

# 基于特征聚类对群三元组损失的车辆再识别

吴燕雄<sup>1,2</sup>, 蔡建美<sup>1</sup>, 滕云田<sup>1,2</sup>

(1. 防灾科技学院电子科学与控制工程学院, 河北三河 065201; 2. 中国地震局地球物理研究所, 北京 100081)

**摘要:** 车辆再识别旨在从多个摄像机拍摄的图像中识别出同一车辆. 本文提出了一种对群三元组损失函数, 以特征中心点替代均值, 并将对群思想和三元组损失相结合, 优化了困难样本的识别. 车辆再识别过程中, 对群损失函数的训练过程扩大了样本规模, 增加了计算量, 且传统对群损失函数无法准确处理困难正样本. 为此, 提出了一种特征聚类对群三元组损失函数. 本方法采用正样本特征聚类中心并改进了三元组损失函数的设计, 从而优化了对群损失函数. 在不扩增输入样本数量的同时提升了算法处理困难样本的能力. 实验表明, 与主流车辆再识别算法相比, 本方法可有效提升车辆再识别的准确率.

**关键词:** 车辆再识别; 视觉特征; 特征聚类对群损失; 三元组损失

**中图分类号:** TP391.4      **文献标识码:** A      **文章编号:** 0372-2112 (2020)12-2444-09

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2020.12.021

## Vehicle Re-identification Using the Coupled Feature Clusters Embedded into Triplet Loss

WU Yan-xiong<sup>1,2</sup>, CAI Jian-xian<sup>1</sup>, TENG Yun-tian<sup>1,2</sup>

(1. School of Electronic Science and Control Engineering, Institute of Disaster Prevention, Sanhe, Hebei 065201, China;

2. Institute of Geophysics, China Earthquake Administration, Beijing 100081, China)

**Abstract:** Vehicle re-identification is the task of identifying the same vehicle across some images captured by multiple cameras. We propose a coupled feature clusters embedded into triplet loss dealing with hard samples. During the vehicle re-identification, the coupled clusters loss suffers from larger computation consumption caused by the extension of the sample scale and the reduction of identification accuracy. Therefore, the coupled feature clusters embedded into triplet loss is proposed. It improves the ability of the algorithm on processing hard samples in terms of selecting feature centers of positive samples based on clustering and the embedded into a triple loss. Experiments show that the algorithm effectively improves the accuracy of vehicle re-identification compared to the vehicle re-identification algorithm based on coupled clusters loss.

**Key words:** vehicle re-identification; visual appearance; coupled feature clusters loss; triplet loss

### 1 引言

随着监控摄像机规模的不断扩大, 监控网络中任何一个摄像头都无法完整覆盖整个监控环境, 因此不同摄像头中锁定有违法嫌疑车辆的需求牵引了车辆再识别技术的深入研究. 车辆再识别算法利用监控系统抓取的车辆图片所包含的表观信息, 运用机器视觉中的特征提取方法提取图像包含的表观特征, 然后通过机器学习中的度量学习方法对提取到的不同图像的表观特征进行相似度的评估并判断不同图片是否拍摄

的同一辆汽车.

根据特征提取方法应用对象的不同, 可以将此方法细分为基于全局表观信息的车辆再识别方法和基于局部表观信息的车辆再识别方法. 基于全局表观信息的车辆再识别方法<sup>[1-6]</sup>特点是提取到的表观特征包含了车辆全身所有部位的信息, 更能反映不同图片中车辆的整体相似程度. 文献[5]提出了一种叫深度相关距离学习 (Deep Relative Distance Learning, DRDL) 的方法, 通过改进传统三元组损失为对群损失更好的扩大类间差异缩小类内差异, 并结合细粒度车型分类构成混合

收稿日期: 2020-04-10; 修回日期: 2020-05-11; 责任编辑: 梅志强

基金项目: 防灾科技学院教学研究与教育改革项目 (No. JY2016B10); 河北省高等学校科学技术研究重点项目 (No. ZD2018304); 中央高校基本科研业务费 (No. ZY20180111); 国家重点研发计划项目 (No. 2018YFC1503801)

网络框架来实现车辆再识别的目标.但是基于对群损失的方法,依然存在传统三元组损失的缺点,即扩大了训练过程输入样本数,消耗大量计算资源且随数据集规模的扩大问题越显著,另外其对于区分困难样本组依然缺乏解决力.文献[7]提出了一种由粗到细的排序损失去学一个结构化的嵌入式特征来更好解决车辆再识别问题.但是存在着仅利用全局特征缺乏针对性和处理困难样本能力待提高的问题.文献[3]提出了一种应用颜色直方图、HOG 特征并使用 SVM 作为车辆匹配识别器的车辆再识别方法.文献[6]应用了行人再识别领域一种基于深度学习的能从不同领域数据中学习一种一般性鲁棒特征的特征提取方法 DGD<sup>[8]</sup>,并据此完成全局表观特征的提取实现车辆的再识别.文献[4]提出了一种由 SIFT 特征、颜色名及 GoogLeNet 特征组成的全局表观特征 FACT 来实现车辆再识别的方法.文献[2]提出了一种名为 Siamese Visual 的应用孪生网络根据全局表观信息来获取车辆表观相似度的车辆再识别方法.

基于局部表观信息的车辆再识别方法<sup>[6,7]</sup>是指特征提取方法的应用对象为图片的各个局部的车辆再识别方法,它的特点是提取到的表观特征只包含了人工选取的某些车辆局部位置的信息,因为加入了人类知识的辅助使得提取到的表观特征更符合人类对车辆特性的认知.但是如果具备区分力的信息出现在人工选取局部区域之外,那么此种方法提取的表观特征也就失去了区分力.文献[7]通过自选的 20 个车辆局部位置并抽取其特征,再结合全局特征形成方向不变性特征来衡量车辆差异,来达到车辆再识别目的.但是自选的 20 个车辆局部位置需要人工对数据集进行二次标注,增加了大量人工成本削弱了现实可行性,针对困难样本组也缺乏较好的自动化处理能力.文献[6]应用行人再识别领域一种克服视角及光照变化的局部特征 LOMO<sup>[9]</sup>来完成车辆再识别任务.

此外,对于融合时空信息的车辆再识别<sup>[10,11]</sup>,虽然时空信息修正可以提高车辆再识别精度,但是由于标注困难等问题,现有数据集大多没有包含时空信息,因此利用时空信息解决车辆再识别问题不具有普遍性,无法应用于超大规模场景中.基于局部信息的车辆再识别<sup>[12,13]</sup>同样需要局部特征区域的标注信息,且算法复杂度相对更高.

由于实际场景中存在大量表观上极为相似的同型号的不同车辆,而且同一辆车拍摄得到的图片又因拍摄角度环境光照等影响其表观上的相似度.故存在不同车辆图片的表观特征向量在特征空间中相距较近,而同一车辆的特征向量在特征空间中相距较远,即正样本与负样本在特征空间的聚散关系不明确.

因此,提高处理困难样本的能力是车辆再识别算法优化的核心目标.为优化车辆再识别算法,提升其处理困难样本的能力,鉴于深度学习在特征自动提取的强势表现,本论文以全局表观信息的车辆再识别为基础,通过样本聚类取代样本均值,优化对群的损失函数,并将对群的设计融入到三元组损失函数中,构造了基于特征聚类对群三元组损失的车辆再识别方法,有效地提升了困难样本的再识别准确率.

## 2 相关工作与问题分析

对于基于全局表观信息的车辆再识别算法,根据采用的损失函数的不同,存在着不同的车辆再识别算法.为了研究如何更好地优化正负样本在特征空间中的聚散关系,本节主要针对其中表现优秀的基于对群损失的车辆再识别算法的现状进行介绍,分析其中还存在的问题并给出相应的解决思路.

### 2.1 相关工作介绍

图 1 即为基于对群损失的车辆再识别算法的框架示意图.算法使用卷积神经网络来提取每一张图片的特征,并使用对群损失替换三元组损失使得训练过程更加稳定且能加快收敛速度.

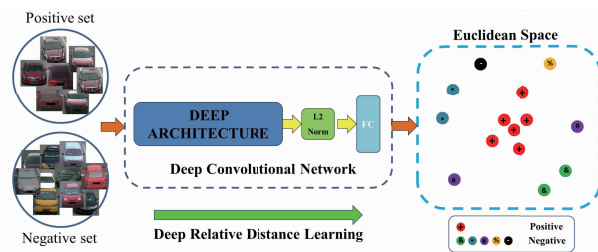


图1 基于对群损失的车辆算法框架

不同于三元组输入的方式,训练使用对群损失的卷积神经网络时,输入由两组不同的图片集组成,一个是正样本集一个是负样本集.正样本集  $X^p = \{x_1^p, \dots, x_{N^p}^p\}$  包含  $N^p$  张拥有相同类别的图片,负样本集  $X^n = \{x_1^n, \dots, x_{N^n}^n\}$  包含  $N^n$  张不同类别的图片.通常,在欧氏特征空间中属于相同类别的样本应该位于一个相同的特征中心的附近,因此正样本集中的样本在特征空间中应该共同组成一个聚集群.而负样本集中的样本在特征空间中应当相对远离这个聚集群,于是就形成了一个对群的形势,图 2 展示了这样的理想情形.

另一方面,对群损失采用了与三元组损失相同的设计思想,即在特征空间中样本的类内距离应当小于类间距离.不过对群损失在具体设计上相对于三元组有两个主要不同点.第一,在选取计算距离所需的标准值时,对群损失使用的是一个由正样本特征算术均值代表的聚集群中心,而不像三元组损失那样使用的是

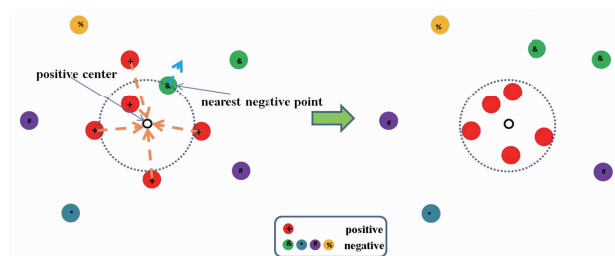


图2 对群损失示意图

随机选取的一个正样本;第二,对群损失函数的定义是基于多样本的而不是像三元组损失那样基于三个样本.上述设计可以确保在反向传播阶段样本的移动方向正确,同时使得所有离特征中心不够近的正样本尽可能的向特征中心靠拢.

## 2.2 存在问题与解决思路

基于表现信息的车辆再识别方法中的基于对群损失的车辆再识别方法,在三元组损失设计思想的基础上优化了标准值的选取并创新性地提出了基于对群结构的损失函数设计最终提升了车辆再识别准确率.但是基于对群损失的方法<sup>[5]</sup>还是存在着和三元组损失类似的缺点,即扩大了训练过程输入样本数.因此这一方法将会消耗大量的计算资源且随训练集数据规模的扩大问题越显著.另外,该方法设计的对群损失也存在问题.对群损失将特征空间中正样本的特征向量求算术均值,并将这一个均值视为正样本的标准值来作为衡量是否同一辆汽车的度量.也就是说,一个样本图片经过抽取特征,当这一特征向量在特征空间中离上述标准值距离越近就表示这一样本和上述标准值代表的车辆越可能是同一辆车.但是简单的求算术均值得到的标准值并不能很好的代表一辆汽车在表现上的本质特征.另外,对群损失函数的设计并没有将困难正样本的情形考虑进去,因此也使得该方法对于困难样本的区分能力不足.

通过上述对基于对群损失的车辆再识别方法的分析,可将其存在的问题归纳为如下两点:

(1)取一组正样本特征向量的算术均值作为正样本的标准值,即正样本特征中心的选取,并不能很好地度量这组正样本从属汽车的表现特征的本质,制约了车辆再识别准确率的提升.

(2)对群损失函数的设计在训练过程中扩大了样本输入数量增加计算量且没有考虑样本集中最不相像的正样本即困难正样本的情形,消耗计算资源过大且对困难正样本的处理能力不足.

针对现有的基于对群损失的车辆再识别方法中存在的上述问题,本论文在后续部分提出了相应解决思路:

(1)针对正样本特征中心不能很好地度量这组正样本从属汽车的表现特征本质的问题,增加考虑正样本内自身存在差异的情况,不是简单取算术平均值来抹平正样本内差异.本文提出基于聚类的正样本特征中心选取方法.通过观察监控汽车图片等实例发现正对车头视角与正对车尾视角差异最大,且基本不存在正对车辆侧面的情形.于是在设计正样本特征中心时,先将正样本做二分类取此两类中心位置间合适位置为正样本特征中心来代表一辆汽车的度量本质,以优化正样本特征中心的选取.

(2)针对对群损失函数在训练过程中扩大了样本输入数量增加计算量且没有考虑困难正样本情形的问题,借助人再识别技术中的改进三元组损失设计思路<sup>[14]</sup>来解决训练过程输入样本过大和未考虑困难正样本的问题,提出了特征聚类对群三元组损失函数来优化损失函数设计,使得离正样本特征中心最远的正样本向正样本特征中心靠拢的同时最具欺骗性的负样本,即离正样本特征中心最近的负样本,离正样本特征中心尽可能的远,从而提升处理困难样本的能力.

## 3 基于特征聚类对群三元组损失的车辆再识别算法

### 3.1 基于聚类的正样本特征中心选取方法

针对基于对群损失的车辆再识别方法选取正样本特征向量算术均值作为正样本特征中心不能很好地度量汽车表现特征本质的问题,本节给出一种基于聚类的正样本特征中心选取方法来优化正样本特征中心的选取.

利用监控车辆图片中正对车头视角与正对车尾视角差异最大的特点.首先,利用无监督的 K-Means<sup>[15]</sup>算法将输入训练的一组正样本集做聚类簇数  $K=2$  的划分;其次,将各簇的特征向量的算术均值作为簇中心;最后,根据簇内正样本个数比值情况确定了一个的正样本特征中心,使得簇内样本个数越多正样本特征中心就越靠近此簇中心.

现有方法<sup>[5]</sup>采取对输入正样本组所有特征向量计算算术均值,其中心点可作如下表示

$$c^p = \frac{1}{N^p} \sum_i^{N^p} f(x_i^p) \quad (1)$$

其中,  $c^p$  表示中心点,  $N^p$  表示正样本图片数量,  $x_i^p$  表示第  $i$  张正样本,  $f(x)$  表示图片  $x$  的神经网络特征表示.

本文采用的优化方法可以由下三式表示出优化后的正样本特征中心.

$$E = \sum_{i=1}^2 \sum_{x \in C_i} \|x - m_i\|_2^2 \quad (2)$$

$$m_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (3)$$

$$c_i^p = \frac{(\alpha \mu_{i1} + \beta \mu_{i2})}{\alpha + \beta} \quad (4)$$

具体获取正样本特征中心的过程可以描述为:首先,输入一组车辆图像,利用深度卷积神经网络提取其视觉表现特征向量  $\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_i, \dots, \mathbf{F}_N)$ ,  $\mathbf{F}_i = (\mathbf{F}_{i1}, \mathbf{F}_{i2}, \dots, \mathbf{F}_{iM})$ , 其中  $N$  为车辆的总数目,  $M$  为同一车辆不同图像的数目. 其次,对每个车辆的所有图像的视觉表现特征向量  $\mathbf{F}_i = (\mathbf{F}_{i1}, \mathbf{F}_{i2}, \dots, \mathbf{F}_{iM})$ ,  $i = 1, 2, \dots, N$ , 运用 K-Means 算法将其划分为 2 个聚类, 得到两个聚类中心  $\mu_{i1}$  和  $\mu_{i2}$ . 最后, 利用公式  $c_i^p = \frac{(\alpha \mu_{i1} + \beta \mu_{i2})}{\alpha + \beta}$ ,

$i = 1, 2, \dots, N$  计算每一车辆的正样本特征中心  $c_i^p$ ,  $i = 1, 2, \dots, N$ , 其中  $\alpha$  和  $\beta$  为聚类中心加权系数, 满足  $N_{C_1} * \beta = N_{C_2} * \alpha$ ,  $N_{C_1}$  和  $N_{C_2}$  表示两个聚类包含的样本个数, 以保证正样本特征中心总是靠近大簇的中心, 更好地代表车辆图像组的主要表现特征.

### 3.2 特征聚类对群三元组损失函数

针对基于对群损失的车辆再识别方法在训练过程中扩大了样本输入数量增加计算量且没有考虑困难正样本情形的问题, 本节给出一种特征聚类对群三元组损失函数来优化对群损失函数.

本文借用行人再识别技术中的改进三元组损失设计思路<sup>[14]</sup>, 即让在特征空间离正样本标准值最远的正样本特征向量尽可能的向标准值靠拢的同时让离正样本标准值最近的负样本特征向量尽可能地远离正样本标准值, 改进了现有方法中的对群损失, 最终得到特征聚类对群三元组损失函数来增强本文算法处理困难样本的能力. 图 3 为特征聚类对群三元组损失设计示意图.

对群损失函数<sup>[5]</sup>可由下式表示

$$L(\mathbf{W}, X^p, X^n) = \sum_i^{N^p} \frac{1}{2} \max \{0, \|f(\mathbf{x}_i^p) - \mathbf{c}^p\|_2^2 + \alpha - \|f(\mathbf{x}_i^n) - \mathbf{c}^p\|_2^2\} \quad (5)$$

其中,  $\mathbf{W}$  表示模型权重,  $X^p$  表示正样本集,  $X^n$  表示负样本集,  $\mathbf{x}_i^n$  表示离中心点最近的负样本,  $\mathbf{c}^p$  即为式(1)中表示的一组正样本特征向量的算术均值. 整个对群损失函数的工作机制是, 首先输入一组同属于一个 ID 的正样本车辆图片和同样数量的一组随机抽取的其他任意 ID 车辆图片构成的负样本一同组成一个对群, 然后选取这个对群中正样本特征向量算术均值为正样本标准值, 训练的过程就是在特征空间让离正样本标准值最远的正样本特征向量尽可能的向标准值靠拢的同时让离正样本标准值最近的负样本特征向量尽可能地远离正样本标准值.

行人再识别技术中的改进三元组损失函数<sup>[14]</sup>可由下式表示.

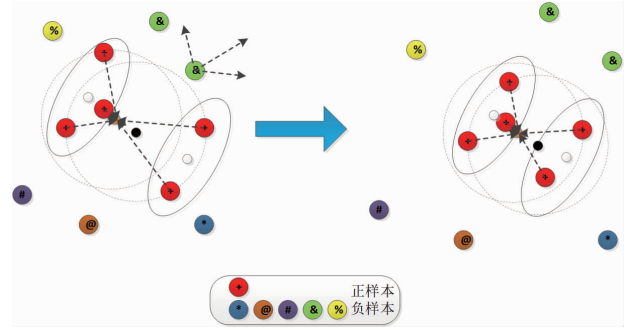


图3 特征聚类对群三元组损失示意图

$$L(\theta; X) = \sum_{i=1}^N \sum_{a=1}^M \max \{0, \max_{q=1, \dots, M} \{ \|f_\theta(\mathbf{x}_{iq}^p) - f_\theta(\mathbf{x}_{ia}^p)\|_2 \} - \min_{\substack{j=1, \dots, N \\ r=1, \dots, M \\ j \neq i}} \{ \|f_\theta(\mathbf{x}_{jr}^n) - f_\theta(\mathbf{x}_{ia}^p)\|_2 \} + m\} \quad (6)$$

其中,  $\theta$  表示模型权重,  $X$  表示一个 Batch 包含的样本集,  $N$  是不同行人种类数目,  $M$  是从属于同一行人的图片数目,  $N * M$  即等于一个 Batch 包含的样本集的图片数量,  $f_\theta(\mathbf{x}_{ia}^p)$  是作为标准值的第  $i$  类第  $a$  张图片的特征向量,  $f_\theta(\mathbf{x}_{iq}^p)$  是作为正样本的第  $i$  类第  $q$  张图片的特征向量,  $f_\theta(\mathbf{x}_{jr}^n)$  是作为负样本的第  $j$  类第  $r$  张图片的特征向量. 整个改进三元组损失函数的工作机制是, 输入一个 Batch 的样本集并让每张图片做一次标准值, 当一张图片做标准值时找出同 ID 其他图片离它最远的图片和一个 Batch 样本集中离它最近的那张其他 ID 图片, 训练过程就是在特征空间让离标准值最远的正样本离标准值尽可能近的同时让离标准值最近的负样本尽可能地远离标准值.

本文采用的特征聚类对群三元组损失函数可以由下式表示. 其设计思路是结合对群结构优于三元组结构的特点<sup>[5]</sup>及改进三元组损失充分挖掘困难样本信息且不用扩大数据输入规模的特点, 集成了一个优于对群损失函数的特征聚类对群三元组损失函数.

$$L(\mathbf{W}, \mathbf{F}) = \sum_{i=1}^N \sum_{l=1}^M \frac{1}{2} \max \{0, \max_{l=1, \dots, M} \{ \|\mathbf{F}_{il}^p - \mathbf{c}_i^p\|_2 \} - \min_{\substack{l=1, \dots, M \\ i=1, \dots, N}} \{ \|\mathbf{F}_{il}^n - \mathbf{c}_i^p\|_2 \} + m\} \quad (7)$$

其中  $\mathbf{W}$  为更新后的深度卷积神经网络参数,  $\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_i, \dots, \mathbf{F}_N)$ ,  $\mathbf{F}_i = (\mathbf{F}_{i1}, \mathbf{F}_{i2}, \dots, \mathbf{F}_{iM})$  为车辆表现特征,  $N$  为车辆的总数目,  $M$  为同一车辆不同图像的数目,  $\mathbf{F}_{il}^p$  是第  $i$  正样本集合的第  $l$  个特征向量,  $\mathbf{F}_{il}^n$  是第  $i$  个负样本集合的第  $l$  个特征向量,  $n$  表示此特征向量属于负样本,  $p$  表示此特征向量属于正样本,  $m$  是离正样本特征中心最远的正样本与正样本特征中心欧氏距离的平方与离正样本特征中心最近的正样本与正样本特征中心欧氏距离的平方的差值的最小边距. 整个特征聚类对群三元组损失函数的工作机制是, 输入一

个 Batch 的样本集并使用基于聚类的正样本特征中心方法求得正样本特征中心并令其作为标准值,当某一 ID 的一组正样本的特征中心做标准值时找出同 ID 其他图片离它最远的图片和一个 Batch 样本集中离它最近的那张其他 ID 图片,训练过程就是在特征空间让离标准值最远的正样本离标准值尽可能近的同时让离标准值最近的负样本尽可能地远离标准值。

### 3.3 基于特征聚类对群三元组损失的车辆再识别算法描述

基于特征聚类对群三元组损失函数的车辆再识别算法与当前主流的基于对群损失函数的方法相比,可以较好地提升处理困难样本的能力,并且适用于道路监控场景应用中的车辆再识别。

本论文算法可做如下简要描述:首先,利用无监督的 K-Means 算法将输入训练的一组正样本集聚成 2 类;其次,将各簇的特征向量的算术均值作为簇中心;然后,根据簇内正样本个数比值情况,选定正样本特征向量的中心  $c_i^p$  其中  $i=1,2,\dots,N,N$  为车辆的总数目;由此,利用正样本特征中心  $c_i^p$  设计得到特征聚类对群三元组损失函数  $L(\mathbf{W}, \mathbf{F})$  并结合 ResNet 深度卷积神经网络,完成对网络模型的训练后获取车辆表观特征  $\mathbf{F}$ ,将查询车辆图片表观特征与目标库车辆图片表观特征欧氏距离值从小到大排序,按照距离值越小与待识别车辆身份相同概率越大的规则完成对两幅图片是否属于同一辆汽车的判定。其中算法主要涉及了一个无监督聚类算法和一个主流的神经网络模型,即 K-Means 算法<sup>[15]</sup>和 ResNet 网络模型<sup>[16]</sup>,本节将对他们逐一做出介绍并给出基于特征聚类对群三元组损失函数的车辆再识别算法的详细流程。图 4 是基于特征聚类对群三元组损失函数的车辆再识别算法示意图。图的含义是通过将不同车辆图片输入到应用了特征聚类对群三元组损失训练好的卷积神经网络,获得不同车辆的表观特征再求特征间的欧氏距离并以此作为车辆的相似度来衡量车辆的相似性。

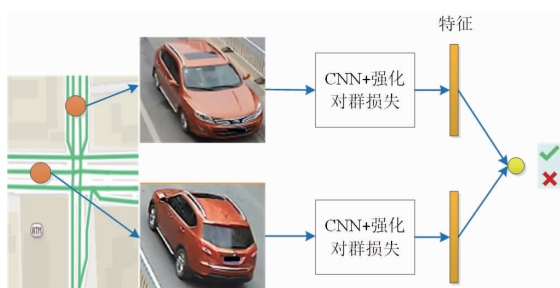


图4 基于特征聚类对群三元组损失的车辆再识别算法示意图

#### 3.3.1 基于 K-Means 算法的车辆样本集聚类

K-Means 算法于 1957 年提出,是一种主流的无监

督聚类分析方法常用于数据挖掘领域。本文采用 K-Means 算法来完成对车辆图片样本集自动划分为主要包含车头信息图片和主要包含车尾信息图片的两个样本集合的目标。

K-Means 算法可描述为输入样本集  $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$  和聚类簇数  $k$ , 获得划分后簇集合  $C = \{C_1, C_2, \dots, C_k\}$  要求满足平方误差  $E = \sum_{i=1}^k \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|_2^2$  取得最小值,其中  $\boldsymbol{\mu}_i$  是簇  $C_i$  内样本向量的均值。因为这个问题是一个 NP 难的,所以在求解过程中往往仅使用贪心策略获得近似解。此求解流程可描述为:步骤 1,从样本集中随机选取  $k$  个样本作为簇中心;步骤 2,当一个样本与某簇中心距离最小时即被划分至此簇;步骤 3,重新计算簇中心  $\boldsymbol{\mu}_i$ ;步骤 4,重复执行步骤 2、3 直到簇中心保持不变;步骤 5,输出簇划分。

本文利用 K-Means 算法在特征空间对车辆图片做二分类并利用获取的两个簇中心求解新的正样本特征中心,帮助设计了特征聚类对群三元组损失函数。具体设计过程在 3.1 和 3.2 节已作出说明。图 5 展示了使用 K-Means 算法做二分类的情形。

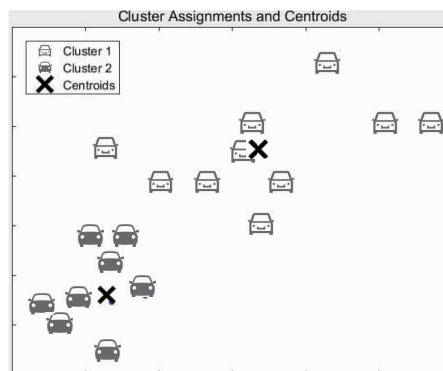


图5 K-Means算法做二分类仿真图

#### 3.3.2 ResNet 网络模型

ResNet 网络模型于 2015 年提出,因其简单实用及在分类任务中的出色效果,被广泛应用于检测、分割、识别等机器视觉领域。本文采用 ResNet 网络模型来提出车辆表观特征。

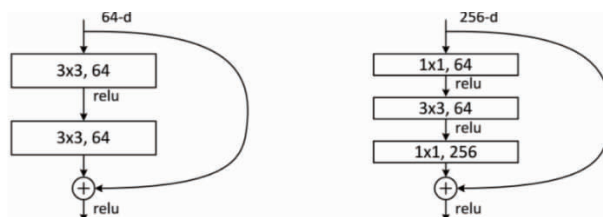


图6 ResNet网络模型中的两种残差单元

ResNet 的提出是为了解决以往深度神经网络存在随着网络加深训练集准确率下降的退化问题。ResNet

网络模型是在 VGG19 网络模型的基础上修改而来. 其改动主要体现在使用步长为 2 的卷积做下采样, 并将全连接层改为全局平均池化层. ResNet 网络模型的一个设计原则是: 为了保持网络层的复杂度, 当特征图大小减小一半时, 将特征图数量增加一倍. 另外 ResNet 网络模型通过抄近路机制引入了残差单元, 如图 6 所示, 为网络获得了残差学习的能力. ResNet 使用的残差单元有两种, 一种对应图左侧的两层网络, 一种对应图右侧的三层网络. 对于 18 和 34 层的 ResNet 网络模型采用两层网络进行残差学习, 对于更深的网络模型采用三层网络进行残差学习. 总结来说, ResNet 网络模型通过残差学习从根本上解决了深度神经网络的退化问题, 为训练出更深的优良网络模型提供了途径. 本文应用 ResNet 网络模型来提取车辆图片的表观特征.

### 3.3.3 详细的算法流程

首先, 利用 ResNet 深度卷积神经网络提取车辆表观特征, 对同一车辆的正样本图像采用基于聚类的正样本特征中心选取方法, 获得每个车辆的正样本特征中心  $c_i^p$ , 其中  $i=1, 2, \dots, N$ ,  $N$  为车辆的总数目. 具体步骤如下:

**步骤 1** 对于输入车辆图像, 利用深度卷积神经网络提取其视觉表观特征向量  $F = (F_1, F_2, \dots, F_i, \dots, F_N)$ ,  $F_i = (F_{i1}, F_{i2}, \dots, F_{iM})$ , 其中  $N$  为车辆的总数目,  $M$  为同一车辆不同图像的数目;

**步骤 2** 对每个车辆所有图像的视觉表观特征向量  $F_i = (F_{i1}, F_{i2}, \dots, F_{iM})$ ,  $i=1, 2, \dots, N$ , 运用 K-Means 算法将其聚为两类, 两个聚类中心分别为  $\mu_{i1}$  和  $\mu_{i2}$ ,  $i=1, 2, \dots, N$ ;

**步骤 3** 利用公式  $c_i^p = \frac{\alpha \mu_{i1} + \beta \mu_{i2}}{\alpha + \beta}$ ,  $i=1, 2, \dots, N$ , 计算每一车辆的正样本特征中心  $c_i^p$ ,  $i=1, 2, \dots, N$ , 其中  $\alpha$  和  $\beta$  为聚类中心加权系数, 满足  $N_{c_1} * \beta = N_{c_2} * \alpha$ ,  $N_{c_1}$  和  $N_{c_2}$  表示两个聚类包含的样本个数, 以保证正样本特征中心总是靠近大簇的中心, 更好地代表车辆图像组的主要表观特征.

其次, 使用正样本特征中心  $c_i^p$ , 设计特征聚类对群三元组损失函数  $L(W, F)$ , 对深度卷积神经网络参数进行重新训练, 得到新的网络模型及车辆表观特征  $F$ , 其中  $W$  为更新后的深度卷积神经网络参数. 具体步骤如下:

**步骤 1** 利用正样本特征中心  $c_i^p$  设计得到特征聚类对群三元组损失函数  $L(W, F) = \sum_{i=1}^N \sum_{l=1}^M \frac{1}{2} \max \{0, \max_{l=1, \dots, M} \{ \|F_{il}^p - c_i^p\|_2^2 \} - \min_{l=1, \dots, M} \{ \|F_{il}^n - c_i^p\|_2^2 \} + m \}$ , 其中  $N$  为车辆的总数目,  $M$  为同一车辆不同图像的数目,  $F_{il}^p$  是第  $i$  正样本集合的第  $l$  个特征向量,  $F_{il}^n$  是第  $i$  个负

样本集合的第  $l$  个特征向量,  $n$  表示此特征向量属于负样本,  $p$  表示此特征向量属于正样本,  $m$  是离正样本特征中心最远的正样本与正样本特征中心欧氏距离的平方与离正样本特征中心最近的正样本与正样本特征中心欧氏距离的平方的差值的最小边距.

**步骤 2** 基于特征聚类对群三元组损失函数对神经网络进行训练, 更新网络参数  $W$ .

**步骤 3** 得到新的网络模型并提取车辆图片新的表观特征  $F$ .

最后, 通过对待识别图像和数据库中车辆图像与提取到的表观特征  $F$  计算欧式距离, 衡量不同车辆图像与待识别图像的表观相似度  $D_i$ , 依照欧氏距离越小其为同一辆车的概率越大的规则, 完成对车辆的再识别.

## 4 实验结果与分析

本节对本论文提出的基于特征聚类对群三元组损失的车辆再识别算法性能进行评价. 首先介绍了本论文的实验环境及训练配置, 接着对基于特征聚类对群三元组损失的车辆再识别算法的有效性进行了验证, 最后将本论文算法与其他车辆再识别算法进行对比.

### 4.1 实验环境及训练配置

实验环境为: GeForce GTX 1080 Ti 显卡, 11GB 显存, Intel Core i7 处理器, 3.6GHz 主频, 64GB 内存; 平台为 Ubuntu 16.04, Cuda 9.0, cudnn 7.0.5.

训练配置为: 采用 ADAM 优化器; 在 VeRi-776<sup>[17]</sup> 数据集上总共训练 150 个 epochs, mini-batch 大小为 144, 一个 batch 包含 24 种 ID, 每个 ID 包含 6 张图片, 样本规格为  $128 \times 256$ ; 在 VehicleID<sup>[5]</sup> 数据集上总共训练 80 个 epochs; mini-batch 大小为 288, 一个 batch 包含 48 种 ID, 每个 ID 包含 6 张图片, 样本规格为  $128 \times 256$ ; 采用 ResNet-50 预训练模型, last\_conv\_stride 设置为 1 或采用 VGG16 预训练模型; 未将特征归一为单位长度; 只使用了水平翻转来做数据扩充.

实验中训练集和测试集均采用两个通用数据集中的默认设置. 其中 VeRi-776 数据集将所有 776 辆不同车辆中的 576 辆车的 37781 张图片作为训练集, 将剩余 200 辆车的 11579 张图片作为测试集. 表 1 中的小样本集、中样本集、大样本集指最常用的车辆再识别数据集之一 VehicleID 数据集中三种不同大小的测试集. VehicleID 数据集分为三种不同大小的测试集, 分别包含 800 辆车的 7332 张图片、1600 辆车的 12995 张图片、2400 辆车的 20038 张图片. 测试集划分如表 1 所示. 三种数据集均为测试集, 因此采用同样的训练模型.

基础网络模型可采用 VGG、GoogLeNet、ResNet、DenseNet 等任何一种基础模型. 对比算法中采用原方法的基础网络模型, 本文算法采用 ResNet-50 作为与训练模型.

表 1 VehicleID 数据集中的测试集结构

测试集规模	小样本集	中样本集	大样本集
图库图片数量	800	1600	2400
查询图片数量	6532	11395	17638

## 4.2 算法有效性验证

为了对本论文提出的特征聚类对群三元组损失函数的有效性进行验证,本文设计了下面的验证实验. 实验采用 top1 和 top5 命中率为评价指标. 本文沿用 CCL 表示,本文所提出的方法使用 CFCETL 来表示,其中特征聚类的改进为 FC,对群三元组损失函数的改进为 CCTL. 实验结果如表 2 所示.

表 2 特征聚类对群三元组损失有效性实验

命中率		小样本集	中样本集	大样本集
Top1	VGG + TripletLoss <sup>[5]</sup>	0.404	0.354	0.319
	VGG + CCL <sup>[5]</sup>	0.436	0.370	0.329
	VGG + CCL + FC	0.437	0.373	0.331
	VGG + CCTL	0.440	0.380	0.337
	VGG + CFCETL	<b>0.449</b>	<b>0.387</b>	<b>0.342</b>
Top5	VGG + TripletLoss <sup>[5]</sup>	0.617	0.546	0.503
	VGG + CCL <sup>[5]</sup>	0.642	0.571	0.533
	VGG + CCL + FC	0.657	0.583	0.544
	VGG + CCTL	0.670	0.591	0.562
	VGG + CFCETL	<b>0.685</b>	<b>0.612</b>	<b>0.575</b>

如表 2 所示,特征聚类对群三元组损失的车辆再识别算法在车辆再识别应用中能获得更好的性能. top1 和 top5 分别提升 1.3% ~ 1.7%、4.1% ~ 4.3% 了. 说明了使用本论文提出的基于聚类的正样本特征中心选取方法的特征聚类对群三元组损失函数,是优于随机选取一个正样本作为标准值的三元组损失及应用正样本算术均值作为标准值的对群损失的. 因此,本论文提出的基于聚类的正样本中心能够更好地度量汽车表现特征本质及特征聚类对群三元组损失能更好处理困难样本,从而有效提升车辆再识别准确率.

## 4.3 与其他算法对比

前面的实验验证了基于特征聚类对群三元组损失的车辆再识别算法的有效性,下面的实验将对基于特征聚类对群三元组损失的车辆再识别算法和五种当前基于表现信息的车辆再识别算法 LOMO<sup>[9]</sup>、DGD<sup>[8]</sup>、FACT<sup>[4]</sup>、Siamese Visual<sup>[2]</sup>、VGG + CCL<sup>[5]</sup> 在 VeRi-776<sup>[17]</sup> 和 VehicleID<sup>[5]</sup> 数据上进行比较. 本文按照车辆与行人再识别领域惯例,如文献[2,5,6]等,采用 mAP、Top1、Top5 和 CMC 曲线作为性能评价指标,mAP 和 CMC 曲线中包含了查全率的评估.

(1) 基于 LOMO 的车辆再识别算法. LOMO 是行人再识别领域提出的一种人工提取的局部特征,致力于解决

行人再识别领域中经常出现的视角及光照变化的问题.

(2) DGD 算法. 此算法是一种基于深度学习的特征提取方法,可以从不同领域数据中学习到一般性的鲁棒的特征. 通过在 VeRi-776<sup>[17]</sup> 及 VehicleID<sup>[5]</sup> 数据集上做再训练,获得一个适用于提取车辆特征的网络模型.

(3) 基于 FACT 的车辆再识别算法. 此算法是联合领域中的一种区分车辆的方法,FACT 由 SIFT 特征、颜色名及 GoogLeNet 特征组成的.

(4) VGG + CCL 算法,即基于对群损失的车辆再识别算法. 此算法通过将再识别领域流行的三元组损失改进为对群损失,并结合 VGG 深度神经网络模型来提取车辆图片的表观特征.

(5) Siamese-Visual 算法. 此算法使用分类用交叉熵损失来计算一对车辆图片的视觉相似度,完成对车辆表观特征的学习.

实验采用 top1 和 top5 命中率及累积匹配曲线图 CMC 作为评价指标. 实验结果如表 3 及图 7 所示,图 7(a)为对比算法的 CMC 曲线图,图 7(b)为本文算法的 CMC 曲线图. 图 8 和图 9 展示了部分实验效果. 图

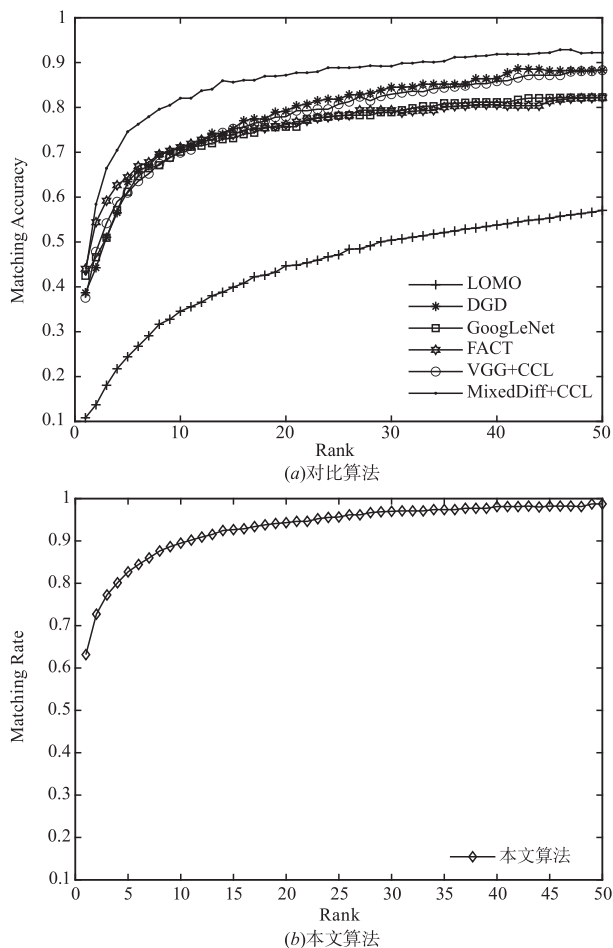


图 7 图库图片数量为 800 时 VehicleID 数据集上不同车辆再识别方法的 CMC 曲线

8 展示的是基于特征聚类对群三元组损失的车辆再识别算法在 VeRi-776<sup>[17]</sup> 数据上的部分实验效果. 第一列为查询图片, 图片带虚线边框表示和查询图片属于同一辆车, 图片带粗实线边框表示和查询图片不属于同一辆车. 图 9 是在 VehicleID<sup>[5]</sup> 数据集上车实验效果. 方框线型表示含义与图 8 相同, 需要说明的是 VehicleID<sup>[5]</sup> 数据集的图库图片里每一辆车只有一张图片, 所以最相似的前五张图片最多有一张是和查询图片属于同一辆车的. 可以看出实验效果不错.

本文基于特征聚类对群三元组损失的车辆再识别

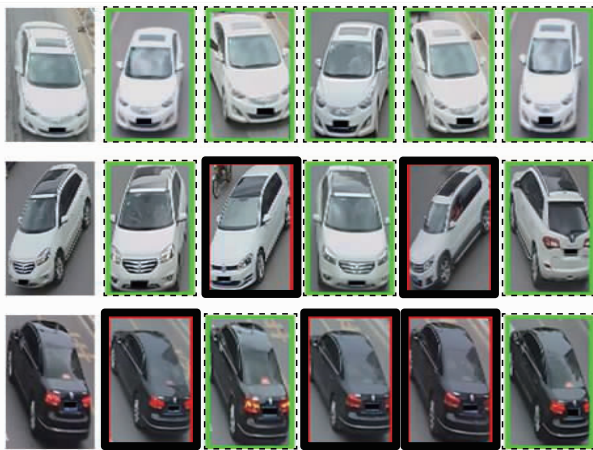


图8 基于特征聚类对群三元组损失的车辆再识别算法在 VeRi-776数据集上的部分实验效果

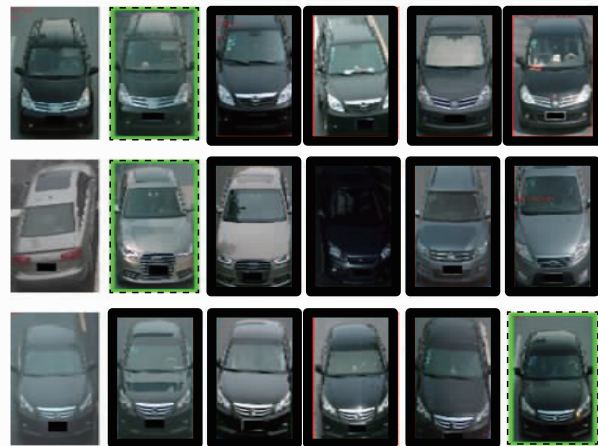


图9 基于特征聚类对群三元组损失的车辆再识别算法在 VehicleID数据集上的部分实验效果

算法分别在 VeRi-776 数据集<sup>[17]</sup> 和 VehicleID 数据集<sup>[5]</sup> 上进行了测试, 通过观察 mAP、top1、top5 评价指标, 本文算法在 VeRi-776 数据集上比 top1 性能最好的 Siamese-Visual 算法的 top1 提升了 13.11%, 比 top5 性能最好的 DGD 算法的 top5 提升了 7.21%, mAP 只比最高的 Siamese-Visual 算法稍稍下降, 在 VehicleID 数据集的小样本集上比四个主流算法中最高的 top1、top5 分别提升了 0.18%、7.57%. 由此说明本论文提出的基于特征聚类对群三元组损失的车辆再识别算法获取的车辆表观特征的精确性相比其他算法是更优的, 从而有效提升了车辆再识别准确率.

表 3 基于表观信息的车辆再识别算法对比实验

对比方法	VeRi-776 数据集 <sup>[17]</sup>			配置	VehicleID 数据集 <sup>[5]</sup>						
	Query = 1678 Test = 11579				方法	小样本集		中样本集		大样本集	
	mAP	Top1	Top5			Top1	Top5	Top1	Top5	Top1	Top5
LOMO	9.78	23.87	39.14	LOMO	19.76	32.01	18.85	29.18	<b>15.32</b>	25.29	
DGD	17.92	50.70	67.52	DGD	44.80	66.28	40.25	65.31	37.33	57.82	
FACT	18.73	51.85	67.16	FACT	49.53	68.07	<b>44.59</b>	64.57	<b>39.92</b>	60.32	
Siamese-Visual	<b>29.48</b>	41.12	60.31	VGG + CCL	43.62	64.84	39.94	62.98	35.68	56.24	
ResNet + CFCETL	29.32	<b>54.23</b>	<b>74.73</b>	ResNet + CFCETL	<b>49.71</b>	<b>75.64</b>	43.76	<b>67.91</b>	38.23	<b>63.05</b>	

## 5 结论

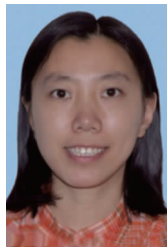
针对仅基于表观信息的车辆再识别方法中基于对群损失的方法存在着训练过程扩大了样本输入数量增加计算量的缺点及对群损失中没有考虑困难正样本情形且对标准值项的设计过于简单的问题, 本论文提出了一种基于特征聚类对群三元组损失的车辆再识别算法. 首先, 利用深度卷积神经网络提取车辆表观特征, 对正样本车辆图片采用基于聚类的正样本特征中心选取方法, 获得正样本特征中心. 其次, 使用正样本特征中心设计特征聚类对群三元组损失函数, 不扩增输入样本数量, 更新深度卷积神经网络参数, 得到新的网络

模型及新的车辆表观特征, 将查询车辆图片表观特征与目标库车辆图片表观特征欧氏距离值从小到大排序, 按照距离值越小与待识别车辆身份相同概率越大的规则完成对两幅图片是否属于同一辆汽车的判定. 实验结果表明本论文提出的基于特征聚类对群三元组损失的车辆再识别算法在处理困难样本上是优于基于对群损失的车辆再识别算法的. 同时, 在当前主流的 VeRi-776 数据集和 VehicleID 数据上, 与当前五种基于表观信息的车辆再识别算法进行了比较, 论文算法相比于当前主流的基于表观信息的车辆再识别算法能更好地提升再识别的准确率.

## 参考文献

- [1] Guo H, Zhao C, Liu Z, et al. Learning coarse-to-fine structured feature embedding for vehicle re-identification [A]. 32th AAAI Conference on Artificial Intelligence [C]. New Orleans, Louisiana, USA: AAAI, 2018. 6853 – 6860.
- [2] Shen Y, Xiao T, Li H, et al. Learning deep neural networks for vehicle Re-ID with visual-spatio-temporal path proposals [A]. 2017 IEEE International Conference on Computer Vision [C]. Venice, Italy: ICCV, 2017. 1900 – 1909.
- [3] Zapletal D, Herout A. Vehicle re-identification for automatic video traffic surveillance [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Las Vegas, Nevada, USA: IEEE, 2016. 25 – 31.
- [4] Liu X, Liu W, Ma H, et al. Large-scale vehicle re-identification in urban surveillance videos [A]. IEEE International Conference on Multimedia and Expo [C]. Seattle, WA, USA: IEEE, 2016. 1 – 6.
- [5] Liu H, Tian Y, Wang Y, et al. Deep relative distance learning: tell the difference between similar vehicles [A]. Computer Vision and Pattern Recognition [C]. Las Vegas, Nevada, USA: IEEE, 2016. 2167 – 2175.
- [6] Zhou Y, Shao L. Viewpoint-aware attentive multi-view inference for vehicle re-identification: computer vision and pattern recognition [A]. Computer Vision and Pattern Recognition [C]. Salt Lake City, UT, USA: IEEE, 2018. 6489 – 6498.
- [7] Wang Z, Wang X, Tang L, et al. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification [A]. 2017 IEEE International Conference on Computer Vision [C]. Venice, Italy: IEEE, 2017. 379 – 387.
- [8] Xiao T, Li H, Ouyang W, et al. Learning deep feature representations with domain guided dropout for person re-identification [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Las Vegas, Nevada, USA: IEEE, 2016. 1249 – 1258.
- [9] Liao S, Hu Y, Zhu X, et al. Person re-identification by local maximal occurrence representation and metric learning [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Boston, Massachusetts, USA: IEEE, 2015. 2197 – 2206.
- [10] Wu C W, Liu C T, Chiang C E, et al. Vehicle re-identification with the space-time prior [A]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops [C]. Salt Lake City, UT, USA: IEEE, 2018. 121 – 128.
- [11] Chen G, Lu J, Yang M, et al. Spatial-temporal attention-aware learning for video-based person re-identification [J]. IEEE Transactions on Image Processing, 2019, 28 (9): 4192 – 4205.
- [12] He B, Li J, Zhao Y, et al. Part-regularized near-duplicate vehicle re-identification [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Long Beach, CA, USA: IEEE, 2019. 3997 – 4005.
- [13] Liu X, Zhang S, Huang Q, et al. Ram: a region-aware deep model for vehicle re-identification [A]. 2018 IEEE International Conference on Multimedia and Expo (ICME) [C]. San Diego, USA: IEEE, 2018. 1 – 6.
- [14] Hermans A, Beyer L, Leibe B. In Defense of the Triplet Loss for Person Re-identification [EB/OL]. <https://arxiv.org/abs/1703.07737>, 2017.
- [15] Yadav J, Sharma M. A review of K-mean algorithm [J]. International journal of engineering trends and technology, 2013, 4(7): 2972 – 2976.
- [16] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Las Vegas, Nevada, USA: IEEE, 2016. 770 – 778.
- [17] Liu X, Liu W, Mei T, Ma H. A deep learning-based approach to progressive vehicle re-identification for urban-surveillance [A]. ECCV 2016 [C]. Amsterdam, Netherlands: Springer, 2016. 869 – 884.

## 作者简介



**吴燕雄** 女, 1983 年出生, 辽宁沈阳人, 于 2006 年和 2009 年分别获得北京航空航天大学学士学位和硕士学位。现为中国地震局防灾科技学院讲师, 并为中国地震局地球物理研究所博士生, 主要研究方向为智能信息处理、智能传感器和观测技术。

E-mail: wuyanxiong@cidp.edu.cn



**蔡建美** 女, 1978 年出生, 河北衡水人, 2001 年于河北科技大学获学士学位, 2003 年于燕山大学获硕士学位, 2010 年于北京工业大学获博士学位。现为防灾科技学院教授, 主要研究方向为: 机器学习、地震仪器监测技术、智能机器人技术。

E-mail: cjxlaq@163.com



**滕云天 (通信作者)** 男, 1966 年出生, 浙江兰溪人, 研究员, 博士生导师。1988 年毕业于电子工北京理工大学程系, 1991 年毕业于国家地震局地球物理研究所获硕士学位, 后留所参加工作。1999 年获北京大学工商管理硕士学位 (MBA), 2001 年在地震局地球物理研究所获博士学位。现任中国地震局地球物理研究所研究员, 防灾科技学院电子科学与控制工程学院学科带头人。研究方向: 地球探测与智能信息技术。

E-mail: tengyt@cea-igp.ac.cn