

# 基于卷积注意力的输电线路防震锤检测识别

李飞<sup>1</sup>, 王超<sup>1</sup>, 浦东<sup>1</sup>, 陈瑞<sup>2</sup>, 张智坚<sup>1</sup>

(1. 南京工程学院 人工智能产业技术研究院, 南京 211167;  
(2. 南京工程学院 信息与通信工程学院, 南京 211167)

**摘要:** 为提高单目标多分类 (SSD, single shot multibox detector) 网络模型对输电线路防震锤的识别准确率, 提出一种融合卷积注意力机制和 SSD 模型相结合的新方法; 该算法采用残差网络 ResNet 作为骨干网络, 引入卷积注意力机制将通道和空间注意力结合, 通过压缩提取中间特征和利用权重系数更好地分辨出前景与背景, 提高对输电线路中防震锤检测的精度和速度; 训练时引入迁移学习策略, 克服了模型训练困难问题; 实验结果表明, 提出的算法不仅提高了检测准确率, 计算效率亦得到了提升; 与经典 SSD 算法相比, 输电线路的防震锤检测准确率提升了 2.5%, 检测速度达到了 12 fps 识别效果明显提升, 证明了新算法的有效性。

**关键词:** 目标检测; 防震锤; SSD; 注意力机制; 残差网络

## Convolutional Attention Mechanism Based Object Detection Method for Vibration Damper

LI Fei<sup>1</sup>, WANG Chao<sup>1</sup>, PU Dong<sup>1</sup>, CHEN Rui<sup>2</sup>, ZHANG ZhiJian<sup>1</sup>

(1. AI Industrial Technology Research Institute, Nanjing Institute of Technology, Nanjing 211167, China;  
2. Institute of Information and Communication, Nanjing Institute of Technology, Nanjing 211167, China)

**Abstract:** In order to improve the detection accuracy of Single Shot MultiBox Detector (SSD) model for vibration damper, an attention mechanism based the detection method is proposed. The ResNet is adopted by the backbone network instead of the VGG network, and the attention mechanism is introduced to improve the detection accuracy and speed for the vibration damper in transmission lines, the intermediate features are extracted through the compression, and the foreground and background is better distinguished by using the weight coefficient. The introduced fused convolutional attention mechanism combines the channel and the spatial attention, the performance jump is relatively obvious, and the computational efficiency is improved. A migration learning strategy is introduced to overcome the model training problem. The experimental results show that the introduced method improves the detection accuracy and efficiency, in comparsion to the SSD detection network model, and the fused convolutional attention mechanism improves the detection accuracy of the seismic hammer in transmission lines by 2.5%, and completes the vibration damper detection at 12fps. The recognition effect is significantly improved, which the new algorithm is effectively proved.

**Keywords:** object detection; vibration damper; SSD; convolutional attention mechanism; residual network

## 0 引言

随着我国国家电网发展规模的不断扩大, 电力传输线路途经的地理环境愈加复杂, 如沼泽、山脉、湖泊、盆地、水库等, 不仅在建设时存在困难, 而且加大了后期维护的难度。遭遇大风等恶劣天气时, 输电线将发生振动跳跃, 导致悬挂点处会反复弯折, 进而可能引发断线, 甚至倒塔事件。为了减少导线因外力因素引起的振动, 输电线路中普遍采用防震锤来吸收能量。但是由于风雨侵蚀、金属生锈等原因, 防震锤可能会出现断裂、滑移等故障, 将失去其原有的牵制力。因此及时检测防震锤发现它的故障并迅

速处理, 这对电网系统的稳定运行具有深远意义<sup>[1]</sup>。

无人机因为其体型小、灵活度高、反应迅速的特点已被广泛应用于输电线路的巡检中。传统方法对无人机拍摄图片进行目标物检测识别时, 主要对输电线路及其部件的颜色、形状以及边缘特征进行处理, 如 Haar 特征、线性反投影 (LBP, local binary pattern) 特征等。文献 [2] 提出了基于分块的 Haar 新特征以及基于区域的 LBP 新特征, 达到减小漏检率的目的。文献 [3] 结合直方图均衡化、形态学处理和 RGB 彩色模型, 实现锈蚀缺陷的检测。文献 [4] 通过提取防震锤的 Haar 特征, 并结合 AdaBoost 算法进行识别, 能够较好地从复杂背景中识别出防震锤。这些

收稿日期: 2021-08-17; 修回日期: 2021-09-09。

基金项目: 国家自然科学基金(61703201); 江苏省自然科学基金(BK20170765); 江苏省研究生科研创新计划(SJ CX21\_0945)。

作者简介: 李飞(1995-), 男, 江苏连云港人, 硕士研究生, 主要从事智能电网及信息技术方向的研究。

陈瑞(1972-), 女, 湖北武汉人, 教授, 硕士研究生导师, 主要从事视频信息处理和目标检测方向的研究。

引用格式: 李飞, 王超, 浦东, 等. 基于卷积注意力的输电线路防震锤检测识别[J]. 计算机测量与控制, 2022, 30(3): 48-53, 105.

传统的图像处理算法具备占用资源小等优点, 但都是针对某种特定的环境, 对于复杂的环境鲁棒性较差。

随着深度学习的快速发展, 基于深度学习卷积神经网络的输电线路部件检测已成为热点。文献 [5] 通过构建深度学习的网络模型提取图像特征, 并设置阈值, 判断原始图像中防震锤故障的概率。文献 [6] 提出结合 DeepLabV3 语义分割网络与防震锤的空间上下文关系对其进行识别与缺陷诊断。文献 [7] 采用 Faster RCNN 卷积神经网络算法对高重叠防震锤区域进行迭代合并, 构建防震锤识别模型。文献 [8] 为了实现高压输电线路部件的缺陷检测和故障诊断, 提出基于改进 YOLOv3 的高压输电线路关键部件目标检测算法。文献 [9] 使用单目标多分类检测器 (SSD, single shot multiBox detector<sup>[10]</sup>) 网络模型进行防震锤的检测, 已取得较好的效果, 但是当图片中存在大量被遮挡的防震锤时, 效果较差。

综合上述的防震锤故障检测算法来看, 深度学习检测目标的方法已经成为主流, 针对上述问题, 鉴于 ResNet<sup>[11]</sup> 残差网络结构比视觉几何组 (VGG, visual geometry group) 网络结构的运算高效且有更强的特征提取能力, 所以本文采用 ResNet-50 作为 SSD 的骨干网络, 同时在特征提取阶段引入文献 [12] 提出的注意力机制, 它结合了空间注意力和通道注意力机制, 能显著提高图像分类和目标检测的正确率。当该算法应用于防震锤实时识别时, 识别准确率能达到 81%, 同时召回率提升了 3.7%。

## 1 相关算法介绍

### 1.1 SSD 目标检测算法

SSD 算法同时借鉴了 YOLO<sup>[13]</sup> 和 Faster R-CNN<sup>[14]</sup> 网络的思想并结合两者的优点。SSD 基于前向传播的卷积神经网络 (CNN, convolutional neural network), 使用锚点框<sup>[15]</sup> 的思想: 原始图像经过卷积神经网络提取特征后, 由非极大值抑制 (NMS, non-maximum suppression) 算法处理, 可以直接回归目标的位置和类别。其网络结构如图 1 所示。

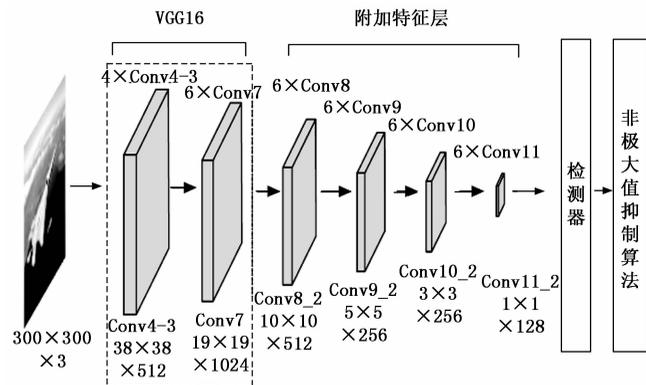


图 1 SSD 网络结构

从图 1 可以看出, SSD 由 VGG16 基础网络和附加特征层网络 (Extra Feature Layers) 两部分构成。卷积层 Conv4\_3、Conv7、Conv8\_2、Conv9\_2、Conv10\_2、Conv11

\_2 构成了特征金字塔结构, 可以在多个尺度上进行目标检测。SSD 将 VGG-16 的 FC6 和 FC7 全连接层替换成 Conv6 和 Conv7 层, 并添加 Conv8\_2、Conv9\_2、Conv10\_2、Conv11\_2 四个卷积层来获取锚点框。这 6 层的特征图相加, 共得到 8732 个锚点框, 如表 1 所示。

表 1 各卷积层锚点框个数

卷积层	特征图尺寸	卷积核个数	锚点框个数
Conv4_3	38×38	4	38×38×4=5 776
Conv7	19×19	6	19×19×6=2 166
Conv8_2	10×10	6	10×10×6=600
Conv9_2	5×5	6	5×5×6=150
Conv10_2	3×3	4	3×3×6=54
Conv11_2	1×1	4	1×1×6=6
			总计:8732

默认框 (default box) 生成及筛选:

SSD 算法的锚点 (anchor) 生成默认框 (default box) 的方法与 Faster R-CNN 算法基本相同。首先将输入的图像划分成 8×8 的网格图像, 设置的默认框会将图像中的目标全部包含进去。通过设置的  $k$  值可以计算默认框的大小, 计算方式如下:

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m - 1}(k - 1), k \in [1, m] \quad (1)$$

其中:  $S_{\min} = 0.2$  为 Conv4\_3 与原图尺度的比例大小,  $S_{\max} = 0.9$  为 Conv11\_2 与原图尺度的比例大小。设置的锚点框宽高比一般为  $a_r = 1, 2, 3, 1/2, 1/3$  从而求得默认框的宽和高, 公式如下:

$$h_k^a = \frac{S_k}{\sqrt{a_r}}, w_k^a = S_k \sqrt{a_r} \quad (2)$$

鉴于过多的默认框会增加计算成本, 而有的默认框中并无目标存在, 通常采用重叠度 (IOU, intersection over union<sup>[16]</sup>) 匹配策略对锚点框进行筛选。IOU 用来计算预测框与真实框的重合比率, 其计算公式如下:

$$IOU = \frac{S_{\text{预}} \cap S_{\text{真}}}{S_{\text{预}} \cup S_{\text{真}}} \quad (3)$$

其中:  $S_{\text{预}}$  代表的是默认锚点框,  $S_{\text{真}}$  代表的是真实框。IOU 匹配策略中, 阈值通常设置为 0.5。

### 1.2 注意力机制

注意力机制的本质就是一组注意力权重系数, 即在目标区域中提取对任务目标更有价值的信息, 同时抑制或忽略某些无关的细节信息。在深层网络中, 注意力机制可以帮助获取某些重要的目标特征, 即对输入图像有针对性主动提取特征中相关性较大的部分, 使更有利于网络模型训练的特征被学习。目前图像处理中最常用的注意力机制分为通道注意力 (Channel Attention)<sup>[17]</sup> 和空间注意力 (Spatial attention)<sup>[18]</sup> 两部分。

#### 1.2.1 通道注意力

通道注意力关注的是 “what” 的问题, 即关注的是这张图上哪些内容是有重要作用的。输入图像经过由卷积核

组成的卷积层得到特征矩阵，卷积核的个数决定了特征矩阵的通道数，但并不是每一个通道对于主要特征的提取都十分有用。通道注意力首先对输入的特征图  $F \in R^{C \times H \times W}$  进行最大值池化和平均池化，将两个特征图进行维度压缩，转发到多层感知器 (Multi-Layer Perceptron, MLP)，共享全连接，将两个特征图基于元素对应相乘的加和操作后，再经过 sigmoid 函数进行激活，得到含有权重的通道注意力特征图，其网络结构如图 2 所示。

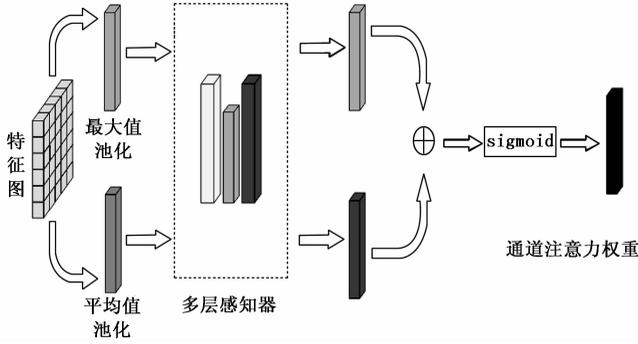


图 2 通道注意力机制

我们用  $C$  代表输入特征图的通道数， $H$  代表输入特征的长， $W$  代表输入特征图的宽， $F$  是输入的特征图， $F_{\max}^c$  和  $F_{\text{avg}}^c$  分别为经过最大值池化和平均值池化的特征图， $W_1$  和  $W_2$  为多层感知机的权重，通道注意力机制的计算方法如下：

$$F = \frac{1}{C} \sum_{i \in C} F(i) + \max_{i \in C} F(i) \quad (4)$$

$$M_c(F) = \sigma(W_1(W_2(F_{\text{avg}}^c)) + W_1(W_2(F_{\max}^c))) \quad (5)$$

### 1.2.2 空间注意力

空间注意力关注的是“where”的问题，即图像在整张图片的哪个位置，空间注意力是通道注意力的补充，通过空间特征的加权来有选择地聚合各个空间特征。空间注意力首先对输入不同的特征图  $F \in R^{C \times H \times W}$  的相同位置进行最大值池化和平均池化，再将特征图进行维度的压缩，然后将两个结果基于通道融合。本文通过一个卷积操作，将结果降维为 1 个通道。再经过 sigmoid 函数操作，得到含有权重的空间注意力特征图  $M_s(F) \in R^{1 \times H \times W}$ ，方法如下：

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\max}^s])) \quad (6)$$

其中： $\sigma$  为 sigmoid 操作。本文经过多次实验对比，选用比  $3 \times 3$  卷积核效果更好的  $7 \times 7$  卷积核。空间注意力的网络结构如图 3 所示。

## 2 融合卷积注意力机制的防震锤检测方法

### 2.1 整体网络架构

针对 SSD 中原来的特征提取网络对小目标定位能力差，且运算量大，检测效率较低的问题，本文采用 ResNet 残差网络结构代替 SSD 的 VGG 网络结构作为目标检测的骨干网络，并在该骨干网络中引入融合卷积注意力机制，通过

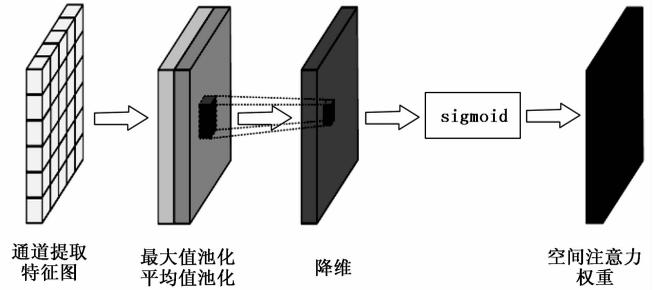


图 3 空间注意力机制

压缩提取中间特征，提高对物体检测的精度和速度。整体网络结构如图 4 所示。

### 2.2 特征提取网络

通常加深网络模型的层数，可以提高模型的学习能力，但是更深的网络模型使用随机梯度下降 (SGD, stochastic gradient descent) 优化算法会变得更困难，同时会出现梯度消失的情况。为了解决这个问题，文献 [11] 提出了残差网络 ResNet，它在速度和精度上都优于 VGG，它的核心是通过建立前面层与后面层之间的“短路连接” (skip connection)，有助于训练过程中梯度的反向传播，能训练出更深的网络。

无人机拍摄输电线路部件图像的背景往往十分复杂，可能有些部件粘连在一起，导致部件无法呈现它原本的特征。而浅层网络 VGG 对特征抓取能力不够强，ResNet-50 残差网络中包含了 50 个 Conv 的操作，能更好地提取图像特征，其结构如表 2 所示。

表 2 ResNet-50 网络结构

层数名	输出尺寸	50-layer
Conv_1	$112 \times 112$	$7 \times 7, 64$ , 步长
Conv2_x	$56 \times 56$	$3 \times 3$ , 最大值池化, 步长 2 $\begin{bmatrix} 7 \times 7, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3_x	$28 \times 28$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4_x	$14 \times 14$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Conv5_x	$7 \times 7$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	$1 \times 1$	Average pool, 1 000-d fc, softmax

输入图像在经过卷积和池化后进入第一层残差块中，在接下来的每一个阶段都要进行经过一次卷积和归一化的操作，残差结构块如图 5 所示。

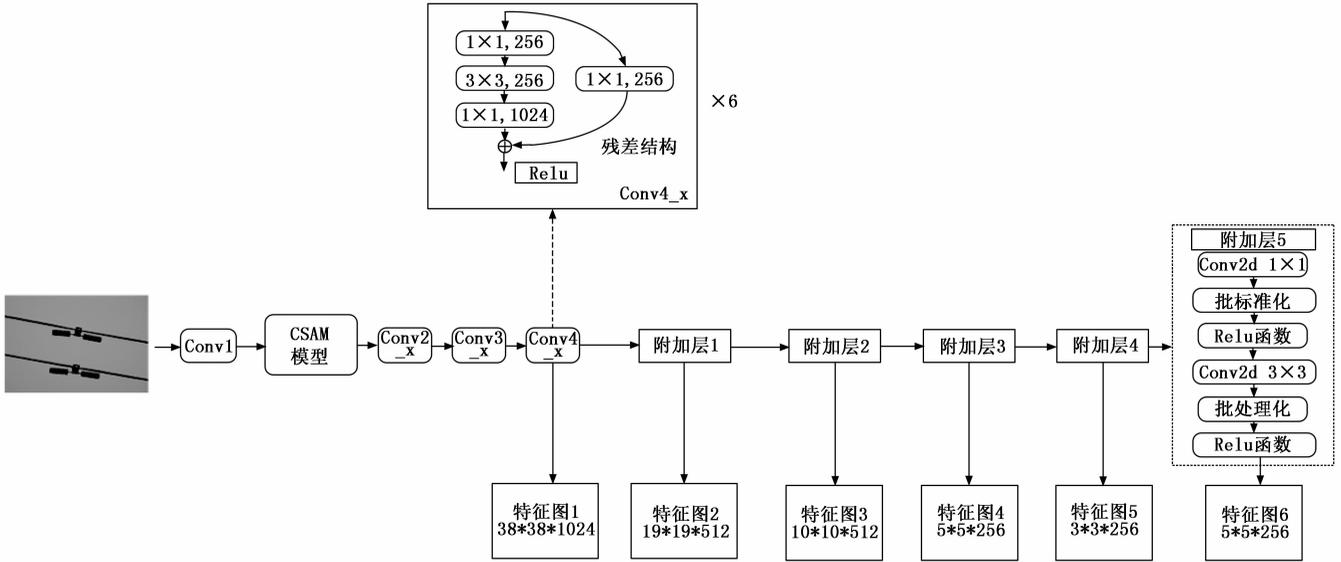


图 4 整体网络结构图

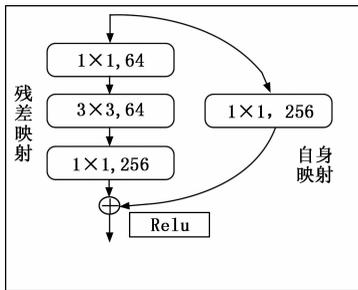


图 5 残差结构块

由图 5 可以看出, 残差学习块主要基于自身映射 (identity mapping) 和残差映射 (residual mapping) 两个映射, 输出是卷积计算部分加上自身映射, relu 函数再次激活。VGG 网络用于提取小目标的特征层在传递信息时, 总会存在特征信息丢失问题, 但 ResNet 残差结构能直接将输入信息直接传递到输出, 可以解决信息丢失问题, 降低学习目标难度, 且 ResNet 模型的参数量比 VGG 网络显著减少, 如表 3 所示。

表 3 VGG 与 ResNet 计算过程中参数量对比

模型	参数量	计算量(GFLOPS)
ResNet	23.518 3 M	3.8
VGG	138.357 6 M	14.2

表 3 中可以看出, VGG 前向传播一次需要 14.2 亿次浮点数据计算, 而 ResNet 仅需要 3.8 亿浮点数据计算, 运算量约减少了 4 倍左右, 且 ResNet 网络模型占用空间更小, 有利于更多图片进行一次性批量训练。

### 2.3 卷积注意力机制模块

为了使图像的有益信息在整个网络框架中传递, 抓取到防震锤的关键特征, 加强关键信息对整张图片的增益,

提高对防震锤的检测能力。本文在残差网络 ResNet-50 的卷积块中引入卷积注意力机制, 结合通道和空间注意力, 记为通道空间注意力模型 (CSAM, channel space attention module), 如图 6 所示。

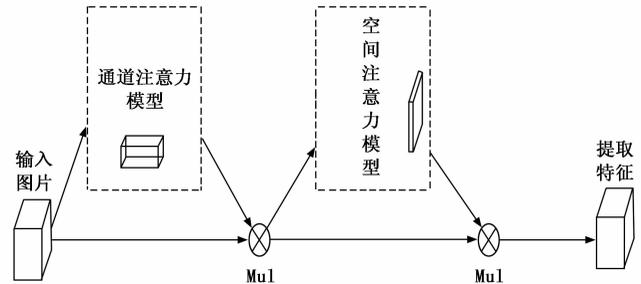


图 6 通道空间注意力模块 (CSAM)

输入的中间特征图为  $I \in \mathbf{R}^{C \times H \times W}$ , 该特征图首先输入通道注意力机制模块, 获得含有权重的通道注意力特征图  $M_c \in \mathbf{R}^{C \times 1 \times 1}$ , 图中 Mu1 代表对应矩阵元素相乘, 将含有权重的通道注意力特征图与原特征图相乘, 获得中间特征图仍然为  $I \in \mathbf{R}^{C \times H \times W}$ , 然后将此中间特征图输入空间注意力机制模块, 获得含有权重的空间注意力特征图  $M_s \in \mathbf{R}^{C \times 1 \times 1}$ , 将获得含有权重的空间注意力特征图与上一层特征图相乘, 得到同时获得空间注意力与通道注意力的特征图。

上述过程也就是将通道注意力与空间注意力相结合, 实现层级间信息的最大化, 引导模型在迭代时获得更显著的防震锤检测网络模型。

CSAM 模块对 ResNet-50 网络可能产生不利影响, 导致预训练模型参数不能匹配新的网络模型, 因此 CSAM 不能直接加在 ResNet-50 网络内部。本文将 CSAM 模块置于 Conv\_1 (7×7) 卷积层之后, 即原始输入 300×300 的彩色 RGB 图像, 通过一个 7×7 卷积层后的特征图作为 CSAM 模块的初始特征图输入。由 CSAM 模块找出特征图

中任意位置之间的空间依赖,对所有位置上的特征加权和更新。相应两个位置之间的特征相似性作为权重用来提升获取主要特征的能力,而不需要关注于它们之间的距离。相较于 SENet<sup>[19]</sup>,CSAM 模块通道注意力中加入的全局最大化在一定程度上弥补了平均值池化 AvgPool 丢失的信息,且在 CSAM 模块空间中,生成的二维空间注意力特征图使用卷积核大小为 7 的卷积层进行编码,避免了选用较小卷积核只关注局部特征的缺点,对保留重要的空间信息很有帮助。同时,本文在 Conv4\_x 的第一个 block 中将 stride 设置为 1,让通过 Conv4\_x 之后的特征图尺寸缩小一半,这样 Conv3\_x 输出特征图尺寸为 512×38×38,Conv4\_x 输出为 1 024×38×38,而不是原来 ResNet 网络中的 1 024×19×19。本文提出的网络在 ResNet 残差结构之后增添了 5 个层后,再加上 ResNet 的 Conv4\_x 的特征图,一共提取 6 张特征图,这些选出的各个层的特征图将被用于种类和位置的预测,如表 4 所示。

表 4 融合卷积注意力及残差网络结构表

层数名	50-layer	输出尺寸
Conv_1	7×7,64,步长	150×150
CSAM 模块		
Conv2_x	3×3,最大池化,步长 2 $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	75×75
Conv3_x	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	38×38
Conv4_x	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	38×38 (特征图 1)
额外添加模块		
层数名	layer	输出尺寸
特征图 2	1×1;BN;3×3,	512×19×19
特征图 3	1×1;BN;3×3,	512×10×10
特征图 4	1×1;BN;3×3,	512×5×5
特征图 5	1×1;BN;3×3,	512×3×3
特征图 6	1×1;BN;3×3,	512×1×1

### 3 实验结果与分析

#### 3.1 实验平台

本文实验所用的软、硬件平台参数配置如表 5 所示。

表 5 实验平台参数设置

软硬件平台	型号参数
操作系统	Ubuntu18.04
CPU	Intel(R)Core(TM) i7-7800X@ 3.50GHz
GPU	GTX1080Ti
显存	12 GB
框架	Pytorch
编程环境	Python

#### 3.2 数据集与实验参数

##### 3.2.1 数据集

本实验采用标准数据集与自建数据集结合的方式来测试网络的性能。标准数据集采用 Pascal VOC2007 和 COCO2017。Pascal VOC2007 训练集共 5 011 幅,测试集共 4 952 幅,数据集包含了 20 个分类且光照、拍摄角度等因素各不相同。COCO2017 数据集是一个大型的、丰富物体检测数据集,由复杂的日常景物截图组成,共标注了含背景在内 81 类经过精确分割进行位置标定的目标。训练集 118 287 张图片,验证集 5 000 张图片,测试集 40 670 张图片。标准数据集可以作为衡量图像分类识别能力的基准,对模型的评判具有一定的意义。

自建数据集由某市供电公司提供,共 8 295 张无人机拍摄的输电线路高清原始图,将其中的 7 465 张图片作为测试集,830 张图片作为验证集,其中共包含防震锤的数量为 11 876 个。

##### 3.2.2 实验参数

在进行训练时,所有的批处理(batchsize)均设置为 64,初始学习率设置为 10<sup>-3</sup>,动量参数设置为 0.9,权重衰减为 2×10<sup>-4</sup>。

对于 VOC2007 数据集,共训练 40 000 次,算法分别在 2 640 次和 3 230 次学习率衰减。对比改进前后算法的收敛性,改进后的 Resnet 结构算法收敛性要略优于原 VGG 结构算法,同时两种算法针对 VOC2007 数据集在 26 000 次左右均已经趋于收敛,如图 7 所示。

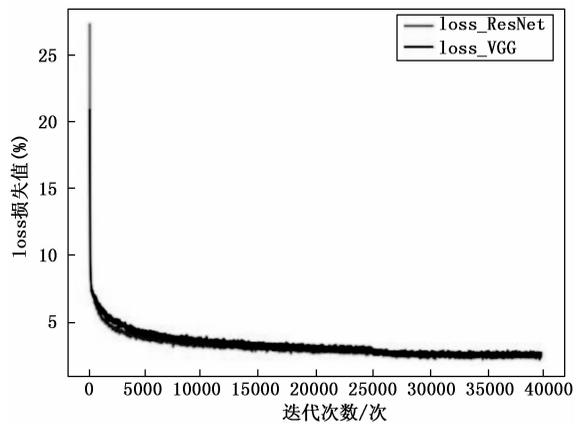


图 7 VOC2007 数据集上的损失曲线对比

对于 COCO 数据集,共训练 440,000 次,以便得到对该数据集的较好拟合和较佳预测。

#### 3.3 实验结果及分析

首先在 VOC2007 数据集上进行测试,其中 mAP@0.5 为网络模型在 IOU 阈值大于 0.5 时,多类预测时每一类的精度(precision)取平均值。通过表 6 对比,可以发现采用残差网络作为骨干网络并增加注意力机制 CSAM 模块,平均精度提升了 1.5% 左右,其中鸟类、羊类等特征明显的有显著提升,约为 5% 左右。测试效果如图 8 所示。

表 6 PASCAL VOC2007 test 数据集上的检测结果

算法	mAP 精度	飞机	自行车	鸟	船	瓶子	巴士	
SSD <sup>[10]</sup>	0.755 6	0.803 4	0.834 6	0.741 0	0.676 9	0.458 5	0.822 9	
本文算法	0.770 7	0.811 1	0.855 3	0.791 1	0.694 2	0.475 0	0.844 2	
		汽车	猫	椅子	奶牛	餐桌	狗	马
SSD <sup>[10]</sup>	0.840 9	0.877 2	0.585 9	0.810 4	0.757 0	0.860 4	0.852 3	
本文算法	0.846 8	0.878 6	0.589 7	0.817 3	0.737 2	0.870 0	0.865 0	
		摩托车	人	盆栽植物	羊	沙发	火车	电视机
SSD <sup>[10]</sup>	0.821 8	0.772 4	0.478 1	0.731 6	0.790 3	0.851 3	0.745 2	
本文算法	0.833 9	0.762 2	0.506 5	0.789 4	0.795 4	0.876 1	0.774 7	



图 8 本文算法在 VOC2007 数据集复杂场景的检测效果

由于算法轻量化提升, 整个算法模型的正向推理时间也有所改善, 检测一张图片时速度约提升了 25 ms, 如表 7 所示。

表 7 算法检测时间对比

模型	推理时间/s
SSD <sup>[10]</sup>	0.107
本文算法	0.082

在 COCO2017 数据集上进行测试, 用以评估算法有效性。从表 8 中可以见到, 本文算法 (SSD + ResNet + CSAM) 相对于几种典型算法在检测精度上均有所提高, 效果见图 9。

表 8 各算法精度对比

模型	骨干结构	AP	AP50	AP75	APS	APM	APL
YOLOv2 <sup>[13]</sup>	DarkNet-19	21.6	44.0	19.2	5.0	22.4	35.5
SSD300 <sup>[10]</sup>	VGG16	23.3	41.2	23.4	5.3	23.2	39.6
SSD512 <sup>[10]</sup>	VGG16	26.8	46.5	27.8	9.0	28.9	41.9
本文算法	ResNet	28.6	48	29.6	9.1	31.1	47.2

在验证算法的有效性后, 将含有防震锤的图片进行训练测试, 在训练时采用了迁移学习<sup>[20]</sup>策略, 可以加快防震锤的网络训练, 优化模型的学习效率, 在相同的时间内能训练出精度更高的网络。除了采用在标准数据集测试算法



图 9 本文算法在 COCO2017 数据集复杂场景的检测效果

模型中的 mAP 平均准确率评价指标外, 还使用了国网运检部规定的 recall 召回率来衡量算法模型找出的防震锤相对总数占比, 得到的结果如表 9 所示。

表 9 算法改进前后防震锤检测结果

模型	mAP	召回率
SSD <sup>[10]</sup>	0.785	0.783
本文算法	0.81	0.82

由表 9 的对比可以看出, 在算法引入融合注意力模块和残差结构块后, 识别防震锤的平均准确率达到 81%, 相较于原来提升了 2.5%。在面对输电线路复杂背景、光线多变等不良因素影响下, 能够减少图像信息丢失, 提取到图像更深层特征信息以提高识别精度, 识别效果明显提升。同时提取特征图经过融合预测, 并结合已有的预训练的卷积神经网络权重, 查全率提升了约 3.7%, 有效解决了因监视点距离防震锤较远造成的目标过小而引起的漏检问题。选取部分不同背景下无人机拍摄的包含防震锤的输电线路图片, 图像实际共拍摄到 20 个防震锤, 原 SSD 算法<sup>[10]</sup>识别到 13 个防震锤目标而本文算法共识别到 17 个防震锤目标, 检测结果对比见图 10。

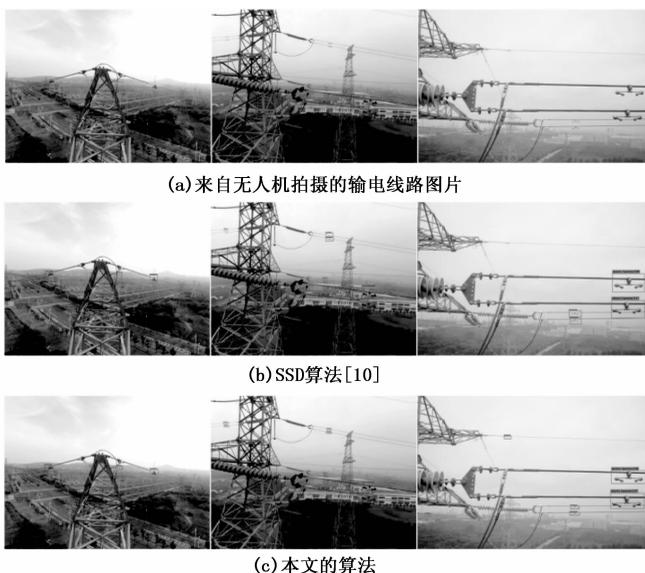


图 10 防震锤检测识别效果

(下转第 105 页)