Uncertainty-Aware Hierarchical Refinement for Incremental Implicitly-Refined Classification

Jian Yang[1*] Kai Zhu[1,*,‡] Kecheng Zheng[2] Yang Cao[1,3,†]
1 University of Science and Technology of China 2 Ant Group
3Institute of Artificial Intelligence, Hefei Comprehensive National Science Center

1.Setting ——Incremental Implicitly-Refined Classification

Is incremental learning always a binary classification option that distinguishes between the old and new class?

existing incremental learning settings

1.Setting —— Incremental Implicitly-Refined Classification

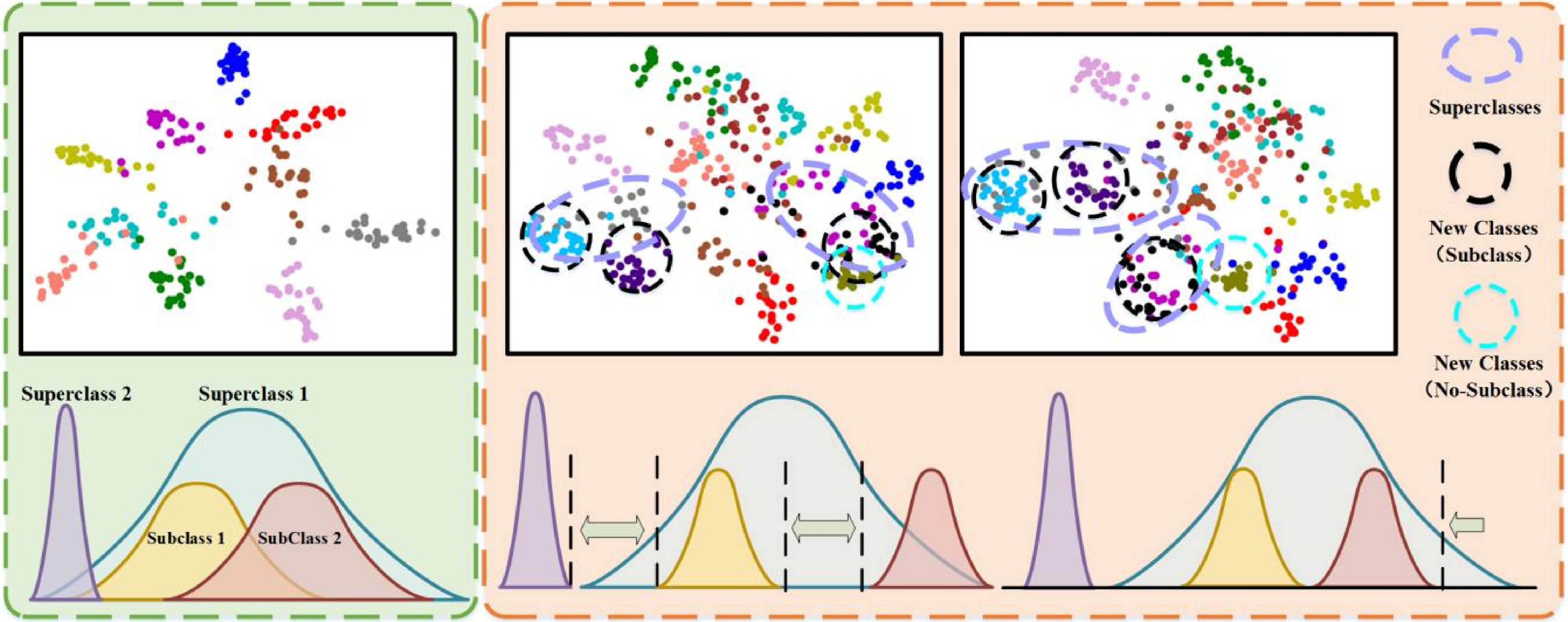
existing incremental learning settings

In real life, people's semantic understanding of the same instance may be gradually enriched as the learning process proceeds.

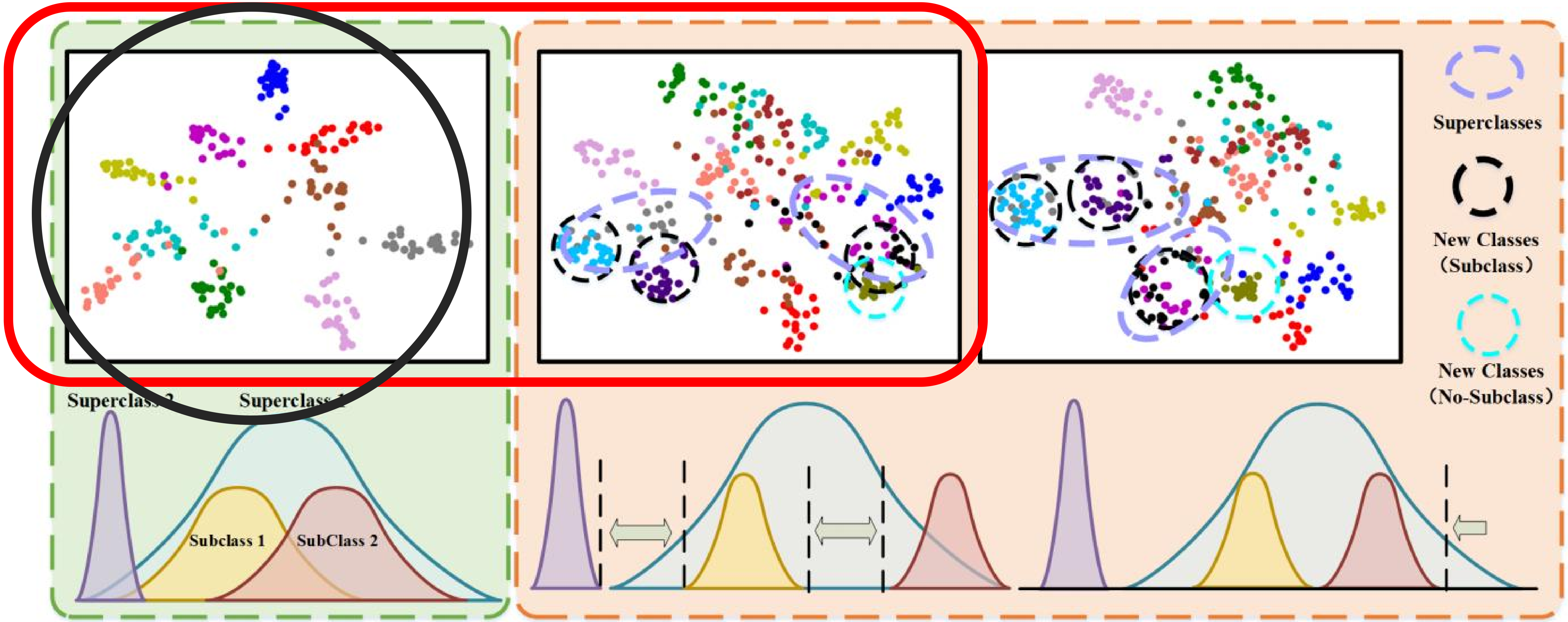
Incremental Implicitly-Refined Classification

2.Motivation

How to discern the semantic inheritance relationship in a hierarchical incremental scenery?

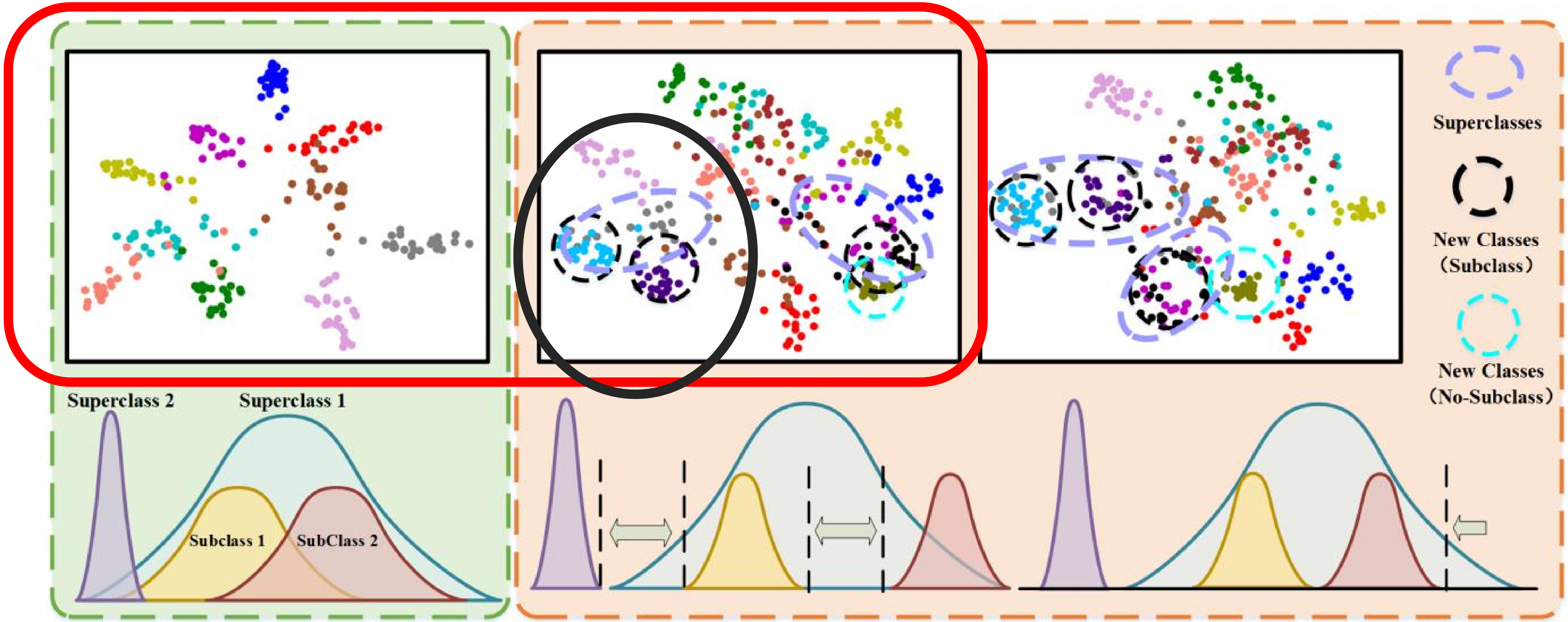


2.Motivation



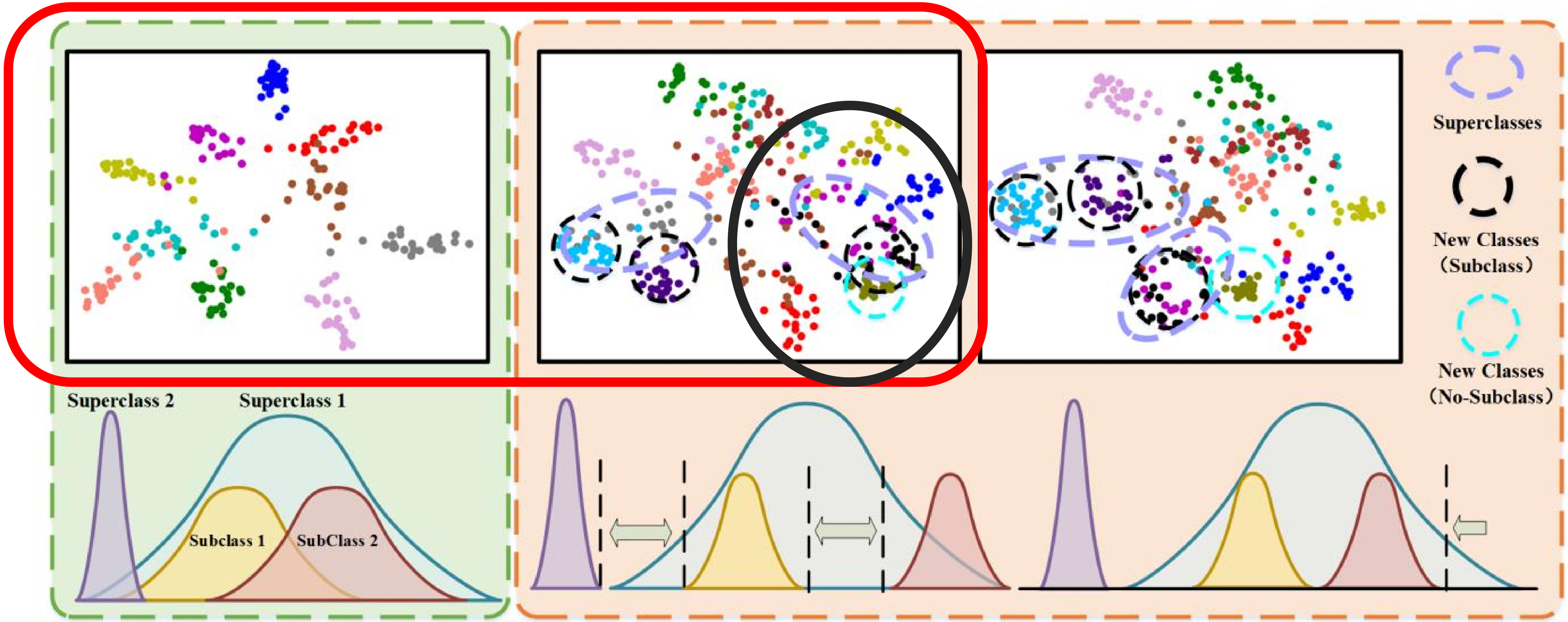
The probability distributions of classes with inheritance relationships show an obvious consistency in the initial phase.

2.Motivation



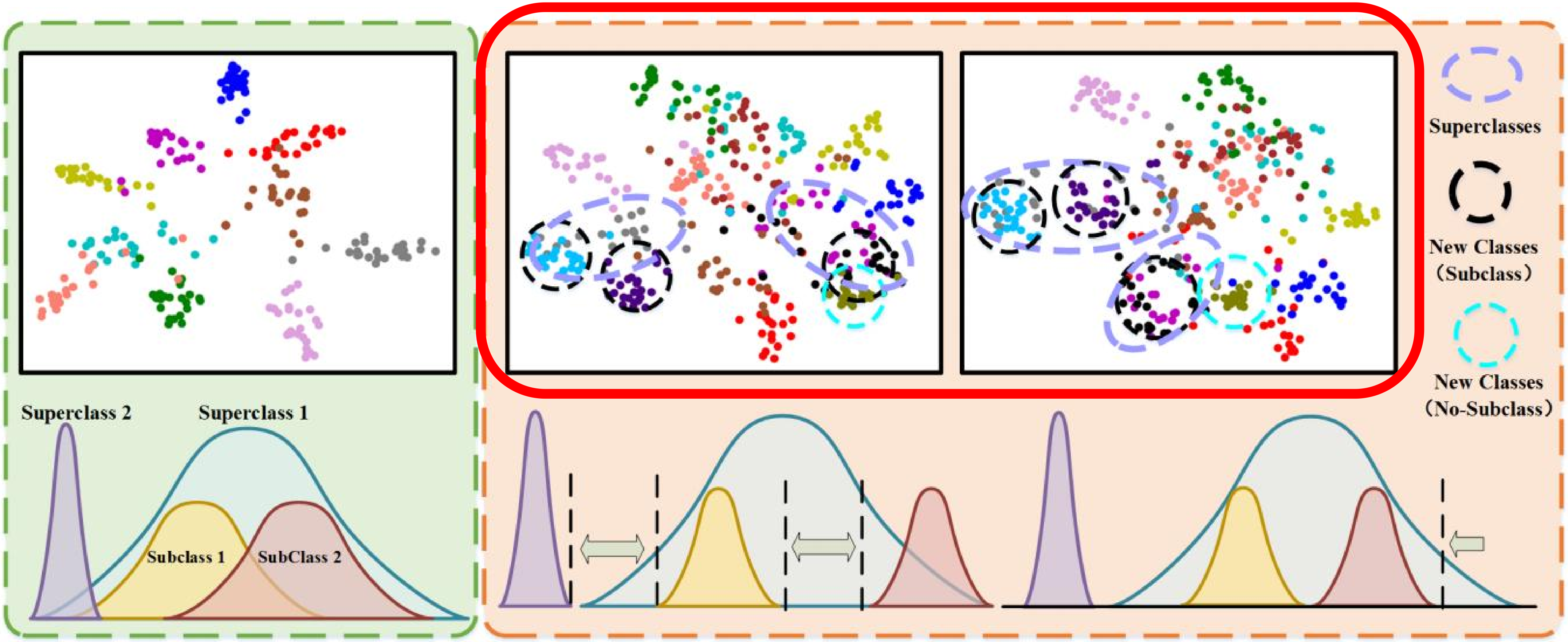
The incremental subclass inherited from a certain old class gradually outgroups under the supervision of new labels.

2.Motivation



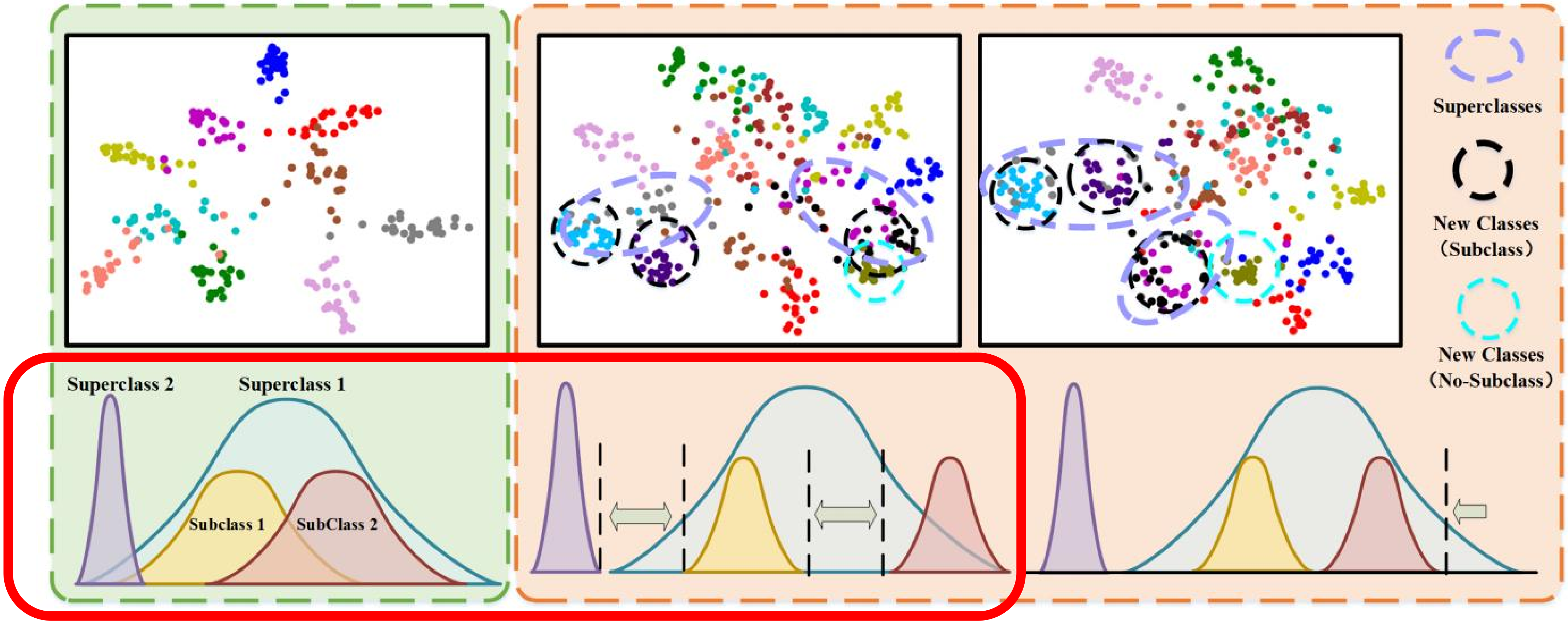
Some incremental classes inherit from none of the existing classes, leading to feature confusion.

2.Motivation



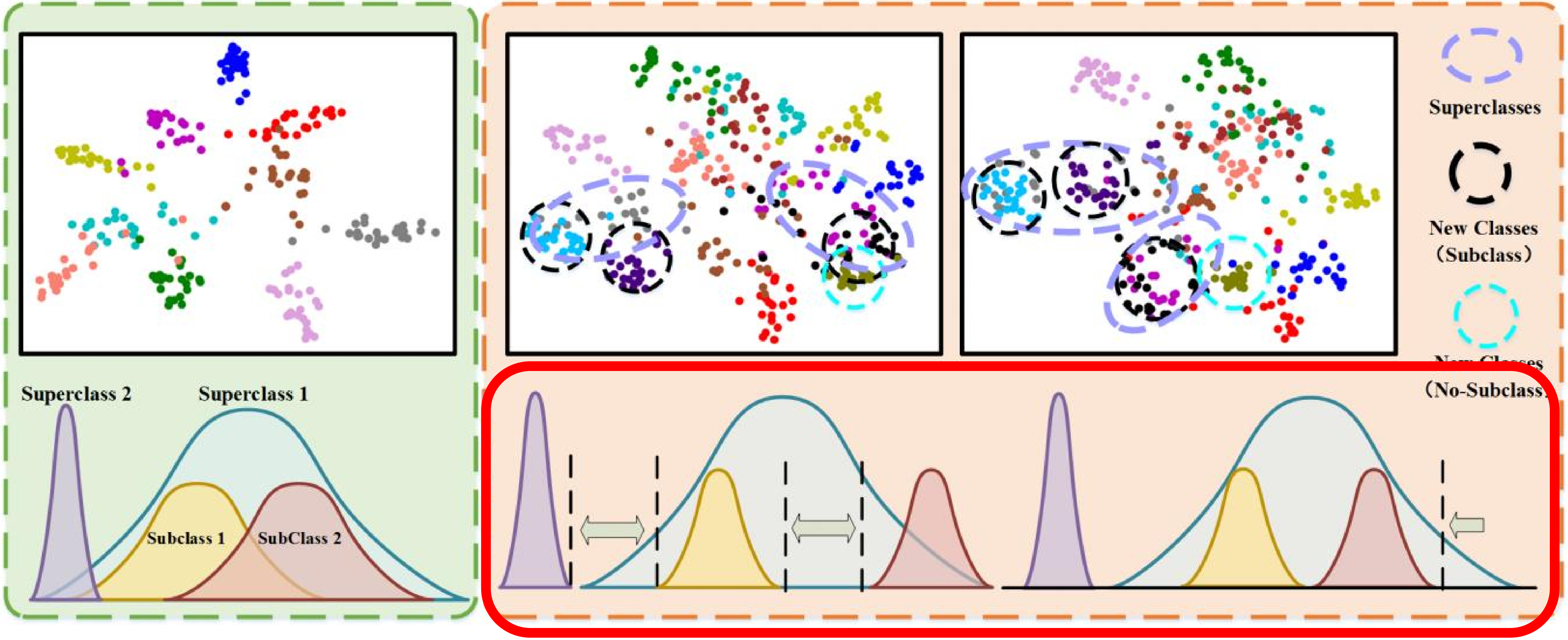
We propose an Uncertainty-Aware Hierarchical Refinement (UAHR) scheme.

2.Motivation



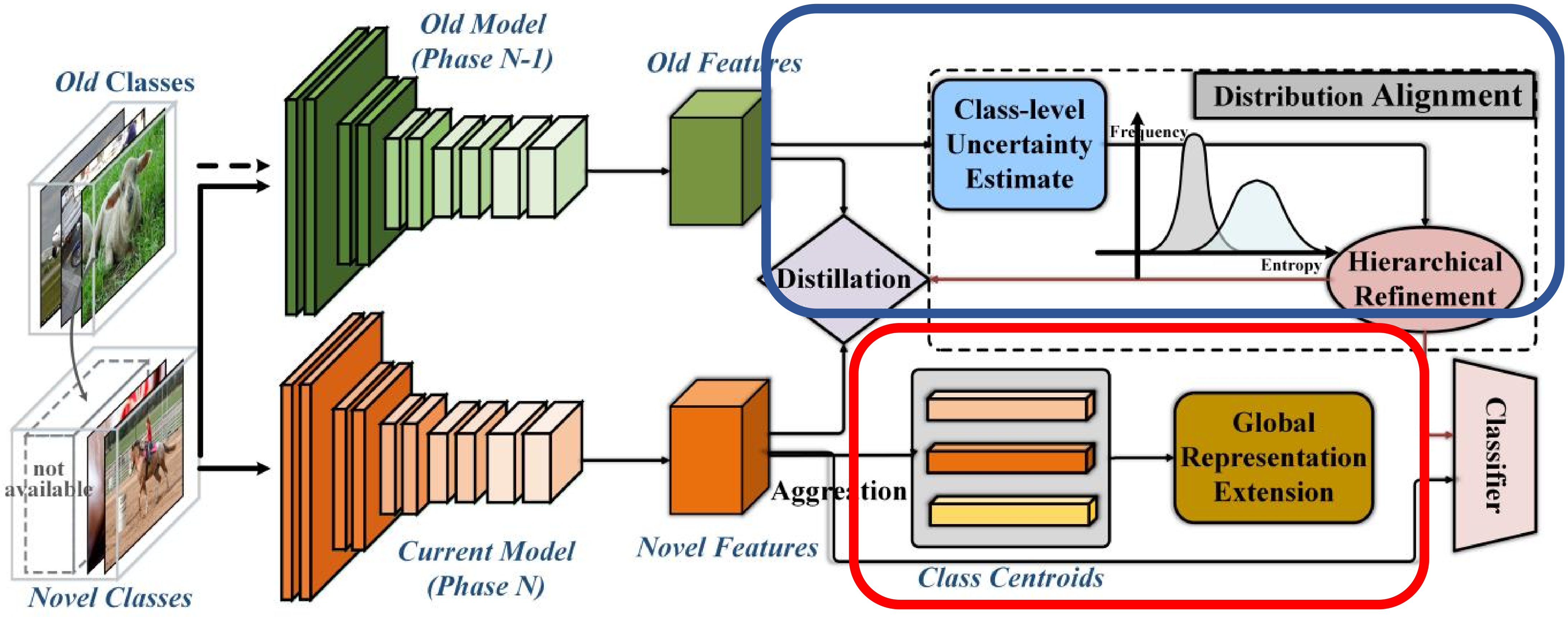
A global representation extension strategy is proposed to widen the distribution distance among all new classes in the embedding space, enhancing their discriminative properties..

2.Motivation



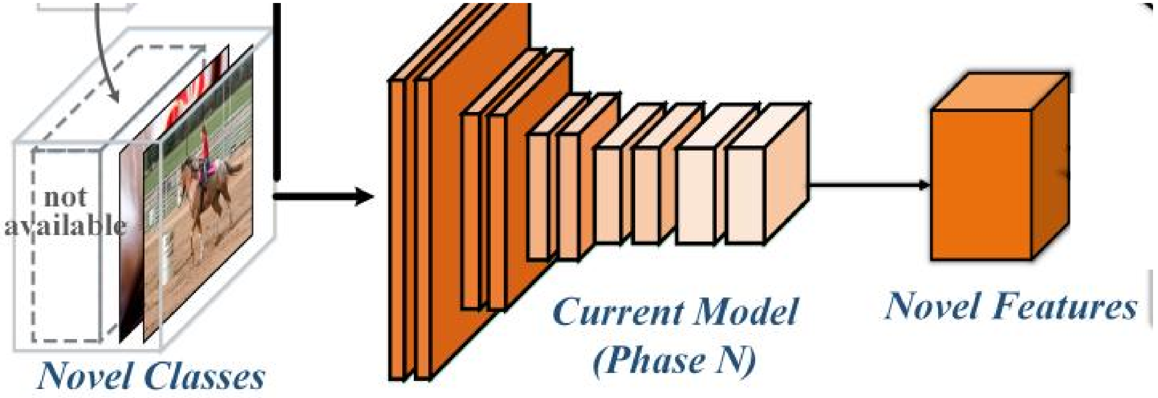
A hierarchical distribution alignment strategy is further proposed to correct the optimization of the shifting subclasses by aligning with the distribution of the whole superclass, ensuring the consistency of the hierarchical uncertainty.

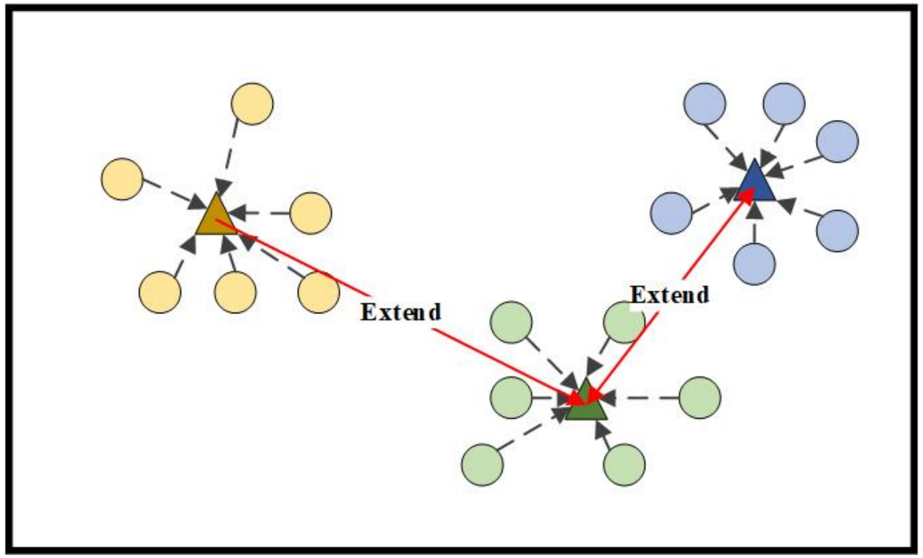
3. Method



Old Classes
Old Model
(Phase N-1)
Old Features
Distribution Alignment
Class-level
Uncertainty
Estimate
Frequency
Entropy
Hierarchical
Refinement
Distillation
not
available
Novel Classes
Current Model
(Phase N)
Novel Features
Aggregation
Global
Representation
Extension
Class Centroids
Classifier

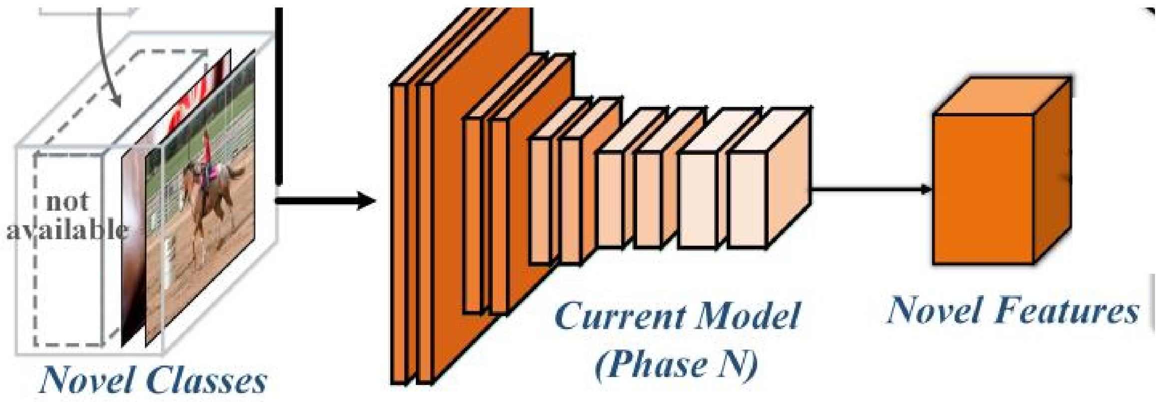
3. Method—— Global Representation Extension（GRE）

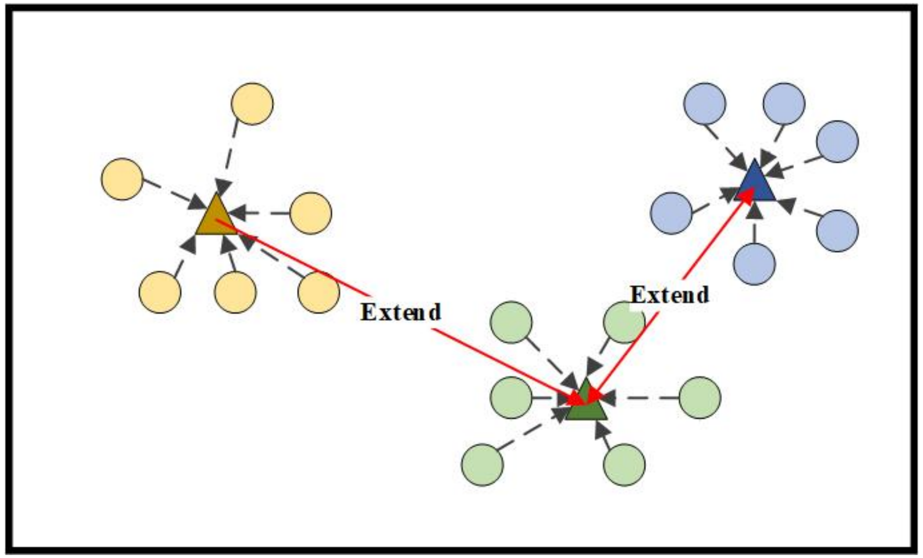




To maintain the stability of the feature space, we use the representation distance as a measure of uncertainty.

3. Method—— Global Representation Extension（GRE）

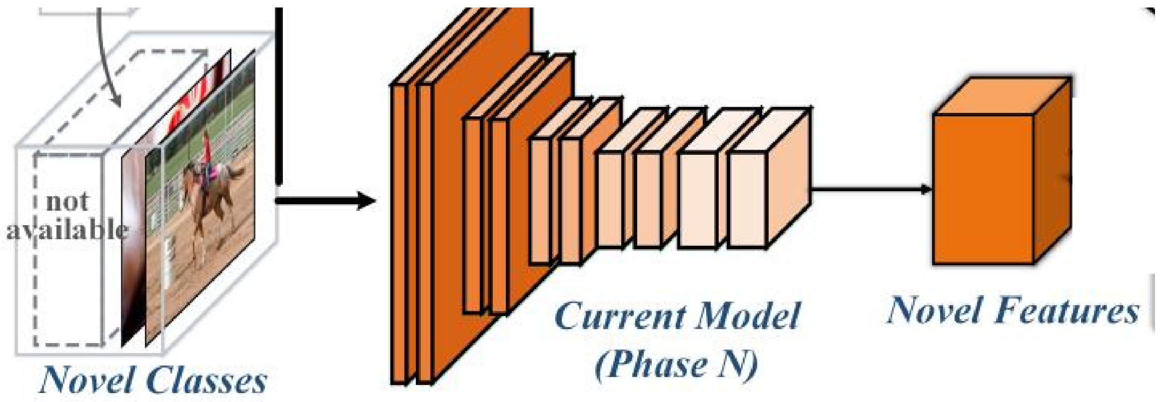


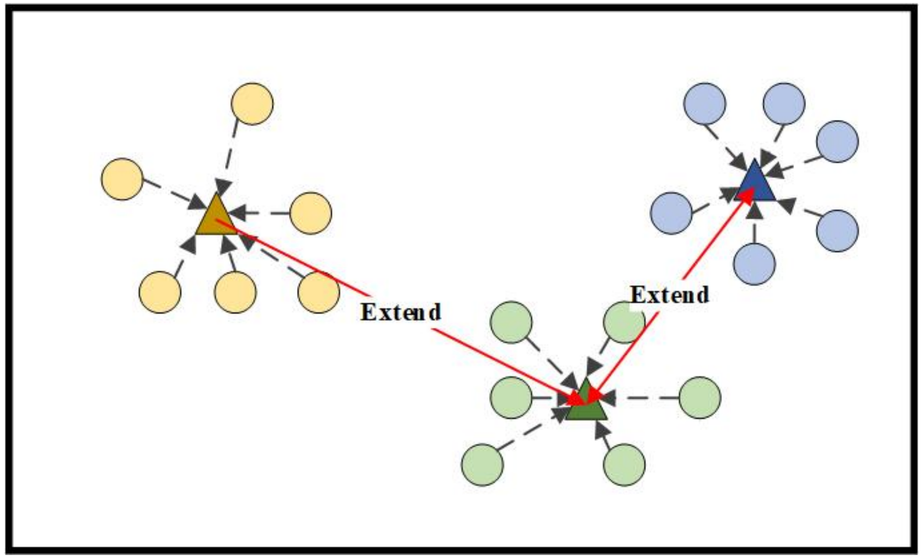


To maintain the stability of the feature space, we use the representation distance as a measure of uncertainty.

$$\mathcal{L}_{div} = \sum_{c=0}^{n_b} K(h_\theta(\boldsymbol{x})_c, h_\theta(\boldsymbol{x})_{j_{near}}) = \sum_{c=0}^{n_b} \exp \boxed{\left[\frac{-\frac{1}{n_d} \| h_\theta(\boldsymbol{x})_c - h_\theta(\boldsymbol{x})_j \|_2^2}{2\sigma^2} \right]}$$

3. Method—— Global Representation Extension（GRE）





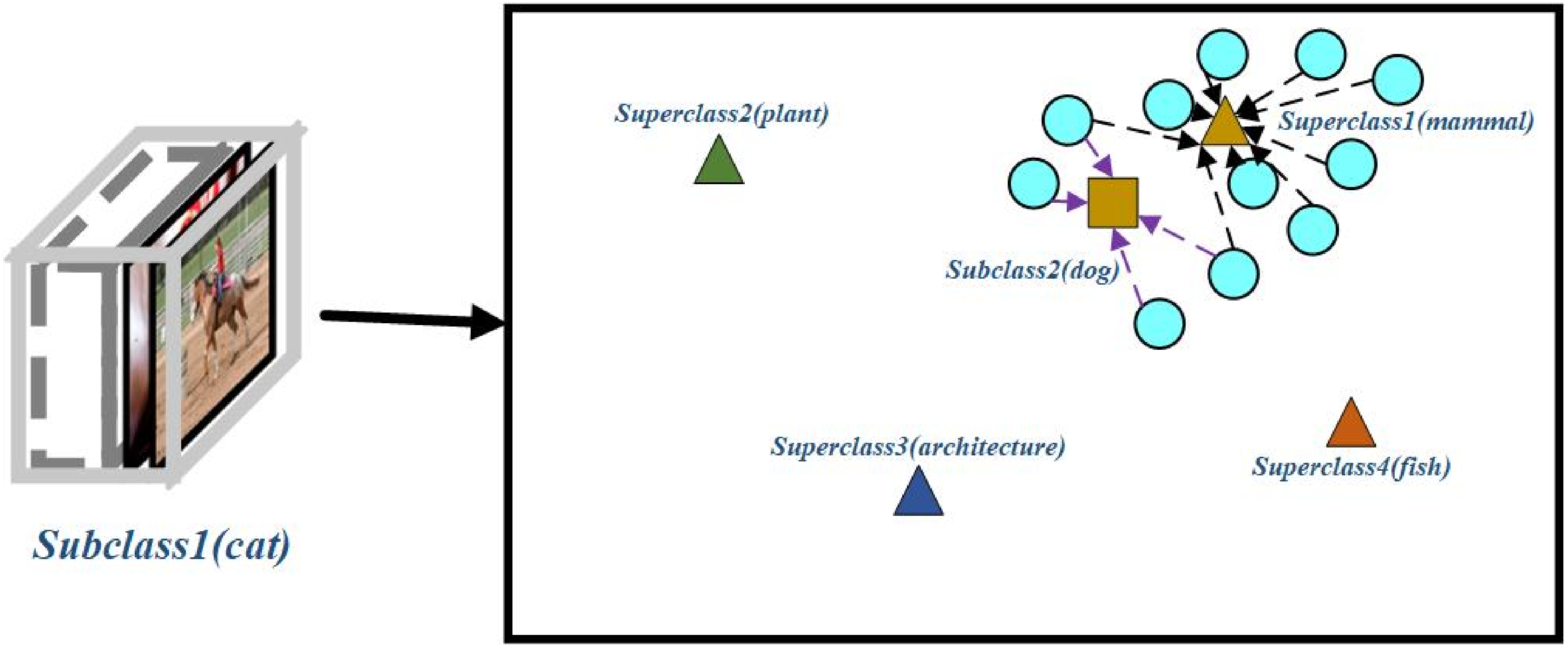
To maintain the stability of the feature space, we use the representation distance as a measure of uncertainty. And we adopt the BCEWithLogitsLoss to optimize the novel classes learning.

$$\mathcal{L}_{div} = \sum_{c=0}^{n_b} K(h_\theta(\boldsymbol{x})_c, h_\theta(\boldsymbol{x})_{j_{near}}) = \sum_{c=0}^{n_b} \exp\left[\frac{-\frac{1}{n_d}\|h_\theta(\boldsymbol{x})_c - h_\theta(\boldsymbol{x})_j\|_2^2}{2\sigma^2}\right]$$

$$\mathcal{L}_{cls} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\mathrm{sigmoid}(f_\theta(h_\theta(\boldsymbol{x}_i)))) + (1-y_i)\log(1-\mathrm{sigmoid}(f_\theta(h_\theta(\boldsymbol{x}_i))))]$$

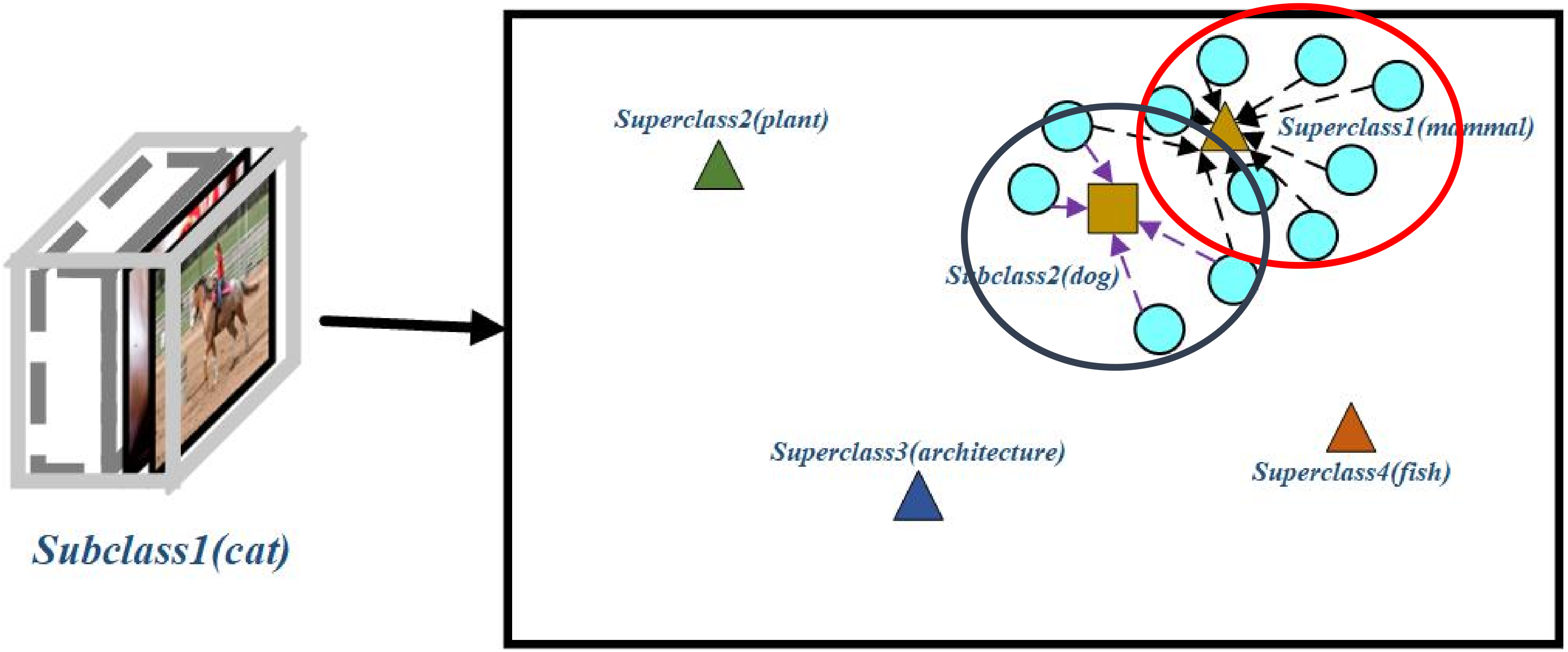
3. Method—— Hierarchical Distribution Alignment（HDA）

The label relationships can be inferred from the hierarchical uncertainty, which is calculated by counting the entropy of new samples on the old model across phases.



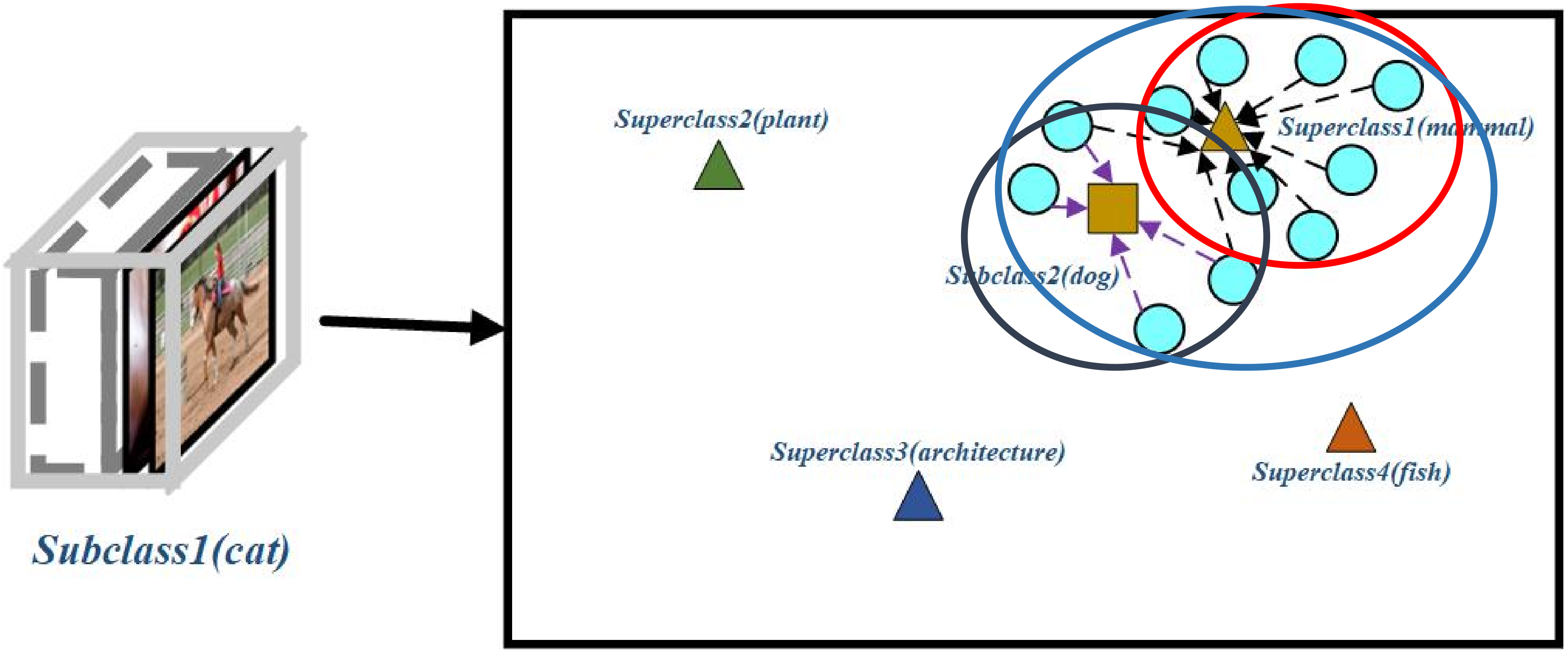
3. Method—— Hierarchical Distribution Alignment（HDA）

The label relationships can be inferred from the hierarchical uncertainty, which is calculated by counting the entropy of new samples on the old model across phases.



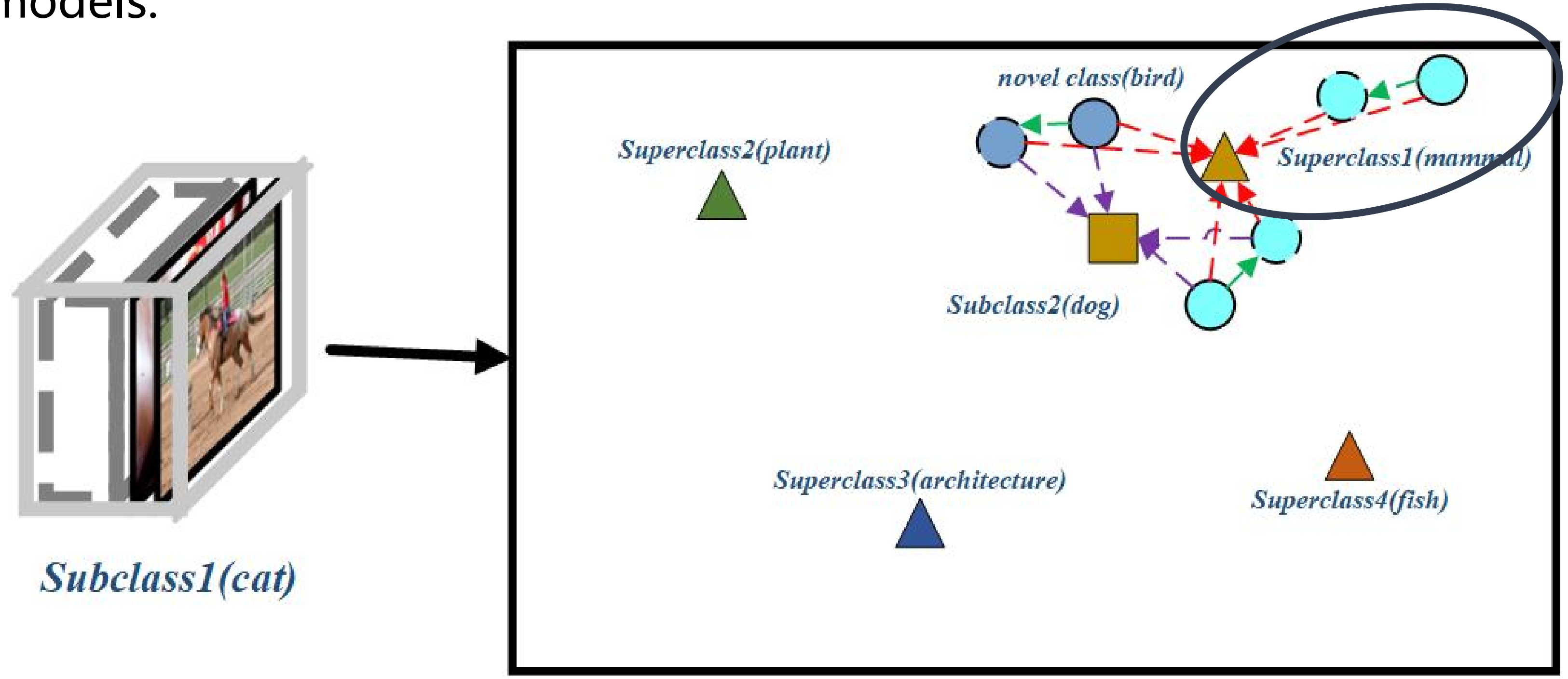
3. Method—— Hierarchical Distribution Alignment（HDA）

The label relationships can be inferred from the hierarchical uncertainty, which is calculated by counting the entropy of new samples on the old model across phases.



3. Method—— Hierarchical Distribution Alignment（HDA）

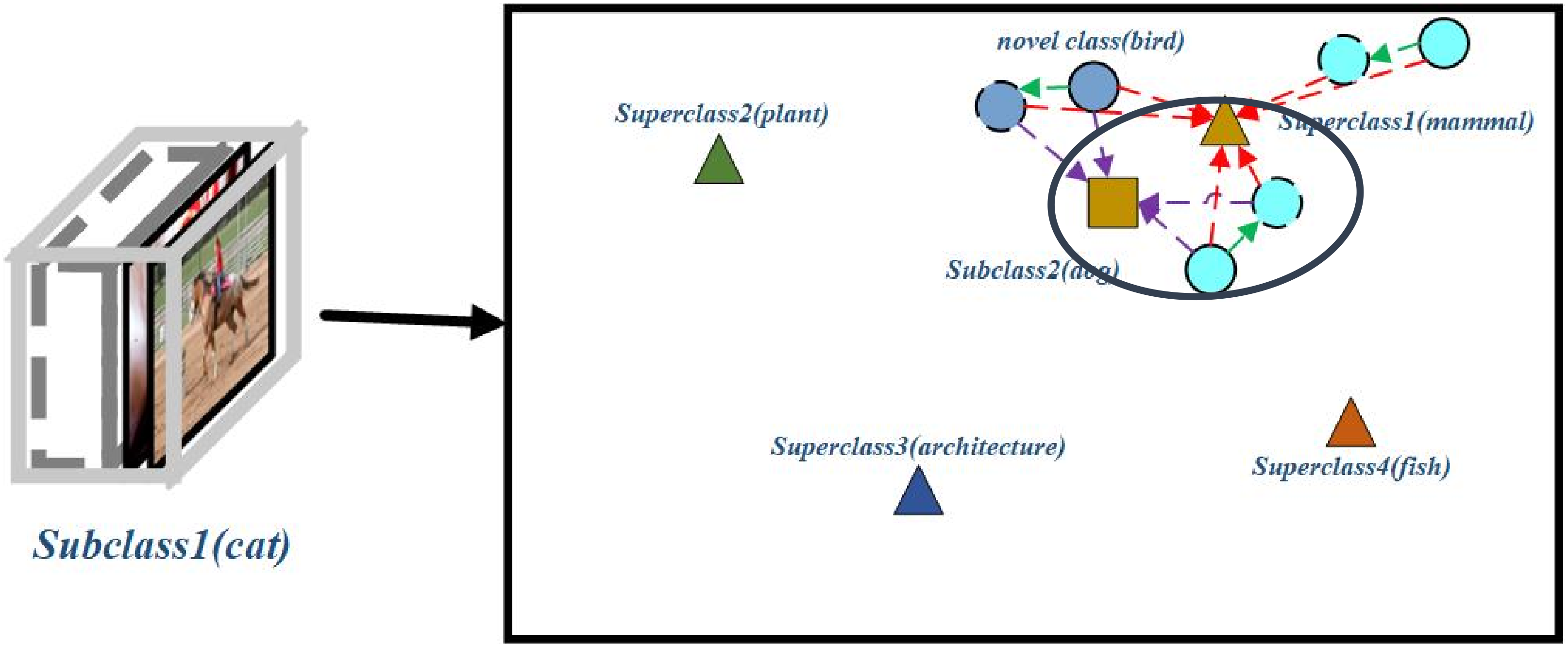
we use the selective distribution alignment distillation mechanism to guide the representation learning of models.



the current new class C is a subclass of an old class, the output entropy value of the class C sample on the old model, which corresponds to the superclass, is added with one margin value.

3. Method—— Hierarchical Distribution Alignment（HDA）

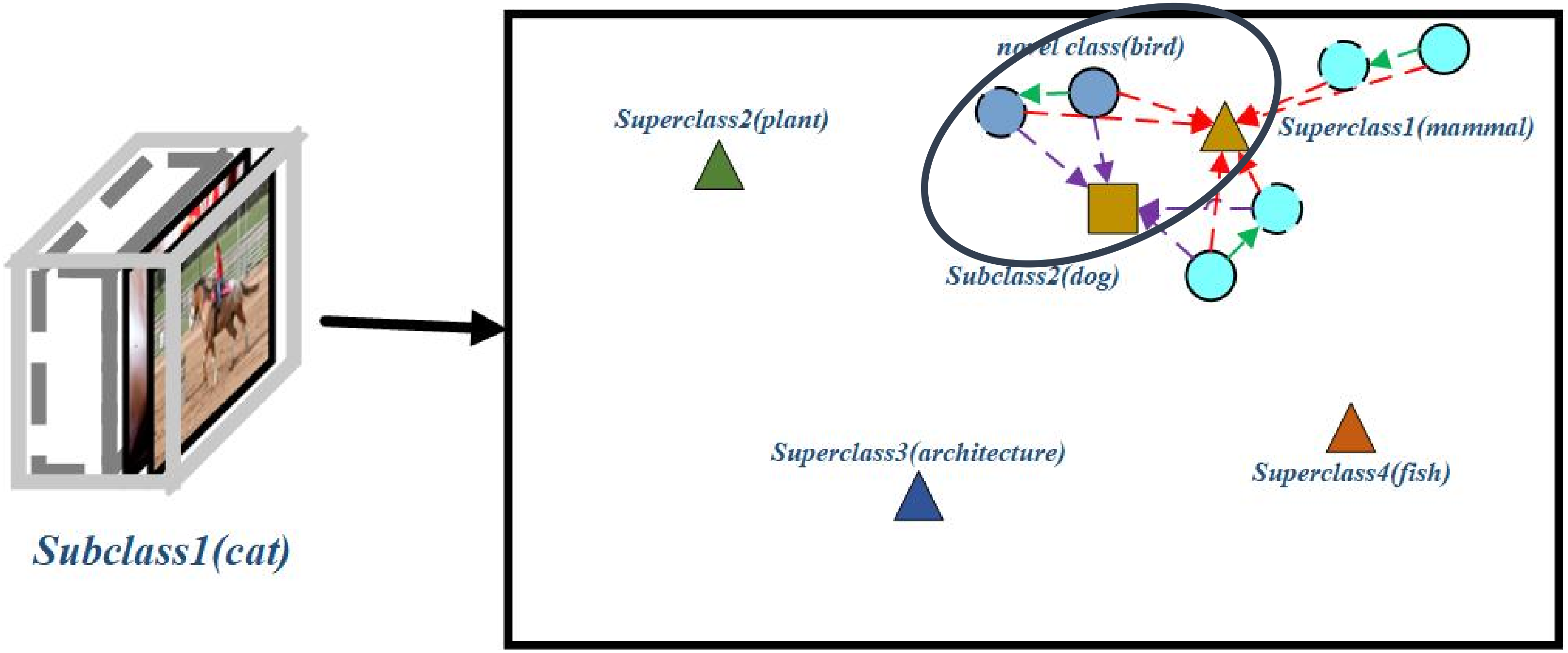
we use the selective distribution alignment distillation mechanism to guide the representation learning of models.



At the same time, the highest output entropy value of the non-superclass is subtracted by a margin value.

3. Method—— Hierarchical Distribution Alignment（HDA）

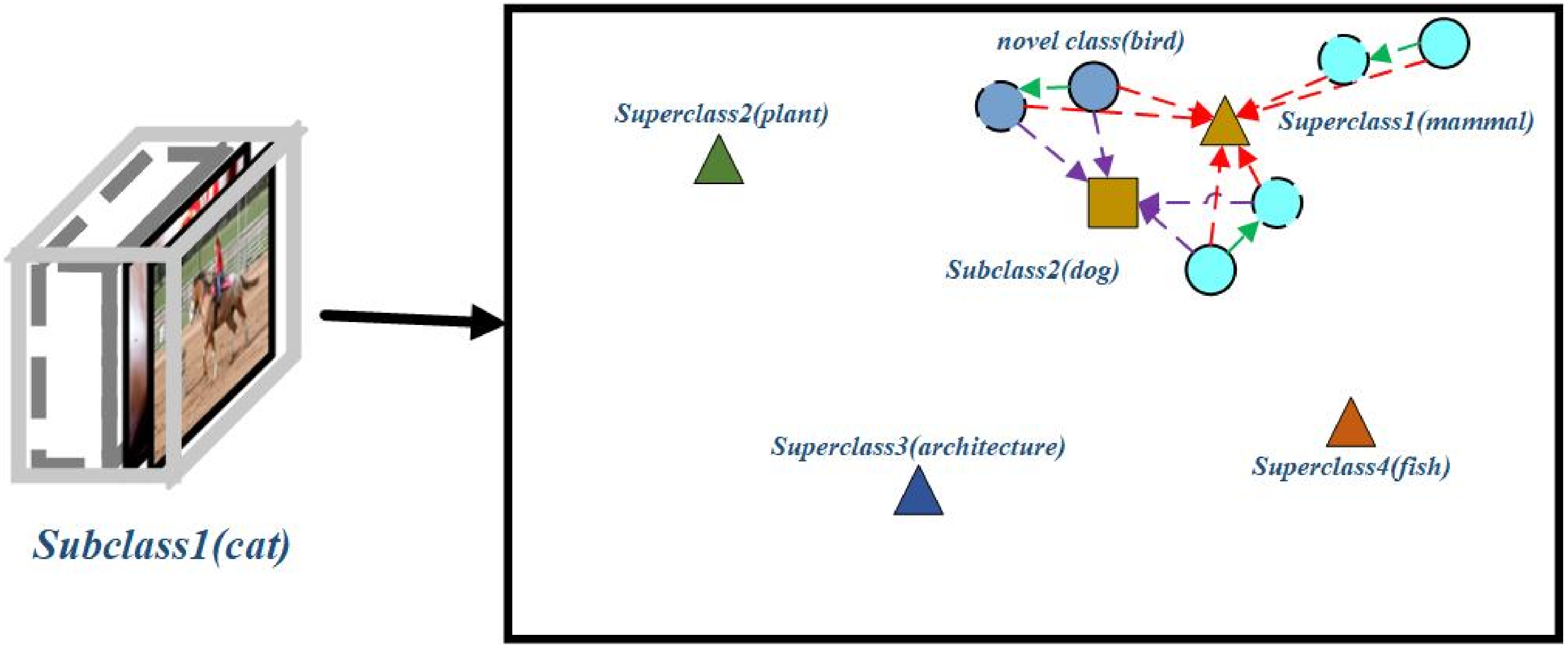
we use the selective distribution alignment distillation mechanism to guide the representation learning of models.



the current new class C is not related to an old class, the highest output entropy by the sample of class C on the old model is subtracted by a margin value.

3. Method—— Hierarchical Distribution Alignment（HDA）

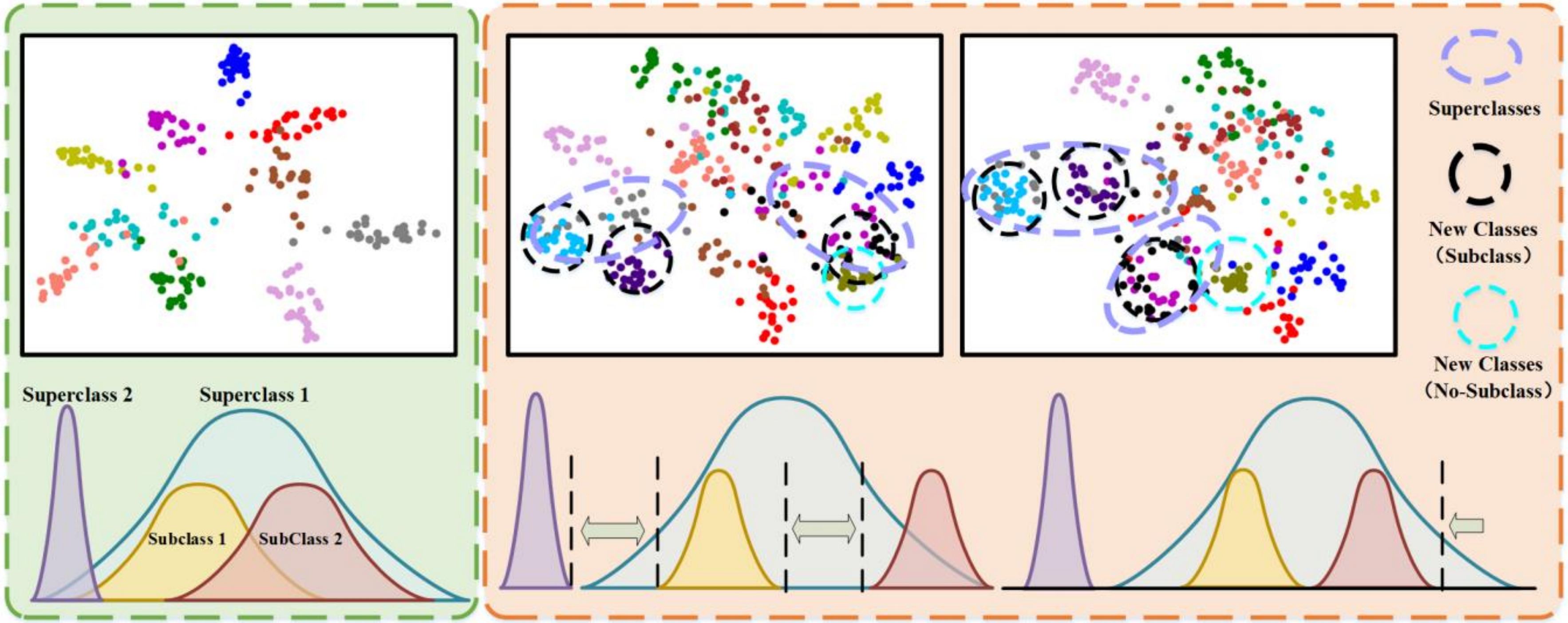
we use the selective distribution alignment distillation mechanism to guide the representation learning of models.



$$\mathcal{L}_{dis} = \mathrm{BCEWithLogitsLoss}(f_\theta(h_\theta(\boldsymbol{x}))[:, : n_{old}], y^{new})$$

y_{new} is the new output of the old model, after performing our hierarchical distribution alignment strategy.

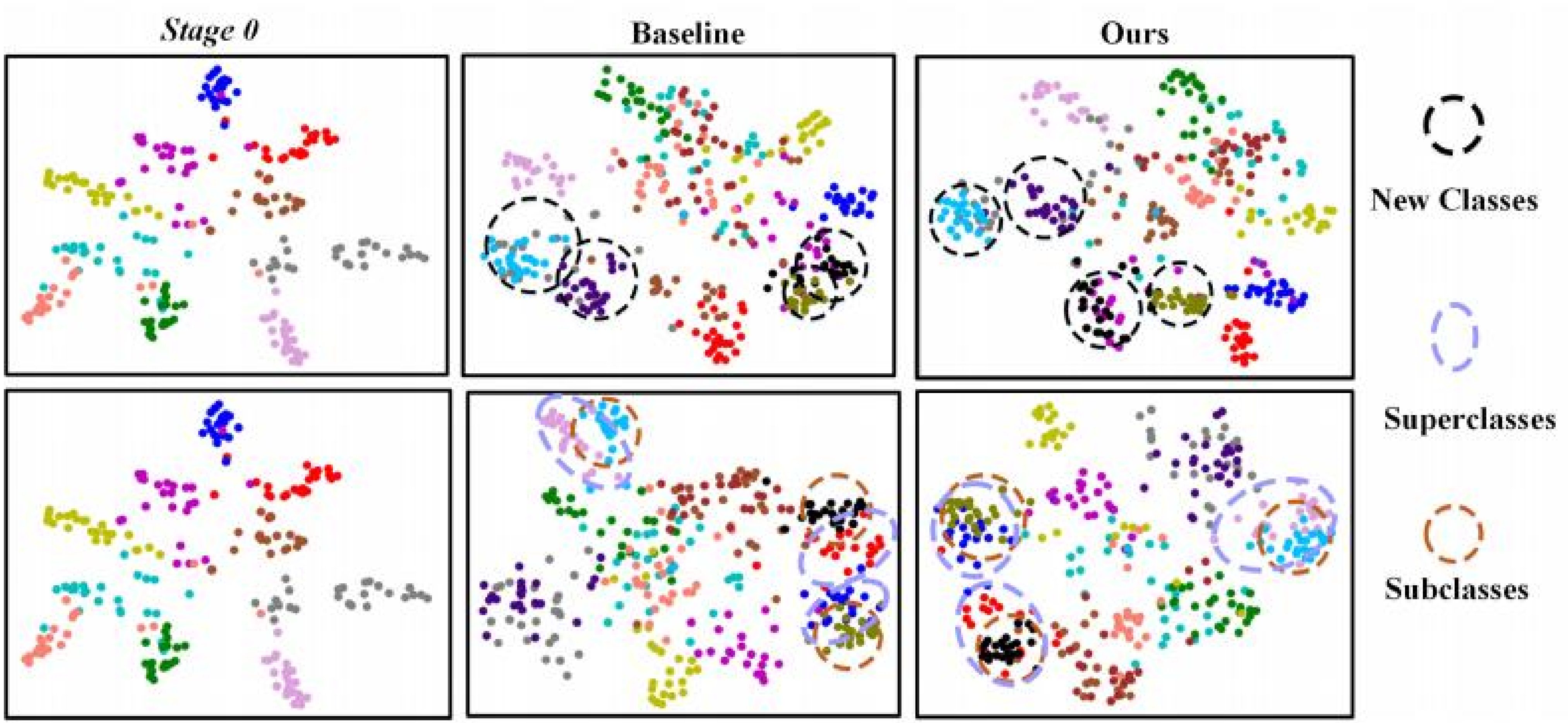
3. Method—— UAHR



$$\mathcal{L}_{all} = \mathcal{L}_{cls} + \mathcal{L}_{dis} + \mathcal{L}_{div} * \gamma$$

γ denotes hyper-parameters for balancing the losses

4. Experiments

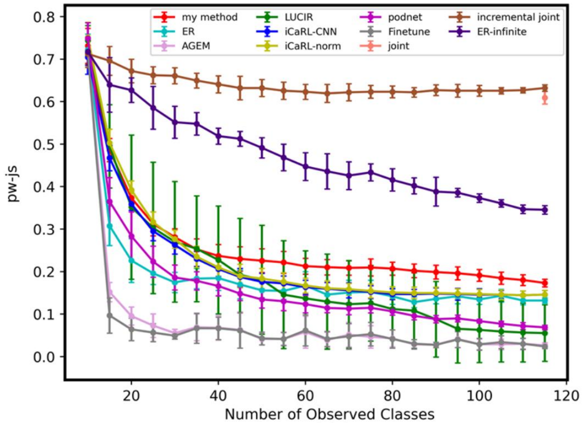


4. Experiments

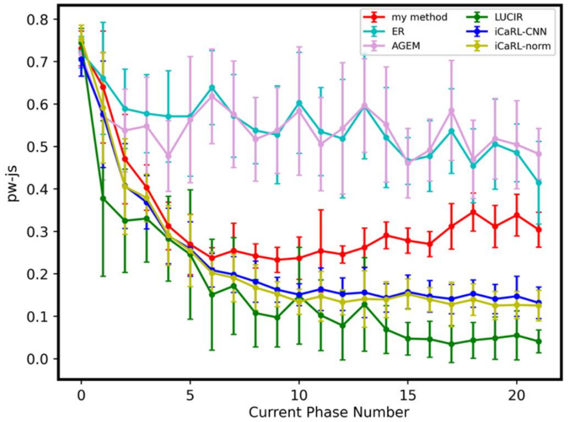
GRE	HDA	IIRC-CIFAR				
		phase 0	phase 5	phase 10	phase 15	phase 21
		78.35	26.48	21.27	18.81	17.78
√		77.04	26.75	21.46	19.38	18.32
	√	77.06	29.31	24.73	22.42	18.38
√	√	77.53	30.11	25.31	23.56	19.05

Table 1: Ablation study of our method on IIRC-CIFAR.

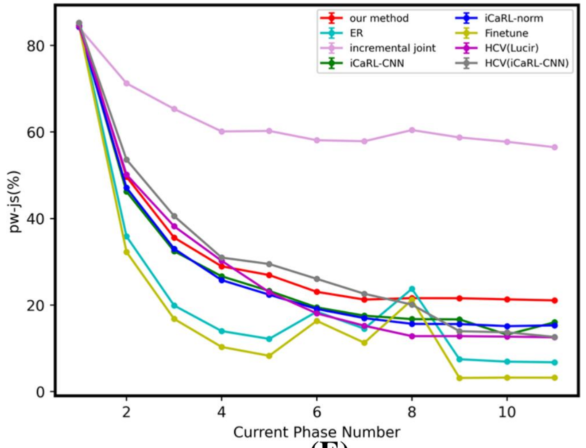
4. Experiments



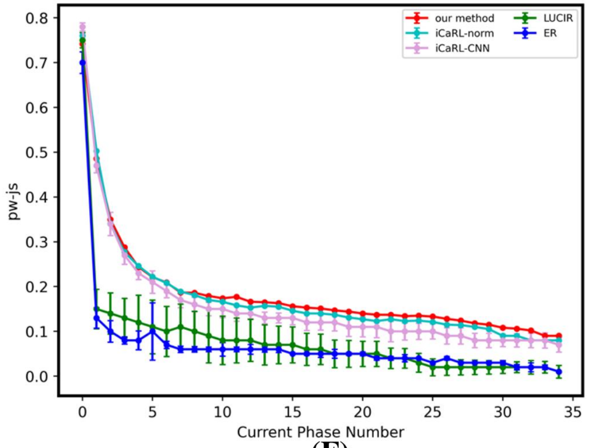
(A)



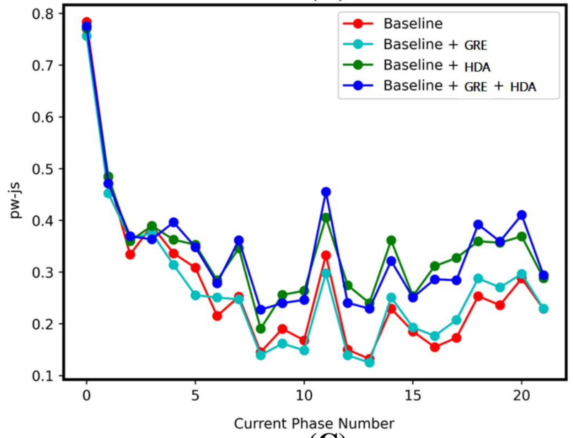
(B)



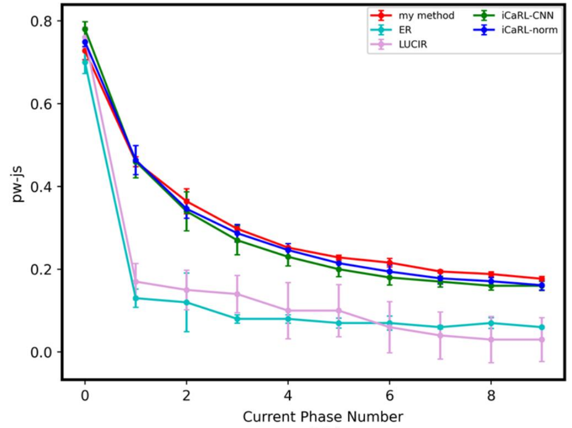
(E)



(F)

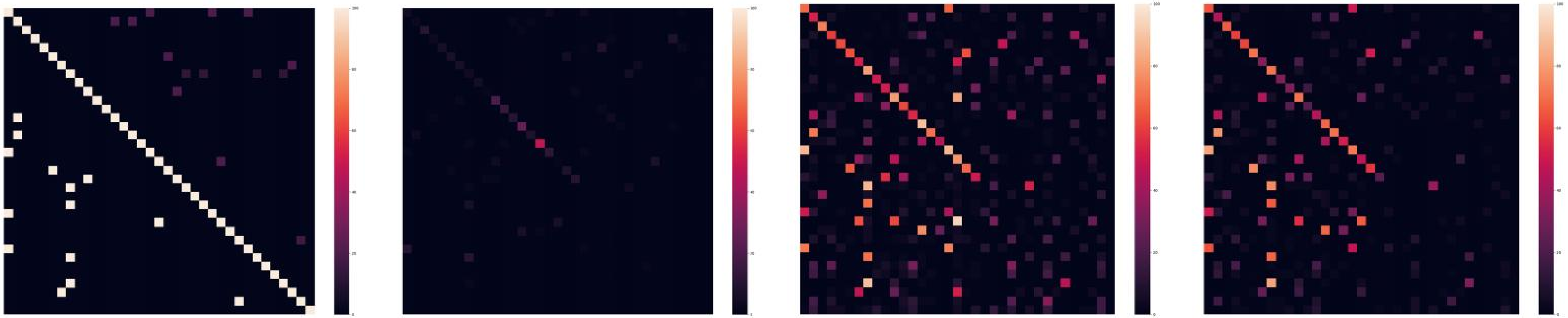


(C)



(D)

4. Experiments



(A)Ground Truth (B)Fine Tune (C)iCaRL-CNN (D)Ours

5. Conclusion

This paper proposes a novel Uncertainty-Aware Hierarchical Refinement scheme for the IIRC task. A global representation extension strategy is presented to enhance the discrimination of incremental classes, and the tricky distillation process is refined with a hierarchical distribution alignment strategy. Consequently, our method involves a multi-level semantic scenery in incremental learning. Experimental results show the superiority of our method in both stability and plasticity.