第35卷 第3期 石家庄铁道大学学报(自然科学版) Vol. 35 No.3
2022年9月 Journal of Shijiazhuang Tiedao University(Natural Science Edition) Sep. 2022

基于YOLOv5深度学习模型的动态靶标识别跟踪方法

李福禄[1]， 吉 喆[1]， 段修生[2]

(1.石家庄铁道大学 机械工程学院，河北 石家庄 050043；
2.河北交通职业技术学院 电气与信息工程系，河北 石家庄 050035)

摘要：针对火炮身管动态靶标识别跟踪精度和实时性不高的问题，提出了一种基于YOLOv5深度学习模型的靶标识别跟踪方法。分析了靶标识别跟踪过程和基本思想，通过网格化模型对靶标样本图像进行多尺度处理，并利用金字塔模型进行融合预测；搭建了YOLOv5网络模型，对组件设置优化；对比了损失函数对锚框识别效果的影响，并选取优化后的CIOU作为模型损失函数；最后对模型进行训练，并利用训练好的模型对动态靶标进行识别跟踪。实验结果可视化分析显示，靶标动态识别跟踪率可达到99.3%，动态实时跟踪效果较好。

关键词：YOLOv5；身管靶标；深度学习；目标识别；动态跟踪

中图分类号：TP391 **文献标志码**：A **文章编号**：2095－0373(2022)03－0111－07

0 引言

身管靶标位置识别与跟踪作为身管位姿解算的先决条件，其识别效果和动态跟踪精度决定了身管位姿解算的最终精度。实际情况中，为了保证身管靶标的识别效果和跟踪精度，往往选取棋盘标志物、圆形标志物等形状规则的物体作为身管靶标。近些年有诸多学者对此开展了一系列研究，苏建东等[1]提出一种基于单目视觉的平面靶标姿态测量模型，但动态跟踪识别效果不理想，存在一定的局限性。齐寰宇等[2]提出火炮身管指向测量方法，但由于靶标识别算法采用霍夫变换检测原理，在靶标识别精度满足的同时，动态跟踪相对滞后。另外还有一些学者采用CCD、经纬仪、全站仪、激光跟踪仪等方法，方法相对繁琐且动态跟踪精度不高[3-8]。

随着近些年基于深度学习的图像处理与目标识别技术的发展，越来越多的领域将深度学习技术与计算机视觉技术进行融合，从而达到更好的机器识别效果。尹靖涵等[8]提出了一种基于深度学习的交通标志识别模型[9]，将计算机视觉与深度学习算法进行融合，有效地解决了雾霾、雨雪等恶劣天气下小型交通标志识别精度低、漏检严重的问题。其与实际靶标图像提取时情况相近，需通过单目摄像机采集不同姿态下动态身管靶标数据，因靶标相对较小，且单目视觉系统采集精度受目标动态变化影响较大[9-10]。受其启发，提出一种基于YOLOv5深度学习算法的动态身管靶标识别跟踪方法。经过实验验证，基于该方法的身管靶标实时识别跟踪效果良好。

1 YOLOv5网络模型搭建

如图1所示，YOLOv5包括：Input(输入端)、Backbone(主干网络)、Neck(颈部网络)基本模块。

收稿日期：2022－04－20 责任编辑：车轩玉 DOI：10.13319/j.cnki.sjztddxxbzrb.20220115
基金项目：河北省教育厅重点基金(ZD2022106)
作者简介：李福禄(1997—)，男，硕士研究生，研究方向为检测技术与自动化装置。E-mail：2677693800@qq.com
李福禄，吉喆，段修生.基于YOLOv5深度学习模型的动态靶标识别跟踪方法[J].石家庄铁道大学学报(自然科学版)，2022，35(3)：111-117.

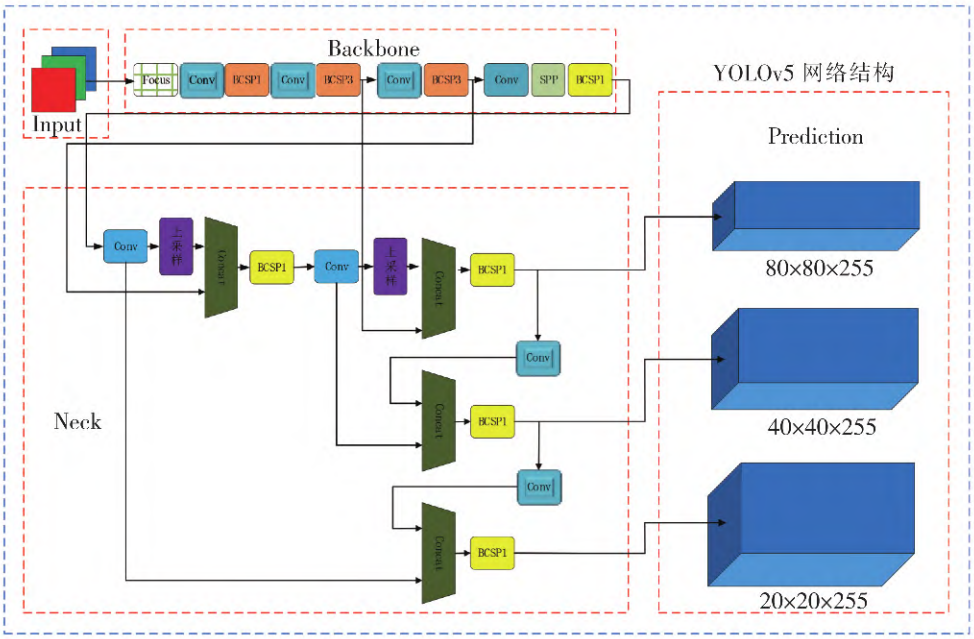


图 1 YOLOv5 网络架构与组件图

(1)Input 模块。Input 端包括 Mosaic 数据增强、自适应锚框计算、图片尺寸处理 3 部分。Mosaic 数据增强:通过对 4 张样本图像进行随机缩放、随机裁剪、随机排布的方式进行拼接,丰富了样本图像的细节数据。自适应锚框计算:在 YOLOv5 模型训练前,需要预先设定初始锚框,然后根据初始锚框预测输出锚框,通过计算对比输出锚框与真实锚框的差值来更新模型网络参数。通过增加自适应锚框计算代码改善了以往锚框值单独计算存在训练结果不优的情况,对样本图片进行尺寸处理,保证所有样本的输入尺寸统一。

(2)Backbone。Backbone 主要由 Focus、Bottleneck CSP 模块构成。Focus 负责把数据切分为 4 份,每份数据都是相当于 2 倍下采样得到的,然后在 channel 维度进行拼接,最后进行卷积操作。Bottleneck CSP 通过对基础层的特征映射图进行复制、dense block 处理等操作,将基础层特征图进行分离,在很大程度上解决了梯度消失带来的影响,并且能够实现特征的传播和复用,降低了网络模型的参数冗余,提高了模型计算速度和识别准确率。

(3)Neck。YOLOv5 的 Neck 模块应用了 Path-Aggregation Network 路径聚合网络(Pannet),如图 2 所示,在从上到下的金字塔网络中(FPN),通过上采样操作处理信息传递过程得到预测特征图。

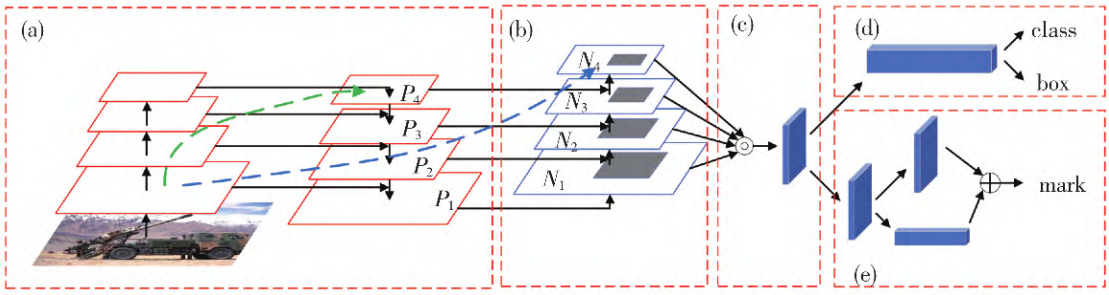


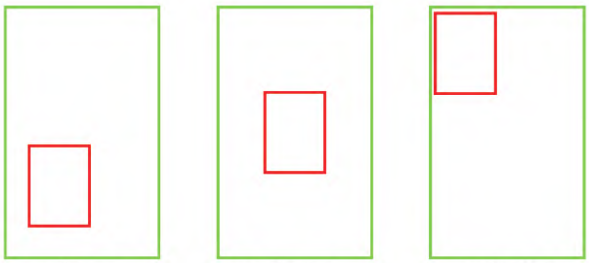
图 2 路径聚合网络 Path-Aggregation Network

Pannet 是基于 Mask R-CNN 和 FPN 搭建的,增强了信息传递效能。通过优化 Pannet 特征提取器的 FPN 结构,改善了特征层自上而下的传递方式。在每个路径中,都以前一级的特征映射作为当前路径的输入,并采用卷积的方式进行处理。然后将特征输出送入 FPN 结构的特征图中,给下层结构提供输入。同时,以自适应特征池化(adaptive feature pooling)方法,重现特征层和候选区的中断信息,并将其合并以防分配混乱。

2　损失函数对比优化

损失函数包括：分类损失(classification loss)用来预测模型分类误差，定位损失(localization loss)用来预测边界框与 GT 之间的误差，置信度损失(confidence loss)，用来预测锚框的目标性[11]。为了选择合适的损失函数，对 YOLO 常用的集中损失函数进行了对比分析。

图 3 为基于 IOU、GIOU、DIOU 识别精度对比，使用 3 种损失函数下的区别，尤其是当目标锚框包括预测框时，DIOU 可以更好地区分不同情况的差异。



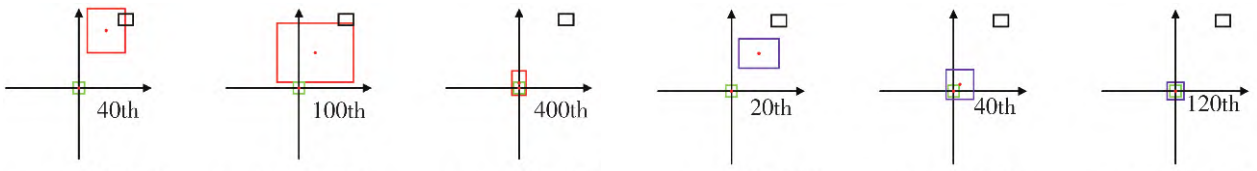
(a)情况 1　(b)情况 2　(c)情况 3

图 3　目标锚框包括预测框的 3 种情况

从表 1 可以看出，基于 IOU 和 GIOU 的损失函数，模型无法区分 3 种情况，基于 DIOU 损失函数的网络模型可以区分出。图 4、图 5 为基于以上 3 种损失函数模型收敛速度对比，可以看出，相对于 GIOU 作为损失函数，DIOU 有着更加优秀的收敛速度。

表 1　3 种情况下损失函数指数

损失函数	(a)	(b)	(c)
IOU	0.75	0.75	0.75
GIOU	0.75	0.75	0.75
DIOU	0.81	0.77	0.75



(a)40th 情况下　(b)100th 情况下　(c)400th 情况下　(a)20th 情况下　(b)40th 情况下　(c)120th 情况下

图 4　以 GIOU 作为损失函数的收敛情况　　图 5　以 DIOU 作为损失函数的收敛情况

YOLOv5 采用二元交叉熵损失函数计算类别概率和目标置信度得分的损失，并且使用 CIOU_Loss 作为 bounding box 回归的损失。CIOU 在 DIOU 的基础上增加了 2 个参数

$$\mathrm{CIOU}=\frac{\rho^2(b,b^{gt})}{c^2}+\alpha v \tag{1}$$

式中，v 为锚框宽度和高度的比例混合度。

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right)^2 \tag{2}$$

$$\alpha=\frac{v}{(1-\mathrm{IOU})+v} \tag{3}$$

如果进一步对 CIOU 进行优化，需要进行以下操作

$$\frac{\partial v}{\partial w}=\frac{8}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right)^2\frac{h}{w^2+h^2} \tag{4}$$

$$\frac{\partial v}{\partial h}=\frac{8}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right)^2\frac{w}{w^2+h^2} \tag{5}$$

式中，w、h 属于$[0,1]$，为了避免梯度爆炸造成计算量大幅增加的情况，一般令 $h/w^2+h^2=1$。

对 IOU、GIOU、CIOU 作为损失函数的回归情况进行对比，如图 6 所示。

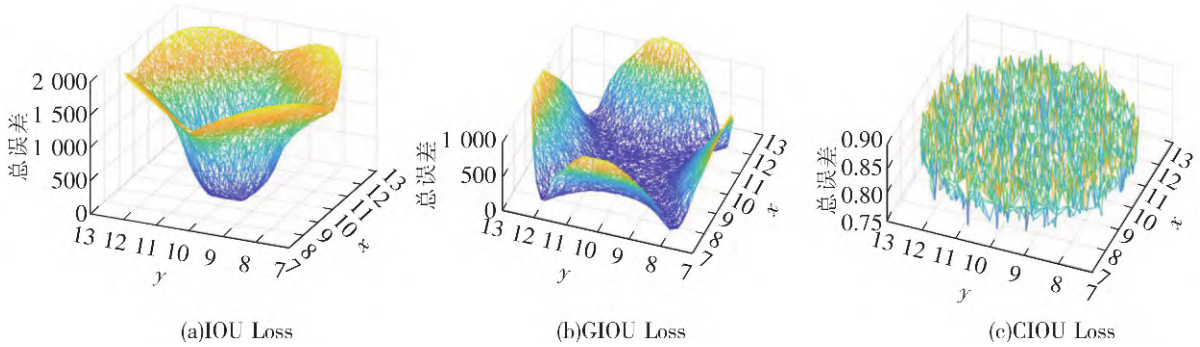


图 6 回归情况对比

DIOU Loss 虽然比 GIOU Loss 和 CIOU Loss 具有更快的收敛速度，但 CIOU Loss 考虑的因素更多，能够更充分地描述锚框的回归。

3 模型训练

3.1 制作数据集

选取 MAKE SENSE 环境制作数据集，首先将需要的样本数据集导入 MAKE SENSE 环境，并精准框选出线性靶标各个标识圆的位置，最终得到带有标识圆位置信息的 labels 文件。

在 labels 文件中从左到右，每一列分别代表目标类别，x_center、y_center、宽度和高度[12]，由于实际情况中只专注于身管靶标的跟踪识别，所以只需设置一类记为 0。图 7 所示的靶标位置示意了 YOLOv5 的 labels 标注文件所需要的目标位置信息，其中图像的宽度记为 width，高度记为 height，目标的左上角坐标记为$(x_{\min},y_{\min})$，目标右下角坐标记为$(x_{\max},y_{\max})$，目标中心坐标记为(x_center,y_center)。由于锚框坐标必须采用规范化的 $xywh$ 格式(从 0 到 1)，所以需要通过公式的计算才能得到 labels 标注文件的数据，计算公式为

图 7 样本标注格式示意图

$$x\_center=\frac{x_{\min}+x_{\max}}{2\text{width}} \tag{6}$$

$$y\_center=\frac{y_{\min}+y_{\max}}{2\text{height}} \tag{7}$$

$$x=\frac{x_{\min}-x_{\max}}{\text{width}} \tag{8}$$

$$y=\frac{y_{\min}-y_{\max}}{\text{height}} \tag{9}$$

3.2 目标框的回归策略

Anchor 给出了目标宽高的初始值，需要回归的是目标真实宽高与初始宽高的偏移量，即预测框中心点相对于对应网格(gridcell)左上角位置的相对偏移值。为了将边界框中心点约束在当前网格中，采用 sigmoid 函数处理偏移值，使预测偏移值在(0,1)范围内，目标框回归示意图如图 8 所示。

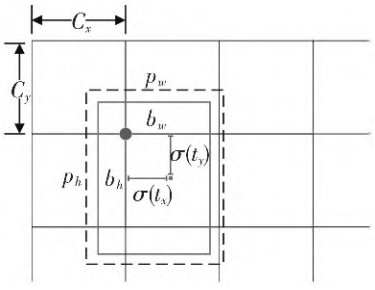


图 8 目标框回归示意图

由于原始的框方程式宽度和高度完全不受限制，所以可能导致失控的梯度、不稳定、NaN 损失并最终完全失去训练[13]。通过 sigmoid 所有模型输出来修补此错误，同时还要确保中心点保持不变。通过设置当前的方程式，将锚点的倍数从最小 0 限制为最大 4，并将锚框-目标匹配更新为基于宽度-高度的倍数，设置标称上限阈值超参数为 4.0。采用跨邻域网格的匹配策略，从而得到更多的正样本，通过从当前网格的上、下、左、右的 4 个网格中找到离目标中心点最近的 2 个网格，再加上当前网格共 3 个网格进行匹配，从而加速收敛。

3.3　训练结果

身管靶标模型训练环境为：Windows10 操作系统、NVIDA GeForce GTX 3050 Ti GPU、AMD Ryzen 75800H@3.2 Hz CPU、16 GB RAM、软件环境 Pytorch1.10、实验语言 Python3.10、运行环境 Pycharm2021.3。经过 300 轮次的训练得到了靶标识别跟踪模型，不同姿态下模型的动态识别跟踪效果如图 9 所示。

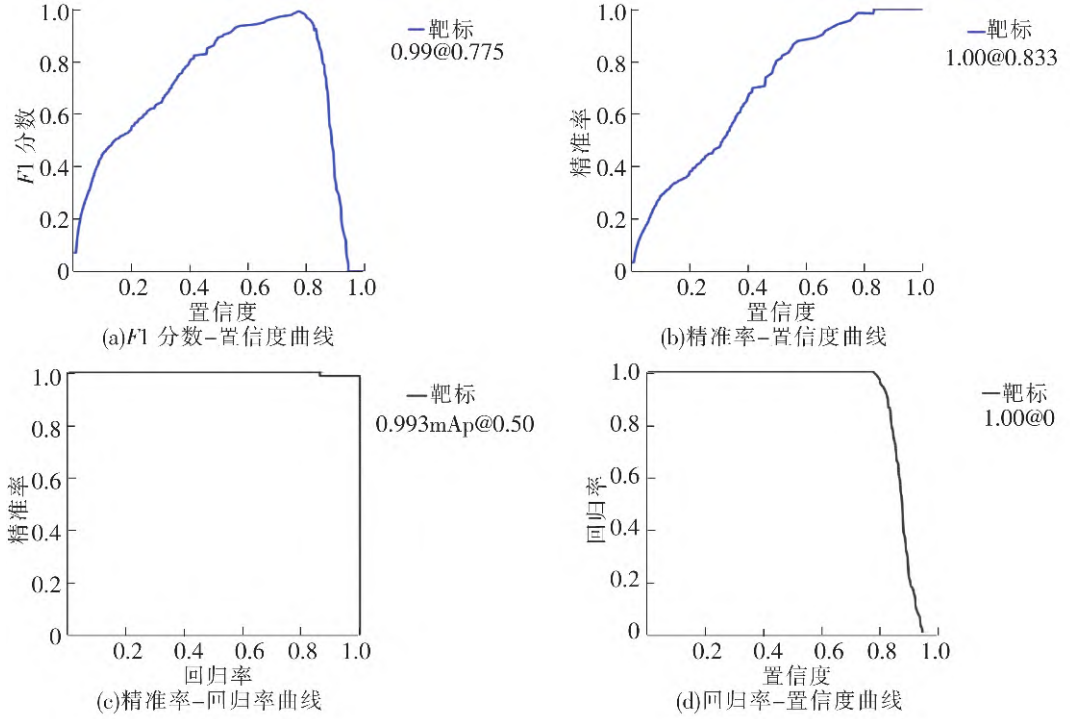
(a)姿态 1　(b)姿态 2　(c)姿态 3　(d)姿态 4　(e)姿态 5

图 9　训练后模型对线性身管靶标样本的检测效果

4　实验数据可视化分析

为了更好地对实验结果进行分析，采用 YOLOv5 模型的数据可视化模块对靶标训练模型进行可视化分析，包括对训练模型的置信度和 $F1$ 分数的关系、识别精准率与召回率关系、识别准确率和置信度关系、以及锚框跟踪损失进行可视化处理。对所涉及参数指标及关系进行介绍并对实验结果展开分析。

如图 10 所示为训练模型 $F1$ 分数、精准率(precision)、回归率(recall)、置信度(confidence)等参数[14]之间的关系曲线，由于只针对靶标进行专一性单类目标识别跟踪，仅需分析身管靶标的模型识别参数，因此只需呈现一条数据曲线。$F1$ 分数是分类效果的衡量指标，在图 10(a)中，0.775 是参数节点，在 0.775 之前随着置信度上调，$F1$ 分数逐渐上升，最大值达到 0.99 的分类精确度，之后开始下降。在图 10(b)中，通过设置置信度参数到 0.833，分类精准率可达到 1，之后分类精准率得到保持。图 10(c)的下围面积反映了模型的分类器性能，图 10(d)中数据回归率达到了 0.993，说明该训练学习器表现良好。综上所述，该训练模型的置信率配置节点于 0.80 左右可以实现良好的模型训练结果，达到 0.993 的识别效果。



(a)$F1$ 分数-置信度曲线　(b)精准率-置信度曲线
(c)精准率-回归率曲线　(d)回归率-置信度曲线

图 10　模型参数特征曲线

将靶标跟踪模型数据导出,对数据进行可视化处理,如图 11 所示为 300 轮次训练的靶标回归损失数据。YOLOv5 使用 GIOU Loss 作为 bounding box 的损失,box 推测为 GIoU 损失函数均值,训练损失率基本在 0.04～0.10 范围之内,训练模型达到了锚框精准度要求;目标检测 Loss 均值在 0.04～0.08 之间;分类 Loss 均值为 0;分类精度在 0.9～1 之间;回归参数 0.9～1 之间;mAP 基本处于 0.9～1.0 之间。从结果可以看出,本实验有着良好的回归效果,靶标锚框跟踪精度达到 0.95,分类跟踪准确率达到 0.993,验证了该方法的有效性和可行性。

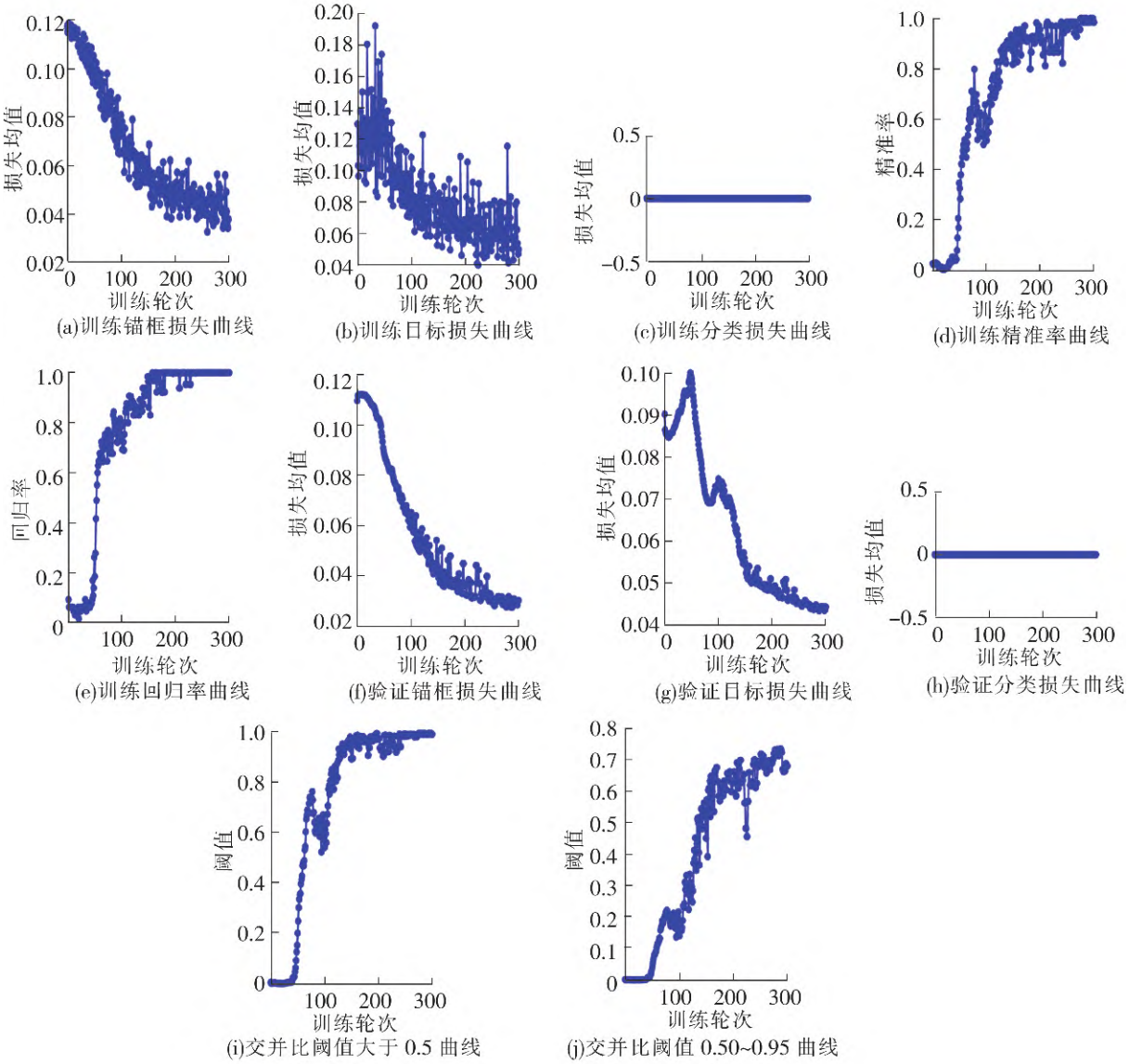


图 11 300 训练轮次中靶标锚框回归损失

5 结论

为解决身管靶标动态识别跟踪检测问题,融合深度学习相关理论和技术,提出了一种基于 YOLOv5 的身管靶标检测方法,经过实验,证明该方法切实可行,动态识别跟踪精度达到了实际需求。该方法较现有的身管靶标识别跟踪方法最大的改进在于动态情况下跟踪检测的实时性,以及靶标锚框动态跟踪精度。由于实际情况中身管靶标与摄像机的距离在 5～10 m 范围,因而不需要考虑遮挡对识别效果的影响。今后仍需要将该方法的原理与身管靶标后续的角点检测、位姿解算相结合,深入研究尺度变换、畸变和数据采集帧率对动态识别跟踪效果的影响,增加方法的普遍适用性。

参 考 文 献

[1]苏建东,段修生,齐晓慧.基于单目视觉的平面靶标姿态测量及其仿真[J].火力与指挥控制,2018,43(7):160-165.
[2]齐寰宇,段修生,马月辉,等.火炮身管指向测量中图像的特征提取与校正[J].火炮发射与控制学报,2021,42(2):34-39.
[3]曾刊,赖文娟,雷雨能.单全站仪调炮精度检测系统[J].四川兵工学报,2013,34(4):18-19.
[4]郭虎生,尹智,王圣旭.双经纬仪调炮精度测量数据的误差分析及补偿[J].计量与测试技术,2018,45(5):79-81.
[5]孟波,王圣旭,游藩, 等.火箭炮调炮精度智能检测系统的设计与实现[J].兵器装备工程学报 2019,40(1):78-82.
[6]刘志鹏,李锴,徐强.基于蒙特卡洛法的单全站仪调炮精度测量系统不确定度研究[J].火炮发射与控制学报,2017,38(2):63-66.
[7]孟繁胜,张玉福,李彬, 等.一种基于激光跟踪仪的火箭炮精度检测方法[J]新技术新工艺,2020(11):76-79.
[8]尹靖涵,瞿绍军,姚泽楷, 等.基于 YOLOv5 的雾霾天气下交通标志识别模型[J/OL].计算机应用:1-10[2022-05-04].http://kns.cnki.net/kcms/detail/51.1307.TP.20210917.1616.008.html.
[9]苏建东.结合扩展卡尔曼滤波的摄像机标定技术[J].信息技术,2014(5):192-195.
[10]张慧娟.单目视觉姿态自动测量方法研究[D]. 武汉:湖北工业大学,2019.
[11]陈令刚.目标识别方法研究[D]. 淮南:安徽理工大学,2018.
[12]Nanxh,Ding L.Review of typical target detection algorithms for deep learning[J]. Application Research of Computers, 2020,37(S2):15-21.
[13]Chen F, Liu Y P, Li S Y. Survey of traffic sign detection and recognition methods in complex environment[J/OL]. Computer Engineering and Applications, 2021,57(16):65-73.
[14]Tsung Y L,Priya G,Ross G,et al.Focal loss for dense object detection[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2018,42(2): 318-327.

Target Recognition and Tracking Method Based on YOLOv5 Deep Learning Model

Li Fulu[1], **Ji Zhe**[1], **Duan Xiusheng**[2]

(1. School of Mechanical Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050043, China;
2. Department of Electrical And Information Engineering, Hebei Communications Vocational and Technical College, Shijiazhuang 050035, China)

Abstract: Aiming at the low accuracy and real-time performance of dynamic target recognition and tracking of gun barrel, a target recognition and tracking method based on yolov5 deep learning model was proposed. The process and basic idea of target recognition were analyzed. The multi-scale processing of target sample image was carried out through grid model, and the fusion prediction was carried out by pyramid model. The research built the yolov5 network model and optimized the component settings. The influence of loss function on the recognition effect of anchor frame was compared, and the optimized CIOU was selected as the loss function of the model. Finally, the model was trained, and the trained model was used to identify and track the dynamic target. The visual analysis of the experimental results shows that the target dynamic recognition and tracking rate can reach 99.3%, and the dynamic real-time tracking effect is good.

Key words: YOLOv5; barrel target; deep learning; target identification; dynamic tracking