

# Cross-Modal Cross-Domain Dual Alignment Network for RGB-Infrared Person Re-Identification



X. Fu, F. Huang, Y. Zhou, H. Ma, X. Xu and L. Zhang, "Cross-Modal Cross-Domain Dual Alignment Network for RGB-Infrared Person Re-Identification," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 10, pp. 6874-6887, Oct. 2022.

报告人：王禹轩、赵继承、张熙林

汇报时间：2024/05/22



# 引言和相关工作

报告人：王禹轩

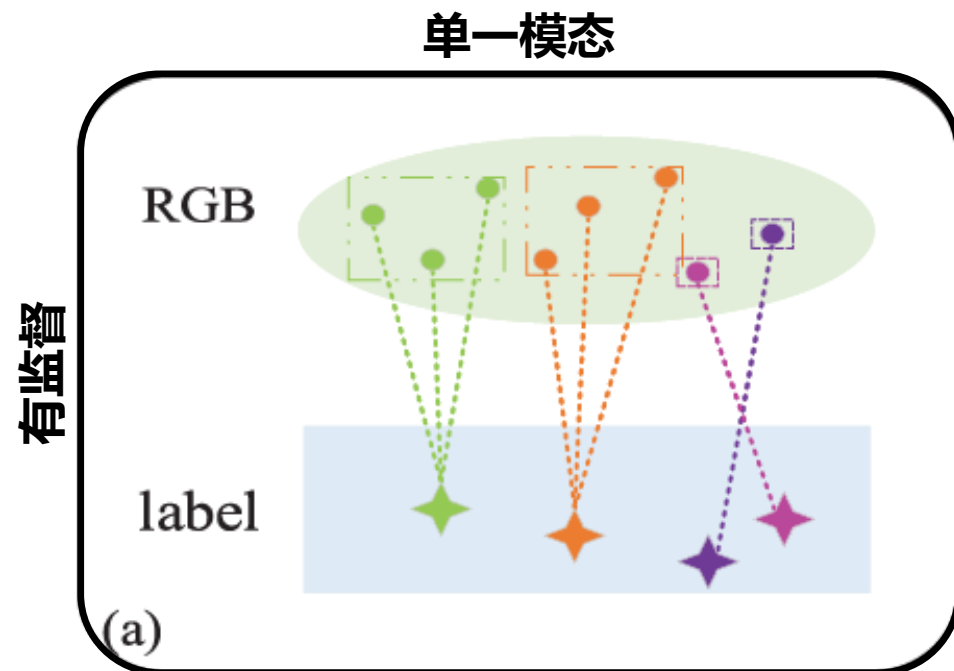
- 人员重新识别 (Re-ID) 是分布式多摄像头监控系统中的一项**基本任务**
- 虽然人员Re-ID的性能已经显著提升，但在**真实场景中的问题**还远远没有解决[28][41][53]



[28] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, "Feature refinement and filter network for person re-identification," *IEEE Trans. Circuits Syst. Video Technol* (2021).

[41] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Proc. 26th ACM Int. Conf. Multimedia* (2018).

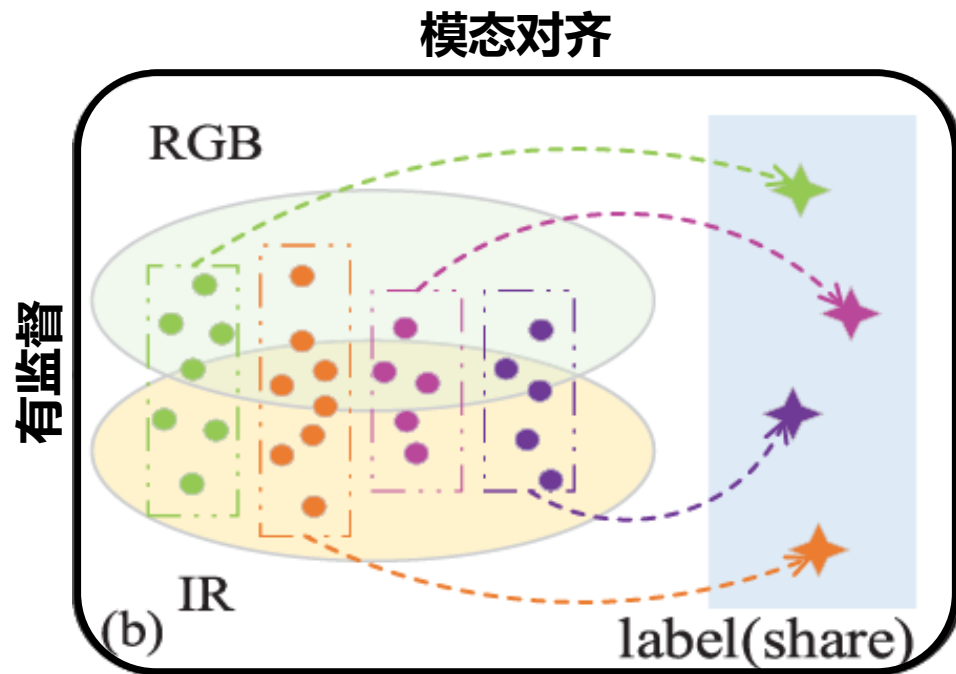
[53] Z. Zhang, C. Lan, W. Zeng, X. Jin, and Z. Chen, "Relation-aware global attention for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2020).



- 在光线不佳环境下，RGB图像**无法清晰捕捉到**人员的外观特征
- 在实际的摄像头监控网络中，大多数摄像头支持自动从可见光**切换到红外光**工作模式

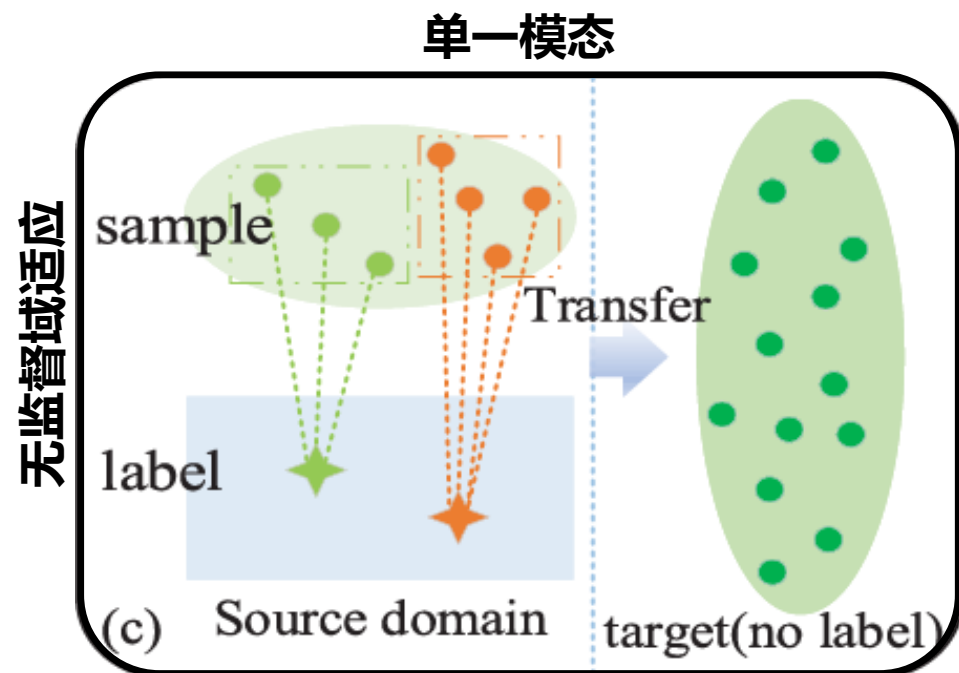


✓ 研究**RGB-红外(RGB-IR)跨模态人员匹配问题**是有意义的



RGB-IR图像对的  
外观存在**显著差异**

- RGB-IR Re-ID面临着固有异质性差距引起的**模态间差异挑战**
- 不同RGB-IR图像对的外观存在**显著差异挑战**
- 虽然已有许多RGB-IR Re-ID方法[6]提出，但是获取成对的RGB-IR数据**耗时又费力**

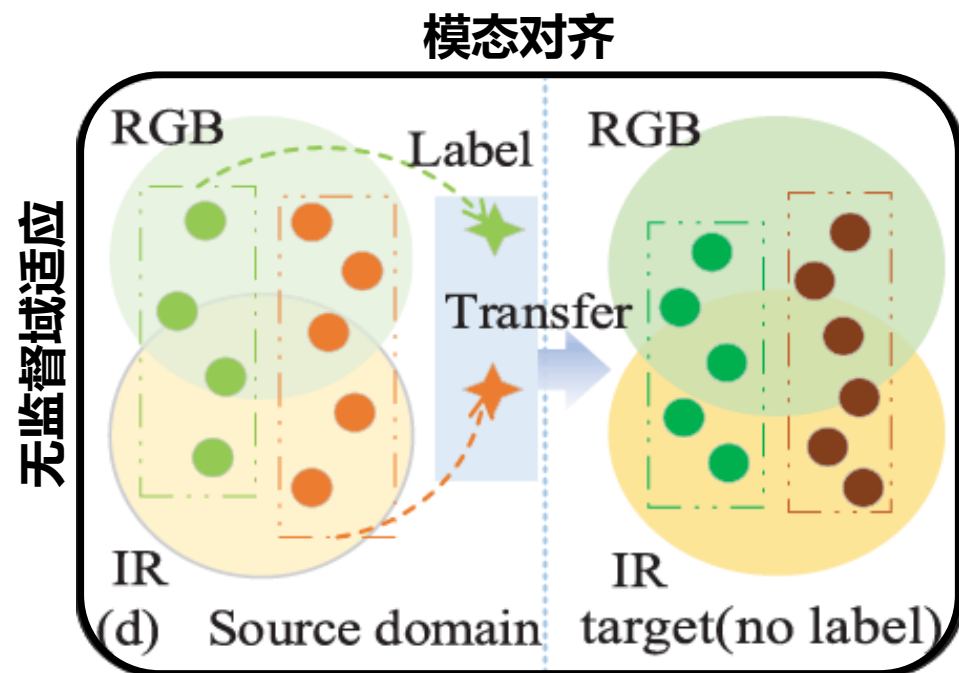


- 不同域的样本因数据集偏差而遭受 **不一致的数据分布**
- 由于缺乏身份注释，目标域内的域内变化对性能更加重要

单一模态下的监督学习问题：RGB-红外跨模态人员匹配是有意义的



无监督跨域人员Re-ID问题：仅仅关注可见光数据的不同数据集之间的域适应是不够的

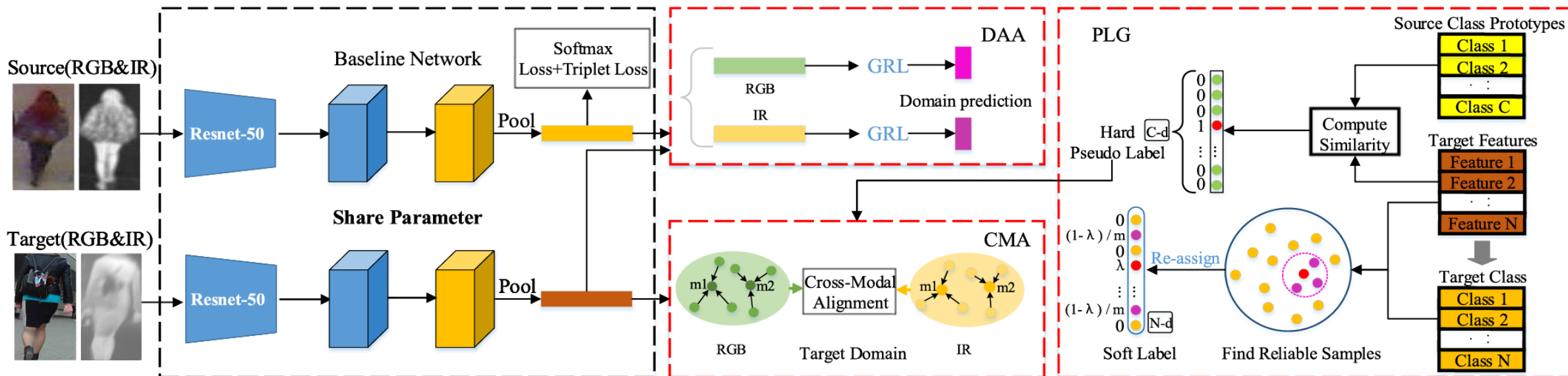


- 应对RGB-IR跨模态人员Re-ID挑战
- 应对跨域适应的挑战



- ✓ 首次将跨模态跨域问题结合，更加符合实际需求  
(获取未标记的数据更容易)





跨模态跨域问题 → 双重对齐网络 (DAN) {

- 域对抗对齐组件 (DAA) : 缩小域间差异, 减少跨域影响
- 目标域伪标签生成模块 (PLG) : 生成伪标签
- 跨模态对齐组件 (CMA) : 减少跨模态的影响

## 论文实验：

- 跨真实域实验：RegDB[27] & ThermalWorld[17]
- 跨真实和合成域实验：RegDB (RGB-红外) → 合成的RegDB

SYSU-MM01 (RGB-近红外) → 合成的SYSU-MM01

## 主要贡献：

- 首次提出**RGB-IR跨模态跨域人员Re-ID问题**，并为之提出一个**三重组件构成的双重对齐网络（DAN）**
- 通过为现有数据集合成目标域，经过实验证明所提出方法的**有效性和可转移性**

[17] V. V. Kniaz, V. A. Knyaz, J. Hladuvka, W. G. Kropatsch, and V. Mizginov, "ThermalGAN: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset," in *Proc. ECCV Workshops* (2018).

[27] D. Nguyen, H. Hong, K. Kim, and K. Park, "Person recognition system based on a combination of body images from visible light and thermal cameras," *Sensors* (2017).c

- Tan等人提出Consecutive Batch DropBlock网络 (CBDB-Net) 来捕获注意力集中且鲁棒的人物描述符[36]
- IDE学习通过分类损失从图像中学习ID级别的全局特征[55]
- PCB和MGN考虑局部特征的影响以增强特征表示的可区分性[35][41]
- Wang等人提出通过学习深度表示与自适应边界列表损失来缓解数据不平衡问题，并通过对四个CNN的组合来增强对姿势变化和部分遮挡的鲁棒性[42]
- Li等人提出部分感知变换器 (PAT) 来解决遮挡人物重识别问题，通过变换器编码器-解码器架构发现多样化的部分[19]
- DotSCN专注于群体重识别，并提出将个体的表示转移到目标G-ReID领域，并学习两个群体成员之间的新耦合表示以实现更好的区分能力[14]

[14] Z. Huang, Z. Wang, C.-C. Tsai, S. Satoh, and C.-W. Lin, "DotSCN: 1013 Group re-identification via domain-transferred single and couple representation learning," *IEEE Trans. Circuits Syst. Video Technol.* (2021).  
[19] Y. Li, J. He, T. Zhang, X. Liu, Y. Zhang, and F. Wu, "Diverse part discovery: Occluded person re-identification with part-aware transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2021).  
[35] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in *Proc. ECCV*, (2018).  
[36] H. Tan, X. Liu, Y. Bian, H. Wang, and B. Yin, "Incomplete descriptor mining with elastic loss for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.* (2022).  
[41] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Proc. 26th ACM Int. Conf. Multimedia* (2018).  
[42] J. Wang, Z. Wang, C. Gao, N. Sang, and R. Huang, "DeepList: Learning deep features with adaptive listwise constraint for person reidentification," *IEEE Trans. Circuits Syst. Video Technol.* (2017).  
[55] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian, "Person re-identification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2017).

- I2LT开发了一个模型来探索文本和图像特征之间的关系，并学习一个转换函数作为桥来直接传递模态间和模态内的标签来注释图像[31]
- Wu等人构建了SYSU-MM01数据集用于RGB-IR人物重识别[44]
- DMiR利用潜在的光谱不变但身份可辨的特征来进行对抗性解耦和模态不变表示学习[13]
- SDL通过非对抗性方法分离光谱信息，同时学习身份可辨特征[16]
- Wu等人利用联合模态和模式对齐网络（MPANet）发现不同模式中的跨模态细微差别以提取可辨特征[45]
- Zhong等人生成中间灰度图像作为辅助信息来给红外图像上色。[59]

➤ 与上述研究不同，我们旨在解决**无监督**RGB-IR跨模态跨域人员Re-ID问题，无需目标域的成对标签

[13] W. Hu, B. Liu, H. Zeng, Y. Hou, and H. Hu, “Adversarial decoupling and modality-invariant representation learning for visible-infrared person re-identification,” *IEEE Trans. Circuits Syst. Video Technol.* (2022).

[16] K. Kansal, A. V. Subramanyam, Z. Wang, and S. Satoh, “SDL: Spectrum-disentangled representation learning for visible-infrared person re-identification,” *IEEE Trans. Circuits Syst. Video Technol.* (2020).

[31] G.-J. Qi, W. Liu, C. Aggarwal, and T. Huang, “Joint intermodal and intramodal label transfers for extremely rare or unseen classes,” *IEEE Trans. Pattern Anal. Mach. Intell.* (2017).

[44] A. Wu, W.-S. Zheng, H.-X. Yu, S. Gong, and J. Lai, “RGB-infrared cross-modality person re-identification,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)* (2017).

[45] Q. Wu et al., “Discover cross-modality nuances for visible-infrared person re-identification,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2021).

[59] X. Zhong, T. Lu, W. Huang, M. Ye, X. Jia, and C.-W. Lin, “Grayscale enhancement colorization network for visible-infrared person re-identification,” *IEEE Trans. Circuits Syst. Video Technol.* (2022).

- DTNs跨异构域传输标记信息，并通过引入丰富的标签来缓解图像训练数据不足的问题[33]
- Tang等人构建了参数和表示共享层以生成特定于域和共享的跨域特征[37]
- M2M-GAN和PTGAN将源域图像转换为目标域风格[20][43]
- Pang等人提出了一种中值稳定聚类 and 全局距离分类方法，通过考虑聚类之间的相似性、聚类中的样本数量以及聚类内的组合相似性[29]
- Liu等人提出了一种UnityStyle适应方法，通过平滑同一摄像头内和不同摄像头之间的风格差异[22]
- Jing等人提出了一种跨域跨模态方法，用于人物搜索，使用时刻对齐模型[15]
- TAL-MIRN考虑了多视图信息推理和具有相同身份的行人图像的域差异[18]

➤ 与上述研究不同，我们所提出的DAN**同时进行域和模态对齐**，这更具挑战性，但对实际应用具有潜力

[15] Y. Jing, W. Wang, L. Wang, and T. Tan, "Cross-modal cross-domain moment alignment network for person search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2020).

[18] H. Li et al., "Triple adversarial learning and multi-view imaginative reasoning for unsupervised domain adaptation person re-identification," *IEEE Trans. Circuits Syst. Video Technol.* (2022).

[20] W. Liang, G. Wang, J. Lai, and J. Zhu, "M2M-GAN: Many-to-many generative adversarial transfer learning for person re-identification," (2018).

[22] C. Liu, X. Chang, and Y.-D. Shen, "Unity style transfer for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2020).

[29] Z. Pang, J. Guo, Z. Ma, W. Sun, and Y. Xiao, "Median stable clustering and global distance classification for cross-domain person re-identification," *IEEE Trans. Circuits Syst. Video Technol.* (2022).

[33] X. Shu, G.-J. Qi, J. Tang, and J. Wang, "Weakly-shared deep transfer networks for heterogeneous-domain knowledge propagation," in *Proc. 23rd ACM Int. Conf. Multimedia* (2015).

[37] J. Tang, X. Shu, Z. Li, G.-J. Qi, and J. Wang, "Generalized deep transfer networks for knowledge propagation in heterogeneous domains," *ACM Trans. Multimedia Comput., Commun., Appl.* (2016).

[43] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (2018).



# 方法

报告人：赵继承

问题：RGB-IR 跨模态跨域（cross-modal、cross-domain）行人再识别

- 域差异：不同数据集受光照、环境、数据采集方式等干扰因素的影响，存在较大的域偏移
- 模态差异：RGB突出颜色和纹理细节信息，IR突出热辐射信息
- 数据获取：获取带标签的成对RGB-IR图像十分困难

➡ 提出了一种 **RGB-IR 跨模态跨域行人 Re-ID 模型**

数据输入

源域：  $D^s = \{I_{p_i}^s, y_i^s\}, i \in [1, \dots, N^s]$

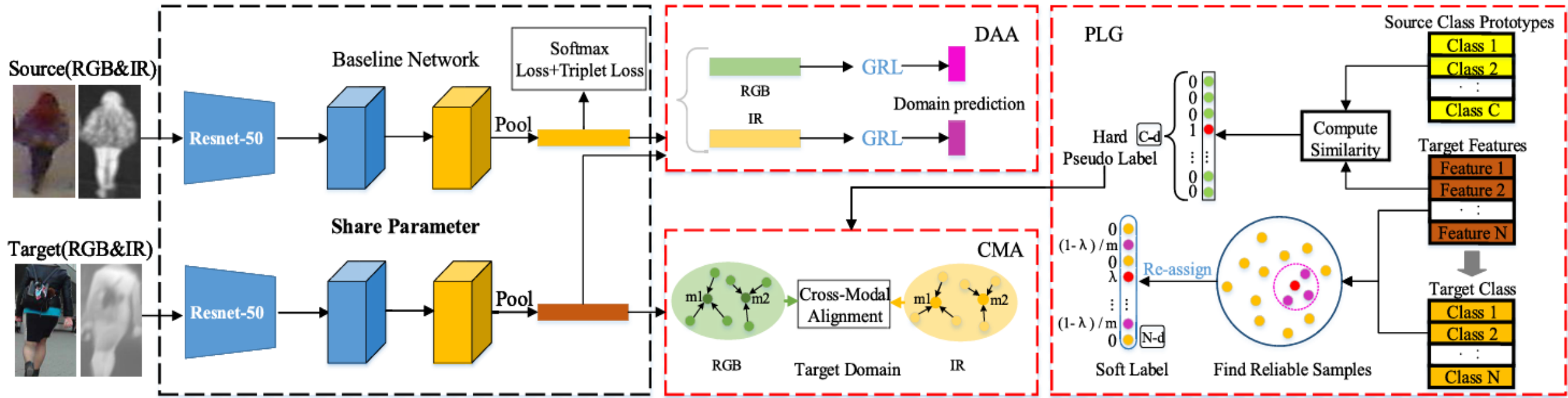
目标域：  $D^t = \{I_{p_i}^t\}, i \in [1, \dots, N^t]$



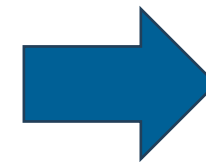
- 域不变特征提取（BN、DAA）
- 跨模态对齐（CMA）
- 无监督的目标域适应（PLG）



## Dual Alignment Network (DAN)



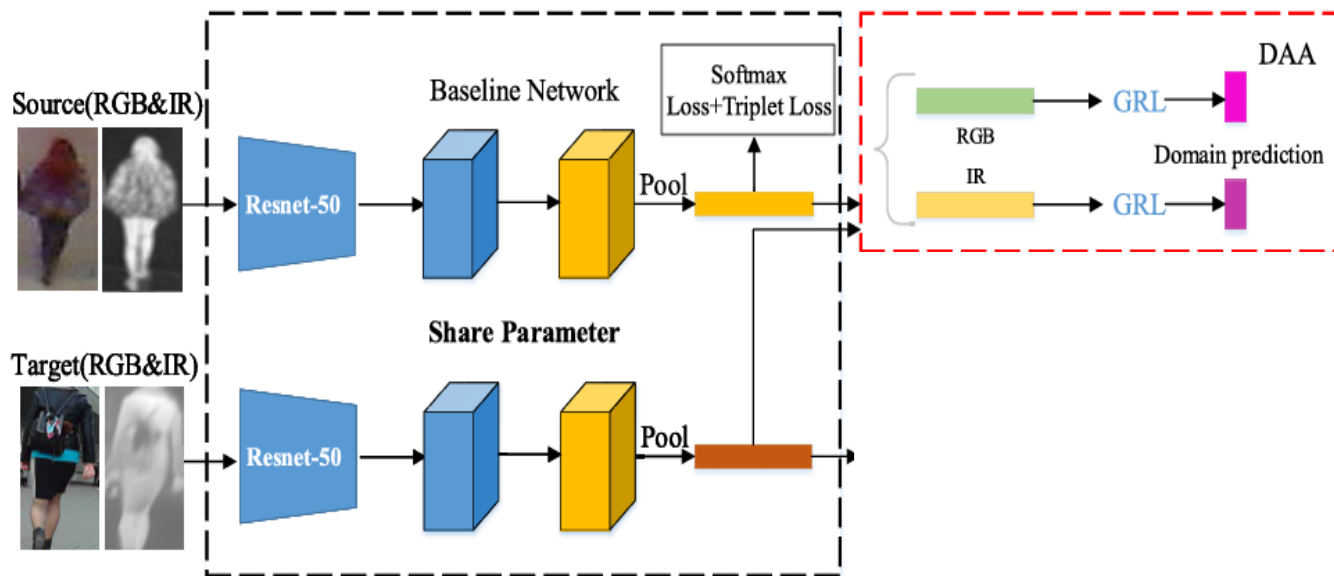
- Domain adversarial alignment(DAA): 降低域差异
- Cross-modal alignment(CMA): 降低模态差异
- Pseudo label generation(PLG): 为目标域数据生成伪标签



- 域不变性
- 模态不变性
- 无监督学习



## Domain adversarial alignment



鉴别器优化目标:

$$\mathcal{L}_{D_p}(\Theta_G, \theta_{D_p}; x_p^f) = -\frac{1}{N_p^s} \sum_{i=1}^{N_p^s} \log(D_p(x_{p_i}^s)) - \frac{1}{N_p^t} \sum_{j=1}^{N_p^t} \log(1 - D_p(x_{p_j}^t)) \quad (3)$$

最终优化目标:

$$\mathcal{L}_{DAA} = \max_{\Theta_G} \min_{\theta_{D_p}} \mathcal{L}_{D_p} \quad (4)$$

## A-distance

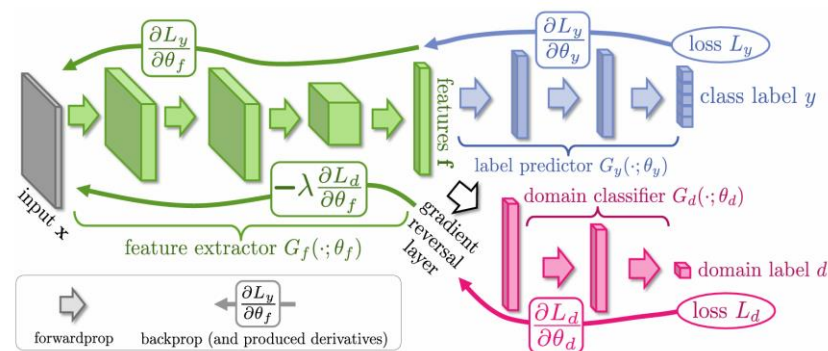
$$d_A(s, t) = 2 \left( 1 - 2 \min_{h \in \mathcal{H}} \text{err}(h(x)) \right) \quad (1)$$

$$h(x^s) \rightarrow 0 \quad h(x^t) \rightarrow 1$$

→ 域差异与分类误差成反比

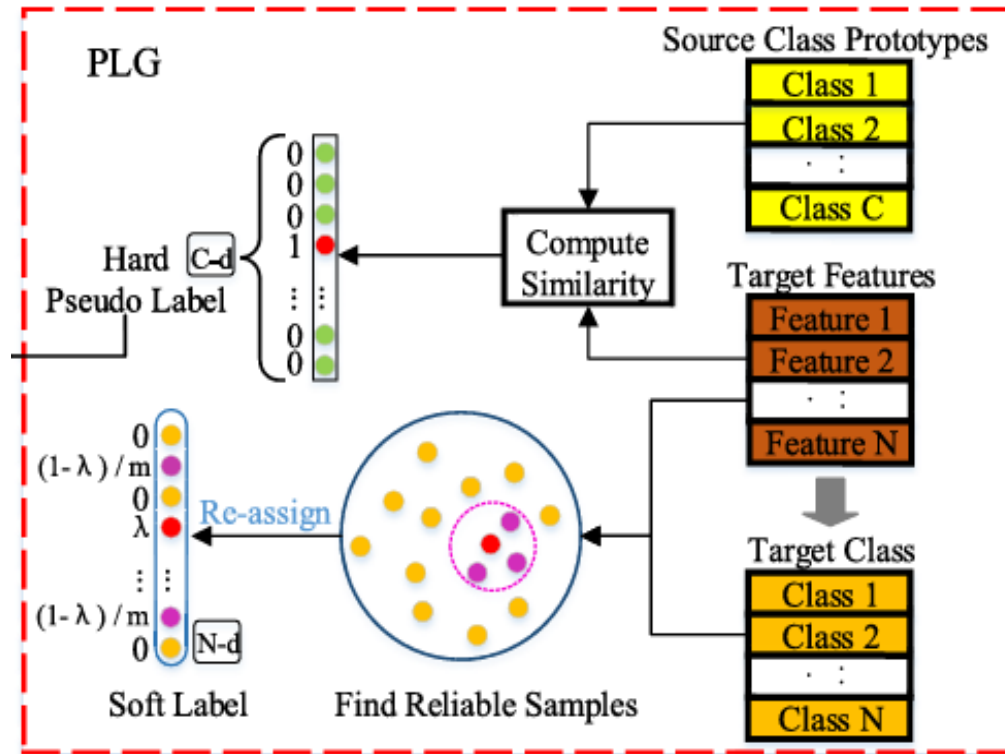
$$\begin{aligned} \min_G d_A(s, t) &\Leftrightarrow \min_G 2 \left( 1 - 2 \min_{h \in \mathcal{H}} \text{err}(h(x)) \right) \quad (2) \\ &\Leftrightarrow \max_G \min_{h \in \mathcal{H}} \text{err}(h(x)) \end{aligned}$$

## Gradient Reversed Layer(GRL)



- 反向传播，梯度方向反转，促进域不变特征学习

## Pseudo Label Generation



为目标域中未标记样本生成伪标签

- 硬伪标签：确切将目标归为一类
- 软伪标签：提供概率分布而不是硬性的类别表示

## 硬伪标签生成

- 计算源域RGB和IR样本的类特征原型

$$\mu_{r_k}^s = \frac{1}{N_{r_k}^s} \sum_{i=1}^{N_{r_k}^s} x_{r_i}^s, \quad \mu_{ir_k}^s = \frac{1}{N_{ir_k}^s} \sum_{i=1}^{N_{ir_k}^s} x_{ir_i}^s \quad (5)$$

$$k \in [1, \dots, C]$$

- 构建查找表存储类原型，计算目标域样本特征与查找表中类原型之间的余弦相似度：

$$y_{r_i}^t = \arg \max_{k \in [1, \dots, C]} (\cos(x_{r_i}^t, \mu_{r_k}^s)) \quad (6)$$

$$y_{ir_i}^t = \arg \max_{k \in [1, \dots, C]} (\cos(x_{ir_i}^t, \mu_{ir_k}^s)) \quad (7)$$

- 取最相似的类为硬伪标签，作为原始监督信号

## 软伪标签生成

- 使用索引为数据  $x_{p_i}^t$  定义特征的初始伪标签:

$$\{y_{p_i}^t = i \mid 1 \leq i \leq N_p^t\}$$

- 使用余弦距离来计算目标域中每个样本与其他样本在特征空间中的相似度，探索样本之间的相似关系。

- 为目标域中的每个样本找到具有相同类的样本为一组， $x_{p_i}^t$  的可靠样本集为:

$$\mathcal{V}_{p_i}^{\text{re}} = \{x_{p_{i1}}^t, x_{p_{i2}}^t, \dots, x_{p_{im}}^t\} \quad \mathcal{Y}_{p_i}^{\text{re}} = \{y_{p_{i1}}^t, y_{p_{i2}}^t, \dots, y_{p_{im}}^t\}$$

- 利用可靠样本集，重新分配软标签:

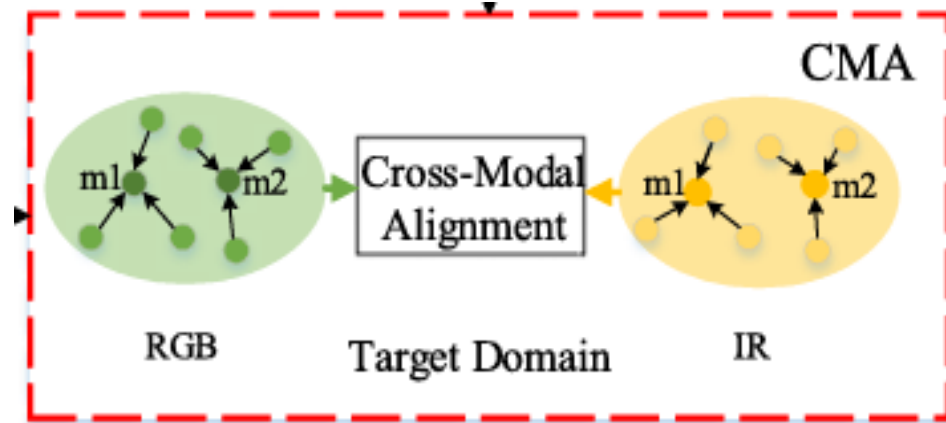
$$r(y_{p_j}^t) = \begin{cases} \lambda, & y_{p_j}^t = y_{p_i}^t \\ (1 - \lambda)/m, & y_{p_j}^t \in \mathcal{Y}_{p_i}^{\text{re}} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

- 重新分配样本标签后，伪标签的目标域数据不再是硬标签值 0 或 1，基于软标签分布特征，网络可以学习之间的相似性

$$\mathcal{L}_{\text{soft}} = \sum_{i=1}^{N^t} -\lambda \log \left( \frac{\exp(\mathbf{w}_i^T x_i^t)}{\sum_{j=1}^{N^t} \exp(\mathbf{w}_j^T x_i^t)} \right) - \frac{1 - \lambda}{m} \sum_{n=1}^m \log \left( \frac{\exp(\mathbf{w}_n^T x_i^t)}{\sum_{j=1}^{N^t} \exp(\mathbf{w}_j^T x_i^t)} \right) \quad (9)$$

其中  $\mathbf{w}$  是最后一个完全分类的参数连接层， $m$  是可靠特征的数量

## Cross-Modal Alignment



源域

分类损失:


$$\mathcal{L}_{cls} = \sum_{i=1}^{Y^s} -q_i^s \log(p_i^s) \begin{cases} q_i^s = 0, & y \neq i \\ q_i^s = 1, & y = i \end{cases} \quad (10)$$

三元组损失:

$$\mathcal{L}_{tri} = [d_p - d_n + \alpha]_+ \quad (11)$$

$d_p$  和  $d_n$  分别是正对和负对的特征距离

- 距是衡量变量分布和形态特征的统计度量

跨模态对齐  减少同类RGB 和 IR 之间的矩差

目标域

方差计算:

$$\sigma_{r_k}^t = \frac{1}{N_{r_k}^t} \sum_{i=1}^{N_{r_k}^t} (x_{r_i}^t - \mu_{r_k}^t)^2 \quad (12)$$

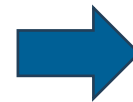
$$\sigma_{ir_k}^t = \frac{1}{N_{ir_k}^t} \sum_{i=1}^{N_{ir_k}^t} (x_{ir_i}^t - \mu_{ir_k}^t)^2 \quad (13)$$

跨模态对齐损失:

$$\mathcal{L}_{CMA} = \sum_{k=1}^C d(\mu_{r_k}^t, \mu_{ir_k}^t) + \eta d(\sigma_{r_k}^t, \sigma_{ir_k}^t) \quad (14)$$

对抗性对齐损失:

$$\mathcal{L}_{DAA} = \max_{\Theta_G} \min_{\theta_{D_p}} \mathcal{L}_{D_p} \quad (4)$$



消除不同域差异

目标域交叉熵损失:

$$\mathcal{L}_{soft} = \sum_{i=1}^{N^t} -\lambda \log \left( \frac{\exp(\mathbf{w}_i^T \mathbf{x}_i^t)}{\sum_{j=1}^{N^t} \exp(\mathbf{w}_j^T \mathbf{x}_i^t)} \right) - \frac{1-\lambda}{m} \sum_{n=1}^m \log \left( \frac{\exp(\mathbf{w}_n^T \mathbf{x}_i^t)}{\sum_{j=1}^{N^t} \exp(\mathbf{w}_j^T \mathbf{x}_i^t)} \right) \quad (9)$$



减少噪声硬标签引起损失

源域分类损失:

$$\mathcal{L}_{cls} = \sum_{i=1}^{Y^s} -q_i^s \log(p_i^s) \begin{cases} q_i^s = 0, & y \neq i \\ q_i^s = 1, & y = i \end{cases} \quad (10)$$



获取判别性的人物特征

源域三元组损失:

$$\mathcal{L}_{tri} = [d_p - d_n + \alpha]_+ \quad (11)$$

目标域跨模态对齐损失:

$$\mathcal{L}_{CMA} = \sum_{k=1}^C d(\mu_{r_k}^t, \mu_{ir_k}^t) + \eta d(\sigma_{r_k}^t, \sigma_{ir_k}^t) \quad (14)$$



确保同一类别内的样本在特征空间中彼此接近

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{tri} + \lambda_1 \mathcal{L}_{DAA} + \lambda_2 \mathcal{L}_{soft} + \lambda_3 \mathcal{L}_{CMA} \quad (15)$$



# 实验与分析

报告人：张熙林

- 不同跨模态数据集跨域评估是不可行的
- 跨域跨模态数据集稀少

## ◆ 添加随机噪声

$$x^t = x + r$$

## ◆ 调整光照强度

$$x^t = \text{Gamma}(\text{HSV}(\text{GF}(x)))$$

## ◆ 改变背景颜色

$$x'_r = IDCT(w_1 \times DCT(x_r) \times H(u, v) + m(u, v))$$

$$H(u, v) = 1 - e^{-\frac{u^2 + v^2}{2R^2}}$$

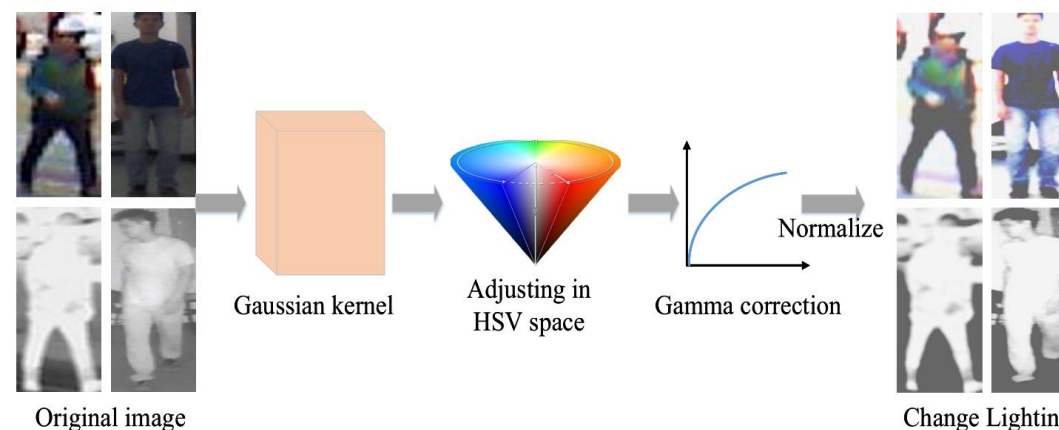


Fig. 5. Pipeline of lighting changes and comparison of image appearance before and after variation.



## 合成数据集的可行性

在单个真实数据集上表现良好的跨模态方法直接在合成的目标域上进行测试时，会观察到极差的性能。



Fig. 4. Taking SYSU-MM01 dataset (RGB-Near infrared) as an example, compared with the original data, the synthesized images (the 2nd row, the 3rd row and the 4th row) has a significant difference from the original images in appearance. Best viewed in color and zooming in.



	数据集	人员	数量	图像对
真实数据集	ThermalWorld	409	8125	可见光、热图像
	RegDB	412	8240	可见光、热图像
	SYSU-MM01	491	45863	可见光、近红外图像
合成数据集	Synthetic RegDB(random)	412	8240	可见光、热图像
	Synthetic RegDB(lightning)	412	8240	可见光、热图像
	Synthetic RegDB(color)	412	8240	可见光、热图像
	Synthetic SYSU-MM01(random)	491	45863	可见光、近红外图像
	Synthetic SYSU-MM01(lightning)	491	45863	可见光、近红外图像
	Synthetic SYSU-MM01(color)	491	45863	可见光、近红外图像

实验遵循无监督自适应的设置

骨干网络: ResNet-50



性能评估: mAP、Rank-1、Rank-5、Rank-10

初始学习率设为0.1。



动量衰减和权值衰减分别设为0.9和 $2e^{-4}$ 。

## 任务

### 跨真实域和真实域

RegDB  ThermalWorld  
ThermalWorld  合成 RegDB

### 跨真实域和合成域

RegDB  合成 RegDB  
SYSU-MM01  合成 SYSU-MM01

## 验证DAN的有效性

- 探索不同损失函数在跨模态跨域设置下的性能

Circle Loss、Instance Loss、Contrastive Loss、Sphere Loss

- 将DAN与SOTA跨模态人员重识别的方法进行比较

DDAG、DGTL、HcTri、MGMRA、CMNAS、LbA、CAJ

- 将DAN与典型的无监督域自适应(UDA)方法及其他域偏移减少的人员再识别方法进行比较

ADDA、GVB、DBSCAN、UNRN、Rectifying

## 跨真实域和真实域

在跨模态的人重新识别任务中，域差异对人员再识别的表现产生了灾难性的影响

DAN模型在与跨模态或跨域的方法的比较中取得更好的结果

表1.TW和RegDB之间的跨域任务性能比较

Method	TW → RegDB			RegDB → TW		
	R-1	R-5	mAP	R-1	R-5	mAP
Target Model	66.89	76.70	62.01	39.77	58.42	23.74
baseline	13.64	21.33	14.47	12.95	25.69	8.13
Circle Loss[34]	3.65	6.74	5.46	4.04	12.27	5.57
Instance Loss[57]	9.04	14.91	10.64	11.70	25.74	6.42
Contrast Loss[58]	6.46	10.46	7.99	8.65	18.64	6.14
Sphere Loss[25]	9.74	15.47	11.26	7.66	20.09	6.52
DDAG[50]	14.07	21.48	15.00	9.43	22.63	6.25
HcTri[24]	6.31	11.02	8.34	8.75	25.48	7.48
MGMRA[4]	8.17	13.94	9.29	12.33	28.28	8.05
CMNAS[2]	8.39	23.56	6.43	8.18	23.72	6.50
DGTL[23]	10.38	17.05	10.64	9.89	22.48	6.19
LbA[30]	4.95	9.01	6.06	16.73	32.78	8.49
CAJ[49]	8.79	16.76	10.05	6.99	18.38	7.66
ADDA[38]	5.19	11.70	6.02	10.31	25.32	7.10
GVB[8]	5.29	10.87	6.16	11.24	27.08	7.93
DBSCAN[10]	10.21	16.02	11.62	14.71	27.60	9.57
UNRN[54]	12.27	20.05	12.87	15.38	32.52	7.36
Rectifying[56]	11.38	18.98	11.03	18.85	34.23	10.53
DAN(ours)	<b>20.41</b>	<b>30.53</b>	<b>19.43</b>	<b>21.65</b>	<b>36.41</b>	<b>11.53</b>

## 跨真实域和合成域

真实数据和合成数据之间的域差异对跨模态人员重新识别性能具有致命影响

### 对颜色鲁棒性较差

在通过添加噪声和改变光照获得的合成目标域上，不同的损失函数具有相似的性能。但是颜色变化的合成域上，这些损失函数表现不佳。

表2. RegDB到合成的RegDB目标域的跨域任务性能比较

Method	RegDB→Synthetic RegDB (random)				RegDB→Synthetic RegDB (lighting)				RegDB→Synthetic RegDB (color)			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
Target Model	54.90	69.47	77.38	49.94	54.71	68.01	74.13	49.58	50.49	65.15	73.83	44.68
baseline	7.28	11.46	15.83	8.06	4.56	8.74	11.99	5.67	10.19	19.03	23.93	10.95
Circle Loss[34]	3.59	8.16	12.52	3.71	4.95	10.73	14.47	5.66	1.31	3.35	5.78	2.25
Instance Loss[57]	5.58	11.31	14.37	5.53	5.39	10.44	15.00	5.65	4.37	7.77	10.87	5.10
Contrast Loss[58]	5.58	10.83	14.95	5.09	6.60	11.89	16.55	6.97	3.35	5.97	8.11	4.53
Sphere Loss[25]	5.87	10.53	13.54	5.46	7.23	13.25	17.57	5.94	5.53	10.24	13.74	4.90
DDAG[50]	7.28	13.64	20.83	6.78	11.21	21.12	28.11	11.31	6.50	12.86	17.18	7.72
HeTri[24]	2.09	5.83	9.17	2.96	16.46	30.39	38.93	16.04	9.95	17.04	21.75	10.13
MGMRA[4]	3.30	7.96	11.46	3.49	18.93	31.55	40.10	17.67	8.25	15.58	21.21	9.79
CMNAS[2]	0.68	3.25	5.49	1.46	11.89	21.07	27.23	12.05	10.44	18.74	24.32	10.99
DGTL[23]	14.13	24.32	28.88	10.16	11.49	21.31	27.23	10.95	9.42	16.41	21.36	8.88
LbA[30]	19.71	33.64	40.78	15.76	14.17	21.80	26.75	13.07	7.18	13.35	18.20	7.93
CAJ[49]	0.73	1.70	2.96	1.21	8.20	16.36	25.68	9.33	6.26	12.43	17.48	7.40
ADDA[38]	6.21	14.08	20.58	5.05	3.11	5.92	9.76	3.99	3.83	7.67	10.68	4.54
GVb[8]	7.91	15.92	22.18	7.17	5.19	9.51	12.86	6.47	5.53	10.53	15.05	5.84
DBSCAN[10]	17.91	31.70	39.08	15.17	9.47	17.62	22.62	10.42	12.82	23.01	29.51	13.39
UNRN[54]	16.94	30.49	38.25	14.77	12.38	18.64	22.91	11.04	8.59	13.69	17.48	8.19
Rectifying[56]	9.08	19.13	25.53	9.01	12.62	19.71	24.32	11.65	7.77	13.54	17.96	7.40
DAN(ours)	29.13	45.63	53.16	23.29	22.82	34.22	42.04	22.14	26.17	38.74	46.36	24.92

## 跨真实域和合成域

CAJ在SYSU-MM01→合成SYSU-MM01(lightning)任务上取得了比DAN稍好的性能

整体来说，DAN模型在合成的目标域上进行跨模态跨域人再识别任务中表现出最好的性能

表3. SYSU-MM01到合成的SYSU-MM01目标域的跨域任务性能比较

Method	SYSU→Synthetic SYSU (random)				SYSU→Synthetic SYSU (lighting)				SYSU→Synthetic SYSU (color)			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
Target Model	33.26	61.68	74.39	34.04	31.80	56.69	72.61	32.49	40.43	69.97	81.46	39.52
baseline	14.98	37.59	52.16	16.98	8.95	25.72	37.76	11.22	12.36	29.18	39.89	10.77
Circle Loss[34]	14.37	33.84	46.50	16.01	6.18	19.31	30.14	8.85	2.80	9.82	16.50	4.14
Instance Loss[57]	14.05	34.70	47.77	15.22	7.48	22.51	34.46	9.82	3.95	12.77	20.52	5.38
Contrast Loss[58]	13.79	37.08	51.61	16.65	8.53	26.83	40.84	12.27	5.42	16.59	25.50	6.36
Sphere Loss[25]	9.82	25.28	35.85	10.77	7.38	22.41	34.63	9.73	3.95	12.08	19.82	4.91
DDAG[50]	11.87	30.57	42.43	12.76	9.71	27.60	40.78	11.81	11.73	26.48	35.63	9.62
HcTri[24]	16.64	36.65	49.65	17.36	11.33	27.08	37.63	12.00	12.99	27.00	34.68	9.93
MGMRA[4]	11.05	26.94	38.16	11.39	11.30	29.60	42.16	12.48	8.51	17.19	24.19	6.75
CMNAS[2]	4.47	15.17	24.97	6.57	5.10	18.25	29.71	7.72	5.55	14.18	21.41	5.51
DGTL[23]	17.93	37.23	48.89	17.61	13.29	34.09	47.00	14.80	14.27	31.01	40.44	11.06
LbA[30]	19.65	42.70	56.22	19.92	13.09	34.68	48.72	14.66	4.74	16.11	25.55	6.08
CAJ[49]	20.60	44.79	58.47	20.94	<b>14.41</b>	<b>35.92</b>	<b>49.38</b>	<b>16.37</b>	9.68	25.03	35.53	9.41
ADDA[38]	16.41	42.31	56.98	17.52	12.42	32.58	46.32	14.28	16.83	40.38	55.11	16.23
GVB[8]	<b>19.56</b>	48.04	62.19	20.29	<b>11.78</b>	31.66	45.41	13.55	<b>14.63</b>	36.33	49.83	14.59
DBSCAN[10]	16.75	37.58	50.66	17.64	6.90	20.35	31.41	8.73	7.64	20.83	29.77	7.49
UNRN[54]	18.38	43.20	56.64	17.61	11.83	29.42	41.89	13.74	4.97	17.70	25.53	6.45
Rectifying[56]	18.01	41.41	56.32	18.48	8.44	24.35	36.42	9.57	3.81	10.99	16.83	4.67
DAN(ours)	<b>25.45</b>	<b>53.96</b>	<b>68.13</b>	<b>25.08</b>	13.49	34.84	49.08	15.35	<b>19.40</b>	<b>43.92</b>	<b>57.18</b>	<b>17.41</b>

$\lambda_1, \lambda_2, \lambda_3$  为权衡参数

$\lambda$  用于在生成伪标签的同时平衡伪标签和可靠类的效果,  $m$  为可靠样本的数量

$\eta$  是一个平衡CMA中不同矩的超参数

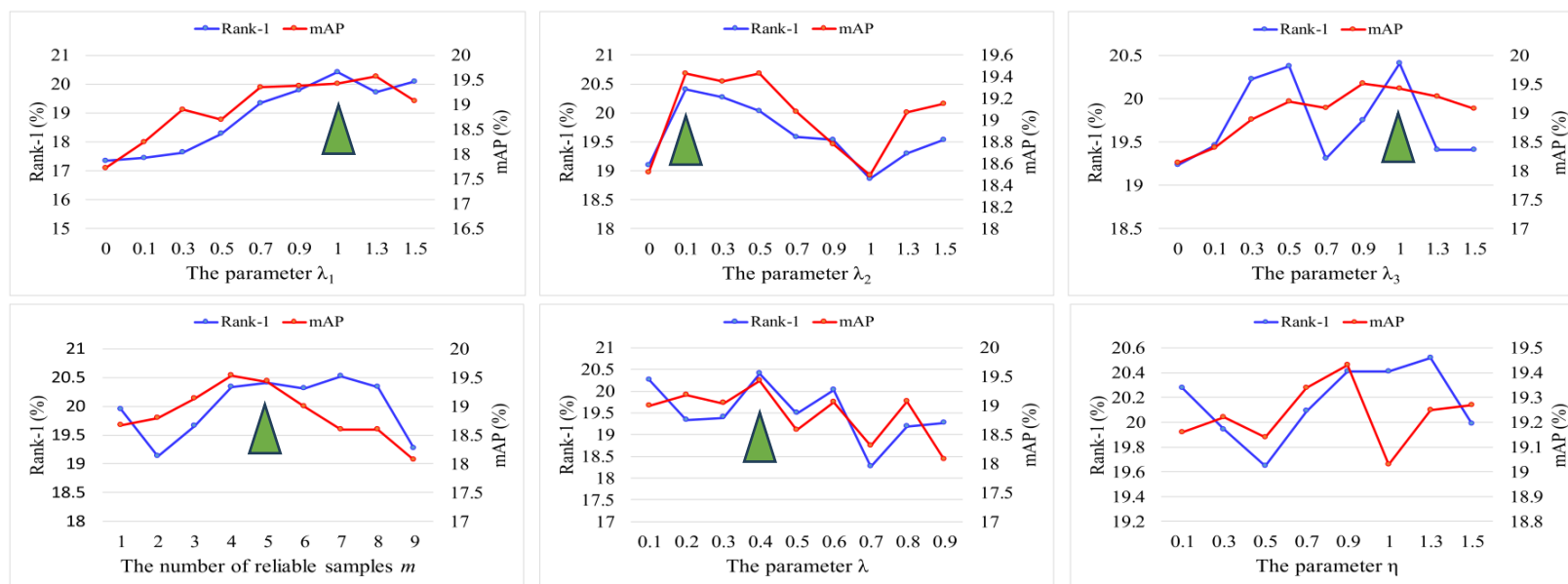


Fig. 6. Hyper-parameters analysis on ThermalWorld→RegDB task.  $\lambda_1, \lambda_2$  and  $\lambda_3$  are trade-off parameters in model Eq.(15).  $m$  is the number of reliable samples and  $\lambda$  balances the effects of the pseudo label and reliable classes in Eq.(9).  $\eta$  balances the effects of the moments in CMA of Eq.(14).

为验证DAN中各组成部分的有效性，进行了消融实验。

这三个模块可以协同学习领域不变和语义对齐的跨模态表示，以提高性能。

表4.在 thermalworld  $\rightarrow$  regdb 任务中使用提出的DAN方法中不同组分的消融分析

Method	Rank-1	Rank-5	mAP
baseline	13.64	21.33	14.47
baseline+DAA	19.08	28.51	18.06
baseline+Softmax (hard label)	15.67	24.99	16.73
baseline+Softmax (soft label)	17.08	26.74	17.45
baseline+CMA	17.18	25.44	17.67
baseline+DAA+CMA+soft label	<b>20.41</b>	<b>30.53</b>	<b>19.43</b>



目标域Top-10排序结果。绿框表示正确结果，红框表示不正确结果。

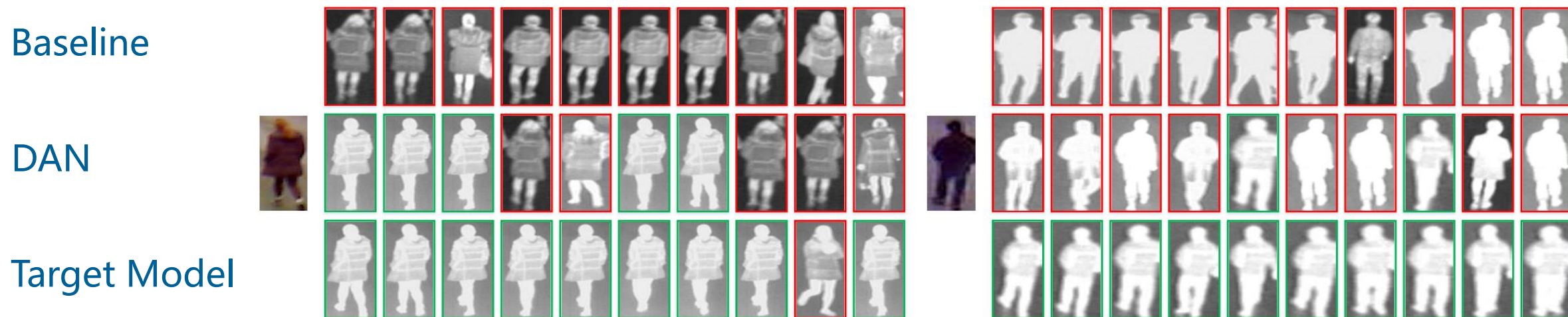


Fig. 7. Visualization of retrieval results of target domain on ThermalWorld→RegDB task. The green boxes indicate correct results and the red boxes are incorrect. The first line is a list of the results of the baseline model without adaptation, the middle row shows the results of our DAN, and the bottom shows the results of the Target Model (the upper bound) trained by ground-truth target labels. Best viewed in color.

添加随机颜色、大小、位置的矩形进行遮挡

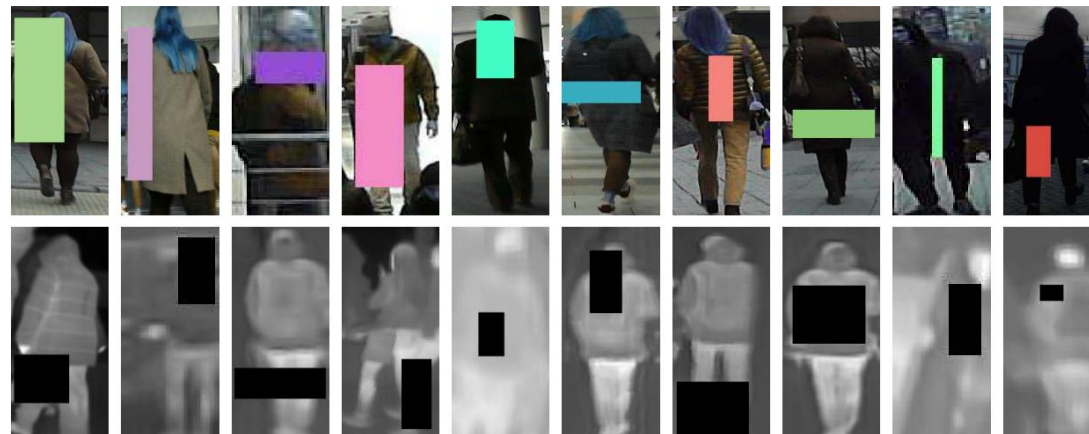


Fig. 8. Randomly occluded images on the ThermalWorld dataset. The first row is RGB-modality data with randomly block occlusions and the second row is Thermal-modality data with randomly block occlusions.

DAN可以显著提高模型对遮挡的适应性。

表5.在RegDB和ThermalWorld上随机遮挡时进行跨域实验的性能表现

Method	RegDB (O)			ThermalWorld (O)		
	R-1	R-5	mAP	R-1	R-5	mAP
baseline	30.68	47.48	24.63	21.96	39.20	11.20
DAN(ours)	<b>47.18</b>	<b>63.40</b>	<b>37.31</b>	<b>27.24</b>	<b>50.75</b>	<b>13.65</b>

- 论文首次尝试对RGB-红外跨模态跨域的人员重新识别进行研究。
- 我们提出了一个新的双对齐网络，它包括域对抗对齐组件、目标域伪标签生成模块和跨模态对齐组件。这三个模块互相配合，促进模型学习域不变和模态不变的人表示。
- 在真实域和合成域上部署了跨域跨模态实验，验证了所提出的方法在各种因素引起的域差异下的有效性。

**Thanks**  
**Q & A**