

DOI:10.19344/j. cnki. issn1671-5276.2024.04.023

基于多轴自注意力的无人机避障模型

王新文¹, 赵伟杰²

(1. 福州大学 先进制造学院, 福建 泉州 362251;

2. 中国科学院 海西研究院泉州装备制造研究中心, 福建 泉州 362216)

摘 要:针对无人机在飞行过程中容易因旋翼碰撞而坠毁的问题,提出利用改进的图像识别模型实现自动预警。将瓶颈多轴自注意力模块(BMSA)嵌入到图像识别模型中进行改进,提升模型对细小物体的识别准确率。多轴自注意力层在低分辨率阶段替换原本卷积层,使得模型能够兼顾局部自注意力和全局自注意力。实验结果表明:改进得到的多轴自注意力的残差网络(MS-ResNet)具有较高的障碍物识别准确率,能实现较好的预警效果。

关键词:图像识别;深度学习;自注意力机制;卷积神经网络;避障模型;无人机

中图分类号:TP391.4 **文献标志码:**A **文章编号:**1671-5276(2024)04-0124-05

UAV Obstacle Avoidance Model Based on Multi-axis Self-attention

WANG Xinwen¹, ZHAO Weijie²

(1. School of Advanced Manufacturing, Fuzhou University, Quanzhou 362251, China;

2. Quanzhou Research Center of Equipment Manufacturing of Haixi Institute, Chinese Academy of Science, Quanzhou 362216, China)

Abstract:To address the proneness of UAV crash due to rotor collision during flight, an improved image recognition model is proposed to achieve automatic warning. A bottleneck multi-axis self-attention module (BMSA) is embedded into the image recognition model for improvement, enabling the model to improve the recognition accuracy of the model for fine objects. The multi-axis self-attentive layer replaces the original convolutional layer in the low-resolution stage, enabling the model to obtain both local self-attention and global self-attention. The experiments show that the improved multi-axis self-attentive residual network (MS-ResNet) has high accuracy of obstacle recognition and achieve a better early warning effect.

Keywords: image recognition; deep learning; self-attention mechanism; convolutional neural network; obstacle avoidance model; UAV

0 引言

基于视觉系统的障碍物预警系统是实现无人机安全飞行的重要研究内容。在无人机飞行过程中,航拍摄像头通常无法采集到无人机旋翼上方图像,导致无人机有旋翼撞击障碍物而坠毁的风险。为了解决这个问题,通过图像识别算法对无人机的飞行路径进行实时避障预警。相比无人机上的单目避障系统^[1]和 SIFT 图像匹配避障方法^[2],采用图像识别网络来实现避障系统计算量较低、精度较高,能够较好地平衡计算耗时和准确率。

图片识别算法在各个领域已经实现广泛的应用,2012 年,深度卷积结构的 AlexNet^[3]在 ILSVRC^[4]分类挑战大赛中取得让人印象深刻的成就。GoogLeNet^[5]在 ILSVRC 大赛上实现了

74.8%的 top-1 准确率,后来提出的 ResNet^[6]相比其他模型实现了非常显著的准确率提升。

近几年来,研究人员开始研究计算机视觉任务中的 Transformer^[7]结构,BoT^[8]基于 ResNet 提出带有多头自注意力模块的基本组成结构,同样可以堆叠得到类 ResNet 的网络模型。TU 等^[9]提出结合局部自注意力和全局自注意力的方法,基于此,本文对 BoT 进行改进,提出多轴自注意力的残差网络(deep residual network of multi-axis self-attention, MS-ResNet)模型。

主要创新和工作如下所述。

1)通过视觉系统实现无人机飞行避障预警,以降低旋翼碰撞的风险。在公开数据集中搜集大量无人机飞行障碍物的数据,对相关图像识别网络进行训练和对比。

2)根据 Max-ViT 中提出的多注意力融合方

基金项目:福建省科技计划引导性项目(2022H0042)

第一作者简介:王新文(1996—),男,湖北黄冈人,硕士研究生,研究方向为图像识别、目标检测,1098637488@qq.com。

法,提出一种嵌入卷积网络中的多轴自注意力模块对图像识别模型进行改进,以增强模型对障碍物的识别准确率。

3)为验证所提方法对图像识别模型的有效性,在 CIFAR、FLOWER-102 和 ImageNet 等图片识别数据集上进行了大量的对比实验,充分验证预警模型的先进性。

1 视觉避障模型

通常无人机的航拍摄像头是向下倾斜的,这样会导致操纵无人机时无法注意到前方或者上方的微小障碍物,以至于无人机的旋翼发生碰撞而坠机。本文提出的无人机视觉预警系统如图 1 所示,在无人机的前方添加一个向上倾斜的摄像头,对前上方的场景进行自动拍摄,再由图像识别系统对危险物体进行实时预警,从而避免无人机的旋翼发生碰撞。图像识别模型在无人机嵌入式设备或者云计算终端上对上倾角摄像头的拍摄图像进行实时分析,达到对障碍物预警的目的。

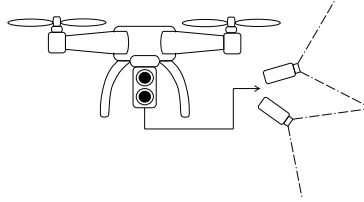


图 1 无人机视觉预警结构

无人机避障系统的预警效果由图像识别模型的准确率来体现。基于视觉系统的避障系统,要求对大尺寸物体和小尺寸物体识别准确率足够

高,因此,下面对图像识别网络进行改进。

图 2(a) 是深度残差网络的 Bottleneck, 其中包含 1 个 1×1 卷积层、1 个 3×3 卷积层再加上 1 个 1×1 卷积层。图 2(b) 为所提方法,将 1 个多轴自注意力模块嵌入到深度残差网络的 Bottleneck 中,对特征图进行空间上的局部建模和全局建模。

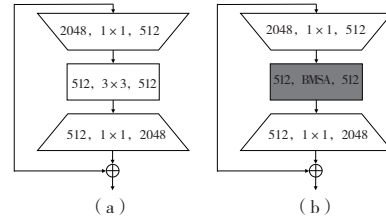


图 2 Bottleneck 对比

具体计算过程如图 3 所示。BMSA 的计算方式分为两步,第一步将特征图分成多个子块(本文中默认是 4 块),被分割的特征子图分别通过多头自注意力层,实现对特征图中局部特征的交互。第二步将前面分割得到的多个特征子图进行错位拼接,再经过全局自注意力层得到结果。错位拼接可以使 BMSA 模块关注到更远距离的交互信息,有利于模型捕捉更加多样化的特征关系。

接下来是方法的数学描述。 X 是每个 Bottleneck 的输入特征图, X_1, X_2, X_3, X_4 分别是第 1、2、3、4 块特征子图,而 $x_i^1, x_i^2, \dots, x_i^n$ ($i=1, 2, 3, 4$) 则是每个特征子图变换成向量后的组成元素,其中 n 为特征图的元素个数。首先进行特征图分割,将输入特征图分割成多个子图(这里默认分为 4 块):

$$X = (X_1, X_2, X_3, X_4) \quad (1)$$

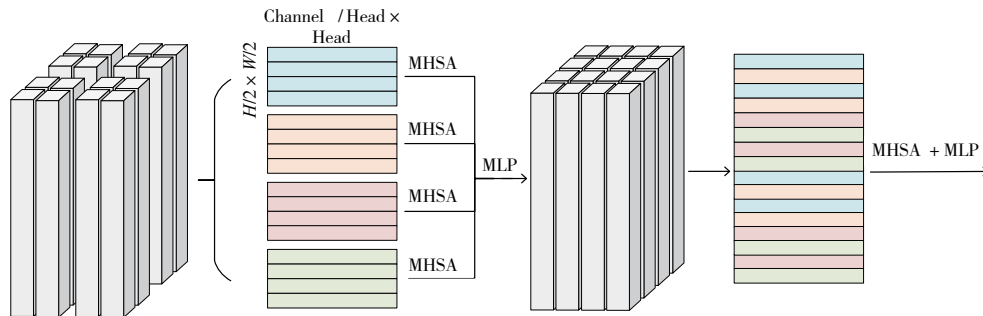


图 3 瓶颈多轴自注意力模块

计算过程中,每个特征图子图可由式(2)表示。

$$X_i = \begin{bmatrix} x_i^1 & x_i^2 & \cdots & x_i^{n/4} \\ x_i^{n/4+1} & \cdots & x_i^{n/2} & \\ \vdots & & \vdots & \\ x_i^{3n/4+1} & \cdots & x_i^n & \end{bmatrix} \quad (i=1, 2, 3, 4) \quad (2)$$

分割后的特征图进入多头自注意力层 (Multi-Head Self-Attention, MHSA), 如式(3)所示。

$$X_i = \text{LN}(F_{\text{MHSA}}(x_i^1, x_i^2, \dots, x_i^n)) \quad (i=1, 2, 3, 4) \quad (3)$$

经过局部自注意力模块后,会再经过 LayerNormalization (LN) 层。LayerNormalization 层

的作用是防止多次的特征图重塑操作。

特征图 X_1 、 X_2 、 X_3 、 X_4 经错位拼接得到和输入特征图同维度的 X^* , X^* 经过全局自注意力层,如式(4)所示。

$$O_{\text{output}} = \text{LN}(F_{\text{MHSA}}(X^*)) = \text{LN}(F_{\text{MHSA}}(x_1^1, x_2^1, x_1^2, x_2^2, \dots, x_4^n)) \quad (4)$$

经过全局自注意力部分的计算,同样再经过一个 LayerNormalization 层,得到一个 BMSA 模块的输出结果 O_{output} 。式(4)中 F_{MHSA} 的计算过程如式(5)所示。

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T)V \quad (5)$$

式中 Q 、 K 、 V 均由 $[x_i^1, x_i^2, \dots, x_i^n]$ 得到。将自注意力建模过程抽象为 $\Phi(X)$, 局部自注意力过程如式(6)所示。

$$\Phi(X) = \text{concat}(\gamma(X_1), \gamma(X_2), \gamma(X_3), \gamma(X_4)) \quad (6)$$

经过当前层 BMSA 的错位拼接后,下一个 BMSA 模块中的局部自注意力过程则可以表示为

$$\Phi(X) = \text{concat}(\gamma(X_1^r), \gamma(X_2^r), \gamma(X_3^r), \gamma(X_4^r)) \quad (7)$$

$X_i^r (i=1, 2, 3, 4)$ 表示经过变形重组得到的特征子图,局部自注意力能捕捉特征图上远距离的特征交互信息,这是 BMSA 表现优于 MHSA 的主要原因。模型中的局部注意力模块能够增加无人机在飞行过程中识别细小的电线、树枝等物体的准确率。

2 实验

2.1 数据集

为了提高模型对无人机飞行障碍物识别的有效性,制作障碍物数据集。从各公开数据集中挑选飞行障碍物的图片,主要有电线、树木、飞行物(鸟类等)和大型障碍物(山体等)4个类别,每个类别大约1500张训练图片和50张验证图片。为了验证避障模型对多种物体的识别能力和泛化能力,还在 CIFAR、FLOWER-102、Tiny-ImageNet 和 ImageNet 上进行对比试验。

2.2 训练细节

对比实验中模型使用 weight decay 为 0.000 1, momentum 为 0.9 的 SGD 优化器进行训练。Batchsize 设置为 128, 初始学习率固定为 0.1, 随着实验进行学习率逐渐衰退。在实验的训练过程中,对所有训练数据运用随机裁剪和随机水平翻转进行数据增强。

1) 障碍物数据集实验

主流的图像识别模型在障碍物数据集上进行对比实验,实验结果如表1所示,改进后的 MS-ResNet 实现 99.10% 的准确率,能够准确地识别无人机在飞行过程中可能遇到的障碍物,从而对场景中的障碍物实时监测,对有危险的飞行路径进行提前预警,并且 MS-ResNet 相比 ResNet、VGG 以及 BoT 等模型可实现更高的准确率。这表明采用 MS-ResNet 作为避障系统的图像识别模型能达到最好的效果,能最准确地对无人机飞行路径进行实时预警,以保障避障系统效果最大化。

表1 预警模型准确率对比

模型	验证准确率/%
ResNet50	96.86
ResNeXt	96.89
VGG16	96.21
DenseNet	96.91
BoT	97.07
MS-ResNet	99.10

随机选择无人机飞行过程中拍摄的图片对预警模型进行测试,以验证避障模型对实际障碍物的识别置信度。效果图如图4所示,图4中(a)、(b)、(c)、(e)、(f)和(g)识别置信度都在90%以上,其中图4(d)和图4(h)对电线和树木的识别置信度分别为73%和77%,能够准确判断出无人机在飞行过程中是否遇到障碍物,达到预警的目的。

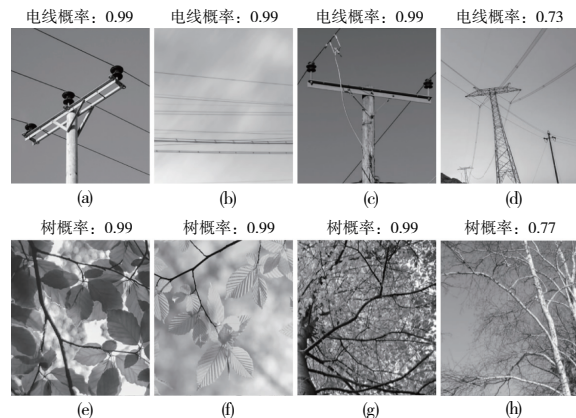


图4 图片识别效果

2) 公开数据集实验

为了进一步验证预警模型的效果,分别在不同分辨率的公开数据集上进行实验。MS-ResNet、BoT、

ResNet-50、ResNeXt-50^[10]、ResNet101、VGG16 和 DenseNet 在 CIFAR^[11]、FLOWER-102^[12]、Tiny-ImageNet^[13]和 ImageNet 数据集上进行训练。为了适应小分辨率数据的尺寸,将网络都进行相同的修改,将下采样次数减少为 3 次。

实验结果如表 2 所示,无论是在 CIFAR 和 Tiny-ImageNet 等小分辨率数据集上,还是在 FLOWER-102 和 ImageNet 等大分辨率数据集上,MS-ResNet 相比同类型的 BoT 实现更好地验证准确率。MS-ResNet 在 CIFAR10、CIFAR100、Tiny-ImageNet 和 ImageNet 上的准确率分别比 BoT 高 0.17、1.1、0.58 和 0.36 个百分点。在数据量较少的 FLOWER-102 上,MS-ResNet 的准确率比 BoT 高 5.62 个百分点。

表 2 CIFAR 和 ImageNet 等数据集实验 单位: %

模型	CIFAR10	CIFAR100	Tiny- ImageNet	FLOWER- 102	ImageNet
ResNet50	96.08	77.15	66.27	75.97	76.00
ResNeXt	96.45	77.33	67.08	74.50	77.00
VGG16	96.27	76.84	67.15	80.67	72.21
DenseNet	96.89	80.96	65.62	80.72	76.2
BoT	96.25	76.50	66.13	69.45	77.00
MS-ResNet	96.42	77.60	66.71	76.07	77.36

考虑到预警系统要求实时性,预警模型的计算效率和准确率同样重要。为了对比模型的计算资源消耗,采集部分模型在 CIFAR100 数据集上的吞吐量(单张 NVIDIA V100 上每秒计算图片数),同时得到模型的参数量对比。如表 3 所示,MS-ResNet 的参数量要稍高于 BoT 和 ResNet50,但是 MS-ResNet 在上述 3 个数据集上得到的验证准确率都高于 BoT 和 ResNet50。这表明 MS-ResNet 的泛化能力要优于 BoT 和 ResNet50,并且 MS-ResNet 的吞吐量仅仅稍低于 BoT 和 ResNet50。

表 3 模型吞吐量和参数量对比

模型	吞吐量(图片数/s)	参数量($\times 10^6$)
ResNet50	2 439	23.71
ResNeXt	1 837	23.18
VGG16	2 985	134.67
DenseNet	1 162	18.29
BoT	2 421	18.98
MS-ResNet	2 083	26.10

表 2 和表 3 数据表明,DenseNet 在上述 3 个数据集上的表现都稍优于 MS-ResNet,但是 DenseNet 的吞吐量远远低于 MS-ResNet。BMSA 在提升模型准确率的同时,没有大幅降低模型的计算效率,也能保证预警系统的实时性,可见 BMSA 嵌入卷积网络的方案是可行的。

3 结语

本文论述了基于图像识别模型对无人机的飞行路径进行实时监测,以达到避障的目的。对 ResNet 和 BoT 进行改进,实验表明改进图像识别模型能够对障碍物实现更准确的识别。将多轴自注意力模块 BMSA 嵌入到原本模型中,引入局部自注意力单元和全局自注意力单元,解决了模型在识别细小物体方面表现不佳的问题,提升了系统的避障准确率。大量实验结果表明,深度卷积网络结合 BMSA 能在图像识别任务中实现更高的验证准确率,同时并没有大幅降低模型的吞吐率。此外,BMSA 相比于同类型的结构表现出了更好的泛化性能,在数据量较少的情况下,BMSA 能帮助深度卷积网络获得巨大的准确率提升。

参考文献:

- [1] 张香竹,张立家,宋逸凡,等. 基于深度学习的无人机单目视觉避障算法[J]. 华南理工大学学报(自然科学版),2022,50(1): 101-108,131.
- [2] 肖英楠,孙抒雨. 基于改进 SIFT 图像匹配的无人机高精度避障算法设计[J]. 机械制造与自动化,2022,51(1): 237-240.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [4] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [5] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 1-9.
- [6] 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, NV, USA: IEEE, 2016: 770-778.
- [7] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale [C]//International Conference on Learning Representations, Vienna, Austria: [s.n.], 2021.

- 收稿日期:2023-02-01

- [4] HICKS R M, HENNE P A. Wing design by numerical optimization [J]. Journal of Aircraft, 1978, 15 (7): 407-412.
- [5] ROHRBACH C. A report on the aerodynamic design and wind tunnel test of a Prop - Fan model [C]//12th Propulsion Conference. Palo Alto, CA, USA. Reston, Virginia: AIAA, 1976: 667-678.
- [6] WAINAUSKI H, VACZY C. Aerodynamic performance of a counter rotating Prop-Fan [C]//22nd Joint Propulsion Conference. Huntsville, AL, USA. Reston, Virginia: AIAA, 1986: 1550-1560.
- [7] HANNIGAN T, WAINAUSKI H. Wind tunnel results of counter rotation prop-fans designed with lifting line and Euler code methods [C]//27th Joint Propulsion Conference. Sacramento, CA, USA. Reston, Virginia: AIAA, 1991: 2499-2505.
- [8] 陈博, 贺象. 国外桨扇技术发展概况 [J]. 燃气涡轮试验与研究, 2020, 33(1): 54-58.
- [9] 袁培博, 李博, 汤宏宇, 等. 一种高速对转桨扇的设计参数对性能的影响 [J]. 机械制造与自动化, 2022, 51(3): 159-163, 177.
- [10] YU M L, WANG Z J, HU H. Airfoil thickness effects on the thrust generation of plunging airfoils [J]. Journal of Aircraft, 2012, 49(5): 1434-1439.
- [11] KELLOGG M, BOWMAN J. Parametric design study of the thickness of airfoils at Reynolds numbers from 60,000 to 150,000 [C]//42nd AIAA Aerospace Sciences Meeting and Exhibit. Reno, Nevada. Reston, Virginia: AIAA, 2004: 1054-1061.

- 收稿日期: 2023-01-09