

第9回ディスカッション

近況報告， 物体検出の継続学習

中部大学大学院修士1年 今井孝洋

<http://mprg.jp>



MPRG

MACHINE PERCEPTION AND ROBOTICS GROUP

- 近況報告
 - MIRU2024参加報告
 - インターン参加
 - ACCV2024のリバットル
- 物体検出の継続学習
 - クラス増分物体検出
 - ドメイン増分物体検出

- 8月やったこと
 - MIRU2024
 - OKIインターン
 - ACCV2024リバトル

2024年8月

<

今日

>

日	月	火	水	木	金	土
28日	29日	30日	31日	8月1日	2日	3日
4日	5日	6日	7日	8日	9日	10日
		MIRU2024				
11日	12日	13日	14日	15日	16日	17日
山の日	振替休日					
18日	19日	20日	21日	22日	23日	24日
	OKIインターン					
25日	26日	27日	28日	29日	30日	31日
OKIインターン						
	ACCV2024-リバトル					
9月1日	2日	3日	4日	5日	6日	7日
ACCV2024-リバトル						

- 期間：2024年8月6日（火）～9日（金）
- 場所：熊本城ホール
- 頂いた指摘・意見
 - マルチクロップの計算コストは？オンライン継続学習なら処理時間って短い方が良いのでは？
 - データセットの使い方は？データストリームとは何か？
 - クラス分類以外のタスクに適応できないのか？
 - 破滅的忘却に対する対処はしないのか？

記憶制御コードの最適化による選択的忘却

琴寄有貴[†] 後藤優太[†] 柴田剛志^{††} 木村昭悟^{†††} 入江豪[†]
[†]東京理科大学 ^{††}日本電気株式会社 ^{†††}日本電信電話株式会社

概要

- 「忘れられる継続学習」の研究
- 記憶/忘却の制御コードを最適化する新たな忘れられる継続学習方式の提案
- 制御コードを最適化することで既存法と比較して精度向上を達成

背景

継続学習で全ての学習済みクラスの精度を保ち続けることが有益とは限らない
⇒データ漏洩やプライバシー保護の問題

ex.顔認証による入館システム

新入社員 → 入館システム → 退職社員

記憶 → 忘却

学習 with Selective Forgetting (LSF) [Shibata+, IJCAI21]

LSFの概要

入力画像 → 特徴空間 → 記憶したいクラス → 忘却したいクラス

特徴空間上の線形性を利用

LSFの問題点

ニーモニックコードはランダムに生成された後変化しない

クラス内特徴と大きく乖離する恐れがある

手法：提案方式

新しいタスク（追加学習） → CNN → “重” 既存手法と同様

古いタスク（記憶保持） → CNN → “重” コードの最適化

古いタスク（忘却保持） → CNN → “大” 既存手法と同様

コードの最適化で期待すること

LSF → Ours

結果

既存手法との精度比較

全てのデータセットにおいて提案法が最高精度を達成

手法	CIFAR-100	CUB
Vanilla	51.79	42.23
LwF	17.23	15.20
LwF+	68.24	44.54
EWIC	48.57	41.36
EWIC+LwF+	49.61	42.08
MAS	67.64	43.42
MAS+LwF+	47.46	45.07
LSF	66.35	47.49
LSF+	73.21	57.97
LSFw	79.60	61.41
Ours	80.45	76.93

タスク数の違いによる精度変化

全てのタスク数においてLSFを上回る精度を達成

Number of task

解像度の違いによる精度変化

解像度が大きいほど精度向上を確認

Resolution

IS2-27

Abstract

Panoptic Lifting [Siddiqui, et al., 2023], a combination of novel view synthesis and segmentation, is useful for 3D scene recognition. However, in practical applications, there may be a lack of training data in some part of the scene and the model will have to be rebuilt. To rebuild the model efficiently, we propose Incremental Learning of Panoptic Lifting as an approach to update the model with new data.

And to prevent Catastrophic Forgetting, the loss of previous knowledge after model update, we introduce a uniform camera pose selection across the scene.

IS2-027

破滅的忘却抑制を考慮したカメラポーズ選択による Panoptic Liftingの増分学習

○ 幸壬 晃 (滋賀大, 理研) 園頭 元春 (理研) 飯山 将見 (滋賀大) 川西 康友 (理研)

テーマ：任意視点画像生成とセグメンテーションの組み合わせ手法

Guardian Robot Project

問題提起：観測データの不足箇所では推論精度が低下する

入力：様々な視点からのRGB, Panoptic Segmentation, カメラポーズ

出力：任意視点からのRGB, Depth, Panoptic Segmentation

問題：破滅的忘却 (Catastrophic Forgetting)

初期値として暗黙的に保持されている学習済みの知識が増分学習によって書き換えられ、消失してしまう現象

学習済み視点の精度が低下

対処法：学習済みデータの明示的な再学習

増分のデータに加え、学習済みデータの一部分を学習に用いることで破滅的忘却を抑制する

最小限のデータで破滅的忘却を抑制できれば十分な精度のモデル構築にかかる時間を低減できる

「カメラ位置的に等間隔な視点」の画像選択法

シーンを空間的に満遍なく観測している画像集合は、三次元再構成のための情報を多く含むと考えられる

例) 1%の選択

例) 30%の選択

提案手法：Panoptic Liftingにおける増分学習

増分学習 (Incremental Learning)

学習済みの知識を保持したまま新たに得られたデータにも適応するモデルの構築

学習済みモデル → 増分データの学習 → 新たなモデル

パラメータを初期値として再利用

問題：破滅的忘却 (Catastrophic Forgetting)

初期値として暗黙的に保持されている学習済みの知識が増分学習によって書き換えられ、消失してしまう現象

学習済み視点の精度が低下

対処法：学習済みデータの明示的な再学習

増分のデータに加え、学習済みデータの一部分を学習に用いることで破滅的忘却を抑制する

最小限のデータで破滅的忘却を抑制できれば十分な精度のモデル構築にかかる時間を低減できる

「カメラ位置的に等間隔な視点」の画像選択法

シーンを空間的に満遍なく観測している画像集合は、三次元再構成のための情報を多く含むと考えられる

例) 1%の選択

例) 30%の選択

実験・結果

3種類のモデルを比較し、増分学習の有効性を検証した

PREモデル (学習済み視点のみ) vs FULLモデル (学習済み視点+増分視点) vs INC+モデル (学習済み視点+増分視点+PREモデルを初期値)

増分視点の知識を持たない 既存手法によるモデル vs 学習済み視点+増分視点 vs PREモデルを初期値

増分視点の知識を持たない 既存手法によるモデル vs 学習済み視点+増分視点 vs PREモデルを初期値

用いたデータセット

Replica (room_0) [Strack, et al., 2019] vs 自作データセット

合計: 100K vs 合計: 454K

学習済み視点: 719 vs 学習済み視点: 168

増分視点: 181 vs 増分視点: 86

検証データ: 1490 vs 検証データ: 1490

増分視点における各モデルの推論精度

PREモデルからの増分学習により精度が向上

モデル	PSNR	mIoU	PQ
PRE	31.3	0.222	0.302
FULL	27.1	0.060	0.472
INC+	28.5	0.078	0.542
INC+	28.3	0.045	0.537
INC+	28.2	0.065	0.488
INC+	28.0	0.447	0.473

学習済み視点における各モデルの推論精度

選択割合を増加させることにより破滅的忘却が抑制

モデル	PSNR	mIoU	PQ
PRE	31.3	0.222	0.302
FULL	27.1	0.060	0.472
INC+	28.5	0.078	0.542
INC+	28.3	0.045	0.537
INC+	28.2	0.065	0.488
INC+	28.0	0.447	0.473

計算時間 (1エポックあたりの平均 [s])

Replica (room_0) vs 自作データセット

約36%に削減 vs 約34%に削減

次の方針

カメラ位置の移動は少なくカメラ方向の移動が大きい領域は選択されにくい

カメラの方向やオブジェクトの配置も考慮した画像選択法の導入

- 期間：2024年8月19日（月）～30日（金）
- 場所：埼玉県蕨市 OKI蕨システムセンター
- インターン中の研究テーマ
 - 工場機械における音響異常検知
- 他の研究テーマ
 - 数値最適化
 - 画像系（物体検出，行動検知など）
 - 音響系



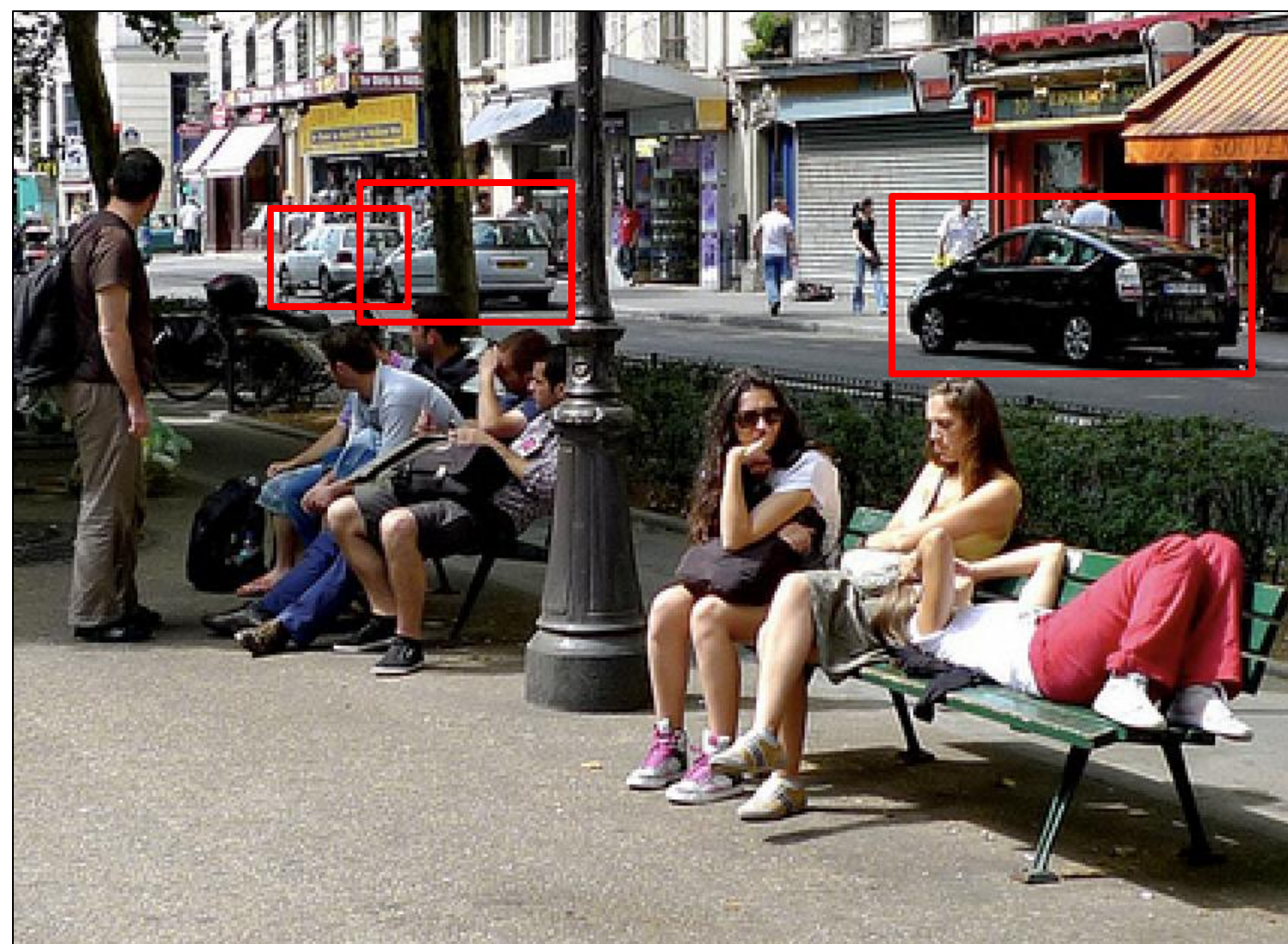
日程次第で長期インターンも受け入れ可能

- 期間：2024年8月27日（火）～9月2日（月）
- 頂いた指摘
 - データ選択
 - なぜコサイン類似度を使用？ユークリッド距離などではだめなのか？
 - データ選択の有効性は？提案手法の有効性は損失関数に依存するのでは？
 - 計算コスト
 - 提案手法の損失関数は計算コストが増加する。オンライン継続学習としてはデメリットではないか？
 - 損失関数
 - TCR損失の有効性は？TCR損失についてアブレーションすべき。
 - 損失関数のオンライン継続学習に対する貢献はあるのか？

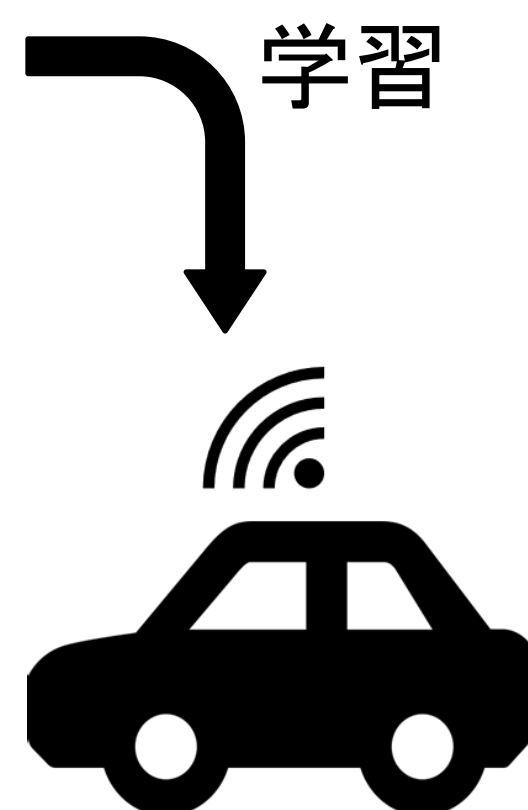
- 継続学習
 - 過去に学習した知識を保持しながら新しい知識を継続的に学習する能力
- 従来の継続学習
 - クラス分類タスクにおける継続学習がほとんど
 - 物体検出やセグメンテーションタスクに関する継続学習は少ない
- 物体検出における継続学習の問題設定
 - クラス増分物体検出 (CIOD)
 - ドメイン増分物体検出 (DIOD)

クラス増分物体検出 (CIOD)

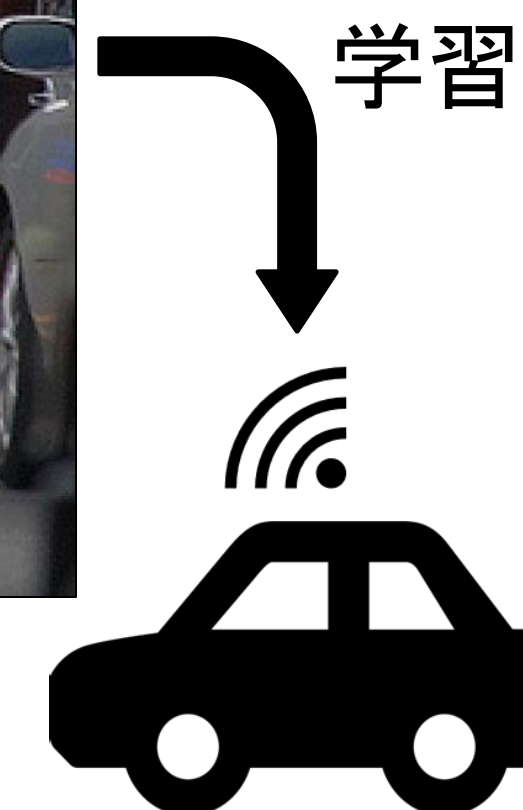
- 検出したい物体クラスが増加する継続学習の問題設定



車にアノテーションした学習データ



人にアノテーションした学習データ



クラス増分物体検出 (CIOD)

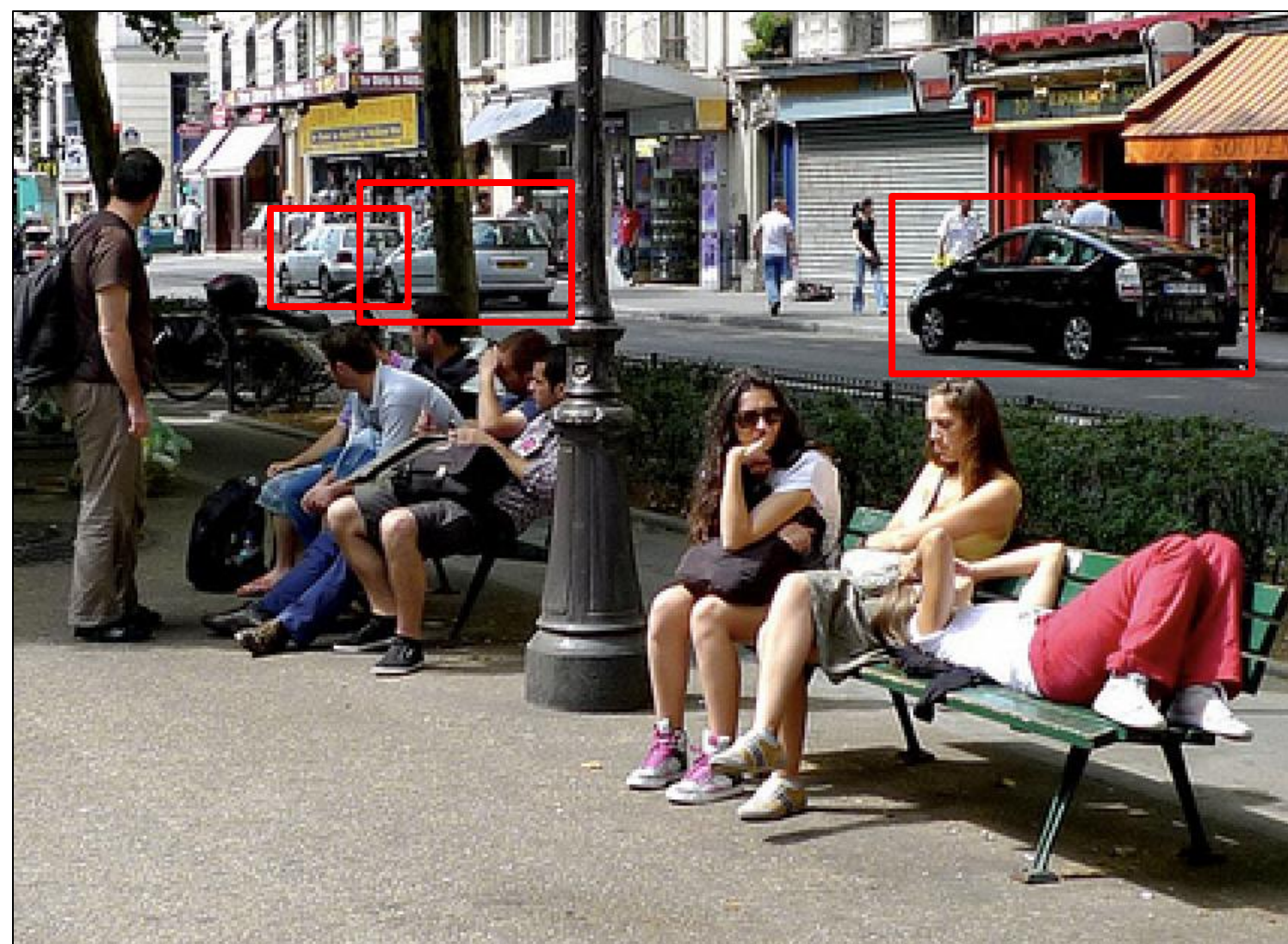
- 検出したい物体クラスが増加する継続学習の問題設定

車 : ○ 人 : —

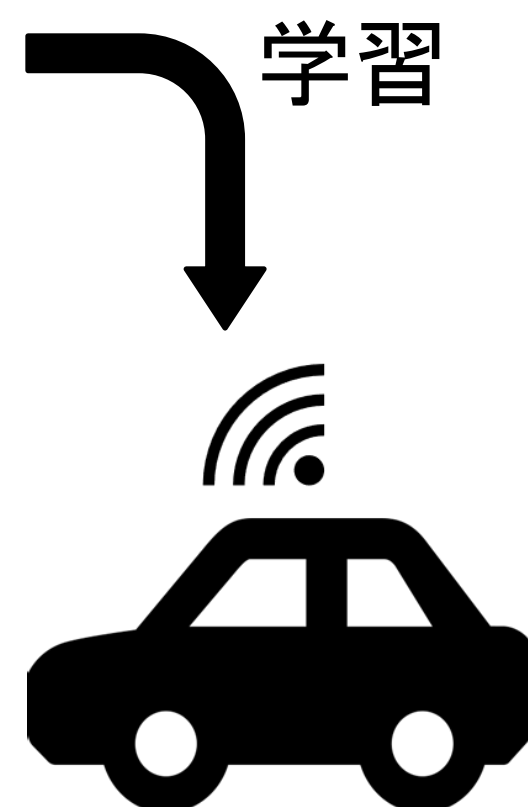
過去に学習した
クラスの知識を保持

新しいクラス
にも適応

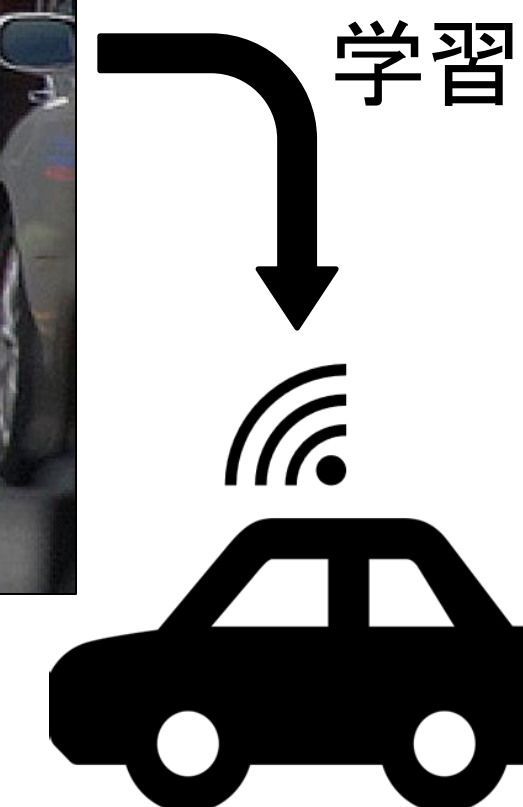
車 : ○ 人 : ○



車にアノテーションした学習データ

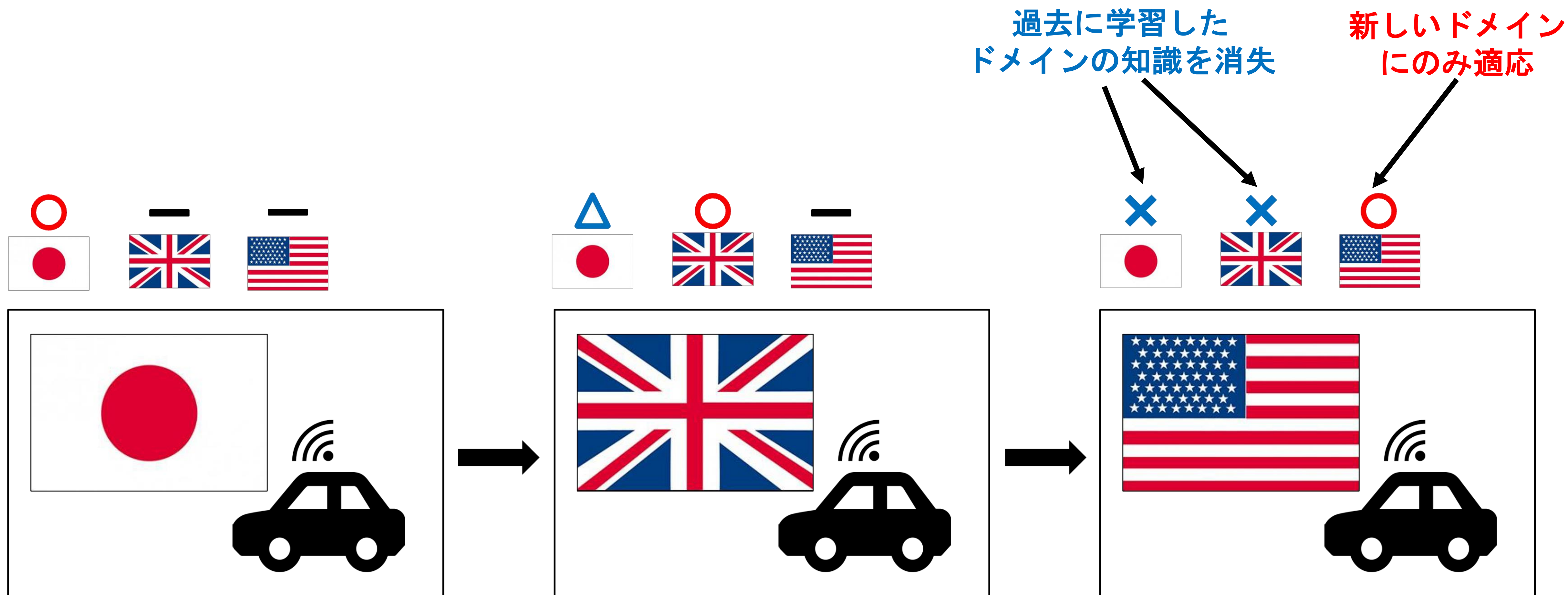


人にアノテーションした学習データ



ドメイン増分物体検出 (DIOD)

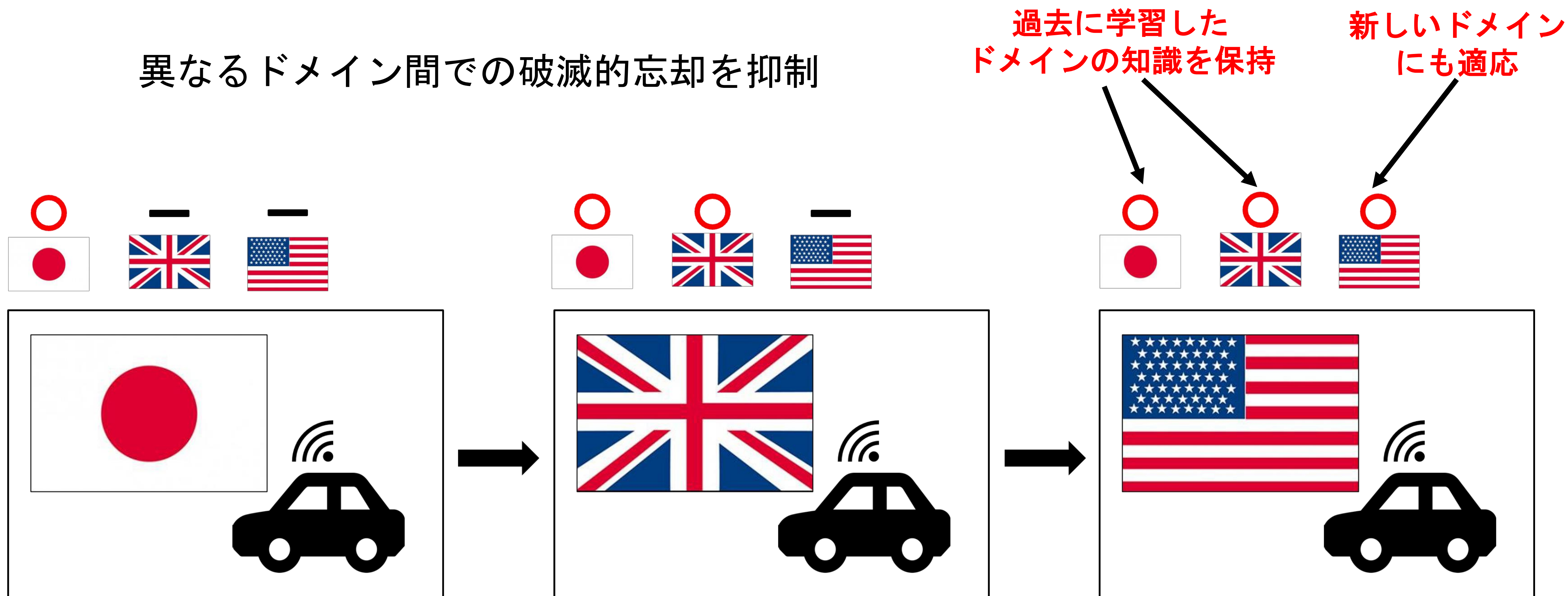
- ドメインが増加する継続学習の問題設定
 - 新しいドメインで学習した時, 古いドメインでの検出精度が低下



ドメイン増分物体検出 (DIOD)

- ドメインが増加する継続学習の問題設定
 - 新しいドメインで学習した時, 古いドメインでの検出精度が低下

異なるドメイン間での破滅的忘却を抑制



- 近況報告
 - MIRU2024参加報告
 - インターン参加
 - ACCV2024のリバttl
- 物体検出の継続学習
 - 物体検出勉強

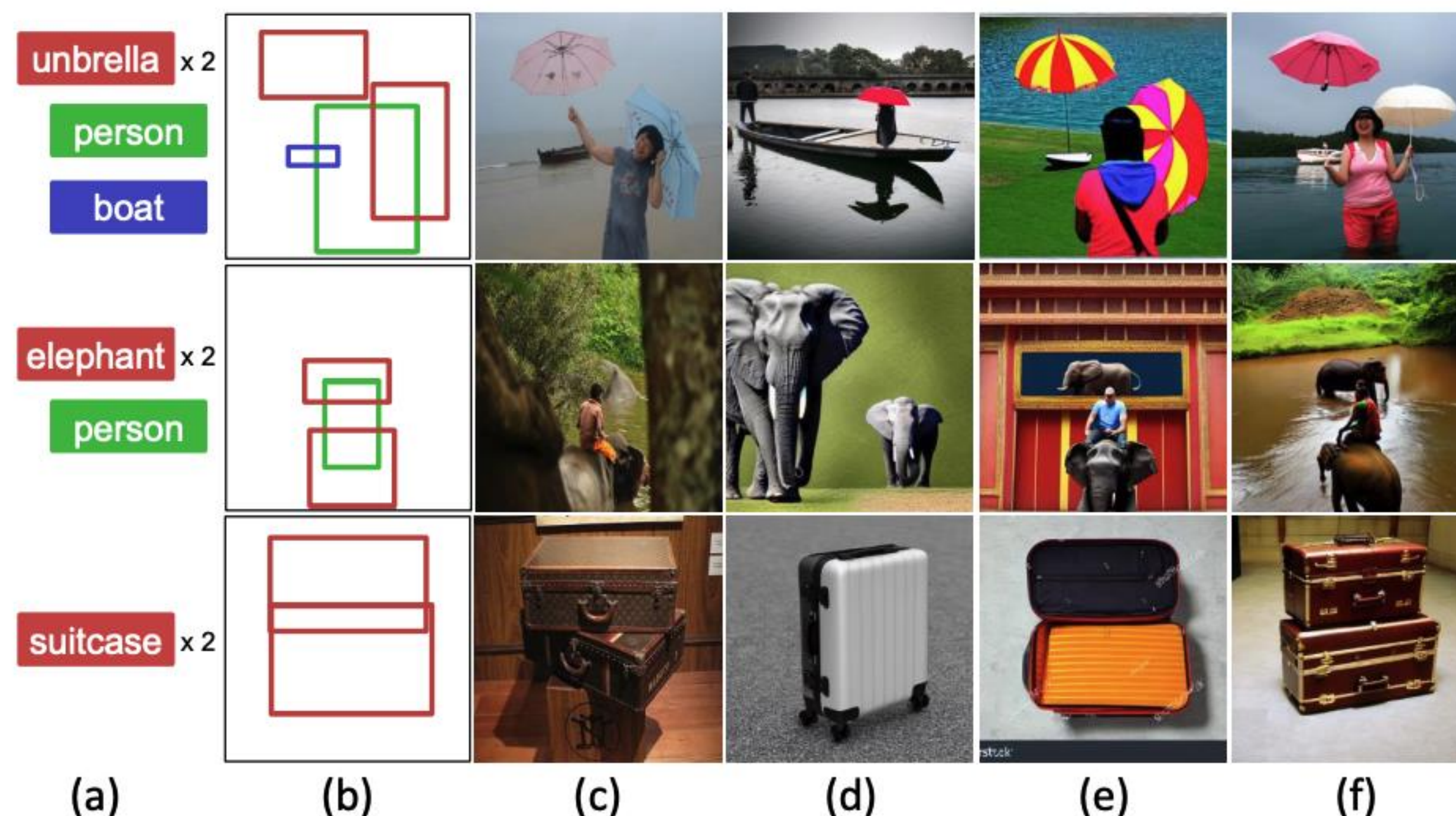
資料置き場

- 事前学習済みの拡散モデルを使用したクラス増加型物体検出法SDDGRを提案
 - Text2imageの生成モデルによって複雑なデータを生成
- 画像生成にテキスト, ラベル, BBoxを用いることで生成データを制御
 - 物体検出に適した学習用データの生成

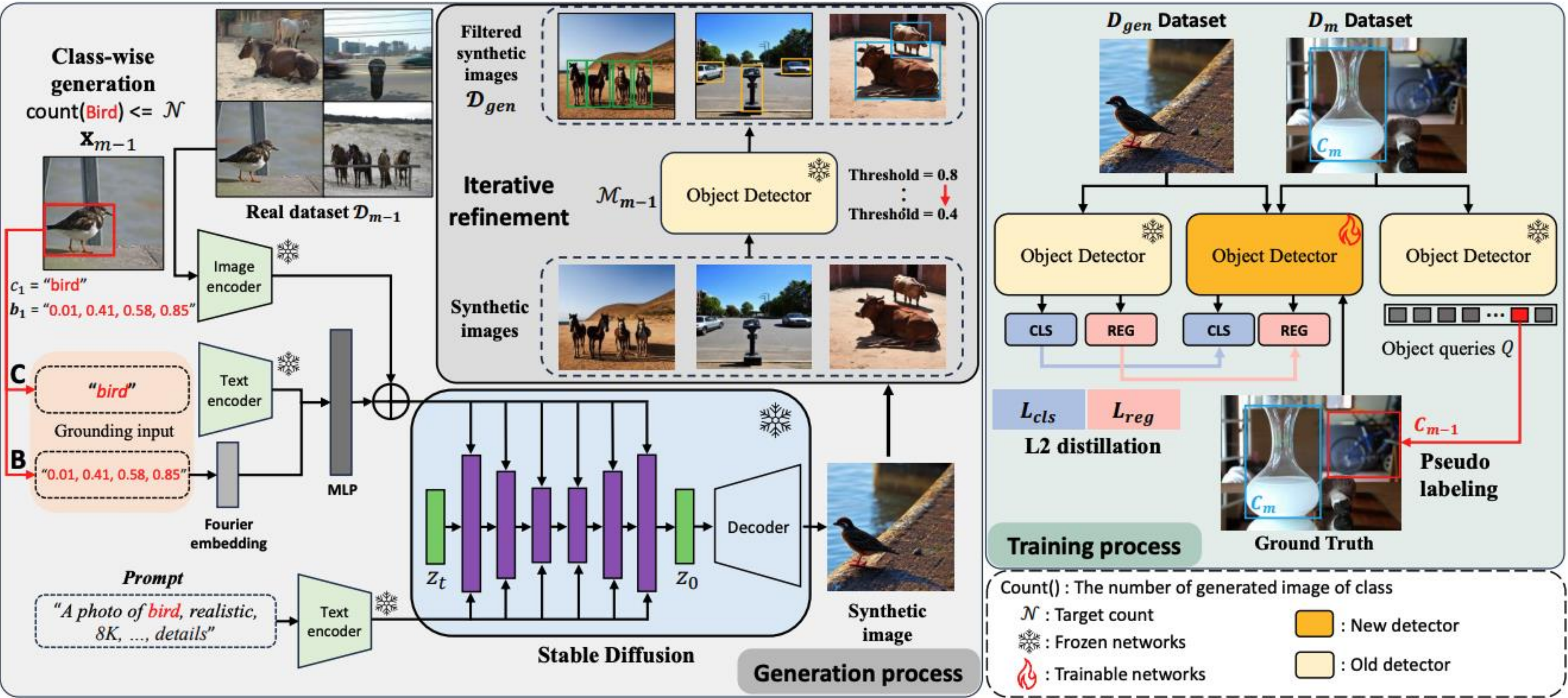
“A photo of two umbrellas, person and boat, realistic, ... details”

“A photo of two elephants and person, ...”

“A photo of two suitcases, ...”

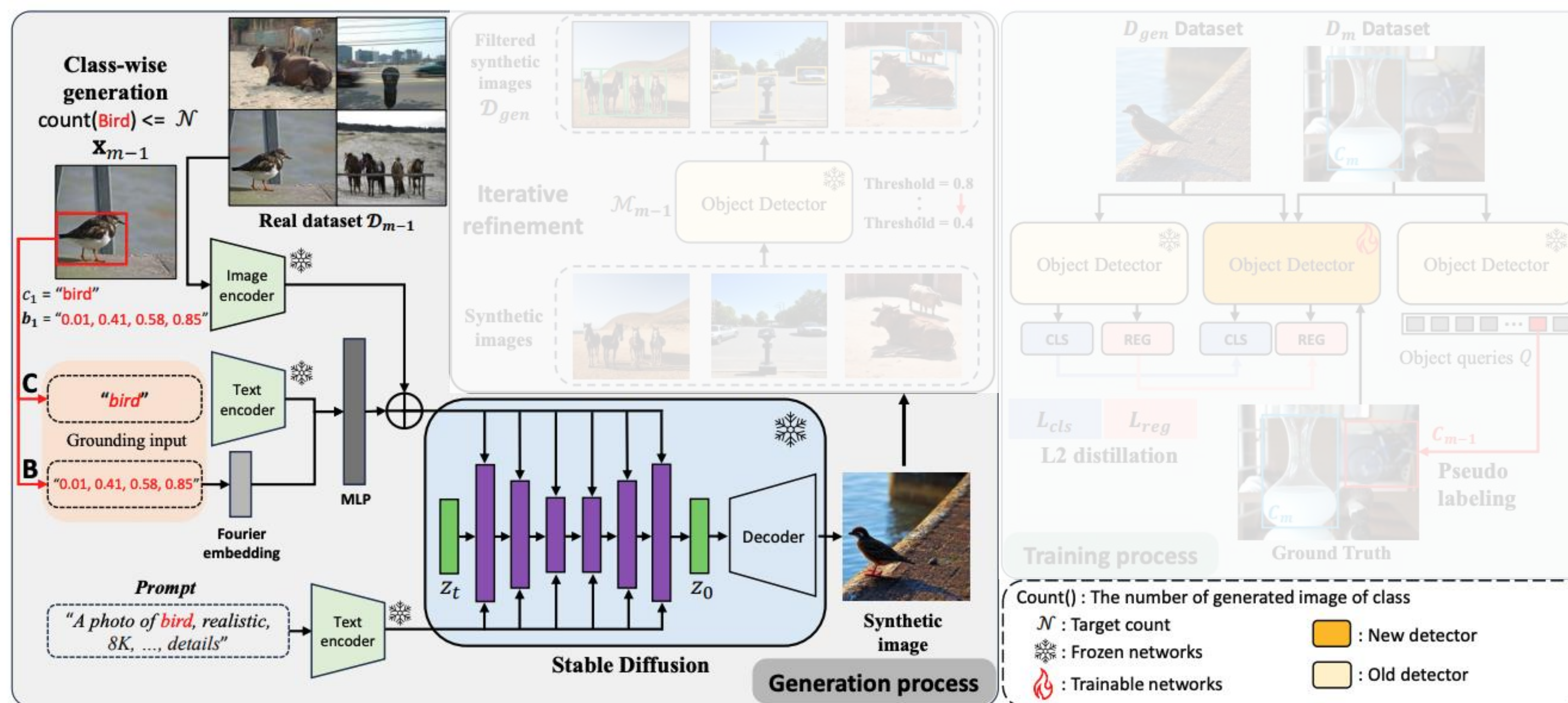


1. 物体検出に適した画像生成
2. 生成画像の選択による品質・忠実度向上
3. 擬似ラベリング
4. 生成画像による知識蒸留



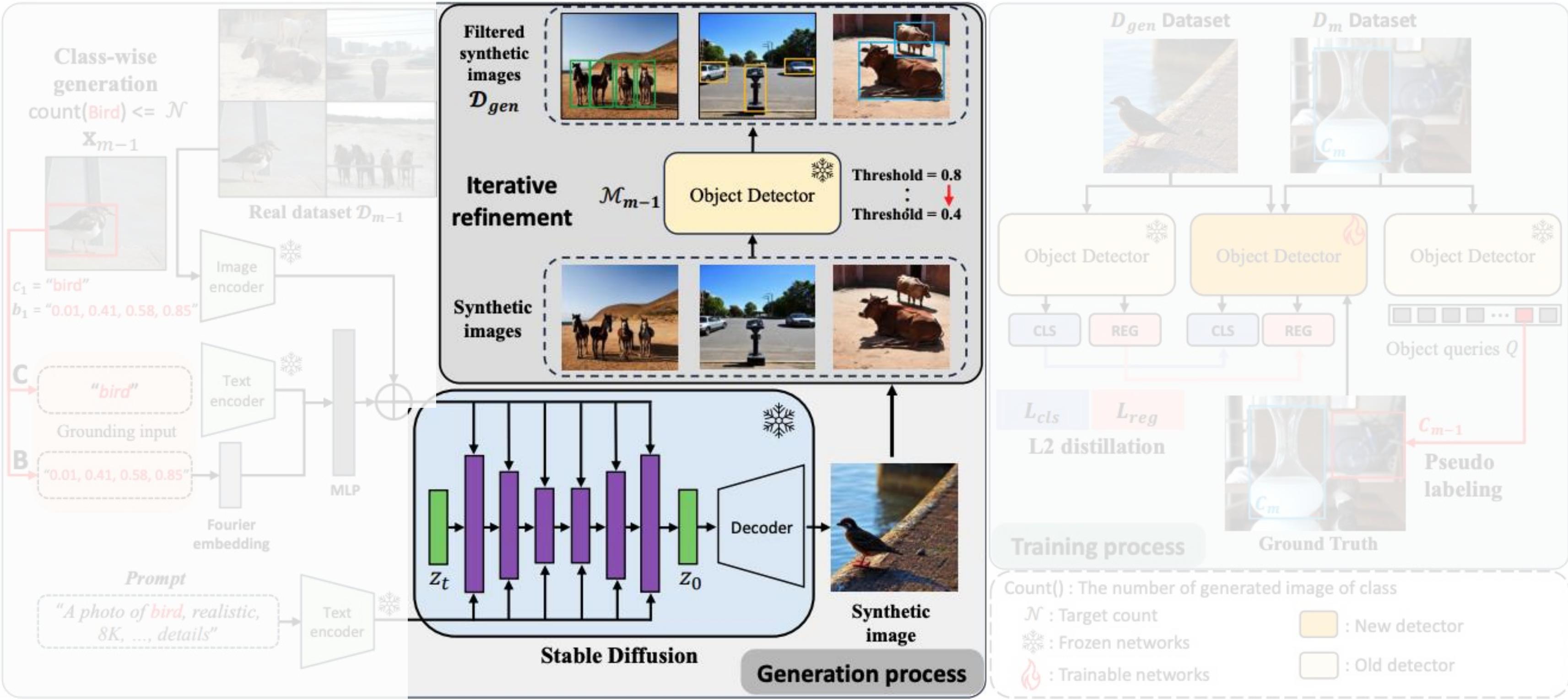
SDDGRの学習パイプライン

- 目的：マルチラベルを持つ複雑な画像を生成
 - クラス，位置座標，テキストプロンプトにより生成画像を制御



生成画像の選択による品質・忠実度向上

- 目的：実画像に近い生成画像のみを選択し学習に使用
 - 生成画像と過去モデルから

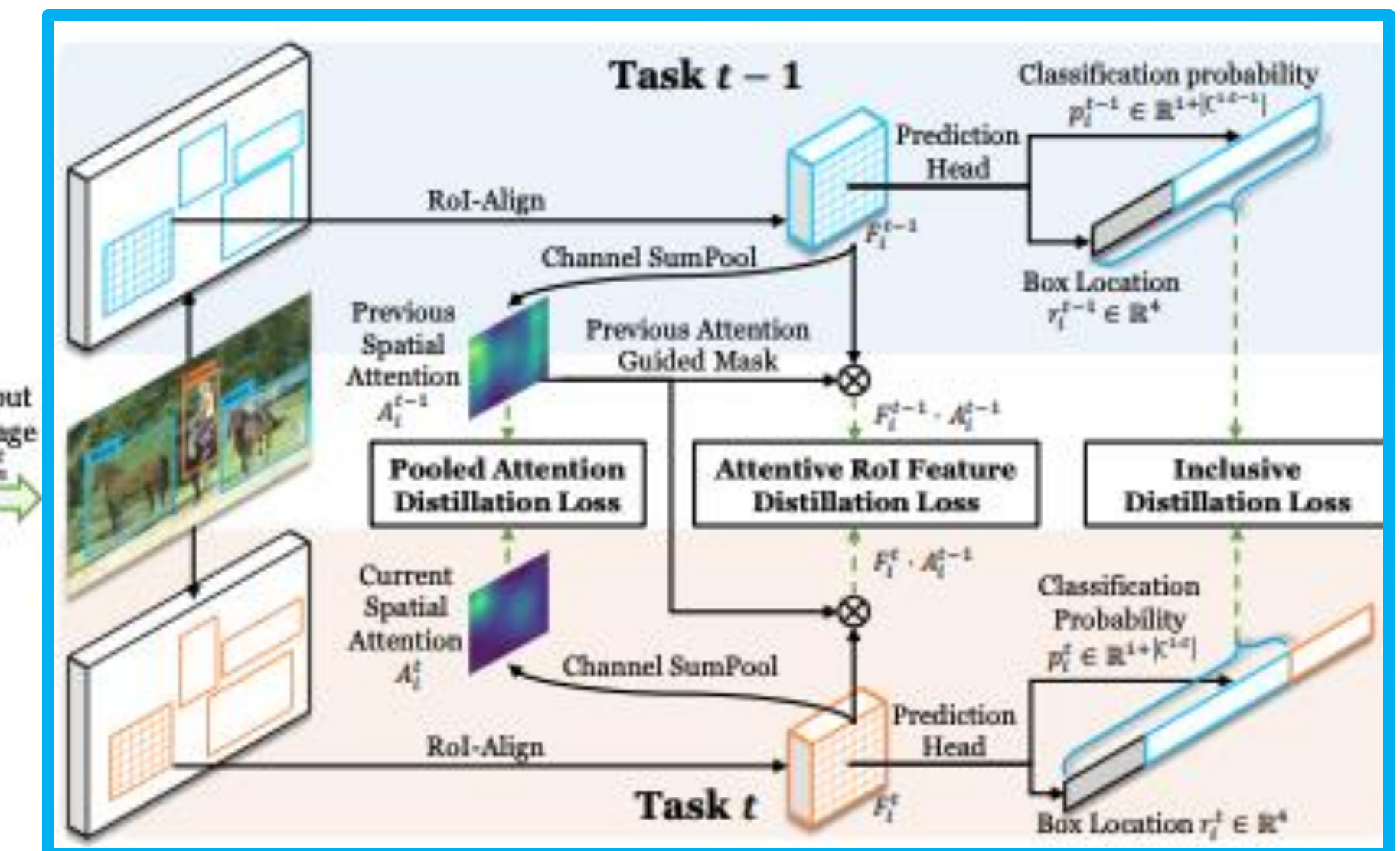
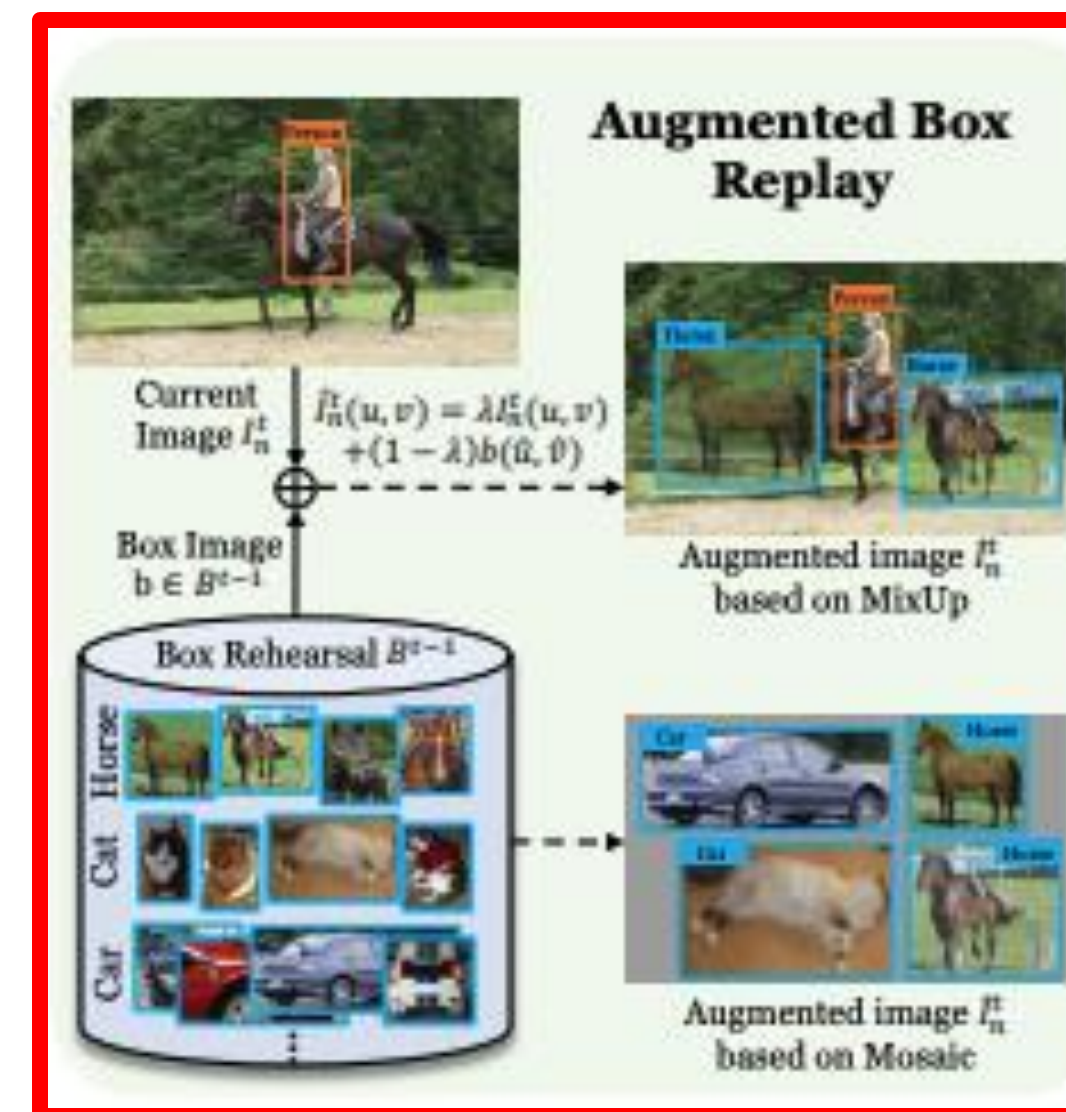


Augmented Box Replay: Overcoming Foreground Shift for Incremental Object Detection [Y. Liu+, ICCV'23]

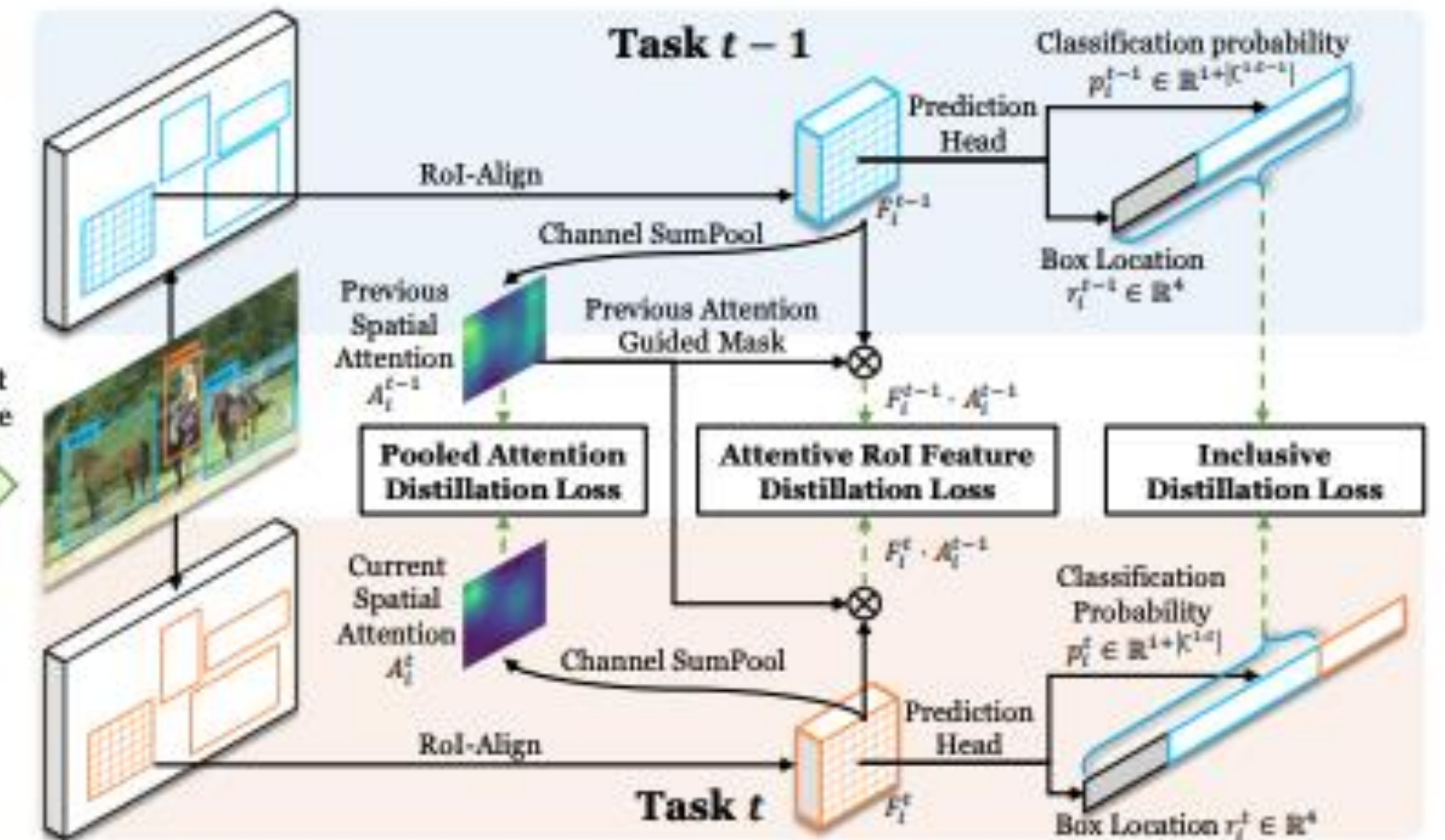
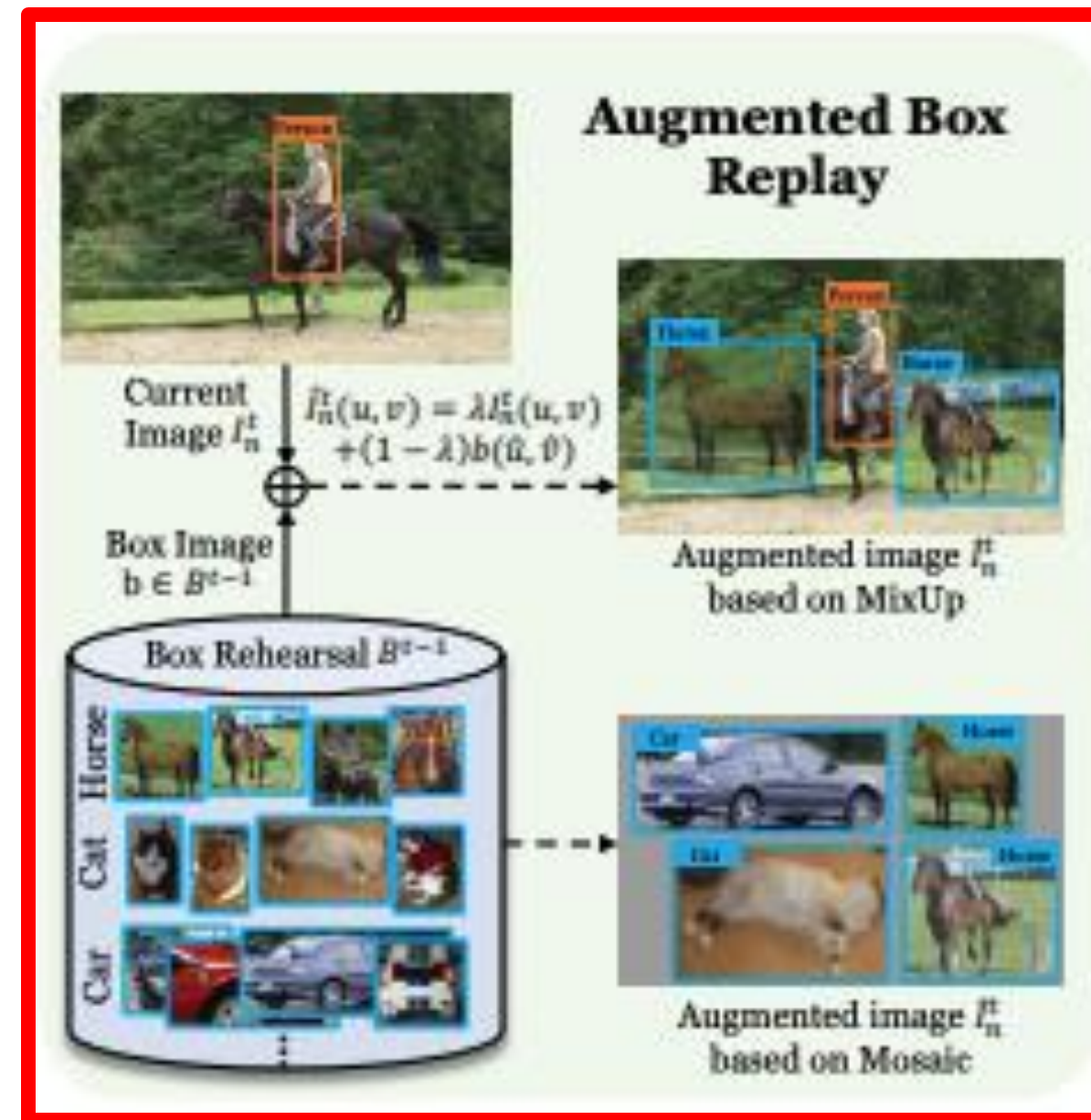
- クラス増加型物体検出における前景シフトに対処した手法を提案
 - 前景シフト：過去タスクのデータの背景に現在タスクのクラス C が存在
 - クラス C を背景と物体の両方として扱う矛盾した学習によってモデルの性能が低下

2つのアプローチから構成

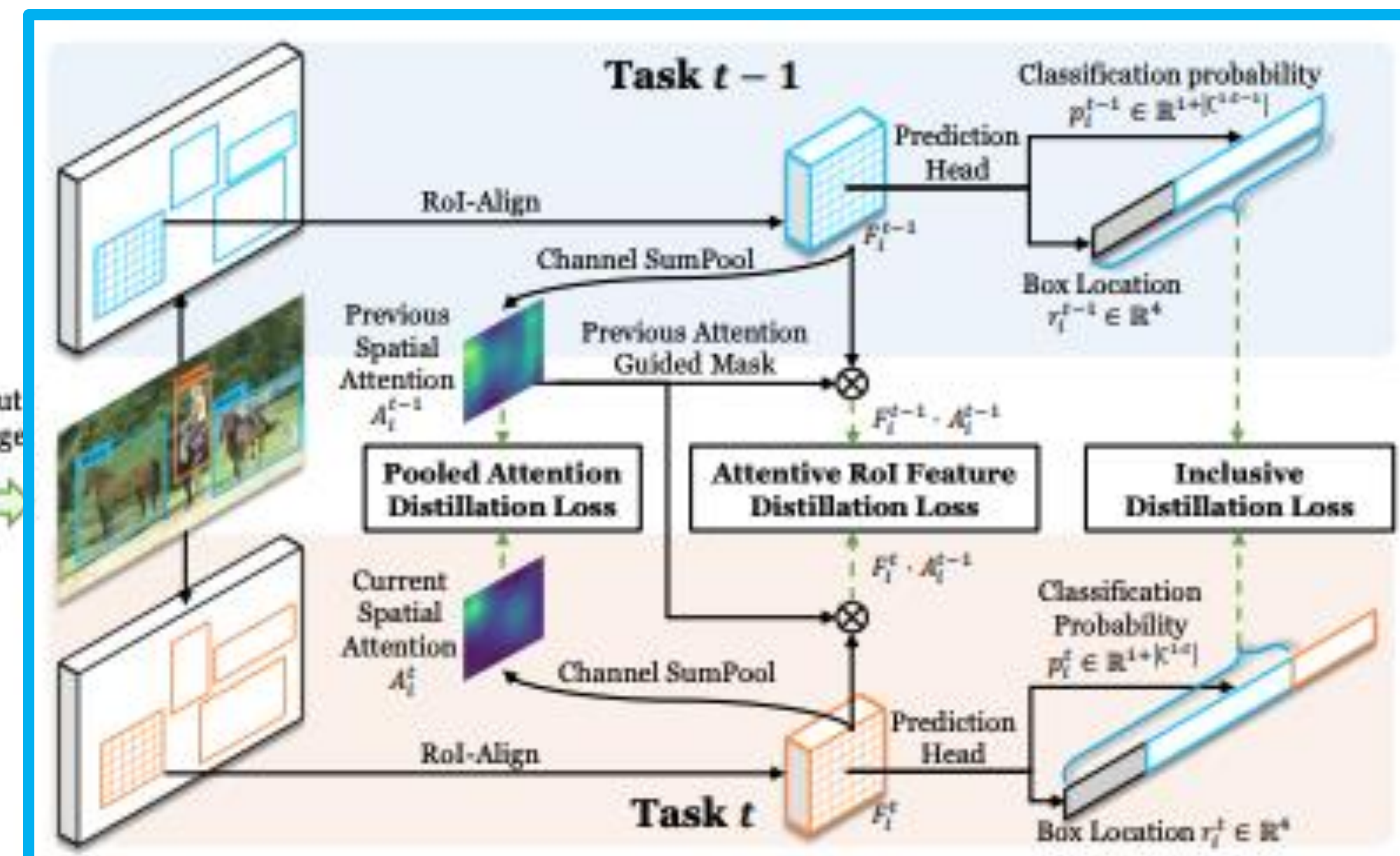
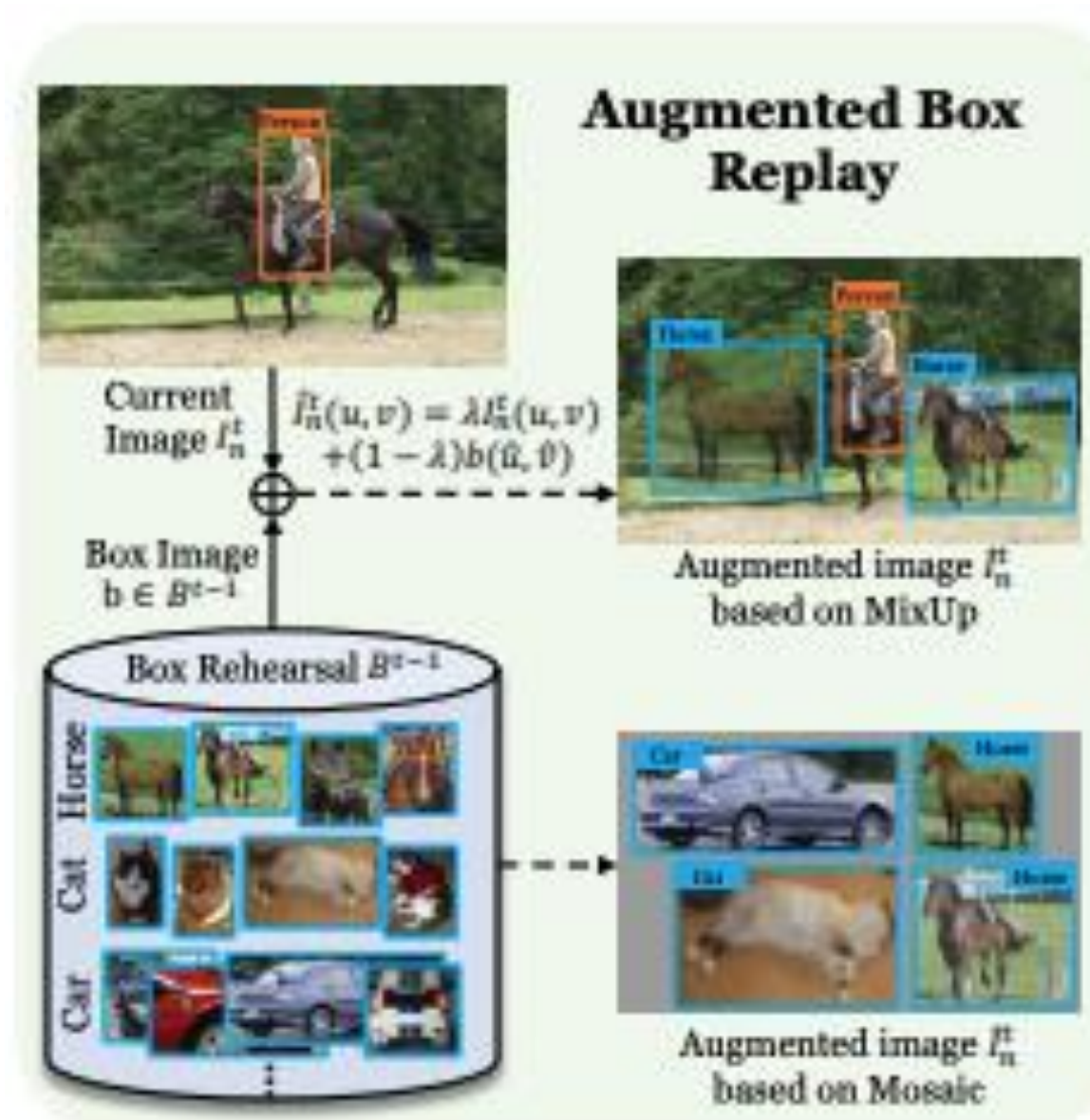
- Augmented Box Replay (ABR)
 - 前景シフトに対処
- Region of Interest Distillationと Inclusive Loss
 - 破滅的忘却に対処



- 目的：新旧タスク間の前景シフトに対処
 - データ拡張によって物体検出に適した学習用データを生成
- 処理手順
 - タスク $t-1$ のデータからオブジェクトのみ切り取ってバッファに保存
 - タスク t のデータとバッファ内のオブジェクトを合成
 - MixUpとMosaicの2種類データ拡張を使用



- 目的：破滅的忘却への対処
 - 新旧モデルのAttention, 特徴マップ, 確率分布による知識蒸留
- 3種類の損失関数を使用
 - Pooled Attention Distillation Loss : Attentionに対する蒸留
 - Attentive RoI Feature Distillation Loss : 特徴マップに対する蒸留
 - Inclusive Distillation Loss : 確率分布に対する蒸留

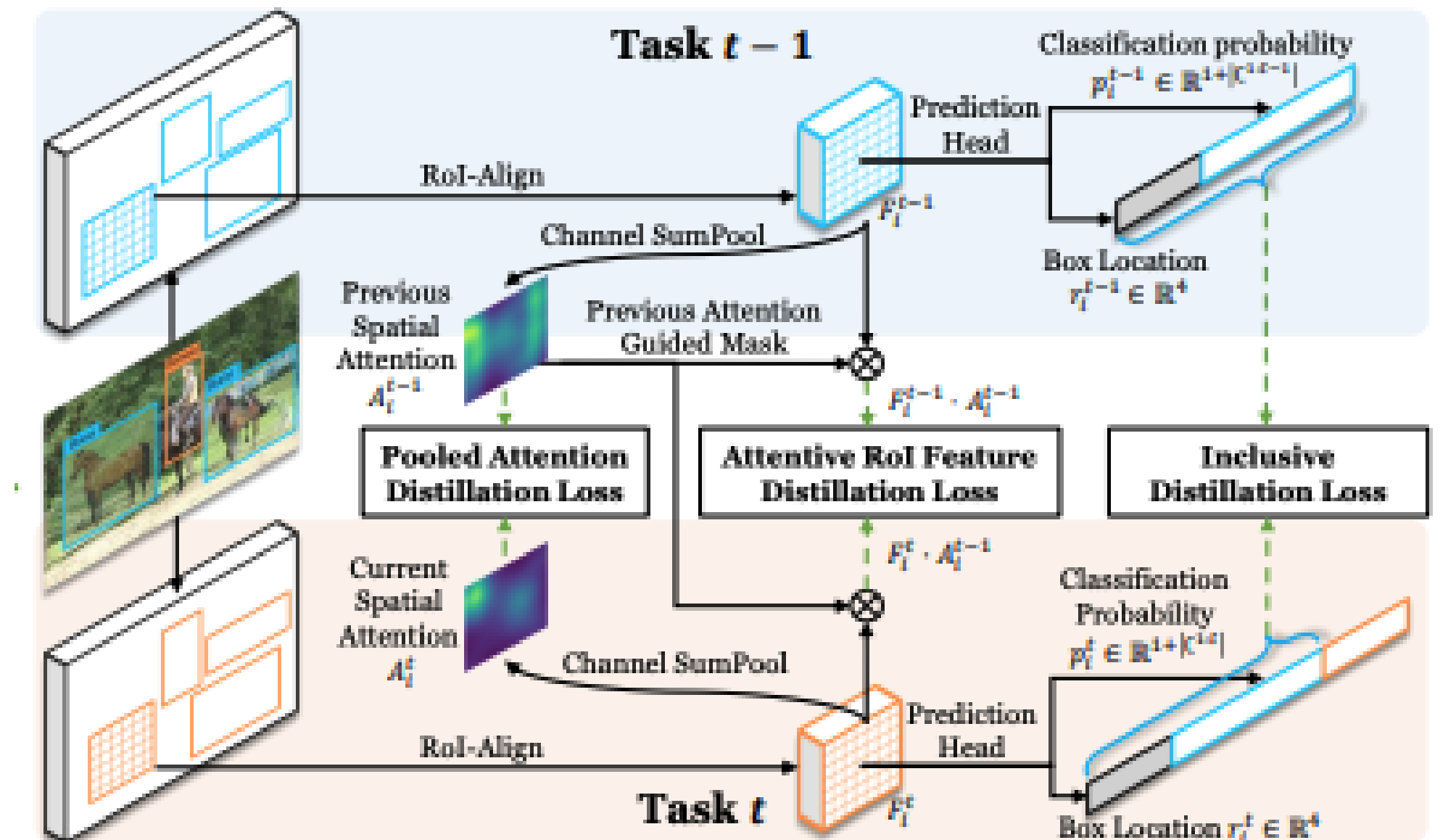


- Pooled Attention Distillation Loss : Attentionに対する蒸留

$$\mathcal{L}_{PAD} = \|A_i^{t-1} - A_i^t\|$$

- Attention RoI Feature Distillation Loss : 特徴マップに対する蒸留

$$\mathcal{L}_{AFD} = \frac{1}{P_n} \sum_{i=1}^{P_n} (F_i^{t-1} - F_i^t)^2 A_i^{t-1}$$



- Inclusive Loss with Background Constraint : ABR に適応した損失
 - クラス分類損失

$$\mathcal{L}_{IC} = \frac{1}{P_n^t} \sum_{i=1}^{p_n^t} c_i \begin{cases} \log \left(p_i^b + \sum_{c=1}^{C^{1:t-1}} p_i^c \right), & (c_i = c^b) \\ \sum_{c=1}^{C^{1:t}} c_i \log p_i^c, & (c_i \in C^{1:t}) \end{cases}$$

- 蒸留損失

$$\mathcal{L}_{ID} = \frac{1}{\Omega} \begin{cases} p_i^{b,t-1} \log \left(p_i^{b,t} + \sum_{c=1}^{C^t} p_i^{c,t} \right), & (c_i = c^b) \\ \sum_{c=1}^{C^{1:t-1}} p_i^{c,t-1} \log p_i^{c,t}, & (c_i \in C^{1:t}) \end{cases}$$

