

文章编号: 1672-8785(2021)12-0034-11

基于深度学习和相关滤波的 飞机跟踪方法

曹文武¹ 陈小林¹ 朱鑫鑫² 王 博¹ 吴志佳¹ 王雨青¹

(1. 中国科学院长春光学精密机械与物理研究所精密仪器与装备研发中心, 吉林 长春 130033;
2. 中国人民解放军 63618 部队, 新疆 库尔勒 841000)

摘 要: 针对飞行员的飞行训练任务进行了飞机的稳定跟踪。为了解决远距离下飞机目标小、天气环境复杂、飞鸟干扰等问题, 提出了一种基于深度学习和相关滤波的飞机跟踪测量方法。首先选取骨干网络并建立深度学习的算法模型, 接着利用大量飞机图像得到用于实际场景的参考模型, 再将模型检测到的特征与相关滤波结合, 从而达到飞机的稳定跟踪效果并生成目标脱靶量。然后根据跟踪和脱靶量信息, 开启激光器并利用激光测距原理来测量飞机的实时距离。最后进行基于光电经纬仪的飞机捕获与跟踪实验, 以验证模型和算法的有效性和可行性。实验结果表明, 通过深度学习和相关滤波获得的目标信息可以用于远距离飞机的捕获与跟踪, 成功地消除了复杂环境和飞鸟的干扰, 实现了飞机的稳定跟踪。

关键词: 图像跟踪; 深度学习; 激光测量; 飞行训练任务

中图分类号: TP391.41 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2021.12.006

Airplane Tracking Method Based on Deep Learning and Correlation Filtering

CAO Wen-wu¹, CHEN Xiao-lin¹, ZHU Xin-xin², WANG Bo¹, WU Zhi-jia¹, WANG Yu-qing¹

(1. Precision Instrument and Equipment R&D Center, Changchun Institute of Optics, Fine Mechanics
and Physics, Chinese Academy of Sciences, Changchun 130033, China;
2. Unit 63618 of PLA, Kuerle 841000, China)

Abstract: The aircraft stability tracking is carried out for the pilot's training mission. In order to solve the problems such as small target, complex weather environment and bird disturbance, an aircraft tracking measurement method based on deep learning and correlation filtering is proposed in this paper. Firstly, the backbone network is selected and the algorithm model of deep learning is established. Then, a reference model for the actual scene is obtained by using a large number of aircraft images. And the features detected by the model are combined with the correlation filtering to achieve the stable tracking effect of the aircraft and generate the

收稿日期: 2021-08-10

作者简介: 曹文武(1993-), 男, 吉林松原人, 硕士, 研究实习员, 主要从事光电经纬仪和计算机视觉方面的研究。E-mail: hitcww@163.com

missing distance of the target. According to the tracking information and missing distance, the laser is turned on and the laser ranging principle is used to measure the real-time range of the aircraft. Finally, an aircraft capture and tracking experiment based on the photoelectric theodolite is carried out to verify the validity and feasibility of the model and algorithm. The experimental results show that the target information obtained by deep learning and correlation filtering can be used to capture and track long-range aircrafts. It successfully eliminates the interference of complex environment and birds, and the stable tracking of aircraft realizes.

Key words: image tracking; deep learning; laser measurement; flight training task

0 引言

目标跟踪是指对图像中感兴趣的目标进行快速捕获, 并追踪该目标的位置。它是图像跟踪任务的重要研究课题, 对靶场实验和军事目标识别等具有重要意义^[1-2]。传统的目标跟踪方法通常根据目标建模或者对目标特征进行跟踪。这类方法没有将复杂背景考虑在内, 抗干扰能力差, 而且算法执行速度慢, 无法满足实时性要求。

随着计算机技术的快速发展, 基于相关滤波和卷积神经网络(Convolutional Neural Networks, CNN)的跟踪方法已经占据了目标跟踪领域的半壁江山。将相关滤波与 CNN 相结合表现出巨大的优势, 具有速度快、表达能力丰富等优点。Danelljan M 等人^[3]于 2016 年针对特征图连续空间域的学习问题, 提出了一种连续卷积特征训练的跟踪方法。他们用深度神经网络 VGG-Net 提取特征, 并通过立方插值将不同分辨率的特征图插值到连续空间域, 再采用 Hessian 矩阵即可求得具有亚像素精度的目标位置。

同样地, 基于深度学习的目标检测算法也可以与相关滤波等算法结合。具有代表性的目标检测方法包括 RCNN (Region-CNN)^[4]、YOLO (You Only Look Once)^[5]、MobileNet^[6]等。它们具有速度快和检测准确性高等优点, 在图像识别等领域应用广泛。Liu S 等人^[7]于 2018 年针对水下鱼类检测和跟踪的问题, 提出了一种将 YOLOv3 算法与并行相关滤波器相结合的跟踪方法, 完成了鱼类在线检测与跟踪任务。Asha C S 等人^[8]于 2018 年针对城市交通的车辆计数问题, 提出了一种基于视频的车辆计

数方法。他们采用 YOLO 算法进行目标检测, 然后用相关滤波器进行跟踪和计数, 并使用 YOLO 框架产生的边界框建立了多目标的相关滤波跟踪器。Dai W 等人^[9]于 2018 年针对遥感图像中的飞机检测问题, 提出了一种基于改进的 YOLOv3 的实时飞机检测算法。该算法的网络由 49 个卷积层组成, 网络之间采用稠密连接; 利用最大值池化来增强稠密块之间的特征传输。由于遥感图像中的飞机为小目标, 将尺度检测从 3 级提高到了 4 级, 并采用稠密连接的方法对不同尺度的特征图进行融合。

为了解决飞机快速飞行中的稳定跟踪问题, 本文研制了一套由图像采集、图像处理 and 图像显示等模块组成的光电经纬仪自动跟踪系统, 训练了一种基于深度学习的可自动识别图像中飞机目标的网络模型, 并设计了一个通过将深度学习的识别结果与相关滤波相结合来稳定可靠跟踪飞机的目标跟踪器。与其他基于深度学习的目标检测方法相比, 本文方法具有速度快、实时性好、准确性高等特点, 能够在复杂背景下跟踪目标并可给出正确的脱靶量。

1 飞机目标跟踪方法

1.1 目标检测框架介绍

CNN 是一类包含卷积计算且具有深度结构的前馈神经网络, 是深度学习的代表算法之一^[10-11]。当前最流行的神经网络是深度卷积神经网络^[12]。本文采用基于深度神经网络的目标检测算法来对目标的类别进行识别, 同时对目标的区域进行预测。该算法的优势主要有以下几点:

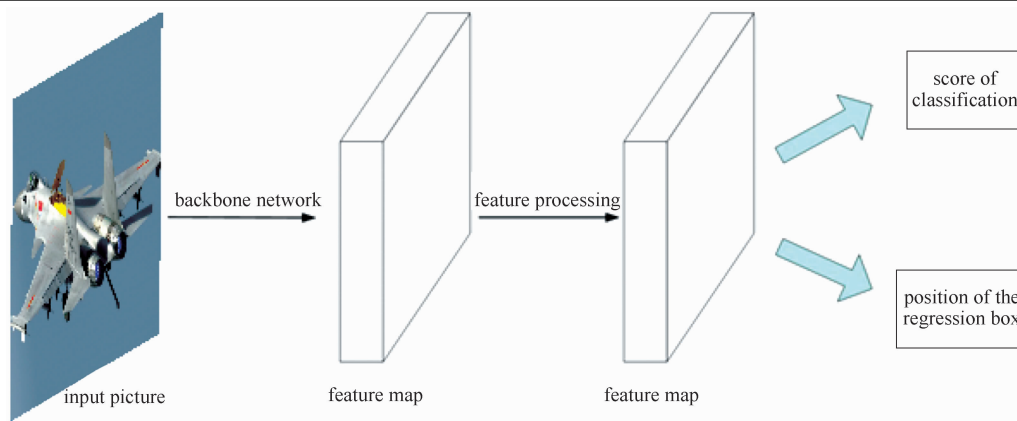


图1 深度学习目标检测框架

(1)自动、准确地捕获待测目标，为智能调焦模块和后续的目标全程稳定跟踪提供保障；

(2)剔除背景中可能对稳定跟踪产生影响的干扰因素(如飞鸟、云朵、塔台等)，确保跟踪全程可靠稳定，同时具备出色的抗干扰能力；

(3)获得目标的区域信息，为特征点选择和目标跟踪提供先验条件。

如图1所示，首先通过将相机实时采集的图像数据送入到骨干网络和特征处理模块中来提取深度特征。获得特征图后有两个处理分支：第一个分支用于实现预测目标类别概率的功能。在本方案中，系统只需预测目标是否为待测飞机即可。若目标检测结果确认无误，则对应该类别的概率很高。第二个分支用于回归出目标区域框的坐标位置。下面介绍自动捕获算法中所使用的骨干网络、特征处理模块以及预测回归算法。

1.2 骨干网络

骨干网络用于提取深度特征图。一般情况下，深度学习模型都是在经典基础网络结构上增加适用于特定任务的模块或分支来实现的。这种经典基础网络结构是网络的基础，因此被称为骨干网络。骨干网络通常会在普适性的图像任务(如图像分类)上进行预训练，从而获得图像语义的先验信息。在预训练后骨干网络的基础上针对特定任务进行参数微调往往能达到更理想的效果。

如图2所示，经典的骨干网络有 LeNet^[13]、AlexNet^[14]、GoogleNet^[15]、VGG^[16]、ResNet^[17]、DenseNet^[18]和 CSPNet^[19]等等，还包括在它们基础上提出的修改和变形。上述几种骨干网络在目标检测领域得到了广泛的应用。为了满足框架实时性的要求，在本方案中使用 CSPNet 作为骨干网络。CSPNet 从网络结构设计的角度来解决以往工作在推理过程中需要很大计算量的问题。该问题是由网络优化中的梯度信息重复导致的。CSPNet 将梯度的变化从头到尾地集成到特征图中，在减少计算量的同时还可保证准确率。

1.3 特征处理模块

为了得到更适用于目标检测任务的特征，在得到骨干网络的深度特征后通常会使用特征处理模块来作进一步处理。本文采用了空间金字塔池化(Spatial Pyramid Pooling, SPP)和路径汇聚网络(Path Aggregation Network, PAN)两种经典的深度特征处理模块。

在目标检测任务中，常常会碰到因目标大小不同而检测不准确的问题。比如，在飞机较小或较大的情况下，可能会出现检测类别不一致的问题。针对该问题，利用 SPP 模块(见图3)来处理特征。它对输入的深度特征图采用不同大小的池化，可获得不同感受野下的特征，从而得到多尺度下的目标信息。这样就可保证不会因输入图像中目标大小多变而影响实际预测的效果。

随着网络层数的加深，对目标语义的抽象

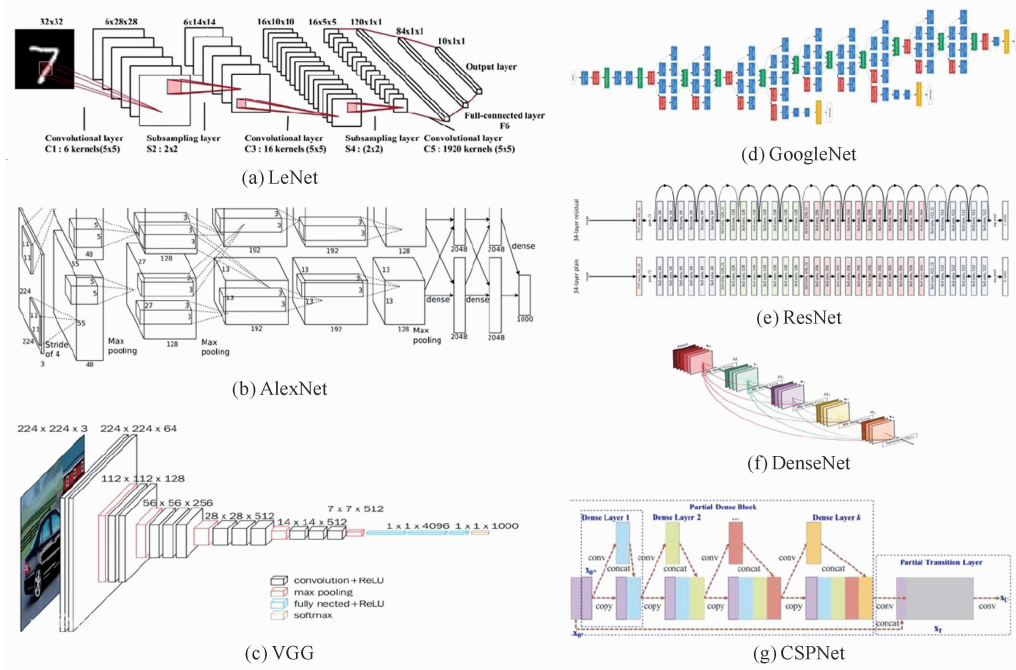


图 2 经典的骨干网络

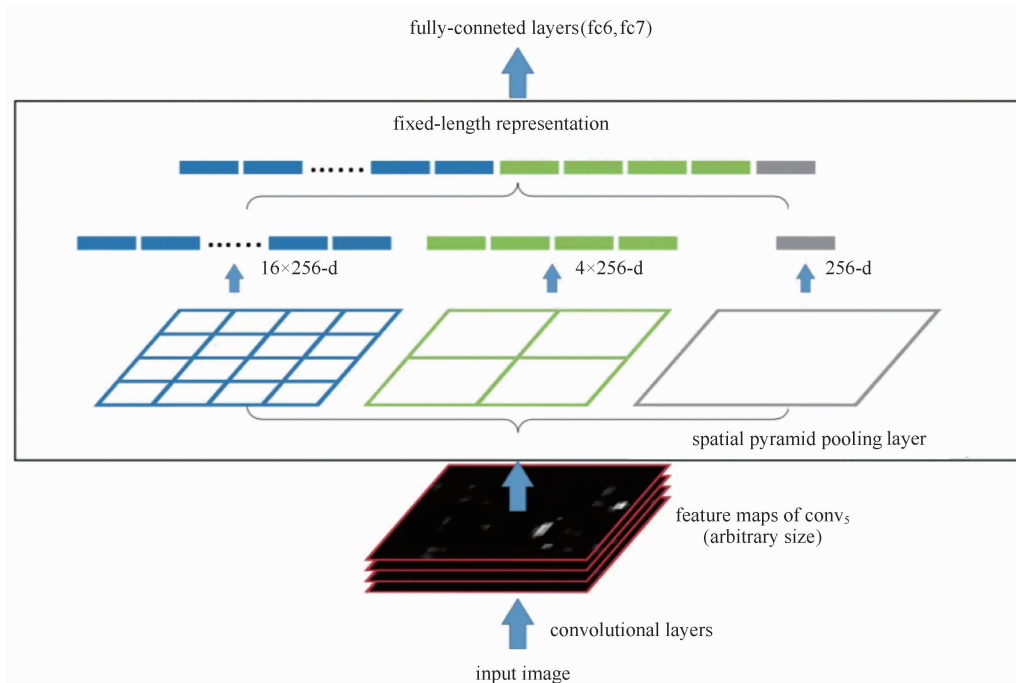


图 3 SPP 结构

能力会越来越强, 但是不准确的定位信息也会随之而来。而深度网络的浅层往往包含目标的细节信息。因此, 本方案采用路径汇聚模块 (见图 4) 对深度特征进行处理。参考图 4 右半部分的处理方式, 采用多个路径汇聚的方式逐渐融合浅层特征并恢复出图像中的细节信息,

从而对目标进行准确定位。

1.4 目标类别预测及目标区域位置回归

一般情况下, 对于网络训练和优化的损失函数表示为

$$L(S, t_{x,y,w,h}) = L_{cls}(S_c) + \alpha L_{reg}(t, t^*) \quad (1)$$

式中, c^* 是真实类别标签, t^* 表示真实边框坐

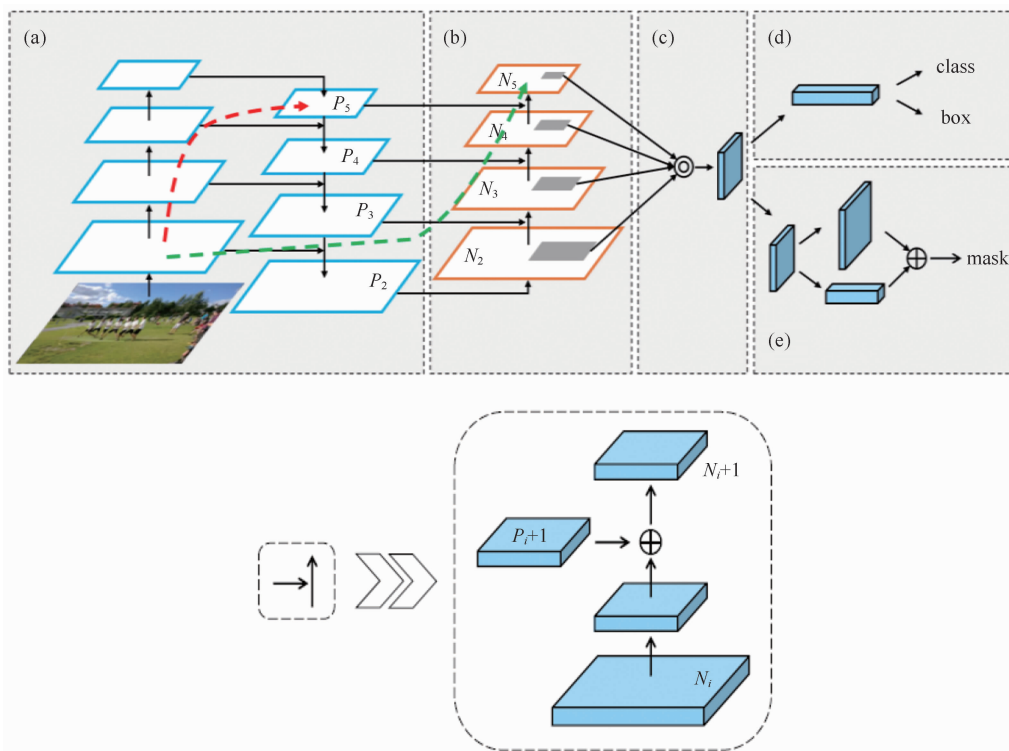


图4 路径汇聚网络结构

标。等号右边第一项对应的是类别标签损失，用于保证网络能学习到正确的目标类别；第二项对应的是目标区域框坐标位置的回归损失，用于约束预测出的目标框的位置。网络训练和优化结束后，可以通过该网络对输入图像进行预测。当待测飞机进入捕获区域并闯入设备视场时，就能够对其完成识别确认工作并准确捕获目标了。

1.5 相关滤波

相关滤波最早应用在通信领域，用于衡量两个信号的相似程度。将其引入图像领域之后，利用滤波模板与输入图像计算相关度，从而形成相关图像。从图像的角度来讲，它就是模板匹配检测，通过在图像中寻找目标模型的最佳匹配位置来检测目标。通用相关滤波检测方法为

$$NC(x, y) = \frac{\sum (|F(x, y) - E(F)| \times |G(x, y) - E(G)|)}{\sqrt{\sum (F(x, y) - E(F))^2 \sum (G(x, y) - E(G))^2}} \quad (2)$$

式中， $F(x, y)$ 为检测子区域； $G(x, y)$ 为目

标模板； $E(F)$ 为检测子区域均值， $E(G)$ 为模板均值； $NC(x, y)$ 为相关值，该值最大的坐标点对应为目标检测点。

由于飞机目标尺寸较大且具有明显形状甚至纹理特征，因此采用相关滤波检测方法检测飞机目标时具有良好的抗噪声、抗遮挡、抗光照变化能力。通过深度学习检测到飞机目标后，跟踪器会转入跟踪状态。这时相关滤波利用深度学习检测的特征和包围框来生成目标跟踪模板，并将它送入跟踪器以使其不断更新和迭代跟踪目标。同时根据目标位置的变化生成脱靶量。将脱靶量送到伺服系统即可完成飞机的稳定跟踪任务，从而解决了跟踪过程中可能出现的云朵、飞鸟等干扰问题。

1.6 目标跟踪器设计和距离测量

根据飞行任务情况，将系统执行过程（见图5）分为两部分。即按照目标降落的空间位置，将航迹划分为捕获区与跟踪区两个区域。

为了提高自动化和智能化程度，设备需要在捕获区根据场景中的实时情况对目标进行自动捕获。也就是说，当目标出现在视场范围内

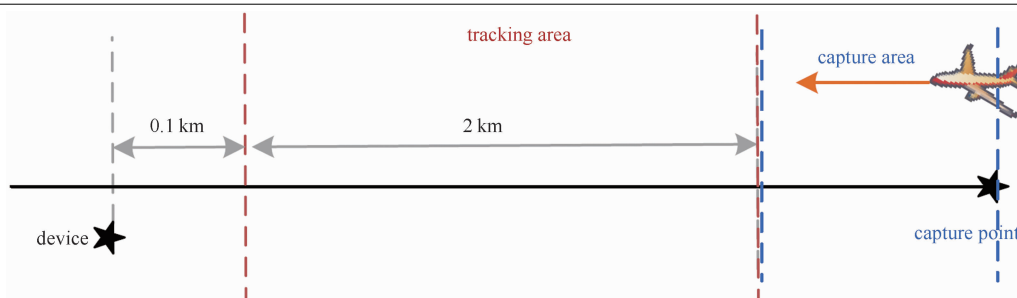


图 5 系统执行过程的示意图

时, 系统要第一时间判断目标是否为待测飞机。若目标是待测飞机, 则执行后续操作, 否则就继续等待。在捕获区捕获到飞机后, 目标转入跟踪区, 系统开始对飞机进行稳定跟踪与距离测量解算。为了确保激光测距的准确度, 需要系统持续稳定跟踪飞机的某一固定区域, 直至飞机降落并行至 100 m 处。如图 5 所示, 系统在等待区自动捕获后转入跟踪状态。飞机接下来会大幅减速, 并正对跑道直线飞行约 2 km, 然后降落。系统重置, 结束跟踪并自动返回等待区域。因此, 目标跟踪环节需要经过飞机转体、降落、由远及近等多种变化情况。系统的目标跟踪模块的技术难点是在初始跟踪区域的自动提取以及在大气湍流、飞机转体、光照变化、尺度变化等影响因素下实现对该区域飞机的稳定跟踪。

为了实现捕获区与跟踪区的全程自动化, 设计了基于相关滤波的自动跟踪器。其整体逻辑原理框图如图 6 所示。开机后, 系统自动指向捕获区域, 等待测量目标出现, 并准备自动捕获。当视场中有疑似目标出现时, 智能调焦模块开始工作, 为神经网络目标识别模块提供大小适宜、稳定清晰的图像以供检测识别。神经网络目标识别模块能够准确判断出该目标是否为待测目标, 并选择继续等待测量目标出现或立即转入跟踪状态。若转入跟踪状态, 智能调焦模块则会全程跟踪, 并根据目标距离信息、清晰度等来保证目标成像尺寸适宜, 从而为确认特征点并转入稳定跟踪提供良好的前提条件。

转入跟踪状态后, 相关滤波模块也会始终

辅助特征点跟踪模块, 以帮助排除跟踪途中可能出现的云朵、塔台等干扰稳定跟踪的因素。激光器会在转入跟踪状态后开启, 测量参数实时计算模块会收集、整理激光测距信息、编码器信息等相关数据, 并根据时间戳确保所有参与计算的数据的同步性。另外还要进行必要的平滑滤波处理以免引入非必要的误差。不断更新和计算目标的距离信息, 直至目标飞抵测量结束点(跟踪测量结束)。设备重新指向捕获区域并等待下一测量目标出现。系统状态重置, 准备再次自动捕获目标。

目标跟踪过程中的抗干扰设计如图 7 所示。首先, 神经网络目标识别模块会不断对跟踪器捕获的目标进行重新检测, 以保证跟踪的目标为飞机。若检测结果不是飞机, 跟踪器则会不断根据结果放开目标, 并重新转入搜索状态, 检测视场中是否有飞机。然后相关滤波模块会将跟踪器捕获的目标作为模板。模板被设计为 30×30 个像素点以适应小目标的跟踪, 并能够根据纹理和颜色等特征检测匹配目标。由于飞机具有偏亮的明暗特性, 而鸟类和塔台具有偏暗的明暗特性, 相关滤波模块被设计成对亮目标更加敏感以提高飞机跟踪的成功率。最后, 具有目标锁定功能的跟踪器不会随意捕获目标: 一旦处于锁定状态, 它会缩小目标识别和相关滤波的检测范围并排除外围干扰, 从而保证飞机跟踪的成功率。

跟踪测量的距离信息主要由激光器来获得。它由激光发射单元、回波接收单元、电气单元、瞄准望远镜和操作手柄组成。在跟踪器输出稳定的跟踪信息和脱靶量后, 激光器会自

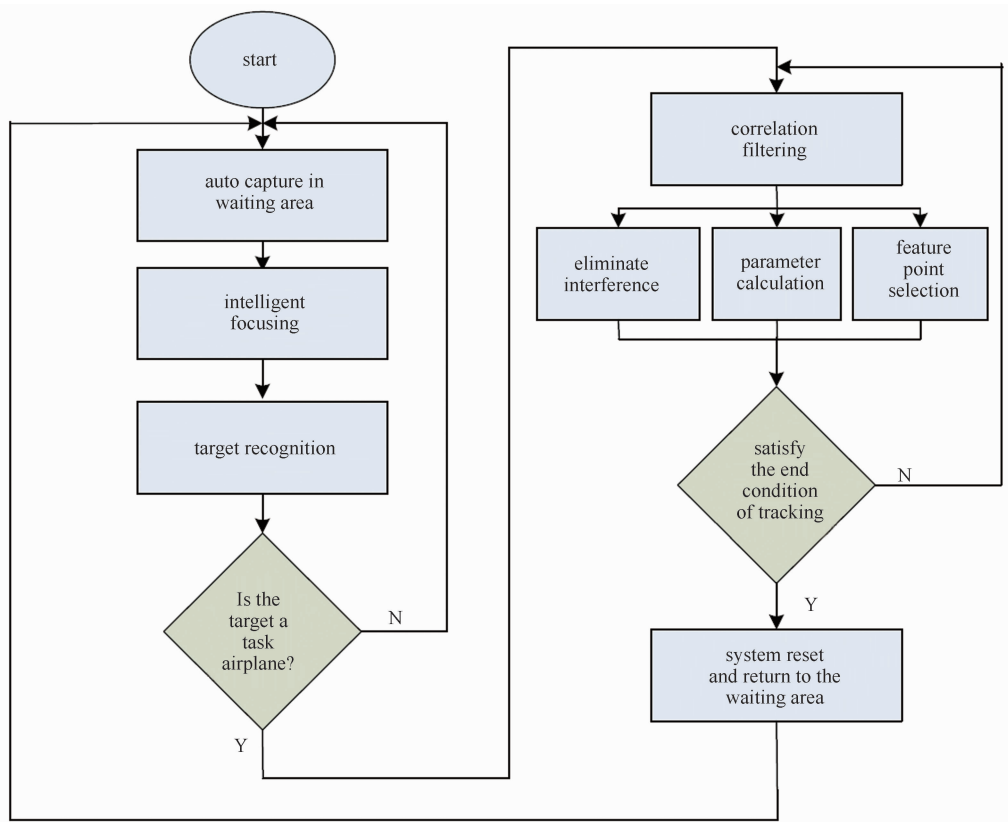


图 6 自动跟踪器的逻辑框图

动开启，并利用激光测距原理来测量飞机与经纬仪之间的距离。通过解算通讯信息，即可获得详细的目标距离信息。本文系统的距离测量精度可以达到 5 m 以内，能够很好地完成距离测量任务。

2 分析与讨论

为了验证算法的可靠性与有效性，使用基于深度学习的目标识别算法对不同大小和角度的真实飞机图像进行了目标捕获测试实验。图 8 所示为算法输出的实验结果。可以看出，基于深度学习的目标检测算法在各种情况下都能对飞机目标有比较理想的、稳定的捕获识别结果。

下面进行一个固定翼飞机降落段的捕获跟踪实验，以验证该方法在目标跟踪过程中的有效性和可靠性。如图 9 所示，在识别飞机并利用特征点匹配进行跟踪的过程中，一只小鸟飞入视场右侧上部，并在几帧后由屏幕左侧穿出视场。跟踪始终保持稳定，未受任何影响。

如图 10 所示，在跟踪过程中，飞机开始朝跑道方向降落。纯净天空背景逐渐切换为空地背景，很多地面建筑及植物等可能会干扰稳定跟踪的目标出现。此时基于深度神经网络的目标识别模块一直在工作，且相关滤波模块对识别后的预测区域进行计算。从图 10 中可以看出，树木、铁塔和大地等未对跟踪造成影响，证明了算法的有效性与可靠性。

为了对比算法的精度和准确性，选择了包括大量的无人机视频、序列和图像的 UAV123 大型数据集。本文将成功率和精度作为评价标准。其中，成功率是指算法预测边界框与真实标注边界框的交并比大于阈值的帧数占总帧数的比率；精度是指算法预测中心与标注中心的距离小于阈值的帧数占总帧数的比率。另外还将成功率和精度作为一次性通过评估(One-Pass Evaluation, OPE)的跟踪指标。本文选取其他先进的跟踪算法(如 DSST、KCF、DCF、MOSSE)与本文算法进行对比，并利用 MATLAB 2016a 软件来实现。选取数据集中光照变

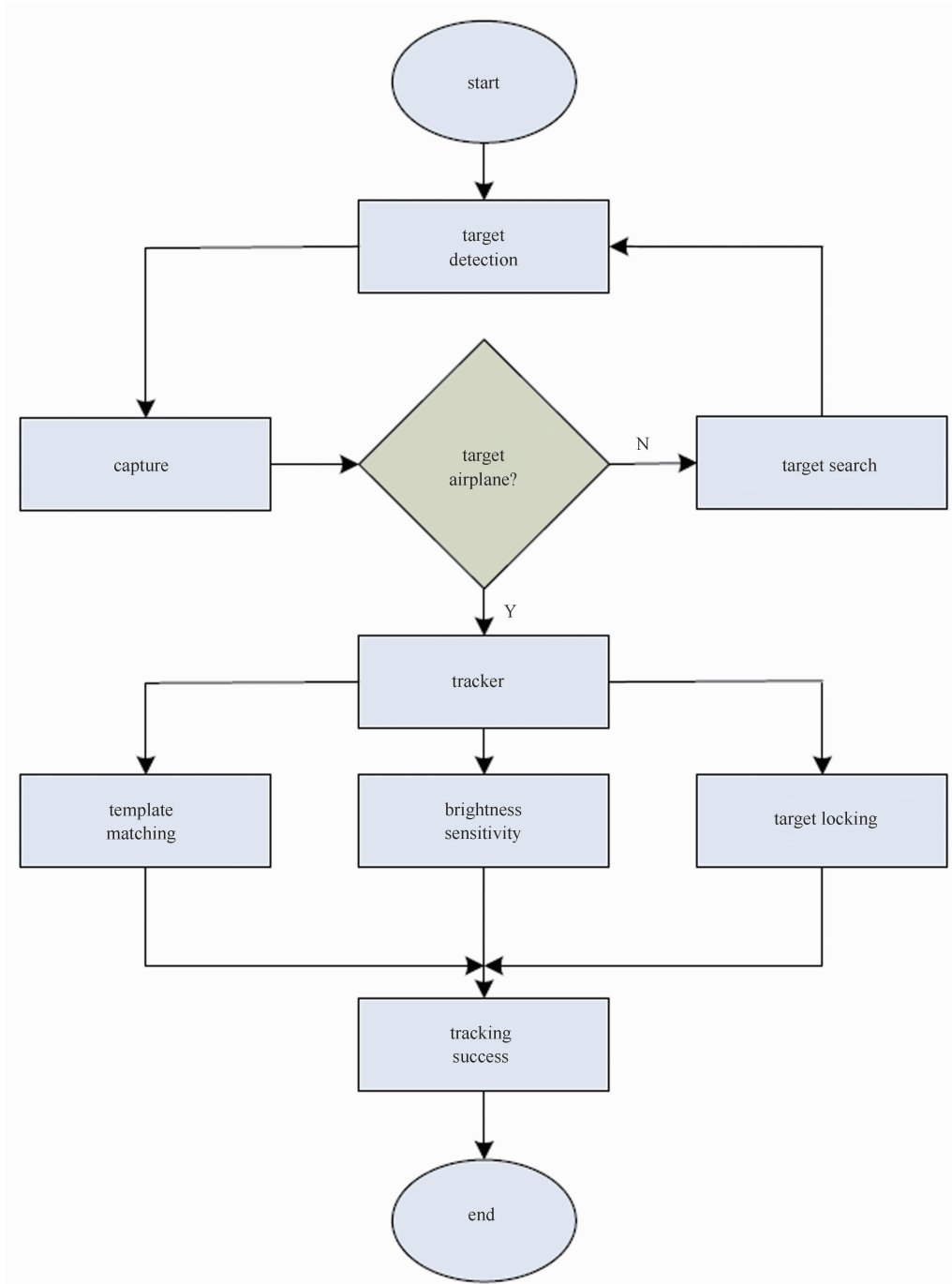


图 7 抗干扰设计的逻辑框图

化和目标形态变化的不同类型视频进行了测试对比, 并评估了成功率和精度的结果(见图 11~图 13)。通过对比图中结果可以看出, 本文算法在光照变化、目标形态变化等干扰情况下具有突出的性能, 优于其他跟踪算法。将深度学习和相关滤波相结合的方式能够提高目标跟踪的成功率和精度, 可以实现目标的稳定跟踪。

3 结论

本文提出了一种将深度学习和相关滤波相结合的飞机检测跟踪测量方法, 并将其用于解决飞机跟踪状态下可能受到飞鸟、云朵、高楼等干扰的问题。通过多组存在干扰物情况下的飞行跟踪实验发现, 基于深度学习和相关滤波的飞机跟踪模型能够更好地适应真实的跟踪情况。



图 8 实际飞机图像的实验结果



图 9 跟踪过程中出现的飞鸟干扰

在提出的深度学习目标检测框架下，选取了 CSPNet 作为骨干网络，并利用 SPP 进行特征处理。最终通过目标类别预测及目标区域位置回归，判断目标是否为飞机并计算位置，从而利用相关滤波不断更新目标信息。结果表明，深度学习后的参考模型适用于对飞机等目标的捕获和跟踪，且基于相关滤波的跟踪器能够成功消除跟踪过程中飞鸟等的干扰，使经纬仪成功完成跟踪和测量任务。本文主要解决了单个目标的跟踪问题，而在多目标跟踪方面仍存在明显不足。下一步的研究重点是优化和改进跟踪算法，使其适用于多目标跟踪领域。

参考文献

[1] Lowell J R. Military applications of localization,

tracking, and targeting [J]. *IEEE Wireless Communications*, 2011, **18**(2): 60–65.

[2] Mazor E, Averbuch A, Dayan J, et al. Interacting multiple model methods in target tracking: a survey [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1998, **34**(1): 103–123.

[3] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking [C]. Amsterdam: European Conference on Computer Vision, 2016.

[4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Columbus: IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[5] Redmon J, Divvala S, Girshick R, et al. You only

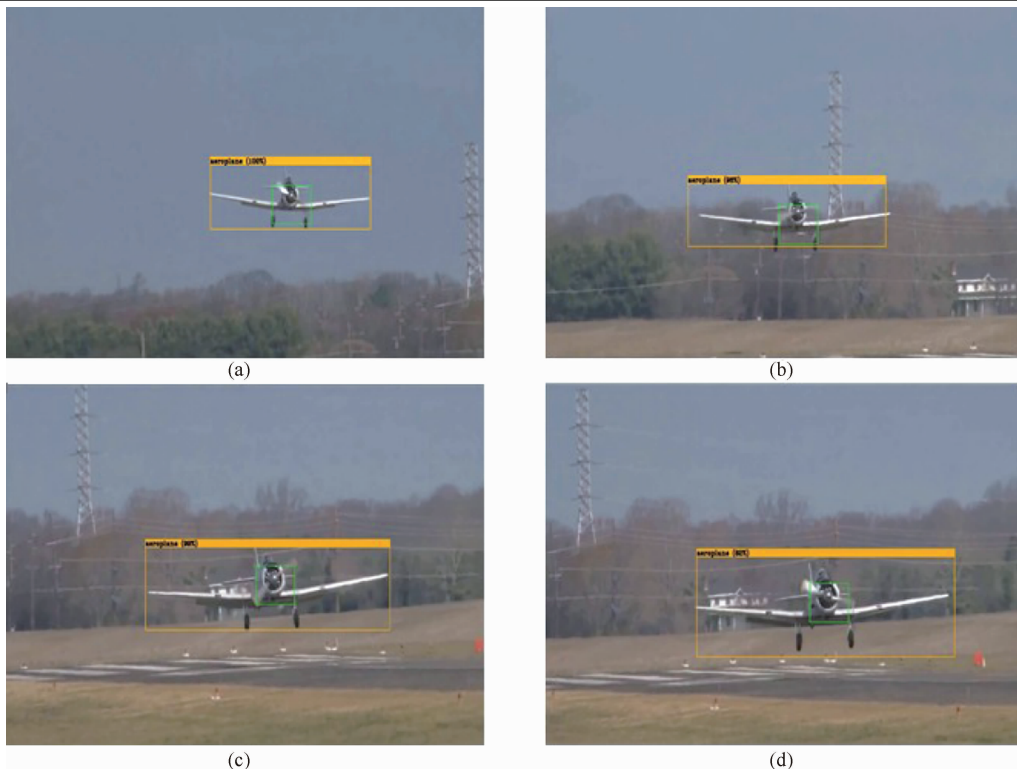


图 10 跟踪过程中出现的塔台和地面景物干扰

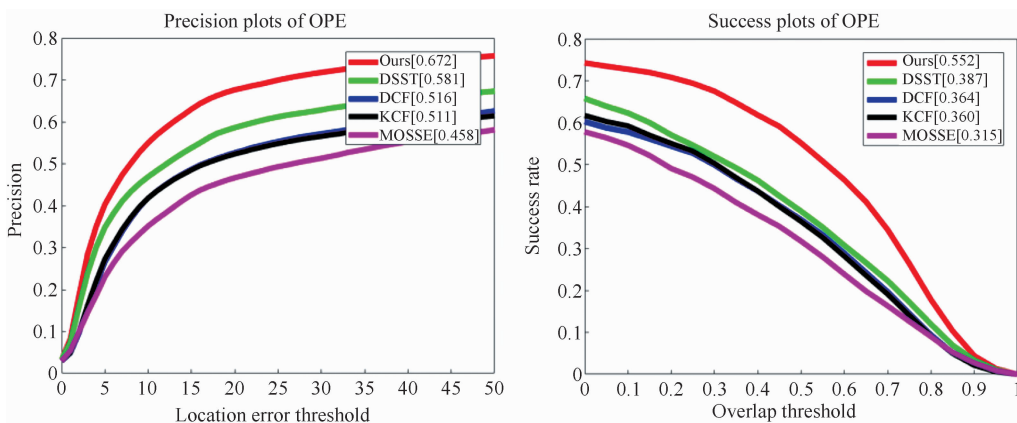


图 11 光照变化数据集一次通过评估结果

look once: unified, real-time object detection [C]. Las Vegas; IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[6] Harjoseputro Y, Yuda I P, Danukusumo K P. MobileNets: efficient convolutional neural network for identification of protected birds [J]. *International Journal on Advanced Science Engineering and Information Technology*, 2020, **10** (6): 2290.

[7] Liu S, Li X, Gao M, et al. Embedded online fish detection and tracking system via YOLOv3 and

parallel correlation filter [C]. Charleston; IEEE Conference on Oceans Mts, 2018.

[8] Asha C S, Narasimhadhan A V. Vehicle counting for traffic management system using YOLO and correlation filter [C]. Bangalore; IEEE International Conference on Electronics, Computing and Communication Technologies, 2018.

[9] Dai W, Jin L, Li G, et al. Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3 [J]. *Opto-Electronic Engineering*, 2018, **45**(12): 12086.

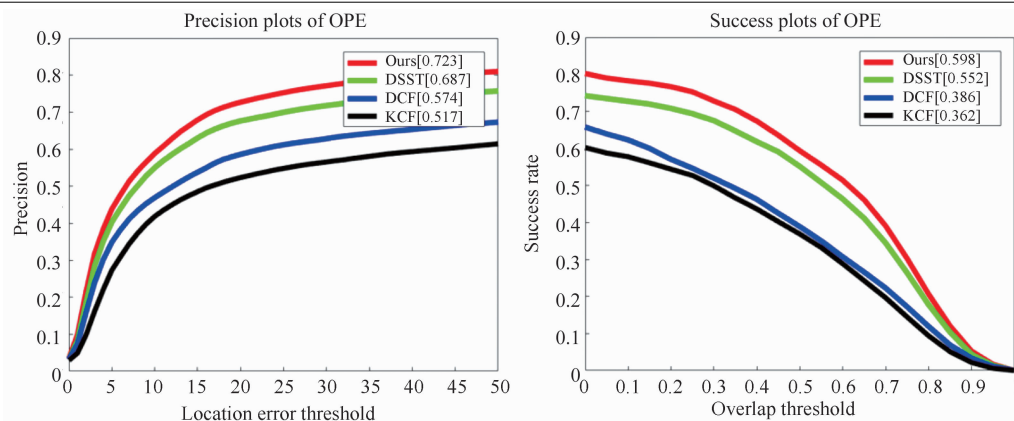


图 12 目标姿态变化数据集一次通过评估结果



图 13 本文算法与其他算法的对比结果

- [10] Zeiler M, Fergus R. Visualizing and understanding convolutional networks [C]. Zurich; European Conference on Computer Vision, 2014.
- [11] Deng L, Yu D. Deep learning; methods and applications [J]. *Foundations & Trends in Signal Processing*, 2014, **7**(3): 1082–1091.
- [12] Shin H C, Roth H R, Gao M, et al. Deep convolutional neural networks for computer-aided detection; CNN architectures, dataset characteristics and transfer learning [J]. *IEEE Transactions on Medical Imaging*, 2016, **35**(5): 1285–1298.
- [13] Lecun Y, Bottou L. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324.
- [14] Gao Z, Wang L, Zhou L, et al. HEp-2 cell image classification with deep convolutional neural networks [J]. *IEEE Journal of Biomedical & Health Informatics*, 2017, **21**(2): 416–428.
- [15] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]. Boston; IEEE Conference

on Computer Vision and Pattern Recognition, 2015.

- [16] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks [C]. Las Vegas; IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [17] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]. Las Vegas; IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [18] Huang G, Liu Z, Laurens V, et al. Densely connected convolutional networks [C]. Honolulu; IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [19] Wang C Y, Liao H, Yeh I H, et al. CSPNet: A new backbone that can enhance learning capability of CNN [C]. Seattle; IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2020.