IDEA: Inverted Text with Cooperative Deformable Aggregation for Multi-Modal Object ReID

CVPR2025

报告人：王宇皓

单位：大连理工大学

发展脉络

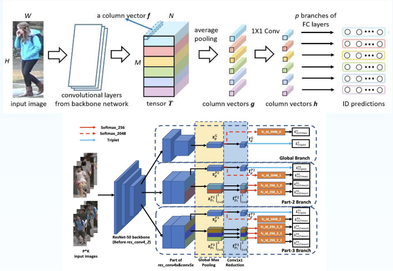
如何提取
行人鲁棒表征?

2006

行人重识别

2012

深度学习



2017

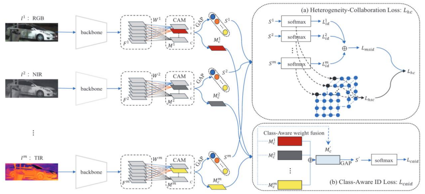
红外跨模
行人重识别

单模态

跨模态

配准的多模态
相机的普及

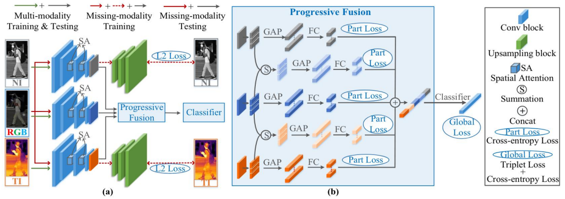
2020



多模态车辆重识别

RGBNT100/RGBN300

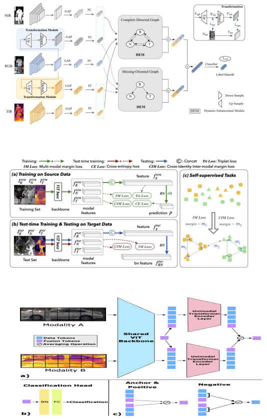
2021



多模态行人重识别

RGBNT201

至今



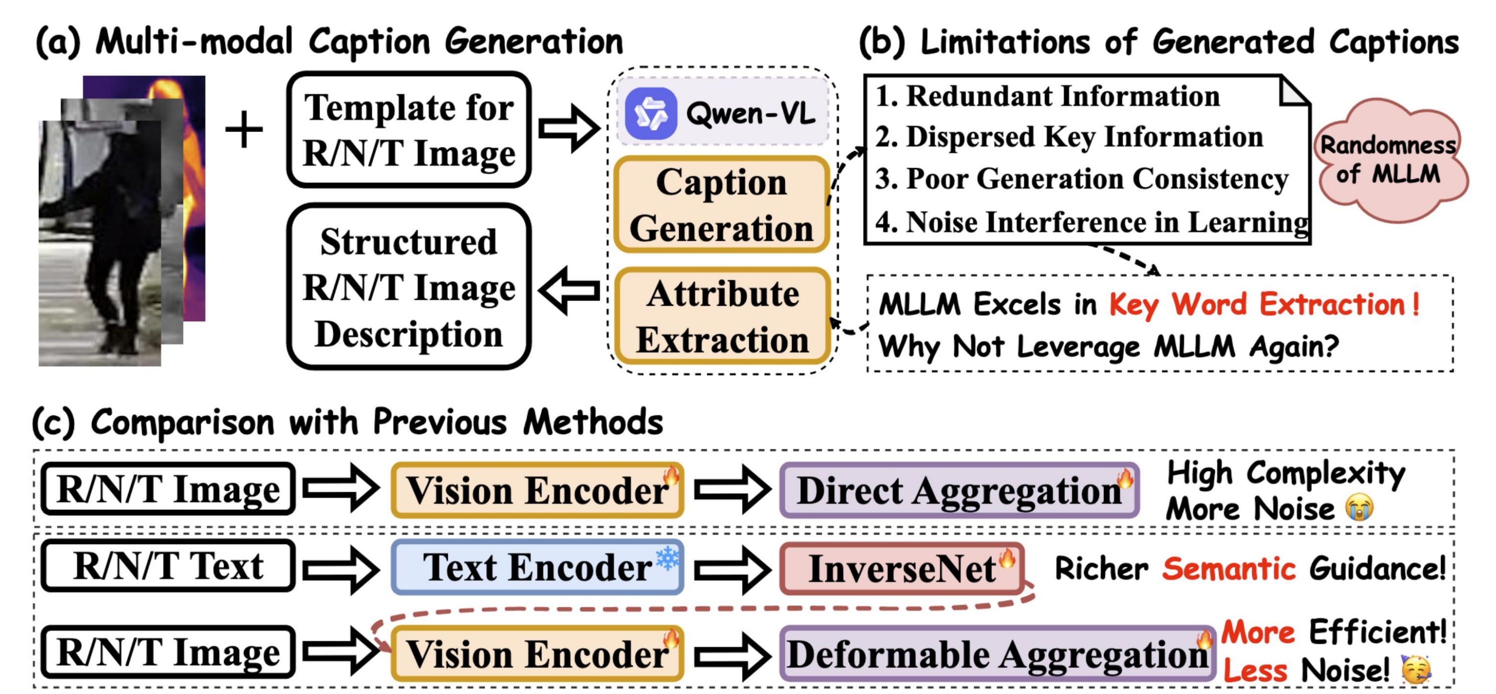
单模态/跨模态

多模态

CVPR、AAAI、NeurIPS....

如何充分融合多模态数据?

问题与动机



(a) Multi-modal Caption Generation

+ Template for R/N/T Image

Qwen-VL

Caption Generation

Attribute Extraction

Structured R/N/T Image Description

(b) Limitations of Generated Captions

1. Redundant Information
2. Dispersed Key Information
3. Poor Generation Consistency
4. Noise Interference in Learning

Randomness of MLLM

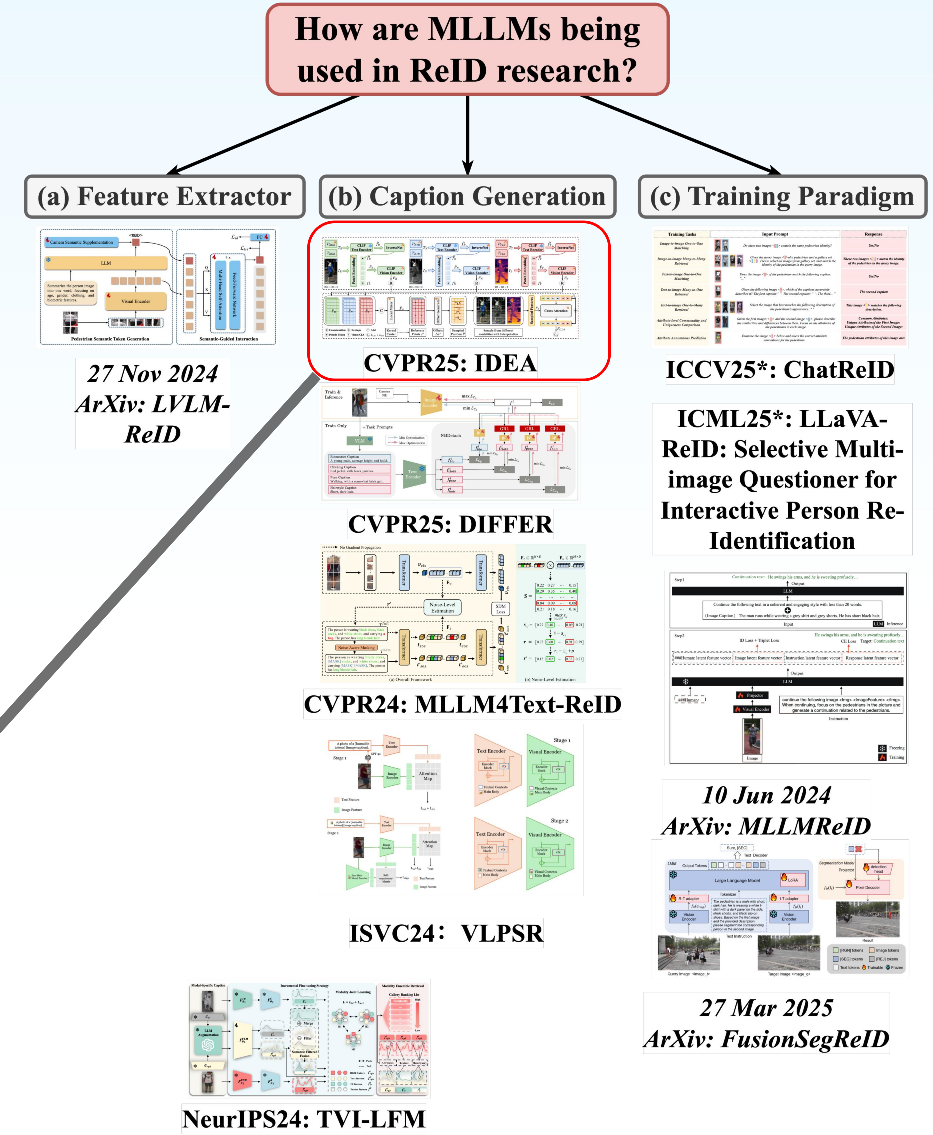
MLLM Excels in Key Word Extraction!
Why Not Leverage MLLM Again?

(c) Comparison with Previous Methods

R/N/T Image ⇨ Vision Encoder ⇨ Direct Aggregation — High Complexity More Noise 😭

R/N/T Text ⇨ Text Encoder ⇨ InverseNet — Richer Semantic Guidance!

R/N/T Image ⇨ Vision Encoder ⇨ Deformable Aggregation — More Efficient! Less Noise! 🥳



How are MLLMs being used in ReID research?

(a) Feature Extractor

(b) Caption Generation

(c) Training Paradigm

27 Nov 2024
ArXiv: LVLM-ReID

CVPR25: IDEA

ICCV25*: ChatReID

CVPR25: DIFFER

ICML25*: LLaVA-ReID: Selective Multi-image Questioner for Interactive Person Re-Identification

CVPR24: MLLM4Text-ReID

10 Jun 2024
ArXiv: MLLMReID

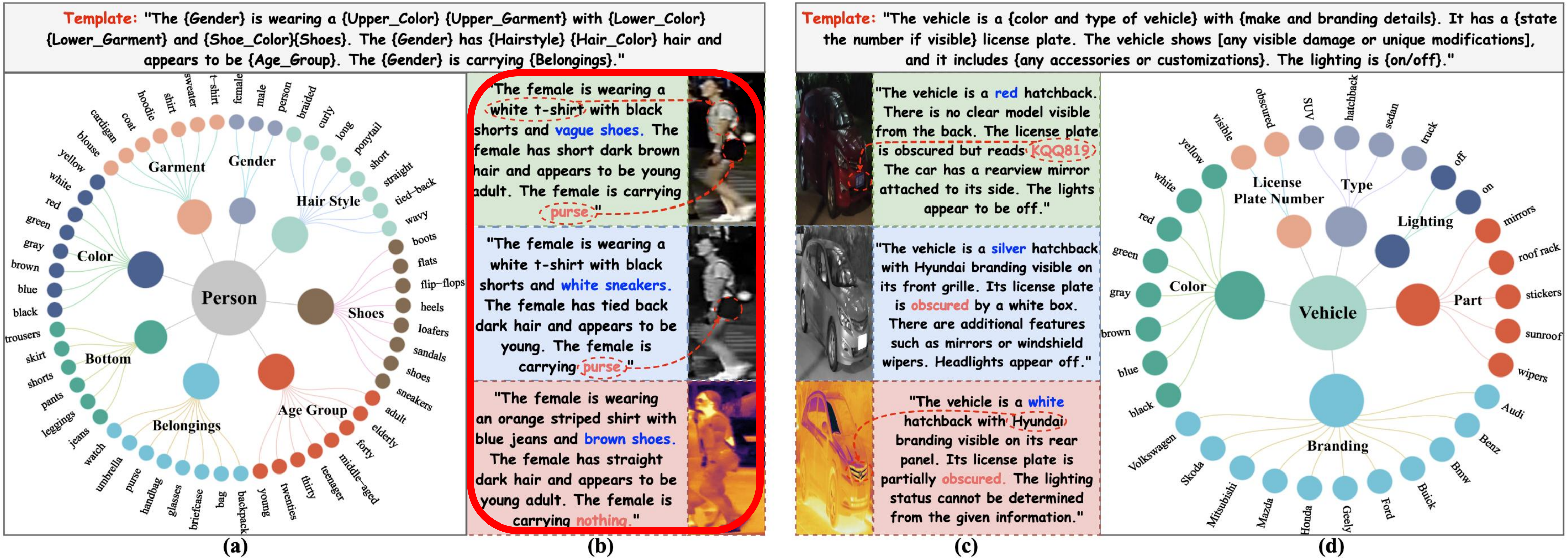
ISVC24：VLPSR

27 Mar 2025
ArXiv: FusionSegReID

NeurIPS24: TVI-LFM

- ➢ 现在ReID领域大家都在怎么用MLLMs?
- ➢ 是否可以将MLLMs引入多模态ReID?
- ➢ 如何做高效的多模态特征聚合?

生成流程



Template: "The {Gender} is wearing a {Upper_Color} {Upper_Garment} with {Lower_Color} {Lower_Garment} and {Shoe_Color}{Shoes}. The {Gender} has {Hairstyle} {Hair_Color} hair and appears to be {Age_Group}. The {Gender} is carrying {Belongings}."

"The female is wearing a white t-shirt with black shorts and vague shoes. The female has short dark brown hair and appears to be young adult. The female is carrying purse."

"The female is wearing a white t-shirt with black shorts and white sneakers. The female has tied back dark hair and appears to be young. The female is carrying purse."

"The female is wearing an orange striped shirt with blue jeans and brown shoes. The female has straight dark hair and appears to be young adult. The female is carrying nothing."

Template: "The vehicle is a {color and type of vehicle} with {make and branding details}. It has a {state the number if visible} license plate. The vehicle shows [any visible damage or unique modifications], and it includes {any accessories or customizations}. The lighting is {on/off}."

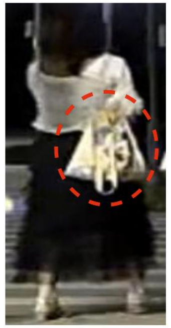
"The vehicle is a red hatchback. There is no clear model visible from the back. The license plate is obscured but reads KQQ819. The car has a rearview mirror attached to its side. The lights appear to be off."

"The vehicle is a silver hatchback with Hyundai branding visible on its front grille. Its license plate is obscured by a white box. There are additional features such as mirrors or windshield wipers. Headlights appear off."

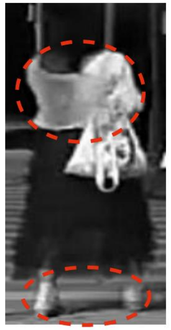
"The vehicle is a white hatchback with Hyundai branding visible on its rear panel. Its license plate is partially obscured. The lighting status cannot be determined from the given information."

- ➢ 每个模态单独标注，对应模版与图像匹配送入MLLMs
- ➢ 原始标注送入MLLMs，按预定义的属性提取关键信息填充模版

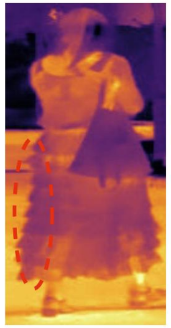
实例展示



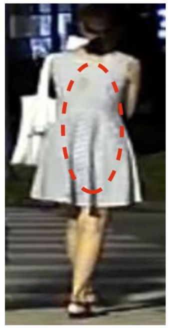
The female is wearing a vague off-white blouse with vague long black skirt and white shoes. The female has black hair and appears to be young adult. The female is carrying beige purse with gold accents.



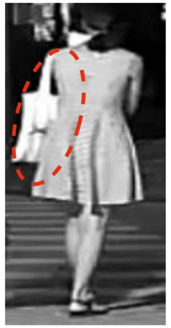
The female is wearing a off-white blouse with black long skirt and heeled shoes. The female has unknown loose waves hair and appears to be middle-aged. The female is carrying white purse with accents.



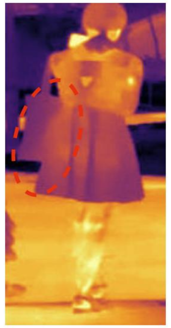
The female is wearing a vague orange with black stripes with vague white with ruffles and white shoes. The female has brown hair and appears to be middle age. The female is carrying handbag.



The female is wearing a blue sleeveless dress with white knee-length hemline and black shoes. The female has long dark hair and appears to be adult (late twenties). The female is carrying large white handbag.



The female is wearing a black and white striped dress with vague flared skirt and sandal shoes. The female has unknown short hair and appears to be young adult. The female is carrying white handbag.

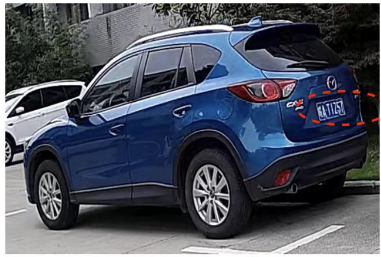


The female is wearing a orange blouse with white skirt and sandal shoes. The female has unknown blonde hair and appears to be 30-40. The female is carrying handbag.

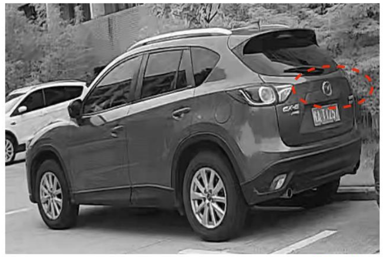
- ➢ 降低标注难度
- ➢ 保留模态特定信息

模态文本冲突

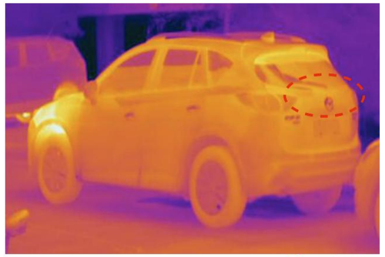
颜色（R/N/T）



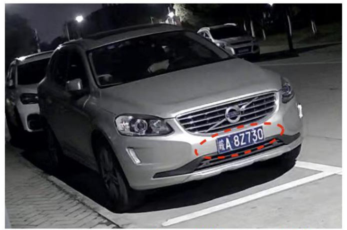
The vehicle is a blue SUV with a \"T1257\" license plate located at the rear. There are no visible damages or modifications, but there may be additional accessories such as a roof rack or bike holder. The lights appear to be off.



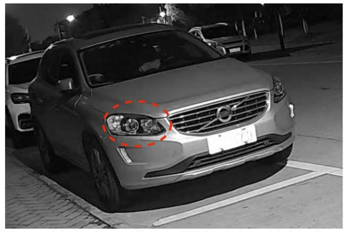
The vehicle is a gray SUV with a Mazda logo. Its license plate is partially obscured but reads \"T1257\". There are no other discernible features such as accessories or additional modifications. The lights appear to be off.



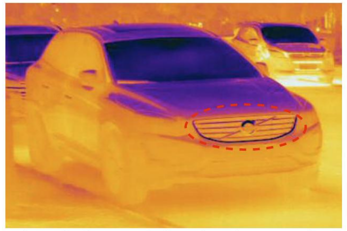
The vehicle is a white SUV with a Mazda logo. Its license plate is partially obscured but appears to be blue with white lettering. There are no visible damages or modifications. The headlights appear to be off, suggesting either early morning or late evening hours.



The vehicle is a silver SUV with a Volvo logo visible on its front grille. Its license plate is clearly readable as \"A82730\". There are no noticeable damages or modifications, but there might be a slight scratch on the rear bumper. No additional accessories like a bike rack or trailer hitch were observed. Headlights appear off.

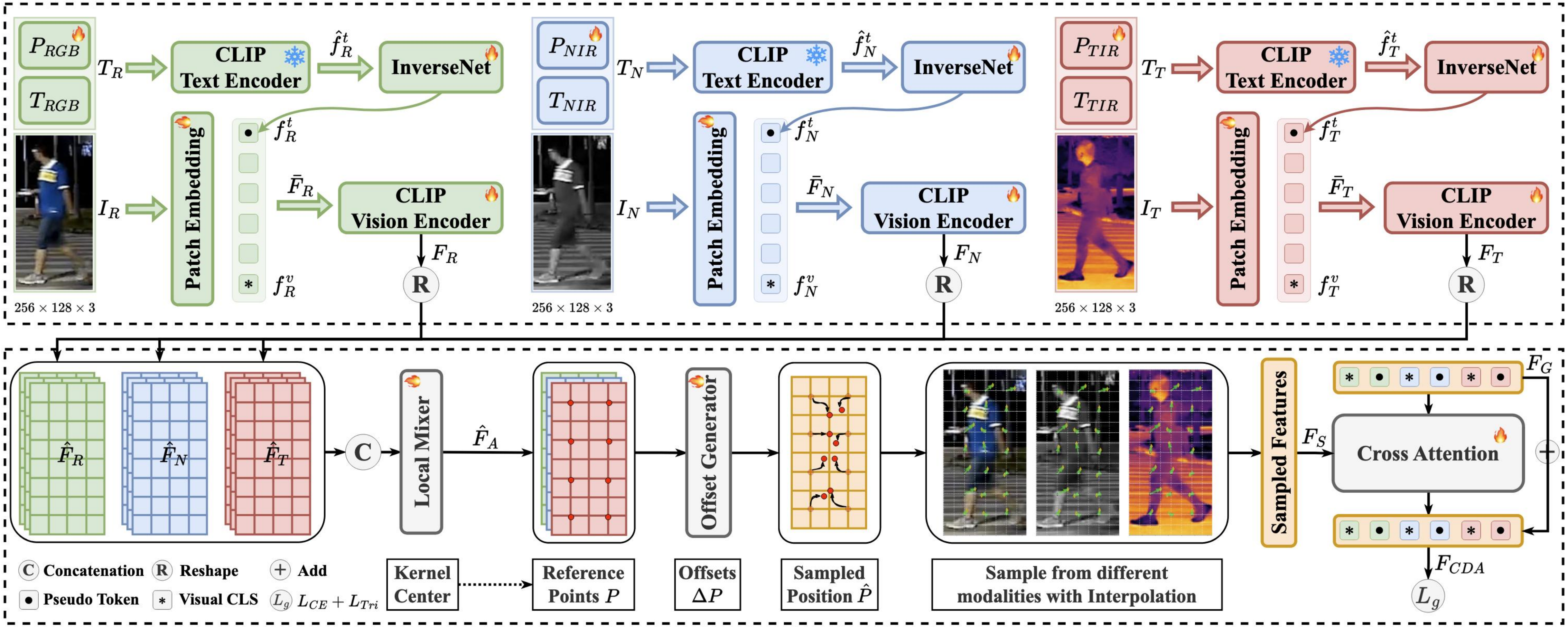


The vehicle is a silver SUV with a Volvo logo visible on its front grille. Its license plate is partially obscured but appears to be white with black lettering. There are no noticeable damages or modifications to the car, and there are no additional accessories or customization seen. The headlights appear to be off, indicating nighttime or low-light conditions.



The vehicle is a dark-colored SUV with a Volvo logo visible on its front grille. Its license plate is obscured from view. There are no noticeable damages or modifications to the car, but there are two small dents on its rear bumper. Additionally, there seems to be a bike rack attached to the back of the SUV. The lighting status cannot be determined.

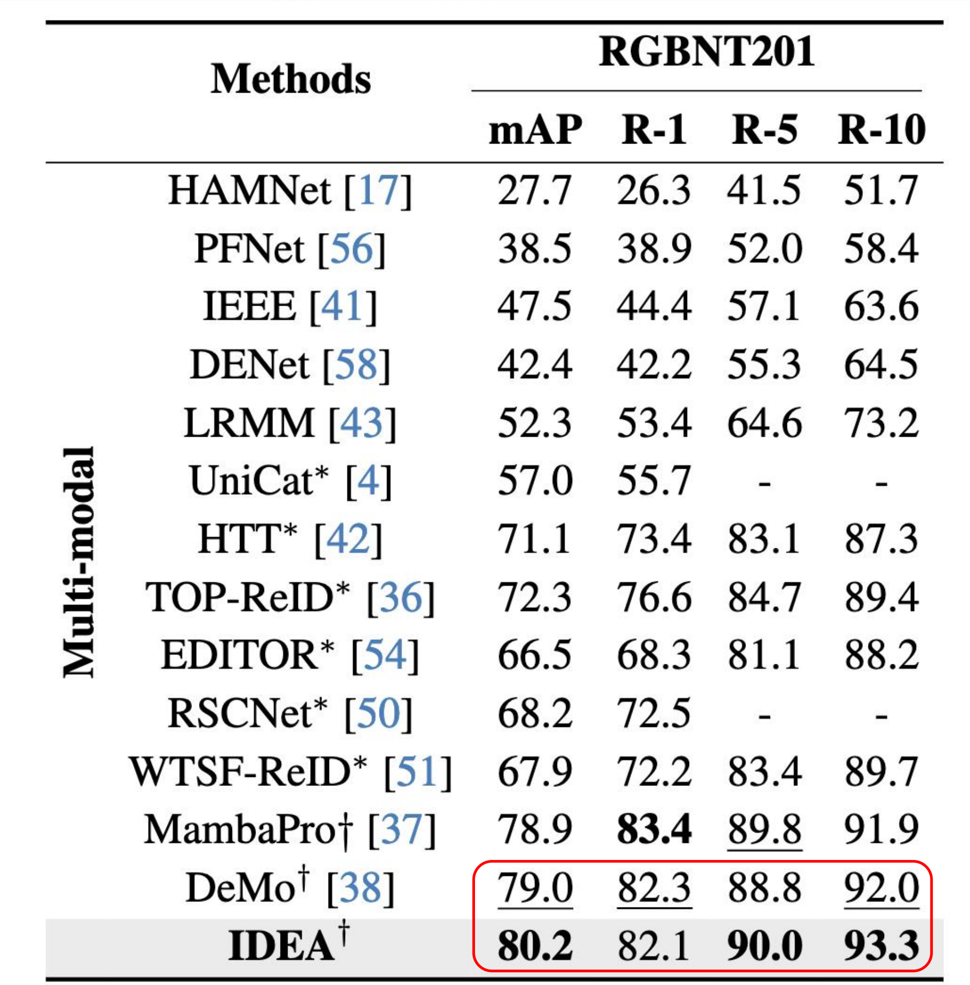
模型框架



- ➢ **Modal Prefixes**
- ➢ **InverseNet**
- ➢ **Cooperative Deformable Aggregation**

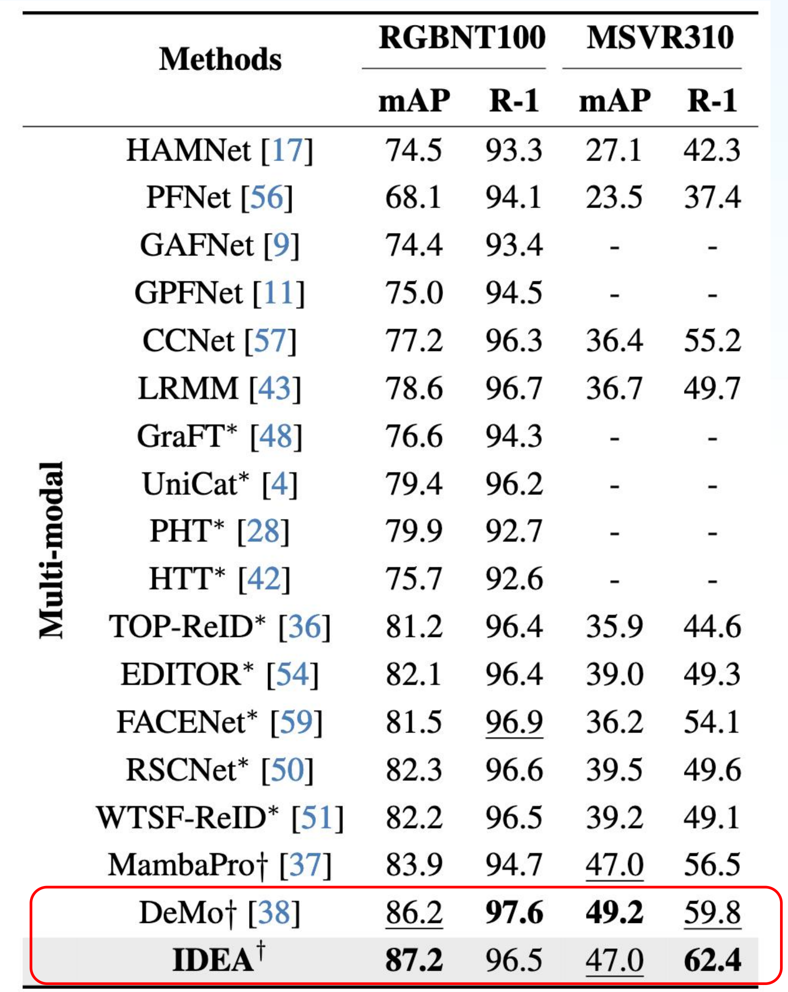
数值实验

➢ 多模态行人重识别测评



	Methods	RGBNT201			
		mAP	R-1	R-5	R-10
Multi-modal	HAMNet [17]	27.7	26.3	41.5	51.7
	PFNet [56]	38.5	38.9	52.0	58.4
	IEEE [41]	47.5	44.4	57.1	63.6
	DENet [58]	42.4	42.2	55.3	64.5
	LRMM [43]	52.3	53.4	64.6	73.2
	UniCat* [4]	57.0	55.7	-	-
	HTT* [42]	71.1	73.4	83.1	87.3
	TOP-ReID* [36]	72.3	76.6	84.7	89.4
	EDITOR* [54]	66.5	68.3	81.1	88.2
	RSCNet* [50]	68.2	72.5	-	-
	WTSF-ReID* [51]	67.9	72.2	83.4	89.7
	MambaPro† [37]	78.9	**83.4**	89.8	91.9
	DeMo† [38]	79.0	82.3	88.8	92.0
	IDEA†	**80.2**	82.1	**90.0**	**93.3**

➢ 多模态车辆重识别测评



	Methods	RGBNT100		MSVR310	
		mAP	R-1	mAP	R-1
Multi-modal	HAMNet [17]	74.5	93.3	27.1	42.3
	PFNet [56]	68.1	94.1	23.5	37.4
	GAFNet [9]	74.4	93.4	-	-
	GPFNet [11]	75.0	94.5	-	-
	CCNet [57]	77.2	96.3	36.4	55.2
	LRMM [43]	78.6	96.7	36.7	49.7
	GraFT* [48]	76.6	94.3	-	-
	UniCat* [4]	79.4	96.2	-	-
	PHT* [28]	79.9	92.7	-	-
	HTT* [42]	75.7	92.6	-	-
	TOP-ReID* [36]	81.2	96.4	35.9	44.6
	EDITOR* [54]	82.1	96.4	39.0	49.3
	FACENet* [59]	81.5	96.9	36.2	54.1
	RSCNet* [50]	82.3	96.6	39.5	49.6
	WTSF-ReID* [51]	82.2	96.5	39.2	49.1
	MambaPro† [37]	83.9	94.7	47.0	56.5
	DeMo† [38]	86.2	**97.6**	**49.2**	59.8
	IDEA†	**87.2**	96.5	47.0	**62.4**

大规模数据集

RGBNT100

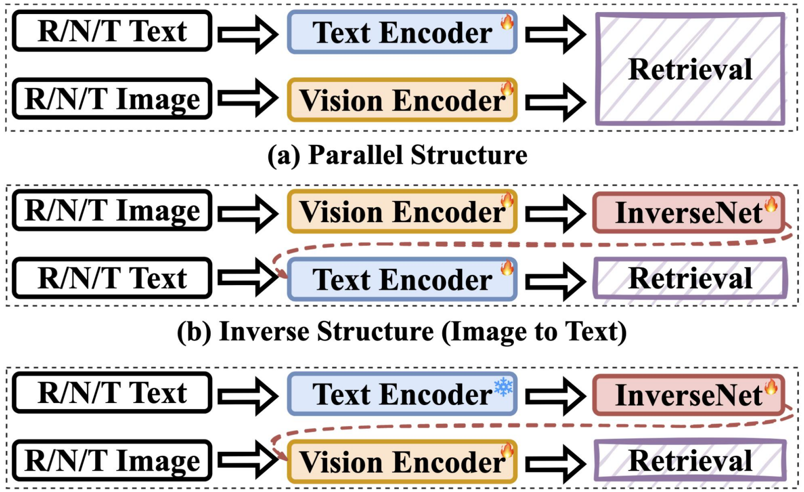
mAP显著提升

消融实验

Index	Modules			Metrics	
	Text	IMFE	CDA	mAP	Rank-1
A	✕	✕	✕	70.3	72.1
B	✓	✕	✕	73.4	75.8
C	✓	✓	✕	77.2	81.1
D	✓	✓	✓	**80.2**	**82.1**

Index	IMFE			Metrics	
	InverseNet	Prefixes	Prompt	mAP	Rank-1
A	✕	✕	-	72.6	75.1
B	✕	✓	✕	73.4	75.8
C	✓	✕	-	73.7	77.3
D	✓	✓	✕	75.4	78.6
E	✓	✓	✓	**77.2**	**81.1**

Index	CDA			Metrics	
	Sample	Cross Attn	Shared Offset	mAP	Rank-1
A	✕	✕	-	76.3	78.7
B	✕	✓	-	77.0	79.8
C	✓	✕	✕	76.8	78.9
D	✓	✕	✓	77.6	80.4
E	✓	✓	✕	79.5	81.7
F	✓	✓	✓	**80.2**	**82.1**



(a) Parallel Structure

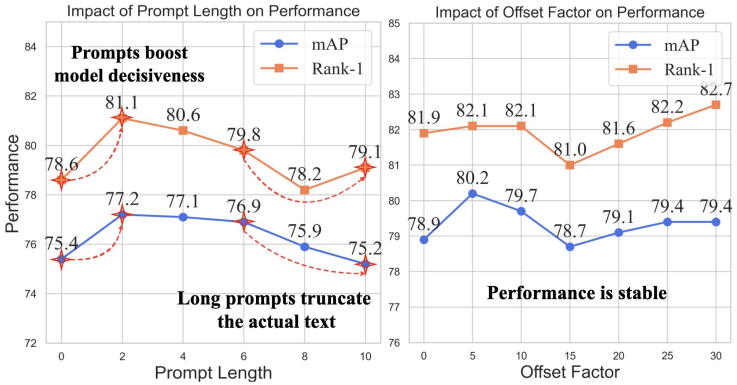
(b) Inverse Structure (Image to Text)

(c) Inverse Structure (Text to Image)

Index	Inverse Direction	Metrics			
		mAP	Rank-1	Rank-5	Rank-10
A	Image to Text	72.0	73.4	83.4	89.2
B	Text to Image	**77.2**	**81.1**	**88.4**	**92.2**

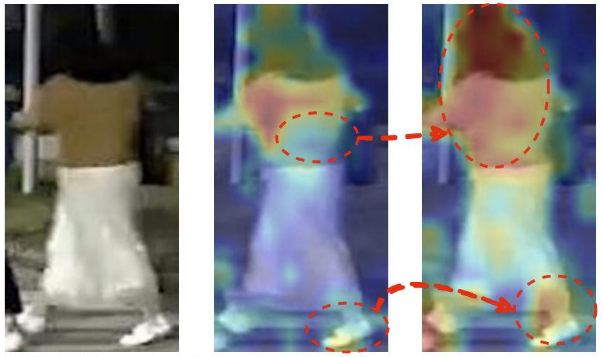
Model	Training Time (h)	Memory (GB)	Time per Epoch (min)
TOP-ReID	0.6650	17.80	0.3325
EDITOR	0.4074	16.41	0.3492
IDEA	**0.3075**	**18.02**	**0.3600**

Methods	Params	RGBNT201		RGBNT100		MSVR310	
	M	mAP	R-1	mAP	R-1	mAP	R-1
HAMNet [17]	78.00	27.7	26.3	74.5	93.3	27.1	42.3
CCNet [57]	74.60	-	-	77.2	96.3	36.4	55.2
IEEE [41]	109.22	49.5	48.4	-	-	-	-
GAFNet [9]	130.00	-	-	74.4	93.4	-	-
UniCat* [4]	259.02	57.0	55.7	79.4	96.2	-	-
GraFT* [48]	101.00	-	-	76.6	94.3	-	-
TOP-ReID* [36]	324.53	72.3	76.6	81.2	96.4	35.9	44.6
EDITOR* [54]	118.55	66.5	68.3	82.1	96.4	39.0	49.3
RSCNet* [50]	124.10	68.2	72.5	82.3	96.6	39.5	49.6
WTSF-ReID* [51]	143.60	67.9	72.2	82.2	96.5	39.2	49.1
MambaPro† [37]	74.20	78.9	**83.4**	83.9	94.7	47.0	56.5
DeMo† [38]	98.79	79.0	82.3	86.2	**97.6**	**49.2**	59.8
IDEA†	91.67	**80.2**	82.1	**87.2**	96.5	47.0	**62.4**

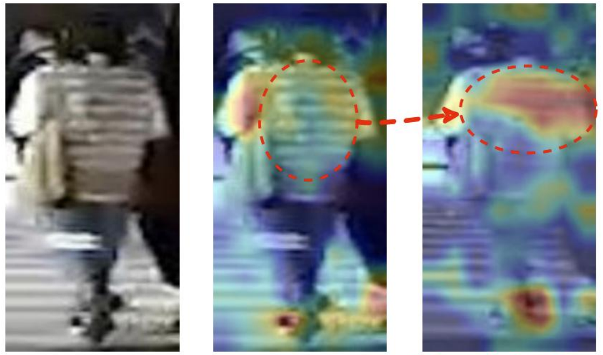


参数量小，收敛快，训练可以在20分钟左右完成

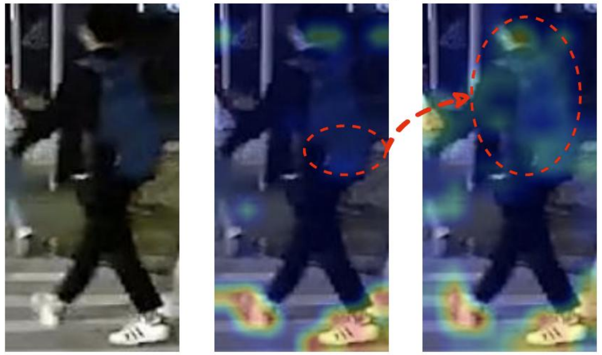
可视化



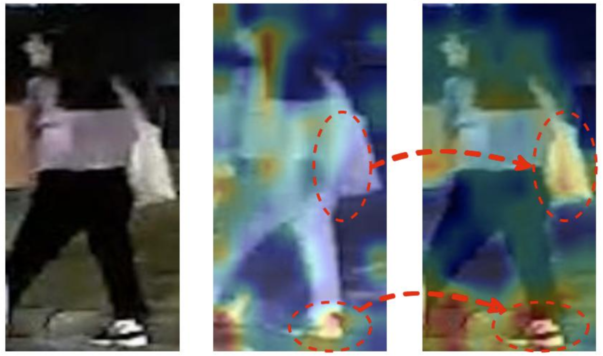
The female is wearing a vague orange blouse with vague long white skirt and white shoes. The female has short black hair and appears to be young adult. The female is carrying earrings.



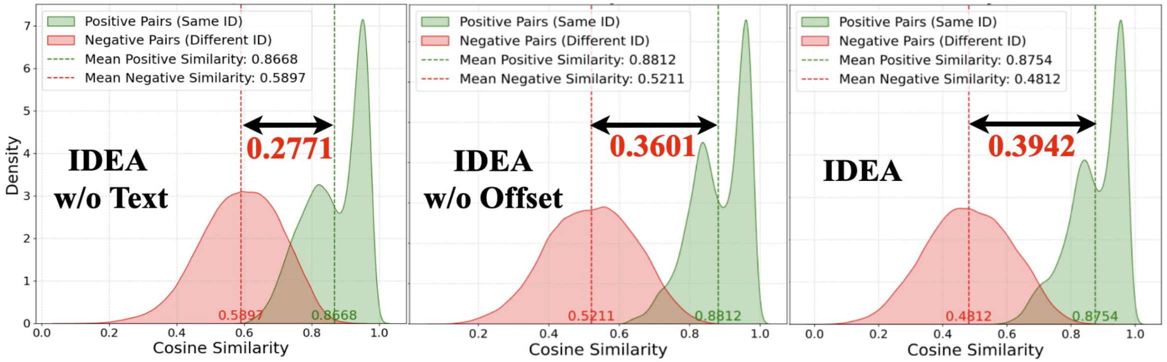
The female is wearing a white black stripe shirt with blue jeans. The female has short dark hair and appears to be around 30 yeas old. The female is carrying beige purse.



The male is wearing a black jacket with blue jeans and white shoes.The male has short dark brown hair and appears to be 20-30 years old. The male is carrying bag.

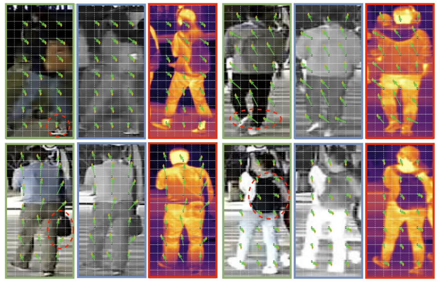


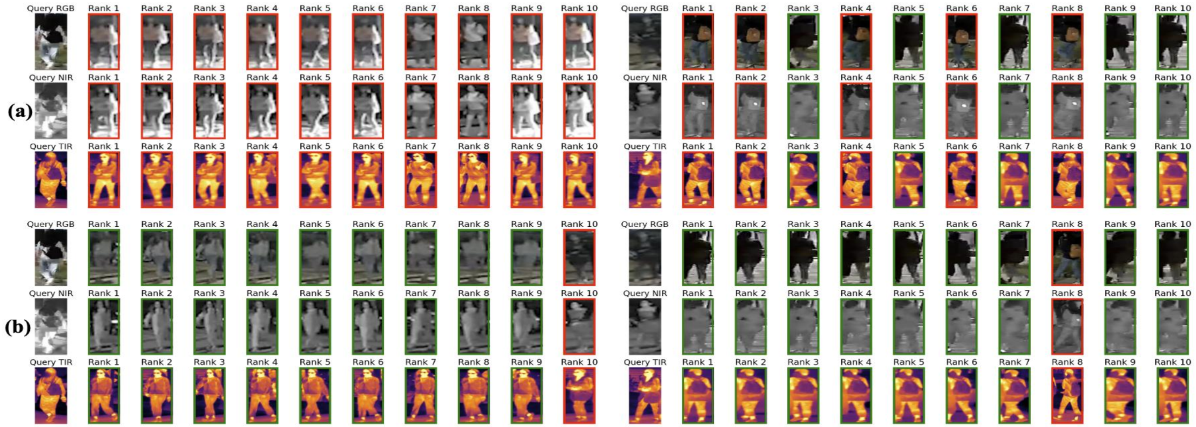
The female is wearing a pink blouse with black stripes and white sneakers. The female has long dark hair and appears to be young. The female is carrying white purse.



Model	mAP	Rank-1	Rank-5	Rank-10
IDEA w/o Text	74.5	75.0	84.8	88.8
IDEA	**80.2**	**82.1**	**90.0**	**93.3**

Model	mAP	Rank-1	Rank-5	Rank-10
IDEA w/o Offset	78.4	81.3	89.7	92.2
IDEA	**80.2**	**82.1**	**90.0**	**93.3**





系列工作

Trans.
CLIP

MLLMs

【AAAI24】 TOP-ReID

- ➢ Token Permutation
- ➢ Modality-Missing

【CVPR24】 EDITOR

- ➢ Token Selection
- ➢ Spatial-Frequency

【AAAI25】 MambaPro

- ➢ Prompt Tuning
- ➢ Mamba Fusion

【AAAI25】 DeMo

- ➢ Feature Decoupling
- ➢ Attention-based MoE

【CVPR25】 IDEA

- ➢ InverseNet via MLLMs
- ➢ Deformable Aggregation

Awesome
Collection

未来展望

1. 扩展数据规模：制作大规模多模态数据集，构建标准的测试工具包

2. 数据配准划分：根据配准情况划分子集，提供多样的研究方向

3. 丰富任务设定：考虑不同情况下的模态缺失/比例，开发强力检索模型

4. 探究数据偏差：深入分析模态懒惰问题，帮助研究者应对数据偏差

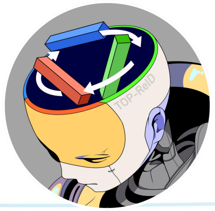
5. 增加描述模态：引入文本、草图、点云和事件相机等多种模态

全天候

多模态

多平台

目标重新识别



……

上游做好多模态的数据融合
为具身智能等多源感知铺垫

IDEA: Inverted Text with Cooperative Deformable Aggregation for Multi-Modal Object ReID

CVPR2025

WeChat

PHP

报告人：王宇皓

单位：大连理工大学