Supplementary Material

Table 5. IoU comparison across entropy thresholds ϵ on 1-point supervision.

Dataset	$\epsilon = 0.1$	$\epsilon = 0.2$	$\epsilon = 0.3$	$\epsilon = 0.6$
NWPU	69.2	**70.25**	66.87	66.05
HRSID 2	58.02	**58.40**	56.75	55.15

7. The Effect of Entropy Threshold ϵ

The entropy threshold ϵ is a critical hyperparameter in our pseudo-label refinement stage, determining the confidence level required for a pixel to be classified as foreground or background during the iterative R^3 self-prompting procedure. As reported in Table 5, we observe a consistent performance peak at $\epsilon = 0.2$ for both NWPU and HRSID datasets, achieving IoU of **70.25** and **58.40**, respectively. While a lower threshold $\epsilon = 0.1$ tends to permit high-entropy regions that introduce noise, increasing the threshold beyond 0.2 restricts and compresses the predictions too harshly, leading to the generation of inaccurate pseudo-labels. This suggests that our pseudo-label generation benefits from conservative entropy filtering; because remote sensing datasets exhibit small objects and cluttered backgrounds where high-entropy pixels are typically ambiguous; enforcing stricter confidence boundaries effectively prevents the propagation of noise during the self-training process.

8. SSA Algorithm

In Algo 1, we provided pseudo code for Soft Semantic Alignment (SSA). The method aligns the embedding space of weak and strong augmentations through normalized cosine similarity, reinforced by a FIFO memory queue. The queue stores a running history of feature representations, which encourages temporal consistency and mitigates noise from single-sample predictions. This mechanism plays a key role in stabilizing the pseudo-label refinement loop in R^3, particularly in the later training stage when the model can skew towards degradation.

Algorithm 1 Soft Semantic Alignment (SSA)

Require: Weak embeddings E^w, strong embeddings E^s, bboxes $\mathcal{B}$, queues $\mathcal{Q}_s$, $\mathcal{Q}_h$ of max size q

$\mathcal{S} \leftarrow \emptyset, \mathcal{H} \leftarrow \emptyset$
$\mathcal{L}_{SSAL} \leftarrow 0$
for $b \in \mathcal{B}$ **do**
  $s_i \leftarrow \text{GetBBoxFeature}(E^w, b)$
  $h_i \leftarrow \text{GetBBoxFeature}(E^s, b)$
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{s_i\}$
  $\mathcal{H} \leftarrow \mathcal{H} \cup \{h_i\}$
end for
$\hat{\mathcal{S}} \leftarrow \text{Normalize}(\mathcal{S})$
$\hat{\mathcal{H}} \leftarrow \text{Normalize}(\mathcal{H})$
if $|\mathcal{Q}_s| = q$ **then**
  $\mathcal{L}_{SSAL} \leftarrow \frac{1}{q} \sum_{i=1}^{q} \left(1 - \hat{s}_i^\top \hat{h}_i\right)$
else
  $\mathcal{L}_{SSAL} \leftarrow 0$
end if
Push $\hat{s}_i$ into $\mathcal{Q}_s$ (FIFO)
Push $\hat{h}_i$ into $\mathcal{Q}_h$ (FIFO)
return $\mathcal{L}_{SSAL}$

9. Qualitative Results of Pseudo Labels

Figures 7 and 8 provide extended qualitative comparisons of the pseudo labels generated by the R^3 method on NWPU and HRSID datasets. Each row visualizes the input image with points, ground truth, SAM prediction, and the refined pseudo-labels.The results highlight several key observations:

- **Fine structures are preserved** more accurately compared to baseline SAM predictions.
- **Noisy backgrounds are suppressed**, especially in cluttered remote sensing scenes.
- **Point-based supervision becomes increasingly effective** after multiple R^3 refinement rounds.
- **Small objects such as ships or vehicles are recovered** even when initially missed by SAM.

These results demonstrate the advantage of the self-prompting mechanism and its robustness across varying scene complexities.

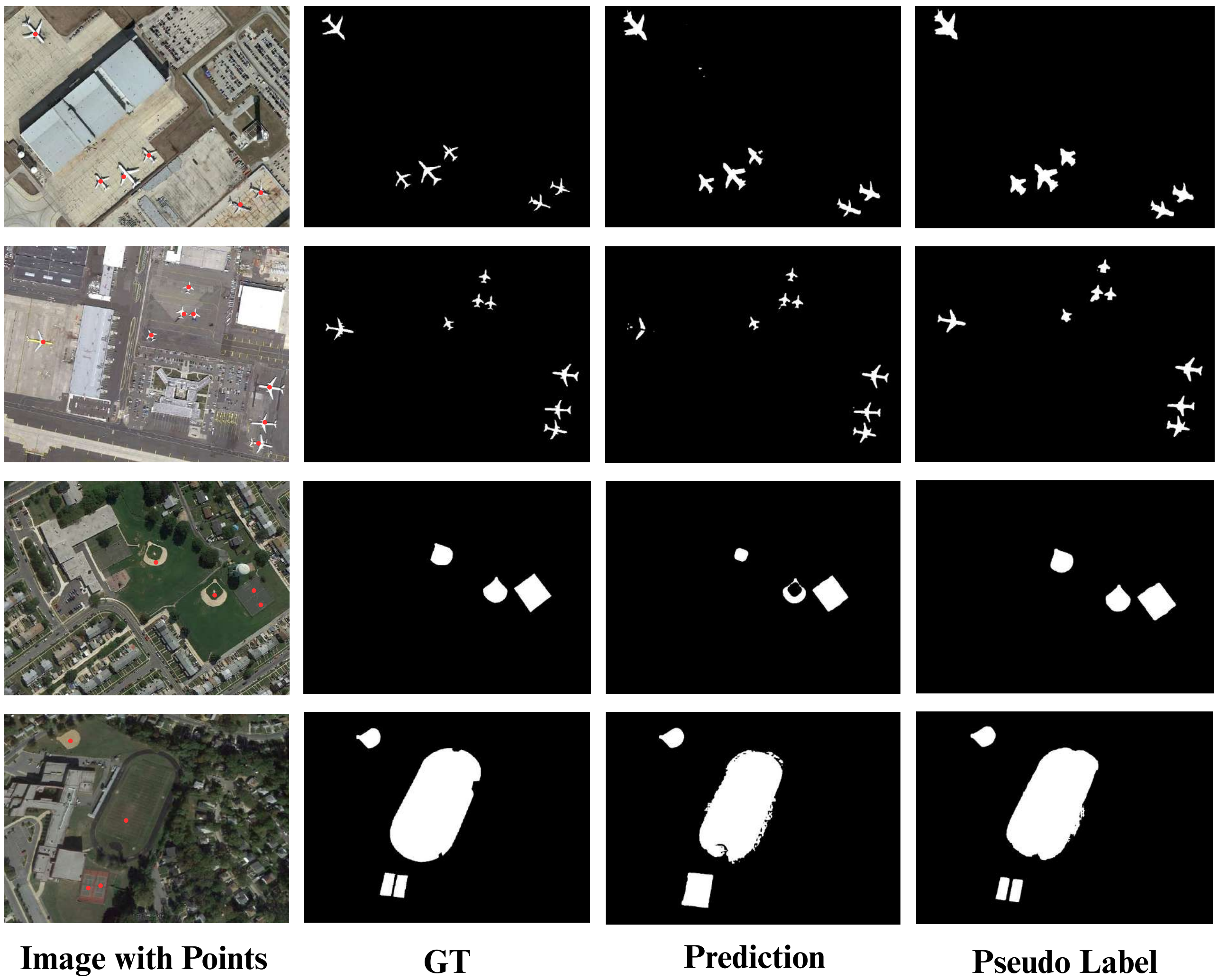

Figure 7. Qualitative results on the NWPU dataset for pseudo labels and corresponding predictions. The refined R^3 pseudo labels provide clearer object boundaries and improved foreground consistency.

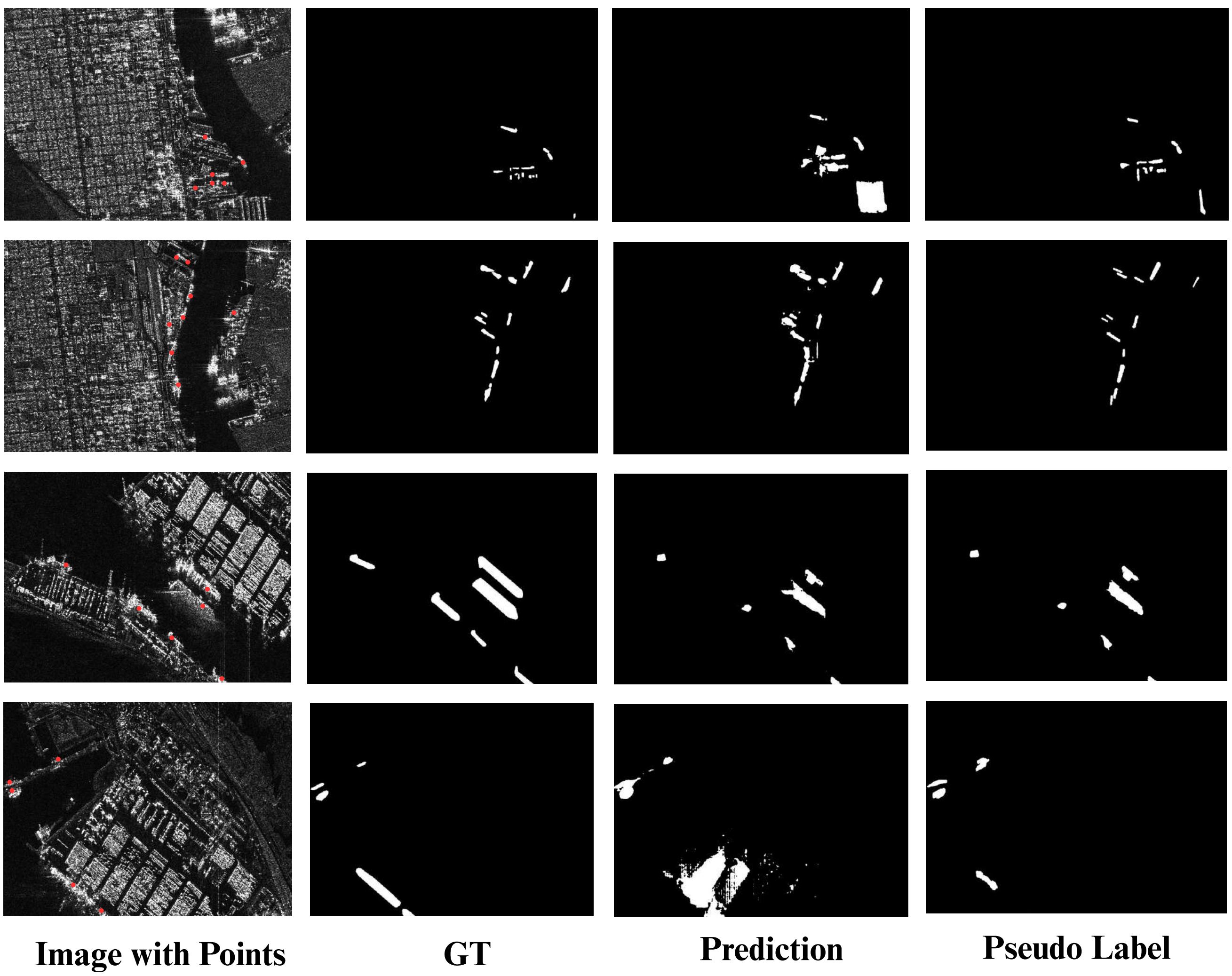

Figure 8. Qualitative results on the HRSID dataset showing improved pseudo labels and predictions from ReSAM. The refinement particularly enhances ship boundary localization.