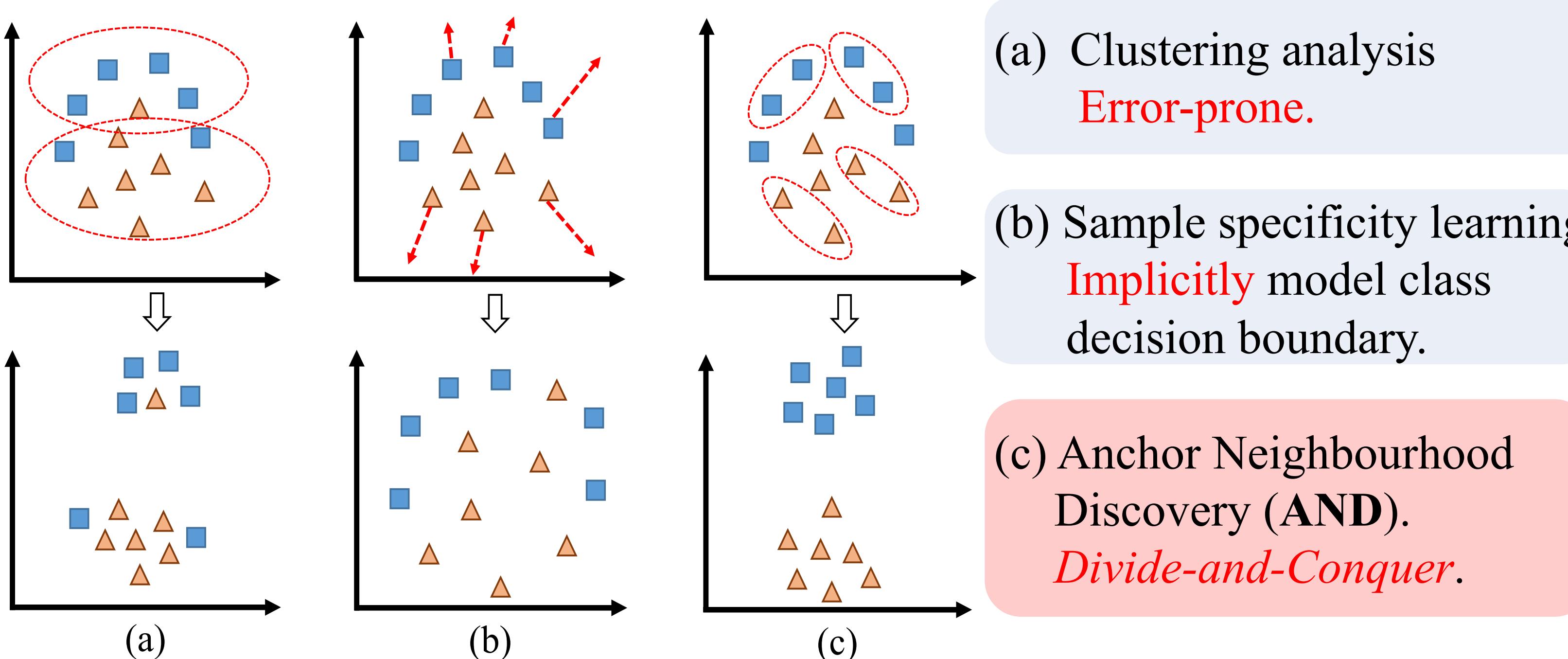
 Queen Mary University of London  
[s.gong@qmul.ac.uk](mailto:s.gong@qmul.ac.uk)
**Xiatian Zhu**

 Vision Semantic Ltd.  
[eddy.zhuxt@gmail.com](mailto:eddy.zhuxt@gmail.com)

## 1 Problem Definition

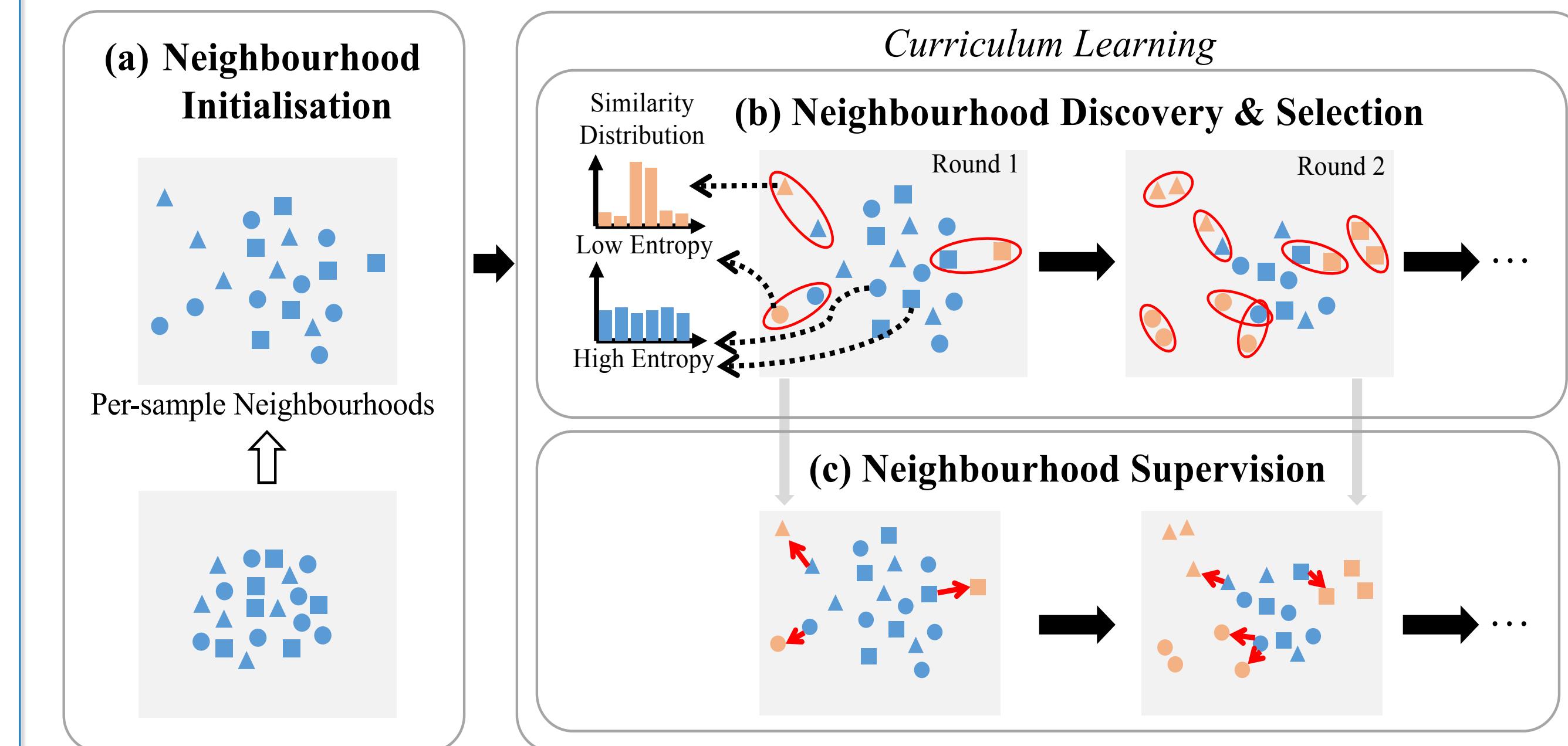


## 2 Motivation



## 3 AND Overview

- Anchor Neighbourhood Discovery



## 5 Neighbourhood Supervision

$$\mathcal{L}^r = \mathcal{L}_{inst} + \mathcal{L}_{AN}$$

- Probability Distribution for Similarity

$$x_i \text{ and } x_j \text{ are drawn from the same class}$$

$$p_{i,j} = \frac{\exp(x_i^T \tilde{x}_j / \tau)}{\sum_{k=1}^N \exp(x_i^T \tilde{x}_k / \tau)} \quad \tilde{x}_i = (1 - \eta) \cdot \tilde{x}_i + \eta \cdot x_i$$

- Unsupervised Neighborhood Supervision

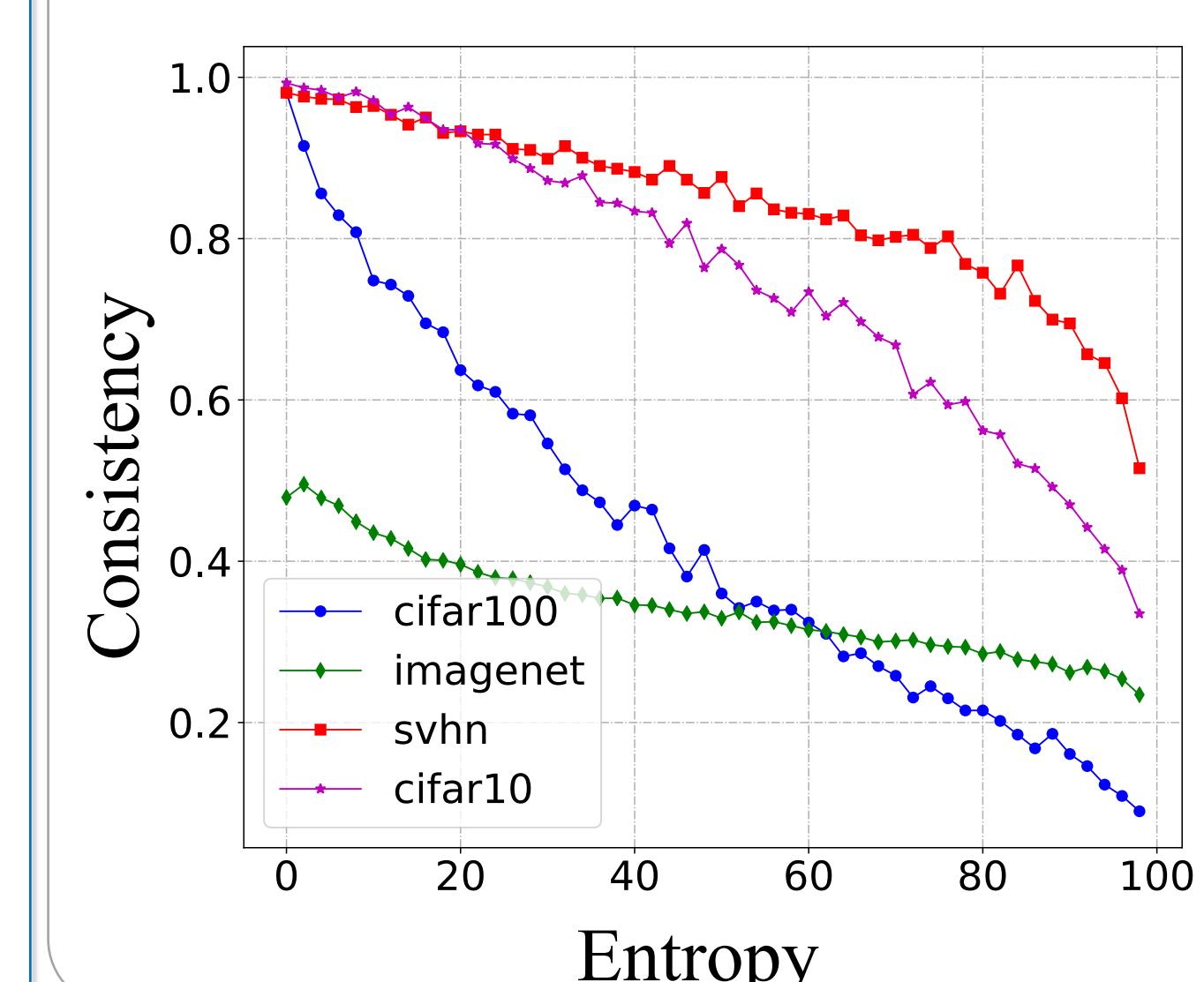
Objective loss function for  $r$ -th round

$$\mathcal{L}_{inst} = - \sum_{i \in B_{inst}^r} \log(p_{i,i}) \quad \mathcal{L}_{AN} = - \sum_{i \in B_{AN}^r} \log(\sum_{j \in N_k(x_i)} p_{i,j})$$

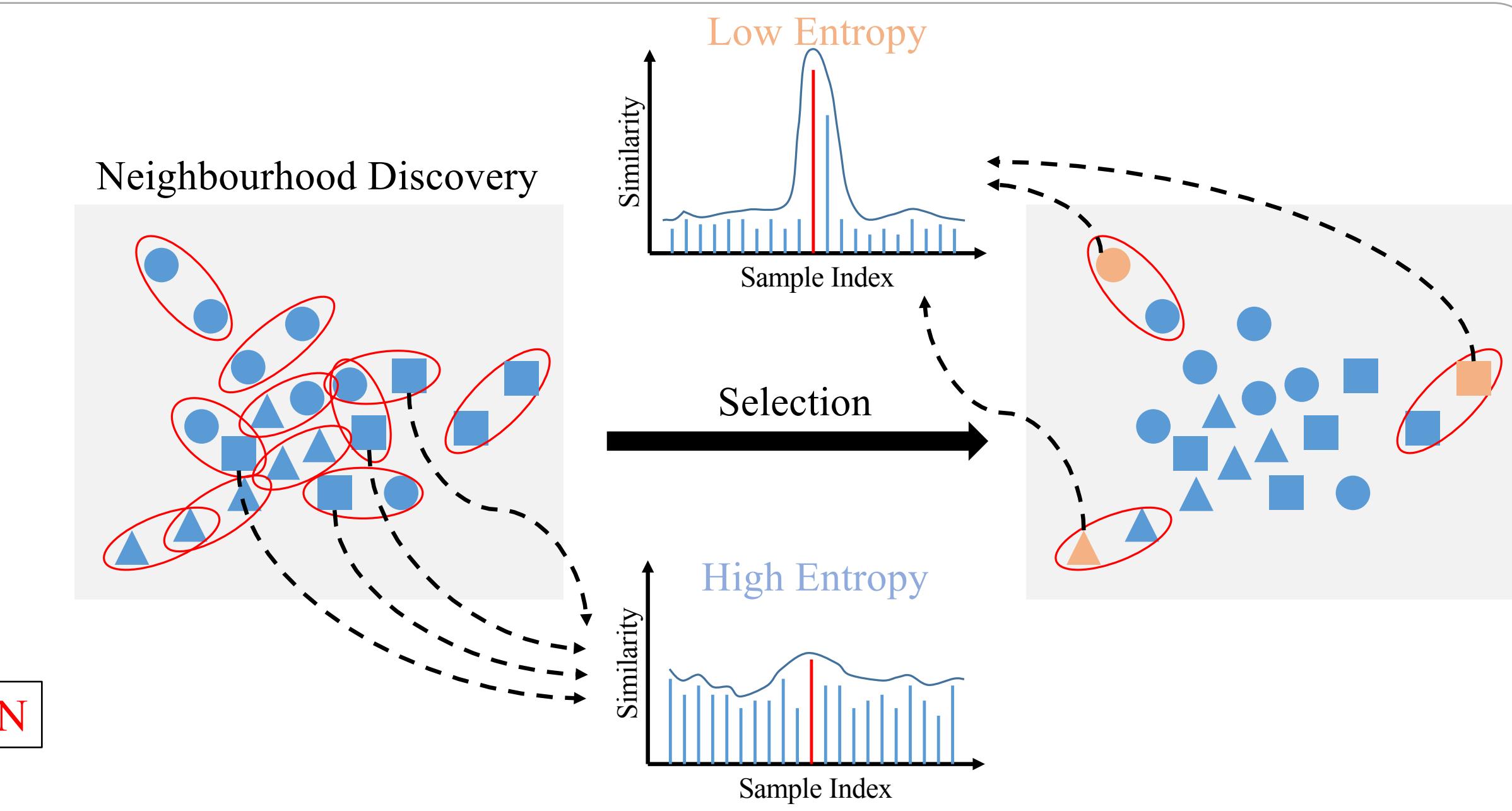
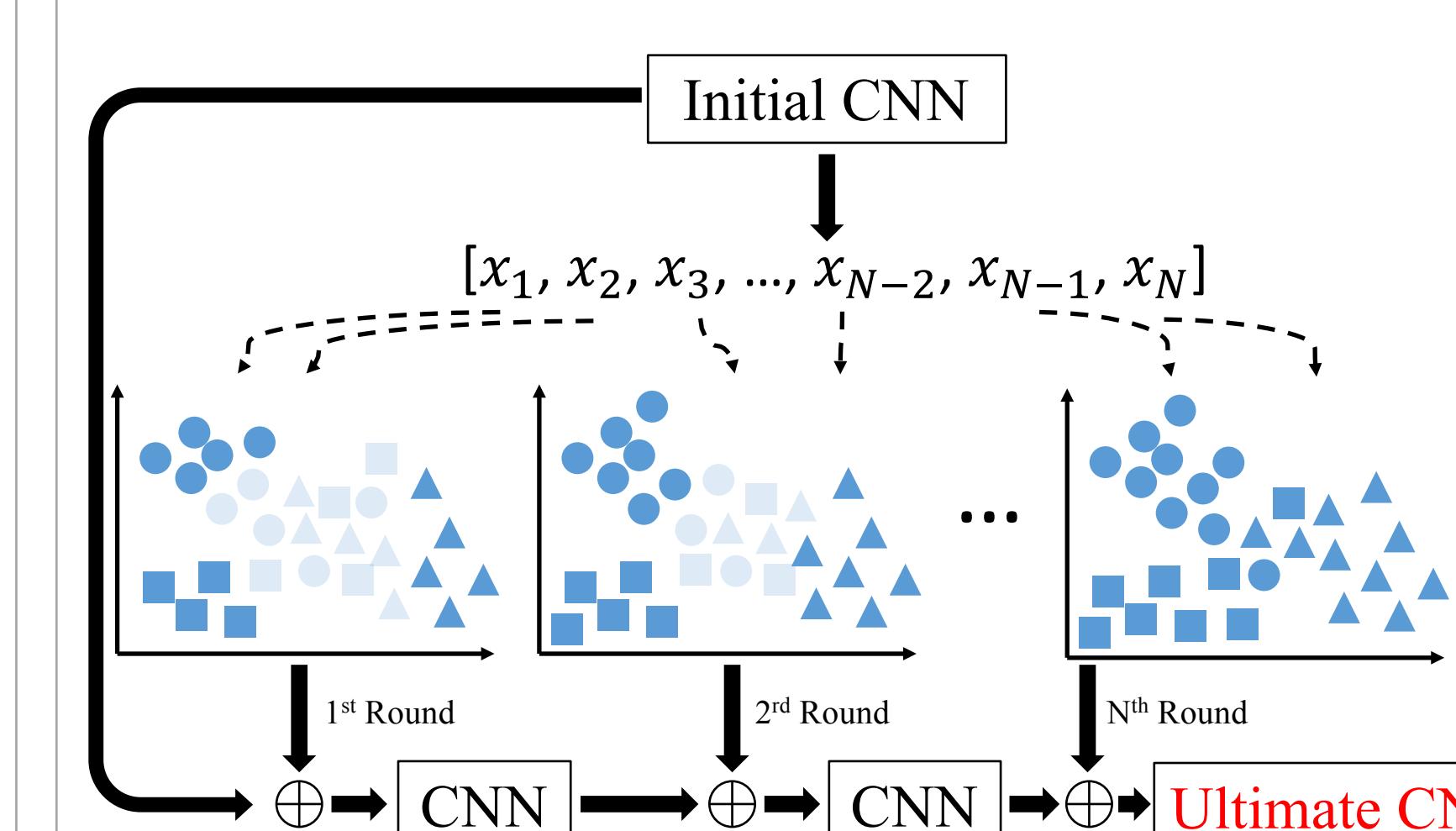
$N_k(x_i)$ :  $k$ -nearest neighbours of sample  $x_i$

## 4 Neighbourhood Discovery & Selection

- Entropy v.s. Consistency



- Curriculum Learning



## 6 Experiments

- Task1: Small scale image classification

Dataset	CIFAR10	CIFAR100	SVHN
Classifier/Feature			
DeepCluster [1]	62.3	22.7	84.9
Instance [2]	60.3	32.7	79.8
<b>AND (Ours)</b>	<b>74.8</b>	<b>41.5</b>	<b>90.9</b>
Classifier/Feature			
DeepCluster [1]	77.9	41.9	92.0
Instance [2]	70.1	39.4	89.3
<b>AND (Ours)</b>	<b>77.6</b>	<b>47.9</b>	<b>93.7</b>

### References

- [1] Caron, M., et al. Deep clustering for unsupervised learning of visual features. Proc. ECCV, 2018.
- [2] Wu, Z., et al. Unsupervised feature learning via non-parametric instance discrimination. Proc. CVPR, 2018.
- [3] Zhang, R., et al. Split-brain autoencoders: Unsupervised learning by cross-channel prediction. Proc. CVPR, 2017.

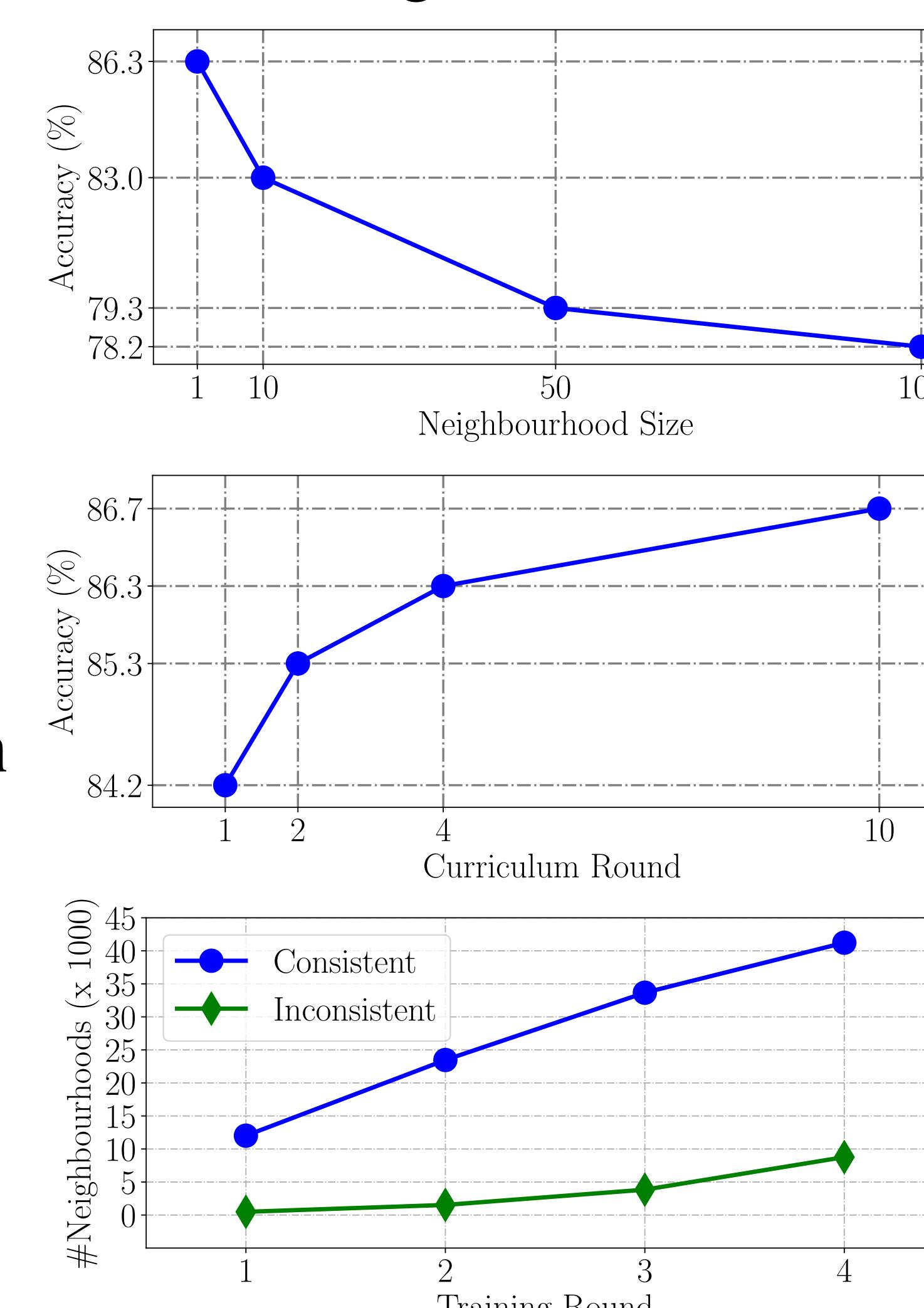
### Task2: Large scale image classification

Classifier	Linear Classifier			$k$ NN
Feature	conv3	conv4	conv5	FC
Split-Brain [3]	35.4	35.2	32.8	11.8
DeepCluster [1]	<b>41.0</b>	<b>39.6</b>	<b>38.2</b>	<b>26.8</b>
Instance [2]	31.8	34.1	35.6	<b>31.3</b>
<b>AND (Ours)</b>	<b>35.9</b>	<b>39.7</b>	<b>37.9</b>	<b>31.3</b>

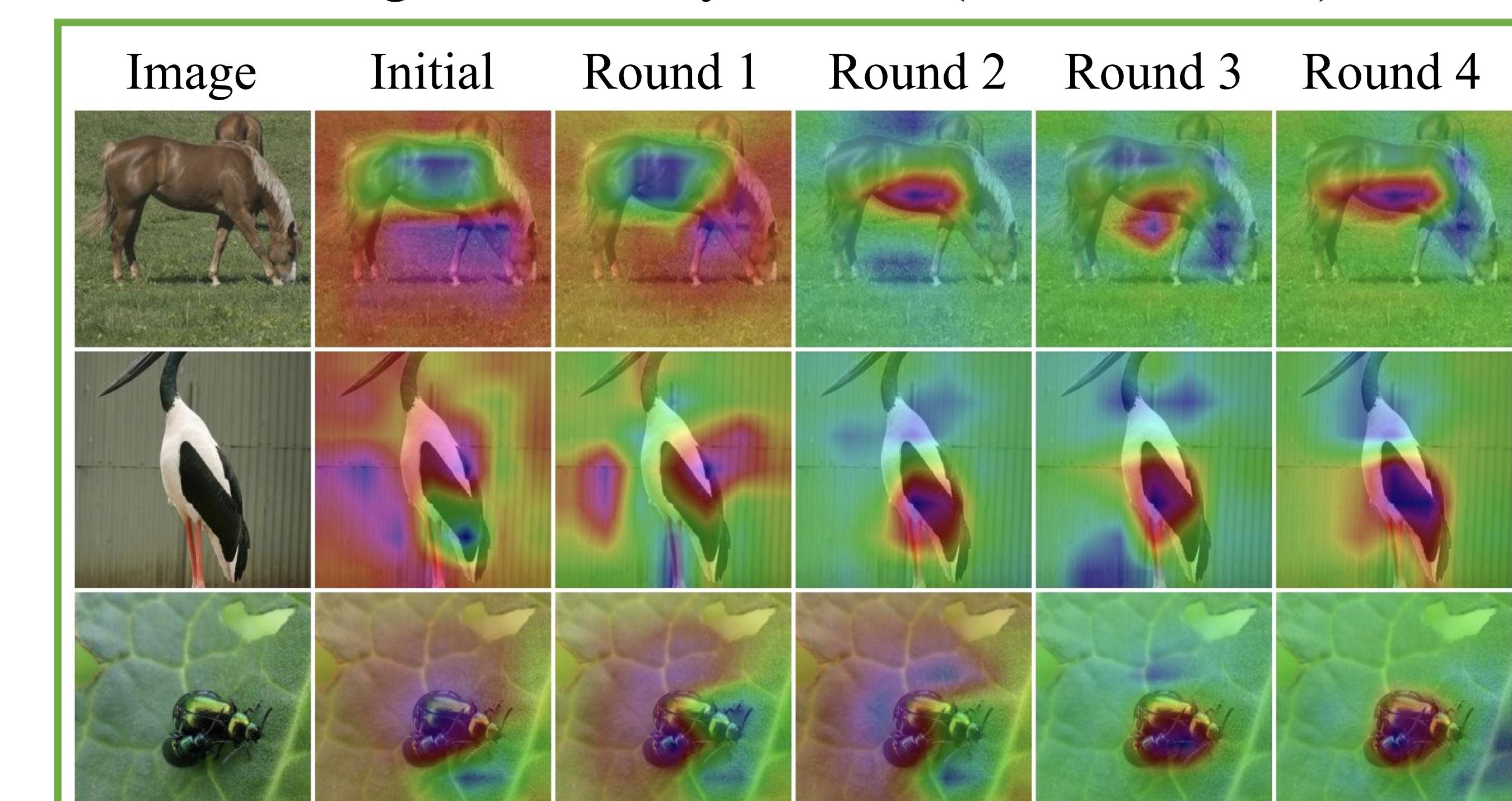
### Task3: Fine-grained image classification

Dataset	CUB200	Dogs
Instance [2]	11.6	27.0
<b>AND (Ours)</b>	<b>14.4</b>	<b>32.3</b>

### Training Robustness



### Learning attention dynamics (Success cases)



### Learning attention dynamics (A failure case)

