Few-shot Adaptive Open-set Object Detection

with Personalized Scene Generation

Yuzuru Nakamura[1] · Yasunori Ishii[1] · Takayoshi Yamashita[2]

[1]Panasonic Holdings · [2]Chubu University

Motivation: deployment is rarely clean or closed

REAL-WORLD DEPLOYMENT

Object detectors degrade when the scene appearance changes.

At deployment time, two additional constraints often occur together:

Only a few target images + **Novel object classes appear**

Example: urban driving scenes with new weather, camera characteristics, and object categories.

Source | **Target**

Clear weather

→

Foggy weather

One city

→

Another city

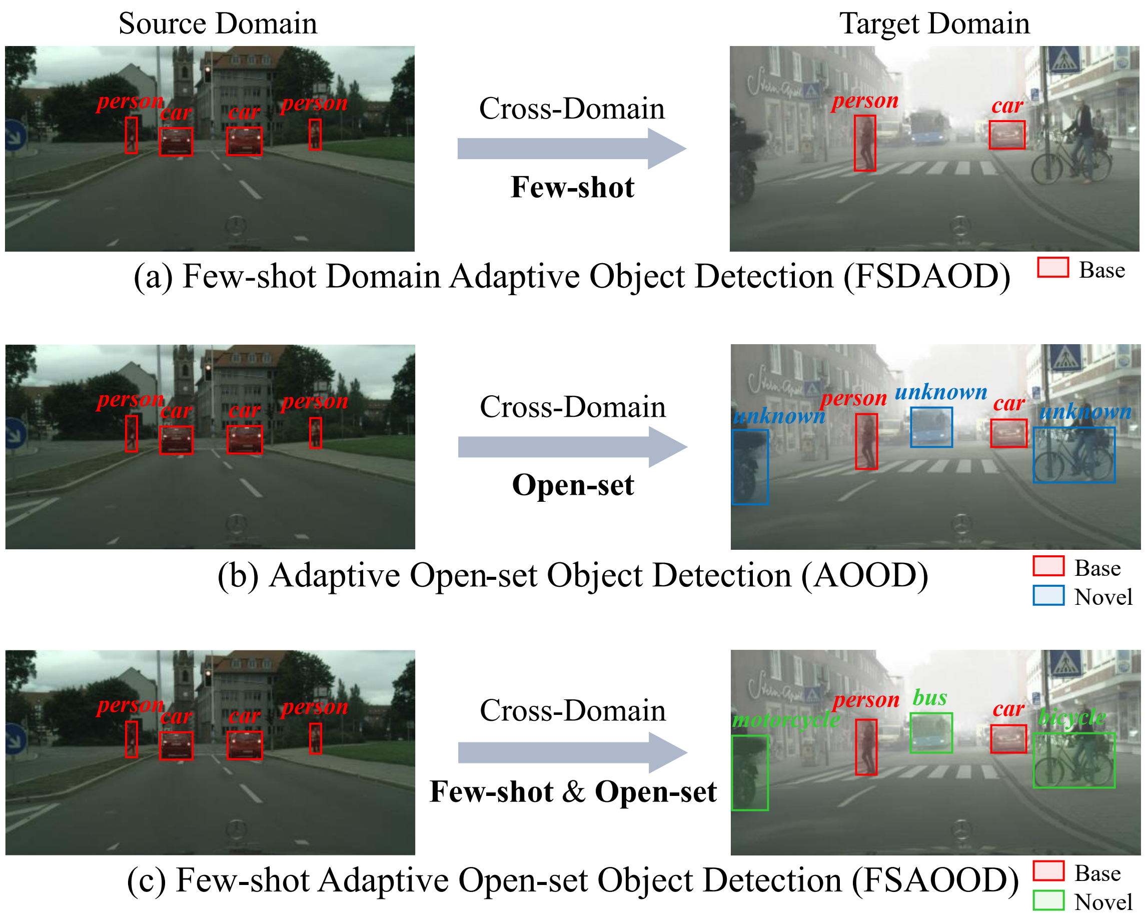
Synthesis

→

Real

Problem setting: few-shot + open-set adaptation

Existing settings cover only part of the challenge.



Problem Setting	Few-shot	Open-set
FSDAOD	✓	✗
AOOD	✗	✓
FSAOOD (Ours)	✓	✓

FSAOOD asks the detector to adapt with few target images while also recognizing novel target classes.

Key idea: generate the missing target-domain data

DATA-CENTRIC VIEW

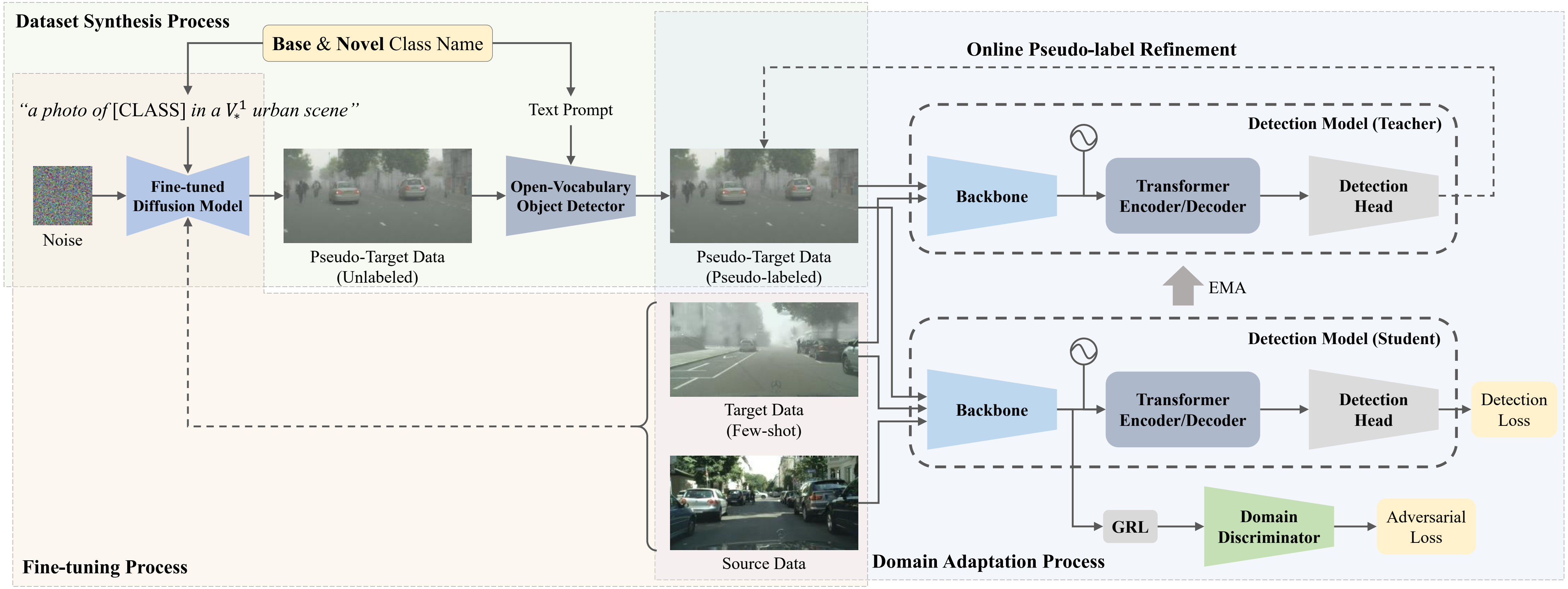
Few target images → Personalize a diffusion model → Generate target-style images with [CLASS]

Prompt control makes novel classes explicit.

“a photo of [CLASS] in a V urban scene”*

[CLASS] controls the object category; **V*** stores the target-domain appearance.

Method overview: generation, labeling, adaptation



1 **Personalize**

Fine-tune diffusion with a few target images.

2 **Synthesize**

Generate many target-style images using class prompts.

3 **Pseudo-label**

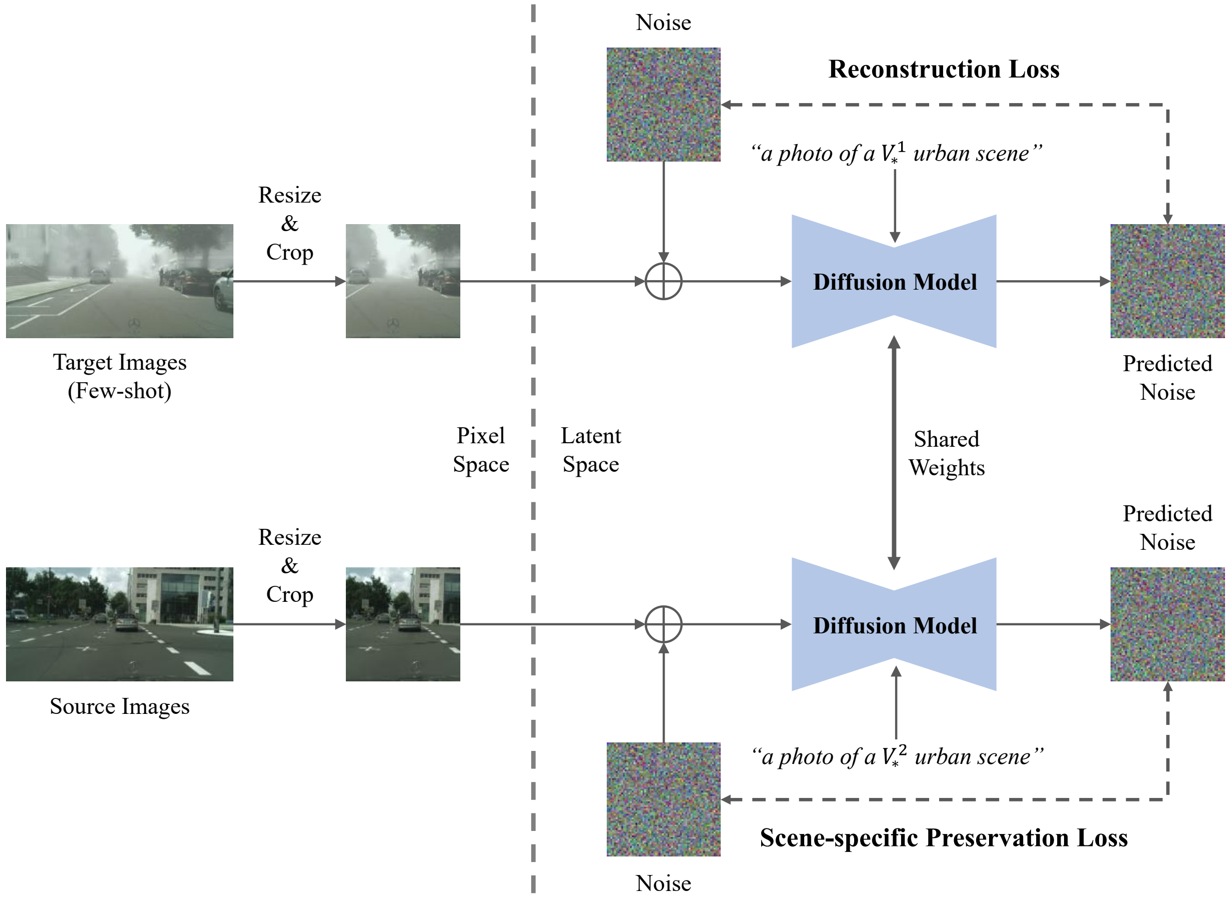
Use an open-vocabulary detector for base and novel classes.

4 **Adapt**

Train with domain adaptation and online refinement.

Panasonic

Personalize: avoiding overfitting during diffusion fine-tuning



Challenge

With only a few target images, fine-tuning can memorize them and reduce diversity.

Proposed solution

Scene-specific preservation loss uses source-domain scenes to preserve appearance and context diversity.

$$\mathcal{L}_{ldm} = \underbrace{\mathbb{E}_{\boldsymbol{z}^T, t, \boldsymbol{\epsilon}\sim\mathcal{N}(0,1)}\left[\left\|\boldsymbol{\epsilon} - \epsilon_\theta(\boldsymbol{z}_t^T, t, \tau_\theta(\boldsymbol{p}^T))\right\|_2^2\right]}_{\text{Target reconstruction}} + \lambda \underbrace{\mathbb{E}_{\boldsymbol{z}^S, t', \boldsymbol{\epsilon}'\sim\mathcal{N}(0,1)}\left[\left\|\boldsymbol{\epsilon}' - \epsilon_\theta(\boldsymbol{z}_{t'}^S, t', \tau_\theta(\boldsymbol{p}^S))\right\|_2^2\right]}_{\text{Source preservation}}$$

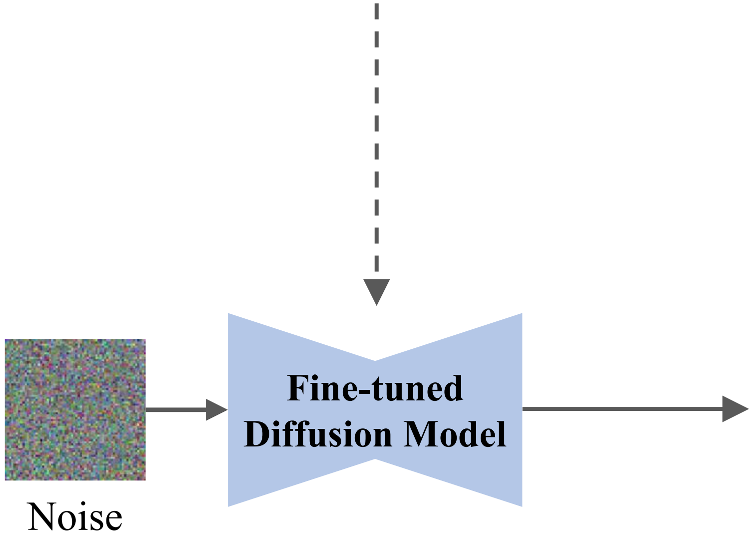
Synthesize: personalized scene generation

After fine-tuning, the unique token captures the target scene style.

a photo of **[CLASS]** in a **V*** urban scene

[CLASS] can be any base or novel category.

V* is learned from only a few target images.



Example generated images after personalization

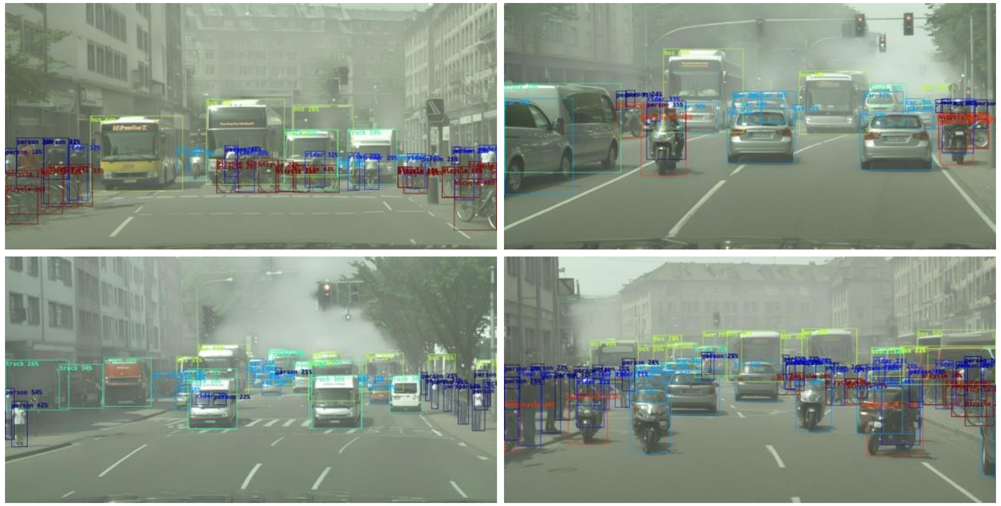
Pseudo-label and adapt

Generated images need bounding boxes and class labels.

Synthetic target image → Open-vocabulary detector → Initial pseudo-labels → Adapted detector

Online refinement

During training, update pseudo-labels using confident predictions from the current model.



Example predicted pseudo-labels by open-vocabulary detector

Experimental setup

We evaluate our approach in both open-set and closed-set settings.

8 target images

10k synthetic images

Problem settings

FSAOOD

Few target images
+ novel target classes

Target shots are unlabeled. The detector must recognize base and novel classes.

FSDAOD

Few target images
+ same class set

Target shots are labeled. The goal is closed-set domain adaptation.

Domain scenarios

Cityscapes → Foggy Cityscapes — adverse weather; 3 base / 3-5 novel classes

Cityscapes → BDD100k — cross city; 3 base / 2-4 novel classes

4 sub-tasks with different base / novel class split settings

Cityscapes → Foggy Cityscapes — adverse weather; 8 classes

SIM10k / KITTI → Cityscapes — sim2real / cross camera; 1 class

Evaluation metrics

mAP_b — mean Average **Precision** for **base** class

mAP_n — mean Average **Precision** for **novel** class

mAR_n — mean Average **Recall** for **novel** class

mAP — mean Average Precision

AP_{car} — Average Precision for car class

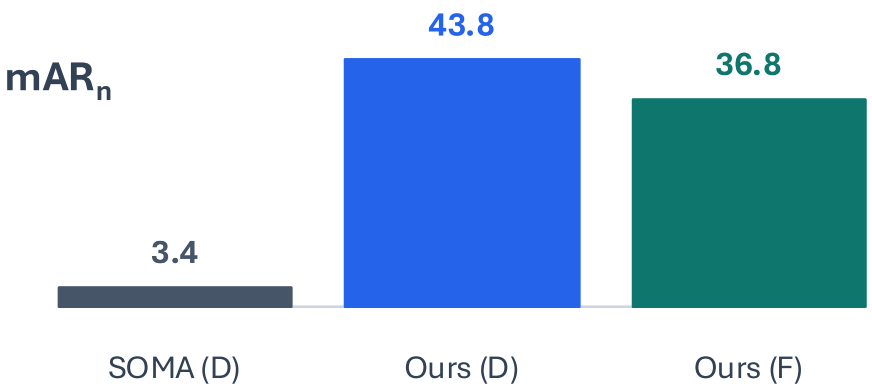
Panasonic

Main results: FSAOOD setting

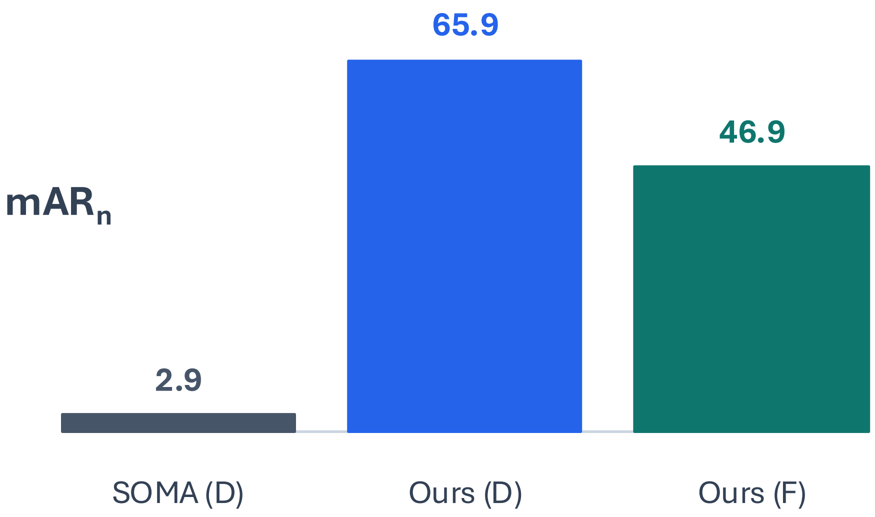
Our method significantly improves novel-class detection in the few-shot open-set setting.

D: Deformable DETR
F: Faster R-CNN

Cityscapes → Foggy Cityscapes
het-sem, 3 novel classes



Cityscapes → BDD100k
het-sem, 2 novel classes



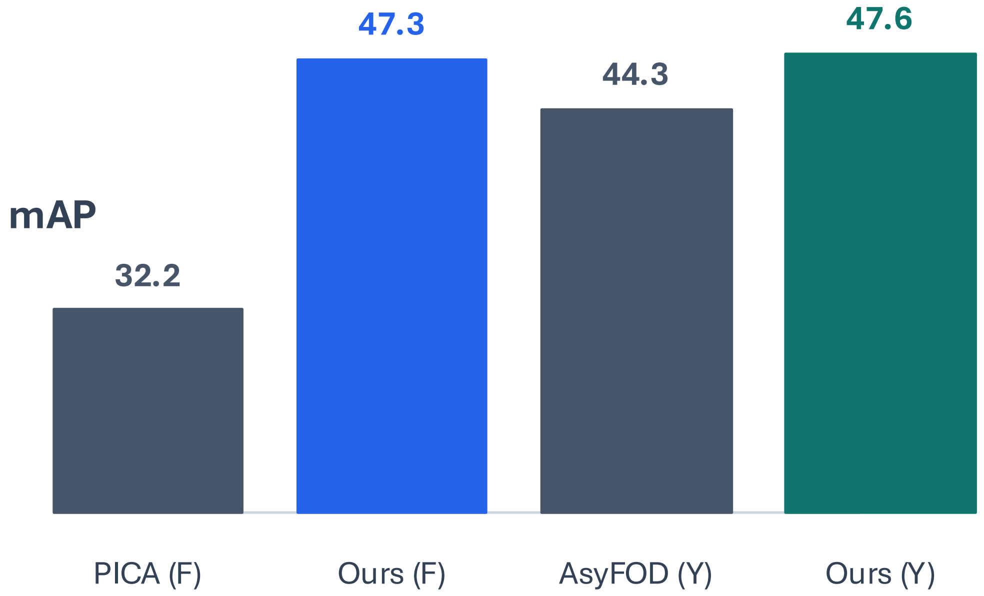
Generating target-style images with explicit novel-class prompts turns unseen classes from "unknown" objects into learnable categories.

W. Li et al. Novel scenes & classes: Towards adaptive open-set object detection. ICCV, 2023.

Main results: FSDAOD setting

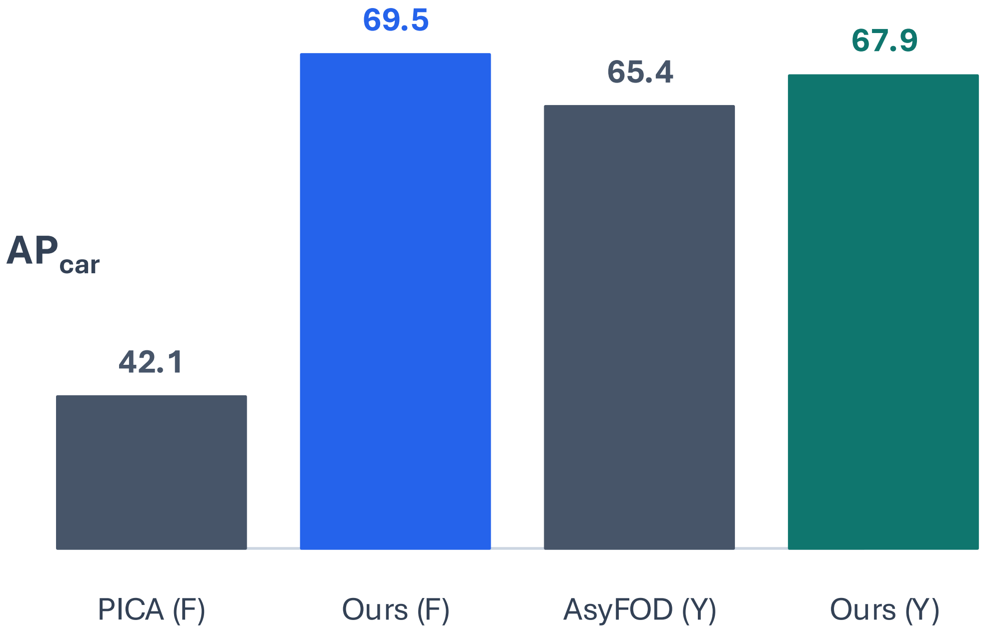
Our data-centric approach is also effective in the conventional closed-set few-shot setting.

Cityscapes → Foggy Cityscapes



SIM10k → Cityscapes

F: Faster R-CNN
Y: YOLOv5



The consistent gains across detector architecture highlight the data-centric nature of our approach.

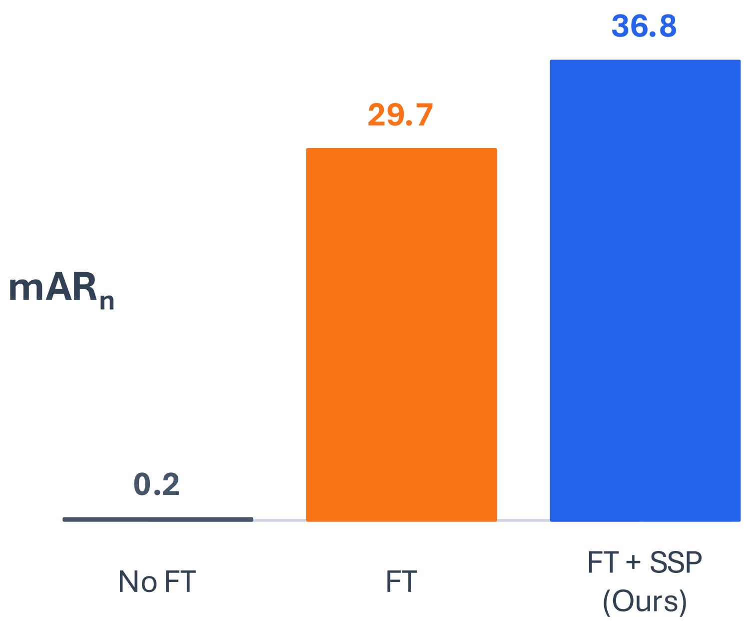
C. Zhong et al. Pica: Point-wise instance and centroid alignment based few-shot domain adaptive object detection with loose annotations. WACV, 2022.
Y. Gao et al. Asyfod: An asymmetric adaptation paradigm for few-shot domain adaptive object detection. CVPR, 2023.

Panasonic

Ablation studies

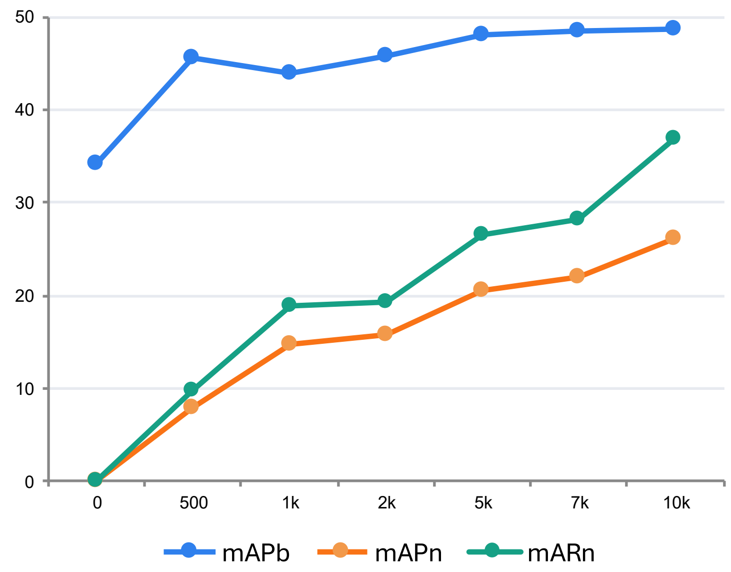
We next examine which factors in personalized data generation drive the performance gains.

1) Diffusion fine-tuning strategy



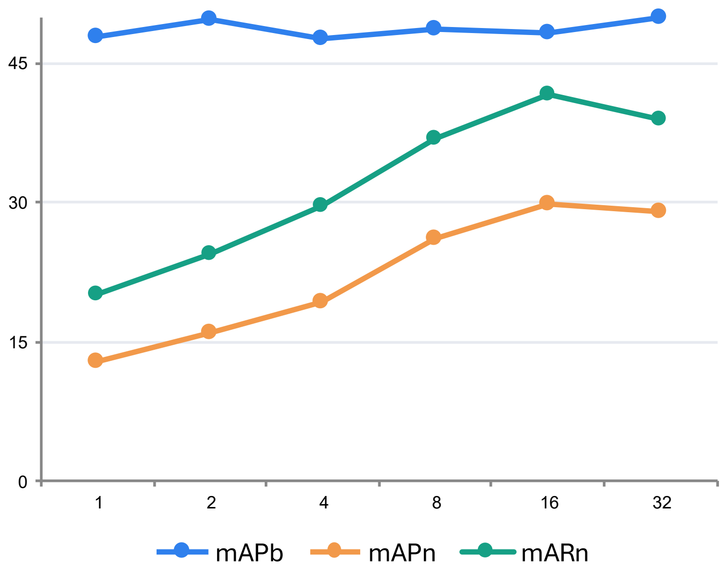
Fine-tuning is essential; SSP further improves diversity and prompt fidelity.

2) Number of synthetic images



Novel-class performance keeps increasing up to 10k synthetic images.

3) Number of target shots



More target shots generally improve novel-class performance.

Effective target-domain personalization, the SSP loss, and sufficient synthetic data are all important for learning novel classes.

Panasonic

Qualitative results

The personalized diffusion model captures target-domain appearance from only a few real images.

The generated images reproduce target-specific characteristics and provide diverse target-style data.

(a) Foggy Cityscapes

(b) BDD100k

(c) Cityscapes (Car only)

Panasonic

Conclusion

Takeaways

1. **New setting**

   FSAOOD combines few-shot target data with novel target classes.

2. **Personalized generation**

   Diffusion models can create target-style data with base and novel classes.

3. **Pseudo-label refinement**

   Open-vocabulary labeling plus online refinement enables detector training.

Future work

Reduce generation cost, hyperparameter tuning, and pseudo-label post-processing.

Merci beaucoup.