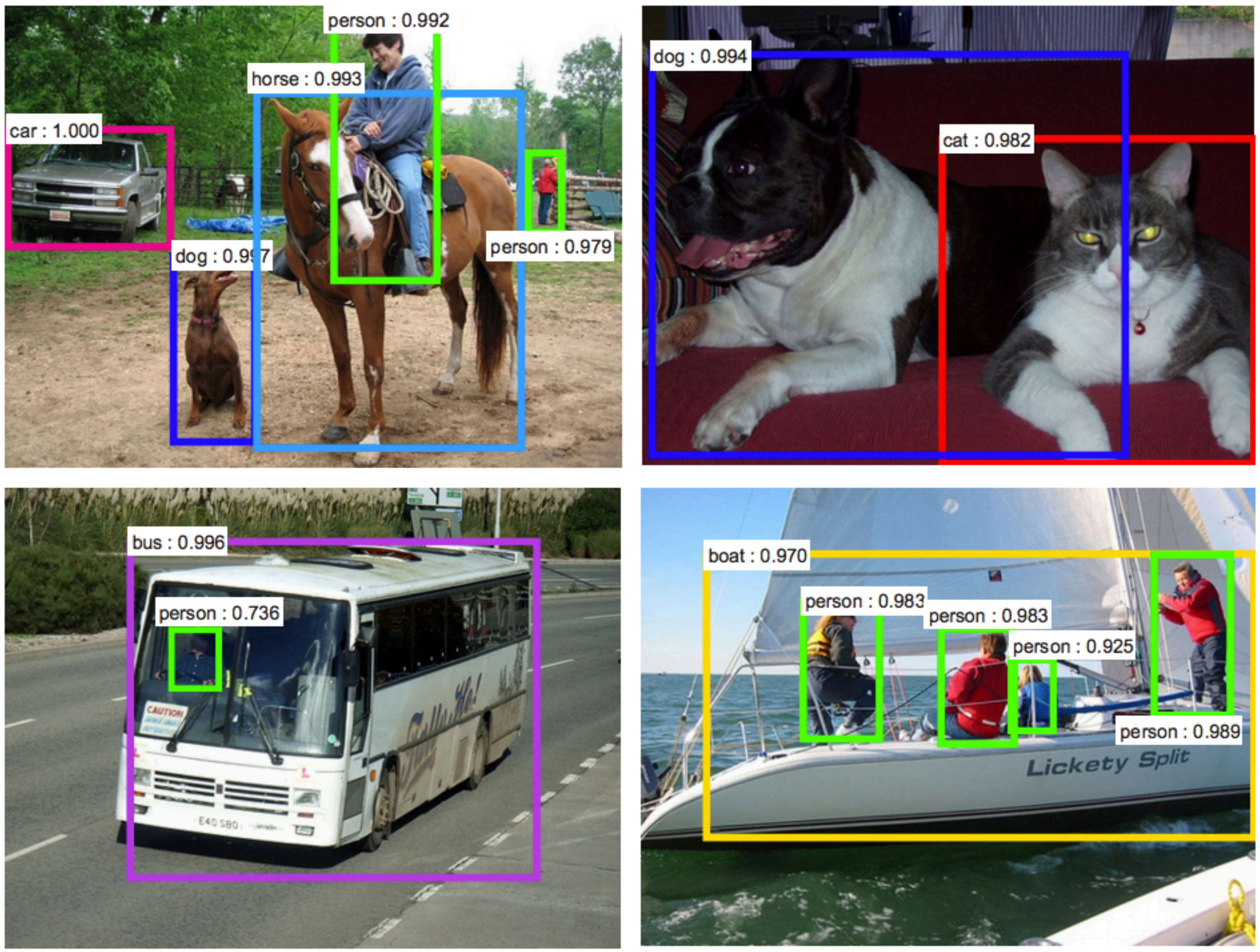
Leveraging Synthetic Data for Mobile Object Detection

Vidit
PhD Candidate
CVLab, EPFL

EPFL

EPFL

when labeled data is available



Object Detectors: <u>Faster RCNN</u> <u>Yolo</u>

when labeled data is available

Object Detectors: Faster RCNN Yolo

Datasets	Examples/class	mAP (Faster RCNN)
MS-COCO 80 classes	10k	42.7
Pascal-VOC 20 classes	500	78.8

when labeled data is available

Object Detectors: Faster RCNN Yolo

Datasets	Examples/class	mAP (Faster RCNN)
MS-COCO 80 classes	10k	42.7
Pascal-VOC 20 classes	500	78.8

Cons:

1. **Annotation Cost**
2. **Hard to find annotations for eg: Medical data**

when few labeled data is available

Object Detectors: Faster RCNN Yolo

Datasets	Examples/ class	mAP (Faster RCNN)	mAP*** (Faster RCNN) ~10 examples/ class
MS-COCO 80 classes	10k	42.7	**4.9**
Pascal-VOC 20 classes	500	78.8	**29**

***Meta R-CNN : Towards General Solver for Instance-level Few-shot Learning CVPR, 2020

when no-real labeled data is available

Leveraging BIM - the digital twin

Synthetic Images

when no-real labeled data is available

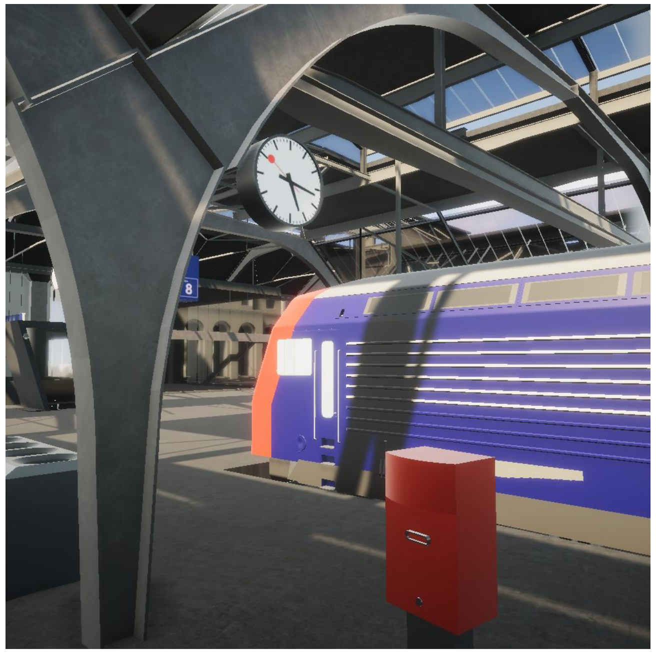
Leveraging BIM - the digital twin

Synthetic Images with day/night scenes, different angles/distances

when no-real labeled data is available

- Domain Adaptation
- Pseudo-Labels

Domain Adaptation

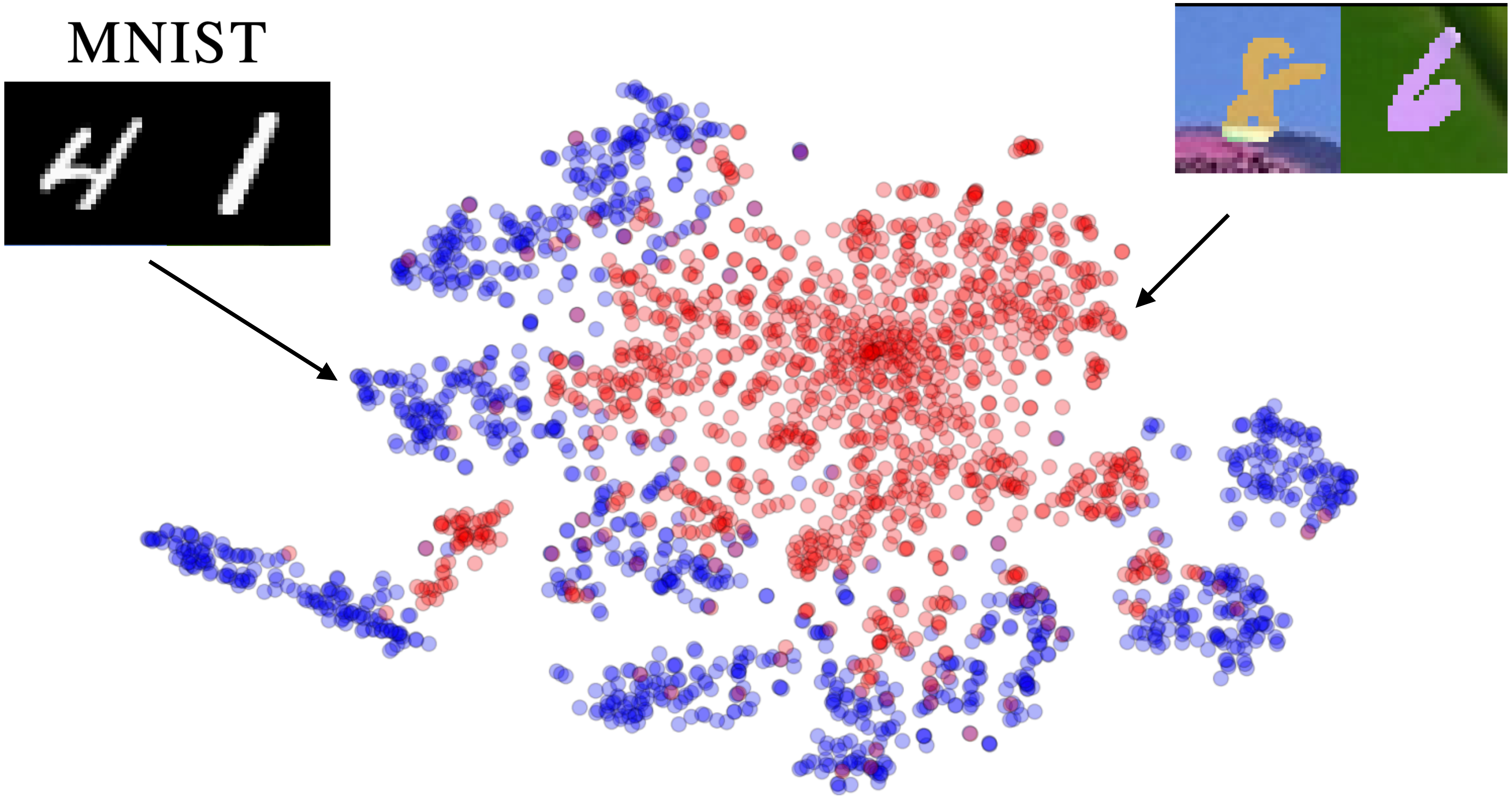


Source Domain
Labels available

Target Domain
No labels

Domain Adaptation

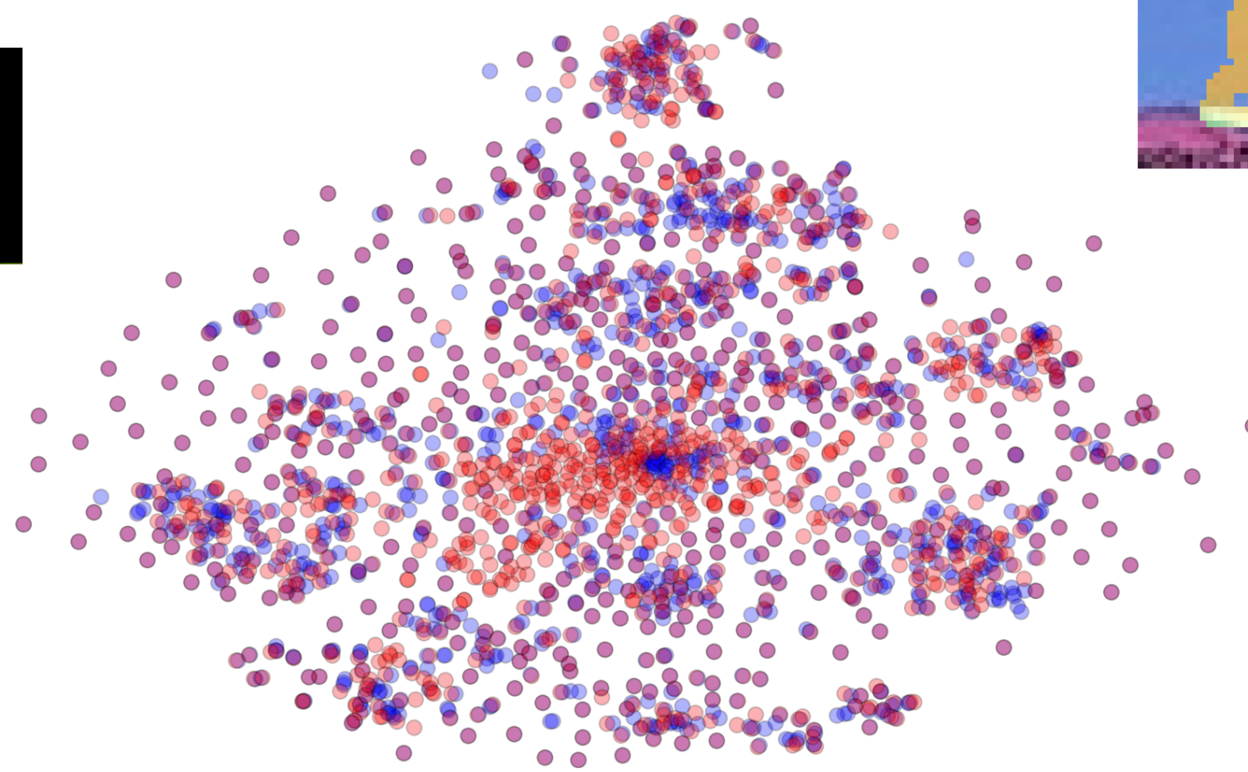
$P_S(x) \neq P_T(x)$



Unsupervised Domain Adaptation by Backpropagation, ICML 2015

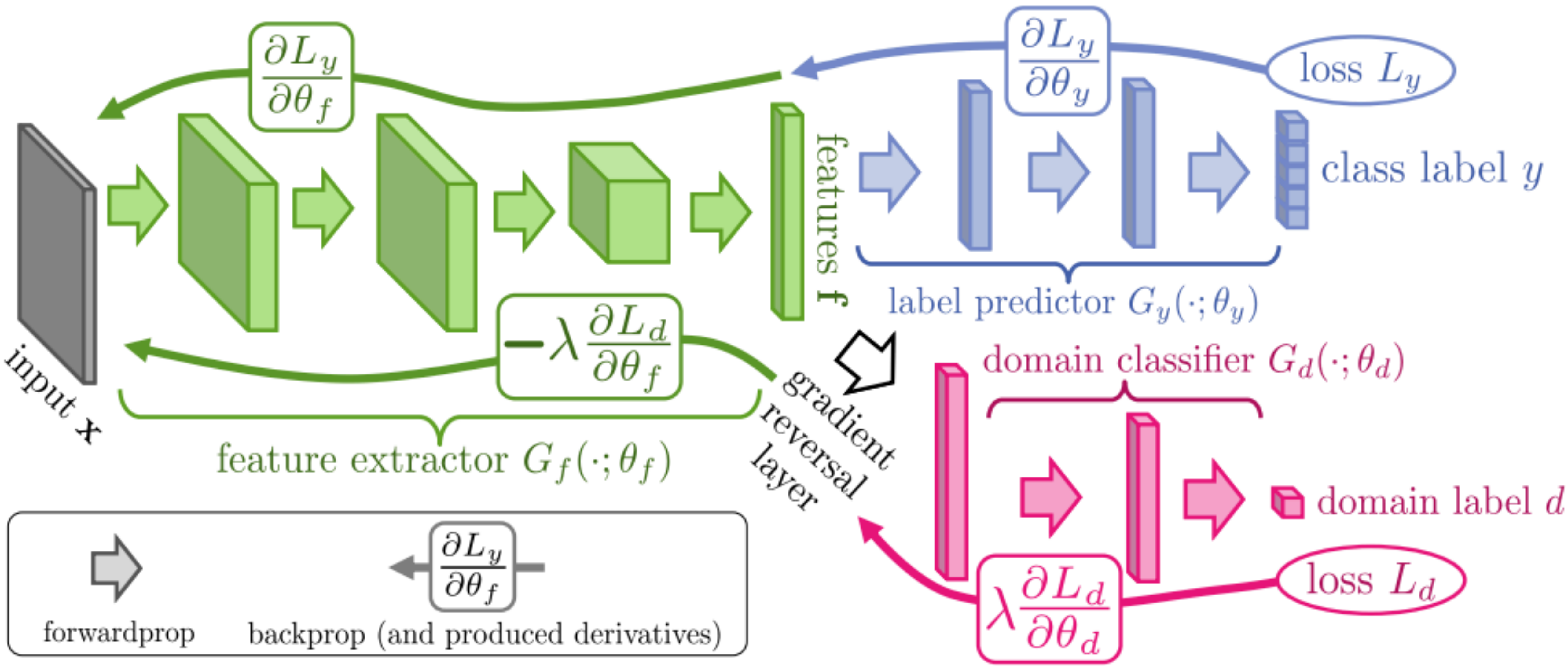
Domain Adaptation

MNIST



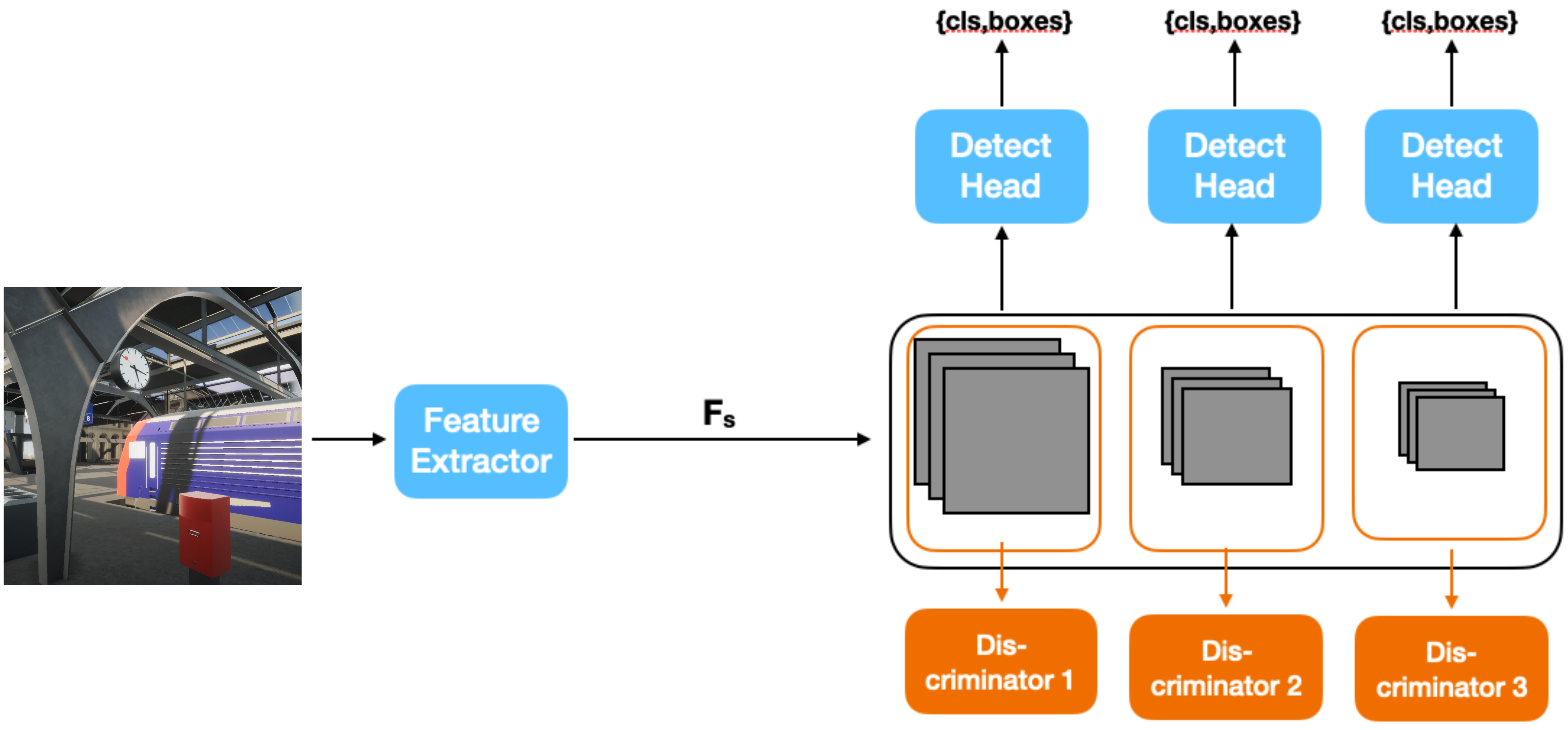
Unsupervised Domain Adaptation by Backpropagation, ICML 2015

Domain Adaptation



Unsupervised Domain Adaptation by Backpropagation, ICML 2015

Domain Adaptation



SSD object detector

Domain Adaptation

Class	mAP(%) mobilenetV3+ Synthetic image only	mAP(%) mobilenetV3+DA + 500 unlabelled images	Gain
Recycling Station	71.4	72.8	1.4
Billettentwerter	74.2	78.2	4
Uhr	66.7	63.9	-2.8
LCD	29.9	46.1	16.2
Automatic Ticket Machine	79.6	79.1	-0.5

~1500 synthetic images ~160 real test images

Domain Adaptation

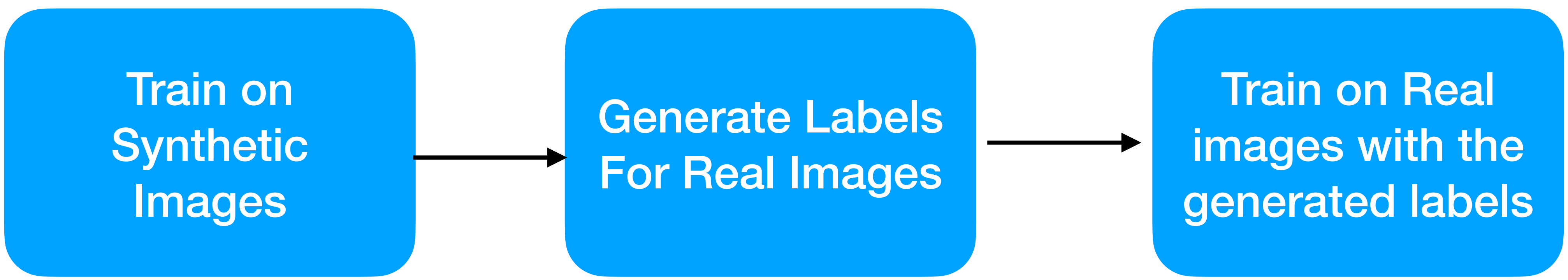
Class	mAP(%) mobilenetV3 With labelled Real images	mAP(%) mobilenetV3+trained on Synthetic Images	mAP(%) mobilenetV3+DA + 1700 unlabelled images	Gain
Recycling Station	72.3	37.8	41.2	3.4
Billettentwerter	86.9	84	75.9	-8.1
Uhr	55.6	41	43.5	2.5

~1500 synthetic images ~600 real test images

Domain Adaptation

Issues

Pseudo-Labels



Keep only the predictions with 50% or more confidence

Pseudo-Labels

Class	mAP(%) mobilenetV3 With labelled Real images	mAP(%) mobilenetV3+trained on Synthetic Images	mAP(%) mobilenetV3+DA + 1700 unlabelled images	mAP(%) mobilenetV3+ Pseudo Labels 800 images	Gain
Recycling Station	72.3	37.8	41.2	44.9	7.1
Billettentwerter	86.9	84	75.9	83.1	-3.8
Uhr	55.6	41	43.5	45.1	4.1

~1500 synthetic images ~600 real test images

YOLOv5 with default augmentations

Classes	Train only on Real Images	Train on Synthetic	Train on Synthetic+ Pseudo Labels 800 images	Gain
RecylingStation	88.1	52.3	57.5	5.2
Billetenwerter	98.9	84.9	94.7	9.8
Uhr	89.6	85.3	85.9	0.6

Train on Synthetic

Train on Synthetic+ Pseudo Labels

Challenges

- Domain Adaptation:
  - + no real-labels needed
  - + Relies on synthetic images reflecting on the use case scenarios
  - - Need for diversity in the synthetic and real data

Challenges

- Pseudo - Labels
  - + Relies on synthetic images reflecting on the use case scenarios
  - + Opportunity to quickly leverage on larger unlabelled data
  - - need to handle false positives and class imbalances

Challenges

- Domain Adaptation:
  - + no real-labels needed
  - + Relies on synthetic images reflecting on the use case scenarios
  - - Need for diversity in the synthetic and real data
- Pseudo - Labels
  - + Relies on synthetic images reflecting on the use case scenarios
  - + Opportunity to quickly leverage on larger unlabelled data
  - - need to handle false positives and class imbalances

Thank You!