DuET: Dual Incremental Object Detection via Exemplar-Free Task Arithmetic

Supplementary Material

A. Detailed Evaluation Protocol and Metrics

This section provides a detailed discussion of the proposed evaluation protocol and metrics for the DuIOD task. Table S1 outlines the training sequence that is followed for four different DuIOD experiments, along with the detailed evaluation protocol that is used to comprehensively evaluate the performance of different object detectors on the respective DuIOD setting. Unlike existing metrics [1, 21, 24] that focus only on catastrophic forgetting, we used the **Retention-Adaptability Index (RAI)**, which balances both knowledge retention and generalisation to unseen categories across evolving domains. We define RAI as the mean of the Average Retention Index (Avg RI) and Average Generalisation Index (Avg GI), which are discussed in the sections below.

$$RAI = \frac{\text{Avg RI} + \text{Avg GI}}{2} \tag{1}$$

A.1. Average Retention Index

For each domain $\mathcal{D}_i$ corresponding to task $\mathcal{T}_i$ where $i \in \{1, \ldots, T-1\}$, we define the Retention Index $RI_{\mathcal{D}_i}$ as:

$$RI_{\mathcal{D}_i} = \frac{\text{mAP}^{\mathcal{T}_T}_{\text{old}}(\mathcal{D}_i[\mathcal{C}_i])}{\text{mAP}^{\mathcal{T}_i}_{\text{new}}(\mathcal{D}_i[\mathcal{C}_i])} \tag{2}$$

Here, $\text{mAP}^{\mathcal{T}_T}_{\text{old}}(\mathcal{D}_i[\mathcal{C}_i])$ denotes the mean Average Precision (mAP) at IoU threshold = 0.5 of the object detector at the final task $\mathcal{T}_T$ on the classes $\mathcal{C}_i$ which were learned from domain $\mathcal{D}_i$, and $\text{mAP}^{\mathcal{T}_i}_{\text{new}}(\mathcal{D}_i[\mathcal{C}_i])$ is the mAP when classes $\mathcal{C}_i$ from domain $\mathcal{D}_i$ were first encountered and learned, at task $\mathcal{T}_i$. The Avg RI is then calculated as:

$$\text{Avg RI} = \frac{1}{T-1} \sum_{i=1}^{T-1} RI_{\mathcal{D}_i}. \tag{3}$$

To illustrate this, consider the multi-phase experiment (Table S1) with training sequence: *Night Sunny [1:2] → Daytime Sunny [3:4] → Daytime Foggy [5:7]*. The Avg RI is computed as the mean of the Retention Index values for Night Sunny (NS) and Daytime Sunny (DS) domains at the final task $\mathcal{T}_3$ as follows:

$$RI_{\text{NS}} = \frac{\text{mAP}^{\mathcal{T}_3}_{\text{old}}(\text{NS}[1{:}2])}{\text{mAP}^{\mathcal{T}_1}_{\text{new}}(\text{NS}[1{:}2])} \quad RI_{\text{DS}} = \frac{\text{mAP}^{\mathcal{T}_3}_{\text{old}}(\text{DS}[3{:}4])}{\text{mAP}^{\mathcal{T}_2}_{\text{new}}(\text{DS}[3{:}4])} \tag{4}$$

$$\text{Avg RI} = \frac{RI_{\text{NS}} + RI_{\text{DS}}}{2} \tag{5}$$

Hence, in this case, a higher Avg RI indicates how effectively the object detector has retained past knowledge from old Night Sunny and Daytime Sunny domains in the final task $\mathcal{T}_3$. Conversely, a lower value indicates significant catastrophic forgetting.

A.2. Average Generalization Index

The Generalisation Index ($GI_{\mathcal{D}_i,\mathcal{T}_j}$) quantifies how well the model detects unseen classes from domain $\mathcal{D}_i$ at task $\mathcal{T}_j$. These classes were not part of the training set for task $\mathcal{T}_j$, meaning the model is required to generalise beyond its explicitly trained classes (see Table S1). For a given domain $\mathcal{D}_i$ at task $\mathcal{T}_j$, the Generalization Index is computed as:

$$GI_{\mathcal{D}_i,\mathcal{T}_j} = \frac{\text{mAP}^{\mathcal{T}_j}_{\text{unseen}}(\mathcal{D}_i[\mathcal{C}_{\text{unseen}}])}{\text{mAP}_{\text{ref}}(\mathcal{D}_i[\mathcal{C}_{\text{unseen}}])} \tag{6}$$

Here, $\text{mAP}^{\mathcal{T}_j}_{\text{unseen}}(\mathcal{D}_i[\mathcal{C}_{\text{unseen}}])$ is the mAP of the model at task $\mathcal{T}_j$ on the unseen classes $\mathcal{C}_{\text{unseen}}$ from domain $\mathcal{D}_i$, and $\text{mAP}_{\text{ref}}(\mathcal{D}_i[\mathcal{C}_{\text{unseen}}])$ is the reference mAP obtained by training the object detector solely on these unseen classes on domain $\mathcal{D}_i$. The Average Generalisation Index (Avg GI) over all relevant domain-task pairs is then computed as:

$$\text{Avg GI} = \frac{1}{\mathcal{N}} \sum_{(\mathcal{D}_i,\mathcal{T}_j)} GI_{\mathcal{D}_i,\mathcal{T}_j} \tag{7}$$

where $\mathcal{N}$ is the total number of unseen-class domain-task pairs considered in the evaluation.

Continuing with the same example, the Avg GI is computed as the mean of the Generalization Index values for unseen classes across the Night Sunny (NS), Daytime Sunny (DS), and Daytime Foggy (DF) domains for a total of five domain-task pairs- two from task $\mathcal{T}_2$ and three from $\mathcal{T}_3$:

$$GI_{\text{NS},\mathcal{T}_2} = \frac{\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{NS}[3{:}4])}{\text{mAP}_{\text{ref}}(\text{NS}[3{:}4])} \quad GI_{\text{DS},\mathcal{T}_2} = \frac{\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{DS}[1{:}2])}{\text{mAP}_{\text{ref}}(\text{DS}[1{:}2])} \tag{8}$$

$$GI_{\text{NS},\mathcal{T}_3} = \frac{\text{mAP}^{\mathcal{T}_3}_{\text{unseen}}(\text{NS}[3{:}4])}{\text{mAP}_{\text{ref}}(\text{NS}[3{:}4])} \quad GI_{\text{DS},\mathcal{T}_3} = \frac{\text{mAP}^{\mathcal{T}_3}_{\text{unseen}}(\text{DS}[1{:}2])}{\text{mAP}_{\text{ref}}(\text{DS}[1{:}2])}$$

$$GI_{\text{DF},\mathcal{T}_3} = \frac{\text{mAP}^{\mathcal{T}_3}_{\text{unseen}}(\text{DF}[1:4])}{\text{mAP}_{\text{ref}}(\text{DF}[1:4])} \tag{9}$$

$$\text{Avg GI} = \frac{GI_{\text{NS},\mathcal{T}_2} + GI_{\text{DS},\mathcal{T}_2} + GI_{\text{NS},\mathcal{T}_3} + GI_{\text{DS},\mathcal{T}_3} + GI_{\text{DF},\mathcal{T}_3}}{5} \tag{10}$$

Hence, in this case, a higher Avg GI indicates better zero-shot generalisation to unseen categories: NS [3:4], DS [1:2], and DF [1:4] across incremental training. Conversely, a lower value suggests that the model is overfitting to seen classes and fails to generalise.

Table S1. Training Sequence & Evaluation Protocol for different DuIOD experiments.

DuIOD Experiment	**Training Sequence**		**Evaluation Protocol**		
	Task	**Class IDs**	**New Classes**	**Old Classes**	**Unseen Classes**
Pascal Series Datasets					
Two Phase VOC [1:10] → Clipart [11:20]	$\mathcal{T}_1$	1-10 from VOC	$\text{mAP}^{\mathcal{T}_1}_{\text{new}}(\text{VOC}[1:10])$	—	—
	$\mathcal{T}_2$	11-20 from Clipart	$\text{mAP}^{\mathcal{T}_2}_{\text{new}}(\text{Clipart}[11:20])$	$\text{mAP}^{\mathcal{T}_2}_{\text{old}}(\text{VOC}[1:10])$	$\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{VOC}[11:20])$ $\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{Clipart}[1:10])$
Multi Phase Watercolor [1:3] → Comic [4:6] → Clipart [7:13] → VOC [14:20]	$\mathcal{T}_1$	1-3 from Watercolor	$\text{mAP}^{\mathcal{T}_1}_{\text{new}}(\text{Watercolor}[1:3])$	—	—
	$\mathcal{T}_2$	4-6 from Comic	$\text{mAP}^{\mathcal{T}_2}_{\text{new}}(\text{Comic}[4:6])$	$\text{mAP}^{\mathcal{T}_2}_{\text{old}}(\text{Watercolor}[1:3])$	$\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{Watercolor}[4:6])$ $\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{Comic}[1:3])$
	$\mathcal{T}_3$	7-13 from Clipart	$\text{mAP}^{\mathcal{T}_3}_{\text{new}}(\text{Clipart}[7:13])$	$\text{mAP}^{\mathcal{T}_3}_{\text{old}}(\text{Watercolor}[1:3])$ $\text{mAP}^{\mathcal{T}_3}_{\text{old}}(\text{Comic}[4:6])$	$\text{mAP}^{\mathcal{T}_3}_{\text{unseen}}(\text{Watercolor}[4:6])$ $\text{mAP}^{\mathcal{T}_3}_{\text{unseen}}(\text{Comic}[1:3])$ $\text{mAP}^{\mathcal{T}_3}_{\text{unseen}}(\text{Clipart}[1:6])$
	$\mathcal{T}_4$	14-20 from VOC	$\text{mAP}^{\mathcal{T}_4}_{\text{new}}(\text{VOC}[14:20])$	$\text{mAP}^{\mathcal{T}_4}_{\text{old}}(\text{Watercolor}[1:3])$ $\text{mAP}^{\mathcal{T}_4}_{\text{old}}(\text{Comic}[4:6])$ $\text{mAP}^{\mathcal{T}_4}_{\text{old}}(\text{Clipart}[7:13])$	$\text{mAP}^{\mathcal{T}_4}_{\text{unseen}}(\text{Watercolor}[4:6])$ $\text{mAP}^{\mathcal{T}_4}_{\text{unseen}}(\text{Comic}[1:3])$ $\text{mAP}^{\mathcal{T}_4}_{\text{unseen}}(\text{Clipart}[1:6])$ $\text{mAP}^{\mathcal{T}_4}_{\text{unseen}}(\text{VOC}[1:13])$
Diverse Weather Series Datasets					
Two Phase Daytime Sunny [1:4] → Night Sunny [5:7]	$\mathcal{T}_1$	1-4 from Daytime Sunny	$\text{mAP}^{\mathcal{T}_1}_{\text{new}}(\text{Daytime Sunny}[1:4])$	—	—
	$\mathcal{T}_2$	5-7 from Night Sunny	$\text{mAP}^{\mathcal{T}_2}_{\text{new}}(\text{Night Sunny}[5:7])$	$\text{mAP}^{\mathcal{T}_2}_{\text{old}}(\text{Daytime Sunny}[1:4])$	$\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{Daytime Sunny}[5:7])$ $\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{Night Sunny}[1:4])$
Multi Phase Night Sunny [1:2] → Daytime Sunny [3:4] → Daytime Foggy [5:7]	$\mathcal{T}_1$	1-2 from Night Sunny	$\text{mAP}^{\mathcal{T}_1}_{\text{new}}(\text{Night Sunny}[1:2])$	—	—
	$\mathcal{T}_2$	3-4 from Daytime Sunny	$\text{mAP}^{\mathcal{T}_2}_{\text{new}}(\text{Daytime Sunny}[3:4])$	$\text{mAP}^{\mathcal{T}_2}_{\text{old}}(\text{Night Sunny}[1:2])$	$\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{Night Sunny}[3:4])$ $\text{mAP}^{\mathcal{T}_2}_{\text{unseen}}(\text{Daytime Sunny}[1:2])$
	$\mathcal{T}_3$	5-7 from Daytime Foggy	$\text{mAP}^{\mathcal{T}_3}_{\text{new}}(\text{Daytime Foggy}[5:7])$	$\text{mAP}^{\mathcal{T}_3}_{\text{old}}(\text{Night Sunny}[1:2])$ $\text{mAP}^{\mathcal{T}_3}_{\text{old}}(\text{Daytime Sunny}[3:4])$	$\text{mAP}^{\mathcal{T}_3}_{\text{unseen}}(\text{Night Sunny}[3:4])$ $\text{mAP}^{\mathcal{T}_3}_{\text{unseen}}(\text{Daytime Sunny}[1:2])$ $\text{mAP}^{\mathcal{T}_3}_{\text{unseen}}(\text{Daytime Foggy}[1:4])$

B. Loss Function Formulation

To ensure effective incremental learning in case of DuIOD, we employ a combination of standard detector loss $\mathcal{L}_{\text{Detector}}$, a modified distillation loss $\mathcal{L}^*_{\text{Distill}}$ (discussed below), and the Directional Consistency Loss $\mathcal{L}_{\text{DC}}$ (discussed in main paper `Section 3.5`). This section details the formulation of $\mathcal{L}_{\text{total}}$ using these loss components.

Knowledge distillation plays a crucial role in mitigating catastrophic forgetting during incremental learning. In our approach, we extend the standard distillation loss ($\mathcal{L}_{\text{Distill}}$) for incremental learning [13] by incorporating a dynamic thresholding mechanism that filters low-confidence classification outputs and high-variance bounding box predictions from the old (previous task) model.

Let $\mathcal{M}_{\theta_{t-1}}$ represent the previous task model, and $\mathcal{M}_{\theta_t}$ be the current model being trained on $\mathcal{T}_t$. Given input data for the current task $\mathcal{X}_t$, the classification outputs and predicted bounding boxes from both models will be:

$$\mathbf{z}_{\text{curr}} = \mathcal{M}_{\theta_t}(x), \quad \mathbf{z}_{\text{old}} = \mathcal{M}_{\theta_{t-1}}(x) \tag{11}$$

where $\mathbf{z} = (\mathbf{c}, \mathbf{b})$, with $\mathbf{c}$ being classification logits and $\mathbf{b}$ the predicted bounding box coordinates.

The classification distillation loss is computed as:

$$\mathcal{L}^*_{\text{Distill}_{\text{cls}}} = \frac{1}{|\mathcal{M}^*_{\text{cls}}|} \sum_{i \in \mathcal{M}^*} \left\| \mathbf{c}^{(i)}_{\text{curr}} - \mathbf{c}^{(i)}_{\text{old}} \right\|^2 \tag{12}$$

where $\mathcal{M}^*_{\text{cls}}$ is the dynamically selected mask that excludes predictions with low confidence scores in $\mathbf{c}_{\text{old}}$:

$$\mathcal{M}^*_{\text{cls}} = \{ i \mid \max(\mathbf{c}^{(i)}_{\text{old}}) \geq \tau_{\text{cls}} \} \tag{13}$$

where τ_{cls} is an adaptive threshold computed as the 75th percentile of $\max(\mathbf{c}_{\text{old}})$ values.

Similarly, the bounding box regression distillation loss is computed by computing the KL divergence between the softmax of bounding box outputs from the current and old models:

$$\mathcal{L}^*_{\text{Distill}_{\text{bbox}}} = \frac{1}{|\mathcal{M}^*_{\text{bbox}}|} \sum_{j \in \mathcal{M}^*_{\text{bbox}}} \mathcal{D}_{\text{KL}}\left(\text{Softmax}(\mathbf{b}^{(j)}_{\text{curr}}) \middle\| \text{Softmax}(\mathbf{b}^{(j)}_{\text{old}}) \right) \tag{14}$$

where $\mathcal{M}^*_{\text{bbox}}$ filters out bounding boxes with high variance in $\mathbf{b}_{\text{old}}$:

$$\mathcal{M}^*_{\text{bbox}} = \{ j \mid \text{Var}(\mathbf{b}^{(j)}_{\text{old}}) \leq \tau_{\text{bbox}} \} \tag{15}$$

where τ_{bbox} is an adaptive threshold computed as the 75th percentile of bounding box variance values.

The final modified distillation loss becomes:

$$\mathcal{L}^*_{\text{Distill}} = \mathcal{L}^*_{\text{Distill}_{\text{cls}}} + \mathcal{L}^*_{\text{Distill}_{\text{bbox}}} \tag{16}$$

$\mathcal{L}_{\text{Detector}}$ depends on the object detector used. In our framework, we augment the detector losses of YOLO11 [10] and RT-DETR [18] object detectors with $\mathcal{L}^*_{\text{Distill}}$ and $\mathcal{L}_{\text{DC}}$ on the Ultralytics [10] pipeline. In the case of YOLO11, the detection loss consists of classification loss, bounding box regression loss, and Distribution Focal Loss

[10], while in the case of RT-DETR, the detection loss follows a Hungarian matching strategy, consisting of classification, bounding box, and Generalized IoU (GIoU) losses [18].

Hence, the total loss for incremental tasks ($t \geq 2$) is computed as:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{Detector}} + \lambda_{\text{Distill}} \mathcal{L}^*_{\text{Distill}} + \lambda_{\text{DC}} \mathcal{L}_{\text{DC}} \tag{17}$$

where λ_{Distill} and λ_{DC} are scaling coefficients that control the impact of distillation and directional consistency losses, respectively.

C. Extended Ablation Studies

C.1. Impact of Loss Components

Table S2. Performance comparison of different model-merging algorithms on VOC [1:10] → Clipart [11:20] depicting the impact of $\mathcal{L}_{\text{DC}}$, with YOLO11n [10] as the base detector. Among columns, best in **bold**, second best *underlined*.

Model-merging algorithm	$\mathcal{L}_{\text{DC}}$	Avg RI (%)	Avg GI (%)	RAI (%)
Fisher-Merging [20]	✗	20.27	17.15	18.71
Fisher-Merging [20]	✓	21.64	24	22.82 (+ 4.11)
MagMax [19]	✗	65.05	28.09	46.57
MagMax [19]	✓	66.79	28.28	47.54 (+ 0.97)
Weight-Averaging [8]	✗	66.42	31.42	48.92
Weight-Averaging [8]	✓	76.12	37.53	56.83 (+ 7.91)
EMR-Merging [7]	✗	67.66	34.4	51.03
EMR-Merging [7]	✓	68.03	36.46	52.25 (+ 1.22)
DuET (Ours)	✗	87.06	37.75	62.41
DuET (Ours)	✓	**87.44**	**44.54**	**65.99** (+ 3.58)

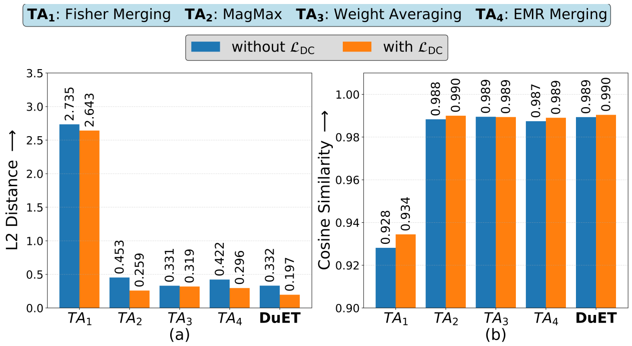


Figure S1. **Impact of $\mathcal{L}_{DC}$ in (a) reducing L2 Distance and (b) improving cosine similarity.** These results are obtained on the VOC [1:10] → Clipart [11:20] experiment using YOLO11n [10] as the base detector with Incremental Head and Sequential Fine-tuning.

Impact of $\mathcal{L}_{DC}$. Continuing the ablations from the main paper (Section 6), in this section, we further investigate the role of $\mathcal{L}_{DC}$. In Table S2, we compare the performance of different model-merging algorithms, with and without $\mathcal{L}_{DC}$. The results show that $\mathcal{L}_{DC}$ consistently improves the RAI across all methods, with an average RAI improvement of **+3.56%** among all merging methods, with DuET achieving the best performance. Moreover, the bar charts in Figure S1 compare the L2 distance and cosine similarity between the merged model weights with both old and current model weights across different model-merging algorithms. The results show that $\mathcal{L}_{DC}$ significantly reduces L2 distance by **43.46%** averaged across all methods, with DuET achieving the lowest values. Lower L2 distance suggests that after incorporating $\mathcal{L}_{DC}$, the merged model lies closer to the original models, ensuring effective knowledge integration from both. Similarly, incorporation of $\mathcal{L}_{DC}$ consistently improves cosine similarity across all methods by **0.23%** average, with DuET achieving the highest values. Higher cosine similarity suggests that $\mathcal{L}_{DC}$ helps the merged model better align with the original models.

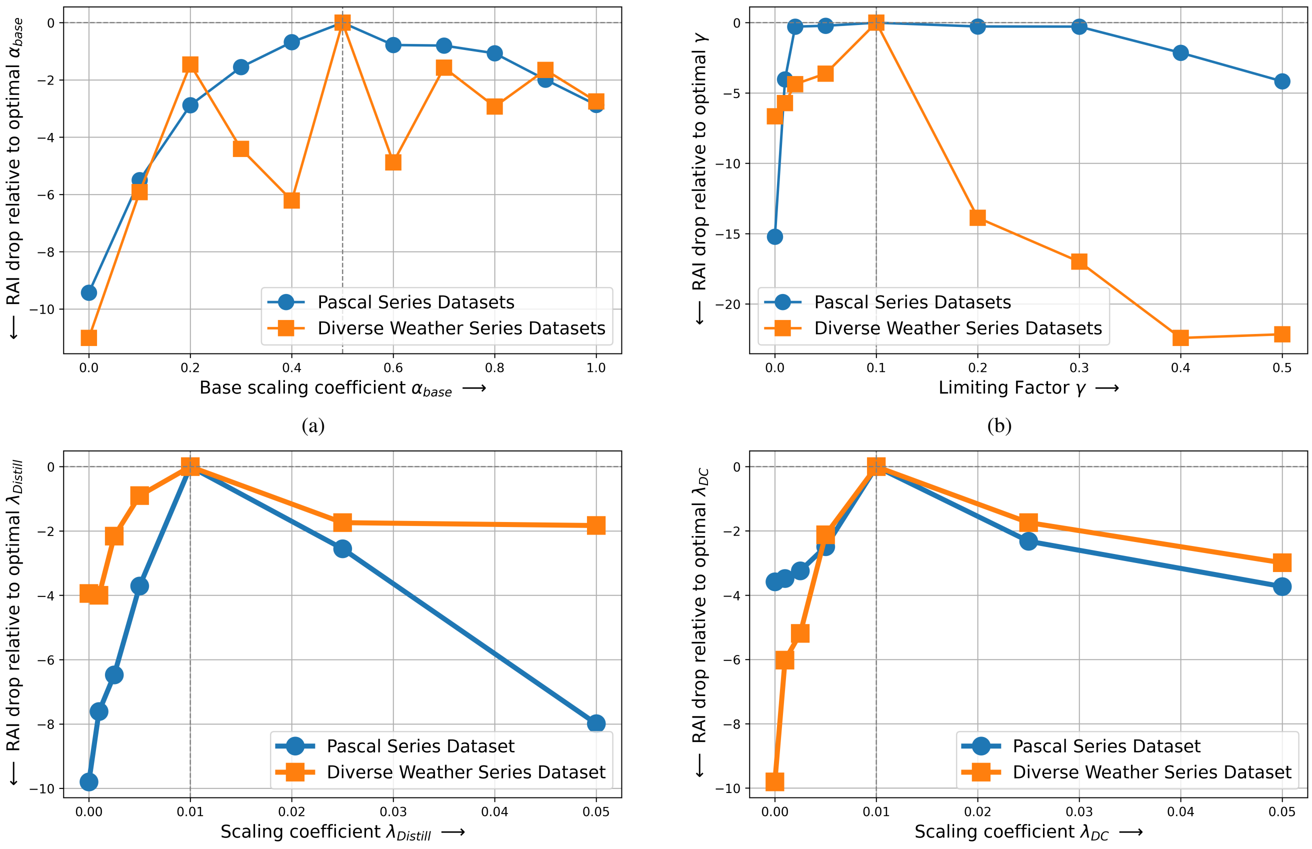
Table S3. Ablation studies of different loss components augmented with detector loss ($\mathcal{L}_{\text{Detector}}$).

Loss Component	Avg RI (%)	Avg GI (%)	RAI (%)
$\mathcal{L}_{Detector} + \mathcal{L}_{\text{Distill}}$	72.64	33.74	53.19
$\mathcal{L}_{Detector} + \mathcal{L}^*_{\text{Distill}}$	87.06	37.75	62.41
$\mathcal{L}_{Detector} + \mathcal{L}^*_{\text{Distill}} + \mathcal{L}_{\text{DC}}$	**87.44**	**44.54**	**65.99**

Impact of $\mathcal{L}^*_{Distill}$. Table S3 presents the ablation studies of different loss components augmented with detector loss ($\mathcal{L}_{\text{Detector}}$). We observe that the inclusion of $\mathcal{L}^*_{\text{Distill}}$ instead of $\mathcal{L}_{\text{Distill}}$ significantly improves all metrics, with a **+14.42%** increase in Avg RI, **+4.01%** increase in Avg GI, and **+9.22%** increase in RAI. The addition of $\mathcal{L}_{\text{DC}}$ brings in additional improvements, leading to the best performance across all metrics.

C.2. Sensitivity Analysis for key hyper-parameters

Figure S2 shows the sensitivity analysis for key hyperparameters used in the DuET approach. Base scaling coefficient α_{base} (Figure S2a) effectively controls the contributions from the old (prior task) model and current model; hence, a value of **0.5** ensures a balanced trade-off between past knowledge retention and new adaptation, while extreme values ($\alpha_{base} < 0.3$ or $\alpha_{base} > 0.7$) significantly degrade RAI. The limiting factor γ (Figure S2b) impacts task-merging, with $\gamma = 0.1$ giving optimal results for both *Pascal Series* and *Diverse Weather Series* datasets. The scaling coefficients $\lambda_{Distill}$ and λ_{DC} (Figures S2c and S2d) control the impact of Distillation and Directional Consistency losses, respectively. We observe that, for both of them, a value of **0.01** gives the best results and effectively helps in mitigating catastrophic forgetting by improving retention while preventing sign conflicts; deviations from these values lead to reduced adaptability and degraded performance



(a)

(b)

(c)

(d)

Figure S2. **Sensitivity analysis of DuET approach to key hyperparameters.** (a) shows the effect of varying base scaling coefficient α_{base}. (b) illustrates the impact of the limiting factor γ on RAI. (c) and (d) depicts the effect of varying scaling coefficients $\lambda_{Distill}$ and λ_{DC} on RAI respectively.

on both series of datasets.

Table S4. Influence of random domain and class permutations across three incremental tasks on the *Diverse Weather Series* dataset.

$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	**Avg RI**	**Avg GI**	**RAI**
Night Sunny [5:7]	Daytime Sunny [1:2]	Daytime Foggy [3:4]	83.49	51.01	67.25
Night Sunny [3:4]	Daytime Sunny [5:7]	Daytime Foggy [1:2]	80.39	51.76	66.08
Night Sunny [1:2]	Daytime Sunny [3:4]	Daytime Foggy [5:7]	88.57	41.92	65.25
Daytime Foggy [1:2]	Night Sunny [3:4]	Daytime Sunny [5:7]	78.34	50.54	64.44
Daytime Sunny [1:2]	Daytime Foggy [3:4]	Night Sunny [5:7]	88.33	35.97	62.15
	Standard Deviation		**4.12**	**6.26**	**1.72**

C.3. Influence of random class-domain order

In real-world incremental learning scenarios, the sequence in which new classes and domains are introduced can influence knowledge retention and generalisation. To check the sensitivity of the proposed DuET approach to such variations, we conducted experiments with shuffled class orders while keeping the same domain progression (top three rows in Table S4) and shuffled domain orders while keeping the same class sets (bottom three rows). We observe that Avg RI remains consistently high across all permutations ($> 78\%$), and there are minor variations in RAI and Avg RI with standard deviations of **1.72** and **4.12**, respectively. However, the slight variation in Avg GI, with a standard deviation of **6.26**, stems from domain shifts affecting generalisation. This indicates that the proposed DuET approach maintains a consistent performance irrespective of the randomness in class-domain orders.

C.4. Complexity Analysis

Table S11 presents a detailed complexity analysis of various methods evaluated in the multi-phase experiment: *Watercolour [1:3]* → *Comic [4:6]* → *Clipart [7:13]* → *VOC [14:20]*. Table S11 compares computational complexity in terms of GFLOPs, trainable parameters (in millions), average inference speed (in milliseconds), and average memory footprint (in gigabytes) across all incremental tasks. While training time (in hours) as evaluated on a single NVIDIA A100-PCIE-40GB on *Daytime Sunny [1:4]* → *Night Rainy [5:7]* experiment is reported in Table S14. The results demonstrate that DuET retains the real-time detection capabilities of YOLO11n [10], effectively transforming it into a robust real-time incremental object detector with only a minimal increase in memory footprint (**0.244 GB**) compared to its unaltered counterpart, Sequential FT (**0.235 GB**). Furthermore, since the proposed DuET approach does not modify the base detector architecture, it preserves the same GFLOPs and the number of trainable parameters.

In contrast to other model-merging algorithms [7, 8, 19, 20], which require storing task vectors—and consequently, model weights—for every task, our approach is designed to be more efficient and scalable. DuET maintains only two shared task vectors at any given task: τ_{old} (derived from the previous phase's model weights) and τ_{curr} (derived from the current phase's model weights), along with the pre-trained model weights. This design utilizes the fact that knowledge from earlier tasks, $\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_{t-2}$, is already encapsulated within the previous phase's weights, θs_{t-1}. Consequently, DuET avoids the overhead of maintaining a complete history of task vectors, resulting in a consistent memory footprint across all incremental tasks ($\mathcal{T}_t, t \geq 2$), while in case of other TA approaches, the memory footprint grows linearly with the number of tasks (Figure S3).

D. Implementation Details

Our implementation is primarily based on the Ultralytics framework[1](v8.3.9), with YOLO11n [10] primarily serving as the base detector, also extending to other variants of YOLO11 and RT-DETR [18]. Following the default configuration provided by Ultralytics, we used AdamW [17] optimiser with auto lr find and OneCycleLR scheduler, keeping a batch size of 64. For every task, we trained the detector for 100 epochs, keeping five warm-up epochs with a higher initial learning rate by a factor of 10. For the base task ($t = 1$), we use the default weight decay of 0.0005, while for incremental tasks ($t \geq 2$), we slightly increase it to 0.001 to prevent overfitting to new tasks and help prevent catastrophic forgetting. The same protocol is used while preparing other baselines for a fair comparison. Moreover, unlike LDB [22] and CL-DETR [16], we keep all layers trainable during incremental training to ensure that shared task vectors effectively capture the shift in shared knowledge across incremental tasks.

E. Comprehensive Results

E.1. Detailed analysis of Quantitative Results:

Tables S7 to S14 present the comprehensive results of various methods on different DuIOD experiments across multiple base detectors. We conducted a total of seven DuIOD experiments—five two-phase and two multi-phase experiments—three from the *Pascal Series* and the remaining four from the *Diverse Weather Series* datasets.

We provide detailed results for the five two-phase experiments: *VOC[1:10]* → *Clipart[11:20]* (Table S7), *Clipart [1:10]* → *VOC [11:20]* (Table S8), *Daytime Sunny [1:4]* → *Night Sunny [5:7]* (Table S12), *Night Sunny [1:4]* → *Daytime Sunny [5:7]* (Table S9) and *Daytime Sunny [1:4]* → *Night Rainy [5:7]* (Table S14). Meanwhile, the results for the two multi-phase experiments—*Watercolor [1:3]* → *Comic [4:6]* → *Clipart [7:13]* → *VOC [14:20]* and *Night Sunny [1:2]* → *Daytime Sunny [3:4]* → *Daytime Foggy [5:7]*—are presented in Tables S10 and S13, respectively.

We evaluate DuET across all DuIOD experiments using five detection backbones: DeformableDETR [26], YOLO11n & YOLO11x [10], and RTDETR-l & RTDETR-x [18]. Our results show that DuET consistently outperforms the baselines in nearly all DuIOD experiments, demonstrating its effectiveness in addressing the DuIOD task. Notably, DuET outperforms both CL-DETR [16] & LDB [22] on their respective backbones (Deformable DETR [26] & ViTDet [12]) with a **+5.97%** and **+11.98%** RAI gain, preserving **80.3%** vs. **66.85%** and **50.99%** vs. **41.74%** Avg RI respectively (refer Tables S7 to S14). This indicates that **gains are method-specific and not backbone dependent**.

E.2. Qualitative Visualizations

In Figure S4, we present qualitative visualisations for various methods on the task sequence: *Watercolour [1:3]* → *Comic [4:6]* → *Clipart [7:13]* → *VOC [14:20]*. The detection results are shown for unseen classes: Watercolour [4:6], Comic [1:3], Clipart [1:6], and VOC [1:13] in the final task, $\mathcal{T}_4$. Our observations indicate that DuET consistently outperforms other methods by accurately detecting most objects across different domains and classes. Notably, in the second row (Comic [1:3]), the `bicycle` class, which was learned in $\mathcal{T}_1$ (Watercolor [1:3]), and the `person` class, introduced in $\mathcal{T}_2$ (Comic [4:6]), are examined. We observe that only DuET successfully retains the knowledge from $\mathcal{T}_1$ and correctly detects the `bicycle` class in the unseen Comic [1:3] domain. A similar trend is observed in the fourth row (VOC [1:13]). Additionally, in the third row, the `car` class, introduced in $\mathcal{T}_1$, is not detected by other methods in the unseen Clipart [1:6] domain, whereas DuET consistently identifies it. These qualitative results further emphasise DuET's effectiveness in adapting to unseen classes across different domains in the DuIOD task. A similar trend is observed in the *Diverse Weather Series* (Figure S5), where DuET consistently outperforms other methods by accurately detecting most objects across various domains and classes.

[1] https://github.com/ultralytics/ultralytics

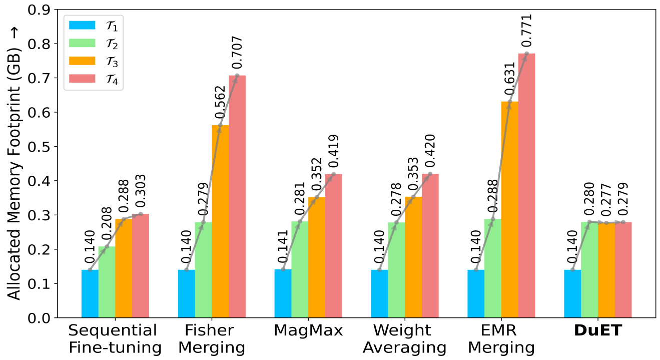


Figure S3. Comparison of allocated memory footprint (in GB) for various model-merging approaches on the multi-phase experiment with four tasks: Watercolor [1:3] ($\mathcal{T}_1$) → Comic [4:6] ($\mathcal{T}_2$) → Clipart [7:13] ($\mathcal{T}_3$) → VOC [14:20] ($\mathcal{T}_4$).

Table S5. **Dataset Statistics:** Class-wise distribution across different domains in *Pascal Series* datasets.

Class ID	Class Name	Watercolor [9]	Comic [9]	Clipart [9]	VOC [4]
1	bicycle	✓	✓	✓	✓
2	bird	✓	✓	✓	✓
3	car	✓	✓	✓	✓
4	cat	✓	✓	✓	✓
5	dog	✓	✓	✓	✓
6	person	✓	✓	✓	✓
7	aeroplane			✓	✓
8	boat			✓	✓
9	bottle			✓	✓
10	bus			✓	✓
11	chair			✓	✓
12	cow			✓	✓
13	diningtable			✓	✓
14	horse			✓	✓
15	motorbike			✓	✓
16	pottedplant			✓	✓
17	sheep			✓	✓
18	sofa			✓	✓
19	train			✓	✓
20	tvmonitor			✓	✓
Total Classes		6	6	20	20
Train Images		1000	1000	500	16551
Val Images		1000	1000	500	4952

Table S6. **Dataset Statistics:** Class-wise distribution across different weather conditions in *Diverse Weather Series* datasets.

Class ID	Class Name	Daytime Sunny [25]	Night Sunny [25]	Daytime Foggy [2, 6]
1	bike	✓	✓	✓
2	bus	✓	✓	✓
3	car	✓	✓	✓
4	motor	✓	✓	✓
5	person	✓	✓	✓
6	rider	✓	✓	✓
7	truck	✓	✓	✓
Total Classes		7	7	7
Train Images		19317	25868	1829
Val Images		8289	7756	688

F. Background Shift

Background shift is a major issue in IOD scenarios [15, 21], where previously learned object categories, if unannotated in subsequent tasks, are treated as background. DuET mitigates this shift by explicitly decomposing model parameters into shared and task-specific components, and then merging these through TA with dynamic, layer-wise retention and adaptation weights (see Section 3.4). This strategy ensures that parameters crucial for previously learned object representations remain stable, thus preventing catastrophic forgetting and minimising the likelihood of previously learned objects being erroneously classified as background when they become unlabeled in subsequent tasks.

G. Dataset Statistics

Following prior works [3, 11, 22, 23], we evaluate the DuET approach on two dataset series: the *Pascal Series* and the *Diverse Weather Series*, which cover diverse environmental conditions and domain variations, respectively. Table S5 presents the class-wise distribution, capturing cross-domain variations across four different domains: Watercolour, Comic, Clipart, and VOC. Following [11, 14], we combined the PASCAL VOC 2007 and 2012 [4] datasets to form the VOC domain, while the Watercolour, Comic, and Clipart domains were taken from [9]. Watercolour and Comic domains consist of six object categories, forming a subset of the 20 object categories present in Clipart and VOC. We used class splits of $10+10$ and $3+3+7+7$ with Class IDs as mentioned in Table S5 to conduct two-phase and multi-phase DuIOD experiments on the *Pascal Series* datasets, respectively. Similarly, Table S6 presents the class-wise distribution across three different weather conditions: Daytime Sunny, Night Sunny, and Daytime Foggy. Following [11, 22], the datasets are taken from BDD100k [25], Foggy Cityscapes [2], and Adverse Weather [6]. Since each domain contains a common set of seven classes, we used class splits of $4 + 3$ and $2 + 2 + 3$, with Class IDs sorted in alphabetical order (as shown in Table S6), to perform two-phase and multi-phase DuIOD experiments on the *Diverse Weather Series* datasets, respectively.

H. Limitations and future works

Since DuET is a task vector-based model-merging approach, it inherits the limitations of existing task vector-based methods, and hence it cannot be generalised to models trained from scratch and requires access to pre-trained object detectors to calculate shared task vectors at each incremental task. This is a common limitation of task vector-based methods. Moreover, DuET merges shared task vectors through a weighted linear interpolation mechanism, which may be suboptimal for highly heterogeneous class shifts or extreme domain variations across incremental tasks. Future work could explore more sophisticated non-linear merging approaches to better capture the shared knowledge across tasks.

Table S7. Results of various methods on VOC [1:10] → Clipart [11:20] with different base detectors. Among columns, best in **bold**, second best *underlined*.

Method	Base Detector	T1	T2: Clipart [11:20]				Avg RI (%)	Avg GI (%)	RAI (%)
		VOC	Old	New	Unseen				
		[1:10]	VOC [1:10]	Clipart [11:20]	Clipart [1:10]	VOC [11:20]			
LDB [22]	ViTDet	74.9$_{\pm0.7}$	50.10$_{\pm0.5}$	22.30$_{\pm0.8}$	8.90$_{\pm0.4}$	9.60$_{\pm0.6}$	66.89$_{\pm0.5}$	18.76$_{\pm0.3}$	42.83$_{\pm0.4}$
DuET (Ours)	ViTDet	74.9$_{\pm0.2}$	54.20$_{\pm0.4}$	17.60$_{\pm0.2}$	17.90$_{\pm0.3}$	11.80$_{\pm0.4}$	72.36$_{\pm0.2}$	32.63$_{\pm0.3}$	52.50$_{\pm0.2}$
CL-DETR [16]	Deformable DETR	56.1$_{\pm0.6}$	38.29$_{\pm0.7}$	9.04$_{\pm0.5}$	9.22$_{\pm0.8}$	3.02$_{\pm0.4}$	68.29$_{\pm0.3}$	40.72$_{\pm0.4}$	54.51$_{\pm0.3}$
DuET (Ours)	Deformable DETR	56.1$_{\pm0.6}$	42.32$_{\pm0.3}$	4.10$_{\pm0.2}$	15.63$_{\pm0.4}$	1.68$_{\pm0.4}$	75.48$_{\pm0.6}$	72.37$_{\pm0.3}$	*73.93*$_{\pm0.5}$
Sequential FT	RTDETR-l	87.1$_{\pm0.6}$	0.00$_{\pm0.0}$	55.00$_{\pm0.7}$	0.00$_{\pm0.0}$	32.20$_{\pm0.5}$	0.00$_{\pm0.0}$	18.85$_{\pm0.6}$	9.43$_{\pm0.4}$
LwF [13]	RTDETR-l	87.1$_{\pm0.6}$	3.13$_{\pm0.2}$	25.90$_{\pm0.8}$	1.30$_{\pm0.3}$	18.50$_{\pm0.4}$	3.59$_{\pm0.7}$	12.30$_{\pm0.5}$	7.95$_{\pm0.6}$
ERD [5]	RTDETR-l	87.1$_{\pm0.6}$	1.74$_{\pm0.3}$	56.00$_{\pm0.8}$	1.42$_{\pm0.2}$	37.80$_{\pm0.5}$	2.00$_{\pm0.1}$	23.74$_{\pm0.6}$	12.87$_{\pm0.8}$
DuET (Ours)	RTDETR-l	87.1$_{\pm0.6}$	46.10$_{\pm0.1}$	68.00$_{\pm0.2}$	28.10$_{\pm0.4}$	62.20$_{\pm0.4}$	52.93$_{\pm0.7}$	68.20$_{\pm0.5}$	60.57$_{\pm0.6}$
Sequential FT	RTDETR-x	89.2$_{\pm0.5}$	0.00$_{\pm0.0}$	56.52$_{\pm0.7}$	0.33$_{\pm0.2}$	30.43$_{\pm0.6}$	0.00$_{\pm0.0}$	17.66$_{\pm0.4}$	8.83$_{\pm0.3}$
LwF [13]	RTDETR-x	89.2$_{\pm0.5}$	20.40$_{\pm0.5}$	28.80$_{\pm0.3}$	17.00$_{\pm0.7}$	17.90$_{\pm0.4}$	22.87$_{\pm0.6}$	28.27$_{\pm0.5}$	25.57$_{\pm0.2}$
ERD [5]	RTDETR-x	89.2$_{\pm0.5}$	22.60$_{\pm0.7}$	55.20$_{\pm0.8}$	3.83$_{\pm0.3}$	32.80$_{\pm0.4}$	25.34$_{\pm0.5}$	22.73$_{\pm0.1}$	24.04$_{\pm0.8}$
DuET (Ours)	RTDETR-x	89.2$_{\pm0.5}$	60.50$_{\pm0.5}$	26.80$_{\pm0.4}$	49.00$_{\pm0.3}$	22.50$_{\pm0.2}$	67.83$_{\pm0.4}$	64.93$_{\pm0.6}$	66.38$_{\pm0.7}$
Sequential FT	YOLO11n	80.4$_{\pm0.3}$	0.60$_{\pm0.1}$	36.70$_{\pm0.8}$	1.02$_{\pm0.3}$	17.40$_{\pm0.4}$	0.75$_{\pm0.2}$	12.86$_{\pm0.4}$	6.81$_{\pm0.3}$
LwF [13]	YOLO11n	80.4$_{\pm0.3}$	58.40$_{\pm0.8}$	3.96$_{\pm0.3}$	28.60$_{\pm0.7}$	5.00$_{\pm0.2}$	72.64$_{\pm0.3}$	33.74$_{\pm0.5}$	53.19$_{\pm0.4}$
ERD [5]	YOLO11n	80.4$_{\pm0.3}$	55.20$_{\pm0.4}$	20.60$_{\pm0.7}$	30.50$_{\pm0.5}$	16.70$_{\pm0.8}$	68.66$_{\pm0.4}$	43.68$_{\pm0.3}$	56.17$_{\pm0.6}$
DuET (Ours)	YOLO11n	80.4$_{\pm0.3}$	70.30$_{\pm0.3}$	8.45$_{\pm0.3}$	33.80$_{\pm0.3}$	12.80$_{\pm0.3}$	**87.44**$_{\pm0.2}$	44.54$_{\pm0.1}$	65.99$_{\pm0.3}$
Sequential FT	YOLO11x	88.4$_{\pm0.5}$	0.00$_{\pm0.0}$	43.50$_{\pm0.8}$	0.00$_{\pm0.0}$	16.10$_{\pm0.6}$	0.00$_{\pm0.0}$	10.13$_{\pm0.7}$	5.07$_{\pm0.4}$
LwF [13]	YOLO11x	88.4$_{\pm0.5}$	57.30$_{\pm0.5}$	38.00$_{\pm0.3}$	40.30$_{\pm0.7}$	30.30$_{\pm0.4}$	64.82$_{\pm0.6}$	*74.57*$_{\pm0.2}$	69.70$_{\pm0.3}$
ERD [5]	YOLO11x	88.4$_{\pm0.5}$	23.70$_{\pm0.7}$	46.80$_{\pm0.8}$	26.00$_{\pm0.5}$	23.60$_{\pm0.3}$	26.81$_{\pm0.4}$	50.66$_{\pm0.2}$	38.74$_{\pm0.8}$
DuET (Ours)	YOLO11x	88.4$_{\pm0.5}$	74.30$_{\pm0.3}$	52.40$_{\pm0.2}$	44.50$_{\pm0.1}$	46.80$_{\pm0.1}$	*84.05*$_{\pm0.5}$	**90.73**$_{\pm0.3}$	**87.39**$_{\pm0.6}$

Table S8. Results of various methods on Clipart [1:10] → VOC [11:20] with different base detectors. Among columns, best in **bold**, second best *underlined*.

Method	Base Detector	T1	T2: VOC [11:20]				Avg RI (%)	Avg GI (%)	RAI (%)
		Clipart	Old	New	Unseen				
		[1:10]	Clipart [1:10]	VOC [11:20]	VOC [1:10]	Clipart [11:20]			
LDB [22]	ViTDet	36.4$_{\pm0.4}$	16.30$_{\pm0.3}$	23.80$_{\pm0.5}$	7.10$_{\pm0.2}$	9.10$_{\pm0.6}$	44.78$_{\pm0.7}$	15.81$_{\pm0.4}$	30.30$_{\pm0.8}$
DuET (Ours)	ViTDet	36.4$_{\pm0.2}$	31.20$_{\pm0.3}$	34.50$_{\pm0.3}$	24.40$_{\pm0.3}$	1.60$_{\pm0.1}$	85.71$_{\pm0.1}$	18.23$_{\pm0.2}$	51.97$_{\pm0.3}$
CL-DETR [16]	Deformable DETR	10.5$_{\pm0.6}$	8.88$_{\pm0.7}$	27.02$_{\pm0.4}$	3.88$_{\pm0.3}$	10.33$_{\pm0.8}$	84.57$_{\pm0.5}$	**54.35**$_{\pm0.2}$	*69.46*$_{\pm0.7}$
DuET (Ours)	Deformable DETR	10.5$_{\pm0.2}$	9.54$_{\pm0.1}$	20.08$_{\pm0.1}$	3.17$_{\pm0.2}$	10.23$_{\pm0.2}$	**90.86**$_{\pm0.1}$	*53.22*$_{\pm0.2}$	**72.04**$_{\pm0.2}$
Sequential FT	RTDETR-l	44.2$_{\pm0.5}$	0.00$_{\pm0.0}$	81.50$_{\pm0.7}$	0.00$_{\pm0.0}$	30.80$_{\pm0.6}$	0.00$_{\pm0.0}$	29.79$_{\pm0.8}$	14.90$_{\pm0.4}$
LwF [13]	RTDETR-l	44.2$_{\pm0.4}$	2.81$_{\pm0.2}$	66.00$_{\pm0.7}$	0.73$_{\pm0.3}$	37.20$_{\pm0.5}$	6.36$_{\pm0.6}$	36.39$_{\pm0.4}$	21.38$_{\pm0.8}$
ERD [5]	RTDETR-l	44.2$_{\pm0.5}$	0.37$_{\pm0.3}$	81.20$_{\pm0.6}$	2.81$_{\pm0.4}$	27.50$_{\pm0.7}$	0.84$_{\pm0.2}$	28.21$_{\pm0.8}$	14.53$_{\pm0.5}$
DuET (Ours)	RTDETR-l	44.2$_{\pm0.1}$	37.80$_{\pm0.2}$	8.17$_{\pm0.1}$	29.40$_{\pm0.2}$	13.20$_{\pm0.2}$	85.52$_{\pm0.1}$	29.64$_{\pm0.2}$	*57.58*$_{\pm0.1}$
Sequential FT	RTDETR-x	47.0$_{\pm0.6}$	0.00$_{\pm0.0}$	81.60$_{\pm0.7}$	0.00$_{\pm0.0}$	35.70$_{\pm0.5}$	0.00$_{\pm0.0}$	37.27$_{\pm0.8}$	18.64$_{\pm0.3}$
LwF [13]	RTDETR-x	47.0$_{\pm0.5}$	2.42$_{\pm0.3}$	64.30$_{\pm0.6}$	1.25$_{\pm0.2}$	35.70$_{\pm0.8}$	5.15$_{\pm0.4}$	37.97$_{\pm0.7}$	21.56$_{\pm0.5}$
ERD [5]	RTDETR-x	47.0$_{\pm0.4}$	0.67$_{\pm0.2}$	82.00$_{\pm0.7}$	0.93$_{\pm0.3}$	34.40$_{\pm0.6}$	1.43$_{\pm0.5}$	36.43$_{\pm0.8}$	18.93$_{\pm0.4}$
DuET (Ours)	RTDETR-x	47.0$_{\pm0.2}$	41.10$_{\pm0.2}$	4.64$_{\pm0.2}$	21.30$_{\pm0.1}$	5.27$_{\pm0.2}$	*87.45*$_{\pm0.2}$	17.44$_{\pm0.2}$	52.45$_{\pm0.2}$
Sequential FT	YOLO11n	47.1$_{\pm0.7}$	0.00$_{\pm0.0}$	73.60$_{\pm0.6}$	0.00$_{\pm0.0}$	29.10$_{\pm0.5}$	0.00$_{\pm0.0}$	30.12$_{\pm0.8}$	15.06$_{\pm0.4}$
LwF [13]	YOLO11n	47.1$_{\pm0.6}$	31.40$_{\pm0.7}$	4.00$_{\pm0.5}$	20.30$_{\pm0.3}$	5.36$_{\pm0.8}$	66.67$_{\pm0.4}$	18.17$_{\pm0.2}$	42.42$_{\pm0.7}$
ERD [5]	YOLO11n	47.1$_{\pm0.5}$	33.20$_{\pm0.3}$	0.72$_{\pm0.6}$	20.70$_{\pm0.4}$	0.63$_{\pm0.7}$	70.49$_{\pm0.2}$	13.53$_{\pm0.8}$	42.01$_{\pm0.5}$
DuET (Ours)	YOLO11n	47.1$_{\pm0.2}$	32.70$_{\pm0.1}$	44.00$_{\pm0.2}$	21.70$_{\pm0.2}$	26.10$_{\pm0.1}$	69.43$_{\pm0.2}$	40.51$_{\pm0.2}$	54.97$_{\pm0.1}$
Sequential FT	YOLO11x	36.3$_{\pm0.5}$	0.00$_{\pm0.0}$	77.50$_{\pm0.6}$	0.00$_{\pm0.0}$	33.50$_{\pm0.8}$	0.00$_{\pm0.0}$	38.15$_{\pm0.7}$	19.08$_{\pm0.4}$
LwF [13]	YOLO11x	36.3$_{\pm0.4}$	25.10$_{\pm0.3}$	0.96$_{\pm0.8}$	13.00$_{\pm0.5}$	1.59$_{\pm0.6}$	69.15$_{\pm0.4}$	9.16$_{\pm0.7}$	39.16$_{\pm0.3}$
ERD [5]	YOLO11x	36.3$_{\pm0.5}$	29.40$_{\pm0.7}$	0.52$_{\pm0.3}$	12.90$_{\pm0.6}$	0.68$_{\pm0.2}$	80.99$_{\pm0.8}$	8.07$_{\pm0.4}$	44.53$_{\pm0.5}$
DuET (Ours)	YOLO11x	36.3$_{\pm0.2}$	19.30$_{\pm0.1}$	1.25$_{\pm0.2}$	6.06$_{\pm0.2}$	3.03$_{\pm0.1}$	53.17$_{\pm0.2}$	6.88$_{\pm0.1}$	30.03$_{\pm0.2}$

Table S9. Results of various methods on Daytime Sunny [1:4] → Night Sunny [5:7] with different base detectors. Among columns, best in **bold**, second best *underlined*.

Method	Base Detector	T1 Daytime Sunny [1:4]	T2: Night Sunny [5:7]				Avg RI (%)	Avg GI (%)	RAI (%)
			Old	New	Unseen				
			Daytime Sunny [1:4]	Night Sunny [5:7]	Night Sunny [1:4]	Daytime Sunny [5:7]			
LDB [22]	VitDet	45.3$_{\pm 0.6}$	0.50$_{\pm 0.3}$	15.10$_{\pm 0.4}$	0.30$_{\pm 0.5}$	16.90$_{\pm 0.7}$	1.10$_{\pm 0.2}$	22.41$_{\pm 0.3}$	11.76$_{\pm 0.6}$
DuET (Ours)	VitDet	45.3$_{\pm 0.2}$	12.48$_{\pm 0.3}$	11.60$_{\pm 0.3}$	4.33$_{\pm 0.2}$	9.60$_{\pm 0.2}$	27.55$_{\pm 0.2}$	28.22$_{\pm 0.1}$	27.89$_{\pm 0.2}$
CL-DETR [16]	Deformable DETR	46.3$_{\pm 0.4}$	27.41$_{\pm 0.5}$	31.94$_{\pm 0.6}$	19.85$_{\pm 0.3}$	32.55$_{\pm 0.4}$	59.20$_{\pm 0.2}$	*54.96*$_{\pm 0.5}$	57.08$_{\pm 0.4}$
DuET (Ours)	Deformable DETR	46.3$_{\pm 0.2}$	39.1$_{\pm 0.1}$	15.06$_{\pm 0.2}$	28.17$_{\pm 0.2}$	4.33$_{\pm 0.1}$	84.45$_{\pm 0.2}$	33.45$_{\pm 0.1}$	58.95$_{\pm 0.2}$
Sequential FT	RTDETR-l	57.2$_{\pm 0.5}$	0.00$_{\pm 0.0}$	77.40$_{\pm 0.7}$	2.52$_{\pm 0.3}$	39.80$_{\pm 0.6}$	0.00$_{\pm 0.0}$	35.36$_{\pm 0.8}$	17.68$_{\pm 0.4}$
LwF [13]	RTDETR-l	57.2$_{\pm 0.4}$	0.15$_{\pm 0.2}$	76.40$_{\pm 0.7}$	0.03$_{\pm 0.1}$	41.50$_{\pm 0.5}$	0.26$_{\pm 0.2}$	35.01$_{\pm 0.8}$	17.64$_{\pm 0.4}$
ERD [5]	RTDETR-l	57.2$_{\pm 0.5}$	0.09$_{\pm 0.1}$	80.50$_{\pm 0.7}$	0.04$_{\pm 0.1}$	39.80$_{\pm 0.6}$	0.16$_{\pm 0.2}$	33.59$_{\pm 0.8}$	16.88$_{\pm 0.4}$
DuET (Ours)	RTDETR-l	57.2$_{\pm 0.2}$	27.30$_{\pm 0.1}$	8.63$_{\pm 0.2}$	20.10$_{\pm 0.2}$	7.88$_{\pm 0.1}$	47.73$_{\pm 0.2}$	21.00$_{\pm 0.1}$	34.37$_{\pm 0.2}$
Sequential FT	RTDETR-x	61.0$_{\pm 0.6}$	0.00$_{\pm 0.0}$	84.80$_{\pm 0.7}$	0.00$_{\pm 0.0}$	40.80$_{\pm 0.5}$	0.00$_{\pm 0.0}$	33.77$_{\pm 0.8}$	16.89$_{\pm 0.3}$
LwF [13]	RTDETR-x	61.0$_{\pm 0.5}$	0.57$_{\pm 0.2}$	79.10$_{\pm 0.7}$	0.61$_{\pm 0.3}$	40.60$_{\pm 0.6}$	0.93$_{\pm 0.2}$	34.03$_{\pm 0.8}$	17.48$_{\pm 0.4}$
ERD [5]	RTDETR-x	61.0$_{\pm 0.4}$	0.81$_{\pm 0.2}$	84.80$_{\pm 0.7}$	0.98$_{\pm 0.3}$	38.90$_{\pm 0.6}$	1.33$_{\pm 0.2}$	32.87$_{\pm 0.8}$	17.10$_{\pm 0.4}$
DuET (Ours)	RTDETR-x	61.0$_{\pm 0.2}$	34.40$_{\pm 0.1}$	6.51$_{\pm 0.2}$	28.10$_{\pm 0.2}$	6.02$_{\pm 0.1}$	56.39$_{\pm 0.2}$	24.15$_{\pm 0.1}$	40.27$_{\pm 0.2}$
Sequential FT	YOLO11n	49.4$_{\pm 0.3}$	0.00$_{\pm 0.0}$	62.20$_{\pm 0.5}$	12.60$_{\pm 0.4}$	35.90$_{\pm 0.3}$	0.00$_{\pm 0.0}$	45.88$_{\pm 0.6}$	22.94$_{\pm 0.3}$
LwF [13]	YOLO11n	49.4$_{\pm 0.2}$	27.60$_{\pm 0.4}$	0.34$_{\pm 0.6}$	21.30$_{\pm 0.3}$	0.67$_{\pm 0.5}$	55.87$_{\pm 0.3}$	21.88$_{\pm 0.7}$	38.88$_{\pm 0.6}$
ERD [5]	YOLO11n	49.4$_{\pm 0.5}$	33.00$_{\pm 0.4}$	34.00$_{\pm 0.3}$	26.10$_{\pm 0.6}$	29.10$_{\pm 0.7}$	66.80$_{\pm 0.5}$	53.04$_{\pm 0.3}$	59.92$_{\pm 0.4}$
DuET (Ours)	YOLO11n	49.4$_{\pm 0.2}$	43.50$_{\pm 0.1}$	22.20$_{\pm 0.3}$	31.60$_{\pm 0.2}$	27.40$_{\pm 0.1}$	88.06$_{\pm 0.2}$	**56.95**$_{\pm 0.1}$	**72.51**$_{\pm 0.2}$
Sequential FT	YOLO11x	64.2$_{\pm 0.6}$	12.50$_{\pm 0.4}$	68.60$_{\pm 0.7}$	18.80$_{\pm 0.3}$	46.00$_{\pm 0.8}$	19.47$_{\pm 0.2}$	47.60$_{\pm 0.5}$	33.54$_{\pm 0.4}$
LwF [13]	YOLO11x	64.2$_{\pm 0.7}$	62.10$_{\pm 0.6}$	0.04$_{\pm 0.2}$	42.40$_{\pm 0.8}$	0.01$_{\pm 0.1}$	*96.73*$_{\pm 0.5}$	27.79$_{\pm 0.4}$	62.26$_{\pm 0.8}$
ERD [5]	YOLO11x	64.2$_{\pm 0.6}$	62.20$_{\pm 0.7}$	0.06$_{\pm 0.2}$	42.70$_{\pm 0.8}$	0.07$_{\pm 0.1}$	95.95$_{\pm 0.5}$	28.04$_{\pm 0.4}$	62.46$_{\pm 0.8}$
DuET (Ours)	YOLO11x	64.2$_{\pm 0.2}$	61.60$_{\pm 0.1}$	12.90$_{\pm 0.2}$	44.70$_{\pm 0.2}$	17.10$_{\pm 0.1}$	**96.88**$_{\pm 0.2}$	42.41$_{\pm 0.1}$	*69.18*$_{\pm 0.2}$

Table S10. Results of various methods on Night Sunny [1:4] → Daytime Sunny [5:7] with different base detectors. Among columns, best in **bold**, second best *underlined*.

Method	Base Detector	T1 Night Sunny [1:4]	T2: Daytime Sunny [5:7]				Avg RI (%)	Avg GI (%)	RAI (%)
			Old	New	Unseen				
			Night Sunny [1:4]	Daytime Sunny [5:7]	Daytime Sunny [1:4]	Night Sunny [5:7]			
LDB [22]	ViTDet	37.0$_{\pm 0.6}$	0.40$_{\pm 0.4}$	18.30$_{\pm 0.7}$	0.10$_{\pm 0.2}$	14.30$_{\pm 0.5}$	1.08$_{\pm 0.3}$	20.66$_{\pm 0.7}$	10.87$_{\pm 0.8}$
DuET (Ours)	VitDet	37.0$_{\pm 0.2}$	5.50$_{\pm 0.1}$	18.33$_{\pm 0.1}$	3.33$_{\pm 0.5}$	14.70$_{\pm 0.2}$	14.86$_{\pm 0.1}$	24.80$_{\pm 0.2}$	19.83$_{\pm 0.1}$
CL-DETR [16]	Deformable DETR	48.8$_{\pm 0.7}$	25.90$_{\pm 0.8}$	45.70$_{\pm 0.5}$	19.88$_{\pm 0.6}$	41.33$_{\pm 0.2}$	53.04$_{\pm 0.7}$	*55.64*$_{\pm 0.3}$	54.34$_{\pm 0.8}$
DuET (Ours)	Deformable DETR	48.8$_{\pm 0.2}$	38.81$_{\pm 0.1}$	4.24$_{\pm 0.2}$	29.09$_{\pm 0.1}$	7.59$_{\pm 0.2}$	79.48$_{\pm 0.1}$	37.69$_{\pm 0.2}$	58.59$_{\pm 0.2}$
Sequential FT	RTDETR-l	70.0$_{\pm 0.6}$	0.00$_{\pm 0.0}$	58.90$_{\pm 0.7}$	0.00$_{\pm 0.0}$	40.80$_{\pm 0.5}$	0.00$_{\pm 0.0}$	24.64$_{\pm 0.8}$	12.32$_{\pm 0.3}$
LwF [13]	RTDETR-l	70.0$_{\pm 0.7}$	8.73$_{\pm 0.8}$	21.80$_{\pm 0.5}$	6.00$_{\pm 0.6}$	8.37$_{\pm 0.4}$	12.47$_{\pm 0.7}$	10.30$_{\pm 0.3}$	11.39$_{\pm 0.8}$
ERD [5]	RTDETR-l	70.0$_{\pm 0.6}$	1.09$_{\pm 0.7}$	57.60$_{\pm 0.5}$	1.86$_{\pm 0.8}$	43.10$_{\pm 0.4}$	1.56$_{\pm 0.7}$	27.65$_{\pm 0.6}$	14.61$_{\pm 0.3}$
DuET (Ours)	RTDETR-l	70.0$_{\pm 0.2}$	48.70$_{\pm 0.2}$	1.86$_{\pm 0.1}$	43.80$_{\pm 0.2}$	1.03$_{\pm 0.1}$	69.57$_{\pm 0.2}$	38.91$_{\pm 0.2}$	54.24$_{\pm 0.2}$
Sequential FT	RTDETR-x	73.3$_{\pm 0.7}$	0.00$_{\pm 0.0}$	58.80$_{\pm 0.8}$	0.00$_{\pm 0.0}$	44.10$_{\pm 0.6}$	0.00$_{\pm 0.0}$	24.39$_{\pm 0.7}$	12.20$_{\pm 0.3}$
LwF [13]	RTDETR-x	73.3$_{\pm 0.6}$	10.70$_{\pm 0.7}$	9.25$_{\pm 0.5}$	6.02$_{\pm 0.6}$	7.66$_{\pm 0.4}$	14.60$_{\pm 0.7}$	9.17$_{\pm 0.3}$	11.89$_{\pm 0.8}$
ERD [5]	RTDETR-x	73.3$_{\pm 0.7}$	7.03$_{\pm 0.8}$	57.00$_{\pm 0.5}$	0.93$_{\pm 0.6}$	42.50$_{\pm 0.4}$	9.59$_{\pm 0.7}$	24.27$_{\pm 0.8}$	16.93$_{\pm 0.3}$
DuET (Ours)	RTDETR-x	73.3$_{\pm 0.2}$	66.30$_{\pm 0.2}$	6.40$_{\pm 0.1}$	58.90$_{\pm 0.2}$	5.26$_{\pm 0.1}$	*90.45*$_{\pm 0.2}$	51.19$_{\pm 0.2}$	*70.82*$_{\pm 0.2}$
Sequential FT	YOLO11n	50.1$_{\pm 0.7}$	0.12$_{\pm 0.8}$	37.60$_{\pm 0.5}$	0.25$_{\pm 0.6}$	25.8$_{\pm 0.4}$	0.24$_{\pm 0.7}$	19.48$_{\pm 0.8}$	9.86$_{\pm 0.3}$
LwF [13]	YOLO11n	50.1$_{\pm 0.6}$	39.00$_{\pm 0.7}$	0.29$_{\pm 0.2}$	33.90$_{\pm 0.8}$	1.51$_{\pm 0.3}$	77.84$_{\pm 0.5}$	35.44$_{\pm 0.7}$	56.64$_{\pm 0.4}$
ERD [5]	YOLO11n	50.1$_{\pm 0.7}$	39.60$_{\pm 0.6}$	0.06$_{\pm 0.2}$	34.20$_{\pm 0.8}$	0.04$_{\pm 0.1}$	79.04$_{\pm 0.5}$	34.65$_{\pm 0.7}$	56.85$_{\pm 0.4}$
DuET (Ours)	YOLO11n	50.1$_{\pm 0.2}$	47.10$_{\pm 0.2}$	20.10$_{\pm 0.1}$	41.30$_{\pm 0.2}$	19.10$_{\pm 0.1}$	**94.01**$_{\pm 0.2}$	**56.03**$_{\pm 0.2}$	**75.02**$_{\pm 0.2}$
Sequential FT	YOLO11x	76.3$_{\pm 0.7}$	0.33$_{\pm 0.8}$	50.20$_{\pm 0.5}$	2.62$_{\pm 0.6}$	44.60$_{\pm 0.4}$	0.43$_{\pm 0.7}$	28.51$_{\pm 0.8}$	14.47$_{\pm 0.3}$
LwF [13]	YOLO11x	76.3$_{\pm 0.8}$	67.50$_{\pm 0.7}$	0.66$_{\pm 0.2}$	50.10$_{\pm 0.8}$	0.23$_{\pm 0.1}$	88.47$_{\pm 0.5}$	39.16$_{\pm 0.7}$	63.82$_{\pm 0.4}$
ERD [5]	YOLO11x	76.3$_{\pm 0.7}$	67.70$_{\pm 0.8}$	0.46$_{\pm 0.2}$	51.30$_{\pm 0.8}$	0.25$_{\pm 0.1}$	88.73$_{\pm 0.5}$	40.10$_{\pm 0.7}$	64.42$_{\pm 0.4}$
DuET (Ours)	YOLO11x	76.3$_{\pm 0.2}$	22.60$_{\pm 0.1}$	43.70$_{\pm 0.2}$	21.60$_{\pm 0.2}$	40.00$_{\pm 0.1}$	29.62$_{\pm 0.2}$	40.58$_{\pm 0.2}$	35.10$_{\pm 0.2}$

Table S11. Computational complexity analysis of various methods evaluated on the multi-phase experiment: Watercolor [1:3] → Comic [4:6] → Clipart [7:13] → VOC [14:20]. The inference speed and memory footprint are averaged across all incremental tasks.

Method	Base Detector	GFLOPs	Trainable Params (M)	Avg. Inference Speed (ms)	Avg. Memory Footprint (GB)
Sequential FT	YOLO11n [10]	6.3	2.58	9.150	0.235
LwF [13]	YOLO11n [10]	6.3	2.58	9.125	0.261
ERD [5]	YOLO11n [10]	6.3	2.58	9.075	0.257
LDB [22]	ViTDet [12]	1829.61	110.52	137.92	1.818
CL-DETR [16]	Deformable DETR [26]	11.77	39.85	39.075	0.789
DuET (Ours)	YOLO11n [10]	6.3	2.58	4.4	0.244

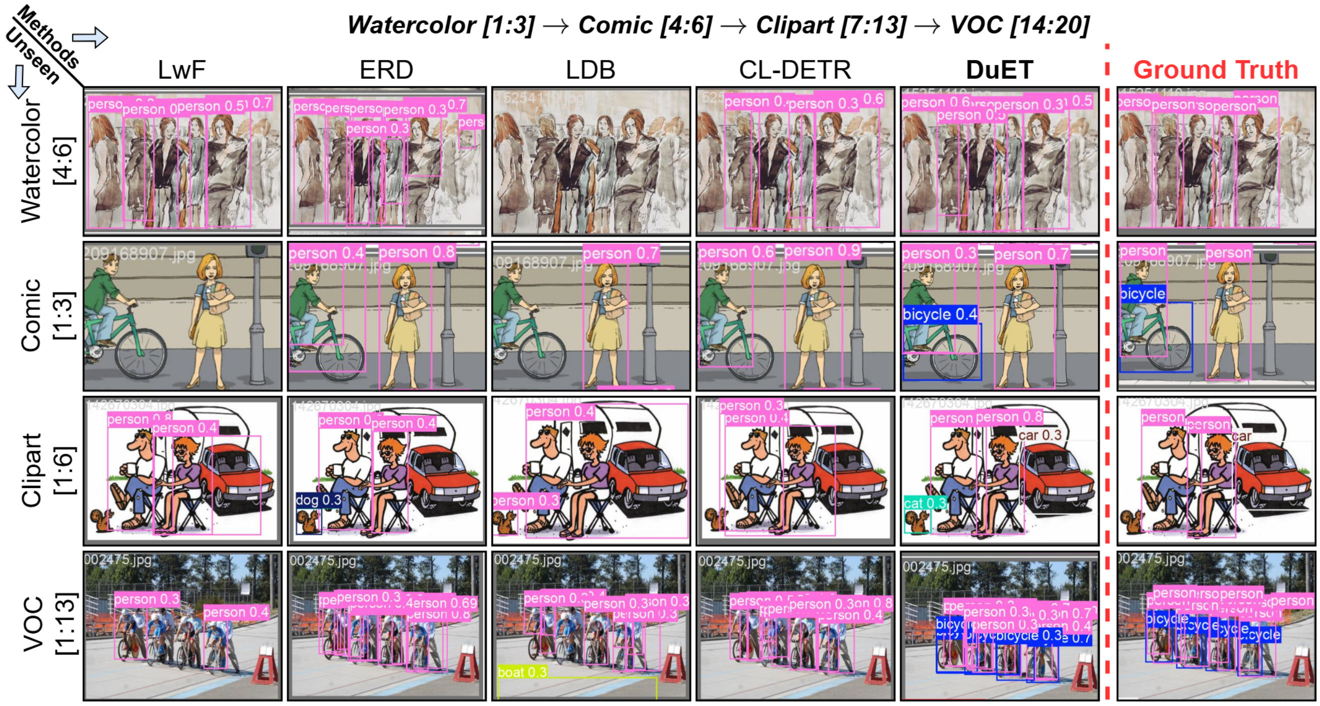


Figure S4. Qualitative comparisons on *Pascal Series* multi-phase experiment: Watercolour [1:3] → Comic [4:6] → Clipart [7:13] → VOC [14:20] for different methods on the DuIOD task. The four rows display detection results on unseen classes: Watercolour [4:6], Comic [1:3], Clipart [1:6], and VOC [1:13] on the final task $\mathcal{T}_4$. (zoomed in for best view).

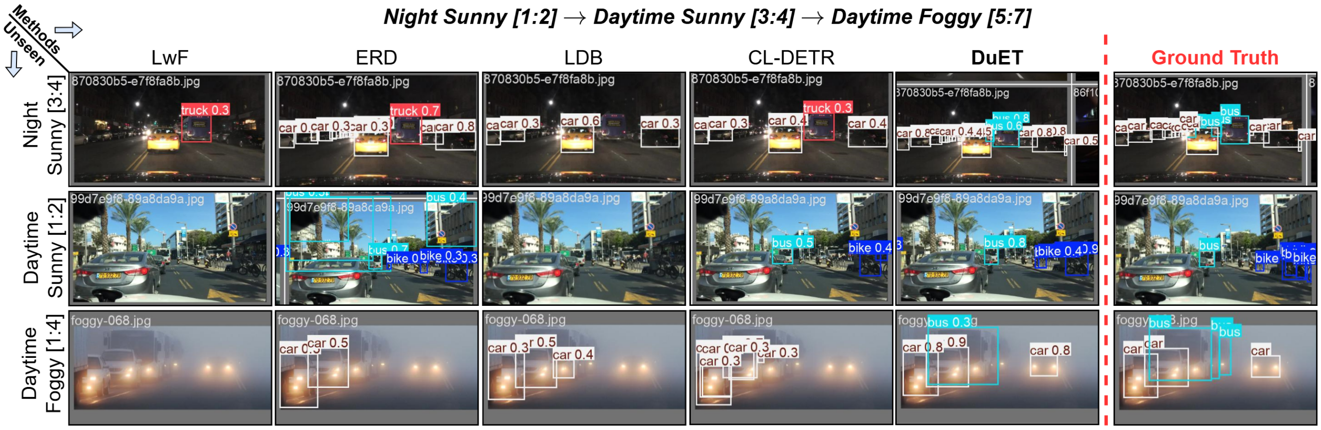


Figure S5. Qualitative comparisons on *Diverse Series* multi-phase experiment: Night Sunny [1:2] → Daytime Sunny [3:4] → Daytime Foggy [5:7] for different methods on the DuIOD task. The four rows display detection results on unseen classes: Night Sunny [3:4], Daytime Sunny [1:2], and Daytime Foggy [1:4] on the final task $\mathcal{T}_3$. (zoomed-in for best view).

Table S12. Results of various methods on Watercolor [1:3] → Comic [4:6] → Clipart [7:13] → VOC [14:20] with different base detectors. Among columns, best in **bold**, second best *underlined*.

Method	Base Detector	Avg RI (%)	Avg GI (%)	RAI (%)
LDB [22]	ViTDet	86.08±0.6	19.57±0.5	52.83±0.4
DuET (Ours)	ViTDet	65.57±0.2	40.44±0.1	53.01±0.2
CL-DETR [16]	Deformable DETR	71.73±0.5	36.63±0.6	54.18±0.4
DuET (Ours)	Deformable DETR	88.54±0.2	34.81±0.1	61.68±0.2
Sequential FT	RTDETR-l	0.00±0.0	7.76±0.4	3.88±0.3
LwF [13]	RTDETR-l	0.40±0.2	13.47±0.7	6.94±0.5
ERD [5]	RTDETR-l	1.05±0.4	17.91±0.6	9.48±0.3
DuET (Ours)	RTDETR-l	24.75±0.2	22.89±0.2	23.82±0.1
Sequential FT	RTDETR-x	0.00±0.0	6.35±0.4	3.18±0.2
LwF [13]	RTDETR-x	65.51±0.8	16.69±0.3	41.10±0.7
ERD [5]	RTDETR-x	0.05±0.1	10.42±0.2	5.24±0.3
DuET (Ours)	RTDETR-x	40.80±0.2	18.88±0.2	29.84±0.1
Sequential FT	YOLO11n	0.00±0.0	11.05±0.5	5.53±0.3
LwF [13]	YOLO11n	52.66±0.6	17.01±0.4	34.84±0.3
ERD [5]	YOLO11n	54.76±0.5	*41.13*±0.4	47.95±0.7
DuET (Ours)	YOLO11n	*89.30*±0.2	**42.60**±0.1	**65.95**±0.2
Sequential FT	YOLO11x	0.00±0.0	10.21±0.4	5.11±0.3
LwF [13]	YOLO11x	10.56±0.3	17.46±0.5	14.01±0.2
ERD [5]	YOLO11x	54.19±0.6	8.49±0.2	31.34±0.8
DuET (Ours)	YOLO11x	**96.72**±0.2	26.49±0.1	*61.61*±0.2

Table S13. Results of various methods on Night Sunny [1:2] → Daytime Sunny [3:4] → Daytime Foggy [5:7] with different base detectors. Among columns, best in **bold**, second best *underlined*.

Method	Base Detector	Avg RI (%)	Avg GI (%)	RAI (%)
LDB [22]	ViTDet	50.50±0.6	5.42±0.7	27.96±0.5
DuET (Ours)	VitDet	39.87±0.2	17.12±0.1	28.50±0.2
CL-DETR [16]	Deformable DETR	*64.26*±0.3	**43.46**±0.5	*53.86*±0.6
DuET (Ours)	Deformable DETR	62.99±0.2	45.11±0.1	54.05±0.2
Sequential FT	RTDETR-l	0.00±0.0	14.97±0.4	7.49±0.3
LwF [13]	RTDETR-l	14.76±0.2	1.24±0.3	8.00±0.5
ERD [5]	RTDETR-l	5.40±0.4	15.83±0.6	10.62±0.3
DuET (Ours)	RTDETR-l	20.76±0.2	10.55±0.2	15.66±0.1
Sequential FT	RTDETR-x	0.00±0.0	22.74±0.4	11.37±0.3
LwF [13]	RTDETR-x	7.62±0.3	8.86±0.2	8.24±0.4
ERD [5]	RTDETR-x	3.02±0.2	19.16±0.3	11.09±0.2
DuET (Ours)	RTDETR-x	27.39±0.2	23.65±0.1	25.52±0.2
Sequential FT	YOLO11n	0.00±0.0	30.51±0.6	15.26±0.5
LwF [13]	YOLO11n	27.94±0.7	23.78±0.5	25.86±0.6
ERD [5]	YOLO11n	44.60±0.6	39.40±0.5	42.00±0.3
DuET (Ours)	YOLO11n	**88.57**±0.2	*41.92*±0.1	**65.25**±0.2
Sequential FT	YOLO11x	0.00±0.0	18.16±0.4	9.08±0.3
LwF [13]	YOLO11x	19.57±0.3	28.44±0.5	24.01±0.2
ERD [5]	YOLO11x	45.85±0.6	37.38±0.2	41.62±0.8
DuET (Ours)	YOLO11x	43.46±0.2	38.37±0.1	40.92±0.2

Algorithm 1: DuET Training Algorithm

```
Input: Pre-trained model weights: $\theta_0$, Sequence of tasks:
       $\{T_1, T_2, \ldots, T_T\}$.
Output: Final model weights: $\theta_T$.
Initialize model with pre-trained weights: $\theta_0$;
for $t = 1, 2, \ldots, T$ do
    if $t = 1$ then
        Train model on task $T_1$ using $\mathcal{L}_{\text{Detector}}$;
        Update weights: $\theta_1 \leftarrow \theta_0 - \eta \cdot \nabla_\theta \mathcal{L}_{\text{Detector}}$;
        Decompose weights:
        $\theta_0 \rightarrow [\theta_{s_0}, \theta_{\tau_0}], \quad \theta_1 \rightarrow [\theta_{s_1}, \theta_{\tau_1}]$;
        Compute shared task vector: $\tau_{s_1} = \theta_{s_1} - \theta_{s_0}$;
        Compute task-specific task vector: $\tau_{\tau_1} \leftarrow \theta_{\tau_1}$;
    else
        Initialize: $\theta_t \leftarrow \theta_{t-1}$ ;
        Train model on task $T_t$ using $\mathcal{L}_{\text{Total}}$;
        Update weights: $\theta_t \leftarrow \theta_{t-1} - \eta \cdot \nabla_\theta \mathcal{L}_{\text{Total}}$;
        Decompose weights:
        $\theta_{t-1} \rightarrow [\theta_{s_{t-1}}, \theta_{\tau_{t-1}}], \quad \theta_t \rightarrow [\theta_{s_t}, \theta_{\tau_t}]$;
        Compute shared task vectors:
        $\tau_{\text{old}} = \tau_{s_{t-1}} = \theta_{s_{t-1}} - \theta_{s_0}$;
        $\tau_{\text{curr}} = \tau_{s_t} = \theta_{s_t} - \theta_{s_0}$;
        Compute task-specific task vectors:
        $\tau_{t_{\text{old}}} = \theta_{\tau_{t-1}}, \quad \tau_{t_{\text{curr}}} = \theta_{\tau_t}$;
        Update shared weights using DuET:
          $(\theta_{s_t})_{\text{incre}} \leftarrow \textbf{DuET}(\tau_{\text{curr}}, \tau_{\text{old}}, \theta_{s_0})$;
        Update task-specific weights:
          $(\theta_{\tau_t})_{\text{incre}} \leftarrow [\tau_{t_{\text{old}}}, \tau_{t_{\text{curr}}}]$;
        Load new updated weights:
          $\theta_t \leftarrow [(\theta_{s_t})_{\text{incre}}, (\theta_{\tau_t})_{\text{incre}}]$;
    end
end
return $\theta_T$;
```

Algorithm 2: DuET Task Arithmetic Algorithm

```
Input: Parameters: Shared pre-trained weights: $\theta_{s_0}$, Old
       Task Vector: $\tau_{\text{old}}$, Current Task Vector: $\tau_{\text{curr}}$
       Hyperparameters: Limiting Factor: $\gamma$, Base
       Scaling Coefficient: $\alpha_{\text{base}}$, Numerical Stability
       constant: $\epsilon$
Output: Updated inremental shared weights: $\theta_{s_t}^{\text{incre}}$
for Model Layer: $l = 1, 2, \ldots, L$ do
    $p_l = \frac{\|\tau_{\text{old}}^l\| - \|\tau_{\text{curr}}^l\|}{\|\tau_{\text{old}}^l + \tau_{\text{curr}}^l\| + \epsilon}$,
    $\delta_l = \gamma \cdot \tanh(p_l)$
    $\alpha_l = \alpha_{\text{base}} + \text{clamp}\,(\delta_l, -\gamma, \gamma)$
    $\beta_l = 1 - \alpha_l$
    $(\theta_{s_t}^l)_{\text{incre}} = \theta_{s_0}^l + \alpha_l \cdot \tau_{\text{old}}^l + \beta_l \cdot \tau_{\text{curr}}^l$
end
$(\theta_{s_t})_{\text{incre}} = \{(\theta_{s_t}^l)_{\text{incre}}\}_{l=1}^L$
return $(\theta_{s_t})_{\text{incre}}$
```

Table S14. Results of various methods on Daytime Sunny [1:4] → Night Rainy [5:7] with different base detectors. Among columns, best in **bold**, second best *underlined*. Training time is reported on single NVIDIA A100-PCIE-40GB.

Method	Base Detector	Training Time ($\mathcal{T}_1 + \mathcal{T}_2$) (hours)	T1 Daytime Sunny [1:4]	T2: Night Rainy [5:7]				Avg RI (%)	Avg GI (%)	RAI (%)
				Old	New	Unseen				
				Daytime Sunny [1:4]	Night Rainy [5:7]	Night Rainy [1:4]	Daytime Sunny [5:7]			
LDB [22]	VitDet	23.73	45.3$_{\pm 0.6}$	1.40$_{\pm 0.3}$	8.10$_{\pm 0.4}$	0.02$_{\pm 0.5}$	16.10$_{\pm 0.7}$	3.09$_{\pm 0.2}$	21.02$_{\pm 0.3}$	12.05$_{\pm 0.6}$
DuET (Ours)	VitDet	24.13	45.3$_{\pm 0.2}$	2.02$_{\pm 0.3}$	16.10$_{\pm 0.1}$	1.58$_{\pm 0.3}$	15.20$_{\pm 0.2}$	4.46$_{\pm 0.2}$	22.47$_{\pm 0.1}$	13.47$_{\pm 0.2}$
CL-DETR [16]	Deformable DETR	91.96	46.3$_{\pm 0.4}$	26.26$_{\pm 0.4}$	9.75$_{\pm 0.5}$	7.41$_{\pm 0.6}$	14.88$_{\pm 0.3}$	56.72$_{\pm 0.4}$	*53.93*$_{\pm 0.2}$	55.33$_{\pm 0.5}$
DuET (Ours)	Deformable DETR	92.63	46.3$_{\pm 0.2}$	26.75$_{\pm 0.1}$	3.33$_{\pm 0.2}$	7.98$_{\pm 0.2}$	13.9$_{\pm 0.1}$	*57.78*$_{\pm 0.2}$	**54.58**$_{\pm 0.1}$	*56.18*$_{\pm 0.2}$
CL-DETR [16]	RT-DETR-l	33.87	57.2$_{\pm 0.4}$	5.11$_{\pm 0.5}$	28.1$_{\pm 0.6}$	5.04$_{\pm 0.3}$	14.3$_{\pm 0.4}$	8.93$_{\pm 0.2}$	20.82$_{\pm 0.5}$	14.88$_{\pm 0.4}$
DuET (Ours)	RT-DETR-l	33.89	57.2$_{\pm 0.4}$	11.8$_{\pm 0.1}$	14.90$_{\pm 0.2}$	12.50$_{\pm 0.2}$	17.30$_{\pm 0.1}$	20.63$_{\pm 0.2}$	36.50$_{\pm 0.1}$	28.57$_{\pm 0.2}$
LwF [13]	YOLO11n	11.01	49.4$_{\pm 0.2}$	21.50$_{\pm 0.4}$	0.17$_{\pm 0.6}$	9.36$_{\pm 0.3}$	0.55$_{\pm 0.5}$	43.52$_{\pm 0.3}$	17.78$_{\pm 0.7}$	30.65$_{\pm 0.6}$
ERD [5]	YOLO11n	11.50	49.4$_{\pm 0.2}$	25.60$_{\pm 0.4}$	16.70$_{\pm 0.3}$	12.20$_{\pm 0.6}$	17.60$_{\pm 0.7}$	51.82$_{\pm 0.5}$	38.96$_{\pm 0.3}$	45.39$_{\pm 0.4}$
DuET (Ours)	YOLO11n	11.02	49.4$_{\pm 0.2}$	43.30$_{\pm 0.1}$	2.67$_{\pm 0.3}$	14.00$_{\pm 0.2}$	8.93$_{\pm 0.1}$	**87.65**$_{\pm 0.2}$	34.18$_{\pm 0.1}$	**60.92**$_{\pm 0.2}$

References

[1] Li Chen, Chunyan Yu, and Lvcai Chen. A new knowledge distillation for incremental object detection. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2019. 1

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 6

[3] Muhammad Sohail Danish, Muhammad Haris Khan, Muhammad Akhtar Munir, M Saquib Sarfraz, and Mohsen Ali. Improving single domain-generalized object detection: A focus on diversification and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17732–17742, 2024. 6

[4] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010. 6

[5] Tao Feng, Mang Wang, and Hangjie Yuan. Overcoming catastrophic forgetting in incremental object detection via elastic response distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9427–9436, 2022. 7, 8, 9, 10, 11

[6] Mahmoud Hassaballah, Mourad A Kenk, Khan Muhammad, and Shervin Minaee. Vehicle detection and tracking in adverse weather using a deep learning framework. *IEEE transactions on intelligent transportation systems*, 22(7):4230–4242, 2020. 6

[7] Chenyu Huang, Peng Ye, Tao Chen, Tong He, Xiangyu Yue, and Wanli Ouyang. Emr-merging: Tuning-free high-performance model merging. *arXiv preprint arXiv:2405.17461*, 2024. 3, 5

[8] Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. *arXiv preprint arXiv:2212.04089*, 2022. 3, 5

[9] Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, and Kiyoharu Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5001–5009, 2018. 6

[10] Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024. 2, 3, 5, 9

[11] Madhu Kiran, Marco Pedersoli, Jose Dolz, Louis-Antoine Blais-Morin, Eric Granger, et al. Incremental multi-target domain adaptation for object detection with efficient domain transfer. *Pattern Recognition*, 129:108771, 2022. 6

[12] Yanghao Li, Hanzi Mao, Ross Girshick, and Kaiming He. Exploring plain vision transformer backbones for object detection. In *European conference on computer vision*, pages 280–296. Springer, 2022. 5, 9

[13] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 2, 7, 8, 9, 10, 11

[14] Xialei Liu, Hao Yang, Avinash Ravichandran, Rahul Bhotika, and Stefano Soatto. Multi-task incremental learning for object detection. *arXiv preprint arXiv:2002.05347*, 2020. 6

[15] Yuyang Liu, Yang Cong, Dipam Goswami, Xialei Liu, and Joost van de Weijer. Augmented box replay: Overcoming foreground shift for incremental object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11367–11377, 2023. 6

[16] Yaoyao Liu, Bernt Schiele, Andrea Vedaldi, and Christian Rupprecht. Continual detection transformer for incremental object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23799–23808, 2023. 5, 7, 8, 9, 10, 11

[17] Ilya Loshchilov, Frank Hutter, et al. Fixing weight decay regularization in adam. *arXiv preprint arXiv:1711.05101*, 5: 5, 2017. 5

[18] Wenyu Lv, Shangliang Xu, Yian Zhao, Guanzhong Wang, Jinman Wei, Cheng Cui, Yuning Du, Qingqing Dang, and Yi Liu. Detrs beat yolos on real-time object detection, 2023. 2, 3, 5

[19] Daniel Marczak, Bartłomiej Twardowski, Tomasz Trzciński,

and Sebastian Cygert. Magmax: Leveraging model merging for seamless continual learning. In *European Conference on Computer Vision*, pages 379–395. Springer, 2025. 3, 5

[20] Michael S Matena and Colin A Raffel. Merging models with fisher-weighted averaging. *Advances in Neural Information Processing Systems*, 35:17703–17716, 2022. 3, 5

[21] Angelo G Menezes, Gustavo de Moura, Cézanne Alves, and André CPLF de Carvalho. Continual object detection: a review of definitions, strategies, and challenges. *Neural networks*, 161:476–493, 2023. 1, 6

[22] 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. 5, 6, 7, 8, 9, 10, 11

[23] Aming Wu and Cheng Deng. Single-domain generalized object detection in urban scene via cyclic-disentangled self-distillation. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 847–856, 2022. 6

[24] Dongbao Yang, Yu Zhou, Aoting Zhang, Xurui Sun, Dayan Wu, Weiping Wang, and Qixiang Ye. Multi-view correlation distillation for incremental object detection. *Pattern Recognition*, 131:108863, 2022. 1

[25] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020. 6

[26] 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. 5, 9