

基于深度学习的航空监视方法研究

王艳明 王宝珠

(河北工业大学 电子信息工程系 天津 300401)

摘要:我国是一个幅员辽阔的国家,地理条件复杂,常规的国土安全巡检方法会耗费大量人力物力。为此,提出了一种基于深度学习的航空监视方法,其利用无人机从高空采集图像,并利用卷积神经网络对采集图像进行分类判断,从而对场景进行监视。其目的在于用人工智能的手段,通过无人机代替人工进行巡检,从而提高国土安全监视效率。为此,本文建立了包含10种不同场景的俯视视角的数据库。通过卷积神经网络模型,对不同场景的图像特征进行学习,使得模型可以分辨不同的场景。为了验证本方法的可行性,本文在10种空基视角的数据库上进行了实验,结果显示其分类准确率达到97%。说明本方法可满足安全监视的需求,为实现智能监视提供了思路。

关键词:航空监视;俯视视角;深度学习;国土安全;人工智能

中图分类号: TP391.4; TN919.81 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Research on aviation monitoring method based on deep learning

Wang Yanming Wang Baozhu

(Department of Electronic Information Engineering, Hebei University of Technology, Tianjin 300401, China)

Abstract: Our nation has a vast territory, which holds variate geology and climate condition. In order to ensure homeland security, it is usually necessary for the relevant personnel to carry out routine inspections, which will consume a lot of manpower and resources. To this end, this paper proposes an aeronautical surveillance method based on deep learning, which uses drones to collect images from high altitude, and uses convolutional neural networks to classify and judge the collected images to monitor the scene. The purpose is to use artificial intelligence to replace the manual inspection by drones, thereby improving the efficiency of homeland security monitoring. To this end, this thesis establishes a database of top-down perspectives containing 10 different scenarios. Through the convolutional neural network model, the image features of different scenes are learned, so that the model can distinguish different scenes. In order to verify the feasibility of this method, this paper carried out experiments on 10 kinds of space-based perspective databases, and the results showed that the classification accuracy reached 97%. This shows that this method can meet the needs of security monitoring and provides ideas for intelligent monitoring.

Keywords: aviation surveillance; view angle; deep learning; homeland security; artificial intelligence

0 引言

我国是一个幅员辽阔的国家,地理气候条件复杂,广袤的中华大地上包含了平原、丘陵、山地、高原、湖泊、河流和海洋等各种地形^[1]。而且人口分布严重不均匀,东部沿海地区人口密度远远大于西部。此外,西部沙漠地带还存在很多无人区^[2]。在这种条件下,如何确保国土安全、监视每一片土地就成为一个问题^[3]。同时,我国也是世界上自然灾害最严重的国家之一。中国的自然灾害具有种类多、发生频率高、灾情严重的特点。水灾、旱灾、地震、台风、雪灾、山体滑坡、泥石流和森林火灾等自然灾害几乎每年都有发

生。当灾害发生时,救灾勘探人员往往很难进入现场,如何尽快了解整个区域的事态并对灾情做出准确评估是整个救灾抢险的关键^[4]。

近年来,随着国内低空空域的逐步开放,我国通用航空事业发展迅猛,成为了经济发展的新增长点^[5]。而这种发展,也促进了低空飞行器在国民经济建设、工业发展中的应用^[6]。无人机在抢险救灾、遥感测绘、节日表演等事宜中都有巨大的应用潜力,在近几年也表现出了越来越重要的作用。由于无人机“飞得高,看得远”,因此,对于我国这样一个面积大、地形复杂且自然灾害频繁的国家,将低空无人机与监视技术相结合,研究基于空基的航空监视,对于国土安

全有着巨大的意义。

对于航空监视,通常的做法是将高清摄像头装载在无人机上,通过航拍采集地面图像信息,并利用计算机视觉的相关算法来进行分析、判断。而这又涉及到计算机视觉中图像分类的问题。传统的计算机图像分类常使用支持向量机(support vector machine, SVM)。而近几年,深度学习技术受到了越来越广泛的关注,其在图像处理、模式识别、场景分类^[7]、目标检测^[8]等很多领域都得到了广泛的应用,越来越多的新的卷积神经网络模型相继被提出,并取得了很好的效果。

基于上述分析,本文融合了当下的前沿知识,对基于深度学习方法的航空监视方法进行了研究,建立了空基视角下不同场景的视频图像数据库,并在该数据集上进行仿真实验加以验证。结果显示,本文所述方法和所建立数据库可以有效应对航空视角下的场景分类,在空基监视中有着较大的应用潜力。

1 问题分析与解决方案

基于视频图像处理技术的航空监视,其最关键的步骤要属图像分类,其处理图像样本数的能力、分类效果的好坏,直接影响监视系统的性能。

1.1 图像分类方法概述

图像分类是根据不同图像所包含的不同信息、不同特征,把不同类别的图像区分开来的方法。它通常利用计算机对图像进行定量分析,以此代替人的视觉判读。

传统的图像分类方法,通常需要人为设计并提取一些特定模式的特征,比如边缘特征、纹理特征、颜色特征等。通过提取这些特征,并将特征很好地组合在一起,最终得到图像的最终表达。在这个阶段,研究人员更多专注于如何设计并提取到更好的特征,以及如何更好地组合这些特征。在这一时期,人们研究出很多高效的表述子,如尺度不变特征变换(scale-invariant feature transform, SIFT)^[9-11]、方向梯度直方图(histogram of oriented gradient, HOG)^[12-13]等。但是,受制于计算量以及描述子本身的性能,该方法很难处理大规模多样本的图像分类问题。

1.2 卷积神经网络

最早的卷积神经网络可以追溯到 1994 年提出的 LeNet5,它可以被视为卷积神经网络的前驱,而且有效地推进了深度学习的发展进程。LeNet5 的网络架构非常基础,特别是发现了图像特征分布于整个图像之中。不过受制于硬件的限制,在当时并没有 GPU 可以用于加速训练,因此当时该网络并未在深度学习领域中引起很大的震动。

从 1998 年到 2010 年期间,神经网络的发展处于一个真空期。但是,随着手机和数码相机的普及,以及互联网的飞速发展,使得图像数据变得越来越多,这让神经网络有了更多的训练样本。同时,计算机的运算能力也在不断发展,图形处理器(graphics processing unit, GPU)成为了通用计

算工具。这些都为神经网络的飞速发展奠定了基础^[14]。

在 2012 年,Alex 团队提出了 AlexNet。该神经网络在 LeNet 的基础上调整了网络架构并加深了深度,在当年的 ImageNet 大赛上以明显优势夺得了冠军。AlexNet 像一个巨大的石头投入平静的湖面一样,对整个计算机视觉领域产生了巨大冲击。之后几年,更多越来越强大的网络被逐渐提出:2013 年,VGG 网络被提出,其网络深度最深达到了 19 层,将卷积神经网络的深度推向了新的高度。由于卷积神经网络能够提取图像的特征,而网络的层数越多,意味着能够提取到更多不同的特征。因此,通常认为,网络越深,得到的特征越丰富、模型的表达能力越强、分类性能越好。

但是,当进一步简单堆叠来增加网络深度时,神经网络的性能并没有增加,反而会出现下降,这种现象被称为网络的退化问题。为此,2015 年,微软亚洲研究院提出了残差神经网络(ResNet)^[15]。其在将网络的深度继续提升的同时,通过构造同构映射形成了残差块,解决了网络退化问题,即能够让网络的性能不会因为深度的增加而下降。

总体来看,人们更多地把卷积神经网络看作一个特征提取器,其替代了传统的人工设计特征的方法,可以像“黑盒子”一样自主提取输入图片的特征。从近几年发展来看,其分类效果明显好于传统的特征设计方法,这说明其提取的特征具有更强的表达能力。因此,在本文也采用了卷积神经网络提取图像特征。

1.3 空基监视技术现状

在我国现阶段,有些管理部门可能会让相关人员搭乘直升机,从空中进行安全检查,但很少利用无人机进行自主空基监视。综合来看,其难点主要有 3 个^[16-18]:

1)地面环境比较复杂,山地、森林、河流、丘陵交错纵横。人工智能系统很难做出准确判断。

2)由于摄像头装载在无人机上,所以采集到的图像都是俯视视角。但是,常规的图像分类数据库,比如 ImageNet,图片通常都是平时视角。如果用此类数据库训练、测试模型,可能会导致模型所提取特征和实际应用的特征分布会有很大差距。

3)俯视视角的图片训练数据集比较缺乏,这为训练一个可以较准确地地区分俯视视角下不同场景的模型增加了很大难度。

为此,本文首先收集了航拍视角下不同场景的视频资料,建立俯视场景的图像数据库。通过这些数据,可以更好地训练模型以应对空基航空监视的需要,也让文中的仿真实验更加具有真实性和代表性。

2 方 法

本文将从空基场景数据库构建、特征提取网络模型的建立、特征融合方法和分类器设计 4 个方面介绍空基场景分类的具体实现方法。其基本流程是从数据集中抽取图像

送入特征提取网络,提取特征后进行特征融合,得到视频的特征表达,再送入分类器进行分类判决。

2.1 空基数据库构建

本文提出,将高清摄像头装载在无人机上,在高空平台上拍摄地面的状况,并根据图像内容进行分类判断。但是,如果用平视视角下的图像训练模型可能很难适应空基监视的任务。为此,本文收集了10个不同场景的俯视图,并将它们处理后用于模型的训练。这样,最大程度地模拟了空基视角下图像的状态,为空基监视提供了良好的训练数据,为得到一个适应性强的模型打下了良好的基础。

本文建立的空基视频数据库包括洪水、森林火灾、湖水、泥石流、高压线、草原、公路、雪地、海啸和火山喷发10个不同类型的场景,每个场景包括10个不同视频。具体视频图片如图1所示。



图1 数据库示例图片

这些视频都是无人机航拍得到的,皆为俯视视角。这些图像资料和无人机巡检中的实际视角是相似的。这样,通过这些数据,可以得到一个适应性较强、具有较强表征能力的分类模型。

在数据集中,不同场景的视频所包含的总帧数有所不同,表1所示为不同类型的场景和数据集中帧数的对应关系。

表1 数据库各场景所含帧数对照表

场景	帧数
洪水	5 822
森林火灾	3 775
湖水	4 992
山体滑坡	2 584
电力线	3 672
草原	2 553
公路	5 385
雪地	5 397
海啸	3 960
火山喷发	2 620

2.2 残差神经网络

由于卷积神经网络具有较强的表达能力,本文将其作为特征提取器来对图片进行处理,并选择了残差神经网络ResNet-50用以提取特征。

在ResNet-50模型中,最核心的模块就是如图2所示的残差块。

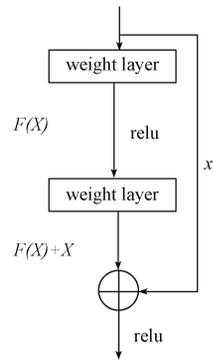


图2 残差块示意图

通过这个模块,可以将要优化的目标函数,从 $H(x)(H(x)=F(x)+x, x$ 是这个残差块的输入)转化为 $H(x)-x$ 。如前所述,简单地在浅网络上堆叠卷积层增加网络深度很难提升网络性能。但是,如果新堆叠的模块通过学习可以达到恒等映射的效果,即使得这一层的输出等于输入。那么,增加模块后至少可以达到浅层网络的效果,即深层网络至少不应该比浅层的效果差。但是事实并非如此,增加新的模块后,网络最终达到了一个更差的效果。因此,ResNet网络构造了如图2所示的残差块,其通过一个恒等映射,将浅层的中间结果直接引入给深层网络。这时候就不需要把上面几层训练到一个等价映射,而是将其逼近于0就可以近似得到浅层网络的结果,这样训练的难度比训练到一个等价映射应该下降了很多。

2.3 多源特征融合方法

由于神经网络每次只能处理单个图片,因此,对于每段场景的视频,首先需要进行解帧,解帧后得到多个图片。训练时,每次从一段视频中随机选取若干帧,并将这些帧依次送入网络,网络最终的输出就是对于该图片提取出的特征。对于ResNet-50而言,其输出是一个长度为4 096维的特征向量。由于一段视频包含多帧,为了得到视频的最终表达,需要将每帧提取到的特征进行融合,最终得到视频的全局表示。

本文尝试了3种不同的方法:求和融合法、点乘融合法和最大值函数融合法。具体表达式如表2所示。式中, $f(X)$ 表示融合后的特征表达; x_i 表示输入的某一帧; $f(x_i)$ 表示融合前每帧的特征表达。通过这3种方式,可以得到3种不同的特征融合结果,并最终得到一个4 096维的表达,其即为该视频的总特征表达。

表2 不同融合方法表达式

名称	表达式
求和融合法	$f(X) = \sum_i f(x_i)$
点乘融合法	$f(X) = \prod_i f(x_i)$
最大值函数融合法	$f(X) = \text{MAX}_i f(x_i)$

2.4 全连接层分类器

对于每一段视频,通过卷积神经网络提取了一段 4 096 维的特征向量,需要一个分类器来区分这些特征,进而将不同类别的视频分开。在这里,本文选择了两个全连接层作为分类器,全连接层(fully connected layers, FC)在整个卷积神经网络中起到“分类器”的作用。全连接层的每一个结点都与上一层的所有结点相连,用来把前边提取到的特征综合起来。由于其全相连的特性,一般全连接层的参数也是最多的。

本文设计的第一个全连接层先将 4 096 维的特征映射到 512 维的中间向量,之后,再将其映射到维度为 10 的标记空间,得到一个维度为 10 的最终向量,称之为分类向量;最终,用 softmax 函数处理这个分类向量。Softmax 函数的公式如式(1)所示。

$$S(p_j) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (1)$$

式中: K 为分类种类数; z_j 表示第 j 类在分类向量中的值; $S(p_j)$ 为 softmax 函数输出结果中对应第 j 类的值,其表示视频属于第 j 类的置信概率。这里, K 的取值为 10,即分别代表 10 类视频。显然,对于 softmax 的输出结果,有 $\sum_{j=1}^K z_j = 1$,这与实际意义也是一致的。置信概率是用来衡量统计推断可靠程度的概率,表示模型认为该视频属于某一类有多大“把握”。如果某一类的值越大,越接近 1,则模型认为输入视频是该类的可能性越大;反之如果越接近 0,则模型认为该视频是某类别的可能性很小。

3 实验及结果分析

根据前文所述,为了更好地仿真空基视角下的场景识别与分类问题,收集并建立了包含空基视角下不同场景的视频图像数据库。实验中,将数据集内图片划分为两部分,从中选取 80% 的图片作为训练集,用来进行模型训练;20% 的图片作为测试集,用来进行仿真验证。具体的仿真验证结果通过分类准确率来判定,即测试集中,模型预测结果和真实结果一致的比例。具体到每次训练和测试时,每段视频选取的帧数为 5。

3.1 不同融合方法比较

首先,验证了不同的融合方法,具体结果如表 3 所示。

表 3 不同融合方法分类精度

名称	分类精度/%
求和融合法	97
点乘融合法	93
最大值函数融合法	90

结果显示,求和融合法的分类精度最高,这说明其能最好地融合多帧之间的特征,构建视频的整体特征表达。

3.2 不同场景分类效果

用求和融合法进一步测试了不同场景的分类准确率,具体结果如表 4 所示。

表 4 不同场景分类精度

场景	分类精度/%
洪水	90
森林火灾	100
湖水	90
山体滑坡	100
电力线	100
草原	100
公路	100
雪地	100
海啸	90
火山喷发	100

结果显示,对于大部分场景,模型都可以准确分类判断。但是,在洪水、湖水和海啸 3 个场景中,该方法均存在一定程度的误判。综合分析认为,洪水、湖水和海啸的场景十分相似,都是以水面为基础,只是在形态上可能存在一些较小的差异。如图 3 所示,从左至右,分别是湖面、海啸和洪水的场景图,可以发现其在图像上具有较高的相似性。因此,模型在判断时出现了一些错误。而对于像森林火灾、火山喷发、草原等一些有明显特征的场景,模型都较好地做出了判断。



图 3 湖水、海啸、洪水示例图

综上所述,该仿真结果说明,该模型可以较好地适应空基视角下场景分类的任务。

4 结 论

本文主要研究了如何利用无人机的空基巡检,代替人工巡检,实现对国土安全的监视。为此,本文在传统图像分类方法的基础上,从两个方面开展了研究工作。一方面,建立了 10 种不同场景的俯视视角的图像数据库;另一方面,利用卷积神经网络对不同场景进行分类。和传统的图像分类研究相比,本文具有更强的工程价值。构建的空基视角的数据库,也让卷积神经网络能更好地在无人机巡检中发挥更好的作用。通过仿真实验,用不同场景的图像对模型进行了验证,结果显示该方法可以较好地地区分不同场景,在无人机巡检中可以较好地应用,有较好的发展前景。

参考文献

- [1] 陈丞.我国的国土安全形势分析及我国国防面临的严峻挑战[J].魅力中国,2014(12):16.
- [2] 周国华.我国自然资源管理的初步研究[J].经济地理,1992(2):24-29.
- [3] 肖兴威.中国森林资源与生态状况综合监测体系建设的战略思考[J].林业资源管理,2004(3):1-5.
- [4] 池肖杰,韩光迎.中国的水资源[J].地理教育,2010(10):17.
- [5] 高启明,金乾生.我国通用航空产业发展特征、关键问题及模式选择[J].经济纵横,2013(4):98-102.
- [6] 陈蓓蓓,曾小舟,阎雷.国际通用航空发展比较及我国通用航空发展策略[J].南京航空航天大学学报(社会科学版),2012,14(2):37-41.
- [7] 田仲富,黎粤华,刘晓义.基于图像的嵌入式森林火灾监控系统研究与设计[J].安徽农业科学,2014,42(13):4105-4107.
- [8] 韩亚伟,张有志,李庆涛,等.动态场景监控系统中人数统计算法的研究[J].计算机应用与软件,2011,28(2):260-262.
- [9] 傅卫平,秦川,刘佳,等.基于 SIFT 算法的图像目标匹配与定位[J].仪器仪表学报,2011,32(1):163-169.
- [10] 方小艳.基于 SIFT 特征点检测与维纳滤波的图像复原算法[J].电子测量技术,2017,40(6):105-108.
- [11] 白廷柱,侯喜报.基于 SIFT 算子的图像匹配算法研究[J].北京理工大学学报,2013,33(6):622-627.
- [12] 蒋美云,郭雷,徐梦珠.基于 HOG 算子的手型特征提取算法[J].计算机应用与软件,2015,32(12):326-329.
- [13] 万源,李欢欢,吴克风,等.LBP 和 HOG 的分层特征融合的人脸识别[J].计算机辅助设计与图形学学报,2015,27(4):640-650.
- [14] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(6):1229-1251.
- [15] 张川.面向图像分类的深度残差网络优化结构研究[D].北京:中国科学院大学,2016.
- [16] 曹林林,李海涛,韩颜顺,等.卷积神经网络在高分遥感影像分类中的应用[J].测绘科学,2016,41(9):170-175.
- [17] 马强,罗喜伶.空基交通监视系统中的运动目标检测方法研究[J].2008 第四届中国智能交通年会,中国青岛,2008:769-774.
- [18] 司朋举,胡伟.一种改进的神经网络车牌识别算法研究[J].电子测量技术,2016,39(10):100-103.

作者简介

王艳明,本科,主要研究方向为空基监视。

E-mail:13311081336@163.com

王宝珠,教授,主要研究方向为智能信息处理、图像处理。

E-mail:w bz_bz_china@126.com