Weakly Supervised Person Re-Identification

Jingke Meng, Sheng Wu, Wei-Shi Zhen

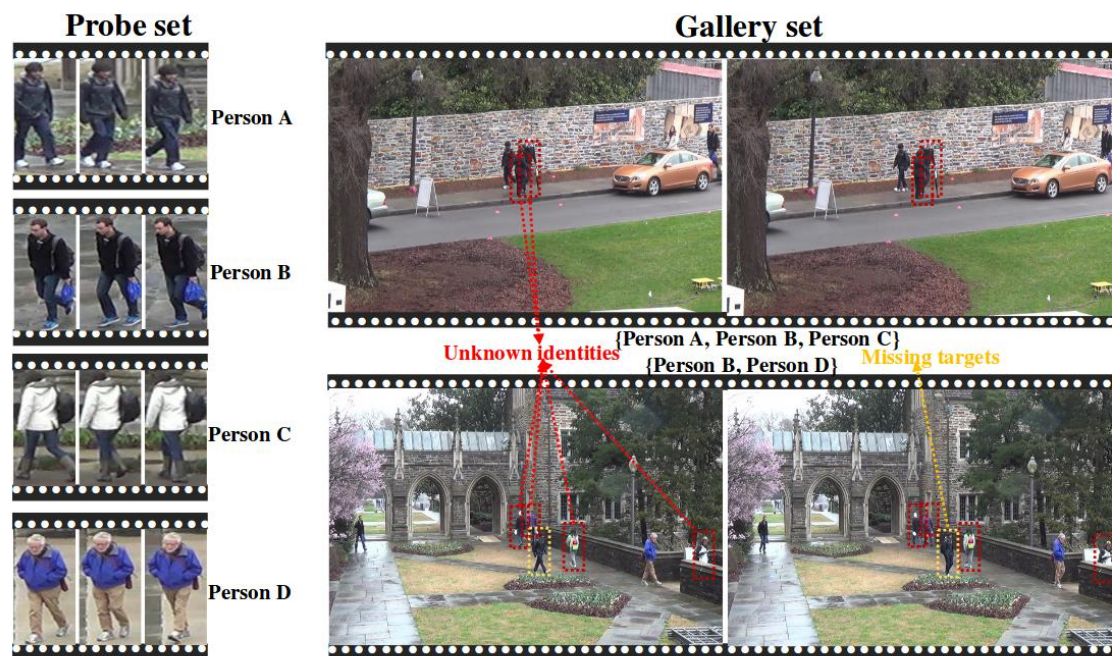Reference: https://arxiv.org/pdf/1904.03832.pdf

引言

Conventional Video Person Re-id

Proposed Video Person Re-id

Probe set

Gallery set

Person A   Person B   Person A   Person B

(a) Conventional fully supervised setting

Probe set   Gallery set

Person A

Person B

{Person A, Person B, Person C}
Unknown identities   {Person B, Person D}   Missing targets

Person C

Person D

(b) Proposed weakly supervised setting

是谁?
每一帧 img = ??

存在谁?
{?, ??} in video

来源于多实例多标签学习(MIML)

一段视频由很多张图组成，假如10000张，那么我们要判断视频里是否包含某一物体，比如气球。单张标注每一帧是否有气球**太耗时，**通常人们**看一遍说这个视频里是否有气球，**就得到了**多实例学习**的数据。

10000帧的数据不是每一个都有气球出现，只要有一帧有气球，那么我们就认为这个数据包是有气球的。只有当所有的视频帧都没有气球，才是没有气球的。

从这里面学习哪一段视频（10000张）是否有气球出现就是多实例学习的问题。

--zhihu: 别拽我红领巾

那么多实例多标签就是 **一段视频里面是否包含几个物体**

对应到行人重识别一段视频，也就是一个包（bag）中，
就是这个bag 里面有谁 谁 谁

Train:

probe set

gallery set

Person D

{Person B, Person D}

Noise: 1. 由于标注者的一些小失误，可能这包里面不只有B/D两个人。
2.或者明明包里面有B这个人，但是标注者没有标注

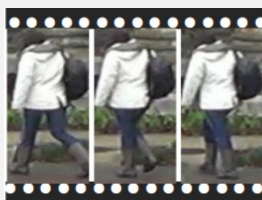**Test:** probe set ＋ gallery set

由于整体的目标是减少打标量。
1. 所以对于测试集的gallery set 的每个包也只是弱的标签。只是说存在谁。并没有抠出bouding box 并说明属于谁
2. 所以在测试集的查找命中为： 找到probe 里面的人在哪个包出现即可

How to do it

1. Mask RCNN将包中的行人检测出来（如果数据集是原始Video）

2. 既然是MIML(多实例多标签学习)的思想迁移过来的，那就先用**MIML的思想**

For probe set every bag:

确定类别

$$\mathcal{L}_p = \frac{1}{N_p} \sum_{X_b \in \mathcal{X}_p} \sum_{i \in \{1, \cdots, n_b\}} \sum_{c \in \{0, \cdots, C\}} (-y_b^c \log \tilde{y}_{b,i}^c),$$

For gallery set every bag:

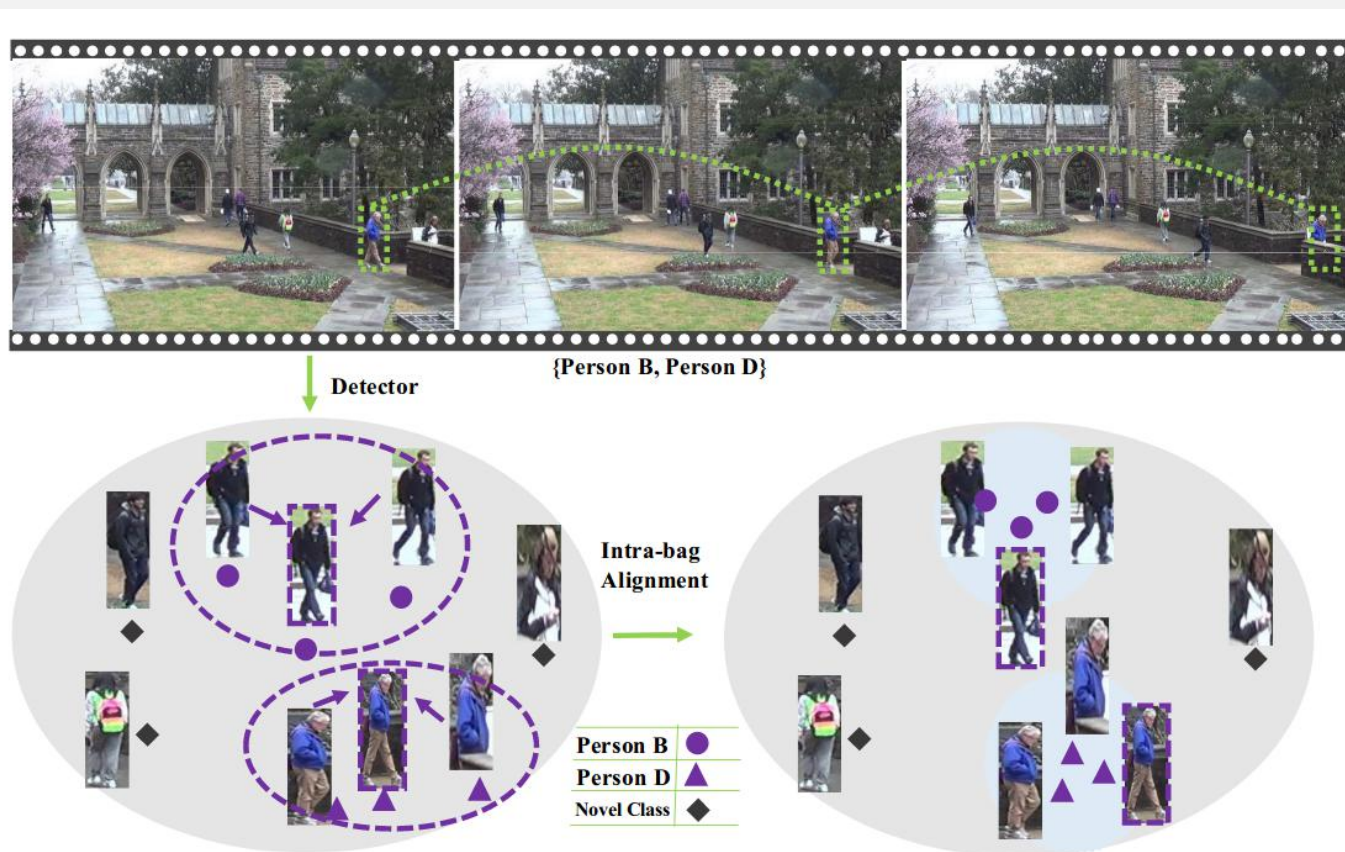

不确定类别

$$\mathcal{L}_g = \frac{1}{N_g} \sum_{X_b \in \mathcal{X}_g} \sum_{c \in \{0, \cdots, C\}} (-y_b^c \log \max\{\tilde{y}_{b,1}^c, \tilde{y}_{b,2}^c, ..., \tilde{y}_{b,n_b}^c\}),$$

最大先验概率

$$\mathcal{L}_C = \mathcal{L}_p + \mathcal{L}_g.$$

How to do it

3. Intra-bag Alignment　前面没有考虑跨摄像头信息

一个人通常出现在包里面的连续几帧，应该group起来

- **Group:**

- 确定中心点：一个batch中，有几张图片属于C类，其中一张属于C类**概率最大**的图片，作为**seed instance**，类似中心。
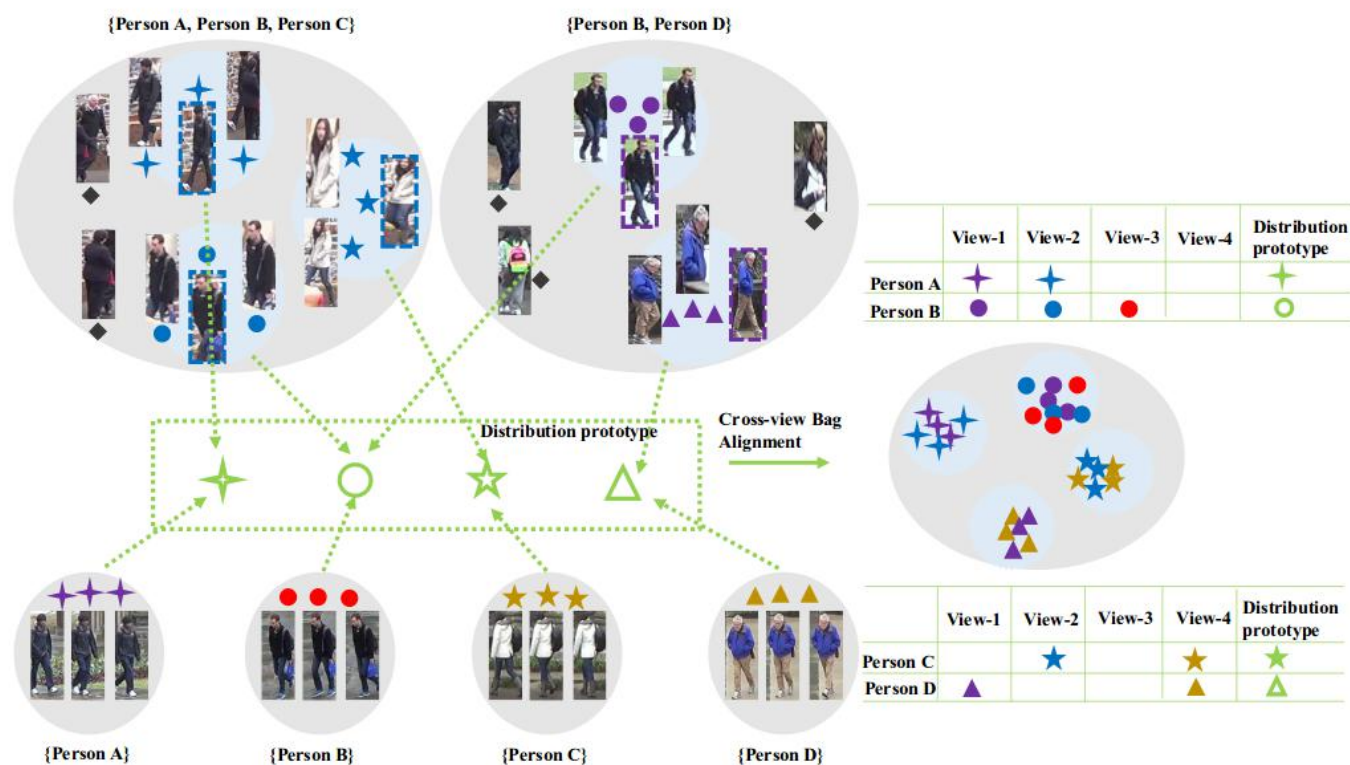- 某些C类图片的概率向seed instance靠近。

$$\mathcal{L}_{IA} = \frac{1}{N_{IA}} \sum_{X_b \in \mathcal{X}_g} \sum_{c \in \{0, \cdots, C\}} \sum_{p \in \mathcal{G}_{b,c}} y_b^c D_{KL}(\tilde{\mathbf{y}}_{b,p} \| \tilde{\mathbf{y}}_{b,q_c}),$$

group 有风险:

$$\mathcal{G}_{b,c} = \{p | \mathbf{x}_{b,p} \in \mathcal{N}_{q_c} \text{ and } \tilde{y}_{b,p}^c \geq \gamma \tilde{y}_{b,q_c}^c\}$$

How to do it

3. Cross-view Bag Alignment

- Group:

- 确定中心点：

$$\mathbf{p}_c^t = \frac{1}{|\mathcal{V}_c|} \sum_{v \in \mathcal{V}_c} \left(\frac{1}{|\mathcal{I}_{c,v}|} \sum_{i \in \mathcal{I}_{c,v}} \tilde{\mathbf{y}}_i \right),$$

$$\hat{\mathbf{p}}_c^t = \alpha \hat{\mathbf{p}}_c^{t-1} + (1 - \alpha) \mathbf{p}_c^t,$$

- 拉近距离：

$$\mathcal{L}_{CA} = \frac{1}{N_{CA}} \sum_{X_b \in (\mathcal{X}_p \bigcup \mathcal{X}_g)} \sum_{c \in \{0, \cdots, C\}} \sum_{i \in \mathcal{I}_c} y_b^c D_{KL}(\tilde{\mathbf{y}}_{b,i} \| \hat{\mathbf{p}}_c^t),$$

跨摄像头同一个人的图片特征应该在同一个**子空间**

How to do it

3. Deep Cross-view MIML Model

$$\mathcal{L}_E = \frac{1}{N_E} \sum_{X_b \in \mathcal{X}_g} \sum_{i \in \{1,\cdots,n_b\}} \sum_{c \in \{0,\cdots,C\}} (-\tilde{y}_{b,i}^c \log \tilde{y}_{b,i}^c).$$

$$\mathcal{L}_{CV-MIML} = \mathcal{L}_C + \delta(\mathcal{L}_{IA} + \mathcal{L}_{CA} + \mathcal{L}_E)$$

Experiments

Dataset	# Camera views	# Identities (training/testing)	# Training BBoxes (probe/gallery)	# Testing BBoxes (probe/gallery)
WL-DukeMTMC-REID	8	880/1695	60,267/923,879	116,128/904,066
WL-PRID2011	2	100/100	11,201/8,191	12,129/8,512
WL-iLIDS-VID	2	150/150	9,731/11,278	12,129/8,512
WL-MARS	6	631/630	38,324/460,236	36,988/472,978

Table 1. Detailed information of the one genuine and three new simulated datasets for the weakly supervised person re-id.

Experiments

WL-DukeMTMC-REID	r=1	r=5	r=10	r=20	mAP
CV-MIML	**78.05**	**90.50**	**93.75**	**95.99**	**59.53**
baseline + IA	74.69	88.50	92.15	94.81	56.97
baseline+CA	72.92	87.96	92.04	94.75	55.30
baseline+entropy	70.56	85.90	90.15	92.68	53.05
baseline	67.26	84.90	89.50	92.68	50.96
WL-PRID2011	r=1	r=5	r=10	r=20	mAP
CV-MIML	**72.00**	**89.00**	95.00	**99.00**	**70.78**
baseline+IA	69.00	**89.00**	93.00	98.00	65.89
baseline+CA	68.00	87.00	**96.00**	98.00	63.72
baseline+entropy	70.00	**89.00**	**96.00**	**99.00**	67.32
baseline	67.00	86.00	95.00	97.00	62.87
WL-iLIDS-VID	r=1	r=5	r=10	r=20	mAP
CV-MIML	**60.00**	80.00	87.33	**96.67**	**56.01**
baseline+IA	55.33	**80.67**	**89.33**	95.33	53.78
baseline+CA	52.67	78.00	88.00	95.33	50.58
baseline+entropy	44.67	69.33	81.33	92.67	44.99
baseline	41.33	70.00	83.33	94.67	42.26
WL-MARS	r=1	r=5	r=10	r=20	mAP
CV-MIML	**66.88**	**82.02**	**87.22**	**91.48**	**55.16**
baseline+IA	62.15	80.44	85.80	89.75	50.27
baseline+CA	63.09	79.97	84.23	88.96	50.61
baseline+entropy	60.88	79.34	85.49	89.43	49.13
baseline	53.47	71.77	79.02	85.49	40.31

Table 2. Ablation study of the proposed CV-MIML method. The matching accuracy values (%) at rank(r) = 1, 5, 10, 20 and mAP are shown on the four datasets. The best results are shown in black boldface font.

Experiments

WL-DukeMTMC-REID	r=1	r=5	r=10	r=20	mAP
MIMLfast[16]	13.63	44.66	55.69	64.78	10.05
DeepMIML[10]	65.37	82.30	86.90	90.68	48.02
Deep CV-MIML	**78.05**	**90.50**	**93.75**	**95.99**	**59.53**
WL-PRID2011	r=1	r=5	r=10	r=20	mAP
MIMLfast[16]	29.00	56.00	72.00	87.00	31.66
DeepMIML[10]	67.00	**90.00**	94.00	99.00	61.80
Deep CV-MIML	**72.00**	89.00	**95.00**	**99.00**	**70.78**
WL-iLIDS-VID	r=1	r=5	r=10	r=20	mAP
MIMLfast[16]	28.00	58.67	69.33	78.67	27.42
DeepMIML[10]	44.00	70.00	81.33	89.33	43.49
Deep CV-MIML	**60.00**	**80.00**	**87.33**	**96.67**	**56.01**
WL-MARS	r=1	r=5	r=10	r=20	mAP
MIMLfast[16]	20.50	37.22	43.06	52.05	11.39
DeepMIML[10]	47.16	70.19	76.18	81.07	36.59
Deep CV-MIML	**66.88**	**82.02**	**87.22**	**91.48**	**55.16**

Table 3. Comparison with state-of-the-art MIML methods. The best results are in black boldface font.

WL-DukeMTMC-REID	r=1	r=5	r=10	r=20	mAP
CAMEL [40]	0.53	0.77	1.18	3.24	0.90
PUL[8]	-	-	-	-	-
EUG[32]	35.93	50.74	59.41	66.96	21.94
Deep CV-MIML	78.05	90.50	93.75	95.99	59.53
PCB[29]	79.82	90.38	93.45	96.17	62.09
WL-PRID2011	r=1	r=5	r=10	r=20	mAP
CAMEL [40]	2.00	11.00	20.00	44.00	4.59
PUL[8]	32.00	58.00	71.00	85.00	35.28
EUG[32]	55.00	83.00	93.00	97.00	53.26
Deep CV-MIML	72.00	89.00	95.00	99.00	70.78
PCB[29]	88.00	97.00	99.00	99.00	87.35
WL-iLIDS-VID	r=1	r=5	r=10	r=20	mAP
CAMEL [40]	4.67	16.00	26.67	43.33	6.26
PUL[8]	20.00	44.00	59.33	76.00	22.56
EUG[32]	26.67	60.67	72.00	86.67	29.86
Deep CV-MIML	60.00	80.00	87.33	96.67	56.01
PCB[29]	72.00	89.33	92.67	96.00	69.87
WL-MARS	r=1	r=5	r=10	r=20	mAP
CAMEL [40]	0.32	1.10	2.52	5.52	0.56
PUL[8]	-	-	-	-	-
EUG[32]	25.87	39.59	46.21	55.21	15.63
Deep CV-MIML	66.88	82.02	87.22	91.48	55.16
PCB[29]	68.14	84.07	86.28	90.54	54.18

Table 4. Comparison with related re-id methods. The $1^{st}/2^{nd}$ best results are indicated in red/blue.

Thanks!!!~