Taxonomy-Aware Class-Incremental Semantic Segmentation for Open-World Perception.

Julia Hindel, Daniele Cattaneo, and Abhinav Valada

***Abstract*— Semantic segmentation models are typically trained on a fixed set of classes, limiting their applicability in open-world scenarios. Class-incremental semantic segmentation aims to update models with emerging new classes while preventing catastrophic forgetting of previously learned ones. However, existing methods impose strict rigidity on old classes, reducing their effectiveness in learning new incremental classes. In this work, we propose Taxonomy-Oriented Poincaré-regularized Incremental-Class Segmentation (TOPICS) that learns feature embeddings in hyperbolic space following explicit taxonomy-tree structures. Additionally, we maintain implicit class relational constraints on the geometric basis of the Poincaré ball. This ensures that the latent space can continuously adapt to new constraints while maintaining a robust structure to combat catastrophic forgetting. Extensive evaluations of TOPICS on the Cityscapes benchmark demonstrate that it achieves state-of-the-art performance. Additional details are available at http://topics.cs.uni-freiburg.de.**

I. Introduction

Class-Incremental Learning (CIL) aims to update a model with new classes at periodic timesteps, balancing learning of new classes while preserving knowledge of old classes [1]. Class-Incremental Semantic Segmentation (CISS) additionally incorporates the background shift, as pixels that belong to old classes are labeled as background in new data samples [2]. State-of-the-art CISS methods restrain the forgetting of old knowledge with data replay [3], distillation [2], [4], or network expansion [5], [6]. In this work, we introduce taxonomy-aware continual semantic segmentation for automated driving scenarios. Our proposed method TOPICS, depicted in Fig. 1, enforces features conform to taxonomy-tree structures in hyperbolic space. To further avoid catastrophic forgetting, we incorporate pseudo-labeling of the background, and we introduce two novel regularization losses.

II. Technical Approach

A. Class-Incremental Semantic Segmentation

CISS aims at training a model f_θ over $t = 1, ..., T$ incremental tasks. Every task is defined by its own disjoint label space $\mathcal{C}^t$ and training dataset $(x^t, y^t) \in \mathcal{D}^t$. The background class b^t includes all pixels whose true semantic class (y) is not included in $\mathcal{C}^t$. We consider the more realistic overlapped setting of CISS where training images (x^t) may include pixels whose dataset ground truth labels are old, current, or future classes. After every task t, the network is challenged to make predictions on $\mathcal{C}^{1:t}$ whereas only true background pixels should not be associated with a semantic class. In contrast to [2]–[4], [8], we do not constrain future classes to originate only from the background in taxonomic CISS. We regard incremental scenarios where future classes are refinements of known classes or the background. In the former case, we define disjoint subsets $\mathcal{D}^t$ according to a fixed ratio, i.e. the same image cannot be observed with different labeling taxonomies at different time steps.

B. Semantic Segmentation with the Poincaré Model

We model the class hierarchy in hyperbolic space due to its favorable property of equidistant node connections on all hierarchy levels. Consequently, distances are inversely proportional to the semantic similarity of classes. The hyperbolic space follows the geometry of constant negative curvature which is defined in the variable c. The Poincaré model is a stereographic projection of the upper sheet of a two-sheeted hyperboloid and is represented by a unit ball, see [10] for more details. The geometric interpretation of multinomial regression in hyperbolic space suggests that every class y is represented as a hyperplane in the Poincaré ball with offset $o_y \in \mathbb{D}_c^N$ and orientation $r_y \in T\mathbb{D}_c^N$ [11].

C. Hierarchical Segmentation

We model the hierarchy of semantic classes in the last layer of the network. We opt for a binary cross-entropy loss to ensure magnitudes of old and new class predictions do not correlate. Specifically, we extend the state-of-the-art hierarchical segmentation loss [12] to multi-hierarchy levels. Therefore, we model leaf nodes and all their ancestors as separate output classes $\mathcal{V}$ and use a combination of ancestor $\mathcal{A}$ and descendant $\mathcal{D}$ logits (s) in the loss function. We follow the tree-min loss $(\mathcal{L}_{TM})$ proposed in [12]. Further, we separately employ a categorical cross-entropy $(\mathcal{L}_{CE})$ on every hierarchy level. This loss penalizes high prediction scores of sibling class descendants. The complete hierarchical loss is defined as $\mathcal{L}_{hier} = \alpha\mathcal{L}_{TM} + \beta\mathcal{L}_{CE}$.

D. Hierarchical Relation Distillation

We employ an InfoNCE loss on the hyperbolic class hyperplanes $\mathcal{H}_t$ to maintain closely grouped classes of the prior model in a similar constellation in the updated model. We define a distance between two classes y^1 and y^2 as the distance between one class offset o_{y^1} and the hyperplane of the other class H_{y^2}. Before beginning the training procedure, we utilize the old model's weights to compute the top k most similar hyperplanes H_y for every offset o_y in $\mathcal{C}^{1:t-1}$. Further, we denote all positive anchors of a class, $k^+_{y_1}$, as the top k

This work was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1597 – 499552394.
Department of Computer Science, University of Freiburg, Germany
The authors thank Kshitij Sirohi for technical discussions.

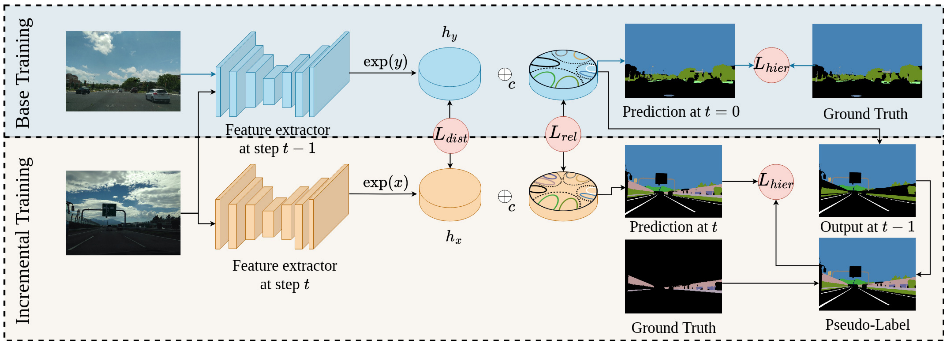

Fig. 1: During base training of TOPICS, features are mapped onto the Poincaré ball before the class hierarchy is explicitly enforced with $\mathcal{L}_{hier}$. In incremental steps, the old model is used to generate pseudo-labels of old classes and to regularize the last layer's weights with $\mathcal{L}_{rel}$ and feature radii with $\mathcal{L}_{dist}$.

TABLE I: Continual semantic segmentation results on Cityscapes in mIoU (%). Tasks defined as $\mathcal{C}^1$-$\mathcal{C}^T$(T tasks) and h class hierarchy increments.

	14-1 (6 tasks)			**10-1 (10 tasks)**			**7-4 (4 tasks)h**			**7-18 (2 tasks)h**		
Method	1-14	15-19	all	1-10	11-19	all	1-7	8-25	all	1-7	8-25	all
PLOP [4]	63.54	15.38	48.33	60.75	27.97	42.96	88.56	18.14	20.75	88.73	15.06	17.99
MiB [2]	66.37	14.36	50.05	61.80	32.97	45.73	77.66	6.61	9.83	90.10	5.71	9.64
MiB + AWT [7]	65.60	19.19	50.72	60.97	35.70	46.55	84.65	10.46	13.64	90.19	5.61	9.56
DKD [8]	68.83	14.70	51.86	66.77	34.52	48.92	89.46	0.56	4.98	89.19	4.29	8.32
MicroSeg [9]	51.35	11.61	38.84	44.37	23.55	32.78	86.39	1.63	5.79	86.37	7.71	11.26
TOPICS (Ours)	**73.03**	**42.47**	**61.74**	**71.37**	**52.62**	**59.36**	**90.02**	**51.31**	**50.69**	**90.33**	**61.62**	**59.98**

smallest absolute distances to o_{y_1} and enforce these relations to be maintained during the incremental training. We apply an InfoNCE-inspired loss:

$$\mathcal{L}_{rel} = -\log \frac{\exp\left(1 - \tau \cdot d_{k^+}/d_{max}\right)}{\sum_{i=0}^{D} \exp\left(1 - \tau \cdot d_i/d_{max}\right)}. \quad (1)$$

with τ being the temperature hyper-parameter.

E. Hyperbolic Distance Correlation

As incremental data is unbalanced with new classes appearing more frequently, we aim to constrain the radii of features to be unchanged between the old and new models. Therefore, we enforce features of the new and old model to be equidistant from the center of the Poincaré ball ($\mathcal{L}_{dist}$).

III. Experimental Evaluation

A. Datasets

We evaluate TOPICS on the Cityscapes [13] dataset. The Cityscapes dataset consists of 19 semantic classes in addition to a void class. For CISS from the background, we adapt the 14-1 (6 tasks) and 10-1 (10 tasks) setting as proposed in [7]. The first 14 or 10 classes are learned during base training while one class is added per incremental step. For CISS from known classes, we learn 7 base classes that correspond to the official sub-categories defined for Cityscapes and increment the model in a 7-4 (4 tasks) or 7-18 (2 tasks) manner.

B. Experimental Setup

In line with prior work [2], [4], [9], we use the DeepLabV3 model with the ResNet-101 backbone which is pre-trained on ImageNet for all the experiments. We employ the Geoopt library [14] to project the Euclidean features to a Poincaré ball with $c = 2.0$. Further, we follow the Möbius approximation defined in [11] for more efficient computations. We train TOPICS for 60 epochs per task with batch size 24 using the Riemannian SGD optimizer with momentum of 0.9 and weight decay of 0.0001. We use a poly learning rate scheduler with initial learning rates of 0.05 for base training and 0.01 in all incremental steps. For the hierarchical loss function, we set $\alpha = 5$ and $\beta = 1$. We train on random non-empty crops of (512,1024) with horizontal flipping.

C. Quantitative Results

We compare TOPICS with five state-of-the-art CISS methods: PLOP [4], MiB [2], MiB+AWT [7], DKD [8] and MicroSeg [9]. For each method, we use the respective author's published code and use the same augmentations outlined in Sec. III-B. We evaluate the models using the mean intersection-over-union (mIoU) metric over all the base classes ($\mathcal{C}_1$) and novel classes ($\mathcal{C}_{2:T}$) separately as an indication of rigidity and plasticity. We present the results in Tab. I. TOPICS outperforms all baselines by at least 9.88pp on the CISS from the background. While the difference in base IoU measures 4.2pp, our method significantly exceeds the benchmarks by at least 16.9pp in terms of novel IoU. Additionally, the baselines are specifically designed for the CISS from background setting only, and thus significantly underperform on the CISS from known classes setting. This highlights the need for solutions tailored to both scenarios. We highlight the versatility of our method to balance plasticity and rigidity in all tested CISS settings.

IV. Conclusion

We present TOPICS, a novel CISS approach that models features conforming to taxonomy-tree structures on the Poincaré ball to balance rigidity and plasticity in incremental learning. TOPICS further maintains implicit class relations between old class hyperplanes and constraints features to have equidistant radii. Our method is one of the early works that uniformly addresses the bifurcation of previously observed classes and incremental classes from the background.

References

[1] D.-W. Zhou, Q.-W. Wang, Z.-H. Qi, H.-J. Ye, D.-C. Zhan, and Z. Liu, “Deep class-incremental learning: A survey,” *arXiv preprint arXiv:2302.03648*, 2023.

[2] F. Cermelli, M. Mancini, S. Rota Bulò, E. Ricci, and B. Caputo, “Modeling the background for incremental learning in semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2020.

[3] S. Cha, b. kim, Y. Yoo, and T. Moon, “Ssul: Semantic segmentation with unknown label for exemplar-based class-incremental learning,” in *Adv. Neural Inform. Process. Syst.*, vol. 34, 2021, pp. 10 919–10 930.

[4] A. Douillard, Y. Chen, A. Dapogny, and M. Cord, “Plop: Learning without forgetting for continual semantic segmentation,” *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 4039–4049, 2020.

[5] C.-B. Zhang, J.-W. Xiao, X. Liu, Y.-C. Chen, and M.-M. Cheng, “Representation compensation networks for continual semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 7053–7064.

[6] J.-W. Xiao, C.-B. Zhang, J. Feng, X. Liu, J. van de Weijer, and M.-M. Cheng, “Endpoints weight fusion for class incremental semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2023, pp. 7204–7213.

[7] D. Goswami, R. Schuster, J. van de Weijer, and D. Stricker, “Attribution-aware weight transfer: A warm-start initialization for class-incremental semantic segmentation,” in *Proc. of the IEEE/CVF Winter Conf. on Applications of Comput. Vis.*, 2023, pp. 3195–3204.

[8] D. Baek, Y. Oh, S. Lee, J. Lee, and B. Ham, “Decomposed knowledge distillation for class-incremental semantic segmentation,” in *Adv. Neural Inform. Process. Syst.*, 2022.

[9] Z. Zhang, G. Gao, Z. Fang, J. Jiao, and Y. Wei, “Mining unseen classes via regional objectness: A simple baseline for incremental segmentation,” in *Adv. Neural Inform. Process. Syst.*, vol. 35, 2022, pp. 24 340–24 353.

[10] M. Nickel and D. Kiela, “Poincaré embeddings for learning hierarchical representations,” *Advances in neural information processing systems*, vol. 30, 2017.

[11] M. G. Atigh, J. Schoep, E. Acar, N. Van Noord, and P. Mettes, “Hyperbolic image segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 4453–4462.

[12] L. Li, W. Wang, T. Zhou, R. Quan, and Y. Yang, “Semantic hierarchy-aware segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 4, pp. 2123–2138, 2024.

[13] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016.

[14] M. Kochurov, R. Karimov, and S. Kozlukov, “Geoopt: Riemannian optimization in pytorch,” *arXiv preprint arXiv:2005.02819*, 2020.