Modality Translation for Object Detection Adaptation Without Forgetting Prior Knowledge

Heitor R. Medeiros, Masih A., Fidel A. G. Peña,
David Latortue, Eric Granger, Marco Pedersoli
LIVIA, Dept. of Systems Engineering, ETS Montreal, Canada

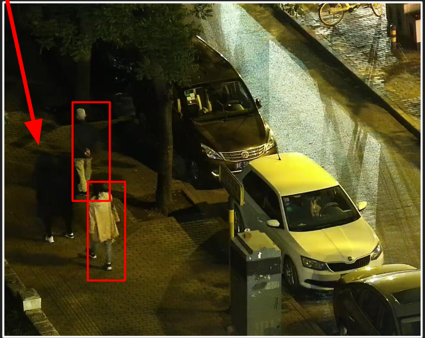
1. Background and Motivation

Object Detection RGB and IR

- RGB and IR can contain complementary information that can be used to improve object detection.

IR sensors are better than RGB for people detection in low-light conditions.

Infrared (IR)

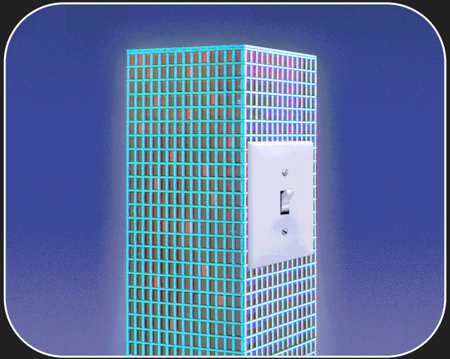


Visible (RGB)

[1] Jia, Xinyu, et al. "LLVIP: A Visible-infrared Paired Dataset for Low-light Vision." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

Self-Driving Cars

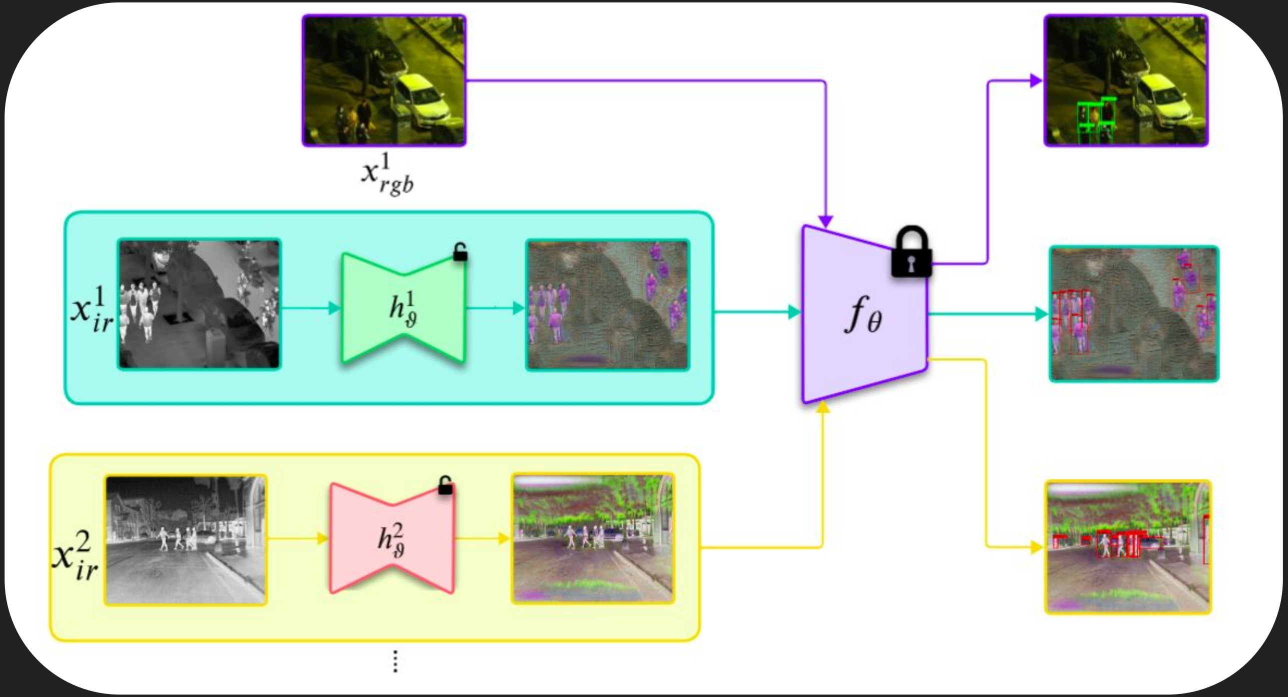
Smart Buildings



[1] Qingyun, Fang, Han Dapeng, and Wang Zhaokui. "Cross-Modality Fusion Transformer for Multispectral Object Detection." arXiv preprint arXiv:2111.00273 (2021).
[2] ADVIDS. "20 Smart and Intelligent building solutions Video Marketing Examples", accessed 21 March 2022, https://blog.advids.co/20-smart-and-intelligent-building-solutions-video-marketing-examples/.
[3] AXIOS, Illustration: Annelise Capossela/Axios, accessed 21 March 2022. https://www.axios.com/coronavirus-smart-city-stalled-projects-852731df-072f-45bf-8218-2f5def57c8d4.html.

2. ModTr

Proposed Model: ModTr



ModTr Loss

$$\mathcal{L}_{det}(\theta) = \frac{1}{|\mathcal{D}|} \sum_{(x,\mathbf{y}) \in \mathcal{D}} \mathcal{L}_{det}[f_\theta(x), \mathcal{Y}].$$

$$\mathcal{L}_{\text{ModTr}}(x, \mathcal{Y}; \vartheta) = \mathcal{L}_{det}\left[f_\theta\left(\Phi(h_\vartheta^d(x), x)\right), \mathcal{Y}\right]$$

3. Results

Comparison with Translation Approaches

Image translation	RGB	Box	Test Set IR (Dataset: LLVIP)		
			FCOS	RetinaNet	Faster R-CNN
Histogram Equal. [15]			31.69 ± 0.00	33.16 ± 0.00	38.33 ± 0.02
CycleGAN [53]	✓		23.85 ± 0.76	23.34 ± 0.53	26.54 ± 1.20
CUT [39]	✓		14.30 ± 2.25	13.12 ± 2.07	14.78 ± 1.82
FastCUT [39]	✓		19.39 ± 1.52	18.11 ± 0.79	22.91 ± 1.68
HalluciDet [31]	✓	✓	28.00 ± 0.92	19.95 ± 2.01	57.78 ± 0.97
$\text{ModTr}_{\odot}$ (ours)		✓	**57.63 ± 0.66**	**54.83 ± 0.61**	**57.97 ± 0.85**
Image translation	**RGB**	**Box**	**Test Set IR (Dataset: FLIR)**		
			FCOS	**RetinaNet**	**Faster R-CNN**
Histogram Equal. [15]			22.76 ± 0.00	23.06 ± 0.00	24.61 ± 0.01
CycleGAN [53]	✓		23.92 ± 0.97	23.71 ± 0.70	26.85 ± 1.23
CUT [39]	✓		18.16 ± 0.75	17.84 ± 0.75	20.29 ± 0.48
FastCUT [39]	✓		24.02 ± 2.37	22.00 ± 2.73	26.68 ± 2.59
HalluciDet [31]	✓	✓	23.74 ± 2.09	22.29 ± 0.45	29.91 ± 1.18
$\text{ModTr}_{\odot}$ (ours)		✓	**35.49 ± 0.94**	**34.27 ± 0.27**	**37.21 ± 0.46**

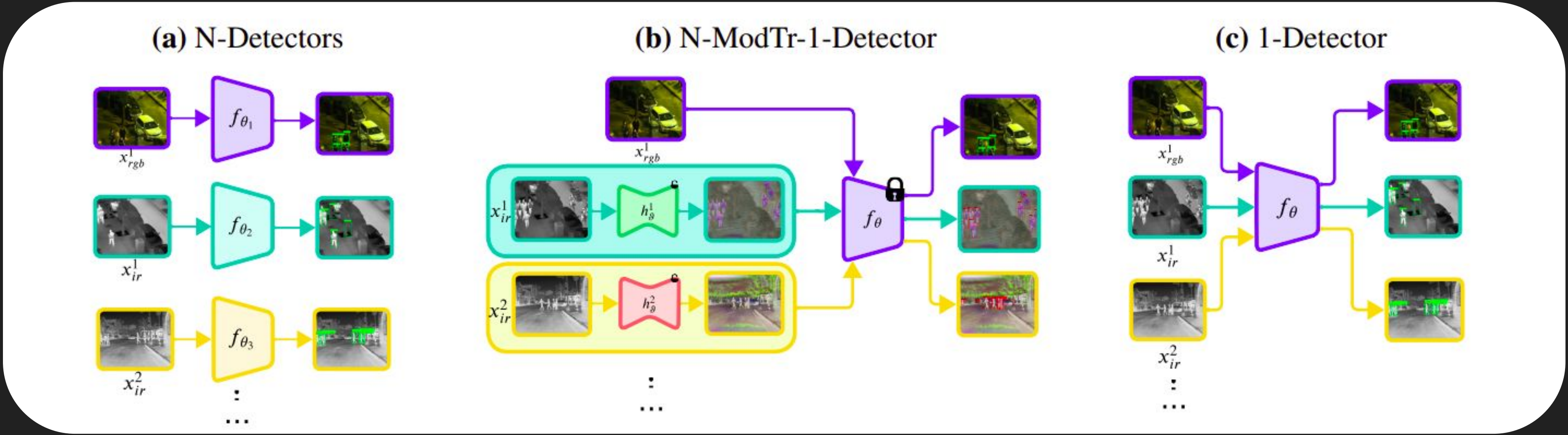
Table 1. IR object detection AP performance with different translation methods.

Translation vs. Fine-tuning

Method	Test Set IR (Dataset: LLVIP)		
	FCOS	**RetinaNet**	**Faster R-CNN**
Fine-Tuning (FT)	57.37 ± 2.19	53.79 ± 1.79	**59.62 ± 1.23**
FT Head	49.11 ± 0.70	44.00 ± 0.28	59.33 ± 2.17
LoRA [19]	47.72 ± 0.58	-	54.83 ± 1.30
$\text{ModTr}_{\odot}$ (ours)	**57.63 ± 0.66**	**54.83 ± 0.61**	57.97 ± 0.85
Method	**Test Set IR (Dataset: FLIR)**		
	FCOS	**RetinaNet**	**Faster R-CNN**
Fine-Tuning (FT)	27.97 ± 0.59	28.46 ± 0.50	30.93 ± 0.46
FT Head	27.40 ± 0.12	26.78 ± 0.70	33.53 ± 0.36
LoRA [19]	-	-	29.44 ± 0.61
$\text{ModTr}_{\odot}$ (ours)	**35.49 ± 0.94**	**34.27 ± 0.27**	**37.21 ± 0.46**

Table 2. AP performance benchmark for different OD fine-tuning strategies.

Knowledge Preservation through Input Modality Translation

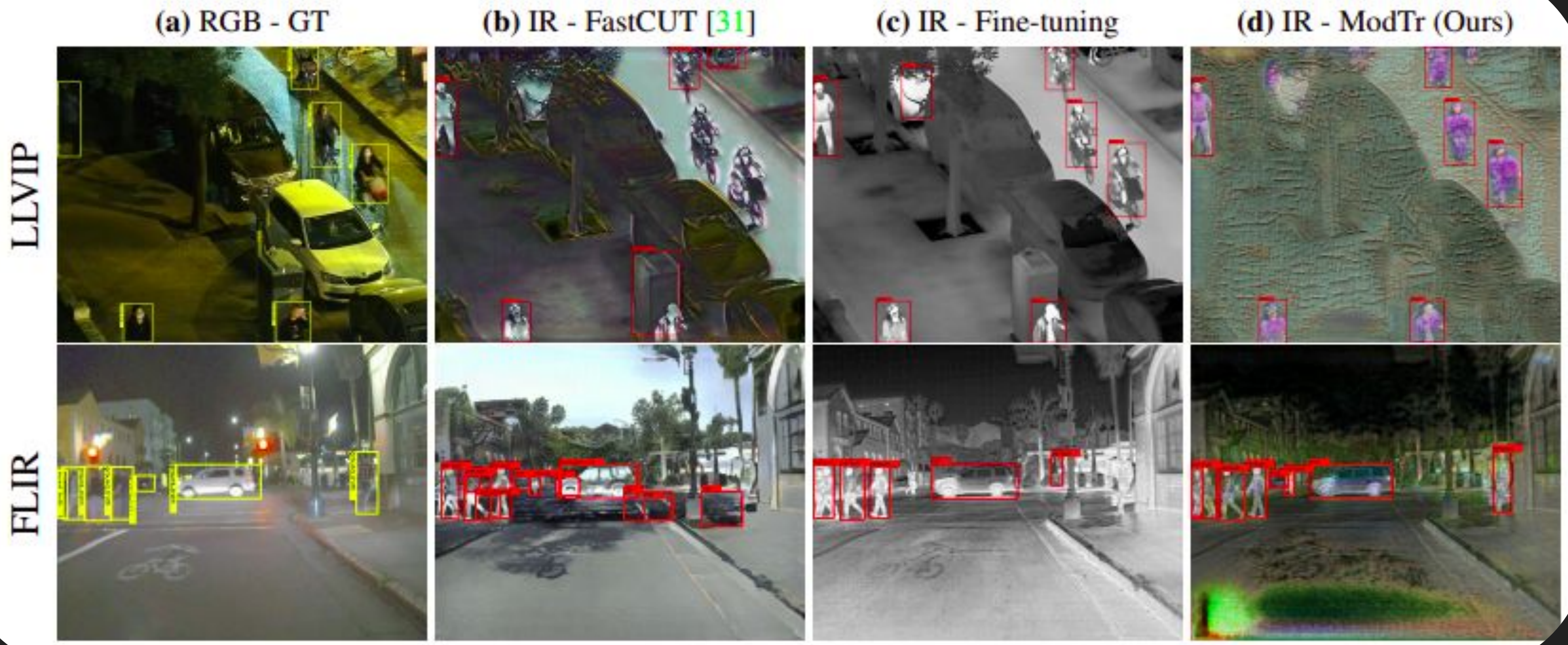


Knowledge Preservation through Input Modality Translation

Detector	Dataset	N-Detectors	1-Detector	N-ModTr-1-Det.
FCOS	LLVIP	57.37 ± 2.19	**58.55 ± 0.89**	57.63 ± 0.66
	FLIR	27.97 ± 0.59	26.70 ± 0.48	**35.49 ± 0.94**
	COCO	**38.41 ± 0.00**	00.33 ± 0.04	**38.41 ± 0.00**
	AVG.	41.25 ± 0.92	28.52 ± 0.47	**43.84 ± 0.53**
RetinaNet	LLVIP	53.79 ± 1.79	53.26 ± 3.02	**54.83 ± 0.61**
	FLIR	28.46 ± 0.50	25.19 ± 0.72	**34.27 ± 0.27**
	COCO	**35.48 ± 0.00**	00.29 ± 0.01	**35.48 ± 0.00**
	AVG.	39.24 ± 0.76	26.24 ± 1.28	**41.52 ± 0.29**
Faster R-CNN	LLVIP	59.62 ± 1.23	**62.50 ± 1.29**	57.97 ± 0.85
	FLIR	30.93 ± 0.46	28.90 ± 0.33	**37.21 ± 0.46**
	COCO	**39.78 ± 0.00**	00.40 ± 0.00	**39.78 ± 0.00**
	AVG.	43.44 ± 0.56	30.60 ± 0.54	**44.98 ± 0.43**

Table 3. Detection performance (AP) of knowledge-preserving techniques.

Qualitative Results



4. Conclusion

Conclusion

- In this work, we present a novel ModTr method for adapting ODs without changing their parameters.

- ModTr benefits from preserving the full knowledge of the detector, which opens the possibility of using the translation network as a node to change the modality for an unaltered detector.

ECCV

EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO 2024

Paper

Code

Heitor R. Medeiros, Masih A., Fidel A. G. Peña,
David Latortue, Eric Granger, Marco Pedersoli
LIVIA, Dept. of Systems Engineering, ETS Montreal, Canada

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE Université du Québec ÉTS Engineering for Industry

LIVIA LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTIFICIELLE

DISTECH CONTROLS

ILLS International Laboratory on Learning Systems

EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO 2024

Modality Translation for Object Detection Adaptation Without Forgetting Prior Knowledge

Heitor R. Medeiros, Masih A., Fidel A. G. Peña,
David Latortue, Eric Granger, Marco Pedersoli
LIVIA, Dept. of Systems Engineering, ETS Montreal, Canada

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE Université du Québec ÉTS Engineering for Industry

LIVIA LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTIFICIELLE

DISTECH CONTROLS™

ILLS International Laboratory on Learning Systems