

# 面向陆战场目标识别的轻量级卷积神经网络

乔梦雨 王鹏 吴娇 张宽

西安工业大学电子信息工程学院 西安 710021

(273021898@qq.com)



**摘要** 在实际陆战场环境中,作战人员无法随身携带 GPU 等大型计算设备,因此较难计算规模较大的神经网络参数,进而导致目标识别网络无法实时工作。现有的轻量级神经网络虽然解决了实时性的问题,但是不能满足准确率的要求。为此,文中提出了一种基于轻量级卷积神经网络的目标识别算法(E-MobilNet)。为了提升网络学习的效果,以现有深度学习的主要目标检测框架 MobileNet-V2 为基础,插入一种 ELU 函数作为激活函数。首先,使用扩张卷积来增加通道数,以获得更多的特征;接着,通过 ELU 函数激活输出特征,这样可以缓解线性部分的梯度消失,并且使非线性部分对输入变化的噪声更鲁棒;然后,通过残差连接的方式组合高层特征与低层特征的输出;最后,将全局池化的输出结果输入 Softmax 分类函数。实验数据表明,在同样的测试集和测试环境下,与现在主流的轻量级深度学习目标识别算法相比,E-MobileNet 识别的准确率和每秒检测的帧率都有所提升。实验数据充分说明,使用 ELU 激活函数和全局池化层减少了参数的数量,增强了模型的泛化能力,提升了算法的鲁棒性,在保证神经网络模型轻量级的基础上有效地提高了目标的识别准确率。

**关键词:** 目标识别;轻量级模型;可分离卷积;残差网络;激活函数

**中图分类号** TP391.9

## Lightweight Convolutional Neural Networks for Land Battle Target Recognition

QIAO Meng-yu, WANG Peng, WU Jiao and ZHANG Kuan

School of Electronic Information Engineering, Xi'an Technological University, Xi'an 710021, China

**Abstract** In an actual land battle environment, people cannot carry large computing devices such as GPUs with them. Therefore, it is more difficult to calculate the large-scale neural network parameters, which further leads to the target recognition network not working in real time. To this end, a target recognition algorithm based on lightweight convolutional neural network (E-MobilNet) is proposed. In order to improve the network learning effect, based on the existing target learning framework MobileNet-V2, an ELU function is inserted as an activation function. Firstly, use the expansion convolution to increase the number of channels to get more features to activate and output through the ELU function, which can alleviate the disappearance of the gradient of the linear part, the nonlinear part is more robust to the noise of the input change. Then, the way of the residual connection combine high-level features with low-level features and then output. Finally, output to Softmax using global pooling. The experimental data shows that compared with the current mainstream lightweight deep learning target recognition algorithm, E-MobileNet has improved the accuracy of recognition and the frame rate per second in the same test environment of the same test set. The experimental data fully demonstrates that the use of the ELU activation function and the global pooling layer reduces the number of parameters, enhances the generalization ability of the model, and improves the robustness of the algorithm. On the basis of ensuring the lightweight of the neural network model, the recognition accuracy of the target is effectively improved.

**Keywords** Target recognition, Lightweight model, Separable convolution, Residual network, Activation function

## 1 引言

目标识别是计算机视觉领域中的一个重要分支<sup>[1]</sup>,其主要研究内容是将一个特殊目标(或一种类型的目标)从其他目标(或其他类型的目标)中区分出来<sup>[2]</sup>。随着科学技术的不断

发展,各个国家在追求和平的同时也保持着高度的军事前瞻性<sup>[3]</sup>。将目标识别技术应用于军事领域,从不同时机、不同分辨率的海量图像、视频大数据中获取位置信息并识别出所需要的军事目标(人、车辆、坦克、武器)类别,从而为指挥员做出决策提供情报支持,显得尤为重要<sup>[4]</sup>。实际情况下,军事目标处

到稿日期:2019-03-15 返修日期:2019-07-09 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61671362);陕西省科技厅重点研发计划(2019GY-022);西安市未央区科技计划项目(2019Z1);陕西省组合与智能导航重点实验室开放基金(SKLIIN-20180201)

This work was supported by the National Natural Science Foundation(61671362), Shaanxi Provincial Science and Technology Department Key R & D Program(2019GY-022), Xi'an Weiyang District Science and Technology Plan Project(2019Z1) and Shaanxi Provincial Combination and Intelligent Navigation Key Laboratory Open Fund(SKLIIN-20180201).

通信作者:王鹏(371723761@qq.com)

于运动状态会存在部分遮挡或背景复杂的情况,从而影响到识别的准确率<sup>[5]</sup>。同时,由于深度学习技术的落地对终端硬件设备的要求较高,只有轻量级的网络才能使应用场景更广泛<sup>[6]</sup>。

近年来,随着计算机硬件的发展,计算机的算力大大提升,基于深度学习的目标识别算法发展迅速。卷积神经网络是一种深度学习技术,具有很好的鲁棒性,非常适合学习图像或视频中的目标的特征。2012年,Hinton等首次参加ImageNet<sup>[7]</sup>图像识别大赛,使用卷积神经网络技术构建了AlexNet<sup>[8]</sup>并获得了第一,其识别率大大超过了传统的目标识别方法(SVM<sup>[9]</sup>)。以此为开端,在目标识别领域越来越多的学者开始研究并使用卷积神经网络技术<sup>[10]</sup>。2016年,Liu等提出了一种端到端的识别算法SSD<sup>[11]</sup>,其提取不同大小的特征图来进行识别,用浅层特征图和深层特征图来分别识别小物体和大物体。2017年,Zhang等提出基于通道混洗的ShuffleNet<sup>[12]</sup>识别算法,通过通道混洗的方法将每个通道的卷积特征进行交互,然后组合输出,同时修改原来的残差网络单元,减少了深度网络的计算量。2018年,Redmon等提出YOLO v3<sup>[13]</sup>识别算法,模型的特征提取器是一个残差模型,由于使用了残差单元,因此模型构建得更深。2018年,Sandler等提出了MobileNet v2<sup>[14]</sup>识别算法,其通过堆叠深度可分离卷积模块与倒置残差结构,在减少模型参数的同时提升了模型的计算速度。

综上所述,目前已有的大部分目标识别算法都是在图形处理器(Graphics Processing Unit, GPU)加速的情况下进行识别的。使用GPU可以构建很深的网络,并提升网络的识

别准确率,但是实时性较低<sup>[15]</sup>。构建浅层的卷积神经网络可以在不使用GPU加速的情况下完成实时监测,但准确率较低<sup>[16]</sup>。针对以上问题,本文在传统的目标检测算法MobileNex v2的基础上,提出了一种基于轻量级的卷积神经网络的目标识别算法。

本文的主要贡献如下:1)使用深度可分离卷积模块,将原来的基本卷积模块按照通道和位置分开进行计算,以减少计算量。2)添加了残差支路,使模型能够有效地学习浅层的特征;组合全局池化层的残差支路与可分离卷积层的特征,提高了模型的鲁棒性;同时,使用ELU激活函数代替传统的ReLU激活函数,先升维再降维,缓解了特征退化。3)最后通过全局池化输出特征,减少了模型的参数,防止了过拟合,在保证检测速度的同时提升了准确率。

## 2 E-MobileNet

本文提出的卷积神经网络模型E-MobileNet如图1所示。首先,输入图像经过两层卷积层;然后,通过设计的残差块(此处5个残差块的结构相同)学习特征;最后,使用softmax函数进行分类。模型的设计如下:1)添加一个AVG Pool层,用于在残差块对元素进行相加,即使用残差连接的方式复用了浅层特征;2)可分离卷积部分采用扩张再压缩的方法,将通道信息与位置信息解耦,大大减小了计算量,缓解了特征退化的情况;3)使用ELU函数作为激活函数,避免了梯度消失的情况,在加快收敛速度的同时提高了准确率;4)使用全局池化的方式将特征映射输出至下一层。

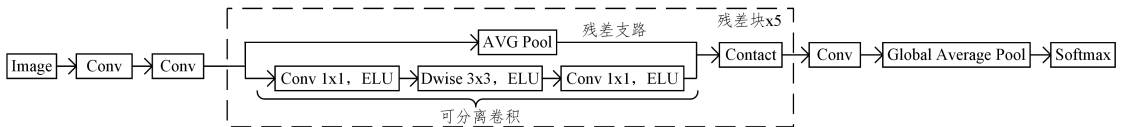


图1 E-MobileNet 的结构

Fig. 1 Structure of E-MobileNet

### 2.1 残差网络

当网络层数变深时,理想状况下网络可以学习到更多的知识,拟合更复杂的函数,从而提高识别的准确率<sup>[17]</sup>。但是,在反向传播过程中,随着层数的变深,多次计算偏导数容易出现梯度弥散或者梯度爆炸,进而导致网络学习不到任何东西。He等<sup>[18]</sup>发现通过恒等映射的方式可以避免上述问题,加入了残差连接后,直接跨越多层将前几层的输入通过恒等映射的方式直接输出。假设函数 $H(x)$ 为几个卷积层要拟合的最终结果, $x$ 表示这些卷积层中第一层的输入。多个非线性函数可以近似地逼近一个复杂函数,等价于残差函数 $H(x)-x$ 。因此,定义残差块为:

$$H(x) = F(x, W_i) + x \quad (1)$$

其中, $x$ 和 $H(x)$ 是残差块的输入量和要拟合的量;函数 $F(x)$ 是要学习的残差映射,可以为任意函数。

标准残差块的结构如图2所示。

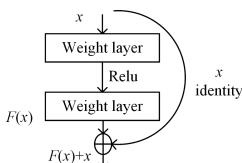


图2 标准残差结构

Fig. 2 Standard residual structure

### 2.2 可分离卷积

可分离卷积为一个一维通道卷积核和一个二维位置卷积核,分别学习图像中的通道信息和位置信息,如图3所示。

使用可分离卷积替代传统卷积模块的主要目的是将空间互相关信息与通道互相关信息分离,在加快计算的同时提高识别率<sup>[19]</sup>。首先,采用深度卷积(大小为 $D \times D \times 1$ )对不同输入通道分别进行卷积操作;然后,采用点卷积(大小为 $1 \times 1 \times C$ )将深度输出后再进行结合。

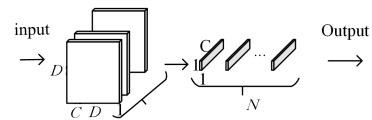


图3 可分离卷积

Fig. 3 Separable convolution

标准卷积结构如图4所示,大小为 $D \times D \times C$ 。分别使用大小为 $D \times D \times 1$ 和 $1 \times 1 \times C$ 的滤波器来减小计算量。具体实现过程:假定输入特征的大小为 $D \times D \times N$ ,输出特征的大小为 $D \times D \times C$ ,其中 $D$ 为图的长和高(假设是相同的), $C$ 和 $N$ 是通道数。采用一个 $D_k \times D_k$ 的标准卷积的计算量为:

$$D_k \times D_k \times M \times C \times D \times D \quad (2)$$

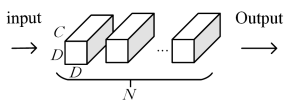


图4 标准卷积

Fig. 4 Standard convolution

而对于可分离卷积来说,首先对  $C$  个输入通道施加大小为  $D \times D$  的滤波器,即  $D \times D \times C \times D_k \times D_k$ ,然后应用  $N$  个大小为  $1 \times 1 \times C$  的卷积滤波器将  $C$  个输入信道组合成  $N$  个输出信道,即  $C \times N \times D_k \times D_k$ 。将大小为  $1 \times 1 \times C$  的特征图中的每个值合并在一起,计算量为:

$$D \times D \times M \times D_k \times D_k + M \times C \times D_k \times D_k \quad (3)$$

这样,相对于标准的卷积结构,可分离卷积结构的计算量得以减少。

$$\frac{D \times D \times M \times D_k \times D_k + M \times C \times D_k \times D_k}{D \times D \times M \times C \times D_k \times D_k} = \frac{1}{C} + \frac{1}{D^2} \quad (4)$$

### 2.3 指数线性单元

由于传统的激活函数 ReLU<sup>[20]</sup> 的输出值没有负值,因此当激活值的均值非 0 时,就产生了一个很小的误差,如果激活值之间不会相互抵消,则会导致下一层的激活单元存在偏差量。如此叠加,激活单元越多,偏差量就越大<sup>[21]</sup>。

本文拟采用指数线性单元(Exponential Linear Units, ELU)来代替传统的 ReLU 激活函数。ELU 函数的表达式如下:

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha(e^x - 1), & \text{if } x < 0 \end{cases} \quad (5)$$

$$f'(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ f(x) + \alpha, & \text{if } x < 0 \end{cases}$$

ELU 函数在 ReLU 函数上进行了改进,如图 5 所示。

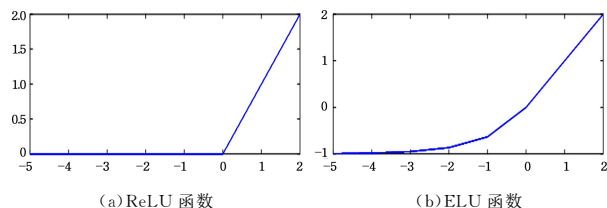


图5 ELU 对 ReLU 的改进函数

Fig. 5 Function ReLU based improvement of function ELU

当参数大于或等于 0 时,计算复杂度低,不需要进行指数运算,学习速度更快,同时增加了模型的非线性特征;当参数小于 0 时,使用一段平滑的函数,这样在函数接近于 0 时也可以输出梯度,达到了 BN 层的效果,而且减小了计算量,在对输入变化的噪声更加鲁棒的同时具有低复杂度。

### 2.4 全局平均池化

在目标识别工作中,全连接层一直是 CNN 分类网络的标配结构<sup>[22]</sup>。在分类工作中,对最后一个卷积层的输出进行矢量化操作,然后将结果输入全连接层,最后降低这个矢量计算的维度,并将其输入 softmax 层中计算分类得分。但是,全连接层的参数数量在整个神经网络中占很大一部分,且很容易过拟合,会影响网络的泛化能力<sup>[23]</sup>。

本文提出使用全局平均池化(Global Average Pooling, GAP)的策略来替代卷积神经网络中全连接层的工作。其思想是为最后一个卷积层输出的特征映射分别生成一个特征映

射,即不需要进行全连接的工作,而是取每个特征映射的平均值输出至 Softmax 层,如图 6 所示。

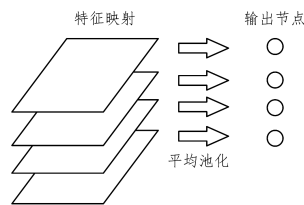


图6 全局平均池化

Fig. 6 Global average pooling

## 3 实验与分析

### 3.1 实验环境

本文在 Ubuntu 16.04 操作系统,深度学习框架 TensorFlow 的基础上,实现了基于轻量级卷积神经网络的陆战场目标识别算法。

为进一步体现算法的准确率及实时率,选取了 MobileNet, ShuffleNet, YOLO, SSD 算法进行对比,每种算法均是在相同的环境下运行。

### 3.2 实验数据集

根据陆战场中的实际情况,本文需要识别的目标种类包含坦克(Tank)、坦克炮塔(Cannon)、装甲车辆(Car)、人(Person)、枪(Gun)等目标种类。

由于目标种类特殊,数据资源较少,因此数据集一部分来自 Imagenet 数据集,还有一部分来源于百度图片中自行下载的图片。

数据集经过裁剪翻转、增加噪点等方法扩充后,每类有 5000 张,总计 25000 张图像。按照 PASCAL VOC 数据集<sup>[24]</sup>的格式进行了人工标注,数据集的部分图片如图 7 所示。

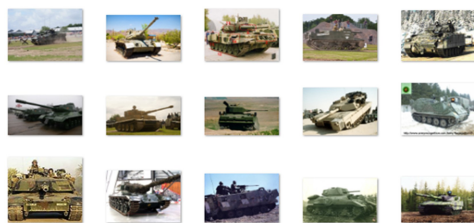


图7 部分数据集

Fig. 7 Partial dataset

### 3.3 模型测试

#### 3.3.1 激活函数对比

本文所提模型拟采用 ELU 激活函数代替常用的 ReLU 函数。为了证明 ELU 激活函数模型的优越性,使用 ReLU 函数、PReLU<sup>[25]</sup> 函数来继续进行对比。激活函数如式(6)一式(8)所示。只更改模型中的激活函数,模型中的其他部分均保持不变,如图 8 所示,并在同样的训练集下进行训练。

$$ELU(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha(e^x - 1), & \text{if } x < 0 \end{cases} \quad (6)$$

其中,  $\alpha$  为调整的参数,控制 ELU 负值部分饱和的时间。

$$ReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \quad (7)$$

$$PReLU(x) = \begin{cases} x_i, & \text{if } x_i \geq 0 \\ \alpha_i x_i, & \text{if } x_i < 0 \end{cases} \quad (8)$$

其中,  $\alpha_i$  是固定的,  $i$  表示不同的通道对应不同的  $\alpha_i$ 。

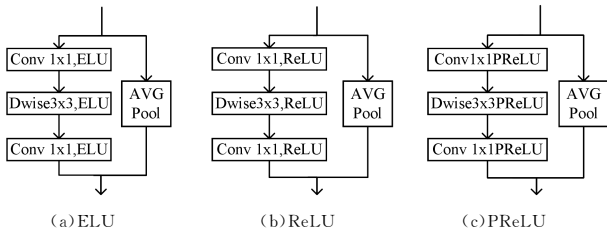


图 8 对比激活函数

Fig. 8 Contrast activation function

由图 9 可以看出,在迭代的前 15 000 次, PReLU 的收敛速度略快于 ELU 和 ReLU;但随着迭代次数的增加,相对于采用 ReLU, PReLU, 采用 ELU 激活函数时最终的损失函数值较小,模型的训练效果更好。使用各个激活函数的模型损失函数值如表 1 所列。

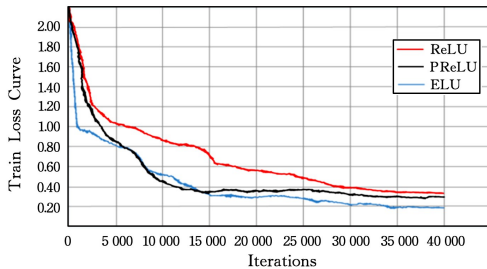


图 9 损失函数的变化

Fig. 9 Loss function change

表 1 损失函数值

Table 1 Loss function value

激活函数	模型迭代次数对应的损失函数值			
	10 000	20 000	30 000	40 000
ELU	0.51	0.31	0.20	0.18
PReLU	0.46	0.38	0.31	0.29
ReLU	0.88	0.56	0.39	0.35

由图 10 可以看出,使用 ELU 激活函数时,模型的准确率相较于 ReLU, PReLU 的准确率提升了 3% 以上,如表 2 所列。

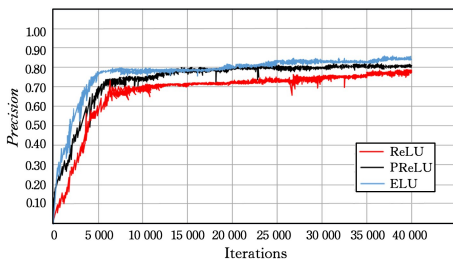


图 10 准确率的变化

Fig. 10 Accuracy change

表 2 准确率对比

Table 2 Accuracy comparison

激活函数	准确率
ELU	0.846
PReLU	0.811
ReLU	0.783

综上所述,使用 ELU 激活函数时,模型效果最优。

### 3.3.2 全局池化对比

为了验证全局池化层和全连接层的效果,在模型 Softmax 层前分别连接一个全连接层和全局池化层,以对比模型的准确率。实验均在同样的测试集和测试环境下进行,结果如表 3 所列。

表 3 算法准确率的对比

Table 3 Algorithm accuracy comparison

方法	准确率
E-MobileNet+全局池化	0.846
E-MobileNet+全连接	0.829

相比于全连接的方法,全局池化方法的准确率略有提升,但减少了很多参数,加快了计算速度。

### 3.4 模型对比实验

为了展现算法的轻量级和在移动端检测的能力,将 E-MobileNet 与当前主流的目标检测模型进行对比(测试环境不采用 GPU 加速,仅使用 CPU),测试集来自 ImageNet 中的部分类别(坦克、坦克炮塔、装甲车辆、人、枪)。用于对比的目标识别模型包括 SSD, ShuffleNet, YOLOv3, MobileNet v2。

上述几种方法在测试集中的平均准确率(mAP)和每秒检测帧数(FPS)结果如表 4 所列。由表 4 中的数据可知,SSD 算法的 mAP 略高于本文算法,但是 FPS 远低于本文算法,达不到实时的效果。YOLOv3 和 ShuffleNet 算法的 mAP 和 FPS 均低于本文方法。MobileNet 与本文算法 FPS 接近,但是准确率较低。在综合识别准确率与检测速度的情况下,本文目标识别方法的效果最好,部分测试结果如图 11 所示。

表 4 算法的平均准确率与帧率

Table 4 Algorithmic mAP and FPS

方法	mAP	FPS
SSD	0.849	2.4
YOLO v3	0.821	20
ShuffleNet	0.733	28
MobileNet v2	0.748	34
E-MobileNet(Ours)	0.846	37



图 11 部分测试结果

Fig. 11 Partial test results

**结束语** 为了解决目前识别方法在移动端识别准确率和实时性难以兼顾的问题,本文提出了一种基于卷积神经网络的实时目标识别算法。该算法加入了 ELU 激活函数,降低了计算复杂度,减弱了输入噪声的影响;添加了残差支路,使得模型能复用浅层的特征;组合平均池化层的残差支路与可分

离卷积层的特征,提高了模型的学习效果;使用全局池化层减少了计算参数,避免了模型的过拟合。实验结果表明,在不使用GPU加速的环境下,本文算法的 *mAP* 和 *FPS* 均优于现有主流识别方法,在保证识别率的基础上加快了识别速度。但由于陆战场环境复杂,导致正负样本不均衡,如果修改损失函数加强对难分类样本的学习,则会使准确率得到一定提升。另外,本文并没有考虑损失函数对目标识别准确率和实时性的影响,后期可以针对这一点做进一步的研究。

## 参 考 文 献

- [1] 李未. 人工智能新时代的群体智能[N]. 中国信息化周报, 2017-09-18(007).
- [2] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [3] CAI M C, LV S K. Preliminary probing into intelligent warfare and its supporting technology system [J]. National Defense Science and Technology, 2017(1): 94-98.
- [4] HUANG J M. Aiming at Intelligent Innovative Command Concept [N]. PLA Newspaper, 2018-09-27 (007).
- [5] WANG Y. Research on Battlefield Target Identification and Situation Intention Forecasting[D]. Wuxi: Jiangnan University, 2015
- [6] CHEN H Y, SU C Y. An Enhanced Hybrid MobileNet[C]// 2018 9th International Conference on Awareness Science and Technology (ICAST). IEEE, 2018: 308-312.
- [7] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2012: 1097-1105.
- [9] PAL M, FOODY G M. Feature selection for classification of hyperspectral data by SVM[J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(5): 2297-2307.
- [10] SHIN H C, ROTH H R, GAO M, et al. Deep convolutional neural networks for computer-aided detection; CNN architectures, dataset characteristics and transfer learning[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1285-1298.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]// European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [12] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6848-6856.
- [13] JU M, LUO H B, WANG Z B. Improved YOLO V3 Algorithm and Its Application in Small Target Detection[J]. Acta Optica Sinica, 2019, 39(7): 0715004.
- [14] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4510-4520.
- [15] LEMLEY J, BAZRAFKAN S, CORCORAN P. Deep Learning for Consumer Devices and Services; Pushing the limits for machine learning, artificial intelligence, and computer vision[J]. IEEE Consumer Electronics Magazine, 2017, 6(2): 48-56.
- [16] MHASKAR H, LIAO Q, POGGIO T. When and why are deep networks better than shallow ones? [C]// Thirty-First AAAI Conference on Artificial Intelligence. 2017.
- [17] WANG G, CHEN J Y, GAO F, et al. Research on the Infrastructure Target Detection of Remote Sensing Image Based on Deep Learning[J]. Radio Engineering, 2018, 48(3): 219-224.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [19] CHOLLET F. Xception; Deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1251-1258.
- [20] JIANG A B, WANG W W. Research on optimization of ReLU activation function[J]. Sensors and Microsystems, 2018, 37 (2): 50-52.
- [21] SHANG W, SOHN K, ALMEIDA D, et al. Understanding and improving convolutional neural networks via concatenated rectified linear units [C] // International Conference on Machine Learning. 2016: 2217-2225.
- [22] SCHWING A G, URTASUN R. Fully connected deep structured networks[J]. arXiv:1503.02351, 2015.
- [23] LIN M, CHEN Q, YAN S. Network in network [J]. arXiv: 1312. 4400, 2013.
- [24] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [25] CAO H Z. Research on Image Classification Based on Improved Convolution Neural Network [D]. Nanning: Guangxi Normal University, 2017.



**QIAO Meng-yu**, born in 1995, postgraduate. His research interests include computer vision and depth of learning.



**WANG Peng**, born in 1978, Ph.D, professor. His principal research interests include computer vision and embedded systems.