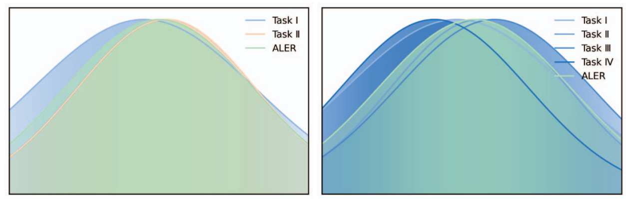
Dual Domain Control via Active Learning for Remote Sensing Domain Incremental Object Detection

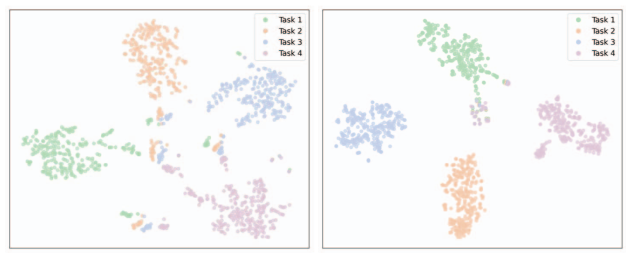
Jiachen Sun *, De Cheng *, Xi Yang †, Nannan Wang
State Key Laboratory of Integrated Services Networks,
School of Telecommunications Engineering, Xidian University, Xi'an 710071, China
sunjc@stu.xidian.edu.cn, {dcheng, yangx, nnwang}@xidian.edu.cn

Abstract

Domain incremental object detection in remote sensing addresses the challenge of adapting to continuously emerging domains with distinct characteristics. Unlike natural images, remote sensing data vary significantly due to differences in sensors, altitudes, and geographic locations, leading to data distribution shifts and feature misalignments. These challenges make it difficult for models to generalize across domains while retaining knowledge from previous tasks, requiring effective adaptation strategies to mitigate catastrophic forgetting. To address these challenges, we propose the Dual Domain Control via Active Learning (Active-DDC) method, which integrates active learning strategies to handle data distribution and model feature shifts. The first component, the Data-based Active Learning Example Replay (ALER) module, combines a high-information sample selection strategy from active learning with the characteristic extreme foreground-background ratio in remote sensing images, enabling the selection of highly representative samples for storage in a memory bank. The second component, the Query-based Active Domain Shift Control (ADSC) module, leverages the query vector, a key element for DETR-based detectors, to implement query active preselection and optimal transport matching, thus facilitating effective cross-domain knowledge transfer. Our method achieves optimal performance in domain incremental tasks across four remote sensing datasets, and ablation studies further validate the effectiveness of both components.



(a) Data shift and the balancing effect of ALER



(b) Feature shift and the balancing effect of ADSC

Figure 1. Data and Feature Shift: (a) Illustrates the data distribution for two-task and four-task settings, along with the combined training data distribution generated by ALER, demonstrating its effectiveness in balancing data distribution shifts. (b) Presents the T-SNE visualization of query features from the Fine-tuning model and the ADSC-trained model on incremental data, highlighting ADSC's capability in mitigating feature shift.

1. Introduction

Deep learning has greatly advanced remote sensing object detection, but domain-incremental learning remains challenging due to shifting imaging conditions, sensor types, and geographic variations. Unlike natural images, remote sensing data exhibit complex distributions, making it difficult for models to generalize across domains while retaining past knowledge. In this setting, models must detect objects in new domains without catastrophic forgetting. However, conventional methods struggle with data distribution shifts and feature misalignment caused by domain changes. These challenges are further amplified in remote sensing, where heterogeneous data sources and extreme foreground-background imbalances complicate adaptation, requiring more effective strategies for knowledge retention and domain adaptation. Traditional object detection methods struggle to generalize across domains, especially when new data from different environments or conditions

*Contributed equally.
†Corresponding author.

becomes available. This issue is particularly pronounced in remote sensing, where large-scale, high-dimensional images introduce both data distribution shifts and feature misalignments.

Incremental object detection methods can be categorized into knowledge distillation-based methods [8, 9, 22, 28, 47], example replay-based methods [8, 9, 16, 26, 27], and adapter-based methods [10, 37, 38, 41, 43]. Knowledge distillation and example replay methods are particularly effective as they do not require additional network parameters, making them popular in incremental learning. While some methods perform well in multi-domain incremental object detection tasks on natural images, remote sensing images present unique challenges, particularly in data and feature distribution.

Remote sensing images are captured under diverse conditions—different sensors, altitudes, and geographic locations—resulting in notable domain shifts across datasets, as shown in Fig. 1(a). Variations in resolution, scale, and spatial layout further hinder model generalization. In contrast, natural images are typically acquired under more consistent settings, leading to relatively stable data distributions.

Remote sensing images also exhibit complex, multi-scale characteristics, including objects with diverse shapes and sizes, along with rich contextual cues such as terrain and vegetation, as illustrated in Fig. 1(b). These factors challenge the extraction and alignment of features across domains. By comparison, natural images tend to contain more standardized objects and consistent contexts, making cross-domain knowledge transfer significantly easier.

To address the above challenges, we propose the Dual Domain Control via Active Learning (Active-DDC) framework. The first component, Data-based Active Learning Example Replay (ALER), mitigates data distribution shifts, as illustrated in fig. 1(a). It employs an active learning strategy to select representative samples, which are stored in a memory bank. By preserving essential information from previous domains, ALER effectively minimizes the impact of data distribution shifts as new domains emerge. The second component, Query-based Active Domain Shift Control (ADSC), addresses the feature shift problem inherent in domain-incremental learning, with its control effect on feature migration depicted in fig. 1(b). By leveraging a query selection strategy combined with optimal transport matching, ADSC facilitates knowledge distillation from the previous domain. The class distillation mechanism helps preserve essential features, ensuring that the model effectively retains critical knowledge while adapting to new domains.

The main contributions of this work are summarized as follows:

- To the best of our knowledge, we are the first to formulate the task of domain-incremental object detection for remote sensing images and propose the Active-DDC method.
- We introduce the Data-based Active Learning Example Replay (ALER) module, which effectively mitigates data migration in incremental tasks and enhances the representativeness of selected examples.
- We propose the Query-based Active Domain Shift Control (ADSC) module, which addresses feature shifts in incremental tasks and enables effective cross-domain knowledge transfer through query-based preselection and optimal transport matching.
- Extensive experiments demonstrate that our approach achieves state-of-the-art performance, with ablation studies further validating the effectiveness of our proposed modules.

2. Related Work

Incremental Object Detection (IOD). As a typical extension of incremental learning, IOD aims to handle both old and new task data appearing simultaneously in images. In this task, Knowledge Distillation (KD) [4, 15, 29, 39] has been widely adopted for its ability to leverage prior knowledge from new training samples, thereby minimizing discrepancies between the responses of the previous and current models [3, 13, 31, 45]. A pioneering approach, ILOD [36], applies knowledge distillation to Fast R-CNN [11] to retain responses for old classes, effectively mitigating catastrophic forgetting.

Subsequently, knowledge distillation has been extended to various object detection frameworks, including SID [32] on CenterNet [46], RILOD [19] on RetinaNet [24], and ERD [9] on GFLV1 [23]. Further advancements have applied this technique to Faster R-CNN [34], with methods such as CIFRCN [14], Faster ILOD [31], DMC [42], BNC [7], and IOD-ML [17]. More recently, knowledge distillation has been integrated into Transformer-based detectors, leading to methods such as CL-DETR [28], DyQ-DETR [43], and Incremental-DETR [8], leveraging architectures like DETR [2], UP-DETR [6], and Deformable DETR [48].

Exemplar Replay (ER) is another key strategy in incremental object detection [1, 12, 33, 35]. ORE [16] uses an exemplar set for fine-tuning, while AFD [26] introduces adaptive sampling, and CL-DETR [28] applies distribution-preserving calibration. ABR [27] further enhances replay with a mosaic-based approach. However, these methods often rely on exemplars with incomplete annotations, overlooking their informativeness and reliability. In domain-incremental object detection, this limitation weakens cross-domain knowledge transfer, making it harder to address data and feature distribution shifts.

Transformer-based Object Detection. Detection Transformer (DETR) [2] redefines object detection as a set prediction problem, using a transformer-based architecture to model object relationships via global attention.

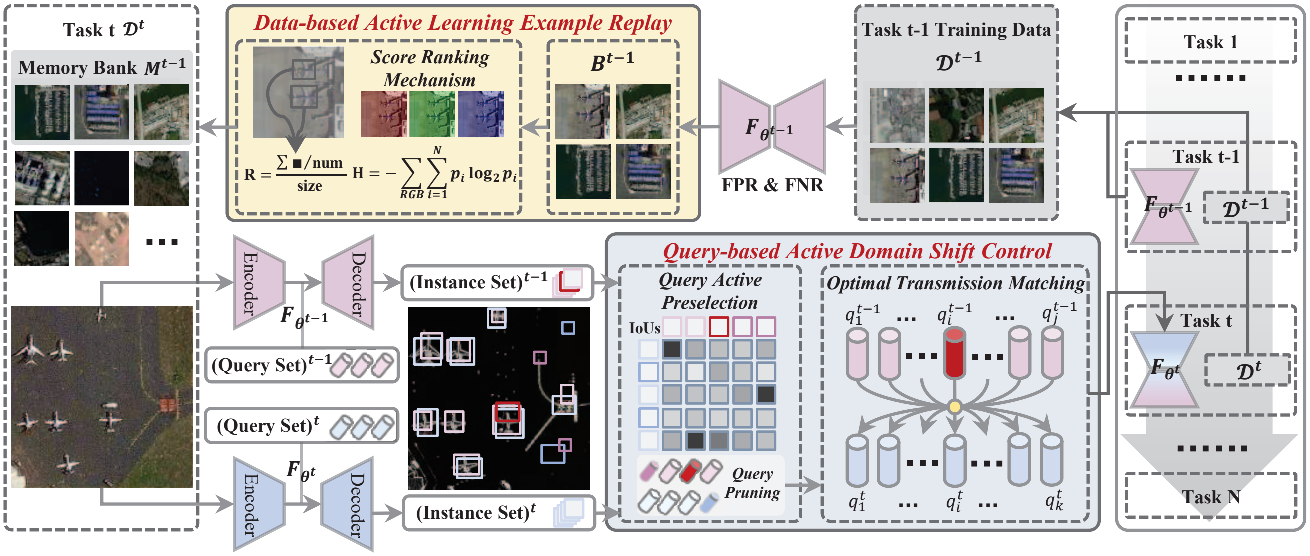


Figure 2. Overview of our Active-DDC framework. During cross-domain adaptation, ALER first selects highly representative samples and stores them in a memory bank. Then, the model is jointly trained using both the memory bank and the new domain data. Finally, ADSC controls query vector shifts to mitigate domain drift. Active-DDC effectively suppresses both data and feature shifts caused by domain changes.

It eliminates the need for hand-crafted components like non-maximum suppression by employing learnable object queries and Hungarian matching [18]. To improve efficiency, Deformable DETR [48] introduces sparse attention on multilevel feature maps, accelerating convergence and enhancing small-object detection. Various DETR variants [6, 20, 25, 30, 49] further refine performance. Our method is built upon the widely adopted Deformable DETR.

3. Method

3.1. Problem Definition

Object detection primarily focuses on accurately identifying and localizing objects of interest within an image. Let $(x, y) \in \mathcal{D}$ denote a dataset $\mathcal{D}$, where x represents the images and y corresponds to their ground truth labels. An ideal object detector $F_\theta(x)$ should generate a set of predictions $\hat{y}$ from the features $f_\theta(x)$ extracted from the input image.

Domain incremental object detection aims to train a model to detect objects across a sequence of T tasks, where each task $\mathcal{D}^t = (x^t, y^t)$ consists of a set of images x^t from the t-th domain. The model should be able to detect objects in new domain images x^t while preserving its ability to detect objects in previously encountered domains $x^{1:t-1}$, without experiencing catastrophic forgetting.

However, domain increment introduces two primary effects: the shift in data distribution and the shift in feature vectors. Specifically, let $\mu_t = \mathbb{E}[x^t]$ represent the mean of the image data in the t-th domain, and $\sigma_t^2 = \text{Var}(x^t)$ represent the variance of image data in the t-th domain. The distribution of image data in the t-th domain can be expressed as $X^t \sim N(\mu_t, \sigma_t^2)$. Similarly, $f_\theta^t(x^t)$ denotes the features extracted by the model F_θ^t in the t-th domain. The shifts in data distribution and feature vectors between the t-th domain and the $(t-1)$-th domain can be defined as $D(X^t, X^{t-1})$ and $D(f_\theta^t(x^t), f_\theta^{t-1}(x^t))$, respectively. These two types of shifts contribute to catastrophic forgetting, ultimately causing the detector to perform poorly on data from incremental domains.

3.2. Overall Framework

Our proposed Dual Domain Control via Active Learning (Active-DDC) framework is designed to tackle the challenges of domain-incremental object detection by mitigating data distribution shifts and model feature shifts through a combination of active learning and knowledge distillation strategies. Active-DDC consists of two key modules: Data-based Active Learning Example Replay (ALER) and Query-based Active Domain Shift Control (ADSC). As illustrated in fig. 2.

To counteract data distribution shifts, the ALER module leverages an active learning-based high-information sample selection strategy, ensuring that the most representative samples are stored in the memory bank for effective knowledge retention. To address model feature shifts, the ADSC module exploits query vectors in DETR-based detectors, implementing query preselection and an optimal transport-based matching mechanism to achieve effective many-to-many query alignment, thereby enhancing cross-domain knowledge transfer.

Our method is applied to Transformer-based detectors, Deformable-DETR and DETR. Compared to DETR, Deformable-DETR introduces deformable attention, which significantly reduces computational cost and enhances detection performance, particularly for small and dense objects.

3.3. Active Learning Example Replay

Example Replay (ER) mitigates data distribution shifts by replaying past samples, reducing variance, preventing overfitting, and accelerating learning. Active learning further enhances this by selecting the most informative samples rather than relying on random selection. To address distribution shifts, we propose Active Learning Example Replay (ALER), which selects representative samples based on factors like representativeness, uncertainty, and difficulty. These samples are stored in a memory bank and trained alongside new task data. By dynamically updating the memory bank, ALER helps retain knowledge from previous domains while adapting to new ones. Figure 2 illustrates the ALER pipeline.

At the start of task t, ALER pre-selects difficult samples based on the false positive rate (FPR) and false negative rate (FNR), which are stored in the memory buffer B^{t-1} for further processing after task $t-1$ training. The memory buffer and bank are denoted as B and M, respectively, with a limited size. Sample selection is crucial for performance. After applying the scoring ranking mechanism, the memory bank M^{t-1} holds highly representative samples. The ranking mechanism considers information complexity, sample difficulty, and foreground-background uncertainty, following active learning principles to score and retain high-scoring samples for joint training with $\mathcal{D}^t$ during task t.

First, based on the active learning strategy of information entropy, images with higher entropy contain more detailed and structured information, while images with lower entropy tend to contain less information and are often simple or flat regions. The information entropy of the image x in the memory buffer is calculated as follows:

$$H(x) = -\sum_{i=1}^{N} p_i \log_2 p_i, \tag{1}$$

where p_i represents the probability of the i-th color level occurring in the image, and N is the total number of possible color levels.

Inspired by the DyQ-DETR [43] method and the extreme foreground-background ratio in remote sensing images, we then calculate the average foreground-background ratio for each image:

$$r_{\text{avg}}(x) = \frac{\sum_{k=1}^{N_{\text{fg}}(x)} \text{size}(\text{fg}(k))}{N_{\text{fg}} \cdot \text{size}(x)}, \tag{2}$$

where $N_{\text{fg}}(x)$ is the number of foreground objects in image x, and size$(\cdot)$ represents the dimension calculation.

Subsequently, the score mapping function maps the average foreground-background ratio to the risk level of the sample:

$$R(x) = k_1 e^{-b_1 \cdot r_{\text{avg}}(x)} + k_2 e^{-(r_{\text{avg}}(x) - c_1)} + c_2, \tag{3}$$

where $r_{\text{avg}}(x)$ represents the average foreground-background ratio of the sample, k, b, and c are the mapping coefficients, respectively. Serving as the risk assessment metric from the DyQ-DETR [43] method. However, our approach places greater emphasis on the difficulty and risk associated with remote sensing images. As a result, the score mapping function assigns a higher value when the average foreground-background ratio $r_{\text{avg}}(x)$ is low.

The final score is composed of both the image information entropy and the foreground-background ratio:

$$S(x) = \alpha H(x) + \beta R(x), \tag{4}$$

where α and β are the information weight and risk weight, respectively. The score $S(x)$ ultimately determines whether image x from the memory buffer B^{t-1} is selected to be loaded into the memory bank M^{t-1} for multiple rounds of training.

During cross-task learning, the model first tests the samples in the memory bank M^{t-1} and stores the difficulty p_{t-1} of the current samples in the memory bank. Then, during the hard sample selection, the difficulty p_t of the new task samples is stored. The ratio of p_{t-1} to p_t is used to determine the number of new task samples to be loaded into the memory bank M^t. When discarding old task samples, the model also uses a score ranking mechanism, where samples with lower scores are discarded.

Based on entropy and foreground-background ratio scores, ALER dynamically selects and discards samples from both new and old tasks. This selective approach helps the model better adapt to new domains while retaining knowledge from previous domains, reducing the impact of data distribution shifts.

3.4. Active Domain Shift Control

In incremental object detection, knowledge distillation helps mitigate catastrophic forgetting by transferring knowledge from previous tasks. In DETR-based detectors, query vectors play a crucial role, representing potential detections that interact with image features to predict objects. To address feature shifts, we propose Active Domain Shift Control (ADSC), which integrates knowledge distillation with active learning. ADSC first preselects highly overlapping queries for distillation, then applies optimal transport for multi-to-multi query matching, enhancing knowledge transfer. This allows the student model to retain essential knowledge while extracting richer semantic information.

In DETR, the query vectors are unordered and dynamic. They do not have any inherent order or explicit category label association, and during training, the query vectors are dynamically adjusted based on the features of the input image. Additionally, the focus of each query may shift as the task progresses. To address this, query Active Preselection is applied to preselect and match query vectors. After the training of task $t-1$ is completed, we use both the teacher model F_θ^{t-1} and student model F_θ^t to perform detection and calculate the Intersection over Union (IoU) between their predicted boxes. Based on the IoU values, we filter out query vectors with significant overlap for subsequent knowledge distillation. This step effectively removes queries with low semantic relevance, thereby reducing ineffective learning.

To preselect queries, we calculate the Intersection over Union (IoU) between each query and the prediction boxes from the teacher model. Let the predicted boxes from the student model in task t be b_s^t, and the predicted boxes from the teacher model in task $t-1$ be b_t^{t-1}. Then, we set a threshold $\text{IoU}_{\text{thresh}}$, and select queries whose IoU exceeds this threshold for distillation. The final query preselection can be expressed as:

$$\begin{aligned}&(Q^t_{\text{selected}}, Q^{t-1}_{\text{selected}}) = \\ &\left\{(Q^t, Q^{t-1}) \mid \text{IoU}(b_s^t, b_t^{t-1}) > \text{IoU}_{\text{thresh}}\right\},\end{aligned} \tag{5}$$

where Q_i is the selected query vector, which satisfies the condition that the IoU with the teacher model's predicted box is greater than the set threshold $\text{IoU}_{\text{thresh}}$.

Our method focuses on the distribution of query vectors between the target and source models. Let ξ_t and ξ_{t-1} represent the query vector distributions of the teacher model F_θ^{t-1} and the student model F_θ^t, respectively. Q_i^t and Q_i^{t-1} are the individual query elements in these distributions. We construct a cost matrix based on the Intersection over Union (IoU) and cosine similarity, where higher IoU overlap and higher cosine similarity imply lower cost.

The optimal transport (OT) distance between the teacher's and student's query vectors can be expressed as:

$$\begin{aligned}&\min_\pi \sum_{i=1}^{m}\sum_{j=1}^{n} C(q_i^t, q_j^{t-1})\pi_{q_i^t, q_j^{t-1}}, \\ &\text{s.t.} \sum_{i=1}^{m} \pi_{q_i^t, q_j^{t-1}} = s_i, \quad \sum_{j=1}^{n} \pi_{q_i^t, q_j^{t-1}} = d_j, \\ &\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j, \\ &\pi_{q_i^t, q_j^{t-1}} \geq 0, \quad i = 1, 2, ..., m, \quad j = 1, 2, ..., n.\end{aligned} \tag{6}$$

where $C(q_i^t, q_j^{t-1})$ represents the cost matrix; $q_i^t \in Q^t_{\text{selected}}$, $q_j^{t-1} \in Q^{t-1}_{\text{selected}}$ denote query vectors from the current task t and the previous task $t-1$, respectively. $\pi_{q_i^t, q_j^{t-1}}$ represents the optimal matching plan for transferring queries; s_i and d_j are the distribution constraints. m and n represent the number of pruned Queries in model t and model $t-1$, respectively.

The cost matrix $C_{q^t, q^{t-1}}$ is determined by both the IoU and cosine similarity:

$$\begin{aligned}&C(q_i^t, q_j^{t-1}) = \\ &\lambda_1 \cdot \text{IoU}(b_i^t, b_j^{t-1}) + \lambda_2(1 - \cos(q_i^t, q_j^{t-1})),\end{aligned} \tag{7}$$

where b_i^t and b_j^{t-1} are the bounding boxes corresponding to queries q_i^t and q_j^{t-1}, λ_1 and λ_2 are weighting hyperparameters that balance the contribution of IoU and cosine similarity in the cost calculation.

By introducing entropy regularization, the Sinkhorn algorithm enables the OT problem to be solved efficiently through smoothing. The optimization problem then becomes:

$$L_{OT}^{\lambda}(q_i^t, q_j^{t-1}) = \min_{\pi \in U(q_i^t, q_j^{t-1})} \langle \pi, C\rangle - \lambda H(\pi), \tag{8}$$

where $\lambda > 0$ is a regularization coefficient and $H(\pi)$ represents the entropy constraint, defined as:

$$H(\pi) = -\sum_{i,j} \pi_{q_i^t, q_j^{t-1}} \log(\pi_{q_i^t, q_j^{t-1}}). \tag{9}$$

This optimal transport method ensures that the student model learns richer knowledge from the teacher model by aligning query vectors between the target and source, thereby enhancing performance in incremental tasks.

After the optimal transport matching of the query vectors between the target and source models, the loss is computed between the matched query vectors. Let q_i^t and q_i^{t-1} be the query vectors from the teacher and student models for task t and task $t-1$, respectively. The matching loss can be expressed as:

$$\mathcal{L}_{mat} = \sum_{\pi_{q^t, q^{t-1}}} \mathcal{L}_{\text{CE}}(q^t, q^{t-1}), \tag{10}$$

where $\pi_{q^t, q^{t-1}}$ is the optimal transport plan indicating the matching between the teacher and student queries.

3.5. Overall Loss

The training loss of Deformable-DETR mainly consists of two components: the classification loss and the bounding box loss:

$$\mathcal{L}_{det} = \mathcal{L}_{cls} + \mathcal{L}_{bbox}, \tag{11}$$

	HRRSD		LEVIR		DIOR		DOTA		Joint Test		
Methods	mAP	AP_{50}	mAP	AP_{50}	mAP	AP_{50}	mAP	AP_{50}	mAP	AP_{50}	AP_{75}
Joint Training (Faster-RCNN)	72.2	97.3	58.7	87.1	51.6	79.0	42.7	65.6	45.9	73.3	51.5
Fine-tuning (Faster-RCNN)	39.5	70.4	22.1	52.0	26.8	59.5	31.8	62.4	32.3	61.1	29.6
Faster-ILOD [31]	42.2	74.9	23.2	55.0	30.9	62.6	32.6	62.1	32.7	62.2	31.1
ORE [16]	46.1	79.2	23.9	56.9	30.8	63.9	32.6	61.0	31.3	60.3	30.0
Joint Training (Deformable-DETR)	74.3	97.8	61.7	88.6	48.7	77.0	36.3	62.1	44.4	71.2	47.3
Fine-tuning (Deformable-DETR)	36.6	71.5	22.1	50.0	26.8	59.5	32.6	61.0	31.3	60.3	30.0
CL-DETR [28]	47.0	78.0	23.6	52.3	28.6	61.0	35.6	60.8	32.6	61.9	31.0
Incre-DETR [8]	33.8	56.6	24.5	55.6	29.7	62.4	36.9	62.9	33.8	64.0	32.2
DyQ-DETR [43]	52.8	86.0	26.9	59.6	26.8	59.5	36.7	61.5	34.0	63.2	32.8
LDB* [37]	46.5	81.9	24.0	52.3	26.9	60.7	36.8	63.0	32.4	63.3	30.8
ours	**49.5**	**86.9**	**29.1**	**63.9**	**32.6**	**69.5**	**35.9**	**62.8**	**36.2**	**67.8**	**35.0**

Table 1. The table presents the performance comparison of sate-of-the-art methods on four remote sensing datasets. Evaluation metrics include mAP, AP_{50}, and AP_{75}. The Joint Test column reports the overall performance across all datasets, indicating the model's ability to retain knowledge from previously encountered domains. The best results for each metric are highlighted in bold. The domain-bias module of the LDB* method is applied to the DETR detector.

	ALER	ADSC	mAP	AP_{50}	AP_{75}
Sequence I	-	-	31.3	60.3	30.0
	✓	-	34.7	65.1	31.4
	-	✓	33.9	64.0	34.9
	✓	✓	36.2	67.8	35.0
Sequence II	-	-	24.5	58.4	14.4
	✓	-	29.1	61.7	26.6
	-	✓	28.3	60.2	27.9
	✓	✓	32.7	62.8	30.8

Table 2. The ablation study results for the ALER and ADSC modules across two different fine-tuning sequences.

Sequence I	HR	LE	DI	DO	JT
Joint Training	97.3	87.1	79.0	65.6	72.3
Fine-tuning	71.9	52.0	59.6	62.4	63.4
ours	86.9	63.9	69.5	62.8	67.8
Sequence II	LE	HR	DO	DI	JT
Fine-tuning	58.4	79.3	47.4	71.7	58.4
ours	63.4	91.5	51.7	73.9	62.8
Sequence III	DI	DO	HR	LE	JT
Fine-tuning	41.0	33.6	58.4	76.5	36.6
ours	48.8	40.5	74.0	85.0	42.5

Table 3. Performance of our method across different sequences.

The final loss of our method is backpropagated by the matching loss and the original detector loss of Deformable DETR:

$$\mathcal{L}_{total} = \mathcal{L}_{mat} + \mathcal{L}_{det}. \tag{12}$$

4. Experiments

4.1. Datasets and Implement Details

To evaluate the effectiveness of our proposed Active-DDC method, we conducted experiments on four widely used remote sensing object detection datasets: HRRSD [44], LEVIR [50], DIOR [21], and DOTA [40].

To simulate domain-incremental object detection in remote sensing, we applied class pruning across all datasets. Specifically, except for LEVIR, we pruned the categories in the other datasets to three classes: ship, airplane, and oil tank. These datasets, collected from heterogeneous sensors and covering diverse and complex detection scenarios, provide a strong benchmark for evaluating our method's performance in domain-incremental object detection. Additionally, images from the DOTA dataset were cropped into a fixed size of 1024×1024 pixels.

Our proposed method, Active-DDC, is implemented under the MMDetection [5] framework in Python. Unless otherwise stated, all the hyper-parameters follow the default settings of the corresponding student model for both training and testing.

4.2. Results and Analysis

To assess the effectiveness of our proposed method, we conduct a comparative evaluation against several state-of-the-art incremental object detection approaches on four widely used remote sensing datasets: HRRSD, LEVIR, DIOR, and DOTA. The evaluation metrics: mAP, AP_{50}, and AP_{75} to ensure a comprehensive performance assessment.

As shown in Table 1, our method consistently outperforms existing incremental object detection techniques across all datasets. Notably, it achieves the highest mAP

Methods	HR	LE	DI	DO	JT
Joint Training	97.8	88.6	77.0	62.1	71.2
Fine-tuning	71.5	50.0	59.5	61.0	60.3
ER [31]	74.3	54.4	61.2	61.0	61.1
ABR* [27]	70.8	51.6	59.4	61.1	60.3
RPC [43]	78.5	56.1	62.0	61.5	64.7
our ALER	**79.8**	**56.5**	**62.3**	**61.7**	**65.1**

Table 4. Comparison of different Example Replay methods.

Methods	HR	LE	DI	DO	JT
Joint Training	97.8	88.6	77.0	62.1	71.2
Fine-tuning	71.5	50.0	59.5	61.0	60.3
DETRDistill [4]	72.7	51.6	59.0	61.2	61.0
KD-DETR [39]	79.5	**54.1**	60.6	61.9	63.4
our ADSC	**79.8**	53.9	**61.6**	**62.0**	**64.0**

Table 5. Comparison of different Knowledge Distillation methods.

on HRRSD (49.5%), LEVIR (29.1%), DIOR (32.6%), and DOTA (35.9%), demonstrating strong adaptability across diverse domains.

Compared to previous methods, our approach significantly surpasses CL-DETR and Incre-DETR, which also employ knowledge distillation, and outperforms DyQ-DETR and LDB, which rely on adaptor-based techniques. For instance, in the joint evaluation, our method outperforms CL-DETR, Incre-DETR, DyQ-DETR, and LDB by 5.9, 3.8, 4.6, and 4.5 percentage points in terms of AP_{50}, respectively. These results highlight the effectiveness of our ALER and ADSC, which collectively enhance feature transfer mitigation in domain-incremental object detection. In the Joint Test, which evaluates the model's ability to retain knowledge from previously encountered domains, our method achieves the best overall performance. This suggests that our approach effectively alleviates catastrophic forgetting while adapting to new domains.

4.3. Ablation Study

Modules ablation study. As shown in Table 2, the ablation study results for the Active Learning Example Replay (ALER) and Active Domain Shift Control (ADSC) modules are presented. The experiments use mAP, AP_{50}, and AP_{75} as evaluation metrics, conducted under the Sequence I and Sequence II settings.

In both Sequence I and Sequence II, using ALER or ADSC individually improves performance, indicating that both modules are highly adaptable to different data sequences. The best performance is achieved when ALER and ADSC are combined in both fine-tuning sequences. In Sequence I, mAP increases from 31.3% to 36.2%, AP_{50} improves by 7.5%, and AP_{75} increases by 5.0%; in Sequence II, mAP increases from 24.5% to 32.7%, AP_{50} improves by 4.4%, and AP_{75} increases by 16.4%. These results demonstrate that our method exhibits strong robustness in domain incremental object detection, effectively mitigating catastrophic forgetting, and consistently improving performance across different task sequences.

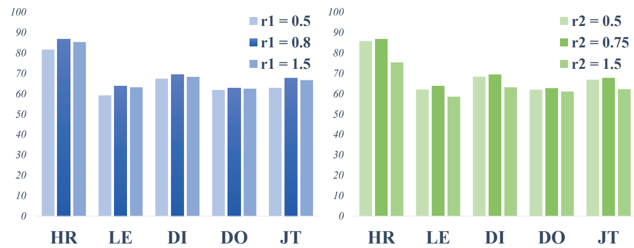


Figure 3. The Impact of Domain Increment in Remote Sensing Images on Data Distribution and Model Feature Shifts, and the Balancing Effect of Our Approach on These Impacts.

Sequences ablation study. As shown in Table 3, our method consistently surpasses Fine-tuning across all sequences and datasets. It enhances individual dataset performance while achieving better overall results in joint tests. The improvements are especially notable in sequences with complex domain shifts, demonstrating the robustness of our approach. However, detection performance remains sensitive to sequence order, influenced by domain difficulty and differences.

ER ablation study. The ER method, a common approach in incremental learning, has many variants. Compared to previous ER methods, our ALER method integrates an informative evaluation strategy based on active learning, which combines the risk and difficulty of remote sensing samples for sampling. This makes our method more advantageous for remote sensing image data with domain shifts. We compare our approach with the ABR [27] method using mixup and mosaic augmentation strategies, as well as the Risk-balanced Partial Calibration (RPC) [43] method. As shown in Table 3, our method outperforms both the ABR and RPC methods across four task datasets and joint tests, demonstrating the superiority of ALER for this task.

KD ablation study. Similarly, knowledge distillation (KD) has been used to mitigate catastrophic forgetting. However, traditional KD methods, which focus on logit or pseudo-label distillation in DETR-based detectors, struggle with feature shifts in domain-incremental object detection. Since queries are unique to DETR-based detectors, query-based KD is gaining attention. KD-DETR [39] introduces dedicated object queries for distillation, using a progressive sampling strategy. DETRDistill [4] constructs soft masks for feature-level distillation by computing the similarity between query embeddings and feature representations. Table 5 compares the performance of DETRDistill, KD-DETR, and our ADSC method in domain-incremental tasks.

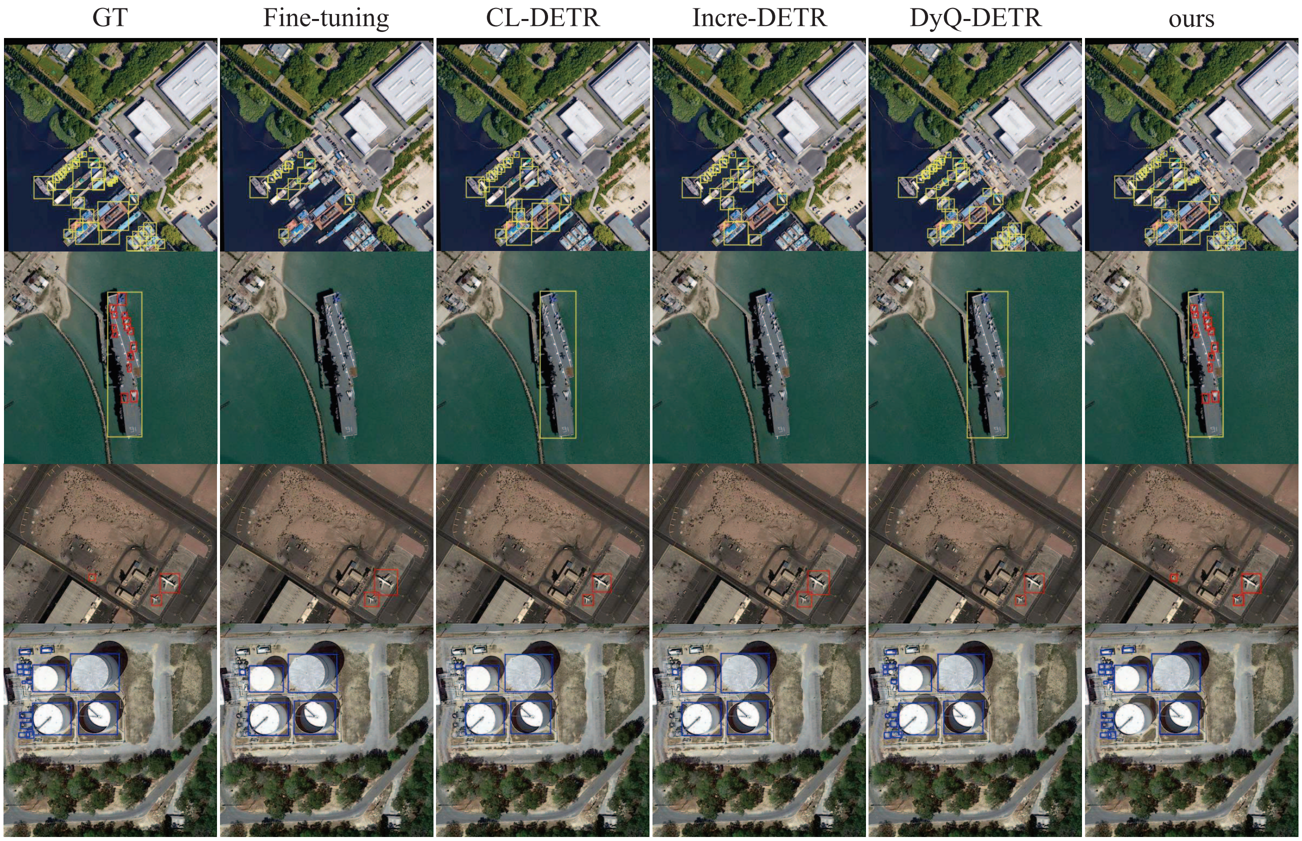


Figure 4. Visual detection results of images. The Ground truth, Fine-tuning, CL-DETR, Incre-DETR, DyQ-DETR and ours method are compared respectively. All methods are sequentially trained and evaluated in the joint test.

4.4. Sensitivity to Hyperparameters

As shown in fig. 3, we conducted a parameter sensitivity analysis. In Eq.(4) and Eq.(7), α and β are the information weight and risk weight. $r_1 = \alpha : \beta$ r_1 represents the ratio between informativeness and risk. A lower r_1 indicates a preference for selecting samples with higher risk but lower informativeness, whereas a higher r_1 favors samples with greater informativeness and lower risk. $r_2 = \lambda_1 : \lambda_2$ represents the ratio between IoU weight and cosine similarity weight. A lower r_2 indicates a preference for selecting samples with higher IoU overlap but lower similarity. Conversely, a higher r_2 favors samples with greater similarity but lower overlap confidence.

4.5. Visualization Analysis

As illustrated in fig. 4, we conducted a series of visualization experiments on detection results across different domains. The visualizations clearly highlight the effectiveness of our proposed method in accurately identifying objects under varying domain conditions. Notably, our approach not only maintains high detection precision in newly introduced domains but also effectively mitigates catastrophic forgetting by retaining critical knowledge from previously learned domains. This demonstrates the model's strong ability to generalize and adapt, ensuring stable, reliable, and consistent performance in domain-incremental object detection scenarios.

5. Conclusion

We propose Active-DDC, a novel framework for domain-incremental object detection in remote sensing images. To address domain shifts and catastrophic forgetting, Active-DDC incorporates two key components: ALER and ADSC. ALER uses active learning to select and store representative samples, preserving essential knowledge from past domains. ADSC enhances cross-domain knowledge transfer by using query-based optimal transport to align features. Together, these modules improve model adaptation across various remote sensing datasets. Experimental results show that Active-DDC outperforms existing methods, achieving state-of-the-art performance. Ablation studies confirm the effectiveness of ALER and ADSC in reducing distribution shifts and enhancing model stability.

6. Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62372348, 62176198, U22A2096 and 62036007, in part by the Key Research and Development Program of Shaanxi under Grant 2024GXZDCYL-02-10, in part by the Key Laboratory of Big Data Intelligent Computing under Grant BDIC-2023-A-004, in part by Shaanxi Outstanding Youth Science Fund Project under Grant 2023-JC-JQ-53, in part by Scientific and Technological Innovation Teams in Shaanxi Province under Grant 2025RS-CXTD-011, in part by the Shaanxi Province Core Technology Research and Development Project under Grant 2024QY2-GJHX-11, in part by the Fundamental Research Funds for the Central Universities under Grant QTZX23042.

References

[1] Jihwan Bang, Heesu Kim, YoungJoon Yoo, Jung-Woo Ha, and Jonghyun Choi. Rainbow memory: Continual learning with a memory of diverse samples. In *CVPR*, pages 8218–8227, 2021. 2

[2] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *ECCV*, pages 213–229, 2020. 2

[3] Fabio Cermelli, Antonino Geraci, Dario Fontanel, and Barbara Caputo. Modeling missing annotations for incremental learning in object detection. In *CVPR*, pages 3700–3710, 2022. 2

[4] Jiahao Chang, Shuo Wang, Hai-Ming Xu, Zehui Chen, Chenhongyi Yang, and Feng Zhao. Detrdistill: A universal knowledge distillation framework for detr-families. In *ICCV*, pages 6898–6908, 2023. 2, 7

[5] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, et al. Mmdetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019. 6

[6] Zhigang Dai, Bolun Cai, Yugeng Lin, and Junying Chen. Up-detr: Unsupervised pre-training for object detection with transformers. In *CVPR*, pages 1601–1610, 2021. 2, 3

[7] Na Dong, Yongqiang Zhang, Mingli Ding, and Gim Hee Lee. Bridging non co-occurrence with unlabeled in-the-wild data for incremental object detection. *NeurIPS*, pages 30492–30503, 2021. 2

[8] Na Dong, Yongqiang Zhang, Mingli Ding, and Gim Hee Lee. Incremental-detr: Incremental few-shot object detection via self-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 543–551, 2023. 2, 6

[9] Tao Feng, Mang Wang, and Hangjie Yuan. Overcoming catastrophic forgetting in incremental object detection via elastic response distillation. In *CVPR*, pages 9427–9436, 2022. 2

[10] Caili Gao, Qisheng Xu, Peng Qiao, Kele Xu, Xifu Qian, and Yong Dou. Adapter-based incremental learning for face forgery detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4690–4694, 2024. 2

[11] Ross Girshick. Fast r-cnn. In *ICCV*, pages 1440–1448, 2015. 2

[12] Akshita Gupta, Sanath Narayan, KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. Ow-detr: Open-world detection transformer. In *CVPR*, pages 9235–9244, 2022. 2

[13] Yu Hao, Yanwei Fu, Yu-Gang Jiang, and Qi Tian. An end-to-end architecture for class-incremental object detection with knowledge distillation. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2019. 2

[14] Yu Hao, Yanwei Fu, Yu-Gang Jiang, and Qi Tian. An end-to-end architecture for class-incremental object detection with knowledge distillation. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2019. 2

[15] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. 2

[16] KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In *CVPR*, pages 5830–5840, 2021. 2, 6

[17] KJ Joseph, Jathushan Rajasegaran, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Incremental object detection via meta-learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 9209–9216, 2021. 2

[18] Harold W Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, pages 83–97, 1955. 3

[19] Dawei Li, Serafettin Tasci, Shalini Ghosh, Jingwen Zhu, Junting Zhang, and Larry Heck. Rilod: Near real-time incremental learning for object detection at the edge. In *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, pages 113–126, 2019. 2

[20] Feng Li, Hao Zhang, Shilong Liu, Jian Guo, Lionel M Ni, and Lei Zhang. Dn-detr: Accelerate detr training by introducing query denoising. In *CVPR*, pages 13619–13627, 2022. 3

[21] Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, pages 296–307, 2020. 6

[22] Xiang Li, Wenhai Wang, Lijun Wu, Shuo Chen, Xiaolin Hu, Jun Li, Jinhui Tang, and Jian Yang. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *NeurIPS*, 33:21002–21012, 2020. 2

[23] Xiang Li, Wenhai Wang, Lijun Wu, Shuo Chen, Xiaolin Hu, Jun Li, Jinhui Tang, and Jian Yang. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *NeurIPS*, pages 21002–21012, 2020. 2

[24] TY Lin, P Goyal, R Girshick, K He, and P Dollar. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 318–327, 2018. 2

[25] Shilong Liu, Feng Li, Hao Zhang, Xiao Yang, Xianbiao Qi, Hang Su, Jun Zhu, and Lei Zhang. Dab-detr: Dynamic anchor boxes are better queries for detr. *arXiv preprint arXiv:2201.12329*, 2022. 3

[26] Xialei Liu, Hao Yang, Avinash Ravichandran, Rahul Bhotika, and Stefano Soatto. Multi-task incremental learning for object detection. *arXiv preprint arXiv:2002.05347*, 2020. 2

[27] Yuyang Liu, Yang Cong, Dipam Goswami, Xialei Liu, and Joost Van De Weijer. Augmented box replay: Overcoming foreground shift for incremental object detection. In *ICCV*, pages 11367–11377, 2023. 2, 7

[28] Yaoyao Liu, Bernt Schiele, Andrea Vedaldi, and Christian Rupprecht. Continual detection transformer for incremental object detection. In *CVPR*, pages 23799–23808, 2023. 2, 6

[29] Yi Liu, Luting Wang, Zongheng Tang, Yue Liao, Yifan Sun, Lijun Zhang, and Si Liu. Knowledge distillation via query selection for detection transformer. *arXiv preprint arXiv:2409.06443*, 2024. 2

[30] Depu Meng, Xiaokang Chen, Zejia Fan, Gang Zeng, Houqiang Li, Yuhui Yuan, Lei Sun, and Jingdong Wang. Conditional detr for fast training convergence. In *ICCV*, pages 3651–3660, 2021. 3

[31] Can Peng, Kun Zhao, and Brian C Lovell. Faster ilod: Incremental learning for object detectors based on faster rcnn. *Pattern Recognition Letters*, pages 109–115, 2020. 2, 6, 7

[32] Can Peng, Kun Zhao, Sam Maksoud, Meng Li, and Brian C Lovell. Sid: Incremental learning for anchor-free object detection via selective and inter-related distillation. *Computer vision and Image Understanding*, page 103229, 2021. 2

[33] Ameya Prabhu, Philip HS Torr, and Puneet K Dokania. Gdumb: A simple approach that questions our progress in continual learning. In *ECCV*, pages 524–540, 2020. 2

[34] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1137–1149, 2016. 2

[35] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. *NeurIPS*, 30, 2017. 2

[36] Konstantin Shmelkov, Cordelia Schmid, and Karteek Alahari. Incremental learning of object detectors without catastrophic forgetting. In *ICCV*, pages 3400–3409, 2017. 2

[37] Xiang Song, Yuhang He, Songlin Dong, and Yihong Gong. Non-exemplar domain incremental object detection via learning domain bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 15056–15065, 2024. 2, 6

[38] Qiang Wang, Yuhang He, Songlin Dong, Xinyuan Gao, Shaokun Wang, and Yihong Gong. Non-exemplar domain incremental learning via cross-domain concept integration. In *ECCV*, pages 144–162, 2024. 2

[39] Yu Wang, Xin Li, Shengzhao Weng, Gang Zhang, Haixiao Yue, Haocheng Feng, Junyu Han, and Errui Ding. Kd-detr: Knowledge distillation for detection transformer with consistent distillation points sampling. In *CVPR*, pages 16016–16025, 2024. 2, 7

[40] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *CVPR*, pages 3974–3983, 2018. 6

[41] Jayeon Yoo, Dongkwan Lee, Inseop Chung, Donghyun Kim, and Nojun Kwak. What how and when should object detectors update in continually changing test domains? In *CVPR*, pages 23354–23363, 2024. 2

[42] Junting Zhang, Jie Zhang, Shalini Ghosh, Dawei Li, Serafettin Tasci, Larry Heck, Heming Zhang, and C-C Jay Kuo. Class-incremental learning via deep model consolidation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1131–1140, 2020. 2

[43] Jichuan Zhang, Wei Li, Shuang Cheng, Ya-Li Li, and Shengjin Wang. Dynamic object queries for transformer-based incremental object detection. *arXiv preprint arXiv:2407.21687*, 2024. 2, 4, 6, 7

[44] Yuanlin Zhang, Yuan Yuan, Yachuang Feng, and Xiaoqiang Lu. Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection. *IEEE Transactions on Geoscience and Remote Sensing*, pages 5535–5548, 2019. 6

[45] Wang Zhou, Shiyu Chang, Norma Sosa, Hendrik Hamann, and David Cox. Lifelong object detection. *arXiv preprint arXiv:2009.01129*, 2020. 2

[46] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. *arXiv e-prints*, pages arXiv–1904, 2019. 2

[47] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. *arXiv preprint arXiv:1904.07850*, 2019. 2

[48] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020. 2, 3

[49] Zhuofan Zong, Guanglu Song, and Yu Liu. Detrs with collaborative hybrid assignments training. In *ICCV*, pages 6748–6758. 3

[50] Zhengxia Zou and Zhenwei Shi. Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images. *IEEE Transactions on Image Processing*, pages 1100–1111, 2018. 6