第 58 卷 第 16 期/2021 年 8 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于改进 MobileNetV2SSDLite 岸桥表面锈蚀检测

韩冬[1*]，唐刚[1]，赵政坤[2]

[1]上海海事大学物流工程学院，上海 201306；

[2]约克大学电子工程学院，英国 约克 YO105DD

摘要 为了使岸桥表面锈蚀检测任务在部署到嵌入式设备和移动设备中获取更快的推理速度，在不牺牲精度的前提下提出改进的轻量化目标检测网络 MobileNetV2SSDLiteV1/V2。改进后的网络采用 5 个卷积层的特征映射作为检测器输入，并使用 3×3 深度卷积预测分类和位置得分。为了减少骨干网络的参数量，将原始 17 个线性残差模瓶颈块结构设计成 14 个，并将分辨率为 256 pixel×256 pixel 的图像作为网络输入，同时改变原始默认框的系数，使先验框的个数减少了 82.51%，接着将所有层进行批处理归一化并从零开始训练网络。以上改进可以使网络参数变为 0.96×10^6，减少至原来的 1/4，网络的浮点运算次数为 0.12×10^9，较原始减少 81.25%，mAP 值高达 77.40%，推理速度达 45 frame/s。

关键词 机器视觉；深度学习；目标检测；岸桥；轻量级网络；边缘计算

中图分类号 F407.42；TP312 **文献标志码** A **doi**：10.3788/LOP202158.1615006

Surface Corrosion Detection of Quayside Crane Based on Improved MobileNetV2SSDLite

Han Dong[1*], Tang Gang[1], Zhao Zhengkun[2]

[1] *School of Logistics Engineering, Shanghai Maritime University, Shanghai* 201306, *China*;

[2] *Department of Electronic Engineering, University of York, York YO*105*DD, UK*

Abstract In order to enable the quayside crane surface corrosion detection task to be deployed to embedded devices and mobile devices to obtain faster inference speed, an improved lightweight object detection network MobileNetV2SSDLiteV1/V2 is proposed without sacrificing accuracy. The improved network uses feature maps of 5 convolutional layers as the detector input, and uses 3×3 depthwise convolution to predict classification and location scores. In order to reduce the number of parameters of the backbone network, the original 17 inverted linear bottleneck block structure is designed into 14, and the image with a resolution of 256 pixel × 256 pixel is used as the network input to change the coefficients of the original default box. The number of is reduced by 82.51%, and then all convolutions are normalized and the network is trained from scratch. The above improvements can make the network parameters become 0.96×10^6, which is reduced to 1/4 of the original. The number of floating-point operations of the network is 0.12×10^9, which is 81.25% less than the original, the mAP value is as high as 77.40%, and the inference speed reaches 45 frame/s.

Key words machine vision; deep learning; object detection; quayside crane; lightweight network; edge computing

OCIS codes 150.1135; 100.3008; 100.2960

收稿日期：2020-11-09；**修回日期**：2020-12-07；**录用日期**：2020-12-17

基金项目：上海市青年科技英才杨帆计划(19YF1419100)

通信作者：*hd19821252578@163.com

1 引　　言

近年来，随着机器学习技术和深度学习技术的飞速发展，人工智能已经渗透到各个领域，包括以循环神经网络（Recurrent Neural Network，RNN）[1]和长短期记忆（Long Short-Term Memory，LSTM）人工神经网络[2]为代表的技术在自然语言处理中的应用，以卷积神经网络为核心的技术在计算机视觉领域中的研究。目标检测作为计算机视觉的基础任务之一，在安防、无人驾驶和机器人中至关重要。目标检测任务是在分类任务的基础上对目标进行定位，在人脸识别[3]、人脸检测[4]、行人检测[5]、车辆检测[6]、交通信号灯检测[7]和车道线检测[8-9]等任务中已有广泛的应用，而且在检测速度和检测精度方面均超越传统手工特征提取与分类器结合的方法。在民生基建中，目标检测常用于道路检测[10]、混凝土裂缝检测[11]、轨道缺陷检测[12]和桥梁裂缝检测[13]等任务。吴之昊等[14]提出了一种基于轻量级 SSD（Single Shot MultiBox Detector）[15]的电力设备锈蚀目标检测方法，该方法基于注意力模型的上采样特征融合策略来弥补缩减模型结构所带来的精度损失。陈立里等[16]提出了改进的具有实时检测能力的 SSD 算法，在计算量有限的平台上可以实时检测并保持精度。李强等[17]提出了基于视觉技术的船舶船身锈蚀检测技术，证明了该技术的检测范围比无人机更大。Yuan 等[18]提出了基于区域建议网络-全卷积神经网络（Region Proposal Network-Full Convolution Network，RPN-FCN）的电力设备锈蚀检测方法，相比于其他方法改进了锈蚀检测的精度。谢学立等[19]针对现有显著性目标检测算法存在的显著区域检测不均匀和边缘表示模糊等问题，提出了一种双注意力循环卷积显著性目标检测算法，该算法可以提高目标检测算法的性能。

本文将深度学习目标检测网络应用于岸桥的健康监测，详细研究单阶段目标检测算法 SSD 结合轻量级卷积神经网络（MobileNetV2）在岸桥锈蚀检测中的应用，同时对现有算法进行改进。原始的 MobileNetV2SSDLite 虽然为轻量级算法，但是其仍然具有 0.64×10^9 次的浮点运算，在一些算力较小的数字信号处理器（Digit Signal Processor，DSP）上的推理速度较慢，而且当网络结构更改时，无法使用预先给出的训练模型，并且当网络从零开始训练时会难以收敛。本文根据这些问题对网络进行裁剪和改进以提高网络的推理速度，并满足从零开始训练的要求，从而提升小目标的检测精度，为嵌入式和移动设备的部署提供条件。

2 MobileNetV2SSDLite 检测模型

MobileNetV2SSDLite 检测模型使用分辨率为 300 pixel×300 pixel 的输入图像，并且在 conv 13、conv 1、layer 19_2_2、layer 19_2_3、layer 19_2_4 和 layer 19_2_5 这 6 个特征映射层上对图像进行检测，检测模型的结构如图 1 所示，其中 DW 为深度卷积。不同于原始的 SSD 检测模型（每层默认框的个数分别为 4、6、6、6、4 和 4，一共可以产生 8732 个先验框），MobileNetV2SSDLite 检测模型的每层默认框的个数分别为 3、6、6、6、6 和 6，一共可以产生 7308 个先验框。

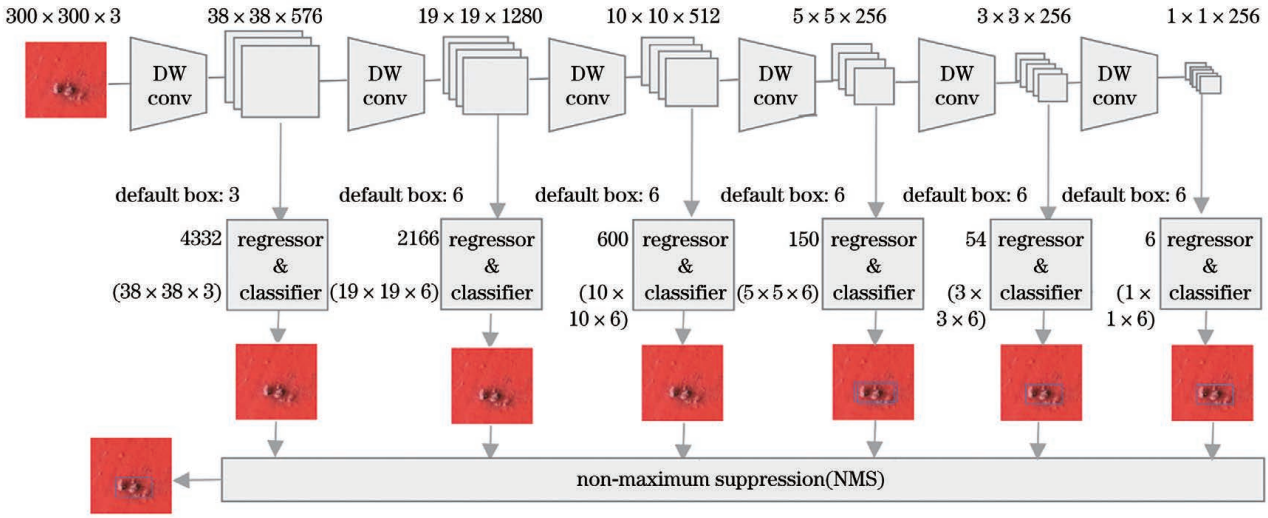

图 1　MobileNetV2SSDLite 检测模型的结构

Fig. 1　Structure of MobileNetV2SSDLite detection model

2.1 先验框的生成和匹配

先验框的设置包括长宽比和尺度两个方面。对于尺度 s 的设置，其遵守线性递增规则，随着特征图大小的降低，s 值的大小呈线性增加。先验框大小相对于原始图的比例可表示为

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(k-1),\ k \in [1,m], \tag{1}$$

式中：m 表示特征图的个数，原始检测器(SSD)采用 6 个特征映射作为输入；$s_{\min}$ 值和 $s_{\max}$ 值分别为 0.20 和 0.95。由(1)式可以得到各个特征图的先验框尺寸，分别为(60 pixel，60 pixel)、(105 pixel，150 pixel)、(150 pixel，195 pixel)、(195 pixel，240 pixel)、(240 pixel，285 pixel)和(285 pixel，300 pixel)，本文是对 5 个特征映射图进行检测并使用分辨率为 256 pixel×256 pixel 的输入图像，因此 $s_{\min}$ 值和 $s_{\max}$ 值分别为 0.2 和 0.8，各个特征映射图的先验框尺寸分别为(51.2 pixel，51.2 pixel)、(89.6 pixel，128.0 pixel)、(128.0 pixel，166.4 pixel)、(166.4 pixel，204.8 pixel)和(204.8 pixel，256.0 pixel)。网络的先验框详细尺寸和特征映射分布如表 1 所示。

表 1　网络先验框的尺寸

Table 1　Size of network priori box

Network structure	Layer	Size/(pixel, pixel)
Original SSD	conv 4_3, conv 7, conv 8_2, conv 9_2, conv 10_2, conv 11_2	(30, 60), (60, 111), (111, 162), (162, 213), (213, 264), (264, 315)
MobileNetV2SSDLite	conv 13, conv 1, layer 19_2_2, layer 19_2_3, layer 19_2_4, layer 19_2_5	(60, 60), (105, 150), (150, 195), (195, 240), (240, 285), (285, 300)
Ours-MobileNetV2SSDLite	conv 13, conv 1, layer 19_2_2, layer 19_2_3, layer 19_2_4	(51.2, 51.2), (89.6, 128.0), (128.0, 166.4), (166.4, 204.8), (204.8, 256.0)

在训练过程中，需要确定与训练图像中真值框相匹配的先验框，该先验框所对应的边界框负责预测。首先，对于每个真值框，找到与其交并比(Intersection over Union，IoU)最大的先验框并使用其与之匹配。其次，将 IoU 值大于 0.5 且剩余未匹配的先验框与真值框匹配。为了保证正负样本尽量平衡，采用难负例挖掘对负样本进行抽样，抽样过程中按照置信度误差进行降序排列，选取误差较大的前 k' 个置信度误差作为训练的负样本，以保证正负样本比例接近 1∶3。

2.2 损失函数

SSD 的损失函数是定位损失 L_{loc} 和置信度损失 L_{conf} 的线性组合，表达式为

$$L(x,c,l,g) = \frac{1}{N}\left[L_{\text{conf}}(x,c) + \alpha L_{\text{loc}}(x,c) + \alpha L_{\text{loc}}(x,l,g)\right], \tag{2}$$

式中：l 表示预测框；g 表示真值框；c 表示置信度；x 表示输入特征；N 表示匹配到真值框的先验框数量；α 为调整置信度损失和位置损失之间的比例，默认 $\alpha=1$。对于置信度损失，其采用的是在多类别置信度上的 Softmax 损失，表达式为

$$L_{\text{conf}}(x,c) = -\sum_{i\in P_c}^{N} x_{ij}^{p}\ln(\hat{c}_i^{p}) - \sum_{i\in N_e}\ln(\hat{c}_i^{0}), \tag{3}$$

其中

$$\hat{c}_i^{p} = \frac{\exp(c_i^{p})}{\sum_{p}\exp(c_i^{p})}, \tag{4}$$

式中：P_c 和 N_e 分别表示正样本和负样本的集合；i 表示搜索框的序号；j 为真实框的序号；p 表示类别序号，当 $p=0$ 时，表示背景；x_{ij}^{p} 表示第 i 个预测框与第 j 个真值框关于类别 k 是否匹配，当 $x_{ij}^{p}=1$ 时，表示第 i 个先验框匹配到第 j 个真值框，真值框的类别为 p；c_i^{p} 表示第 i 个搜索框对应类别 p 的置信度。(3)式的第一项为正样本的损失函数，第二项为负样本的损失函数。对于位置回归损失，其采用的是 Smooth L1 损失，表达式为

$$L_{\text{loc}}(x,l,g) = \sum_{i\in P_c}^{N}\sum_{m'\in(X_C,Y_C,w,h)} x_{ij}^{k} S_{\text{SmoothL1}}(l_i^{m'} - \hat{g}_j^{m'}), \tag{5}$$

其中

$$\begin{cases}\hat{g}_j^{X_C} = (g_j^{X_C} - \mathrm{d}_i^{X_C})/\mathrm{d}_i^{w}\\ \hat{g}_j^{Y_C} = (g_j^{Y_C} - \mathrm{d}_i^{Y_C})/\mathrm{d}_i^{h}\end{cases}, \tag{6}$$

$$\begin{cases}\hat{g}_j^{w} = \ln(g_j^{w}/\mathrm{d}_i^{w})\\ \hat{g}_j^{h} = \ln(g_j^{h}/\mathrm{d}_i^{h})\end{cases}, \tag{7}$$

式中：(X_C,Y_C)为补偿后默认框的中心；(w,h)为默认框的宽和高；$S_{\text{SmoothL1}}(\cdot)$表示 Smooth L1 损失函数。

3　改进的 MobileNetV2SSDLite 岸桥表面锈蚀检测

轻量级分类网络 MobileNetV2[20] 是 MobileNetV1[21] 的升级，MobileNetV2 不仅使用深度可分离卷积来减少计算量，同时引入反残差线性结构来提高网络特征的提取能力，当 MobileNetV2 的计算量为 1/4 的 MobileNet 时，就能达到与 MobileNet 相同的性能。MobileNetV2 在每个反残差模块后采用线性激活函数，解决了低维空间经过非线性激活函数输出后存在的特征损失问题。

3.1　深度可分离卷积

深度可分离卷积层由深度卷积层和逐点卷积层构成。深度卷积层使用单个卷积核对每一个输入通道进行卷积，得到输入通道数的深度后使用逐点卷积层进行处理，即应用一个简单的 1×1 卷积对深度卷积中的输出进行线性结合。深度卷积层对每一个通道使用一种卷积核，卷积过程可表示为 $\hat{G}_{k,l,m}=\sum_{i',j'}\hat{K}_{i',j',m''}\cdot F_{k^*+i'-1,l'+j'-1,m'}$，即 $\hat{K}$ 中第 m'' 个卷积核应用于 F 中的第 m' 个通道来产生第 m' 个通道的卷积输出特征图 $\hat{G}$，其中 $\hat{K}$ 为深度卷积核的尺寸，F 为逐通道卷积操作，$\hat{G}$ 为逐深度卷积的输出，$k^*\times l'$ 为深度卷积输出特征图的尺寸，i' 和 j' 为深度卷积核的索引。深度卷积层的计算量为 $D_K\cdot D_K\cdot M\cdot D_F\cdot D_F$，其中 D_F 为特征图大小，D_K 为深度卷积的卷积核大小，M 为输入通道数量。对于标准卷积层，深度卷积层的效果明显，但其只对输入通道进行卷积，并未对其进行组合来产生新的特征，因此使用 1×1 卷积对深度卷积层的输出进行线性组合以产生新的特征。深度可分离卷积层的计算量为 $D_K\cdot D_K\cdot M\cdot D_F\cdot D_F+M\cdot N\cdot D_F\cdot D_F$，即深度卷积和 1×1 逐点卷积的和，其中 N 为输出通道数量。将卷积分为滤波和线性组合的过程，得到的计算量为 $\frac{D_K\cdot D_K\cdot M\cdot D_F\cdot D_F+M\cdot N\cdot D_F\cdot D_F}{D_K\cdot D_K\cdot M\cdot D_F\cdot D_F}=\frac{1}{N}+\frac{1}{D_K^2}$，MobileNetV1 使用 3×3 深度可分离卷积的计算量比标准卷积少了 1/8～1/9，说明 3×3 深度可分离卷积的准确率只有极小的下降。标准卷积和深度可分离卷积的过程如图 2 所示。

3.2　反残差网络结构

在 MobileNetV1 网络中反残差结构的基础上引入了旁路连接结构，残差结构的作用是先降维后升维，首先使用 1×1 卷积进行压缩，然后使用 3×3 卷积进行特征提取，最后使用 1×1 卷积将整个通道数还原。该结构减少了 3×3 卷积模块的计算量，提高了整个残差模块的计算效率。反残差模块的作用是先升维后降维，首先使用 1×1 卷积进行通道扩张，然后使用 3×3 深度卷积进行特征提取，最后使用 1×1 卷积将通道数压缩，目的是使 3×3 深度卷积层提取更多的特征，过程与残差结构相反，如图 3 所示。

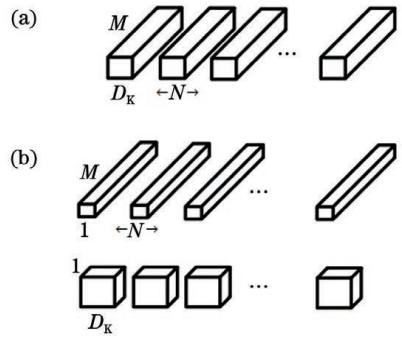

图 2　标准卷积和深度可分离卷积的过程。(a)标准卷积；(b)深度可分离卷积

Fig. 2　Process of standard convolution and depth separable convolution. (a) Standard convolution; (b) depth separable convolution

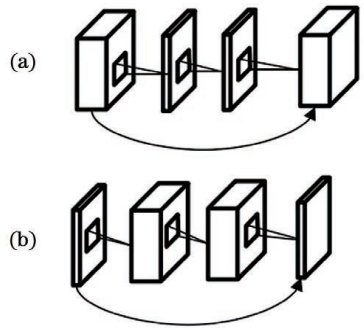

图 3　不同残差结构。(a)残差结构；(b)反残差结构

Fig. 3　Different residuals. (a) Residual structure; (b) inverted residual structure

改进后的反残差模块在每个深度卷积层和逐点卷积层后进行归一化处理，如图 4 所示，其中 BN 为批处理归一化。该模块分为两个部分，即步长分别为 1 和 2 的情况。当滑动步长为 1 时，使用 Shortcut 结构对特征进行融合。

3.3　改进后的 MobileNetV2SSDLite 网络结构

MobileNetV2 网络升维是因为深度卷积层没有改变通道数的能力，所以在低维上的表现较差，为此使用基础模块先对网络进行升维以提高准确率。当使用 1×1 逐点卷积进行通道扩张时，扩张系数均

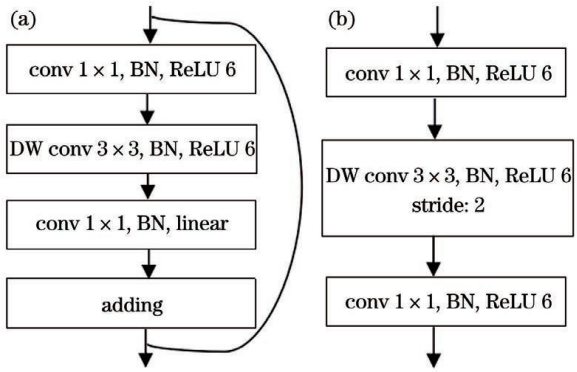

图 4　反残差模块。(a)步长为 1;(b)步长为 2

Fig. 4　Inverted residual block. (a) Stride is 1; (b) stride is 2

为 6。为了减少网络的参数量,将 17 个线性瓶颈结构减少到 14 个,最终减少了 12 个卷积计算,基础网络结构的参数如表 2 所示。

改进后的网络采用 5 个特征映射层进行检测,以分辨率为 256 pixel×256 pixel 的图像作为输入,每层默认框的个数分别为 3、6、6、6 和 6,一共有 1278 个先验框,减少了 82.51%,使用 3×3 深度卷积在每个特征映射层上来预测分类得分和位置得分,并在卷积层后加上 BN(Batch Normalization)层,先验框的计算过程如图 5 所示,网络结构参数如表 3 所示。

表 2　基础网络结构的参数

Table 2　Parameters of basic network structure

Backbone	Bottleneck group	Expand ratio	Times of repetition	Total bottleneck
Original MobileNetV2	7	1, 6, 6, 6, 6, 6, 6	1, 2, 3, 4, 3, 3, 1	17
Ours-MobileNetV2	8	1, 6, 6, 6, 6, 6, 6, 6	3, 2, 1, 2, 1, 1, 2, 2	14

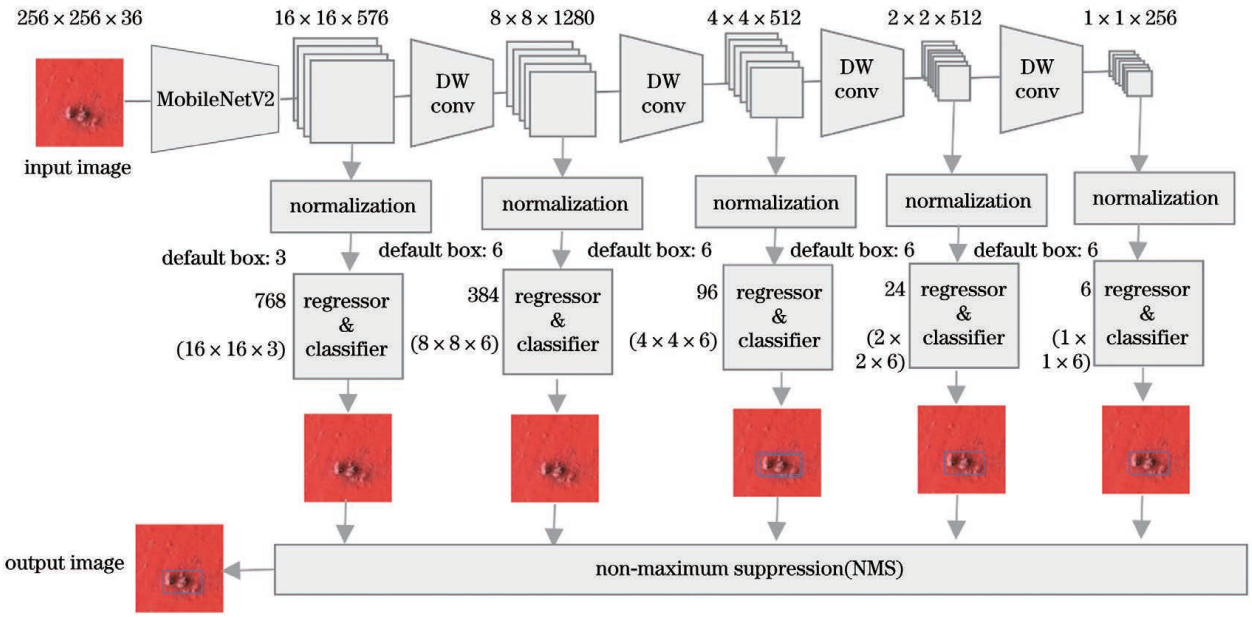

图 5　改进后的网络结构

Fig. 5　Structure of improved network

表 3　网络结构的参数

Table 3　Parameters of network structure

Network structure	Size of input image /(pixel×pixel)	Number of default boxes	Feature map size /(pixel×pixel)	Number of prior boxes
Original SSD	300×300	4, 6, 6, 6, 4, 4	38×38, 19×19, 10×10, 5×5, 3×3, 1×1	8732
MobileNetV2SSDLite	300×300	3, 6, 6, 6, 6, 6	38×38, 19×19, 10×10, 5×5, 3×3, 1×1	7308
Ours-MobileNetV2SSDLite	256×256	3, 6, 6, 6, 6	16×16, 8×8, 4×4, 2×2, 1×1	1278

3.4 训练方式

网络采用了从零开始训练、通道数量减半和无偏置训练以及多尺度的输入图像综合训练的方式，训练过程如下。

1）从零开始训练。原始网络只在特征提取层的每个卷积层后使用 BN，训练过程中初始学习率为 0.0001，批处理数量为 64，权重衰减率为 0.00001，损失值在 10 附近震荡，无法收敛，实验证明这是脱离预训练模型所导致的。通过调整超参数来增加岸桥数据集的样本容量，网络均未收敛，最后在检测部分所有卷积层后均加入 BN 层，将初始学习率调整至 0.001，迭代至 2 万步损失值减少至原来的 1/4，使用较大的学习率来迭代网络，收敛速度提高了近 3 倍，迭代至 1 万步损失值下降到 3 以内，迭代至 6 万步收敛至 1 上下，这大大地缩短了训练时间，同时提高了精度。

2）通道减半和无偏置训练。通过 3.3 节对网络的改进，参数量减少至 3.32×10^6。为了进一步减少网络的计算量以及在嵌入式设备上更快运行，考虑将网络的所有通道数减半，在训练过程中舍弃偏置，此时网络的计算量减少至 0.96×10^6，减少至原来的 1/4，实验证明这种方法只稍微降低了精度，平均精度均值（mean Accuracy Precision mAP）值下降了 1.78%，但是大大地提高了网络前向传播的速度。

3）多尺度图像训练。为了提高岸桥小目标的识别精度，首先考虑特征融合的方法，将高层的特征映射信息融合到低层上，通过低层更丰富的特征信息来提升小目标的检测精度，其次使用较大分辨率的岸桥图像作为输入，但这两种方法均会增加计算量。因此本文将图像等比缩放至不同尺度，该方法能够提高小目标的检测精度，最终岸桥图像随机裁剪成 128 pixel×128 pixel、160 pixel×160 pixel、192 pixel×192 pixel 和 224 pixel×224 pixel。

4 实验细节及结果分析

实验部分包括岸桥数据的预处理、Lmdb（caffe 数据集格式）数据集的制作、网络的训练和测试以及不同骨干网络的性能对比。通过轻量级骨干网络的性能对比实验，分析 MobileNetV2、SqueezeNet[22]、MobileNetV1[21]以及 ShuffleNet[23]的参数量、浮点运算次数以及 mAP 等。

4.1 数据获取

岸桥数据是在上海海事大学南京港机上采集获取的，图像尺寸分别为 320 pixel×240 pixel 和 256 pixel×256 pixel，图像大小为 25 kB。

4.2 数据的预处理

数据的预处理过程分为数据集的前处理和数据增强两个部分。网络输入接口的尺寸为 256 pixel×256 pixel，开发板采集到的图像尺寸分别为 320 pixel×240 pixel 和 256 pixel×256 pixel，其中 320 pixel×240 pixel 岸桥图的宽高比不是 1∶1，而且在输入网络的过程中会产生拉伸变形，影响识别的准确率，因此将所有的 320 pixel×240 pixel 岸桥图片将宽等比例缩小至 256 pixel，并将高向下填充至 256 pixel。受到 YOLO-V3（You Only Look Once-V3）[24]的启发，使用多尺度训练的方式来处理岸桥数据以提高小目标的检测精度。将岸桥图像随机裁剪成 128 pixel×128 pixel、160 pixel×160 pixel、192 pixel×192 pixel 和 224 pixel×224 pixel，并保留原始尺寸为 256 pixel×256 pixel 的图像，可以实现 5 种尺度的岸桥图像训练。网络输入接口的尺寸被设为 256 pixel×256 pixel，当不同尺度的岸桥图片输入网络中时，会将图像尺寸统一到 256 pixel×256 pixel，那些裁剪过多并且小于 256 pixel×256 pixel 的图像中的目标在整个图像中的占比会增大，实现了在不增加网络开销的情况下将小目标的识别问题转换为一般的识别问题，从而提高网络的精确度和召回率。

前处理得到的岸桥图像保证了输入图像满足 1∶1 的宽高比，而且在训练过程中不会发生数据拉伸和压缩变形，最终获得了 2522 张多尺度的岸桥锈蚀图像，按照 0.8（训练验证集）和 0.2（测试集）的比例制作 Lmdb 岸桥数据集。

前处理帮助网络实现了多尺度训练，前处理后网络还会对输入的岸桥数据集进行增强处理，包括颜色扭曲、随机裁剪、水平翻转和随机采集块域，数据增强图像如图 6 所示。颜色扭曲是通过调节对比度来增强样本；随机裁剪是在图像中随机裁剪出一个样本；水平翻转是将原图以 0.5 的概率随机水平翻转；随机采集块域是将随机裁剪出的块域缩小后填入图像中。

4.3 实验环境

在 Ubuntu 16.04 系统中搭建 OpenCV4.1、Python2.7、Caffe-ssd、Cuda10.0 和 Cudnn7.5 作为软件环境，硬件环境为 Titan XP GPU。

4.4 网络训练

使用并行式计算和从零开始训练的方式来训练

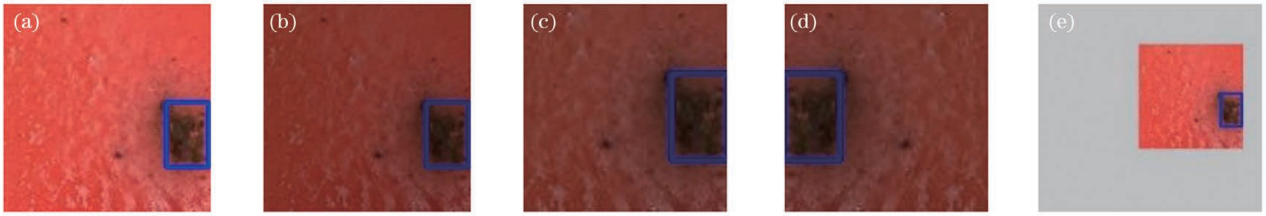

图 6　数据增强后的图像。(a)原始图像;(b)颜色扭转;(c)随机裁剪;(d)水平翻转;(e)随机采样

Fig. 6　Images after data enhancement. (a) Original image; (b) color distortion; (c) random cropping; (d) horizontal flip; (e) random sampling

网络,初始学习率为 0.001,权值衰减系数为 0.0005,优化器为 RMSProp。

4.5　实验结果分析

为了评估所提方法对岸桥锈蚀检测的有效性,使用精确率、召回率和 mAP 三种标准的实验评测指标来评估网络性能,分别表示为

$$P=\frac{x_{\mathrm{TP}}}{x_{\mathrm{TP}}+x_{\mathrm{FP}}},\tag{8}$$

$$R=\frac{x_{\mathrm{TP}}}{x_{\mathrm{TP}}+x_{\mathrm{FN}}},\tag{9}$$

$$M_{\mathrm{mAP}}=\int_0^1 P(R)\mathrm{d}(R),\tag{10}$$

式中:x_{TP} 为真正例的样本个数;x_{FP} 为假正例的样本个数;x_{FN} 为假反例的样本个数。

不同网络的性能对比结果如表 4 所示,其中 FLOPs 为浮点运算次数,FPS 为每秒传输帧数。在面向嵌入式和移动设备的模型部署过程中,YOLO[25] 系列对比了 YOLOV3-Tiny[24]。SSD 则分析对比了轻量级算法 SqueezeNet、MobileNetV1 和 ShuffleNet,MobileNetV2-SSDLite 网络改进后的网络即为 MobileNetV2SSDLiteV1 和 MobileNetV2SSDLiteV2(MobileNetV2SSDLiteV1/V2),其中后者无偏置且通道数减半,而且在大幅度减少网络参数和浮点运算次数的同时产生较小的精度损失,使模型的推理速度更快。从表 4 可以看到,MobileNetV2SSDLiteV2 的参数量相比于 MobileNetV2SSDLite 减少约为 1/4,mAP 相对于 MobileNetV2SSDLiteV1 下降 1.78 个百分点。训练过程中,输入图像尺寸为 256 pixel×256 pixel 的 MobileNetV2SSDLite 无法采用预训练模型,而且损失值不收敛,因此未给出 mAP 值。从表 4 可以看到,MobileNetV2SSDLiteV2 在 GPU 上的推理速度达到 45 frame/s;浮点运算次数为 0.12×10^9,较原始减少 81.25%,说明浮点运算量最少,综合性能最好,适合嵌入式和移动端的应用。

表 4　不同网络的性能对比

Table 4　Performance comparison of different networks

Network	Params / 10^6	FLOPs / 10^9	mAP / %	FPS
ShuffleNet-SSD	8.26	1.23	74.13	42
YOLOV3-Tiny	7.90	0.81	72.14	40
MobileNetV1SSD	5.64	0.79	73.63	40
SqueezeNet-SSD	5.52	0.51	76.30	43
MobileNetV2SSDLite	3.32	0.64		
MobileNetV2SSDLiteV1	**3.30**	**0.44**	**77.40**	**40**
MobileNetV2SSDLiteV2	**0.96**	**0.12**	**75.62**	**45**

图 7 为每个检测网络的 mAP 值变化曲线。从图 7 可以看到,最大 mAP 值是 MobileNetV2SSDLiteV1 网络,即 77.40%;2×10^4 次迭代后,每种网络的性能逐渐稳定且基本收敛,随着训练的继续,每种网络的性能差异逐渐减少,轻量级算法在检测精度上的差异较小,为此主要衡量网络性能好坏的标准落在模型计算量和运算速度上。因此本文提出的 MobileNetV2SSDLiteV1/V2 具有实际意义,将参数量减少至原始网络的 1/4,使该网络模型变得更加轻量。

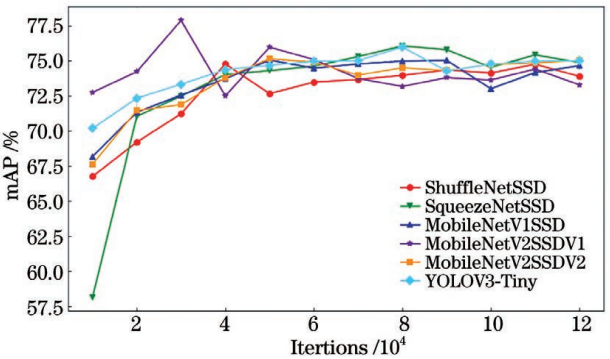

图 7　不同模型的性能对比

Fig. 7　Performance comparison of different models

网络的参数量和浮点运算次数如图 8 所示。从图 8 可以看到,MobileNetV2 相比于

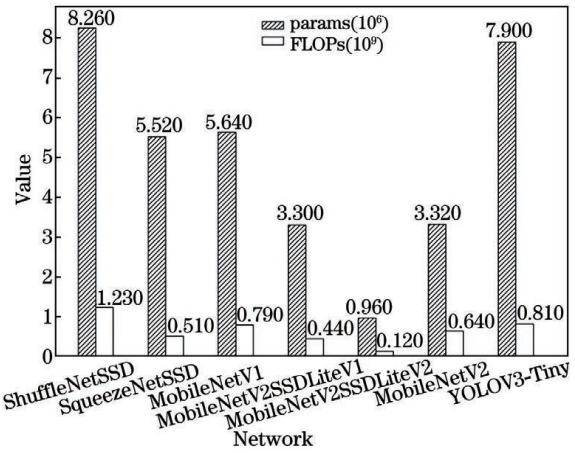

图 8　网络的参数量和浮点运算次数

Fig. 8　Number of network parameters and number of floating point operations

MobileNetV1 减少了近 19% 的网落开销，而 MobileNetV2SSDLiteV1 相比于 MobileNetV2SSDLite 减少了 31%的 FLOPs，MobileNetV2SSDLiteV2 相比于 MobileNetV2 减少了 81.25%的 FLOPs，同时具有 45 frame/s 的推理速度，并且比 SqueezeNet 和 ShuffleNet 具有更快的速度和更高的精度。

为了进一步衡量所提方法对网络综合性能的影响，本文进行了一系列的对比实验，结果如表 5 所示。

从表 5 可以看到，当使用数据增强或多尺度训练时，网络的性能提升明显，mAP 值分别提升了 2.57 个百分点和 2.65 个百分点；当使用整体通道减半和无偏置训练时，可以提升网络的推理速度，但是精度会有一定的损失。

表 5　各种方法对网络性能的影响

Table 5　Impact of various methods on network performance

Data enhancement	Half channel	No bias	Multi-scale	mAP /%	Params /10^6	FPS
No	No	No	No	73.78	3.30	40
No	No	No	Yes	76.43	3.30	40
Yes	No	No	No	76.35	3.30	40
No	Yes	No	No	72.33	1.58	44
No	No	Yes	No	71.98	2.68	41
Yes	No	No	Yes	77.40	3.30	40
Yes	Yes	Yes	Yes	75.62	0.96	45

目标检测通常将精度-召回率(P-R)、精度均值(AP)和 mAP 作为衡量网络模型好坏的综合评价指标。为了进一步评价不同网络模型的综合性能，采用 11 点法绘制不同网络的 P-R 曲线，如图 9 所示。11 点法是在最大积分方法中找到 11 个最大的精确值，然后结合召回率绘制 P-R 曲线，11 点的 AP 和 mAP 可表示为

$$A_{11\text{point}}=\frac{1}{11}\left(\sum x\right),\ x\in P_{\text{maxprecision}},\tag{11}$$

$$M_{\text{mAP}}=\frac{1}{n}\sum A,\tag{12}$$

式中：n 为类别总数。图 9 中的 a～f 分别为 MobileNetV2SSDLiteV2、YOLOV3-Tiny、MobileNetV1SSD、SqueezNetSSD、MobileNetV2SSDLiteV1 和 ShuffleNetSSD。由(10)式可知，AP 是精度在召回率上的积分，即 AP 的几何意义是 P-R 曲线在横坐标和纵坐标上围成的面积，面积越大，AP 值越大。从图 9(c)可以看到，曲线 e 的面积最大，其将其他曲线包围在内，表明 MobileNetV2SSDLiteV1 网络具有最大的 mAP 值，同时该网络在图 9(a)和图 9(b)的 AP 值均大于其他网络。对比图 9(a)和图 9(b)可以看到，裂纹类的 AP 值比腐蚀类小，原因在于裂纹类包含较多的小目标困难样本[图 10(a)]，这种小目标的检测可能会发生漏检或置信度低等情况，从而降低裂纹类的召回率。图 10(b)为目标区域较大的简单样本，不易发生漏检和错检情况，所以精度和召回率均很大。图 10(c)的样本容易发生错检，所以精度降低；并且还可能发生漏检，所以召回率降低。同时图 10(a)和图 10(c)也表明 MobileNetV2SSDLiteV1 能够识别小目标，而置信度得分较低的问题可作为未来研究的改进方向。

图 11 为不同网络模型的检测效果。从图 11(b)可以看到，其他网络 SqueezeNet-SSD 发生错检和漏检，其将 image 1 的“圆形雨点”噪声误判为腐蚀，将image2的“条形雨点”误判为裂纹，而且

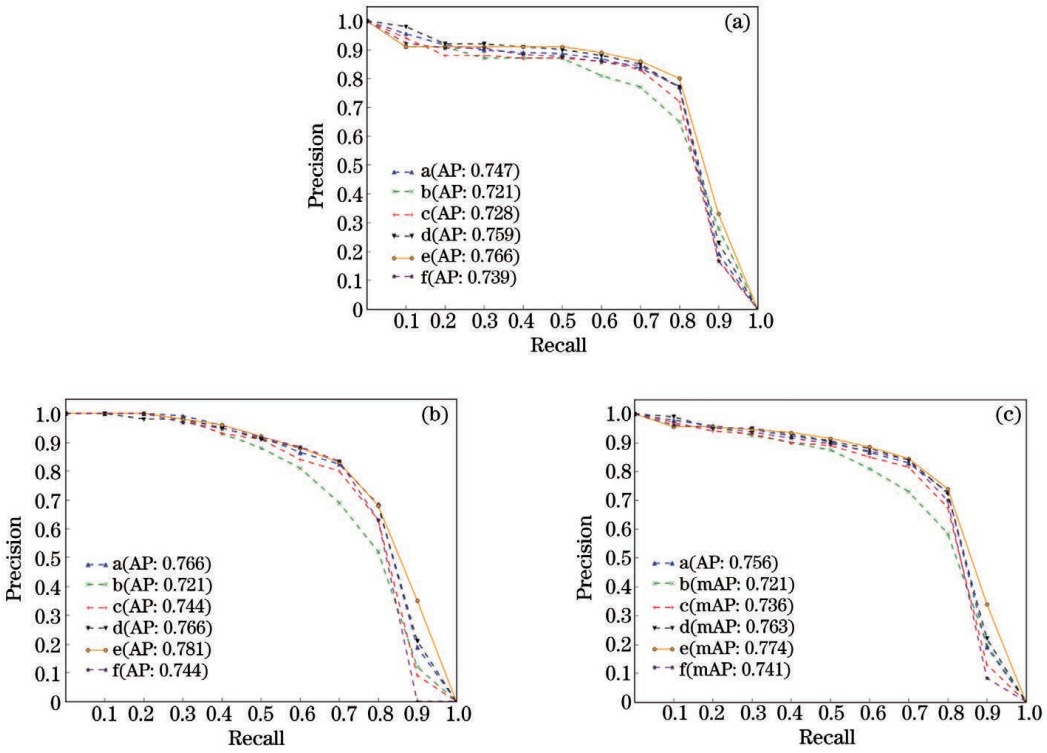

图 9　各种网络在不同情况下的性能曲线。(a)裂纹;(b)腐蚀;(c)总体

Fig. 9　Performance curves of various networks under different conditions. (a) Crack; (b) erosion; (c) overall

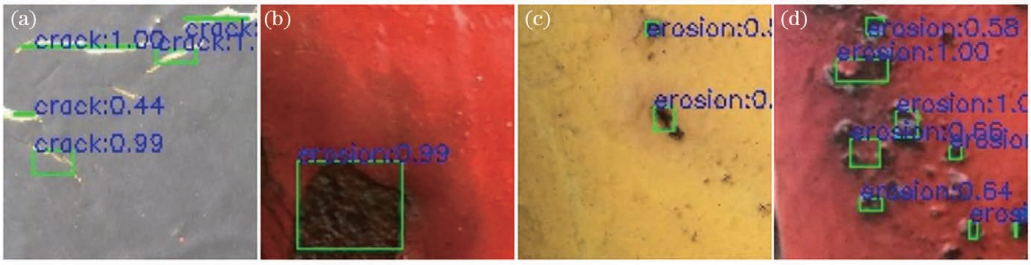

图 10　困难样本和简单样本。(a)困难样本 1;(b)简单样本;(c)困难样本 2;(d)困难样本 3

Fig. 10　Hard samples and easy sample. (a) Difficult sample 1; (b) simple sample; (c) difficult sample 2; (d) difficult sample 3

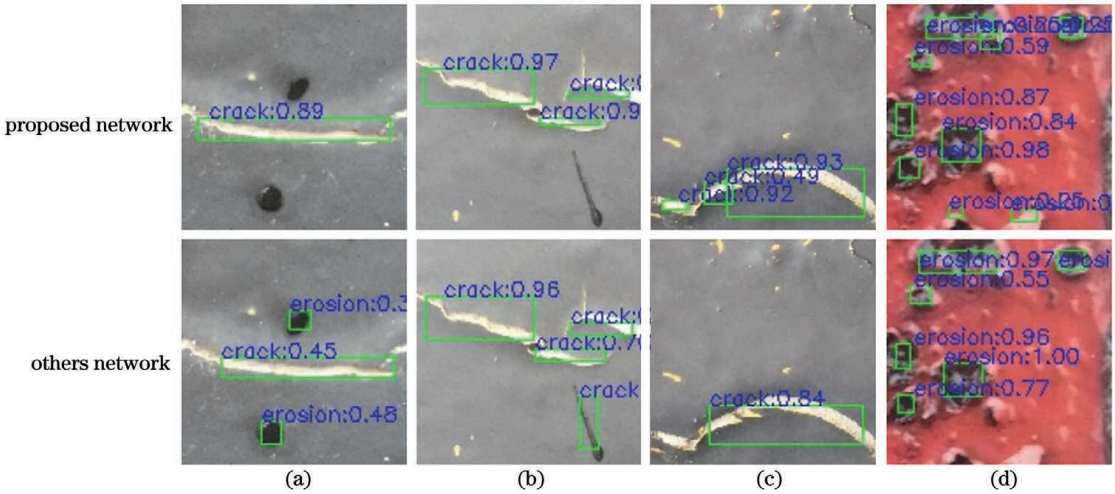

图 11　不同网络的检测结果。(a)image 1;(b)image 2;(c)image 3;(d) image 4

Fig. 11　Detection results of different networks. (a) image 1; (b) image 2; (c) image 3; (d) image 4

真实目标的预测置信度低于所提的网络；其他网络在 image 3 和 image 4 上分别发生裂纹和腐蚀的漏检，而所提网络在 image 3 和 image 4 上检测出三个裂纹和 9 个腐蚀，而其他网络只检测出一个裂纹和 6 个腐蚀。

图 12 为岸桥数据在 MobileNetV2SSDLiteV2 网络上的测试结果，分别为带状腐蚀、点状腐蚀和块状腐蚀。从图 12 可以看到，岸桥腐蚀的差异较大，尺度变化大，裂纹的宽度和形状不一，腐蚀的面积不同且每类目标的背景不同，而所提网络能够适应岸桥复杂的场景，对高中低三种岸桥腐蚀尺度均能检测，从而保证了模型部署的精度，并且可以保证模型部署的速度，最快可达到 45 frame/s。

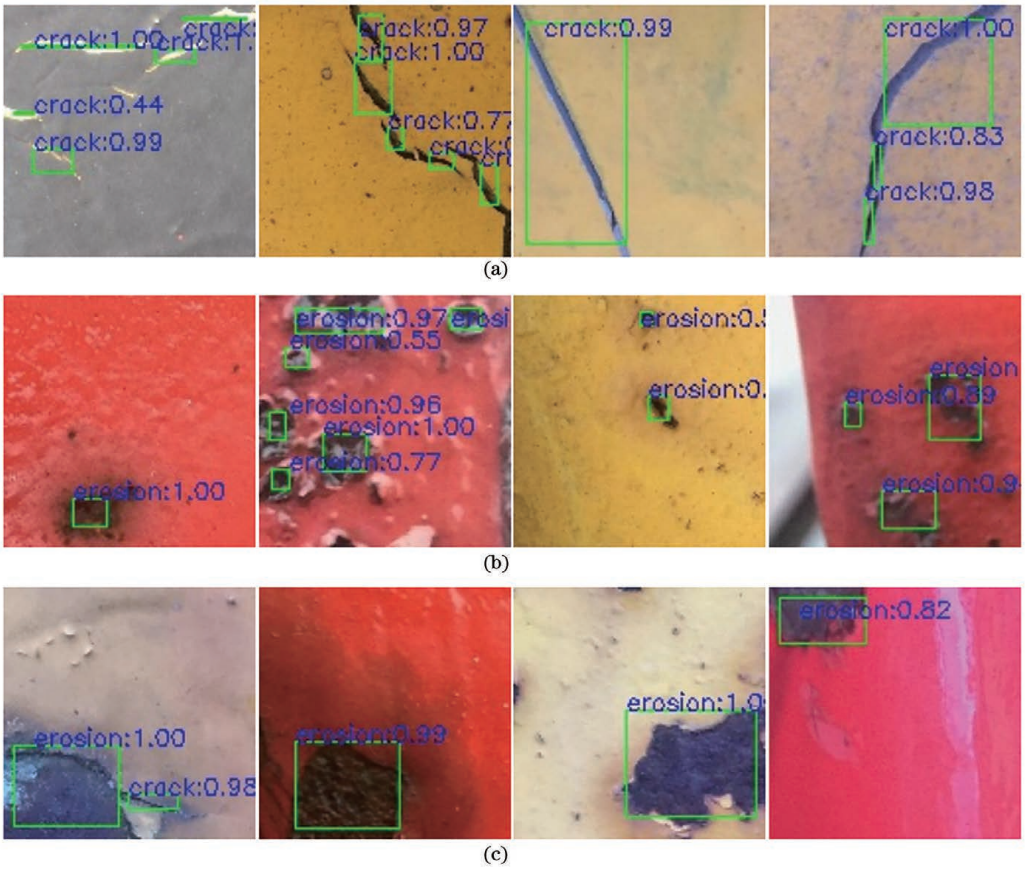

图 12　岸桥锈蚀的检测结果。(a)带状腐蚀；(b) 点状腐蚀；(c)块状腐蚀

Fig. 12　Detection results of corrosion of quay bridge. (a) Banded corrosion; (b) pitting corrosion; (c) block corrosion

5　结　　论

本文从三个方面改进 MobileNetV2SSDLite 网络。一是在网络的结构设计方面，对网络层进行修改，在骨干网络部分设计 14 个线性反残差瓶颈结构，将所有卷积层进行批处理归一化并从零开始训练网络，在检测器的改进部分设计 5 层多尺度特征映射并用于分类和定位。二是在数据的预处理方面，对岸桥图像进行多尺度前处理以实现多尺度训练，并使用数据增强来提升网络性能。三是使用不同的方法改进卷积层，如将原来网络的所有通道数减半，使用无偏置来训练网络。最后在实验部分进行网络性能分析并对比其他 4 种网络的性能，通过网络参数、浮点运算次数以及平均精度均值来衡量轻量级算法的综合性能，绘制每个轻量级算法的 P-R 曲线。实验结果显示改进后的网络相比于其他轻量网络速度更快，精度更高。

设计轻量化的神经网络是模型加速的一种方法，同时为了加快神经网络在硬件设备上的运行速度，在模型部署的过程中还会有很多的优化，如将模型定点化，即将 32 位单精度浮点型的模型量化成 8 位整型等，以及后续的优化工作。由于篇幅限制，本文未对其进行进一步的研究。此外，该网络在以下方面还有优化的空间：使用网络加速的方法，如网络的剪枝、低秩分解和知识蒸馏等。通过设计二值神经网络(BNN)的方法来加速推理；通过神经网络搜索(NAS)的方式来减少人为干预以设计更快更强壮的神经网络，并将其作为未来网络优化的研究方向。

参 考 文 献

[1] Koutník J, Greff K, Gomez F, et al. A clockwork RNN[EB/OL]. (2014-02-14)[2020-11-08]. https://arxiv.org/abs/1402.3511v1.

[2] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks [C] // Advances in Neural Information Processing Systems, December 8-13, 2014, Montreal, Quebec, Canada. New York: Curran Associates, 2014.

[3] Madhavan S, Kumar N. Incremental methods in face recognition: a survey [J]. Artificial Intelligence Review, 2021, 54(1): 253-303.

[4] Sun K, Li Q M, Li D Q. Face detection algorithm based on cascaded convolutional neural network[J]. Journal of Nanjing University of Science and Technology, 2018, 42(1): 40-47.
孙康, 李千目, 李德强. 基于级联卷积神经网络的人脸检测算法[J]. 南京理工大学学报, 2018, 42(1): 40-47.

[5] Hou Y L, Song Y Y, Hao X L, et al. Multispectral pedestrian detection based on deep convolutional neural networks[C]//2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), October 22-25, 2017, Xiamen, China. New York: IEEE Press, 2017: 17484506.

[6] Huang L L, Barth M. Vehicle detection vehicle detection, tightly coupled LIDAR and computer vision integration for[M]//Meyers R A. Encyclopedia of Sustainability Science and Technology. New York: Springer, 2012: 11455-11481.

[7] Diaz M, Cerri P, Pirlo G, et al. A survey on traffic light detection[M]//Murino V, Puppo E, Sona D, et al. New trends in image analysis and processing-ICIAP 2015 Workshops. Lecture notes in computer science. Cham: Springer, 2015, 9281: 201-208.

[8] Chen H Z, Jin Z L. Research on real-time lane line detection techonogy based on machine vision[C]// 2010 International Symposium on Intelligence Information Processing and Trusted Computing, October 28-29, 2010, Huanggang, China. New York: IEEE Press, 2010: 528-531.

[9] Huang G, Liu X L. Automatic extraction and classification of road markings based on deep learning [J]. Chinese Journal of Lasers, 2019, 46(8): 0804002.
黄刚, 刘先林. 基于深度学习的道路标线自动提取与分类方法[J]. 中国激光, 2019, 46(8): 0804002.

[10] Brust C A, Sickert S, Simon M, et al. Convolutional patch networks with spatial prior for road detection and urban scene understanding[C]//Proceedings of the 10th International Conference on Computer Vision Theory and Applications, March 11-14, 2015, Berlin, Germany. Setúbal: Science and and Technology Publications, 2015: 510-517.

[11] Yamaguchi T, Hashimoto S. Fast crack detection method for large-size concrete surface images using percolation-based image processing[J]. Machine Vision and Applications, 2010, 21(5): 797-809.

[12] Lad P, Pawar M. Evolution of railway track crack detection system[C]//2016 2nd IEEE International Symposium on Robotics and Manufacturing Automation (ROMA), September 25-27, 2016, Ipoh, Malaysia. New York: IEEE Press, 2016: 16657739.

[13] Yeum C M, Dyke S J. Vision-based automated crack detection for bridge inspection[J]. Computer-Aided Civil and Infrastructure Engineering, 2015, 30(10): 759-770.

[14] Wu Z H, Xiong W H, Ren J F, et al. Corrosion object detection of power equipment based on lightweight SSD[J]. Computer Systems & Applications, 2020, 29(2): 262-267.
吴之昊, 熊卫华, 任嘉锋, 等. 基于轻量级 SSD 的电力设备锈蚀目标检测[J]. 计算机系统应用, 2020, 29(2): 262-267.

[15] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.

[16] Chen L L, Zhang Z D, Peng L. Real-time detection based on improved single shot MultiBox detector[J]. Laser & Optoelectronics Progress, 2019, 56(1): 011002.
陈立里, 张正道, 彭力. 基于改进 SSD 的实时检测方法[J]. 激光与光电子学进展, 2019, 56(1): 011002.

[17] Li Q, Shi P. Ship hull corrosion detection based on vision technology[J]. Ship Science and Technology, 2020, 42(2): 190-192.
李强, 时鹏. 基于视觉技术的船舶船身锈蚀检测[J]. 舰船科学技术, 2020, 42(2): 190-192.

[18] Yuan J R, Xue B, Zhang W S, et al. RPN-FCN based Rust detection on power equipment [J]. Procedia Computer Science, 2019, 147: 349-353.

[19] Xie X L, Li C X, Yang X G, et al. Salient object detection algorithm based on dual-attention recurrent convolution[J]. Acta Optica Sinica, 2019, 39(9): 0915005.
谢学立, 李传祥, 杨小冈, 等. 双注意力循环卷积显

著性目标检测算法[J]. 光学学报, 2019, 39(9): 0915005.

[20] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residualsand linear bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.

[21] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2020-11-08]. https://arxiv.org/abs/1704.04861.

[22] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size[EB/OL]. (2016-11-04)[2020-11-08]. https://arxiv.org/abs/1602.07360.

[23] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6848-6856.

[24] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2020-11-08]. https://arxiv.org/abs/1804.02767.

[25] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.