EW-DETR: Evolving World Object Detection via Incremental Low-Rank DEtection TRansformer

Supplementary Material

A. EUMix Mathmematical Formulation

For completeness, we detail the formulation of the Entropy-Aware Unknown Mixing (EUMix) module in this section. For a given query i at task t, let

$$\mathbf{z}_i^{\text{cls}} = \left[\mathbf{z}_i^{\text{known}},\, z_i^{\text{unk}}\right] \tag{1}$$

denote the raw classification logits over the $|\mathcal{K}^t|$ known classes and the single unknown class, where $\mathbf{z}_i^{\text{known}} \in \mathbb{R}^{|\mathcal{K}^t|}$ and $z_i^{\text{unk}} \in \mathbb{R}$. We write $z_{i,c}^{\text{known}}$ for the logit of known class $c \in \mathcal{K}^t$. In addition, let z_i^{obj} be the objectness logit produced by the Query-Norm Objectness Adapter for the same query.

We first compute the maximum known-class confidence

$$p_i^{\text{known,max}} = \max_{c \in \mathcal{K}^t} \sigma\big(z_{i,c}^{\text{known}}\big), \tag{2}$$

where $\sigma(\cdot)$ denotes the logistic sigmoid. If the model is highly confident in some known class, $p_i^{\text{known,max}}$ is close to 1, whereas ambiguous or out-of-distribution objects yield lower values. We convert this observation into a calibrated gap g_i that measures how much probability mass is available for the unknown class:

$$g_i = \big(1 - p_i^{\text{known,max}}\big)^{\gamma}, \qquad \gamma = \text{softplus}(\theta_\gamma), \tag{3}$$

where θ_γ is a learned scalar and $\gamma > 0$ acts as a temperature on the gap. When $\gamma > 1$ the gap is sharpened, so that only strongly uncertain predictions yield a significant g_i; when $\gamma < 1$ the transition is smoother, which is beneficial if known-class logits are noisy.

We interpret the product of objectness and g_i as an objectness-derived unknown probability:

$$p_{\text{obj},i}^{\text{unk}} = \sigma\big(z_i^{\text{obj}}\big)\, g_i, \tag{4}$$

which is high exactly when the model believes there is an object at the query location but no known class explains it confidently. In parallel, we convert the learned unknown logit into a probability, allowing a learnable bias b_{obj} to compensate for the fact that the unknown logit rarely sees positive supervision:

$$p_{\text{cls},i}^{\text{unk}} = \sigma\big(z_i^{\text{unk}} + b_{\text{obj}}\big). \tag{5}$$

EUMix combines these two estimates through a learnable mixing coefficient

$$\alpha = \sigma(\theta_\alpha), \tag{6}$$

where θ_α is a scalar parameter. The final unknown probability is

$$p_{\text{final},i}^{\text{unk}} = \alpha\, p_{\text{cls},i}^{\text{unk}} + (1 - \alpha)\, p_{\text{obj},i}^{\text{unk}}, \tag{7}$$

which is then converted back to a logit:

$$z_{\text{final},i}^{\text{unk}} = \text{logit}\big(p_{\text{final},i}^{\text{unk}}\big), \tag{8}$$

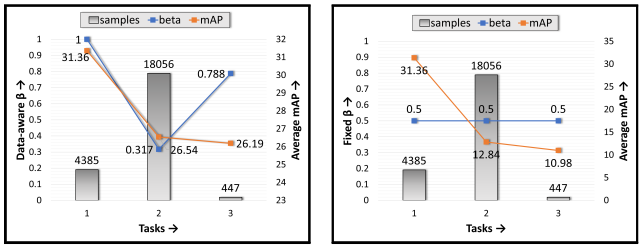
where logit denotes the inverse of the logistic sigmoid. The mixing weight α is initialised to favour the classifier and is learned end-to-end; if the classifier becomes reliable on unknowns, α naturally increases, but in early tasks or under strong domain shift the model can lean more heavily on the objectness–gap signal.

The final logits fed into the detection loss are then

$$\mathbf{z}_i^{\text{final}} = \left[\mathbf{z}_{\text{final},i}^{\text{known}},\, z_{\text{final},i}^{\text{unk}}\right]. \tag{9}$$

All parameters in this module, $\theta_\gamma, \theta_\alpha, \theta_\lambda, b_{\text{obj}}$, are trained jointly with the rest of the network using exactly the same detection loss as the base detector, without any explicit supervision on the unknown category. Their role is purely to reshape the logit space so that unknown evidence coming from objectness and classification uncertainty is translated into calibrated unknown scores.

B. Addressing data imbalance across tasks



(a) (b)

Figure S1. Effect of (a) data-aware vs. (b) fixed merging.

A distinctive challenge in EWOD is the severe data imbalance across tasks, where different domains and class distributions yield vastly different sample sizes. As illustrated in Table S2, the Diverse Weather benchmark exhibits severe data imbalance across tasks, with Task $\mathcal{T}_2$ (Night Sunny) containing approximately 40 times more training samples than Task $\mathcal{T}_3$ (Night Rainy). This imbalance poses a critical challenge for knowledge consolidation in exemplar-free incremental learning: naively merging task-specific updates

into the aggregate adapter can either cause catastrophic forgetting when large tasks dominate, or prevent adaptation when small tasks are under-weighted. In Figure S1a, our data-aware merging coefficient β_t (Eq. 3, main) dynamically adjusts based on the ratio between current and aggregate samples. At Task $\mathcal{T}_1$, $\beta_1 = 1.0$ initialises the aggregate adapter entirely from the first task. At Task $\mathcal{T}_2$, the large training sample count relative to $\mathcal{T}_1$ results in a low $\beta_2 = 0.317$, limiting the influence of this data-rich task to prevent it from overwhelming prior knowledge. Consequently, the model retains strong performance (average mAP: 26.54), while if we compare it to fixed β (Figure S1b) case, the uniform weighting causes the data-rich Task $\mathcal{T}_2$ to dominate the aggregate adapter, resulting in severe performance degradation, resulting in severe performance degradation: the average mAP drops from 31.36 to 12.84 at $\mathcal{T}_2$. Hence, the fixed strategy fails to preserve knowledge from data-scarce tasks ($\mathcal{T}_1$, $\mathcal{T}_3$) as their contributions are insufficiently weighted during merging. The effectiveness of data-aware merging is further validated in Figure S5b (`Appendix G.3`), where we ablate different data-aware merging bounds $(\beta_{\min}, \beta_{\max})$ for the merging coefficient β_t. The fixed-β baseline ($\beta_{\min} = \beta_{\max} = 0.5$) achieves the lowest FOGS score (54.04) with catastrophic GSS collapse (0.02), confirming that ignoring data imbalance severely impairs both retention and generalisation. Hence, data-aware merging is crucial for balancing stability and plasticity in EWOD under heterogeneous data distributions.

C. Query norm across tasks

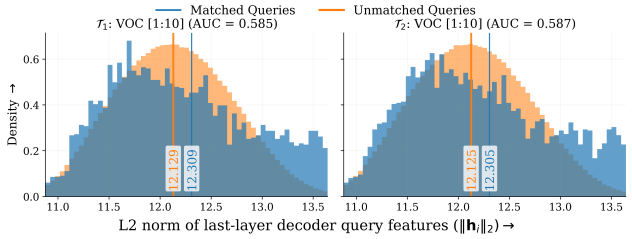


Figure S2. Histogram plots of last-layer decoder query norms for VOC [1:10] → Clipart [11:18].

To quantitatively validate the norm-objectness hypothesis used by QNorm-Obj (Section 3.4, main), we plot the ℓ_2 norm of last-layer decoder query features $\|\mathbf{h}_i\|_2$ and split queries at inference into **matched** (assigned to a ground-truth box by Hungarian matching) and **unmatched** (background). Figure S2 shows that matched queries consistently attain higher norms than unmatched ones, both after training on $\mathcal{T}_1$ (VOC [1:10]: 12.309 > 12.129) and when evaluating the same prior classes after training on $\mathcal{T}_2$ (VOC [1:10]: 12.305 > 12.125). Moreover, $\|\mathbf{h}_i\|_2$ maintains stable separation across tasks (ROC-AUC = 0.585 in $\mathcal{T}_1$ and 0.587 in $\mathcal{T}_2$), supporting that query norms provide a persistent, class-agnostic objectness signal across tasks, as utilized by QNorm-Obj (Section 3.4, main).

D. EWOD Protocol and Metrics

Table S1 illustrates the proposed EWOD training and evaluation protocol on the Diverse Weather benchmark. EWOD evaluation protocol builds upon the dual-incremental schedule of DuIOD [11], while integrating open-world constraints from OWOD methods [4, 8, 19]. As shown in Table S1, the protocol spans multiple tasks $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_T$, where each task introduces new classes from a new domain.

Training Protocol. During training on task $\mathcal{T}_t$, only the current task's known classes $\mathcal{K}_t$ receive supervision through bounding box annotations. Crucially, instances of previously learned classes $\{\mathcal{K}_1 \cup \ldots \cup \mathcal{K}_{t-1}\}$ and truly novel objects remain unlabeled in the training set, mimicking real-world scenarios where exhaustive annotation is impractical. Additionally, certain classes are intentionally withheld from training across all tasks (e.g., ~~truck~~ in Table S1) to serve as consistent unknown objects throughout the evaluation.

Evaluation Protocol. At task $\mathcal{T}_t$, the detector is evaluated on its ability to: (i) detect all previously learned classes $\{\mathcal{K}_1 \cup \ldots \cup \mathcal{K}_t\}$ across all encountered domains $\{\mathcal{D}_1, \ldots, \mathcal{D}_t\}$, (ii) identify unseen objects as *"unknown"*, and (iii) generalize to domain shifts by detecting current task classes across mixed domains.

To illustrate this protocol, consider the 3-task scenario in Table S1: *Daytime Sunny [1:2]* → *Night Sunny [3:4]* → *Night Rainy [5:6]*. In $\mathcal{T}_1$, the model learns bike, bus from Daytime Sunny and is evaluated on these classes plus ***unknown*** detection in Daytime Sunny. In $\mathcal{T}_2$, the model learns car, motor from Night Sunny (without annotations for bike, bus) and is evaluated on bike, bus + car, motor + ***unknown*** across both Daytime Sunny and Night Sunny domains. Finally, in $\mathcal{T}_3$, the model learns person, rider from Night Rainy (excluding ~~truck~~) and is evaluated on all six learned classes plus unknowns across all three seen domains: Daytime Sunny, Night Sunny, and Night Rainy.

Hence, EWOD protocol ensures that the detector is challenged with: (1) retaining past knowledge without explicit supervision, (2) rejecting unknowns while maintaining precision on known classes, and (3) adapting to domain shifts for both old and new classes.

FOGS: Forgetting-Openness-Generalisation Score As discussed in `Section 4.2` (main), existing metrics either focus on individual EWOD dimensions or fail to capture their interplay. To capture the coupled failure modes in EWOD and facilitate holistic comparison, we introduce FOGS as the mean of three calibrated sub-scores:

$$\text{FOGS} = \frac{\text{FSS} + \text{OSS} + \text{GSS}}{3} \tag{10}$$

Table S1. EWOD Training and Evaluation Protocol in Diverse Weather benchmark. Blue, orange, and olive denote classes introduced in $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$, respectively; ~~red~~ indicates classes excluded during training in all tasks but treated as ***unknown*** during testing.

Daytime Sunny [1:3] → Night Sunny [4:6]				
Task	**Train Classes**	**Train Domain**	**Test Classes**	**Test Domains**
$\mathcal{T}_1$	Daytime Sunny [1:3] bike, bus, car	Daytime Sunny	Daytime Sunny [1:3] bike, bus, car + ***unknown***	Daytime Sunny
$\mathcal{T}_2$	Night Sunny [4:6] motor, person, rider, ~~truck~~	Night Sunny	Daytime Sunny [1:6] + Night Sunny [1:6] bike, bus, car + motor, person, rider + ***unknown***	Daytime Sunny + Night Sunny
Daytime Sunny [1:2] → Night Sunny [3:4] → Night Rainy [5:6]				
Task	**Train Classes**	**Train Domain**	**Test Classes**	**Test Domains**
$\mathcal{T}_1$	Daytime Sunny [1:2] bike, bus	Daytime Sunny	Daytime Sunny [1:2] bike, bus + ***unknown***	Daytime Sunny
$\mathcal{T}_2$	Night Sunny [3:4] car, motor	Night Sunny	Daytime Sunny [1:4] + Night Sunny [1:4] bike, bus + car, motor + ***unknown***	Daytime Sunny + Night Sunny
$\mathcal{T}_3$	Night Rainy [5:6] person, rider, ~~truck~~	Night Rainy	Daytime Sunny [1:6] + Night Sunny [1:6] + Night Rainy [1:6] bike, bus + car, motor + person, rider + ***unknown***	Daytime Sunny + Night Sunny + Night Rainy

where each sub-score quantifies a distinct dimension of EWOD performance. Higher FOGS indicates better overall performance across forgetting resistance, open-set robustness, and domain generalisation.

FSS: Forgetting Sub-Score As discussed in `Section 4.2 (main)`, FSS quantifies the detector's ability to retain performance on previously learned classes across all cumulative tasks. As defined in Eq. 12 (main), FSS measures retention by comparing the performance on previously learned classes at each task with their average initial performance when first introduced.

To illustrate the computation, consider the 3-task scenario from Table S1: *Daytime Sunny [1:2] → Night Sunny [3:4] → Night Rainy [5:6]*. Here, we have $T = 3$ tasks, and FSS is computed by averaging retention ratios across tasks $\mathcal{T}_2$ and $\mathcal{T}_3$:

$$\text{FSS} = \frac{1}{2}\left(\frac{\text{mAP}^{\mathcal{T}_2}_{\text{prev}}(\text{bike, bus})}{\text{mAP}^{\mathcal{T}_1}_{\text{curr}}(\text{bike, bus})} + \frac{\text{mAP}^{\mathcal{T}_3}_{\text{prev}}(\text{bike, bus, car, motor})}{\frac{1}{2}\left(\text{mAP}^{\mathcal{T}_1}_{\text{curr}}(\text{bike, bus}) + \text{mAP}^{\mathcal{T}_2}_{\text{curr}}(\text{car, motor})\right)}\right) \tag{11}$$

The first term measures how well the detector retains bike, bus (learned in $\mathcal{T}_1$ from Daytime Sunny) when evaluated at $\mathcal{T}_2$ while the second term measures retention of all previously learned classes bike, bus, car, motor when evaluated at $\mathcal{T}_3$. A higher FSS indicates better retention of past knowledge, while lower values indicate catastrophic forgetting.

OSS: Openness Sub-Score OSS captures the detector's open-set behaviour by combining three complementary metrics from OWOD literature [4, 8, 19], as defined in Eq. 13 (main). For the 3-task scenario in Table S1, we compute OSS by averaging the openness scores across all three tasks:

$$\text{OSS} = \frac{1}{3}\sum_{t=1}^{3}\left(\frac{\text{U-Recall}_t + (1 - \text{WI}_t) + \frac{1}{1+\text{A-OSE}_t\,/\,\text{GT}_{\text{unk, t}}}}{3}\right) \tag{12}$$

Higher OSS values indicate robust unknown detection with minimal impact on known class precision, while lower values suggest open-set collapse.

GSS: Generalisation Sub-Score GSS evaluates the detector's ability to adapt to domain shifts, as defined in Eq. 14 (main). GSS measures how well newly learned classes transfer across the mixed domains encountered so far.

For the 3-task scenario from Table S1, GSS is computed by averaging the cross-domain performance for tasks $\mathcal{T}_2$ and $\mathcal{T}_3$:

$$\text{GSS} = \frac{1}{2}\left(\text{mAP}^{\mathcal{T}_2}_{\text{curr}}(\text{car, motor}) + \text{mAP}^{\mathcal{T}_3}_{\text{curr}}(\text{person, rider})\right) \tag{13}$$

Here, $\text{mAP}^{\mathcal{T}_2}_{\text{curr}}(\text{car, motor})$ essentially measures whether the detector can generalise car, motor to Daytime Sunny, despite being trained only on Night Sunny. Similarly, $\text{mAP}^{\mathcal{T}_3}_{\text{curr}}(\text{person, rider})$ measures the performance on classes person, rider (learned from Night Rainy) when evaluated across all three Daytime Sunny, Night Sunny and Night Rainy domains, testing cross-domain adaptability for these newly learned classes. Higher GSS reflects stronger domain generalisation, while lower values indicate domain-specific overfitting.

Table S2. Class-wise distribution on Diverse Weather benchmark: Daytime Sunny [1:2] → Night Sunny [3:4] → Night Rainy [5:6]; ~~red~~ indicates classes (i.e., ~~truck~~) excluded during training in all tasks but treated as ***unknown*** during testing. Best viewed in colour.

Class ID	Class Name	Training			Testing		
		$\mathcal{T}_1$: Daytime Sunny	$\mathcal{T}_2$: Night Sunny	$\mathcal{T}_3$: Night Rainy	$\mathcal{T}_1$: Daytime Sunny	$\mathcal{T}_2$: Daytime Sunny + Night Sunny	$\mathcal{T}_3$: Daytime Sunny + Night Sunny + Night Rainy
1	bike	✓			✓	✓✓	✓✓✓
2	bus	✓			✓	✓✓	✓✓✓
3	car		✓			✓✓	✓✓✓
4	motor		✓			✓✓	✓✓✓
5	person			✓			✓✓✓
6	rider			✓			✓✓✓
–	~~truck~~						
7	***unknown***				✓	✓	✓
# Classes		2	2	2	2+**1**	2+2+**1**	2+2+2+**1**
# images		4709	18459	471	10219	26754	29527
# annotations		6772	169460	1341	116785	197711	364231

Table S3. Class-wise distribution on three Pascal Series benchmarks: $\mathcal{T}_1$: VOC [1:10] → $\mathcal{T}_2$: {Clipart [11:18] / Watercolor [11:14] / Comic [11:14]}. Note that while Watercolor and Comic classes are listed here as IDs 15–18 for visual continuity, they are strict subsets of the whole label space and are mapped to IDs 11–14 during experiments; ~~red~~ indicates classes (i.e., ~~dog~~, ~~person~~) excluded during training in all tasks but treated as ***unknown*** during testing. Best viewed in colour.

Class ID	Class Name	Training				Testing			
		$\mathcal{T}_1$: VOC	$\mathcal{T}_2$: Clipart	$\mathcal{T}_2$: Watercolor	$\mathcal{T}_2$: Comic	$\mathcal{T}_1$: VOC	$\mathcal{T}_2$: VOC + Clipart	$\mathcal{T}_2$: VOC + Watercolor	$\mathcal{T}_2$: VOC + Comic
1	aeroplane	✓				✓	✓✓	✓	✓
2	boat	✓				✓	✓✓	✓	✓
3	bottle	✓				✓	✓✓	✓	✓
4	bus	✓				✓	✓✓	✓	✓
5	chair	✓				✓	✓✓	✓	✓
6	cow	✓				✓	✓✓	✓	✓
7	diningtable	✓				✓	✓✓	✓	✓
8	horse	✓				✓	✓✓	✓	✓
9	motorbike	✓				✓	✓✓	✓	✓
10	pottedplant	✓				✓	✓✓	✓	✓
11	sheep		✓				✓✓		
12	sofa		✓				✓✓		
13	train		✓				✓✓		
14	tvmonitor		✓				✓✓		
15	bicycle		✓	✓	✓		✓✓	✓✓	✓✓
16	bird		✓	✓	✓		✓✓	✓✓	✓✓
17	car		✓	✓	✓		✓✓	✓✓	✓✓
18	cat		✓	✓	✓		✓✓	✓✓	✓✓
–	~~dog~~								
–	~~person~~								
19	***unknown***					✓	✓	✓	✓
# Classes		10	8	4	4	10 + **1**	10 + 8 + **1**	10 + 4 + **1**	10 + 4 + **1**
# images		8909	165	1072	1150	6041	7774	8657	16630
# annotations		16181	270	1661	3214	14976	16502	8719	18151

E. Dataset Statistics

As mentioned in `Section 4.1 (main)`, we evaluate our approach on two dataset series that support evolving open-world object detection across domain-incremental tasks. Here, we provide detailed statistics for both benchmarks.

Diverse Weather benchmark Table S2 presents the class-wise distribution across three weather conditions: Daytime Sunny, Daytime Foggy, and Night Rainy, sourced from BDD-100k [17], FoggyCityscapes [1], and Adverse-Weather [5]. The benchmark follows a 2+2+2 class split across three incremental tasks $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$. Each task introduces two new object categories while encountering data from a new weather domain. The testing phase at each task evaluates on the union of all domains seen up to that point, including instances of the ~~truck~~ class treated as ***unknown*** in all tasks to simulate open-world scenarios. The dataset exhibits significant domain shift across weather conditions,

with Daytime Foggy providing substantially more annotations compared to the other two domains, reflecting realistic data availability patterns across different environmental conditions.

Pascal Series benchmark Table S3 illustrates the class distribution across four visually distinct domains: Pascal VOC [3], Clipart, Watercolor, and Comic [7] for three different Pascal Series benchmarks: VOC [1:10] → Clipart [11:18], VOC [1:10] → Watercolor [11:14], and VOC [1:10] → Comic [11:14]. Each benchmark follows a two-task incremental learning setup where $\mathcal{T}_1$ trains on VOC with 10 base classes, followed by $\mathcal{T}_2$ introducing a different domain variant. The Clipart configuration introduces 8 new classes following class split of 10 + 8, while Watercolor and Comic configurations each introduce 4 classes following class split of 10 + 4. Notably, Watercolor and Comic contain only 6 classes in total, which are strict subsets of the 20-class label space shared by VOC and Clipart. Consequently, these domains lack annotations for the 14 classes present in VOC and Clipart. Similar to the Diverse Weather Series, the ~~dog~~ and ~~person~~ classes are excluded during training across all tasks and treated as *unknown* during testing. The testing phase for each $\mathcal{T}_2$ configuration evaluates on the union of VOC and the respective domain, with a significant data imbalance as reflected in the counts of images and annotations.

F. Implementation Details

We instantiate EW-DETR on top of DETR-based detectors (RF-DETR-N [14] and Deformable DETR [20]), retaining the original detector pipelines and losses unchanged. As discussed in `Section 3 (main)`, for each task, the backbone and transformer encoder–decoder layers are kept frozen, along with Aggregate LoRA adapters. For Incremental LoRA Adapters, we set rank r = 16, while for data-aware merging coefficient calculation, we keep $(\beta_{min}, \beta_{max})$ = (0.2, 0.8), as detailed in `Appendix G.3`. All methods are based on DINO-pretrained ResNet-50 backbones [6, 18], while the recent RF-DETR-N supports DINOv2-S [12, 14]. To ensure a fair comparison across baselines, we keep each method's trainable components consistent with their original implementations; only the dataloaders and training loops are adapted to meet the exemplar-free EWOD requirements. For [11], we additionally incorporate Energy Based Unknown Identification (EBUI) from [8] to handle unknowns; hence, we see non-zero OSS for [11] in Figure S4. For all methods in each task, the detector is trained for 100 epochs using the AdamW optimiser with an initial learning rate of 1e-4, weight decay of 1e-4, and batch size of 16 on a single NVIDIA H100 80GB card. All methods follow the evaluation protocol detailed in `Appendix D` for a fair comparison.

G. Additional Results & Ablations

G.1. Complexity Analysis

Table S4. Computational complexity analysis of various methods evaluated on Pascal Series (VOC [1:10] → Clipart [11:18]) benchmark, all evaluated on a single NVIDIA H100 80GB card.

Method	Underlying Detector	Trainable Params (M)	GFLOPs	Avg. Inference Speed (ms)	Avg. Memory Footprint (GB)
ORE [8]	Faster RCNN [13]	32.96	1665.03	60.41	3.75
OW-DETR [4]	D-DETR [20]	24.22	157.71	59.82	1.33
PROB [21]	D-DETR [20]	23.99	135.31	97.71	1.32
CAT [9]	D-DETR [20]	24.25	135.34	112.54	1.35
ORTH [15]	RandBox [16]	105.9	2073.57	1151.19	1.26
DuET [11]	D-DETR [20]	24.22	135.3	150.61	1.41
OWOBJ [19]	D-DETR [20]	23.99	135.3	80.85	1.32
EW-DETR	D-DETR [20]	0.46	171.23	131.92	1.55
EW-DETR	RF-DETR [14]	1.8	32.22	57.38	0.32

Table S4 presents a comprehensive computational complexity analysis of EW-DETR against recent methods on the Pascal Series benchmark. Since Incremental LoRA adapters (`Section 3.3, main`) freeze the base model and transformer encoder-decoder weights, EW-DETR achieves a drastic reduction in trainable parameters: only **0.46M** when built upon Deformable DETR [20] and **1.8M** when built upon RF-DETR [14], compared to 23.99M–105.9M for other methods. This represents a **98.1%** parameter reduction relative to standard Deformable DETR-based approaches, making EW-DETR highly suitable for resource-constrained deployment scenarios. Moreover, under EWOD, Deformable DETR-based methods with the standard six encoder–decoder layers consistently collapse to near-zero mAP, while a single encoder–decoder layer yields stable results for all; thus, we adopt the single-layer setting throughout. This reduced the number of trainable parameters; hence, we observe ∼24M trainable parameters for D-DETR-based methods in Table S4.

Regarding comparison of FLOPs, the D-DETR variant requires 171.23 GFLOPs, which is moderately higher than other D-DETR-based methods (135.3–157.71 GFLOPs) due to the QNorm-Obj and EUMix modules, yet substantially lower than ORE [8] (1665.03 GFLOPs) and ORTH [15] (2073.57 GFLOPs). However, the RF-DETR variant achieves remarkable efficiency with only 32.22 GFLOPs. Inference speed remains practical, with EW-DETR (D-DETR) averaging 131.92 ms per image, comparable to other D-DETR methods, while EW-DETR (RF-DETR) achieves 57.38 ms, making it among the fastest approaches evaluated. Following [14], we use PyTorch's `FlopCounterMode` to calculate FLOPs for all methods.

Memory footprint analysis reveals that EW-DETR (D-DETR) requires 1.55 GB, slightly higher than other D-DETR methods (1.32–1.41 GB) due to the dual LoRA adapter architecture maintaining both aggregate and task-specific adapters during training. However, EW-DETR (RF-DETR) depicts a memory efficiency at only 0.32 GB, significantly lower than all compared methods. Hence,

overall complexity analysis demonstrates that the EW-DETR framework is suitable for real and dynamic evolving world scenarios where continuous adaptation, storage constraints, and deployment efficiency are required.

G.2. Ablation analysis for random task sequences

Table S5. Sensitivity analysis for random task permutations on Diverse Weather benchmark.

$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	FSS	OSS	GSS	FOGS
Daytime Sunny [5:6]	Night Sunny [1:2]	Night Rainy [3:4]	73.30	73.45	24.62	57.12
Daytime Sunny [3:4]	Night Sunny [5:6]	Night Rainy [1:2]	75.03	71.31	17.94	54.76
Daytime Sunny [1:2]	Night Sunny [3:4]	Night Rainy [5:6]	73.63	73.43	18.68	55.25
Night Rainy [1:2]	Daytime Sunny [3:4]	Night Sunny [5:6]	71.79	72.50	28.67	57.65
Night Sunny [1:2]	Night Rainy [3:4]	Daytime Sunny [5:6]	71.93	70.69	21.25	54.62
Daytime Sunny [7,2]	Night Sunny [3,4]	Night Rainy [5,6]	73.65	71.71	14.96	53.44
Daytime Sunny [1,7]	Night Sunny [3,4]	Night Rainy [5,6]	74.94	69.84	14.74	53.17
Daytime Sunny [1,2]	Night Sunny [7,4]	Night Rainy [5,6]	73.36	70.2	17.72	53.76
Standard Deviation across tasks:			**1.19**	**1.39**	**4.81**	**1.65**

In real-world, evolving environments, the order in which new classes and domains appear can vary unpredictably. To evaluate the robustness of EW-DETR under such conditions, we conducted experiments with different task permutations on the Diverse Weather benchmark. Following [11], we randomly shuffled both the class assignments and the domain order across five distinct configurations. As shown in Table S5, EW-DETR delivers consistent performance across all permutations. FSS remains highly stable, with a standard deviation of only **1.19**, indicating that the proposed Incremental LoRA Adapters effectively preserve previously learned knowledge regardless of task order. Similarly, OSS exhibits very low standard deviation (**1.11**), demonstrating that the QNorm-Obj and EUMix modules maintain strong unknown-detection capabilities across different sequences. However, GSS shows comparatively larger variation, which could be attributed to significant data imbalance across domains: because GSS measures how well current-task classes generalise to all previously seen domains, configurations in which data-rich domains appear in later tasks naturally achieve higher GSS. Conversely, when data-scarce domains appear in later tasks, GSS declines noticeably. Despite these variations in individual sub-scores, the overall FOGS metric remains stable with a standard deviation of **1.26**, confirming that EW-DETR maintains consistent holistic performance across the three critical EWOD dimensions irrespective of task ordering. Moreover, we test **unknown-class sensitivity** by changing which category is withheld as the stationary unknown prior (last three rows of Table S5). The performance remains stable under this change: OSS varies only modestly (69.84–71.71), and FOGS remains tightly bounded (53.17–53.76), indicating that QNorm-Obj and EUMix do not rely on a particular choice of unknown class to maintain open-set robustness.

G.3. Ablation analysis for key hyperparameters

LoRA rank (r). Figure S5a demonstrates the sensitivity of EW-DETR to LoRA rank r, which governs the capacity of the Incremental LoRA Adapters. FSS remains remarkably stable across all ranks (94.95–97.86). OSS similarly exhibits robustness (77.4–79.28), indicating that the QNorm-Obj and EUMix operate reliably regardless of adapter capacity. However, GSS varies more significantly, peaking at $r = 16$ (8.42) before declining at higher ranks, suggesting that excessive capacity may lead to domain-specific overfitting. The overall FOGS metric peaks at $r = 16$ (61.08) while maintaining a manageable parameter count (1.8 M), representing the optimal trade-off between model capacity and generalisation. Hence, we select $r = 16$ as the default (`Appendix F`) for EW-DETR.

Data-aware merging bounds ($\beta_{\min}, \beta_{\max}$). Figure S5b illustrates how the data-aware merging coefficient bounds impact the stability-plasticity trade-off during LoRA adapter fusion. The fixed-β baseline ($\beta_{\min} = \beta_{\max} = 0.5$), which ignores task-specific data imbalance, exhibits severe GSS collapse (0.02) and the lowest FOGS (54.04), demonstrating that uniform merging fails under heterogeneous data distributions. Configuration biased towards stability $(0.2, 0.4)$ achieve the highest FSS (98.25) by limiting the influence of new tasks, effectively preventing forgetting but at the cost of reduced plasticity (GSS: 6.66). Conversely, plasticity-biased configurations $(0.6, 0.8)$ sacrifice some retention (FSS: 86.63) but enable better adaptation to current domains (GSS: 12.55). Overall, $(\beta_{\min}, \beta_{\max}) = (0.2, 0.8)$ emerges as the optimal choice for EWOD, and hence is adopted as the default setting (`Appendix F`), achieving a balanced trade-off with FOGS score of 61.08.

G.4. Comprehensive Results

Figure S4 presents a holistic comparison of all evaluated methods across the three critical dimensions of EWOD. EW-DETR (RF-DETR) demonstrates superior performance and achieves an average FOGS of **52.33**, which represents a **57.24%** improvement over the next-best method ORTH [15] (FOGS: 33.28). Notably, EW-DETR (RF-DETR) achieves the highest Forgetting Sub-Score (FSS: **75.69**), indicating exceptional retention of previously learned classes without exemplar replay. While PROB [21] demonstrates strong Openness performance (OSS: 66.67), it suffers from severe catastrophic forgetting (FSS: 0.61), highlighting the fundamental trade-off that exemplar-free methods must navigate. EW-DETR successfully balances all three dimensions, achieving competitive OSS (**67.3**) while maintaining stability (FSS: **75.69**) and reasonable generalisation (GSS: 14.02). In contrast, all OWOD methods [4, 8, 9, 15, 19, 21] exhibit near-zero forgetting scores, demonstrating their inability to operate under exemplar-free constraints, while DuET [11], despite being an exemplar-free method, achieves moderate FSS (25.9) in comparison to EW-DETR (75.69).

Table S6. Analysis of unknown object confusion on Pascal Series benchmark: VOC [1:10] → Clipart [11:18]. The table compares all methods using unknown class confusion metrics, including U-Recall, WI and A-OSE. Best results per column in **bold**, second-best *underlined*.

Method	$\mathcal{T}_1$: VOC [1:10]			$\mathcal{T}_2$: Clipart [11:18]			Metrics			
	U-Recall (↑)	WI (↓)	A-OSE (↓)	U-Recall (↑)	WI (↓)	A-OSE (↓)	FSS (↑)	OSS (↑)	GSS (↑)	FOGS (↑)
ORE-EBUI [8]	9.84	0.0985	12916	6.97	0.0592	**852**	0	55.48	*11.37*	22.28
OW-DETR [4]	16.25	0.0851	55228	8.07	0.0644	28689	11.42	40.47	7.96	19.95
PROB [21]	52.73	0.1603	**12700**	*46.27*	0.0216	*1529*	0	*67.58*	0.27	22.62
CAT [9]	20.23	*0.083*	57094	8.1	0.0566	28045	29.52	41.29	8.05	26.29
ORTH [15]	*63.92*	0.757	32519	41.61	*0.0213*	5160	5.83	51.06	**32.44**	29.78
DuET [11]	0	0.0714	80714	0	0.0622	35214	41.05	35.49	1.46	26
OWOBJ [19]	45.93	0.155	14947	35.72	0.0383	3898	0	60.73	0.51	20.41
EW-DETR D-DETR	44.98	0.1382	20236	41.45	0.1096	2086	*64.86*	61.67	7.92	*44.82*
EW-DETR RF-DETR	**77.35**	**0.012**	*12773*	**78.23**	**0.0038**	2251	**96.19**	**78.62**	8.42	**61.08**

Table S6 provides a detailed breakdown of the unknown class confusion metrics: Wilderness Impact (WI) [2] and Absolute Open-Set Error (A-OSE) [10], that contribute to the overall Openness Sub-Score (OSS). As shown in Table S6, EW-DETR (RF-DETR) achieves the lowest WI values across both tasks (**0.012** in $\mathcal{T}_1$ and **0.0038** in $\mathcal{T}_2$), indicating minimal precision degradation when unknown objects are present. Moreover, it attains high Unknown Recall values (**77.35** in $\mathcal{T}_1$ and **78.23** in $\mathcal{T}_2$), demonstrating its effectiveness in identifying unknown objects without auxiliary supervision, while attaining competitive A-OSE values in comparison to other methods.

G.5. Qualitative Visualizations

Figure S3 presents qualitative comparisons on the VOC [1:10] → Clipart [11:18] benchmark, revealing distinct failure modes of each approach. OWOBJ successfully detects unknown objects (black bounding boxes) but suffers from severe catastrophic forgetting, missing several previously learned instances, confirming its reliance on exemplar replay. DuET, operating under closed-world assumptions, absorbs unknown objects into the background. In contrast, EW-DETR correctly identifies unknown objects while simultaneously maintaining accurate detection of all previously learned instances across both domains, demonstrating an effective balance between open-set robustness, catastrophic forgetting mitigation, and cross-domain generalisation, which validates the quantitative findings in `Table 1 (main)`.

H. Takeaways and Future Directions

EW-DETR demonstrates that DETR-based detectors can effectively operate in evolving-world settings through parameter-efficient incremental adaptation, achieving strong retention (FSS: 75.69) and robust open-set detection (OSS: 67.3) without storing any previous data. However, domain generalisation (GSS: 14.02) remains modest, indicating room for improvement in cross-domain transfer under exemplar-free constraints. This work establishes EWOD as a challenging yet practical paradigm that bridges continual learning, domain adaptation, and open-world recognition. We hope our framework and comprehensive evaluation protocol inspire further research in this direction.

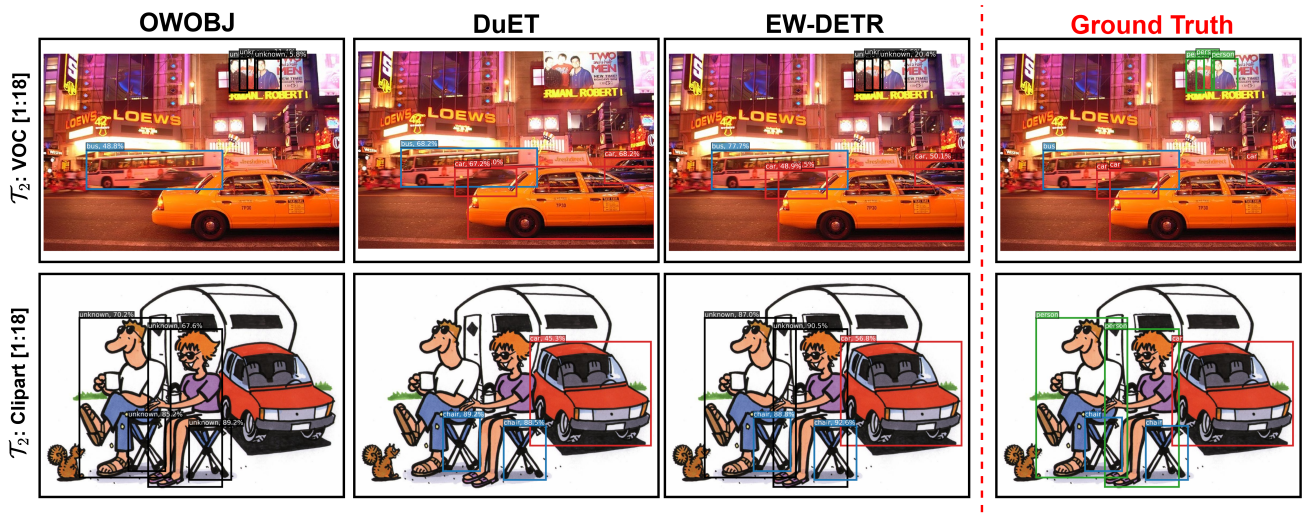


Figure S3. Qualitative comparison of EW-DETR with other methods on: VOC [1:10] → Clipart [11:18]. Best viewed in colour with zoom.

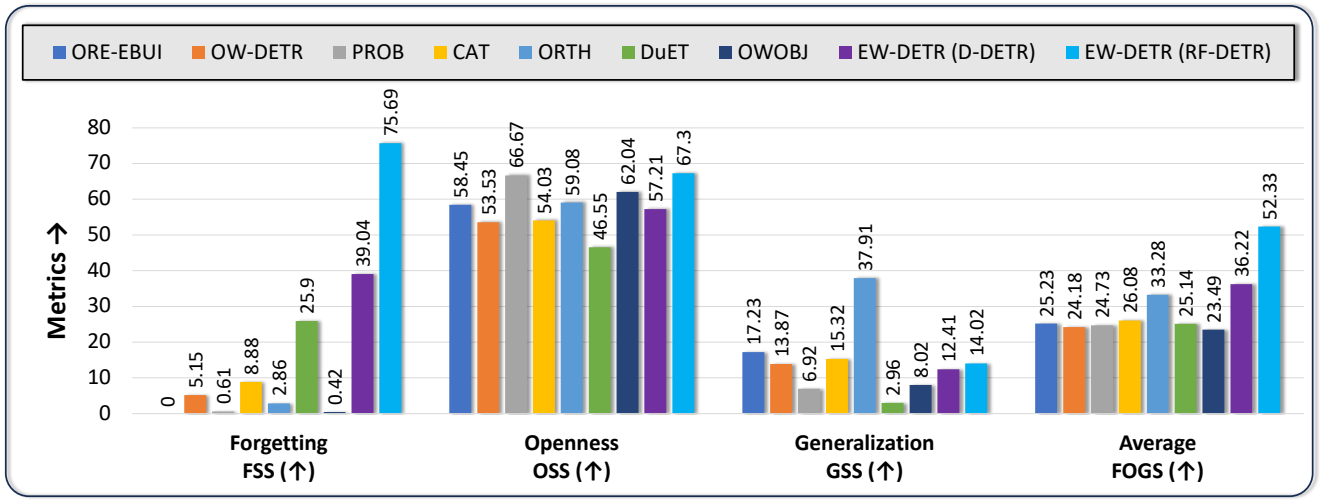
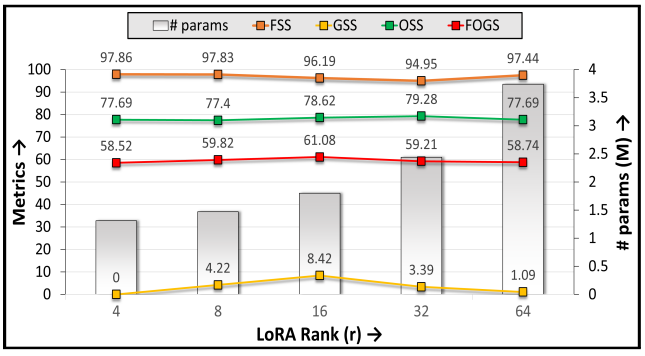
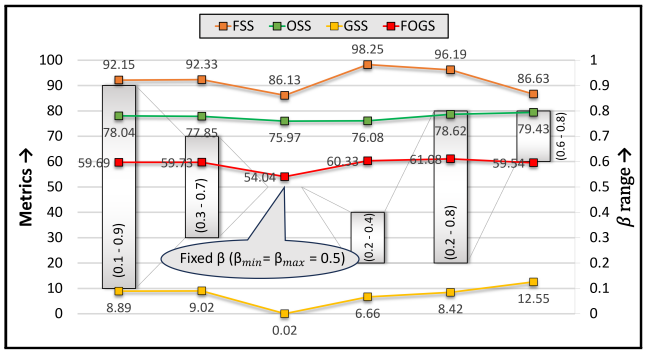


Figure S4. Comprehensive comparison of various methods across all three dimensions of EWOD: Forgetting, Openness, and Generalisation, quantified by FSS, OSS, and GSS, respectively, averaged across all EWOD experiments.



(a) Sensitivity to LoRA rank r



(b) Sensitivity to data-aware merging bounds $(\beta_{\min}, \beta_{\max})$

Figure S5. **Ablation analysis for key hyperparameters in the EW-DETR framework.** (a) Sensitivity to LoRA rank r. (b) Sensitivity to data-aware merging bounds $(\beta_{\min}, \beta_{\max})$.

References

[1] 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. 4

[2] Akshay Dhamija, Manuel Gunther, Jonathan Ventura, and Terrance Boult. The overlooked elephant of object detection: Open set. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1021–1030, 2020. 7

[3] 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. 5

[4] Akshita Gupta, Sanath Narayan, KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. Ow-detr: Open-world detection transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9235–9244, 2022. 2, 3, 5, 6, 7

[5] 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. 4

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[7] 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. 5

[8] KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5830–5840, 2021. 2, 3, 5, 6, 7

[9] Shuailei Ma, Yuefeng Wang, Ying Wei, Jiaqi Fan, Thomas H Li, Hongli Liu, and Fanbing Lv. Cat: Localization and identification cascade detection transformer for open-world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19681–19690, 2023. 5, 6, 7

[10] Dimity Miller, Lachlan Nicholson, Feras Dayoub, and Niko Sünderhauf. Dropout sampling for robust object detection in open-set conditions. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3243–3249. IEEE, 2018. 7

[11] Munish Monga, Vishal Chudasama, Pankaj Wasnik, and Biplab Banerjee. Duet: Dual incremental object detection via exemplar-free task arithmetic. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3121–3131, 2025. 2, 5, 6, 7

[12] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 5

[13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 5

[14] Isaac Robinson, Peter Robicheaux, Matvei Popov, Deva Ramanan, and Neehar Peri. Rf-detr: Neural architecture search for real-time detection transformers, 2025. 5

[15] Zhicheng Sun, Jinghan Li, and Yadong Mu. Exploring orthogonality in open world object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17302–17312, 2024. 5, 6, 7

[16] Yanghao Wang, Zhongqi Yue, Xian-Sheng Hua, and Hanwang Zhang. Random boxes are open-world object detectors. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6233–6243, 2023. 5

[17] 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. 4

[18] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022. 5

[19] Shan Zhang, Yao Ni, Jinhao Du, Yuan Xue, Philip Torr, Piotr Koniusz, and Anton van den Hengel. Open-world objectness modeling unifies novel object detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 30332–30342, 2025. 2, 3, 5, 6, 7

[20] 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

[21] Orr Zohar, Kuan-Chieh Wang, and Serena Yeung. Prob: Probabilistic objectness for open world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11444–11453, 2023. 5, 6, 7