

# 改进的 Cascade R-CNN 算法在目标检测上的应用

张娜<sup>1</sup>, 包梓群<sup>1</sup>, 罗源<sup>1</sup>, 吴彪<sup>2</sup>, 涂小妹<sup>3</sup>

(1. 浙江理工大学计算机科学与技术学院(人工智能学院), 浙江杭州 310018; 2. 浙江理工大学理学院, 浙江杭州 310018;  
3. 浙江广厦建设职业技术大学, 浙江东阳 322100)

**摘要:** 针对 Cascade R-CNN 目标检测算法中存在检测精度较低以及目标遮挡问题, 本文提出一种改进的 Cascade R-CNN 网络目标检测算法. 该算法在主干网络 ResNet101 中引入可切换空洞卷积模块(Switchable Atrous Convolution, SAC), 该模块主要由两个全局上下文模块以及 SAC 组件构成, 采用 SAC 组件以不同的空洞卷积率对特征进行卷积, 并使用 Switch 函数收集特征来提高特征提取能力. 同时, 在 ResNet101 残差网络中引入坐标注意力机制(Coordinate Attention, CA), 该机制将位置信息嵌入通道注意力中, 用于更好地获取方向感知和位置感知信息, 进而提高目标检测精度. 此外, 针对目标遮挡问题, 引入 Repulsion Loss 损失函数. 该损失函数主要由吸引项和排斥项组成, 吸引项使得预测框和匹配上的目标框尽可能接近, 排斥项使得预测框远离错误目标, 进而减少非极大值抑制(Non-Maximum Suppression, NMS)的误检, 提高目标检测中遮挡问题的检测精度. 实验结果表明, 在公开的科大讯飞 AI 挑战赛数据集上, 与原算法测试性能相比, 改进的 Cascade R-CNN 网络对该数据集检出率增长了 2.39%, 改进算法的识别精度有一定的提高.

**关键词:** Cascade R-CNN; 可切换空洞卷积; Repulsion Loss; 目标检测; 目标遮挡

**基金项目:** 浙江省重点研发计划项目(No.2020C03094); 国家级大学生创新创业训练计划项目(No.202010338024); 浙江省教育厅一般科研项目(No.Y202147659)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2023)04-0896-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220116

## Application of Improved Cascade R-CNN Algorithm in Target Detection

ZHANG Na<sup>1</sup>, BAO Zi-qun<sup>1</sup>, LUO Yuan<sup>1</sup>, WU Biao<sup>2</sup>, TU Xiao-mei<sup>3</sup>

(1. School of Computer and Technology (School of Artificial Intelligence), Zhejiang Science Technology University, Hangzhou, Zhejiang 310018, China;

2. School of Science, Zhejiang Science Technology University, Hangzhou, Zhejiang 310018, China;

3. Zhejiang Guangsha Vocational and Technical University of Construction, Dongyang, Zhejiang 322100, China)

**Abstract:** An improved target detection algorithm based on Cascade R-CNN network is proposed to solve the problems of low detection accuracy and target occlusion in the target detection algorithm based on Cascade R-CNN. The algorithm introduces a switchable atrous convolution (SAC) module into the backbone ResNet101, which is composed of two global context modules and SAC components. The SAC component is used to convolution the features with different void convolution rates, and the Switch function is used to collect the features to improve the ability of feature extraction. At the same time, coordinate attention (CA) is introduced in ResNet101 residual network, which embeds position information into channel attention, and is used to obtain direction and position information better to improve the accuracy of target detection. In addition, aiming at the problem of target occlusion, this paper introduces the repulsion loss function, which is mainly composed of the attraction term and the exclusion term. The attraction term makes the prediction box and the target box on the matching as close as possible, and the exclusion term makes the prediction box away from the wrong target, thereby reducing the false detection of non-maximum suppression (NMS) and improving the detection accuracy of the occlusion problem in object detection. The experimental results show that the detection rate of the improved Cascade R-CNN network is 2.39% higher than that of the original algorithm on the open IFLYTEK Challenge dataset, the recognition accuracy of the improved algorithm is improved to a certain extent.

**Key words:** Cascade R-CNN; switchable atrous convolution; repulsion loss; target detection; target occlusion

Foundation Item(s): Key R&D Project of Zhejiang Province (No.2020C03094); National College Students Innovation and Entrepreneurship Training Program (No.202010338024); Education General Scientific Research Project of Zhejiang Province (No.Y202147659)

## 1 引言

随着深度学习的快速发展,基于深度学习的目标检测算法被广泛应用于各个领域.目前主流的目标检测算法分为两大类:(1)基于候选区域的双阶段目标检测算法;(2)基于直接回归的单阶段目标检测算法.单阶段检测算法在准确率和运行速度上能达到一个均衡,是目前目标检测中使用较多的一种检测框架.

单阶段目标检测算法主要有 SSD<sup>[1]</sup> 系列和 YOLO<sup>[2]</sup> 系列.曲优等<sup>[3]</sup>提出了基于锚框变换的单阶段旋转目标检测方法,该算法在 YOLOv3 的基础上,设计了新的特征提取网络 Darknet-53-Dense,提高了特征提取能力,并且在检测头部网络提出了一种锚框变换网络(Anchor Transformation Network, ATN),将初始的水平锚框变换为旋转锚框,进而提高目标检测精度.侯少麒等<sup>[4]</sup>提出了基于空洞卷积金字塔的目标检测算法,将不同尺寸的空洞卷积层引入特征金字塔(Feature Pyramid Networks, FPN)中,构建混合感受野模块(Hybrid Receptive Field Module, HRFM),通过增大感受野获取更多全局特征信息,解决目标的遮挡问题.张云佐等<sup>[5]</sup>提出了反残差目标检测算法,该算法在 YOLOv4 的基础上,提出了一种高效的轻量级卷积模型(Lightweight Ghost Convolution, LGC),采用更少的参数获取更多的特征图,提高网络对全局特征信息的提取能力,进而提升目标检测的整体性能.Zhao 等<sup>[6]</sup>提出了基于改进 YOLOv3 的目标检测算法,在原有算法的网络结构上添加卷积层模块对样本进行目标背景分类,并调整特征图上的锚框大小,通过输出目标背景概率,过滤掉背景概率值低于设定阈值的样本,从而解决原算法中存在的正负样本比例失衡的问题.Zhai 等<sup>[7]</sup>提出了一种基于 DenseNet<sup>[8]</sup>和特征融合的改进的 Single Shot multibox Detector(SSD)目标检测算法,在 SSD 网络的基础上根据 DenseNet 的稠密连接设计了特征提取网络 DenseNet-S-31-1 来增强模型的特征提取能力,并且在多尺度检测部分,引入了多尺度特征层的融合机制,将网络结构中的底层语义特征和高层语义特征结合起来,进而提高模型性能.

双阶段目标检测算法主要有 RCNN<sup>[9]</sup> 系列.Liu<sup>[10]</sup>提出了一种改进的 Faster R-CNN<sup>[11]</sup> 目标检测算法,通过引入难例样本挖掘<sup>[12]</sup>和交替训练两种策略使模型在精度方面有了一定提升.李晓光等<sup>[13]</sup>提出了面向多尺度目标检测的改进的 Faster R-CNN 算法,通过采用多层次提取特征策略提取多尺度目标特征<sup>[14]</sup>和使用多通

道方法生成多尺度目标候选框使目标检测算法的精度有一定的提高.基于上述单阶段或双阶段的方法进行检测,在一定程度上可以提高目标检测精度.然而,以上改进算法在检测目标差异大、背景复杂以及目标遮挡的情况下,检测精度较低.

针对目标检测算法中在检测目标差异大、背景复杂以及目标遮挡的情况下检测精度较低的问题,本文提出了改进的 Cascade R-CNN<sup>[15]</sup> 目标检测算法.该算法在原有的 ResNet101<sup>[16]</sup> 的主干网络中,引入可切换空洞卷积模块<sup>[17]</sup>.该模块由两个全局上下文模块对特征进行加权融合,再通过可切换空洞卷积模块(Switchable Atrous Convolution, SAC)组件以不同的空洞卷积率对特征进行卷积,并使用 Switch 函数收集特征,提高基础网络的特征提取能力,进而提高目标检测的精度.同时,在 ResNet101 残差网络中引入坐标注意力机制,该机制将位置信息嵌入通道注意力中,用于更好地获取方向感知和位置感知信息,进而提高目标检测精度.此外,基于 Smooth<sub>L1</sub><sup>[18]</sup> 的回归损失函数主要是吸引损失函数,即该类损失函数主要是使预测框和匹配上的目标尽可能接近,而无法解决同类物体相互遮挡的问题.本文采用 Repulsion Loss<sup>[19]</sup> 函数在要求预测框靠近目标框的同时,也要求预测框远离其他不属于目标的真实框,并且降低了 NMS<sup>[20]</sup> 对阈值的敏感性,进而解决目标遮挡问题.

## 2 相关工作

### 2.1 Cascade R-CNN 算法概述

Cascade R-CNN 网络主要由特征提取网络 ResNet101,特征金字塔网络 FPN 和 RPN<sup>[21]</sup>(Region Proposal Network, RPN)网络层,以及级联检测器组成.特征提取网络 ResNet101 用于提取图片特征,将原图经过 Conv1, Conv2, Conv3, Conv4 以及 Conv5 卷积操作以及不同层次的特征融合得到不同尺度的特征图 P2, P3, P4 以及 P5.然后将 P2, P3, P4 以及 P5 不同尺度的特征图输入 RPN 中用于得到候选目标区域,将所得的候选目标区域做 ROI Align<sup>[22]</sup>操作后,得到统一大小的 Region of Interest(ROI)特征图. Cascade R-CNN 网络结构图如图 1 所示.

在检测阶段,与 Faster R-CNN 不同的是, Cascade R-CNN 使用级联检测器的方式进行检测,由三个检测器分别设置不同的阈值进行检测.每个检测器由 ROI Align、全连接层、分类得分 C 和边框回归位置坐标 B 组

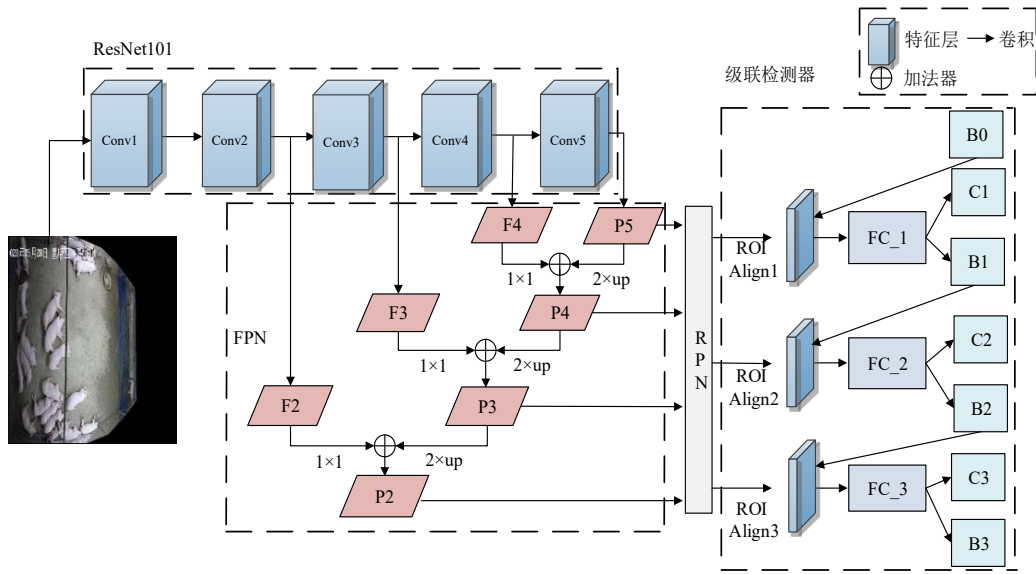


图1 Cascade R-CNN 网络结构

成. 检测时通过前一阶段检测器输出的边框回归 B 对候选目标区域进行重采样, 逐步提升 IoU 阈值训练得到新的分类得分 C 和边框回归 B, 最终提升样本质量和网络训练效果.

### 2.2 边框回归

如图 2 所示, 标注框  $P$  为预定义的 anchor<sup>[23]</sup>, 标注框  $G$  为目标框, 标注框  $G'$  为模型输出的预测框, 其最终目的是将预测框更加接近目标框. 当候选框与目标框的 IoU 较大时, 从候选框到预测框的变换  $d(\cdot)$  可以认为近似的线性变换. 定义目标框的中心  $(G_x, G_y)$ , 宽高  $(G_w, G_h)$ , 候选框中心  $(P_x, P_y)$ , 宽高  $(P_w, P_h)$ , 建立回归模型如式(1)所示:

$$\begin{cases} G'_x = P_x + P_w d_x(P) \\ G'_y = P_y + P_h d_y(P) \\ G'_w = P_w \exp(d_w(P)) \\ G'_h = P_h \exp(d_h(P)) \\ d_*(P) = \mathbf{w}_*^T \phi(P) \end{cases} \quad (1)$$

其中,  $\phi(P)$  为候选框特征信息,  $\mathbf{w}_*$  为要学习的参数.

### 2.3 损失函数

Cascade R-CNN 的损失函数主要由两部分组成, 分别是分类误差<sup>[24]</sup>和坐标回归误差<sup>[25]</sup>, 如式(2)所示:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \quad (2)$$

其中,  $p_i$  为 anchor 预测为目标概率,  $p_i^*$  为真实框的概率,  $t_i = (t_x, t_y, t_w, t_h)$  是一个向量, 表示预测框的 4 个参数化坐标,  $t_i^*$  是真实框的坐标向量,  $N_{\text{cls}}$  和  $N_{\text{reg}}$  都表示样

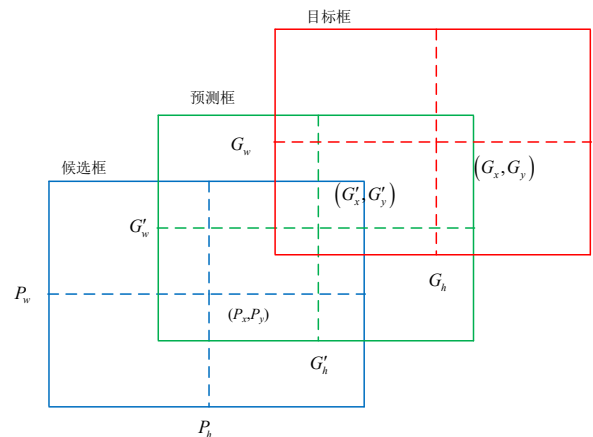


图2 目标回归框图

本总数量,  $\lambda$  为权重平衡因子.  $L_{\text{cls}}(p_i, p_i^*)$  是预测类别置信度与目标类别的误差, 损失函数采用交叉熵损失函数.  $L_{\text{reg}}(t_i, t_i^*)$  为边框回归损失, 采用 Smooth<sub>L1</sub> 作为损失函数, 如式(3)~(5)所示:

$$L_{\text{cls}}(p_i, p_i^*) = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (3)$$

$$L_{\text{reg}}(t_i, t_i^*) = \text{Smooth}_{L1}(t_i - t_i^*) \quad (4)$$

$$\text{Smooth}_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \quad (5)$$

## 3 改进的 Cascade R-CNN 模型

针对 Cascade R-CNN 网络中存在检测率低下以及目标遮挡的问题, 本文提出了改进的 Cascade R-CNN 网络结构图, 如图 3 所示. 改进的 Cascade R-CNN 算法在 ResNet101 中引入可切换空洞卷积模块, 该模块由 3 部

分组成,包括两个全局上下文模块和一个 SAC 组件,采用 SAC 组件以不同的空洞卷积率对特征进行卷积,并使用 Switch 函数收集特征来提高特征提取能力.在 ResNet101 残差网络中引入坐标注意力机制,该机制将位置信息嵌入通道注意力中,用于更好地获取方向感知和位置感知信息,进而提高目标检测精度.此外,

针对目标遮挡问题,引入 Repulsion Loss 损失函数,该损失函数主要由吸引损失和排斥损失组成,吸引损失使得预测框和匹配上的目标框尽可能接近,排斥损失使得预测框远离错误目标,进而减少非极大值抑制的误检,提高目标检测中遮挡问题的检测精度.

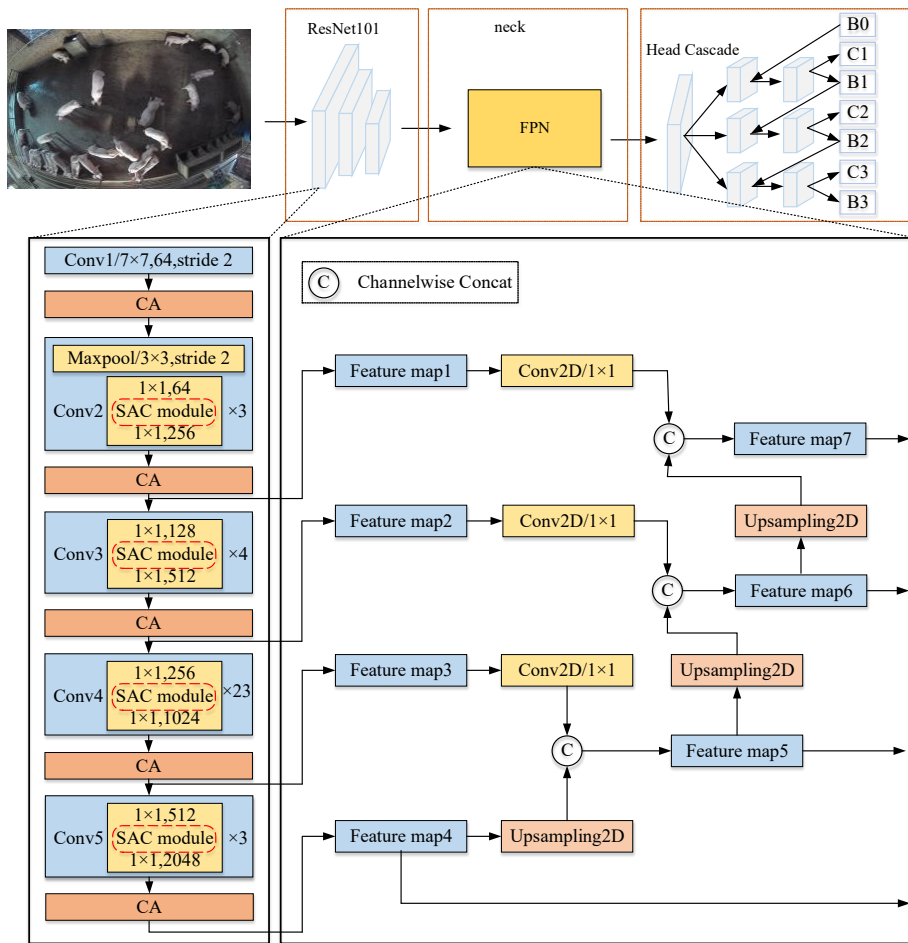


图3 改进的 Cascade R-CNN 网络结构图

如图3所示,在网络结构上,对 Cascade R-CNN 的主干网络 ResNet101 进行改进,分别在 Conv1, Conv2, Conv3, Conv4 以及 Conv5 的层次间加入 CA 模块,在 Conv2, Conv3, Conv4 以及 Conv5 模块内加入 SAC 模块,通过在各个卷积模块中加入注意力机制,强化特征提取网络去学习有用的特征,抑制无用的特征,提高特征的提取能力.

### 3.1 可切换空洞卷积模块

可切换空洞卷积模块主要由三个部分组成,如图4所示:两个全局上下文信息模块,即 Pre-Global Context 模块和 Post-Global Context 模块;一个 SAC 组件. SAC 组件处于两个全局上下文信息模块之间.首先,输入特征经过 Pre-Global Context 模块,该模块将输入特征使用全

局平均池化层进行压缩处理,然后使用  $1 \times 1$  的卷积得到输出,再将该输出与输入特征进行相加,得到 Pre-Global Context 模块处理后的输出. SAC 组件主要由三个分支组成,分别是空洞卷积率为 1 的形变空洞卷积、空洞卷积率为 3 的形变空洞卷积和 Switch 函数.将 Pre-Global Context 处理后的结果分别使用空洞卷积率为 1 和空洞卷积率为 3 的形变卷积进行卷积,再将得到的结果分别与 Switch 函数得到的结果进行相乘,得到的结果再经过 Post-Global Context 模块处理后得到最后的特征输出.可切换空洞卷积模块的重点是使用 Switch 函数计算出的值来确定卷积核选择不同的空洞率. Switch 函数的数值由卷积核为  $5 \times 5$  的平均池化层后接一个  $1 \times 1$  的卷积层组成.

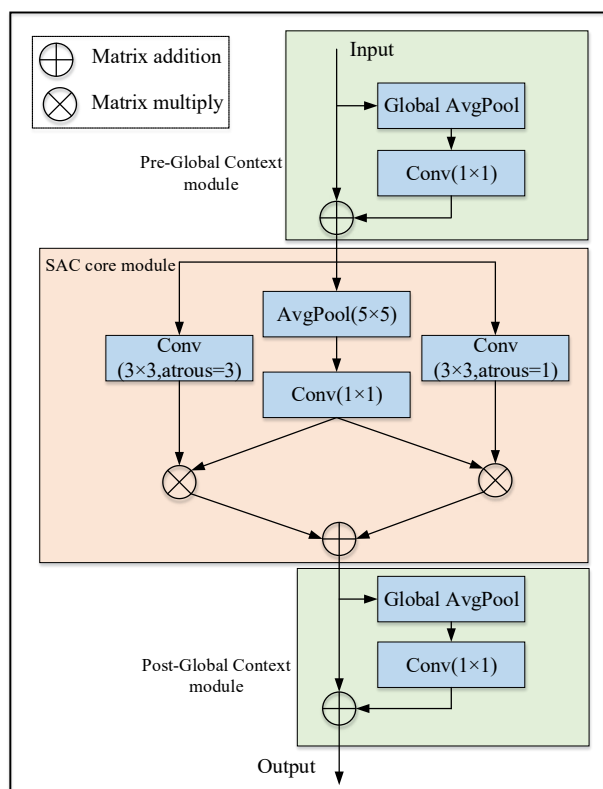


图4 可切换空洞卷积模块

如图3所示,将ResNet101中所有BlockNeck中的3×3的卷积层都替换为可切换空洞卷积模块,来实现不同空洞率情况下卷积操作的切换.不同空洞率的空洞卷积可以捕捉到不同感受野的目标,从而使网络可以自适应学习调整选择不同感受野的卷积结果.SAC算法如算法1所示.

### 3.2 Coordinate Attention

Coordinate Attention注意力机制主要由两个部分组成:一个是坐标信息嵌入,另一个是注意力生成.坐标信息嵌入是将全局池化解为一对一的一维特征编码操作.具体做法是对于输入特征图 $T$ (维度为 $B \times C \times H \times W$ ),其中 $B$ 是批次, $H$ 是 $T$ 的高度, $W$ 是 $T$ 的宽度, $C$ 是 $T$ 的通道数.使用尺寸为 $(H, 1)$ 和 $(1, W)$ 的池化核分别沿着水平方向和垂直方向对每个通道进行编码. $z^h$ 和 $z^w$ 分别为高度为 $h$ 和宽度为 $w$ 的第 $c$ 个通道的输出,如式(6)和式(7)所示:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} T_c(h, i) \quad (6)$$

$$4z_c^w(w) = \frac{1}{H} \sum_{0 \leq i < H} T_c(j, w) \quad (7)$$

上述变换是沿着两个方向进行特征聚合的,返回的是一对方向感知注意力图.为了更好地利用这两个方向上的特征聚合信息,级联这两个特征图,然后使用一个 $1 \times 1$ 卷积运算 $F_1$ ,生成对空间信息在水平方向和垂

#### 算法1 可切换空洞卷积算法

输入:输入张量 $X$ ,大小为 $B \times C \times H \times W$ , $X$ 表示ResNet101中的特征图.  
输出:输出张量 $X'$ .

(1)计算Pre-Global Context模块的结果 $x_{pre}$ :

$$x_{pre} \leftarrow \text{Conv2d}(\text{AdaptiveAvgPool2d}(X)) + X$$

/\*将张量 $X$ 进行平均池化和卷积,再与输入相加.AdaptiveAvgPool2d为平均池化函数;Conv2d为卷积函数\*/

(2)计算Switch函数的结果 $S$ :

$$S \leftarrow \text{Conv2d}(\text{AdaptiveAvgPool2d}(x_{pre}))$$

/\*将张量 $x_{pre}$ 进行平均池化和卷积.AdaptiveAvgPool2d为平均池化函数,Conv2d为卷积函数\*/

(3)计算SAC组件中空洞率为3的卷积,输出结果为 $P_{out_l}$ :

$$P_{out_l} \leftarrow \text{deform\_conv}(x_{pre}, \text{dilation}=3)$$

/\*将张量 $x_{pre}$ 进行可形变卷积,空洞率为3\*/

(4)计算SAC组件中空洞率为1的卷积,输出结果为 $P_{out_r}$ :

$$P_{out_r} \leftarrow \text{deform\_conv}(x_{pre}, \text{dilation}=1)$$

/\*将张量 $x_{pre}$ 进行可形变卷积,空洞率为3\*/

(5)计算Switch函数选择切换的结果 $P_{out}$ :

$$P_{out} \leftarrow S \times P_{out_r} + (1 - S) \times P_{out_l}$$

/\*利用Switch计算出结果与空洞率为1和空洞率为3的结果进行求积\*/

(6)计算Post-Global Context模块的结果 $X'$ :

$$x_{post} \leftarrow \text{Conv2d}(\text{AdaptiveAvgPool2d}(P_{out})) + P_{out}$$

/\*将张量 $P_{out}$ 进行平均池化和卷积,再与张量 $P_{out}$ 相加.AdaptiveAvgPool2d为平均池化函数,Conv2d为卷积函数\*/

直方向的中间特征图 $f$ ,如式(8)所示:

$$f = \sigma(F_1([z^h, z^w])) \quad (8)$$

沿着空间维度将 $f$ 分成两个张量 $f^h$ 和 $f^w$ ,再利用两个 $1 \times 1$ 卷积运算 $F_h$ 和 $F_w$ 将特征图 $f^h$ 和 $f^w$ 变换到和输入 $T$ 同样的通道,再经过Sigmoid激励函数得到两个张量 $g^h$ 和 $g^w$ ,如式(9)和式(10)所示:

$$g^h = \sigma(F_h(f^h)) \quad (9)$$

$$g^w = \sigma(F_w(f^w)) \quad (10)$$

其中, $g^h$ 和 $g^w$ 分别经过Sigmoid激励函数( $\sigma$ 表示Sigmoid函数)归一化到0~1范围内的数值,将 $g^h$ 和 $g^w$ 拓展维度,作为原始特征图 $T$ 的注意力权重,得到最终的输出特征 $T'$ ,输出如式(11)所示:

$$T'_c(i, j) = T_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (11)$$

如图5所示,对于输入特征图 $T$ (维度为 $B \times C \times H \times W$ ),分别对水平坐标方向和垂直坐标方向的每个方向做平均池化,然后级联这个方向上的张量,进行卷积运算.

最后沿着空间维度将处理后的张量切成单独的张量,将这两个单独的张量作为注意力权重与原有的输入特征图 $T$ 相乘得到最后的输出.CA的具体算法流程如算法2所示.

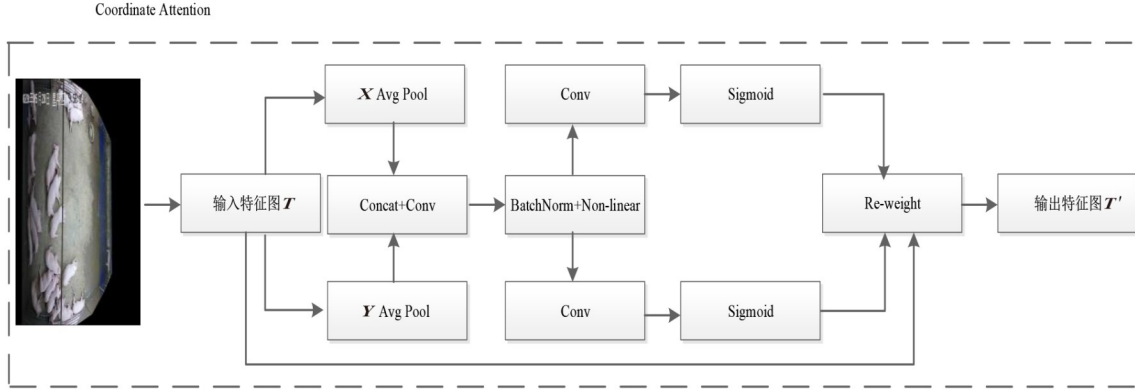


图5 Coordinate Attention 模块

算法2 Coordinate Attention 算法

输入:输入张量  $T$ ,大小为  $B \times C \times H \times W$ ,  $T$  表示 ResNet101 中的特征图.  
 输出:输出张量  $T'$ .

(1)初始化  $C, H, W: H, W \leftarrow T.size()[-3:]$   
 /\*其中  $C, H, W$  分别为特征图的通道数、高、宽\*/

(2)计算  $x_h, x_w: x_h \leftarrow AdaptiveAvgPool2d((1, None))(T)$ ,  
 $x_w \leftarrow AdaptiveAvgPool2d((None, 1))(T)$ ;  
 /\*将张量  $T$  分别沿着水平方向和垂直方向进行平均池化, AdaptiveAvgPool2d 为平均池化函数\*/

(3)计算  $x_{cat}: x_{cat} \leftarrow cat([x_h, x_w], dim=2)$ ;  
 /\*将  $x_h$  和  $x_w$  这两个张量沿着通道进行级联\*/

(4)计算  $P_{out}: P_{out} \leftarrow relu6(BatchNorm2d(Conv2d(x_{cat})))$   
 /\*将步骤(3)中的  $x_{cat}$  进行卷积、归一化函数、激励函数得到结果张量  $P_{out}$ . Conv2d, BatchNorm2d, relu6 分别为卷积函数、归一化函数、激励函数\*/

(5)计算  $x'_h, x'_w: x'_h, x'_w \leftarrow split(P_{out}, [H, W], dim=2)$   
 /\*将步骤(4)中的张量  $P_{out}$  进行切分得到张量  $x'_h, x'_w$ . split 为切片函数\*/

(6)转置  $x'_w: x'_w \leftarrow permute(x'_w, 0, 1, 3, 2)$   
 /\*将步骤(5)中的  $x'_w$  进行转置,转置后的维度与  $x'_h$  一致. permute 为维度转置函数\*/

(7)计算  $P_{out_h}, P_{out_w}: P_{out_h} \leftarrow sigmoid(Conv2d(x'_h)), P_{out_w} \leftarrow sigmoid(Conv2d(x'_w))$  /\* $P_{out_h}$  和  $P_{out_w}$  分别是由  $x'_h$  和  $x'_w$  经过卷积和 sigmoid 函数激活函数得到的张量\*/

(8)计算  $T': T' \leftarrow T \times P_{out_h} \times P_{out_w}$   
 /\*将输入张量  $T$  与步骤(7)中的  $P_{out_h}$  和  $P_{out_w}$  相乘得到最终的结果\*/

3.3 Repulsion Loss 损失函数

Repulsion Loss 主要由三个模块组成,分别是使得预测框与目标框尽可能接近的损失  $L_{Attr}$ 、使预测框  $P$  和周围的目标框  $G$  尽可能远离的损失  $L_{RepGT}$ 、使得预测框  $P$  和周围的预测框尽可能远离  $L_{RepBox}$ , 如式(12)所示:

$$L = L_{Attr} + \alpha L_{RepGT} + \beta L_{RepBox} \quad (12)$$

其中,  $L_{Attr}$  模块的作用是使预测框和它的目标框尽可能

接近.  $L_{Attr}$  采用的是通用目标检测中的回归损失,即  $Smooth_{L1}$  损失函数.  $L_{Attr}$  模块如式(13)所示:

$$L_{Attr} = \frac{\sum_{p \in P_+} Smooth_{L1}(B^p, G^p_{Attr})}{|r_+|} \quad (13)$$

其中,  $r_+$  表示所有正样本的集合,正样本的划分是根据设定的 IoU 阈值来划分的;  $G^p_{Attr}$  为每一个检测框  $P$  匹配一个有最大 IoU 值的真实目标框;  $B^p$  表示检测  $P$  做回归偏移后得到的预测框.

式(13)利用  $Smooth_{L1}$  求边框回归损失,其中  $Smooth_{L1}$  的如式(14)和式(15)所示:

$$L_{loc}(t^u, v) = \sum_{i \in \{x, y, w, h\}} Smooth_{L1}(t^u_i - v_i) \quad (14)$$

$$Smooth_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & otherwise \end{cases} \quad (15)$$

由式(13)和式(14)可以得出  $B^p$  和  $G^p_{Attr}$  的左上角坐标以及宽高  $(x, y, w, h)$ , 分别进行  $Smooth_{L1}$  计算,然后累加求和. 通过该损失函数让预测框与目标框的距离不断接近,从而使预测框与目标框不断接近.

式(12)中,  $L_{RepGT}$  是使预测框  $P$  和周围的目标框  $G$  尽可能远离.  $L_{RepGT}$  如式(16)所示:

$$L_{RepGT} = \frac{\sum_{p \in P_+} Smooth_{ln}(IoG(B^p, G^p_{Rep}))}{|p_+|} \quad (16)$$

$$Smooth_{ln} = \begin{cases} -\ln(1-x) & x \leq s \\ \frac{x-s}{1-s} - \ln(1-s) & x > s \end{cases} \quad (17)$$

其中,  $G^p_{Rep}$  表示周围的目标框是除了已经匹配上的目标框以外的 IoU 最大的目标框.

式(12)中,  $L_{RepBox}$  是使预测框  $p_i$  和周围的预测框  $p_j$  尽可能远离,  $p_i$  和  $p_j$  分别对应不同的预测框,两个预测框之间的距离采用 IoU 进行评估.  $L_{RepBox}$  如式(18)所示:

$$L_{\text{RepBox}} = \frac{\sum_{i \neq j} \text{Smooth}_{\text{In}}(\text{IoU}(B^{p_i}, B^{p_j}))}{\sum_{i \neq j} 1[\text{IoU}(B^{p_i}, B^{p_j}) > 0]} + \varepsilon \quad (18)$$

## 4 实验设计与结果分析

### 4.1 实验环境以及数据集构建

为了分析和评估本文提出的算法的性能,本节在公开的科大讯飞 AI 挑战赛猪只数据集上进行对比实验。

实验中计算机的配置如下:CPU 为 Intel(R) Core (TM) i7-4790K CPU;主频为 4.00 GHz;内存为 8 GB;系统为 Linux 4.4.0-130-generic x86\_64 位;GPU 为 NVIDIA Corporation GM204 [GeForce GTX 970].

数据集 A:从公开科大讯飞 AI 挑战赛猪只数据集中采集了 3 600 张猪只图像,每张图像上大致有 30 多只猪,其中每张图片中猪只的姿势、体型、大小都不尽相同. 使用了 2 880 张作为训练集,720 张作为验证集. 部分数据集样本如图 6 所示。

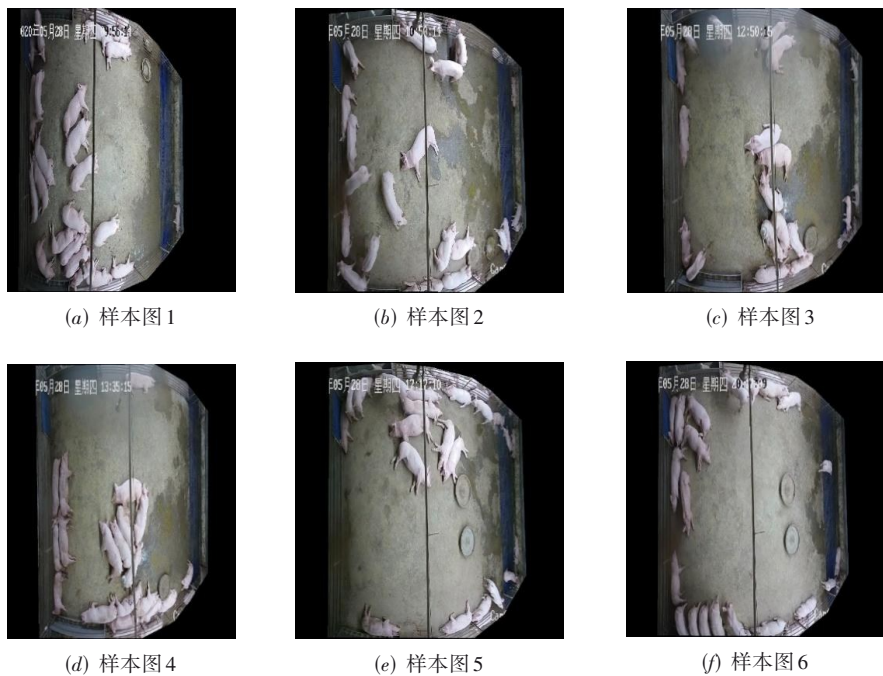


图6 科大讯飞 AI 挑战赛猪只数据集部分样例

数据集 B:从公开科大讯飞 AI 挑战赛猪只数据集中采集了 3 600 张猪只图像,挑选出图像存在遮挡的数据集 845 张作为测试集,用于验证数据集 A 训练出来的模型的精度。

### 4.2 消融实验

#### 4.2.1 评价指标

目标检测的任务是对感兴趣的目标进行定位和分类,通常采用平均准确率(mean Average Precision, mAP)来对目标检测算法优劣进行评价,其计算如式(19)所示:

$$\begin{cases} P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \\ AP_c = \frac{1}{N_c} \sum_{r_c \in R_c} p(r_c) \\ mAP = \frac{1}{N} \sum AP_c \end{cases} \quad (19)$$

其中,TP 表示真阳性样本数,即正确样本预测为正样本,也就是在目标检测中预测框与真实框的 IoU 大于阈值的样本;FP 表示假阳性样本数,即错误样本预测为正样本,也就是在目标检测中预测框与真实框的 IoU 小于阈值的样本;FN 表示假阴性样本数,即正确样本预测为负样本,没有检测到框; $P$  表示精度,指的是所有被预测为正类的结果中预测正确的结果所占的比例; $R$  表示召回率,指的是结果中正样本占全部正样本的比例; $N_c$  表示类别为  $c$  时检测到为正样本的个数; $p(r_c)$  表示类别为  $c$  在召回率为  $r_c$  时的精度; $AP$  表示平均精度,mAP 表示平均精度均值。

#### 4.2.2 不同 Repulsion Loss 损失权重系数的对比实验

本节实验采用 mmdetection 工具进行实验,特征提取网络 backbone 使用 ResNet101,使用 ImageNet 的预训练模型,学习率设置为 0.001 25, Batchsize 设置为 4,训练 50 个 epoch. 基础网络采用 Cascade R-CNN+ ResNet101 进行实验,通过改变 Repulsion Loss 损失函

数的权重系数,即改变  $\alpha$  和  $\beta$  的值,实验结果如表 1 所示.

表 1 验证改变 Repulsion Loss 损失函数的权重系数的 mAP 增长情况

权重系数	mAP/%
0.5,1	94.21
1,1	94.53
1,0.5	94.33

如表 1 所示,当采用 Cascade R-CNN+ResNet101+Repulsion Loss 进行实验,权重系数为 0.5 和 1 时,测试科大讯飞 AI 挑战赛猪只数据集 A 的 mAP 为 94.21%,当权重系数为 1 和 1 时,测试科大讯飞 AI 挑战赛猪只数据集 A 的 mAP 为 94.53%,mAP 值增长了 0.3%. 当权重系数为 1 和 0.5 时,测试科大讯飞 AI 挑战赛猪只数据集 A 的 mAP 为 94.33%,mAP 值又降低了 0.2%. 说明 Repulsion Loss 损失函数的权重系数的大小对 Cascade R-CNN 网络的精度有一定影响.

#### 4.2.3 不同 SAC 空洞率的对比实验

在实验过程中,采用 Cascade R-CNN+ResNet101+SAC 进行实验时,通过改变 SAC 模块中的空洞率在公开的科大讯飞 AI 挑战赛猪只数据集 A 进行测试,mAP 值均有所差异. 当改变 SAC 模块中的空洞率时,得到如表 2 所示的结果.

表 2 验证改变 SAC 空洞率后 mAP 的增长情况

空洞率	mAP/%
1,1	94.12
1,2	94.25
1,3	94.56
2,3	94.29
3,3	94.32

如表 2 所示,当采用 Cascade R-CNN+ResNet101+SAC 进行实验,改变 SAC 空洞率为 1 和 1 时,测试科大讯飞 AI 挑战赛猪只数据集 A 的 mAP 为 94.12%,当空洞率为 1 和 2 时,测试科大讯飞 AI 挑战赛猪只数据集 A 的 mAP 为 94.25%,mAP 增长了 0.13%;当空洞率为 1 和 3 时,测试科大讯飞 AI 挑战赛猪只数据集 A 的 mAP 为 94.56%,mAP 值增长了 0.31%;当空洞率为 2 和 3 时,测试科大讯飞 AI 挑战赛猪只数据集 A 的 mAP 为 94.29%,mAP 值又下降了 0.27%. 说明 SAC 的空洞率的大小对 Cascade R-CNN 网络的精度有一定影响.

### 4.3 对比实验

#### 4.3.1 不同模型的对比实验

在实验过程中,分别在网络中加入可切换空洞卷积、坐标注意力机制以及 Repulsion Loss 损失函数,使用公开的科大讯飞 AI 挑战赛猪只数据集 A 和 B 进行测

试,mAP 值均有所差异. 当加入不同的模块时,得到如表 3 所示的结果.

表 3 分别验证加入 SAC,CA 以及 Repulsion Loss 损失函数后在数据集 A 和数据集 B 的 mAP 增长情况

模型	数据集	mAP/%
Cascade R-CNN+ResNet101	A	94.03
	B	90.21
Cascade R-CNN+ResNet101+SAC	A	94.56
	B	90.48
Cascade R-CNN+ResNet101+SAC+CA	A	95.29
	B	91.56
Cascade R-CNN+ResNet101+SAC+CA+Repulsion Loss	A	96.42
	B	92.34

如表 3 所示,将 SAC,CA 及 Repulsion Loss 分别加入到 Cascade R-CNN 网络进行实验. 当采用 Cascade R-CNN+ResNet101 进行实验时,测试科大讯飞 AI 挑战赛猪只数据集 A 和数据集 B 的 mAP 分别为 94.03% 和 90.21%. 当采用 Cascade R-CNN+ResNet101+SAC 进行实验时,测试科大讯飞 AI 挑战赛猪只数据集 A 和数据集 B 的 mAP 为 94.56% 和 90.48%,mAP 分别增长了 0.53% 和 0.27%. 当采用 Cascade R-CNN+ResNet101+SAC+CA 进行实验时,测试科大讯飞 AI 挑战赛猪只数据集 A 和数据集 B 的 mAP 为 95.29% 和 91.56%,mAP 值又分别增长了 0.73% 和 1.08%. 当采用 Cascade R-CNN+ResNet101+SAC+CA+Repulsion Loss 进行过实验时,测试科大讯飞 AI 挑战赛猪只数据集 A 和数据集 B 的 mAP 分别为 96.42% 和 92.34%,mAP 值又分别增长了 1.13% 和 0.78%. 说明 SAC,CA 以及 Repulsion Loss 对 Cascade R-CNN 网络的精度有一定提高,同时在遮挡数据集上精度也有一定的提高.

#### 4.3.2 不同算法的对比实验

在实验过程中,分别将改进的 Cascade R-CNN 算法与其他目标检测算法进行对比实验,使用 mAP 作为评价标准. 实验结果如表 4 所示.

表 4 分别验证其他目标检测算法和改进 Cascade RCNN 算法的 mAP 增长情况

目标检测算法	mAP/%
Cascade R-CNN+ResNet101+SAC+CA+Repulsion Loss	96.42
Cascade R-CNN+ResNet101	94.03
Faster R-CNN+ResNet101	91.25
Mask R-CNN+ResNet101	92.43
Retinanet+ResNet101	89.21
SSD+ResNet101	86.63

如表 4 所示,将改进的 Cascade R-CNN 算法与 Faster R-CNN,Mask R-CNN,RetinaNet 及 SSD 算法进行

对比实验,当采用改进的 Cascade R-CNN 算法进行实验时,测试科大讯飞 AI 挑战赛猪只数据集 A 的 mAP 为 96.42%,当采用 Faster R-CNN, Mask R-CNN, RetinaNet 及 SSD 算法进行实验时,测试科大讯飞 AI 挑战赛猪只数据集 A 的 mAP 分别为 91.25%, 92.43%, 89.21%, 86.63%, mAP 分别增长了 5.17%, 3.99%, 7.21%, 9.79%。说明改进的 Cascade R-CNN 算法较双阶段算法以及单阶段算法在精度上都有一定提高。

### 4.3.3 改进算法在主流数据集上的对比实验

在实验过程中,分别将改进的 Cascade R-CNN 算法用于 VOC 数据集上进行对比实验,使用 mAP 作为评价标准。实验结果如表 5 所示。

如表 5 所示,将改进的算法用于 VOC2007 数据集进行测试,当使用 Cascade R-CNN+ResNet101 进行实验发现,测试数据集 mAP 为 79.6%。当又加入 SAC 模块后测试数据集 mAP 为 80.1%, mAP 增长了 0.5%。当又加入 CA 模块后,测试数据集 mAP 为 81.3%, mAP 增

表 5 分别验证改进算法在 VOC2007 数据集上的 mAP 增长情况

模型	mAP/%
Cascade R-CNN+ResNet101	79.6
Cascade R-CNN+ResNet101+SAC	80.1
Cascade R-CNN+ResNet101+SAC+CA	81.3
Cascade R-CNN+ResNet101+SAC+CA+Repulsion Loss	81.7
Faster R-CNN+ResNet101	78.8
RetinaNet+ResNet101	79.4

长了 1.2%。当又加入 Repulsion Loss 后,测试数据集 mAP 为 81.7%, mAP 又增长了 0.4%。说明改进的 Cascade R-CNN 算法对 VOC 数据集中目标检测的精度有一定的提高。

### 4.3.4 检测效果

通过在 Cascade R-CNN 网络中加入 SAC, CA 以及 Repulsion Loss 等改进机制,检测猪只数据集的效果比原算法的效果有了一定提升。数据集检测效果如图 7 所示。

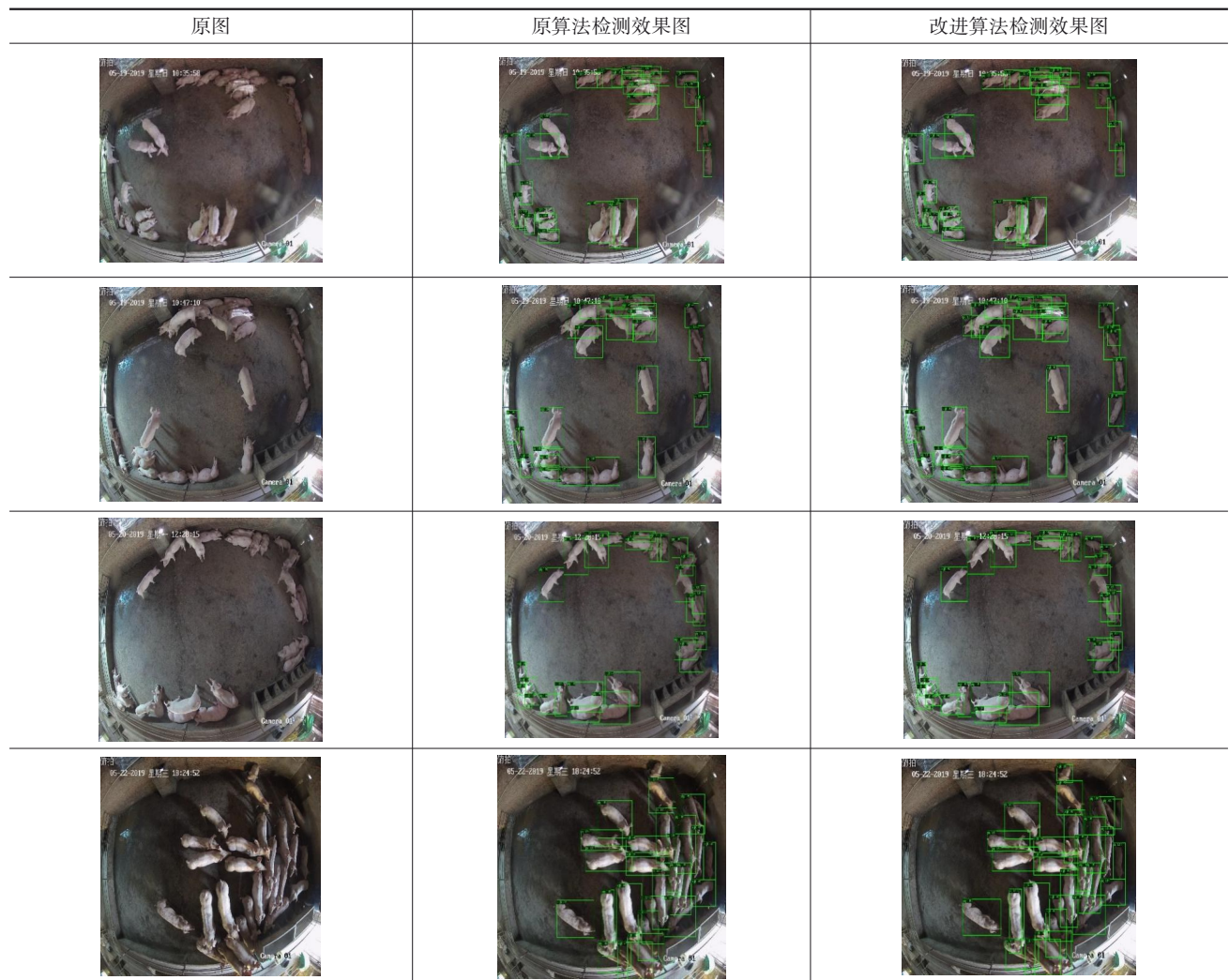


图 7 改进算法检测结果与原算法检测结果对比图

## 5 结束语

本文基于 Cascade R-CNN 网络,在网络结构的主干网络中引入可切换空洞卷积模块和坐标注意力机制来提高主干网络的特征提取能力.同时,在损失函数中引入 Repulsion Loss 损失函数用于解决目标遮挡的问题.与原算法以及其他单阶段和双阶段算法相比,改进的算法在检测精度和解决遮挡问题上有一定的提升.

### 参考文献

- [1] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision - ECCV 2016. Amsterdam: Springer, 2016: 21-37.
- [2] 辛文斌, 郝惠敏, 卜明龙, 等. 基于 ShuffleNetv2-YOLOv3 模型的静态手势实时识别方法[J]. 浙江大学学报(工学版), 2021, 55(10): 1815-1824, 1846.  
XIN W B, HAO H M, BU M L, et al. Static gesture real-time recognition method based on ShuffleNetv2-YOLOv3 model[J]. Journal of Zhejiang University (Engineering Science), 2021, 55(10): 1815-1824, 1846. (in Chinese)
- [3] 曲优, 李文辉. 基于锚框变换的单阶段旋转目标检测方法[J]. 吉林大学学报(工学版), 2022, 52(1): 162-173.  
QU Y, LI W H. Single-stage rotated object detection network based on anchor transformation[J]. Journal of Jilin University (Engineering and Technology Edition), 2022, 52(1): 162-173. (in Chinese)
- [4] 候少麒, 梁杰, 殷康宁, 等. 基于空洞卷积金字塔的目标检测算法[J]. 电子科技大学学报, 2021, 50(6): 843-851.  
HOU S Q, LIANG J, YIN K N, et al. Object detection algorithm based on atrous convolutional pyramid[J]. Journal of University of Electronic Science and Technology of China, 2021, 50(6): 843-851. (in Chinese)
- [5] 张云佐, 李文博, 郑婷婷. 基于 LGC 的反残差目标检测算法[J/OL]. 北京航空航天大学学报, 2021. DOI:10.13700/j.bh.1001-5965.2021.0452.  
ZHANG Y Z, LI W B, ZHENG T T. An Inverse Residual Object Detection Algorithm Based on LGC[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2021. DOI: 10.13700/j.bh.1001-5965.2021.0452. (in Chinese)
- [6] ZHAO Q, LI B Q, LI T W. Target detection algorithm based on improved YOLOv3[J]. Laser & Optoelectronics Progress, 2020, 57(12): 121502.
- [7] ZHAI S P, SHANG D R, WANG S H, et al. DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion[J]. IEEE Access, 2020, 8: 24344-24357.
- [8] IANDOLA F, MOSKEWICZ M, KARAYEV S, et al. DenseNet: Implementing efficient ConvNet descriptor pyramids[EB/OL]. (2014-04-07) [2022-01]. <https://arxiv.org/abs/1404.1869>.
- [9] CHEN X, GUPTA A. An implementation of faster RCNN with study for region sampling[EB/OL]. (2017-02-07) [2022-01]. <https://arxiv.org/abs/1702.02138>.
- [10] LIU Y. An improved faster R-CNN for object detection [C]//2018 11th International Symposium on Computational Intelligence and Design (ISCID). Hangzhou: IEEE, 2019: 119-123.
- [11] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [12] 张焯, 许艇, 冯定忠, 等. 基于难分样本挖掘的快速区域卷积神经网络目标检测研究[J]. 电子与信息学报, 2019, 41(6): 1496-1502.  
ZHANG Y, XU T, FENG D Z, et al. Research on faster RCNN object detection based on hard example mining[J]. Journal of Electronics & Information Technology, 2019, 41(6): 1496-1502. (in Chinese)
- [13] 李晓光, 付陈平, 李晓莉, 等. 面向多尺度目标检测的改进 Faster R-CNN 算法[J]. 计算机辅助设计与图形学学报, 2019, 31(7): 1095-1101.  
LI X G, FU C P, LI X L, et al. Improved faster R-CNN for multi-scale object detection[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(7): 1095-1101. (in Chinese)
- [14] 陈科圻, 朱志亮, 邓小明, 等. 多尺度目标检测的深度学习研究综述[J]. 软件学报, 2021, 32(4): 1201-1227.  
CHEN K Q, ZHU Z L, DENG X M, et al. Deep learning for multi-scale object detection: A survey[J]. Journal of Software, 2021, 32(4): 1201-1227. (in Chinese)
- [15] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6154-6162.
- [16] 陈莹, 龚苏明. 改进通道注意力机制下的人体行为识别网络[J]. 电子与信息学报, 2021, 43(12): 3538-3545.  
CHEN Y, GONG S M. Human action recognition network based on improved channel attention mechanism[J]. Journal of Electronics & Information Technology, 2021, 43(12): 3538-3545. (in Chinese)
- [17] QIAO S Y, CHEN L C, YUILLE A. DetectoRS: Detecting objects with recursive feature pyramid and switchable

atrous convolution[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 10208-10219.

- [18] SENHAJI K, RAMCHOUN H, ETTOUIL M. Training feedforward neural network via multiobjective optimization model using non-smooth L1/2 regularization[J]. Neurocomputing, 2020, 410: 1-11.
- [19] WANG X L, XIAO T T, JIANG Y N, et al. Repulsion loss: Detecting pedestrians in a crowd[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7774-7783.
- [20] 孟子尧, 陈斯佳, 吕天予, 等. 基于深度学习的肾小球病理图像识别与分类[J]. 计算机辅助设计与图形学学报, 2021, 33(6): 947-955.
- MENG Z Y, CHEN S J, LYU T Y, et al. Recognition and classification of glomerular pathological images based on deep learning[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(6): 947-955. (in Chinese)
- [21] 谭建豪, 郑英帅, 王耀南, 等. 基于中心点搜索的无锚框全卷积孪生跟踪器[J]. 自动化学报, 2021, 47(4): 801-812.
- TAN J H, ZHENG Y S, WANG Y N, et al. AFST: Anchor-free fully convolutional Siamese tracker with searching center point[J]. Acta Automatica Sinica, 2021, 47(4): 801-812. (in Chinese)
- [22] 朱煜, 方观寿, 郑兵兵, 等. 基于旋转框精细定位的遥感目标检测方法研究[J]. 自动化学报, 2023, 49(2): 415-424.
- ZHU Y, FANG G S, ZHENG B B, et al. Research on remote sensing target detection method based on fine positioning of rotating frame[J]. Acta Automatica Sinica, 2023, 49(2): 415-424.
- [23] 赵琰, 赵凌君, 匡纲要. 基于注意力机制特征融合网络的 SAR 图像飞机目标快速检测[J]. 电子学报, 2021, 49(9): 1665-1674.
- ZHAO Y, ZHAO L J, KUANG G Y. Attention feature fusion network for rapid aircraft detection in SAR images [J]. Acta Electronica Sinica, 2021, 49(9): 1665-1674. (in Chinese)
- [24] 姜正申, 刘宏志, 付彬, 等. 集成学习的泛化误差和 AUC 分解理论及其在权重优化中的应用[J]. 计算机学报, 2019, 42(1): 1-15.
- JIANG Z S, LIU H Z, FU B, et al. Decomposition theories of generalization error and AUC in ensemble learning with application in weight optimization[J]. Chinese Journal of Computers, 2019, 42(1): 1-15. (in Chinese)

- [25] 李书和, 张奕群, 王东升, 等. 数控机床热误差的建模与预补偿[J]. 计量学报, 1999(1): 49-52.

LI S H, ZHANG Y Q, WANG D S, et al. Modeling and precompensation of thermal errors in CNC machine tools [J]. Acta Metrologica Sinica, 1999(1): 49-52. (in Chinese)

#### 作者简介



张娜 女, 1977年生, 浙江杭州人. 硕士, 副教授. 主要研究方向为智能信息处理.  
E-mail: zhangna@zstu.edu.cn



包梓群 男, 2001年生, 浙江东阳人. 浙江理工大学本科生. 主要研究方向为图像处理、智能信息处理.  
E-mail: 359020134@qq.com



罗源 男, 1995年生, 湖北安陆人. 硕士研究生. 主要研究方向为图像处理、智能信息处理.  
E-mail: 993807182@qq.com



吴彪 男, 1989年生, 浙江杭州人. 博士. 主要研究方向为计算机视觉与模式识别.  
E-mail: biaoowuzg@zstu.edu.cn



涂小妹 女, 1995年生, 湖北黄冈人. 硕士. 主要研究方向为图像处理、智能信息处理.  
E-mail: txm\_95@163.com