



Contact



HiPoSeg

Hierarchical **P**rototype Learning for Semantic **S**egmentation

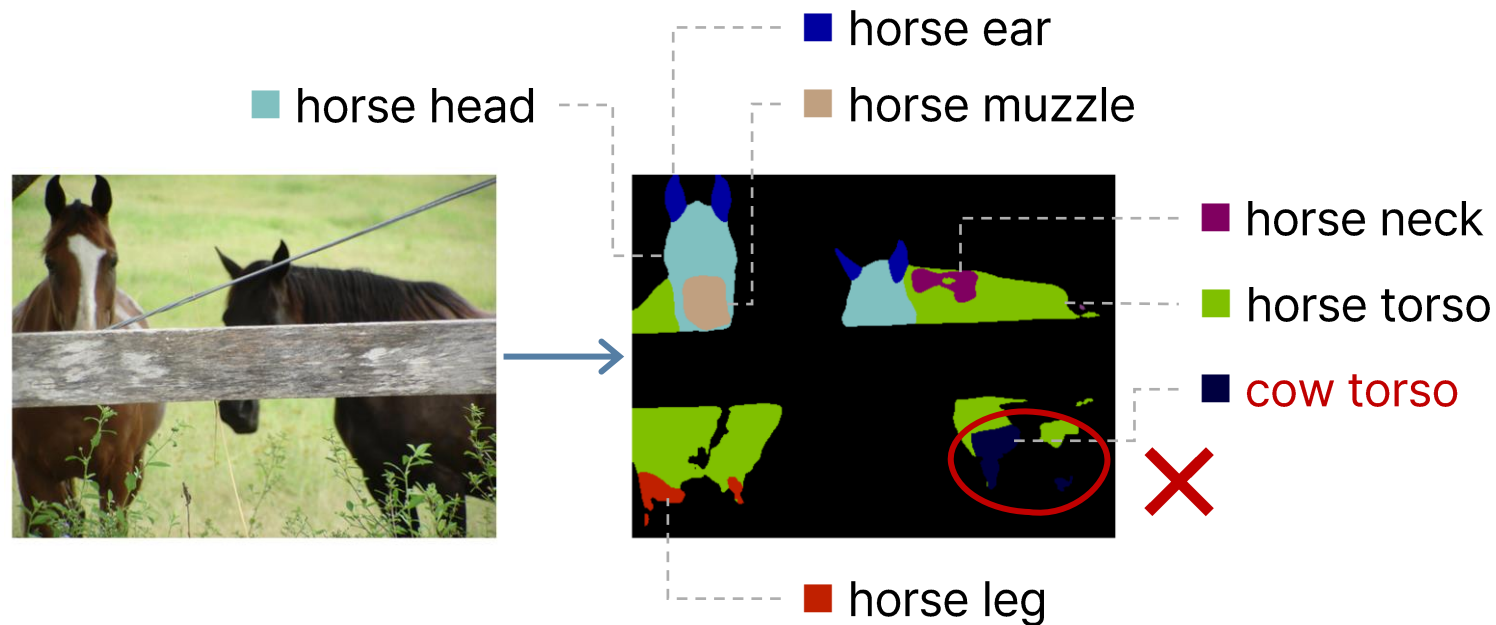
Seoha Lim*, Jinmyeong Kim*, Jieun Kim and Sung-Bae Cho

Yonsei University

{seoha815, jmkim_, lilly9928, sbcho}@yonsei.ac.kr

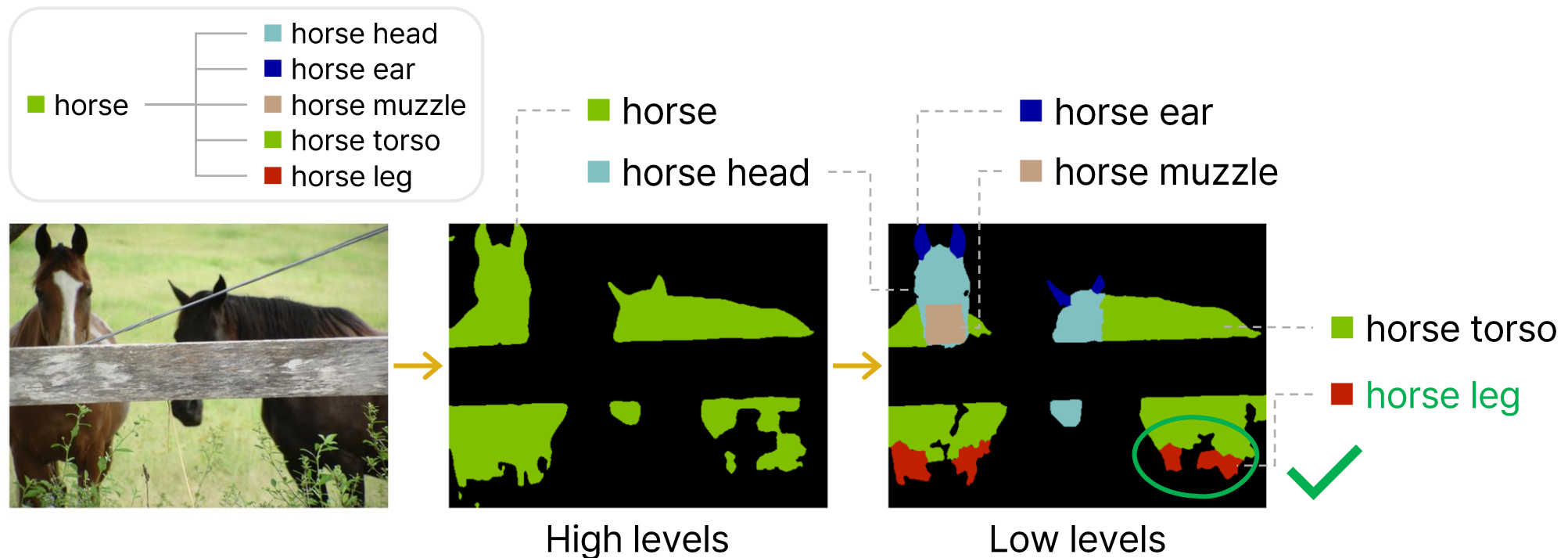
Why Flat Segmentation Fails?

- Models treat all classes **independently** → no structural context
- Visually similar parts across different objects are **easily confused**
- **Fine-grained / rare / small parts** are frequently misclassified



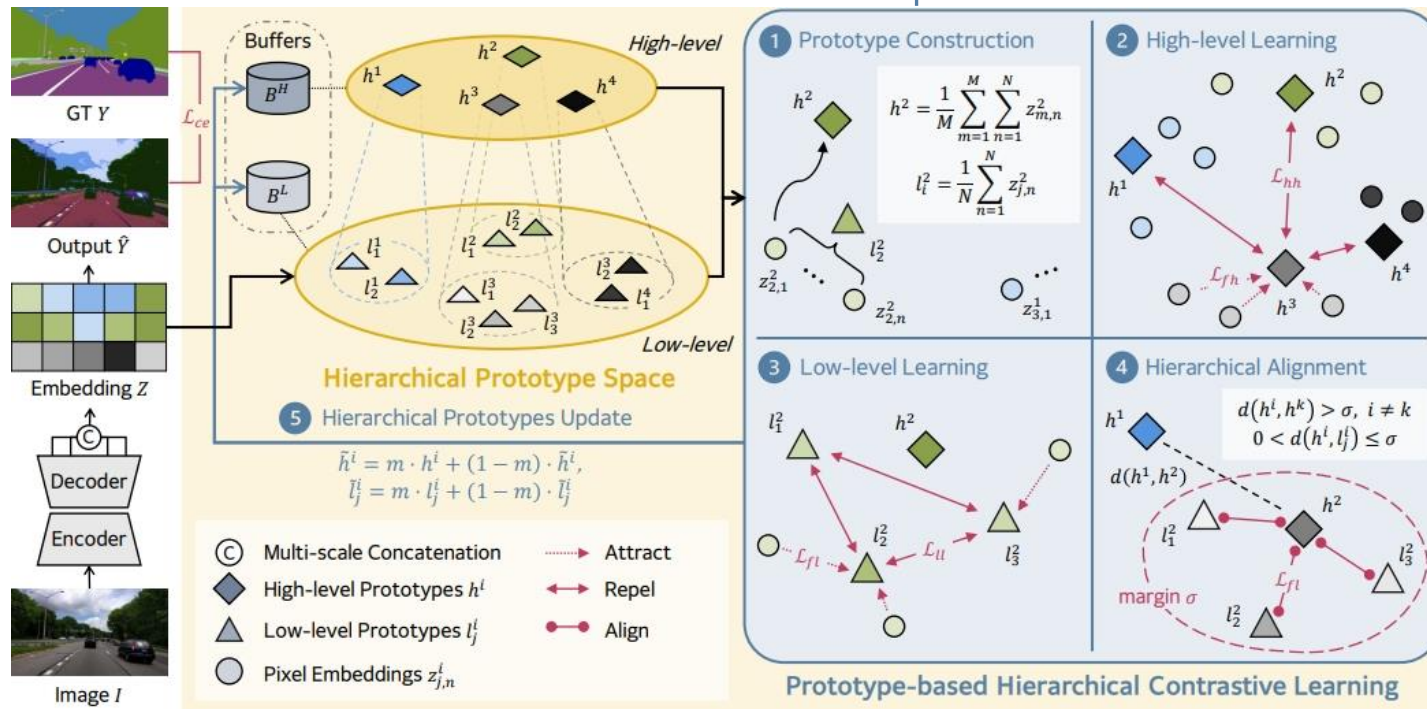
Coarse-to-Fine via Hierarchical Prototypes

- **High-level prototypes** → object-level semantics
- **Low-level prototypes** → part-level features
- **Alignment constraint** → child stays near parent; unrelated parents stay apart

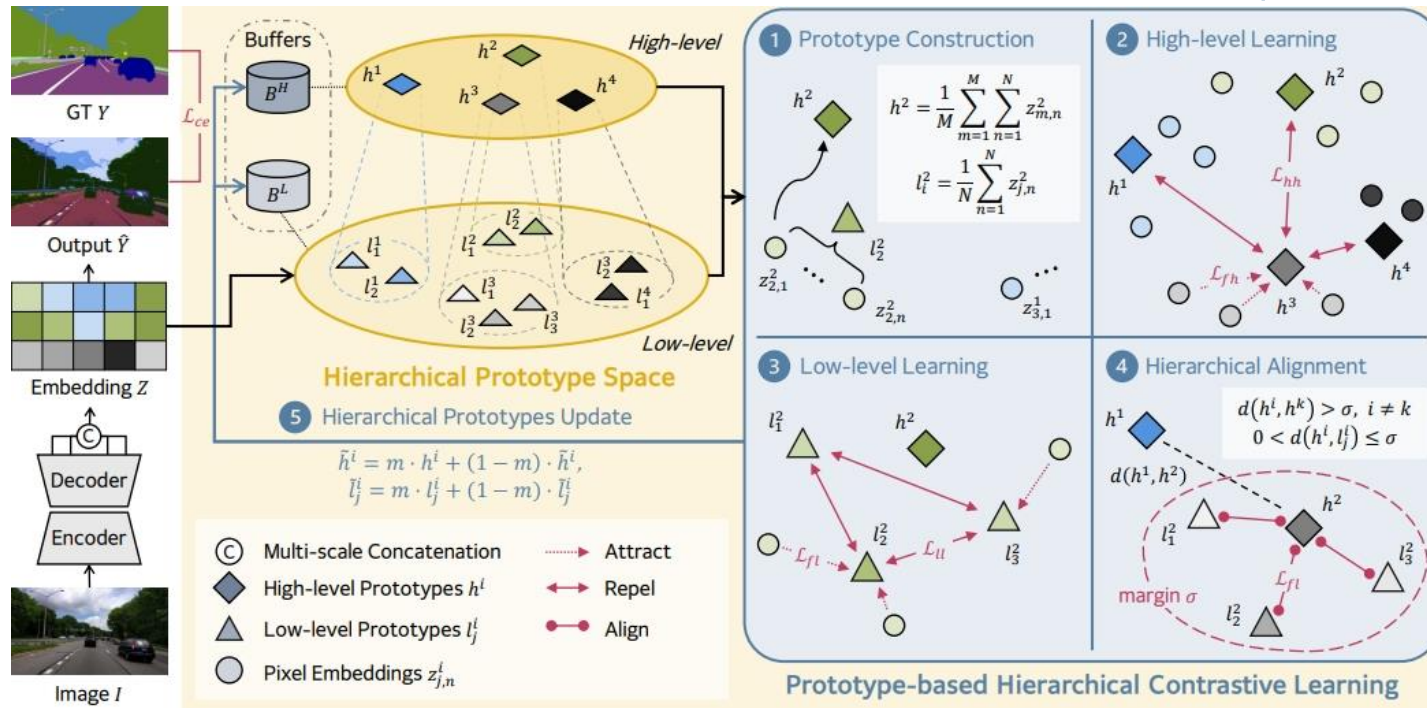


Hierarchical Prototype Learning

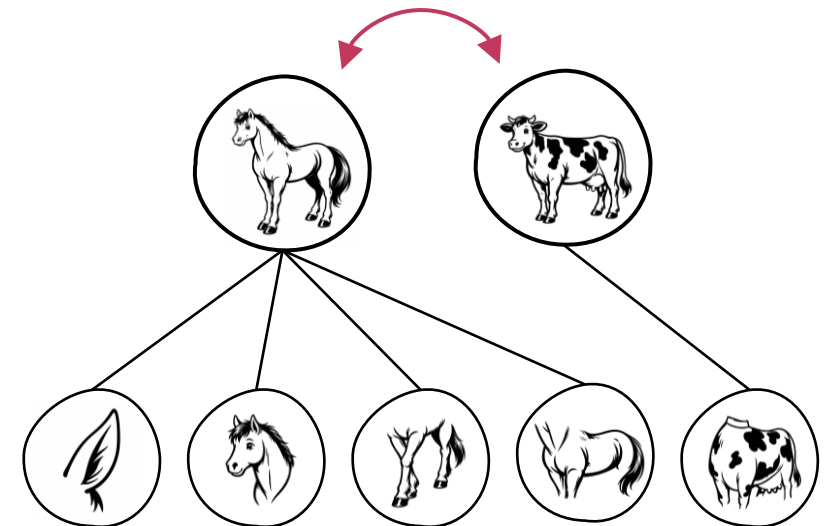
① Group pixel embeddings by class label at both levels



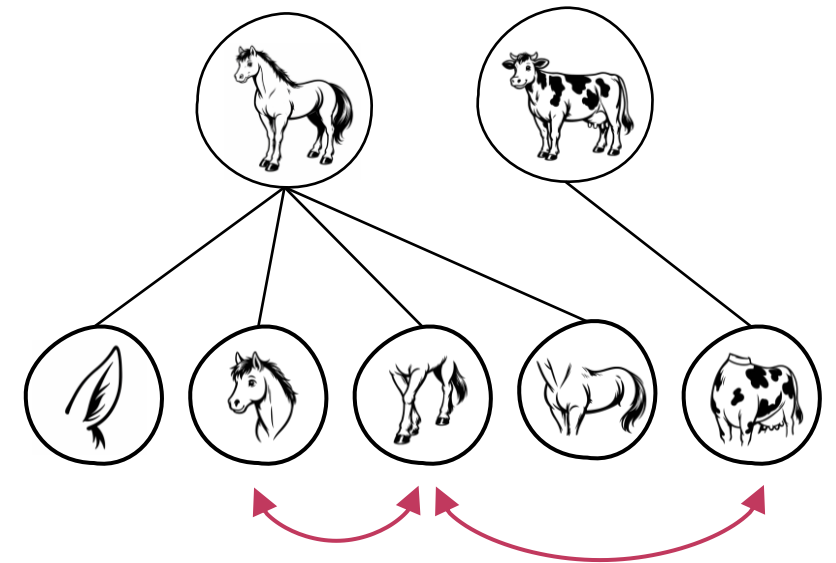
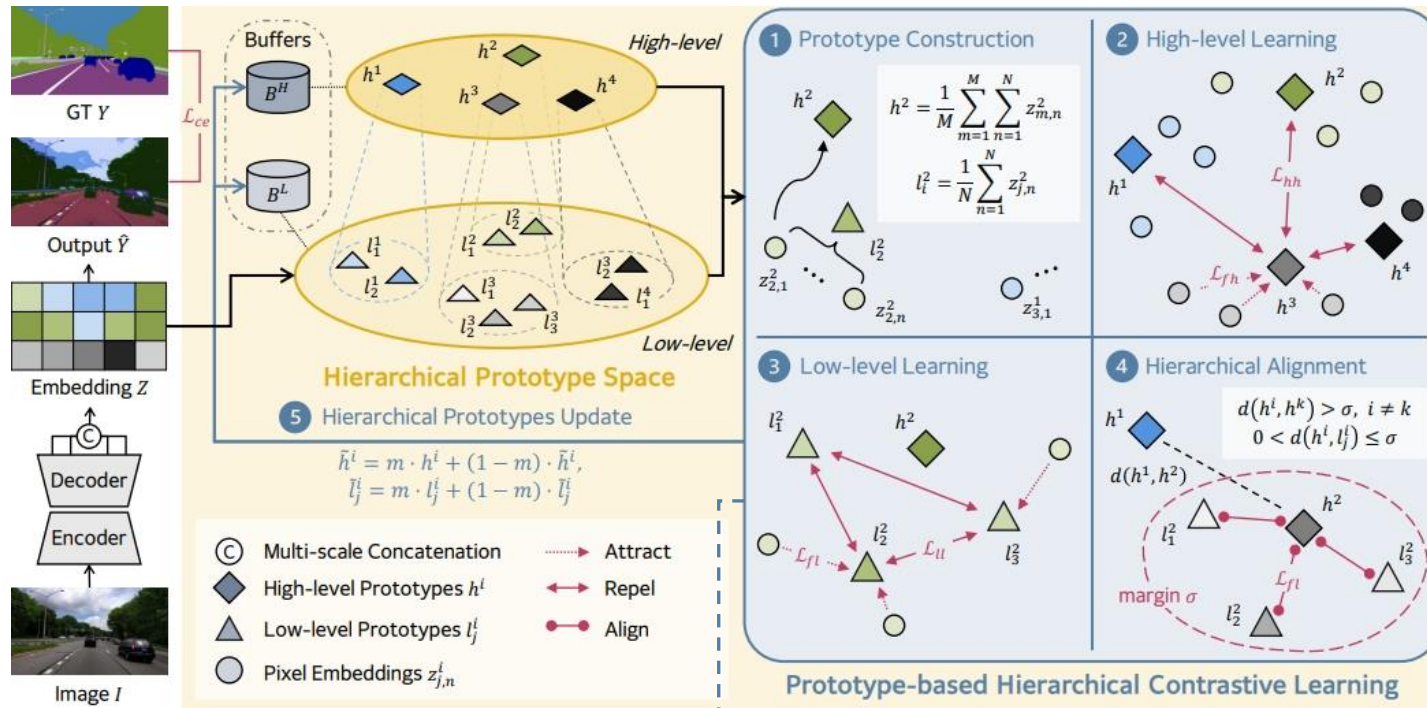
Hierarchical Prototype Learning



② Pull features to correct parent prototype, repel others



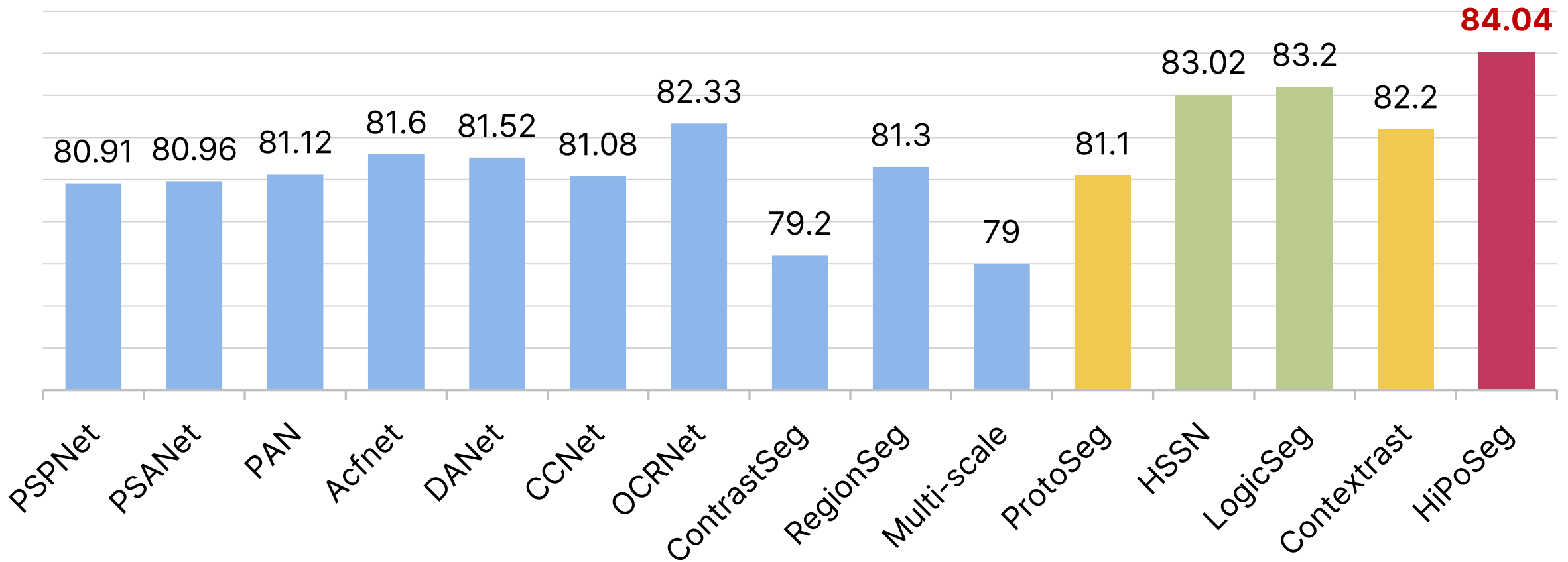
Hierarchical Prototype Learning



3 Same pull-push for fine-grained parts

Consistent SOTA Across All Benchmarks

Quantitative Results (mIoU, %)

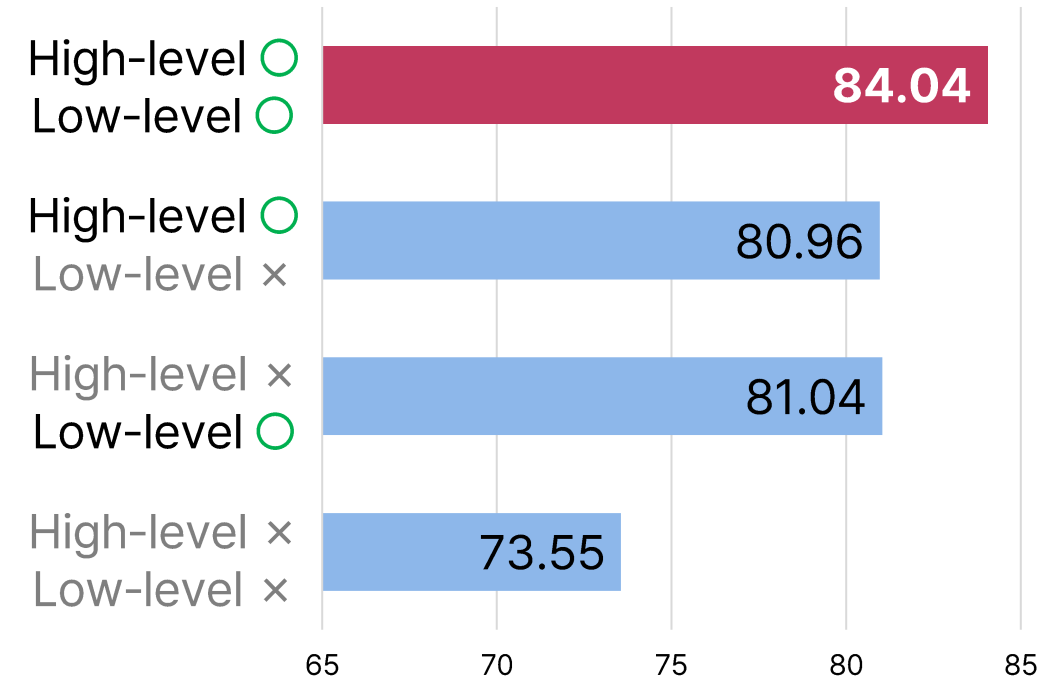


Alignment is the Key

Ablation of Loss Terms

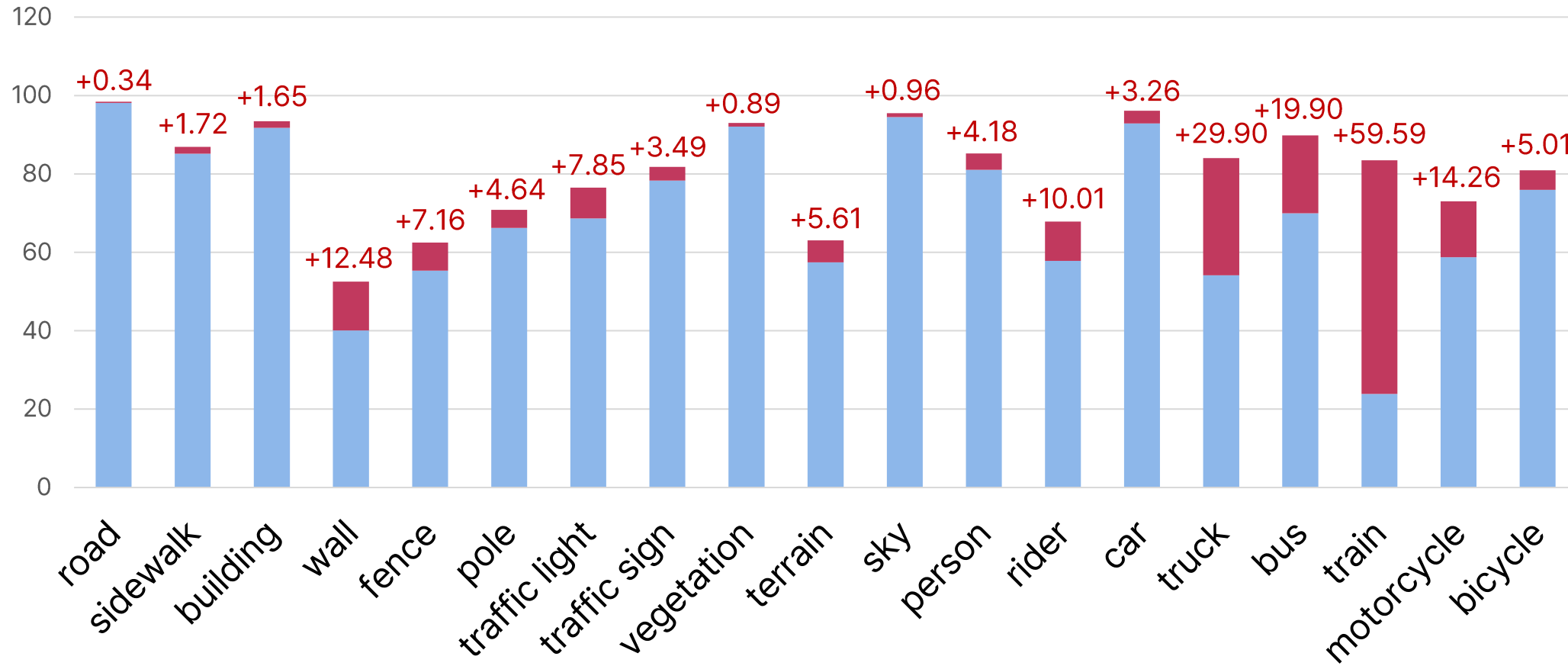
| CE loss | High-level loss | Low-level loss | Alignment loss | mIoU (%) |
|---------|-----------------|----------------|----------------|-----------------------|
| ✓ | | | | 73.55 |
| ✓ | ✓ | | | 80.96 (+7.41) |
| ✓ | | ✓ | | 81.04 (+7.49) |
| ✓ | ✓ | ✓ | | 79.16 (+5.61) |
| ✓ | ✓ | ✓ | ✓ | 84.04 (+10.49) |

Ablation of Hierarchy Levels



Low-level IoU Comparison

Class-wise Comparison between Baseline and HiPoSeg



HiPoSeg is better on rare/small parts

