

# 随机缩放混合与跨尺度特征增强的任务对齐目标检测算法

王国刚, 李佳琪

(山西大学 物理电子工程学院, 太原 030006)

**摘要:** 针对 TOOD 算法鲁棒性差、特征金字塔顶层丢失部分语义信息、不同尺度特征层存在语义差距的问题, 提出随机缩放混合与跨尺度特征增强的任务对齐目标检测算法; 该算法提出  $2 \times 4$  混合增强方法, 丰富训练样本, 提高模型的泛化性和鲁棒性; 构造多重残差特征增强模块, 自适应融合顶层不同尺度的上下文信息, 减少最高层语义信息的损失; 构建堆叠金字塔卷积模块, 缩小不同尺度特征层之间的语义差距, 提升多尺度特征的融合效果; Pascal VOC 数据集上的实验结果表明, 所提算法的均值平均精度、查准率、查全率分别比 TOOD 算法提高了 3.76%、15.71%、6.28%; 而且该算法的  $F_1$  值与均值平均精度均优于 6 种主流对比算法。

**关键词:** 目标检测; TOOD; 缩放混合; 特征增强; 金字塔卷积

## Task-aligned Object Detection Algorithm Based on Random Scaling Mixture and Cross-Scale Feature Enhancement

WANG Guogang, LI Jiaqi

(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

**Abstract:** Aimed at the disadvantages of poor robustness, some semantic information loss in the top-level feature layer of feature pyramid network in the TOOD algorithm, and existing a semantic gap for feature layers with different scales, a task-aligned object detection algorithm based on random scaling mixture and cross-scale feature enhancement is proposed. The  $2 \times 4$  hybrid augmentation method is used to enrich the training samples, and improve the generalization and robustness of the model. The multiple residual feature enhancement module is constructed to adaptively fuse the context information with different scales at the top level, and reduce the semantic information loss at the highest level. Moreover, the proposed method constructs the stacked pyramid convolution module, reduces the semantic gap among the different scale features, and improves the effect of multi-scale feature fusion. Experimental results on the Pascal VOC data set show that the mean average precision, precision, and recall of the proposed algorithm are 3.76%, 15.71% and 6.28% respectively higher than that of the TOOD algorithm. The presented algorithm is superior to six mainstream comparison algorithms in the  $F_1$  value and mean average precision.

**Keywords:** object detection; TOOD; scaling mixture; feature enhancement; pyramid convolution

### 0 引言

近年来, 随着理论研究的发展及硬件设备的更新迭代, 计算机视觉领域的突破层出不穷。目标检测是计算机视觉中一项基本且富有挑战的任务, 其目的是从自然图像中定位和识别出感兴趣的目标。目标检测已广泛应用于人脸识别<sup>[1-2]</sup>、自动驾驶<sup>[3-4]</sup>、异常检测<sup>[5-8]</sup>等领域。

传统目标检测通常通过人工设计的特征描述算子提取特定任务下的样本特征以检测目标位置和类别<sup>[9]</sup>。此类方法通过滑动窗口的遍历确定目标的大小和位置。对于简单场景下的目标检测任务, 传统方法可以获得较好的检测精度, 故这种方法在早期的目标检测领域曾占据重要地

位。然而, 在一些复杂场景下, 传统目标检测方法往往难以获得令人满意的检测结果。一方面, 人工设计的特征描述算子往往难以全面、准确地捕捉目标在不同场景下的复杂变化; 另一方面, 基于滑动窗口的传统算法对目标位置缺乏针对性, 且容易生成大量冗余的候选框, 这不仅降低了检测准确性, 还增加了时间复杂度。

2012年, 文献[10]提出的 AlexNet 模型展示了卷积神经网络在 CV 领域应用的巨大潜力。2014年, 文献[11]提出的 R-CNN 算法首次将深度学习应用于目标检测领域。与传统方法相比, 采用深度学习技术的目标检测方法可以在初始数据中自动学习目标特征, 并通过端到端的方式预

收稿日期: 2024-03-11; 修回日期: 2024-04-15。

基金项目: 国家自然科学基金(11804209); 山西省自然科学基金(201901D111031, 201901D211173); 山西省高校科技创新计划(2019L0064, 2020L0051)。

作者简介: 王国刚(1977-), 男, 博士, 副教授。

引用格式: 王国刚, 李佳琪. 随机缩放混合与跨尺度特征增强的任务对齐目标检测算法[J]. 计算机测量与控制, 2024, 32(9): 225-233.

测更加精确的位置和类别。这些方法主要分为两类：一类是基于回归的单阶段方法，如 SSD<sup>[12]</sup>和 RetinaNet<sup>[13]</sup>，该类方法直接将样本输入到卷积神经网络以进行目标的分类和定位；另一类是基于候选区域的两阶段方法，如 Faster R-CNN<sup>[14]</sup>和 Mask R-CNN<sup>[15]</sup>，该类方法先生成候选框，然后对生成的候选框进行分类和定位。然而，这些方法在分类和定位任务的学习机制上存在差异，这可能导致两个任务之间的预测存在一定程度的空间错位。

为解决此问题，TOOD<sup>[6]</sup>算法引入面向对齐的学习方法，增加分类和定位任务之间的交互，调整任务不可知的样本分配。这优化了分类和定位任务学习到的特征在空间上的分布，提高了模型的检测性能。TOOD算法旨在解决分类和定位任务学习机制不同的问题。该算法已经在医学图像处理<sup>[17]</sup>、城市交通管理<sup>[18]</sup>和智能驾驶<sup>[19]</sup>领域得到了应用。

TOOD算法首次提出任务对齐概念。任务对齐是指分类和定位任务之间的协调性或同步性，其目的是为了提高这两个任务之间的交互；它主要通过任务对齐检测头（T-Head, task-aligned head）和任务对齐学习（TAL, task alignment learning）实现。一方面，T-Head先采用多层卷积层以学习任务交互特征，并通过层注意力机制动态计算层级任务特征以避免分类和定位任务的特征冲突；T-Head再引入空间概率图和空间偏移图来调整分类得分和目标边界框的空间分布，从而实现这两个任务的联合优化。另一方面，TAL设计了新的样本分配方案和任务对齐损失函数，这联合优化了不同任务的最优锚点，提高了分类和定位的准确性。然而，TOOD算法鲁棒性差，特征金字塔顶层丢失部分语义信息，不同尺度特征层存在语义差距。

针对 TOOD 算法存在的缺陷，提出随机缩放混合与跨尺度特征增强的任务对齐目标检测算法（TRSFCME, task-aligned object detection algorithm based on random scaling mixture and cross-scale feature enhancement）。该算法提出 2×4 混合增强方法，扩充训练样本，提高模型的鲁棒性和泛化能力；设计多重残差特征增强模块，自适应融合多尺度上下文信息，弥补最高层语义信息损失；构造堆叠金字塔卷积模块，捕捉不同尺度特征层之间的相互关系，提高多尺度特征融合的效果。实验结果表明，TRSFCME 算法的查准率、查全率、F<sub>1</sub> 值、均值平均精度均优于 6 种主流对比算法。

### 1 TOOD 算法概述

与最近提出的单阶段目标检测器<sup>[20-21]</sup>类似，TOOD 采用“backbone-FPN-head”结构，其网络配置参见表 1。

表 1 TOOD 的网络配置

骨干网络	特征融合网络	预测网络
ResNet-50	FPN	T-head

由表 1 可知，TOOD 的骨干网络和特征融合网络分别为 ResNet-50 和 FPN。在特征提取方面，ResNet-50 通过残差网

络引入恒等映射等操作，有效地解决了由于网络深度增加而导致的过拟合、梯度消失等问题。这使得 TOOD 算法可以从更深的网络中提取更精确的特征信息。在特征融合方面，FPN 利用横向连接和自上而下的结构以融合不同尺度的上下文信息。这使得低层特征层的语义特征得到了增强。

通过 TAL 机制，T-head 预测网络自动调整分类得分和预测框来对齐分类和定位任务，从而输出分类得分最高且预测框更精确的目标检测图。如图 1 所示。

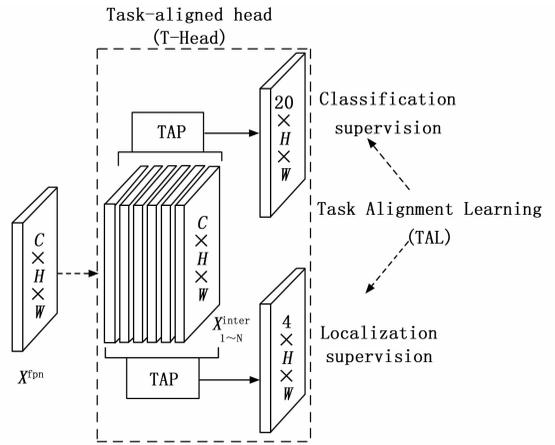


图 1 预测网络

## 2 随机缩放混合与跨尺度特征增强的任务对齐目标检测算法

### 2.1 2×4 混合增强方法

为使模型具有更强的鲁棒性，提出 2×4 混合增强方法。随机选取 8 个训练样本，随机等分为两组，随机裁剪混合缩放后的组内样本。再将每组生成的样本加权求和，以获取混合增强后的新样本。

设第  $n$  组中的第  $m$  个样本为  $A_{mn}$ ，按最大尺度等比例缩放后的样本为  $A'_{mn}$ 。若  $A_{mn}$  的高和宽分别为  $I_{mn}$  和  $J_{mn}$ ， $A'_{mn}$  的高和宽分别为  $I'_{mn}$  和  $J'_{mn}$ ，则它们之间的关系如公式 (1) ~ (2) 所示。

$$\begin{cases} H_n = \max_{m \in \{1,2,3,4\}} I_{mn} \\ W_n = \max_{m \in \{1,2,3,4\}} J_{mn} \end{cases} \quad (1)$$

$$\begin{cases} r_{mn} = \max\left\{\frac{H_n}{I_{mn}}, \frac{W_n}{J_{mn}}\right\} \\ I'_{mn} = r_{mn} H_n \\ J'_{mn} = r_{mn} W_n \\ m \in \{1,2,3,4\} \end{cases} \quad (2)$$

其中： $n \in \{1,2\}$ 。

$A'_{mn}$  的高和宽分别缩放  $\frac{1}{r_{mn}}$  倍后，获得新样本  $A''_{mn}$ 。随机变量  $\mu, \nu$  分别表示  $H_n \times W_n$  混合样本  $Z_n$  行和列的归一化混合分界点。设  $\mu$  和  $\nu$  均服从  $(0.5, 0.6)$  上的均匀分布，其概率密度函数如公式 (3) 所示。混合样本  $Z_n$  可由公式 (4) 得到：

$$f(x) = \begin{cases} 10 & 0.5 < x < 0.6 \\ 0 & \text{其它} \end{cases} \quad (3)$$

$$Z_n = A''_{n1} \odot \begin{pmatrix} I_{\lfloor \mu H_n \rfloor, \lfloor \nu W_n \rfloor} & 0 \\ 0 & 0 \end{pmatrix} + A''_{n2} \odot \begin{pmatrix} 0 & I_{\lfloor \mu H_n \rfloor, \lfloor (1-\nu) W_n \rfloor} \\ 0 & 0 \end{pmatrix} + A''_{n3} \odot \begin{pmatrix} 0 & 0 \\ I_{\lfloor (1-\mu) H_n \rfloor, \lfloor \nu W_n \rfloor} & 0 \end{pmatrix} + A''_{n4} \odot \begin{pmatrix} 0 & 0 \\ 0 & I_{\lfloor (1-\mu) H_n \rfloor, \lfloor (1-\nu) W_n \rfloor} \end{pmatrix} = \begin{pmatrix} (A''_{n1})_{\lfloor \mu H_n \rfloor, \lfloor \nu W_n \rfloor} & (A''_{n2})_{\lfloor \mu H_n \rfloor, \lfloor (1-\nu) W_n \rfloor} \\ (A''_{n3})_{\lfloor (1-\mu) H_n \rfloor, \lfloor \nu W_n \rfloor} & (A''_{n4})_{\lfloor (1-\mu) H_n \rfloor, \lfloor (1-\nu) W_n \rfloor} \end{pmatrix} \quad (4)$$

其中:  $\odot$  表示矩阵点乘,  $I$  为全 1 矩阵,  $n \in \{1, 2\}$ 。

不妨设  $H_1 > H_2, W_1 > W_2$ 。令  $M = \max\{H_1, H_2\}, N =$

$\max\{W_1, W_2\}, Z'_2 = \begin{pmatrix} Z_2 & 0 \\ 0 & 0 \end{pmatrix}$ 。先利用凸组合的方式对  $Z_1、$

$Z'_2$  加权求和, 再对此和的每个元素向下取整, 以生成尺度为  $M \times N$  的样本  $Y$ , 如公式(5) ~ (6) 所示。凸组合的权重  $\lambda$  是一个随机变量, 服从参数为 1.5 的 *Beta* 概率分布, 简记为  $\lambda \sim \text{Beta}(1.5, 1.5)$ 。

$$Y' = \lambda Z_1 + (1-\lambda) \begin{pmatrix} Z_2 & 0 \\ 0 & 0 \end{pmatrix} \quad (5)$$

$$Y = \lfloor Y' \rfloor \quad (6)$$

设训练样本  $A_{mn}$  的标签为  $l_{mn}$ , 则  $2 \times 4$  混合后得到的样本标签  $\bar{l}$  可由公式 (7) 求得。

$$\begin{aligned} \bar{l} = & \lambda\mu\nu l_{11} + \lambda\mu(1-\nu)l_{12} + \lambda\nu(1-\mu)l_{13} + \\ & \lambda(1-\mu)(1-\nu)l_{14} + \mu\nu(1-\lambda)l_{21} + \\ & \mu(1-\lambda)(1-\nu)l_{22} + \nu(1-\lambda)(1-\mu)l_{23} + \\ & (1-\lambda)(1-\mu)(1-\nu)l_{24} \end{aligned} \quad (7)$$

### 2.2 多重残差特征增强模块

TOOD 特征融合网络中, C5 层经一系列卷积和激活生成高层特征层  $P_5、P_6、P_7$ , 并采用自上而下的方式逐渐与低层融合。一方面, 由于融合了不同尺度的上下文信息, 低层特征层的语义特征得到了增强; 另一方面, C5 层通道数减少, 丢失了一部分信息, 且未融合低层特征, 只包含单一尺度的上下文信息。

针对 C5 层信息丢失, 上下文信息尺度单一的问题, 构

建多重残差特征增强模块 (MRFEM, multiple residual feature enhancement module)。该模块采用残差特征增强 (RFE, residual feature enhancement) 方法以自适应融合来自 C5 层的不同尺度的上下文信息。这丰富了 TRSFCME 特征融合网络的多尺度语义信息, 提高了检测性能。RFE 如图 2 所示。

若  $\alpha_k、Z$  分别为缩放因子和 RFE 输入特征层, AAP 为自适应平均池化操作, 则多尺度特征层  $P_k$  可由公式 (8) 求得。 $P_k$  先经卷积和上采样操作得到  $X_k$ , 再通过拼接、卷积和激活操作生成权重  $W_k^1$ , 如公式 (9) ~ (10) 所示。重复拼接  $W_k^1$  后, 得权重  $W_k$ 。 $W_k$  与  $X_k$  点乘求和以获得 RFE 的输出  $Y$ , 见公式 (11) ~ (12):

$$P_k = AAP(Z, \alpha_k) \quad (8)$$

$$X_k = \text{Upsample}[\text{conv}_{(1,1)}(P_k)] \quad (9)$$

$$W_k^1 = \text{sigmoid}\{\text{conv}_{(3,1)}[\text{conv}_{(1,1)}(\text{concat}(X_1, X_2, X_3))]\} \quad (10)$$

$$W_k = \begin{pmatrix} W_k^1 \\ W_k^1 \\ \vdots \\ W_k^1 \end{pmatrix} \quad (11)$$

$$Y = \sum_{k=1}^3 X_k \odot W_k \quad (12)$$

其中:  $Z \in R^{C \times h \times w}, P_k \in R^{C \times \alpha_k h \times \alpha_k w}, X_k \in R^{C' \times h \times w}, W_k^1 \in R^{1 \times h \times w}, W_k \in R^{C' \times h \times w}, Y \in R^{C' \times h \times w}, k \in \{1, 2, 3\}$ ;  $\text{conv}_{(i,1)}$  表示步长为 1 的  $i \times i$  卷积。

### 2.3 堆叠金字塔卷积模块

特征金字塔网络提取不同尺度的特征以聚合不同层的上下文信息。然而, 基于 FPN 的目标检测方法很少融合不同尺度特征层之间的信息。针对此类方法几乎未涉及多尺度特征层之间相关性的问题, 构建堆叠金字塔卷积 (SP-Conv, stacked pyramid convolutions) 模块, 如图 3 所示。该模块构造金字塔卷积 (PConv, pyramid convolutions) 以捕捉不同尺度特征层之间的相互关系。

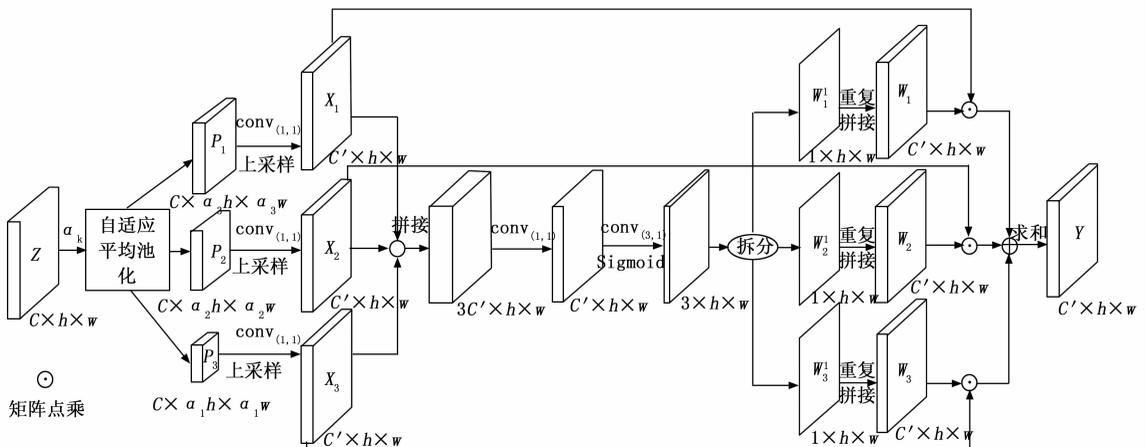


图 2 RFE 示意图

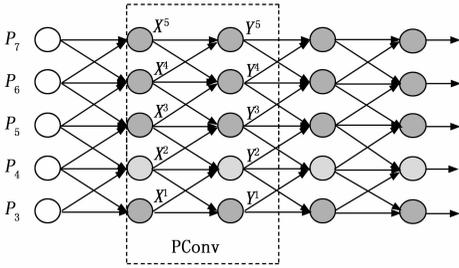


图 3 SPCConv 模块结构图

$P_3、P_4、P_5、P_6、P_7$  依次通过 4 个串联的 PConv 组件以得到 SPCConv 模块的输出。令  $\delta_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}$ 。设  $X^l、Y^l$  分别为 PConv 组件第  $l$  层的输入和输出特征层。利用不同步长的卷积操作, PConv 组件将当前层  $X^l$  及其上下两层  $X^{l+1}、X^{l-1}$  的特征信息进行融合, 得到输出特征层  $Y^l$ , 如公式 (13) 所示:

$$Y^l = \delta_{l5} \cdot [\text{conv}_{(3,0.5)}(X^{l+1})] + \text{conv}_{(3,1)}(X^l) + \delta_{l1} \cdot [\text{conv}_{(3,2)}(X^{l-1})] \quad (13)$$

其中:  $l \in \{1, 2, 3, 4, 5\}$ ;  $\text{conv}_{(3,j)}$  表示步长为  $j$  的  $3 \times 3$  卷积。

步长为 0.5 的  $3 \times 3$  卷积可通过步长为 1 的  $3 \times 3$  卷积和双线性上采样实现, 故可由公式 (14) 表达 PConv 组件的信息传递过程:

$$Y^l = \delta_{l5} \cdot \{\text{Upsample}[\text{conv}_{(3,1)}(X^{l+1})]\} + \text{conv}_{(3,1)}(X^l) + \delta_{l1} \cdot [\text{conv}_{(3,2)}(X^{l-1})] \quad (14)$$

SPCConv 模块利用 PConv 组件来探索不同尺度特征层之间的相互作用。这缩小了特征层之间的语义差距, 提升了多尺度特征融合的效果。

### 2.4 TRSFCME 网络模型

TRSFCME 算法的网络模型如图 4 所示。提出  $2 \times 4$  混

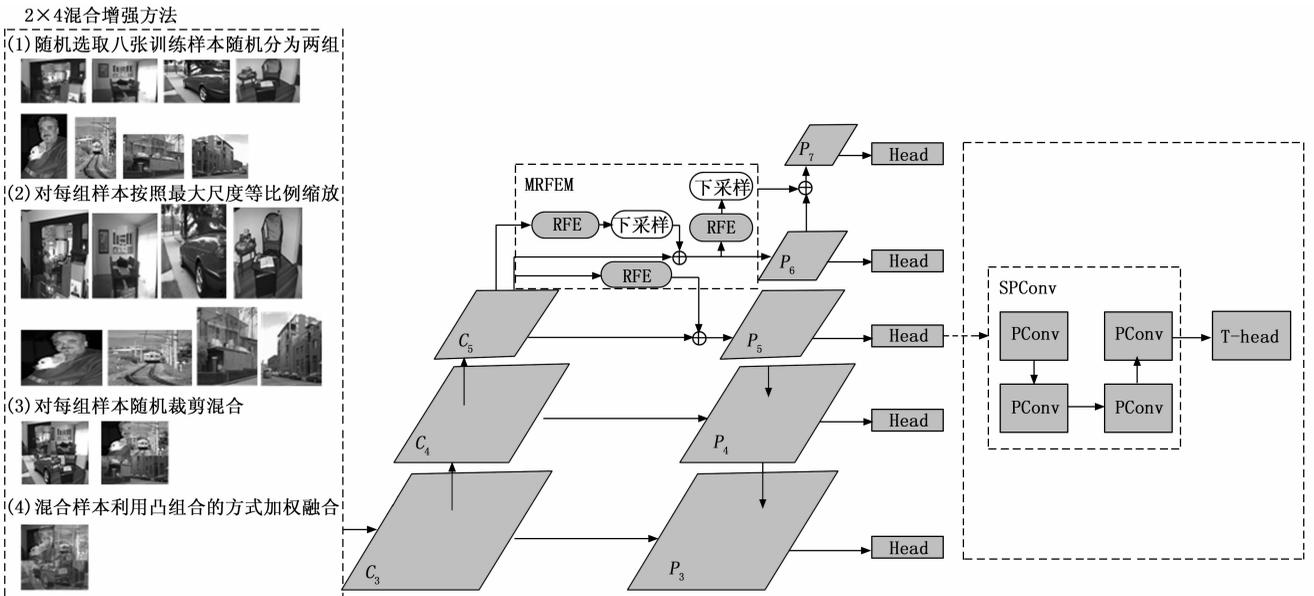


图 4 TRSFCME 网络模型示意结构

合增强方法, 丰富训练样本, 提高模型的鲁棒性与泛化性; 构建多重残差特征增强模块, 自适应融合最高层不同尺度的上下文信息, 减少顶层语义信息的损失; 设计堆叠金字塔卷积模块, 缩小不同尺度特征层之间的语义差距, 提升多尺度特征的融合效果。

## 3 实验结果及分析

### 3.1 实验细节与数据集

实验采用的 CPU 为 Intel (R) Xeon (R) Platinum 8255C CPU @ 2.50 GHz, GPU 为 NVIDIA GeForce RTX 3080, 深度学习框架为 Pytorch 1.7.0、MMDetection 2.21.0<sup>[32]</sup>。

训练阶段, 采用 SGD 算法优化网络参数。epoch、Batch size 分别设置为 12 和 4, 动量因子、权重衰减因子、初始学习率分别设置为 0.9、0.000 1、 $2.5 \times 10^{-3}$ 。初始 500 次迭代, 学习率以线性方式预热, 预热系数为 0.001; 且在第 8、第 11 个 epoch 结束时, 学习率调整为上一轮次的 0.1 倍。

实验采用的训练集为 Pascal VOC 2007 和 Pascal VOC 2012 的训练集和验证集, 测试集为 Pascal VOC 2007 的测试集。

### 3.2 评价指标

实验选取  $F_1$  值、平均精度 (AP, average precision)、均值平均精度 (mAP, mean average precision)、查准率 (P, precision)、查全率 (R, Recall)、浮点运算次数 (FLOPs, floating point operations) 参数大小作为评价指标; 并绘制  $P-R$  曲线和  $F_1$  柱形图来比较不同算法的检测性能。

### 3.3 定量分析

表 2 给出 TRSFCME 算法与其它 6 种算法的  $mAP$  和各类别 AP 的对比情况。由表 2 可知, TRSFCME 算法 16 个类别的 AP 均高于 6 种对比算法; 且 TRSFCME 算法的  $mAP$  达

到了 82.02%, 分别比 ATSS、Faster R-CNN、Dynamic Head、Generalized Focal Loss、Sparse R-CNN、TOOD 算法的  $mAP$  提升了 7.01%、2.89%、2.01%、5.85%、3.88%、3.76%。

表 2 7 种算法精度对比

算法	$mAP$ /%	各类别平均精度 ( $AP$ /%)						
		Aero	Bicycle	Bird	Boat	Bottle	Bus	Car
ATSS	75.01	80.43	80.75	73.14	61.63	63.53	82.00	86.44
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		85.71	58.69	82.54	67.29	81.20	83.32	81.73
		Person	P-plant	Sheep	Sofa	Train	Tv	
		82.80	42.51	77.52	72.60	81.68	74.80	
Faster R-CNN	79.13	80.47	86.18	79.63	69.81	68.24	86.02	87.89
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		88.33	63.50	<b>87.05</b>	74.45	<b>88.35</b>	<b>88.70</b>	80.54
		Person	P-plant	Sheep	Sofa	Train	Tv	
		80.02	52.44	79.09	79.70	83.69	78.64	
Dynamic Head	80.01	86.50	83.25	78.90	72.64	67.69	85.53	88.55
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		88.61	64.70	85.74	75.72	85.48	87.54	83.70
		Person	P-plant	Sheep	Sofa	Train	Tv	
		84.96	53.06	83.55	77.33	87.19	79.69	
Generalized Focal Loss	76.17	83.17	82.04	75.15	65.03	64.57	84.10	86.55
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		87.01	58.63	81.52	67.52	84.55	85.59	80.12
		Person	P-plant	Sheep	Sofa	Train	Tv	
		82.99	50.04	76.07	71.61	82.25	74.84	
Sparse R-CNN	78.14	84.20	84.30	78.61	67.93	63.09	81.62	86.39
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		87.77	61.82	83.88	73.25	85.60	85.16	84.47
		Person	P-plant	Sheep	Sofa	Train	Tv	
		82.37	49.14	81.31	80.06	85.29	76.55	
TOOD	78.26	81.02	84.53	76.35	68.46	67.83	84.96	87.09
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		87.27	63.86	83.10	72.20	85.64	85.79	84.22
		Person	P-plant	Sheep	Sofa	Train	Tv	
		84.56	47.34	75.88	<b>85.63</b>	85.29	77.81	
TRSF CME	<b>82.02</b>	88.93	<b>86.80</b>	<b>83.30</b>	<b>73.89</b>	<b>72.25</b>	<b>87.23</b>	<b>88.81</b>
		Cat	Chair	Cow	D-table	Dog	Horse	M-bike
		<b>88.68</b>	<b>68.03</b>	86.02	<b>78.34</b>	87.19	88.21	86.12
		Person	P-plant	Sheep	Sofa	Train	Tv	
		<b>85.97</b>	<b>54.66</b>	<b>83.63</b>	79.39	<b>87.87</b>	<b>83.66</b>	

注: 加粗数字为最优结果, 带下划线的数字为次优结果。

为比较 7 种算法的综合性能, 实验采用 P-R 曲线和  $F_1$  柱形图进行对比, 如图 5~6 所示。

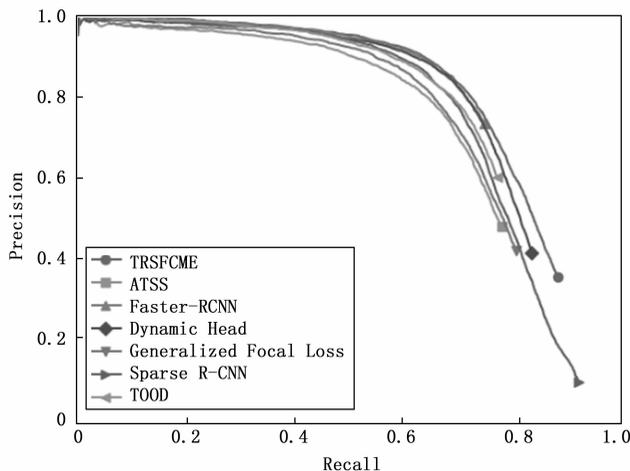


图 5 7 种算法 P-R 曲线比较

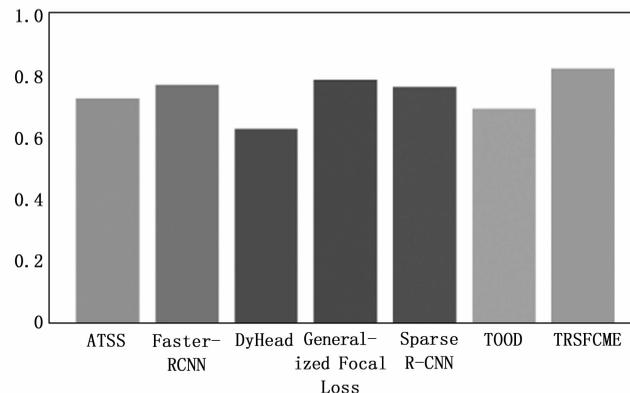


图 6 7 种算法  $F_1$  值比较

P-R 曲线越靠近右上角, 算法性能越好。由图 5 可看出, TRSF CME 算法的 P-R 曲线最靠近右上角。故 TRSF CME 算法的检测性能优于其它 6 种对比算法。

$F_1$  值越大, 算法综合性能越好。由图 6 可看出, TRSF CME 算法的  $F_1$  值最大, 综合性能最优。

表 3 给出七种算法在五项评价指标上的结果。由表 3 可知, 七种算法中, TRSF CME 算法的均值平均精度、 $F_1$  值、查全率、查准率均为最优。

表 3 7 种算法的客观评价指标对比

算法	$mAP$ /%	参数个数/M	$F_1$	$P$ /%	$R$ /%
TRSF CME	<b>82.02</b>	38.50	<b>0.82</b>	<b>72.21</b>	<b>94.86</b>
ATSS	75.01	<u>31.93</u>	0.72	60.82	88.21
Faster R-CNN	79.13	41.22	0.63	49.21	87.50
Dynamic Head	<u>80.01</u>	38.71	0.77	67.01	90.49
Generalized Focal Loss	76.17	32.08	<u>0.78</u>	69.42	88.99
Sparse R-CNN	78.14	105.97	0.76	63.77	94.03
TOOD	78.26	<b>31.84</b>	0.69	56.50	88.58

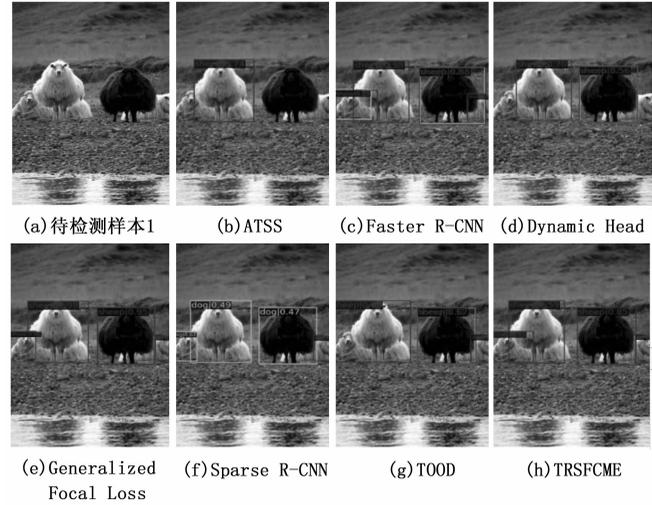
注: 加粗数字和下划线数字分别表示最优、次优结果。

### 3.4 定性分析

为评估算法的检测性能, 在 Pascal VOC 2007 测试集上

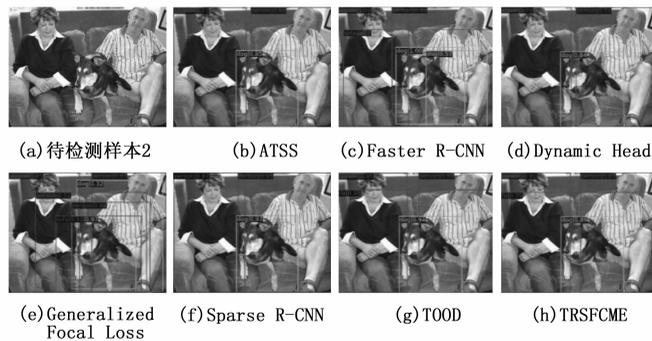
进行定性分析。图 7~11 为 7 种算法分别对 5 个待检测样本的检测结果。

待检测样本 1 包含 4 个 sheep 目标。该样本场景存在遮挡现象，实验结果如图 7 所示。由图 7 可知，ATSS、Dynamic Head、TOOD 算法漏检了 sheep 目标；Faster-RCNN、Dynamic Head 和 TOOD 算法存在重复检测现象；Sparse R-CNN 算法将 sheep 目标错检为 dog。只有 Generalized Focal Loss 和 TRSFCME 算法没有出现重复检测、错检、漏检问题。



(a) 为待检测样本 1, (b) ~ (h) 为不同算法的实验结果  
图 7 检测结果

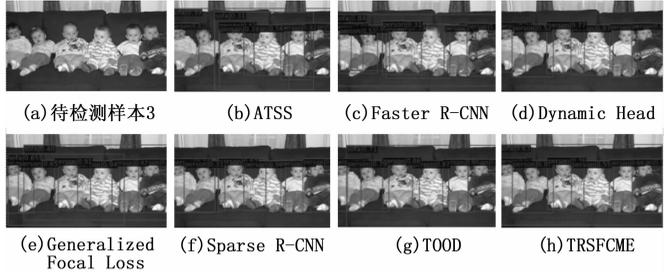
待检测样本 2 包含 2 个人物目标、1 个 dog 目标和 1 个 sofa 目标。样本场景存在目标重叠、遮挡现象，实验结果如图 8 所示。由图 8 可知，ATSS、Dynamic Head 和 Sparse R-CNN 算法均漏检了 sofa 目标；Faster R-CNN 算法存在重复检测问题；Generalized Focal Loss 算法存在定位不准和重复检测现象；只有 TOOD 和 TRSFCME 算法没有出现漏检、重复检测、定位不准确的问题。



(a) 为待检测样本 2, (b) ~ (h) 为不同算法的实验结果  
图 8 检测结果

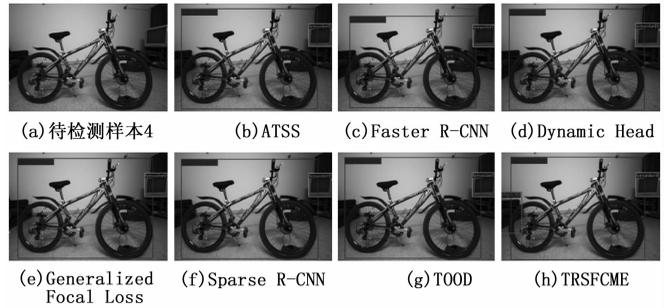
待检测样本 3 包含 6 个人物目标和 1 个 sofa 目标。样本场景中，目标个数较多，存在遮挡、重叠问题，实验结果如图 9 所示。由图 9 可知，ATSS、Faster R-CNN 和 Generalized Focal Loss 算法均存在重复检测问题；ATSS 算

法漏检了 person 目标，Dynamic Head 和 Sparse R-CNN 算法均漏检了 sofa 目标；Faster R-CNN 和 TOOD 算法分别对 person 目标和 sofa 目标的定位存在偏差；TRSFCME 算法无重复检测、漏检和定位不准确的问题，检测效果最优。



(a) 为待检测样本 3, (b) ~ (h) 为不同算法的实验结果  
图 9 检测结果

待检测样本 4 包含 1 个 bicycle 目标、1 个 tvmonitor 目标和 1 个 sofa 目标。样本场景中，光照不足、目标尺度变化大，实验结果如图 10 所示。由图 10 可知，ATSS、Faster R-CNN、Generalized Focal Loss、Dynamic Head、Sparse R-CNN 和 TOOD 算法均漏检了 sofa 目标。TRSFCME 算法没有出现漏检问题，故检测效果最好。



(a) 为待检测样本 4, (b) ~ (h) 为不同算法的实验结果  
图 10 检测结果

待检测样本 5 包含 5 个人物目标和 2 个 boat 目标。该样本场景复杂，尺度变化大，实验结果如图 11 所示。由图 11 可知，Dynamic Head 算法漏检了 person 目标；ATSS 算法漏检了 person 和 boat 目标；Faster R-CNN、Generalized Focal Loss、Sparse R-CNN 和 TOOD 算法均漏检了 boat 目标。只有 TRSFCME 算法没有出现漏检现象，检测效果最好。

### 3.5 消融实验

为评估 TRSFCME 算法中  $2 \times 4$  混合增强方法、MR-FEM、SPConv 模块的有效性，进行了 4 组消融实验，实验结果如表 4~7 所示。

表 4 给出了  $2 \times 4$  混合增强方法、MRFEM、SPConv 模块的消融实验结果。由表 4 第 1、8 行可知，TOOD 算法在 Pascal VOC 2007 测试集上的  $mAP$  为 78.26%，融合  $2 \times 4$  混合增强方法、MRFEM 和 SPConv 模块的 TRSFCME 算法的  $mAP$  为 82.02%，TRSFCME 算法的  $mAP$  比 TOOD 算法提高了 3.76%。由表 4 第 5、8 行可知，缺少  $2 \times 4$  混

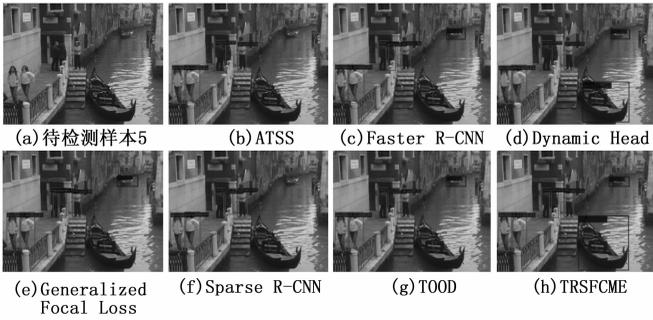


图 11 (a) 为待检测样本 5, (b) ~ (h) 为不同算法的实验结果

合增强方法时,  $mAP$  下降了 1.22%, 这说明  $2 \times 4$  混合增强方法丰富了训练样本的信息, 提高了模型的鲁棒性和泛化能力。由表 4 第 6、8 行可看出, 缺少 MRFEM 时,  $mAP$  下降了 0.35%, 这说明 MRFEM 减少了顶层的语义信息损失, 丰富了特征融合网络的多尺度上下文信息。由表 4 第 7、8 行可看出, 缺少 SPConv 模块时,  $mAP$  下降了 0.33%, 这说明 SPConv 模块缩小了不同尺度特征层之间的语义差距, 提升了多尺度特征的融合效果。由表 4 第 1~4 行可知, 把  $2 \times 4$  混合增强方法、MRFEM 和 SPConv 模块分别引入到 TOOD 算法, 算法的  $mAP$  分别提升了 1.79%、0.39%、2.14%, 这说明  $2 \times 4$  混合增强方法、MRFEM 和 SPConv 模块对网络的性能都有所提升, 且 SPConv 模块的构建对网络性能的提升贡献最大。

表 4 TRSFCME 算法的消融实验

SPConv	MRFEM	$2 \times 4$ 混合增强方法	$mAP/\%$
			78.26
✓			80.40
	✓		78.65
		✓	80.05
✓	✓		80.80
✓		✓	81.67
	✓	✓	81.69
✓	✓	✓	82.02

表 5 给出了  $2 \times 4$  混合增强方法的消融实验结果。由表 5 可知, 与基准算法相比, Mixup、Mosaic、 $2 \times 4$  混合增强方法使 TOOD 算法的  $mAP$  分别提高了 0.63%、0.77%、1.79%。可以看出, 3 种数据增强方法中,  $2 \times 4$  混合增强方法的效果最好。这验证了  $2 \times 4$  混合增强方法的有效性。

表 5  $2 \times 4$  混合增强方法的消融实验

模型	$mAP/\%$
TOOD	78.26
TOOD+Mixup	78.89
TOOD+Mosaic	79.03
TOOD+ $2 \times 4$ 混合增强方法	80.05

表 6 给出了 MRFEM 的消融实验结果。由表 6 第 1、2、3 行可知, 与基准网络相比, 全局平均池化使  $mAP$  提高了

0.03%, 而全局最大池化使  $mAP$  降低了 0.19%, 这表明平均池化比最大池化更鲁棒。由表 6 第 3、4 行可知, 相较于全局平均池化, AAP 使  $mAP$  提高了 0.04%, 这说明 AAP 能保留更多有利于目标检测的信息。

由表 6 第 4、7 行可知, 相较于 sum, RFE 使  $mAP$  提高了 0.04%, 这说明 RFE 模块的多尺度特征的融合方式优于 sum。由表 6 第 5~14 行可知, 当 MRFEM 和 RFE 的  $\alpha$  设置相同时, 采用 MRFEM 的  $mAP$  均高于 RFE, 这表明 MRFEM 比 RFE 能更好地融合多尺度语义信息。由表 6 第 3、4 部分可知,  $\alpha$  的取值影响算法性能; 随着  $\alpha$  设置数量的增加, 算法性能逐渐提升, 复杂度逐渐增加。综合考虑算法性能和复杂度的影响, TRSFCME 算法采用 0.1、0.2、0.3 的  $3\alpha$  设置。

表 6 MRFEM 的消融实验

融合方式	池化方式	$\alpha$	$mAP/\%$	FLOPs(G)
基准网络			78.26	110.32
sum	GMP		78.07	110.52
sum	GAP		78.29	110.52
sum	AAP	0.1,0.2,0.3	78.33	110.35
RFE	AAP	0.1	78.29	111.02
RFE	AAP	0.1,0.2	78.30	111.72
RFE	AAP	0.1,0.2,0.3	78.37	112.47
RFE	AAP	0.1,0.2,0.3,0.4	78.40	114.22
RFE	AAP	0.1,0.2,0.5	78.32	112.53
MRFEM	AAP	0.1	78.31	111.04
MRFEM	AAP	0.1,0.2	78.45	111.76
MRFEM	AAP	0.1,0.2,0.3	78.65	112.50
MRFEM	AAP	0.1,0.2,0.3,0.4	78.77	114.57
MRFEM	AAP	0.1,0.2,0.5	78.43	112.57

注: 该模块的融合方式为 sum 或 RFE 时, 仅与 P5 层融合; AAP 为自适应平均池化; GMP 为全局最大池化; GAP 为全局平均池化;  $\alpha$  为不同缩放比例的设置。

表 7 给出了 SPConv 模块的消融实验结果。由表 7 可知, 与 FPN 相比, PA-Net、Libra、SPConv 的  $mAP$  分别提升了 0.83%、0.47%、1.78%, FLOPs 分别增加了 3.58 G、0.98 G、0.32 G。可以看出, 改进 FPN 的各融合模块中, 构建的 SPConv 模块复杂度最小, 性能最好。这验证了 SPConv 模块的有效性。

表 7 SPConv 模块的消融实验

融合方式	$mAP/\%$	FLOPs(G)
FPN	78.26	110.32
PA-Net	79.09	113.90
Libra	78.73	111.62
SPConv	80.40	110.64

#### 4 结束语

本文提出随机缩放混合与跨尺度特征增强的任务对齐目标检测算法。该算法提出  $2 \times 4$  混合增强方法, 丰富训练

样本的信息,提高模型的鲁棒性和泛化能力。构建多重残差特征增强模块,自适应融合多尺度上下文信息,弥补顶层特征层语义信息的损失;设计堆叠金字塔卷积模块,减小特征层之间的语义差异,提高多尺度特征融合的性能。实验结果表明,所提算法在 Pascal VOC 数据集上的查准率、查全率、 $F_1$  值、均值平均精度均优于 6 种主流对比算法。

#### 参考文献:

- [1] YAN C, ZHANG Y, ZHANG Q, et al. Privacy-preserving on-line AutoML for domain-specific face detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 4134 - 4144.
- [2] XU Y, YIN Y, JIANG L, et al. TransEditor: transformer-based dual-space gan for highly controllable facial editing [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 7673 - 7682.
- [3] HU Y, YANG J, CHEN L, et al. Planning-oriented autonomous driving [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 17853 - 17862.
- [4] HOU Y. An end-to-end convolutional neural network model for autonomous driving [C] // 2023 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), IEEE, 2023: 360 - 365.
- [5] ZHANG X, LI S, LI X, et al. DeSTSeg: segmentation guided denoising student-teacher for anomaly detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 3914 - 3923.
- [6] BATZNER K, HECKLER L, KÖNIG R. EfficientAD: accurate visual anomaly detection at millisecond-level latencies [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 128 - 138.
- [7] JEONG J, ZOU Y, KIM T, et al. WinCLIP: zero-/Few-Shot anomaly classification and segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 19606 - 19616.
- [8] JIANG A, ZHANG W Q, DENG Y, et al. Unsupervised anomaly detection and localization of machine audio: a gan-based approach [C] // ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023: 1 - 5.
- [9] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32 (9): 1627 - 1645.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the Association for Computing Machinery, 2017, 60 (6): 84 - 90.
- [11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2014: 580 - 587.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // Proceedings of the European Conference on Computer Vision, 2016: 21 - 37.
- [13] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980 - 2988.
- [14] REN S, HE K, 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, 2015, 39 (6): 1137 - 1149.
- [15] KAIMINGH, GEORGIAG, PIOTRD, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 37 (12): 2663 - 2673.
- [16] FENG C, ZHONG Y, GAO Y, et al. TOOD: task-aligned one-stage object detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 3490 - 3499.
- [17] TURECKOVA A, TURECEK T, OPLATKOVA Z K. ICIP 2022 Challenge: PEDCMI, TOOD Enhanced by Slicing-Aided Fine-Tuning and Inference [C] // 2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022: 4292 - 4295.
- [18] OU K, DONG C, HUANG W, et al. UAV Vehicle Detection Method and Application Based on Improved TOOD [C] // 2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAD). IEEE, 2023: 46 - 50.
- [19] OU K, DONG C, LIU X, et al. Drone-TOOD: A Lightweight Task-Aligned Object Detection Algorithm for Vehicle Detection in UAV Images [J]. IEEE Access, 2024, 12: 41999 - 42016.
- [20] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9759 - 9768.
- [21] LI X, WANG W, WU L, et al. Generalized Focal Loss: learning qualified and distributed bounding boxes for dense object detection [J]. Advances in Neural Information Processing Systems, 2020, 33: 21002 - 21012.
- [22] DAI X, CHEN Y, XIAO B, et al. Dynamic Head: unifying object detection heads with attentions [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7373 - 7382.
- [23] SUN P, ZHANG R, JIANG Y, et al. Sparse R-CNN: end-to-end object detection with learnable proposals [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14454 - 14463.
- [24] GUO C, FAN B, ZHANG Q, et al. AugFPN: improving multi-scale feature learning for object detection [C] // Pro-

ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 12595–12604.

- [25] WANG X, ZHANG S, YU Z, et al. Scale-equalizing pyramid convolution for object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 13359–13368.
- [26] GE Z, LIU S, WANG F, et al. Yolox: exceeding yolo series in 2021 [J]. arXiv preprint arXiv: 2107.08430, 2021.
- [27] ZHANG Z, HE T, ZHANG H, et al. Bag of freebies for training object detection neural networks [J]. arXiv preprint arXiv: 1902.04103, 2019.
- [28] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018; 8759–8768.
- [29] PANG J, CHEN K, SHI J, et al. Libra R-CNN: towards balanced learning for object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 821–830.
- [30] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [C] // IEEE

Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020, 57 (5): 9–12.

- [31] FENG C, ZHONG Y, HUANG W. Exploring classification equilibrium in long-tailed object detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021; 3417–3426.
- [32] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 10781–10790.
- [33] ZHANG H, WANG Y, DAYOUB F, et al. VarifocalNet: an iou-aware dense object detector [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021; 8514–8523.
- [34] CAI Q, PAN Y, WANG Y, et al. Learning a unified sample weighting network for object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 14173–14182.
- [35] CHEN K, WANG J, PANG J, et al. MMDetection: open mmlab detection toolbox and benchmark [J]. ArXiv Preprint ArXiv: 1906.07155, 2019.

(上接第 224 页)

门体尺寸大, 拆拆困难, 温度范围宽, 充气/放气频繁, 开启方式多, 直接影响围带使用寿命, 系统先后出现围带老化局部开裂泄漏、围带气嘴脱落、围带放气后膨胀部分没回落导致围带与门体刮擦等问题。由于每条围带都是单件定制研制, 而橡胶产品又不能囤积较多备件, 研制成本高。为提高充气围带使用寿命, 必须持续做好充气围带的运行维护, 科学规范运行维护, 及时发现影响围带安全使用的问题, 如门体密封面不能有尖锐物损伤围带、定期开展围带不受压充气/放气测试、供气管路与围带气嘴拔插、围带放气后延时开门和门体打开过程中密封面不与围带刮擦等。同时与外协单位交流协作, 及时跟踪国内外充气围带新材料和制作工艺, 与时俱进优化设计、生产工艺, 降低围带研制成本, 研制更适合风洞使用环境的充气密封围带。

#### 参考文献:

- [1] 李周复. 风洞试验手册 [M]. 北京: 航空工业出版社, 2015.
- [2] 贺德馨. 风洞天平 [M]. 北京: 国防工业出版社, 2001.
- [3] 王勋年. 低速风洞试验 [M]. 北京: 国防工业出版社, 2002.
- [4] 真空设计手册 [M]. 3 版. 北京: 国防工业出版社, 2004.
- [5] 手册编辑委员会. 机械工作手册: 机械零部件设计卷 [M]. 2 版. 北京: 机械工业出版社, 1996.
- [6] 徐 灏. 机械设计手册 [M]. 2 版. 北京: 机械工业出版社, 2000.
- [7] 抽充气圆管试验装置的密封结构设计 [J]. 润滑与密封, 2008, 33 (11): 59–63.
- [8] 王继常. 真空技术及应用系列讲座第十讲: 真空密封 [J]. 真空, 2001 (1): 51.
- [9] 谭 晶, 杨卫民, 丁玉梅, 等. O 形橡胶密封圈密封性能的有

限元分析 [J]. 润滑与密封, 2006 (7): 65–69.

- [10] JING, YANG W M, DING Y M, et al. Finite element analysis of the sealing performance of O-ring seal structure [J]. Lubrication Engineering, 2006 (9): 65–69.
- [11] 候 琳, 易 辉, 何晓哲, 等. 纤维增强层硅橡胶充气密封带的研制 [J]. 特种橡胶制品, 2013, 34 (4): 51–55.
- [12] 褚 衡, 刘庆丰, 严海彪, 等. 橡胶制品纤维骨架材料的研究进展 [J]. 特种橡胶制品, 2006, 27 (3): 53–57.
- [13] 周仕明, 李道奎, 唐国金. 充气式自密封气囊结构的非线性有限元分析 [J]. 润滑与密封, 2010, 35 (10): 59–63.
- [14] 唐淋伟, 马东平, 于凤举, 等. 跨超音速风洞喷管段柔壁新型充气密封围带研制 [J]. 研润滑与密封, 2021, 46 (6): 126–130.
- [15] 赵 照, 熊建军, 张平涛, 等. 分布式光纤测温的某风洞喷雾耙温度场测量 [J]. 计算机测量与控制, 2021, 29 (3), 14–17.
- [16] 熊建军, 刘 锡, 冉 林, 等. 基于控制律的电加热防除冰系统设计与验证 [J]. 测控技术, 2021, 40 (2): 130–134.
- [17] 熊建军, 梁 鉴, 郭 龙, 等. 某风洞试验段上壁面自动顶盖装置设计与应用 [J]. 计算机测量与控制, 2020, 28 (12): 135–138.
- [18] 熊建军, 倪章松, 李 刚, 等. 3 米×2 米结冰风洞动力系统设计与应用 [J]. 测控技术, 2018, 37 (z): 192–196.
- [19] 朱鹏霏, 李 丽, 常泽宇, 等. 基于 BMS 与云平台的动力电池健康管理体系统 [J]. 计算机测量与控制, 2023, 31 (9): 190–198.
- [20] 王丰磊, 程国防, 魏宏建, 王宇江. 阵地电力系统跨平台集中监测设计实现 [J]. 计算机测量与控制, 2023, 31 (9): 36–41.