

硕士学位论文

基于卫星图像序列的初生对流检测算法研究

**RESEARCH ON PRIMARY CONVECTION
DETECTION ALGORITHM BASED ON
SATELLITE IMAGE SEQUENCE**

韩佳成

哈尔滨工业大学

2021 年 12 月

国内图书分类号：TM301.2
国际图书分类号：62-5

学校代码：10213
密级：公开

工学硕士学位论文

基于卫星图像序列的初生对流检测算法研究

硕士研究生：韩佳成
导 师：徐晓飞教授
申 请 学 位：工学硕士
学 科：计算机技术
所 在 单 位：计算机科学与技术学院
答 辩 日 期：2021 年 12 月
授 予 学 位 单 位：哈尔滨工业大学

Classified Index: TM301.2

U.D.C: 62-5

Dissertation for the Master's Degree in Engineering

RESEARCH ON PRIMARY CONVECTION DETECTION ALGORITHM BASED ON SATELLITE IMAGE SEQUENCE

Candidate:	Han Jiacheng
Supervisor:	Professor Xu Xiaofei
Associate Supervisor:	Associate Professor Li Xutao
Academic Degree Applied for:	Master of Engineering
Specialty:	Computer Technology
Affiliation:	School of Mechatronics Engineering Mechatronics Engineering
Date of Defence:	December, 2022
Degree-Conferring-Institution:	Harbin Institute of Technology

摘要

预防强对流天气是天气预报的重要课题，也关乎人民人身财产安全。对于预防强对流天气，通常采用的方法是采取精度更高的预报措施，对于已经形成了的对流进行跟踪预测，对于还在形成的对流云进行判别分析。早期的强对流天气预警主要是基于雷达资料，初生对流的定义是在强对流天气发生之前，大气各种物理量处于某种特殊的状态，多普勒天气雷达第一次观测到对流云产生的反射率因子大于等于 $35\text{dBZ}^{[1]}$ 。雷达图像的生成主要依靠雷达回波产生，这就导致只有具有雷达观测站的区域才有能够采集的雷达数据。单纯依靠雷达数据做对流研究预警存在两个问题，一是雷达站的回波区域并不能做到对于全国的全覆盖，因此对于全局对流的形成和移动不能做到最为合理和及时的判断；二是雷达数据本身的局限性，导致其信息不足以对初生对流进行合理判断。

因此我们可以利用卫星云图对对流云进行进一步预测分析。卫星云图具有长时效性，范围广等特点。强对流天气是以大尺度天气系统为背景，大尺度天气系统影响或决定着中小尺度天气系统的生成、发展和移动过程。利用卫星云图可以有效监控大尺度天气系统，从而对初生对流进行有效预测。

关键词：卫星图像；深度学习；天气预报

Abstract

The prevention of severe convective weather is an important topic of weather forecast, and also related to the safety of people's lives and property. For the prevention of severe convective weather, the usual method is to take more accurate forecasting measures, to track and predict the formed convection, and to discriminate and analyze the convective clouds that are still forming. The early warning of severe convective weather is mainly based on radar data. The definition of primary convection is that before the occurrence of severe convective weather, various physical quantities of the atmosphere are in a special state, and the reflectivity factor produced by convective clouds observed by Doppler Weather Radar for the first time is greater than or equal to 35dbz^[1]. The generation of radar image mainly depends on radar echo, which results in radar data acquisition only in the area with radar observation station. There are two problems in the research and early warning of convection based on radar data. One is that the echo area of radar station can not cover the whole country, so the formation and movement of global convection can not be judged reasonably and timely; the other is that the information of radar data itself is limited, which makes it impossible to judge the primary convection reasonably.

Therefore, we can make further prediction and analysis of convective clouds by using satellite cloud images. Satellite cloud image has the characteristics of long time effect and wide range. The strong convective weather is based on the large-scale weather system, which influences or determines the generation, development and movement of the mesoscale weather system. The satellite cloud images can be used to monitor the large-scale weather system effectively, so as to predict the primary convection effectively.

Keywords: Satellite Image, Deep Learning, Weather Forecast

目 录

摘 要	I
Abstract	II
第 1 章 课题的来源及研究的目的和意义	1
1.1 课题来源	1
1.2 研究的目的和意义	1
第 2 章 国内外研究现状及分析	3
2.1 国内外研究现状	3
2.2 国内外文献综述及简析	3
2.2.1 基于自编码器的视频预测方法研究现状	3
第 3 章 基于卫星资料的初生对流像素级检测算法设计	5
3.1 基于卫星资料的初生对流数据集构建	5
3.1.1 融合 DenseConnection 的 RCNN 模型结构	6
结 论	9
参考文献	10
哈尔滨工业大学学位论文原创性声明和使用权限	11

第1章 课题的来源及研究的目的和意义

1.1 课题来源

国家重点研发计划“基于国产卫星的历史数据再构建研究”。

1.2 研究的目的和意义

强对流天气指的是突然发生、强度剧烈，常伴有短时强降水、雷电、大风、冰雹、龙卷等强烈对流性灾害的天气。强对流天气空间尺度不大、生命周期短暂，却能在短时间内释放强大的力量，具有极强的破坏力，是人们不得不预防的灾害性天气。强对流天气时效短暂，危害大，其产生的强降雨往往在较短时间内就可以达到暴雨的量级，是引发城市内涝、山洪爆发以及山体滑坡和泥石流之类的地质灾害的主要源头。其中，短时大风可以达到 8 级以上，破坏树木或房屋；雷电则对电力电信设施和生命财产构成威胁，引发雷击火险和森林火灾；冰雹从 5 毫米到 10 厘米大小不等，最大的可以达到 30 厘米，可直接摧毁地里的庄稼。

强对流是空气强烈的垂直运动而导致出的天气现象。强对流的形成主要是因为云团受到的太阳辐射导致的云团热胀，密度减小，从而发生向上的垂直运动，在运动过程中气温逐渐降低，空气中包含的水蒸气凝结成为水滴，水滴在下降过程中又会被其他上升气流携升，如此反复最后积累成大水滴直至高空气流无力支持其重量，最后下降成雨。

预防强对流天气是天气预报的重要课题，也关乎人民人身财产安全。对于预防强对流天气，通常采用的方法是采取精度更高的预报措施，对于已经形成了的对流进行跟踪预测，对于还在形成的对流云进行判别分析。早期的强对流天气预警主要是基于雷达资料，初生对流的定义是在强对流天气发生之前，大气各种物理量处于某种特殊的状态，多普勒天气雷达第一次观测到对流云产生的反射率因子大于等于 35dBZ[1]。雷达图像的生成主要依靠雷达回波产生，这就导致只有具有雷达观测站的区域才有能够采集的雷达数据。单纯依靠雷达数据做对流研究预警存在两个问题，一是雷达站的回波区域并不能做到对于全国的全覆盖，因此对于全局对流的形成和移动不能做到最为合理和及时的判断；二是雷达数据本身的局限性，导致其信息不足以对初生对流进行合理判断。

因此我们可以利用卫星云图对对流云进行进一步预测分析。卫星云图具有长

时效性，范围广等特点。强对流天气是以大尺度天气系统为背景，大尺度天气系统影响或决定着中小尺度天气系统的生成、发展和移动过程。利用卫星云图可以有效监控大尺度天气系统，从而对初生对流进行有效预测。

强对流天气严重威胁人民生命财产安全，如果能够有效监测初生对流，那么将会大幅提高对流的预警时效，减少强对流天气带来的损失。

随着计算机算力的提升，利用深度学习结合卫星数据对于解决初生对流的检测的问题带来了新的解决方案。卫星数据具有很长的时序性，数据量充足，国产静止卫星监控的范围辽阔，分辨率高，足够作深度学习的数据集来源。除此之外，我国卫星数据还具有较多的历史信息，通过对于历史数据的信息的充分挖掘，可以得到对流的初生、移动、变化、消亡等特征。因此可以针对卫星数据做时序图像检测或预测。

对卫星数据做初生对流检测研究，初生对流检测与强对流检测不同，此时的云团虽然内部有一定变化，但是利用通常的阈值法不能很好的进行检测，对流云在检测过程中也会出现将卷云误判为对流的情况，初生对流表现在卫星图像上利用传统的方法很难取得很好的效果。所以要依靠深度学习的方法深度挖掘对流在形成前的云的特征，在对流形成之前将初生对流检测出来。有对初生对流的基本判断下，可以达到更为及时的强对流预警。

在实现初生对流的检测，需要对卫星云图做最基本的对流的检测及预测，卫星云图本身就带有对流的基本信息，通过对流目标检测的方式将对流检测出来，通过上一帧中的对流信息及图形信息，预测下一帧中的对流信息，从而得到这一帧中的初生对流可能存在的区域。

第2章 国内外研究现状及分析

2.1 国内外研究现状

国内外研究普遍采用的初生对流的定义是多普勒天气雷达首次检测到对流云反射率回波大于等于 35dB。基于卫星资料的初生对流检测算法主要采用多普勒天气雷达检测的初生对流作为准则。利用卫星资料的初生对流检测方法主要采用的是传统的多通道阈值方法，通过对于不同波长的红外亮温通道的卫星图像进行判断。初生对流的检测的前提需要时序的对流检测以及对对流的跟踪。

基于卫星图像的传统对流识别方法主要是采用的阈值法^[2]，阈值法是最简单最常用的图像分割方法，通常适用于目标和背景差别明显的特征判识。对于卫星图像的长波红外通道，判定对流的采用的是红外亮温阈值，早期采用的方法有温度阈值、面积阈值和温度梯度阈值等。阈值法最大的问题是阈值的划定并不存在一个统一的标准，收到季节、海拔和经纬度的影响，不同地域的不同时间的阈值划分也应有所差异。从应用情况看，国内外相关的研究所采用的阈值通常在 240K-250K 之间。

基于卫星图像序列的传统对流目标的跟踪主要包括面积重叠法和统计法^[3]。传统方法首先采用阈值法识别对流云团，根据亮温值连续低于某阈值进行判别利用云团轮廓码保存对流云团边界。采用多种阈值保存不同阈值下标注的对流云团结果。面积重叠法是假定对流云团不发生形变和旋转的前提下，利用连续两个时刻的对流云团边界计算对流云团的运动矢量。相重叠的面积越大，表明同一对流云团的可能性越高，再利用线性外推对对流云团进行预测，通过统计法的方法对生成的对流云团进行修正。

除了较为成熟的传统方法，对流外推本质是视频预测问题，初生对流的检测可以看作图像分割问题。图像语义分割方法以及视频预测方面的方法也能够对初生对流的检测给予很多启发性的方法。

2.2 国内外文献综述及简析

2.2.1 基于自编码器的视频预测方法研究现状

自编码器是一种以无监督的方式来学习数据表征的神经网络，通常用来做数

据降维。自编码器通常分为编码器和解码器两部分，编码器将数据编码为潜在变量，解码器将潜在变量重建为原数据。自编码器有很多变体，例如降噪自编码器、稀疏自编码器、变分自编码器 (VAE)。因为自编码器可以高效地进行数据降维，相当一部分视频预测模型采用了自编码器架构，如图 2-1 所示。

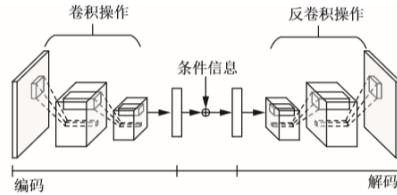


图 2-1 基于自编码器的视频预测模型架构

自编码器架构下表现较好的是 ConvLSTM，与 LSTM 不同，ConvLSTM 在计算过程中用卷积操作代替了矩阵乘法，如公式 2-1 所示，很好的利用卷积提取图像中的空间信息和利用 LSTM 提取时序信息。后来在 ConvLSTM 的基础上，又提出了训练代价更低，参数量更小的 ConvGRU，在参数量和训练时间大大减少的前提下，精度能够得到基本保障。

$$\begin{aligned}
 i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\
 f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\
 C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\
 o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{oi} \circ C_{t-1} + b_o) \\
 H_t &= o_t \circ \tanh(C_t)
 \end{aligned} \tag{2-1}$$

第3章 基于卫星资料的初生对流像素级检测算法设计

3.1 基于卫星资料的初生对流数据集构建

初生对流 (CI) 的物理定义为强对流天气发生前，大气各种物理量处于某种特殊的状态，换句话说初生对流的未来发展趋势应该是强对流。基于卫星资料的初生对流的标注工作，主要采用带有多通道时序的阈值法，通过当前时刻与未来 15 分钟和 30 分钟（即两帧图像）的卫星红外 1 (10.8um, IR1)、红外 2 (12um, IR2) 与水汽 (6.25um,WV) 三个通道的信息，来标注初生对流的位置。其中共有 8 个阈值，其中 F1 到 F8 的分别为用以上信息设置的阈值限定指标，其中的物理意义代表了不同尺度下的云顶高度、云顶高度变化以及积云形态。其中满足任意 7 个阈值的像元，将会被标注为初生对流。

其中标注的阈值如下表 3.1 所示。

表 3-1 初生对流阈值指标

序号	CI 预报指标	预报指标的判断阈值
F1	IR1 亮温差	< 0
F2	WV 与 IR1 亮温差	35 ~ 10
F3	IR2 与 IR1 亮温差	25 ~ 5
F4	IR1 亮温变化率	< -4/(15 min)
F5	IR1 亮温变化率	$T/(30 \text{ min}) < T/(15 \text{ min})$
F6	IR1 亮温降低至 0 的时间	< 30 min
F7	WV 与 IR1 亮温差变化率	> 3/(15 min)
F8	IR2 与 IR1 亮温差变化率	> 2/(15 min)

通过此多通道时序的阈值法，构造了 2018 年 FY-4a 卫星全国范围内的初生对流数据集，选取了其中对流出现较为频繁的 6-10 月的华南地区作为此项课题的数据集。数据集的输入数据为当前时刻 t 的卫星数据以及当前时刻前 3 帧的卫星数据序列，即 $t-15, t-30, t-45$ 。数据集的输出数据为当前时刻 t 为基准标注的初生对流掩码图像。其中 t 时刻的初生对流掩码是利用上述多通道阈值法通过 t 时刻， $t+15$ 时刻，以及 $t+30$ 时刻标注出来的。以 2018 年 7 月 15 日为例，数据集一条数据可视化的效果如 3.1 所示。

表 3-2 多通道云图序列数据展示图

通道	t-45	t-30	t-15	t
水汽通道 WV				
红外通道 IR1				
红外通道 IR2				

表 3-3 初生对流可视化类别标注即掩码展示图

可视化类别标注结果	类别标注掩码

经过统计分析，每张类别标注图像中存在初生对流的平均比率为 0.79%，平均每张图像存在 7457 个初生对流样本像元。可见其中的类别比例严重失衡。

3.1.1 融合 DenseConnection 的 RCNN 模型结构

对于像素级分类这一问题，本课题采用了 RCNN 模型作为基础模型进行初生对流的预测工作。初生对流多存在于云团系的边缘附近，因此要判识初生对流，模型至少需要一个像元附近的周围的像元来判别是否在云团边缘，从而更好的判识初生对流。一般初生对流所联通的区域很小，通常仅仅是一个或者几个像元相联通，而其中一个像元单独存在且不连通的样本占绝大多数。因此本课题所采用的模型是能够擅长处理像素级分类的循环卷积神经网络 RCNN。

RCNN 模型是逐步缩小检测范围的多层卷积神经网络结构，相比 CNN 能够

复用不同尺度特征，有效提取像元的空间信息。RCNN 模型相比于语义分割模型（Unet、Deeplab）等对于处理单个像元分类的任务要更加优秀，语义分割模型更擅长处理大范围联通样本分类问题。

基础的 RCNN 模型如图 3-1 所示：

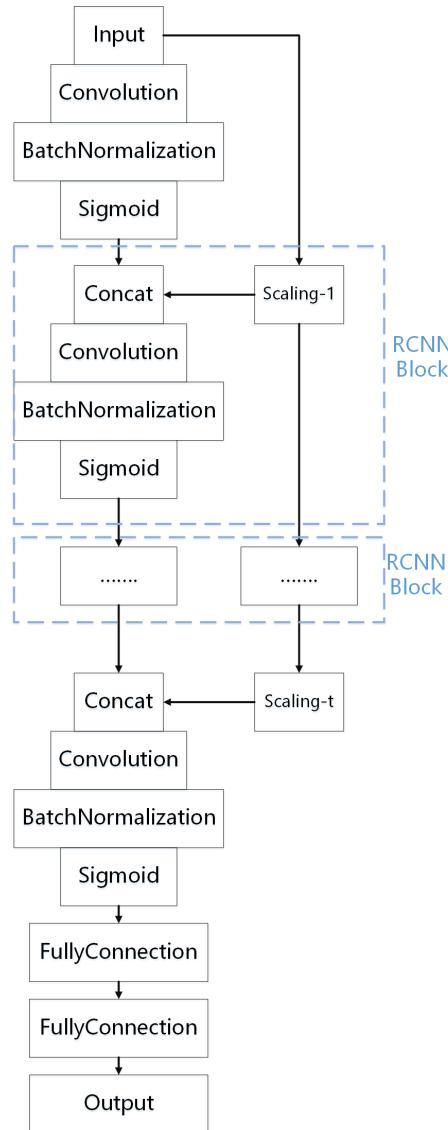


图 3-1 RCNN 模型框架

RCNN 模型中基本的结构为 RCNN Block，由一次 2D 卷积一次批标准化以及一层激活函数组成。RCNN Block 每层所接受的输入为上一层的结果以及原始图像缩放后经过通道连接的张量。由 3-2，可见其中对于越靠近像元中心的像元，其 RCNN 的模型对其的附加的权重就越大，然而真实情况下单独依靠像元中心并不能判识初生对流，甚至无法判识云团边缘。不同层级的 RCNN 卷积核用于提取不同尺度

的特征，随着层数加深卷积核所提取的特征所概况的空间信息越少，而很多情况下初生对流所依赖的特征所属于较初始的低阶特征，因此这里引入 DenseConnection 结构，每次卷积都保留上一层所有的特征图，对特征依次传递保留，如 3-3 所示。

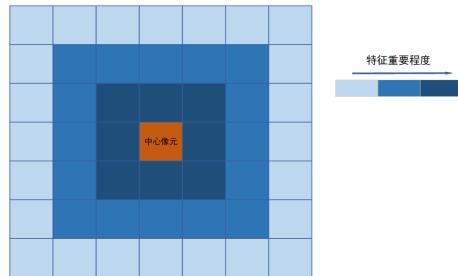


图 3-2 特征重要程度

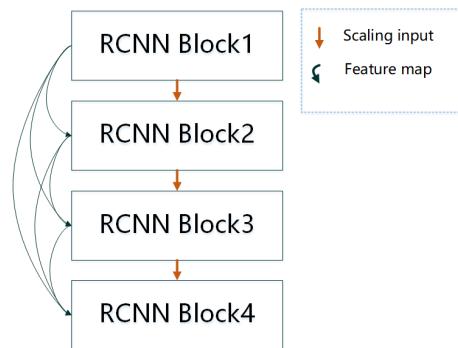


图 3-3 Dense-RCNN 连接结构

结 论

本文算法效果如表。

参考文献

- [1] Mecikalski J R, Bedka K M. Forecasting Convective Initiation by Monitoring the Evolution of Moving Cumulus in Daytime GOES Imagery[J]. Monthly Weather Review, 2006, 134(1) : 49-78.
- [2] 覃丹宇, 方宗义. 利用静止气象卫星监测初生对流的研究进展[J]. 气象, 2014(1) : 9-19.
- [3] 张春桂, 周乐照, 林炳青. 基于卫星和雷达资料的对流云团识别跟踪 and Tracking of Convective Cloud Clusters Based on Satellite and Radar Data[J]. 气象科技, 2017, 045(3) : 485-491.

哈尔滨工业大学学位论文原创性声明和使用权限

学位论文原创性声明

本人郑重声明：此处所提交的学位论文《基于卫星图像序列的初生对流检测算法研究》，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名：_____ 日期：_____ 年 _____ 月 _____ 日

学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

(1) 学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；(2) 学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；(3) 研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名单位为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：_____ 日期：_____ 年 _____ 月 _____ 日

导师签名：_____ 日期：_____ 年 _____ 月 _____ 日