Proceedings Paper:
Pereira, E.M., Guimarães, F.G. and Dos Santos, J.A. (2025) MO-SHW: hierarchy-aware multi-objective optimization for open-world segmentation. In: 36th British Machine Vision Conference 2025, {BMVC} 2025, Sheffield, UK, November 24-27, 2025. 36th British Machine Vision Conference 2025, (BMVC 2025), 24-27 Nov 2025, Sheffield, UK. BMVA.

MO-SHW: Hierarchy-Aware Multi-Objective Optimization for Open-World Segmentation

Erico M. D. A Pereira[1]
emarco.pereira@dcc.ufmg.br
Frederico G. Guimarães[1]
fredericoguimaraes@ufmg.br
Jefersson A. dos Santos[2]
jefersson@ieee.org

[1] Computer Science Department
Universidade Federal de Minas Gerais
Belo Horizonte - MG, Brazil

[2] School of Computer Science
University of Sheffield
Sheffield, United Kingdom

Abstract

The exploitation of hierarchical information by vision models has shown significant benefits in various segmentation tasks. However, this remains largely unexplored in open-world scenarios, where models must cope with unknown, evolving, and underrepresented labeled class spaces. Most existing hierarchy-aware segmentation approaches are not readily applicable to open-world settings. This is primarily because they rely on architectural modifications that are incompatible with the design constraints of open-world models. Moreover, hierarchy-aware losses are challenging to integrate into such pipelines, as they often conflict with task-specific objectives and exacerbate optimization complexity in already multi-objective training environments. In this work, we demonstrate that hierarchy-aware losses can be effectively leveraged in open-world models when optimized under a multi-objective learning framework. Specifically, we show that gradient-based multi-objective optimization methods, such as multi-objective gradient descent (MOGD), are well-suited for jointly optimizing hierarchical and task-specific objectives, leading to better overall performance. To support this, we propose SHW, a novel hierarchy-aware loss function based on the Wasserstein distance. SHW is lightweight, model-agnostic, and encourages intra-class compactness and inter-class separation across multiple semantic levels. The integration of SHW with MOGD yields a general, model-agnostic framework that enables the effective exploitation of semantic hierarchies in open-world segmentation tasks, improving the performance of several recent methods.

1 Introduction

In computer vision tasks, the ability to understand and interpret a scene goes beyond recognizing individual objects. True visual understanding involves identifying the semantic relationships between the objects present in the scene. These relationships encompass several aspects, such as spatial positioning (how objects are arranged or oriented relative to each other), functional interactions (the roles or actions that entities perform or are capable of performing), and categorical associations (how objects are grouped based on similar or shared attributes). Such relationships provide context, enabling a more refined and accurate interpretation of visual data. They often appear as hierarchies, which are structures that gradually connect broad, generic categories to increasingly detailed and specific concepts. Understanding the context of hierarchical relationships is fundamental in semantic segmentation tasks, where images are partitioned into regions labeled with different object classes. As the system learns typical spatial arrangements and inter-object relationships, it becomes capable of producing more precise segmentation boundaries, with more accurate and less uncertain label assignments.

In open-world tasks [24, 66], systems must adapt to an unpredictable and constantly changing environment. In such scenarios, models must simultaneously address the following challenges: handling imbalanced data distributions (imbalanced learning [46]), effectively learning underrepresented classes (few-shot learning [51]), detecting samples from unknown classes (open-set recognition [42, 59]), identifying what those unknown classes are (novel class discovery [53]), and incrementally learning new classes without forgetting previously acquired knowledge (class-incremental learning [64]). By exploring how classes are hierarchically related, it becomes possible to share knowledge among them, allowing models to leverage previously learned patterns from other classes to make more informed predictions under limited data conditions. This knowledge transfer benefits the learning of minority and few-shot classes, since the model can exploit common features of a broader category to infer about a specific class, as well as the learning of incremental classes, as the previously acquired general knowledge can assist in assimilating new related concepts. Furthermore, the understanding how classes relate to each other enables the model to construct embeddings that are coherent with semantic relationships, bringing semantically related classes closer together and pushing unrelated ones apart, thus facilitating the recognition of pixels belonging to unknown labels and the identification of novel classes.

Although some works in the semantic segmentation literature have explored semantic hierarchies [27, 28, 32], most of these methods are designed for closed-world problems, in which all classes are previously known and well-represented, limiting their applicability in real-world scenarios. Consequently, such approaches struggle to address open-world challenges, which involve the presence of unknown or underrepresented classes and the need for incrementally learning new classes. Existing open-world methods either overlook hierarchical structures or address only isolated challenges. To the best of our knowledge, the exploitation of class hierarchies remains unexplored in comprehensive open-world scenarios.

According to some of the most effective approaches [30, 31, 40] of hierarchical learning in closed-world settings, one of the most common and effective ways to exploit semantic hierarchies is through the use of hierarchy-aware losses. However, incorporating them into existing open-world pipelines presents several challenges: I) Typically, open-world models already include several loss terms. Adding another one increases training complexity and makes it more difficult to achieve consistent results, both in terms of maintaining strong performance on the main task and in effectively acquiring the knowledge provided by the hierarchical loss. II) The hierarchical losses proposed for semantic segmentation are incompatible with most open-world models, as they require specific modifications in the model architecture that are prohibitive of being made in most open-world pipelines [27], such as changing the meaning or structure of their output, or are conflicting with their main loss objectives [40].

Toward overcoming challenge I, we investigate the idea of handling the several loss terms of the open-world model as different objectives and employ model optimization through multi-objective gradient descent, instead of aggregating them in a single objective (a.k.a.

scalarization technique [22, 26]) as done by most works of the deep learning literature. Our motivation is grounded in insights from the multi-objective optimization literature [11, 21, 45], which demonstrate that jointly optimizing multiple objectives enables a more effective exploration of the solution space, leading to outcomes that better satisfy each objective individually. Based on this, we hypothesize that adopting a multi-objective framework can enhance the optimization of hierarchical-aware losses without hindering, but also improving the optimization of the primary objectives already present in open-world models.

To address challenge II, we introduce SHW (Semantic Hierarchy-Aware Wasserstein), a novel loss function designed to enhance the hierarchical structure of learned feature representations. SHW explicitly enforces inter-class separation and intra-class compactness across multiple semantic levels, encouraging a structured embedding space where pixel representations align with the underlying class hierarchy. By operating directly on the output of the feature extractor, SHW remains agnostic to the model architecture, requiring no alterations to the model architecture. This design ensures seamless integration into existing open-world learning pipelines, regardless of the task-specific loss functions already in use, thereby preserving compatibility. Furthermore, SHW leverages the Wasserstein distance [41, 58] to compare class distributions. Unlike traditional distance metrics, the Wasserstein formulation captures both the magnitude and the structural geometry of differences in the feature space. This is particularly beneficial in open-world scenarios, where data distributions are non-stationary, imbalanced, and subject to the emergence of previously unseen classes. The use of Wasserstein distance allows SHW to be sensitive to these complexities, providing more stable gradients, improved generalization to underrepresented or novel classes, and better robustness to distributional shifts [35, 37], all of which are critical for scalable and adaptive open-world systems.

Through the integration of our solutions for challenges I and II, we construct a unified framework (MO-SHW) that enables open-world models to effectively leverage semantic class hierarchies. Empirical results show that our framework effectively improves closed-world segmentation and consistently enhances the performance of open-world models yielding significant improvements for state-of-the-art methods in the sub-tasks of open-set semantic segmentation and few-shot class-incremental semantic segmentation.
Summary of contributions:

- i We are the first to enable the effective exploitation of class hierarchies in comprehensive open-world semantic segmentation, resulting in consistent improvements in performance across multiple sub-tasks.

- ii We introduce a model-agnostic hierarchy-aware representation learning framework that augments open-world semantic segmentation pipelines with awareness of semantic hierarchy. The framework offers seamless *plug-and-play* integration, requires no modifications to the model architecture, ensures ease of adoption, and introduces only minor computational and memory overhead.

- iii We show that employing gradient-based multi-objective optimization significantly improves training outcomes when using hierarchy-aware loss functions, by better balancing competing objectives.

- iv We propose **SHW**, a novel Wasserstein-based hierarchy-aware loss function that enforces hierarchical consistency among pixel embeddings in open-world models.

2 Background and Related Work

2.1 Open-World Semantic Segmentation

In real-world scenarios, systems usually operate in open long-tailed data scenarios [38], where they have to continuously adapt and learn from limited and unseen data. When these challenges are simultaneously encountered in the same scenario, they are studied under the umbrella of Open World Recognition [38, 66] - a paradigm in machine learning that aims to overcome the limitations of closed-world models by enabling them to recognize when faced with something unknown and incorporate this new information into their learning process. Ideally, open-world models must deal with several subtasks, such as imbalanced learning [7], few-shot learning [10], open-set recognition [42], novel class discovery [63] and class-incremental learning [60]. However, in the context of semantic segmentation, only a few methods have been proposed to jointly handle more than one subtask. Some methods [8, 16] separate the open-world problem in three sequential steps: (1) Open-Set Semantic Segmentation (6.1.1), to classify pixels from known classes and recognize unknown; (2) Manual labeling of unknown pixels (with human supervision), to group them in novel classes; and (3) Few-Shot Class-Incremental Semantic Segmentation (6.1.2), to learn these novel classes. Some other works [50, 54] replace step (2) with the use of automatic learning strategies to identify novel classes, a.k.a. Novel Class Discovery. **Open-World methods exploiting class hierarchies.** While some existing methods leverage hierarchical relationships among classes to address open-world challenges, they are tailored to tackle only specific subtasks in isolation. For example, certain incremental learning approaches [20, 25] utilize class hierarchy information to facilitate knowledge expansion. However, these methods are generally evaluated on closed-world benchmarks and do not address key open-world challenges such as the handling of underrepresented or unknown classes. Our proposed methodology is task-agnostic and seamlessly integrates into existing open-world pipelines providing plug-and-play exploitation of semantic class hierarchies to a wide range of models and subtasks. To the best of our knowledge, our work is the first to provide hierarchy-aware learning across a broader spectrum of subtasks in the open world.

2.2 Hierarchy-Aware Semantic Segmentation

Several works in the semantic segmentation literature have been designed with the purpose of exploring the information of class hierarchies. They usually exploit hierarchical information through modifications in the network architecture, specific loss objectives, or learning embeddings with hierarchical coherence. As they were typically designed for closed-world problems, they are not properly equipped to effectively operate in more realistic data scenarios. **Hierarchy-Aware Architectures** [32, 33, 39, 56] encode the class hierarchy directly into the network structure. They typically construct structured neural modules, which replicate the hierarchy semantics, to replace the class-agnostic segmentation head of existing standard segmentation models, producing a hierarchy-aligned architecture. This often causes significant architectural modifications, hindering generality as the model becomes more specific and complex, and leading to scalability issues, as the number of parameters usually increases with the size of the hierarchy. **Hierarchy-Aware Losses** [40] are designed to incorporate the class hierarchy into the learning process by directly penalizing predictions that violate hierarchical relationships. These objectives are parameterized by the class hierarchy, ensuring that the model's predictions respect the structured relationships between classes. However, incorporating hierarchical losses can introduce additional complexity to already intricate pipelines, such as those designed for open-world problems, burdening models that already feature multiple loss terms, complicating the training process, and consequently hindering performance improvements. **Hierarchy-Aware Embeddings** [27, 30, 31] aim to capture multiple levels of class abstraction within a single embedding representation, ensuring that similar or related classes are positioned closer together in the embedding space, thus reflecting their hierarchical relationships. Specifically, some approaches [27] parameterize the hierarchical embedding space using hierarchical models, which intuitively organize embeddings to mirror the structure of the class hierarchy. These models are often computationally intensive and require adjustments in the network architecture to accommodate different hierarchies. This leads to increased complexity and resource demands, making them less practical for large-scale applications. **Employing Hierarchy-Aware Approaches in Open-World Pipelines**. Although the closed-world literature offers a variety of hierarchy-aware methods (*e.g.* loss functions) that are relatively easy to integrate and have shown promising result, these approaches are generally unsuitable for open-world segmentation models. For instance, HSSN* [30, 31] introduces the Focal Tree-Min Loss, which enforces hierarchy-consistent predictions, and the Tree-Triplet Loss, which imposes hierarchical separation margins on pixel embeddings across distinct semantic classes. However, both rely on a multi-label classification paradigm that requires architectural changes incompatible with most open-world segmentation frameworks. Likewise, HCE [40] proposes a hierarchy-aware cross-entropy formulation to account for inter-class hierarchical relations. However, in open-world settings, where label spaces may be incomplete, dynamic, or unknown, such objectives often conflict with the primary task losses, limiting their applicability. To the best of our knowledge, no current hierarchy-aware semantic segmentation method is directly compatible with open-world models without imposing significant constraints or introducing incompatibilities.

3 Methodology

With the goal of enabling the effective exploitation of semantic hierarchies in open-world segmentation models, we propose MO-SHW (Fig. 1) - a unified model optimization framework that integrates two complementary components: the SHW loss (Section 3.2) and a gradient-based multi-objective optimization (Section 3.1). This integration combines the structural benefits of SHW, which enforces hierarchical organization in pixel embeddings with seamless adaptability, with the broader exploratory capacity of multi-objective optimization, which allows for a more balanced and principled trade-off among competing objectives. By employing gradient-balancing methods, our approach ensures the simultaneous and efficient optimization of both task-specific and auxiliary hierarchy-aware losses. The final framework is model-agnostic, lightweight, and easily integrable into existing open-world pipelines. As demonstrated in our experiments, this joint formulation improves segmentation performance across several open-world methods and different subtasks.

3.1 Multi-Objective Optimization for Hierarchy-Aware Open-World Models

Deep learning models are traditionally trained using single-objective loss functions that aggregate multiple objectives into a single scalar value through weighted sums - a method

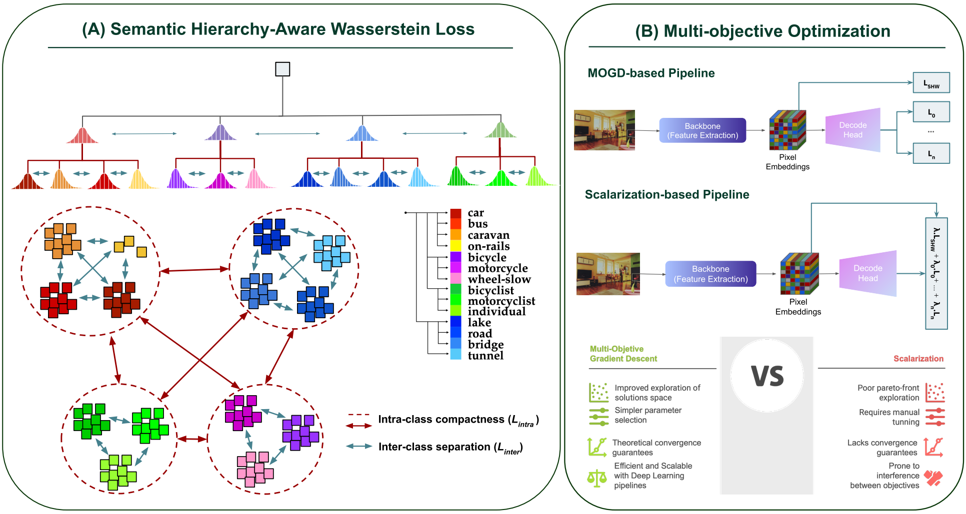

Figure 1: Overview of the proposed MO-SHW methodology. It comprises two key components: (A) the Semantic Hierarchical Wasserstein (SHW) loss (Sec.3.2), which operates directly on pixel class distributions explicitly enforcing inter-class separation and intra-class compactness across multiple semantic levels to encourage a structured embedding space where pixel representations are aligned with the underlying class hierarchy; and (B) a multi-objective optimization strategy (Sec. 3.1) based on Gradient Descent (MOGD). Unlike scalarization, which aggregates multiple objectives into a single weighted loss, MOGD treats each objective independently, offering several advantages in terms of solution-space exploration, parameter tuning, and convergence guarantees.

known as Scalarization [22, 26]. However, this approach has several limitations that can hinder effective model training in complex, multi-objective optimization scenarios. Scalarization often requires manual tuning of loss term weights, lacks convergence guarantees to Pareto-optimal solutions, and is prone to suboptimal trade-offs - especially in non-convex or conflicting objective settings [22]. It also struggles to identify "unsupported" yet valuable solutions and is susceptible to gradient interference between tasks, which can destabilize training and reduce optimization efficiency [14].

These limitations are especially critical in open-world scenarios, where models are required to simultaneously optimize a diverse set of objectives (*e.g.* segmentation accuracy, novelty detection, and semantic consistency). The challenge is further exacerbated when integrating additional objectives, such as hierarchy-aware losses, alongside task-specific losses. In such complex multi-objective settings, scalarization tends to oversimplify the optimization landscape, often leading to suboptimal solutions or unstable training dynamics.

To address these limitations, we reformulate model training as a multiobjective optimization (MOO) problem. Our hypothesis is that treating each objective as a distinct optimization target allows for a more principled and balanced training process. Rather than collapsing all objectives into a single function, a MOO formulation enables simultaneous optimization of multiple loss functions, which we expect to improve model optimization across both standard and hierarchy-informed evaluation metrics alongside with its overall segmentation performance.

Among the available MOO strategies (detailed in Appendix 6.1.3), we adopt a gradient-balancing algorithm due to its ability to effectively balance theoretical robustness with computational efficiency. Unlike Pareto-based algorithms, which are computationally intensive, or loss-weighting heuristics, which lack formal convergence guarantees, gradient-based approaches scale well with modern deep learning pipelines while providing theoretically grounded convergence to Pareto-stationary solutions [11]. Specifically, we adopt Aligned-MTL [47] as our gradient-balancing multi-objective optimization algorithm due to its strong empirical performance in recent comparative studies [34].

3.2 SHW: Semantic Hierarchy-aware Wasserstein Loss

Our goal is to design a loss function that encourages the model to learn feature representations aligned with the semantic structure of the class hierarchy, bringing embeddings of semantically similar classes closer together while pushing apart those of semantically distant classes. To ensure compatibility with a wide range of model architectures, the proposed loss should not depend directly on the model's output (e.g., predicted labels) or require any modifications to the architecture itself. To this end, we propose leveraging the Wasserstein distance D_w (Appendix 6.1.4), approximated via DbTSW [52], to act directly on pixel embeddings, reorganizing their structure in the feature space by comparing the distances between their associated class distributions. The use of this distance is particularly advantageous for open-world learning, as it enables the model to better capture semantic relationships [35, 36], handle uncertainty, and deal more effectively with novel or out-of-distribution inputs [13].

We introduce a contrastive learning objective, L_{SHW} (Eq. 3), designed to compare class distributions across multiple levels of a class hierarchy $\mathcal{H}$, guided by two core principles: promoting intra-class proximity (Eq. 1) and enforcing inter-class separation (Eq. 2). The first component, L_{intra}, encourages each class node v to move closer to its parent class v_p, while simultaneously pushing it away from all other non-parent classes $u \in \Gamma(v)$ at the same higher hierarchical level. The second component, L_{inter}, aims to increase the separation between v and its immediate siblings $s \in S(v)$, that is, the other child nodes sharing the same parent. The overall loss is defined as:

$$L_{intra}(v) = -\log \frac{\exp(-D_w(v, v_p))}{\sum_{u \in \Gamma(v)} \exp(-D_w(v, u))} \quad (1) \qquad L_{inter}(v) = -\log \frac{1}{\sum_{s \in S(v)} \exp(-D_w(v, s))} \quad (2)$$

$$L_{SHW} = \sum_{v \in V} \lambda_l(v) \left(\frac{1}{|\Gamma(v)|} L_{intra}(v) + \frac{1}{|S(v)|} L_{inter}(v) \right) \quad (3)$$

where λ_l is a scaling parameter that imposes a fixed penalty based on a node's position in the class hierarchy. This is grounded in the intuition that distances in the feature space should reflect the degree of semantic similarity: pairs of similar classes at deeper (more specific) levels of the hierarchy should be pulled closer together, while pairs at higher (more general) levels may tolerate greater separation. Accordingly, we apply stronger penalties to classes at lower levels of the hierarchy and weaker penalties to those at higher levels. In this sense, we define $\lambda_l(v) = \exp(1/(|L| - level(v)))$, where L is set of hierarchical levels, following the formulation shown to outperform alternative penalty schemes in prior work [61].

4 Experimental Evaluation

4.1 Multi-Objective Optimization for Hierarchy-Aware Losses

Table 1: Quantitative results of the comparison between model optimization strategies employing hierarchical-aware losses in the Cityscapes [12] dataset. Values in green indicate MOGD gains against Scalarization.

Objective Functions	Scalarization			Multi-objective GD		
	CE loss ↓	HA loss ↓	mIoU% ↑	CE loss ↓	HA loss ↓	mIoU% ↑
CE (flat model)	0.1891	—	80.97	—	—	—
CE + HCE [40] [VISIGRAPP'20]	0.1830	0.1957	81.09	0.1697 (-0.0133)	0.1763 (-0.0194)	**81.88** (+0.81)
CE + HSSN* [31] [TPAMI'23]	0.1755	0.3364	81.62	0.1659 (-0.0095)	0.3105 (-0.0259)	**82.09** (+0.47)
CE + SHW (Section 3.2)	0.1673	-2.1563	82.03	0.1431 (-0.0242)	-2.4216 (-0.2653)	**83.19** (+1.16)

This experiment aims to evaluate the effectiveness of our proposed multi-objective optimization strategy (described in Section 3.1) for training models with additional hierarchy-aware losses, compared to the conventional single-objective scalarization approach. We evaluate these optimization strategies using three distinct loss functions: HCE [40], a widely adopted formulation of hierarchy-aware loss in computer vision; HSSN* [31], currently one of the best-performing methods for hierarchy-aware semantic segmentation in closed-world settings; and our proposed SHW loss (see Section 3.2). **Experimental Setup:** Given that HCE and HSSN* are not readily adaptable to open-world models (as discussed in Section 2.2), this evaluation is conducted within a standard closed-world setting using the Cityscapes dataset and the DeepLabV3+ architecture with a ResNet-101 backbone trained for 90K iterations. See Appendix 6.2 for other details. **Quantitative Results:** As shown in Table 1, the multi-objective optimization strategy led to consistent and substantial improvements across all evaluated loss functions. These results demonstrate that the proposed approach can identify solutions that more effectively optimize the targeted objectives, all within the same training budget. The observed improvements in individual loss metrics align with insights from the multi-objective optimization literature [22], which suggest that simultaneously optimizing multiple objectives enables a more thorough exploration of the solution space, allowing superior trade-offs and better fulfillment of each objective. These findings provide strong empirical support for our hypothesis that multi-objective optimization not only enables the seamless incorporation of hierarchy-aware losses into open-world segmentation models, but also improves overall performance by jointly enhancing both auxiliary (hierarchical) and primary (task-specific) learning. objectives. *SHW in closed-world*: Although the primary objective of the proposed SHW loss is its employment in open-world tasks (Section 4.2), experimental results also reveal strong performance in closed-world settings. Specifically, SHW achieves a notable improvement of nearly 2% in mIoU under the multi-objective optimization framework when compared to standard flat segmentation. **Qualitative Results:** please see Appendix 6.3.2.

4.2 Evaluating the Proposed Framework in Open-World Pipelines

The following experiments are designed to evaluate the impact of our proposed framework (MO-SHW), which combines the SHW loss with a multi-objective optimization strategy using a gradient-balancing algorithm, on open-world semantic segmentation pipelines. To ensure a comprehensive evaluation, we adopt a standard protocol in the open-world semantic segmentation literature and conduct independent assessments for each of the key subtasks: open-set semantic segmentation and few-shot class-incremental semantic segmentation. A more detailed description of the experimental setups is provided in Appendix 6.2.

Table 2: Quantitative evaluation of open-set semantic segmentation on the StreetHarzards [19] dataset.

Method	Anomaly Segmentation		Closed-Set	Open-Set Segmentation		
	AuPRC↑	FPR_{95} ↓	mIoU↑	O-IoU_{t5} ↑	O-IoU_{t6} ↑	O-IoU↑
DML [8] [ICCV'21]	14.7	17.3	53.9	-	-	-
MSP [ICLR'17]	7.5	27.9	65.0	32.7	40.2	35.1
ODIN [ICLR'18]	7.0	28.7	65.0	26.4	33.9	28.8
OE [ICLR'19]	14.6	17.7	61.7	43.7	44.1	43.8
OOD-H [GCPR'19]	19.7	56.2	66.6	33.7	34.3	33.9
Energy [NurIPS'20]	12.9	18.2	63.3	41.7	44.9	42.7
ReAct [NurIPS'21]	10.9	21.2	62.7	33.0	36.2	34.0
OH*MSP [CoRR'21]	18.8	30.9	66.6	43.3	44.2	43.6
ML [ICML'22]	11.6	22.5	65.0	39.6	44.5	41.2
AEM [CVPRw'23]	30.7	99.7	71.3	35.3	54.6	41.4
RbA [ICCV'23]	50.1	96.9	73.2	10.4	12.1	10.9
DH [17] [ECCV'22]	30.2	13.0	63.0	46.1	45.3	45.8
DH + MO-SHW (ours)	**35.9**	**12.1**	**66.3**	**47.2**	**46.6**	**46.9**
M2A [44] [CVPR'23]	58.1	**14.9**	72.3	59.9	59.7	59.8
M2A + MO-SHW (ours)	**58.7**	15.3	**74.1**	**60.7**	**60.2**	**60.3**

Open-Set Semantic Segmentation (OSSS). For this subtask, we selected DenseHybrid [17] and Mask2-Anomaly [44], two of the most recent and best-performing methods in the literature, as base models. We integrated our framework into their respective pipelines and compared the resulting models against their original versions as well as other prominent baselines from the open-set segmentation literature. *Quantitative Results:* Table 2 presents the comparative results. Overall, the MO+SHW-augmented models consistently outperform their original counterparts and competing baselines, with pronounced improvements observed for the DenseHybrid variant. The anomaly segmentation metrics reveal that incorporating MO-SHW enhances the models' ability to detect out-of-distribution pixels. Specifically, DenseHybrid benefited from a notable reduction in false positives, while Mask2Anomaly experienced a marginal increase. In line with the observations of Section 4.1, the proposed framework substantially improved performance in closed-world sem. segmentation, suggesting improved dense prediction capabilities. Furthermore, the open-set recognition metrics indicate that MO-SHW introduces modest but meaningful gains in the models' ability to identify unknown classes. These findings confirm that MO-SHW effectively boosts the general predictive performance of models in open-set semantic segmentation tasks, by simultaneously enhancing the recognition of both known (closed-set) and unknown (open-set) categories.

Few-Shot Class-Incremental Semantic Segmentation (FSCISS). We integrated the our framework into the pipelines of two traditional methods for this task: PIFS [9] and GAPS [43]. The performance of the resulting MO-SHW-augmented models was then compared against their original versions as well as other notable approaches from the FSCILSS literature. *Quantitative Results:* As reported in Table 3, the MO-SHW-extended models consistently outperform their respective baselines across all evaluation metrics. Notably, improvements in $mIoU_{new}$ demonstrate that our approach enhances the models' capacity to learn novel

Table 3: Quantitative evaluation for few-shot class-incremental semantic segmentation in the PASCAL-5^i [1] dataset.

Method	1-shot			5-shot		
	$\text{mIoU}_{base}\uparrow$	$\text{mIoU}_{new}\uparrow$	HM↑	$\text{mIoU}_{base}\uparrow$	$\text{mIoU}_{new}\uparrow$	HM↑
Fine-Tunning	47.2	3.9	7.2	58.7	7.7	13.6
EHNet [49] [ACM-MM'22]	68.4	18.8	29.1	67.9	31.3	43.5
CaLNet [48] [ACM-MM'23]	74.2	17.4	28.2	74.7	30.1	42.9
SRAA [65] [MMM'24]	66.4	18.8	29.3	64.3	28.7	39.7
MBCL [62] [ISCAS'24]	65.0	19.2	29.6	66.1	28.5	40.0
PIFS [9] [BMVC'21]	64.1	16.9	26.7	64.5	27.5	38.6
PIFS + MO-SHW (ours)	**65.9**	**18.5**	**28.8**	**65.3**	**29.2**	**39.9**
PIFS + GAPS [43] [CVPRw'23]	66.8	23.6	34.7	68.2	43.9	53.4
PIFS + GAPS + MO-SHW (ours)	**68.5**	**25.6**	**37.2**	**68.9**	**46.4**	**55.4**

classes with limited supervision. Concurrently, gains in $\text{mIoU}_{\text{base}}$ indicate improved retention of previously acquired knowledge, suggesting that MO-SHW not only facilitates the acquisition of new concepts but also mitigates the effects of catastrophic forgetting. These results highlight the effectiveness of the MO-SHW framework in supporting continual learning under few-shot constraints, thereby improving both adaptability to novel information and stability of prior knowledge in open-world segmentation scenarios.

5 Conclusion

In this paper, we address the challenge of incorporating hierarchical understanding into open-world semantic segmentation. We proposed a novel framework that enables supervised open-world models to effectively leverage semantic class hierarchies, thereby improving both the robustness and generalization of their predictions. Our framework introduces two key contributions: (i) a novel and adaptable Wasserstein-based hierarchy-aware loss function (SHW) which enforces hierarchical consistency by promoting intra-class similarity and inter-class separation across multiple levels of abstraction without architecture modifications; and (ii) the integration of this loss into a multi-objective optimization paradigm using gradient-balancing methods, allowing simultaneous and balanced optimization of hierarchical and task-specific objectives. Our approach has several advantages: it is model-agnostic, seamlessly integrates into existing pipelines, introduces minimal computational overhead, and maintains compatibility with the training dynamics of open-world segmentation tasks. Through experiments on several benchmarks, we empirically demonstrated that our framework consistently enhances the performance of leading open-world segmentation models, including recent methods for open-set segmentation and few-shot class-incremental learning. These results show that semantic hierarchies offer a valuable source of structural prior knowledge, which, when properly integrated, can significantly improve open-world reasoning in visual scene understanding.

References

[1] Shaban Amirreza, Bansal Shray, Liu Zhen, Essa Irfan, and Boots Byron. One-shot learning for semantic segmentation. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2017.

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. In *International Conference on Machine Learning (ICML)*, 2017.

[3] Arturs Backurs, Yihe Dong, Piotr Indyk, Ilya Razenshteyn, and Tal Wagner. Scalable nearest neighbor search for optimal transport. In *International Conference on machine learning*. PMLR, 2020.

[4] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Simultaneous semantic segmentation and outlier detection in presence of domain shift. In *Pattern Recognition: 41st DAGM German Conference*, 2019.

[5] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Dense open-set recognition based on training with noisy negative images. *Image and vision computing*, 2022.

[6] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 2015.

[7] Patrik Olã Bressan, José Marcato Junior, José Augusto Correa Martins, Maximilian Jaderson de Melo, Diogo Nunes Gonçalves, Daniel Matte Freitas, Ana Paula Marques Ramos, Michelle Taís Garcia Furuya, Lucas Prado Osco, Jonathan de Andrade Silva, et al. Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping. *International Journal of Applied Earth Observation and Geoinformation*, 2022.

[8] Jun Cen, Peng Yun, Junhao Cai, Michael Yu Wang, and Ming Liu. Deep metric learning for open world semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2021.

[9] Fabio Cermelli, Massimiliano Mancini, Yongqin Xian, Zeynep Akata, and Barbara Caputo. Prototype-based incremental few-shot segmentation. In *The 32nd British Machine Vision Conference*, 2021.

[10] Zhaobin Chang, Yonggang Lu, Xingcheng Ran, Xiong Gao, and Xiangwen Wang. Few-shot semantic segmentation: a review on recent approaches. *Neural Computing and Applications*, 2023.

[11] Weiyu Chen, Xiaoyuan Zhang, Baijiong Lin, Xi Lin, Han Zhao, Qingfu Zhang, and James T Kwok. Gradient-based multi-objective deep learning: Algorithms, theories, applications, and beyond. *arXiv preprint arXiv:2501.10945*, 2025.

[12] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[13] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[14] Stanford CS224N, Default Project, . Mentor, and Manasi Sharma. Unitary scalarization or gradient surgery? best practices for multitask fine-tuning. 2023.

[15] M Cuturi. Lightspeed computation of optimal transportation distances. *Advances in Neural Information Processing Systems*, 2013.

[16] Hexin Dong, Zifan Chen, Mingze Yuan, Yutong Xie, Jie Zhao, Fei Yu, Bin Dong, and Li Zhang. Region-aware metric learning for open world semantic segmentation via meta-channel aggregation. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, 2022.

[17] Matej Grcić and Siniša Šegvić. Hybrid open-set segmentation with synthetic negative data. *IEEE transactions on pattern analysis and machine intelligence*, 2024.

[18] Matej Grcić, Josip Šarić, and Siniša Šegvić. On advantages of mask-level recognition for outlier-aware segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[19] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joseph Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. In *International Conference on Machine Learning*, 2022.

[20] Julia Hindel, Daniele Cattaneo, and Abhinav Valada. Taxonomy-aware continual semantic segmentation in hyperbolic spaces for open-world perception. *IEEE Robotics and Automation Letters*, 2024.

[21] S. S. Hotegni, M. Berkemeier, and Sebastian Peitz. Multi-objective optimization for sparse deep multi-task learning. *2024 International Joint Conference on Neural Networks (IJCNN)*, 2023.

[22] Yuzheng Hu, Ruicheng Xian, Qilong Wu, Qiuling Fan, Lang Yin, and Han Zhao. Revisiting scalarization in multi-task learning: A theoretical perspective. *ArXiv*, 2023.

[23] Chengjia Jiang, Tao Wang, Sien Li, Jinyang Wang, Shirui Wang, and Antonios Antoniou. Few-shot class-incremental semantic segmentation via pseudo-labeling and knowledge distillation. In *2023 4th International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*, 2023.

[24] Mayank Kejriwal, Eric Kildebeck, Robert Steininger, and Abhinav Shrivastava. Challenges, evaluation and opportunities for open-world learning. *Nature Machine Intelligence*, 2024.

[25] Hyoseo Kim and Junsuk Choe. Weakly-supervised incremental learning for semantic segmentation with class hierarchy. *Pattern Recognition Letters*, 2024.

[26] Vitaly Kurin, Alessandro De Palma, Ilya Kostrikov, Shimon Whiteson, and Pawan K Mudigonda. In defense of the unitary scalarization for deep multi-task learning. *Advances in Neural Information Processing Systems*, 2022.

[27] Hyeongjun Kwon, Jinhyun Jang, Jin Kim, Kwonyoung Kim, and Kwanghoon Sohn. Improving visual recognition with hyperbolical visual hierarchy mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[28] Loic Landrieu and Vivien Sainte Fare Garnot. Leveraging class hierarchies with metric-guided prototype learning. In *British Machine Vision Conference (BMVC)*, 2021.

[29] Tam Le, Makoto Yamada, Kenji Fukumizu, and Marco Cuturi. Tree-sliced variants of wasserstein distances. *Advances in neural information processing systems*, 2019.

[30] Liulei Li, Tianfei Zhou, Wenguan Wang, Jianwu Li, and Yi Yang. Deep hierarchical semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[31] Liulei Li, Wenguan Wang, Tianfei Zhou, Ruijie Quan, and Yi Yang. Semantic hierarchy-aware segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[32] Zhiheng Li, Wenxuan Bao, Jiayang Zheng, and Chenliang Xu. Deep grouping model for unified perceptual parsing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[33] Xiaodan Liang, Hongfei Zhou, and Eric Xing. Dynamic-structured semantic propagation network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.

[34] Baijiong Lin, Weisen Jiang, Feiyang Ye, Yu Zhang, Pengguang Chen, Ying-Cong Chen, Shu Liu, and James T Kwok. Dual-balancing for multi-task learning. *arXiv preprint arXiv:2308.12029*, 2023.

[35] Xiaofeng Liu, Yuzhuo Han, Song Bai, Yi Ge, Tianxing Wang, Xu Han, Jane You, and Jun Lu. Importance-aware semantic segmentation in self-driving with discrete wasserstein training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

[36] Xiaofeng Liu, Wenxuan Ji, Jundong You, Georges Fakhri, and Jonghye Woo. Severity-aware semantic segmentation with reinforced wasserstein training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[37] Xiaofeng Liu, Yunhong Lu, Xiongchang Liu, S. Bai, Site Li, and J. You. Wasserstein loss with alternative reinforcement learning for severity-aware semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[38] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X Yu. Open long-tailed recognition in a dynamic world. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[39] Panagiotis Meletis and Gijs Dubbelman. Training of convolutional networks on multiple heterogeneous datasets for street scene semantic segmentation. In *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018.

[40] Bruce R Muller and William AP Smith. A hierarchical loss for semantic segmentation. In *VISIGRAPP*, 2020.

[41] Kimia Nadjahi. *Sliced-Wasserstein distance for large-scale machine learning: theory, methodology and extensions*. PhD thesis, Institut polytechnique de Paris, 2021.

[42] Ian Nunes, Camila Laranjeira, Hugo Oliveira, and Jefersson A dos Santos. A systematic review on open-set segmentation. *Computers & Graphics*, 2023.

[43] Ri-Zhao Qiu, Peiyi Chen, Wangzhe Sun, Yu-Xiong Wang, and Kris Hauser. Gaps: Few-shot incremental semantic segmentation via guided copy-paste synthesis. *CVPRW*, 2023.

[44] Shyam Nandan Rai, Fabio Cermelli, Barbara Caputo, and Carlo Masone. Mask2anomaly: Mask transformer for universal open-set segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[45] J. D. Rojas, O. Arrieta, and Ramón Vilanova. Multiobjective optimization. *Industrial PID Controller Tuning*, 2021.

[46] Manisha Saini and Seba Susan. Tackling class imbalance in computer vision: a contemporary review. *Artificial Intelligence Review*, 2023.

[47] Dmitry Senushkin, Nikolay Patakin, Arseny Kuznetsov, and Anton Konushin. Independent component alignment for multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[48] Leo Shan, Wenzhang Zhou, and Grace Zhao. Incremental few shot semantic segmentation via class-agnostic mask proposal and language-driven classifier. In *Proceedings of the 31st ACM International Conference on Multimedia*, 2023.

[49] Guangchen Shi, Yirui Wu, Jun Liu, Shaohua Wan, Wenhai Wang, and Tong Lu. Incremental few-shot semantic segmentation via embedding adaptive-update and hyper-class representation. In *Proceedings of the 30th ACM international conference on multimedia*, 2022.

[50] Matteo Sodano, Federico Magistri, Lucas Nunes, Jens Behley, and Cyrill Stachniss. Open-world semantic segmentation including class similarity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[51] Yisheng Song, Ting Wang, Puyu Cai, Subrota K Mondal, and Jyoti Prakash Sahoo. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys*, 2023.

[52] Hoang V. Tran, Minh-Khoi Nguyen-Nhat, Huyen Trang Pham, Thanh Chu, Tam Le, and Tan Minh Nguyen. Distance-based tree-sliced Wasserstein distance. In *International Conference on Learning Representations*, 2025.

[53] Colin Troisemaine, Vincent Lemaire, Stéphane Gosselin, Alexandre Reiffers-Masson, Joachim Flocon-Cholet, and Sandrine Vaton. Novel class discovery: an introduction and key concepts. *arXiv preprint arXiv:2302.12028*, 2023.

[54] Svenja Uhlemeyer, Matthias Rottmann, and Hanno Gottschalk. Towards unsupervised open world semantic segmentation. In *Uncertainty in Artificial Intelligence*, 2022.

[55] Cédric Villani. *Optimal Transport: Old and New*. Springer, 2008.

[56] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. In *Proceedings of the European conference on computer vision (ECCV)*, 2018.

[57] Makoto Yamada, Yuki Takezawa, Ryoma Sato, Han Bao, Zornitsa Kozareva, and Sujith Ravi. Approximating 1-wasserstein distance with trees. *Transactions on Machine Learning Research*, 2022.

[58] Makoto Yamada, Yuki Takezawa, Guillaume Houry, Kira Michaela Düsterwald, Deborah Sulem, Han Zhao, and Yao-Hung Tsai. An empirical study of self-supervised learning with wasserstein distance. *Entropy*, 2024.

[59] Zhen Yang, Jun Yue, Pedram Ghamisi, Shiliang Zhang, Jiayi Ma, and Leyuan Fang. Open set recognition in real world. *International Journal of Computer Vision*, 2024.

[60] Bo Yuan and Danpei Zhao. A survey on continual semantic segmentation: Theory, challenge, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[61] Shu Zhang, Ran Xu, Caiming Xiong, and Chetan Ramaiah. Use all the labels: A hierarchical multi-label contrastive learning framework. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022.

[62] Yuxuan Zhang, Miaojing Shi, Taiyi Su, and Hanli Wang. Memory-based contrastive learning with optimized sampling for incremental few-shot semantic segmentation. In *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2024.

[63] Yuyang Zhao, Zhun Zhong, Nicu Sebe, and Gim Hee Lee. Novel class discovery in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022.

[64] Da-Wei Zhou, Qi-Wei Wang, Zhi-Hong Qi, Han-Jia Ye, De-Chuan Zhan, and Ziwei Liu. Class-incremental learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[65] Yuan Zhou, Xin Chen, Yanrong Guo, Jun Yu, Richang Hong, and Qi Tian. Advancing incremental few-shot semantic segmentation via semantic-guided relation alignment and adaptation. In *International Conference on Multimedia Modeling*, 2024.

[66] Fei Zhu, Shijie Ma, Zhen Cheng, Xu-Yao Zhang, Zhaoxiang Zhang, and Cheng-Lin Liu. Open-world machine learning: A review and new outlooks. *arXiv preprint arXiv:2403.01759*, 2024.