

中文图书分类号：TP24

密 级：公开

UDC：621

学 校 代 码：10005



硕 士 学 位 论 文

MASTERAL DISSERTATION

论 文 题 目：雪龙号极地科考船全天相机系统的设计与研究

论 文 作 者：崔顺

学 科：机械工程

指 导 教 师：苏丽颖 副教授

论文提交日期：2019年5月

UDC: 621
中文图书分类号: TP24

学校代码: 10005
学 号: S201601079
密 级: 公开

北京工业大学工学硕士学位论文

题 目: 雪龙号极地科考船全天相机系统的设计与研究

英文题目: DESIGN AND RESEARCH OF THE ALL-SKY
CAMERA SYSTEM OF THE XUELONG
POLAR SCIENTIFIC RESEARCH SHIP

论 文 作 者 : 崔 顺
学 科 专 业 : 机 械 工 程
研 究 方 向 : 机 构 及 机 器 人 系 统 分 析 与 控 制
申 请 学 位 : 工 学 硕 士
指 导 教 师 : 苏 丽 颖 副 教 授
所 在 单 位 : 机 械 工 程 与 应 用 电 子 技 术 学 院
答 辩 日 期 : 2019 年 5 月
授 予 学 位 单 位 : 北 京 工 业 大 学

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京工业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已体现在论文中作了明确的说明并表示了谢意。

签 名： 崔顺
日 期： 2019年5月27日

关于论文使用授权的说明

本人完全了解北京工业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内 容，可以采用影印、缩印或其他复制手段保存论文。

(保密的论文在解密后应遵守此规定)

签 名： 崔顺 日 期： 2019 年 5 月 27 日
导师签名： 苏研新 日 期： 2019 年 5 月 27 日

摘要

在天文观测领域,云量是天文台站选址和可见光成像质量的重要影响因素之一。全天相机是目前国际上测量云量的通行做法之一,具有观测自动化、成本低等特点。然而现有全天相机系统的鲁棒性较差,对复杂环境的适应能力较弱,且缺少后期数据处理能力。针对这些问题,本文设计研发了一种适用于高温、寒冷、高湿度、高腐蚀性环境下的无人值守全天相机系统,拥有自主拍摄、远程数据传输、数据处理等功能,并将其搭载在“雪龙号”极地科考船上,随船完成了第8次北极科考和第34次南极科考,充分验证了设备在极端环境下的可靠性。此外,基于设备拍摄的云图数据进行了后期处理算法和数据发布网站的开发。本文主要包括以下几个部分:

(1) 全天相机系统硬件设计:完成了设备的硬件结构设计、硬件选型和硬件集成,并对设备进行可靠性检验,确保设备能够安全、稳定地进行工作。将全天相机系统搭载在极地科考船雪龙号上,随船航行并进行拍摄,验证系统性能,获取全天空地基云图数据。

(2) 全天相机系统软件设计:设计了观测组件、存储组件和维保组件,并使用C#语言编写实现。将这些软件部署在系统的工控计算机中,实现相机自动化拍摄、数据存储、数据传输等功能。

(3) 全天空地基云图数据分类算法设计:对系统获取的全天空地基云图数据进行数据分析和人工标注,制作雪龙号全天空地基云图数据集。实现了基于卷积神经网络的数据分类算法设计,并基于数据集对模型进行了训练和优化。此外,对云南丽江天文观测站全天相机数据进行了人工标注和预处理,并使用该数据对模型进行异分布数据集的验证,证明算法具有良好的可迁移性。

(4) 云点自动识别算法对比:基于系统获取的少云类全天空地基云图,使用Python语言编写程序,验证云点识别算法,包括固定阈值法、自适应阈值法以及k-means聚类算法,对三种算法的准确性以及适用场景进行对比分析和总结。

(5) 全天相机观测数据发布平台设计与实现:使用Django框架、MySQL数据库、HTML5和CSS3等技术设计搭建了雪龙号全天相机观测数据发布平台,为相关科研人员提供便捷查询、查看、下载雪龙号全天空地基云图数据集的web服务。

关键词: 全天相机系统; 全天空地基云图; 云图分类; 云量分析

Abstract

In the field of astronomical observation, cloud cover is one of the important factors affecting the location of the observatory station and the quality of visible light imaging. The all-sky camera is one of the current practices for measuring cloud volume in the world. It has the characteristics of automatic observation and low cost. However, the existing all-sky camera system is less robust, has a weaker adaptability to complex environments, and lacks post-data processing capabilities. In response to these problems, this paper designs and develops an unattended all-sky camera system suitable for high temperature, cold, high humidity and high corrosive environments. It has functions of self-photographing, remote data transmission, data processing, etc. On the Xuelong polar scientific research ship, the 8th Arctic expedition and the 34th Antarctic expedition were completed with the ship, which fully verified the reliability of the equipment in extreme environments. In addition, post-processing algorithms and data publishing sites were developed based on cloud image data captured by the device. This article mainly includes the following parts:

(1) Hardware design of the all-sky camera system: The hardware structure design, hardware selection and hardware integration of the equipment are completed, and the reliability of the equipment is checked to ensure that the equipment can work safely and stably. The all-sky camera system is mounted on the Xuelong Polar Scientific Ship, and the ship is voyaged and photographed to verify the system performance and obtain the full sky foundation cloud image data.

(2) All-sky camera system software design: Designed observation components, storage components and maintenance components, and written in C# language. The software is deployed in the industrial computer of the system to realize automatic camera shooting, data storage, data transmission and other functions.

(3) Design of the classification algorithm of the full-sky foundation cloud image data: data analysis and manual annotation of the full-sky foundation cloud image data acquired by the system, and the Xuelong full-sky foundation cloud image data set. The data classification algorithm based on convolutional neural network is designed, and the model is trained and optimized based on the data set. In addition, the all-sky camera data of the Lijiang Astronomical Observatory in Yunnan was manually labeled and preprocessed, and the data was used to verify the model's heterogeneous dataset, which proved that the algorithm has good mobility.

(4) Cloud point automatic recognition algorithm comparison: Based on the system-acquired cloud-based full-sky foundation cloud map, the program is written in Python language to verify the cloud point recognition algorithm, including fixed threshold

method, adaptive threshold method and k-means clustering algorithm. Compare and summarize the accuracy of the three algorithms and the applicable scenarios.

(5) Design and implementation of the all-sky camera observation data release platform: using the Django, MySQL, HTML5 and CSS3 and other technical design to build the Xuelong all-sky camera observation data release platform, providing relevant researchers with convenient query, view, download Xuelong full sky foundation cloud map dataset web service.

Keywords: All-sky camera system, full-sky foundation cloud map, cloud map classification, cloud volume analysis

目 录

摘 要.....	I
Abstract.....	III
第 1 章 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.3 课题来源及论文研究内容.....	5
1.3.1 课题来源.....	5
1.3.2 本文主要研究内容.....	5
第 2 章 全天相机系统硬件设计.....	7
2.1 引言.....	7
2.2 全天相机系统硬件设计方案及箱体设计.....	7
2.3 全天相机系统硬件设备集成.....	10
2.4 全天相机系统可靠性验证试验.....	11
2.4.1 低温耐受性能.....	11
2.4.2 高温耐受性能.....	12
2.4.3 除雾除霜性能.....	13
2.4.4 成像防抖性能.....	13
2.5 本章小结.....	13
第 3 章 全天相机系统无人值守软件设计.....	15
3.1 引言.....	15
3.2 全天相机系统无人值守软件设计方案.....	15
3.2.1 观测组件原理及曝光参数计算算法.....	15
3.2.2 存储组件设计方案.....	19
3.2.3 维保组件设计方案.....	21
3.2.4 系统整体设计方案.....	21
3.3 全天相机系统无人值守软件实现过程.....	21
3.3.1 观测组件实现.....	21
3.3.2 存储组件实现.....	23
3.4 软件工具与软件环境.....	25
3.5 本章小结.....	25
第 4 章 全天空地基云图数据处理.....	27

4.1 引言.....	27
4.2 全天空地基云图数据介绍.....	27
4.3 全天空地基云图数据分类.....	28
4.3.1 分类标准.....	28
4.3.2 分类方法及理论.....	30
4.3.3 卷积神经网络模型介绍.....	32
4.3.4 数据预处理.....	33
4.3.5 卷积神经网络模型训练.....	35
4.3.6 实验平台信息.....	37
4.3.7 结果分析.....	37
4.4 泛化能力验证.....	38
4.4.1 验证数据集介绍.....	39
4.4.2 验证数据预处理.....	39
4.4.3 模型验证结果分析.....	40
4.5 云量检测方法验证和对比分析.....	42
4.5.1 固定阈值法.....	42
4.5.2 基于最大类间方差的自适应阈值法.....	43
4.5.3 k-means 聚类方法.....	43
4.5.4 云点识别结果及分析.....	44
4.6 本章小结.....	45
第 5 章 雪龙号全天相机数据发布平台.....	47
5.1 引言.....	47
5.2 相关技术概述与开发环境介绍.....	47
5.3 平台总体设计.....	49
5.4 全天相机数据发布网站首页.....	50
5.5 全天相机数据展示界面.....	52
5.5.1 数据展示界面交互功能.....	53
5.5.2 数据展示界面实现方法.....	54
5.6 全天相机数据下载界面.....	54
5.6.1 数据下载界面交互功能.....	55
5.6.2 数据下载界面实现方法.....	58
5.7 平台使用注意事项.....	59
5.8 本章小结.....	60
结 论.....	61

参考文献.....	63
附录.....	67
攻读硕士学位期间发表的学术论文.....	71
致 谢.....	73

第1章 绪论

1.1 研究背景与意义

对于天文选址，云量是需要首先考虑的因素之一^[1]。天文观测目标发射出的光线到达地基望远镜的终端时，会受到大气中云的散射和吸收，大气中云量的多少，决定了观测数据质量的好坏，从而决定了天文可用时间和天文观测可视域。目前观测云的方式主要有卫星观测和地基观测两种，相较于卫星观测，地基观测是区域云量检测的重要方法，也是天文观测台站获取云量信息的主要方式。然而，云量的地基检测长期以来仍依靠天文工作者的目视识别，为减轻天文工作者的工作量，避免因观测员本身主观因素带来的判别误差，实现云量的自动化检测尤为重要。

全天相机是实现云量自动化检测的主要仪器。现阶段主流的全天相机主要由鱼镜头和互补性氧化金属半导体（Complementary Metal-Oxide Semiconductor, CMOS）或电荷耦合元件(Charge Coupled Device, CCD)组成，通过全天相机拍摄的全天空地基云图具有很高的空间和时间分辨率，其分析结果能够准确反映当地的云量覆盖与变化特征，适合用于全天云量信息监测。

天文观测选址一般在高原、极地、沙漠等极端环境地区开展，对全天相机系统的可靠性要求很高，且这些地区难以实现人员的长期值守，为此，实现一款能够适应高寒、高温、高湿度等极端环境的无人值守全天相机系统能够解决天文观测选址工作的一个痛点，而现有全天相机系统均不能完全满足需求。在设计研发符合上述条件的全天相机系统后，需要对设备运行的鲁棒性能进行多种不同恶劣场景的验证。

极地科考船雪龙号是相机鲁棒性能很好的验证平台。雪龙号是一艘北极供应船，1993年由俄罗斯的乌克兰赫尔松船厂建造，中国购进后改装成为一艘集科考、运输、破冰等多功能于一体的综合极地科学考察专用船，实物如图 1-1。雪龙号从 1994 年代替“极地”号服役至 2018 年，共已成功完成 8 次北极航行和 34 次南极航行，是中国南北极科考的重要组成部分，并且承担着大量的南北极科考任务。雪龙号每年进行两次极地科考，航行期间，随船搭载的全天相机系统将经历赤道地区 60 摄氏度以上的高温和极地零下 30 摄氏度的低温；承受长期的风吹日晒，并经受高湿度高盐分地侵蚀；此外，还会受到船只本身震动以及在经过西风带时船只大幅度摆动带来的成像挑战。这些因素均为全天相机系统的研究提供了绝好的验证环境，并且在雪龙号航行期间，随船搭载的全天相机系统可以获取

较为丰富的全天空地基云图数据，也将有利于后续地基云图相关算法的开发。



图 1-1 雪龙号极地考察船

Fig. 1-1 XueLong Polar survey ship

1.2 国内外研究现状

获取全天空图像是测量全天空云的前提^[2]。近些年来，数字图像采集设备和数字图像处理技术得到迅猛发展，国内外涌现了许多全天空地基云观测设备^[3]。在目前已有的全天空测云仪器中，大都采用 CCD 和镜头的组合方式来获取全天空地基云图。主要方式大致分为 3 种：第 1 种是 CCD 相机和鱼镜头组合对天空进行拍摄，这类仪器主要有美国加利福尼亚大学研制的全天空成像仪（Whole Sky Imager, WSI）^[4]，中国科学院研制的可见光全天空成像仪（All Sky Imager, ASI）^[5]，中国气象科学研究院研制的地基全天空云自动观测系统（Total Sky Cloud Imager, TCI）^[6]，霍娟等研究开发的地基全天空可见光成像观测系统（All-sky Digital Camer）^[7]等，如图 1-2。这种方法设计模式简单、设计结构稳定并且拍摄一次便可获得全天空图像。但是，由于鱼镜头的焦距通常是 16mm 或更短，并且视角接近或等于 180°，所以通过鱼镜头拍摄的云图在天边位置会发生一定畸变，将增大后续云量分析或云状识别工作的难度。第 2 种是通过球面镜获取天空镜像后，使用相机向下拍摄球面镜获取全天空图像，代表仪器是美国 YANKEE 公司研制的全天空成像仪（Total Sky Imager, TSI）^[8]，如图 1-3。这种方法的优点是镜头向下间接拍摄全天空图像，有利于保护镜头，但球面镜可能受外界环境影响造成影像的失真，在较暗的环境下也不容易得到高质量的全天空图像。第 3 种方式是采用 CCD 镜头旋转扫描拍摄天空并通过拼接得到全天空图像，全天红外云测量系统（Whole Sky Infrared cloud Measuring System, WSIRCMS）^[9]就是采用这种方法对八方位及天顶方向进行扫描，将得到的 9 幅辐射值图进行拼接从而得到全天空辐射值分布，如图 1-4。此方法解决了采用大视场非致冷红外焦

平面组件实现全天空红外辐射的定量测量问题，但由于采用扫描拼图技术，对设备的转动性能要求较高，同时获得的全天空图像会有 1-2 个像元的拼缝存在。

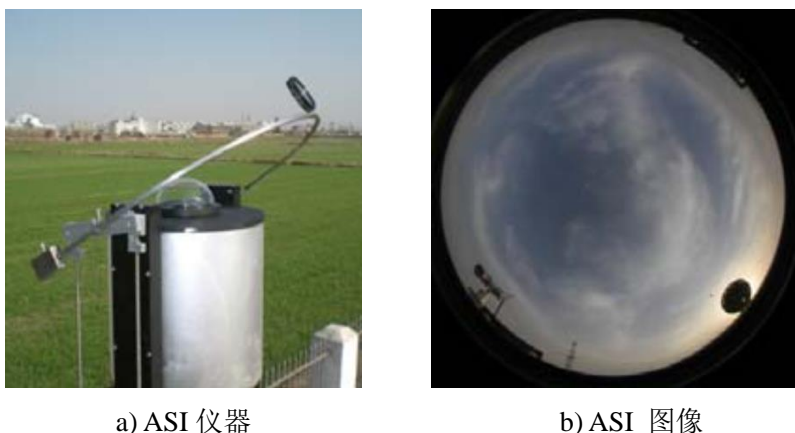


图 1-2 ASI 的仪器图片以及拍摄的全天空图像(安徽寿县)

Fig. 1-2 All Sky Imager and All Sky Images(Shou Xian, An Hui). a) All Sky Image. b) All Sky Images

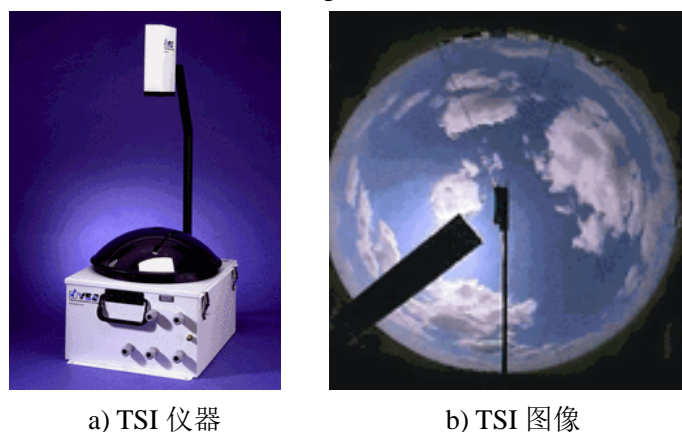
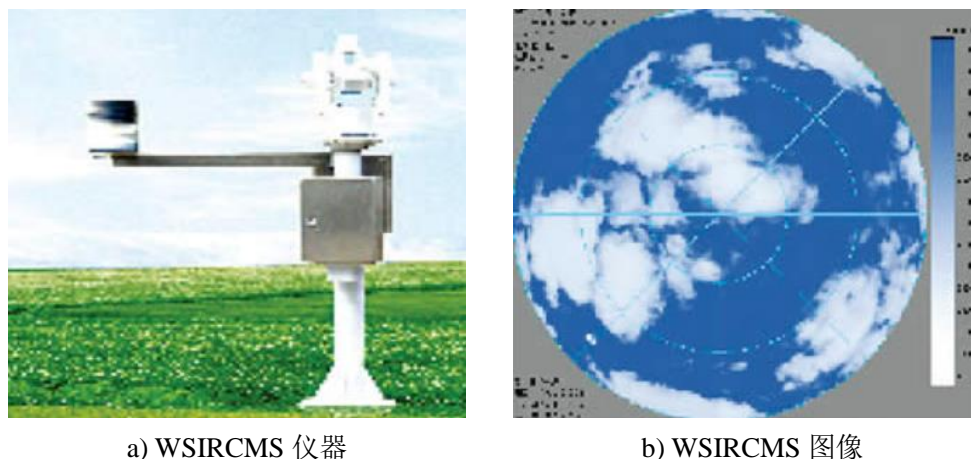


图 1-3 TSI 的仪器图片以及拍摄的全天空图像

Fig. 1-3 Total Sky Imager and Total Sky Images. a) Total Sky Imager. b) Total Sky Images

在天文观测方面，目前世界上大部分天文台均配备了全天相机，如位于智利阿塔卡玛的拉斯坎帕纳斯天文台(Las Campanas Observatory)，位于美国亚利桑那州图森市霍普金斯山顶的弗雷德·劳伦斯·惠普尔天文台(Fred Lawrence Whipple Observatory)，位于美国亚利桑那州基特峰的基特峰国家天文台(Kitt Peak National Observatory)，以及欧洲南方天文台(European Southern Observatory，ESO)在世界上的多个观测基地，这些大型天文台均配备了自主设计的全天相机。此外这些全天相机系统都配备了相应的网站进行信息公开，通过访问网页可以获得该全天相机最近一段时间内的地基云图，并且可以查看该区域这段时间内的云量信息及其变化趋势。但这些设备的硬件细节不公开，软件代码也没有开放共享，因此，观测台址的相关科研技术人员往往难以对设备在工作中出现的问题进行分析和解决^[10]。



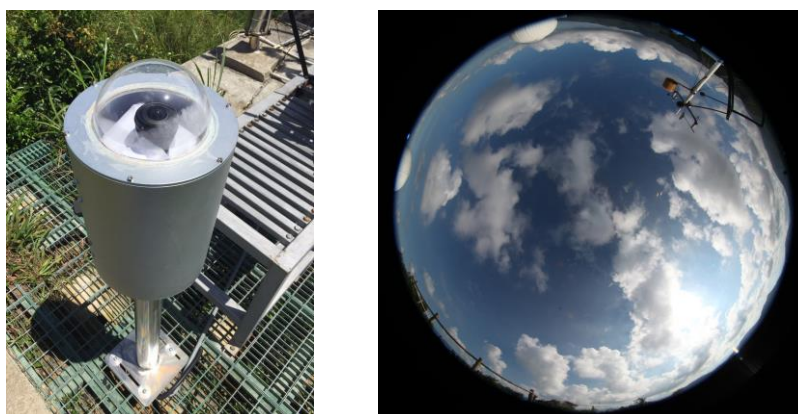
a) WSIRCMS 仪器

b) WSIRCMS 图像

图 1-4 WSIRCMS 的仪器图片以及拍摄的全天空图像

Fig. 1-4 Whole Sky Infrared cloud Measuring System and Whole Sky Infrared cloud Measuring Images. a) Whole Sky Infrared cloud Measuring System. b) Whole Sky Infrared cloud Measuring Images

国内的天文观测台站，如河北兴隆观测站、丽江观测站、青海观测站等也都拥有自主设计的全天相机系统，丽江天文观测站采取佳能 EOS 600D 和适马鱼镜头搭建了自已的全天相机系统，拍摄的全天图像清晰度好、质量高。青海观测站采取同样的相机和镜头搭建全天相机系统，并且通过使用 ARM 单片机进行控制，提高系统的集成度，从而提升了系统的稳定性，使之前平均一周 1 次的故障率降为平均一年 1 到 2 次^[11]。深圳市天文台采用 EOS 5D Mark II 单反相机和佳能 EF 8mm-15mm F4.0 变焦鱼镜头搭建全天相机系统，如图 1-5，并采用基于最大类间方差的自适应阈值算法，对云进行有效识别，实现了云图数据简单的后处理^[12]。总体来说，目前的全天相机系统在数据获取过程中容易受到环境的影响，稳定性能较差，并且在获取数据后也往往并不具备数据分析的能力。因此，一套鲁棒性高、自适应能力强并具备后期数据处理能力的自动化全天相机系统在天文观测领域中有着很强的应用需求。



a) 深圳市天文台全天相机

b) 全天图像

图 1-5 深圳市天文台全天相机以及拍摄得到的全天空图像

Fig.1-5 Shenzhen Astronomical Observatory All-sky camera and All Sky Images. a) Shenzhen Astronomical Observatory All-sky camera. b) All Sky Image

1.3 课题来源及论文研究内容

1.3.1 课题来源

本课题受国家自然科学基金（11573019，11803055），国家自然科学基金委员会-中国科学院天文联合基金（U1531246，U1731125，U1731243），中国科学院“十三五”信息化建设专项（XXH13503-03-107），国家科技部国家科技基础条件平台项目“国家地球系统科学数据共享服务平台”、“国家基础科学数据共享服务平台”（DKA2017-12-02-07）的资助。

1.3.2 本文主要研究内容

本研究以雪龙号极地科考船为验证平台，设计实现并验证了一款适用于高原、极地、沙漠等极端环境下的无人值守全天相机系统。系统可根据程序设定完成定时拍摄、自动备份数据等功能，并且能够在故障时自动恢复（非不可抗力除外）。设备在随船航行期间获取了大量的全天空地基云图，基于这些数据本研究完成了云图自动化分类算法的开发以及云点识别算法的对比验证。另外，基于 web 应用开发技术，开发了用于数据发布的 web 应用平台，实现了已获取数据的开放共享。该设备的设计研发与后续算法的开发完善均将为天文选址以及天文观测提供重要帮助。本文主要分为如下几个部分：

第一章：绪论。介绍了本文的研究背景、研究意义、国内外研究现状以及论文主要的研究内容。首先介绍了全天相机系统在天文选址和天文研究中的意义，并对极地科考船雪龙号进行介绍，阐述了雪龙号对于全天相机系统的验证作用。然后对国内外地基云图的自动化获取方法进行了分析和比较，并且对国内外天文台站的全天相机系统做了详细介绍。最后简要的给出了本文各章节的主要研究内容。

第二章：硬件设计。详细分析了全天相机系统将要面对的恶劣环境，并在此基础上进行了设备的硬件结构设计，介绍了设备选型以及设备集成，并在设备实际随船工作前进行了设备的可靠性验证，确保设备在恶劣的实际工作环境中能够正常工作。最后将设备安装到雪龙号甲板上进行随船拍摄。

第三章：软件设计。随船工作过程中，设备自动化拍摄的稳定性以及图像数据的安全性是需要考虑的重点。通过比例-积分控制器调节拍摄参数使相机稳定地进行自动化拍摄，采取多重备份的数据安全存储方法保证数据的安全性，最后添加计数器复位指令和心跳信号程序，保证相机在故障情况下能够及时重启或通知相应工作人员对设备进行检查和维修。

第四章：数据处理及应用。对设备随船拍摄过程中获取的数据进行分析和有效的人工分类，将其划分为夜晚、雨雪、过曝光、晴空、少云、多云 6 类，创建雪龙号全天相机数据集。针对分类问题设计搭建卷积神经网络模型对图像数据进行自动化分类，完成了数据预处理、模型训练、模型验证等关键步骤。使用云南丽江天文观测站的异分布数据集对模型进行认证，证明了模型良好的可迁移性能。基于少云类图像对几种常用的云点识别算法进行验证，并对算法的准确性以及适用场景进行了分析和总结。

第五章：雪龙号全天观测数据发布平台。基于 Django 框架、MySQL 数据库、Apache 服务器以及其他 web 应用开发技术，如 Ajax、Bootstrap、HTML5、CSS3 等搭建了数据发布 web 平台，实现了数据来源项目内容介绍、数据展示以及数据下载等功能，最后对平台使用时的注意事项做了一些总结。

第2章 全天相机系统硬件设计

2.1 引言

硬件系统是全天相机系统能否适应极地科考和海上航行过程中所经历恶劣环境的基础和关键。在整个硬件系统的设计实现过程中,针对工作环境进行合适硬件结构设计、器件选型和箱体设计均能增强设备对外部环境的适应能力,保证设备正常稳定地工作。对生产出的设备进行可靠性验证试验可以进一步确保设备登船后的正常工作。

2.2 全天相机系统硬件设计方案及箱体设计

从雪龙号第8次北极科考和第34次南极科考的走航线路(图2-1)可以看出,雪龙号航行期间将穿越温带、亚热带、热带、寒带等不同气候带,并且还要经过南北半球的西风带。期间设备需要能承受 -30°C 的寒冷至 60°C 的高温,并能够在镜头处结霜时自动除霜,同时大洋上的高湿度和高盐分也会对设备造成腐蚀侵袭,此外船体的晃动和震动也要求设备具备抗震能力。因此,在随船航行过程中,检验相机能否完成全自动拍摄的同时,外在环境条件也可以对设备的耐高温、耐低温、防潮、防结霜、防震等功能进行验证。

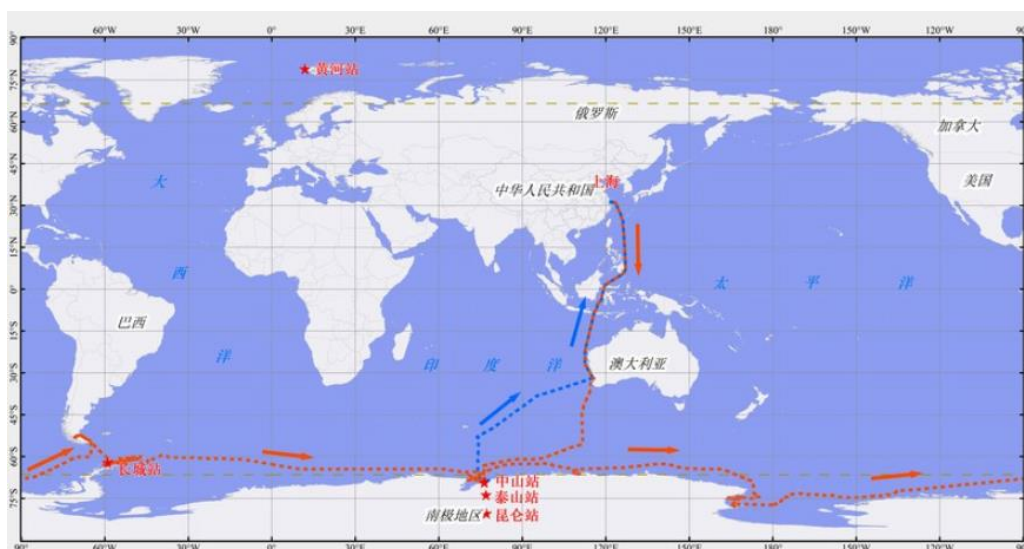
本研究设计的硬件系统包括设备箱体、鱼眼镜头、CMOS相机、工控计算机、风道风扇、电子复位器6个部分。鱼眼镜头和CMOS相机是系统的拍摄模块,用于图像的获取;工控计算机是控制模块,内置第三章中设计的自动化拍摄算法,可实现相机的无人值守;风道风扇是温控模块,主要用于调节硬件系统内部的温度;箱体则可以减轻内部器件所受外部环境的影响。

其中最核心的器件是鱼眼镜头、CMOS相机和工控计算机。镜头决定了拍摄成像的质量,是相机中最重要的部件。其中鱼眼镜头具有超大视场,能够实现全空域包容和全时域实时信息的获取,符合天文观测时对区域云量实时监测的需求。

设备采用的鱼眼镜头型号是富士FE185C086HA-1,如图2-2。相机具体参数为2.7mm定焦、光圈范围为F1.8-F16,可视角度为 $185^{\circ}\times 185^{\circ}$ 。镜头具有分辨率高、成像效果好、拍摄性能稳定等特点,适用于本次开发的全天观测设备。



a) 雪龙号第 8 次北极航行航线



b) 雪龙号第 34 次南极航行航线

图 2-1 雪龙号航行线路图

Fig. 2-1 The sailing route of the XueLong. a)The 8th Arctic sailing route of the XueLong. b) The 34th Antarctic sailing route of the XueLong



图 2-2 富士 FE185C086HA-1 鱼镜头

Fig. 2-2 The Fisheye Lens of Fujifilm FE185C086HA-1

CCD 与 CMOS 是被普遍采用的两种图像传感器，它们都是使用感光二极管将光信号转化为电信号，实现图像数据转换为数字数据的过程，其主要差异在于数字数据传送方式的不同。CMOS 传感器相较于 CCD 传感器具有低能耗、低成本、高度整合等特点，因此，本设备采用了 QHYCCD 公司的 QHY5III174 国产 CMOS 相机，其核心部件为 Sony IMX174 Exmor CMOS 图像传感器，如图 2-3，具有高帧率、低读出噪声、高量子效率的特性，芯片尺寸为 1/1.2inch，有效像素为 1936×1216，满阱电荷数大于 32 Ke。



a) QHY5III174 相机

b) Sony IMX174 Exmor CMOS 图像传感器

图 2-3 QHY5III174 及其采用的 CMOS 图像传感器

Fig. 2-3 QHY5III174 and CMOS image sensor. a) QHY5III174 camera. b) Sony IMX174 Exmor CMOS image sensor

工控计算机主要用于相机的拍摄控制、数据存储和数据传输。工控计算机采用基于 X86 架构的小型化计算单元（NUC，Next Unit of Computing），该计算单元采用了 Intel M 系列低功耗 CPU，拥有 16G 运行内存，512G 固态存储，安装 Windows 2008 R2 操作系统。该计算单元具有功耗低、运行速度快、存储量大、工作性能稳定等特点，符合全天相机系统对工控计算机的性能要求。

器件选择后将箱体划分为上中下三部分空间，整体结构设计如图 2-4 所示，底层容纳工控计算机（含电子复位器）和电源适配器、中层容纳 CMOS 相机和风道风扇、上层容纳干燥盒及鱼镜头。当设备运行时，工控计算机会产生大量热量，设备中层两端的风扇通过一个从下往上吹，另外一个从上往下吹的方式，将底层工控机产生的热量引导至机体上层，实现设备内部热量的循环流动，并且可以对箱体顶部亚克力圆顶进行加热，达到除霜除雾的目的。

之后对箱体进行设计，设备箱体是容纳全天相机系统的容器，箱体也是由三部分组成，最上方是亚克力半球形透明薄壳，亚克力材料具有透明度高、稳定性好、耐高温、耐腐蚀等特点，很适合做镜头的保护罩；全天相机外壳材料采用铝合金，具备良好的耐腐蚀性，并且相较于不锈钢材料，铝合金更加轻便；全天相机后盖也使用铝合金材料加工制造。三部分均通过螺母进行连接。工程图见附录 A。

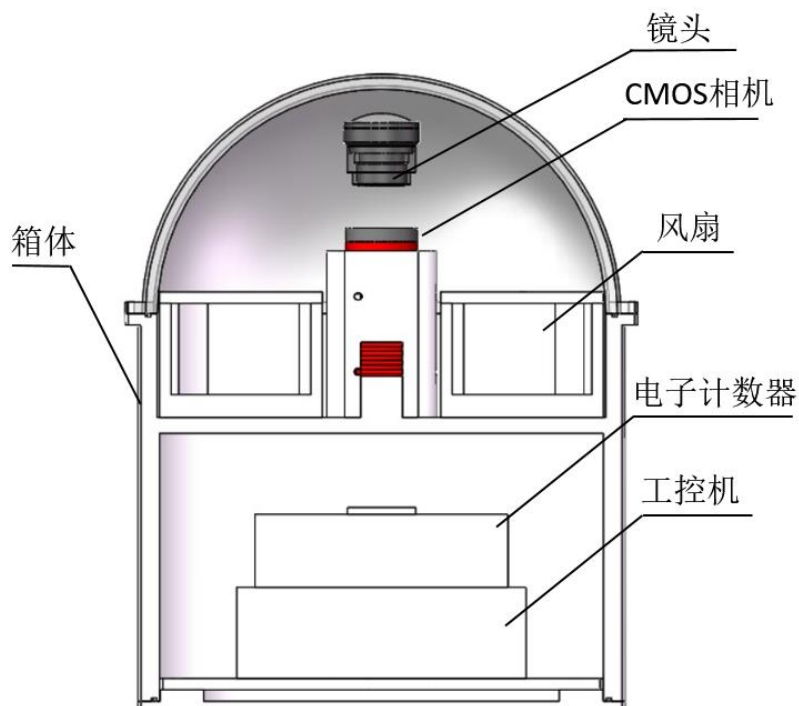


图 2-4 设备整体结构图

Fig. 2-4 Equipment design

2.3 全天相机系统硬件设备集成

本节对硬件设备的集成方式做较为详细的阐述。箱体底层的工控计算机（含电子复位器）和与其对应的电源适配器通过压板固定在箱体最底层，压板之间采用螺钉进行固定。中层的 CMOS 相机和风扇同样使用螺钉进行固定，鱼眼镜头和 CMOS 相机使用相机转接环进行连接，干燥盒通过螺钉固定在箱体中间的板上。最上方的亚克力圆顶与箱体同样使用螺钉进行连接。设备实物如图 2-5。



图 2-5 全天相机系统实物图

Fig.2-5 All-sky observation equipment

箱体上共留有设备开关、网络接口、电源接口、顶部开口和底部开口 5 个开放处，设备开关、网络接口、电源接口均采用了专业级别的防水元件；顶部开口在安装亚克力圆顶时，使用了密封胶条进行防水处理；底部开口也使用密封胶完全密封；在设备安装部署时，网络 and 电源接口均采用密封胶加固密封。防水部分如图 2-6。



a) 防水胶圈（红色部分） b) 防水电源接口（红色）及网线接口（黄色）

图 2-6 防水部件

Fig.2-6 Waterproof parts. a) Waterproof rubber ring (red part). b) Waterproof power interface (red) and network cable interface (yellow)

2.4 全天相机系统可靠性验证试验

在设备实际投入使用前，对设备低温耐受性能、高温耐受性能、除露除霜性能以及成像防抖性能进行检测，确保设备能够在雪龙号上正常稳定的工作。在试验完成后将设备安装在雪龙号上，随船航行并获取全天空地基云图。

2.4.1 低温耐受性能

设备采用的器件均具备良好的低温耐受性能，相机、CMOS、工控计算机均能在-40 摄氏度的环境下正常工作，工控计算机运行时产生的热量，也将通过风道风扇对整个设备起到一定的升温作用。在低温耐受测试中，将冷柜温度设定在-30 摄氏度，设备在其中运行 24 小时，记录其测光、拍摄、数据读取的性能（如图 2-7），实验过程中设备性能稳定，拍摄过程正常，证明了设备具备良好的低温耐受性能。

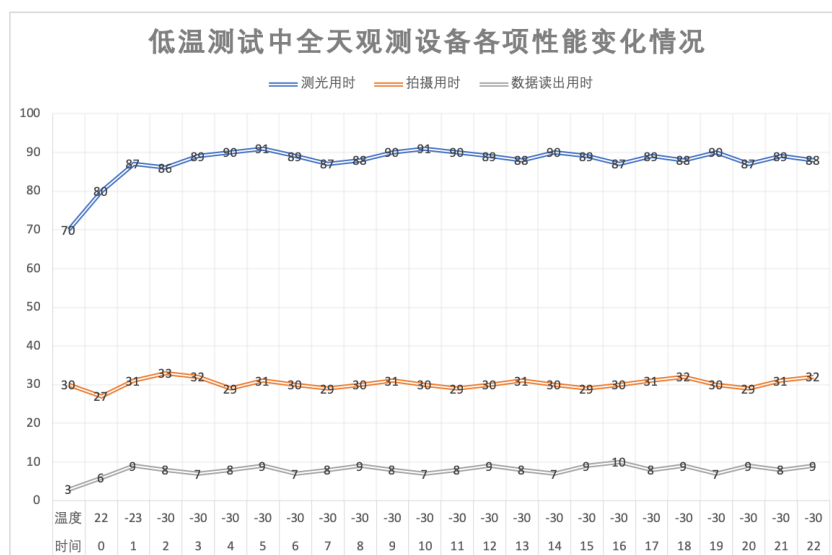


图 2-7 低温测试中全天相机系统各项性能变化情况

Fig. 2-7 Performance changes of all-sky observation equipment in low temperature environment

2.4.2 高温耐受性能

雪龙号在热带航行时，环境温度可以达到 40 摄氏度以上，极端环境会达到 60 摄氏度。基于合适的器件选型和良好的热交换设计，设备本身产生的热量不容易发生积累，从而使设备温度不会大幅高于环境温度。在高温环境测试中，将烘干箱的温度设置在 80 度，设备在其中运行 24 小时，记录其测光、拍摄、数据读取的性能（如图 2-8），实验过程中设备性能稳定，拍摄过程正常，证明了设备具备良好的高温耐受性能。

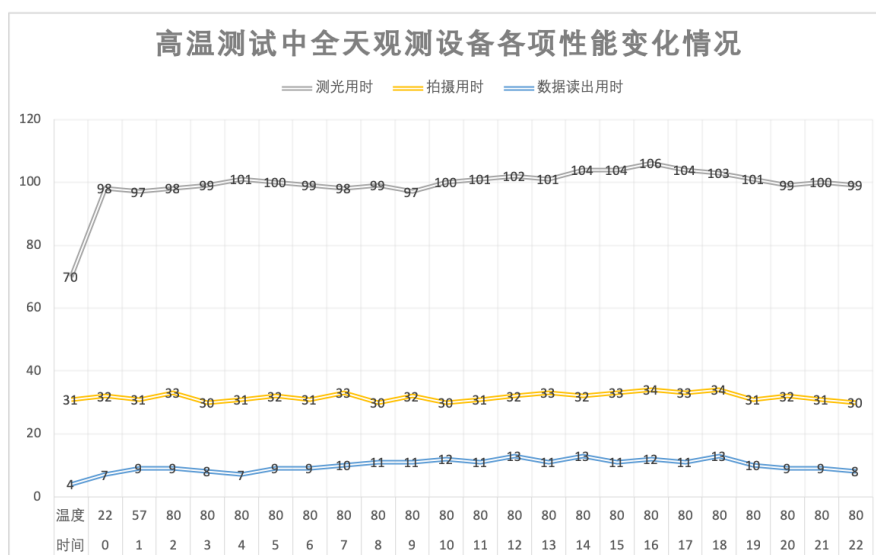


图 2-8 高温测试中全天相机系统各项性能变化情况

Fig. 2-8 Performance changes of all-sky observation equipment in high temperature environment

2.4.3 除雾除霜性能

箱体中的风道风扇能够将工控计算机产生的热量引导至亚克力圆顶处,对其进行持续加热,避免结雾。在极地环境中,圆顶结霜难以避免,结霜图像如图 2-9,通过该加热方式能够加速霜雾消散的过程,但主要除霜方式仍是阳光照射和自然空气流动。



图 2-9 结霜时拍摄的图片

Fig.2-9 Picture taken when the lens is frosted

2.4.4 成像防抖性能

由于全天相机系统的主要观测对象为地基云,拍摄时光线往往较为充足,所以不需要长时间曝光成像。船体的震动和摆动对成像质量的影响,可以通过降低曝光时间来消除。设备选取的 CMOS 最短曝光时间为 20ms,能够抵消震动带来的成像影响。

2.5 本章小结

本章首先介绍了雪龙号在极地科考过程中所经历的恶劣自然环境能为全天相机系统提供良好的验证效果。之后对设备整体硬件结构设计、相关硬件设备的选型、箱体设计和硬件集成做了详细介绍。设备可靠性试验,从高温耐受性能、低温耐受性能、防抖性能以及自动除雾除霜功能等方面进行分析和测试,确保设备在恶劣环境中可以正常稳定的工作。最后将设备安装在雪龙号甲板上,随船航行期间将对设备性能进行真实环境中的验证。

第3章 全天相机系统无人值守软件设计

3.1 引言

无人值守软件系统是全天相机系统的大脑，是实现系统自动化拍摄的基础，它与硬件系统相辅相成，共同实现图像数据拍摄的自动化。由于全天相机系统安放在雪龙号的科考平台上，位于船体的最顶层，在整个航行期间运行是无人值守的。因此，设备需要有自动测光、自动拍摄、自动存储、自动数据传输、自动故障恢复的能力。为此本文提出了全天相机系统的无人值守软件体系，包括观测组件、存储组件和维保组件，设计了合适的内部算法，并通过 C#语言进行实现。最终将软件部署在全天相机系统的工控机上，实现全天相机系统的自动化拍摄。

3.2 全天相机系统无人值守软件设计方案

自动测光、自动拍摄、自动存储、自动数据传输和自动故障恢复等能力是全天相机系统实现无人值守的关键。自动测光是指相机能够自动测出当前天空的亮度值，并根据亮度值调整曝光参数，从而拍摄得到最佳质量的图像数据；自动拍摄是指设备能够按照一定的频率自动曝光获取观测数据；自动数据存储和传输是指设备可以按照一定的策略完成观测数据的存储过程，并将其传输至雪龙号船载服务器上进行备份；自动故障恢复是指当设备遇到非不可逆性故障时（主要是软件故障）能够自动恢复至正常状态。为实现上述功能，本章将全天相机系统的无人值守软件体系分为观测组件、存储组件、维保组件三个部分，各部分主要功能如图 3-1 所示。

3.2.1 观测组件原理及曝光参数计算算法

观测组件负责控制 CMOS 进行样张拍摄，之后通过对样张图像进行计算分析从而得到曝光参数，最后控制 CMOS 根据得到的曝光参数进行拍摄。观测组件由自动曝光参数计算服务和拍摄服务组成，如图 3-2，其核心是自动曝光参数计算服务。

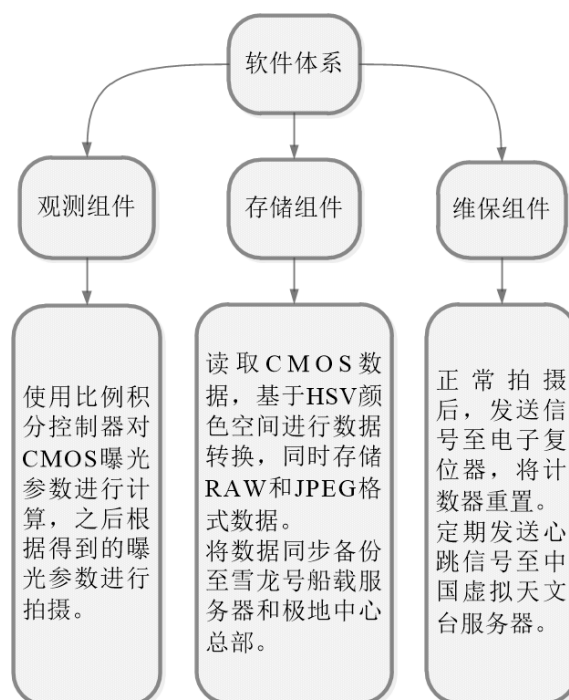


图 3-1 全天相机系统软件体系

Fig. 3-1 All-sky observation equipment software system

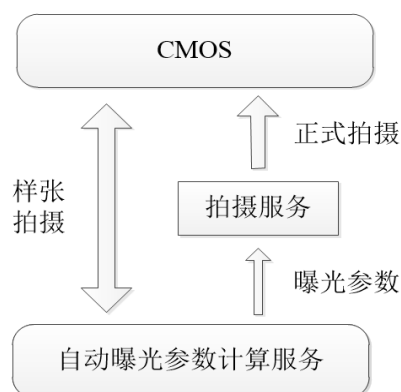


图 3-2 观测组件结构

Fig. 3-2 Observation component structure

全天相机系统在进行拍摄时，对获取图像质量有影响的主要因素有三个，分别是曝光时间、镜头光圈值和增益，这三个参数统称为曝光参数。其中曝光时间是指 CMOS 感光的时长，以毫秒为单位；镜头光圈值是指镜头通光量的大小，光圈的大小将决定光线穿过镜头的强弱，用 f 值表示，常用的光圈有 $f1.4$ 、 $f4.0$ 、 $f11$ 、 $f16$ 等， f 后的数值和光圈孔径的大小相反， f 后数值越大，光圈的孔径越小，进光量越少，反之则孔径越大，进光量越多；增益是指 CMOS 的光敏感程度，也就是在相机拍摄时通常指的 ISO 值，增益值越高，对光越敏感，但是噪点会随之增加，为拍摄质量高的图像，通常将其设置为一个较低的值。曝光参数关系如

图 3-3 所示。

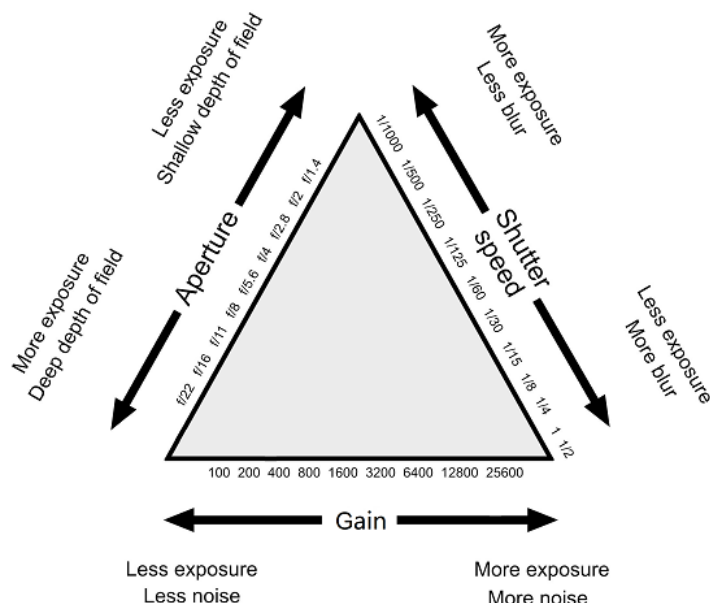


图 3-3 曝光参数三角

Fig. 3-3 Exposure parameter triangle

确定曝光参数的主要方式是通过测光值，常用的测光方式包括使用光度计测光和通过图像计算测光，由于使用光度计需要额外添加硬件设备，为此我们通过拍摄测光图像进行计算测光。调整曝光参数的目标是使最终拍摄的图像不会过曝或过暗，一般来说，图像达到一个中间灰度值即可，即 18% 灰度，其对应的平均 RGB 值为 118、118、118。常用的图像测光方法是在正式拍摄前，采用不同的预设曝光参数多重曝光得到多张测光图像，再计算图像中央 60%-80% 的亮度。这种方法适用于一般拍摄，但在极端光照环境，如户外阳光直射和黑夜环境，这种方法效果不佳。对于全天相机系统，通过鱼眼镜头在 CMOS 上的成像是一个圆形区域，其外为黑色无数据区域，如果直接对测光图片进行测光，会导致测光值偏低，进而致使曝光参数错误，使得最终拍摄图像过曝光。

为此，我们引入了模板法进行测光修正，模板法可以在测光时剔除无关区域的亮度影响，使我们感兴趣的区域（天空）得到最佳的曝光度。模板法是将测光图像中不适合进行测光计算的部分用模板减掉，比如鱼眼镜头产生的黑色区域、设备视野范围内的建筑物以及太阳在天空中的位置。黑色区域和视野内建筑是已知的，太阳位置可以通过时间和纬度计算出来，从而生成动态模板。

我们采用图像亮度直方图来进行测光图像曝光度的判断。把图像的亮度分为 0 到 255 共 256 个数值，亮度越高，值越大，最暗的纯黑色区域用 0 代表，最亮的纯白色区域用 255 代表，中间的数字则是代表不同亮度的灰色。采用亮度直方图可以很好的表现一张图像的曝光程度，将直方图的亮度值划分为 5 档，分别表

示 CMOS 对不同光照强度的反应，过曝光的图片其直方图将会向右倾斜，而曝光不足的图片其直方图将向左倾斜。所以，理想影像的直方图应主要分布在中间部分，由于鱼镜头在 CMOS 上的成像是一个圆形区域，其外为黑色无数据区域，所以在直方图最左侧均出现像素堆积，如图 3-4。

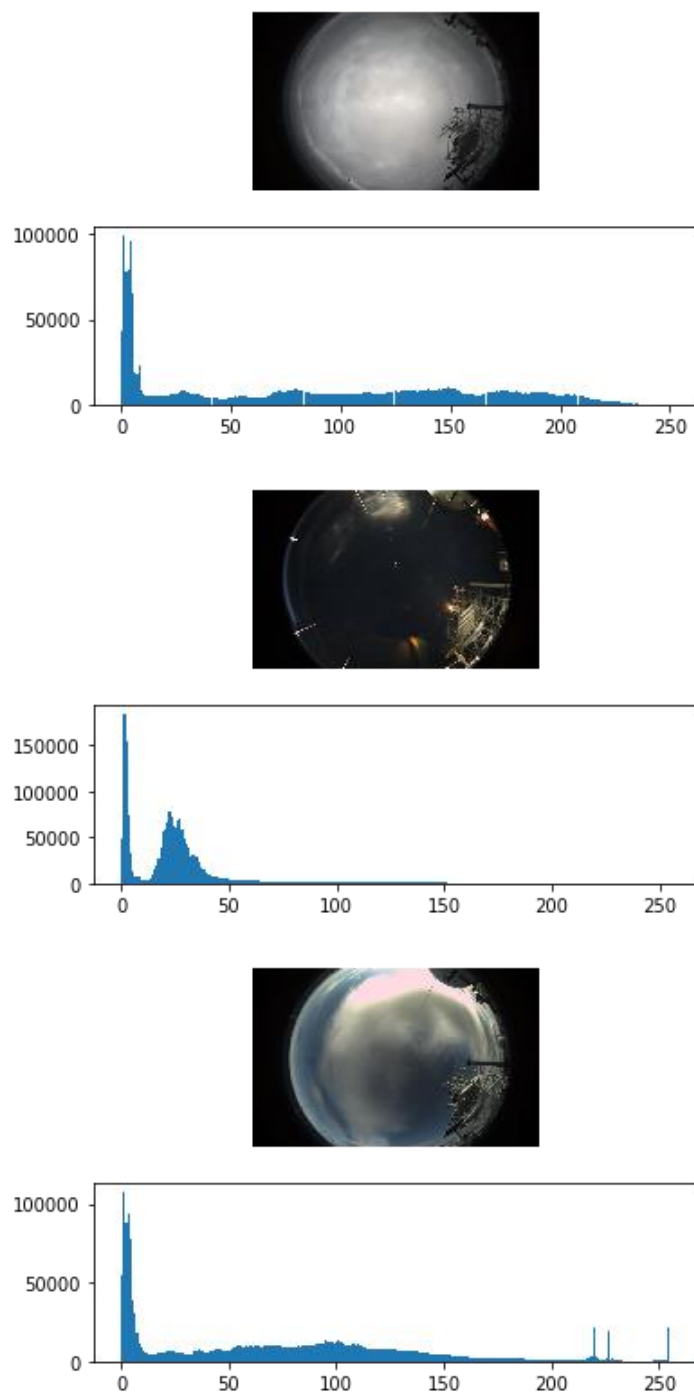


图 3-4 全天相机在不同时间拍摄的图片及其直方图

Fig. 3-4 Pictures and histograms taken at different times by All-sky camera

设备中镜头的光圈固定在 $f11$ ，所以可调的曝光参数分别是增益值和曝光时间。采用比例-积分控制器（PI Controller）来分别控制这两个参数，在一定时间

内不断进行调整得到合适的增益和曝光参数，使得图像的平均亮度值达到相应的预设值。PI Controller 是一个反馈回路控制方法，它将控制反馈得到的值和参考值进行比较，并将其差值用于计算新的输入值，通过不断迭代，使反馈值和参考值尽可能达到一致。它适用于线性且动态特性不随时间变化的系统，而曝光参数的控制符合上述条件。PI Controller 的比例误差单元(P)、积分误差单元(I)分别对应目前误差、过去累计误差，从而使得系统的调整，即达到预期平均亮度值的过程，可以用控制器对误差的反应快慢、控制器过冲的程度来表示，其计算公式如下：

$$e(t) = SP - PV \quad (3-1)$$

$$u(t) = K_c e(t) + \frac{K_c}{\tau_i} \int_0^t e(t) dt \quad (3-2)$$

式中 SP ——图像平均亮度值的期望值；

PV ——图像平均亮度值的测量值；

$e(t)$ ——两者的误差；

K_c ——控制增益 (controller gain)；

τ_i ——积分时间常量。

K_c 作为乘数作用于控制器的比例误差单元和积分误差单元，其大小控制各误差单元的增幅。 $u(t)$ 为反馈值，即对应曝光时间和增益值。

对于 K_c 的选取，需要在实践调节中得到最佳值，对于设备采用的 CCD 和镜头来说，所选取的 K_c 分别是：

$$K_c(Exposure) = 25 \cdot c + 20 \quad (3-3)$$

$$K_c(Gain) = 10 \cdot c + 15 \quad (3-4)$$

式中 c ——减去模板后的图像（感兴趣的图像）占总图像的像素比例；

在实践中，我们对增益值和曝光时间分别应用比例积分控制。由于增益值越小图像中的噪点越少，所以首先将增益值设为 1，对曝光时间进行比例积分控制计算，并将反馈得到的曝光时间用于测光拍摄，计算测光图像的平均亮度值，并将其和参考值进行比较得到误差，迭代该过程，直至测光图像平均亮度值达到预计值。当曝光时间达到最大值，平均亮度值仍未达到预计时，则固定曝光时间，继续对增益值进行比例积分控制计算，直到测光图像平均亮度值达到预计值。

3.2.2 存储组件设计方案

设备完成拍摄后，工控计算机通过 CMOS 的固件接口将数据读入到计算机内存。读出的原始数据为灰度图像，为了使之更好的反映出全天景象的特征要素，需要对其进行基于 HSV 颜色空间的转换，经过转换后，数据保存为 RAW 格式

并写入固态存储。**RAW** 格式是一种无损图像存储格式，能够完整记录图像场景的光照强度和颜色等物理信息，对于后期的数据处理非常有利，缺点是占用的物理空间较大，每幅图像占用空间约 6.49MB。在存储为 **RAW** 格式的同时，也会把数据压缩成 **JPEG** 格式，并将其尺寸压缩为 625X1000 作为快视图，每幅图像占用空间约 40KB。数据存储的文件组织形式是以拍摄时间构成的树状结构，其根目录为拍摄时间的年份，之后的文件名依次是拍摄时的月份、日期和协调世界时（Coordinated Universal Time, UTC）。

存储过程中通过采取多重备份策略来保证全天相机系统的数据完整性。在工控计算机中部署的备份服务，会使全天相机系统在完成拍摄后，自动把生成的 **RAW** 格式数据和压缩过后的 **JPEG** 格式数据复制存储在工控机内的另一块硬盘上；此外，在雪龙号机房中还部署了一套备份服务器，工控计算机的备份服务通过 **FTP**（File Transfer Protocol）传输协议将这些数据同步发送至备份服务器中；备份服务器的网络与极地中心总部的机房连接，会定时将一部分数据传输到极地中心总部进行备份。备份策略如图 3-5 所示。

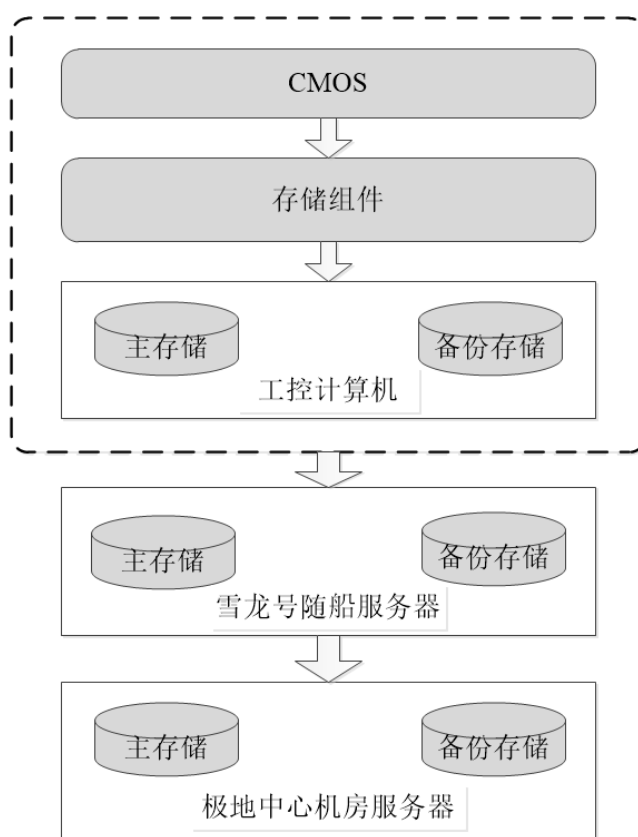


图 3-5 多重备份策略的数据流向

Fig. 3-5 Data flow of multiple backup strategies

3.2.3 维保组件设计方案

维保组件主要由电子计数器复位功能和心跳信号程序组成。在全天相机系统的观测服务中，包含了一条计数器复位指令，在每次正常拍摄后，观测服务会发送该指令到设备上的电子计数器，将计数器重置。而如果未能正常拍摄，说明拍摄服务出现软件故障，此时则不会发送复位指令，当电子计数器达到一定计数后，会自动启动工控计算机的重启脚本，使运行在工控计算机上的所有程序重启，从而达到恢复软件服务的目的。

心跳信号是每隔一段时间向互联的另一方发送一个很小的数据包，通过对方回复情况判断互联双方之间的通讯链路是否已经断开的方法。设备中的心跳信号程序会定期发送心跳信号至中国虚拟天文台的服务器，该心跳信号为一条 FTP 命令，其内容是创建一个以发送命令的 UTC 时间为文件名的文件夹，该命令每 5 分钟发送一次（与拍摄时间间隔一致）。陆上值班人员会定期查看命令所创建的文件夹，若设备出现故障且未能实现自动恢复，则最后一个文件夹创建的时间会与当前 UTC 时间间隔较远，此时说明设备存在不可逆故障（如电源故障、网络故障等），陆上值班人员会及时通知随船工程师进行故障排查。

3.2.4 系统整体设计方案

无人值守软件系统流程图如图 3-6，流程图中主要包含观测组件和储存组件，可实现全天相机系统拍摄和存储的自动化。维保组件相对独立于其他两个组件且逻辑实现简单，所以论文中不做过多描述。

3.3 全天相机系统无人值守软件实现过程

本节主要对观测组件和存储组件的实现过程做较为详细的描述。

3.3.1 观测组件实现

相机购买时提供 `cameraController` 类，可通过 `getCameraController()` 函数对相机进行控制，其中最主要的函数是 `SetCameraParameters()`，可通过它设置拍摄参数。

`SetCameraParameters(double exposureTime, double gain)` 函数：

参数说明：

`exposureTime`——曝光时间。

`gain`——增益值。

编写 AllSkySurveyService 类,用于检查相机状态和控制相机的拍摄。OnStart() 函数用于连接相机并循环执行拍摄命令; ChkSrv()函数用于判断工控计算机是否和相机连接成功; OnStop()函数用于断开与相机的连接。ExposureComputer 类则是观测组件的核心,主要函数为自动曝光参数计算函数 ComputeExposureTime(), 主要功能是实现 3.2.1 节中提到的比例-积分控制算法; DoComputeTest()函数则是计算图像亮度值并验证其是否符合预期要求。

ComputeExposureTime(List<Point> parray)函数:

参数说明: List<Point> parray——图像亮度值列表。

ExposureComputer 类部分定义如下:

```
class ExposureComputer
{
    double minDurationTime;
    double maxDurationTime;
    double targetBrightness;
    .....
```

ComputeExposureTime()函数部分代码如下:

```
public double ComputeExposureTime(List<Point> parray)
{
    if (parray.Count < 2)
    {
        Console.WriteLine("点的数量小于 2, 无法进行线性回归");
        return 0;
    }
    .....
```

DoComputeTest ()函数部分代码如下:

```
public void DoComputeTest()
{
    CameraController cameraController = new CameraController();
    cameraController.Connection();
    List<Point> array = new List<Point>();
    double brightness = 0;
    double exposure_time = 0;
    .....
```

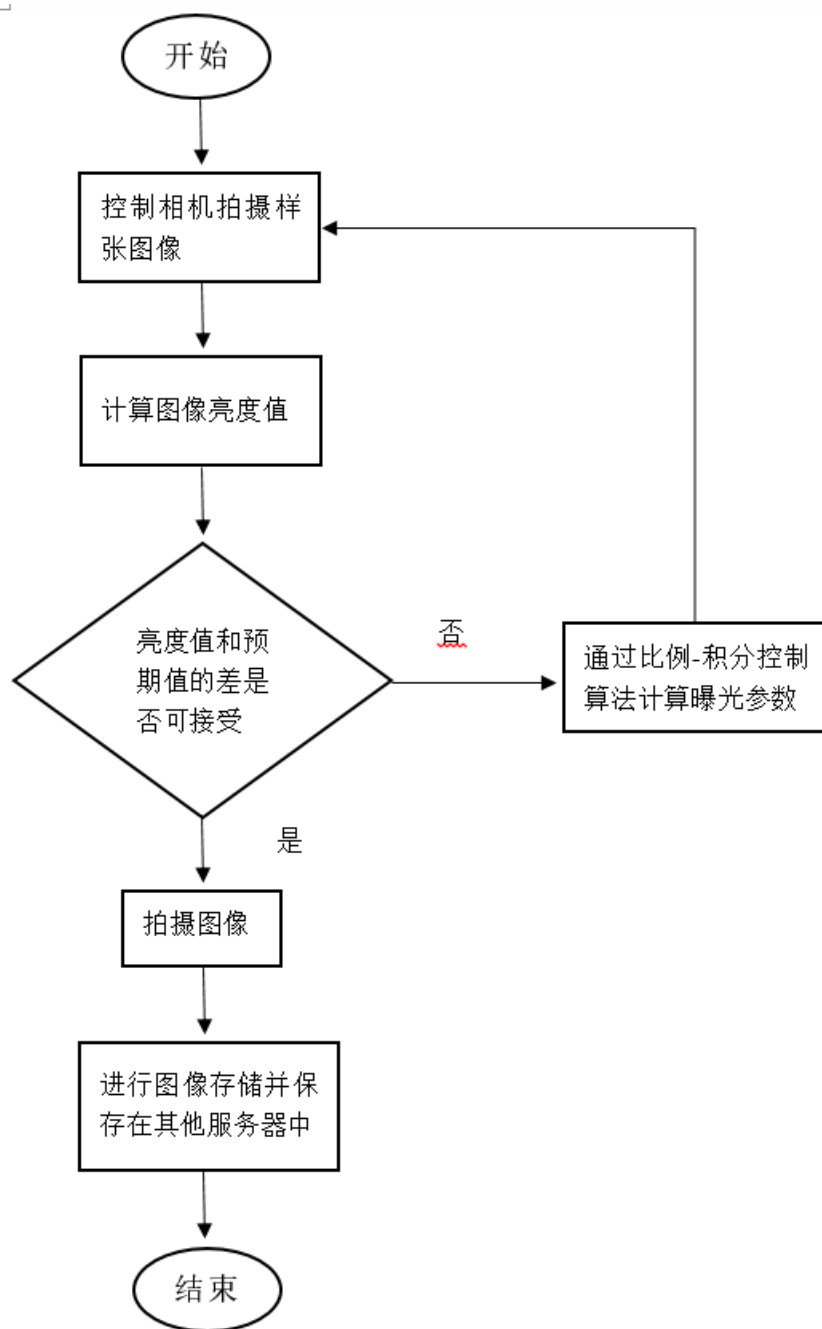


图 3-6 无人值守软件系统流程图

Fig. 3-6 Unattended software system flow chart

3.3.2 存储组件实现

首先编写 ASCDataTransmissionService 类，包含 OnStart()函数、OnCreated()函数、IsFileReady()函数和 OnStop()函数。OnStart()实现服务的开启，默认路径为“C:\\ASCData”；OnCreated()创建存储路径；IsFileReady()进行循环判断文件是否需要传输，如果需要则调用传输命令；OnStop()函数实现服务关闭，一般不进行

调用。

之后完成编写 `FtpController` 类，通过 FTP 传输服务来实现文件或文件夹的创建、删除、重命名、上传、下载等功能。包括 `FtpController()`, `download()`, `upload()`, `delete()`, `rename()`, `createDirectory()`, `getFileCreatedDateTime()`。 `FtpController()` 用于实现对 FTP 服务器的连接； `download()` 实现从服务器上下载文件的功能； `upload()` 实现文件上传到服务器的功能； `delete()` 用于删除文件； `rename()` 实现文件重命名； `createDirectory()` 可创建文件夹； `getFileCreatedDateTime()` 则是用于获取创建文件时的 UTC 时间信息。

`FtpController(string hostIP, string userName, string password)` 函数：

参数说明：

`hostIP`——FTP 服务器 IP 地址。

`userName`——FTP 服务器用户名。

`password`——FTP 服务器密码。

`download(string remoteFile, string localFile)` 函数：

参数说明：

`remoteFile`——服务器上文件的文件名。

`localFile`——本地文件名。

`upload (string localFile, string remoteFile)` 函数：

参数说明：

`remoteFile`——服务器上文件的文件名。

`localFile`——本地文件名。

`delete(string deleteFile)` 函数：

参数说明： `deleteFile`——要删除文件的文件名。

`rename(string currentFileNameAndPath, string newFileName)` 函数：

参数说明：

`currentFileNameAndPath`——需修改名称文件的文件路径。

`newFileName`——修改后文件名。

`createDirectory(string newDirectory)` 函数：

参数说明： `newDirectory`——需创建文件夹名。

`getFileCreatedDateTime(string fileName)` 函数：

参数说明： `fileName`——文件名。

图像格式转换通过相机自带函数库实现，文章不做具体描述。

3.4 软件工具与软件环境

软件工具与软件环境的选择是软件编写前的一个重要步骤,关系到功能具体实现方式、基本函数库调用方法、执行时间和最终软件的应用环境。本文采用微软公司的 Visual Studio 进行软件开发,其中包含了 Visual C++、Visual C#、Visual Basic 等多种开发语言。Visual C#多用于软件开发,具有简单、功能强大、类型安全且面向对象等优点,所以我们使用 Visual C#进行软件系统的编写。

3.5 本章小结

本章介绍了全天相机系统无人值守软件程序的整体结构,包括观测组件、储存组件以及维保组件,各部分组件协同运行,确保了全天相机系统的自动化拍摄和数据的安全存储。对使用到的测光方法、曝光参数计算算法、存储方式、多重备份策略以及当遇到不可预测的问题时设备自动恢复拍摄和对运行状态进行及时反馈等功能做了详细描述,完成了系统整体的实现方案。之后对观测组件和储存组件的具体实现方法做了详细描述。最后介绍了软件工具和软件环境。

第4章 全天空地基云图数据处理

4.1 引言

云图数据的后处理不仅可以帮助天文学家对观测环境进行实时云量分析,也是实现望远镜智能化观测的基础。但现有的全天观测系统大都不配备相应的云量分析算法。本文基于此次拍摄的全天空地基云图数据,使用深度学习技术开发了一种适应性强、鲁棒性高的图像分类算法,可将云图根据其应用场景进行自动分类。之后使用少云类全天空地基云图数据对几种常用的云点识别算法进行了实验验证和对比分析。通过上述方法,实现了云图数据的自动化分类及部分云图的云量计算和分析。

4.2 全天空地基云图数据介绍

雪龙号于 2017 年 7 月 20 日至 2017 年 10 月 9 日开展了第 8 次北极科考航行、2017 年 11 月 8 日至 2018 年 4 月 21 日完成了第 34 次南极科考航行。在此期间,我们设计研发的全天相机系统出色的完成了数据获取的任务,北极航行期间获取 13484 张全天影像、南极航行期间获取约 53000 张全天影像。航行沿途经过区域广、环境因素复杂,拍摄时的时间、天气、光照等因素也在不断变化。因此,雪龙号全天设备获取的全天云图像数据具有类型丰富、覆盖性强等特点。部分数据如图 4-1 所示。

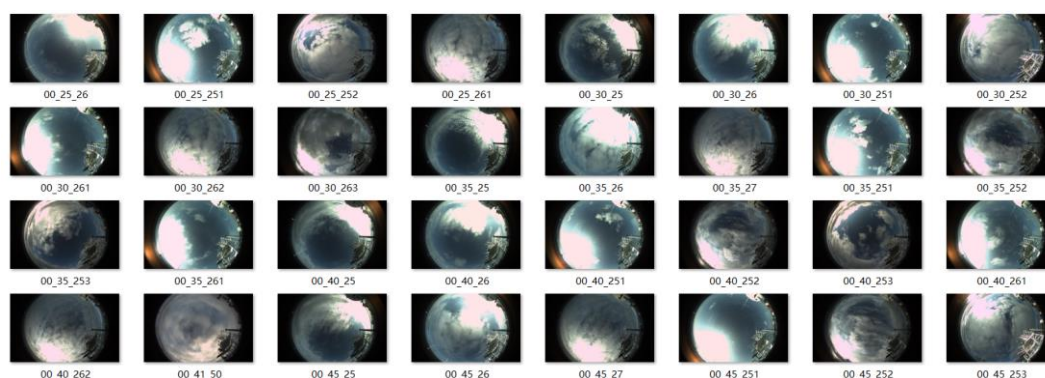


图 4-1 部分图像数据

Fig. 4-1 Partial image data

4.3 全天空地基云图数据分类

地基云图在获取过程，难免会受到时间、天气、光照的影响，获取图像的内容、质量也不尽相同。在进行面向天文选址或者实现望远镜自动化观测的云量检测工作时，针对不同时段（如白天与黑夜）的全天图像，云量检测方法也不同。因此，在进行云量检测之前，去除不可靠图像（雨雪覆盖镜头、过曝光），区分图像场景（白天或夜晚）是云量分析工作的基础。

4.3.1 分类标准

根据应用场景，对数据进行分析，最终将云图划分为夜晚（dark）、雨雪（rain_snow）、过曝光（overlight）、晴空（bluesky）、少云（cloudimage）、多云（overcast）6类。

昼夜会给图像带来不同的特性，研究的内容和方法也往往不同。对于白天拍摄的全天空地基云图，往往通过图像可见光波段的性质来确定有云点^[13]；夜间则将测得的星场图像与计算的星象图作比较，确定哪些已知的明亮恒星被云遮盖，从而识别云^[14]。因此，昼夜图像分类将便于之后的科学研究，我们首先将夜晚拍摄的图像划分为一类，如图 4-2 所示。



图 4-2 夜晚类典型图像

Fig. 4-2 Typical image of dark

对白天拍摄的图像进行更为细致的划分。拍摄过程中，镜头往往会受到恶劣天气的影响，雨雪覆盖镜头时会严重影响成像的质量，使得这部分图像很难应用于之后的云量检测工作。因此，将受雨雪影响的图像划分为单独一类。雨雪类图像如图 4-3。



图 4-3 雨雪类典型图像

Fig. 4-3 Typical image of rain_snow

正午光线充足，在强光照射下拍摄图片容易发生曝光的现象，将这类数据划分为单独一类，可区分出更有研究价值的全天空云图数据。当曝光面积大于全天图像有效面积的三分之一时，认为该图片属于过曝光类。过曝光类典型图像如图 4-4 所示。



图 4-4 过曝光类典型图像

Fig. 4-4 Typical image of overlight

在去除掉质量较差的图像后，根据云量的多少将剩余图像划分为三类，分别是晴空类、多云类和少云类。晴空类不含云，整个天空几乎完全被云覆盖定义为多云类，其他则均认为是少云类。部分图像在太阳所在位置会发生过曝的现象，出现的白色区域较难与云区分，此时根据数据拍摄时间与区域形状确定该区域是过曝过还是云。各类别典型代表图像如下，图 4-5 为晴空类，图 4-6 为少云类，图 4-7 为多云类。



图 4-5 晴空类典型图像

Fig. 4-5 Typical image of bluesky

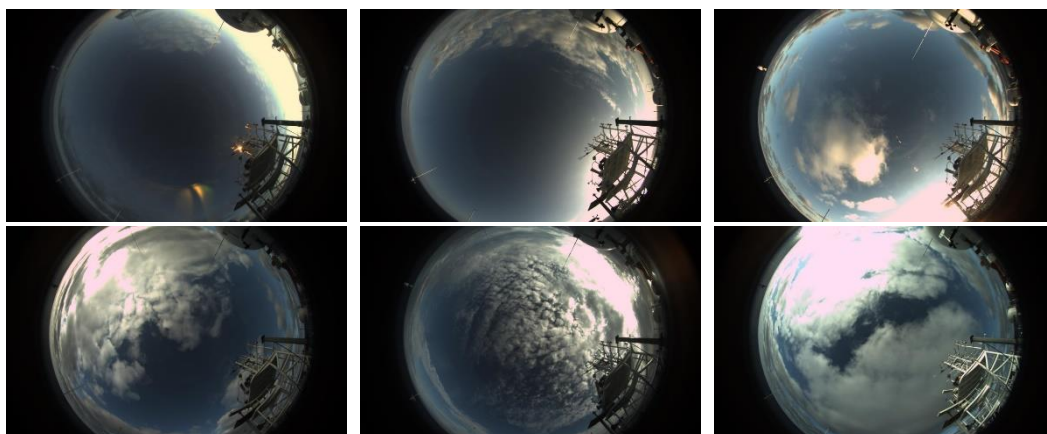


图 4-6 少云类典型图像

Fig. 4-6 Typical image of cloudimage



图 4-7 多云类典型图像

Fig. 4-7 Typical image of overcast

4.3.2 分类方法及理论

图像分类任务要求识别一张图像属于哪一类别或主要包含哪一类事物。一般来说，物体分类算法通过手工特征，如局部特征包括尺度不变特征转换 (Scale Invariant Feature Transform, SIFT)^[15]、方向梯度直方图(Histogram of Oriented Gradient, HOG)^[16]、局部二值模式 (Local Binary Pattern, LBP)^[17]等或者特征学习方法对整个图像进行全局描述，然后使用分类器，如 K-最近邻 (k-NearestNeighbor, KNN)^[18]、支持向量机 (Support Vector Machine, SVM)^[19]等进行判断。近些年来基于机器学习的特征学习表达方式成为复杂图像分类任务中特征提取的主流方法，代表型的算法包括有词袋模型(Bag of Words)、深度学习模型等^[20]。

选取合适的分类方法是数据分类过程中的关键问题。深度学习^[21]的概念于 2006 年由 Geoffrey Hinton 等人提出，相对于传统机器学习的方法，深度学习可以让机器自主学习高维的数据特征^[22]，并在大数据的支撑下得到泛化的数据模型，减轻了繁琐的人工寻找特征的步骤，所以近年来发展十分迅速。仿造生物视觉机制^[23]构建的卷积神经网络 (Convolutional Neural Network, CNN) 作为深度学习中一种重要的框架在图像识别领域拥有着巨大的优势，它可以直接使用图像的像素点作为输入，通过训练来提取最有效的特征，并且对缩放、平移、旋转等畸变具有不变性，有着很好的泛化效果。而卷积^[24]的权值共享结构，可以大幅减少神经网络的参数量^[25]，在防止过拟合的同时降低了神经网络模型的复杂度。

CNN 具有多层网络结构，用于图像分类时其基本结构主要包括卷积层、池化层和全连接层^[26]。卷积层 (convolutional layers) 是 CNN 的核心，主要用于图像的特征提取。卷积层通过使用多个特定尺寸的卷积核与上一层的输入图像进行卷积运算，得到多张特征图。卷积运算即是让卷积核沿着输入图片的坐标横向或纵向滑动，与相对应的数据进行卷积运算，随着卷积核的滑动，可以得到一个新的二维特征激活图，激活图的值为卷积核在输入图片不同空间位置的响应^[27]，卷积操作如图 4-8 所示。

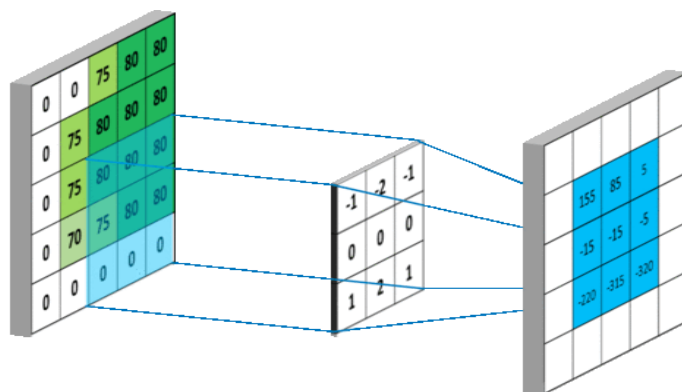


图 4-8 卷积操作示意图

Fig. 4-8 Convolution operation diagram

步长 (stride) 是卷积核滑动的距离, 卷积核的尺寸和步长的大小决定了激活图的尺寸。如输入图片尺寸为 5×5 像素, 卷积核尺寸为 3×3 像素, 当 stride 为 1 时, 每次运算卷积核在纵向或者横向移动 1 个像素, 最终输出图片的尺寸将会是 3×3 , 当 stride=2 时, 输出图片尺寸将会是 2×2 。输出图片尺寸可以通过以下公式得出:

$$y = \frac{x - k}{stride} + 1 \quad (4-1)$$

式中 y ——输出图片尺寸;

x ——输入图片尺寸;

k ——卷积核的尺寸;

$stride$ ——步长。

从式 (4-1) 可以看出, 步长不一定能够整除整张图像的像素宽度, 因此卷积运算的过程被分为两种, 分别记为“VALID”和“SAME”, “VALID”是不越过边缘取样的方式, 输出图像的面积小于输入图像的像素宽度, “SAME”则是越过边缘取样的方式, 输出图像的面积和输入图像的像素宽度保持同等大小^[28]。

在神经元中, 对输入进行加权、求和后, 会被作用一个激活函数 (Activation functions), 或者在卷积层后通常也会添加激活函数。激活函数能将非线性特性引入到网络中, 对人工神经网络模型学习、理解非常复杂的非线性函数来说具有十分重要的作用。常用的激活函数有 Sigmoid^[29]函数、tanh^[30]函数和 ReLU^[31]函数。ReLU 是如今较为常用的激活函数, 数学表达式为:

$$f(x) = \max(0, x) = \begin{cases} 0 & x < 0 \\ x & x > 0 \end{cases} \quad (4-2)$$

相比起 Sigmoid 和 tanh, ReLU 具有如下优点:

- 1) 能使网络在训练过程更加快速的收敛。

2) 有效缓解了梯度消失的问题，另外当输入值为正数时，不存在梯度饱和问题。

3) 预先不经过无监督预训练，整个网络也能表现出较好的能力。

4) 使神经网络具有稀疏表达能力。**ReLU** 不会同时激活所有的神经元。从 **ReLU** 函数的数学表达式可以看出，当输入值为负值，**ReLU** 函数的输出值将会是 0，该神经元不传递该处的梯度，神经元未被激活。在每次计算中，只有输入值为正数的神经元才会被激活，因此，神经网络具有稀疏性。未被激活的神经元减少了大量的计算量，使网络训练过程更加高效且快速。

池化层（**Pooling layers**）一般用于特征图像的下采样，在保留主要特征的同时，减少下一层的参数和计算量，防止模型过拟合。池化方式分为两种：最大池化(**max pooling**) 以及平均池化(**average pooling**)。如图 4-9，采用 2×2 大小池化核进行池化操作的示意图。

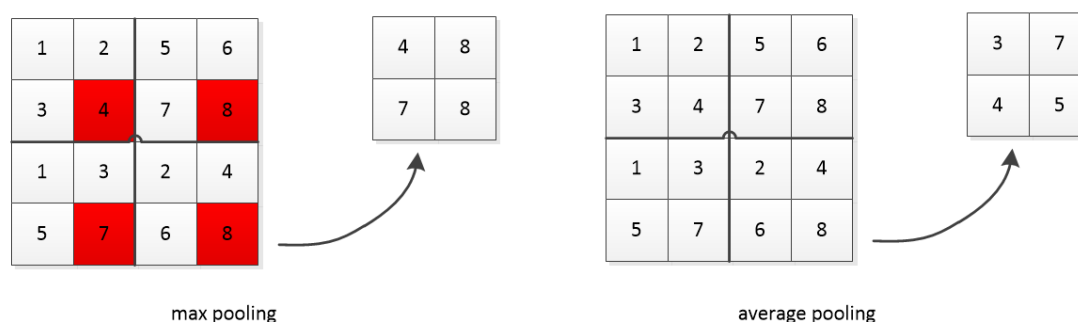


图 4-9 最大池化和平均池化示意图

Fig.4-9 Max pooling and average pooling

全连接层（**fully connected layers, FC**）一般由多个神经元组成，在整个卷积神经网络中起到特征分类的作用。卷积层、池化层和激活函数层等操作完成特征的提取工作，用最合适、少量的特征将图像进行表示，是将原始数据映射到特征空间的过程。全连接层则是将学习到的特征和样本标记空间进行对比，找到它们之间的映射关系，从而完成样本的分类任务。全连接层在实际使用中不仅可以通过神经元的传递方式实现，也可通过卷积操作来实现，对于上一层是卷积层的全连接层可以转化为卷积核为 $h \times w$ 的全局卷积， h 和 w 分别为前层卷积结果的高和宽^[32]。

4.3.3 卷积神经网络模型介绍

AlexNet^[33]由 Hinton 和他的学生 Alex Krizhevsky 共同设计，它将 CNN 的基本原理应用到了很深很宽的网络中，在 2012 年的 ImageNet 图像分类竞赛中

AlexNet 以显著优势获得冠军。本章在 AlexNet 的启发下进行卷积神经网络的设计和搭建。实验中通过不断地测试，最终搭建模型的结构如图 4-10 所示。卷积层数设定为 2，卷积核大小为 5*5，步长为 1。每个卷积层后均连接一个大小为 3*3，步长为 2 的最大池化层，之后连接两个全连接层，神经元个数分别为 384 和 192，Softmax^[34]层在网络的最末端，用于计算各类别输出值的概率值，并根据概率值的大小确定最终的分类结果。每个池化层和全连接层后都添加激活函数 ReLU，CNN 网络详细参数配置如表 4-1。

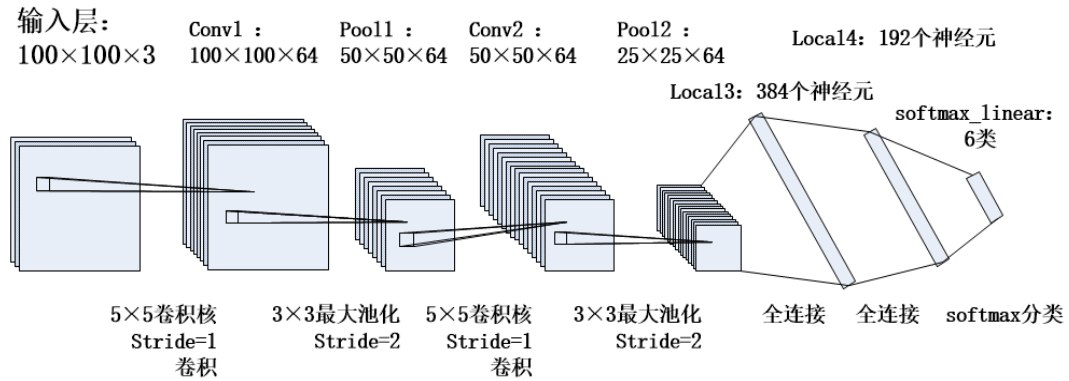


图 4-10 卷积神经网络的结构

Fig.4-10 The structure of a convolutional neural network

表 4-1 CNN 网络详细参数配置

Table 4-1 CNN network detailed parameter configuration

Layer	Input	Kernel	Stride	Output	Parameters
Conv1	3*100*100	5*5	1	64*100*100	64*(5*5*3+1)
Pool1	64*100*100	3*3	2	64*50*50	0
Conv2	64*50*50	5*5	1	64*50*50	64*(5*5*64+1)
Pool2	64*50*50	3*3	2	64*25*25	0
FC3	40000	1*1	1	384	384*(40000+1)
FC4	384	1*1	1	192	192*(384+1)
FC5	192	1*1	1	6	6*(192+1)
Total	—	—	—	—	15542790

4.3.4 数据预处理

在对数据进行分类训练之前，为使数据符合模型的输入要求，并最大限度的包含有效信息，需要进行下列的预处理操作：

(1) 划分训练集和测试集。对全天相机数据进行人工判读和标注，用于模型训练和测试。人工标注数据共 6160 张，其中 overligh (1124 张)、overcast (1128

张)、rain_snow (1117 张)、cloudimage (1115 张)、dark (1119 张)、bluesky (557 张)。随机选取 bluesky500 张, 其余每类 1000 张作为训练集, 余下图像作为测试集, 测试集和训练集没有交叉图像, 测试集数据不用于对模型的训练, 仅用于判断模型对未知图像的分类准确度。

(2) 获取感兴趣区域 (region of interest, ROI)。将 1000×625 大小的图像, 以图像中心为基准进行裁剪, 去掉两端无关信息, 得到 800×625 的输出图像 (如图 4-11)。

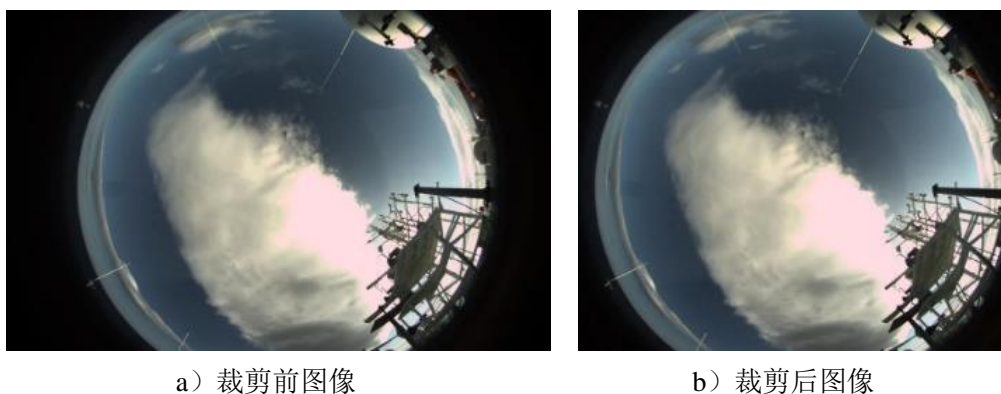


图 4-11 全天图像进行裁剪获取 ROI 区域的过程

Fig.4-11 Crop all-sky images to get the ROI. a) Image before cropping. b) Cropped image

(3) 数据增强 (Data Argumentation)。充足的数据量可以有效防止模型过拟合现象的发生。在数据量较小的情况下, 可以通过数据增强的方法, 在保证数据特征不变的基础上对图像数据进行几何变换, 来增加数据量。主要数据增强操作包含旋转反射变换(Rotation/reflection)、翻转变换(flip)、缩放变换(zoom)、平移变换(shift)、尺度变换(scale)、对比度变换(contrast)、噪声扰动(noise)、颜色变换(color)等。结合数据特征, 本研究对训练集采取如下数据增强的方法: 首先对训练集图像做偏移处理, 分别以右下、右上、中心、左下、左上为坐标端点, 裁剪大小为 700×550 区域; 再通过左右翻转操作, 将数据扩充 10 倍, 整个过程如图 4-12 所示, 第一行分别是以图像右下, 右上, 中心, 左下, 左上端点为坐标原点裁剪后的图像, 第二行是对第一行图像进行左右翻转操作后得到的图像。对数量较少的 bluesky 类, 在经过上述操作过程后将所有图像再进行上下翻转, 最终得到每类 10000 张图像, 从而确保训练数据各种类样本在数量上保持均衡。

(4) 将测试集图像以图像中心为坐标原点, 裁剪大小为 700×550 的区域, 保证训练集图像和测试集图像在输入模型时具有相同的尺度和图像信息, 提高测试准确率。

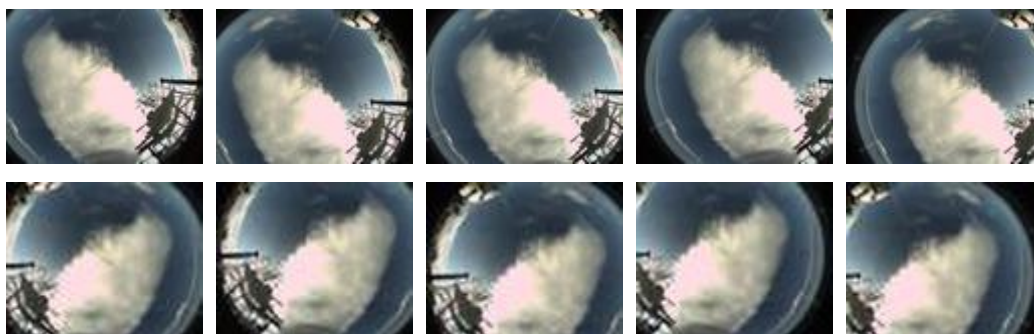


图 4-12 数据增强操作示例

Fig.4-12 Data augmentation example

(5) 使用 `python` 的第三方库 `cv2` 中的 `resize()` 函数将所有图像均缩放为 128×128 ，通过减小图像尺度，来减轻模型的训练负担，降低模型对硬件设备计算能力的需求。

(6) 读取训练集图像数据信息，并在训练之前对数据进行随机截取，大小为 100×100 ，数据量相当于扩大了 $(128-100)^2$ 倍，之后再随机上下翻转，改变图像对比度、亮度、饱和度、色调等特征，进一步扩大数据量。使用上述数据增强手段将有效减轻过拟合现象，提升模型的泛化能力。

(7) 对图像进行标准化处理^[34]，可以降低输入图像的冗余性，去除输入特征间的相关性，使网络对图片的动态范围变化不敏感。主要操作如下式：

$$img_{out} = \frac{img_{in} - mean}{adjusted_{stddev}} \quad (4-3)$$

$$adjusted_{stddev} = \max \left(stddev, \frac{1.0}{\sqrt{NumElements}} \right) \quad (4-4)$$

式中 img_{in} ——输入图像的 RGB 三通道像素值；

$mean$ ——三通道像素的均值；

$stddev$ ——三通道像素的标准差；

$NumElements$ ——三通道各自的像素个数。

图像经过标准化处理后，即可输入到模型中进行训练或预测。

4.3.5 卷积神经网络模型训练

加载训练模型，按一定大小的批次 (`batch=128`) 输入图像，对每个 `batch` 计算前向传播的输出值和损失值，之后通过反向传播算法计算梯度，最后采取随机梯度下降法更新网络参数。

网络训练过程中，随着训练批次的增加，损失值 (`loss`) 逐渐下降 (如图 4-13)，准确率 (`accuracy`) 不断上升 (如图 4-14)。由图 4-14 可以看出，在选取合

适大小的训练批次以及学习率的情况下，模型 loss 值平稳下降，在训练迭代次数（epoch）达到 50 之后，下降速度明显放缓，模型逐渐收敛；当 epoch 达到 100 时，loss 逐渐趋近于 0.15，并不再有明显下降，模型完成收敛。训练集以及测试集的 accuracy 整体趋势较为相像。在 epoch 达到 50 之后，测试集 accuracy 逐渐趋于平稳，训练集 accuracy 在 95% 与 100% 之间不断波动，这是由于训练集图像在输入模型前进行随机的对比度、亮度变换等操作，导致在当前迭代次数下，模型并没有完全拟合所有可能出现的图像数据。训练在 epoch 达到 220 之后，测试集 accuracy 稳定为 95.5%，不再发生变化，表示模型已经收敛

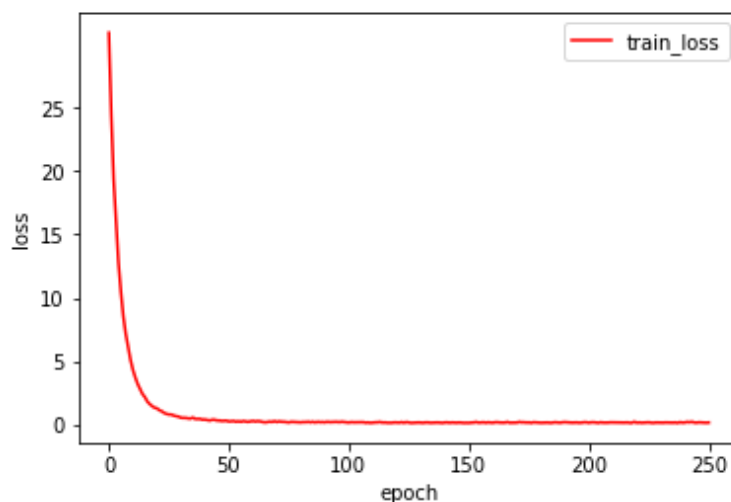


图 4-13 训练过程中损失值变化曲线

Fig. 4-13 Loss curve of training

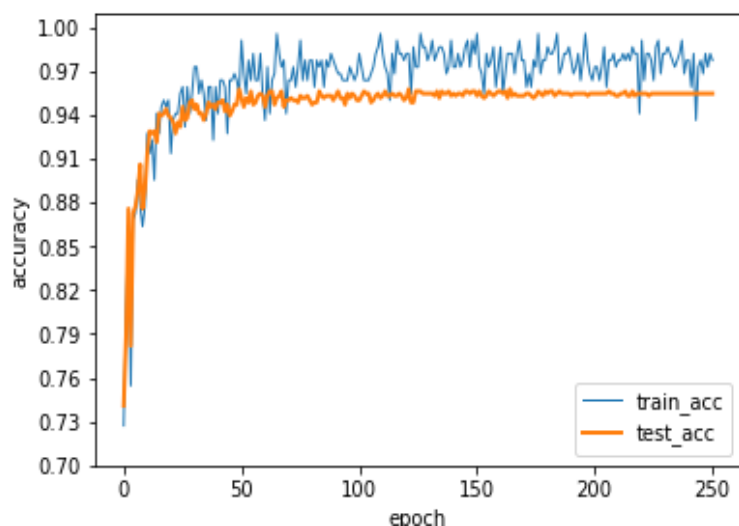


图 4-14 训练过程中准确率变化曲线

Fig.4-14 Accuracy curve of training

4.3.6 实验平台信息

本章实验在中国虚拟天文台的虚拟机上完成，此服务器 CPU 芯片型号是 Intel(R) Xeon(R) CPU E5-2682 v4 @ 2.50GHz，并配有 16G 内存和 260G 固态硬盘；GPU^[36]型号是 Tesla P4，内存 8G，由 Nvidia 公司生产；安装 ubuntu16.04.5 英文版操作系统。深度学习模型搭建运行在使用 Anaconda3 搭建的虚拟环境中，环境中添加 CUDA Toolkit 8.0 深度学习工具和 cudnn6.0 深度学习加速库。此外，编程工具选用 Vim，编程语言为 Python，第三方算法库包括 pandas 0.20.3、numpy 1.13.1、math、os、PIL、sklearn、time、matplotlib、re、argparse 等。

深度学习开发框架使用 Tensorflow，Tensorflow 由 Google 公司推出，在 2015 年 11 月正式开源。具有以下特征^[37]：

1) 高灵活性。Tensorflow 对神经网络进行了抽象封装，具有灵活的编程接口，支持各类模型的组合与扩展，能够快速实现各种算法。

2) 高执行性能。Tensorflow 支持分布式扩展，利用分布式技术进行并行计算可以有效提高计算的吞吐量，提升训练效率。

3) 可移植性。Tensorflow 可在 Windows、MacOS、Linux、iOS、Android 等平台上使用，便于将同一算法移植到不同的设备上使用，并可选择在 CPU 或 GPU 上运行^[38]。

4) 可复现性。深度学习算法多是基于概率论发展起来的学科，因此很多实验具有一定程度的随机性。Tensorflow 可保存训练过程中的各个参数及网络状态，保证模型可以复现。

4.3.7 结果分析

由于样本种类较多，并且存在数量不均衡的特点。因此，我们采用精确率 (Precision)、召回率 (Recall) 以及调和值 F1-Score 三个指标对模型预测结果进行评价。精确率表示预测为正的样本里真正正样本的比例，即在预测为正样本的结果中，会有两种可能，一种是预测正确的，预测样本真实值为正类，预测结果也为正类 (TP)，另一种是预测错误的，样本真实值为负类但被预测为正类 (FP)，精确率 P 的表达公式如下：

$$P = \frac{TP}{TP + FP} \quad (4-5)$$

召回率是针对某类样本而言的，它表示正类样本被成功预测的概率，对于所有正类样本的预测结果，同样也有两种可能，一种是正确的预测，即正类样本被成功预测 (TP)，另一种是错误的预测，即将正类样本预测为负类 (FN)，召回

率 R 的表达式如下:

$$R = \frac{TP}{TP + FN} \quad (4-6)$$

F1-Score (F1) 是精确率和召回率的调和均值, 相当于精确率和召回率的综合评价指标, 可由公式如下:

$$\frac{2}{F_1} = \frac{1}{P} + \frac{1}{R} \quad (4-7)$$

将式 (4-7) 变换后 F1 可表示为:

$$F_1 = \frac{2TP}{2TP + FP + FN} \quad (4-8)$$

将验证集数据用训练好的 CNN 模型进行分类预测, 结果中各项类别的精确率、召回率、和 F1-Score 如图 4-15 所示:

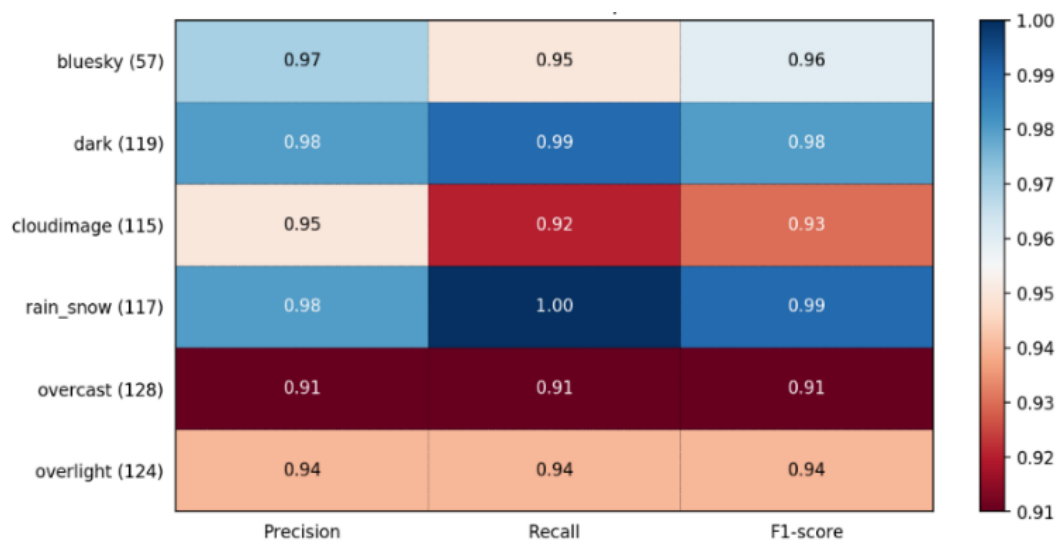


图 4-15 各项类别的精确率、召回率和 F1-Score

Fig.4-15 Precision、Recall and F1-Score for each category derived from CNN classifications

从图 4-15 可以看出, CNN 方法在各类别云图的预测中都取得了非常好的效果, 其分类精确率、召回率和 F1-score 均达到 90% 以上, 平均分类准确率达到 95% 以上, 证明了模型优异的分类能力。

4.4 泛化能力验证

在天文观测领域中, 由于全天空地基云图的获取过程并没有统一的拍摄设备和拍摄标准, 且拍摄时图像受环境因素影响较大, 所以不同天文观测台站、不同全天相机、不同时段拍摄的图像往往差别较大, 属于异分布的数据集。使用其他全天空地基云图数据集对分类模型进行验证, 可进一步判断模型的泛化能力和可迁移性能。

4.4.1 验证数据集介绍

本研究采用了丽江观测站全天相机数据集对模型的分类能力进行验证。中国科学院云南天文台丽江观测站是我国南方重要的天文观测基地，目前拥有多台望远镜：2.4m 望远镜、1.8m 望远镜、伽玛暴与光学暂现源观测网 4 号站(BOOTES-4)和台湾自动化望远镜 (Taiwan Automated Telescope, TAT) 等。全天相机为这些望远镜的观测提供实时的云量信息，在天文观测中起着必不可少的辅助作用。

选取 2016 年 11 月 26 日到 2016 年 12 月 26 日丽江观测站全天相机拍摄的全天空云图数据对模型进行验证，共有 5807 张图像，单张图像分辨率为 720×480 ，按照 4.3.1 节中描述的分类标准对其进行人工分类，其中 dark 类别 4257 张、bluesky 类别 454 张、overlight 类别 614 张、cloudimage 类别 321 张、overcast 类别 115 张、rain_snow 类别 46 张。部分数据如图 4-16。



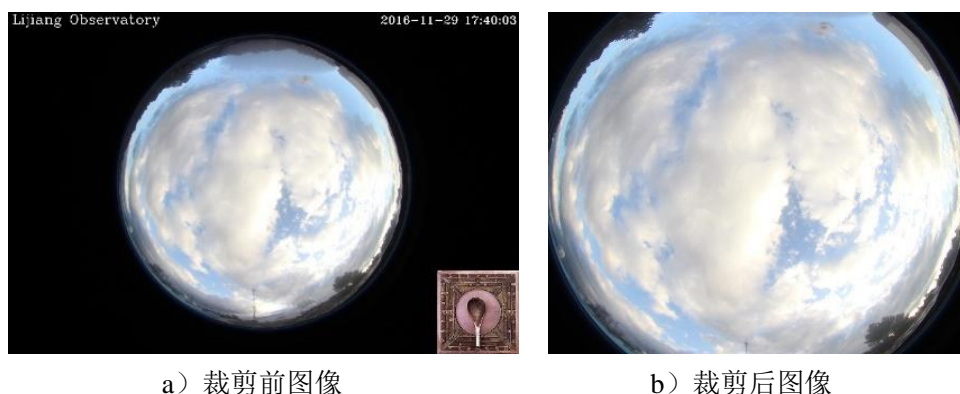
图 4-16 丽江观测站部分全天相机数据

Fig. 4-16 Partial all-sky camera data from Lijiang Observatory

4.4.2 验证数据预处理

在验证数据输入到分类网络前，需要对数据进行预处理。拍摄过程中，由于拍摄设备、相机参数的不同，图像尺度会有所差异。对图像进行预处理操作，使得验证数据集在输入模型前尺度信息尽可能与训练数据一致，可以提高模型对数据的分类准确率。

原始图像(如图 4-17a)分辨率为 720×480 ，以图像中心为基准，截取 400×320 大小的图像(如图 4-17b)，获取与训练数据近似长宽比的有效信息。之后再压缩至 128×128 ，最后做中心裁剪，得到分辨率为 100×100 的图像，保持输入图像与训练集图像尺度的一致。经过 4.3.5 中的标准化处理后，输入到模型中进行类别预测。



a) 裁剪前图像

b) 裁剪后图像

图 4-17 对丽江站全天相机数据进行裁剪

Fig.4-17 Crop all-sky images of Lijiang station. a) Image before cropping. b) Cropped image

4.4.3 模型验证结果分析

模型对验证集各项类别分类结果的精确率、召回率、和 F1-Score 如图 4-18 所示。

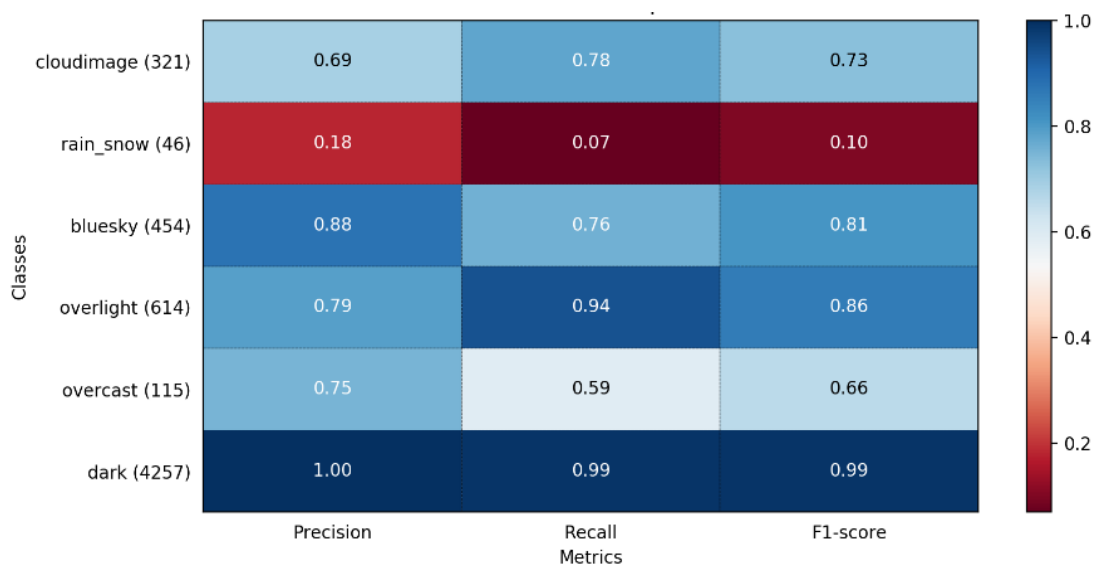


图 4-18 验证集数据各项类别的精确率、召回率和 F1-Score

Fig. 4-18 Precision、Recall and F1-Score for each category of validation set

预测结果的混淆矩阵如图 4-19 所示，横坐标表示各类别的预测值，纵坐标表示各类别的真实值。结合图 4-18 和 4-19 可以看出，模型在异分布数据集上有着不错的表现，对数量最多的 dark 类判定精准，并且对 cloudimage、bluesky、overlight 和 overcast 类别也有较好的区分能力。但对数量最少的 rain_snow 类别预测结果较差，并且 bluesky 的召回率、overcast 的召回率、cloudimage 的准确率相对较低，接下来将对预测结果较差的部分进行详细分析。

	dark	overcast	overlight	bluesky	rain_snow	cloudimage
dark	4210	6	5	23	1	12
overcast	0	68	38	0	0	9
overlight	1	14	576	7	10	6
bluesky	0	0	22	345	2	85
rain_snow	0	2	28	10	3	3
cloudimage	1	1	58	9	1	251

图 4-19 测试集数据预测结果的混淆矩阵

Fig. 4-19 Confusion matrix for test set data

从分类结果的混淆矩阵（图 4-19）中可以看出，rain_snow 大部分预测为 overlight，误判图像如图 4-20 所示。可以看出，图像上存在着严重的过曝光现象，模型更加关注到过曝光的特征，可能是误判的主要原因。此外，测试集中 rain_snow 类雨滴数量相较训练集明显减少，也可能是 rain_snow 类预测结果整体较差，准确率较低的主要原因。

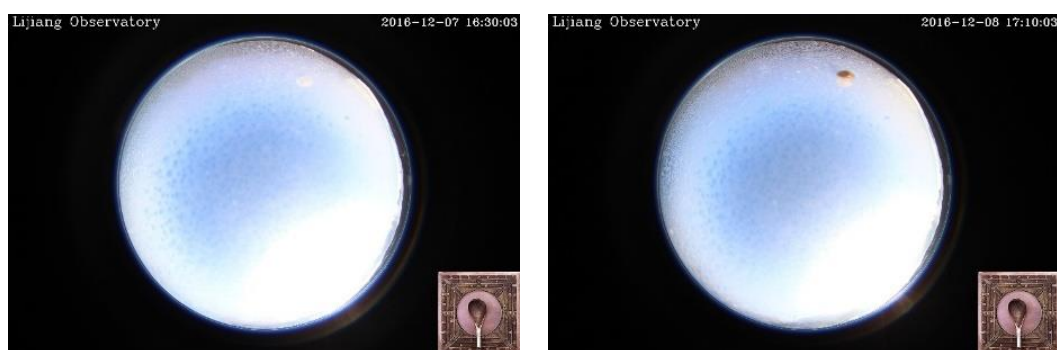


图 4-20 误判为 overlight 类的 rain_snow 类图像

Fig. 4-20 Rain_snow misjudged as overlight

由图 4-19 可知，bluesky 召回率、cloudimage 准确率较低的主要原因是一部分 bluesky 被预测为 cloudimage，典型误判图像如图 4-21，根据对图像的观察，这些图像与训练集中的 bluesky 有较大差异，图像整体偏暗，并且存在较为明显三个污点，可能是导致误判的原因。

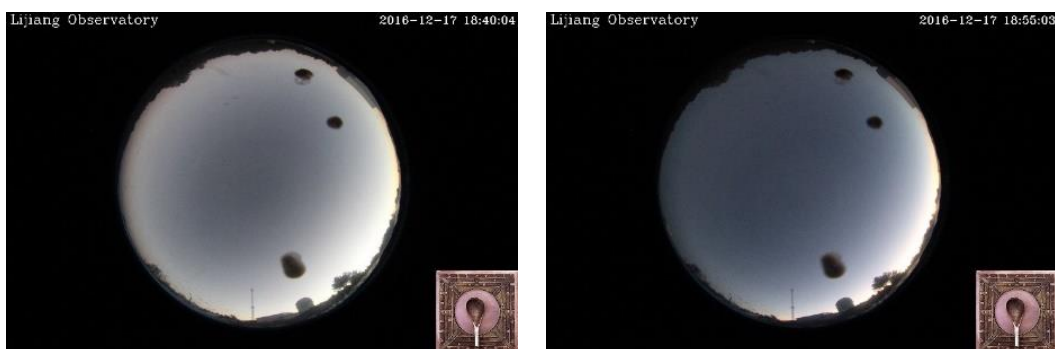


图 4-21 误判为 cloudimage 类的 bluesky 类图像

Fig.4-21 Bluesky misjudged as cloudimage

overcast 的召回率较低，由图 4-19 可知，overcast 主要误判为 overlight，典型误判图像如图 4-22 所示，误判图像亮度较高，整体特征与训练集中 overlight 类较为相近。这种区别主要是由于相机拍摄时的参数设置不同所导致。

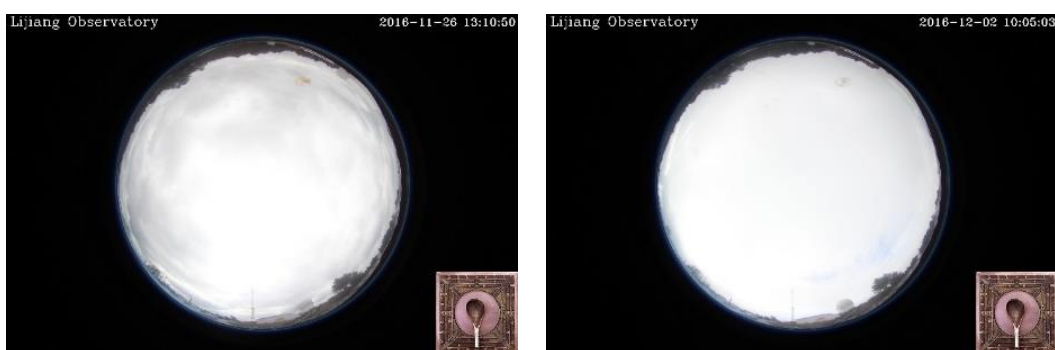


图 4-22 误判为 overlight 类的 overcast 类图像

Fig. 4-22 Overcast misjudged as overlight

4.5 云量检测方法验证和对比分析

判断区域内云量以及云所在位置是云量检测的关键问题。云点识别的一般方法是先计算云图红色波段和蓝色波段灰度值的比值，再通过一定的阈值判断该图像上每一点是否为云。K均值聚类算法（k-means）是一种无监督的聚类算法，它可以将相似的对象归到同一个簇中，也适用于云天对象的区分。本研究从雪龙号全天空地基云图数据集中挑选出部分典型云图对以下三种云点识别方法进行了验证：包括固定阈值的红蓝波段比值方法、最大类间方差法确定自适应阈值的红蓝波段比值方法^[39]以及k-means聚类方法^[40]。

4.5.1 固定阈值法

阈值分割法是图像分割中的经典方法。图像中的目标景物和背景在灰度上往往是不同的，而同种类景物灰度值却比较接近，因此，通常通过设置阈值来将图

像上所有的像素点分成若干类,来实现目标景物与背景的分割。霍娟等人将蓝色通道亮度和红色通道亮度的比值作为判断云的灰度特征,并使用经验阈值对特征值进行判别,完成目标(云)和背景(天空)的区分。具体实现方法是先计算图像整体的红蓝两个波段的比值,再以经验阈值1.30作为分割标准,对图像进行分割,大于1.30的像素点预测为云,小于1.30的像素点预测为天空。固定阈值法实现方式简单,但是提取特征弱、适应性差,当云量较多或云量较少时使用经验阈值往往会出现预测不准确的情况。

4.5.2 基于最大类间方差的自适应阈值法

由Otsu^[41]提出的基于类间方差最大的分割算法一直被认为是分割阈值自动选取的最优方法。算法具体实现如下:

选取Th为目标与背景的分割阈值后,分别计算整幅图像前景点和背景点的个数、灰度值总和,并通过下述公式,计算得到前景和背景图像的方差g:

$$u = w_0 \times u_0 + w_1 \times u_1 \quad (4-9)$$

$$g = w_0 \times (u_0 - u)^2 + w_1 \times (u_1 - u)^2 \quad (4-10)$$

联立上面两式可得:

$$g = w_0 \times w_1 + w_1 \times (u_0 - u_1)^2 \quad (4-11)$$

即:

$$g = \frac{w_0}{1 - w_0} \times (u_0 - u)^2 \quad (4-12)$$

式中 w_0 ——前景点占图像的比例;

u_0 ——前景点平均灰度值;

w_1 ——背景点占图像的比例;

u_1 ——背景点平均灰度值;

u ——图像的总平均灰度值;

g ——前景和背景图像的方差。

每个Th均会对应一个方差g,最大g所对应的Th值在Otsu算法中被认为是最可以区分前景和背景的阈值,即最佳分割阈值。在云量检测中可以使用Otsu算法计算出最佳的红蓝波段比值的分割阈值,从而实现云与天空的区分。

4.5.3 k-means 聚类方法

k-means是一种简单的无监督聚类算法,应用十分普遍。k-means算法首先在样本空间中随机选取k(分类数目)个样本值作为起始的聚类中心,即确定k个聚

类，然后计算每个样本与各个聚类中心之间的距离，通过比较距离大小把样本值分配给距离最小的聚类。此时，每个聚类中心以及分配给它们的样本值共同组成一个聚类。当所有样本值均被分配完毕后，计算每个聚类的均值并将其确定为新的聚类中心。不断重复该过程直到聚类中心不再发生变化，所有样本值均不再被重新分配或者迭代达到预定的次数。

k-means 算法具体实现方法如下：

- 1) 随机选择 k 个点作为聚类中心， k 为分类数目。
- 2) 对整个样本空间进行遍历，每个样本值划分给距离最近的中心点。
- 3) 分别计算每个聚类的均值并将计算得出的 k 个均值作为新的聚类中心。
- 4) 重复第二步到第三步，当中心点不再发生变化或执行了足够多的迭代次数，停止迭代过程，得出分类结论。

4.5.4 云点识别结果及分析

使用 3.3.1 节中提到的模板法去除黑色区域与建筑物等干扰信息。添加模板过程如图 4-23 所示。云点识别算法均只应用在模板图像内白色区域所对应的位置，黑色部分不用于计算。



图 4-23 使用模板法去除无关信息

Fig.4-23 Template method removes irrelevant information. a) Original image. b) Template image. c) Image after adding template

分别使用 `cloudimage` 类中云量适中、云量较少和云量较多的三类图像对上述算法进行验证。结果如图 4-24，从左到右图像处理方法分别为：未处理、**k-means** 聚类法、固定阈值法和自适应阈值法，其中白色代表云，蓝色代表天空。

由图 4-24 的结果可以得出以下结论：

- 1) **k-means** 聚类方法识别效果差，不适用于云点识别。
- 2) 对于阈值法，太阳或过曝光区域容易被识别为云。
- 3) 固定阈值法的自适应能力较差，容易发生大面积的识别错误。
- 4) 当天空中的云量占比很高或很低时，蓝红波段处理后的直方图可能不呈双峰分布，自适应阈值法计算的阈值容易出现偏差，影响云量检测的精度。

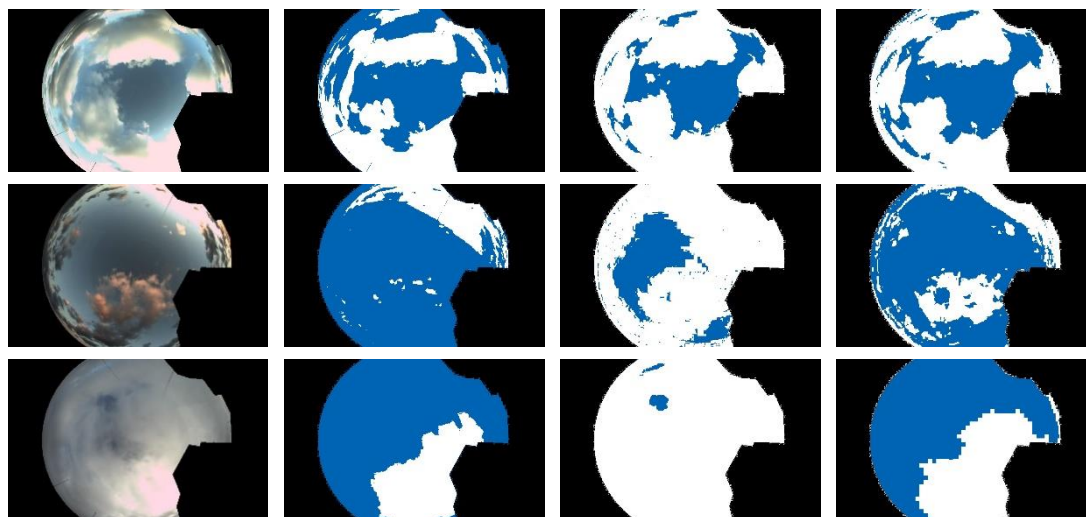


图 4-24 云点识别方法验证效果

Fig. 4-24 Cloud point recognition method verification effect

上述三种方法均在一定限制条件下有效，对于变化多样的全天空云图适应性较弱、实用性能较差。实现鲁棒性高、适应性强的分类方法将是一下步工作的重点。

4.6 本章小结

本章首先详细介绍了全天相机系统获取数据的数据量、数据格式和数据特点，并根据应用场景对其进行人工分类，创建了雪龙号全天相机数据集。设计开发了用于图像分类的卷积神经网络算法模型，完成了数据预处理、模型训练以及结果分析等任务。之后使用云南丽江天文观测站全天相机数据对模型进行验证，证明了算法的有效性和良好的适应能力。最后总结了一些主流的全天空地基云图云量检测方法，对方法进行验证、比较它们的优缺点并提出了新的展望。

第5章 雪龙号全天相机数据发布平台

5.1 引言

天文数据往往处于开放共享的状态,这有利于科学研究的快速发展。全天相机系统在随雪龙号航行过程中已经获取了大量的全天空地基云图数据,并且在未来可能获得更多的数据。开发一个可以用来存储、展示和共享本次数据的 web 平台可以极大地提升数据的价值。本文基于 Django 架构、MySQL 数据库以及众多 web 网站编写技术,设计搭建了易于访问、用户友好的雪龙号全天相机数据发布平台,实现了此次全天相机系统获取数据的开放共享。本章将对搭建平台使用到的相关技术、平台的总体设计、平台的具体实现内容和使用时的注意事项做详细介绍。

5.2 相关技术概述与开发环境介绍

搭建平台主要使用了如下技术: Django 框架、MySQL 数据库和 Apache 服务器,并使用 jQuery、Ajax、Bootstrap、HTML5、CSS3、JavaScript 和 Python 等技术或编程语言编写前端页面和后端逻辑应用程序。

Django 是一个开源的 web 开发框架^[42],使用 python 语言编写,拥有用户多、第三方库丰富、开发文档详细完善等优点。Django 搭建 web 应用的主要模式为 MTV^[43]模式,即模型(Model)、模板(Template)和视图(Views)。模型是数据存取层,处理与数据相关的所有事务。模板是表现层,决定如何在页面或其他类型文档中进行显示。视图是业务逻辑层,包含存取模型及调取恰当模板的相关逻辑。Model 与数据库实现 ORM(对象-关系映射),使用时无需关注数据库中数据的结构层次,直接通过操作 Model 对象来操作数据库^[44]。Django 架构如图 5-1 所示。

MySQL 是一个关系型数据库管理系统。MySQL 数据库具有如下优点:

- 1) 免费开源。
- 2) 可以对大型数据进行处理。
- 3) 支持常见的 SQL 语句规范。
- 4) 软件小巧、安装简单、可移植性高、运行效率高。
- 5) 调试、管理、优化简单。

6) 可以与 Django 框架实现无缝对接,适用于雪龙号全天相机数据发布平台的开发。

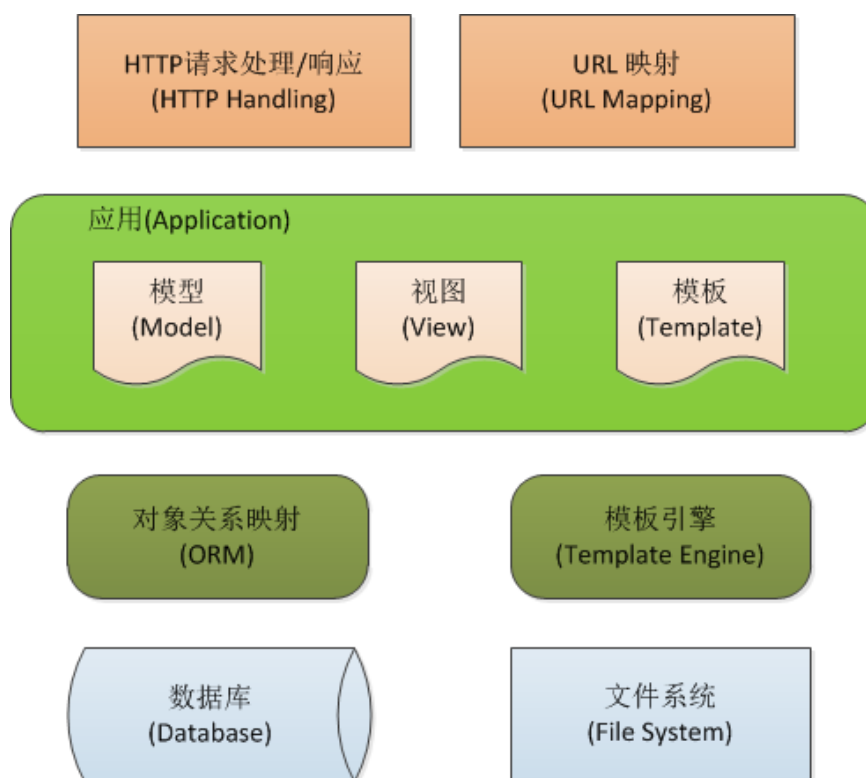


图 5-1 Django 架构总览图

Fig. 5-1 Django architecture overview map

Apache 是一款应用广泛的 Web 服务器软件，它可以运行在 Windows7、Windows10、MacOS、Linux 等几乎所有主流的操作系统上，并且用户群体广泛、安全性高。采用 Apache 作为数据平台的服务器软件可以确保网站的兼容性和稳定性。

超文本标记语言 (HTML5) 是网页制作必备的编程语言，用于编写网页的基本内容。HTML5 包含了使用者需要获取的主要内容，包括文字、图片、视频等。CSS3 的作用是增加网页的效果，网页上动态文字的色彩、字体、动画效果等均通过 CSS3 来实现，CSS3 的存在使得网站变得丰富多样。

JavaScript 是一种脚本语言，通过 JavaScript 用户可以和网页进行交互，增强用户的可操作性，从而形成良好的用户体验。jQuery 是一个 JavaScript 库，对 JavaScript 进行了更高级的封装，实现了很多常用的功能，可以减轻开发者使用 JavaScript 时的编程负担。Ajax 则是一种与服务器交换数据的技术，在前端网页与后台服务器进行数据交换时，使用 Ajax 技术并不会更新整个网页，只会对需要进行交互的部分进行更新，从而提高网页刷新效率，增强用户体验^[45]。jQuery 提供了丰富的函数库用于 Ajax 开发，通过使用 jQuery 可以快速的将 Ajax 技术应用到实际工作中^[46]。

Bootstrap 是基于 HTML、CSS、JavaScript 开发的前端开发框架，拥有应用简

单、响应式布局、跨设备、跨浏览器等优点，使用它可以快速高效的搭建可用于多设备的 Web 应用。Bootstrap 的响应式布局设计，让一个网站可以兼容不同分辨率的设备，页面根据浏览器显示的大小进行自适应缩放，从而给用户提供更好的视觉使用体验。使用 Bootstrap 搭建的界面清新、简洁，排版利落大方。

开发环境：Windows10 操作系统，Inter(R) Core(TM) i5-4210U CPU @1.70GHz，8G 内存，1000G 机械硬盘，200G 固态硬盘。编程时使用到的 IDE 是 PyCharm，使用到的后端编程语言是 Python；网站运行服务器：Windows Server 2012 R2 操作系统，Inter(R) Xeon(R) E5-2682 v4 @2.50GHz，4G 内存，500G 机械硬盘。

5.3 平台总体设计

基于展示、共享本次数据的目的，本章搭建的雪龙号全天相机数据发布平台主体由网站首页、数据展示界面和数据下载界面三部分组成：平台首页主要包括雪龙号全天相机系统项目和拍摄数据的整体情况介绍，并包含平台标签 (logo)、标题栏、数据公开声明、版权和致谢等内容，科研人员可以在这里快速了解项目内容；数据展示页面可查看图像信息和拍摄图像时雪龙号所处位置的信息和气象信息；数据下载页面会将数据下载链接和数据下载方式提供给需求者。除上述三个主要页面外，平台还包括登录注册和 china-vo 两个跳转链接功能，通过点击登录注册可以跳转到 china-vo 平台的登录注册界面，通过点击 china-vo 可以跳转到 china-vo 平台网站首页，平台主体结构和各部分功能如图 5-2 所示。

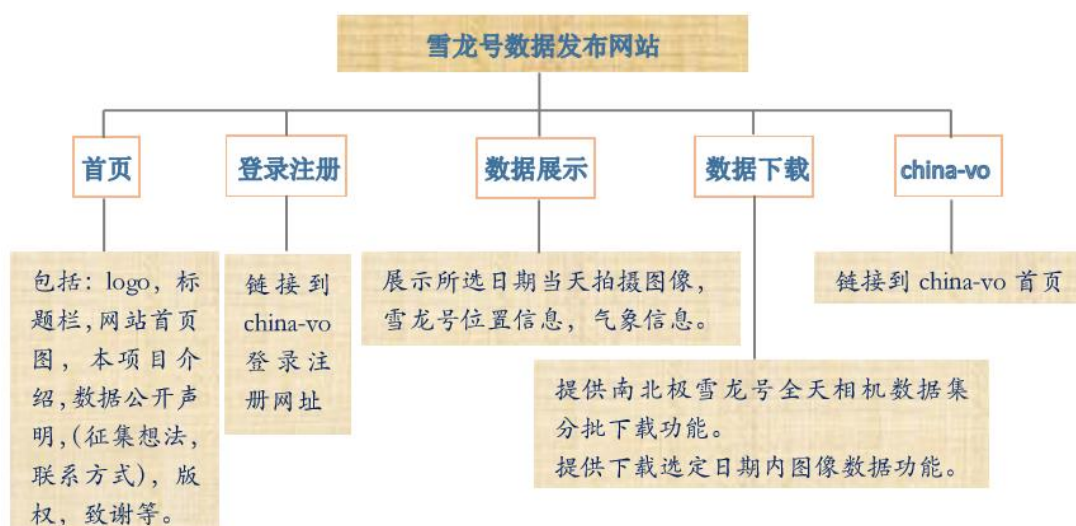


图 5-2 网站设计结构

Fig. 5-2 Website design structure

5.4 全天相机数据发布网站首页

可通过 <http://47.93.115.68:8083/publish/index/> 进行访问。首页主要包括全天相机系统介绍及数据介绍，内容如图 5-3 所示，主要由如下几部分组成：

(1) 雪龙号介绍：雪龙号是中国的极地考察船，集科考、运输、破冰等多功能于一体，承担着中国大量的南北极科考任务。

(2) 项目介绍：安装于雪龙号甲板上的全天观测设备,每隔 5 分钟进行一次全自动拍摄，在第 8 次北极航行(N8)与第 34 次南极航行(S34)过程中获取了大量的全天空云图数据。

(3) 全天设备介绍：设备由中国虚拟天文台(China-VO)独立研制开发，拍摄全自动进行，随船航行拍摄过程中经历了和高温等恶劣环境考验，具有很好的鲁棒性。

(4) 数据及平台介绍：全天观测设备拍摄全自动运行，具有良好的成像性能。N8 航行期间获取图像约 13000 张，S34 航行期间获取约 53000 张，图像分辨率为 1000*625，格式为 JPG。

(5) 部分典型代表图像展示：展示部分典型代表图像，包括有晴空图像、多云图像和少云图像。

(6) 应用设想：分为科普教育、科学研究以及集思广益三个方面。科普教育方面讲述了可以根据图像时间上的连续性制作动画，展示南北极天空或天气变化过程；科学研究方面阐述了数据可用于云检测相关算法研究和已实现的云图种类的自动化识别；集思广益则是希望其他科研工作者可以去挖掘本次数据更多的价值。

(7) 数据公开声明：为促进全天空地基云图相关算法研究以及数据在相关科普教育领域的推广，数据对国内外科研人员免费公开。

(8) 致谢：本项目得到中国虚拟天文台和中国天文数据中心提供的数据资源和技术支持。中国天文数据中心得到国家科技部国家科技基础条件平台项目“地球系统科学数据共享服务平台”和“国家基础科学数据共享服务平台”的资助。感谢国家天文台-阿里云天文大数据联合研究中心对本项工作的支持。

在网站最上方的标题栏中提供数据展示网页和数据下载网页的跳转入口。并添加中国虚拟天文台logo。

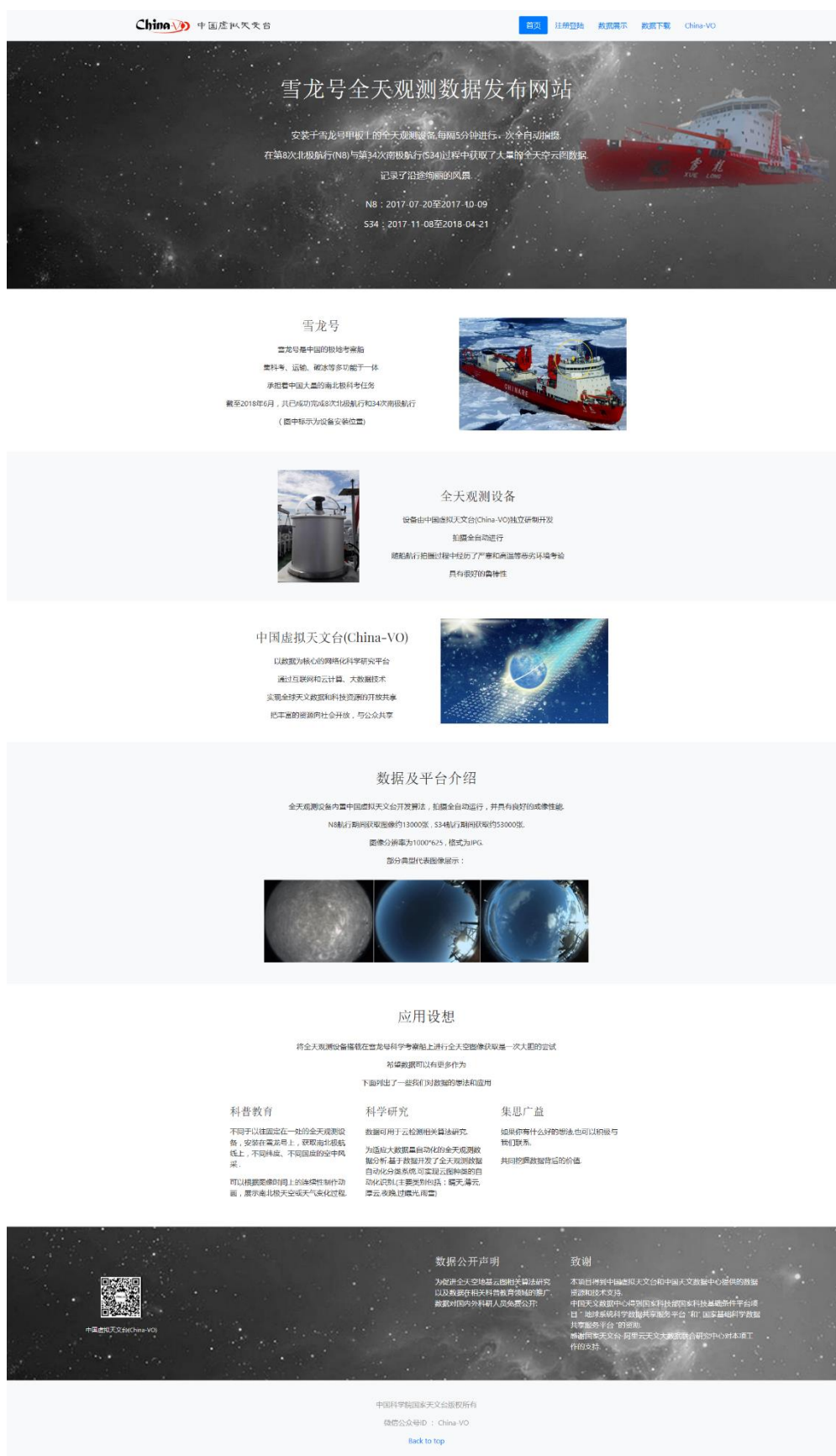


图 5-3 雪龙号全天相机数据发布网站首页

Fig. 5-3 The home page of XueLong all-sky camera data release website

5.5 全天相机数据展示界面

可通过 <http://47.93.115.68:8083/publish/display/> 进行访问，或通过点击标题栏中的数据展示选项进行跳转。网站初始内容包括展示界面功能介绍、部分数据图像、相机位置信息以及相机气象信息。界面默认展示最后一天（2018年4月21日）全天相机拍摄的图像信息以及当天相机所处位置的气象与位置信息，单天展示图像数量小于20张。位置信息包括UTC时间、经度、纬度、航速和航向。气象信息包括UTC时间、温度、湿度、气压、风向和风速，页面如图5-4所示。

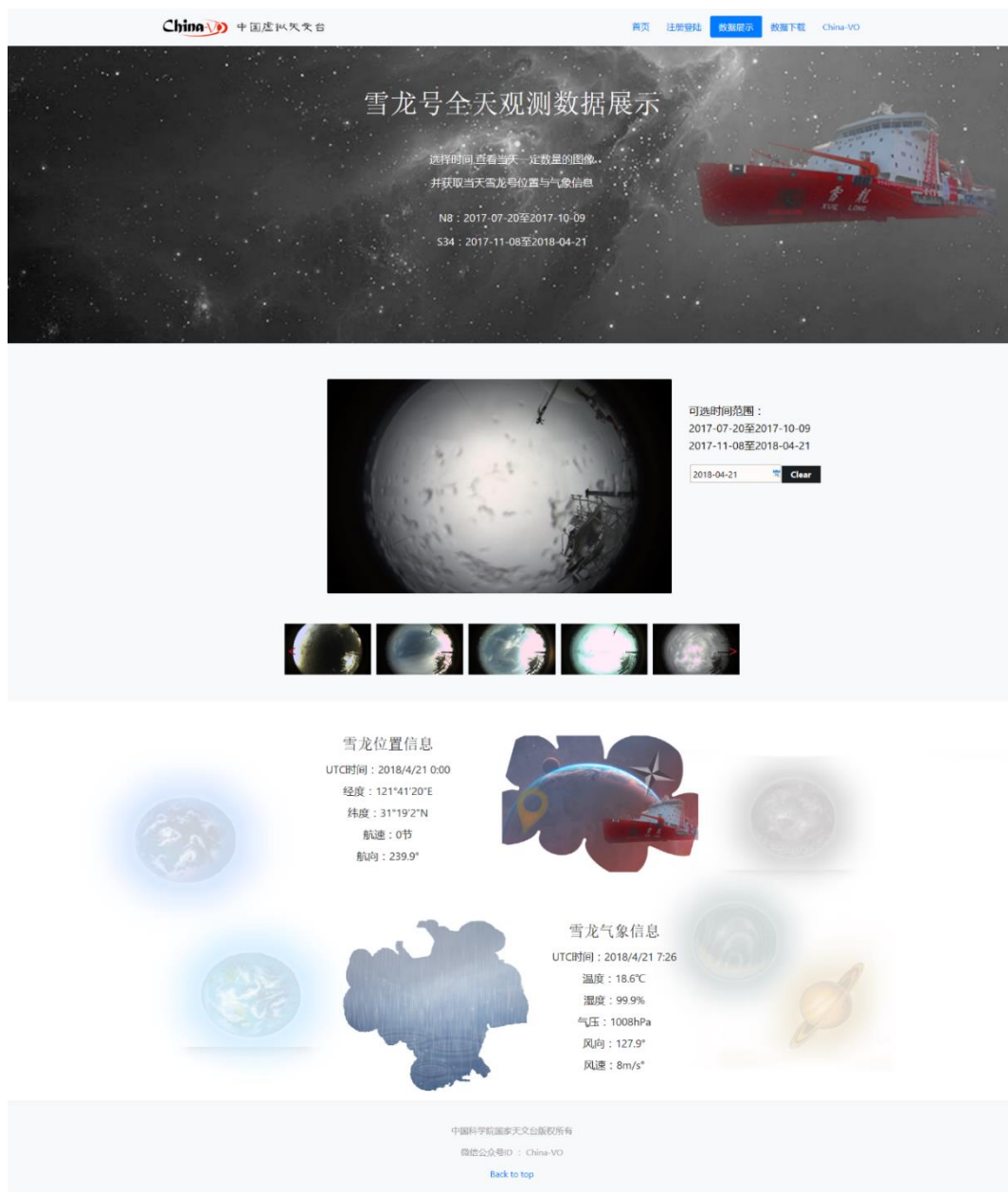


图 5-4 雪龙号全天相机数据展示界面

Fig. 5-4 Data display interface of XueLong all-sky camera data release website

5.5.1 数据展示界面交互功能

通过点击日期文本框，如图 5-5 所示，可弹出日历。选择时间后，可以在页面看到所选日期当天一定数量的图像信息，相机位置以及气象信息。选择日期在提示可选时间范围外，图像、气象、位置信息均不会发生改变。

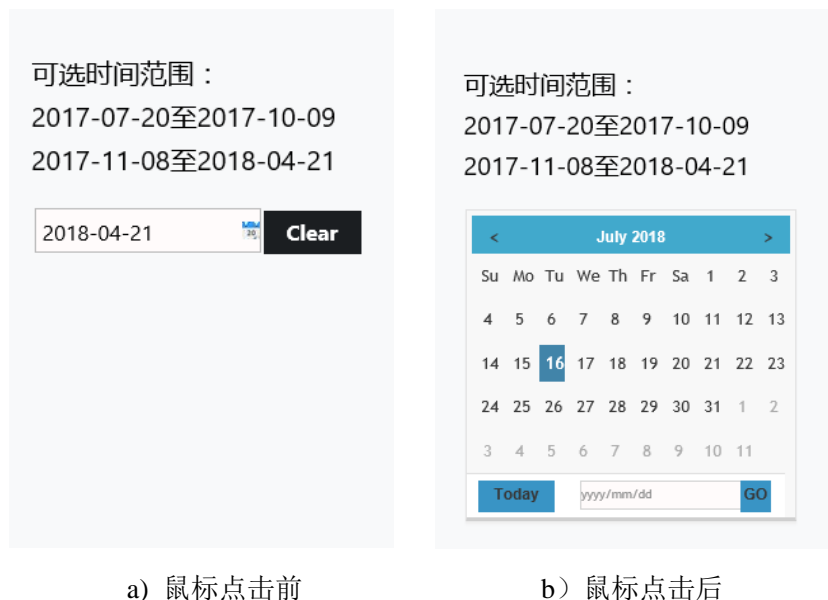


图 5-5 日期选择区

Fig. 5-5 Date selection area. a) Display before mouse click. b) Display after mouse click

获取目标日期图像后，通过点击图5-6中的左右红色箭头来滚动当天展示出的所有图像。通过单击滚动框中的图像，可将其放大显示到滚动框上方的大图展示区域。

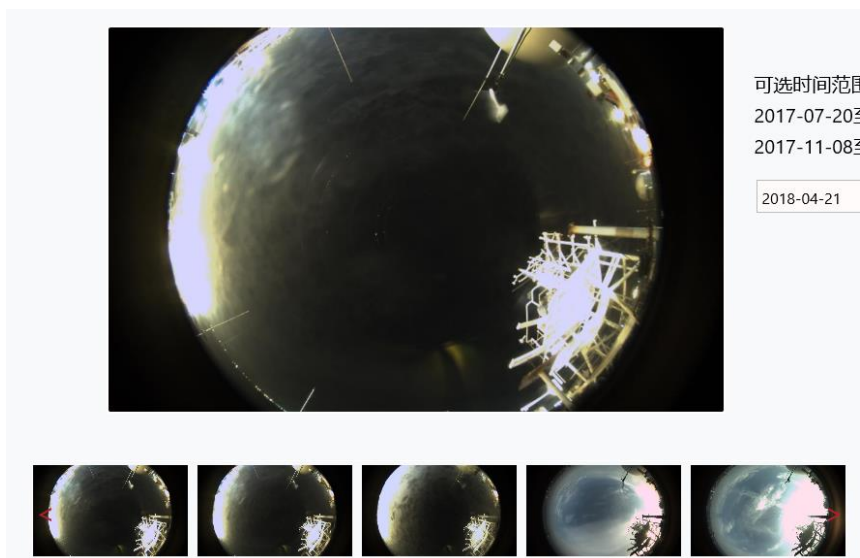


图 5-6 数据图像展示区

Fig. 5-6 Image display area

5.5.2 数据展示界面实现方法

日历插件通过 jQuery 技术生成，位置信息时间段包括 2017-07-20 至 2017-10-10 和 2017-11-08 至 2018-04-21，位置信息表详细设计如表 5-1；气象信息时间段是 2017-11-20 至 2017-04-12，气象信息表详细设计如表 5-2，位置信息表和气象信息表均保存在 MySQL 数据库中。图像数据保存在服务器上，当在日期文本框中选择日期时，前端将会发送时间信息与更新加载命令给后台，后台应用程序根据时间信息获取对应的图像、气象、位置信息数据并通过 Ajax 技术将其异步更新展示到网页上。

表5-1 位置信息数据结构表

Table 5-1 Location information data structure table

字段	类型	备注
id	int(11)	主键
UTC	varchar(25)	UTC 时间
longitude	varchar(20)	经度
latitude	varchar(20)	纬度
speed	varchar(10)	航速（节）
course	varchar(10)	航向（度）

表5-2 气象信息数据结构表

Table 5-2 Meteorological information data structure table

字段	类型	备注
id	int(11)	主键
UTCtime	varchar(25)	UTC 时间
temp	varchar(20)	温度
hum	varchar(20)	湿度
press	varchar(20)	气压
mwinddir	varchar(20)	风向
mwindspeed	varchar(20)	风速

5.6 全天相机数据下载界面

可通过 <http://47.93.115.68:8083/publish/download/> 访问全天相机数据下载界面，或通过点击标题栏中数据下载选项跳转。下载界面包括下载功能介绍、图像数据直接下载功能和获取自定义时间段内图像数据的功能。网页内容如图 5-7 所

示。

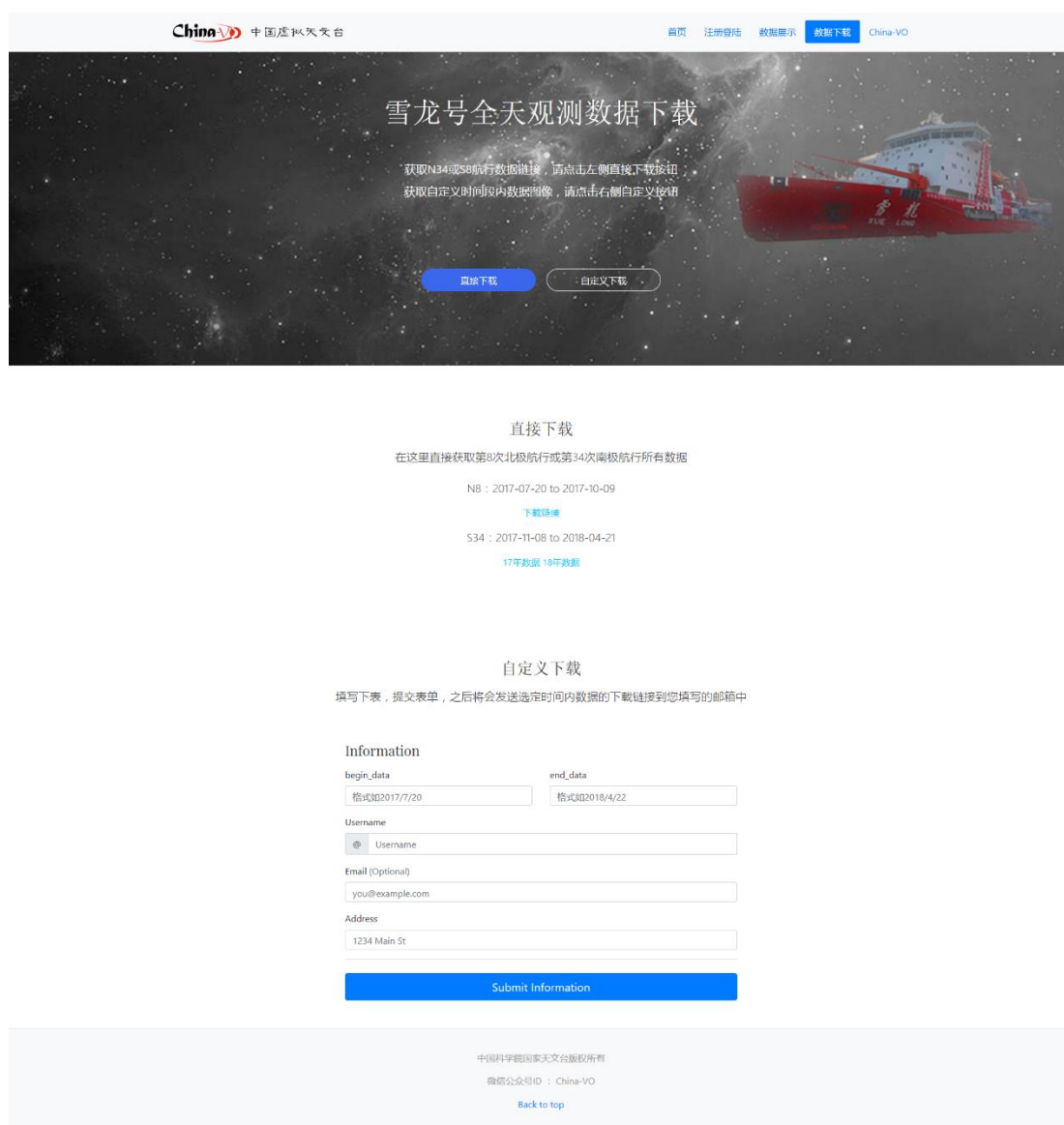


图 5-7 全天相机数据下载界面

Fig. 5-7 Data download interface of XueLong all-sky camera data release website

5.6.1 数据下载界面交互功能

如图 5-8 所示，在下载功能介绍部分，点击直接下载按钮将跳转到直接下载功能区域，点击自定义下载将跳转到自定义下载功能区域，或者通过向下翻动页面到达相应功能区域。



图 5-8 下载功能介绍及跳转功能区

Fig.5-8 Download function introduction and jump function area

直接下载功能分别提供 N8 航线（2017-07-20 至 2017-10-09）数据下载链接、S34 航线 17 年（2017-11-08 至 2017-12-31）数据下载链接和 18 年（2018-01-01 至 2018-04-21）数据下载链接，压缩包大小分别约为 546MB、907MB、1.06GB。如图 5-9，单击蓝色部分，将弹出下载框，可通过单击绿色下载按钮直接对数据进行下载。



图 5-9 直接下载功能区

Fig.5-9 Direct download function

除上述直接下载全部数据的功能外，设计了自定义下载数据的功能，用户可以通过此功能下载自己感兴趣时间段内的数据，节约资源以及下载时间。填写需求表单并提交后系统将会发送对应数据的下载链接到填写的邮箱中，表单如图 5-10 所示。通过鼠标右键单击 Submit information 按钮提交下载申请，如果表单内容填写不完整，将会在页面上展示对应框的具体填写要求，如图 5-11 中的红字部分。由于时间信息是数据压缩时的主要考虑因素，并且错误的时间信息可能会导致数据压缩失败，所以如果表单中填写的起止时间不在正确的数据获取时间范围内，设计的程序将会弹出对应的时间信息提示框（如图 5-12），用户需要关闭提示框并按要求重新填写时间信息。

自定义下载

填写下表，提交表单，之后将会发送选定时间内数据的下载链接到您填写的邮箱中

Information

begin_data

end_data

Username

Email (Optional)

Address

July 2018

Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

[Submit Information](#)

图 5-10 自定义下载功能区

Fig. 5-10 Custom download ribbon

自定义下载

填写下表，提交表单，之后将会发送选定时间内数据的下载链接到您填写的邮箱中

Information

begin_data
Valid start data is required.

end_data
Valid end data is required.

Username
Your username is required.

Email (Optional)
Please enter a valid email address for shipping updates.

Address
Please enter your shipping address.

[Submit Information](#)

图 5-11 表单填写不完整提示信息

Fig. 5-11 Form fill incomplete message

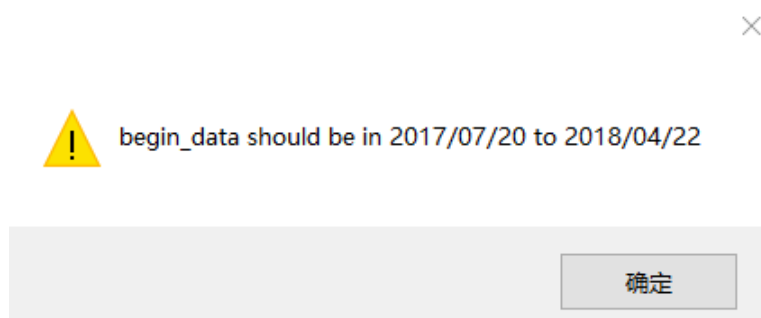


图 5-12 起止时间不满足要求提示信息

Fig. 5-12 Time does not meet the request prompt information

信息按要求完整填写后，点击提交表单按钮，将会弹出提示框，显示申请成功，后台应用把数据打包完成后将发送下载链接到申请邮箱中，如图5-13。

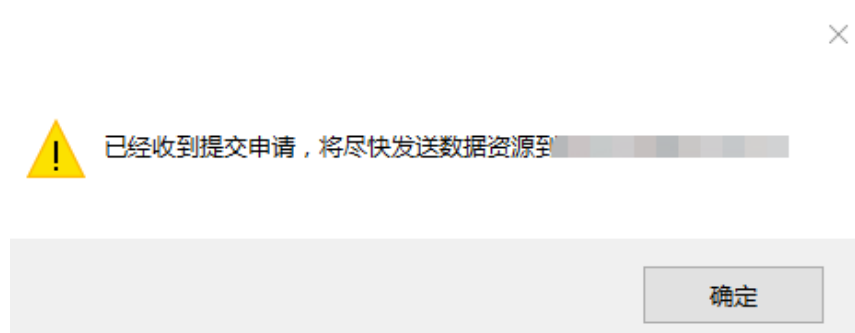


图 5-13 申请提交成功提示

Fig. 5-13 Application submission success prompt

5.6.2 数据下载界面实现方法

为实现上述功能，首先将南北极数据压缩为 ZIP 格式文件，因为整体数量较大，所以采取分批压缩的方法，北极航行期间拍摄数据压缩包名为 North_Polo.zip，大小为 546.58MB，南极航行期间拍摄数据较多，对 17 年和 18 年拍摄图像分别进行压缩，17 年数据压缩包名为 South_Pole2017.zip，大小是 907.64MB，18 年数据压缩包名为 South_Pole2018.zip，大小是 1.62GB，压缩后生成对应的下载链接，可通过直接下载方式访问下载。

当选择自定义下载功能，完整填写信息并提交后，申请信息将会保存到 MySQL 数据库中的表 publish_downloadinf 中，publish_downloadinf 结构详细设计如表 5-3，表 5-3 中的 status 字段将更新为 create。

后台主要包括压缩程序和定时清除压缩包程序。当压缩程序检查到 status=create 时，会根据当前字段时间段信息压缩对应图像，并将 status 更新为 compression。压缩完成后将下载链接信息发送给当前字段邮箱并将 status 更新为 send，另外程序会将发送邮件时的时间记录到表 5-3 中的 email_time 字段中。回

复内容包括压缩数据时间段信息、提示链接过期信息以及下载链接，回复内容如图 5-14 所示：

压缩时间为 2018-01-10 至 2018-01-11。
 链接将在七天后过期，请及时下载。
 下载链接为：
http://47.93.115.68:8083/static/south/data/2018-01-10_2018-01-11.zip

图 5-14 自动回复内容示例

Fig. 5-14 Auto reply content example

表5-3 用户提交信息数据结构表

Table 5-3 User submission information data structure table

字段	类型	备注
id	int(11)	主键
begin_time	varchar(50)	开始日期
end_time	varchar(50)	结束日期
username	varchar(50)	用户名称
email	varchar(254)	邮箱信息
address	varchar(50)	联系地址
status	varchar(20)	压缩包状态
create_time	varchar(50)	压缩包创建时间
email_time	varchar(50)	发送下载链接到指定邮件时间

为保证服务器的正常运行，后台添加了定时清除压缩包程序。该程序每天0点运行一次，将生成时间超过十五天的压缩包删除，删除压缩包后表5-3中的status字段被更新为delete，代表该压缩包已被删除。

5.7 平台使用注意事项

本软件适用于 Windows7、Windows10、MacOS、Linux 等操作系统，Internet Explorer 9.0 以上的浏览器；本软件是网页服务，用户打开浏览器，输入网站地址直接点击运行即可开启本软件；点击页面底端的 back to top 可返回到当前页面起始位置；无法访问网页时，尝试检查网络，并尝试刷新页面；点击链接后如果无法立即下载或不显示弹窗，可能是因为压缩包较大，下载前计算将消耗一定时间，所以需等待片刻；点击链接无法下载时，尝试刷新页面，重新点击下载链接进行下载。

5.8 本章小结

本章搭建了全天空地基云图数据发布网站。首先介绍了平台的总体设计、使用到的相关技术和平台开发环境。之后介绍了平台的各个模块，主要包括数据展示和数据下载功能，用户可以在数据展示部分查询和查看自己感兴趣的图像和对应拍摄时间下的气象信息和位置信息，数据下载功能则是根据数据量的大小进行了切分，并提供更为人性化的自定义下载功能。最后对平台使用时的注意事项做了一些总结。

结 论

本文完成了全天相机系统的软硬件设计开发并在验证设备可靠性的同时获取了大量的全天空地基云图,基于云图数据完成了云图分类算法的开发和常用云量检测算法的验证,最后开发了用于数据发布的 web 应用平台。本文设计的全天相机系统可以部署在高原、极地、沙漠等候选台址,在无人值守的情况下进行自动观测,结合后续开发的相关算法,将大大降低天文选址的人力支出,有望能够为天文选址工作提供关键支撑,为我国天文事业的发展提供助力。研究的具体内容分为如下几个部分:

(1) 本研究设计研发了一种适用于高温、极寒、高湿度、高腐蚀性环境下的无人值守全天相机系统,完成了设备的硬件设计和软件开发,可应用于天文观测选址工作中全天空地基云图数据的获取。

(2) 将设备搭载在雪龙号极地科考船上以验证其在全球不同极端环境下的可靠性。设备跟随雪龙号完成了第八次北极科考和第三十四次南极科考,经受住了赤道酷暑、极地严寒以及大洋上高湿度高盐分环境的考验,圆满完成了全球不同区域全天空云图的采集任务。该实践证明,设备能够在极端环境下实现无人值守的全天空云图自动采集,将是天文观测选址工作的一件利器。

(3) 本研究获取并创建了雪龙号全天空地基云图数据集,该数据集图像类型丰富,适用于云点识别、云量计算、云高计算和云状识别等方向的研究。

(4) 在本研究中,基于深度学习方法开发了鲁棒性强、可迁移性能好的云图分类算法,从结果上看,卷积神经网络算法在云图分类方面效果优秀,能够很好的区分出不同数据质量和不同时相的云图。之后使用雪龙号全天空地基云图数据集对云量检测的一些常用算法进行验证,比较了它们的优缺点和适用场景。

(5) 基于 web 技术,开发了用于数据发布的雪龙号全天相机数据发布平台,通过平台数据可以快速高效的分享给相关科研工作者,实现了数据的开放共享。该平台具有易于访问、用户友好等特点。

本文虽然围绕全天相机系统完成了很多工作,但是仍有许多不足之处,有待下一步改进。例如:

(1) 虽然云图分类算法取得了较好的分类效果,但是云量检测算法效果不佳,更加准确的云量检测算法是下一步工作的重点。在有太阳出现的图像中,如何区分太阳和云也是未来算法优化的一个重点。

(2) 云高计算和云状识别等方向本文未做研究,但是基于雪龙号全天空地基云图数据集,可进行对应算法的研究和开发。

(3) 雪龙号全天相机数据发布平台现阶段仅包含数据展示和数据下载功能，后续也可将云图分类和云量检测算法嵌入到平台中，提供云图自动化分类、自动云量分析等 web 服务。

参考文献

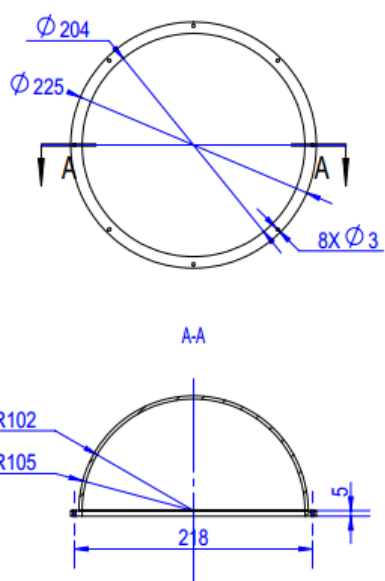
- [1] 尹佳, 姚永强, 王红帅 等. 天文选址的夜间云量处理方法[J]. 天文学报, 2012, 53(3): 230-239.
- [2] 高太长, 刘磊, 赵世军 等. 全天空测云技术现状及进展[J]. 应用气象学报, 2010, 21(1):101-109.
- [3] 张振. 基于深度学习的全天空云图分类方法研究[D]. 2017.
- [4] Shields J E , Karr M E , Tooman T P , et al. The Whole Sky imager-A Year of Progress[J]. Ann R Coll Surg Engl, 1999, 95(1):1-1.
- [5] 霍娟, 吕达仁. 全天空数字相机观测云量的初步研究[J]. 南京气象学院学报, 2002, 25(2):242-246.
- [6] 张阳, 吕伟涛, 马颖. 基于球瓣旋转遮光结构的地基全天空云自动观测系统[J]. 中国专利, 2009.
- [7] 万霞, 吕达仁, 霍娟. 基于可见光全天空图像云型识别的初步研究[C]. 第二十八届中国气象学会年会论文集, 2011:1-14.
- [8] Long C N , Slater D W , Tooman T . Total Sky Imager Model 880 Status and Testing Results[J]. Office of Scientific & Technical Information Technical Reports, 2001.
- [9] 孙学金, 高太长, 翟东力等. 基于非制冷红外焦平面阵列的全天空红外测云系统. 红外与激光工程, 2008, 37(5):761-764.
- [10] 彭焕文, 辛玉新, 和寿圣, 等. 丽江天文观测站全天相机介绍[J]. 天文研究与技术, 2015, 12(1): 89-95.
- [11] 田健峰, 邓李才, 闫正洲, 等. 中国 SONG 项目节点全天云量监测方案[J]. 天文学报, 2016, 57(3):366-375.
- [12] 梅林, 吴翔, 周作超. 深圳市天文台全天云图监测系统的设计与实现[J]. 天文研究与技术, 2018, 15(2): 232-239.
- [13] 杨俊, 吕伟涛, 马颖, 等. 基于局部阈值插值的地基云自动检测方法[J]. 气象学报, 2010, 68(6):1007-1017.
- [14] 漆随, 刘涛, 胡桐, 等. 地基云观测技术及装备研究进展[J]. 山东科学, 2014, 27(6):1-9.
- [15] Lowe D G . Distinctive Image Features from Scale-Invariant Keypoints[J]. IJCV, 2004, 60.
- [16] Dalai N , Triggs B . Histograms of Oriented Gradients for Human Detection. CVPR, 2005:886
- [17] Song T , Xin L , Gao C , et al. Grayscale-Inversion and Rotation Invariant Texture Description Using Sorted Local Gradient Pattern[J]. IEEE Signal Processing Letters, 2018, PP(99):1-1.
- [18] Dencœux T. A k -Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory[M]// Classic Works of the Dempster-Shafer Theory of Belief Functions. Springer Berlin Heidelberg, 2008:804-813.
- [19] Cortes C , Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3):273-297.
- [20] 黄凯奇, 任伟强, 谭铁牛. 图像物体分类与检测算法综述[J]. 计算机学报, 2014, 37(6):1225-1240.
- [21] Lecun Y , Bengio Y , Hinton G . Deep learning[J]. Nature, 2015, 521(7553):436.
- [22] Forsyth C , Graesser A , Jr P I P , et al. Discovering Theoretically Grounded Predictors of

- Shallow vs. Deep- level Learning[J]. 2014.
- [23] Hubel D H , Wiesel T N . Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. *Journal of Physiology*, 1962, 160(1):106-154.
- [24] Lecun Y , Kavukcuoglu K , Clement Farabet . Convolutional Networks and Applications in Vision[C]// *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. IEEE, 2010:253-256.
- [25] Poultney C S , Chopra S , Cun Y L . Efficient learning of sparse representations with an energy-based model[J]. *Advances in neural information processing systems*, 2006:1137-1144.
- [26] Rawat W, Wang Z . Deep Convolutional Neural Networks for Image Classification:A Comprehensive Review[J]. *Neural Computation*, 2017, 29(9):1.
- [27] Hinton G E . Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair[C]// *International Conference on International Conference on Machine Learning*. Omnipress, 2010.
- [28] 高震宇. 基于深度卷积神经网络的图像分类方法研究及应用[D]. 中国科学技术大学. 2018.
- [29] Han J , Moraga C. The influence of the sigmoid function parameters on the speed of backpropagation learning[C]// *International Workshop on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 1995.
- [30] Fan E . Extended tanh-function method and its applications to nonlinear equations[J]. *Physics Letters A*, 2000, 277(4-5):212-218.
- [31] Xu B , Wang N , Chen T , et al. Empirical Evaluation of Rectified Activations in Convolutional Network[J]. *Computer Science*, 2015.
- [32] 王晓斌, 黄金杰, 刘文举. 基于优化卷积神经网络结构的交通标志识别[J]. *计算机应用*, 2017, 37(2):530-534.
- [33] Krizhevsky A , Sutskever I , Hinton G E . ImageNet classification with deep convolutional neural networks[C]//*International Conference on Neural Information Processing Systems*. Curran Associates Inc. 2012:1097-1105.
- [34] Bishop C M . *Pattern Recognition and Machine Learning (Information Science and Statistics)*[M]. Springer-Verlag New York, Inc. 2006.
- [35] Ioffe S , Szegedy C . Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 2015.
- [36] André R. Brodtkorb, Hagen T R , Martin L. Sætra. Graphics processing unit (GPU) programming strategies and trends in GPU computing[J]. *Journal of Parallel and Distributed Computing*, 2013, 73(1):4–13.
- [37] 郑泽宇, 顾思宇. *TensorFlow 实战 Google 深度学习框架*[M]. 北京: 电子工业出版社, 2017.
- [38] 黄文坚, 唐源. *TensorFlow 实战*[M]. 北京: 电子工业出版社, 2017.
- [39] 杨俊, 吕伟涛, 马颖, 等. 基于自适应阈值的地基云自动检测方法[J]. *应用气象学报*, 2009, 20(6):713-721.
- [40] Krishna K , Murty M N . Genetic K-means algorithm[J]. *IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society*, 1999, 29(3):433.
- [41] Otsu N . A threshold selection method from gray-level histograms[J]. *Automatica*, 1975, 11(285-296): 23-27.
- [42] Chun W . *Python Web Development with Django*[M]// *Python Web development with Django*,

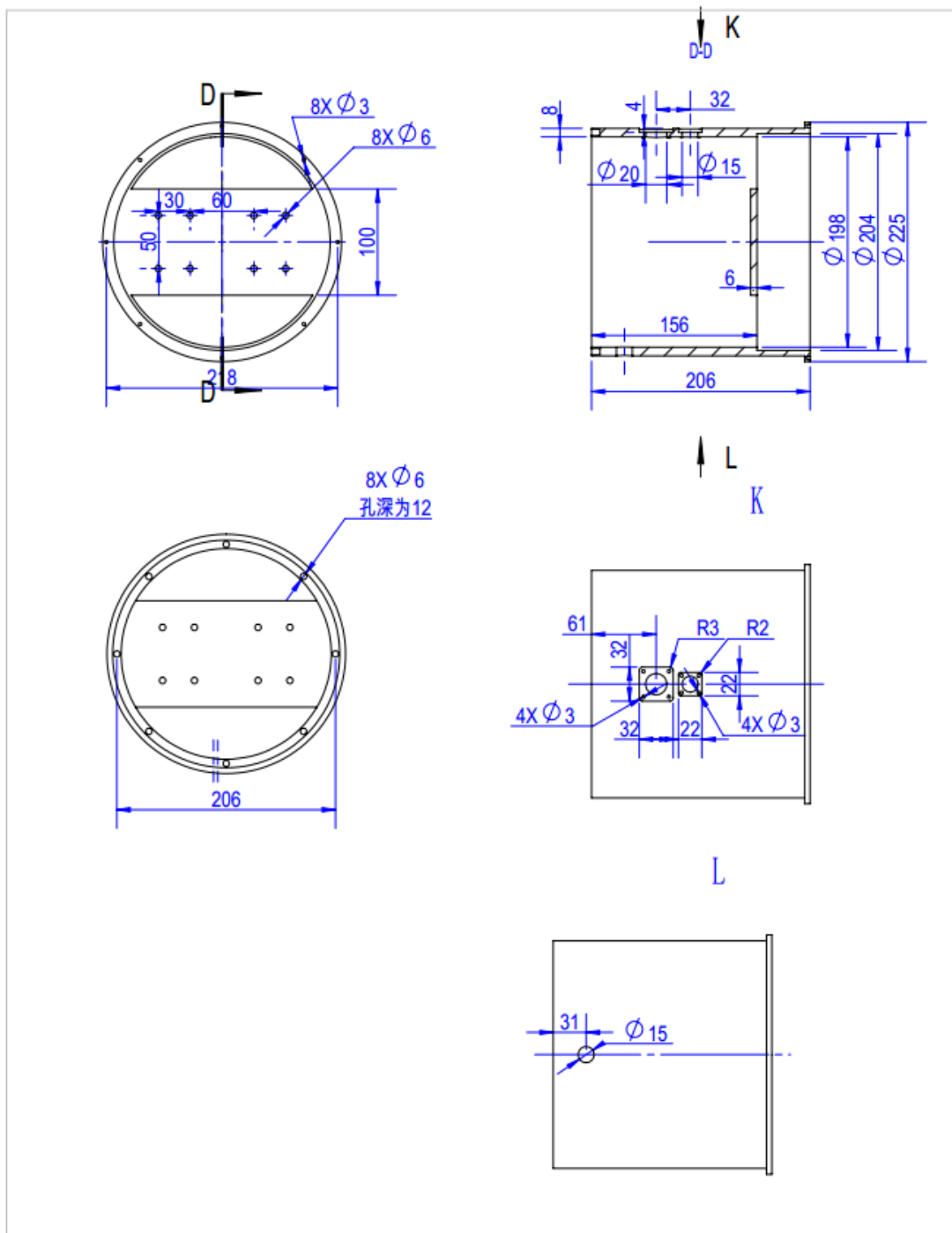
- 2009.
- [43] Borchers P H . Python:a language for computational physics[J]. Computer Physics Communications, 2007, 177(1-2):199-201.
- [44] 陈林. 基于 Django 的英语在线视频教学系统的设计与实现[D]. 厦门大学, 2014.
- [45] 李红娟, 温喆. 基于 J2EE 与 Ajax 的 Web 应用架构设计[J]. 制造业自动化, 2011, 33(9):127-129.
- [46] 周杨. AJAX 应用的典型设计模式[J]. 计算机系统应用, 2011, 20(1):128-132.

附录

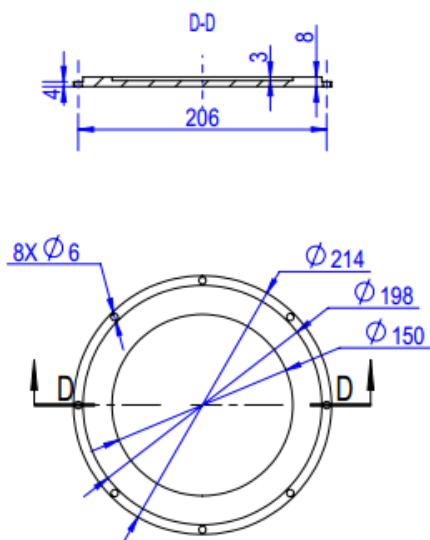
附录 A:



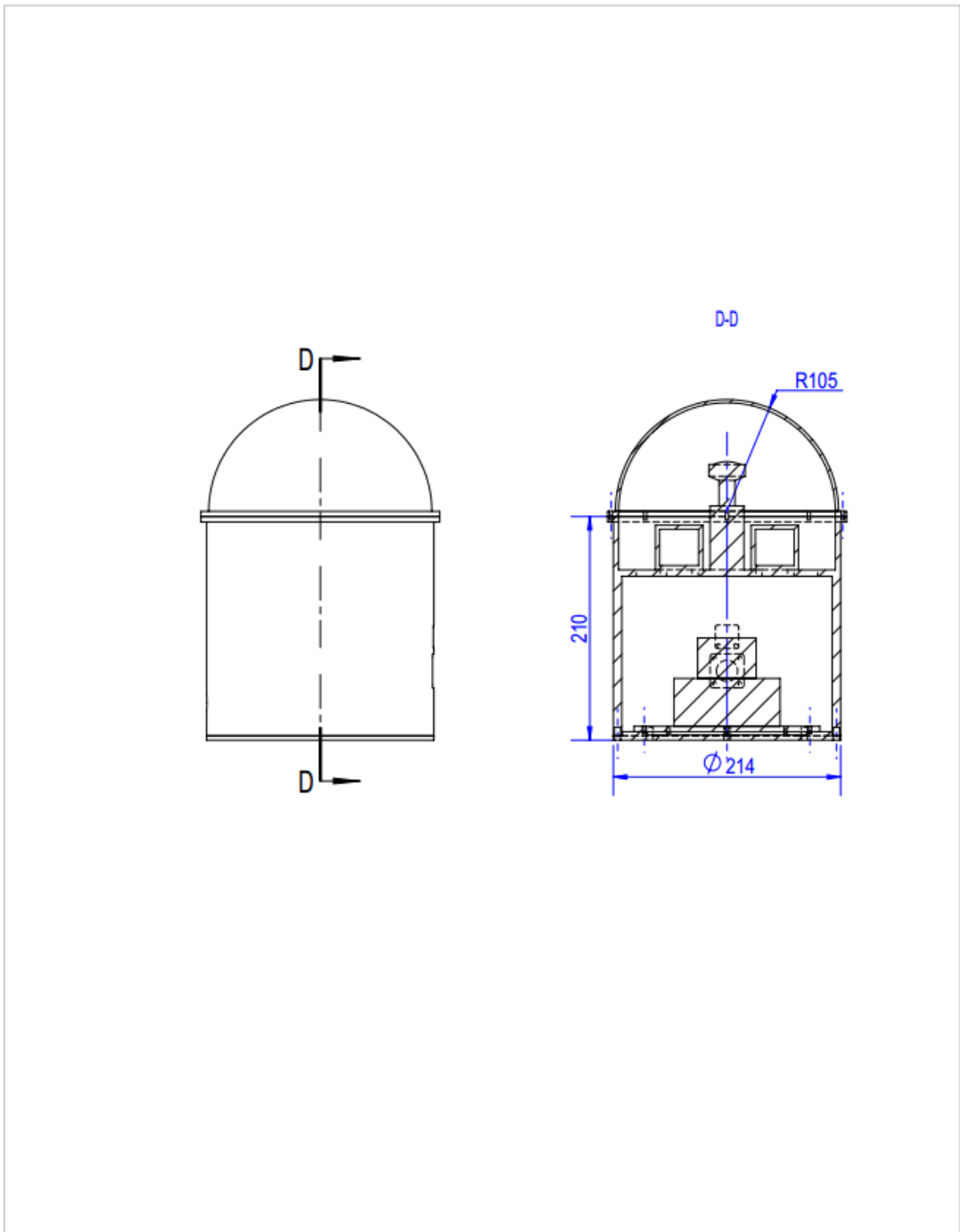
					亚克力半球形透明薄壳					
标记	处数	更改文件号	签名	日期				图样标记	重量	比例
设计			工艺				1:5			
绘图			标准化							
校对			批准							
审核			日期		第 张	共 张				



					全天相机外壳		
标记	处数	更改文件号	签名	日期			1:5
设计			工艺				
绘图			标准化				
校对			批准				
审核			日期				
					第 张	共 张	



					全天相机后盖		
标记	处数	更改文件号	签名	日期	图样标记	重量	比例
设计			工艺				1:5
绘图			标准化				
校对			批准				
审核			日期		第 张	共 张	



					全天相机设备			
标记	处数	更改文件号	签名	日期		图样标记	重量	比例
设计			工艺					1:5
绘图			标准化					
校对			批准					
审核			日期		第 张	共 张		

攻读硕士学位期间发表的学术论文

发表的学术论文：

1 崔顺, 许允飞, 苏丽颖, 崔辰州等.基于卷积神经网络的全天空地基云图分类研究[J].天文研究与技术.

申请的专利：

1 苏丽颖, 崔顺.全天相机数据发布平台 V1.0. 软件著作权.登记号：2018SR836626 ,2018-10-19.

致 谢

三年的时间转瞬即逝，我也即将完成我的研究生学业，回想过去，既有开题时的雄心壮志，也有接触新领域时的彷徨困惑，在我遇到困难时，老师、同学、家人给我的关心鼓励让我有了继续下去的动力，在此我向所有对我给予帮助和关心的人表示衷心的感谢！

首先我要感谢我的导师苏丽颖副教授和校外导师崔辰州研究员，两位恩师的悉心指导和教诲，将使我终生受益，深厚的学术底蕴、严谨的治学态度、丰富的实践经验、求实的学术作风和慈爱的长者风范，将是我终生效仿的典范。尽管工作繁忙，但对我在学习、工作和生活上的关怀、鼓励和鞭策，将会铭记于心。

特别感谢国家天文台的许允飞老师，在课题研究过程中给予了我极大的支持和指导，老师和我亦师亦友的关系也使我在三年的学习中不断感受到关怀和鼓励，在此表示由衷的感谢。

感谢国家天文台的樊东卫老师、韩军老师、陶一寒老师对我学习上的指导和帮助。同时感谢台里李正、韩叙、李珊珊、米琳莹、杨丝丝、李长华、张磊等在学习和生活上的关心和照顾。

感谢余跃庆、高国华、刘婧芳老师在课题方面给予的支持和指导，在课题的研究过程中，学院的领导和老师也给予了热情的指导与帮助，在此表示衷心的感谢和崇高的敬意。

在平时的学习生活中，感谢实验室张双博士、孔凡臣博士、王川中、刘旋、王建、程艳霞、陈振伟、以及师弟张会鹏、师妹魏洪苗等，在我生活及论文撰写过程中提供的帮助和支持。他们营造出的良好学习氛围是我研究得以顺利进行的保障，真诚无私的友谊和团结互助的情谊也助我解除了一些生活上的困难。

向我的室友兼挚友田宝华表示感谢，他永远是我学习的榜样，我们的友谊也将天长地久。

也对审阅此文的所有专家一并表示感谢！

谨以此文的完成来表示家人在研究生学习期间给予我的全力支持，家人的支持是我工作的所有源泉！

最后再次向所有提供无私帮助的人们表示衷心的感谢！

崔顺
2019年5月

