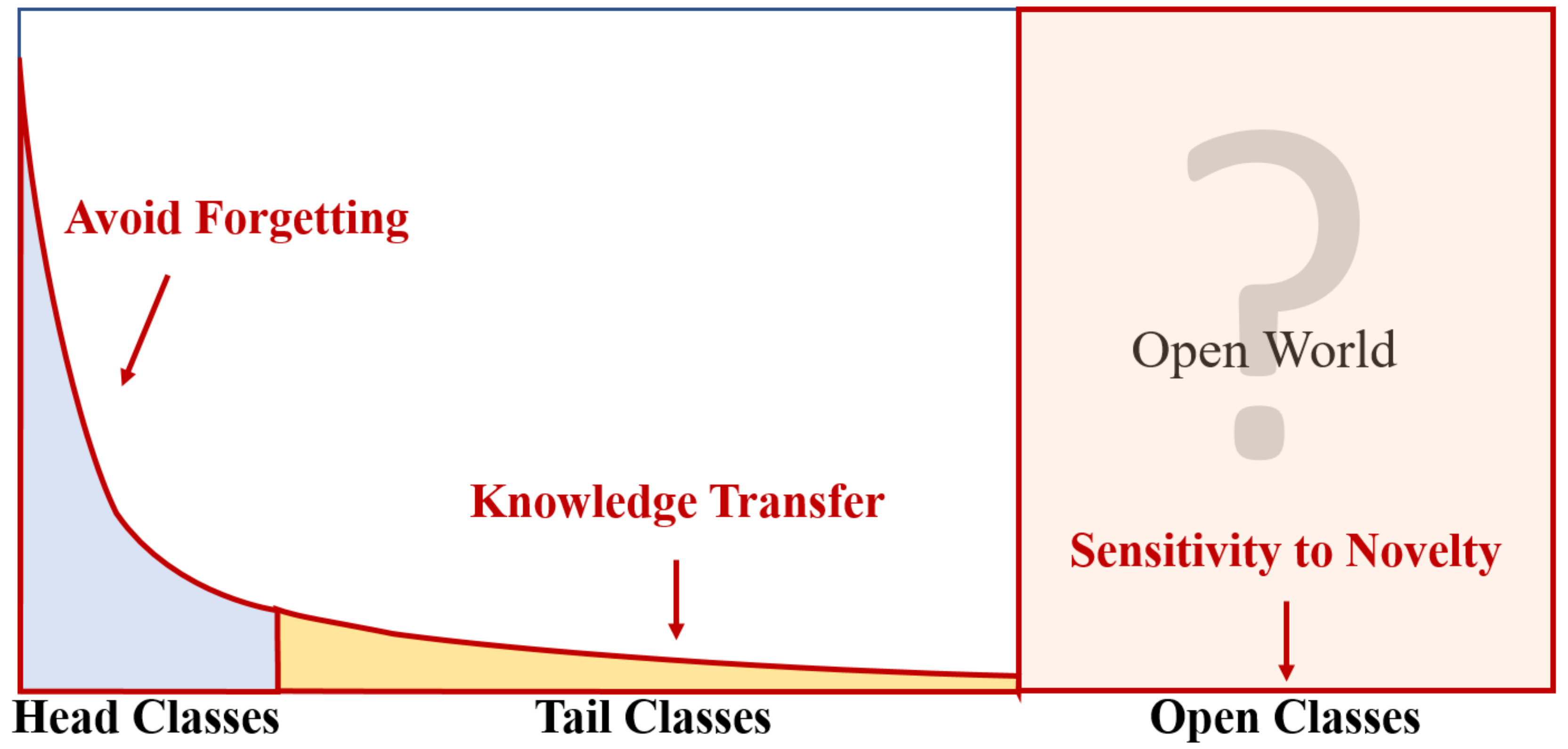
Large-Scale Long-Tailed Recognition in an Open World

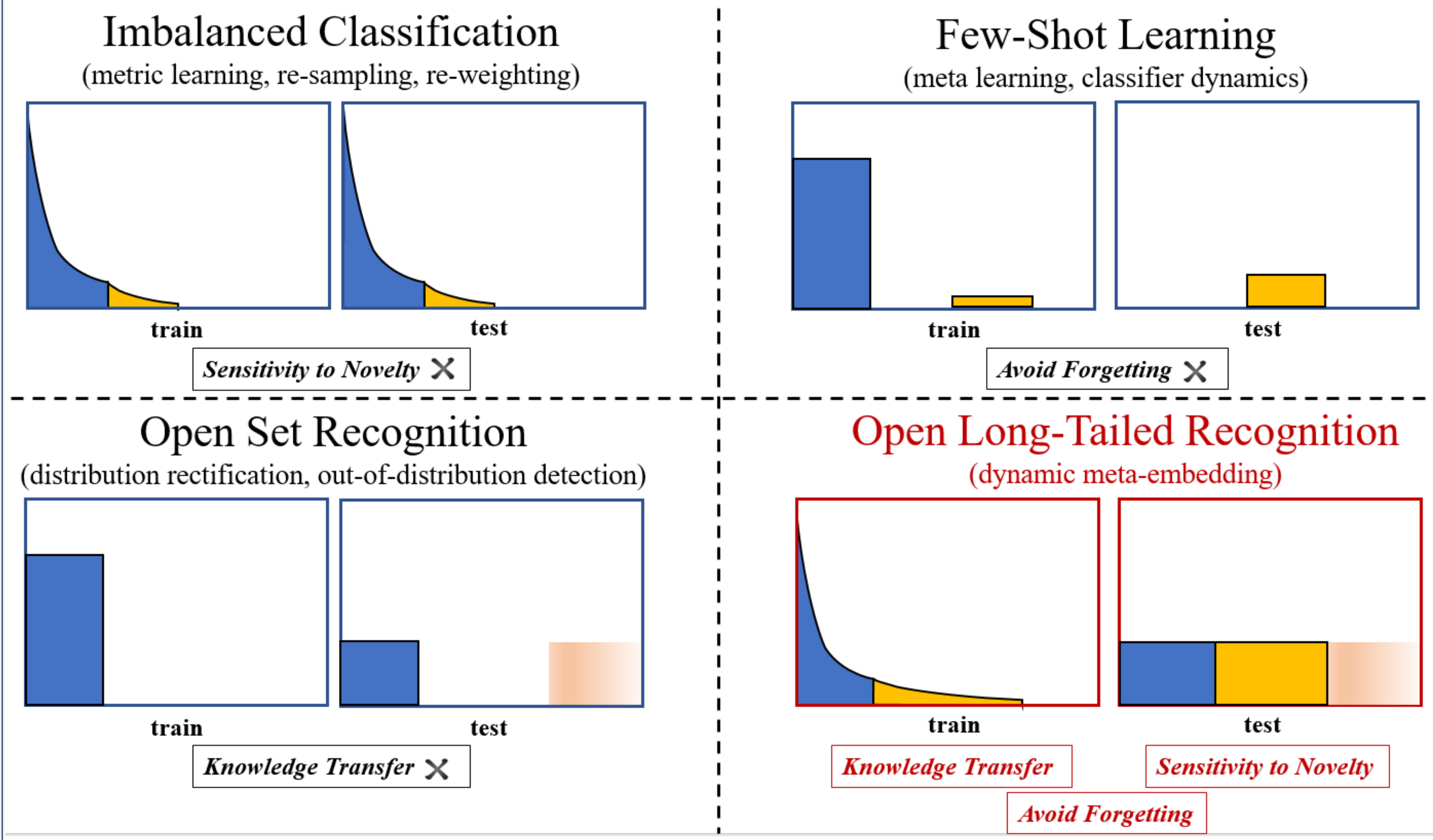
Ziwei Liu*, Zhongqi Miao*, Xiaohang Zhan, Jiayun Wang, Boqing Gong, Stella X. Yu
The Chinese University of Hong Kong & UC Berkeley / ICSI

Open Long-Tailed Recognition



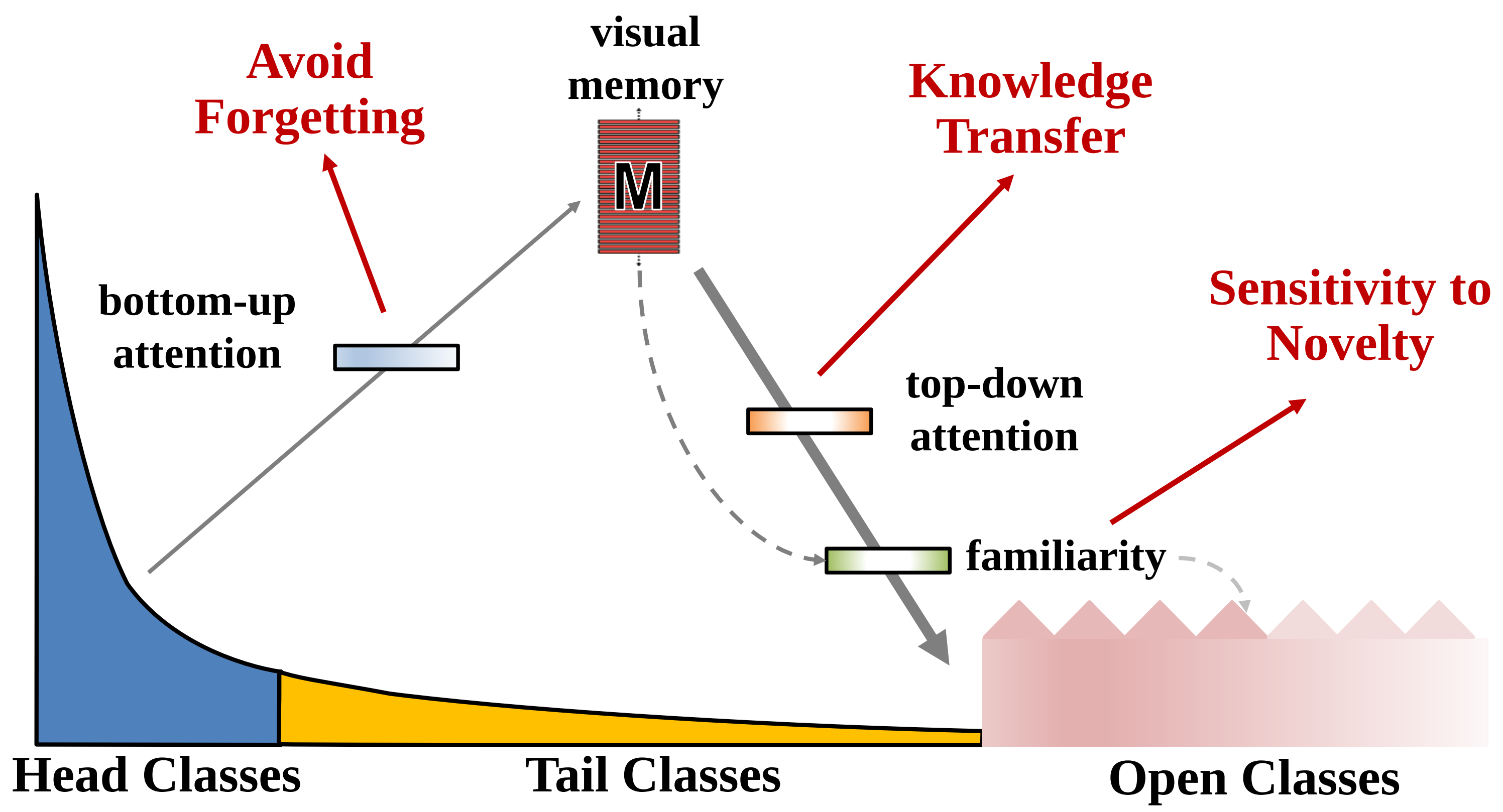
- the whole spectrum of real-world visual recognition

Relation to Existing Tasks



- **A full treatment to the whole spectrum of real-world visual recognition**
- **A task requiring improvement to different aspects of the existing deep neural networks**

Approach Intuition

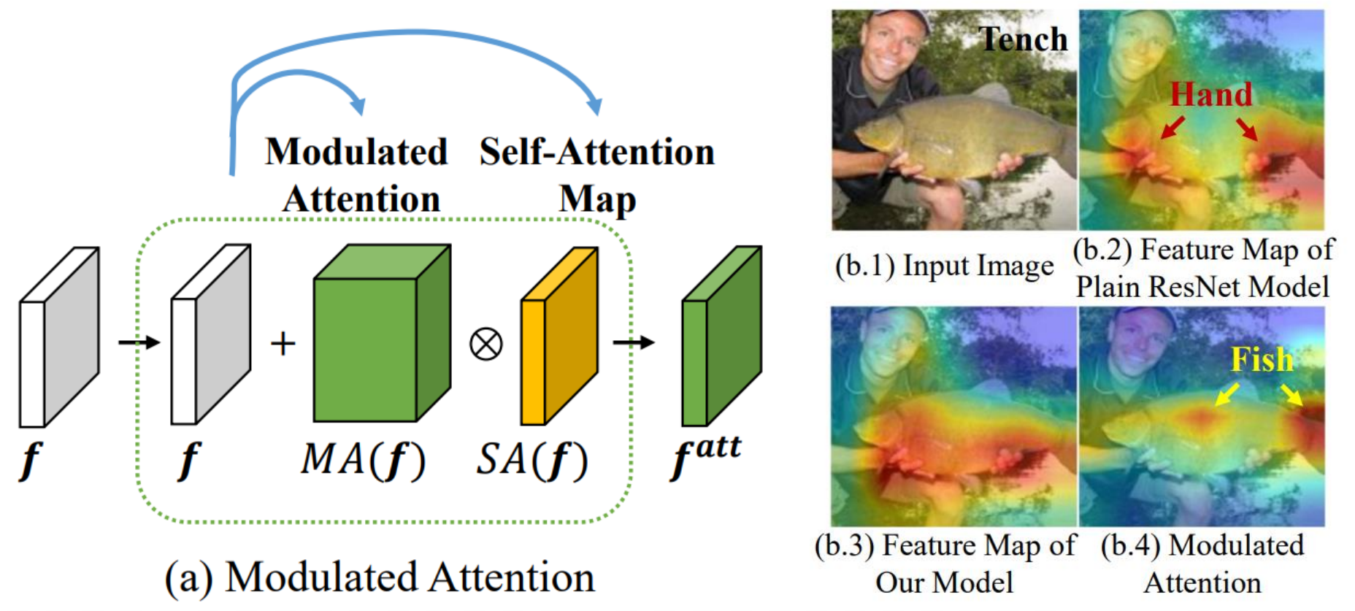


- **incorporate *memory* and *attention* into learning**

Module Explanation

bottom-up attention
(spatial self-attention)

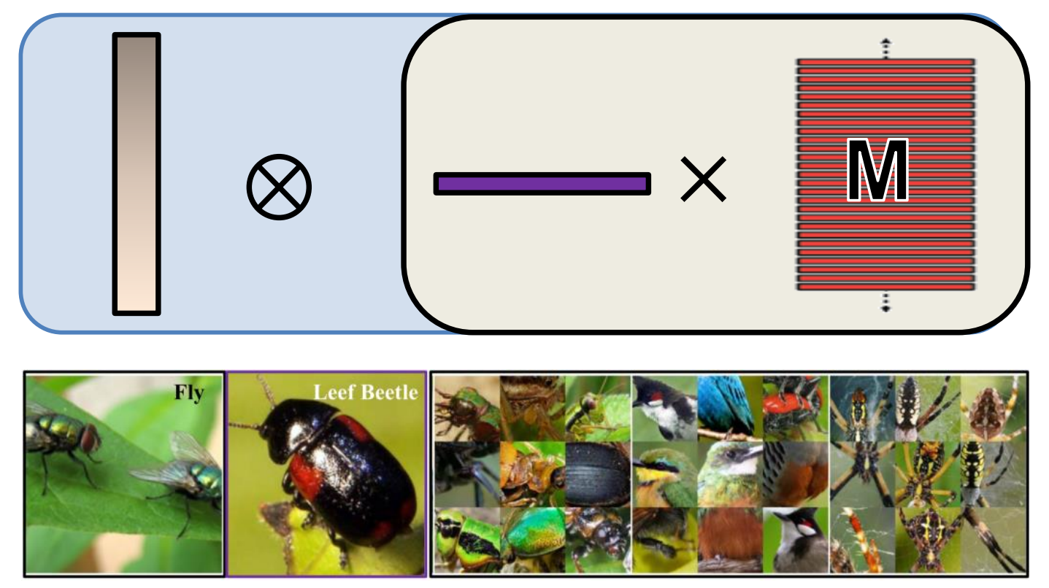
Avoid Forgetting



(a) Modulated Attention

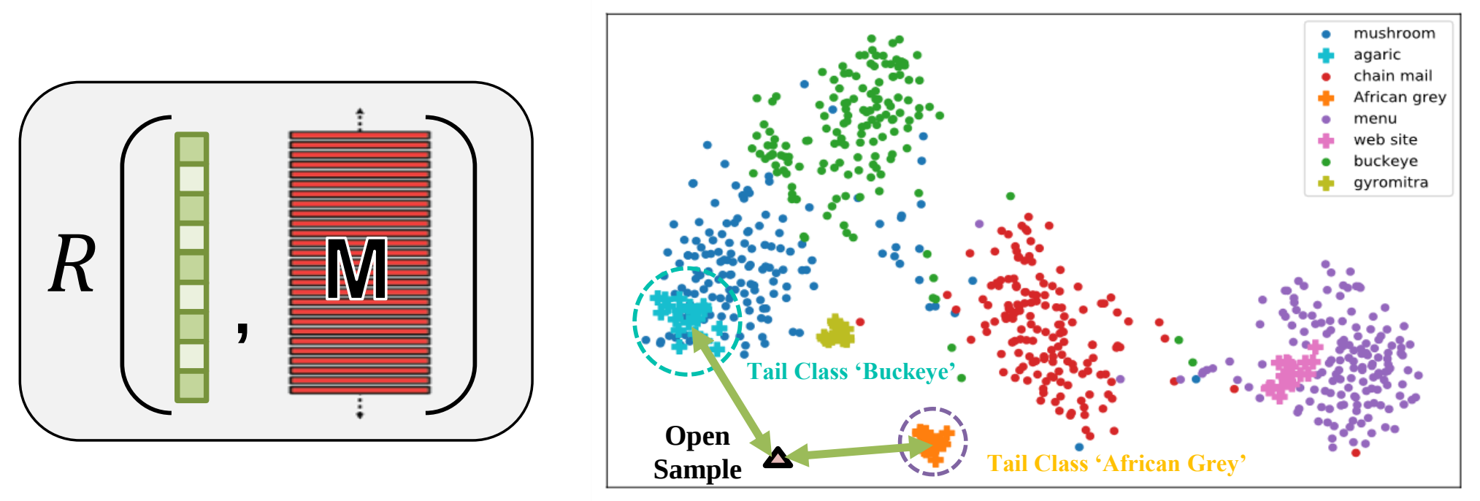
top-down attention
(memory-induced features with attention selection)

Knowledge Transfer

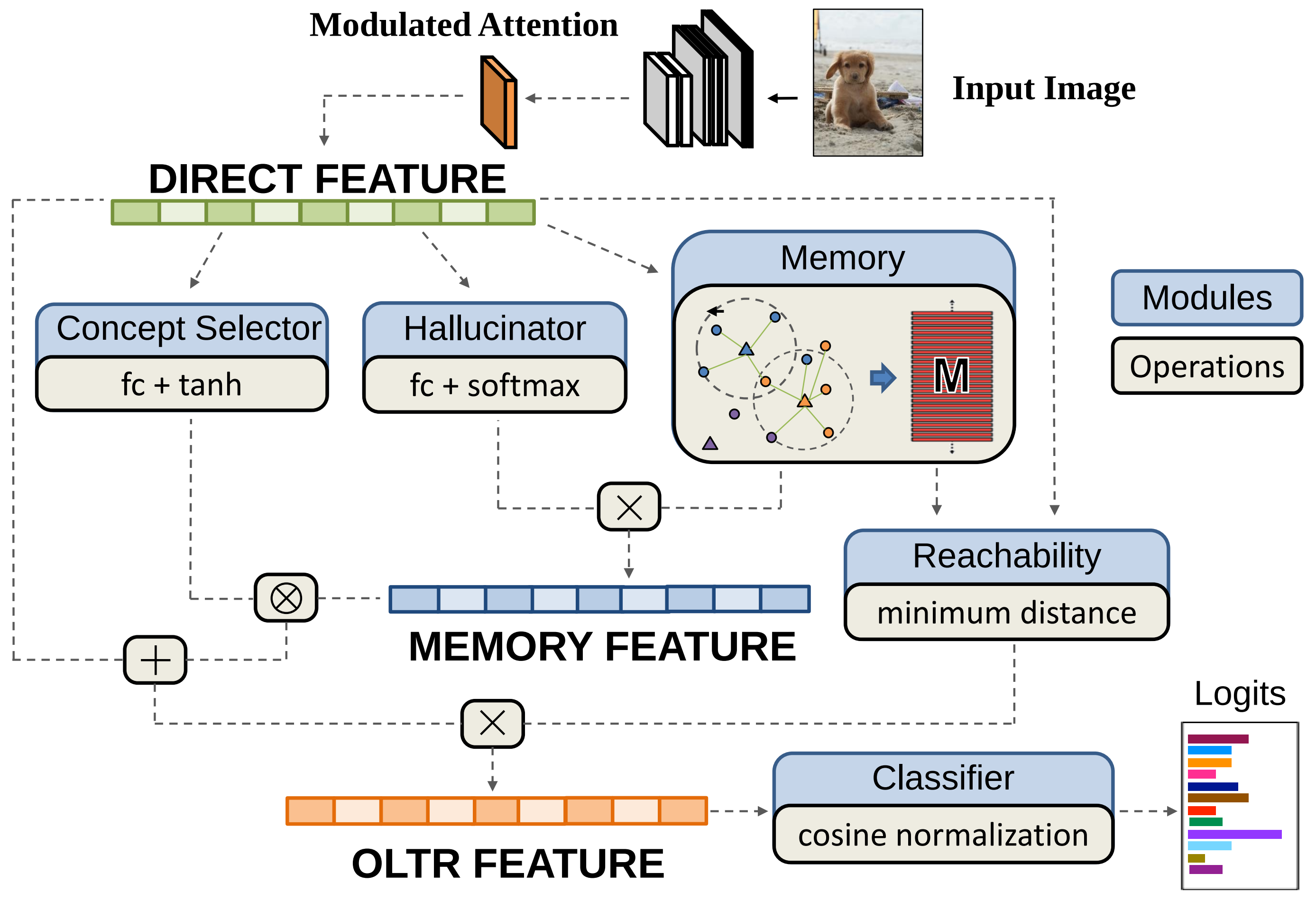


familiarity
(reachability to the learned memory)

Sensitivity to Novelty



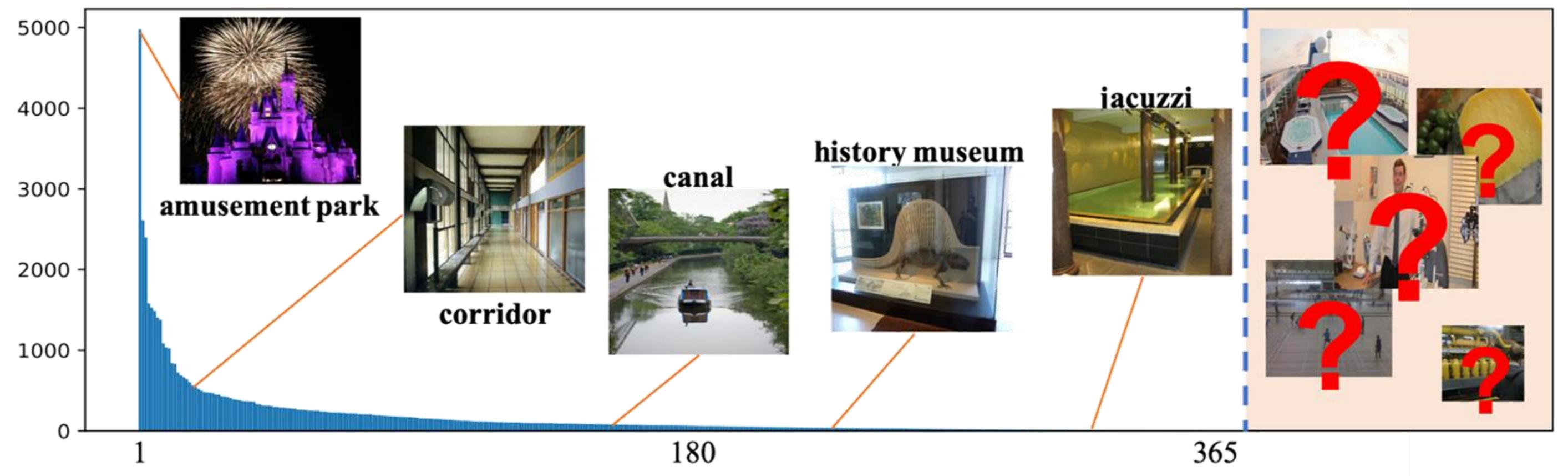
Overall Architecture



Benchmarks

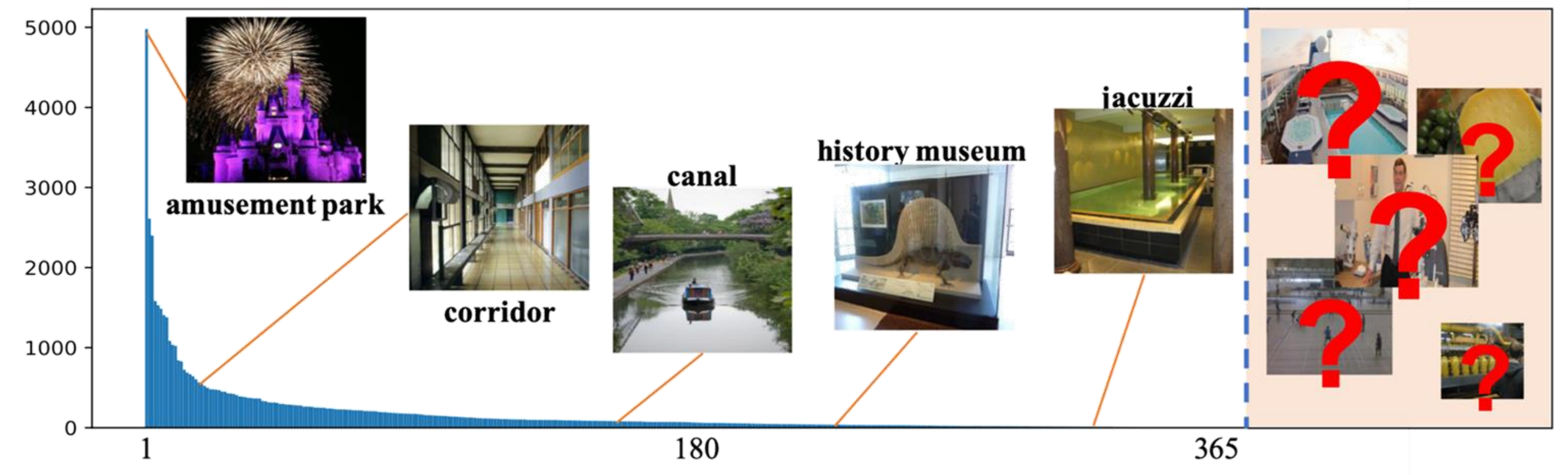
ImageNet-LT Benchmark

Absolute Performance Gain: ~20%



Places-LT Benchmark

Absolute Performance Gain: ~10%



head class > 1000 samples, tail class < 5 samples

Experimental Results

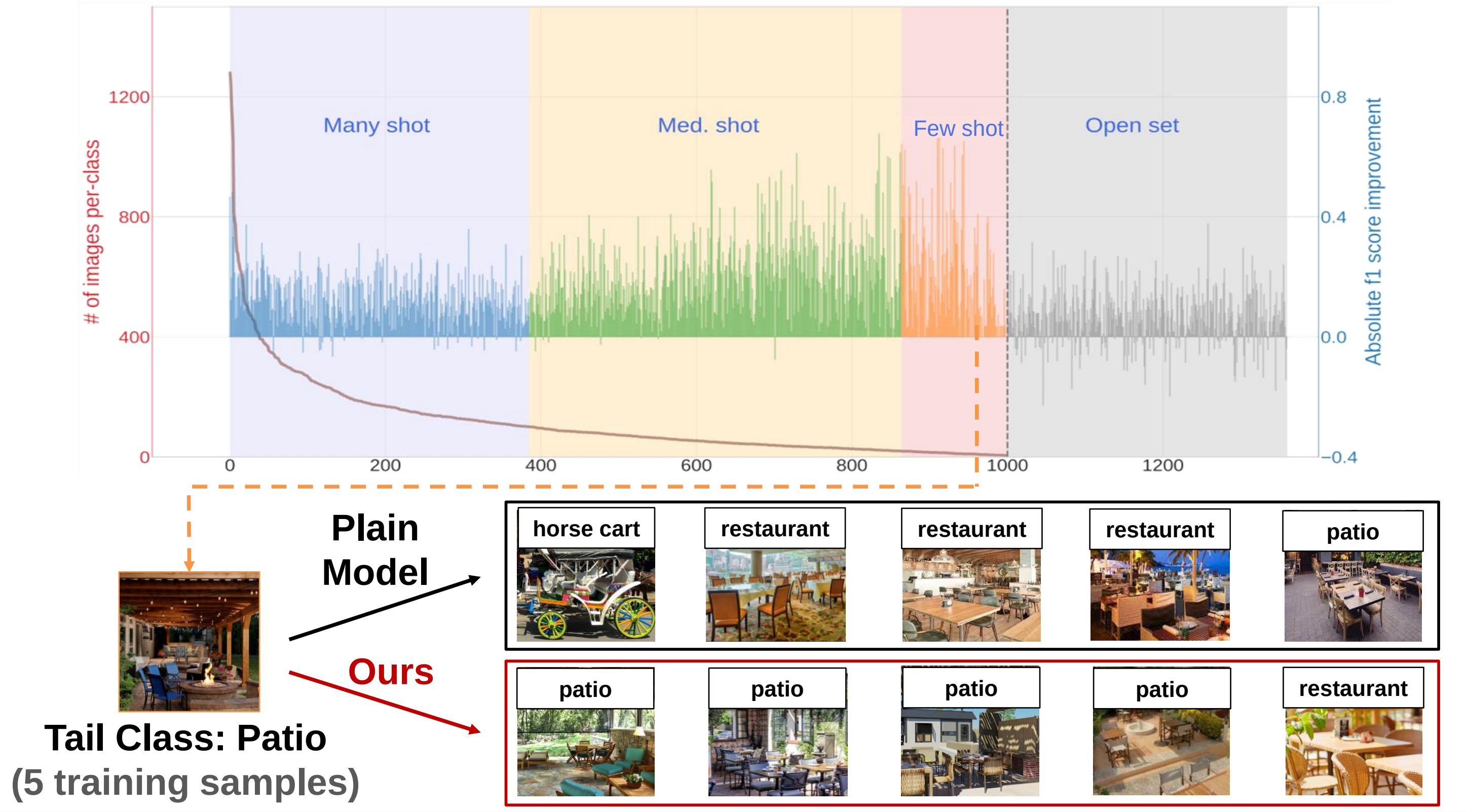
top-1 classification accuracy on ImageNet-LT

Backbone Net ResNet-10 Methods	closed-set setting > 100 Many-shot	 ⩽ 100 & > 20 Medium-shot	 < 20 Few-shot	 Overall	open-set setting > 100 Many-shot	 ⩽ 100 & > 20 Medium-shot	 < 20 Few-shot	 F-measure
Plain Model [19]	40.9	10.7	0.4	20.9	40.1	10.4	0.4	0.295
Lifted Loss [34]	35.8	30.4	17.9	30.8	34.8	29.3	17.4	0.374
Focal Loss [28]	36.4	29.9	16	30.5	35.7	29.3	15.6	0.371
Range Loss [60]	35.8	30.3	17.6	30.7	34.7	29.4	17.2	0.373
+ OpenMax [2]	-	-	-	-	35.8	30.3	**17.6**	0.368
FSLwF [14]	40.9	22.1	15	28.4	40.8	21.7	14.5	0.347
Ours	**43.2**	**35.1**	**18.5**	**35.6**	**41.9**	**33.9**	17.4	**0.474**

top-1 classification accuracy on Places-LT

Backbone Net ResNet-152 Methods	closed-set setting > 100 Many-shot	 ⩽ 100 & > 20 Medium-shot	 < 20 Few-shot	 Overall	open-set setting > 100 Many-shot	 ⩽ 100 & > 20 Medium-shot	 < 20 Few-shot	 F-measure
Plain Model [19]	**45.9**	22.4	0.36	27.2	**45.9**	22.4	0.36	0.366
Lifted Loss [34]	41.1	35.4	24	35.2	41	35.2	23.8	0.459
Focal Loss [28]	41.1	34.8	22.4	34.6	41	34.8	22.3	0.453
Range Loss [60]	41.1	35.4	23.2	35.1	41	35.3	23.1	0.457
+ OpenMax [2]	-	-	-	-	41.1	35.4	23.2	0.458
FSLwF [14]	43.9	29.9	**29.5**	34.9	38.1	19.5	14.8	0.375
Ours	44.7	**37**	25.3	**35.9**	44.6	**36.8**	**25.2**	**0.464**

Across-the-board improvement on all class types



Tail Class: Patio
(5 training samples)

Conclusions

- **New Task** towards real-world visual recognition
  Open Long-Tailed Recognition
- **New Approach** with memory-augmented network
  Dynamic Meta-Embedding
- **New Benchmarks** for future research
  ImageNet-LT Places-LT MS1M-LT