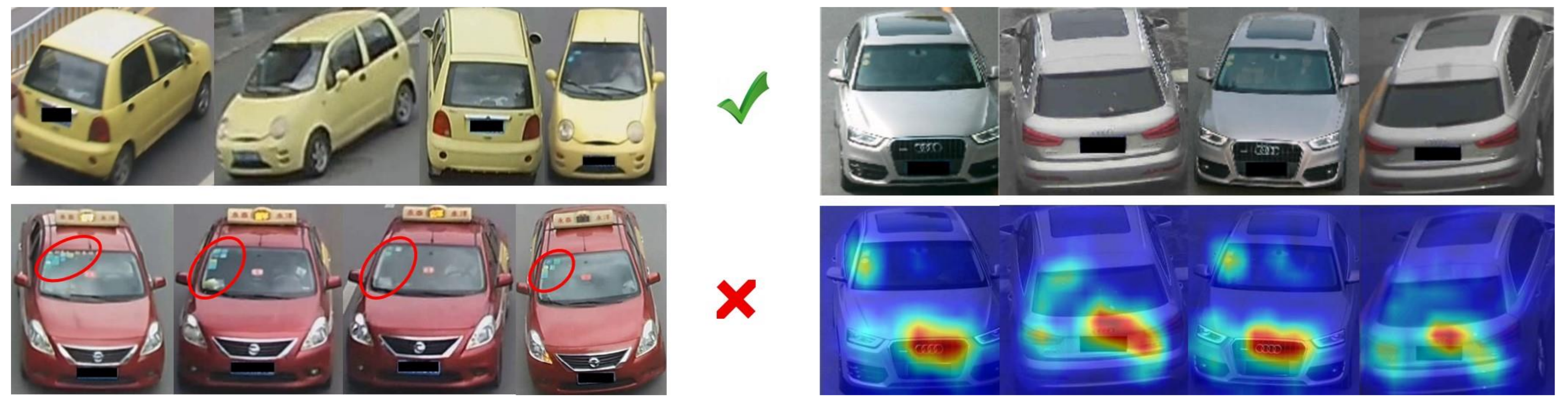
SCAN: Spatial and Channel Attention Network for Vehicle Re-Identification

Shangzhi Teng[1], Xiaobin Liu[2] , Shiliang Zhang[2] , Qingming Huang[1]

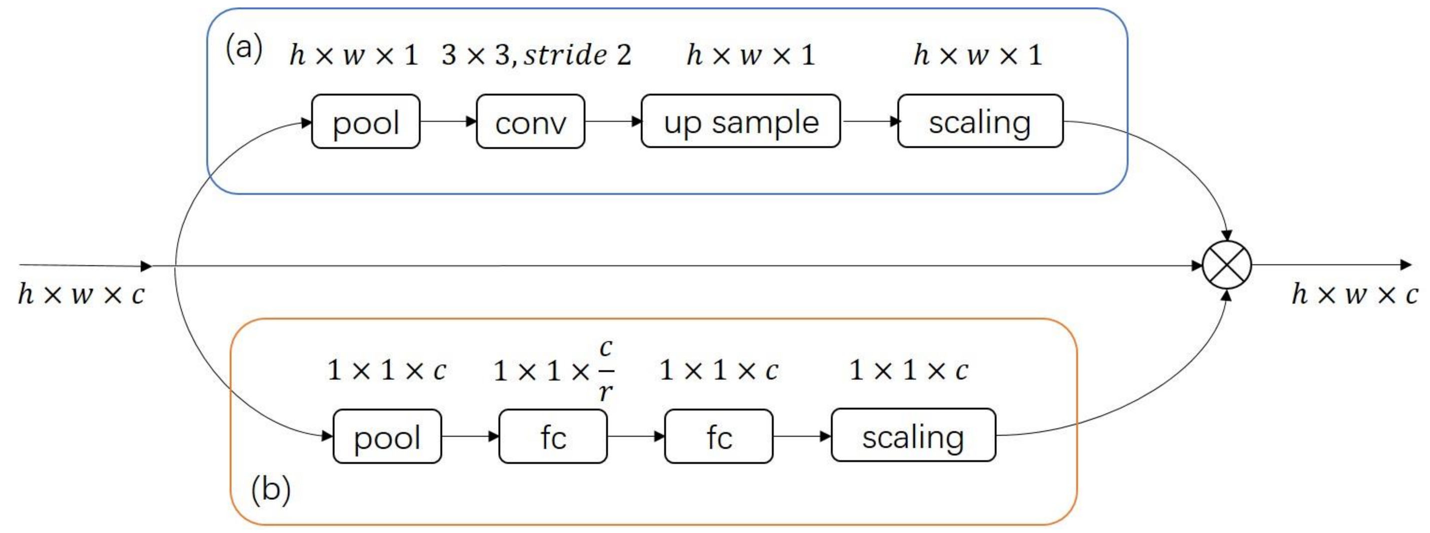
[1] University of Chinese Academy of Sciences, Beijing, CN [2] Peking University, Beijing, CN

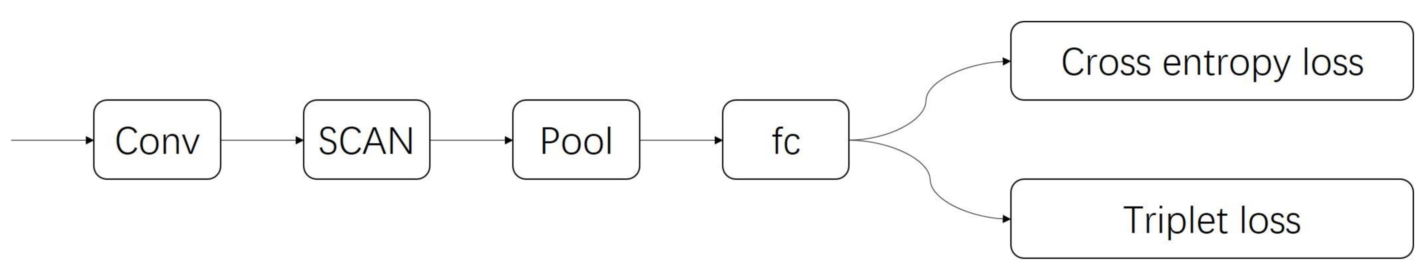
Motivation

1. Compared with global appearance, some local regions could be more discriminative.
2. It is difficult to align every local part between two vehicle images of different viewpoints.
3. To automatically discover discriminative regions on vehicles and discriminative channels in networks, we propose a Spatial and Channel Attention Network (SCAN) based on DCNN.



Spatial and Channel Attention Network





➢ **Spatial Attention Branch**

The structure of SAB is illustrated in (a). channel-wise global average pooling, a convolutional layer of 3×3 filter with stride 2, an up sampling layer and a scaling conv layer (1 parameter).

➢ **Channel Attention Branch**

The structure of CAB is illustrated in (b). Motivated by SENet .

➢ **Vehicle ReID by SCAN**

We add our SCAN after the conv5 layer in VGG_CNN_M 1024 and the conv5_3 layer in VGG16. we use the 512-D fully connected layer as the vehicle feature. We calculate L2 distance between query images and each gallery image using this 512-D deep feature.

Experiments

➢**Datasets**

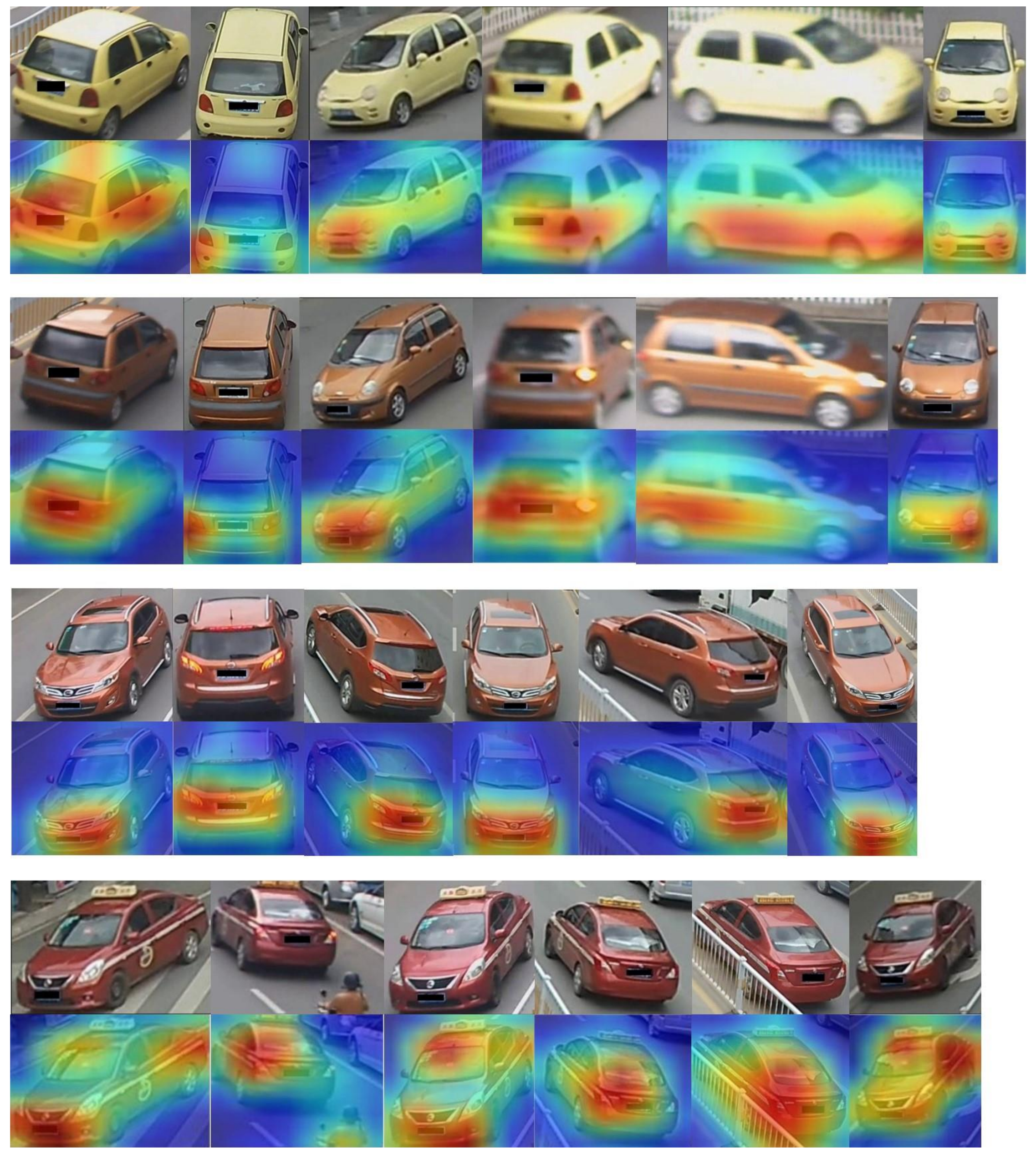
Dataset	Train ID/image	Probe ID/image	gallery ID/image
VeRi-776	576/37778	200/1678	200/11579
VehicleID	13164/100182	2400/17638	2400/2400

➢**Results on VeRi-776 and VehicleID**

VeRi-776	Top-1(%)	Top-5(%)	mAP(%)
KEPLER [16]	48.2	64.3	33.53
FACT [14]	50.95	73.48	18.49
FACT+Plate+STR [14]	61.44	78.78	27.77
OIFE [23]	68.3	89.7	51.42
VAMI [32]	77.03	90.82	50.13
light vgg m	76.02	86.05	38.94
light vgg m+SCAN	**82.24**	**90.76**	**49.87**
light vgg16	76.82	86.71	39.91
light vgg16+SCAN	**79.92**	**88.32**	**50.15**

VehicleID	Top-1(%)	Top-5(%)
VGG + Triplet Loss [11]	31.9	50.3
VGG + CCL [11]	32.9	53.3
Mixed Diff + CCL [11]	38.2	61.6
OIFE [23]	67.0	82.9
VAMI [32]	47.34	70.29
light vgg m	44.14	65.21
light vgg m+SCAN	55.73	71.73
light vgg16	60.63	72.67
light vgg16+SCAN	63.52	77.53
light vgg16+SCAN∗	**65.44**	**78.47**

Visualisation of our spatial attention



Spatial attention branch locates some spatial regions of vehicles, which approximately corresponds to headlights, taillights, vehicle signs, and vehicle marks.

Comparisons of the number of parameters

Model	NP (million)
VGG M 1024	86.2
VGG16	127.2
light vgg m	**6.8**
light vgg16	**7.9**
SCAN	**0.033**

Conclusions

➢We proposed an end-to-end trainable framework, namely Spatial Channel Attention Network (SCAN), for joint learning attention weights and feature representation.

➢With our SCAN model we could explore discriminative regions and channels for powerful feature extraction without extra annotations.

➢Our two baseline network are all lightweight CNN architectures. So it's easy to embed our model in mobile devices.

References

[5] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. arXiv preprint arXiv:1709.01507 (2017)

[23] Wang, Z., et al. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. pp. 379-387 (CVPR2017).

[32] Zhou, Y., Shao, L.: Aware attentive multi-view inference for vehicle reidentification. pp. 6489-6498 (CVPR2018).