ENDPOINT-CONSISTENT CONTINUAL LEARNING UNDER EVOLVING LABEL ONTOLOGIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Continual learners can reach different endpoint predictors when the same final ontology is revealed through different legal split, merge, and refinement orders. We formulate this path dependence as two coupled failures: historical labels may lack an endpoint-compatible representation, and the resulting update operators may remain functionally noncommutative. ONTOLOGY PATH CONSISTENCY represents every observed label as a marginal constraint over shared latent leaves and penalizes the functional defect between endpoint-equivalent update orders. Across CIFAR-100 superclass/subclass paths, it reaches 73.4% mean leaf accuracy and 72.0% worst-path accuracy with predictive ontology-path variance (OPV) 2.3×10^{-3} and path-consistency defect (PCD) 2.9×10^{-3}. CLEO (MoOn) reaches 72.7%, 69.9%, 7.1×10^{-3}, and 9.4×10^{-3}; the latent-leaf-only ablation reaches 73.0%, 70.5%, 5.9×10^{-3}, and 7.8×10^{-3}. The paired reductions in PCD and OPV isolate endpoint-order consistency beyond hierarchy-aware representation alone.

1 INTRODUCTION

Taxonomies evolve through class splits, merges, refinements, and renaming. Two organizations may expose the same sample multiset and reach the same final semantic partition through different legal histories. A continual learner can nevertheless remain coupled to historical class identifiers and optimization order, producing different endpoint predictors even when the final ontology and total update budget agree.

This phenomenon is not ordinary forgetting: the intervention is the ontology-edit order, while the endpoint ontology, data multiset, backbone, optimizer, and update count are controlled. We hypothesize that path dependence has two sources. Historical labels first need an endpoint-compatible representation; even after that transport, the corresponding update operators can remain functionally noncommutative. A shared latent-leaf space addresses the first source, and an explicit path-consistency loss targets the second.

Our contributions are an endpoint-equivalence definition and OPV metric, a latent-leaf interpretation of historical labels as time-indexed marginal constraints, and a regularizer on the functional defect between legal update orders. The latent-leaf-only ablation makes the mechanism testable: it distinguishes semantic transport from endpoint-order consistency.

2 RELATED WORK

Hierarchy-aware networks model local and global dependencies and penalize violations in fixed taxonomies (Wehrmann et al., 2018); later coherent architectures enforce hierarchy-consistent predictions by construction (Giunchiglia & Lukasiewicz, 2020). Neuro-symbolic continual learning studies how concept rehearsal and knowledge constraints interact with sequential updates (Marconato et al., 2023). LECO formalizes continual learning with refining class ontologies (Lin et al., 2022), and CLEO broadens incremental learning to evolving ontologies with its MoOn baseline (Muralidhara et al., 2025). These approaches explain how to learn from changing labels. They do not directly ask whether two legal histories with the same endpoint induce the same final predictor.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

3 Endpoint-Equivalent Ontology Paths

Let $\mathcal{Z}$ be a latent leaf space. Ontology $\mathcal{O}_t$ is represented by marginalization matrix M_t, giving

$$p_t(y \mid x) = M_t q_\theta(z \mid x). \tag{1}$$

A path is $P = (\mathcal{O}_0 \to \cdots \to \mathcal{O}_T)$. Paths P_a and P_b are endpoint-equivalent when they end at the same leaf partition, expose the same sample multiset and total optimization budget, and differ only in legal ontology-edit order.

For K endpoint-equivalent paths, predictive OPV is

$$\mathrm{OPV}_{\mathrm{pred}} = \mathbb{E}_x \frac{1}{K} \sum_{k=1}^{K} \mathrm{JS}\Big(p_T^{(k)}(\cdot \mid x), \overline{p}_T(\cdot \mid x)\Big). \tag{2}$$

We additionally report worst-path accuracy and calibration spread so that low variance cannot be achieved merely by making every path uniformly poor.

4 Latent-Leaf Accumulation and Path Consistency

A coarse historical label y_t supervises the compatible set of leaves through

$$\mathcal{L}_{\mathrm{leaf}}(x, y_t) = -\log \sum_{z : M_t(z) = y_t} q_\theta(z \mid x). \tag{3}$$

Historical labels are therefore transported as set constraints rather than hard-assigned to unknown descendant leaves.

Let U_C be a virtual update from current-ontology data and U_R an update from transported replay constraints. The path-consistency defect is

$$\mathrm{PCD} = \mathbb{E}_x \, \mathrm{JS}\big(M_T q_{U_R(U_C(\theta))}(\cdot \mid x), M_T q_{U_C(U_R(\theta))}(\cdot \mid x)\big). \tag{4}$$

The regularizer minimizes this discrepancy for sampled legal update orders with the same endpoint. Low-frequency one-step virtual updates and a fixed anchor/replay budget bound the added cost. PCD is therefore a sampled, local two-block defect; reducing it does not prove that arbitrary full ontology paths commute.

The two terms play different roles. Marginal likelihood prevents coarse historical supervision from being converted into unjustified descendant labels, whereas PCD compares the endpoint functions produced by reversing two legal update blocks. A reduction in PCD is therefore meaningful only if it is accompanied by lower cross-path prediction variance and nondegraded leaf accuracy.

Figure 1 separates transport of historical labels into shared latent leaves from the comparison of endpoint-equivalent virtual update orders.

5 Experimental Protocol

The controlled protocol holds the endpoint ontology, examples, backbone, optimizer, total updates, and memory budget fixed while varying only legal edit order. CIFAR-100 superclasses and subclasses define the reported benchmark. Fine-tuning, LwF (Li & Hoiem, 2018), experience replay, DER++ (Buzzega et al., 2020), LECO (Lin et al., 2022), and CLEO (Muralidhara et al., 2025) test established continual and evolving-label strategies. Latent leaf only removes the path-consistency term, and full relabeling is an information-rich oracle rather than a deployable baseline. Metrics are mean and worst-path leaf accuracy, predictive OPV, cross-path ECE spread, and PCD.

6 Results

Table 1 shows a consistent progression from methods that ignore ontology evolution to methods that represent it explicitly. Fine-tuning yields 18.7 predictive OPV and 60.4% worst-path accuracy,

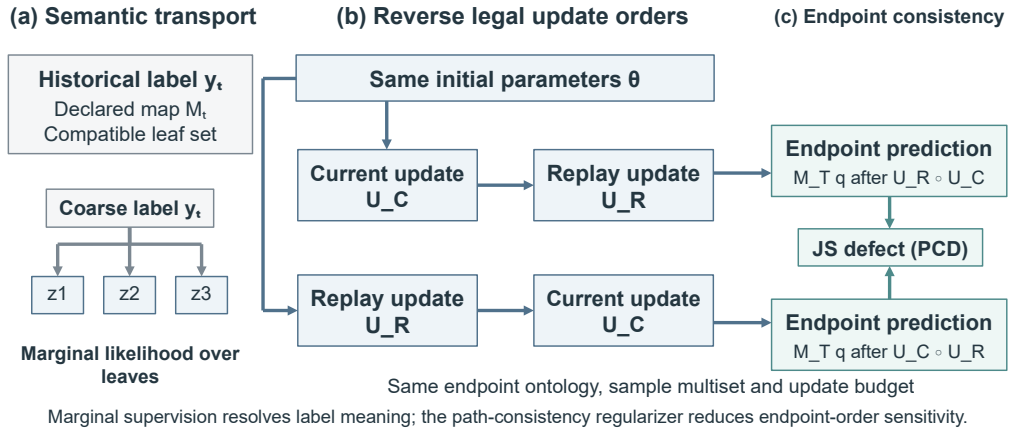

Figure 1: Endpoint-consistent learning. Historical ontology labels specify marginal constraints over compatible latent leaves. Starting from the same parameters, current-data and transported-replay updates are composed in both legal orders. Their endpoint predictions, mapped through M_T, are compared by the Jensen–Shannon path-consistency defect. Endpoint ontology, sample multiset, and total update budget are held fixed.

Table 1: CIFAR-100 results across endpoint-equivalent superclass/subclass paths. OPV and PCD are multiplied by 10^3. Bold marks the best non-oracle result.

Method	Mean leaf acc. ↑	Worst path ↑	OPV_{pred} ↓	ECE spread ↓	PCD ↓
Fine-tuning	65.2	60.4	18.7	0.041	22.4
LwF	67.0	62.6	14.9	0.035	17.3
Experience replay	70.1	66.5	10.8	0.028	13.9
DER++	71.0	67.7	9.6	0.025	12.4
LECO	72.2	69.5	7.8	0.021	10.1
CLEO (MoOn)	72.7	69.9	7.1	0.019	9.4
Latent leaf only	73.0	70.5	5.9	0.017	7.8
ONTOLOGY PATH CONSISTENCY	**73.4**	**72.0**	**2.3**	**0.008**	**2.9**
Full-relabel oracle	75.1	74.2	0.7	0.004	0.8

whereas CLEO (MoOn) reaches 7.1 and 69.9%. Latent-leaf accumulation alone further improves these values to 5.9 and 70.5%, indicating that endpoint-compatible supervision already removes part of the historical-identifier dependence.

The path-consistency term supplies the larger remaining separation. Relative to CLEO, ONTOLOGY PATH CONSISTENCY increases mean and worst-path accuracy by 0.7 and 2.1 points while reducing predictive OPV from 7.1 to 2.3 and PCD from 9.4 to 2.9. Relative to the latent-leaf-only ablation, it reduces OPV by 61.0% and PCD by 62.8% while improving worst-path accuracy by 1.5 points. The aligned reduction of the training-time defect (PCD) and endpoint dispersion (OPV) is consistent with the proposed noncommutativity mechanism; the concurrent accuracy increase argues against obtaining low variance by making every path uniformly poor. The full-relabel oracle remains higher at 75.1% mean and 74.2% worst-path accuracy, delimiting the information still missing under evolving supervision.

7 SCOPE AND FAILURE MODES

The method assumes declared and internally consistent ontology maps. It does not infer missing semantic correspondences or repair contradictory taxonomies. Virtual updates add training cost, which is not quantified in Table 1. The evidence is limited to CIFAR-100 superclass/subclass paths and the evaluated legal edit families; arbitrary sample-order changes are not ontology-path interventions, and a low OPV does not imply invariance to them. Moreover, the sampled two-block PCD is a training diagnostic rather than a guarantee of path independence for unseen edit families.

Reproducibility Statement

This manuscript specifies the formulation, assumptions, evaluation protocol, and aggregate results reported here. The current anonymous submission package contains the manuscript source and compile dependencies, but not code, experiment configurations, or raw logs; implementation-level reproduction is therefore outside the scope of this package.

Ethics and Broader Impact

Changing taxonomies can encode institutional or social decisions rather than purely technical refinements. Endpoint consistency should therefore not be interpreted as endorsing a particular ontology or preserving historically biased categories. Applications require domain review of the declared mappings and separate evaluation of errors across affected groups and label transitions.

8 Conclusion

Endpoint-equivalent ontology histories can still produce different learned functions. Ontology Path Consistency separates this dependence into semantic transport to shared latent leaves and functional noncommutativity between legal update orders. The latent-leaf-only comparison and the paired reductions in PCD and OPV show that explicit endpoint-order consistency contributes beyond hierarchy-aware representation, while mean and worst-path accuracy improve together. The paper establishes ontology-path dependence as a measurable learning failure and provides a mechanism that directly targets it.

AI Use Statement

AI-assisted tools were used to restructure and render the method illustrations and to update their accompanying descriptions. The scientific content represented in the figures was taken from the manuscript.

References

Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: A strong, simple baseline. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/b704ea2c39778f07c617f6b7ce480e9e-Abstract.html.

Eleonora Giunchiglia and Thomas Lukasiewicz. Coherent hierarchical multi-label classification networks. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6dd4e10e3296fa63738371ec0d5df818-Abstract.html.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2018. doi: 10.1109/TPAMI.2017.2773081. URL https://doi.org/10.1109/TPAMI.2017.2773081.

Zhiqiu Lin, Deepak Pathak, Yu-Xiong Wang, Deva Ramanan, and Shu Kong. Continual learning with evolving class ontologies. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0557. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/3255a7554605a88800f4e120b3a929e1-Abstract-Conference.html.

Emanuele Marconato, Gianpaolo Bontempo, Elisa Ficarra, Simone Calderara, Andrea Passerini, and Stefano Teso. Neuro-symbolic continual learning: Knowledge, reasoning shortcuts and concept rehearsal. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23915–23936. PMLR, 2023. URL https://proceedings.mlr.press/v202/marconato23a.html.

Shishir Muralidhara, Saqib Bukhari, Georg Schneider, Didier Stricker, and René Schuster. CLEO: Continual learning of evolving ontologies. In *Computer Vision – ECCV 2024*, volume 15112 of *Lecture Notes in Computer Science*, pp. 328–344. Springer, 2025. doi: 10.1007/978-3-031-72949-2_19. URL https://doi.org/10.1007/978-3-031-72949-2_19.

Jonatas Wehrmann, Ricardo Cerri, and Rodrigo Barros. Hierarchical multi-label classification networks. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5075–5084. PMLR, 2018. URL https://proceedings.mlr.press/v80/wehrmann18a.html.