

基于 SqueezeNet 卷积神经网络的车辆检测

魏泽发¹, 崔华²

(1. 长安大学教育技术与网络中心, 陕西 西安 710064; 2. 长安大学信息工程学院, 陕西 西安 710064)

摘要: 在智能交通系统中, 针对车辆目标检测算法可移植性不高、检测速度较慢等问题, 提出了一种基于 SqueezeNet 卷积神经网络的车辆检测方法。通过融合 SqueezeNet 与 SSD (single shot multibox detector) 算法的车辆检测方法, 在 UA-DETRAC 数据集上进行训练, 实现了车辆目标的快速检测, 提升了模型的可移植性, 缩短了单帧检测时间。实验结果表明, 所提模型在保证准确率的同时, 模型单帧检测时间可达 22.3 ms, 模型大小为 16.8 MB, 相较于原 SSD 算法, 模型大小减少了约 8/9。

关键词: 智能交通; 卷积神经网络; SqueezeNet; 车辆检测

中图分类号: TP389.1

文献标识码: A

doi: 10.11959/j.issn.2096-3750.2020.00175

Vehicle detection based on SqueezeNet convolutional neural network

WEI Zefa¹, CUI Hua²

1. Educational Technology and Network Center, Chang'an University, Xi'an 710064, China

2. School of Information Engineering, Chang'an University, Xi'an 710064, China

Abstract: In the intelligent transportation system, aiming at the problem of low portability and speed of detection in vehicle target detection algorithm, a vehicle detection method based on SqueezeNet convolutional neural network was proposed. In order to realize the rapid detection of vehicle targets, improve the portability and shorten the detection time of the single frame, the model was trained on the UA-DETRAC dataset by fusing the SqueezeNet with the single shot multibox detector (SSD) algorithm. The experimental results showed that the time of the single frame detection could reach 22.3 ms and the model size was 16.8 MB. Compared with the original SSD algorithm, the model size was reduced by about 8/9. At the same time, the accuracy of the proposed model was guaranteed.

Key words: intelligent transportation, convolutional neural network, SqueezeNet, vehicle detection

1 引言

随着智能交通系统的不断发展和完善, 实现车辆的精准、快速检测成为智能交通系统中的关键。车辆检测是自动驾驶的重要组成部分, 在减少甚至避免交通事故方面具有重要意义。同时, 车辆检测可以为交通管理部门提供准确的数据支撑, 在交通管制、拥堵检测和信号配时等方面有广泛应用。

传统的车辆检测算法需要对图像进行预处理,

然后对整幅图像进行滑动窗口遍历操作, 通过初步判断车辆目标可能会出现的位置, 人工设计车辆目标的某种特征, 如常见的方向梯度直方图 (HOG, histogram of oriented gradient) 特征^[1]、尺度不变特征变换 (SIFT, scale-invariant feature transform) 特征^[2]、哈尔特征 (Haar-link feature) ^[3]和加速稳健特征 (SURF, speeded up robust feature) ^[4]等, 最后将特征送入支持向量机 (SVM, support vector machine) ^[5]或 Adaboost^[6]分类器进行分类, 完成检测

收稿日期: 2020-06-18; 修回日期: 2020-07-06

通信作者: 魏泽发, wzf@chd.edu.cn

基金项目: 陕西省科技厅重点研发计划 (No.2018ZDXM-GY-047)

Foundation Item: The Key R&D Plan of Shaanxi Provincial Science and Technology Department (No.2018ZDXM-GY-047)

任务。但是人工设计特征存在很大弊端，在设计过程中会过于依赖以往经验，且算法在陌生场景下的表现效果不好，检测算法的稳健性不强，极大地阻碍了检测算法的应用。

随着深度学习在各个领域取得了较大突破，目标检测将面临新的发展机遇。深度学习目标检测算法在检测准确率上有巨大的提升，主要得益于卷积神经网络强大的特征提取能力。卷积神经网络可以自动提取目标中的关键特征信息，不需要人工设计和经验，只需要向网络中传递足够多的图像数据即可，因此，该算法具有更高的稳健性，更适应不同场景，在实际应用中具有较大优势。

基于深度学习的目标检测算法主要包括两类，即 one-stage 方法和 two-stage 方法。其中，two-stage 方法首先需要使用生成算法产生一系列候选框，然后在候选框上进行回归和分类操作，该方法的特点是检测准确率高、检测速度较慢。2014年，Girshick等^[7]提出了 R-CNN (region-CNN) 目标检测算法，利用选择性搜索 (selective search)^[8]方法生成候选框，然后将候选框送入卷积神经网络提取特征，最后将特征输入 SVM 分类器进行回归运算。为了改善 two-stage 方法中存在的问题，研究人员提出了 one-stage 方法，其主要思路是摒弃候选框生成过程，直接利用卷积神经网络在图像数据上进行卷积操作，然后在后续特征图上不断提取信息，最后从提取出的信息中取部分特征层完成最终的检测任务，该方法的检测速度较快，但是检测准确率较低。2016年，YOLO (you only look once) 目标检测系列算法^[9-11]很好地解决了算法的实时性问题，通过将检测和分类两个过程整合为一个过程，在每个特征单元上预测检测框的位置和类别，然后结合图像中的背景信息在整个图像特征上进行预测，虽然检测准确率下降了，但该算法为未来的研究工作提供了很好的思路 and 方向。SSD 检测算法^[12]很好地结合了 R-CNN 目标检测算法和 YOLO 目标检测系列算法的优势，继承了 R-CNN 算法的 anchor 机制和 YOLO 算法的回归思想，通过在多个尺度的特征图上生成不同长宽比的候选框，实现对各种尺寸目标的检测。相比于 R-CNN 目标检测算法和 YOLO 目标检测系列算法，SSD 算法在各个领域中的应用更广泛。

本文提出一种基于 SqueezeNet^[13]与 SSD 算法融合的车辆目标检测算法，利用轻量级卷积神经网络 SqueezeNet 的特性，在保证同等检测准确率的同时，

大幅度减少模型的参数量，且缩短了模型的单帧检测时间。这为将算法移植到如现场可编程逻辑门阵列 (FPGA, field programmable gate array) 这类移动开发板上提供了可能，也意味着服务器在同等时间内可以处理更多数据，大幅度提升了服务器的利用效率。

2 车辆检测技术及原理

2.1 SqueezeNet 网络结构

轻量级卷积神经网络 SqueezeNet 的特点在于 fire module 结构，fire module 结构如图 1 所示，该结构由 squeeze 层和 expand 层构成。其中，squeeze 层全部为尺寸是 1×1 的卷积核，特征图经过 squeeze 层后，特征通道数减少，且相比于尺寸是 3×3 的卷积核参数量减少了 $8/9$ 。同时，squeeze 层的输出将作为 expand 层的输入，通道数的减少将使整个网络的参数量大幅度减少；expand 层中包含 1×1 的卷积核和 3×3 的卷积核，特征图分别经过两种卷积核后通过 concat 操作输出，特征图在经过 fire module 结构后，其尺寸不会发生变化，但是相较于特征图只经过一种 3×3 卷积核后，其尺寸不变但参数量得到了很大的缩减，这是 SqueezeNet 被称为轻量级网络的根本原因。

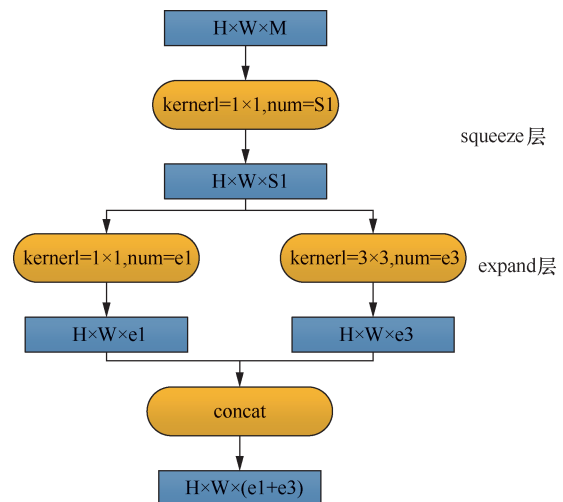


图1 fire module 结构

2.2 车辆目标检测原理

SSD 算法在目标检测任务中采用多尺度的特征图进行预测，同时将在多个特征层中进行预测框的回归和类别预测，SSD 算法的核心理念可以总结为以下 3 点。

1) 不同尺度特征图产生先验框

在卷积神经网络结构中，由于卷积和池化操

作,特征图的尺寸会随着网络的加深而不断变小。浅层卷积层分辨率高,提取的特征包含更多的细节信息,如边缘、形状等几何信息;深层卷积层分辨率低,提取的特征包含更多的语义信息,所表达的信息更抽象。由于在检测任务中目标尺寸不同,如果仅使用浅层卷积层的特征信息进行预测,会使得小目标的检测效果不好,因此,SSD算法采用多尺度特征图进行预测,即利用具有较大感受野的深层卷积特征预测大目标,利用具有较小感受野的浅层卷积特征预测小目标。

2) 利用卷积层完成检测,摒弃全连接层

在目标检测算法的研究过程中,网络中的全连接层并没有被舍弃,导致网络中的参数量很大,参数过多使得特征无法有效向后传递且模型的训练较困难。因此,SSD算法将原来的全连接层改为卷积层,同时在网络后增加若干个卷积层,更有利于图像数据中特征的表达。

3) 根据任务设置先验框

SSD算法继承了Faster R-CNN中anchor的概念,以每个特征层为单元设置不同尺度、不同长宽比的先验框,在训练时比较预测边界框与先验框之间的位置差异,计算两者的损失函数。先验框的生成规则为:首先以每个单元的中心为准,生成矩形框,其中正方形先验框的边长最小值为 \min_size ,最大值为 $\sqrt{\min_size \times \max_size}$;然后在网络中配置一个参数为纵横比($aspect_ratio$),并结合上述的最小值和最大值得到两个矩形框的长和宽,分别为 $\sqrt{aspect_ratio \times \min_size}$ 和 $1/\sqrt{aspect_ratio \times \min_size}$,每个特征图中的最大值和最小值由式(1)确定。

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(k-1) \quad (1)$$

其中, m 表示特征图个数, k 的取值范围是 $1 \sim m$, s_k 表示先验框大小相对于图片的比例, s_{\min} 和 s_{\max} 分别表示比例的最小值和最大值。

SSD算法会在多个尺度的特征图上生成不同尺寸的先验框,将这些先验框和预测框进行匹配并计算,实现正、负样本的划分,最终完成训练,更新了模型中的参数。

2.3 SqueezeNet与SSD算法融合

SSD算法由基础网络、检测网络和分类网络3部分组成,SqueezeNet与SSD算法融合首先将检测算法的基础网络由VGG(visual geometry group)

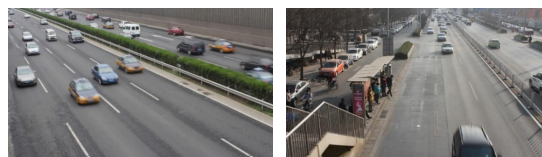
替换为SqueezeNet,使得整个检测架构的参数量大幅度减少。在上节提到了SSD检测算法需要对尺度特征图进行检测,因此,特征层的选取很关键。根据原始SSD算法的设计理念,本文在SqueezeNet的基础上增加了fire10和fire11两个卷积层,并且选用fire5、fire9、fire10和fire11这4个卷积层输出的特征图进行分类和检测框的回归计算。

3 实验结果与分析

3.1 UA-DETRAC数据集

UA-DETRAC数据集^[14]共有100个视频序列,视频采用EOS 550D相机在全国范围内24个不同路口进行拍摄,视频帧率为25帧/s,图像的分辨率为960 ppi×540 ppi。拍摄场景包含很多常见的交通场景,如城市道路场景、十字路口场景和交通拥堵场景等,UA-DETRAC数据集如图2所示。数据分为训练集和测试集,训练集DETRAC-Train-Images包含60个视频序列,每个序列存放在一个文件夹中,被裁剪成 n 个jpg格式的图片;测试集包含40个视频序列,但不包含真实的标签值。

在使用UA-DETRAC数据集进行训练前,需要将数据集转换成VOC数据集的数据格式。首先,将DETRAC-Train-Annotations-XML文件夹中每个xml文件转换成VOC数据集中xml的样式,样式中应包含图片名称、图片尺寸、目标类别(car格式)以及每个目标的位置(包括xmin、ymin、xmax和ymax共4个坐标值),转换后根据生成的xml文件将对应的、单独存放的图片迁移出来并存放在一个总文件夹中。由于截取同一个视频序列得到的图像有重叠,因此,本文从所有图像中随机抽取60%的图像作为训练样本,最后生成trainval.txt、test.txt、train.txt、val.txt共4个文件,用于模型训练。



(a) 场景一

(b) 场景二

图2 UA-DETRAC数据集

3.2 模型训练

由于训练卷积神经网络需要进行大量的矩阵运算,普通计算机无法完成庞大的计算任务,因此,本次训练采用了由NVIDIA开发的Titan X1080 Ti

显卡, 采用 caffe (convolutional architecture for fast feature embedding) [15]深度学习框架, 同时配置 Python、OpenCV 和 CUDA (compute unified device architecture) 等软件包和软件库。

模型训练主要分为前向传播和反向传播两个过程。1) 前向传播: 图像数据经过若干卷积操作, 每一层输出的特征图将作为下一层的输入, 依次向后传递, 其中会经过激活层的非线性映射和池化层的降维, 最终网络可以学习到图像数据中包含的关键信息。2) 反向传播: 随机梯度下降是反向传播过程中最关键的因素, 通过随机梯度下降算法计算输出分类结果和数据标签之间的误差, 从而将误差依次向前传递, 更新每一层中的权重, 当误差收敛到一定程度时, 则训练完成。

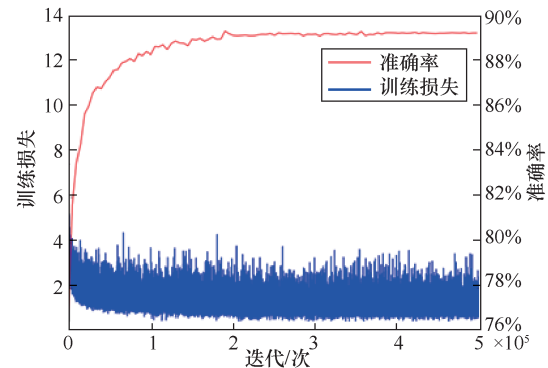
在训练时, 设置基础学习率为 0.001, 学习率随着迭代次数每 200 000 次降为原来的 1/10, 总迭代次数为 500 000 次, 即共迭代 100 轮次 (epoch) (所有样本完成一次训练过程为一个轮次), 动量 (momentum) 是指训练损失函数值下降的趋势, 其值为 0.9, 权重衰减值为 0.000 5。数据集图片共有 49 251 张, 其中 39 400 张图片用于训练, 9 851 张图片用于测试。模型训练结果如图 3 所示, 由图 3 可知, 两个模型训练最终都达到收敛, 但以 SqueezeNet 为基础模型的训练结果曲线更平滑, 震荡较小, 且两者最终的训练准确率几乎一致。

3.3 模型测试

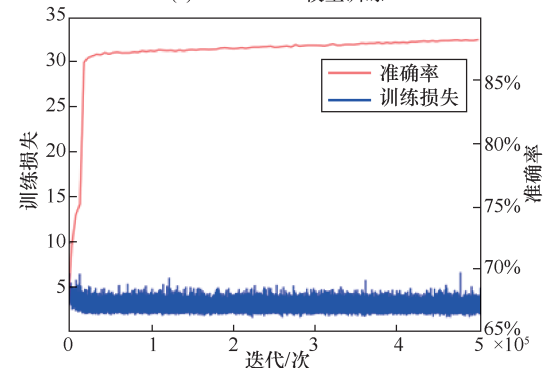
在实验中, 本文研究了以 MobileNet、ShuffleNet 和 SqueezeNet 为基础网络时与 SSD 算法融合的结果, 与原始模型对比结果如表 1 所示。从表 1 可以看出, MobileNet 和 SqueezeNet 作为基础网络时可以达到和原始网络近乎一致的准确率, 但是 SSD-MobileNet 与 SSD-SqueezeNet 相比模型更大、训练时间以及单张测试时间更长, 因此, SSD-SqueezeNet 的结果优于 SSD-MobileNet 的结果。同样可以看到, 尽管 SSD-ShuffleNet 的模型大小、训练时间和单张测试时间基本和 SSD-SqueezeNet 的相同, 但是准确率低很多。

从模型大小、训练时间、单张测试时间和准确率这 4 个指标综合来看, 以 SqueezeNet 为基础模型的检测算法最终的训练准确率可达 88.14%, 和原始模型的准确率几乎一致, 但是模型的单张测试时间仅为 18.2 ms, 相比于原始算法的检测时间缩短了近 1/2, 同时模型大小仅为 11.6 MB, 是原始模型大小的

1/9。由此得出结论, 将 SqueezeNet 卷积网络与 SSD 检测算法融合可以实现目标的精准、快速检测, 且模型大小仅为十几兆, 这为将模型移植至微型计算机或者移动开发板上提供了可能。



(a) SSD-VGG16模型训练



(b) SSD-SqueezeNet模型训练

图3 模型训练结果

表1 与原始模型对比结果

模型名称	模型大小/MB	训练时间/h	单张测试时间/ms	准确率
SSD-VGG16	95.5	65	30.6	89.2%
SSD-MobileNet	22.2	57	26.3	88.7%
SSD-ShuffleNet	14.5	36	16.1	82.4%
SSD-SqueezeNet	11.6	42	18.2	88.14%

3.4 改进算法结果分析

3.3 节的实验结果表明, 将 SqueezeNet 网络与 SSD 算法融合的检测效果较好。但是总体来看, 模型的检测准确率仍有较大的提升空间, 同时模型对于数据集中的小目标车辆检测效果不好, 为了改善上述问题, 本文采取两种操作: 1) 增加 fire5、fire9、fire10、fire11 这 4 个特征层之间的联系; 2) 增加 fire5、fire9、fire10、fire11 每个特征层输出特征图的数量, 使其能够提取更多有效特征, 完成针对小目标的检测, 从而提升检测准确率, 改进后算法的特征金字塔结构如图 4 所示。

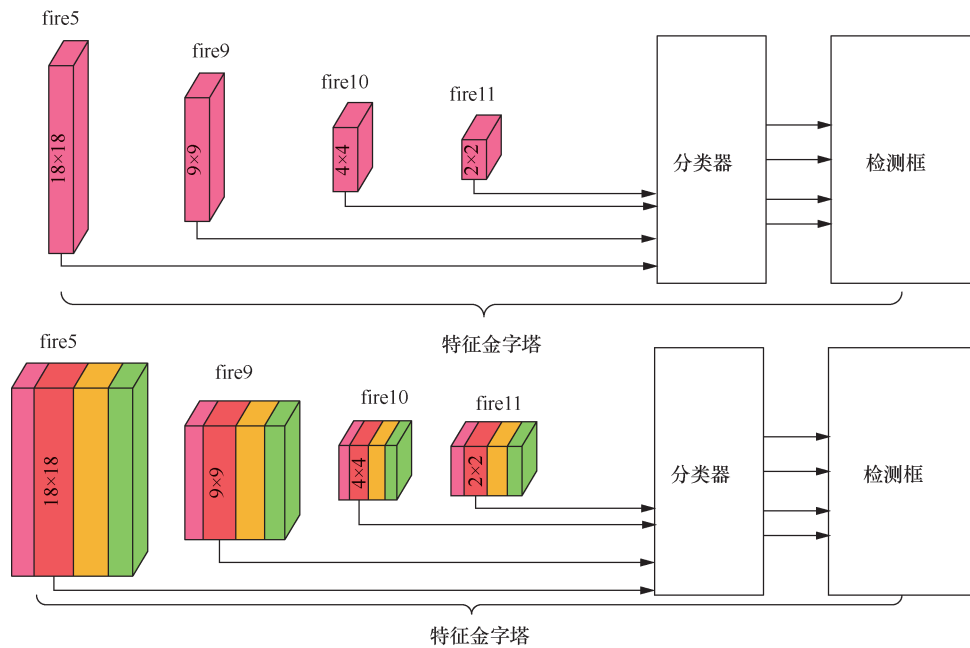


图 4 改进后算法的特征金字塔结构

图 4 主要分为上、下两部分，上部分为 SSD-SqueezeNet 的原始结构，即在 fire5、fire9、fire10、fire11 这 4 个特征层上生成先验框，然后输入分类器中完成目标检测任务；下部分为针对 SSD-SqueezeNet 的改进，分为两个阶段。1) 首先从左向右依次在每个特征层上进行池化操作，然后将上层池化结果和下层进行拼接，整体再进行池化并且和下一层拼接，依次往复直到最后一层；2) 从右向左进行反卷积^[16]操作，将反卷积后的结果和上一层拼接，依次向上直到第一层。需要注意的是，在进行拼接前应进行归一化操作，采用 Batch Normalization^[17]方式完成归一化。

在训练时，模型的环境、参数设置和 3.2 节中的设置保持一致，由于模型改进借鉴了 Rainbow-SSD 算法^[18]的设计理念，因此，将训练得到的模型命名为 SSD-SqueezeNet-R，改进前、后模型对比结果如表 2 所示。从表 2 的统计结果可以看出，虽然 SSD-SqueezeNet-R 在模型大小和单张测试时间等性能方面略差，但改进后模型的检测准确率得到提升，可达 90.63%，提升了 2.49 个百分点。

因此，可以得知模型改进方法是有效的。

4 结束语

本文提出了一种基于 SqueezeNet 卷积神经网络的车辆目标检测算法，解决了在智能交通系统中关于车辆目标检测算法可移植性不高、检测速度较慢等问题，该算法通过将 SqueezeNet 与 SSD 算法融合，同时根据融合后的模型表现采取针对性改进方式，在保证模型检测准确率的同时，降低网络参数量，缩短模型的单帧检测时间，实现了对车辆目标的精准、快速检测，为模型的移植、开发提供了可能。因此，利用该技术的车辆目标检测算法可以应用于交通摄像头、车载相机等设备拍摄的场景，对自动驾驶以及交通管理部门工作效率的提升具有重要意义。

参考文献:

[1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[J]. IEEE Computer Society, 2015, 1: 886-893.
 [2] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):

表 2 改进前、后模型对比结果

模型名称	模型大小/MB	训练时间/h	单张测试时间/ms	准确率
SSD-SqueezeNet	11.6	42	18.2	88.14%
SSD-SqueezeNet-R	16.8	50	22.3	90.63%

- 91-110.
- [3] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[J]. Institute of Electrical and Electronics Engineers Computer Society, 2001, 1(1): 1511-1518.
- [4] BAY H, ESS A, TUYTELAARS T, et al. SURF: speeded up robust features[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [5] SÁNCHEZ A V. Advanced support vector machines and kernel methods[J]. Neurocomputing, 2003, 55(1): 5-20.
- [6] VIOLA P, JONES M J. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014: 580-587.
- [8] UIJLINGS J R R, SANDE K E A V D, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Computer Vision and Pattern Recognition. 2016: 779-788.
- [10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 6517-6525.
- [11] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv: 1804.02767, 2018.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[M]. Berlin: Springer, 2016.
- [13] IANDOLA F N, HAN S, MOSKEWICZ M, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size[J]. International Conference on Learning Representations. 2017.
- [14] WEN L Y, DU D W, CAI Z W, et al. UA-DETRAC: a new benchmark and protocol for multi-object detection and tracking[J]. Computer Vision and Image Understanding, 2020, 193: 102907.
- [15] JIA Y Q, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[C]//ACM International Conference on Multimedia. ACM, 2014: 675-678.
- [16] ZEILER M D, TAYLOR G W, FERGUS R. Adaptive deconvolutional networks for mid and high level feature learning[C]//2011 International Conference on Computer Vision. IEEE, 2011: 2018-2025.
- [17] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. Journal of Machine Learning Research, 2015: 448-456.
- [18] JISOO J, HYOJIN P, NOJUN K. Enhancement of SSD by concatenating feature maps for object detection[J]. Computer Vision and Pattern Recognition, 2017.

[作者简介]



魏泽发（1992-），男，河北衡水人，长安大学助理工程师，主要研究方向为智能交通、深度学习、图像处理等。



崔华（1977-），女，陕西西安人，博士，长安大学教授，主要研究方向为图像分析、数据挖掘、机器学习、深度学习及其在智能交通中的应用研究等。

高级顾问

国际电信联盟秘书长

赵厚麟

中国科学院院士

李 未 姚建铨 朱中梁 陈国良 吴培亨 何积丰
吴一戎 徐宗本 梅 宏 王 巍 吕 建 郝 跃
陆建华 房建成 黄 如

中国工程院院士

倪光南 孙 玉 薛禹胜 邬贺铨 贲 德 邬江兴
柴天佑 刘韵洁 戴 浩 于 全 吴曼青 段宝岩
王广基 何 友 费爱国 桂卫华 余少华 姜会林
钱 锋 樊邦奎

第一届编辑委员会

主任委员：尹 浩

常务副主任委员：朱洪波

副主任委员：

刘 多 郭英杰 刘海涛 刘云浩 刘明亮 宋 彤 刘华鲁

委 员：（按姓氏笔画排序）

丁进良 丁志国 丁恩杰 于宏毅 卫 国 马华东 马建国
马建峰 王 伟 王 翔 王文博 王永建 王汝言 王红刚
王丽宏 王现斌 王金龙 王承祥 王重钢 王祖林 王新兵
韦 岗 牛志升 毛军发 仇洪冰 卢光跃 兰巨龙 吕卫锋
朱 艳 朱 琦 乔 辉 刘 飞 刘 云 刘元安 江 涛
安建平 孙 怡 孙玲玲 阳春华 杜文莉 李安民 李建东
李宣东 李德识 杨 旻 杨 健 杨龙祥 时龙兴 吴启晖
冷甦鹏 汪一鸣 沈八中 沈连丰 宋令阳 张 平 张 伟
张 彦 张 瑞 张云勇 张文生 张宏科 张英海 张钦宇
张海林 张朝阳 张登银 陆宇平 陈 岚 陈 钟 陈山枝
陈少平 陈仪香 陈后金 陈如山 陈前斌 陈晓华 陈积明
范九伦 范平志 林金朝 易东山 罗军舟 金 石 金 海
周兴社 周建江 郑 淦 郑志彬 郑福春 赵小强 赵晓晖
胡清阳 胡瑞敏 钟章队 施 毅 姜 斌 洪 文 姚 力
夏香根 钱 毅 高 跃 高飞飞 郭 庆 郭贵生 黄继杰
崔曙光 隆克平 续合元 葛 宁 葛俊祥 程崇虎 曾凡仔
谢 斌 詹 杰 潘 炜

物联网学报

Chinese Journal on Internet of Things



2020年9月30日出版 定价：60.00元

ISSN 2096-3750



9 772096 375201