Paper Reading

李星泽 2018-12-22

Outline

- Introduction to Person Re-identification
- Part-Aligned Bilinear Representations for Person Re-id(ECCV18)
- End-to-End Deep Kronecker-Product Matching (CVPR 18)
- Spatial-Temporal Person Re-id (AAAI 19)

Person Re-identification

Identification

- Trainset and testset have same labels.

Re-identification

- Trainset and testset have different labels.
- It can be viewed as a retrieval task.
- Give a probe, calculate the similarities between probe and gallery, and arrange the gallery in descending order.

Person Re-identification

CMC(Cumulative Matching Characteristics)

- CMC-K is the percentage that there exists matched person in the top-K similar person.

MAP(Mean Average Precision)

- K matched person, their ranking positions are P_i

$$AP = \frac{1}{K}\sum_{i=1}^{K}\frac{i}{P_i}$$

probe

ranking results

precision	1/1	2/2	2/3	2/4	3/5

AP = (1/1+2/2+3/5) / 3 = 13/15

Part-Aligned Bilinear Representations for Person Re-identification

Yumin Suh[1], Jingdong Wang[2], Siyu Tang[3,4], Tao Mei[5], and Kyoung Mu Lee[1]

[1] ASRI, Seoul National University, Seoul, Korea
[2] Microsoft Research Asia, Beijing, China
[3] Max Planck Institute for Intelligent Systems, Tübingen, Germany
[4] University of Tübingen, Tübingen, Germany
[5] JD AI Research, Beijing, China
{n12345,kyoungmu}@snu.ac.kr, jingdw@microsoft.com,
stang@tuebingen.mpg.de, tmei@jd.com

Motivation:

Body part misalignment is one of the key challenges in person re-id. State-of-the-art pose estimation solutions are still not perfect. Also, bounding box-based schemes lack fine-grained part localization within the boxes. Our approach learns to represent the human poses as part maps and combine them directly with the appearance maps to compute part-aligned representations.

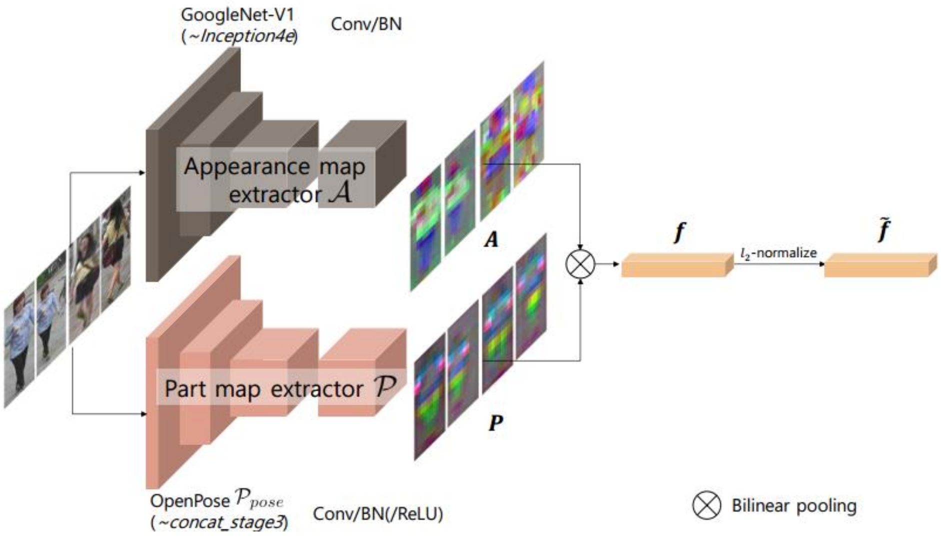
Part-Aligned Bilinear Representation

$$\mathbf{f} = \text{pooling}_{xy}\{\mathbf{f}_{xy}\} = \frac{1}{S}\sum_{xy}\mathbf{f}_{xy}, \qquad \mathbf{f}_{xy} = \text{vec}(\mathbf{a}_{xy} \otimes \mathbf{p}_{xy}),$$

$$\text{vec}(\mathbf{a} \otimes \mathbf{p}) = [(p_1\mathbf{a})^\top \ (p_2\mathbf{a})^\top \ \ldots (p_{c_P}\mathbf{a})^\top]^\top,$$

$$\tilde{\mathbf{f}} = \frac{\mathbf{f}}{\|\mathbf{f}\|_2}. \qquad \tilde{\mathbf{a}}_{xy} = \frac{\mathbf{a}_{xy}}{\sqrt{\|\mathbf{f}\|_2}} \text{ and } \tilde{\mathbf{p}}_{xy} = \frac{\mathbf{p}_{xy}}{\sqrt{\|\mathbf{f}\|_2}}.$$

$$\tilde{\mathbf{f}}_{xy} = \text{vec}(\tilde{\mathbf{a}}_{xy} \otimes \tilde{\mathbf{p}}_{xy}), \qquad \tilde{\mathbf{f}} = \frac{1}{S}\sum_{xy}\tilde{\mathbf{f}}_{xy}.$$

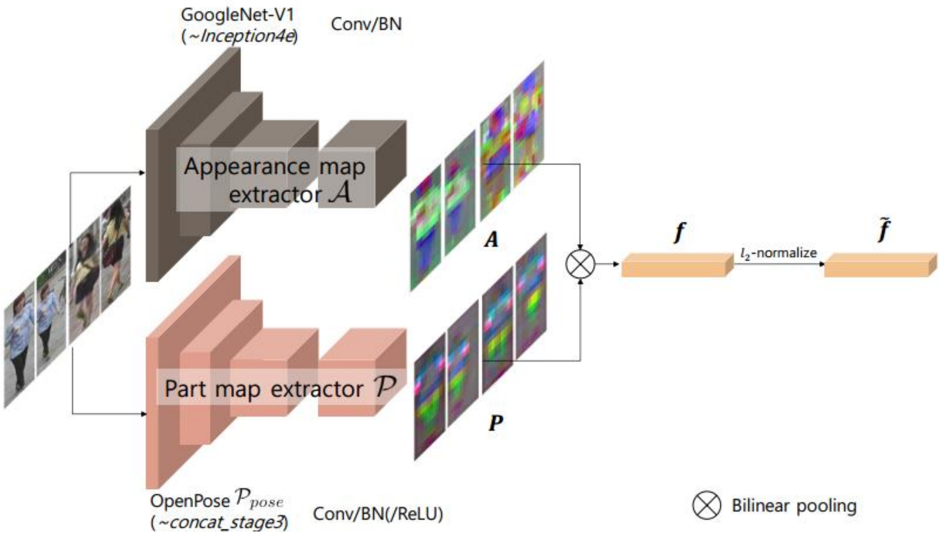


Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee, “Part-Aligned Bilinear Representations for Person Re-identification.” ECCV, 2018.

Part-Aligned Bilinear Representation

$$\tilde{\mathbf{f}} = \frac{1}{S}\sum_{xy}\tilde{\mathbf{f}}_{xy}.$$

$$\begin{aligned}\mathrm{sim}_I(\mathbf{I},\mathbf{I}') = <\tilde{\mathbf{f}},\tilde{\mathbf{f}}'> &= \frac{1}{S^2} < \sum_{xy}\tilde{\mathbf{f}}_{xy}, \sum_{x'y'}\tilde{\mathbf{f}}'_{x'y'} > \\ &= \frac{1}{S^2}\sum_{xy}\sum_{x'y'} < \tilde{\mathbf{f}}_{xy}, \tilde{\mathbf{f}}'_{x'y'} > \\ &= \frac{1}{S^2}\sum_{xy}\sum_{x'y'} \mathrm{sim}(\tilde{\mathbf{f}}_{xy}, \tilde{\mathbf{f}}'_{x'y'}),\end{aligned}$$



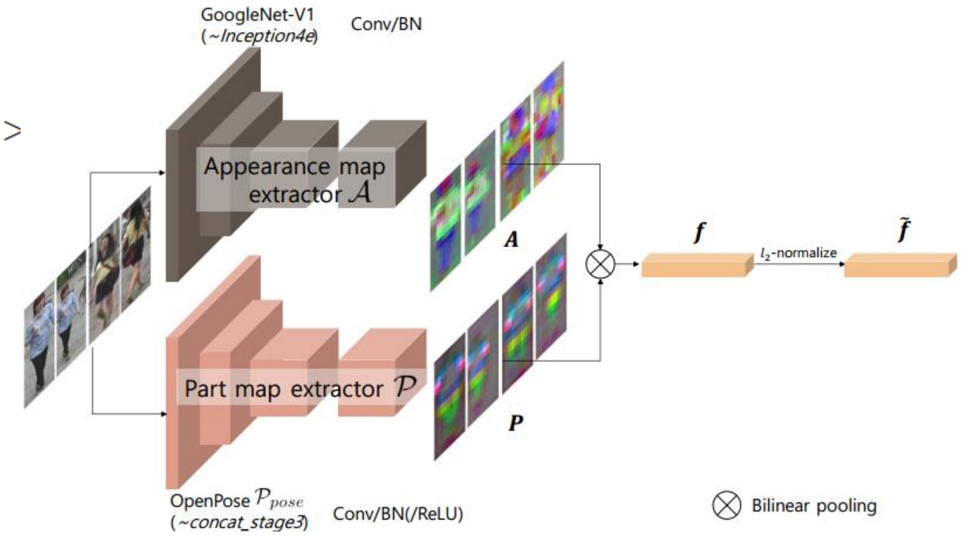
Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee, “Part-Aligned Bilinear Representations for Person Re-identification.” ECCV, 2018.

Part-Aligned Bilinear Representation

$$\mathrm{vec}(\mathbf{a} \otimes \mathbf{p}) = [(p_1\mathbf{a})^\top \; (p_2\mathbf{a})^\top \; \ldots (p_{c_P}\mathbf{a})^\top]^\top,$$

$$\begin{aligned}\mathrm{sim}(\tilde{\mathbf{f}}_{xy}, \tilde{\mathbf{f}}'_{x'y'}) &=< \mathrm{vec}(\tilde{\mathbf{a}}_{xy} \otimes \tilde{\mathbf{p}}_{xy}), \mathrm{vec}(\tilde{\mathbf{a}}'_{x'y'} \otimes \tilde{\mathbf{p}}'_{x'y'}) > \\ &=< \tilde{\mathbf{a}}_{xy}, \tilde{\mathbf{a}}'_{x'y'} >< \tilde{\mathbf{p}}_{xy}, \tilde{\mathbf{p}}'_{x'y'} > \\ &= \mathrm{sim}(\tilde{\mathbf{a}}_{xy}, \tilde{\mathbf{a}}'_{x'y'})\,\mathrm{sim}(\tilde{\mathbf{p}}_{xy}, \tilde{\mathbf{p}}'_{x'y'}).\end{aligned}$$

$$\mathrm{sim}_I(\mathbf{I}, \mathbf{I}') = \frac{1}{S^2} \sum_{xyx'y'} \mathrm{sim}(\tilde{\mathbf{a}}_{xy}, \tilde{\mathbf{a}}'_{x'y'})\,\mathrm{sim}(\tilde{\mathbf{p}}_{xy}, \tilde{\mathbf{p}}'_{x'y'}).$$



Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee, “Part-Aligned Bilinear Representations for Person Re-identification.” ECCV, 2018.

Part-Aligned Bilinear Representation

Table 1. Accuracy comparison on Market-1501

	Sinlge Query					Multi Query				
Rank	1	5	10	20	mAP	1	5	10	20	mAP
Varior et al. 2016 [58]	61.6	-	-	-	35.3	-	-	-	-	-
Zhong et al. 2017 [86]	77.1	-	-	-	63.6	-	-	-	-	-
Zhao et al. 2017 [76]	80.9	91.7	94.7	96.6	63.4	-	-	-	-	-
Sun et al. 2017 [53]	82.3	92.3	95.2	-	62.1	-	-	-	-	-
Geng et al. 2016 [16]	83.7	-	-	-	65.5	89.6	-	-	-	73.8
Lin et al. 2017 [31]	84.3	93.2	95.2	97.0	64.7	-	-	-	-	-
Bai et al. 2017 [2]	82.2	-	-	-	68.8	88.2	-	-	-	76.2
Chen et al. 2017 [9]	72.3	88.2	91.9	95.0	-	-	-	-	-	-
Hermans et al. 2017 [19]	84.9	94.2	-	-	69.1	90.5	96.3	-	-	76.4
+ re-ranking	86.7	93.4	-	-	81.1	91.8	95.8	-	-	87.2
Zhang et al. 2017 [74]	87.7	-	-	-	68.8	91.7	-	-	-	77.1
Zhong et al. 2017 [87]	87.1	-	-	-	71.3	-	-	-	-	-
+ re-ranking	89.1	-	-	-	83.9	-	-	-	-	-
Chen et al. 2017 [8] (MobileNet)	90.0	-	-	-	70.6	-	-	-	-	-
Chen et al. 2017 [8] (Inception-V3)	88.6	-	-	-	72.6	-	-	-	-	-
Ustinova et al. 2017 [57] (Bilinear)	66.4	85.0	90.2	-	41.2	-	-	-	-	-
Zheng et al. 2017 [79] (Pose)	79.3	90.8	94.4	96.5	56.0	-	-	-	-	-
Zhao et al. 2017 [75] (Pose)	76.9	91.5	94.6	96.7	-	-	-	-	-	-
Su et al. 2017 [50] (Pose)	84.1	92.7	94.9	96.8	65.4	-	-	-	-	-
Proposed (Inception-V1, R-CPM)	88.8	95.6	97.3	98.6	74.5	92.9	97.3	98.4	99.1	81.7
Proposed (Inception-V1, OpenPose)	**90.2**	**96.1**	**97.4**	**98.4**	**76.0**	93.2	97.5	98.4	99.1	82.7
+ dilation	**91.7**	**96.9**	**98.1**	**98.9**	**79.6**	**94.0**	**98.0**	**98.8**	**99.3**	**85.2**
+ re-ranking	**93.4**	**96.4**	**97.4**	**98.2**	**89.9**	**95.4**	**97.5**	**98.2**	**98.9**	**93.1**

Table 4. Accuracy comparison on DukeMTMC

Rank	1	5	10	20	mAP
Zheng et al. [85]	67.7	-	-	-	47.1
Tong et al. [67]	68.1	-	-	-	-
Lin et al. [31]	70.7	-	-	-	51.9
Schumann et al. [47]	72.6	-	-	-	52.0
Sun et al. [53]	76.7	86.4	89.9	-	56.8
Chen et al. [8] (MobileNet)	77.6	-	-	-	58.6
Chen et al. [8] (Inception-V3)	79.2	-	-	-	60.6
Zhun et al. [87]	79.3	-	-	-	62.4
+ re-ranking	84.0	-	-	-	78.3
Proposed (Inception V1, OpenPose)	82.1	90.2	92.7	95.0	64.2
+ dilation	**84.4**	**92.2**	**93.8**	**95.7**	**69.3**
+ re-ranking	**88.3**	**93.1**	**95.0**	**96.1**	**83.9**

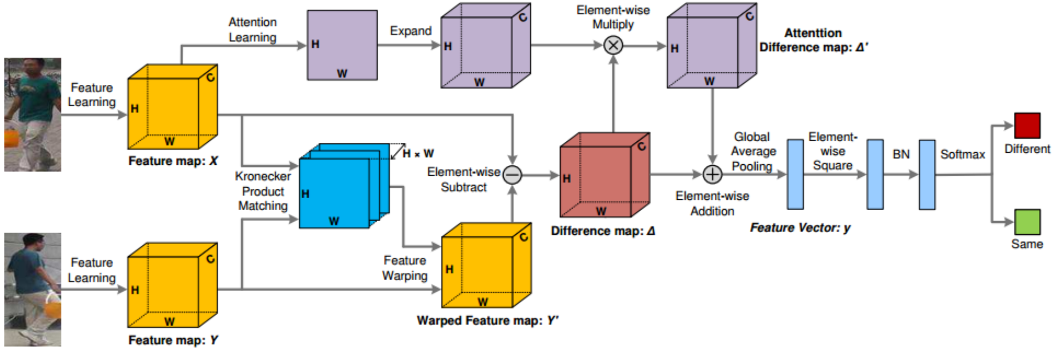
Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee, "Part-Aligned Bilinear Representations for Person Re-identification." ECCV, 2018.

End-to-End Deep Kronecker-Product Matching for Person Re-identification

Yantao Shen[1*] Tong Xiao[1*] Hongsheng Li[1†] Shuai Yi[2] Xiaogang Wang[1†]
[1] CUHK-SenseTime Joint Lab, The Chinese University of Hong Kong
[2] SenseTime Research
[1]{ytshen, xiaotong, hsli, xgwang}@ee.cuhk.edu.hk
[2]yishuai@sensetime.com

Motivation:

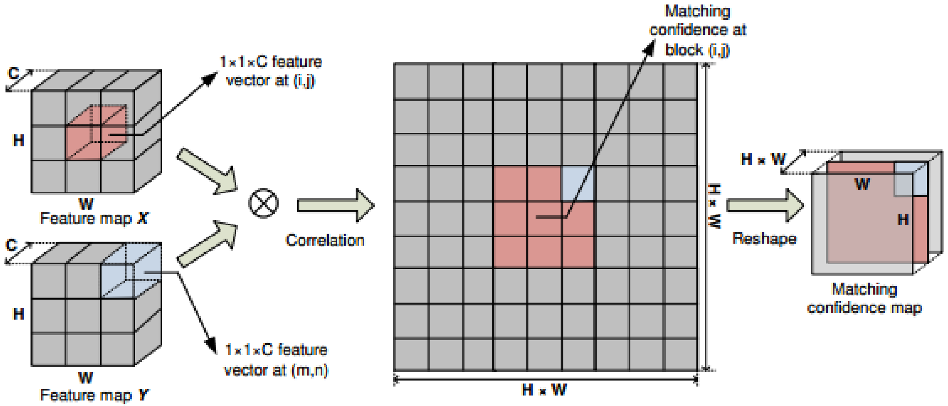
The global average pooling abandons valuable spatial information. Direct vectorization faces misalignment problems. Use Kronecker Product Matching to recover probabilistic correspondences between spatial regions across two images.



Kronecker Product Matching

$$\beta_{i,j} = \frac{\exp(x_i^T y_j / \tau_{\text{KPM}})}{\sum_{k=1}^{M} \exp(x_i^T y_k / \tau_{\text{KPM}})}$$

$$\Delta = X - Y\tilde{K}^T.$$



Yantao Shen, Tong Xiao, Hongsheng Li, Shuai Yi, Xiaogang Wang, " End-to-End Deep Kronecker-Product Matching for Person Re-identification." CVPR, 2018.

Kronecker Product Matching

Methods	Ref	Market-1501 [42]			
		mAP	top-1	top-5	top-10
DGD [36]	CVPR'16	31.9	59.5	-	-
CADL [18]	CVPR'17	47.1	73.8	-	-
P2S [48]	CVPR'17	44.3	70.7	-	-
MSCAN [14]	CVPR'17	53.1	76.3	-	-
SSM [2]	CVPR'17	68.8	82.2	-	-
SpinNet [39]	CVPR'17	-	76.9	91.5	94.6
JLML [16]	IJCAI'17	65.5	85.1	-	-
VI+LSRO [45]	ICCV'17	66.1	84.0	-	-
OL-MANS [47]	ICCV'17	-	60.7	-	-
PDC [30]	ICCV'17	63.4	84.1	92.7	94.9
PA [40]	ICCV'17	63.4	81.0	92.0	94.7
SVDNet [31]	ICCV'17	62.1	82.3	92.3	95.2
Ours		**75.3**	**90.1**	**96.7**	**97.9**

Methods	Ref	DukeMTMC [27]			
		mAP	top-1	top-5	top-10
BoW+KISSME [42]	ICCV'15	12.2	25.1	-	-
LOMO+XQDA [17]	CVPR'15	17.0	30.8	-	-
ACRN [28]	CVPR'17	52.0	72.6	84.8	88.9
Basel+LSRO [45]	ICCV'17	47.1	67.7	-	-
SVDNet [31]	ICCV'17	56.8	76.7	86.4	89.9
Ours		**63.2**	**80.3**	**89.5**	**91.9**

Yantao Shen, Tong Xiao, Hongsheng Li, Shuai Yi, Xiaogang Wang, " End-to-End Deep Kronecker-Product Matching for Person Re-identification." CVPR, 2018.

Spatial-Temporal Person Re-identification

Guangcong Wang[1], **Jianhuang Lai**[1,2,3*], **Peigen Huang**[1], **Xiaohua Xie**[1,2,3]
[1]School of Data and Computer Science, Sun Yat-sen University, China
[2]Guangdong Key Laboratory of Information Security Technology
[3]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education
{wanggc3, huangpg}@mail2.sysu.edu.cn, {stsljh, xiexiaoh6}@mail.sysu.edu.cn

Motivation:

Most methods neglect spatial-temporal constraint. Spatial-temporal constraint eliminates lots of irrelevant target images in gallery, and thus significantly alleviates the appearance ambiguity problem.

Spatial-Temporal Person Re-id

- Histogram

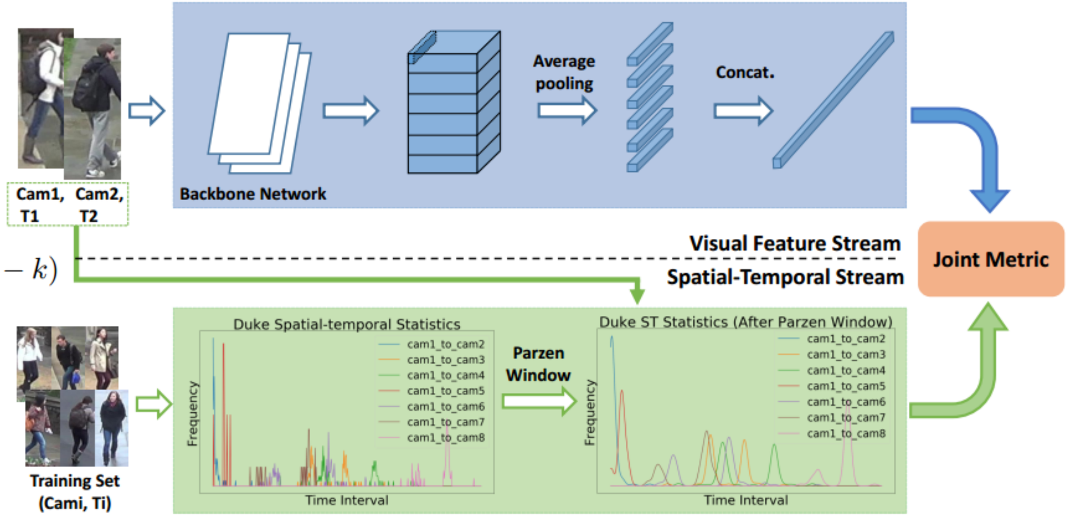
$$\hat{p}(y=1|k,c_i,c_j) = \frac{n^k_{c_i c_j}}{\sum_l n^l_{c_i c_j}}$$

- Parzen Window Smoothing

$$p(y=1|k,c_i,c_j) = \frac{1}{Z}\sum_l \hat{p}(y=1|l,c_i,c_j)K(l-k)$$

- Laplace Smoothing

$$p_\lambda(Y=d_k) = \frac{m_k+\lambda}{M+D\lambda}$$



Guangcong Wang, Jianhuang Lai, Peigen Huang, Xiaohua Xie, "Spatial-Temporal Person Re-identification." AAAI, 2019.

Spatial-Temporal Person Re-id

- Visual Feature Distance

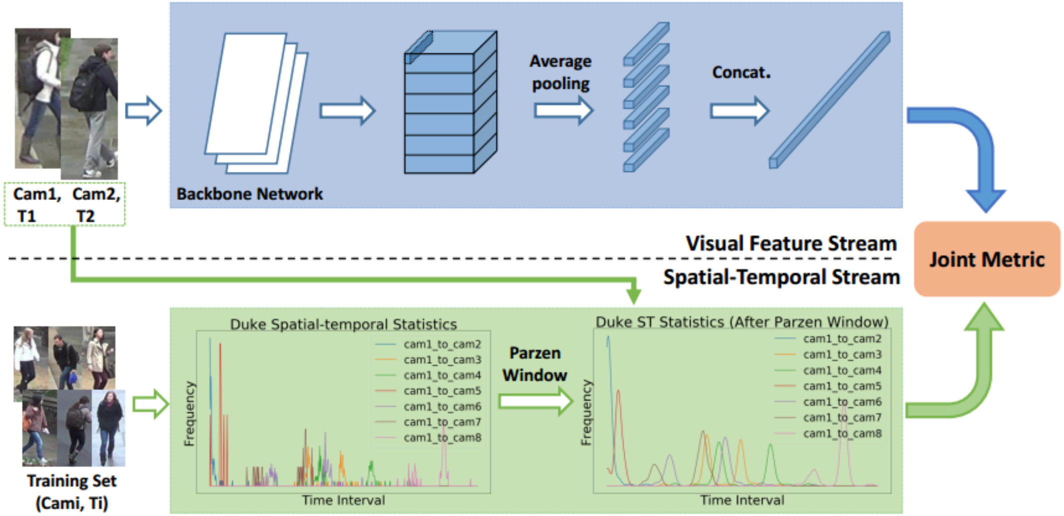
$$s(\mathbf{x_i}, \mathbf{x_j}) = \frac{\mathbf{x_i} \cdot \mathbf{x_j}}{||\mathbf{x_i}||||\mathbf{x_j}||}$$

- Spatial-Temporal probability

$$p(y = 1|k, c_i, c_j)$$

- Joint Metric

$$p_{joint} = f(s; \lambda_0, \gamma_0) f(p_{st}; \lambda_1, \gamma_1)$$



Guangcong Wang, Jianhuang Lai, Peigen Huang, Xiaohua Xie, "Spatial-Temporal Person Re-identification." AAAI, 2019.

Spatial-Temporal Person Re-id

Market1501

SSDAL	39.4	-	-	19.6
APR	84.3	93.2	95.2	64.7
Human Parsing	93.9	98.8	99.5	-
Mask-guided	83.79	-	-	74.3
Background	81.2	94.6	97.0	-
PDC	84.1	92.7	94.9	63.4
PSE+ECN	90.3	-	-	84.0
MultiScale	88.9	-	-	73.1
Spindle Net	76.9	91.5	94.6	-
Latent Parts	80.3	-	-	57.5
Part-Aligned	81.0	92.0	94.7	63.4
PCB(*)	91.2	97.0	98.2	75.8
TFusion-sup	73.1	86.4	90.5	-
st-ReID	97.2	**99.3**	99.5	86.7
st-ReID+RE	**98.1**	**99.3**	**99.6**	87.6
st-ReID+RE+re-rank	98.0	98.9	99.1	**95.5**

DukeMTMC

PAN	71.6	-	-	51.5
SVDNet	76.7	-	-	56.8
HA-CNN	80.5	-	-	63.8
APR	70.7	-	-	51.9
Human Parsing	84.4	91.9	93.7	71.0
PSE+ECN	85.2	-	-	79.8
MultiScale	79.2	-	-	60.6
PCB(*)	83.8	91.7	94.4	69.4
st-ReID	94.0	97.0	97.8	82.8
st-ReID+RE	94.4	**97.4**	**98.2**	83.9
st-ReID+RE+re-rank	**94.5**	96.8	97.1	**92.7**

Guangcong Wang, Jianhuang Lai, Peigen Huang, Xiaohua Xie, ”Spatial-Temporal Person Re-identification.” AAAI, 2019.

Thank you