

基于图像识别的

低空动态目标碰撞预测方法论述

Discussion on Collision Prediction
Methods for Low-altitude Dynamic
Targets Based on Image Recognition

孙一卓¹, 陈麒², 张欢³, 吴炎瑾³, 宋昊阳³ (1. 中国电信深圳分公司, 广东深圳 518000; 2. 中通信息服务有限公司, 广东深圳 518000; 3. 广东南方电信规划咨询设计院有限公司, 广东深圳 518000)

Sun Yizhuo¹, Chen Lin², Zhang Huan³, Wu Yanjin³, Song Haoyang³ (1. China Telecom Shenzhen Branch, Shenzhen 518000, China; 2. Zhong Tong Information Service Co., Ltd., Shenzhen 518000, China; 3. Guangdong Southern Planning & Designing Institute of Telecom Co., Ltd., Shenzhen 518000, China)

摘要:

目前无人机已在各个领域得到了广泛的应用。基于无人机图像传感器的动态目标检测在理论研究和实际应用中占据关键地位, 备受研究人员的关注。深入研究了无人机的工作场景, 针对其特殊的运行环境进行了数学建模并加以分析。通过简化场景, 提出了一种利用轻量级卷积神经网络模型进行图像识别的方法, 在降低计算成本的同时保证对碰撞的准确预测, 从而为无人机的飞行过程提供更为可靠的保障。

Abstract:

Currently, UAVs have been widely used in various fields. Dynamic target detection based on UAV image sensor has attracted extensive attention from researchers because of its key position in theoretical research and practical application. It deeply studies the working scenarios of UAVs, and gives targeted mathematical modeling and analysis for their special operating environments. Then by simplifying the complexity of the scenes, it proposes a method for image recognition using a lightweight neural network model, which reduces computing costs while ensuring accurate prediction of collisions, thereby providing a more reliable safeguard for the UAVs' flight process.

Keywords:

Low altitude economy; Urban situation; UAV vision; Image recognition

关键词:

低空经济; 城市环境; 无人机视觉; 图像识别

doi: 10.12045/j.issn.1007-3043.2024.12.013

文章编号: 1007-3043(2024)12-0078-05

中图分类号: TP391

文献标识码: A

开放科学(资源服务)标识码(OSID):



引用格式: 孙一卓, 陈麒, 张欢, 等. 基于图像识别的低空动态目标碰撞预测方法论述[J]. 邮电设计技术, 2024(12): 78-82.

0 引言

在现代社会中, 无人机的应用范围正在不断扩大, 涵盖了从航拍摄影到快递交付各个领域。然而, 随着无人机数量的增加, 空中交通的拥挤程度也在不断上升, 飞行器碰撞问题日益突出。飞行器碰撞不仅会导致财产损失, 还可能威胁地面人员的安全, 对飞行器碰撞进行准确预测成了至关重要的任务。

图像识别技术^[1-5]在解决这一挑战性问题的过程

中发挥着重要作用。通过分析来自无人机摄像头的图像数据, 可以识别空中的其他飞行器、障碍物以及环境的变化。这为飞行器的自动飞行系统提供了关键信息, 使它们能够实时做出决策, 避免潜在的碰撞。

在以往的工作中, 大量的研究人员专注于无人机视觉领域, 并取得了卓越的研究成果。例如, ZHANG J 等人^[6]利用轻量级 CNN 架构在无人机上实现高精度的运动目标检测; LI G Q 等人^[7]优化了轻量级卷积神经网络, 使无人机目标检测速度有了更加明显的提升; LI T 等人^[8]采用空心卷积架构对特征图像进行重采样, 以此提高了模型的特征提取能力和无人机的目标

收稿日期: 2024-10-16

检测性能。

以上的研究致力于实现无人机对地面目标的高精度运动检测,然而,在空中目标检测方面,过多的计算资源却未必能够得到有效的利用。二者主要有以下2个方面的差异。首先,地面目标通常较小且距离较远,要实现对动态目标的准确识别,需要进行复杂的前景背景分离预处理操作,同时还需要识别分析更高分辨率的图像。相比之下,空中动态目标的碰撞预测首先面临较少的噪声干扰,因此背景变化相对较小,无需特别的预处理步骤。其次,当动态目标即将与观测者的无人机发生碰撞时,其在无人机摄像头中的视角会迅速扩大,这一显著特征使我们无需极高的图像精度即可直接捕获到对应的特征数据。因此,本文针对空中动态目标碰撞环境的这2个显著特点,进行了一定的分析,并设计了相应的碰撞预测方法。

1 碰撞分析

本章对无人机碰撞情况进行了数学建模。如图1所示,无人机a为观测源,无人机b位于同一水平高度与无人机a相向而行,二者相距 L_1 ,无人机b高为 h ,且被无人机a观测到的视角大小为 θ_1 ;后方建筑高为 H ,与无人机a相距 L_2 ,且被观测到的视角大小为 θ_2 。通过计算可知,无人机a的观测视角大小 θ 与观测目标距离 L 及高度 H 的关系可近似表示为式(1)和式(2)。

$$\theta_1 = \arccos \left[1 - \frac{2h^2}{(h + L_1)^2} \right] \quad (1)$$

$$\theta_2 = \arccos \left[1 - \frac{2H^2}{(H + L_2)^2} \right] \quad (2)$$

图2所示为上述2个公式的函数曲线(设常数 $h=0.5\text{ m}$, $H=10\text{ m}$)。从图2可以看出,在观测者与观测目

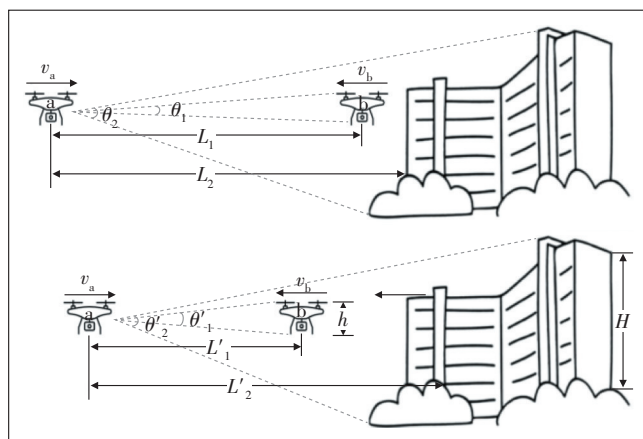


图1 无人机碰撞情况示意

标的距离缩小到某一特定范围时,观测视角 θ 开始加速增大;一旦二者的距离进入碰撞区间,视角 θ 便迅速上升至 180° 。同时可以注意到,当距离相同时,观测目标 h 的视角 θ 整体变化幅度比观测目标 H 更大。这是因为观测目标 H 的初始视角远大于 h ,因此 H 的视角 θ 整体变化幅度较小。

图2左侧也可以描述为无人机b保持不动且与建筑位于同一垂直方向($x=0$)时,观测源无人机a主动靠近二者时视角的变化情况。而图2右侧则更加贴合现实情况,该图描述了当观测源位于垂直方向 $x=0$ 的位置并保持不动,而建筑物位于垂直方向 $x=-5$ 的位置时,作为观测目标的无人机b从两侧接近观测源时视角 θ 的变化。在该图的碰撞区间内,由于建筑物与观测源之间保持一定的距离,因此视角的变化非常微小。然而,观测源对于观测目标的视角却急剧增加,从而形成了明显的视觉特征。这种特性使得图像识别方法更适用于空中碰撞环境这一特殊场景,并为后文提出的方法架构提供了理论基础。

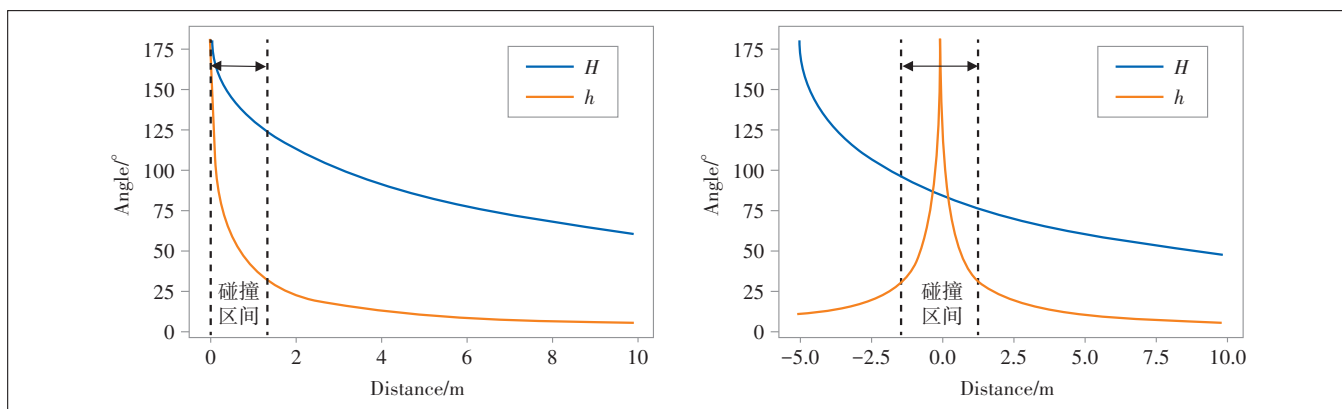


图2 观测者视角(θ)与观测目标距离(L)的关系曲线

2 系统组成

2.1 数据采集和预处理

a) 关键帧选取。为减少视频中相邻帧间的冗余信息,需要从整段视频中选取具有代表性的关键帧,有如下几种方法。

(a) 基于时间间隔。定期抽取帧,保证在时间序列上均匀取样。

(b) 基于运动。挑选画面运动较为明显的帧,突出动态特征。

(c) 基于内容。比较帧间的图像特征,选取较具代表性的帧,以更全面地涵盖图像内容。

b) 去噪。去噪通过滤除图像中的高斯、椒盐、斑点等噪声,提高后续分析的准确度和鲁棒性。去噪可选用的滤波器包括均值滤波、中值滤波、高斯滤波和双边滤波,需根据噪声类型与水平进行选择 and 参数调整,使图像在平滑处理后仍能保留关键纹理和特征。

c) 校正。无人机摄像头成像可能出现镜头畸变、色差或亮度不均等问题,通过相机标定和失真校正,可得到更准确、几何形状更真实的图像。校正后通常需对图像进行重新采样,以确保其尺寸、分辨率与原始图像一致,为后续处理奠定高质量基础。

d) 尺度调整。模型通常要求固定大小的输入图像,可通过双线性或双三次插值将图像缩放至指定分辨率。在此过程中,需在尽量保持图像清晰度和细节的前提下实现尺寸变化。

e) 数据标准化。将像素值映射到统一范围(如 $[0, 1]$ 或 $[-1, 1]$),可减少不同图像间数据分布差异对模型训练的不利影响。同时,对彩色图像的每个通道分别进行归一化(减去均值,除以标准差)操作,使各通道具有相似统计特性,提升模型对颜色特征的敏感度和稳定性。

f) 数据增强。在模型训练中,为提升泛化能力和鲁棒性,常对训练数据进行随机旋转、裁剪、镜像翻转、添加噪声、调整亮度对比度以及改变色彩空间等操作。通过这些多样化的变换,模型可适应更多真实场景下的变化,从而在测试和实际应用中表现更佳。

2.2 模型选择

本节将基于本文方法所涉及的应用场景,筛选出与之适配的图像识别深度学习架构,并对其进行深入解析。在后续章节中,将进行各选定模型的性能指标的比较实验。

2.2.1 Faster R-CNN

Faster R-CNN (Faster Region-based Convolutional Neural Network)^[9]是一种用于目标检测的深度学习架构,引入了可训练的区域建议网络(Region Proposal Network, RPN),从而实现了端到端的目标检测,并在准确性和速度上都取得了显著的提升。Faster R-CNN的主要组成部分如下。

a) RPN。RPN是一个全卷积神经网络,负责生成候选区域。RPN可在图像上滑动一个小窗口,生成不同大小和宽高比的候选区域。然后,RPN根据每个候选区域的特征表示和交并比(IoU)来评估其是否包含目标,并为每个候选区域生成一组边界框修正信息。

b) 共享卷积特征。整个网络共享卷积层的特征提取部分,这样可以避免多次计算相同的特征,提高了计算效率。

c) 目标分类和定位。Faster R-CNN使用分类器和回归器来对生成的候选区域进行目标分类和定位。这些分类和回归任务都在卷积特征的基础上进行,不需要重新提取特征。

2.2.2 YOLOv3

YOLOv3 (You Only Look Once version 3)^[10]是YOLO(You Only Look Once)系列的第3个版本,是一种单阶段目标检测器,具有出色的速度和准确性,适用于各种实时应用。YOLOv3的主要组成部分如下。

a) Backbone(Darknet-53)。YOLOv3的特征提取网络使用了一个名为Darknet-53的骨干网络,它是一个53层的卷积神经网络。这个网络用于从输入图像中提取丰富的特征,以供后续的检测头部使用。Darknet-53的结构类似于ResNet,但具有自己的一些特点,例如使用了更多 3×3 的卷积核。

b) 分级预测。每个尺度的检测头部由卷积层组成,用于生成目标检测的预测结果。每个检测头部生成的预测包括如下几个部分。

(a) 目标类别概率(Class Probability)。对每个检测头部,它会生成一个目标类别概率分数,表示检测到的物体属于哪个类别。

(b) 目标边界框位置偏移(Bounding Box Offset)。每个检测头部还会生成目标边界框的位置偏移信息,用于确定目标的准确位置。

(c) 目标置信度分数(Confidence Score)。目标置信度分数表示模型对每个检测框是否包含物体的置信程度。

c) 锚框。YOLOv3使用多尺寸和宽高比的锚框(anchor boxes)来预测目标边界框。每个检测头部使用不同的锚框,以适应不同尺度和宽高比的目标,这样可以更好地捕捉不同类型的目标。

2.2.3 PeleeNet

PeleeNet(Pelee: A Real-Time Object Detection System on Mobile Devices)^[11]是一种用于在移动设备上运行目标检测的深度学习架构。该架构具有高准确性和实时性,同时模型尺寸较小,可适应移动设备的计算和存储限制。PeleeNet的主要组成如下。

a) 深度可分离卷积(Depthwise Separable Convolution)。深度可分离卷积将标准卷积拆分为深度卷积和逐点卷积,以减少参数数量和计算复杂度,同时保持了良好的特征表示能力。

b) 密集连接(Dense Connection)。密集连接使得每个卷积层都连接到前面所有的卷积层,有助于信息传递和梯度的流动。

c) SSD头部。PeleeNet的输出特征图被送入SSD头部,用于目标检测。SSD是一种单阶段(single-shot)目标检测器,可以同时预测不同尺寸和宽高比的目标框。SSD头部由一系列卷积和池化层组成,用于生成目标类别的分数和边界框的位置信息。对于每个位置,SSD头部会生成一组锚框,并为每个锚框分配类别概率和位置偏移。

3 方法架构

本章基于前文的内容,详细阐述了低空动态目标碰撞预测方法的深度学习架构(见图3)。

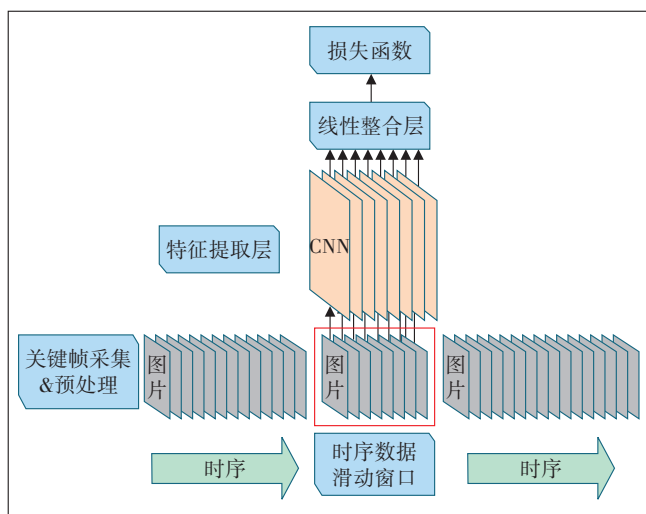


图3 方法架构

3.1 时序数据滑动窗口

为了使模型能够学习到碰撞前观测源记录的图像变化信息,本文采用了一种时间序列数据处理策略,即时序数据滑动窗口。该窗口包含了一系列连续的关键帧图像,并被整合成一个批次(Batch)。随后,将这个批次输入到后续的特征提取层,进行信息的抽取和分析。这一设计的核心目的是使深度学习模型能够更为敏感地感知连续图像数据的信息变化趋势,从而提高系统对观测目标的动态监测能力。

3.2 特征提取层

特征提取层利用前文提供的深度学习模型,通过并行计算的方式处理输入窗口内的每一关键帧的图像数据,进而从中提取观测目标的动态变化信息。由于本文筛选的深度学习模型普遍具有规模小、速度快和精度高等特点,因此非常适用于当前场景的并行计算任务。

3.3 线性整合层

本文在特征提取层上引入了一个线性整合层,其主要职责是整合来自上一层并行计算产生的多个特征数据,随后进行对比分析,从而真正实现连续帧之间的动态信息对比。假设输入层时序窗口大小Batch= N ,也就是说,该层的输入向量维度为 N ,设该层拥有 M 个神经元节点,则其具体表达式如(3)所示。

$$y_i = \tanh \left(\sum_{j=0}^N w_{ij} x_{ij} + b \right), \quad 0 \leq i \leq M \quad (3)$$

其中, i 为输出层神经元序号, j 为输入层神经元序号, x_{ij} 为第 j 个输入层神经元到第 i 个输出层神经元的输入值, w_{ij} 为第 j 个输入层神经元到第 i 个输出层神经元的权重, y_i 为第 i 个输出层神经元的输出值, b 为偏置权重。

3.4 损失函数

笔者选用了log函数作为该深度学习框架的损失函数,其计算如式(4)所示。

$$\text{output} = \log \left(\sum_{i=0}^M w_i y_i + b \right) \quad (4)$$

其中, i 为输出层神经元序号, w_i 为第 i 个输出层神经元的输出权重, y_i 为第 i 个输出层神经元的输出值, b 为偏置权重。

需要强调的是,在模型的训练阶段,当时序窗口图像数据中的观测目标进入到碰撞区间时,损失函数的输出值被设定为1,而其他情况下则被设定为0。碰

撞区间为本文方法的超参数,采用经验取值。

4 实验结果

本次实验将前文给出的3种深度学习模型(Faster R-CNN、YOLOv3和PeeleNet)分别代入方法架构中的特征提取层,以评估它们在无人机碰撞预测任务中的性能。实验使用了一个包含15万张图像的标准无人机碰撞仿真数据集作为总的训练数据集。从总数据集中随机选取了15 000张图像作为校验集,用于模型的中期性能评估。另外选取了20 000张图像作为测试集,用于最终评估模型的性能。所有模型均在具有GPU加速的计算环境中进行训练。实验采用Adam优化器,初始学习率设置为0.001,并随着训练的进展进行调整。实验结果主要关注模型在校验集和测试集上的平均精度(mAP)、处理速度(FPS,帧每秒)、模型大小(MB)等一系列指标,具体如表1所示。

表1 无人机碰撞预测模型性能比较

模型	参数规模/ 百万	校验集 mAP/%	测试集 mAP/%	处理速度/ FPS
Faster R-CNN	50	82	85	8
YOLOv3	62	86	88	15
PeeleNet	5.4	79	83	35

根据实验结果,YOLOv3在碰撞预测准确率方面表现最佳,而PeeleNet在处理速度和模型尺寸上具有明显优势。Faster R-CNN展现了平衡的性能,但在处理速度上稍逊一筹。综合考虑无人机的实时性和资源限制,尽管PeeleNet的准确率略低于其他2个模型,但它可能是最适合的选择。同时,YOLOv3的检测准确率最高,更适合需要高精度检测的应用场景。

5 结论

本文首先对低空环境下飞行器的碰撞情况进行了数学建模,并对通过图像识别系统来预测低空动态目标碰撞情况的技术进行了可行性分析。其次,详细列举了系统搭建过程中所需的各个算法组件、模型结构和技术要素,提出了一种利用轻量级卷积网络模型进行动态目标碰撞预测的方法架构。最后,将3种不同的深度学习模型分别用作系统的特征提取层,比较它们对系统整体性能的影响,并对结果进行了分析。

值得强调的是,当前的技术条件无法实现短距离精确定位,本文的方法填补了这一不足。本文的方法可在短距离无法精确定位的情况下,通过图像识别技

术来预测碰撞风险,为未来低空经济的实施添加了安全保障。

参考文献:

- [1] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.
- [3] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//ICLR 2015.San Diego:ICLR, 2015:1-14.
- [4] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas:IEEE, 2016:770-778.
- [5] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 32nd International Conference on Machine Learning. Lille: PMLR, 2015:448-456.
- [6] ZHANG J, LIANG X, WANG M, et al. Coarse-to-fine object detection in unmanned aerial vehicle imagery using lightweight convolutional neural network and deep motion saliency[J]. Neurocomputing, 2020, 398:555-565.
- [7] LI G Q, ZHANG J W, ZHANG M, et al. Efficient depthwise separable convolution accelerator for classification and UAV object detection[J]. Neurocomputing, 2022, 490:1-16.
- [8] LI T, LV X Y, LIAN X F, et al. YOLOv4_drone: UAV image target detection based on an improved YOLOv4 algorithm[J]. Computers & Electrical Engineering, 2021, 93:107261.
- [9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(1):142-158.
- [10] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2024-03-05]. <https://arxiv.org/abs/1804.02767?context=cs.LG.html>.
- [11] WANG R J, LI X, LING C X, et al. Pelee: a real-time object detection system on mobile devices[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook:Curran Associates Inc., 2018:1967-1976.

作者简介:

孙一卓,毕业于澳洲昆士兰大学,工程师,硕士,主要从事网络架构规划和网络配置维护等工作;陈麒,毕业于东北大学,工程师,博士,主要从事人工智能领域的技术研发和算法设计工作;张欢,毕业于北京邮电大学,博士,主要从事低空经济和网络安全关键技术研究 and 方案设计工作;吴炎瑾,毕业于布里斯托大学,硕士,主要从事飞控与空域管理算法、通信协议设计工作;宋昊阳,毕业于哈尔滨工业大学,工程师,硕士,主要从事空地一体化网络、信息基础设施、智能计算相关的研究工作。