基于神经网络的军用飞机识别

陈玉熙 李洪仁 赵 航 李育泓 姚俊豪 付晓明 李 瀚

摘 要 为了提高对图片和视频中的军用飞机的目标检测速度，以便对网络上关于军用飞机的敏感信息进行监管，在 YOLOv5 的模型基础上，进行了针对化改进。首先对数据集进行了 Mosaic-9 数据增强处理，并采用 MobileNet V3 Small 取代 YOLOv5s 的 ResNet 特征提取网络，以丰富小目标样本和提高特征提取速度；其次通过 Label Smoothing 防止过拟合，通道剪枝以降低神经网络的复杂度和冗余度; 最后利用 TensorRT 对底层硬件资源进行优化，提高检测速度。实验使用 Military Aircraft Detection Dataset 军用飞机数据集，在 NVIDIA Tesla P100 16GB GPU 上进行模型训练和验证，结果表明，上述改进，在保持模型较高的预测准确度情况下，有效提高了模型的训练速度及检测速度。

关键词 机器学习，目标检测，数据增强，通道剪枝，标签平滑，模型加速
中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.01.2021.00001

Title

Yuxi Chen Hongreng Li Hang Zhao Yuhong Li Junhao Yao Xiaoming Fu Han Li

Abstract In order to improve the target detection speed of military aircraft in pictures and videos for the regulation of sensitive information about military aircraft on the web, targeting improvements were made based on YOLOv5's model. Firstly, Mosaic-9 data enhancement was performed on the dataset, and MobileNet V3 Small was used to replace YOLOv5s ResNet feature extraction network to enrich small target samples and improve feature extraction speed; secondly, overfitting was prevented by Label Smoothing, and channel pruning was used to reduce the complexity and redundancy of the neural network; finally TensorRT is used to optimize the underlying hardware resources to improve the detection speed. The experiments use Military Aircraft Detection Dataset for model training and validation on NVIDIA Tesla P100 16GB GPU. The results show that the above improvements, while maintaining the high prediction accuracy of the model, effectively improve the training speed and detection speed of the model.

Key words machine learning; target detection; data enhancement; channel pruning; label smoothing; model acceleration

1 引言

近年来，物联网、大数据、人工智能等技术的成熟与运用标志着人类开启了信息化时代，为人们带来了便利的同时，也凸显了互联网的划时代意义。根据第 47 次《中国互联网络发展状况统计报告》数据显示，2020 年 12 月，中国网民人数共 9.89 亿，有超过 70% 的中国公民开始使用互联网。其中，约 99% 的网民使用手机上网。互联网发展至今，信息传播速度之快，范围之广，导致了互联网海量的信息无法得到及时的监管与处理。网络信息安全作为一种非传统信息安全形态，是总体国家安全体系的重要组成部分。

2004 年，四川一军迷偷拍我军用飞机照片泄密被判刑。2009 年 8 月，福州市黄某拍摄军机视频并发布到网上，有关部门制止时这段视频已经播放 1.5 万次，最终由于故意泄露国家机密罪被判刑。2018 年，有以色列军人多次将以军最新服役的 F-35 隐身战斗机、一款秘密坦克以及地下军事基地的照片泄露到互联网上。国家机密军事信息泄露将会给国家安全带来巨大威胁。因此，除了加大国家安全信息保护宣传外，还要采取必要的技术措施来对互联网内容进行监管，坚决避免与国家安全有关的军事设施照片和视频出现在互联网上。其中，对互联网上机密军事飞机的图片和视频的监管更是重中之重。在机密军事飞机的相关照片和视频上传到互联网上之前，需要采取技术手段将其识别并拦截。

国内外已经在军事飞机检测和识别方面提出了很多方法。文献[1] 是基于主成分分析（PCA）和图像匹配的飞机识别算法，在对物体识别前先经过大量的图像处理操作，如均值滤波、直方图均衡化

收稿日期：2021-11-25；修改日期：2021-11-25 陈玉熙，E-mail: . 李洪仁，E-mail: . 赵航，E-mail: . 李育泓，E-mail: . 姚俊豪，E-mail: . 付晓明，E-mail: . 李瀚，E-mail: .

等，再在机场图像中使用图像分割技术对飞机进行分割，接着在各个分割的区域中采用 PCA，最后与模板库进行匹配。文献[2] 提出了一种采用飞机的几何特征、主成分分析特征、HU 不变矩特征等多种特征混合来自动识别飞机类型的方案，使用支持向量机（SVM）作为预分类器，使用多特征融合的决策方案来进行识别。文献[3] 提出了一种基于声波的最快间歇信号检测（Quickest Intermittent Signal Detection）方法，并将该方法应用到飞机检测方面。

随着深度学习（Deep Learning）[4] 技术的兴起，提供了一种端到端的方式来对图像中的目标进行识别，即在一端输入原图像，在另一端输出目标的预测类别。文献[5] 使用卷积神经网络对飞机图像进行了识别，识别效果较好，采用工具生成大量的飞机图像。文献[6] 提出了一种使用卷积神经网络改进的多标签网络结构（MLCNN），通过在不同深度层中设置各个标签的分类器，解决标签之间存在包含关系时的识别问题。

本文在深度学习的基础上，将 yolov5 目标检测网络应用到军事飞机的识别。yolov5 作为 yolo 系列最新的目标检测网络，在保证准确率的情况下，相比于上一代 yolov4 更加快速和灵活，应用于军事飞机目标检测具有巨大的优势。此外，本文还采取了一系列改进措施：使用 Mosaic-9 数据增强方法，丰富数据集目标的同时，增加了小目标样本，提升了网络训练速度；采用 MobileNet V3 Small 取代 YOLOv5s 的 Backbone 特征提取网络；在 YOLOv5 的 Prediction 层引入 Label Smoothing 标签平滑方法；采用结构化剪枝方法使得卷积神经网络轻量化。同时，通过实验与改进前的方法进行对比，证明了本文方法的有效性。

2 相关工作

2016 年 Redmon 等人[7] 提出了 YOLO 算法，将分类、定位、检测功能集成在一个神经网络当中，只需经过一次计算，就可以直接得到图像中目标的边界框和类别概率。输入图片被划分成 S×S 个网格，只需判断是否有目标中心落在网格内，以及预测出 B 个边界框信息，然后选择合适的置信度阈值去除那些存在可能性低的边界框。虽然 YOLO 算法完全舍弃了候选区域生成步骤，极大提高了检测速率，能满足实时目标检测的速度要求，但由于其网络设计比较粗糙，远远达不到实时目标检测的精度要求，而且存在目标不能精准定位、容易漏检，小目标和多目标检测效果不好等问题。

2017 年 Redmon 等人[8] 提出了 YOLO9000（YOLOv2）算法，对 YOLO 算法进行了一系列改进，重点解决召回率低和定位精度差的问题。它借鉴了 Faster R-CNN 算法的 Anchor 机制，移除了网络中的全连接层，使用卷积层预测检测框的位置偏移量和类别信息。而且不同于原 Anchor 机制的手工设计，它利用 K-Means 聚类方式在训练集中学习最佳的初始 Anchor 模板。不仅如此，YOLOv2 添加了一个 pass-through 层，将浅层的特征图连接到深层的特征图，使网络具有了细粒度特征。此外，YOLOv2 可以采用多种数据集联合优化训练的方式，利用 WordTree 方法在 ImageNet 分类数据集和 MSCOCO 检测数据集上同步训练，实现超过 9000 个目标类别的实时检测任务。

2018 年 Redmon 等人[9] 提出了 YOLOv3 算法，它借鉴残差网络中跳跃连接的思路，构建了名为 DarNet-53 的 53 层基准网络，该网络只采用 3×3 和 1×1 的卷积层，具有与 ResNet-152 相仿的分类准确率，但大大减少了计算量；为了处理多尺度目标，采用了 3 种不同尺度的特征图来进行目标检测，每个特征图都是高层与浅层特征图融合所得；在预测类别时，使用 Logistic 回归方法代替 Softmax 方法，使得每个候选框可以预测多个类别，支持检测具有多个标签的对象。YOLOv3 算法能满足实时检测任务的精度与速率的要求，成为了当前工程界首选的目标检测算法之一。

2020 年 Bochkovskiy 等人[10] 提出了 YOLOv4 算法，在 YOLOv3 的基础上不断进行改进和开发。YOLOv4 可以使用传统的 GPU 进行训练和测试，并能够获得实时的，高精度的检测结果。与其他最先进的目标检测器的比较的结果 YOLOv4 在与 EfficientDet 性能相当的情况下，推理速度比其快两倍，相比 YOLOv3 的 AP 和 FPS 分别提高了 10% 和 12%。2021 年 Jocher[11] 提出了 YOLOv5，整合了计算机视觉领域的先进技术，大大提高了灵活性和速度。

3 关键技术

3.1 Mosaic 数据增强

Mosaic 数据增强方法是一种常见的数据增强，由 Alexey Bochkovskiy 等人[10] 提出，其主要思想就是将四张图片进行随机裁剪，再拼接到一张图上作

为训练数据，这样做的好处是丰富了图片的背景，并且四张图片拼接在一起变相地提高了 batch size，在进行 batch normalization 的时候也会计算四张图片，所以对本身 batch size 不是很依赖，单块 GPU 就可以训练 Yolov4，训练的过程如图1所示

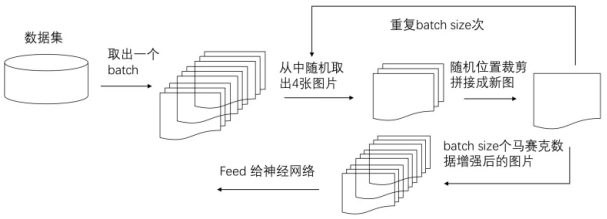

图 1　Mosaic 数据增强流程

相较于 Mixup[12]、Cutout[13]、CutMix[14]，Mosaic 数据增强在训练过程中具有以下优点：

1. 在训练过程中不会出现非信息像素，从而能够提高训练效率；

2. 保留了 regional dropout 的优势，能够关注目标的 non-discriminative parts；

3. 通过要求模型从局部视图识别对象，对 cut 区域中添加其他样本的信息，能够进一步增强模型的定位能力；

4. 不会有图像混合后不自然的情形，能够提升模型分类的表现；

5. 训练和推理代价保持不变。

而在本次实验中采用 Mosaic 方法的增强版——Mosaic-9，即对 9 张图片随机裁剪、随机缩放、随机排列组合成一张图片，其细节如图2所示。Mosaic-9 数据增强利用了 9 张图片，对 9 张图片进行拼接，每一张图片都有其对应的框，将 9 张图片拼接之后就获得一张新的图片，同时也获得这张图片对应的框，然后将这样一张新的图片传入到神经网络当中去学习，相当于一次性传入 9 张图片进行学习，这极大丰富了检测物体的背景，且在标准化计算的时候就会计算 9 张图片的数据。操作完成之后然后再将原始图片按照 9 块相对位置进行摆放。完成 9 张图片的摆放之后，利用矩阵的方式将 9 张图片它固定的区域截取下来，然后将它们拼接起来，拼接成一张新的图片，新的图片上含有框等一系列的内容。拼接完成之后得到的新的一张图片，在拼接的时候部分图会被相邻近的图覆盖掉了，拼接的时候很有可能也会把另外的图中的框给覆盖掉，这些问题都会在最后的对框进行处理：当图片的框（或者图片本身）超出两张图片之间的边缘（即设置的分割线）的时候，就需要把这个超出分割线的部分框或者图片的部分）处理掉，进行边缘处理。

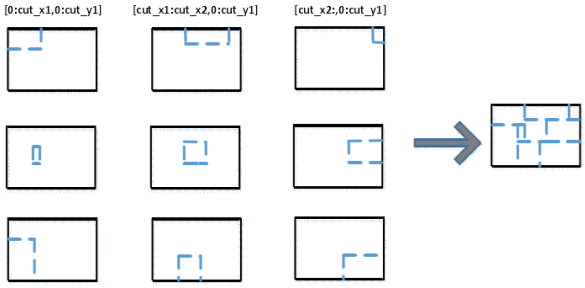

图 2　Mosaic-9 数据增加细节

3.2　MobileNet

MobileNet 是 Google 提出的一种用于移动和嵌入式视觉应用的称为的高效模型。MobileNet 基于一种流线型的架构，该架构使用深度可分离卷积 (Depthwise separable convolution) 来构建轻量级的深度神经网络。深度可分卷积是一种分解卷积的形式，它将标准卷积分解为深度卷积和一个被称为点卷积的 1×1 卷积。其不仅可以降低模型计算复杂度，而且可以大大降低模型大小[15]。

MobileNet V2 是在 2018 年提出，对 V1 的卷积单元进行了改进，主要引入了线性瓶颈 (Linear bottleneck) 和反向残差 (Inverted residuals)[16]：

1. 线性瓶颈即去掉卷积单元中最后一个 ReLU 函数，不使用 ReLU 激活，而用线性变换来代替原本的非线性激活变换。这样做是为了避免 ReLU 对低维度运算时对特征的破坏，造成信息的丢失，使得卷积层参数为空；

2. 反向残差即先升维到高维空间进行卷积操作来提取特征，随后再进行降维。由于深度卷积本身没有改变通道的能力，如果输入通道很少，深度卷积只能在低维度上工作，于是使用点卷积升维，在一个更高维的空间中进行卷积操作来提取特征，最后再次使用点卷积进行降维，以增加非线性通道转换的能力。

MobileNet V3 结构上相较于 V2 而言，引入了 SE 模块、修改了尾部结构、修改了通道数、改用了 h-swish 激活函数[17]：

1. 引入 SE 模块，即通过显式地建模网络卷积特征通道之间的相互依赖关系，来提高网络所产生表示的质量。

2. 修改尾部结构，即将平均池化层前的 1×1 卷积放置层后，因为 1×1 卷积会占据大量的运算时间，放置在池化层后可以减少计算量。

3．修改通道数，即将第一个卷积核的通道数由 32 修改为 16。

4．为提高网络准确性，可以使用 swish 非线性激活函数替换 ReLU 激活函数，其公式为：

$$\text{swish}(x) = x \cdot \sigma(x) \tag{1}$$

但由公式可知，计算 swish 需要计算 sigmoid 函数，这会提高计算成本。而 ReLU6 函数在许多软硬件框架中都已实现，易于量化部署，即使以 16 位浮点数或 8 位整型低精度运算时性能也较好，且计算推理速度快。于是使用 h-swish 激活函数来近似 swish 函数，其公式如下：

$$h-\text{swish}[x] = \frac{x \cdot (\text{ReLU6}(x+3))}{6} \tag{2}$$

经过试验验证，如图3所示，h-swish 激活函数并不会降低网络精度，即使用该近似能在保持精度的情况下加快速度。

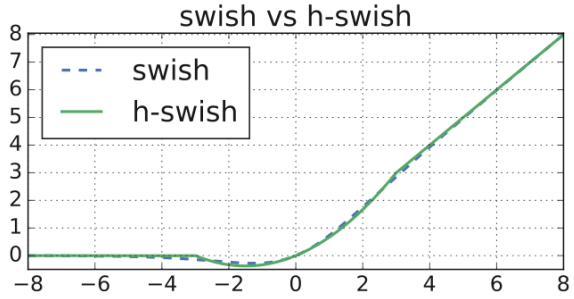

图 3　swish 与 h-swish 激活函数对比

MobileNet V3 是目前最新于是效果最好的版本，同时其推出 Large 版本和 Small 版本，适用于不同算力资源的情况。为实现 YOLOv5s 网络的轻量化、小型化，最终本实验采用 MobileNet V3 Small 取代 YOLOv5s 的 Backbone 特征提取网络来进行特征提取。

3.3　通道剪枝

深度卷积神经网络 (CNN) 的部署在许多现实世界的应用很大程度上受到其高计算成本的阻碍，这是因为 CNN 优异的性能表现通常来源于上百万可训练的参数，而且在推理期间 CNN 的中间激活值和响应存储空间甚至需要比存储模型参数的空间还要大，以及因为在高分辨率图片上卷积操作可能会出现计算密集从而使计算时间很长。

通道剪枝是通过在网络中以一种简单但有效的方式强制信道级稀疏性来实现的，直接适用于现代 CNN 架构，在训练过程中引入了最小的开销，并且生成的模型不需要特殊的软硬件加速器。通道剪枝可以减小模型大小、减少运行时的内存占用、在不影响精度的同时降低计算操作数，使得其广泛应用在实际场景中[18]。

目前的剪枝算法分为结构化枝和非结构化剪枝。非结构化剪枝的方法需要对稀疏连接的网络进行量化和编码才能减少模型的实际存储空间，而且需要采用专业的硬件设备和计算方式才能实现模型推理加速。采用结构化剪枝作为卷积神经网络的轻量化方法，通道剪枝以模型重构的方式筛选神经网络中存在的一些冗余连接，这些结构对于模型性能的贡献很小，去掉这部分神经元能够有效降低模型复杂度，同时几乎不会对网络的精度产生影响，甚至还能改善网络的综合性能。

本实验在神经网络中的 BN （batch normalization) 层引入可学习的参数 γ 和 β 加快网络的训练和收敛速度，通过平移和缩放对通道数据进行归一化处理，在迭代训练中学习网络的特征分布，公式如下所示：

$$\hat{z} = \frac{z_{\text{in}} - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}, z_{\text{out}} = \gamma\hat{z} + \beta \tag{3}$$

模型通道剪枝示意图如图4所示：

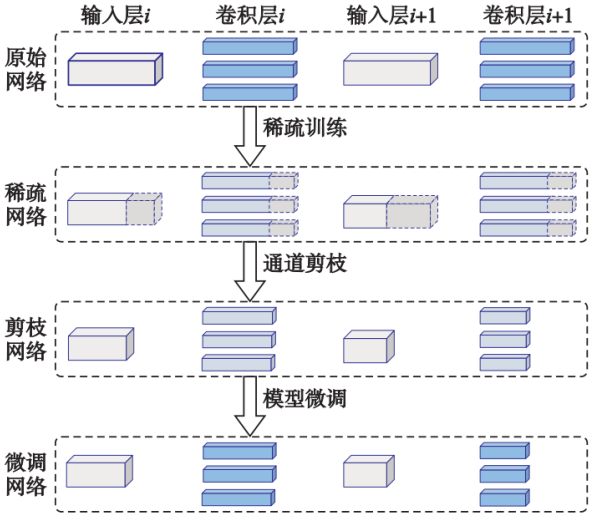

图 4　模型通道剪枝示意图

3.4　Label Smoothing

Label Smoothing[19] 标签平滑最早应用于分类算法中，后面引入目标检测算法中。目标检测算法分为分类与回归两个分支，其主要作用于分类分支，属于正则化方法中的一种。本质上，标签平滑

将帮助模型围绕错误的标签数据进行训练，从而提高其健壮性和性能。标签平滑可以降低模型的可信度，并防止模型下降到过拟合所出现的损失的深度裂缝里，其公式为

$$q_i' = (1-\varepsilon)q_i + \frac{\varepsilon}{K} \tag{4}$$

其中 q_i 表示真实标签，ε 是一个非常小的常数，K 代表分类的类别数。经过 Label Smoothing 后能过减少模型过度依赖标签的问题，有效改善标签准确性不高的情况，故在 Prediction 层中引入 Label Smoothing 标签平滑方法。

3.5　TensorRT 优化

NVIDIA TensorRT 是基于 Nvidia CUDA 编程模型的 SDK，主要用于高性能深度学习推理[20]。具体优化结构如图5所示，TensorRT 提供 api 和解析器来从所有主要的深度学习框架中导入经过训练的模型，经过权重与激活精度校准、层与张量融合、内核自动调整、动态张量显存和多流执行，然后生成可部署在各种环境的优化运行时引擎。[21]

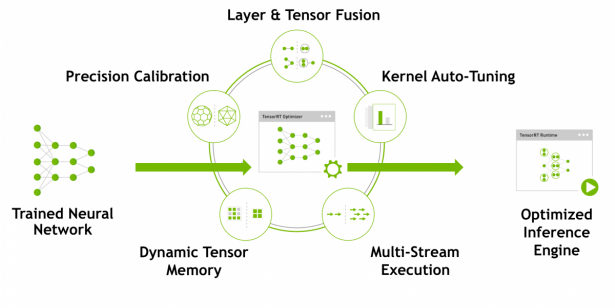

图 5　TensorRT 优化原理

在我们的项目中，为了使 yolo 训练的模型进一步轻量化，我们将模型送入 TensorRT 中优化产生 Engine 引擎，然后再应用在 GPU 推理中，优化流程如图6所示。在最后的 GPU 推理阶段，优化后的引擎被反序列化解析，当推理请求发出时，输入数据从 CPU 复制到 GPU，推理完成后再以异步方式返回结果至 CPU。

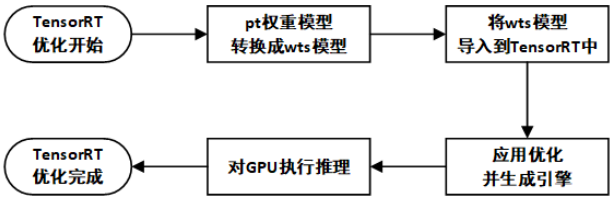

图 6　TensorRT 处理模型流程图

4　实验结果与分析

我们的实验基于 Military Aircraft Detection Dataset，该数据集包含 36 种，5062 张军用飞机图片，包含中、美、俄、欧等国家热门机型。最终经过 TensorRT 加速后的模型虽然会损失一定的精度，但检测速度得到了极大的提升。采用 FP16 精度对模型优化后的对比效果如图7所示：

图 7　未使用 TensorRT 优化（左）与使用 TensorRT 优化（右）检测对比

5　结论

本文提出了一种基于改进 YOLOv5 网络的军用飞机快速检测方法，主要的改进工作为：

1. 使用 Mosaic-9 增加了小目标样本，提升网络训练速度。

2. 采用 MobileNet V3 Small 取代 YOLOv5s 的 Backbone 特征提取网络。

3. 通道剪枝，改善网络的综合性能。

4. 引入 Label Smoothing 标签平滑方法。通过 soft one-hot 来加入噪声，起到抑制过拟合的效果。

5. 使用 TensorRT 优化 GPU 硬件资源

通过对 5000 张军用飞机照片的训练和测试，实验结果表明，最终 mAP_0.5 时精确度稳定在 0.75，mAP_0.5:0.95 时也能超过 0.6，precision 能达到 0.75～0.80，召回率也能超过 0.65。可见 YOLOv5 模型下的美军用飞机识别精度也并没落下。

虽然使用神经网络可以快速而准确的完成军用飞机图像识别和检测任务，但还存在一些不足。当数据量较大时，一定程度上占用资源较多，对 GPU 的性能要求也较高。因此，关于军用飞机识别与检测方面的工作未来可着眼于以下几点：

1. 采用性能高的设备。本文中的实验是基于显存为 16G 的 GPU 上完成的。未来工作中，在扩大研究目标和图像数据集的同时，使用高效设备代替低效设备。

2. 提高模型的分类精度。在本文中使用端到端的方式来完成军用飞机的图像分类和目标检测问题。在未来的工作中可结合传统的特征提取方法来提高模型的分类能力。

3. 对目标检测算法进行优化。对目标检测算法进行深层次理解，尝试优化目标检测算法也是未来工作之一。

4. 试着将这项工程落实到移动端，提高人均防范意识，区别我国和美方的飞机差异，做到不大范围传播泄露。

参考文献

[1] 邵大培, 张艳宁, 魏巍. 基于 PCA 和图像匹配的飞机识别算法[J]. 2009:143-146.

[2] 胡燕, 李元祥, 郁文贤. 基于多特征决策融合的 SAR 飞机识别[J]. 现代电子技术, 2016, 39(21):50-55,60.

[3] JAMES J, FORD J J, MOLLOY T L. Quickest detection of intermittent signals with application to vision-based aircraft detection[J/OL]. IEEE Transactions on Control Systems Technology, 2019, 27(6):2703-2710. DOI: 10.1109/TCST.2018.2872468.

[4] 张荣, 李伟平, 莫同. 深度学习研究综述[J]. 信息与控制, 2018, 47(4):385-397,410.

[5] 欧阳瑞麒, 雍杨, 王兵学. 卷积神经网络在飞机类型识别中的应用[J]. 兵工自动化, 2017, 36(12):71-75.

[6] 孙振华, 李新德. 基于卷积神经网络的多标签飞机识别算法[J]. 计算机应用与软件, 2018, 35(9):270-274.

[7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.: s.n.], 2016.

[8] REDMON J, FARHADI A. Yolo9000: better, faster, stronger[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2017: 7263-7271.

[9] FARHADI A, REDMON J. Yolov3: An incremental improvement [C]//Computer Vision and Pattern Recognition. [S.l.]: Springer Berlin/Heidelberg, Germany, 2018: 1804-2767.

[10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[11] JOCHER G. TPC-W yolov5[Z]. [S.l.: s.n.], 2021.

[12] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond empirical risk minimization[Z]. [S.l.: s.n.], 2018.

[13] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[Z]. [S.l.: s.n.], 2017.

[14] YUN S, HAN D, OH S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features[Z]. [S.l.: s.n.], 2019.

[15] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[Z]. [S.l.: s.n.], 2017.

[16] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[Z]. [S.l.: s.n.], 2019.

[17] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [Z]. [S.l.: s.n.], 2019.

[18] LIU Z, LI J, SHEN Z, et al. Learning efficient convolutional networks through network slimming[Z]. [S.l.: s.n.], 2017.

[19] DAI J, LI Y, HE K, et al. R-fcn: Object detection via region-based fully convolutional networks[Z]. [S.l.: s.n.], 2016.

[20] NVIDIA. NVIDIA TensorRT 可编程推理加速器[Z]. [出版地不详: 出版者不详], 2021.

[21] ABBASIAN H, PARK J, SHARMA S, et al. Speeding up deep learning inference using tensorrt[Z]. [S.l.: s.n.], 2020.