Learning to Select, Learning to Judge: Active Preference Alignment for Mars Terrain Segmentation

Supplementary Material

6. Prompt Generator (YOLOv10) Details

6.1. Training and Inference

Data construction from masks. We derive detection labels from the segmentation masks used in the main paper. For each binary mask, we extract all connected components and convert each component into a tight axis-aligned bounding box $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$. Components with small area are removed to suppress speckles ($A < A_{\min} = 64\,\text{px}$). Overlapping boxes are merged using non-maximum suppression (NMS) with an IoU threshold of 0.7. The detector is single-class ("rock"), and we keep the same scene-level splits as in the main paper to avoid leakage.

Model and training recipe. We adopt YOLOv10 with an anchor-free, decoupled head and train for 150 epochs at an input size of 512×512. Optimization uses AdamW (weight decay 5×10^{-4}), cosine learning-rate decay from 1×10^{-3}) after a 5-epoch warmup, and batch size 16 on a single RTX 4090.

6.2. Training Objective and Convergence

Training Objective. The YOLOv10 detection head is optimized with a composite loss over box regression, classification/objectness, and distributional regression:

$$\mathcal{L} = \lambda_{\text{box}}\,\mathcal{L}_{\text{box}} + \lambda_{\text{cls}}\,\mathcal{L}_{\text{cls}} + \lambda_{\text{dfl}}\,\mathcal{L}_{\text{dfl}}.$$

Here, $\mathcal{L}_{\text{box}}$ is an IoU-family regression loss (implemented as CIoU), $\mathcal{L}_{\text{cls}}$ is a BCE-based classification/objectness loss with focal weighting to mitigate foreground–background imbalance, and $\mathcal{L}_{\text{dfl}}$ is the Distribution Focal Loss (DFL) for boundary distribution regression. We train for 150 epochs; heavy augmentations are disabled in the last 10 epochs to stabilize convergence. An exponential moving average (EMA) of the weights is maintained, and the final checkpoint is selected by the best validation mAP@50:95.

Convergence behavior. As shown in Fig. 5, training/validation losses ($box/cls/dfl$) decrease rapidly in early epochs and then flatten, without train–val divergence or oscillation. Precision and recall rise quickly and stabilize, while mAP@50 and mAP@50:95 steadily increase before reaching a plateau. Overall, the curves indicate stable optimization and good generalization, yielding proposal sets with high coverage and low redundancy for downstream SAM prompting.

6.3. Detection Quality

Tab. 4 shows that the prompt generator achieves high recall with acceptable precision. YOLOv10 reaches 81.2% recall / 78.4% precision with $AP@[50:95] = 55.1$, indicating reliable coverage and adequate localization for SAM prompts. Inference is fast (5.6 ms per image), so proposal generation is not the bottleneck. Fig. 6 illustrates YOLOv10 detections that are later used as SAM prompts. We observe:

- High recall and multi-scale coverage. Salient rock outcrops and small pebbles are both captured. Scenes with sparse geology still receive adequate proposals.
- Geometric robustness. Despite strong perspective and tilt, axis-aligned boxes cover most objects well enough for prompt encoding. Perfect box tightness is not required for SAM prompting.

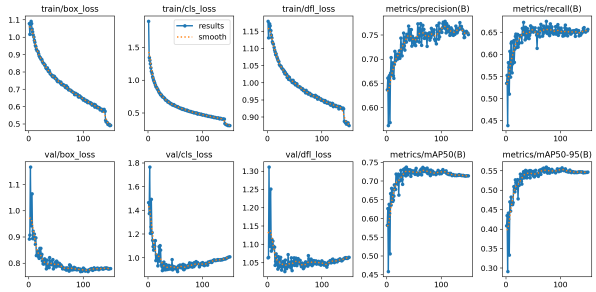

Figure 5. YOLOv10 training dynamics. Top: training losses (box/cls/dfl) and precision/recall; Bottom: validation losses and mAP@50 / mAP@50–95.

Table 4. Detection quality of the YOLOv10 prompt generator.

	AP@50	AP@75	AP@[50:95]	Prec.@0.25	Rec.@0.25	Latency (ms)
YOLOv10	72.5	58.3	55.1	78.4	81.2	5.6

7. OOD Generalization

Protocol. We evaluate out-of-distribution (OOD) transfer with no target labels or tuning. Models are trained on a source domain and evaluated on disjoint target domains. We report per-image mIoU averaged over three seeds and compute 95% confidence intervals via scene-level bootstrap (1,000 resamples). We do not include ZhuRong$\rightarrow(\cdot)$ as a source due to its very small and homogeneous distribution (< 300 usable images, mostly plains), which yields unstable OOD estimates.

Findings. As shown in Tab. 5, APAS consistently achieves the highest OOD performance across both transfer directions.

- Curiosity$\rightarrow$SimMars6k: APAS reaches 74.0 ± 0.5 mIoU, surpassing the best baseline (SAM-YOLO, 72.5 ± 0.7)

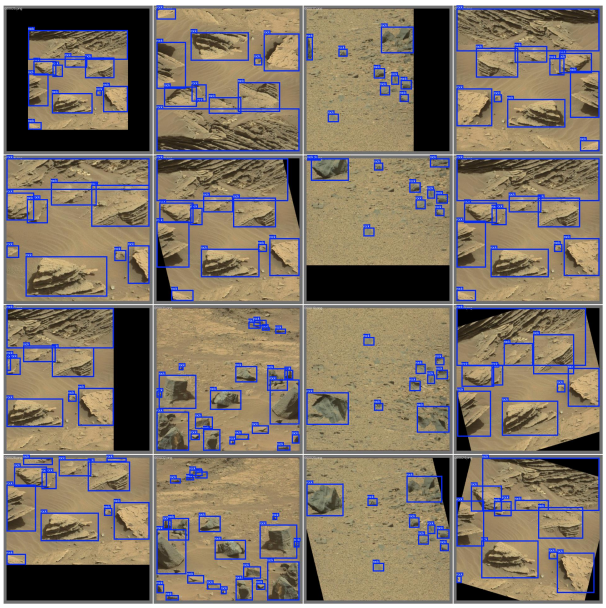

Figure 6. YOLOv10 detections visualization on Curiosity scenes (blue boxes).

Table 5. OOD generalization performance (mIoU ↑; mean ± 95% CI). 'Best Specialist" denotes the strongest task-specific segmenter among our specialist baselines.

Train → Test	SFT-Only	SAM-YOLO	Best Specialist	**APAS (ours)**
Curiosity → SimMars6k	71.2 ± 0.6	72.5 ± 0.7	67.8 ± 0.6	74.0 ± 0.5
SimMars6k → Curiosity	68.1 ± 0.5	70.3 ± 0.8	64.4 ± 0.7	74.2 ± 0.3

by $+1.5$ mIoU. The confidence intervals do not overlap, indicating statistically reliable gains. This direction is relatively easy (real→synthetic), and the smaller gap reflects that the synthetic domain is less diverse.

- SimMars6k→Curiosity: APAS achieves 74.2 ± 0.3 , clearly outperforming SAM-YOLO (70.3 ± 0.8) and SFT-Only (68.1 ± 0.5) by large margins of $+3.9$ and $+6.1$ mIoU respectively. This sim→real setting is the most challenging, and APAS's improvement highlights its robustness to domain shift and noisy supervision.

8. Case Study Analysis

The images presented in Fig. 7 showcase two distinct cases of terrain segmentation, illustrating the performance differences between various models. The red boxes highlight the areas where APAS demonstrates its strength in handling segmentation tasks that other models struggle with.

Case I: Simple Rock Segmentation. In Case I, the task involves segmenting a relatively small rock formation with clear boundaries. While the Light4Mars, S5Mars, and nnWnet models fail to capture the full rock structure, producing under-segmented or fragmented outputs, APAS (Our Method) succeeds in segmenting the entire object accurately, retaining fine details that are lost by other models. The red box highlights an area where APAS outperforms others, capturing the subtle nuances of the rock boundaries that other models miss, particularly SwinUnet, which has difficulty with smaller structures.

Case II: Complex Terrain with Multiple Objects. In Case II, the terrain is more complex, containing numerous rocks of varying sizes scattered across the scene. However, the GT only labels a single, large rock. This creates an ambiguity in segmentation, as the models need to decide whether to treat these multiple rocks as separate objects or to group them into a single entity.

In this case, the SAM series models (SAM1 and SAM2) along with our APAS method tend to group the scattered rocks into a single cohesive segment, reflecting a more unified understanding of the scene. SAM1, SAM2, and APAS all lean towards treating the multiple objects as part of one larger segment, which aligns with the inherent uncertainty in the GT label, where a more comprehensive approach to grouping the rocks is needed. In contrast, the static expert models strictly adhere to the GT, resulting in under-segmentation of the multiple rocks and leaving some rocks unsegmented or only partially segmented. This demonstrates how APAS and SAM series models can adapt to such ambiguous cases by producing more cohesive and accurate segmentations despite the inherent limitations of the GT.

The red box highlights the region where APAS performs better by unifying smaller, fragmented segments into a coherent whole, while static models fail to capture the full complexity of the scene.

9. Hyperparameter Analysis

Scope. We analyze the sensitivity of APAS to main hyperparameters, focusing on (1) the uncertainty threshold ζ in APS and (2) the preference strength β in DPO optimization.

Effect of uncertainty threshold ζ. As shown in Tab. 6, the uncertainty threshold ζ plays a pivotal role in controlling the balance between sample diversity and signal strength.

Impact of Smaller ζ Values: When ζ is set too low (*e.g.*, $\zeta = 0.05$), the model retains a larger number of samples. While this results in more data being processed, a significant portion of these samples comes from low-uncertainty, low-gradient regions, which contribute less to the learning process. As a result, despite the larger sample size, the improvement in performance is marginal compared to higher values of ζ. This suggests diminishing returns, where adding more samples without sufficient uncertainty does not effectively contribute to training, and the computational cost increases.

Effect of Moderate ζ Values: At $\zeta = 0.08$, performance peaks with 299 high-uncertainty samples, yielding a Dice score of 88.7 and IoU of 79.5, which is the best perfor-

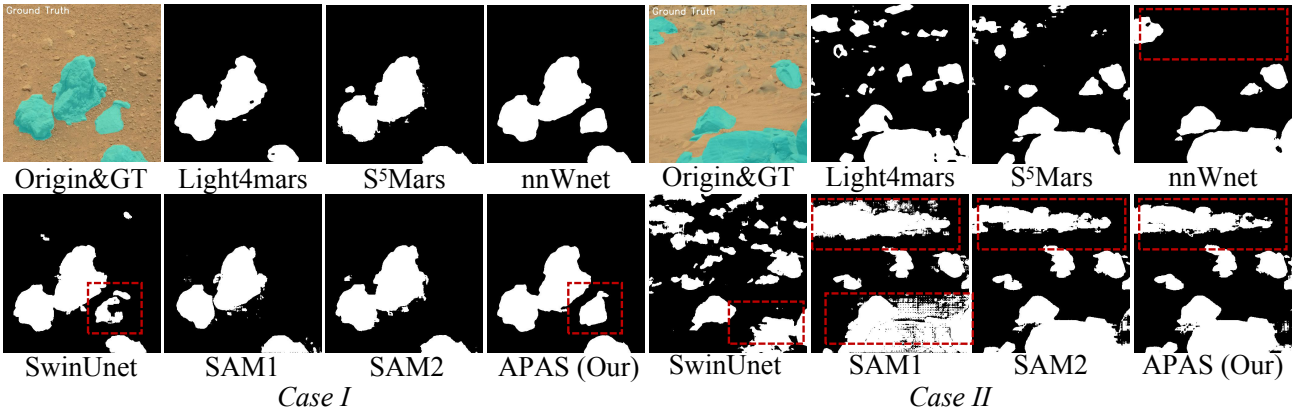

Figure 7. Case Study. Case I (left): Simple rock segmentation. Case II (right): Complex terrain with multiple rocks.

Table 6. Impact of uncertainty threshold ζ on sample count and performance.

ζ	**Samples Retained (Pairs)**	**Dice ↑**	**IoU ↑**
0.05	1046	88.9	79.3
0.06	706	88.8	79.3
0.07	466	88.6	79.2
0.08	299	**88.7**	**79.5**
0.09	197	88.3	79.2
0.10	138	87.8	78.1
Random samples	299	87.1	77.8

mance achieved in this experiment. This threshold strikes a practical balance, selecting a manageable number of high-quality samples that help the model focus on the most uncertain and informative regions. The model benefits from a combination of diversity and concentrated signal, which leads to improved learning efficiency and performance.

Impact of Larger ζ Values: As the threshold is raised (*e.g.*, $\zeta = 0.10$), the number of retained samples decreases drastically (only 138 pairs), which results in underfitting. While these selected samples are highly uncertain, the reduced dataset leads to poorer model performance. This shows that overly stringent thresholds can prune too much data, reducing the diversity of training samples and ultimately hindering the model's generalization ability.

Comparison with Random Sampling: In the final row of Tab. 6, we also observe the performance of randomly selected samples, where 299 samples were chosen without regard to their uncertainty. This baseline method results in lower performance compared to the uncertainty-guided selection with $\zeta = 0.08$, which reinforces that the selection process, based on uncertainty, is crucial in identifying and prioritizing informative samples. This also confirms that the performance gains are not simply due to using more data, but rather due to the strategic selection of high-value samples.

Table 7. Sensitivity of APAS to the DPO preference strength β. Results on Curiosity→SimMars6k (mIoU ↑).

β	0.00	0.05	0.10	0.15	0.25	0.50
mIoU	71.2	73.5	**74.0**	73.7	72.8	70.5

Table 8. Intrinsic evaluation of the **QualityNet** module on their ability to correctly classify the rank of $K = 3$ candidates.

Judge Training Objective	**Top-1 Acc.**	**Pairwise Acc.**
Rating	68.4%	80.1%
Ranking	**81.2%**	**92.5%**

Effect of preference strength β. The coefficient β regulates the trade-off between staying close to the SFT reference policy and optimizing toward preference-aligned updates. As reported in Table 7, moderate values ($\beta \in [0.05, 0.15]$) yield the best balance between stability and alignment strength. Excessively large β values destabilize early training, while β=0 degenerates to pure SFT. Our default β=0.1 provides near-optimal mIoU and convergence speed.

10. QualityNet Evaluation.

To further examine why the Ranking objective is superior, we evaluate the intrinsic accuracy of QualityNet on an independent validation set (Tab. 8). The Ranking-based Judge attains significantly higher *Top-1* accuracy (81.2% *vs.* 68.4%) and Pairwise accuracy (92.5% *vs.* 80.1%), indicating it learns a more reliable representation of the relative quality ordering among candidate masks. These results explain its stronger downstream impact in APAS: by providing consistent and noise-robust preference supervision, the Ranking Judge drives more stable preference optimization and higher final segmentation accuracy.