# CLIMB-3D: Class-Incremental Imbalanced 3D Instance Segmentation

Vishal Thengane[1]
v.thengane@surrey.ac.uk

Jean Lahoud[2]
jean.lahoud@mbzuai.ac.ae

Hisham Cholakkal[2]
hisham.cholakkal@mbzuai.ac.ae

Rao Muhammad Anwer[2]
rao.anwer@mbzuai.ac.ae

Lu Yin[1]
l.yin@surrey.ac.uk

Xiatian Zhu[1]
xiatian.zhu@surrey.ac.uk

Salman Khan[2, 3]
salman.khan@mbzuai.ac.ae

[1] University of Surrey,
Guildford, UK

[2] Mohamed bin Zayed University of
Artificial Intelligence,
Abu Dhabi, UAE

[3] Australian National University,
Canberra, Australia

## Abstract

While 3D instance segmentation (3DIS) has advanced significantly, most existing methods assume that all object classes are known in advance and uniformly distributed. However, this assumption is unrealistic in dynamic, real-world environments where new classes emerge gradually and exhibit natural imbalance. Although some approaches address the emergence of new classes, they often overlook class imbalance, which leads to suboptimal performance, particularly on rare categories. To tackle this, we propose **CLIMB-3D**, a unified framework for **CL**ass-incremental **Imb**alance-aware **3D**IS. Building upon established exemplar replay (ER) strategies, we show that ER alone is insufficient to achieve robust performance under memory constraints. To mitigate this, we introduce a novel pseudo-label generator (PLG) that extends supervision to previously learned categories by leveraging predictions from a frozen model trained on prior tasks. Despite its promise, PLG tends to be biased towards frequent classes. Therefore, we propose a class-balanced re-weighting (CBR) scheme that estimates object frequencies from pseudo-labels and dynamically adjusts training bias, without requiring access to past data. We design and evaluate three incremental scenarios for 3DIS on the challenging ScanNet200 dataset and additionally validate our method for semantic segmentation on ScanNetV2. Our approach achieves state-of-the-art results, surpassing prior work by up to 16.76% mAP for instance segmentation and approximately 30% mIoU for semantic segmentation, demonstrating strong generalisation across both frequent and rare classes. Code is available at: `https://github.com/vgthengane/CLIMB3D`.

# 1 Introduction

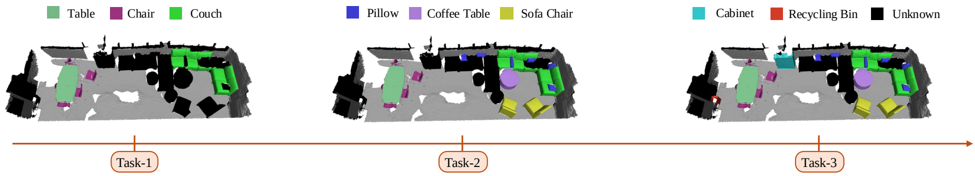


Figure 1: **Overview of the CI-3DIS setting.** New object categories are introduced incrementally with each task. After every phase, the model must recognise both newly added and previously learned classes. For instance, in Task 2, categories such as *Pillow*, *Coffee Table*, and *Sofa Chair* are introduced, while the model is expected to retain recognition of earlier classes like *Table*, *Chair*, and *Couch*.

3D instance segmentation (3DIS) is a fundamental task in computer vision that involves identifying precise object boundaries and class labels in 3D space, with broad applications in graphics, robotics, and augmented reality [4, 39]. Traditional 3DIS methods, including top-down [28, 70, 78], bottom-up [26, 73], and transformer-based approaches [59], perform well under the assumption that all classes are available and balanced during training. However, this assumption does not hold in real-world environments, where new object classes emerge over time and exhibit natural imbalance.

This gap motivates the need for Class-Incremental Learning (CIL) [17, 64], which supports learning new categories while retaining knowledge of previous ones [51]. While CIL has seen success in 2D image tasks [1, 43, 54, 60], extensions to 3D point clouds remain limited, often focusing on object-level classification [13, 18, 48]. Recent works on scene-level class-incremental 3DIS (CI-3DIS) [4] and class-incremental semantic segmentation (CI-3DSS) [74] show promise, but depend heavily on large exemplar [56] and overlook class imbalance, limiting their practicality.

To address this, we propose **CLIMB-3D**, a unified framework for **CL**ass-incremental **IMB**alance-aware **3D**IS that jointly tackles catastrophic forgetting and class imbalance. The framework begins with Exemplar Replay (ER), storing a small number of samples from past classes for later replay. However, this alone does not yield promising results under the strict memory constraints required in CI-3DIS. To address this, we introduce a Pseudo-Label Generator (PLG), which uses a frozen model from previous task to generate supervision for earlier classes. However, we observed that PLG tends to favour frequent classes while ignoring under-represented ones. To mitigate this, we propose a Class-Balanced Re-weighting (CBR), which estimates class frequencies from the pseudo-labels to enhance the learning of rare classes without accessing past data. Together, these components form a practical and effective solution for real-world 3DIS.

To evaluate CLIMB-3D, we designed three benchmark scenarios for CI-3DIS based on the ScanNet200 dataset [57]. These scenarios simulate incremental learning setup under natural class imbalance, where new classes emerge based on (A) object frequency, (B) semantic similarity, or (C) random grouping. Additionally, for comparison with existing methods, we evaluate our approach on CI-3DSS using the ScanNetV2 dataset [16]. Experimental results show that our method significantly reduces forgetting and improves performance across both frequent and rare categories, outperforming previous methods in both settings. Fig. 1 illustrates the CI-3DIS setup.

In summary, our contributions are: (i) a novel problem setting of imbalanced class-incremental 3DIS with an effective method to balance learning and mitigate forgetting; (ii) three benchmarks modelling continual object emergence with natural imbalance; and (iii) strong experimental results, achieving up to 16.76% mAP improvement over baselines.

## 2 Related Work

**3D Instance Segmentation.** Various methods have been proposed for 3DIS. Grouping-based approaches adopt a bottom-up pipeline that learns latent embeddings for point clustering [9, 21, 22, 29, 37, 44, 70, 78]. In contrast, proposal-based methods follow a top-down strategy by detecting 3D boxes and segmenting objects within them [19, 26, 46, 73, 76]. Recently, transformer-based models [68] have also been applied to 3DIS and 3DSS [59, 64], inspired by advances in 2D vision [10, 11]. However, these approaches require full annotations for all classes and are not designed for progressive learning, where only new class annotations are available and previous data is inaccessible. Another line of work aims to reduce annotation costs by proposing weakly supervised 3DIS methods based on sparse cues [12, 27, 72]. While effective with limited annotations, these methods assume a fixed set of classes and are susceptible to catastrophic forgetting in incremental settings.

**Incremental Learning.** Continual learning involves training models sequentially to mitigate catastrophic forgetting. One common strategy is model regularisation, which constrains parameter updates using techniques such as elastic weight consolidation or knowledge distillation [1, 24, 35, 43, 60]. Another approach is exemplar replay, where past data is stored or generated to rehearse older tasks while learning new ones [5, 7, 30, 54]. A third direction involves dynamically expanding the model architecture or using modular sub-networks to accommodate new tasks without interfering with old ones [31, 42, 53, 58, 71, 79]. Recent efforts in class-incremental 3D segmentation [63, 74] have shown early promise but often rely on basic architectures and focus primarily on semantic segmentation. Other works [4, 36] explore continual learning in 3D settings, though they often depend on large memory buffers. In contrast, our work introduces a tailored framework for 3D instance segmentation that effectively transfers knowledge across tasks, even under class imbalance.

**Long-tailed Recognition.** Imbalanced class distributions lead to poor recognition of rare (long-tail) classes. To address this, re-sampling methods balance the data via input [8, 20, 50, 52, 61] or feature space [3, 14, 40], avoiding naïve under-/over-sampling. Loss re-weighting offers another direction, using class-based [6, 15, 23, 32, 33, 69] or per-example [45, 55, 62] adjustments to ensure fair contribution. Parameter regularisation improves generalisation through weight constraints [2], albeit requiring careful tuning. Other approaches leverage transfer learning [77, 81], self-supervision [41, 75], or contrastive learning [34, 41, 82] to improve rare-class representations. Long-tailed recognition is well-studied in 2D with large-scale datasets [15, 49, 67]; in 3D, ScanNet200 [16, 57] enables similar efforts. Prior 3D work focused on re-weighting, re-sampling, and transfer learning [57], while regularisation remains underexplored. CeCo [80] balances class centres via auxiliary loss but overlooks sampling and augmentation. Lahoud et al. [38] propose adaptive classifier regularisation for 3D segmentation, outperforming prior approaches without threshold tuning. However, neither method considers incremental learning. In contrast, we jointly address long-tail and continual 3D semantic segmentation.

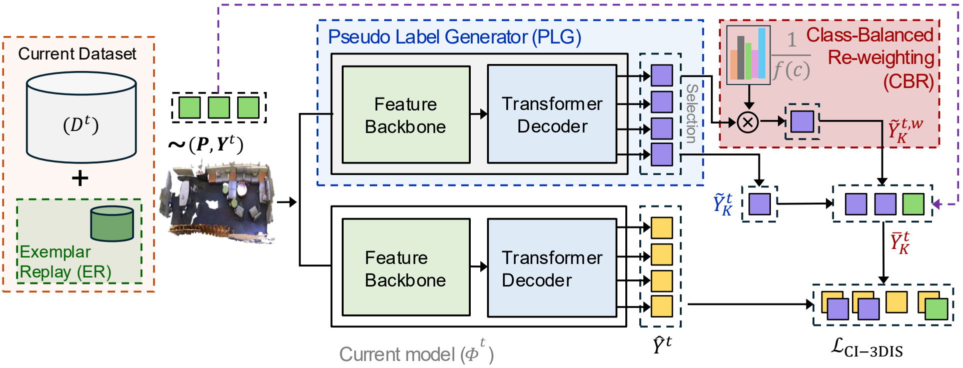


Figure 2: Overview of **CLIMB-3D** for CI-3DIS. The model incrementally learns new classes across sequential phases. During task $t$, point clouds $\mathbf{P}$ and their corresponding labels $\mathbf{Y}^t$ are sampled from a combination of the current training dataset $D^t$ and Exemplar Replay (ER), which maintains a small memory of past examples. These are then passed to the current model $\Phi^t$ to produce predicted labels $\mathbf{Y}^t$. The Pseudo-Label Generator (PLG) selects the top-$K$ predictions from the previous model $\Phi^{t-1}$. These pseudo-labels are then weighted based on class frequency $f(c)$ using Class-Balanced Re-weighting (CBR), and the top-$K$ re-weighted labels are selected to form the balanced pseudo-label set $\bar{\mathbf{Y}}^t$. This pseudo-label set is then concatenated with the ground-truth labels to form a final augmented supervision set $\overline{\mathbf{Y}}^t$ for task $t$, which is used to optimise the model $\Phi^t$ using Eq. (3).

# 3 Methodology

## 3.1 Problem Formulation

**3DIS.** The objective of this task is to detect and segment each object instance within a point cloud. Formally, the training dataset is denoted as $\mathcal{D} = \{(\mathbf{P}_i, \mathbf{Y}_i)\}_{i=1}^N$, where $N$ is the number of training samples. Each input $\mathbf{P}_i \in \mathbb{R}^{M\times 6}$ is a coloured point cloud consisting of $M$ points, where each point is represented by its 3D coordinates and RGB values. The corresponding annotation $\mathbf{Y}_i = \{(m_{i,j}, c_{i,j})\}_{j=1}^{J_i}$ contains $J_i$ object instances, where $m_{i,j} \in \{0,1\}^M$ is a binary mask indicating which points belong to the $j$-th instance, and $c_{i,j} \in \mathcal{C} = \{1,\dots,C\}$ is the semantic class label of that instance. Given $\mathbf{P}_i$, the model $\Phi$ predicts instance-level outputs $\hat{\mathbf{Y}}_i = \{(\hat{m}_{i,j}, \hat{c}_{i,j})\}_{j=1}^{\hat{J}_i}$, where $\hat{m}_{i,j}$ and $\hat{c}_{i,j}$ denote the predicted mask and category label for the $j$-th instance, respectively. The number of predicted instances $\hat{J}_i$ varies depending on the model's inference. The model is optimised using the following objective:

$$\mathcal{L}_{\text{3DIS}}(\mathcal{D};\Phi) = \frac{1}{|\mathcal{D}|}\sum_{(\mathbf{P},\mathbf{Y})\in\mathcal{D}} \frac{1}{|\mathbf{Y}|}\sum_{(m_j,c_j)\in\mathbf{Y}} \left(\mathcal{L}_{\text{mask}}(m_j,\hat{m}_j) + \lambda\cdot\mathcal{L}_{\text{cls}}(c_j,\hat{c}_j)\right), \tag{1}$$

where $\mathcal{L}_{\text{cls}}$ and $\mathcal{L}_{\text{mask}}$ are the average classification and mask losses over all instances, and $\lambda$ controls their trade-off.

**CI-3DIS.** Unlike conventional 3DIS, CI-3DIS involves sequential learning, where object categories are introduced incrementally over $T$ training tasks. Each task $t$ introduces a disjoint set of classes $\mathcal{C}^t$, with $\mathcal{C} = \bigcup_{t=1}^T \mathcal{C}^t$ and $\mathcal{C}^t \cap \mathcal{C}^{t'} = \emptyset$ for $t \neq t'$. During each task $t$, the model receives a dataset $\mathcal{D}^t = \{(\mathbf{P}_i, \mathbf{Y}_i^t)\}_{i=1}^N$, where each coloured point cloud $\mathbf{P}_i \in \mathbb{R}^{M \times 6}$ contains $M$ points, and the corresponding annotation $\mathbf{Y}_i^t = \{(m_{i,j}^t, c_{i,j}^t)\}_{j=1}^{J_i^t}$ includes instance masks $m_{i,j}^t \in \{0,1\}^M$ and semantic class labels $c_{i,j}^t \in \mathcal{C}^t$. The model $\Phi^t$, initialized from $\Phi^{t-1}$, is trained on $\mathcal{D}^t$ to predict instance-level outputs $\hat{\mathbf{Y}}_i^t = \{(\hat{m}_{i,j}^t, \hat{c}_{i,j}^t)\}_{j=1}^{\hat{J}_i^t}$, where each predicted semantic label $\hat{c}_{i,j}^t$ belongs to the cumulative set of all classes observed so far: $\bigcup_{k=1}^t \mathcal{C}^k$. The key challenge in this setup is to learn new classes without forgetting those encountered in previous tasks, despite limited supervision and the absence of past labels.

## 3.2 Method: CLIMB-3D

**Overview.** As shown in Fig. 2 and formulated in Sec. 3.1, the proposed framework for CI-3DIS follows a *phase-wise* training strategy. In each phase, the model is exposed to a carefully curated subset of the dataset, simulating real-world scenarios discussed in Sec. 3.3. Training naïvely on such phased data leads to *catastrophic forgetting* [51], where the model forgets prior knowledge when learning new tasks. To address this, we first incorporate *Exemplar Replay* (**ER**) [5], a strategy inspired by 2D incremental methods [1, 43, 54] and recently extended to 3D settings [4], to mitigate forgetting by storing a small set of representative samples from earlier phases. However, ER alone is insufficient to achieve robust performance. To address this, two additional components are introduced: a *Pseudo-Label Generator* (**PLG**), which leverages a frozen model to generate supervision signals for previously seen classes, and a *Class-Balanced Re-weighting* (**CBR**) module that compensates for class imbalance across phases. Each of these components is described in detail below [1].

**Exemplar Replay (ER).** Inspired by Buzzega et al. [5], ER addresses the issue of catastrophic forgetting by allowing the model to retain a subset of data from earlier phases. During phase $t$, the model is trained not only on the current task data $\mathcal{D}^t$, but also on a set of exemplars $\mathcal{E}^{1:t-1}$, collected from previous phases. These exemplars are accumulated incrementally: $\mathcal{E}^{1:t-1} = \mathcal{E}^1 \cup \cdots \cup \mathcal{E}^{t-1}$. Here, $\mathcal{E}^t$ denotes the exemplar subset stored after phase $t$, and $|\mathcal{E}^t|$ denotes the exemplar replay size. The combined replay dataset is denoted as: $\mathcal{D}^t_{\text{ER}} = \mathcal{D}^t \cup \mathcal{E}^{1:t-1}$. Epoch training consists of two stages: first, the model is trained on the current data, $\mathcal{D}^t$; then it is trained using exemplars from $\mathcal{E}^{1:t-1}$.

While previous CI-3DIS approaches rely on large exemplar sets to mitigate forgetting [4], such strategies become impractical under memory constraints. Consequently, although ER aids knowledge retention, it is insufficient on its own for robust performance with a limited exemplar budget. Therefore, PLG and CBR are introduced to better preserve past representations and enhance generalisation across phases.

**Pseudo-Label Generator (PLG).** In the CI-3DIS setting, during phase $t$ (where $t > 1$), although ground-truth labels for previously seen classes are unavailable, the model from the previous phase $\Phi^{t-1}$, which preserves knowledge of past tasks, is retained. We use this model to generate pseudo-labels for previously seen classes, thereby providing approximate supervision during training. This allows the current model $\Phi^t$, to retain prior knowledge and reduce forgetting, even without access to ground-truth annotations.

Given a point cloud and label pair $(\mathbf{P}, \mathbf{Y}^t)$, from $\mathcal{D}^t$, the previous model $\Phi^{t-1}$, generates pseudo-labels for the previously seen classes as $\hat{\mathbf{Y}}^{1:t-1} = \Phi^{t-1}(\mathbf{P})$, which we denote as $\tilde{\mathbf{Y}}^t$

[1] For clarity, modifications introduced by each component (ER, PLG, and CBR) are highlighted in green, blue, and red, respectively.

for brevity. However, we observe that utilising all predictions from $\Phi^{t-1}$ introduces noisy or incorrect labels, which degrade overall performance. To mitigate this, we select the top-$K$ most confident instance predictions, denoted as $\tilde{\mathbf{Y}}^t_K$, and concatenate them with the ground-truth labels of the current task to obtain the supervision set: $\bar{\mathbf{Y}}^t = \mathbf{Y}^t \parallel \tilde{\mathbf{Y}}^t_K$. Following this, the loss is computed over both the current task's ground-truth labels and the top-$K$ confident pseudo-labels from earlier phases, enabling the model to retain prior knowledge while adapting to new classes.

**Class-Balanced Re-weighting (CBR).** While ER and PLG contribute to preserving prior knowledge, we observe that the model's predictions are biased towards frequent classes, resulting in the forgetting of rare categories. This issue is further amplified by the top-$K$ pseudo-labels selected by PLG, which predominantly represent dominant classes in the data. To address this, we propose a *Class-Balanced Re-weighting* (CBR) scheme that compensates for class imbalance using object frequency statistics. At task $t$, only the current dataset $\mathcal{D}^t$ and its label distribution $p^t(c)$ over classes $c \in \mathcal{C}^t$ are available. Datasets and class distributions from earlier tasks $\{\mathcal{D}^i, p^i(c)\}_{i=1}^{t-1}$ are no longer accessible, which makes it challenging to directly account for class imbalance across all previously seen tasks. Therefore, we propose to leverage the pseudo-label predictions of the frozen model $\Phi^{t-1}$ as a proxy for the class distribution across previously learned categories.

During each training iteration, the previous model $\Phi^{t-1}$ is applied to the current input $\mathbf{P}$ to generate pseudo-labels $\tilde{\mathbf{Y}}^t$, from which we accumulate class-wise frequency statistics for previously learned classes. Let $\mathcal{C}^{1:t} = \bigcup_{i=1}^{t} \mathcal{C}^i$ represent the union of all classes seen up to task $t$. At the end of each epoch, we compute the overall class frequency $\mathbf{f} \in \mathbb{R}^{|\mathcal{C}^{1:t}|}$ by combining the pseudo-label distribution $\tilde{p}^t(c)$ derived from $\tilde{\mathbf{Y}}^t$ with the ground-truth label distribution $p^t(c)$ from the current task. The resulting frequency vector $\mathbf{f} = [f(c)]_{c \in \mathcal{C}^{1:t}}$ is then used to compute class-wise weights $\mathbf{w} = [w(c)]_{c \in \mathcal{C}^{1:t}}$, which guide balanced pseudo-label selection and help the model focus on rare classes.

In the next epoch, the model $\Phi^{t-1}$ produces soft pseudo-label predictions $\tilde{\mathbf{Y}}^t$ (as described in PLG) , where each component $\tilde{\mathbf{Y}}^t[c]$ denotes the confidence score for class $c \in \mathcal{C}^{1:t-1}$. To promote balanced pseudo-label selection, class-wise re-weighting is applied using $w(c) = \frac{1}{f(c)+\varepsilon}$, where $\varepsilon$ is a small constant for numerical stability. The adjusted predictions are computed by applying the class-wise weights as: $\tilde{\mathbf{Y}}^{t,w}[c] = w(c) \cdot \tilde{\mathbf{Y}}^t[c], \quad \forall c \in \mathcal{C}^{1:t-1}$. Top-$K$ selection is then applied to both the original and the re-weighted scores, yielding two pseudo-label sets: $\tilde{\mathbf{Y}}^t_K$ and $\tilde{\mathbf{Y}}^{t,w}_K$, respectively. The final augmented target set, $\overline{\mathbf{Y}}^t$, is constructed by concatenating the ground-truth labels $\mathbf{Y}^t$ with both sets of pseudo-labels:

$$\overline{\mathbf{Y}}^t = \mathbf{Y}^t \parallel \tilde{\mathbf{Y}}^t_K \parallel \tilde{\mathbf{Y}}^{t,w}_K. \tag{2}$$

Although the above re-weighting mitigates bias from prior tasks, it does not address class imbalance within the current task $t$. To resolve this, the class-balanced weights $w_c$ are extended to incorporate statistics from both previously seen and current classes, i.e., over $\mathcal{C}^{1:t}$ rather than only $\mathcal{C}^{1:t-1}$. The updated weights, denoted as $w'_c$, are used to re-weights both pseudo-label selection and classification loss, ensuring balanced supervision across all seen categories, including rare ones. This unified strategy reduces bias and enhances performance across phases. The pseudo-code for constructing the supervision set in Eq. (2), which incorporates both PLG and CBR components, is provided in Appendix A.

**Final Objective.** The final training objective of **CLIMB-3D**, incorporating ER, PLG, and

CBR, is formulated in Eq. (1) as:

$$\mathcal{L}(\mathcal{D}^t_{\mathrm{ER}};\Phi^t) = \frac{1}{|\mathcal{D}^t_{\mathrm{ER}}|}\sum_{(\mathbf{P},\mathbf{Y}^t)\in\mathcal{D}^t_{\mathrm{ER}}} \frac{1}{|\overline{\mathbf{Y}}^t|}\sum_{(\overline{m}^t_j,\overline{c}^t_j)\in\overline{\mathbf{Y}}^t} \left(\mathcal{L}_{\mathrm{mask}}(\overline{m}^t_j,\hat{m}^t_j) + \mathbf{w}''_c\cdot\mathcal{L}_{\mathrm{cls}}(\overline{c}^t_j,\hat{c}^t_j)\right), \qquad (3)$$

where $\mathbf{w}''_c = \mathbf{w}'_c \cdot \lambda_{\mathrm{cls}}$ denotes the scaled weight vector, and $\lambda_{\mathrm{cls}}$ is a hyperparameter for balancing the loss terms.

### 3.3 Benchmarking Incremental Scenarios

While CI-3DIS methods offer a wide range of practical applications, they frequently rely on the assumption of uniform sample distribution, which rarely holds in real-world settings. In practice, the number of object categories, denoted by $\mathcal{C}$, is often large and characterised by substantial variability in category frequency, shape, structure, and size. To address these challenges, we propose three incremental learning scenarios, each designed to capture a different facet of real-world complexity. The design of these scenarios is detailed below; for further information and illustrations, refer to Appendix E of the supplementary material.

Split-A**: Frequency Scenarios.** This scenario acknowledges that datasets are often labelled based on the frequency of categories. To accommodate this, we propose a split where the model learns from the most frequent categories and subsequently incorporates the less frequent ones in later stages. By prioritising the training of frequently occurring categories, the model can establish a strong foundation before expanding to handle rare categories.

Split-B**: Semantic Scenarios.** Beyond frequency, semantic similarity is crucial in real-world deployments. While objects often share visual or functional traits, models may encounter semantically different categories in new environments. To simulate this challenge, we introduce the Split-B scenario. Here, categories are grouped based on their semantic relationships, and the model is incrementally trained on one semantic group at a time. This setup encourages the model to generalise across semantically similar categories and adapt more effectively when exposed to new ones. Unlike the Split-A scenario, which organises learning based on category frequency, the Split-B scenario may contain both frequent and infrequent categories within a single task, focusing on semantic continuity and transfer.

Split-C**: Random Scenarios.** In some cases, data labelling is driven by the availability of objects rather than predefined criteria. To capture this, we introduce the Split-C scenario, which represents a fully random setting where any class can appear in any task, resulting in varying levels of class imbalance. By exposing the model to such diverse and unpredictable distributions, we aim to improve its robustness in real-world situations where labelled data availability is inconsistent.

These incremental scenarios are designed to provide a more realistic representation of object distributions, frequencies, and dynamics encountered in the real world.

# 4 Experiments

## 4.1 Setup

**Datasets.** We evaluate **CLIMB-3D** on ScanNet200 [57], which comprises 200 object categories and exhibits significant class imbalance, which makes it ideal for simulating and assessing real-world scenarios. In addition, we compare our approach against existing incremental learning methods using ScanNetV2 [16] in the 3DSS setting. For this evaluation, we follow the standard training and validation splits defined in prior works [74] to ensure consistency with existing methods.

**Evaluation Metrics.** We evaluate our method using *Mean Average Precision* (mAP), a standard metric for 3DIS which provides a comprehensive measure of segmentation quality by accounting for both precision and recall. For comparison with existing 3DSS methods, we report the *mean Intersection over Union* (mIoU), which quantifies the overlap between predicted and ground truth segments. To assess the model's ability to mitigate catastrophic forgetting in incremental settings, we use the *Forgetting Percentage Points* (FPP) metric, as defined in [47]. FPP captures performance degradation by measuring the accuracy drop on the initially seen categories between the first and final training phases.

Detailed descriptions of the incremental scenarios (Split-A, Split-B, Split-C) and implementation are provided in Appendix B.

## 4.2 Results and Discussion

Table 1: Comparison between the proposed method and the baseline in the 3DIS setting, evaluated using $\text{mAP}_{25}$, $\text{mAP}_{50}$, mAP, and FPP after training across all phase.

| Scenarios | Methods | Average Precision ↑ | | | FPP ↓ | |
|---|---|---|---|---|---|---|
| | | $\text{mAP}_{25}$ | $\text{mAP}_{50}$ | mAP | $\text{mAP}_{25}$ | $\text{mAP}_{50}$ |
| Split-A | Baseline | 16.46 | 14.29 | 10.44 | 51.30 | 46.82 |
| | **CLIMB-3D** (Ours) | **35.69** | **31.05** | **22.72** | **3.44** | **2.63** |
| Split-B | Baseline | 17.22 | 15.07 | 10.93 | 46.27 | 42.1 |
| | **CLIMB-3D** (Ours) | **35.48** | **31.56** | **23.69** | **8.00** | **5.51** |
| Split-C | Baseline | 25.65 | 21.08 | 14.85 | 31.68 | 28.84 |
| | **CLIMB-3D** (Ours) | **31.59** | **26.78** | **18.93** | **9.10** | **7.89** |

To evaluate our proposed approach, we construct a baseline using ER [5] for the 3DIS setting, as no existing incremental baselines are available for this task. In contrast, for the 3DSS setting, where incremental baselines do exist, we compare our method against three prior approaches [43, 60, 74].

**Results on CI-3DIS.** Tab. 1 compares CLIMB-3D against the ER baseline across the three CI-3DIS scenarios, evaluated after all phases using $\text{mAP}_{25}$, $\text{mAP}_{50}$, overall mAP, and FPP. In the Split-A scenario, characterised by significant distribution shifts, our method achieves gains of +19.23%, +16.76%, and +12.28% in $\text{mAP}_{25}$, $\text{mAP}_{50}$, and overall mAP, respectively, while drastically reducing forgetting with FPP scores of 3.44% ($\text{mAP}_{25}$) and 2.63% ($\text{mAP}_{50}$), both over 45 points lower than the baseline. Under the Split-B scenario, with semantically related new categories, improvements of +18.26%, +16.49%, and +12.76% are observed alongside a reduction in forgetting to 8.00% and 5.51% from baseline levels above 42%. For the Split-C scenario, featuring increasing geometric complexity, CLIMB-3D maintains solid gains of +5.94%, +5.70%, and +4.08% across mAP metrics, with forgetting reduced by over 22 points. Overall, these results demonstrate robust forward transfer, minimal forgetting, and stable learning, highlighting the method's effectiveness for scalable continual 3D semantic understanding.

Table 2: Comparison between the proposed method and existing baselines in the 3DSS setting on ScanNetV2 [16], evaluated using mIoU.

| Methods | Phase=1 | Phase=2 | All |
|---|---|---|---|
| EWC [60] | 17.75 | 13.22 | 16.62 |
| LwF [43] | 30.38 | 13.37 | 26.13 |
| Yang *et al.* [74] | 34.16 | 13.43 | 28.98 |
| **CLIMB-3D** (Ours) | **69.39** | **32.56** | **59.38** |

**Results on CI-3DSS.** Although our primary focus is on CI-3DIS, we also evaluate our method under the CI-3DSS setting to enable comparisons with existing approaches. To do so, we adapt our predictions to assign each point the label corresponding to the highest-confidence mask and exclude background classes (e.g., floor and wall), as these are not part of the semantic segmentation targets. Tab. 2 presents a comparison between our method and existing baselines on the ScanNet V2 dataset [16], evaluated using mIoU across two training phases and overall. Although originally designed for instance segmentation, our method generalises effectively to semantic segmentation, achieving substantial gains of +35.23% mIoU in Phase 1 and +19.1% in Phase 2. Overall, it achieves 59.38% mIoU, significantly higher than the $\sim$30% mIoU of prior methods, highlighting its robustness and transferability.

Refer to Appendix C, D, and F of the supplementary material for detailed analyses of per-phase performance, rare-class evaluation, and qualitative result comparisons, respectively.

## 4.3 Ablation

Table 3: Ablation study illustrating the impact of each component in a three-phase setup. Each split (**s**) and corresponding number indicates data introduced at that phase. Best results are highlighted in **bold**.

| Row | Modules | p=1 ↑ | p=2 ↑ | | | p=3 ↑ | | | | FPP ↓ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | s1 | s1 | s2 | All | s1 | s2 | s3 | All | |
| 1. | Oracle | - | - | - | - | 55.14 | 30.77 | 25.30 | 37.68 | - |
| 2. | Naïve | 56.82 | 0.00 | 28.09 | 14.15 | 0.00 | 0.00 | 19.67 | 5.80 | 56.82 |
| 3. | + ER | 56.82 | 18.51 | 32.81 | 25.72 | 10.38 | 9.43 | 24.27 | 14.28 | 46.44 |
| 4. | + PLG | 56.82 | 50.00 | **34.39** | 42.13 | 49.78 | 11.41 | 26.47 | 29.28 | 7.04 |
| 5. | + CBR | 56.82 | **54.67** | 33.75 | **44.13** | **54.19** | **12.02** | **26.55** | **31.05** | **2.63** |

We conduct an ablation study to evaluate the contribution of each component in our framework. As an upper bound, we report the performance of an *Oracle* model, which is trained jointly on the full dataset. For the incremental setting, we follow the previously defined splits and first train the model naively across phases. We then incrementally add each module to isolate its impact. Tab. 3 presents results for the `Split-A` scenario using both mAP$_{50}$ and FPP metrics.

**Naïve Training.** When trained naïvely without any dedicated modules, the model suffers from severe catastrophic forgetting, as evident in row 2. The model entirely forgets previously learned classes upon entering new phases, with performance dropping to zero for earlier splits and FPP rising to 56.82.

**Effect of ER.** Adding exemplar replay (row 3) partially alleviates forgetting by maintaining a buffer of past examples. This yields notable gains for **s1** in Phase 2 (+18.51%) and Phase 3 (+10.38%), and also improves **s2** in Phase 3 (+9.43%). However, the overall forgetting remains substantial (FPP: 46.44), showing ER alone is insufficient.

**Effect of PLG.** The addition of the pseudo-label generator (PLG), which generates labels for previous classes by retaining a copy of the model from an earlier phase, facilitates knowledge retention and forward transfer. As shown in row 4, PLG significantly reduces forgetting and enhances performance on current tasks. For **s1**, it improves mAP$_{50}$ by 31.49% in Phase 2 and 39.40% in Phase 3 over exemplar replay. Overall, PLG yields a 15.00% increase in performance and reduces forgetting by 39.40%.

**Effect of CBR.** Finally, class-balanced re-weighting (CBR) mitigates class imbalance during both pseudo-labelling and current task learning by adjusting each class's contribution based on its frequency. As shown in row 5, CBR enhances **s1** retention over PLG in Phases 2 and 3 (+4.67% and +4.41%), and further improves performance on **s2** and **s3**. It also achieves the lowest FPP of 2.63, reflecting strong forgetting mitigation. Overall, CBR offers the best trade-off between retaining past knowledge and acquiring new tasks, outperforming PLG and ER by 4.41% and 43.81%, respectively.

### 4.4 Limitations & Future Work

While CLIMB-3D achieves strong performance across CI-3DIS and CI-3DSS settings, several limitations remain and warrant discussion. The experiments are currently limited to indoor datasets (ScanNet200 and ScanNetV2), therefore evaluating performance on outdoor scenes would help assess generalisability across diverse environments. The current setup also considers only three incremental phases, whereas real-world applications often involve longer sequences; incorporating such extended setups would better reflect practical challenges. Moreover, the model operates in a uni-modal setting. Integrating multi-modal cues, such as vision-language models, could further enhance performance, as suggested by recent 2D and 3D studies [25, 65, 66].

## 5 Conclusion

This work addresses the challenge of catastrophic forgetting in class-incremental 3D instance segmentation by introducing a modular framework that integrates exemplar replay (ER), a pseudo-label generator (PLG), and class-balanced re-weighting (CBR). The proposed method incrementally adapts to new classes while preserving prior knowledge, all without requiring access to the full dataset. Extensive experiments on a three-phase benchmark validate the individual and combined effectiveness of each component: ER enables efficient memory-based retention; PLG facilitates knowledge preservation through pseudo-supervision; and CBR mitigates class imbalance to enhance learning stability. Together, these components significantly reduce forgetting and improve segmentation performance across all phases. By achieving a strong balance between stability and plasticity, our approach advances continual 3D scene understanding and sets a new baseline for future research in this area.

# References

[1] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory aware synapses: Learning what (not) to forget. In *ECCV*, pages 139–154, 2018.

[2] Shaden Alshammari, Yu-Xiong Wang, Deva Ramanan, and Shu Kong. Long-tailed recognition via weight balancing. In *CVPR*, pages 6897–6907, 2022.

[3] Shin Ando and Chun Yuan Huang. Deep over-sampling framework for classifying imbalanced data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 770–785. Springer, 2017.

[4] Mohamed El Amine Boudjoghra, Salwa Al Khatib, Jean Lahoud, Hisham Cholakkal, Rao Anwer, Salman H Khan, and Fahad Shahbaz Khan. 3d indoor instance segmentation in an open-world. *NeurIPS*, 36, 2024.

[5] Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: a strong, simple baseline. *NeurIPS*, 33:15920–15930, 2020.

[6] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *NeurIPS*, 32, 2019.

[7] Hyuntak Cha, Jaeho Lee, and Jinwoo Shin. Co2l: Contrastive continual learning. In *ICCV*, pages 9516–9525, 2021.

[8] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[9] Shaoyu Chen, Jiemin Fang, Qian Zhang, Wenyu Liu, and Xinggang Wang. Hierarchical aggregation for 3d instance segmentation. In *ICCV*, pages 15467–15476, 2021.

[10] Bowen Cheng, Alex Schwing, and Alexander Kirillov. Per-pixel classification is not all you need for semantic segmentation. *NeurIPS*, 34:17864–17875, 2021.

[11] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *CVPR*, pages 1290–1299, 2022.

[12] Julian Chibane, Francis Engelmann, Tuan Anh Tran, and Gerard Pons-Moll. Box2mask: Weakly supervised 3d semantic instance segmentation using bounding boxes. In *ECCV*, pages 681–699. Springer, 2022.

[13] Townim Chowdhury, Mahira Jalisha, Ali Cheraghian, and Shafin Rahman. Learning without forgetting for 3d point cloud objects. In *Advances in Computational Intelligence: 16th International Work-Conference on Artificial Neural Networks, IWANN 2021, Virtual Event, June 16–18, 2021, Proceedings, Part I 16*, pages 484–497. Springer, 2021.

[14] Peng Chu, Xiao Bian, Shaopeng Liu, and Haibin Ling. Feature space augmentation for long-tailed data. In *ECCV*, pages 694–710. Springer, 2020.

[15] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *CVPR*, pages 9268–9277, 2019.

[16] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *CVPR*, 2017.

[17] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE PAMI*, 44(7):3366–3385, 2021.

[18] Jiahua Dong, Yang Cong, Gan Sun, Bingtao Ma, and Lichen Wang. I3dol: Incremental 3d object learning without catastrophic forgetting. In *AAAI*, volume 35, pages 6066–6074, 2021.

[19] Francis Engelmann, Martin Bokeloh, Alireza Fathi, Bastian Leibe, and Matthias Nießner. 3d-mpa: Multi-proposal aggregation for 3d semantic instance segmentation. In *CVPR*, pages 9031–9040, 2020.

[20] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis: A dataset for large vocabulary instance segmentation. In *CVPR*, pages 5356–5364, 2019.

[21] Lei Han, Tian Zheng, Lan Xu, and Lu Fang. Occuseg: Occupancy-aware 3d instance segmentation. In *CVPR*, pages 2940–2949, 2020.

[22] Tong He, Chunhua Shen, and Anton Van Den Hengel. Dyco3d: Robust instance segmentation of 3d point clouds through dynamic convolution. In *CVPR*, pages 354–363, 2021.

[23] Yin-Yin He, Peizhen Zhang, Xiu-Shen Wei, Xiangyu Zhang, and Jian Sun. Relieving long-tailed instance segmentation via pairwise class balance. In *CVPR*, pages 7000–7009, 2022.

[24] Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7), 2015.

[25] Yining Hong, Haoyu Zhen, Peihao Chen, Shuhong Zheng, Yilun Du, Zhenfang Chen, and Chuang Gan. 3d-llm: Injecting the 3d world into large language models. *NeurIPS*, 36:20482–20494, 2023.

[26] Ji Hou, Angela Dai, and Matthias Nießner. 3d-sis: 3d semantic instance segmentation of rgb-d scans. In *CVPR*, pages 4421–4430, 2019.

[27] Ji Hou, Benjamin Graham, Matthias Nießner, and Saining Xie. Exploring data-efficient 3d scene understanding with contrastive scene contexts. In *CVPR*, pages 15587–15597, 2021.

[28] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *ICML*, pages 4904–4916. PMLR, 2021.

[29] Li Jiang, Hengshuang Zhao, Shaoshuai Shi, Shu Liu, Chi-Wing Fu, and Jiaya Jia. Pointgroup: Dual-set point grouping for 3d instance segmentation. In *CVPR*, pages 4867–4876, 2020.

[30] Nitin Kamra, Umang Gupta, and Yan Liu. Deep generative dual memory network for continual learning. *arXiv preprint arXiv:1710.10368*, 2017.

[31] Zixuan Ke, Bing Liu, and Xingchang Huang. Continual learning of a mixed sequence of similar and dissimilar tasks. *NeurIPS*, 33:18493–18504, 2020.

[32] Salman Khan, Munawar Hayat, Syed Waqas Zamir, Jianbing Shen, and Ling Shao. Striking the right balance with uncertainty. In *CVPR*, pages 103–112, 2019.

[33] Salman H Khan, Munawar Hayat, Mohammed Bennamoun, Ferdous A Sohel, and Roberto Togneri. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE transactions on neural networks and learning systems*, 29(8): 3573–3587, 2017.

[34] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *NeurIPS*, 33:18661–18673, 2020.

[35] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[36] Theodora Kontogianni, Yuanwen Yue, Siyu Tang, and Konrad Schindler. Is continual learning ready for real-world challenges? *arXiv preprint arXiv:2402.10130*, 2024.

[37] Jean Lahoud, Bernard Ghanem, Marc Pollefeys, and Martin R Oswald. 3d instance segmentation via multi-task metric learning. In *ICCV*, pages 9256–9266, 2019.

[38] Jean Lahoud, Fahad Shahbaz Khan, Hisham Cholakkal, Rao Muhammad Anwer, and Salman Khan. Long-tailed 3d semantic segmentation with adaptive weight constraint and sampling. In *International Conference on Robotics and Automation (ICRA)*, pages 5037–5044, 2024. doi: 10.1109/ICRA57147.2024.10610029.

[39] Xin Lai, Jianhui Liu, Li Jiang, Liwei Wang, Hengshuang Zhao, Shu Liu, Xiaojuan Qi, and Jiaya Jia. Stratified transformer for 3d point cloud segmentation, 2022.

[40] Shuang Li, Kaixiong Gong, Chi Harold Liu, Yulin Wang, Feng Qiao, and Xinjing Cheng. Metasaug: Meta semantic augmentation for long-tailed visual recognition. In *CVPR*, pages 5212–5221, 2021.

[41] Tianhao Li, Limin Wang, and Gangshan Wu. Self supervision to distillation for long-tailed visual recognition. In *ICCV*, pages 630–639, 2021.

[42] Xilai Li, Yingbo Zhou, Tianfu Wu, Richard Socher, and Caiming Xiong. Learn to grow: A continual structure learning framework for overcoming catastrophic forgetting. In *ICML*, pages 3925–3934. PMLR, 2019.

[43] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE PAMI*, 40(12): 2935–2947, 2017.

[44] Zhihao Liang, Zhihao Li, Songcen Xu, Mingkui Tan, and Kui Jia. Instance segmentation in 3d scenes using semantic superpoint tree networks. In *ICCV*, pages 2783–2792, 2021.

[45] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, pages 2980–2988, 2017.

[46] Shih-Hung Liu, Shang-Yi Yu, Shao-Chi Wu, Hwann-Tzong Chen, and Tyng-Luh Liu. Learning gaussian instance segmentation in point clouds. *arXiv preprint arXiv:2007.09860*, 2020.

[47] Yaoyao Liu, Bernt Schiele, Andrea Vedaldi, and Christian Rupprecht. Continual detection transformer for incremental object detection. In *CVPR*, pages 23799–23808, 2023.

[48] Yuyang Liu, Yang Cong, Gan Sun, Tao Zhang, Jiahua Dong, and Hongsen Liu. L3doc: Lifelong 3d object classification. *ICIP*, 30:7486–7498, 2021.

[49] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X Yu. Large-scale long-tailed recognition in an open world. In *CVPR*, pages 2537–2546, 2019.

[50] Dhruv Mahajan, Ross Girshick, Vignesh Ramanathan, Kaiming He, Manohar Paluri, Yixuan Li, Ashwin Bharambe, and Laurens Van Der Maaten. Exploring the limits of weakly supervised pretraining. In *ECCV*, pages 181–196, 2018.

[51] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pages 109–165. Elsevier, 1989.

[52] Seulki Park, Youngkyu Hong, Byeongho Heo, Sangdoo Yun, and Jin Young Choi. The majority can help the minority: Context-rich minority oversampling for long-tailed classification. In *CVPR*, pages 6887–6896, 2022.

[53] Jathushan Rajasegaran, Munawar Hayat, Salman Khan, Fahad Shahbaz Khan, and Ling Shao. Random path selection for incremental learning. *NeurIPS*, 3, 2019.

[54] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *CVPR*, pages 2001–2010, 2017.

[55] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *ICML*, pages 4334–4343. PMLR, 2018.

[56] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *NeurIPS*, 32, 2019.

[57] David Rozenberszki, Or Litany, and Angela Dai. Language-grounded indoor 3d semantic segmentation in the wild. In *ECCV*, 2022.

[58] Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. *arXiv preprint arXiv:1606.04671*, 2016.

[59] Jonas Schult, Francis Engelmann, Alexander Hermans, Or Litany, Siyu Tang, and Bastian Leibe. Mask3D: Mask Transformer for 3D Semantic Instance Segmentation. In *International Conference on Robotics and Automation (ICRA)*, 2023.

[60] Joan Serra, Didac Suris, Marius Miron, and Alexandros Karatzoglou. Overcoming catastrophic forgetting with hard attention to the task. In *ICML*, pages 4548–4557. PMLR, 2018.

[61] Li Shen, Zhouchen Lin, and Qingming Huang. Relay backpropagation for effective learning of deep convolutional neural networks. In *ECCV*, pages 467–482. Springer, 2016.

[62] Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. Meta-weight-net: Learning an explicit mapping for sample weighting. *NeurIPS*, 32, 2019.

[63] Yuanzhi Su, Siyuan Chen, and Yuan-Gen Wang. Balanced residual distillation learning for 3d point cloud class-incremental semantic segmentation. *arXiv preprint arXiv:2408.01356*, 2024.

[64] Jiahao Sun, Chunmei Qing, Junpeng Tan, and Xiangmin Xu. Superpoint transformer for 3d scene instance segmentation. *arXiv preprint arXiv:2211.15766*, 2022.

[65] Vishal Thengane, Salman Khan, Munawar Hayat, and Fahad Khan. Clip model is an efficient continual learner. *arXiv preprint arXiv:2210.03114*, 2022.

[66] Vishal Thengane, Xiatian Zhu, Salim Bouzerdoum, Son Lam Phung, and Yunpeng Li. Foundational models for 3d point clouds: A survey and outlook. *arXiv preprint arXiv:2501.18594*, 2025.

[67] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *CVPR*, pages 8769–8778, 2018.

[68] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *NeurIPS*, 30, 2017.

[69] Tong Wang, Yousong Zhu, Chaoyang Zhao, Wei Zeng, Jinqiao Wang, and Ming Tang. Adaptive class suppression loss for long-tail object detection. In *CVPR*, pages 3103–3112, 2021.

[70] Weiyue Wang, Ronald Yu, Qiangui Huang, and Ulrich Neumann. Sgpn: Similarity group proposal network for 3d point cloud instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2569–2578, 2018.

[71] Zifeng Wang, Tong Jian, Kaushik Chowdhury, Yanzhi Wang, Jennifer Dy, and Stratis Ioannidis. Learn-prune-share for lifelong learning. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 641–650. IEEE, 2020.

[72] Saining Xie, Jiatao Gu, Demi Guo, Charles R Qi, Leonidas Guibas, and Or Litany. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In *ECCV*, pages 574–591. Springer, 2020.

[73] Bo Yang, Jianan Wang, Ronald Clark, Qingyong Hu, Sen Wang, Andrew Markham, and Niki Trigoni. Learning object bounding boxes for 3d instance segmentation on point clouds. *NeurIPS*, 32, 2019.

[74] Yuwei Yang, Munawar Hayat, Zhao Jin, Chao Ren, and Yinjie Lei. Geometry and uncertainty-aware 3d point cloud class-incremental semantic segmentation. In *CVPR*, pages 21759–21768, June 2023.

[75] Yuzhe Yang and Zhi Xu. Rethinking the value of labels for improving class-imbalanced learning. *NeurIPS*, 33:19290–19301, 2020.

[76] Li Yi, Wang Zhao, He Wang, Minhyuk Sung, and Leonidas J Guibas. Gspn: Generative shape proposal network for 3d instance segmentation in point cloud. In *CVPR*, pages 3947–3956, 2019.

[77] Xi Yin, Xiang Yu, Kihyuk Sohn, Xiaoming Liu, and Manmohan Chandraker. Feature transfer learning for face recognition with under-represented data. In *CVPR*, pages 5704–5713, 2019.

[78] Biao Zhang and Peter Wonka. Point cloud instance segmentation using probabilistic embeddings. In *CVPR*, pages 8883–8892, 2021.

[79] Tingting Zhao, Zifeng Wang, Aria Masoomi, and Jennifer Dy. Deep bayesian unsupervised lifelong learning. *Neural Networks*, 149:95–106, 2022.

[80] Zhisheng Zhong, Jiequan Cui, Yibo Yang, Xiaoyang Wu, Xiaojuan Qi, Xiangyu Zhang, and Jiaya Jia. Understanding imbalanced semantic segmentation through neural collapse. In *CVPR*, pages 19550–19560, 2023.

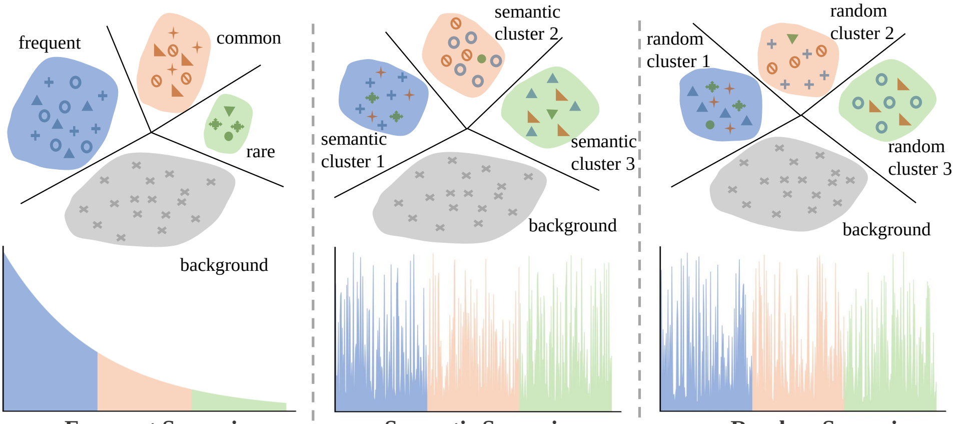
[81] Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *CVPR*, pages 9719–9728, 2020.

[82] Jianggang Zhu, Zheng Wang, Jingjing Chen, Yi-Ping Phoebe Chen, and Yu-Gang Jiang. Balanced contrastive learning for long-tailed visual recognition. In *CVPR*, pages 6908–6917, 2022.

# CLIMB-3D: Class-Incremental Imbalanced 3D Instance Segmentation

Supplementary Material

## Appendix A Illustrating Incremental scenarios



**Frequent Scenario** **Semantic Scenario** **Random Scenario**

Figure 3: Tasks are grouped into incremental scenarios based on object frequency, semantic similarity, and random assignment. 🟦, 🟧, and 🟩 denote different tasks; shapes indicate object categories; ⬜ marks the background. **Left:** Grouped by category frequency. **Middle:** Grouped by semantic similarity (e.g., similar shapes). **Right:** Randomly grouped, mixing semantic and frequency variations.

## Appendix B Per Phase Analysis

We extend the analysis from Tab. 1 to Tab. 4 to highlight the impact of our proposed method on individual splits across various scenarios. The results clearly demonstrate that our model consistently retains knowledge of previous tasks better than the baseline. For Split-A, our model shows improvement throughout the phase. In Phase 3 of (**s2**), although both the baseline and our method exhibit a performance drop, our method reduces forgetting significantly compared to the baseline. The Split-B scenario, while more complex than Split-A, achieves comparable results due to semantic similarity among classes within the same task. In Phase 2, our model achieves overall all 43.13% mAP$_{50}$ compared to 24.53% on baseline, a similar trend is observed in Phase 3, where our method not only consistently improves learning but also enhances retention of previous information. After all three tasks, our method achieves an overall performance of 31.56% AP50, compared to 15.07% for the baseline. In the Split-C scenario, the first-stage model struggles due to the increased complexity introduced by random grouping. In Phase 2, while the baseline focuses on learning the current task, it suffers from severe forgetting of prior knowledge. Conversely, our method balances new task learning with the retention of earlier information. By Phase 3, the model effectively con-

Table 4: Comparison of results in terms of mAP$_{50}$ with the proposed CLIMB-3D across three different scenarios. Each scenario is trained over three phases (phase $= 1,2,3$), introducing a single split **s** at a time. Results highlighted in orange correspond to the proposed method, and the best results for each scenario are shown in **bold**.

| Scenarios | Methods | phase=1 | phase=2 | | | phase=3 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | s1 | s1 | s2 | All | s1 | s2 | s3 | All |
| Split-A | Baseline | 56.82 | 18.51 | 32.81 | 25.72 | 10.38 | 9.43 | 24.27 | 14.28 |
| | **CLIMB-3D** | 56.82 | **54.67** | **33.75** | **44.13** | **54.19** | **12.02** | **26.55** | **31.05** |
| Split-B | Baseline | 51.57 | 13.32 | **42.21** | 24.53 | 9.55 | 12.45 | **26.78** | 15.07 |
| | **CLIMB-3D** | 51.57 | **46.74** | 37.45 | **43.13** | **46.06** | **15.95** | 26.68 | **31.56** |
| Split-C | Baseline | 36.40 | 7.74 | **37.62** | 22.32 | 7.55 | 15.96 | **40.41** | 21.08 |
| | **CLIMB-3D** | 36.40 | **32.63** | 33.38 | **33.00** | **28.51** | **17.11** | 34.64 | **26.78** |

solidates **s1** and maintains strong performance across all task splits. Overall, our proposed method improves mAP by 5.6%.

In this supplementary material, we first demonstrate the performance gains on rare classes achieved by incorporating the IC module in Appendix C. Next, we provide detailed split information for all scenarios, based on class names, in Appendix D. Finally, we present a qualitative comparison between the baseline method and our proposed approach in Appendix E.

# Appendix C Evaluation on Rare Categories

The proposed imbalance correction (IC) module, as detailed in Section 4.2, is designed to address the performance gap for rare classes. To assess its impact, we compare its performance with the framework which has exemplar replay (ER) and knowledge distillation (KD). Specifically, we focus on its ability to improve performance for rare classes, which the model encounters infrequently compared to more common classes.

The results, shown in Tab. 5 and Tab. 6, correspond to evaluations on Split-A for *Phase 2* and *Phase 3*, respectively. In *Phase 2*, we evaluate classes seen 1–20 times per epoch, while *Phase 3* targets even less frequent classes, with observations limited to 1–10 times per epoch.

As illustrated in Tab. 5, the IC module substantially improves performance on rare classes in terms of mAP$_{50}$ in Phase 2 of Split-A. For instance, classes like recycling bin and trash bin, seen only 3 and 7 times, respectively, shows significant improvement when the IC module is applied. Overall, the IC module provides an average boost of 8.32%, highlighting its effectiveness in mitigating class imbalance.

Similarly, Tab. 6 presents results for *Phase 3*, demonstrating significant gains for infrequent classes. For example, even though the classes such as piano, bucket, and laundry basket are observed only once, IC module improves the performance by 52.30%, 10.40%, and 13.60%, respectively. The ER+KD module does not focus on rare classes like shower and toaster which results in low performance, but the IC module compensates for this imbalance by focusing on underrepresented categories. On average, the addition of the proposed

Table 5: Results for classes observed by the model 1–20 times during an epoch, evaluated on Split-A for Phase 2, in terms of mAP$_{50}$.

| Classes | Seen Count | ER + PLG | ER + PLG + CBR |
|---|---|---|---|
| paper towel dispenser | 2 | 73.10 | 74.90 |
| recycling bin | 3 | 55.80 | 60.50 |
| ladder | 5 | 53.90 | 57.10 |
| trash bin | 7 | 31.50 | 57.30 |
| bulletin board | 8 | 23.30 | 38.20 |
| shelf | 11 | 48.00 | 50.50 |
| dresser | 12 | 44.00 | 55.80 |
| copier | 12 | 93.30 | 94.50 |
| object | 12 | 3.10 | 3.30 |
| stairs | 13 | 51.70 | 67.70 |
| bathtub | 16 | 80.30 | 86.60 |
| oven | 16 | 1.50 | 3.30 |
| divider | 18 | 36.40 | 45.00 |
| column | 20 | 57.30 | 75.00 |
| **Average** | - | 46.66 | **54.98** |

IC module into the framework outperforms ER+KD by 12.13%.

## Appendix D Incremental Scenarios Phases

Tab. 7 presents the task splits for each proposed scenario introduced in Section 4.3 using the ScanNet200 dataset. The three scenarios, Split-A, Split-B, and Split-C, are each divided into three tasks: Task 1, Task 2, and Task 3. Notably, the order of classes in these tasks is random.

## Appendix E Qualitative Results

In this section, we present a qualitative comparison of the proposed framework with the baseline method. Fig. 4 illustrates the results on the Split-A evaluation after learning all tasks, comparing the performance of the baseline method and our proposed approach. As shown in the figure, our method demonstrates superior instance segmentation performance compared to the baseline. For example, in row 1, the baseline method fails to segment the sink, while in row 3, the sofa instance is missed. Overall, our framework consistently outperforms the baseline, with several missed instances by the baseline highlighted in red

Table 6: Results for classes observed by the model 1–10 times during an epoch, evaluated on Split-A for Phase 3, in terms of $\text{mAP}_{50}$.

| **Classes** | **Seen Count** | **ER+KD** | **ER+KD+IC** |
|---|---|---|---|
| piano | 1 | 7.10 | 59.40 |
| bucket | 1 | 21.10 | 31.50 |
| laundry basket | 1 | 3.80 | 17.40 |
| dresser | 2 | 55.00 | 55.40 |
| paper towel dispenser | 2 | 32.50 | 35.50 |
| cup | 2 | 24.70 | 30.30 |
| bar | 2 | 35.40 | 39.50 |
| divider | 2 | 28.60 | 42.40 |
| case of water bottles | 2 | 0.00 | 1.70 |
| shower | 3 | 0.00 | 45.50 |
| mirror | 8 | 56.00 | 68.80 |
| trash bin | 4 | 1.10 | 2.70 |
| backpack | 5 | 74.50 | 76.70 |
| copier | 5 | 94.00 | 96.80 |
| bathroom counter | 3 | 3.90 | 20.30 |
| ottoman | 4 | 32.60 | 36.20 |
| storage bin | 3 | 5.10 | 10.50 |
| dishwasher | 3 | 47.40 | 66.20 |
| trash bin | 4 | 1.10 | 2.70 |
| backpack | 5 | 74.50 | 76.70 |
| copier | 5 | 94.00 | 96.80 |
| sofa chair | 6 | 14.10 | 43.50 |
| file cabinet | 6 | 49.20 | 57.60 |
| tv stand | 7 | 67.70 | 68.60 |
| mirror | 8 | 56.00 | 68.80 |
| blackboard | 8 | 57.10 | 82.80 |
| clothes dryer | 9 | 1.70 | 3.20 |
| toaster | 9 | 0.10 | 25.90 |
| wardrobe | 10 | 22.80 | 58.80 |
| jacket | 10 | 1.20 | 4.10 |
| **Average** | - | 32.08 | **44.21** |

circles.

In Fig. 5, we present the results on Split-B, highlighting instances where the baseline method underperforms, marked with red circles. For example, in row 2, the baseline method

Table 7: Classes grouped by tasks for each proposed scenario on the ScanNet200 dataset labels. The three scenarios Split-A, Split-A, and Split-C are each divided into three tasks: Task 1, Task 2, and Task 3.

| Split_A | | | Split_B | | | Split_C | | |
|---|---|---|---|---|---|---|---|---|
| **Task 1** | **Task 2** | **Task 3** | **Task 1** | **Task 2** | **Task 3** | **Task 1** | **Task 2** | **Task 3** |
| chair | wall | pillow | tv stand | cushion | paper | broom | fan | rack |
| table | floor | picture | curtain | end table | plate | towel | stove | music stand |
| couch | door | book | blinds | dining table | soap dispenser | fireplace | tv | bed |
| desk | cabinet | box | shower curtain | keyboard | bucket | blanket | dustpan | soap dish |
| office chair | shelf | lamp | bookshelf | bag | clock | dining table | sink | closet door |
| bed | window | towel | tv | toilet paper | guitar | shelf | toaster | basket |
| sink | bookshelf | clothes | kitchen cabinet | printer | toilet paper holder | rail | doorframe | chair |
| toilet | curtain | cushion | pillow | blanket | speaker | bathroom counter | wall | toilet paper |
| monitor | kitchen cabinet | plant | lamp | microwave | cup | plunger | mattress | ball |
| armchair | counter | bag | dresser | shoe | paper towel roll | bin | stand | monitor |
| coffee table | ceiling | backpack | monitor | computer tower | bar | armchair | copier | bathroom cabinet |
| refrigerator | whiteboard | toilet paper | object | bottle | toaster | trash bin | ironing board | shoe |
| tv | shower curtain | blanket | ceiling | bin | ironing board | dishwasher | radiator | blackboard |
| nightstand | closet | shoe | board | ottoman | soap dish | lamp | keyboard | vent |
| dresser | computer tower | bottle | stove | bench | toilet paper dispenser | projector | toaster oven | bag |
| stool | board | basket | closet wall | basket | fire extinguisher | potted plant | paper bag | paper |
| bathtub | mirror | fan | couch | fan | ball | coat rack | structure | projector screen |
| end table | shower | paper | office chair | laptop | hat | end table | picture | pillar |
| dining table | blinds | person | kitchen counter | person | shower curtain rod | tissue box | purse | range hood |
| keyboard | rack | plate | shower | paper towel dispenser | paper cutter | stairs | tray | coffee maker |
| printer | blackboard | container | closet | oven | tray | fire extinguisher | couch | handicap bar |
| tv stand | rail | soap dispenser | doorframe | rack | toaster oven | case of water bottles | telephone | pillow |
| trash can | radiator | telephone | sofa chair | piano | mouse | water bottle | shower curtain rod | decoration |
| stairs | wardrobe | bucket | mailbox | suitcase | toilet seat cover dispenser | ledge | trash can | printer |
| microwave | column | clock | nightstand | rail | storage container | shower head | closet wall | object |
| stove | ladder | stand | washing machine | container | scale | guitar case | cart | mirror |
| bin | bathroom stall | light | picture | telephone | tissue box | kitchen cabinet | hat | ottoman |
| ottoman | shower wall | pipe | book | stand | light switch | poster | paper cutter | water pitcher |
| bench | mat | guitar | sink | light | crate | candle | storage organizer | refrigerator |
| washing machine | windowsill | toilet paper holder | recycling bin | laundry basket | power outlet | bowl | vacuum cleaner | divider |
| copier | bulletin board | speaker | table | pipe | sign | plate | mouse | toilet |
| sofa chair | doorframe | bicycle | backpack | seat | projector | person | paper towel roll | washing machine |
| file cabinet | shower curtain rod | cup | shower wall | column | candle | storage bin | laundry detergent | mat |
| laptop | paper cutter | jacket | toilet | bicycle | plunger | microwave | calendar | scale |
| paper towel dispenser | shower door | paper towel roll | copier | ladder | stuffed animal | office chair | wardrobe | dresser |
| oven | pillar | machine | counter | jacket | headphones | clothes dryer | whiteboard | bookshelf |
| piano | ledge | soap dish | stool | storage bin | broom | headphones | laundry basket | tv stand |
| suitcase | light switch | fire extinguisher | refrigerator | coffee maker | guitar case | toilet seat cover dispenser | shower door | closet rod |
| recycling bin | closet door | ball | window | dishwasher | dustpan | bathroom stall door | curtain | plant |
| laundry basket | shower floor | hat | file cabinet | machine | hair dryer | speaker | folded chair | counter |
| clothes dryer | projector screen | water cooler | chair | mat | water bottle | keyboard piano | suitcase | bench |
| seat | divider | mouse | wall | windowsill | handicap bar | cushion | hair dryer | ceiling |
| storage bin | closet wall | scale | plant | bulletin board | purse | table | mini fridge | piano |
| coffee maker | bathroom stall door | power outlet | coffee table | fireplace | vent | nightstand | dumbbell | closet |
| dishwasher | stair rail | decoration | stairs | mini fridge | shower floor | bathroom vanity | oven | cabinet |
| bar | bathroom cabinet | sign | armchair | water cooler | water pitcher | laptop | luggage | cup |
| toaster | closet rod | projector | cabinet | shower door | bowl | shower wall | bar | laundry hamper |
| ironing board | structure | vacuum cleaner | bathroom vanity | pillar | paper bag | desk | pipe | light switch |
| fireplace | coat rack | candle | bathroom stall | ledge | alarm clock | computer tower | bathroom stall | cd case |
| kitchen counter | storage organizer | plunger | mirror | furniture | music stand | soap dispenser | blinds | backpack |
| toilet paper dispenser | | stuffed animal | blackboard | cart | laundry detergent | container | toilet paper dispenser | windowsill |
| mini fridge | | headphones | trash can | decoration | dumbbell | bicycle | coffee table | box |
| tray | | broom | stair rail | closet door | tube | light | dish rack | book |
| toaster oven | | guitar case | box | vacuum cleaner | cd case | clothes | guitar | mailbox |
| toilet seat cover dispenser | | hair dryer | towel | dish rack | closet rod | machine | seat | sofa chair |
| furniture | | water bottle | door | range hood | coffee kettle | furniture | clock | shower curtain |
| cart | | purse | clothes | projector screen | shower head | stair rail | alarm clock | bulletin board |
| storage container | | vent | whiteboard | divider | keyboard piano | toilet paper holder | board | crate |
| tissue box | | water pitcher | bed | bathroom counter | case of water bottles | floor | file cabinet | tube |
| crate | | bowl | floor | laundry hamper | coat rack | bucket | ceiling light | window |
| dish rack | | paper bag | bathtub | bathroom stall door | folded chair | stool | ladder | power outlet |
| range hood | | alarm clock | desk | ceiling light | fire alarm | door | paper towel dispenser | power strip |
| dustpan | | laundry detergent | wardrobe | trash bin | power strip | sign | shower floor | bathtub |
| handicap bar | | object | clothes dryer | bathroom cabinet | calendar | recycling bin | stuffed animal | column |
| mailbox | | ceiling light | radiator | structure | poster | shower | water cooler | fire alarm |
| music stand | | dumbbell | shelf | storage organizer | luggage | jacket | coffee kettle | storage container |
| bathroom counter | | tube | | potted plant | | bottle | kitchen counter | |
| bathroom vanity | | cd case | | mattress | | | | |
| laundry hamper | | coffee kettle | | | | | | |
| trash bin | | shower head | | | | | | |
| keyboard piano | | case of water bottles | | | | | | |
| folded chair | | fire alarm | | | | | | |
| luggage | | power strip | | | | | | |
| mattress | | calendar | | | | | | |
| | | poster | | | | | | |
| | | potted plant | | | | | | |

incorrectly identifies the same sofa as separate instances. Similarly, in row 5, the washing machine is segmented into two instances by the baseline. In contrast, the proposed method delivers results that closely align with the ground truth, demonstrating its superior performance

Similarly, Fig. 6 highlights the results on Split-C, where classes are encountered in random order. The comparison emphasizes the advantages of our method, as highlighted by red circles. The baseline method often misses instances or splits a single instance into

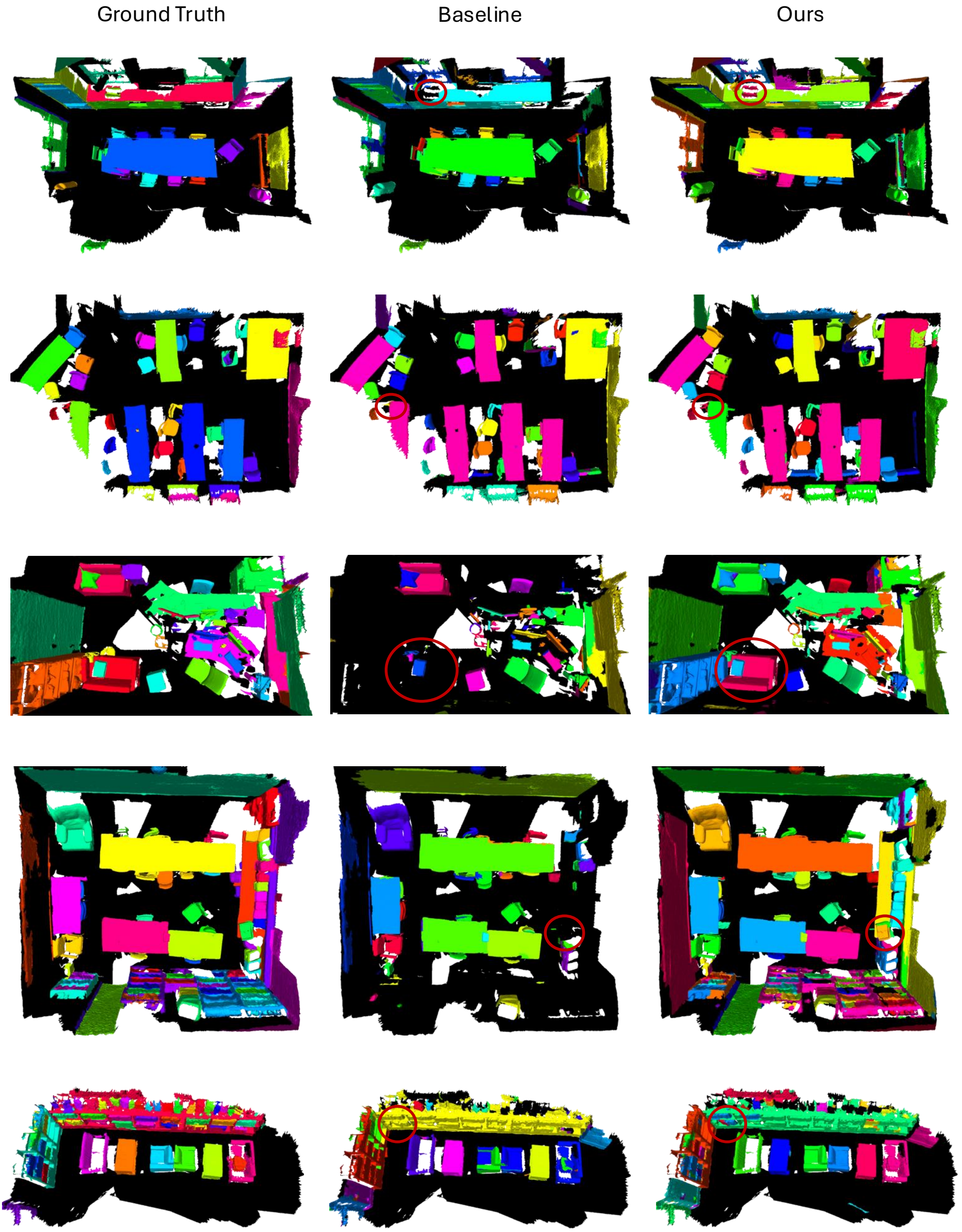


Figure 4: Qualitative comparison of ground truth, the baseline method, and our proposed framework on the `Split-A` evaluation after learning all tasks.

multiple parts. In contrast, our approach consistently produces results that are closely aligned with the ground truth, further underscoring its effectiveness.

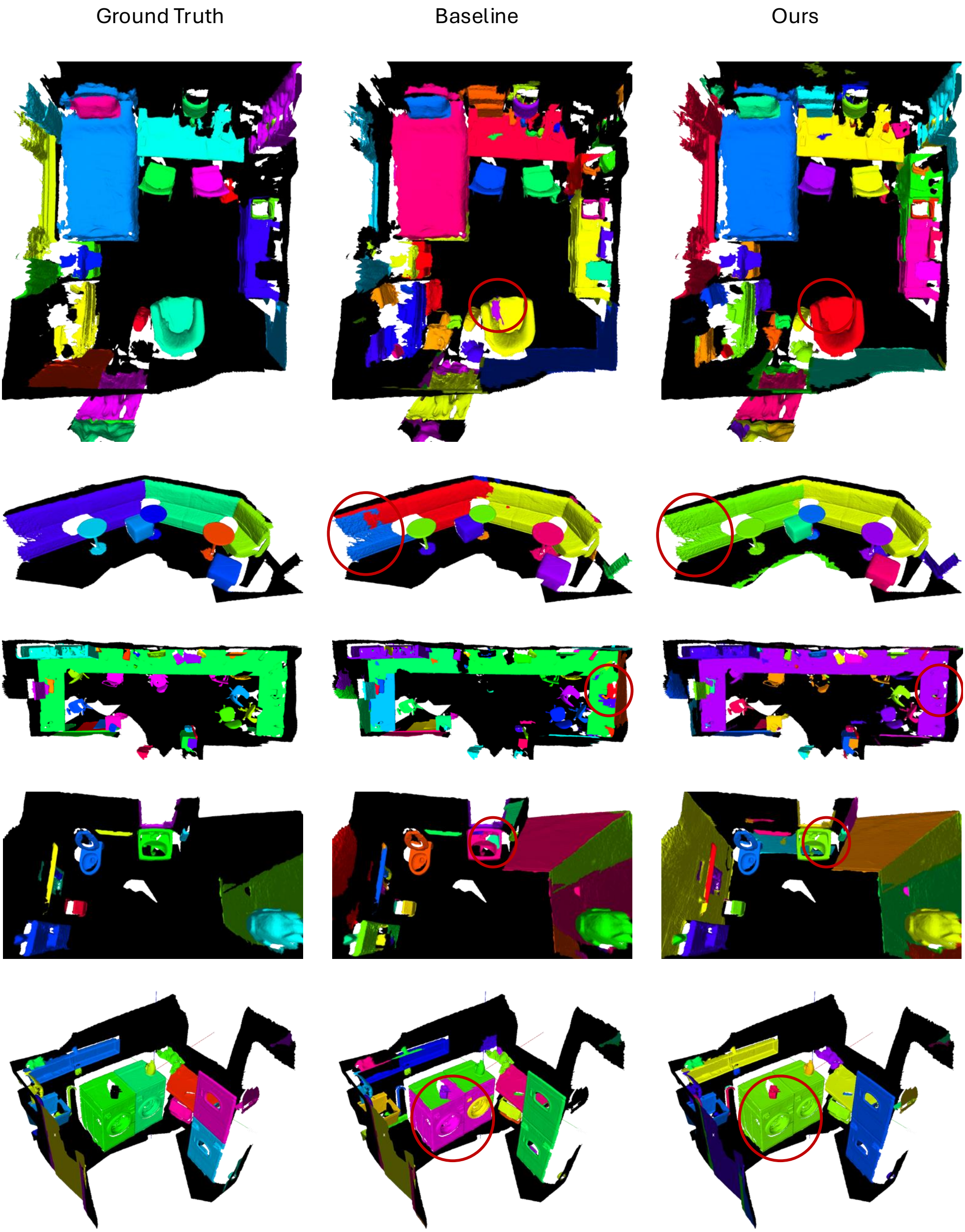


Figure 5: Qualitative comparison of ground truth, the baseline method, and our proposed framework on the Split-B evaluation after learning all tasks.

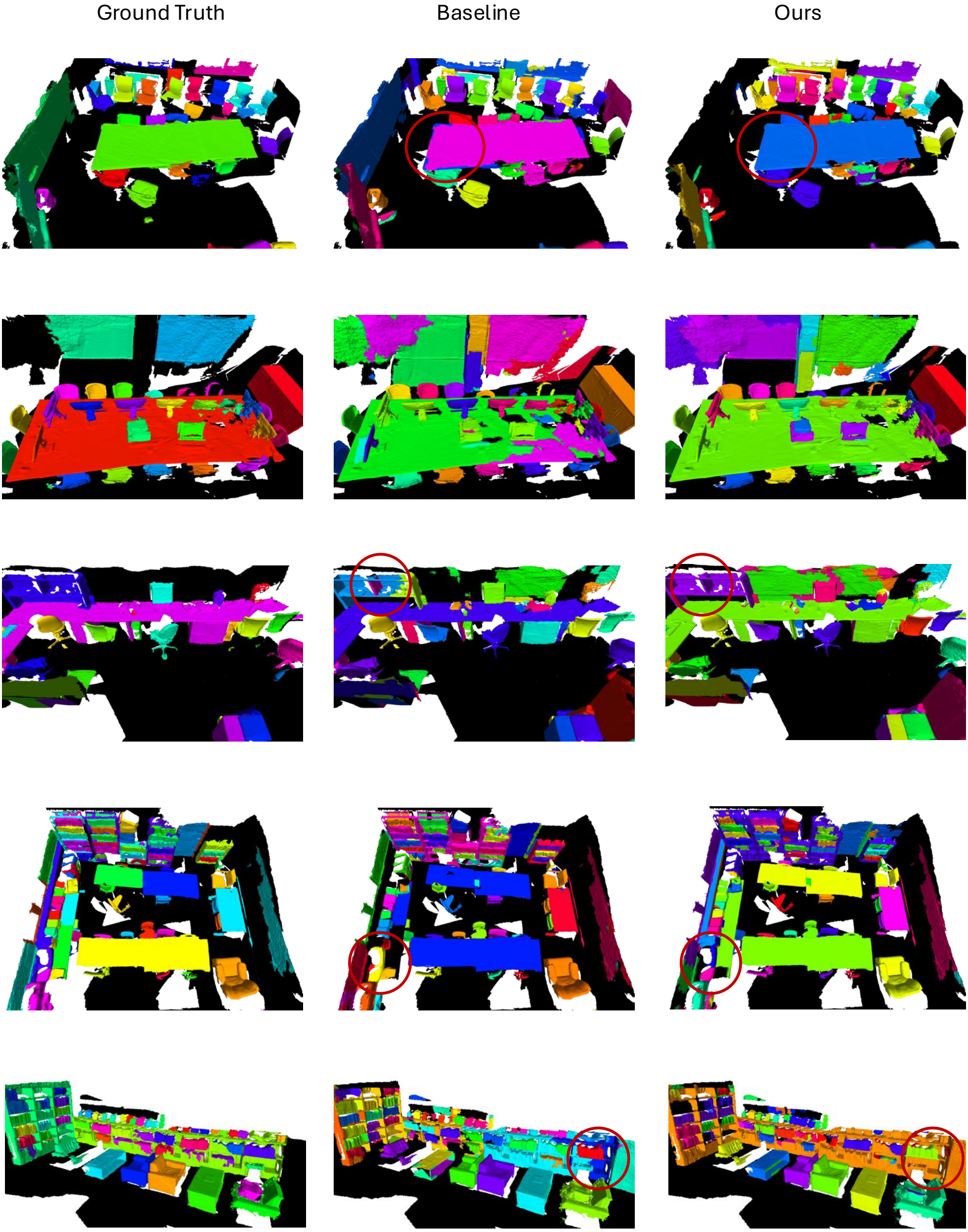


Figure 6: Qualitative comparison of ground truth, the baseline method, and our proposed framework on the Split-C evaluation after learning all tasks.