Hierarchical Semantic Tree Anchoring for CLIP-based Class-Incremental Learning

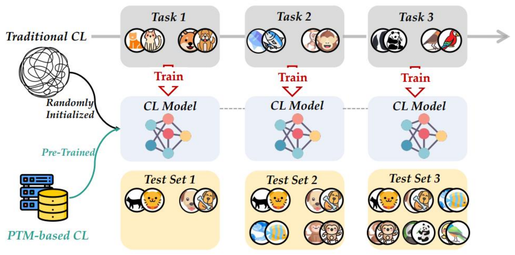
Tao Hu, Lan Li, Zhen-Hao Xie, Da-Wei Zhou
Nanjing University hut@lamda.nju.edu.cn

NEURAL INFORMATION PROCESSING SYSTEMS

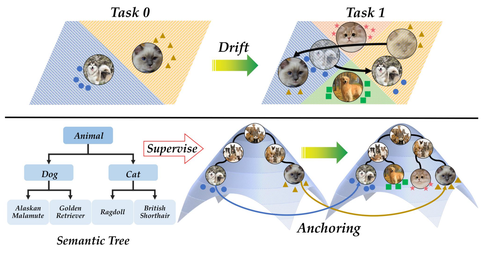
Highlights

- Hierarchical semantic trees provide **structured supervision** for CLIP-based class-incremental learning.
- Hierarchy-aware modeling **preserves parent-child relations** in hyperbolic space.
- Null-space projection **stabilizes** the shared mapper and **reduces interference** between old and new classes.
- Extensive experiments show **state-of-the-art** performance

Class-Incremental Learning with Pre-Trained Model



Motivation



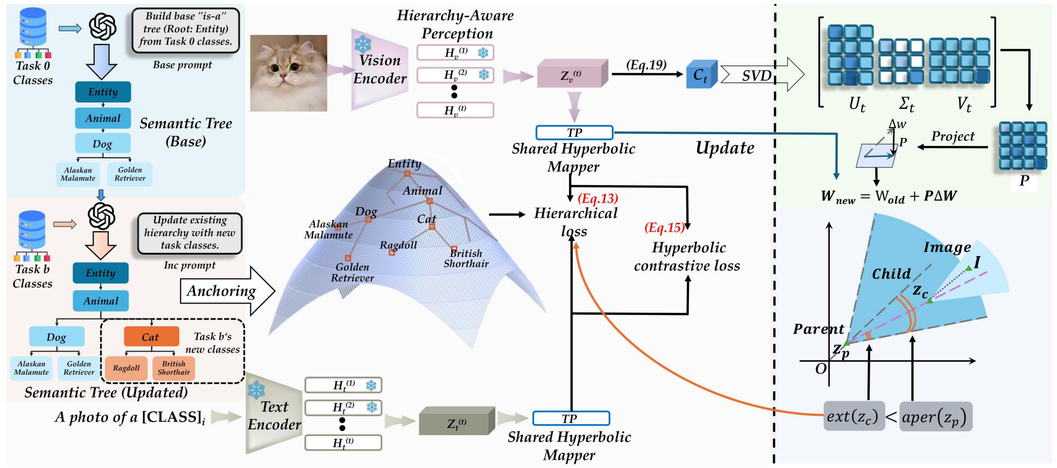
Existing CLIP-based CIL often overlooks the natural *hierarchy and suffers from feature drift as new classes arrive.*

Can we mitigate feature drift in CLIP-based CIL by leveraging semantic hierarchies to anchor fine-grained representations across tasks?

HASTEN Framework

Key idea: Use semantic hierarchies as stable anchors for CLIP representations, while constraining shared updates to reduce forgetting.

Setting: Exemplar-free CLIP-based CIL: train on current-task data only; recognize all seen classes without task IDs at inference.



Incremental Semantic Tree -- Build Stable Anchors : GPT-5 incrementally constructs an is-a hierarchy using only observed class names. Existing parent-child relations remain unchanged, while cached and frozen virtual-node embeddings provide stable semantic anchors across tasks.

Hierarchy-Aware Anchoring — Preserve Hierarchical Relations : Task-specific visual and textual modules adapt frozen CLIP features, with previous modules frozen and their outputs accumulated. A shared TP maps them into Lorentz hyperbolic space, where entailment constraints preserve text–text and text–image hierarchies, while contrastive learning maintains cross-modal alignment.

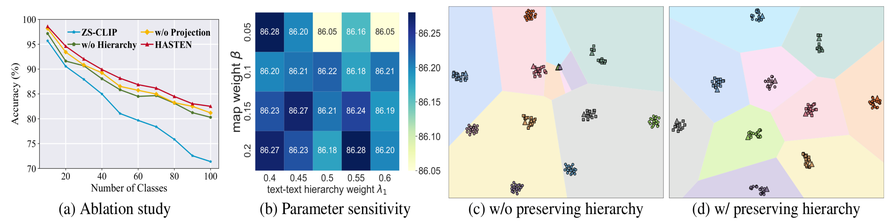
$$\mathcal{L}_{\text{hier}}^{\text{txt}\to\text{txt}} = \mathcal{L}_{\text{entail}}(\mathbf{z}_t^p, \mathbf{z}_t^c) + \text{SP}\left(\delta - d_{\mathcal{B}}(\mathbf{z}_t^p, \mathbf{z}_t^c)\right)$$

Null-Space Stabilization — Protect Previous Mappings: Since TP is shared across tasks, its updates may disturb old representations. We estimate an approximate null space from prior-task feature covariance via SVD and project TP gradients onto it, reducing interference while retaining adaptability to new classes.

$$\Delta W_{\text{proj}} = V_\perp V_\perp^\top \Delta W,$$
$$X^o \Delta W_{\text{proj}} = X^o V_\perp V_\perp^\top \Delta W = (X^o V_\perp)(V_\perp^\top \Delta W) \approx 0$$

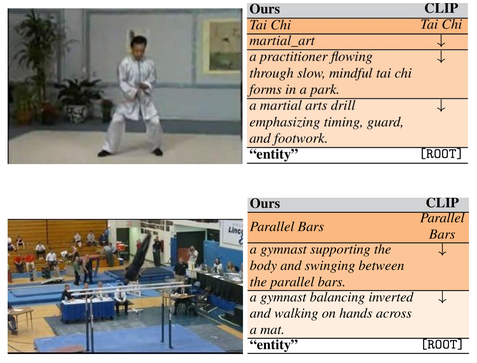
State-Of-The-Art Results

Method	Aircraft B0 Inc10		Aircraft B50 Inc10		CIFAR100 B0 Inc10		CIFAR100 B50 Inc10		Cars B0 Inc10		Cars B50 Inc10	
	$\overline{\mathcal{A}}$	$\mathcal{A}_B$	$\overline{\mathcal{A}}$	$\mathcal{A}_B$	$\overline{\mathcal{A}}$	$\mathcal{A}_B$	$\overline{\mathcal{A}}$	$\mathcal{A}_B$	$\overline{\mathcal{A}}$	$\mathcal{A}_B$	$\overline{\mathcal{A}}$	$\mathcal{A}_B$
Finetune	3.16	0.96	1.72	1.05	7.84	4.44	5.30	2.46	3.14	1.10	1.54	1.13
SimpleCIL [88]	59.24	48.09	53.05	48.09	84.15	76.63	80.20	76.63	92.04	86.85	88.96	86.85
ZS-CLIP [56]	26.66	17.22	21.70	17.22	81.81	71.38	76.49	71.38	82.60	76.37	78.32	76.37
L2P [76]	47.19	28.29	44.07	32.13	82.74	73.03	81.14	73.61	76.63	61.82	76.37	65.64
DualPrompt [75]	44.30	25.83	46.07	33.57	81.63	72.44	80.12	72.57	76.26	62.94	76.88	67.55
CODA-Prompt [64]	45.98	27.69	45.14	32.28	82.43	73.43	78.69	71.58	80.21	66.47	75.06	64.19
PROOF [90]	63.81	56.14	59.47	57.10	86.77	79.11	83.32	79.73	90.74	86.51	88.00	85.58
RAPF [30]	50.38	23.61	40.47	25.44	86.14	78.04	82.17	77.93	82.89	62.85	75.87	63.19
CLG-CBM [84]	66.05	55.93	59.26	55.39	81.82	74.57	68.61	65.55	78.94	74.26	80.11	73.62
HASTEN	**67.23**	**56.32**	**61.47**	**57.19**	**88.64**	**82.51**	**85.45**	**82.38**	**93.06**	**88.93**	**90.23**	**88.46**



(a) Ablation study (b) Parameter sensitivity (c) w/o preserving hierarchy (d) w/ preserving hierarchy

Qualitative Traversal



HASTEN traces specific actions to generic concepts.

Conclusion

HASTEN uses hierarchy-aware external knowledge to stabilize CLIP-based class-incremental learning, and achieves **state-of-the-art results** across 9 benchmarks.

Wechat Toolbox Code