Journal of Intelligent & Robotic Systems (2026) 112:45
https://doi.org/10.1007/s10846-026-02380-2

RESEARCH

Adaptive Active Learning with Dynamic Uncertainty-Diversity Balancing for Object Detection

Joonsun Auh[1,2] · Changsik Cho[1] · Seon-tae Kim[1,2]

Received: 18 November 2025 / Accepted: 26 February 2026

Abstract

Recent advances in deep learning have positioned object detection as a critical technology across various domains, including autonomous driving, video surveillance, and robotics. Despite their high accuracy, object detection models rely heavily on large, meticulously annotated datasets. However, annotating data for object detection is significantly more time-consuming and expensive than for standard image classification tasks, creating substantial barriers in both research and industry. To address this challenge, we propose an adaptive active learning framework aimed at reducing annotation costs without compromising model performance. Although active learning is known for its ability to minimize labeling efforts, applying it to object detection presents unique challenges owing to the need to account for both uncertainty and diversity. Our approach estimates uncertainty using object confidence scores and quantifies diversity based on the number of classes per image across the unlabeled dataset. Moreover, our framework dynamically adjusts the weighting between uncertainty and diversity throughout training. Experiments on a Unmanned Aerial Vehicle (UAV) dataset and a real-world industrial dataset involving high-voltage electrical cables demonstrated performance improvements of 1.1% and approximately 10%, respectively, under the same annotation budget. These results demonstrate the potential of our framework to significantly lower annotation costs while maintaining high detection performance, rendering it well-suited for real-world industrial applications.

Keywords Active learning · Uncertainty · Diversity · Object detection

1 Introduction

Recent advancements in artificial intelligence have made computer vision a pivotal technology in various industries and everyday applications. Among its many capabilities, object detection, which automatically identifies the location and category of objects within images, has become essential across numerous applications [1–3], including autonomous driving [4, 5], intelligent surveillance systems [6, 7], robotics [8–10] and augmented reality (AR) [11, 12].

However, achieving high performance with deep learning–based object detection models requires large, accurately labeled datasets. In contrast to image classification, which assigns a single label to an entire image, object detection involves identifying and annotating multiple objects along with their precise locations. This increases annotation complexity, requiring more time, expertise, and cost, particularly for real-world datasets that often feature cluttered scenes, occlusions, and a wide range of object types. To overcome these challenges, active learning has gained attention as a promising approach to improve model performance with reduced labeling effort [13–16]. Active learning enables the model to select the most informative samples for annotation, thereby optimizing performance under limited resources. Approaches typically fall into three categories: uncertainty-based, diversity-based, and hybrid methods. Uncertainty-based methods prioritize samples the model is least confident about, with uncertainty often measured using confidence

✉ Seon-tae Kim
stkim10@etri.re.kr

Joonsun Auh
jsauh@etri.re.kr

Changsik Cho
cscho@etri.re.kr

1 On-Device Artificial Intelligence Research Division, Electronics and Telecommunications Research Institute (ETRI), 218, Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea

2 Department of Artificial Intelligence, University of Science and Technology (UST), 217, Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea

Published online: 26 March 2026

scores in object detection. Diversity-based methods focus on selecting varied samples that represent different data distributions or classes, thereby improving generalization. However, each of these methods has limitations when used independently: uncertainty-based strategies may select only ambiguous samples and overlook diversity, while diversity-based strategies might exclude highly informative, uncertain samples. Hence, hybrid methods have been proposed to combine the strengths of both approaches. However, most existing hybrid strategies use a fixed ratio to balance uncertainty and diversity, failing to adapt to changes in data distribution or model learning state throughout training. As both the dataset characteristics and model behavior evolve during training, static sampling strategies may be suboptimal. Despite the importance of adaptability, few studies have explored dynamic sampling approaches for active learning in object detection. In this study, we propose a novel adaptive data sampling framework for object detection based on active learning using the YOLOv7 model [17]. The model detects objects in an unlabeled pool, assigns an *uncertainty score* using object confidence, and calculates diversity by counting the number of object classes in each image. Our method dynamically adjusts the weighting between uncertainty and diversity during training, prioritizing uncertain samples in the early stages and gradually shifting focus toward more representative, diverse samples as training progresses.

The contributions of this paper are as follows:

1. **Active learning with a dynamic uncertainty–diversity ratio**. In contrast to prior work that relies on fixed ratios or a single selection strategy, our approach adaptively balances uncertainty and diversity based on both the data distribution and the training stage. This dynamic adjustment improves model efficiency and accuracy under constrained annotation budgets.
2. **Object detection integrated with active learning.** This study applies active learning to object detection tasks using the YOLOv7 model. Given the high labeling costs associated with object detection, we developed an active learning framework designed to improve model performance efficiently under a limited annotation budget. Specifically, uncertainty and diversity were quantified based on the confidence scores and the number of object classes within each image obtained through YOLOv7, maximizing the efficiency of data selection.
3. **Experimental validation using UAV and industrial data.** In addition to the commonly used UAV datasets, this study used industrial datasets for the experiment. The results confirmed that the proposed adaptive active learning method effectively detects objects for real-world data and demonstrates their applicability.

2 Related Work

Active learning improves the performance of models under limited annotation budgets by selecting the most useful data during training [18, 19]. As mentioned earlier, it is typically categorized into uncertainty-based, diversity-based, and hybrid methods.

2.1 Uncertainty-Based Method

In uncertainty-based methods [20–22], uncertainty plays a critical role in sample selection. Uncertainty refers to the model's doubt about its predictions, while confidence indicates the estimated likelihood that a prediction is correct. These concepts are inversely related, lower confidence corresponds to higher uncertainty. The key idea is that labeling uncertain samples improves model performance effectively. One common approach is least confidence sampling [23, 24], which selects samples with the lowest confidence. Margin sampling [25, 26] considers the difference between the top two predicted class probabilities; a smaller margin indicates greater uncertainty. Entropy-based sampling [27, 28] selects samples with the highest entropy, assuming that they will benefit the model most. Although uncertainty-based methods effectively identify ambiguous samples near the decision boundary, they often ignore diversity (Fig. 1). As these methods focus only on prediction uncertainty, they may repeatedly select similar samples, causing redundancy. Labeling redundant samples leads to diminishing returns and inefficient use of annotation resources.

2.2 Diversity-Based Method

Diversity-based methods aim to select samples that represent the overall data distribution. The objective is to train models on diverse data rather than similar examples. By selecting broadly representative samples, models better capture the underlying data distribution, improving generalization and reducing overfitting. Two main approaches exist: distance-based and clustering-based sampling. Distance-based methods select samples far from the labeled set, as these likely represent unexplored regions. For example, *Deep Active Learning over the Long Tail* [29] uses the Farthest-First Traversal strategy, selecting the sample farthest from the existing labeled set iteratively.

Clustering-based methods group unlabeled data into clusters and select representative samples from each, assuming that samples within clusters share similar properties. Known examples include Core-set selection [30] and *Batch Active Learning by Diverse Gradient Embeddings* (BADGE) [31], which selects diverse batches using gradient-based embed-

Fig. 1 Concept behind uncertainty-based methods. Uncertainty-based methods are active learning strategies that prioritize unlabeled instances near the decision boundary of the model

dings and k-means++ clustering [32]. Although diversity-based methods effectively cover different regions of the data, they may not always select the most informative samples, limiting model improvement (Fig. 2).

2.3 Hybrid Method

The hybrid method combines uncertainty-based and diversity-based approaches to mitigate their individual limitations.

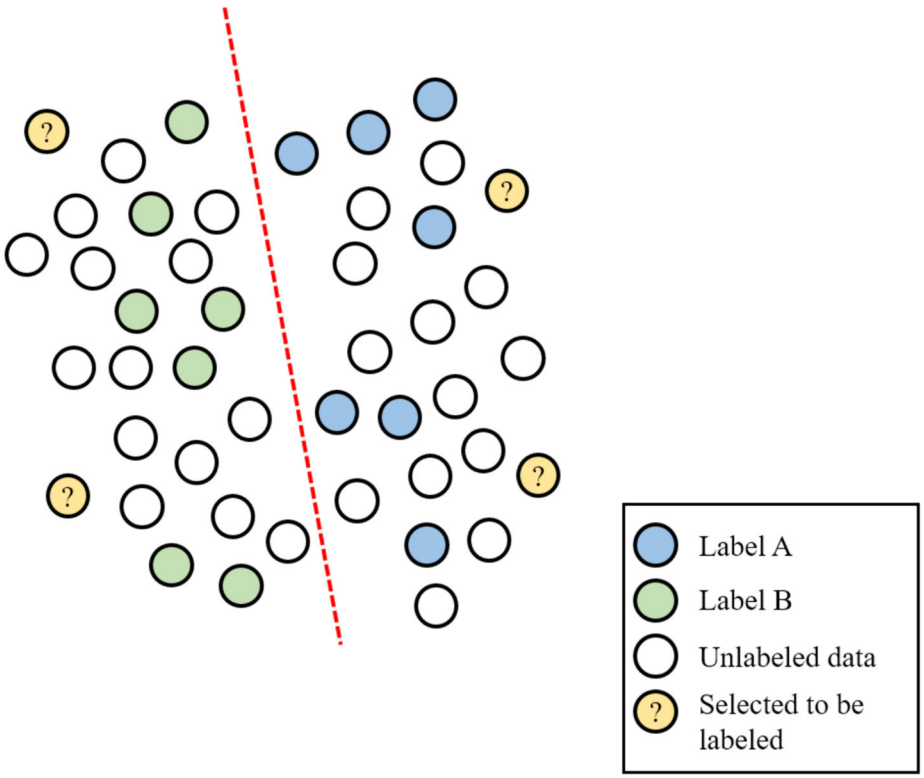

Fig. 2 Concept behind diversity-based methods. Selected samples should be maximally dissimilar from the existing training data and from one another to improve data coverage

Uncertainty-based sampling focuses on ambiguous samples near the decision boundary but can lead to redundancy and inefficient use of labeling budgets. Conversely, diversity-based sampling improves data coverage but may include less informative samples. By merging both, hybrid active learning selects samples that are both uncertain and diverse, reducing redundancy and improving generalization. This renders hybrid methods particularly effective for complex tasks, such as, object detection, where both uncertainty and diversity are important.

Diverse Uncertainty Aggregation (DUA) [33] (Fig. 3) was proposed for single-stage object detection on UAV images. It addresses class imbalance by employing class-aware uncertainty aggregation instead of relying solely on model confidence. This approach assigns greater weight to low-performing classes. Based on evaluation using YOLOv7, it prioritizes training on classes with higher uncertainty (i.e., lower confidence). Through this weighting mechanism, DUA attempts to combine uncertainty and diversity in its model.

3 Proposed Method

3.1 Problem Definition

DUA evaluates 10% of the annotation budget using YOLOv7's test and assigns weights to the lowest-performing classes based on the results. However, some classes that generally perform well may show lower performance in certain images owing to challenging objects.

If these isolated cases receive higher weights, lower-performing classes might not receive sufficient attention. For example, in Fig. 4(a), the model shows high confidence for cars but low confidence for motorcycles. In contrast, Fig. 4(b) shows reduced car confidence owing to challenging objects. Using average confidence to determine weights, as in DUA, may incorrectly assign higher weights to cars, even if motorcycle detection generally performs worse. Moreover, DUA sacrifices part of the annotation budget for evaluations at each iteration, resulting in a loss of labeling resources, which is particularly problematic when budgets are limited. Therefore, the goal of the proposed method is to create a hybrid active learning framework that explicitly measures diversity without sacrificing the annotation budget.

3.2 Adaptive Mixed Active Learning

This study proposes a hybrid active learning framework to optimize sample selection efficiency in object detection. The proposed method combines uncertainty-based and diversity-based strategies, improving model performance.

The overview of the proposed method is shown in Fig. 5. In contrast to DUA, this method does not sacrifice annotation budget for evaluation, thus maximizing data efficiency. It estimates uncertainty using confidence scores from YOLOv7 directly, eliminating the separate evaluation step used by DUA. DUA uses YOLOv7's test for evaluation. It selects uncertain, low-performing samples based on the evaluation results and assigns weights accordingly. The proposed method, however, employs YOLOv7's detect directly on the unlabeled data pool. In contrast to YOLOv7's test, which requires ground truth for evaluation, YOLOv7's detect does not need ground truth. This enables direct use of YOLOv7's detect results for uncertainty estimation, preserving annotation resources for model training.

The proposed method begins by detecting objects using YOLOv7's detect on the entire unlabeled dataset. Based on the detection results, referred to as *detect label*, both uncertainty and diversity are calculated for each image. Using *detect label*, uncertainty and diversity scores for each image are derived. The uncertainty of an object is defined as follows:

$$uncertainty_{object} = 1 - confidence_{object}, \quad (1)$$

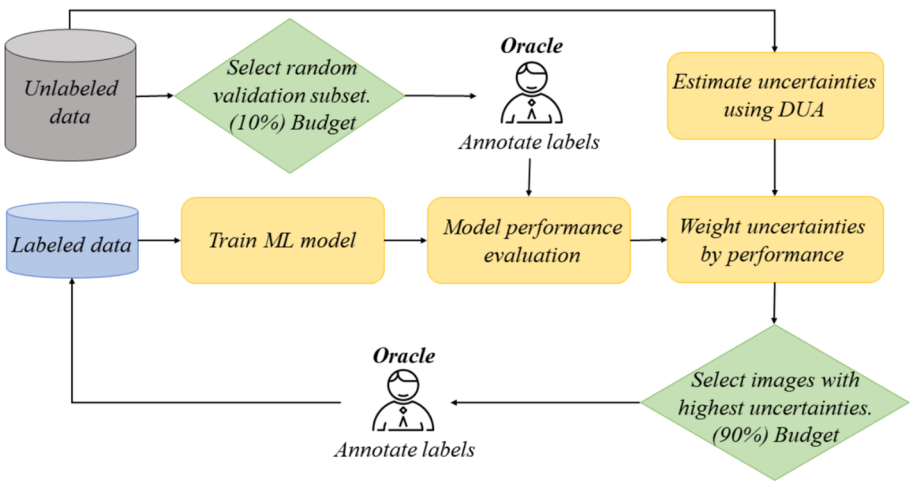

Fig. 3 Process of the Diverse Uncertainty Aggregation (DUA) model. The DUA evaluates 10% of the annotation budget with YOLOv7's test, and weights the classes based on the evaluation results

(a) (b)

Fig. 4 Test results of DUA. The same model was used for testing. The model detect cars more accurately than motorcycles. However, the uncertainty for the car class can become higher with a few hard samples

where $confidence_{object} \in \mathbb{R}_0^+$. Equation (1) evaluates uncertainty. Low confidence indicates high uncertainty about the object. Image uncertainty is calculated by aggregating confidence scores of all detected objects in an image, as follows:

$$image_{uncertainty} = \sum_{1}^{n} Uncertainty_{object}, \tag{2}$$

where $Image_{uncertainty} \in \mathbb{R}_0^+$, n denotes the number of objects in the image. Diversity is defined as the number of distinct object classes within an image; thus, more classes imply higher diversity. This definition aims to help the model learn various visual features and data distributions. Diversity can be calculated as follows:

$$D_i = |C(I_i)|, \tag{3}$$

where I_i denotes an i_{th} image, $C(I_i)$ denotes the set of different object classes in I_i, $|\cdot|$ denotes the cardinality of the set. Subsequently, the number of samples initially targeted is selected in order of descending uncertainty and diversity. The ratio between uncertainty and diversity depends on an *uncertainty score*, dynamically changing during training.

AdaMix-AL was inspired by the relative importance of uncertainty and diversity. Based on the *detect label*, an *uncertainty score* is evaluated as shown in Fig. 6. The *uncertainty score*, deciding the sampling ratio throughout training, is calculated as follows:

$$\delta = f(\alpha, \beta) = (\alpha - \beta) \times (100/\beta), \tag{4}$$

where $\delta \in \mathbb{R}_0^+$ denotes the *uncertainty score*. $\alpha, \beta \in \mathbb{R}_0^+$ denote the 60th and 80th values, respectively, in the list ordered by uncertainty. The rationale for selecting the 60th and 80th values is illustrated in Fig. 7. Although the distribution range is initially large, there are many outliers. Furthermore, as the training iterations progress, the distribution becomes compressed. The 60th and 80th values

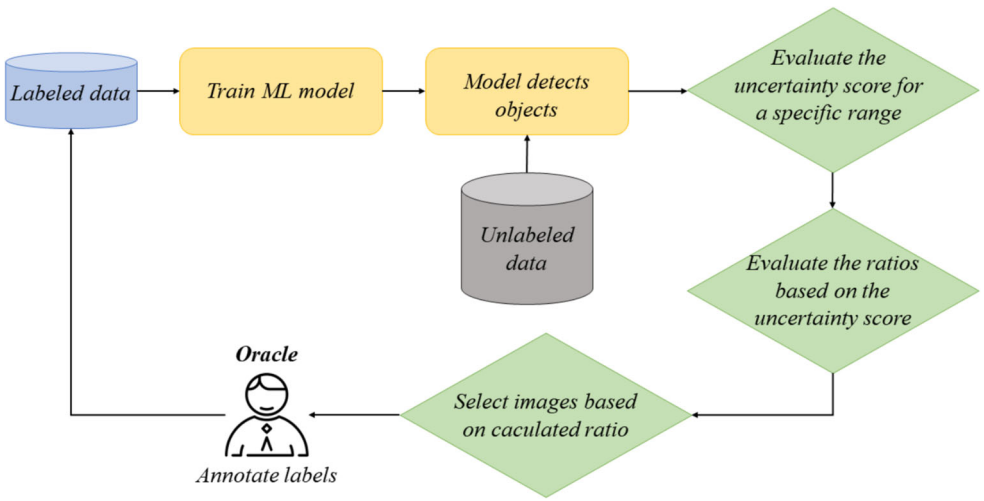

Fig. 5 Overview of the AdaMix-AL. In contrast to DUA, there is no 10% annotation budget sacrifice, and data is selected dynamically at a rate after evaluating *uncertainty score* based on inference results

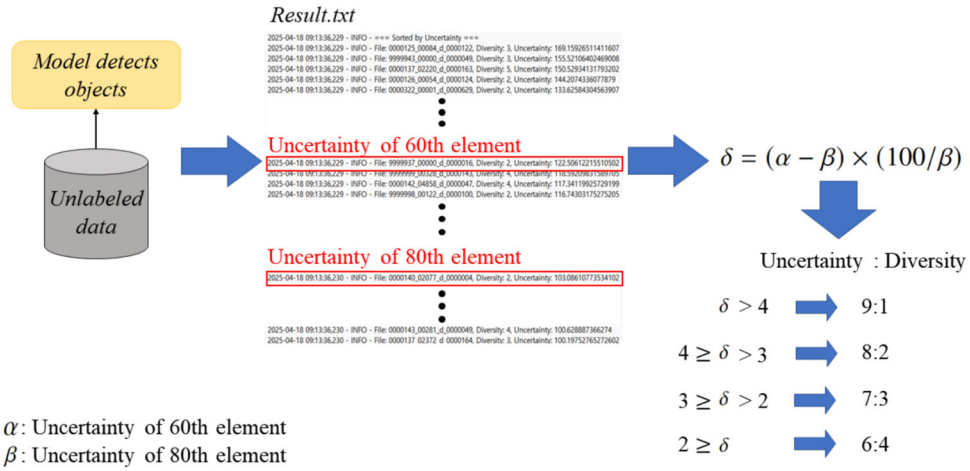

Fig. 6 Process of evaluating *uncertainty score*. The ratio of uncertainty and diversity changes depending on the value of the *uncertainty score* δ. The figure illustrates the process when the annotation budget is 100

were selected because this range mitigates the influence of extreme outliers while still effectively representing the gradient of the compressed distribution. In Eq. 4, the initial term $(\alpha - \beta)$ represents the absolute difference in uncertainty between two positions within the ranked list. A larger difference indicates a sharp gradient in *uncertainty score* across the selection range, signifying that uncertainty remains a critical criterion for annotation. Conversely, a smaller difference suggests a more uniform distribution of uncertainty, indicating limited additional benefit from selecting samples based solely on uncertainty. In such cases, integrating other criteria, such as diversity, becomes essential. The second term, $(100/\beta)$, serves as a normalization factor, ensuring the *uncertainty score* is consistently interpretable across datasets with varying uncertainty ranges. This normalization enables comparability of *uncertainty score* regardless of inherent differences between datasets. In practice, the two datasets analyzed, UAV and WEDA, exhibited significantly different uncertainty distributions. The WEDA dataset comprises real high-voltage electrical cable industry data that we collected, with details provided in Section 4.1.2. Specifically, as shown in Fig. 8, the maximum uncertainty in the UAV dataset ranged from 170 to 45, whereas in the WEDA dataset (Fig. 9), it varied between 3.5 and 0.4. Despite these variations, the proposed method successfully maintained consistent evaluation of the *uncertainty score* across both datasets.

The computed δ values are classified into four ranges: $\delta > 4$, $4 \geq \delta > 3$, $3 \geq \delta > 2$, and $\delta \leq 2$. When δ exceeds 4, the sampling ratio is assigned as 90% uncertainty and 10% diversity. For δ values from 4 down to 3, the proportion changes to 80% uncertainty and 20% diversity. Between 3 and 2, the ratio adjusts to 70% uncertainty and 30% diversity, and for values below 2, it shifts further to 60% uncertainty and 40% diversity. By dynamically

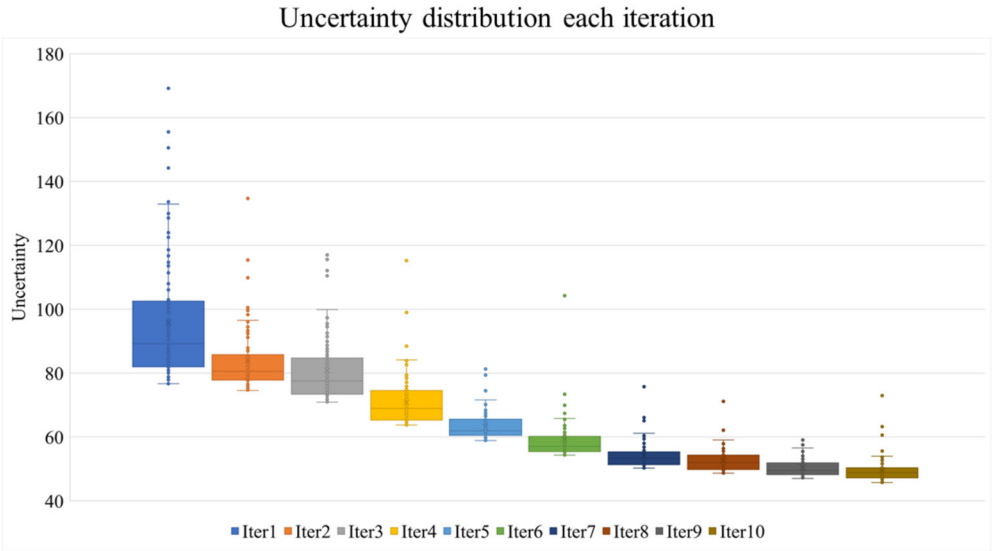

Fig. 7 Uncertainty distribution of UAV dataset. Each color represents an iteration

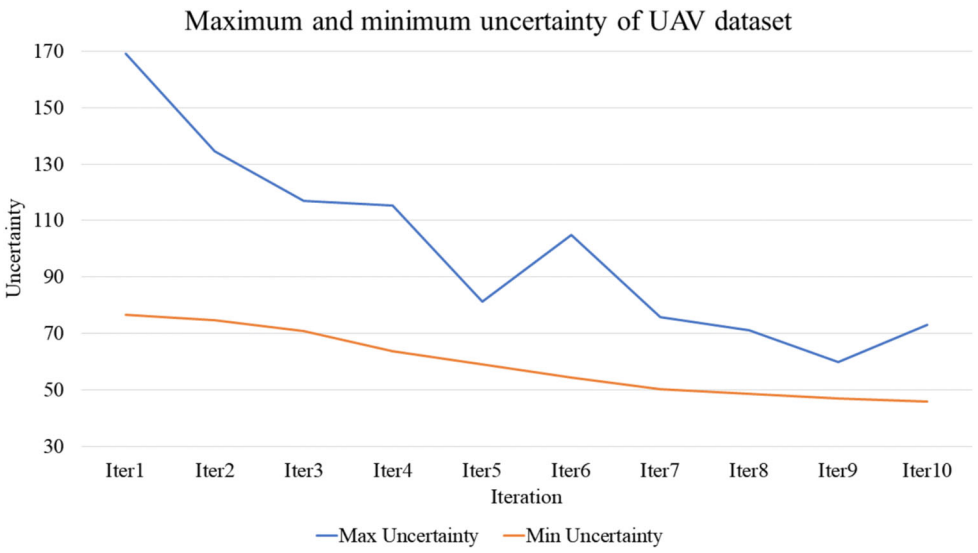

Fig. 8 Maximum and minimum uncertainty of UAV dataset. The maximum and minimum uncertainty of UAV dataset of the list of descending order of uncertainty at each iteration

adjusting these ratios, AdaMix-AL adapts effectively to the evolving data conditions throughout the active learning process, optimizing annotation resources and improving model accuracy simultaneously.

4 Experiment

4.1 Dataset

4.1.1 VisDrone-2019 (UAV)

The VisDrone-2019 dataset [34] is a comprehensive benchmark designed to evaluate object detection and tracking in aerial imagery, primarily captured by unmanned aerial vehicles (UAVs). Developed by the AISKYEYE research group at Tianjin University, it enables research on practical visual understanding from drone perspectives. Figure 10 shows a sample illustration from the dataset. It covers diverse urban and rural environments, captured from various altitudes and angles under different weather and lighting conditions, providing realistic scenarios for assessing computer vision techniques. The dataset includes 10,209 static images and 288 video sequences, totaling over 260,000 annotated frames. Each image or frame contains labeled bounding boxes for ten object categories: pedestrian, person, car, bus, van, truck, motorcycle, bicycle, awning-tricycle, and tricycle. Data were collected from 14 Chinese cities, ensuring variation in urban layouts, road structures, crowd densities, and object sizes. Images and videos were captured using multiple drone platforms at different heights and speeds, introducing variability in spatial resolution and motion characteristics. The dataset is divided into training, validation, and testing subsets to support fair benchmarking of detection and track-

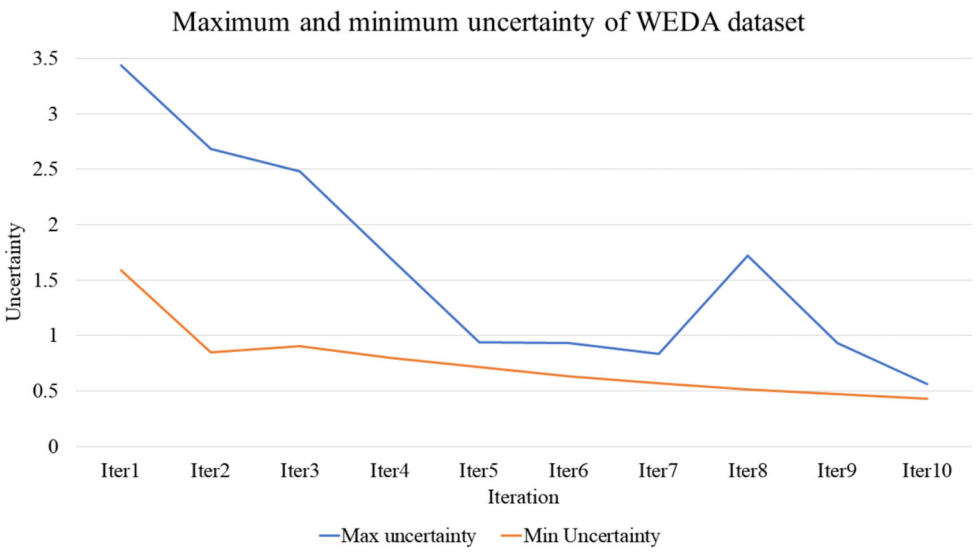

Fig. 9 Maximum and minimum uncertainty of WEDA dataset. The maximum and minimum uncertainty of WEDA dataset of the list of descending order of uncertainty at each iteration

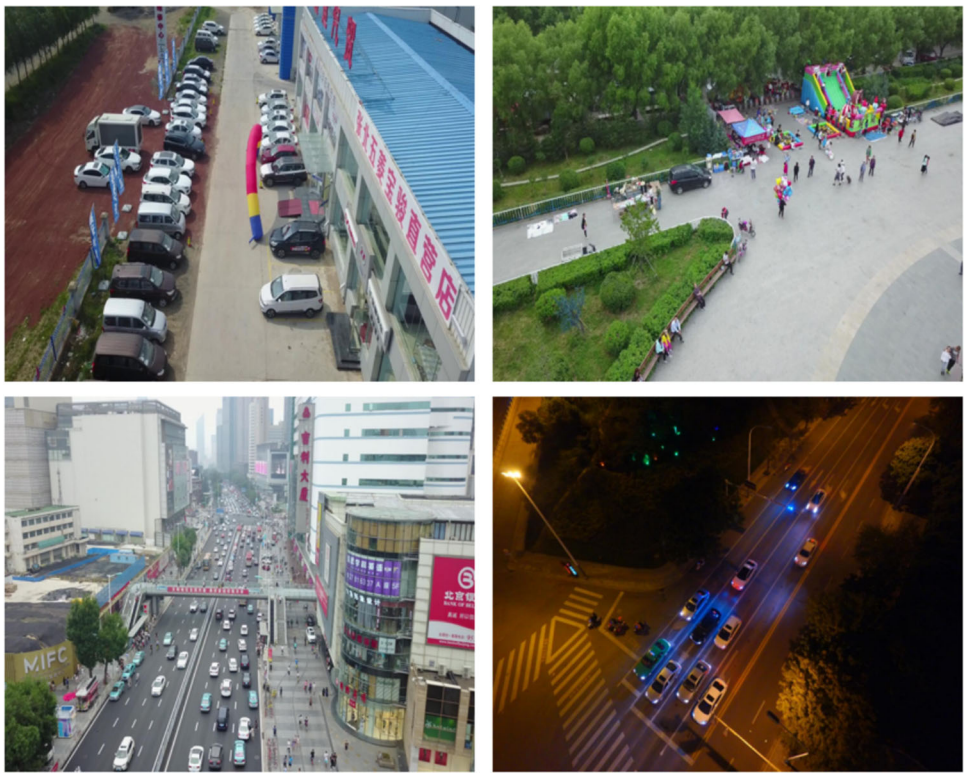

Fig. 10 Sample images of the VisDrone-2019 dataset

ing algorithms. Widely adopted in computer vision research, VisDrone-2019 has been the official dataset for several international challenges. In this study, it is used to evaluate the proposed methodology, as its detailed annotations, complexity, and diversity render it highly suitable for benchmarking sample selection algorithms.

4.1.2 High-Voltage Electrical Cable (WEDA)

The high-voltage electrical cable dataset, known as the WEDA dataset, is an industrial dataset created by WEDA Corporation using real-world data. It includes four defect classes: PO, UNDER_CUT, LF, and IP. PO denotes porosity, which is gas bubbles or voids in the insulation or conductive layers. UNDER_CUT denotes depressions or melted edges along cable joints. LF denotes Lack of Fusion, areas where adjacent materials are not fully bonded. IP denotes Incomplete Penetration, that is, zones where fusion did not fully penetrate through the interface. Figure 11 shows an illustrative example from this dataset. The WEDA dataset consists of 3,077 images, divided into training (2,154 images), validation (616 images), and test (307 images) subsets. Each image represents a segment of cable surface captured under standardized imaging conditions. All images are annotated to indicate the presence or absence of surface defects, specifically visual irregularities.

4.2 Experiment Results

In this study, YOLOv7 was used as the baseline object detection model without architectural modifications. All experiments followed the official YOLOv7 implementation for training, detection, and testing. Hyperparameters were kept identical to the default configuration to ensure a fair comparison with the baseline method, DUA. For the **UAV dataset**, the model was pre-trained on the *Microsoft COCO: Common Objects in Context* (COCO) dataset [35]. Active learning started with an initial set of 100 randomly labeled images, followed by 11 iterations, each adding 100 samples. For the WEDA dataset, the same COCO-pretrained YOLOv7 model was used. Due to domain differences between WEDA and COCO, transfer learning was applied before active learning. An initial subset of 500 WEDA images was fine-tuned for 200 epochs, after which 100 samples were selected per iteration for six iterations. All experiments were conducted five times, and the reported results correspond to the average performance.

4.2.1 Results on UAV Dataset

Four sampling strategies were evaluated on the UAV dataset: the reproduced baseline model, DUA, a fully uncertainty-based model (100Un), a fixed 50:50 uncertainty-diversity

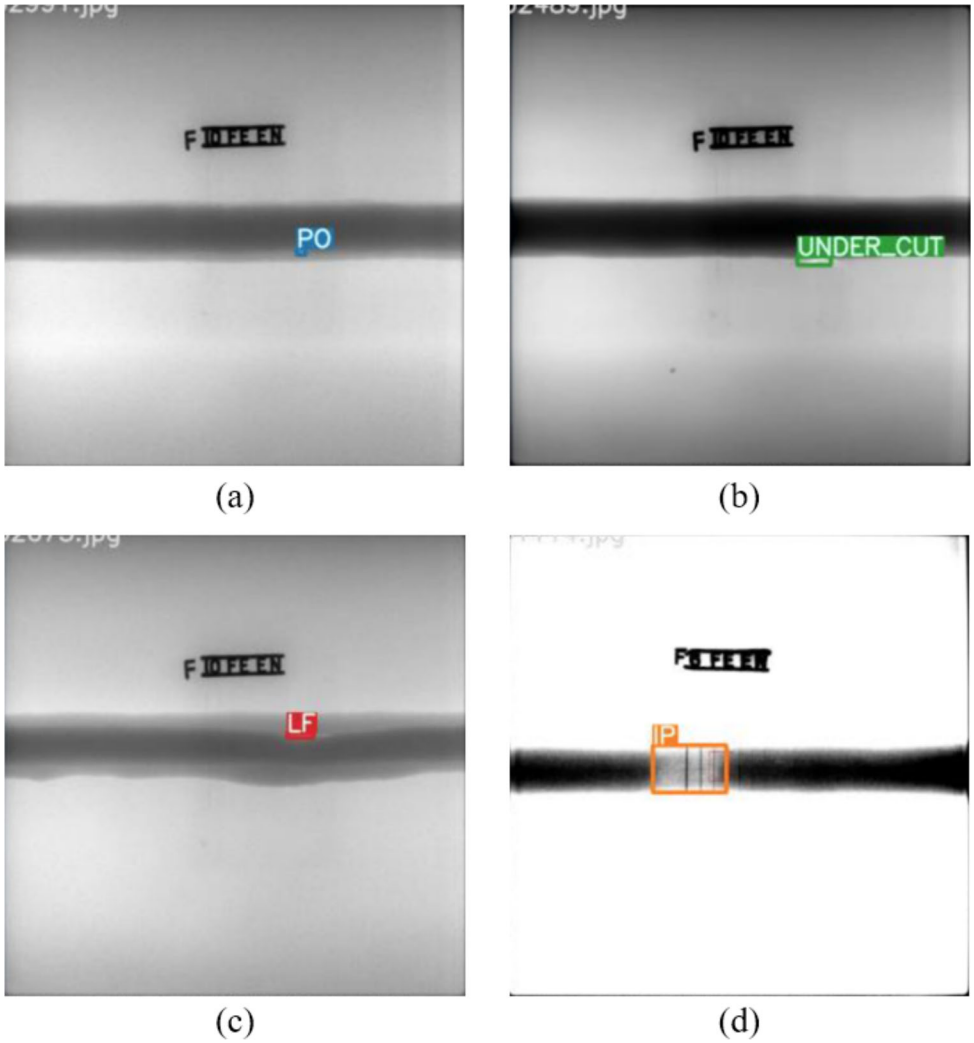

Fig. 11 Sample images of the WEDA dataset; (**a**) denotes a PO, (**b**) denotes an UNDER_CUT, (**c**) denotes an LF, (**d**) denotes an IP

model (50Un50Div), and AdaMix-AL, which adaptively adjusts the ratio based on *uncertainty score*.

Table 1 represents the main results of the experiment in the UAV. The 100Un and reproduced DUA models showed similar performance. The 50Un50Div model improved performance by approximately 0.7%. AdaMix-AL achieved a 1.1% performance improvement over DUA by dynamically adjusting the sampling ratio throughout the training. Fig-

Table 1 Main result comparison on the UAV dataset. The table presents an accuracy of mAP@0.5

Method	DUA	100Un	50Un50Div	AdaMix-AL	Core-set
All	27.5	27.5	28.2	**28.6**	26.2
Pedestrian	30.2	29.3	31.4	32.2	26.3
People	19.6	20.2	20.9	20.9	18.1
Bicycle	9.98	9.11	10.5	9.07	5.77
Car	70.6	68.8	71	71.1	68.4
Van	26	25.9	26.6	26.3	23.6
Truck	23.6	24.6	23.3	21.7	28.2
Tricycle	14.3	15.1	15.8	15.1	11
Awning-tricycle	9.19	10.6	10.2	12.1	9.51
Bus	44.5	44.1	44.7	48.4	48.1
Motor	26.6	26.9	27.4	28.6	23.3

The bold text indicates the best performing model among each experiment or model

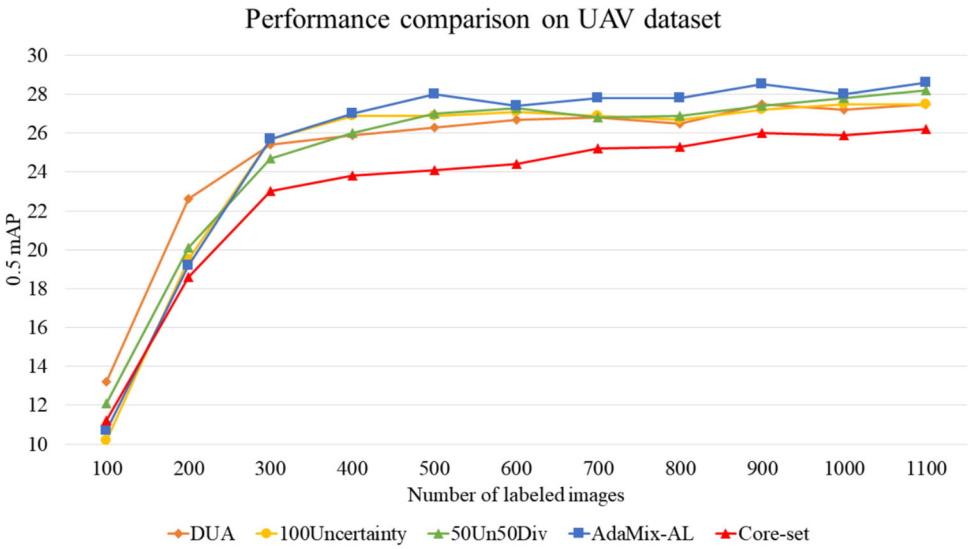

Fig. 12 Model performances at each iteration

ure 12 represents the performance across iterations. The x-axis represents the number of labeled images, which increases by 100 in each iteration, and the y-axis represents mAP@0.5. As the initial 100 images were randomly selected, the early performance differences are large. From 300 labeled images onward, AdaMix-AL consistently outperforms other models. The outperformance of the AdaMix-AL continues throughout later iterations. The performance of the AdaMix-AL, which was trained with 500 labeled images, already outperformed the DUA trained with 1100 labeled images. The performance of AdaMix-AL was 28%, and the performance of DUA was 27.5%. This result indicates that AdaMix-AL analyzes unlabeled data more efficiently. It also suggests that combining uncertainty and diversity yields better results, particularly when their ratio is adjusted dynamically.

4.2.2 Results on WEDA Dataset

The WEDA dataset differs significantly from COCO, requiring more labeled data in the early training phase.

When trained with only 100 labeled images for 100 epochs in the first iteration, the model failed to detect objects. To address this, the initial iteration used 500 labeled images and trained for 200 epochs. Subsequent iterations added 100 images each and trained for 100 epochs. Table 2 reports the results of transfer learning using the COCO-pretrained YOLOv7 model. With the full WEDA training set, the model achieved 92.8% accuracy at 600 epochs.

Table 3 shows the performance comparison of the three models tested on WEDA. DUA achieved its highest accuracy (83.5%) in the second iteration. 50Un50Div reached 91.3% in the third iteration, while AdaMix-AL achieved 93.2% in the fourth iteration. AdaMix-AL consistently outperformed DUA across all iterations, with performance differences ranging from 4% to 18%. These results confirm that AdaMix-AL is better suited for real industrial data than DUA and demonstrate the potential of applying active learning to industrial applications.

4.3 Ablation Study

Additional experiments were conducted to support these findings. The first experiment aimed to identify the optimal number of labeled images required in early iterations on the UAV dataset. As the model was pre-trained on COCO, transfer learning was necessary.

The objective was to determine the effect of the number of labeled images in the first iteration on model performance and establish the minimum number of labeled images needed for reliable detection. This step was critical because both

Table 2 Transfer learning results for the WEDA dataset of the YOLOv7 model pretrained on the COCO dataset

Epochs	mAP@0.5	mAP@0.5:0.95
200	87.7	44.8
300	90.5	46.9
400	90.5	46.4
500	91.2	47.9
600	**92.8**	**49.5**
700	91.6	49
800	92.3	48.9
900	90	48.7

The bold text indicates the best performing model among each experiment or model

Table 3 Main comparison in the WEDA dataset. AdaMix-AL achieved a performance improvement compared to the baseline model, DUA

Iter	Labeled images	Method	mAP@0.5	mAP@0.5:0.95
0	500	DUA	51.5	22.4
		50Un50Div	71.1	32
		AdaMix-AL	65.7	23.9
		Core-set	43.3	18.6
1	600	DUA	68.1	29.2
		50Un50Div	85	40.7
		AdaMix-AL	83.1	39.4
		Core-set	67.9	27.4
2	700	DUA	**83.5**	**38.4**
		50Un50Div	90.5	44
		AdaMix-AL	87.8	43.8
		Core-set	74.3	33.8
3	800	DUA	82.3	35.4
		50Un50Div	91.3	42.1
		AdaMix-AL	88.8	44.8
		Core-set	79.1	37.6
4	900	DUA	81.7	36.6
		50Un50Div	**91.3**	**42.3**
		AdaMix-AL	**93.2**	**45.9**
		Core-set	84.6	41.3
5	1000	DUA	73.9	33.9
		50Un50Div	90.4	45.6
		AdaMix-AL	91.6	44.6
		Core-set	84.5	41.2
6	1100	DUA	77.9	34.3
		50Un50Div	88.9	44.6
		AdaMix-AL	91.2	45.9
		Core-set	**87.5**	**45.2**

The bold text indicates the best performing model among each experiment or model

50Un50Div and AdaMix-AL rely on YOLOv7's detection results to select images for annotation. Therefore, even with limited initial labels, the model must detect classes accurately in the unlabeled set. Table 4 shows results when starting with 100 labeled images with an addition of 100 per iteration up to 1,100 total. The experiment in Table 5 shows results when starting with 500 labeled images. Both experiments produced similar performance because YOLOv7 was pre-trained on COCO, which contains 80 diverse classes, many overlapping with the 10 UAV classes (e.g., UAV includes "pedestrian" and "person," while COCO uses a general "person" class). Due to this overlap, increasing the initial labeled set provided minimal benefit. Therefore, the initial labeled set was set to 100 to minimize labeling cost. Another experiment investigated the effect of diversity on model performance. To observe the impact of diversity, we selected the uncertainty-based model(100Un). The first iteration started with 500 labeled images. In the first experiment (100Un), 500 images were randomly selected. In the second experiment (100Un with uniform), 500 images were selected to maximize class uniformity. The model trained with the uniform initial image outperformed the other by 1.4% (Table 6). This result indicates that diversity affects model performance, making the degree of diversity in the initial images important.

Based on these results, the next experiment selected samples solely based on diversity for the first two iterations. The results (Fig. 13) compare three models with different uncertainty-to-diversity ratios. Each method used diversity-only sampling for the first two iterations, and then applied a specific ratio. For example, 7:3 means 70% uncertainty and

Table 4 The performance of 50Un50Div and 100Un starting with 100 labeled images. The results show that training started with an initial set of 100 labeled images from the UAV dataset

Iter	Labeled images	Method	mAP@0.5	mAP@0.5:0.95
0	100	100Un	13.4	6.12
		50Un50Div	12.1	6.23
1	200	100Un	22.5	11.2
		50Un50Div	20.1	10.3
2	300	100Un	26	13.3
		50Un50Div	24.7	12.7
3	400	100Un	25.7	13.3
		50Un50Div	26	13.6
4	500	100Un	26.8	14
		50Un50Div	27	14.3
5	600	100Un	27.2	14.4
		50Un50Div	27.3	14.5
6	700	100Un	27.2	14.4
		50Un50Div	26.8	14.3
7	800	100Un	27.2	14.5
		50Un50Div	26.9	27.4
8	900	100Un	27.2	14.6
		50Un50Div	27.4	14.8
9	1000	100Un	27	14.6
		50Un50Div	27.8	14.9
10	1100	100Un	**27.5**	**14.8**
		50Un50Div	**28.2**	**15.1**

The bold text indicates the best performing model among each experiment or model

Table 5 Performance of 50Un50Div and 100Un starting with 500 labeled images. The results show that training started with an initial set of 500 labeled images from the UAV dataset

Iter	Labeled images	Method	mAP@0.5	mAP@0.5:0.95
0	500	100Un	25.2	13.8
		50Un50Div	25.2	13.8
1	600	100Un	25.2	13.7
		50Un50Div	25.7	13.8
2	700	100Un	25.7	14
		50Un50Div	26.5	14.5
3	800	100Un	26.2	14.3
		50Un50Div	27.1	14.6
4	900	100Un	25.8	14.3
		50Un50Div	27.4	14.8
5	1000	100Un	26.1	14.1
		50Un50Div	27.4	14.9
6	1100	100Un	**26.9**	**14.6**
		50Un50Div	**27.8**	**15.1**

The bold text indicates the best performing model among each experiment or model

Table 6 Performance of 100Un with and without initial uniform labeled images. Performance of the models depending on whether the classes of the first 500 images are uniform or not

Iter	Labeled images	Method	mAP@0.5	mAP@0.5:0.95
0	500	100Un	25.2	13.8
		100Un(With uniform)	27.3	14.7
1	600	100Un	25.2	13.7
		100Un(With uniform)	27.7	14.9
2	700	100Un	25.7	14
		100Un(With uniform)	27.6	14.8
3	800	100Un	26.2	14.3
		100Un(With uniform)	28	15.1
4	900	100Un	25.8	14.3
		100Un(With uniform)	28.2	15.2
5	1000	100Un	26.1	14.1
		100Un(With uniform)	**28.3**	**15.2**
6	1100	100Un	**26.9**	**14.6**
		100Un(With uniform)	28.2	15.2

The bold text indicates the best performing model among each experiment or model

30% diversity. The goal was to examine the effect of this ratio on performance. With 7:3, the model achieved 28.4%; with 8:2, 28.1%; and with 9:1, 28%. This indicates that the ratio between uncertainty and diversity influences performance. In Fig. 13, the 9:1 ratio improved fastest in the early stage (after 400 labeled images, as the first two iterations were diversity-only). However, as training progressed, the 7:3 ratio achieved the highest performance. This supports our hypothesis that diversity becomes more important as training progresses.

The final experiment on the UAV dataset tested whether adjusting the sampling ratio during training improves performance. Based on the uncertainty distribution of the UAV dataset (Fig. 7), a threshold of 65 was chosen to change the ratio at the training midpoint. If the 80th uncertainty value exceeded this threshold, a 9:1 ratio was applied; otherwise, a 7:3 ratio was applied. This adaptive strategy achieved 29%, the highest among all settings (Fig. 14), suggesting that adjusting the ratio during training improves performance. To automate this process, we introduced (4) to calculate the *uncertainty score*, enabling dynamic ratio control for any dataset. Although AdaMix-AL showed slightly lower overall performance compared to manual tuning, it required no manual intervention and surpassed the manually tuned model by the 8th iteration, showing strong potential.

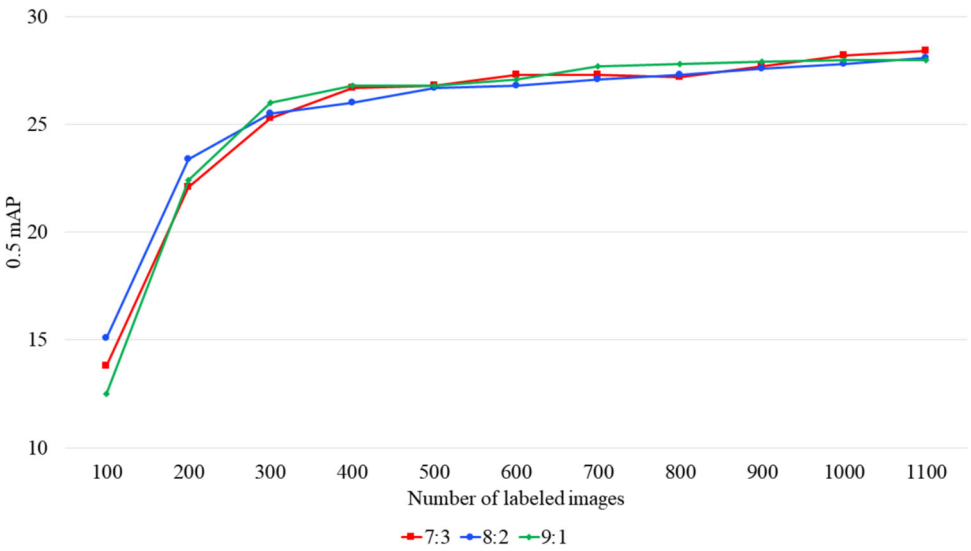

Fig. 13 Object detection performance with the UAV dataset according to the ratio of the uncertainty and diversity. Each color represents the ratios of 7:3, 8:2, and 9:1, respectively

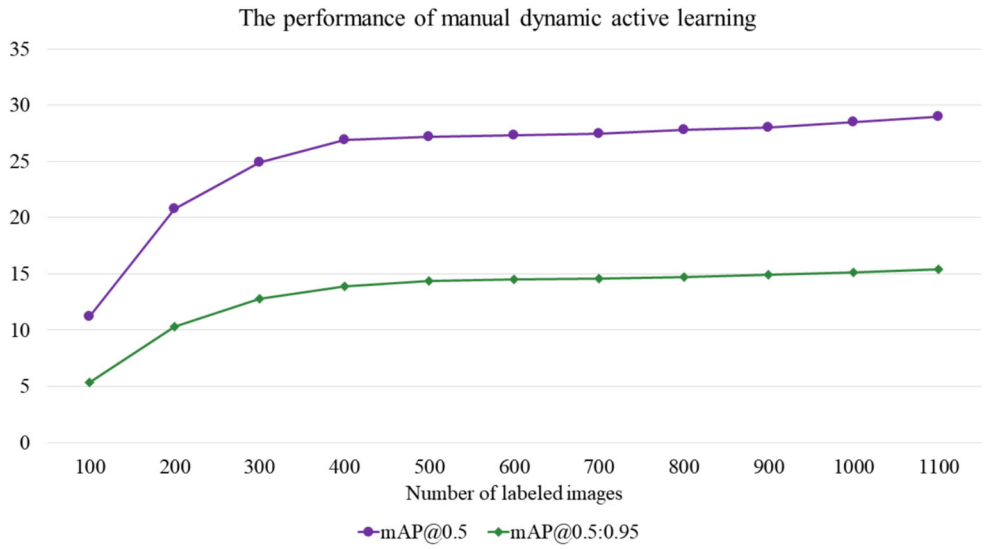

Fig. 14 Result of evaluating the ratio of uncertainty to diversity using a predetermined uncertainty threshold

5 Conclusion

This study introduces AdaMix-AL, a hybrid active learning framework combining uncertainty-based and diversity-based methods for improved object detection. It removes the need to sacrifice annotation budget for evaluation by using YOLOv7's detect instead of test, making full use of available data. AdaMix-AL applies an adaptive sampling strategy that adjusts the uncertainty-diversity ratio based on *uncertainty score*. When uncertainty is high, the model selects more uncertain samples; as uncertainty decreases, it prioritizes diversity. This helps the model explore a broader feature space and maintain improvements even when uncertainty-based methods become less effective. AdaMix-AL achieved a 1.1% improvement on the UAV dataset and 11.5% on the WEDA dataset. Experiments on UAV and WEDA datasets confirmed that the proposed method outperforms existing approaches such as DUA in both annotation efficiency and accuracy. AdaMix-AL reduces labeling costs while improving performance, and its adaptive nature renders it suitable for various data types. This work contributes to scalable active learning frameworks for object detection with promising cross-domain results. Although AdaMix-AL improves efficiency and accuracy, future research is needed. One direction is task-aware sampling strategies. AdaMix-AL currently targets object detection using general uncertainty and diversity measures. However, uncertainty characteristics differ across tasks. For example, in semantic segmentation, uncertainty may be better captured at the pixel or region level. Future work could design strategies tailored to specific tasks, extending applications beyond object detection. Moving in this direction will make active learning more adaptive, task-specific, and semantically aware. In addition, while the current study employs YOLOv7, future work could extend the experiments by adopting more powerful detection frameworks or backbone architectures with richer feature representations. Such representations, including those inspired by transformer-based or large-scale pre-trained vision models, may further enhance uncertainty and diversity estimation. Furthermore, future research could investigate combining such detectors with the Segment Anything Model (SAM) [36, 37] to reduce annotation costs. Such extensions would also enable uncertainty and diversity to be estimated at finer spatial resolutions, such as regions or pixels, rather than being limited to bounding-box-level information. This progress will improve efficiency and enable better analysis of unlabeled data in real-world applications, ultimately contributing to scalable and intelligent learning systems across domains.

Acknowledgements This work was supported by an Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government(MSIT) (No. RS-2024-00460194, Development of Telepresence and Pass-Through Service Technologies for XRDevices and No. 2021-0-00766, Development of Integrated Development Framework that supports Automatic Neural NetworkGeneration and Deployment optimized for runtime environment).

Author Contributions Joonsun Auh: Conceptualization, Data curation, Formal analysis, Investigation, Software, Resources, Methodology, Validation, Visualization, Writing Changsik Cho: Methodology, Formal analysis, Resources, Investigation, Funding acquisition Seon-tae Kim: Conceptualization, Methodology, Formal analysis, Resources, Writing, Supervision, Project administration, Funding acquisition.

Data Availability This paper used the VisDrone-2019 and WEDA datasets. The VisDrone-2919 datasets are available at https://github.com/VisDrone/VisDrone-Dataset. The WEDA datasets and code generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflicts of Interest The authors declare that they have no conflict of interest.

Competing interests The authors declare no competing interests.

References

1. Zhang, X., Feng, Y., Zhang, S., Wang, N., Mei, S.: Finding nonrigid tiny person with densely cropped and local attention object detector networks in low-altitude aerial images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. **15**, 4371–4385 (2022)
2. Zhang, X., Feng, Y., Zhang, S., Wang, N., Lu, G., Mei, S.: Robust aerial person detection with lightweight distillation network for edge deployment. IEEE Trans. Geosci. Remote Sens. **62**, 1–16 (2024). https://doi.org/10.1109/TGRS.2024.3421310
3. Zhang, X., Feng, Y., Wang, N., Lu, G., Mei, S.: Transformer-based person detection in paired RGB-T aerial images with VTSaR dataset. IEEE J. Select. Topics Appl. Earth Observ. Remote Sens., **18**, 5082–5099 (2025). https://doi.org/10.1109/JSTARS.2025.3526995
4. Michaelis, C., Mitzkus, B., Geirhos, R., Rusak, E., Bringmann, O., Ecker, A.S., Bethge, M., Brendel, W.: Benchmarking robustness in object detection: Autonomous driving when winter is coming (2019). arXiv:1907.07484
5. Wang, L., Zhang, X., Song, Z., Bi, J., Zhang, G., Wei, H., Tang, L., Yang, L., Li, J., Jia, C., et al.: Multi-modal 3D object detection in autonomous driving: A survey and taxonomy. IEEE Transactions on Intelligent Vehicles. **8**(7), 3781–3798 (2023)
6. Jeon, H., Kim, H., Kim, D., Kim, J.: PASS-CCTV: Proactive anomaly surveillance system for CCTV footage analysis in adverse environmental conditions. Expert Systems with Applications, **124391** (2024)
7. Kim, H., Jeon, H., Kim, D., Kim, J.: Elevating urban surveillance: A deep CCTV monitoring system for detection of anomalous events via human action recognition. Sustain. Cities Soc. **114**, 105793 (2024)
8. Liu, G., Hu, Y., Chen, Z., Guo, J., Ni, P.: Lightweight object detection algorithm for robots with improved YOLOv5. Eng. Appl. Artif. Intell. **123**, 106217 (2023)
9. Song, K., Wang, J., Bao, Y., Huang, L., Yan, Y.: A novel visible-depth-thermal image dataset of salient object detection for robotic visual perception. IEEE/ASME Trans. Mechatron. **28**(3), 1558–1569 (2022)
10. Singh, K.J., Kapoor, D.S., Thakur, K., Sharma, A., et al.: Computer-vision based object detection and recognition for service robot in indoor environment. Computers, Materials & Continua **72**(1) (2022)
11. Ghasemi, Y., Jeong, H., Choi, S.H., Park, K.-B., Lee, J.Y.: Deep learning-based object detection in augmented reality: A systematic review. Comput. Ind. **139**, 103661 (2022)
12. Łysakowski, M., Żywanowski, K., Banaszczyk, A., Nowicki, M.R., Skrzypczyński, P., Tadeja, S.K.: Real-time onboard object detection for augmented reality: Enhancing head-mounted display with YOLOv8. In: 2023 IEEE International Conference on Edge Computing and Communications (EDGE), pp. 364–371 (2023). IEEE
13. Bengar, J.Z., Weijer, J., Twardowski, B., Raducanu, B.: Reducing label effort: Self-supervised meets active learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1631–1639 (2021)
14. Chan, Y.-C., Li, M., Oymak, S.: On the marginal benefit of active learning: Does self-supervision eat its cake? In: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3455–3459 (2021). IEEE
15. Wang, X., Chi, X., Song, Y., Yang, Z.: Active learning with label quality control. PeerJ Computer Science. **9**, 1480 (2023)
16. Bangert, P., Moon, H., Woo, J., Didari, S., Hao, H.: Active learning performance in labeling radiology images is 90% effective. Frontiers in radiology (2021)
17. Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7464–7475 (2023)
18. Settles, B.: Active learning literature survey. (2009)
19. Takezoe, R., Liu, X., Mao, S., Chen, M.T., Feng, Z., Zhang, S., Wang, X., et al.: Deep active learning for computer vision: Past and future. APSIPA Transactions on Signal and Information Processing **12**(1) (2023)
20. Lewis, D.D.: A sequential algorithm for training text classifiers: Corrigendum and additional data. In: Proceedings of the ACM SIGIR Forum, vol. 29, pp. 13–19 (1995). ACM New York, NY, USA
21. Lewis, D.D., Catlett, J.: Heterogeneous uncertainty sampling for supervised learning. In: Machine Learning Proceedings **1994**, pp. 148–156. Elsevier, (1994)
22. Gal, Y., Islam, R., Ghahramani, Z.: Deep Bayesian active learning with image data. In: Proceedings of the International Conference on Machine Learning, pp. 1183–1192 (2017). PMLR
23. Tong, S., Koller, D.: Support vector machine active learning with applications to text classification. J. Mach. Learn. Res. **2**(Nov), 45–66 (2001)
24. Yoo, D., Kweon, I.S.: Learning loss for active learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 93–102 (2019)
25. Elhamifar, E., Sapiro, G., Yang, A., Sasrty, S.S.: A convex optimization framework for active learning. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 209–216 (2013)
26. Roth, D., Small, K.: Margin-based active learning for structured output spaces. In: Proceedings of the Machine Learning: ECML 2006: 17th European Conference on Machine Learning Berlin, Germany, September 18–22, 2006 Proceedings 17, pp. 413–424 (2006). Springer
27. Joshi, A.J., Porikli, F., Papanikolopoulos, N.: Multi-class active learning for image classification. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2372–2379 (2009). IEEE
28. Luo, W., Schwing, A., Urtasun, R.: Latent structured active learning. Adv. Neural Inf. Proc. Syst. **26** (2013)
29. Geifman, Y., El-Yaniv, R.: Deep active learning over the long tail (2017). arXiv:1711.00941

30. Ash, J.T., Zhang, C., Krishnamurthy, A., Langford, J., Agarwal, A.: Deep batch active learning by diverse, uncertain gradient lower bounds (2019). arXiv:1906.03671
31. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach (2017). arXiv:1708.00489
32. Arthur, D., Vassilvitskii, S.: k-means++: The advantages of careful seeding. Technical report, Stanford (2006)
33. Yamani, A., Alyami, A., Luqman, H., Ghanem, B., Giancola, S.: Active learning for single-stage object detection in UAV images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1860–1869 (2024)
34. Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., Ling, H.: Detection and tracking meet drones challenge. IEEE Trans. Pattern Anal. Mach. Intell. **44**(11), 7380–7399 (2021)
35. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: Proceedings of the Computer vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13, pp. 740–755 (2014). Springer
36. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026 (2023)
37. Sozio, A., Rizzo, A., Mariano Scarrica, V., Patrizio Ciro Aucelli, P., Anfuso, G., Barracane, G., Dimuccio, L.A., Ferreira, R., La Salandra, M., Staiano, A., et al.: An innovative SAM-ViT based tool for the automatic detection of litter items on sandy beaches. EGU General Assembly **2024** (2024)

Joonsun Auh received his Ph.D. degree in artificial intelligence from the University of Science & Technology, UST, Rep. of Korea, in 2025 and the M.S. degree in computer science from Yonsei University, Rep. of Korea, and the B.S. degree from Chungnam National University, Rep. of Korea in 2018. He has been a postdoctoral researcher with the Electronics and Telecommunication Research Institute, Rep. of Korea, since September 2025. His research interests include image processing, visual representation, active learning, and self-supervised learning.

Changsik Cho received his Ph.D. from the Department of Computer Science, Chungnam National University, Rep. of Korea, in 2011, BS and MS degrees from Kyungpook National University, Rep. of Korea, in 1993 and 1995. In January 1995, he joined the Electronics and Telecommunications Research Institute, Rep. of Korea, where he is currently a principal researcher. His research interests include AutoML, MLOps, and No-code neural network development tools.

Seon-tae Kim received a BS degree from the Department of Electrical and Electronics Engineering, Korea Advanced Institute of Science and Technology, in 1997, an MS degree from Seoul National University, in 2000, and a PhD degree from Korea University, Rep. of Korea, in 2012. He has been a Principal Researcher with the Electronics and Telecommunication Research Institute, Rep. of Korea, in February 2000. Since 2021, he has also been a professor in the Department of Artificial Intelligence at the University of Science and Technology. His research interests include image processing, visual representation, lightweight OS, and sensor networks.