

基于深度学习的水下光学图像分类识别方法研究*

汤永涛¹ 林鸿生¹ 刘 兴²

(1. 海军士官学校 蚌埠 233012)(2. 南京航空航天大学 南京 210016)

摘 要 水下光学探测在沿海港口的防御系统中具有不可忽视的作用。针对水下光学图像的目标分类问题,论文搜集了水下潜水员、水中生物与水下机器人等数据,利用深度学习方法对水下场景小数据集的目标分类进行研究,采用迁移学习和轻量级网络设计两种方法对数据集进行训练及测试,实验表明,大型网络的迁移学习可以得到接近 100% 的分类准确率,但是网络的参数量巨大,不适于水下场景的部署;论文设计的轻量级网络也可以得到 98.67% 的准确率,且网络参数量很小,在水下场景的部署中具有很大的优势。

关键词 水下光学图像;分类识别;深度学习;迁移学习;轻量级网络

中图分类号 TP39 **DOI:** 10. 3969/j. issn. 1672-9730. 2022. 08. 040

Research on Underwater Optical Image Classification and Recognition Based on Deep Learning

TANG Yongtao¹ LIN Hongsheng¹ LIU Xing²

(1. Naval Petty Officer School, Bengbu 233012)

(2. Nanjing University of Aeronautics and Astronautics, Nanjing 210016)

Abstract Underwater optical detection plays an important role in the defense system of coastal ports. To solve the problem of classification for underwater optical image, the data of underwater divers, underwater organisms and underwater robots are collected from the internet, then deep learning method is used to study the classification problem of small data sets. Transfer learning and lightweight network design are used to train and test the data set. The results show that, accuracy of large-scale networks can be close to 100% in transfer learning. But they are not suitable for underwater scene deployment because the network parameters are huge. The results also show that, accuracy of our lightweight network can reach 98.67%, and it has great advantages in underwater scene deployment owing to its small parameters.

Key Words underwater optical image, classification, deep learning, transfer learning, lightweight network

Class Number TP39

1 引言

我国沿海分布有大量港口,在国际形势风云变化的情况下,港口的防御压力巨大。敌对势力可利用水下蛙人、机器人以及小型水下运载器等从水下趁虚而入,对重要设施进行破坏^[1]。目前水下安全监视主要利用水声技术对水下目标进行探测、分类、定位和跟踪。但声呐系统易受水下环境及目标属性等条件制约,很难获得准确、直观的目标识别结果。相对于水声而言,水下光学探测具有成像分辨率高、细节丰富、成像速度快等优点。因此,水下

光学探测逐渐成为水下探测的一种有效辅助手段^[1-2]。

但是,由于水和水体中的杂质对光信号存在着吸收、散射等效应,水下光学图像普遍存在着对比度低、噪声突出等问题^[3-4]。这给水下目标的自动分类识别带来困难。

深度学习在图像分类识别方面表现非常出色。自 2012 年 Krizhevsky 等提出 AlexNet^[5]以来,学者们开发出了多种用于分类识别的神经网络结构,如 VGGNet^[6]、GoogleNet^[7]、ResNet^[8]、ResNeXt^[9]、NASNet^[10]、EfficientNet^[11]等。这些网络结构在

* 收稿日期:2022 年 2 月 15 日,修回日期:2022 年 3 月 27 日

作者简介:汤永涛,男,副教授,研究方向:电子对抗和水下图像分类识别。林鸿生,男,博士研究生,讲师,研究方向:电子对抗和水下图像分类识别。刘兴,男,博士研究生,讲师,研究方向:人工智能。

ImageNet 数据库上能得到很不错的准确率(各种经典网络在 ImageNet 数据库测试集上的准确率见图 1(a))。但这些分类网络在实际应用中存在两个问题:一是用于训练小型数据集时容易出现过拟合,因为它们都是针对大型数据库进行设计的,网络结构复杂,参数量巨大(见图 1(b));二是较难应用于移动设备,移动设备硬件性能受到限制,无法与服务器等大型硬件相比,这些大型网络由于模型巨大,推理时间较长,在移动设备上应用会受到限制,实时性不能满足要求。

在偏向于应用到移动设备的网络中,也出现了像 MobileNet^[12~14]、ShuffleNet^[15]、GhostNet^[16]等高效的轻量级网络。但它们的分类准确率与大型网络相比,仍有不小的差距。图 1 为各种网络的准确率与参数量比较,以白框表示的为 MobileNet、ShuffleNet、GhostNet。

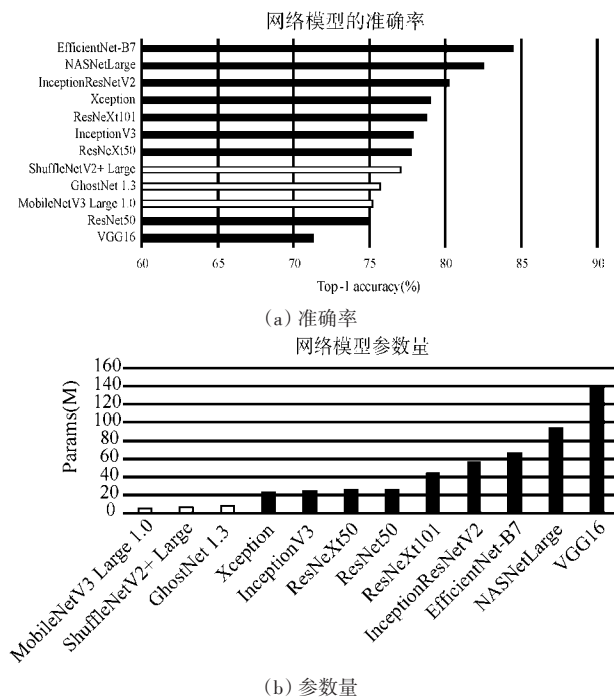


图 1 经典网络的性能对比

本文在上述背景的基础上,针对水下光学图像小数据集目标分类的问题,利用深度学习方法进行研究,从迁移学习与轻量化网络设计两个方面进行。

2 理论基础

2.1 迁移学习

随着各种网络架构越来越成熟,迁移学习的重要性在逐渐提高。大型网络在大型数据集上的训练对于普通研究人员来说很难实现(硬件设备,训练时间很难得到保证),而越来越多的企业在网络

上分享了他们在大型数据集上训练过的网络模型,这为普通研究人员节省了大量训练时间。

迁移学习定义为:给定一个源域 D_s 和一个学习任务 T_s ,一个目标域 D_t 和一个学习任务 T_t ,迁移学习的目的是使用在 D_s 和 T_s 上的知识帮助提高在目标域 D_t 上的预测函数 $f_T(x)$ 的学习^[17]。迁移学习的目的是利用现有的训练过的网络模型,改进目标域的学习效果。本文利用迁移学习训练的过程如图所示。首先对预训练模型进行随机初始化后,利用小数据集进行训练,训练稳定后,取出网络模型进行推理,最后利用测试集对其性能进行测试。

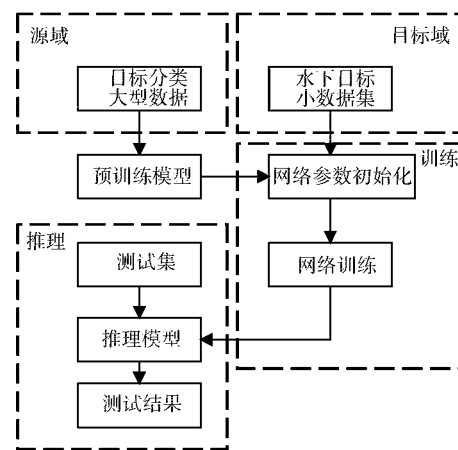


图 2 迁移学习示意图

2.2 深度可分离卷积

单个卷积核具有两个空间维度(宽度和高度)和一个通道维度,在卷积过程中需要同时映射跨通道相关性和空间相关性。在卷积核数量较多的情况下,计算量巨大。

深度可分离卷积(Depthwise separable convolution)^[18]把映射空间相关性和映射通道间相关性的任务分离,不同的通道独立起来,先进行特征提取,再进行特征融合,这种方式充分利用模型参数进行特征学习,大幅降低计算量,从而让网络效率更高。

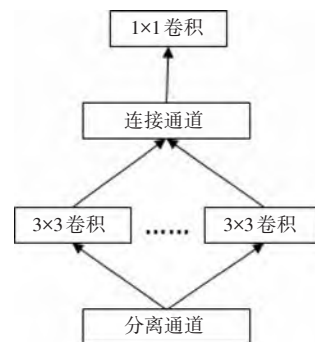


图 3 深度可分离卷积示意图

典型的深度可分离卷积如图 3 所示。对于 n 个通道的卷积操作,首先对每一个通道独立进行空间

卷积,然后使用concat操作把各个通道连接在一起,最后进行逐点卷积(1×1卷积)实现跨通道融合。

3 建立数据集

本文分类目标为潜水员、机器人及小型水下运载器、水下生物等三种。主要搜集方式为因特网。考虑到水下目标有近有远,搜集到的图像分辨率有大有小,最后水下生物图片为505张(分辨率从40×38到743×497,分辨率为100×100以下的有176张),潜水员图片为535张(分辨率从44×52到867×549,分辨率为100×100以下的有110张),水下运载器图片为337张(分辨率从53×68到890×629,分辨率为100以下的有20张)。

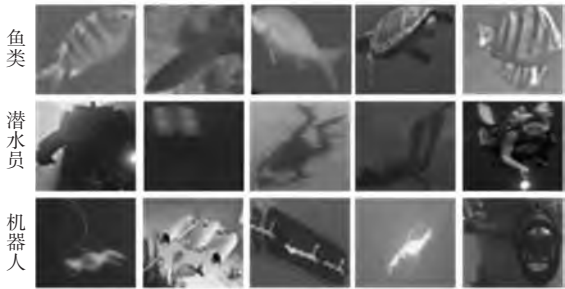


图4 本文数据样例

最后,采用旋转、镜像等方式对数据进行扩增,作为训练和测试的数据集(水下生物为1002张,潜水员为1005张,水下机器人为996张)。

4 网络结构设计

本文场景为水下,数据较难收集,属于小数据集的应用场景。针对小数据集的分类模型设计,主要有两种方法:一种是迁移学习,另外一种是设计轻量化的模型,重新训练。

4.1 迁移学习设计

大型数据库的分类识别模型训练,对硬件要求较高,而且常常需要训练较长时间。迁移学习的好处在于节省训练时间,同时对训练的硬件设备要求较低,而且还能得到比较好的预测结果。

ImageNet是一个大型数据库,各个经典网络在ImageNet测试数据集的准确率以及网络的计算复杂度如图5所示。其中FLOPs(floating point operations),表示浮点运算次数,代表模型的计算复杂度。

从图中可以看到,计算复杂度较小,且准确率较高的网络模型有很多。本文选取计算复杂度在6GFLOPs以下,且准确率达到77%以上的网络进行

迁移学习。

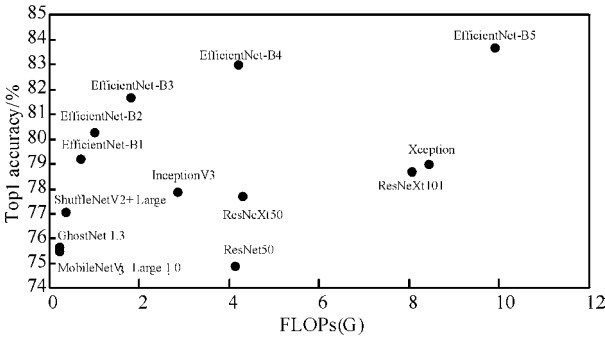


图5 经典网络的准确率和复杂度

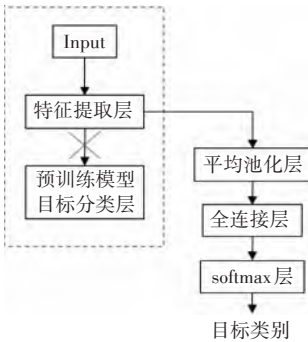


图6 迁移学习设计

用于图像分类的卷积网络架构包括了两大部分:特征提取层(它们包含一系列卷积和池化层)和分类器(一般是密集连接的 fully connected layer)。由于本文分类器与预训练模型分类器不同,所以进行迁移学习时,对预训练模型分类层进行改变,具体如图6所示。其中全连接层分为2层,分别有256个节点和3个输出节点(分类目标为3类)。送入网络进行训练的图像统一设置为128×128(宽度×高度,单位为像素)大小的RGB图像,且经过归一化处理。

4.2 轻量化模型设计

深度学习模型在水下目标的分类识别中的应用偏向于移动设备的部署,这就要求网络模型要较小,计算速度要快。

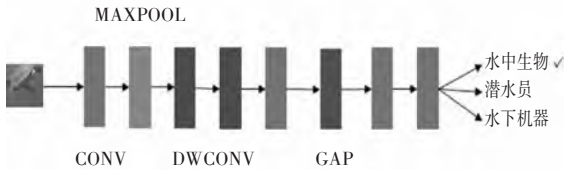


图7 本文网络结构

本文设计的网络结构如图7所示。主要由卷积层(CONV)、池化层(MAXPOOL)和深度可分离卷积模块(DWCONV)组成。具体卷积核数量如表1所示。

表1 ShuffleNetV2的测试结果

名称	卷积核尺寸	步长	输出尺寸	BN	RELU
Input	-	-	128×128×3	-	-
Conv2D	64×7×7	2	64×64×64	P	P
MaxPool	64×3×3	2	32×32×64	-	-
DWConv	128×3×3	1	16×16×128	-	P
DWConv	256×3×3	1	8×8×256	-	P
Conv2D	960×1×1	1	1×1×960	P	P
GAPool	-	-	960	-	-
Conv2D	480×1×1	1	1×1×480	P	P
Conv2D	3×1×1	1	1×1×3	P	-
Reshape	-	-	3	-	-

考虑到分辨率较小的图片比较多,所以输入选择统一缩放到128×128大小。第一个卷积层用于提取图片的特征,选择7×7的大卷积核,感受野比较大,特征提取效果较好。DWCONV模块结构如图8所示,分为2个通道,分别提取不同细节特征。最后,用最大平均池化(GAP)和1×1卷积核代替分类层的全连接,减少计算量。

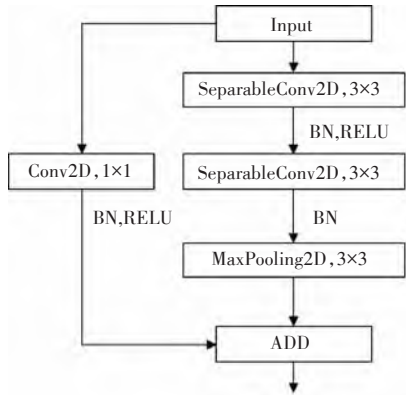


图8 DWCONV模块结构

5 实验结果

本文各种算法基于keras和pytorch框架实现,在Ubuntu中用Python进行实验。训练平台采用CPU为intel i7 7700K,显卡为nvidia gtx 1080,内存为16G的台式机。训练过程中采用GPU加速。

5.1 迁移训练

迁移训练时,对各种网络在ImageNet数据集下训练得到的模型进行初始化;优化算法采用可以自适应的Adam算法;初始学习率设置为0.0001;损失函数采用交叉损失熵;Batchsize设置为32;epochs设置为20;每个epoch迭代200次;每次epoch开始的时候,对训练数据进行打乱。

其中两个典型的网络EfficientNet-B1和InceptionV3的训练曲线如图9所示,其中方格点标示的

曲线为准确率曲线,圆点标示的曲线为损失曲线。红色线为测试集的曲线,蓝色的为训练集的曲线。

从中可以看到,EfficientNet系列的测试集曲线比较平稳,InceptionV3的测试集曲线有点波动。说明EfficientNet系列的预训练模型的泛化能力较强。

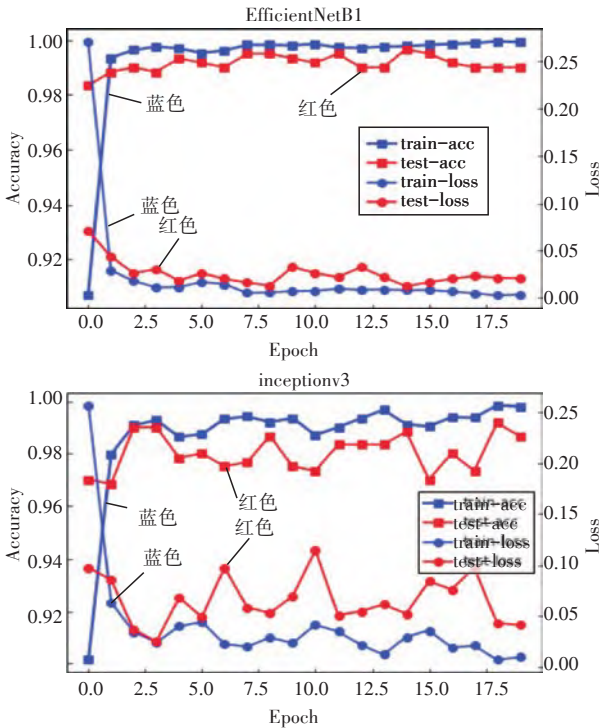


图9 迁移学习的训练曲线

训练结束后,对测试集进行测试,其中EfficientNet-B3、ShuffleNetV2和InceptionV3网络的测试结果如下所示。可以看到,EfficientNet-B3准确率非常高,601个图像,推理结果只错了1个。InceptionV3推理能力也非常不错,ShuffleNetV2相对来说推理能力弱了一点。

表2 InceptionV3的测试结果

预测值	实际值		
	0	1	2
InceptionV3			
0	204	1	0
1	0	201	0
2	2	2	191
EfficientNet-B3			
0	205	0	0
1	0	201	0
2	1	0	194
ShuffleNetV2			
0	199	4	1
1	5	190	6
2	6	8	182

各个网络的参数量、计算复杂度及测试集的准

确率如表所示。可以看到,参数量较大的网络都取得很不错的准确率。参数量较小的网络 Shuffle-NetV2 与其他网络相比,仍有差距。

表3 InceptionV3的测试结果

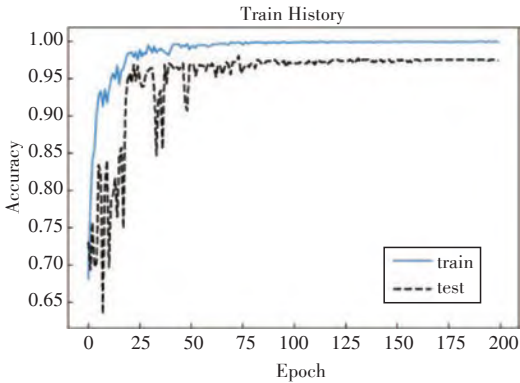
Model	Params(M)	FLOPs (M)	acc
EfficientNet-B1	6.90	169.85	0.9967
EfficientNet-B2	8.13	197.89	0.9950
EfficientNet-B3	11.18	292.08	0.9983
EfficientNet-B4	18.13	459.55	0.9950
InceptionV3	22.33	1525.06	0.9917
ResNext50	23.57	2621.34	0.9933
ShuffleNetV2	1.26	96.56	0.98003

总的来说,迁移学习省时省力,而且能得到很不错的结果。但是同时也受预训练模型的限制,网络模型固定,在部署到移动设备时比较不方便。所以重新设计网络模型,进行自我训练就显得尤为重要。

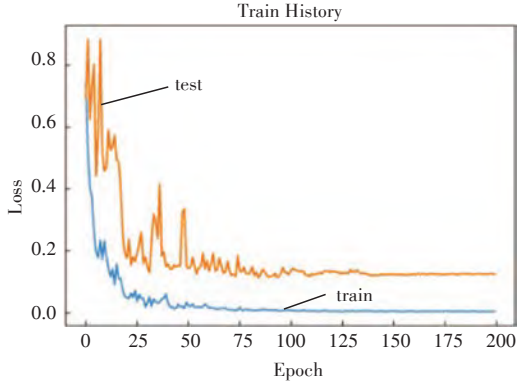
5.2 轻量化模型训练

本文网络进行训练时,优化算法采用可以自适应的 Adam 算法;初始学习率设置为 0.0008,然后采用可变学习率,每 20 个 epoch 学习率减小一半;损失函数采用交叉损失熵;Batchsize 设置为 32;epoch 数量设置为 200;每个 epoch 迭代 300 次;每次 epoch 开始的时候,对训练数据进行打乱。

训练曲线如图 10 所示。



(a) 准确率曲线



(b) 损失曲线

图 10 训练曲线

从图 10 中可以看到,在迭代训练 100 次以后,训练的准确率接近 100%,损失也接近 0。同样的,在迭代训练 100 次以后,测试集的准确率接近 99%,损失趋近 0.14。

训练结束后,对测试集进行测试,各个网络的测试结果如表 4 所示。

表4 Varysimplenet的测试结果

预测值	实际值		
	0	1	2
0	204	0	1
1	1	200	0
2	3	3	189

本文方法得到的混淆矩阵见表 2。分类错误数量为 8 个,正确数量为 593。准确率达到 98.67%。

计算复杂度方面,本文网络参数数量为 0.899M,

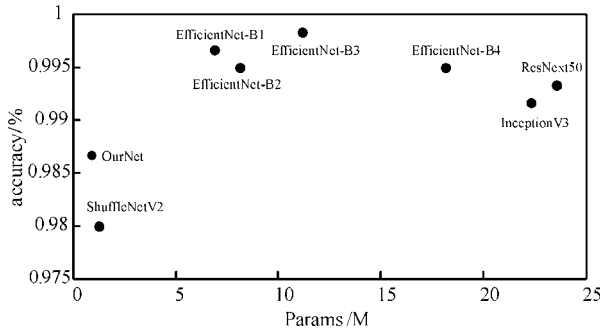


图 11 与迁移学习的性能比较

本文网络模型的性能与其他网络模型的性能比较见图 11。从中可以看到,本文模型参数量较少,而且取得较不错的测试结果,在移动设备部署中具有很不错的优势。

6 结语

本文研究了水下光学图像的分类识别方法。从迁移学习与自我设计的网络两个方面出发,都取得了不错的目标分类结果。总的来说,迁移学习省时省力,而且能得到很不错的结果。但是受预训练模型的限制,在部署到移动设备时比较不方便。轻量化设计的网络比较灵活,可以控制参数量,而且在小数据集的训练方面,小网络适应性更好,泛化能力也比较强。

参考文献

[1] 张寅权, 张爽, 孙春健, 等. 水下目标监视系统发展综述[J]. 海洋信息, 2019, 34(1): 11-18.
[2] HOU W. Ocean Sensing and Monitoring: Optics and Other Methods[M]. Washington: SPIE Press, 2013: 87-116.

- [3] 张清博, 张晓晖, 韩宏伟. 一种基于深度卷积神经网络的水下光电图像质量优化方法[J]. 光学学报, 2018, 38(11): 96-104.
- [4] 白继嵩. 水下图像处理及目标分类关键技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2017: 9-11.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25(2): 1097-1105.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]// International Conference on Learning Representations. San Diego: IEEE, 2015: 1-14.
- [7] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 1-9.
- [8] WANG T, WU D, COATES A, et al. End-to-end text recognition with convolutional neural networks [C]// International Conference on Pattern Recognition, Tsukuba: IEEE, 2012: 3304-3308.
- [9] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks [C]// 2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition, USA: IEEE, 2017: 5987-5995.
- [10] Zoph B, Vasudevan V, Shlens J, et al. Learning transferable architectures for scalable image recognition [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, USA: IEEE, 2018: 8697-8710.
- [11] TAN M, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks [EB/OL]. <https://arxiv.org/abs/1905.11946.pdf>, 2020.
- [12] Howard A G, ZHU Menglong, CHEN Bo, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. <https://arxiv.org/abs/1704.04861>, 2017.
- [13] M. Sandler, A. Howard, M. Zhu, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, USA: IEEE, 2018: 4510-4520.
- [14] Andrew Howard, Grace Chu, et al. Searching for MobileNetV3 [C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Korea: IEEE, 2019: 1314-1324.
- [15] X. Zhang, X. Zhou, M. Lin, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, USA: IEEE, 2018: 6848-6856.
- [16] K. Han, Y. Wang, Q. Tian, et al. GhostNet: More Features from Cheap Operations [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA: IEEE, 2020: 1577-1586.
- [17] Pan S J, Yang Q. A Survey on Transfer Learning [J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 22(10): 1345-1359.
- [18] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning [M]. USA: The MIT Press, 2016: 362-363.

(上接第84页)

的实现[J]. 电讯技术, 2009, 49(8): 59-64.

- [11] 杨冰. 机载任务多链路资源分配研究[D]. 成都: 电子科技大学, 2020.
- [12] R. A. J. Castillo, R. Grünheid, G. Bauch, et al. Communication Analysis Between an Airborne Mobile User and a Terrestrial Mobile Network [J]. IEEE Transactions on

Vehicular Technology, 2018, 67(4): 3457-3465.

- [13] A C Nearchou. A differential evolution approach for the common due date early/tardy job scheduling problem [J]. Computers & Operations Research, 2008, 35(4): 1329-1343.
- [14] 鸥鑫, 龚远强. 一体化系统资源调度算法研究与仿真 [J]. 电子信息对抗技术, 2020(6): 53-56.