CiTIUS
Centro Singular de Investigación
en **Tecnoloxías Intelixentes**

FBK
FONDAZIONE
BRUNO KESSLER

UNIVERSITÀ
DI TRENTO

ICCV
OCT 19-23, 2025
HONOLULU
HAWAII

Superpowering Open-Vocabulary Object Detectors for X-ray Vision

International Conference on Computer Vision, ICCV 2025

Oct 19 – 23th, 2025, Honolulu, Hawai'i

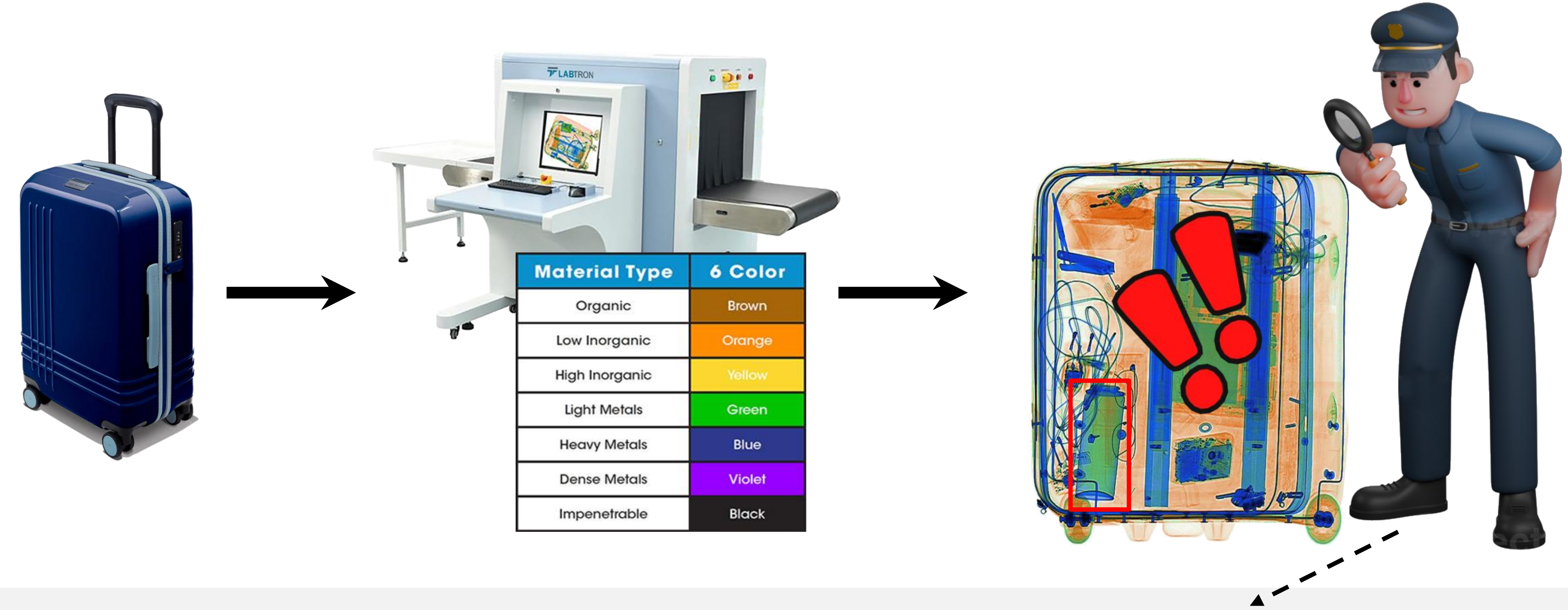
P. Garcia-Fernandez, L. Vaquero, M. Liu, F. Xue,
D. Cores, N. Sebe, M. Mucientes, E. Ricci

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CAMPUS VIDA
CAMPUS DE EXCELENCIA INTERNACIONAL

Motivation – X-ray scanners

❖ X-Ray scanners **are everywhere**. They **map materials** into a color-coded image based on object density

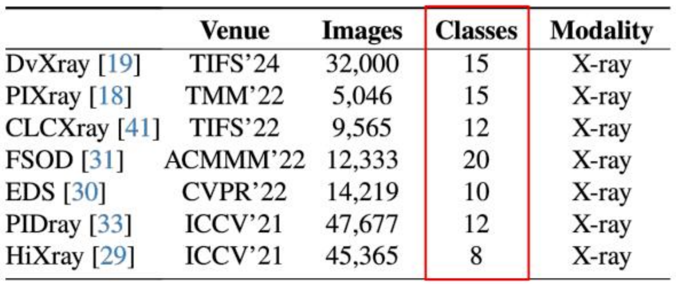


Material Type	6 Color
Organic	Brown
Low Inorganic	Orange
High Inorganic	Yellow
Light Metals	Green
Heavy Metals	Blue
Dense Metals	Violet
Impenetrable	Black

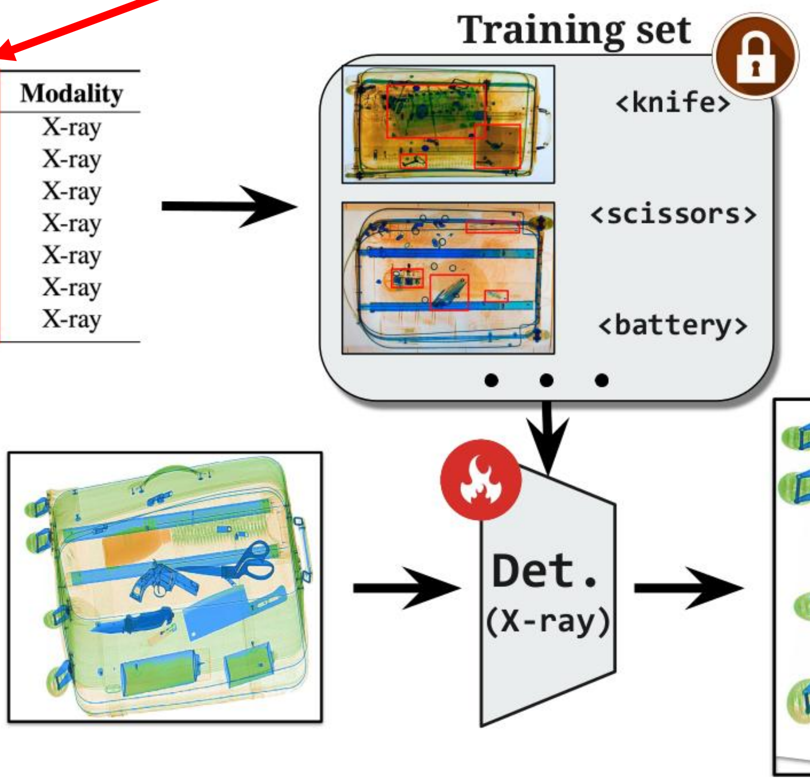
🔎 Continuous **expert** human **oversight** is **required** for **detecting** dangerous objects

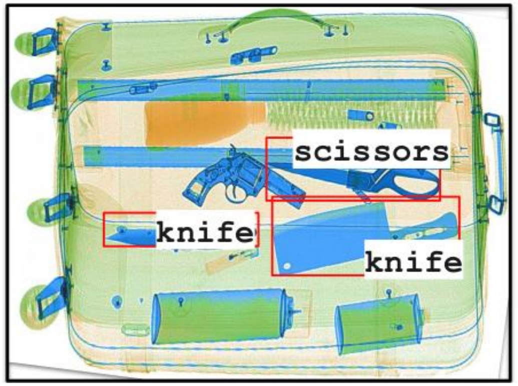
Motivation – X-ray detection

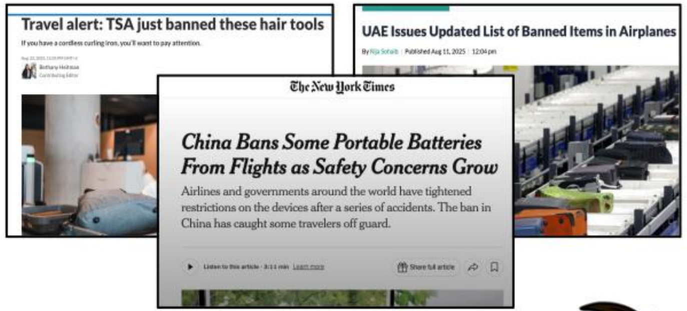
❖ Current X-ray detectors are **limited** by the **categories** in their **training** datasets



	Venue	Images	Classes	Modality
DvXray [19]	TIFS'24	32,000	15	X-ray
PIXray [18]	TMM'22	5,046	15	X-ray
CLCXray [41]	TIFS'22	9,565	12	X-ray
FSOD [31]	ACMMM'22	12,333	20	X-ray
EDS [30]	CVPR'22	14,219	10	X-ray
PIDray [33]	ICCV'21	47,677	12	X-ray
HiXray [29]	ICCV'21	45,365	8	X-ray







Travel alert: TSA just banned these hair tools

If you have a cordless curling iron, you'll want to pay attention.

Bethany Heitman
Contributing Editor

UAE Issues Updated List of Banned Items in Airplanes

By Rija Sohail | Published Aug 11, 2025 | 12:04 pm

The New York Times

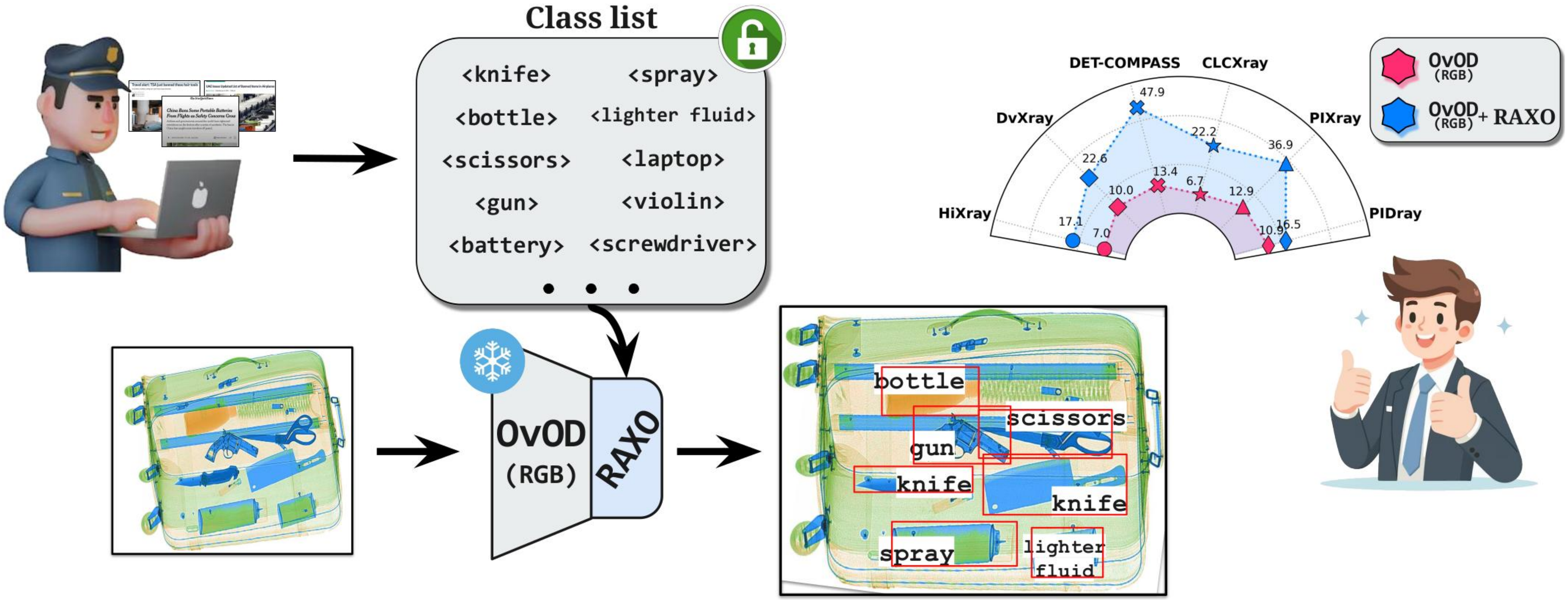
China Bans Some Portable Batteries From Flights as Safety Concerns Grow

Airlines and governments around the world have tightened restrictions on the devices after a series of accidents. The ban in China has caught some travelers off guard.

🔎 Labeled **X-ray data** is **scarce**, making them **unable** to **adapt** to **new classes**

Motivation – Open-Vocabulary X-ray detection

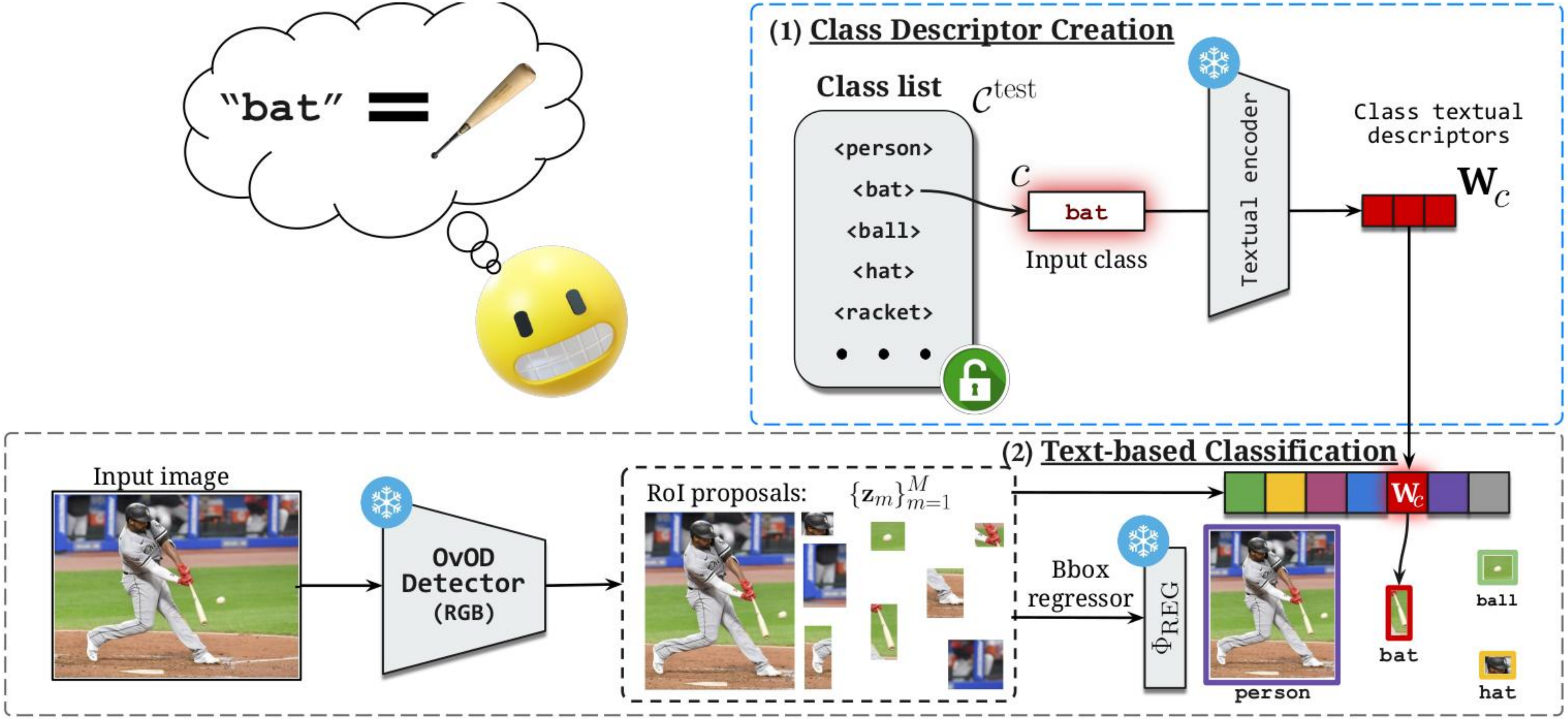
❖ We **introduce** the **task** of **OvOD for X-ray** imaging and **propose RAXO**



👉 **RAXO** is a **training-free** method that **adapts** off-the-shelf **RGB OvOD** models to **X-ray**

Motivation – OvOD in a Nutshell

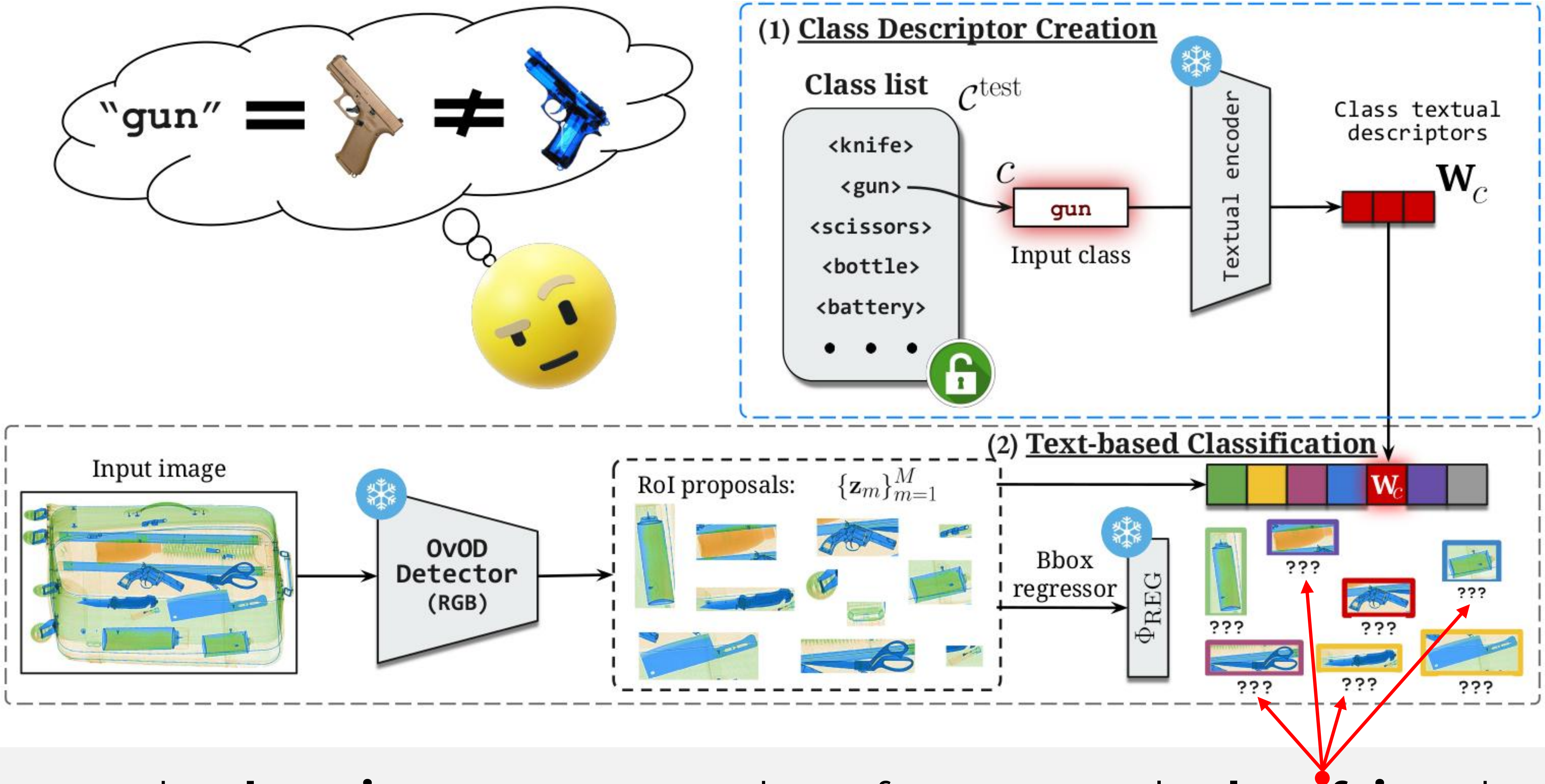
- OvOD detectors first generate candidate **regions of interest**. Then Regions' **visual features** are compared with text embeddings of the classes.



Motivation – OvOD in a Nutshell

Centro Singular de Investigación en **Tecnoloxías Intelixentes**

❖ RGB OvOD detectors can still localize plausible regions on X-ray images

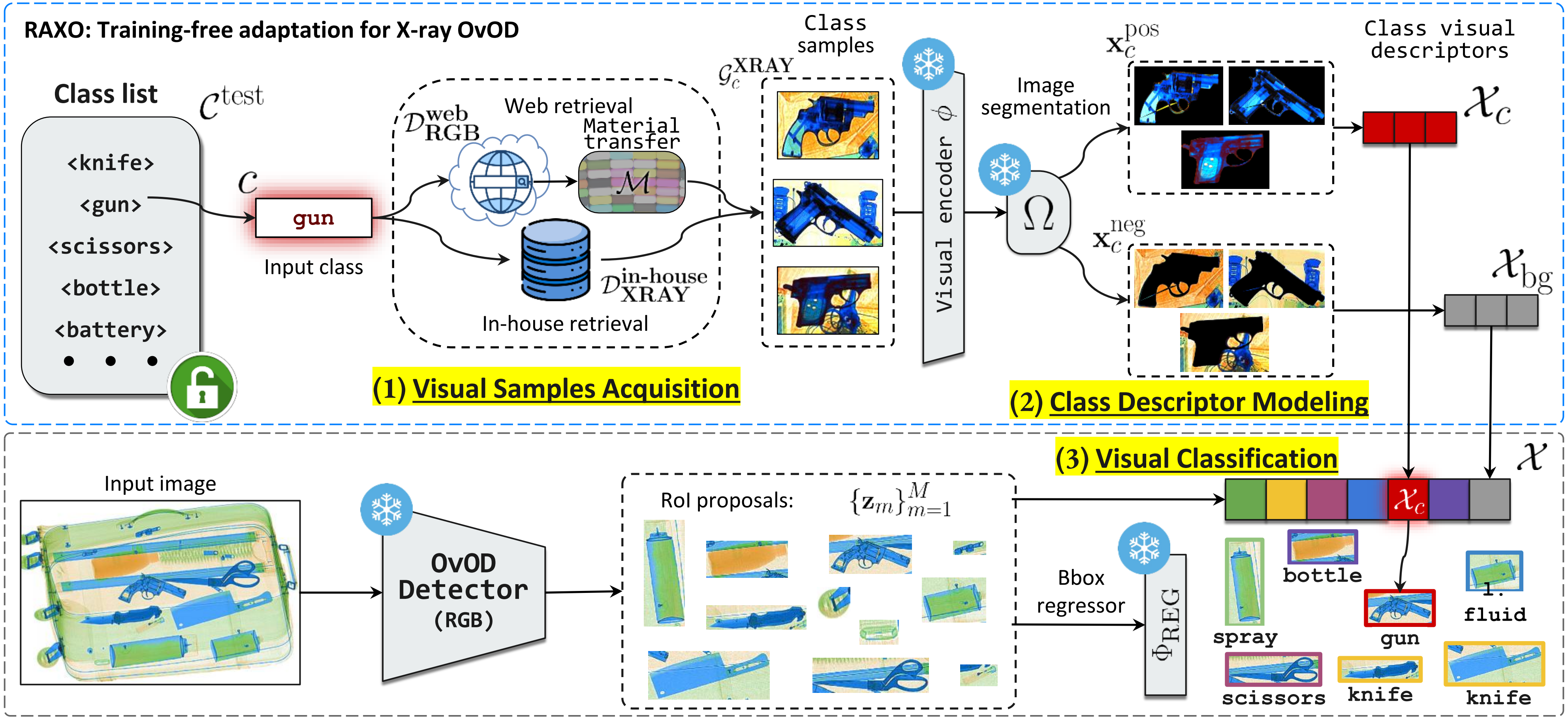


🔎 **However**, the **domain gap prevents** them from correctly **classifying** the objects

RAXO

Centro Singular de Investigación en Tecnoloxías Intelixentes

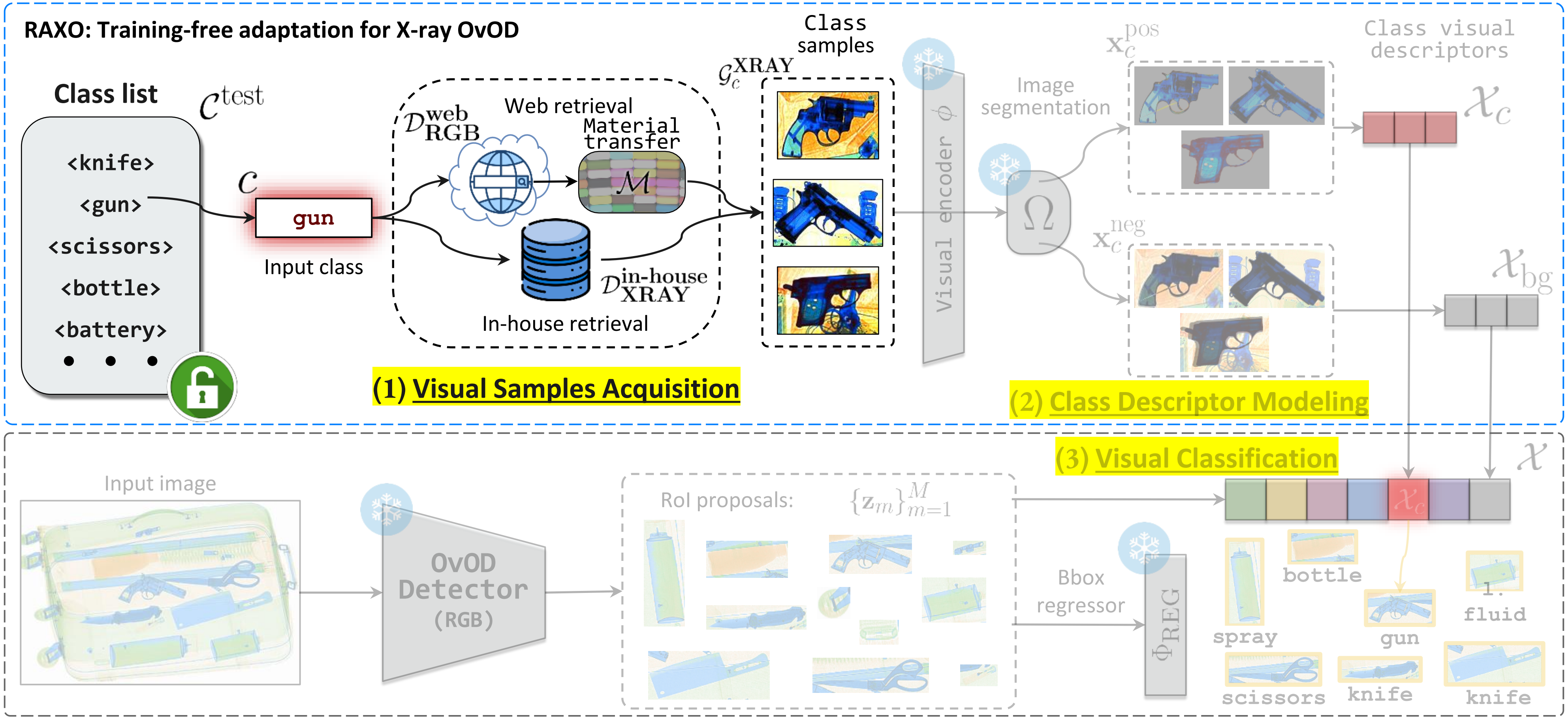
- ❖ RAXO **replaces** the **text-based classifier** in an OvOD with our **visual-based** classifier
- ❖ It is **training-free** and most of it runs **offline** once, introducing **negligible overhead**



RAXO: Training-free adaptation for X-ray OvOD
Class list
$\mathcal{C}^{\text{test}}$
<knife>
<gun>
<scissors>
<bottle>
<battery>
• • •
c
gun
Input class
$\mathcal{D}^{\text{web}}_{\text{RGB}}$
Web retrieval
Material transfer
$\mathcal{M}$
$\mathcal{D}^{\text{in-house}}_{\text{XRAY}}$
In-house retrieval
$\mathcal{G}_c^{\text{XRAY}}$
Class samples
Visual encoder ϕ
Image segmentation
Ω
$\mathbf{x}_c^{\text{pos}}$
$\mathbf{x}_c^{\text{neg}}$
Class visual descriptors
$\mathcal{X}_c$
$\mathcal{X}_{\text{bg}}$
(1) Visual Samples Acquisition
(2) Class Descriptor Modeling
(3) Visual Classification
$\mathcal{X}$
Input image
OvOD Detector (RGB)
RoI proposals: $\{\mathbf{z}_m\}_{m=1}^{M}$
Bbox regressor
Φ_{REG}
$\mathcal{X}_c$
spray
bottle
gun
l. fluid
scissors
knife
knife

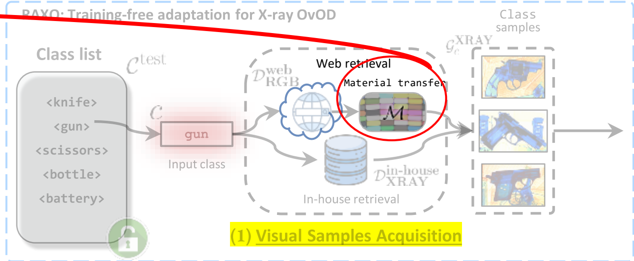
RAXO – Visual Samples Acquisition

- We **build** X-ray **class descriptors** using a **dual-source** (web&in-house) retrieval strategy. **In-house** data is already **X-ray** and **web images** are **adapted** with a **material transfer**

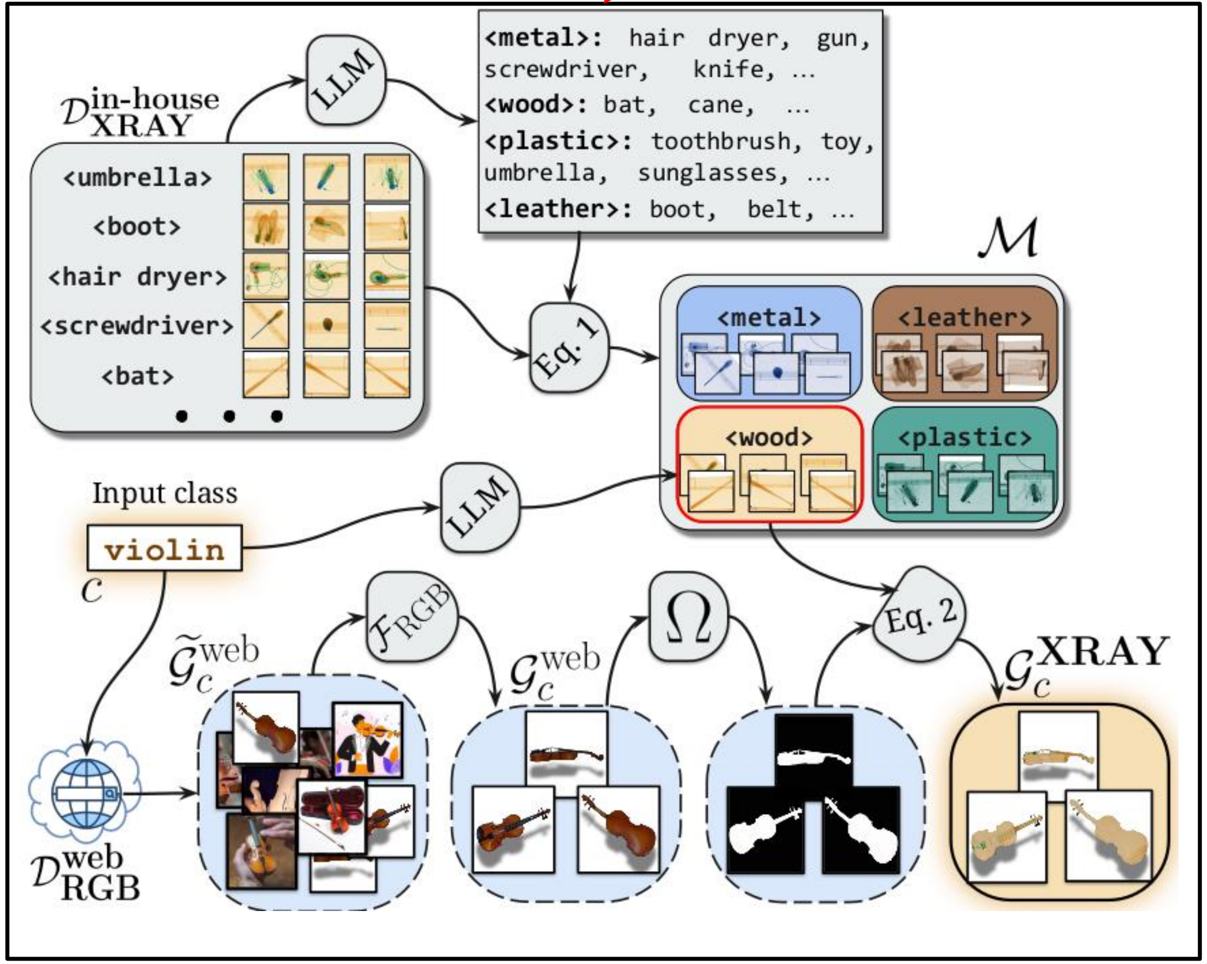

RAXO: Training-free adaptation for X-ray OvOD
Class list
<knife>
<gun>
<scissors>
<bottle>
<battery>
$\mathcal{C}^{\text{test}}$
c
gun
Input class
$\mathcal{D}_{\text{RGB}}^{\text{web}}$
Web retrieval
Material transfer
$\mathcal{M}$
$\mathcal{D}_{\text{XRAY}}^{\text{in-house}}$
In-house retrieval
$\mathcal{G}_c^{\text{XRAY}}$
Class samples
(1) Visual Samples Acquisition
Visual encoder ϕ
Image segmentation
Ω
$\mathbf{x}_c^{\text{pos}}$
$\mathbf{x}_c^{\text{neg}}$
Class visual descriptors
$\mathcal{X}_c$
$\mathcal{X}_{\text{bg}}$
(2) Class Descriptor Modeling
(3) Visual Classification
$\mathcal{X}$
$\mathcal{X}_c$
Input image
OvOD Detector (RGB)
RoI proposals: $\{\mathbf{z}_m\}_{m=1}^{M}$
Bbox regressor
Φ_{REG}
spray
bottle
gun
l. fluid
scissors
knife
knife

RAXO – Material Transfer Mechanism

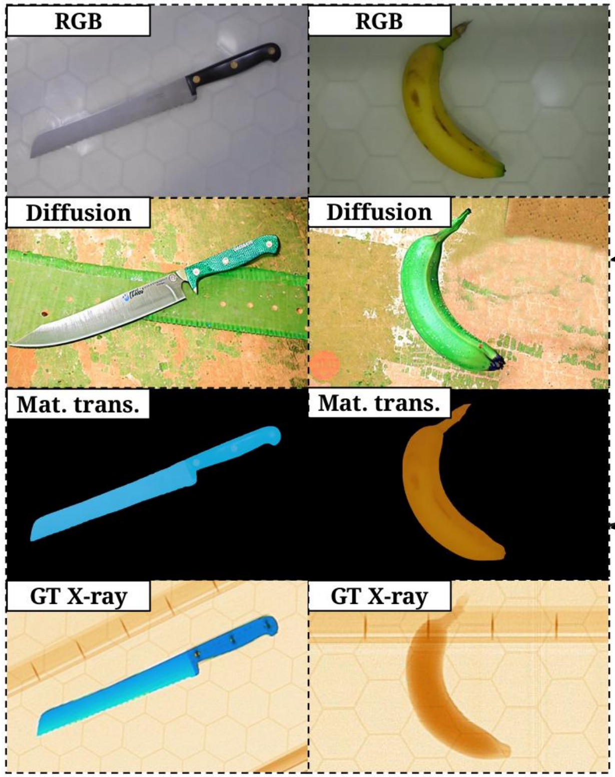
❖ Build a **material database *M*** by **clustering** in-house objects based on **materials**



Material Transfer



<metal>: hair dryer, gun, screwdriver, knife, …
<wood>: bat, cane, …
<plastic>: toothbrush, toy, umbrella, sunglasses, …
<leather>: boot, belt, …

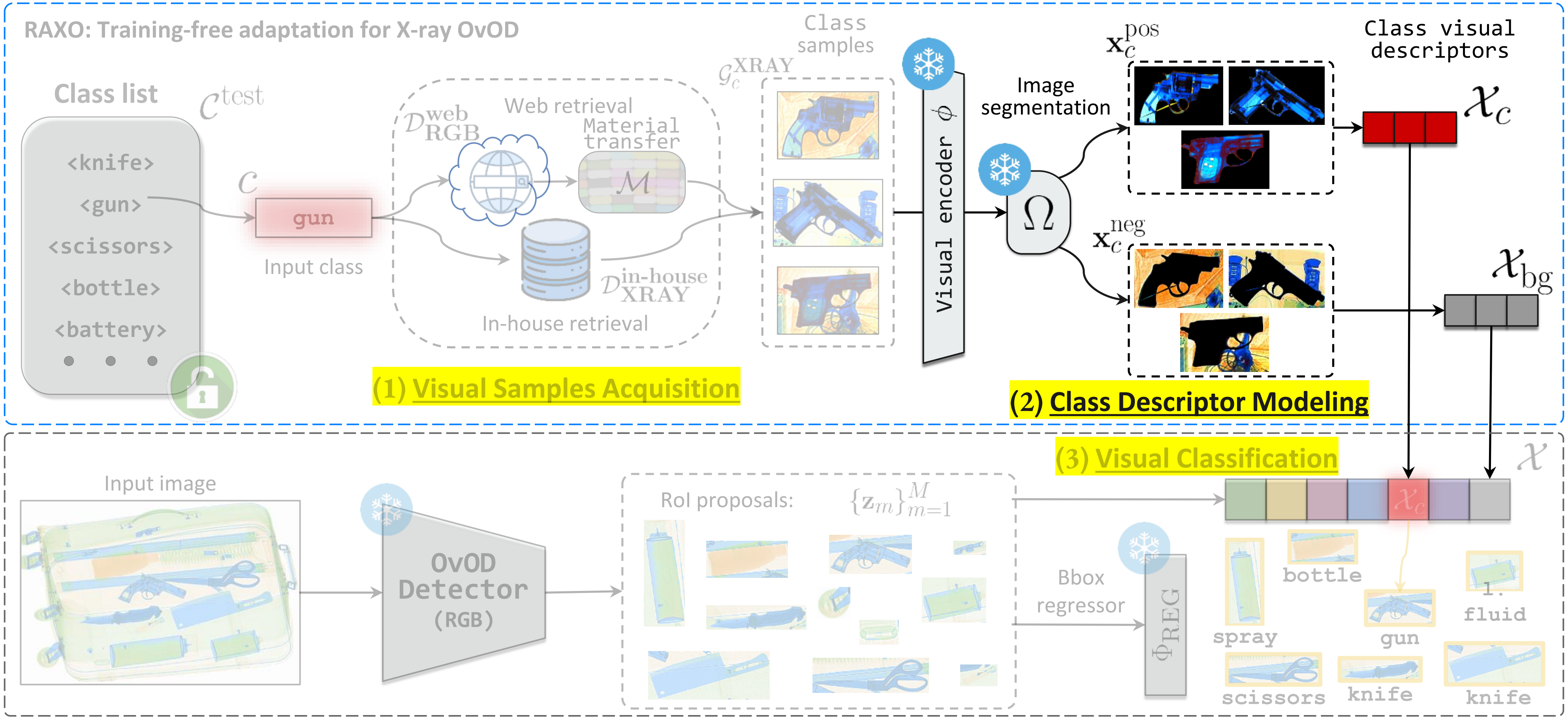


Style transfer (StyleShot) does **not** work!

Our material transfer **works!**

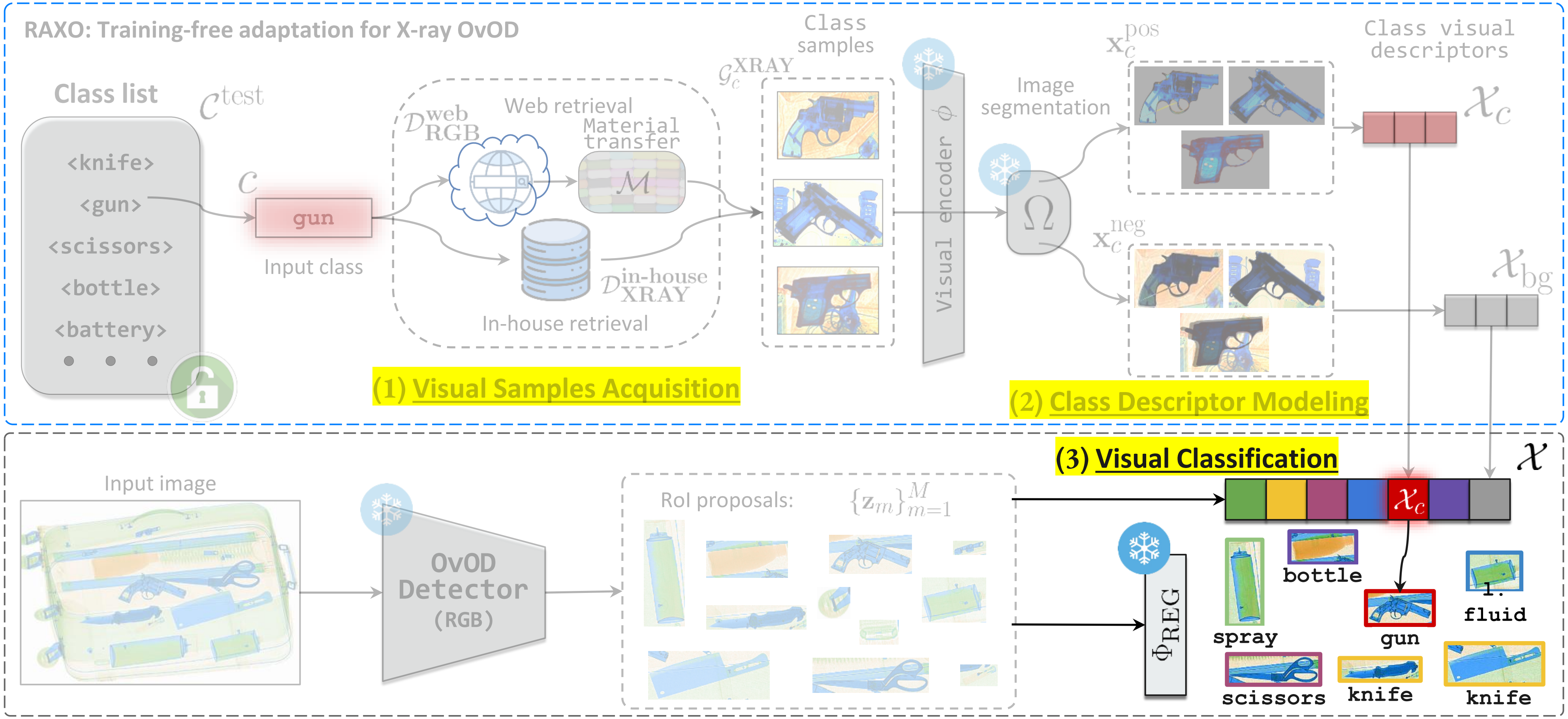
RAXO – Class Descriptor Modeling

- For each sample, extract **positive** and **negative** features using **segmentation** masks
- Samples are **concatenated** into class descriptor, **maintaining intra-class variability**



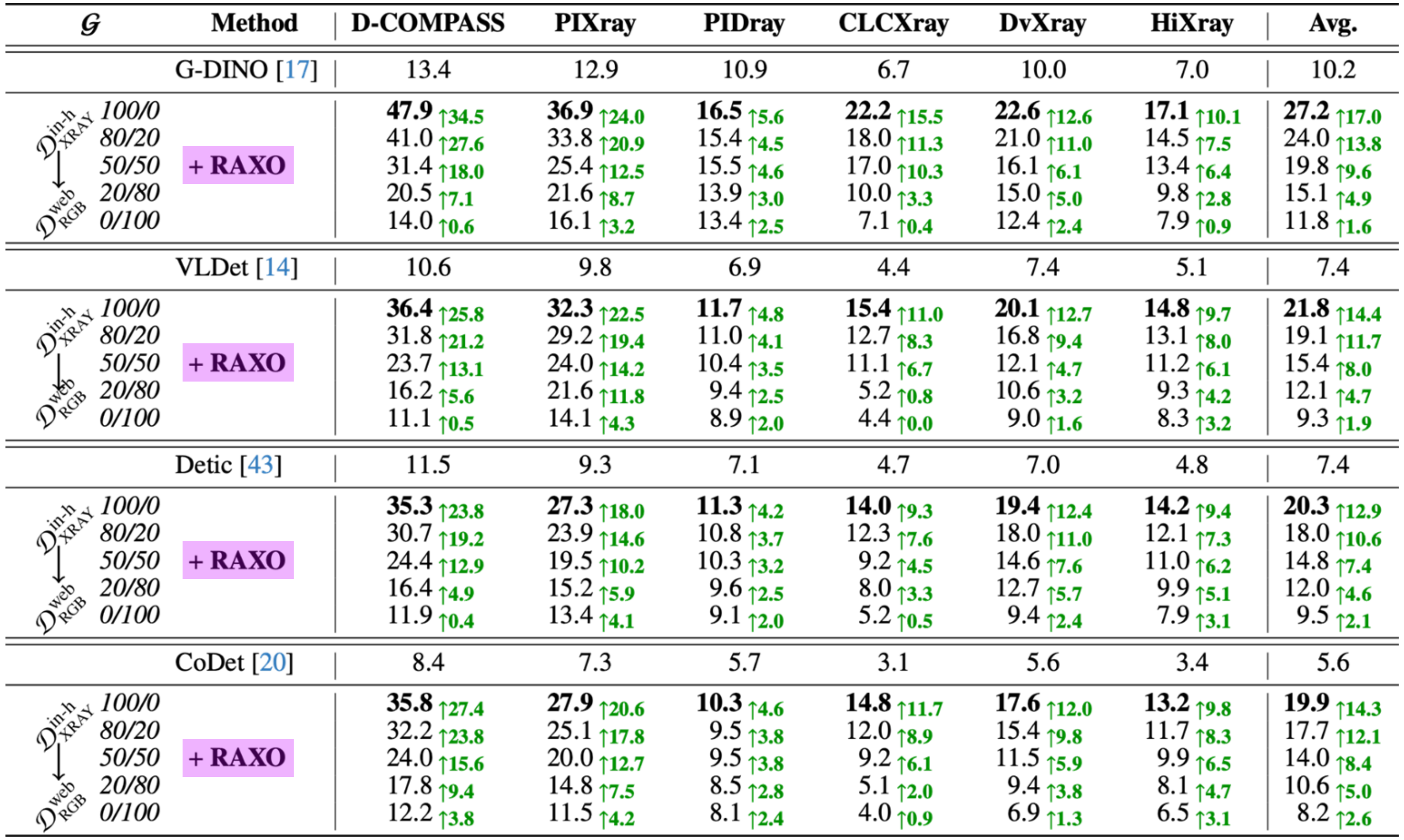
RAXO – Visual Classification

- For each sample, extract **positive** and **negative** features using **segmentation** masks
- Samples are **concatenated** into class descriptor, **maintaining intra-class variability**

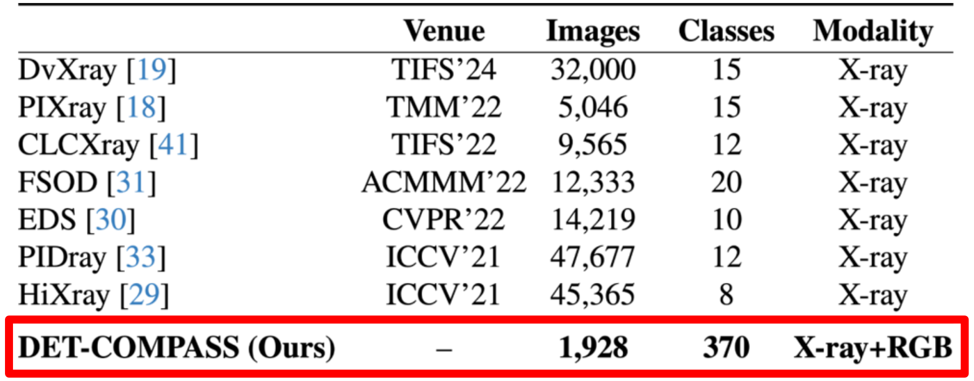


Main Results and Our DET-COMPASS Dataset

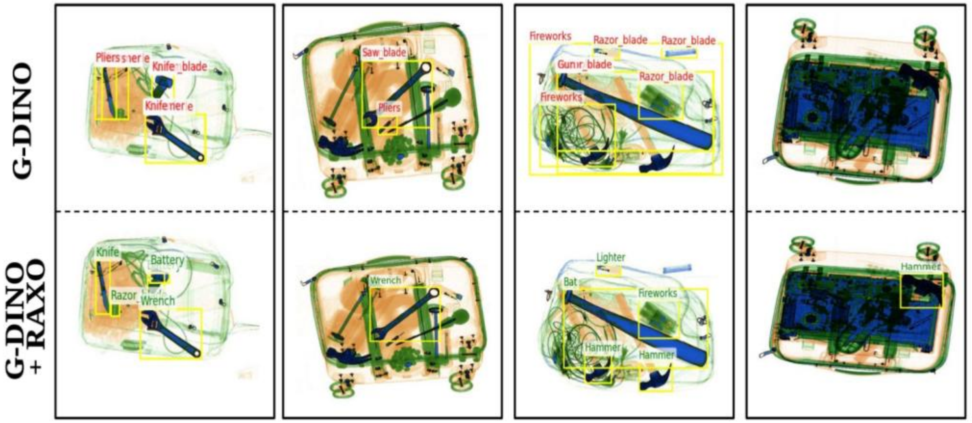
- **RAXO improves** 4 RGB OvOD models across 6 benchmarks, **without training**
- We also **introduce DET-COMPASS** the **most diverse X-ray detection** benchmark



$\mathcal{G}$	Method	D-COMPASS	PIXray	PIDray	CLCXray	DvXray	HiXray	Avg.
	G-DINO [17]	13.4	12.9	10.9	6.7	10.0	7.0	10.2
$\mathcal{D}^{\text{in-h}}_{\text{XRAY}} \downarrow \mathcal{D}^{\text{web}}_{\text{RGB}}$ 100/0	+ RAXO	**47.9** ↑34.5	**36.9** ↑24.0	**16.5** ↑5.6	**22.2** ↑15.5	**22.6** ↑12.6	**17.1** ↑10.1	**27.2** ↑17.0
80/20		41.0 ↑27.6	33.8 ↑20.9	15.4 ↑4.5	18.0 ↑11.3	21.0 ↑11.0	14.5 ↑7.5	24.0 ↑13.8
50/50		31.4 ↑18.0	25.4 ↑12.5	15.5 ↑4.6	17.0 ↑10.3	16.1 ↑6.1	13.4 ↑6.4	19.8 ↑9.6
20/80		20.5 ↑7.1	21.6 ↑8.7	13.9 ↑3.0	10.0 ↑3.3	15.0 ↑5.0	9.8 ↑2.8	15.1 ↑4.9
0/100		14.0 ↑0.6	16.1 ↑3.2	13.4 ↑2.5	7.1 ↑0.4	12.4 ↑2.4	7.9 ↑0.9	11.8 ↑1.6
	VLDet [14]	10.6	9.8	6.9	4.4	7.4	5.1	7.4
$\mathcal{D}^{\text{in-h}}_{\text{XRAY}} \downarrow \mathcal{D}^{\text{web}}_{\text{RGB}}$ 100/0	+ RAXO	**36.4** ↑25.8	**32.3** ↑22.5	**11.7** ↑4.8	**15.4** ↑11.0	**20.1** ↑12.7	**14.8** ↑9.7	**21.8** ↑14.4
80/20		31.8 ↑21.2	29.2 ↑19.4	11.0 ↑4.1	12.7 ↑8.3	16.8 ↑9.4	13.1 ↑8.0	19.1 ↑11.7
50/50		23.7 ↑13.1	24.0 ↑14.2	10.4 ↑3.5	11.1 ↑6.7	12.1 ↑4.7	11.2 ↑6.1	15.4 ↑8.0
20/80		16.2 ↑5.6	21.6 ↑11.8	9.4 ↑2.5	5.2 ↑0.8	10.6 ↑3.2	9.3 ↑4.2	12.1 ↑4.7
0/100		11.1 ↑0.5	14.1 ↑4.3	8.9 ↑2.0	4.4 ↑0.0	9.0 ↑1.6	8.3 ↑3.2	9.3 ↑1.9
	Detic [43]	11.5	9.3	7.1	4.7	7.0	4.8	7.4
$\mathcal{D}^{\text{in-h}}_{\text{XRAY}} \downarrow \mathcal{D}^{\text{web}}_{\text{RGB}}$ 100/0	+ RAXO	**35.3** ↑23.8	**27.3** ↑18.0	**11.3** ↑4.2	**14.0** ↑9.3	**19.4** ↑12.4	**14.2** ↑9.4	**20.3** ↑12.9
80/20		30.7 ↑19.2	23.9 ↑14.6	10.8 ↑3.7	12.3 ↑7.6	18.0 ↑11.0	12.1 ↑7.3	18.0 ↑10.6
50/50		24.4 ↑12.9	19.5 ↑10.2	10.3 ↑3.2	9.2 ↑4.5	14.6 ↑7.6	11.0 ↑6.2	14.8 ↑7.4
20/80		16.4 ↑4.9	15.2 ↑5.9	9.6 ↑2.5	8.0 ↑3.3	12.7 ↑5.7	9.9 ↑5.1	12.0 ↑4.6
0/100		11.9 ↑0.4	13.4 ↑4.1	9.1 ↑2.0	5.2 ↑0.5	9.4 ↑2.4	7.9 ↑3.1	9.5 ↑2.1
	CoDet [20]	8.4	7.3	5.7	3.1	5.6	3.4	5.6
$\mathcal{D}^{\text{in-h}}_{\text{XRAY}} \downarrow \mathcal{D}^{\text{web}}_{\text{RGB}}$ 100/0	+ RAXO	**35.8** ↑27.4	**27.9** ↑20.6	**10.3** ↑4.6	**14.8** ↑11.7	**17.6** ↑12.0	**13.2** ↑9.8	**19.9** ↑14.3
80/20		32.2 ↑23.8	25.1 ↑17.8	9.5 ↑3.8	12.0 ↑8.9	15.4 ↑9.8	11.7 ↑8.3	17.7 ↑12.1
50/50		24.0 ↑15.6	20.0 ↑12.7	9.5 ↑3.8	9.2 ↑6.1	11.5 ↑5.9	9.9 ↑6.5	14.0 ↑8.4
20/80		17.8 ↑9.4	14.8 ↑7.5	8.5 ↑2.8	5.1 ↑2.0	9.4 ↑3.8	8.1 ↑4.7	10.6 ↑5.0
0/100		12.2 ↑3.8	11.5 ↑4.2	8.1 ↑2.4	4.0 ↑0.9	6.9 ↑1.3	6.5 ↑3.1	8.2 ↑2.6



	Venue	Images	Classes	Modality
DvXray [19]	TIFS'24	32,000	15	X-ray
PIXray [18]	TMM'22	5,046	15	X-ray
CLCXray [41]	TIFS'22	9,565	12	X-ray
FSOD [31]	ACMMM'22	12,333	20	X-ray
EDS [30]	CVPR'22	14,219	10	X-ray
PIDray [33]	ICCV'21	47,677	12	X-ray
HiXray [29]	ICCV'21	45,365	8	X-ray
DET-COMPASS (Ours)	–	**1,928**	**370**	**X-ray+RGB**



Any questions?

See you on October 23, at Session 5

POSTER id: 1914

Superpowering Open-Vocabulary Object Detectors for X-ray Vision

P. Garcia-Fernandez, L. Vaquero, M. Liu, F. Xue, D. Cores, N. Sebe, M. Mucientes, E. Ricci

CiTIUS
Centro Singular de Investigación
en Tecnoloxías Intelixentes

FBK
FONDAZIONE
BRUNO KESSLER

UNIVERSITÀ
DI TRENTO

ICCV
OCT 19-23, 2025
HONOLULU
HAWAII