

文章编号: 2095-2163(2020)08-0026-07

中图分类号: TP-751

文献标志码: A

# 基于改进 SSD 算法的濒危动物检测

蒋飞, 王霄

(贵州大学 电气工程学院, 贵阳 550025)

**摘要:** 针对在野生濒危动物保护中,存在着对保护动物监测难度大、准确率低等问题。本文利用 PyTorch 框架,以 VGG16 为基础的 SSD 算法,提出了一种对濒危物种的动物识别与检测方法。在基础网络中,加入了注意力机制,并且在训练过程中使用迁移学习和数据扩增方法,提升了检测速度与精度。通过对数据集的测试,此方法不仅准确率达到了 96.6%,而且其检测速度与模型大小符合预期,适用于对濒危动物的检测识别。

**关键词:** 濒危动物检测; 迁移学习; SSD 算法; 深度学习

## Endangered animal detection based on improved SSD algorithm

JIANG Fei, WANG Xiao

(School of Electrical Engineering, Guizhou University, Guiyang 550000, China)

**[Abstract]** In the protection of wild endangered animals, there are problems such as difficulty in monitoring protected animals and low accuracy. Based on the PyTorch framework, a method for identifying and detecting endangered species of animals based on the VGG16 SSD algorithm is proposed. In the basic network, an attention mechanism is added, and migration learning and data amplification methods are used in the training process to improve the detection speed and accuracy. The test on the data set shows that the accuracy of this method reaches 96.6%. And its detection speed and model size are in line with expectations, which is suitable for detection and identification of endangered animals.

**[Key words]** endangered animal detection; migration learning; SSD algorithm; deep learning

### 0 引言

随着人类活动空间的持续扩大以及对野生动物的捕获,生物的种类正在以每日 160 种的速度不停的消失。目前,有 5%~20%的树木物种和脊椎动物正面临灭绝的危险,而且灭绝的速度正以上百倍的速度增长。有 257 种野生动物被列入《国家重点保护野生动物名录》,其中有 120 多种野生动物被列入《濒危野生动植物种国际贸易公约》<sup>[1]</sup>。因此,对动物的保护特别是濒危动物的保护便显得尤为重要,而对于濒危物种的识别与检测便成为一个热门且存在挑战的研究方向。生态学领域的动物图像检测与识别一直是模式识别领域的一个重要课题。Slavomir Matuska 等人<sup>[2]</sup>使用混合局部描述算子与 SVM<sup>[3]</sup>的传统方法,对狐狸、鹿、野猪、棕熊、狼等 5 种动物进行分类;Ramanan 等人采用几何特征拼接与纹理特征,实现了对视频中的长颈鹿的识别<sup>[4]</sup>;张公伯等利用简单矩形特征训练 Adaboost 分类器,实现了对视频中狮子的检测与跟踪<sup>[5]</sup>。这种基于

动物的检测识别都是采用图像特征识别的方法,而特征的提取均采用人工提取的方式获得。这些方法对于场景简单,物种种类较少的情况来说,可以获得较好的结果。但在自然界中,背景环境复杂,这种运用手工的特征设计与提取的方法,通过训练后便存在鲁棒性低、泛化能力差等问题。

近年来,深度学习在语音识别、文字处理,特别是图像检测与识别中获得了突飞猛进的发展,并且逐步应用于人们的日常生活中。卷积神经网络模型通过对图像深度特征的自动提取,在大规模图像识别任务中表现出色。2014 年,Chen 等人公布了第一篇使用 CNN(Convolutional Neural Networks)进行动物物种分类的论文<sup>[6]</sup>。其在 RCT 数据集上使用了 3 个卷积层和 3 个池化层的标准 CNN 网络,但准确率只有 38%,存在很大的改进空间。2016 年 Gomez 等人使用 SS(the Snapshot Serengeti)数据集上训练物种分类<sup>[7]</sup>,对 AlexNet、VGG<sup>[8]</sup>、GoogleNet<sup>[9]</sup>和 ResNet 等 8 种不同的 CNN 框架进行比较,其最好的

**基金项目:** 国家自然科学基金(61861007,61640014);贵州省工业攻关项目(黔科合支撑[2019]2152);黔科合人才团队((2015)4014);物联网理论与应用案例库(KCALK201708);自动化专业卓越工程师计划(ZYS 2015004)。

**作者简介:** 蒋飞(1995-),男,硕士研究生,主要研究方向:图像识别;王霄(1985-),男,博士,副教授,硕士生导师,主要研究方向:无线传感器网络、嵌入式系统。

**通讯作者:** 王霄 Email: 76186684@qq.com

**收稿日期:** 2020-06-15

表现只有 57%。2017 年 Norouzzadeh 等人在 SS 数据集上的分类结果达到了 92%<sup>[10]</sup>, 相比上述结果, 其准确率得到了大幅度的提升。

但是图像的识别分类并不能满足现实的需求, 图像分类只产生出一个结果, 一幅图中有何种动物。在现实环境下, 一幅图片中可能存在多个物种的动物, 并且需要把每一种动物在背景环境特别复杂的情况下准确的识别出来。因此, 本文设计的网络主要由以下部分构成:

(1) 以 VGG 为基础网络构建的 SSD 目标检测模型。

(2) 在 VGG 基础网络中嵌入注意力机制。

(3) 数据集图片的扩增。

(4) 利用迁移学习训练图片。

## 1 SSD 基础网络

SSD<sup>[11]</sup> 是一种标准的目标检测算法, 主要由两部分构成, 前面部分是传统的图像分类卷积标准体系结构。例如, VGG 结构、ResNet 结构。在任何分类层之前均可以被截断, 将其称为基础网络。基础网络用于获取图片中的相对应特征。后面的网络同样是由卷积层构成, 用于提取更加深层次的特征, 称之为辅助结构。其组合结构如图 1 所示。

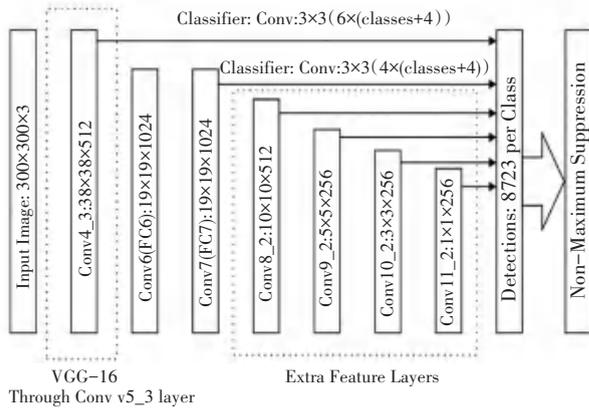


图 1 SSD 原始结构图

Fig. 1 SSD original structure diagram

基础结构与辅助结构相结合, 会生成大量的边界框, 大多数的边界框中并没有包含相对应的物体在里面。在后续操作中, 将对这些边界框中的物体进行分类评分以及定位评分。运用非极大抑制的方法选择最优的边界框和分类, 得到最终的和定位结果。

SSD 使用的是一种多尺度滑动窗口检测原理, 窗口在图像上划过, 并在每个位置识别该窗口是否包含任何感兴趣的对象。大窗口对应图像中大型的物体, 小窗口对应图像中比较小的物体, 因此可以对多个尺度和类型的目标进行检测。

对于网络顶部的多个卷积层和池化层, 获取到重要的要素特征图, 在每一个像素点平铺上不同大小的边界框, 通过预测每一个边界框对于真实边界宽的位置来进行学习。具体来说, 对于给定位置  $k$  中的每个边界框, 计算  $c$  种类别的得分和相对于原始默认框形状的 4 个偏移量。这将导致总共  $(c+4)k$  个滤镜应用在特征图中的每个位置周围, 从而生成  $m \times n$  特征图的  $(c+4)kmn$  输出。在几个特征图中使用不同的默认框的大小, 基本能将图片中所有需要检测的物体包含进去。在后续的操作中, 去掉多余的边界框, 可以离散的在图片上输出所有有用的空间。

## 2 CBAM 注意力机制

在图像识别中, 对通道上和空间上的处理重视程度一致。对于一些特征含量较少的通道和空间, 这种处理方式将导致计算资源的浪费。因此, 需要根据每个通道和空间中特征的多少, 计算资源的分配。可以保留更多的目标特征, 抑制无关信息, 提高检测精度。

近几年, 注意力机制在深度学习的各个领域被广泛使用, 该机制模拟人类的视觉机制, 将关注重点放在图像中辅助判断的部分信息, 并忽略掉不相关的信息。CBAM (Convolutional Block Attention Module) 注意力机制是 Sanghyun Woo 等人在 2018 年提出的一种用于前馈卷积神经网络的简单而有效的注意力模块<sup>[12]</sup>。注意力机制由人观察一个场景的灵感而来。人们在观察一个场景时, 不会把场景中的每一个细节都分毫不差地记忆下来, 但总会先关注到比较突出或者事先想要关注的部分。Hu 等介绍了一种利用通道之间关系的模块。在“挤压和激励”模块中, 使用全局平均池化功能, 来计算通道注意力。但是, 这些是次优功能, 错过了对空间的关注, 而空间的关注在“聚焦”确定位置中起着重要作用。因此在 SeNET (Squeeze - and - Excitation Networks)<sup>[13]</sup>基础上进行改进, 形成了 CBAM 注意力机制模块。

如图 2 所示, CBAM 注意力模块由通道注意力模块和空间注意力模块结合构成。给定一个中间特征图  $F \in R^{C \times H \times W}$  作为输入, CBAM 依次计算出一维通道注意力图  $M_c \in R^{C \times 1 \times 1}$  和二维空间注意力图  $M_s \in R^{1 \times H \times W}$ , 其过程可以概括为:

$$F' = M_c(F) \otimes F,$$

$$F'' = M_s(F') \otimes F'.$$

其中,  $\otimes$  表示逐元素相乘,  $F''$  是最终的输出。图 2 描绘了每个注意力图的计算过程。

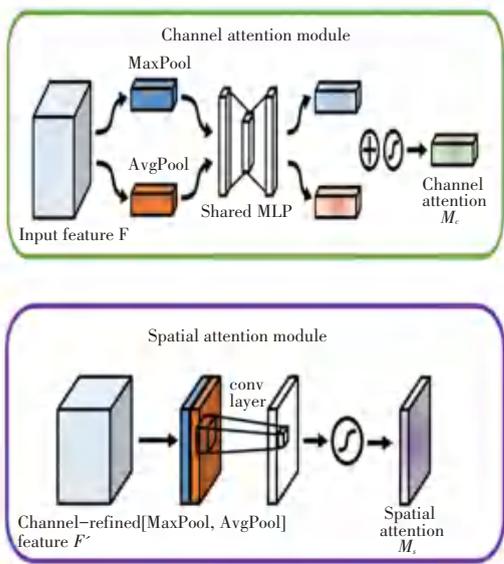


图 2 通道和空间注意力子模块

Fig. 2 Channel and spatial attention sub-module

### 3 改进的 SSD 注意力网络

为了提高模型的检测精度,在图 1 所示的 SSD 算法中引入注意力机制,引入注意力机制后的 SSD 网络结构如图 3 所示。

由图 3 可见,通过在 conv3->conv4 中、conv7->conv8->conv9->conv10->conv11 之间,加入注意力机制,使每一层的权重系数从平均权重值变成了按重要程度分布的权重值。权重系数的调整,一方面能够提升对动物重要特征提取的效率,另一方面,在后续的动物 loss 回归进行精确定位时,可以提高 loss 计算的准确率。

#### 3.1 默认框设计

在 SSD 网络结构中,为了识别特征图中的图像,将一组默认边界框与每个要素特征单元相关联,以用于基础网络的多个特征图。默认框(Default box)作用在每个映射单元上,生成一系列不同长宽比的默认框,之后计算默认框与真实框之间的偏移值,以及每个默认框中包含物体的分数(score)。分数的计算过程为:对于图片中每个真实物体,找到与其 IOU(模型默认框和真实框之间的交并比)最大的默认框。默认框匹配步骤如图 4 所示,在同一个栅格(cell)中,SSD 会使用高宽比不同的默认框进行匹配(高宽比可由式(5)计算)。匹配到物体之后,再进行分类和回归操作。这将导致总共  $(c + 4)k$  个滤镜,应用在特征图中的每个位置周围,从而产生  $m \times n$  特征图的  $(c + 4)kmn$  输出。将这些输出通过非极大抑制算法,选择出最优的方案。

对于不同特征层生成的默认框尺度是不同的。默认框尺度与原图的比例计算如公式(4):

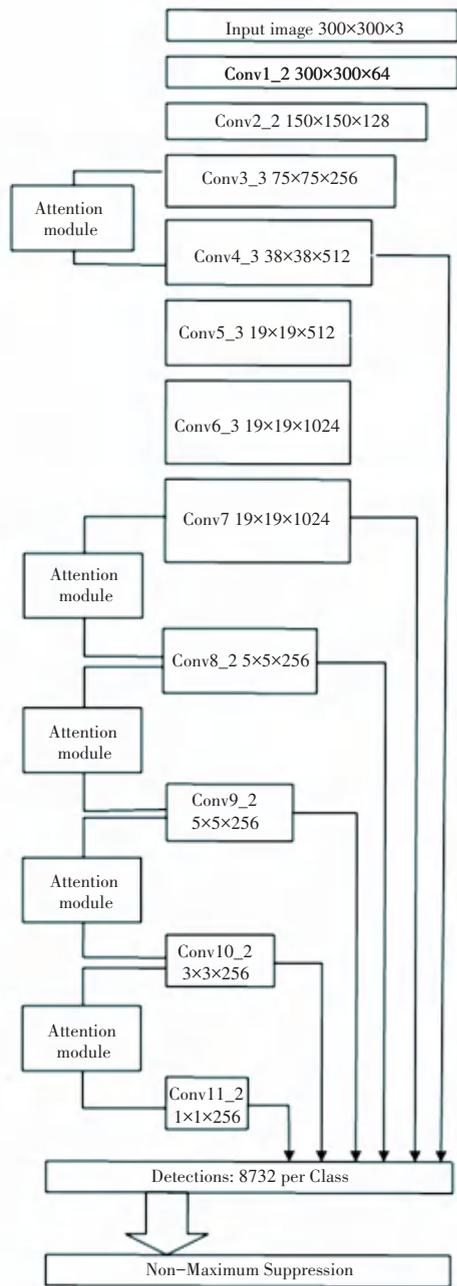
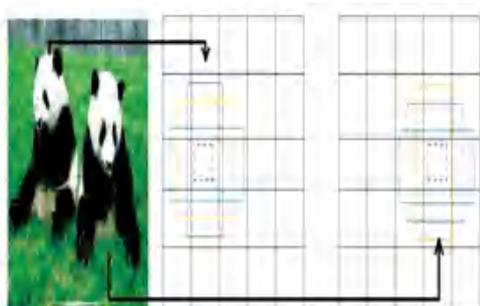


图 3 引入注意力机制的 SSD 网络结构图

Fig. 3 SSD network structure diagram with the introduction of attention mechanism



(a) Image with GT boxes (b) Feature map with 5x5 (c) Feature map with 5x5

图 4 默认框匹配步骤

Fig. 4 Default box matching steps

$$s_k = s_{min} + \frac{s_{max} - s_{min}}{m - 1}(k - 1), \quad k \in [1, m]. \quad (4)$$

其中,  $m$  指特征图个数, 且第一个预测特征图  
的先验框尺度比例为单独设置, 取值  $\frac{s_{min}}{2} = 0.1$ 。  $s_{min}$   
指最小尺度比例, 取值为 0.2;  $s_{max}$  指最大尺度比例,  
取值 0.95。先验框的高度与宽度计算公式为:

$$w_k^\alpha = S_k \sqrt{\alpha_r}, \quad h_k^\alpha = S_k / \sqrt{\alpha_r}. \quad (5)$$

其中,  $w_k^\alpha$  为默认框的宽度,  $h_k^\alpha$  为默认框的高  
度,  $s_k$  为默认框的尺度,  $\alpha_r \in \{1, 2, 3, \frac{1}{2}, \frac{1}{3}\}$ 。

最后一个默认框尺度由公式  $S_k' = \sqrt{S_k S_{k+1}}$  构  
成, 设置长宽比为  $\alpha_r = 1$ , 计算得出。其中,  $s_k$ 、 $S_{k+1}$   
分别为当前层与之后一层长宽比,  $\alpha_r = 1$  的小正方  
形默认框尺度。根据公式(4)、(5)计算得到的各特  
征层默认框尺寸, 见表 1。

表 1 各层默认框尺寸

Tab. 1 Default frame size of each layer

图层编号	图层宽和高	默认框尺寸	默认框数量
1	38x38	21   1/2, 1, 2   ; $\sqrt{21 \times 45}$ ;   1	38x38x4
2	19x19	45   1/3, 1/2, 1, 2, 3   ; $\sqrt{45 \times 99}$ ;   1	19x19x6
3	10x10	99   1/3, 1/2, 1, 2, 3   ; $\sqrt{99 \times 153}$ ;   1	10x10x6
4	5x5	153   1/3, 1/2, 1, 2, 3   ; $\sqrt{153 \times 207}$ ;   1	5x5x6
5	3x3	207   1/2, 1, 2   ; $\sqrt{207 \times 261}$ ;   1	3x3x4
6	1x1	261   1/2, 1, 2   ; $\sqrt{261 \times 315}$ ;   1	1x1x4

### 3.2 回归定位

利用 SSD 算法进行回归预测时, 用到了二个指  
标: 一是对默认框进行回归操作时的坐标误差损失  
 $L_{loc}$ ; 第二个对物体的种类进行预测时的分类误差损  
失  $L_{conf}$ 。两个指标相结合, 用来评估网络对物体识  
别定位的适用性大小, 如式 6 所示。

$$L_{(x,c,l,g)} = \frac{1}{N}(L_{conf}(x,c) + \alpha L_{loc}(x,l,g)). \quad (6)$$

其中,  $N$  是默认框的个数;  $\alpha$  是坐标误差的权  
重参数, 默认为 1。通过更改  $\alpha$ , 可以改变坐标误差  
在梯度下降过程中的重要程度。

分类置信度损失函数  $L_{conf}$ 。在本实验中使用的  
是 softmax loss。其中包含两部分, 一个是对需要检  
测物体的置信度损失, 另一个是不需要检测物体的  
置信度损失, 如式 7 所示。

$$L_{conf}(x,c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0),$$

$$\text{where } \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}. \quad (7)$$

在(7)式中,  $x_{ij}^p$  表示第  $i$  个默认框与类别  $p$  的第  
 $j$  个真实框匹配时取 1, 否则取 0;  $c_i^p$  为第  $i$  个默认框  
属于类别  $p$  的预测置信度分数,  $\hat{c}_i^p$  为第  $i$  个默认框属

于类别  $p$  的预测置信度概率<sup>[18]</sup>, 需要满足  $\sum_1^i \hat{c}_i^p = 1$ 。  
SSD 模型坐标误差损失  $L_{loc}$  的计算如式(8)所示:

$$L_{loc}(x,l,g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^{k, Smooth_{L1}}(l_i^m - \hat{g}_j^m) \cdot \quad (8)$$

$$\hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx}) / d_i^w,$$

$$\hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy}) / d_i^h,$$

$$\hat{g}_j^{cx} = \log(g_j^{cx} - d_i^{cx}) / d_i^w,$$

$$\hat{g}_j^{cy} = \log(g_j^{cy} - d_i^{cy}) / d_i^h,$$

$$Smooth_{L1} = \begin{cases} 0.5x^2 & |x| < 1.5 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

其中,  $l_i^m$  为第  $i$  个预测框  $m$  参数的相对偏移量;  
 $\hat{g}_j^m$  为第  $j$  个真实框  $m$  参数相对于默认框的偏移量;  
 $\hat{g}_j^m$ 、 $d_i^m$  分别对应真实框和默认框的  $m$  参数;  $Smooth_{L1}$   
为绝对损失函数, 常用于回归。

图 5 为 SSD 模型的训练过程。其中 G (Ground  
Truth Box) 代表真实框, D (Default Box) 代表默认框。  
在训练过程中, 随着训练的迭代, 训练的损失值会变  
得越来越小。定位损失及分类损失的的大小, 即默认  
框(D)相对于真实框(G)中心( $cx, cy$ )、宽( $w$ )和  
高( $h$ )的偏移量和分类的准确率; 定位损失减小的  
过程, 即 D 慢慢移动接近 G 的过程。



图5 真实框和默认框

Fig. 5 Real frame and default frame

#### 4 迁移学习

受人类跨任务传递知识能力的启发,迁移学习旨在利用来自源域的知识,提高学习性能或最小化目标域中需要标记的示例数量<sup>[14]</sup>。当使用有限的、带标签的示例来处理任务时,迁移学习尤其重要。事实证明,迁移学习具有广泛的适用性:

迁移学习的关键思想,是通过利用以前的方法来增强从源域到目标域的迁移学习效率、迁移学习经验,以优化它们之间的迁移方式。为了实现这一目标,将其分两个阶段建立了迁移学习方式。

第一阶段:将每种迁移学习经验编码为三个部分。将一对源域和目标域之间的转移知识参数化为潜在特征因子,并提高性能。从所有经验中学习反射函数,该函数将一对域以及它们之间传递的知识,映射到性能改进上。

第二阶段:优化在新到达的一对域之间传输的内容,以使与性能提高相匹配的学习反射函数的值最大化。

该框架为利用先前的迁移学习经验提高迁移学习效率打开了一扇新门。迁移学习无需系统专业知识即可以系统和自动的方式发现更多可转让的知识。如图6所示,VOC2007数据集中有大量物体,如狗、猫、鹦鹉等,与本文所需检测的动物有大量相似区域。因此,使用VOC2007的数据集,在SSD中进行预训练,进行迁移学习,加快了识别的收敛性以及提高检测的准确率。

运用迁移学习后,网络输入的图片被处理成 $300 \times 300$ 大小的图片。训练过程中,conv1-conv4为冻结参数预训练,后续再经过卷积和注意力模型获取动物关键性特征,得到分类置信度以及定位置信度,经过非极大抑制方法后,获得最终输出。后续流程如图7所示。



图6 迁移学习过程

Fig. 6 Transfer learning process

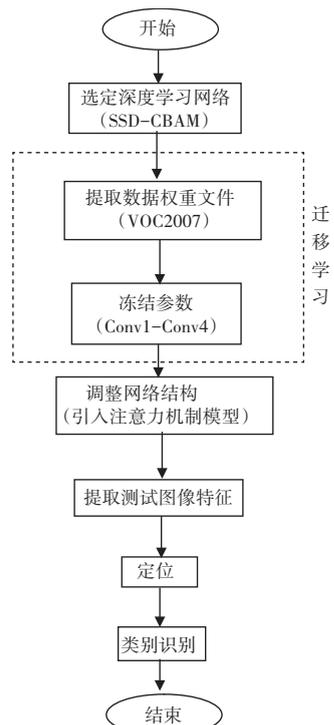


图7 训练流程图

Fig. 7 Training flowchart

### 5 实验结果与分析

#### 5.1 平台配置与数据集

本文实验使用的开发平台为 win10, 平台具体配置见表 2。

表 2 系统平台配置

Tab. 2 System platform configuration

名称	配置
CPU	Intel i5-5200U 2.20GHZ
内存	8GB
GPU	NVIDIA GeForce 920M 4.0G
GPU 加速库	CUDA 9.0cuDNN v7.1
深度学习框架	Pytorch1.1.0

有效的数据扩充不仅能够扩充训练样本数量, 还能增加训练样本的多样性。本文中同比例增加样本数量的方法有: 亮度增强、对称变换、旋转 0-90 角度以及翻转图片的方法。经过数据扩展后的图像如图 8 所示。



图 8 原图像与扩展后的图像

Fig. 8 Original image and expanded image

实验中, 运用 Pascal voc2007 数据权重集。大熊猫、大象、孔雀、金丝猴以及老虎图片各 500 张, 并用三维标注软件 labeling 做好标签。其标签分别对应 panda、elephant、peacoke、monkey 和 tiger。且按照 7 : 3 比例划分训练集和数据集。在训练前, 将所有输入网络的图片缩放处理成 300×300 像素大小的图片。

在训练过程中, 损失函数由分类置信度和坐标误差构成。一般情况下, 采集到的负样本数量, 即背景样本数量会远远大于其它样本数量, 这样会导致模型的准确度大大降低。为此, 采用目标样本数: 背景样本数 = 1 : 3 的方法来解决。

#### 5.2 实验结果与分析

如图 9 所示, 是本文算法 SSD-CBAM 与未改进算法的准确率曲线图。从图中可以看出, 本文算法

的准确率明显高于其它算法, 并在训练的较短步数内, 达到了较高的准确率。

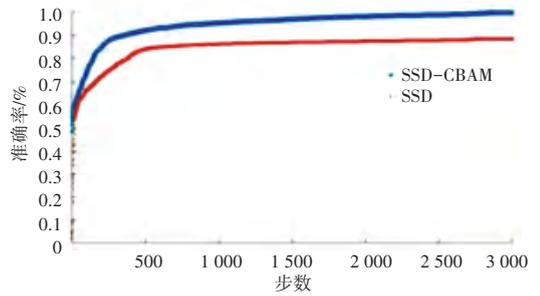


图 9 不同算法准确率对比

Fig. 9 Comparison of accuracy of different algorithms

表 3 为几个目标检测算法对濒危动物检测时的精确率、召回率以及检测速度的相关数据。从数据中可以看出, 本文算法 SSD-CBAM 虽然在检测速度上比 YOLO3<sup>[15]</sup> 及 SSD 低, 但是在保证检测速度的前提下, 精确率比其它方法有明显的优势, 在召回率上也有一定的提高。

表 3 不同算法性能参数

Tab. 3 Performance parameters of different algorithms

算法	精确率/%	召回率	平均每秒检测张数
Faster-RCNN	92.6	87.5	5
YOLO3	88.5	85.3	40
SSD	93.4	93.5	29
SSD-CBAM	96.0	94.1	23

图 10 为本文算法 SSD-CBAM 对濒危动物的检测效果。从图中可以看出: 对于金丝猴、老虎这样非常清晰的目标, 具有高达 98% 以上的置信度; 像熊猫和大象这样的重叠物体也能准确检测出来; 对有遮挡的凤凰, 虽然其置信度只有 64%, 但是预测仍然正确, 说明网络具有极强的鲁棒性。

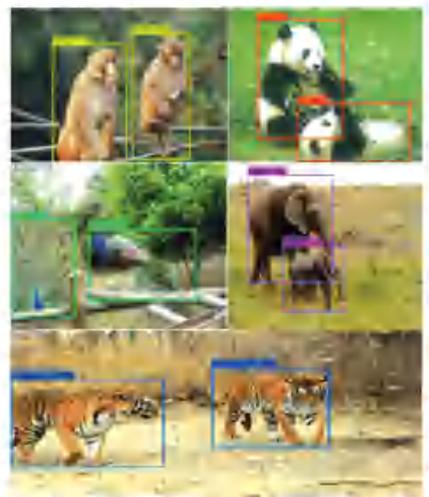


图 10 识别五种动物准确率

Fig. 10 The accuracy of identifying five animals

## 6 结束语

本文提出改进的 SSD-CBAM 网络,能够准确的识别濒危物种。其准确率,召回率以及检测速度都可满足实际要求,为濒危物种的保护以及可持续发展提供了保障。而且当训练集扩充到足够大时,识别精度也得到提升。但在实验中也发现一些不足,如当动物在图片中所占比例太小或者身体只有一部分出现在图片中时,其准确率及对动物检测时的置信度都有一定的下降。进一步的研究重点是对有遮挡动物的识别。

## 参考文献

- [1] 姜桂萍. 分析我国濒危野生动物保护现状与前景展望[J]. 中国农村教育, 2018(18): 66-67.
- [2] MATUSKA S, HUDEC R, KAMENCAY P, et al. Classification of wild animals based on SVM and local descriptors[J]. AASRI Procedia, 2014, 9: 25-30.
- [3] CORTES C, VAPNIK V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
- [4] RAMANAN D, FORSYTH D A, Barnard K. Detecting, localizing and recovering kinematics of textured animals[C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005.
- [5] 张公伯, 谷显良, 朱和贵. 基于 Adaboost 的动物二分类识别方法[J]. 计算机与数字工程, 2017, 45(04): 720-726+767.
- [6] CHEN G, HAN T X, HE Z, et al. Deep convolutional neural network based species recognition for wild animal monitoring[C]//2014 IEEE international conference on image processing (ICIP). IEEE, 2014: 858-862.

- [7] GOMEZ A, DIEZ G, SALAZAR A, et al. Animal Identification in Low Quality Camera-Trap Images Using Very Deep Convolutional Neural Networks and Confidence Thresholds[C]// International Symposium on Visual Computing. Springer International Publishing, 2016.
- [8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]// Proceeding of the IEEE Conference on ICLR 2015.
- [9] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [10] NOROUZZADEH M S, NGUYEN A, KOSMALA M, et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning[J]. Proceedings of the National Academy of Sciences of the United States of America, 2018.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
- [12] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [13] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP(99).
- [14] 高明, 陈玉涵, 张泽慧, 等. 基于新型空间注意力机制和迁移学习的垃圾图像分类算法[J/OL]. 系统工程理论与实践; 1-15 [2020-11-08].
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. Youonlylookonce: Unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

(上接第 25 页)

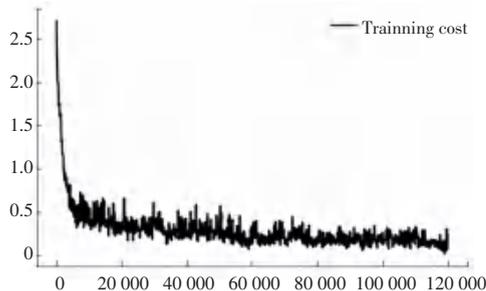


图 3 MNIST 上训练损失变化率

Fig. 3 Change rate of training loss on MNIST

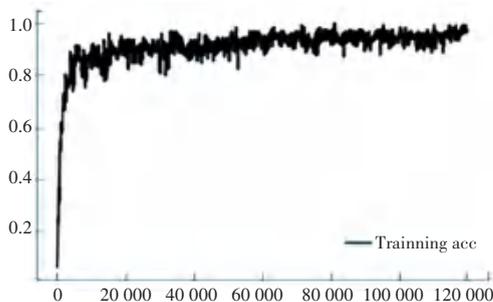


图 4 MNIST 上训练准确率变化率

Fig. 4 Accurate change rate of training on MNIST

表 1 0-9 测试数字预测识别率

Tab. 1 0-9 test digital prediction recognition rate		%								
数字	0	1	2	3	4	5	6	7	8	9
识别率	97.6	97.1	98.8	97.4	98.2	97.6	98.5	98.8	98.7	99.7

由表 1 可以看出,对 600 张手写数字进行预测识别后,每个数字的正确识别率都达到了 97% 以上,平均正确识别率达到了 98.24%。

## 3 结束语

该文介绍了一种手写数字识别的方法,使用深度神经网络 DNN 来训练模型,用交叉熵代价函数代替均方差成本函数。最后对模型进行测试,由表 1 可知,测试所得到的数字正确识别率均达到 97% 以上,本文设计的网络模型达到了较好的识别效果

## 参考文献

- [1] 楼怡航. 基于数字图像处理的车牌识别技术[J]. 电子制作, 2019(Z1): 72-75.
- [2] 吕国豪, 罗四维, 黄雅平, 等. 基于卷积神经网络的正则化方法[J]. 计算机研究与发展, 2014(9).
- [3] 白亚龙. 基于深度神经网络的图像识别系统的研究与改进[D]. 哈尔滨: 哈尔滨工业大学, 2014.
- [4] 邓长银, 张杰. 基于改进 LeNet-5 模型的手写数字识别[J]. 信息通信, 2018(1): 109-112.