



计算机工程与应用  
Computer Engineering and Applications  
ISSN 1002-8331, CN 11-2127/TP

## 《计算机工程与应用》网络首发论文

题目: 轻量化高精度卷积神经网络的安全帽识别方法  
作者: 陈柳, 陈明举, 薛智爽, 罗仕胜  
网络首发日期: 2020-10-26  
引用格式: 陈柳, 陈明举, 薛智爽, 罗仕胜. 轻量化高精度卷积神经网络的安全帽识别方法. 计算机工程与应用.  
<https://kns.cnki.net/kcms/detail/11.2127.TP.20201023.1727.009.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 轻量化高精度卷积神经网络的安全帽识别方法

陈柳, 陈明举\*, 薛智爽, 罗仕胜

四川轻化工大学 人工智能四川省重点实验室, 四川 自贡 643000

**摘要:** 由于施工环境的复杂性, 基于机器视觉的安全帽识别方法常常出现误检与漏检的情况。为提高复杂环境下安全帽识别的准确率, 同时满足实时性的要求, 本文提出一种基于视觉感受野特性的轻量化高精度卷积神经网络。该卷积神经网络以 RFBnet 网络为基础, 增加特征金字塔网络模块, 使神经网络同时兼顾浅层语义信息和深层语义信息的表示能力, 以实现复杂施工环境下不同形态与大小安全帽的识别。进而采用 SE-Ghost 模块在保持网络特征提取能力不变的情况下, 对网络的主干网络结构进行轻量化。为验证本文方法的性能, 将本文基于感受野特性的轻量化卷积神经网络和当前主要卷积神经网络进行实验对比, 结果表明, 本文模型的检测准确率较 YOLO-v3、RFBnet-300 和 RFBnet-512 网络分别提高了 1.6、3.62 和 0.98, 检测速度达到 20 帧/s。

**关键词:** 安全帽检测; 卷积神经网络; 感受野结构; 特征金字塔

**文献标志码:** A **中图分类号:** TP391.4 **doi:** 10.3778/j.issn.1002-8331.2006-0354

陈柳, 陈明举, 薛智爽, 等. 轻量化高精度卷积神经网络的安全帽识别方法. 计算机工程与应用  
CHEN Liu, CHEN Mingju, XUE Zhishuang, et al. Lightweight and High-precision Convolutional  
Neural Network for Helmet Recognition Method. Computer Engineering and Applications

## Lightweight and High-precision Convolutional Neural Network for Helmet Recognition Method

CHEN Liu, CHEN Mingju\*, XUE Zhishuang, LUO Shisheng

School of Information Engineering, Sichuan University of Science & Engineering, Zigong,  
Sichuan 643000, China

**Abstract:** Due to the complexity of the construction environment, machine vision-based helmet recognition methods often have false detections and missed detections. In order to improve the accuracy of helmet recognition in complex environments and satisfy the real-time requirement, this paper proposes a deep convolutional neural network based on the characteristics of the receptive field. The convolutional neural network is based on the RFBnet network, and a feature pyramid network module is embedded to the neural network to represent the shallow semantic information and the deep semantic information simultaneously, so as to improve the accuracy of helmet recognition with different shapes and sizes in complex construction environments. SE-Ghost module is used to lighten the network backbone and reduce the complexity of the network without network feature extraction capability unchanged. In the experiment, the proposed receptive field-based convolutional neural network is compared with current main convolutional neural network, the results show that the proposed network has higher accuracy than YOLO-v3, RFBnet-300 and RFBnet-512 networks, and increased 1.6, 3.62 and 0.98 percent respectively,

**基金项目:** 四川省科技厅项目 ( No.2018GZDZX0043, No.2019YJ0476, No.2019YJ0477 ) ; 人工智能四川省重点实验室开放基金项目 (No.2020RZY02)。

**作者简介:** 陈柳(1995-), 男, 硕士研究生, 研究领域为计算机视觉; 陈明举(1982-), 男, 博士, 副教授, 研究领域为图像处理, E-mail: mailcmj@163.com; 薛智爽(1995-), 男, 硕士研究生, 研究领域为图像处理; 罗仕胜(1995-), 男, 硕士研究生, 研究领域为计算机视觉。

while the detection speed reached 20 frame/s.

**Key words:** helmet detection; convolutional neural network; receptive field structure; feature pyramid

## 1 引言

随着电子信息技术的发展,机器视觉技术已逐渐应用到社会生产生活的各个领域。在电力生产、施工现场以及交通等领域,未正确佩戴安全帽是造成安全事故发生主要因素之一。通过机器视觉技术能实现对施工作业现场安全帽的检测与识别,减少安全事故的发生。

近年来,一些科研人员对机器视觉技术实现安全帽识别进行相应的研究,并获得了一些有效的研究成果。在传统机器学习算法研究方面,主要通过安全帽的形状与颜色特性进行识别。如 Park 等人<sup>[1]</sup>利用颜色直方图与空间关系匹配实现对安全帽的检测。李琪瑞等人<sup>[2]</sup>通过 ViBe 运动目标检测,并结合 SVM 的特征分类算法,以判断是否佩戴安全帽。刘晓慧等人<sup>[3]</sup>通过头部区域的 Hu 矩特征进行训练得到分类器,从而实现对安全帽的判断。传统的基于机器学习的安全帽检测算法,检测速度较快,但严重依赖人为特征设计和选取,同时由于特征单一、泛化能力较差,无法对复杂施工环境下的目标进行有效检测,导致检测准确率较低。

基于深度学习的卷积神经网络技术在图像处理中表现出较好的性能,越来越多的科研人员将卷积神经网络应用于图像的分类识别中,并取得了较好的识别效果。基于神经网络的目标检测主要分为两阶段网络和单阶段网络方法。两阶段网络通过区域提取和分类两个阶段去实现识别,具有较高的检测精度,但由于检测过程存在大量的数学计算从而导致检测速度较慢<sup>[4][5]</sup>。单阶段网络使用目标的高级特征直接进行识别,避免复杂的特征提取与检索算法,大大减少运算时间<sup>[5]</sup>,但由于单阶段网络对各级网络特征单独进行预测,底层特征缺少语义与全局信息,造成小物体漏检的情况发生<sup>[6]</sup>。

由于生产作业的场景错综复杂,采用单阶段卷积神经网络对作业人员是否佩戴安全帽进行识别,存在漏检现象。采用两阶段

网络存在运算时间长,无法满足实时性的要求。为满足对作业场景佩戴安全帽识别的精度和速度的要求,本论文以单阶段检测网络 RFBnet<sup>[7]</sup>为基础,引入多尺度特征金字塔结构增强对小目标的检测能力,并轻量化网络模块提高检测效率。实验中采用 SCUT-HEAD 数据集进行训练和测试,将本文改进的轻量化 RFBnet 网络与 RFBnet-300<sup>[7]</sup>、RFBnet-512<sup>[7]</sup>、YOLO-v3<sup>[8]</sup>三种检测方法进行对比试验,结果表明,本文改进的 RFBnet 检测方法能够保持较快的检测速度,同时能够实现较高的检测精度。

## 2 RFBnet 卷积神经网络基本结构

强大的特征提取能力在目标检测网络中变得至关重要,提取精确的目标特征需要增加卷积神经网络的层数,然而性能的不断提高带来的却是高额的计算成本,导致模型检测速度降低。受人类视觉系统中感受野(Receptive Field, RF)结构的启发而诞生的 RFBnet 网络是一种简单而高效的目标检测网络。

RFBnet 目标检测网络以 VGG-16 网络<sup>[9]</sup>作为提取特征的主干网络,将 VGG-16 网络最后的三个全连接层采用空洞卷积层 Conv6 和 Conv7 进行替换,并对池化层 Pool5 的步长和大小进行改变,实现在不增加网络运算量的前提下使卷积层能够提取到更多的特征信息,实现图像语义信息准确分割<sup>[10]</sup>。另外,为了更好地增强浅层卷积神经网络的特征表示能力,有效提取目标的特征, RFBnet 模拟人类浅层视网膜图中的群体感受野(population Receptive Field, pRF),构建以多分支卷积模块与膨胀卷积层为组件的 RF Block,以提取目标更精细的特征<sup>[7]</sup>。RFBnet 在 Conv4\_3 层连接小群体感受野的 RF Block,在 Conv7 层之后连接多个 RF Block,如图 1 所示。根据网络输入图片的大小, RFBnet 又有 RFBnet-300 和

RFBnet-512 两个版本,以适用输出特征图不同分辨率的要求。

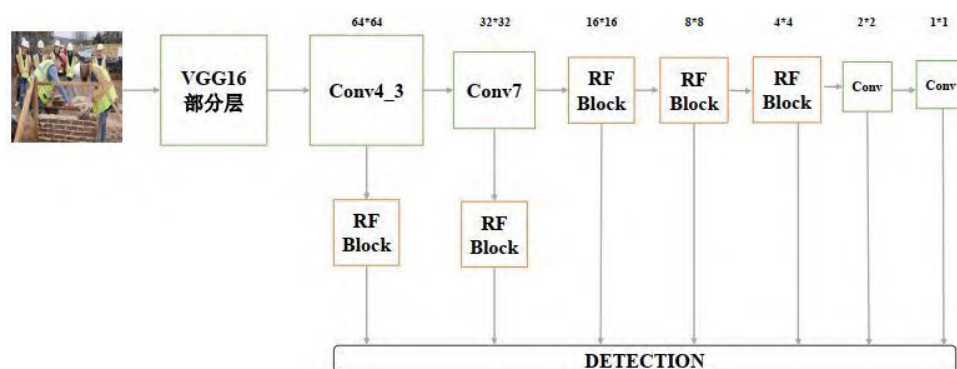


图 1 RFBnet-512 结构

Fig.1 Structure of RFBnet-512

由上分析可知, RFBnet 采用空洞卷积层与 RF Block 模块实现以较少的层数增强浅层卷积神经网络的特征表示能力, 然而 RFBnet 强调同层之间特征表示, 忽略了高层特征与底层特征之间的相互联系, 使得浅层特征表示能力有限, 从而造成复杂环境下小目标的漏检。

### 3 一种轻量化高精度卷积神经网络

针对施工环境的复杂性, 为达到安全帽检测精度的要求, 本文选择在检测效果更佳的 RFBnet-512 的基础上进行改进。为了更好地利用不同层次特征之间的联系, 有效地增强浅层特征, 将特征金字塔 (Feature Pyramid Networks, FPN) 模块引入到 RFBnet 中。同时保持网络特征提取能力不变的情况下减少计算量, 本文构建一种轻量化模块代替网络中的部分卷积模块, 从而建立一种实现安全帽检测的轻量化高精度卷积神经网络。

#### 3.1 引入 FPN 的 FP+RFBnet

在目标检测网络中, RFBnet 通过卷积层和池化层的不断堆叠, 导致深层特征图分

辨率较低, 不能对小目标进行有效检测。特征金字塔 FPN 模块<sup>[11]</sup>通过深层特征进行上采样, 然后将上采样得到的特征与次一级特征图进行融合, 实现高层特征向浅层特征融合, 使浅层特征也具有较强的语义信息, 并且在多个融合尺度上进行预测, 增强对小目标的检测能力<sup>[11]</sup>, 其网络结构如图 2 所示。

为了实现适用于复杂环境下多形态安全帽的检测, 并降低安全帽漏检的概率, 本文将 FPN 模块引入到 RFBnet 结构中。RFBnet 网络层数较大的深层特征图不含有小目标信息, 在过小的特征图上进行上采样融合, 反而会使目标的检测效果变差, 影响整个网络的识别能力。本文针对施工环境安全帽图像的特性, 在 RFBnet 中输出  $8 \times 8$  特征图的特征层处添加自上而下的 FPN 结构, 通过上采样生成  $16 \times 16$  的特征图, 然后与低一层  $16 \times 16$  的特征图进行对应融合, 依次自上向下融合直到 Conv4\_3 层。改进后的 FP-RFBNet 结构如图 3 所示。



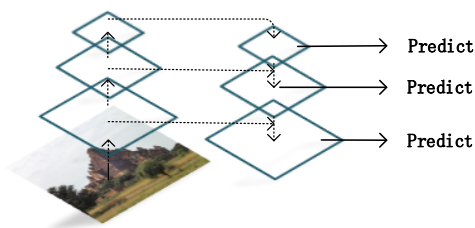


图2 FPN 网络  
Fig.2 FPN network

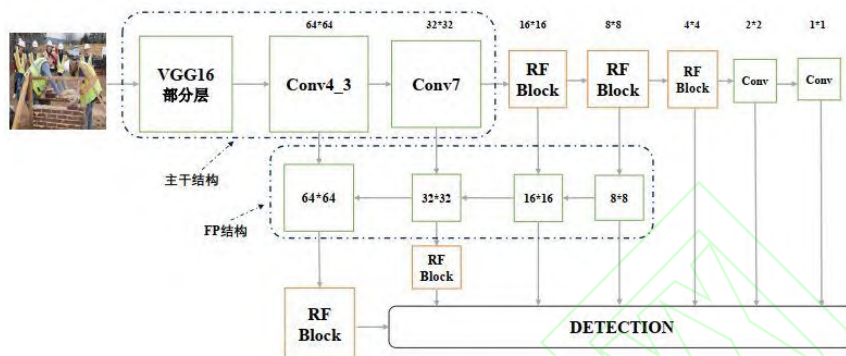


图3 FP-RFBNet 网络  
Fig.3 FP-RFBNet network

### 3.2 SE+Ghost 轻量化的SG+FP-RFB net

由于引入FPN模块的FP-RFBNet网络，增加了网络的结构复杂性，从而导致检测速度变慢，为了提高网络的检测速度，同时保持检测性能不变的情况下，提出一种SG（SE+Ghost）模块来对FP-RFBNet网络进行轻量化。

SE注意力模块通过学习自动获取每个特征通道的重要程度，然后根据该重要程度去提升有用的特征并抑制表达能力不强的特征。SE模块首先对特征进行压缩得到一个多维的全局感受野的特征，然后通过自动学习确定每一维度特征的权重，最后通过将学习到的特征逐通道加权到先前的特征上，实现对原始特征的重新标定<sup>[12]</sup>。在RFBnet网络中，卷积存在大量的浮点运算并产生大量的冗余特征图，而这些冗余的特征对于目标检测等任务是至关重要的<sup>[13][14]</sup>。Ghost模块由两个卷积块构成，首先通过较小卷积核生成少量的原始特征，然后对原始特征通过“廉价”的线性运算来产生大量的冗余特征，从而实现花费很小的计算代价达到普通

卷积实现的效果<sup>[15]</sup>。

这里，将SE模块和Ghost模块组合得到SG模块（SE-Ghost），SG模块不仅实现RFBnet网络中冗余特征图的提取，且提取的特征图体现特征通道重要程度，SG模块其结构如图4所示。

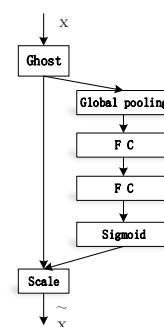


图4 SG 结构  
Fig.4 Structure of SG

在SG结构中，首先将Ghost模块输出的特征经过Squeeze全局平均池化得到压缩特征，然后将特征维度降低到输入的1/16，经过激活函数后通过两个全连接层FC将维度提升到原来的水平，接着通过Sigmoid激活函数将每个特征层的权重归一化到[0,1]之间，最后通过Scale操作来将归一化后的权重加权到每个通道的特征上，得到具有特

征通道重要程度信息的特征层。

为了减小网络的参数量与计算量,用构建的轻量高效的 SG 模块替换 FP-RFBnet 主干网络中的部分卷积模块,使其能够聚合不同感受野的特征和不同特征层之间的相关性,使得改进后的网络更好地实现对较小的目标的检测,在提取安全帽特征时更加高

效。另外,FP-RFBnet 在 Conv7 之后连接了 4 个 RF Block,由于深层特征分辨率较低,同时,RF Block 相较于普通卷积运算量较大,为了进一步轻量化网络,用普通卷积替换 FP-RFBnet 中最后两个 RF Block。改进后的 FP-RFBnet 网络命名为 SG-FP-RFBnet,其主干结构如图 5 所示。

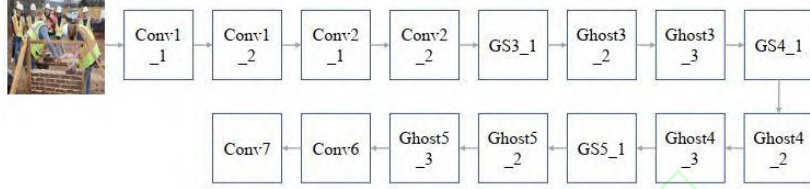


图 5 SG-FP-RFBnet 主干网络

Fig.5 backbone network of SG-FP-RFBnet

## 4 实验

### 4.1 实验环境

在服务器上搭建进行训练所需要的环境,服务器硬件配置为 CPU(Inter Xeon)、GPU(Nvidia TITAN Xp)、主板(超微 X10DRG-Q)、内存(镁光 256G)、硬盘(2TB);软件配置为 Windows10 专业版操作系统、CUDA、Anaconda、Pytorch 等。实验数据采用开源的安全帽 SCUT-HEAD 数据库(<https://github.com/njvisionpower/Safety-Helmet-Wearing-Dataset>),该数据集下有 6878 张图片,其中 6328 张作为训练集,550 张作为测试集。为了提高 SG-FP-RFBnet 的鲁棒性并防止模型过拟合,对训练样本图片进行随机裁剪和放大、亮度和镜像变换等操作。为了说明 SG-FP-RFBnet 网络的有效性,在相同的实验环境下采用相同的学习参数与数据集对 RFBnet-512、RFBnet-300、YOLO-v3 分别进

行训练,得到相对应的训练模型。

### 4.2 识别结果与分析

选取复杂施工环境的样本图片用上述网络模型进行测试,并对测试结果进行对比分析。图 6(a)给出了不同遮挡、干扰和大小的 4 幅代表性测试图像。图 6(b)、6(c)、6(d)、6(e) 分别是测试样本在 YOLO-v3、RFBnet-300、RFBnet-512 和 GS-FP-RFBnet 方法检测下得到的结果。从图中可以看出,RFBnet-300 的检测效果最差,在 4 幅图中都存在弱小安全帽的漏检,这是由于其提取图像特征信息的能力较弱造成的。YOLO-v3 网络尺度表示和特征提取能力有限,对图 1、图 2 中的弱小安全帽存在漏检。相比 YOLO-v3 与 RFBnet-300 网络,RFBnet-512 检测准确率有一定的提高,但在图 3 中出现了远处弱小安全帽的漏检。在所有算法中,本文的 GS-FP-RFBnet 实现了 4 图中所有安全帽的准确检测,取得最好的检测效果,说明了建立的 GS-FP-RFBnet 网络有效地实现了图像多尺度语义信息的表示与图像特征的提取,从而提高弱小目标的提取能力。



图 1

图 2

图 3

图 4

(a) 原始样本

(a) Original sample





(b) 图 1 识别结果

(b) Therecognition resultsofPicture 1



(c) 图 2 识别结果

(c) Therecognition resultsofPicture 2



(d) 图 3 识别结果

(d) Therecognition resultsofPicture 3



(e) 图 4 识别结果

(e) Therecognition resultsofPicture4

图 6 模型对比结果

Fig.6 Model test comparison results

(从左往右依次为 YOLO-v3、RFBnet-300、RFBnet-512、GS-FP-RFBnet 方法识别结果)

(From left to right are the recognition results of YOLO-v3, RFBnet-300, RFBnet-512, GS-FP-RFBnet)

为了从客观指标上说明 GS-FP-RFBnet 网络的性能,本文从检测精度、速度以及模型参数量等方面来验证改进的效果,利用 SCUT-HEAD 测试数据集分别对上文中的四种检测方法进行测试,识别率结果如图 7 和表 1 所示。图 7 给出了四种检测方法在 SCUT-HEAD 测试数据集上的平均准确率 (Average Precision, AP) 和平均准确率均值 (Mean Average Precision, MAP)。GS-FP-RFBnet 在安全帽 (hat) 和人脸 (person)检测中获得的 AP 与 MAP 值最高,

相较于 YOLO-v3 、 RFBnet-300 和 RFBnet-512, MAP 分别提高了 1.6、3.62 和 0.98,说明 GS-FP-RFBnet 网络中的特征金字塔网络模块可以有效地提高安全帽检测的精度。表 1 给出了四种检测方法的速度与参数量,可以看出 GS-FP-RFBnet 虽然每秒检测的帧数相较于 FRBnet-512 约有下降,但达到 20frame/s,具有一定的实时性。模型存储参数量得到了极大的降低,相较于 YOLO-v3、RFBnet-300 和 RFBnet-512 分别减少了 183.5M、77.5M 和 80.5M,可以节省

大量的存储空间。

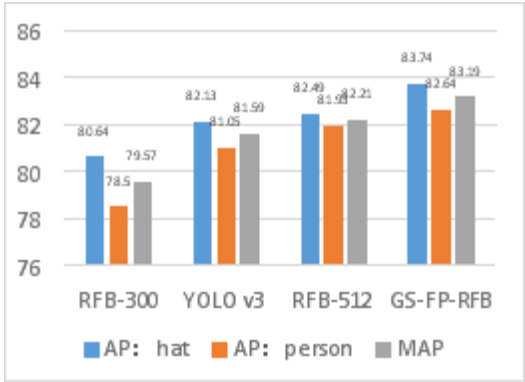


图 7 模型精度对比

Fig.7 Model accuracy comparison

表 1 参数对比结果

Table 1 Parameter comparison results

检测模型	测试样本数量	检测速度/(frame/s)	参数量/MB
YOLO-v3	550	32	235
RFBnet-300	550	27	129
RFBnet-512	550	22	132
<b>GS-FP-RFBnet</b>	<b>550</b>	<b>20</b>	<b>51.5</b>

综述，可见本文将特征金字塔网络模块 FPN 引入到 RFBnet 网络中，使神经网络融合浅层和深层语义信息，实现复杂施工环境下安全帽的精确识别。采用 SE-Ghost 模块，对网络的主干网络结构进行轻量化，大幅度降低网络的存储空间。建立的 GS-FP-RFBnet 实现高精度安全帽的识别，并满足实时性的要求。

## 5 结论

本文针对工业复杂环境下的安全帽佩戴检测问题，构建 FPN 结构来增强网络对于小目标检测的能力，采用 SG 模块对网络进行轻量化，从而建立轻量化高精度的安全帽佩戴检测网络 GS-FP-RFBnet。通过实验证明，该网络在复杂施工场景下，能有效地实现安全帽佩戴识别，极大地缩小网络的存储空间，运算速度满足实时要求，识别准确率相对其它网络有一定的提高。因此，本文提出的 GS-FP-RFBnet 网络具有较大的学术参考与应用价值。

## 参考文献：

- [1] PARK M W, ELSAFTY N, ZHU Z. Hardhat-wearing detection for enhancing on-site safety of construction workers[J]. Journal of Construction Engineering & Management, 2015, 141(9): 04015024.1-04015024.16.
- [2] 李琪瑞.基于人体识别的安全帽视频检测系统研究与实现[D].电子科技大学,2017.  
  
Li Qirui. A research and implementation of safety-helmet video detection system based on human body recognition[D]. School of Computer science & engineering,2017.
- [3] 刘晓慧,叶西宁.肤色检测和 Hu 矩在安全帽识别中的应用[J].华东理工大学学报(自然科学版),2014,40(3):365-370.  
  
Liu Xiaohui, Ye Xining. Skin color detection and hu moments in helmet recognition research[J]. Journal of East China University of Science and Technology(Natural Science Edition),2014,40(3):365-370.
- [4] 王生生,王萌,王光耀.基于深度神经网络剪枝的两阶段遥感图像目标检测[J].东北大学学报(自然科学版),2019,40(2):174-179.  
  
Wang Shengsheng, Wang Meng, Wang Guangyao.Deep neural network pruning based two-stage remote sensing image object detection[J]. Journal of Northeastern University (Natural Science), 2019, 40(2): 174-179.
- [5] 刘俊明, 孟卫华.基于深度学习的单阶段目标检测算法研究综述[J/OL].航空兵器: 1-11 [2020-06-10].<http://ffiy208e51c2dd88406685526280e50de659hw9pxp65vbuv66pqx.ffffb.suse.cwkeji.cn:999/kcms/detail/41.1228.TJ.20191124.1632.007.html>.



- Liu Junming, Meng Weihua. Research on single-stage object detection algorithm based on deep learning[J/OL]. Aero Weaponry: 1-11 [2020-06-10].<http://ffiy208e51c2dd88406685526280e50de659hw9pxp65vbuv66pqx.fffb.suse.cwkeji.cn:999/kcms/detail/41.1228.TJ.20191124.1632.007.html>.
- [6] 黄泽桑. 基于深度学习的目标检测技术研究[D].北京邮电大学,2019.
- Huang Zesang. Object detection based on deep learning[D]. Beijing University of posts and telecommunications, 2019.
- [7] LIU S,HUANG D,WANG Y. Receptive field block net for accurateand fast object detection[C]. Computer Vision-ECCV, 2018: 404-419.
- [8] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 89-95.
- [9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. in International Conference on Learning Representations, May 2015.
- [10] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional netsand fully connected CRFs[J]. Computer Science, 2014(4):357-361.
- [11] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu:IEEE, 2017: 2117-2125.
- [12] HU J, SHEN L, AIBANIE S, et al. Squeeze-and-excitation networks[J].2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 7132-7141, doi: 10.1109/CVPR.2018.00745.
- [13] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[J]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea(South), 2019: 1314-1324.
- [14] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices[J].2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018: 6848-6856.
- [15] HAN K, WANG Y, TIAN Q, et al,Ghostnet: more features from cheap operations[J]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020: 1577-1586.