From Easy to Hard: Progressive Active Learning Framework for Infrared Small Target Detection with Single Point Supervision

Supplementary Material

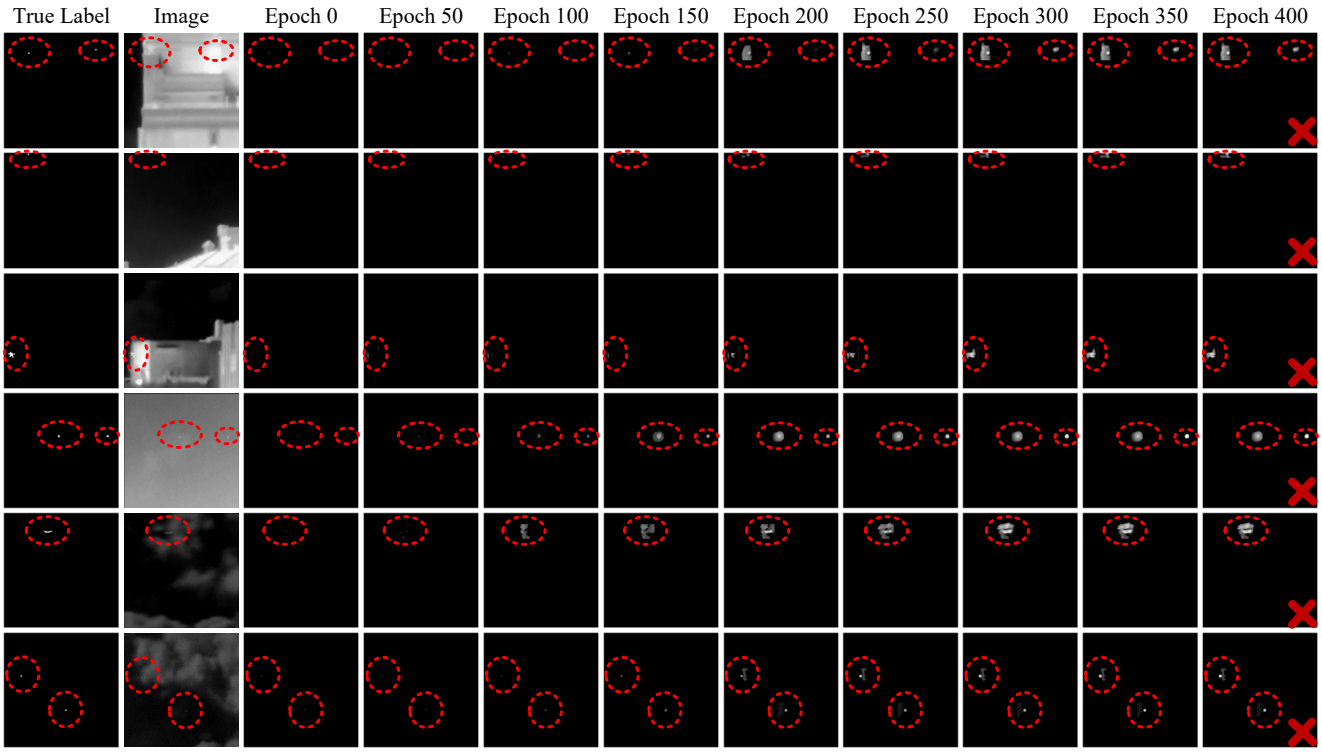

Figure 1. Label evolution results of MSDA-Net equipped with the LESPS framework on the NUDT-SIRST dataset. As the number of epochs increases, the labelled area expands excessively and does not shrink, which will affect the final detection performance.

In this supplementary material, we offer extra details and additional results to complement the main paper. In Sec. A, we provide a presentation and analysis of the risk of excessive label evolution in the LESPS framework. In Sec. B, we provide a detailed introduction to the used edge-enhanced difficulty-mining (EEDM) loss. In Sec. C, we provide a more detailed explanation on why "from easy to hard" fits this task and more visualizations. In Sec. D, we provide a detailed performance comparison with other methods (MCLC [1], LELCM [2]). In Sec. E, we provide more ablation experiments to fully explore the performance of our proposed Progressive Active Learning (PAL) framework. In Sec. F, we provide more quantitative comparative experiments on different datasets. In Sec. G, we provide more qualitative results to further verify the superiority of the proposed PAL framework.

A. Excessive Label Evolution in the LESPS

In exploring the LESPS framework [3], we find that it has the risk of excessive label evolution. From Fig. 1, when using the LESPS framework, if the pseudo-label has an overly large annotation of the target area during evolution, the area will not shrink, but will either remain the same or expand further. The reason is that the label evolution rule does not consider the shrinkage problem of the target annotation in the pseudo-label after it is too large. The LESPS framework is designed to generate reasonable candidate regions using an adaptive threshold rule, which effectively avoids cumulative errors. However, it can only prevent the continued large-scale expansion and cannot prevent the initial occurrence of errors. At the same time, it ignores the fact that the calculation of the adaptive threshold is based on the target area in the pseudo-label rather than the prediction result of the current iteration. Therefore, when the target annotation of the pseudo-label is small and the annotation of the current prediction result is too large, there is a risk of over-expansion in the target annotation of the updated pseudo-label. Since there is no design for shrinking the annotation area in its pseudo-label update strategy, an overly expanded annotation area cannot be shrunk. Therefore, to

Table 1. Performance comparison of the PAL and MCLC on the SIRST3 dataset with coarse point.

Net	Method	SIRST3-Test				Net	Method	SIRST3-Test			
		IoU	*nIoU*	P_d	F_a			*IoU*	*nIoU*	P_d	F_a
ACM	MCLC	48.17	49.94	85.45	110.30	DNA	MCLC	54.57	59.94	87.04	102.87
	PAL	**51.51**	**54.07**	**92.89**	**39.18**		PAL	**67.20**	**70.20**	**96.15**	**10.86**
ALC	MCLC	51.05	53.14	82.99	85.10	GGL	MCLC	55.54	61.96	88.24	129.56
	PAL	**57.11**	**60.22**	**93.95**	**37.20**		PAL	**68.52**	**71.69**	**97.14**	**16.69**
MLCL	MCLC	52.26	58.06	89.57	136.18	UIU	MCLC	54.56	62.21	87.97	164.06
	PAL	**64.87**	**69.40**	**94.95**	**24.43**		PAL	**69.05**	**71.53**	**96.81**	**15.45**
ALCL	MCLC	53.82	58.69	86.38	109.44	MSDA	MCLC	54.71	60.89	88.90	132.39
	PAL	**66.29**	**68.18**	**94.75**	**18.79**		PAL	**69.38**	**71.55**	**97.41**	**16.34**

Table 2. Performance comparison of the PAL and MCLC on the SIRST3 dataset with centroid point.

Net	Method	SIRST3-Test				Net	Method	SIRST3-Test			
		IoU	*nIoU*	P_d	F_a			*IoU*	*nIoU*	P_d	F_a
ACM	MCLC	47.87	49.57	85.51	133.21	DNA	MCLC	55.92	62.53	87.97	110.28
	PAL	**51.51**	**53.73**	**92.82**	**35.98**		PAL	**66.97**	**70.63**	**96.28**	**14.66**
ALC	MCLC	49.82	52.85	85.65	105.89	GGL	MCLC	55.82	62.04	87.91	107.26
	PAL	**55.01**	**57.93**	**93.94**	**31.63**		PAL	**67.83**	**70.27**	**95.68**	**16.28**
MLCL	MCLC	53.83	59.31	88.04	99.71	UIU	MCLC	55.97	62.49	87.71	117.33
	PAL	**66.38**	**69.25**	**95.08**	**15.94**		PAL	**69.05**	**70.01**	**95.68**	**21.10**
ALCL	MCLC	54.31	60.35	88.24	125.57	MSDA	MCLC	56.00	62.09	90.90	93.61
	PAL	**65.99**	**70.59**	**95.22**	**20.81**		PAL	**69.21**	**72.40**	**97.01**	**15.70**

Table 3. Performance comparison of the PAL and LELCM on three individual datasets with coarse point.

Net	Method	NUAA-SIRST			NUDT-SIRST			IRSTD-1K		
		IoU	P_d	F_a	*IoU*	P_d	F_a	*IoU*	P_d	F_a
UIU	LELCM	52.11	**91.61**	51.36	50.35	89.27	46.64	50.02	87.33	**19.96**
	PAL	**62.25**	91.25	**34.99**	**74.89**	**98.62**	**7.10**	**61.70**	**92.57**	21.60
DNA	LELCM	53.98	88.39	36.67	56.56	91.96	23.31	58.04	87.78	**24.36**
	PAL	**66.57**	**90.49**	**27.44**	**73.29**	**98.10**	**22.08**	**58.40**	**91.25**	24.52

Table 4. Performance comparison of the PAL and LELCM on three individual datasets with centroid point.

Net	Method	NUAA-SIRST			NUDT-SIRST			IRSTD-1K		
		IoU	P_d	F_a	*IoU*	P_d	F_a	*IoU*	P_d	F_a
UIU	LELCM	53.21	89.24	59.68	54.84	90.27	59.21	50.68	90.47	43.92
	PAL	**66.43**	**96.20**	**14.54**	**74.58**	**98.20**	**4.67**	**61.30**	**91.25**	**36.61**
DNA	LELCM	58.71	**92.26**	38.49	58.30	89.46	22.43	55.23	87.15	22.67
	PAL	**66.16**	91.63	**24.42**	**73.24**	**97.99**	**9.10**	**59.53**	**88.89**	**20.14**

reduce the risk, we introduce a decay factor, which helps achieve a dynamic balance between the expansion and contraction of target annotations.

B. The EEDM Loss

For the SIRST detection task, the lack of intrinsic features makes it difficult to accurately locate the target area [4, 6]. Therefore, we introduce an edge-enhanced difficulty-mining (EEDM) loss [5] to constrain the network optimization. The EEDM loss consists of two parts: edge pixel enhancement and difficult pixel mining. Taking a single image as an example, firstly, we obtain the target edge contour in the binary pseudo-label. Secondly, the binary cross-entropy loss is used to obtain the loss value of each pixel and form a loss matrix. Finally, we weight the edge contours extracted from the labels and apply this weighting matrix to the calculated loss matrix to obtain the loss value of each point after edge weighting. The expressions are as follows:

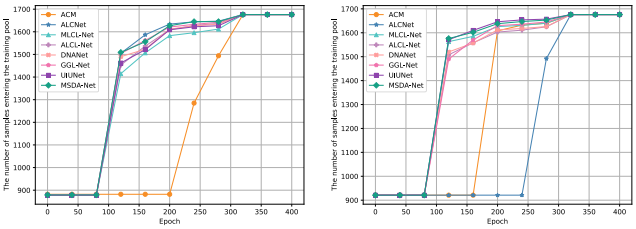

Figure 2. Training pool sample count on the SIRST3 dataset every 40 epochs. Left: coarse point. Right: centroid point.

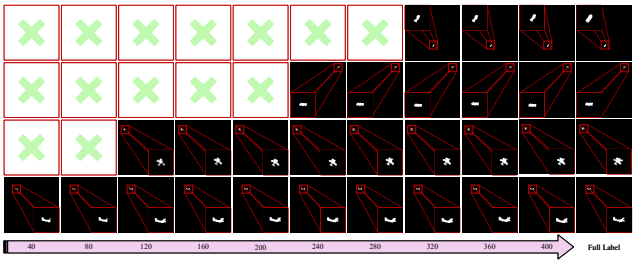

Figure 3. Pseudo-label evolution results of MSDA-Net equipped with the PAL framework on the SIRST3 dataset.

$$L_{ij} = W_{ij}(-(T_{ij}\log(P_{ij}) + (1 - T_{ij})\log(1 - P_{ij}))) \tag{1}$$

$$W_{ij} = \alpha \cdot E_{ij} + (1 - E_{ij}) \tag{2}$$

where E_{ij} denotes the edge extracted from the binary pseudo-label, edge pixels are marked as 1 and non-edge pixels are marked as 0. P_{ij} denotes the prediction result. T denotes the pseudo-label after binarization. α is the edge weighting coefficient, which is set to 4 [5].

Subsequently, difficult pixel mining is performed. Firstly, the loss values of each point are sorted. Secondly, the set of loss values that are greater than or equal to the median is obtained. Finally, the final loss is obtained by calculating the mean loss of difficult pixels. The expressions are as follows:

$$L_{EEDM} = \frac{1}{|S|} \sum_{(i,j \in S)} L_{ij} \tag{3}$$

$$S = \{(i,j) | L_{ij} \geq median(L_{ij})\} \tag{4}$$

where L_{EEDM} is the output, S is the difficult pixel set.

On the one hand, EEDM loss promotes the network to increase its sensitivity to target edges by using edge information as an additional constraint. On the other hand, it uses difficult pixel mining to help the model better focus on difficult-to-detect target areas, thereby preventing small targets from being submerged by the background.

C. Why "From Easy to Hard" Fits this Task?

For this task, the target regions are usually very small and low-contrast, which makes them highly sensitive to pseudo-

Table 5. Batch size investigation on the SIRST3 dataset. *Coarse* denotes coarse point supervision. *Centroid* denotes centroid point supervision.

Batch size	MSDA-Net Coarse + PAL				MSDA-Net Centroid + PAL			
	IoU	$nIoU$	P_d	F_a	IoU	$nIoU$	P_d	F_a
40	67.33	70.90	96.41	20.27	67.31	70.18	96.48	28.77
32	68.03	71.30	97.28	18.67	68.43	71.60	96.08	14.89
24	**70.31**	71.53	95.15	**13.78**	69.16	71.98	95.55	**7.82**
16	69.38	**71.55**	**97.41**	16.34	**69.21**	**72.40**	**97.01**	15.70
8	68.60	71.12	96.61	24.20	68.83	71.84	96.48	14.88

Table 6. Update period investigation on the SIRST3 dataset. *Coarse* denotes coarse point supervision. *Centroid* denotes centroid point supervision.

Update period	MSDA-Net Coarse + PAL				MSDA-Net Centroid + PAL			
	IoU	$nIoU$	P_d	F_a	IoU	$nIoU$	P_d	F_a
1	66.35	68.21	95.88	29.37	64.61	47.27	82.66	41.45
3	**70.01**	70.65	96.68	20.18	**69.65**	71.51	**97.41**	**13.62**
5	69.38	71.55	**97.41**	16.34	69.21	**72.40**	97.01	15.70
7	69.04	71.79	96.88	15.06	68.53	72.11	97.01	22.07
10	68.57	**71.80**	97.08	**14.38**	68.33	71.16	96.94	26.89

Table 7. Learning rate investigation on the SIRST3 dataset. *Coarse* denotes coarse point supervision. *Centroid* denotes centroid point supervision.

Learning rate	MSDA-Net Coarse + PAL				MSDA-Net Centroid + PAL			
	IoU	$nIoU$	P_d	F_a	IoU	$nIoU$	P_d	F_a
$1e^{-2}$	-	-	-	-	-	-	-	-
$5e^{-3}$	**69.53**	**71.69**	**97.81**	17.13	68.61	70.21	96.88	19.81
$1e^{-3}$	69.38	71.55	97.41	16.34	69.21	**72.40**	**97.01**	**15.70**
$5e^{-4}$	68.83	70.71	95.88	**16.23**	**69.24**	72.34	96.35	23.05
$1e^{-4}$	65.31	70.87	94.88	16.94	65.41	70.09	95.02	18.07
$5e^{-5}$	59.68	65.70	92.09	19.38	61.64	67.33	94.49	19.14

Table 8. Missed detection rate threshold investigation on the SIRST3 dataset. *Coarse* denotes coarse point supervision. *Centroid* denotes centroid point supervision.

T_{miss}	MSDA-Net Coarse + PAL				MSDA-Net Centroid + PAL			
	IoU	$nIoU$	P_d	F_a	IoU	$nIoU$	P_d	F_a
Change	**69.38**	71.55	**97.41**	16.34	**69.21**	**72.40**	97.01	15.70
0.2	69.30	**71.69**	96.88	19.05	69.07	70.89	**97.54**	16.32
0.4	69.14	71.14	96.21	**14.00**	69.06	71.38	96.94	16.52
0.6	68.73	71.61	96.61	15.82	68.91	71.41	96.41	17.69
0.8	67.41	71.06	96.08	15.61	68.42	71.14	96.08	**15.64**

label noise. This challenge is further exacerbated with single point supervision, as sparse annotations provide little spatial guidance. Training directly on unreliable pseudo-labels not only leads to degraded performance but also causes semantic drift due to misleading supervision. The "from easy to hard" idea can alleviate this problem, so we systematically introduce it into the SIRST detection with single point supervision for the first time and build a Progressive Active Learning (PAL) framework. Fig. 2 and Fig. 3 show that our PAL can gradually introduce harder samples and generate more refined pseudo-labels. The adaptive modifications to the characteristics of this task are mainly reflected in the easy-sample pseudo-label generation (EPG) strategy and fine dual-update strategy. The former automatically screens "easy samples" based on target characteristics through local brightness and edge information. The latter combines point labels with model feedback to regulate sample introduction and label evolution.

D. Compared with Other Methods

To further verify the performance of the proposed PAL framework, this section further presents the performance comparison with other methods. On the one hand, we compare with the MCLC (Monte Carlo Linear Clustering) method [1], which is a static pseudo-label generation method. On the other hand, we compare with the LELCM (Label Evolution framework based on Local Contrast Measure) method [2], which is a dynamic pseudo-label evolution method. Notably, the MCLC will additionally use the area information of the true target area to classify the target type ("Point", "Spot", "Extended") when generating pseudo-labels. However, the task setting of this study is that the training set only has point labels. To make a fair comparison and ensure that MCLC can cover all targets in the selected area, all targets in the MCLC experiment are assigned the "Extended" type. In addition, since the code of LELCM is not available, we directly use the results in the paper for comparison.

Comparison with the the MCLC method. To fully compare the performance of PAL and MCLC, we conduct comparative experiments on the SIRST3 dataset with different point labels and multiple networks. From Tab. 1, compared with MCLC, using our proposed PAL improves the IoU by 3.34%-14.67% and the P_d by 5.38%-10.96% on the SIRST3 dataset with coarse point. From Tab. 2, compared with MCLC, using our proposed PAL improves the IoU by 3.64%-13.21% and the P_d by 6.11%-8.31% on the SIRST3 dataset with centroid point. In summary, compared with MCLC, our PAL framework has significantly better performance.

Comparison with the the LELCM method. We further compare the performance of PAL and LELCM on three individual datasets with different point labels. The experimental results are shown in Tab. 3 and Tab. 4. From the results on three individual datasets with coarse point, compared with LELCM, using the proposed PAL improves the IoU by an average of 12.67% and the P_d by an average of 4.32%. From the results on three individual datasets with centroid point, compared with LELCM, using the proposed PAL improves the IoU by an average of 11.71% and the P_d by an average of 4.22%. In summary, compared with LELCM, our PAL framework has significantly better performance.

Table 9. IoU (%), $nIoU$ (%), P_d (%) and F_a (10^{-6}) values of different methods achieved on the SIRST3 dataset with centroid point labels. *NUAA-SIRST-Test*, *NUDT-SIRST-Test* and *IRSTD-1K-Test* denote the decompositions of *SIRST3-Test* to verify the robustness of the model. *DLN Centroid* denotes DLN-based methods under centroid point supervision.

Scheme	Description	SIRST3-Test				NUAA-SIRST-Test				NUDT-SIRST-Test				IRSTD-1K-Test			
		IoU	$nIoU$	P_d	F_a	IoU	$nIoU$	P_d	F_a	IoU	$nIoU$	P_d	F_a	IoU	$nIoU$	P_d	F_a
ACM	DLN Full	64.93	64.89	94.88	20.97	67.61	67.57	92.02	10.36	65.77	65.65	95.87	15.95	61.92	59.55	94.28	28.07
	DLN Centroid + LESPS	38.38	36.39	91.16	60.84	42.25	40.72	**88.97**	46.79	36.50	34.12	92.06	53.77	40.49	39.27	90.24	70.56
	DLN Centroid + PAL (Ours)	**51.51**	**53.73**	**92.82**	**35.98**	**57.87**	**58.00**	88.59	**28.40**	**54.73**	**54.76**	**94.50**	**21.79**	**42.51**	**45.79**	**91.25**	**49.80**
ALCNet	DLN Full	65.69	66.68	95.02	34.60	70.33	69.49	94.68	13.72	66.64	68.32	95.56	25.65	61.60	58.28	93.60	47.77
	DLN Centroid + LESPS	46.48	43.82	89.44	38.72	52.27	50.36	89.73	36.84	44.62	41.32	89.10	25.53	47.19	45.11	90.24	50.14
	DLN Centroid + PAL (Ours)	**55.01**	**57.93**	**93.94**	**31.63**	**62.51**	**64.42**	**90.87**	**28.74**	**58.43**	**58.46**	**94.71**	**23.81**	**44.78**	**49.29**	**91.92**	**38.89**
MLCL-Net	DLN Full	78.44	82.01	95.22	17.99	71.28	74.38	91.25	34.57	89.36	90.11	97.14	13.44	63.25	63.38	92.59	17.16
	DLN Centroid+ LESPS	37.28	36.95	90.90	45.13	40.29	40.51	89.73	49.39	35.25	35.08	92.06	43.36	40.72	39.34	88.22	45.42
	DLN Centroid + PAL (Ours)	**66.38**	**69.25**	**95.08**	**15.94**	**67.25**	**69.70**	**93.16**	**20.99**	**71.73**	**72.44**	**97.25**	**8.80**	**54.89**	**58.24**	**89.90**	**20.44**
ALCL-Net	DLN Full	79.38	81.09	96.08	18.64	72.79	75.36	93.54	20.92	87.68	88.21	97.57	13.03	68.36	63.66	93.60	22.64
	DLN Centroid + LESPS	55.63	54.13	94.15	26.29	59.99	60.30	93.92	**20.79**	56.03	52.49	94.92	13.67	52.17	52.95	**91.92**	38.24
	DLN Centroid + PAL (Ours)	**65.99**	**70.59**	**95.22**	**20.81**	**68.27**	**72.76**	**95.06**	29.57	**72.07**	**73.55**	**96.93**	**13.10**	**52.90**	**58.48**	89.90	**24.75**
DNANet	DLN Full	81.96	85.90	97.54	9.11	78.06	80.50	97.72	15.71	92.70	93.37	99.05	5.77	64.80	66.94	92.59	10.04
	DLN Centroid + LESPS	55.24	61.30	90.37	19.48	57.61	63.93	91.63	**6.86**	60.91	63.64	91.85	33.62	41.78	50.74	84.51	**11.29**
	DLN Centroid + PAL (Ours)	**66.97**	**70.63**	**96.28**	**14.66**	**71.01**	**73.33**	**96.96**	23.05	**71.79**	**73.67**	**98.94**	**13.35**	**53.83**	**57.71**	**87.21**	13.42
GGL-Net	DLN Full	82.06	85.34	97.74	12.76	78.40	79.99	96.96	20.72	92.07	92.34	99.26	4.87	66.68	67.88	93.60	17.08
	DLN Centroid + LESPS	56.66	54.79	93.55	24.41	60.06	60.45	92.02	**12.14**	56.34	53.24	94.92	26.52	55.39	53.89	**90.57**	26.06
	DLN Centroid + PAL (Ours)	**67.83**	**70.27**	**95.68**	**16.28**	**72.47**	**73.63**	**95.82**	26.34	**70.99**	**72.72**	**98.20**	**13.51**	**58.30**	**58.59**	87.54	**15.79**
UIUNet	DLN Full	83.14	85.29	97.34	15.44	76.64	78.43	95.82	12.69	92.94	93.29	98.41	4.11	70.04	66.19	95.29	25.55
	DLN Coarse + LESPS	49.63	48.07	88.97	54.56	57.54	57.12	86.31	59.20	47.88	45.28	90.05	27.97	48.91	47.65	87.88	75.23
	DLN Coarse + PAL (Ours)	**69.05**	**70.01**	**95.68**	**21.10**	**70.71**	**72.12**	**91.63**	**23.39**	**70.63**	**72.17**	**96.93**	**4.78**	**65.11**	**60.67**	**95.29**	**33.95**
MSDA-Net	DLN Full	83.46	85.97	97.41	17.15	74.81	78.61	95.06	30.94	93.62	94.03	99.26	9.67	70.98	67.16	93.60	19.51
	DLN Centroid + LESPS	53.57	50.34	92.43	29.05	56.98	56.00	90.87	**13.24**	51.72	48.17	94.18	20.27	55.94	51.48	88.22	40.67
	DLN Centroid + PAL (Ours)	**69.21**	**72.40**	**97.01**	**15.70**	**70.60**	**72.69**	**96.20**	26.55	**74.17**	**75.61**	**98.20**	**9.81**	**58.63**	**61.47**	**93.94**	**17.61**

E. More Ablation Experiments

In this section, we study more influencing factors in detail, including the batch size, update period of the refined dual-update strategy, learning rate, and missed detection rate threshold.

1) Batch size. To explore the impact of batch size on the performance of the final generated model, we explore the PAL framework with different batch size settings. From Tab. 5, when the batch size is set too large or too small, the final generated model will experience a slight performance degradation. Specifically, when the batch size is set too large, the number of model updates will be reduced and each update will be based on a large number of samples, making the gradient update smoother and reducing the randomness of the model, which makes the model more likely to overfit on the training data. When the batch size is set too small, each gradient update is based only on a small number of data samples, resulting in a large variance in the gradient estimate and large fluctuations in the gradient direction, which makes it easy to fall into a local optimum. On the whole, the final model with different batch size settings has relatively stable results, which verifies the stability of the proposed PAL framework. Based on the results in Tab. 5, the batch size is uniformly set to 16 in the experiments.

2) Update period of the refined dual-update strategy. To explore the impact of the update period of the refined dual-update strategy on the performance of the final generated model, we explore the PAL framework with different update period settings. The experimental results are shown in Tab. 6. Except that the update period is set to 1, the other settings have relatively stable results, which illustrates the robustness of the proposed PAL framework. The significant decrease in performance when the update period is set to 1 is because hard samples need to be trained for appropriate epochs after entering the training pool so that the model can fully learn the newly entered hard samples. It is just like when students face difficult content, they need to spend a certain amount of time to recognize, understand and apply it flexibly. If the time given is too short, students will not be able to deeply understand the knowledge, which will lead to a certain degree of knowledge confusion. In addition, the smaller the update period is set, the more time it takes to train. In the experiments, the update period is set to 5.

3) Learning rate. To explore the impact of the learning rate on the performance of the final generative model, we explore the PAL framework with different learning rate settings. As shown in Tab. 7, the results are consistent with the relationship between the learning rate settings and performance changes in general deep learning networks. When the learning rate is set too high ($1e^{-2}$), the update step size of

Table 10. IoU (%), $nIoU$ (%), P_d (%) and F_a (10^{-6}) values of different methods achieved on the separate NUAA-SIRST, NUDT-SIRST, and IRSTD-1k datasets with centroid point labels. *(213:214)*, *(663:664)* and *(800:201)* denote the division of training samples and test samples. *DLN Centroid* denotes DLN-based methods under centroid point supervision.

Scheme	Description	NUAA-SIRST (213:214)				NUDT-SIRST (663:664)				IRSTD-1K (800:201)			
		IoU	$nIoU$	P_d	F_a	IoU	$nIoU$	P_d	F_a	IoU	$nIoU$	P_d	F_a
ACM	DLN Full	65.67	63.74	90.11	24.01	65.33	65.12	95.87	12.92	60.45	53.70	92.26	46.06
ALCNet	DLN Full	66.41	65.18	91.63	35.26	69.74	70.67	97.46	11.15	62.47	55.25	88.89	36.79
MLCL-Net	DLN Full	74.68	76.50	95.82	28.74	94.03	93.97	98.73	7.72	64.86	63.35	91.25	23.23
	DLN Centroid + LESPS	32.69	32.59	82.89	**20.85**	34.11	32.00	89.52	46.17	46.59	45.16	86.87	29.42
	DLN Centroid + PAL (Ours)	**67.64**	**70.28**	**93.92**	45.00	**72.67**	**73.87**	**98.31**	**7.79**	**59.14**	**59.21**	**91.25**	**25.79**
ALCL-Net	DLN Full	72.22	72.64	94.68	35.06	92.80	93.01	99.05	2.21	65.56	65.03	91.58	9.68
	DLN Centroid + LESPS	35.56	32.99	92.02	**32.45**	46.08	43.19	86.88	41.00	45.77	42.80	86.53	**20.46**
	DLN Centroid + PAL (Ours)	**65.10**	**66.41**	**93.16**	44.11	**71.55**	**72.55**	**97.25**	**9.81**	**59.82**	**53.92**	**87.54**	22.38
DNANet	DLN Full	76.40	78.32	96.20	20.72	95.17	95.19	98.94	2.00	69.06	65.22	91.58	11.56
	DLN Centroid + LESPS	16.89	19.50	61.98	45.14	39.39	45.53	86.14	291.02	50.14	49.95	87.54	**16.13**
	DLN Centroid + PAL (Ours)	**66.16**	**66.68**	**91.63**	**24.42**	**73.24**	**74.41**	**97.99**	**9.10**	**59.53**	**57.62**	**88.89**	20.14
GGL-Net	DLN Full	75.47	75.93	97.34	20.65	94.86	94.93	99.47	1.03	69.09	65.52	92.59	13.68
	DLN Centroid + LESPS	54.85	53.82	90.49	21.20	48.23	46.86	89.95	50.23	48.82	42.67	81.48	**19.66**
	DLN Centroid + PAL (Ours)	**64.79**	**65.05**	**94.68**	**20.51**	**73.42**	**74.71**	**98.41**	**5.68**	**61.73**	**54.95**	**82.49**	21.22
UIUNet	DLN Full	78.02	76.85	96.58	14.68	95.07	95.10	98.73	0.21	70.94	64.32	91.25	10.57
	DLN Centroid + LESPS	26.05	25.27	54.75	37.52	41.60	39.43	85.93	68.23	41.40	40.42	87.88	84.93
	DLN Centroid + PAL (Ours)	**66.43**	**69.49**	**96.20**	**14.54**	**74.58**	**75.43**	**98.20**	**4.67**	**61.30**	**55.37**	**91.25**	**36.61**
MSDA-Net	DLN Full	76.73	77.78	96.20	21.75	95.27	95.18	99.15	1.72	70.98	65.70	93.94	33.95
	DLN Centroid + LESPS	48.63	48.70	87.45	37.59	32.90	30.59	83.92	112.97	48.67	46.41	85.86	26.72
	DLN Centroid + PAL (Ours)	**64.56**	**66.42**	**91.25**	**32.04**	**73.45**	**74.65**	**98.52**	**11.84**	**65.33**	**60.18**	**92.93**	**23.50**

the model will become larger, resulting in unstable parameter updates during training, which in turn leads to training collapse. When the learning rate is set too low ($5e^{-5}$), the update step size of the model becomes smaller, which leads to slow network optimization and easy to fall into local optimality. In addition, when the learning rate is set to $5e^{-5} - 5e^{-3}$, the results of the final generated model are stable. This verifies the robustness of the proposed PAL framework. In the experiment, we set the learning rate uniformly to $1e^{-3}$.

4) Missed detection rate threshold. During this research, we discover an interesting phenomenon: there are few false detections in the detection results of single-frame infrared small target (SIRST) based on DLNs. This can also be found from the order of magnitude ($1e^{-6}$) of the F_a. At the same time, the falsely detected areas will be eliminated in the coarse outer updates. Therefore, we focus on exploring the threshold of the missed detection rate in the model enhancement phase. To further explore the impact of the missed detection rate threshold setting in the coarse outer updates on the performance of the final generated model, we explore the PAL framework with different missed detection rate threshold settings. The experimental results are shown in Tab. 8. Compared with using a fixed missed detection rate threshold, using a variable value has relatively better detection results. At the same time, when the missed detection rate threshold is set larger and larger, the final performance gradually decrease. A larger threshold setting means that more harder samples will enter the training pool for training in the early of the model enhancement phase. Combined with the final results, it shows that hard samples should be input reasonably and gradually from simple to difficult in the model enhancement phase. This further verifies the effectiveness of our proposed progressive active learning idea. For the missed detection rate threshold, we set its initial threshold to 0.2 and gradually increase it to 1 as the number of epochs increases in the experiment.

F. More Quantitative Results

Considering that the main papers only conduct experiments on various datasets with coarse point labels, to further verify the effectiveness of our PAL framework, we conduct additional experiments on the SIRST3, NUAA-SIRST, NUDT-SIRST and IRSTD-1K datasets with centroid point labels in this section.

Evaluation on the SIRST3 dataset with centroid point labels. As shown in Tab. 9, consistent with the use of coarse point labels, when the networks (UIUNet, MSDA-Net) with obvious performance advantages under full supervision are embedded into the LESPS framework for single point supervision tasks, the potential performance advantages of these networks cannot be effectively exploited. However, the performance change trend of each SIRST detection network equipped with our PAL framework under single-point supervision is basically consistent with that of the network

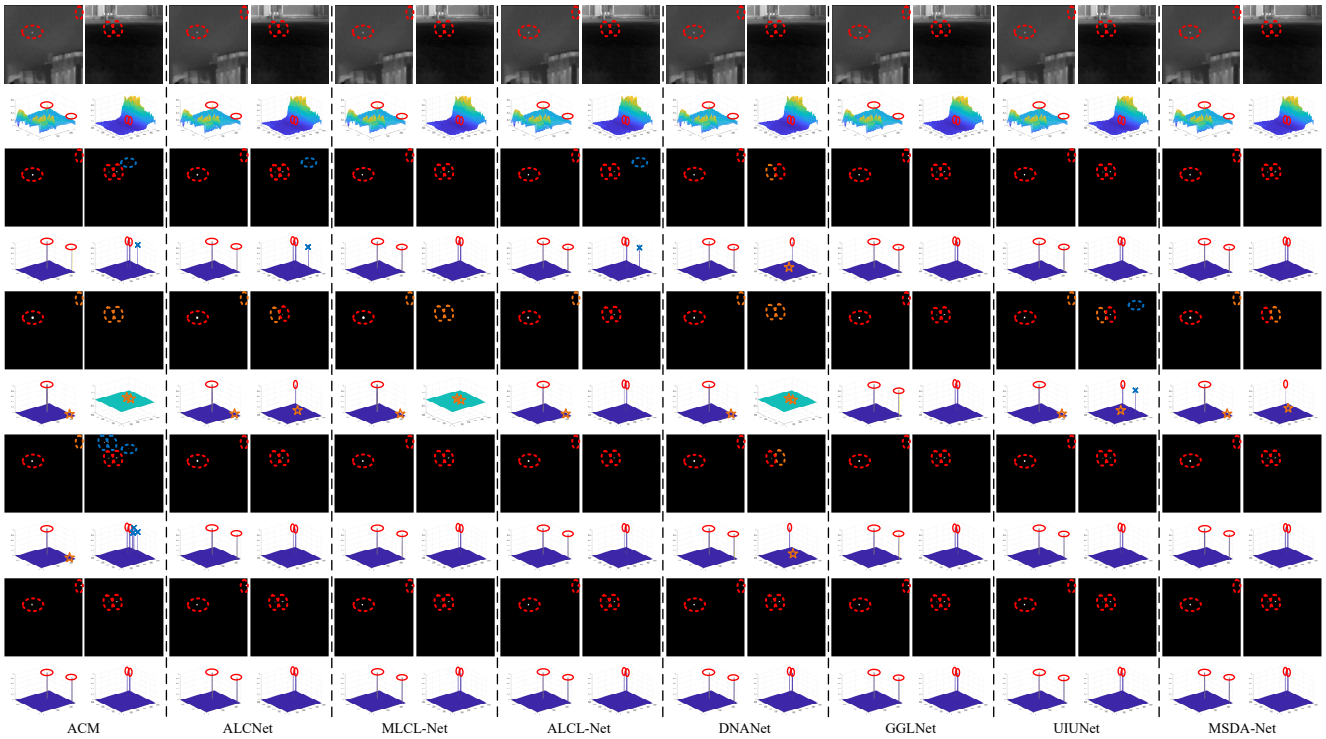

Figure 4. Visualization of several excellent methods on the SIRST3 dataset with coarse point labels. *Red* denotes the correct detections, *blue* denotes the false detections, and *yellow* denotes the missed detections. Every two rows from top to bottom: *Image*, *DLN Full*, *DLN Coarse + LESPS*, *DLN Coarse + PAL*, *True label*.

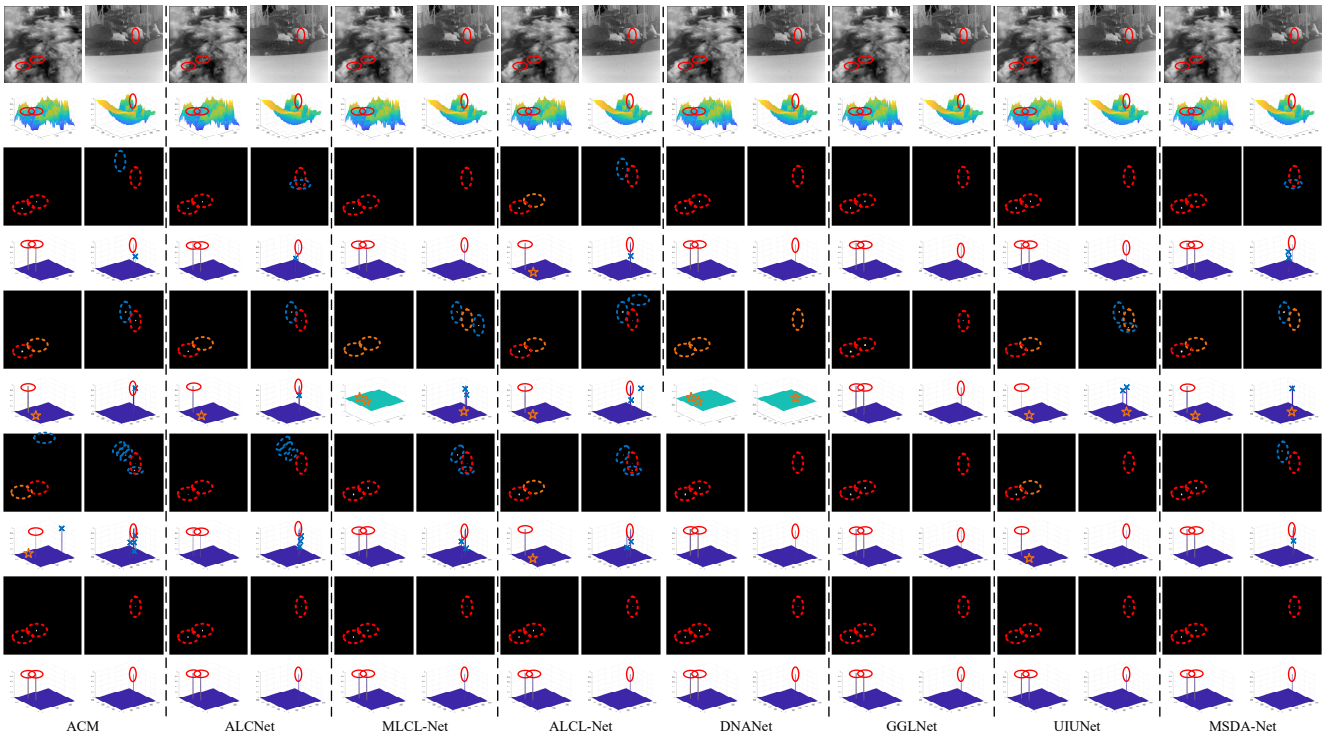

Figure 5. Visualization of several excellent methods on the SIRST3 dataset with centroid point labels. *Red* denotes the correct detections, *blue* denotes the false detections, and *yellow* denotes the missed detections. Every two rows from top to bottom: *Image*, *DLN Full*, *DLN Centroid + LESPS*, *DLN Centroid + PAL*, *True label*.

under full supervision. Our PAL framework can build an efficient and stable bridge between full supervision and single point supervision tasks. Meanwhile, compared with the LESPS framework, using our PAL framework improves the IoU by 8.53%-29.10%, the nIoU by 9.33%-32.30%, and the P_d by 1.07%-6.71% on the comprehensive SIRST3-Test. The improvement is very obvious. Compared with the fully supervised task on the SIRST3-Test, our PAL framework can reach 79.33%-84.63% on IoU, 82.08%-87.05% on nIoU, and has comparable performance on P_d. In addition, by observing the results of the three decomposed test subsets, our PAL framework has a stable performance that is better than that of the LESPS framework and is in line with the full supervision performance trend. These verify that the proposed PAL framework has excellent generalizability and robustness for SIRST detection tasks in multiple scenes and multiple target types.

Evaluation on three individual datasets with centroid point labels. To further explore the stability of the PAL framework when there are few training samples and centroid point labels are used, we conduct separate experiments on the NUAA-SIRST, NUDT-SIRST, and IRSTD-1K datasets. From Tab. 10, consistent with the results using coarse point labels, the performance of DLNs equipped with the PAL framework is significantly better than that with the LESPS framework. At the same time, some DLNs equipped with the LESPS framework will experience the phenomenon of "model invalidity" where the final generated model does not meet the F_a requirements. Specifically, compared with the LESPS framework, using the PAL framework improves the IoU by 9.39%-49.27%, the nIoU by 7.67%-47.18%, and the P_d by 1.01%-41.45%. The improvement is very significant. In addition, except for some minor differences, the performance change trend of each SIRST detection network equipped with our PAL framework under single point supervision is basically consistent with that of the network under full supervision. These results fully verify that our PAL framework still has excellent robustness in the single point supervised SIRST detection task with a small number of training samples.

G. More Qualitative Results

To further qualitatively compare and analyze the performance of the proposed PAL framework, in this section, we present detailed visualizations of the detection results of multiple methods on multiple datasets using either coarse point labels or centroid point labels.

Visualization on the SIRST3 dataset. As shown in Fig. 4 and Fig. 5, we can find that whether using coarse point labels or centroid point labels, DLNs equipped with our PAL framework are significantly better than those equipped with the LESPS framework. From the 3D results, the LESPS framework easily leads to a large number of missed detections in difficult scenarios, whereas the PAL framework can solve this problem. In addition, for some images, the target-level detection effect of DLNs equipped with the PAL framework under single point supervision is even better than that under full supervision. From the 2D results, DLNs equipped with the PAL framework are significantly better than the LESPS framework in pixel-level segmentation. These results fully demonstrate the effectiveness of our proposed PAL framework for the SIRST detection task with single-point supervision.

Visualization on three individual datasets. As shown in Figs. 6 to 17, we provide a detailed visualization of various methods, different point labels and different training frameworks. First, when the LESPS framework is used, there is a significant decrease in detection performance in some networks, such as Fig. 10, Fig. 11, Fig. 16, and Fig. 17. This shows that the LESPS framework is prone to unstable performance when facing a dataset with few samples. Secondly, DLNs equipped with the PAL framework generally have more refined segmentation effects than LESPS. Finally, compared with the detection results with full supervision, DLNs equipped with the PAL can achieve similar results with single point supervision. These results fully demonstrate the robustness of our proposed PAL framework on the single point supervised SIRST detection task with a small number of samples.

References

[1] Boyang Li, Yingqian Wang, Longguang Wang, Fei Zhang, Ting Liu, Zaiping Lin, Wei An, and Yulan Guo. Monte carlo linear clustering with single-point supervision is enough for infrared small target detection. In *ICCV*, pages 1009–1019, 2023. 1, 3

[2] Dongning Yang, Haopeng Zhang, Ying Li, and Zhiguo Jiang. Label evolution based on local contrast measure for single-point supervised infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 1, 3

[3] Xinyi Ying, Li Liu, Yingqian Wang, Ruojing Li, Nuo Chen, Zaiping Lin, Weidong Sheng, and Shilin Zhou. Mapping degeneration meets label evolution: Learning infrared small target detection with single point supervision. In *CVPR*, pages 15528–15538, 2023. 1

[4] Chuang Yu, Yunpeng Liu, Shuhang Wu, Xin Xia, Zhuhua Hu, Deyan Lan, and Xin Liu. Pay attention to local contrast learning networks for infrared small target detection. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. 2

[5] Jinmiao Zhao, Zelin Shi, Chuang Yu, and Yunpeng Liu. Infrared small target detection based on adjustable sensitivity strategy and multi-scale fusion. *arXiv preprint arXiv:2407.20090*, 2024. 2

[6] Jinmiao Zhao, Zelin Shi, Chuang Yu, and Yunpeng Liu. Multi-scale direction-aware network for infrared small target detection. *arXiv preprint arXiv:2406.02037*, 2024. 2

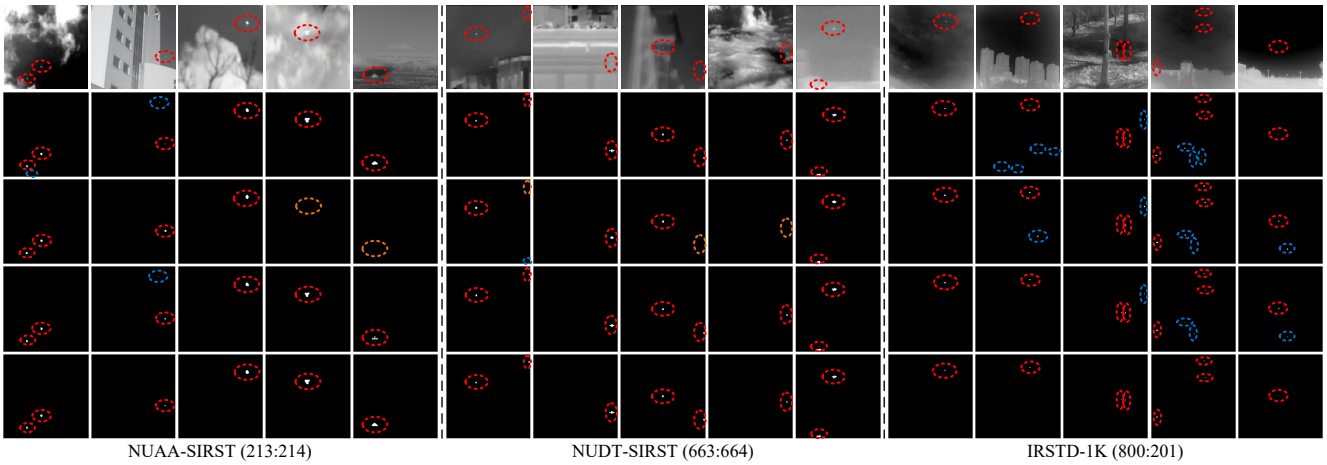

Figure 6. Visualization of MLCL-Net on the SIRST3 dataset with coarse point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Coarse + LESPS*, *DLN Coarse + PAL*, *True label*.

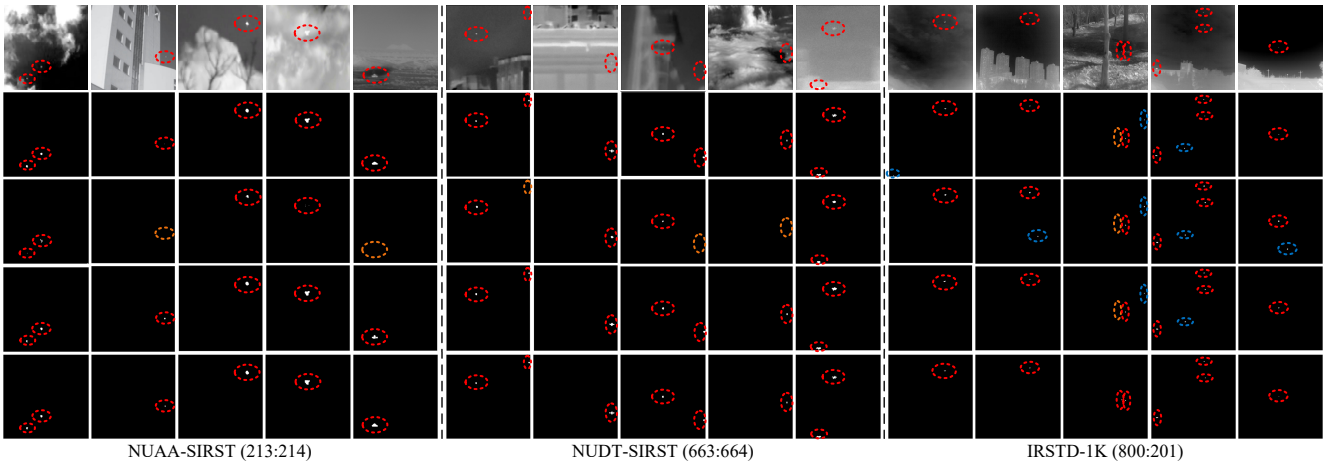

Figure 7. Visualization of ALCL-Net on the SIRST3 dataset with coarse point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Coarse + LESPS*, *DLN Coarse + PAL*, *True label*.

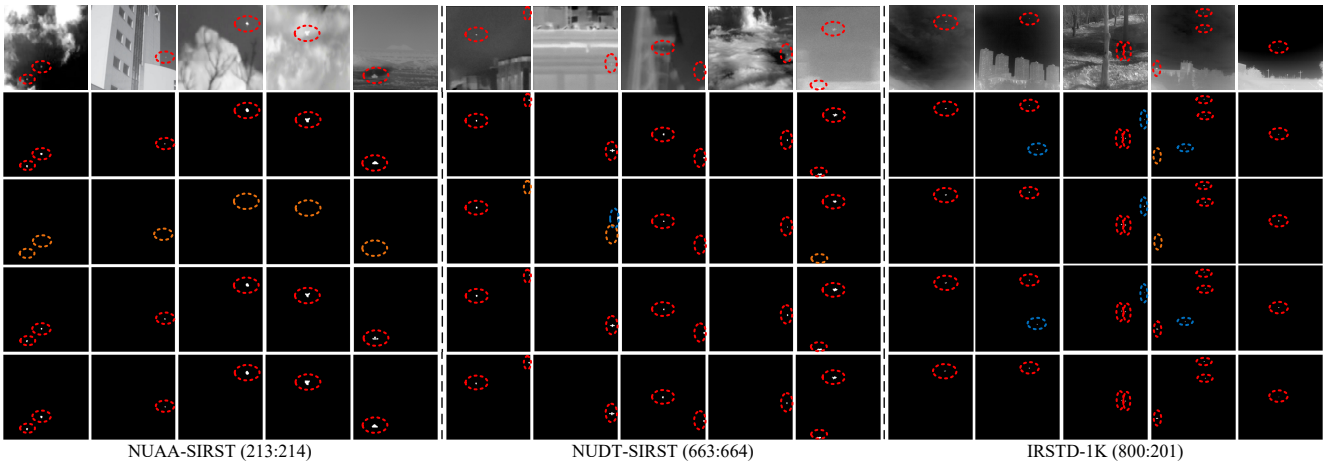

Figure 8. Visualization of DNANet on the SIRST3 dataset with coarse point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Coarse + LESPS*, *DLN Coarse + PAL*, *True label*.

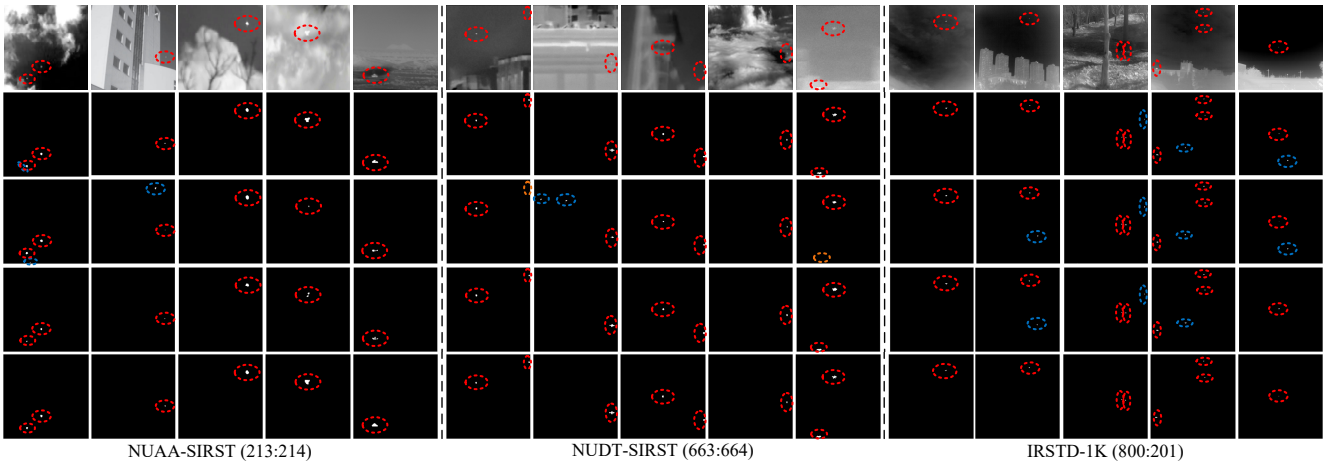

Figure 9. Visualization of GGL-Net on the SIRST3 dataset with coarse point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Coarse + LESPS*, *DLN Coarse + PAL*, *True label*.

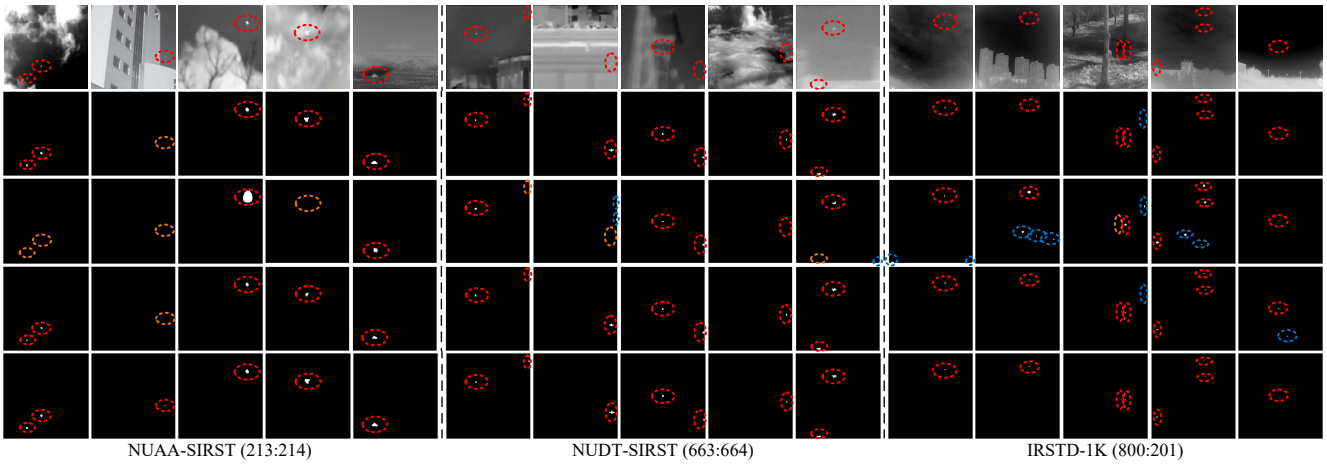

Figure 10. Visualization of UIUNet on the SIRST3 dataset with coarse point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Coarse + LESPS*, *DLN Coarse + PAL*, *True label*.

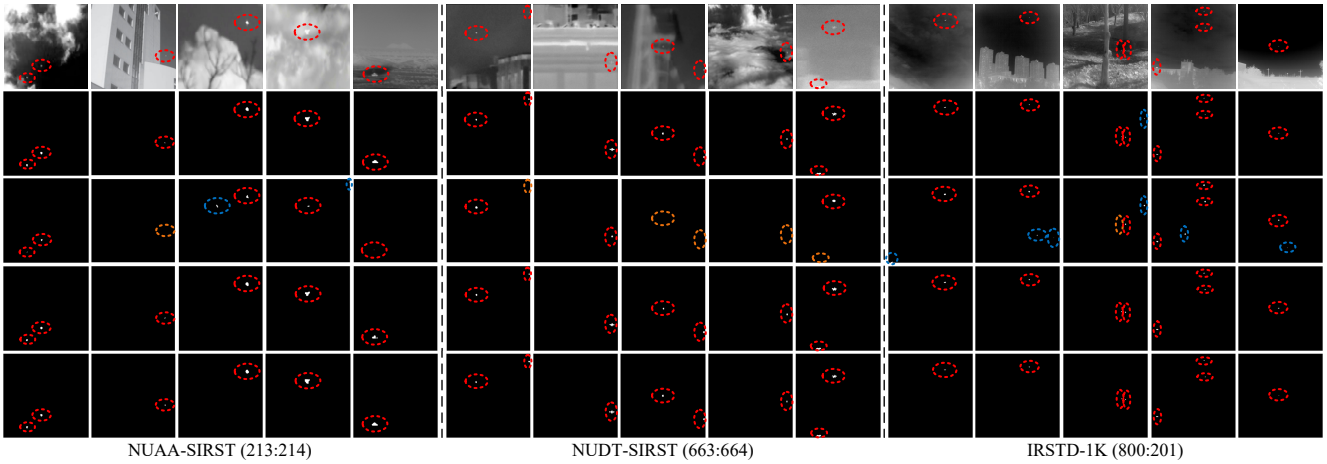

Figure 11. Visualization of MSDA-Net on the SIRST3 dataset with coarse point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Coarse + LESPS*, *DLN Coarse + PAL*, *True label*.

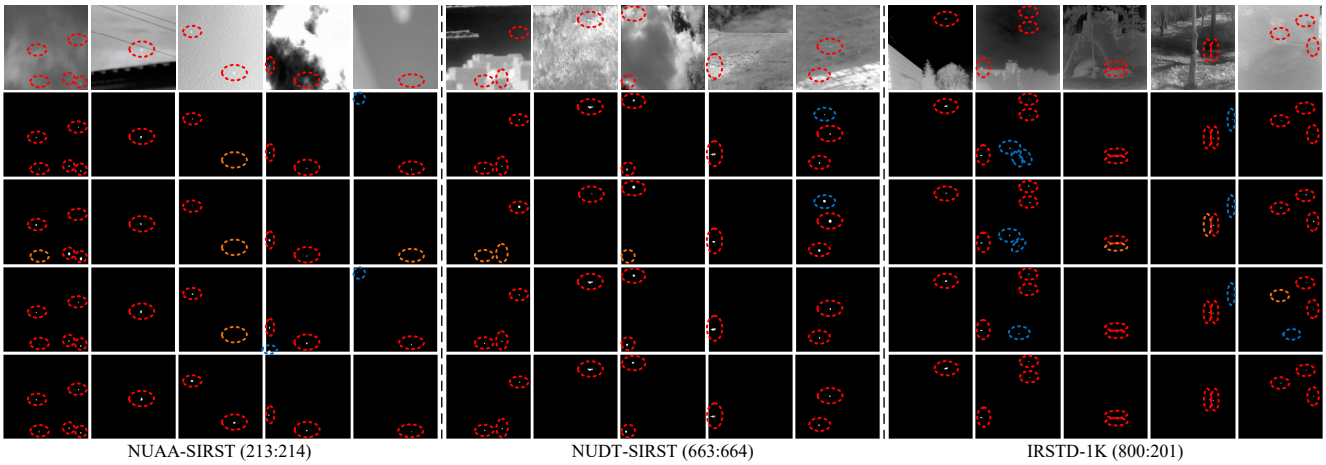

Figure 12. Visualization of MLCL-Net on the SIRST3 dataset with centroid point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Centroid + LESPS*, *DLN Centroid + PAL*, *True label*.

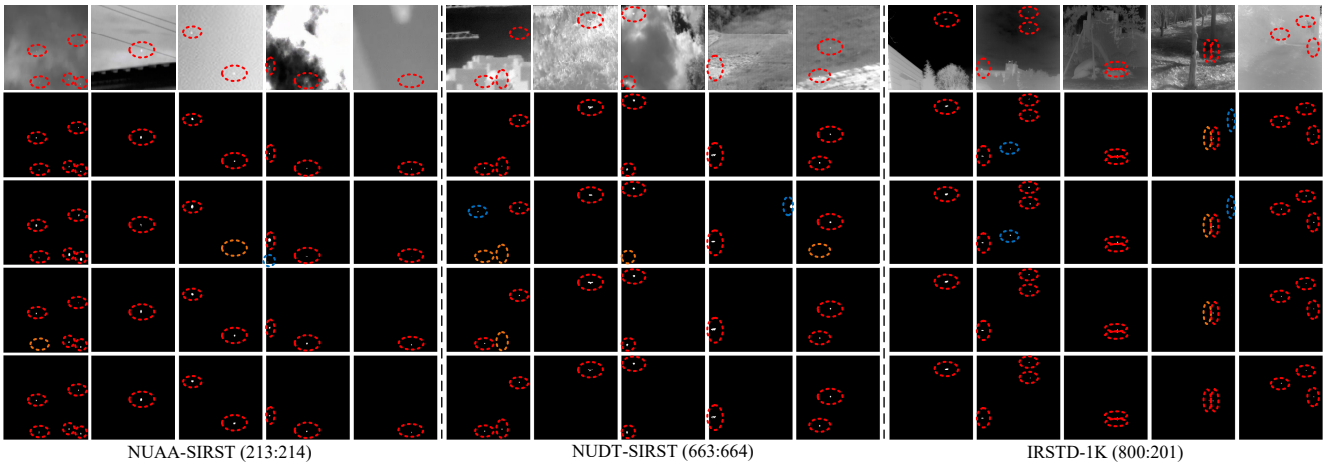

Figure 13. Visualization of ALCL-Net on the SIRST3 dataset with centroid point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Centroid + LESPS*, *DLN Centroid + PAL*, *True label*.

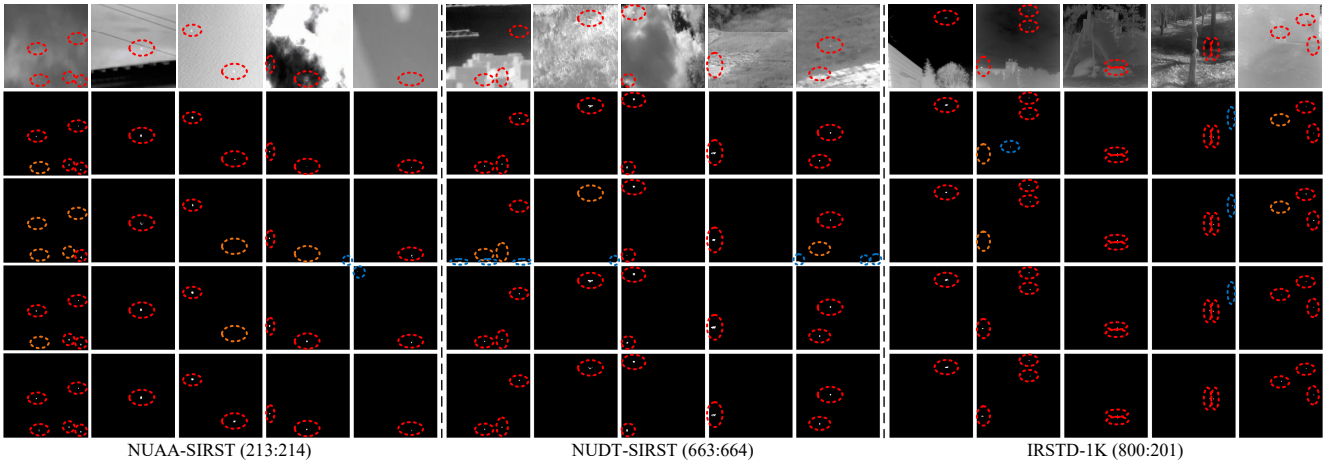

Figure 14. Visualization of DNANet on the SIRST3 dataset with centroid point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Centroid + LESPS*, *DLN Centroid + PAL*, *True label*.

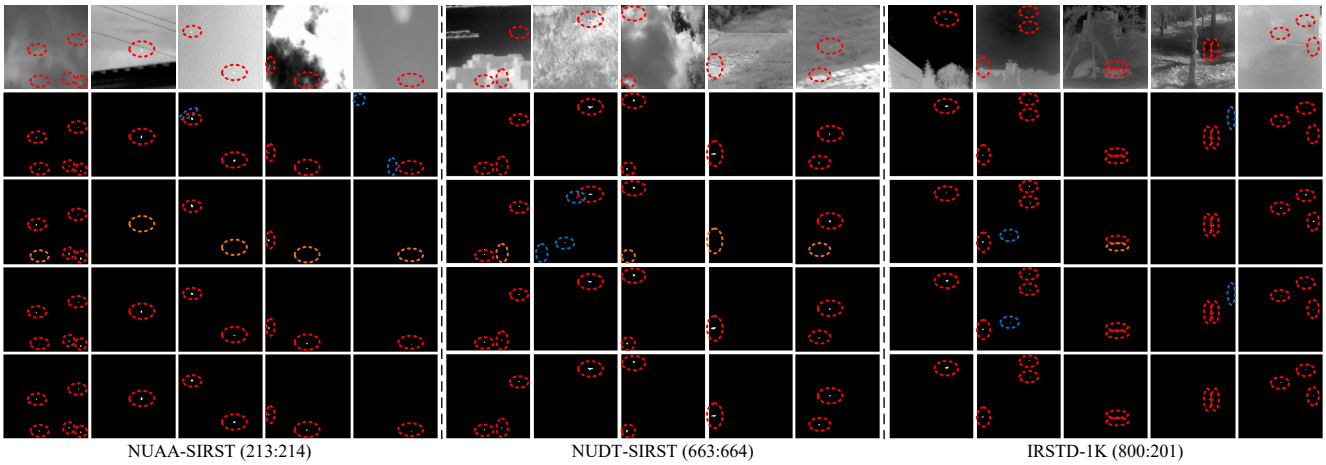

Figure 15. Visualization of GGL-Net on the SIRST3 dataset with centroid point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Centroid + LESPS*, *DLN Centroid + PAL*, *True label*.

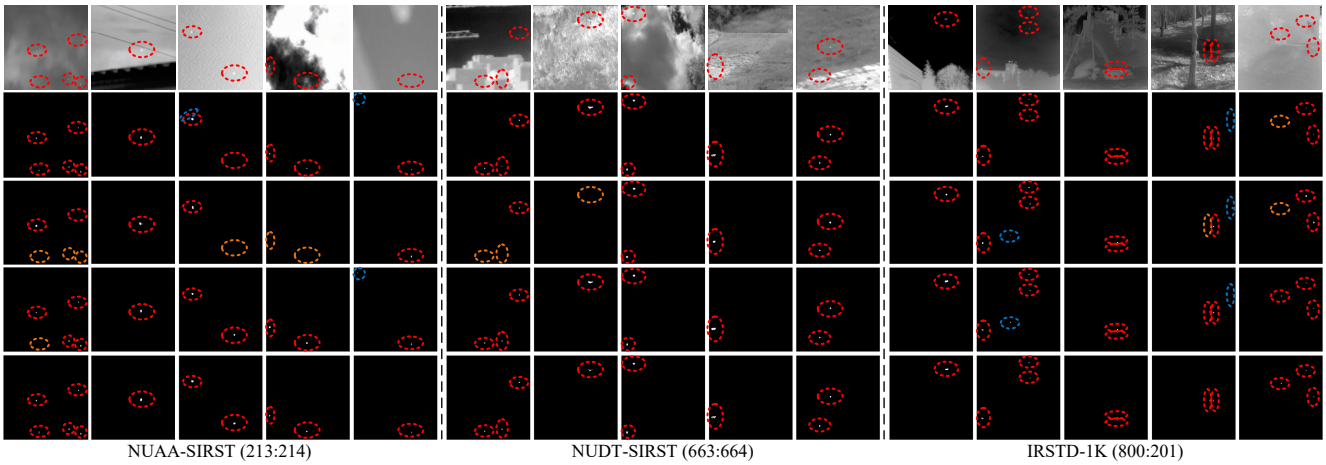

Figure 16. Visualization of UIUNet on the SIRST3 dataset with centroid point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Centroid + LESPS*, *DLN Centroid + PAL*, *True label*.

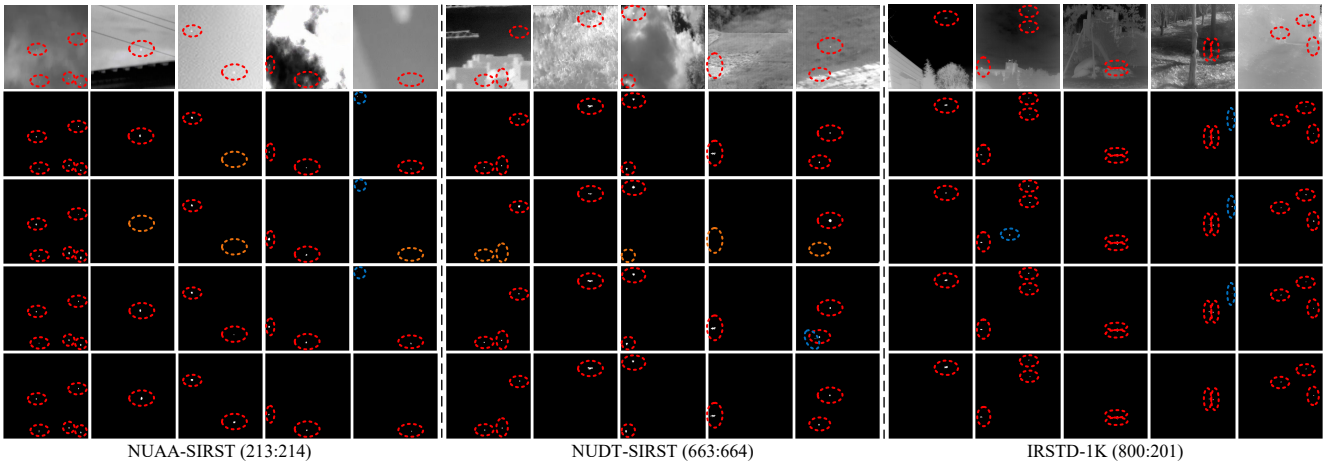

Figure 17. Visualization of MSDA-Net on the SIRST3 dataset with centroid point labels. *Red*, *blue*, and *yellow* denotes correct detections, false detections, and missed detections. From top to bottom: *Image*, *DLN Full*, *DLN Centroid + LESPS*, *DLN Centroid + PAL*, *True label*.