Monte Carlo Linear Clustering with Single-Point Supervision is Enough for Infrared Small Target Detection (Supplementary Material)

Boyang Li[1], Yingqian Wang[1], Longguang Wang[2], Fei Zhang[3],
Ting Liu[1], Zaiping Lin[1]✉, Wei An[1], Yulan Guo[1]
[1]National University of Defense Technology, [2]Aviation University of Air Force,
[3]Shanghai Jiao Tong University
{liboyang20, linzaiping}@nudt.edu.cn

Section I introduces the details of the annotation cost experiment. Section II discusses the results of the color distance change experiment. Section III provides more target probability maps (TPMs) on different SIRST datasets. Section IV presents additional ablation studies on the other three SIRST datasets. More quantitative results on different SIRST datasets are shown in Section V. Moreover, we developed an offline webpage to summarize the pipeline, methods, and visual results of our paper. Readers can refer to the attached files for more details.

I. Details of Annotation Cost Experiment

In Section 3.1 of the main body of our paper, we report the annotation cost of four common weakly-supervised and one fully-supervised approaches. Here, we describe how we used the labeling software to generate the various kinds of annotations in Figure II. The detailed annotation cost statistics are shown in Figure I.

As shown in Figure II, we use Adobe Photoshop 2019 as labeling software. Randomly-selected 100 images from the NUAA-SIRST [1] and NUDT-SIRST [2] datasets are used for annotation cost evaluation. Averages annotation cost from 3 trails are reported in Figure I. Given an input image, we first localize the small targets and then zoom in the targets located small region, as shown in Figure II (a)-(c). Then, multiple types of label are placed in the target region by corresponding labelling manners. For single point annotation (shown in Figure II (d)), we place the single-point label on the center of the target. For multiple points annotation (shown in Figure II (e)), the first point is also placed on the center of the target. Then, the remaining 4 points are randomly placed in the four corners of the target. Scribble annotation is a curve that passes through the central region of the target (shown in Figure II (f)). Bounding boxes annotation (shown in Figure II (g)) is a box that fits tightly to the target. Pixel level annotation (shown in Figure II (h)) is achieved by labelling every pixels in the image.

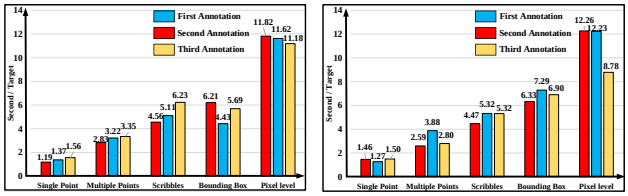


Figure I. Annotation time cost on the (a) NUAA-SIRST and (b) NUDT-SIRST datasets. We take 1.4, 3.1, 5.1, 6.5, and 11.2 seconds per target to generate single point, multiple points, scribbles, bounding boxes, and pixel-level annotations, respectively. Averages annotation cost from 3 trails are reported.

Figure I reports the annotation cost of five kinds of annotations by re-labelling the NUAA-SIRST [1] and NUDT-SIRST [2] datasets. We take 1.4, 3.1, 5.1, 6.5, and 11.2 seconds per target to generate single point, multiple points, scribbles, bounding boxes, and pixel-level annotations, respectively. Single-point supervision can reduce about 87% annotation time as compared to the pixel-level annotation approach.

II. Details of Color Distance Experiment

In Section 3.3 of the main paper, we provide qualitative results of color distance change experiment on Figure 6 (b). Here, we provide more details for this experiment.

As described in the main body, we want to verify the opinion that random noise helps to increase the distance between background and target region. In this way, the misclustered target regions can be pushed away from the false clustering center and thus return back to the true clustering center. To support this opinion, we first calculated three kinds of color distance (i.e., maximum color distance with false clustering center, minimum color distance with false clustering center, color distance with true clustering center) and drew three kinds of curves (i.e., upper boundary of $\Delta(\mathcal{D}_c(\mathcal{C}^F, M_{pred}))$, lower boundary of $\Delta(\mathcal{D}_c(\mathcal{C}^F, M_{pred}))$,

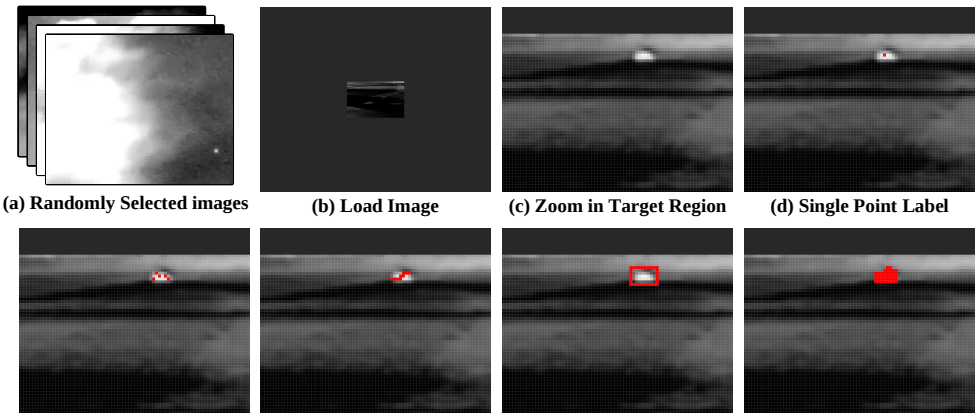


Figure II. Detailed process of achieving four common weakly-supervised and one fully-supervised labels. Adobe Photoshop 2019 is used as labelling software.

and $\Delta(\mathcal{D}_c(\mathcal{C}^T, M_{pred}))$) with the increasing of noise intensity in Figure 6 (b).

We can observe from Figure 6 (b) that, with the increase of noise intensity, the color distance between edge areas and true clustering center (i.e., $\Delta(\mathcal{D}_c(\mathcal{C}^T, M_{pred}))$) is always smaller than the lower boundary of $\Delta(\mathcal{D}_c(\mathcal{C}^F, M_{pred}))$. This demonstrates that random noise with proper intensity has higher probability of pushing the target away from the false clustering center and helping them return back to the true clustering center. Note that, the above curves are drawn by averaging the results from 10 trails. Although random noise cannot always introduce true guidance for misclustered regions in each experiment, the average result of multiple experiments presents a robust trend that random noise can effective guide the misclustered regions to return back to the true clustering center.

III. MCLC Process on Different Datasets

Here, we introduce more visual TPMs during Monte Carlo linear clustering (MCLC) process in Figure III. Clustering results at iteration 1, 2, 20, 100 of MCLC process and the denseCRF refined pseudo masks on the NUAA-SIRST [1], IRSTD-1k [4], NUDT-SIRST [2], and NUDT-SIRST-sea [3] datasets are shown in Figure III. Although easily producing inaccurate results at the beginning of clustering (e.g., iteration number less than 20), MCLC can gradually recover a reliable clustering result.

IV. Ablation Study on Different Datasets

In Section 4.3 of the main body of our paper, some ablation studies (e.g., Type and Intensity of Noise, Number of Clustering Center, Influence of Labeling Position Deviation) were conducted on the NUAA-SIRST dataset [1] only. Here, we present the experimental results on the other datasets (IRSTD-1k [4], NUDT-SIRST [2], and NUDT-SIRST-sea [3]) to verify the generalization our method.

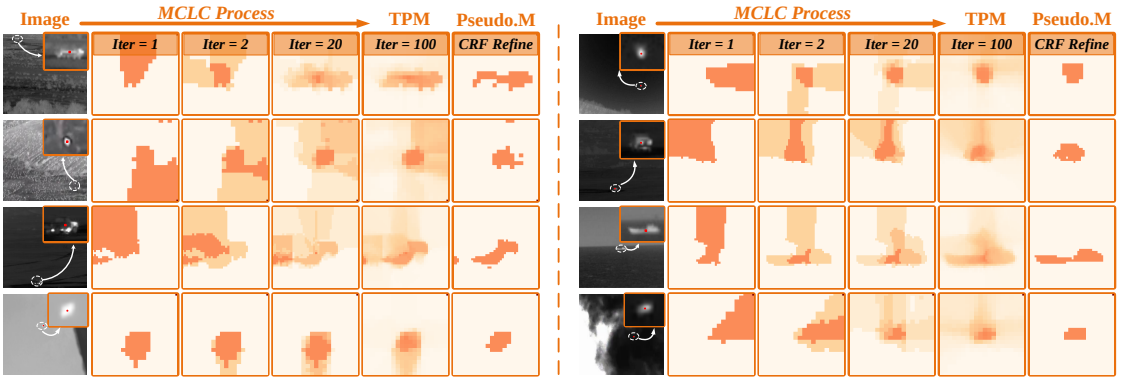
Figure IV (a1)-(d1) show the change trend of IoU with respect to different noise intensity under three types of common noise (i.e., salt, pepper, and Gaussian) on four datasets. With the increase of noise intensity, the IoU of MCLC with denseCRF under salt and noise increases rapidly at the beginning. After that, excessive intensity value reduces the saliency of target region and thus results in the decrease of IoU. Moreover, Gaussian noise causes huge performance decrease under any intensity. Similar change trend can also be found on the other datasets (IRSTD-1k, NUDT-SIRST, and NUDT-SIRST-sea). These results disclose the fact that proper type and intensity of noise are essential to MCLC on all datasets.

Figure IV (a2)-(d2) report the IoU of the results generated by MCLC with denseCRF, MCLC with fixed threshold, and LCA under different number of clustering center. It can be observed that the IoU firstly shows a rapid increasing trend with the increase number of clustering center. Afterwards, the quality of TPM gradually decreases when the number of clustering center further increases. Similar change trend can be also found on the other datasets (IRSTD-1k, NUDT-SIRST, and NUDT-SIRST-sea). The above results demonstrate that inappropriate number of clustering center will result in over-small or over-large area of initial search region, and thus introduce negative effect on MCLC for all datasets.

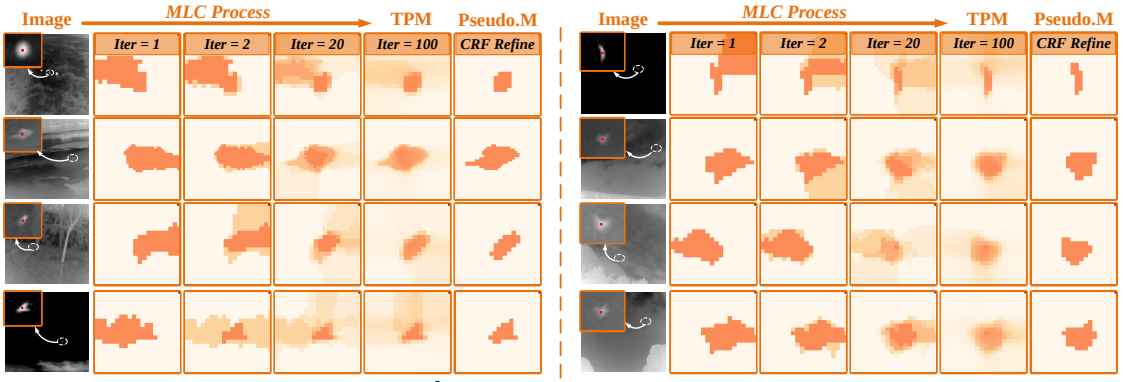
As shown in Figures IV (a3)-(d3), with the increase of label position deviation, IoU value gradually decreases, but still maintains at a high level even with five pixels deviation. Similar change trend can be also found on the other datasets (IRSTD-1k, NUDT-SIRST, and NUDT-SIRST-sea). That is because, our proposed Monte Carlo regularization method enables the model to seek for robustness representation from repetitive Monte Carlo clustering.

V. Quantitative Results on Different Datasets

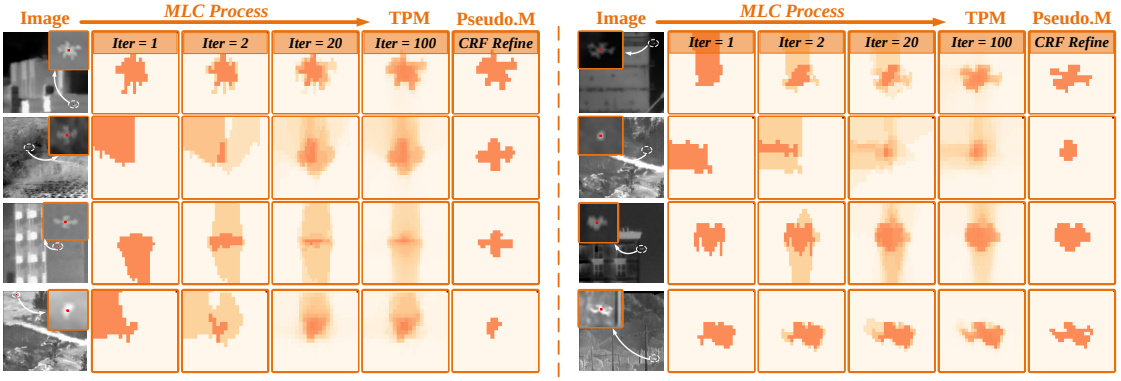
Figures V and VI show the qualitative results of our single-point supervised method and the compared fully-



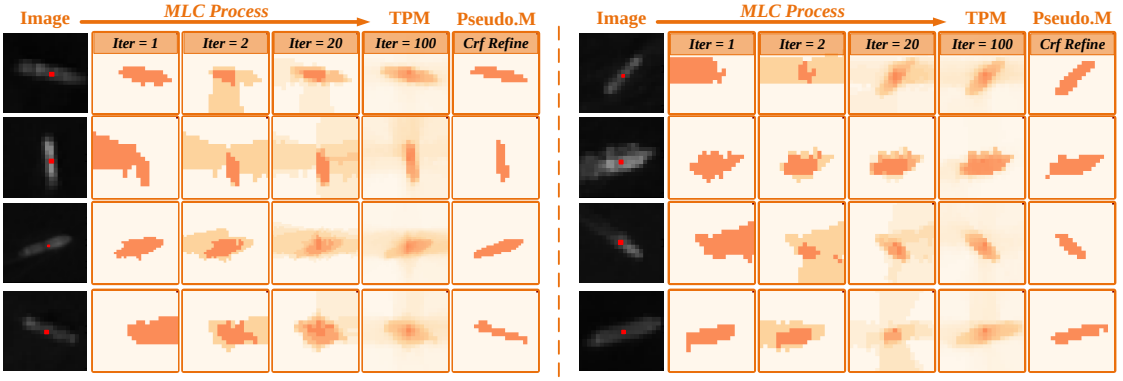
(a) MCLC Process on NUAA-SIRST



(b) MCLC Process on IRSTD-1k

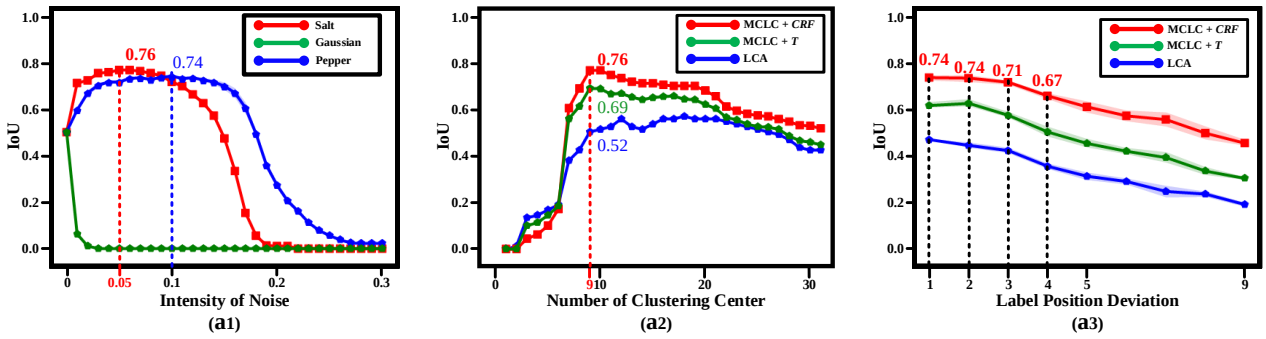


(c) MCLC Process on NUDT-SIRST

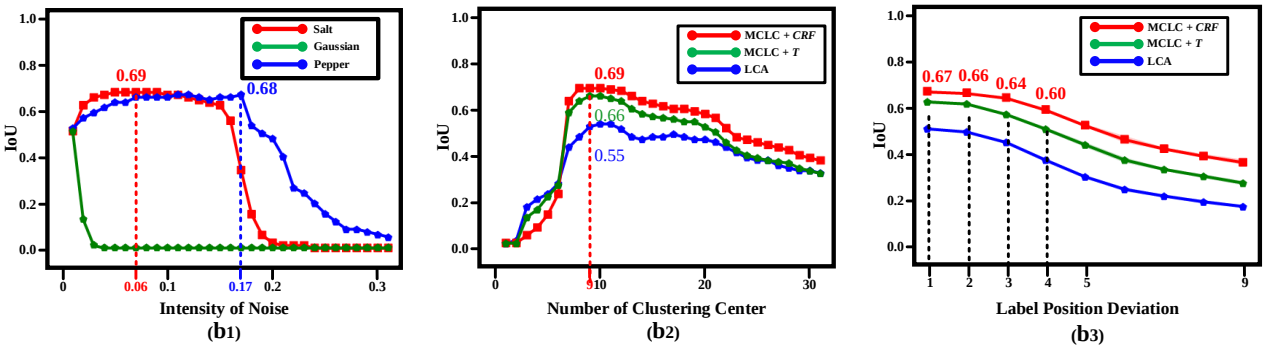


(d) MCLC Process on NUDT-SIRST-sea

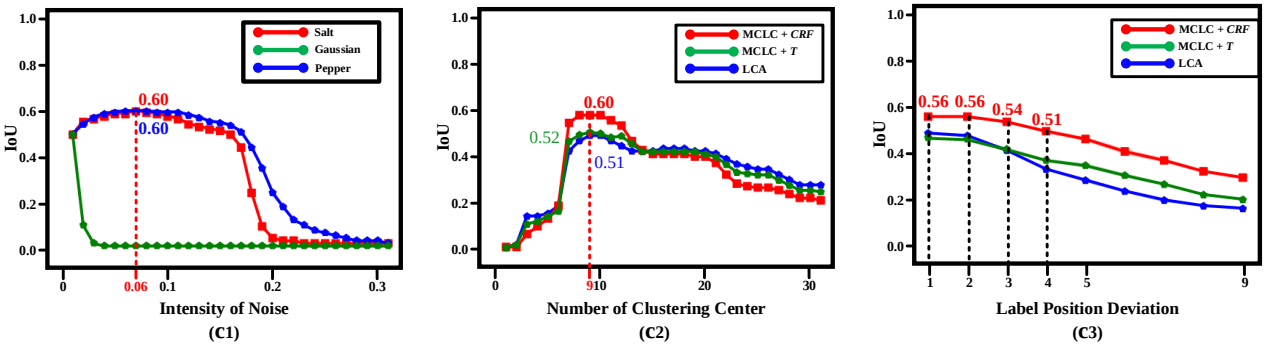
Figure III. Examples of target probability map and the corresponding refined pseudo masks during the MCLC process on four datasets.



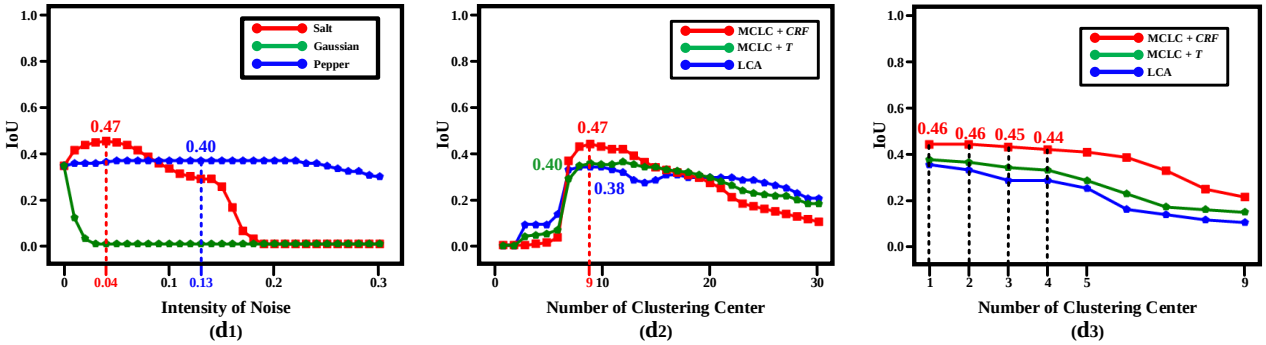
(a) Results on the NUAA-SIRST Dataset



(b) Results on the IRSTD-1k Dataset



(c) Results on the NUDT Dataset



(d) Results on the NUDT-SIRST-sea Dataset

Figure IV. *IoU* scores achieved by our method with (a1)-(d1) different types and intensity of additional noise, (a2)-(d2) different number of clustering center, and (a3)-(d3) different label position deviation on four SIRST datasets.

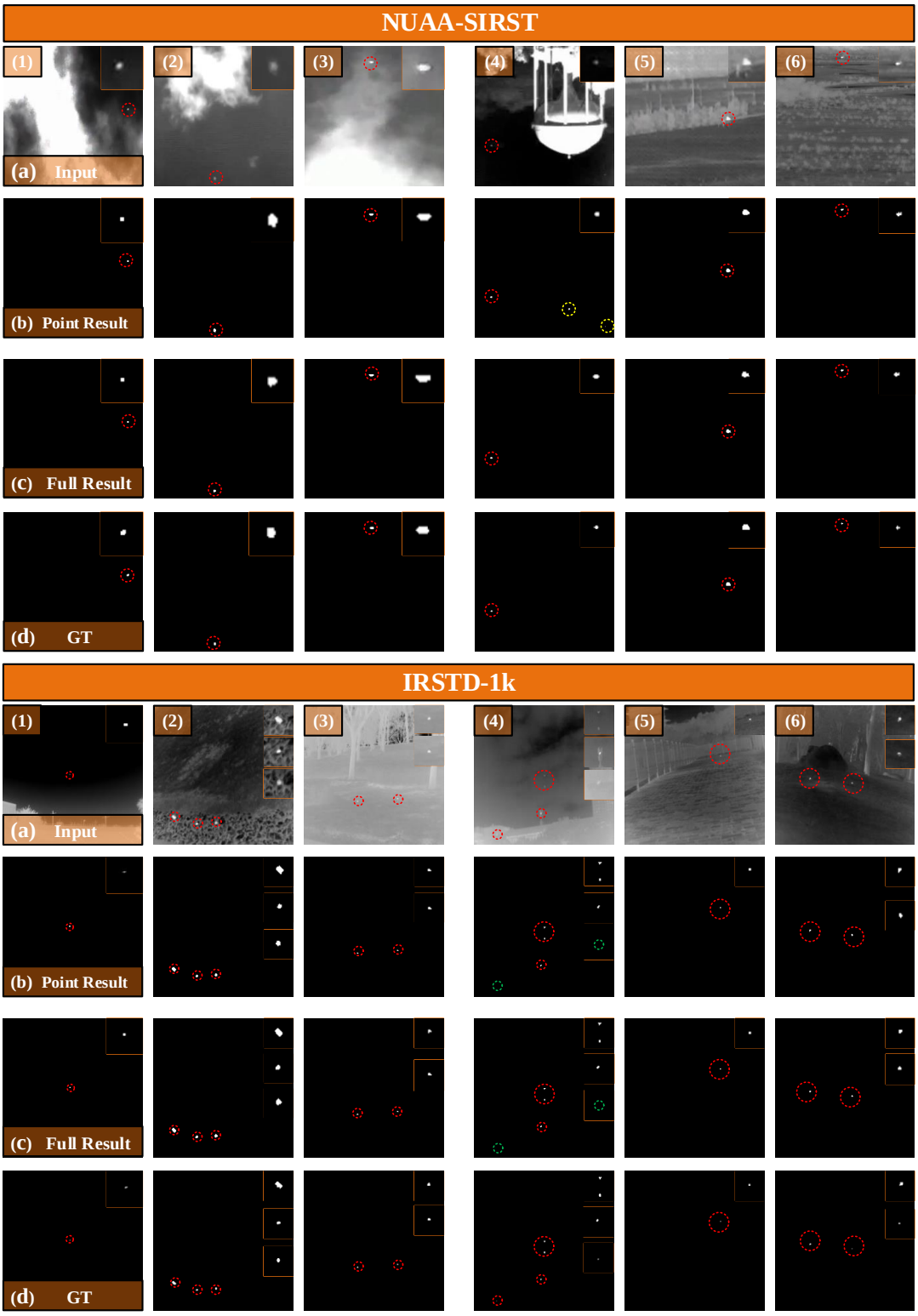


Figure V. Qualitative results achieved by different SIRST detection methods on the NUAA-SIRST and IRSTD-1k datasets under (b) point-level supervision, (c) pixel-level supervision. For better visualization, the target area is enlarged in the right-top corner. The correctly detected target, false alarm, and miss detection areas are highlighted by red, yellow, and green dotted circles, respectively. Our MCLC enables the network to achieve comparable performance to the fully-supervised results with only single-point annotation.

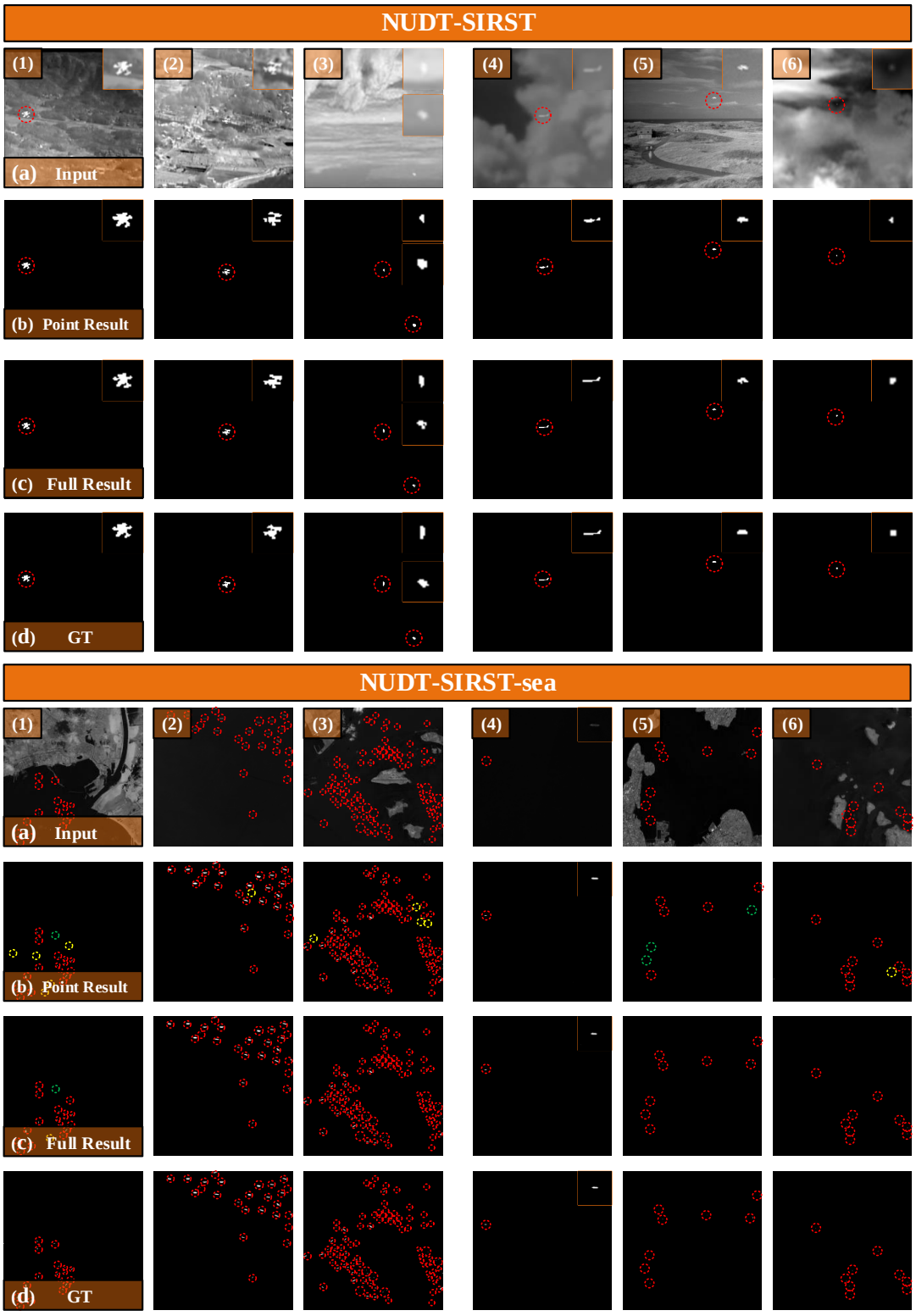


Figure VI. Qualitative results achieved by different SIRST detection methods on the NUDT-SIRST and NUDT-SIRST-sea datasets under (b) point-level supervision, (c) pixel-level supervision. For better visualization, the target area is enlarged in the right-top corner. The correctly detected target, false alarm, and miss detection areas are highlighted by red, yellow, and green dotted circles, respectively. Our MCLC enables the network to achieve comparable performance to the fully-supervised results with only single-point annotation.

supervised methods on different SIRST datasets [1–4].

References

[1] Yimian Dai, Yiquan Wu, Fei Zhou, and Kobus Barnard. Asymmetric contextual modulation for infrared small target detection. In *WACV*, pages 950–959, 2021. 1, 2, 7

[2] Boyang Li, Chao Xiao, Longguang Wang, Yingqian Wang, Zaiping Lin, Miao Li, Wei An, and Yulan Guo. Dense nested attention network for infrared small target detection. *IEEE TIP*, 2022. 1, 2, 7

[3] Tianhao Wu, Boyang Li, Yihang Luo, Yingqian Wang, Chao Xiao, Ting Liu, Jungang Yang, Wei An, and Yulan Guo. Mtu-net: Multi-level transunet for space-based infrared tiny ship detection. *arXiv preprint arXiv:2209.13756*, 2022. 2, 7

[4] Mingjin Zhang, Rui Zhang, Yuxiang Yang, Haichen Bai, Jing Zhang, and Jie Guo. Isnet: Shape matters for infrared small target detection. In *CVPR*, pages 877–886, 2022. 2, 7