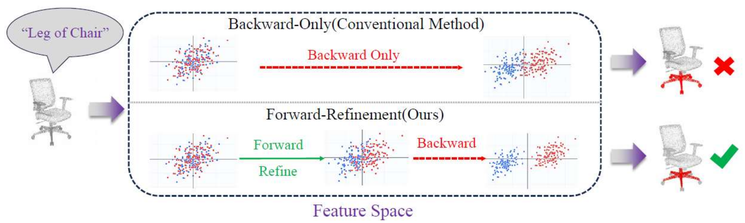
Open-Vocabulary 3D Part Segmentation with Semantic Propagation Hawkes Process

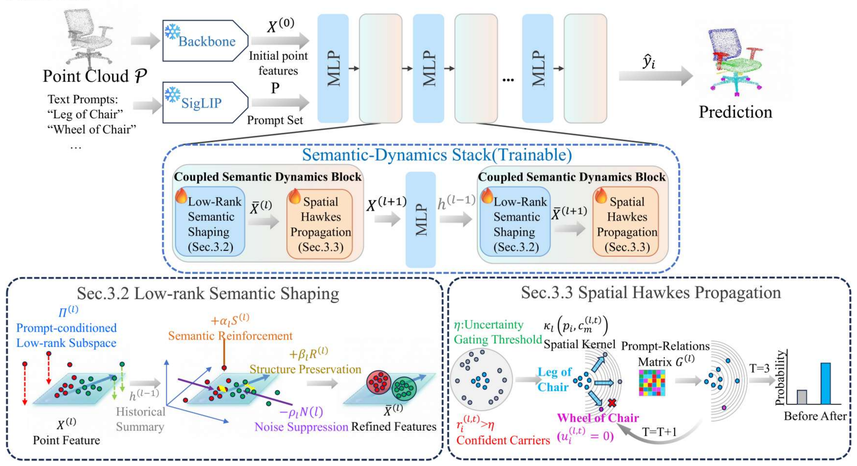
1 Motivation



- Open-vocabulary 3D part segmentation assigns fine-grained part labels to points from free-form or object-conditioned prompts, far beyond recognizing local geometric patterns.
- Repeated structures, sparse observations and irregular sampling split points of the same part, while locally similar regions may belong to different parts; under prompts such as "leg of a chair", local geometry alone cannot determine the semantics.
- Existing methods resolve inference by one-shot point–text similarity: semantic organization is shaped implicitly by the training loss, there is no explicit forward mechanism, and the final loss constrains outputs only indirectly.
- High-dimensional point features are contaminated by geometry-irrelevant perturbations, which amplify intra-part variance and blur the point–prompt correspondence.
- Context aggregation typically lacks explicit modeling of point-wise semantic confidence, allowing low-confidence ambiguous points to participate in information propagation, which can lead to undirected context mixing and even semantic contamination.
- The gap is therefore not more inter-point communication but forward semantic control: first stabilize prompt-relevant states, then propagate only reliable semantic carriers.
- These two mechanisms are tightly coupled throughout the network: low-rank semantic shaping organizes a propagation-ready semantic state, while the propagated context in turn refines subsequent shaping.

2 Method

Framework



- Given a point cloud and a prompt set from a frozen SigLIP encoder, L coupled blocks are stacked on frozen point features and the model predicts $\hat{y}_i = \text{argmax}_k \lambda_{ik}$ after the last block and step.
- Low-rank semantic shaping restricts feature updates to a prompt-conditioned rank-r subspace: αS reinforces prompt-consistent directions, −ρN suppresses off-subspace drift, and βR preserves local structural detail.
- Spatial Hawkes-style propagation turns confident responses ($r \geq \eta$) into sparse semantic events and accumulates their evidence over T finite steps, so ambiguous regions receive context from reliable carriers rather than from all points.
- Increments are weighted by point uncertainty, a locality kernel κ over a support radius, and a prompt-relation matrix G, so propagation refines relative prompt preference instead of inflating scores.
- Feedback coupling: the pooled propagated context conditions the next block, so shaping → scoring → carrier extraction → propagation → feedback all happen in a single pass.
- Trained end-to-end with the final segmentation loss, with discrete routing kept stop-gradient; no test-time iterative optimization and no post-processing.

Shaping

$$\bar{X}^{(l)} = X^{(l)} + \alpha_l S^{(l)} - \rho_l N^{(l)} + \beta_l R^{(l)}$$

Propagation

$$\lambda_{ik}^{(l,t+1)} = \lambda_{ik}^{(l,t)} + u_i^{(l,t)} \sum_{m=1}^{M_{l,t}} s_m^{(l,t)} \, \kappa_l\left(p_i, c_m^{(l,t)}\right) G^{(l)}_{k\, q_m^{(l,t)}}$$

3 Results

Table 1: Main benchmark comparison (object mIoU, %, higher is better). Part denotes free-form part prompts, and Obj-Part denotes object-conditioned prompts of the form *part of object*. Find3D reports the original results; Find3D* is its aligned-suite reproduction. PartSLIP++† denotes the fine-tuned canonical-view PartNetE setting. Best and second-best results are in bold and underlined.

I: 3DCompat200

Method	Venue	3DCompat200-Coarse				3DCompat200-Fine			
		Canonical		Rotated		Canonical		Rotated	
		Part	Obj-Part	Part	Obj-Part	Part	Obj-Part	Part	Obj-Part
PointCLIPV2	ICCV'23	14.16	14.09	13.39	13.17	7.18	7.25	7.20	7.43
PartSLIP	CVPR'23	6.54	1.27	8.03	1.74	1.56	0.23	4.44	0.71
PartSLIP++	arXiv'23	17.62	6.12	15.82	5.89	5.68	3.79	4.76	3.41
Find3D	ICCV'25	25.36	31.72	26.23	32.47	12.57	12.69	12.79	12.73
Find3D*	ICCV'25	45.00	36.89	46.02	38.14	24.97	17.72	26.35	18.57
CoSMo3D	CVPR'26	<u>54.52</u>	<u>47.51</u>	<u>54.55</u>	<u>47.74</u>	<u>31.29</u>	<u>24.48</u>	<u>30.97</u>	<u>24.46</u>
Ours	–	**57.94**	**52.82**	**57.69**	**52.65**	**35.90**	**29.63**	**35.35**	**29.16**

II: ShapeNetPart and PartNetE

Method	Venue	ShapeNetPart				PartNetE			
		Canonical		Rotated		Canonical		Rotated	
		Part	Obj-Part	Part	Obj-Part	Part	Obj-Part	Part	Obj-Part
OpenMask3D	NeurIPS'23	10.37	8.94	14.56	6.75	11.24	12.54	11.67	11.93
PartSLIP++†	arXiv'23	6.46	1.43	6.03	0.94	**32.71**	5.12	**23.03**	3.87
PointCLIPV2	ICCV'23	20.22	16.91	18.19	16.88	9.70	11.28	10.22	10.32
Find3D	ICCV'25	24.09	28.39	23.71	29.64	16.38	16.86	17.16	17.62
CoSMo3D	CVPR'26	<u>33.31</u>	<u>36.16</u>	<u>32.84</u>	<u>34.20</u>	17.19	<u>17.59</u>	18.17	<u>18.48</u>
Ours	–	**36.13**	**38.04**	**34.86**	**37.98**	<u>20.28</u>	**21.83**	<u>21.61</u>	**23.00**

Qualitative comparison

