# Detect, Augment, Compose, and Adapt: Four Steps for Unsupervised Domain Adaptation in Object Detection

BMVC 2023

Mohamed Lamine Mekhalfi, Davide Boscaini, Fabio Poiesi
Fondazione Bruno Kessler {mmekhalfi, dboscaini, poiesi}@fbk.eu
github.com/MohamedTEV/DACA

CENTER FOR DIGITAL INDUSTRY — FONDAZIONE BRUNO KESSLER — TeV TECHNOLOGIES OF VISION

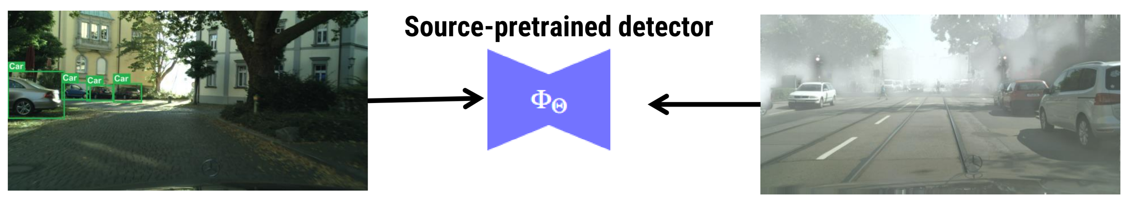
## Motivation

Adapting a source-trained detector to unlabelled target domain

**What is unsupervised domain adaptation in object detection (UDA)?** Adapt a detector by leveraging source data (images & annotations from domain A) and target data (only images, from domain B).

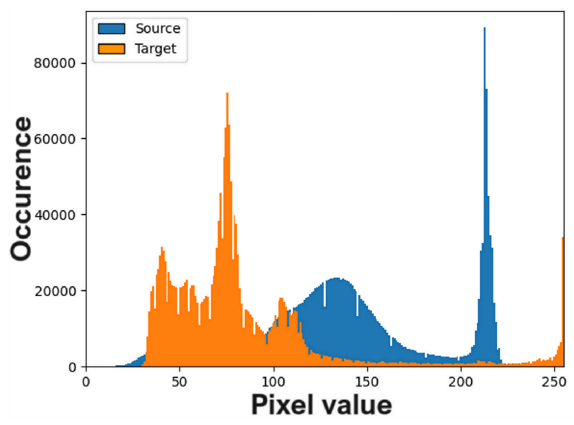
**Offline training procedure:** Pretrain the model on source data


**Online training procedure:** Unsupervised domain adaptation with source (images&groundtruth) and target (only images)



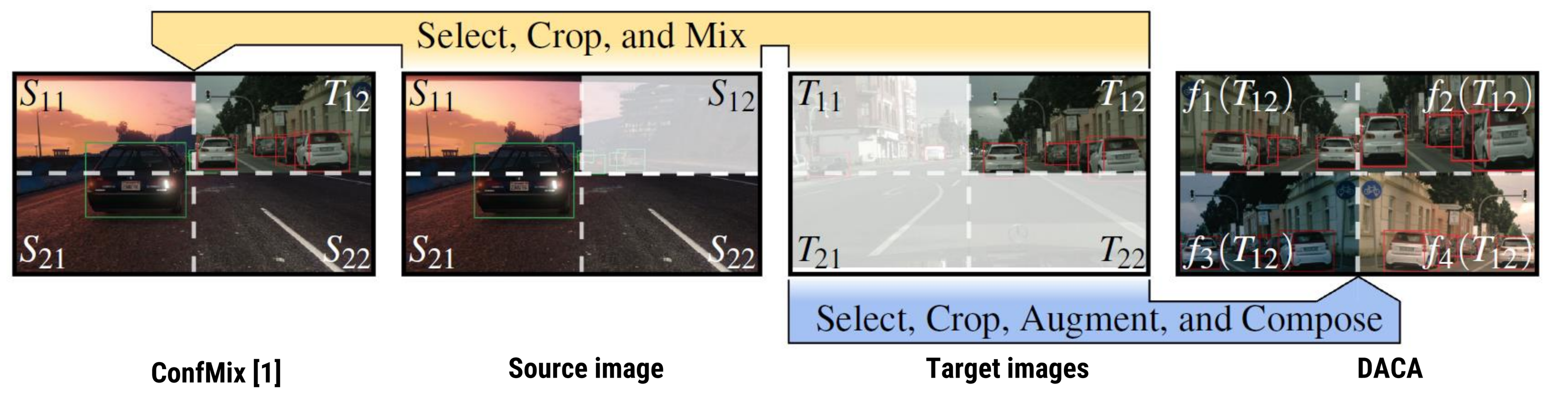

**Challenges of UDA in object detection:**

- Distribution mismatch across domains.
- Error accumulation (i.e., false positives as pseudo-labels) during self-training.
- No target annotations.
- Calibration: model detection thresholds across domains may differ due to domain gap.




**DACA vs ConfMix [1]:**

**ConfMix:** Mixes source and target images **VS** **DACA:** Does not mix but composes target images.




**Contributions:**

- DACA is the first alternative to mix up approaches that does not mix images from different domains, but instead generates difficult and informative composite images only from the unsupervised target images.
- DACA generates the composite image based on augmented versions of the target image region with the most confident detections, making the adaptation more effective.

## Our Approach

Self-supervision with challenging composite images

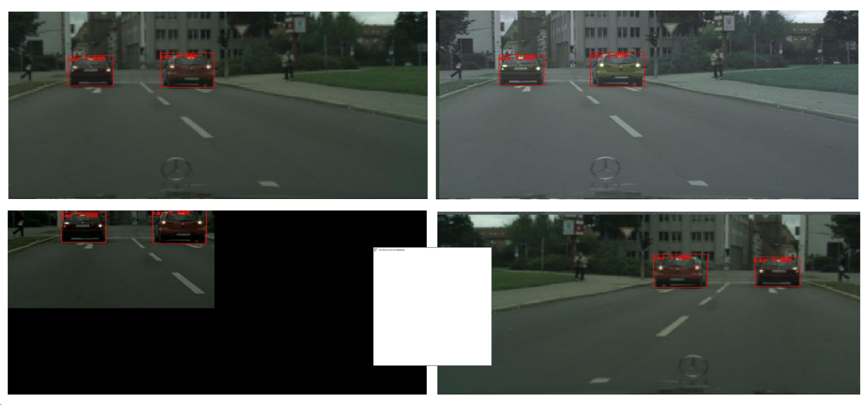
- **What is DACA ? An UDA approach that composes target images via random augmentations (only during training phase) and leverages self-training to adapt the model to the target domain.**
- **Why are the images challenging ? Because they stem from random augmentations, yet they present new-to-learn knowledge for the detector.**

### Step1: Detect

Draw pseudo-detections from the target image

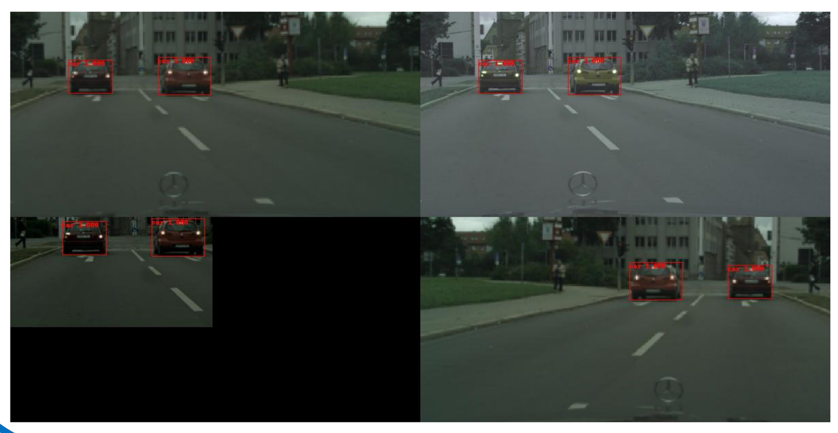
### Step 2: Augment

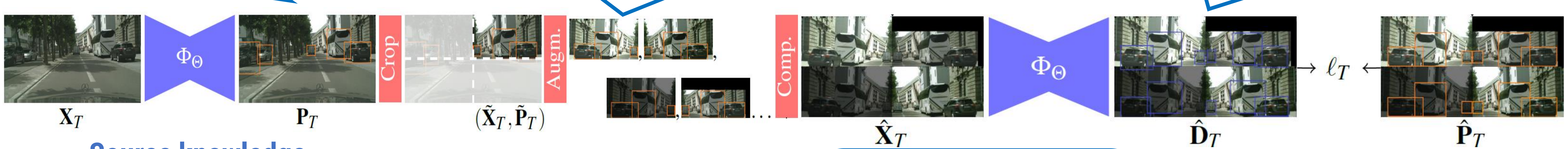
Random augmentations of the most confident target region (i.e., average confidence of all the detections ) & its detections.



### Step 3: Compose

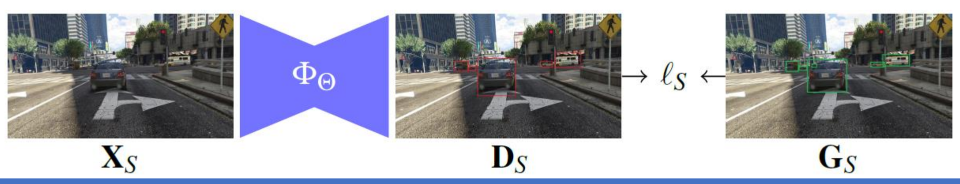
Combine the generated augmentations into a composite image






### Source knowledge

Maintain source supervision to prevent catastrophic forgetting [2]
Source knowledge is maintained during adaptation via consistency loss w.r.t source groundtruth.




Source knowledge → **Step 4: Adapt** ← Target knowledge

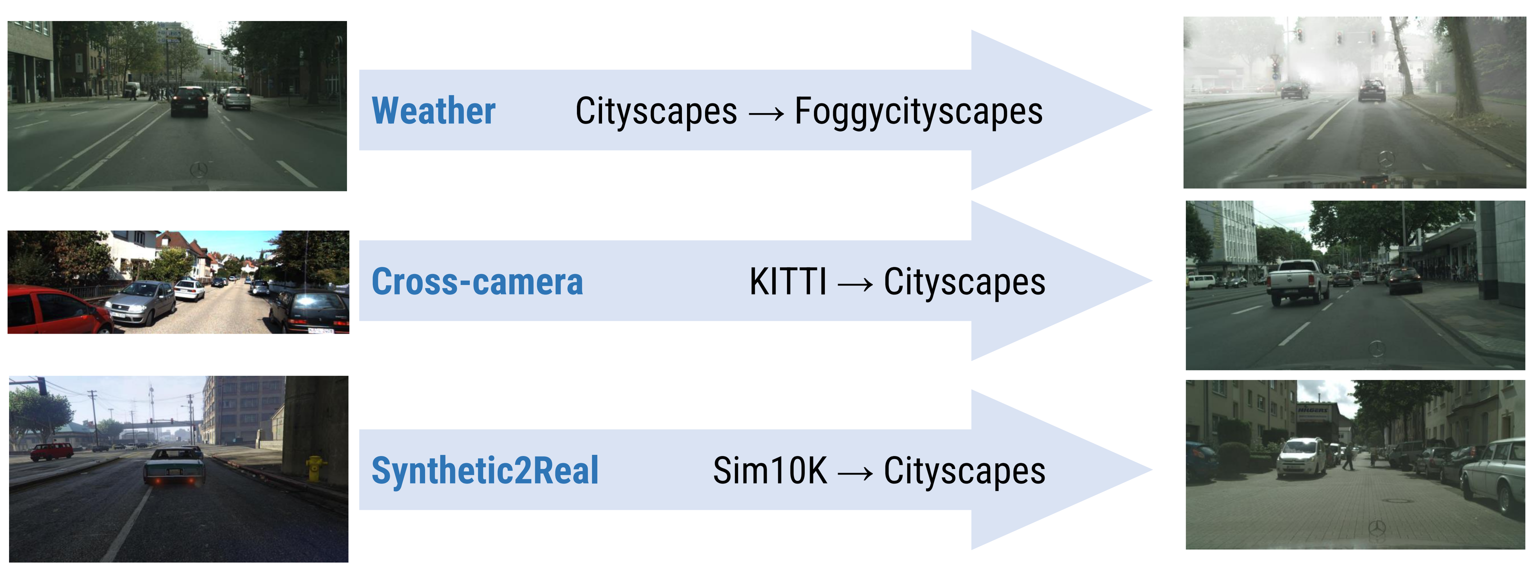
Backpropagate total loss. Source loss maintains source knowledge whilst target loss increments knowledge towards the target

$$\ell = \ell_S + \ell_T$$

## Results

DACA is superior to SOTA in two adaptation scenarios

**Adaptation scenarios & datasets:**




**Quantitative results:**

**Detection performance (AP) for the Car class.**

| Method | Detector | Backbone | S→C | K→C | C→F | Average |
|---|---|---|---|---|---|---|
| Source only | YOLOv5 | Darknet53 | 50.4 | 42.9 | 54.9 | 49.4 |
| Target only | YOLOv5 | Darknet53 | 69.5 | 69.5 | 67.9 | 69.0 |
| ConfMix [1] | YOLOv5 | Darknet53 | 56.2 | 51.6 | 63.0 | 56.9 |
| DACA (ours) | YOLOv5 | Darknet53 | **60.6** | **54.2** | 63.0 | **59.3** |

**Detection performance (AP) for the C→F adaptation benchmark.**

| Method | Detector | Backbone | Person | Rider | Car | Truck | Bus | Train | Motorcycle | Bicycle | mAP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Source only | YOLOv5 | Darknet53 | 39.2 | 38.0 | 54.9 | 12.4 | 33.1 | 06.2 | 19.9 | 33.6 | 29.7 |
| Target only | YOLOv5 | Darknet53 | 45.6 | 43.0 | 67.9 | 30.2 | 48.0 | 39.4 | 30.3 | 37.5 | 42.7 |
| ConfMix [1] | YOLOv5 | Darknet53 | **44.0** | **43.3** | 63.0 | **30.1** | **43.0** | 29.6 | 25.5 | **34.4** | 39.1 |
| DACA (ours) | YOLOv5 | Darknet53 | 41.9 | 40.8 | 63.0 | 29.4 | 42.2 | **37.2** | **27.8** | 33.0 | **39.4** |

**Ablations:**

**List of augmentations**

| Acronym | Transformation |
|---|---|
| HF | HorizontalFlip |
| RC | BBoxSafeRandomCrop |
| B | Blur |
| CJ | ColorJitter |
| D | Downscale |
| BC | RandomBrightnessContrast |

**Effect of the number of augmented regions**

| #regions | C→F | K→C | S→C | Avg. |
|---|---|---|---|---|
| 1 | 35.4 | 51.5 | 56.6 | 47.8 |
| 2 | 38.3 | 52.2 | 58.5 | 49.7 |
| 3 | 39.1 | 53.1 | 60.2 | 50.8 |
| 4 | **39.4** | **54.2** | **60.6** | **51.4** |

**Effect of transformations**

| Trans. | C→F | K→C | S→C | Avg. |
|---|---|---|---|---|
| None | 33.5 | 52.8 | 57.4 | 47.9 |
| HF | 38.0 | 52.9 | 59.7 | 50.2 |
| RC | 34.3 | 52.2 | 59.4 | 48.7 |
| B | 35.9 | 53.2 | 58.4 | 49.2 |
| CJ | 34.6 | 52.5 | 58.2 | 48.5 |
| D | 35.3 | 54.1 | 59.8 | 49.8 |
| BC | 33.9 | 52.6 | 57.5 | 48.0 |
| HF+B | 38.9 | 53.9 | 59.3 | 50.7 |
| HF+D | 35.4 | 53.3 | 56.7 | 48.5 |
| D+B | 36.8 | 53.5 | 59.9 | 50.0 |
| HF+D+B | 37.8 | 54.0 | 57.1 | 49.6 |
| All | **39.4** | **54.2** | **60.6** | **51.4** |

**Effect of grid layout**

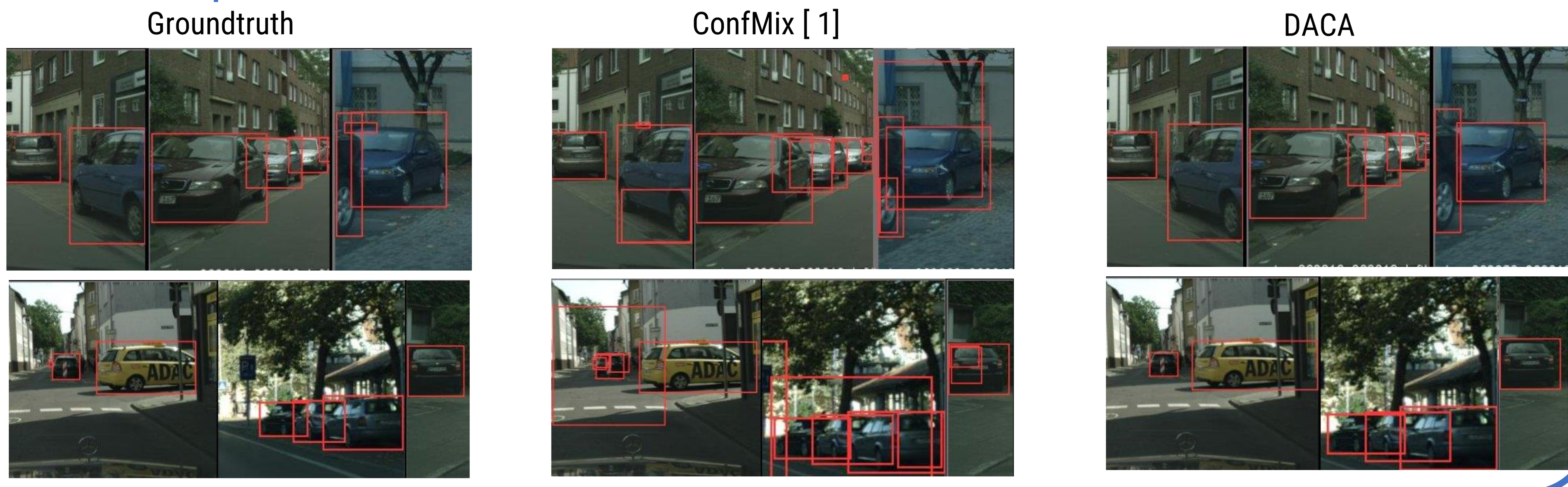
| Layout | C→F | K→C | S→C | Avg. |
|---|---|---|---|---|
| 3x3 | 37.8 | 51.2 | 57.7 | 48.9 |
| 2x3 | 38.5 | 51.7 | 58.7 | 49.6 |
| 3x2 | 38.6 | 53.6 | 59.9 | 50.7 |
| 2x2 | **39.4** | **54.2** | **60.6** | **51.4** |

**Conclusions**

- Augmentation is a efficient way to produce challenging target images to perform UDA via self-training.
- To address the problem of false positive accumulation, style-transfer techniques [3] can be applied to lessen style shift.

**Qualitative examples:**

Groundtruth — ConfMix [ 1] — DACA



**References:**
[1] G. Mattolin, L. Zanella, E. Ricci, and Y. Wang. ConfMix: Unsupervised Domain Adaptation for Object Detection via Confidence-based Mixing. In WACV, 2023.
[2] R. Kemker, M. McClure, A. Abitino, T. Hayes, and C. Kanan. Measuring catastrophic forgetting in neural networks. In AAAI, 2018.
[3] Y. Yang and S. Soatto. FDA: Fourier Domain Adaptation for Semantic Segmentation. In CVPR, 2020.

**Acknowledgement:**
We are very grateful to the support by European Union's Horizon Europe research and innovation programme under grant agreement No. 101092043, project AGILEHAND (Smart Grading, Handling and Packaging Solutions for Soft and Deformable Products in Agile and Reconfigurable Lines).