CoMBO: Conflict Mitigation via Branched Optimization for Class Incremental Segmentation

Supplementary Material

Contents

8. More Analysis

In this section, we present additional experimental results to further examine the effectiveness of the proposed components, including an ablation study of parameter freezing on queries and QCR modules, component ablations within CISS, an ablation study of IKD, and an analysis of prediction results from different layers of class embeddings. These experiments provide us with a deeper understanding of the individual contributions and interactions of each component, shedding light on their specific roles and the ways in which they enhance overall system performance.

8.1. Ablation Study of Parameter Freezing

As mentioned in Fig. 2, the parameters of the query embeddings Q and the QCR modules corresponding to the previous incremental classes $C^{2:t-1}$ are frozen during step t. To evaluate the effectiveness of this parameter-freezing strategy in incremental learning, we conducted a series of experiments. As shown in Tab. 5, freezing both the query embeddings Q and QCR modules $f_{QCR,\tilde{c}}$ with $\tilde{c} \in C^{2:t-1}$ results in an improvement of 0.4% compared to the unfreezing method. This demonstrates the strategy's efficacy in retaining knowledge of old classes while accommodating new classes. Thus, the results prove that the parameter freezing strategy avoids disturbing the impressionable query embeddings. Besides, it is important to note that freezing the parameters of queries does not mean keeping the queries of Q_l static when $l > 1$. The optimization of Q_l mainly affects features from the pixel-decoder, where the model integrates knowledge of new classes to enhance feature extraction. Furthermore, the QCR module $f_{QCR,\tilde{c}}$, as a lightweight adapter, encounters challenges similar to query embeddings, where limited pseudo labels of previous incremental classes $C^{2:t-1}$ could result in overfitting and misguidance, causing 1.3% decreasing on these incremental classes. In such cases, freezing the relevant parameters offers a simpler and more effective alternative to distillation, mitigating these risks and maintaining performance stability.

Q	f_{QCR}	**100-10** (6 steps) 1-100	101-150	all
✓	✓	40.3	25.2	35.2
✓		40.7	25.0	35.5
	✓	41.0	23.9	35.3
		40.8	25.2	**35.6**

Table 5. Ablation study of the parameter freezing strategy on the query embeddings Q and QCR modules $f_{QCR,\tilde{c}}$ of previous incremental classes $\tilde{c} \in C^{2:t-1}$. The experiments are conducted on the CIPS 100-10 task of the ADE20K. Note that the "✓" denotes learnable parameters.

8.2. Component Ablations in CISS

In this section, we evaluate the components of our proposed framework, including the Half-Distillation-Half-Learning Strategy, Importance-based Knowledge Distillation (IKD), and Query Conflict Reducing (QCR) module, in the 100-10 scenario of the CISS task. We analyze various combinations of these components and present the results in Tab. 6. The baseline approach employs standard losses from Mask2Former [17] with pseudo-labeling. Under the same experimental setup, the inclusion of the HDHL strategy leads to a significant improvement in overall mIoU by 3.4%, highlighting its ability to seamlessly integrate logits from new classes into the existing logits distribution while avoiding conflicting losses. The introduction of IKD leads to an impressive 4.1% increase in mIoU for old classes, showcasing its effectiveness in mitigating catastrophic forgetting by selectively distilling important knowledge. Additionally, by incorporating the QCR module for branched optimization, the proposed CoMBO further enhances performance on both old and new classes, with respective mIoU improvements of 0.8% and 0.2%, surpassing the limitations of previous state-of-the-art methods in balancing these two aspects. As a result, the overall mIoU increases by 4.6% compared to the baseline. These findings emphasize the efficacy of the proposed approach CoMBO and its associated components in reducing conflicts within model structures and losses, achieving substantial performance gains.

HDHL	IKD	QCR	**100-10** (6 steps) 1-100	101-150	all
			42.9	23.6	36.5
✓			46.5	26.8	39.9
✓	✓		47.0	27.5	40.5
	✓	✓	46.1	27.1	39.8
✓	✓	✓	47.8	27.7	**41.1**

Table 6. Ablation study of the main components on task 100-10 of CISS. Baseline in the 1st row uses vanilla losses with pseudo-labeling.

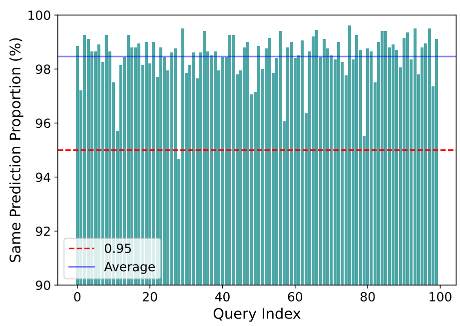

Figure 6. Proportion of samples with the same classification predictions between $\mathcal{E}_{cls,L-1}$ at layer $L-1$ and $\mathcal{E}_{cls,L}$ at layer L without the QCR module. The results indicate that nearly all embeddings from the queries have more than 95% samples with the same predictions, with an average proportion exceeding 98%.

8.3. Ablation Study of IKD

Table 7 presents the ablation study on the operations with the Importance-based Knowledge Distillation (IKD). The experiments are conducted on the ADE20K dataset under the CISS 100-10 scenario. This study evaluates the impact of three key operations in IKD: **Importance**, which represents importance of each query on the previous classes, **Weight**, which determines whether the importance vector is weighted in each step based on the number of classes, and **Norm**, which denotes whether min-max normalization is applied to the importance vector. The 1st row represents the baseline setup, where the distillation importance of all queries are uniformly set to 1.0, without applying either weighting or normalization. The results reveal the following trends. Without any additional operations (1st row), the model achieves mIoU scores of 46.8%, 26.6%, and 40.1% for old classes (1-100), new classes (101-150), and all classes, respectively. Applying importance and min-max normalization (3rd row) improves the performance on new classes (27.9% compared to 26.6%), resulting in a slight increase in overall mIoU to 40.3%. Using importance and weighting (4th row) remarkably enhances the old class mIoU to 48.0%, while maintaining comparable performance on new classes. Finally, combining both weighting and min-max normalization (5th row) achieves the best overall performance, with the mIoU of 41.1%, including balanced improvements for both old (47.8%) and new classes (27.7%). These results highlight the complementary roles of weighting and normalization in improving the performance of the IKD.

Importance	**Weight**	**Norm**	**100-10** (6 steps) 1-100	101-150	all
			46.8	26.6	40.1
✓			47.2	27.5	40.6
✓		✓	46.5	27.9	40.3
✓	✓		48.0	26.4	40.8
✓	✓	✓	47.8	27.7	**41.1**

Table 7. Ablation study on operations of the IKD module. The experiments are conducted on the CISS 100-10 task of the ADE20K. Note that the "✓" denotes whether the operation is utilized.

8.4. Analysis of Class Embeddings

We introduce the QCR module in Sec. 4.1, where the classification prediction from the class embedding $\mathcal{E}_{cls,L-1}$ of layer $L-1$ determines whether using QCR and which QCR should be selected according to the class prediction. Therefore, the premise of using the QCR module effectively is that the classification prediction from $\mathcal{E}_{cls,L-1}$ is the same as the classification prediction from $\mathcal{E}_{cls,L}$ of layer L. Only the inheritable prediction between adjacent layers could enable the class-specific adaptation from the QCR module focusing on its corresponding class. We record the proportion of the samples with the same classification prediction results between the $\mathcal{E}_{cls,L-1}$ and $\mathcal{E}_{cls,L}$ w/o QCR module of current classes in Fig. 6. The result shows that almost all the results from the corresponding queries meet the requirement on more than 95% samples, and the average proportion reaches above 98%. These statistics support the premise of our proposed QCR module, and the ablation studies in Sec. 6.1 and Sec. 8.2 show the effectiveness of QCR module that provides a more harmonious branched optimization structure.

8.5. Hyperparameters setting

We present ablation studies of λ_{KL} and λ_{IKD} in Tab. 8, where we analyze their impact on the CIPS 100-10 task of the ADE20K dataset. For λ_{KL}, performance improves as the value increases from 1 to 5, reaching the highest score of 35.6. However, further increasing λ_{KL} to 7 and 10 results in a slight decline, suggesting that excessive regularization may restrict model flexibility. Similarly, for λ_{IKD}, performance peaks at 35.61 when $\lambda_{IKD} = 3$, while larger values (5 and 10) show diminishing returns or slight degradation. Additionally, the λ_{cls} setting follows Mask2Former [17].

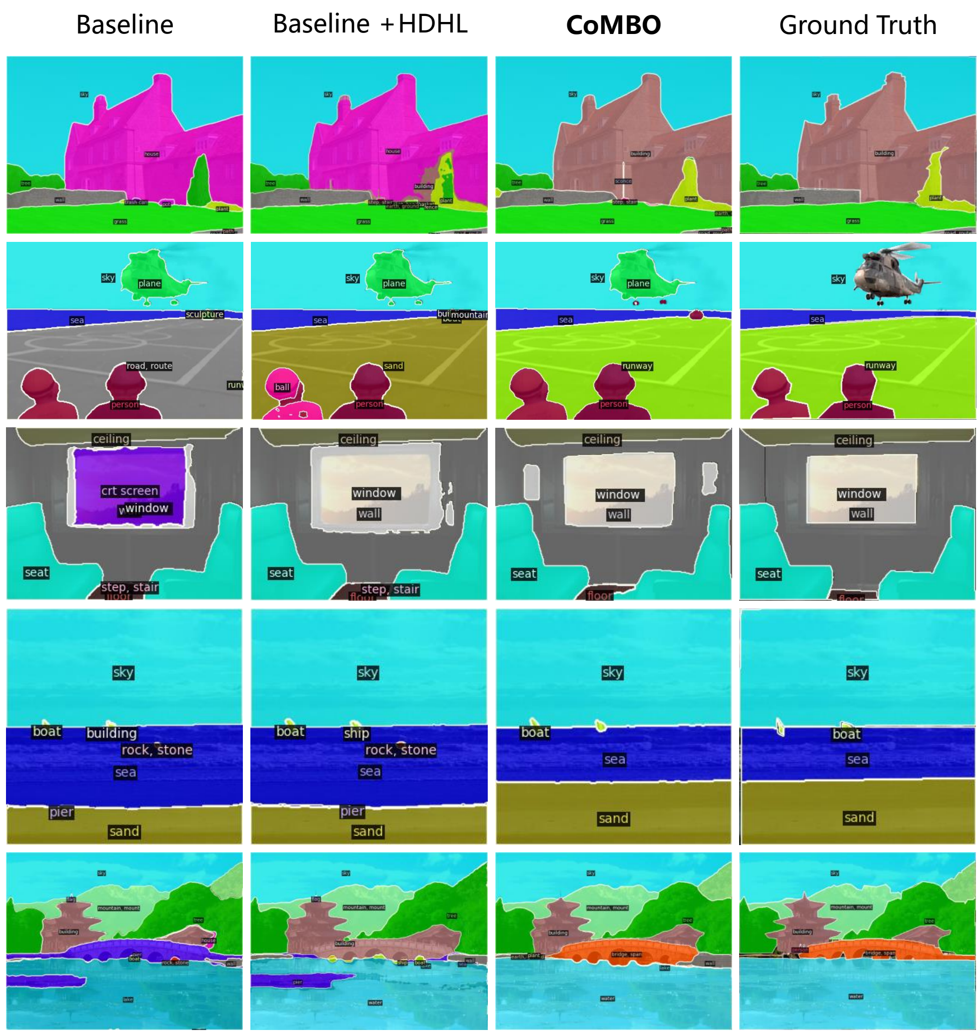

Figure 7. Qualitative results of CoMBO comparing to Baseline, Baseline+HDHL on 100-10 CISS task of ADE20K. Each class is uniquely represented by a specific color, making both boundary accuracy and correct color alignment with the ground truth essential for evaluation.

λ_{KL}	1	3	5	7	10
100-10	33.4	34.9	**35.6**	35.3	34.5
λ_{IKD}	0	1	3	5	10
100-10	34.88	35.38	**35.61**	35.60	34.93

Table 8. Ablation study on hyperparameter λ_{KL} and λ_{IKD}.

9. Additional Qualitative Results

In this section, we perform additional qualitative analysis by contrasting our proposed CoMBO method (3$^{\text{rd}}$ column) with both the Baseline (1$^{\text{st}}$ column) and Baseline+HDHL (2$^{\text{nd}}$ column) on the 100-10 scenario in the CISS, as shown in Fig. 7 and Fig. 8. In the 1$^{\text{st}}$ and 2$^{\text{nd}}$ columns, the Baseline fails to accurately recognize the old classes after the incremental learning, leading to incomplete or incorrect predictions for objects such as *Building* (1$^{\text{st}}$ row) and *runway* (2$^{\text{nd}}$ row). While Baseline+HDHL shows some improvement in segmenting new classes, it struggles with preserving the masks of initial classes, resulting in the misclassification of *sand* (4$^{\text{th}}$ row) and *Bridge* (5$^{\text{th}}$ row). In contrast, our CoMBO method (3$^{\text{rd}}$ column) successfully identifies the incremental classes, such as *stool* (4$^{\text{th}}$ row of Fig. 8), while maintaining accurate predictions for the old classes, as evidenced by the precise segmentation of *Building* (1$^{\text{st}}$ row) and *Earth* (7$^{\text{th}}$ row of Fig. 8). Additionally, CoMBO achieves finer boundary details for the segments, demonstrating improved refinement capabilities. These results highlight CoMBO's superior performance in reducing the conflict between the retention of old class knowledge and the acquisition of new class information.

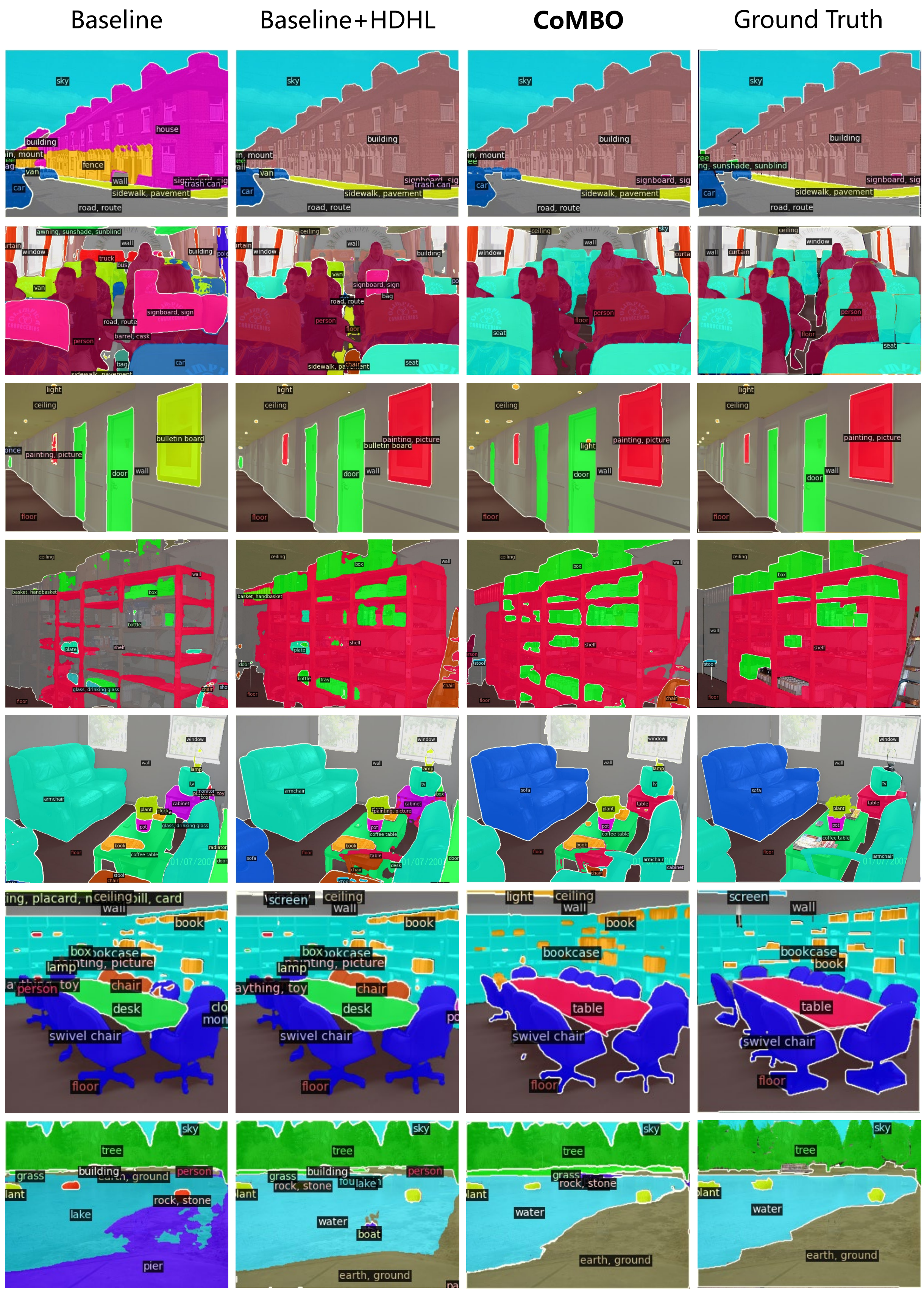

Figure 8. Qualitative results of CoMBO comparing to Baseline, Baseline+HDHL on 100-10 CISS task of ADE20K. Each class is uniquely represented by a specific color, making both boundary accuracy and correct color alignment with the ground truth essential for evaluation.