Self and Channel Attention Network for Person Re-Identification

Asad Munir, Niki Martinel, Christian Micheloni
University of Udine

Problem and Related Work

- Re-identify a person in the field of view of non-overlapping cameras.
- The re-id task is similar to image classification task with having different identities in training and testing (identities mismatching)
- Discriminative and sharp features needed to get a better similarity score. The general networks ignore the similar features at distant locations.

Dataset Images

- Existing methods are usually trained with single classifiers.
- Attention based methods don't take into count the fact that the person re-id datasets are blurry and noisy. So, they are unable to learn sharp and salient features.

Contributions

- We proposed multi classifiers training to learn the most discriminative features with multiple classifiers instead of single classifier.
- Introduction of Self Attention (SA) module in the baseline network to make it rely on non-local similarities instead of local mechanism of convolution filters.
- Introduction of Channel Attention (CA) module for learning sharp and discriminative features for better matching.

Proposed SCAN

Self Attention

Feature vector x is composed into three parts using 1X1 convolution layers

$$f(x) = w_f x, \quad g(x) = w_g x, \quad h(x) = w_h x$$

Then calculate similarity between the two patches by taking dot product.

$$s_{ij} = f(x_i)^T g(x_j)$$

Compute attention maps by applying softmax.

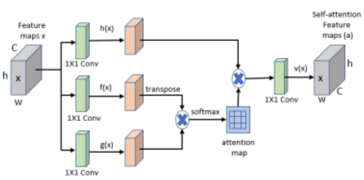
$$\alpha_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^{N} \exp(s_{ij})}$$

Calculate final attention with the whole (third) patch.

$$o_j = v\left(\sum_{i=1}^{N} \alpha_{j,i} h(x_i)\right)$$

Final Transformed features are obtained by

$$y_i = \gamma o_i + x_i$$



Channel Attention

Convolution operation is written as

$$u_c = k_c * X = \sum_{n=1}^{C'} k_c^n * x^n$$

Calculate channel descriptor with GAP

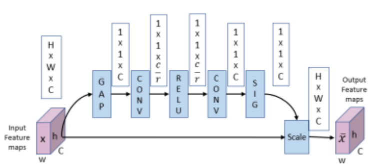
$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j)$$

Apply sigmoid non-linearity after the reduction layers

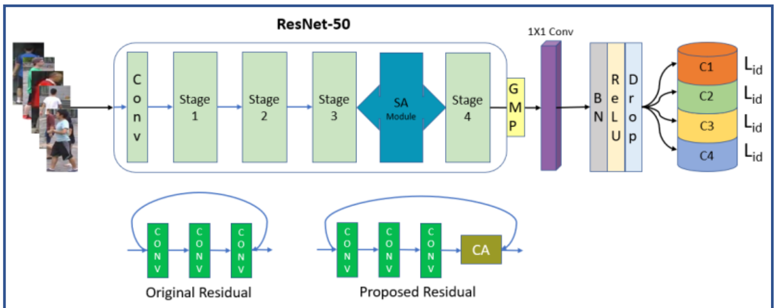
$$n = \sigma\big(g(z, W)\big) = \sigma\big(W_2 \delta(W_1 z)\big)$$

Final transformed output is

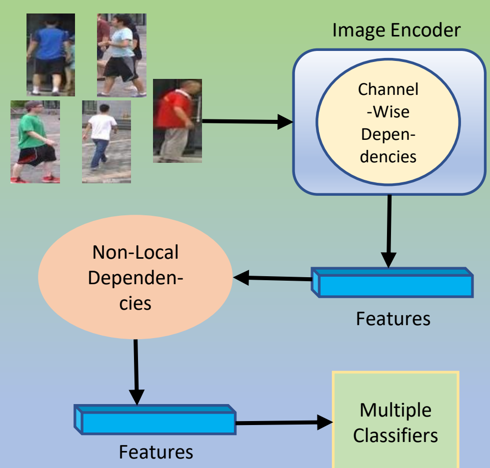
$$\bar{x}_c = n_c . u_c$$



This work was supported by EU H2020 MSCA through Project ACHIEVE-ITN (Grant No 765866)



Overview Of The Proposed Approach



Experimental Results

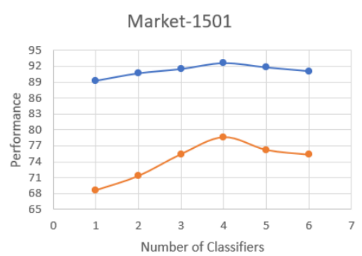
Market1501
Cameras: 6
IDs: 751 &750
Train Imgs: 12936
Test Imgs: 19281
Market Queries 3368

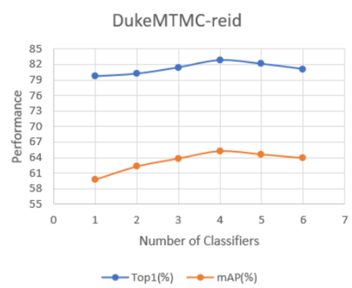
DukeMTMC-reID
Cameras: 8
IDs: 702 & 702
Train Imgs: 16522
Test Imgs: 17661
Queries : 2228

Results on Market1501

Methods	Reference	Market-1501		
		Rank-1(%)	*Rank-5(%)*	*mAP(%)*
SpindleNet [12]	CVPR17	76.9	91.5	-
Part-Aligned [13]	ICCV17	81.0	92.0	63.4
HydraPlus-Net [16]	ICCV17	76.9	91.3	-
LSRO [10]	ICCV17	84.0	-	66.1
SVDNet [37]	ICCV17	82.3	92.3	62.1
DPFL [38]	ICCV17	88.9	92.3	73.1
PSE [39]	CVPR18	87.7	94.5	69.0
HA-CNN [18]	CVPR18	91.2	-	75.5
AACN [17]	CVPR18	85.9	-	66.9
MLFN [40]	CVPR18	90.0	-	74.3
DuATM [41]	CVPR18	91.4	97.1	76.6
DKP [42]	CVPR18	90.1	96.7	75.3
GCSL [43]	CVPR18	93.5	-	81.6
PCB [14]	ECCV18	92.3	97.2	77.4
OGSL [44]	ICPR18	87.1	-	70.2
PRFF [45]	ICPR18	86.3	94.8	69.4
IDCL [9]	CVPRW19	93.1	-	78.9
PyrNet [6]	CVPRW19	93.6	98.2	81.7
CASN(IDE) [19]	CVPR19	92.0	-	78.0
SFT [46]	ICCV19	93.4	97.4	82.7
SCAN(ID)	-	94.1	97.7	82.1
SCAN(ID+Tri)	-	94.2	97.8	83.6

Effect of multiple Classifiers





Results on DukeMTMC-reID

Methods	Reference	DukeMTMC-reid		
		Rank-1(%)	*Rank-5(%)*	*mAP(%)*
Verif-Identif [47]	TOMM18	68.9		49.3
LSRO [10]	ICCV17	67.7	-	47.1
SVDNet [37]	ICCV17	76.7	86.4	56.8
DPFL [38]	ICCV17	73.2	-	60.6
PSE [39]	CVPR18	79.8	89.7	62.0
HA-CNN [18]	CVPR18	80.5	-	63.8
AACN [17]	CVPR18	76.8	-	59.2
MLFN [40]	CVPR18	81.0	-	62.8
DuATM [41]	CVPR18	81.8	90.2	68.6
DKP [42]	CVPR18	80.3	89.5	63.2
GCSL [43]	CVPR18	84.9	-	69.5
PCB [14]	ECCV18	81.8	-	66.1
OGSL [44]	ICPR18	76.3	-	63.7
PRFF [45]	ICPR18	72.1	83.8	53.4
IDCL [9]	CVPRW19	83.9	-	68.2
CASN(IDE) [19]	CVPR19	84.5	-	67.0
SCAN(ID)	-	84.9	92.0	69.2
SCAN(ID+Tri)	-	85.3	92.7	71.0

Component Analysis

Networks	Components			Market		Duke	
	CA	SA	Multi-C	mAP	R1	mAP	R1
baseline	✗	✗	✗	68.6	89.3	59.8	79.7
multi-C	✗	✗	✓	78.6	92.6	65.2	82.8
CA-baseline	✓	✗	✓	81.6	93.5	68.9	83.9
SA-baseline	✗	✓	✓	80.8	93.8	68.2	84.5
SCAN (ID)	✓	✓	✓	82.1	94.1	69.2	84.9
SCAN (ID+Tri)	✓	✓	✓	**83.6**	**94.2**	**71.0**	**85.3**

Conclusion

- With multiple classifiers and losses, proposed network learns robust global features at the added convolutional layers.
- To capture the non-local dependencies, we introduced self-attention(SA) module to enhance the similarity learning.
- To learn the salient and sharp features from degraded person re-identification data, the Channel-Attention (CA) module is introduced in the network.
- The proposed SCAN model learns the most discriminative, sharp and salient features for feature matching.

EU H2020 MSCA ACHIEVE-ITN (Grant No. 765866)