IEEE International Conference On Multimedia and Expo (ICME) 2018

E2BoWs: An End-to-End Bag-of-Words Model via Deep Convolutional Neural Network

Xiaobin Liu[1], Shiliang Zhang[1], Qingming Huang[2], Wen Gao[1]
[1]School of Electronic Engineering and Computer Science, Peking University, Beijing, 100871, China
[2]School of Computer and Control Engineering, University of Chinese Academy of Sciences, China
{ xbliu.vmc, slzhang.jdl, wgao }@pku.edu.cn, qmhuang@ucas.ac.cn

Abstract

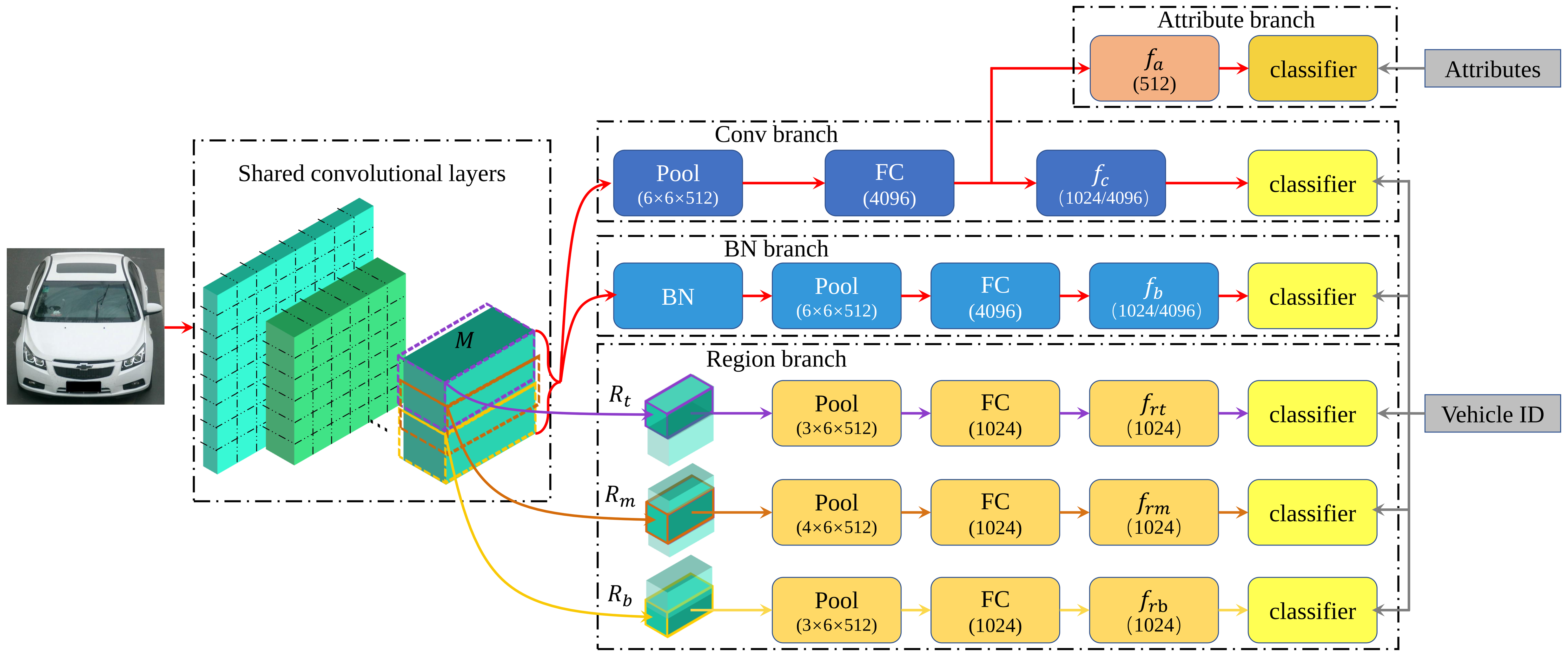
Different vehicles with similar global appearance. The local differences are highlighted with red circles.

Previous works on vehicle ReID mainly extract global features. However, some different vehicles may have similar global appearance, making it hard to distinguish them. Compared with global appearance, some local regions may be more distinctive. To embed the detailed visual cues, we propose a Region-Aware deep Model (**RAM**). Specifically, besides global features, RAM also extracts features from a series of local regions. This encourages the model to learn discriminative features. We also introduce a novel learning algorithm that jointly uses vehicle IDs, types/models and colors to train the RAM. This strategy fuses more cues for training and results in more discriminative features. Extensive experiments on two large-scale vehicle Re-ID datasets, *i.e.*, *VeRi* and *VehicleID*, show our methods achieve promising performance.

Contributions

1. We propose RAM to jointly learn deep features from both the global appearance and local regions. Learned features are more discriminative to detailed local cues than ones in previous works.
2. Color and model cues are additionally used to jointly train the deep model. The final concatenated feature achieves promising performance in comparison with recent ones.

RAM structure



- Conv branch extracts **global** features as previous works do.
- BN branch embeds a BN layer before pooling layer to extract complementary **global** features.
- Region branch extracts **local** features from three overlapped parts of feature maps.
- Attribute branch extracts **attribute** features learned by attributes classifiers.

Model training

We training the model in a step-by-step manner.
Step-1 first trains the *baseline* model only having the Conv branch.
Step-2 adds the BN branch to the baseline model. Model trained in this step is denoted as *BN*.
Step-3 further adds the Region branch to model BN. Model trained in this step is denoted as *BN+R*.
Step-4 adds Attribute branch to model *BN+R*. This final model is denoted as *RAM*.

Experiments

1. Performance comparison of features learned by different models on *VeRi*.

Models	mAP	Top-1	Top-5
Baseline	0.550	0.848	0.931
BN	0.581	0.871	0.940
BN+R	0.609	0.887	0.941
RAM	**0.615**	**0.886**	**0.940**

2. Performance comparison of features learned by different models on *VehicleID*.

Models	Top-1			Top-5		
	Small	Medium	Large	Small	Medium	Large
Baseline	0.694	0.673	0.632	0.892	0.820	0.795
BN	0.722	0.705	0.666	0.904	0.853	0.832
BN+R	0.747	0.720	0.674	0.908	0.863	0.842
RAM	**0.752**	**0.723**	**0.677**	**0.915**	**0.870**	**0.845**

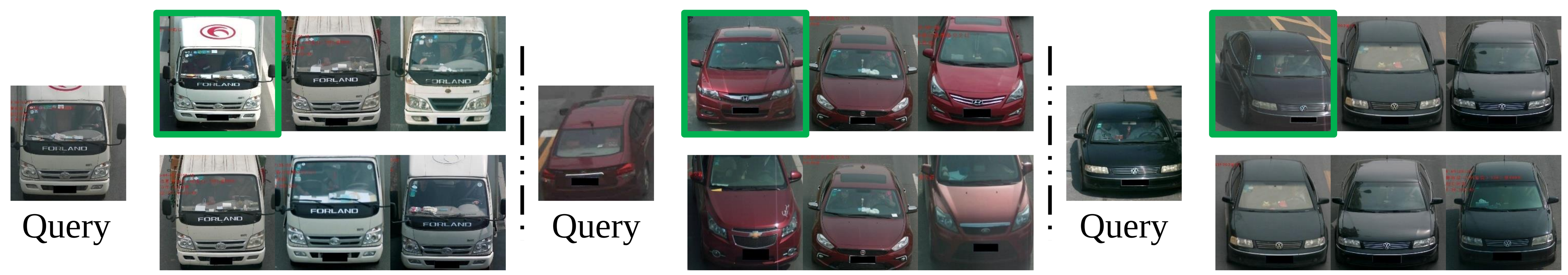
3. Comparison with recent works on *VeRi*.

Models	mAP	Top-1	Top-5
FACT[14]	0.199	0.597	0.753
FPSS[4]	0.278	0.614	0.788
SCPL[5]	0.583	0.835	0.900
OIF[6]	0.480	0.659	0.877
OIF+SF[6]	0.514	0.683	0.897
RAM	**0.615**	**0.886**	**0.940**

4. Comparison with recent works on *VehicleID*.

Models	Top-1			Top-5		
	Small	Medium	Large	Small	Medium	Large
VGGT[3]	0.404	0.354	0.319	0.617	0.546	0.503
VGGCCL[3]	0.436	0.370	0.329	0.642	0.571	0.533
MD+CCL[3]	0.490	0.428	0.382	0.735	0.668	0.616
OIF	-	-	0.670	-	-	0.829
RAM	**0.752**	**0.723**	**0.677**	**0.915**	**0.870**	**0.845**

Examples of returned images on *VehicleID* by *RAM* (in the first line) and *baseline* (in the second line):



Acknowledgements

This work is supported by National Science Foundation of China under Grant No. 61572050, 91538111, 61620106009, 61429201, and the National 1000 Youth Talents Plan.

Code
Person-Vehicle ReID Demo
VMC Team
Xiaobin Liu's homepage