NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2026

S²-RL

Sample-Set Dual Reinforcement Learning for Generative Semantic Segmentation Dataset Distillation

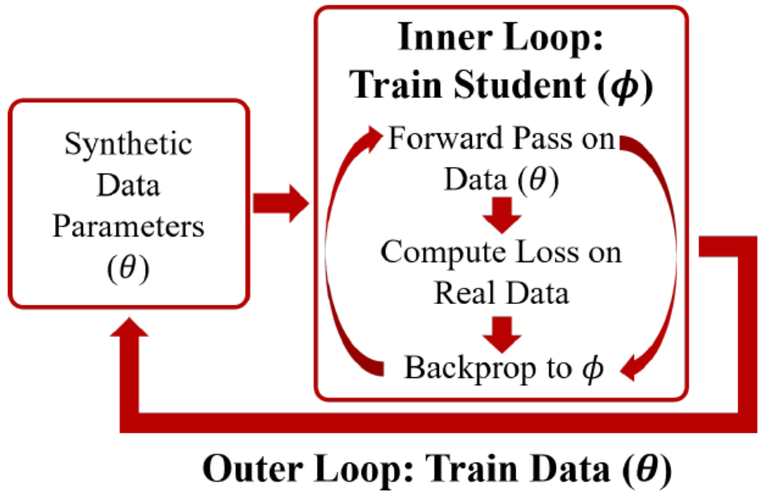
Haoyu Wang[1] Fei Zhou[2] Qingqing Qiu[3] Lei Zhang[1] Wei Wei[1] Chen Ding[4]

[1] Northwestern Polytechnical University [2] The Chinese University of Hong Kong
[3] Zhejiang University [4] Xi'an University of Posts & Telecommunications

Project

Dataset distillation for dense prediction

A compact training set must preserve classes and their spatial relationships.



(a) Optimization-based methods



(b) Proxy-based methods

Our objective

Generate informative, novel scene layouts from class-list prompts.

Dense gradients

Unrolling student training is memory-intensive.

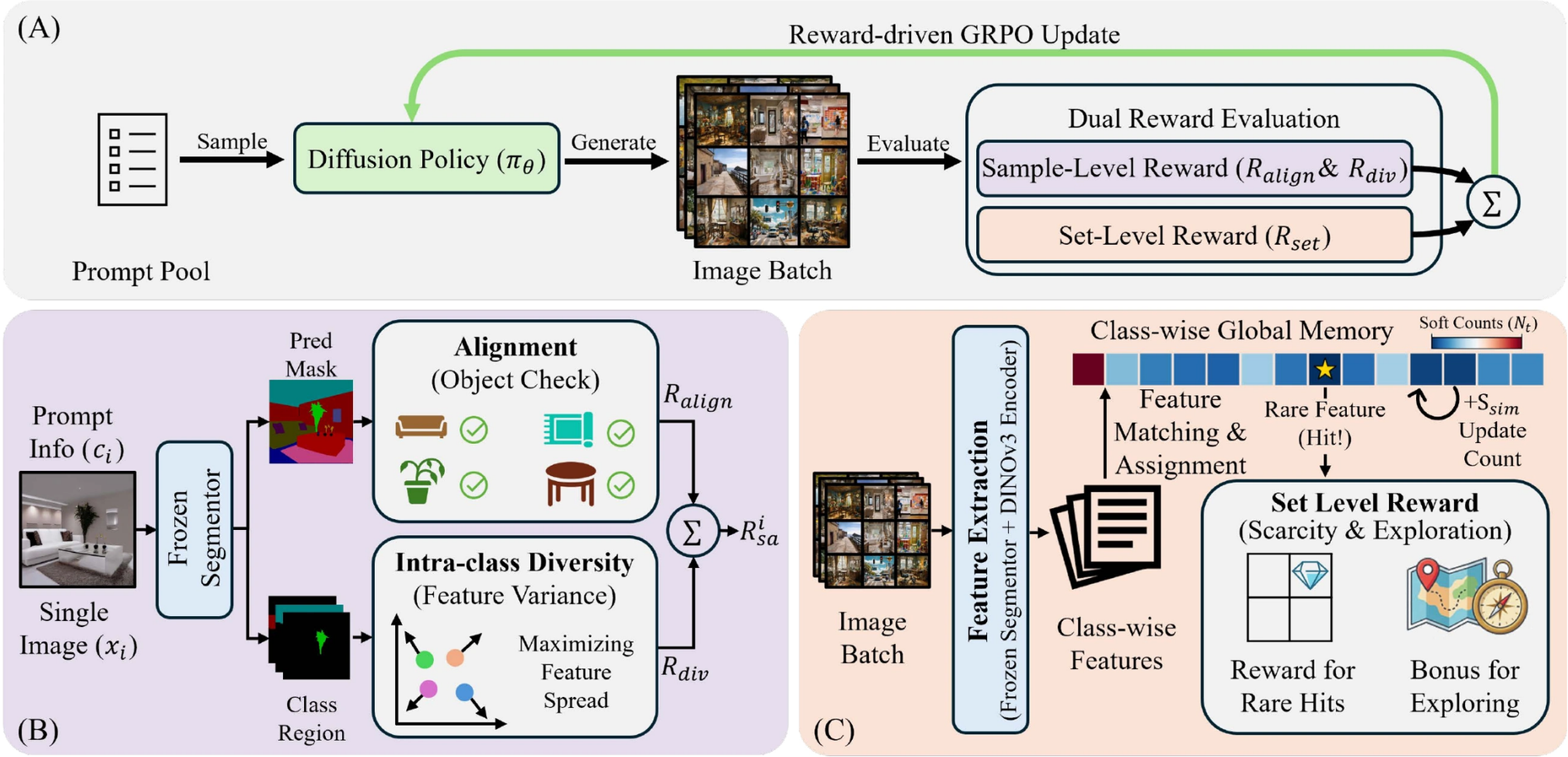
Fixed layouts

Mask conditioning limits spatial configurations.

Small data budgets

Each image should add complementary knowledge.

S²-RL framework



Generation

Class lists specify content. The policy chooses the layout.

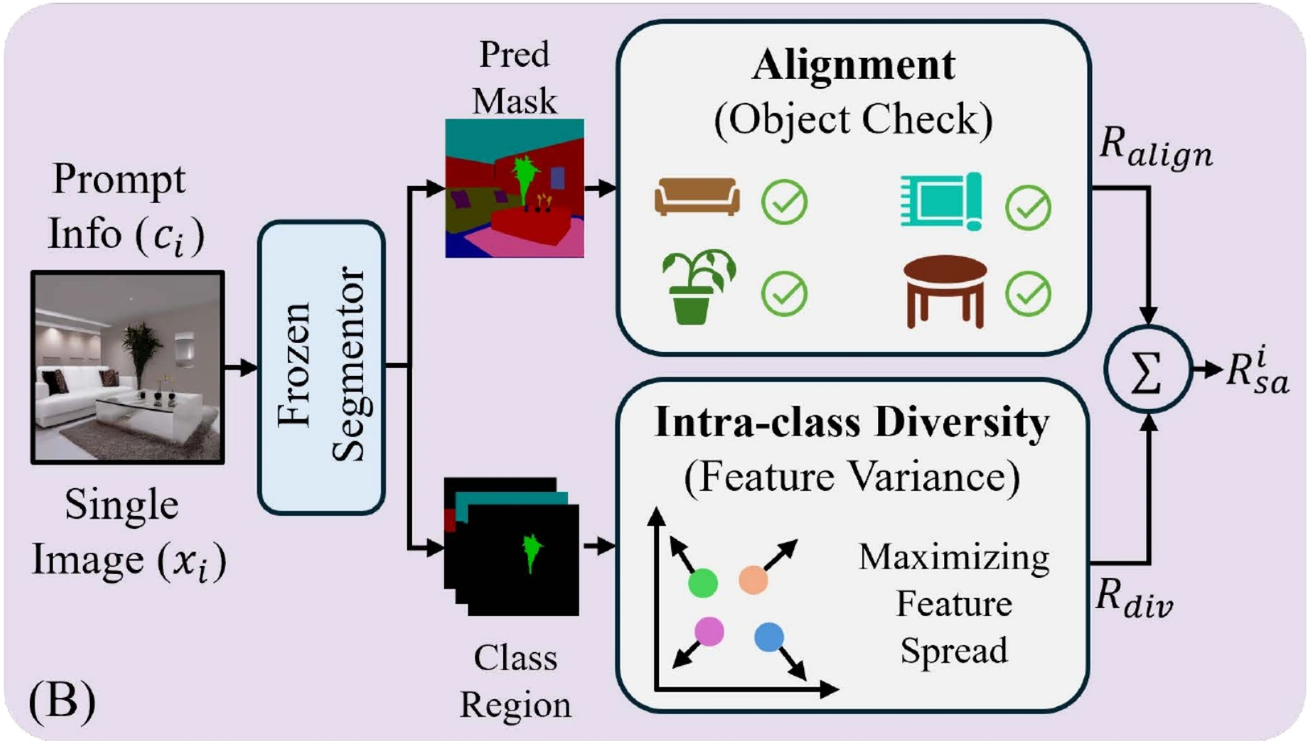
Policy update

GRPO combines sample and set rewards to update SD3.5.

Candidate screening and teacher pseudo-labels produce the final training set.

Sample-level reward: informative individual images

Semantic alignment and intra-class feature diversity provide complementary feedback.



Semantic alignment

Check requested categories and their spatial occupancy using frozen segmentors.

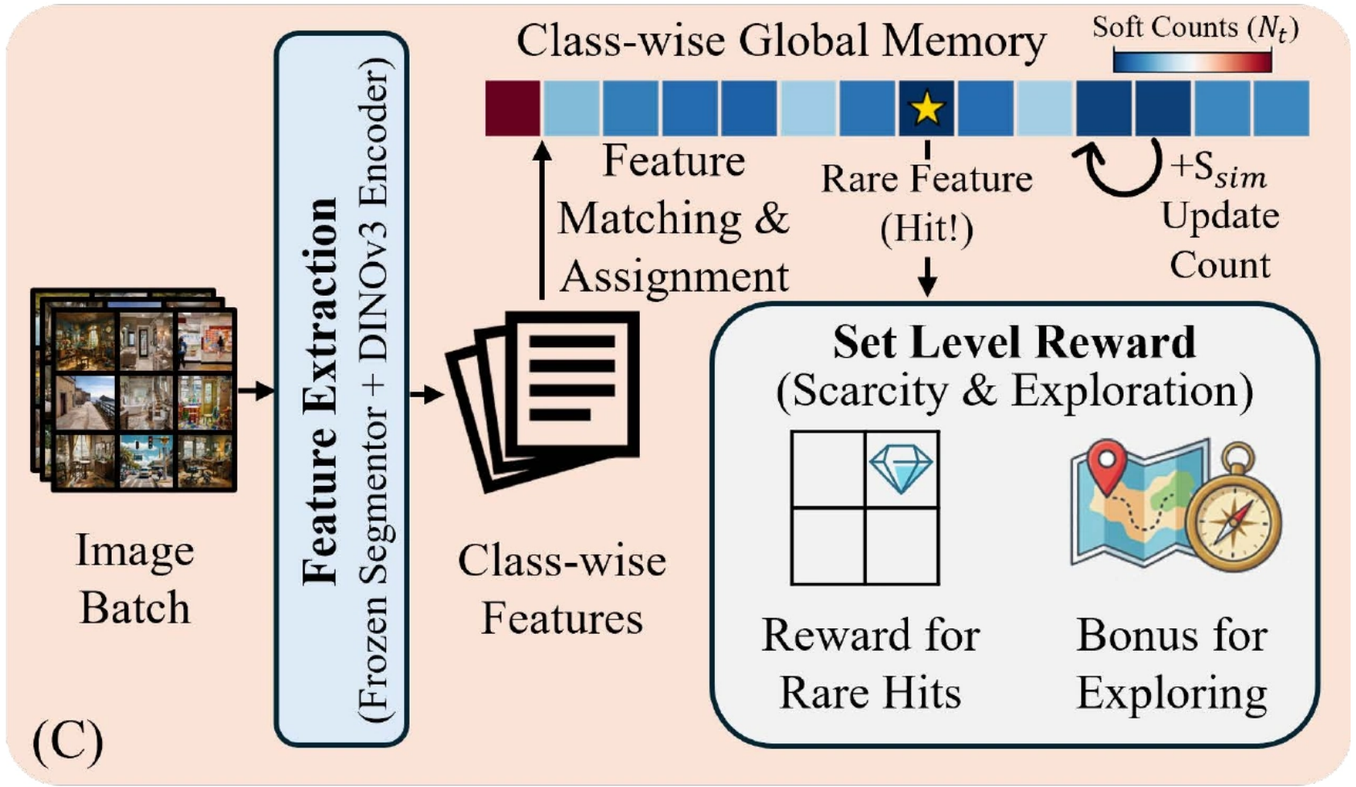
Intra-class diversity

Penalize excessively similar local features within the same semantic class.

Both terms evaluate the content of an individual generated image.

Set-level reward: coverage across generated samples

Prototype memory rewards semantic modes that previous samples cover less well.



Class-wise prototypes

Up to 50 representatives per class from real training regions.

Scarcity + exploration

Larger rewards for underrepresented prototypes.

Within each class

Explore different appearances of the same semantic category.

Class-specific semantic prototypes define the arms of a contextual bandit.

ADE20K: gains across all three data budgets

SegFormer-B5. Mean ± standard deviation, mIoU (%).

Method	0.25% data	0.5% data	1.0% data
Random	9.59 ± 0.36	14.92 ± 0.45	19.95 ± 0.67
Uniform	12.05 ± 0.68	18.59 ± 0.47	24.10 ± 0.55
K-Center	12.56 ± 0.55	17.92 ± 0.93	23.90 ± 0.96
Herding	10.54 ± 0.17	15.44 ± 0.21	21.00 ± 0.21
SD3.5	13.54 ± 0.56	18.09 ± 0.90	21.66 ± 0.40
UniDD*	12.05 ± 0.64	17.81 ± 0.90	22.70 ± 0.56
S²-RL	**15.60 ± 0.59**	**21.52 ± 0.56**	**25.47 ± 0.24**

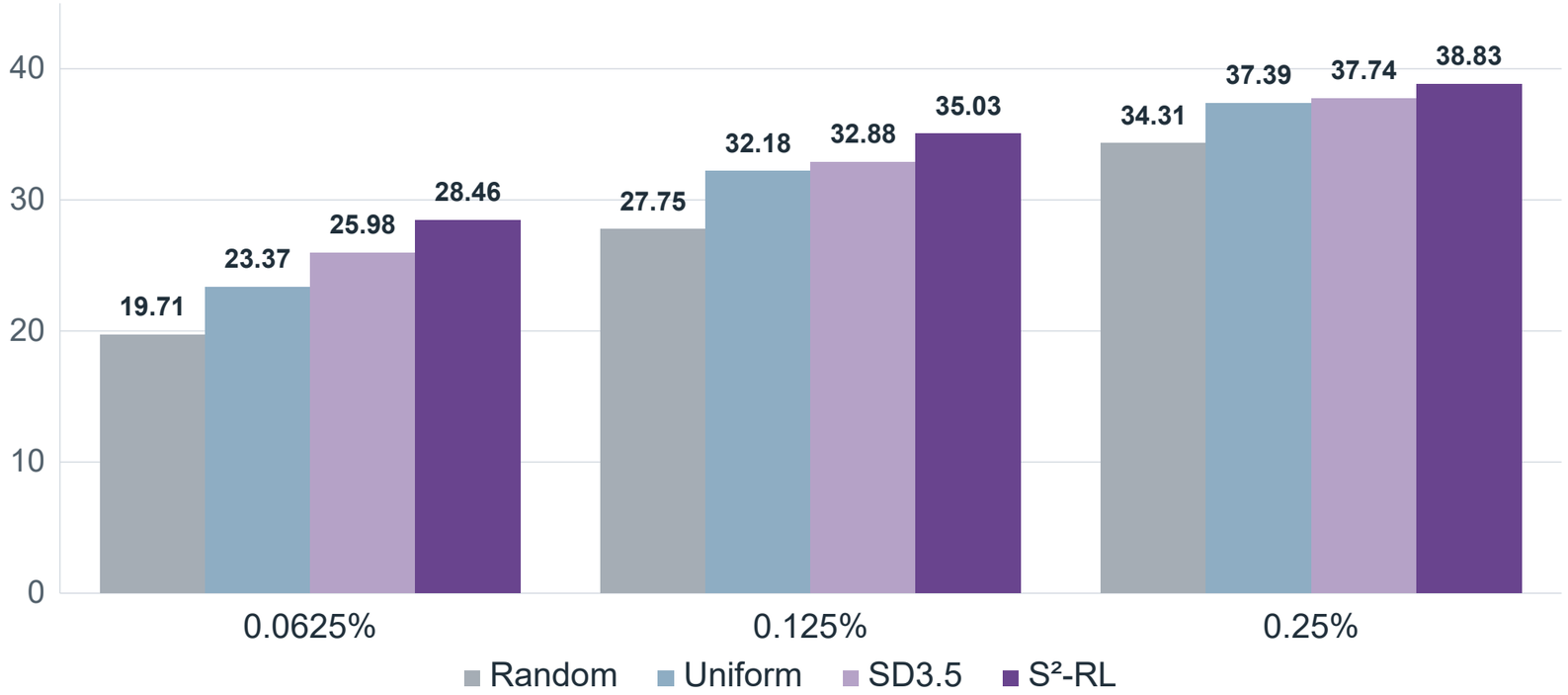
+2.06 points over SD3.5 at 0.25% data

+2.93 points over Uniform at 0.5% data

5 groups × 5 student seeds. Full-data: 49.13 mIoU. UniDD*: reproduced UniDD-inspired baseline.

COCO: gains persist across data budgets

SegFormer-B5. Mean mIoU (%).



38.83

mIoU at 0.25% data

+2.48

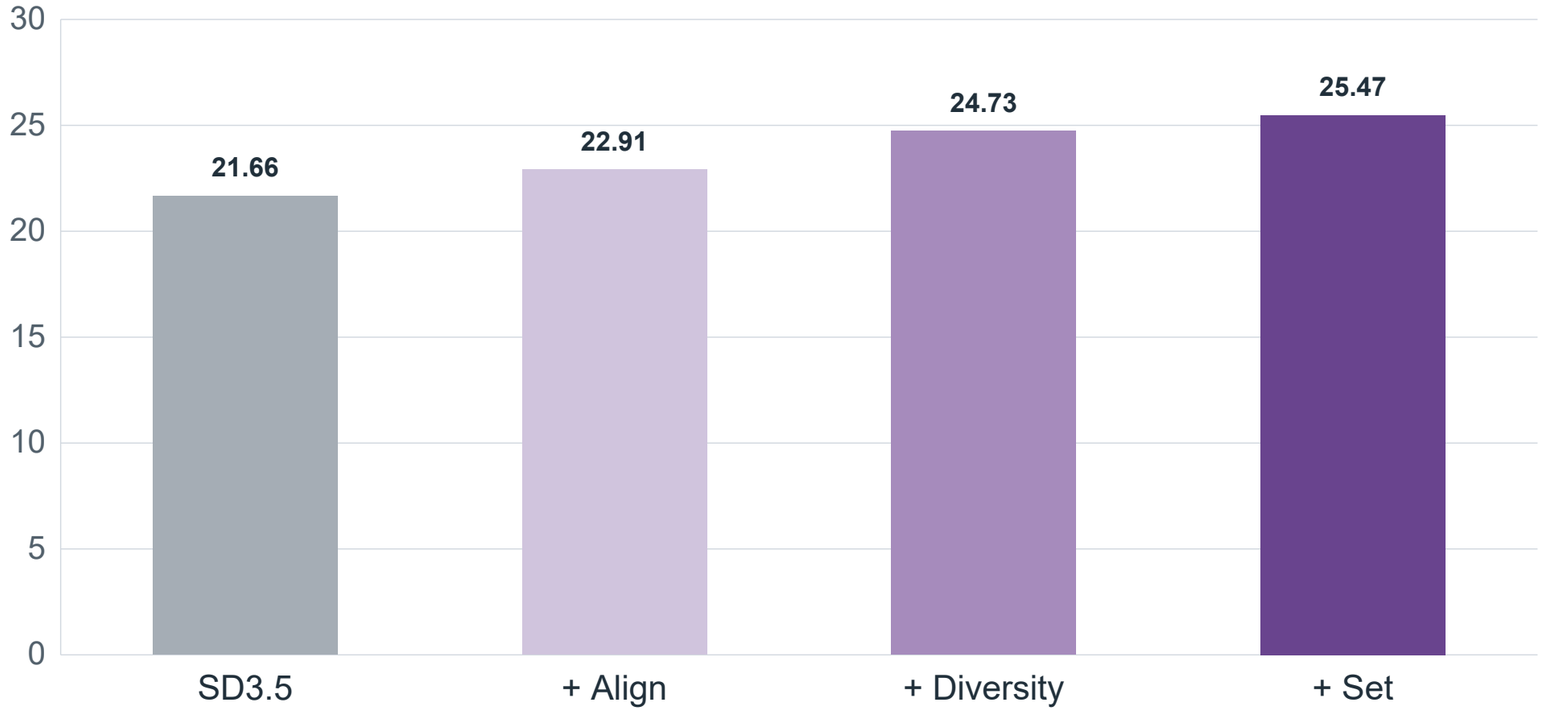
points over SD3.5 at 0.0625% data

Full-data training reaches 61.04 mIoU. A substantial performance gap remains.

The same data budget and downstream protocol apply to each comparison.

Ablation: both reward levels improve training utility

ADE20K at 1% data. SegFormer-B5. Rewards are added cumulatively.



Sample rewards

21.66 to 24.73 mIoU with alignment and intra-class diversity.

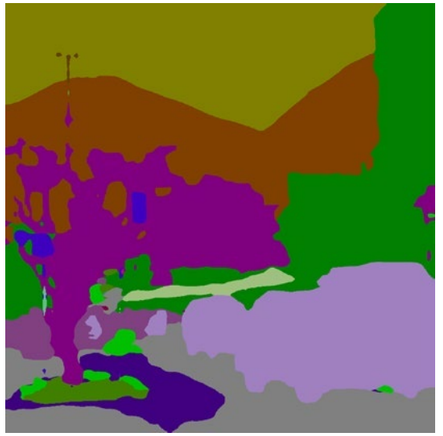
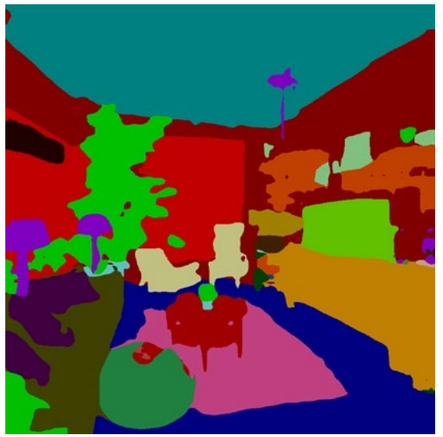
+0.74

additional mIoU points from set-level feedback

The full objective combines informative images with coverage across the set.

Richer scene content under the same class prompts

Generated images and teacher pseudo-labels. Selected examples from Figure 4.

	SD3.5 image	Pseudo-label	S^2-RL image	Pseudo-label
Street building, tree, road, sidewalk, mountain, plant, car, truck, awning, van				
Living room wall, floor, ceiling, window, cabinet, plant, sofa, shelf, armchair, lamp, cushion, blinds, coffee table, flower, television, pouf, pot				

Qualitative examples illustrate content; downstream mIoU measures training utility.

Conclusions and resources

Sample quality

Align requested categories and diversify local features within each semantic class.

Set coverage

Use prototype memory to reward underrepresented semantic modes.

Practical boundary

Policy adaptation costs about 48 GPU-hours.
Teacher pseudo-labels can contain errors.



Paper



Project



Code

Thank you