

文章编号: 2095-2163(2022)03-0087-05

中图分类号: TP312

文献标志码: A

改进 SSD 算法对鸟类目标检测研究

唐鑫鑫, 陆安江, 彭熙瞬, 高海韬

(贵州大学 大数据与信息工程学院, 贵阳 550025)

摘要: 为了更加高效化、精确化的保护鸟类, 针对传统的人工鸟类识别方式, 本文提出利用 SSD 算法模型自动识别检测分类。首先, 收集不同种类鸟类图片, 并对图片灰度化、直方图均衡化以及归一化操作, 利用 labelImg-master 工具制作数据集; 其次, 改进特征提取网络 ResNet50, 在其最后一个卷积层后面增加一个 Dropout 层和 3 层全连接层, 用 Leaky-reLU 激活函数取代原残差学习单元的 ReLU 激活函数, 用改进的 ResNet50 替代 VGG 网络融入 SSD 模型; 最后, 在同样的条件下, 用 SSD 模型、ResNet50 替换 VGG 的 SSD 模型以及改进 ResNet50 替换 VGG 的 SSD 模型做目标检测实验, 通过几个不同评估标准对比表明, 改进 SSD 模型在鲁棒性、平均检测精度以及小目标检测性能方面都有有效改善。

关键词: 鸟类; SSD 模型; VGG 网络; ResNet50; 目标检测

Improved SSD algorithm for bird target detection

TANG Xinxin, LU Anjiang, PENG Xishun, GAO Haitao

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

[Abstract] In order to protect birds in a more efficient and accurate way, the SSD algorithm model is proposed to automatically identify, detect and classify the traditional artificial bird identification methods. First, the pictures of birds of different species were collected, and the graying, histogram equalization and normalization operations were carried out on these pictures. Next, the data set was made with labelimg-master tool. Next, the feature extraction network ResNet50 is improved by adding a Dropout layer and three fully connected layers behind the last convolution layer of ResNet50 and replacing the ReLU activation function of the original residual learning unit with the Leaky-ReLU activation function. With improved ResNet50 instead of VGG network into SSD model. Finally, we compared the SSD model, the SSD model of ResNet50 replacing VGG, and the SSD model of improved ResNet50 replacing VGG under the same conditions. Through the comparison of several different evaluation criteria, it is shown that the improved algorithm can effectively improve the convergence speed, average detection accuracy and small target detection performance.

[Key words] birds; SSD model; VGG network; ResNet50; target detection

0 引言

社会经济的快速发展, 对生态环境资源索取使用过度, 自然环境被严重破坏。鸟类是自然生态中重要的一员, 据国际鸟类联盟最新调查研究显示, 全球八分之一的鸟类面临灭绝的危险, 生存状况日益恶化, 数量下降呈现恐怖趋势, 保护鸟类刻不容缓。目前, 鸟类的识别工作仍主要依靠人工先验知识, 时间成本过大。通过人工智能技术构建一个智能鸟类识别分类系统, 能够节约人力成本, 为鸟类保护贡献一份力量。鸟类识别算法主要分为两类: 基于部件的多级分类算法和基于端到端的分类算法^[1]。基于部件的多级分类算法代表有 Wah^[2] 最先提出的基于传统特征的词包分类模型; Donahue 等人^[3] 提出的基于 CNN 的 DeCAF 特征; Zhang 等人^[4] 提出的

自局部检测到特征提取均采用 CNN 架构的 Part-based R-CNN 等。基于端到端的分类算法有 Fu 等人^[5] 提出的 RA-CNN 模型, 基于递归神经网络增强局部注意力及特征表达; Li 等人^[6] 提出幂归一化协方差矩阵 MPN-COV; 魏秀参^[7] 提出的基于 FCN 学习分割模型的 Mask-CNN 算法等等。

本文使用改进过的 Resnet50 网络 (称作 DL-Resnet50) 替换 SSD 的前置网络 VGG, 对自制的鸟类数据集进行目标识别检测分类。

1 SSD 模型的基本原理及结构

Wei Liu 在 ECCV 2016 上提出了 SSD 模型。该模型全称 Single Shot MultiBox Detector, 意思为单镜头多盒检测器。SSD 基于深度学习, 是人工智能下的一种目标检测模型, 其前置网络为 VGG-16, 是目

基金项目: 国家自然科学基金 (61865002)。

作者简介: 唐鑫鑫 (1997), 女, 硕士研究生, 主要研究方向: 计算机视觉、图像处理; 陆安江 (1979), 男, 博士, 副教授, 主要研究方向: 图像处理、物联网应用。

通讯作者: 陆安江 Email: 1321670647@qq.com

收稿日期: 2021-10-29

前流行的目标检测框架之一。该算法会在图片上均匀地产生若干个不同大小的候选框,且其长宽比均

不相同,使用卷积层将图像的特征提取出来,再进行回归和分类。SSD 模型网络结构如图 1 所示。

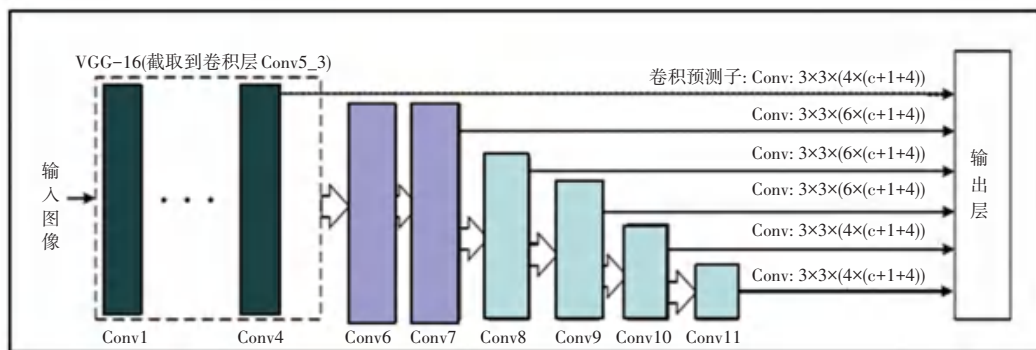


图 1 SSD 模型网络结构图

Fig. 1 Network structure diagram of SSD model

SSD 采用 VGG-16 为前置网络,将 VGG-16 的 2 个全连接层 FC6 和 FC7 替换为 3×3 卷积 Conv6 和 1×1 卷积 Conv7,将池化层 Pool_5 由 2×2 池化修改为 3×3 池化,移除 Dropout 层和全连接层 FC8,并添加 4 个新的卷积层: Conv8、Conv9、Conv10、Conv11。其处理过程中如激活函数全部采用 ReLU,且均在卷积操作结束后使用,池化操作不使用激活函数,做卷积与池化交替计算,假设输入的图像尺寸为 $224 \times 224 \times 3$,则每一步的计算与图像尺寸见表 1。

表 1 VGG16 计算过程图像变化

Tab. 1 Image changes during VGG16 calculation process

输入图像尺寸	卷积核及次数	池化域	输出图像尺寸
$224 \times 224 \times 3$	64 个, 3×3 , 2 次	无	$224 \times 224 \times 64$
$224 \times 224 \times 64$	无	2×2	$112 \times 112 \times 64$
$112 \times 112 \times 64$	128 个, 3×3 , 2 次	无	$112 \times 112 \times 128$
$112 \times 112 \times 128$	无	2×2	$56 \times 56 \times 128$
$56 \times 56 \times 128$	256 个, 3×3 , 3 次	无	$56 \times 56 \times 256$
$56 \times 56 \times 256$	无	2×2	$28 \times 28 \times 256$
$28 \times 28 \times 256$	512 个, 3×3 , 3 次	无	$28 \times 28 \times 512$
$28 \times 28 \times 512$	无	2×2	$14 \times 14 \times 512$
$14 \times 14 \times 512$	512 个, 3×3 , 3 次	无	$14 \times 14 \times 512$
$14 \times 14 \times 512$	无	2×2	$7 \times 7 \times 512$

最后的输出图像与两层 $1 \times 1 \times 4$ 096, 一层 $1 \times 1 \times 1$ 000 进行全连接,即共有 3 层全连接层,最后再通过 ReLU 激励函数激活,输出的预测结果可以有 1 000 个^[8]。

不同特征图的每个特征点上都有一组默认先验框与之对应,因此 SSD 模型预测的时候计算默认先验框的位置偏移和置信度^[9]。

SSD 的损失函数值包含两个部分:置信度 (L_{conf}) 和定位 (L_{loc}), 而损失函数就是这两者的加权和。置信度损失 L_{conf} 的计算公式(1)如下:

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0) \quad (1)$$

其中, Pos 表示正样本(目标); Neg 表示负样本

(背景); \hat{c}_i^p 是 $\frac{\exp(c_i^p)}{\sum_p (c_i^p)}$ 的简化表示,是 Softmax 激活函数产生的。

定位损失 L_{loc} 的计算公式(2):

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in (c_x, c_y, w, h)} x_{ij}^k smooth_{L1}(l_i^m - g_j^m) \quad (2)$$

其中, x_{ij}^k 表示第 i 个预测框与第 j 个真实框关于类别 k 是匹配关系; $smooth_{L1}(l_i^m - g_j^m)$ 表示平滑 L1 范数; l_i^m 为预测框; g_j^m 为真实框。

将其加权求和就是总的损失函数,计算公式(3)如下:

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (3)$$

SSD 算法需要计算两部分,即相应的先验框与目标类别的置信度和相应的位置回归。

其中,参数 $x \in (0, 1)$, 表示默认框和实际检测框的匹配是否成功;参数 α 作用类似于权重,用来平衡置信损失和位置损失; c 是分类置信度; l 代表预测框; g 代表真实标签框; N 是与该类别的校准框匹配的默认框数量,当 $N = 0$ 时,说明损失为 0。

2 SSD 模型改进

2.1 SSD 模型存在的问题

SSD 的优点在于采取多尺度特征融合策略,利用多个卷积层的输出参与预测,提升了检测精度;直接使用默认框进行检测,没有候选区域生成的过程,提升了检测速度。但对小目标的鲁棒性差,会有识别不出来的问题。根据 VGG 网络的特性,增加网络层数后提升的效果并不明显,因为出现了网络退化的问题,退化问题阻碍了网络收敛,不仅增加了训练误差,也增加了测试误差,降低了网络精度。

为了提升网络训练的识别准确率,最简单直接的方式是增加网络层数,以增加对目标特征的学习次数。但若只是盲目增加网络层数,问题并没有得到解决;相反地,还会导致错误率的突然上升,产生此问题的原因是发生了梯度消失^[10]。

2.2 改进 ResNet50 网络结构

ResNet50 网络中提出的残差结构 (Residual Block) 主要目的是解决网络退化问题^[11], ResNet50 的残差网络学习单元如图 2(a) 所示。

ResNet50 网络由多个残差模块构建而成,其函

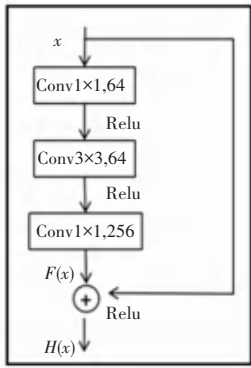
数表达式为式(4):

$$H(x) = F(x) + x \quad (4)$$

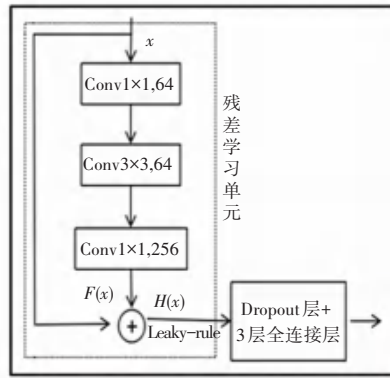
其中, x 为输入; $F(x)$ 表示残差映射; $H(x)$ 表示残差网络的特征输出。

当 $F(x) = 0$ 时,表示该卷积层做恒等映射;当 $F(x) > 0$,表示该卷积层学习到新的特征信息,从而保证反向传播时的梯度传递,有效解决了网络训练过程中存在的梯度消失和网络退化问题。

本文对 ResNet50 网络的改进结构如图 2(b) 所示,将改进后 ResNet50 的网络称为 DL-ResNet50。



(a) ResNet50 残差学习单元



(b) ResNet50 改进残差学习单元

图 2 两个残差学习单元对比

Fig. 2 Comparison of two residual learning units

由于 ResNet50 网络训练过程中易出现过拟合,为了进一步提高该网络的识别精度,本文在原网络最后一个卷积层后加入 Dropout 层和 3 层全连接层,用 Leaky-relu 函数取代原网络的 ReLU 函数。

加入 Dropout 层和 3 层全连接层后,可以让 ResNet50 网络过拟合的机率有效下降,改善增强模型泛化能力。当网络输入 $x < 0$ 时,ReLU 函数无法更新,从而模型不能训练学习特征。而 Leaky-relu 函数在输入 $x < 0$ 时输出持续变化,更新权重继续学习。改进的 SSD 模型工作流程如图 3 所示。

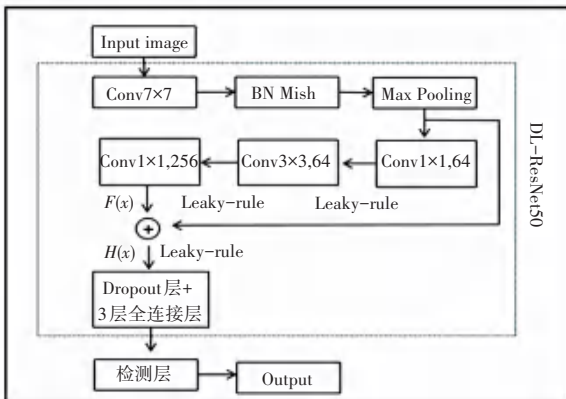


图 3 改进 SSD 模型工作流程

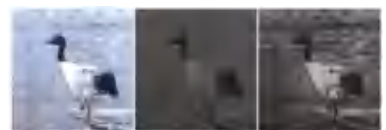
Fig. 3 Improved SSD model workflow

3 实验结果分析

3.1 数据预处理

本次数据集从文献[12]所记录的鸟类中随机选取 10 种,分别是黑颈鹤、灰鹤、斑头雁、黄眉柳莺、白鹭、大斑啄木鸟、山斑鸠、喜鹊、赤麻鸭以及乌鸦,共计 4 830 张图片,为了保证模型识别效果,防止模型过拟合,对数据图像进行镜像翻转处理,增加数据集数量,将数据集按照模型比例 6 : 4 : 1 分配划分为训练集、验证集、测试集。

为了更好的训练模型,对收集的图像数据灰度化、均衡化以及归一化处理,如图 4 所示,其中(a)为原图,(b)表示经过加权均值灰度化的图,(c)表示经过直方图均衡化后的灰度图。对比(b)和(c)可以明显看出图中的目标与背景的差异变大。再利用 labellmg-master 工具在图中框出目标位置及类型,即可获取目标的像素信息。



(a) 原图 (b) 灰度图 (c) 均衡化灰度图

图 4 图片处理对比

Fig. 4 Comparison of image processing

3.2 实验结果对比

本文实验平台选取 tensorflow 开源框架, Windows10 系统, python3.6 编程语言, cuda9.0, cudnn7.0, NVIDIA GTX2060 显卡。

本文实验以模型损失值曲线变化、平均交并比 (Average Intersection Over Union, Avg IOU) 以及 AP 曲线来作为评估依据, 分别在 SSD、SSD+ResNet50 以及 SSD+DL-ResNet50 3 个模型下做实验对比。3 个模型实验训练参数一致, 训练批次大小为 8, 迭代 300 次, 初始学习率为 0.001, 权值衰减系数为 0.0005。损失值曲线变化如图 5 所示。

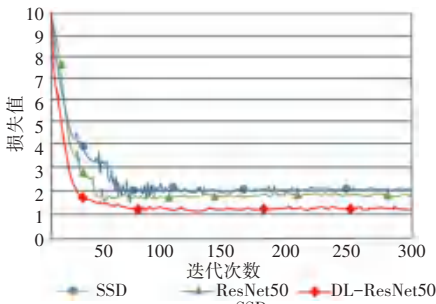


图 5 3 种模型的损失值曲线

Fig. 5 Loss value curves of the three models

本文所提算法 SSD+DL-ResNet50 (红色线) 收敛速度最快且最终损失值最小, 稳定在 1.3 左右。SSD+ResNet50 (绿色线) 次之, 原 SSD 模型最慢损失值也最高, 约 2.1, 本文算法在损失值这一项上降低了约 0.8。

IOU 在目标检测中常表示预测框和真实框相交面积与相并面积的比值, 比值越接近 1 效果越好, 公式(5)。

$$IoU = \frac{A_{pred} \cap A_{true}}{A_{pred} \cup A_{true}} \quad (5)$$

其中, A_{pred} 表示目标预测框面积, A_{true} 表示目标真实检测框的面积。

通过实验结果可得 3 个模型的交并比变化曲线, 如图 6 所示。由图 6 可知, 随着迭代次数的增加, 3 种模型的交并比值都逐渐增加, 但是由红色线可知本文改进模型 DL-ResNet50 SSD 的值均高于绿色线 ResNet50 SSD 模型和蓝色线原 SSD 模型。

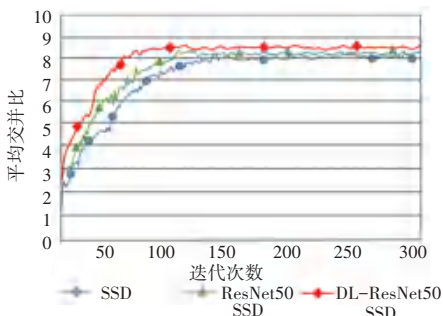


图 6 3 种模型交并比变化曲线

Fig. 6 Intersection ratio change curves of the three models

本文选择准确率和召回率来作为目标检测的性能评估标准, 准确率就是测试检测正确的目标在所有样本中所占的比例, 计算表达如式(6)。召回率则表示在所有真实目标中被正确检测出来的占比, 计算表达如式(7)。

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

其中, $Precision$ 表示准确率; $Recall$ 表示召回率; TP 表示模型正确检测的目标个数; FP 表示模型错误检测的目标个数; FN 表示模型漏检的正确目标个数。

实际上, 准确率和召回率往往是此消彼长, 难以兼得, 因此将计算得到的 P, R 值绘制成 PR 曲线图, 综合考虑 PR 曲线下的面积 AP 值大小来评估各类别检测性能, AP 值越大表示模型检测精度越高, 即性能越好, 计算公式(8)。

$$AP = \int_0^1 PdR \quad (8)$$

针对多个类别 N , 使用平均值 mAP 评估模型整体性能, 公式(9)。

$$mAP = \frac{1}{N} \sum_{k=1}^N AP_k \quad (9)$$

斑头雁和灰鹤两种类别的鸟在原 SSD 模型、用 ResNet50 替换 VGG 网络的 SSD 模型还有本文提出的改进 ResNet50 取代 VGG 的 SSD 模型下的 AP 变化值如图 7 所示。对比分析, 本文所提算法的 AP 值更大, 模型精确度更高。

SSD、SSD 中已有的 VGG 网络替换为 ResNet50 以及改进 ResNet50 取代 VGG 的 SSD 模型下的检测精度对比见表 2, 可知本文的改进算法在总体的检测精度上比原 SSD 提升 5 个点

表 2 3 种模型检测精度

Tab. 2 Detection accuracy of the three models

模型	检测精度/%
SSD(VGG)	78.1
SSD(ResNet50)	80.4
SSD(DL-ResNet50)	83.6

检测结果如图 8 所示, 其中 (a) 是原 SSD 模型的检测结果, 可以看到有些被遮挡的目标没有被检测到, (b) 图是改进后的 SSD 模型检测结果, 虽然有的被遮挡目标也没有被检测出, 但是增加了 2 个目标检测框。

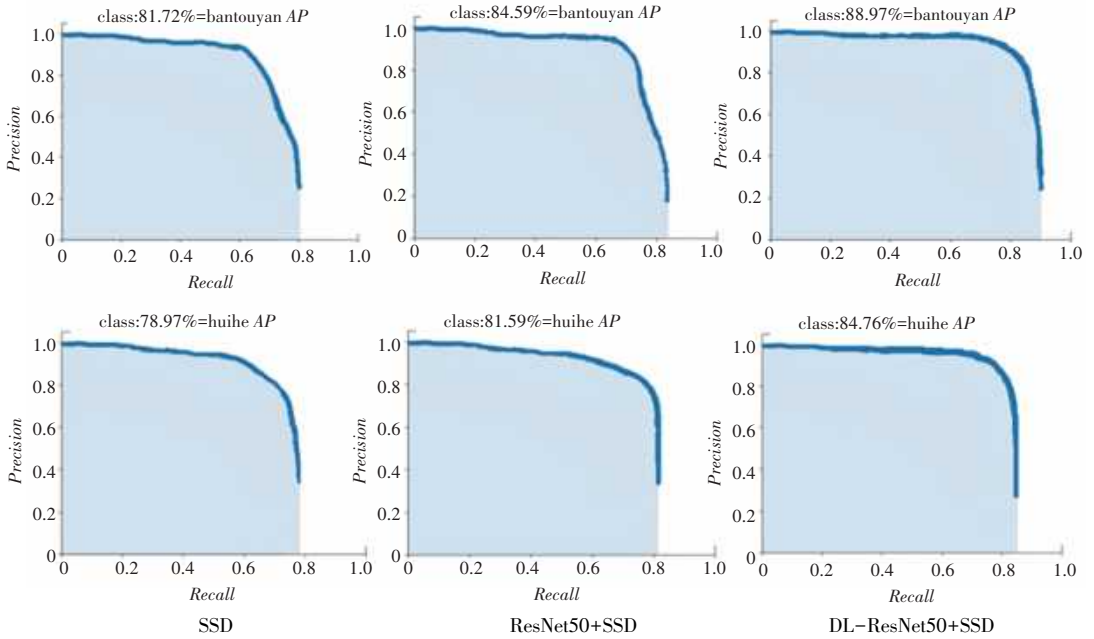


图 7 3 种网络的 AP 曲线

Fig. 7 AP curves of the three networks



(a) SSD 模型检测结果



(b) 改进 SSD 模型检测结果

图 8 检测结果

Fig. 8 Detection results

参考文献

[1] 智绪晔. 基于深度学习的鸟类图像识别研究[D].大连:大连理工大学,2019.

[2] WAH C, BRANSON S, WELINDER P, et al. The Caltech-UCSD Birds-200-2011 Dataset[J]. california institute of technology, 2011.

[3] DONAHUE J, JIA Y, VINYALS O, et al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition [C]//International Conference on Machine Learning, 2014; 647-655.

[4] ZHANG N, DONAHUE J, GIRSHICK R, et al. Part-Based R-CNNs for Fine-Grained Category Detection [C]//European Conference on Computer Vision, 2014; 834-849.

[5] FU J, ZHENG H, MEI T. Look closer to See better; Recurrent attention convolutional neural network for fine-grained image recognition [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017; 4438-4446.

[6] LI P, XIE J, WANG Q, et al. Is Second-order Information Helpful for Large-scale Visual Recognition? [C]//International Conference on Computer Vision, 2017; 2070-2078.

[7] 魏秀参. 深度学习下细粒度级别图像的视觉分析研究[D].南京:南京大学,2018.

[8] 阮士峰. 基于改进 SSD 算法的行人佩戴口罩检测研究[J]. 科技经济导刊, 2020, 28(35): 9-13.

[9] 朱良宽, 邵思协, 景维鹏, 等. PSO 优化三维 Otsu 森林冠层图像快速分割算法[J]. 哈尔滨理工大学学报, 2019, 24(5): 128-133.

[10] HE F, LIU T, TAO D. Why ResNet Works? Residuals Generalize [J]. IEEE Trans Neural Netw Learn Syst, 2020, 31(12): 5349-5362.

[11] KAIMING He, et al. Deep Residual Learning for Image Recognition[J]. arXiv:1512.03385, 2015.

[12] 罗祖奎, 刘文, 李振吉, 等. 贵州草海冬季鸟类群落特征[J]. 华东师范大学学报(自然科学版), 2012(4): 102-111.

4 结束语

本文提出一种用改进 ResNet50 网络, 取代 SSD 模型的基准网络 VGG 的 SSD 模型算法对鸟类目标进行检测研究。通过实验对比可知, 本文提出的算法在模型收敛速度、检测精度以及多目标检测方面都有改善提升。本文数据集种类只有 10 类, 因此下一步的研究工作是增加数据集种类以及数量, 将算法融合进系统之中, 实现一个完整的鸟类识别分类系统。