Self and Channel Attention Network for Person Re-Identification

Presenter: Asad Munir

Authors : Asad Munir, Niki Martinel, Christian Micheloni

Affiliation: University of Udine (Machine Learning and Perception Lab)

Marie Skłodowska-Curie Actions

ACHIEVE ITN

Grant No. 765866

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Person Re-Identification

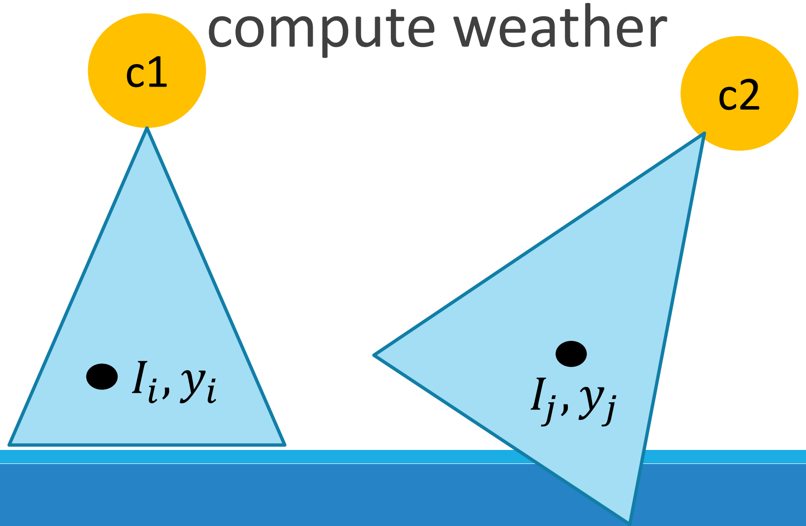
Given a dataset of N persons

$$D_{Tr} = \{I_k, y_k\}_{k=1}^{N}$$

Where I_k and y_k are the person image and ID of the k^{th} person in the dataset.

For a given pair of person images $\{I_i, I_j\}$, the task of re-identification is to compute weather

$$y_i = y_j \text{ or } y_i \neq y_j$$



Problems and Contributions

- ***Identities mismatching*.**

We proposed multi classifiers training to learn the most discriminative features with multiple classifiers instead of single classifier.

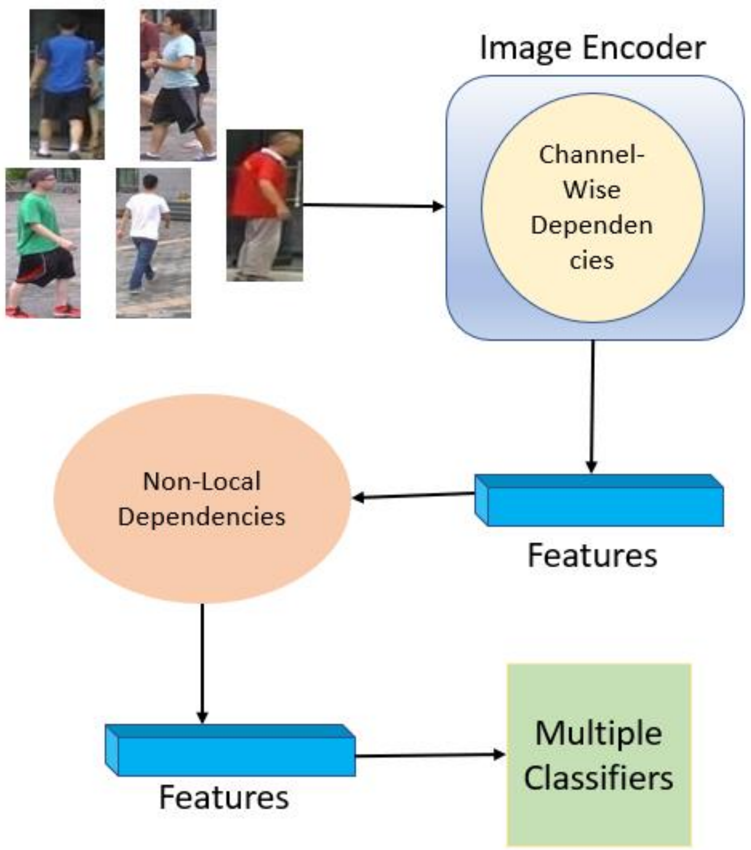
- ***Non-Local Dependencies*.**

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.

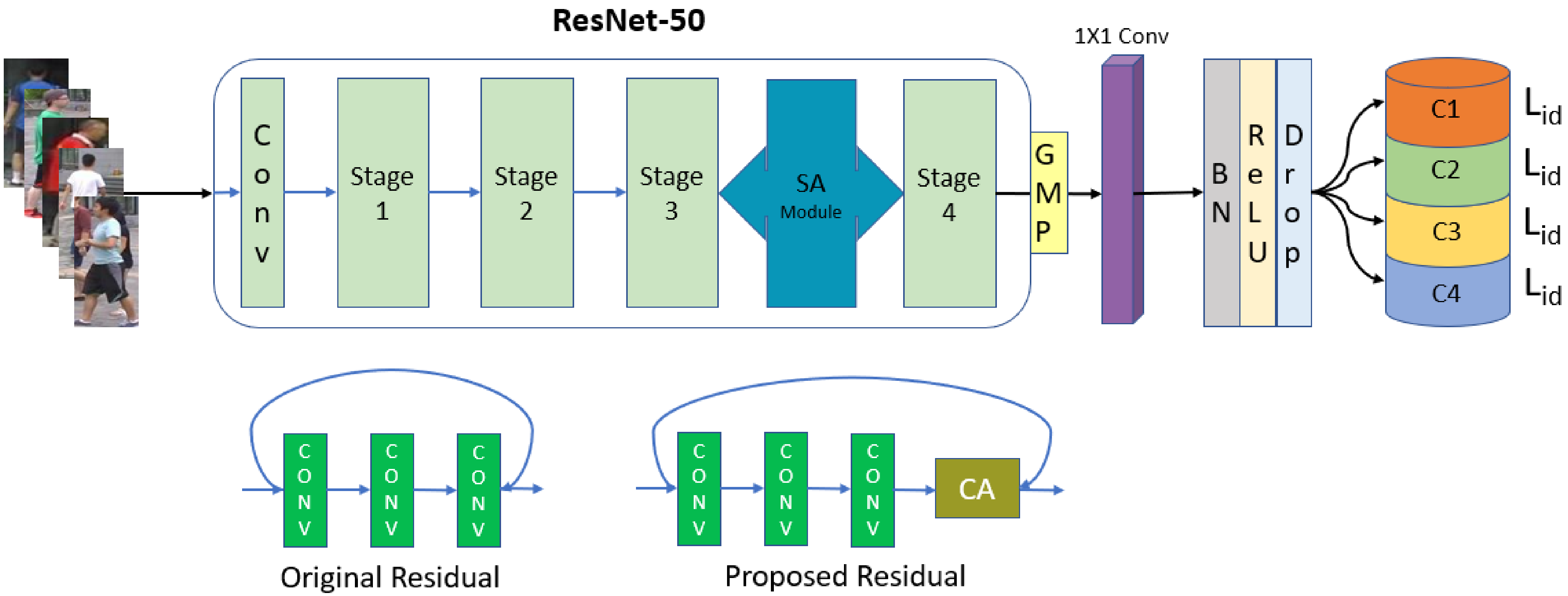
Introduction of Channel Attention (CA) module for learning sharp and discriminative features for better matching.

Methodology

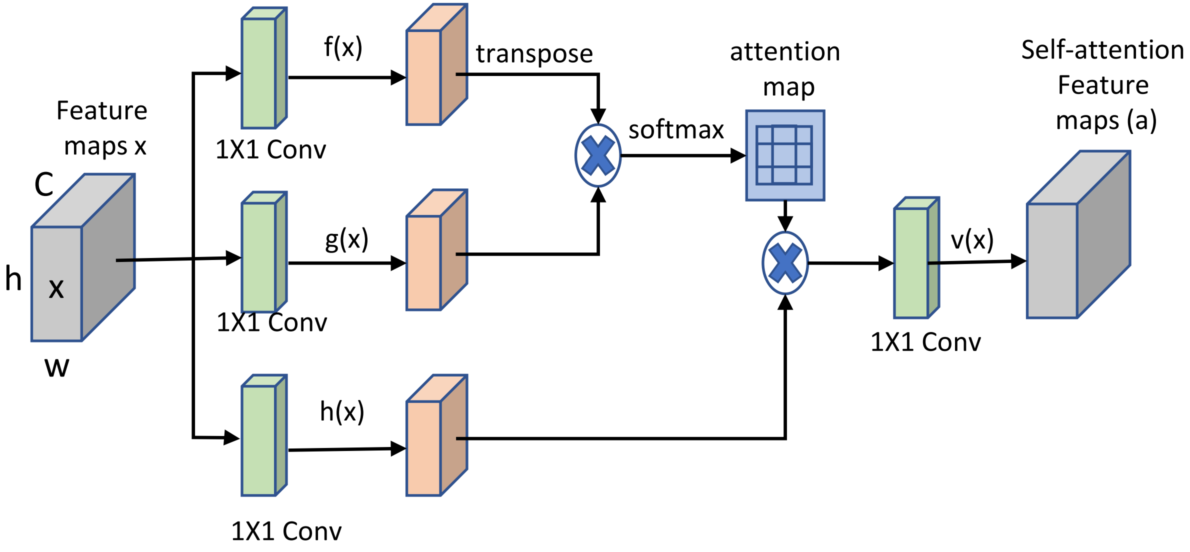
- Multi Branch (classifier) Training
  - Using multiple classifiers layers instead of single layer.
- Self Attention (SA) Module.
  - Features are passed through SA module to apply self attention.
  - Performs better on small spatial size of features.
- Channel Attention (CA) Module.
  - At every residual connection CA module is applied to increase discriminability of learned features.

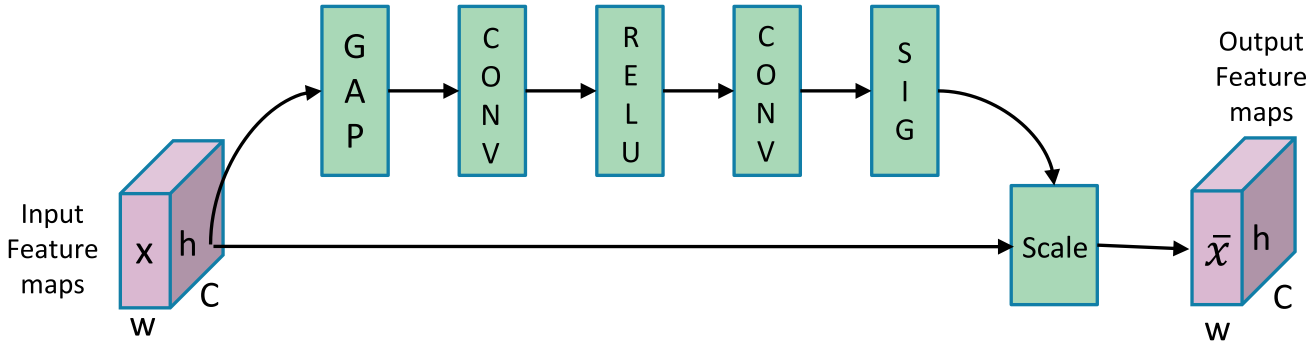


Proposed Network



Self and Channel Attention Mechanism





is a dot product

1X1 Conv are the reduction layers

Experiments

Datasets:

- **Market1501**
  - The dataset consists of images from 6 cameras
  - 12936 images in training set with 751 identities
  - 19281 images in testing set with 750 identities (3368 queries)

- **DukeMTMC-reID**
  - The dataset consists of images from 8 cameras
  - 16522 images in training set with 702 identities
  - 17661 images in testing set with 702 identities (2228 queries)

Experiments

- Comparisons to the state-of-the-art re-id methods on Market-1501. The top 1 and 2 results are in red and blue.

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

Experiments

- Comparisons to the state-of-the-art re-id methods on DukeMTMC-ReID. The top 1 and 2 results are in red and blue.

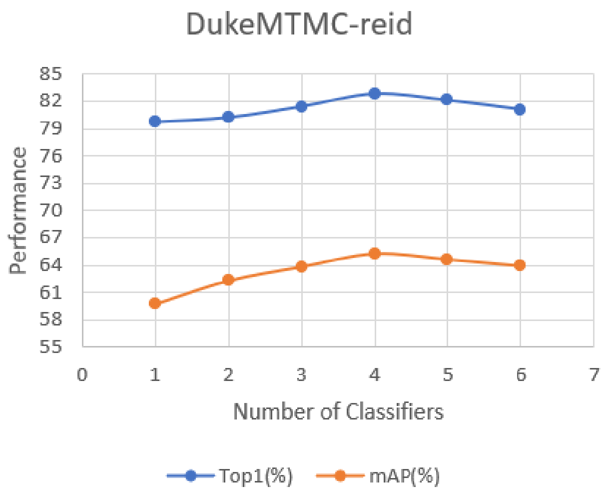
Methods	Reference	DukeMTMC-reid		
		Rank-1(%)	*Rank-5(%)*	*mAP(%)*
Verif-Identif [47]	TOMM18	68.9		49.3
LSRO [10]	ICCV17	67.7	-	47.1
SVDNet [37]	ICCV17	76.7	86.4	56.8
DPFL [38]	ICCV17	73.2	-	60.6
PSE [39]	CVPR18	79.8	89.7	62.0
HA-CNN [18]	CVPR18	80.5	-	63.8
AACN [17]	CVPR18	76.8	-	59.2
MLFN [40]	CVPR18	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

- Multi Classifier training.
- Effect of Self and Channel attention modules.

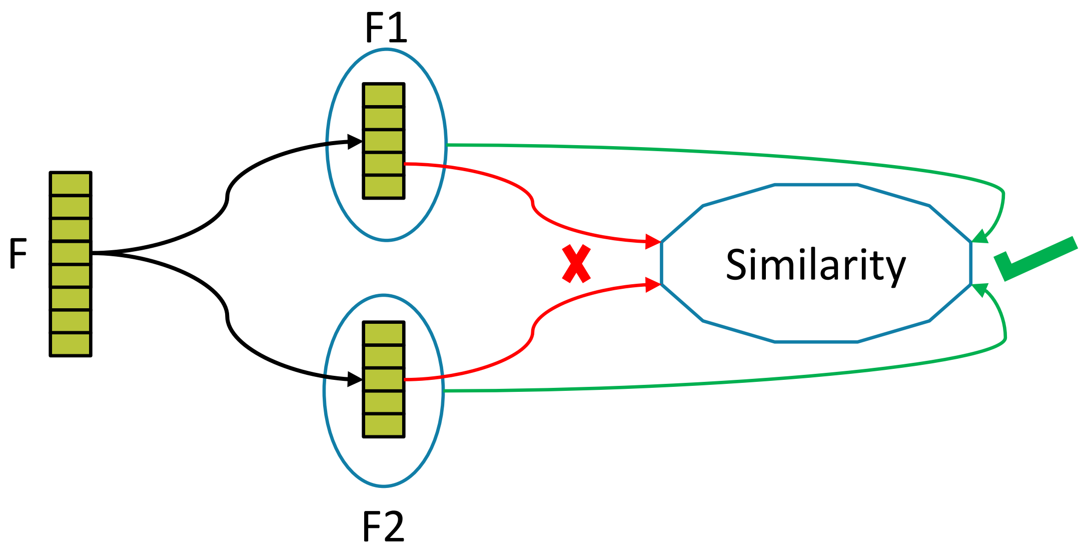
Networks	Components			Market		Duke	
	CA	SA	Multi-C	mAP	R1	mAP	R1
baseline	✗	✗	✗	68.6	89.3	59.8	79.7
multi-C	✗	✗	✓	78.6	92.6	65.2	82.8
CA-baseline	✓	✗	✓	81.6	93.5	68.9	83.9
SA-baseline	✗	✓	✓	80.8	93.8	68.2	84.5
SCAN (ID)	✓	✓	✓	82.1	94.1	69.2	84.9
SCAN (ID+Tri)	✓	✓	✓	**83.6**	**94.2**	**71.0**	**85.3**

- Impact of multi classifiers on two benchmark datasets.



Limitations

- High memory usage due to splitting of features and extra operations.
- Missing cross feature similarities.
- Red arrows for features at different levels.



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.

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

THANK YOU

Questions?

asad.munir@uniud.it