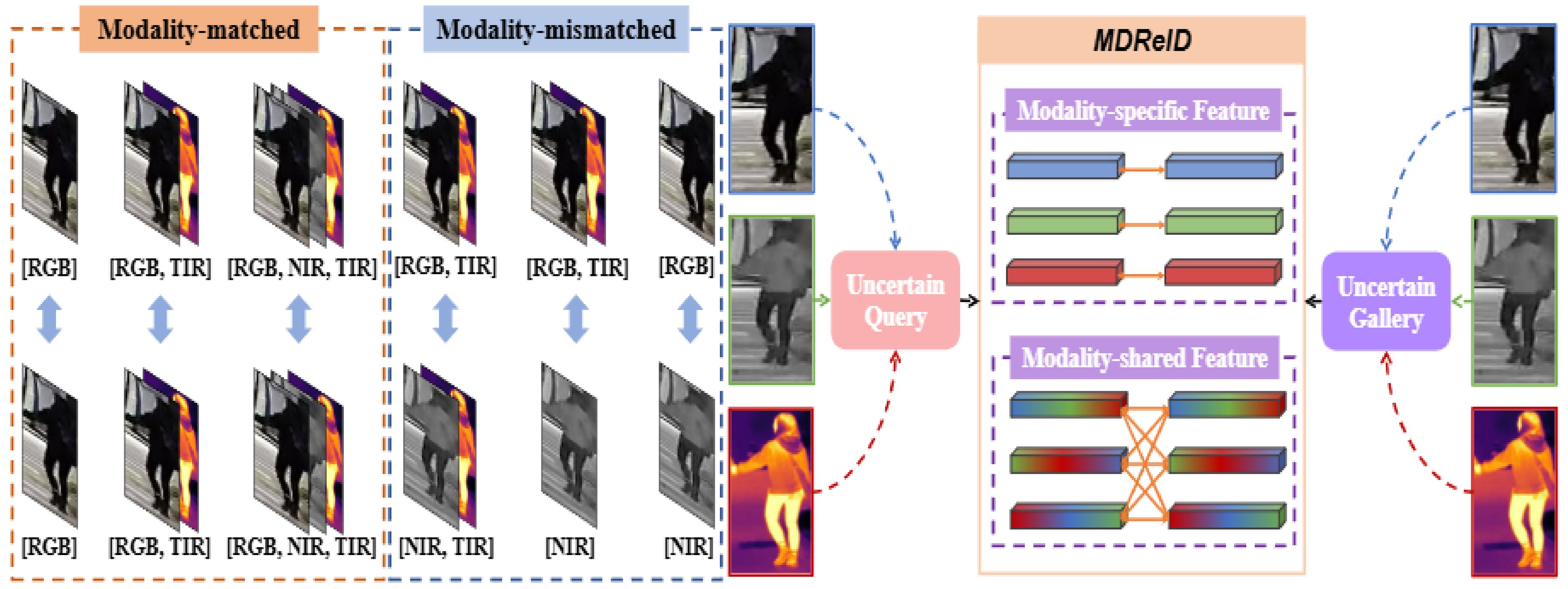
MDReID: Modality-Decoupled Learning for Any-to-Any Multi-Modal Object Re-Identiffcation

Yingying Feng[1], Jie Li[2], Jie Hu[3], Yukang Zhang[2], Lei Tan[3], Jiayi Ji[2,3],
[1]School of Computer Science and Engineering, Northeastern University, Shenyang, China
[2]School of Informatics, Xiamen University, Xiamen, China
[3]National University of Singapore, Singapore

NEURAL INFORMATION PROCESSING SYSTEMS
NeurIPS | 2025

Motivation



(a) Examples of modality-(mis)matched ReID scenarios. (b) The motivation of proposed MDReID.

(a) Though the availability of spectral modalities (e.g., RGB, NIR, TIR) varies across queries and galleries, recent methods only focus on the modality-matched scenarios, which limits their practical applicability.

(b) MDReID overcomes the rigidity of modality constraints by disentangling modality-shared and modality-specific features, enabling effective matching between queries and galleries from arbitrary modalities.

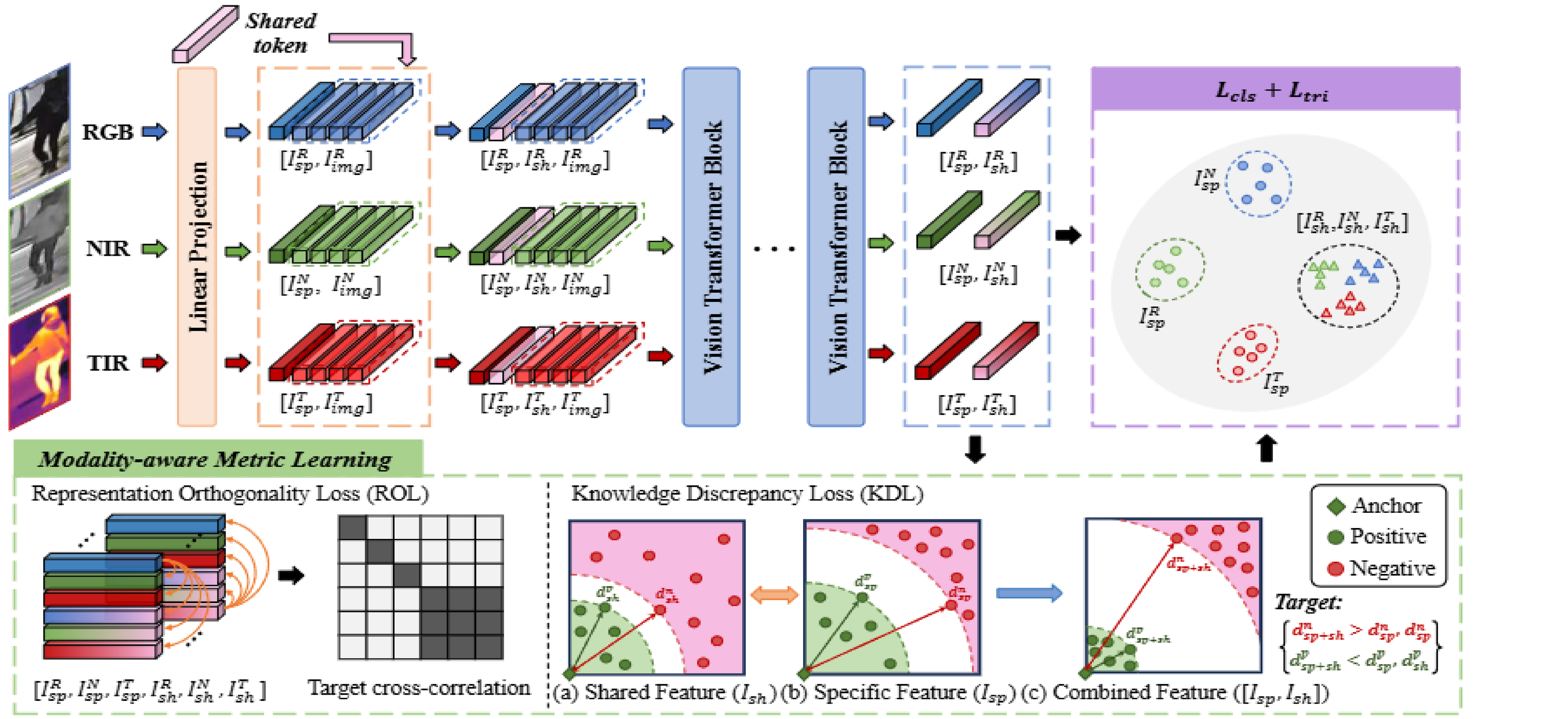
Contribution

(1) We propose MDReID, a flexible any-to-any object re-identification framework that supports retrieval across arbitrary query-gallery modality combinations, addressing the practical limitations of strictly aligned multi-modal datasets.

(2) We introduce a Modality-Decoupled Learning (MDL) strategy, coupled with a Modality-aware Metric Learning (MML) strategy. MDL explicitly disentangles modality-shared and modality-specific representations, while MML further strengthens this disentanglement by introducing a representation orthogonality loss and a knowledge discrepancy loss, encouraging the two components to encode distinct and complementary information.

Extensive experiments on multi-spectral object ReID datasets (RGBNT201, RGBNT100, MSVR310) demonstrate the superior adaptability and performance of the MDReID across diverse scenarios. MDReID achieves significant mAP improvements of 9.8%, 3.0%, and 11.5% in general modality-matched scenarios, and average gains of 3.4%, 11.8%, and 10.9% in modality-mismatched scenarios, respectively.

The Proposed MDReID



MDReID is designed to support retrieval across arbitrary modality combinations. It disentangles features into shared and specific components to boost performance in both matched and mismatched scenarios. Additionally, by leveraging representation orthogonality loss (ROL) and knowledge discrepancy loss (KDL), MDReID refines feature separation and enhances retrieval robustness.

Experiments

Method	RGBNT201 mAP	RGBNT201 R-1	RGBNT201 R-5	RGBNT201 R-10	Method	RGBNT100 mAP	RGBNT100 R-1	MSVR310 mAP	MSVR310 R-1
Single									
MUDeep [32]	23.8	19.7	33.1	44.3	DMML [33]	58.5	82.0	19.1	31.1
HACNN [34]	21.3	19.0	34.1	42.8	BoT [35]	78.0	95.1	23.5	38.4
MLFN [36]	26.1	24.2	35.9	44.1	Circle Loss [37]	59.4	81.7	22.7	34.2
PCB [38]	32.8	28.1	37.4	46.9	HRCN [39]	67.1	91.8	23.4	44.2
OSNet [40]	25.4	22.3	35.1	44.7	TransReID [2]	75.6	92.9	18.4	29.6
CAL [41]	27.6	24.3	36.5	45.7	AGW [15]	73.1	92.7	28.9	46.9
Multi									
HAMNet [7]	27.7	26.3	41.5	51.7	GAFNet [20]	74.4	93.4	-	-
PFNet [19]	38.5	38.9	52.0	58.4	GraFT [42]	76.6	94.3	-	-
IEEE [8]	47.5	44.4	57.1	63.6	GPFNet [23]	75.0	94.5	-	-
DENet [22]	42.4	42.2	55.3	64.5	PHT [30]	79.9	92.7	-	-
UniCat [31]	57.0	55.7	-	-	UniCat [31]	79.4	96.2	-	-
HTT [9]	71.1	73.4	83.1	87.3	CCNet [21]	77.2	96.3	36.4	55.2
EDITOR [10]	66.5	68.3	81.1	88.2	EDITOR [10]	82.1	96.4	39.0	49.3
RSCNet [26]	68.2	72.5	-	-	RSCNet [26]	82.3	**96.6**	39.5	49.6
TOP-ReID [11]	72.3	76.6	84.7	89.4	TOP-ReID [11]	81.2	96.4	35.9	44.6
MDReID (Ours)	**82.1**	**85.2**	**90.3**	**92.6**	**MDReID (Ours)**	**85.3**	95.6	**51.0**	**68.9**

	Methods	RT-to-NT mAP	RT-to-NT R-1	RT-to-N mAP	RT-to-N R-1	R-to-N mAP	R-to-N R-1	R-to-NT mAP	R-to-NT R-1	Average mAP	Average R-1
RGBNT201	EDITOR [10]	27.3	27.9	2.80	0.0	4.0	2.3	4.3	4.1	8.5	7.5
	TOP-ReID [11]	43.0	44.6	14.4	13.3	15.4	**14.0**	11.9	8.7	18.2	18.0
	MDReID (Ours)	**53.1**	**51.7**	**16.7**	**13.8**	**16.6**	11.1	**15.1**	**10.9**	**21.6**	**19.1**
RGBNT100	EDITOR [10]	42.1	59.9	2.6	0.8	2.8	1.5	2.7	1.5	11.9	15.5
	TOP-ReID [11]	59.0	81.8	21.9	28.5	26.2	34.0	21.7	25.4	26.8	36.1
	MDReID (Ours)	**69.4**	**85.5**	**39.6**	**47.9**	**45.4**	**56.4**	**37.6**	**41.8**	**38.6**	**47.4**
MSVR310	EDITOR [10]	6.4	11.0	2.2	2.5	1.6	0.2	1.7	1.5	2.5	3.4
	TOP-ReID [11]	18.4	30.5	12.9	19.5	13.7	21.3	12.6	18.8	11.2	17.8
	MDReID (Ours)	**35.1**	**52.5**	**24.7**	**34.5**	**28.6**	**39.8**	**27.0**	**36.0**	**22.1**	**31.7**

(a) **Ablation study.** MDL, L_{ROL}, and L_{KDL} indicate the Modality Decoupled Learning, Representation Orthogonality Loss (ROL), and Knowledge Discrepancy Loss (KDL), respectively.

Index	MDL	L_{ROL}	L_{KDL}	mAP	R-1
1	✕	✕	✕	27.8	27.1
2	✓	✕	✕	39.4	38.2
3	✓	✓	✕	41.2	40.8
4	✓	✕	✓	39.9	40.9
5	✓	✓	✓	**43.2**	**42.3**

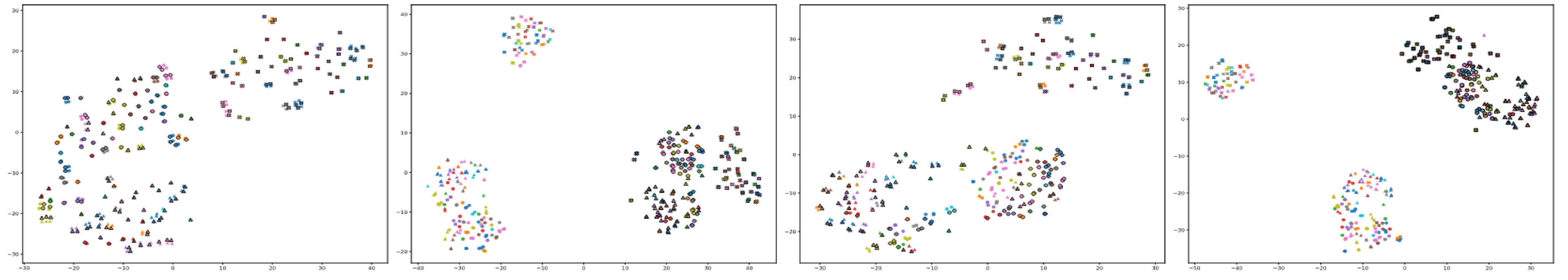
(b) **Performance under different** w_1. The optimal performance achieves when w_1 is set to 1.5.

w_1	mAP	R-1
0.5	39.6	38.1
1.0	40.3	40.0
1.5	**41.2**	**40.8**
2.0	38.9	38.1
3.0	38.2	37.0

(c) **Performance under different** w_2. The optimal performance achieves when w_2 is set to 5.25.

w_2	mAP	R-1
4.5	41.7	41.1
5.0	40.2	39.6
5.25	**43.2**	**42.3**
5.5	41.6	41.3
6.0	41.9	40.4

Visulization



(a) MDReID (MDL) (b) MDReID (MDL + ROL) (c) MDReID (MDL + KDL) (d) MDReID (MDL + ROL + KDL)

In (a), using only MDL leads to noticeable feature overlap and unclear separation between modality-specific and shared components. In (c), adding KDL slightly improves clustering and separation. In contrast, (b) shows ROL significantly enhances feature orthogonality, creating clearer boundaries. Finally, (d) reveals that combining ROL and KDL yields the most structured and disentangled feature space, with shared features well-separated from modality-specific ones. These results confirm ROL and KDL's complementary roles in refining the representation space.