Domain Adaptation of Virtual and Real Worlds for Pedestrian Detection

Dr. Antonio M. López

www.cvc.uab.es/adas

www.cvc.uab.es/~antonio

www.cvc.uab.es/domainadaptation

September 19th, 2014

Workshop on

"Generalization and reuse of machine learning models over multiple contexts"

Outline ◄

1. Autonomous driving.
2. Pedestrian (Object) detection: self-training.

3. Virtual-world training.
4. Virtual and real world adaptation: cool-world.

5. Virtual-world DPM (VDPM).
6. Domain adaptation of VDPM.

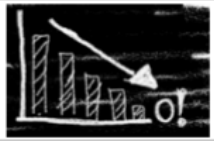
7. Random Forest of Local Experts: Adaptation.

8. Conclusions.

Outline ◄

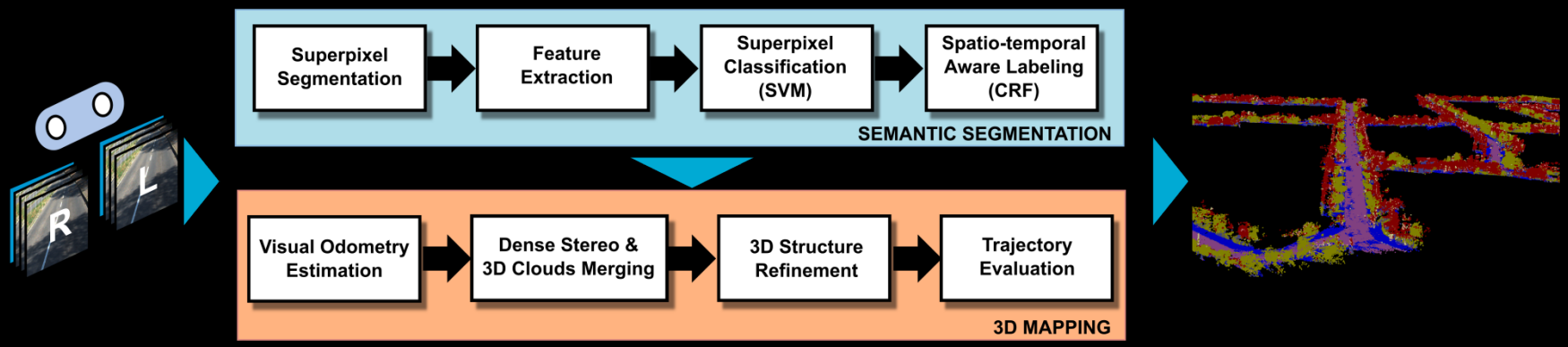
1. Autonomous driving.
2. Pedestrian (Object) detection: self-training.
3. Virtual-world training.
4. Virtual and real world adaptation: cool-world.
5. Virtual-world DPM (VDPM).
6. Domain adaptation of VDPM.
7. Random Forest of Local Experts: Adaptation.
8. Conclusions.

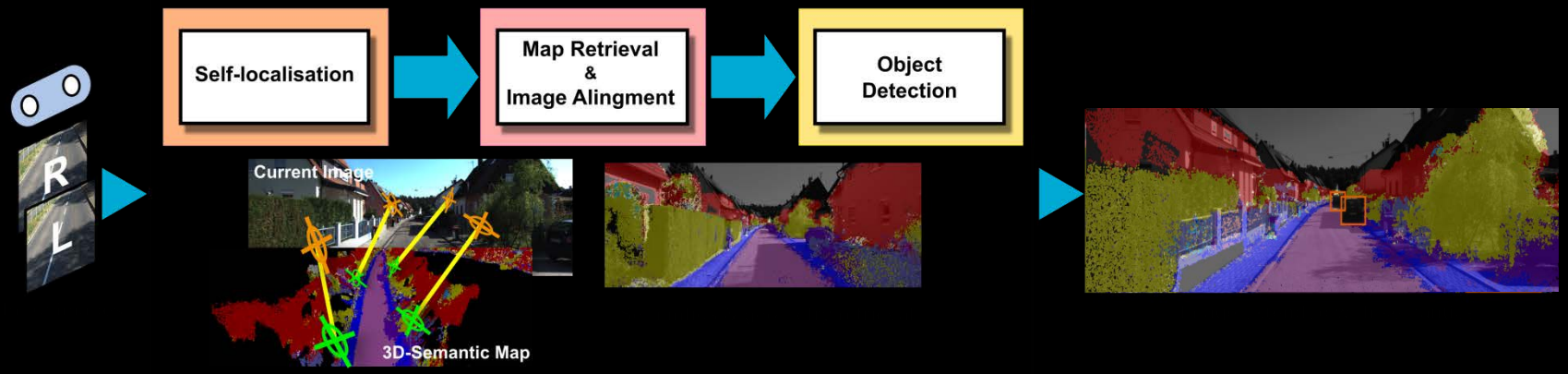
Autonomous Driving ◄

1	*Zero Emission*	- Optimization of traffic flow management - Reduction of fuel cons. and CO2 emission	
2	*Demographic change*	- Support unconfident drivers - Enhance mobility for elderly people	
3	*Vision Zero*	- Potential for more safety by avoidance of human driving errors	
4	*Increasing traffic density*	- Optimization of traffic flow management - Convenient, time efficient driving via automation	

* Dietmar Peters and Arne Bartels. Automated Driving - State of the art and future challenges. 7th Simposio Vehículo Inteligente, Barcelona, 15th May 2013

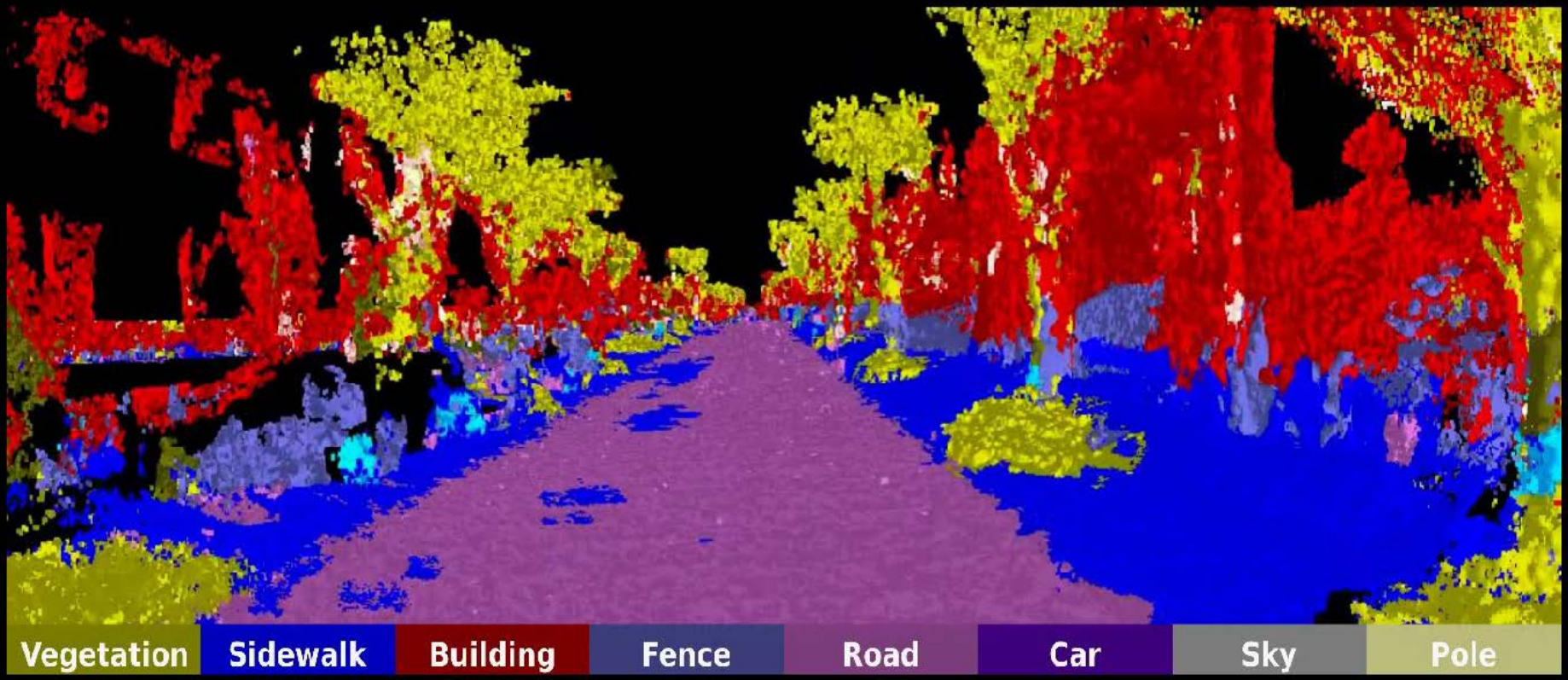
Autonomous Driving: Offline-Online Strategy ◄





Autonomous Driving: Offline-Online Strategy ◄

Offline: semantic 3D map generation



Autonomous Driving: Offline-Online Strategy ◄

Online: real-time 3D scene understanding

Focus of this Talk: Pedestrian Detection◀

Outline ◄

1. Autonomous driving.
2. Pedestrian (Object) detection: self-training.

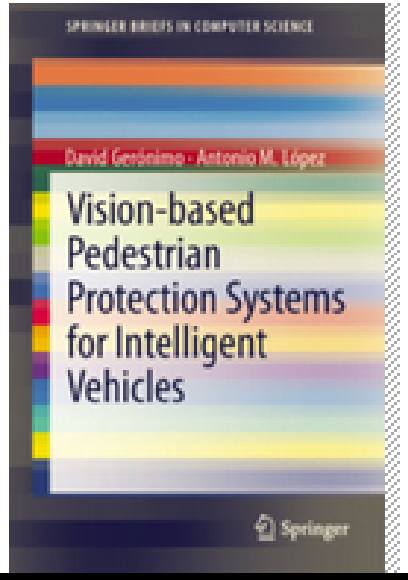
3. Virtual-world training.
4. Virtual and real world adaptation: cool-world.

5. Virtual-world DPM (VDPM).
6. Domain adaptation of VDPM.

7. Random Forest of Local Experts: Adaptation.

8. Conclusions.

Pedestrian Det.: Hot Topic of the Last Decade



Vision-based Pedestrian Protection Systems for Intelligent Vehicles

Series: » SpringerBriefs in Computer Science

Geronimo, David, Lopez, Antonio M.

2014, X, 114 p. 42 illus.

Free Preview

Available Formats:

EURO NCAP

www.euroncap.com

Pedestrian Detection: Demo◀

Pedestrian Detection: 2D Still Images ◀

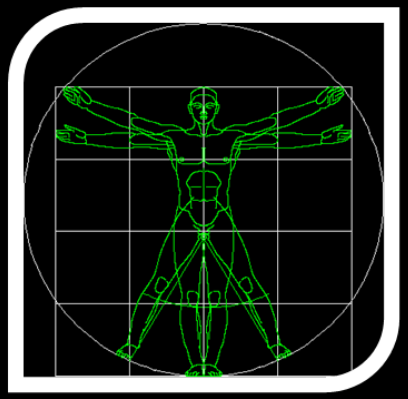
Training Data

Positive samples

Negative samples

Feature Extraction

Pedestrian Detection: 2D Still Images ◄

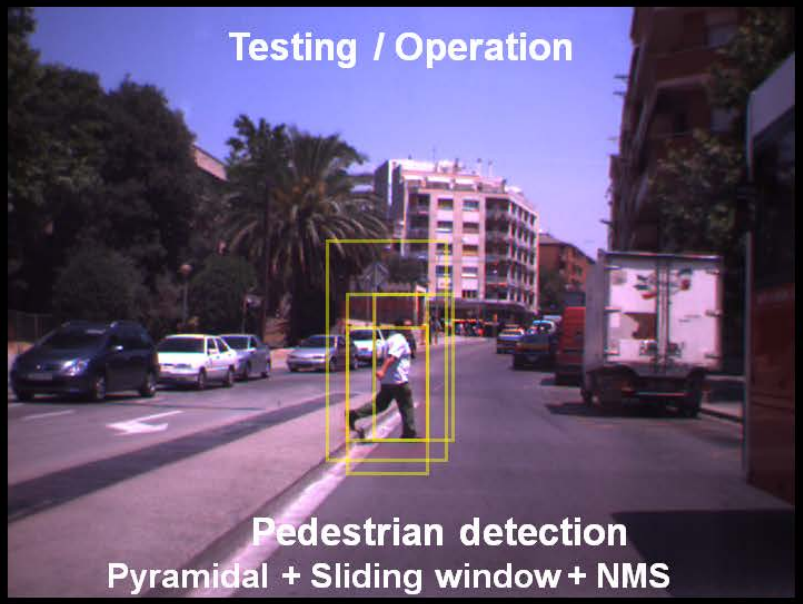


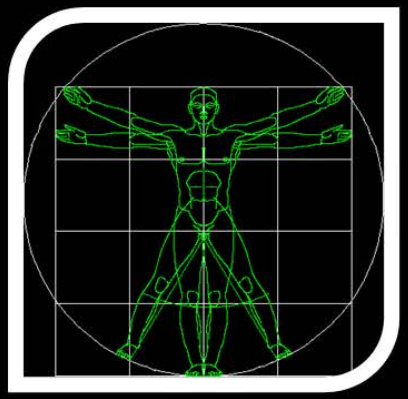
Classifier

Feature Extraction

Machine learning

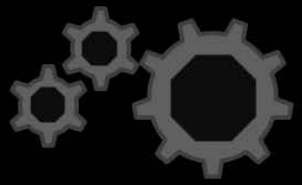
Pedestrian Detection: 2D Still Images ◄





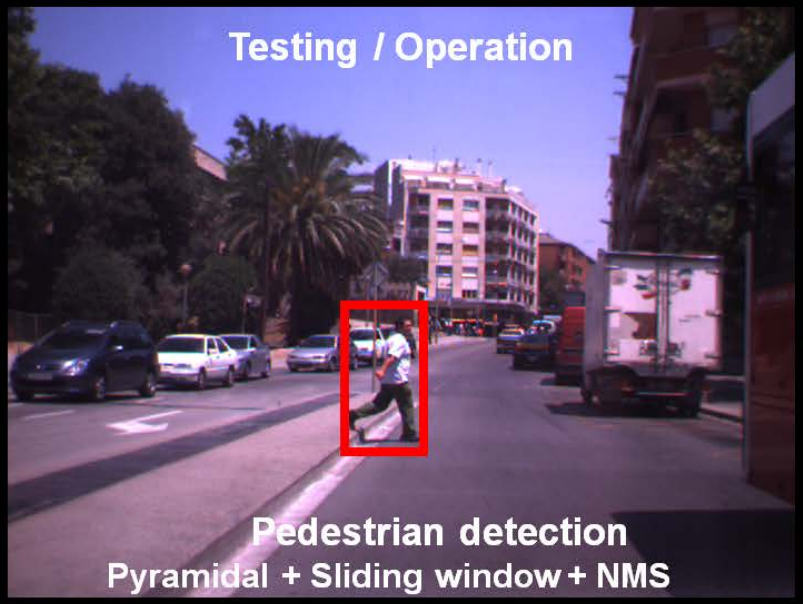
Classifier

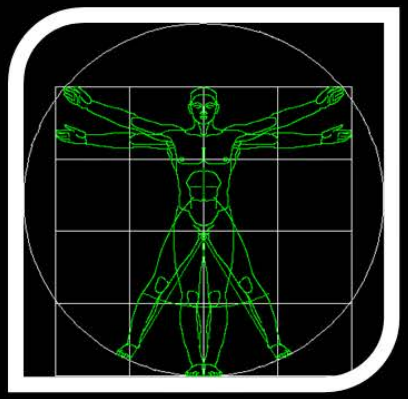
Feature Extraction



Machine learning

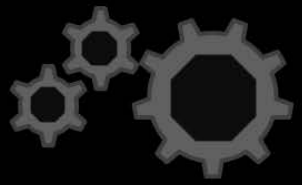
Pedestrian Detection: 2D Still Images ◄





Classifier

Feature Extraction



Machine learning

Pedestrian Detection: please, annotate first◀

Pedestrian Detection: please, annotate first◀

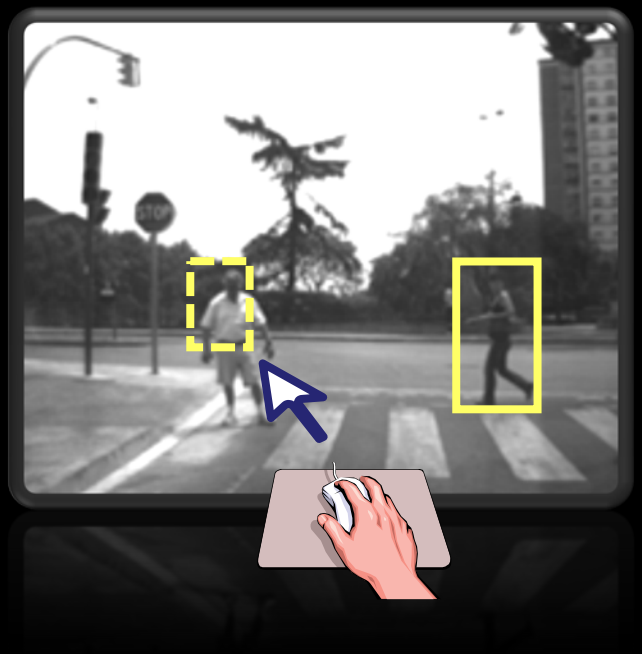
Examples **Counterexamples**

Daimler (real)

M. Enzweiler and D. Gavrila. **Monocular pedestrian detection: survey and experiments**. *TPAMI* 2009.

Different training and testing sets provided.

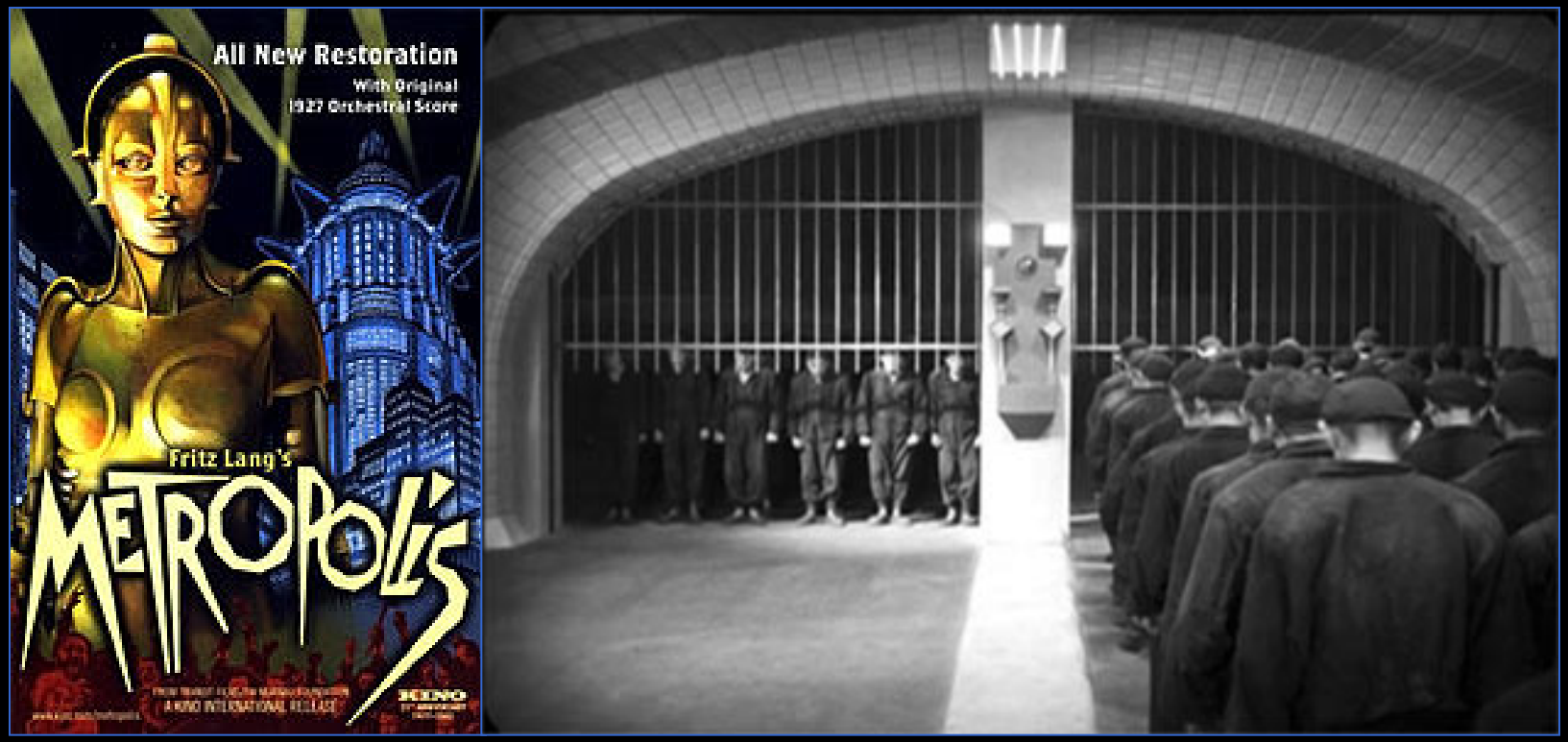
Manual annotation



Annotation (of examples) is a tiresome and subjective manual task, prone to errors.

E.g., Daimler ➔ A couple of millions.

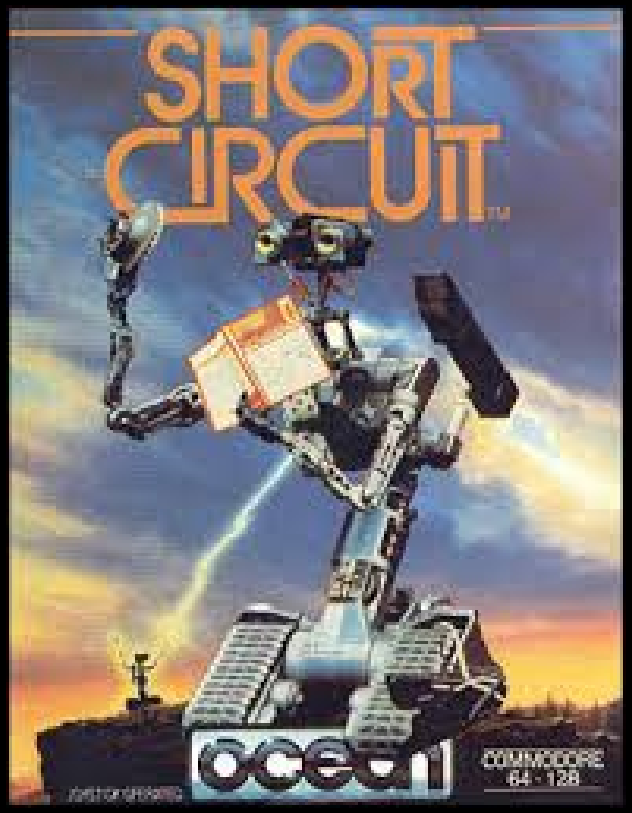
Current solution: web based annotation tools ◄



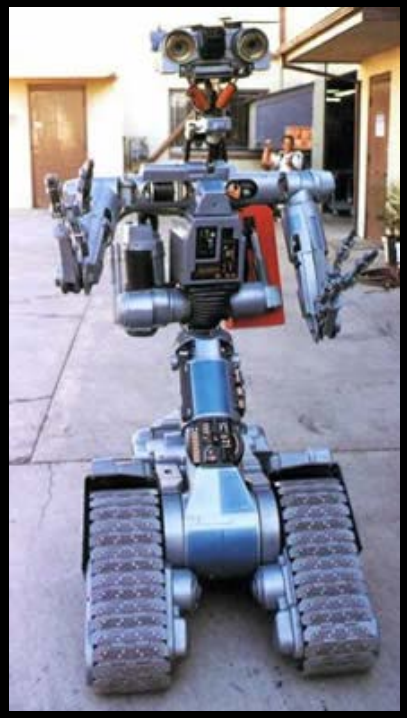
No free lunch:

- Deal with annotation errors (misunderstood task, malicious workers, tiresome activity).
- Devising: reward strategy, annotation guidelines, etc.

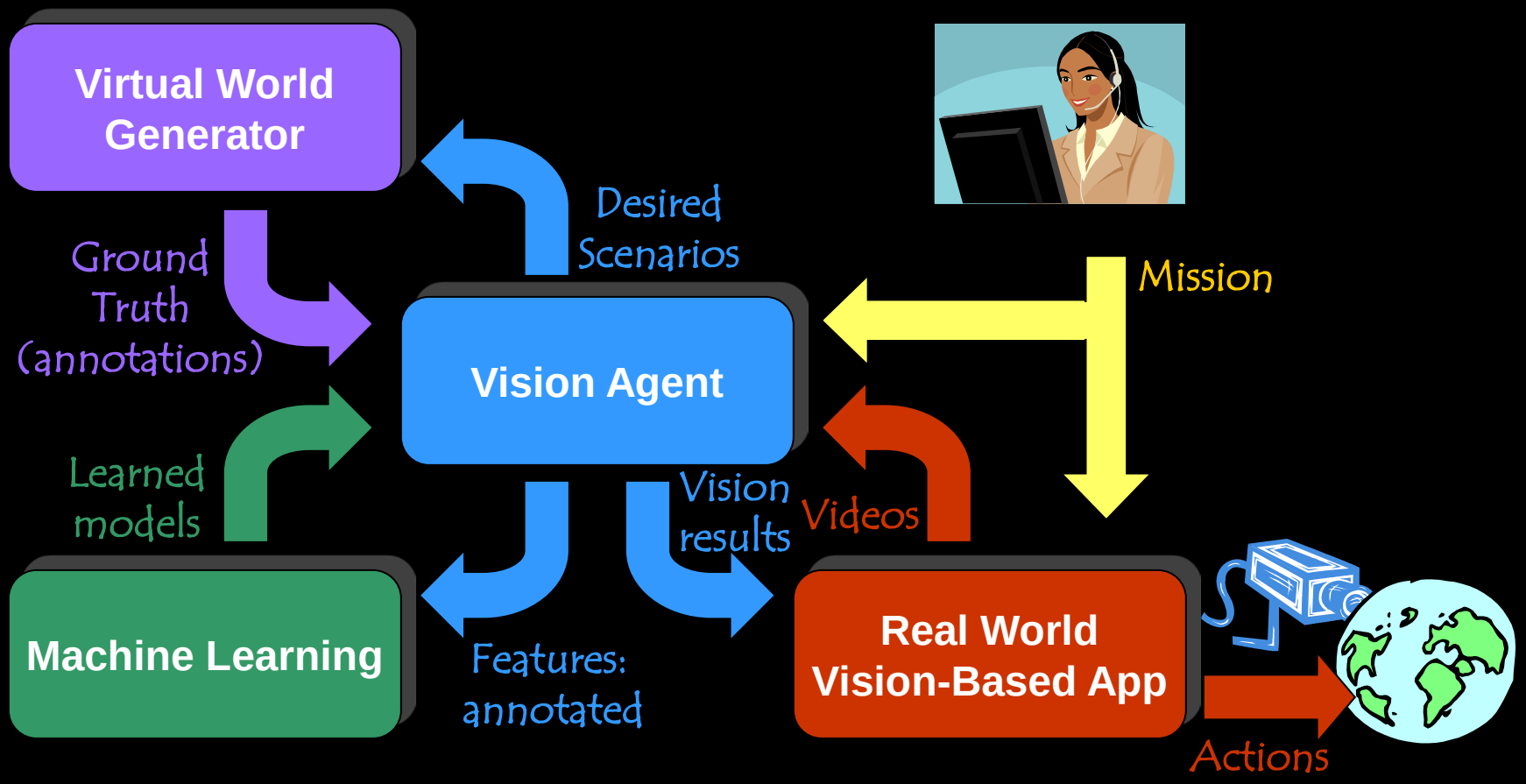
Pedestrian Detection: is self-training possible?◀



Realistic computer graphics◀



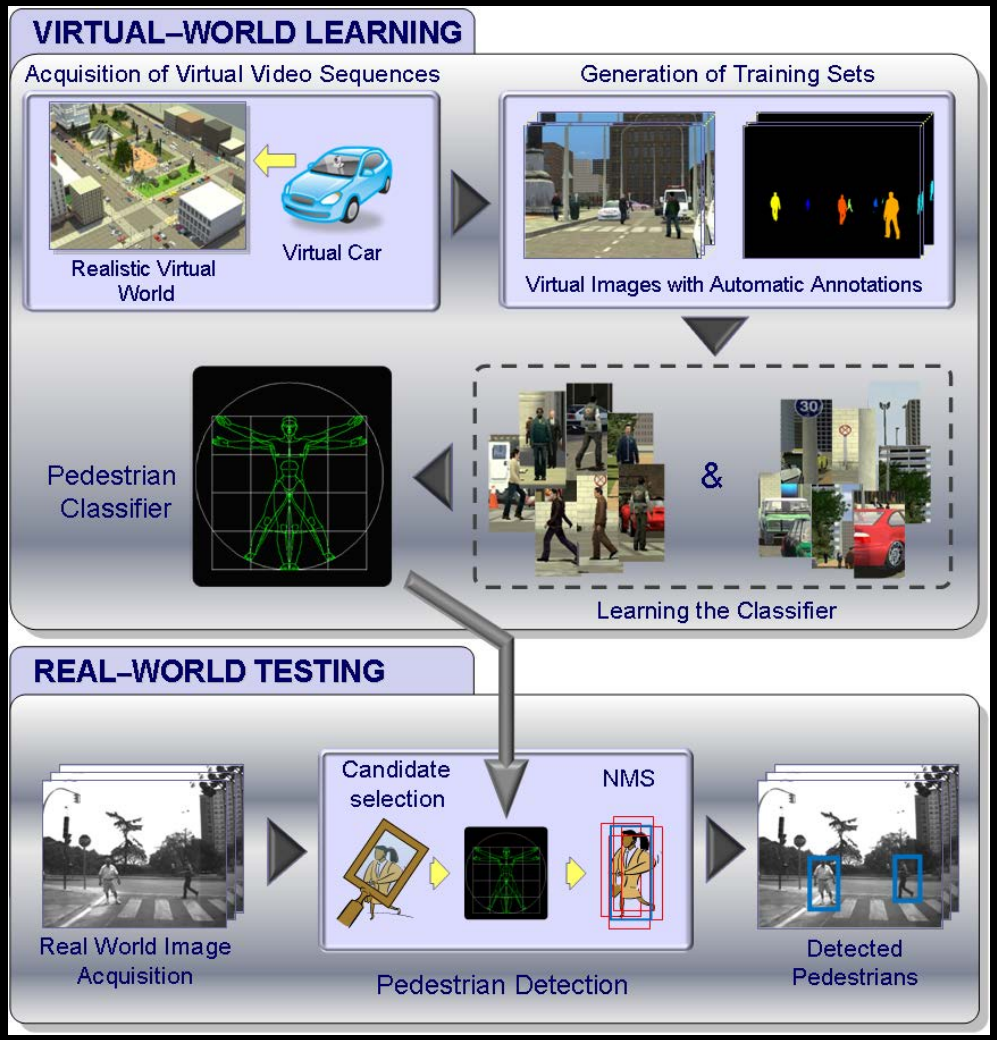
Self-training idea for Object Detection



Outline ◄

1. Autonomous driving.
2. Pedestrian (Object) detection: self-training.
3. Virtual-world training.
4. Virtual and real world adaptation: cool-world.
5. Virtual-world DPM (VDPM).
6. Domain adaptation of VDPM.
7. Random Forest of Local Experts: Adaptation.
8. Conclusions.

“Learning appearance in virtual world for operating in real one” Is it possible? ◄



“Close the circle between modern Computer Animation and Computer Vision”

Advantages:

- Gather automatically annotated samples.
- Such annotations are precise.

J. Marín, D. Váquez, D. Gerónimo, A.M. López. **Learning appearance in virtual scenarios for pedestrian detection.** CVPR 2010.

Virtual-world based on a videogame ◄

A kind of "Truman's Show" world, but virtual.

Virtual-world based on a videogame ◄

Examples

Counterexamples

Virtual

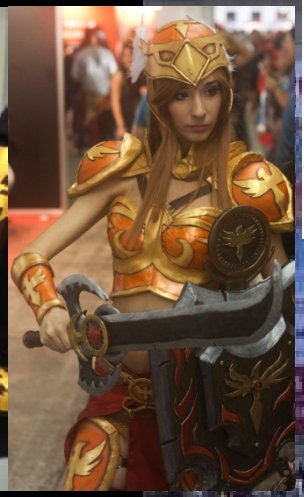
Virtual-world based on a videogame ◄

Examples

Counterexamples

Virtual

Virtual-world based on a videogame ◄

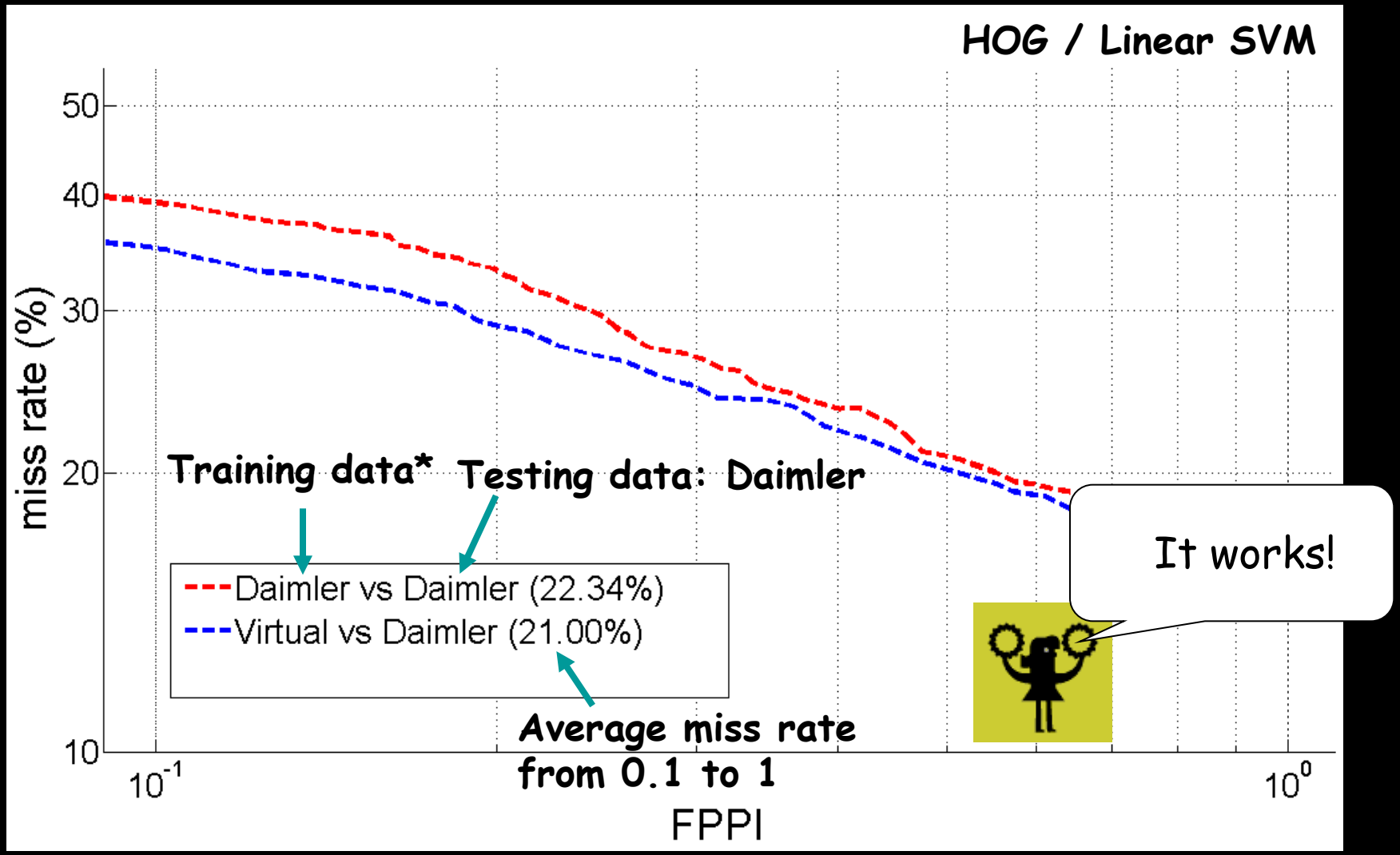


Virtual-world based on a videogame ◄

Virtual-world based on a videogame ◄

J. Marín, D. Gerónimo, D. Vázquez, A.M. López. **Pedestrian Detection: Exploring Virtual Worlds.** In Handbook of Pattern Recognition: Methods and Application.

Results: testing in Daimler real-world data ◄



(*) Training sets size as well as other issues have been carefully set for fair comparisons.

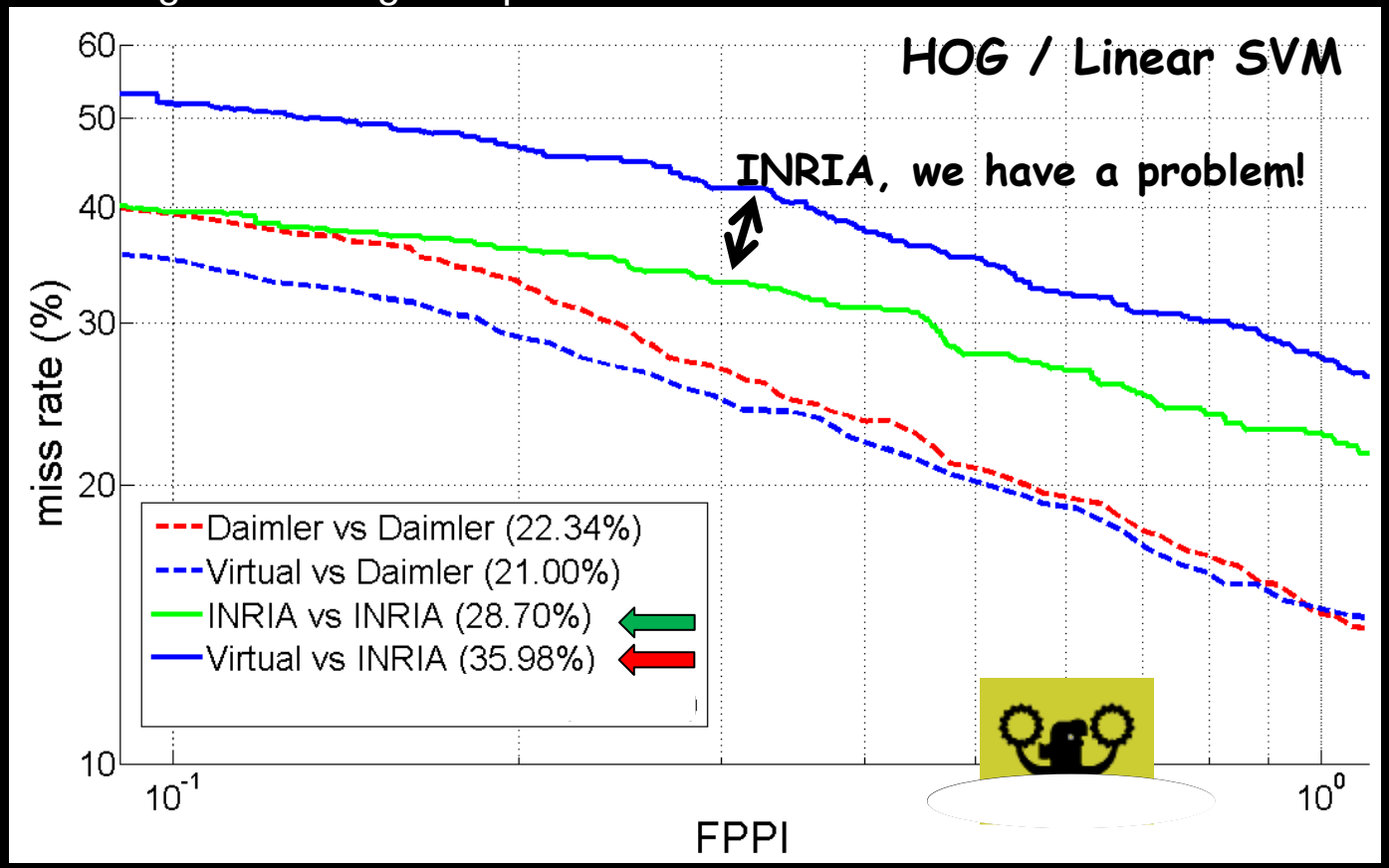
Results: testing in INRIA real-world data ◄

Examples **Counterexamples**

INRIA (real)

N. Dalal and B. Triggs. **Histograms of oriented gradients for human detection**. *CVPR* 2005.

Different training and testing sets provided.



Did any body experienced a similar problem? ◄

Google

the classifier drops accuracy in the new testing set, I'm frustrated

Did any body experienced a similar problem? ◄

Google

the classifier drops accuracy in the new testing set, I'm frustrated

Frustratingly Easy Domain Adaptation

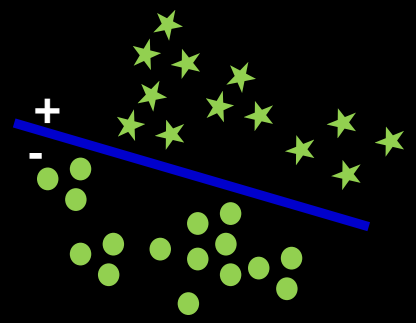
Hal Daumé III
School of Computing
University of Utah
Salt Lake City, Utah 84112
me@hal3.name

1 Introduction

The task of domain adaptation is to develop learning algorithms that can be easily ported from one domain to another—say, from newswire to biomedical documents. This problem is particularly interesting in NLP because we are often in the situation that we have a large collection of labeled data in one "source" domain (say, newswire) but truly desire a model that performs well in a second "target" domain. The approach we present in this paper is based on the idea of transforming the domain adaptation learning problem into a standard supervised learning problem to which any standard algorithm may be applied (eg., maxent, SVMs, etc.). Our transformation is incredibly simple: we augment the feature space of both the source and target data and use the result as input to a standard learning algorithm.

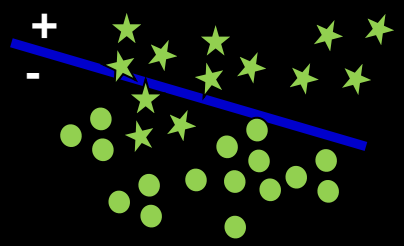
Domain adaptation◀

Domain shift: training and testing data follow different probability distributions

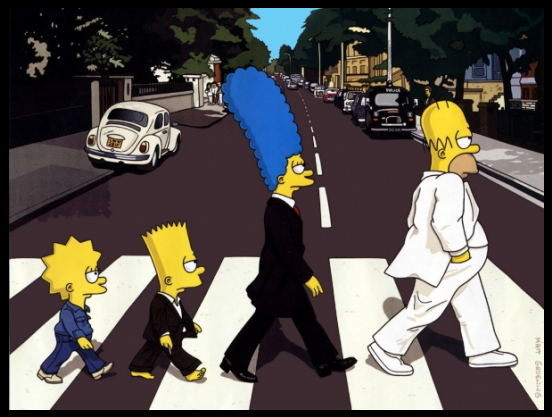


Source domain (training)

Domain Adaptation



Target domain (testing)



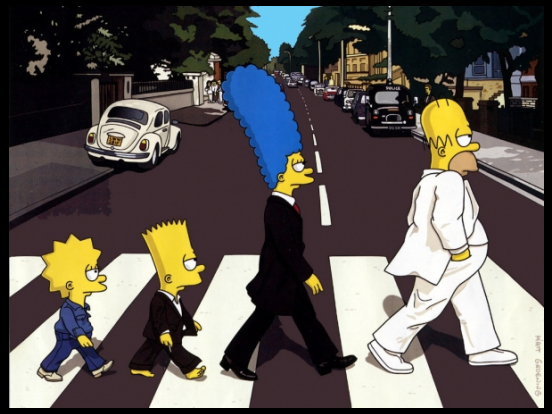
Domain adaptation: source knowledge (data or models) &
target data with few or no manual annotations.

Domain adaptation◀

Domain shift: training and testing data follow different probability distributions

Source domain: Virtual world

Target domain: Real world



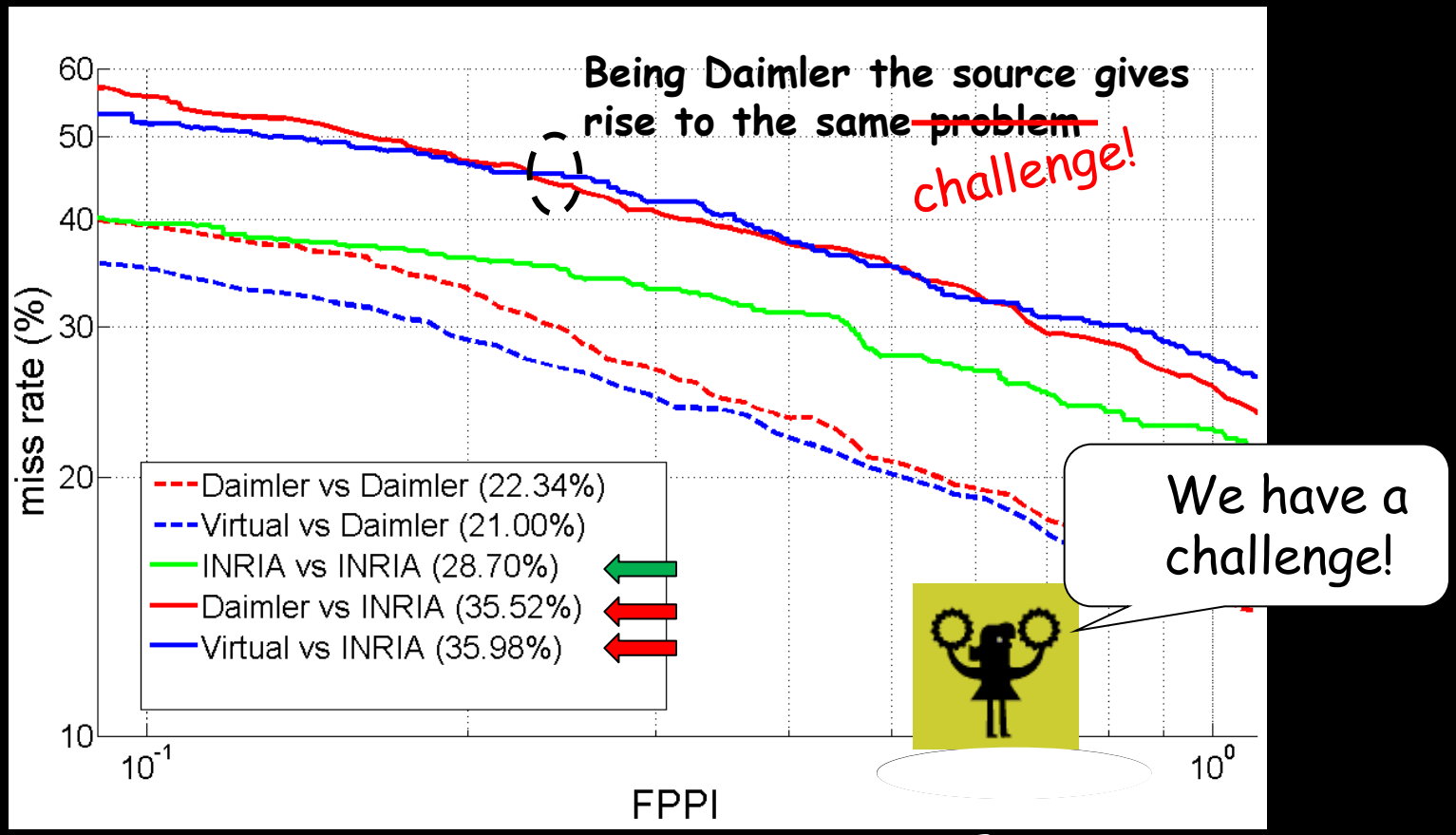
Domain adaptation: source knowledge (data or models) & target data with few or no manual annotations.

Source domain = virtual world; is this the problem? ◄

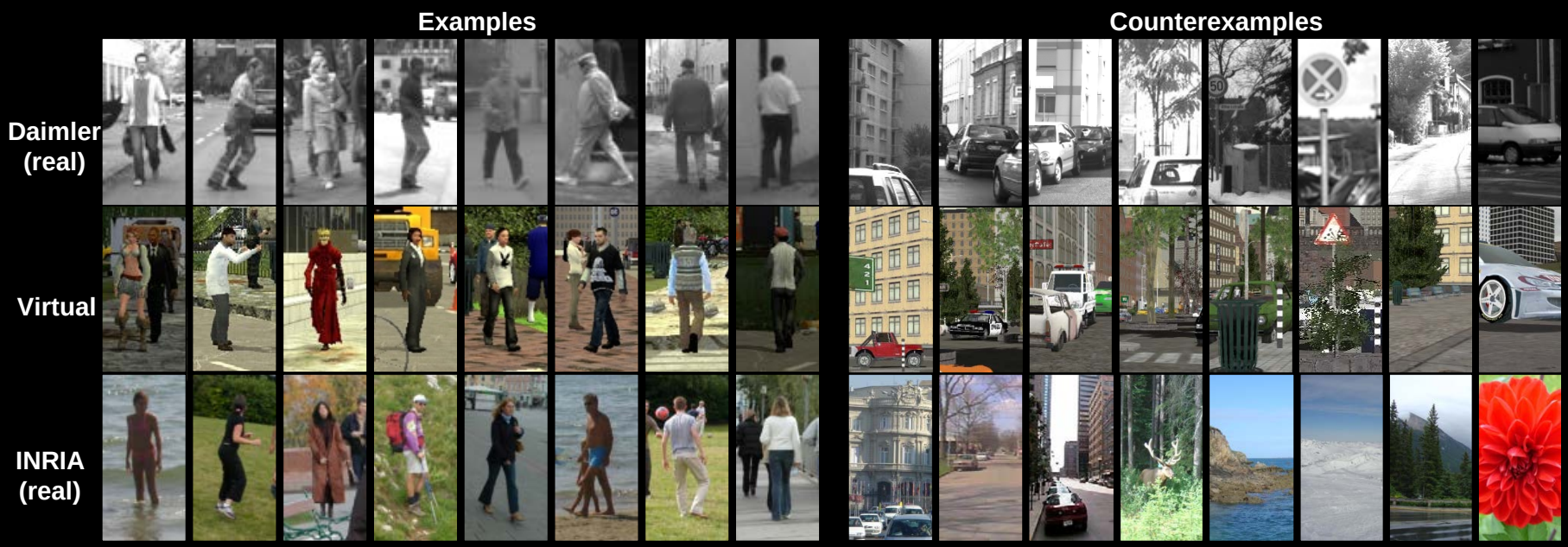
Examples **Counterexamples**

INRIA (real)

Different training and testing sets provided.



Dataset shift◀



Pedestrians:

- Similar poses in Virtual and Daimler.
- Similar clothes in each dataset, but not equal.

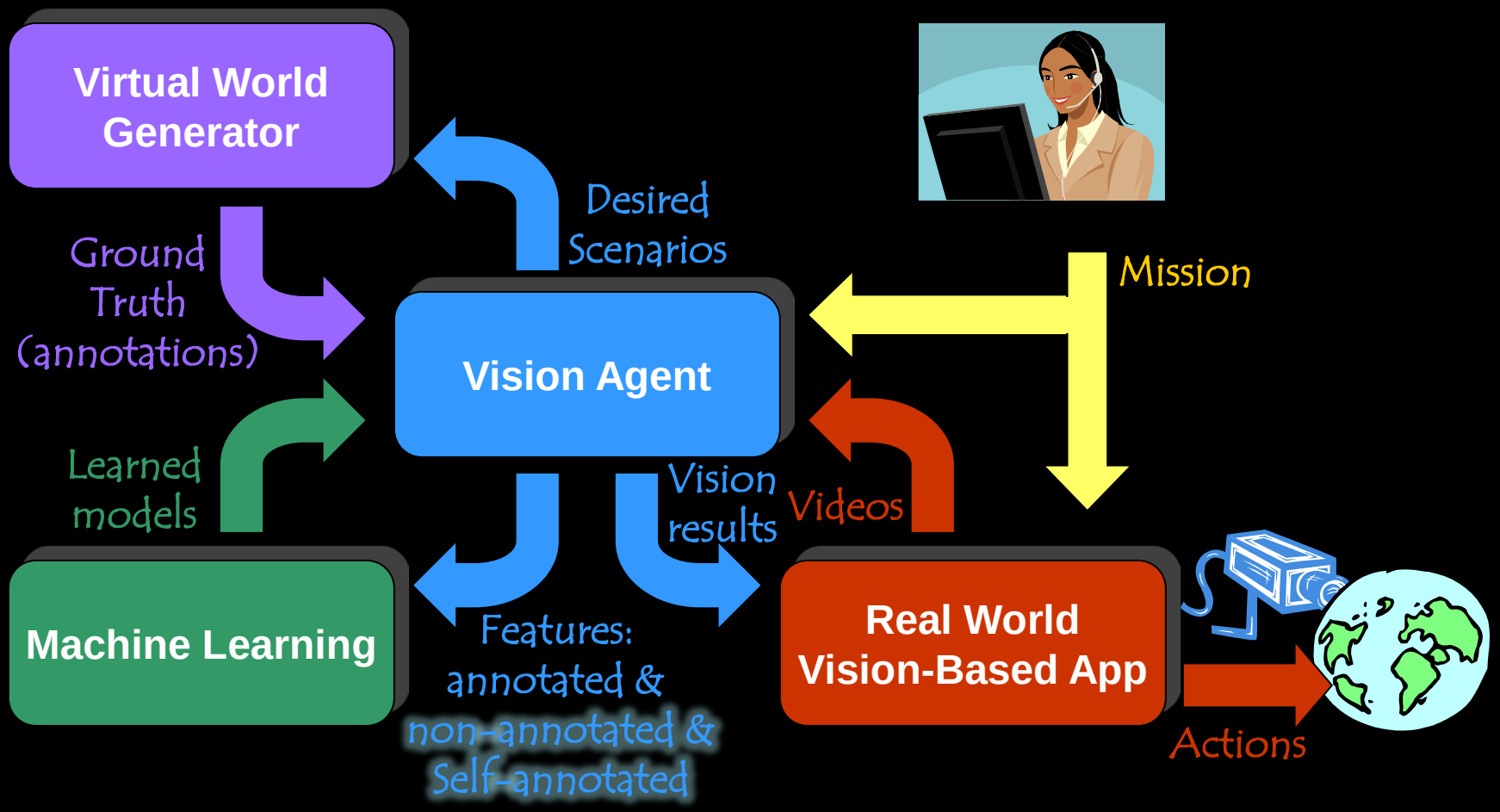
Backgrounds:

- Virtual and Daimler images were acquired onboard from urban scenarios.
- INRIA contains photos from different scenarios: beach, countryside, city, etc.

Sensors:

- Images captured by different eyes.

Self-training idea for Object Detection



Outline ◄

1. Autonomous driving.
2. Pedestrian (Object) detection: self-training.

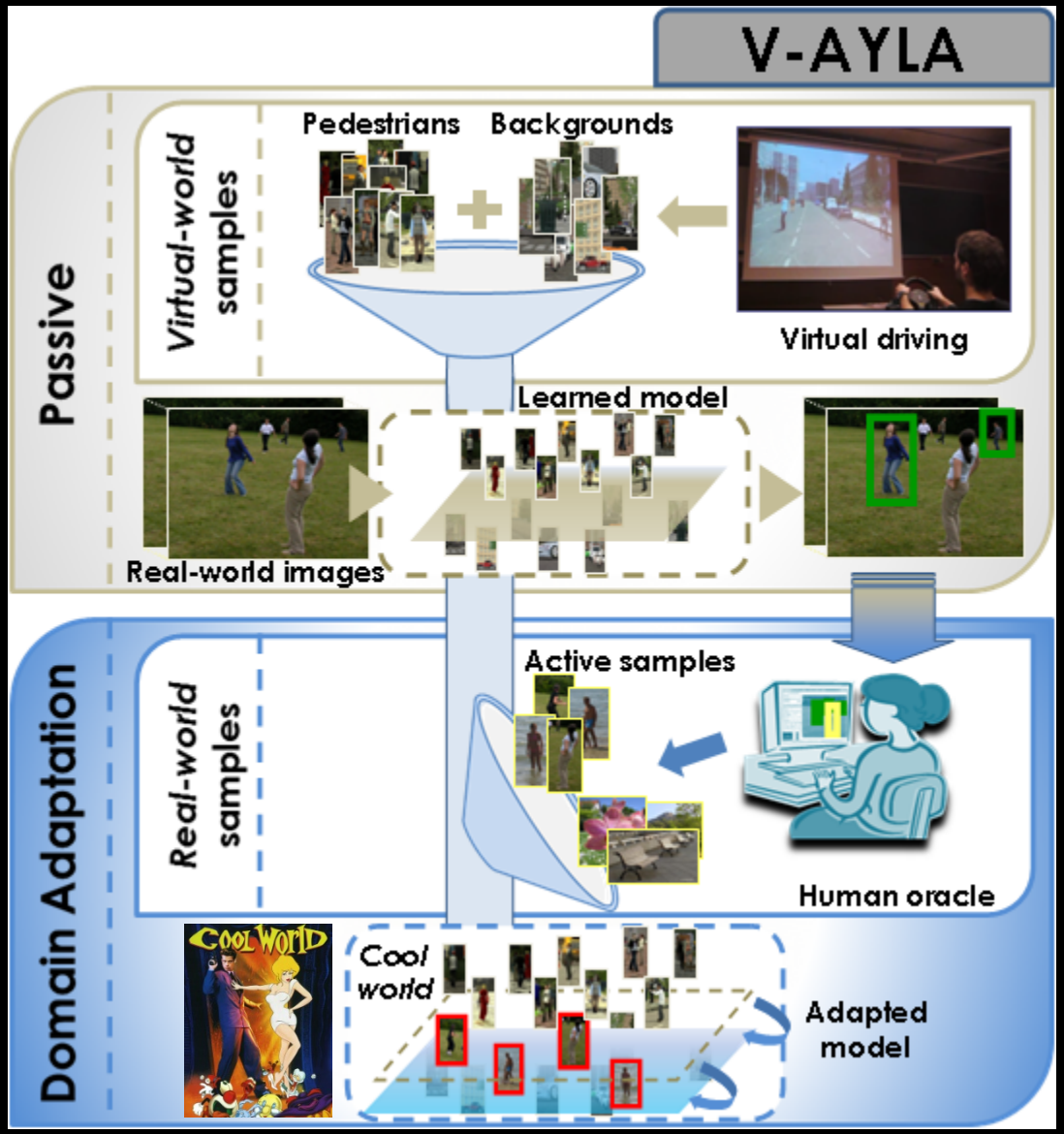
3. Virtual-world training.
4. Virtual and real world adaptation: cool-world.

5. Virtual-world DPM (VDPM).
6. Domain adaptation of VDPM.

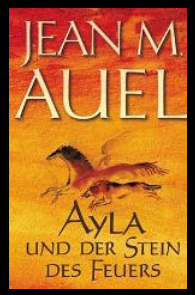
7. Random Forest of Local Experts: Adaptation.

8. Conclusions.

Proposed (pragmatic) framework: V-AYLA ◄



Virtual-world
Annotations,
Yet
Learning
Adaptively.



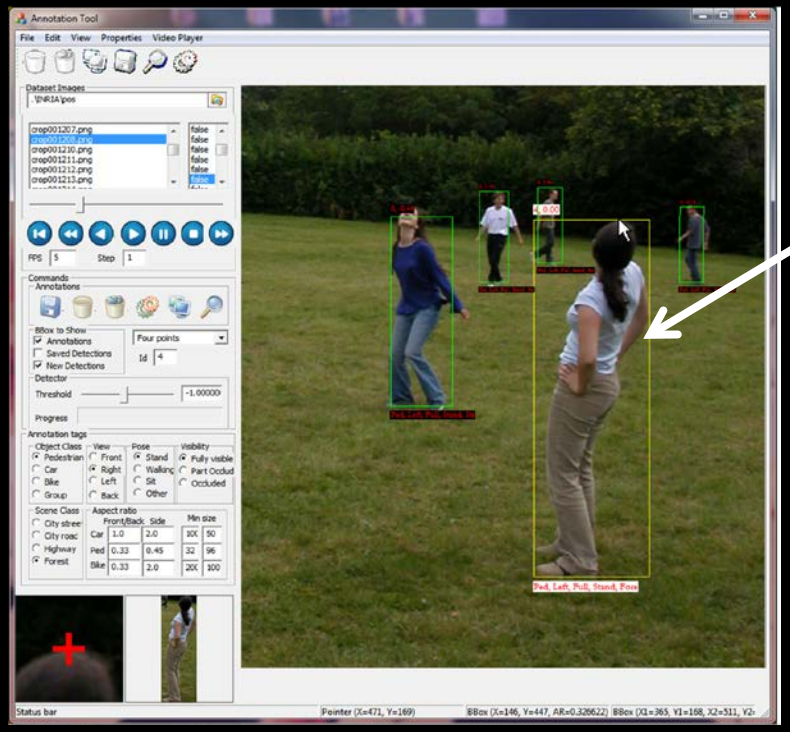
Basics:

- Supervised domain adaptation.
- Collect a few real examples using active learning.
- Learn in cool world: many virtual examples a few real ones.

D. Váquez, A.M. López, D. Ponsa, J. Marín. **Virtual worlds and Active Learning for Human Detection.** ICMI 2011.

D. Váquez, A.M. López, D. Ponsa, J. Marín. **Cool world: domain adaptation of virtual and real worlds for human detection using active learning.** NIPS 2011, D.A. Workshop.

V-AYLA: Experiments ◄



Non-detected

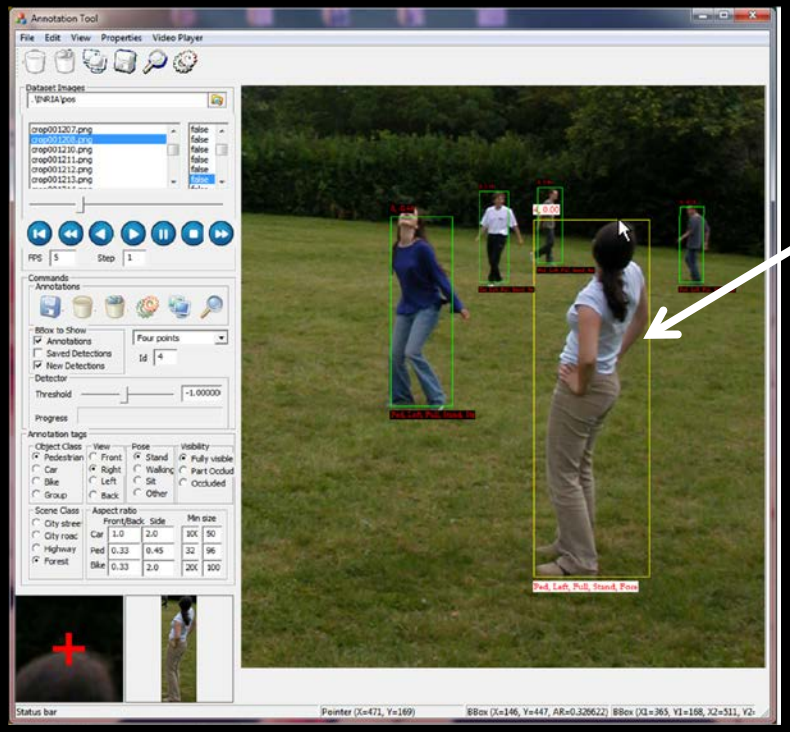
Detecting and annotating in the real-world training set

REAL (training)

↓

REAL (testing)

V-AYLA: Experiments ◄



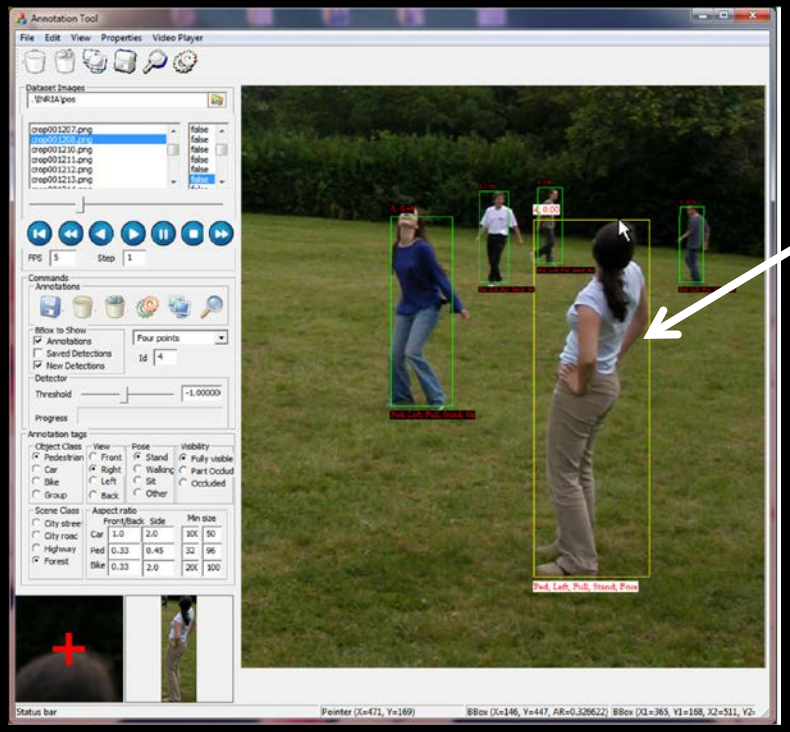
Non-detected

VIRTUAL (training)

REAL (testing)

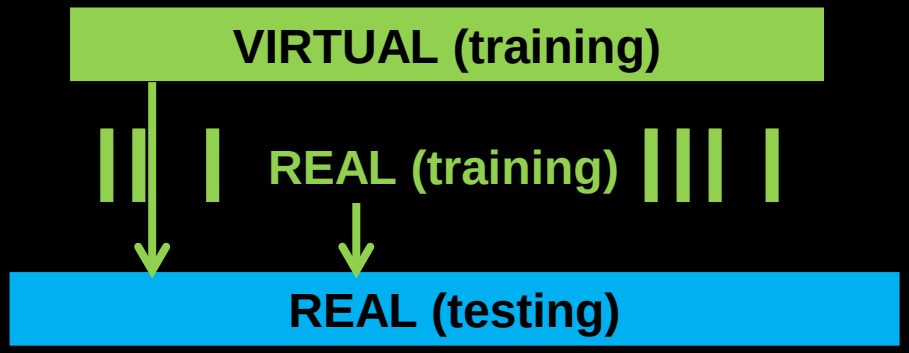
Detecting and annotating in the real-world training set

V-AYLA: Experiments ◄



Non-detected

Detecting and annotating in the real-world training set

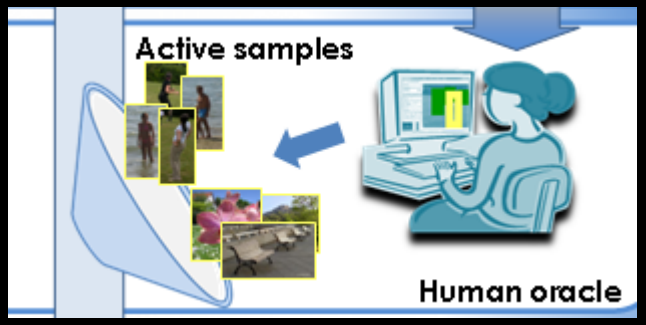


V-AYLA: Experiments-520 ◄

- Cool world:

  - ➢ Original ($\mathbf{x}_{s\&t}$) and Augmented ($<\mathbf{x}_s, \mathbf{x}_{s\&t}, \mathbf{x}_t>$).

- Human oracle:



  - ➢ Rnd, Act+, Act-, Act±, Act~

- Features:

  - ➢ HOG, LBP, HOG+LBP, with Linear SVM.

- Source datasets: Virtual. Only for training.
- Target datasets: Daimler, INRIA ….. Caltech. Train and test subsets.

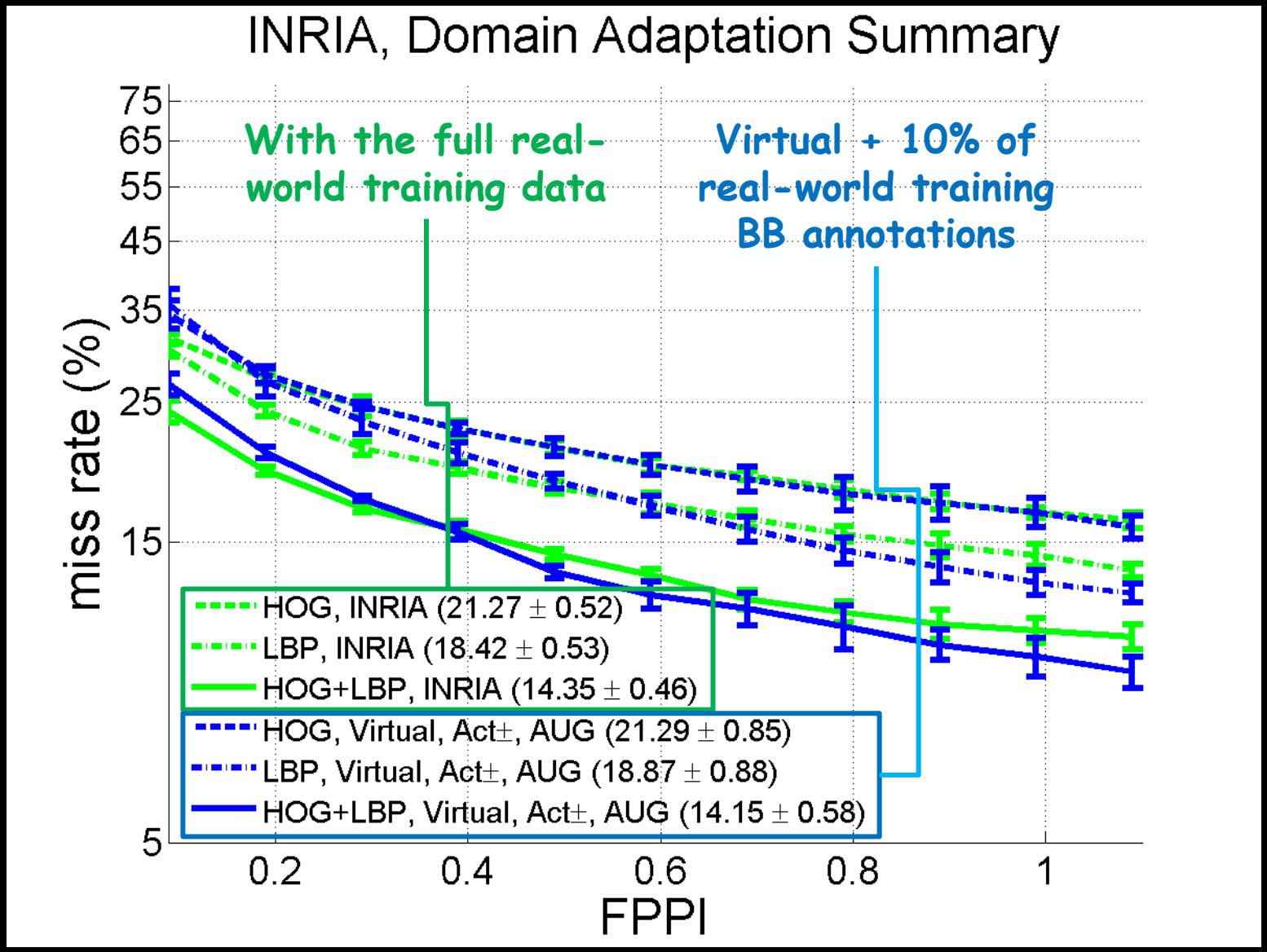
- Allowed BBs annotation effort in the real world: 10%.

V-AYLA: Experiments-520 ◄

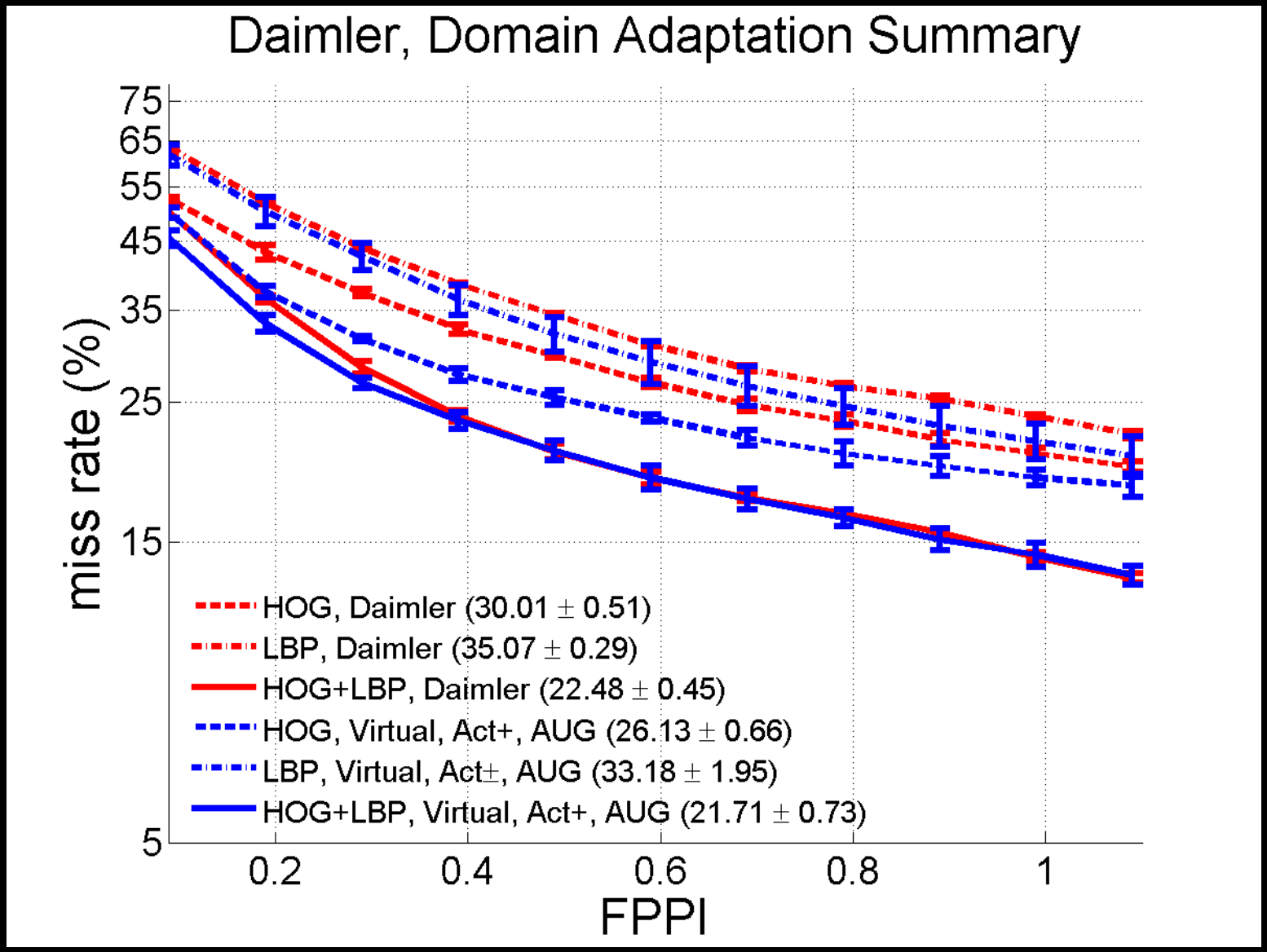
Passive Learning	$\mathcal{T}_{\mathcal{X}}^{tr}$	$\mathcal{T}_{\mathcal{R}}^{tt}$	
		$\mathcal{I}$	$\mathcal{D}$
HOG	$\mathcal{D}$	38.46±0.45	30.01±0.51 ⋆35.62±0.33
	$\mathcal{I}$	**21.27**±0.52 ⋆27.86±0.60	41.12±1.01
	$\mathcal{V}$	32.47±0.47	30.64±0.43
LBP	$\mathcal{D}$	39.54±0.55	**35.07**±0.29 ∘50.03±0.36
	$\mathcal{I}$	**18.42**±0.53 ∘34.53±0.82	35.40±0.70
	$\mathcal{V}$	28.87±0.70	45.21±0.49
HOG + LBP	$\mathcal{D}$	32.28±0.47	**22.48**±0.45 ∘38.04±0.46 •28.85±0.52
	$\mathcal{I}$	14.35±0.46 ∘23.92±0.81	26.22±0.85
	$\mathcal{V}$	23.81±0.53	28.27±0.48

(⋆) Dalal *et al.* implementation [24].
(∘) Wang *et al.* impl. [25], without occlusion handling.
(•) Training with the 15,660 pedestrians (Sect. 2.1).

V-AYLA: Experiments-520 ◄



V-AYLA: Experiments-520 ◄



V-AYLA: Experiments ◄

INRIA 100%: 14.35 ± 0.46

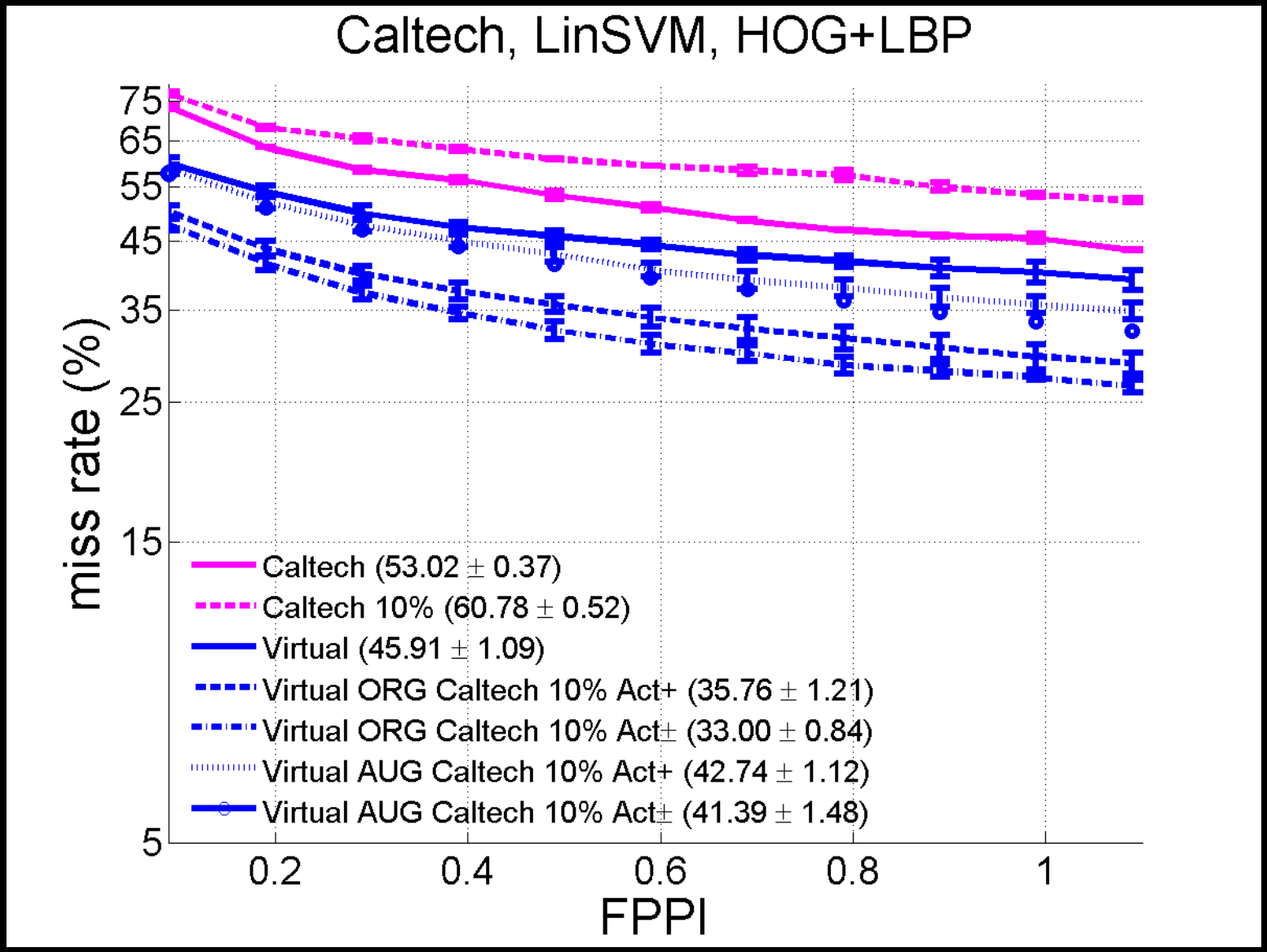
HOG+LBP				
INRIA ($\mathcal{T}_{\mathcal{I}}^{tt}$)	Act+	Act− Act∼	Rnd	Act±
ORG	16.65 ± 0.74	19.34 ± 0.60 19.61 ± 0.51	18.56 ± 0.61	15.10 ± 0.91
AUG	14.70 ± 0.63	17.46 ± 0.63 15.47 ± 0.89	15.07 ± 1.29	$\mathbf{14.15} \pm 0.58$
Daimler ($\mathcal{T}_{\mathcal{D}}^{tt}$)	Act+	Act∼	Rnd	Act±
ORG	22.85 ± 0.43	24.15 ± 0.73	23.64 ± 0.57	22.18 ± 0.65
AUG	$\mathbf{21.71} \pm 0.73$	27.43 ± 1.31	22.20 ± 1.26	23.79 ± 1.01

Daimler 100%: 22.48 ± 0.45

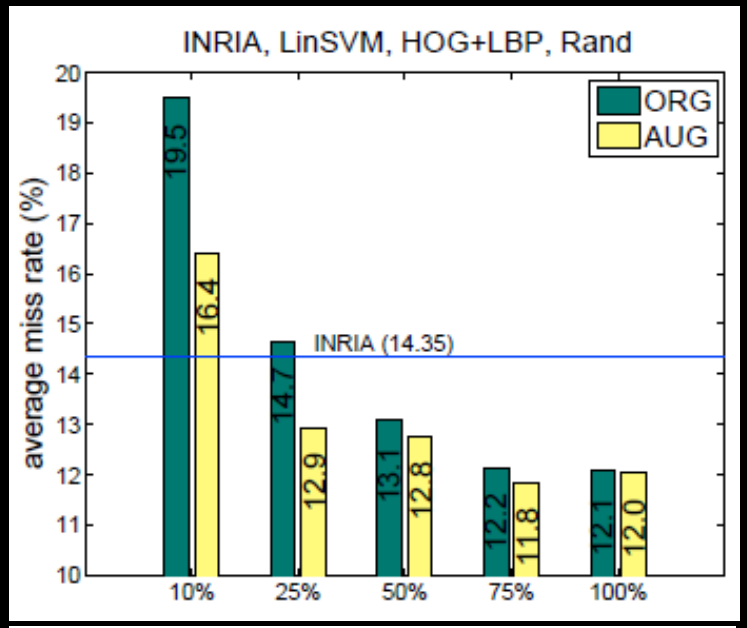
V-AYLA: Experiments-520 ◄

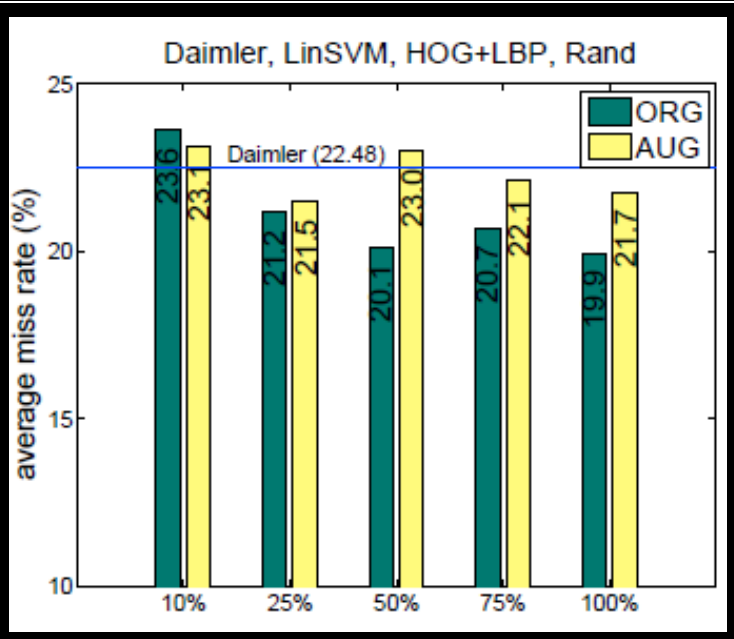
INRIA ($\mathcal{T}_{\mathcal{I}}^{tt}$)	HOG	LBP	HOG+LBP
$\mathcal{T}_{\mathcal{V}}^{tr}$	32.47 ± 0.47	28.87 ± 0.70	23.81 ± 0.53
$10\% \mathcal{T}_{\mathcal{I}}^{tr}$	30.81 ± 1.51	26.56 ± 1.96	18.89 ± 1.24
Act+/AUG	22.47 ± 1.01	22.83 ± 0.92	14.70 ± 0.63
Δ_1	10.00	06.04	09.11
Δ_2	08.34	03.73	04.19
Daimler ($\mathcal{T}_{\mathcal{D}}^{tt}$)	HOG	LBP	HOG+LBP
$\mathcal{T}_{\mathcal{V}}^{tr}$	30.64 ± 0.43	45.21 ± 0.49	28.27 ± 0.48
$10\% \mathcal{T}_{\mathcal{D}}^{tr}$	34.64 ± 1.31	41.13 ± 1.36	30.96 ± 1.59
Act+/AUG	26.13 ± 0.66	34.69 ± 1.15	21.71 ± 0.73
Δ_1	04.51	10.52	06.56
Δ_2	08.51	06.44	09.25

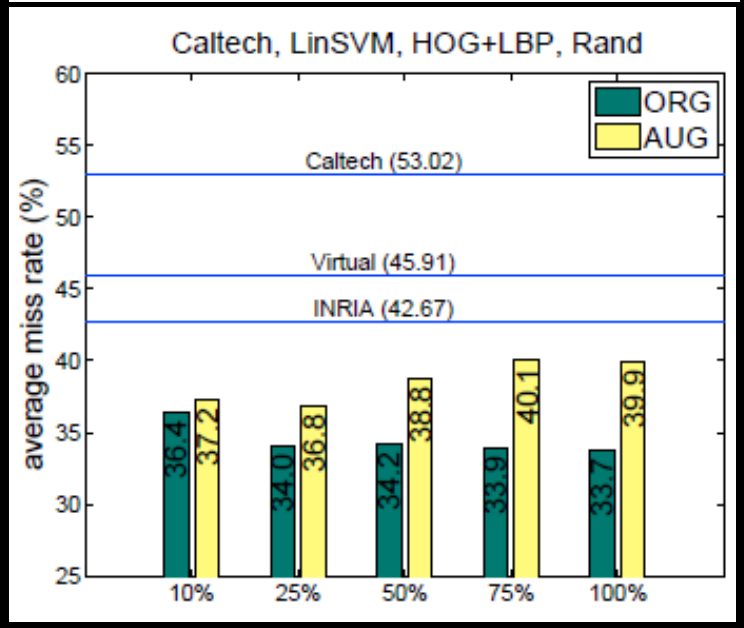
V-AYLA: Experiments-520 ◄



V-AYLA: Experiments-520 ◄







V-AYLA: Conclusions ◄

- Virtual world: virtual-world data is useful for training models that will operate in real-world if domain shift is solved.

- Cool world: For few target data AUG slightly outperforms ORG, when keeping adding target data there is not significant difference.

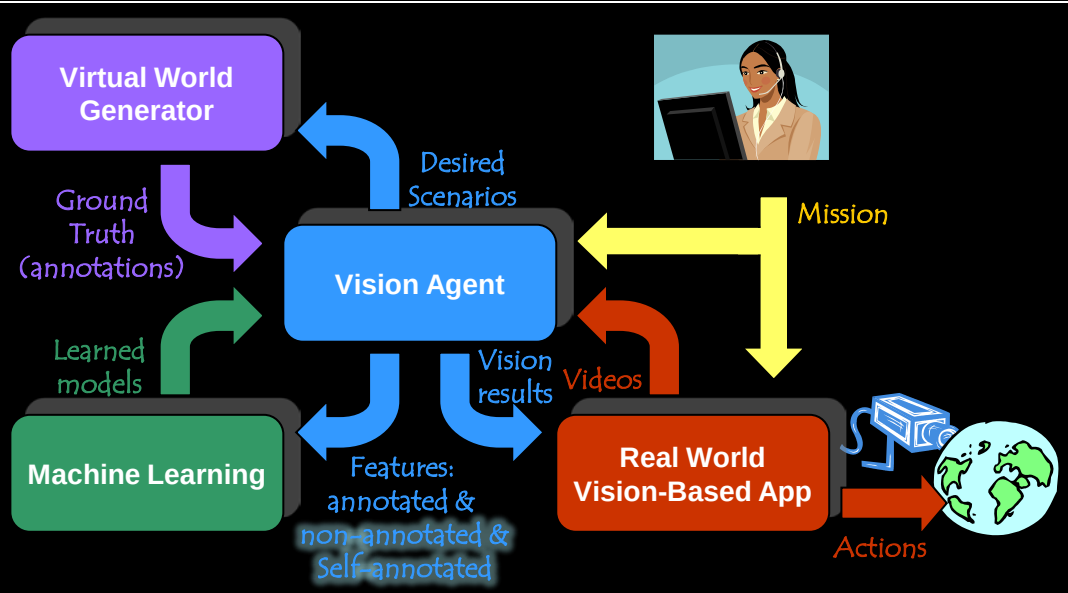
- Oracle (1): Rnd, Act+, and Act± perform similar, but Act+ and Act± can teach the annotators.

- Oracle (2): Act- performs slightly worse but still allows to “recover” large steps of accuracy with the advantage of not requiring the annotation of BBs.

- Oracle (3): Act- and Act~ are similar ➔ good BBs from detections.

➔ For HOG, LBP, HOG+LBP with Linear SVM, adaptation of real and virtual worlds have been achieved.

V-AYLA: Conclusions ◄



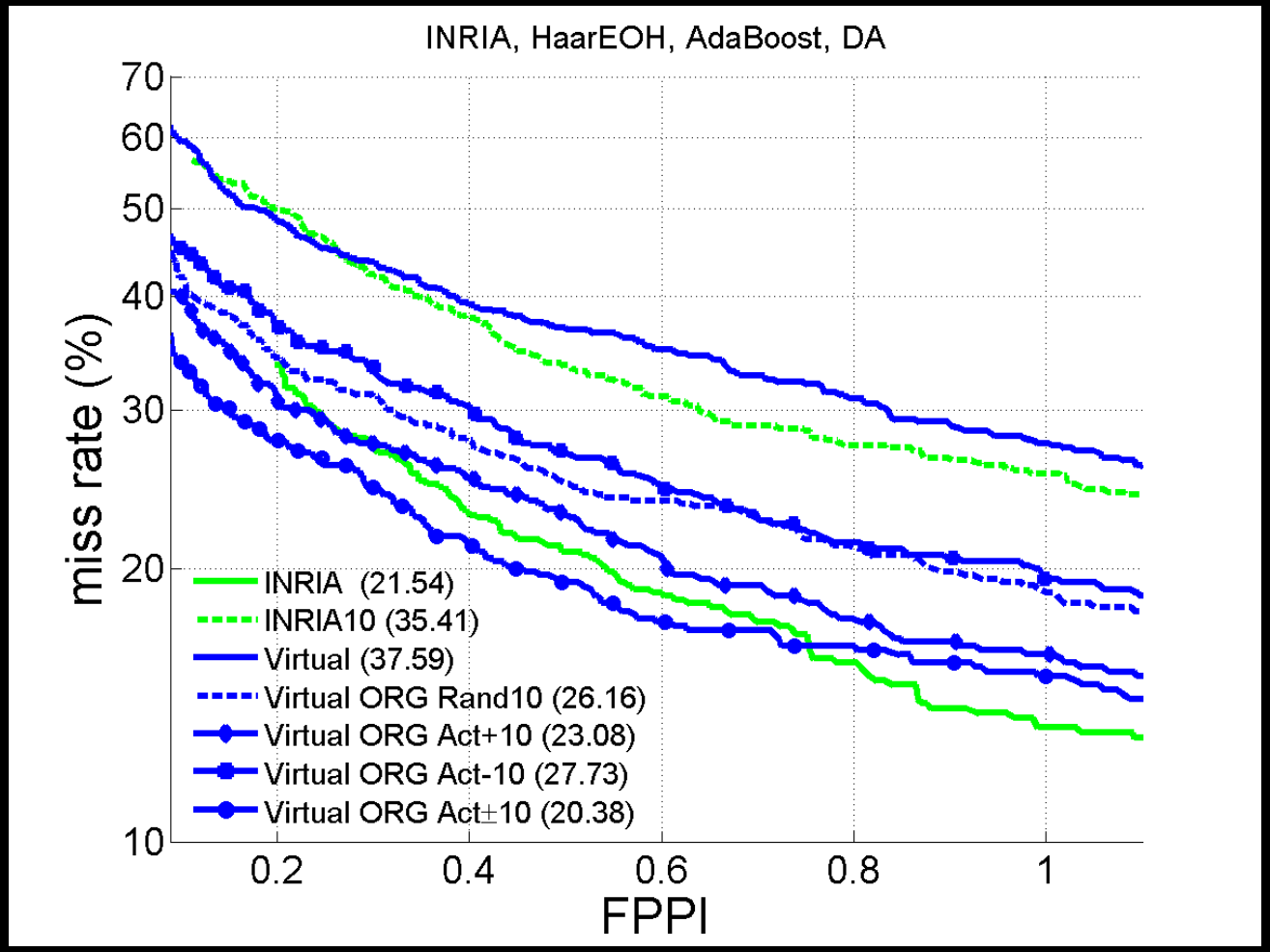
- <u>Oracle (2)</u>: Act- performs slightly worse but still allows to "recover" large steps of accuracy with the advantage of not requiring the annotation of BBs.

- <u>Oracle (3)</u>: Act- and Act~ are similar ➔ good BBs from detections.

➔ For HOG, LBP, HOG+LBP with Linear SVM, adaptation of real and virtual worlds have been achieved.

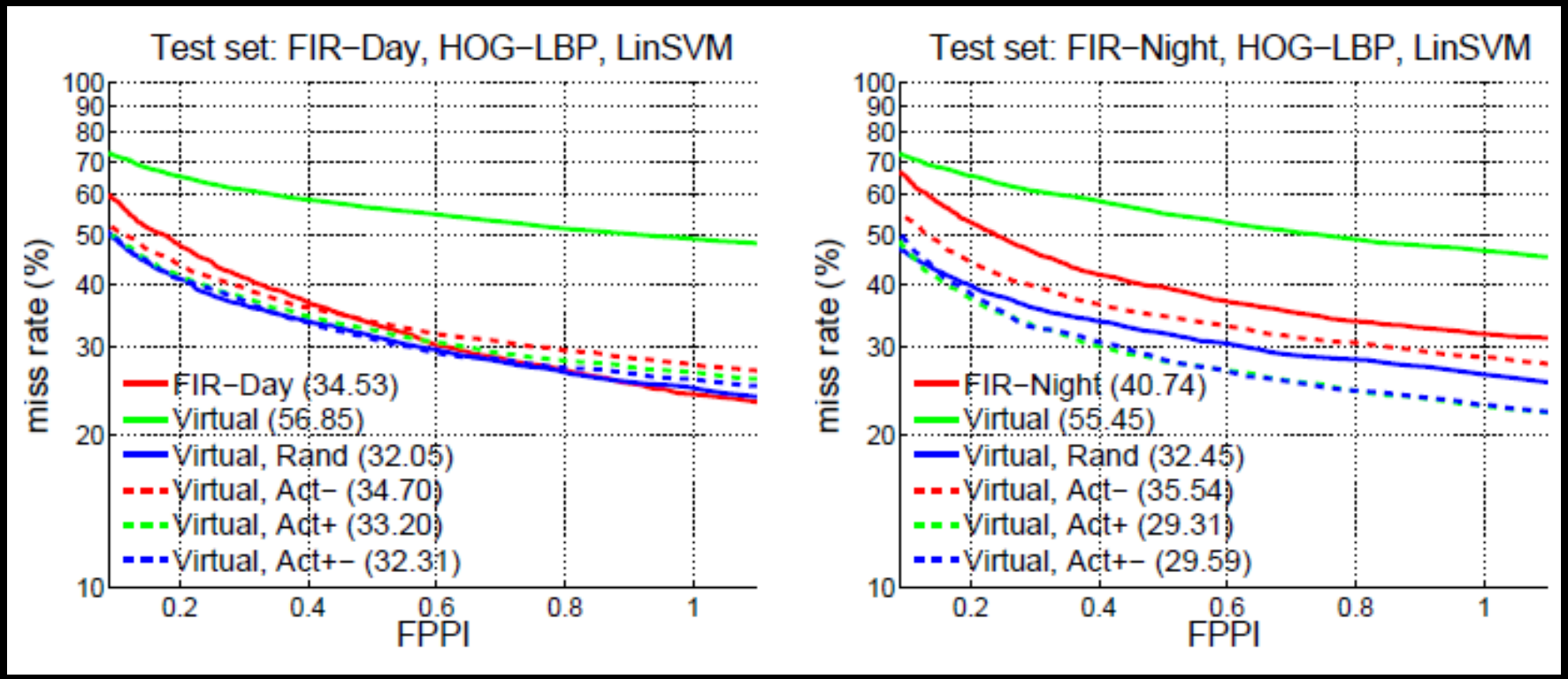
V-AYLA: More Experiments ◄

- We tested other settings:
  - ➢ Haar, EOH, Haar+EOH, features with Real-AdaBoost.



V-AYLA: More Experiments ◄

V-AYLA: More Experiments ◄



Datasets	FIR	
	Day	Night
Train	3110 (4548)	2198 (4333)
Test	2880 (2304)	2883 (2883)

Outline ◄

1. Autonomous driving.
2. Pedestrian (Object) detection: self-training.

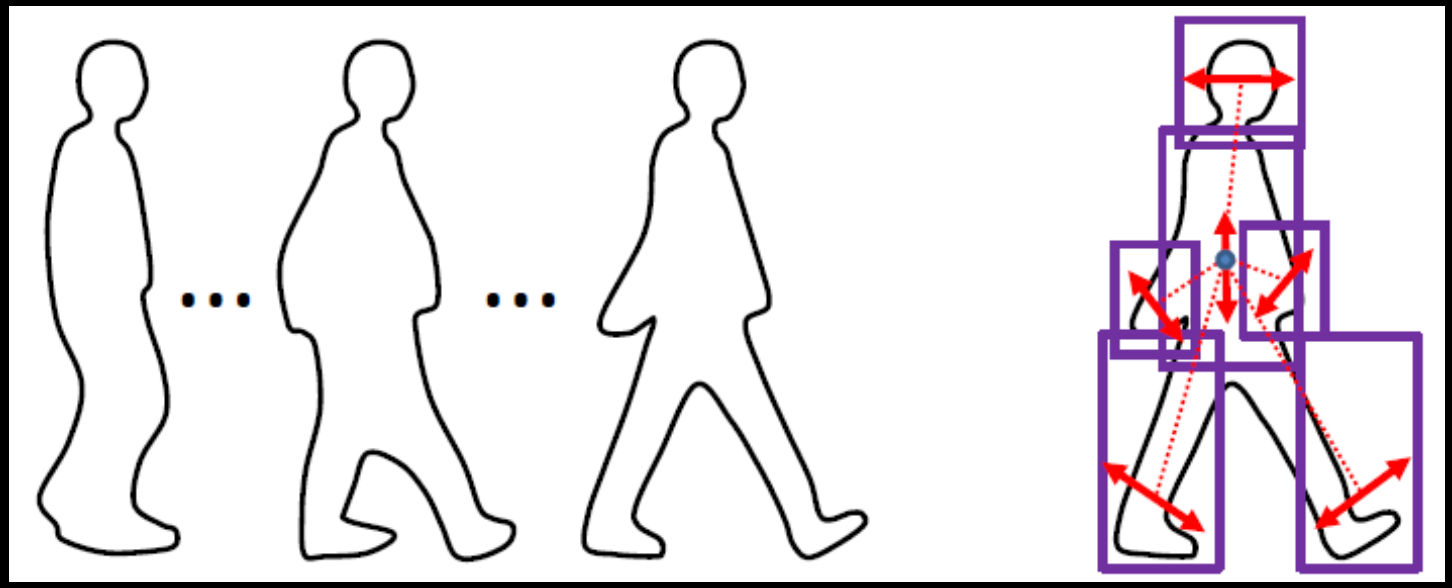
3. Virtual-world training.
4. Virtual and real world adaptation: cool-world.

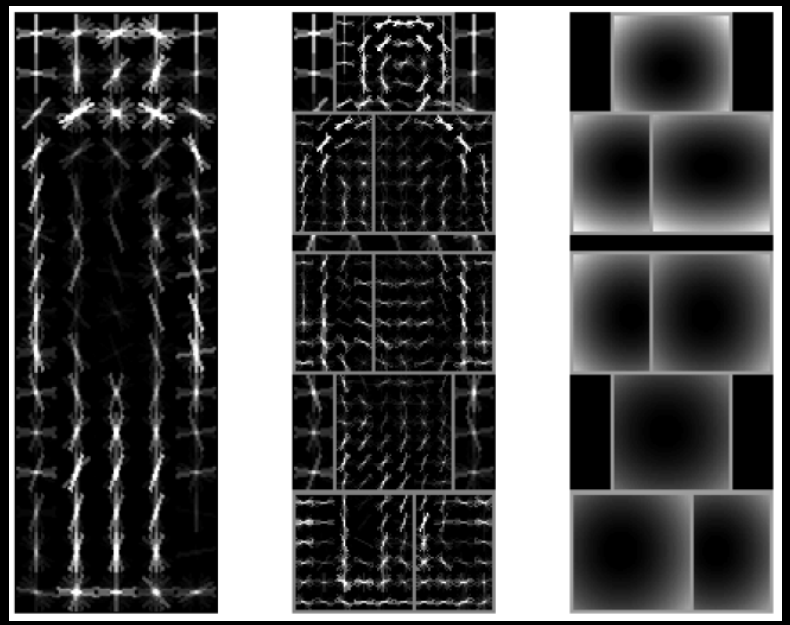
5. Virtual-world DPM (VDPM).
6. Domain adaptation of VDPM.

7. Random Forest of Local Experts: Adaptation & Demo.

8. Conclusions.

DPM: deformable part-based model◀



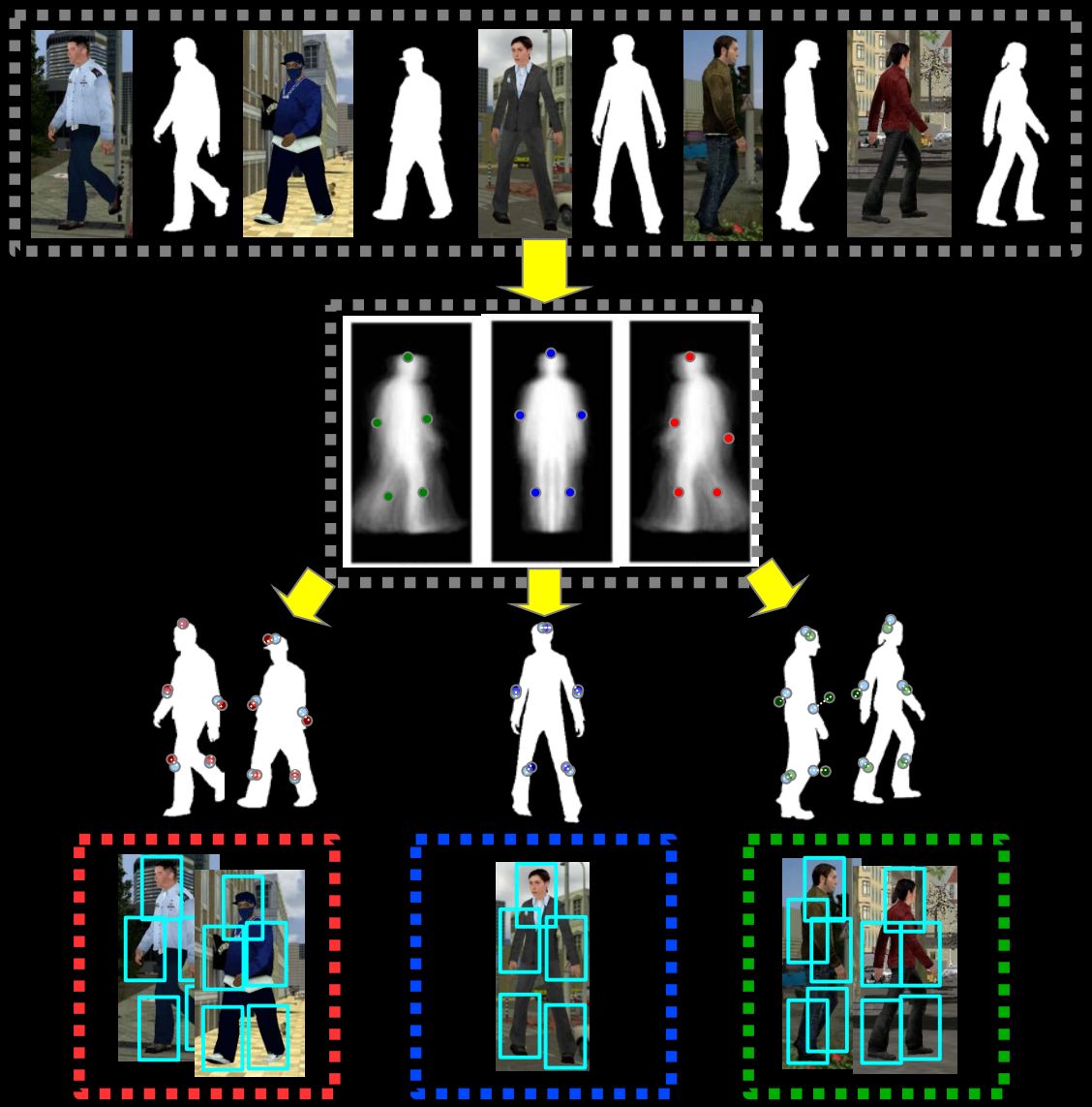


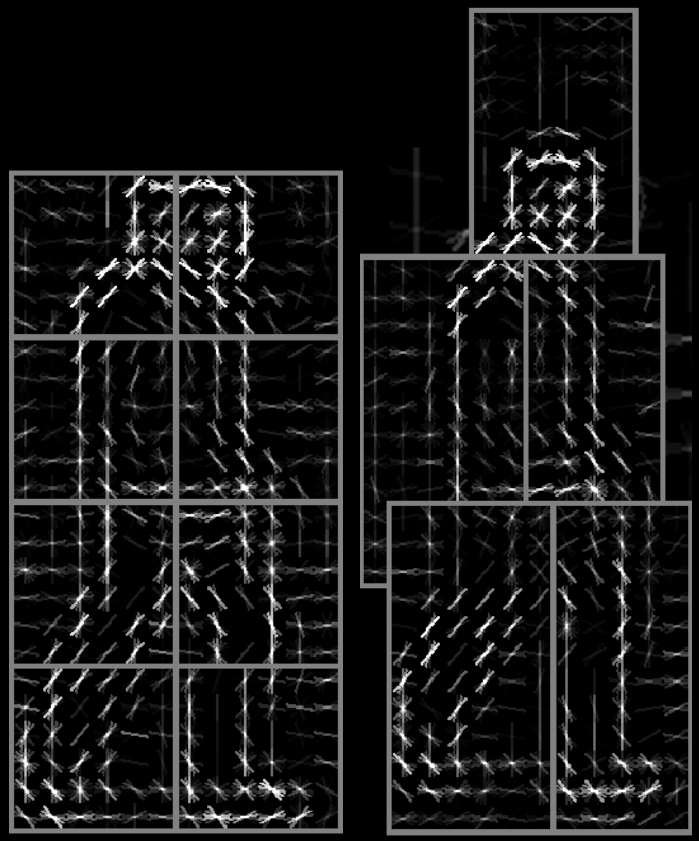
Appearance (HOG)
Parts / root
Deformation
Components

Learning: Latent SVM

VDPM: Learning with part-level supervision◀

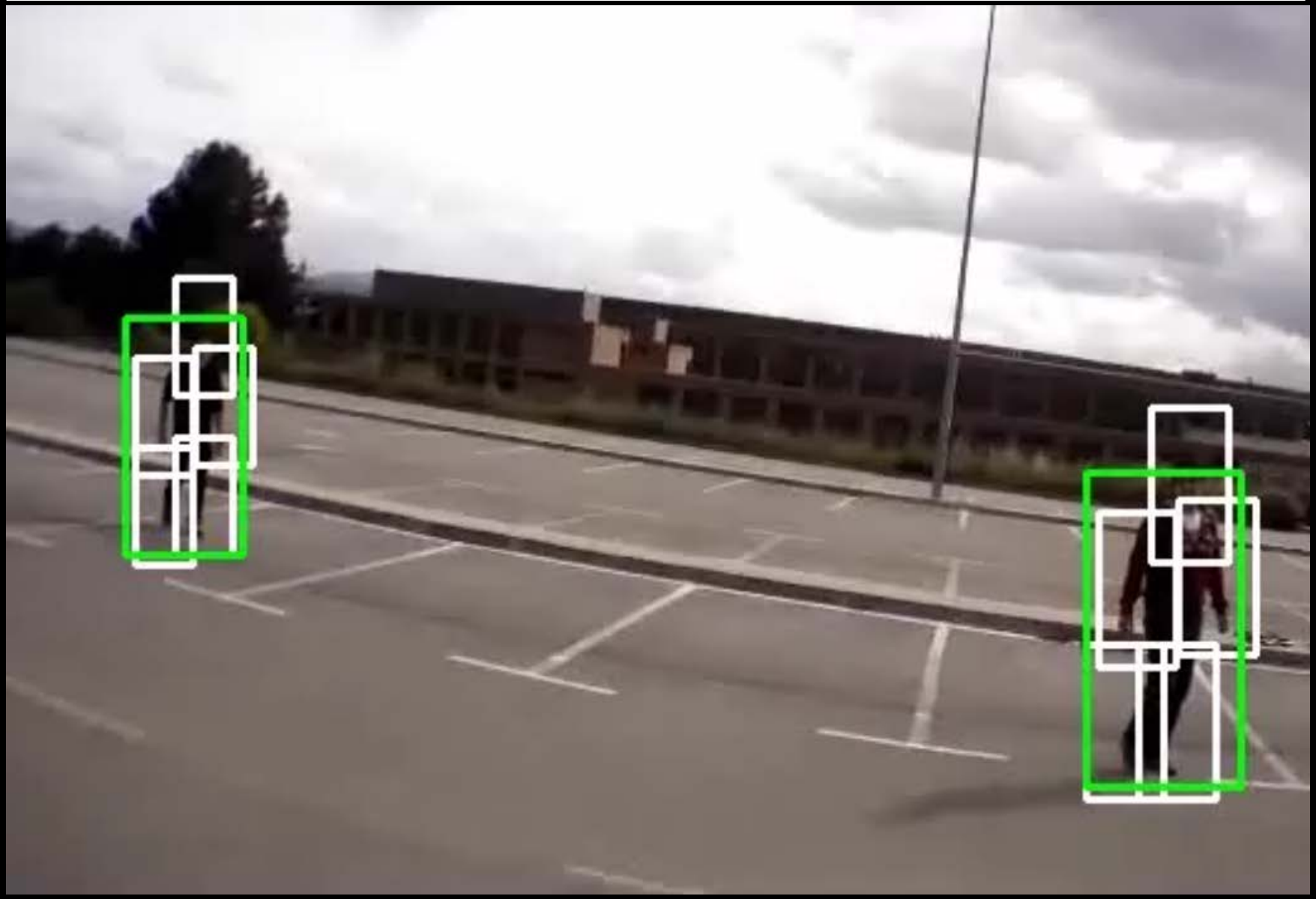
Source domain: virtual data ➔ no latent variables





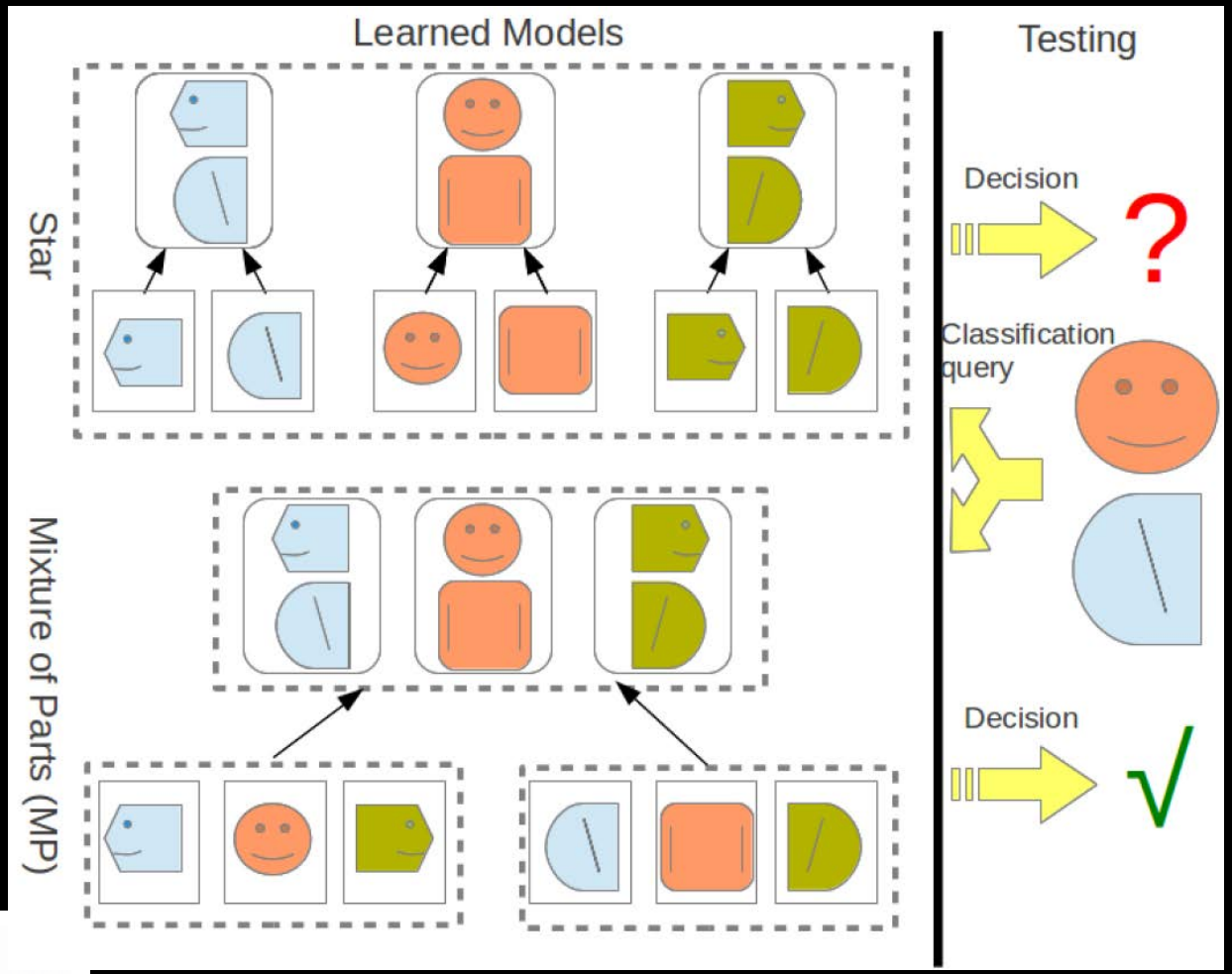
DPM **VDPM**

VDPM: Learning with part-level supervision◀



VDPM: Learning with part-level supervision◀

VDPM with Parts Sharing



IEEE TRANSACTIONS ON
INTELLIGENT TRANSPORTATION SYSTEMS
A PUBLICATION OF THE IEEE INTELLIGENT TRANSPORTATION SYSTEMS COUNCIL

Learning a Part-based Pedestrian Detector in Virtual World

Jiaolong Xu, David Vázquez, Antonio M. López *Member, IEEE,* Javier Marín and Daniel Ponsa

Outline ◄

1. Autonomous driving.
2. Pedestrian (Object) detection: self-training.

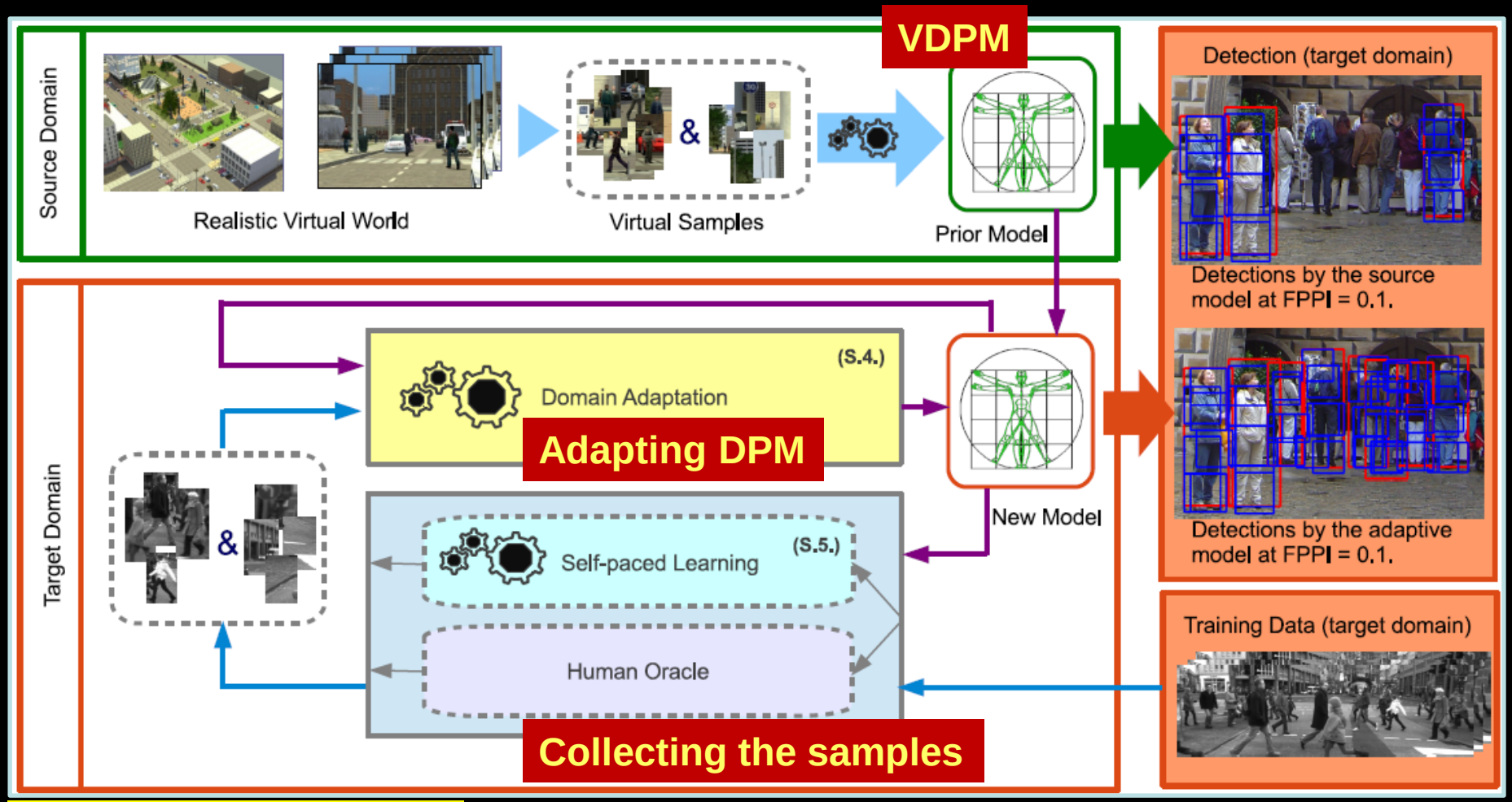
3. Virtual-world training.
4. Virtual and real world adaptation: cool-world.

5. Virtual-world DPM (VDPM).
6. Domain adaptation of VDPM.

7. Random Forest of Local Experts: Adaptation.

8. Conclusions.

V-AYLA meets VDPM◀



IEEE TRANSACTIONS ON
PATTERN ANALYSIS A
MACHINE INTELLIGENCE

Domain Adaptation of Deformable Part-Based Models

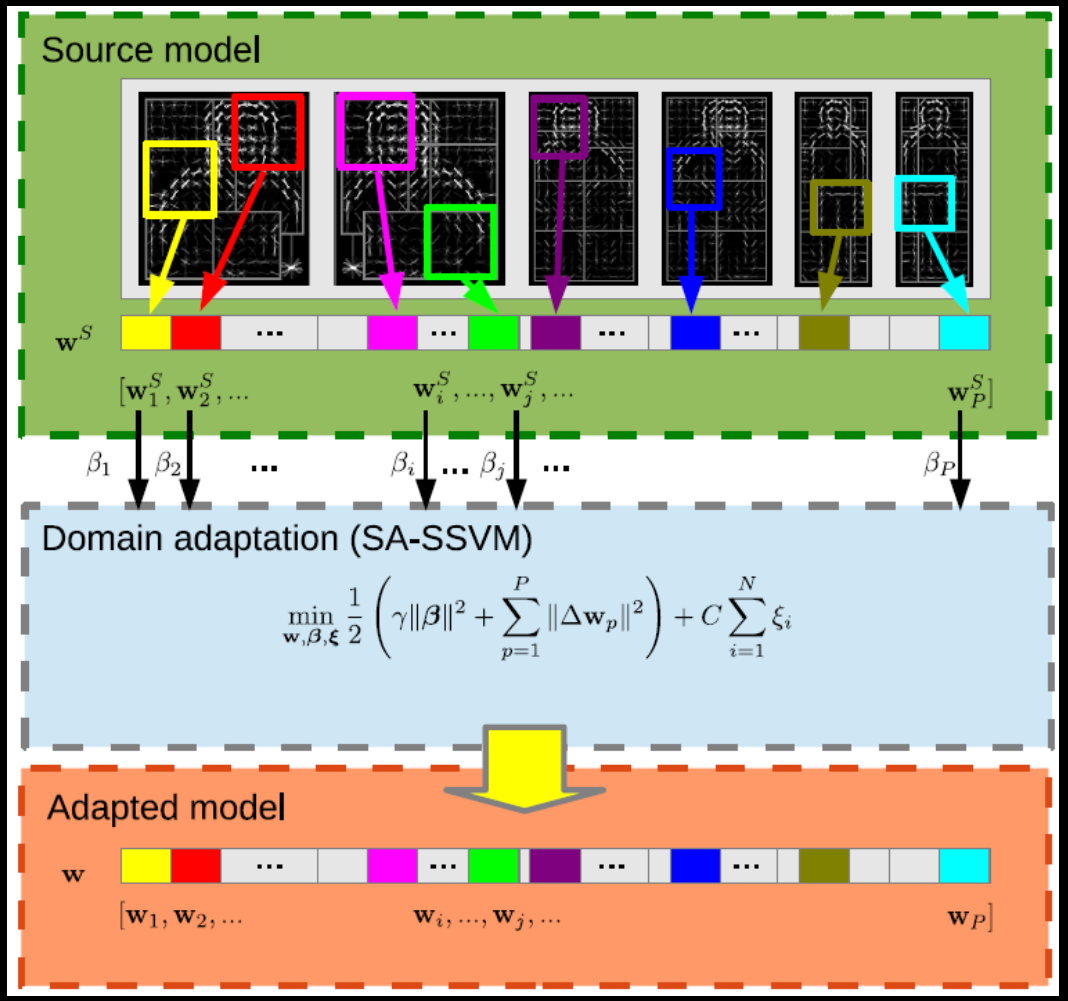
Jiaolong Xu, *Student Member, IEEE*, Sebastian Ramos, *Student Member, IEEE*, David Vázquez, *Member, IEEE*, Antonio M. López, *Member, IEEE*

Avoid Retraining with the Source Data

Adaptive SVM ➔ Structure Aware SSVM (SA-SSVM)◀

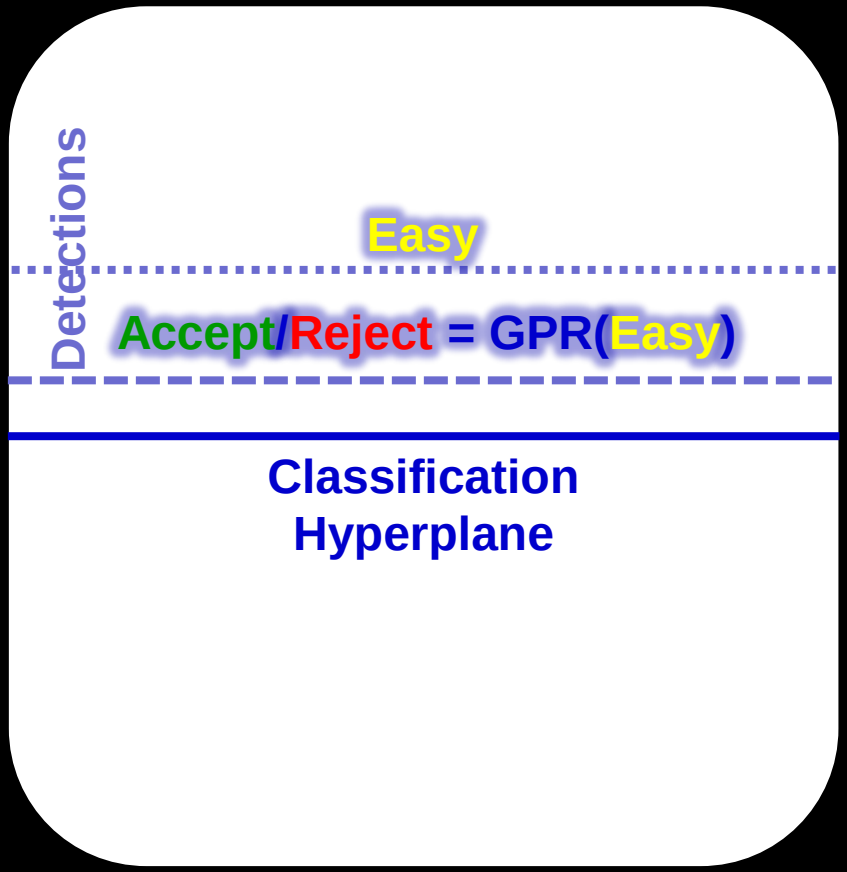
$$\min_{w,\xi} \frac{1}{2}\left\|w^T - w^S\right\|^2 + C\sum_{i=1}^{N}\xi_i$$

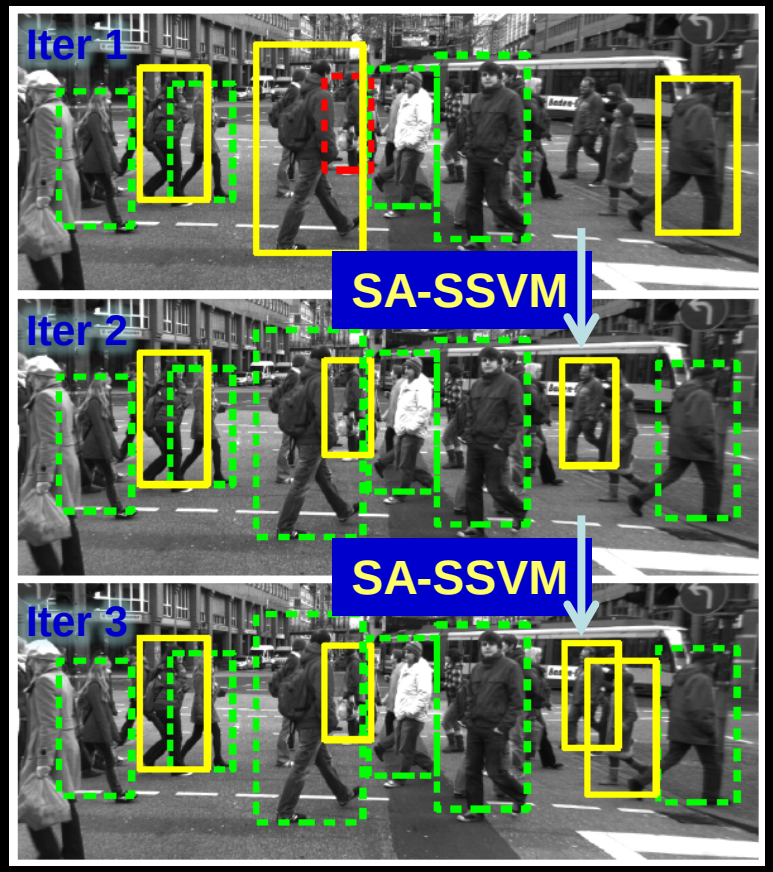
Adaptive SVM



Structure Aware A-SSVM (SA-ASSVM)

Self-detections for SA-SSVM◄



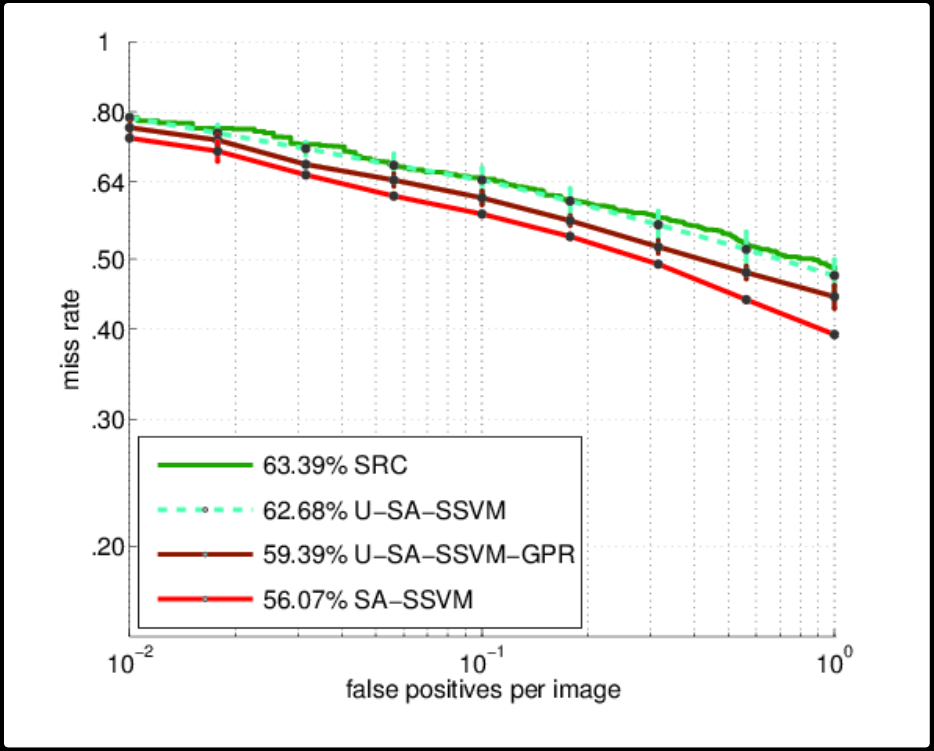


Self-paced Learning Inspired +
Gaussian Process Regression (GPR)

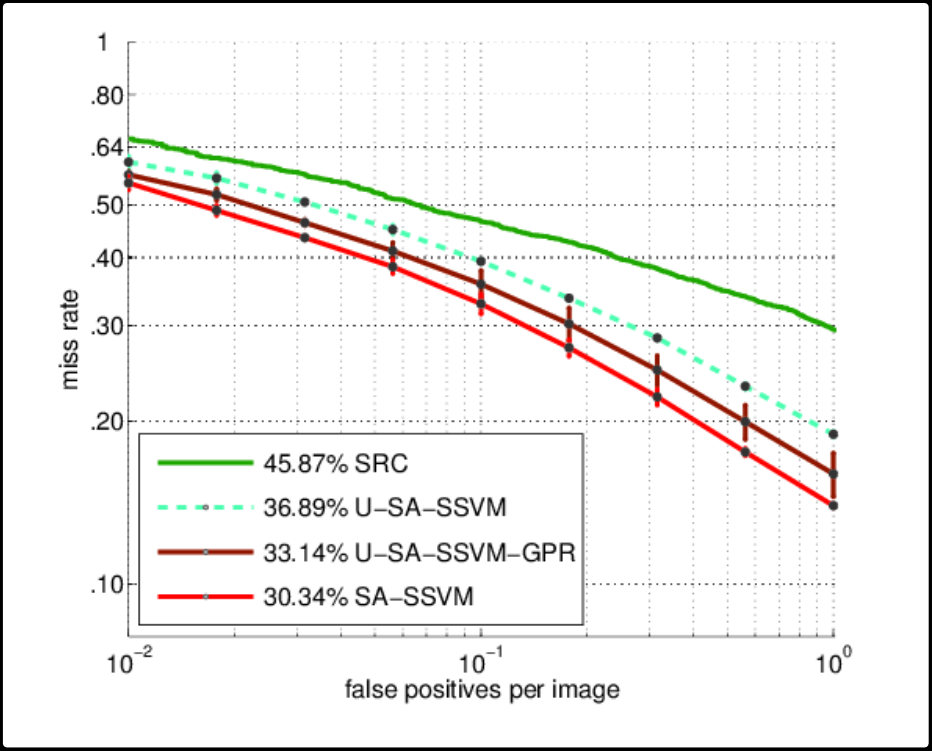
ACCEPTED by GPR
EASY (input GPR)
REJECTED by GPR

SA-SSVM experiments

Caltech



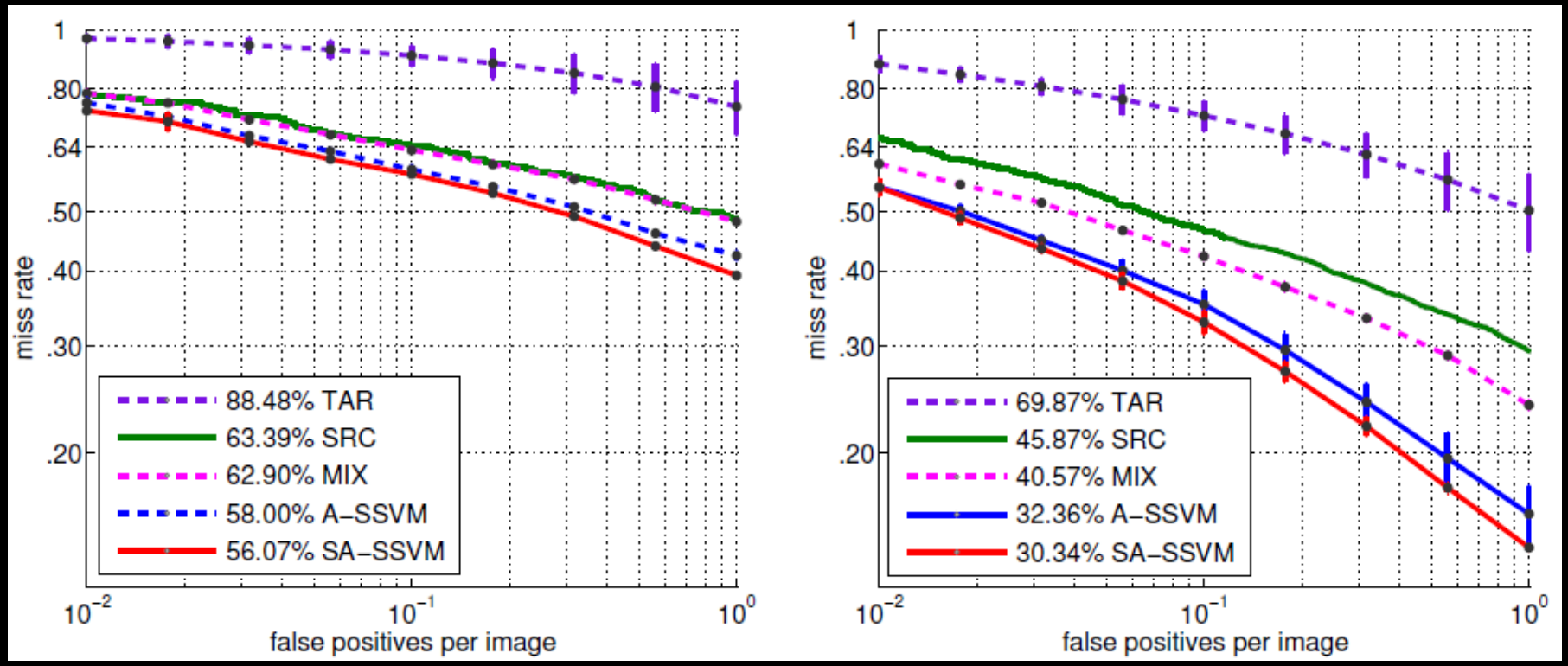
CVC02



SRC: VDPM
SA-SSVM: SA-SSVM with 100 manually annotated
U-SA-SSVM: SA-SSVM with self-paced learning over 500 images
U-SA-SSVM-GPR: As previous with GPR

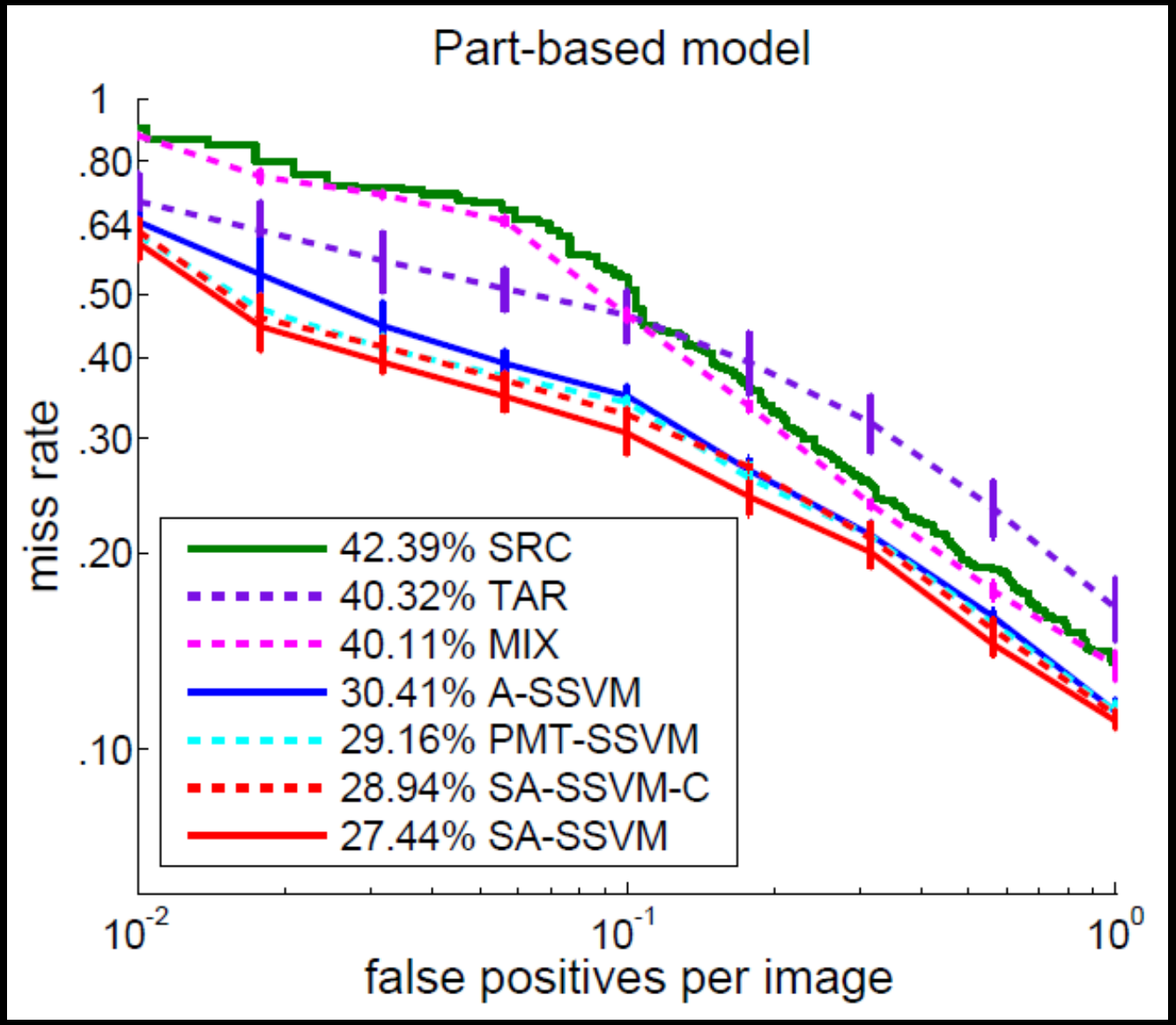
SA-SSVM experiments ◄

Caltech **CVC02**



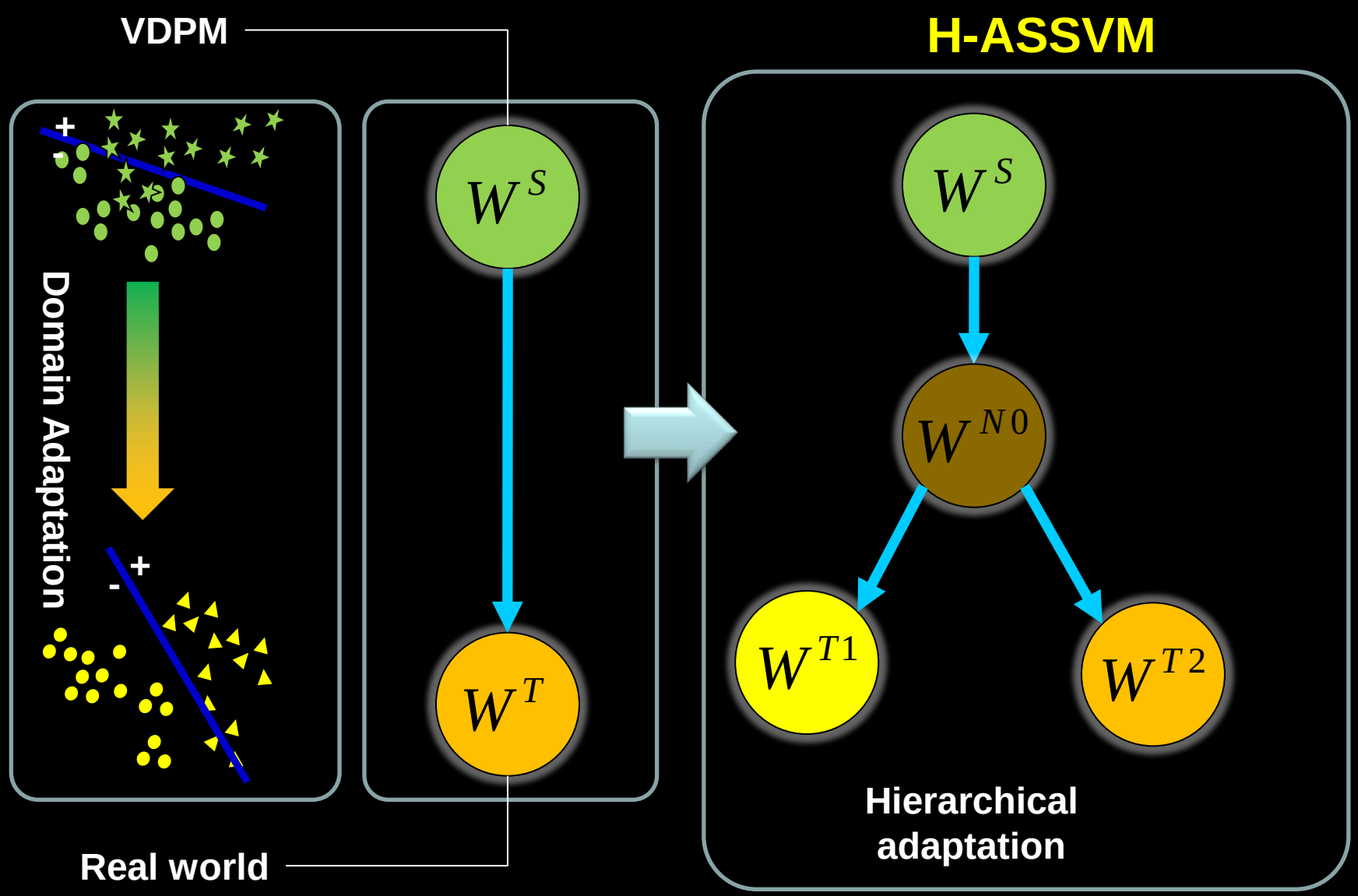
MIX: DPM with virtual data + 100 manually annotated real-world pedestrians
TAR: DPM with 100 manually annotated real-world pedestrians
A-SSVM: A-SVM applied to the DPM with the previous real-world data

SA-SSVM experiments ◄



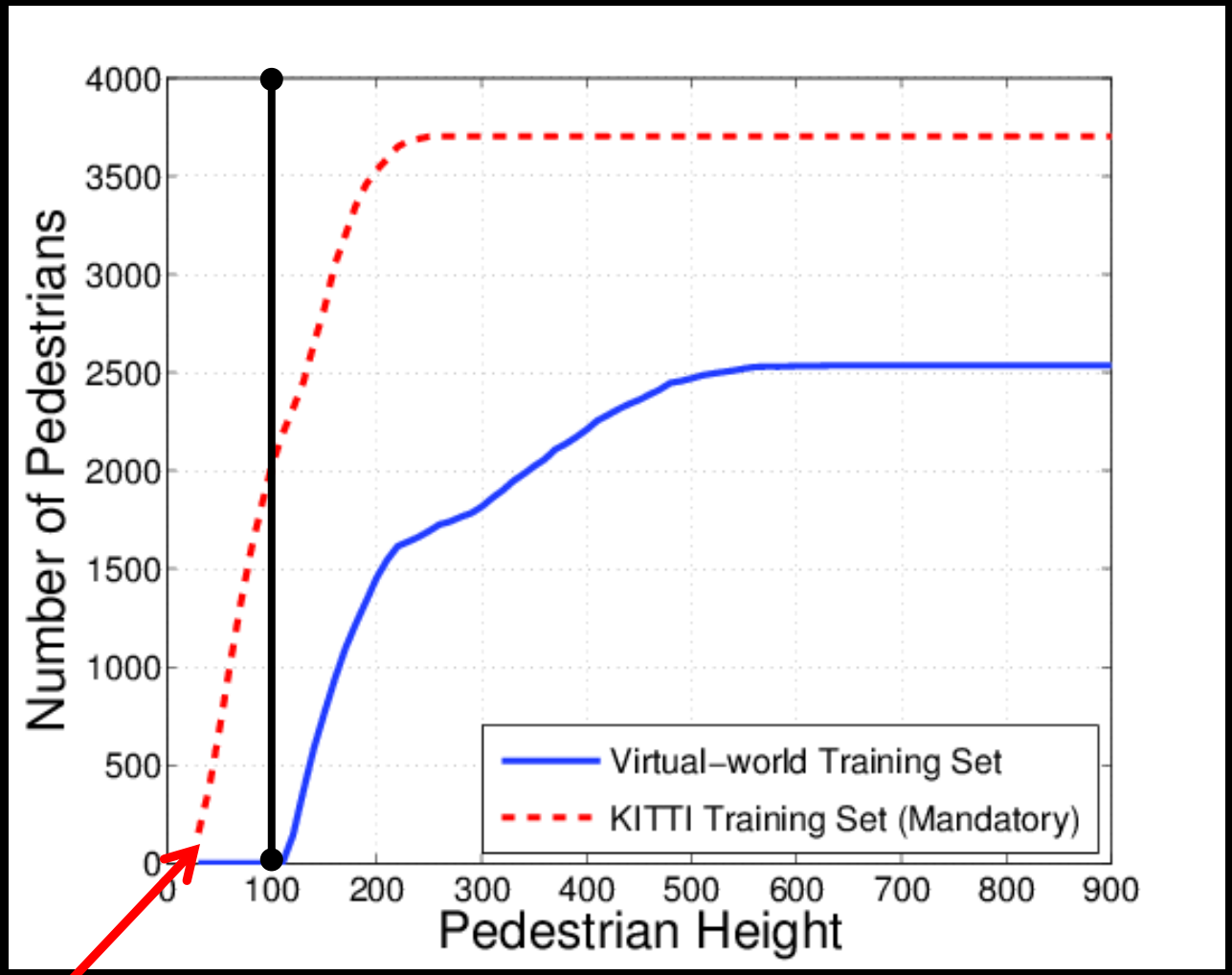
PASCAL➔ INRIA

Hierarchical A-SSVM (HA-SSVM)◀



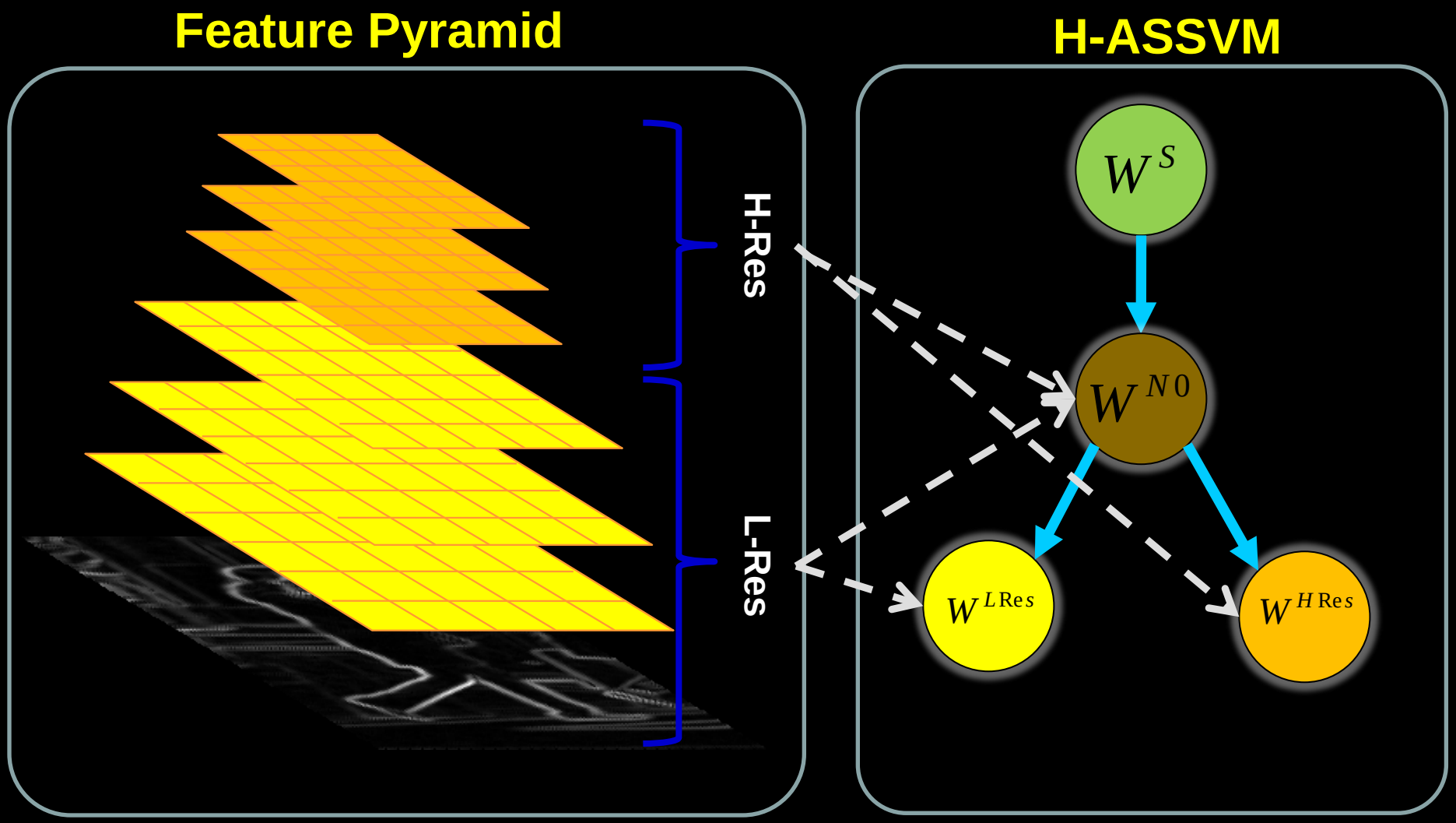
KITTI pedestrian detection challenge◀

ICCV2013 Workshop: Reconstruction Meets Recognition Challenge



KITTI minimum pedestrian height during detection: 25 pixels

Resolution Adaptation of VDPM to KITTI◀



HA-SSVM: KITTI pedestrian detection challenge◀

Settings:

1. **Source domain: virtual-world.**
   a. **Pedestrians: 1267.**
   b. **Background: 2000 images.**
2. **Target domain: KITTI.**
3. **Initial pedestrian classifier: VDPM.**
4. **Domain adaptation (using KITTI training set):**
   a. **Pedestrians: 200 (~11%).**
   b. **Background: 2000 images (~26%).**
5. **With these KITTI data we run DA-DPM.**

HA-SSVM: KITTI pedestrian detection challenge◀

Rank on KITTI dataset (during ICCV'2013)

(Average Precision %)

Method	Easy	Moderate	Hard
DA-DPM	**56.36 (↑ 8)**	**45.51 (↑ 6)**	**41.08 (↑ 5)**
LSVM-MDPM-sv	47.74	39.36	35.95
LSVM-MDPM-us	45.50	38.35	38.78
mBoW	44.28	31.37	30.62

http://www.cvlibs.net/datasets/kitti/eval_object.php

Experimental results

Training on KITTI dataset

DPM

1851 pedestrians

7518 bg images

> 10 hours

VS

H-ASSVM

200 pedestrians

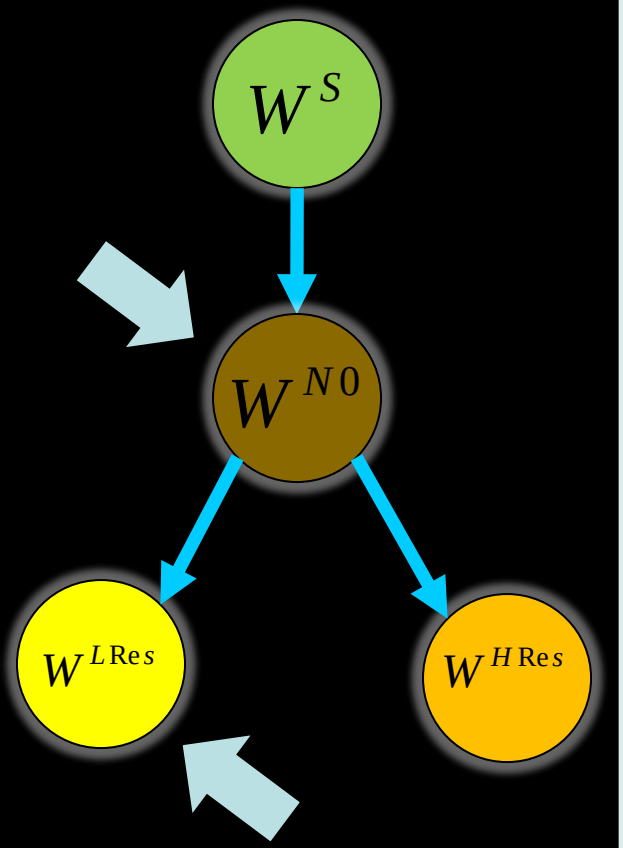
2000 bg images

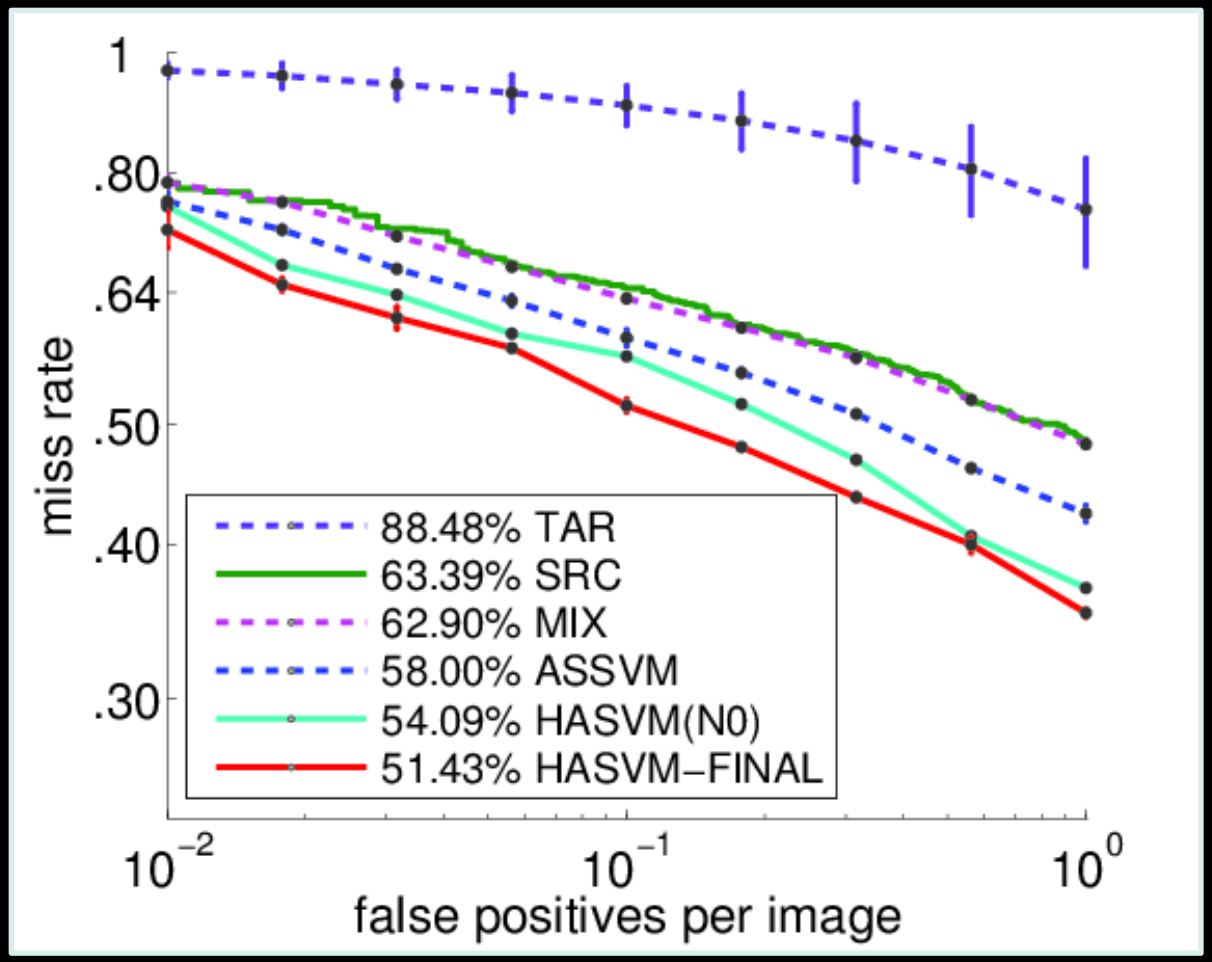
~20 minutes

HA-SSVM experiments

Evaluation on Caltech testing dataset

(Average miss rate %)





Outline ◄

1. Autonomous driving.
2. Pedestrian (Object) detection: self-training.

3. Virtual-world training.
4. Virtual and real world adaptation: cool-world.

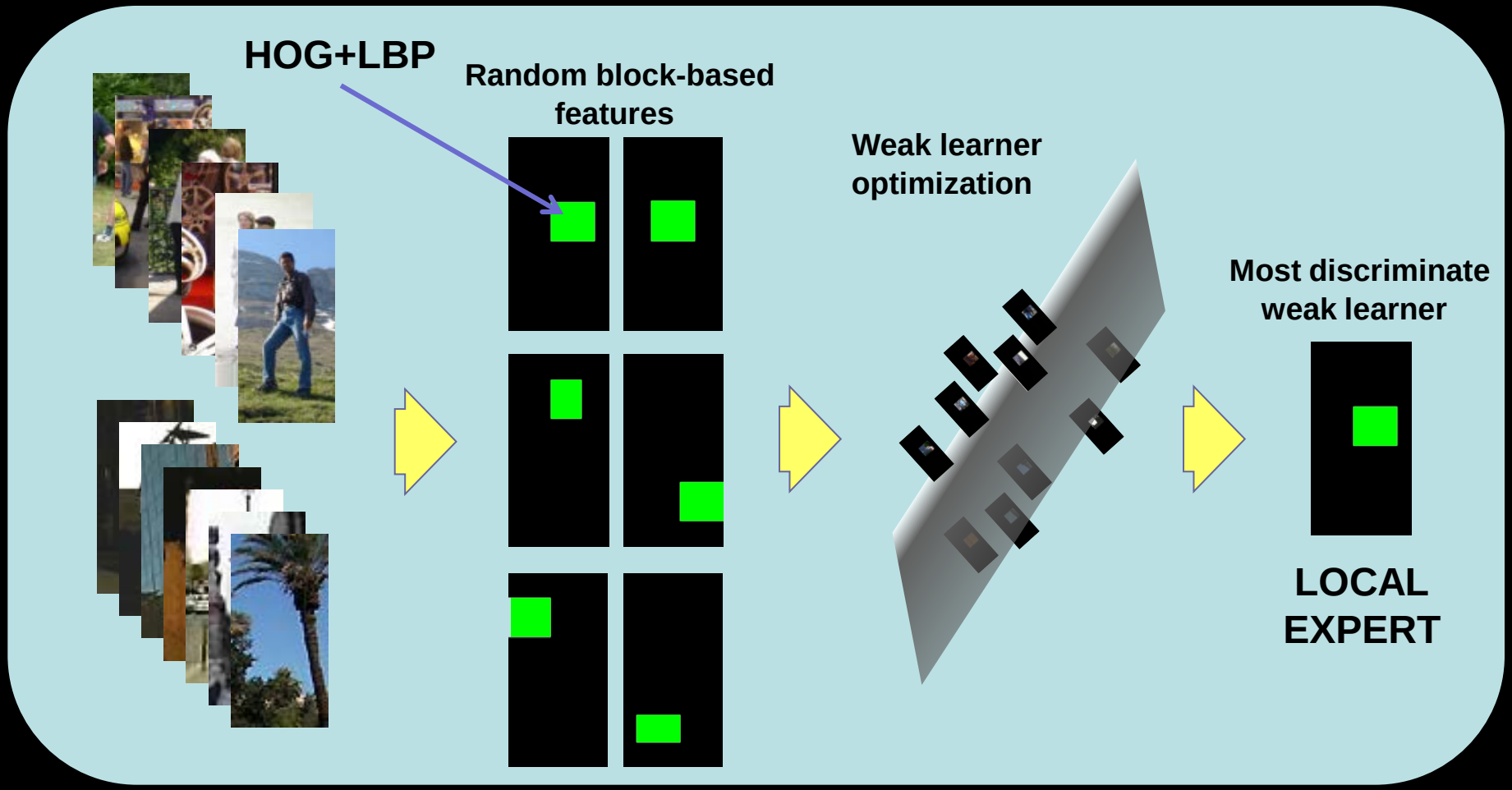
5. Virtual-world DPM (VDPM).
6. Domain adaptation of VDPM.

7. Random Forest of Local Experts: Adaptation.

8. Conclusions.

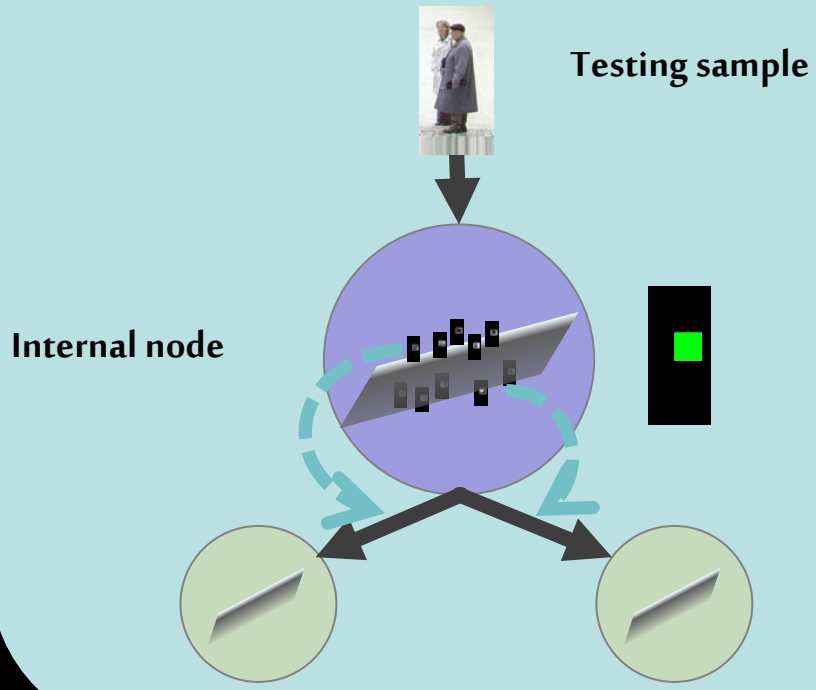
Random Forest of Local Experts

J. Marín, D. Vázquez, A.M. López. J. Amores, B. Leibe. **Random Forest of Local Experts for Pedestrian Detection.** ICCV 2013.

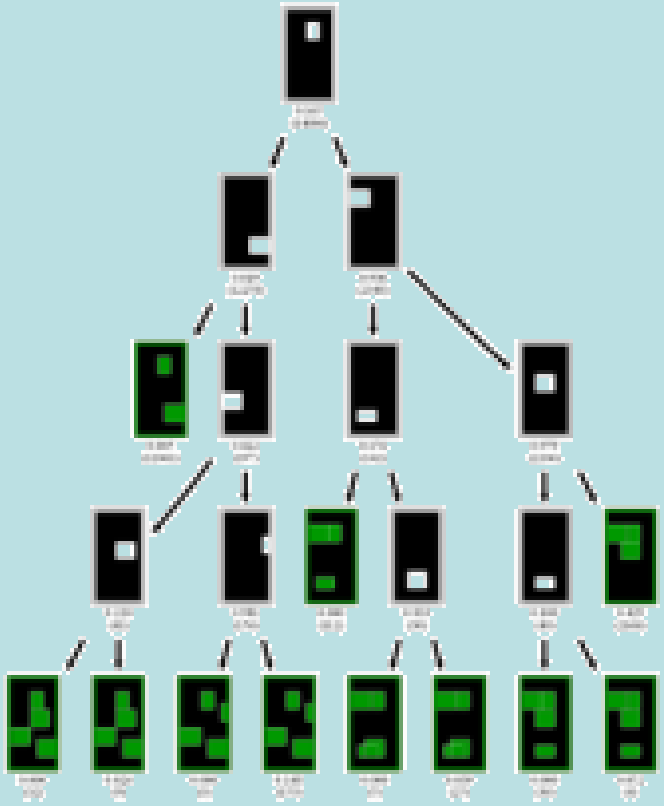


Random Forest of Local Experts

Local expert (weak learner)

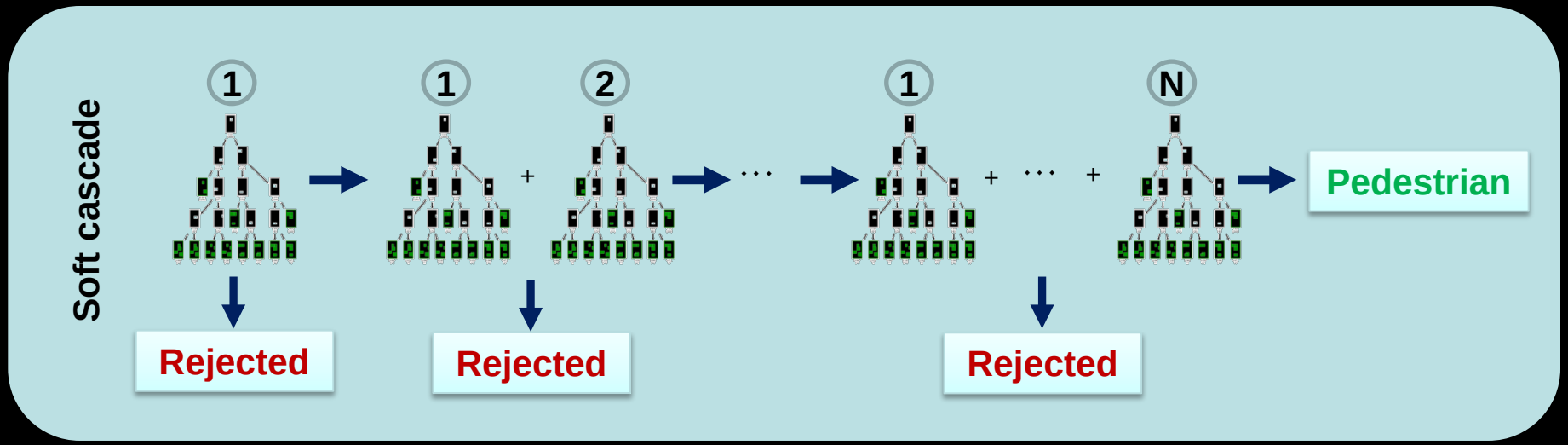


Decision Tree

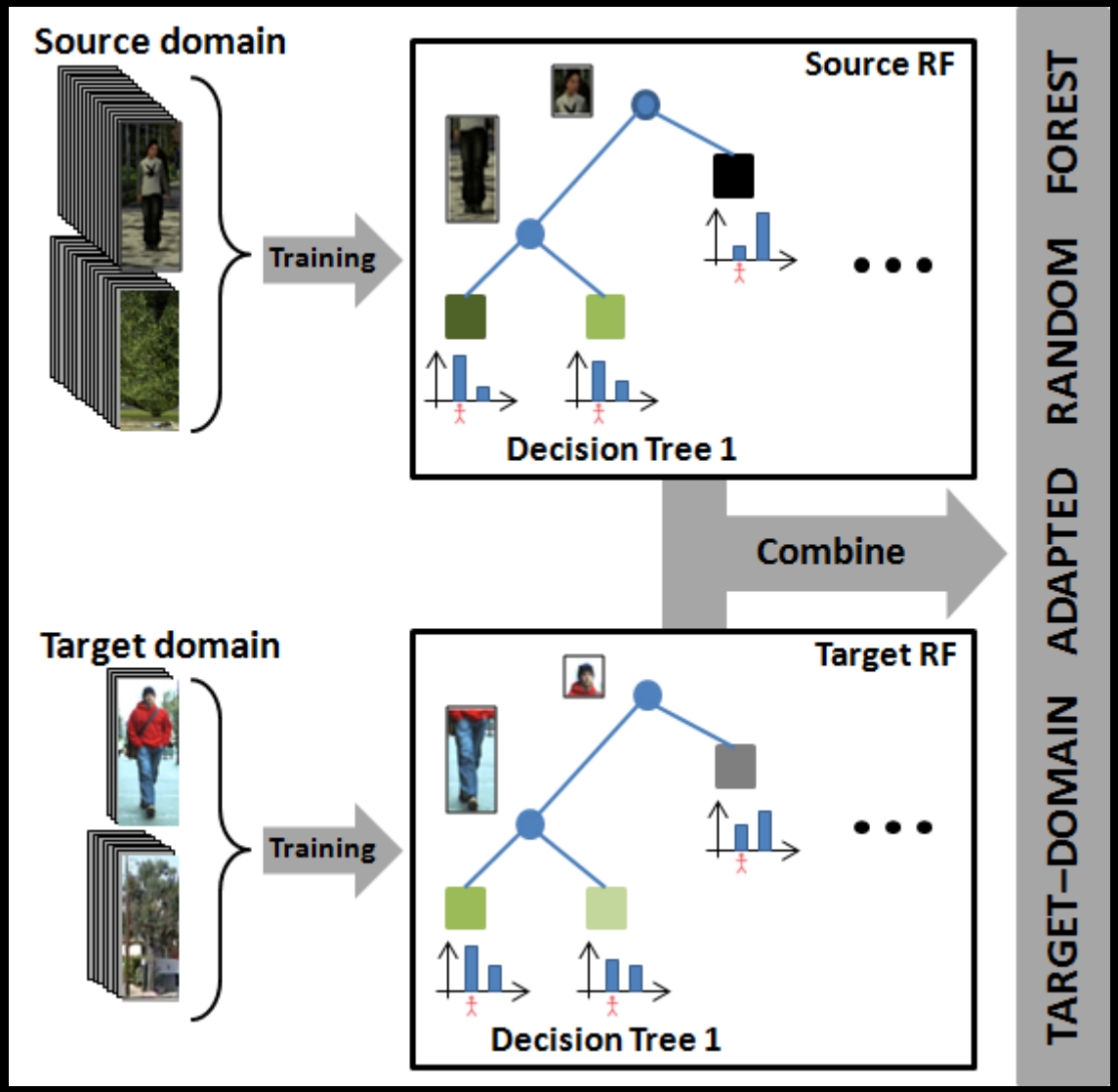


Random Forest of Local Experts ◀

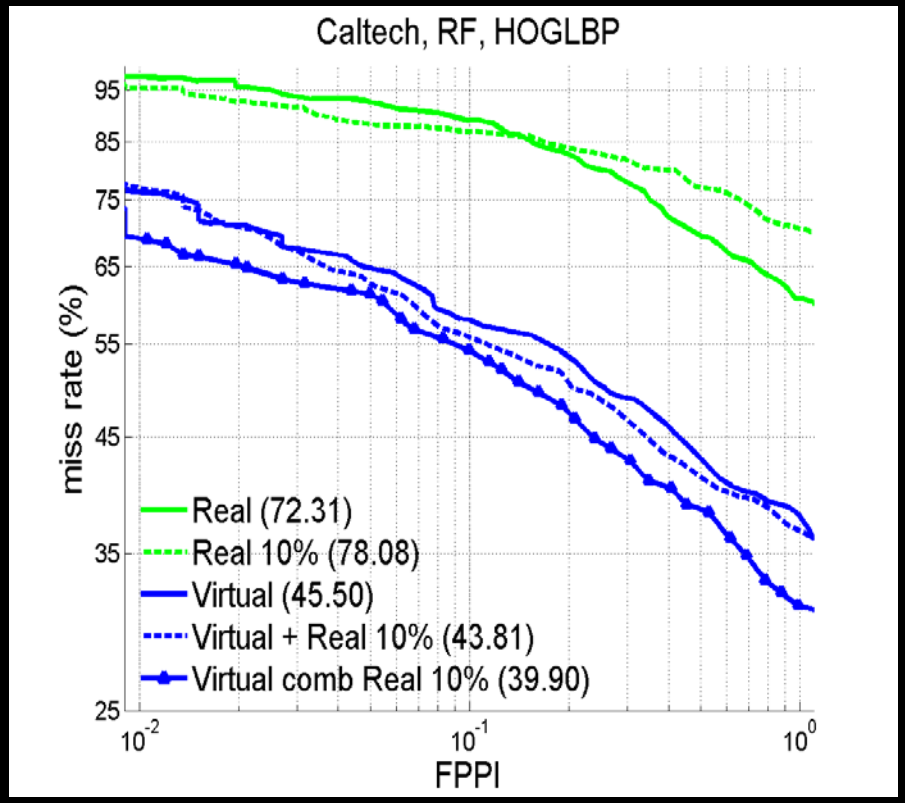
Detection

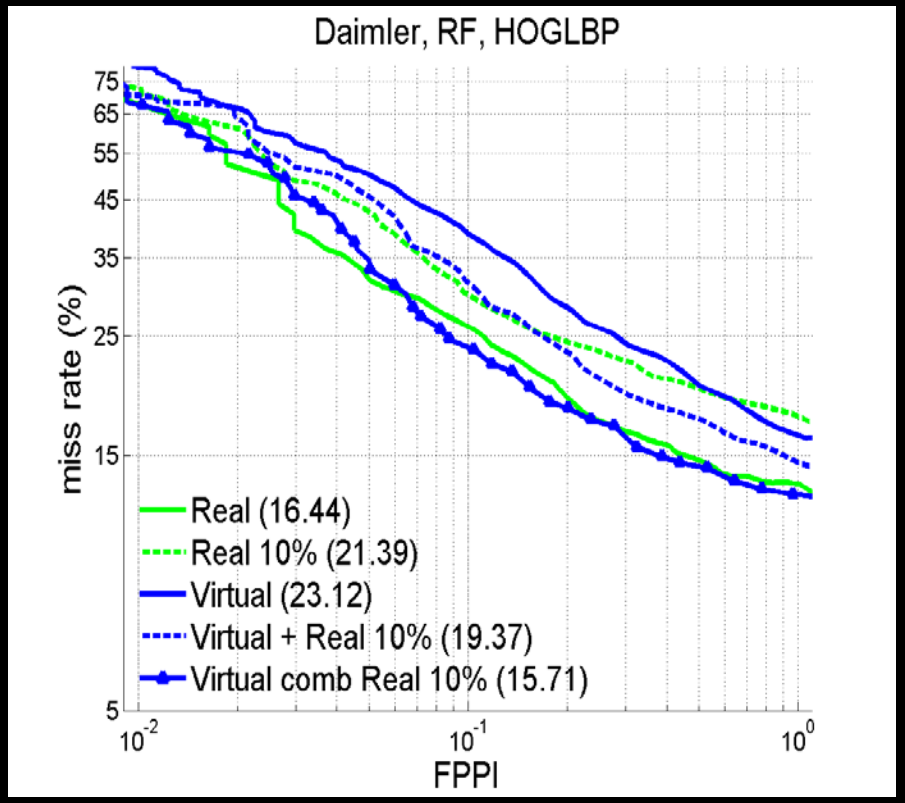


Frustratingly Easy RF of Local Experts ◀



Experiments: Frustratingly Easy RF of Local Experts◀





Outline ◄

1. Autonomous driving.
2. Pedestrian (Object) detection: self-training.

3. Virtual-world training.
4. Virtual and real world adaptation: cool-world.

5. Virtual-world DPM (VDPM).
6. Domain adaptation of VDPM.

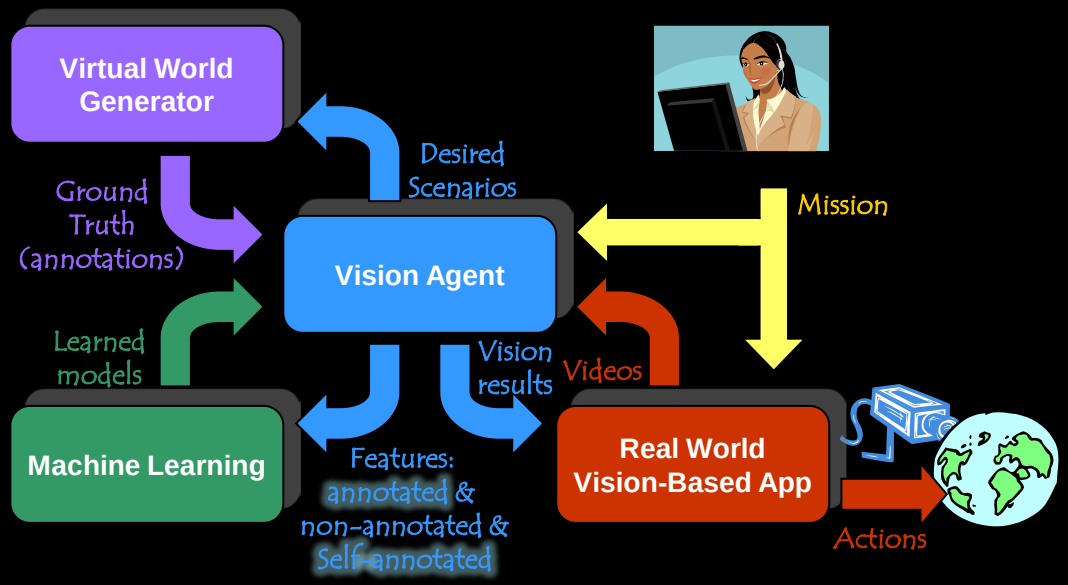
7. Random Forest of Local Experts: Adaptation.

8. Conclusions.

Conclusions

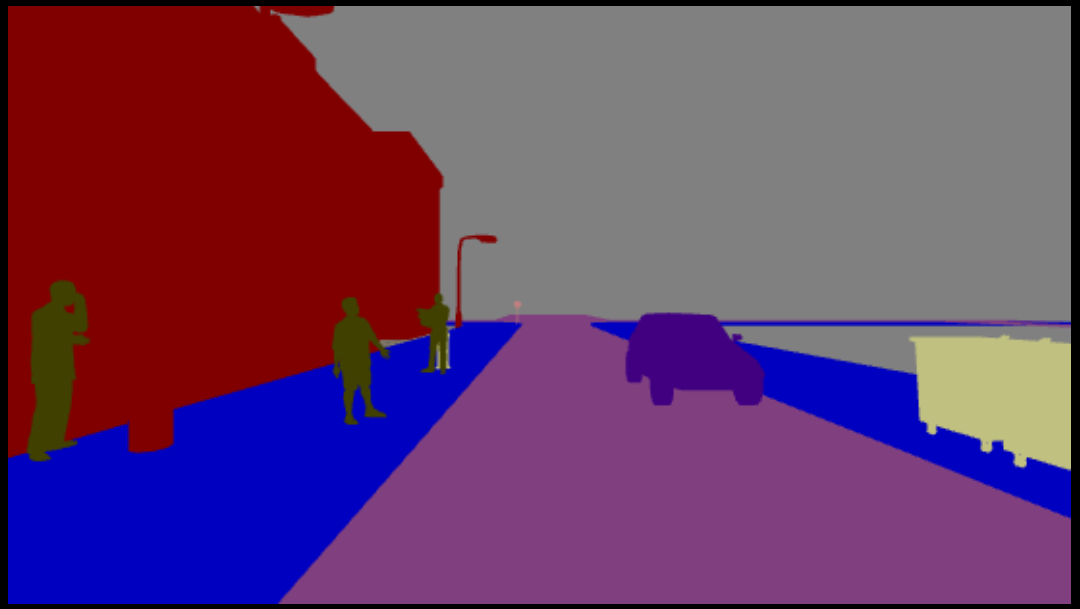
1. Virtual-world data is useful for training detectors that will operate in real-world if domain shift is solved.

2. Our holistic, DPM, and patch-based detectors have been successfully adapted for operating in real-world data by using only a few amount of the available training pedestrians.



3. For more experiments: unsupervised; bilinear classifiers; exemplar LDA, etc., please visit our web page.

Future work



Note

TASK-CV: Transferring and Adapting Source Knowledge in Computer Vision

ECCV 2014, Zürich, Switzerland

Organizers

Antonio M. López, CVC/UAB
Kate Saenko, UMass Lowell
Francesco Orabona, TTI Chicago
José Antonio Rodríguez, XRCE
David Vázquez, CVC
Sebastian Ramos, CVC/UAB
Jiaolong Xu, CVC/UAB

Many Thanks!!!